LOAN DOCUMENT

PHOTOGRAPH THIS SHEET

AD-A261 993

DTIC ACCESSION NUMBER

LEVEL

1

INVENTORY

AFOSR-TR-93-0115

DOCUMENT IDENTIFICATION

Dec 92

DISTRIBUTION STATEMENT A

Approved for public release.
Distribution Unlimited

DISTRIBUTION STATEMENT

HANDLE WITH CARE

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 1 0 1993
S C D

DATE ACCESSIONED

DTIC QUALITY INSPECTED 1

DATE RETURNED

93 3 10 004

~~93 3 4 078~~

DATE RECEIVED IN DTIC

93-05122

REGISTERED OR CERTIFIED NUMBER

PHOTOGRAPH THIS SHEET AND RETURN TO DTIC-FDAC

DTIC FORM 70A JUN 90

DOCUMENT PROCESSING SHEET

PREVIOUS EDITIONS MAY BE USED UNTIL STOCK IS EXHAUSTED.

LOAN DOCUMENT

UNITED STATES AIR FORCE

SUMMER RESEARCH PROGRAM -- 1992

SUMMER FACULTY RESEARCH PROGRAM (SFRP) REPORTS

VOLUME 5B

WRIGHT LABORATORY

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	3. REPORT TYPE AND DATES COVERED
	28 Dec 92	Annual 1 Sep 91 - 31 Aug 92

4. TITLE AND SUBTITLE

1992 Summer Faculty Research Program (SFRP)
Volumes 1 - 16 V-5B

5. FUNDING NUMBERS

F49620-90-C-0076

6. AUTHOR(S)

Mr Gary Moore

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

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

AFOSR-TR 93 0115

9. SPONSORING/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

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

UNLIMITED

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The purpose of this program is to develop the basis for cintinuing research of interest to the Air Force at the institution of the faculty member; to stiumlate 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.

14. SUBJECT TERMS

15. NUMBER OF PAGES

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	**UNCLASSIFIED**	**UNCLASSIFIED**	**UL**

NSN 7540-01-280-5500

UNITED STATES AIR FORCE

SUMMER RESEARCH PROGRAM -- 1992

SUMMER FACULTY RESEARCH PROGRAM (SFRP) REPORTS

VOLUME 5B

WRIGHT LABORATORY

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 Air Force Office of Scientific Research (AFOSR) Summer Research Program. The current volume, Volume 5B of 16, presents part two of the final research reports of faculty (SFRP) participants at Wright Laboratory.

Reports presented herein 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

Wright Laboratory

Wright Laboratory (cont'd)

Wright Laboratory (cont'd)

Wright Laboratory (cont'd)

ENHANCEMENT OF THE TIME RESPONSE OF LINEAR CONTROL VIA FUZZY LOGIC AND NONLINEAR CONTROL

Charles E. Hall, Jr.
Assistant Professor
Department of Mechanical and Aerospace Engineering

North Carolina State University
Raleigh, NC 27695-7910

Final Report for:
Summer Research Program
Wright Laboratory

Sponsored by:
Air Force Office Of Scientific Research
Bolling Air Force Base, Washingron, D.C.

July, 1992

ENHANCEMENT OF THE TIME RESPONSE OF LINEAR CONTROL SYSTEMS VIA FUZZY LOGIC AND NONLINEAR CONTROL

Charles E. Hall, Jr.
Assistant Professor
Department of Mechanical and Aerospace Engineering
North Carolina State University

ABSTRACT

The use of linear feedback, of either outputs or state variables, is the standard technique for control of dynamical systems. In many problems the gain of the feedback is a constant that was determined by analysis of the behavior of the dynamics. While usually this yields acceptable results for many systems it can produce unacceptable results, due to either large disturbances or the invalidity of the linear approximation, and thus gain scheduling is used in order to improve system performance. The techniques of fuzzy logic control and nonlinear control offer methods of increasing the system's performance without the need of gain scheduling. Fuzzy logic is a linguistic based system and accordingly it is much simpler and faster to implement than standard control techniques. It was found that fuzzy logic controllers produce deterministic maps and thus can be reduced off-line to these deterministic maps for real time implementation in a control system. Fuzzy logic control systems are not as linguistic based as the proponents contend. There is a large numerical base that needs to be established for each system, which is more complex than normal control techniques. Analysis and simulations involving two types of fuzzy logic based systems were run and compared to similar linear systems. Further examination of fuzzy logic systems were terminated since they were not producing desirable results. Nonlinear control techniques were examined for an improved linear system response. It was found that a linear and cubic operation on the error signal to a first order scalar differential equation improved the system's performance. This technique was not, by itself, applicable to higher order systems, but since the systems of interest were for the control of aircraft, this nonlinear technique was applied to the elevator servo for control of the short-period longitudinal mode of the Lambda URV. This method did provide desirable results.

ENHANCEMENT OF THE TIME RESPONSE OF LINEAR CONTROL SYSTEMS VIA FUZZY LOGIC AND NONLINEAR CONTROL

Charles E. Hall, Jr.

INTRODUCTION

The use of linear feedback, of either outputs or state variables, is the standard technique for control of dynamical systems. In many problems the gain of the feedback is a constant that was determined by analysis of the behavior of the dynamics. While usually this yields acceptable results for many systems this produces unacceptable results, due to either large disturbances or the invalidity of the linear approximation, and thus gain scheduling is used in order to improve system performance. The techniques of fuzzy logic control and nonlinear control offer methods of increasing the system's performance without the need of gain scheduling. This paper examines both of these techniques for the enhancement of the performance of linear systems.

Section 1 is a review of current fuzzy logic control theory, and the basic concepts of fuzzy logic control are presented. Section 2 applies fuzzy logic control to two linear control systems, a Type 0 and a Type 1 system. Section 3 develops two stratagems of nonlinear feedback control with applications to the linear systems presented in Section 2. Section 4 presents an application of one of the nonlinear feedback controllers to the Lambda Unmanned Research Vehicle. Conclusions are presented in Section 5.

1. FUZZY LOGIC CONTROL

The mathematical basis of Fuzzy Logic will not be presented here as there is sufficient background material availible [1-5]. The application of Fuzzy Logic to form Fuzzy Logic Control (FLC) is reviewed [6-7]. The concepts of FLC are understandable without a thorough understanding of the mathematical basis of Fuzzy Logic. This will present no problem to the reader.

The main advantage of FLC touted by its proponents is that it is a linguistic rule based system and thus it does not require the elaborate analysis in generating a feedback control law. The linguistic rule base is a common sense approach to controlling the system and thus knowledge of the dynamics of the system is

not required. The linguistic rule base is represented by the following rule

IF low THEN go up.

While this linguistic base is heavily emphasized in the literature, the fuzzification and defuzzification, which requires numeric values, is deemphasized [3-7]. Analysis of the process of fuzzification/defuzzification and the linguistic rule base interaction results in a deterministic system. This deterministic formulation of a FLC system is further examined.

The basis of FLC is the ability of a object to have a partial membership to a set. Given an n-dimensional vector (which represents the input to the FLC), $\mathbf{x}$, and a set of sets, $A = \{A_i | i = 1 \text{ to } k\}$ the measure of $\mathbf{x}$ being a member of a particular set A_i is represented by a_i.

$$a_i = \mu_{A_i}(\mathbf{x})$$

Each i is associated with the i'th rule of the rule base. Where a_i has allowable values in the range of [0,1]. The variable, a_i, is termed a fuzzy variable, and the measurement of membership is reffered to as fuzzification of the variable $\mathbf{x}$. It is easily shown that the k-dimensional vector, $\mathbf{A}$, formed from the a_i is the result of a deterministic map of the n-dimensional vector onto the space $[0, 1]^k$.

$$\mu_A : \mathbf{x} \rightarrow [0, 1]^k$$

Fuzzification is a unique process once decided on, but there are an infinite number of possible shapes to the sets A_i. Typical shapes are shown in Fig 1, which for clarity is shown for a scalar input $\mathbf{x}$. The shapes of the sets A_i perform the same function as window functions that are common to signal processing [8].

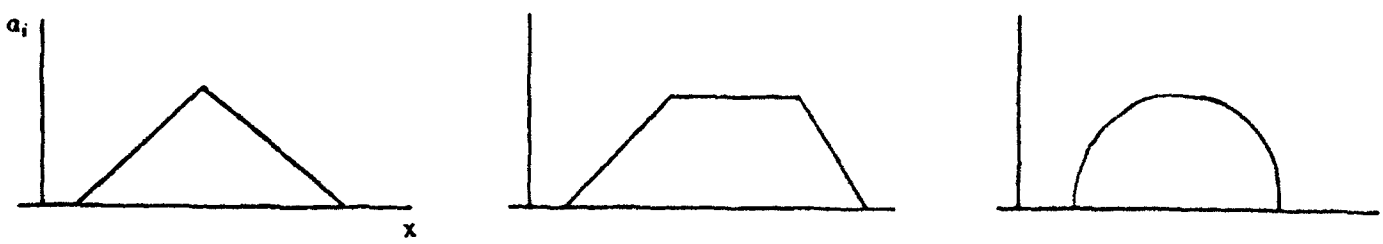


Fig 1. Three possible set shapes.

If the window functions are constructed of a sequence of line segments, it is easy to show that there is a function that maps the n-dimensional vector $\mathbf{x}$ into a scalar and that this map has a structure similar to that of the equation of a line.

$$a_i = \mu_{A_i}(\mathbf{x}) = \mathbf{m}_i \cdot \mathbf{x} + \beta_i$$

The operator, $\cdot$ is the normal inner product operator. Where $\mathbf{m}_i$ is a vector of the local slopes, and β_i is a vector based on the intercepts. The vectors $\mathbf{m}_i$ and β_i are piecewise constant function of $\mathbf{x}$. This form of fuzzification of the variable $\mathbf{x}$ will be used in the remainder of this paper, unless otherwise noted.

Let $\mathbf{B}$ be a k-dimensional fuzzy variable of the m-dimensional vector $\mathbf{y}$. As with the inputs to the fuzzy system, $b_j = \mu_{B_j}(\mathbf{y})$. The fuzzy vector, $\mathbf{B}$ represents k outputs of the linguistic rule base and thus are related to the fuzzy inputs, $\mathbf{A}$, through the rule base. The linguistic rule base is represented by a matrix.

$$\mathbf{B} = M \circ \mathbf{A}$$

Where M is a $k \times k$ matrix. The $\circ$ operator is a type of matrix multiplication that is accomplished by a max-min operation, *which in the literature is called fuzzy matrix multiplication.*

$$b_j = \max_{(1 \le i \le k)} \min(a_i, m_{ij})$$

Since the various m_{ij} are constants, they can be represented by an equation of a line, a line of constant value. It thus follows that the b_j are given by.

$$b_j = \mathbf{n}_j \cdot \mathbf{x} + \gamma_j$$

Where the $\mathbf{n}_j$ and γ_j are piecewise constant functions of $\mathbf{x}$. This is another mapping of $\mathbf{x}$ onto a scalar space with the locus of the map being line segments.

Defuzzification is the process of mapping the fuzzy vector, $\mathbf{B}$, to the vector $\mathbf{y}$. As with the numerous fuzzification methods, ther are many methods of defuzzification. Two of these methods were examined, Tsukamoto's method and Fuzzy Centroid Defuzzification.

Tsukamoto's method of defuzzification is the simplest and easiest applied. Tsukamoto's method is a weighted sum of the fuzzy variables. For the j'th rule of the rule base there is an associated output vector,

$\mathbf{y}_j$. Thus to defuzzify the output, the b_j is the weight applied to the rule associated output vector $\mathbf{y}_j$. Tsukamoto's method is given by the equation.

$$\mathbf{y} = \frac{\sum_{j=1}^{k} b_j \mathbf{y}_j}{\sum_{j=1}^{k} b_j}$$

A substitution for the b_j can be made yielding.

$$\mathbf{y} = \frac{\sum_{j=1}^{k} \mathbf{n}_j \cdot \mathbf{x}\mathbf{y}_j + \gamma_j \mathbf{y}_j}{\sum_{j=1}^{k} \mathbf{n}_j \cdot \mathbf{x} + \gamma_j}$$

If either; 1)Only one rule is activated, or 2)The coefficient for each component of $\mathbf{x}$, x_i, in the demoninator is zero, $\sum_{j=1}^{k} \mathbf{n}_j \cdot \hat{\mathbf{e}}_i = 0$ then the resultant map of the overall fuzzy system is piecewise linear in $\mathbf{x}$.

$$\mathbf{y} = K\mathbf{x} + \mathbf{k}_0$$

Where K is a gain matrix premultiplying the vector $\mathbf{x}$, in the usual sense of vector multiplication, and $\mathbf{k}_0$ is an offset vector. The gain matrix and offset vector are piecewise constant functions of the variable $\mathbf{x}$ and they can be calculated off-line for a given fuzzy control system. A typical FLC, with Tsukamoto's defuzzification which results in a piecewise linear system, output verses the input is shown in Fig 2.

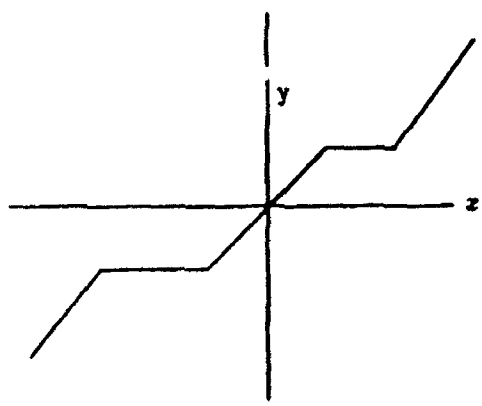


Fig 2. Outout of a Tsukamoto FLC.

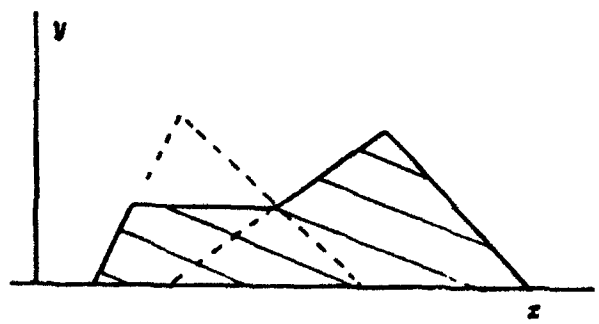


Fig 3. Fuzzy Centroid Defuzzification.

The fuzzy centroid defuzzification(FCD) is based on the centroid of the area of the fuzzy output. Fig 3 shows the area in which two rules are activatied by the input to the FLC. The shapes for the particular $\mathbf{y}_j$

for Fig 3 were choosen to be triangular, but this need not be the case. The shaded area is the area of which the centroid is to be calculated. The clipping that is observed on the first triangle is due to the input for that particular rule. For a scalar input and output to the FLC yields the overall result of the form of the following equation where x is the input to the FLC.

$$y = \frac{n_3 x^3 + n_2 x^2 + n_1 x + n_0}{d_2 x^2 + d_1 x + d_0}$$

Where the coefficients of the numerator and demoninator are solely determined by the shape and values of the particular input and output set A and B. For triangular input and output sets the coefficients are complicated and will not be further expanded in this paper, but they can be calculated offline. In the general for an output window whose shape is of order, ν, FCD yields an input-output map of the form.

$$y = \frac{\sum_{i=0}^{\nu+2} n_i x^i}{\sum_{i=0}^{\nu+1} d_i x^i}$$

For triangular output windows $\nu = 1$, since the window is formed of line segments. Due to the complicated structure of the coeficients it would be best to calculate them with a package such as MACSYMA. Fig 4 shows the function of y verses x of a FLC using centroid defuzzification method with scalar input and scalar output to the controller and triangular output windows. The curve is between adjacent peaks of the output sets.

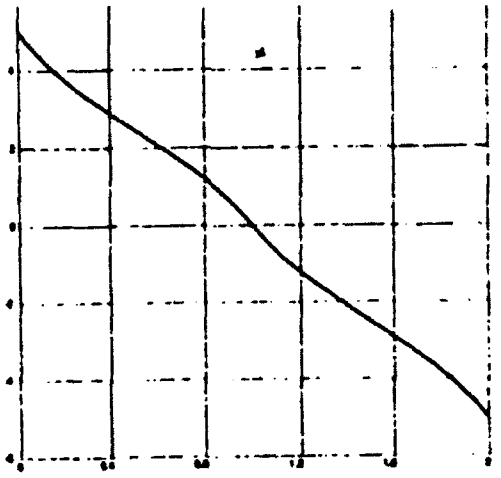


Fig 4. Output of Fuzzy Centroid Method.

While only these two methods of defuzzification were examined it, is expected that the other methods will also yield a deterministic map between the FLC inputs and outputs. While the scalar case was primarily

examined, expansion to the case of the multiple inputs and outputs is straight forward. With the reduction of the FLC with either defuzzification via Tsukamoto's method or centroid method to a deterministic map the FLC can be implemented with the respective maps in an IF-THEN-ELSE structure with identical results.

The major fallacy of Fuzzy Logic Controllers is that it is a linguistic rule based system. From the example rule one must ask how is it encoded into the M matrix, since the $m_{i,j}$ are numeric values. The fuzzy logic proponents further qualify the rule base by adjectives such as: small, large, and very large. But what is large low or large up? The literature that was examined did not include information on the assignment of numeric values to the linguistic rules. If anything was mentioned about the assignment of numeric values, it was usually done in a single sentence that related how it took a large amount of computer time, performing simulations, to assign the numeric values to the fuzzy system in order to obtain an acceptable level of system performance. It is easy to see that once the rule base is set, the shape of the window functions for both inputs and the outputs leave a large number of variables to modify. The shape of these window functions which directly affect the peformance of the system and shape of these window functions are vastly more complex than one or two feedback gain constants.

While the determination of the fuzzy system, for acceptable system performance, is a formidable task it does have it merits and therefore it has the ability to assist the control engineer. Use of the linguistic rule base can, in many complicated systems, clarify the partitioning of the output or state space into regions for application of standard control techniques, such as gain scheduling. The actual numeric values that are used in the partitioning of the space would still need to be determined, but an initial use of easily determined values or generic values would give the control engineer an initial concept on how to approach the problem. This would be best accomplished by the window functions, for both input and output functions, to be restricted to triangular or trapezoidal in shape. The task could be further simplied by a compiler for the rule base. This compiler could be written in a language such as "C". The compiler would operate on a file containing the rule base and the necessary information of the space and it would generate the partitioned space. The control law generation should not include many of the concepts of Fuzzy Logic Control, but should be based on the more standard control techniques.

2. FLC APPLICATION TO TYPE 0 AND 1 LINEAR SYSTEMS

Two linear systems were used in simulation of FLC. These linear systems were of Type 0 and Type 1 structure and respresented by the following equations.

$$G_0(s) = \frac{1}{(s+1)(s+3)(s+4)}$$

$$G_1(s) = \frac{1}{s(s+3)(s+4)}$$

These systems were chosen for their easy analysis. Two FLC's were designed, one employed Tsukamoto's defuzzification method and the other system used centroid defuzzification. These were made comparable to a linear feedback with a gain of 10. The FLC systems were run, one at a time, in parallel with the linear system in order to compare the results of the two systems. All simulations were run with on the EASY5 simulation and analysis package[8]. Fig 5 shows the block diagram of the simulation along with blocks to calculate the comparisons. Both the linear and FLC systems were started with the same initial conditions and given the identical inputs. In the simulations in which noise was injected into the systems, there was only one noise generator which was fed into both systems as shown in Fig 5. For the simulations that did not include noise, the summer junctions that inserted the noise into the feedback path from the output were set such that the coefficient for the noise signal was zero.

The FLC using Tsukamoto's method of defuzzification was designed so that it yielded a linear system that was identical with the reference linear system, the results of the simulations will not be presented here. But it was used primarily for verification of the fuzzification and rule implementation. Simulations were then performed using the centroid method of defuzzification.

Fig 6 and 7 are typical results for the simulations conducted with a Type 0 linear system. The sign difference on the fuzzy and linear controllers is compensated by summing junction coefficient. There was no noise injected into the simulations results presented in Fig 6 and 7. The input to the systems for this particular simulation was a step function selected such that the output of the purely linear system was 1. It should be noted that the steady state error of the system employing fuzzy logic control was a function of x.

If the overall gain of the FLC was modified to give an error similar to the strictly linear system, the steady state error would be different for other values of the inputs to the system.

One way of eliminating the steady state error was to use a Type 1 linear system in the simulations, which are shown in Fig 8 and 9. Noise was injected into these systems. The noise had a standard deviation of 0.05. The input for this particular simulation was 10. The purely linear system had, as expected, a zero steady state error. The system with the FLC did not possess a zero steady state error, even when no noise was present. This was due to nonlinearities inherent in the FLC. While initially this appears confusing, it should be noted that the output of the FLC is -10 which when added to the input results in a zero error signals.

During the simulations of the FLC and comparison linear system it was noted about the amount of time that was required for a run. Two purely linear systems were run and the time noted. The time was only a rough measure. The FLC system took about 2.5 times longer to execute that the two linear systems. This was due to the numeric intensive centroid defuzzification. There are several fuzzy logic engines on the market which would decrease the time required.

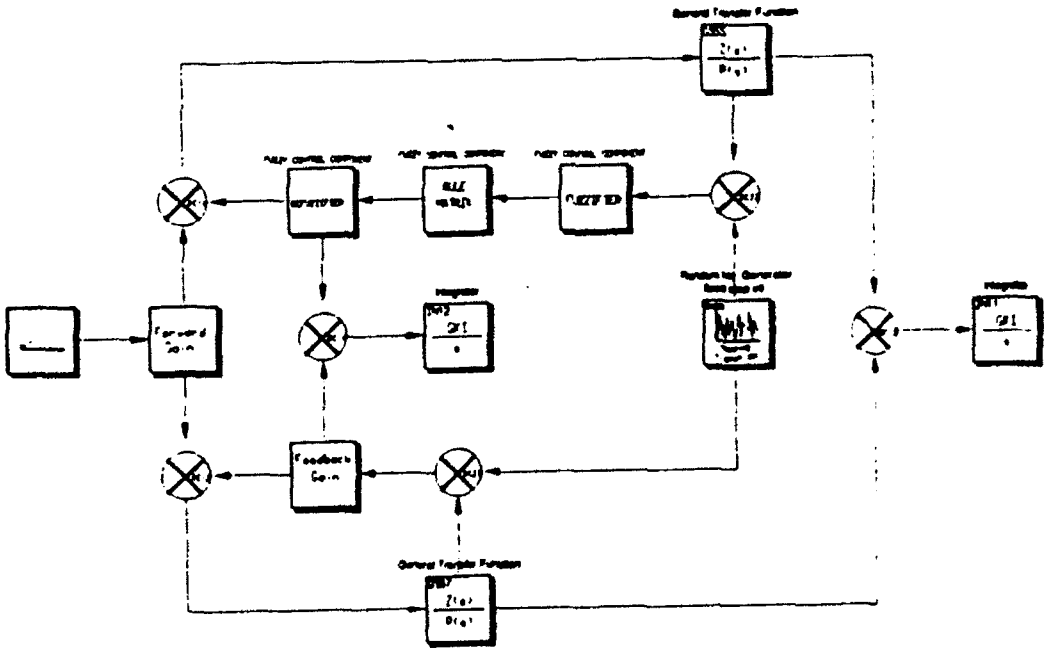


Fig 5. FLC/Linear Simulation.

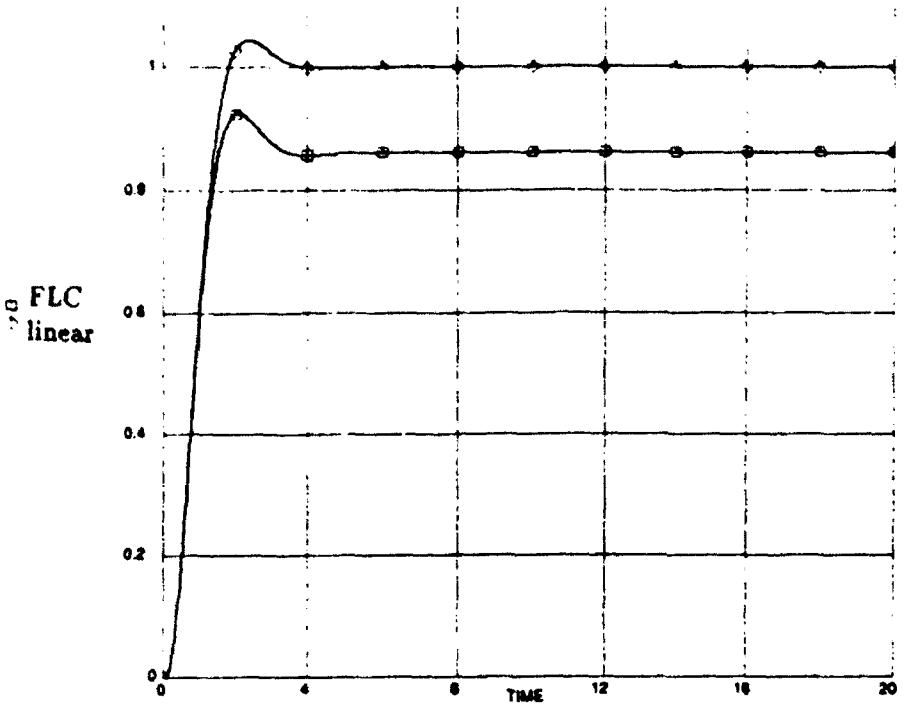


Fig 6. Type 0 system response.

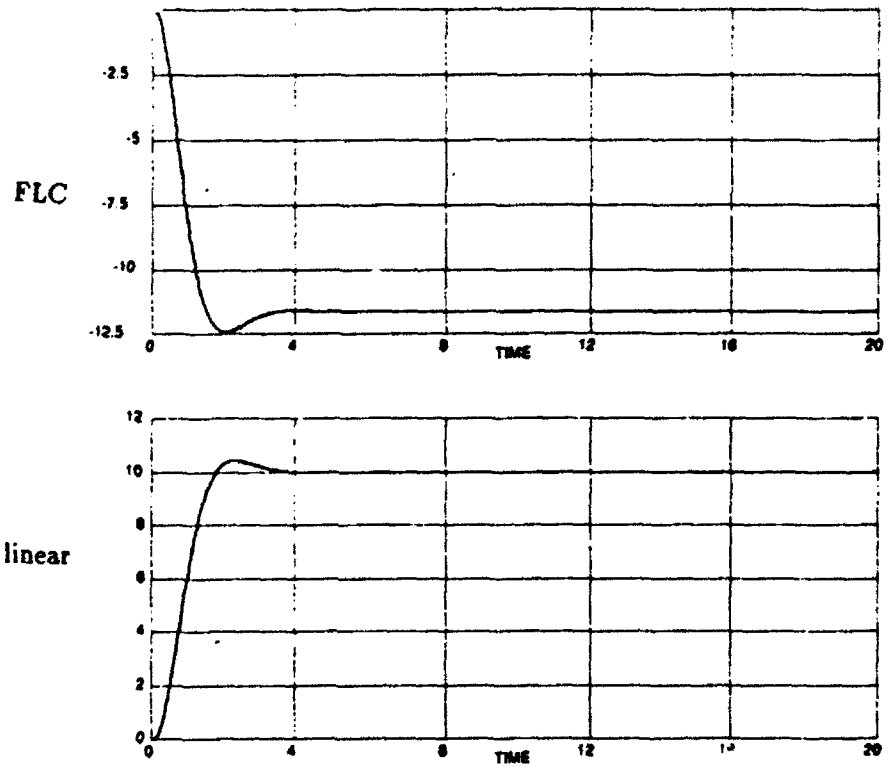


Fig 7. Feedback signals for Fig 6.

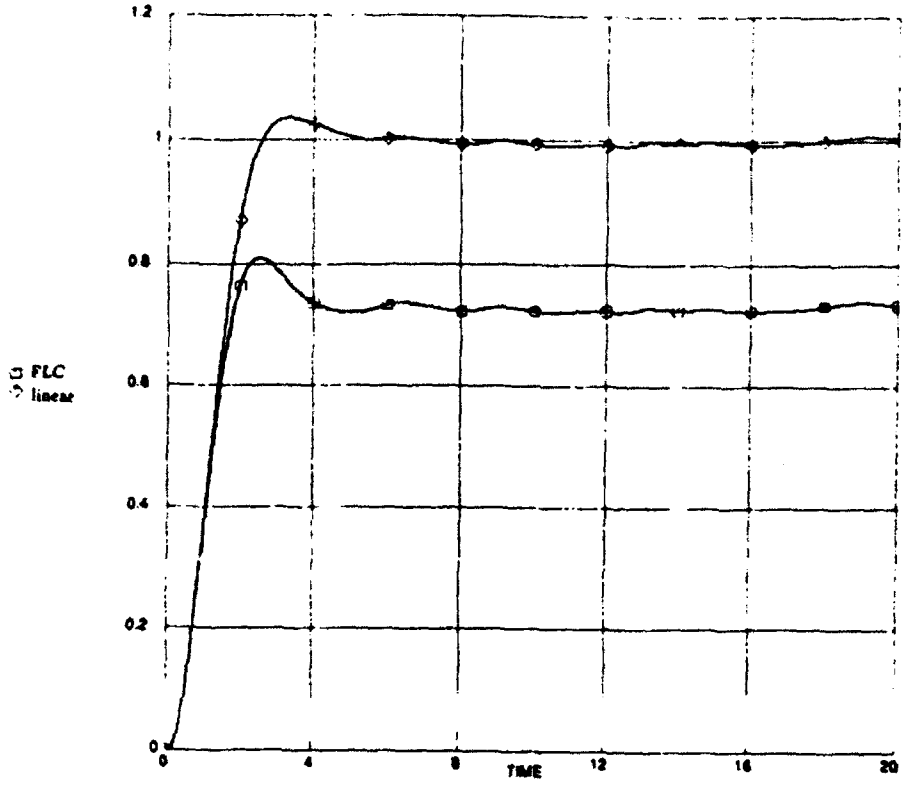


Fig 8. Type 1 system response.

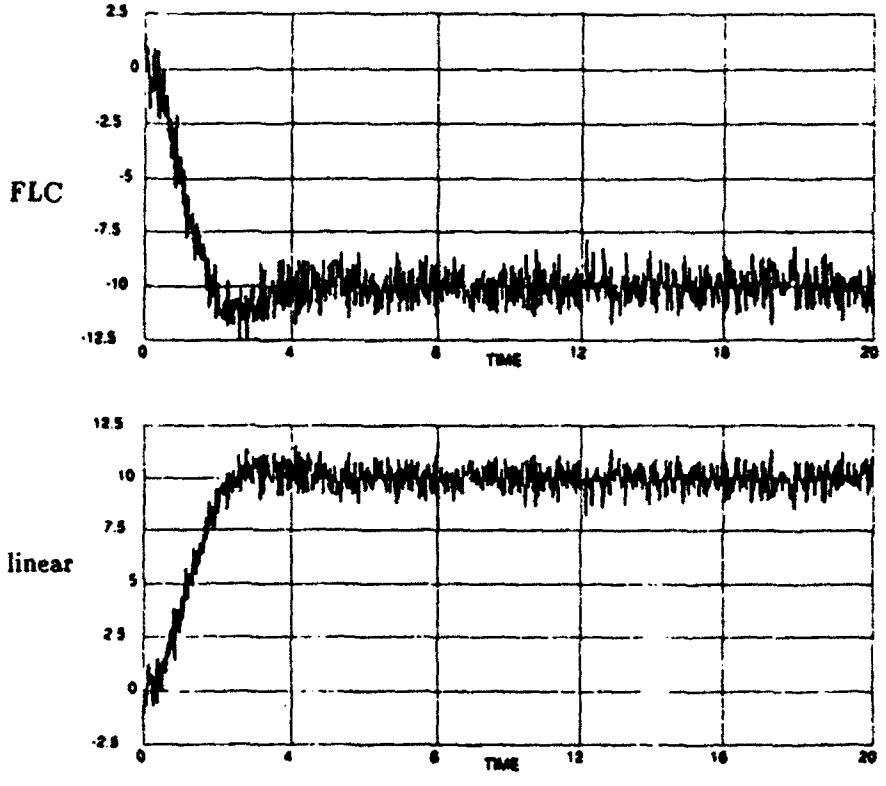


Fig 9. Feedback signals for Fig 8.

4. NONLINEAR FEEDBACK CONTROL

The application of nonlinear control to linear control systems present a dilemma in that the first choice for feedback would be to generate a system that would minimize a quadratic cost function. For nonlinear control systems this implies solving the Hamilton-Jacobi Equation. But when the Hamilton-Jacobi Equation is applied to a constant coefficient linear system the result is the Algebraic Ricotti Equation. Thus, the application of the Hamilton-Jacobi Equation to a linear system results in the familiar Linear Quadratic Regulator.

Thus application of nonlinear feedback to a linear system requires an approach that is not based on optimal control. It was desired to increase the system's response to a disturbance, without using gain scheduling or at least gain scheduling in the usual sense. All of the nonlinear feedback methods that were examined were functions that possessed even or odd symmetry about zero, and were thus applied to the error signal of the system.

One method that was examined was a feedback gain that was based on a inverted Gaussian curve. The gain was given by the following equation.

$$k = k_{\infty} - k_{diff} e^{-x^2/\sigma^2}$$

This gain function has the characteristics that as x gets far from zero the effective gain is k_{∞}, and for x close to zero the effective gain is $k_{\infty} - k_{diff}$. The variable σ relates to how rapidly the effective gain changes from $k_{\infty} - k_{diff}$ to k_{∞}. This system is effectively a continuous gain schedule. Simulations of this system, with comparison to the associated linear system given in Section 2 were run. Results indicated that this feedback method did increase the response speed to a large disturbance, but they were not appreciably better than that of a gain scheduled linear system. Further efforts related to these method were abandoned.

A more promising technique was linear and cubic feedback, as given by the following equation.

$$u = k_{lin} x + k_{cubic} x^3$$

For the first order system

$$\dot{x} = ax + bu$$

This feedback law always increases the speed of the system's response as long as the feedback increases the system's stability. For the first order system the time response is given by.

$$\frac{x}{\sqrt{-a + bk_{lin} + bk_{cubic}x^2}} = \frac{x_0}{\sqrt{-a + bk_{lin} + bk_{cubic}x_0^2}} e^{(a-bk_{lin})t}$$

It can be seen that the time component in the exponent is the same as that for the linear system. A nonlinear transformation of x to z results in a linear system.

$$z = \frac{x}{\sqrt{-a + bk_{lin} + bk_{cubic}x^2}}$$

$$z = z_0 e^{(a-bk_{lin})t}$$

It is easy to show that the response of the system with the added cubic feedback is always faster than the associated linear system. The system is stable independent of the value of x.

Unfortunately, the results are not as productive with higher dimensional systems as it was with the 1-dimensional first order system. The results of the time solution for the n-dimensional first order inhomogeneous differential equation are either an infinite series or an extremely complicated equation. In addition, if the root locus is examined it is seen that, for the 1-dimensional first order equation the root locus is a real locus located to the left of the open loop pole. While for the Type 0 system given in Section 2, ignoring the cubic feedback term, shows that the system is stable providing $-12 \leq k \leq 140$. With the application of the cubic feedback, we can set $k = bk_{lin} + bk_{cubic}x^2$ and it is seen that for large values of x the system will become unstable. For example, with $b = 1$, $k_{lin} = 0$ and $k_{cubic} = 0.1$ the region of stability is limited on one side to $x \leq 37.4$. The magnitude of k_{cubic} can be limited such that the system does not leave the desired stability region, but the effect of the cubic feedback is reduced for moderate values of x. This is primarily due to the rapid increase of x^3. Simulations of the Type 0 and 1 systems with the cubic feedback added proved to be unsatisfactory, and thus it was not examined further.

It may appear that the cubic feedback was unacceptable, except for 1-dimensional first order systems or possibly 2-dimensional first order systems that have a restricted zero arrangement. The ultimate goal of this program are for control system applications that are applied to aircraft systems. For aircraft control systems, control of the aircraft is through the use of aerodynamic control surfaces that are actuated by either electric

or hydraulic servos. These surface actuators are usually modelled as 1-dimensional first order systems and thusly they are candidates for a linear and cubic feedback law. Initial simulations indicated that this was a productive approach to the problem.

Servos are generally supplied as a complete package, but an add on controller can be added in order to implement the linear and cubic control law. The modified servo system is shown in Fig 10. For the given 1-dimensional first order system, $b = -a$. With u being the input to the servo and u_{com} being the command input to the servo, in order to obtain the linear and cubic control law for the servo the LC Control block must implement the following.

$$u = u_{com} + k_{cubic}(u_{com} - x)^3$$

This control law can be modified to affect the exponential component in the time response.

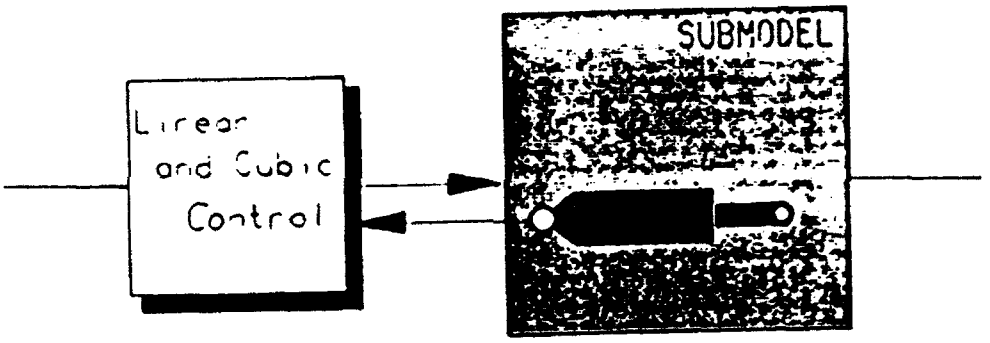


Fig 10. Servo with cubic feedback added.

Simulations indicated that the linear and cubic feedback implemented on the actuator servo for the previously given Type 0 and 1 system indicated improved system performance. The results of these simulations will not be presented here, as the next section presents the results of the simulations based on the Lambda URV, which is a Type 1 system with one zero.

An attempt was made at solving the Hamilton-Jacobi Equation of the Type 0 and 1 linear systems that included the linear and cubic servo. Due to the nonlinearities injected in the servo model, the Hamilton-Jacobi Equation does not reduce to the Ricotti Equation. Unfortunately, at this time these efforts have been unsuccessful.

4.THE LAMBDA URV APPLICATION

The Control System Development Branch of Wright Laboratory at Wright-Patterson AFB uses a remotely piloted vehicle, the Lambda Unmanned Research Vehicle (URV), for flight testing of new flight control systems. The aerodynamic model for the Lambda URV[10] was used in the short period approximation for longitudinal motion[11]. The servo model for the elevator servo was also obtained[10]. The servo model included position limits, but not rate limits. Rate limits typical for servo of the type used on the Lambda URV were implemented, 60°/s. The damping of the short period mode was lower than desired, a pitch damper was used to make this an acceptable value[10]. This lead-lag compensator was also implemented on the simulation of the Lambda URV.

As was the case in the previous simulations, two identical systems were run in parallel. One system was the standard Lambda URV, while the second system was augmented by a servo with linear and cubic feedback. Simulations with the inclusion of noise are not included here. An acceptable value of the cubic feedback term in the modified servo was used. Optimization of the servo response (ie. k_{cubic}) was not conducted. The simulations were run such that the the systems started in a disturbed state and it was the controllers goal to return their respective systems to the zero state. This was done in order to obtain useful data from the error analysis block which employed an l^1 norm, to produce a positive definite value for deviations from zero, comparisons were established so that a negative value indicated that the system with the modified servo was closer to the desired value. Some comparisons were integrated for knowledge pertaining to cummulative performance instead of a instantaneous value. An l^1 norm was chosen for easy comparison to the particular variable of interest.

The simulation presented in Fig 12-14 present system with the only difference being the modified servo.It can be seen in the simulation that the modified servo increases the systems response and that the aircraft spends less time away from the zero value, therefore less drag. Also the deviation of the elevator from zero is also reduced. Both of these factors indicate a decrease in drag due to the disturbance, and thus a smaller loss of energy.

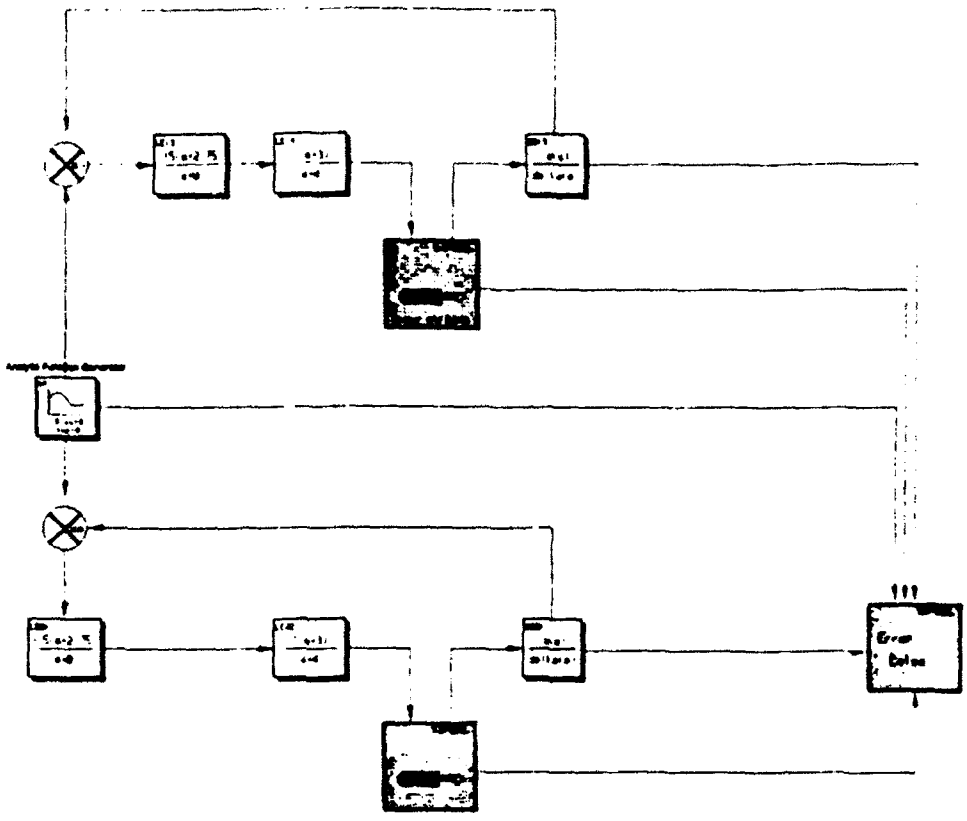


Fig 11. Lambda URV simulation block diagram.

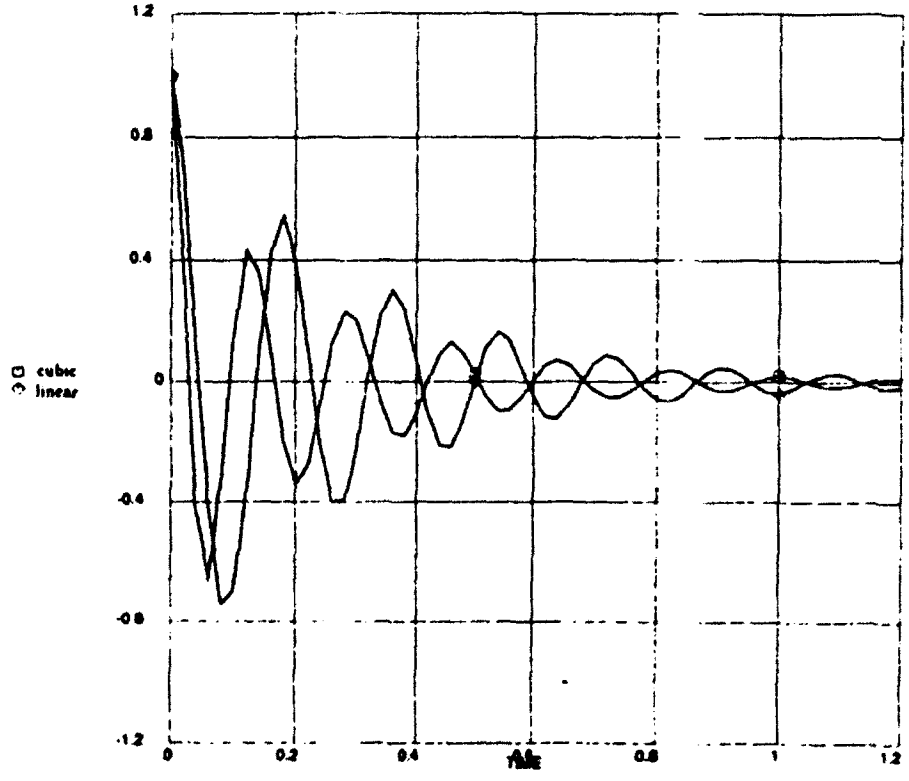


Fig 12. System output for initial disturbance of 1°.

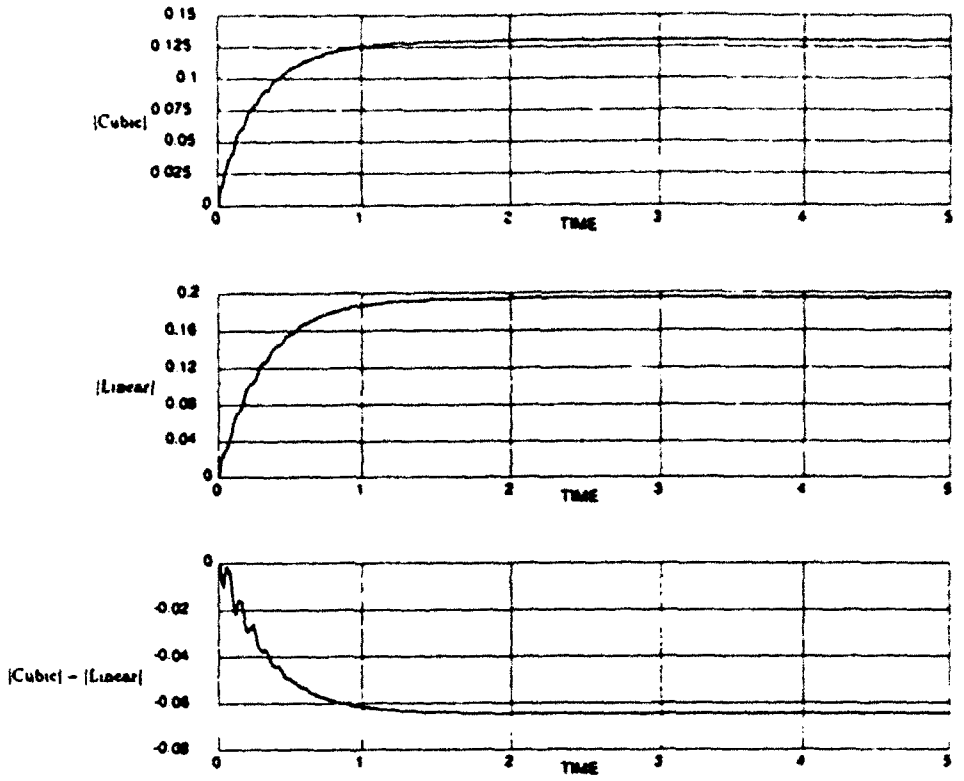


Fig 13. Output of integrated error signal.

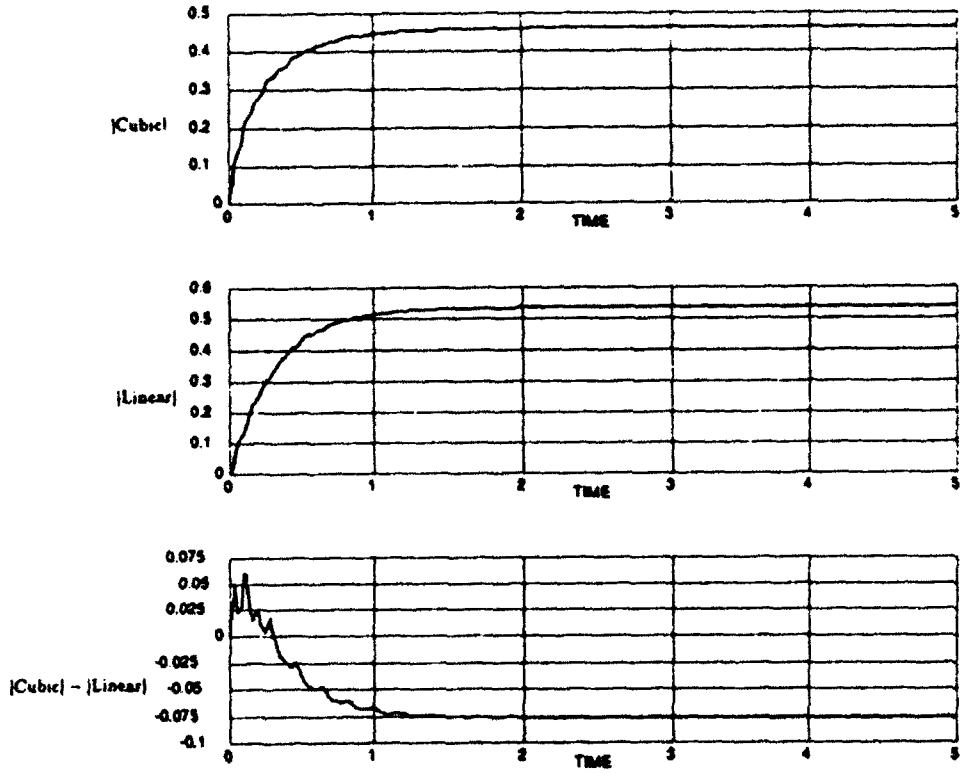


Fig 14. Servo integrated output signals.

CONCLUSIONS

The Fuzzy Logic Controllers while initially appearing attractive did not provide as many advantages as were expected at the beginning of this program. The numerical aspects of fuzzification and defuzzification, and the numeric translation of the linguistic rule base had been highly glossed over in the literature. These numerical aspects left many variables for the control engineer to adjust in the design of a FLC. The general nonlinearities of the FLC also generated problems in the use of FLC. In addition the extra on-line computer time involved in the implementation of a FLC, neccessitates the addition of a fuzzy logic engine which implies an increase in cost and hardware for an implementation of a FLC, unless the process under control were of a slow nature. Implementation of a FLC onto an already existing system with a standard digital control could be accomplished easily through the off-line processing of the FLC to a set of deterministic maps, and then these maps could be applied either in a look-up table or an IF-THEN-ELSE structure.

There were two positive attributes to the FLC structure. One would be for a partitioning of a space into regions via the use of a "Fuzzy System Compiler". This would allow for the input of linguistic rules, some space and value data and the output would be the regions of validity of the various control laws. Further refinement of the control laws could then be performed in the more standard methods. This technique would be extremely useful for complex problems, such as carrier landings of aircraft. The other area of use is for the situations were little or nothing is known of the dynamics of the system under investigation, but for which rules could be established, for example spin recovery of an aircraft.

An application of nonlinear feedback to enhance the time response for a linear system did provide some encouraging preliminary results. These techniques should be further examined as there probably exist better functions to use. Also, the inability of the nonlinear feedback technique that was employed on the simulations of the Lambda URV to operated on the entire system was a draw back. Other functions and methods should alieviate this problem.

ACKNOWLEDGEMENTS

I thank Capt. Stuart Sheldon and Mr. Duane Rubertus of the Control System Development Branch at Wright Laboratory for their help and assistance on this research. Thanks are due to Ms. Andrea Fellows at

RDL for her assistance with the Summer Faculty Research Program. This research was sponsored by the Air Force Office of Scientific Research.

REFERENCES

1. Berenji, Hamid R., **Fuzzy Logic Controllers**, to appear in Yager, R.R., and Zadeh, L.A., **Fuzzy Logic Applications in Intelligent Systems**. Kluwer Academic Publishers, 1991.

2. Kosko, B., **Nueral Nets and Fuzzy Systems**. ©1992

3. Berenji, Hamid R., et al, **A Hierarchical Approach to Designing Approximate Reasoning-Based Controllers for Dynamic Physical Systems**. Priceedings of the Sixth Conference on Uncertainty in Artificial Intelligence. 27-29 July 1990, Cambridge, Mass.

4. Berenji, Hamid R., **Refinement of Approximate Reasoning-based Controllers by Reinforcement Learning**. Proceedings of the Eighth International Workshop, 27-29 June 1991, Evanston, Ill.

5. Langari, G.R., and Tomizuka, M., **Stability of Fuzzy Linguistic Control Systems**. IEEE Conference on Decisions and Control, Hawaii, December 1990.

6. Lee, C. and Berenji, H.R., **An Intelligent Controller Based on Approximate Reasoning and Reinforcement Learning**. Published by the IEEE Computer Society, ©1989.

7. Steinberg, M.L. and DiGirolamo, R.D., **Neural Nerwork and Fuzzy Logic Technology for Naval Flight Control Systems**. NADC-91080-60, Naval Air Development Center (Code 6012), Warminster, PA, 1991.

8. Stanley, **Digital Signal Processing**.

9. **EASY5x Users Guide**, Boeing Computer Services, ©1987

10. Swift, 1LT Gerald A., **Model Identification and Control System Design for the Lambda Unmanned Research Vehicle**. MS Thesis AFIT/GAE/ENY/91S-4. School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB, OH, Sept 91.

11. Etkin, Bernard, **Dynamics of Flight-Stability and Control**. Second edition. Wiley and Sons, New York, New York, ©1982

Microstructural Evolution of Ti-23.2Al-24.4Nb

Ian W. Hall
Chairman, Materials Science Program
and
Associate Professor of Mechanical Engineering

Spencer Laboratory
University of Delaware
Newark, DE 19716

Final Report for:
Summer Research Program
Wright Laboratory, WPAFB

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

September 1992

Microstructural Evolution of Ti-23.2Al-24.4Nb

Ian W. Hall
Chairman, Materials Science Program and
Associate Professor of Mechanical Engineering
Spencer Laboratory
University of Delaware
Newark, DE 19716

Abstract

A Ti-23.2Al-24.4Nb alloy has been subjected to long term heat treatments at temperatures within its projected use range. The microstructure was studied by transmission electron microscopy and the phase changes which occurred were investigated as the heat treatment temperature was increased from 500°C to 800°C. It was found that the alloy transformed from an initial microstructure which was predominantly ordered beta to a predominantly orthorhombic structure containing a few percent of beta and alpha-2 phases. Furthermore, it appears that the orthorhombic phase undergoes an ordering reaction within the temperature range of interest.

Microstructural Evolution of Ti-23.2Al-24.4Nb

Ian W. Hall

1. Background

The need for new materials for high temperature applications has led to the development of many promising intermetallic systems: titanium aluminides are one group of such materials. Early interest in titanium aluminides quickly identified their advantages, such as strength at high temperature, oxidation resistance and low density, but also pointed out their disadvantages, principally low ductility and toughness at room temperature. More recent work has shown that proper alloying can improve the ductility and, since there is cause for believing that the properties of monolithic aluminides can become acceptable, there is now the interest in producing composites of these intermetallics.

Titanium aluminides were initially based on either Ti_3Al or TiAl, so-called α_2 or γ, but more recent work has explored other compositions and the phase relationships and stabilities are, of course, much more complex. One of the particularly promising new intermetallic alloys is based on a variation of the α_2 phase, with the composition approximately Ti_2AlNb (Ti-25at.%Al-25at.%Nb). This material is being extensively studied and considerable progress has been made in outlining its structure.

Banerjee *et al.* [1] first identified a new orthorhombic phase in Ti-25at%Al-12.5at.%Nb alloy (all compositions hereafter are given in atomic %). They later investigated the structure, tensile deformation and fracture of the same alloy and presented a schematic pseudo-binary Ti_3Al-Nb phase diagram [2]. Rowe *et al.* [3] have investigated the mechanical properties of intermetallics near the composition Ti-25Al-25Nb and found that compositions near Ti-24Al-25Nb yielded the best combination of properties with a yield stress estimated at ~800MPa up to ~750°C. The optimum structure was a dual phase orthorhombic+ordered beta microstructure. Later, Rowe *et al.* [4] also showed that the creep resistance was comparable to currently available titanium aluminides. Since these results all indicate that alloys based upon this composition may be suitable for use as a matrix material for intermetallic matrix composites, it was decided to conduct a microstructural investigation of the long term stability of the alloy at temperatures up to its maximum potential use temperature. This report presents the results of that study.

2. Experimental

The material used in this study was Ti-23.2%Al-24.4%Nb, corresponding closely to the desired stoichiometric composition Ti_2AlNb. Initial production and processing was by Timet® and consisted of hot pack rolling, grinding and pickling to remove surface contamination. Thereafter, the material was further processed to sheet by T.I. Inc. of Attleboro, Mass.. Although the process is proprietary it is known that the alloy was reduced in successive passes at room temperature with interstage annealing treatments at a temperature in the range 980-1010°C. The final pass, after the last annealing treatment, consisted of a light reduction (or 'kiss pass') to final dimensions. The sheet was 2.0mm thick and 133mm wide.

Samples of this starting material were wrapped in tantalum foil, and then encapsulated in quartz tubes with titanium sponge as a getter. Isothermal heat treatments of 500 and 1000 hours at 500°, 600°, 700°, and 800°C were carried out followed by air cooling. Samples were prepared for optical and scanning electron microscopy, X-ray microanalysis and transmission electron microscopy (TEM). This report focuses principally upon the transmission electron microscopy study which constituted the main activity of this investigator. Details of the other studies can be found in a forthcoming report [5]

Specimens for TEM were prepared by mechanical thinning to ~100μm, followed by dimpling on a Gatan Dimple Grinder (merely to ensure perforation in the center of the foil). They were then electropolished in a Fischione twin jet polisher at 15V and -40°C using an electrolyte of 250ml ethanol, 150ml ethylene glycol monobutyl ether and 15ml perchloric acid. It is remarked here that, although results were satisfactory for all but the 800°C heat treatments, other researchers report using the normal ethanol, butanol, perchloric acid electrolyte which may provide superior results for heat treated conditions which prove difficult to polish.

Samples were sent for oxygen and nitrogen analysis: the results, presented below as Table 1, show that significant interstitial contamination only occurred after the longest heat treatment at the highest temperature. The TEM study showed no obvious effects arising due to interstitial pick-up.

Table 1. Oxygen/Nitrogen Analyses

Sample	wt% Oxygen	wt% Nitrogen
As received	0.115 ± 0.002	0.017±0.001
500°C for 500 hours	0.116±0.004	0.018
500°C for 1000 hours	0.115±0.002	0.017±0.002
800°C for 500 hours	0.114±0.003	0.013±0.003
800°C for 1000 hours	0.140±0.003	0.020±0.001

3. Results

The following microstructural observations will be presented in order of increasing heat treatment temperature and heat treatment time, after a description of the initial material.

The structures had been previously examined by SEM to determine the major features of the microstructure [5] and the initial structure is illustrated in Figure 1. It can be seen that the as-received structure consisted of three phases, namely, i) the matrix with a grain size of approximately 20μm, ii) large particles approximately 2-5μm in length and ~2μm wide, and iii) smaller needlelike particles ~1μm in length and ~0.5μm wide. The large particles were identified by microprobe analysis as α_2 and the smaller ones were believed to be the orthorhombic phase. After ageing for 500 hours at 500°C the matrix grains were clearly delineated by a network of α_2 particles and a contrast effect apparently associated with the grain boundaries suggested the occurrence of a transformation. After heat treatment for 500hrs at 600°C, the α_2 had spheroidized, the needles of orthorhombic phase were less clearly in evidence and a lighter phase had begun to appear as small equiaxed grains ~0.25μm in diameter. After treatment at 700°C the matrix appeared to have undergone recrystallization and the light phase had increased in volume fraction while the needles of orthorhombic phase had effectively disappeared. Treatment for 500 hours at 800°C led to two morphologies of the (presumably) same light phase, namely, similarly oriented needle-like precipitates within the grains and grain boundary precipitates as illustrated in Figure 2.

3.a) As received condition

TEM shows that the microstructure consists of three phases. The matrix consists of large grains ~ 20μm in diameter which are subdivided into smaller subgrains about 0.5-1μm in diameter, Fig. 3. The subgrains form no doubt from the successive roll/anneal cycles during production. Distributed within the matrix and aligned in the rolling direction are α_2 particles ~2μm in width and with an aspect ratio of ~2 as shown in Fig. 4. The third phase is present as short aligned plates, constituting ~5% by volume of the alloy, which usually do not coincide with the rolling direction and which, therefore, are presumed to have formed during the interstage annealing treatment and been unaffected by the final light rolling pass.

Considering first the major (matrix) phase, selected area diffraction patterns can be indexed on the basis of a body centered cubic cell with a lattice parameter of 3.23Å. Many, but by no means all, of the patterns exhibit clear and sharp superlattice reflections at the 010 matrix reflection positions, Fig. 5, indicating the presence of an ordered structure and also show pronounced diffuse streaking along <110>. The streaking arises from a tweed effect which is clearly visible at high magnifications. Other patterns exhibit much weaker superlattice reflections but show correspondingly stronger streaking in <110> and additional reflections in positions which can not arise from β or β_o. Figure 6 shows both superlattice reflections at $(100)_\beta$ etc. and the pronounced diffuse streaking in <112> and reflections with spacings at $\frac{1}{2}(112)_\beta$ etc.. Since the nature of the diffraction patterns obtained from individual grains shows much variation and the appearance of what appear be irrational reflections, it is considered probable that the matrix consists of a mixture of disordered beta and ordered beta phases, referred to as β and β_o respectively, and also that another transformation is also in its early stages.

The large, clearly visible second phase grains are identified by electron diffraction as α_2, Fig. 7. The α_2 is very heavily faulted on $\{10\bar{1}0\}$ as shown in Fig. 8, probably representing the early stages of a phase transformation. The needle-like orthorhombic phase will be illustrated in the next heat treated condition

3.b) 500° for 500hrs.

The α_2 particles are still visible and are unchanged in appearance. The small needle-like grains are still there, apparently without transformation and are identified as the orthorhombic phase. Fig. 9(a) shows a general view of these particles and Fig. 9(b) shows a higher magnification view of a single one. The particle has well-defined edges on two sides and is probably a needle which cuts through the plane of the foil at a shallow angle. The indexed $[1\bar{1}0]$ zone axis diffraction pattern from this latter particle is shown in Fig. 10.

In the matrix, a complex transformation of the matrix has clearly begun with complex faulting and lamellar features appearing across entire grains in well-defined crystallographic directions in four variants, Fig. 11. In some regions the transformation occurs over a large area in a single variant as illustrated in Fig. 12(a) and the corresponding diffraction pattern Fig. 12(b) which shows the expected corresponding streaking of the spots.

All matrix diffraction patterns are now indexable as the orthorhombic phase, however, the differences between this and the β are subtle. Fig. 13(a) is an indexed $[100]_o$ matrix pattern, still containing vestigial spots from the β matrix, which should be compared with the $[100]_\beta$ pattern of the as-received matrix in Fig. 5. The similarity is striking since the two are related by only a very minor change in lattice parameter.

Similarly Fig. 13(b) is a $[1\bar{1}0]$ orthorhombic pattern which, it can be seen, closely resembles a $<111>_\beta$ pattern with a ~4% change in one of the lattice parameters. Furthermore, comparison of Figure 13(b) with the pattern from the same zone axis of the small acicular particle described just above, Fig. 10, shows a clear difference in relative intensities between the spots which is indicative of a structure factor difference between the two crystals even though both are orthorhombic. This must be due to differences in the atomic configurations in the respective orthorhombic structures, one of which arises due to a low temperature process involving only short diffusion distances whereas the other is the result of a high temperature annealing treatment where diffusion is much less restricted. This gives rise to a highly ordered orthorhombic structure after low temperature ageing and a more highly disordered orthorhombic structure after high temperature ageing. The observation of both types of orthorhombic phase in the same microstructure is interesting since it indicates that it is easier to nucleate ordered

orthorhombic phase from ordered beta than to follow the disorder→order transformation.

3.c) 500° for 1000hrs.

After 1000 hours the structure had evolved considerably and, although many areas still showed the complex faulting and lamellar transformation product several microns in extent, Fig. 14, many regions now exhibited discrete precipitates. These regions were typically located near original prior β grain boundaries and consisted of small, new recrystallized grains with extensive precipitation within them. This is illustrated in Fig. 15 which shows a prior β grain boundary lying horizontally across the center of the micrograph and 0.5-1μm grains nucleated there; extensive precipitation is evident in the adjoining grains. The grains which have nucleated at the boundaries are often entirely single phase orthorhombic with no trace of intragranular precipitation, Fig. 16.

The precipitates themselves are very thin lamellae and will be further illustrated at a later stage of their development below.

3.d) 600° for 500hrs.

The transformation is now quite well advanced and the planar faults and striations, barely visible previously after the 500° treatments, are now well resolved as lamellar precipitates. They are extremely thin and are ribbon-like with partial dislocations visible along their length. Many are quite straight while others appear to be rather wavy, Fig. 17. Despite repeated attempts, it has proved impossible to obtain diffraction patterns from these features and their identity as either faults or discrete precipitates is still uncertain.

The structure at the prior β grain boundaries resembles that of the 500°/1000h specimens and the transformation there again appears to be giving rise to single phase grains and recrystallized grains with lamellar precipitates, Fig. 18. The α_2 and the small acicular particles of the orthorhombic phase, which were both present in the starting material, are still visible and unchanged in appearance.

In summary, the structure after 600°/500h appears analogous to the 500°/1000h specimens.

3.e) 700° for 500hrs.

Figure 19 shows a general view of the structure after this treatment, the principal difference being a modest coarsening of the structure present after 600° treatments. Precipitation is still very heavy and fine-scale in most regions except in the grain boundary grains where single phase grains appear to be undergoing a discontinuous precipitation process, Fig. 20. The α_2 grains from the starting material can still be found but they are much reduced in size, Fig. 21.

3.f) 800° for 500hrs.

The structure had totally recrystallized, yielding a grain size of approximately 5-10μm. The majority of the grains contained well oriented needle-like precipitates, Fig. 22(a): the precipitates were typically 100nm in thickness, Fig. 22(b). Large regions of the matrix consisted entirely of monolithic precipitate-free grains but these did, however, exhibit a contrast effect which may indicate an ordering or further transformation reaction, Fig. 23. Other small equiaxed grains are described more fully in the next section.

In addition, occasional grains of α_2 were also observed, although much less commonly than in the 700°C treated material

3.g) 800° for 1000hrs.

The structure was basically the same as that after 500 hours at 800°C, the needle-like precipitates had not coarsened appreciably, Fig. 24. The sample appeared to consist of two types of grains, the first being the 5-10μm grains with crystallographically oriented precipitates and the second being small equiaxed grains ~0.5μm in diameter, Fig. 25. These were both studied using convergent beam and conventional electron diffraction. The matrix was shown to be orthorhombic with approximate lattice parameters of a=0.605nm, b=0.961nm and c=0.465nm. Figures 26(a) and (b) are from $[102]_o$ and $[001]_o$ matrix zone axes respectively.

The needles and the small equiaxed grains were both found to consist of β-phase. Figures 27(a) and (b) show [100] patterns clearly exhibiting the expected

four-fold symmetry and, furthermore, showing no evidence of superlattice reflections which would be present if the structure were ordered.

Again, occasional grains of α_2 were found, but the total volume fraction is estimated at <5 volume%.

4. Discussion

If this material is to find use in high temperature applications, its thermal stability and equilibrium microstructure at temperature are of fundamental concern. The results of this investigation indicate that, for this specific composition, the equilibrium microstructure at moderately elevated temperatures will consist of orthorhombic, beta and α_2 phases. The structure present after 500hrs at 800°C appears to be essentially unchanged after a further 500 hours, indicating that a stable three phase microstructure should be obtainable. Such a structure should present many possibilities for thermomechanical treatment to obtain suitable microstructures for intermetallic matrix composites.

The various transformations within the Ti-Al-Nb system have been studied by several investigators [6-11] and a broad understanding of the processes has been formulated. The phase transformations which occur depend, of course, upon the precise composition, and the transformation temperature. In the present study the transformations which have been encountered are:

i) $\beta \rightarrow$ Orthorhombic (O)

ii) $\beta_o \rightarrow$ O

iii) $\alpha_2 \rightarrow$ O

iv) O $\rightarrow \beta$

and combinations thereof. Of these, i) & ii) were the major ones and occurred as the as-received material was heated up: iii) occurred as the α_2 present in the initial material decreased in volume fraction: iv) occurred as the needles were precipitated in the large O grains at 800°C.

At this point it is useful to consider the sequence of phase transformations which may be expected as the material is heated up. The starting structure consisted principally of β_o with some disordered β, ~10vol% of α_2, and ~4vol% orthorhombic phase. The latter 3 phases were presumably present at the annealing temperatures used in production of the sheet and the β_o formed during the cooling from this temperature. In addition, electron diffraction, Fig. 6, showed that both

the β and the β_o had already begun to transform even during the final cooling, leading to the appearance of the tweed structure and to evidence of the orthorhombic phase. With increasing time and temperature thereafter, the transformation of β and $\beta_o$to an orthorhombic phase is essentially complete after 500°/500h. However, intensity changes continue to occur in electron diffraction patterns (compare Figs.10 & 13(b)) indicating that transformation continues up to 800°C. The lamellar precipitates have not been unambiguously identified, however, they are probably precursors to the β needles which have been identified definitely after treatment at 800°C. The final structure consists, therefore, of O, β and α_2 phases although it is not at all clear at this point why disordered β forms instead of the expected β_o.

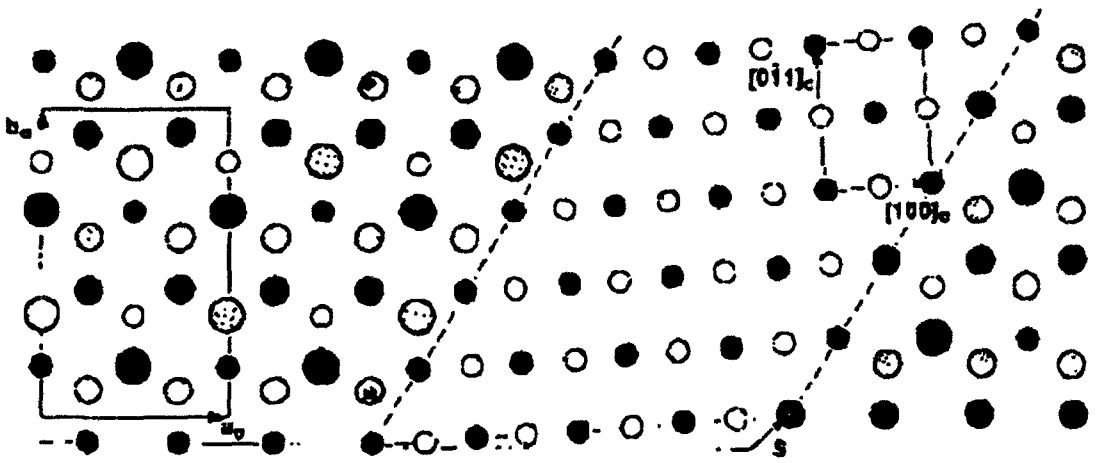
The diffraction patterns obtained allow orientation relationships to be derived for most of the phases involved in the transformations, the principal one (which can be derived from Fig. 5, for example) being as follows for either precipitation of O from β or β_o (at 500°C) and precipitation of β from O (at 800°C):

$$(001)_o \parallel (0\bar{1}1)_\beta$$
$$[100]_o \wedge [100]_\beta \approx 2\text{-}3°$$

The orientation relationship between the orthorhombic and beta phases, and the β habit plane has been studied experimentally and theoretically by Bendersky *et al.* [11] using TEM and a minimization of strain energy technique. They found by TEM an O.R. of the type:

$$(110)_o \parallel (\bar{2}\bar{1}1)_\beta$$
$$[001]_o \parallel [001]_\beta$$

with a {112} habit plane. Calculations broadly agreed with these observations but it was pointed out that the accuracy of matching and the deviation from any specific orientation relationship is a sensitive function of precise composition and lattice occupancy. They presented the model below showing the atomic matching to be expected across such an interface in which $[001]_\beta \wedge [100]_o$ is ~2-3°, a conclusion which is corroborated by the present study.



(after Bendersky *et al.*, Ref [11])

It appears, then, that the Ti-Al-Nb system presents the possibility of other order-disorder transformations in the orthorhombic phase in addition to the already well known $\beta \rightarrow \beta_o$ transformation. Even the $\beta \rightarrow \beta_o$ transformation deserves further study since, as pointed out above, the needles which form in the orthorhombic phase after heat treatment at 800°C appear to be β and not β_o. *Investigation of the occurrence of such* transformations could involve high temperature X-ray or *in-situ* TEM studies in order to determine the phase boundaries and structures. However, the implications for mechanical properties should not be overlooked, especially if order/disorder transformations should occur in the temperature regime of interest for structural applications. A transformation of this type could make the ductility of the alloy strongly dependent upon temperature and thermal history, especially under conditions of thermal cycling.

Finally, it is suggested that now the microstructure of the monolithic alloy is reasonably well understood, future work should be directed to the effects of processing into composite form. Microstructural effects of the unavoidable presence of residual stresses on the order/disorder transformations may may be anticipated. Also, interstitial element solution from potential reinforcing fibers may be a further source of influence on the microstructure.

5. References

[1] D. Banerjee, A.K. Gogia, T. K. Nandy and V. A Joshi, Acta Metall., 36, 1988, p. 871.

[2] A. K. Gogia, D. Banerjee and T. K. Nandy, Metall. Trans. A, 21A, 1990, p. 609.

[3] R. G. Rowe, in "Microstructure/Property Relationships in Titanium Aluminides and Alloys", (Y.-W. Kim & R. R. Boyer. eds.) TMS, Warrendale, PA, (1991) p. 387.

[4] R. G. Rowe, D. G. Konitzer, A. P. Woodfield and J. C. Chesnutt, Mat. Res. Soc. Symp. Proc. Vol. 213, (1991) Materials Research Society, p. 703.

[5] I. W. Hall, J. Menon, D. Miracle, in preparation, WL/MLLM, W-PAFB.

[6] L. M. Hsiung and H. N. G. Wadley, Scripta Met. et Mat., 27, (1992) p. 605.

[7] H. T. Weykamp, D. R. Baker, D. M. Paxton and M. J. Kaufman, Scripta Met. et Mat., 24, (1990) p. 445.

[8] R. Strychor, J. C. Williams and W.A. Soffa, Metall. Trans. A, 19A, (1988), p. 225.

[9] B. Mozer, L. A Bendersky, W. J. Boettinger and R. G. Rowe, Scripta Met. et Mat., 24, (1990) p. 2363.

[10] L. A. Bendersky, B. P. Burton, W. J. Boettinger and F. S. Biancaniello, Scripta Met. et Mat., 24, (1990) p. 1541.

[11] L. A. Bendersky, W. J. Boettinger and A. Roytburd, Acta Metall. et Mater., 39, (1991) p. 1959.

6. Figure Captions

Fig. 1. Microstructure of as-received material.

Fig. 2. Microstructure after heat treatment at 800°C for 500hrs.

Fig. 3. TEM micrograph of as-received material showing small subgrain size.

Fig. 4. α_2 grains in as-received material.

Fig. 5. [001] matrix SADP showing superlattice reflections and streaking along <110>.

Fig. 6. $[1\bar{1}0]$ matrix SADP showing streaking, superlattice and extra reflections.

Fig. 7. [0001] SADP from α_2.

Fig. 8. α_2 grains showing heavy faulting on $\{10\bar{1}0\}$.

Fig. 9(a). Particles of orthorhombic phase, (some are labelled O) in material heat treated at 500°C for 500 hours (500°/500h).

Fig. 9(b). Higher magnification view of an O particle.

Fig. 10. $[001]_o$ SADP from orthorhombic crystal.

Fig. 11. Complex transformation in matrix, (500°/500h).

Fig. 12(a) Large area of similarly oriented transformation product.

Fig. 12(b) SADP from Fig. 12(a).

Fig. 13(a). $[100]_o$ zone axis pattern.

Fig. 13(b) $[1\bar{1}0]_o$ zone axis pattern.

Fig. 14. Transformation after 500°/1000h.

Fig. 15. Recrystallization at prior β grain boundary (horizontal across center of micrograph).

Fig. 16. Monolithic orthorhombic crystal at prior β grain boundary.

Fig. 17. Lamellar precipitates in 600°/500h material.

Fig. 18. Orthorhombic grains at prior β grain boundary.

Fig. 19. General view of 700°/500h material.

Fig. 20. Discontinuous precipitation in orthorhombic phase at prior β grain boundary.

Fig. 21. α_2 grain and lamellar precipitates in orthorhombic matrix.

Fig. 22(a). General view of 800°/500h showing needle-like precipitates in orthorhombic matrix.

Fig. 22(b). Higher magnification view showing the apparently needle-like precipitates (800°/500h).

Fig. 23. Apparent fine-scale transformation within the orthorhombic matrix.

Fig. 24. Needle-like precipitates after 800°/1000h.

Fig. 25. Type of small crystal occurring at orthorhombic grain boundary triple points, etc..

Fig. 26(a) $[102]_o$ SADP from matrix.

Fig. 26(b). $[001]_o$ SADP from matrix.

Fig. 27(a). $[001]_\beta$ SADP from one of the needles.

Fig. 27(b). Corresponding $[001]_\beta$ CBDP showing the expected 4mm symmetry.

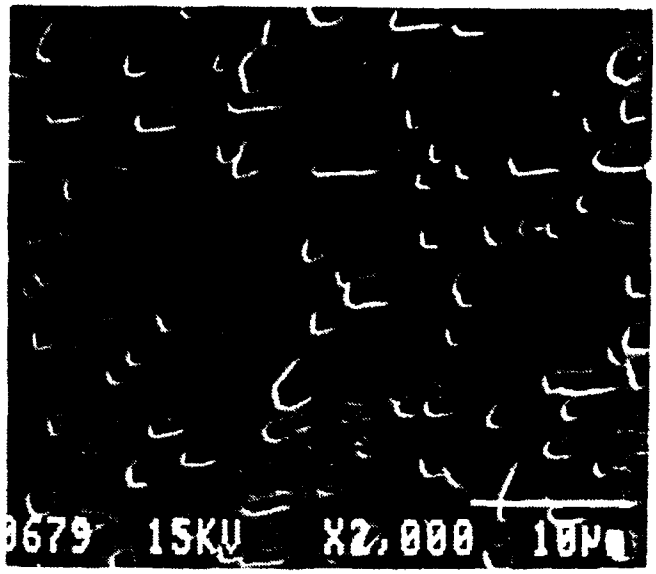


Fig. 1. Microstructure of as-received material

Fig. 2. Microstructure after heat treatment at 800°C for 500hrs.

Fig. 3. TEM micrograph of as-received material showing small subgrain size.

Fig. 4. α_2 grains in as-received material.

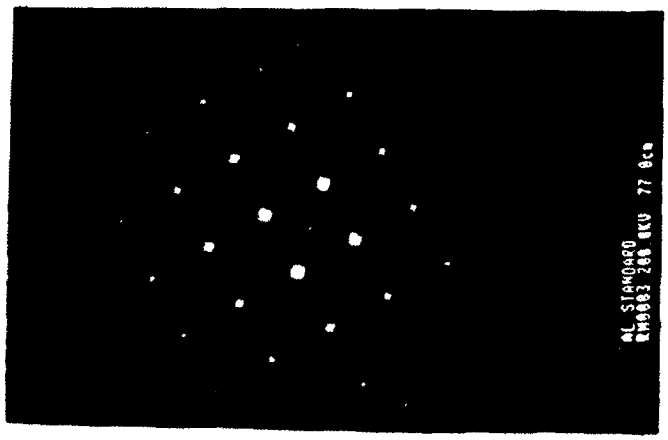


Fig. 5. [001] matrix SADP showing superlattice reflections and streaking along <110>.

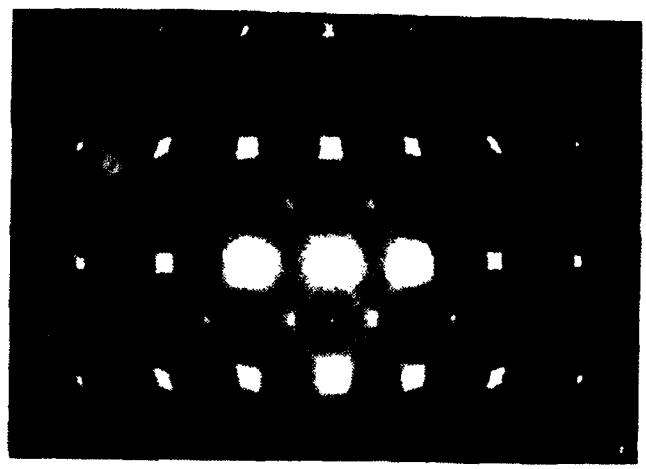


Fig. 6. [1$\bar{1}$0] matrix SADP showing streaking, superlattice and extra reflections.

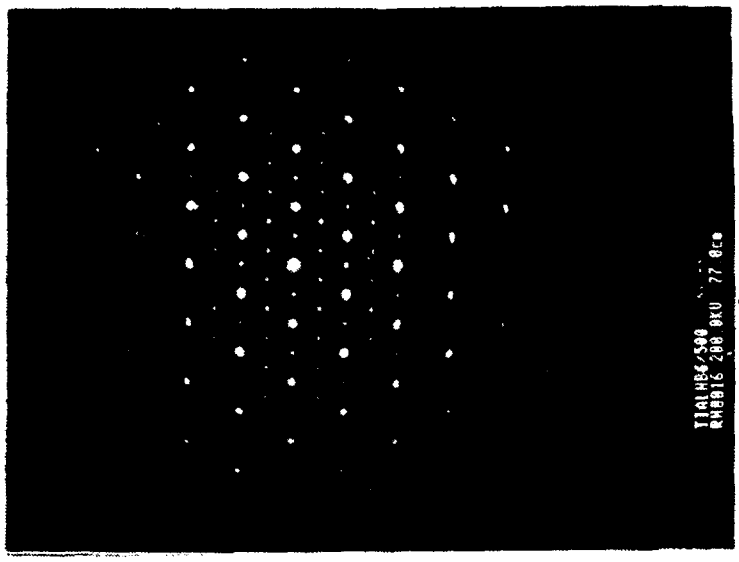


Fig. 7. [0001] SADP from α_2.

Fig. 8. α_2 grains showing heavy faulting on {10$\bar{1}$0}.

Fig. 9(a). Particles of orthorhombic phase, view (some are labelled O) in material heat treated at 500°C for 500 hours (500°/500h).

Fig. 9(b). Higher magnification of an O particle.

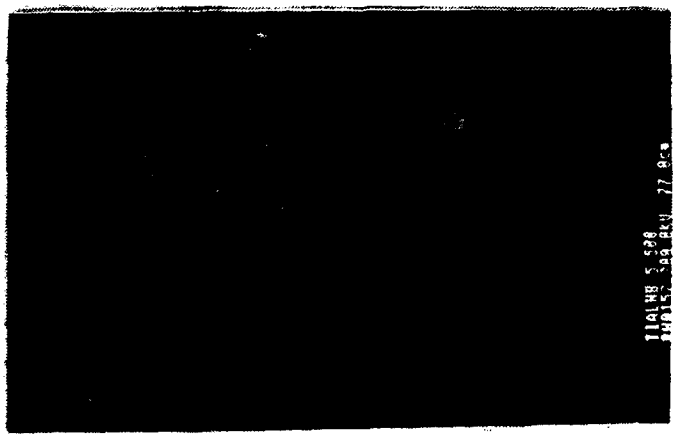


Fig. 10. $[1\bar{1}0]_o$ SADP from orthorhombic crystal.

Fig. 11. Complex transformation in matrix, (500°/500h).

Fig. 12(a) Large area of similarly oriented transformation product.

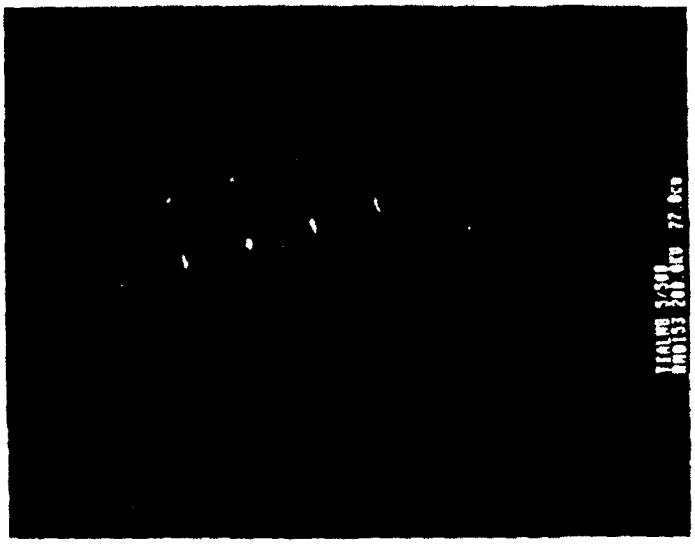


Fig. 12(b) SADP from Fig. 12(a).

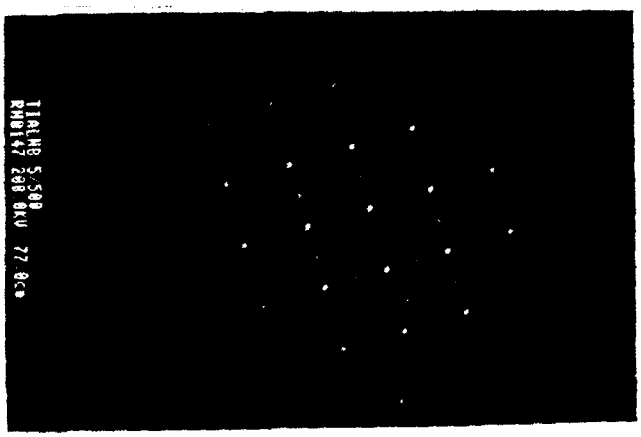


Fig. 13(a). $[100]_o$ zone axis pattern.

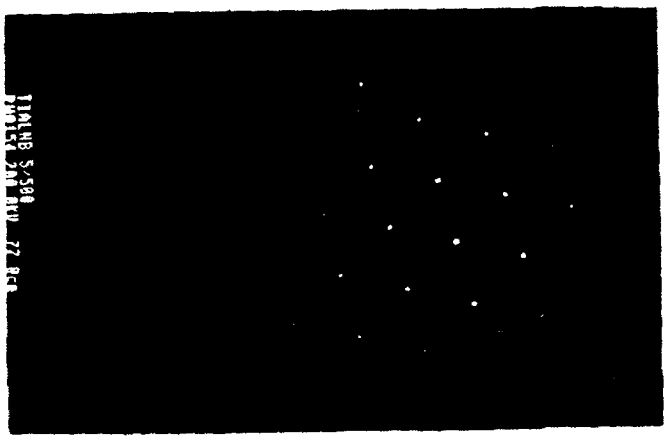


Fig. 13(b) $[1\bar{1}0]_o$ zone axis pattern.

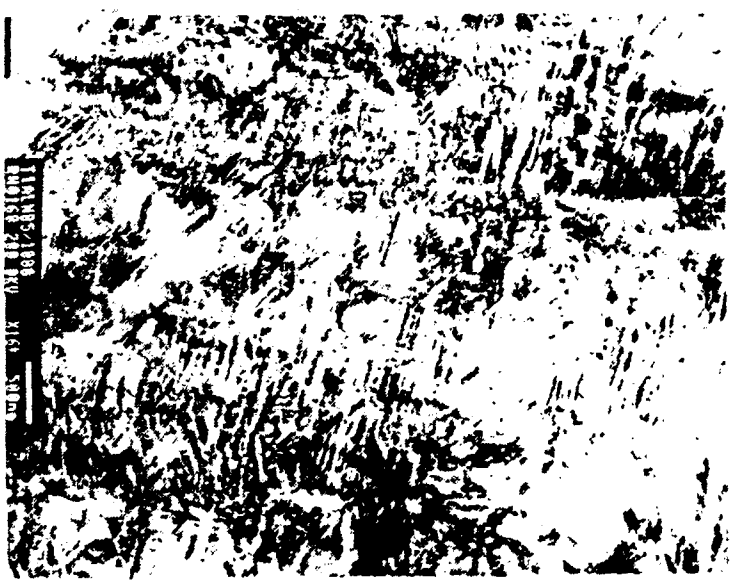


Fig. 14. Transformation after 500°/1000h.

Fig. 15. Recrystallization at prior β grain boundary (horizontal across center of micrograph).

Fig. 16. Monolithic orthorhombic crystal at prior β grain boundary.

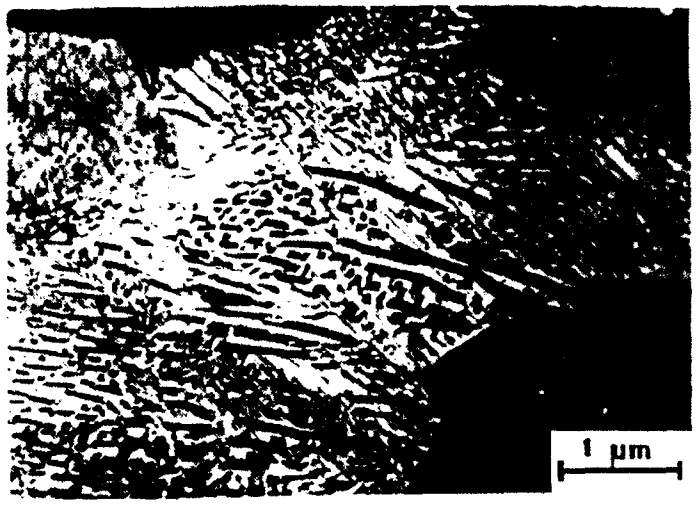


Fig. 17. Lamellar precipitates in 600°/500h material.

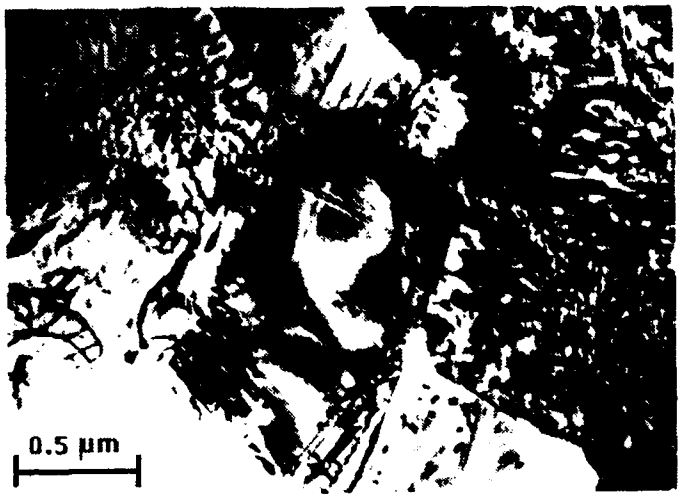


Fig. 18. Orthorhombic grains at prior β grain boundary.

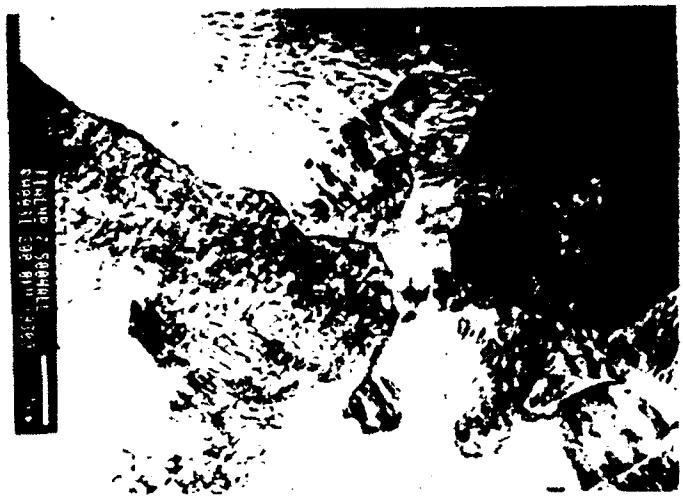


Fig. 19. General view of 700°/500h material.

Fig. 20. Discontinuous precipitation in orthorhombic phase at prior β grain boundary.

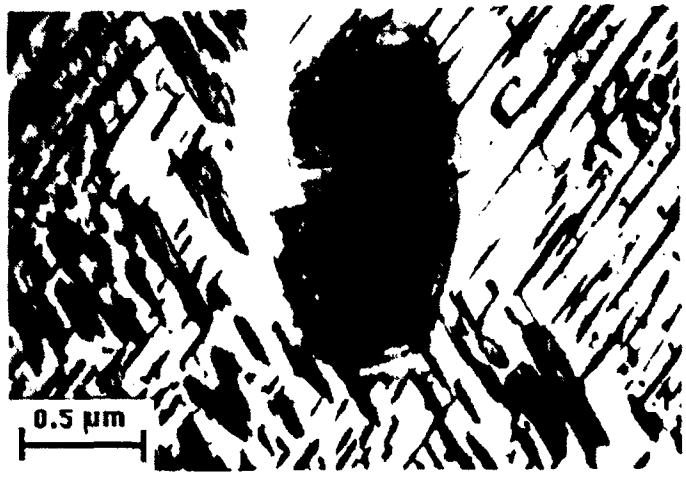


Fig. 21. α_2 grain and lamellar precipitates in orthorhombic matrix.

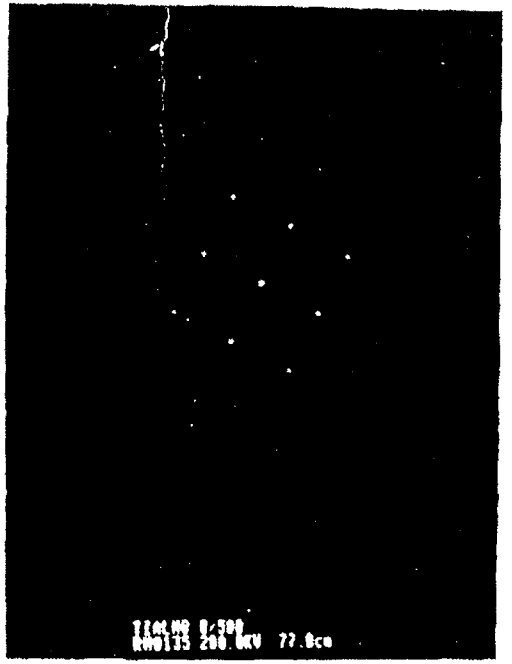
Fig. 22(a). General view of 800°/500h showing needle-like precipitates in orthorhombic matrix.

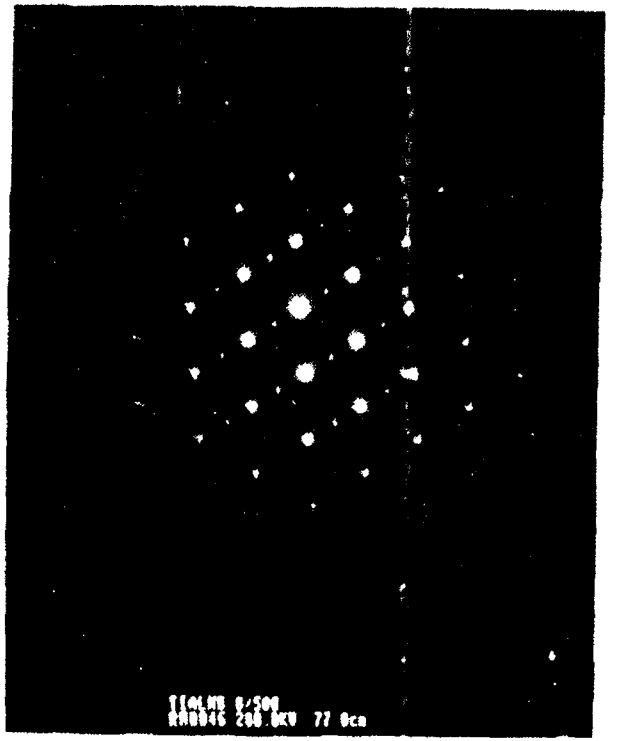
Fig. 22(b). Higher magnification view showing the apparently needle-like precipitates (800°/500h).

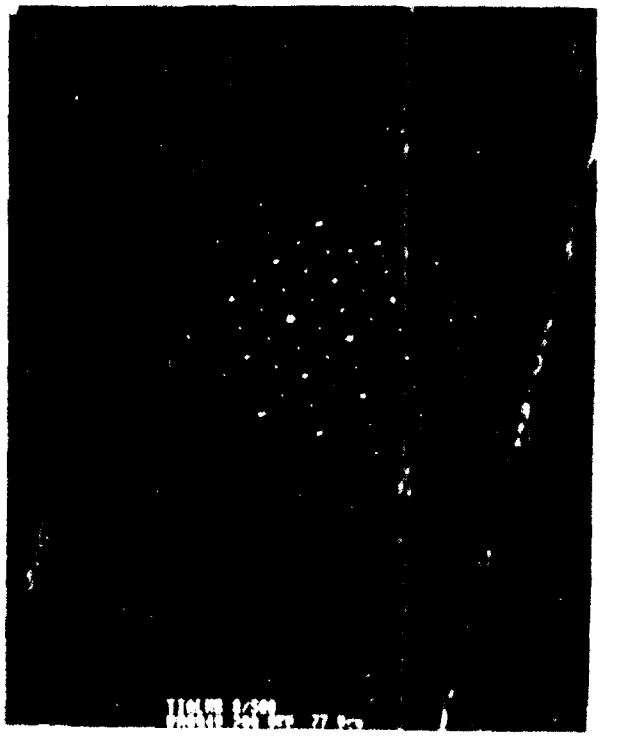
Fig. 23. Apparent fine-scale transformation within the orthorhombic matrix.

Fig. 24. Needle-like precipitates after 800°/1000h.

Fig. 25. Type of small crystal occurring at orthorhombic grain boundary triple points, etc..







INVESTIGATION OF THE COMBUSTION CHARACTERISTICS OF SWIRLED INJECTORS IN A CONFINED COANNULAR SYSTEM WITH A SUDDEN EXPANSION

Paul O. Hedman, Professor
Department of Chemical Engineering

David L. Warren, Master Candidate
Department of Mechanical Engineering

Brigham Young University
Provo, Utah 64602

Final Report for:
Summer Research Program
Aero Propulsion and Power Laboratory
Wright-Patterson AFB, OH 45433

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

September 1992

INVESTIGATION OF THE COMBUSTION CHARACTERISTICS OF SWIRLED INJECTORS IN A CONFINED COANNULAR SYSTEM WITH A SUDDEN EXPANSION

Paul O. Hedman, Professor
David L. Warren, Master Candidate
Departments of Chemical and Mechanical Engineering
Brigham Young University

ABSTRACT

This report contains a brief summary of work done during the 1992 AFOSR summer faculty research program to investigate the flow and combustion characteristics of a burner designed to "specifically reproduce recirculation patterns and LBO processes that occur in a real gas turbine combustor." The Pratt & Whitney Task 150 Combustor uses a swirling fuel injector from an actual turbojet engine installed in a sudden expansion combustor which closely simulates the geometry of a jet engine combustor. The Task 150 burner has been configured so the geometry around the injector is nearly axisymmetric, but incorporates quartz windows permitting optical (laser-based) measurements to be made in the flame.

The primary effort during this summer's AFOSR sponsored research program for faculty and graduate students was to use OH^- laser induced fluorescence (LIF) imaging to investigate OH^- ion concentrations and laser Doppler anemometry (LDA) to make gas velocity measurements. These two techniques were employed in both the Task 100 and Task 150 combustors. Preliminary analysis of the Task 150 OH^- images, combined with flow split information obtained during last year's summer research program, revealed basic mixing patterns. The images also further defined characteristic flame shapes previously measured. LDA measurements have quantified the axial, radial, and tangential velocity components in the combustor. Preliminary analysis of iso-axial velocity contours have identified the major recirculation zones. This data will allow streamlines to be determined which will give a good understanding of the flow field within the combustor. This data is also useful for model validation. Measurements of the fuel equivalence ratio at lean blow out as a function of air flow rate with nitrogen dilution were also obtained with the Task 100 burner.

INVESTIGATION OF THE COMBUSTION CHARACTERISTICS OF SWIRLED INJECTORS IN A CONFINED COANNULAR SYSTEM WITH A SUDDEN EXPANSION

Paul O. Hedman and David L. Warren

I. INTRODUCTION

This report presents a brief summary of results of an investigation to determine the flame characteristics of a single swirling fuel injector from an actual Pratt & Whitney jet engine installed in a burner with a sudden expansion (Pratt & Whitney Task 150 combustor). This work provides a bridge between the combustion characteristics of confined, coannular fuel and air jets discharged into a sudden expansion (Pratt & Whitney Task 100 combustor) and the characteristics of a linear array with four swirling fuel injectors installed in a rectangular combustion chamber simulating a segment of a real jet engine combustor (Pratt & Whitney Task 200 combustor). The advantage of the Task 150 combustor is it allows the actual combustion characteristics of a real injector to be investigated in a simple geometry where various diagnostic measurements (primarily laser-based optical measurements) can be more easily made. The Task 100 and 150 combustor configurations have been specifically developed to study the phenomenon of lean blowout (LBO) in modern annular aircraft gas turbine combustors. The combustor has been carefully designed (Sturgess, et al. 1990) to "specifically reproduce recirculation patterns and LBO processes that occur in a real gas turbine combustor."

The Task 100 combustor consists of a 29 mm diameter central fuel jet surrounded by a 40 mm diameter annular air jet. The jets are located in the center of a 150 mm diameter duct, creating a rearward-facing bluff body with a step height of 55 mm at the exit plane. The combustor test section incorporates flat quartz windows to accommodate laser and other optical access, but uses a metal shell with metal corner fillets to reduce the vorticity concentration and its effect on the bulk flow field in the combustor. This box-section combustor with corner fillets allows reasonable optical access while providing a cross section that approximates a two-dimensional axisymmetric cross section. The bluff body provides a recirculation region which stabilizes the flame.

The Task 150 combustor configuration utilizes the basic Task 100 hardware, but replaces the confined, coannular jets with an insert and a swirling fuel injector from a Pratt & Whitney jet engine. A schematic drawing of the Task 150 Combustor is shown in Figure 1. A drawing showing the installation of the fuel injector in greater detail is presented in Figure 2.

Two different fuel injectors were used for this study, a high swirl injector, and a low swirl injector. The high swirl injector used in this study was referred to by representatives of Pratt & Whitney as a "bill of materials injector" used in production engines. The low swirl injector was reported to match the characteristics of the injectors supplied by Pratt & Whitney for use in the Task 200 combustor.

The objective of the project was to determine the combustion and flow characteristics of the Task 150 burner over a range of operating conditions. Specifically, the study was to characterize the broad operating characteristics of the Task 150 burner with both high and low swirl injectors. The initial characterization (Hedman and Warren, 1991) included the following experimental work: flow meter calibration, checkout experiments, flame characterization experiments which were recorded on film and video tape, lean blow out measurements, determination of the flow partitioning between the injector passages as a

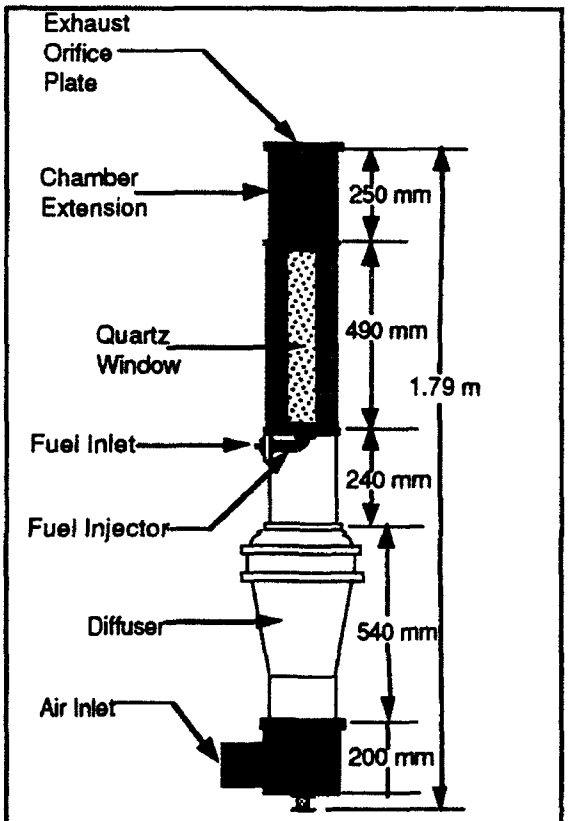


Figure 1 - Task 150 Pratt and Whitney Combustor

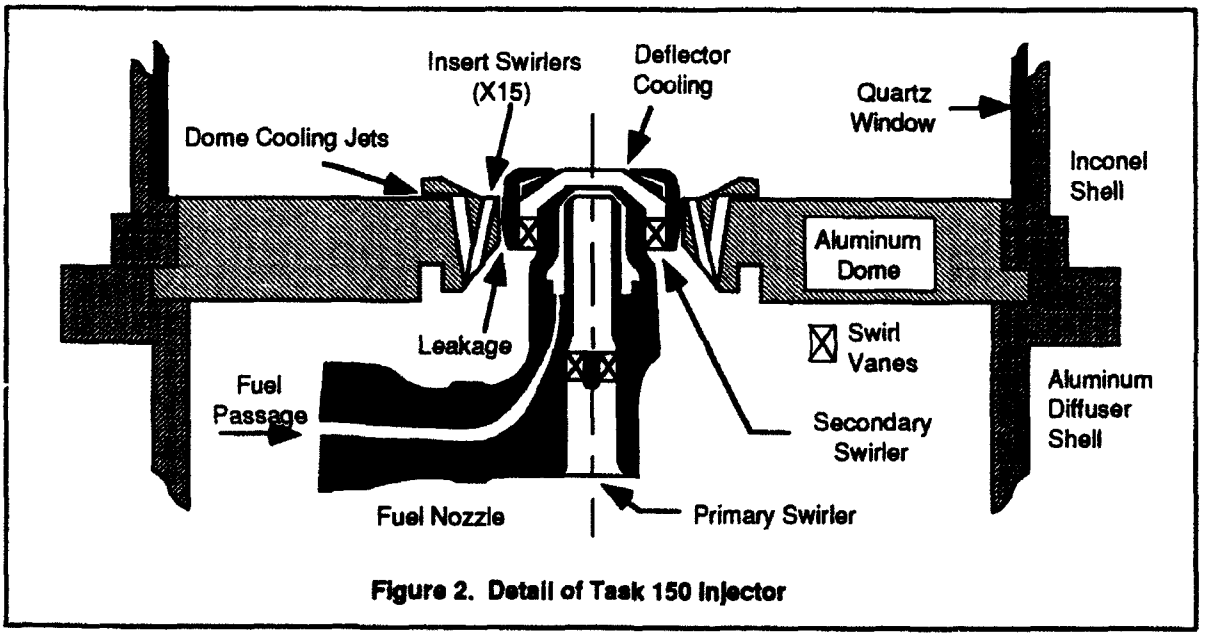


Figure 2. Detail of Task 150 Injector

function of air flow rate, and measurements of wall differential pressure and wall temperature measurements. The results of this study relate to a flame blowout modeling study being conducted by other investigators from Pratt & Whitney (Sturgess, et al. 1990). A secondary purpose of this study has been to collect data to be used in validating a computer code which predicts the flame phenomena and blowout limits (Sturgess, et al. 1990).

The experimental efforts during the 1992 summer faculty research program (this study) were focused on three areas: 1) The collection of OH^- concentrations images using laser induced fluorescence (LIF) to document the instantaneous location of the flame reaction zones, 2) The collection of extensive sets of laser Doppler anemometer (LDA) data to quantify the velocity flow fields existing in the burner, and 3) a limited amount of additional lean blow out data with nitrogen dilution in the Task 100 burner to supplement data obtained last year (Hedman and Warren, 1991) in the Task 150 burner.

The page constraints of this report prevent a detailed summary and discussion of all of the experimental results obtained. Consequently, only example results are presented. A more complete compilation of the data collected has been prepared in the form of an appendix which has been supplied to the Air Force sponsor at Wright Patterson Air Force Base. A complete analysis of the data obtained during the 1991 (Hedman, and Warren, 1991) and 1992 summer research programs (This study), and the data collected in two AFOSR research initiation grants (Pyper and Hedman, 1991; Pyper, Warren, and Hedman, 1992) will be analyzed in detail and reported in the Master of Science theses of Mr. David L. Warren (Warren, 1993) and Mr. David K. Pyper (Pyper, 1993).

II. RESULTS

Laser Induced Fluorescence (LIF) Imaging of OH^-

LIF (Laser Induced Florescence) is a single photon process. A dye laser is tuned to a frequency which causes a particular ion or molecule to fluoresce at a different frequency. Some of these emitted light photons are then recorded and their image preserved with a camera. In these experiments, OH^- ions were excited with an ultra-violet (283 nm) sheet produced by a tunable dye laser pumped with a 10 ns pulse from a Nd-Yag laser. This laser sheet passed through centerline of the reactor. An intensified CCD camera, located normal to the laser sheet, captured the 75 mm high two-dimensional uv (308 nm) image. This nearly instantaneous map of OH^- concentration was then stored by a Macintosh computer (Figure 3).

OH^- ions are commonly chosen for LIF because they are important markers in hydrocarbon flames. These ions are produced in large quantities during the combustion process, therefore being a good indicator of flame fronts. However, in some circumstances, these ions may persist long distances downstream of the

actual flame front, limiting their usefulness after the initial flame zone. Also, as in any laser diagnostic technique, there is a concentration level below which OH^- ions will not be detected. With these limitations in mind, conclusions based solely on OH^- LIF images are clearly suspect.

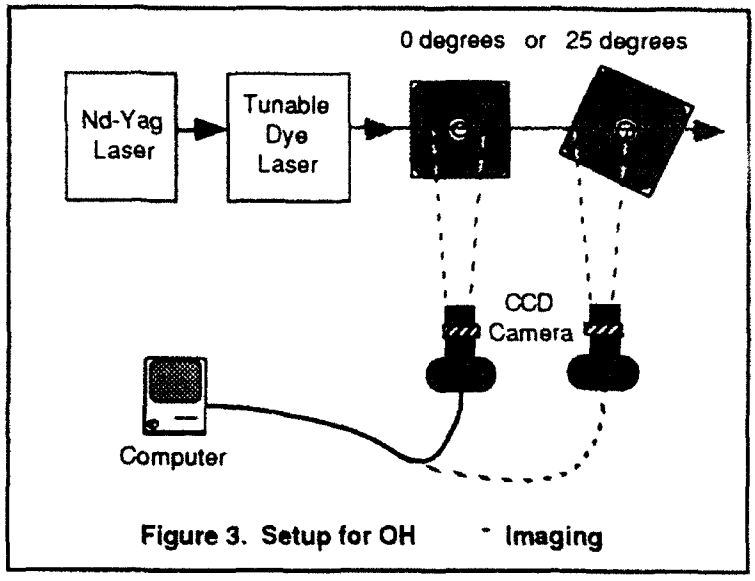


Figure 3. Setup for OH - Imaging

The experimental behavior of the Pratt & Whitney Task 100 combustor is quite different than the predictions by computational fluid dynamics (CFD) programs. These computer codes predict the flame to be anchored in the jet shear layer for all fuel equivalence ratios (Pratt & Whitney Aircraft Co., 1991). However, the flame has been experimentally observed (Roquemore, et al. 1991) to attach to the backward facing step just outside the air tube at fuel equivalence ratios (ϕ) in excess of 1.08. In these operating conditions, a small "coke bottle" shaped flame pilots a thicker flame sheet which is much lower than the CFD programs predict. The waisting is predicted by the CFD code, but the mixture is too lean to burn. Roquemore postulates a discrete and intermittent process is responsible for the entrainment of the fuel into the step recirculation zone. This type of transport would require passageways in the flame in order to deliver the unburned fuel from the fuel tube to the step recirculation zone. These passageways would appear as a region with little to no OH^- ions present.

Much of what Roquemore observed can also be found in the OH^- images collected in this study of the flame in the Task 100 burner. As he observed in the well-attached flame (ϕ=1.56), the OH^- ions appear in vortex structures being shed off the backwards step. These structures were very clear in the images collected. Once again, in the lean condition, OH^- ions were not observed below 150 mm in the reactor, with relatively small amounts between 150-200 mm and very large amounts beyond. It cannot be said Roquemore's theory was either definitively validated or contradicted by any data collected in this study. The Task 100 OH^- images were not included in this report because of limited space but will be further analyzed (Warren, 1993).

Although theoretical work for the Task 150 in all its complexity does not exist, swirl stabilized flames have been studied extensively by both empirical and theoretical methods. This research has shown the funnel-like structure shown in Figure 4. The swirl causes the gases to be flung outward. These gases are

replaced by other gases drawn from downstream. The OH^* images taken of the Pratt & Whitney Task 150 burner (Figures 5,6, and 7) dramatically illustrate these characteristics of swirling flames and the highly variable nature of the flame shape, particularly with high swirl nozzle at the lower flow rate. In addition, the effect of the insert jets is conspicuous. It is informative to correlate these OH^* images with the information known about the partitioning of the air flow rates through the various passageways of the nozzle obtained in earlier studies (Hedman and Warren, 1991). The local fuel equivalence ratios shown in Table 1 were calculated from the air flow through each of the different passageways and the total fuel flow. Implicit in these calculations are two assumptions. First, the fuel is assumed to mix uniformly within each combination of partitions before mixing with remaining air. Second, the fuel blockage effects (which would change the partitioning as a function of fuel flow) observed previously (Hedman and Warren, 1991) are assumed to be negligible. At this air flow (500 slpm), LBO occurs at a fuel equivalence ratio of about 0.5.

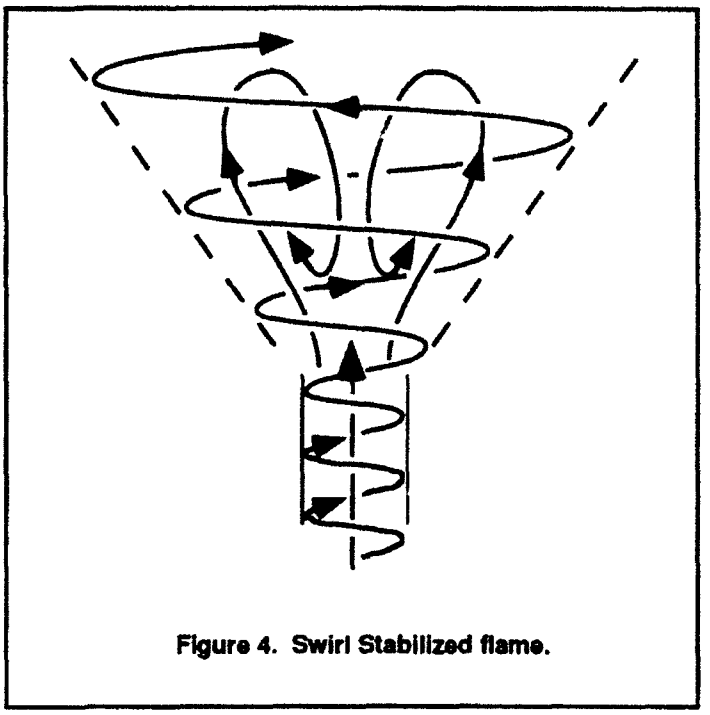


Figure 4. Swirl Stabilized flame.

Table 1. Local Fuel Equivalence Ratios. These ratios were calculated from the total fuel flow and the combined air flows for each combination.

Overall Phi	Primary Swirler Phi	Primary+Secondary Swirlers Phi	Primary+Secondary Swirlers and Insert Jets Phi
0.62	4.17	1.11	0.80
1.08	7.26	1.94	1.39
1.49	10.02	2.67	1.92

With these assumptions in mind, and knowing the flammability limits (ϕ) of propane are roughly 0.5 to 2.5, some conclusions can cautiously be applied to these images. In every case, the air in the primary swirler alone does not provide sufficient oxidizer to permit combustion. Thus, the fuel must mix with at least the secondary swirled air before combustion is possible. As shown in Figure 5, with a fuel equivalence ratio of 0.62, the funnel structure expected in a swirl stabilized flame is clearly evident. The equivalence ratio of

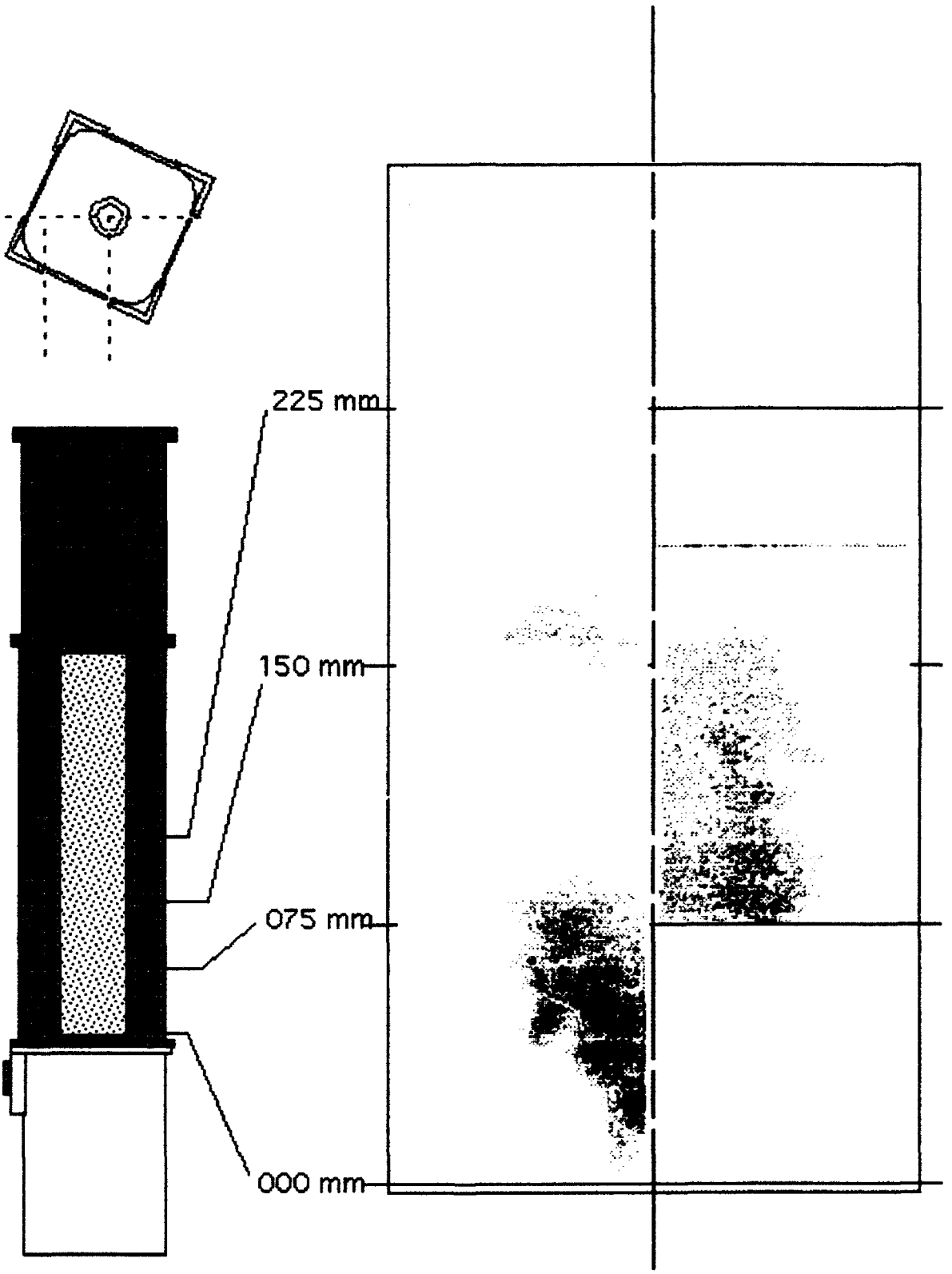


Figure 5. OH Image for Task 150-HS, Air 500 slpm, Phi 0.62

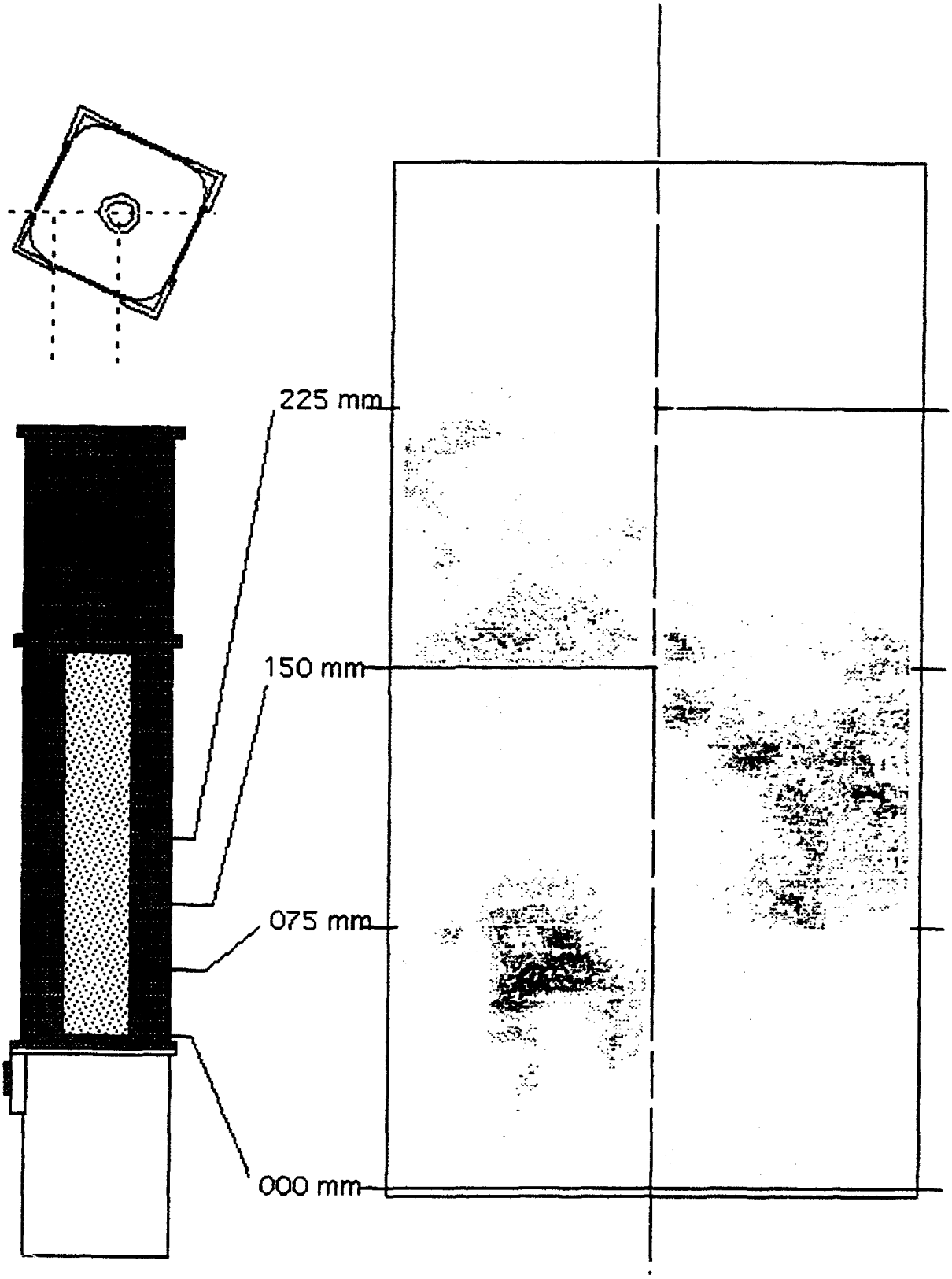


Figure 6. OH Image for Task 150-HS, Air 500 slpm, Phi 1.08

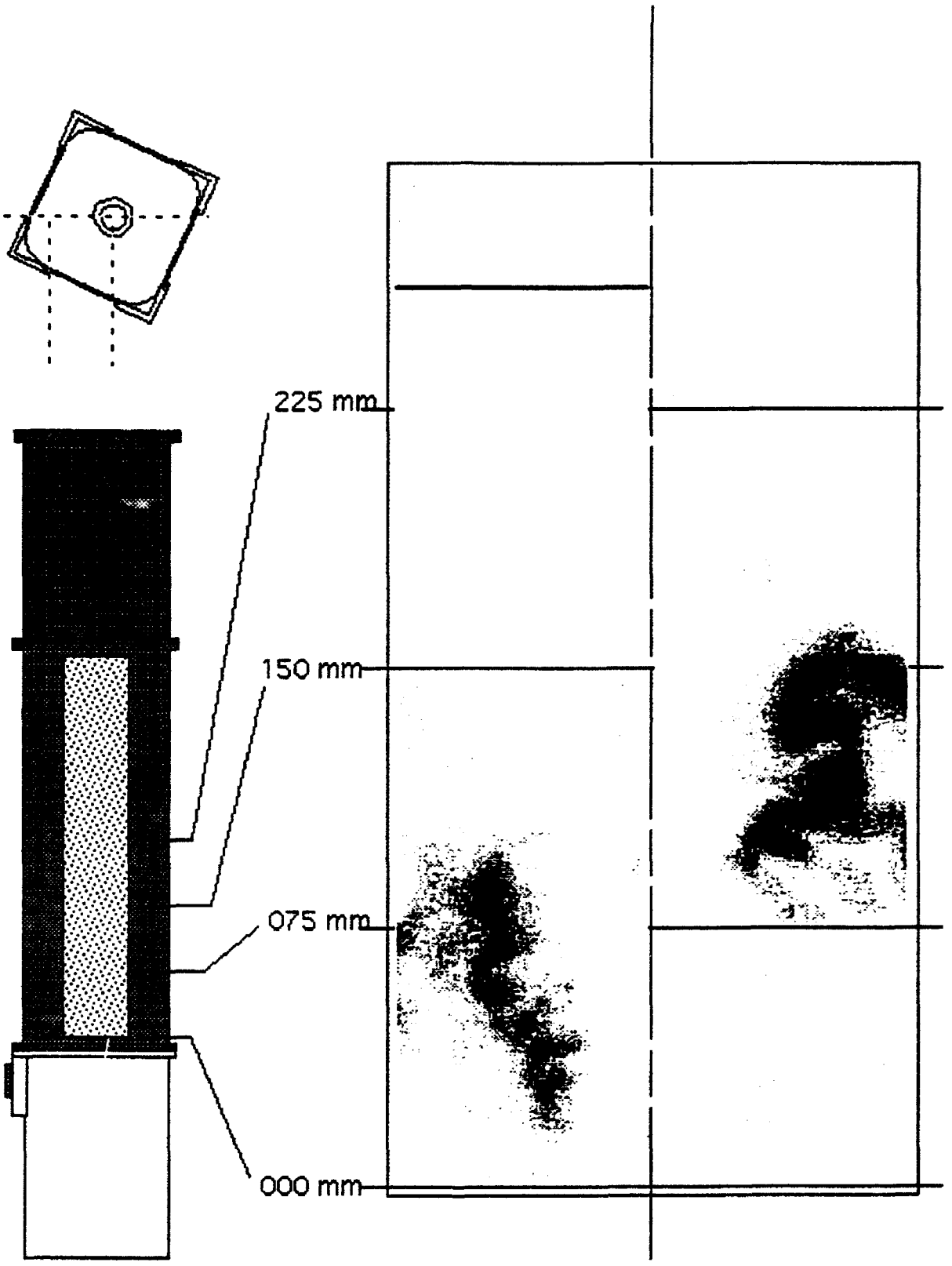


Figure 7. OH Image for Task 150-HS, Air 500 slpm, Phi 1.49

the two swirled jets is 1.11, indicating little air from the insert jets is needed to complete the combustion. Although Figure 5 shows high concentrations of OH^* ions extending above the funnel, visual observations reveal the visible flame region is apparently only a thin sheet, much like a horn, or funnel, with a rounded cusp.

As the overall fuel equivalence ratio is increased to 1.08 (Figure 6), the swirled air/fuel mixture was still within flammability limits. What changed was the amount of fuel left for to mix with the insert jets. This additional fuel, as shown in Figure 6, apparently burned on the shoulders of the funnel. Comparison with Figure 5 highlights the increase of OH^* ions in this area.

Finally, as the overall fuel equivalence ratio is increased to 1.49 (Figure 7), the fuel-rich swirled air can no longer support combustion. The characteristic funnel of a swirl stabilized flame is no longer visible. The combustion is only taking place in the presence of air from the insert and dome jets. This can be illustrated by superimposing Figure 7 with Figure 5. These two images are negatives of each other--where one is black the other is white. This supports the original assumption of the fuel mixing with each passageway in turn from the inside-out.

Laser Doppler Anemometer (LDA) Measurements

A laser Doppler anemometer (LDA) was used to make extensive measurements of gas velocity in the burner at five separate experimental conditions. The experimental conditions used are summarized in Table 2.

Table 2. Summary of Experimental Conditions for LDA Measurements

Hardware Configuration	Flow Condition	Air Flow rate slpm(70 F)	Propane Flow rate slpm (0 C)	Nitrogen Flow rate slpm (0 C)	Equivalence Ratio
Task 150-HS	Cold	500		14	0.72
Task 150-HS	Hot	500	14		0.72
Task 150-HS	Cold	500		29	1.49
Task 150-HS	Hot	500	29		1.49
Task 100	Cold	1000		23	0.59

The focus of the LDA measurements made during this year's summer faculty research program was the Task 150 burner with the high swirl injector installed. Measurements were made near lean blow-out (ϕ=0.72), and also fuel rich conditions (ϕ=1.49) with the flame attached to the dome and insert jets. LDA measurements were also obtained in isothermal, non-reactive flows where nitrogen was substituted for the propane fuel. This has allowed the effect of the flame temperature on the flow field and gas velocities to be determined at least two of the test conditions used.

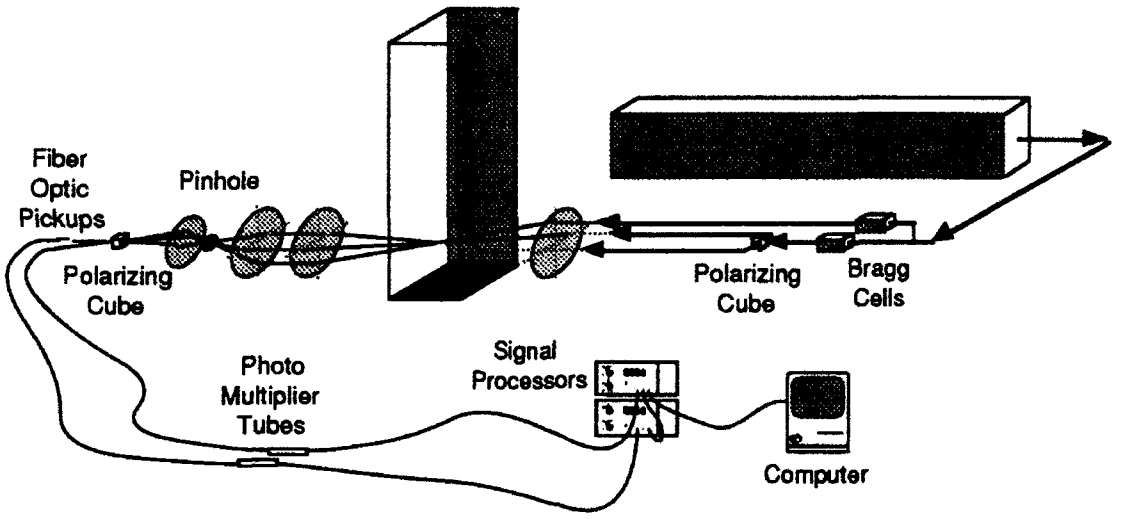


Figure 8. LDA Setup

A series of LDA measurements in the Task 100 burner were conducted as part of a previous summer faculty research program (Hedman, 1990). Subsequently, researchers at SRL, Inc. conducted a more complete set of LDA measurements on the Task 100 burner (Post and Vilimpoc, 1992). These earlier LDA data were thought to be somewhat in error because the LDA filters had not been set to measure negative velocities correctly. The problem in the LDA system was corrected for this summer's program. Consequently, one set of cold flow LDA data was taken at a geometry (Task 100) and flow condition (1000 slpm air, 23 slpm N2) which was identical to the earlier experiments (Post and Vilimpoc, 1992) in an effort to determine the extent of error in the earlier sets of LDA data. It was hoped much of the earlier data (Hedman, 1990; Post and Vilimpoc, 1992) could be appropriately used.

A schematic of the LDA experimental set-up is presented in Figure 8. The beam from the argon-ion laser was split into two beams, frequency shifted (40 and 34 MHz), polarized, and focused into a diagnostic volume in the test section. The forward-scattered LDA signals for the radial and axial velocity components were focused into fiber optic cables and passed to a photo multiplier tube to be amplified and converted to electrical signals. These electrical signals were collected with TSI, Inc counters and analyzed with a Macintosh IIfx computer. A Le Croy 9314L Quad 300 MHz oscilloscope was used to monitor the Doppler bursts to help in the alignment of the LDA system and to insure quality data was being collected. Even with careful alignment, there was still optical noise which had to be filtered using data analysis programs prepared by researchers at SRL, Inc. (Goss, 1992)

A brief experimental study was conducted to evaluate the effect of the number of points taken at a given test location on the accuracy of the gas velocity measurement. Figure 9 shows a set of axial and tangential velocity measurements taken at an axial location of 11.3 mm from about the centerline to a radial location of about 19 mm. Three different sets of data are shown, a set with 5000 points collected at each radial location, a set with 2000 points collected, and a set with 1000 points collected. In general, there was little difference observed in the mean axial and tangential velocities determined from the different number of points in the data sets. However, the fluctuating velocity components (i.e. rms velocities) are expected to be better described by the data sets containing the largest number of points. Nevertheless, for this study, it was deemed advisable to use 1000 data points at each test location. This allowed a greater number of experimental conditions and geometries to be evaluated during the limited time period afforded by the 12 week summer faculty research program. Detailed analysis of this data should give considerable insight and direction into future experiments where increased accuracy of the rms velocity data may warrant data collection with a larger number of samples.

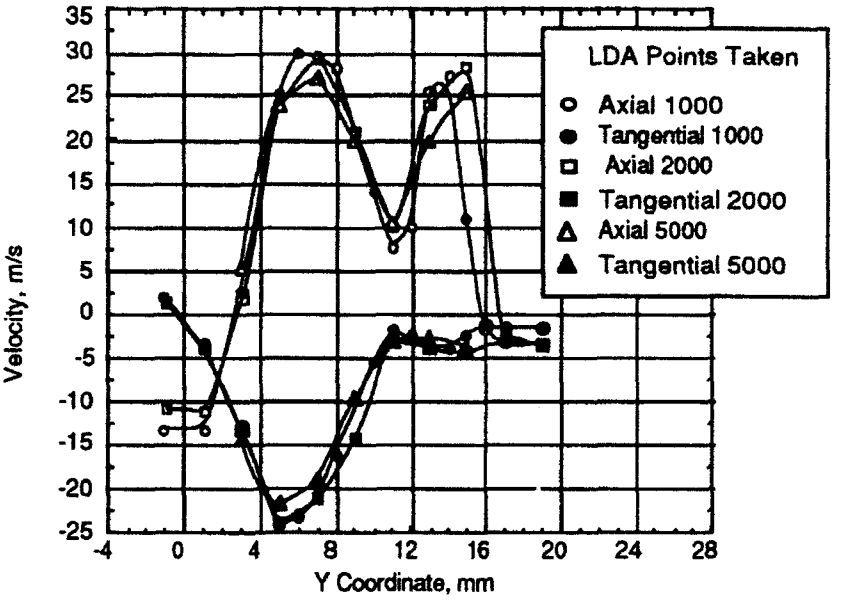


Figure 9. Effect of Number of LDA Data Points Taken (Z=11.3 mm, Task 150-HS, 29 slpm C_3H_8, 500 slpm Air, ϕ=1.49)

The axial and tangential velocity data shown in Figure 9 were measured in close proximity to the discharge of the high swirl injector. The sharp peak in axial velocity between about 2 mm and 11 mm corresponds to the outlet of the secondary swirler (See Figure 2). A corresponding strong tangential component exists in this same location. The second peak in the axial velocity between 11 and 17 mm is associated with the flow through the insert jets. Since the insert jets were set at discrete intervals around the burner, it was possible to measure velocities between the insert jets which did not show this sharp peak. It is interesting to note the tangential velocity in the region of the insert jets is significantly reduced from those directly in line with the discharge from the swirl vanes.

Two component velocity data were obtained for each of the test conditions described above (Table 2). The burner was translated with respect to the laser diagnostic volume in an X, Y, and Z coordinate system. Translation in the Z coordinate direction allowed variation in the axial location. Translation in the X or Y axis allowed variations with radius to be determined. For these tests, the X or Y translations were done along a coordinate centerline. Translation in the X coordinate direction along the Y coordinate centerline allowed axial and radial velocity data to be obtained. The edge of the windows limited translation in this coordinate direction to about ± 30 mm. Translation in the Y coordinate direction along the X coordinate centerline allowed axial and tangential velocity data to be obtained. As the diagnostic volume was brought near the quartz windows, significant optical noise was added to the Doppler signals. The quartz windows were approximately ± 76 mm from the center of the reactor. The optical noise from the windows generally limited data collection to ± 65 mm, although with especially clean windows, it was sometimes possible to get good data at ± 70 mm. Typically, data was collected at 0.5 or 1.0 mm radial increments where the velocity gradients were large. Data was collected at up to 10 mm increments where velocity profiles were relatively flat. A typical set of data was taken at axial locations of 10, 15, 20, 25, 50, 75, 100, 125, 150, 200, and 240 mm above the dome of the reactor. Occasionally, other intermediate locations were examined where large velocity gradients or other interesting behavior were found. A small error in the location of the reactor center caused some of the data to be taken at locations slightly removed from the prescribed test location. Correction of this location error has caused some of the data to be reported at other than the desired integer number test locations. The X, Y, Z positions as reported in the data sets are the corrected locations. In general, it was felt the reactor location was known to about 0.5 mm in the Y direction and about 0.1 mm in the X and Z directions. This error was mostly due to the size and shape of the diagnostic volume where the laser beams were focused. Hard copies and electronic files of all LDA data collected were provided to the laboratory focal point, Dr. W.M. Roquemore.

An example of the LDA data is presented in Figure 10. This figure presents representative axial velocity profiles collected for the Task 150-HS combustor burning 14 slpm propane with 500 slpm air (ϕ=0.72). Only six of the eleven radial plots collected are shown. Nevertheless, the sharp peaks associated with the burner in the region near the injector (discussed above) are clearly evident. The decay of the high velocity region associated with the injector as one moves from one down stream to the next is also apparent. Although the radial profile is becoming flatter at each down stream location, it has not reached the classic flat turbulent profile by the 201.3 mm location. Although not shown, the strong tangential velocities associated with the injector (See Figure 9), rapidly decay and diffuse to the outer edges of the reactor becoming nearly uniform (ca -1 m/s) across the duct by the 201.3 mm axial location.

The relatively fine grid of axial, radial and tangential velocity data will allow the nominal flow streamlines through the reactor to be estimated. This will allow the major flow structure and the recirculation patterns to be determined. However, there was not sufficient time to analyze even an example case for this report because of the reporting deadlines.

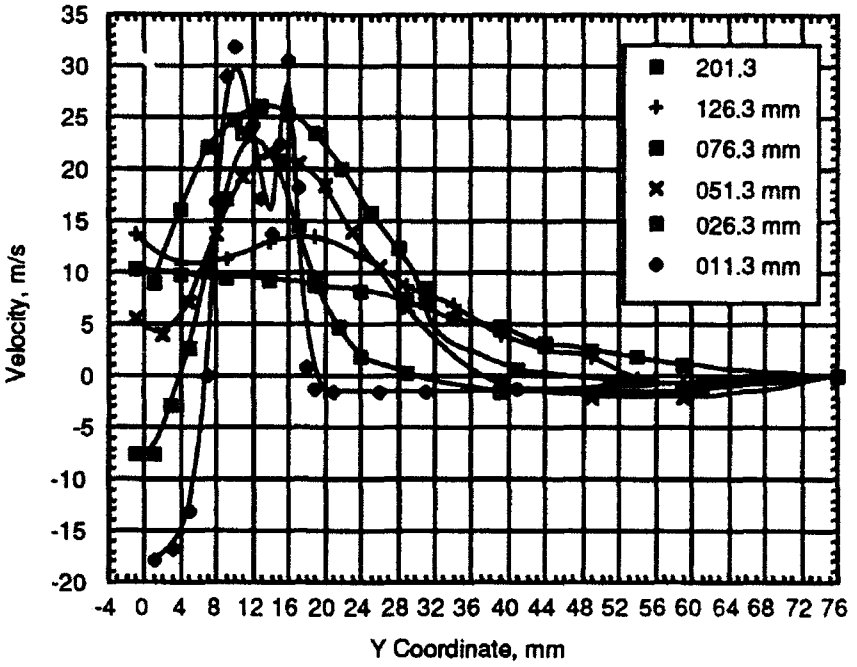


Figure 10. Example of Axial Velocity Task 150-HS, 14 slpm C_3H_8, 500 slpm Air

This analysis will be made and reported in the Master's Thesis of David L. Warren (Warren, 1993). In an effort to get some understanding of the flow patterns in the burner, an analysis of the zero axial velocity contours for one of the experimental cases (Task 150-HS, 14 slpm N_2, and 500 slpm Air) was prepared. The results are shown in Figure 11. The zero axial velocity contours are shown as solid lines on a projection of the burner and reactor chamber. Although uncertain at this point, dashed lines have been used to show the outline of the recirculation zones. This analysis shows four separate recirculation zones in the reactor. There is a recirculation zone located directly over the injector, which seems to be associated with the swirlers. There is another recirculation zone at the bottom edge of the reactor driven by the dome jets. A major recirculation zone is associated with the zero axial velocity line the begins at the edge of the injector and angles upward to the wall of the reactor. These recirculation zones are consistent with the visual observations and video images taken of the reactor (Hedman and Warren, 1991). There is also a fourth recirculation zone located high up in the reactor. The velocities in this zone are rather low, but it clearly exists and brings fluid from the outer edges of the reactor back to nearly the center of the reactor. The recirculation zone near the primary swirler and downstream recirculation zone create a region of nearly zero axial velocity between them which may contribute to the stabilization of the flame.

Lean Blow Out Experiments

An extensive set of lean blow out (LBO) experiments were conducted during the previous summer faculty research program (Hedman and Warren, 1991) in the Task 150 configuration, with and without the addition of nitrogen to the air stream. The added nitrogen affects the loading parameter, and also simulates a reduced ambient pressure (i.e. a reduced partial pressure of oxygen). A limited set of lean blow out experiments were conducted in this summers program in the Task 100 burner in order to complete the data set. Figure 12 presents the results of this LBO investigation. Lean blow outs were measured at air flow rates of 500, 1000, 2000, and 3200 slpm air flow rate with various levels of nitrogen dilution up to about 870 slpm. The test facility limited the maximum amount of nitrogen which could be flowed without major facility modification to about 900 slpm. This data will be used to complement data being taken at Brigham Young University (BYU) under last year's research initiation grant (Pyper, Warren, and Hedman, 1992). The fuel equivalence ratio at lean blow out without nitrogen are consistent with the recent data taken at BYU, but are somewhat higher (ca 0.55 to 0.65) compared to previous data (0.45-0.55) taken at Wright Patterson AFB (Hedman, 1990). Investigation into these differences is continuing.

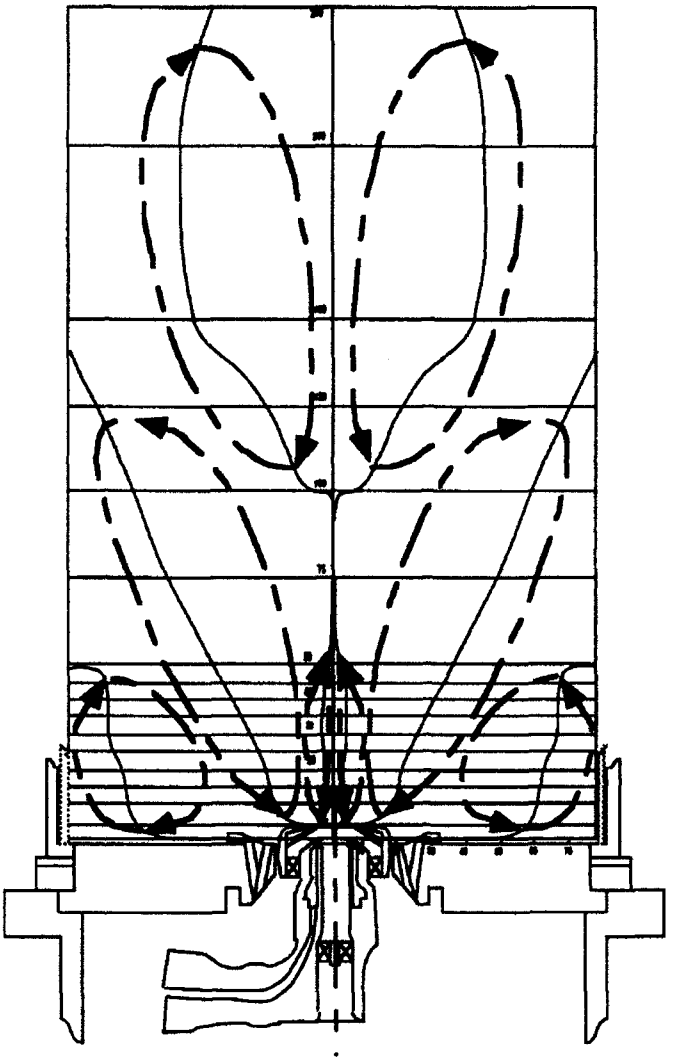


Figure 11. Zero Axial Velocity Contours, Task 150-HS, 14 slpm N2, 500 slpm Air

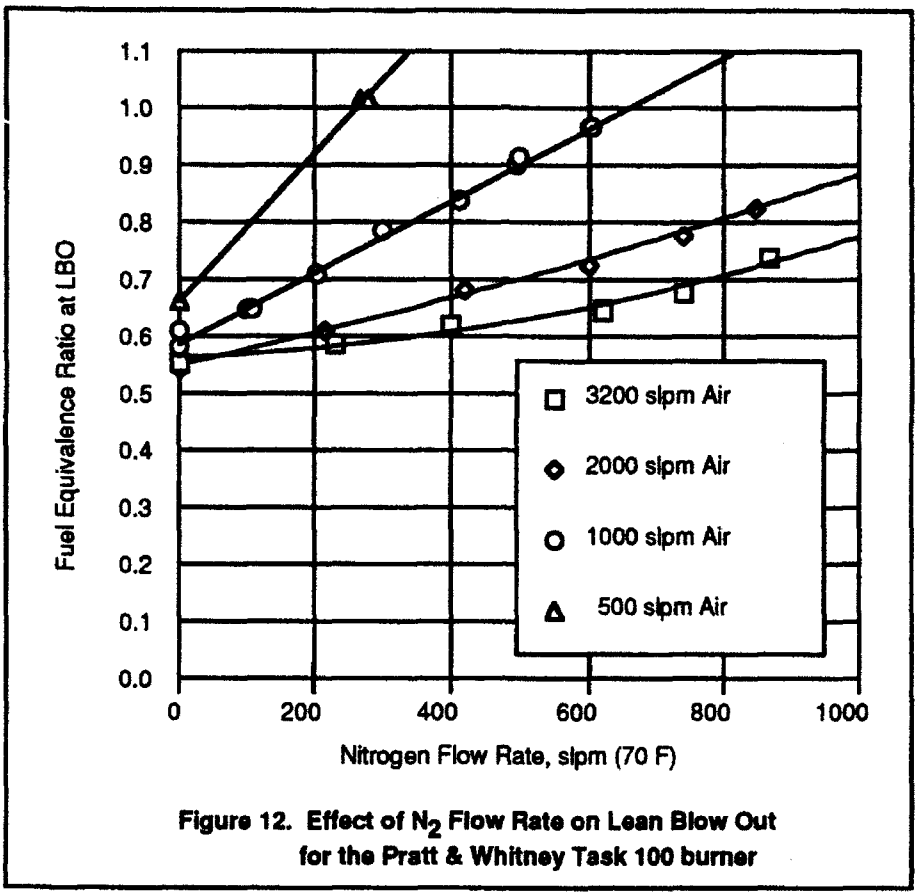


Figure 12. Effect of N_2 Flow Rate on Lean Blow Out for the Pratt & Whitney Task 100 burner

III. RECOMMENDATIONS:

Suggestions for Follow-on Research.

There have been considerable insights into the operational characteristics gained from this summer's faculty and graduate student research program, the previous two summer's faculty research programs (Hedman, 1990; Hedman and Warren, 1991), and from the two companion research initiation studies (Pyper and Hedman, 1991; Pyper, et al., 1992) already completed. There is yet much to be done before a full understanding of the burner is achieved. There are several suggested investigations to further these insights. Additional work needs to be done on the flow partitioning and fuel blockage effects in the low and high swirl nozzles. The air flow rates to the primary and secondary swirlers were thought to have a major impact on the operation of the combustor. Much needed data on this characteristic could be obtained if the facility were modified to allow the air flow to the various injector passages to be independently controlled.

Work needs to be done to quantify the volume of a flame and the amount of air and fuel which is actually entering the flame zone. This is needed in order to define an appropriate loading parameter for the Task 150 burner, and to better determine the actual fuel equivalence ratio in the "well mixed" flame zone. Additional laser sheet lighted images (Mie scattering from a seeded flow) would be very useful to provide qualitative mappings of the flow fields. This information is essential in order to better understand the characteristics of the eddies and other turbulent structures. Local gas temperature data and mixture fraction measurements from the Task 100 and/or Task 150 Combustors would be very useful information for comparison to model predictions. The temperature data could be obtained using the CARS diagnostic system and Burner already in place at BYU. Gas mixture fraction is difficult to obtain with laser-based diagnostic techniques, but could be done with gas sampling probes. This area needs further investigation.

Eventually, it would be desirable to investigate the feasibility of introducing a liquid fuel into the Task 150 combustor. The facility modifications to accomplish this task would be very simple. Whether the windows would remain clear enough for laser diagnostics when operating with a potential sooting fuel is uncertain and needs to be investigated.

Application for a Mini Research Initiation Grant.

An application for a mini research grant will be prepared as a result of the summer fellow research program. The proposed project would include the measurements on the Task 150 provided to BYU and installed under previous research initiation grants (Pyper and Hedman, 1991; Pyper, et al., 1992). The proposed study will include: additional flame characterization measurements using film photographs and video images, CARS gas temperature measurements, sheet lighted film photographs from MIE scattering from a seeded flow, and investigation into the feasibility of using liquid fuels in the Task 150 combustor. The proposed BYU project would fill a need to investigate the swirling burner in a single burner configuration. In addition, to basic combustion measurements, a future program at BYU could begin to investigate the performance of advanced combustor designs, and the formation of NO_X pollutants at the higher fuel equivalence ratios associated with higher performance gas turbines.

ACKNOWLEDGMENTS

We wish to thank the Aero Propulsion and Power Laboratory, Wright Patterson Air Force Base, and the Air Force Office of Scientific Research for sponsorship of this research. The timely help of Mr. Gary Moore and other individuals at Research and Development Laboratories, Inc. is greatly appreciated. Special recognition of the help of Mr. Milton H. Danishik (RDL) is acknowledged. Not only did he provide extraordinary help in providing information about living accommodations and base information, but he became a good friend as we shared the same office.

The very successful experience gained during this summer research program is in large measure due to the very excellent support of my sponsor, Dr. W. M. (Mel) Roquemore (WRDC/POSF). The help of K. Y. Hsu from the University of Iowa in making the OH^- image measurements is gratefully acknowledged. Special thanks are also given to Mr. Dale Shouse and Mr. Melvin Russell who provided much needed support with instrumentation and the burner and facility modifications. The very well-qualified research staff at Systems Research Laboratory, Inc. (SRL) is gratefully appreciated. Special thanks are in order for Dr. Larry Goss, Mr. Benjamin Sarka, and Mr. Daryl Trump of SRL. Their willingness to help, to answer questions, and the help they provided made this summer's effort truly rewarding.

IV. REFERENCES

Goss, L.P., Personal Communication, Systems Research Laboratories, Inc., Dayton, Ohio (1992)

Hedman, P.O., "Investigation of the Combustion Characteristics of a Confined Coannular Jet with a Sudden Expansion," Final Report, 1990 AFOSR Summer Faculty Research Program, Universal Energy Systems, Inc., Dayton, Ohio (1990)

Hedman, P.O.and Warren D.L., "Investigation of the Combustion Characteristics of Swirled Injectors in a Confined Coannular System with a Sudden Expansion," Final Report, 1991 AFOSR Summer Faculty Research Program, Research and Development Laboratories, Inc., Culver City, California (1991)

Post, M.E. and Vilimpoc, V., Personal Communication, Systems Research Laboratories, Inc., Dayton, Ohio (1992)

Pratt & Whitney Aircraft Co., "Combustion LBO Model Evaluation R&D Status Report and Program Schedule," Report Number FR 20204-13, Air Force Contract Number F33615-87-C-2822 CDRL 1, Quarterly R&D Status Report, Pratt and Whitney Government Engine Business, West Palm Beach, Florida, (January 25,1991)

Pyper, D.K. and Hedman, P.O., "Investigation of the Combustion Characteristics of Confined Coannular Swirling Jets with a Sudden Expansion," Final Report, 1990 AFOSR Research Initiation Program, Universal Energy Systems, Inc., Dayton, Ohio (1991)

Pyper, D.K. Warren, D.L., Blackham, M.S., Christensen, J., and Hedman, P.O., "Investigation of the Combustion Characteristics of Confined Coannular Swirling Jets with a Sudden Expansion [Task 150 Combustor]," Final Report, 1991 AFOSR Research Initiation Program, Research and Development Laboratories, Inc., Culver City, California (1992)

Pyper, D.K., Master of Science Thesis, Mechanical Engineering Department, Brigham Young University, Provo, Utah (in progress) (1993)

Roquemore, W.M., Reddy, V.K., Hedman, P.O., Post, M.E., Chen, T.H., Goss, L.P., Trump, D., Vilimpoc, V., and Sturgess, G.J., "Experimental and Theoretical Studies in a Gas-Fueled Research Combustor," AIAA 91-0639, 29th Aerospace Sciences Meeting, Reno, Nevada (January 7-10, 1991)

Sturgess, G.J., D.G. Sloan, A.L. Lesmerises, S.P. Henneghan and D.R. Ballal. (June 1990), "Design and Development of a Research Combustor for Lean Blowout Studies", 35th International Gas Turbine and Aeroengine Congress and Exposition. Brussels, Belgium.

Warren, D.L., Master of Science Thesis, Mechanical Engineering Department, Brigham Young University, Provo, Utah (in progress) (1993)

STRESS WAVE PROPAGATION THROUGH THE THICKNESS OF GRAPHITE/EPOXY LAMINATED PLATES USING PVDF SENSORS

David Hui
University of New Orleans
Dept of Mechanical Engineering
New Orleans, LA 70148

Final Report for
Summer Research Program
Wright Laboratory

sponsored by
Air Force Office of Scientific Research
Wright Patterson AFB, Ohio 45433

ABSTRACT

The object of this study was to determine the stress wave (or pulse) propagation through the thickness of a graphite-epoxy laminated plate. This was part of an overall study to understand the damage of these plates under normal projectile impact. Upon a sharp impact by a tiny spherical steel ball, the stress wave propagated from the impact point into the rest of the material. It was found that the embedded polyvinylidene fluoride (PVDF) sensors enabled prediction of the wave velocities and wave attenuation.

INTRODUCTION

The impact behavior of laminated plates is an important topic because composite plates are known to respond to impact loading and energy dissipation in a very different way than metallic plates. In fact, impact resistance is one of the most serious weaknesses of composite material plates. Excellent recent survey articles on this topic was reported by [1] and [2]. The impact of metallic plates by spherical balls was reported by [3], [4] and more recently by [5].

The concept of embedded strain gages was employed by [6, 7] to study the deformation and damage of composite laminates under impact loading. The characteristic features of the strain records are associated with specific failure modes of the laminates. The load history, imparted energy and transient strains at various locations through the thickness were obtained. Wave propagation in transversely impacted composite laminates was obtained by [8] and [9].

The objective of this work is to examine the wave velocities and wave attenuation in the thickness direction using the PVDF sensors that are embedded in the interior of the laminate. Upon a sharp impact, the stress wave propagates from the impact point into the rest of the material. Immediately below the impact point over a small area, the stress wave can be assumed to propagate with a plane front in the thickness direction. A plane pressure sensor of relatively small dimension

Dr. David Hui, University of New Orleans, Dept. of Mechanical Engineering, New Orleans, LA 70148

Dr. Piyush Dutta, U.S. Army Cold Regions Research and Engineering Laboratory, 72 Lyme Road, Hanover, NH 03755

will respond to this propagating stress wave front. A series of such sensors embedded in various depths in the interfaces of the laminate will respond to this incoming stress pulse in the sequence at which the sensors meet the pulse. A measurement of the time difference between the start of successive pulses, divided by the distance between the sensors, would give the velocity of the stress wave between sensors.

The present impact problem is concerned with an extremely short duration pulse so that the wavelength is short relative to the thickness of the individual lamina. Such short-wave length pulse is especially needed in laminates that contain relatively few layers, because one needs to examine the reflection, transmission and superposition of wave through a laminated plate.

The present work is concerned with the impact of tiny spheres on laminated plates. Such impact does not cause damage of the composite plates. The validity of the techniques of using embedded sensors is demonstrated. Further, knowing the velocity of the stress wave will enable one to measure the Young's modulus in the thickness direction, which is known to be hard to measure. The drop test enables one to determine the compressive stresses of a particular lamina as the wave propagates toward the free back surface and the superposition of the tensile stresses as the wave is reflected from free back surface.

The use of piezoelectric polymer as a material that transforms an electric field to a small mechanical deformation directly through a readjustment of internal polarization is well known [10]. The polyvinylidene fluoride (PVDF) piezoelectric sensors are embedded in the interior of the laminated plate. The purpose of these sensors is to enable one to "look" inside the composite specimens and to determine the sequence and propagation of the stress waves. Of particular concern is the duration of the stress wave as it crosses a lamina due to impact and the reflection of waves from the back free surface. It appears that the data collection instruments (CREATEC and NICOLET) were sensitive enough to measure the stress pulse through the charge in the PVDF and hence the force applied to the sensor within the laminate.

The wavelength of the pulse is assumed to be short relative to the lamina thickness but long relative to the diameter of the individual fiber. Thus, the material is governed by the effective properties of the equivalent homogeneous material. The "interface" effects are neglected and the analysis is thus identical to that of the homogeneous material.

The laminated plate consists of four sets of layers (each set consists of seven layers). The plate dimensions are 4 in. by 4 in. The plate is clamped in a fixture so as to become a circular plate with a diameter of 3 in. The fixture was used in previous experiments involving Hopkinson bar tests [11].

WAVE SPEED AND ATTENUATION

As a first approximation, the "interface" effects are neglected and the composite plates can be modeled as transversely isotropic materials. The wave speed in the thickness direction [9] is

$$v = \left[\left(c_{13t}^2 + c_{23t}^2 \right) / 2 \right]^{1/2} \tag{1}$$

where $c_{13t} = \left(G_{13/\rho} \right)^{1/2} \approx \left(G_{12/\rho} \right)^{1/2}$ (2)

and $$c_{23t} = \left(G_{23/\rho}\right)^{1/2} \approx \left(\frac{E_{22}}{2\rho(1+\nu_{23})}\right)^{1/2} \quad (3)$$

In the above, E_{22} is Young's modulus in the direction perpendicular to the fiber, G_{12}, G_{13} and G_{23} are the shear moduli, ν_{23} is Poisson's ratio and ρ is the density of the material. Further subscripts, 1, 2, 3, refer to the fiber direction, in-plane perpendicular to the fiber and out-of-plane perpendicular to the fiber, respectively. Note that the above velocities are independent of the frequency so that for plane waves, a pulse shape composed of a spectrum of frequencies can propagate without distortion of its shape. The laminae are manufactured with a 28-layer stacking sequence (0_7, 45_7, 90_7, 0_7).

The experimental wave speed was measured by any two of the three embedded sensors. The wave speed, measured by the two sensors that are seven plies apart and with the pulse duration of 0.5 ms (that is, 500 ns) can be found from

ν = lamina thickness/propagation time
= (0.041/12)/(500x10^{-9})
= 6833 ft/sec

For an eight-layer stacking sequence of $(0_2/\pm45)_s$ graphite epoxy specimens, the wave speed in the thickness direction was calculated by [9] to be 6525 ft/sec. This value agrees well with the present experimental result. The wave propagation is accompanied by attenuation of its amplitude for three major effects: (1) geometric attenuation, (2) interfacial friction between fibers and matrix and (3) the interlaminar friction between adjacent plies. Geometric attenuation is due to the spreading out of the wave in a spherical direction starting from the point of impact. Normally, the geometric attenuation is predominant over the remaining effects. However, since the distance of wave propagation in the thickness direction is so small compared to the planar direction, it is not necessarily the dominant factor.

The dispersion effects are expected to be small because all layers (with different stacking orientation) are identical in the thickness direction. Careful calibration of the PVDF sensor is needed. Since the forces on the PVDF sensors are proportional to the charge, proportional constants are determined that are a function of the applied force.

The analysis is confined to the transient stage since one wishes to determine the process of damage in the first crucial nanoseconds. The subsequent vibration problem in the transition from a transient to steady-state stage may also be important. Since one is interested in the initial process of damage, particular emphasis is placed on the wave velocities and attenuation in the transient stage. The subsequent steady-state vibration problem is also of interest as it gives the complete damping process involving wave attenuation; this will be investigated in the future.

EXPERIMENTAL WORK

The experimental setup for the drop test is shown in Figure 1. The sample square plate is clamped in a fixture to produce a circular plate with diameter being 3 in. A schematic diagram for the impactor and the clamped circular plate is shown in Figure 2.

The PVDF sensors of 28 μm (1 μm = 10^{-6} m) thickness are embedded at every seven-ply interval where adjacent plies change the orientation. Two thin lead wires of diameter 41 standard gauge are soldered onto each PVDF sensor and extended out to the edge of the plate and are

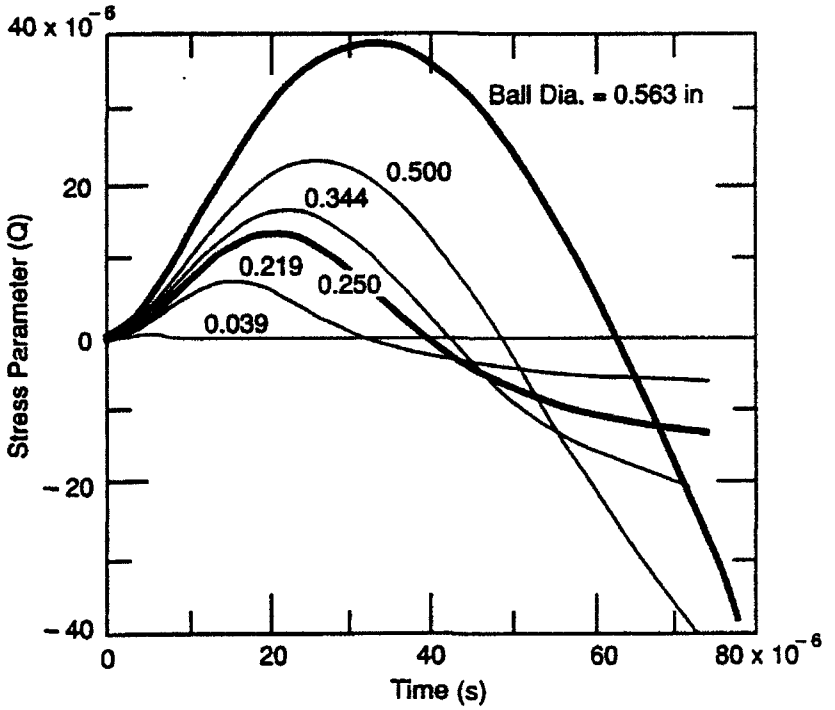


Figure 3. Stress parameter vs. time for different ball diameters.

connected to the digital oscilloscope. The velocity measurements are made by monitoring the pressure sensed by two consecutive sensors with a high resolution digital oscilloscope capable of sampling at 50-ns intervals. Attenuation and energy of the stress pulses are monitored with a four-channel oscilloscope having a sampling rate of 500 ns.

The stress waves are generated by dropping spherical steel balls of six different sizes (0.038-in., 0.219-in., 0.250-in., 0.344-in., 0.500-in., and 0.563-in. diameters). The resulting stress vs. time curves for each of these balls are presented in Figure 3. The wave forms are recorded on the high speed digital oscilloscope. The repeatability of the wave forms from the PVDF sensors from eight consecutive drop ball impacts is demonstrated in Figure 4.

RESULTS

Figure 5 shows the front part of the plot of two waves generated by the drop of a 0.038-in.-diam. steel ball. The time difference between the start of the two stress waves shown in Figure 5 is the time that the wave has taken to propagate from the first sensor to the next sensor through a thickness of seven plies (0.041 in.). The velocity is computed by dividing this distance by the time interval.

Figure 6 shows the amplitude decay of the stress wave as it propagates through the three consecutive sensors located at seven-ply intervals. It can be seen that as the wave propagates, the peak amplitude decays. These tests are repeated with six different sizes of steel balls. For each ball, five measurements of the wave form decay are made. This figure depicts the wave form recorded from the drop of a single size ball (0.039 in.). One vertical division represents 0.292×10^{-3} psi and a horizontal division represents 1.953 μs.

Figure 1. Experimental setup for drop ball stress waves measurement.

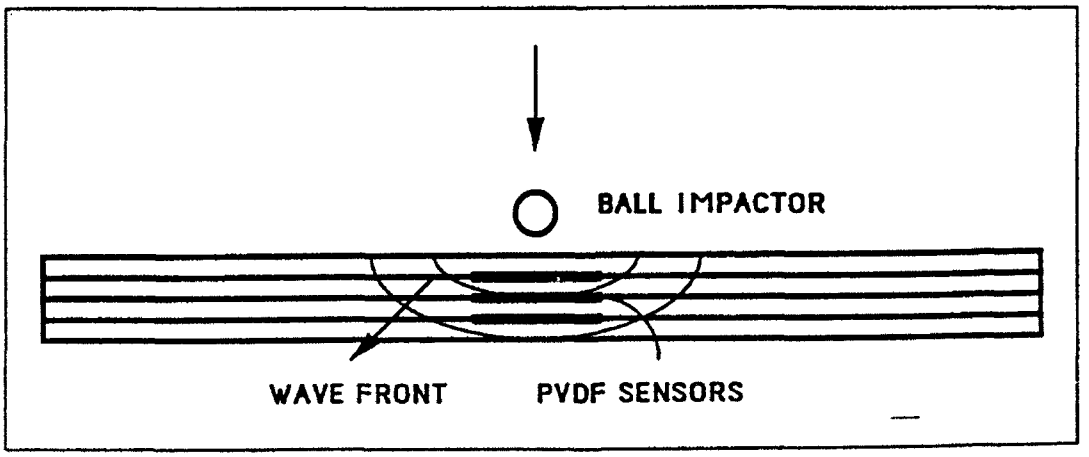


Figure 2. Schematic diagram of the embedded sensors and the clamped circular plates and typical stress versus time as sensed by each of the embedded sensors.

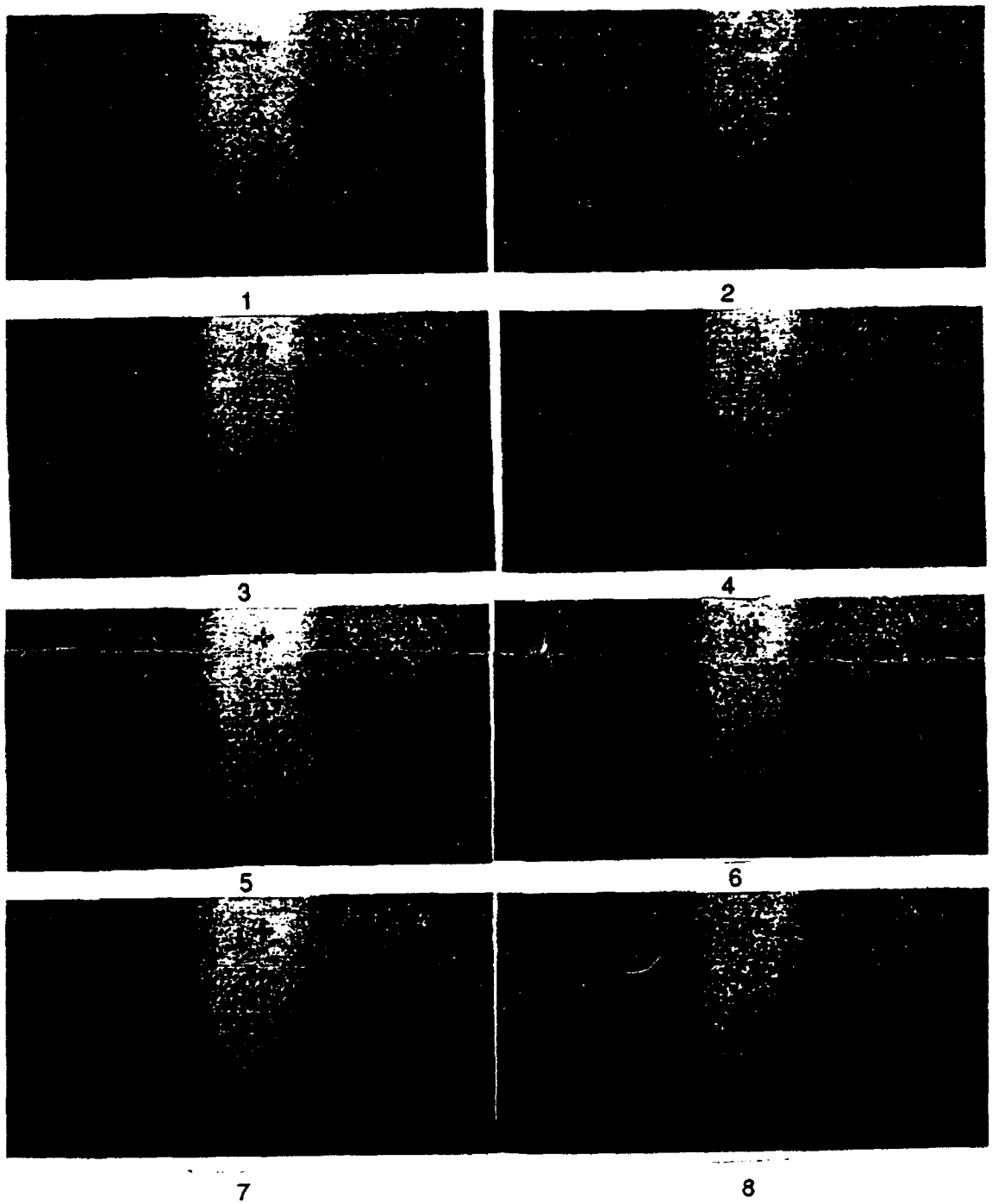


Figure 1. Wave forms from the PVDF sensors show repeatability of the test in eight consecutive drop ball impacts.

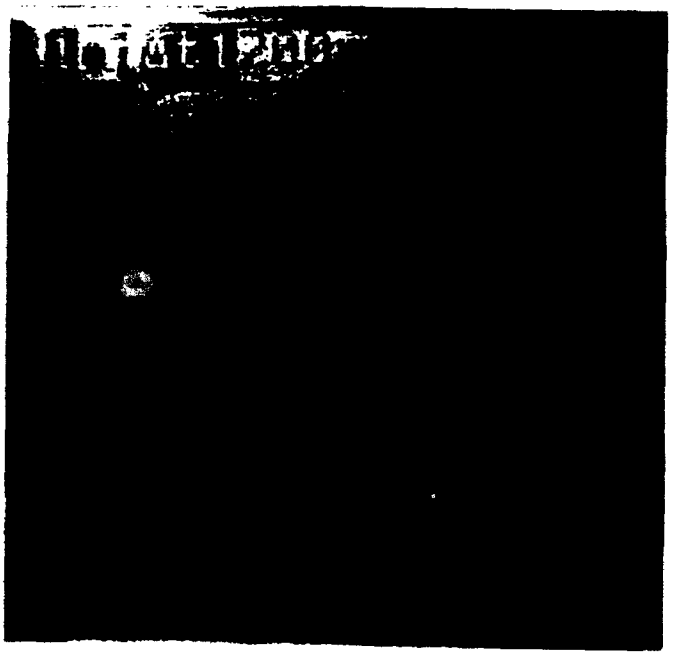


Figure 5. An enlarged photograph from the oscilloscope showing the delay in the starting times from the two sensors.

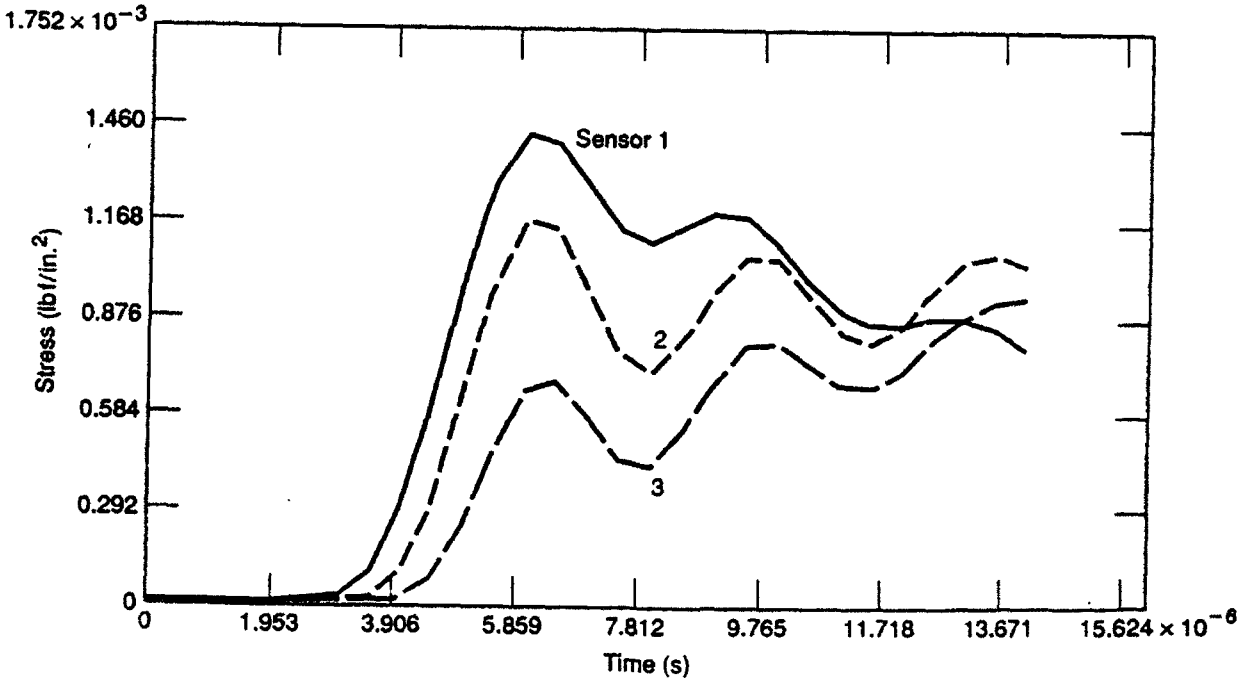


Figure 6. Amplitude decay of the stress wave as it propagates through the three consecutive sensors located at seven ply intervals.

CONCLUSIONS

The embedded PVDF sensors were found to be effective for studying the wave propagation in the thickness direction of the laminated plates. The interference from reflected waves can be minimized or even eliminated by reducing the size of the impactor balls. The method provides a way to determine Young's modulus in the thickness direction from the measured wave speed. This method also allows an estimation of the wave attenuation in the thickness direction.

ACKNOWLEDGMENTS

This work was financially supported by the Wright Laboratory (Wright Patterson Air Force Base, Ohio) AFOSR Summer Research Grant and by the U. S. Army Cold Regions Research and Engineering Laboratory, Hanover, New Hampshire.

REFERENCES

1. Abrate, S. 1991. "Impact on Laminated Composite Materials," ASME Applied Mechanics Review, Volume 44, No. 4, April, pp. 155-190.

2. Cantwell, W.J. and J. Morton. 1991. "The Impact Resistance of Composite Materials - A Review," Composites, Vol. 22, No. 5, September, pp. 347-362.

3. Goldsmith, W. 1960. "Impact: The Theory and Physical Behavior of Colliding Solids," Edward Renold, London.

4. Goldsmith, W. and P.T. Lyman. 1960. "The Penetration of Hard-Steel Spheres into Plane Metal Surfaces," ASME J. of Applied Mechanics, December, pp. 717-725.

5. Sondergaard, R., K. Chaney, and C.E. Brennan. 1990. "Measurements of Solid Spheres Bouncing Off Flat Plates," ASME J. of Applied Mechanics, Vol. 57, September, pp. 694-699.

6. Daniel, I.M. and S.C.Wooh. 1985. "Embedded Gages for Study of Transient Deformation and Dynamic Fracture in Composites," Proc. of Fall Meeting of SEM, Grenelefe, FL, November, pp. 62-68.

7. Daniel, I.M. and S.C.Wooh. 1990. "Deformation and Damage of Composite Laminates under Impact Loading," Impact Response and Elastodynamics of Composites, edited by A.K. Mal and Y.D.S. Rajapakse, AMD-Volume 116, Winter Annual Meeting, Nov. 25-30, pp. 11-26.

8. Kim, B.S. and, F.C. Moon. 1979. "Impact Induced Stress Waves in an Anisotropic Plate," AIAA Journal, Vol. 17, pp. 1126-1133.

9. Daniel, I.M., T. Liber and R.H. LeBedz. 1979. "Wave Propagation in Transversely Impacted Composite Laminates," Experimental Mechanics, Vol. 19, No. 1, pp. 9-16.

10. Cundari, M. and B. Abedian. 1991. "The Dynamic Behavior of a Polyvinylidene Fluoride Piezoelectric Motional Device," Smart Structures and Materials, G.K. Haritos and A.V. Srinivasan, ed. ASME Ad-volume 24, AMD-Volume 123, Winter Annual Meeting, Atlanta, Dec 1-6, 1991, pp. 25–31.

11. Altamirano, M. 1991. "Experimental Investigation of High and Low Impact Energy Absorption of AS4/3502 Graphite/Epoxy Panels." MS thesis, University of New Orleans, May.

APPENDIX A1. THE PVDF SENSORS

The PVDF sensor being used has the following properties:

Thickness $= 28 \times 10^{-6}$ m $= 1.102 \times 10^{-3}$ in.)
Lead wires = 41 swg
Readout = digital oscilloscope
Speed $= 50 \times 10^{-9}$ sec/div

Number of channels = 2 (for velocity study)
= 4 (for attenuation study)

Lamina thickness (seven plies) $= 1.04 \times 10^{-3}$ m = 0.041 in.
Sensor to lamina thickness = 1/37.2

A schematic diagram of the PVDF film is shown in Figure A1. The PVDF sensor measurement principle is shown in Figure A2. The applied stress σ can be found from

$$\sigma = Q/(Ad)$$

where Q is the charge, A is the sensor area and d is the charge sensitivity. Further,

$$d\sigma/dt = (1/(Ad))\, dQ/dt$$

and the current I can be found from

$$dQ/dt = I = V/R$$

where V is voltage and R is resistance. Finally,

$$d\sigma/dt = V/(RAd).$$

Integrating, one obtains,

$$\sigma(t) = (1/(RAd)) \int_0^t v\, dt.$$

For application purposes,

$R = 10^6$ ohms

$d = 33 \times 10^{-12}$ (C/m^2)/(N/m^2)

$A = 1.88 \times 10^{-4}$ m^2

Thus, one obtains

$$\sigma(t) = 6.119 \times 10^8 \, (1/(RAd)) \int_0^t v \, dt \ (Pa)$$

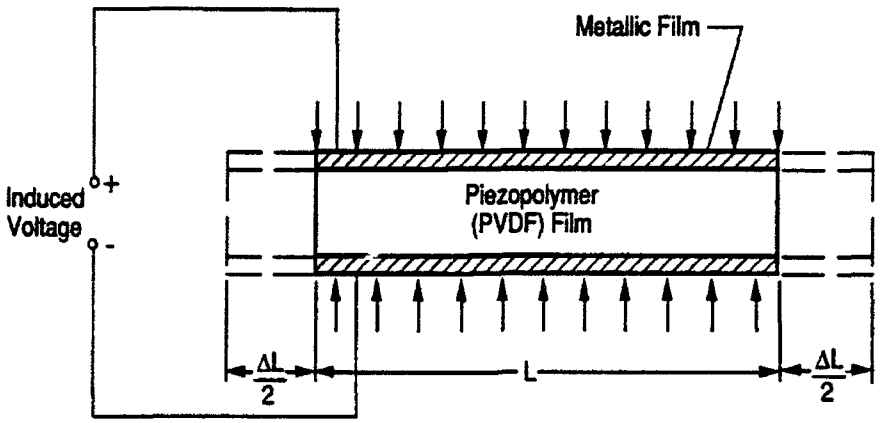


Figure A1. Schematic of the PVDF film inducing electric charge from an applied mechanical stress.

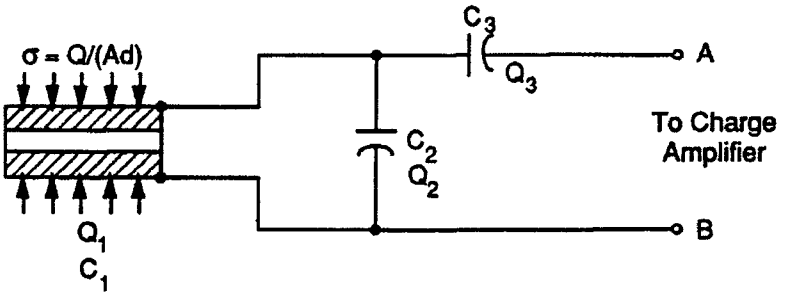


Figure A2. Principle of measurement of stress from PVDF sensors.

PRELIMINARY MISSILE AUTOPILOT DESIGN USING REACTION JETS AND AERODYNAMIC CONTROL

Mario Innocenti
Associate Professor
Department of Aerospace Engineering

Auburn University, 211 Aerospace Building
Auburn, AL 36849-5338

Final Report for:
AFOSR Summer Faculty Research Program
Wright Laboratory / Armament Directorate
Eglin Air Force Base, FL 32542

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

August 1992

PRELIMINARY MISSILE AUTOPILOT DESIGN USING REACTION JETS AND AERODYNAMIC CONTROL

Mario Innocenti
Associate Professor
Department of Aerospace Engineering
Auburn University, AL 36849-5338

Abstract

The feasibility of combining traditional aerodynamic control with reaction jets is studied, in the framework of missile autopilot design. The purposes of reaction jets are to increase the missile turning capabilities in critical phases of flight, as well as to help the traditional aerodynamic control in configurations where the fin size is reduced because of limited storage volume. Due to the nonlinear characteristics of both controller and airframe dynamics during fast maneuvers, a control strategy based on variable structure systems is used. A control law is synthesized for the pitch loop and some initial results are presented and compared with a traditional controller based on linear quadratic techniques.

PRELIMINARY MISSILE AUTOPILOT DESIGN USING REACTION JETS AND AERODYNAMIC CONTROL

Mario Innocenti

INTRODUCTION

Future missile systems will possess higher turn rates and larger maneuverability envelopes, while simultaneously requiring reduced storage and signature volumes. In this respect, efforts are under way to develop alternate means of missile control as opposed to purely aerodynamic control.

Several technology payoffs can be envisioned if alternate control strategies could be implemented, among which:

- decreased stowage volume for internal carriage, especially important for the type of fighters currently being developed,
- increased maneuverability and off-boresight capability for improved all-aspect defensive shield,
- high angle of attack launch capability to take advantage of improved aircraft agility,
- better end-game accuracy.

These goals pose difficult challenges to the control system in all phases of flight. During separation, an increase in pitch-up tendencies is expected due to lack of sufficient aerodynamic stabilization and high angles of attack. In the midcourse phase, the system is required to perform fast 180 degrees turns to account for defense and engagement against tail positioned threats. During the end-game, the reduced aerodynamic control effectiveness must be appropriately compensated in order to generate sufficient load factors.

The limitation in cross section and volume drastically reduces the amount of aerodynamic effectiveness of the missile, which must be compensated and/or improved by using alternate technologies. Primary options suggest propulsive control in the form of thrust vectoring (TVC) and reaction jets thrusters (RCS) leading to a generic configuration shown in figure 1.

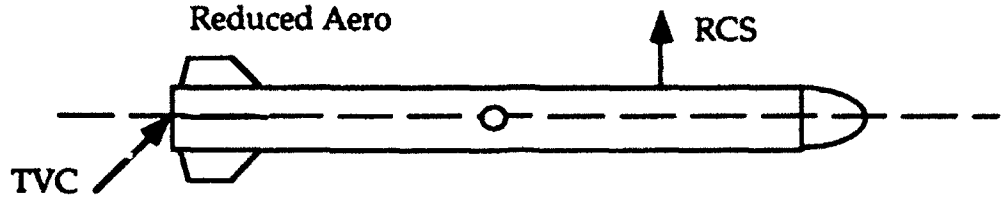


Figure 1. Generic Control Capabilities

In order to gain appreciation for some of the problems involving reaction jets control and its blending with traditional aerodynamic control, it was decided to concentrate on evaluating the guidance and autopilot issues during an " over the shoulder " flight task representative of the 180 degree off-boresight capability of the system. Typically, this task would allow for defense against tail and fly-by threats as shown schematically in figure 2.

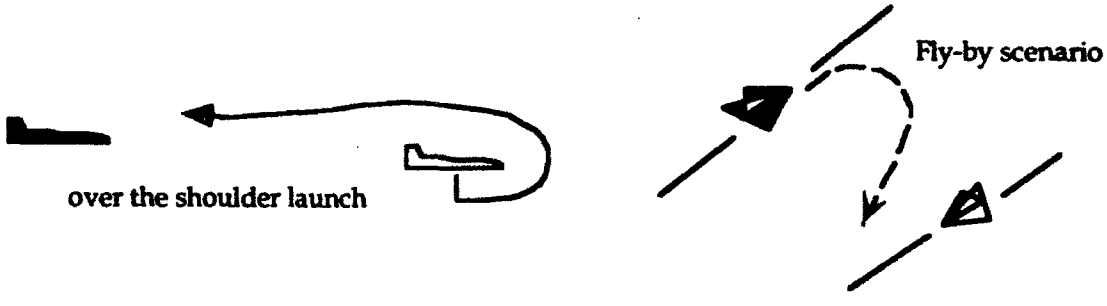


Figure 2. Midcourse Trajectory Capabilities

The challenges to the guidance and control systems are several and require a much greater effort than the one available during the summer research program, however some of the goals could be specified to lead to a preliminary analysis of the autopilot. A 180 degree off-boresight trajectory is sought with turning rates of the order of 80 deg/sec capable of pointing the missile in the opposite direction as

quickly as possible following a minimum radius turn path and in a time frame of the order of two seconds.

Clearly, the specifications are both in terms of guidance as well as autopilot requirements. The guidance aspects deal with the generation of an appropriate flight path along which the missile turns in minimum time changing its heading of up to 180 degrees. The autopilot aspects deal with the creation of forces and moments on the missile capable of generating accelerations and attitude rates required by the guidance system. Appropriate blending of aerodynamic and reaction jet controls is required since, during the trajectory, the missile will possibly experience complete loss of lifting capabilities due to angles of attack reaching values higher than 90 degrees.

The next section of the report will briefly develop the analytical aspects of a representative guidance problem without actually solving it. The main results of the work are described in the autopilot section together with some preliminary assessment of RCS capabilities.

GUIDANCE ISSUES

In this section, the point mass equations of motion of a generic missile will be derived, which could be used in the development of optimal trajectories for the present problem. The maneuver of interest has been described in the previous section and it consists of a turning trajectory from some initial condition to a final time characterized by the missile flying in a direction opposite to that at the initial time. The maneuver is to be performed in a minimum time, or in a time frame of two seconds. During the maneuver, the attitude change should be caused by a combination of RCS and aerodynamic controls, while the main engine will be used to recover the loss of dynamic pressure due to post stall angle of attack and increase of drag. Although a lot of work has been done with regard to this type of guidance problem, [1], [2], most of the results deal with coordinated flight assuming zero sideslip and side forces. This is not the case here, where skid capabilities are allowed and in fact may be necessary in order to reduce the time to achieve the final state.

A point mass model of the system is used and constant mass is assumed. The

applied forces acting on the missile are due to aerodynamics, gravity, main engine thrust (vectored) and reaction jets thrust. The coordinate systems used are standard for this problem [3] and consist of wind, body, and inertial axes identified by $[\bullet]_w$, $[\bullet]_b$ and $[\bullet]_I$ respectively. Defining the following angular quantities

$$\begin{cases} \mu = \text{velocity roll angle} \\ \gamma = \text{flight path angle} \\ \chi = \text{heading angle} \\ \varepsilon = \text{TVC wind angle of attack} \\ \nu = \text{TVC wind sideslip angle} \end{cases} \qquad \begin{cases} \phi = \text{body roll angle} \\ \theta = \text{body pitch angle} \\ \psi = \text{body yaw angle} \\ \alpha = \text{angle of attack} \\ \beta = \text{sideslip angle} \end{cases}$$

we have the following expressions for the applied forces:

$$\bar{F}_A = -D\,\bar{i}_w - Y\,\bar{j}_w - L\,\bar{k}_w \tag{1}$$

$$\bar{T}_M = T_M\,[\cos\varepsilon\cos\nu\,\bar{i}_w + \cos\varepsilon\sin\nu\,\bar{j}_w - \sin\varepsilon\,\bar{k}_w]$$

$$\bar{F}_g = g\,[-\sin\gamma\,\bar{i}_w + \sin\mu\cos\gamma\,\bar{j}_w + \cos\mu\cos\gamma\,\bar{k}_w]$$

where T_M, F_A and F_g are engine thrust, aerodynamic and gravitational forces respectively. Although the reaction jet force is negligible compared to the other three, it can be included in the equations of motion as shown below

$$\bar{T}_R = [-T_{RY}\sin\beta + T_{RZ}\sin\alpha\cos\beta\,]\bar{i}_w + [\,T_{RY}\cos\beta + T_{RZ}\sin\alpha\sin\beta\,]\bar{j}_w + T_{RZ}\cos\alpha\,\bar{k}_w \tag{2}$$

where T_{RY} and T_{RZ} are expressed in body axes. Using Newton's law of motion the standard set of six first order nonlinear differential equations can be obtained as:

$$\dot{V}_T = \frac{1}{m}[T_M\cos\varepsilon\cos\nu - D - mg\sin\gamma] + \frac{T_{RXw}}{m}$$

$$\dot{\gamma} = \frac{1}{mV_T}[T_M(\sin\varepsilon\cos\mu - \cos\varepsilon\sin\mu) + L\cos\mu + Y\sin\mu - mg\cos\gamma + T_{RZw}\cos\mu - T_{RXw}\sin\mu]$$

$$\dot{\chi} = \frac{1}{mV_T\cos\gamma}[T_M(\sin\varepsilon\sin\mu - \cos\varepsilon\cos\mu\sin\upsilon) + L\sin\mu - Y\cos\mu + T_{RZw}\sin\mu + T_{RXw}\cos\mu]$$

$$\dot{x} = V_T\cos\gamma\cos\chi \quad , \quad \dot{y} = V_T\cos\gamma\sin\chi \quad , \quad \dot{h} = V_T\sin\gamma$$

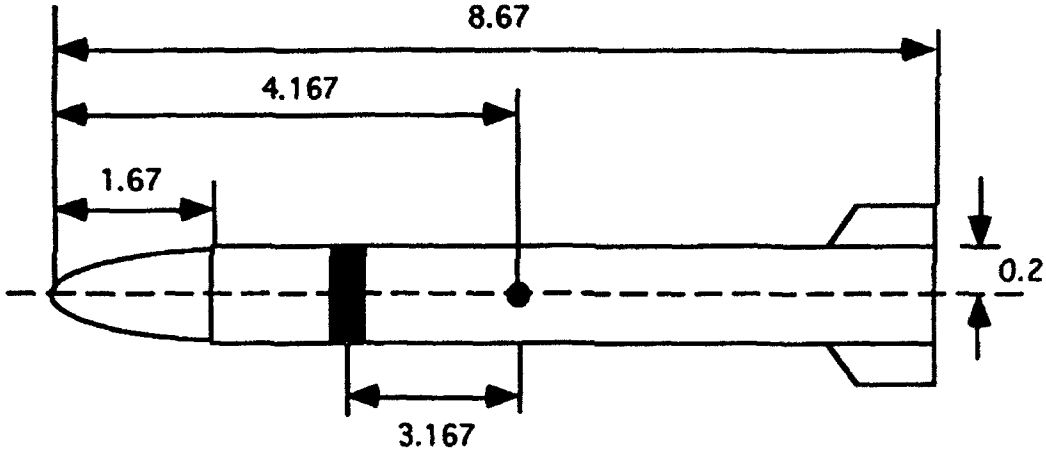
The specific energy rate of change could also be used if necessary. Traditionally, the control variables available for trajectory computation are the engine thrust level T_M and the velocity roll angle μ. In the present case, additional control variables include the thrust vectoring angles ε and ν plus the missile angle of attack and sideslip necessary to perform the required path. The latter two angles enter in the optimization process via aerodynamic forces L, D, and Y in the traditional form

$$L = \frac{1}{2}\rho S V_T^2 C_{L\alpha}\alpha \quad , \quad D = \frac{1}{2}\rho S V_T^2 [C_{D0} + k C_{L\alpha}^2 \alpha^2] \quad , \quad Y = \frac{1}{2}\rho S V_T^2 C_{y\beta}\beta$$

A typical approach to the trajectory problem could be stated as follows: find the control strategies for thrust, thrust direction, velocity roll angle, angle of attack and sideslip such that the system follows a minimum time trajectory from some initial condition to a final condition characterized by a final flight path angle of 180 degrees. Other trajectories could be generated as well by limiting α and α values and/or by requiring the final time to be specified (for instance t_f = 2 seconds).

AUTOPILOT SYNTHESIS

The autopilot is responsible for the generation of appropriate forces and moments in response to guidance commands, usually issued in terms of desired acceleration time histories. In this particular case, a lot of preliminary work had to be done because of the lack of frozen configuration, undetermined location of the reaction jets and absence of a meaningful aerodynamic database. To this end a working configuration had to be chosen in order to derive geometric and inertia properties as well as some preliminary aerodynamic characteristics using the missile DATCOM code. It must be noted that the aero properties could only be estimated up to medium angles of attack ($\alpha < 35$ deg). A sketch of the chosen configuration with the flight conditions considered is shown below.



Lref = .4167 ft	Flight Conditions:
Sref = .1367 ft	M = .3 , .6 , .8 , 2.0
m = 7.0 sl	h = 10,000 ft
Iy = Iz = 51 sl ft^2	trim alpha = 30 deg (5 deg at M = 2.0)
Ix = .229 sl ft^2	trim beta = 0 deg
X configuration	

Figure 3. Missile Configuration

The following aerodynamic characteristics were obtained using the DATCOM, where the symbols represent the standard aerodynamic derivatives. Due to the limited time available only the pitch channel has been evaluated, in addition, no

flexibility effects are include in the model. Future research could concentrate on this latter aspect, together with the coupling between pitch and yaw channel at high angles of attack, where vortex shedding will produce aerodynamic asymmetries.

Table 1. Aerodynamic Coefficients from DATCOM

Coefficient	M = 0.3	M = 0.6	M = 0.8	M = 2.0
Cnalpha	18.41	16.898	14.99	8.05
Cndelta	11.683	11.368	11.151	6.429
Cmalpha	-85.83	-83.887	-80.45	-24.038
Cmdelta	-116.92	-113.77	-111.52	-64.067

The equations of motion for the pitch channel (short period approximation) are given by:

$$\begin{cases} \dot{\alpha} = -\dfrac{QS_{REF}}{mV_T} C_{N\alpha}\,\alpha + q - \dfrac{QS_{REF}}{mV_T} C_{N\delta}\,\delta_E + \dfrac{T_{RCS}}{mV_T} u_T \\ \dot{q} = \dfrac{QS_{REF}\,L_{REF}}{I_y} C_{m\alpha}\,\alpha + \dfrac{QS_{REF}\,L_{REF}}{I_y} C_{m\delta}\,\delta_E + \dfrac{L_{RCS}\,T_{RCS}}{I_y} u_T \end{cases} \qquad (3)$$

T_{RCS} is the amount of thrust provided by the reaction jets (1000 lbs in this example), and L_{RCS} is the distance between reaction jets and center of mass of the missile. In addition to (3), two additional equations are used to model the actuators relative to the elevator δ_E and reaction jet thrust ratio u_T (u_T = 1 implies maximum thrust).

$$\begin{cases} \dot{\delta}_E = -a\delta_E + b\delta_C \\ \dot{u}_T = -\dfrac{1}{\tau} u_T + \dfrac{1}{\tau} u_C \end{cases} \qquad (4)$$

Nominal values for the actuator constants are: a = b = 1/180; τ = 1/500.

Introducing the normal load factor A_Z as $A_Z = V_T[\alpha - q]$

we can rewrite (3) and (4) in state space form

$$\dot{\bar{x}} = A\bar{x} + B\bar{u} \quad \text{with } \bar{x} = [A_Z \;\; q \;\; \delta_E \;\; u_T]^T \text{ and } \bar{u} = [\delta_C \; u_C]^T \tag{5}$$

Using radians to measure the angles and g's as unit for the load factor, the matrices A and B are given by:

$$A = \begin{bmatrix} Z_\alpha & \frac{V_T Z_\alpha}{g} & -\frac{V_T Z_\delta}{ga} & -\frac{V_T Z_T}{g\tau} \\ \frac{gM_\alpha}{V_T Z_\alpha} & 0 & M_\delta - \frac{M_\alpha Z_\delta}{Z_\alpha} & M_T - \frac{M_\alpha Z_T}{Z_\alpha} \\ 0 & 0 & -\frac{1}{a} & 0 \\ 0 & 0 & 0 & -\frac{1}{\tau} \end{bmatrix} \qquad B = \begin{bmatrix} \frac{V_T Z_\delta}{gb} & \frac{V_T Z_T}{g\tau} \\ 0 & 0 \\ \frac{1}{b} & 0 \\ 0 & \frac{1}{\tau} \end{bmatrix}$$

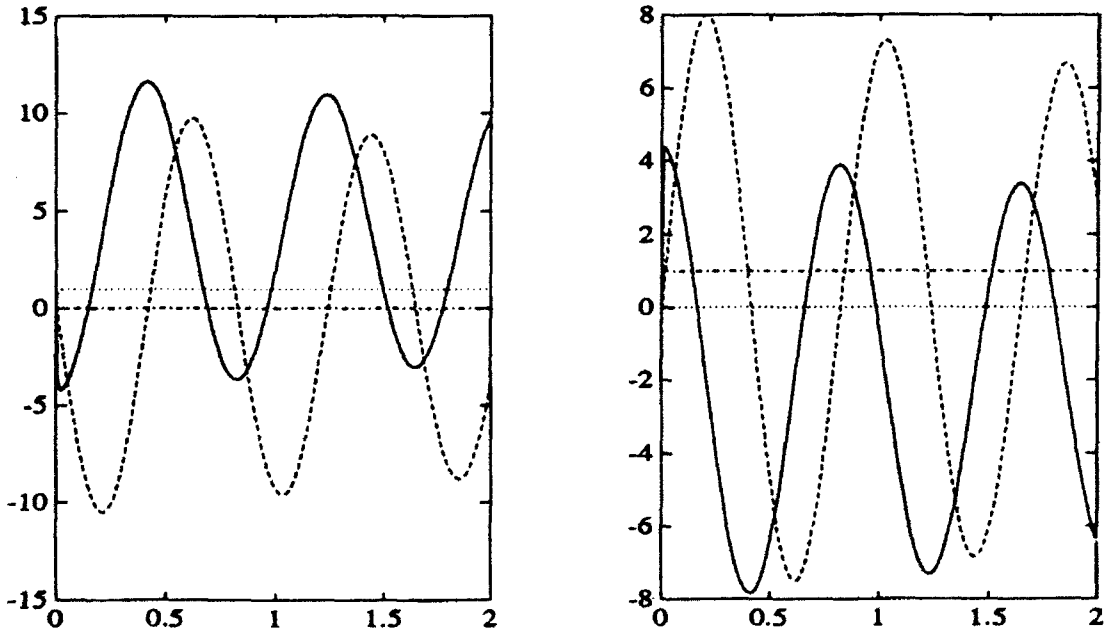
The autopilot synthesis will be performed using the variable structure control approach (VSC). Variable structure control has been studied since the seminal work by Emel'yanov in the early sixties [4] and a typical aerospace application can be found in [5].

The basic concept behind VSC consists in the design of a control law that forces the system's state to converge to and stay on a hypersurface in the state's dimensional space, such surface in called sliding hyperplane. Once on the surface, the system is said to be in a sliding mode, it remains there and reaches the origin irrespective of uncertainties and parameter variations satisfying Erzberger's model following conditions. The control law will assume a switching strategy changing sign instantaneously so as to keep the state in a sliding mode whenever it tends to depart the hypersufrace. Such switching leads to chattering, which is one of the properties of variable structure systems.

There are two main problems of VSC that need to be addressed in order to complete the design: the first step is the determination of the appropriate sliding

hyperplane, the second step is the determination of the control law. The selection of sliding surfaces can be made apriori thus leading to the so-called hierarchical control structure [6], or it can result from the choice of system behavior during sliding [5], which is the method used here. The determination of the control law structure, again is up to the designer. Typically VSC laws can be implemented using relay-type controllers with amplitude (gains) depending on the speed of response desired for the system and uncertainty levels. In this work a unit vector for the control law is used following the methodology described in [5]. The following paragraphs will describe an example application of VSC to the problem at hand, keeping in mind the limited effort of the work.

The autopilot synthesis deals with the determination of a control law which would allow the system to follow a g-command. The model is tested at a flight condition corresponding to M = 0.8 and altitude of 10,000 feet. The model of the system is given by (5), which represent the short period approximation of a rigid missile. The characteristics of the open loop system to a unit step control input are shown below.



where the left plots indicate the response to the elevator and the right plots the

response to reaction jets commands respectively. The main purpose of the above time histories is to indicate the lightly damped nature of the system. In fact, the fin size is much smaller than traditional all-aero controlled systems, thus providing almost negligible natural damping.

In order to achieve zero steady state error in commanded acceleration, an integrator is introduced in the loop. Define an additional state z_1 such as

$$\dot{z}_1 = A_Z$$

the augmented system becomes:

$$\dot{\chi} = A_a \chi + B_a \bar{u} \quad \text{with } \chi = \left[\int A_Z dt \quad A_Z \quad q \quad \delta_E \quad u_T \right]^T$$

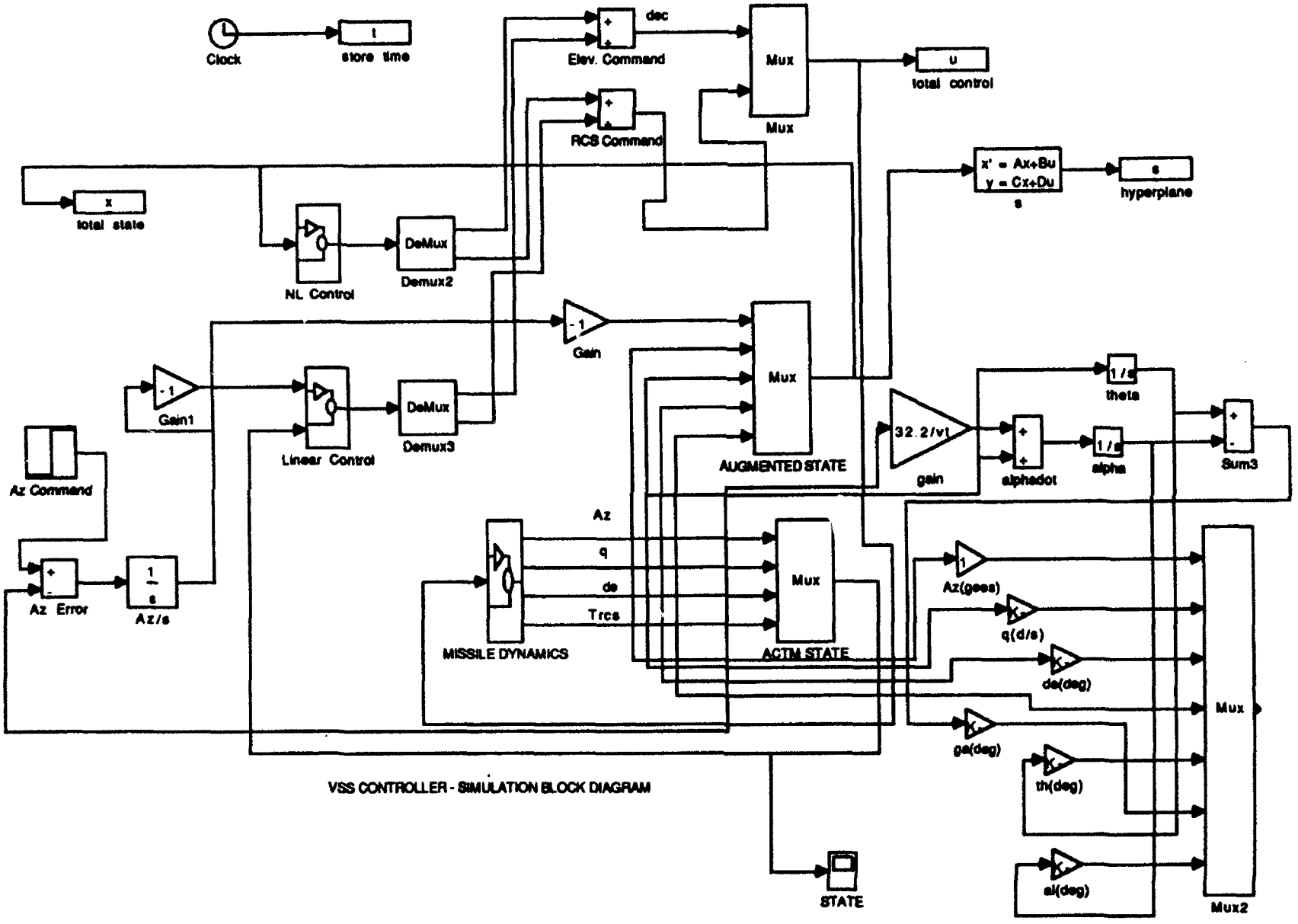
with A_a and B_a given by

$$A_a = \left[\begin{array}{c|c} 0 & 1000 \\ \hline \begin{matrix} 0 \\ 0 \\ 0 \end{matrix} & A \end{array}\right] \qquad B_a = \left[\begin{array}{c} 00 \\ \hline B \end{array}\right]$$

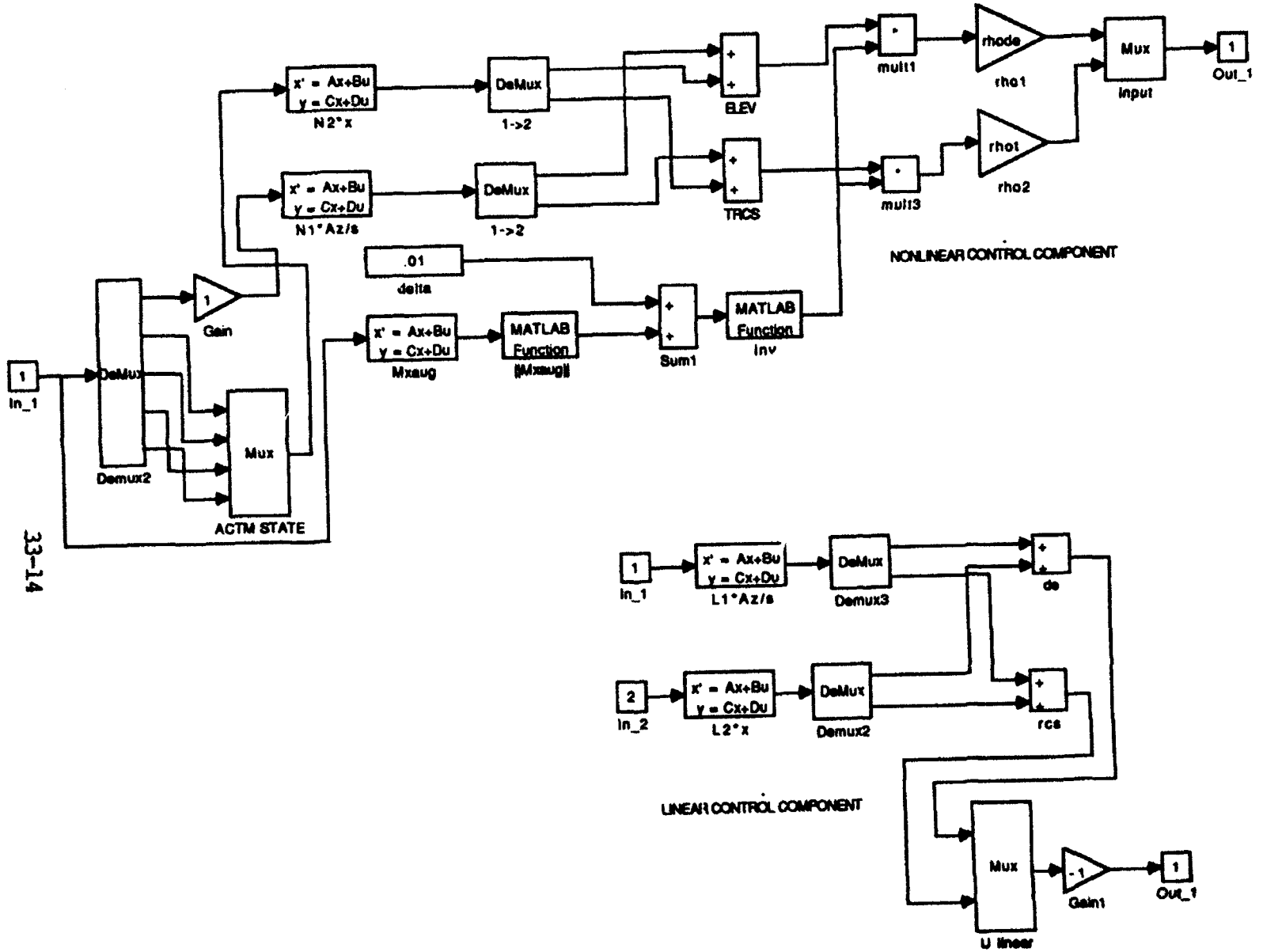
The VSC control law is derived according to the procedure used in [5] and is given by a linear component plus a nonlinear term responsible for getting the system to the sliding hyperplane in a finite time.

$$\bar{u}(t) = L\chi + \begin{bmatrix} \rho_\delta & 0 \\ 0 & \rho_{RCS} \end{bmatrix} \frac{N\chi}{|M\chi| + \delta} = \bar{u}^L + \bar{u}^{NL}$$

A schematic block diagram of the controlled system is shown in the next two pages, including the implementation of the control law.



VSS CONTROLLER - SIMULATION BLOCK DIAGRAM


x' = Ax+Bu
y = Cx+Du
N2*x
DeMux
1->2
ELEV
mult1
rhode
rho1
Mux
input
Out_1
x' = Ax+Bu
y = Cx+Du
N1*Az/s
DeMux
1->2
TRCS
mult3
rhot
rho2
NONLINEAR CONTROL COMPONENT
.01
delta
In_1
DeMux
Demux2
Gain
Mux
ACTM STATE
x' = Ax+Bu
y = Cx+Du
Mxaug
MATLAB Function
||Mxaug||
Sum1
MATLAB Function
Inv
In_1
x' = Ax+Bu
y = Cx+Du
L1*Az/s
DeMux
Demux3
de
In_2
x' = Ax+Bu
y = Cx+Du
L2*x
DeMux
Demux2
rcs
LINEAR CONTROL COMPONENT
Mux
U linear
-1
Gain1
Out_1

We remind the reader that, during this simulation, the reaction jet controller is modelled as a continuous gain function with no on-off characteristics. Further study is necessary to obtain a more precise and realistic modelling of the thrusters, such as using relays with deadband and histeresis.

The next four plots show the capability of the system to follow a 2g command in load factor.

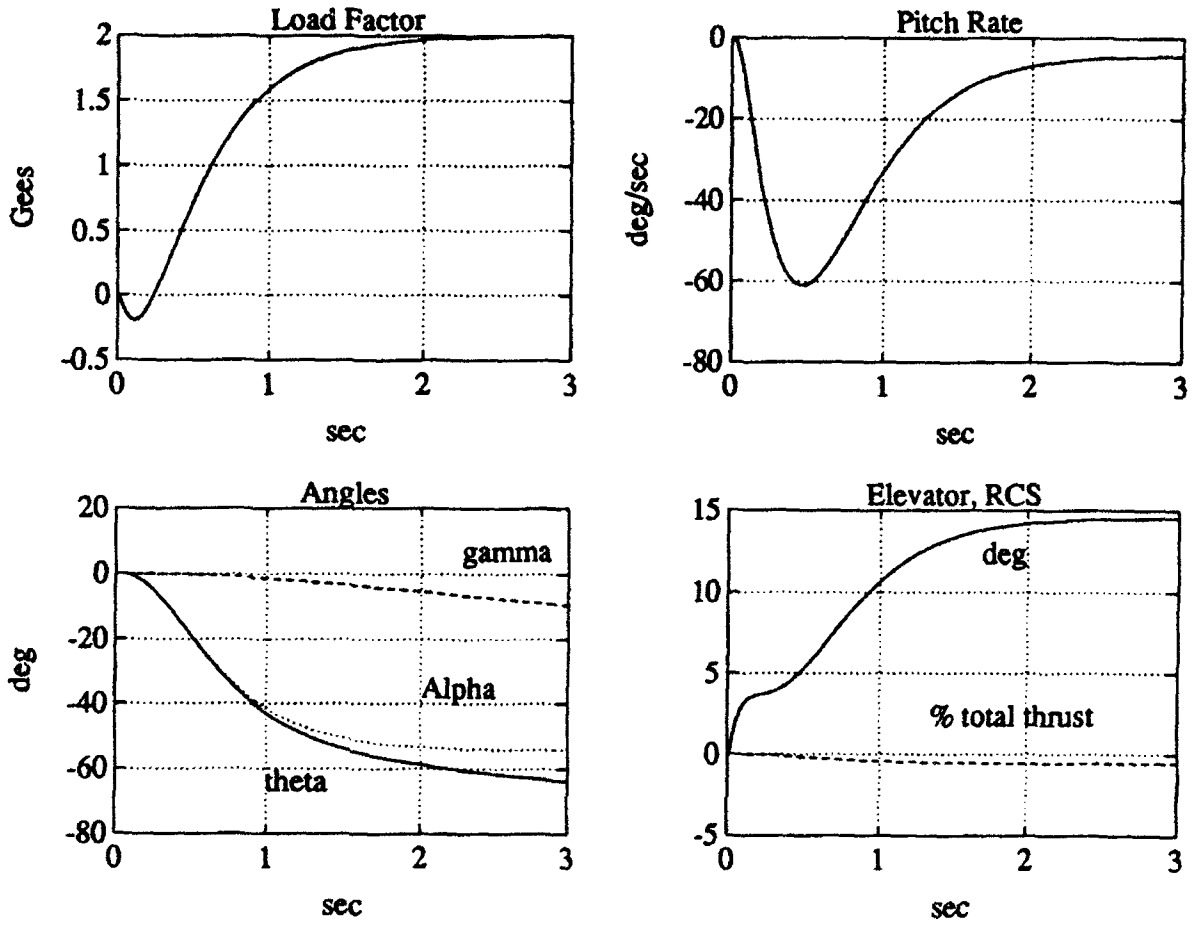


Figure 4. Response to Azc = 2g. Mach = 0.8 Configuration

The system is well behaved and the RCS conponent operates up to about 500 pounds.

Since the main objective of the controller is the capability of generating enough attitude so as to rotate the missile of about 180 degrees, an ideal block

command sequence is simulated and shown in the next four plots. The example is relative to a flight condition corresponding to Mach 0.3. The speed is low in order to reduce the drag originated by angle of attack values up to 90 degrees and higher. The example serves feasibility purposes only since the aerodynamic properties have not been changed from their trim values. Further study is suggested to investigate such dynamic behavior.

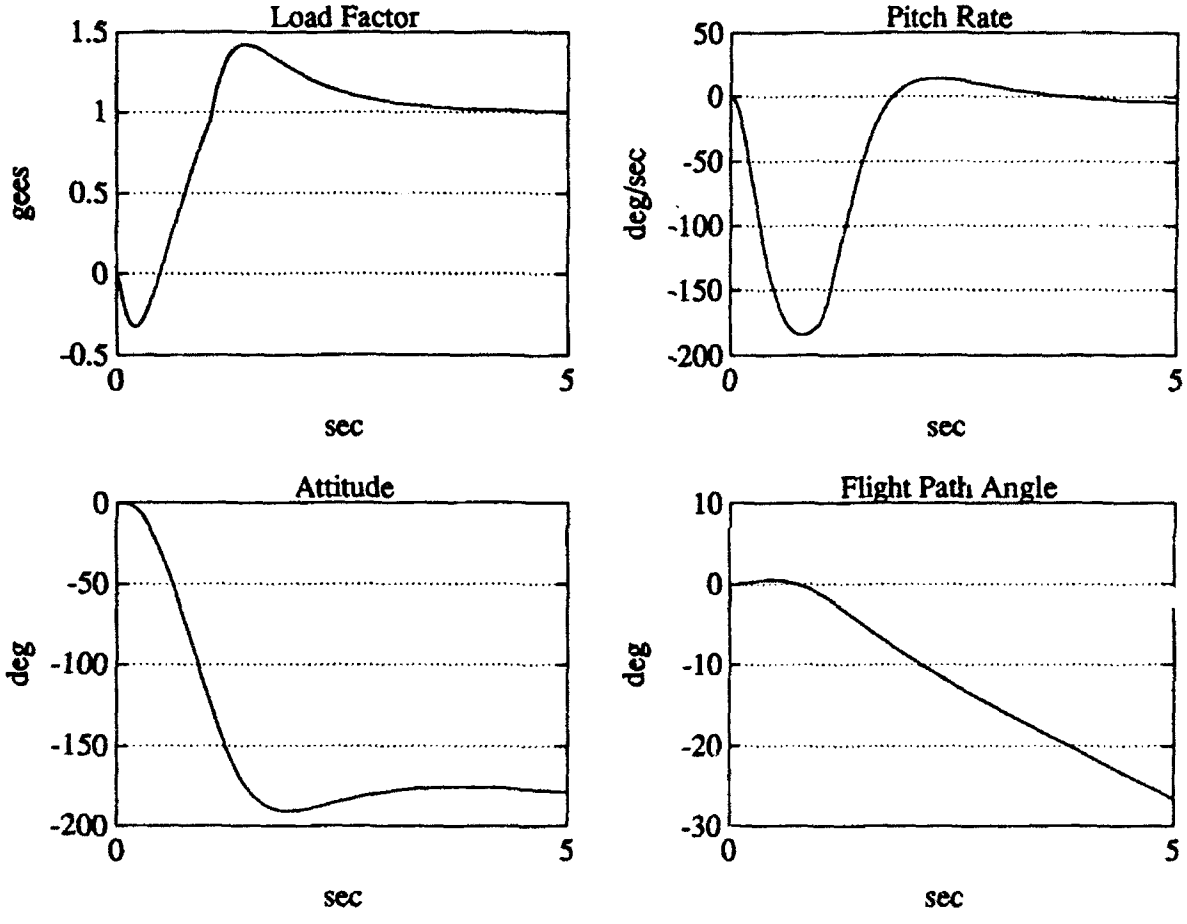


Figure 5. Response to a block g command. Mach = 0.3

the system is given a 2g command up to 1 second, after which the g demand is reduced to unity. The desired attitude is achieved very quickly, however the flight path angle follows with a delay that moves the angle of attack to value well beyond post stall. A more accurate simulation should account for the loss of lift, speed and the increase of drag experienced when the angle of attack passes the stall value. At

this point, the aerodynamic control effectiveness ceases and the only effector left is due to the reaction jet system. In this regime, flow separation and vortex asymmetries generate significant side forces so that pitch and yaw control can not be studied separately. The absence of data relative to these regimes does not allow, at this point, additional speculations although some analytical models are under investigation and will be further developed if additional funds become available.

CONCLUSIONS

A preliminary investigation of variable structure control has been presented, applied to a missile system with aerodynamic and propulsive control. The control system shows promise in terms of handling the ineherent nonlinearities due to RCS actuation. Further studies are suggested to investigate the autopilot capabilities in generating trajectories that allow the missile to change its orientation of values up to 180 degrees.

ACKNOWLEDGEMENTS

The author wishes to thank Mr. Fred A. Davis and all the personnel of WL/MNAV and WL/MNAG for their help and fruitful discussions.

REFERENCES

[1] Humphreys, R.P., Hennig, G.R., Bolding, W.A., Hegelson, L.A., "Optimal 3-Dimensional Minimun Time Turns for an Aircraft", The Journal of Astronautical Sciences, Vol. XX, No. 2, pp. 88-112, Sept.-Oct., 1972.

[2] Schultz, R.L., "Three-Dimensional Trajectory Optimization for Aircraft", Journal of Guidance, Control and Dynamics, Vol. 13, No. 6, Nov.-Dec., 1990.

[3] Miele, A., "Theory of Flight Paths", 1962.

[4] Emel'yanov, S.V., Design of VAriable Structure Control Systems with Discontinuous Switching Functions", Eng. Cybern., 1, 1964.

[5] Innocenti, M., "Controls Design Challenge: A Variable Structure Approach", AIAA-91-2676, AIAA Guidance, NAvigation and Control Conference, New Orleans, LA August 12-14, 1991.

LASER IMAGING AND RANGING (LIMAR) PROCESSING

Jack S.N. Jean
Assistant Professor
Department of Computer Science and Engineering
Wright State University
Dayton, Ohio 45435

Louis A. Tamburino
Avionics Directorate
Wright Laboratory
Wright-Patterson AFB, Ohio 45433

Ahmet A. Coker
Department of Computer Science and Engineering
Wright State University
Dayton, Ohio 45435

Final Report for:
Summer Research Program
Wright Laboratory

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

September 1992

LASER IMAGING AND RANGING (LIMAR) PROCESSING

Jack S.N. Jean
Assistant Professor
Department of Computer Science and Engineering
Wright State University

Louis A. Tamburino
Avionics Directorate
Wright Laboratory

Ahmet A. Coker
Department of Computer Science and Engineering
Wright State University

Abstract

The LIMAR (Laser IMaging and Ranging) project is a Wright Laboratory effort to develop an advanced imaging and ranging system for robotics and computer vision applications. LIMAR embodies a concept for the fastest possible three-dimensional camera. It eliminates the conventional scanning processes by producing a registered pair of range and intensity images with data collected from two video cameras. The initial prototype system was assembled and successfully tested at Wright Laboratory's Avionics Directorate in 1992. This prototype LIMAR system used several frame grabbers to capture the demodulated LIMAR image signals from which the range and intensity images were subsequently computed on a general purpose computer. The prototype software did not address the errors which are introduced by differential camera gain, misalignment, and distortion. The tasks performed during this Summer Research Program include (1) modeling and developing algorithms to correct the distortion introduced by using two cameras and (2) design of special purpose hardware to convert, in real-time, the outputs from the two cameras into a fully registered range and intensity image.

LASER IMAGING AND RANGING (LIMAR) PROCESSING

Jack S.N. Jean, Louis A. Tamburino, and Ahmed A. Coker

1 Introduction

The LIMAR (Laser IMaging and Ranging) project is a Wright Laboratory effort to develop an advanced imaging and ranging system for robotics and computer vision applications. LIMAR, the invention of Dr. Louis A. Tamburino of Wright Laboratory and Dr. John Taboada of the USAF School of Aerospace Medicine, embodies a concept for the fastest possible three-dimensional camera. It eliminates the conventional scanning processes by producing a registered pair of range and intensity images with data collected from two video cameras. The initial prototype system was assembled and successfully tested at Wright Laboratory's Avionics Directorate in 1992. The prototype LIMAR system used several frame grabbers to capture the demodulated LIMAR image signals from which the range and intensity images were subsequently computed on a general purpose computer. In this report, we describe several efforts for future enhancements to the LIMAR processing system.

The original intent of this research effort was to design a customized LIMAR processor to generate both the range and intensity images in real time. This new ability is needed to facilitate future robotic and automatic vision applications. The processor design effort was extended to incorporate error correction for camera distortions and misalignments. This in turned required investigations into calibration techniques and development of alignment algorithms that were not implemented or explored in the original LIMAR prototype evaluation. In the current investigation, a LIMAR processor was designed with an XILINX Development System to use programmable gate arrays, in order to reduce costs and facilitate rapid implementation.

LIMAR Device Overview As shown in Figure 1, the LIMAR device contains a laser, a polarization modulator, a beam splitter, and a processing unit. The laser shines light on the object which reflects the light back to the modulator. The modulator changes the polarization of the returned light. The polarization change depends on when the light reaches the modulator, or equivalently, depends on the distance between the object and the LIMAR device. Note that the light considered here is not a single spot but an area where the polarization of each single spot can

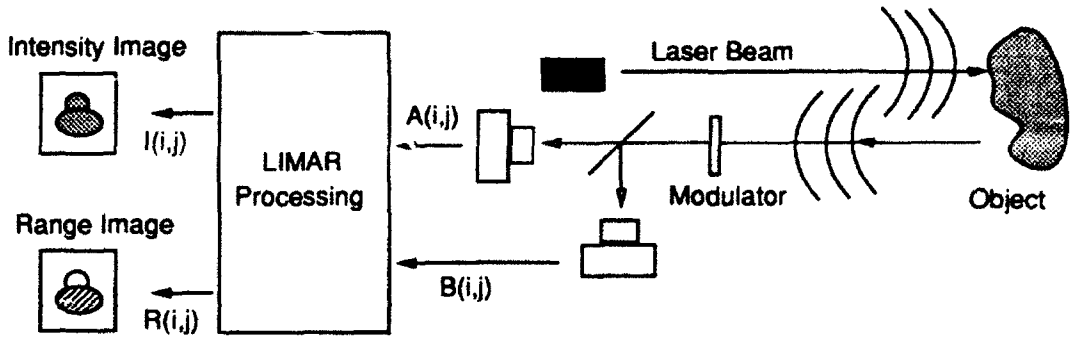


Figure 1: LIMAR system overview

be different from that of others. The beam splitter separates the light into two bundles of light, each captured by a camera, and the degree of separation is a function of the polarization of each spot. With the separation, some computation can be performed to extract the polarization for each single spot, and therefore, to calculate the distance of each spot on the object surface to the LIMAR device. In addition, the intensity, or the reflectivity, of each object spot can be obtained.

In an ideal case, the intensity image $\mathbf{I}(i,j)$ and the range image $\mathbf{R}(i,j)$ are as follows.

$$\mathbf{I}(i,j) = \mathbf{A}(i,j) + \mathbf{B}(i,j) \tag{1}$$

$$\mathbf{R}(i,j) = c\,\mathbf{f}(\,tan^{-1}\sqrt{\frac{\mathbf{A}(i,j)}{\mathbf{B}(i,j)}}\,) + d \tag{2}$$

where $\mathbf{A}(i,j)$ and $\mathbf{B}(i,j)$ are the two captured images from cameras, c and d are two constants, and the function $f(\cdot)$ is the inverse of the characteristic function associated with the modulator. In other words, the two output images can be computed pixel by pixel from the two input images in the ideal case.

Tasks Performed in the Summer Previously a prototype with two cameras was set up to take images. However, since the processing necessary to get the range image was performed on a general purpose machine, it was not possible to grab the range image in real time. Furthermore, the distortion introduced by using two cameras has not been investigated before. The tasks performed during this summer include (1) modeling and developing algorithms to correct the distortion introduced by using two cameras and (2) design special purpose hardware so that range images can be grabbed in real time.

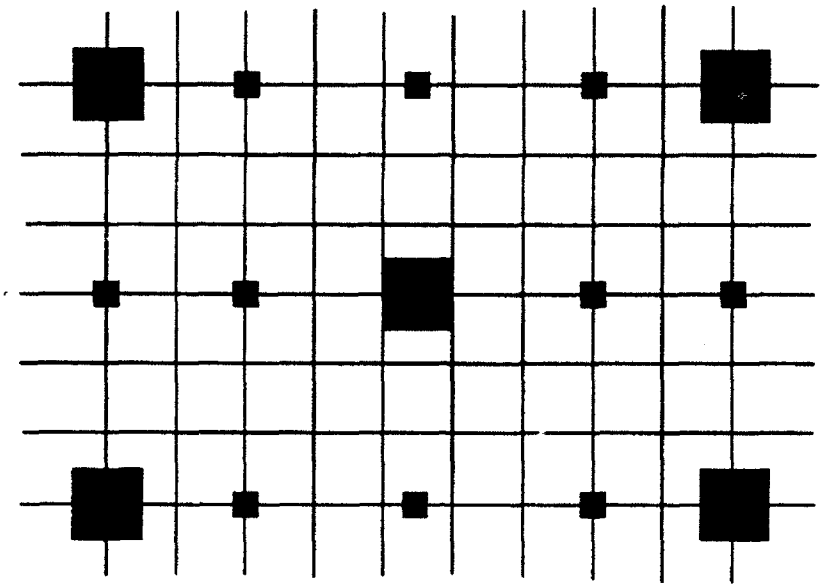


Figure 2: The computer generated test pattern used in the experiments.

In Section 2, we address the problem of image alignment utilizing two different approaches. In the first approach, we describe algorithms that automatically extract a set of corresponding points from two images and find a polynomial correction between them. In the second approach, we utilize a gradient descent technique to improve the polynomial correction. Section 3 describes the LIMAR processing system design.

2 Image Alignment

Because the two input images are grabbed with two cameras, they may not be aligned well with respect to each other. Furthermore, the two cameras may introduce different distortions. So there is a need to first align one of them to the other before the pixel by pixel operations as indicated in equations 1 and 2 can be performed. To explore the problem, a test pattern as shown in Figure 2 was generated and a laboratory environment was set up to take pictures of the test pattern. The two cameras were applied one by one and, each time the lighting, camera focusing, and camera positions were manually adjusted to maximize the image contrast and to align two images as close as possible. The resulting two 450×512 images are shown in Figure 3. The image quality is not good. Since high quality cameras are expected to be used in future LIMAR devices, the two images are considered as the worst case to be processed by a LIMAR device.

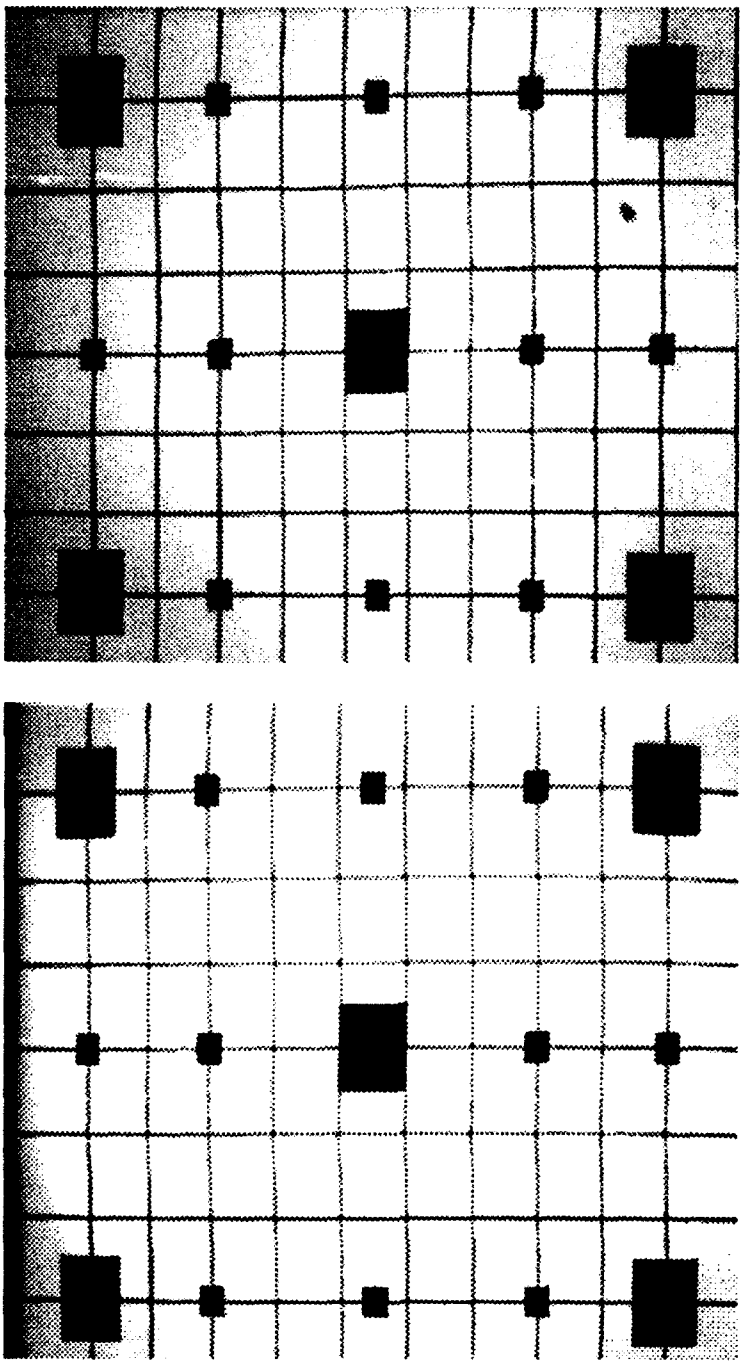


Figure 3: The input images used in the alignment process. They were grabbed with two different cameras. The top one is to be aligned with the bottom one.

The alignment problem was successfully modeled as a least square problem to estimate the coefficients of two third-degree polynomials. Algorithms were then developed to solve the problem.

2.1 Least Square Formulation

Let the image size be $M \times N$, where M is the image height in terms of pixels and N is the image width. Suppose x and y are the pixel indices used for an image *before* alignment and u and v are the pixel indices used for the image *after* alignment. The x and u indices, each ranging from 0 to $N-1$, are for horizontal axes and the y and v indices, each ranging from 0 to $M-1$, are for vertical axes. For the two images as shown in Figure 3, we may assume one is represented in (x, y) domain while the other is in (u, v) domain. Or we can label the two images as $\mathbf{A}(x, y)$ and $\mathbf{B}(u, v)$, respectively. The image $\mathbf{A}(x, y)$ is to be aligned, or to be transformed into another image $\mathbf{A}'(u, v)$, so that $\mathbf{A}'(u, v)$ can "match" well with $\mathbf{B}(u, v)$.

Given the image $\mathbf{A}(x, y)$, two separate steps are required to obtain $\mathbf{A}'(u, v)$.

1. *Position Mapping:* For each pixel location (u, v), where is the corresponding location (x, y) in the original image?

2. *Interpolation:* Since the original image $\mathbf{A}$ is defined only on the $M \times N$ rectangular grid, it is possible that no pixel value is defined at the location (x, y) which is obtained from the position mapping. For example, x or y may be negative or they may not be integer numbers. Some generalization of the image $\mathbf{A}$, or how interpolation of $\mathbf{A}$ on the (x, y) domain is to be performed, should be defined.

Position Mapping The mapping defines, for each pixel location (u, v) in the aligned image, the corresponding location (x, y) in the original image. In general, the mapping can be represented in terms of bi-variate polynomials. For example, a third-degree polynomial mapping is illustrated below.

$$x = a_0 + a_1 u + a_2 v + a_3 u^2 + a_4 uv + a_5 v^2 + a_6 u^3 + a_7 u^2 v + a_8 uv^2 + a_9 v^3 \quad (3)$$

$$y = b_0 + b_1 u + b_2 v + b_3 u^2 + b_4 uv + b_5 v^2 + b_6 u^3 + b_7 u^2 v + b_8 uv^2 + b_9 v^3 \quad (4)$$

The alignment problem is to find a suitable polynomial degree and a corresponding set of polynomial coefficients so as to "match" one image to the other. In reality, the coefficients can be estimated by identifying some correspondence in the two test images. For example, for each pixel

at a line intersection location (x, y) in the image **A**, there is a corresponding pixel location (u, v) in the image **B**. Suppose it is possible to identify P such related pixel pairs. Or equivalently, for some pixels (u_i, v_i) $(i = 1, 2, \ldots, P)$ on one image, their corresponding pixels (x_i, y_i) on the other image can be identified. Then the mappings for all the P pairs of pixels can be represented as the following set of linear equations with the polynomial coefficients as unknowns.

$$
\begin{bmatrix}
1 & u_1 & v_1 & u_1^2 & u_1v_1 & \cdots & v_1^3 \\
1 & u_2 & v_2 & u_2^2 & u_2v_2 & \cdots & v_2^3 \\
1 & u_3 & v_3 & u_3^2 & u_3v_3 & \cdots & v_3^3 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
1 & u_P & v_P & u_P^2 & u_Pv_P & \cdots & v_P^3
\end{bmatrix}
\begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \\ a_7 \\ a_8 \\ a_9 \end{bmatrix}
=
\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_P \end{bmatrix}
$$

$$
\begin{bmatrix}
1 & u_1 & v_1 & u_1^2 & u_1v_1 & \cdots & v_1^3 \\
1 & u_2 & v_2 & u_2^2 & u_2v_2 & \cdots & v_2^3 \\
1 & u_3 & v_3 & u_3^2 & u_3v_3 & \cdots & v_3^3 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
1 & u_P & v_P & u_P^2 & u_Pv_P & \cdots & v_P^3
\end{bmatrix}
\begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \\ b_7 \\ b_8 \\ b_9 \end{bmatrix}
=
\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_P \end{bmatrix}
\qquad (5)
$$

Since there are 20 coefficients for the third-degree polynomials and there are many more pairs of corresponding pixels, i.e., $P > 40$, the above linear equations are overdetermined and the coefficients can be solved with a least square technique.

Bilinear Interpolation Once the position mapping, $x(u, v)$ and $y(u, v)$, is estimated, the image after alignment, $\mathbf{A}'(u, v)$, should be determined from the given image $\mathbf{A}(x, y)$. With the position

mapping, the pixel $\mathbf{A}'(u,v)$ is to be mapped to a location $(x(u,v), y(u,v))$ on the $\mathbf{A}$ image. Note that $\mathbf{A}(x(u,v), y(u,v))$ may not be well defined. For example, it may fall outside the image array, it may be between two pixels, or it may fall in an area enclosed by four pixels. When it falls outside the image array, the corresponding $\mathbf{A}'$ value is set to be zero (white). When it is between two pixels, a linear interpolation of the two pixels is used as the value of $\mathbf{A}'$. When it falls in an area enclosed by four pixels, a bilinear interpolation is used for the $\mathbf{A}'$ value. More specifically, if $(x,y) = (x(u,v), y(u,v))$, $x \in [0, N-1)$, and $y \in [0, M-1)$, let x_I and y_I be the integer parts of x and y, respectively, and x_F and y_F be the fractional parts of x and y, respectively. Then

$$\mathbf{A}'(u,v) = a \times \mathbf{A}(x_I, y_I) + b \times \mathbf{A}(x_I + 1, y_I) + c \times \mathbf{A}(x_I, y_I + 1) + d \times \mathbf{A}(x_I + 1, y_I + 1)$$

where the four parameters a, b, c, and d are as follows.

$$\begin{aligned} a &= (1 - x_F)(1 - y_F) \\ b &= x_F(1 - y_F) \\ c &= (1 - x_F)y_F \\ d &= x_F y_F \end{aligned}$$

2.2 Extraction of Line Intersections

The test pattern was designed so that pixels at line intersections can be identified *automatically* and used to produce the P pairs of corresponding pixels. The following procedure was applied to each image to locate the 56 (the value of P in this case) line intersections.

1 Identify all the 15 black squares and estimate the center of each square by executing Steps 1.1 to 1.3.

1.1 Perform horizontal projections. Three strips of areas can be reliably identified from the resulting intensity distribution. Each corresponds to one row of black squares.

1.2 For each strip of area obtained from Step 1.1, perform a vertical projection *within the strip* and identify five areas, each corresponding to one black square.

1.3 For each black square obtained from Step 1.2, perform a horizontal projection to tighten the top and the bottom of the black area. The center is then estimated from the four boundaries of the black area.

2 From the centers obtained in Step 1, estimate line intersection locations. This can be done since the topology of the test pattern is known.

3 Refine the estimate by executing Steps 3.1 and 3.2.

3.1 (Pixel-level Refinement) On a 15×15 area with the estimated line intersection (obtained from Step 2) as the center, move the following 9×9 window around and convolve with the image. (The window serves as a detector of line intersection. For each pixel, it computes the sum of four 3-pixel areas, each along one side of the pixel.) Find the location where the convolved result is a maximum under the constraint that none of the four 3-pixel areas falls below some threshold.

```
0 0 1 1 1 0 0 0
0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0
1 0 0 0 0 0 0 0 1
1 0 0 0 0 0 0 0 1
1 0 0 0 0 0 0 0 1
0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0
0 0 0 1 1 1 0 0 0
```

3.2 (Subpixel Refinement) To get a subpixel estimate of the line intersection, a 7×7 window centered around the pixel obtained from Step 3.1 is formed. Along each side of the window, the position of a maximum is recorded. The average of the two horizontal maximum locations is used as the horizontal index of the line intersection. Similarly, the average of the two vertical maximum locations is used as the vertical index of the line intersection. Therefore the precision of the intersection is up to 0.5 pixel distance.

Similar technique was used by Dijak [1] for a camera calibration problem and Step 3 is from [1] with a slight modification. However, his process of extracting pairs of pixels required human interaction. The reason the current procedure requires no human interaction is because of the special test pattern used. The black square areas are large enough that they can be located very reliably to provide an estimate of the intersections.

With the 56 pairs of pixels, the least square formulation as in Eq. 5 was used together with an IMSL library function dlsqrr() to calculate the polynomial coefficients and the resulting error. This was done for polynomials of various degrees and the results are summarized in Table 1. It is clear that both the third-degree polynomial and the fourth-degree one were able to provide

Table 1: The absolute error for the x-index and the y-index after applying the estimated coefficients to the 56 pairs of line intersection pixels. The unit used for error is number of pixels.

		Max. Error	Ave. Error
2nd degree	x	3.288	1.351
	y	1.710	0.585
3rd degree	x	1.174	0.354
	y	0.788	0.266
4th degree	x	1.153	0.333
	y	0.680	0.243

subpixel accuracy for the alignment of the 56 pairs of pixels. The second-degree polynomial introduced an average error of 1.351 pixel for the x-index which is not suitable for the problem. Therefore, the third-degree polynomial was adopted in the project and the results are shown in Figure 4.

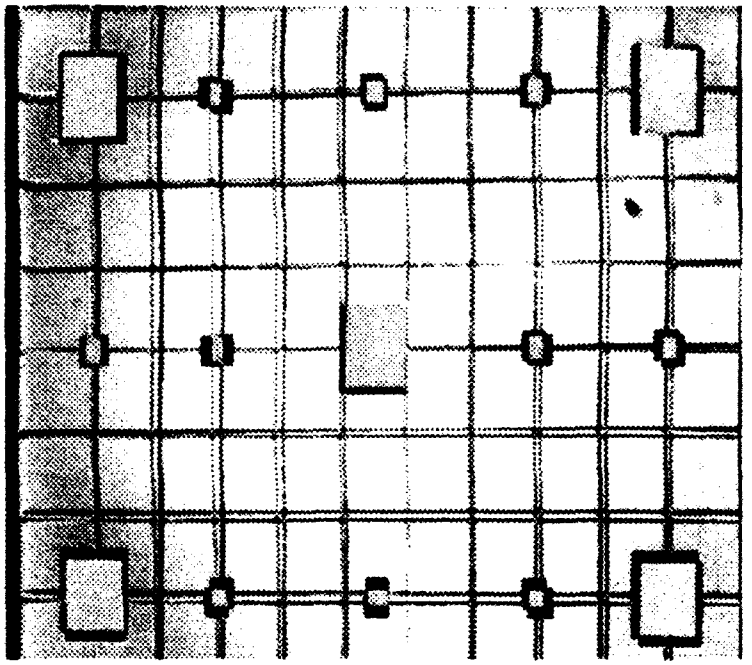
2.3 Gradient Descent Improvement

As demonstrated in the previous section, the coefficients obtained through least square estimation in average produce less than 0.4 pixel error in both x and y directions. This is true only for the 56 pairs of pixels. There is no guarantee for the remaining pixels. Therefore another algorithm, a gradient descent algorithm, was applied to slightly improve the coefficients. The algorithm tries to reduce the absolute error *over the whole image*.

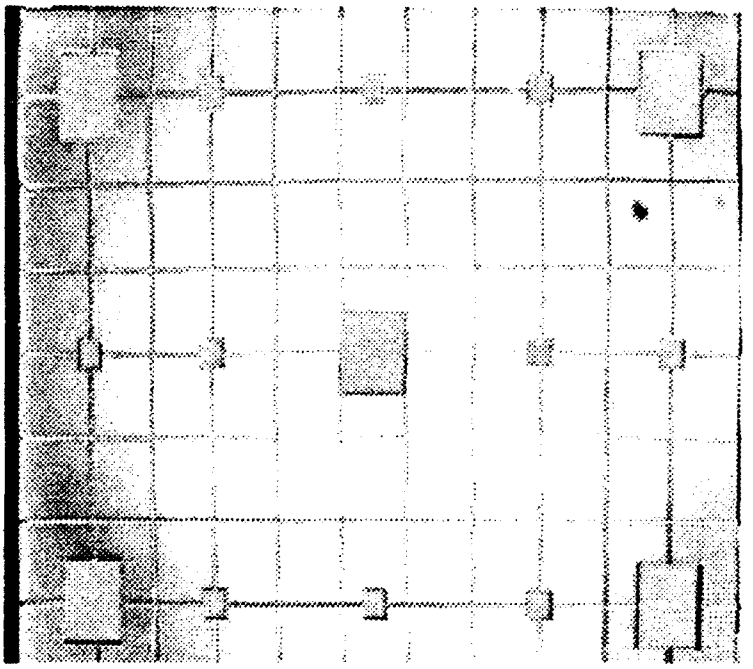
$$\mathrm{E} = \sum_{u=0}^{N-1} \sum_{v=0}^{M-1} |\mathbf{A}'(x(u,v), y(u,v)) - \mathbf{B}(u,v)| \tag{6}$$

where **A** and **B** are the two input images and $\mathbf{A}'(x(u,v), y(u,v))$ is the aligned image of **A**.

The algorithm starts from the coefficients obtained from least square estimation and gradually improves the coefficients in an iterative manner. At each iteration, it looks for the gradient descent direction $-(\frac{\partial \mathrm{E}}{\partial a_0}, \frac{\partial \mathrm{E}}{\partial a_1}, \ldots, \frac{\partial \mathrm{E}}{\partial a_9}, \frac{\partial \mathrm{E}}{\partial b_0}, \frac{\partial \mathrm{E}}{\partial b_1}, \ldots, \frac{\partial \mathrm{E}}{\partial b_9})$ by estimating all the 20 elements of the direction vector. For example, to estimate $\frac{\partial \mathrm{E}}{\partial a_i}$ from a current set of coefficients which has absolute error as E_{cur}, the coefficient a_i is increased by a very small amount δ, the resulting absolute error, E_i, is computed, and



(a)



(b)

Figure 4: The absolute difference between the two images (a) before alignment and (b) after one of the image is aligned with least square estimated polynomials.

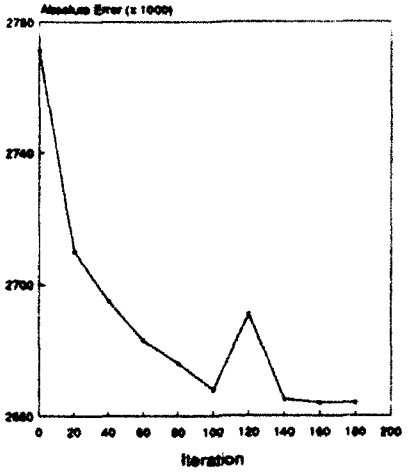


Figure 5: The absolute error versus iteration. The least square estimated coefficients are used initially.

$$-\frac{\partial \mathrm{E}}{\partial a_i} = \frac{\mathrm{E}_{cur} - \mathrm{E}_i}{\delta} \tag{7}$$

Therefore the following iterative coefficient updating is performed on a_i. (The coefficient b_i is similarly updated.)

$$\Delta a_i(t+1) = m\Delta a_i(t) - \eta\frac{\partial \mathrm{E}}{\partial a_i} = m\Delta a_i(t) + \frac{\eta}{\delta}(\mathrm{E}_{cur} - \mathrm{E}_i) \tag{8}$$

$$a_i(t+1) = a_i(t) + \Delta a_i(t+1) \tag{9}$$

where $m(= 0.6)$ is a momentum used to avoid local minima and increase convergence rate [2]. The parameter η is the learning rate which, depending on E_{cur}, is dynamically changed to improve solution quality. The result of applying the algorithm is depicted in Figure 5 where the absolute error is plotted as a function of iteration. It shows the decrease of the absolute error is quite significant within the first 200 iterations, which will not take more than 4 minutes with the proposed real-time hardware. Note that the improvement comes from the effect of minor adjustment of coefficients over the whole image array and its effects cannot be clearly identified by looking at the aligned images.

3 System Hardware Overview

3.1 Computational Requirement

Before the design of the hardware, a study of the computational requirements was performed. The study is based on the following assumptions:

- For each camera, there are 30 frames of images per second.
- Image size is 512×512.
- The polynomials $x(u, v)$ and $y(u, v)$ are re-evaluated for every frame of image. (The alternative is to precompute $x(u, v)$ and $y(u, v)$ for all u and v and store them in a lookup table. That was ruled out because of the need to perform iterative refinement of coefficients.)
- Two lookup tables are used for Eq. 2. (The equation contains time-consuming operations such as arctangent, square root, and division. Table lookup is faster since the two inputs $a = \mathbf{A}(i, j)$ and $b = \mathbf{B}(i, j)$ are both 8-bit integers. The reason for using two lookup tables instead of one is because it is desirable to be able to modify $f(\cdot)$, the inverse characteristic function of the modulator, in real time. The first table contains 2^{16} values for precomputed $(tan^{-1}\sqrt{\frac{a}{b}})$. A smaller table is then used to store $cf(\cdot) + d$.)
- A lookup table is used to compensate the disparity of the camera gains. The ratio of gains for each pixel location is precomputed and stored.

From the assumptions, it can be found out that

1. For the alignment of one image, the computations involved are the evaluation of the two polynomials and the bilinear interpolation. For each polynomial, nine MACs (multiplication and accumulation) are needed per pixel. The bilinear interpolation requires four MACs per pixel. This adds to 22 MACs per pixel, or 165 million MACs per second.
2. To compensate gain disparity, one of the images needs one memory access and one multiplication per pixel. That means 7.5 million memory accesses and 7.5 million multiplications per second.
3. To get the range image, two memory accesses per image pixel pair are required. To get the intensity image, one addition per image pixel pair is necessary. That means 7.5 memory accesses and 7.5 million additions per second.

Suppose each of the operation takes only one instruction cycle, the total computational requirement is 195 MIPS (million instructions per second). If the image size were 1K × 1K, then it becomes 780 MIPS. Considering a commercial DSP (Digital Signal Processing) chip normally can only provide 20 to 30 MIPS, we decided to implement an application specific hardware design for LIMAR processing.

3.2 Processing System Design

From the above analysis, it is clear that most computations are spent in the evaluation of polynomials. Therefore a special hardware was designed for that purpose with XILINX field programmable gate array (FPGA) chips. FPGA chips are user programmable and convenient for prototyping purposes. The polynomial generator is able to evaluate one polynomial once every clock cycle. With two polynomial generators, one for each polynomial, a hardware processing unit for the LIMAR device is illustrated in Figure 6.

Two input images **A** and **B** are scanned and input to the processing unit. The image **A** goes through an alignment process and the image **B** goes through a gain adjustment process. The functionality of each component is explained below.

- Four dual port memory chips, each of size $8K \times 8$, are used to store the image **A**. The image is partitioned and stored into the four chips so that the four pixels used for bilinear interpolation are always located in different chips. This is achieved by (1) assigning odd rows to two chips and even rows to the other two chips and (2) for the two chips sharing the same row, assigning odd columns to one chip and even columns to the other chip. The partitioning increases the memory bandwidth by four times for the bilinear interpolation. In real-time processing, while the input image is scanned and put into one port of one chip, the bilinear interpolation is performed by reading from the other ports of *all* the four chips. To avoid contention, several rows of image pixels should be stored into the chips before the interpolation operations start. This introduces some delays in the processing latency but has no influence on the processing rate.

- Four multipliers and three adders are used for bilinear interpolation.

- A memory is used to store the gain ratios between two images. Because the images go through the beam splitter and two cameras, they may have different gains at various places. The gain difference will be compensated pixel by pixel through the usage of a multiplier

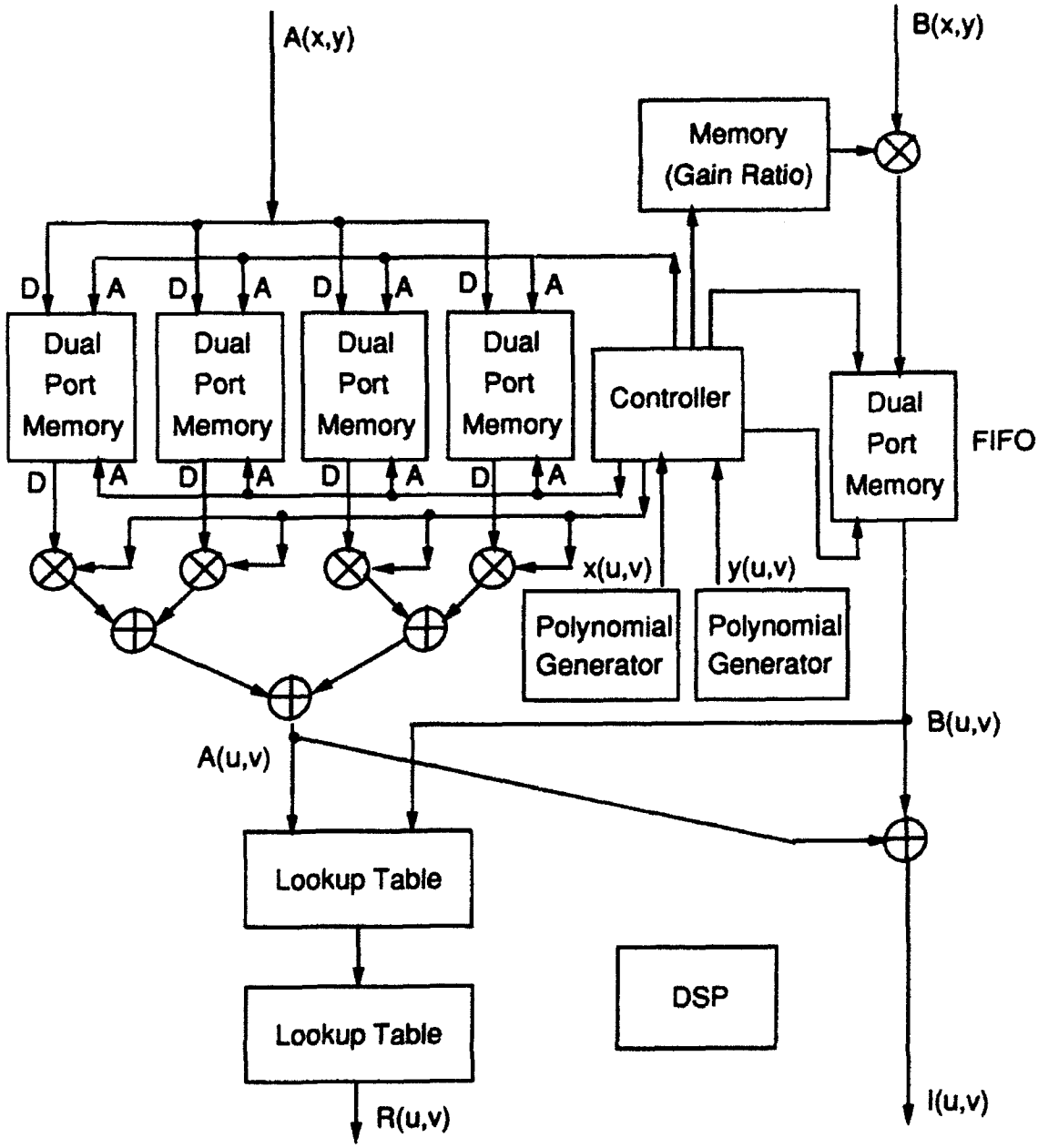


Figure 6: Block diagram for LIMAR processing

and the memory. This memory area should be initialized by the DSP chip. For each pixel location, the ratio $\frac{\mathbf{A}(x,y)}{\mathbf{B}(x,y)}$ should be precomputed based on more than one image frame to reduce noise effect, then stored in memory.

- A dual port memory is used as a FIFO (First Come First Out) to compensate for the delay introduced in the alignment process.
- Two lookup tables are used for obtaining the range image $\mathbf{R}(u,v)$ and one adder is used to generate the intensity image $\mathbf{I}(u,v)$.
- The DSP chip is responsible for (1) estimating the coefficients with the least square approach, (2) initializing the two polynomial chips for the iterative improvement algorithm, (3) initializing the memory for compensating intensity disparity, and (4) initializing the two polynomial chips for real-time processing.
- The controller, which can be implemented with programmable logic arrays, provides the address necessary for all the memory accesses.

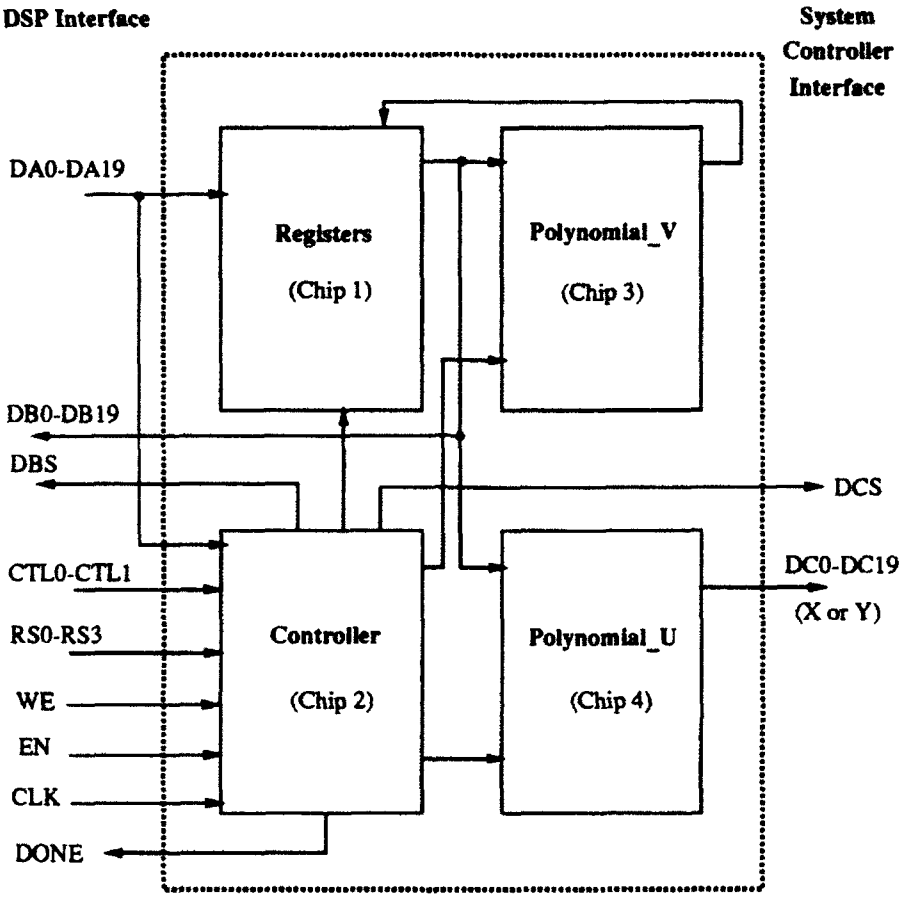
3.3 Polynomial Generator

The polynomial generator, or processor, consists of four XILINX 4003PC-10 programmable gate arrays (see Figure 7.) It is a special purpose processor that can evaluate a third-degree bivariate polynomial with given polynomial coefficients and generate an output every clock cycle. In addition to the polynomial coefficients, the processor must be provided with the upper bounds of the two variables that are inputs to the polynomial. (The two bounds are the image width and height in terms of pixels.) The polynomial processor uses these bounds to internally generate a sequence of pairs (u,v) starting from (0,0) as inputs to the polynomial. When an upper bound is reached, the next value of the variable starts from zero. Once both variables reach their upper bounds, both are set to zero and the cycle starts over again.

The processor has 16 20-bit registers and uses no multiplier at all due to a recursive formulation of the polynomials. Because of the formulation, if the polynomial is

$$p(u,v) = a0 + a1 * u + a2 * v + a3 * u^2 + a4 * u * v + a5 * v^2 + a6 * u^2 * v + a7 * u * v^2 + a8 * u^3 + a9 * v^3$$

then the registers should be initialized to the values as shown in Figure 7(b).



(a)

register 0: 2*a6	register 6: a8+a3+a1	register 11: a0
register 1: 6*a8+2*a3	register 7: 0	register 12: 6*a8
register 2: 0	register 8: 6*a9	register 13: image width
register 3: 0	register 9: 6*a9+2*a5	register 14: image height
register 4: 2*a7	register 10: a2+a5+a9	register 15: reserved for control
register 5: a7+a6+a4		

(b)

Figure 7: (a) Polynomial generator design and (b) Initial values for the 16 20-bit registers

The processor has three 20-bit data buses. The first data bus, DA0-DA19, is used for input to the registers. The second data bus, DB0-DB19, can be used for reading out the contents of the registers. The third data bus, DC0-DC19, is used for outputting the polynomial result. Two strobe pins, DBS and DCS, are used for indicating valid outputs on the DB bus and the DC bus, respectively. The two control pins, CTL1 and CTL0, are reserved for defining the operations to be performed. These operations can be

RESET	0 0 :	resets the processor to initial state,
START	0 1 :	starts execution of the processor to generate output,
READ	1 0 :	reads out the register contents,
WRITE	1 1 :	writes to the registers.

READ and WRITE operations are used in conjunction with the 4-bit address bus. The address bus is used to select which register to read from or write to. The write pin should become high when a valid input to the buses becomes available. The processor will start generating outputs once the operation START is made available to its data bus and the write pin is set high and then low. The execution of the START operation will not start if the write pin is left high. The RESET operation must be used after the power-on of the chip. This will allow the polynomial processor to be in the correct state prior to the execution of the other operations.

4 Conclusion

The LIMAR device, which is conceptually the fastest image ranging device, utilizes two cameras to grab images which are required to be in full registration. This presents several technical problems which we have addressed in this effort to design a real-time LIMAR processor that converts the image pair into range and intensity images. Because the original LIMAR prototype system was unavailable during this study, it was necessary to simulate the collection of image pairs using two different cameras. These experiments were necessary to determine the polynomial form for correcting differences between images which should be spatially registered except for distortions and variations introduced by the two cameras that captured them. The experimental results indicate that a third order polynomial is required to map one image into the other. The mapping algorithm also incorporates bilinear interpolation to improve pixel to pixel mapping with the polynomial. In addition, a recursive gradient descent technique was also derived to improve the polynomial corrections. Furthermore, the computational requirement was analyzed and a special polynomial

processor was designed using XILINX FPGA chips. With the chips, a system which can perform LIMAR processing in real time is proposed. The inclusion of camera registration algorithms into the resultant design addresses differential camera gain, misalignment, and distortion, and represents a significant enhancement to the future LIMAR development program at the Avionic Directorate of Wright Laboratory.

Acknowledgement

We wish to thank Edward Gliatti, Dale Nelson, Mason Friar, and Fred Meyer of the Avionics Directorate of Wright Laboratory for their cooperative efforts in providing the needed resources during this research effort.

References

[1] Jerome T. Dijak, "A Method for Correcting Geometric Distortion In Video Cameras," Internal Report, AF Wright Aeronautical Laboratories.

[2] Robert A. Jacobs, "Increased Rates of Convergence Through Learning Rate Adaptation," *Neural Networks*, Vol. 1, pp. 295-307, 1988.

DESCRIPTION AND RECOGNITION OF RADAR TARGETS USING WAVELETS

Ismail Jouńy

Department of Electrical Engineering
Lafayette College
Easton, PA, 18042

Final Report for:
Summer Research Program
Wright-Patterson Laboratories

Sponsored by:
Air Force Office of Scientific Research
Wright-Patterson Air Force Base, Dayton, OH

August 1992

DESCRIPTION AND RECOGNITION OF RADAR TARGETS USING WAVELETS

Ismail Jouny
Assistant Professor
Department of Electrical Engineering
Lafayette College

Abstract

A method for characterizing radar signatures using wavelets is developed based on the principle of scattering centers. The adopted target representation is based on the colored bright spots approach which is closely related to wavelet transform. Scattering features extracted using wavelets are directly related to target geometry and can be used for target identification. An optimal M-ary target classification technique utilizing wavelet decomposition is presented. The proposed target analysis and recognition techniques are tested using experimentally measured radar data.

Description And Recognition of Radar Targets Using Wavelets

Ismail Jouny

I. Introduction

The radar system adopted in this study operates by transmitting a single frequency signal for a short period of time. The returned backscattered signal is recorded as amplitude and phase, then a new signal with incremented frequency is transmitted and so on. A practical radar system utilizes about 64 frequencies and records the backscattered signal at each frequency. The result is often described as the target frequency response which represents the Radar Cross Section (RCS) of the target as a function of frequency. The inverse Fourier Transform of the recorded data represents "the target impulse response" and is displayed either as a function of time "t" or range "r", where r=ct/2, and "c" is the speed of light. The class of radar systems considered in this study also includes the chirp based radar systems. Furthermore, radars that utilize more than 64 frequencies can be characterized as Ultra Wide Band (UWB) radars and the resulting target response contains significant scattering information that is particularly suited for wavelet decomposition. Scattering from a radar target increases or decreases with frequency and can be modeled as

$$X(\omega) = \sum_{p=1}^{P} \sum_{m=-M_1}^{m=M_2} x_{pm} (j\omega)^{(-m)} e^{-j t_p \omega} \qquad (1)$$

and the target "impulse response" h(t), represents a generalized description of its scattering behavior and can be approximated as

$$x(t) = \sum_{p=1}^{P} \sum_{m=-M_1}^{m=M_2} x_{pm} \delta^{(m)} (t - t_p) \qquad (2)$$

where P indicates the number of scattering centers and M_1, M_2 denote the level of differentiation or integration (dispersion) respectively. The above expression approximates the target impulse response by a weighted sum of impulses, ramps,..., and doublets. For example, it is shown in [1] that the impulse response of a spherical target which includes direct specular scattering and a creeping wave can be approximated by

$$-(d/2c)\delta^{(0)}(t)+0.5\delta^{(-1)}(t)-0.44\delta^{(-3)}(t)+1.2\delta^{(-4)}(t)-1.9\delta^{(-5)}(t)+$$

$$6.1\delta^{(-2)}(t-2.6)-4.7\delta^{(-3)}(t-2.6)+15.7\delta^{(-4)}(t-2.6)-6.7\delta^{(-5)}(t-2.6)$$

where "d" is the target diameter. Furthermore, Rayleigh scattering is directly proportional to frequency and can be modeled by doublets. The significance of equation (1), however, extends beyond the principle of approximating a target impulse response by a weighted sum of integrated and differentiated impulse functions, and is directly related to the geometry of the target. The above target model is derived from the Geometrical Theory of Diffraction (GTD), and presents a better interpretation of the target scattering mechanisms. For example, based on GTD, scattering from a tip (corner) decreases linearly with frequency [2] and can be modeled as a step function, while scattering from a curved plate increases linearly with frequency [2] and can be modeled as a doublet. Other scattering mechanisms that increase or decrease with frequency can be attributed to flat plates, dihedrals, trihedrals, curved edges, etc. Therefore, by estimating the parameters x_{pm} for each scatterer, it is possible to identify the geometry of the target scattering component. As a special case, assume that x_{pm} is negligible except for m=1, then the p-th scatterer is likely to be a tip (corner).

II. Radar Signature Analysis

This study addresses the problem of estimating the parameters t_p and x_{pm} without prior knowledge of the number of scatterers P and, based on these estimates, attempts to identify the geometrical aspects of the target and uses the extracted information for target identification in a maximum likelihood sense. Recently, the problem of estimating the location of target scattering centers and their respective amplitudes, in the framework of spectral analysis, has received considerable attention [2,3,4,5,11,15,17]. Methods using Fourier Transform and high resolution spectral estimation algorithms have been used to estimate the parameters t_p. The Fourier transform based techniques are limited in resolution and are often incapable of resolving two closely related scatterers, but perform adequately when used for target recognition particularly under low signal-to-noise ratios. On the other hand, high resolution techniques

provide highly accurate estimates of the target scattering centers under high SNR but their performance deteriorates severely at low SNR. High resolution spectral estimation techniques such as Prony method, MUSIC, ARMA modeling, Maximum Entropy method have been used for target signature analysis [2,3,4,15,17], and scattering mechanisms identification. These methods produce accurate estimates of the location of scatterers based on a point scattering model and hence ignore the effects of differentiation and integration associated with each scattering component.

A method for estimating the delay parameters t_p and their amplitudes x_{pm}, is proposed in [2], where the location of scatterers t_p is estimated using the Prony method, the parameters x_{pm} are then estimated by taking the logarithm of both sides of (1) and using a bank of FIR filters to isolate all scattering components after being demodulated to the zero-time reference. This algorithm produces highly accurate estimates of the scattering centers, but less reliable estimates of the differentiation and integration parameters x_{pm}.

III. Target Signature Analysis With Wavelets

Description of sonar and radar returns as a weighted sum of impulses, ramp functions, step functions, etc has been addressed in [1], where a bank of matched filters is used to isolate and detect the target scattering components. Since differentiation results in an increase in the frequency response of the filters needed thus increasing its complexity, a specific transmitted signal was developed so that the time-bandwidth product of the matched filter remains constant. Altes [1], derived a class of signals {u(t)} that are compressed in time when differentiated and whose Fourier Transform is given by

$$U_{n,g,c}(\omega) = A\omega^n \exp\left(-\frac{\ln^2(\omega)}{2\ln(g)}\right) \exp\left(j\frac{2\pi c \ln(\omega)}{\ln g}\right) \qquad (4)$$

for $\omega \geq 0$, where g,n, and c are constants. The signal u(t) encompasses hyperbolic frequency modulation with a real part that resembles the chirp signal and has been used in early pulse compression methods (nonlinear FM techniques [16]). The matched filters used in [1] are then defined as

where E_m is the energy of the mth filter response. An important and desirable property of the

$$V_m(\omega) = (j\omega)^m U(\omega) / \sqrt{E_m} \tag{5}$$

derived signal u(t) is optimal doppler tolerance [1]. In addition to the estimation of the target scattering parameters x_{pm} even without prior knowledge, the system proposed in [1] is optimal for target detection and classification in a maximum likelihood sense. The estimates of the parameters x_{pm} are given by

$$[X] = A^{-1} r \tag{6}$$

where the vector r is the real portion of the output of the bank of matched filters, and the elements a_{kn} of the matrix A are defined as [1]

$$a_{km} = Re\left[\frac{1}{2\pi}\int_{-\infty}^{\infty} v_k^*(\omega)\, v_m(\omega)\, d\omega\right] \tag{7}$$

The estimates of the scattering elements x_{pm} are optimal in a minimum mean square sense. Flandrin et al [8], noted that the matched filter based system proposed by Altes [1] is equivalent to evaluating the wavelet transform of the target impulse response. The wavelet needed however, is the autocorrelation of the Altes signal u(t). Moreover, it is shown that such a wavelet is admissible [8] and its shape partially resembles the Grossman Morlet wavelet [8] for specific choices of the constants n,g, and c. The multiscale representations of the impulse response by wavelet decomposition correspond to different levels of integration and differentiation and hence different geometrical substructures of the target. It is shown in [8] that by using the autocorrelation of u(t) as a wavelet, then the scale parameter (a) is defined as

$$a = g^{-m/2}. \tag{8}$$

(note that g=4 corresponds to a dyadic grid). Hence, small scales emphasize higher order differentiations (m > 0), and large scales correspond to higher order integrations (m < 0). Notice that the scale parameter ($a = 2^{-m}$) where (m) is the level of integration and differentiation.

The wavelet derived from the Altes signal is bandpass and, by properly choosing the parameters n and g (the parameter c does not appear in the wavelet expression), can be made almost orthogonal to its own dilations and translations. For orthogonality to be established, it

can be shown that the parameter $m_{j,k}$ defined as

$$m_{j,k}=\frac{1}{2^j}\int_{\omega=0}^{\omega=\infty} |U(\omega)\,U(\frac{\omega}{2^j})|^2 e^{-\frac{j\omega k}{2^j}} d\omega \qquad (9)$$

must approach zero for all nonzero values of the integers j and k. Although some choices of the parameters n and g may minimize the parameter $m_{j,k}$, in general, wavelets constructed from the autocorrelation of u(t) do not constitute an orthonormal basis and may not produce exact estimates of the scattering coefficients.

Let w(t) be the analyzing wavelet of the target impulse response given in (2), then the wavelet transform of x(t) is

$$W_x(t,a)=\frac{A}{\sqrt{a}}\sum_{p=1}^{P}\sum_{m=-M_1}^{M_2}(-1)^m x_{pm} w^{(m)}(\frac{t-t_p}{a}) \qquad (10)$$

where A is a constant. Furthermore, if w(t) is chosen as the autocorrelation of u(t) then

$$W_x(t,a)=\frac{A}{\sqrt{a}}\sum_{p=1}^{P}\sum_{m=-M_1}^{m=M_2}(-1)^m x_{pm} w(\frac{g^{m/2}t-t_p}{a}) \qquad (11)$$

Also, because w(t) represents the impulse response of the matched filter used in [1], the target impulse response is approximated by [1,2]

$$x(t)=A\sum_{p=1}^{P}\sum_{m=-M_1}^{m=M_2}(-1)^m x_{pm} w(g^{m/2}(t-pn)) \qquad (12)$$

where $(n=t_p g^{-m/2})$. Equation (12) suggests that for an accurate approximation of target response, the autocorrelation of the Altes signal has to be very narrow for m>0, and very wide for m<0. Also, Equation (12) provides means for representing the target impulse response by a set of $P(M_1+M_2+1)$ coefficients.

The wavelet expansion of the impulse response of a radar target is, by definition, given as

$$x(t)=a^{-j/2}\sum_j\sum_k c_{j,k} w(a^{-j}(t-kT)) \qquad (13)$$

where T is some sampling interval. Let w(t) be a narrow wavelet that, upon compression, approaches derivatives of the delta function, then by comparing Equations 12 and 13, the target

scattering parameters x_{pm} can be estimated using wavelet expansion techniques. The scale parameter (j) used in wavelets is associated with the differentiation/integration parameter (m). The time shift parameter KT is associated with the location of target scattering centers t_p and need to be carefully selected. Let $k_1, k_2, \ldots, k_P$ be defined such that $t_1=k_1T$, $t_2=k_2T$, , . . . , $t_P=k_{PT}$, then the impulse response x(t) is approximated as

$$x(t) \approx C\sum_{k=k_1}^{k=k_P}\sum_{j=-M_1}^{j=M_2} x_{kj}2^{-j/2}w\left(\frac{t-kT}{2^{-j}}\right) \qquad (14)$$

where (C) is a constant. A prior knowledge of the location of scattering centers (can be obtained using Fourier Transform or any of the methods mentioned in Section II) expedites the estimation of the parameters x_{pm}. If the parameters t_p are not know beforehand then the wavelet expansion coefficients are evaluated over a dense time-scale grid. Estimates of scattering parameters $\hat{x}_{pm}$ that are below a certain threshold can then be ignored.

Assuming that a wavelet $w_{m,n}(t)=w(2^{-m}t-np)$ forms an orthonormal basis and the corresponding coefficients form a tight frame [9], where

$$x_{pm}=2^{-m/2}\int x(t)\, w^*(2^{-m}(t-pn))\, dt \qquad (15)$$

then the decomposition of target responses using wavelets has several merits that enhance the process of target recognition. These rewards are: 1) reduction in the number of data samples needed to represent a target, 2) accurate reconstruction using the parameters x_{pm}, and utilizing a set of compactly supported functions generated by dilation and compression of a single wavelet, 3) optimal detection and classification in noise enhanced by the orthonormal basis of wavelets, 4) identification of the geometry of the target scattering parameters by examining the decomposition coefficients x_{pm}, and then associating large scales with integration and small scales with differentiation, 5) unlike high resolution spectral estimation methods, wavelet decomposition does not require a priori information of the number of target scattering centers P.

The backscattered radar signal is often modulated by an unknown frequency due to mismatch between the target zero-time reference and the radar time origin. Such modulation results in translation (shift) in the impulse response of the target. This phenomenon diminishes the

contribution of phase information to target recognition. It is often the case that phase independent recognition methods (shift invariant), such as cross-correlation, are used to identify unknown radar targets, thus producing suboptimal classification results. Notably, dyadic wavelet transform is shift invariant [6] when evaluated over a continuous time interval, and optimal target recognition performance is reachable. Subsampling [7] and the discrete wavelet transform as defined in Equation (12), however, produces target scattering parameters that are not shift invariant, and hence cannot be used in target signature decomposition except when absolute phase information is available.

To examine the significance of wavelet decomposition in radar signature analysis, consider two hypothetical targets each consisting of three synthetic scatterers target with impulse responses

$$x_1(t) = A_1\delta(t-4) + B_1\delta^{(1)}(t) + C_1\delta^{(-1)}(t+3) \qquad (16)$$

$$x_2(t) = A_2\delta(t-4) + B_2\delta^{(2)}(t) + C_2\delta^{(-2)}(t+3) \qquad (17)$$

where the constants A,B,C are chosen such that all responses are comparable in magnitude. Note that both targets have similar scattering centers and differ only in the level of integration and differentiation associated with each scatterer. Figure 2 shows the wavelet transform of $x_2(t)$ impulse responses at different scales. Notice that while the response of the specular scatterer remains feasible at all scales, the response of the integration based scatterer diminishes at small scales and that of the differentiation based scatterer increases in magnitude. At large scales, the response of the differentiation scatterer decreases and that of integration increases. The significance of the wavelet transform of target impulse responses when examined at different resolution scales is more apparent when higher order differentiations and integrations are present. Clearly, the wavelet transform of target responses can be used to discriminate between scatterers whose responses increase with frequency (differentiation) and those whose responses decrease with frequency (integration). The wavelet used in Figure 1 is the Grossmen-Morlet wavelet as given in [6,8]. The interpretation of the result in Figure 1 is related to the fact that wavelet compression enhances the response of differentiating scatterers and wavelet dilation accents the

response of integrating scatterers, while the response of specular scatterers is not affected by either dilation or compression. However, it is not clear from Figure 2, as equation 12 suggests, that a direct correspondence exists between the level of integration/differentiation and the dilation/compression scale (a) used.

Patterns of responses similar to those shown in Figure 1 were observed when the wavelet derived from the altes signals was used as well as when using the Daubechies 4,6 and 8 wavelets [10]. The synthetic target response shown in Figure 1 is based on hypothetical scattering data in the frequency range 1-12 GHz with increments of 50 MHz. The simulated targets consist of three scatterers located at (-3), 0, and 4 (nanoseconds). The signal generated was corrupted with additive Gaussian noise with 18 dB SNR.

Figure 2 shows the impulse response of a real radar target model, namely, the Concord aircraft. The data was experimentally measured in a compact range [18] in the frequency range 1.5-12 GHz. Notice that by examining the target impulse response at different scales, it is possible to identify scatterers that have differentiating features and those that are characterized by integration. Upon exploring the nature of each scatterer and comparing its response and location with the geometry of the target model on the basis of GTD, it is clear that the behavior of each scatterer when examined at different scales is in accordance with its geometry.

Figure 3 shows the original and the reconstructed impulse responses of a sphere. The reconstruction was accomplished using the scattering parameters given in Equation (3) combined with dilated and compressed replicas of the wavelet derived using the autocorrelation of the Altes signal. Figure 3 suggests that the Altes wavelet, though it does not constitute an orthonormal expansion basis, can be used for characterizing radar signatures. On the other hand, almost exact replicas of the target shown in Figure 3 could be reconstructed using the Daubechies wavelets [9], or the spline wavelet [7]. The results in Figures 2 and 3 clearly indicate that signatures of radar targets can be completely characterized using wavelets that form orthonormal basis and are dilated/compressed upon integration/differentiation.

IV. Target Recognition Using Wavelets

Radar signature analysis encompasses the estimation of the number of target scattering centers P, their location along the target $ct_p/2$, and the associated dispersion level (frequency

dependency) or the level of integration and differentiation m, M_1, M_2 which determines the geometrical structure of the illuminated target. Estimates of such parameters can be used to identify the target under examination with minimal number of data. These parameters, or features, contribute in different proportions to the identification of unknown targets, particularly those targets that do not match the prototype at all times (before and after launching a missile for example) and are not easily identifiable using statistical pattern recognition techniques that are based on the raw scattering data. Furthermore, acquiring estimates of these parameters may be useful in a laboratory environment where electromagnetic scattering features of canonical objects are of interest [1,2,17]. It is important to emphasize, however, that the target model given in (1) excludes any multiple interactions between scattering components. Some of the interactions can be identified using higher order spectral analysis techniques [11].

By virtue of the direct relation between target scattering parameters and the wavelet expansion coefficients, wavelet expansion of the impulse responses of radar targets can be used for reliable target identification. The optimality of target recognition using wavelets is related to the fact that compactly supported wavelet functions that are orthogonal to their own dilations have already been developed [7,9,10].

Consider the problem of classifying an unknown radar target represented by the data vector X, to belong to a set of L equiprobable targets. It is assumed that the returned radar signal is corrupted with additive white Gaussian noise with known variance σ^2. The unknown target is assumed to be at zero degrees elevation and at an azimuth $\theta \in [\theta_1, \theta_N]$. Let the received radar signal be

$$y(t) = x(t) + n_t \qquad (18)$$

where x(t) is the target impulse response and n_t is additive white gaussian noise. Let the vectors

Y, X, and n represent samples of the signals y(t), x(t), and n_t. Assume that the vectors $X_1, X_2, \ldots\ldots$ and X_L represent the noiseless impulse responses of all likely targets. Also, let the vectors A and B and N denote the wavelet expansion coefficients of the signals Y and X and

n respectively. Then by the linearity of wavelet transform A=B+N. Furthermore, the elements of A are Gaussian random variables with E{A}=B and variance σ^2. Also, notice that the elements of the vector A are identically independent random variables. Let a_{jk} and $b_{j,k}$ be the elements of A, and B respectively then [13]

$$E\{a_{jk}\} = b_{jk}$$

$$Cov\{a_{jk}\} = \frac{\sigma^2}{2^j}$$

Hence, the wavelet expansion coefficients of the unknown target impulse response represented by A are independent and identically distributed random variables with a diagonal matrix as their covariance matrix. Consequently, in addition to its physical relevance as a scattering identification tool, wavelet based classifier have simple structure and require minimal computation.

Let $p(Y/i,\theta)$ represents the likelihood function of the vector Y given target (i) and azimuth θ. Assume that the noiseless impulse response of each radar target is recorded at several azimuth angles. Then, $p(Y/i)$ can be determined using a mixed Gaussian distribution,

$$p(Y/i) = \frac{1}{N}\sum_{\theta} p(Y/i,\theta) \qquad (19)$$

assuming that all N azimuth angles are equally likely. A decision regarding the identity of the unknown target is then based on maximizing the likelihood function p(Y/i), i.e. choose j such that $p(Y/j) = \max_i\{p(Y/i)\}$.

The above test assumes equally likely hypotheses [16]. If some targets are more likely than others, then the likelihood test may be changed accordingly. It is also possible to employ a sequential decision technique [16] for the purpose of minimizing the average number of observations for a given probability of error. Knowing that the random variables Y are i.i.d, the above test reduces to maximizing

where $a^l_{i,k}$ is the kth wavelet expansion coefficient for the ith target at azimuth l, and K=total number of wavelet coefficients needed to represent the target .

$$p(Y/i) = \frac{1}{N}\sum_{l=1}^{l=N} (2\pi\sigma^2)^{\frac{-K}{2}} e^{\frac{-\sum_{k=1}^{K} (y_k - a_{i,k}^{l})^2}{2\sigma^2}} \qquad (20)$$

Note that if the target azimuth is known the above test reduces to a minimum distance classifier (nearest neighbor). Furthermore, if the dyadic wavelet transform is used without subsampling, then the above classifier is shift invariant and does not require a priori knowledge of the target zero-time reference (no need for absolute phase information). The above decision criterion assumes that all targets have the same number of wavelet expansion parameters K. In the experimental phase of this study, K is determined by the target with the largest number of wavelet coefficients.

Backscattered radar signals representing the Radar Cross Section of five commercial aircraft models recorded in the frequency range 1.5-12 GHz with increments of 50 MHz and at azimuth position $\theta \in [-50, 50]$ degrees with increments of 10 degrees were used in the experimental phase of this study. Figure 4 shows the classification performance of the wavelet based classifier for different azimuth uncertainties. The results in Figure 4 were obtained using 250 experiments with the azimuth position of the unknown target chosen uniformly between -30 to 30 degrees with increments of 10 degrees.

Figure 5 shows a comparison between classification using normalized cross-correlation and using the wavelet based classifier. Both classifiers assume relative phase information. Note that the wavelet based classifier outperforms the cross-correlation classifier, with both classifier requiring comparable computational effort. Two experiments are shown in Figure 7, corresponding to complete and partial azimuth information. Notice that performance of the cross-correlation classifier, in comparison to that of the wavelet classifier, degrades considerably as the target azimuth position becomes uncertain. Figures 4 and 5 indicate that wavelet based classifiers are relatively robust with respect to azimuth. Figure 6 shows classification using all three methods (assuming absolute phase information) but with an extraneous scatterer positioned at t=-.5 ns along the range of the unknown target and has a signal power of 2 dB. The example in Figure 6 simulates the effects of mismatch between the catalog and the unknown target (for example, release of attachments). Figure 6 shows that the wavelet based classifier is superior to

both the raw data based system and the cross-correlation classifiers.

An alternative example is shown in Figure 7 where recognition is made based on a comparison between the impulse response of the unknown target and that of the catalog target. The impulse response of the catalog target, however, is constructed from its wavelet expansion coefficients. Reconstruction from wavelet expansion coefficients does not always produce exact replicas of the target impulse response. The purpose of this experiment is to show the degradation in classification performance due to reconstruction error. Figure 11 shows that such an error has minimal effect on target classification. Furthermore, Figure 11 suggests that the size of catalog radar database may be reduced using wavelet expansion without noticeable degradation in recognition performance.

V. Conclusions

Description and recognition of radar targets using wavelet expansion techniques provides significant insight into the mechanism of scattering from radar targets and contributes to the identification of geometrical features of unknown targets. A wavelet based classifier is not computationally intensive and produces reliable recognition results that are, under certain measurement scenarios, superior to those of other classifiers. The proposed recognition scheme is optimal in a maximum likelihood sense and is particularly suited for recognizing targets with extraneous scatterers and with unknown azimuth position. Wavelet decomposition helps discriminating between scatterers whose response increases with frequency and those whose response decreases with frequency. Furthermore, the proposed scatterer identification method does not require a priori knowledge of the number scattering centers located along the target.

References

[1] R. A. Altes, "Sonar for Generalized Target Description and its Similarity to Animal Echolocation Systems", Journal of the Acoustic Society of America, Vol. 59, No. 1, pp 97-105, 1975.

[2] R. Carriere and R. L. Moses, "High Resolution Parametric Modeling of Canonical Radar Scatterers with Application to Target Identification," IEEE International Conference on Systems Engineering, Dayton, OH, July 1991.

[3] R. Carriere and R. L. Moses, "High Resolution Target Modeling Using a Prony Estimator," IEEE Transactions on Antenna and Propagation, Vol. 40, No. 1, pp 13-18, January, 1992.

[4] R. T. Compton, "Two Dimensional Imaging of Radar Targets with The Music Algorithm," The Ohio State University ElectroScience Laboratory, Report, No. 719276-14, December 1987.

[5] I.Jouny, "Target Description Using Wavelet Transform", International Conference on Signal Processing, San Francisco, CA, March 1992.

[6] S. Kadambe and G. F. Boudreaux-Bartels, "Application of The Wavelet Transform for Pitch Detection of Speech Signals", IEEE Transactions on Information Theory, Vol. 38, No. 2, March 1992.

[7] S. G. Mallat, " A Theory for Multiresolution Signal Decomposition: The wavelet Transform," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 11, No. 7, pp 674-693, July 1989.

[8] P. Flandrin, F. Magand, and M. Zakharia, "Generalized Target Description and Wavelet Decomposition," IEEE Transactions on Signal Processing, Vol. 38, No. 2, pp 350-352, February 1990.

[9] I. Daubechies, "Orthonormal Basis of Compactly Supported Wavelets," Communications on Pure and Applied Mathematics, Vol. XLI, pp 909-996, 1988.

[10] G. Strang, "Wavelets and Dilation Equations: A Brief Introduction," SIAM, Vol. 31, No. 4, pp 614-627, December 1989.

[11] E. K. Walton and I. Jouny, "Bispectral Analysis of Radar Signatures with Application to Target Classification," Radio science, Vol. 25, No. 2, pp 101-113, March 1990.

[12] O. Rioul and M. Vitterli, "Wavelets and Signal Processing," IEEE Signal Processing Magazine, pp 14-38, October 1991.

[13] R. Kronland-Martinet, J. Morlet, and A. Grossman, "Analysis of Sound Patterns Through Wavelet Transform," International Journal of Pattern Recognition and Artificial Intelligence, Vol. 1, No. 2, pp 273-302, 1987.

[14] R. Altes, "Some Invariance Properties of Wideband Ambiguity Functions," Journal of The Acoustical Society of America, Vol. 53, No. 4, pp 1154-1160.

[15] E. K. Walton, "Far Field Measurements and Maximum Entropy Analysis of Lossy Material on a Conducting Plate," IEEE Transactions on Antenna and Propagation, Vol. 37, No. 8, pp 1042-1047, August 1989.

[16] I. Jouny and F. D. Garber, "M-ary Sequential Hypotheses Testing with Application to Target Classification," IEEE Transactions on Aerospace and electronic Systems, April, 1992.

[17] W. Steedly and R. L. Moses, "High Resolution Exponential Modeling of Fully Polarized Radar Returns," IEEE Transactions on Aerospace and Electronic Systems, Vol. 27, No. 3, pp 459-469, May 1991.

NOTE: An expanded version of this report and related material may be available at the Wright-Patterson Avionics Laboratory, Radar Branch AARM-3. For further information, please contact Dr. Jack Bell at (513)-255-6071.

Acknowledgement: The author wishes to thank the Air Force Office of Scientific Research for the opportunity to do this work. Many thanks to the in-house staff at the radar branch AARM-3, at the Wright-Patterson AFB Avionics Lab for many fruitful discussions.

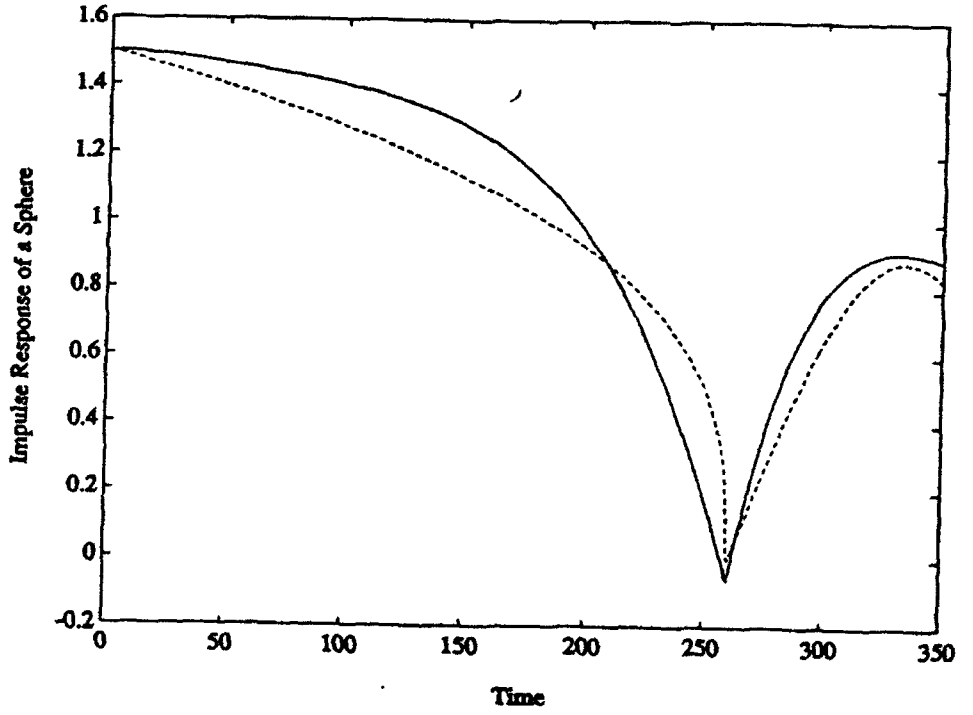


Figure 3: The original impulse response (solid) of a sphere target and its reconstruction using wavelet derived from the Altes signal.

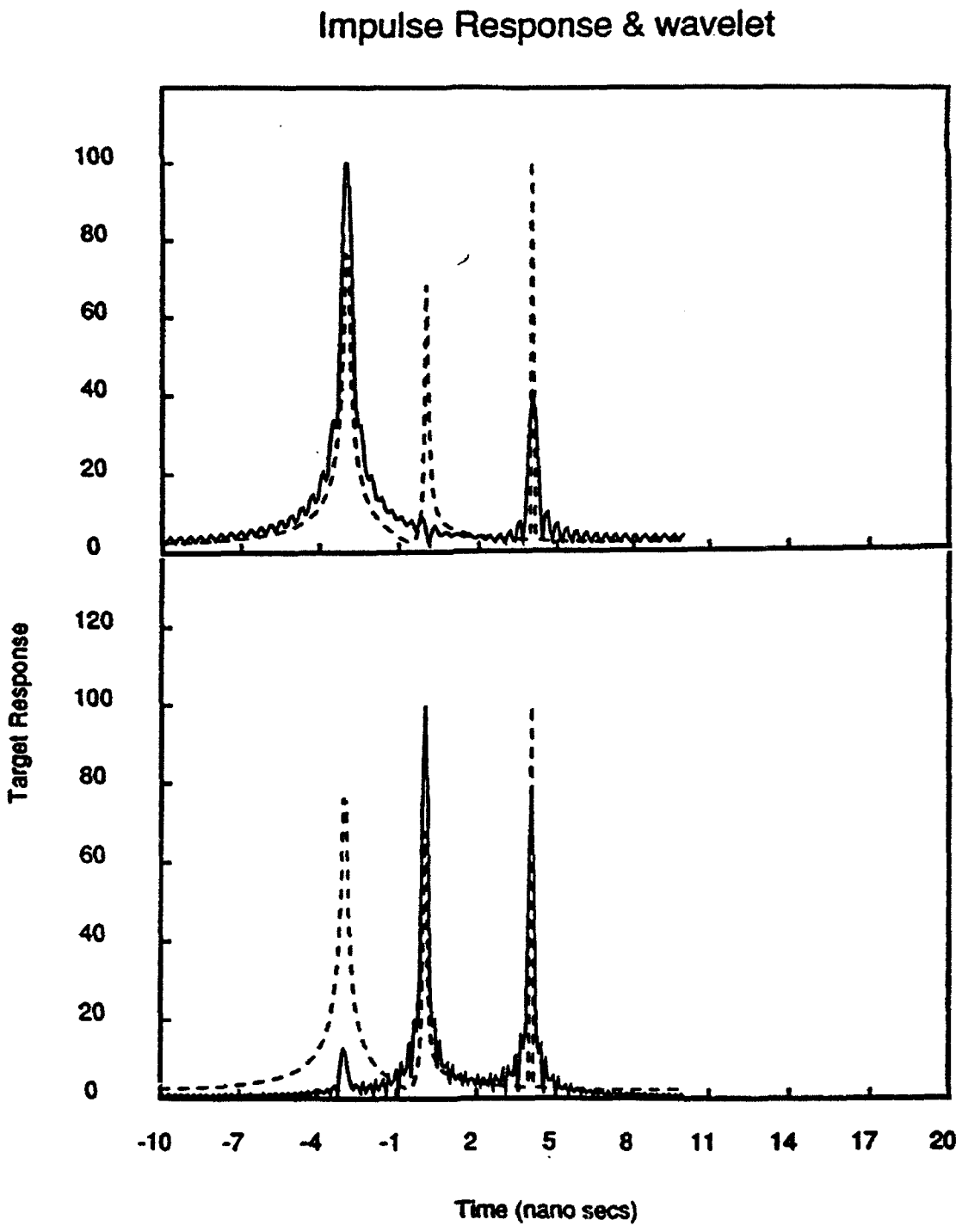


Figure 1: The impulse responses (dotted line) and their wavelet transforms (examined at different scales 2^j) of the hypothetical target $x_2(t)$.

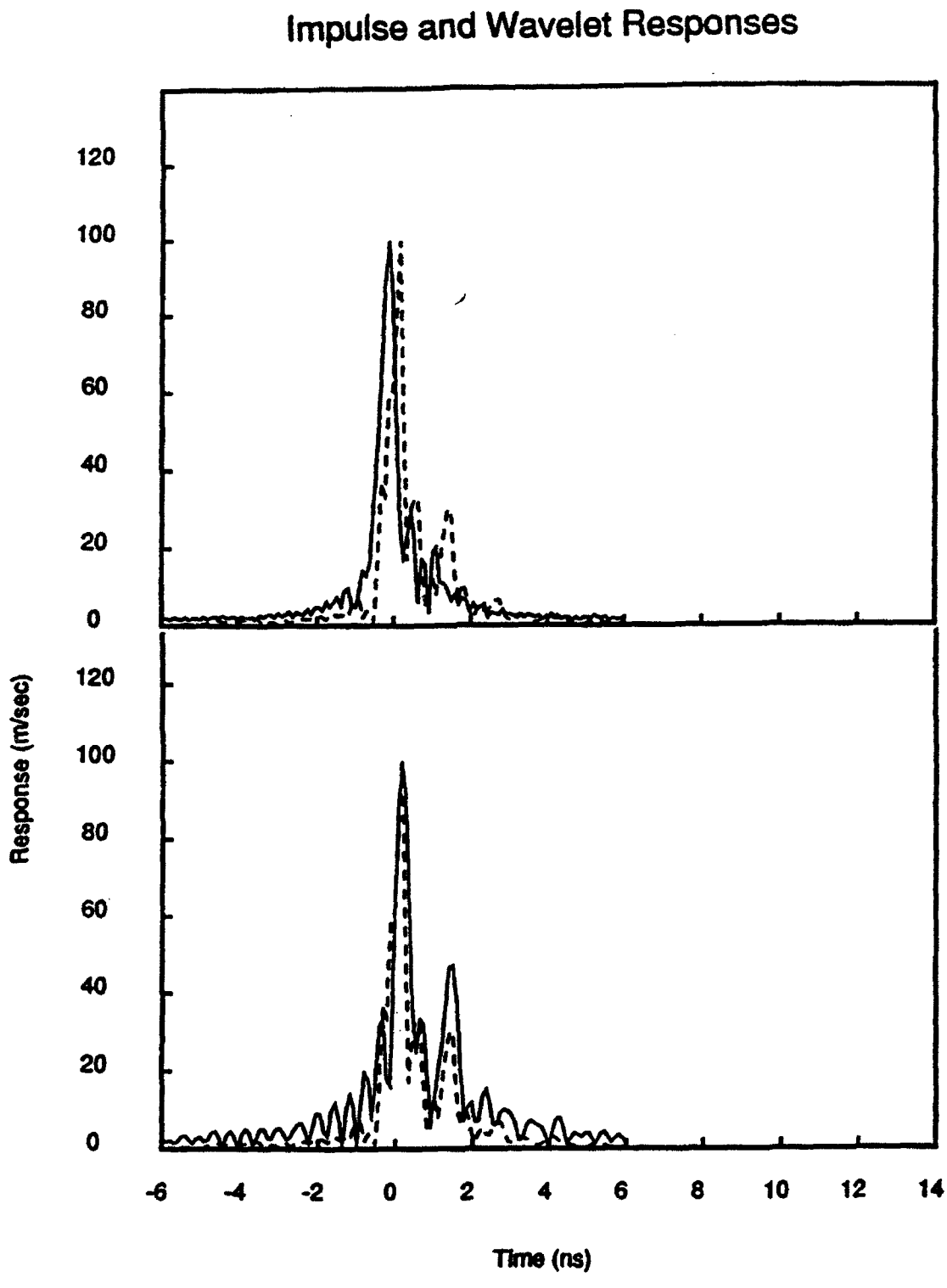


Figure 2: The impulse response (dotted) and its wavelet transform (examined at different scales) of a Concord model aircraft.

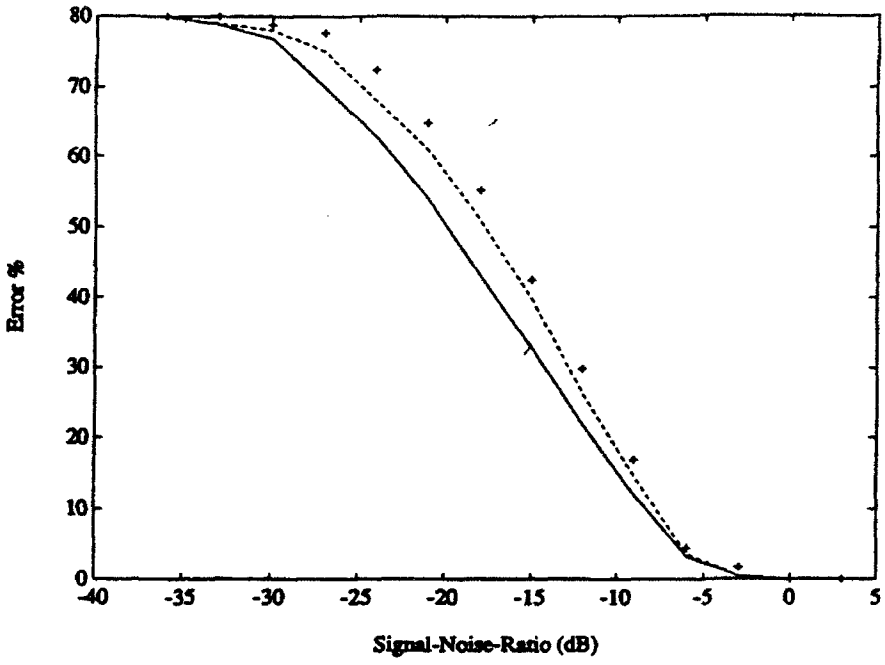


Figure 4: Classification performance of wavelet based target recognition system for different azimuth ambiguities. Target azimuth completely known (solid), known within ±10 degrees (dashed), and known within ±20 degrees (marked by +).

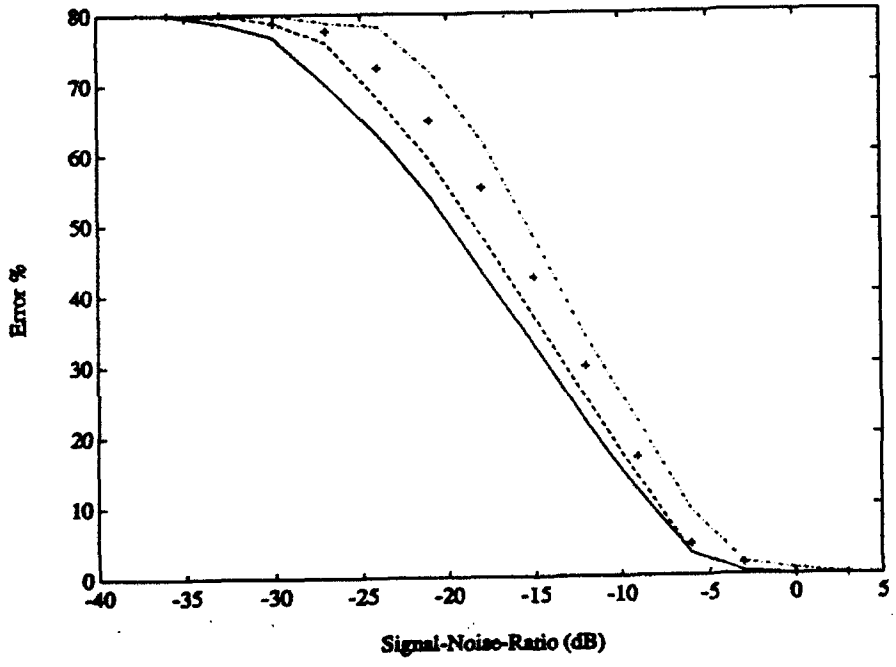


Figure 5: Comparison between target classification using wavelets and using cross-correlation technique. Target azimuth completely known: wavelet classifier (solid) and cross-correlation (marked by +). Target azimuth known within ±20 degrees: wavelet classifier (dashed) and cross-correlation classifier (dashed-dotted).

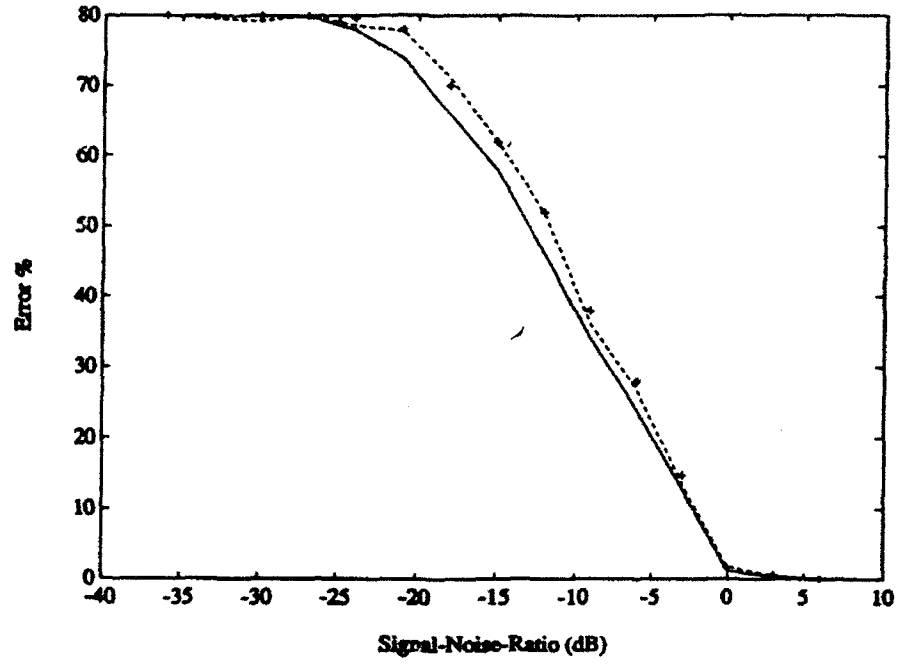


Figure 6: Classification performance of three different techniques: wavelets (solid), maximum likelihood applied to the raw data (dashed), and cross-correlation (marked by +). An extraneous scatterer with 2 dB power was placed at t= -0.5 ns along the unknown target.

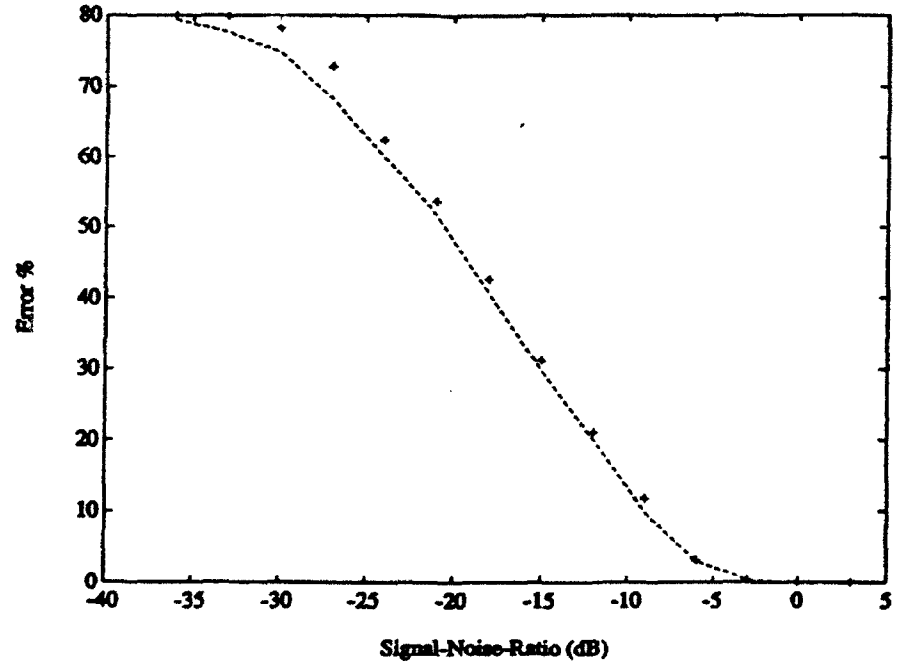


Figure 7: Performance of the maximum likelihood classifier using raw radar data (dashed) and using data generated from the targets wavelet expansion coefficients (marked by +).

AXISYMMETRIC THERMOELASTIC RESPONSE OF A COMPOSITE CYLINDER CONTAINING AN ANNULAR MATRIX CRACK AND A FRICTIONAL INTERFACE

Autar K. Kaw
Associate Professor
Department of Mechanical Engineering

University of South Florida
ENG 118, 4202 East Fowler Avenue
Tampa, FL 33620-5350.

Final Report for:
Summer Research Program
Materials Laboratory
Wright-Patterson AFB

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

August 1992

AXISYMMETRIC THERMOELASTIC RESPONSE OF A COMPOSITE CYLINDER CONTAINING AN ANNULAR MATRIX CRACK AND A FRICTIONAL INTERFACE

Autar K. Kaw
Associate Professor
Department of Mechanical Engineering
University of South Florida

Abstract

The effect of a frictional interface on the response of a unidirectional ceramic matrix composite under a remote axial tensile strain is studied. The geometry of the composite is approximated by a concentric cylinder model with a transverse annular crack in the matrix. The fiber-matrix interface follows the Coulomb friction law. On applying the boundary and the continuity conditions at the interface, the solution is obtained in terms of three coupled integral equations.

An asymptotic analysis of the coupled integral equations will reveal the nature of the singularity of the stress and strain functions in the fiber and the matrix. The extent of the interfacial damage and the stress fields in the fiber and the matrix will be studied as a function of the fiber volume fraction, the elastic moduli of the constituents, the thermal expansion coefficients of the constituents, the coefficient of friction, the temperature change and the remote uniform axial strain. The objective of the study is to give an idea of what features accelerate or inhibit the interfacial damage.

An AFOSR Research Initiation proposal is being submitted for continuing the work initiated in this report.

AXISYMMETRIC THERMOELASTIC RESPONSE OF A COMPOSITE CYLINDER CONTAINING AN ANNULAR MATRIX CRACK AND A FRICTIONAL INTERFACE

Autar K. Kaw

INTRODUCTION

Ceramic matrix composites are becoming attractive as load bearing structures for high temperature and corrosive atmosphere applications. Although these composites have higher ultimate strength and strain than monolithic ceramics, matrix cracking followed by interfacial failure is a critical issue in their use. When the matrix crack reaches the fiber, the interface may open or slip. This blunts the crack, and slows and arrests the propogation of the matrix crack. Although this increases the fracture toughness of the composite, the damage in the interface reduces the axial strength, the transverse strength and the axial compressive strength of the composite (Steif, 1984). Because of these conflicting requirements, it becomes important to fully understand the mechanics of matrix fracture in ceramic matrix composites.

The pioneering work on the micromechanical failure modeling of matrix cracking of ceramic matrix fiber reinforced composites was done by Aveston, Cooper and Kelly (1971), commonly called the ACK theory. The ACK theory has also been extended and improved by several authors. These theories are reviewed by McCartney (1987). These theories have given a better understanding of the strength and toughness of ceramic matrix composites. They have also led to common misconceptions in brittle matrix composites, such as, that the steady state cracking mode is responsible for matrix crack initiation (Kim and Pagano, 1990). These theories are based on simplifications and may ignore coupling between different mechanisms.

In this paper, the authors have developed a solution which fully accounts for all the equations of equilibrium and stress-strain, and boundary

conditions for an axisymmetric composite cylinder. The fiber-matrix interface is considered to be imperfect. The interface can have open, slip and stick zones and their lengths are found automatically based on material, geometry and loading (thermal and mechanical) conditions. The model also accounts for the finite fiber volume fractions and the mismatch of thermal expansion coefficients and the elastic moduli of the fiber and the matrix.

Several axisymmetric elasticity solutions exist for crack problems in homogeneous solids under a uniaxial tensile load. A brief literature survey of these solutions is given in Kaw and Pagano (1992). The problem of a crack in a composite cylinder was first attempted by Ford (1973). He solved the problem of a composite cylinder made of a solid cylinder (fiber) bonded to a surrounding hollow cylinder (matrix) of infinite outer radius. A penny shaped crack was assumed in the fiber. The composite cylinder was subjected to axial tension. He approximated the fiber by a solid cylinder and the matrix by a hollow cylinder of infinite outer radius. The fiber and the matrix were considered to be perfectly bonded. The problem was formulated in terms of a the axial displacements of the crack surface. The boundary condition of a traction free surface was satisfied only in the least square sense. The stress intensity factors at the crack tips were studied as a function of the fiber-matrix moduli ratio.

Only recently, Wijeyewickrema, et. al. (1991) solved the same problem of a damaged composite cylinder, except for that the damage considered was now in the form of an annular crack in the matrix. The composite cylinder was subjected to a uniform axial tensile strain at infinity. The solution was obtained in terms of a Cauchy singular integral equation. The stress intensity factors at the crack tips were studied as a function of the ratio of the fiber to matrix moduli, and the location and the length of the crack. Wijeyewickrema and Keer (1992) extended this problem to include a finite volume fraction of the fibers by assuming a finite outer radius of the outer cylinder.

The interface between the fiber and the matrix was assumed to be perfect in the above studies (Ford, 1973, Wijeyewickrema, et.al, 1991, Wijeyewickrema and Keer, 1992a) of axisymmetric composite problems. In ceramic matrix composites, the interface may not be perfect. Matrix cracks can result in slipping and opening of fiber-matrix interfaces. This phenomenon blunts the crack and significantly increases the fracture toughness of the composite system. Hence, one needs to study the problem which accounts for this phenomena at the interface. Some studies which have

relaxed the assumption of the perfect interface in the composite cylinder failure model include the work of Kaw and Pagano (1992), McCartney (1990), Wijeyewickerma and Keer (1992) and Schweitert and Steif (1991).

Kaw and Pagano (1992) included an imperfect interface in the composite cylinder model by approximating the interface by distributed shear springs. They also included the effect of both thermal expansion and elastic moduli mismatch of the constituents for the perfect bond case. Wijeyewickerma and Keer (1992) included the effect of an imperfect interface by assuming a finite slip length along the interface. In that study, they assumed the length of the slip zone as well as the value of the shear stress in the slip zone. McCartney (1990) allowed the frictional interface to follow the Coulomb frictional law but his model approximated two stress-strain equations and the displacement boundary conditions in the average sense. Steif, et. al. (1988, 1991) accounted for the frictional interface as well as all the equations of elasticity. Dollar and Steif (1988) solved the problem of a layered model with the fiber layer with a crack which impinges on the interface. They assumed the interface to follow Coulomb's friction law but assumed the fiber and the matrix to have the same moduli.

The assumption of equal material properties for the fiber and matrix was relaxed later by Schweitert and Steif (1991). They solved the axisymmetric problem of a solid cylinder (fiber) surrounded by a hollow cylinder of infinite outer radius (matrix). The fiber-matrix interface followed the Coulomb friction law. The composite geometry was subjected to a remote mechanical axial strain and a remote compressive radial stress. The remote radial stress represented residual stresses due to thermal expansion mismatch of the fiber and the matrix. In the present study, Schweitert and Steif's (1991) model is extended as follows.

The present model accounts for a finite fiber volume fraction, allows the crack to be in the matrix rather than in the fiber, accounts for the thermal expansion mismatch of the fiber and the matrix directly rather than by a constant remote radial load on the matrix. Moreover, the matrix crack also can be internal, edge and/or touching the interface.

Also, the mathematical formulation of the problem does not follow the novel boundary element procedure of Schweitert and Steif (1991). Instead, the formulation is based on the Fourier Integral Transforms technique. Coupled with some new definitions of unknown functions in the formulation, it reduces the problem to a set of three coupled integral equations. These

equations have the unknowns in terms of the matrix crack opening displacement, and the radial and the axial displacement jump along the interface in the open and slip zones. Since all these unknowns are non-zero for a finite length, unlike in Schweitert and Steif's (1991) model, it allows application of traditional numerical algorithms and more convenient convergence studies. Also the stresses in the fiber at the interface can be calculated directly without any extrapolation techniques such as used in their study.

In the sections to follow is the formulation of the model given in Figure 1. The boundary and interface conditions are applied to give three coupled integral equations. An asymptotic analysis will be conducted on the integral equations, if required, to find the nature of the unknown functions at the end points. The numerical technique to solve the coupled integral equations will be based on the asymptotic analysis results. In the results, the damage zones (slip and open zones) at the interface and the stress concentration in the fiber will be found for different composite systems. Several parametric studies will be done, particularly, as a function of the fiber volume fraction and the coefficient of friction.

FORMULATION

The geometry of the composite cylinder consists of an infinitely long fiber bonded to an annular matrix of finite outer radius (Figure 1). This geometry approximates a representative volume element (RVE) of a composite in a hexagonal array. The cylindrical coordinates are denoted by r,θ and z, where u_r and u_z are the radial and axial displacements, respectively. The normal and shear stresses are denoted by σ_{rr}, σ_{zz}, $\sigma_{\theta\theta}$ and σ_{rz}. The indices 0 and 1 stand for the fiber and the matrix, respectively.

The fiber is approximated by a linearly elastic, isotropic, homogeneous and infinitely long solid cylinder of radius a, shear modulus μ_0, Poisson's ratio ν_0, Young's modulus E_0, coefficient of thermal expansion, α_0. The matrix is approximated by a linearly elastic, isotropic and homogeneous and infinitely long annular cylinder of inner radius a and outer radius d, shear modulus μ_1, Poisson's ratio ν_1, Young's modulus, E_1, coefficient of thermal expansion, α_1. The matrix consists of an annular crack of length 'c_1-b_1'

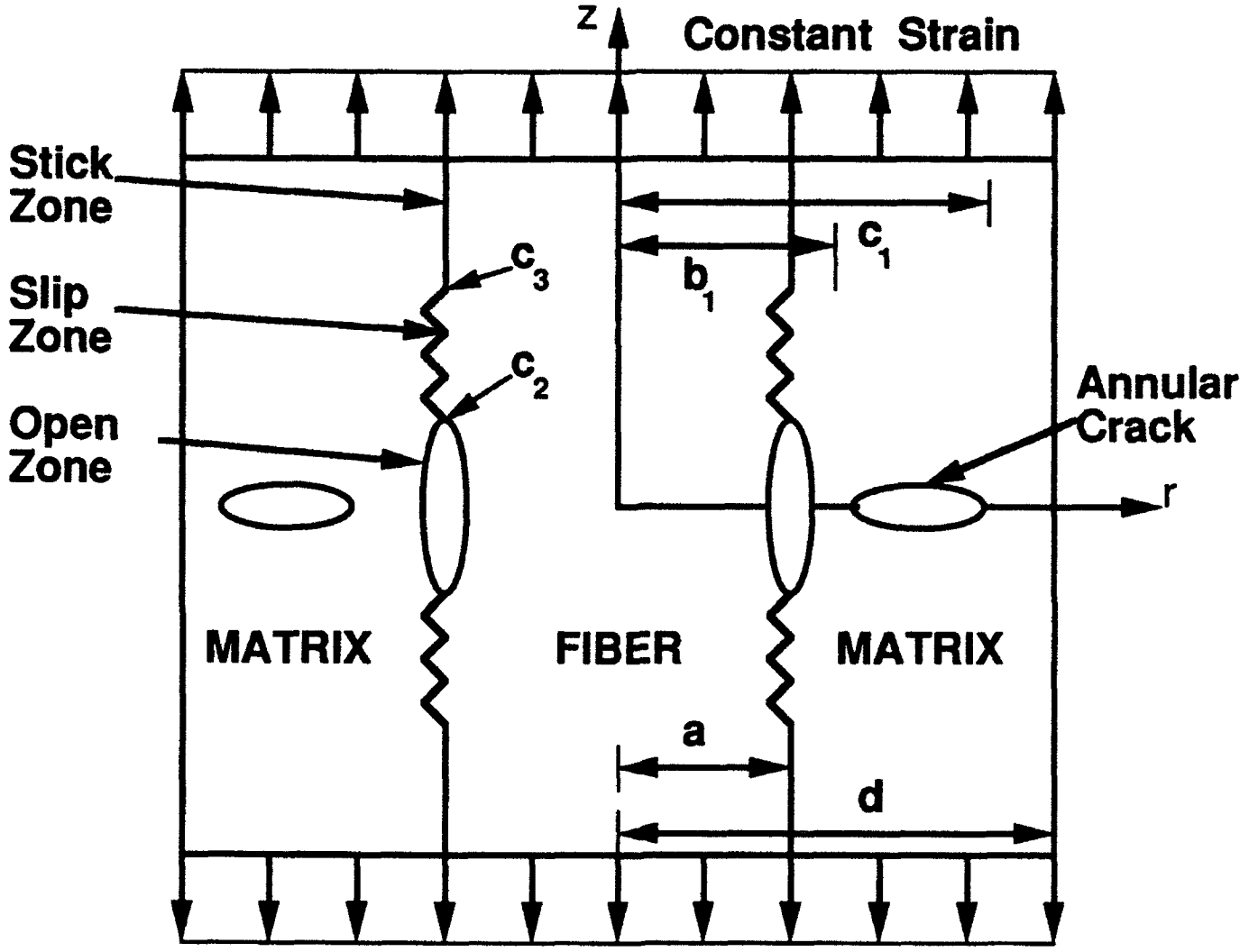


Figure 1. Annular Matrix Crack in Brittle Matrix Composites in Presence of a Frictional Interface and under Thermomechanical Loading.

$(a \le b_1 < c_1 \le d)$ in the z=0 plane, at a distance of 'b_1-a' from the interface. The volume fraction of the fibers is given by a^2/d^2.

The bond between the fiber and the matrix is considered to be imperfect and is approximated appropriately by open, slip and stick zones. The open zones are traction free; the slip zones follow the Coulomb friction law; and the stick zones have continuous displacements, and radial and shear stresses. The radial stress in the slip and stick zones are compressive. The length of the open, slip and stick zones is not prescribed. The length of the open zone is denoted by 'c_2', while the length of the slip zone is denoted by 'c_3-c_2'. The kinetic and sliding friction coefficient are considered to be equal.

The composite is subjected to an axial remote strain, ϵ_0 on the ends plus a constant temperature change, ΔT. The temperature change, ΔT and Poisson's ratio mismatch are assumed to lead to an overall compressive stress at the interface in the uncracked body. This maintains radial contact at the interface in the undamaged zones of the interface.

The displacement and stress field for axially symmetric problems with a temperature change, (ΔT) and applied remote axial strain, (ϵ_0) with an annular matrix crack is given as follows.

The displacement field for the matrix is

$$u_r^1(r,z) = -\frac{1}{2\mu_1}\Big(\frac{2}{\pi}\int_0^\infty [f_1(s)\, I_1(rs) + f_2(s)\, rs\, I_0(rs) - f_3(s)\, K_1(rs) - f_4(s)\, rs\, K_0(rs)]\, s^2 \cos(sz)\, ds - \int_0^\infty f_7(p)\, p^3(1-2\nu_1-zp)e^{-zp}\, J_1(rp)\, dp\Big) + u_r^{1T}(r,z) + u_r^{1M}(r,z), \qquad (1.a)$$

$$u_z^1(r,z) = \frac{1}{2\mu_1}\Big(\frac{2}{\pi}\int_0^\infty [f_1(s)\, I_0(rs) + f_2(s)\, [4(1-\nu_1)\, I_0(rs) + rs\, I_1(rs)] + f_3(s) K_0(rs) + f_4(s)\, [rs\, K_1(rs) - 4(1-\nu_1)K_0(rs)]]\, s^2 \sin(sz)\, ds - \int_0^\infty f_7(p)\, p^3[2(1-\nu_1)+zp]e^{-zp} J_0(rp)\, dp\Big) + u_z^{1T}(r,z) + u_z^{1M}(r,z) \qquad (1.b)$$

The stress field for the matrix is given by

$$\sigma^1_{rr}(r,z) = \frac{2}{\pi}\int_0^\infty \{f_1(s)[-I_0(rs)+I_1(rs)/(rs)]+ f_2(s)[(2\nu_1-1)I_0(rs) - rsI_1(rs)]- f_3(s)\,[K_0(rs)+K_1(rs)/(rs)] + f_4(s)[(1-2\nu_1)K_0(rs)-rsK_1(rs)]\}\, s^3\cos(sz)\, ds - \int_0^\infty f_7(p)\, p^3[p(1-zp)J_0(rp)-\frac{1-2\nu_1-zp}{r} J_1(rp)]e^{-zp}dp + \sigma^{1T}_{rr}(r,z) + \sigma^{1M}_{rr}(r,z), \quad (2.a)$$

$$\sigma^1_{zz}(r,z) = \frac{2}{\pi}\int_0^\infty \{f_1(s)I_0(rs)+f_2(s)\,[2(2-\nu_1)I_0(rs)+rsI_1(rs)] +f_3(s)\,K_0(rs) + f_4(s)[\,-2(2-\nu_1)K_0(rs)+rs\,K_1(rs)]\}s^3\cos(sz)\, ds + \int_0^\infty f_7(p)\, p^4(1+zp)e^{-zp}\, J_0(rp)\, dp + \sigma^{1T}_{zz}(r,z) + \sigma^{1M}_{zz}(r,z), \quad (2.b)$$

$$\sigma^1_{rz}(r,z) = \frac{2}{\pi}\int_0^\infty \{f_1(s)I_1(rs)+f_2(s)[rsI_0(rs)+2(1-\nu_1)I_1(rs)] - f_3(s)K_1(rs) + f_4(s)[\,2(1-\nu_1)K_1(rs)-rs\,K_0(rs)]\}s^3\sin(sz)\, ds + \int_0^\infty f_7(p)\, p^5 z e^{-zp}\, J_1(rp)\, dp, \quad (2.c)$$

$$\sigma^1_{\theta\theta}(r,z) = \frac{2\,\mu_1(1+\nu_1)}{r}\,[u^1_r(r,z) - \alpha_1\,\Delta T\, r] + \nu_1[\sigma^1_{rr}(r,z) + \sigma^1_{zz}(r,z)]. \quad (2.d)$$

In equations (5-6), the functions f_i(i=1,..4,7) are the unknown functions, $K_n()$ is the modified Bessel's functions of the second kind of order n, $I_n()$ is the modified Bessel's functions of the first kind, and $J_n()$ is the Bessel's functions of the first kind of order n.

The displacement field for the fiber is given by

$$u^0_r(r,z) = -\frac{1}{2\mu_0}\left[\frac{2}{\pi}\int_0^\infty \{f_5(s)\, I_1(rs) + f_6(s)\, rs\, I_0(rs)\}\, s^2\cos(sz)ds\right] + u^{0T}_r(r,z) + u^{0M}_r(r,z), \quad (3.a)$$

$$u_z^0(r,z) = \frac{1}{2\mu_0}\left[\frac{2}{\pi}\int_0^\infty \{f_5(s)\, I_0(rs) + f_6(s)[4(1-\nu_0)I_0(rs) + rs\, I_1(rs)\}\right.$$

$$\left. s^2 \sin(sz)\, ds\right] + u_z^{0T}(r,z) + u_z^{0M}(r,z). \qquad (3.b)$$

The stress field for the fiber is given by

$$\sigma_{rr}^0(r,z) = \frac{2}{\pi}\int_0^\infty \{f_5(s)[-I_0(rs)+I_1(rs)/(rs)]$$

$$+f_6(s)[(2\nu_0-1)I_0(rs)-rsI_1(rs)]\}\, s^3 \cos(zs)\, ds$$

$$+ \sigma_{rr}^{0T}(r,z) + \sigma_{rr}^{0M}(r,z), \qquad (4.a)$$

$$\sigma_{zz}^0(r,z) = \frac{2}{\pi}\int_0^\infty \{f_5(s)I_0(rs)+f_6(s)\,[2(2-\nu_0)I_0(rs)+rsI_1(rs)]\}$$

$$s^3\cos(zs)\, ds + \sigma_{zz}^{0T}(r,z) + \sigma_{zz}^{0M}(r,z), \qquad (4.b)$$

$$\sigma_{rz}^0(r,z) = \frac{2}{\pi}\int_0^\infty \{f_5(s)I_1(rs)+f_6(s)[rsI_0(rs)+2(1-\nu_0)I_1(rs)]\}s^3\, \sin(zs)ds, \qquad (4.c)$$

$$\sigma_{\theta\theta}^0(r,z) = \frac{2\,\mu_0(1+\nu_0)}{r}\,[u_r^0(r,z) - \alpha_0\,\Delta T\, r]$$

$$+ \nu_0[\sigma_{rr}^0(r,z) + \sigma_{zz}^0(r,z)]. \qquad (4.d)$$

In equations (3) and (4), the functions f_i, i=5,6 are the unknown functions.

In the above equations (1-4), σ_{rr}^{iT}, σ_{zz}^{iT}, u_r^{iT}, u_z^{iT}, σ_{rr}^{iM}, σ_{zz}^{iM}, u_r^{iM}, u_z^{iM}, i=0,1 are stresses and displacements due to the thermal expansion mismatch and remote strain in the undamaged composite cylinder. The superscript 'T' corresponds to the effect of temperature change, while the superscript 'M' corresponds to the effect of the axial remote tensile strain. These stresses and displacements are added to the integral transform forms to preserve the far field stresses since the integral transforms forms contribute vanishing stresses as $z\to\infty$.

Unlike the case of a perfect bond, superposition of stresses is not admissible in the current problem. This is because the current model follows the non-linear law because the interface can have a slip zone governed by the Coulomb friction law.

The continuity conditions at the interface between the fiber and the matrix at r=a are given by

$$\sigma^0_{rr}(a,z) = \sigma^1_{rr}(a,z),\ 0 \le |z| < \infty, \tag{5.a}$$

$$\sigma^0_{rz}(a,z) = \sigma^1_{rz}(a,z),\ 0 \le |z| < \infty. \tag{5.b}$$

Also at the interface (r=a) between the fiber and the matrix

Open zone

$$\sigma^0_{rz}(a,z) = 0,\ 0 < |z| < c_2, \tag{5.c}$$

$$\sigma^0_{rr}(a,z) = 0,\ 0 < |z| < c_2, \tag{5.d}$$

constrained by

$$u^1_r(a,z) - u^0_r(a,z) > 0,\ 0 < |z| < c_2, \tag{5.e}$$

Slip zone

$$u^0_r(a,z) = u^1_r(a,z),\ c_2 < |z| < c_3, \tag{5.f}$$

$$|\sigma^0_{rz}(a,z)| = |\rho\ \sigma^0_{rr}(a,z)|,\ c_2 < |z| < c_3, \tag{5.g}$$

constrained by

$$\mathrm{sgn}\left\{\frac{d}{dt}\left[u^1_z(a,z) - u^0_z(a,z)\right]\right\} = \mathrm{sgn}\left\{\sigma^0_{rz}(a,z)\right\},\ c_2 < |z| < c_3, \tag{5.h}$$

$$\sigma^0_{rr}(a,z) < 0,\ c_2 < |z| < c_3, \tag{5.i}$$

Stick zone

$$u^0_r(a,z) = u^1_r(a,z),\ c_3 \le |z| < \infty, \tag{5.j}$$

$$u^0_z(a,z) = u^1_z(a,z),\ c_3 \le |z| < \infty, \tag{5.k}$$

constrained by

$$\sigma^0_{rr}(a,z) < 0,\quad c_3 \le |z| < \infty, \tag{5.l}$$

$$|\sigma^0_{rz}(a,z)| < |\rho\ \sigma^0_{rr}(a,z)|,\ c_3 \le |z| < \infty. \tag{5.m}$$

The friction coefficient 'ρ' is assumed to be constant in the slip zone. Equation (5.f) suggests that the radial contact is maintained in the slip zone. This assumption is justified as the uncracked composite cylinder is assumed to have compression at the interface due to the mismatch of thermal expansion coefficients and Poisson's ratios of the fiber and the matrix.

Equation (5.h) is the condition of positive energy dissipation which implies that the shear stress and the increment of axial slip be consistent.

The boundary conditions at the matrix edge r=d are given by

$$\sigma_{rz}^{1}(d,z) = 0, \quad 0 \le |z| < \infty, \tag{6.a}$$

and

EITHER

$$u_r^{1}(d,z) = u_r^{1M}(d,z) + u_r^{1T}(d,z), \quad 0 \le |z| < \infty, \tag{6.b}$$

OR

$$\sigma_{rr}^{1}(d,z) = 0, \quad 0 \le |z| < \infty. \tag{6.c}$$

Equations (6.b) and (6.c) maximize and minimize the constraint on the composite cylinder, respectively. Equation (6.b) is preferred in this study since it allows the slope of the crack surface $\partial u_z^1(r,0)/\partial r$ at r=d to be equal to zero for the edge crack problem. Near the crack surface the lateral restraint from the neighboring fibers in the vicinity of the crack will be different from as given by equation (6.b), but the two conditions (6.b) and (6.c) are equivalent for points at r=d which are removed from the crack and interfacial damage surfaces (Pagano and Brown, 1992).

The shear stresses in the composite cylinder at z=0 are

$$\sigma_{rz}^{0}(r,0) = 0, \quad r < a, \tag{7.a}$$

$$\sigma_{rz}^{1}(r,0) = 0, \quad a < r < d. \tag{7.b}$$

The shear stress conditions (7.a) and (7.b) are automatically satisfied by the shear stress equations (2.c) and (4.c).

The other boundary conditions at z=0 are

$$u_z^{0}(r,0) = 0, \quad 0 < r \le a, \tag{8.a}$$

$$u_z^{1}(r,0) = 0, \quad a < r \le b_1, \; c_1 < r < d, \tag{8.b}$$

$$\sigma_{zz}^{1}(r,0) = 0, \quad b_1 < r < c_1, \tag{8.c}$$

subject to

$$u_z^{1}(r,0) \ge 0, \quad b_1 < r < c_1. \tag{8.d}$$

Introduce a new unknown slope function $\phi_1(r)$ as follows

$$\frac{\mu_1}{1-\nu_1} \frac{\partial}{\partial r} u_z^{1}(r,0) = \phi_1(r), \quad b_1 < r < c_1. \tag{9}$$

From equations (1.b), (8.b) and (9), one gets

$$p^3 f_7(p) = \int_{b_1}^{c_1} t_1 \phi_1(t_1)\, J_1(pt_1)\, dt_1. \tag{10}$$

Introduce two new displacement jump functions as

$$\phi_2(z) = \mu_0\, [u_r^1(a,z) - u_r^0(a,z)],\ 0 \le |z| < \infty, \tag{11}$$

$$\phi_3(z) = \mu_0\, [u_z^1(a,z) - u_z^0(a,z)],\ 0 \le |z| < \infty. \tag{12}$$

Now note from equations (11), (12), (5.f), (5.j) and (5.k) that

$$\phi_2(z) = 0,\ c_2 < |z| < \infty, \tag{13}$$

$$\phi_3(z) = 0,\ c_3 < |z| < \infty. \tag{14}$$

The above manipulation is done because the functions, ϕ_1, ϕ_2 and ϕ_3 are now unknown only on a finite domain (b_1,c_1), $(0,c_2)$ and $(0,c_3)$, respectively.

Conducting some algebraic manipulations while using exact integrals and recursion formulas given in (Kaw and Pagano, 1992: Appendix C), the six equations (5.a-b, 6.a, 6.b or 6.c, 11, 12) result in the following simultaneous set of six equations,

$$\sum_{j=1}^{6} A_{ij}\, f_j(s) = \frac{1}{s^3} \int_{b_1}^{c_1} \phi_1(t_1)\, h_{i1}(s,t_1)\, dt_1 + \frac{1}{s^3} \int_{b_2=0}^{c_2} \phi_2(t_2)\, h_{i2}(s)\, \mathrm{Cos}(st_2)\, dt_2$$

$$+ \frac{1}{s^3} \int_{b_3=0}^{c_3} \phi_3(t_3)\, h_{i3}(s)\, \mathrm{Sin}(st_3)\, dt_3,\ i=1,..6, \tag{15.a-f}$$

where A_{ij}, $j=1,..6, i=1,..,6$, and h_{ik}, $i=1,..,6$, $k=1,2,3$, are given as follows.

$$A_{11} = I_0(as) - \frac{I_1(as)}{as},\quad A_{12} = (1-2\nu_1)\, I_0(as) + as\, I_1(as),$$

$$A_{13} = K_0(as) + \frac{K_1(as)}{as},\quad A_{14} = -(1-2\nu_1)\, K_0(as) + as\, K_1(as),$$

$$A_{15} = -I_0(as) + \frac{I_1(as)}{as},\quad A_{16} = -(1-2\nu_0)\, I_0(as) - as\, I_1(as),$$

$$A_{21} = -I_1(as),\ A_{22} = -as\, I_0(as) - 2(1-\nu_1) I_1(as),\ A_{23} = K_1(as),$$

$$A_{24} = as\, K_0(as) - 2(1-\nu_1) K_1(as),\ A_{25} = I_1(as),$$

$$A_{26} = -as I_0(as) + 2(1-\nu_0) I_1(as),$$

If equation (6.b) holds, A_{3j}'s are given by

$A_{31}= I_1(ds)$, $A_{32}= ds\ I_0(ds)$

$A_{33}= -K_1(ds)$, $A_{34}= -ds\ K_0(ds)$, $A_{35}= 0$, $A_{36}= 0$,

If equation (6.c) holds, A_{3j}'s are given by,

$$A_{31}= -I_0(ds) + \frac{I_1(ds)}{ds}\ ,\quad A_{32}= (2\nu_1-1)\ I_0(ds) - ds\ I_1(ds),$$

$$A_{33}= -K_0(ds) - \frac{K_1(ds)}{ds}\ ,\quad A_{34}= (1-2\nu_1)\ K_0(ds) - ds\ K_1(ds),$$

$A_{35}=0$, $A_{36}=0$,

$A_{41} = I_1(ds)$, $A_{42} = ds\ I_0(ds)+2(1-\nu_1)\ I_1(ds)$, $A_{43}= -K_1(ds)$,

$A_{44} = -ds\ K_0(ds) + 2(1-\nu_1)\ K_1(ds)$, $A_{45} = 0$, $A_{46} = 0$,

$$A_{51}= -\frac{\bar{\mu}}{2}\ I_1(as),\quad A_{52}= -\frac{\bar{\mu}}{2}\ as\ I_0(as),\quad A_{53} = \frac{\bar{\mu}}{2}\ K_1(as),$$

$$A_{54}= \frac{\bar{\mu}}{2}\ as\ K_0(as),\quad A_{55} = \frac{1}{2}\ I_1(as),\quad A_{56} = \frac{1}{2}\ as\ I_0(as),$$

$$A_{61}= \frac{\bar{\mu}}{2}\ I_0(as),\quad A_{52}= \frac{\bar{\mu}}{2}\ [4(1-\nu_1)I_0(as)+asI_1(as)],$$

$$A_{63} = \frac{\bar{\mu}}{2}\ K_0(as),\quad A_{64}= \frac{\bar{\mu}}{2}\ [\ as\ K_1(as)-4(1-\nu_1)K_0(as)],$$

$$A_{65} = -\frac{1}{2}\ I_0(as),\quad A_{66} = -\frac{1}{2}\ [4(1-\nu_0)I_0(as)+as\ I_1(as)],\quad \bar{\mu} = \frac{\mu_0}{\mu_1}. \qquad (16)$$

$$h_{11}(s,t_1)=s\{t_1sI_0(as)K_0(t_1s) + I_0(as)\ K_1(t_1s) - \frac{t_1}{a}\ I_1(as)K_0(t_1s)$$
$$-[as+\frac{2(1-\nu_1)}{as}\ ]\ I_1(as)K_1(t_1s)\},$$

$$h_{21}(s,t_1)=s\{asI_0(as)K_1(t_1s)-t_1sI_1(as)K_0(t_1s)\},$$

If equation (6.b) holds,

$$h_{31}(s,t_1)=s\{-t_1sI_0(t_1s)K_1(ds) +ds\ I_1(t_1s)K_0(ds)$$
$$+ 2(1-\nu_1)I_1(t_1s)K_1(ds)\},$$

If equation (6.c) holds,

$$h_{31}(s,t_1)=-s\{t_1sI_0(t_1s)K_0(ds) -I_1(t_1s)K_0(ds)+ \frac{t_1}{d}I_0(t_1s)K_1(ds)$$

$$-[ds+\frac{2(1-\nu_1)}{ds}]I_1(t_1s)K_1(ds)),$$

$$h_{41}(s,t_1)=s(-t_1sI_0(t_1s)K_1(ds)+dsI_1(t_1s)K_0(ds)),$$

$$h_{51}(s,t_1) = -\frac{\bar{\mu}s}{2}(-asI_0(as)K_1(t_1s)+t_1sI_1(as)K_0(t_1s)$$
$$+2(1-\nu_1)I_1(as)K_1(t_1s)),$$

$$h_{61}(s,t_1)= -\frac{\bar{\mu}s}{2}(-t_1sI_0(as)K_0(t_1s)+2(1-\nu_1)I_0(as)K_1(t_1s)$$
$$+asI_1(as)K_1(t_1s)),$$

$$h_{12}(s)=h_{22}(s)=h_{32}(s)=h_{42}(s)=h_{62}(s)=0,\ h_{52}(s)=s,$$

$$h_{13}(s)=h_{23}(s)=h_{33}(s)=h_{43}(s)=h_{53}(s)=0,\ h_{63}(s)=s. \qquad (17)$$

The six equations (15.a-f) can be solved for $f_i(j=1,..6)$ and expressed as

$$f_i(s) = \frac{1}{s^3}\left[\int_{b_1}^{c_1}\phi_1(t_1)\ f_{i1}(t_1,s)\ dt_1 + \int_{b_2=0}^{c_2}\phi_2(t_2)\ f_{i2}(t_2,s)\ \mathrm{Cos}(st_2)\ dt_2\right.$$
$$\left. + \int_{b_3=0}^{c_3}\phi_3(t_3)\ f_{i3}(t_3,s)\ \mathrm{Sin}(st_3)\ dt_3\right],\ i=1,..,6, \qquad (18)$$

$$f_{i1}(t_1,s) = \sum_{j=1}^{6}\frac{B_{ij}(s)\ t_1h_{i1}(t_1,s)}{\Delta(s)},\ i=1,..,6,$$

$$f_{i2}(t_2,s) = \frac{B_{i5}(s)\ h_{52}(s)}{\Delta(s)},\ i=1,..,6,$$

$$f_{i3}(t_3,s) = \frac{B_{i6}(s)\ h_{62}(s)}{\Delta(s)},\ i=1,..,6, \qquad (19.a\text{-}e)$$

where $\Delta(s)$ is the determinant and $B_{ij}(i=1,..,6;\ j=1,..,6)$ are the elements of the adjoint of the coefficient matrix A_{ij} of equation (15).

Substituting values for $f_i(s)$ from equation (18) in the boundary condition (8.c) and the interface conditions (5), the solution to the boundary value problem is given in terms of three simultaneous integral equations

$$\frac{1}{\pi}\int_{b_1}^{c_1}\frac{\phi_1\ (t_1)}{(t_1-r\ )}\ dt_1 + \frac{1}{\pi}\sum_{j=1}^{3}\int_{b_j}^{c_j}\phi_j(t_j)\ K_{1j}(r,t_j)\ dt_j = -\sigma_{zz}^{1M}(r,0)-\sigma_{zz}^{1T}(r,0),$$

$$b_1<r<c_1,$$

$$\frac{2}{\pi}\sum_{j=1}^{3}\int_{b_j}^{c_j}\phi_j(t_j)\,K_{2j}(z,t_j)\,dt_j = -\sigma_{rr}^{OM}(a,z)-\sigma_{rr}^{OT}(a,z),\quad b_2<z<c_2,$$

$$\frac{2}{\pi}\sum_{j=1}^{3}\int_{b_j}^{c_j}\phi_j(t_j)\,K_{3j}(z,t_j)\,dt_j + \langle z-c_2\rangle\,\frac{2}{\pi}\sum_{j=1}^{3}\int_{b_j}^{c_j}\phi_j(t_j)\,K_{2j}(z,t_j)\,dt_j$$

$$= \langle z-c_2\rangle\,\rho\,[\sigma_{rr}^{OM}(a,z)+\sigma_{rr}^{OT}(a,z)],\quad b_3<z<c_3, \qquad (20.a\text{-}c)$$

where

$$K_{11}(r,t_1) = K_{111}(r,t_1) + K_{112}(r,t_1), \qquad (21.a)$$

$$K_{111}(r,t_1) = \frac{m(r,t_1)-1}{t_1-r} + \frac{m(r,t_1)}{t_1+r}, \qquad (21.b)$$

$$m(r,t_1) = E\left(\frac{r}{t_1}\right),\quad r<t_1,$$

$$= \frac{r}{t_1}E\left(\frac{t_1}{r}\right) + \frac{t_1^2-r^2}{rt_1}K\left(\frac{t_1}{r}\right),\quad r>t_1, \qquad (21.c)$$

$$K_{112}(r,t_1) = \int_0^{\infty} 2\,k_{11}(r,t_1,s)\,ds, \qquad (21.d)$$

$$K_{12}(r,t_2) = \int_0^{\infty} 2\,k_{12}(r,t_2,s)\,\mathrm{Cos}(st_2)\,ds, \qquad (21.e)$$

$$K_{13}(r,t_3) = \int_0^{\infty} 2\,k_{13}(r,t_3,s)\,\mathrm{Sin}(st_3)\,ds, \qquad (21.f)$$

$$k_{1j}(r,t_j,s) = f_{1j}(t_j,s)\,I_0(rs) + f_{2j}(t_j,s)\,[2(2-\nu_1)I_0(rs)+rsI_1(rs)]$$
$$+f_{3j}(t_j,s)\,K_0(rs) + f_{4j}(t_j,s)\,[-2(2-\nu_1)K_0(rs)+rsK_1(rs)]),\quad j=1,2,3, \qquad (21.g)$$

$$K_{21}(z,t_1) = \int_0^{\infty} k_{21}(t_1,s)\,\mathrm{Cos}(sz)\,ds, \qquad (21.h)$$

$$K_{22}(z,t_2) = \int_0^{\infty} k_{22}(t_2,s)\,\mathrm{Cos}(st_2)\,\mathrm{Cos}(sz)\,ds, \qquad (21.i)$$

$$K_{23}(z,t_3) = \int_0^\infty k_{23}(t_3,s)\ \mathrm{Sin}(st_3)\ \mathrm{Cos}(sz)\ ds, \tag{21.j}$$

$$k_{2j}(t_j,s) = f_{5j}(t_j,s)\ [-I_0(as) + I_1(as)/(as)] + f_{6j}(t_j,s)\ [(2\nu_0-1)I_0(as) - asI_1(as)],\ j=1,2,3, \tag{21.k}$$

$$K_{31}(z,t_1) = \int_0^\infty k_{31}(t_1,s)\ \mathrm{Sin}(sz)\ ds. \tag{21.l}$$

$$K_{32}(z,t_2) = \int_0^\infty k_{32}(t_2,s)\ \mathrm{Cos}(st_2)\ \mathrm{Sin}(sz)\ ds, \tag{21.m}$$

$$K_{33}(z,t_3) = \int_0^\infty k_{33}(t_3,s)\ \mathrm{Sin}(st_3)\ \mathrm{Sin}(sz)\ ds, \tag{21.n}$$

$$k_{3j}(t_j,s) = f_{5j}(t_j,s)\ I_1(as) + f_{6j}(t_j,s)[asI_0(as) + 2(1-\nu_0)I_1(as)],\ j=1,2,3, \tag{21.o}$$

$$\langle z-c_2\rangle = 0 \text{ if } z<c_2$$

$$= 1 \text{ if } z>c_2. \tag{21.p}$$

where E() and K() are complete elliptic integrals of the first and second kind, respectively. Note that the functions k_{22}, k_{23}, k_{32}, k_{33} are only functions of 's' and are written as functions of 's' and 't_j' for purposes of condensed notations in equations (21.g), (21.k) and (21.o).

The integral equations (20.a-c) are also subject to the crack closure condition

$$\int_{b_1}^{c_1} \phi_1(t_1)\ dt_1 = 0, \tag{22}$$

for the case of cracks which close at the tips. These cases include internal cracks and cracks up to the interface for a perfect bond. For other cases, such as cracks up to the imperfect bond or edge cracks, the crack closure condition is not applicable.

ACKNOWLEDGEMENT

This work was supported by the 1992 AFOSR Summer Faculty Program at Wright Patterson Air Force Base, (WPAFB), Ohio through Research and Development Laboratory, California.

REFERENCES

Aveston, J., Cooper, G.A., and Kelly, A., 1971, "The Properties of Fiber Composites", Proceedings of National Physical Laboratory, IPC Science and Technology Press, Guilford, pp. 15-26.

Ford, E.F., 1973, Stress Analysis of a Broken Fiber Imbedded in an Elastic Medium, NSF Grant No. 33576 Grant Report, Harvard University.

Kaw, A.K. and Pagano, N.J., 1992, "Axisymmetric Thermoelastic Response of a Composite Cylinder Containing an Annular Matrix Crack", to be published.

Kim, R. and Pagano, N.J, 1991, "Crack Initiation in Unidirectional Brittle Matrix Composites", Journal of the American Ceramic Society, Vol. 74, 1082-1090.

McCartney, L.N., 1987, "Mechanics of Matrix Cracking in Brittle Matrix Fiber Reinforced Composites," Proceedings of the Royal Society of London, Series A, Vol. A409, pp. 329-350.

McCartney, L.N., 1990, "New Theoretical Model of Stress Transfer Between Fiber and Matrix in Uniaxially Fiber Reinforced Composites", Proceedings of the Royal Society of London, Series A, Vol. A425, pp. 215-244.

Pagano, N.J., and Brown III, H.W., "The Full Cell Cracking Mode in Unidirectional Brittle Matrix Composites", Composites, in press.

Steif, P.S., 1984, "Stiffness Reduction Due to Fiber Breakage", Journal of Composite Materials, Vol. 18, pp. 153-172.

Schweitert, H.R., and Steif, P.S., 1991, "Analysis of a Broken Fiber in a Weakly Bonded Composite", International Journal of Solids and Structures, Vol. 28, pp. 283-297.

Wijeyewickrema, A.C., Keer, L.M., Hirashma, K., and Mura, T., 1991, "The Annular Crack Surrounding an Elastic Fiber in a Tension Field", International Journal of Solids and Structures, Vol. 27, pp. 315-328.

Wijeyewickrema, A.C., and Keer, L.M., 1992a, "Matrix Crack Interaction in Fiber Reinforced Brittle Matrix Composites", International Journal of Solids and Structures, Vol. 29, pp. 559-570.

Wijeyewickrema, A.C., and Keer, L.M., 1992b, "Matrix Cracking in a Fiber Reinforced Composite with Slip at the Fiber-Matrix Interface", to be published.

ON THE FAILURE MECHANISMS IN TITANIUM ALUMINIDE COMPOSITES

Demitris A. Kouris
Associate Professor
Department of Mechanical and Aerospace Engineering

Arizona State University
Tempe, AZ 85287-6106

Final Report for:
Summer Research Program
Wright Laboratory

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

September 1992

ON THE FAILURE MECHANISMS IN TITANIUM ALUMINIDE COMPOSITES

Demitris A. Kouris
Associate Professor
Department of Mechanical and Aerospace Engineering
Arizona State University

I. ABSTRACT

Certain mechanisms of failure associated with titanium aluminide composites were studied. The focus of the analysis was directed towards the thermomechanical behavior of unidirectional composites under the influence of transverse loading. Recent experimental investigations indicated that there exist considerable differences in the mechanical behavior of composites with coated fibers, as compared to the uncoated systems. In addition, transverse mechanical loading yielded longitudinal matrix cracking in the regions of closely spaced fibers. A simple analytical model and finite element calculations were utilized in order to qualify the experimental results. Some preliminary thoughts on the issue of fiber fracture were also discussed, an issue that is currently under investigation.

II. FIBER/MATRIX INTERFACES

1.0 High Temperature Materials

Because of their low density and high strength at temperatures up to ~800°C, titanium aluminide alloys are the most likely materials to advance the high temperature performance of aerospace structures. For highly stressed components such as compressor disks in turbine engines, reinforcement of these alloys with high strength continuous fibers is necessary. The most suitable material appears to be SiC, although fiber coatings are necessary in order to obtain optimum interfacial properties and to control interfacial reactions under severe service and fabrication conditions.

Most previous studies of titanium aluminide composites have involved either β-stabilized titanium alloy matrices or matrices based on the α_2 Ti_3Al phase. High stiffness, strength and fatigue resistance have been achieved in both systems (Johnson et al. 1990, Revelos and Smith 1992). However, they are both limited in application; the β-stabilized matrices by high temperature stability and the α_2 - based matrices by severe degradation during thermal fatigue in air (Revelos and Smith 1992)

Alloys based on the orthorhombic Ti_2AlNb phase have been shown by Banerjee et al. (1988) and Rowe (1991) to possess higher strength, fracture toughness, and ductility at low temperature, and better environmental stability compared to other titanium aluminide alloys such as Ti_3Al based materials. Moreover, Smith et al. (1992) have shown that composites of an orthorhombic based alloy reinforced with SCS6 SiC fibers exhibit improved properties and are resistant to environmental degradation. Several Air Force programs are under way to develop SiC reinforced composites with this matrix material.

2.0 Interfaces

The properties of the interface between the fibers and matrix dictate the overall properties of high temperature ceramic and intermetallic matrix composites. Generally, weak interfaces that allow debonding are required in order to achieve optimum longitudinal properties in unidirectionally reinforced composites. In ceramic matrix composites, matrix cracking is the first damage to develop. If the fiber/matrix interface is

sufficiently weak to deflect an incident crack into the interface rather than allow it to penetrate the fibers, and if the fibers are sufficiently strong relative to the sliding resistance of the interface then the strength of the composite is dictated by the fiber strength and is insensitive to damage and notches. Moreover, the stress strain curve is non-linear prior to the peak stress, and non catastrophic beyond the peak. On the other hand, if the interface is strong enough for the first matrix crack to penetrate the fiber, the composite behaves as a monolithic brittle material with linear stress strain curve to failure, and with strength that is sensitive to pre existing damage and notches. The condition for crack deflection has been derived in terms of the fracture energies of the interface (G_i) and the fibers (G_f) debonding occurs under most conditions if $G_i/G_f \leq 0.25$.

In unidirectionally reinforced intermetallic matrix composites fiber fracture can occur before failure of the matrix under monotonic loading. Interfacial debonding then leads to improved strength by allowing the damage in the matrix to be spread out rather than being concentrated near the fiber cracks. However the largest potential benefit from reinforcement of intermetallic matrix composites is in the improved resistance to fatigue cracking. Under cyclic loading, fatigue cracks initiate in the matrix. If interfacial debonding occurs, the reinforcing fibers remain spanning the crack faces and restrict the range of displacements that are transmitted to the crack tip, thereby reducing the crack growth rate. With optimum interfacial properties this reduction in growth rate due to crack bridging can be dramatic, either arresting a crack that would otherwise accelerate indefinitely, or reducing its growth rate to a low constant velocity. The crack bridging effect responsible for this improvement in properties is discussed further in the next section.

The need to produce weak interfaces for optimum longitudinal properties conflicts with the requirements for high transverse strength where strong interfaces are beneficial. Therefore a compromise is needed, either in the interface properties or in the design requirements. This topic has received very little attention. For instance, very little is known of the effect of interfacial friction on transverse strength.

Most of the preceding comments apply at high temperatures as well as low temperature, although an added complication of rate dependence enters the problem. At intermediate temperatures fatigue crack growth occurs in the matrix, but the growth law becomes time dependent as well as cycle dependent as localized creep occurs near the crack tip. In addition, the interfacial properties that dictate crack bridging become time-

dependent. At high temperatures, where creep occurs throughout the matrix, fatigue cracking is suppressed. In this temperature range the requirements for interfacial properties are not as clear. For short reinforcing fibers , or for transverse loading of continuous fiber composites, strongly bonded interfaces are essential to obtain load transfer. However for longitudinally loaded continuous fiber composites, load transfer occurs over small distances, even when interfaces are weakly bonded.

III. DAMAGE MECHANISMS DUE TO TRANSVERSE LOADING

1.0 Experimental Observations

Metal matrix composites consisting of titanium aluminide matrices reinforced by unidirectional SiC fibers exhibit superior mechanical properties under the influence of loading parallel to the fibers. However, these improved properties are achieved at the expense of properties in the transverse direction. Given the fact that in structural components we usually cannot avoid the presence of transverse loads, it becomes essential to investigate the factors controlling transverse properties. The issue of transverse loading is been investigated during our current, AFOSR supported project. This program involves analytical and numerical studies at Arizona State University as well as experimental investigations performed at the Rockwell International Science Center.

Recent experiments at Rockwell involved two composites with identical fibers and matrices (super α_2/SCS6), but different interfacial compositions. Both systems were consolidated by Textron, using a foil/fiber/foil method. One composite contained three rows of standard SCS6 fibers, while the other contained four rows of SCS6 fibers that had been coated (PVD) with Ag and Ta. The fiber volume fractions were 37% and 30%, respectively. Beams (50 x 2 x 1 mm) that were cut from the composite sheets were tested in unidirectional tension, perpendicular to the direction of the fibers. Measurements were obtained from optical micrographs by utilizing a computerized image analysis system.

Observations of the microstructure indicated that the composite with uncoated fibers contained reaction layers (~1 μm thick) around the individual fibers. These layers were not found in the composite with the Ta/Ag fiber coating. However, the matrix of the coated system was enriched in the β-phase within a layer of ~10μm around each fiber.

The stress-strain curves obtained indicate that the stresses and strains at failure are much smaller in transverse loading than in longitudinal loading. Nevertheless, the transverse failure stress and strain are considerably higher in the composite with the Ta/Ag fiber coating, compared to the uncoated system. The values of Young's modulus are (a) 200 GPa under longitudinal loading; (b) 150 GPa under transverse loading in the composite with uncoated fibers, and (c) 125 GPa under transverse loading in the composite with coated fibers.

Under the influence of transverse loading, the first damage to occur was sliding of the fibers relative to the matrix, driven by the residual stresses (compressive in the fibers, tensile in the matrix). The sliding displacement increased with increasing applied stress. Sliding began at applied stress of 40 MPa in the composite with uncoated fibers and 60 MPa in the composite with coated fibers.

Tensile separation of the fiber/matrix interface was observed in the stress range of 150-180 MPa in the coated system and at 100-150 MPa in the composite with uncoated fibers. The locations of interfacial separation were different in the two composites. In the composite containing uncoated fibers, about 70% of the fibers separated between the outer layer of the fiber and the reaction layer of the matrix. In the coated system, however, 80% of the fibers separated at the carbon layer between the outer layers of the fibers. The remaining fibers in both systems exhibited debonding at both interfaces. There was no correlation between the separation load and the distance between fibers.

The first observable matrix damage was deformation in the regions between closely spaced fibers located along a line parallel to the applied load. This deformation was first observed at an applied stress of 100 MPa in the composite with uncoated fibers and 128 MPa in the composite with the Ta/Ag coating. At higher loads (150 MPa and 180 MPa, respectively) longitudinal cracks (parallel to the loading) were observed, initiating adjacent to the fibers and growing stably between the fibers as the load increased. At closely spaced fibers (separation $\leq$ 20 μm), a pair of longitudinal cracks formed symmetrically with respect to the center line of the fibers. These cracks formed at about 30% of the fibers in both composites, mostly at fibers that were closely spaced ($\leq$ 70 μm).

As the applied load increased, transverse cracks were observed in the composite with uncoated fibers. These cracks initiated in the reaction zone of the matrix adjacent to the fibers and propagated stably across the regions between pairs of fibers. The linking of these transverse cracks and the previously debonded interfaces resulted in the failure of the composite. However, in the composite with the Ta/Ag coating, transverse cracks were

not observed prior to failure, even though the applied stress at failure was more than double compared to the one corresponding to the uncoated system.

The above description of the experimental observations has been originally included in a manuscript by D. Marshall, W. Morris and B. Cox (Rockwell Science Center) and is a part of our second-year report to AFOSR. These observations established the required framework for our modeling effort.

2.0 Modeling

The purpose of the theoretical and numerical analysis was to study the essential factors associated with the deformation mechanisms discussed above and model the behavior of the composite under the effect of thermo-mechanical loading. Using the fiber relaxation measurements obtained after removing the matrix by etching, the residual stresses were calculated using a concentric cylinder model. The results are listed below in Table 1. The presence of the Ta/Ag coatings did not influence the values of the residual stresses, as shown by the misfit strains being equal in the two composites. The differences in residual stresses (Table 1) arise solely from the different volume fractions of fibers in the two composites.

Table 1. Residual Stresses and Frictional Sliding Resistance

Composite	Measured Relaxation e_f	Misfit Strain e^T	Normal Interface Stress σ_r^R (MPa)	Tangential Interface Stress σ_t^R (MPa)	Axial Fiber Stress σ_f^R (MPa)	Interface Sliding Resistance τ_o (MPa)
Super-α_2/ SCS6	0.00156	0.00203	-260	565	-800	70
Super-α_2/ Ag-Ta/ SCS6	0.00203	0.00204	-300	557	-1020	100

The frictional stresses listed above were determined from single fiber pulling experiments (Marshall et al., 1992).

In addition to the concentric cylinder model, the residual stresses were evaluated using finite elements and an analytical solution involving a pair of closely spaced fibers. The finite element analysis involved square and rectangular arrays of fibers and yielded hoop versus radial stress ratios between 2.0 and 2.18. These values are consistent with results of Table 1 and were based on the independently measured moduli of the matrix E_m (E_m = 80 GPa) and the fibers E_f (E_f = 414 GPa).

The aim of the analytical approach was to determine the local stresses and displacements in the vicinity of two fibers under the influence of the residual field. The solution was obtained in closed form and the radial interfacial stress was found to be -255 MPa (versus -260 MPa that was obtained via the concentric cylinder model).

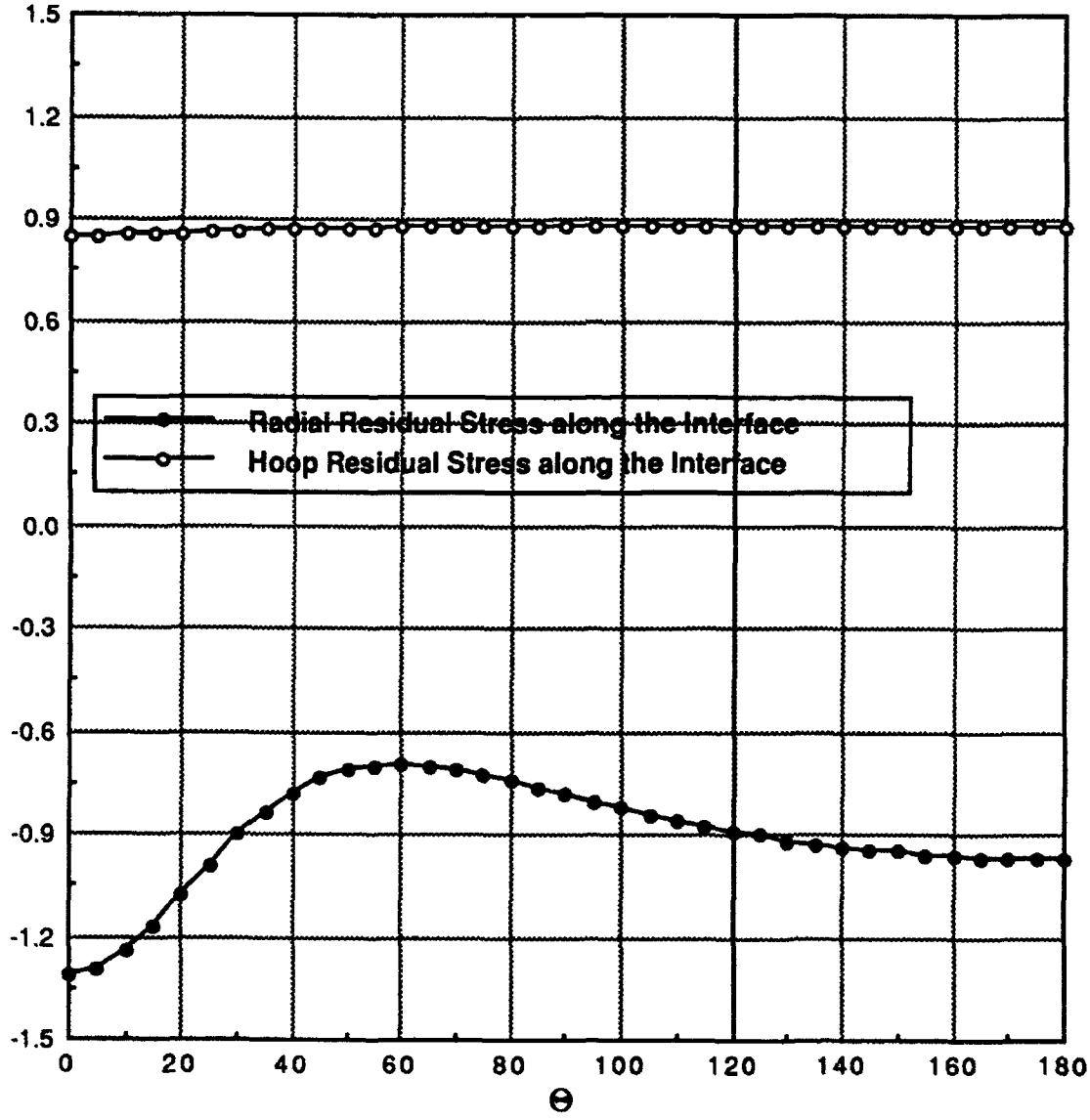


Figure 1. Normalized residual stresses along the fiber interface.

The experimental observations (Marshall et al., 1992) indicated that interfacial separation occurred at applied stress levels of 100-150 MPa in the uncoated system and at 150-180 MPa in the composite with Ta/Ag coatings. These stress levels are the ones at which the stress-strain curves become nonlinear; this suggests that interfacial separation is responsible for the initial departure from linearity. It is also clear that the strength of the interface is negligibly small. Based on these observations, one would anticipate the initial separation to occur at stress levels (due to uniaxial tension) that are equal (and opposite) to the residual field.

A simple plane strain analysis that considers a single fiber under uniaxial tension yields the following radial stress at the interface:

$$\sigma_r = 1.3\, T_x \qquad (1)$$

where T_x is the applied uniaxial tension. For separation to occur, the stress given by equation (1) should equilibrate the corresponding residual stress given in Table 1. Therefore:

$$1.3\, T_x - 260 \text{ MPa} = 0 \Rightarrow T_x = 200 \text{ MPa} \qquad (2)$$

It is clear that this simplistic approach that does not consider the interaction between fibers overestimates the value of the applied stress for separation. Similar erroneous results are obtained if one approximates the fiber interaction by superimposing the single fiber solution.

An exact solution was sought for the fiber interaction problem shown in Fig. ?. Here "a" denotes the radius of the fiber cross section and "d" corresponds to the distance between the interface and the middle point of the line which connects the two fiber centers. The mathematical formulation of the problem was based on the approach followed by Kouris and Tsuchida (1991).

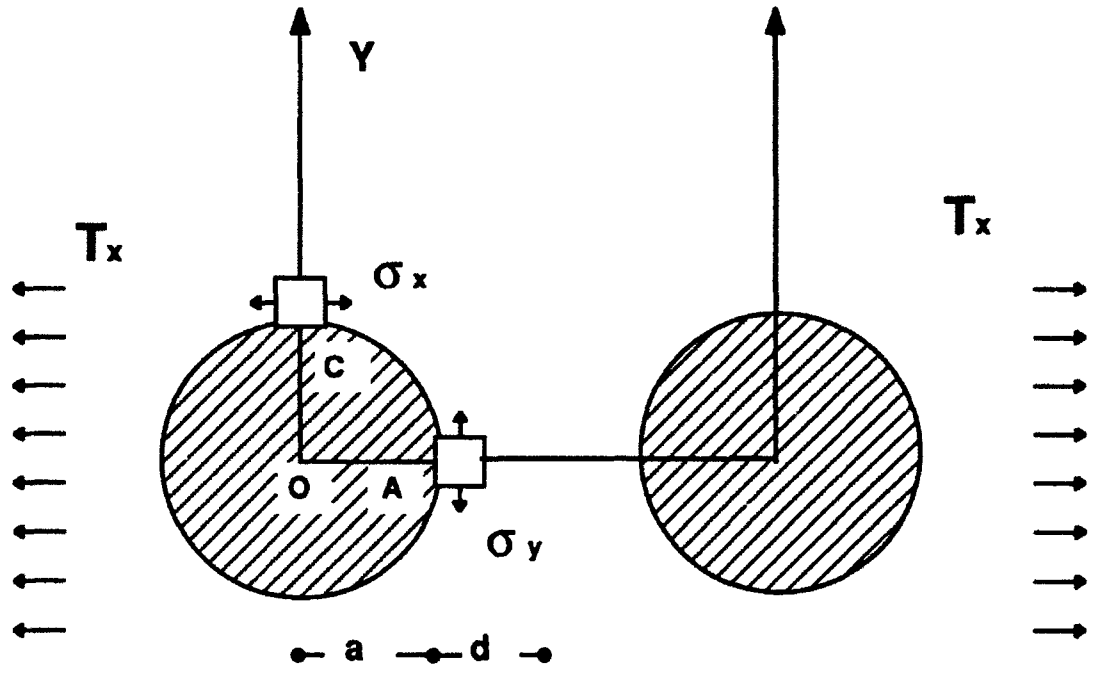


Figure 2. Interaction between two adjacent fibers

The problem was investigated by utilizing the Papcovich-Neuber displacement potentials. The displacement field $\underline{u}$ consists of the addition of the local displacements $\underline{u_1}$ and $\underline{u_2}$ that correspond to the two centers of the fiber cross sections. Consequently, the displacement fields $\underline{u_1}$ and $\underline{u_2}$ are expressed by:

$$2G\underline{u_i} = \text{grad}\,[\,\phi_{0i} + x_i\,\phi_{1i} + y_i\,\phi_{2i}\,] - 4\,(1-\nu)\,[\,\phi_{1i}\;,\;\phi_{2i}\,]\,,\;(i=1,2) \qquad (3)$$

where ϕ_{0i} , ϕ_{1i} and ϕ_{2i} are arbitrary harmonic functions. The potentials chosen for the matrix are:

$$\begin{cases} \phi_{0i} = p_0\,[\,F_0^i \log\rho_i + \sum_{n=1}^{\infty} A_n^i\;\rho_i^{-n}\cos n\theta_i] \\ \phi_{1i} = p_0 \sum_{n=1}^{\infty} B_n^i\;\rho_i^{-n}\cos n\theta_i \\ \phi_{2i} = 0 \end{cases} \qquad (4)$$

and for the fiber cross sections:

$$\begin{cases} \phi_{0i} = p_0 \sum_{n=1}^{\infty} C_n^i \, \rho_i^{\,n} \cos n\theta_i \\ \phi_{1i} = p_0 \sum_{n=1}^{\infty} D_n^i \, \rho_i^{\,n} \cos n\theta_i \\ \phi_{2i} = 0 \end{cases} \tag{5}$$

where (ρ, θ) indicate polar coordinates. Using the potential functions described above, the unknown series coefficients can be obtained by enforcing the boundary conditions along the fiber/matrix interfaces. Consequently, stresses and displacements can be determined.

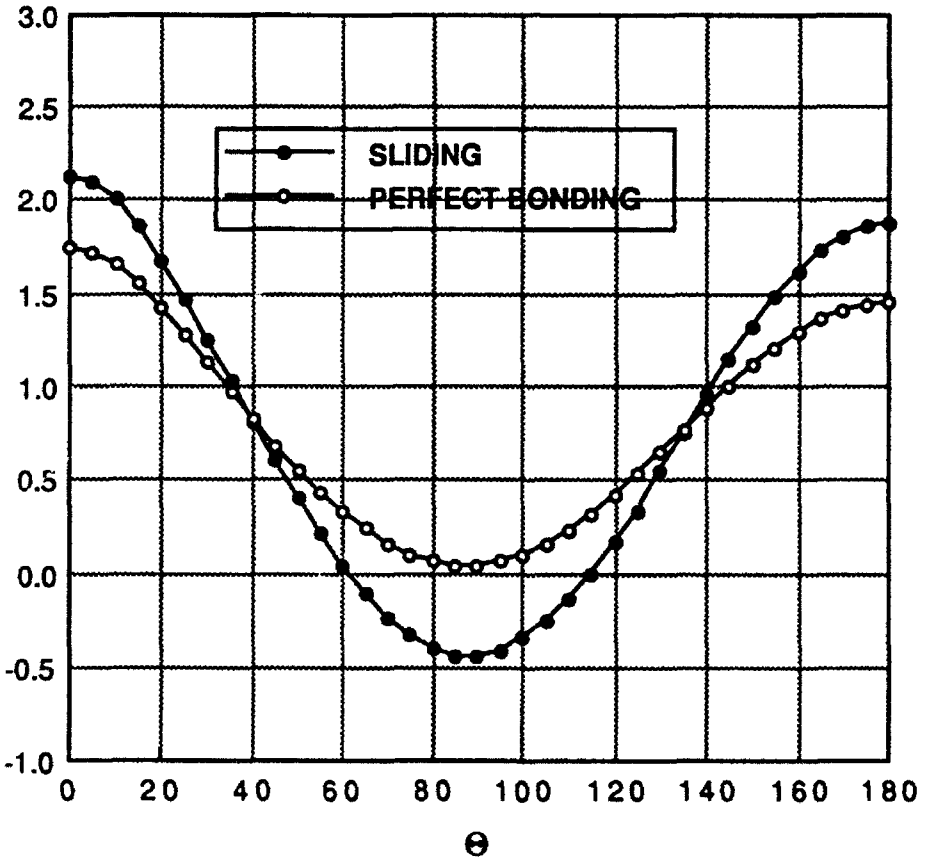


Figure 3. Radial stress σ_r / T_x along the fiber interface.

For a relative distance ratio d/a = 0.43 (which corresponds to ~60 μm) the solution indicates that:

$$\sigma_r = 2.13\, T_x \qquad (6).$$

Therefore, the uniaxial tension that corresponds to separation can be determined:

$$2.13\, T_x - 260\ \text{MPa} = 0 \Rightarrow T_x = 122\ \text{MPa} \qquad (7).$$

These value of T_x is consistent with the experimental observations discussed earlier.

The finite element calculations (square and rectangular arrays) indicated that the value of the radial stress varies between 1.35 T_x and 2.0 T_x, depending on the relative fiber distance. Therefore, the value of T_x was in the range of 130 MPa to 190 MPa.

The next phase of the deformation process was associated with the presence of longitudinal cracks in the matrix, between closely spaced fibers. At the early stages of the investigation, this phenomenon was viewed with considerable skepticism. This was due to the fact that it is difficult to intuitively understand how the first cracks can appear in the direction of the applied load. To complicate the issue, initial calculations based on approximate solutions did not support this experimental fact. However, the problem was clarified by utilizing the exact solution mentioned above.

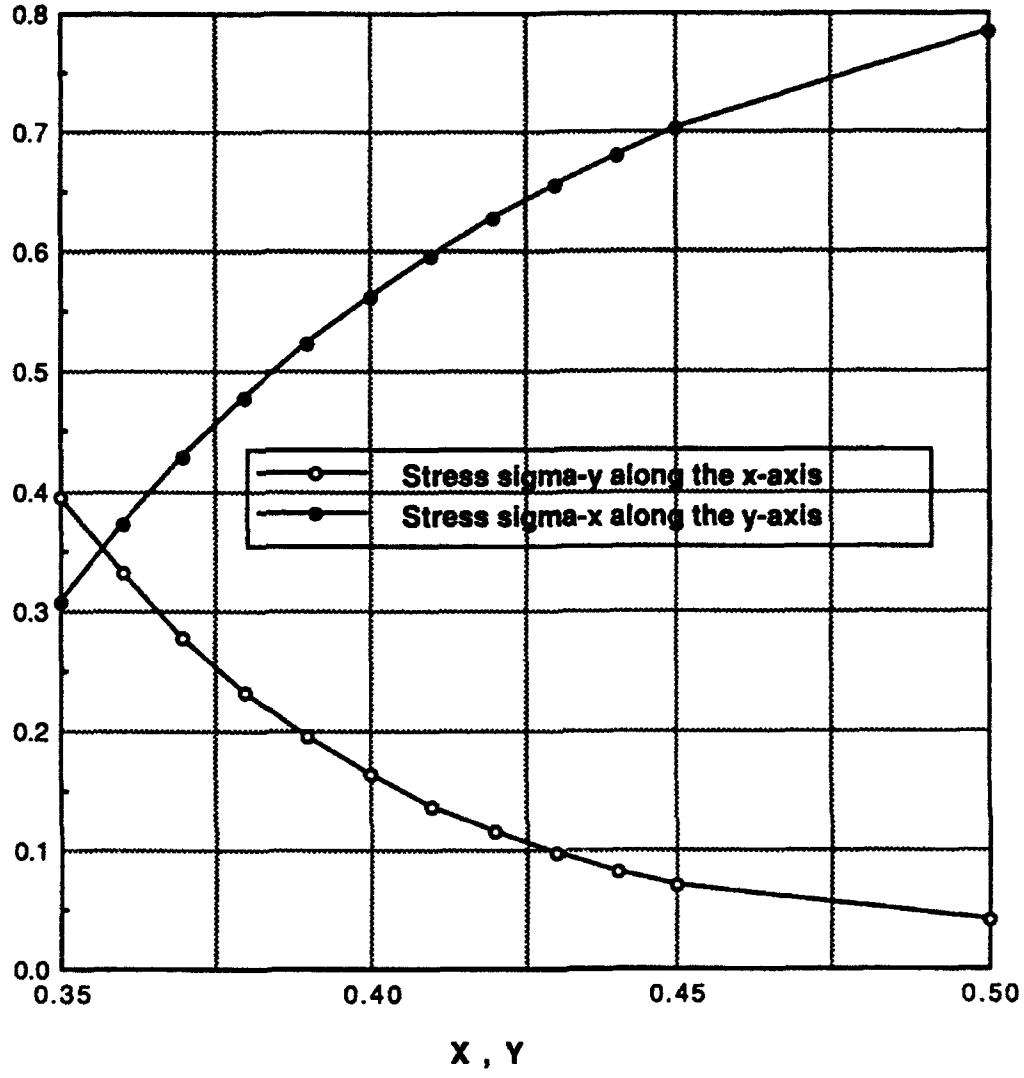


Figure 4. Variation of the stresses corresponding to longitudinal and transverse cracking.

The two-fiber interaction problem allows for a discontinuity of the normal displacement and yields a hoop stress at point A (Fig. 2) which is about 1.7 times higher than the hoop stress at point C. These stresses correspond to the sites of longitudinal and transverse cracking, respectively. It is therefore concluded that longitudinal cracking should occur before the initiation of the catastrophic transverse cracks. This determination was confirmed using finite element analysis; in this case, the above mentioned ratio of the hoop stresses at A and C was about 2.0. When the relative distance between fibers is such

that $0.05 < d/a < 0.67$, the value of the hoop stress at A varies between $0.4\ T_x$ and $0.6\ T_x$. Our experimental observations indicated that longitudinal cracking was present for $T_x > 150$ MPa. Given the residual hoop stress of 560 MPa, the local stress level corresponding to longitudinal cracking was 620~650 MPa.

In order to account for the initiation of transverse cracking, we investigated the variation of the hoop stress σ_x along the y-axis (Fig. 4) as a function of fiber distance. It was found that σ_x generally attains its maximum value at the middle of the distance between rows of fibers. However, if the fibers are very closely packed ($d/a = 0.05$), σ_x along the y-axis is almost uniform. Another conclusion of the analysis was that as the distance between fiber rows increases, so does σ_x. For $d/a = 0.43$ it was found that σ_x is equal to $0.6\ T_x$. The value of the applied tension that corresponds to transverse cracking is about 200 MPa. Consequently, the local stresses in the matrix are within the 740 - 800 MPa range.

IV. SOME PRELIMINARY THOUGHTS ON THE ISSUE OF FIBER FRACTURE

A recent experimental study by Mall et al. (1992) conducted at the Materials Laboratory (Wright Patterson Air Force Base) investigated the thermomechanical fatigue behavior of a β 21-S / SCS-6 composite. The investigation involved three sets of experiments : (a) in-phase thermomechanical fatigue (TMF) from 150°C to 650°C, (b) out-of-phase TMF from 150°C to 650°C, and (c) isothermal fatigue at 650°C. By monitoring the evolution of strains and moduli during the tests, it was concluded that fatigue lives corresponding to (a), (b) and (c) above are controlled by different mechanisms. In cases (a) and (c) fatigue lives were primarily dependent upon the maximum fiber stress, while in case (b) fatigue life was determined by the matrix stress range.

One of the very interesting findings of this effort is related to the evolution of the total mechanical strain as a function of the number of cycles. During the loading cycles close to the end of life, it was apparent that the composites exhibited a considerable increase in the amount of accumulated strain. This increase was accompanied by a very modest decrease in the stiffness (modulus) of the composite. Initially, the phenomenon was attributed to extensive matrix creep (primarily) and subsequent fiber fracture. The issue of fiber fracture was considered as a secondary one for, mainly, two reasons. First,

it appeared (according to the initial interpretation) that the fiber breakage process was occurring during a small fraction of the total life. In addition, one usually associates fiber breakage with a considerable loss of stiffness. However, our preliminary investigation suggests otherwise.

The question of fiber fracture in relation to the subsequent strain increase and loss of stiffness was considered via a simple cylinder model, shown in Fig. 5.

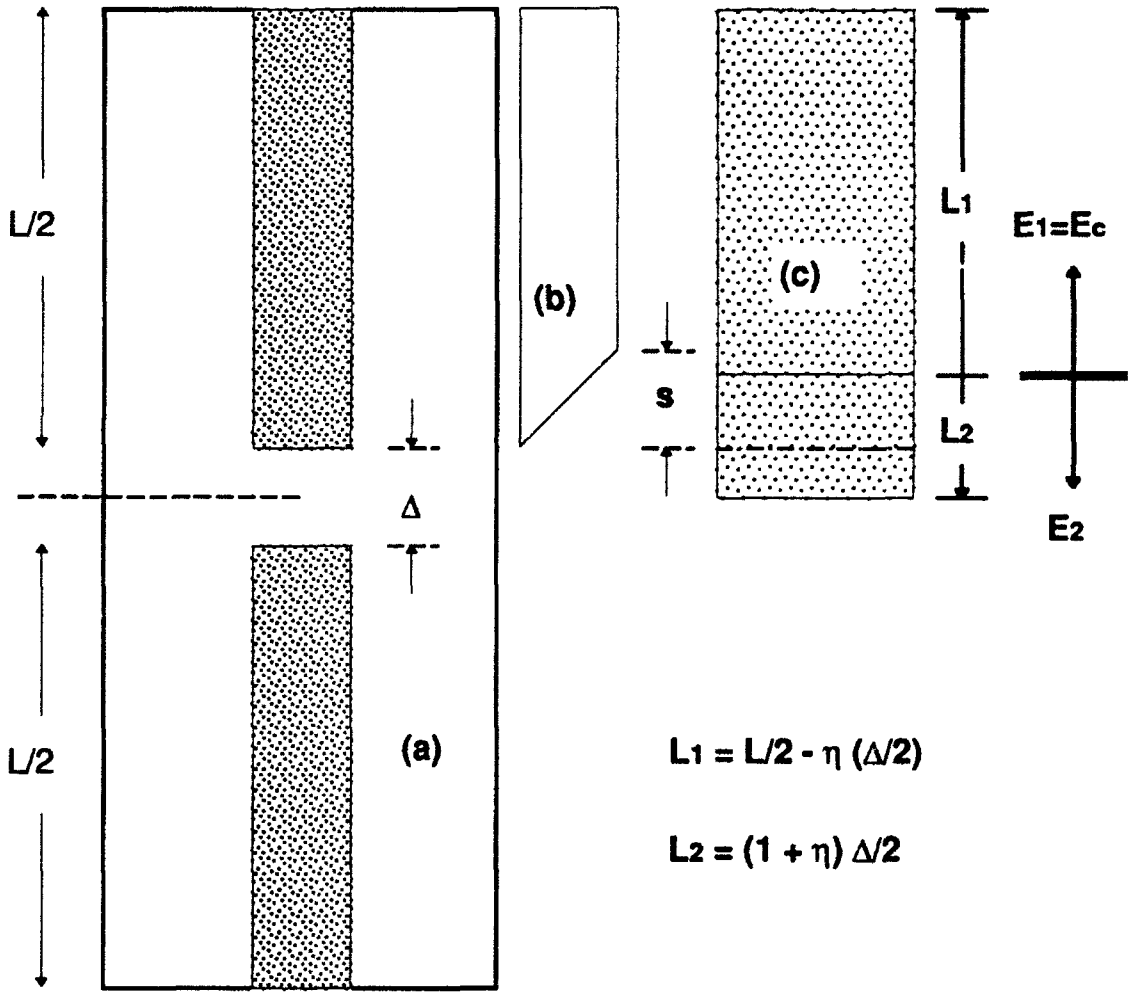


Figure 5: A simple cylinder model for the analysis of fiber fracture.

Fig. 5(a), (b) and (c) indicate the geometry of the system, the distribution of fiber stress, and the regions of effective modulus, respectively. Here E_c denotes the modulus of the composite (unbroken fibers) and E_2 corresponds to the modulus of the matrix:

$$E_c = E_f V_f + E_m (1 - V_f) \text{ and } E_2 = E_m (1 - V_f) \tag{8}$$

where V_f indicates the fiber volume fraction and the subscripts f and m denote fiber and matrix quantities, respectively.

The effective lengths L_1 and L_2 are given in terms of the fracture gap Δ and a numerical factor η ($1 \leq \eta \leq 2$):

$$L_1 = \frac{L}{2} - h \frac{\Delta}{2} \quad \text{and} \quad L_2 = (1 + h) \frac{\Delta}{2} \tag{9}$$

If we denote the strain by ε_o and the initial elongation (corresponding to the unbroken fiber) by δ_o, the model yields the new elongation δ (after the fiber fracture):

$$\frac{\delta}{\delta_o} = 1 - \frac{s}{L} + \frac{1}{\varepsilon_o} \frac{\Delta}{L} \tag{10}$$

In addition, the modulus E of the composite after fiber fracture can be determined in terms of D/L and the properties of the constituents:

$$E = E_c \frac{1 + (L_2 / L_1)}{1 + (L_2 / L_1) (E_c / E_2)} \tag{11}$$

Expressions (10) and (11) require the two ratios s / L and Δ / L. The first is associated with the fiber stress distribution (Fig. 5(b) in the vicinity of the crack and the second with the gap between the two parts of the fiber.

In order to realistically estimate the values of the above mentioned ratios, the analytical approach of McCartney (1992) was followed, which can predict the stress transfer between the matrix and the fibers, in the presence of a fiber crack. For a given macroscopic strain ε_o, stresses and displacements along the fiber/matrix interface were evaluated.

The axial displacements of the fiber and the matrix are shown in Fig. 6, for a macroscopic strain $\varepsilon_o = 0.2\%$. It can be observed that at a distance from the crack which

is between four and five fiber radii, the displacements assume the value of the ideal composite (no broken fibers).

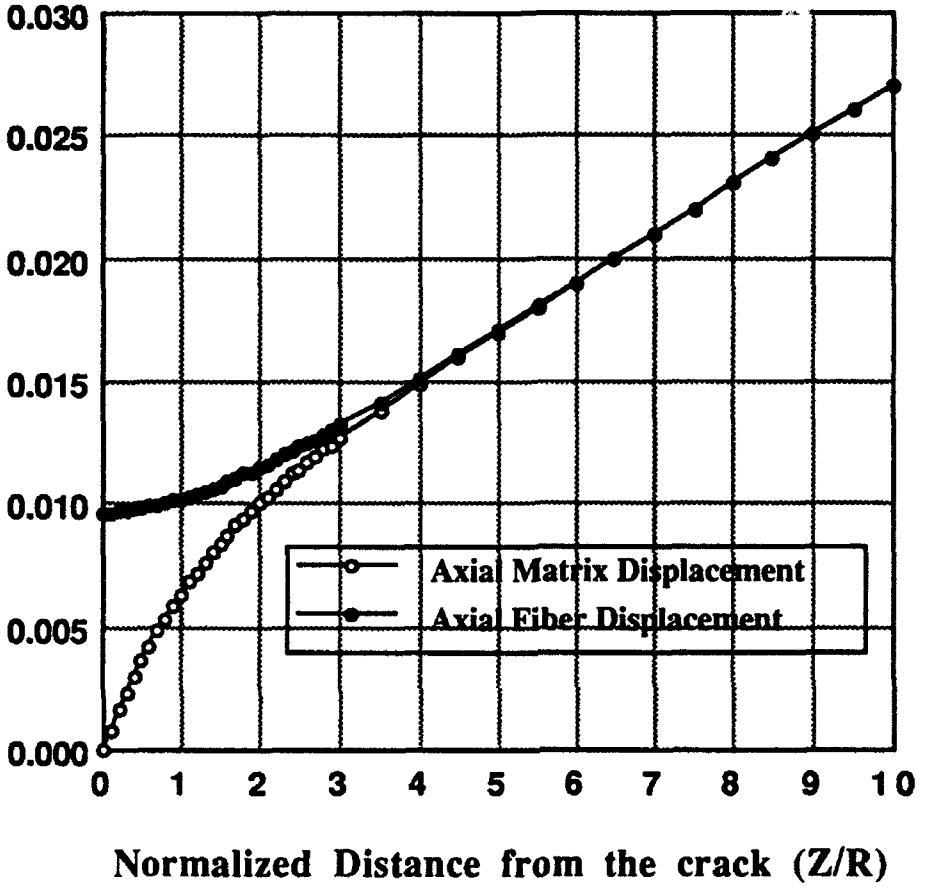


Figure 6. Fiber and matrix displacements vs. distance from the crack.

The same observation can be made for the distribution of the axial stresses shown in Fig. 7.

It was, therefore, concluded that the values of s/L and D/L can be assumed to be:

$$\frac{s}{L} \approx 10\% \qquad \text{and} \qquad 1‰ \leq \frac{D}{L} \leq 1\%. \qquad (12)$$

Based on the estimates given in (12), the elongation ratio and the composite modulus, as described in (10) and (11), can be evaluated. It was found that fiber fracture

can accommodate considerable strain increases (~45%) with a minimal loss of stiffness (~5%).

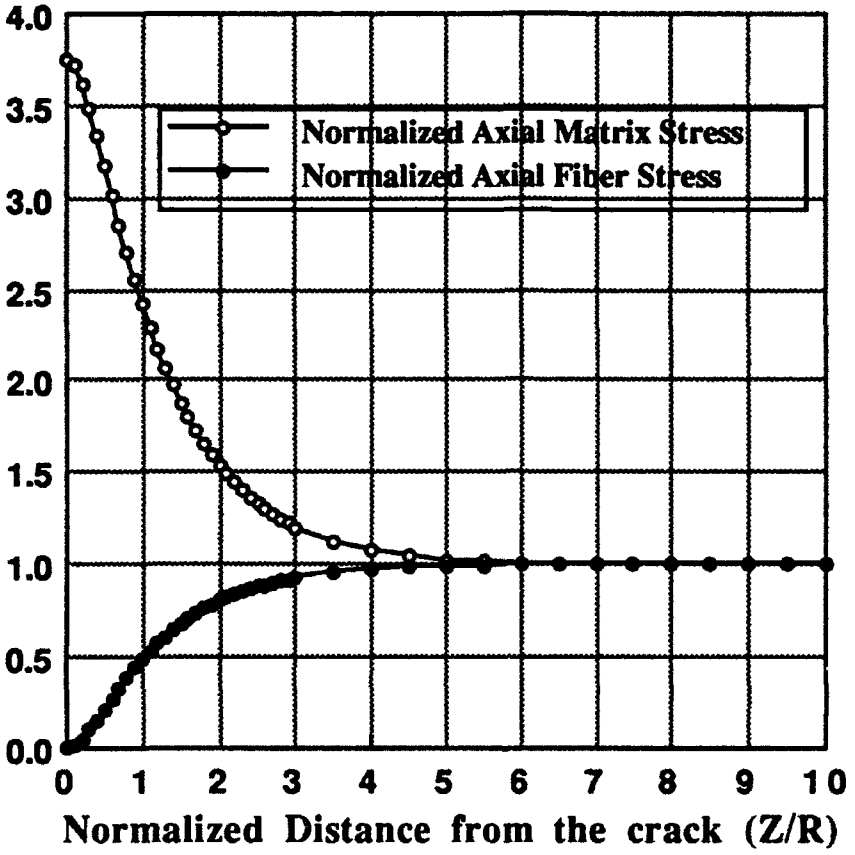


Figure 7. Axial stresses vs. distance from the crack.

These results are consistent with preliminary experimental observations which involve detection of fiber cracks utilizing acoustic emission techniques.

The issue of fiber fracture is one that has not received the appropriate attention, even though it is associated with a critical mechanism of fatigue failure. It is a subject that unravels quite a number of interesting micromechanics problems.

V. REFERENCES

BANERJEE, D., GOGIA, A. K., NANDI, T. K., and JOSHI, V. A., 1988, Acta Met., 36(4), 871.

JOHNSON, W. S., LUBOWINSKI, S. J., and HIGHSMITH, A. L., 1990, ASTM STP 1080, edited by J. M. Kennedy, H. Moeller, and W. S. Johnson.

KOURIS, D. A. and TSUCHIDA, E., 1991, Mechanics of Materials, 12, 131.

MALL, S., HANSON, D. G., NICHOLAS, T., and RUSS, S. M., 1992, in Symposium on Constitutive Behavior of High Temperature Composites, ASME WAM, to appear.

MARSHALL, D. E., SHAW, M. C., and MORRIS, W. L., 1992, Acta Met., 33, 2013.

McCARTNEY, L. N., 1989, Proc. R. Soc. Lond. A 425, 215.

REVELOS, W. C., and SMITH, P. R., 1992, Met. Trans. A 23A, 587.

ROWE, R. G., 1991, in Microstructure/Property Relationships in Titanium Aluminides and Alloys, edited by Y. W. Kim and R. R. Boyer, TMS, 387.

SMITH, P. R., GRAVES, J. A., RHODES, C. G., JAMES, M. R., and PORTER, J. R., 1992, Intermetellic Matrix Composites - II, Proc. Symp. Mat. Res. Soc. Spring Meeting (in press).

A NEW TECHNIQUE FOR MEASURING RAYLEIGH AND LAMB WAVE VELOCITIES IN METALS, GRAPHITE-EPOXY AND METAL MATRIX COMPOSITES

Tribikram Kundu
Associate Professor
Department of Civil Engineering and Engineering Mechanics

University of Arizona
Tucson, Arizona 85721

Final Report for:
Summer Research Program
Wright-Patterson Laboratory

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

July 1992

A NEW TECHNIQUE FOR MEASURING RAYLEIGH AND LAMB WAVE VELOCITIES IN METALS, GRAPHITE-EPOXY AND METAL MATRIX COMPOSITES

Tribikram Kundu
Associate Professor
Department of Civil Engineering and Engineering Mechanics
University of Arizona

Abstract

A new technique is developed to measure Rayleigh and Lamb wave velocities in metals and composites. In this technique two transducers are positioned above the specimen in a pitch-catch orientation. The time of flight of the signal from the transmitter to the receiver is recorded. Then the rate of change of this time as the distance between the reflector and the transducer varies is experimentally determined. This rate remains constant when leaky Rayleigh or Lamb waves are generated, otherwise it varies. Thus surface waves can be detected in an indirect manner. An expression is derived to relate the surface wave velocity to the signal flight time change rate with the transducer specimen distance. Using this expression Rayleigh and Lamb wave velocities have been accurately determined in isotropic metals and anisotropic composites.

A NEW TECHNIQUE FOR MEASURING RAYLEIGH AND LAMB WAVE VELOCITIES IN METALS, GRAPHITE-EPOXY AND METAL MATRIX COMPOSITES

Tribikram Kundu

INTRODUCTION

Surface waves in a material can reveal important information about the material. It has been shown by several investigators that the wave velocity as well as the attenuation property in a material is affected when the material is subjected to fatigue (micro cracks develop) or placed in hostile environments (corrosion occurs, hence material properties change) or the material is damaged due to any other cause such as excessive loading (micro and macro cracks develop) [Mal, Yin and Bar-Cohen (1991), Zhang and Achenbach (1991), Achenbach, Komsky and Zhang (1990), Kinra and Dayal (1987, 1988)]. For a layered material the surface coating thickness and its properties can be obtained by measuring the Rayleigh wave velocity through the layer [Rosen (1987)]. In addition, it has been used to study the anisotropy in a material [Kushibiki et.al.(1982a,1982b,1983)], surface hardening and residual stresses [Lu, Maxfield and Kuramoto (1990), Egle (1980)] among other important surface related properties. Thus an accurate surface wave velocity measurement technique is necessary for studying material behavior nondestructively. Today several techniques are available for measuring Rayleigh wave velocity in a material. These techniques require longitudinal and shear wave transducers or knife-edge buffer or acoustic microscope or optical interferometer or ultrasonic micro-spectrometer. A review of these techniques can be found in Kundu and Maxfield(1992). The proposed new technique only requires two ordinary immersion type ultrasonic transducers as described in the following section.

PROPOSED TECHNIQUE

In the proposed technique two transducers are positioned above the specimen in a pitch-catch orientation as shown in Fig.1a. In this orientation one transducer acts as the transmitter and the other one acts as the receiver. Ultrasonic signal can travel from the transmitter to the receiver by either path TCR or path TABR, see Fig.1a. Path TCR is followed when surface waves are not generated and TABR is followed when surface waves are generated.

Let us denote the angle of incidence by θ and the radius of the circular arc on which transducers are placed by R_c, and the distance between the center of this circular arc and the reflecting surface by z. Then from simple geometry one can see that the length TC is given by

$$TC=[z^2+R_c^2-2zR_c\cos\theta]^{1/2} \qquad (1)$$

So the time taken by the signal to travel from the transmitter to the receiver is given by

$$t=\frac{2\cdot TC}{\alpha_f}=\frac{2}{\alpha_f}[z^2+R_c^2-2zR_c\cos\theta]^{1/2} \qquad (2)$$

where α_f is the acoustic wave velocity in the coupling fluid. From equation (4) one can obtain

$$\frac{dt}{dz}=\frac{2(z-R_c\cos\theta)}{\alpha_f[z^2+R_c^2-2zR_c\cos\theta]^{1/2}} \qquad (3)$$

Similarly one can obtain the path length TABR from simple geometry

$$TA+AB+BR=2[R_c-\frac{z}{\cos\theta}+z\tan\theta] \qquad (4)$$

So the time of travel along this path is given by

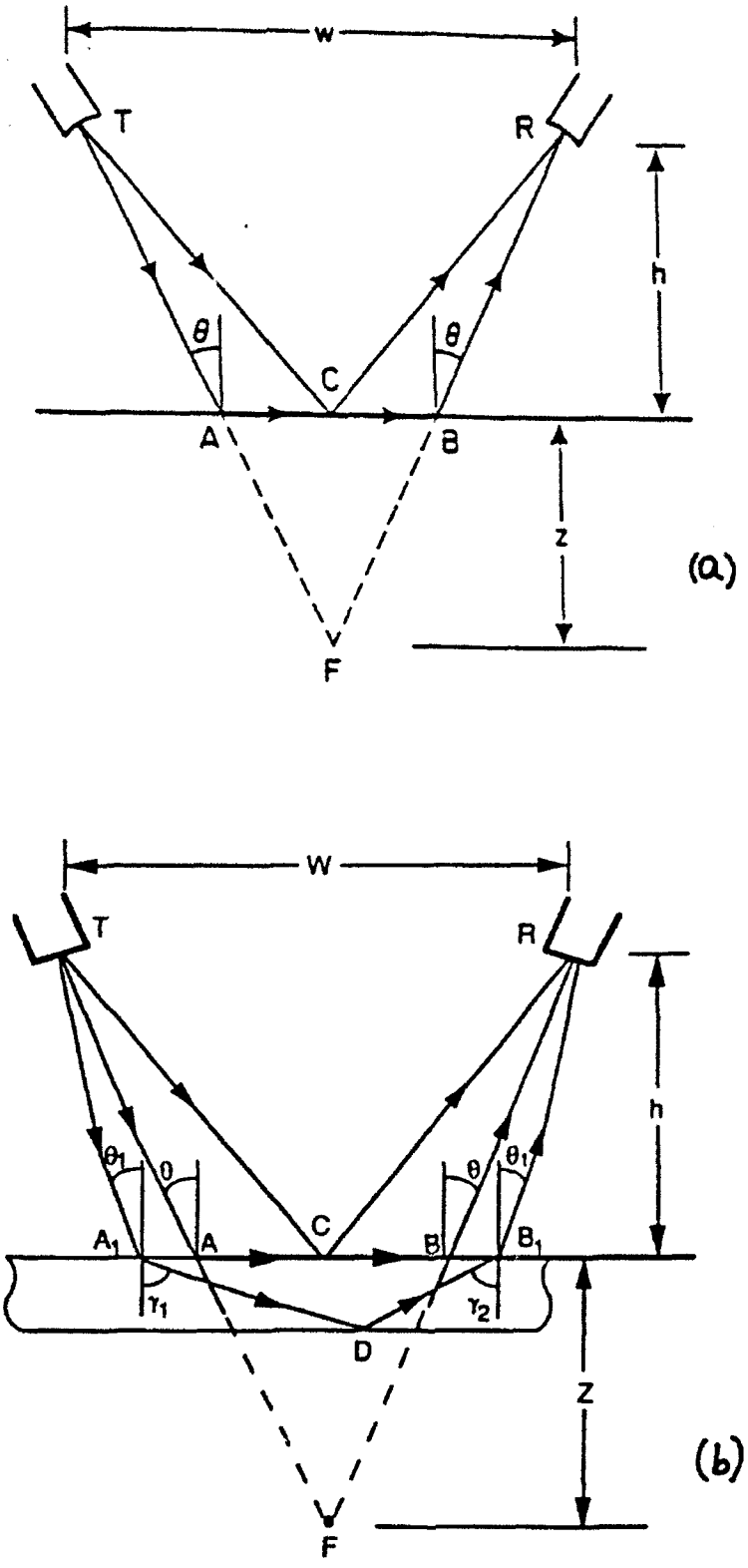


Fig.1: Schematic diagram of the transmitter-receiver-reflector arrangement showing different signal paths during its flight from the transmitter to the receiver, (a) solid half-space reflector, (b) solid plate reflector.

$$t=\frac{2}{\alpha_f}\left(R_c-\frac{z}{\cos\theta}\right)+\frac{2z.\tan\theta}{C_s} \qquad (5)$$

where C_s is the surface wave velocity and t is the time of travel. From equation (7) one can obtain

$$\frac{dt}{dz}=-\frac{2}{\alpha_f\cos\theta}\left(1-\frac{\alpha_f\sin\theta}{C_s}\right) \qquad (6)$$

It should be noted here that when surface waves are not generated dt/dz varies with z (Eq.6) but when surface waves are generated dt/dz is independent of z (Eq.8).

Eq.(8) can be rewritten in the following form

$$C_s=\frac{\alpha_f.\sin\theta}{1+\frac{\alpha_f}{2}\left(\frac{dt}{dz}\right)\cos\theta} \qquad (7)$$

The above equation can be used to obtain the Rayleigh wave velocity in a solid.

The situation becomes much more complex when a plate is used as a reflector instead of a solid half space. As shown in Fig.1b now a beam can reach the receiver after being reflected at the bottom surface of the plate in addition to the two possible paths described earlier. Angles γ_1 and γ_2 will be identical if both rays A_1D and B_1D are either longitudinal or shear wave. However, if one of them is longitudinal and the other one is shear wave then γ_1 and γ_2 will be different. If θ_1 is the angle corresponding to which the signal is reflected by the back face of the plate then for a change of z of amount δz the change of travel path in the coupling fluid should be $\delta z/\cos\theta_1$; the path length in the solid doesn't change. Hence, change in the flight time is given by

$$\delta t = -\frac{\delta z}{\alpha_f \cos\theta_1} \tag{8}$$

or

$$\frac{dt}{dz} = -\frac{1}{\alpha_f \cos\theta_1} \tag{9}$$

In this case also dt/dz is independent of z, as was the case when the surface wave was generated (Eq.8). Hence, if dt/dz is found to be constant one should first decide if it is due to surface (Rayleigh or Lamb) wave or reflection from the back face of the plate. Reflection from the back face of the plate can only occur when the transducer position angle θ is close to θ_1 required to generate longitudinal or shear wave inside the plate.

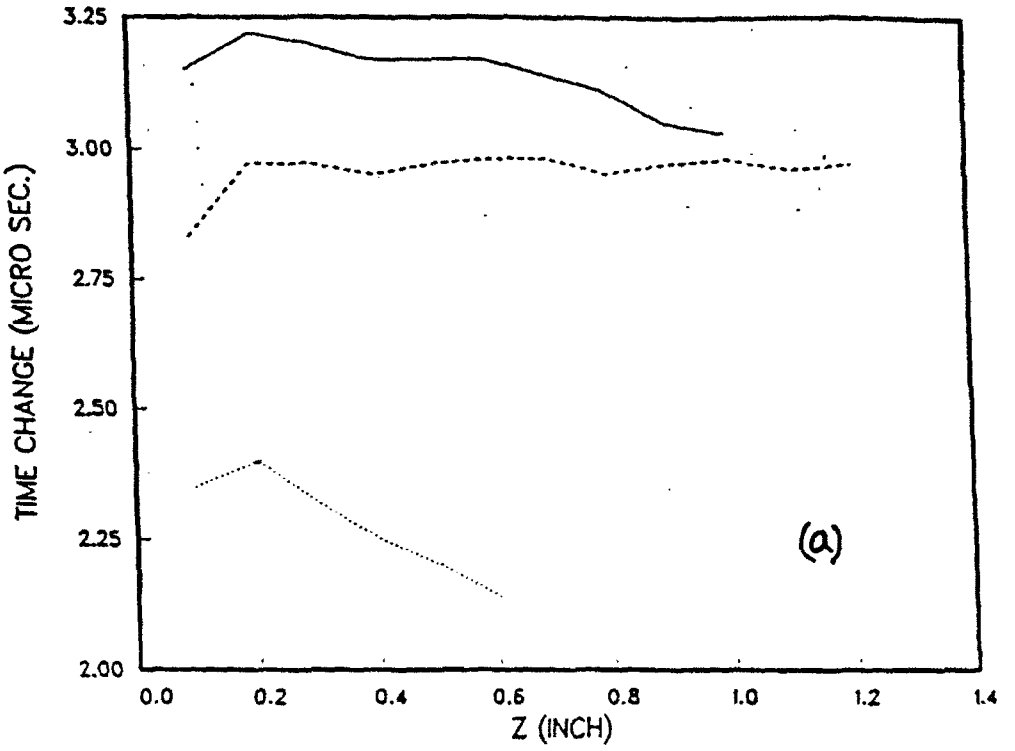
EXPERIMENTAL RESULTS

A number of experiments have been carried out with different types of specimens - isotropic metallic blocks, isotropic metallic plates and anisotropic composite plates. Rayleigh wave velocities in metallic blocks and Lamb wave velocities in plates have been experimentally obtained and compared with their expected values. In this section some of these results are presented.

Rayleigh Wave Velocity Measurement: The first experiment is carried out with a titanium (Ti) block, Ti64. Two identical unfocused transducers (2.2 MHz frequency, 0.25 inch diameter) are used as transmitter and receiver in a pitch-catch position as shown in Fig.1, and the Ti block is used as the reflector. The incident angle as well as the reflection angle is set at 20°, 30°, and 45° for three different sets of experimental readings. For every incident angle the

vertical distance (h) between the transducers and the specimen (see Fig.1) is first set at a desired value. Then the horizontal distance (w) between the transducers is adjusted to get the maximum reflected amplitude. This position is identified as z=0 position. The reflected waveform corresponding to this position is seen on the oscilloscope screen and the time corresponding to a specific peak (or dip) of the waveform is recorded. The transducer-receiver arrangement is then lowered by 0.1 inch (δz=0.1 inch). It shifts the reflected waveform in the time axis. The horizontal shift (i.e. time shift) of the same peak or dip is noted. This process of increasing z by 0.1 inch step size and recording the time change associated with a specific peak of the wave form is continued up to z = 1 inch or until the reflected signal becomes too weak to be accurately detected. The plots of the time change against z for three different angles of incidence are shown in Fig.2a. Solid, dashed and dotted lines are for 20°, 30° and 45° incident angles respectively. One can see from this figure that the solid and dotted lines decrease with z but the dashed line remains almost constant. Hence, from Eqns. (5) and (8) one can conclude that for 30° incident angle the Rayleigh wave has been generated. The average value of the flat region of the dashed line is 2.931 μsec. This value corresponds to δz=0.1 inch. So dt/dz is equal to -1.154 μsec/mm, minus sign is taken because time of flight decreases as z increases. Substitution of this value in Eq.(9) gives Rayleigh wave velocity equal to 2.92 km/sec, when the longitudinal wave velocity in water (α_f) is 1.49 km/sec (obtained experimentally). The published Rayleigh velocity in Ti is 2.89 km/sec [Metrotek, 1982]. The difference between these two values is only 1%.

Instead of using Eq.(9) if one uses Snell's law



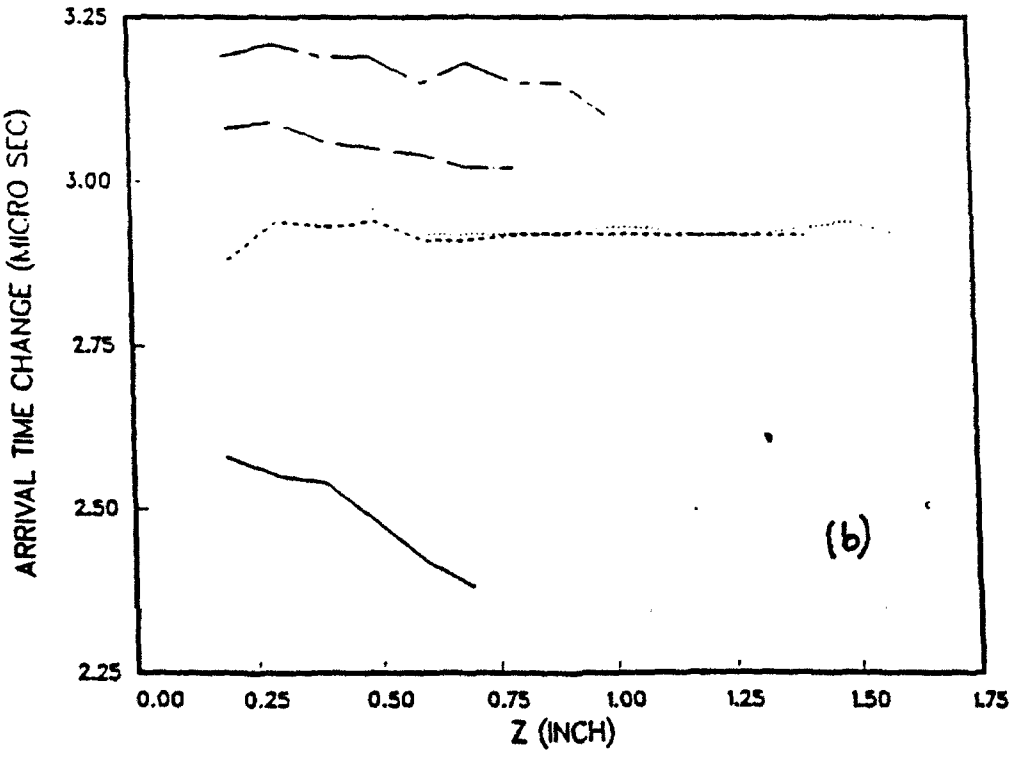


Fig.2: Arrival time change versus z (dt-z) curve for titanium block. Figures (a) and (b) are for two different titanium blocks. Different lines are for different angles of incidence.

$$C_s = \frac{\alpha_f}{\sin\theta_c} \qquad (10)$$

to compute the Rayleigh velocity then one obtains C_s = 2.98 km/sec, for $\theta_c = 30^\circ$, an error of 3.1%.

Fig.2b shows similar curves for another Ti block. This is taken to study how Rayleigh velocity varies from one specimen to another. Solid, dashed, dotted, chain-dotted and chain-dashed lines are for $\theta = 40^\circ$, 32.5°, 30°, 25° and 20° respectively. Clearly, Rayleigh waves have been generated for 32.5° (dt/dz = -1.150 μsec/mm) and 30° (dt/dz = -1.151 μsec/mm) incidence. When these angles and corresponding dt/dz values are substituted in Eq.(9) the Rayleigh wave velocity is obtained as 2.89 km/sec for both these two cases. It matches exactly with the published value. These experimental results clearly show that a slight change (a few degrees) in the incident angle from the critical Rayleigh angle doesn't alter the Rayleigh wave velocity computation by the present technique. On the other hand the computation becomes very much angle dependent if Eq.(12) is used to calculate the Rayleigh velocity instead of Eq.(9). From Eq.(12) these values come out to be 2.77 (4.15% error) and 2.98 km/sec (3.11% error) for 32.5° and 30° angles. Hence, the main advantage of the proposed technique is that an approximate knowledge of the critical angle is sufficient to compute the Rayleigh wave velocity accurately.

Similarly Rayleigh velocities of copper (Cu), tungsten (W) and aluminum (Al) were obtained (plots are not shown in this report because of page limitation but can be found in Kundu and Maxfield [1992]). Rayleigh wave velocities in these specimens are listed in table 1 in the next page.

Table 1: Materials and Their Rayleigh Wave Velocities.

Mat. Type	Published C_s (km/sec)	θ for const.dt/dz	C_s from Eq(9) and (%error)	C_s from Eq(12) and (%error)
Ti	2.89	30°	2.92(1%)	2.98(3.1%)
Ti	2.89	30°	2.89(0%)	2.98(3.1%)
Ti	2.89	32.5°	2.89(0%)	2.77(4.2%)
Cu	2.15	45°	2.11(1.9%)	2.15(0%)
Cu	2.15	43.35°	2.11(1.9%)	2.17(1%)
W	2.65	32.5°	2.62(1.1%)	2.77(4.5%)
W	2.65	35°	2.61(1.5%)	2.60(1.9%)
Al	2.92	30°	2.97(1.7%)	2.98(2%)
Al	2.92	32°	2.94(0.7%)	2.81(3.8%)
Al	2.92	35°	2.91(0.3%)	2.60(11%)

Lamb Wave Velocity Measurement: The technique described above for Rayleigh wave velocity computation can in principle be extended to the Lamb wave velocity determination also. However, in this case complication increases because of the dispersive nature of the Lamb waves. Lamb wave dispersion function has multiple roots at higher frequencies. For an isotropic homogeneous plate Lamb wave dispersion equation can be given by a closed form expression [Mal and Singh (1991)]. This dispersion equation is solved for 0.0635 inch (1.61 mm) thick aluminum plate (longitudinal wave velocity = 6.32 km/sec, shear wave velocity = 3.13 km/sec, and density = 2.7 gm/cc), and the Lamb wave roots are plotted by cross marks in Fig.3. To verify these theoretical curves, Lamb wave roots are also generated experimentally by conventional tone burst frequency sweep

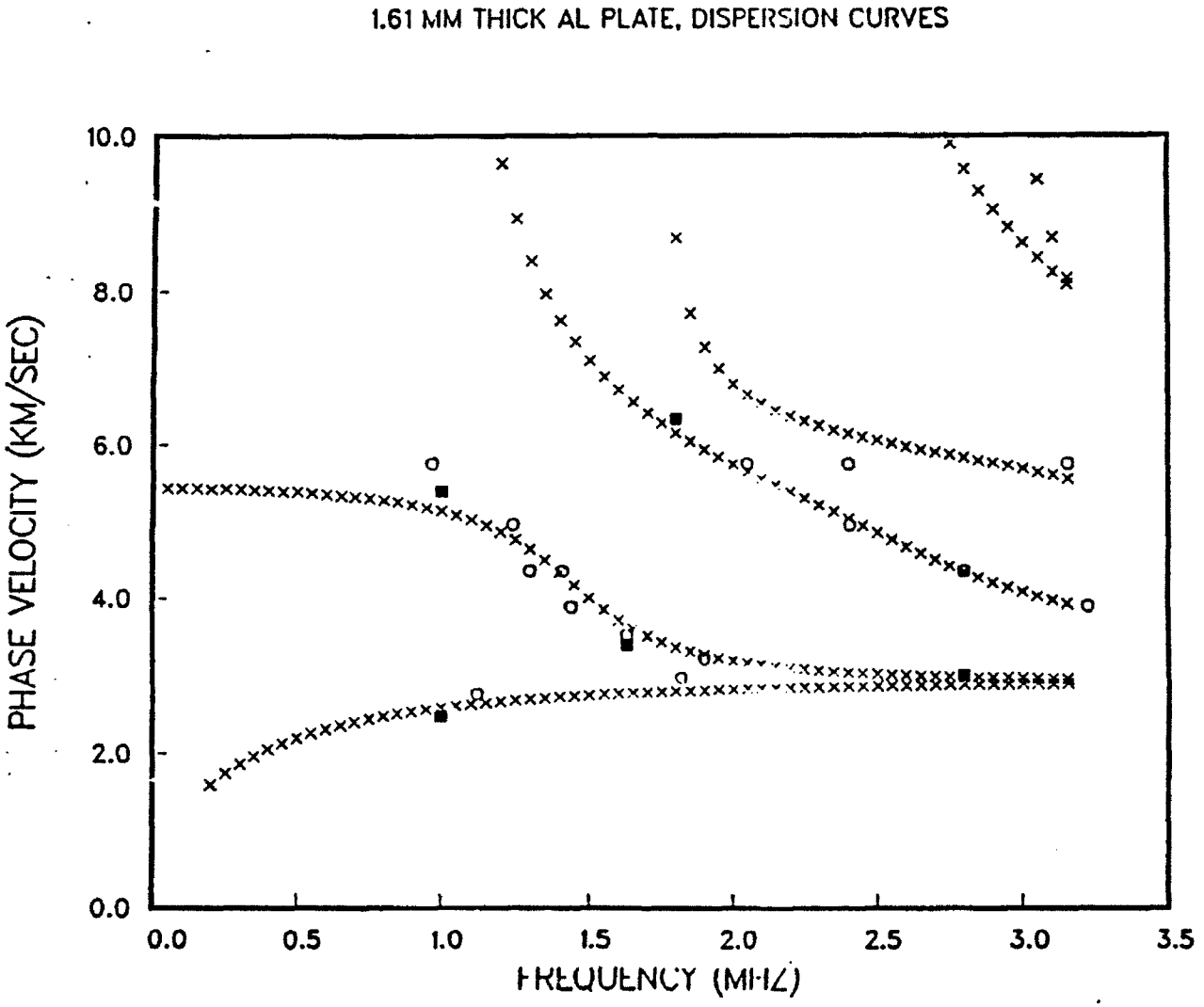


Fig.3: Dispersion curves for 0.063 inch (1.61 mm) thick aluminum plate. Cross marks are obtained analytically, circles are from frequency sweep experimental technique and black squares are from the present technique.

technique. In this technique two broad band transducers are used in the pitch-catch orientation (Fig.1). The transmitter is excited by a signal function generator which generates continuous wave forms (tone burst) and varies the signal frequency continuously between two limits (frequency sweeping). The reflected signal amplitude (vertical axis) versus frequency (horizontal axis) is seen on an oscilloscope screen. For certain incident angle if a Lamb wave mode is generated then energy leaks through the fluid-solid interface in the form of leaky Lamb waves (LLW) and a dip (local minimum) is observed in the amplitude-frequency plot of the reflected signal. Corresponding phase velocity can be obtained from Eq.(12) simply replacing θ_c in that equation by the incident angle θ.

The frequency sweeping was carried out using two sets of transducers, working in two intervals 0.5-1.6 MHz and 1-3.5 MHz. Angle θ was changed from 15° to 45° at an interval of 2.5°. Thus, a square region in Fig.3 bounded by 0.5 MHz and 3.5 MHz along the horizontal axis, and 2.11 km/sec ($\theta=45^\circ$) and 5.76 km/sec ($\theta=15^\circ$) along the vertical axis was scanned for Lamb wave roots. Small circles in Fig.3 are the Lamb wave roots obtained in this manner. These markers match quite well with the theoretical dispersion curves except for a couple of points. It can be noted here that in the regions where the dispersion curves are flat and almost horizontal (first antisymmetric and symmetric modes at frequency greater than 2 MHz and first symmetric mode for frequency less than 0.75 MHz) frequency sweep technique cannot detect any Lamb wave root. This is expected because for the phase velocities for which dispersion curves become horizontal leaky Lamb waves are generated over a wide range of frequencies instead of a discrete number of frequency values. Hence, no local minimum is observed at any one frequency

although Lamb waves are generated. This problem of missing roots in the flat regions can be avoided by the new technique. Black square markers of Fig.3 are obtained by the present technique. Readers are referred to Kundu and Maxfield (1992) for a detailed description on how the black squares have been obtained experimentally.

Similar technique can be followed to generate and detect Lamb waves in anisotropic plates as well. Lamb wave dispersion curves are experimentally obtained for 0.083 inch thick unidirectional graphite-epoxy composite plate (Fig.4), 0.04 inch thick unidirectional metal matrix composite plate [SiC whiskers in titenium matrix] (Fig.5), and 0.044 inch thick multilayered (0-90 lay out) metal matrix composite plate (Fig.6). From these figures one can see that there is a significant difference in the dispersion curves in the two perpendicular directions, along the fiber and across the fiber, for unidirecitonal composite plates. The difference in the 0° and 90° directions is more in Fig.4 than in Fig.5. This is because the elastic property difference between the fiber material and the matrix material is more for the graphite-epoxy composite plate than the SiC-Ti composite plate. When the fibers are oriented in both 0° and 90° directions the dispersion curves for different angles of orientation of the plate become almost identical (Fig.6). It indicates that the plate has become more isotropic in this case.

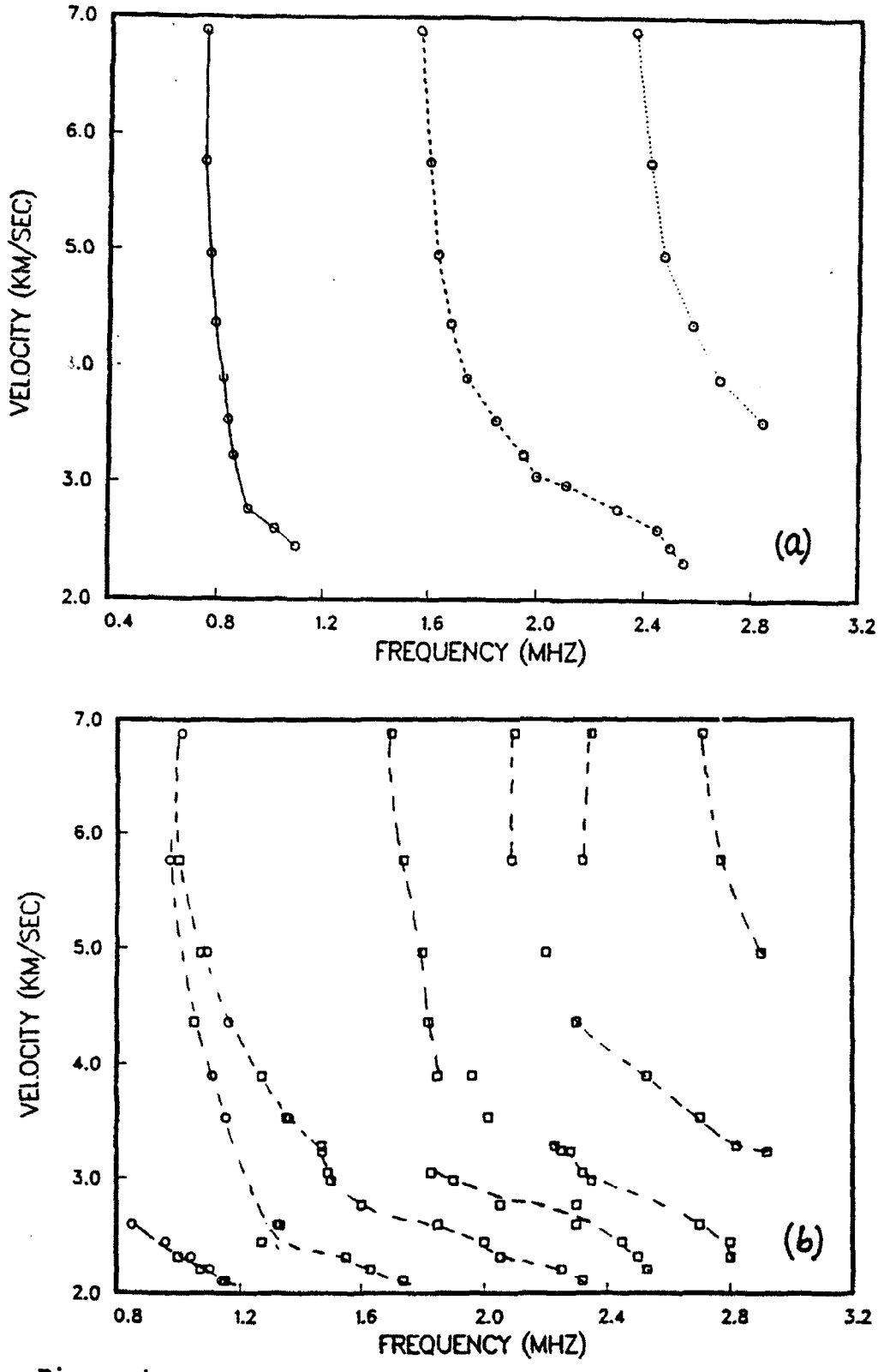


Fig.4: Dispersion curves for 0.083 inch thick unidirectional graphite-epoxy composite plate. (a) along the fiber, (b) across the fiber.

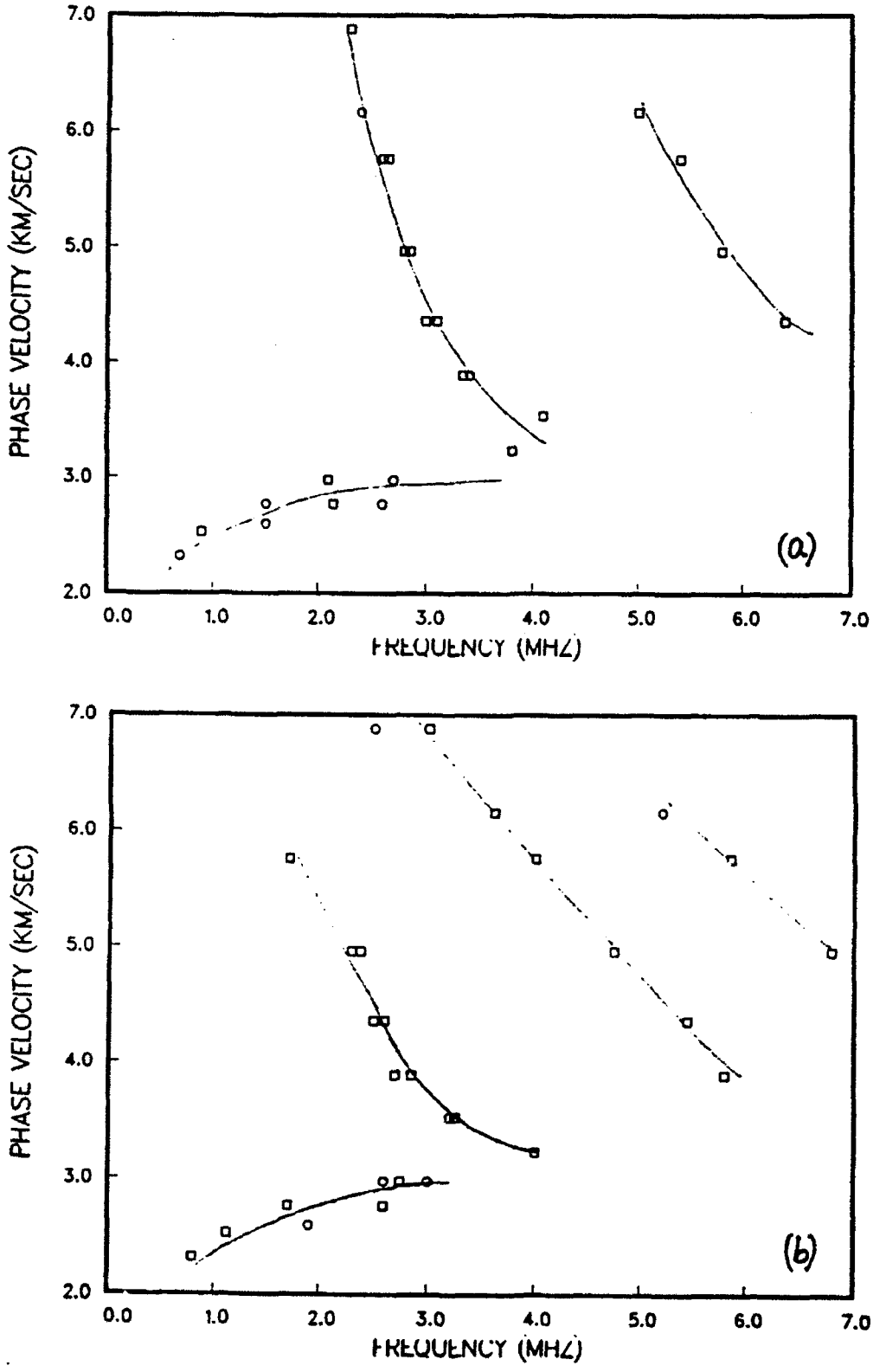


Fig.5: Dispersion curves for 0.04 inch thick unidirectional metal matrix composite (SiC^{W}/Ti) plate. (a) along the fiber, (b) across the fiber.

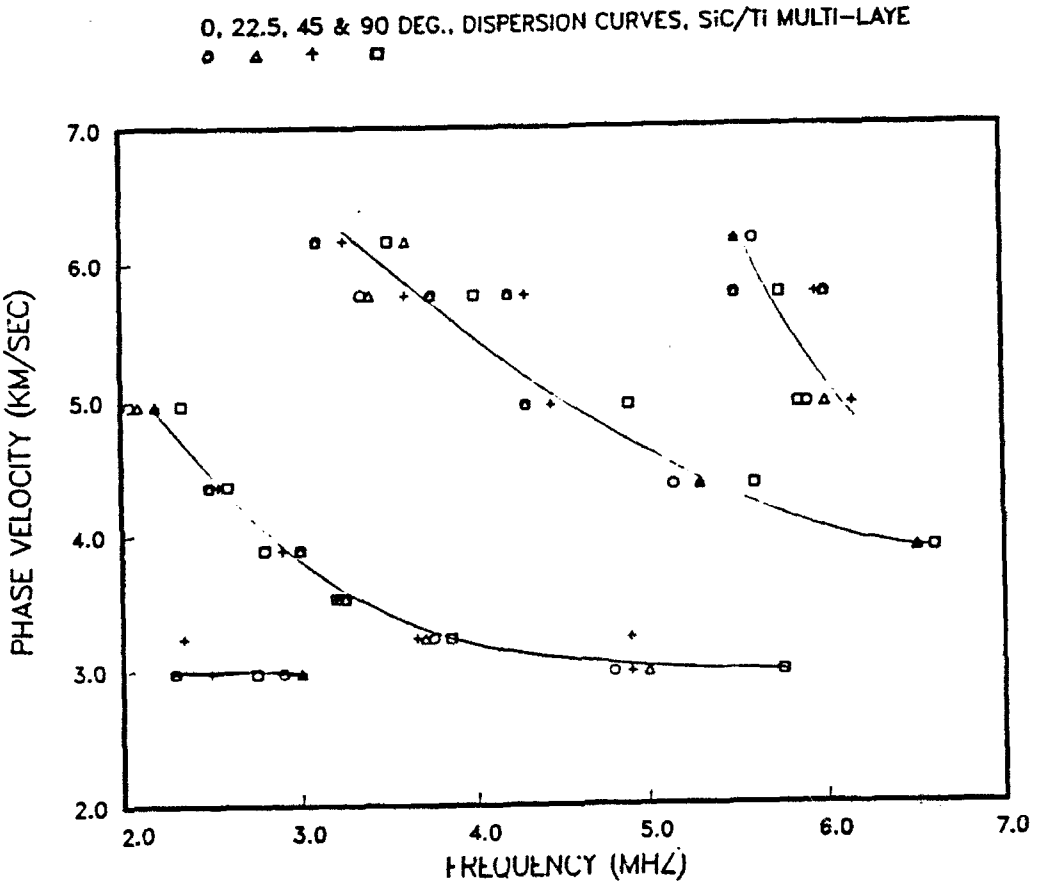


Fig.6: Dispersion curves for 0.045 inch thick multilayered (0-90 lay out) metal matrix composite (SiC^{W}/Ti) plate. Circles, triangles, crosses and squares are 0°, 22.5°, 45°, and 90° orientations of the plate. The orientation angle is measured relative to the fiber direction at the top layer.

CONCLUDING REMARKS

A new method to detect Rayleigh and Lamb waves is developed. The technique is verified by generating and detecting these waves in a variety of materials. The advantage of this technique over other available techniques is its simplicity. In this technique no special type transducer is needed to generate or detect Rayleigh or Lamb waves in a material. The Rayleigh wave velocity in a homogeneous material can be measured very accurately by this technique as shown in this paper. One major advantage of this technique over the conventional tone burst frequency sweep technique for detecting Lamb wave modes is that it is capable of detecting the modes in the region where the dispersion curves become almost horizontal (parallel to the frequency axis) which are missed by the conventional technique. Among the shortcomings of this technique one can point out that it takes comparatively longer time to carry out the experiment. For dispersive Rayleigh and Lamb waves if two modes are very close to each other this technique may fail to distinguish them.

ACKNOWLEDGMENT

This research was supported by the Summer Faculty Research Program of Air Force Office of Scientific Research (Contract Number F49620-90-C-0076). The experiment was carried out at the Wright-Patterson Material Laboratory in Dayton, Ohio. Continuous encouragement of Drs. P. K. Bhagat and T. Moran of the Wright-Patterson laboratory, and their helpful suggestions are gratefully acknowledged. The author would also like to thank Mr. M. Ruddell for his technical help during the experiment.

REFERENCES

Achenbach, J.D., Komsky, I., and Zhang, Ch., 1990, "Ultrasonic Surface Wave Technique to Determine Fatigure Damage", in Damage Mechanics in Engineering Materials, Vol. AMD 109, Pub. ASME, New York, USA, p.227-236.

Egle, D. M., 1980, "Using the Acoustoelastic Effect to Measure Stress in Plates", Report #UCRL-52914, Lawrence Livermore Laboratory.

Kinra, V. K., and Dayal, V., 1988, "New Technique for Ultrasonic-Nondestructive Evaluation of Thin Specimens", Experimental Mechanics, Vol.28, p.288-297.

Kinra, V. K., and Dayal, V., 1987, "Ultrasonic Nondestructive Evaluation of Fiber-Reinforced Composite Material - A Review", Sadhana, Vol.11, p.419-432.

Kundu, T., and Maxfield, B., 1992, "A New Technique for Measuring Rayleigh and Lamb Wave Velocities", Journal of the Acoustical Society of America, submitted.

Kushibiki, J., Ohkubo, A., and Chubachi, N., 1982a, "Effect of Leaky SAW Parameters on V(z) Curves Obtained by Acoustic Microscopy", Electronics Letters, Vol.18, p.668.

Kushibiki, J., Ohkubo, A., and Chubachi, N., 1982b, "Material Characterization by Acoustic Line Focus Beam", Acoustic Imaging, Pub. Plenum Press, New York, Vol.12, p.101.

Kushibiki, J., Matsumoto, A., and Chubachi, N., 1983, "Material Characterization by Acoustic Microscope with Line-Focus Beam", Acoustic Imagir Pub. Plenum Press, New York, Vol.13, p.193.

Lu, W. Y., Maxfield, B., and Kuramoto, A., 1990, "Ultrasonic Velocity Measurement by Correlation Method", Proceedings of the Spring Conference on Experimental Mechanics, Albuquerque, New Mexico, June 3-6, 1990.

Mal, A. K., and Singh, S. J., 1991, Deformation of Elastic Solids, Pub. Prentice Hall.

Mal, A. K., Yin, C.,-C., and Bar-Cohen, Y., 1991, "Ultrasonic Nondestructive Evaluation of Cracked Composite Laminates", Composites Engineering, Vol.1, p.85-101.

Metrotek Inc., 1982, Table of Ultrasonic Properties, Report # AN23.

Zhang, Ch., and Achenbach, J. D., 1991, "Effective Wave Velocity and Attenuation in a Material with Distributed Penny Shaped Cracks", Internation Journal of Solids and Structures, Vol.27, p.751-767.

Fatigue Damage Accumulation of Angle-Plied Cord-Rubber Composites

Byung-Lip ("Les") Lee and Jeffrey A. Smith
Associate Professor and Graduate Student
Department of Engineering Science and Mechanics

Pennsylvania State University
227 Hammond Building
University Park, PA 16802

Final Report for:
Summer Faculty Research Program
Air Force Wright Laboratory
Flight Dynamics Directorate, Vehicle Subsystems Division
(USAF Researchers: John P. Medzorian, Dr. Mangal Chawla)

Sponsored by:
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

September, 1992

Fatigue Damage Accumulation of Angle-Plied Cord-Rubber Composites

by

Byung-Lip ("Les") Lee and Jeffrey A. Smith
Associate Professor and Graduate Student
Department of Engineering Science and Mechanics

Pennsylvania State University
University Park, PA 16802

ABSTRACT

Current phase of the study was undertaken to examine tensile fatigue behavior of cord-rubber composites representing bias tire carcass under various frequencies up to the level which closely simulates loading during high-speed take-off of aircraft. At a given stress amplitude, the use of higher cyclic frequency was found to affect strain response and heat build-up characteristics of composites significantly. The lower level of *initial strain* observed at higher frequency stems clearly from strain rate dependence of deformation of rubber matrix composites. The temperature profile of the specimens subjected from 20 to 30 Hz loading showed that hysteretic heating under these conditions may lead to thermal fatigue failure as well as chemical degradation influencing both fiber-matrix adhesion strength and matrix strength. The involvement of material degradation process was indicated by the fact that gross failure of composites requires lower value of *dynamic creep* when the frequency is increased. At the same time, the use of higher frequency resulted in exponentially higher *dynamic creep rate*. For load range tested, the *energy loss per cycle* was found to be nearly constant and independent of the frequency. As a result, the *energy loss per unit time* became linearly proportional to the frequency. Since fatigue life of carcass composites is linearly proportional to the inverse of frequency, it was postulated that the rate of energy loss determines the lifetime of composites. The study revealed that the specimen surface temperature may not describe such critical parameters as internal heating particularly at the point of crack initiation or true heat dissipation rate.

I. INTRODUCTION

As described in our previous papers (*1-3*), an initiative is underway to realign the aircraft tire test procedure where the emphasis is placed on more cost-effective and less empirical prediction of fatigue lifetime of tires. The operational life of aircraft tires is currently certified by costly dynamometer testing in which the tires are subjected to various combinations of speed and footprint load representing typical operating conditions in the field (*4,5*). Depending on the severity of loading conditions, cumulative damage is induced in critical regions of tires and eventually develops into catastrophic failure (*5,6*). For bias-ply aircraft tires, one of the most susceptible locations for damage initiation and accumulation has been the *carcass* in the *shoulder* area. The dynamometer testing clearly provides an accelerated means of evaluating the structural durability and integrity of aircraft tires. However, the usefulness of dynamometer tests in current form is limited by their empirical nature. The test results reflect merely the sensitivity of each particular tire design and construction to a given set of loading conditions, unless underlying mechanisms of property degradation, damage accumulation and structural failure of tires are identified.

In implementing the proposed initiative, our research effort has been concentrated in defining the deformation and fracture mechanisms of angle-plied cord-rubber composite specimens which simulate the material elements of bias aircraft tire carcass in the shoulder area. The initial goal of our study is to identify the stress, strain or temperature parameters that control the process of fatigue damage accumulation for tire carcass composites under *uniaxial tension*, *biaxial tension* and *out-of-plane bending*. Eventually, it is hoped a fatigue law can be formulated that will serve to predict the life of the tire carcass on a more analytical basis and thereby to complement empirical nature of dynamometer tests. For a better determination of the failure modes, our study initially utilized the *model* composites reinforced by steel wire cables. Subsequently the research program included nylon fiber cord-reinforced composites which represent the actual *aircraft tire carcass*. Under uniaxial cyclic loading which represents fluctuating circumferential tension in the footprint region of tires, these angle-plied composite specimens were found to exhibit high level of interply shear deformation.

The interply shear deformation behavior of angle-plied composites were analyzed by a number of investigators in the past (*1,2,7-13*). Interply shear strain develops in angle-ply laminates when the constituent plies exhibit in-plane shear deformation of opposite direction but the action is prevented by mutual constraint due to interply bonding.

Compared with the case of fiber-reinforced plastic composites (*13*), cord-rubber composites exhibit unusually high level of interply shear strain which results from the load-induced change of reinforcement angle allowed by extreme compliance of rubber matrix. Our previous study (*1*) showed that, at an axial tensile strain of 10 percent, an interply shear strain of around 30 percent develops in the nylon cord-reinforced composites representing the aircraft tire carcass with an initial reinforcement angle of +/-38 degree. Experimental study of the load-displacement response and interply shear strain variations was accompanied by detailed stress analysis based on finite element method (*1,2*). Linear elastic orthotropic or isotropic material elements were used for modeling. Our predictions were in reasonably good agreement with the experimental results.

Above a critical value of interply shear strain, these angle-plied composites were found to exhibit localized failure initiated in the form of *cord-matrix debonding* as observed by other investigators (*14,15*). Debonding is started around the cut ends of fibrous cord reinforcements at the edge of the finite width coupons. This phenomenon is justified since the maximum interply shear strain occurs at the edge of the specimen. A new finding by our previous study (*1*) was that the critical load for the onset of fiber-matrix debonding constitutes a threshold level for semi-infinite fatigue life, i.e. *fatigue endurance limit*, of the composites. The physical meaning of an endurance limit is that, with cyclic stresses lower than this threshold, fiber-matrix debonding is never initiated nor developed. Under cyclic tensile stresses exceeding the endurance limit, fiber-matrix debonding was found to be progressively worsened and developed into *matrix cracking* and *delamination* leading to gross failure of the composites.

The damage accumulation in the forms of debonding, matrix cracking and delamination was accompanied by local strain increase (referred as *dynamic creep* hereafter), heat generation and acoustic emission (AE). The dynamic creep rate and the rate of temperature increase were inversely proportional to the fatigue life according to a power law (2). In monitoring of AE, distinctly different rates of signal accumulation could be assigned to the debonding and delamination failure modes (2). Our study examined the effects of different stress parameters on the fatigue resistance of composites to a limited extent (*2,3*). Among them, *stress amplitude* was found to play a dominant role in determining fatigue lifetime. When the minimum cyclic stress is near zero, the S-N (stress amplitude vs fatigue life) curve of the composites exhibits a steeper slope with shortened fatigue life in the normal loading range. However, the study was restricted to the case of low-frequency cyclic loading.

Reviewing the progress of our study so far, it is clear that more systematic efforts are needed to investigate the following two areas of interest for aircraft tire carcass composites: (a) the effects of stress, strain and temperature *history* on the fatigue fracture mechanisms particularly under high-frequency cyclic loading; (b) *monitoring* of damage accumulation process. Under high-frequency cyclic loading, heat build-up due to hysteretic nature of constituent materials may play more prominent roles in controlling the fatigue life of cord-rubber composites. Understanding of the subject will be an important step for the future study which plans to predict fatigue lifetime of composites under complex sequences of load and frequency (*3*). Our study of fatigue fracture mechanisms of tire carcass composites should also include the development of damage monitoring techniques as its integral part. When fully established, these experimental methodologies for damage monitoring are expected to be *in-situ* or *real-time* means of predicting residual life of tire carcass under fatigue loading.

Within this context, our study has been expanded to examine tensile fatigue behavior of aircraft tire carcass composites under various frequencies up to the level which closely simulates loading during high-speed take-off of aircraft. The study has also assessed the effectiveness of the measurement of strain or temperature in monitoring of fatigue damage accumulation under high frequency. This paper reports preliminary findings of these efforts.

II. OBJECTIVES OF THE RESEARCH EFFORT

The present study was undertaken to investigate the effects of stress, strain and temperature history on the fatigue fracture mechanisms of aircraft tire carcass composites particularly under *high-frequency* loading and to assess the effectiveness of the measurement of strain or temperature in monitoring of damage accumulation process.

III. EXPERIMENTS

Angle-plied composite laminate specimens were prepared from the calendered plies used for construction of actual standard-grade carcass of *KC-135 aircraft tire*. The composite system was made of carbon black-filled proprietary rubber compound matrix and 1260/2

nylon cord reinforcement laid at an angle of +/-38 degree (Table 1). The materials were supplied by The Goodyear Tire & Rubber Company (Akron, OH). To avoid tension-bending coupling, the laminates were constructed with a symmetric ply lay-up. The end tabs were added to the specimens to prevent grip failures that could be experienced during mechanical testing. Coupon specimens of 19mm width were machined from these panels and edges were polished on a grinding wheel.

In simulation of circumferential loading of a tire in the footprint region, composite coupon specimens were subjected to cyclic uniaxial tension. Cyclic testing was performed under a broad range of frequency from 2 to 30 Hz with fixed values of stress range (6.1 MPa) and minimum stress (1.5 MPa). The specimens were run until gross failure occurs. In the case of 5 Hz loading, further variation of stress range was implemented with a constant minimum stress of 1.5 MPa in order to define S-N curve.

Load and displacement were measured at small time intervals during cyclic testing. Typical data acquisition consisted of 20 points per cycle and 4 cycles once every 20 seconds. Force-displacement data recorded during testing allowed for the generation of hysteresis loops for individual cycles. Fatigue testing was performed with continuous monitoring of heat generation. Temperature build-up as a function of fatigue life was recorded initially with a thermocouple attached to the specimen surface. Later the temperature was monitored using Inframetrics 525 thermography equipment. The surface temperature was measured across the width of the specimen and the image of the profile was recorded on a video system.

IV. RESULTS AND DISCUSSION

As stated earlier, the scope of our study has been expanded to examine fracture behavior of aircraft tire carcass composites under high-frequency cyclic loading. Fatigue lifetime of composites was measured under a broad range of frequency from 2 to 30 Hz with the values of stress range and minimum stress fixed at 6.1 and 1.5 MPa respectively. In the case of 5 Hz loading, a S-N curve was constructed by varying stress range (Figure 1). The modes of deformation and failure of composites were not affected by the use of higher frequency. As observed in the past study, angle-plied carcass composites exhibited a high level of interply shear strain. Since the maximum cyclic stresses used in the study exceeded the fatigue endurance limit, fiber-matrix debonding was readily induced in the

first cycle around the cut ends of reinforcing cords. Accompanied by the increase of cyclic strain (i.e. dynamic creep) and heat build-up, the cord-matrix debonding was progressively worsened and developed into matrix cracking and delamination eventually leading to gross failure of the composites. Although no alteration of failure sequence was observed by the increase of frequency, the strain response and heat build-up characteristics which are indicatives of damage accumulation were found to be strongly dependent upon the level of frequency.

The process of dynamic creep undergoes three stages (Figure 2). After an initial stepwise increase in strain, the strain increases at a progressively slower rate until reaching a *steady state* at about 20% of the fatigue life. In this steady-state region, the cyclic strain increases at a constant rate and the damage takes the form of cord-matrix debonding and matrix cracking with no delamination observed. Above 80% of the fatigue life, partial delamination appears at the specimen edge and the cyclic strain increases at a progressively higher rate eventually leading to a catastrophic failure. The increase of frequency was found to affect dynamic creep behavior in several contrasting ways: lower *initial strain*, lower *extent of dynamic creep at gross failure*, and exponentially higher *rate of dynamic creep* (Figures 2 through 5). Here the extent of dynamic creep denotes the increase of displacement or strain beyond the initial instantaneous deformation. Actual values or the values extrapolated from the steady-state region were used to define the strain at gross failure. The same trend was observed in both cases. The dynamic creep rate was estimated from the slope of steady-state region.

The lower level of initial strain observed at higher frequency stems clearly from *strain rate dependence* of deformation of rubber matrix composites. It is well-known fact that, as the rate of elongation increases, the moduli of rubbers or rigid polymers increase due to their viscoelastic nature. An increase in speed of testing is therefore similar to a decrease of temperature in influencing the modulus of rubber matrix of composites. In high-frequency cyclic loading, higher strain rate is expected to produce initially lower strain at given cyclic stress as in the case of lower temperature. However, this trend is opposed by the effect of *hysteretic heating* as soon as cyclic loading starts. In Figure 6, the peak value of surface temperature profile across the width of coupon was plotted against the normalized time. The normalized time is defined as the time at given point divided by the time for gross failure. This figure clearly shows that the surface temperatures of composite specimens over their fatigue lives increase with higher frequency. Smaller temperature difference between the results of 20 and 30 Hz loading suggests that the state of material approaches

near the onset of chemical degradation. In fact, the temperature profile of the specimens subjected to 20 and 30 Hz loading showed the lack of steady-state region and their temperature reached about 150 C (300 F) which is the cure temperature of composites.

Aside from the effect of *chemical degradation*, for some polymers in which the strength decreases with temperature, hysteretic heating could lead to so-called *thermal fatigue failure*. In addition to the decrease in strength of rubber matrix, hysteretic heating may result in the degradation of fiber-matrix adhesion and thereby influence the process of damage accumulation in cord-rubber composites. It is certainly premature to speculate which degradation mechanism(s) will be responsible for the dependence of fatigue lifetime of aircraft tire carcass composites on the frequency. However, regardless of its exact nature, the involvement of material degradation process was indicated by the fact that gross failure of composites requires lower value of *dynamic creep* when the frequency is increased (Figure 4). Data corresponding to 2 Hz shows deviation from the rest of frequencies tested. This implies that more testing is needed at very low frequencies to verify the trend. As shown in Figure 5, the use of higher frequency also resulted in exponentially higher *dynamic creep rate* which reflects increasingly more rapid accumulation of damage in carcass composites. When fatigue lifetime was plotted against loading frequency on a logarithmic scale in Figure 7, the least square curve fit assumed a power law relationship with an exponent of nearly -1 indicating that they are inversely proportional. In other words, fatigue life of composites is linearly proportional to the inverse of frequency i.e. the time per unit cycle.

The above data suggest that the use of higher frequency results in the decrease of fatigue life of cord-rubber composites by simply shortening the time to reach a preset level of dynamic creep for gross failure. The extent of dynamic creep for gross failure, in turn, is lowered by increasing cyclic frequency, presumably because greater amount of hysteretic heating leads to more degradation of fiber-matrix adhesion strength or matrix strength. Although heating of the specimen is its most pronounced symptom, the amount of hysteresis itself is a fundamental parameter for expressing mechanical energy loss in a material system. In our study, the area within the hysteresis loop was used to determine directly the amount of energy loss for each composite specimen during a particular cycle. In addition to hysteresis loop area, the value of tan δ, where δ is the phase angle difference between stress and strain, was calculated to express energy loss relative to energy stored in a material system. The following relationship was assumed: $\delta = \sin^{-1} (A/\pi \sigma_0 \varepsilon_0)$ where

A is the area inside the hysteresis loop, and σ_0 and ε_0 are the stress and strain amplitudes respectively.

Hysteresis loop area and tan δ were measured at regular intervals throughout the fatigue life of aircraft tire carcass composites. The values are given in Table 2 for the specimens subjected to cyclic loading with different frequencies. Typical hysteresis curves under the frequencies of 2 and 20 Hz are shown in Figures 8 and 9 respectively. The values of energy loss in terms of hysteresis loop area and tan δ were found to be relatively constant with slightly higher tan δ at lower frequencies. According to Gehman (*16*), tan δ should be in the range of 0.1 to 0.2. Dodge and Clark (*17*) also reported that tan δ should decrease slightly with increasing frequency. Our results seem to be consistent with the past research and theory. The information indicates that, for load range tested, the *energy loss per cycle* does not vary noticeably with frequency. Higher frequency means greater number of loading cycles per unit time and therefore the *energy loss per unit time* becomes linearly proportional to the frequency. Since fatigue life of carcass composites is linearly proportional to the inverse of frequency, it can be postulated that the amount of energy loss per unit time i.e. the rate of energy loss determines the fatigue life of composites.

The facts discussed so far strongly suggest that hysteretic heating and resultant thermal degradation of materials may be dominating factors in determining the fatigue life of aircraft tire carcass composites. However, it should be noted that a much less clear correlation exists between the specimen temperature and fatigue life (Figure 6) compared with the case between the frequency and fatigue life (Figure 7). Presumably the specimen surface temperature, which this study relies on, cannot describe such critical parameters as internal heating particularly at the point of crack initiation or true heat dissipation rate. In view of this limitation, our future study will take more rigorous approaches in measuring heat generation as well as dynamic creep under cyclic loading. Through-thickness distribution of temperature rather than nominal surface temperature will be obtained by embedding thermocouples at various locations of the specimens. Based on more accurate measurements of local strain and internal temperature, dynamic creep rate and heat dissipation rate will be redefined and correlated with fatigue life of tire carcass composites in the form of power law. By examining the frequency dependence of these power law factors, the contributions of thermal fatigue will be assessed in determining the lifetime of composites.

V. CONCLUDING REMARKS

Our research effort has been concentrated in defining the deformation and fracture mechanisms of angle-plied cord-rubber composite specimens which simulate the material elements of bias aircraft tire carcass in the shoulder area. The initial goal of the study is to identify the stress, strain or temperature parameters that control the process of fatigue damage accumulation for tire carcass composites under various loading conditions. Eventually, it is hoped a fatigue law can be formulated that will serve to predict the life of the tire carcass on a more analytical basis and thereby to complement empirical nature of dynamometer tests. Current phase of the study was undertaken to examine tensile fatigue behavior of carcass composites under various cyclic frequencies up to the level which closely simulates loading during high-speed take-off of aircraft. The study has also assessed the effectiveness of the measurement of strain or temperature in monitoring of fatigue damage accumulation under high frequency.

The modes of deformation and failure of composites were not affected by the use of higher frequency. As observed in our past study, angle-plied carcass composites exhibited a high level of interply shear deformation which induces fiber-matrix debonding. Accompanied by the increase of cyclic strain (i.e. dynamic creep) and heat build-up, the cord-matrix debonding was progressively worsened and developed into matrix cracking and delamination eventually leading to gross failure of the composites. The use of higher frequency, however, was found to affect strain response and heat build-up characteristics of composites at a given loading condition significantly. The lower level of *initial strain* observed at higher frequency stems clearly from strain rate dependence of deformation of rubber matrix composites. This trend is opposed by the effect of hysteretic heating as soon as cyclic loading starts. As a result of hysteretic heating, the surface temperatures of composite specimens over their fatigue lives increased with higher frequency.

The temperature profile of the specimens subjected from 20 to 30 Hz loading showed that hysteretic heating may lead to thermal fatigue failure as well as chemical degradation. Hysteretic heating is expected to influence both fiber-matrix adhesion strength and matrix strength. Regardless of its exact nature, the involvement of material degradation process was indicated by the fact that gross failure of composites requires lower value of *dynamic creep* when the frequency is increased. At the same time, the use of higher frequency resulted in exponentially higher *dynamic creep rate* which reflects increasingly more rapid accumulation of damage in carcass composites. Fatigue lifetime of composites at a given

loading condition was found to be inversely proportional to the frequency and linearly proportional to the time per unit cycle. This fact suggests that the use of higher frequency results in the decrease of fatigue life of carcass composites by simply shortening the time to reach a preset level of dynamic creep for gross failure.

Our study revealed that the specimen surface temperature may not describe such critical parameters as internal heating particularly at the point of crack initiation or true heat dissipation rate. In view of this limitation, the amount of mechanical energy loss was directly determined from hysteresis loop for each composite specimen during a particular cycle. For load range tested, the *energy loss per cycle* was found to be nearly constant and independent of the frequency. As a result, the *energy loss per unit time* becomes linearly proportional to the frequency. Since fatigue life of carcass composites is linearly proportional to the inverse of frequency, it was postulated that the amount of energy loss per unit time i.e. the rate of energy loss determines the fatigue life of composites. Our future study plans to take more rigorous approaches in measuring heat generation as well as dynamic creep under cyclic loading.

VI. RECOMMENDATIONS

The following recommendations can be made for our continuing study on this subject:

(1) Continue to examine the dependence of fatigue lifetime of aircraft tire carcass composites on *stress*, *strain* and *temperature history*. Define respective roles of *stress amplitude* and *minimum cyclic stress* and *frequency* in controlling heat generation and a critical level of dynamic creep at gross failure.

(2) Assess the contribution of viscoelastic properties of constituent materials to the dynamic creep process of composites by performing static *creep* loading experiments at elevated temperatures.

(3) Assess the mechanisms of *thermal fatigue* in determining the lifetime of carcass composites under high-frequency cyclic loading. Determine how the process of damage accumulation interacts with material property changes by measuring residual strength or stiffness of composites at various points of fatigue life.

(4) Derive empirical criteria for the prediction of the fatigue lifetime of cord-rubber composites under the various combinations of stress amplitude and mean stress.

(5) Confirm that the extent of cumulative damage is independent of *load sequence* under high-frequency cyclic loading and *frequency sequence*. Establish the damage models for the prediction of fatigue life under random spectrum loading.

(6) Establish the measurement of local strain change, heat generation or acoustic emission (AE) as a viable experimental technique for *real-time monitoring* of the damage accumulation process. Correlate AE energy release rate with the corresponding strain energy release rate.

(7) Assess the failure modes of internally-pressurized composite tube under *cyclic tension* and/or *out-of-plane bending* mode to induce localized delamination and to avoid straining of the reinforcing cords as a dominant mode of deformation.

(8) Develop test methodologies for the laboratory simulation of other types of failures processes besides shoulder delamination of aircraft tires. Other common types of failure processes include *bead* and *lower sidewall area failure.*

(9) Examine the effects of footprint load, inflation pressure and speed on the mileage to failure, deflection and temperature history of aircraft tires and assess the failure modes of tires based on fractography. Correlate these results with the fatigue resistance data of cord-rubber composites.

ACKNOWLEDGEMENTS

We wish to thank the Air Force Systems Command and the Air Force Office of Scientific Research for sponsorship of this research. The Research and Laboratories, Inc. helped us in all administrative aspects of this program. We sincerely appreciate continuing support and encouragement for this research work from Mr. John P. Medzorian, Dr. Mangal Chawla, Mr. Paul C. Ulrich, Mr. Aivars V. Petersons and Dr. Arnold H. Mayer at the Vehicle Subsystems Division of Wright Laboratory Flight Dynamics Directorate. Mr. Douglas Wolfe of SRL Inc. played an essential role in this program by performing various mechanical testing. Finally our sincere thanks should go to Dr. Alfredo G. Causa, Dr. Yao M. Huang and Mr. Ronald Burgan at The Goodyear Tire & Rubber Co. for providing composite specimens.

REFERENCES

(*1*) B. L. Lee, J. P. Medzorian, P. M. Fourspring, G. J. Migut, M. H. Champion, P. M. Wagner and P. C. Ulrich, "Study of Fracture Behavior of Cord-Rubber composites for Lab Prediction of Aircraft Tire Durability", *SAE Int'l*, Aerospace Technology Conference, Paper #901907, Long Beach, CA (1990).

(*2*) B. L. Lee, J. P. Medzorian, P. K. Hippo, D. S. Liu and P. C. Ulrich, "Fatigue Lifetime Prediction of Angle-Plied Fiber-Reinforced Elastomer Composites as Pneumatic Tire Materials", *ASTM Committee E-9*, Second Symposium on Advances in Fatigue Lifetime Predictive Techniques, Pittsburgh, PA (1992).

(*3*) D. S. Liu and B. L. Lee, "Cumulative Fatigue Damage of Cord-Rubber Composites", *The Tire Society*, Eleventh Annual Meeting, Akron, OH (1992).

(*4*) Personal communications with the researchers in tire industry.

(*5*) S. N. Bobo, "Fatigue Life of Aircraft Tires", *Tire Science and Technology*, Vol. 16, No. 4, p.208 (1988).

(*6*) S. K. Clark, "Loss of Adhesion of Cord-Rubber Composites in Aircraft Tires", *Tire Science and Technology*, Vol. 14, No. 1, p.33 (1986).

(*7*) A. Y. C. Lou and J. D. Walter, "Interlaminar Shear Strain Measurements in Cord-Rubber Composites", *Experimental Mechanics*, Vol. 18, 457 (1978).

(*8*) J. D. Walter, "Cord-Reinforced Rubber" in Mechanics of Pneumatic Tires edited by S. K. Clark, U.S. Department of Transportation, Washington D.C. (1982).

(*9*) D. O. Stalnaker, R. H. Kennedy and J. L. Ford, "Interlaminar Shear Strain in a Two-Ply Balanced Cord-Rubber Composites", *Experimental Mechanics*, Vol. 20, p.87 (1980).

(*10*) J. L. Ford, H. P. Patel and J. L. Turner, "Interlaminar Shear Effects in Cord-Rubber Composites", *Fiber Science and Technology*, Vol. 17, p.255 (1982).

(*11*) H. Rothert, B. Nguyen and R. Gall, "Comparative Study on the Incorporation of Composite Material for Tyre Computation", Composite Structures (Proc. of 2nd Int'l Conf. on Composite Structures, Paisley College of Technology, Scotland, UK), p.549, Applied Sci. Publ. (1983).

(*12*) R. J. Cembrola and T. J. Dudek, "Cord/Rubber Material Properties", *Rubber Chemistry and Technology*, Vol. 58, p.830 (1985).

(*13*) Interlaminar Response of Composite Materials edited by N. J. Pagano, Composite Materials Series Vol. 5, Elsevier Science Publ. Co., New York, NY (1989).

(*14*) R. F. Breidenbach and G. J. Lake, "Mechanics of Fracture in Two-Ply Laminates", *Rubber Chemistry and Technology*, Vol. 52, p.96 (1979).

(*15*) R. F. Breidenbach and G. J. Lake, "Application of Fracture Mechanics to Rubber Articles Including Tyres", *Philosophical Trans. Royal Soc. London*, Vol. A299, p.189 (1981).

(*16*) S. D. Gehman, "Rubber Structures and Properties" in Mechanics of Pneumatic Tires edited by S. K. Clark, U.S. Department of Transportation, Washington D.C. (1982).

(*17*) R. N. Dodge and S. K. Clark, "Properties of Aircraft Tire Materials", *SAE Int'l*, Aerospace Technology Conference, Paper #881358, Anaheim, CA (1988).

TABLE 1

Specifications of Tire Carcass Composites

Reinf Composition	1260/2 nylon cord
Reinf Modulus	2.07 GPa
Reinf Angle	-38, +38, +38, -38 deg
Reinf Radius	0.33 mm
Reinf End Count	11 ends per cm width
Matrix Composition	proprietary
Matrix Modulus	5.51 MPa
Specimen Width	19.05 mm
Specimen Thickness	5.08 mm
Gage Length	101.6 mm

TABLE 2

Frequency Dependence of Mechanical Loss
During Fatigue of Tire Carcass Composites

		Hysteresis Loop Area (N-m)			
Freq	@	25%	50%	75%	99% Fatigue Life
2 Hz		.462	.473	.529	.592
5		.438	.454	.458	.620
10		.467	.466	.461	.507
20		.454	.442	.457	.516
30		.553	.598	.584	.544

TABLE 2 (Cont'd)

Frequency Dependence of Mechanical Loss
During Fatigue of Tire Carcass Composites

Freq	@	Loss Tangent tanδ 25%	50%	75%	99% Nf
2 Hz		.128	.128	.123	.141
5		.119	.120	.121	.118
10		.117	.117	.113	.109
20		.117	.112	.114	.114
30		.118	.120	.120	.104

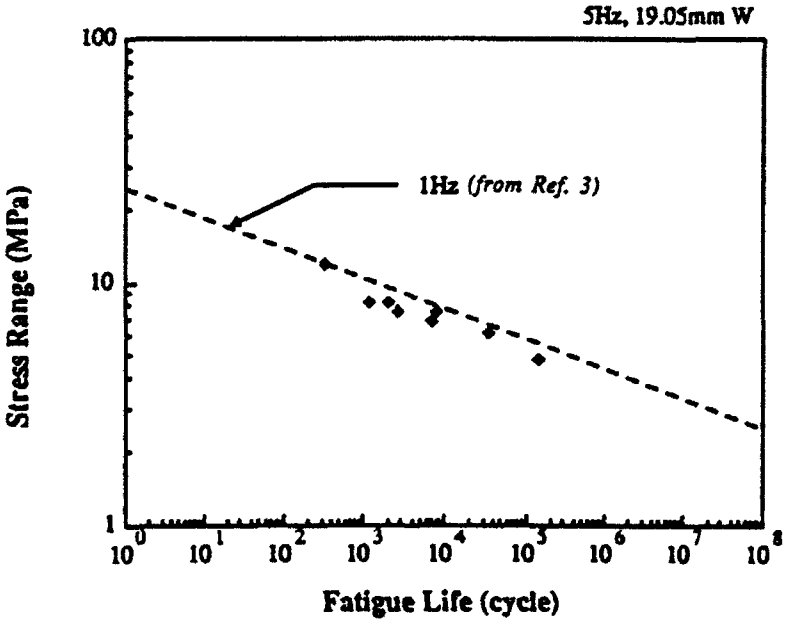


Figure 1

Stress Range vs Fatigue Life (S-N) Curve for Cord-Rubber Composites
(Frequency = 5 Hz)

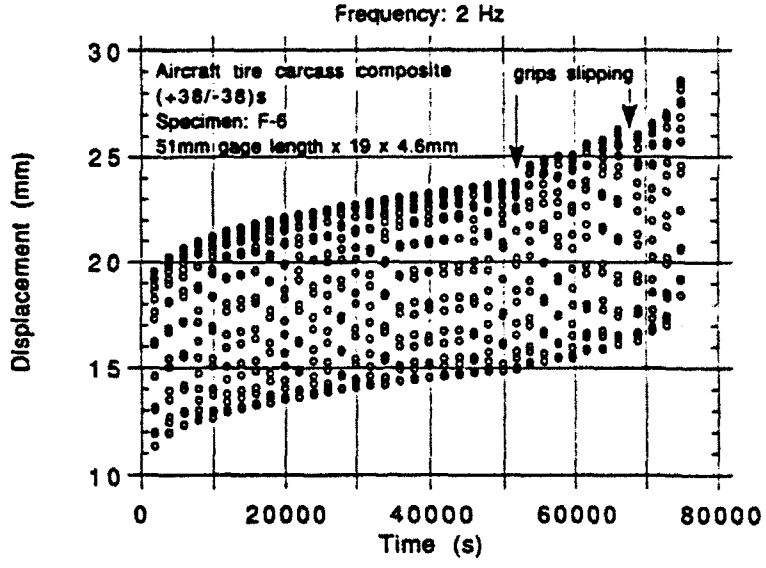


Figure 2

Displacement vs Time under Cyclic Loading
(6.1 MPa Stress Range, 1.5 MPa Minimum Stress, **2 Hz**)

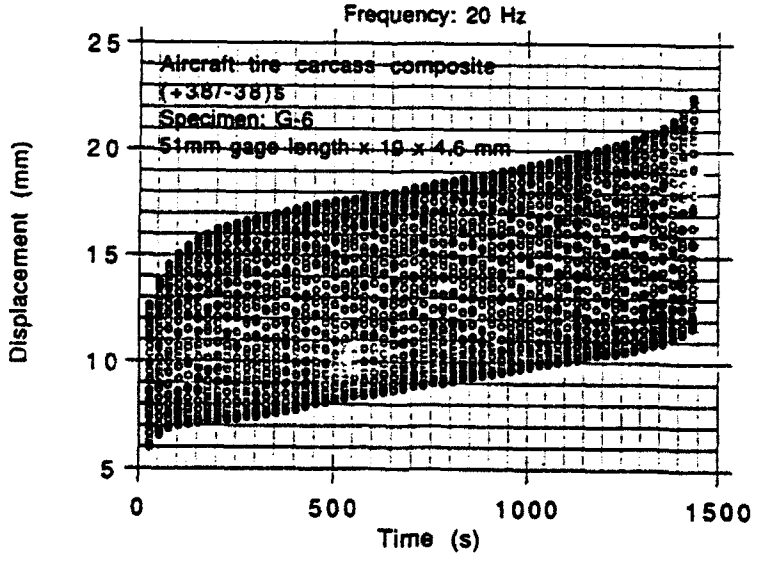


Figure 3

Displacement vs Time under Cyclic Loading
(6.1 MPa Stress Range, 1.5 MPa Minimum Stress, **20 Hz**)

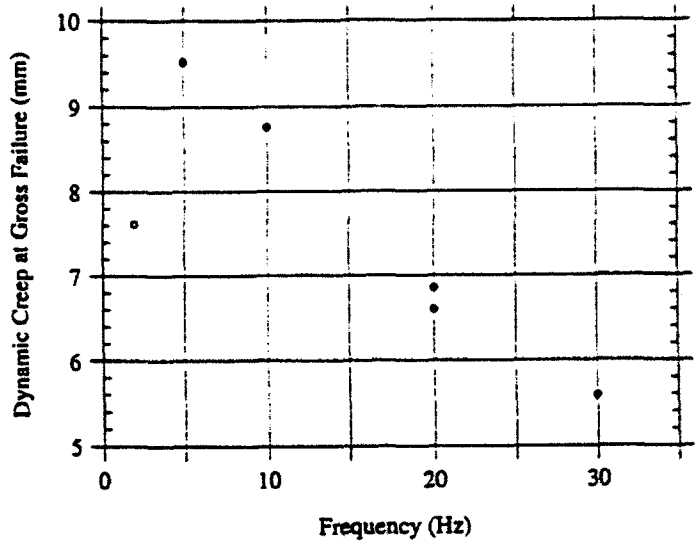


Figure 4

Dynamic Creep at Gross Failure vs Frequency
(6.1 MPa Stress Range, 1.5 MPa Minimum Stress)

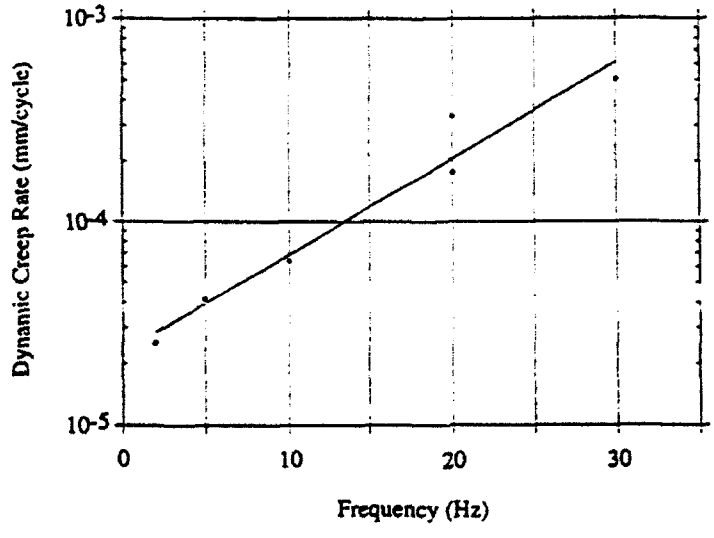


Figure 5

Dynamic Creep Rate vs Frequency
(6.1 MPa Stress Range, 1.5 MPa Minimum Stress)

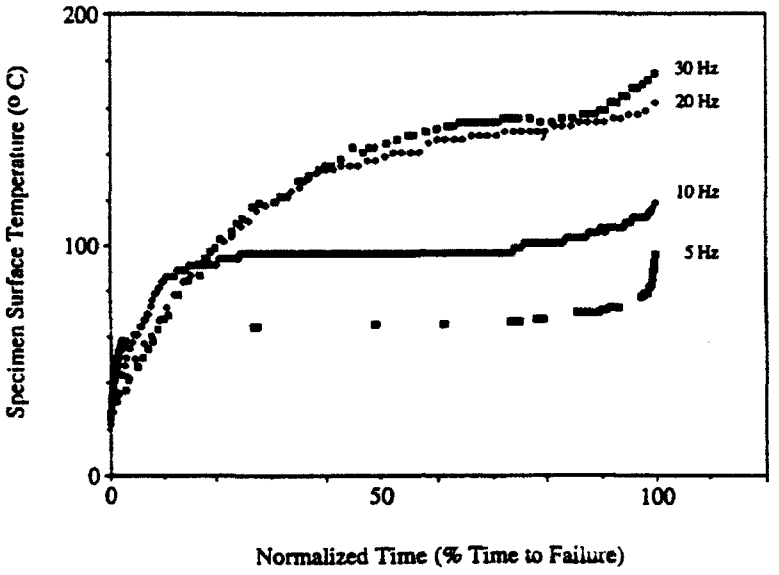


Figure 6

Specimen Surface Temperature vs Normalized Time under Cyclic Loading
(6.1 MPa Stress Range, 1.5 MPa Minimum Stress)

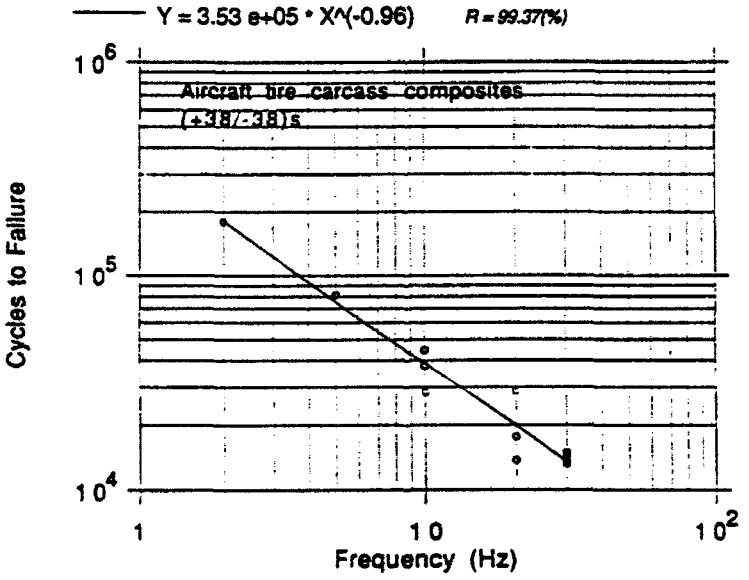


Figure 7

Cycles to Failure vs Frequency
(6.1 MPa Stress Range, 1.5 MPa Minimum Stress)

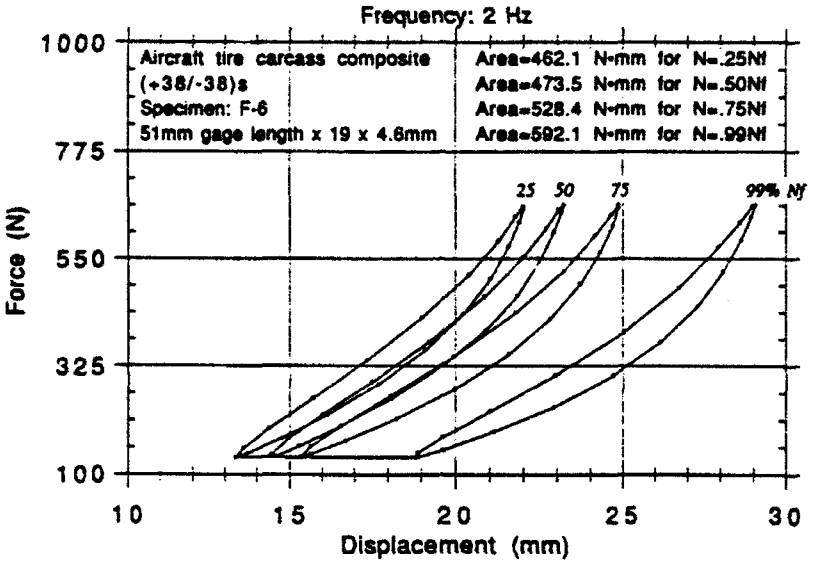


Figure 8

Hysteresis Curves at 25, 50, 75 and 99 % of Fatigue Life
(6.1 MPa Stress Range, 1.5 MPa Minimum Stress, 2 Hz)

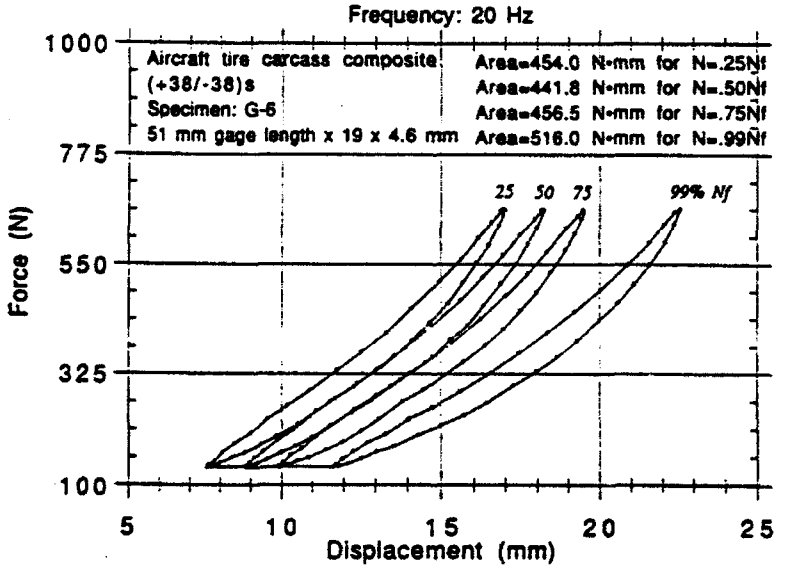


Figure 9

Hysteresis Curves at 25, 50, 75 and 99 % of Fatigue Life
(6.1 MPa Stress Range, 1.5 MPa Minimum Stress, 20 Hz)

A PHYSICS-BASED HETEROJUNCTION BIPOLAR TRANSISTOR MODEL INCLUDING HIGH-CURRENT AND THERMAL EFFECTS

Juin J. Liou
Associate Professor
Department of Electrical & Computer Engineering

University of Central Florida, Orlando, FL 32816

Final Report for:
Summer Research Program
Wright Laboratory

Sponsored by:
Air Force Office of Scientific Research
Wright-Patterson Air Force Base, Ohio

August 1992

A PHYSICS-BASED HETEROJUNCTION BIPOLAR TRANSISTOR MODEL INCLUDING HIGH-CURRENT AND THERMAL EFFECTS

Juin J. Liou
Associate Professor
Department of Electrical & Computer Engineering
University of Central Florida, Orlando, FL 32816

Abstract

We present a detailed, analytical model to predict the d.c. and high-frequency performance of AlGaAs/GaAs heterojunction bipolar transistors (HBTs). The model is developed based on the device configuration and relevant device physics such as current-induced base pushout and thermal effects associated with high collector current density in the HBT. The characteristics of the current gain, cutoff frequency, and maximum frequency versus the collector current density, which is a function of the applied voltage as well as the corresponding temperature in the HBT, are calculated. Our results suggest that the conventional HBT model, which assumes the HBT temperature is the same as that of the ambient, can overestimate the current gain, cutoff frequency, and maximum frequency considerably when the collector current density is high. Furthermore, we have shown that the experimentally observed HBT high-current behavior, like the rapid fall-off of the current gain and cutoff frequency, can be accurately predicted if thermal effect is properly accounted for in the analysis. The model predictions compare favorably with the results obtained from a model which solves numerically the Poisson and continuity equations coupled with the heat transfer mechanism.

A PHYSICS-BASED HETEROJUNCTION BIPOLAR TRANSISTOR MODEL INCLUDING HIGH-CURRENT AND THERMAL EFFECTS

Juin J. Liou

INTRODUCTION

The down scaling of heterojunction bipolar transistor (HBT) structure has increased the collector current density to the range of 10^5 to 10^6 A/cm^2 under normal bias conditions. This high-current handling capability makes HBTs very attractive to high-power, high-frequency microwave amplifier applications [1]. Recently, output power up to 12.5 W cw have been demonstrated from a monolithic, 2-stage HBT power amplifier [2]. Such a high power level can inevitably generate a large amount of heat in the HBT and therefore result in a much higher temperature in the HBT than that of the ambient [3]. Since all the physical properties of the HBT are strongly influenced by the thermal voltage $V_T = KT/q$ (K is the Boltzmann constant and T is the absolute temperature in the device), the performance of HBTs used in power amplifiers will be affected significantly by the thermal effect.

This report presents a physics-based, analytical HBT model including high-current and thermal effects. The model can predict three figures of merit commonly used in analyzing the performance of an HBT: the d.c. current gain, cutoff frequency, and maximum frequency. The required parameters are the device configuration (eg., layer thickness and doping concentration) and material parameters (eg., free-carrier mobility). The model predictions will be compared with the results simulated from a recently developed numerical program [3] which solves the Poisson and continuity equations coupled with the heat transfer mechanism. Like its numerical counterpart, the analytical model developed could provide HBT designers the physical trends into the thermal effect on the HBT performance. The analytical model, however, is more suitable for circuit simulations because of its simplicity and flexibility in implementation. Furthermore, the inclusion of thermal effect in device simulation often suffers numerical instability, and such a problem can be easily averted in our analytical solutions.

MODEL DEVELOPMENT

Throughout the analysis, we will consider an N/p$^+$/n AlGaAs/GaAs/GaAs single HBT with graded emitter-base junction (Fig. 1). We first model the collector current density J_c. Since the emitter-base junction is graded and the conduction band spike is removed, the conventional drift-diffusion model is used here to describe the free-carrier charge transport across the junction:

$$J_c = qD_n\Delta n(0)/(X_B + \Delta X_B), \tag{1}$$

where D_n is the electron diffusion coefficient, X_B is the metallurgical base thickness, ΔX_B is the current-induced base pushout, and $\Delta n(0)$ is the excess minority free-carrier concentration at the emitter junction [4]:

$$D_n = V_T[7200/(1 + 5.5\times10^{-17}N_B)^{0.233}](300/T)^{2.3}, \tag{2}$$

$$\Delta n(0) = n_{0,E}\exp[(V_{bi,BE} - V_{j,BE})/V_T], \tag{3}$$

Here $V_T = kT/q = 0.026(T/300)$ is the thermal voltage, $n_{0,E}$ is the equilibrium electron concentration in the emitter, $V_{j,BE}$ is the emitter-base junction voltage, and $V_{bi,BE}$ is the base-emitter junction built-in potential.

The onset collector current density J_0 for base pushout can be derived from the Poisson equation using the boundary condition of the electric field at the base-collector metallurgical junction being zero [5]:

$$J_0 = qv_s[n_{0,C} + 2\varepsilon_C(V_{bi,BC} - V_{j,BC})/qX_C^2], \tag{4}$$

where v_s is the drift saturation velocity, ε_C is the dielectric permittivity in the collector, $V_{bi,BC}$ is the base-collector junction built-in potential, $V_{j,BC}$ is the base-collector junction voltage, and X_C is the collector thickness.

The current-induced base pushout is given by [6]

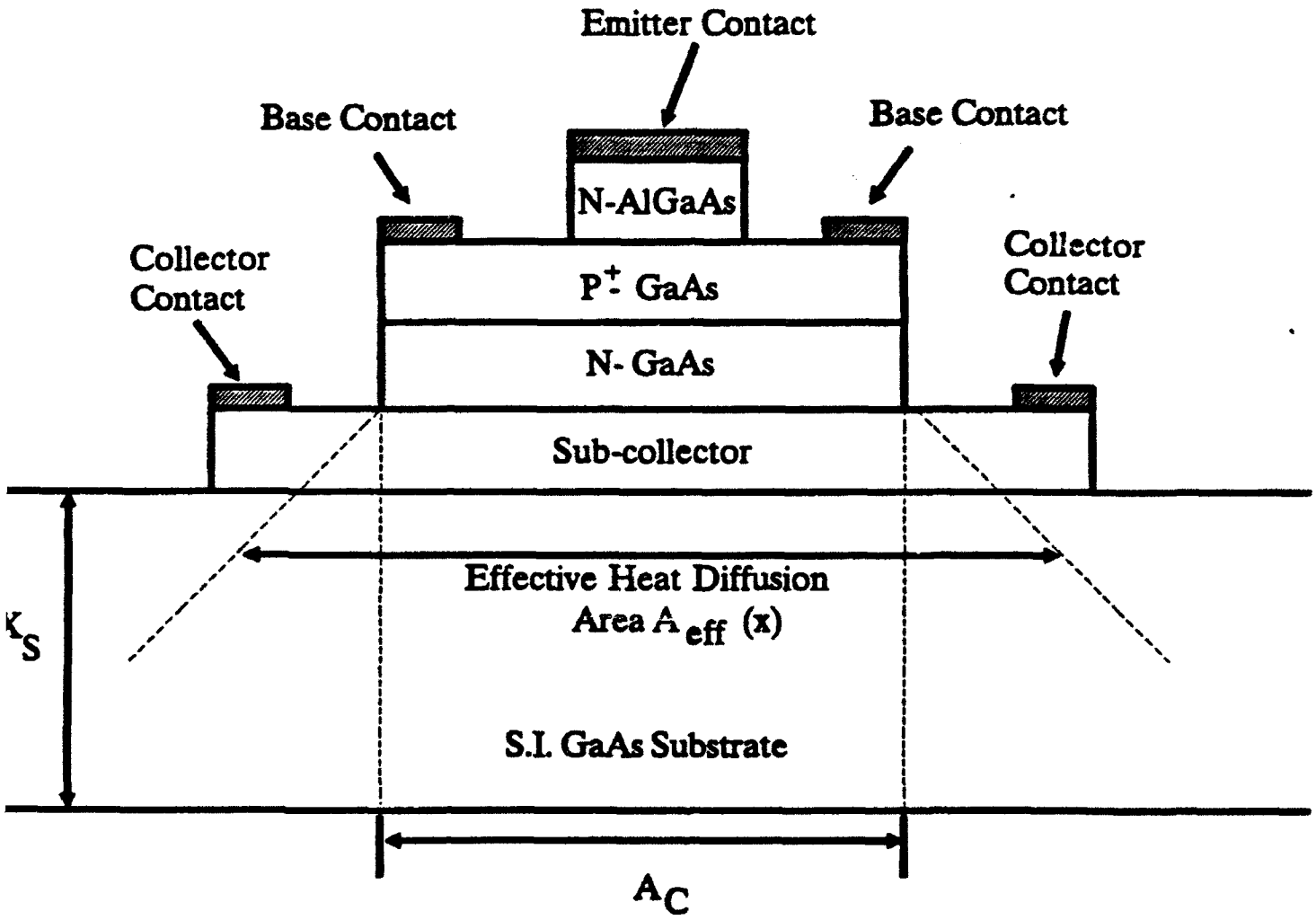


Fig. 1 HBT device structure illustrating the effective area through which the heat generated in the intrinsic HBT is dissipated.

$$\Delta X_n = W_c\{1 - [(J_0 - qv_s n_{0,c})/(J_c - qv_s n_{0,c})]^{0.5}\}, \quad (5)$$

for $J_c > J_0$, and $\Delta X_n = 0$ otherwise.

The heat P_n (W) generated in the HBT is

$$P_n = J_c V_{ce} A_e, \quad (6)$$

where A_e is the emitter area and $V_{ce} = V_{be} - V_{bc}$ is the collector-emitter applied voltage. Since the size of the intrinsic HBT is much smaller compared to that of the extrinsic HBT, the heat generated in the HBT is primarily dissipated through the S.I. GaAs substrate. Thus P_n is related to the thermal resistance R_{th} of the substrate as

$$T - T_0 = P_n R_{th}, \quad (7)$$

where $T_0 = 300$ K is the ambient temperature. Assuming the heat dissipated throughout the S.I. substrate with a lateral diffusion angle θ (Fig. 1) yields [7]

$$R_{th} = 1/K_s \int_0^{X_s} dx/A_{eff}(x) = 1/K_s \int_0^{X_s} dx/[A_C + 2Z\tan(\theta)x], \quad (8)$$

where $K_s = 0.47$ W/K-cm is the GaAs thermal conductivity, Z is the HBT width, A_c is the collector area, and X_s is the thickness of the S.I. substrate.

The base current density J_B of the AlGaAs/GaAs HBT can be written as

$$J_B = J^* \exp(V_{j,be}/V_T) + J^{**} \exp(V_{j,be}/2V_T) + qD_p \Delta p(0)/X_E, \quad (9)$$

where D_p is the hole diffusion coefficient in the emitter, and Δp is the excess minority free-carrier concentration at the emitter junction edge [4]:

$$D_p = V_T[380/(1 + 3.2\times10^{-17}N_E)^{0.266}](300/T)^{2.7}, \quad (10)$$

$$\Delta p(0) = p_{0,B}\exp[(V_{bi,BE} + \Delta V - V_{j,BE})/V_T], \quad (11)$$

where ΔV is the valence band discontinuity at the emitter-base hetero-interface. In (9), J^* and J^{**} are the pre-exponential constants of the kT-like and 2kT-like recombination current densities, respectively.

Finally the common-emitter d.c. current gain β is defined as $\beta = J_C/J_B$.

Two figures of merit for the HBT microwave performance are the cutoff frequency f_T and maximum frequency f_{max}. The cutoff frequency is related to four different delay times [8]

$$f_T = 1/[2\pi(\tau_E + \tau_{BT} + \tau_{CT} + \tau_C)], \quad (12)$$

where τ_E, τ_{BT}, τ_{CT}, τ_C are the emitter charging time, base transit time, collector transit time, and collector charging time, respectively. The emitter charge time is

$$\tau_E = r_E(C_{jE} + C_{jC}), \quad (13)$$

where r_E is the emitter resistance ($r_E = V_T/J_C + r_{EC}$, and r_{EC} is the emitter contact resistance) and C_{jE} and C_{jC} are the emitter-base and base-collector junction capacitances, respectively:

$$C_{jE} = \varepsilon_E/X_{jE} \quad \text{and} \quad C_{jC} = \varepsilon_C/X_{jC}, \quad (14)$$

where ε_E and ε_C are the emitter and collector dielectric permittivity, and X_{jE} and X_{jC} are the emitter-base and base-collector space-charge layer thicknesses.

The diffusion dominated base transit time, including the base pushout, is

$$\tau_{BT} = (X_B + \Delta X_B)^2/2D_n. \quad (15)$$

Assuming the free carriers in the base-collector space-charge layer travel with a drift saturation velocity v_s, we can model the collector transit time as

$$\tau_{CT} = X_{jc}/\gamma v_s, \tag{16}$$

where $\gamma = 2.5$ is used to account for the effect of velocity overshoot in the base-collector junction.

The collector charging time is

$$\tau_c = r_c C_{jc}, \tag{17}$$

where r_c is the collector resistance.

The maximum frequency of oscillation is given by [9]

$$f_{MAX} = 1/[8\pi (r_b C_{jc})_{eff}/f_\tau]^{0.5}, \tag{18}$$

where $(r_b C_{jc})_{eff}$ is the effective base-resistance collector-capacitance time constant.

ILLUSTRATIONS AND DISCUSSIONS

An N/p$^+$/n AlGaAs/GaAs HBT with the configuration listed in Table I under forward-active operation is considered. Figure. 2 plots the collector and base current densities versus the base-emitter applied voltage calculated from the present model which includes thermal effects (hereafter called present model) and the model which assumes a constant device temperature of 300 K (hereafter called isothermal model). Also shown are the results obtained from the numerical model reported in [3]. Good agreement is found between the present model and numerical simulation. The isothermal, present, and numerical models predict the same current densities when the applied voltage is relatively small, but large discrepancies arise when the collector current density is high. This is because the thermal effect becomes important when the collector current density is high,

TABLE 1
HBT DEVICE STRUCTURE USED IN CALCULATIONS

	Thickness (Å)	Type	Doping Density ($/cm^3$)	AlAs Fraction in $Al_xGa_{1-x}As$
Emitter	1700	n	$5X10^{17}$	0.3
Base	1000	p^+	$1X10^{19}$	0
Collector	3000	n	$5X10^{16}$	0

Emitter Area = 1x10 μm^2

Collector Area = 5x20 μm^2

S. I. Substrate Thickness = 100 μm

Lateral Heat Diffusion Angle (θ) = 60 degrees

Emitter Contact Resistance = 10^{-6} Ω-cm^2

Base Contact Resistance = 10^{-6} Ω-cm^2

Collector Contact Resistance = 10^{-6} Ω-cm^2

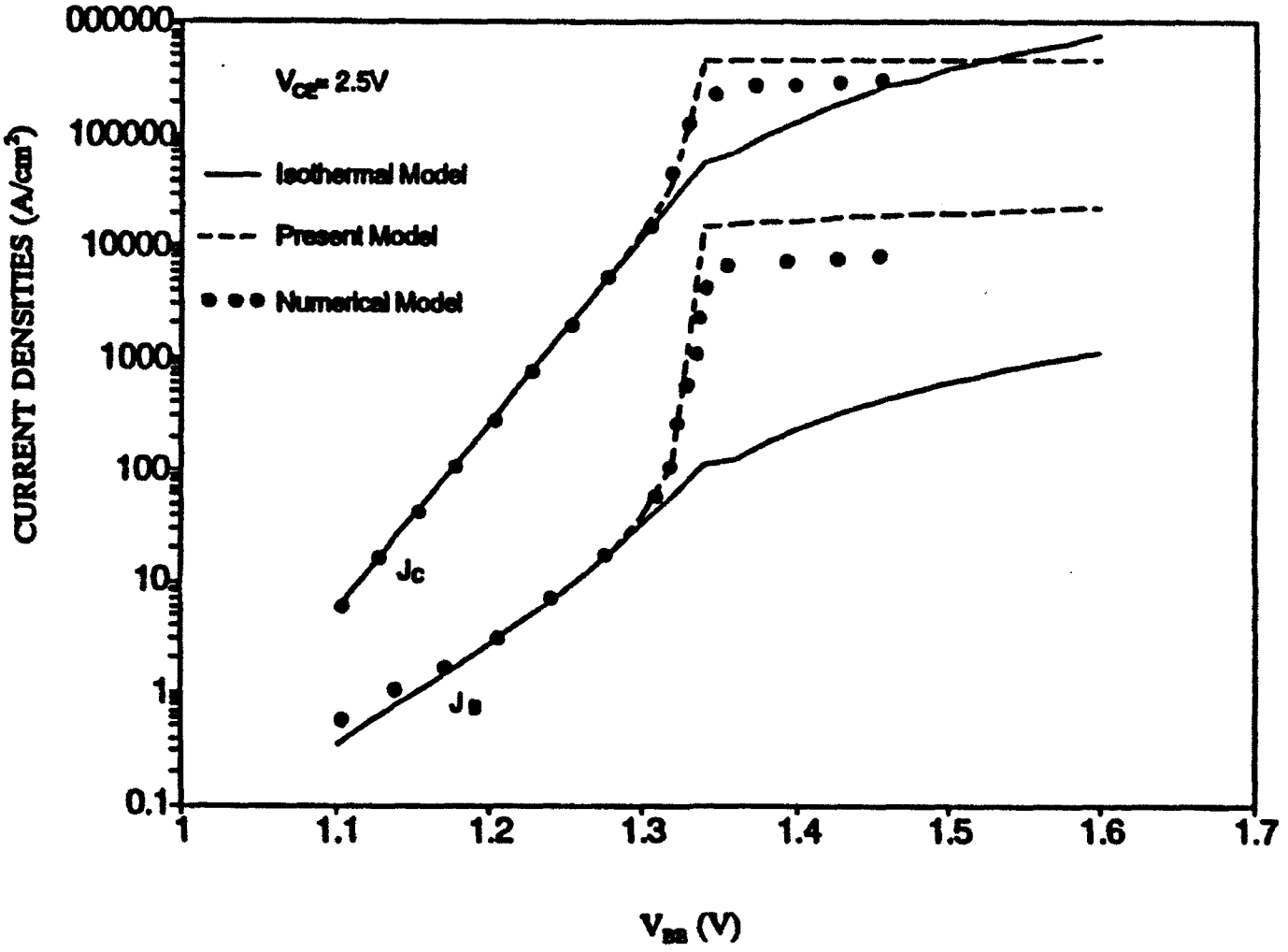


Fig. 2 Gummel plot obtained from the present model, isothermal model, and numerical simulation.

which generates a large amount of heat in the HBT and thus results in a much higher lattice temperature in the HBT than the ambient temperature. Note that when thermal effect is included, the collector and base current densities become constant at about 1.34 V (Fig. 2). The reason is that the emitter-base junction barrier height $V_{BT} = V_{bi,BE} - V_{j,BE}$ ($V_{bi,BE}$ is smaller as the temperature is increased) is flattened ($V_{BT} = 0$) at such high voltages, thus limiting the free-carrier injection in the base to be a constant equals to the emitter doping concentration. This does not occur in the isothermal model, however, because $V_{bi,BE}$ is always larger than $V_{j,BE}$ if the temperature is assumed to be 300 K. For the collector-emitter voltage considered, the temperature in the HBT increases rapidly as J_C is increased beyond 10^4 A/cm^2 and reaches 500 K when the device is operated at a maximum collector current density of about 3×10^5 A/cm^2. Figures 3 and 4 compare the current gains and cutoff frequencies as a function of J_C, respectively, predicted by the present model, isothermal model, and numerical simulation. The results for the maximum frequency are given in Fig. 5.

We next examine the effects of collector-emitter voltage V_{CE} on the HBT performance. The forward-active J_C-V_{CE} characteristics calculated from the present model compare favorably with those obtained from numerical simulations (Fig. 6). The I-V characteristics exhibit a negative slope when the base current density is high (or V_{BE} is large), a phenomenon commonly observed in AlGaAs/GaAS d.c. measurements. The dependencies of the current gain, cutoff frequency, and maximum frequency on V_{CE} are given in Figs. 7, 8, and 9, respectively. It is interesting to see that while both β and f_T are reduced monotonically by the increasing in V_{CE}, f_{MAX} actual increases slightly but falls off at a lower J_C when V_{CE} is increased.

The results clearly show that the commonly observed rapid fall-off behavior of current gain, cutoff frequency, and maximum frequency at high current density can be accurately predicted by the present model developed based on the simple drift-diffusion theory including the thermal effect. The rapid fall-off, which

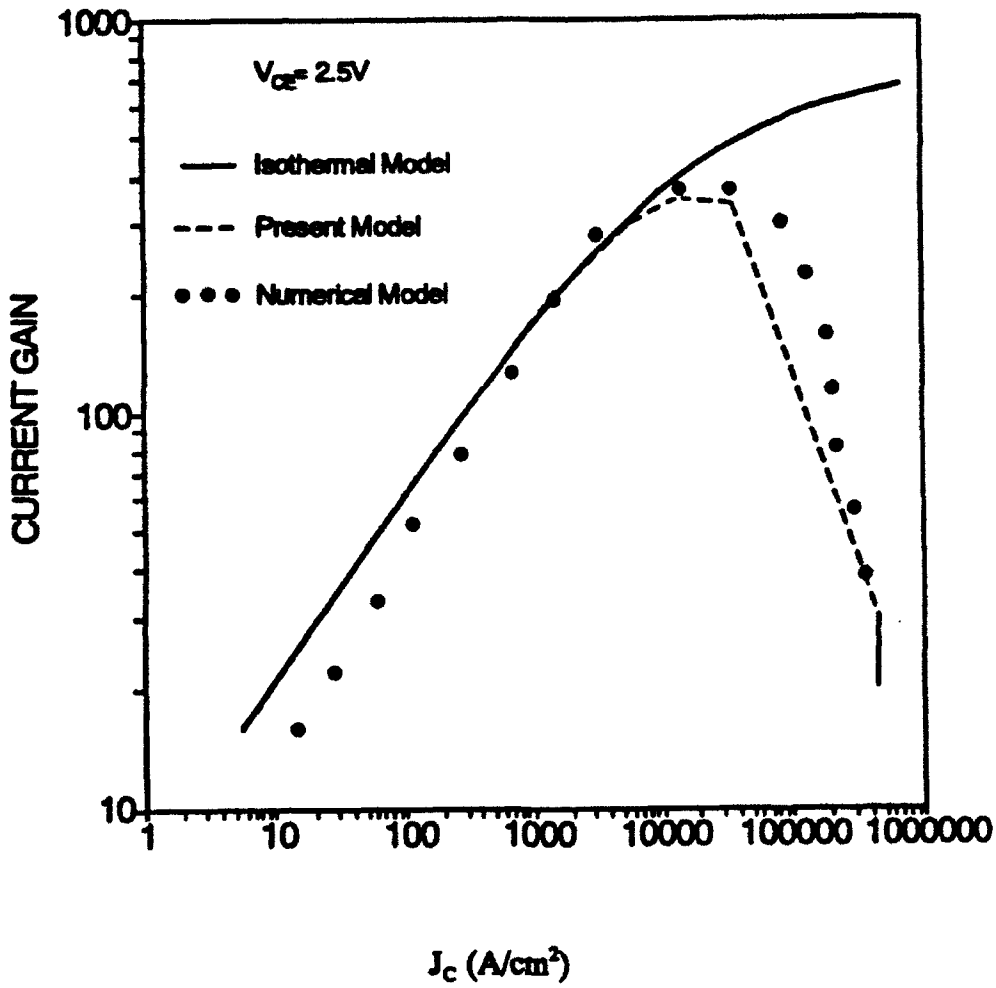


Fig. 3 Current gains calculated from the present model, isothermal model, and numerical model.

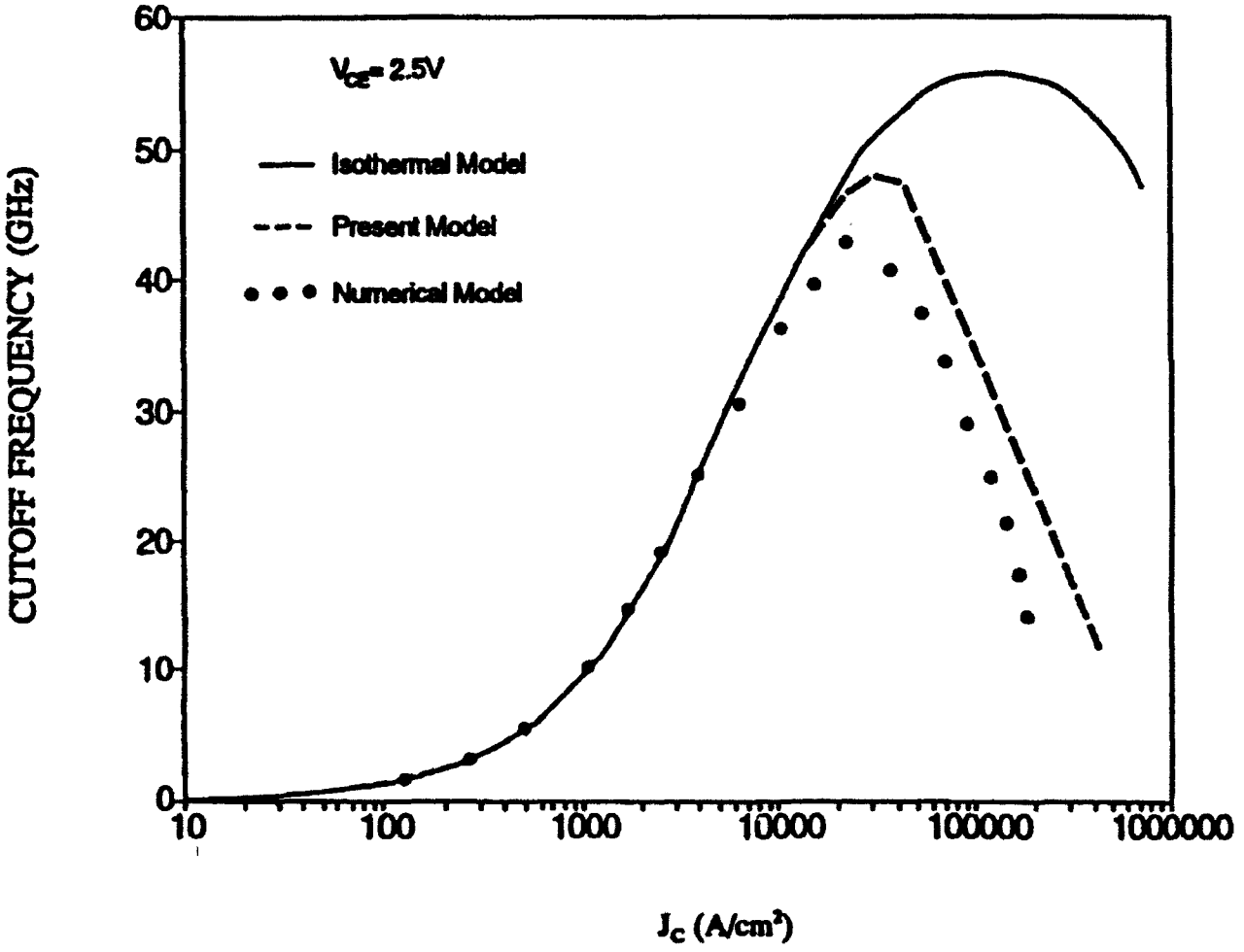


Fig. 4 Cutoff frequencies calculated from the present model, isothermal model, and numerical model.

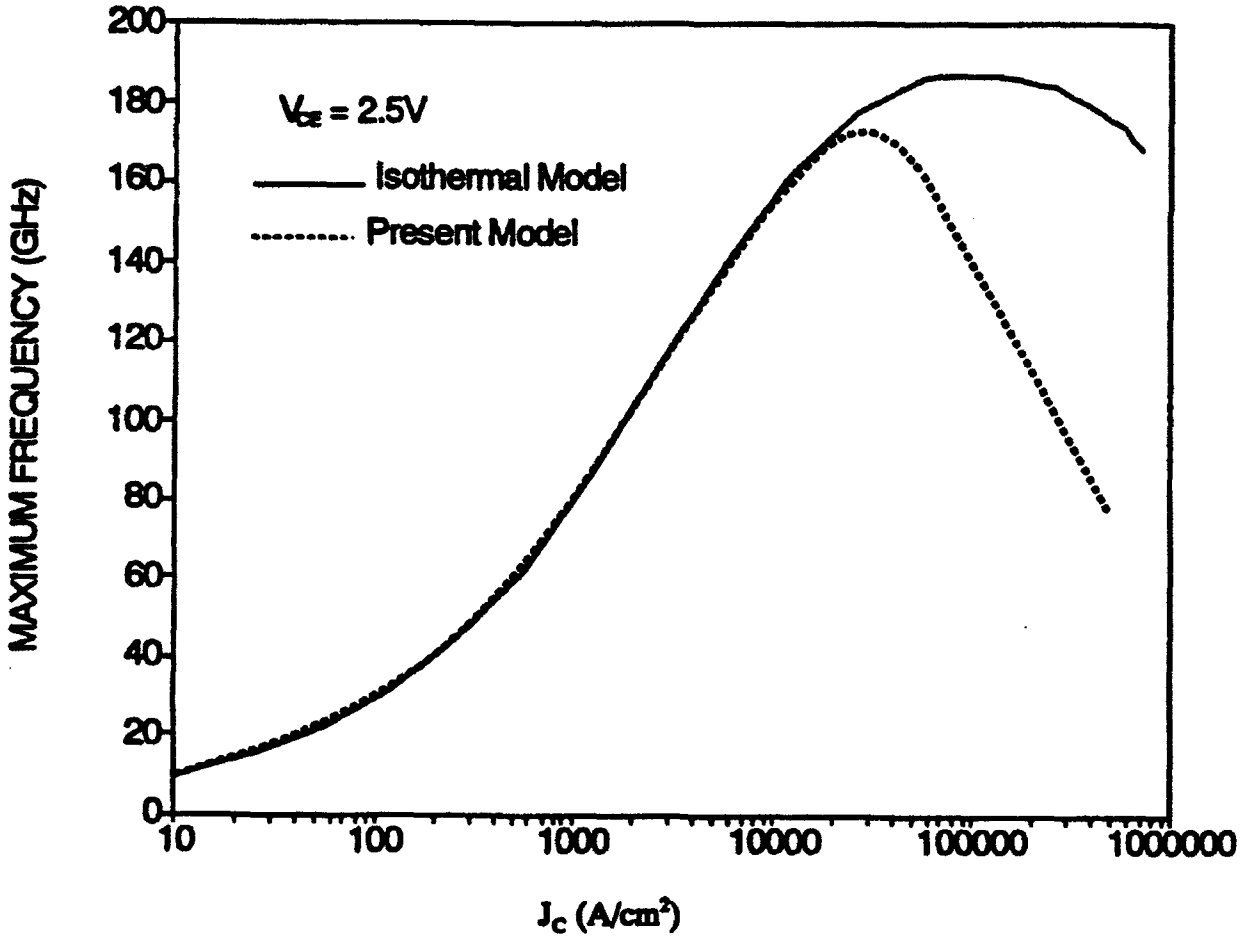


Fig. 5 Maximum frequencies calculated from the present model and isothermal model.

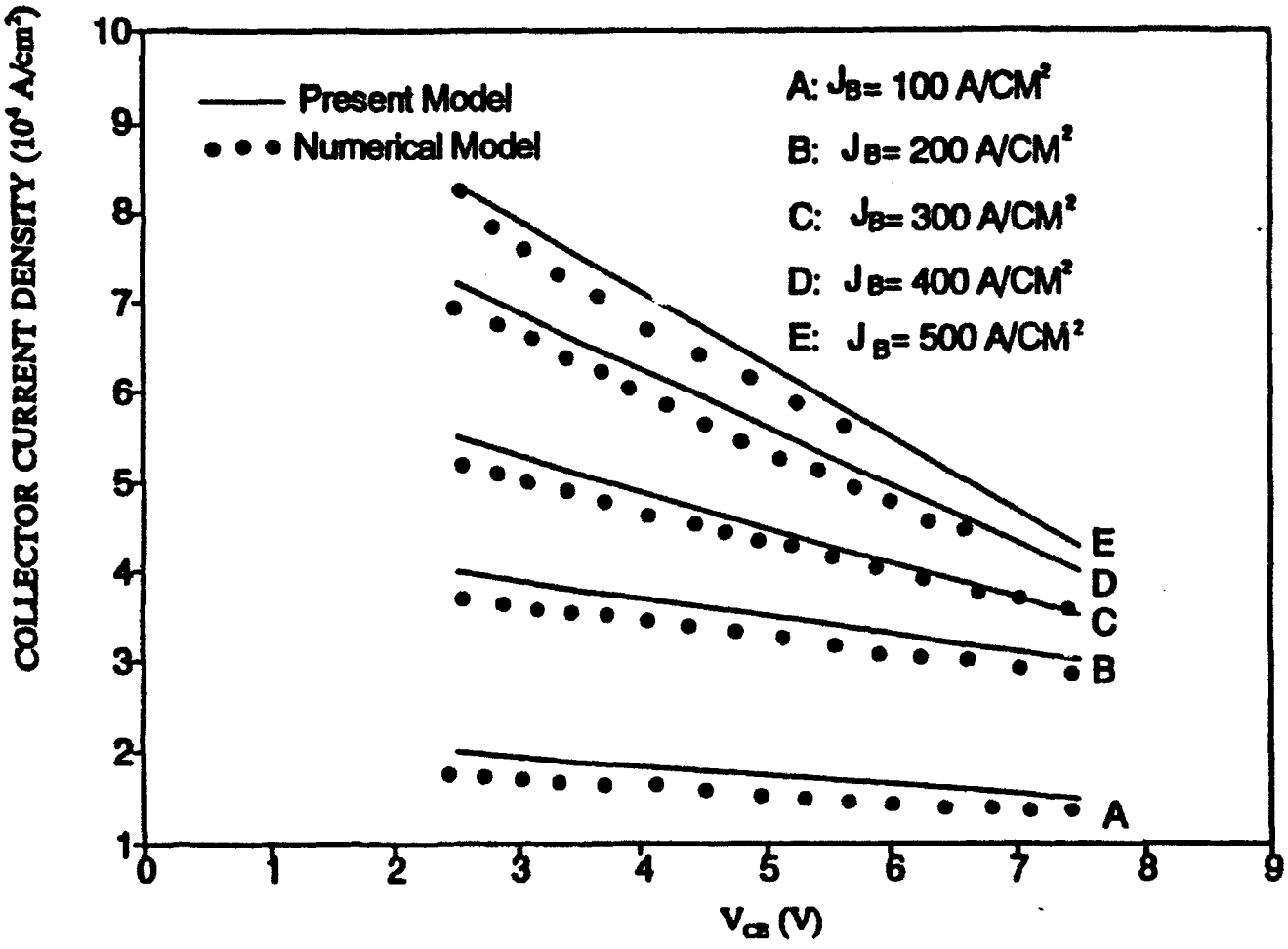


Fig. 6 Forward-active current-voltage characteristics as a function of the base current density.

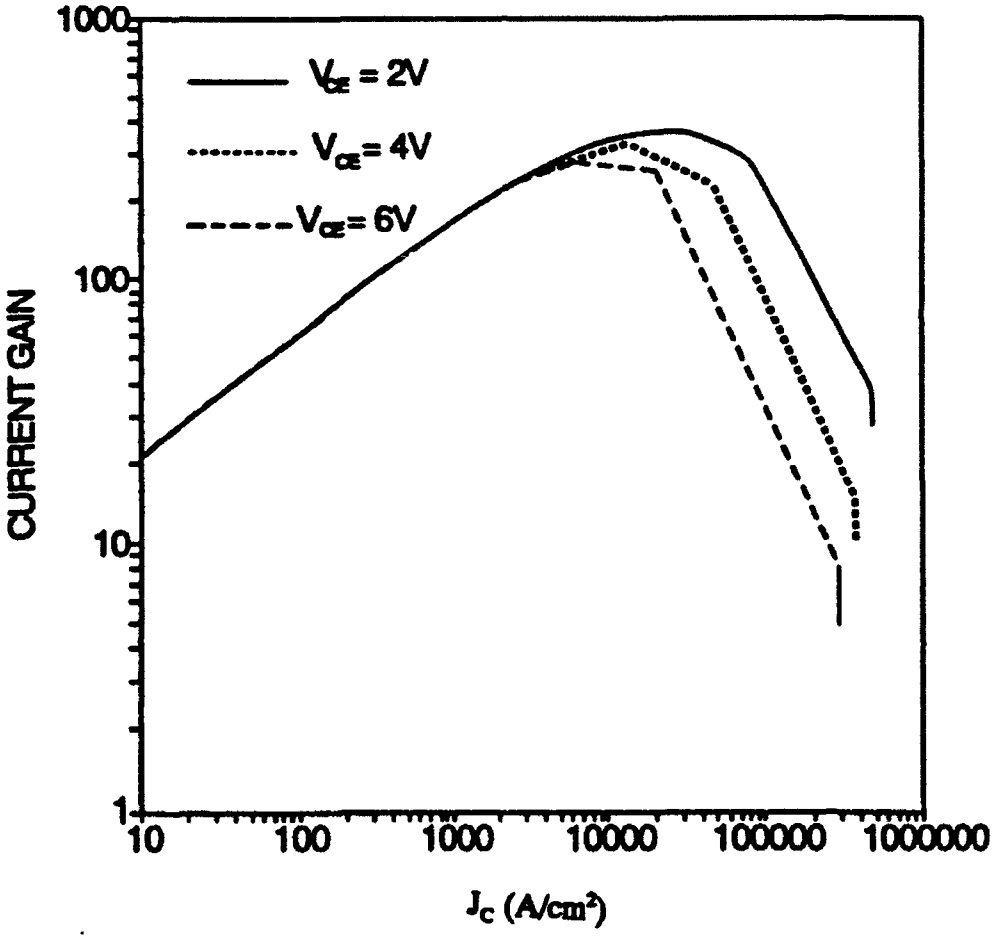


Fig. 7 Current gains calculated from the present model for three different V_{CE}.

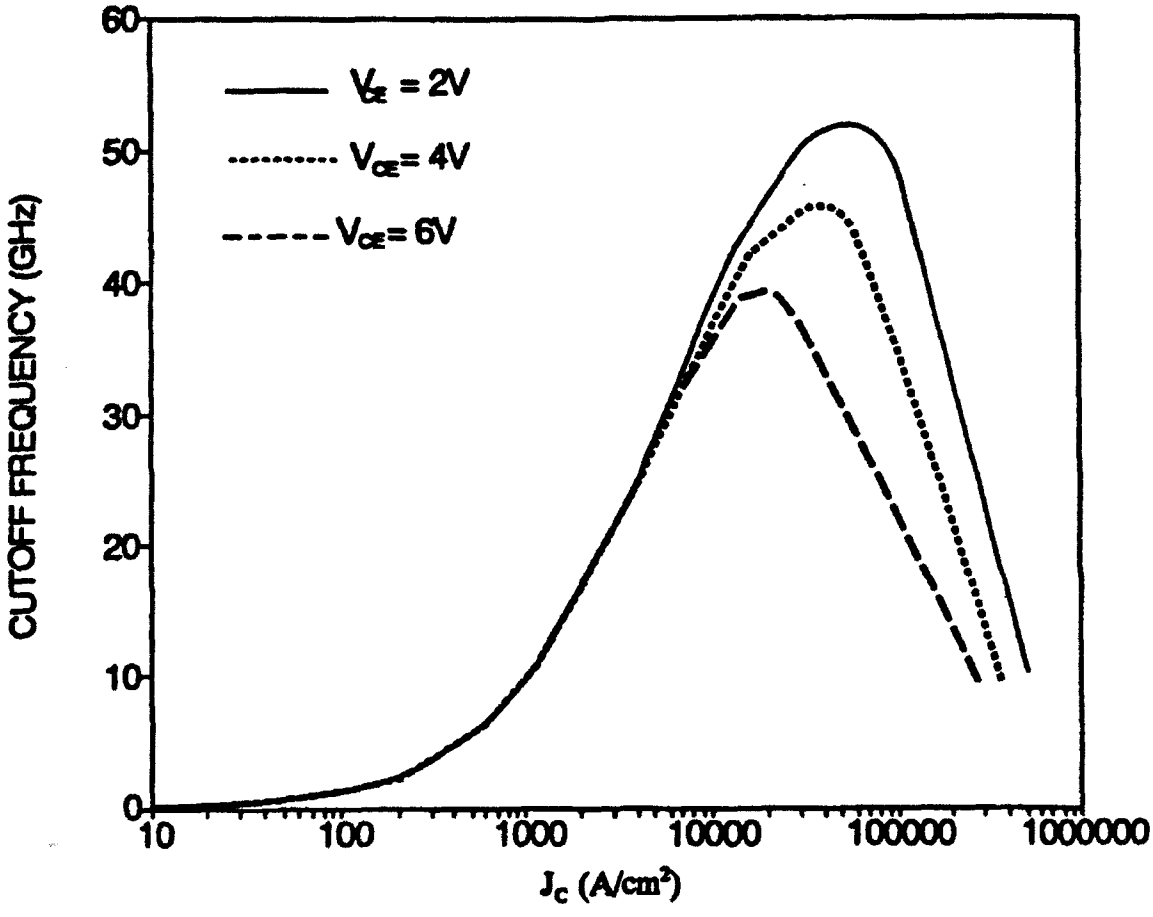


Fig. 8 Cutoff frequencies calculated from the present model for three different V_{CE}.

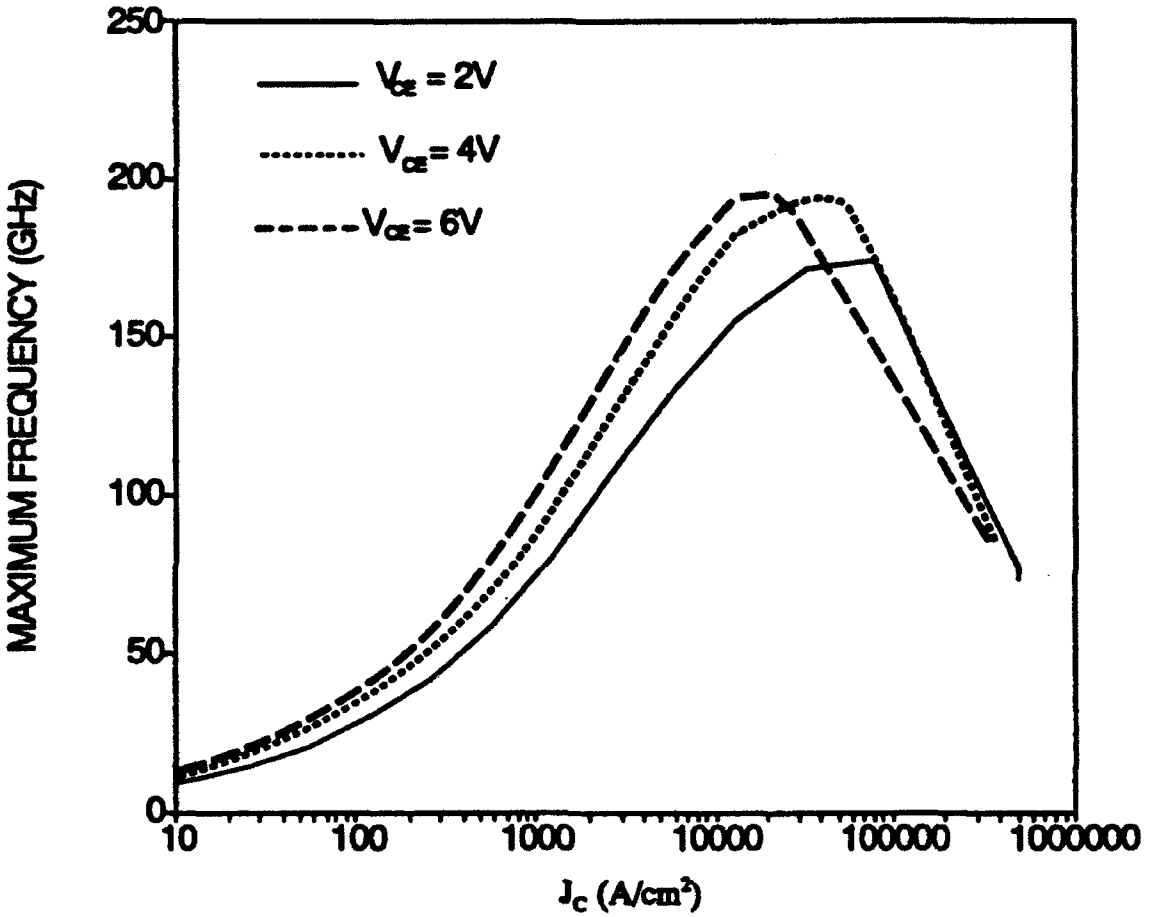


Fig. 9 Maximum frequencies calculated from the present model for three different V_{ce}.

degrades the HBT performance, is caused mainly by the high temperature associated with the high collector current and large collector-emitter applied voltage. The slower fall-off predicted by the isothermal model results from the voltage drops in the emitter and base quasi-neutral regions and current-induced base pushout occur at high current densities.

CONCLUSIONS

The thermal effect is important to the AlGaAs/GaAs HBT performance due to the poor thermal conductivity of GaAs and the high current handling capability of such devices. Aimed toward predicting the HBT behavior and physical trends, we have developed a physics-based, analytical HBT model including high-current as well as thermal effects. Three HBT figures of merit (eg, current gain, cutoff frequency, and maximum frequency) are calculated, and the results are in good agreement with those obtained from a model which solves numerically the Poisson and continuity equations coupled with heat transfer equation. When comparing the present model with the isothermal model, we found that the thermal effect is the main mechanism contributing to the commonly observed rapid fall-off behavior in HBTs operated at high current region.

REFERENCES

[1] B. Bayraktaroglu, R. D. Hudgens, M. A. Khatibzadeh, and H. Q. Tserng, "2.5 W cw X-band heterojunction bipolar transistor," IEEE MTT-S International Microwave Symposium Dig. 1989.

[2] M. A. Khatibzadeh, B. Bayraktaroglu, and T. Kim, "12 W monolithic X-band HBT power amplifier," IEEE MTT-S International Microwave Symposium Dig. 1992.

[3] L. L. Liou, C. I. Hunag, and J. Ebel, "Numerical studies of thermal effects on heterojunction bipolar transistor current-voltage characteristics using one-dimensional simulation," Solid-St. Electron., vol. 35, pp. 579, 1992.

[4] D. A. Sunderland and P. L. Dapkus, "Optimizing N-p-n and P-n-p heterojunction bipolar transistors for speed," IEEE Trans. Electron Devices, vol. ED-34, pp. 367-377, 1987.

[5] D. L. Bowler and F. A. Lindholm, "High current regimes in transistor collector regions," IEEE Trans. Electron Devices, vol. ED-20, pp. 257, 1973.

[6] C. T. Kirk, Jr., "A theory of transistor cutoff frequency falloff at high current densities," IRE Trans. Electron Devices, vol. ED-9, pp. 164, 1962.

[7] J. P. Holman, Heat Transfer, New York: McGraw-Hill, 1981.

[8] J. J. Liou, F. A. Lindholm, and B. S. Wu, "Modeling the cutoff frequency of single heterojunction bipolar transistors subjected to high collector -layer current," J. Appl. Phys., vol. 67, pp. 7125, 1990.

[9] H. F. Cooke, "Microwave transistors: theory and design," Proc. IEEE, vol. 59, p.1163, 1971.

A SWITCHED RELUCTANCE MOTOR DRIVE USING MOSFETS, HCTL-1100, and MC6802 MICROPROCESSOR

Shy-Shenq P. Liou
Assistant Professor

Lawrence Vo
Graduate Student

Division of Engineering

San Francisco State University
1600 Holloway Avenue
San Francisco, CA 94132

Final Report for:
AFOSR Summer Research Program
Wright Patterson Laboratory

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

September 1992

A Switched Reluctance Motor Drive Using MOSFETs, HCTL-1100, and MC6802 Microprocessor

Shy-Shenq P. Liou
Assistant Professor

Lawrence Vo
Graduate Student

Division of Engineering
San Francisco State University

Abstract

A switched reluctance motor drive is designed, built, and tested using the MOSFETs as the power switches. The control function of this drive is done by a general purpose motion control chip HCTL-1100 from Hewlett Packard. The interface between the HCTL-1100 and the user is through a Motorola microprocessor MC-6802. A bang-bang current control circuit is also built into the drive to limit the motor current to be less than or equal to the rated motor current. Simple assembly program can enable the user to input the command velocity, command position, and command profile to the HCTL-1100 motion control chip. The test run of this drive is very successful.

A Switched Reluctance Motor Drive Using MOSFETs, HCTL-1100, and MC6802 Microprocessor

Shy-Shenq P. Liou and Lawrence Vo

1. Introduction

The high temperature capability and fault tolerant characteristic of the relatively new switched reluctance motor make it an ideal candidate for many aircraft applications. Unlike conventional motor drives, the switched reluctance motor drive requires position sensing and complicated control strategy because of the inherent distinct stator and rotor magnetic structures. An one horsepower switched reluctance motor drive is designed, built, and tested. The MOSFETs are chosen as the power switches, the number crunching or control task is taken care of by a general purpose motion control LSI chip, HCTL-1100 from Hewlett Packard. Finally the interface between the users and the HCTL-1100 is via an evaluation module containing a Motorola MC-6802 microprocessor. Final testing indicated that this switched reluctance motor drive works very well. The design details of the drive will be explained fully in the following sections.

2. System Overview

The system block diagram for the switched reluctance motor drive is shown in Figure 1. As mentioned earlier, the HCTL-1100 motion control IC chip is used to handle all the controls and number crunching tasks. There are four control modes HCTL-1100 can perform; position control, proportional velocity control, integral velocity control, and trapezoid profile control. By inputing different values into the 64 8-bit registers of the HCTL-1100, different control mode can be realized easily. In our design, HCTL-1100 receives commands from the

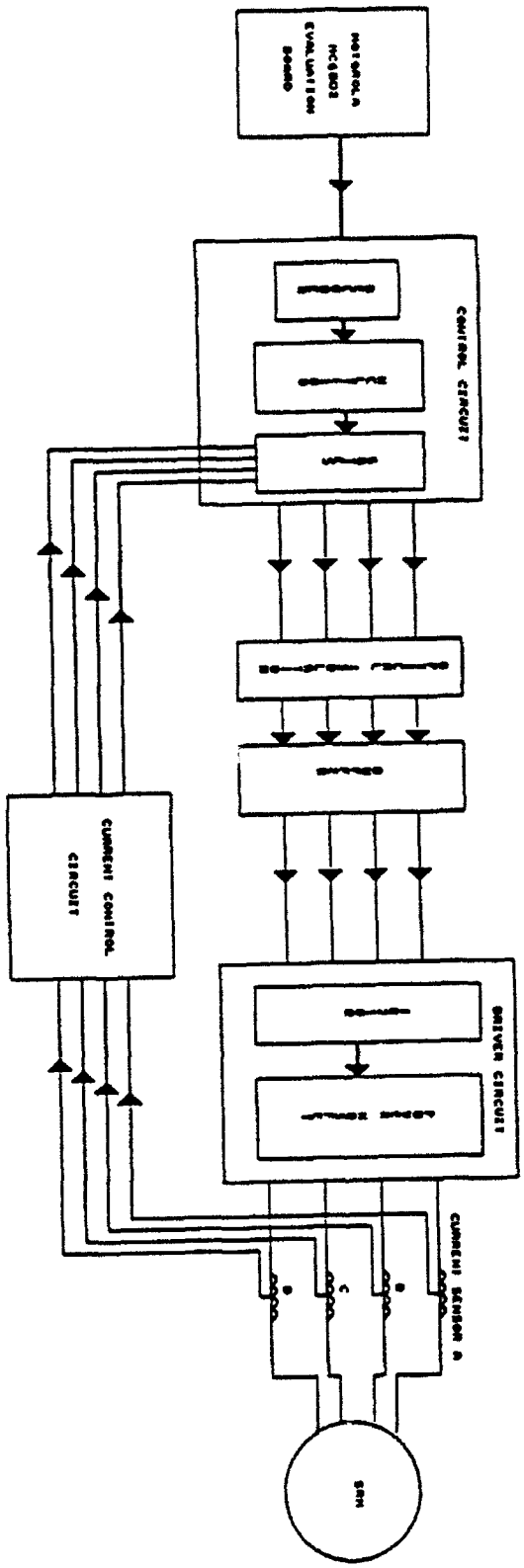


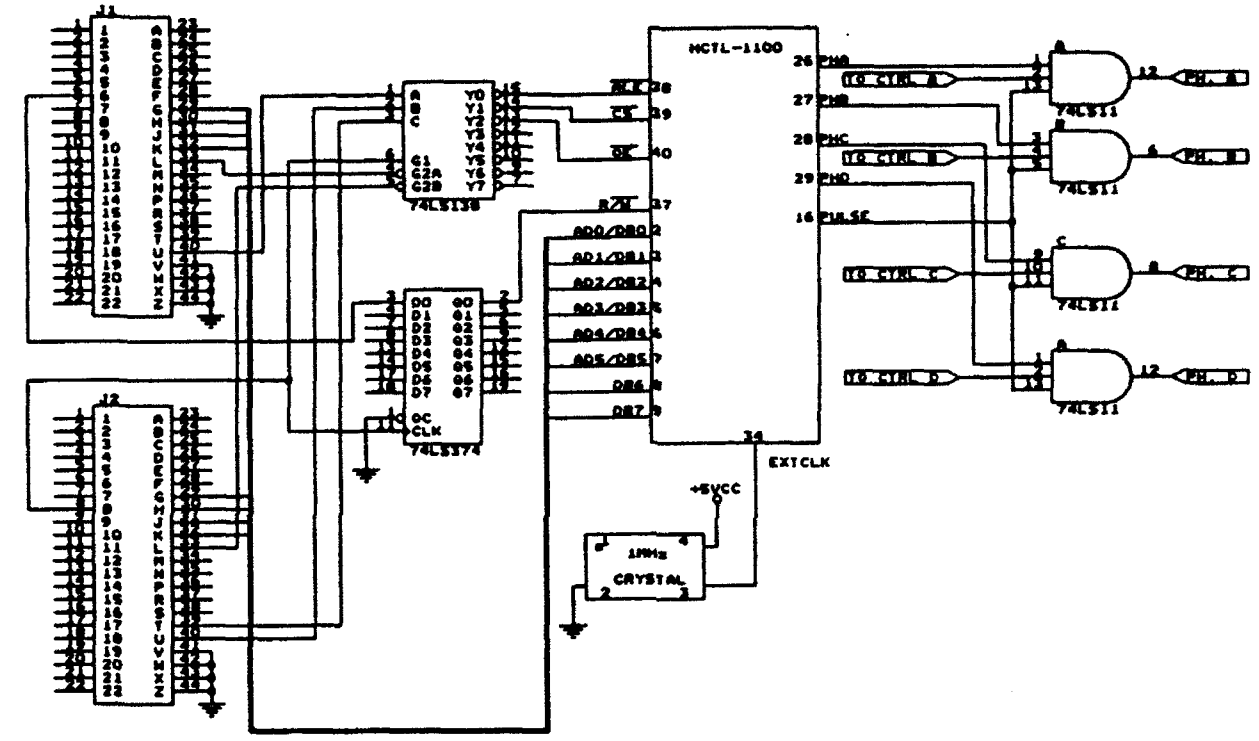
Figure 1: Block Diagram for the Switched Reluctance Motor Drive

users via an evaluation module with the Motorola MC6802 microprocessor. The interfacing, i.e. the READ and WRITE cycles, between the MC6802 microprocessor and HCTL-1100 is realized through the combination of hardware and software in assembly language. The control signal at logic level generated by HCTL-1100 is then fed to the driver circuit for the power MOSFETs via an optical isolator and buffer which provide the isolation, protection, and noise immunity. The output signals from the driver circuit are used to drive the power switches, the MOSFETs which in turns drive the four phase windings of the switched reluctance motor.

To incorporate current control into the control algorithm, a bang-bang control scheme is adopted. A current sensing circuit using hall effect current sensor is designed and built. The output signals from the current sensing circuits are ANDed with the outputs from the HCTL-1100 to generate the input signals to the driver circuit. This closes the control loop eventually.

3. Interfacing between MC6802 and HCTL-1100

The interfacing between the MC6802 and HCTL-1100 is done through the hardware and software. The detailed diagram is shown in Figure 2. The ALE/CS nonoverlapped timing diagram is chosen because of the simplicity. A 74LS138 decoder is used to generates the ALE and CS signals whenever address $0000 and $0001 are referenced respectively during the READ (get one byte from the MC6802 into HCTL-1100) cycle. For the WRITE (read one byte from the HCTL-1100 into MC6802) cycle, referencing the address $0002 generates the OE signal which is needed in addition to the ALE and CS in order to complete the data transfer. R/W signal from the MC6802 is fed to a D-Q flipflop 74LS374 first and then to the R/W pin of HCTL-1100 to satisfy the timing requirement between CS and OE signals during both READ and WRITE cycles. To synchronize all the timing, the E clock (0.89 Mhz approximately) from



NOTE: J1 AND J2 ARE FROM THE SYSTEM BUS PINOUTS OF 6802 EVALUATION MODULE
TO CTRL A, B, C AND D ARE THE OUTPUTS OF THE CURRENT CONTROL CIRCUITS

Figure 2: Interface Circuit between MC6802 and HCTL-1100

the MC6802 is used for both 74LS138 decoder chip and 74LS374 D-Q flipflop. As shown in Figure 2, the clock used by the HCTL-1100 is derived from a crystal oscillator at 1 Mhz.

3.1 READ Cycle

To send one byte of data from MC6802 to HCTL-1100, the software needed is shown as follow

```
LDX  #$080A      Enter $0A into register 08
STX  $0000       Generates ALE and then CS signals
```

In this sequence, the register X of MC6802 is loaded with a two byte data, $080A. First byte represents the address of the register of HCTL-1100 and the second byte is the content which is to sent to the register of HCTL-1100 whose address is specified by the first byte of data of the first instruction, LDX #$080A. The second instruction of the sequence will send the first byte of the register X to the data bus with address $0000 shown at the address bus which in turn generates the ALE signal through the 74LS138 decoder to start the whole READ cycle. The $08 on the data bus informs the HCTL-1100 that the address of the register which should take the data is register 08. The content which should go to the register 08 is sent by the MC6802 through the second part of the second instruction, STX $0000. In second part of this instruction, the address is upgraded by one so $0001 is shown in the address bus which generates a CS signal through the address decoder to the HCTL-1100. And $0A shows up on the data bus which is the content HCTL-1100 will take into the register 08. The measured timing diagram for the READ cycle is shown in Figure 3. The key point is only two instruction sequence is enough to send one byte data into HCTL-1100 through MC6802 microprocessor.

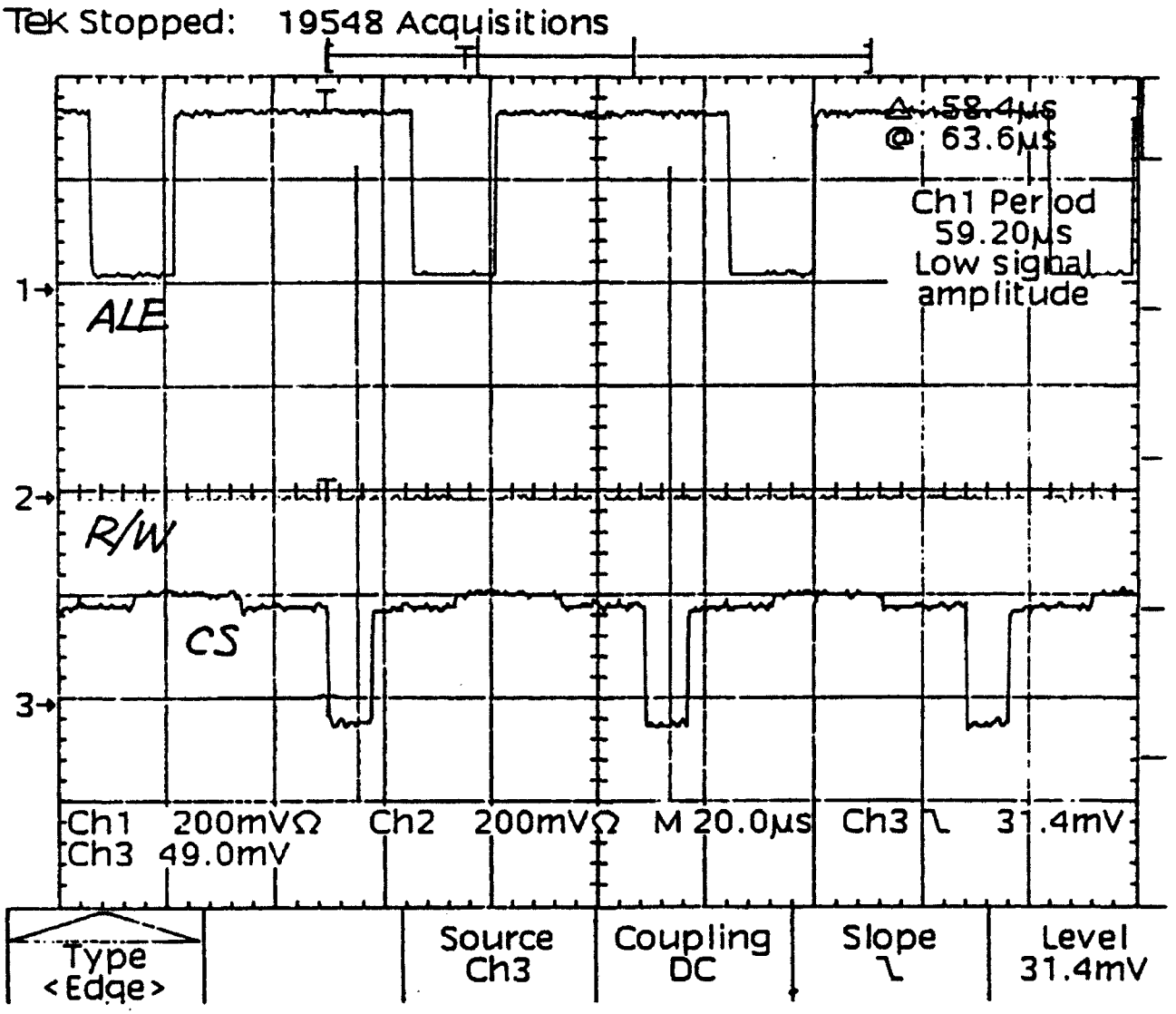


Figure 3: Timing Diagram for the READ cycle

3.2 WRITE Cycle

To get (read) one byte from the HCTL-1100, the necessary software is as follow

LDAA	#$09	Specify which register is chosen
STAA	$0000	Generate ALE signal
LDAA	$0001	Generate CS signal
NOP		For timing matching
NOP		For timing matching
LDAA	$0002	Generate OE signal and read data from data bus

First instruction specify which register the MC6802 wants from HCTL-1100. Second and third instructions generate ALE and CS signals in sequence because address $0000 and $0001 show up on the address bus. Two NOPs do exactly nothing more than delay the timing between CS and OE signal which is produced by the last instruction LDAA $0002 (when $0002 shows up on the address bus). In the meantime, the content of register 09 was put on the data bus by HCTL-1100 and is read into the MC6802. This completes the WRITE cycle. The measured timing diagram is shown in Figure 4. It only takes six instruction sequence to complete a WRITE cycle (get one byte of data from HCTL-1100).

4. Example for Proportional Velocity Control

A software example for a proportional velocity is shown as follow

```
LDX  #$0706     $06 into status register 07
STX  $0000
LDX  #$0F40     $40 into sample time register 0F
STX  $0000
LDX  #$183C     $3C into ring register 18
STX  $0000
```

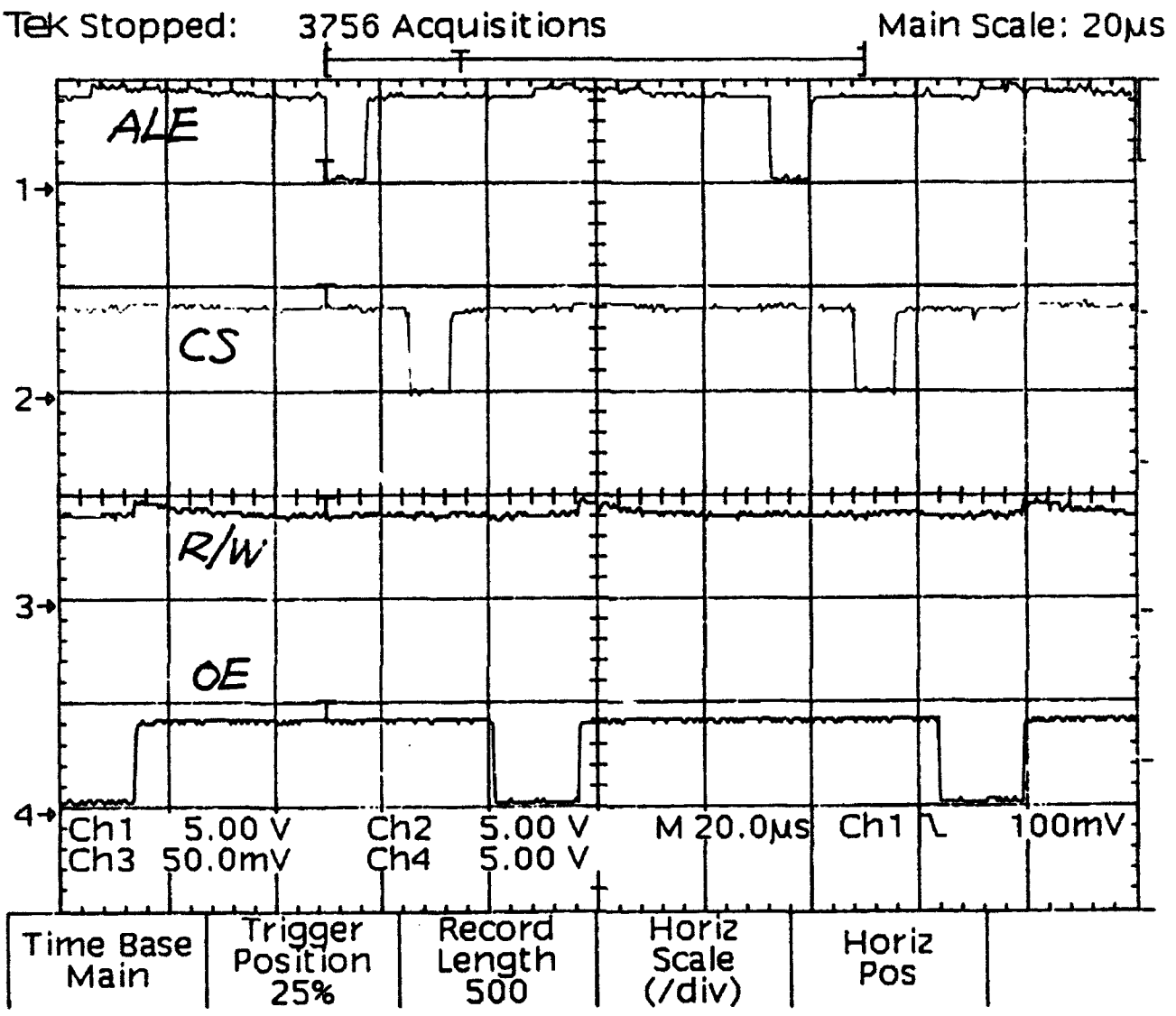


Figure 4: Timing Diagram for the WRITE Cycle

```
LDX  #$1A0F     $0F into X register 1A
STX  $0000
LDX  #$1B00     $00 into Y register 1B
STX  $0000
LDX  #$1C00     $00 into Offset register 1C
STX  $0000
LDX  #$1F00     $00 into advance register 1F
STX  $0000
LDX  #$1900     $00 into velocity timer register 19
STX  $0000
LDX  #$0503     $03 into program counter register 05
STX  $0000
LDX  #$2320     $20 into command velocity LSB
STX  $0000
LDX  #$2403     $03 into command velocity MSB
STX  $0000
LDX  #$000B     $0B into flag register/system ramps to
                specified speed
STX  $0000
```

The previous example inputs appropriate data into relevant registers of HCTL-1100 to do a proportional velocity control. How many instructions are needed depends on the desired control mode. Most likely not all 64 registers need data from the user.

5. Current Control Circuit

Current Control Circuit consists of power supply and current control circuit shown in Figures 5 and 6 respectively.

Phase current is sensed by the "LEM module LA 50-P." Since the maximum current of SRM is 4 Amps, the 11 turns of primary current helps to obtain the optimum accuracy of the phase current of SRM. Use R = 110 Ohms (see LEM module LA 50-P data sheet), I_0 = 48 mA, V_{in} = 5.3 Volts.

The hysteresis comparator is designed to control the phase current of SRM. The upper limit voltage is set at V_{in} = 5.3 Volts, while the lower limit is set at 5.0 Volts. Once the current over

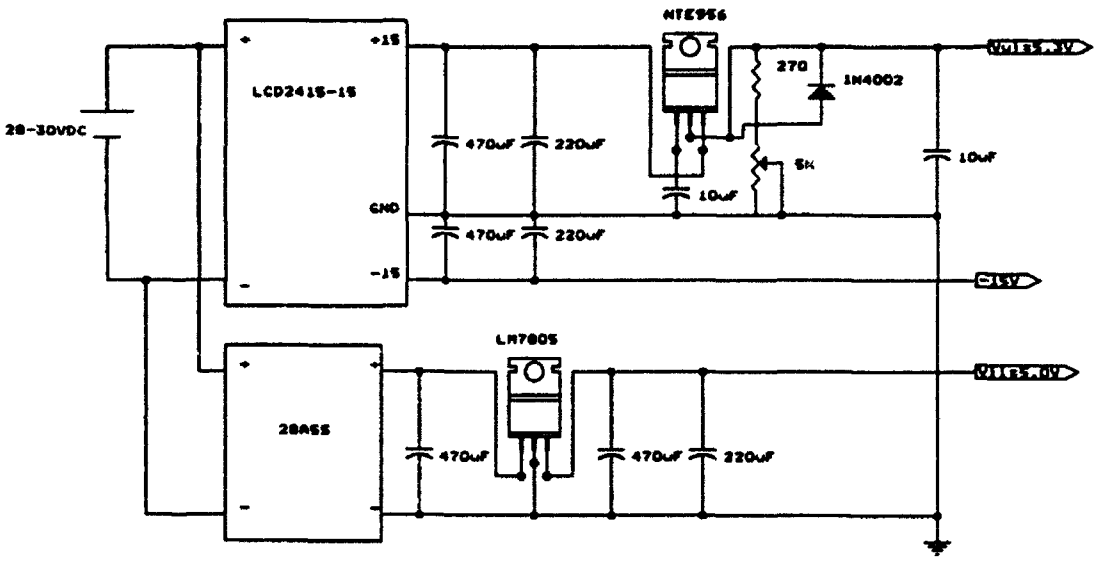
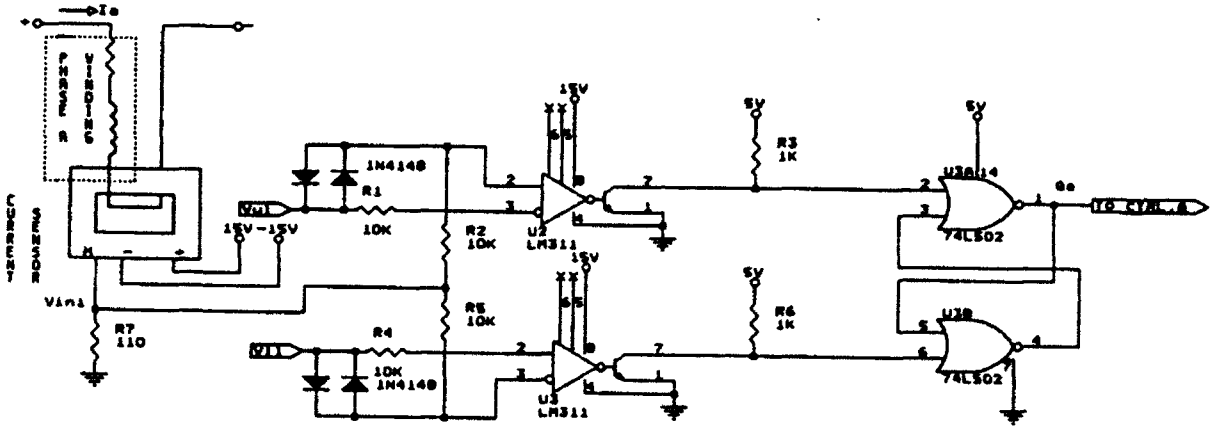
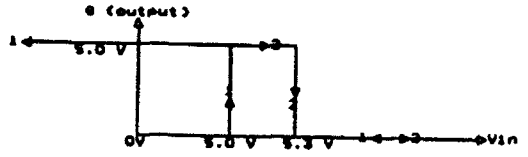


Figure 5: Power Supply Circuit Diagram for Current Control Circuit



1. USE 0.1uF CERAMIC CAPACTOR FOR EACH SUPPLY TO GND
2. THE CURRENT CONTROL CIRCUIT IS THE SAME FOR PHASE B, C AND D



COMPARATOR HYSTERYSIS GRAPH

Figure 6: Circuit Diagram for Current Control Circuit

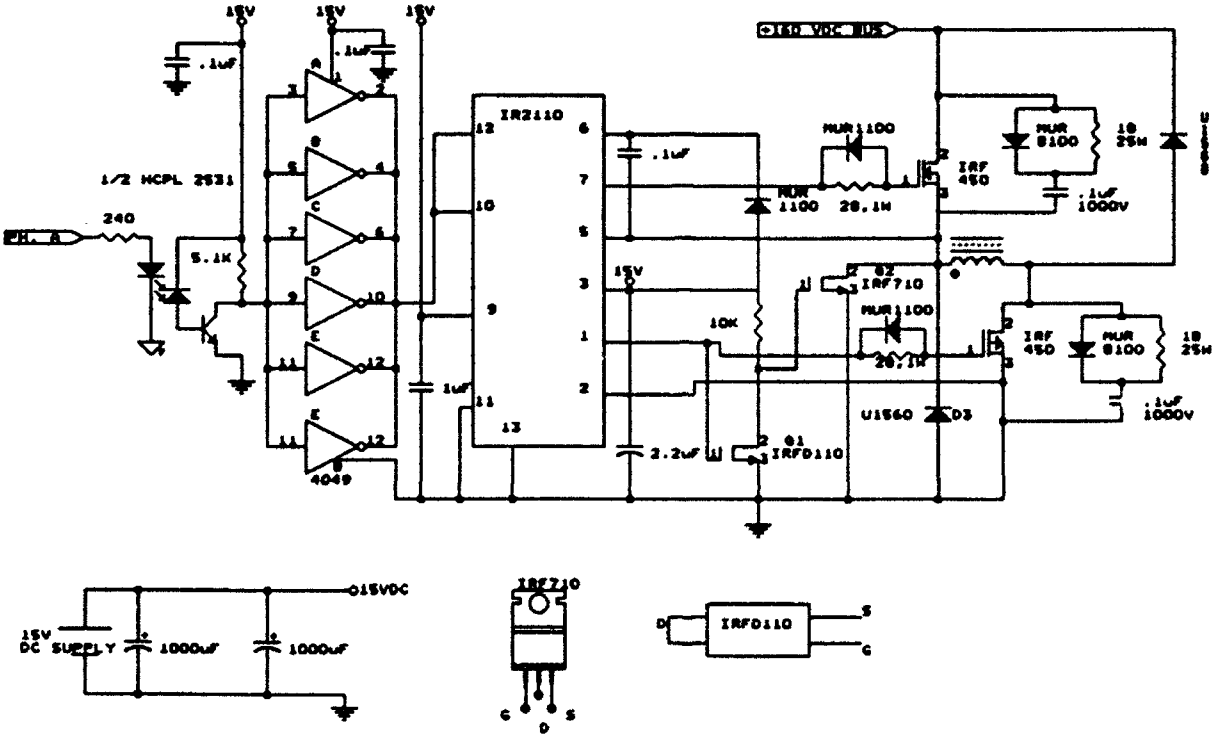
4.0 Amps, the output of current control circuit will be zero to turn the gating signal of MOSFET off; then the phase current reduces until it reaches the lower limit, the output of current control circuit will be high to turn the gating signal on again. the hysteresis graph of current control circuit is shown in Figure 6.

The clamp diodes are added in the comparator circuit to protect the Op-amp when the input signal exceed the common mode voltage.

6. MOSFET Driver Circuit

The IR2110 is used to drive the totem-pole of N-channel MOSFET since it provides very fast switching speed, low power dissipation and can operate on the bootstrap principle. The circuit is shown in Figure 7. The upper MOSFET is turned on by charging 0.1 uF capacitor to full DC supply. In another way, the IR2110-pin 5 must be pulled down to ground in order for 0.1 uF capacitor to be charged. This operation only occurs when D3 turns on. Since the turn-on time of the D3 is short, the 0.1 uF capacitor does not have enough time to charge up for sometimes. The three components are added into the circuit such as R=10K, Q1=IRFD110, Q2=IRF710 to make sure the 0.1 uF capacitor is fully charged. The principal operation of the additional circuit is that when the gating signal is high, the circuit operates as normal. However, when the gating signal is low the Q1 is turned off to pull the gate of Q2 to +15 Volts which turns Q2 on eventually. As result of this operation, IR2110-pin5 is pulled down to zero volt for this instant and 0.1 uF capacitor is fully charged.

To minimize the inductance in the gate drive loop, the twisted wires are used to connect between IR2110 and the gate and the source of MOSFET. The parallel combination of MUR1100, fast switching diode, and 28 Ohms resistor is to discharge energy storing at the gate-



NOTE: THERE ARE FOUR IDENTICAL DRIVER CIRCUITS FOR SRM
BUT ONLY SINGLE 15VDC POWER SUPPLY FOR ALL FOUR PHASES

Figure 7: Circuit Diagram for the IR2110 Driver for MOSFETs

source capacitor and to reduce the ringing of the gate signal respectively. These extra components can introduce the switching losses and rise of junction temperature of the MOSFET. However, in our experiment, the temperature at rated voltage and rated current is in the range of 20 to 25 degrees C with the snubber circuit for each MOSFET. The terminal strip connection diagram for motor phase windings and 160 volts DC bus is shown in Figure 8.

6. Experimental Results

The load currents and voltages of the top MOSFETs between drain-source for phase A is shown in Figure 9. In Figure 10, the graph is obtained for the worst case in which the two phases of SRM are turned on at the same time. In these two cases, the load currents is measured with 22 Ohms resistor in series with 113 uH inductor. The graphs we obtain are theoretically expected except when MOSFET is turned off. At this time, The load current goes below zero due to reverse recovery current of diode. The drain-source voltage of the MOSFET is zero when it is ON while the load current rises up to its maximum value at the end of period. When MOSFET id OFF, the drain-source voltage is equal to DC bus voltage and the load current return to zero.

Figure 11 shows the recorded phase current waveform for phase A winding of the switched reluctance motor when the proportional velocity control assembly language program as stated in section 4 is used to run the motor at specified speed. Based on this waveform, the designed switched reluctance motor drive works very well.

1. A, B, C AND D: CURRENT SENSORS FOR PHASE A, B, C AND D
2. POSITIVE CURRENT OF THE SENSOR IS DOWNWARD (IN THE PAPER)

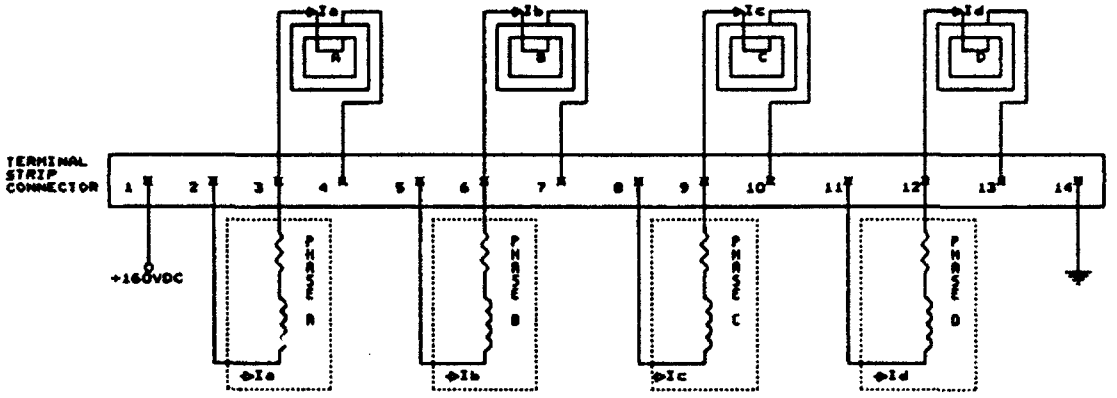


Figure 8: Terminal Connection Diagram for Power MOSFET Module

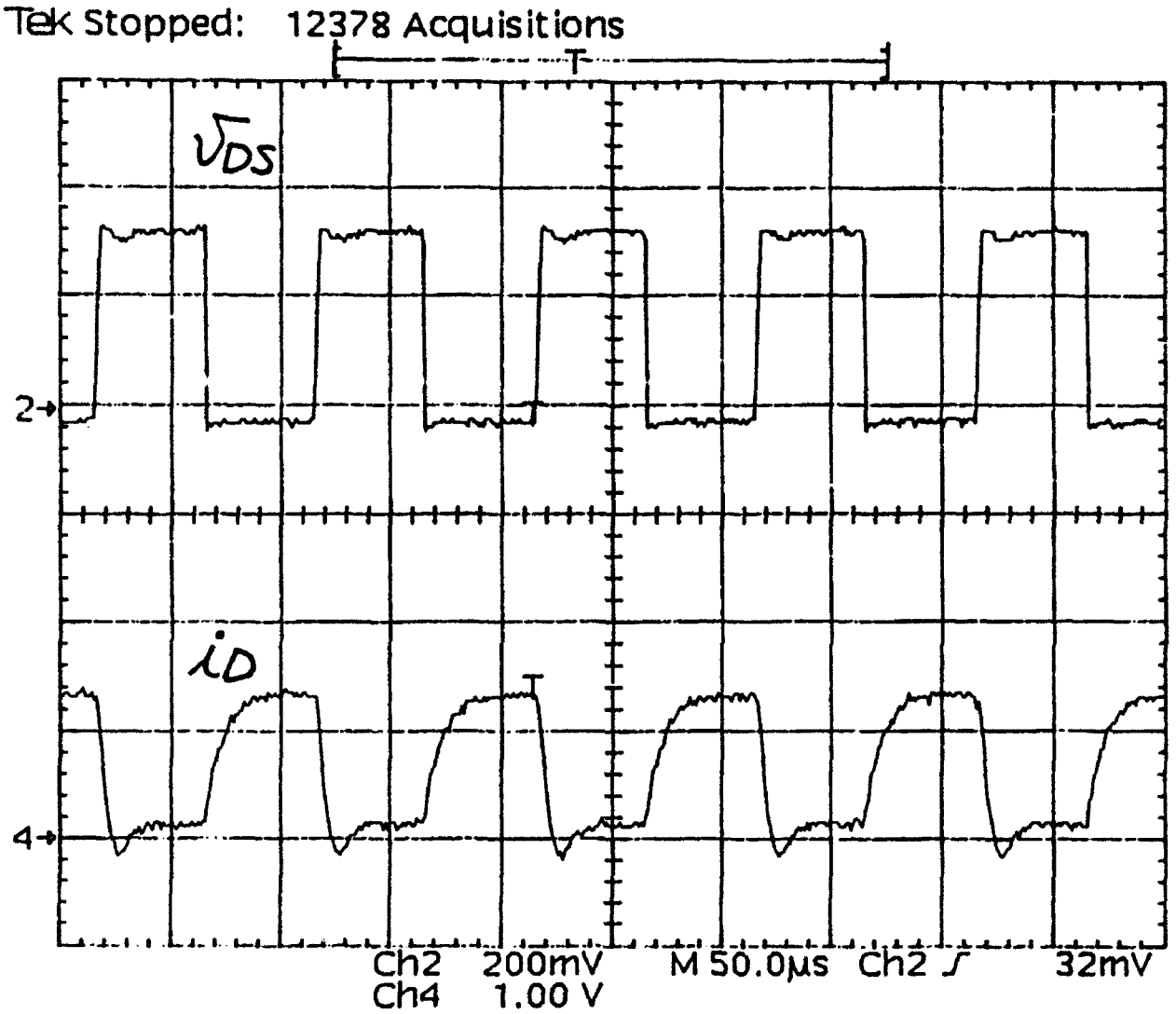


Figure 9: Drain Current and Drain to Source Voltage of the MOSFET

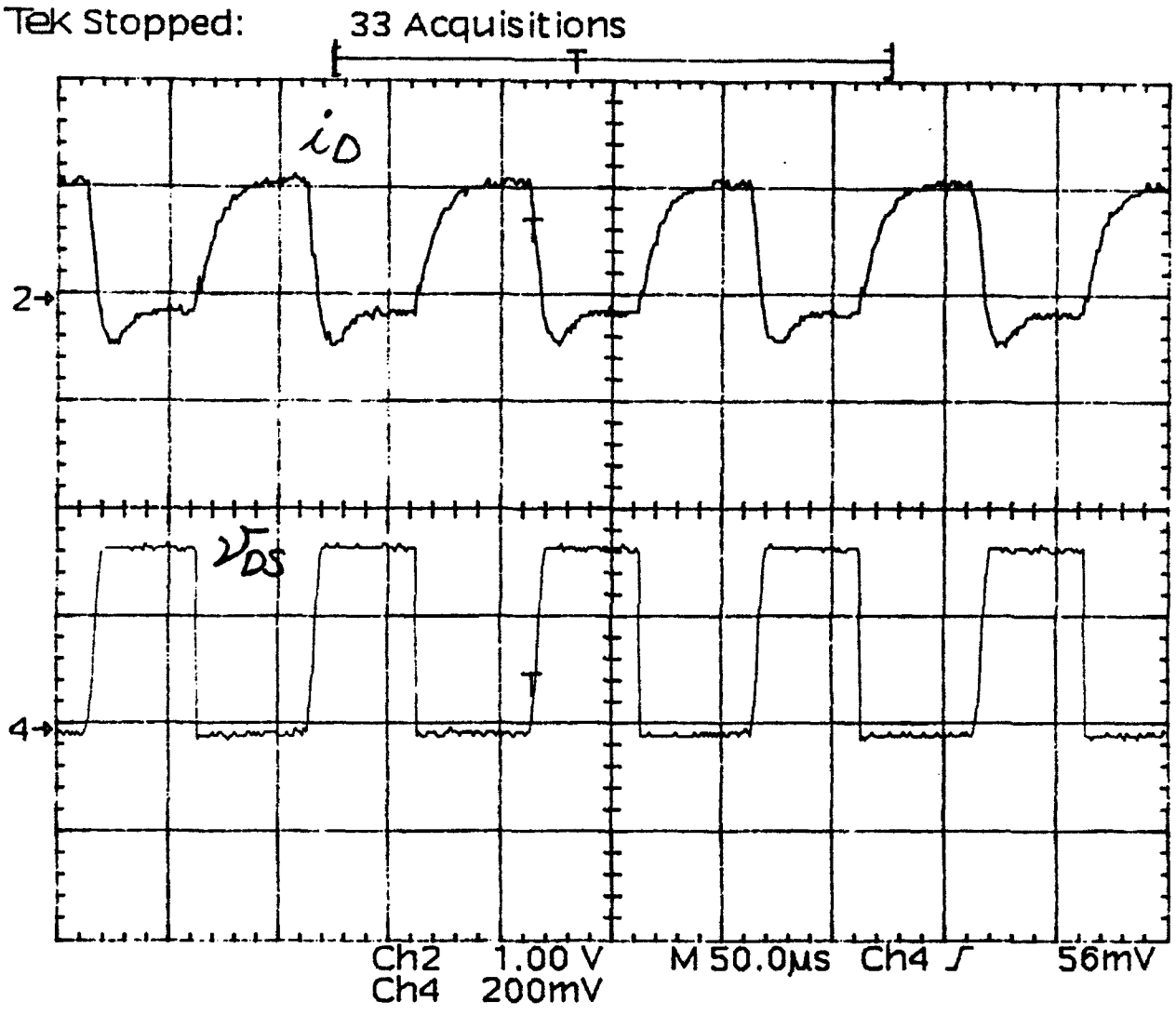


Figure 10: Drain Current and Drain to Source Voltage for Two Phase ON

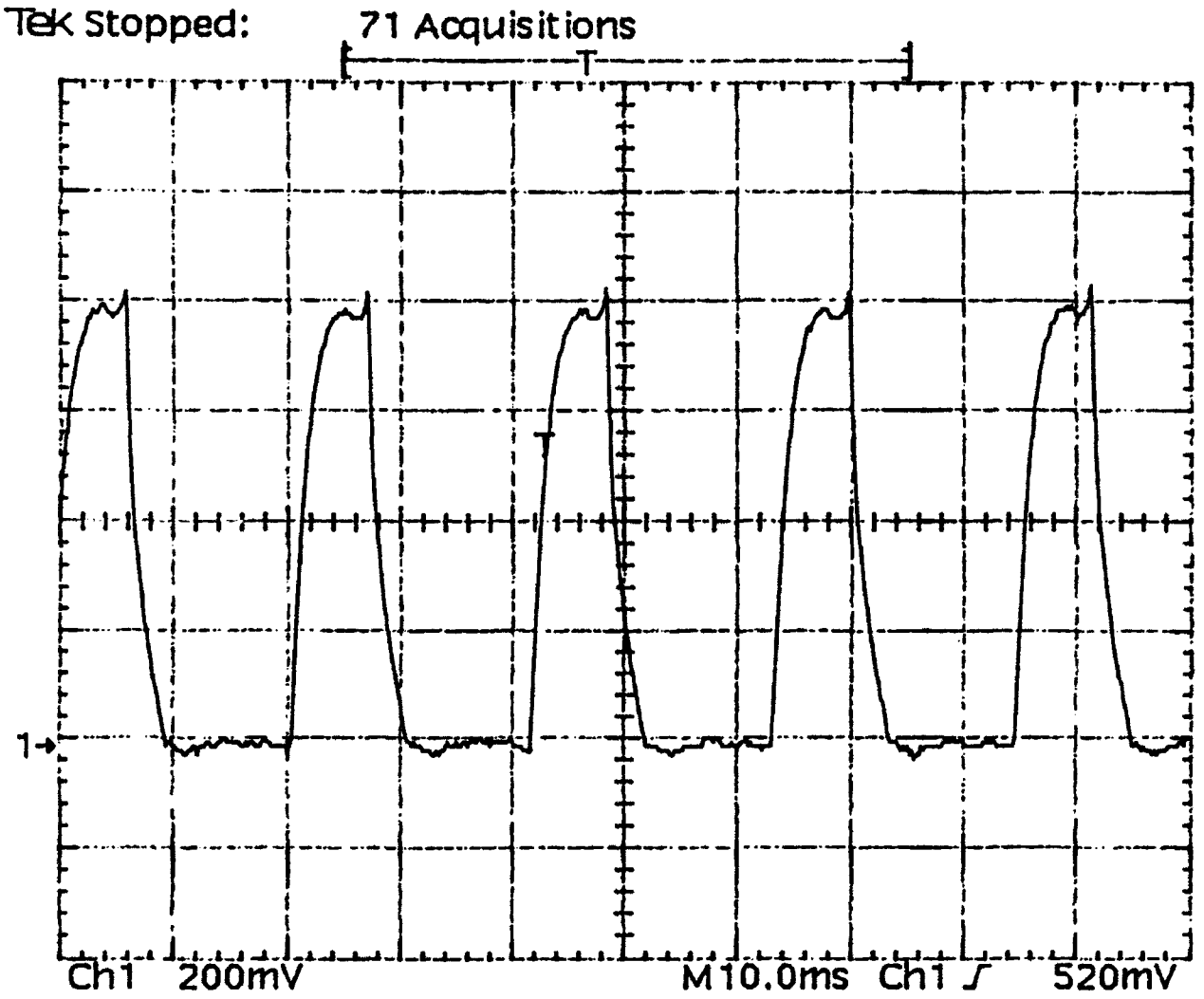


Figure 11: Current Waveform of Phase A of Switched Reluctance Motor

THERMAL ANALYSIS AND MOLECULAR WEIGHT DISTRIBUTION OF TRIARYL PHOSPHATES

Christopher C. Lu
Associate Professor
Department of Chemical and Materials Engineering

University of Dayton
300 College Park Avenue
Dayton, OH 45469

Final Report for:
Summer Research Program
Research & Development Laboratories

Sponsored by:
Lubrication Branch
Aero Propulsion and Power Directorate
Wright Laboratory
(AFMC)

July 1992

Laboratory Focal Point:
Robert L. Wright Jr.
Senior Project Engineer,WL/POSL
WPAFB,OH 45433-6563

THERMAL ANALYSIS AND MOLECULAR WEIGHT DISTRIBUTION OF A TRIARYL PHOSPHATE SOLUTION

Christopher C. Lu
Associate Professor
Department of Chemical and Materials Engineering
University of Dayton

ABSTRACT

Thermogravimetric analysis (TGA) and differential scanning calorimetry (DSC) were used to study the chemical and physical behavior of a triaryl phosphate (TAP) solution which has potential for use as a high temperature, vapor phase lubricant for turbine engines. The heat of vaporization of the TAP solution was found to be 20.8 kcal/mol which is very close to a literature value for tricresyl phosphate (TCP). Gel Permeation Chromatography (GPC) was used to measure the molecular weight distribution of the TAP solution and showed that larger molecular weight molecules were formed when the TAP solution was heated under oxygen and nitrogen environments. This study suggests that the molecular weight increase is due to polymerization. As would be expected, the kinematic viscosity of the TAP solution also increases with heating. This is believed to be the reason that TAP vapor deposition films can be used as a high load/high speed bearing lubricant for low traction operation at temperatures of at least 670°C.

THERMAL ANALYSIS AND MOLECULAR WEIGHT DISTRIBUTION FOR A TRIARYL PHOSPHATE SOLUTION

INTRODUCTION

Triaryl phosphates (TAP), such as tricresyl phosphate (TCP), have long been used as anti-wear additives for turbine engine lubricants. Recently, Wedeven (ref.1) conducted a traction study between a rotating ball and disc specimen and found that by using a mixture of TCP and perfluoroalkylpolyether (PFPAE) fluids delivered to the tribological contact zone as a vapor, the traction coefficient was lowered to a range of 0.02 to 0.08 over a temperature range of 100-670°C. Klaus, et al. (ref.2) also reported that the deposition rate of TCP vapor condensates on iron and stainless steel surfaces increased when temperature increased from 500°C to 700°C.

The potential for application of vapor lubricants in advanced, high temperature engines appears promising if further development is anywhere nearly as successful as early research testing. However, further understanding of physical and chemical behavior of lubricants in liquid and vapor phases is needed. The main objective of this study is to develop basic understandings for the physical and chemical changes of a TAP solution under a heating environment, and to investigate its rheological and tribological behavior as a vapor phase lubricant.

First, Thermogravimetric Analysis (TGA) was used to study the physical and chemical changes of TAP solution at temperatures of 170°C to 300°C for 3 hours under dry air and nitrogen environments. The residual samples obtained from the TGA were then analyzed for molecular weight distribution using gel permeation chromatography (GPC). The chemical and physical changes observed from the TGA test were also compared with Differential Scanning Calorimetry (DSC) experiments.

Secondly, the TGA test results were analyzed and the data was used to calculate the reaction rate constant and the activation energy of the TAP solution by an Arrhenius method. The DSC results were also analyzed to calculate the heat of vaporization of the TAP solution. Finally, using concepts of rheology and tribology, reasons were postulated to explain why thermally deposited films of vapor can be used as a lubricant to decrease the traction coefficient in a rolling/sliding contact at relatively high operating temperatures.

RESULTS AND DISCUSSIONS

1. Thermogravimetric Analyser (TGA)

The TAP solution (about 50 mg) was placed into an aluminum sample pan in a TA Instruments® model 951 TGA. The instrument was programmed to raise the temperature, within a few seconds, to a desired value and to run isothermally for 3 hours; the tests were run in both dry air and nitrogen environments. The TGA plots are shown in Figures 1 through 4, and the results are tabulated in Table 1.

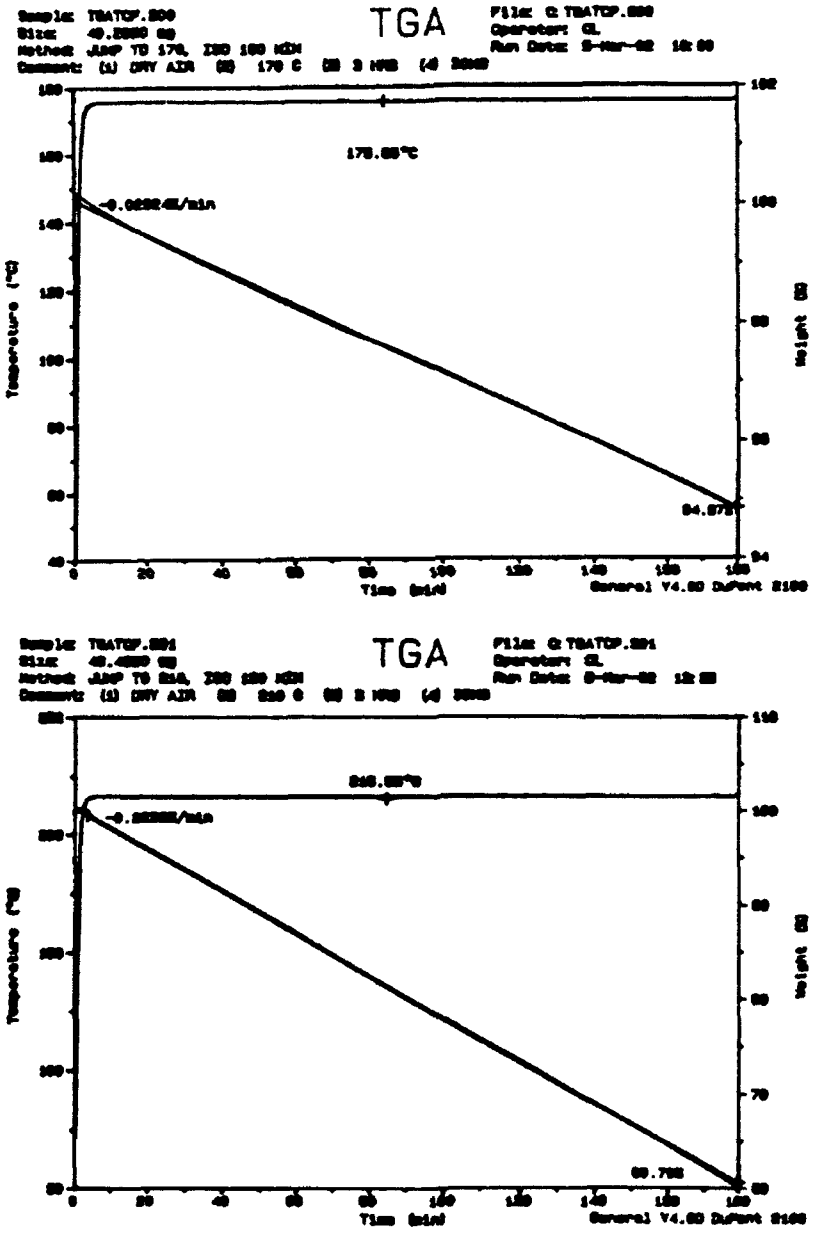


Fig.1: TGA Runs for Dry Air at 175°C and 217°C

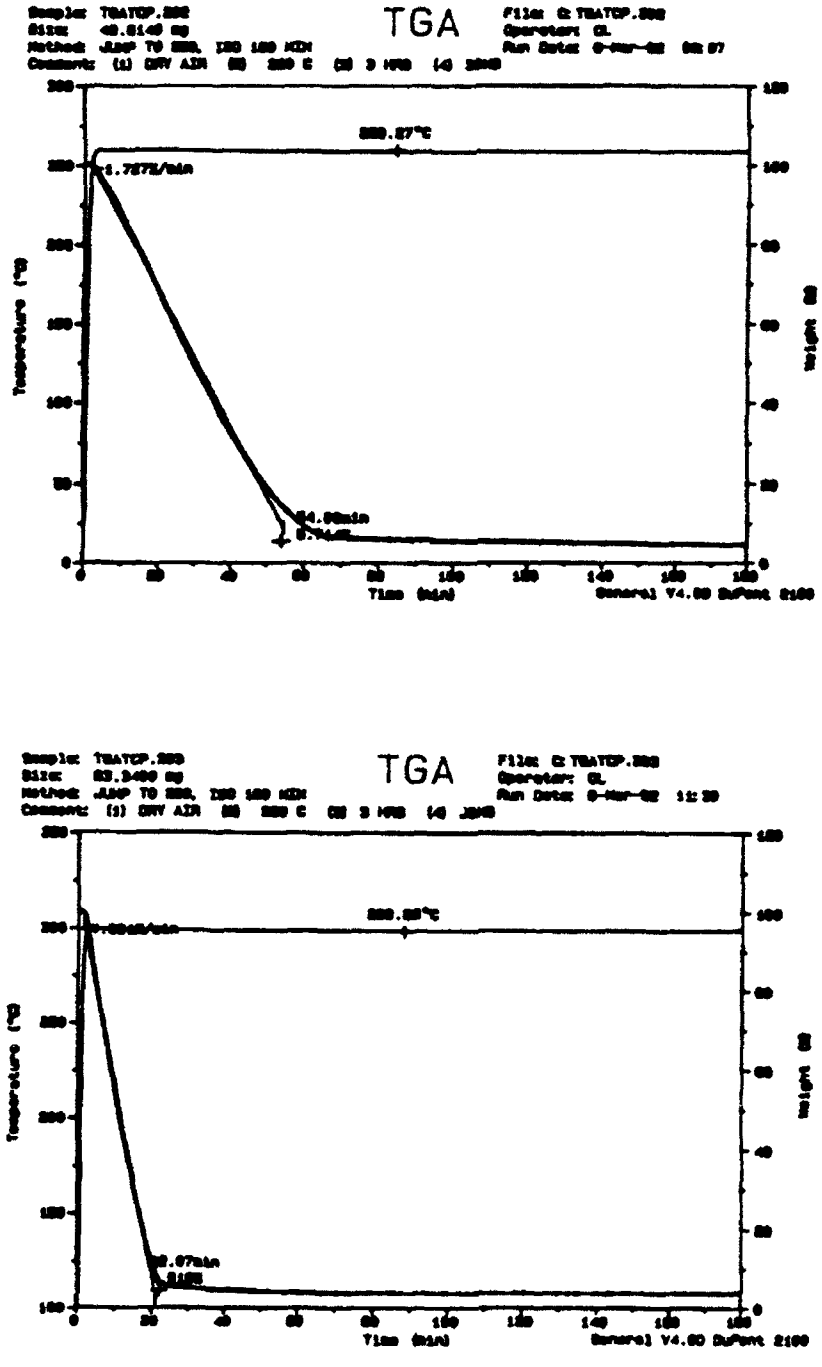


Fig.2: TGA Runs for Dry Air at 260°C and 300°C

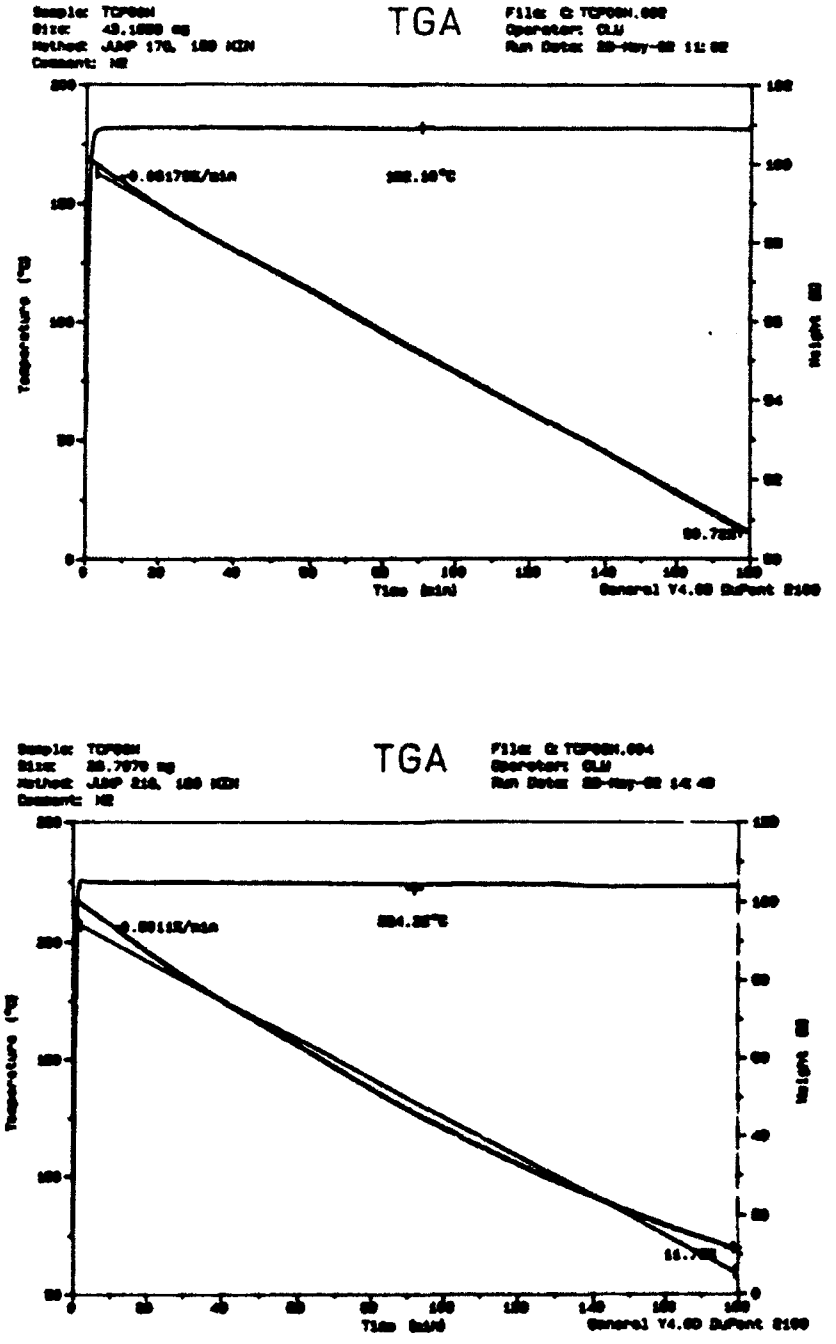
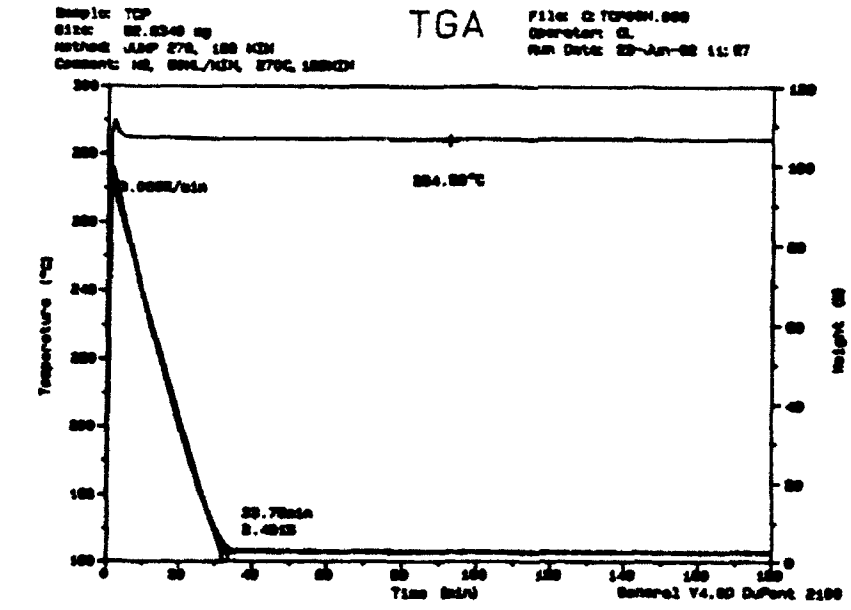


Fig.3: TGA Runs for N_2 at 182°C and 224°C



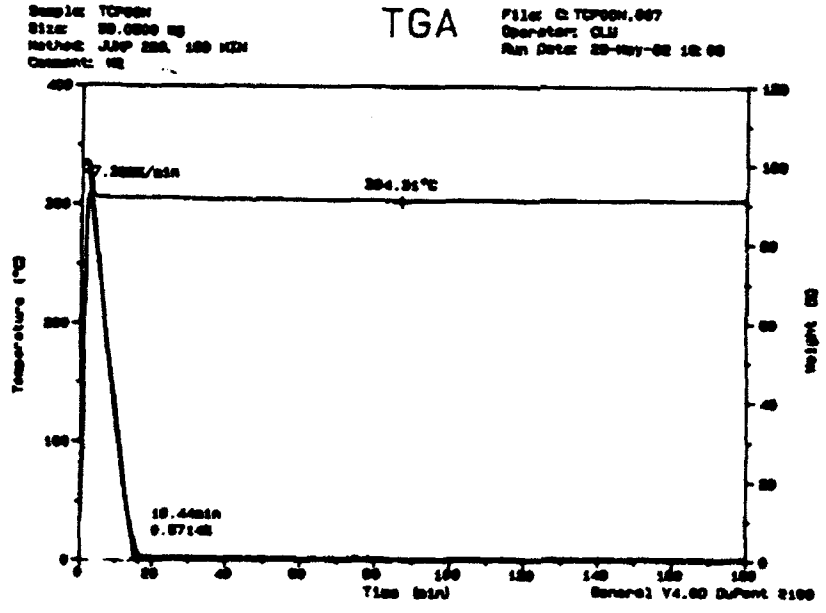


Fig.4: TGA Runs for N_2 at 285°C and 304°C

Table 1. TGA Test Results

	Temperature(°C)	% mass remaining	(minutes)
Dry Air	175	94.9%	(180)
	217	6[illegible].7%	(180)
	260	5.74%	(54)
	300	4.7%	(22)
N_2	182	90.7%	(180)
	224	11.5%	(180)
	285	2.43%	(34)
	304	0.6%	(16)

The results obtained from the TGA were used to calculate the rate constant for the vaporization of TAP as following:

$$dm/dt = -km^n, \text{ for a zero order reaction, } n=0,$$

$$dm/dt = -k,$$

$$dm = -k\,dt,$$

$$m_0 - m = kt,$$

$$(m_0 - m)/m_0 = (k/m_0)t.$$

According to the Arrhenius equation:

$$(k/m_0) = (A/m_0)\exp(-E/RT),$$

$$\ln(k/m_0) = \ln(A/m_0) - (E/RT),$$

where A is the frequency factor and E is the activation energy.

The values of k/m_0 are obtained from the TGA results which can be seen in Figures 1 through 4; and the values of 1/T correspond to each operating temperature. These numbers are tabulated in Table 2, and plotted in the semi-log scale as shown in Figure 5.

Table 2. Data for Arrhenius Equation

	Sample Temperature (°C)	k/m_0 (1/min)	1/T (1/°K)
Dry Air	175	2.94E-4	2.23E-3
	217	2.24E-3	2.04E-3
	260	1.73E-2	1.88E-3
	300	4.876E-2	1.75E-3
N_2	182	5.18E-4	2.2E-3
	224	5.011E-3	2.01E-3
	285	3.06E-2	1.79E-3
	304	7.325E-2	1.73E-3

The above data are linearly fitted using the least squared method, and the results are:

for dry air : y = 1.114E+7 exp(-1.091E+4 * x), and

for nitrogen: y = 2.823E+6 exp(-1.014E+4 * x).

The equation and the line shown in Figure 5 are based on all eight data points since they are very close to a straight line.

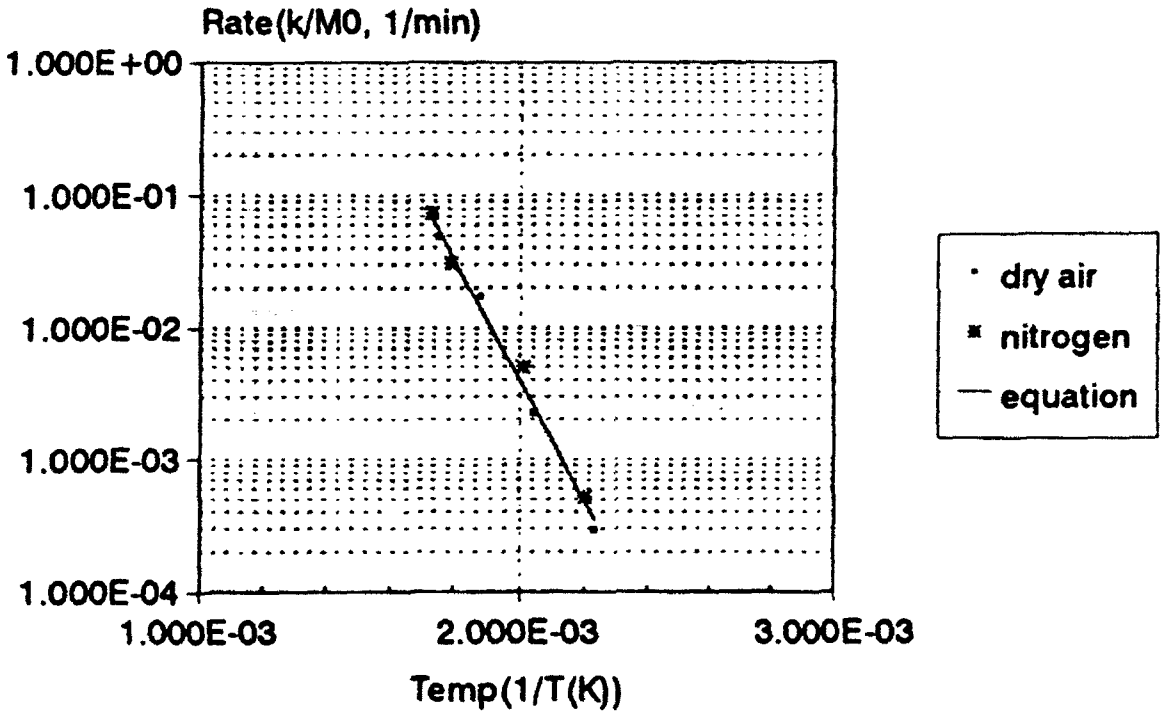
Using the above equations, the activation energy is calculated as:

$$E_{dry\ air} = (1.09E+4)(1.98) = 21.6 \text{ kcal/mol, and}$$

$$E_{nitrogen} = (1.014E+4)(1.98) = 20.07 \text{ kcal/mol.}$$

The result shows that only 5.1% of TAP solution evaporates during 3 hrs at 175°C in the dry air, while 95.3% evaporates during first 22 minutes at 300°C. The evaporation process is faster in the nitrogen environment because a limited oxidation probably takes place for TAP in the presence of nitrogen. We postulate that the oxidation reaction promotes TAP molecules to polymerize, as a result, the molecular weight of molecules grows larger and the evaporation of the solution becomes slower. This is the reason why the TAP solution evaporates slower in the oxygen enviroment and faster in nitrogen. In order to verify our postulation we used the gel permeation chromatography (GPC) to measure the molecular weight distribution of the TAP residues obtained from the TGA tests; these results will be discussed in the next section.

It is interesting to mention that Makki, et al. (ref.3) studied the deposition rate of TCP vapor on the stainless steel surface and calculated the value of the activation energy to be 13.85 kcal/mol for a high induction period and equal to 37.79 kcal/mol for a low induction period. Heat of vaporization of tritolyl phosphate (which has the same chemical formula of TCP) is 20.84 kcal/mol as listed in the CRC handbook. It was also found by integration of the vaporization peak in the DSC experiment that the heat of vaporization was about 20 kcal/mol.



y=(5.901E+6)*exp((-1.056E+4)*x)

Fig.5: Rate Constant for Arrhenius Equation

2. Gel Permeation Chromatograpy(GPC)

Gel permeation chromatography (GPC) was used to measure the molecular weight distribution of TAP residues obtained from the TGA test. Analyses were made on a HP 1084B Hewlett Packard liquid chromatograph with a 500 Angstrom and two 100 Angstrom GPC columns (PL-Gel, Polymer Laboratories, 5 micron pore sizes), tetrahydrofuran (THF) as the solvent, and using a 254 nm ultraviolet detector. The molecular weight was calculated from the retention time using the following formula:

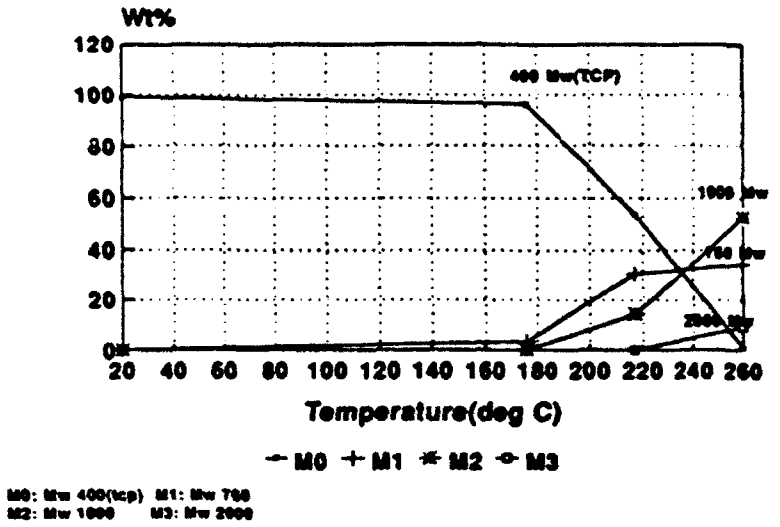
$$\log(Mw) = -0.1195*Time(minutes) + 5.2035.$$

This formula was calibrated by using the molecular weight of several polyphenyl ethers, e.g., 4P3E as 354, 5P4E as 446 and 6P5E as 538. The molecular weight of fresh TAP solution was measured by GPC and calculated as 398 by the above formula; formula weight of tricresyl phosphate is 368. The molecular weights of TAP among those TGA residue samples were analyzed by GPC and ranged between 380 and 412, and was rounded 400 in this study. The result of molecular weight distribution is listed on Table 3.

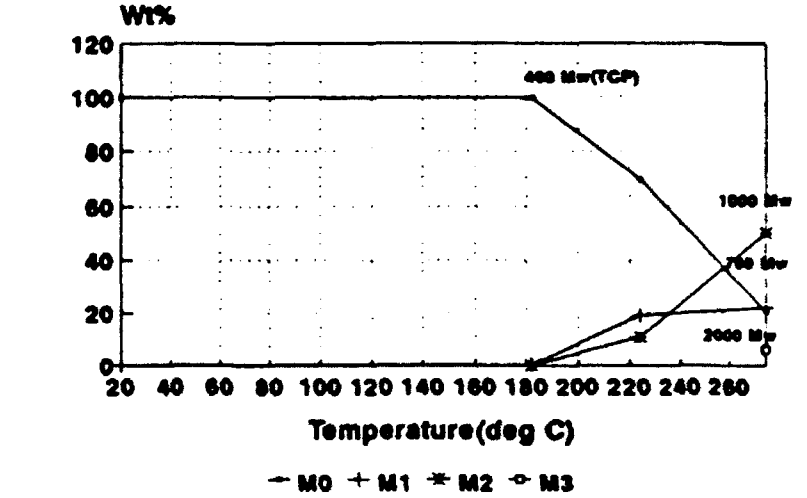
Table 3. Mw Distribution

TAP sample from TGA Test	M_0 (Mw 400)	M_1 (Mw 760)	M_2 (Mw 1000)	M_3 (Mw 2000)
Dry Air				
fresh TCP	99.3 wt%	0.24wt%		
A(175°C, 3hr)	96.1	3.4		
B(217°C, 3hr)	53.8	30.5	14.3	
C(260°C, 3hr)	0.8	34.5	52.5	9.15
D(300°C, 3hr)	SAMPLE DID NOT DISSOLVE WITH THF			
Nitrogen				
E(182°C,3hr)	99.7	0.3		
F(224°C,3hr)	69.8	19	11	
G(275°C,3hr)	20.5	22	50	6.2
H(304°C,3hr)	ALL TAP EVAPORATED IN TGA TEST			

The results show that when TAP solution is heated from 170°C to 300°C the TCP molecule (Mw of 400) gradually decreases and the higher molecular weight molecules gradually increase. As postulated in the TGA study, the oxidation promotes TAP molecules to polymerize and to form larger molecular weight molecules. In the nitrogen environment, the polymerization reaction is still taking place, but is less dramatic. The molecular weight distribution of the TAP solution in both air and nitrogen are also plotted in Figure 6.



(air)



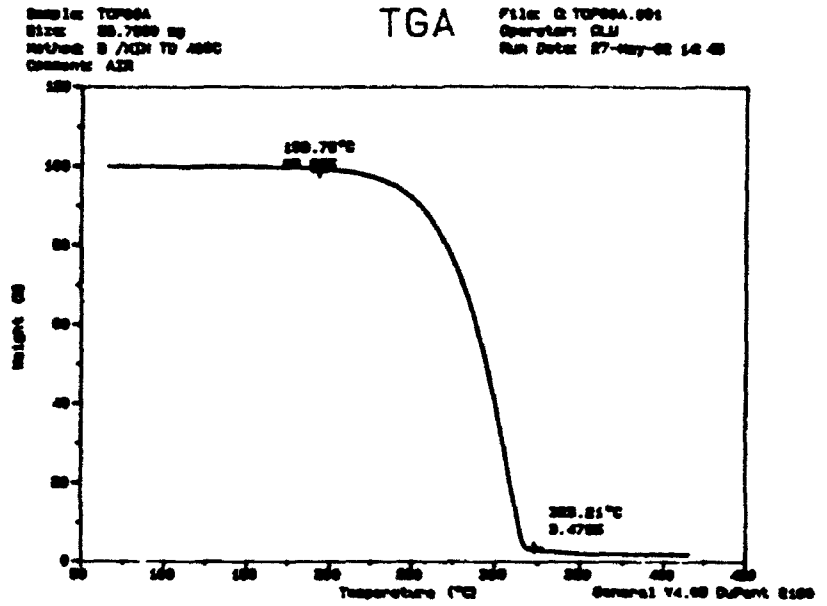
(nitrogen)

Fig.6: Molecular Weight Distribution

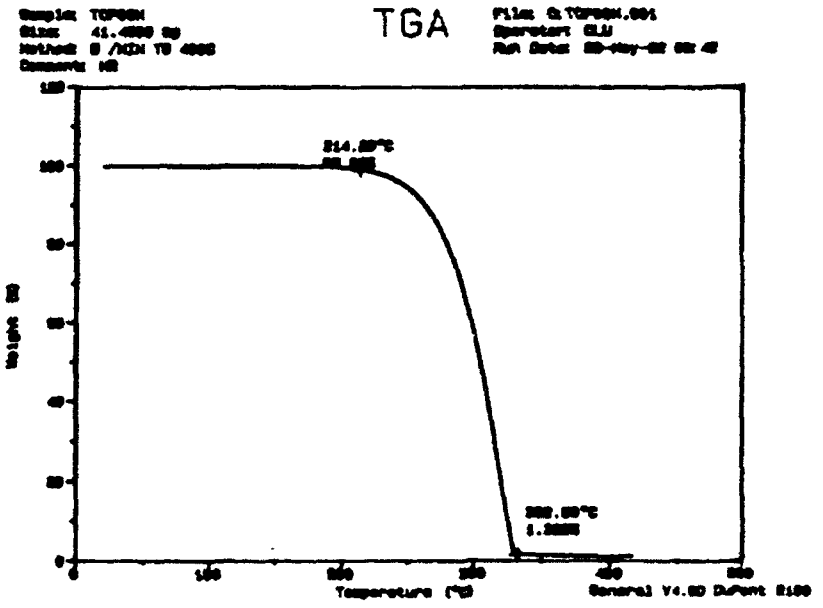
3. Differential Scanning Calorimetry (DSC)

The TAP solution of about 7 to 10 mg was placed in an aluminum sample pan and heated to 600°C at a rate of 10°C/min using the TA Instruments® model 910 DSC. The purpose of the DSC test was to compare the evaporation temperature and the heat of vaporization with those previously observed in the TGA test.

Figure 7 was one of the TGA tests to evaluate the evaporation temperature of the TCP solution, the solution was heated to 400°C at a rate of 5°C/min. The result shows that the TCP solution starts to evaporate at about 200°C and ends at about 325°C.



(air)



(nitrogen)

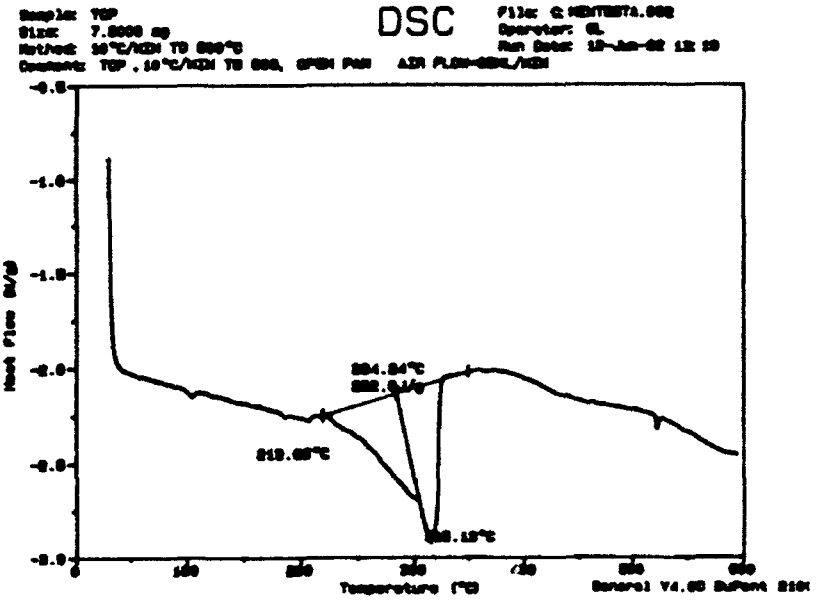
Fig.7: TGA Runs for Dry Air and Nitrogen (Evaporation of TCP)

The DSC results shown in Figure 8 indicate that the evaporation of TAP solution starts at about 215°C and ends at about 310°C, these values agree closely with the TGA results discussed in the previous section. The DSC results were also used to analyze and calculate the heat of vaporization of the TAP solution as:

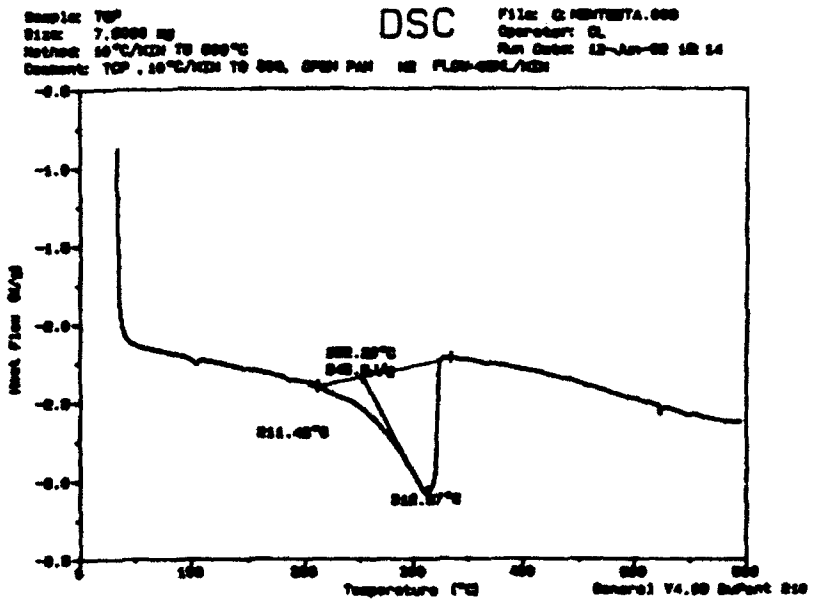
$$\Delta H_v = (222.6 \text{joule/g})(1 \text{ cal}/4.19 \text{ joule})(368 \text{ g/mol})$$
$$= 19.55 \text{ kcal/mol} \quad \text{(for dry air), and}$$

$$\Delta H_v = (245.8)(1/4.19)(368) = 21.59 \text{ kcal/mol (for nitrogen).}$$

In the previous section, the Arrhenius equation was used to calculate an activation energy of evaporation equal to 21.6 kcal/mol for dry air and 20.07 kcal/mol for nitrogen.



(air)



(nitrogen)

Fig.8: DSC Runs for Dry Air and Nitrogen (Evaporation and Heat of Vaporization)

CONCLUSION

From the thermal analysis using TGA and DSC we concluded that the TAP solution polymerizes when the solution is heated in an air or nitrogen environment. As the larger molecular weight molecules are produced during polymerization, the viscosity of the solution increases. When the TAP vapor condensate is used as a lubricant, it can be assumed that the lubricant is formed in the liquid phase. Therefore, the rheological characteristics of TAP polymeric solution can be used to study the tribological behavior of the TAP vapor condensate as a lubricant. For a lubrication process, the viscosity of a lubricant plays an important and complex role for in determining the traction coefficient of a tribo-system. For a highly viscous lubricant, a high operating temperature is required to reduce the viscosity of the lubricant, so as to reduce the traction coefficient, as well as the lubricant film thickness. However, the film thickness of a lubricant has to be great enough to maintain a lubrication process above the boundary lubrication region. This concept explains why the TAP vapor condensate or thermal deposition film is able to operate at high temperatures of about 670°C by lowering the traction coefficient to 0.02 at that point. Furthermore, when a synthetic polymeric lubricant is used at a high temperature and a high shear rate operating condition, such as in a high speed bearing, the lubricant forms a very thin hard viscoelastic film. When a lubricant changes to a viscoelastic material, the conventional concepts of rheology regarding viscosity and temperature can no longer be directly applied.

ACKNOWLEDGEMENTS

The authors wish to thank the Air Force Office of Scientific Research (AFOSR) 1992 Summer Faculty Research Program and the University of Dayton Faculty Sabbatical Program for making funds available for this research opportunity. We also thank Mr. Micheal A. Keller from the University of Dayton Research Institute (UDRI) for measurement of the molecular weight distribution of TAP residues. Thanks also are given to the engineering staff of the Lubrication Branch, Aero Proplusion and Power Directorate, Wright Laboratory, Air Force Materiel Command, Wright-Patterson Air Force Base, Ohio.

REFERENCES

1. Extended Temperature Range Lubrication with Vaporized Liquids. by Wedeven Associates,Inc. Final Report(6March 1991).

2. Formation of Lubricating Films at Elevated Temperatures From the Gas Phase. NIST Special Publication 744 by E.E.Klaus, J.L.Duda, and S.K. Naidu,(September 1988).

3. Vapor Phase Deposition on High Temperature Surfaces, presented at the 35th STLE/ASME Tribology Conference in Fort Lauderdale, Florida,by J.F.Makki and E.E.Graham,(October16-19,1989).

EFFECTS OF FREE-STREAM TURBULENCE AND SURFACE RIBLETS ON HEAT TRANSFER IN A LINEAR TURBINE CASCADE

Paul K. Maciejewski

Assistant Professor
Mechanical Engineering Department
University of Pittsburgh
Pittsburgh, PA 15261

Final Report for:
Aero Propulsion and Power Laboratory
Wright-Patterson AFB, OH 45433

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

July 1992

EFFECTS OF FREE-STREAM TURBULENCE AND SURFACE RIBLETS ON HEAT TRANSFER IN A LINEAR TURBINE CASCADE

Paul K. Maciejewski

Assistant Professor
Mechanical Engineering Department
University of Pittsburgh

ABSTRACT

The present study is an experimental investigation of the effects of free-stream turbulence and surface riblets on heat transfer in a linear turbine cascade. The primary goal of the study is to determine if surface riblets will reduce the average heat transfer rate in a cascade in the absence and in the presence of free-stream turbulence. A smooth, airfoil shaped, constant temperature heat transfer test surface was inserted into the AFIT linear cascade facility, where heat transfer tests were run at three levels of Reynolds number and two levels of free-stream turbulence. The heat transfer test surface was then removed from the facility so that riblets could be engraved on its surface. The newly ribleted heat transfer surface was then re-inserted into the cascade facility, where a second set of heat transfer tests were run at the same set of conditions used during the testing of the test surface while it was smooth. The test results indicate that, under certain conditions, surface riblets reduce the average heat transfer rate in the cascade by 6%.

EFFECTS OF FREE-STREAM TURBULENCE AND SURFACE RIBLETS ON HEAT TRANSFER IN A LINEAR TURBINE CASCADE

Paul K. Maciejewski

INTRODUCTION

Ribleted surfaces are believed to reduce surface heat transfer rates, under certain conditions. Epstein et al. (1991) report a maximum reduction in local heat transfer rates of 5% on a flat plate when the riblet height is approximately 5 to 15 wall units. However, riblet heights greater than 30 wall units increase heat transfer rates on a flat plate by up to 20%. Presumably, for a fixed physical riblet size, riblets may increase, decrease, or have no effect on the average heat transfer rate from a turbine blade, depending on the Reynolds number of the flow. The results of Epstein et al. are presented in Figure 1.

The results of Epstein et al. open the possibility of a modest but significant reduction of heat transfer on turbine blades if the blades are manufactured with surface riblets. However, two issues need to be considered further. The first is that heat transfer reduction occurs only for specific ratios for riblet height to sub layer thickness. If the riblet height is above this range, the heat transfer will be expected to increase. Given that the thickness of the boundary layer will vary along the length of a turbine blade, for a fixed riblet size the ratio of riblet height to sub layer thickness will also vary, so that the region of heat transfer reduction may be limited to only a portion of the surface. The second issue is that in actual engines one expects significant levels of free-stream turbulence to be present. The effects of free-stream turbulence on turbine blade heat transfer may eliminate any reductions in heat transfer that may occur on the blades due to riblets in the absence of free-stream turbulence.

OBJECTIVE

The primary objective of this study is to determine if surface riblets will reduce the average heat transfer rate in a linear cascade either in the absence and/or in the presence of free-stream turbulence.

DESIGN OF THE EXPERIMENT

The present experiments were designed to answer the following two questions: (1) Do riblets reduce the average heat transfer on a turbine blade when the approaching flow has no free stream turbulence?, and (2) Does the presence of free stream turbulence in the approaching flow alter the effects of riblets on the average heat transfer from a turbine blade?

An instrumented heat transfer surface was designed for testing in the AFIT cascade facility, shown schematically in Figure 2. A photograph of the test surface is provided in Figure 3. Since the maximum expected reduction in the heat transfer rate on the test turbine blade is only 5%, the experiment was designed to minimize the uncertainty in measured average Stanton number. The entire blade was 4.5 inches wide, but only the center 2 inches were used for heat transfer measurements in order to eliminate end wall effects. The outer portions of the blade were heated and served to guard the center portion of the blade against conduction losses. The same blade was used both for baseline testing with a smooth surface, then for further testing with a ribleted surface.

The blade was tested in the cascade facility prior to the engraving of the riblets, both with and without grid-generated free-stream turbulence (designated below by Grid and No Grid), and at three different Reynolds numbers (designated below by Low, Intermediate and High). After the testing with the smooth surface was completed, riblets were engraved on the test surface and it was placed back in the cascade facility to be tested at the same conditions at which the smooth surface was tested. Details of the cascade facility and the turbulence grid are provided by Galassi (1989). Low, Intermediate, and High Reynolds number correspond to 50,000; 125,000; and 350,000 based on the velocity at the inlet to the cascade and the chord of the blade.

At the High Reynolds number, one might expect the riblets to augment heat transfer, since the riblet height relative to the boundary layer thickest will be greater than y+=20 over most of the turbulent section of the boundary layer on the blade. At the Low Reynolds number, one might expect a possible reduction in the average heat transfer rate from the blade.

EXPERIMENTAL UNCERTAINTY

The primary response variable for the present study will be the ratio of two Stanton numbers measured on the same test surface (first smooth then with riblets) operated at the same conditions. Consequently, the only variables which contribute to the uncertainty of these ratios are the temperature difference between the test surface and the free-stream, the power supplied to the surface, the free-stream mean velocity, and the properties of air. The uncertainty in the difference in temperature between the heated surface and the free-stream is 1%. The power supplied to the test surface is the product of the voltage drop across the heating element in the surface and the current through the heating element in the surface. The uncertainty in the voltage drop across the heating element is 0.5%. The uncertainty in the current through the heating element is 0.5%. The resulting uncertainty in the power to the heating element is 0.7%. The uncertainty in the mean velocity is 3%. The uncertainty in the properties of air are assumed to be negligible. The resulting uncertainty in the ratio of two Stanton numbers is 4.5%

FLOW FIELD CHARACTERISTICS

In order to isolate the effect of surface riblets on heat transfer in the cascade facility, the test facility was set up to operate under the same conditions for the test runs on the surface with riblets as it was for the test runs on the surface without surface riblets. The smooth and ribleted test surfaces are to be compared only under the same test conditions, i.e., under the influence of an inlet flow field with the same characteristics.

Table 1, below, shows that for the cases of Intermediate and High Reynolds number, the free-stream conditions at the inlet of the cascade upstream of the test surface for the runs with the ribleted surface closely match the free-stream conditions at the inlet of the cascade upstream of the test surface for the runs when the heat transfer surface was smooth. However, for the Low Reynolds number configuration, there is a significant difference in the free-stream conditions between the runs with surface riblets and the runs without surface riblets. For Low Reynolds number and No Turbulence Grid, the

smooth surface experienced a 1.4% level of free-stream turbulence while the ribleted surface experienced a 5.1% level of free-stream turbulence. For Low Reynolds number and the Turbulence Grid, the smooth surface experienced a 16.5% level of free stream turbulence while the ribleted surface experienced only a 12.3% level of free-stream turbulence. These significant discrepancies in the behavior of the test facility were due to modifications that were made to the test facility between the block of experiments with the smooth surface and the block of experiments with the ribleted surface. As a result, only the comparisons to be made at Intermediate and High Reynolds number may be interpreted unambiguously. Any comparisons to be made for Low Reynolds number must note the discrepancies in free-stream turbulence conditions between tests on the smooth and tests on the ribleted surface.

Table 1 Nominal Turbulence Intensity for Experimental Runs

Reynolds number	Tu Grid	Riblets	Tu (%)
Low	no	no	1.4
		yes	5.1
	yes	no	16.5
		yes	12.3
Intermediate	no	no	0.8
		yes	1.2
	yes	no	12.3
		yes	11.6
High	no	no	1.4
		yes	1.2
	yes	no	10.5
		yes	10.7

Figure 4a shows the profile of mean velocity across the inlet to the cascade upstream of the test surface for test runs with and without surface riblets for the case designated High Reynolds number. The two profiles are identical to within the experimental uncertainty. Figure 4b shows the profile of mean velocity across the exit of the cascade downstream of the test surface for the same test configurations given in Figure 4a. Again, the two profiles are nearly identical to each other.

Figure 5a shows the profile of turbulence level across the inlet to the cascade upstream of the test surface for test runs with and without surface riblets for the case designated High Reynolds number. The two profiles are identical to within the experimental uncertainty. Figure 5b shows the profile of turbulence level across the exit of the cascade downstream of the test surface for the same test configurations given in Figure 5a. Again, the two profiles are nearly identical to each other.

Figure 6a shows the distribution of u-component velocity (normalized by the standard deviation in the sample) upstream of the stagnation point of the test surface for test runs with and without surface riblets for the case designated High Reynolds number. The two PDF's are identical to within the experimental uncertainty, and both compare well with the gaussian distribution also shown in Figure 6a. Figure 6b shows the distribution of u-component velocity downstream of the trailing edge of the test surface for the same test configurations given in Figure 6a. Again, the two PDF's are identical to within the experimental uncertainty, although each of the PDF's appears to be slightly non-gaussian.

Figure 7a shows the normalized power spectral density function of the u-component velocity upstream of the stagnation point of the test surface for test runs with and without surface riblets for the case designated High Reynolds number. The two PSD's are identical to within the experimental uncertainty. Figure 7b shows the normalized power spectral density function of the u-component velocity downstream of the trailing edge of the test surface for the same test configurations given in Figure 7a. The two PSD's in Figure 7b are similar, although there is some indication that the presence of the riblets reduces the power in the high frequency portion of the spectrum.

The comparisons of the characteristics of the free-stream velocity field at the inlet of the cascade for the test runs with and without surface riblets for the cases designated Intermediate (not shown here in detail) and High Reynolds number establish that any differences in heat transfer between the test runs with and without surface riblets should be attributed to the absence or presence of the surface riblets and not to other factors.

HEAT TRANSFER RESULTS

The heat transfer surface was operated by supplying power to a resistance heating element embedded in an aluminum turbine blade. Figure 8 displays the temperature distribution along the length of the blade for test runs with and without surface riblets for the a case designated High Reynolds number. The temperature distributions for the two cases shown in Figure 8 are nearly identical to each other. This agreement suggests that any effect of surface riblets on heat transfer is distributed more or less uniformly over the length of the blade. The agreement between the data sets in Figure 8 also suggest that conduction and radiation heat transfer loss mechanisms experienced by the test surface will be the same for the surface both with and without surface riblets.

Table 2, below, provides a comparison of Stanton number in the presence of free-stream turbulence to Stanton number in the absence of free-stream turbulence. Free-stream turbulence increases heat transfer in the cascade for all Reynolds numbers and for both smooth and ribleted surfaces. Also, at High Reynolds number, the increase in Stanton number associated with free-stream turbulence is significantly higher for the surface with riblets. Recall that the cases at Low Reynolds number should be disregarded, since the case with riblets is most likely grossly underestimated due to the peculiarities in the test facility noted above.

Table 2 Ratio of St with Turbulence Grid to St without Turbulence Grid

Reynolds number	Smooth	Riblets
Low	1.29	1.16
Intermediate	1.24	1.28
High	1.34	1.41

Table 3, below, provides a comparison between Stanton number on the test surface with surface riblets and Stanton number on the smooth test surface. If one only considers the cases designated by Intermediate and High Reynolds number, the only significant effect of riblets is to reduce the heat transfer

rate by 6% for the case Intermediate Reynolds number and No Grid. The table does indicate that there is a 3% reduction in heat transfer associated with surface riblets for the case Intermediate Reynolds number and Grid, and a 4% increase in heat transfer associated with surface riblets for the case High Reynolds number and Grid. However, each of these would-be effects fall within the +/- 4.5% uncertainty interval associated with the ratio of two Stanton numbers in this study.

Table 3 Ratio of St with Surface Riblets to St without Surface Riblets

Reynolds number	No Grid	Grid
Low	0.99	0.89
Intermediate	0.94	0.97
High	0.99	1.04

The heat transfer results described above are captured in Figure 9. Recall that the case designated Low Reynolds number, No Grid, and Riblets actually corresponds to an inlet flow with a free-stream turbulence level of 5.1% as opposed to an inlet flow with a free-stream turbulence level of 1.4% for the case designated Low Reynolds number, No Grid, and Smooth Also, recall that the case designated Low Reynolds number, Grid, and Riblets actually corresponds to an inlet flow with a free-stream turbulence level of 12.3% as opposed to an inlet flow with free-stream turbulence level of 16.5% for the case designated Low Reynolds number, Grid, and Smooth.

DISCUSSION

The presents study shows that surface riblets will reduce heat transfer in a turbine cascade, under certain conditions. The clearest case of heat transfer reduction is illustrated by the runs designated Intermediate Reynolds number and No Grid, where the heat transfer rate on the ribleted surface is only 94% of the heat transfer rate on the smooth surface. It also seems reasonably clear that increases in heat transfer associated with free-stream turbulence are higher for a surface with riblets than for a smooth

surface at fixed chord Reynolds number. To see this latter point, compare the entries in each of the two columns in Table 2 for Intermediate and High Reynolds number. This effect might be explained by the thinning of the boundary layer typically associated with free-stream turbulence, thereby increasing the ratio of riblet height to boundary layer thickness at fixed chord Reynolds number, which in turn augments heat transfer.

It is unfortunate that the cases designated Low Reynolds number needed to be discarded in the present study, since it is likely that these case would have shown the largest reductions in heat transfer to be associated with surface riblets. The test facility should be debugged and the cases with surface riblets should be re-run.

CONCLUSIONS

In the present investigation, surface riblets have been observed to reduce the average heat transfer rate on a turbine blade in a linear cascade by 6% for an inlet flow without free-stream turbulence and at a chord Reynolds number of 125,000. The present data also suggest that it may be possible that riblets reduce heat transfer on a turbine blade in the presence of a flow with free-stream turbulence.

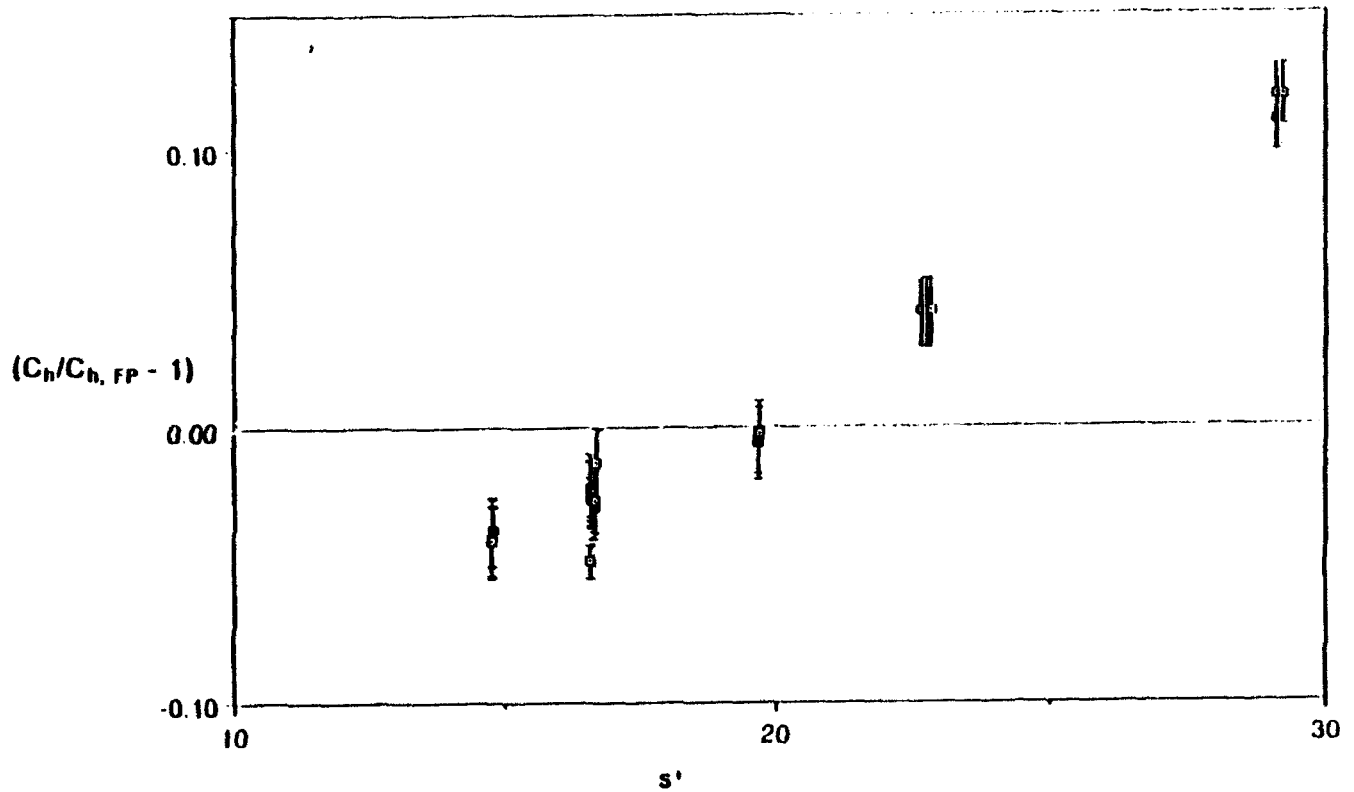
ACKNOWLEDGMENTS

Use of the cascade test facility was facilitated by Dr. Paul King of the Air Force Institute of Technology. The test surfaces were manufactured by Mr. Dave Driscoll, also associated with the Air Force Institute of Technology.

REFERENCES

Galassi, L., 1989, Turbulence Scale Effects on Heat Transfer in a Linear Turbine Cascade, MS Thesis, AFIT.

Epstein, A.H., Guenette, G.R., Stone, T.D., and Steptoe, W.J., 1991, The Effect of Riblets on Heat Transfer in the Incompressible Turbulent Boundary Layer over a Flat Plate, Final Report on Contract F-33615-87-C-2729, Section I.



Effect of Riblets on Integral Stanton Number. The graph above shows the effect of the riblets used in this experiment on the total heat transfer coefficient of the test sections versus riblet spacing in wall units. The results show heat transfer reduction below s+ of 20 with maximum reduction of 5% at 15 wall units. The general trends of the effect of riblets on heat transfer are the same as those on drag.

Figure 1 Effects of surface riblets on heat transfer on a flat plate. [from Epstein et al. (1991)]

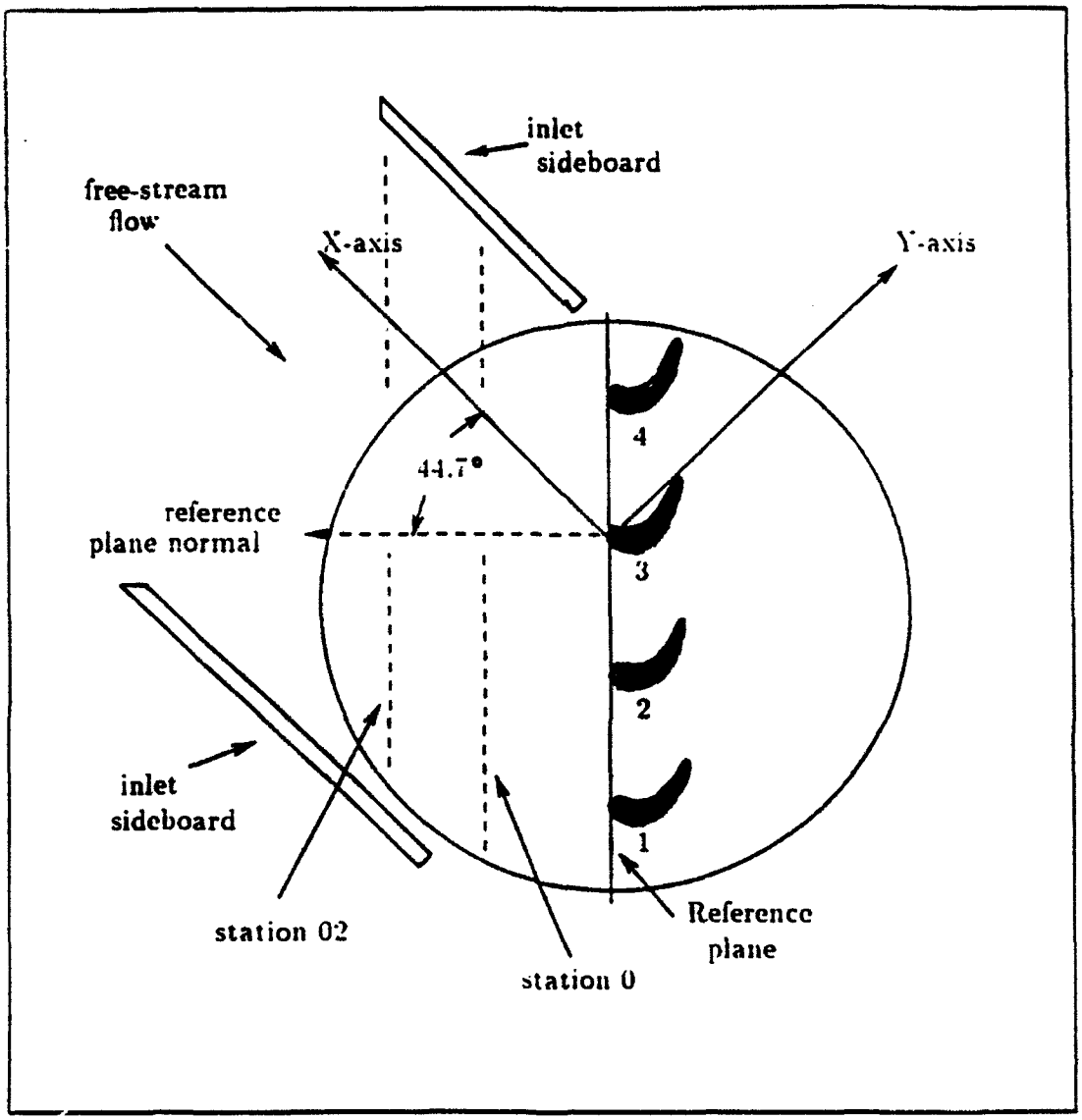


Figure 2 Schematic of the AFIT linear cascade test facility. [from Galassi (1989)]

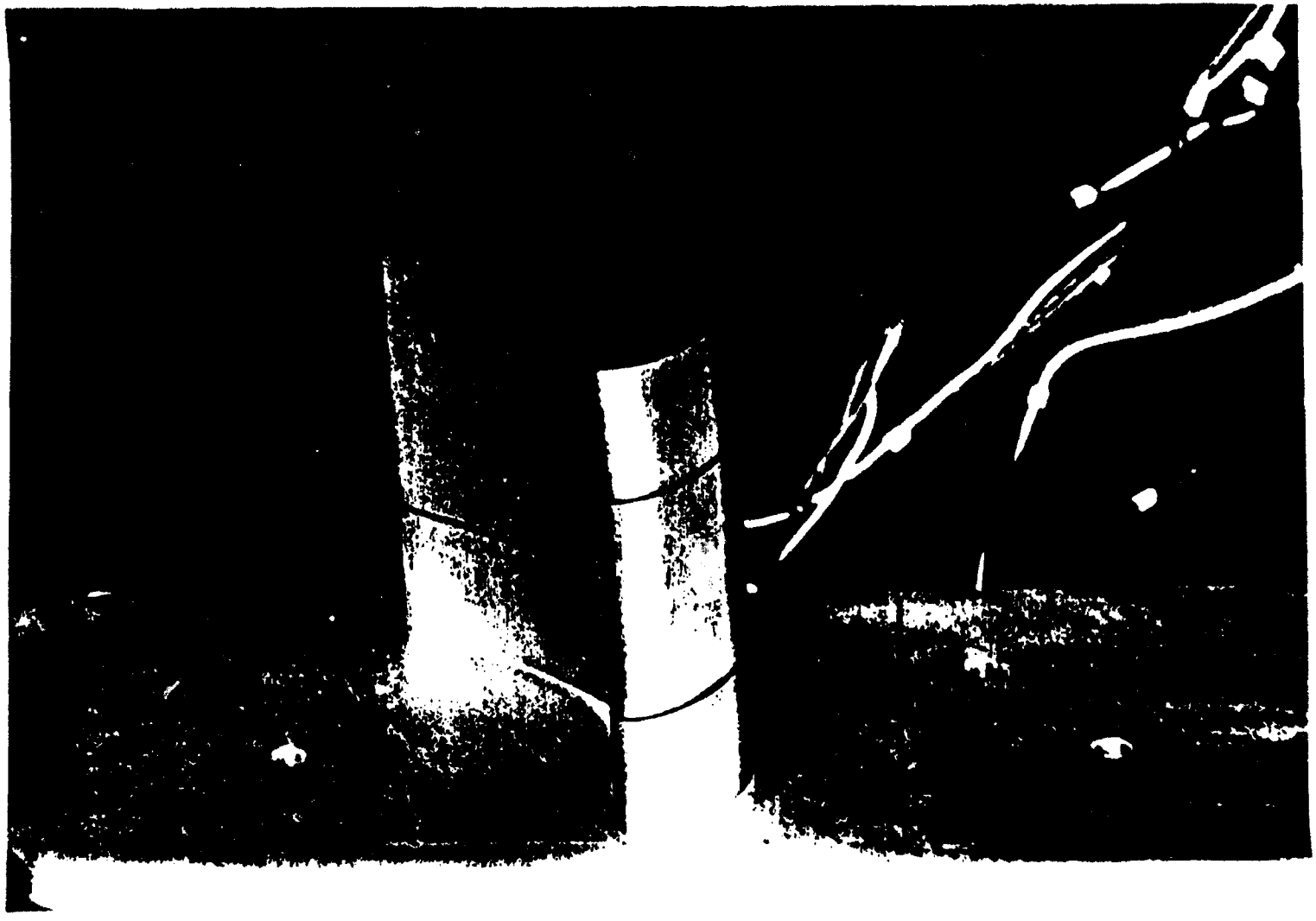


Figure 3 Photograph of the test surface prior to the engraving of the riblets.

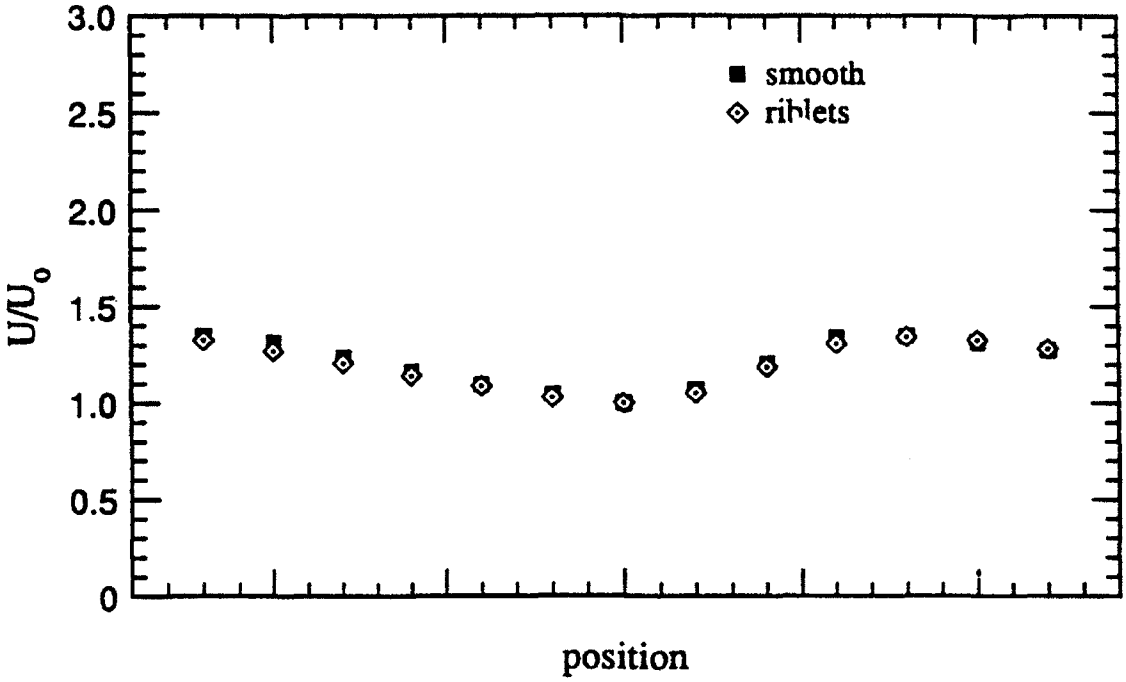


Figure 4a Mean velocity upstream of the test surface.

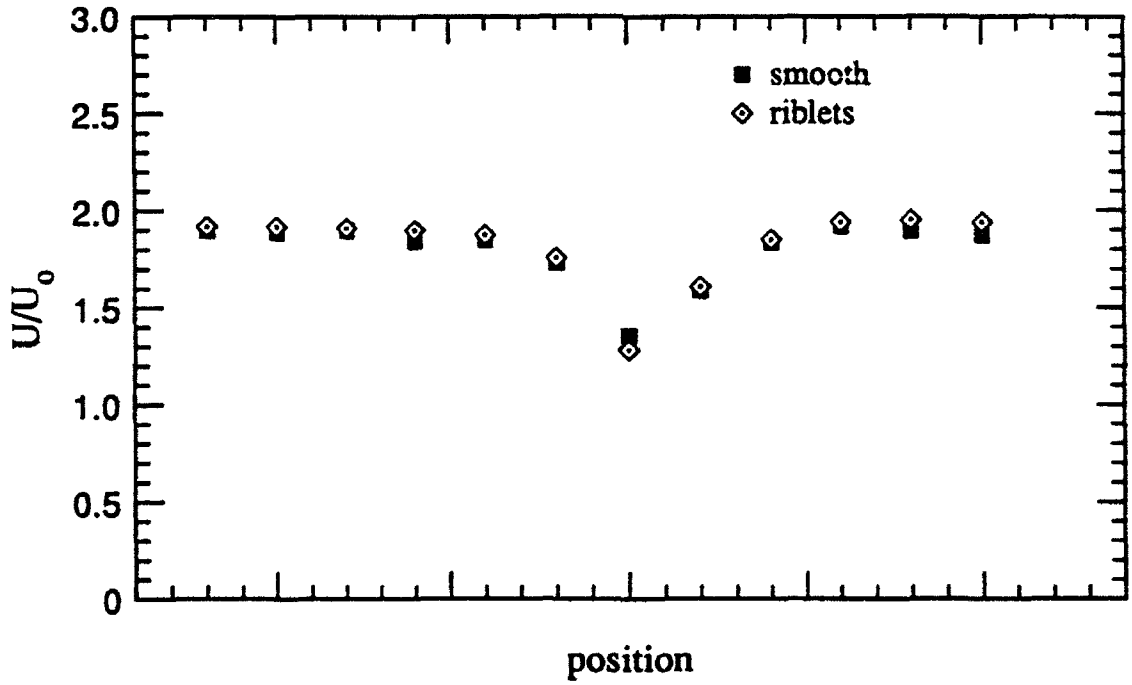


Figure 4b Mean velocity downstream of the test surface.

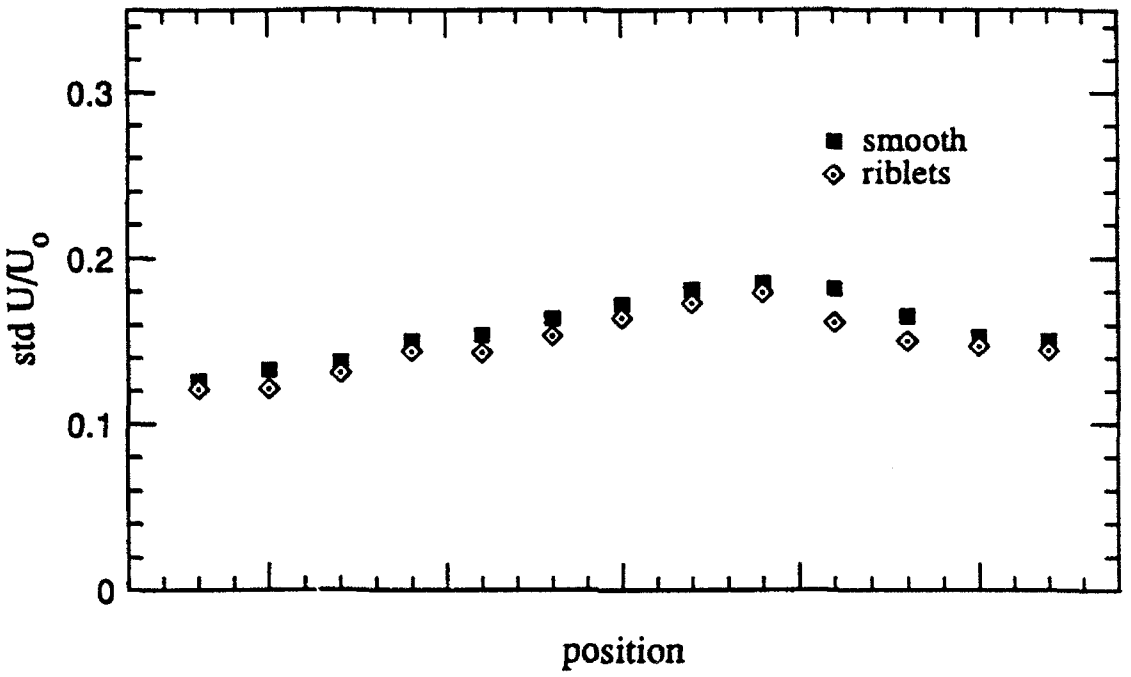


Figure 5a Turbulence level upstream of the test surface.

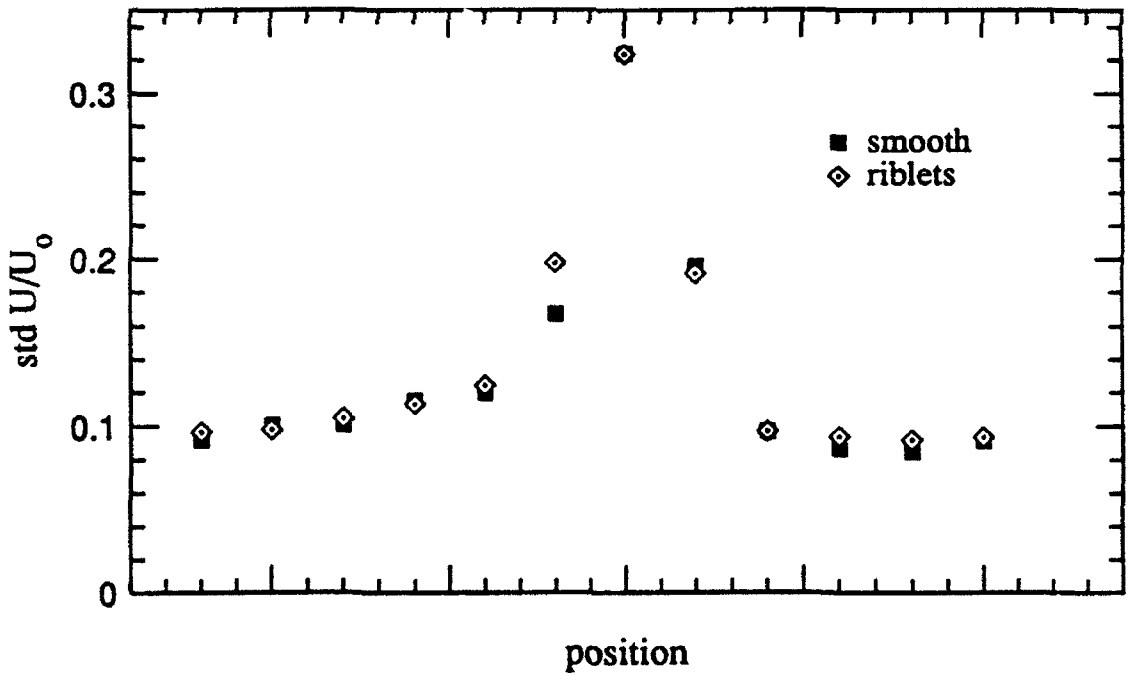


Figure 5b Turbulence level downstream of the test surface.

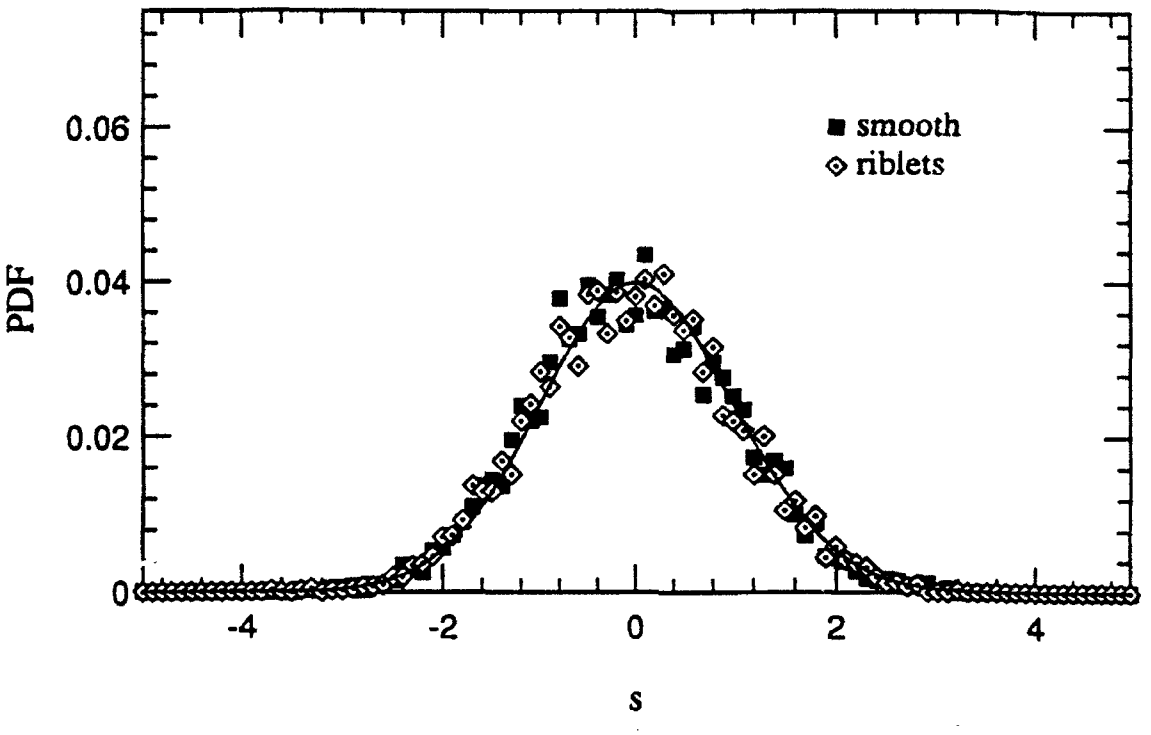


Figure 6a Distribution of velocity upstream of the stagnation point.

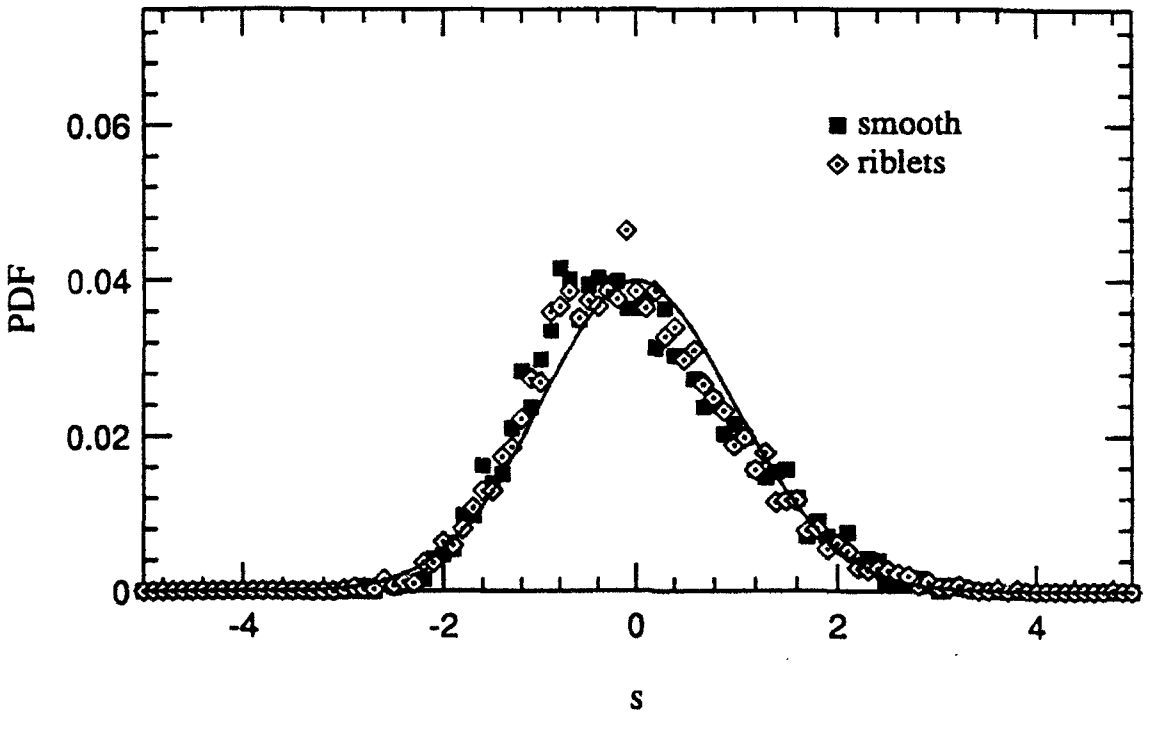


Figure 6b Distribution of velocity downstream of the trailing edge.

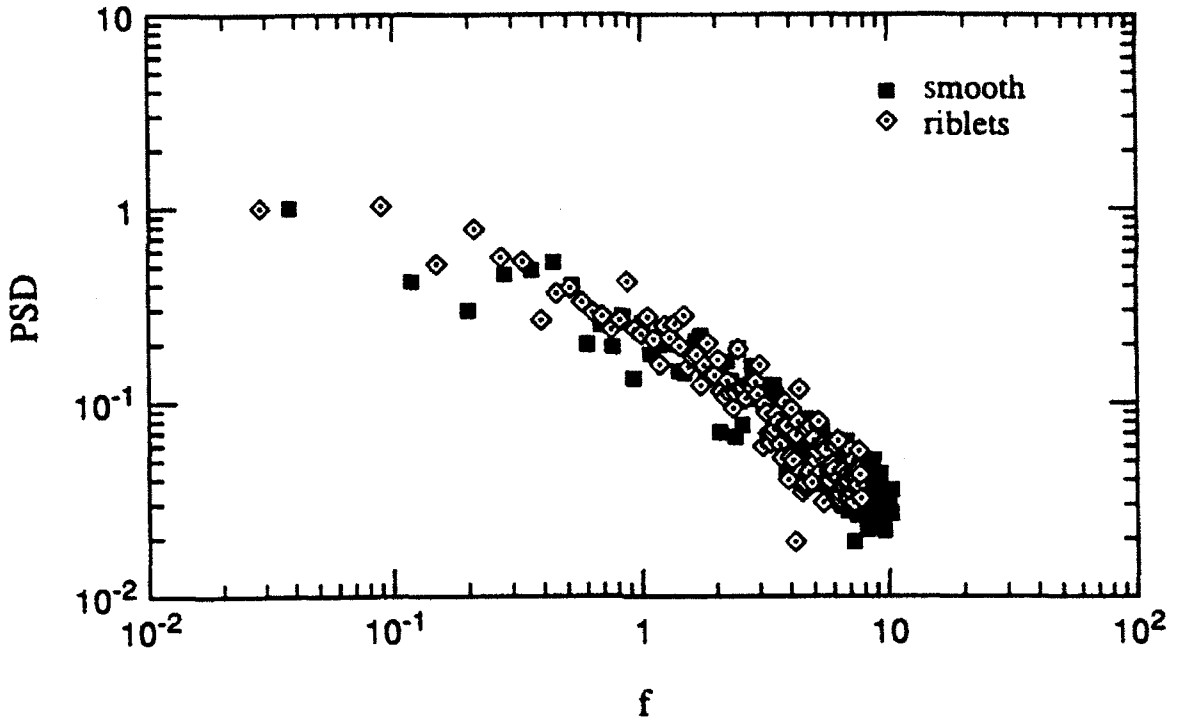


Figure 7a Power spectrum of velocity upstream of the stagnation point.

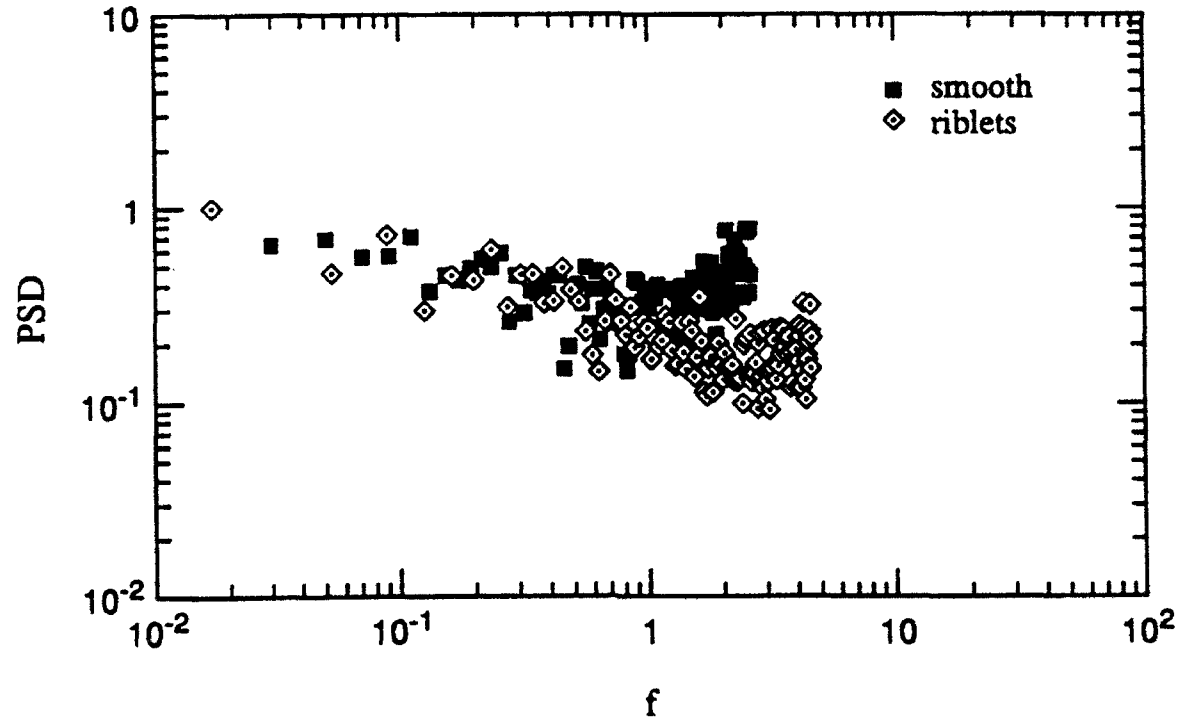


Figure 7b Power spectrum of velocity downstream of the trailing edge.

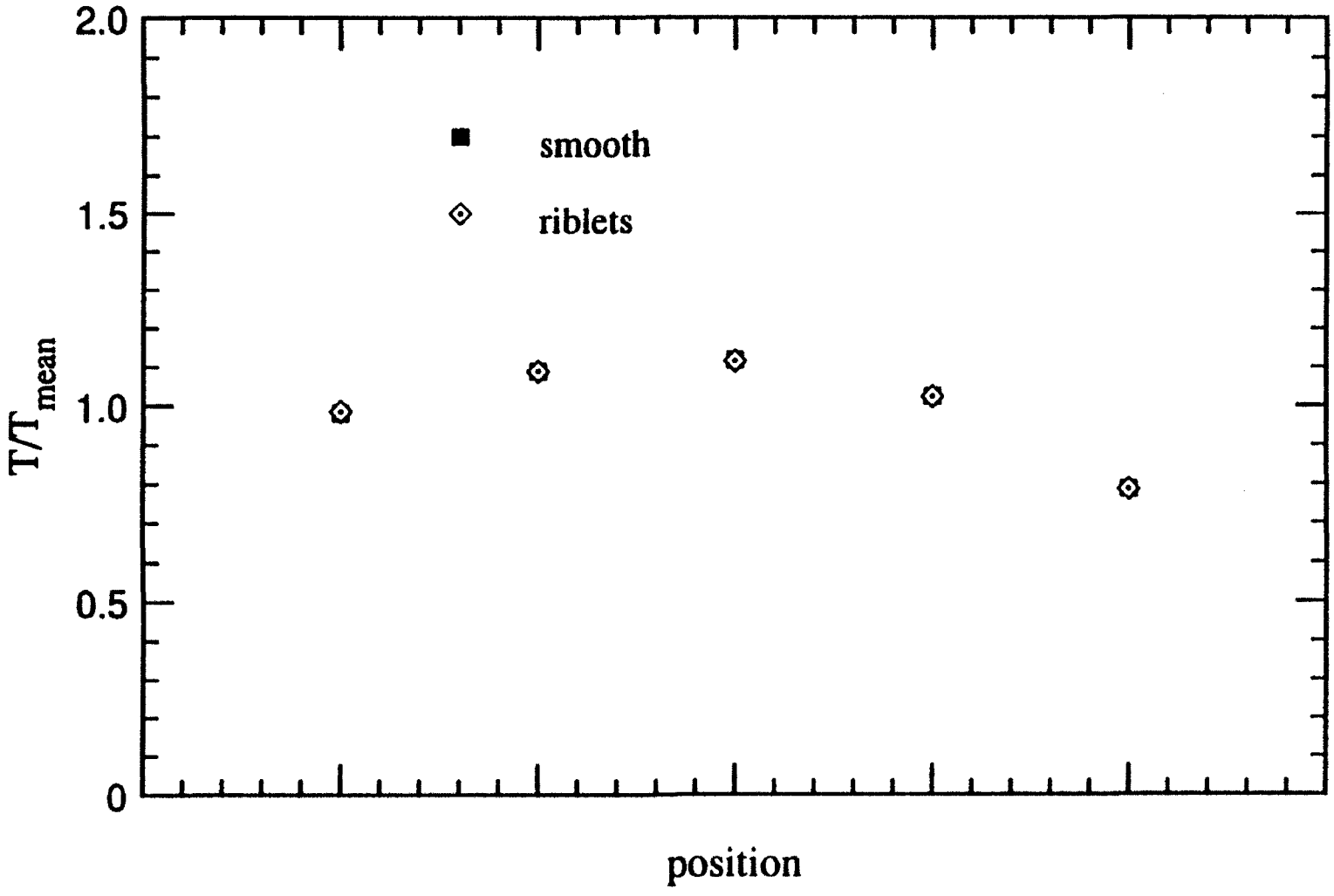


Figure 8 Temperature distribution along the length of the test surface.

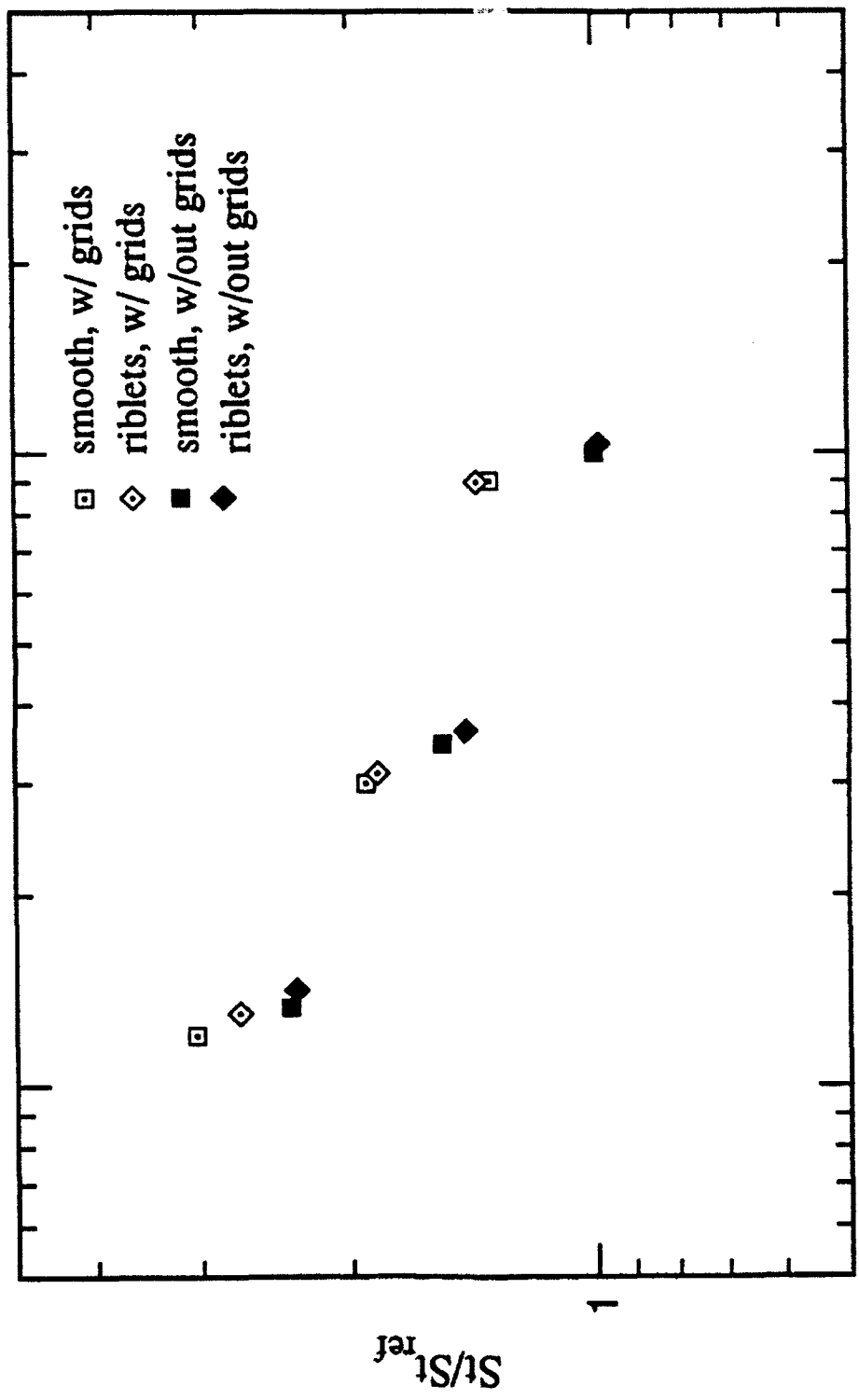


Figure 9 The effects of Reynolds number, free-stream turbulence and surface riblets on heat transfer.

TRADE-OFF ANALYSIS OF SENSOR FUSION METHODOLOGIES FOR AN X-BAND AND W-BAND RADAR SENSOR SUITE

Charlesworth R. Martin
Associate Professor
Department of Physics and Engineering
Norfolk State University
Norfolk, VA 23504

Final Report for:
Summer Research Program
Air Force Armament Laboratory

Sponsored by:
Air Force Office of Scientific Research
Eglin Air Force Base, Florida

September 1992

TRADE-OFF ANALYSIS OF SENSOR FUSION METHODOLOGIES FOR AN X-BAND AND W-BAND RADAR SENSOR SUITE

Charlesworth R. Martin
Associate Professor
Department of Physics and Engineering
Norfolk State University

Abstract

Various methods for implementing multi-sensor fusion for a dual mode RF (Radio Frequency) sensor suite were reviewed. The primary objective being to ascertain which methodology holds the most promise for practical implementation. The criteria for best being (a) Performance capability, (b) Simplicity of algorithm, (c) Ease of implementation, and (d) The stage of the process at which it makes possible for fusion to take place. Preliminary results seem to suggest that the Bayesian Fusion methodology is the scheme that emerges as the most practical. The paper concludes with hardware implementation for the dual RF Bayesian sensor system fusion algorithm.

TRADE-OFF ANALYSIS OF SENSOR FUSION METHODOLOGIES FOR AN X-BAND AND W-BAND RADAR SENSOR SUITE

Charlesworth R. Martin

underbarINTRODUCTION

The demand for a more robust medium range air-to-air missile system with optimized target detection and tracking capability has given impetus to Seeker Designers to investigate the capabilities and processing requirements of various multi-mode sensor designs. Previous works have focused on sensor selection trade-off between several levels of fusion, and the use of widely dissimilar sensor suite such as infrared (IR) and radar sensors in a quest to improve performance and enhance immunity to countermeasures [1,2,3,4].

The dual RF-RF sensor suite has both advantages and disadvantages over a dual mode IR-RF sensor suite. The major advantages are:

(a) Reduced processing requirements

(b) Minimal impact on radome geometry since both RF-sensors are likely to share a common aperture.

(c) Fusion can be more readily implemented at any stage of the fusion process.

(d) Less complicated fusion algorithms.

The major disadvantage is that the RF-RF sensor suite, in principle, has less immunity to countermeasures because it can be defeated by RF jamming, whereas the IR-RF sensor suite would reqiuire both RF and flares.

The purpose of this report is to (a) Present a comparison of different fusion algorithms for the RF-RF sensor suite, (b) To discuss the advantages and challenges of fusion at the feature and decision levels, and (c) To give hardware inplementation for the Bayesian

sensor system fusion algorithm.

underbarDETECTION AND TRACKING

The two most important tasks that a missile seeker for medium range air-to-air engagement must perform are detection and tracking. Detection is the process of testing the observed vector **R** against established hypotheses to determine which of the hypotheses most closely matches the observed vector. Tracking is an estimation problem. The parameter E(t) is determined, in some optimal sense, based on the observed vector **R** as represented in Figure 1. The feature estimated could be range or range-rate.

All detection schemes embody a priori target and clutter characteristics. This may be in the form of probability density functions, rules and confidences, trained adaptive weights, or optimal weighting matrices.

The detection problem can be formulated as binary hypotheses testing (as summarized by Van Trees [5] where two probability distributions represent the two possible states of the observed vector **R**, signal present or absent. The decision process must decide one of two possible hypotheses, H_0 (no signal present) and H_1 (signal present). The a priori probabilities for these cases are represented by P_0 and P_1, respectively. A decision criterion for the signal receiver (sensor) must provide for the four possible outcomes [6]:

Outcome	True State	Decision Made	Description of Decision
a	H_0	H_0	Correct, no signal
b	H_0	H_1	Error, false alarm
c	H_1	H_0	Miss error
d	H_1	H_1	correct detection

The decision rule partitions the two distributions into four regions whose areas are the conditional probabilities of the outcomes. For each decision a cost is established and the average cost over all four decisions is minimized to derive the classical Bayesian criteria. The four costs in the binary case, C_{00}, C_{01}, C_{10}, and C_{11} are assigned to outcomes a, b, c, and d, respectively. The risk C is defined as the sum of the product of cost and conditional probabilities for each outcome:

$$C = C_{00}P_0P_a + C_{01}P_0P_b + C_{10}P_1P_c + C_{11}P_1P_d \tag{1}$$

The cost function is partitioned into fixed and variable terms. The components of the variable terms are separated to establish the test at which cost is minimized. This occurs when the ratio of the values of the two conditional density functions (the likelihood ratio) is related to a constant (the threshold test η) i.e.,

$$\Lambda(\mathbf{R}) = \frac{P(\mathbf{R}/H_1)}{P(\mathbf{R}/H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{2}$$

$$\eta = \frac{P_0(C_{10} - C_{00})}{P_1(C_{01} - C_{11})} \tag{3}$$

The likelihood ratio is a scalar function, and it is often more convenient to use the logarithm of this ratio and threshold test for calculation of the decision threshold.

The result (2) can be used to implement post-detection fusion of the RF-RF sensor suite as shown in figure (2) or pre-detection fusion as shown in figure (3). For pre-detection fusion the sensor report vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ are concatenated to form the combined report ($\mathbf{r}$, i.e., $\mathbf{r}(\mathbf{r}_1, \mathbf{r}_2)$.

The results thus far presented can be expanded in a straight forward way to handle the multiple-observation space.

SIMPLIFIED IMPLEMENTABLE MULTISENSOR DETECTION ALGORITHM

The simplified Independent Detection Algorithm is a form of the likelihood test that avoids the Goussman assumption, but is limited to at most second order statistics [4]. The target estimate $\hat{\mathbf{d}}$ is formed by linearly operating on the received vector $\mathbf{R}$ with a weighting matrix $\mathbf{W}$,

$$\mathbf{W}\vec{\mathbf{R}} = \vec{\mathbf{d}}^1 \tag{4}$$

The weighting matrix is found by constructing the error between the actual desired signal, $\mathbf{d}$, and the desired target estimate, $\hat{\mathbf{d}}$,

$$\vec{\epsilon} = \vec{\mathbf{d}} - \mathbf{W}\vec{\mathbf{R}} \tag{5}$$

The vector error is minimized by minimizing the trace of its outer product (its correlation matrix),

$$t_r\left[\vec{\epsilon}\vec{\epsilon}^T\right] = t_r\, E\left[(\vec{\mathbf{d}} - \mathbf{W}\vec{\mathbf{R}})(\vec{\mathbf{d}}^T - \vec{\mathbf{R}}^T\mathbf{W}^T)\right] \tag{6}$$

The notation is simplified by letting the correlation matrix between the received vector, $\mathbf{R}$, and the desired target signal, $\mathbf{d}$, be

$$\begin{aligned} \mathbf{S}_{rd} &= E[\vec{\mathbf{R}}\vec{\mathbf{d}}^T] \\ \mathbf{S}_{rd}^T &= E[\vec{\mathbf{d}}\vec{\mathbf{R}}^T] \end{aligned} \tag{7}$$

and the autocorrelation of the received vector be

$$\mathbf{S}_{rr} = E[\vec{\mathbf{R}}\vec{\mathbf{R}}^T] \tag{8}$$

and the autocorrelation matrix of the desired target vector be

$$\mathbf{S}_{dd} = E[\vec{\mathbf{d}}\vec{\mathbf{d}}^T] \tag{9}$$

Evaluation of equation (6) with these definitions and finding the trace of the error correlation matrix, gives

$$t_r E[\vec{\epsilon}\vec{\epsilon}^T] = t_r\left(\mathbf{S}_{dd} - \mathbf{W}\mathbf{S}_{rd} - \mathbf{S}_{rd}^T\mathbf{W}^T + \mathbf{W}\mathbf{S}_{rr}\mathbf{W}^T\right) \tag{10}$$

Minimizing this function with respect to the weighting matrix $\mathbf{W}$, gives

$$\frac{\partial}{\partial \mathbf{W}} t_r E\left[\bar{\epsilon}\bar{\epsilon}^T\right] = 0 - \mathbf{S}_{rd} - \mathbf{S}_{rd}^T + 2\mathbf{W}\mathbf{S}_{rr} \tag{11}$$

which after solving for $\mathbf{W}$, leads to an optimal weighting matrix solution

$$\mathbf{W} = \frac{1}{2}\left(\mathbf{S}_{rd} + \mathbf{S}_{rd}^T\right)\mathbf{S}_{rr}^{-1} \tag{12}$$

The weighting matrix solution $\mathbf{W}$, is comprised of correlation matrices $\mathbf{S}_{rr}$ and $\mathbf{S}_{rd}$. $\mathbf{S}_{rr}$ represents the interdependence of the received vector components;

$$\mathbf{S}_{rr} = \begin{bmatrix} \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & E[\mathbf{R}_{i-1}\mathbf{R}_{i-1}] & E[\mathbf{R}_{i-1},\mathbf{R}_i] & E[\mathbf{R}_{i-1},\mathbf{R}_{i+1}] & \cdots \\ \cdots & E[\mathbf{R}_i,\mathbf{R}_{i-1}] & E[\mathbf{R}_i,\mathbf{R}_i] & E[\mathbf{R}_i,\mathbf{R}_{i+1}] & \cdots \\ \cdots & E[\mathbf{R}_{i+1},\mathbf{R}_{i-1}] & E[\mathbf{R}_{i+1},\mathbf{R}_i] & E[\mathbf{R}_{i+1},\mathbf{R}_{i+1}] & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix} \tag{13}$$

where, $\mathbf{R}_i$ is the *ith* component of the received vector, with the correlation function $\mathbf{R}_i$ to $\mathbf{R}_j$ given as

$$E[\mathbf{R}_i,\mathbf{R}_j] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbf{R}_i \cdot \mathbf{R}_j \cdot f(\mathbf{R}_i,\mathbf{R}_j)\mathbf{dR}_i\mathbf{dR}_j \tag{14}$$

where $f(\mathbf{R}_i,\mathbf{R}_j)$ in the joint probability density function between the *ith* and *jth* components of the received vector $\mathbf{R}$.

Although the fusion algorithm (4) has strong formalism, it is quite difficult in principle to implement because of the need to ascertain the joint probability density functions $f(\mathbf{R}_i,\ \mathbf{R}_j)$ between the *ith* and *jth* components of the received vector, $\mathbf{R}$. The appeal for this algorithm, however, in bolstered by its apparent insensitivity to non-commeasurate sensor suites.

ARTIFICIAL INTELLIGENCE TECHNIQUES

Some of the most recently proposed sensor fusion techniques reviewed have moved away from classical statistically based approaches. This is to void the need for joint conditional probability density functions to describe targets and clutter backgrounds. These techniques are variously called, Neural Networks, Fuzzy Set Techniques, Knowledge-Based Approaches, and are cloaked under the umbrella of AI techniques. These techniques lack the formalism so evident with the statistically based approaches. Their development are not sufficiently matured to warrant serious consideration at this time.

DEMPSTER-SHAFER

The Dempster-Shafer formalism circumvents the requirement by the Bayesian approach for prior probabilities that to some prove troublesome to attain. With Dempster-Shafter formalism, all initial beliefs can be assigned to the uncommitted state and redistributed by its rule as new evidence arrives. Formalism for decision making using functions and other measures of uncertainty are not yet sufficiently developed to guide practical sensor fusion designs [6].

CONCLUSION

The axiom-based decision logic of probability theory gives the Bayesian approach a definitive edge in guiding practical sensor design. The decision to fuse decisions as in figure (2) or feature vectors as in figure (3) requires detail study. At issue are: cost of implementation one approach versus the other, performance trade-offs, and difficulty of implementation.

ACKNOWLEDGEMENTS

I wish to express my gratitude to Martin Moorman, Larry E. Lewis, Ellis J. Bondoeau Jr., and Dr. Dennis Goldstein. The camaraderie with the other summer faculty research fellows all contributed to make my summer at the Air Force Armament Laboratory's Advanced Guidance Division, Eglin AFB, an enjoyable one.

Finally, I gratefully acknowledge the financial support received from the USAF Summer Research Program in the form of a Summer Faculty Award.

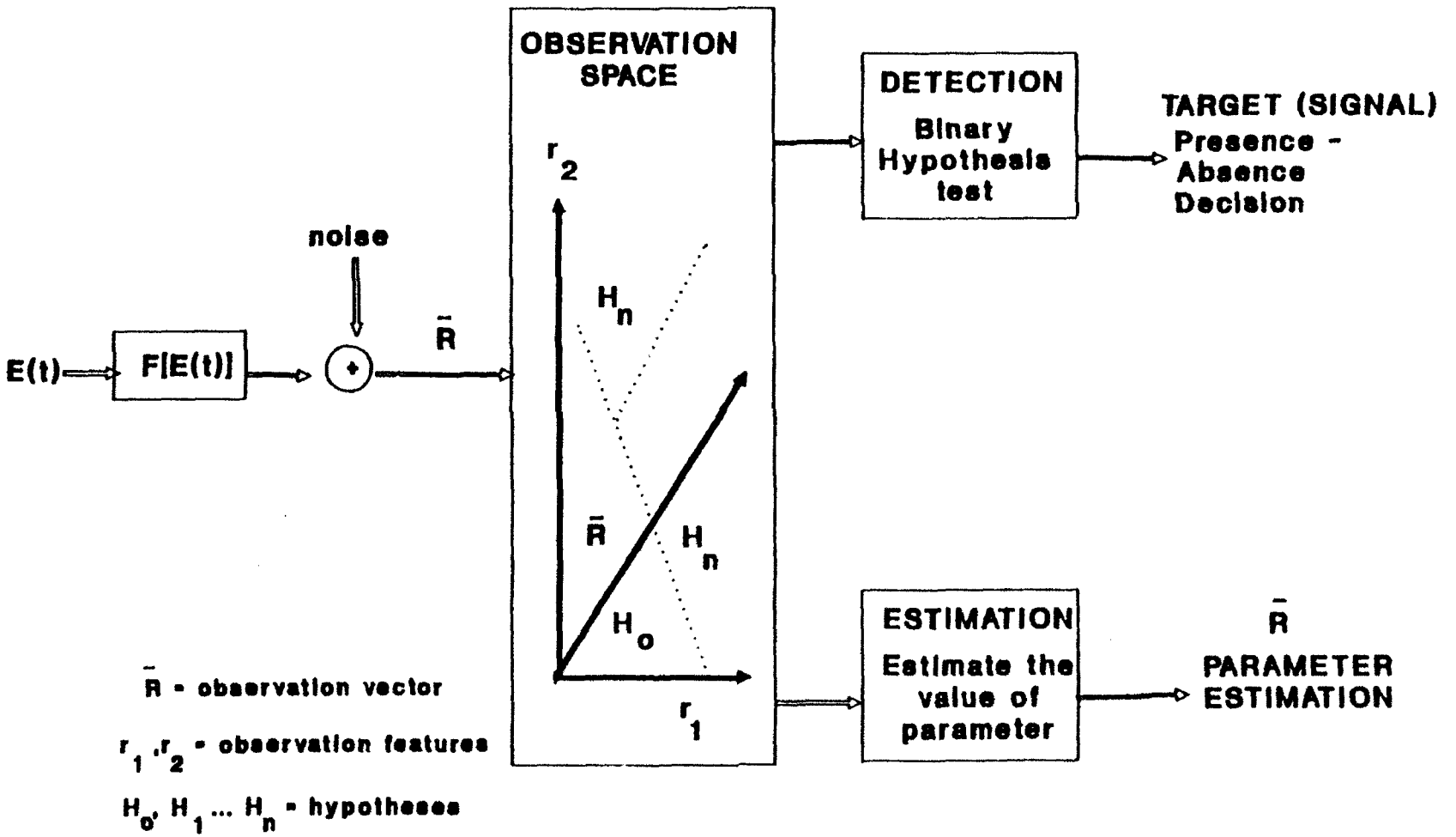


FIGURE 1. MODEL OF SENSOR FUNCTION

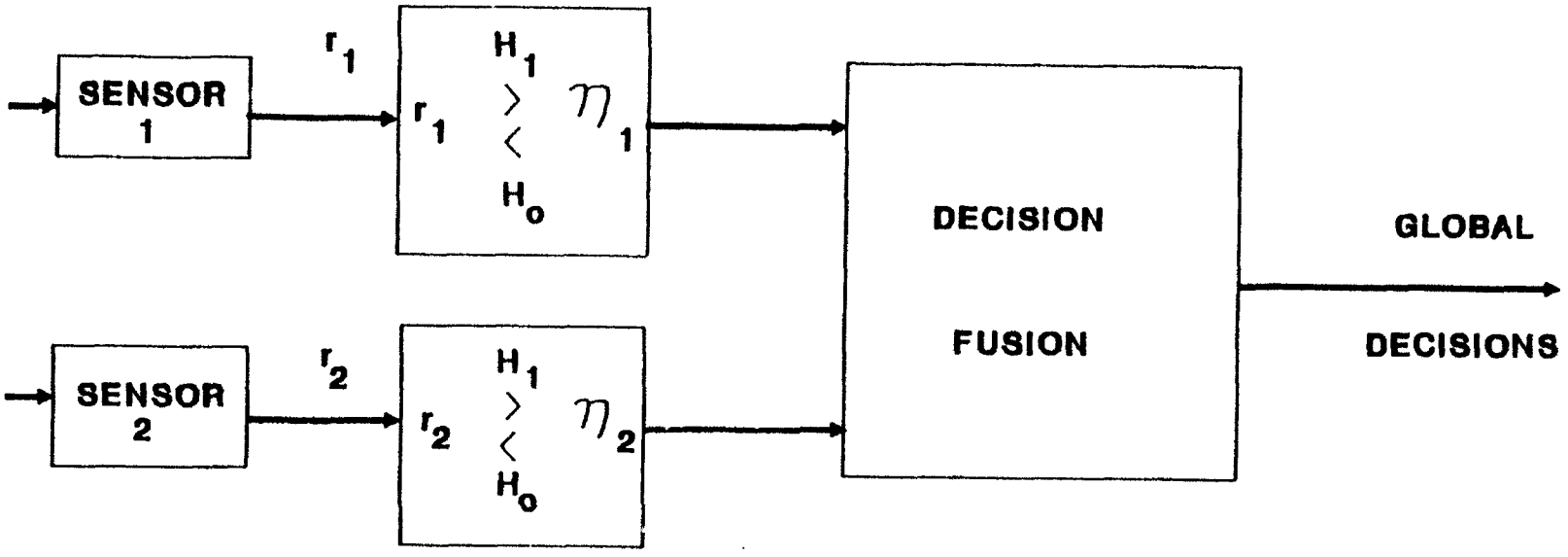


FIGURE 2. POST-DETECTION FUSION

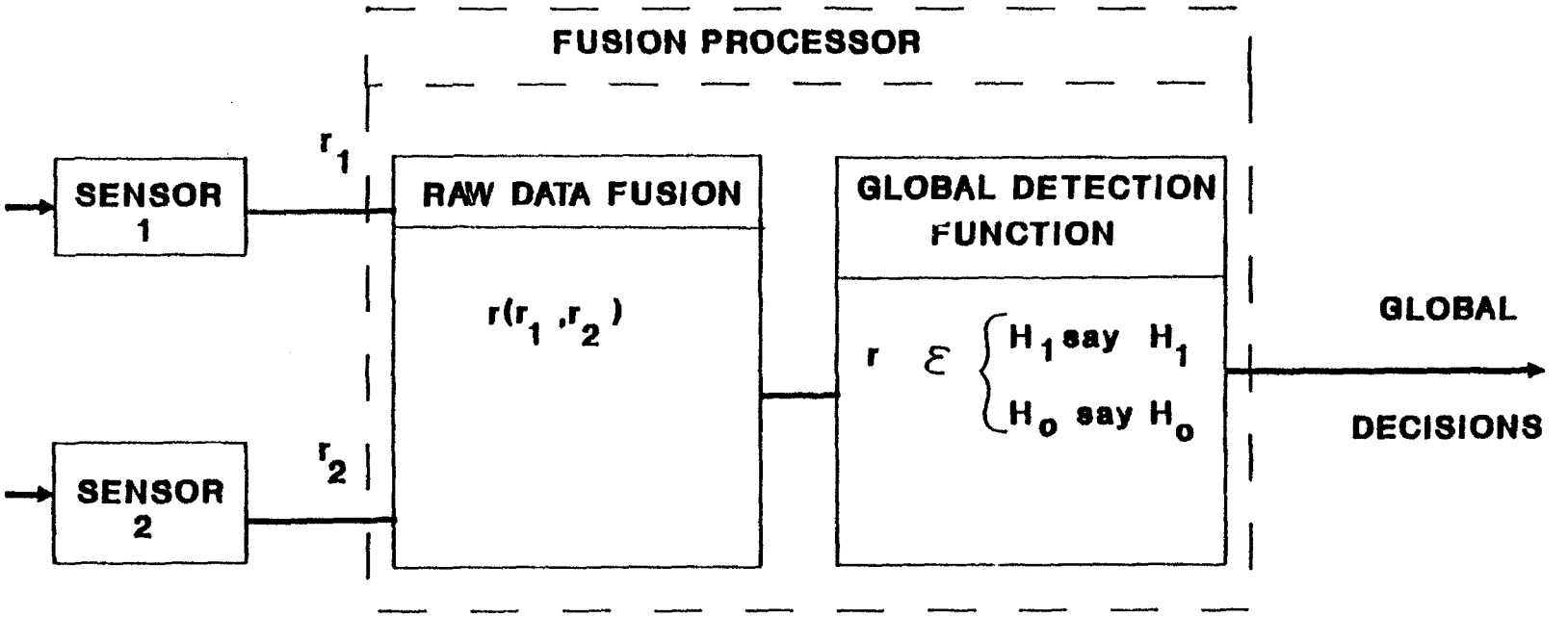


FIGURE 3. PRE-DETECTION FUSION IMPLEMENTATION

REFERENCES

1. J. A. Roecker and C. D. McGillem, Comparison of Two Sensor Tracking Methods Based on State Vector Fusion and Measurement Fusion, IEEE Trans., AES-24, pp. 447-449, (1988).

2. "Optimal Decision Fusion in Multiple Sensor Systems" by Stellios C. A. Thomopoulos et. al., IEEE Trans. on Aerospace and Electronic Systems. Vol. AES-23, No. 5, pp. 644-653, Sept. 1987.

3. Samuel S. Blackman, Multiple-Target Tracking with Radar Applications, Artech House, Inc., Norwood, MA, 1986.

4. IR/RF Raw Data Fusion By: J. Brandstadt, R. Gouse, E. McDermott, and C. Wilder. Report WL-TR-92-7018, April 1992.

5. Van Trees, H. L., Detection, Estimation, and Modulation Theory, Part I, John Wiley and Sons, Inc., New York, 1968.

6. "Multisensor Data Fusion", by Edward Waltz and James Llinas, Artech House Inc., Norwood, MA (1990).

7. Sensor Fusion for Airborne Platform, by D. Rai, GACIAC PR-89-01, pp. 71-81, (1989).

8. Charlesworth R. Martin, "Practical Considerations for a First Cut Multi-Sensor Fusion Analysis", Final Report, 1991 USAF Summer Faculty Research Program.

9. Jim D. Echard, "Estimation of Radar Detection and False Alarm Probabilities", IEEE Trans. on Aerospace and Electronic Systems, Vol. 27, No. 2, March 1991.

EFFICIENT ANALYSIS OF PASSIVE MICROSTRIP ELEMENTS FOR MMIC'S

Krishna Naishadham
Assistant Professor

and

Todd W. Nuteson
Graduate Associate

Department of Electrical Engineering
Wright State University
Dayton, OH 45435

Final Report for:
AFOSR Summer Research Program
Wright Laboratory, WPAFB

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

September 1992

EFFICIENT ANALYSIS OF PASSIVE MICROSTRIP ELEMENTS FOR MMIC'S

Krishna Naishadham, Assistant Professor
Todd W. Nuteson, Graduate Associate
Department of Electrical Engineering
Wright State University

ABSTRACT

Passive microstrip elements such as meander lines are analyzed by the full–wave, space domain moment method at microwave and millimeter wavelengths. Redundant calculations in the moment matrix are eliminated by utilizing various symmetries. At lower microwave frequencies quasi–static approximations of the Green's functions are invoked to simplify the analysis. The Green's functions are in general Sommerfeld–type integrals, which are computationally intensive. In this paper, closed–form analytical approximations of the integrals, recently developed by Chow et al., Aksun and Mittra, are utilized to increase the efficiency of the algorithm such that a circuit of moderate electrical size can be analyzed in reasonable time. Sample computed results are presented for the scattering (S) parameters of meander lines and multi–turn spiral inductors. Computed results compare reasonably well with measurements on a vector network analyzer.

EFFICIENT ANALYSIS OF PASSIVE MICROSTRIP ELEMENTS FOR MMIC'S

Krishna Naishadham and Todd W. Nuteson

I. INTRODUCTION

Microstrip elements occur in Microwave and Millimeter–wave Integrated Circuits (MMIC's) to facilitate various circuit objectives. The step–in–width impedance transition, mitered L–shaped corner, open–end and gap discontinuities are some examples [1]. At lower microwave frequencies the electromagnetic behavior of these circuit elements can be analyzed by quasi–static techniques, which neglect radiation and surface–wave coupling, and which do not accurately account for metallization and dielectric losses [2] – [4]. With increasing operating frequencies, these effects can become significant, and therefore, improved models are necessary to predict them. Computational methods based on full–wave analysis may be used for an accurate solution of the current distribution or the fields associated with the circuit, including characterization of all of the above–mentioned parasitic effects. These methods include the method of moments (MoM) [5] – [7], the finite element method [8], finite difference – time domain (FD–TD) methods [9], and the transmission line matrix (TLM) method [10]. For dense circuits and large microwave packages, full–wave methods, in spite of their versatility to handle various geometries and material parameters, have the disadvantage of formidable memory size and CPU time. Therefore, it is desirable to investigate means of improving the efficiency of full–wave methods so that a circuit of moderate electrical size can be analyzed in reasonable time. This paper describes an efficient MoM technique, based on reducing the computation time by utilizing various symmetries in the problem formulation, for the analysis of printed transmission line (or microstrip) elements of arbitrary shape (see Fig. 1). Efficiency is further enhanced by utilizing simple quasi–dynamic (or near–field) and far–field approximations of the Green's functions where applicable [7, pp. 164–166]. For moderate distances between source and observation points, the Green's functions manifest as Sommerfeld–type integrals, which are computationally intensive. We avoid computation of these integrals by invoking the recently developed closed–form microstrip Green's functions [11], [12].

In the method of moments, the boundary value problem for the current distribution on the surface of the conductor is formulated as a Mixed (scalar and vector) Potential Integral Equation (MPIE). The microstrip surface is segmented into rectangular cells, and special basis functions and a testing procedure tailored to planar geometries are used to compute the current distribution in each cell [6], [13]. A coaxial probe with its center conductor embedded in the dielectric substrate is used as excitation (source) for the

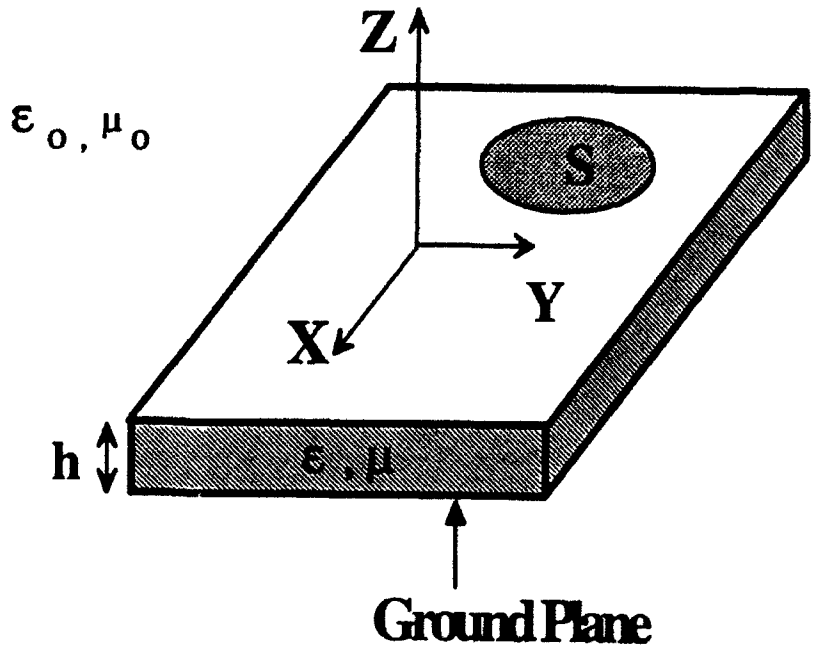


Fig. 1. An arbitrarily shaped circuit configuration (with surface S) printed on the laterally unbounded air–dielectric interface of a microstrip geometry. ϵ_0, μ_0 denote the constitutive parameters of free space, while ϵ, μ are those of the dilectric substrate.

currents [13]. From the current distribution, the multiport impedance matrix and the scattering (S) parameters of the discontinuity are computed [5], [14]. The current distribution and the S–parameters are computed for meander lines and spiral inductors, etched on gallium arsenide (GaAs) substrate. The computed S–parameters of the spiral inductor are compared with measurements on an automated vector network analyzer over a wide frequency band.

II. METHODOLOGY

A. Geometry

The geometry of the trace configuration is shown in Fig. 1. The transverse dimensions of the substrate and the ground plane are assumed to be infinite. The trace conductor is assumed to have a thickness much smaller than the substrate thickness, and much larger than the skin depth at the operating frequency. The trace conductor, with a planar surface S, resides on the air–dielectric interface at z=0. The dielectric substrate, in general lossy with constitutive parameters ϵ and μ, has a thickness h. The medium $z > 0$ is assumed to be free space, characterized by the parameters (μ_0, ϵ_0). A time–dependence of exp(jωt), where ω is the radian frequency, is assumed and suppressed.

B. Mixed Potential Integral Equation

MPIE is determined from the boundary condition that the total tangential electric field on the trace surface S is zero [5] – [7]:

$$\mathbf{a}_z \times \left[\overset{s}{\mathbf{E}}(\rho) + \overset{i}{\mathbf{E}}(\rho)\right] = \mathbf{a}_z \times Z_S \mathbf{J}_S(\rho), \quad \rho \in S. \tag{1}$$

The position vector $\rho = a_x x + a_y y$ locates the observation point on the trace, where $\mathbf{a}_p$ denotes the unit vector along the p–direction. The conductor losses are computed by modeling the microstrip conductor as a resistive sheet, with a frequency – dependent surface impedance Z_S (ohms/square) [14]. $\mathbf{E}^i$ is the excitation field and $\mathbf{E}^s$ is the scattered field, given in terms of the magnetic vector potential $\mathbf{A}$ and the Lorentz gauge related electric scalar potential V as

$$\mathbf{E}^s = -j\omega\mathbf{A} - \nabla V \tag{2}$$

where

$$\mathbf{A}(\rho) = \int_S \overset{=}{\mathbf{G}}_A(\rho;\rho') \cdot \mathbf{J}_S(\rho')\, ds' \tag{3}$$

$$V(\rho) = \int_S G_V(\rho; \rho') \, q_s(\rho') \, ds' \tag{4}$$

The surface current density J_S and the surface charge density q_S are related by the continuity equation. Only the surface currents residing on the air–dielectric interface are considered. $\overline{\overline{G}}_A$ is the dyadic Green's function for the vector potential, determined by solution of the boundary value problem for a horizontal electric dipole located on the air–dielectric interface as [7, pp. 145–160]

$$\overline{\overline{G}}_A = a_x a_x G_A + a_y a_y G_A \tag{5}$$

where, for source and observation points on the air–dielectric interface separated by the distance $R = |\rho - \rho'|$, we have [7, eq. (79)]

$$G_A = \frac{\mu_0}{2\pi} \, PV \int_0^\infty J_0(\alpha R) \frac{\alpha}{D_{TE}} \, d\alpha. \tag{6}$$

$J_0(x)$ denotes the zeroth order Bessel function of the first kind with argument x, and PV denotes the principal value.

The scalar Green's function G_V is the scalar potential due to a unit point charge located on the microstrip surface, and is given by [7, eq. (80)]

$$G_V = \frac{1}{2\pi\epsilon_0} \, PV \int_0^\infty J_0(\alpha R) \frac{N\alpha}{D_{TE} D_{TM}} \, d\alpha \tag{7}$$

where

$$N = u_0 + u \tanh uh \tag{8}$$

$$D_{TE} = u_0 + u \coth uh \tag{9}$$

$$D_{TM} = \epsilon_r u_0 + u \tanh uh \tag{10}$$

$$u_0 = \sqrt{\alpha^2 - k_0^2}, \tag{11}$$

$$u = \sqrt{\alpha^2 - k^2}, \tag{12}$$

$k_0 = \omega\sqrt{\mu_0\epsilon_0}$ and $k = \omega\sqrt{\mu\epsilon}$ refer to wavenumbers in air and dielectric, respectively. It is important to note the translational invariance ($|\rho-\rho'|$ dependence) of the Green's functions. This factor can be utilized to advantage in reducing the fill–time of the moment matrix for arbitrarily shaped traces. The improper integrals in (6) and (7) for the Green's functions should be interpreted in the sense of Cauchy principal value because of pole singularities of the integrand on the real axis. These singularities occur at the zeros of D_{TE} and D_{TM}, and their residues contribute, respectively, to the TE and TM surface wave–excited radiation. The integrals in (6) and (7), known as Sommerfeld integrals, converge slowly due to the oscillatory nature of the Bessel function J_0. Numerical integral schemes for such integrals are generally computer–intensive. One scheme involves numerical integration along the real axis between the zeros of the integrand using the method of averages [7, pp. 170–176], with the singularity in the path of integration being handled by the singularity extraction technique. In spite of the efficiency of this method, it is advantageous to utilize the closed–form approximations of the Sommerfeld integrals, recently developed by Chow et al. [11], Aksun and Mittra [12].

D_{TE} and D_{TM} have real zeros only in the interval $k_0 < \alpha < k$. For $f < c / 4h(\epsilon_r-1)^{1/2}$, where c is the velocity of light in free space, only D_{TM} has a zero in this interval. As an example, for a 64 mil thick fiberglass ($\epsilon_r = 4.7$) printed circuit board, this upper frequency limit is 24 GHz. For frequencies less than 24 GHz, therefore, only the dominant TM surface wave needs to be considered.

C. Approximation of the Green's Functions

To analyze the trace geometry in Fig. 1, it would be advantageous to use quasi–dynamic approximations of the Green's functions when applicable. These approximations are essentially the near–field ($k_0R \to 0$) approximations of the vector potential [7, eq. (103)] and the scalar potential [7, eq. (102)] Green's functions. At low frequencies, or for small R, the microstrip vector potential Green's function reduces to that of free space while the scalar potential Green's function contains, in addition to the free space term, a simple correction factor that depends on the permittivity of the substrate. When the distance R between source and observation points approaches about quarter wavelength in the dielectric, the Green's functions may be approximated by their far–field asymptotic approximations [7, eqs. (104) and (106)]. The advantage of utilizing these simple approximations for the Green's functions is that the need for numerical integration of the oscillatory, divergent Sommerfeld integrals in (6) and (7) is obviated. However, these approximations are limited to a rather restricted range of parameters including the distance between source and observation points, substrate dielectric constant, thickness,

losses and frequency. In the next section, alternative closed–form approximations of the Sommerfeld integrals, valid for any substrate parameters and the distance R, are presented.

D. Closed–Form Green's Functions

The closed–form Green's functions for a thick microstrip substrate have been derived in [11], [12]. The spatial domain Green's functions for the scalar and vector potentials may be written as G = A + B + C, where A represents the contributions from a few quasi–dynamic (real) images which are dominant in the near–field region, B represents the surface wave contributions (in the form of complex residues) which are dominant in the far–field region, and C represents the contributions from complex images which are important in the intermediate–field region. Explicit expressions for A, B and C may be found in [11] for both scalar and vector potentials. The contributions from the complex images can be approximated in terms of complex exponentials using Prony's method [11]. With these closed–form expressions of the Green's functions, the numerical integration of the Sommerfeld integrals can be avoided completely, and significant savings in computer time can be accomplished.

In Fig. 2, we show a comparison of the Green's functions computed by using (a) the closed–form expressions, (b) numerical integration based on the method of averages, and (c) the quasi–static (near–field) approximations. The substrate has the parameters ϵ_r = 12.85, h = 100 μm, and the frequency is 10 GHz. Good agreement is observed between all the three computations. The quasi–static approximations of the scalar potential Green's functions are reasonably accurate until the onset of the dominant TM surface wave, indicated by a dip in the magnitude curve or by 90° phase shift. The vector potential is not affected by this surface wave (for the substrate listed, only one TM surface wave pole exists), hence the quasi–static approximations of G_a are accurate over a much higher range of distances than those applicable for G_Ψ.

III. NUMERICAL CONSIDERATIONS

A. Method of Moments

The MPIE in (1) is solved by the MoM using rooftop subsectional basis functions [6] to represent the surface current density. The current density $\mathbf{J_s} = \mathbf{a_x}J_x + \mathbf{a_y}J_y$ in (3) is expanded in the orthogonal x and y directions as

$$J_x(\rho) \approx \sum_{p=1}^{N_x} J_p^x \Lambda_p^x \tag{13}$$

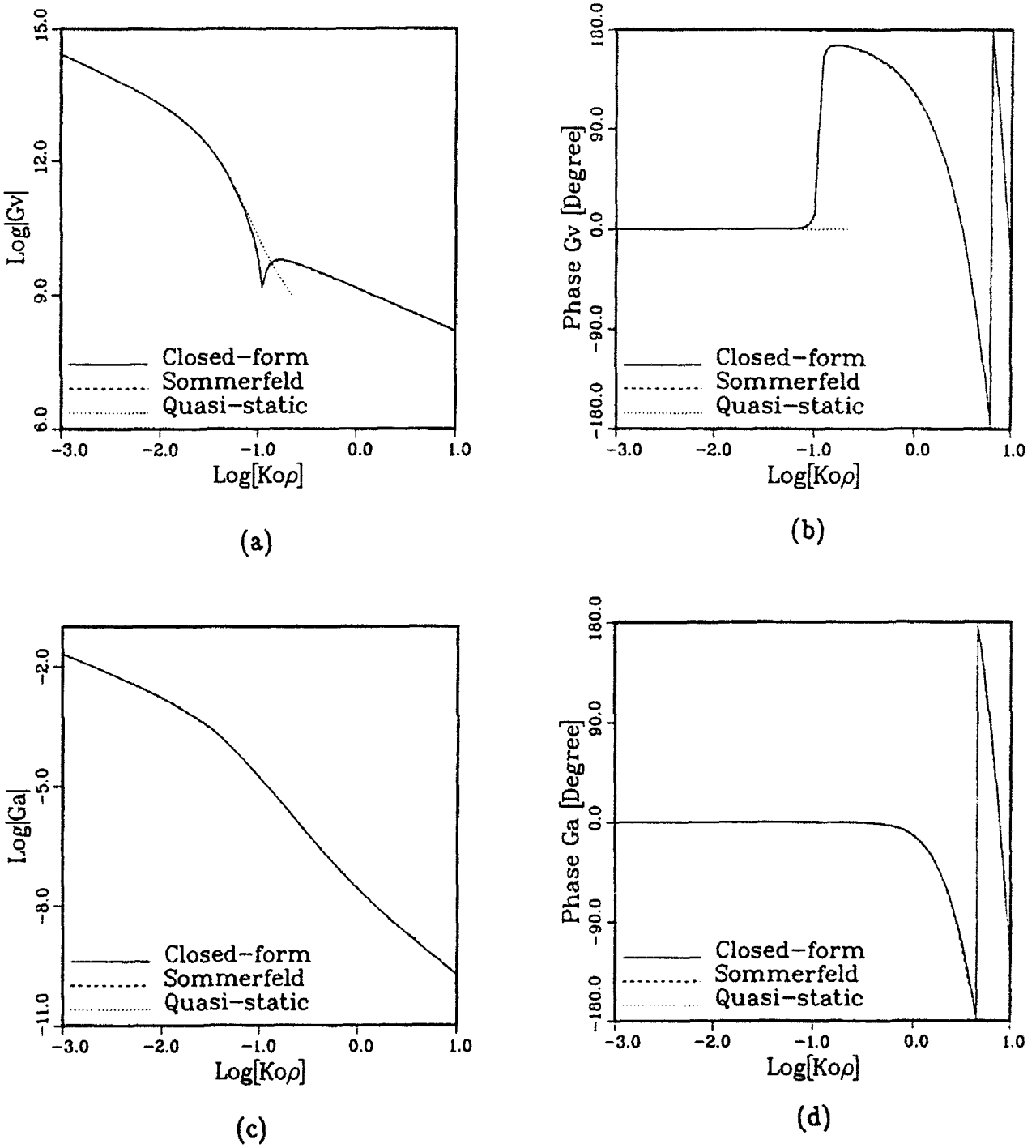


Fig. 2. Comparison of the quasi–static (or quasi–dynamic) and closed–form approximations of the Green's functions with numerical integration of the fully dynamic Sommerfeld integrals. ϵ_r = 12.85, h = 100 μm, f = 10 GHz. (a) Scalar potential magnitude, (b) scalar potential phase, (c) vector potential magnitude, and (d) vector potential phase.

$$J_y(\rho) \approx \sum_{q=1}^{N_y} J_q^y \Lambda_q^y \tag{14}$$

where $\rho = (x,y)$, J_p^x and J_q^y are the unknown expansion coefficients to be determined. The basis functions, each of which overlaps two contiguous rectangular cells, are represented by triangles in the direction of the current and by pulses in the orthogonal direction [6].

In order to solve for the unknowns J_p^x and J_q^y, a system of linear equations is obtained by testing (1) along the x and y directions by one–dimensional pulse functions (line testing). The tested equations can be written out in matrix form as

$$\begin{bmatrix} [Z_{mp}^{xx}] & [Z_{mq}^{xy}] \\ [Z_{np}^{yx}] & [Z_{nq}^{yy}] \end{bmatrix} \begin{bmatrix} [J_p^x] \\ [J_q^y] \end{bmatrix} = \begin{bmatrix} [V_m^x] \\ [V_n^y] \end{bmatrix},$$

$$\begin{aligned} m,p &= 1,2,\ldots,N_x \\ n,q &= 1,2,\ldots,N_y \end{aligned} \tag{15}$$

where the submatrix $[Z_{st}^{uv}]$ denotes the contribution of u–directed testing of the field produced by v–directed current basis elements. The subscripts t and s refer to the individual basis element and the test path, respectively. The matrix elements consist of scalar and/or vector potential integrals defined over the support of a pulse function and a rooftop function, respectively [13]. The column vectors $[V_m^x]$ and $[V_n^y]$, of length N_x and N_y, respectively, correspond to x and y directed testing of the excitation field. Similarly, $[J_p^x]$ and $[J_q^y]$ refer to current expansion coefficients associated with each rooftop function. Explicit expressions for the elements of the moment matrix [Z] and of the excitation vector [V] may be found elsewhere [13].

B. Excitation and S – Parameters

A coaxial probe excitation model, as developed in [13], is used in the MoM algorithm. A current I = 1 A is assumed to flow uniformly through the center conductor. Upon attachment to the microstrip patch, this current spreads radially across a charge cell. A special attachment mode for the current density is defined such that the net charge density on the excitation cell is constant [13]. Since the attachment mode is of the same form as the current basis functions, the elements of the excitation vector in (15) can be determined by a simple modification of the matrix elements in (15).

To find the S–parameters, the surface currents on the connection lines and the discontinuity are computed by MoM. Thus, one finds currents entering the ports, whose reference plane locations are defined to be at the intersection between the transmission lines and the discontinuity. The potential of each port with respect to ground plane (assumed to be at zero potential) is computed by integrating the vertical component of the electric field in the substrate produced by the currents flowing on the microstrip surface:

$$V_i = \sum_{p=1}^{N_x+N_y} \int_{-h}^{0} E_z^p(\rho_i,\rho_p';z)\, dz, \quad i = 1,2,... \tag{16}$$

In (16), the position vectors ρ_i and ρ_p' refer to the (x,y) coordinates of the i^{th} reference port location and the p^{th} current basis function, respectively. The electric field E_z^p is determined by differentiating the vector and scalar potentials. Once the port voltages and currents are known, the open–circuit (or Z) parameters and the S–parameters can be computed using standard circuit theory relations [1]. If the S–parameters are desired at a different reference plane location, a deembedding procedure such as that discussed in [14] may be used.

C. *Efficiency Considerations*

The redundancies present in the moment matrix for equal size cells can be taken advantage of in reducing its fill–time. A careful examination of the matrix elements in (15) reveals (upto) a sixteen–fold redundancy for the scalar potential contributions! Therefore, a significant reduction in the matrix fill–in time can be achieved by computing each scalar potential integral and placing it in the appropriate matrix element as it is calculated.

The Green's functions depend on the distance $R = |\rho - \rho'|$ between source and observation points. Depending on the trace geometry being examined, there may be only a few distinct source–observer distances. For example, on a rectangular patch divided into M×N charge cells, there are only MN such distances, or equivalently MN distinct potential integrals, whereas the number of matrix elements in which these integrals occur is $2MN - M - N$. Therefore, we could precompute and store the MN distinct partial potentials in a two–dimensional array, and retrieve them later to fill the entire moment matrix. Considerable time savings can be accomplished in this manner. The computational efficiency is further enhanced by computing and storing the Green's functions for a few source–observer distances, and using parabolic interpolation on the stored values, for other distances.

When the observation point is within the source cell, the Green's functions become singular as 1/R [13]. In this case, the singularity extraction technique is used to smoothen the integrand in the Green's functions for numerical integration purposes. The extracted singular part is integrated analytically in closed form. Details may be found in [13].

In Fig. 3, we compare the CPU time required for an efficient fill–in of the moment matrix using the method discussed above, with the time involved in filling the complete matrix without the utilization of any symmetry. The structure investigated is a rectangular patch. For a large matrix, the CPU time for the complete matrix fill is at least one order of magnitude higher than the time required for the efficient fill–in. As an example, to fill a 500 × 500 matrix, the complete fill–in requires 2000 (s) whereas the efficient fill–in takes about 100 (s). For comparison, the time involved in inverting the moment matrix by standard Gauss elimination is also plotted as a function of matrix order. The time required for complete matrix fill is higher than the inversion time by one to two orders of magnitude. By extrapolation to large matrix orders, this dominance of the complete matrix fill–time is expected to continue until the order becomes 3000 or so. However, the efficient fill–in requires times much less than the inversion times, especially for large matrix orders. In order to expedite the MoM solution for complex circuits, the problem of efficient inversion of dense matrices needs to be addressed concurrently with efficient fill–in schemes.

IV. RESULTS

In this section, we present sample results on the the magnitude and phase of the S–parameters as functions of frequency for a few representative lossy microstrip elements. The metallization is gold deposited on 100 μm thick GaAs substrate ($\epsilon_r = 12.85$). The conductivity of metallization has been adjusted by a factor of 0.75 σ_{gold} to account for surface roughness effects.

Fig. 4 displays the magnitude and phase of S_{21} for the initial test case of a lossy transmission line. The dimensions of the line are shown along with those of the deembedded transmission line sections in Fig. 4a. The results computed by our MoM program compare well with those predicted by EESof's commercial CAD software package Libra (the magnitude curves are practically coincident). The characteristic impedance of the transmission line can be computed as $Z_0 = \sqrt{Z_{11}^2 - Z_{12}^2}$ (see Fig. 4a for the reference plane locations), where Z_{11} and Z_{12} are the open–circuit parameters at port 1. This gives a value of 52 Ω − 54 Ω with a small imaginary part in the frequency range 1 GHz − 19 GHz, and compares

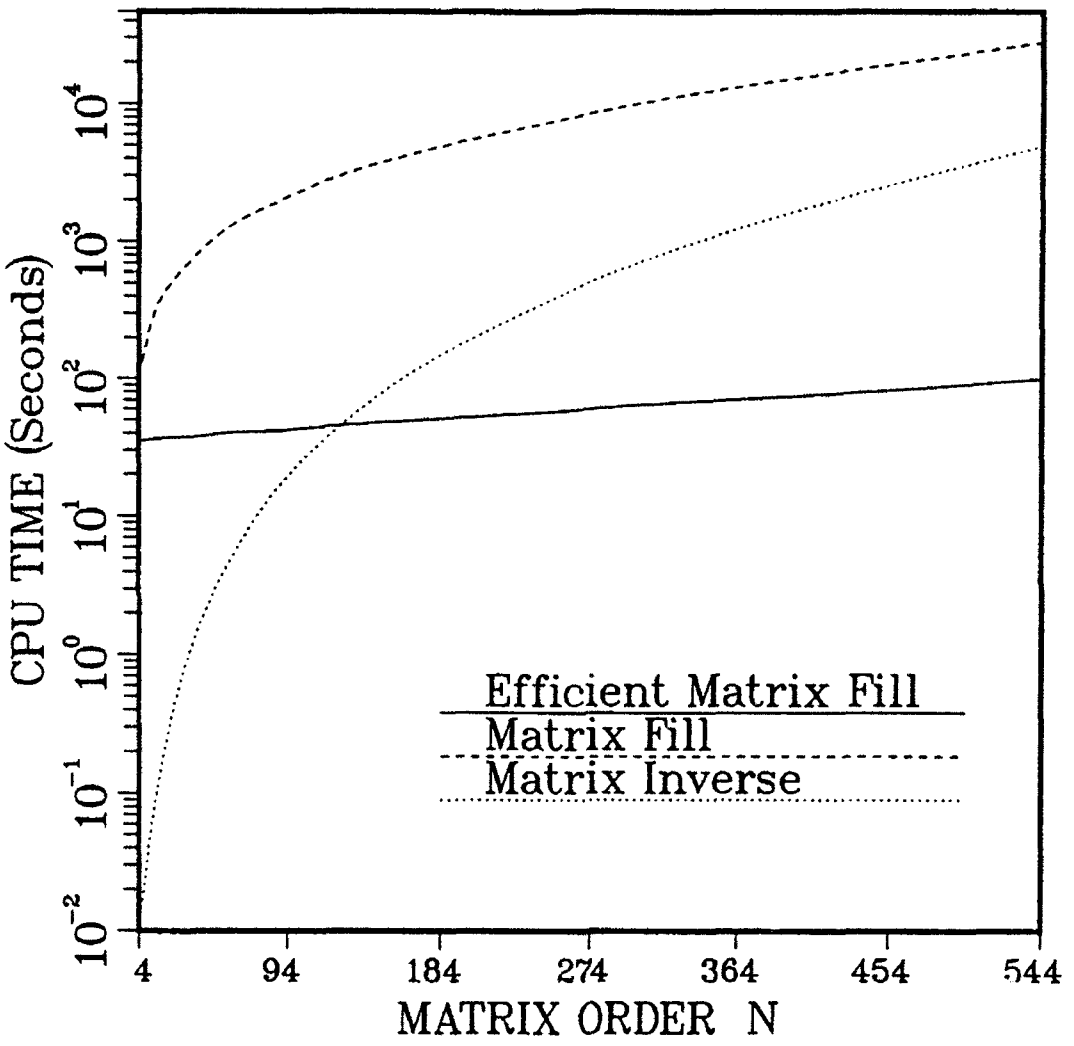
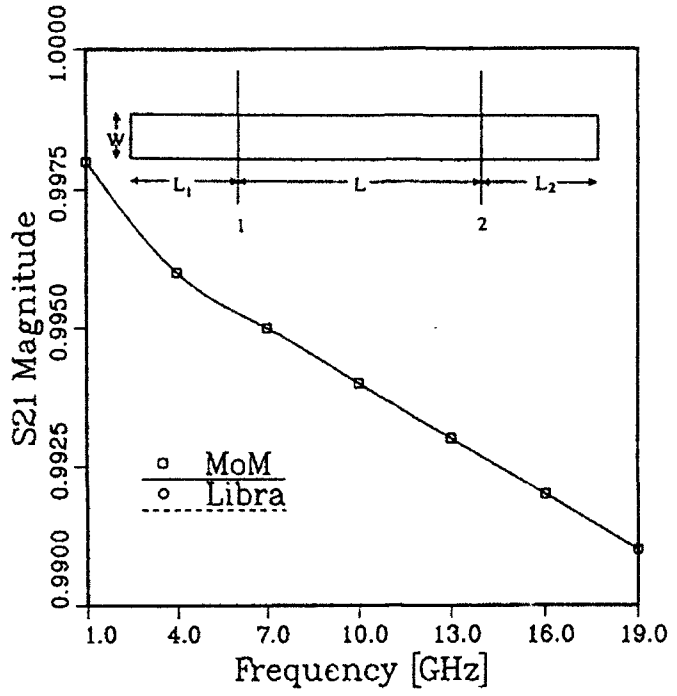
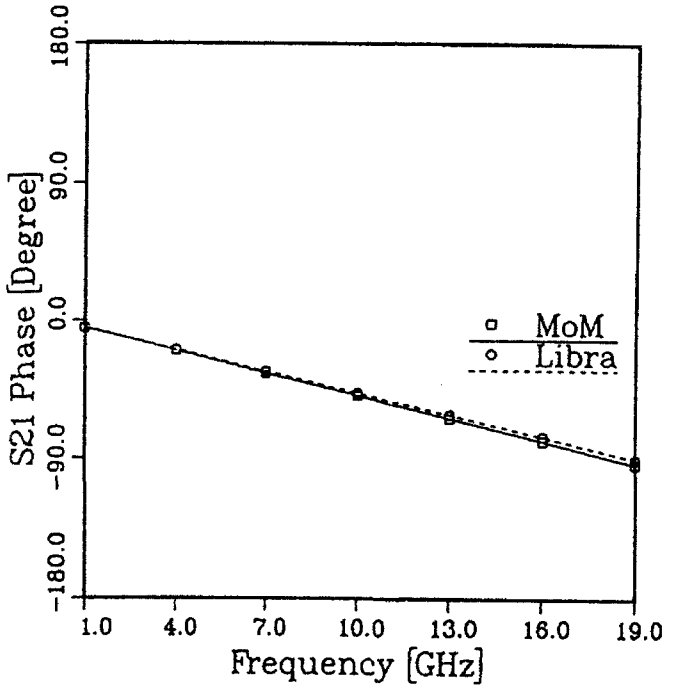


Fig. 3. Comparison of moment matrix fill and matrix inversion CPU times on VAX–8550 computer as a function of the matrix order. Efficient matrix fill refers to the method described in Sec. III.



(a)



(b)

Fig. 4. (a) Magnitude, (b) phase of S_{21} for a lossy transmission line section as functions of the frequency. W = 75 μm, L = 1500 μm, $L_1 = L_2 = 675$ μm.

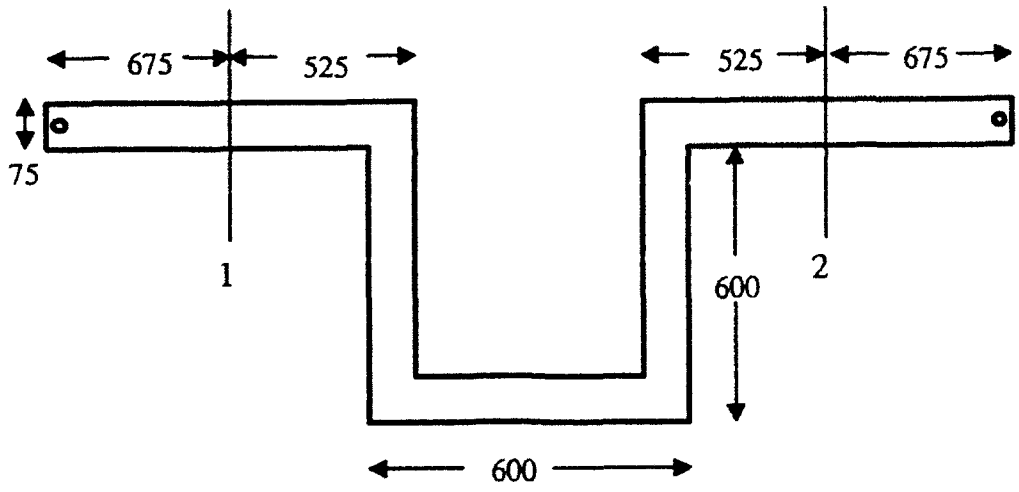
reasonably well with the nominal value of 50 Ω. The magnitude and phase of S_{11} also compare well with Libra's output, and are not shown for brevity.

Next, we examine the elements whose geometry is depicted in Fig. 5. The deembedded transmission line sections are shown for the meander line. The magnitude and phase of S_{21} and S_{11} for the meander line, computed by the method discussed in this paper, are compared with results from Libra in Fig. 6. The agreement between the two in phase is excellent, but the magnitudes at the higher frequencies do not compare as well. This may be either due to the quasi–static nature of the analysis used in Libra, which may not model the dispersive effects at the higher frequencies accurately, or due to the fact that we have used an analytical result, based on quasi–TEM theory, for the current distribution on the short–circuited transmission line in the deembedding procedure [14]. Nevertheless, the agreement is reasonable over a wide frequency band.

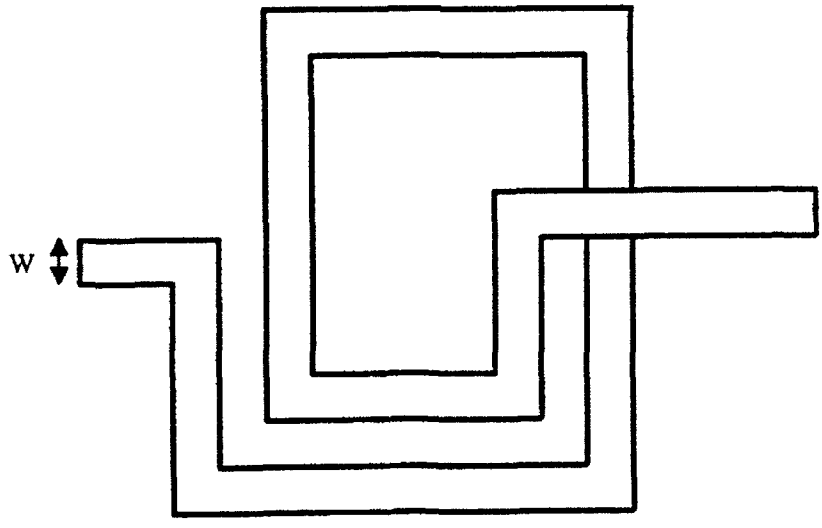
We now consider the 1.5 turn spiral inductor depicted in Fig. 5. The S–parameters of the inductor are computed by the method discussed in Sec. III. The magnitude and phase of computed S_{11} and S_{21} are compared with measurements on an automated vector network analyzer in Fig. 7. A reasonably good agreement is observed between the full–wave result and the measurements, considering the complexity of the structure. Parametric studies have also been made by varying the surface impedance Z_S in order to examine the influence of conductor losses. It is observed that conductor losses do not have significant impact on the overall characterization of the elements.

V. SUMMARY AND CONCLUSIONS

An efficient full–wave moment method has been developed for the analysis of microstrip circuit elements of arbitrary shape. The surface of the microstrip conductors has been divided into rectangular cells and rooftop basis functions have been used with a line testing MoM procedure to determine the unknown surface currents. The algorithm can accommodate both conductor and dielectric losses. The utility of the MoM algorithm in a CAD environment has been illustrated by computation of the S–parameters for a meander line and a 1.5 turn spiral inductor. Several techniques are used to increase the efficiency of the computer program so that a circuit of moderate electrical size can be analyzed in reasonable time. These include the utilization of (a) quasi–dynamic and far–field approximations of the Green's functions where applicable, (b) the closed–form microstrip Green's functions, and (c) the various symmetries in the problem formulation which facilitate an efficient fill–in of the moment matrix for electrically large circuits. As an example, the meander line involves about 200 unknown current



(a)



(b)

Fig. 5. Geometry of (a) symmetrical meander line along with the deembedded transmission line sections (all dimensions are in μm), and (b) 1.5–turn spiral inductor (W = 25 μm).

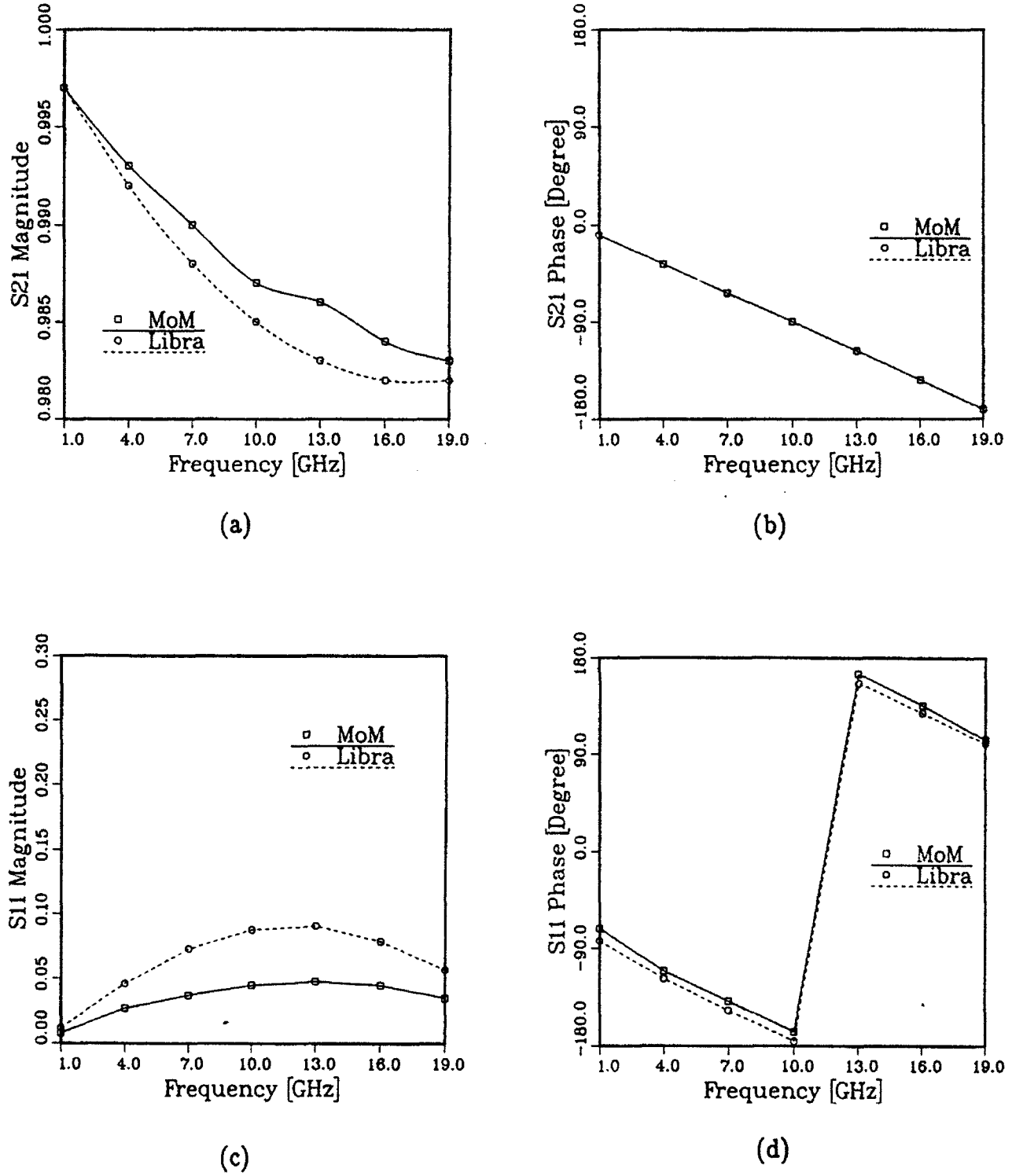


Fig. 6. (a) Magnitude of S_{21}, (b) phase of S_{21}, (c) magnitude of S_{11}, and (d) phase of S_{11}, as functions of frequency, for the meander line. Results computed by using Libra are also shown for comparison.

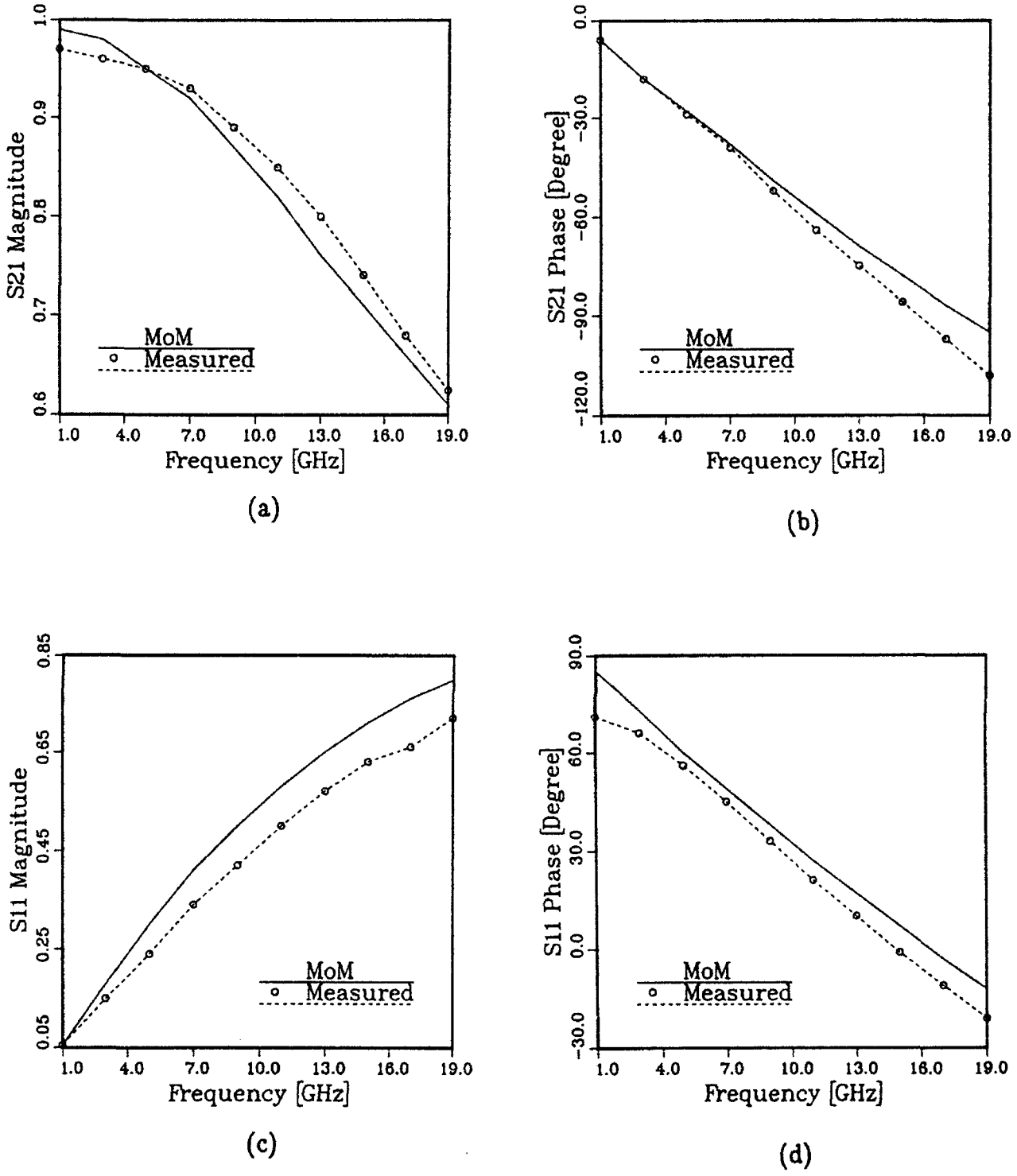


Fig. 7. Comparison of measured S–parameters with computed values for the spiral inductor. (a) Magnitude of S_{21}, (b) phase of S_{21}, (c) magnitude of S_{11}, and (d) phase of S_{11}.

coefficients, and takes about 1.5 CPU minutes per frequency on a VAX 8550 mainframe. The computed S–parameters compare reasonably well with results computed by using commercial software, and with measurements on a network analyzer.

ACKNOWLEDGEMENTS

The authors would like to thank Mr. Greg Creech, Wright Laboratory, for supplying the measured results on the spiral inductor. The authors are grateful to Dr. J. R. Mosig, Ecole Polytechnique Federale de Lausanne, and Dr. R. C. Hall, Ball Aerospace Corporation, for many helpful discussions.

REFERENCES

[1] K. C. Gupta, R. Garg, and R. Chadha, *Computer–Aided Design of Microwave Circuits*, Artech House, Norwood, MA, 1981, Chapter 6.

[2] P. Silvester and P. Benedek, "Microstrip discontinuity capacitances of right–angle bends, T–junctions, and crossings," *IEEE Trans. Microwave Theory Tech.*, vol. MTT–21, pp. 341–346, 1971.

[3] A. F. Thomson and A. Gopinath, "Calculation of microstrip discontinuity inductances," *IEEE Trans. Microwave Theory Tech.*, vol. MTT–23, pp. 648–655, 1975.

[4] E. J. Denlinger, "Losses of microstrip lines," *IEEE Trans. Microwave Theory Tech.*, vol. MTT–28, pp. 513–522, 1980.

[5] R. F. Harrington, *Field Computation by Moment Methods*, Macmillan, New York, 1968.

[6] A. W. Glisson and D. R. Wilton, "Simple and efficient numerical methods for problems of electromagnetic radiation and scattering from surfaces," *IEEE Trans. Antennas Propag.*, vol. AP–28, no. 5, pp. 593–603, 1980.

[7] J. R. Mosig, "Integral Equation Technique," in *Numerical Techniques for Microwave and Millimeter–Wave Passive Structures*, (T. Itoh, ed.), John Wiley, New York, pp. 133–213, 1989.

[8] J. B. Davies, "The Finite Element Method," in *Numerical Techniques for Microwave and Millimeter-Wave Passive Structures*, (T. Itoh, ed.), John Wiley, New York, pp. 33–132, 1989.

[9] X. Zhang, J. Fang, K. K. Mei, and Y. Liu, "Calculation of the dispersive characteristics of microstrips by the time–domain finite–difference method," *IEEE Trans. Microwave Theory Tech.*, vol. MTT–36, no. 2, pp. 263–267, 1988.

[10] W. J. R. Hoefer, "The Transmission Line Matrix (TLM) Method," in *Numerical Techniques for Microwave and Millimeter-Wave Passive Structures*, (T. Itoh, ed.), John Wiley, New York, pp. 496–591, 1989.

[11] Y. L. Chow, J. J. Yang, D. G. Fang, and G. E. Howard, "A closed–form spatial Green's function for the thick microstrip substrate," *IEEE Trans. Microwave Theory Tech.*, vol. MTT–39, no. 3, pp. 588–592, 1991.

[12] M. I. Aksun and R. Mittra, "Investigation of radiation characteristics of microstrip etches," Technical Report No. UILU–ENG–92–2210, University of Illinois at Urbana–Champaign, March 1992.

[13] J. R. Mosig and F. E. Gardiol, "General integral equation formulation for microstrip antennas and scatterers," *Proc. Inst. Elec. Eng.*, pt. H, vol. 132, no. 7, pp. 424–432, 1985.

[14] A. Skrivervik and J. R. Mosig, "Impedance matrix of multiport microstrip discontinuities including radiation effects," *AEU (Electronics and Communication)*, vol. 44, no. 6, pp. 453–461, 1990.

THIRD ORDER NONLINEAR OPTICAL PROPERTIES OF STRAINED LAYER SEMICONDUCTORS WITH APPLICATION TO OPTICAL WAVEGUIDES

M. J. Potasek
Research Professor
Department of Applied Physics

Columbia University
New York, NY 10027

Final Report for:
Summer Research Program
Wright-Patterson

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

September 1992

THIRD ORDER NONLINEAR OPTICAL PROPERTIES OF STRAINED LAYER SEMICONDUCTORS WITH APPLICATION TO OPTICAL WAVEGUIDES

M. J. Potasek
Research Professor
Department of Applied Physics
Columbia University

Abstract

Nonlinear optical phenomena are of interest because of their wide variety of applications in many fields. Therefore I have investigated the waveguiding properties of strained layer III-V semiconductors. In view of the semiconductor growth facilities here as well as measurement capabilities, particular attention was paid to InGaAs/GaAs multiple quantum well materials. Using the numerical beam propagation method, I calculated the coupled waveguide characteristics for these materials and examined the effects of various physical parameters.

THIRD ORDER NONLINEAR OPTICAL PROPERTIES OF STRAINED LAYER SEMICONDUCTORS WITH APPLICATION TO OPTICAL WAVEGUIDES

M. J. Potasek

Introduction

The use of semiconductors for optics is advantageous because of the prospects of integration of several components such as lasers, detectors and modulators. In most cases research has been performed on electro-optic devices. These devices often make use of the quantum confined Stark effect in multiple quantum wells in which one observes enhanced changes in absorption as a result of quantum confinement of the exciton. However because of electron-hole recombination dynamics, electro-optic switching is often slow. Therefore all optical processes are being pursued in which picosecond or femtosecond switching can be observed. Recent advances have utilized the third order nonlinear susceptibility. This susceptibility encompasses the intensity dependent change in the refractive index whose coefficient is denoted by n_2. In intensity dependent media, the phase of the wave changes as a function of distance giving rise to phenomena which are utilized for various applications. Recently several devices have used various configurations. Some devices have included two-channel nonlinear directional couplers (1-3), nonlinear Mach-Zehnder interferometer (4,5), polarization switches in birefringent fibers (6,7) two core fiber nonlinear directional couplers (8,9) and semiconductor multiple quantum well (MQW) waveguides (10,11).

Recently, experimental advances have been made in all-optical GaAs/GaAlAs MQW nonlinear directional couplers. These measurements used a mode-locked frequency doubled Nd:YAG laser to pump a femtosecond dye laser (10). Two components were observed in the kinetics. One component was the subpicosecond optical Stark effect and the slow component was due to the usual

long lived real carrier population. Other forms of optical phenomena in these configurations was observed (12-18).

However in order to understand the waveguide phenomena one has to investigate the wave equations from Maxwell's equations. The general Maxwell's equations are given by Eq (1)

$$\nabla\times\nabla\times E + \frac{1}{\varepsilon_o c^2}\frac{\partial^2 D}{\partial t^2} = 0 \tag{1}$$

where E is the electromagnetic field and D is the displacement vector. The displacement vector is expanded in terms of E in Eq. (2)

$$D = \varepsilon_o \chi * E$$

$$= \varepsilon_o \int_{-\infty}^{t} dt_1\, \chi^{(0)}(t - t_1) E(t_1) \tag{2}$$

$$+ \varepsilon_o \int dt_1 \int dt_2 \int dt_3\, \chi^{(3)}(t - t_1) E(t_1) E(t_2) E(t_3)$$

where electro-optic effects have been excluded. For the slowly varying envelope of the field, denoted by q, the propagation of light is governed by Eq. (3)

$$iq_z + \tfrac{1}{2}\beta_2 q_{tt} + \delta|q|^2 q = 0 \tag{3}$$

where $\delta = n_2\omega_0/c$

<u>THIRD ORDER NONLINEAR OPTICAL PROPERTIES IN STRAINED LAYER III-V SEMICONDUCTORS</u>

Recently great attention has been paid to the measurement of third-order nonlinear optical properties, such as the nonlinear index of refraction in

GaAs/GaAlAs. For example, recent measurements of third-order nonlinearity in GaAs/GaAlAs MQW waveguides yield a value of -10^{-12} to -10^{-13} cm^2/W as compared to a value of 10^{-16} cm^2/W for optical fibers. Large values of n_2 are advantageous because less light intensity is required to achieve a desired level of performance. Recent progress has indicated that improvements in nonlinear optical coefficients can be made in semiconductors by using strained layer heterostructures, particularly InGaAs/GaAs and InGaAs/AlGaAs. These materials offer band gap engineering flexibility often not attainable in lattice matched materials. Inportant parmeters such as the well/barrier combination, the level of strain and the the material concentration in layers can be varied to modify the physical parameters such as the band-gap, electro-optic and optical properties. Other important aspects include: 1) These materials can have band gaps in the technologically important regions of 0.85 um and 1.5 um. 2) These are ternary compounds without the additional complication of added phosphorous as in the common compound InGaAsP. 3) Recent theoretical computations indicate that these compounds may have large optical nonlinearities.

For thin layers the lattice mismatch induces internal strain rather than dislocations (17,18). The presence of built in strain affects the structural aspects of the materials and influences their electronic properties (19) through strain induced changes of the band structure. Such structures have been fabricated and shown to have novel electronic properties(17,20). Recently, enhanced optical properties have been theoretically predicted (21,22).

In the InGaAs/GaAs materials the InGaAs is the well material and the GaAs is the barrier material (23,24). The smaller bandgap InGaAs well material is under compression and the GaAs barriers are unstrained. For small amounts of Indium in the material, the barrier heights are relatively low (for example, for 12% In the electron barrier height is about 80 meV.) This can make superlattice effects more prominent (25).

However there is an important contrast between the strained layers grown

along the [100] and [111] crystallographic axes. For strained-layer superlattices with [111] growth axis, the orientation of the lattice constant mismatch induced strains results in polarization fields directed along the growth axis. Group III-V semiconductors are piezoelectric and strains can lead to electric polarization fields.

Since one of the materials is in biaxial compression and the other is in biaxial tension, the electric polarization vectors are of opposite sign. The internally generated electric fields modify significantly the electronic properties of the superlattice and since the internal fields are screened by photogenerated carriers, they can lead to large optical nonlinearities. Both electronic energy levels and wavefunctions arealtered by internal fields. These changes in the wavefunctions lead to changes in optical matrix elements.

Recently, optical measurements have resulted in strong evidence for the presence of strain-generated electric fields in [111] oriented InGaAs/GaAs strained-layer superlattices (26). Experiments showed that the lowest energy intrinsic transition shifts to the red by about 20 meV in a [111] sample as compared with a [100] sample. This spectral difference can only be fitted by including the strain-generated electric fields.

In InGaAs/AlGaAs, the fields shift the conduction band state to lower energy and the valence band to higher energy, thereby reducing the band gap. Calculations for InGaAs/AlGaAs with the Ga containing layers half as thick as the Al layers show that the strain-induced electric field can significantly change the optical properties. This is a type I superlattice in which the smaller band gap Ga alloy is in biaxial compression. The light hole bands are split away from the heavy hole bands by strain. Thus the band-edge optical properties will be dominated by heavy hole to conduction band transitions. Calculations show that very large third order nonlinear optical properties may be obtained in these systems (22) in relations to other systems (27).

NONRESONANT NONLINEARITY

The resonant nonlinearities have been investigated in bulk GaAs and MQW (28-30). The mechanisms include bandfiling, state filing and screening of the Coulombic interaction (31,32). Recent nonlinearities can be extremely large, however recovery times are slow because they are mediated by carrier recombination. Measurements of these nonlinearities can be difficult because several processes including virtual and real carriers and thermal effects can contribute. Nonresonant nonlinearities can have a rapid recovery time because they are not associated with population transfers.(33) Nonresonant studies with high pump intensities have investigated the optical Stark effect in QW (34-36).

A variety of techniques have been used to measure the intensity dependent nonlinear index changes, such as four-wave mixing (37), nonlinear waveguide couplers (38), nonlinear Fabry-Perot (39), fringe shift interferometry (40), Mach-Zehnder interferometry (41) and beam propagation measurements (42).

Two-Photon Absorption

A significant effect is posed by TPA (43) because it is essentially a nonlinear loss mechanism. Recent measurements have investigated the influence of TPA on waveguide performance . Its efects will be considered in more detail in the next sections.

ALL OPTICAL PHENOMENA IN SEMICONDUCTOR WAVEGUIDES

The optical switching in semiconductor waveguides is governed by the coupled equations

$$iq_{1z} + \tfrac{1}{2}\beta_2 q_{1tt} + \delta\,|q_1|^2 q_1 + kq_2 = -i\beta\,|q_1|^2 q_1 \qquad (4)$$

$$iq_{2z} + \tfrac{1}{2}\beta_2 q_{2tt} + \delta\,|q_2|^2 q_2 + kq_1 = -i\beta\,|q_2|^2 q_2$$

These equations were solved using the fast Fourier transform numerical beam propagation method (NBP). This numerical procedure consists of propagating the field for a distance d/2 with dispersion only, multiplying the result by a nonlinear term that represents the effect of nonlinearity over the whole segment length d, and then propagating the field for the remaining distance d/2 with dispersion only. In effect, the nonlinearity is assumed to be lumped at the midplane of each segment.

Using the NBP method, I investigated the behavior of the previous equations. The calculations were performed on a SUN workstation. Figure 1 shows a plot of the switching fraction as a function of peak power for light incident into one waveguide only for various values of TPA. The curves correspond to various ratios of TPA to n_2. The solid line corresponds to no TPA, the dashed line corresponds to a ratio of 0.1, the dot-dashed line corresponds to a ratio of 0.2 and the dotted line corresponds to a ratio of 0.5. This figure clearly shows the degradation of the output as TPA increases.

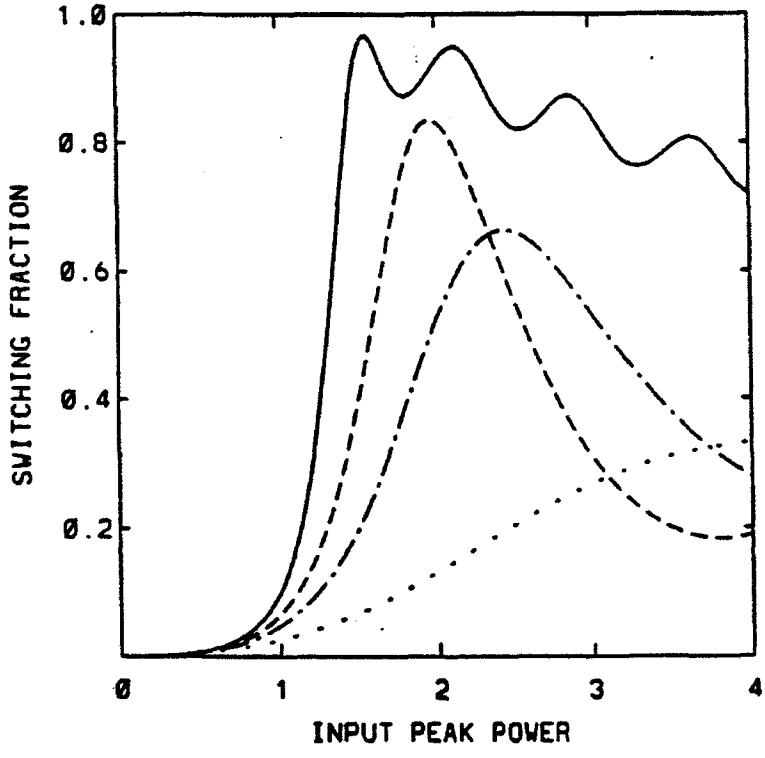


Fig. 1

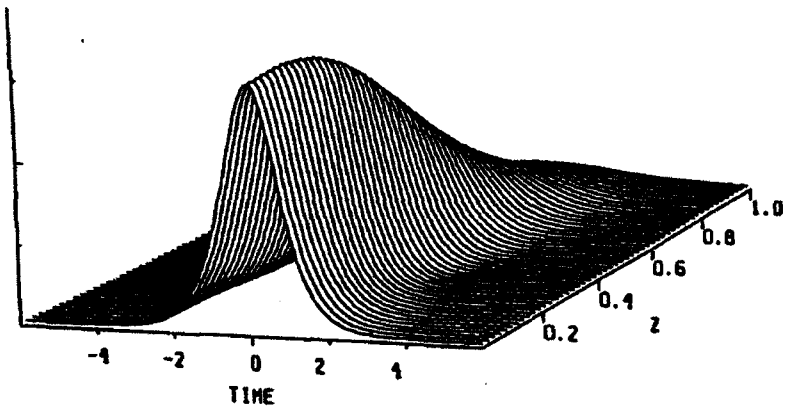
Low Incident Power

For low input power, the system operates linearly and in one coupling length the light switches to the other waveguide. These results are shown in Fig. 2. The three dimensional portions of the figure show the intensity of the optical beam for a given input pulse width as a function of propagation distance. Section (1) corresponds to light into one waveguide and section (2) corresponds to the other coupled waveguide with no incident light. The lower sections of the figure are the contour plots for the two waveguides. From Fig. 2, it can be seen that light into one waveguide is almost totally transferrred to the other waveguide. This is due to the linear coupling parameter k.

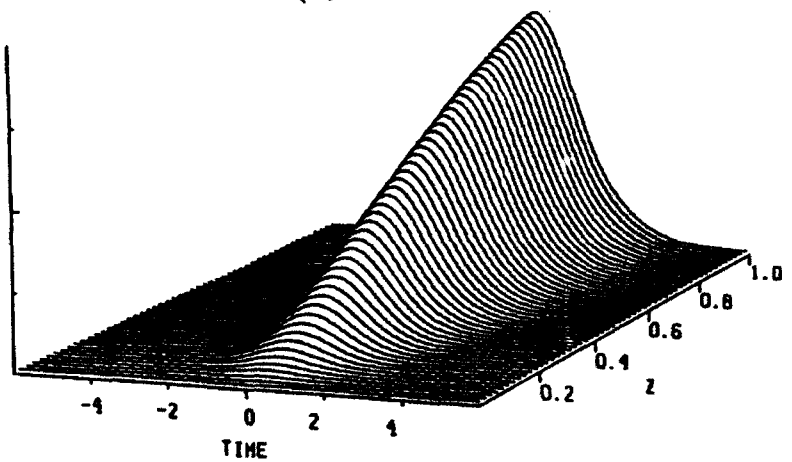
High Incident Power

However as the power increases the nonlinearity introduces a phase shift in such a manner as to counteract the linear coupling. These results are shown in Fig. 3. It is seen that light into one waveguide almost totally remain in that waveguide. However due to the various linear and nonlinear interactions the temporal shape of the pulse is distorted. Further calculations were performed to improve the performance of the system.

(1)


-4
-2
0
2
4
TIME
0.2
0.4
0.6
0.8
1.0
Z

(2)


-4
-2
0
2
4
TIME
0.2
0.4
0.6
0.8
1.0
Z

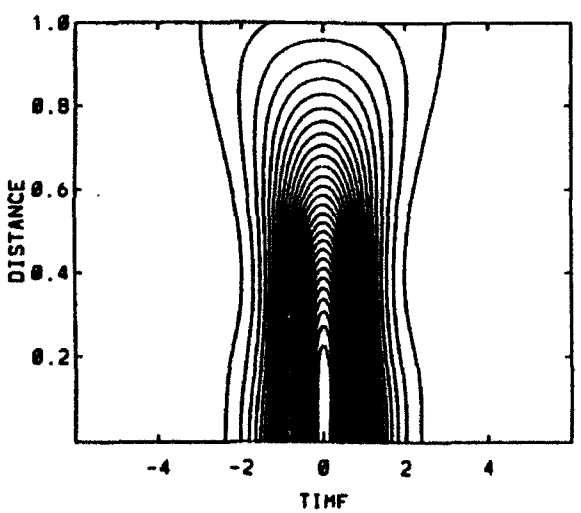
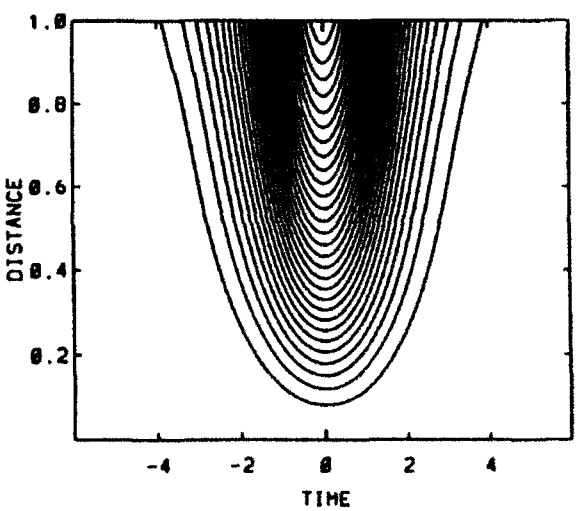
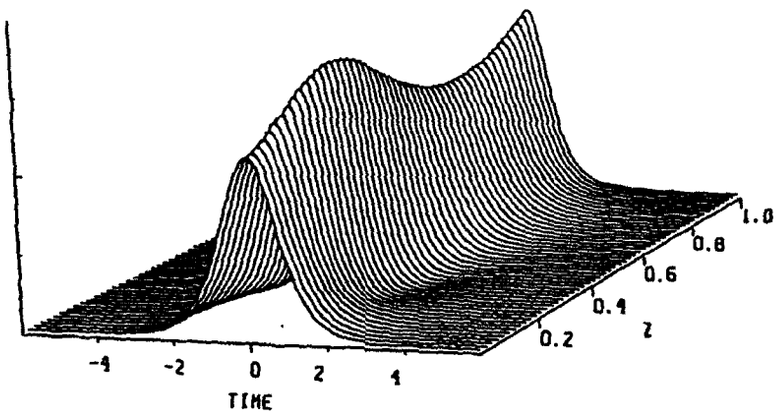

1.0
0.8
0.6
0.4
0.2
DISTANCE
-4
-2
0
2
4
TIME

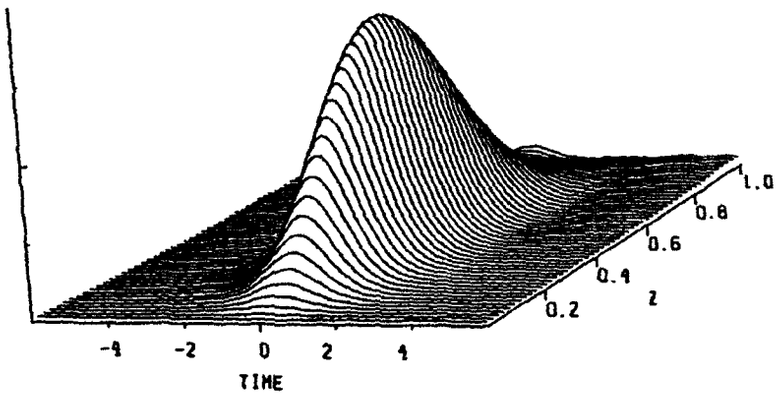
Fig. 2

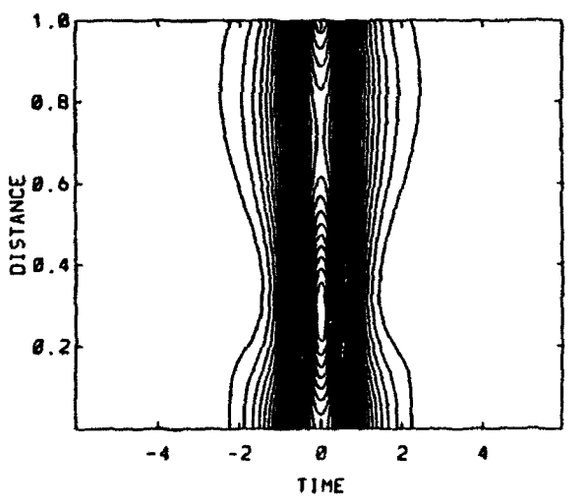

1.0
0.8
0.6
0.4
0.2
DISTANCE
-4
-2
0
2
4
TIME

(1)


-4
-2
0
2
4
TIME
0.2
0.4
0.6
0.8
1.0
Z

(2)


-4
-2
0
2
4
TIME
0.2
0.4
0.6
0.8
1.0
Z


1.0
0.8
0.6
0.4
0.2
DISTANCE
-4
-2
0
2
4
TIME

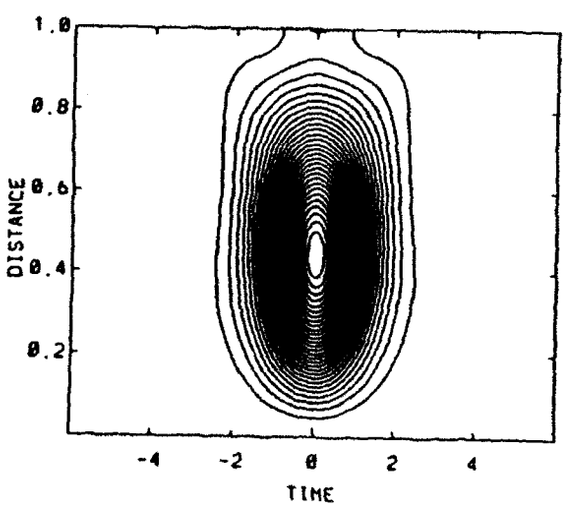

1.0
0.8
0.6
0.4
0.2
DISTANCE
-4
-2
0
2
4
TIME

Fig. 3

Gain Mechanisms

In order to compensate for TPA, I investigated the effects of gain. However because TPA is a nonlinear loss, the gain mechanism would be more complex. Introducing the parameters γ_0 for the small signal gain and γ_2 for the gain dispersion, the coupled set of equations is obtained

$$\begin{aligned} iq_{1z}+\tfrac{1}{2}\beta_2 q_{1tt}+\delta(|q_1|^2+\sigma|q_2|^2)q_1+kq_2 \\ =i\gamma_0 q_1-\frac{i}{2}\gamma_2 q_{1tt}-i\beta(|q_1|^2+\sigma|q_2|^2)q_1 \end{aligned}$$

$$\begin{aligned} iq_{2z}+\tfrac{1}{2}\beta_2 q_{2tt}+\delta(|q_2|^2+\sigma|q_1|^2)q_2+kq_1 \\ =i\gamma_0 q_2-\frac{i}{2}\gamma_2 q_{2tt}-i\beta(|q_2|^2+\sigma|q_1|^2)q_2 \end{aligned} \tag{5}$$

The exact solution for this case is given by

$$q_1(z,t)=\frac{1}{\sqrt{2}}(\sin\theta e^{ikz}-\cos\theta e^{i\phi}e^{-ikz})q_0 e^{iRz}Q(\omega t)$$

$$q_2(z,t)=\frac{1}{\sqrt{2}}(\sin\theta e^{ikz}+\cos\theta e^{i\phi}e^{-ikz})q_0 e^{iRz}Q(\omega t)$$

where $Q(\omega t)=\mathrm{sech}(\omega t)\exp\{i\alpha\ln[\mathrm{sech}(\omega t)]\}$

$$\begin{aligned} &\alpha^2-B\alpha-2=0\,, \\ &B=3\frac{\beta_2\delta+\gamma_2\beta}{\gamma_2\delta-\beta\beta_2}\,, \\ &\omega^2=\frac{2\gamma_0}{2\beta_2\alpha+\gamma_2(1-\alpha^2)}\,, \\ &R=[\tfrac{1}{2}\beta_2(1-\alpha^2)-\gamma_2\alpha]\omega^2\,, \\ &|q_0|^2=\frac{\omega^2}{\delta}[\tfrac{1}{2}\beta_2(2-\alpha^2)-\tfrac{1}{2}\gamma_2\alpha]\,, \end{aligned} \tag{6}$$

Figure 4 shows a plot of the results for light incident into one waveguide. It can be seen that the pulse shape is well preserved. This result will be true for any value of TPA and hence represents an important finding.

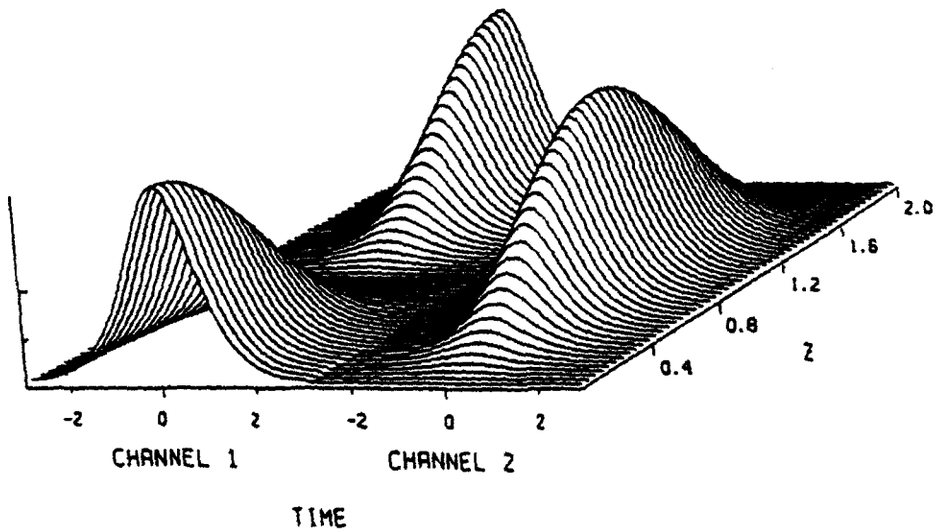


Fig. 4

Summary

This research has set the stage for future experimental work in this area. In particular, the results will be investigated using a Ti:sapphire laser system with samples grown by molecular beam epitaxy. Band structure calculations will provide important guidance for the determination of the semiconductor well thickness and alloy concentration. Further interactions with researchers at Wright-Patterson AFB is envisioned.

BIBLIOGRAPHY

1. S. M. Jensen, IEEE J Quantum ELectron. QE18, 1580 (1982).
2. Y. Silberberg and G.I. Stegeman, Appl. Phys. Lett. 50, 801 (1987).
3. S. Wabnitz, E.M. Wright, C.T. Seaton, and G.I. Stegeman, Appl. Phys.. Lett 49,11, (1986).
4. L. Thylen, N. Finlayson, C.T. Seaton, and G.I. Stegeman, Appl. Phys. Lett. 51,1304 (1987).
5. M.N. Islam, S.P. Dijaili, and J. Gordon, Opt. Lett. 13,518 (1988).
6. A. Mecozzi, S. Trillo, S. Wabnitz, and B. Diano, Opt. Lett. 12,275 (1987).
7. B. Daino, G.Gregori, and S. Wabnitz, Opt. Lett. 11,42 (1986).
8. D. D. Gusovskii, B. Diano, A.A. Maier, V.B. Neustruev, E.I. Shkloskii, and I.A. Sheherbakov, Sov. J. Quantum. Electron. 15, 1523 (1985).
9. S. R. Friberg, Y.Silberberg, M.K. Oliver, M. J. Andrejco, M.A. Saifi, and P.W. Smith, Appl. Phys. Lett. 51, 1135 (1987).
10. P. Li Kam Wa, P.N. Robson, J.S. Roberts, M.A. Pate, amd J.P. David, Appl. Phys. Lett, 52, 2013 (1988).
11. P. Li Kam Wa, A. Miller, J. S. Roberts, and P.N. Robson, Integrated Photonics Meeting, Hilton Head, S.C., Mar. 25-28, 1990, paper WH7.
12. P. Li Kam Wa, J.E. Sitch, N. J. Mason, J.S. Roberts, and P.N. Robson, Electron. Lett. 21, 26 (1985).
13. P.Li Kam Wa, P.N. Robson, J.P.R. David, G. Hill, P. Mistry, M.A. Pate, and J.S. Roberts, Electron. Lett. 22, 1129 (1986).
14. P.R. Berger, Y. Chen, P. Bhattacharya, J. Pamulapati and G. C. Vezzoli, Appl. Phys. Lett. 52, 1125 (1988).
15. G.I. Stegeman, E. Caglioti, S. Trillo, and S. Wabnitz, Opt. Comm. 53, 281 (191987).
16. E. Caglioti, S Trillo, S. Wabnitz, and G.I. Stegeman, J. Opt. Soc. Am. B 5, 472 (1988).
17. G. Bastard, C. Delalande, Y. Guldner and P. Voisin in Advances in Electronics and Electron Physics, P. W. Hawkes, ed. (Academic Press, NY, 1988).

18. J. W. Matthews and A.E. Blakeslee, J. Cry. Growth 27, 118 (1974).
19. F. H. Pollak and M. Cardona, Phys. Rev. 172, 816 (1968).
20. J. Y. Marzin and J. M. Gerard, Superlattices and Microstruct. 5, 51 (1989).
21. C. Mailhoit and D. L. Smith, Phys. Rev. B 33, 8360 (1986).
22. D. L. Smith and C. Mailhoit, Phy. Rev. Lett. 58, 1264 (1987).
23. G. C. Osbourn, J. Vac. Sci. Tech. 21, 459 (1982).
24. S.T. Picraux, L. R. Dawson, G.C. Osbourne, and W.K. Chu, Appl. Phy. Lett. 43, 930 (1983).
25. K. J. Moore, G. Duggan, K. Woodbridge, and C. Roberts, Phys. Rev. B 41, 1090 (1990).
26. B. K. Laurich, K. Elcess, C. G. Fonstad, J.G. Beery, C. Mailhoit, and D. L. Smith, Phys. Rev. Lett. 62, 649 (1989).
27. D. S. Chemla, I. Bar Joseph, J. M. Kau, T.Y. Chang, C. Klingshirn, G. Livescu, and D.A.3. Miller, IEEE J Quantum Electron QE24, 1664 (1988).
28. Y. H. Lee, A. Chavez-Pirson, B.K. Rhee, and H. Gibbs, Appl. Phys. Let. 49, 1505 (1986).
29. D.S. Chemla, D.A.B. Miller, P.W. Smith, A.C. Gossard, and W.Wiegmann, IEEE J. Quantum Electron QE20, 265 (1984).
30. D.S. Chemla and D.A.B. Miller, J. Opt. Soc. Am B2, 1155 (1985).
31. H. Haug and S. Schmitt-Rink, J. Opt. Soc. Am. B2, 1135 (1985).
32. S. Schmitt-Rink, .D. S. Chemla and D.A. B. Miller, Phys. Rev. B 32, 6601 (1985).
33. M. Yamanishi andf M. Kurosaki, IEEE J. Quantum Electron, 24, 325 (1988).
34. A. Mysyrowicz, D. Hulin, A. Antonetti, A. Migus, W.T. Masselink, and H. Morkoc, Phys. Rev. Lett. 56, 2748 (1986).
35. A. VonLehmen. D.S. Chemla, J.E. Zucker, and J. P. Heritage, Opt. Lett. 11, 609 (1989).
36. W. H. Knox, D.S. Chemla, D.A.B. Miller, J.B. Stark, and S. Schmitt-Rink, Phys. Rev. Lett. 62, 1189 (1989).
37. W. K. Burns and N. Bloembergen, Phys. Rev. B4, 3437 (1971).
38. R. Jin, C. L. Chuang, H. M. Gibbs, S.W. Koch, J.N. Polky, and G.A. Pubanz, Appl. Phys. Lett, 53, 1791 (1988).

39. Y.H. Lee, A. Chavez-Pirson, S.W. Koch, H.M. Gibbs, S.H. Park, J. Morhange, A. Jeffery, N. Peyghambarian, L. Banyai, A.C. Gossard and W. Wiegmann, Phys. Rev. Lett. 57, 2446 (1986).
40. G. R. Olbright and N. Peyghambarian, Appl. Phys. Lett., 48, 1184 (1986).
41. D. Cotter, C.N. Ironside, B.J. Ainslie, and H.P. Girdlestone, Opt. Lett. 14, 317 (1989).
42. M. Sheik-babae, A.A. Said, and E. W. VanStryland, Opt. Lett., 14, 955 (1989).
43. H.Q. Le, H.K. Choi and C.A. Wang, Appl. Phys. Lett. 57, 212 (1990).

DEVELOPMENT OF CONTROL DESIGN METHODOLOGIES FOR FLEXIBLE (HIGH ORDER) MISSILE SYSTEMS WITH MULTIPLE HARD NONLINEARITIES

Armando A. Rodriguez
Assistant Professor
Deptartment of Electrical Engineering

Arizona State University
Tyler & Rural
Tempe, AZ 85287-5706

Final Report for:
Summer Research Program
Wright Laboratory

Sponsored by:
Air Force Office of Scientiffic Research
WL/MNAG, Eglin Air Force Base, FL.

September 1992

DEVELOPMENT OF CONTROL DESIGN METHODOLOGIES FOR FLEXIBLE (HIGH ORDER) MISSILE SYSTEMS WITH MULTIPLE HARD NONLINEARITIES

Armando A. Rodriguez
Assistant Professor
Deptartment of Electrical Engineering
Arizona State University

1 Abstract

The primary goal of this research endeavour has been to initiate a program directed at the development of practical methods for designing robust full envelop autopilots for *flexible missile systems with multiple hard nonlinearities*. The methods developed will permit control engineers to systematically generate full envelop controllers on the basis of very complex (high order) missile models.

The research has been driven by the ultimate goal of designing a full envelop autopilot for an Extended Medium-Range Air-to-Air Technology (EMRAAT) Bank-to-Turn (BTT) missile with significant structural modes. Of particular interest is the terminal phase of an air-to-air intercept during which control surfaces may saturate in both position and rate. This is critical because structural modes may be excited and, more importantly, missile destabilization may be induced. Assuring robust performance during such a high control scenario has guided the path of this research.

Contents

DEVELOPMENT OF CONTROL DESIGN METHODOLOGIES FOR FLEXIBLE (HIGH ORDER) MISSILE SYSTEMS WITH MULTIPLE HARD NONLINEARITIES

Armando A. Rodriguez

2 Motivation

This research is motivated by two problems which have received considerable attention in recent years and will receive far greater attention as 21st century aerospace technology is developed. The first problem is that of controlling flexible systems. The second is controlling systems with multiple hard nonlinearities. In particular, this research has been driven by the lack of systematic procedures to develop full envelop autopilots for flexible (high order) missiles with multiple hard nonlinearities.

2.1 Controlling Flexible (High Order) Missile Systems

Because of recent advances in material science, many aerospace systems (e.g. missiles, airplanes, space platforms, etc.) are being designed using composites. Use of such materials typically provides strength and mission required radar signature and/or thermal properties. Such materials also result in much lighter vehicles/platforms. The trade-off in going to composites has been systems with more pronounced structural modes. Developing systematic procedures for designing full envelope controllers for such systems is one focus of the proposed research. This problem becomes particularly difficult when the structural modes conflict with system bandwith requirements: a more flexible missile makes the control problem more challenging. Also, if one is trying to accurately model turbulence for a missile, then the CFD (Computational Fluid Dynamics) models can be very large [3] and *performance based model reduction* becomes a critical issue. Such control/modelling issues have been motives for this research.

2.2 Controlling Systems with Multiple Hard Nonlinearities

Traditionally, control design for aerospace systems has been based on gain scheduling [11]. Gain scheduling is a method whereby linear controllers, based on linear models of the true nonlinear system, are "pieced" together over the operating envelope of the system. The initial designs often do not take into account the effects of hard nonlinearities such as saturating actuators and rate limiters. One motive for this research has been the lack of systematic methods to modify the original designs so that such nonlinearities can be accommodated.

For physical reasons, system designers often want to limit certain variables (e.g. angle of attack, sideslip angle, etc.). Typically, adhoc modifications are employed and extensive simulations must be performed to justify the modifications. A procedure for systematizing this process over the full envelope of the system has been another motive for this research.

3 A Methodology for Infinite-Dimensional Systems

Flexible systems when properly modelled are governed by partial differential equations (e.g. Euler Bernoulli Beam [24]). Such systems are said to be infinite-dimensional [48]. The lack of systematic procedures for controlling such systems motivated the initial stages of this research.

3.1 Introduction

The problem of designing compensators for infinite-dimensional plants has received considerable attention during the past decade. Some relevant works are [15],[16], [19]-[27], [30]-[32], [34],[35], [38]-[41], [43]-[51], [54], [56]-[61].

This initial stage of this research has been motivated by the following question: How can a finite-dimensional controller, satisfying specific $\mathcal{H}^\infty$ design specifications, be systematically obtained for a given infinite-dimensional plant ? Two approaches to this problem have appeared in the literature.

1. A Design/Approximate approach in which a controller is designed using infinite-dimensional techniques. If the controller is infinite-dimensional, a finite-dimensional approximation is obtained. This approach is addressed in [20]-[27] and some of the other aforementioned references. Because the Design/Approximate approach is difficult to apply to general plants, it was not condidered throughout the course of this research. Instead, an Approximate/Design approach has been pursued.

2. In the Approximate/Design approach, the plant is first approximated by a finite-dimensional model and then a finite-dimensional controller is designed based on this model. This is a widely used approach. However, this approach generally comes with no guarantees. The key difficulties which have arisen can be attributed to the fact that these performance measures are often not continuous with respect to plant perturbations, even when the uniform topology is imposed [20], [54].

   In [31], [54], and [56],[57], the authors address issues related to the Approximate/Design approach. In [56], [57], for example, finite-dimensional techniques are used to obtain a near-optimal infinite-dimensional compensator. To obtain the near-optimal compensator, an iterative procedure is presented. The procedure, however, assumes that the optimal performance measure is known, that a certain function is continuous on the extended imaginary axis, and that a particular spectral factorization can be performed.

This section considers the problem of designing near-optimal finite-dimensional compensators for stable infinite-dimensional plants. Standard weighted $\mathcal{H}^\infty$ sensitivity and mixed-sensitivity measures are used to define the notion of optimality. The method of solution is based on finite-dimensional techniques applied to finite-dimensional approximants of the original plant. It is shown in this section, for the sensitivity minimization problem, that it is sufficient to obtain approximants of the plant on compact sets, provided that the inner part of the plant is approximated in some sense. This problem is shown not to exist when the control input is penalized. Constructive algorithms, for both of the above cases, are presented. New results on the convergence of actual closed loop transfer functions are also given.

The remainder of this section is organized as follows. For notation and mathematical details, the reader is referred to [48]. Subsection 3.2 contains a precise statement of three fundamental problems to be addressed. Subsection 3.3 presents new results on $\mathcal{H}^\infty$ model matching problems. Subsection 3.4 contains the solution to the $\mathcal{H}^\infty$ sensitivity problems. An example is given in subsection 3.4. Subsection 3.5 presents the solution to the $\mathcal{H}^\infty$ mixed-sensitivity problems. Finally, section 3.6 summarizes the paper and presents directions for future research.

3.2 Statement of Three Fundamental Problems

In this subsection three fundamental problems associated with the proposed *Approximate/Design* approach are precisely defined for weighted $\mathcal{H}^\infty$ sensitivity and mixed-sensitivity performance criteria. First, basic assumptions and definitions which will be used throughout the sequel are stated.

Throughout the section, focus is placed exclusively on $\mathcal{L}^2$ finite-gain stable plants [17]. This is done in order to better illustrate the main ideas. The case in which the plant is unstable is treated in [50].

Let $P(s) \in \mathcal{H}^\infty$ denote a stable transfer function of a scalar infinite-dimensional plant. Also, let $\{P_n(s)\}_{n=1}^\infty \subset R\mathcal{H}^\infty$ denote a sequence of stable finite-dimensional approximants of P; the sense of which is to be stated shortly.

Let P denote some infinite-dimensional system and K a compensator. Given $F \in \mathcal{H}^\infty$, let $K(F,Q)$ generate all compensators which stabilize F when Q is varied over $\mathcal{H}^\infty$ [58]. Let $J(\cdot, K(\cdot,\cdot)) : \mathcal{H}^\infty \times \mathcal{H}^\infty \times \mathcal{H}^\infty \rightarrow R_+$ denote a cost function or performance measure. It will be referred to as the J - measure. The *optimal performance* for the infinite-dimensional plant P with respect to the J - measure is then defined as follows (see figure 6).

Definition 3.1 (Optimal Performance)

$$\mu_{opt} \stackrel{\text{def}}{=} \inf_{Q \in \mathcal{H}_0^\infty} J(P, K(P,Q)).$$

This is the *target* problem to be solved, when P is infinite-dimensional. Direct approaches for solving this problem have been proposed by various researchers, e.g. see [20]-[27], [39]-[41].

The approach taken here requires that P be approximated by a finite-dimensional system P_n. This, then motivates the following finite-dimensional optimzation problem.

Definition 3.2 (Expected Performance)

$$\mu_n \stackrel{\text{def}}{=} \inf_{Q \in R\mathcal{H}_0^\infty} J(P_n, K(P_n,Q)).$$

In the context of this work, μ_n will be referred to as the *expected performance*. This terminology for μ_n is motivated by the fact that the numbers μ_n are typically used to guide engineers during the design process.

Let Q_n be any near-optimal or optimal solution for the problem in definition 3.2. It should be noted that, in general, any optimal Q_n will be improper (e.g. when J is a weighted sensitivity criterion and P_n is strictly proper [28]).

By the Youla, et al parameterization [58], it follows that Q_n generates an "internally" stabilizing compensator K_n for P_n (see figure 2). This compensator is given by: $K_n \stackrel{\text{def}}{=} K(P_n, Q_n) = \frac{-Q_n}{1-P_nQ_n}$. Because in general, K_n may not be near-optimal with respect to the problem in definition 3.1, and in fact not even stabilizing for P , nor strictly proper, a different finite-dimensional compensator is needed.

This motivates the definition of a "roll-off operator" $r : Q_n \rightarrow \tilde{Q}_n$ which maps Q_n to $\tilde{Q}_n \in R\mathcal{H}_0^\infty$. From the above, it is clear that the roll-off operator r must be chosen intelligently. How this should be done will be answered in what follows.

Given this, one then considers the feedback system obtained by substituting the finite-dimensional compensator generated by $\tilde{Q}_n$, namely $\tilde{K}_n \stackrel{\text{def}}{=} K(P_n, \tilde{Q}_n) = \frac{-\tilde{Q}_n}{1-P_n\tilde{Q}_n}$, into a closed loop system with the infinite-dimensional plant P (see figure 3). Assuming "internal" stability can be shown [17], this then motivates the following "natural" definition for the *actual performance*.

Definition 3.3 (Actual Performance)

$$\tilde{\mu}_n \stackrel{\text{def}}{=} J(P, \tilde{K}_n).$$

Given the above discussion, the *Approximate/Design J-Problem* is defined as follows.

Problem 3.1 (Approximate/Design)
Find conditions on the approximants $\{P_n\}_{n=1}^\infty$ and on the roll-off operator r such that $\lim_{n\rightarrow\infty} \tilde{\mu}_n = \mu_{opt}$.

This problem will be solved in the sequel for two distinct $\mathcal{H}^\infty$ performance measures J; one is a weighted $\mathcal{H}^\infty$ sensitivity measure and the other is a weighted $\mathcal{H}^\infty$ mixed-sensitivity measure.

Each of these measures define an *Approximate/Design* problem. These problems are difficult for several reasons. First, the above measures can be shown to be discontinuous with respect to perturbations in the plant P, even when the uniform topology on $\mathcal{H}^\infty$ is imposed [54]. It should be noted, however, that even if these measures were continuous in the uniform topology, there are many infinite-dimensional plants which cannot be approximated uniformly by real-rational functions. A delay, for example, is not continuous on the extended imaginary axis and hence cannot be approximated uniformly by a sequence of $R\mathcal{H}^\infty$ functions. A second difficulty can be attributed to the fact that the above measures generally have optimal solutions which exhibit "bad" properties. More specifically, one can show that the infimizers for these measures are typically unbounded and hence generate unbounded ("improper") compensators.

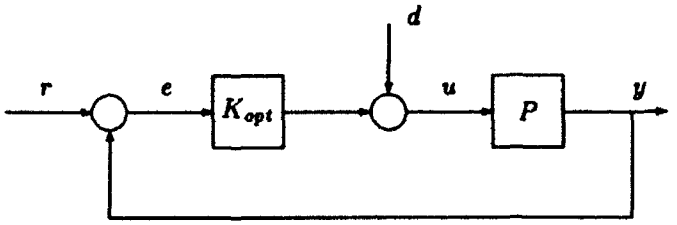


Figure 1: Optimal Infinite-Dimensional Feedback Loop: $\mu_{opt} = J(P, K_{opt})$

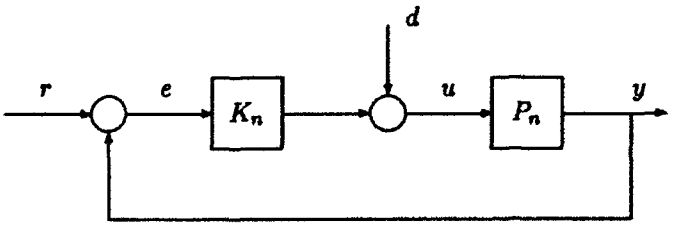


Figure 2: Purely Finite-Dimensional Feedback Loop: $\mu_n = J(P_n, K_n)$

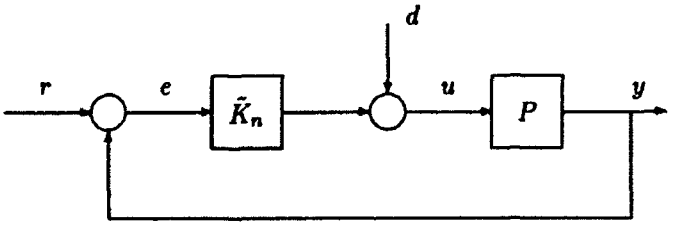


Figure 3: Actual Near-Optimal Feedback Loop: $\tilde{\mu}_n = J(P, \tilde{K}_n)$

In practice, one would like to be able to compute μ_{opt} using finite-dimensional algorithms. For the above sensitivity/mixed-sensitivity criteria, the literature has typically considered infinite-dimensional eigenvalue/eigenfunction problems in order to compute μ_{opt} [20, pp. 28-31], [61, pp. 308]. With an ultimate intention of providing purely finite-dimensional techniques for computing μ_{opt}, the following "purely" finite-dimensional problem is considered.

Problem 3.2 (Purely Finite-Dimensional)
Find conditions on the approximants $\{P_n\}_{n=1}^{\infty}$ such that $\lim_{n\to\infty} \mu_n = \mu_{opt}$. In the context of this work, this problem will be referred to as the *Purely Finite-Dimensional J-Problem.*

Finally, the following problem will also be considered in the sequel.

Problem 3.3 (Loop Convergence)
Given that $\lim_{n\to\infty} \tilde{\mu}_n = \mu_{opt}$ and $\lim_{n\to\infty} \mu_n = \mu_{opt}$, when do the "loop shapes" of the actual near-optimal system "approximate" those of the optimal, or a near-optimal, infinite-dimensional system? In the context of this work, this problem will be referred to as the *Loop Convergence J-Problem.*

3.3 $\mathcal{H}^\infty$ Model Matching Problems

In this subsection two $\mathcal{H}^\infty$ *Model Matching Problems* are examined. The first is an infinite-dimensional *Model Matching Problem* (MMP). The second is a finite-dimensional MMP.

3.3.1 Infinite-Dimensional $\mathcal{H}^\infty$ MMP

In this subsection, the so-called $\mathcal{H}^\infty$ *Model Matching Problem* is examined. The solution to this problem will provide insight needed to solve the sensitivity and mixed-sensitivity problems being addressed in the paper. The $\mathcal{H}^\infty$ Model Matching Problem is defined in terms of the following optimization problem:

Definition 3.4 (Infinite-Dimensional Model Matching Problem)

$$\mu_o \stackrel{\text{def}}{=} \inf_{Q\in\mathcal{H}_0^\infty} \|T_1 - T_2 Q\|_{\mathcal{H}^\infty}$$

where T_1, $T_2 \in \mathcal{H}^\infty$. Given this, the $\mathcal{H}^\infty$ *Model Matching Problem* is that of finding a *near-optimal* Q-parameter in $\mathcal{H}_0^\infty$; i.e. given an $\epsilon > 0$, however small, find $Q_o \in \mathcal{H}_0^\infty$ such that $\mu_o \le \|T_1 - T_2 Q_o\|_{\mathcal{H}^\infty} \le \mu_o + \epsilon$.

Since $T_2 \in \mathcal{H}^\infty$, it possesses an inner-outer factorization over $\mathcal{H}^\infty$ [53, pp. 344]. This permits one to write $T_2 = T_{2_i} T_{2_o}$ where T_{2_i} is inner in $\mathcal{H}^\infty$ and T_{2_o} is outer in $\mathcal{H}^\infty$. It should be noted that obtaining an inner-outer factorization for a general $\mathcal{H}^\infty$ function T_2 is an unresolved problem. This issue is the subject of on-going research [21].

Throughout this subsection, the following technical assumption will be made on the functions T_1 and T_2.

Assumption 3.1 (T_1 and T_2)
(1) $T_1 \in \mathcal{C}_e$.
(2) T_{2_o} has a finite number of distinct zeros on the extended imaginary axis.
(3) Each imaginary zero of T_{2_o} has finite algebraic multiplicity [48].
(4) $T_{2_o} \in \mathcal{C}_e$ [1].
(5) T_{2_i} is continuous on the imaginary axis, except possibly at infinity [2].

Although it is assumed that T_{2_i} is continuous on the finite imaginary axis, it will become apparent in construction 3.1 that this assumption can be relaxed to accommodate finite discontinuities. Also, because T_{2_i} is permitted to be discontinuous at infinity, this work removes the continuity assumptions required in [56], [57].

In order to gain insight into the $\mathcal{H}^\infty$ *Model Matching Problem*, one usually starts with some version of the following proposition [28], [42, pp. 46]. The proposition emphasizes, and quantifies, the non-invertibility of T_{2_i}.

[1] The case where T_{2_o} is invertible in $\mathcal{H}^\infty$ is trivial and will not be addressed.
[2] T_{2_i} can contain, for example, a delay and/or an infinite Blaschke product [33, pp. 63 - 68; pp. 132].

Proposition 3.1 (Inner Problem)

$$\inf_{Z\in\mathcal{H}^\infty} \|T_1 - T_2 Z\|_{\mathcal{H}^\infty} = \left\|\Gamma_{T_1 T_{2_i}^*}\right\|$$

Moreover, there exists $Z_o \in \mathcal{H}^\infty$, not necessarily unique, which attains the infimum.

The existence of an infimizer Z_o can be proved using weak* convergence arguments [14, pp. 134-137], [36, pp. 128], [29, pp. 85 - 86]. The fact that the infimum is $\left\|\Gamma_{T_1 T_{2_i}^*}\right\|$ follows from Nehari's Theorem [38]. Conditions are given in [26] under which Z_o is unique. The reader is also referred to [20, pp. 26 - 31].

The following lemma gives further insight into the $\mathcal{H}^\infty$ *Model Matching Problem.*

Lemma 3.1 (Bounds on μ_o)

$$\max\{|T_1(j\infty)|,\ \max_k |T_1(j\omega_k)|,\ \left\|\Gamma_{T_1 T_{2_i}^*}\right\|\} \le \mu_o \le \|T_1\|_{\mathcal{H}^\infty}$$

Lemma 3.1 gives a lower bound on μ_o which depends on T_{2_i} and on the imaginary zeros of T_{2_o}. This supports the conjecture that a near-optimal Q should invert T_{2_o}, away from its imaginary zeros, in the closed right half plane. In the sequel, it will be shown that the above lower bound is actually equal to μ_o. The lemma also shows that since $T_1 \in \mathcal{H}^\infty$, μ_o will necessarily be finite.

The interest in the "inner problem" will be to use its solution Z_o to construct near-optimal solutions for the $\mathcal{H}^\infty$ *Model Matching Problem.* The idea will be to appropriately modify ("roll-off") $T_{2_o}^{-1} Z_o$ so that the resulting function is admissible [3] and near-optimal. This modification will be carried out using the following two $\mathcal{H}^\infty$ functions.

Definition 3.5 (Roll-Off Functions)

$$f_m(s) \stackrel{\text{def}}{=} \left(\frac{s}{s+\frac{1}{m}}\right)^{\frac{m_0}{m}} \prod_{k=1}^{l-1} \left(\frac{s^2+\omega_k^2}{s^2+2\frac{1}{m}s+\omega_k^2+\frac{1}{m}^2}\right)^{\frac{m_k}{m}} \left(\frac{m}{s+m}\right)^{\frac{m_{l+1}}{m}}$$

and

$$g_n(s) \stackrel{\text{def}}{=} \left(\frac{s}{s+\frac{1}{n}}\right)^{m_0} \prod_{k=1}^{l-1} \left(\frac{s^2+\omega_k^2}{s^2+2\frac{1}{n}s+\omega_k^2+\frac{1}{n}^2}\right)^{m_k} \left(\frac{n}{s+n}\right)^{m_l}$$

where $m, n \in Z_+$.

It should be noted that f_m is irrational even if the m_k are integers. g_n will be irrational if any one of the m_k are non-integers. The next theorem shows how these functions can be used to construct a near-optimal solution for the $\mathcal{H}^\infty$ *Model Matching Problem.*

Theorem 3.1 (Near-Optimal Irrational Solution)
Let $Q_{m,n} \stackrel{\text{def}}{=} T_{2_o}{}^{-1} Z_o f_m g_n$ where f_m and g_n were defined in definition 3.5. Given $\epsilon > 0$, however small, there exists $M \stackrel{\text{def}}{=} M(\epsilon, T_1) \in Z_+$ and $N \stackrel{\text{def}}{=} N(\epsilon, M, \|T_1\|_{\mathcal{H}^\infty}) \in Z_+$ such that

$$\mu_o \le \|T_1 - T_2 Q_{M,n}\|_{\mathcal{H}^\infty} \le \mu_o + 3\epsilon$$

for all $n \ge N$. Consequently, given $\epsilon > 0$, there exists a single function $f \in \mathcal{H}_0^\infty \cap C_e$ which depends on ϵ and $\|T_1\|_{\mathcal{H}^\infty}$, is independent of Z_o, and such that

$$\mu_o \le \|T_1 - T_2 Q_o\|_{\mathcal{H}^\infty} \le \mu_o + 3\epsilon$$

where $Q_o \stackrel{\text{def}}{=} Z_o f \in \mathcal{H}_0^\infty$. Moreover, f has the form $f = T_{2_o}^{-1} f_M g_N$. Finally,

$$\mu_o = \max\{|T_1(j\infty)|,\ \max_k |T_1(j\omega_k)|,\ \left\|\Gamma_{T_1 T_{2_i}^*}\right\|\}.$$

It will be critical later on that the function f can be constructed without knowing the function Z_o which, in general, may be difficult to obtain.

[3] In general, $T_{2_o}{}^{-1} Z_o$ will have poles on the extended imaginary axis.

3.3.2 Sequences of Finite-Dimensional $\mathcal{H}^\infty$ MMP's

In this subsection, the infinite-dimensional $\mathcal{H}^\infty$ *Model Matching Problem* is solved by approximating T_1 and T_2 by the $\mathcal{RH}^\infty$ functions, T_{1_n} and T_{2_n}, respectively. The construction of these approximants is now presented.

Construction 3.1 (Construction of $\{T_{1_n}\}_{n=1}^\infty$ and $\{T_{2_n}\}_{n=1}^\infty$)
Let $\{T_{1_n}\}_{n=1}^\infty \subset \mathcal{RH}^\infty$ be any sequence which uniformly approximates T_1 in $\mathcal{H}^\infty$; i.e.

$$\lim_{n\to\infty} \|T_{1_n} - T_1\|_{\mathcal{H}^\infty} = 0. \tag{1}$$

Let $\{T_{2_{n_i}}\}_{n=1}^\infty \subset \mathcal{RH}^\infty$ be any sequence of inner functions which approximates T_{2_i} uniformly on compact frequency intervals of the imaginary axis; i.e.

$$\lim_{n\to\infty} \left\|(T_{2_{n_i}} - T_{2_i})X_\Omega\right\|_{\mathcal{H}^\infty} = 0 \tag{2}$$

for each $\Omega \in R_+$, however large. It will be said that $T_{2_{n_i}}$ is a *compact approximant* for T_{2_i}. Let $\{T_{2_{n_o}}\}_{n=1}^\infty \subset \mathcal{RH}^\infty$ be any sequence of outer functions which approximates T_{2_o} uniformly in $\mathcal{H}^\infty$; i.e.

$$\lim_{n\to\infty} \left\|T_{2_{n_o}} - T_{2_o}\right\|_{\mathcal{H}^\infty} = 0. \tag{3}$$

Finally, define

$$T_{2_n} \stackrel{\text{def}}{=} T_{2_{n_o}} T_{2_{n_i}}. \tag{4}$$

The sequences $\{T_{1_n}\}_{n=1}^\infty$ and $\{T_{2_{n_o}}\}_{n=1}^\infty$ are guaranteed to exist since $T_1, T_{2_o} \in \mathcal{H}^\infty \cap C_e$. The sequence $\{T_{2_{n_i}}\}_{n=1}^\infty$ is guaranteed to exist since T_{2_i} is assumed to be continuous on the finite imaginary axis. If T_{2_i} has discontinuities on the finite imaginary axis, then the convergence in (2) would be required to hold only on compact subsets of the imaginary axis which exclude the discontinuities. It should be emphasized that T_{2_i} may, and in general will, be discontinuous at infinity. This case is thoroughly addressed in this paper.

A consequence of the above construction is that T_{2_n} approximates T_2 uniformly on compact frequency intervals of the imaginary axis; i.e. $\lim_{n\to\infty} \|(T_{2_n} - T_2)X_\Omega\|_{\mathcal{H}^\infty} = 0$ for any given $\Omega \in R_+$. This compact convergence property will be exploited in what follows.

Given the above construction, it is natural to consider the purely finite-dimensional $\mathcal{H}^\infty$ *Model Matching Problem* defined by the following optimization problem.

Definition 3.6 (Purely Finite-Dimensional Model Matching Problem)

$$\mu_n \stackrel{\text{def}}{=} \inf_{Q\in\mathcal{RH}_0^\infty} \|T_{1_n} - T_{2_n}Q\|_{\mathcal{H}^\infty}$$

Let $Z_n \in \mathcal{RH}^\infty$ be the unique solution [28, pp. 66-74] of

$$\min_{Z\in\mathcal{RH}^\infty} \left\|T_{1_n} - T_{2_{n_i}}Z\right\|_{\mathcal{H}^\infty} = \left\|T_{1_n} - T_{2_{n_i}}Z_n\right\|_{\mathcal{H}^\infty} = \left\|\Gamma_{T_{1_n}T_{2_{n_i}}^*}\right\|.$$

The following theorem establishes the existence of near-optimal real-rational solutions for the finite-dimensional and infinite-dimensional model matching problems being considered. More importantly, the theorem shows how to construct near-optimal real-rational solutions for these problems. The theorem also contains very useful new convergence results which represent a major contribution of this work.

Theorem 3.2 (Sub-Optimal Real-Rational Solutions & Properties)
Suppose that $|T_1(j\infty)| \le \left\|\Gamma_{T_1T_{2_i}^*}\right\|$. Let $\tilde{Q}_n = Z_n f_n \in \mathcal{RH}_0^\infty$ where $f_n \in \mathcal{RH}_0^\infty$ (not to be confused with f_m in definition 3.5) approximates the function f from theorem 3.1 uniformly in $\mathcal{H}^\infty$; i.e $\lim_{n\to\infty} \|f_n - f\|_{\mathcal{H}^\infty} = 0$. Given $\epsilon > 0$, however small, it follows that there exists $N_1 \stackrel{\text{def}}{=} N_1(\epsilon, T_1, T_2, f) \in Z_+$ such that

$$\mu_n \le \left\|T_{1_n} - T_{2_n}\tilde{Q}_n\right\|_{\mathcal{H}^\infty} \le \mu_o + 11\epsilon$$

for all $n \geq N_1$. Moreover, $\lim_{n\to\infty} \left\|(T_2 - T_{2_n})\tilde{Q}_n\right\|_{\mathcal{H}^\infty} = 0$ and the following "upper-semicontinuity" property holds $\lim_{n\to\infty} \sup_{k\geq n} \mu_k \leq \mu_o$. Finally, there exists $N_2 \stackrel{\text{def}}{=} N_2(\epsilon, T_1, T_2, f) \in Z_+$ such that $\mu_o \leq \left\|T_1 - T_2\tilde{Q}_n\right\|_{\mathcal{H}^\infty} \leq \mu_o + 14\epsilon$ for all $n \geq N_2$. If $T_{2_{n_o}}$ is constructed so that its zero structure on the extended imaginary axis is identical to that of T_2 (i.e. $T_{2_{n_o}}(j\omega) = 0$ if and only if $T_2(j\omega) = 0$), then $\mu_o \leq \lim_{n\to\infty} \inf_{k\geq n} \mu_k$ and hence $\lim_{n\to\infty} \mu_n = \mu_o$. Suppose $Z_o \in \mathcal{H}^\infty$ is the unique solution to $\min_{Z\in\mathcal{H}^\infty} \|T_1 - T_{2_i}Z\|_{\mathcal{H}^\infty} = \|T_1 - T_{2_i}Z_o\|_{\mathcal{H}^\infty}$. It then follows that $\{Z_n\}_{n=1}^\infty$ converges to Z_o in the weak* topology on $\mathcal{H}^\infty$. Moreover, the sequence $\{Z_n\}_{n=1}^\infty$ converges uniformly to Z_o on all compact frequency intervals. Given $\epsilon > 0$, let $Q_{sub} = Q_o \in \mathcal{H}_0^\infty$ be the sub-optimal solution constructed in theorem 3.1, satisfying $\mu_o \leq \|T_1 - T_2Q_{sub}\|_{\mathcal{H}^\infty} \leq \mu_o + 3\epsilon$. The sequence $\tilde{Q}_n$ constructed in theorem 3.2 converges to Q_{sub} uniformly in $\mathcal{H}^\infty$. In addition, since $|T_1(j\infty)| \leq \left\|\Gamma_{T_1T_{2_i}^*}\right\|$, it follows that $\lim_{n\to\infty} \left\|\Gamma_{T_1T_{2_{n_i}}^*}\right\| = \left\|\Gamma_{T_1T_{2_i}^*}\right\|$ and $\lim_{n\to\infty} \left\|\Gamma_{T_{1_n}T_{2_{n_i}}^*}\right\| = \left\|\Gamma_{T_1T_{2_i}^*}\right\|$.

This theorem will help solve the three $\mathcal{H}^\infty$ sensitivity problems defined earlier.

Comment 3.1 (Comparisons to Other Research)
This theorem shows how near-optimal finite-dimensional solutions can be constructed for the infinite-dimensional model matching problem. To construct the solutions, the "optimal performance" μ_o need not be known apriori. The theorem addresses the important case where T_{2_i} has an essential singularity at infinity and can only be approximated uniformly on compact frequency intervals. In so doing, it addresses the case where $T_1T_{2_i}^*$ is not continuous on the extended imaginary axis, cannot be approximated uniformly, and whose associated Hankel operator is non-compact [51]. Moreover, how well T_{1_n} and T_{2_n} need to approximate T_1 and T_2 in order to achieve a certain level of "performance", can be determined apriori without solving or partially solving the infinite-dimensional model matching problem.

The above comments are made to contrast the present results and assumptions with those in [56]. In [56], a procedure is given for constructing near-optimal solutions. In that work, however, it is assumed that the "optimal performance" is known and that $T_1T_{2_i}^*$ is continuous on the extended imaginary axis. Although the result presented in theorem 3.2 does not make these constraining assumptions, the theorem does suffer, as do the results in [56], from the fact that an inner-outer factorization of the potentially general $\mathcal{H}^\infty$ function T_2 is required. Obtaining such a factorization, in general, may be quite difficult and is the topic of on-going research [21].

It should be noted that obtaining a sequence of approximants $T_{2_{n_o}}$ which possesses the same zero structure as T_{2_o} may be difficult in general.

It should be noted that unless the optimal solution Z_o, to the infimization in proposition 3.1, is unique, convergence of the sequence Z_n to a particular optimal solution is quite difficult to show. With uniqueness, such properties are more tractable.

The spectral portion of the theorem is a new result to the literature [51]. It removes the continuity assumption which is needed in [56], [57]. It can be applied in cases where the Hankel operator $\Gamma_{T_1T_{2_i}^*}$ is non-compact; e.g. $T_1 \in R\mathcal{H}^\infty$ proper and $T_{2_i} = e^{-s}$ [47]-[51] (cf. Hartman's Theorem on compact Hankel operators [42, pp. 46]). In such cases, the non-compact Hankel operator cannot be approximated by finite-rank operators [51]. Moreover, it is not apparent how one would approximate a non-compact Hankel operator. It also applies in instances where T_{2_i} has an infinite number of poles and/or zeros; e.g. an infinite Blaschke product. To verify the inequality criterion in (4), the lemma which follows may prove helpful.

The approach usually taken in the literature to compute $\left\|\Gamma_{T_1T_{2_i}^*}\right\|$ is to solve an infinite-dimensional eigenvalue/eigenfunction problem [20, pp. 28-31], [60], [61, pp. 308]. Theorem 3.2, however, implies that in order to estimate $\left\|\Gamma_{T_1T_{2_i}^*}\right\|$, all that is needed is to solve a sequence of finite-dimensional eigenvalue/eigenvector problems [51]. This follows from (5) and the fact that computing $\left\|\Gamma_{T_{1_n}T_{2_{n_i}}^*}\right\|$ involves solving two algebraic Lyapunov equations and an associated finite-dimensional eigenvalue/eigenvector problem [28, pp. 66-74].

It should be noted that the condition $|T_1(j\infty)| \leq \left\|\Gamma_{T_1T_{2_i}^*}\right\|$, stated in theorem 3.2, in general, will not hold. Sufficient conditions are given in [48] under which this inequality holds.

The following example illustrates that even if T_{2_n} uniformly approximates T_2, one can obtain undesirable results. The example illustrates that it is necessary that $T_{2_{n_i}}$ approximate T_{2_i} in some reasonable sense.

Example 3.1 (Must Approximate Inner Part)
Let $T_1 = \frac{s+1}{s+\beta}$, where $0 < \beta < 1$ and $T_2 = \frac{e^{-s}}{s+1}$. The associated infinite-dimensional model matching problem is defined by $\mu_{opt} \stackrel{\text{def}}{=} \inf_{Q \in \mathcal{H}_0^\infty} \|T_1 - T_2 Q\|_{\mathcal{H}^\infty}$. It has been shown in [20] that $\mu_{opt} = \max\{ |T_1(j\infty)|, \left\|\Gamma_{T_1 T_{2_i}^*}\right\| \} = \left\|\Gamma_{T_1 T_{2_i}^*}\right\| > |T_1(j\infty)| = 1$ where $T_{2_i} = e^{-s}$. Now let $T_{1_n} = T_1$ and $T_{2_n} = (\frac{n}{s+n})^n \frac{1}{s+1}$. It can be shown that T_{2_n} uniformly aprroximates T_2 on the extended imaginary axis. The approximants T_{2_n} are thus terrific, based on "intuition". The associated purely finite-dimensional optimization problem is given by $\mu_n \stackrel{\text{def}}{=} \inf_{Q \in R\mathcal{H}_0^\infty} \|T_{1_n} - T_{2_n} Q\|_{\mathcal{H}^\infty}$. It can be shown that $\mu_n = \max\{ |T_{1_n}(j\infty)|, \left\|\Gamma_{T_{1_n} T_{2_{n_i}}^*}\right\| \} = |T_{1_n}(j\infty)| = 1$ where $T_{2_{n_i}} = 1$. This example thus provides a case where $\lim_{n\to\infty} \|T_{2_n} - T_2\|_{\mathcal{H}^\infty} = 0$ and yet $\lim_{n\to\infty} \mu_n \neq \mu_{opt}$. This is because $T_{2_{n_i}} = 1$, in no way, approximates $T_{2_i} = e^{-s}$. The example thus illustrates that it is important for $T_{2_{n_i}}$ to approximate T_{2_i} in some reasonable sense.

In this subsection the $\mathcal{H}^\infty$ *Model Matching Problem* was defined and near-optimal solutions were constructed for it. Sequences of appropriately formulated finite-dimensional $\mathcal{H}^\infty$ *Model Matching Problems* were also considered in an effort to avoid the complex infinite-dimensional $\mathcal{H}^\infty$ *Model Matching Problem*. It was shown that near-optimal real-rational solutions could be constructed for the infinite-dimensional $\mathcal{H}^\infty$ *Model Matching Problem* from those resulting from the finite-dimensional problems. The ideas presented in the construction will be heavily exploited in subsequent subsections.

3.4 Solution to $\mathcal{H}^\infty$ Sensitivity Problems

In this subsection, solutions are presented for the $\mathcal{H}^\infty$ *Approximate/Design, Purely Finite-Dimensional, and Loop Convergence Sensitivity Problems* . First, some definitions and assumptions are given.

In what follows, a weighted $\mathcal{H}^\infty$ sensitivity problem will be considered. The following assumption on the sensitivity weighting function W will be made.

Assumption 3.2 (Sensitivity Weighting Function)
$W \in R\mathcal{H}^\infty$ is outer.

It must be noted that the weighting function W need not roll-off as required in [56]. This is very important in weighted $\mathcal{H}^\infty$ sensitivity problems since the weighting function is typically chosen to have a shape which is the inverse of the desired sensitivity shape.

Given the previous assumption, the notion of a *sensitivity* is defined as follows.

Definition 3.7 (Sensitivity)
Let $F, G, H \in \mathcal{H}^\infty$, and $\|(F-G)H\|_{\mathcal{H}^\infty} < 1$.The *sensitivity* of the pair (F, K(G,H)) is defined as follows:

$$J(F, K(G,H)) \stackrel{\text{def}}{=} \left\| \frac{W}{1 - FK(G,H)} \right\|_{\mathcal{H}^\infty}$$

Since $K(G,H) \stackrel{\text{def}}{=} \frac{-H}{1-GH}$, it follows that $J(F, K(G,H)) = \left\| \frac{W(1-GH)}{1-(G-F)H} \right\|_{\mathcal{H}^\infty}$. From this it follows that if $\|(F-G)H\|_{\mathcal{H}^\infty} < 1$, $K(G,H)$ "internally" stabilizes F and J is well defined. This is a consequence of the *Small Gain Theorem* [17]. Moreover, if $F = G$, then this "small gain" assumption is automatically satisfied, and hence J is well defined. This is because $K(F,H)$ "internally" stabilizes F for any $H \in \mathcal{H}^\infty$, by definition.

Since $P \in \mathcal{H}^\infty$, it possesses an inner-outer factorization in $\mathcal{H}^\infty$ which will be denoted $P = P_i P_o$ [53, pp. 344]. It should be noted that the computation of this inner-outer factorization is non-trivial for general infinite-dimensional plants. This issue is currently under study. The following will be assumed about the infinite-dimensional plant P.

Assumption 3.3 (Infinite-Dimensional Plant)
(1) P_o has a finite number of distinct zeros on the extended imaginary axis.
(2) Each imaginary zero of P_o has finite algebraic multiplicity.
(3) $P_o \in \mathcal{C}_e$[4] .
(4) P_i is continuous on the imaginary axis, except possibly at infinity.

Appropriate substitution into definitions 3.1 and 3.7 results in the following expression for the *optimal performance*, μ_{opt}.

$$\mu_{opt} \stackrel{\text{def}}{=} \inf_{Q \in \mathcal{H}_0^\infty} \left\| \frac{W}{1 - PK(P,Q)} \right\|_{\mathcal{H}^\infty} = \inf_{Q \in \mathcal{H}_0^\infty} \|T_1 - T_2 Q\|_{\mathcal{H}^\infty}$$

where $T_1 \stackrel{\text{def}}{=} W$ and $T_2 \stackrel{\text{def}}{=} WP$. It is emphasized that this expression defines an infinite-dimensional optimization problem. The objective here is to show how to solve this problem through approximation rather than a direct solution.

Similarly, from definitions 3.2 and 3.7, one obtains the following expression for the *expected performance*, μ_n.

$$\mu_n \stackrel{\text{def}}{=} \inf_{Q \in R\mathcal{H}_0^\infty} \left\| \frac{W}{1 - P_n K(P_n,Q)} \right\|_{\mathcal{H}^\infty} = \inf_{Q \in R\mathcal{H}_0^\infty} \|T_{1_n} - T_{2_n} Q\|_{\mathcal{H}^\infty}$$

where $T_{1_n} \stackrel{\text{def}}{=} W$ and $T_{2_n} \stackrel{\text{def}}{=} WP_n$.

Now lets suppose that P_n is constructed as follows:

Construction 3.2 (Construction of $\{P_n\}_{n=1}^\infty$)
Let $\{P_{n_i}\}_{n=1}^\infty \subset R\mathcal{H}^\infty$ be any sequence of inner functions which approximates P_i uniformly on compact frequency intervals of the imaginary axis; i.e. $\lim_{n\to\infty} \|(P_{n_i} - P_i)X_\Omega\|_{\mathcal{H}^\infty} = 0$ for each $\Omega \in R_+$, however large. It will be said that P_{n_i} is a *compact approximant* for P_i. Let $\{P_{n_o}\}_{n=1}^\infty \subset R\mathcal{H}^\infty$ be any sequence of outer functions which approximates P_o uniformly in $\mathcal{H}^\infty$; i.e. $\lim_{n\to\infty} \|P_{n_o} - P_o\|_{\mathcal{H}^\infty} = 0$. Finally, define $P_n \stackrel{\text{def}}{=} P_{n_o} P_{n_i}$.

The above construction represents another contribution of this research in the sense that if the inner part is not appropriately approximated, then serious problems can occur. Given the construction, T_{1_n} and T_{2_n} satisfy the conditions in construction 3.1. Consequently, from theorem 3.2 it follows that given $\epsilon > 0$, there exists $\tilde{Q}_n \in R\mathcal{H}_0^\infty$ and $N \stackrel{\text{def}}{=} N(\epsilon) \in Z_+$ such that $\mu_{opt} \le \left\|T_1 - T_2 \tilde{Q}_n\right\|_{\mathcal{H}^\infty} \le \mu_{opt} + 14\epsilon$ for all $n \ge N$.

Given the above, and provided that $\left\|(P_n - P)\tilde{Q}_n\right\|_{\mathcal{H}^\infty} < 1$ for n large enough, then the finite-dimensional strictly proper compensator $\tilde{K}_n \stackrel{\text{def}}{=} K(P_n, \tilde{Q}_n) = \frac{-\tilde{Q}_n}{1 - P_n \tilde{Q}_n}$ will "internally" stabilize the infinite-dimensional plant P and hence the *actual performance*, $\tilde{\mu}_n$ is well defined. An expression for it is then obtained from definitions 3.3 and 3.7 to be

$$\tilde{\mu}_n = \left\| \frac{W}{1 - P\tilde{K}_n} \right\|_{\mathcal{H}^\infty} = \left\| \frac{W(1 - P_n \tilde{Q}_n)}{1 - (P_n - P)\tilde{Q}_n} \right\|_{\mathcal{H}^\infty}.$$

Given this, the following theorem provides a solution to the three $\mathcal{H}^\infty$ sensitivity problems defined earlier.

Theorem 3.3 (Solution to $\mathcal{H}^\infty$ Sensitivity Problems)
Suppose that $|W(j\infty)| \le \|\Gamma_{WP_i *}\|$. $\tilde{K}_n$ will "internally" stabilize P for all but a finite number of n and $\lim_{n\to\infty} \tilde{\mu}_n = \mu_{opt}$; i.e. the *actual performance* approaches the *optimal performance* as the approximants get "better". If P_{n_o} is constructed so that its zero structure on the extended imaginary axis is identical to that of P_o, then $\mu_{opt} \le \lim_{n\to\infty} \inf_{k \ge n} \mu_k$ and hence $\lim_{n\to\infty} \mu_n = \mu_{opt}$. Independent of this, it also follows that $\lim_{n\to\infty} \left\|\Gamma_{WP_{n_i} *}\right\| = \|\Gamma_{WP_i *}\|$. Suppose $Z_{opt} \in \mathcal{H}^\infty$ is the unique solution to $\min_{Z \in \mathcal{H}^\infty} \|W - P_i Z\|_{\mathcal{H}^\infty} = \|W - P_i Z_{opt}\|_{\mathcal{H}^\infty}$. Given this, the sequence $\{Z_n\}_{n=1}^\infty$ converges uniformly to Z_{opt} on all compact frequency intervals. Also, if $Q_{sub} = Q_o$ is constructed as in theorem 3.1, then $\tilde{Q}_n$ converges to Q_{sub} uniformly in $\mathcal{H}^\infty$.

[4]The case where P_o is invertible in $\mathcal{H}^\infty$ is trivial and will not be further discussed.

An example is now presented to illustrate the main ideas presented thus far.

Example 3.2 (An Irrational Outer Part)
If the plant is, for example, $P = \frac{e^{-s}}{\sqrt{s+1}}$ and the weighting function is $W = \frac{s+1}{s+\beta}$, where $0 < \beta < 1$, then one can identify $T_1 = W$ and $T_2 = WP$. The associated infinite-dimensional optimization problem is given by $\mu_{opt} \stackrel{\text{def}}{=} \inf_{Q \in \mathcal{H}_0^\infty} \|T_1 - T_2 Q\|_{\mathcal{H}^\infty}$. It has been shown in [20] that $\mu_{opt} = \max\{ |T_1(j\infty)|, \left\|\Gamma_{T_1 T_{2_i}^*}\right\| \} = \left\|\Gamma_{T_1 T_{2_i}^*}\right\| > |T_1(j\infty)| = 1$ where $T_{2_i} = e^{-s}$ and $T_{2_o} = \frac{W(s)}{\sqrt{s+1}}$. Let Z_o be any solution to the infimization defined in proposition 3.1. A near-optimal irrational solution to the associated infinite-dimensional $\mathcal{H}^\infty$ *Model Matching Problem* is then given by $Q_o = Z_o f$ where $f = T_{2_o}^{-1}(\frac{1000}{s+1000})^{\frac{1}{k}} \frac{B}{s+B}$ provided that A and B are appropriately selected (cf. proof of theorem 3.1).

Now, let P_{n_i} denote the (n,n) Pade' Approximant for $P_i = e^{-s}$ [13]. It has been shown that P_{n_i} is inner in $R\mathcal{H}^\infty$ and uniformly approximates P_i on compact frequency intervals [32]. Also let P_{n_o} denote any sequence of outer $R\mathcal{H}_0^\infty$ functions which approximate $P_o = \frac{1}{\sqrt{s+1}}$ uniformly. Next, identify $T_{2_{n_i}} = P_{n_i}$ and $T_{2_{n_o}} = WP_{n_o}$. Then, let f_n denote any $R\mathcal{H}_0^\infty$ which approximates f uniformly. Given this, a near-optimal real-rational solution is then given by $\tilde{Q}_n = Z_n f_n$ provided n is taken sufficiently large. It should be noted that the approximating sequences P_{n_o} and f_n are guaranteed to exist since P_o and f are continuous on the extended imaginary axis [48]. With the above construction, the finite-dimensional compensator $\tilde{K}_n \stackrel{\text{def}}{=} \frac{-\tilde{Q}_n}{1-P_n\tilde{Q}_n}$ will be nearly-optimal for n sufficiently large.

This completes the discussion of the three fundamental $\mathcal{H}^\infty$ *Sensitivity Problems* considered in this paper. Solutions to the analogous $\mathcal{H}^\infty$ *Mixed-Sensitivity Problems* are now presented.

3.5 Solution to $\mathcal{H}^\infty$ Mixed-Sensitivity Problems

In this subsection, solutions are provided to the $\mathcal{H}^\infty$ *Approximate/Design, Purely Finite-Dimensional, and Loop Convergence Mixed-Sensitivity Problems.* To define the problems two *weighting functions*, W_1 and W_2 will be needed.

In this subsection, new results are presented for a mixed-sensitivity criterion which penalizes the transfer function associated with the control as well as the sensitivity function. By doing so, one will be able to obtain proper compensators to start with. Obtaining a near-optimal finite-dimensional compensator which is strictly proper then becomes the central issue. It will be seen that to achieve this objective, all one need do is extend the ideas presented in subsection 3.4 for the sensitivity problem. To provide a logical flow of ideas, this subsection will follow the presentation given for the sensitivity problem in subsection 3.4.

In what follows the following "standard" assumption on the weighting functions W_1 and W_2 will be made.

Assumption 3.4 (Mixed-Sensitivity Weightings)
$W_1, W_2, W_1^{-1}, W_2^{-1} \in R\mathcal{H}^\infty$.

The above implies that W_1 and W_2 have no poles or zeros in the extended closed right half plane.

To begin the development, one starts with the following definition.

Definition 3.8 (Mixed-Sensitivity)
Let $F, G, H \in \mathcal{H}^\infty$, and $\|(F-G)H\|_{\mathcal{H}^\infty} < 1$. We then define the *mixed-sensitivity* of the pair $(F, K(G,H))$ as follows: $J(F, K(G,H)) \stackrel{\text{def}}{=} \left\| \frac{\begin{pmatrix} W_1 \\ W_2 K(G,H) \end{pmatrix}}{1 - FK(G,H)} \right\|_{\mathcal{H}^\infty}$.

By definition of $K(G,H)$, it follows that $J(F,K(G,H)) = \left\| \frac{\begin{pmatrix} W_1(1-GH) \\ W_2H \end{pmatrix}}{1-(G-F)H} \right\|_{\mathcal{H}^\infty}$. We thus see that when $\|(F-G)H\|_{\mathcal{H}^\infty} < 1$, then $K(G,H)$ "internally" stabilizes F and J is well defined. This follows from the *Small Gain Theorem* [17]. We also see that if $F = G$, then this "small gain" assumption is automatically satisfied, and hence J is well defined. This is because $K(F,H)$ "internally"stabilizes F for any $H \in \mathcal{H}^\infty$, by definition.

In what follows, the following "compact convergence" assumption on the plant approximants $\{P_n\}_{n=1}^\infty$ will be made.

Assumption 3.5 (Finite-Dimensional Plant Approximants)

$$\lim_{n\to\infty} \|(P_n - P)X_\Omega\|_{\mathcal{H}^\infty} = 0$$

for each $\Omega \in R_+$ where $\{P_n\}_{n=1}^\infty$ is uniformly bounded in $\mathcal{H}^\infty$.

It must be pointed out that this assumption has several advantages over those of the previous subsection. First, it is not necessary to approximate the inner part of the plant P_i as it was in addressing the $\mathcal{H}^\infty$ sensitivity problems. An inner-outer factorization of the infinite-dimensional plant is thus not necessary. Furthermore, the above approximation may be carried out on the basis of frequency response data. This makes the assumption very weak and practically appealing.

Given that $K(P,Q) \stackrel{\text{def}}{=} \frac{-Q}{1-PQ}$, the *optimal performance* μ_{opt} is then given by the following expression:

$$\mu_{opt} = \inf_{Q\in\mathcal{H}_0^\infty} \left\| \frac{\begin{pmatrix} W_1 \\ W_2K(P,Q) \end{pmatrix}}{1-PK(P,Q)} \right\|_{\mathcal{H}^\infty} = \inf_{Q\in\mathcal{H}_0^\infty} \left\| \begin{pmatrix} W_1(1-PQ) \\ W_2Q \end{pmatrix} \right\|_{\mathcal{H}^\infty}.$$

It should be emphasized that this expression defines an infinite-dimensional optimization problem. This problem has been addressed directly (using infinite-dimensional techniques) by [19], [25], [35], [39], [40], [41], and indirectly (using finite-dimensional techniques) by [57]. In this paper, it is shown that this infinite-dimensional problem can be avoided. For purposes of discussion suppose that $Q_o \in \mathcal{H}^\infty$ satisfies $\left\| \begin{pmatrix} W_1(1-PQ_o) \\ W_2Q_o \end{pmatrix} \right\|_{\mathcal{H}^\infty} \leq \mu_{opt} + \epsilon$. It must be emphasized that it will not be assumed that Q_o is known or that it is strictly proper.

Similarly, the *expected performance* μ_n is given by the following expression:

$$\mu_n = \inf_{Q\in R\mathcal{H}_0^\infty} \left\| \frac{\begin{pmatrix} W_1 \\ W_2K(P_n,Q) \end{pmatrix}}{1-P_nK(P_n,Q)} \right\|_{\mathcal{H}^\infty} = \inf_{Q\in R\mathcal{H}_0^\infty} \left\| \begin{pmatrix} W_1(1-P_nQ) \\ W_2Q \end{pmatrix} \right\|_{\mathcal{H}^\infty}.$$

At this point, it is instructive to make some observations. First, note that $\mu_{opt} \leq \|W_1\|_{\mathcal{H}^\infty}$ and $\mu_n \leq \|W_1\|_{\mathcal{H}^\infty}$. This shows, for example, that μ_n is a uniformly bounded sequence.

Now, let Q_n achieve the infimum or be near-optimal; i.e. $\left\| \begin{pmatrix} W_1(1-P_nQ_n) \\ W_2Q_n \end{pmatrix} \right\|_{\mathcal{H}^\infty} \leq \mu_n + \epsilon$. The γ-iteration procedures described in [28, pp. 105-131] and [18] can be used to compute Q_n. Since $W_2^{-1} \in \mathcal{H}^\infty$, it then follows that

$$\|Q_n\|_{\mathcal{H}^\infty} \leq \|W_2^{-1}\|_{\mathcal{H}^\infty}\|W_2Q_n\|_{\mathcal{H}^\infty} \leq \|W_2^{-1}\|_{\mathcal{H}^\infty}(\mu_n+\epsilon) \leq \|W_2^{-1}\|_{\mathcal{H}^\infty}(\|W_1\|_{\mathcal{H}^\infty}+\epsilon).$$

This shows that, in general, Q_n will lie in $\mathcal{H}^\infty$. More importantly, it shows that Q_n is uniformly bounded in $\mathcal{H}^\infty$. Since $Q_n \in \mathcal{H}^\infty$, it will generate a compensator $K_n \stackrel{\text{def}}{=} K(P_n,Q_n) = \frac{-Q_n}{1-P_nQ_n}$ which generally is not strictly proper. Also, K_n will not necessarily stabilize the infinite-dimensional plant P [43]. In an attempt

to remedy these shortcomings, a "roll-off operator" $r : Q_n \to \tilde{Q}_n \in R\mathcal{H}_0^\infty$ is defined. The exact form of r, as will be seen, is obtained in a manner analogous to that used for the weighted $\mathcal{H}^\infty$ sensitivity problem. To construct r, all that is required is the roll-off function $h_m \stackrel{\text{def}}{=} (\frac{1}{s+1})^{\frac{1}{m}}$. More specifically, this function can be used to modify Q_o and Q_n to obtain "near-optimal Q's" which are strictly proper and yield near-optimal strictly proper compensators K_{sub} and $\tilde{K}_n$, respectively.

Given the above, and provided that $\left\|(P_n - P)\tilde{Q}_n\right\|_{\mathcal{H}^\infty} < 1$ for n large enough, then the strictly proper finite-dimensional compensator $\tilde{K}_n \stackrel{\text{def}}{=} K(P_n, \tilde{Q}_n) = \frac{-\tilde{Q}_n}{1-P_n\tilde{Q}_n}$ will "internally" stabilize the infinite-dimensional plant P and hence the *actual performance*, $\tilde{\mu}_n$ is well defined and given by

$$\tilde{\mu}_n = \left\| \frac{\begin{pmatrix} W_1 \\ W_2 K(P_n, \tilde{Q}_n) \end{pmatrix}}{1 - PK(P_n, \tilde{Q}_n)} \right\|_{\mathcal{H}^\infty} = \left\| \frac{\begin{pmatrix} W_1(1 - P_n\tilde{Q}_n) \\ W_2\tilde{Q}_n \end{pmatrix}}{1 - (P_n - P)\tilde{Q}_n} \right\|_{\mathcal{H}^\infty}.$$

Given this, the following theorem provides a solution to the *Approximate/Design Mixed-Sensitivity Problem*, the *Purely Finite-Dimensional Mixed-Sensitivity Problem*, and the *Loop Convergence Mixed-Sensitivity Problem*.

Theorem 3.4 (Solution to $\mathcal{H}^\infty$ Mixed-Sensitivity Problems)
There exists a single function $f \in \mathcal{H}_0^\infty \cap C_e$ (independent of Q_o) such that if $Q_{sub} = Q_o f$, then

$$\left\| \begin{pmatrix} W_1(1 - PQ_{sub}) \\ W_2 Q_{sub} \end{pmatrix} \right\|_{\mathcal{H}^\infty} \leq \mu_{opt} + \epsilon$$

and $K_{sub} \stackrel{\text{def}}{=} K(P, Q_{sub})$ is a near-optimal irrational compensator for the infinite-dimensional plant P. Also, if $f_n \in R\mathcal{H}_0^\infty$ (not to be confused with f_m in definition 3.5) is any sequence which uniformly approximates f, and $\tilde{Q}_n = Q_n f_n$, then

$$\lim_{n\to\infty} \left\|(P_n - P)\tilde{Q}_n\right\|_{\mathcal{H}^\infty} = 0.$$

Consequently, $\tilde{K}_n \stackrel{\text{def}}{=} K(P_n, \tilde{Q}_n)$ will "internally" stabilize P for all but a finite number of n. Moreover, the *actual performance* will approach the *optimal performance*; i.e. $\lim_{n\to\infty} \tilde{\mu}_n = \mu_{opt}$. Also, the *expected performance* will approach the *optimal performance*; i.e. $\lim_{n\to\infty} \mu_n = \mu_{opt}$. Finally, if Q_o is the unique optimal solution of the infinite-dimensional problem , then all of the actual sub-optimal loop shapes associated with $(P, \tilde{K}_n)$ converge uniformly to the sub-optimal infinite-dimensional loop shapes associated with (P, K_{sub}).

Given the previous theorem, some comments are in order. First, it is important to note that no infinite-dimensional inner-outer factorization is required as for the sensitivity problem.

Since [57] addresses similar *Approximate/Design* issues, it is important to point out the distinguishing features of this work relative to [57]. It is important to emphasize, for example, that

(1) It suffices for P_n to approximate P uniformly, only on compact frequency intervals.

(2) Unlike in [57], no infinite-dimensional spectral factorization is required in this work.

(3) Unlike in [57], the optimal performance need not be known apriori in order to construct the compensators presented in this work. A consequence of this is that given a near-optimality tolerance, how well the plant must be approximated to achieve the desired tolerance can be specified apriori.

(4) Unlike in [57], the near-optimal compensators obtained in this work are finite-dimensional.

This completes the discussion of the three fundamental $\mathcal{H}^\infty$ *Mixed-Sensitivity Problems* considered in this section.

4 A Methodology for Systems with Hard Nonlinearities

While a Research Associate at Eglin Air Force Base, the principal investigator worked on the problem of enhancing performance for systems with multiple saturating actuators [52]. The methods developed are based on the work of [9] and the more recent work of [7].

To present the main ideas, lets suppose that P is a linear time invariant (LTI), multiple-input multiple-output (MIMO), stable system. The unstable case is discussed in [9]. Also suppose that K is a LTI MIMO compensator which has been designed so that the closed loop system in figure 4 has desirable properties (e.g. robust performance, etc.).

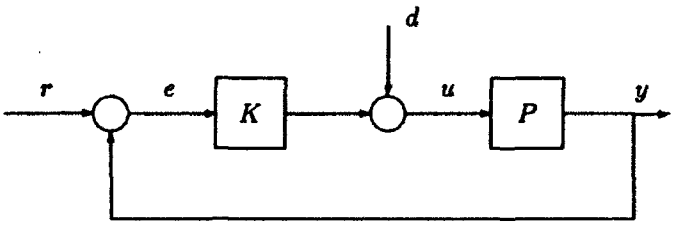


Figure 4: Visualization of Nominal Closed Loop System

Now consider the system in figure 6. In this figure,

$$y(t) = (P_{act}u)(t) = p(t) * sat(u(t))$$

where $p(t)$ is the impulse response matrix of P, $*$ denotes convolution, and $sat(u(t))$ denotes a saturation operator; i.e. multiple saturating actuators; one on each input channel. K_{mod} represents the modified compensator and is defined by the following relation:

$$u(t) = d(t) + (K_{mod}e)(t) = d(t) + k(t) * \lambda(t)e(t)$$

where $k(t)$ is the impulse response matrix of K and $\lambda(t) = \lambda(e(t), x_k(t)) \in [0, 1]$ is a time-varying scalar gain which depends on the error signal $e(t)$ and on the compensator state $x_k(t)$.

In [9], a procedure for computing λ is given. The idea behind the procedure is simple. If the system saturates reduce the gain λ. Since λ is a scalar, such gain reduction preserves the relative coordination of the controls (i.e. the directionality properties of the original design). If the system is not saturated, it should be allowed to operate as intended with $\lambda = 1$. The procedure requires a state space representation of the compensator and guarantees bounded-input bounded-output (BIBO) stability. The procedure, however, requires much computation [52] (see Appendix 8 for abstract). Figure 5 shows how the error governor method of [9] works for $P = \frac{1}{s+1}$, $K = 1/s$, r = 10, and saturation limits of +9 and -9. The linear response is very good, the saturated response causes the integrator to wind up, and the effect of λ is to prevent this.

While at Eglin Air Force Base, the principal investigator noticed that [7] may be used to significantly reduce the required computations. This will be addresed in [52].

5 Significance of Work

The methods which have been developed in this research will systematize the design of full envelope controllers for *flexible (high order) missile systems with multiple hard nonlinearities*. Systematizing the design process should significantly reduce design time and lower system development costs. Preliminary results show clearly that the methods will increase system performance by permitting operation over the full envelope and in some cases extending this envelope. Finally, the methods developed will provide useful information to guide designers during the initial stages of system development, design, and conceptualization. Given this, the contributions of this research to missile control, and more globally to aerospace system development, should be significant.

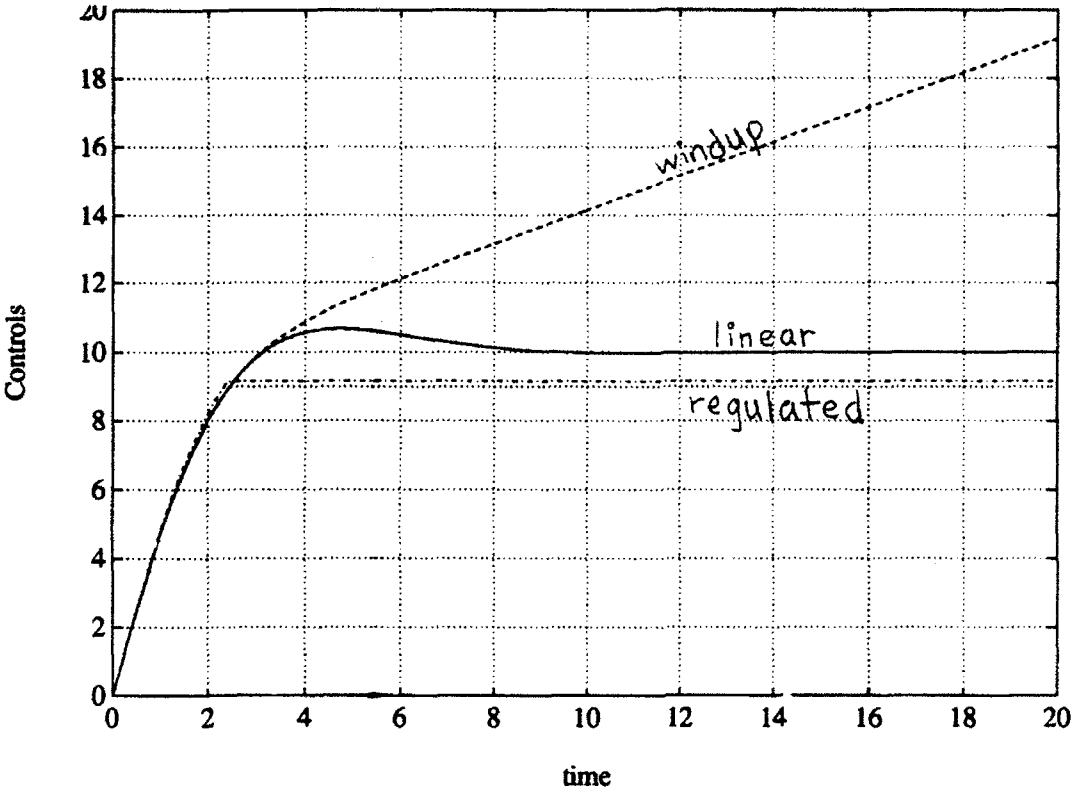


Figure 5: r = 10; satlev = 9

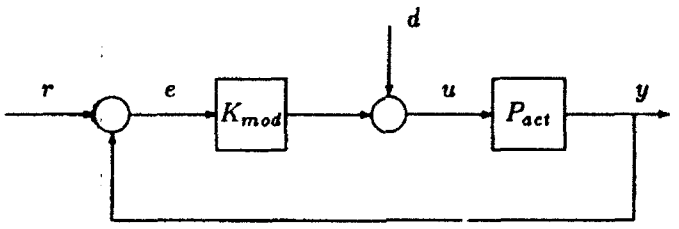


Figure 6: Visualization of Modified Compensator in System with Multiple Saturating Actuators

6 Summary & Directions for Future Research

During the course of this research, methods have been developed to facilitate the desgin of full envelope autopilots for flexible (high order) missiles with multiple hard nonlinearities.

Future directions for research include applying the methods developed here to EMRAAT [1], [52]. How to model EMRAAT when there are pronounced flexible modes is an issue. This brings up the question of *performance based model reduction* [46]. Applying the algorithms of [7], [8] to EMRAAT to accomodate saturations, rate limits, and other hard nonlinearities is a key direction to be pursued. Using these algorithms to design a robust α-limiter is another key direction. Since autopilots are gain scheduled [11]-[12], it is necessary to robustify the individual linear designs. This can be done using, for example, the new $\mathcal{L}^1/\mu$-synthesis design methodology [4]-[6]. The final autopilot should be tested with a full six (6) degree-of-freedom simulation which includes structural modes and all nonlinear effects.

Since the methods developed are general, they may be applied to a very high order structural control problem. An example of such a problem is controlling the Air Forces' Space Integrated Control Experiment (SPICE) located on Kirtland Air Force Base [2]. This laser beam expander is modelled as a 260th order

system. The idea is to perform fast slewing maneuvers without exciting the structural modes.

7 Bibliography

References

[1] I.A. Hirsch, M.A. Langehough, J.A. Bossi, et al. , "Advanced Robust Autopilot," Air Force Armament Laboratory, Eglin AFB, Florida, AFATL-TR-89-64, November 1989.

[2] "Proposal for Space Integrated Control Experiment (SPICE)", Prepared by Lockheed Missiles & Space Company, Inc. (in collaboration with Honeywell Satellite Systems) in response to Air Force RFP F29601-88-R-0025, LMSC-F278345, Sunnyvale, CA 94088-3504, August 1988.

[3] AIAA, *Aerospace America*, Focus '92, Computational Fluid Dynamics, Parts I-II, January-February, 1992.

[4] M.A. Dahleh, "Design of Multivariable Feedback Controliers: ℓ^1-Optimal Systems," PhD Thesis, Rice University, November 1986.

[5] M.A. Dahleh and J.B. Pearson, Jr., "$\mathcal{L}^1$-Optimal Compensators for Continuous-Time Systems," *IEEE Trans AC*, AC-32, No. 10, October 1987, pp. 889-895.

[6] M.A. Dahleh, M.H. Khammash, J.B. Pearson, Jr., I.D. Bobillo "$\mathcal{L}^1$ Robust Control," Workshop No. 7, June 23, 1992, American Control Conference, Chicago, IL.

[7] E.G. Gilbert, and K.T. Tan, "Linear Systems with State and Control Constraints: The Theory and Application of Maximal Output Admissible Sets," *IEEE Trans Automatic Control*, Vol AC-36, No. 9, September 1991, pp. 1008–1020.

[8] K.T. Tan and E.G. Gilbert, "Linear Systems with State and Control Constraints: The Theory and Application of Maximal Output Admissible Sets," submitted for publication in *IEEE Trans Automatic Control*, received October 1992.

[9] P. Kapasouris, "Design for Performance Enhancement in Feedback Control Systems with Multiple Saturating Nonlinearities," LIDS MIT, PhD Thesis, LIDS-TH-1757, March 1988.

[10] J.S. Shamma, "Analysis and Design of Gain Scheduled Control Systems," LIDS MIT, LIDS-TH-1770, PhD, May 1988.

[11] J.S. Shamma and J.R. Cloutier, "A Linear Parameter Varying Approach to Gain Scheduled Missile Autopilot Design," *Proceedings of the 1992 American Control Conference*, Chicago, IL, June, 1992.

[12] G.A. Baker, Jr., *Essentials of Pade Approximants*, Academic Press, NY 1975.

[13] J.B. Conway, *A Course in Functional Analysis*, Springer-Verlag, NY 1985.

[14] R.F. Curtain and K. Glover, "Robust stabilization of infinite-dimensional systems by finite dimensional controllers," *Systems and Control Letters 7*, 1986, 41-47.

[15] R.F. Curtain, "A synthesis of time and frequency domain methods for the control of infinite-dimensional systems: a system theoretic approach," to appear in the series of SIAM Frontiers in Applied Mathematics, 1989.

[16] C.A. Desoer and M. Vidyasagar, *Feedback Systems: Input-Output Properties*, Academic Press, Inc, NY, 1975.

[17] J. C. Doyle, K. Glover, P.P. Khargonekar and B.A. Francis, "State-Space Solutions to Standard $\mathcal{H}^2$ and $\mathcal{H}^\infty$ Control Problems," *IEEE Trans AC*, Vol 34, *No 8*, August 1989.

[18] F. Fagnani and S.K. Mitter, "An operator-theoretic approach to the mixed-sensitivity minimization problem," LIDS-P-1804, MIT, August 1988.

[19] D.S. Flamm, "Control of delay systems for minimax sensitivity," PhD Thesis, LIDS-TH-1560, MIT, June 1986, Cambridge, MA.

[20] D.S. Flamm and K. Klipec, "Numerical computation of inner factors for distributed parameter systems," preprint.

[21] D.S. Flamm and S.K. Mitter, "$\mathcal{H}^\infty$ sensitivity minimization for delay systems," *Systems and Control Letters 9*, 1987, 17-24.

[22] D.S. Flamm and S.K. Mitter, "Approximation of ideal compensators for delay systems," *Linear Circuits, Systems and Signal Processing: Theory and Application* C.I. Byrnes, C.F. Martin and R.E. Saeks (editors), Elsevier Science Publishers B.V., 1988, 517-524.

[23] D.D. Flamm, "A Model of a Damped Flexible Beam," *ISS Report No. 54*, June 14, 1990; Revised July 26, 1990.

[24] D.S. Flamm and H. Yang, "On $\mathcal{H}^\infty$-optimal mixed-sensitivity minimization for general distributed plants," Submitted for publication, April 1990.

[25] C. Foias, A. Tannenbaum, and G. Zames, "Weighted sensitivity minimization for delay systems," *IEEE Trans AC*, AC-31, 1986, 763-766.

[26] C. Foias, A. Tannenbaum, and G. Zames, "Sensitivity minimization for arbitrary SISO distributed plants," *Systems and Control Letters 8*, 1987, 189-195

[27] B.A. Francis, *A Course in H_∞ Control Theory*, Springer-Verlag, 1987.

[28] J.B. Garnett, *Bounded analytic functions*, Academic Press, 1981.

[29] K. Glover, R.F. Curtain, and J.R. Partington, "Realization and approximation of linear infinite-dimensional systems with error bounds," *SIAMJ. Control and Optimization*, 26, July 1988,863-898.

[30] G. Gu, P.P. Khargonekar, and E.B. Lee, "Approximation of infinite-dimensional systems," *IEEE Trans Automatic Control*, Vol AC-34, June 1989.

[31] K. Glover, J. Lam, and J.R. Partington, "Rational approximation of a class of infinite-dimensional systems," to appear in Math Contr Sig Syst.

[32] K. Hoffman, *Banach spaces of analytic functions*, Prentice-Hall, Inc, Engelwood Cliffs, NJ, 1962.

[33] E.W. Kamen, P.P. Khargonekar, and A. Tannenbaum, "Stabilization of time delay systems with finite-dimensional compensators," *IEEE Trans AC*, Vol AC-30, Jan., 1985, 75-78.

[34] K. Lenz, H. Ozbay, A. Tannenbaum, J. Turi, and B. Morton, "Robust control design for a flexible beam using a distributed parameter H^∞ method," *CDC*, Tampa, Florida, December 1989.

[35] D.G. Luenberger, *Optimization by Vector Space Methods*, John Wiley & Sons, Inc., 1968.

[36] J.R. Munkres, *Topology: A first course*, Prentice Hall, 1975.

[37] Z. Nehari, On bounded linear forms, Ann. of Math, Vol 65, 1957, 153-162.

[38] H. Ozbay, "H^∞ Control of Distributed Systems: A Skew Toeplitz Approach," Phd Thesis, University of Minnesota, June 1989.

[39] H. Ozbay, M.C. Smith, and A. Tannenbaum, "Control design for unstable distributed plants," submitted to *ACC* 1990.

[40] H. Ozbay and A. Tannenbaum, "On the structure of suboptimal H^∞ controllers for distributed plants," submitted for publication, 1989.

[41] J.R. Partington, *An Introduction to Hankel Operators*, London Mathematical Society Student Texts 13, Cambridge University Press, 1988.

[42] A.A. Rodriguez and M.A. Dahleh, "A Finite-Dimensional Approach to Infinite-Dimensional Multi-block $\mathcal{H}^\infty$ Optimal Control Problems," Technical Report No. P-1950, LIDS, MIT, Cambridge MA, February 1990.

[43] A.A. Rodriguez, "Control of Infinite-Dimensional Systems Using Finite-Dimensional Techniques," PhD Thesis, LIDS, MIT, Cambridge MA, August 1990.

[44] A.A. Rodriguez and M.A. Dahleh, "Wiener-Hopf Control of Stable Infinite-Dimensional Systems," *Proceedings of the American Control Conference*, Boston, MA, June 1991.

[45] S.H. Mahloch and A.A. Rodriguez, "Approximation of Infinite-Dimensional Systems for Control Design," Arizona State University, MS thesis, in preparation.

[46] A.A. Rodriguez, M.A. Dahleh, and S.H. Mahloch, "Hankel Norm Computation for Infinite-Dimensional $\mathcal{H}^\infty$ Model Matching Problems," *Proceedings of the American Control Conference*, Chicago, IL, June 1992.

[47] A.A. Rodriguez and M.A. Dahleh, "$\mathcal{H}^\infty$ Control of Stable Infinite-Dimensional Systems using Finite-Dimensional Techniques," submitted for publication in *IEEE Transactions on Automatic Control*, July, 1992.

[48] A.A. Rodriguez and M.A. Dahleh, "Weighted $\mathcal{H}^\infty$ Optimization for Stable Infinite-Dimensional Systems using Finite-Dimensional Techniques," *Proceedings of the 29th IEEE Conference on Decision and Control*, Honolulu, Hawaii, December 1990.

[49] A.A. Rodriguez, "Design of $\mathcal{H}^\infty$ Optimal Finite-Dimensional Controllers for Unstable Infinite-Dimensional Systems," Submitted for publication to *AUTOMATICA*, August, 1992.

[50] A.A. Rodriguez and M.A. Dahleh, "On the Computation of Induced Norms for Non-Compact Hankel Operators Arising From for Distributed Control Problems," To appear in the *Systems & Control Letters*, 1992.

[51] A.A. Rodriguez and J.R. Cloutier, "Control of a Bank-to-Turn-Missile with Saturating Actuators", Submitted for publication in the *Proceedings of the 1993 American Control Conference*.

[52] W. Rudin. *Real and Complex Analysis*, McGraw-Hill, Inc., 1987.

[53] M. Smith, "Well-posedness of $\mathcal{H}^\infty$ optimal control problems," *Siam J Control and Optimization*, Vol 28, March 1990, 342-358.

[54] M. Vidyasagar, *Control Systems Synthesis: A Factorization Approach*, MIT press, Cambridge, MA, 1985.

[55] E.N. Wu and E.B. Lee, "Feedback minimax synthesis for distributed systems," *CDC*, Austin, Texas, December 1988, 492-496.

[56] E.N. Wu, "Distributed Parameter Control Systems Synthesis: An Iterative Approach," *Proceedings of the American Control Conference*, San Diego, CA, June 1990, pp 1589-1595.

[57] D.C. Youla, H.A. Jabr, and J.J. Bongiorno, "Modern Wiener-Hopf design of optimal controllers-part 2: The multivariable case," *IEEE Trans AC*, Vol AC-21, June 1976.

[58] G. Zames, "Feedback and optimal sensitivity: model reference transformations, multiplicative seminorms, and approximate inverses," *IEEE Trans AC*, Vol AC-26, No 2, April 1981, 301-319.

[59] G. Zames and S.K. Mitter, "A note on essential spectra and norms of mixed Hankel-Toeplitz operators," *Systems and Control Letters 10*, 1988, 159-165.

[60] K. Zhou and P.P. Khargonekar, "On the weighted sensitivity minimization for delay systems," Systems and Control Letters, 1987, 307-312.

DETERMINATION OF THE OPERATIONAL CHARACTERISTICS OF A PHASE-DOPPLER DROPLET ANALYZER AND APPLICATION TO A RAMJET FUEL-INJECTION RESEARCH TUNNEL

Larry A. Roe
Assistant Professor
Mechanical Engineering Department

Virginia Polytechnic Institute and State University
Blacksburg, VA 24061-0238

Final Report for:
Summer Research Program
Advanced Propulsion Division
Aeropropulsion and Power Directorate
Wright Laboratories

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

September 1992

DETERMINATION OF THE OPERATIONAL CHARACTERISTICS OF A PHASE-DOPPLER DROPLET ANALYZER AND APPLICATION TO A RAMJET FUEL-INJECTION RESEARCH TUNNEL

Larry A. Roe
Assistant Professor
Mechanical Engineering Department
Virginia Polytechnic Institute and State University

Abstract

A single-component, phase-Doppler particle analyzer was assembled, validated and characterized in bench-top testing, and applied to an experimental fuel injection tunnel. The influences of instrument settings, including laser beam intersection angle, diffraction grating speed, photomultiplier voltage, and filter settings, were evaluated. It was concluded that the system produces reliable droplet size information over the range of interest and good results for average velocity. Indicated results for RMS velocity were not felt to be generally reliable under the conditions tested. Data sufficient to produce three-dimensional contour plots of the spray pattern produced by normal injection of water into a crossflowing air stream were acquired for air Mach numbers of 0.2 and 0.4. Additional data were obtained for Mach 0.3. A preliminary analysis of the data indicates that droplet diameters are generally larger on the upper periphery of the plume, with large densities of smaller droplets in the central core of the spray. Small droplets, of low number density, were found near the tunnel floor. Full three-dimensional mapping of the acquired data is recommended.

DETERMINATION OF THE OPERATIONAL CHARACTERISTICS OF A PHASE-DOPPLER DROPLET ANALYZER AND APPLICATION TO A RAMJET FUEL-INJECTION RESEARCH TUNNEL

Larry A. Roe

INTRODUCTION

The injection of liquids into high-speed crossflows is the primary process dominating combustion characteristics of both subsonic combustion and supersonic combustion ramjet engines operating on liquid fuel. The development of the fuel spray from the initial injection conditions occurs through breakup of the injected fuel stream into droplets. The sizes and trajectories of these droplets determine their subsequent vaporization, which strongly determines the combustion characteristics of the engine system. Thrust, range, fuel consumption, emissions, and exhaust signature are all directly related to these combustion characteristics.

Computer modeling of liquid-fueled combustors requires detailed information about the fuel spray. This information has not generally been available, as jet breakup models have not been demonstrated to be reliable and the experimental data are sparse. The solution to this problem requires detailed, spatially resolved, experimental measurements of droplet size distributions, velocities, and flow rates throughout the spray region. The objective of the research program was to provide the Advanced Propulsion Division the capability for obtaining these measurements.

A commercially manufactured system for obtaining droplet size and velocity had previously been purchased but had not been used by division personnel. This instrument, a phase-Doppler particle analyzer (PDPA) manufactured by Aerometrics, uses laser diagnostic techniques similar to those utilized in laser Doppler anemometers. Such systems are not turnkey installations, and their proper utilization requires a comprehensive understanding of the operating principles and a systematic characterization of the sensitivity of the apparent results to instrument operating parameters. In essence, different results can be obtained simply by changing instrument

settings. This report documents the sensitivity of results to instrument parameters, identifies operational procedures necessary to acquire valid data, and describes results obtained for crossflow injection of water into an airstream.

DESCRIPTION AND CHARACTERIZATION OF PHASE-DOPPLER PARTICLE ANALYZER

The PDPA system as manufactured by Aerometrics is based on a development by Bachalo and Houser (1). A more thorough description of the application of this technique is provided by Bachalo et al. (2). A complete description of the operational principles is well beyond the scope of this report, but a brief summary is required for an understanding of the instrument characterization sections of this document.

A PDPA is essentially a single-component laser Doppler anemometer (LDA) with multiple photodetectors and additional signal processing capability. As an LDA system, it is of the standard dual-beam type. Two laser beams intersect at a small angle in the region where measurements are to be obtained. The crossing of these two beams defines a probe volume; droplets passing through this zone scatter light simultaneously from the two beams. When this scattered light reaches the receiving optics (lens or photodetector), it forms an interference pattern (Figure 1), which moves in space due to the droplet motion. The temporal frequency at which the interference fringes sweep across the surface of the detector is related to the transmitting optics (laser wavelength and beam intersection angle) and the droplet velocity component in the plane of the intersecting beams. Since the optical configuration is known, the velocity component can be determined.

Additional information is required for droplet sizing. This is provided by additional photodetectors. Essentially, these photodetectors image different regions of the interference pattern as it sweeps across the receiving optics. The photodetectors all observe the same temporal frequency (the Doppler frequency related to velocity) but observe different spatial frequencies since the interference pattern fringes are not parallel. This

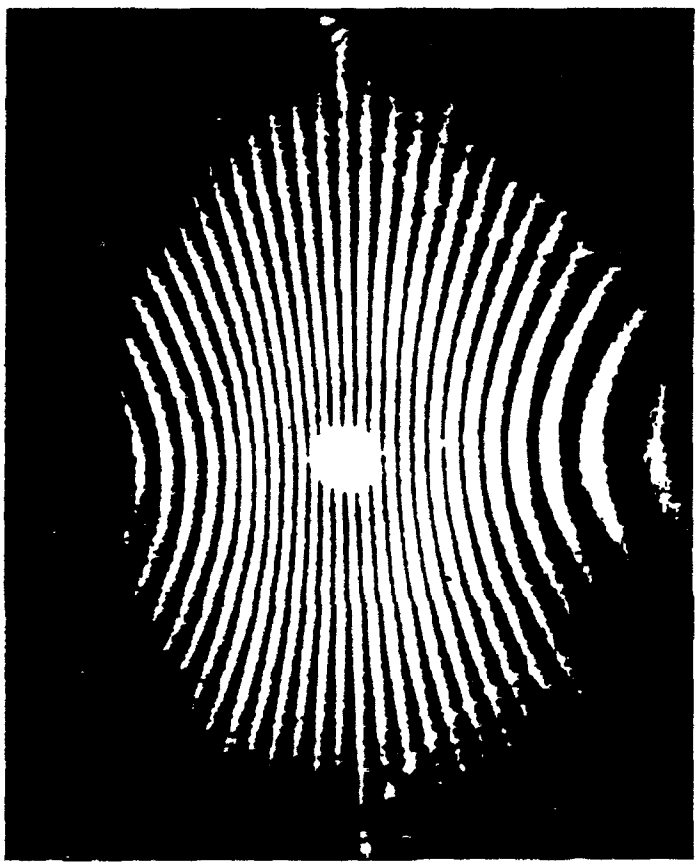


Figure 1. Interference pattern created by droplet in probe volume.

translates to a phase difference between detectors. This phase difference is related to the curvature of the fringe pattern, which is related to droplet diameter (and a long list of other parameters, which are generally known). With sophisticated signal processing and data analysis, the droplet diameter (assumed spherical) can be determined.

The primary instrument operating parameters which can be expected to influence the results are: the incident beam intersection angle, frequency shifting of the incident beams by a rotating diffraction grating, receiving optics alignment, photomultiplier (PMT) voltage, and filtering of the output signal. The sensitivity of the results to these settings was evaluated using a small-diameter spray from a TSI atomizer, a uniform droplet generator as developed by Switzer (3), and a glass fiber mounted on a rotating wheel.

Photomultiplier voltage can be expected to affect instrument operation and delivered results. Large droplets scatter more light than small droplets and therefore provide a higher signal to the photodetectors. PMT output is related to both the incident light intensity and the applied tube voltage. As the voltage is increased, the PMT will become sensitive to lower light levels (smaller droplets). The light scattered from significantly larger droplets will then be of such a level as to saturate the photomultipliers, causing a loss of signal. It is to be expected, then, that as the PMT voltage is increased, the observed size distribution will shift toward smaller and smaller droplets. The question then becomes one of selecting some appropriate operating point.

In the uniform droplet generator tests, PMT voltage did not significantly affect the results, as would be expected. Greater velocity fluctuations were observed at the highest voltage settings. For the atomizer test series, however, trends were seen. In a series of tests increasing the PMT voltage from 400V to 650 V, the indicated mean diameter was seen to increase from 1.58 microns to 2.31 microns. The velocity increased from 4.85 to 6.48 m/s, and the number of attempts required to obtain 1000 good data points increased form 1632 to 10452. The diameter trend was unexpected, but

it was considered that the range of droplet sizes provided by the atomizer was at the very lowest end of the applicable range of the PDPA instrument. This parameter was reevaluated in the injection stand before taking actual data.

Receiving optics alignment was also found to affect the indicated results. The imaged probe volume was moved to various positions on or near the slit which serves as a spatial filter. It was observed that the image had to be visibly quite misaligned to cause a significant problem. For these cases, the data rejection rate was also quite high (greater than 90 percent). High rejection rates are judged to be a good indicator of improper instrument operation in general.

Processor filter settings did not affect the indicated results as long as the velocity distribution was within the values indicated on the data acquisition screen. However, the manufacturer's software also provides a history of rejected data points, including observed data which falls outside of the filter settings. This history was observed to be unreliable and should not be relied upon to set either the velocity or size ranges.

Incident beam spacing and frequency shifting are best discussed as coupled parameters since both directly affect the structure of the fringe pattern. Reduced beam spacing causes the beams to converge at a narrower angle. The system is then sensitive to larger particles, but the fringe spacing is increased and there are fewer fringes in the measuring region. Fewer fringes lowers the velocity resolution. The effective number of fringes can be increased by frequency shifting, as this essentially creates a moving fringe pattern. Unsteadiness in the operation of the grating motor and uncertainty about its rotational speed add inaccuracies to velocity determination. Trends were observed across the available range of combinations of beam spacing and frequency shifting. The narrowest beam spacing gives velocity values which are approximately 5 percent low. Any frequency shifting widens the measured velocity distribution, increasing the indicated velocity fluctuation unrealistically. The velocity RMS data do not appear reliable in general. Overall system accuracy is estimated to be

approximately 7 percent for mean velocity and 5 percent on diameter for droplets larger than 2 microns. The observation of droplets smaller than about 2 microns, in a wide-distribution spray, is not recommended.

APPLICATION OF THE PDPA TO A RESEARCH FUEL-INJECTION STAND

After some operational experience had been acquired and the general characteristics of the instrument documented, the PDPA was installed on a three-axis optical table to evaluate a simulated fuel spray in the fuel injection stand (Figure 2) in test cell 18. This required the design and fabrication of some additional mounting hardware to maintain the necessary optical alignments. Initial testing with water injection at an air Mach number of 0.4 was conducted to evaluate the general spray characteristics and the sensitivity of the results to instrument configuration. After optical alignment, measurements were taken at a fixed location and various PMT settings. Photomultiplier voltages below 320 gave very truncated distributions, with essentially no sensitivity to droplets smaller than about 15 microns. Voltages above 380 caused an excessive sensitivity to droplets smaller than the 2 to 3 micron lower limit for which full measurement confidence has been established. Since a major portion of the spray mass is associated with large droplets and the breakup of these large droplets is the primary concern of the injection studies, an operating PMT voltage of 350 was established for the first experiments, although it was realized that some adjustment to this may be required for local conditions.

Initial spray mapping was done for normal injection of water (flow rate 1.68 lbm/min) through a 0.030-inch diameter orifice, into a Mach 0.4 air stream. Incident beam spacing was set at the intermediate of the three available positions. Frequency shifting was not utilized. Approximately 4100 droplets were sampled at each location in a spatial matrix that extended from 1 inch to 7 inches downstream of the injection point. The number of points in the matrix was varied with the size of the spray cone as it expanded in the downstream direction. Full cross-sections of the spray pattern were obtained

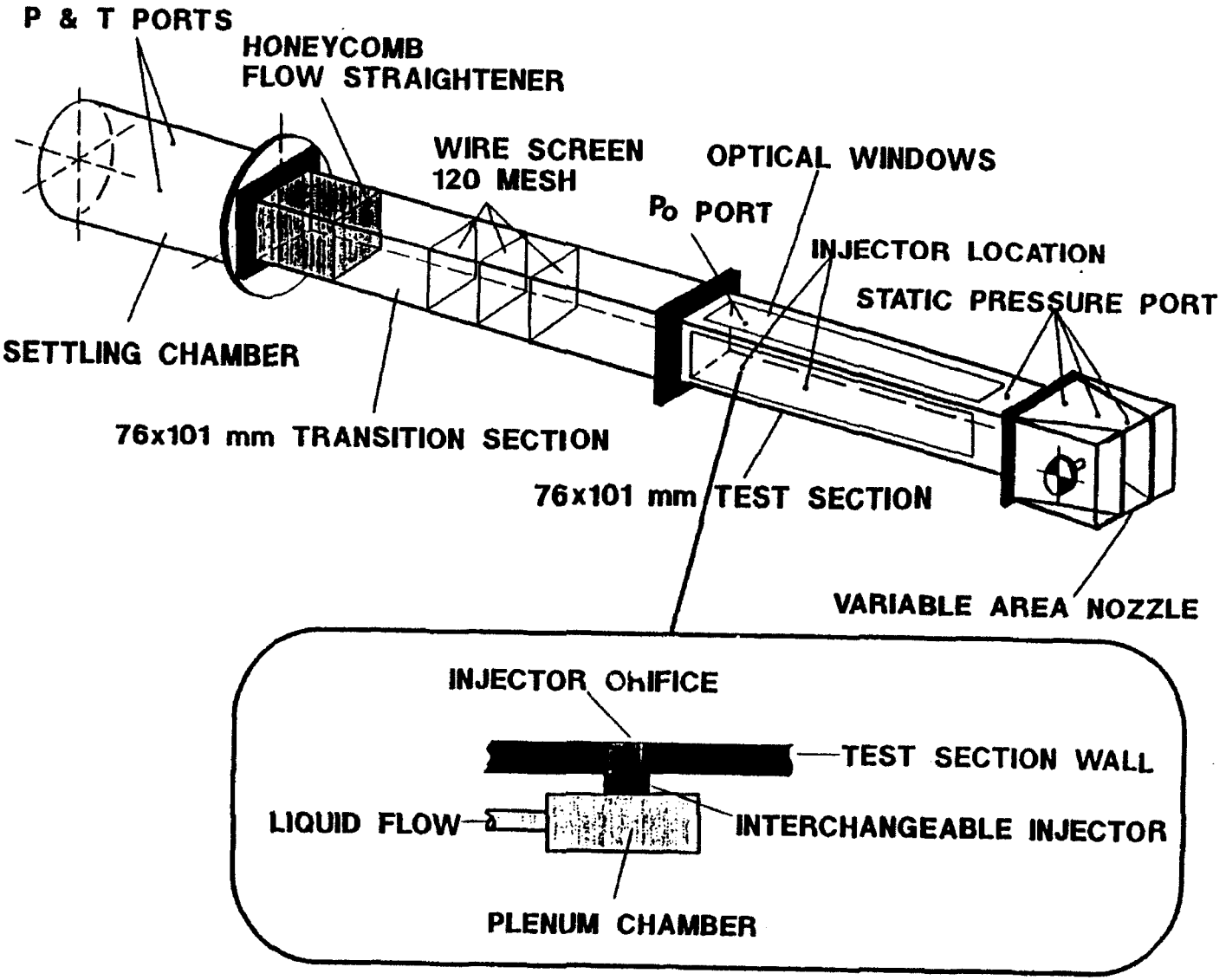


Figure 2. Research fuel injection stand, showing test section with optical windows.

at 1.00, 2.00, 3.00, and 5.00 inches downstream of the injection point, and vertical centerline profiles at 4.00 and 7.00 inches.

Subsequent testing was conducted at an air velocity of Mach 0.2 and a water flow rate of 0.839 lbm/min. The average droplet diameter was much larger for this test series, requiring the minimum incident beam spacing, and a lower PMT potential of 320V. Frequency shifting of 20 m/s was applied to the incident beams to expand the available range of velocities. Full spray cross-sections were obtained 2.00 and 5.00 inches downstream of the injector and vertical centerline profiles were measured at 0.50, 1.00, 3.00, 4.00, and 6.00 inches.

Vertical centerline profiles were obtained at distances of 1.00, 2.00, 3.00, 4.00, 5.00, and 6.00 inches for an air Mach number of 0.3 and injectant flow rate of 1.26 lbm/min. The intermediate beam spacing was utilized, frequency shifting was not employed, and the PMT voltage was 320V.

The analysis of these data will require considerable time. Specific data concerning the acceleration of (nominal) 180-micron-diameter droplets in the Mach 0.2 airstream were provided to Air Force researchers for comparison to data previously obtained through particle image velocimetry. Preliminary observations of potential interest include an observation that the largest droplets were generally on the upper part of the spray and at the outer portion of the pattern. The central section of the spray contained smaller droplets at a much higher number density. Typical results from a preliminary analysis of the data are presented here.

Figure 3 shows droplet diameter and velocity statistics for the Mach 0.4 flow at a point along the centerline of the duct axis (Y = 0.0), 3.00 inches downstream of the injector (X = 3.00), and 1.4 inches above the tunnel floor (Z = 1.4). This position is in a fairly sparsely populated part of the spray near the upper limit. The diameters are seen to range from about 3 microns to 100 microns, with an arithmetic mean of 37.0 microns. Droplets smaller than 3 microns are likely to be present, but are too small to be visible to the receiving optics at the PMT voltage used. Figure 4 shows the velocity-

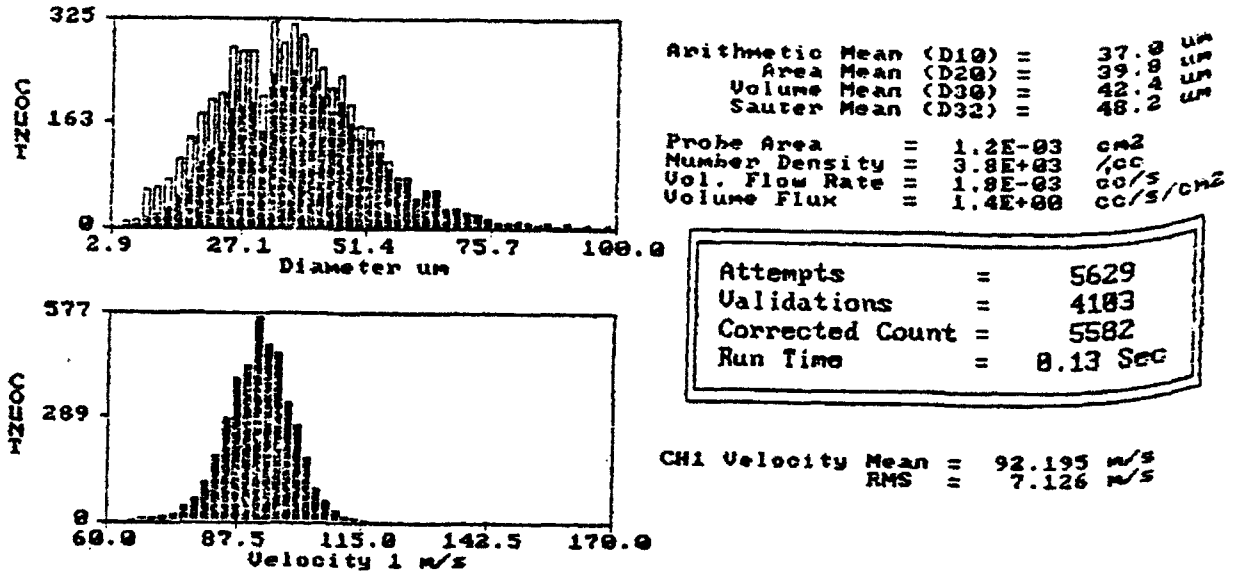


Figure 3. Typical distributions of droplet size and velocity at a fixed point in the spray. (X = 3.00, Y = 0.0, Z = 1.40).

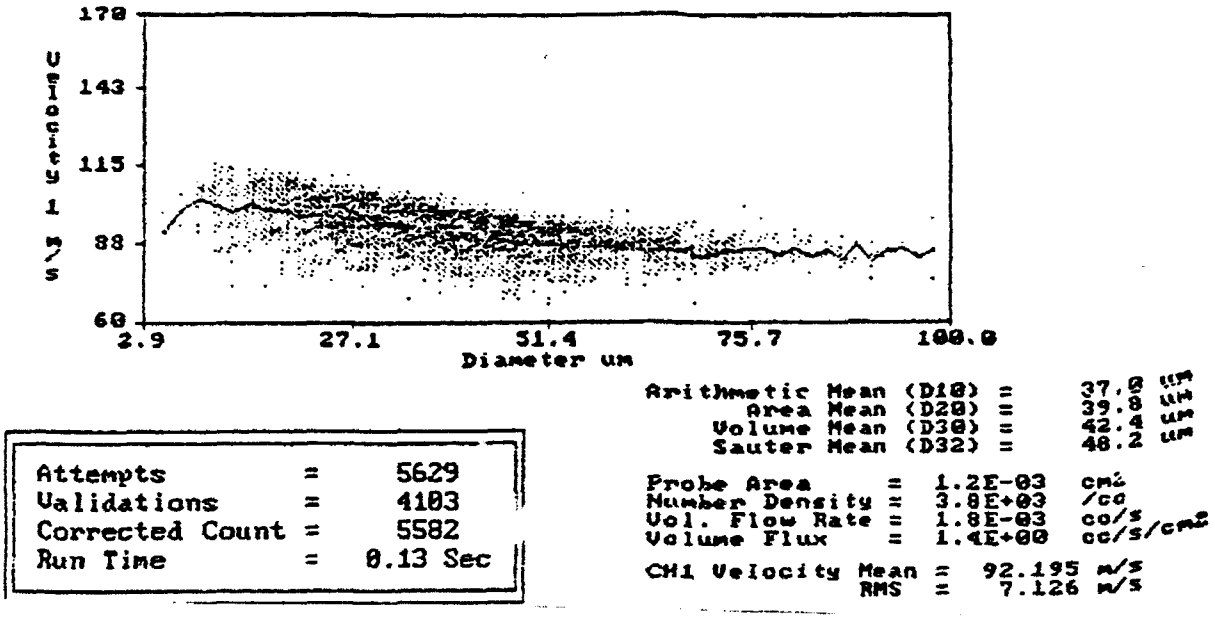


Figure 4. Velocity-diameter correlation. (X = 3.00, Y = 0.0, Z = 1.40).

diameter correlation at this same location. As anticipated, the larger droplets have a lower average velocity in the downstream direction.

A series of such statistics were combined to provide the profile data shown in Figure 5. This represents a series of points in a vertical traverse (variable Z), on the tunnel centerline (Y = 0.0) and 3.00 inches downstream of the injector (X = 3.00). It can be seen that the average diameter varies from a low of 12.57 microns near the floor to a high of 46.764 microns at the upper edge of the spray. The number density is low near the floor and also in the upper region, with a maximum at Z = 1.2. The volume flux is highest at Z = 1.4. Although there are fewer droplets at Z = 1.4 than at lower elevations, the average diameter is larger, leading to the peak in volume flux.

Figure 6 represents a horizontal traverse (variable Y) at a fixed elevation (Z = 1.00), 3.00 inches downstream of the injector. It is again seen that the larger droplets are at the periphery of the spray pattern, with higher number densities of smaller droplets in the center. The spray is not completely symmetric, of course, but the general trends are the same on either side of the tunnel centerline at Y = 0.0. It should also be noted that the centerline of the tunnel is not necessarily the centerline of the plume, as drifting to one side or the other is typical in such systems.

SUMMARY

A single-component phase-Doppler particle analyzer was assembled and validated in bench-top testing, and later applied to an experimental fuel injection tunnel. It was concluded that the system produces reliable data for droplet size over the range from 3 to 300 microns and good results (7 percent) for average velocity. RMS velocity results were not felt to be reliable. Data sufficient to produce three-dimensional mapping of spray distributions produced by normal injection of water into a crossflowing air stream were acquired. A preliminary analysis of the data indicates that droplet diameters are generally larger on the upper periphery of the plume, with smaller

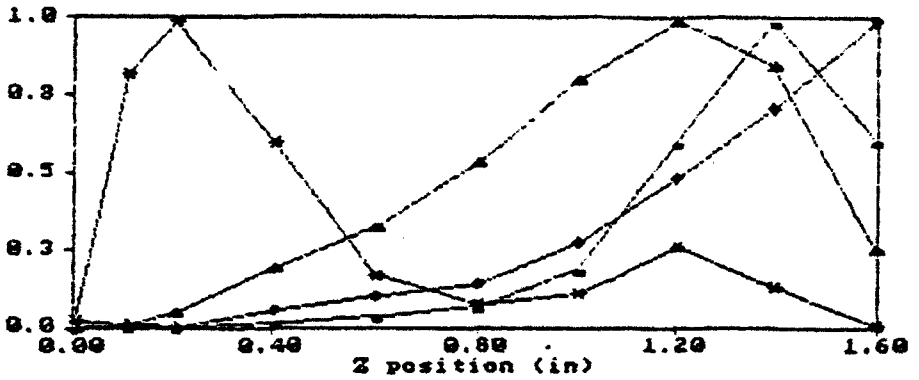


Figure 5. Dependence of droplet parameters on distance above test section floor. (X = 3.00, Y = 0.0, Z variable).

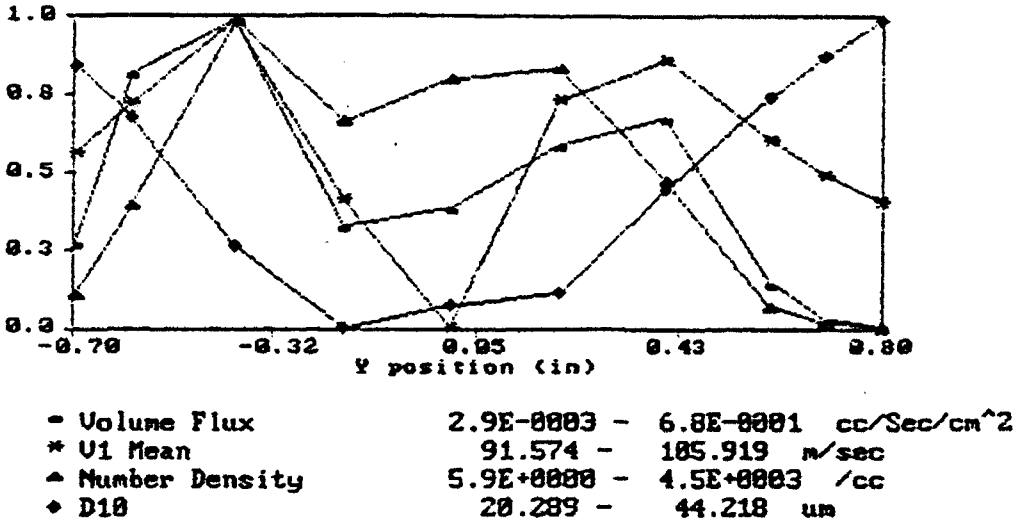


Figure 6. Dependence of droplet parameters on lateral position across tunnel. (X = 3.00, Z = 1.00, Y variable).

droplets near the tunnel wall and in the central core of the spray. It is recommended that full three-dimensional mapping of the acquired data be initiated.

ACKNOWLEDGMENTS

The author wishes to express his appreciation to all Aeropropulsion and Power Directorate personnel who contributed to this research effort. Special appreciation goes to Gary Switzer of SRL and Cindy Obringer of the Fuels and Lubricants Division for assistance in the initial setup and validation of the instrumentation. Advanced Propulsion Division personnel who contributed greatly to the fuel-injection phase of the program include Charles Smith and Tsong Chen. The technical guidance and support of Abdi Nejad are especially appreciated.

REFERENCES

1. Bachalo, W. D., and M. J. Houser, "Phase/Doppler Spray Analyzer for Simultaneous Measurements of Drop Size and Velocity Distributions," Optical Engineering, vol 23, no 6, pp. 583-590, Sept/Oct 1984.

2. Bachalo, W. D., A. Brena de la Rosa, and R. V. Sankar, "Diagnostics for Fuel Spray Characterization," Combustion Measurements, N. Chigier, ed., pp. 229-278, Hemisphere, 1991.

3. Switzer, G. L., "A Versatile System for Stable Generation of Uniform Droplets," Review of Scientific Instruments, vol 62, no 11, pp. 2765-2771, Nov 1991.

A STUDY OF MILLIMETER-WAVER RADAR AND INFRARED SENSOR FUSION USING NEURAL NETWORKS

Thaddeus A. Roppel
Assistant Professor
Department of Electrical Engineering

Auburn University
200 Broun Hall
Auburn, AL 36849-5201

Final Report for:
Summer Research Program
Wright Laboratory

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

September 1992

A STUDY OF MILLIMETER-WAVER RADAR AND INFRARED SENSOR FUSION USING NEURAL NETWORKS

Thaddeus A. Roppel
Assistant Professor
Department of Electrical Engineering
Auburn University

Abstract

The application of an artificial neural network (ANN) to the problem of sensor fusion was studied. Pixel-level data from captive flight tests was used to form the training and testing data sets. The neural network simulation program Aspirin was used to perform back-propagation training. Infrared (IR) pixel images from the available data sets were reduced from 128 x 544 pixels to 10 x 10 pixels after manually rotating and segmenting the images.

During the first phase of the experiment, a two-layer ANN consisting of 100 inputs, 20 hidden layer nodes, and two output nodes was trained to distinguish between two target types, a T-62 tank and a Lance missile launcher. A total of 14 IR target scenes were available, of which 9 were used for training the ANN and 5 were used for testing the ANN after training. Testing showed that the trained ANN was able to distinguish between the two target types 100% of the time with high confidence.

During the second phase of the experiment, the IR data was augmented by including an additional 128 pixels of millimeter-wave (mmw) radar data which had been obtained simultaneously with the IR data. Inclusion of the mmw radar data reduced the overall rms error in the ANN output layer by 14% compared with using the IR data alone.

A STUDY OF MILLIMETER-WAVER RADAR AND INFRARED SENSOR FUSION USING NEURAL NETWORKS

Thaddeus A. Roppel

1.0 INTRODUCTION

This report is based on research conducted from June 15 - September 11, 1992 at the Wright Laboratory Armament Directorate, Eglin AFB, Florida under the AFOSR summer faculty research program. Mr. Ellis Boudreaux was the laboratory focal point for this work.

The research reported here can be divided into three roughly sequential tasks: background literature review, data compilation, and neural network experiments. A brief review of the pertinent literature is included in the Discussion section of this report. The data compilation is described in the Methodology section. The remainder of this report focuses on the neural network experiments and the results.

2.0 DISCUSSION

The specific problem under conside. ation is to determine whether the performance of an IR/MMW dual-mode sensor system can be improved by using a neural network early in the data stream to detect and classify targets. This study addressed one of the most fundamental sensor fusion issues, namely, the performance of a neural-network based system provided with both IR and MMW data vs. a neural-network based system provided with only IR data.

The background literature review for this problem yielded useful insights into the current status of work on this problem. Representative articles are included in references (1-4). The general concensus is that neural network sensor fusion is most effective when applied prior to final feature extraction, but after some preprocessing of the raw sensor data. A number of questions remain unanswered in the literature, such as the efficiency of implementing scale, translation, and rotation-invariant transformations in the neural network compared with implementing these transformations before the ~etwork. Also, a great deal of effort has been devoted to the problem of coregistration of radar and IR pixel data. Finally, there seems to be general agreement that the best results are obtained using neural networks trained for very specific applications (e.g., air-to-ground with clearly defined targets), rather than trying to train one neural network for all possible fusion applications. The system overview for this research is shown in Figure 1.

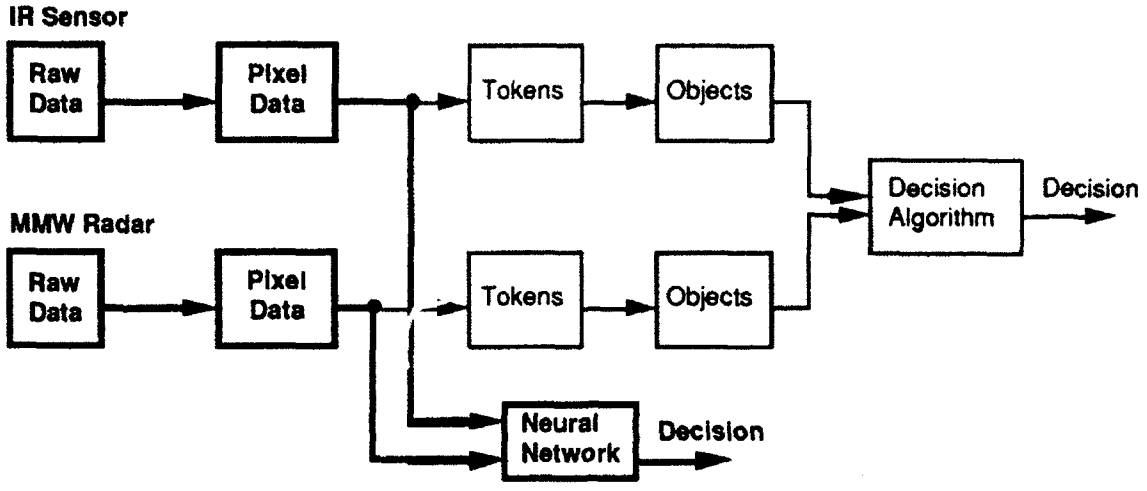


Figure 1. Overview of the data flow for the experiments in this research (bold) compared with conventional fusion approach.

3.0 METHODOLOGY

The experimental work required designing an experimental matrix, converting the available data into a format suitable for input into the neural network simulator, training the neural network, and then post-processing to analyze the results.

3.1 Experimental Matrix

There were 38 target scenes available in the pixel data set. The three most often represented targets were the T-62 tank (12 scenes), the M-113 APC (9 scenes), and the LANCE missile launcher (5 scenes). For the experiments reported here, it was decided to use four of the LANCE scenes and ten of the T-62 scenes. Furthermore, seven of the T-62 scenes and two of the LANCE scenes were used to train the neural network, while the remaining three T-62 scenes and two LANCE scenes were used to test the neural network after training. The experimental design is shown in Figure 2.

Total Number of Scenes: 14

	Training Scenes	Testing Scenes
Lance Missile Launcher	2	2
T-62 Tank	7	3

Procedure:
(1) Train to 0.1 maximum error with IR data only.
(2) Train to 0.1 maximum error with both IR and MMW data.
(3) Compare rms error resulting from testing the neural networks trained in (1) and (2).
(4) Compare the confusion matrices resulting from testing the neural networks trained in steps (1) and (2).

Figure 2. Experimental design for the sensor fusion investigation.

3.2 Data Preparation

The data used in these experiments, referred to as the Westinghouse Data, was collected during captive flight tests using a hybrid MMW / I^2R common 5-1/4" aperture sensor. The MMW data had a format similar to that of Synthetic Aperture Radar (SAR). The formatting technique is referred to as Doppler Beam Sharpening (DBS). In this format, a 20 milliradian field of view is compressed to approximately 2 milliradians. This degree of compression enables registration of MMW radar pixels with those obtained from the HgCdTe imaging infrared focal plane array. The MMW data was organized as a 128 x 96 pixel array of elevation vs. range data.

Each of the target scenes used for training and testing consisted of 128 x 544 pixels of IR data, and 8 sets (4 dwells x 2 polarization states) of 128 x 96 pixels each of MMW data. For these experiments, the IR data was reduced to 10 x 10 pixels, and the MMW data was reduced to 8 x 16 pixels. For the IR data, the reduction was accomplished by manually segmenting the images to a 40 x 40 pixel array, and then applying neighborhood averaging. For the MMW data, the reduction was accomplished entirely by neighborhood averaging. Thus, of the available 167,936 pixels in each scene, 228 pixels were used as inputs to the neural network. Each pixel was represented by eight bits. For presentation to the neural network, the integer pixel values in the range from 0 to 255 were scaled to floating point values in the range from 0.05

to 0.95. The neural network desired output values were set to either 0.05 (target not present in input scene) or 0.95 (target present in input scene).

3.3 Neural Network Training

The neural network simulator used was the Aspirin / MIGRAINES code authored at MITRE Corporation, which is distributed free of charge. This work used version 5.0 of the code. The code was installed and run on a Sun Sparc 2 workstation. The neural network architecture is shown in Figure 3. The code was run with the learning rate set to 0.2 and the inertia set to 0.95. The error bound was set to 0.1. Two training runs were performed. In the first run, only the 100 pixels of IR data from each target scene were used as inputs. In the second run, the IR pixel data for each scene was augmented with the 128 pixels of MMW data available for that scene, for a total of 228 neural network inputs.

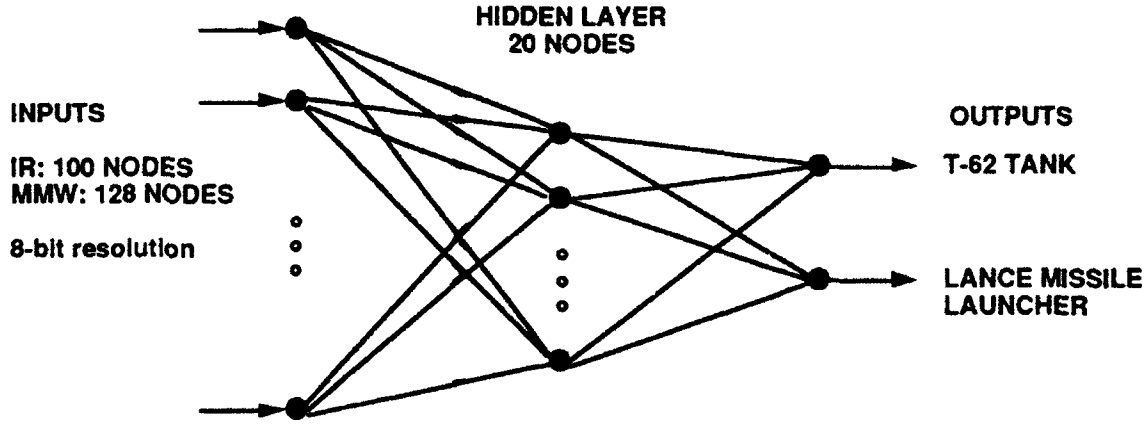


Figure 3. Neural network architecture. The sensor fusion neural network consists of 228 inputs, 20 hidden layer nodes, and two output nodes. There are 4622 connections, including bias inputs for the hidden and output nodes.

3.4 Neural Network Testing

After training, the five test data sets were propagated through the network. For each scene, the rms error between the desired output and the actual output for each output node was computed. Also, the overall rms error computed over all the test scenes was used as a quantitative measure of performance. This allowed a simple quantitative comparison to be made between the case with only IR data and the case with both IR and MMW data.

4.0 RESULTS

Table 1 summarizes the training statistics for the neural network. Table 2 summarizes the performance of the trained neural network in classifying the test data.

Table 1. Training statistics for the neural network for two cases: IR input only and both IR and MMW input.

Parameter	IR only	IR and MMW
Learning Rate	0.2	0.2
Inertia	0.95	0.95
Error Bound	0.1	0.1
Total Epochs to Train this Network	30	20
Number of Connections in Network	2.062×10^3	4.622×10^3
Elapsed Compute Time	1.46	2.51
Average Performance (connections per second)	3.81×10^5	3.31×10^5

Table 2. Neural network performance summary.

Target	Type	IR only			IR and MMW Sensor Fusion		
		Desired Outputs	Actual Outputs	Percent Error	Desired Outputs	Actual Outputs	Percent Error
1	Lance	0.95	0.8697	-8.03	0.95	0.8697	-8.03
		0.05	0.1303	8.03	0.05	0.1303	8.03
2	Lance	0.95	0.8625	-8.75	0.95	0.9377	-1.23
		0.05	0.1375	8.75	0.05	0.0623	1.23
3	T-62	0.05	0.0579	0.79	0.05	0.0947	4.47
		0.95	0.9421	-0.79	0.95	0.9053	-4.47
4	T-62	0.05	0.0365	-1.35	0.05	0.0319	-1.81
		0.95	0.9630	1.30	0.95	0.9681	1.81
5	T-62	0.05	0.0382	-1.18	0.05	0.0437	-0.63
		0.95	0.9612	1.12	0.95	0.9563	0.63
Overall rms percent error				1.70			1.46

Table 2 shows the raw error that results from testing the trained neural network. From the results shown in Table 2, it is clear that the performance of the neural network with only IR input is very good, as indicated by the 1.7 percent overall error. The performance improves when the MMW data is fused with the IR data, in which case the overall error is reduced to 1.46 percent. These values indicate a relative error reduction of 14.1% due to sensor fusion. Table 3 shows the target classification that results from thresholding the neural network outputs at 0.5. In this case, for example, if the "Lance" node output is greater than 0.5 and the "T-62" node output is less than 0.5, the logically correct classification is determined to be "Lance." In fact, the neural network has been trained so that a relatively wide noise margin could be applied, where, for, example, a node output greater than 0.8 would represent "TRUE" and a node output of less than 0.2 would represent "FALSE." In this dual threshold case, the noise margin would be the difference between the high threshold and the low threshold, or 0.6.

Table 3. Confusion matrix showing neural network classifier performance. The data are valid for both the IR-only case and the IR / MMW fusion case.

	Neural Net Classification	
	Lance	T-62
Actual Scene Lance	4	0
T-62	0	10

From Table 3, it can be observed that the neural network has learned to discriminate the two target types with 100% accuracy using only the IR data, and the inclusion of the MMW data cannot therefore be expected to improve the classification rate in this case. Future experiments are required in which classification is not 100% accurate for the IR-only case so that a quantitative measure of the net effect of sensor fusion on classification performance can be made.

5.0 CONCLUSIONS AND DIRECTIONS FOR FUTURE WORK

This study has shown that a neural network can be trained to classify IR targets with high accuracy. Furthermore, the neural network has been shown to provide error reduction when used to implement sensor fusion at the pixel level using IR and MMW radar data. A number of questions requiring further investigation are raised by these results. Possible directions for future work to address these issues are summarized below:

(1) Compare the performance of neural-network based sensor fusion with conventional methods based, for example, on Bayesian statistical methods.

(2) Measure the neural network performance on "harder" image problems, for example, the original IR pixel images from this data set with no segmentation or rotation pre-processing.

(3) Measure the performance on non-target images, i.e., images that contain neither T-62 nor Lance target scenes.

(4) Design and implement the neural network in hardware.

These and related topics will be pursued in the author's continuing work at Auburn University.

6.0 REFERENCES

1. GACIAC Space PR89-01, Proc. of the Second National Symposium on Sensors and Sensor Fusion Vol. II: Unclassified Papers Sponsored by Office of the Undersecretary of Defense Acquisition (Research Advanced Technology), March 27-31, 1989.

2. Sensor Fusion II, Proc. of the SPIE, Orlando, FL, 28-29 March, 1989.

3. Joe R. Brown, Dan Bergondy, and Sue Archer, "Multi-sensor ATR using Neural Nets," Proc. of the Second Workshop on Neural Networks: WNN-AIND 91, SPIE Volume 1515, pp. 319-322, Auburn, AL, Feb. 11-13, 1991.

4. Personal communication with Dr. Kwang Min, Eglin AFB, FL.

TOWARD A CHARACTERIZATION OF THE DEBRIS CLOUD CREATED IN A HYPERVELOCITY IMPACT ON A THIN PLATE

Dr. William P. Schonberg
Associate Professor
Civil Engineering Program

University of Alabama in Huntsville
Huntsville, Alabama 35899

FINAL REPORT

Submitted To:
Technology Assessment Branch
Analysis and Strategic Defense Division
Armament Directorate
Wright Laboratory
Eglin Air Force Base, Florida

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

September 1992

TOWARD A CHARACTERIZATION OF THE DEBRIS CLOUD CREATED IN A HYPERVELOCITY IMPACT ON A THIN PLATE

Dr. William P. Schonberg
Associate Professor
Civil Engineering Program
University of Alabama in Huntsville

Abstract

Current semi-analytical lethality assessment models typically fall into one of two general categories: discrete particle models and expanding shell models. Discrete particle models account for and track only a small number of solid fragments in a debris cloud generated by a high speed impact. As such, they are best suited for impact scenarios in which melting and/or vaporization of the projectile and target materials do not occur. Alternatively, expanding shell models typically assume that all of the debris cloud material is homogeneously distributed over a uniformly expanding spherical shell and are therefore applicable only in those impact situations where melting and/or vaporization of the projectile and target materials do occur. A need clearly exists to bridge the gap between the discrete particle models and the expanding shell models. Specifically, a lethality assessment model that considers the creation and subsequent effects of debris clouds containing all three states of matter is needed. This report describes the results of an investigation into the composition of the material in a debris cloud generated by a hypervelocity projectile impact. The work completed represents the first step in a long-term research program whose overall objective is to develop a general model of the response of a target structure to a hypervelocity impact over the 4-16 km/sec impact velocity regime.

TOWARD A CHARACTERIZATION OF THE DEBRIS CLOUD CREATED IN A HYPERVELOCITY IMPACT ON A THIN PLATE

William P. Schonberg

INTRODUCTION

The response of a target to a kinetic energy weapon (KEW) impact can be said to consist of two basic and distinct types of response: 'local response' and 'global response'. Local response is primarily due to the intense loading associated with a hypervelocity impact. For KEW impacts, material damage occurs very quickly (on the order of microseconds) and is limited to an area near the impact site. At sufficiently high impact velocities, shatter, melting, and/or vaporization of the materials can occur. For an aluminum-on-aluminum impact, the projectile and target materials will begin to shatter, melt, and vaporize at impact velocities of approx. 3.2, 5.6, and 10.4 km/sec, respectively [1,2].

Global response can refer to any one of a number of global phenomena that occur over a longer period of time (on the order of milliseconds), under less intense loads, and over a larger area of the target structure. In KEW impacts, one or more debris clouds are created during the initial impact on the outer wall of a target. These debris clouds spread out as they move through target voids and eventually impact an inner wall or interior component of the target structure. Depending on the impact velocity and the relative material properties of the projectile and target, these debris clouds can contain solid, melted, and/or vaporized projectile and target materials. Typical global responses include the denting, buckling, or tearing of an internal missile component such as a fuel or oxidizer tank.

The research described in this report is the first step in a long-term research program whose overall objective is to develop a general model of

the response of a target structure to a KEW impact over the anticipated impact velocity regime of 4-16 km/sec. The objectives of the work performed thus far were to 1) characterize the shock loading and release of the projectile and target materials due to a hypervelocity impact in the 4-16 km/sec impact velocity regime; 2) estimate the percentages of solid, liquid, and gaseous materials in the debris cloud created in a hypervelocity impact; and, 3) estimate the amount of mass in each of the three states of matter for the debris cloud material. A more detailed description of the work done will be presented in a forthcoming Air Force Technical Report.

LETHALITY ASSESSMENT MODEL REQUIREMENTS

The key to conducting an accurate lethality assessment is the use of a robust assessment methodology. The methodology should incorporate all the significant response and damage mechanisms which result from all hypervelocity weapon-target interactions. To accurately determine the total damage level sustained by an impacted target, a lethality assessment methodology must include the effects of discrete and simultaneous debris cloud fragment impacts, as well as impulsive target debris cloud loadings. Discrete or simultaneous impacts by individual fragments can pose a lethal threat to the inner wall or to an interior component of a target, depending on the fragments' speed, density, and trajectory, and on the density and strength of the target inner wall or interior component material. Individually, the molten and/or vaporous fragments in a debris cloud may not do significant damage; however, as a whole, they can produce a significant impulsive loading over a relatively large area inside the target. This in turn can result in further damage to the target at later times. Clearly then, to accurately assess the total damage to a target impacted by a KEW, the amounts and types of debris in a debris cloud produced by a hypervelocity

impact must be known.

A number of empirical and semi-analytical procedures have been developed over the past decade to determine the lethal effectiveness of KEW systems. While these procedures are capable of assisting engineers and system architects in optimizing weapon designs and in performing cost trade-off studies, they are significantly limited in their characterization of the material in the debris clouds created by hypervelocity impacts. Unfortunately, very little impact test data for relatively massive projectiles (on the order of 10 gms or more) is available at speeds above 8 km/sec. This makes it difficult to properly charcterize the nature of the material in the debris clouds over the entire impact velocity regime of interest. Electrostatic devices which can launch small particles to speeds as high as 100 km/sec exist, but these systems can only launch micron-size particles [3,4]. Other electric gun systems have launched Kapton flyer plates to speeds of 11 km/sec, but cannot reach that velocity with chunky projectiles [5]. Thus, existing lethality assessment models must be used with a fair amount of caution, especially in scenarios involving impact velocities greater than those attainable in experiments.

Current semi-analytical lethality assessment models usually fall into one of two broad groups: discrete particle models [6-17] and expanding shell models [18-22]. Discrete particle models typically account for only solid fragments [6-9,14-17], or track only a small number of discrete fragments [10-13] in the debris cloud created by a high speed impact. These models are best suited for applications in which the debris clouds generated by the initial impact contain only a relatively small number of fragments and in which melting or vaporization of the projectile and target materials do not

occur. The expanding shell models typically assume that all of the debris cloud material is homogeneously distributed over a uniformly expanding spherical shell. These models are applicable only in those impact situations where complete projectile and target material vaporization occurs.

It is evident, therefore, that the need exists to bridge the gap between the discrete particle models, which consider only a finite number of solid fragments, and the expanding shell models, which are valid only when complete vaporization occurs. Specifically, a lethality assessment model that considers the creation and subsequent effects of debris clouds containing all three matter states is needed. FATEPEN [6-9], KAPP-II [11-13], and PEN-4 [16,17] are discrete particle lethality assessment models which can be modified to include the effects of non-solid debris cloud constituents.

CHARACTERIZATION OF DEBRIS CLOUD MATERIAL

The research efforts discussed in this report were directed at the development of a procedure that would extend the applicability of existing discrete particle lethality assessment methodologies to impact scenarios in which the projectile and target materials were expected to melt and/or vaporize. Specifically, the work performed consisted of a series of tasks directed at determining to first-order accuracy the amount of projectile and target material in a debris cloud that is solid, molten, and/or vaporized. Projectiles considered were metallic monolithic right circular cylinders with a length-to-diameter ratio of 2 and which normally impacted thin flat metallic target plates with a zero angle of yaw.

Shock Loading and Release Analysis

Consider the impact of a cylindrical projectile on a flat target plate. Upon impact, strong shock waves are set up in the projectile and target materials. The pressures associated with these shocks typically exceed the

strengths of the projectile and target materials by several orders of magnitude. For example, in an 8 km/sec aluminum-on-aluminum impact, the ratio of the impact pressure (116.5 GPa=1.15 MBar) to the strength of the material (310 MPa for aluminum 6061-T6) is approximately 375, or roughly 2.5 orders of magnitude. As the shock waves propagate, the projectile and target materials are heated adiabatically and non-isentropically. The release of the shock pressures occurs isentropically through the action of rarefaction waves that are generated as the shock waves interact with the free surfaces of the projectile and target. This process leaves the projectile and target materials in high energy states and can cause either or both to fragment, melt or vaporize, depending on the material properties, geometric parameters, and the velocity of impact. At very early times during the impact event, only the area in the immediate vicinity of the impact site is affected by the impact. For the projectile and target geometries considered in this study, the shock waves can be considered to be initially planar. This allows one-dimensional relationships to be used for analyzing the creation and release of shock pressures.

The shock pressures, energies, etc., in the projectile and target materials were calculated using the three shock-jump conditions, a linear relationship between the shock wave velocity and particle velocity in each material, and continuity of pressure and velocity at the projectile/target interface. In order to calculate the release of the projectile and target materials from their respective shocked states, an appropriate equation-of-state was needed for each material. To keep the analysis relatively simple, the Mie-Gruneisen [23] and Tillotson [24] equations-of-state (EOS) were used in this study.

The Mie-Gruneisen EOS is an accurate thermodynamic description of most metals in the solid regime and is relatively easy to use. Based on its thermodynamic origins, it cannot be expected to give accurate results in the expanded liquid regime or in the vapor regime. Experience has shown, however, that it does yield fairly accurate end-state results even when there is a small percentage of molten material present [1]. The Tillotson EOS is asymptotically correct in the compression and expansion regimes and reproduces many of the isentropic release features observed with much more complicated equations-of-state [25].

A one-dimensional shock loading and release process was used to determine the end state of the projectile and target material portions experiencing shock loading and release. However, because of its inherent limitations, the Mie-Gruneisen EOS was eventually abandonded in favor of the Tillotson EOS. The internal energies in the shocked and released portions of the projectile and target materials were calculated using the Tillotson EOS and were translated into temperature increases using classical thermodynamics. Figures 1-4 compare the results of the release process for aluminum-on-aluminum impacts at three different energy levels using the Mie-Gruneisen and Tillotson equations-of-state.

In Figure 1, the release process as described by the Mie-Gruneisen EOS and the Tillotson EOS are nearly identical. This is to be expected for relatively low energy impact (i.e. those impacts in which the materials return to a solid matter state after release). Figure 2 shows the dramatic difference between using the Mie-Gruneisen EOS and the Tillotson EOS for very high energy impacts (i.e. those impacts in which the materials vaporize). The Mie-Gruneisen EOS cannot account for the expansion of the gaseous state and terminates the release process at a much lower specific volume

than the Tillotson EOS.

Figure 3 highlights one of the difficulties in using the Tillotson EOS. This difficulty occurs under impact conditions that are not violent enough to vaporize the material, yet are strong enough to cause the material to melt and be in an energy state that is near incipient vaporization. Under these conditions there is a jump in the release isentrope generated by the Tillotson EOS which results in a final volume that is artificially high.

It is proposed that the spurious behavior of the Tillotson EOS in moderately high energy impacts can be eliminated by subtracting the magnitude of the jump from the pressure values generated using the original Tillotson EOS equations. Implementation of this corrective action in this impact energy regime caused the release processes to terminate at specific volume values that were much more reasonable. It is noted that this correction had no effect when the impact energy was relatively low or very high. Figure 4 shows the result of implementing the jump correction for a 10 km/sec aluminum-on-aluminum impact.

The differences in the final specific volumes obtained in aluminum-on-aluminum impacts using the Mie-Gruneisen, Tillotson, and Modified Tillotson equations-of-state are shown in Figure 5. For low energy impacts (below approx. 9 km/sec), the results are, as expected, nearly identical. For very high energy impacts (above approx. 18 km/sec), the final values predicted by the Tillotson EOS and the Modified Tillotson EOS (upper curve) overlap and exceed those predicted by the Mie-Gruneisen EOS (lower curve) due to the gaseous expansion of the released material at those impact velocities. The odd behavior in the final values of specific volume due to the jump in the Tillotson EOS begins for aluminum at approximately 9 km/sec (upper curve in

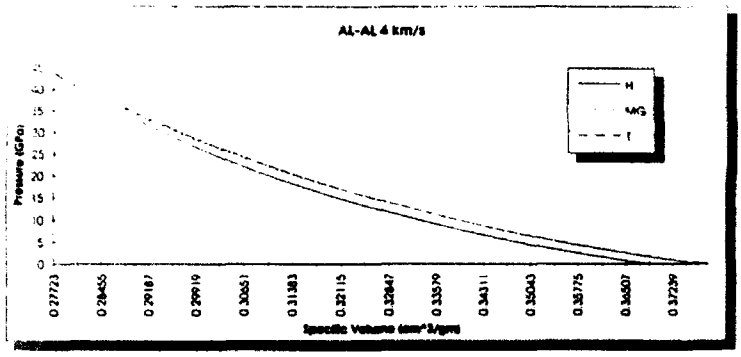


Figure 1. Low Energy Impact Shock Loading and Release Curves (H ... Hugoniot, MG ... Mie-Gruneisen EOS Release Isentrope, T ... Tillotson EOS Release Isentrope)

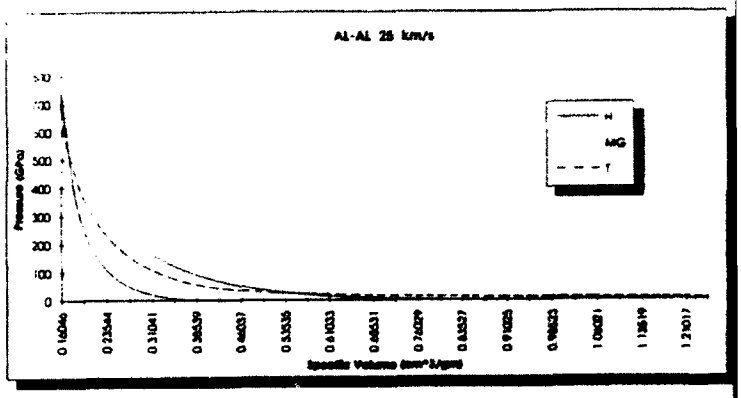


Figure 2. High Energy Impact Shock Loading and Release Curves (H ... Hugoniot, MG ... Mie-Gruneisen EOS Release Isentrope, T ... Tillotson EOS Release Isentrope)

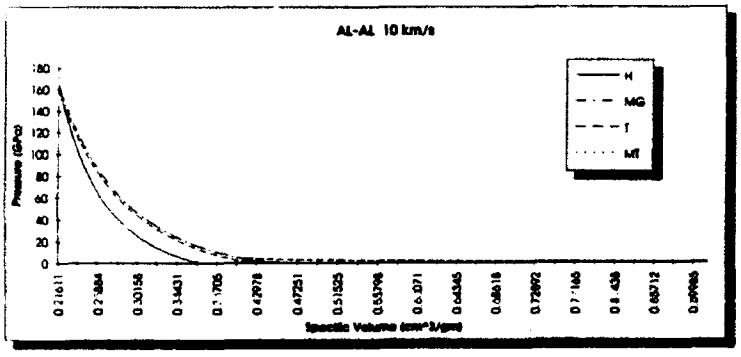


Figure 3. Moderate Energy Impact Shock Loading and Release Curves (H ... Hugoniot, MG ... Mie-Gruneisen EOS Release Isentrope, T ... Tillotson EOS Release Isentrope, MT ... Modified Tillotson EOS Release Isentrope)

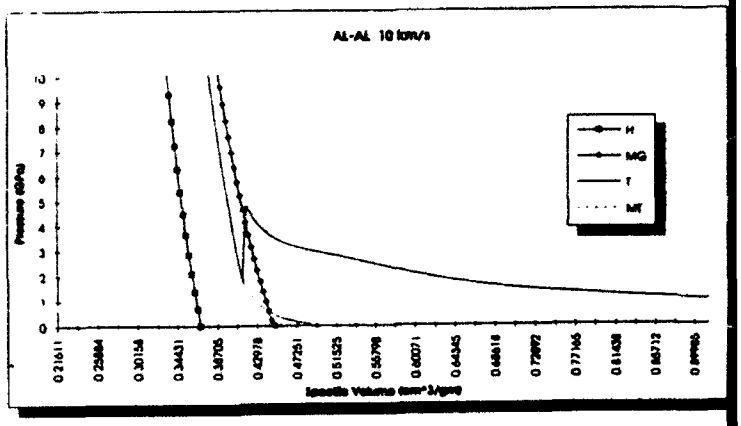


Figure 4. Close-up of Moderate Energy Impact Shock Loading and Release Curves (H ... Hugoniot, MG ... Mie-Gruneisen EOS Release Isentrope, T ... Tillotson EOS Release Isentrope, MT ... Modified Tillotson EOS Release Isentrope)

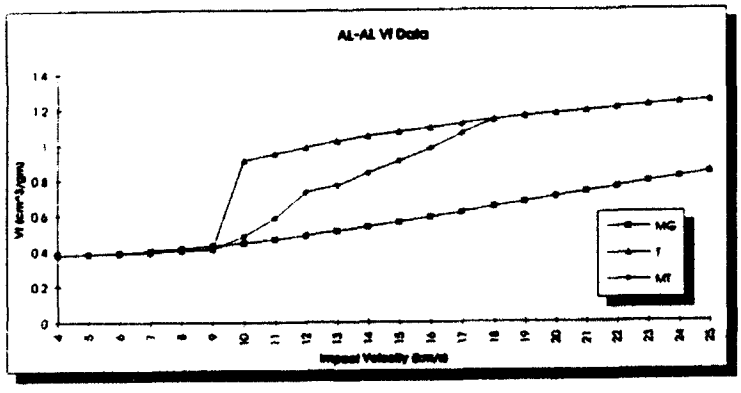


Figure 5. Final Specific Volume vs. Impact Velocity (MG ... Mie-Gruneisen EOS Volume, T ... Tillotson EOS Volume, MT ... Modified Tillotson EOS Volume)

Figure 5). However, the Modfified Tillotson EOS (middle curve) produces a smooth transition as the material changes from a solid state (below approx. 6 km/sec) to a liquid state (between approx. 6 and 11 km/sec) to a gaseous state (above approx. 11 km/sec). It is this modified version of the Tillotson EOS (hereafter referred to as the Modified Tillotson EOS) that was used throughout the remainder of this study.

Computing the Percentages of Solid, Liquid, and Gaseous Debris Cloud Material

Once the residual internal energies in the shocked and released portions of the projectile and target materials had been obtained, the percentages of the various states of matter in the resulting debris cloud were estimated using the following procedure. This procedure requires the knowledge of the materials' solid and liquid specific heats (C_{ps},C_{pl}), their melting and boiling points (T_m,T_v), and their heats of fusion and vaporization (H_f,H_v) in addition to the residual internal energy (E_r).

If $E_r<C_{ps}T_m$, then all of the shocked and released materials was considered to remain in a solid matter state. If $C_{ps}T_m<E_r<C_{ps}+H_f$, then the quantity $(E_r-C_{ps}T_m)/H_f$ represented the fraction of the shocked and released material that was melted. The remaining shocked and released material was assumed to be in solid form. If $C_{ps}T_m+H_f<E_r< C_{ps}T_m+H_f+C_{pl}(T_v-T_m)$, then all of the shocked and released material was considered to be in a liquid state. If $C_{ps}T_m+H_f+C_{pl}(T_v-T_m)<E_r< C_{ps}T_m+H_f+C_{pl}(T_v-T_m)+H_v$, then the quantity $(E_r-[C_{ps}T_m+H_f+C_{pl}(T_v-T_m)])/H_v$ represented the fraction of the shocked and released material that was vaporized. The remaining shocked and released material was considered to be in liquid form. If $C_{ps}T_m+H_f+C_{pl}(T_v-T_m)+H_v<E_r$ then all of the shocked and released material was vaporized.

Computing the Masses of the Solid, Liquid, and Gaseous Debris Cloud Material

The material in the debris cloud created by the initial impact consists

of the target material removed by the impact and the impacting projectile mass. While the mass of the projectile material in the debris cloud was known a priori, the mass of the target material in the debris cloud had to be determined by multiplying the target hole-out area by the target thickness and the target material density. The size of the hole created in the target plate by the initial impact was calculated using an empirical equation for target hole diameter [11]. While the empirical nature of the equation mandates its use only within the impact velocity regime for which it was designed, the results obtained for velocities outside the prescribed regime were not unreasonable.

To calculate the masses of the various states of the projectile and target materials in the debris cloud, the amounts of shocked and released target and projectile material had to be determined. These quantities were obtained by determining the locations in the target plate and in the projectile where the rarefaction waves had overtaken the corresponding shock wave [26]. It was the material through which both the shock wave and the release wave had travelled that was shocked and released and which was therefore either melted or vaporized, depending on the particulars of the impact event. Any material beyond the point at which the rarefaction wave had overtaken the shock wave was assumed, for the purposes of this study, not to have been shocked and to have remained in a solid matter state. If the point at which the release wave had overtaken the shock wave was beyond the thickness of the target plate or the length of the projectile, then all of the target and/or projectile material had been shocked and released.

Once the projectile and target mass contributions to the debris cloud and the fractions of these masses that were shocked and released were ob-

tained, the masses of the target and projectile materials in each of the three states of matter were computed by multiplying each matter state percentage by the appropriate total shocked and released mass. The mass of the solid shocked and released material (if any) was then added to the mass of the unshocked material (if any) to obtain the total mass of the solid component of the material in the debris cloud.

The major limitation of this procedure is the assumption that no further projectile and/or target loading and unloading had occurred beyond the point where the release waves had overtaken the corresponding shock wave. This is not completely correct since the shock wave does not simply cease to exist once it is overtaken by a rarefaction wave. Rather, its magnitude decreases over a finite amount of time and a finite extent of material. Some additional projectile and target material will be heated and possibly melted until the strength of the shock wave diminishes to a point below which melt due to plastic deformation no longer occurs. However, the procedure set forth does allow the calculation of first-order accurate mass quantities for projectile and target materials in the three states of matter.

RESULTS AND DISCUSSION

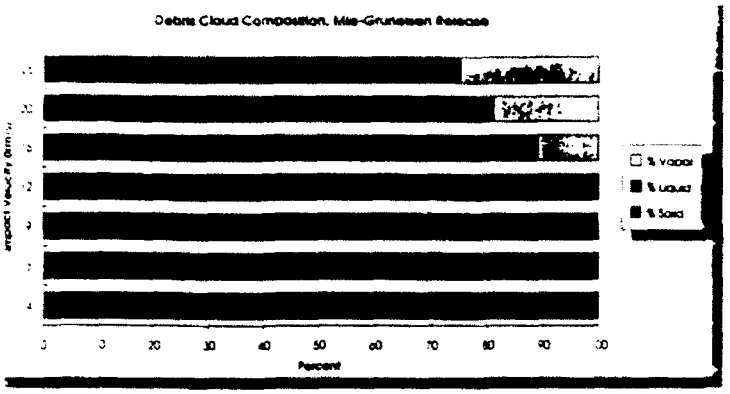
A FORTRAN program was written to implement the various procedures described in the preceding section. The details of the program will be discussed in a forthcoming Air Force Technical Report, which will also include the source code, input files, and sample output. While the Tillotson EOS is relatively straightforward to implement, its use requires a fair amount of familiarity with its peculiarities. These will also be discussed in detail in the Technical Report.

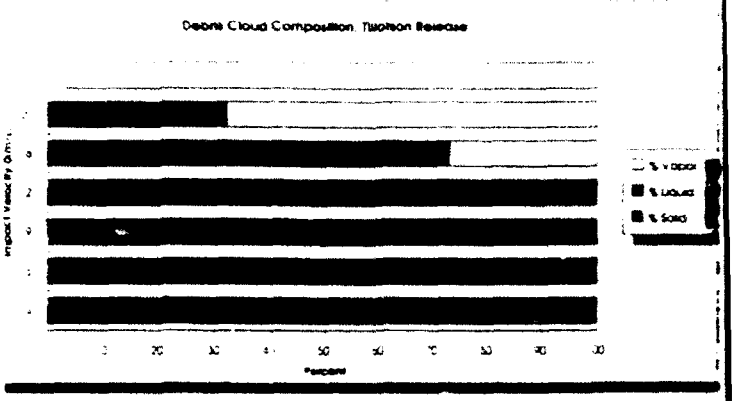
As can be seen in Figure 6, the Mie-Gruneisen EOS predicted only a small amount of vaporized material at an impact velocity as high as 25

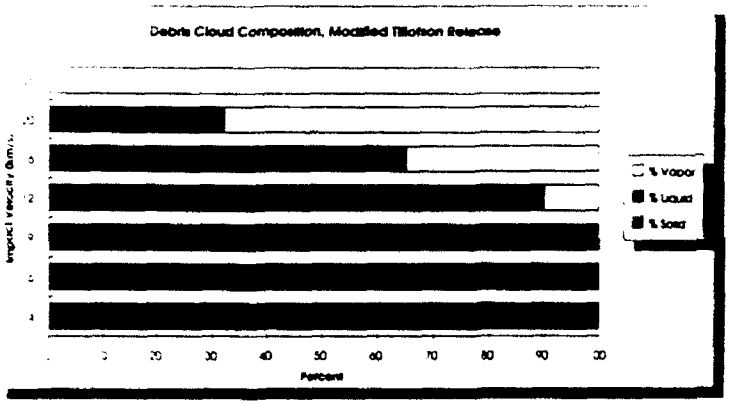
km/sec. However, both the Tillotson and the Modified Tillotson equations-of-state predicted that the aluminum was completely vaporized at an impact velocity between 20 and 25 km/sec. This difference is due to the fact that the Mie-Gruneisen EOS did not account for the expansion of the material as it nears vaporization and completed the release process with the material in a much lower energy state than the Tillotson EOS.

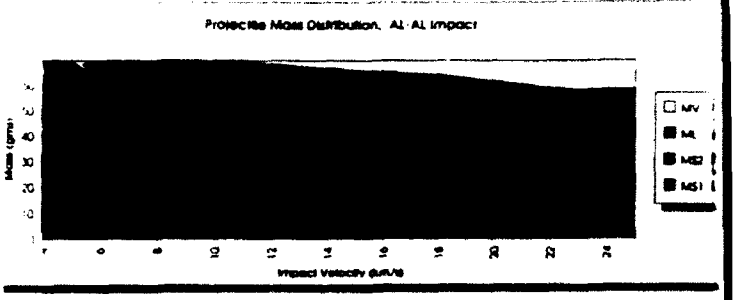
Comparing Figures 7 and 8 reveals that the Tillotson and the Modified Tillotson equations-of-state agreed in the percentages of the various states of matter at speeds below below approx. 9 km/sec and above approx. 18 km/sec. However, within the moderate impact energy regime, the Modified Tillotson EOS predicted vaporization to begin at an impact velocity that was lower than that predicted by the Tillotson EOS. Had this characterization scheme been used in an actual lethality assessment for an impact velocity between approx. 9 and 18 km/sec, the result would have been conservative since there would have been fewer potentially lethal solid fragments remaining in the debris cloud.

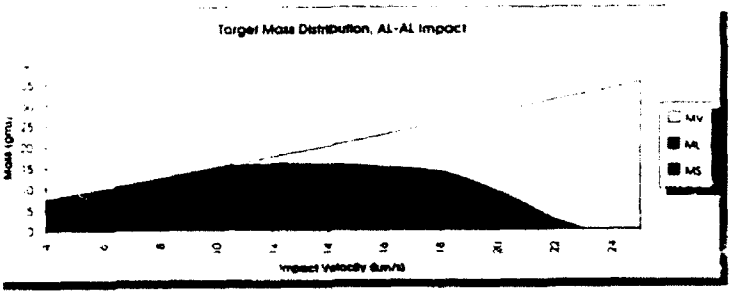
In Figure 9, the total projectile mass remained constant because the projectile length and diameter were fixed in all of the impact scenarios considered. The solid dark region represents the mass of the projectile that was unshocked and therefore was not subjected to melting and/or vaporization. This quantity increased with impact velocity because the speed of the rarefaction wave in the projectile increased at a faster rate than did the speed of the shock wave in the projectile. As the impact velocity increased, the rarefaction wave caught up with the shock wave within a shorter period of time. This in turn increased the amount of the projectile material that was not subject to melting and/or vaporization. The remaining


Debris Cloud Composition, Mie-Grunesen Release
% Vapor
% Liquid
% Solid
Percent


Debris Cloud Composition, Tillotson Release
% Vapor
% Liquid
% Solid
Percent


Debris Cloud Composition, Modified Tillotson Release
% Vapor
% Liquid
% Solid
Percent


Projectile Mass Distribution, AL-AL Impact
MV
ML
MS2
MS1
Impact Velocity (km/s)


Target Mass Distribution, AL-AL Impact
MV
ML
MS
Impact Velocity (km/s)

shaded areas in Figure 9 show the amounts of the shocked and released projectile material in each of the three matter states as the impact velocity increased from 4 to 25 km/sec.

Figure 10 shows that the amount of target material in the debris cloud increased as impact velocity increases due to the growth in target hole size that accompanied an increase in impact velocity. For the projectile and target geometries considered, all of the target material was shocked and released. Hence, there is no solid dark area in Figure 10, only the three lighter-shaded areas which show the amounts of shocked and released target material in each of the three states of matter.

SUMMARY AND RECOMMENDATIONS

A robust lethality assessment methodology must include the effects of discrete particle impacts as well as the response of the target to impulsive debris cloud loadings. A first-order accurate scheme has been implemented to determine the amount of material in each of the three states of matter in a debris cloud created by a hypervelocity impact on a thin target. A modified version of the Tillotson EOS was used to calculate the residual energy in the projectile and target materials upon release from their respective shocked states. Elementary thermodynamic principles were used to determine the percentages of shocked and released projectile and target materials that were melted and/or vaporized during the release process. Using assumed projectile and target geometries, these percentages were then used to calculate the mass of the projectile and target materials in solid, liquid, and gaseous form. Based on the work completed thus far, the following recommendations are offered for continuing the development of a lethality assessment model that would be applicable in impact scenarios where material melt and/or vaporization can be expected to occur.

Once the masses of the solid, liquid, and vaporous debris cloud constituents have been determined, the next step in the characterization of the debris clouds created in a hypervelocity impact would be to determine the nature of the debris cloud solid fragment population. This includes calculating the number of projectile and target material fragments, as well as their sizes, speeds, and trajectories. In addition to the fragmentation models in FATEPEN, PEN-4, and KAPP-II, the fragmentation models developed by Grady, et.al. [27-29] can be used to predict the number of fragments that would result from a KEW impact. The predictions of the various fragmentation models can be compared against one another and against available experimental data to determine which fragmentation model is best suited for use in a lethality assessment methodology.

After a satisfactory first-order accurate procedure that characterizes debris cloud composition is completed, the accuracy of the procedure needs to be improved. This includes modifying the methods presented herein to include the impact of non-monolithic projectiles that are more representative of actual KEW geometries and the impact of yawed and/or obliquely incident projectiles. The method of calculating the percentages of projectile and target material in the three states of matter should also be replaced with a more rigorous thermodynamic procedure.

A shock wave attenuation procedure [30,31] should be implemented to obtain more accurate mass values for the material that is melted and/or vaporized in a high speed KEW impact. Such a procedure will result in a residual energy profile along the length of the projectile and through the thickness of the target. Energy levels at various positions can then be compared to energy levels necessary to begin material melt or vaporization.

An impulsive loading algorithm for the target should be developed to account for the effects of the non-solid debris cloud constituents as well as the solid non-perforating debris cloud fragments. This effort requires as input the masses and velocities of the non-solid debris cloud materials, the area of the inner wall over which the impulsive loading is applied, and the geometric and material properties of the inner wall, including the spacing between the outer and inner walls and the orientation of the inner wall with respect to that of the outer wall. Issues to be addressed include whether the impacts of the target and projectile debris cloud materials need to be considered separately or can be considered simultaneously, whether the effects of the molten and vaporous debris cloud components need to be considered separately or can be combined, and how to account for the decreasing time of the load application and the increasing area over which it is applied as the initial impact velocity increases.

The impulsive loading algorithm can be validated at velocities attainable using existing hypervelocity launchers by comparing the predictions of the algorithm with available impact test data. The algorithm can be modified if necessary until a satisfactory level of accuracy is reached. It can then be combined with the debris cloud characterization scheme and a suitable fragmentation model to yield an improved, robust lethality assessment method for high speed KEW impacts.

REFERENCES

1. Anderson, C.E., Trucano, T.G., and Mullin, S.A., "Debris Cloud Dynamics", Int. J. Impact Engng., Vol. 9, No. 1, pp. 89-113, 1990.

2. Hopkins, A.K., Swift, H.F., and Lee, T.W., "Material Phase Transformations Effects Upon Performance of Spaced Bumper Systems", J. Spacecraft Rockets, Vol. 9, No. 5, pp. 342-345, 1972.

3. Idzorek, G.C., Keaton, P.W., Stradling, G.L., Callopy, M.T., Curling, H.L., and McColl, D.B., "Data Acquistion System for a Hypervelocity-Micro-

particle-Impacts Laboratory", Int. J. Impact Engng., Vol. 10, pp. 261-270, 1990.

4. Iglseder, H., and Idenbergs, E., "Crater Morphology at Impact Velocities Between 8 and 17 km/sec", Int. J. Impact Engng., Vol. 10, pp. 271-280, 1990.

5. Osher, J., Gathers, R., Chau, H., Lee, R., Pomykal, G., and Weingart, R., "Hypervelocity Acceleration and Impact Experiments with the LLNL Electric Guns", Int. J. Impact Engng., Vol. 10, pp. 439-452, 1990.

6. Yatteau, J.D., High Velocity Multiple Plate Penetration Model, NSWC-TR-82-123, Dahlgren, Virginia, 1982.

7. Yatteau, J.D., Modifications to Program FATE - Fragment Residual Mass Calculations, Final Report, Denver Research Institute, University of Denver, Denver, Colorado, 1983.

8. Yatteau, J.D., Zernow, R.H., and Recht, R.F., Compact Fragment Multiple Plate Penetration Model, Volume I: Model Description, NSWC-TR-91-399, Dahlgren, Virginia, 1991.

9. Yatteau, J.D., Zernow, R.H., and Recht, R.F., Compact Fragment Multiple Plate Penetration Model, Volume II: Computer Code User's Manual, NSWC-TR-91-399, Dahlgren, Virginia, 1991.

10. Penetration Mechanics Handbook for Kinetic Energy Penetrators, 61-JTCG/ME-77-16-Rev-1, 1977.

11. Greer, R., and Hatz, M., KAPP-II user's Manual, Version 1.1, Kaman Sciences Corporation, K92-17U(R), Colorado Springs, Colorado, 1992.

12. Snow, P., KAPP - Kaman Analytical Penetration Program, Kaman Sciences Corporation, K85-7U(R), Colorado Springs, Colorado, 1985.

13. Cohen, L., Kaman New Analytical Penetration Program (KNAPP) Space-Based Interceptor Modelling Effort, AFATL-TR-90-02, Eglin AFB, Florida, 1990.

14. Lundeberg, J.F, Stern, P.H., Bristow, J.R., Meteoroid Protection for Spacecraft Structures, NASA-CR-54201, Washington, D.C., 1965.

15. Burch,. G.T., Multi-Plate Damage Study, AFATL-TR-67-116, Eglin AFB, Florida, 1967.

16. Henderson, B.J., and Zimmerschied, A.B., Very High Velocity Penetration Model, NSWC-TR-83-189, Dahlgren, Virginia, 1983.

17. Bjorkman, M.D., Geiger, J.D., and Wilhelm, E.E., Space Station Integrated Wall Design and Penetration Damage Control, Task 3: Theoretical Analysis of Penetration Mechanics, Boeing Aerospace Corporation, Final Report, Contract NAS8-36426, Seattle, Washington, 1987.

18. Swift, H.F., Bamford, R., and Chen, R., "Designing Space Vehicle Shields for Meteoroid Protection: A New Analysis", Adv. Space Research, Vol. 2, No.

12, 1983, pp 219-234.

19. Richardson, A.J., "Theoretical Penetration Mechanics of Multi-SHeet Structures Based on Discrete Debris Particle Modelling", J. Spacecraft, Vol. 7, No. 4, 1970, pp 486-489.

20. Swift, H.F., "On Predicting Projectile Breakup During Thin Plate Impact", Int. J. Impact Engng., Vol. 10, 1990, pp. 579-585.

21. Grady, D.E., and Passman, S.L., "Stability and Fragmentation of Ejecta in Hypervelocity Impact", Int. J. Impact Engng., Vol. 10, 1990, pp. 197-212.

22. Lawrence, R.J., "A Simple Model for the Optimization of Stand-Off Hypervelocity Particle Shields", Int. J. Impact Engng., Vol. 5, 1987, pp. 451-461.

23. Rice, M.H., McQueen, R.G., and Walsh, J.M, "Compression of Solids by Strong Shock Waves", Solid State Physics, Vol. 6, Seitz, F. and Turnbull, D., eds., Academic Press, New York, 1958.

24. Tillotson, J.H., "Metallic Equations of State for Hypervelocity Impact", General Dynamics, General Atomic Division, Report No. GA-3216, 1962.

25. Mullin, S.A., Littlefield, D.L., Anderson, C.A., Velocity Scaling for Lethality Applications, Southwest Research Institute, Final Report, Project No. 06-4438, 1992.

26. Maiden, C.J., Gehring, J.W., and McMillan, A.R., Investigation of Fundamental Mechanism of Damage to Thin Targets by Hypervelocity Projectiles, General Motors Defense Research Laboratory, TR-63-225, Santa Barbara, California, 1963.

27. Grady, D.E., "Local Inertial Effects in Dynamic Fragmentation", J. Appl. Phys., Vol. 53, No. 1, 1982, pp. 322-325.

28. Grady, D.E., "Fragmentation of Solids Under Impulsive Stress Loading", J. Geophys. Res., Vol. 86, No. B2, 1981, pp. 1047-1054.

29. Grady, D.E., and Kipp, M.E., "Geometric Statistics and Dynamic Fragmentation", J. Appl. Phys., Vol. 58, No. 3, 1985, pp. 1210-1222.

30. Fowles, G.R., "Attenutation of the Shock Wave Produced in a Solid by a Flying Plate", J. Appl. Phys., Vol. 31, No. 4, 1960, pp. 655-661.

31. Cohen, L., "Integrated Technology Support for Debris Cloud Material State Modelling", Science Applications International Corporation, Tech. Inf. Memo. WU4/TIM92-2, Shalimar, Florida, 1992.

QUANTUM MECHANICAL INVESTIGATIONS OF MOLECULAR STRUCTURE AND CONFORMATION IN PERFLUOROPOLYALKYLETHERS

Martin Schwartz
Professor
Department of Chemistry

University of North Texas
215 W. Sycamore
Denton, TX 76203

Final Report for:
AFOSR Summer Research Program
Wright Laboratory

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

September 1992

QUANTUM MECHANICAL INVESTIGATIONS OF MOLECULAR STRUCTURE AND CONFORMATION IN PERFLUOROPOLYALKYLETHERS

Martin Schwartz
Professor
Department of Chemistry
University of North Texas

Abstract

Molecular Orbital calculations have been performed on perfluoroethylmethyl ether [PFEME] and perfluorodimethoxymethane [PFDMOM] in order to determine the geometries and energies of their equilibrium and transition state structures.

In PFEME it was found that the CCOC skeleton in the "trans" conformer is twisted by 17° from 180°, and there is a similar rotation of the perfluoromethoxy fluorines about the terminal CC bond, giving the molecule an overall helical structure. It was also determined that the energy of the gauche conformers (relative to the trans form) is unusually high,resulting in virtually 100% trans conformers at normal temperatures.

The structure of "trans" PFDMOM was also found to be helical with a twist of both the COCO skeletal and FCOC terminal dihedral angles from 180°. It was further observed that rotation about both of the internal CO bonds is much more facile (low barriers and roughly equal equilibrium energies) than the equivalent internal rotation in PFEME.

These results indicate that the rigidity and, hence, viscosity of linear perfluoropolyalkylethers should rise with an increasing proportion of -OCCO- linkages in the chain, a conclusion that is in agreement with experimental observations. These quantum mechanical data will be utilized in future molecular dynamics simulations of the molecular motions and bulk properties of PFPAE liquids.

QUANTUM MECHANICAL INVESTIGATIONS OF MOLECULAR STRUCTURE AND CONFORMATION IN PERFLUOROPOLYALKYLETHERS

Martin Schwartz

I. INTRODUCTION

Perfluoropolyalkylether (PFPAE) fluids possess the viscoelastic, thermal and lubricity properties necessary to serve as effective, stable liquid phase lubricants.[1,2] No currently available commercial PFPAE lubricants, however, are capable of operation at the temperature extremes and oxidative conditions required for lubrication of high performance gas turbine jet engines.

The viscoelastic properties of polymer fluids such as the PFPAE's are, of course, intimately connected to the chain flexibility in these systems which is, in turn, dependent upon the potential energy barriers to internal rotation about single bonds in the polymer. This summer, Dr. Harvey Paige (of the Materials Directorate at Wright Laboratory) and I, in collaboration with Ms. Christine Williamson of my research group, have been engaged in two computational chemistry research projects on PFPAE homolog molecules, designed to obtain a better understanding of chain mobility in the perfluoroether polymers.

In the first investigation, we studied the torsional potential to internal rotation about the central C-O bond in perfluoroethylmethyl ether (PFEME), $CF_3CF_2OCF_3$. The results have been compared to those from earlier calculations on perfluorobutane[3,4] (PFB) and ethylmethyl ether[5] (EME). We are in the process of completing a second study of internal rotation about about the two central C-O bonds in perfluorodimethoxymethane (PFDMOM), $CF_3OCF_2OCF_3$. These results have been compared both to those reported for dimethoxymethane[6] (DMOM) and to our work on PFEME.

II. CALCULATIONS

Ab initio molecular orbital calculations were performed on a Cray X-MP/216 computer using the *Gaussian 90* MO program.[7]

A. PFEME

Equilibrium and saddle point geometries were gradient optimized[8] at the SCF level using the 6-31G(d)[9,10] basis set. Hartree-Fock energies were also obtained with the 6-311G(d)[10,11] basis, using the 6-31G(d) geometries. In addition, single point second order Møller-Plesset[12] (MP2) energy calculations were performed with both basis sets.

Vibrational frequencies were calculated for all conformers using the 6-31G(d) basis set. Transition state geometries were confirmed by the presence of a single imaginary frequency.

B. PFDMOM

Since the skeleton conformation of PFDMOM depends upon the dihedral angles of both of the central C-O bonds, the calculation was performed in two parts. First, a grid of energies was established with all combinations of ϕ_1 and ϕ_2 ranging from -180° to +180° in increments of 30°. Because of the large number of independent conformations (thirteen), the geometry optimizations were performed with the 6-31G*(C,O) basis set, which is the 6-31G basis[9,10] augmented by polarization functions on the carbon and oxygen atoms. This was followed by single point energy calculations with the 6-31G* basis using the 6-31G*(C,O) geometries; for brevity, the energies from these calculations are not contained in this report.

Second, once the approximate positions of the potential energy minima and barriers were located, a number of the precise stationary states were obtained using the 6-31G* basis set; these results are contained in Table 4. Additional single point MP2 energies were obtained for all grid points and stationary states. However, again for brevity, these are not shown.

RESULTS AND DISCUSSION

A. PFEME

Geometries

Selected geometric parameters for all equilibrium and transition state conformations of PFEME are displayed in Table 1. In the table and following discussion, the atoms are numbered as shown in Figure 1A.

The most striking result for the optimized geometries is that the skeletal dihedral angle, $\phi(C_1C_2OC_3)$, of the equilibrium trans conformer is twisted from 180°, by approximately 17°. The perfluoromethoxy group in this rotamer is similarly twisted [$\phi(F_{3A}C_3OC_2)$=162°] from a purely trans configuration, giving an overall helical structure to the molecule; this rotamer is labeled as "Twist-Trans" , TT, in Table 1. Similarly, the stable gauche conformer is labelled G, and the three transition states are (T)* [ϕ=180°], (GT)* [180° < ϕ < 60°] and (GG′)* [ϕ=0°].

Dixon[3,4] and Van Catledge[3] have earlier reported that the structure of the trans conformation of PFB is helical, characterized by a twist of the carbon skeleton, ϕ(CCCC)=164.6°,[4] and rotation of the fluorines about both terminal CC bonds, ϕ(FCCC)=170.2°.[4] They have suggested that the calculated helicity in PFB, and in other perfluoroalkanes, may result from a tendency to lower the repulsive interactions between CF dipoles on the 1 and 3 carbons, which are precisely parallel in the perfectly trans structure.

Most significantly, while fluorines on the perfluoromethoxy carbon of PFEME are twisted from 180° in the TT rotamer (*vide supra*), there is virtually no rotation of the fluorines on the C_1 carbon about the C_1C_2 axis [$\phi(F_{1A}C_1C_2O)$=179.1°]. This result lends supportive evidence for Dixon's explanation of the helicity in perfluoroalkanes since, as seen clearly in Figure 1A, 1,3 CF bond dipole repulsions are absent in trans PFEME.

Pacansky, *et al.*[13] have reported the geometry of the TT rotamer of PFEME determined with the 3-21G[14] basis set. Their CC bond length is shorter (by 0.03 Å), and their calculated CO and CF bonds are longer (by 0.01-0.02 and 0.01-0.03 Å, respectively) than those found here. Bond angles agree to within 1-2° on average. The calculated twist of the skeleton using the smaller basis

Table 1. Conformation Dependence of Selected Geometric Parameters in PFEME[a,b]

Parameter	TT	(T)*	(GT)*	G	(GG')*
$R(C_1C_2)$	1.529	1.530	1.537	1.535	1.543
$R(C_2O)$	1.358	1.358	1.365	1.364	1.363
$R(C_3O)$	1.358	1.359	1.355	1.355	1.351
$R(C_1F_{1A})$	1.311	1.311	1.311	1.310	1.310
$R(C_1F_{1B})$	1.310	1.310	1.310	1.310	1.312
$R(C_1F_{1C})$	1.310	1.310	1.311	1.310	1.309
$R(C_2F_{2A})$	1.319	1.320	1.314	1.314	1.320
$R(C_2F_{2B})$	1.320	1.319	1.318	1.319	1.318
$R(C_3F_{3A})$	1.301	1.301	1.301	1.308	1.301
$R(C_3F_{3B})$	1.309	1.308	1.307	1.313	1.309
$R(C_3F_{3C})$	1.307	1.308	1.310	1.301	1.308
$<(C_1C_2O)$	107.9	107.5	114.4	115.1	119.8
$<(C_2OC_3)$	121.4	122.5	125.7	125.2	130.3
$<(C_2C_1F_{1A})$	109.1	109.0	108.9	109.0	107.9
$<(OC_3F_{3A})$	107.1	106.6	106.5	106.9	106.6
$\phi(C_1C_2OC_3)$	162.8	180.0	76.8	61.7	0.0
$\phi(F_{1A}C_1C_2O)$	179.1	180.1	165.8	171.1	169.2
$\phi(F_{3A}C_3OC_2)$	162.3	180.0	172.8	160.4	166.1

a) Bond lengths are in angstroms and angles in degrees.
b) See Figure 1A for atom numbering.

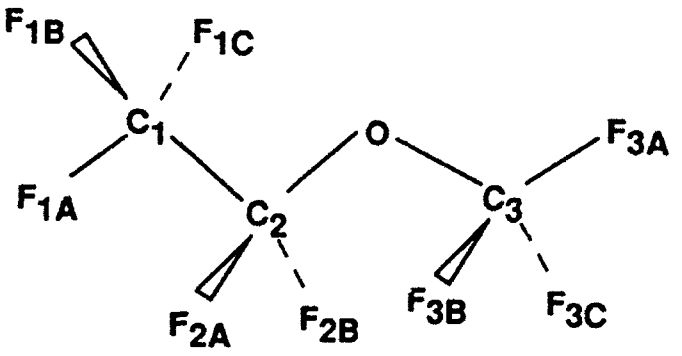


Fig. 1A. Structure and atom numbering in trans PFEME

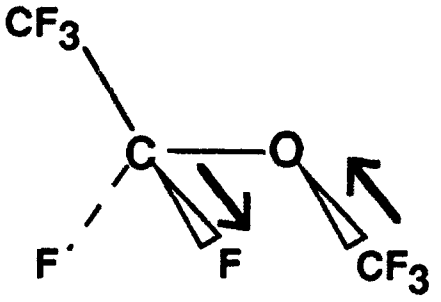


Fig. 1B. Structure of PFEME at the (GT)* transition state

was similar to that obtained here, with $\phi(C_1C_2OC_3)$=164.6°, but the rotation of the perfluoromethoxy group was substantially higher [$\phi(F_{3A}C_3OC_2$=147.6°].[13] Smart and Dixon[15] have also calculated the geometry of the TT equilibrium conformer using a DZ+D_c basis set, which is a double zeta basis with polarization functions on carbon. Their results agree even more closely with those obtained here, particularly for the dihedral angles [$\phi(C_1C_2OC_3)$=160.9° and $\phi(F_{3A}C_3OC_2$=161.2°]. The observed differences, which are relatively small, are expected when comparing geometries calculated with and without polarization functions on the various atoms.

Not surprisingly, since the (T)* energy barrier is quite low (*vide infra*), the geometric parameters of the TT equilibrium conformer and the (T)* saddle point agree closely (Table 1), with the exception of the dihedral angles, which are all 180° in the transition state.

From the table, it is observed that both the COC and CCO angles vary in the order, (T)* ≈ TT < (GT)* ≈ G < (GG')*, reflecting the increased 1,4 atomic repulsions with decreasing CCOC dihedral angle. Similarly, $R(C_1C_2)$ and $R(C_2O)$ exhibit modest increases with diminishing dihedral angle; this trend is not, however, observed for $R(C_3O)$, which is slightly longer in the TT and (T*) rotamers.

Significantly, it may be seen in the last column of the table that, unlike in PFB,[3,4] the terminal fluorines remain twisted from 180° in the (GG')* conformation of PFEME. This may reflect greater F-F repulsions in the syn conformation of the ether, resulting from the shorter CO bond lengths.

It is of interest to compare the geometry of PFEME with that reported earlier for ethylmethyl ether (EME),[5] using the same [6-31G(d)] basis set. Both of the CO bond lengths in the fluorinated ether are substantially shorter than in EME, in which $R(C_2O)$=1.417 Å and $R(C_3O)$=1.390 Å in the trans conformer; in contrast, the C-C bond lengths are approximately the same [$R(C_1C_2)$=1.516 Å in EME]. Too, the COC bond angle is markedly greater in PFEME [$<(C_2OC_3)$=114.2° in trans EME], whereas the CCO bond angles are

approximately equal in the two species [$<(C_1C_2O)$=108.6° in EME].

The shortening of the CO bond lengths and increase in the COC angle in the fluoroether may be explained on the basis of relative bond polarities. The calculated difference between the carbon and oxygen Mulliken charges [q(C)-q(O)] is 1.7-2.1 in PFEME,[16] whereas it is only 0.6-0.8 in EME.[17] Therefore, one expects a more ionic and, hence, shorter CO bond in the fluorinated compound. Consequently, the COC bond angle should increase to moderate the otherwise enhanced electrostatic and/or van der Waal's interactions between fluorines on the 2 and 4 carbons. The same trends in bond lengths and angles and in Mulliken charges are found in calculations on dimethyl ether and its fluorinated analogue.[18]

Energies

Total HF and MP2 energies (in au's) of the five equilibrium and saddle point conformations of PFEME, calculated with the two basis sets (using the 6-31G(d) geometries) are presented in Table 2A. Energies (in kcal/mol) relative to the TT rotamer are given in 2B. For comparison, relative conformational energies for EME[5] and PFB[4,19] are also contained in the latter half of the table.

One sees from the table that the (T)* transition state is only slightly higher in energy than the TT minimum, with $\Delta E \approx 0.3$-0.4 kcal/mol at the SCF level, and $\Delta E \approx 0.5$-0.6 cal/mol with correlation energy corrections.

One observes, also, that relative energies of the G, (GT)* and (GG')* conformations are all lower at the MP2 than the HF level, within a given basis set. This trend may be explained by analysis of Table 2A, from which it is found that, with the 6-31G(d) basis for example, the correlation energy correction (in kcal/mol) varies in the order, (T)* [-1190.2 < TT [-1190.4] < (GT)* [-1191.0] < G [-1191.4] < (GG')* [-1191.5]. It is reasonable that inclusion of electron correlation will preferentially stabilize the more structurally congested conformations [lower ϕ(CCOC)] in which the electronic repulsions are greatest.

Table 2. Calculated Conformational Energies in PFEME[a]

Method	TT	(T)*	(GT)*	G	(GG')*
A. Total Energies (hartrees)					
HF/6-31G(d)	-984.007 346	-984.006 812	-984.002 264	-984.002 692	-983.995 404
HF/6-311G(d)	-984.268 451	-984.267 804	-984.262 704	-984.263 112	-984.255 622
MP2/6-31G(d)	-985.904 547	-985.903 717	-985.900 419	-985.901 509	-985.894 362
MP2/6-311G(d)	-986.470 460	-986.469 558	-986.465 422	-986.466 529	-986.458 868
B. Relative Energies (kcal/mol)					
PFEME/HF/6-31G(d)	0.00	+0.34	+3.19	+2.92	+7.49
PFEME/HF/6-311G(d)	0.00	+0.41	+3.61	+3.35	+8.05
PFEME/MP2/6-31G(d)	0.00	+0.52	+2.59	+1.91	+6.39
PFEME/MP2/6-311G(d)	0.00	+0.57	+3.16	+2.47	+7.27
EME/HF/6-31G(d)[b]	--	0.00	+2.56	+1.67	+6.84
PFB/HF/DZ+P[c]	0.00	+0.15	+2.35	+1.47	+8.30
EME/MP2/6-31G(d)[b]	--	0.00	+2.67	+1.40	+7.00
PFB/MP2/DZ+P[c]	0.00	+0.38	+2.41	+1.48	+8.02

a) All energies were calculated using the HF/6-31G(d) geometries.
b) From Ref. 5.
c) From Ref. 4.

The principal effect of increasing the size of the basis is to increase relative energies of the G, (GT)* amd (GG') conformations. Again, this trend arises from analysis of Table 2A, where it is found that the larger basis sets preferentially stabilize the TT and (T)* states in comparison to the above conformations. Thus, the effects of increasing basis size and introducing electron correlation tend to offset one another.

A principal focus of this investigation is to perform a comparison of the torsional potential for rotation about the central bond in PFEME to those obtained earlier in the fluoroalkane and in the nonfluorinated ether. The three energy curves, calculated at the HF/6-31G(d) level for PFEME and EME and at the equivalent HF/DZ+P[19] level for PFB, are shown superposed (displaced from one another for clarity) in Figure 2. As noted, the SCF and MP2 energies of the stationary state conformers of the latter two molecules are also tabulated at the bottom of Table 2B.

One observes from both figure and table that the rotational potential of PFEME is markedly different from that of either of the other molecules in the vicinity of the G and (GT)* conformations. Most striking is that the dihedral angle of the (GT)* barrier is shifted by over 40° below the nominal angle of 120° (see also Table 1). In order to verify this highly unusual behavior, we have performed additional HF/6-31G(d) optimizations at various fixed values of $\phi(C_1C_2OC_3)$ ranging from 70° to 120°. The resultant energies, plotted also in Figure 2, confirm the position of the (GT)* transition state.

A possible explanation for the large shift in the torsional angle of this transition state may be found by examination of the structure of PFEME at $\phi(C_1C_2OC_3)$=120° (Figure 1B). One observes from the figure that, at this angle, the C_3O and one of the C_2F bond dipoles are precisely antiparallel, which would lead to a stabilization of this configuration, not expected in either PFB or EME, and thus result in a shift of the position of the energy maximum.

A very important feature of the potential energy diagram of PFEME is that the relative energy of its G equilibrium conformers is greater than that

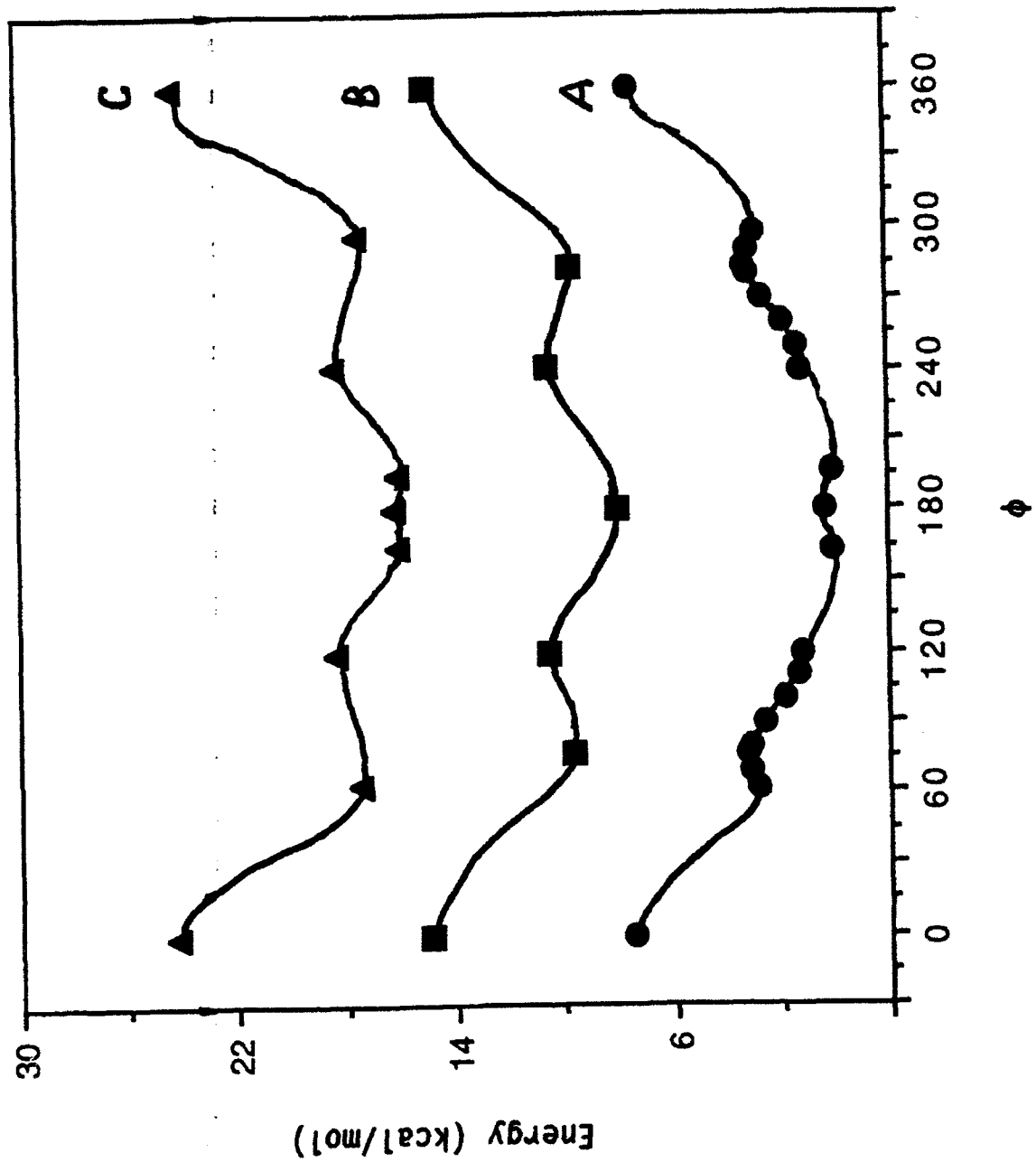


Fig. 2. The torsional potential in (A) PFEME - Circles; (B) EME - Squares (from ref. 5); (C) PFB - Triangles (from ref. 4).

The potential curves for EME and PFB are displaced upward by 8 and 16 kcal/mol, respectively, for clarity of presentation.

in either of the two other molecules. The explanation for this difference may reside in the closer approach distance between terminal fluorine atoms, resulting from the lower CO and greater CF (compared to CH) bond lengths, which would destabilize the gauche conformation in PFEME.

The consequence of the unusually high energy of the two G conformations of PFEME is that, from the Boltzmann distribution, one would have more than 98% of the molecules in the T conformation. Thus, if a PFPAE has numerous $-O-CF_2CF_2-O-CF_2-O-$ linkages, its conformation in the liquid would tend to be elongated (higher percentage of trans bonds) and stiff (since only one of the three conformations about the CO bond is thermally accessible).

As seen clearly in Figure 2, the overall effect of the displaced position of the (GT)* transition state and the high energy of the G conformer is to result in a very shallow, narrow energy minimum for the gauche conformation of PFEME. From Table 2B, the energy barrier, E[(GT)*]-E[G], varies from only 0.26 to 0.69 kcal/mol, dependent upon the level and method of calculation.

Vibrational frequencies for all five stationary states of PFEME were calculated using the 6-31G(d) basis set and are displayed, for reference, in Table 3; these values have been multiplied by the normal 0.90 scale factor to account for the effects of vibrational anharmonicity and electron correlation.

B. PFDMOM

Selected geometric parameters (and energies) of various of the stationary states in PFDMOM are displayed in Table 4. As observed in the above comparison of PFEME and EME, the CO bond lengths are shorter (by approximately 0.03 Å) and COC bond angles are greater (by roughly 8°) than those reported by Wiberg and Murcko[6] in their investigation of dimethoxymethane (DMOM). Again, these differences can be attributed directly to the greater bond ionicity and steric repulsions in the perfluorinated ether.

Both bond lengths and angles are relatively insensitive to the molecular conformation with the exception of $<(O_1C_2O_2)$, which rises if either or both of

Table 3. Vibrational Frequencies in PFEME[a,b]

Vib. No.	TT	(T)*	(GT)*	G	(GG')*
1	35	25i	28i	41	74i
2	54	57	41	53	48
3	83	109	113	102	89
4	122	117	174	158	148
5	202	211	189	197	199
6	215	218	221	213	218
7	301	301	308	312	318
8	327	323	335	332	338
9	343	344	344	342	346
10	364	364	363	365	366
11	419	420	440	434	421
12	507	515	457	458	489
13	515	515	514	516	509
14	542	541	546	547	539
15	587	588	585	585	590
16	612	614	606	608	615
17	644	649	650	647	617
18	669	667	706	709	682
19	745	745	724	718	717
20	837	834	793	789	776
21	905	905	929	932	941
22	1119	1121	1113	1114	1120
23	1224	1225	1232	1231	1238
24	1244	1240	1259	1261	1251
25	1280	1281	1271	1269	1262
26	1288	1288	1288	1284	1268
27	1291	1290	1290	1291	1292
28	1302	1301	1301	1305	1306
29	1343	1342	1349	1345	1332
30	1463	1463	1434	1437	1430

a) Frequencies are in units of $cm.^{-1}$
b) Calculated frequencies have been scaled by the factor 0.9.

Table 4. Stationary State Geometries and Energies in PFDMOM.[a]

	Nominal Dihedral Angles						
Param.	180,180(A)	180,180(B)	180,120	180,60(A)	180,60(B)	180,0	60,60
$R(C_1O_1)$	1.358	1.357	1.357	1.359	1.357	1.360	1.356
$R(C_1O_2)$	1.354	1.355	1.359	1.361	1.361	1.358	1.362
$R(C_2O_2)$	1.354	1.355	1.357	1.354	1.355	1.356	1.362
$R(C_3O_2)$	1.358	1.357	1.355	1.358	1.357	1.356	1.356
$<(O_1C_2O_2)$	105.0	104.8	106.4	109.7	108.6	110.8	114.2
$<(C_1O_1C_2)$	121.1	121.2	121.5	121.4	121.4	121.4	122.6
$<(C_2O_2C_3)$	121.1	121.2	122.7	122.2	121.8	124.9	122.6
$\phi(F_{1A}C_1O_1C_2)$	163.3	196.5	196.5	197.0	163.0	163.4	188.3
$\phi(F_{3A}C_3O_2C_2)$	163.2	163.5	179.2	157.7	195.8	178.5	188.3
$\phi(C_1O_1C_2O_2)$	160.2	195.7	191.7	196.4	161.6	161.1	68.6
$\phi(O_1C_2O_2C_3)$	160.2	164.4	123.5	47.7	84.8	2.4	68.6
E(au)+1058	-0.885960	-0.885408	-0.883838	-0.885956	-0.885426	-0.883248	-0.884607
ΔE(kcal/mol)[b]	0.00	0.35	1.33	0.00	0.34	1.70	0.85

a) Bond lengths are in anstroms and angles in degrees.
b) Energy relative to the 180,180(A) stationary state.

the torsional angles are 60° or 0°. This effect is due to repulsions between the two terminal perfluoromethyl groups.

It is most interesting that, as found for PFEME, the Trans-Trans form (nominally with $\phi_1=\phi_2=180°$) of PFDMOM is actually helical in character, with a rotation of the two terminal CF_3 groups [ϕ(FCOC)$\neq$180°] as well as a twist of both of the COCO dihedral angles from 180°, the value found in DMOM.[6] Significantly, there are two separate helical conformations for the Trans-Trans conformer [180,180(A) and 180,180(B)], which are close to one another in energy. Analysis of the two structures reveals that in the second configuration, there is an actual reversal in the direction of the helix between the two ends of the molecule. This helix reversal is also observed in the Trans-Gauche [180,60] conformation which, too, has two distinct conformations of similar energy (Table 4).

In Figure 3 is shown the potential energy curve for rotation of the second C-O dihedral angle in PDMOM while the first torsion is held constant at (A) $\phi_1=60°$ and (B) $\phi_1=180°$, respectively. One observes quite clearly that when the first C-O bond is trans (180°), then the second bond is free to assume the trans or either of the gauche conformations with equal facility [$\Delta E_{GT}\approx 0$]. Even when $\phi_1=60°$, then ϕ_2 can be either 60° or 180°, although $\phi_2=-60°$ (=300°) is thermally forbidden [$\Delta E\approx 8$ kcal/mol]. The cause of the very high energy when $\phi_1=60°$ and $\phi_2=-60°$ is that in this conformation, the two terminal perfluoromethyl groups lie on the same side of the central O-C-O plane, which causes severe steric repulsions in the molecule.

A comparison of the potential curves for PFEME and PFDMOM reveals that, whereas only one of the three equilibrium conformations is populated in the former molecule, the CO bond conformation in the latter species can be in either two or all three of the equilibrium states. This result has substantial ramifications on the conformational flexibility of perfluoropolyalkylether polymers (*vide infra*).

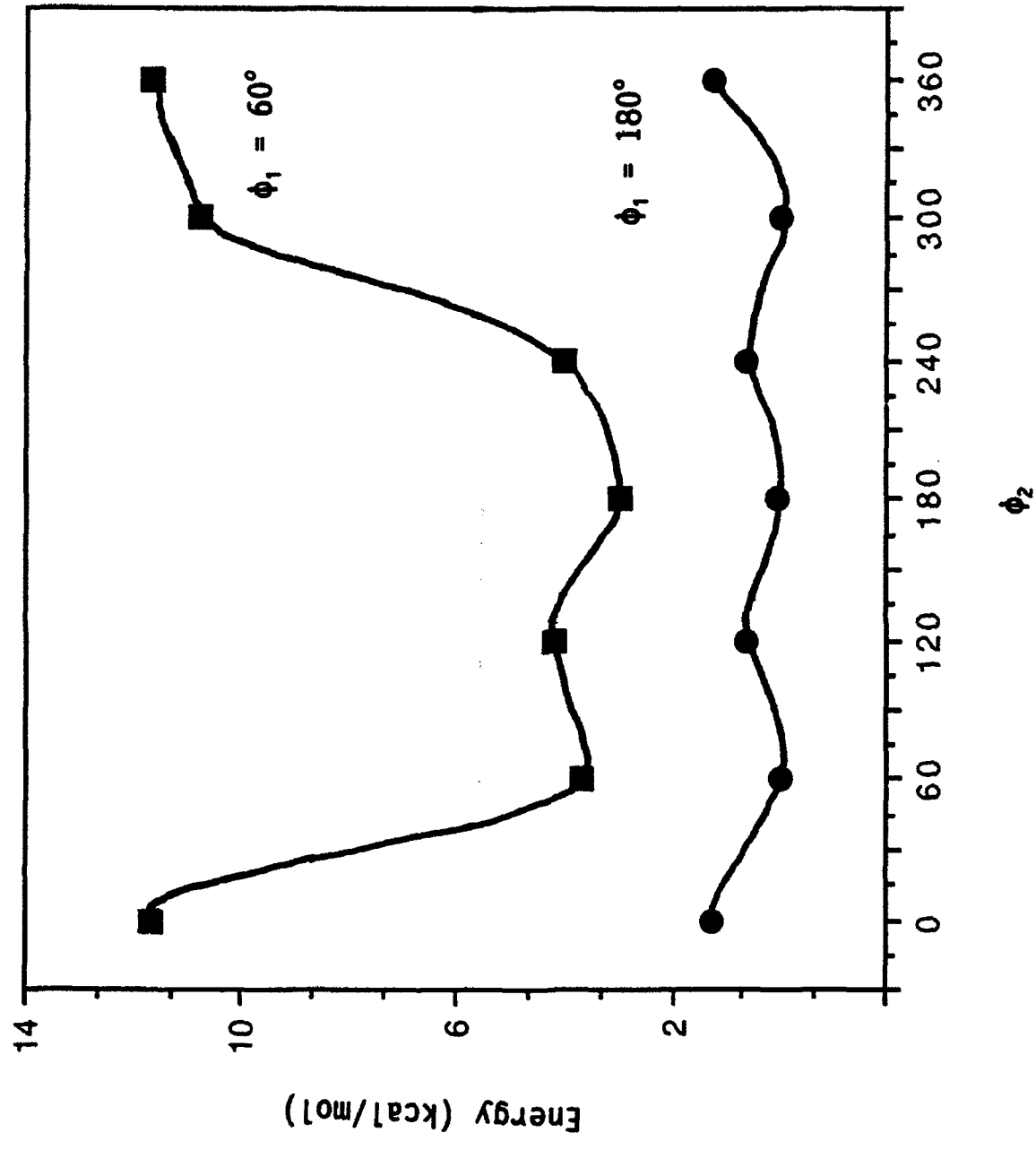


Fig. 3. The torsional potential for the rotation of ϕ_2 with (A) $\phi_1 = 60°$ and (B) $\phi_1 = 180°$.

The potential curve for (A) has been displaced upward by 3 kcal/mol for clarity of presentation.

IV. SUMMARY AND CONCLUSIONS

Molecular geometries and energies of all equilibrium and transition state conformations of perfluoroethylmethyl ether were determined by *ab initio* molecular orbital calculations using the 6-31G(d) basis set. The CCOC skeleton in the "trans" conformer is twisted by 17° from 180°. There is a similar rotation of the perfluoromethoxy fluorines about the terminal CC bond (by 18°), but no twisting of fluorines on the other terminus of the molecule. These results indicate that, as suggested earlier for perfluorobutane, the helical structure of the trans conformer in PFEME is induced by dipolar repulsions between CF bonds on alternant carbons.

The energy of the gauche conformers of PFEME (relative to the twisted trans structure) is unusually high [2.9 kcal/mol], from which it may be determined that one would have virtually 100% trans molecules at thermally accessible temperatures.

The structures and energies of perfluorodimethoxymethane [PFDMOM] were determined as a function of the dihedral angles about both of the central C-O bonds. As found for PFEME, the CO bonds were shorter and the COC angles greater than in the non-fluorinated species. It was also determined that, here too, the molecules adopt a helical configuration in the nominally Trans-Trans state.

It was also determined that in PFDMOM, if one of the two skeletal dihedral angles is 180° (trans), then the second angle can adopt any of the three equilibrium values with equal facility; if the first angle is 60°, then the second CO torsional angle can still adopt two of the three values. Thus, this molecule is substantially more flexible in PFEME, in which all of the CO bonds are trans.

The most significant conclusion is that if a perfluoropolyalkylether polymer has a high percentage of $-O-CF_2CF_2-O-$ linkages, one expects that it will be very rigid. If, in contrast, the majority of the linkages are of the type $-O-CF_2-O-$, then the polymer will exhibit a much higher degree of

flexibility and, hence, lower viscosity. These conclusions can be used to explain the general empirical observation that, for a given molecular weight, the viscosity of a PFPAE fluid increases with rising C:O ratio, since this ratio is a measure of the relative number of the former type of linkage.

VI. FUTURE INVESTIGATIONS

We intend to extend the above studies to the molecule, perfluoro-1,2-dimethoxyethane, $CF_3OCF_2CF_2OCF_3$. This will permit us both to verify the results obtained this summer on PFEME and to determine the potential surface for rotation about C-C bonds in linear PFPAE's. We also wish to model the internal rotation in branched perfluoroethers (such as Krytox). Therefore, we shall study the molecules, $CF(CF_3)_2OCF_3$ and $CF_3OCF_2CF(CF_3)OCF_3$.

We plan to use the results of these quantum mechanical investigations to obtain accurate potential energies for bond stretching, bending and internal rotation, which will then be utilized in molecular dynamics simulations of perfluoroether polymers in order to obtain *a priori* predictions of the dependence of bulk properties such as viscosity, density, thermal expansion and isothermal compressibility as a function of molecular structure in PFPAE lubricants.

REFERENCES

1. Snyder, C. E., Jr.; Dolle, R. E., Jr. *ASLE Trans.* **1975**, *19*, 171.
2. Snyder, C. E., Jr.; Gschwender, L. J.; Tamborski, C. *Lubr. Eng.* **1981**, *37*, 344.
3. Dixon, D. A.; Van Catledge, F. A. *Int. J. Supercomput. Applic.* **1988**, *2*, 52.
4. Dixon, D. A. *J. Phys. Chem.* **1992**, *96*, 3698.
5. Tsuzuki, S.; Tanabe, K. *J. Chem. Soc., Faraday Trans.* **1991**, *87*, 3207.
6. Wiberg, K. B.; Murcko, M. A. *J. Am. Chem. Soc.* **1989**, *111*, 4821.

7. *Gaussian 90*, Revision F; Frisch, M. J.; Head-Gordon, M.; Trucks, G. W.; Foresman, J. B.; Schlegel, H. B.; Raghavachari, K.; Robb, M.; Binkley, J. S.; Gonzalez, C.; Defrees, D. J.; Fox, D. J.; Whiteside, R. A.; Seeger, R.; Melius, C. F.; Baker, J.; Martin, R. L.; Kahn, L. R.; Stewart, J. J. P.; Tolpiol, S.; Pople, J. A.; Gaussian, Inc.: Pittsburgh, PA, 1990.

8. Pulay, P. In *Applications of Electronic Structure Theory*; Schaefer, H. F., III, Ed.; Plenum Press: New York, 1977; p 153.

9. (a) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* 1982, *56*, 2257. (b) Hariharan, P. C.; Pople, J. A.; *Theor. Chim. Acta* 1973, *28*, 213.

10. Frisch, M. J.; Pople, J. A.; Binkley, J. S. *J. Chem. Phys.* 1984, *80*, 3265.

11. (a) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* 1980, *72*, 650. (b) McLean, A. D.; Chandler, G. S. *Ibid.* 1980, *72*, 5639.

12. Møller, C.; Plesset, M. S. *Phys. Rev.* 1934, *46*, 618.

13. Pacansky, J.; Miller, M.; Hatton, W.; Liu, B.; Scheiner, A. *J. Am. Chem. Soc.* 1991, *113*, 329.

14. Pietro, W. J.; Francl, M. M.; Hehre, W. J.; Defrees, D. J.; Pople, J. A.; Binkley, J. S. *J. Am. Chem. Soc.* 1982, *104*, 5039, and references contained therein.

15. Smart, B. E.; Dixon, D. A. "Heterolytic C-F Bond Energies and Stabilities of Poly(perfluoroethers," article preprint. **CHANGE** - Harvey, can we get a better reference?

16. Calculated Mulliken charges on the skeletal atoms of PFEME (the TT conformation) are: $q(C_1)$= +1.08; $q(C_2)$= +1.00; q(O)= -0.70; $q(C_3)$= +1.36.

17. Calculated Mulliken charges on the skeletal atoms of EME are: $q(C_1)$= -0.49; $q(C_2)$=0.02; q(O)= -0.60; $q(C_3)$= -0.16. Williamson, C. L.; Marshall, P; Schwartz, M., unpublished results.

18. Williamson, C. L.; Paige, H. L.; Schwartz, M., unpublished results.

19. Conformational energies for PFB, taken from ref. 4 and contained in Table 2B, were obtained by single point calculations (using DZ+D_c geometries) with a DZ+P basis set, which is a double zeta basis with polarization functions on all atoms.

DETECTION AND ADAPTIVE FREQUENCY ESTIMATION FOR DIGITAL MICROWAVE RECEIVERS

Faculty Associate : Arnab K. Shaw
Assistant Professor
Electrical Engineering Department
Wright State University
Dayton, OHIO-45435

Graduate Associate : Stven Nunes
Electrical Engineering Department
California State University, Chico

Final Report for :
Summer Research Program
Wright Laboratory, Avionics Directorate

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

October 1992

Detection and Adaptive Frequency Estimation for Digital Microwave Receivers

Arnab K. Shaw
Assistant Professor
Electrical Engineering Department
Wright State University

Sven Nunes
Graduate Student
Electrical Engineering Department
California State University, Chico

Abstract

Detection of presence of targets and estimation of Angles-of-Arrival (AOA) are two of the important tasks of a digital EW receiver. In this report, the time-domain detection problem and an adaptive frequency/AOA estimation scheme have been studied. For the detection problem, detection thresholds have been derived for square-law detectors in cases of single and multiple observation samples. The adaptive frequency/AOA estimation scheme has been studied to analyze its behavior at various noise and input conditions. Simulation results indicate that the use of higher estimation model order than the actual order improves the bias and variance of the estimates at the cost of longer convergence time.

DETECTION AND ADAPTIVE FREQUENCY ESTIMATION FOR DIGITAL MICROWAVE RECEIVERS

Arnab K. Shaw and Steven Nunes

I. Introduction :

In Electronic Warfare (EW) environments, microwave receivers play a major role in *passive* identification and localization of unknown targets emitting high-frequency electro-magnetic signals. EW signals cover a relatively wide bandwidth, typically in the range of 0.2 to 15 GHz, and existing microwave receivers utilize mostly analog signal processing tools and techniques [1-3]. In fact, there are no EW receivers that process microwave radar signals entirely in the digital domain. With the emergence of increasingly faster and inexpensive digital computers and high-speed A/D converters, it is expected that digital processing of microwave radar signals would most certainly be the way of the future.

For the past several years the Electronic Support Measure (ESM) group at the EW division of the Avionics Directorate, WPAFB, Ohio, has been engaged in the task of research and development of suitable DSP algorithms and architectures for all-digital microwave receivers [4, 5, 12] for USAF use. The major goals of this summer research were to develop and implement various Digital Signal processing algorithms in order to study their effectiveness and usefulness to the existing Digital Receiver Program. The signal processing algorithms that have been studied are :

- Time-Domain Detection of targets in the presence of noise,
- Adaptive Frequency/AOA estimation.

The time-domain detection problem is considered for single and multiple samples. Detection thresholds and Probability of Detection based on Neyman-Pearson Criterion have been derived. For frequency/AOA estimation, the effectiveness of Direct-Adaptive-Frequency-Estimation (DAFE) algorithm [5] has been studied for realistic signal/noise conditions. Some modification of DAFE is also suggested for improving its performance.

This report is arranged as follows : In Section II, the overall problems associated with digital receiver design are discussed. In Section III, the time-domain detection problem is addressed in detail. In Section IV, the performance of the DAFE algorithm is studied. Finally, in Section V some concluding remarks are given with some discussions on possible future work.

II. The Digital Microwave Receiver Design Problem :

The primary task of a microwave receiver is to gather data for sorting of signals and identification of the type of the radar emitting the received signal. Based on these information, jamming, weapon delivery or other options are considered. In order to perform these tasks, the receiver must analyze the received radar pulses and measure or estimate the following six parameters : Angle-of-

Arrival (AOA), Radio Frequency (RF), Time of Arrival (TOA), Pulse Amplitude (PA), Pulse Width (PW) and Polarization (P). These parameters may be useful in more than one stages of receiver operation. For example, AOA, RF, TOA, PW and Polarization data are used for signal sorting, whereas the RF, TOA, PW and P are utilized for source identification purposes. PA threshold may also be used for detecting the presence of any source signal. For jamming or subsequent weapons delivery, all these parameters need to be analyzed properly. Along with the standard requirements of most conventional radars, the EW receiver design problem is complicated by the fact that no prior knowledge about the input signal is available at the receiver. The nature of the problem also requires that measurements and decisions be taken at or near real-time and in an entirely passive mode of operation [1-4].

In the past two decades, many classes of radar and sonar receivers have been converted from conventional analog technology to purely digital or hybrid systems [16], but EW receivers are yet to make that transition. Unlike other sonar/radar receivers [16], there are essentially two major technological limiting factors that have been holding back analog-to-digital transition of EW receivers. Firstly, if A/D converters are used at the operating frequency range, the Nyquist rate would necessitate sampling at the GHZ range and secondly, the digital hardware or firmware must have the capacity to process such high data rate and produce effective results at near real-time. Considering the high data-rate of this application, computationally simpler algorithms for signal detection and estimation must be identified and implemented. It is expected that the time-domain signal detection and the frequency/AOA estimation algorithms performed during the summer project will be useful in achieving the desired computational efficiency.

III. Time-Domain Detection :

Almost all existing AOA/RF estimation algorithms assume that the signal is already present in the observed data. But in the passive mode of operations of EW applications, source signals may not be present at all within the observation window, or the signals may fill only a part of the estimation window. In either case, any frequency estimation algorithm would essentially produce erroneous or noise frequencies because the observed signal would not satisfy the model assumed by the estimation algorithm. Considering the relatively high computational burden, any estimation method should be invoked only when a detection scheme indicates high probability of presence of threat.

Since EW receivers do not have any prior knowledge about the frequency/amplitude/phase of the received signals, conventional *matched filters* can not be used in this case. An obvious solution would be to perform the detection in the frequency-domain, *i.e.*, the presence of targets can be determined by thresholding of FFT-peaks. The frequency-domain approaches are robust but have certain disadvantages in that a decision can be made only after a block of data has been collected. Furthermore, a lot of computational power may be wasted if FFT is taken continuously, even when

no target is present. Instead, we plan to incorporate a time-domain detection scheme that can detect targets in real-time using a single observation or a small number of samples. Once a preliminary decision is taken, FFT or more sophisticated frequency/AOA estimation algorithm can be invoked, if desired.

III.1 : Signal and Noise Model

Microwave radars signals can be modeled as,

$$x(n) = A\cos(\omega_c k + \theta) + n(k) \tag{1a}$$

$$= A\cos(\omega_c k)\cos\theta - A\sin(\omega_c k)\sin\theta + n_I(k)\cos(\omega_c k) + n_Q(k)\sin(\omega_c k) \tag{1b}$$

where, $n(k)$ denotes narrowband noise samples. To perform the time-domain detection, the received real data is first converted into a complex analytic signal. This is achieved by passing the real signal through a Hilbert Transformer to form the in-phase (I) and quadrature (Q) components of the complex analytic signal. When no signal is present, the I and Q components may be represented as,

$$X_I(k) = n_I(k) \tag{2a}$$

$$X_Q(k) = n_Q(k). \tag{2b}$$

On the other hand, in the presence of signal, the corresponding components are given as :

$$X_I(k) = A\cos\theta + n_I(k) \tag{3a}$$

$$X_Q(k) = A\sin\theta + n_Q(k). \tag{3b}$$

Since the amplitude, frequency and the phase of the received signal are unknown, the detection criterion has to rely on thresholding of the amplitude (PA) of the analytic signal. The frequency and phase can be ignored for detecting only the presence of a target signal. The amplitude threshold can not be based on minimizing the total probability of error because the exact amplitude of the signal is unknown at the receiver. Furthermore, the probability of False Alarm (P_{FA}) must also be kept very low (10^{-6} or smaller). Hence, the best detection scheme would be to calculate the threshold by setting the P_{FA} to a constant. The thresholds for Square-Law detector have been derived next for single and multiple samples within a pulse.

III.2 : Square Law Detector

The noise is assumed to be narrowband and Gaussianly distributed with zero-mean and variance $= \sigma^2$. Hence, for the no-signal case of (2) the I/Q noise samples are distributed as :

$$X_I(k) = N(0,\sigma^2) \tag{4a}$$

$$X_Q(k) = N(0,\sigma^2). \tag{4b}$$

In the following derivation, the time-variable k will be suppressed until the multiple samples case is considered.

III.2.a : Single Sample Case

Assuming independent noise samples, when no signal is present, the joint probability density function (PDF) of the I/Q channel outputs are given by :

$$f(X_I, X_Q) = \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2}(x_I^2 + x_Q^2)}. \tag{5}$$

Let,

$$X_I = R\cos\alpha \tag{6a}$$

$$X_Q = R\sin\alpha. \tag{6b}$$

Using the Jacobian of this transformation, the joint PDF for this polar form can be shown to be :

$$f(r,\alpha) = \frac{r}{2\pi\sigma^2} e^{-\frac{r^2}{2\sigma^2}} u(r). \tag{7}$$

From this the marginal for the Envelope (R) is given by,

$$f_R(r) = \int_0^{2\pi} f(r,\alpha)d\alpha = \frac{r}{2\sigma^2} e^{-\frac{r^2}{2\sigma^2}} u(r), \tag{8}$$

which is known as the Rayleigh PDF.

III.2.a.1 : The PDF and Characteristic Function with No Signal

A square-law detector forms the following quantity,

$$Z \triangleq X_I^2 + X_Q^2 = R^2 \tag{9}$$

which needs to be compared to a threshold to decide the presence/absence of a radar target. Since, $\frac{dZ}{dR} = 2R = 2\sqrt{Z}$, the PDF of the Square-Law output when no signal (denoted as, $\overline{s}$) is given as :

$$f_Z(z|\overline{s}) = \frac{f_R(\sqrt{z})}{2\sqrt{z}} = \frac{1}{2\sigma^2} e^{-\frac{z}{2\sigma^2}} u(z), \tag{10}$$

which is the Exponential PDF. The Characteristic Function (CF) is defined as the Fourier Transform of the Density function :

$$\begin{aligned} \mathbf{C}_Z^{\overline{s}}(\omega) &\triangleq \mathcal{F}[f_Z(z|\overline{s})] \\ &= \frac{1}{2\sigma^2}\int_{-\infty}^{\infty} e^{-\frac{z}{2\sigma^2}}\ e^{-j\omega z} dz \\ &= \frac{1}{1 + j2\omega\sigma^2} \end{aligned} \tag{11}$$

III.2.a.2 : The PDF and Characteristic Function in Presence of Signal

When target is present, *i.e.*, in case of (3), the I/Q samples are distributed as :

$$X_I(k) = N(A\cos\theta, \sigma^2) \tag{12a}$$

$$X_Q(k) = N(A\cos\theta, \sigma^2). \tag{12b}$$

In this case, the joint probability density function (PDF) is given by :

$$f(x_I, x_Q) = \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2}[(x_I - A\cos\theta)^2 + (x_Q - A\sin\theta)^2]}. \tag{13}$$

Once again, using the Jacobian of the transformation, the polar-form joint PDF can be shown to be :

$$f(r, \alpha|s) = \frac{r}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2}[r^2 + A^2 - 2Ar\cos(\alpha - \theta)]} u(r). \tag{14}$$

Integrating over α, the marginal PDF of the Envelope is given by,

$$f_R(r|s) = \int_0^{2\pi} f(r, \alpha|s) d\alpha \tag{15a}$$

$$= \frac{r}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2}(A^2 + r^2)} \int_0^{2\pi} e^{-\frac{2Ar}{2\sigma^2}\cos(\alpha - \theta)} d\alpha, \tag{15b}$$

$$= \frac{r}{\sigma^2} e^{-\frac{1}{2\sigma^2}(A^2 + r^2)} I_0\left(\frac{Ar}{\sigma^2}\right) \tag{15c}$$

where, $I_0(\cdot)$ denotes Bessel Function of the zero-th kind. The PDF in (15c) is known as the Rician.

Similar to the no-signal case in (9)-(10), the PDF of the Square-Law output Z with signal-plus-noise is given as :

$$f_Z(z|s) = \frac{f_R(\sqrt{z})}{2\sqrt{z}} = \frac{1}{2\sigma^2} e^{-\frac{(A^2 + z)}{2\sigma^2}} I_0\left(\frac{A\sqrt{z}}{\sigma^2}\right). \tag{16}$$

In this case, the Characteristic Function can be found as follows :

$$C_Z^s(\omega) \triangleq \mathcal{F}[f_Z(z|s)]$$

$$= \frac{1}{2\sigma^2} e^{-\frac{A^2}{2\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{z}{2\sigma^2}} I_0\left(\frac{A\sqrt{z}}{\sigma^2}\right) e^{-j\omega z} dz \tag{17}$$

The following Fourier Transform pair can be found in [CF, page-79, pair 655.1] :

$$\mathcal{F}\left[e^{-\rho g} I_0\left(\frac{2\sqrt{g}}{\sqrt{\lambda}}\right)\right] \longrightarrow \frac{1}{\omega + \rho} e^{\frac{1}{\lambda(\omega + \rho)}}. \tag{18}$$

Using (18) and with appropriate substitutions, $C_Z^s(\omega)$ is given by :

$$C_Z^s(\omega) = \frac{1}{1 + j2\omega\sigma^2} e^{-\frac{j\omega A^2}{1 + j2\omega\sigma^2}} \tag{19}$$

III.2.a.3 : The Neyman-Pearson Criterion with a Single Sample :

For this one-dimensional case, the decision that the signal is present is taken if the likelihood-ratio [17] :

$$\ell = \frac{f_Z(z|s)}{f_Z(z|\overline{s})} > k(P_{FA}) \tag{20}$$

where k is a constant that depends on the probability of False-Alarm P_{FA}. From this relationship it may appear that in order to find the decision threshold, one would need to know or estimate the signal. But one of the most attractive consequence of Neyman-Pearson criterion is that for a given predetermined P_{FA}, the threshold can be set by integrating $f(z|\overline{s})$ over the region where the signal is present [11, 17].

III.2.a.4 : Probability of False Alarm

If the threshold is denoted as γ, the false-alarm probability can be calculated as,

$$\begin{aligned} P_{FA}^1 &= \int_\gamma^\infty f_Z(z|\overline{s})dz \\ &= \frac{1}{2\sigma^2}\int_\gamma^\infty e^{-\frac{z}{2\sigma^2}}dz \qquad \text{from (10),} \\ &= e^{-\frac{\gamma}{2\sigma^2}} \end{aligned} \tag{21}$$

III.2.a.5 : Detection Threshold

Taking natural logarithm of both sides of (21), the detection threshold is given as,

$$\gamma = -2\sigma^2 \ln P_{FA}^1. \tag{22}$$

III.2.a.6 : Probability of Detection

If the square-law output z of is greater than γ from (22), then the decision is taken that source target is present. Hence the probability of detection can be calculated from :

$$\begin{aligned} P_D^1 &= \int_\gamma^\infty f_Z(z|s)dz \\ &= 1 - \int_0^\gamma f_Z(z|s)dz \\ &= 1 - \frac{1}{2\sigma^2}e^{-\frac{A^2}{2\sigma^2}}\int_0^\gamma e^{-\frac{z}{2\sigma^2}} I_0\left(\frac{A\sqrt{z}}{\sigma^2}\right)dz. \end{aligned} \tag{23}$$

By letting, $v^2 = \frac{z}{2\sigma^2}$ and with appropriate substitutions,

$$P_D^1 = 1 - e^{-\frac{A^2}{2\sigma^2}}\int_0^{\sqrt{\frac{\gamma}{2\sigma^2}}} ve^{-v^2} I_0\left(2\sqrt{\frac{A^2}{2\sigma^2}}v\right)dv. \tag{24}$$

But this integral possesses the form of an Incomplete Toronto Function [13, pp-348] which is defined as follows :

$$T_B(m,n,r) \triangleq 2r^{n-m+1}e^{-r^2}\int_0^B t^{m-n}e^{-t^2}I_n(2rt)dt \tag{25}$$

Hence P_D^1 can be written in a more compact form as :

$$P_D^1 = 1 - T_{\sqrt{\frac{1}{2\sigma^2}}}\left(1,0,\sqrt{\frac{A^2}{2\sigma^2}}\right). \tag{26}$$

III.2.b : Multiple Samples Case

The detector performance can be expected to improve if the decisions can be based on multiple observations within a pulse. The question would then be how to combine the multiple samples in order to come up with an inference. For the Envelope Detection case, Tsui and Sharpin have recently derived an M-out-of-N scheme where the presence of target is decided if at least M samples out of a total of N exceed the detection threshold [12]. In this work we take a different approach where decisions are taken based on the sum of N squared samples. This approach is more akin to traditional CW detection schemes where integration over N pulses is performed for making a decision [13].

Let Y be the random variable formed with the sum of N independent squared samples, *i.e.*,

$$Y \triangleq \sum_{k=1}^{N} Z(k), \tag{27}$$

where, the PDF and CF of $Z(k)$ were derived in III.2.a.

III.2.b.1 : The PDF and Characteristic Function of Y with No Signal

When no signal is present, the PDF of Y which is formed as the sum of N *independent* samples, is given by the following convolution :

$$f_Y(y|\overline{s}) \triangleq f_Z(z_1|\overline{s}) \star f_Z(z_2|\overline{s}) \star \ldots \star f_Z(z_N|\overline{s}) \tag{28}$$

where, each of the $Z(k)$'s has identical distribution. Direct convolution of N PDFs appears complicated. but it is well-known that convolution in PDF-domain implies multiplication in the CF-domain. Consequently, the CF of Y is given by,

$$C_Y^{\overline{s}}(\omega) = \prod_{k=1}^{N} C_Z^{\overline{s}}(\omega) = [C_Z^{\overline{s}}(\omega)]^N$$

$$= \frac{1}{(1 + j2\omega\sigma^2)^N} \qquad \text{using (11)} \tag{29}$$

Using the inverse Fourier Transform pair-431 [Campbell and Foster, pp-44j], the PDF of Y is,

$$f_Y(y|\overline{s}) = \frac{y^{N-1}e^{-\frac{y}{2\sigma^2}}}{(2\sigma^2)^N(N-1)!}u(y) \tag{30}$$

III.2.b.2 : The PDF and Characteristic Function of Y in Presence of Signal

Using arguments similar to those in the previous subsection, when signal is present, the CF of Y is given by,

$$\begin{aligned} C_Y^s(\omega) &= \prod_{k=1}^{N} C_Z^s(\omega) = [C_Z^s(\omega)]^N \\ &= \frac{1}{(1 + j2\omega\sigma^2)^N} e^{-\frac{jN\omega A^2}{1 + j2\omega\sigma^2}} \end{aligned} \tag{31}$$

Once again, using the inverse Fourier Transform pair-650.0 [Campbell and Foster, pp-77], the PDF of Y is,

$$f_Y(y|s) = \frac{1}{2\sigma^2}\frac{y}{NA^2}^{\frac{N-1}{2}} e^{-\frac{1}{2\sigma^2}(NA^2 + y)} I_{N-1}\left(\frac{A\sqrt{Ny}}{\sigma^2}\right)u(y) \tag{32}$$

III.2.b.3 : The Neyman-Pearson Criterion with Multiple Samples :

For this N-dimensional case, the decision that the signal is present is taken if the likelihood-ratio [17] :

$$\ell_Y = \frac{f_Y(y|s)}{f_Y(y|\overline{s})} > k(P_{FA}). \tag{33}$$

For a given predetermined P_{FA}, the threshold can be set by integrating $f(y|\overline{s})$ over the region where the signal is present [11, 17].

III.2.b.4 : Probability of False Alarm

For γ denoting the threshold, the false-alarm probability is,

$$\begin{aligned} P_{FA}^N &= \int_{\gamma}^{\infty} f_Y(y|\overline{s})dy \\ &= \frac{1}{2\sigma^2}\int_{\gamma}^{\infty} \frac{e^{-\frac{y}{2\sigma^2}}y^{N-1}}{(N-1)!}dy \qquad \text{from (30)}, \\ &= 1 - \frac{1}{2\sigma^2}\int_{0}^{\gamma} \frac{e^{-\frac{y}{2\sigma^2}}y^{N-1}}{(N-1)!}dy \qquad (34a) \\ &= 1 - I\left(\frac{d}{2\sigma^2\sqrt{N}}, N-1\right) \qquad (34b) \end{aligned}$$

where, $I(\cdot)$ denotes Incomplete Gamma Function which is defined as,

$$I(u,t) \triangleq \int_0^{u\sqrt{1+t}} \frac{e^{-v}v^t}{t!}dv. \tag{35}$$

III.2.b.5 : Detection Threshold

For a given P_{FA}, the threshold γ can be determined numerically with a computer or using available plots/tables [13].

III.2.b.6 : Probability of Detection

If the sum-of-squares y is greater than the threshold γ determined from (35), then the decision is taken that source target is present. Hence the probability of detection can be calculated from :

$$\begin{aligned} P_D^N &= \int_{\gamma}^{\infty} f_Y(y|s)dy \\ &= 1 - \int_0^{\gamma} f_Y(y|s)dy \\ &= 1 - \frac{e^{-\frac{NA^2}{2\sigma^2}}}{2\sigma^2}\left(\frac{2\sigma^2}{NA^2}\right)^{\frac{N-1}{2}} \int_0^{\gamma} \frac{y}{\sigma^2}^{\frac{N-1}{2}} e^{-\frac{y}{2\sigma^2}} I_{N-1}\left(\frac{A\sqrt{Ny}}{\sigma^2}\right)dy \end{aligned} \tag{36}$$

By letting, $v^2 = \frac{y}{2\sigma^2}$ and with appropriate substitutions,

$$P_D^N = = 1 - \frac{e^{-\frac{NA^2}{2\sigma^2}}}{2\sigma^2}\left(\frac{2\sigma^2}{NA^2}\right)^{\frac{N-1}{2}} \int_0^{\sqrt{\frac{\gamma}{2\sigma^2}}} v^{N-1}e^{-v^2} I_{N-1}\left(2\sqrt{\frac{NA^2}{2\sigma^2}}v\right)dv. \tag{37}$$

This integral also possesses the form of an Incomplete Toronto Function defined in (25). Hence P_D^N can be written in a more compact form as :

$$P_D^N = 1 - T_{\sqrt{\frac{\gamma}{2\sigma^2}}}\left(2N-1, N-1, \sqrt{\frac{NA^2}{2\sigma^2}}\right). \tag{38}$$

III.3 : Envelope Detector

The PDF of the envelope (R) for a single sample was found in (8). Hence, for a given P_{FA}, the detection threshold is,

$$\gamma = \sqrt{-2\sigma^2 \ln P_{FA}}. \tag{39}$$

The calculation of threshold with N observation samples can be shown to be [12],

$$\gamma = T\sqrt{N\left(2-\frac{\pi}{2}\right)} + N\sqrt{\frac{\pi}{2}} \quad \text{where,} \tag{40a}$$

T is found approximately from,

$$P_{FA} = 0.5(1-\phi^{-1}(T)) \tag{40b}$$

and $\phi(\cdot)$ denotes the error function. More details for the Envelope Detector case can be found in [12].

It may be noted that unlike the square law and envelope detection threshold calculations for conventional radars [13], the discretized schemes presented here do not use matched filtered output but use the sampled data directly.

IV. : Adaptive Frequency/AOA Estimation :

Among the six parameters to be estimated by an EW receiver, the AOA and RF information are probably most important for sorting, identification and jamming. The multiple complex exponential model can be expressed as,

$$x(n) \triangleq \sum_{i=1}^{p} A_i e^{j\omega_i n}. \tag{5}$$

Estimation of the ω_i's is particularly difficult because of the non-linear nature of the optimization problem. A rich body of literature is available for high-resolution AOA and frequency estimation [4-11]. But most of the well-known high-resolution techniques appear to be computationally expensive for EW receivers. Recently, an adaptive frequency estimation algorithm that produces good estimates even at low SNR has been developed with digital EW receiver application in mind [5].

The adaptive algorithm is especially appropriate for EW requirements because the frequency estimates are updated at each sampling instant with a small number of computations. This algorithm makes use of the following recursive relationship followed by sinusoids :

$$e^{j\omega n} = 2\cos\omega e^{j\omega(n-1)} - e^{j\omega(n-2)}. \tag{41}$$

This relationship implies that a subsequence formed as,

$$x^{(1)}(n) \triangleq x(n) - 2\cos\omega_1 x(n-1) + x(n-2) \tag{42}$$

would have no contribution from ω_1 if the frequency is known or its estimate is available. In that case the subsequence $x^{(1)}(n)$ can be used in the same manner to form another subsequence $x^{(2)}(n)$ which will not have contributions from ω_1 and ω_2, if both are known. If this process is continued with true signal without any noise, the p-th such subsequence $x^{(p)}(n)$ will be identically zero. For noisy signals and with estimates of frequencies replacing the true values, $x^{(p)}(n)$ will be the modeling error $e(n)$. It has been shown in [5] that the frequencies can be updated as :

$$\cos\hat{\omega}_i(n+1) = \cos\hat{\omega}_i(n) + \mu(e^*(n)x(n-1) + e(n)x^*(n-1)) \quad \text{for, } i = 1,2,\ldots,p. \tag{43}$$

The adaptations for the amplitudes go as follows :

$$A_1 = \frac{x(n) - \sum_{k=1}^{p-1} A_{k+1} e^{j\omega_{k+1} n}}{e^{j\omega_1 n}} \tag{44a}$$

$$A_i = \frac{x^{(i-1)}(n) - \sum_{l=1}^{p-i} 2^{i-1} e^{j\omega_{p-l+1}(n-i+1)} \prod_{k=1}^{i-1}(\cos\omega_{p-l+1} - \cos\omega_k)}{2^{i-1} e^{j\omega_i(n-i+1)} \prod_{k=1}^{i-1}(\cos\omega_i - \cos\omega_k)} \quad i \neq 1, i \neq p, \tag{44b}$$

$$A_p = \frac{x^{(p-1)}(n)}{2^{M-1} e^{j\omega_p(n-p+1)} \prod_{k=1}^{p-1}(\cos\omega_p - \cos\omega_k)}. \tag{44c}$$

This adaptive scheme is attractive because of its computational simplicity as well as its real-time adaptive capabilities. But in simulations we found that when the model orders chosen in the algorithm were identical to the actual number of sinusoids, the estimates showed some bias and large variance. Interestingly, it was also found that if larger than true model order is used in the adaptation algorithm, the estimation bias as well the variance reduced considerably for the true frequencies. But the problem then would be to separate the actual frequencies from the extra ones. An obvious solution for this problem would be to discard the frequencies with lowest amplitudes produced by the amplitude adaptation scheme in (44). These approach worked quite well in most cases, though the performance was not entirely consistent. Some simulation examples that demonstrate these improvements in bias and variance are shown below for the two signals case.

IV.a : Simulation Studies on the Adaptive Algorithm

The frequency estimation part of the Direct Adaptive Frequency Estimation algorithm (DAFE) has been coded using the MATLAB software package. It may be noted that there were some inconsistencies in the amplitude adaptation equations in the viewgraphs of the paper which were supplied by the author of [5]. The corrected equations are given above in (44). The performance of DAFE was evaluated at various SNR levels to study the accuracy of estimation (both frequency and amplitude), the number of iterations needed for convergence, and for determination of 'appropriate' values for the convergence constant, μ.

Fig. 1a - Fig. 1d show the convergence of the frequency updates when 2 frequencies (at 0.1 and 0.3) were present at SNR levels 1000dB (no-noise), 30dB, 20dB and 10dB, respectively. Fig. 1e - Fig. h show the corresponding Amplitude adaptation results. The value of μ=.01 was used in these experiments. Fig. 2a and Fig. 2b display the frequency adaptation results at SNR values of 20dB and 10dB, respectively, when the data also had one additional frequency at 0.2. Fig. 2c and Fig. 2d display the amplitude adaptation results at SNR values of 20dB and 10dB, respectively, when the data had three frequencies at 0.1, 0.3 and 0.4. Fig. 3a - Fig. 3d display the corresponding results when the data had four unknown frequencies at 0.1, 0.2, 0.3 and 0.4. μ was 0.0007 for all the experiments in Fig. 2 and Fig. 3. The amplitude of each frequency was 2 in all cases. The value of μ needed for convergence varied depending upon the strength of the noise, the strength of the signal and the number of frequencies present in the incoming signal. The values given here were chosen by trial and error until a μ-value worked. It may be noted that these values are not necessarily the optimum ones for convergence and may be increased. But care must be taken so as not to increase the μ values too much, which may lead to chaotic behavior of the algorithm. The plots show that for two signals, the algorithm converges within 50-60 iterations at all SNR values, but it took about 1000 adaptations to converge when there were three frequencies. As evidenced by these plots, the algorithm performed reasonably well in terms of accuracy, even at low SNR values. The time of convergence was small in most cases, but in a few it was perhaps too large.

The amplitude adaptation plots indicate that the average amplitude estimations are quite accurate for various SNRs. The time of convergence is small in some cases and large in others. We also found that the time of convergence for amplitude was longer than that for frequency estimation, possibly due to the amplitude equations' dependence on the frequency estimates.

The results in Fig. 1-3 indicate that the performance of DAFE degrades at low SNR. This can be observed from the relatively high variability in the converged values at the tail-end of the plots, especially at 10dB SNR level. The sample-means computed with the frequency/amplitude estimates obtained at the 50-100 of the tail-end of the adaptation schemes were also found to be biased. This phenomenon is not entirely uncommon in other spectrum estimation algorithms [4, 6, 10, 11]. In fact, with true model order, the performance of many well-known algorithms such as, Prony's method, Autocorrelation/Covariance methods, Pisarenko's Method and others, deteriorate considerably as low SNR. A commonly used solution in such cases has been to use higher model orders than the actual order in the estimation algorithm. Whether such strategy would have the desired effect of reduction in bias and variance of the frequency/amplitude estimates from DAFE, was of considerable interest and the results are shown in the final set of plots.

In Fig. 4 and 5, the data generated was exactly identical to the scenario depicted in Fig. 1. The only differences are that the algorithm produces 3 and 4 frequency-amplitude pairs for Fig. 4 and Fig. 5, respectively. Interestingly, the amplitude adaptations show that only two amplitudes converge to the true values (=2), whereas the extra amplitude/s converge to very small values. Hence, the frequencies corresponding the small amplitudes can be discarded. It should be further noted that with model order of three, though the variability in the estimates reduced, the bias became rather high. But, with model order = 4, both the variance and the biases were considerably low at all SNR levels. Furthermore, the amplitude levels of the extraneous frequencies were also considerably lower than the true ones. The convergence times were also found to be of the same order as in Fig. 1, where true model-orders were chosen by the estimation algorithm.

Discussion : Overall, the accuracy of the algorithm under various noise conditions was quite good. When the adaptive updating were continued for long time, the frequency and amplitude estimates became very accurate over time (correct to four decimal places). Increasing the order was successful in the sense that frequencies not present in the incoming signal could be eliminated by thresholding of the amplitude estimates. However, a not of caution should be added that the time of convergence may be unpredictable and erratic over various input conditions (Fig. 4, for example).

Choosing the value of μ was a process of trial and error. When too large, the frequency estimates would sometimes be complex and sometimes the adaptations would not converge and grow erratic instead. When too small, the convergence time would increase. The amplitude estimates were only affected by μ through their dependence on the frequency estimates. The amplitude estimates sometime became complex when noise was present in the incoming signal, but the magnitudes were always close to the proper values for the actual frequencies. The frequency estimates were always

between -pi and pi and 2*pi was added when the estimates were negative.

V : Concluding Remarks and Future Work

In this report, we have addressed two of the most important aspects of digital EW receiver, namely, Time-Domain Detection and Adaptive Frequency/Amplitude Estimation. The MATLAB codes for both algorithms have been provided to the program monitor and the results of a more comprehensive simulation study has also been provided separately. Apart from the problems described in this report, the authors have also thoroughly studied the effect of quantization and hard-limiter on the estimates obtained from the well-known Prony's Algorithm [4, 6, 11]. The results could not be included in this report because of space-limitation, but the program codes and simulation results have been submitted separately to the Program Monitor.

The digital receiver design offers considerable challenge due to the constraints imposed by the nature of the problem scenario itself. Effects of A-D conversion of such high-frequency wide-bandwidth EW signals are not well understood. The availability of hardware technology would also limit the effectiveness of any receiver. We did not incorporate quantization and hard-limiting of the data in the detection and adaptation schemes presented in this report but these constraints must be taken into account in the receiver.

We have not considered the frequency-domain detection problem which can be addressed in many different ways. If tone frequencies exist within the observation window, they will show up as peaks along with the noise spectra in the estimated spectra of the received signal. The detection problem then would be to distinguish between signal and noise peaks in the spectra. For calculating the threshold, the statistics of the signal and noise spectra need to be analyzed. A detection scheme based on periodogram peaks would also be useful as providing crude frequency estimates which can be utilized to initiate other frequency estimation schemes such as the cyclic algorithm [9]. On the other hand, the baseband envelope signal could also be used (as in time-domain case) for detecting the presence of radar signals. The spectral shape of the envelope signal should be that of noise spectra whether a pulse is present in the estimation window or not. In that case, the detection threshold could be based on the energy of the spectra which would be much larger when a pulse is present. Clearly, the performance would depend on the strength of the source signals when compared to the background noise. The frequency-domain detection and tracking appear to be open problems and further theoretical research in these lines may have useful applications in other problems also.

Digital EW receivers can be expected to offer some major advantages over their analog counterparts. Foremost among these is the almost lossless storage capability of digital memories which can eliminate the dependence on lossy analog delay lines. Digital processors and memory chips are relatively inexpensive, compact in size and have low weight and the trends are towards even further reductions. Digital signal processing algorithms and digital computing technology have matured

tremendously and offer a wide range of capabilities. Parallel processing, pipelining, RISC, VLSI design, systolic architecture, vectorization and array processing, fault tolerant computing and etc., are only some of the well-known aspects of digital computing that the last few decades of research have produced. As our research progresses, we intend to study if some of these ideas can be incorporated in the digital receiver in order to improve the efficiency and accuracy of its performance.

REFERENCES

[1] J. B. Y. Tsui, *Microwave receivers with Related Components*, National Technical Information Center, 1982, Peninsula, Los Altos, CA, 1985.

[2] J. B. Y. Tsui, *Microwave Receivers with Electronic Warfare Applications*, John Wiley and Sons., New York, 1986.

[3] D. Curtis Schleher, *Introduction to Electronic Warfare*, Artech House, MA, 1986.

[4] J. B. Y. Tsui, *Digital Microwave Receivers : Theory and Applications*, Artech House, MA, 1989.

[5] J. Y. Cheung, "A Direct Adaptive Frequency Estimation Technique," *30th Midwest Symposium on Circuits and Systems*, New York, Aug., 1987.

[6] S.M. Kay, *Modern Spectral Estimation: Theory and Applications*, Prentice Hall, Englewood Cliffs, NJ, 1988.

[7] R. Kumaresan, L. L. Scharf and A. K. Shaw, "An Algorithm for Pole-Zero Modeling and Spectral Estimation," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol.ASSP-34, pp. 637-640, June, 1988.

[8] Y. Bressler and A. Macovski, "Exact Maximum Likelihood Parameter Estimation of Superimposed Exponential Signals in Noise," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-34, no. 10, pp. 1081-1089, Oct., 1988.

[9] A. K. Shaw, "A Novel Cyclic Algorithm for Maximum-Likelihood Frequency Estimation," *IEEE International Conference on Systems Engineering*, Dayton, OH, Aug., 1991.

[10] D. W. Tufts and R. Kumaresan "Frequency Estimation of Multiple Sinusoids : Making Linear Prediction Perform Like Maximum Likelihood," *Proceedings of the IEEE*, vol. 70, pp. 975-989, Sept., 1982.

[11] L. L. Scharf, *Statistical Signal Processing - Detection, Estimation and Time Series Analysis*, Addison-Wesley, Reading, MA, 1990.

[12] J. B. Y. Tsui and D. Sharpin, Unpublished Report on Time-Domain Detection for Digital Receivers, June, 1992.

[13] J. V. DiFranco and W. L. Rubin, *Radar Detection*, Artech House, Inc., Dedham, MA.

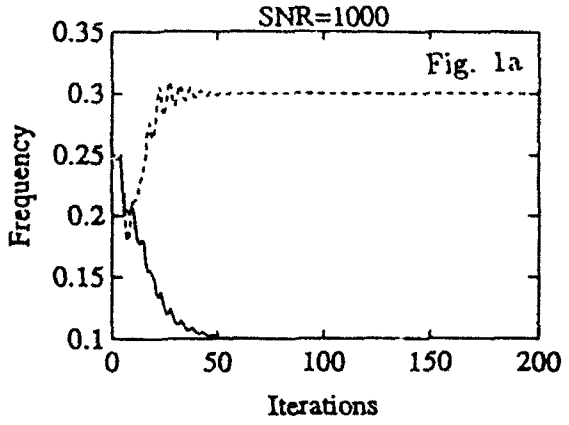
[14] J. I. Marcum, "A Statistical Theory of Target Detection by Pulsed Radar," *Trans. IRE Prof. Group on Information Theory*, IT-6; vol. 2, pp. 59-267, April, 1960.

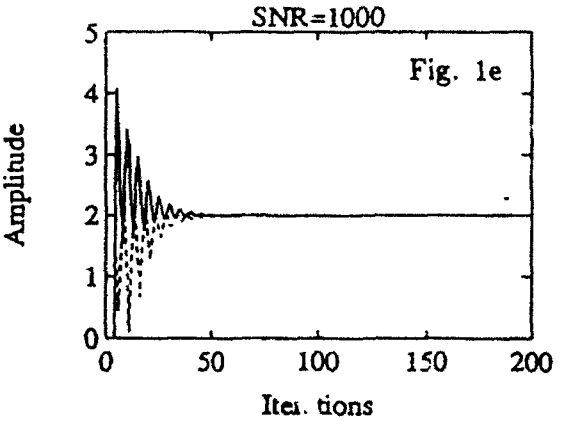
[15] G. A. Campbell and R. M. Foster, *Fourier Integrals for Practical Applications*, Van Nostrand, Princeton, NJ, 1948.

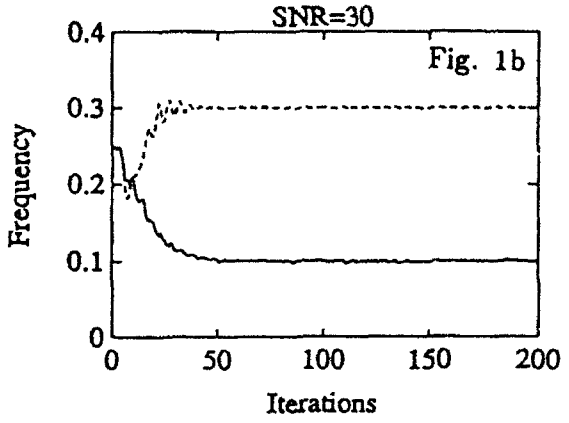
[16] W. C. Knight, R. G. Pridham and S. M. Kay, "Digital Signal Processing of Sonar," *Proceedings of the IEEE*, vol. 69, no. 11, Nov. 1981.

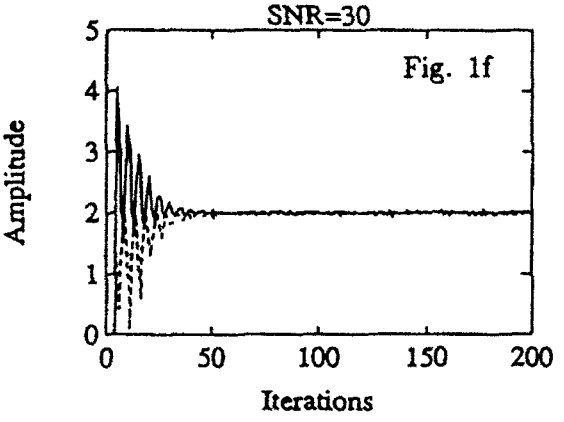
[17] M. Schwartz and L. Shaw, *Signal Processing : Discrete Spectral Analysis, Detection, and Estimation*, McGraw-Hill, New York, 1975.

Frequency Estimates | Order 2 | Amplitude Estimates

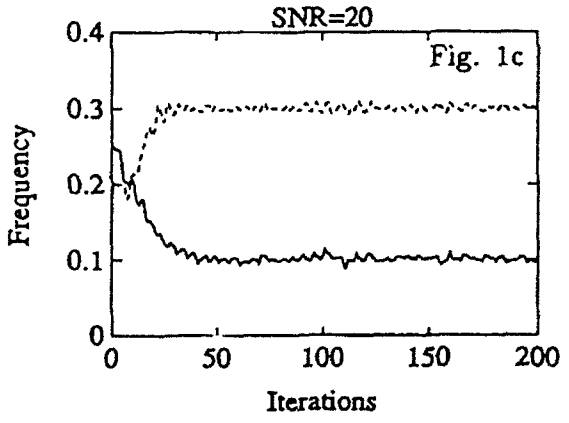


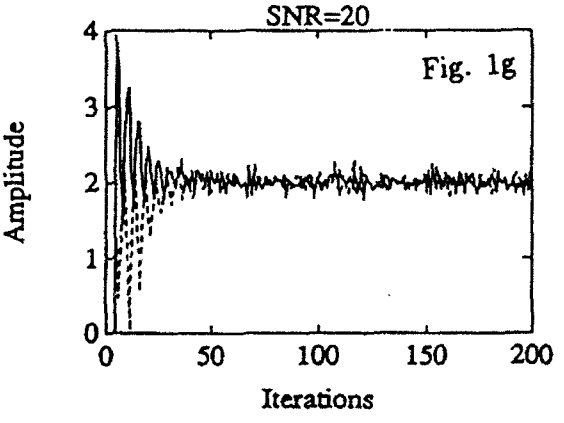


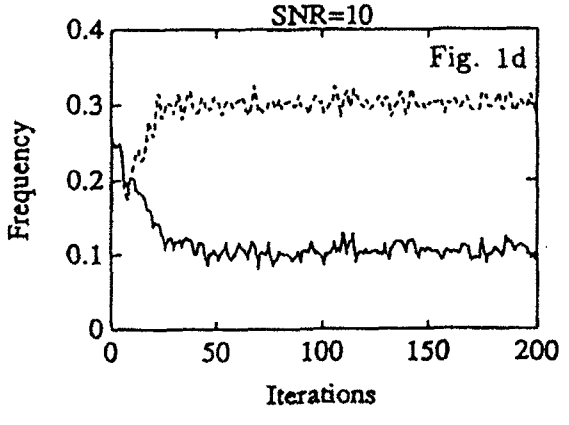


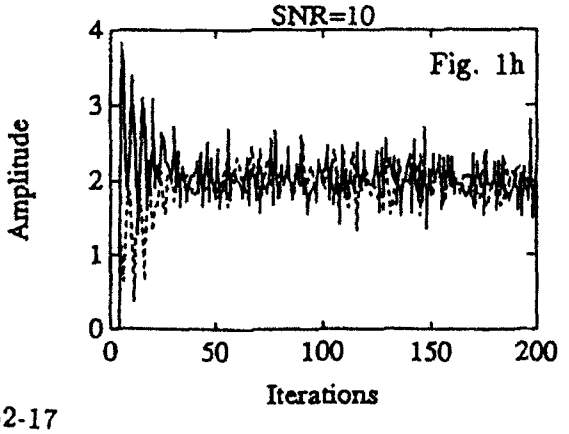


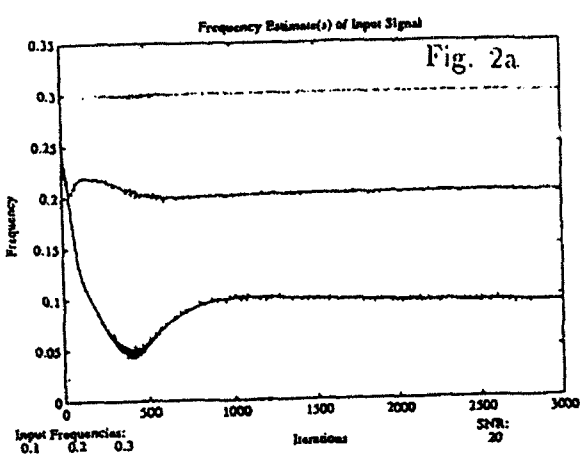
Frequency Estimates | Order 2 | Amplitude Estimates

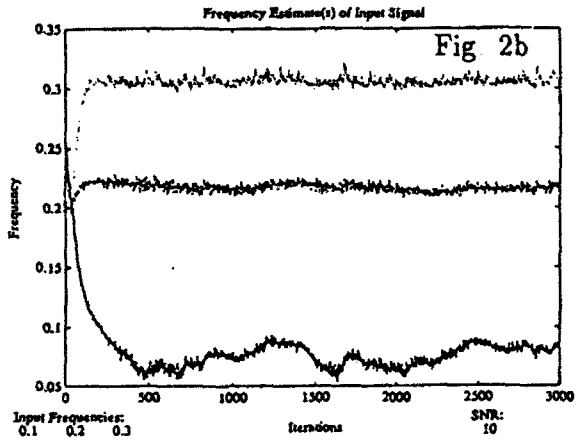


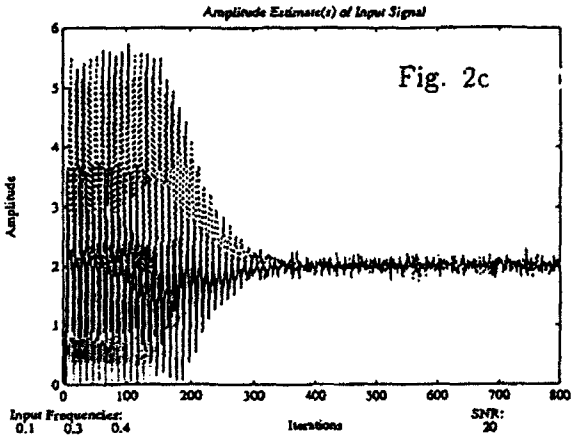


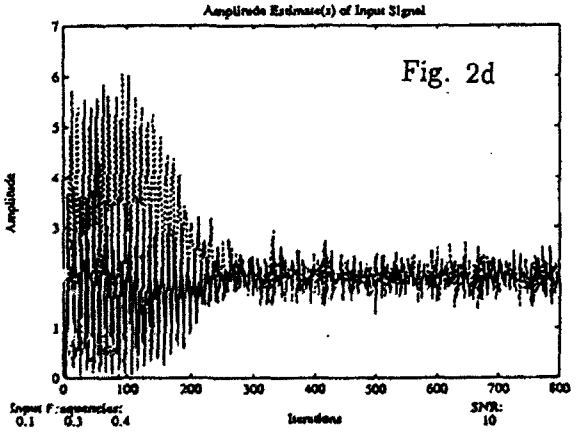


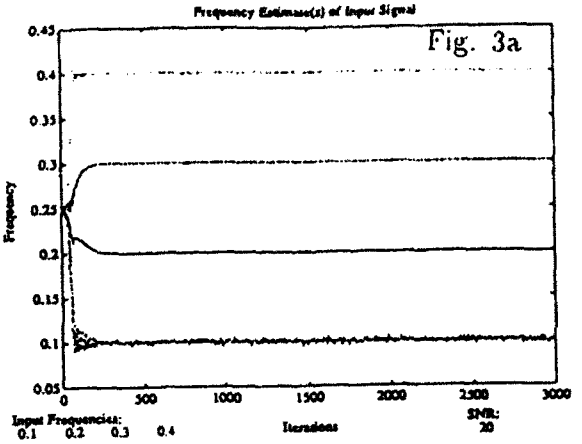


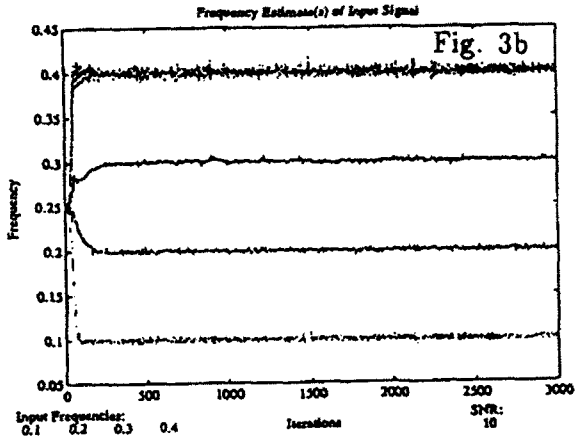

Frequency Estimate(s) of Input Signal
Fig. 2a
Frequency
Iterations
Input Frequencies:
0.1 0.2 0.3
SNR:
20

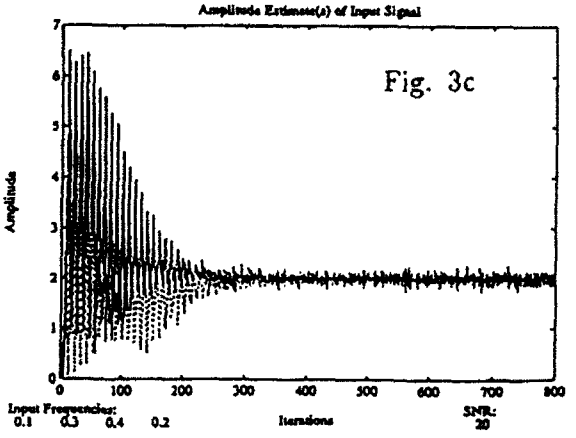

Frequency Estimate(s) of Input Signal
Fig. 2b
Frequency
Iterations
Input Frequencies:
0.1 0.2 0.3
SNR:
10

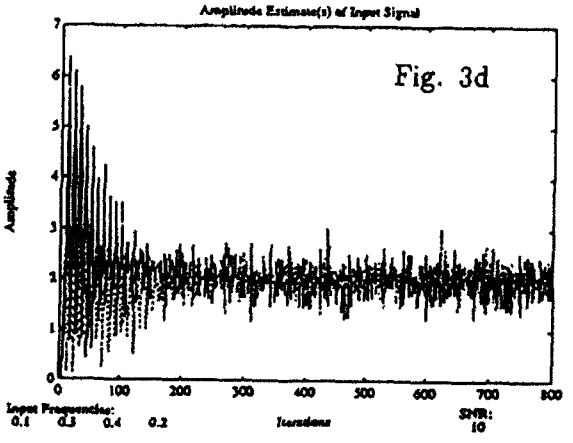

Amplitude Estimate(s) of Input Signal
Fig. 2c
Amplitude
Iterations
Input Frequencies:
0.1 0.3 0.4
SNR:
20


Amplitude Estimate(s) of Input Signal
Fig. 2d
Amplitude
Iterations
SNR:
10

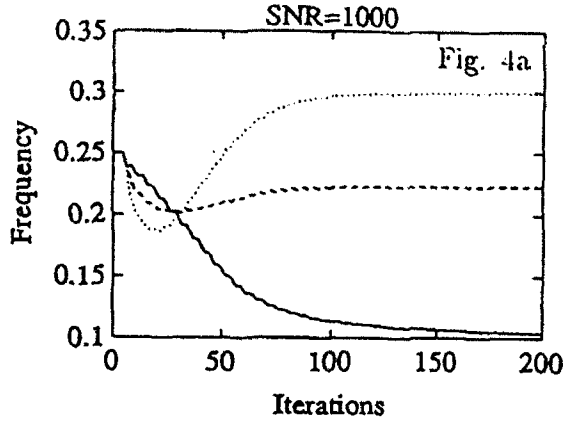

Frequency Estimate(s) of Input Signal
Fig. 3a
Frequency
Iterations
Input Frequencies:
0.1 0.2 0.3 0.4
SNR:
20

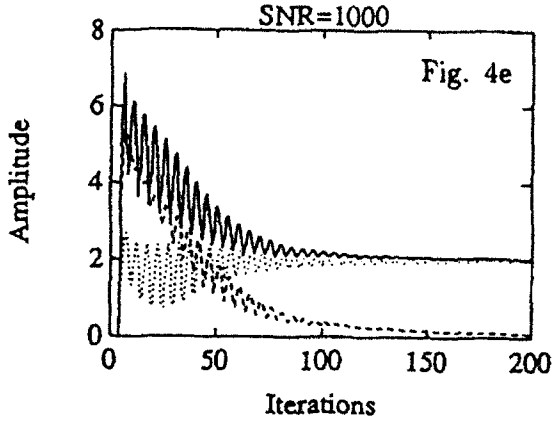

Frequency Estimate(s) of Input Signal
Fig. 3b
Frequency
Iterations
Input Frequencies:
0.1 0.2 0.3 0.4
SNR:
10

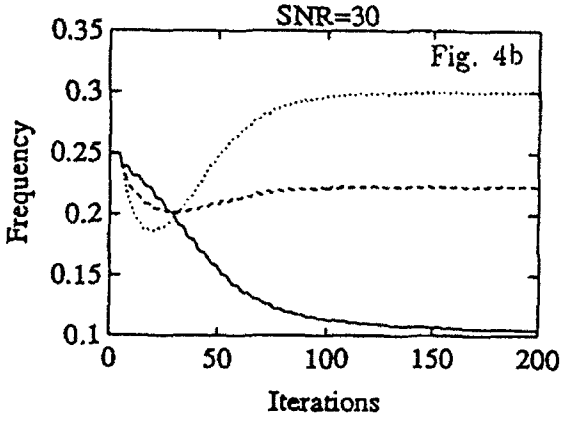

Amplitude Estimate(s) of Input Signal
Fig. 3c
Amplitude
Iterations
Input Frequencies:
0.1 0.3 0.4 0.2
SNR:
20

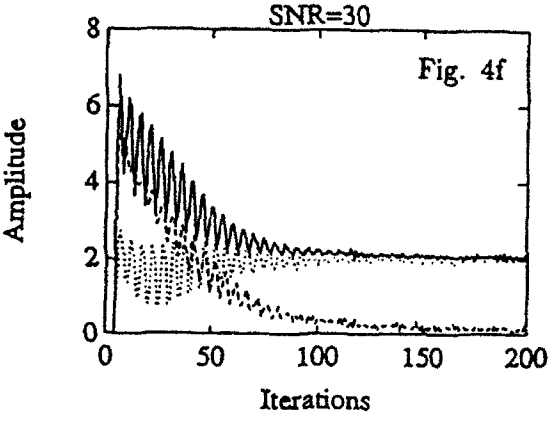

Amplitude Estimate(s) of Input Signal
Fig. 3d
Amplitude
Iterations
Input Frequencies:
0.1 0.3 0.4 0.2
SNR:
10

Frequency Estimates | Order 3 | Amplitude Estimates

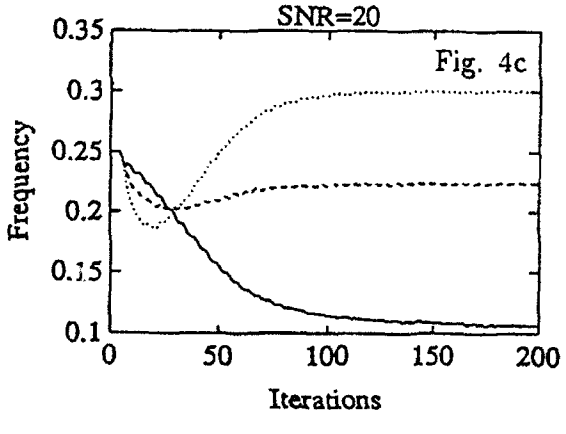


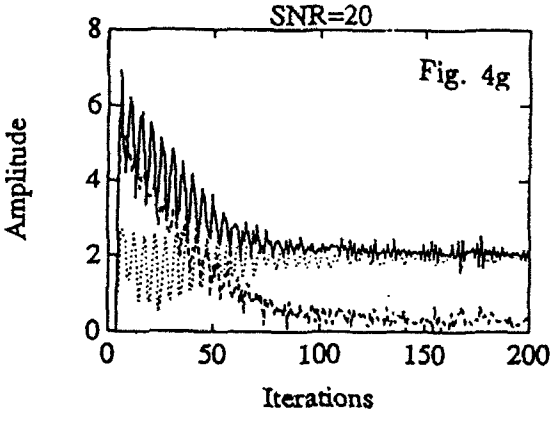


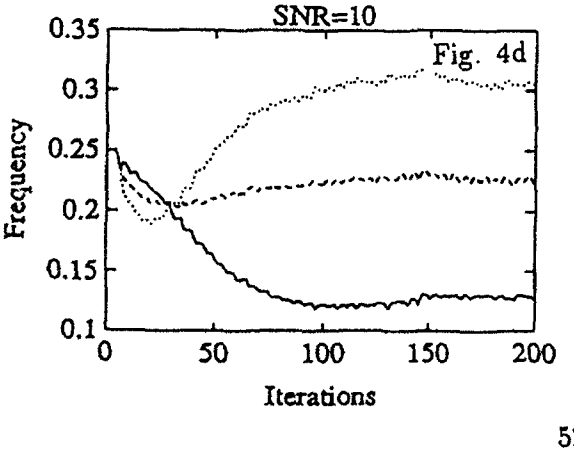


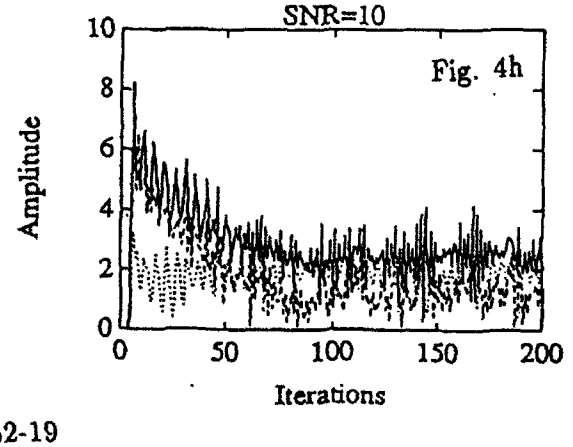


Frequency Estimates | Order 3 | Amplitude Estimates

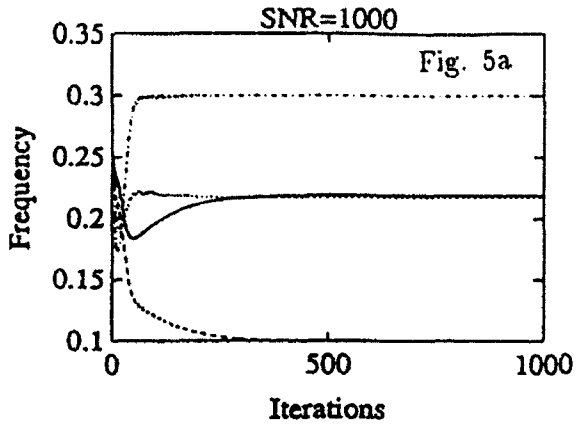


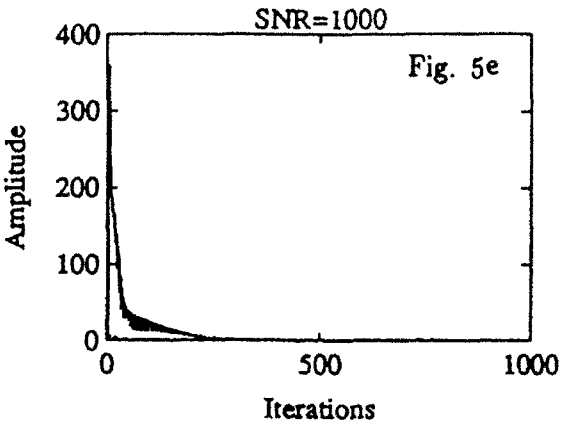


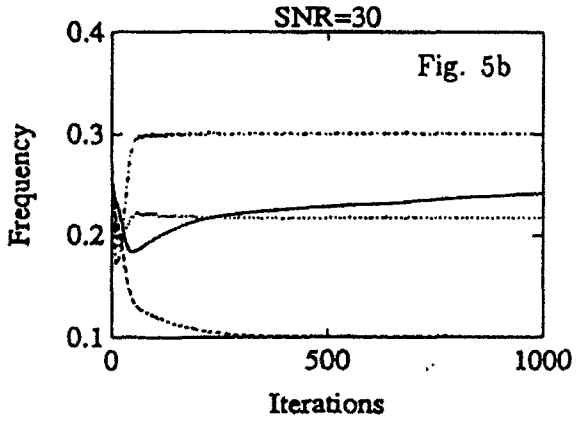


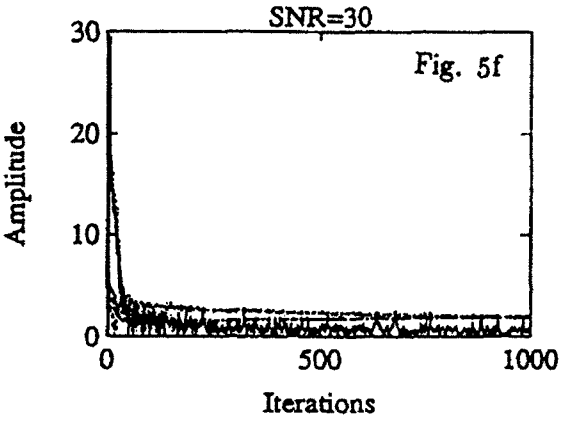


Frequency Estimates | Order 4 | Amplitude Estimates

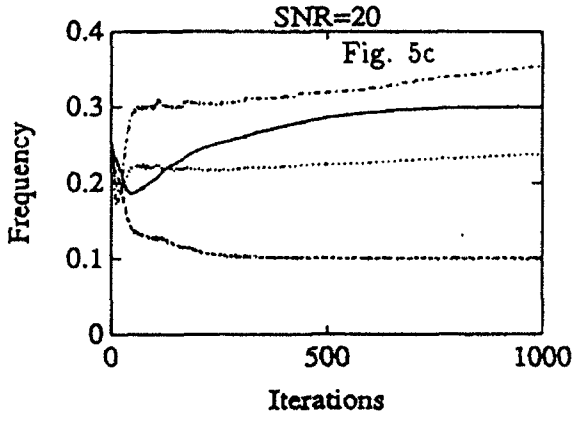


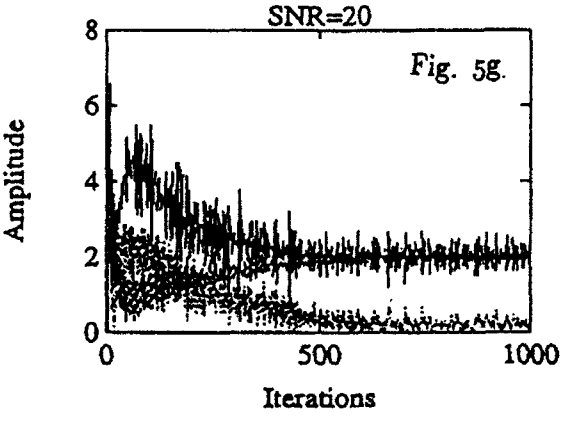


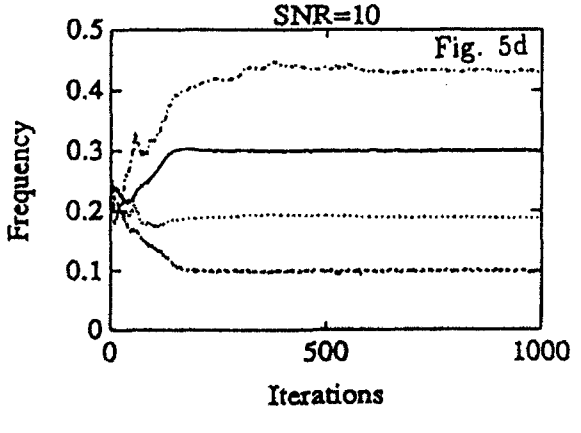


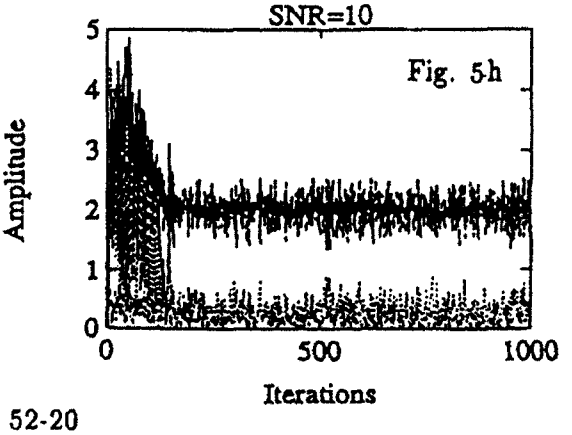


Frequency Estimates | Order 4 | Amplitude Estimates









HARDWARE IMPLEMENTATION OF THE ANVIL ALGORITHMS: A STUDY OF THE APPROACH

Janusz A. Starzyk
Professor
Department of Electrical
and Computer Engineering

Ohio University
347 Stocker
Athens, OH 45701

Final Report for:
AFOSR Summer Research Program
Wright Laboratory

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

August 1992

HARDWARE IMPLEMENTATION OF THE ANVIL ALGORITHMS: A STUDY OF THE APPROACH

Janusz A. Starzyk
Professor
Department of Electrical
and Computer Engineering
Ohio University

Abstract

Hardware implementation of the Artificial Neural VIsion Learning (ANVIL) system algorithms has been studied. Major stages of ANVIL processing steps are represented as separate functional blocks. They are designed as specialized neural networks with the interconnection weights learned off-line, during ANVIL training and associative mapping operations. Tradeoff between hardware requirements and computation efficiency was a motivating factor behind three variants considered during this study. A completely parallel system was analyzed and was found impractical for implementation. A hybrid system, which combines parallel processing in neural network layers with sequential scanning, can be implemented with reasonable hardware costs and perform within an acceptable time delay. Finally, a system of object identification at a given location was analyzed. This task was found to be much easier than the combined task of the object location and identification. Different technologies were considered and my recommendation is to combine MOS technology with the gallium arsenide ACT devices.

HARDWARE IMPLEMENTATION OF THE ANVIL ALGORITHMS: A STUDY OF THE APPROACH

Janusz A. Starzyk

INTRODUCTION

The Artificial Neural VIsion Learning (ANVIL) system, developed by Booz, Allen & Hamilton Inc. [1], is a software tool for automatic target recognition. It analyzes two dimensional photographs of three dimensional targets in order to detect and identify a target in a realistic environment. Targets are photographed under variable lighting conditions, background, and occlusions, and are approached from different elevations, azimuth angles, and distances. The three stages of ANVIL operations on selected images are training, associative mapping, and performance. The training stage is used to automatically extract and learn discriminating features of training objects. Associative mapping is used to establish connections between neural network layers for feature spatial locations, as well as object detection, identification and position within a tested image. The performance stage is used to identify targets on test images, therefore it realizes a computer simulation of a neural network operation.

In this study, results of the first two stages are used to construct a neural network with a specific architecture and interconnection weights. First, the general hardware stages needed to implement the ANVIL algorithms are discussed. Next, a neural network processor level architecture is presented. Then, specific hardware blocks are proposed and the implementation technology is recommended.

ANVIL STAGES

The ANVIL algorithms corresponding to the main operations in the performance stage are as follows:

1. Double variance computation
2. Feature extraction
3. Spatial location
4. Object detection

In order to explain architectural hardware requirements a brief description of these operations is conveyed first.

1. The double variance computation results in fuzzy line images, which yield edge detection more robust to feature distortion than sharp edge detection. Two variances are computed in fields of 3x3 and 5x5 pixels respectively, using the following equation:

$$\sigma_k^2 = \frac{\sum_{p,q \in N_{kij}} (x_{pq} - \overline{x_{kij}})^2}{Card\ N_{kij}} \qquad (1)$$

where N_{kij} is a 3x3 or 5x5 neighborhood of the pixel (i,j) for k=1 or k=2 respectively, Card(.) is the cardinality function. Then the double variance is found as

$$V_{ij} = \sqrt{\sigma_1^2 \times \sigma_2^2} \qquad (2)$$

This stage is not used in the ANVIL algorithms applied to optical images. However it will benefit the recognition of the infrared images. These images do not have clearly defined edges and are not obscured by the background to the same degree as the optical images are.

2. The feature extraction layer performs two operations: intensity gating and similarity computation. Intensity gating results in the output of the feature extraction layer equal to 1.0 if all the pixels in the feature window have values less than the intensity threshold. In ANVIL, with 8 bit gray level intensity values, this threshold is set to 20 on the scale from 0 to 255 (0 corresponding to the black and 255 to the white pixels). Otherwise, the output of the feature extraction layer is computed using the similarity computation. The learned features are combinations of the edge images and segmentation masks. Segmentation masks have external edges three pixels wide. These masks are used to divide the features into internal and external parts. Similarity between a tested image and a feature mask is computed in two stages. First, internal and external edge similarity errors are computed using

$$e_{tot} = \sin^2\theta_k = 1 - \left[\frac{(x_k\ y_k)}{|x_k|\ |y_k|}\right]^2 \qquad (3)$$

where k = 1,2, and x_k, y_k are the pixel vectors of the observed image and a learned feature, which fall under the internal (k = 1) and external (k = 2) masks. Then, the weighted average of the two errors defined as the fth feature value at (i,j)th pixel position is computed using

$$F_{i,j,f} = \frac{p \sin^2\theta_1 + q \sin^2\theta_2}{p + q} \quad (4)$$

where p and q are the numbers of pixels in the internal and the external masks, respectively. Similarity computation is performed with each learned mask for all possible feature positions in the test image. The result is a set of feature extraction clusters, where (i,j)th value in the cluster f corresponds to $F_{i,j,f}$ or 1.0 resulting from the intensity gating.

3. The spatial location layer applies the spatial location layer weight masks of the learned features to each position of the feature extraction clusters and finds a minimum error within each mask. It stores the minimum spatial location error value and the position at which it was found. The spatial location error with respect to the mth feature activation of the training pattern p at the pixel position (i,j) is computed using

$$S_{i,j,m,p} = \min_{q,r} \left[\frac{F_{i+q,j+r,f}}{D_{q,r,m,p}} \right] \quad (5)$$

where (q,r) is the location of the feature relative to the center of the training pattern, and $D_{q,r,m,p}$ is the spatial location layer weight mask, which depends on training parameters - scaling, elevation, azimuth, and a relative location of the feature (q,r) within the spatial location mask

$$D_{q,r,m,p} = \begin{cases} 1 - e_{3D}\, \rho_{qrmp} & if \ \rho_{qrmp} \le 1 \\ 0 & otherwise \end{cases} \quad (6)$$

where

$$\rho_{qrmp} = \sqrt{\left(\frac{h_{qrmp}}{a_{mp}}\right)^2 + \left(\frac{v_{qrmp}}{b_{mp}}\right)^2} \quad (7)$$

a_{mp} and b_{mp} are the lengths of the major and the minor axes of the spatial location mask (ellipsoid), h_{qrmp} and v_{qrmp} are the distances from the (q,r) pixel location to the minor and the major axes.

4. The object detection layer combines outputs from the spatial location layer in clusters which correspond to all training patterns, chooses candidate targets, and applies the user specified threshold to identify targets and their locations. The first stage of the

object detection finds similarity with the training patterns by computing object detection clusters. Each object detection cluster corresponds to a different training pattern p and its (i,j)th element is computed by

$$O_{i,j,p} = \frac{\sum_{m=1}^{N_p} (S_{i,j,m,p} - T_{m,p})}{N_p} \tag{8}$$

where $T_{m,p}$ is the value of the mth feature learned for the training pattern p, N_p is the number of activations of the learned features in the pattern p.

In addition to the object detection error obtained from (8), the system estimates and validates the actual target position. The relationship between the training point position and its transformed image is expressed by the transformation matrices

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = s \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & \sin\phi \\ 0 & -\sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} \tag{9}$$

Using two dimensional projections on the x-y image plane the transformed image moves to new coordinates

$$\begin{aligned} x &= s\ (x_0 \cos\theta + y_0 \sin\theta \sin\phi - z_0 \sin\theta \cos\phi) \\ y &= s\ (y_0 \cos\phi + z_0 \sin\phi) \end{aligned} \tag{10}$$

Having projections of x and y values for several features of a target image, and relating them to their trained positions x_0 and y_0, the transformation parameters s, θ, ϕ and the depth information z_0 can be estimated. Then the estimated values are validated by evaluating (10) for all feature activations and the difference between the observed and the computed locations is used to compute the validation error

$$e_{i,j,m,p} = \sqrt{\left(\frac{e_a}{a_{mp}}\right)^2 + \left(\frac{e_b}{b_{mp}}\right)^2} \tag{11}$$

where e_a and e_b are the differences between the observed and the calculated distances from the selected pixel location to the minor and the major axes of the spatial location mask, respectively. Root mean square value of the validation errors of all feature

activations is added to the object detection error. The object detection clusters are then combined to a single cluster. Each element of this cluster indicates which training object is the most similar to the tested image at a given object position. Finally spatial exclusivity and the output threshold are applied to declare object detections. Positions of the detected objects are indicated.

HARDWARE IMPLEMENTATION STUDY

To set a stage for a system level discussion of these concepts let us introduce the basic building blocks, their hardware requirements, and terminology. Hardware estimation is illustrated with example dimensions of the image and feature windows, as well as the object and spatial location mask sizes. The following system level blocks are defined (their detailed discussion follows the selected architectural concept.):

1. Two dimensional array of shift registers (SRA). This array stores digital data, shifts them to the right, left, or upwards, and feeds them to D/A converters. SRAs form efficient circuits for a parallel scan of memory areas.
2. D/A and A/D converters used in selected areas of SRAs for local signal conversions.
3. Similarity computation block. Used for feature extraction and object detection. Uses weights learned during training. Can be conveniently implemented on a commercial neural network.
4. Winner takes all circuit. Selects an input line with the minimum value. Used in the spatial location and the object detection stages.
5. Spatial exclusion circuit. It is a clustering circuit which selects a local minima and suppresses the neighborhood outputs.

The following symbols are used to indicate different dimensions of the required hardware (numbers in brackets indicate their example values)

IxI	- image size (128x128 pixels)
WxW	- feature window (11x11 pixels)
OxO	- object size (31x31 pixels)
LxL	- spatial location mask (10x10 pixels)
F	- number of features learned (26)
M	- number of activations or spatial location masks (52)
P	- number of training patterns (8)

Below, a system level description of the architectural concept is discussed. Hardware requirements are expressed by the introduced symbols with the example values in brackets.

HYBRID HARDWARE IMPLEMENTATION OF ANVIL

The hybrid realization of ANVIL algorithms has seven functional layers: input layer, feature extraction layer, spacial location layer, object detection layer, winner takes all layer, object identification layer, spacial exclusion layer. The block level structure of the hybrid architecture is shown in Fig. 1. It's functional layers are described next.

1. The input layer stores IxI (16,384) image pixels in the two dimensional shift register array with Ix(2I-W) (31,360) memory locations (see Fig. 2 a)). The upper central part of the SRA is overlapped with an array of D/A circuits with WxW (121) D/A converters. Each D/A converter takes its input from a single memory location and applies its analog output to the corresponding input of the feature extraction layer. There are two phases of operation in the input layer:

I. The image input phase. I (128) cycles are needed to scan-in the input image in parallel using I (128) analog inputs connected to A/D converters.

II. The image scanning phase. During each clock cycle the entire memory is shifted, and new data provided to the inputs of the FE layer. The total number of scanning cycles is (I+1-W)x(I+1-W) (13,924). Fig. 2 b) illustrates the shift operation of the input layer's SRA by tracing shifts of a single pixel in the image plane.

2. The feature extraction (FE) layer consists of a single neural network trained to recognize the learned features. It has WxW (121) inputs and F (26) outputs. Each output is connected to an A/D converter. At each clock cycle the FE network computes the feature errors and delivers them to the outputs. The FE layer works synchronously with the image scanning phase.

3. The spatial location (SL) layer consists of M (52) neural networks, each finding a weighted feature value at different object locations. Each neural network has a single analog input connected to an A/D converter, an SRA with (I+1-W)x(2I-2W+1) (27,730) memory elements, and a winner takes all circuit, which finds a minimum of the weighted inputs from a spatial location mask (5). The upper central part of the SRA is overlapped

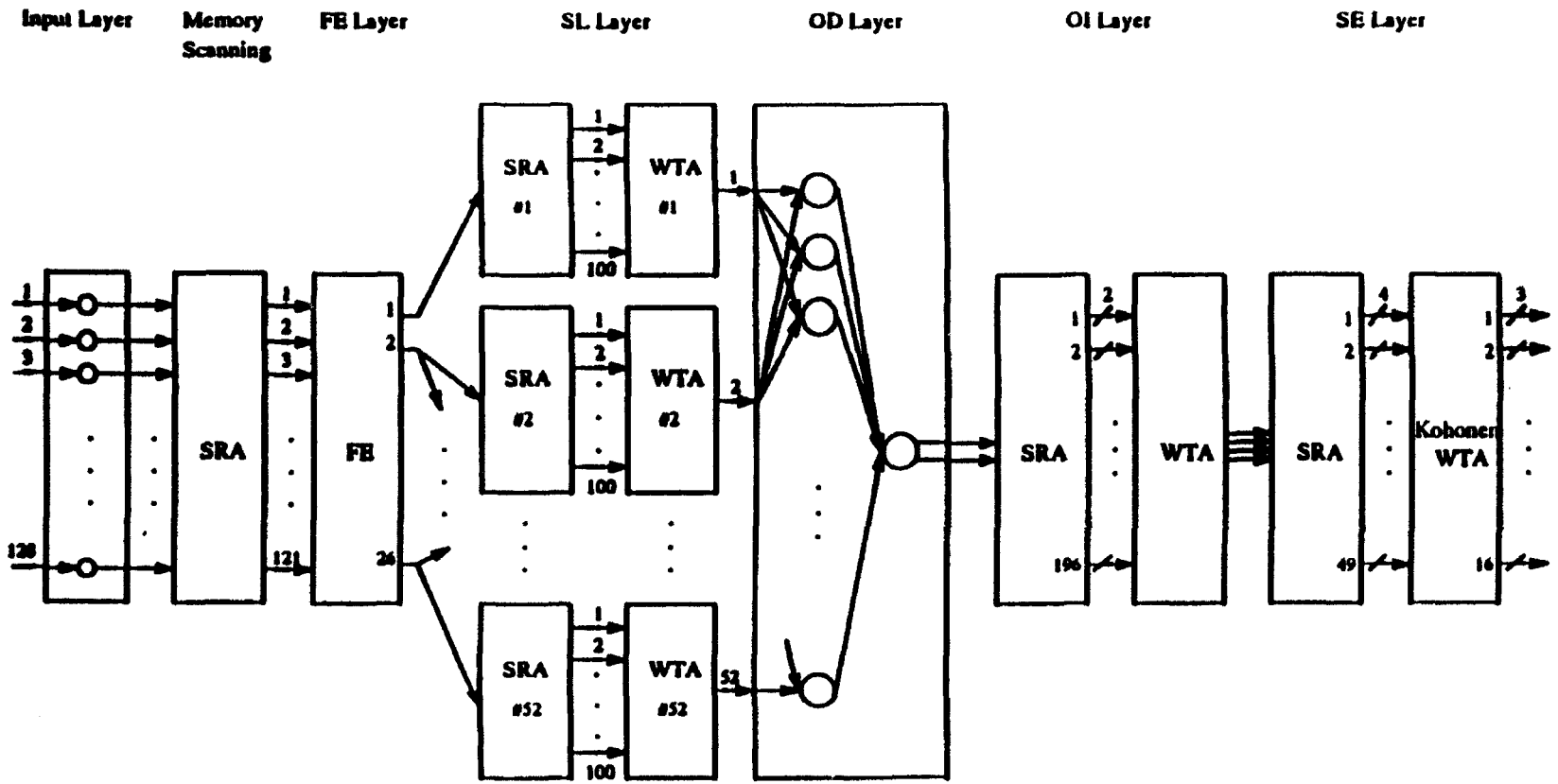


Fig. 1 Hybrid Neural Networks for ANVIL

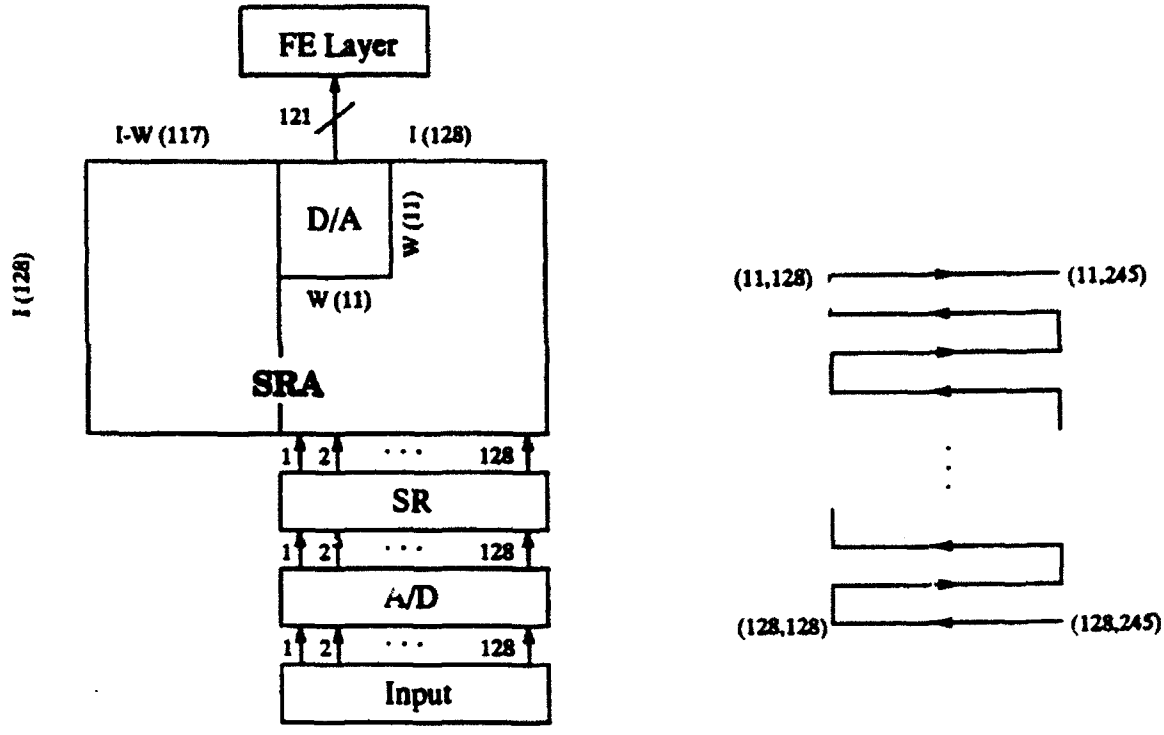


Fig. 2 a) Input SRA and FE layer, b) SRA shifting scheme

with an array of D/A circuits with LxL (100) D/A converters. Each D/A converter takes its input from a single memory location and applies its analog output to the corresponding input of the winner take all circuit [2-4]. The total number of analog outputs from the spatial location layer is M (52). The structure of a spatial location neural network is shown in Fig. 3 a). There are two phases of operation in the FE layer:

I. The image scanning phase. During the image scanning phase the input data are fed through an A/D converter to the SRA. In each clock cycle the entire memory is shifted, and new data provided to the input of the SRA. After (I+1-W)x(I+1-W) (13,924) clock cycles the right part of the SRA is filled with the feature values. The SRA shifting in the image scanning phase is illustrated in Fig. 3 b).

II. Spatial location scanning. In the next (I+1-O)x(I+1-O) (9,604) clock cycles the feature values are shifted and passed through a winner takes all circuit. Weights of the WTA circuits are determined by the spatial location weight masks (6). Each WTA circuit in the SL layer takes up to LxL (100) analog inputs and produces a single output. The SRA shifting in the spatial location scanning phase is illustrated in Fig. 3 c).

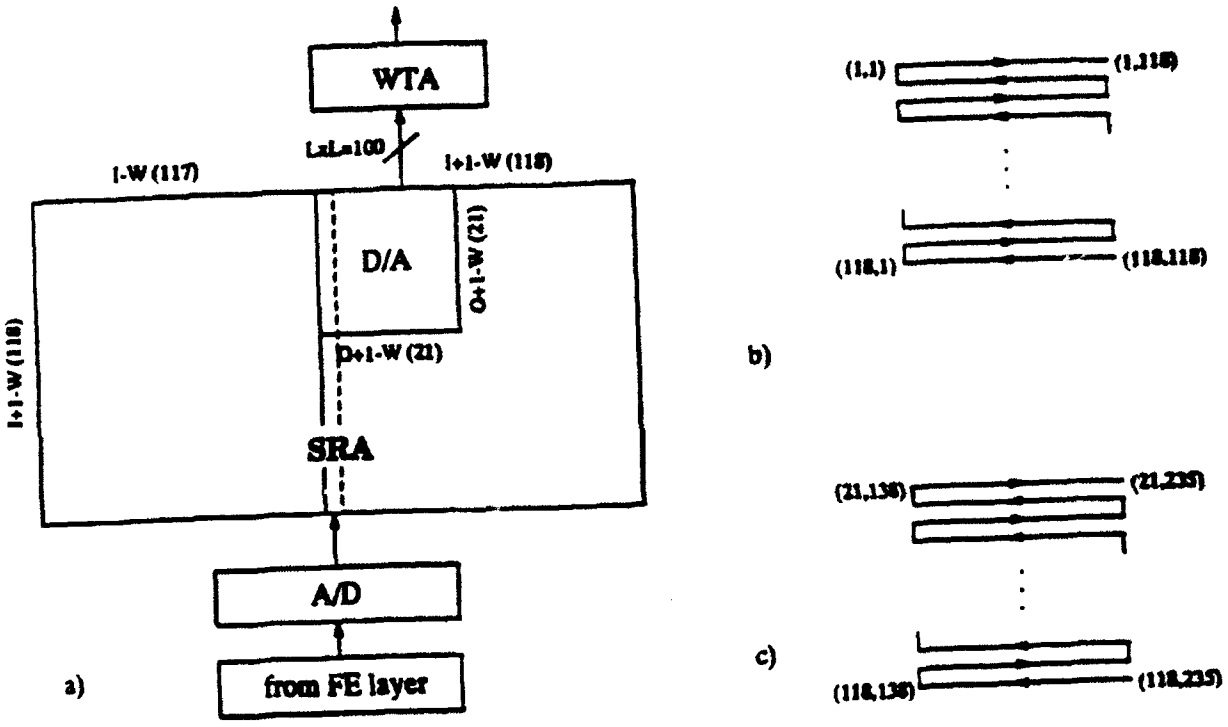


Fig. 3 a) Spatial location neural network, b) SRA shifting in image scanning phase, c) SRA shifting in spacial location scanning phase.

4. The object detection (OD) layer and the winner takes all layer are combined in a single two layer neural network. The network has M (52) analog inputs, P (8) neurons in the hidden layer, and two outputs. Each neuron in the hidden layer finds a weighted sum of trained feature values for a single object at different object locations. For current implementation of ANVIL algorithms, simple averages of the input feature values are required to compute the object detection error.

More accurate results are obtained if the target position is estimated and validated. This implementation requires that outputs of the winner takes all circuits from the spacial location layer contain the location of the winning signal and the feature location (center of the feature mask). An additional neural network is used to obtain transformation parameters s, θ, ϕ and depth z_0. Then the validation errors (11) are computed and added to the object detection error. The combined errors are the outputs of the hidden layer. Output signals from the neurons in the hidden layer are fed to a WTA circuit, which finds a minimum out of the P (8) neuron outputs. The winning neuron number and its output

signal value are passed to the two outputs. The organization of the object detection layer is shown in Fig. 4.

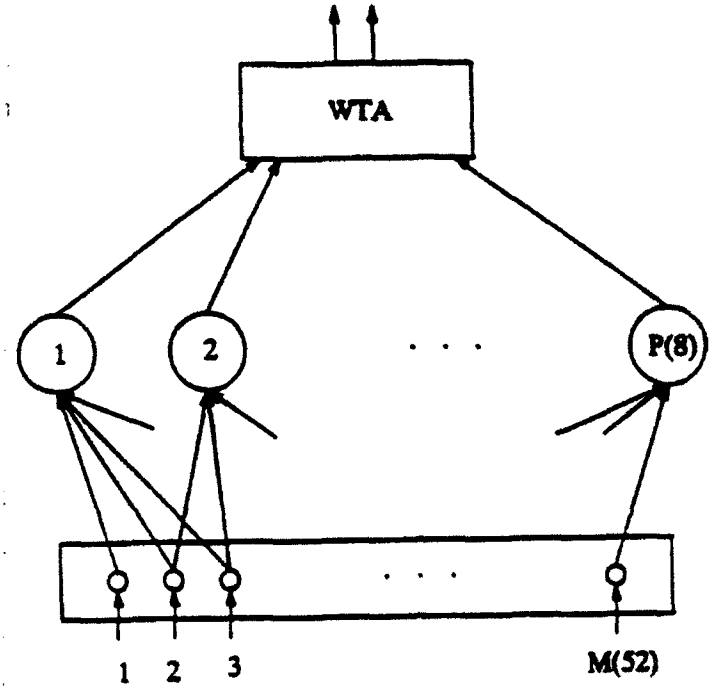


Fig. 4 Object detection layer

5. The object identification layer and the spatial exclusion layer are also combined in a single neural network (see Fig. 5). The neural network has two analog inputs connected through A/D converters to an SRA with 2x(I+1-O)x(2I+1-2O) (38,220) memory elements. The upper central part of the SRA is overlapped with an array of D/A circuits with 14x14 (196) D/A converters. Each D/A converter takes its input from a single memory location and applies its analog output to the corresponding input of the winner takes all circuit. The outputs from the WTA circuit are connected through A/D converters to a small 7x13 SRA. Finally the left part of the SRA is connected through a 7x7 array of D/A converters to the spatial exclusion (SE) circuit. There are three phases of operation in this neural network:

I. Spatial location scanning. In the (I+1-O)x(I+1-O) (9,604) clock cycles of the spatial location scanning phase the winning neuron number and its output value from the OD layer are fed through A/D converter to two SRAs. In each clock cycle the entire memory is shifted, and new data provided to the input of the SRAs.

The SRA shifting in the spatial location scanning phase is illustrated in Fig. 5 b).

II. Object identification scanning. A winner takes all circuit with the output threshold applied is used to identify an object in a 14x14 window. The selected winner (the object number and the winning signal value) and its location are stored in a small SRA (7x13), and after 14 clock cycles a new 14x14 window is examined and the results stored. After {E[(I+1-O)/14]xE[(I+1-O)/14]x14-14] (672) clock cycles, where E[x] stands for the smallest integer greater or equal to x, the object identification scanning is finished and the small SRA filled with data. The operation in the object identification scanning is illustrated in Fig. 5 c).

III. Spatial exclusion operation. All outputs from the object identification layer, which are stored in the small SRA are fed to an SE network. The purpose of this network is to avoid detection of two overlapping objects. Each input corresponds to a subarray of the object identification layer. All inputs to the SE layer are arranged into a rectangular array (7x7). The network suppresses outputs in a neighborhood of a winning signal. This suppression must correspond to the size of a single object. For instance, using 14x14 subarrays and the object size 31x31, the suppressed signals are in the direct neighborhood of the selected signal. The organization of the spatial exclusion operation is shown in Fig. 6.

The hybrid neural network shown in Fig. 1 may detect, locate, and identify up to 16 simultaneous objects. The network uses scanning to reduce the hardware requirements. It has five phases of operation, which take 24,329 clock cycles. Estimating the clock cycle to be on a range of delay time of existing neural networks (5-6 usec), total operation of the hybrid neural network can be completed in less than 150 msec.

ACT TECHNOLOGY

The estimated performance and hardware requirements for the implementation of ANVIL algorithms was based on the bipolar CMOS and CHMOS technologies, taking into account existing neural network chips and their limitations. An alternative solution may be obtained using the new acoustic charge transport (ACT) technology developed by EDI [5-8]. An ACT chip is a gallium arsenide device which performs more than 46 billion multiply and accumulate operations per second. Currently available commercial digital signal processing chips operate at rates around 40 million multiply and accumulate operations per second, which is three orders of magnitude slower than the ACT device.

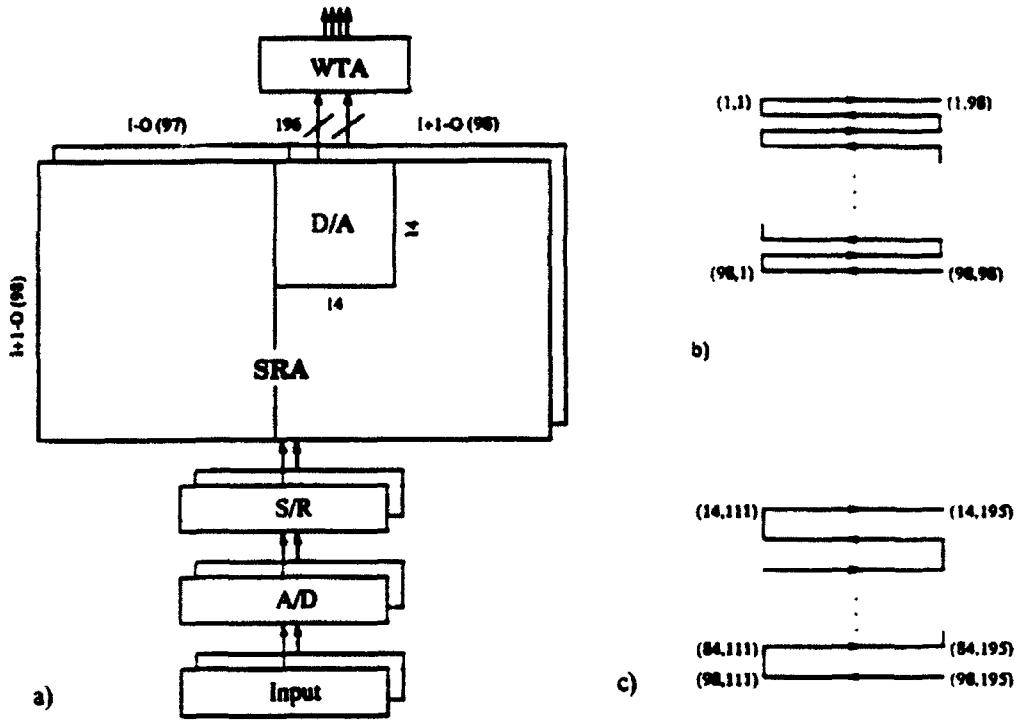


Fig. 5 a) Object identification SRA, b) SRA shifting in spatial location scanning, c) SRA shifting in object identification scanning.

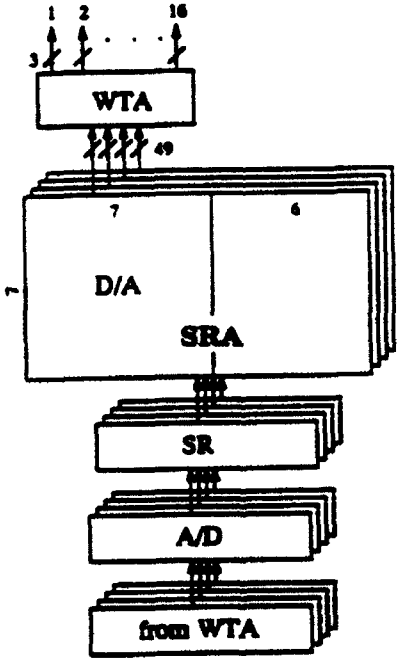


Fig. 6 Spatial exclusion operation.

The ACT device can be used as a programmable transversal filter with the values of the weighting coefficients (or taps) set by a digital controller [9]. A diagram of the ACT programmable transversal filter is shown in Fig. 7. A standard ACT programmable filter has 128 taps. Each tap is controlled by a digital controller, has signed four bits accuracy and can be set up in 100 nsec per tap. Fig. 8 shows the circuit organization of a single tap with weight controlled by the values $D_0,...,D_3$, and sign decided by D_s. Nondestructive sense electrodes (NDS) pick-up the value of the delayed input signals which pass under them, and the weighted signals are added together. As a result, the inner product of the input vector and the vector stored in the tap weights is computed.

The ACT programmable transversal filter used as a neural network element is a recommended choice whenever the scanning of the input array is combined with matching the input with given vectors. For that reason the ACT filter can be used very effectively in the hybrid hardware implementation of ANVIL algorithms at the feature extraction and the spacial exclusion operations. This implementation has seven functional layers labeled as in the hybrid system. Networks in the input and feature extraction layers are slightly different than the corresponding networks in the hybrid system.

The hybrid neural network for ANVIL algorithms which uses ACT technology has six phases of operation with the following time requirements:

I. The image input phase takes I (128) clock cycles
II. The weight loading phase takes Fx40 (1040) usec
III. The image scanning phase takes 2xIx(I+1-W)xF (785,408) clock cycles
IV. Spatial location scanning takes (I+1-O)x(I+1-O) (9,604) clock cycles
V. Object identification scanning takes {E[(I+1-O)/14]xE[(I+1-O)/14]x14-14} (672) clock cycles.
VI. Spatial exclusion operation takes one clock cycle

The first and the third phases can be clocked with 32.7 MHz frequency so the time needed to complete the first three phases is 25 msec. This time can be reduced to less than 2.5 msec if W (11) ACT devices are used for feature extraction. Estimating the clock cycle for the remaining phases to be on a range of delay time of the winner takes all circuit (100-200 nsec [3]), total operation of the hybrid neural network with W ACT devices can be completed in less than 5 msec.

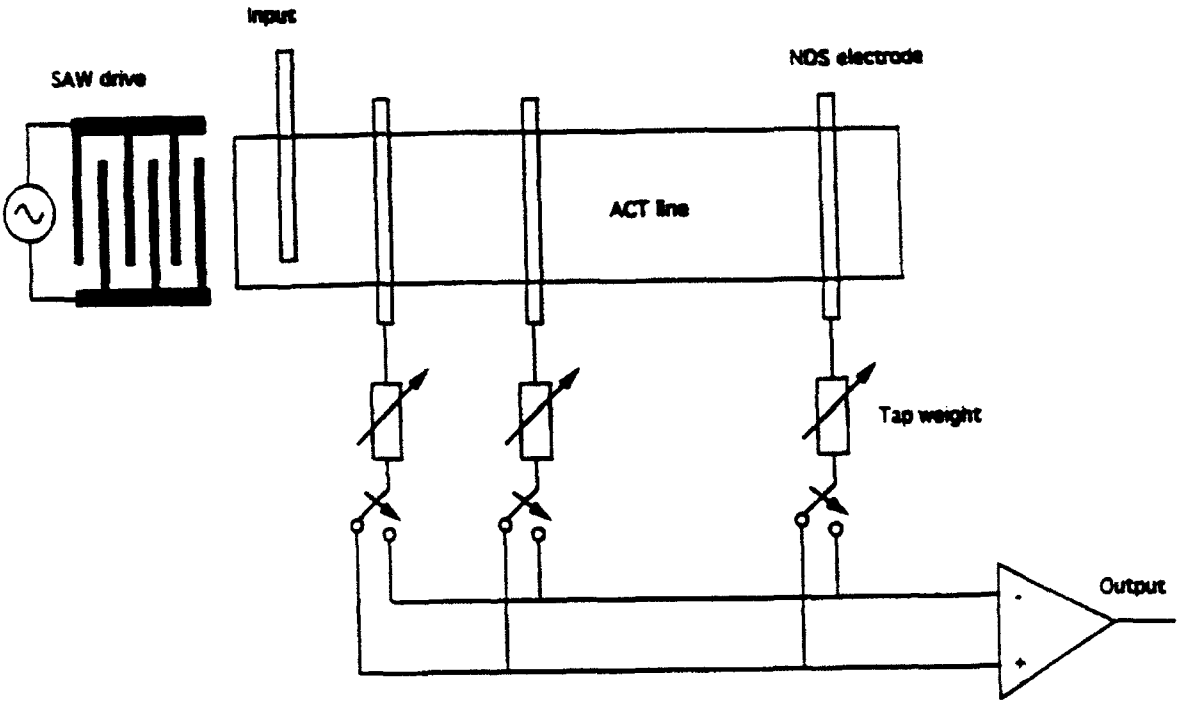


Fig. 7 The ACT programmable filter

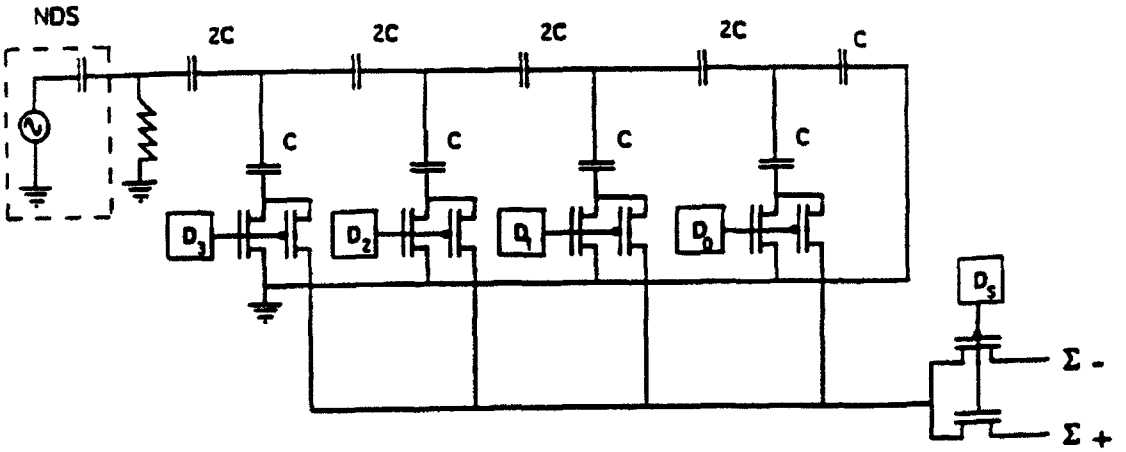


Fig. 8 Tap control circuit

RECOMMENDATIONS

Neural network technology is sufficiently mature for the implementation of sophisticated algorithms like ANVIL, and for providing results in an amount of time which is acceptable for many tasks including target recognition. The hybrid realization of ANVIL algorithms is the most suitable solution for the intended system design and is recommended for practical implementation.

The critical element in deciding detailed system organization is the choice of technology. Different circuits have been used in neural network designs including analog [11-14], digital [15], hybrid [16,17], switched capacitor [18], pulse modulated [19,20] and optoelectronic [21,22] networks. Most of them were based on well established MOS technologies. Related development was concentrated on analog and associative memories for neural networks [23,24].

MOS technology is well researched and has reached the stage where large neural networks can be integrated onto a single chip. For instance, a feature detection step in ANVIL algorithms can be implemented using a single commercial chip [25]. System compatibility is another important issue. MOS circuits can be integrated with digital memories, winner take all circuits, and other elements of the system needed to design the ANVIL algorithms.

ACT devices are based on the gallium arsenide technology and can speed-up the neural network computation by two to three orders of magnitude. While this speed may not be necessary for the target recognition based on the visual images, it is a definite advantage when one considers other applications such as signal identification, processing of radar signals, spread-spectrum communication, or voice recognition. A neural network organization of systems developed for such applications can be performed, and training of such systems can be completed, using the algorithms developed in the ANVIL system. The most similar to ACT devices, in the type of operations performed, is the convolution chip recently designed in digital CMOS technology [26]. Like the ACT it can calculate a full inner product of the input signal vector and the stored tap weights in one clock cycle, however, it is limited to 24 taps with 4 bits accuracy (or 8 taps with 12 bit accuracy) and runs with a clock speed of 20 MHz. This is still two orders of magnitude worse performance than that of a single ACT device.

The guidelines for implementation of the ANVIL system can be stated as follows:

1. Support research efforts in application of the ACT technology. Anza Research Inc. from Cupertino, CA has recently completed the first phase of SBIR project "Neural network image recognition using acoustic charge transport devices" [10]. Their research has demonstrated that the ACT devices can perform like neural network processing cells. They are willing to undertake the research in practical implementation of the ANVIL algorithms as the second phase of their SBIR project.
2. Support the MOS realization of the ANVIL algorithms using commercial chips and developing other chips necessary for memory scanning combined with the winner take all operations. Chips which combine the memory scanning with the winner take all will be useful elements of either system, as the ACT technology is limited to the programmable convolution evaluation and at present must be complemented with MOS chips.

It should be realized that the MOS and ACT technology interface is not an easy task. Reaching the maximum achievable speed of the ACT devices may require the development of special memory buffers and the memory access time of the MOS part must be fast enough to support the assumed processing rate. If no practical options on fast memory can be identified, alternative solutions such as multiplexing between ACT devices could be investigated.

ACKNOWLEDGEMENTS

I wish to thank to the Air Force Office of Scientific Research for sponsorship of this program and to Research and Development Laboratories for its administration. John Mayhan and Dale Nelson of the Information Processing Technology Branch in the Avionics Directorate defined the research areas and provided support and technical discussions. Dale Nelson was also instrumental in identifying additional software support for this research and establishing contact with researchers in Anza Research Inc. Brian Andrews was very helpful in facilitating the use of various computing resources.

Dr. John Hines and Darrel Barker of the Design Branch in the Solid State Electronics Directorate provided information and assistance in using the system level design and simulation tools necessary for successful implementation of the studied architectures. I would like to thank Dr. Jim Noyes from Wittenberg University for stimulating discussions. The entire staff of the Information Processing Technology Branch has provided a very enjoyable working atmosphere for this research.

REFERENCES

[1] ANVIL 1990-1991 Interim Progress Report, Booz, Allen & Hamilton Inc., Arlington, VA, 1991.

[2] R. Perfetti, "Winner-take-all circuit for neurocomputing application," IEE Proc., Pt. G, vol. 137, no. 5, pp. 353-359, Oct. 1990.

[3] L. G. Johnson and S. M. S. Jalaleddine, "MOS implementation of winner-take-all network with application to content-addressable memory," Electronics Lett., pp. 957-958, July 18, 1991.

[4] J. Lazzaro, S. Ryckebusch, M. A. Mahowald, and C. A. Mead, "Winner-take-all networks of O(N) complexity," in advances in Neural Information Processing Systems 1, D. Touretzky, Ed., San Mateo, CA: Kaufmann, 1989, pp. 703-711.

[5] M. J. Hoskins, "Acoustic Charge Transport in Gallium Arsenide", Ph.D. Thesis, University of Illinois, 1983.

[6] J. E. Bales, M. J. Hoskins, and P.H. Sahm, "A GaAs ACT/IC programmable wide-band analog signal processor", 1990 IEEE GaAs IC Symposium, pp. 23-26, Oct. 1990.

[7] E. G. Bogus, M. J. Hoskins, and B. J. Husinger, "Input I-V and sampling time characteristics of the ACT device", IEEE Transactions on Electron Devices, vol. 38, no. 4, pp. 831-839, April 1991.

[8] E. G. Bogus, M. J. Hoskins, and B. J. Husinger, "Sampling time effects in the ACT device", IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control, vol. 38, no. 4, pp. 344-349, July 1991.

[9] R. W. Miller and R. J. Kansy, "Acoustic charge transport digitally programmable transversal filter development", 1990 IEEE MTT-S Symposium Proceedings, vol. 3, pp. 1111-1114, May 1990.

[10] P. D. Wasserman, "Neural Network Image Recognition using Acoustic Charge Transport Devices," DARPA SBIR Report, Contract # DAAH01-91-C-R281, Feb. 1992.

[11] H. P. Graf and L. D. Jackel, "Analog electronic neural network circuits," IEEE Circuits and Devices Mag., July 1989, pp. 44-45.

[12] K. A. Boahen, P. O. Pouliquen, A. G. Andreou, and R. E. Jenkins, "A heteroassociative memory using current-mode MOS analog VLSI circuits," IEEE Trans. Circuits Syst., vol. 36, no. 5, May 1989, pp. 747-755.

[13] M. A. Maher, S. P. Deweerth, M. A. Mahowald, and C. A. Mead, "Implementing neural architectures using analog VLSI circuits," IEEE Trans. Circuits and Syst., vol.

36, no. 5, May 1989, pp. 643-652.

[14] P. Mueller, J. Van der Spiegel, D. Blackman, T. Chiu, T. Clare, J. Dao, C. Donham, T.-P. Hsieh, and M. Loinaz, "A programmable analog neural computer and simulator", Advances in Neural Information Processing Systems 1, 1989, pp. 712-719.

[15] M. Yasunaga, N. Masuda, M. Yagyu, M. Asai, M. Yamada, and A. Masaki, "Design, fabrication and evaluation of a 5-inch wafer scale neural network LSI composed of 576 digital neurons," Proc. Int. Joint Conf. on Neural Networks, vol. II, June 1990, pp. 527-535.

[16] B. Lee, B. J. Sheu, "Design of a neural-based A/D converter using modified Hopfield network," IEEE J. Solid-State Circuits, vol. 24, no. 4, Aug. 1989, pp. 1129-1135.

[17] H. P. Graf and D. Henderson, "A reconfigurable CMOS neural network," IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, Feb. 1990, pp. 144-145.

[18] A. Rodriguez-Vazquez, R. Dominguez-Castro, A. Rudea, J. L. Huertas, and E. Sanchez-Silencio, "Nonlinear switched-capacitor neural networks for optimization problems," IEEE Trans. Circuits and Syst., vol. 37, no. 3, Mar. 1990, pp. 384-398.

[19] A. Murray and A. V. W. Smith, "Asynchronous VLSI neural networks using pulse-stream arithmetic," IEEE J. of Solid-State Circuits, vol. 23, no. 3, June 1988.

[20] J. E. Tomberg and K. K. K. Kaski, "Pulse density modulation technique in VLSI implementation of neural network algorithms," IEEE J. of Solid-State Circuits, vol. 25, no. 5, Oct. 1990, pp. 1277-1286.

[21] N. H. Farhat, "Optoelectronic neural networks and learning machines," IEEE Circuits and Devices Mag., Sept. 1989, pp. 32-41.

[22] N. H. Farhat, "Optoelectronic analog of self-programming neural nets: architecture and methodologies for implementing fast stochastic learning by simulated annealing," Appl. Optics, vol. 26, no. 23, Dec. 1987, pp. 5093-5103.

[23] Y. Horio and S. Nakamura, "Analog memories for VLSI neurocomputing," in Artificial Neural Networks, E. Sanchez-Sinencio and C. Lau, Ed., New York, NY: IEEE Press, 1992.

[24] S. Jalaleddine and L. Johnson, "Associative IC memories with relational search and nearest-match capabilities," IEEE J. Solid-state Circuits, vol. 27, no. 6, pp. 892-900, June 1992.

[25] "80170NX Electronically Trainable Analog Neural Network", Intel Corp., 1991.

[26] D. Reuver and H. Klar, "A configurable convolution chip with programmable coefficients," IEEE J. of Solid-State Circuits, vol. 27, no. 7, July 1992.

WAVELET ANALYSIS OF ULTRASONIC SIGNALS

Theresa A. Tuthill
Assistant Professor
Electrical Engineering Department

University of Dayton
300 College Park
Dayton, OH 45469-0226

Final Report for:
Summer Research Program
Wright Laboratory

Sponsored by:
Air Force Office of Scientific Research
Wright Patterson Air Force Base, Dayton, OH

September 1992

WAVELET ANALYSIS OF ULTRASONIC SIGNALS

Theresa A. Tuthill
Assistant Professor
Electrical Engineering Department
University of Dayton

Abstract

Wavelet theory provides a new approach to time-frequency analysis of signals. Waveforms are decomposed into various scales of a single "wavelet." Dilations and translations of the wavelet provide increased resolution over standard Fourier transform techniques and are useful in detecting transient signals. The wavelet transform was applied to ultrasonic signals to enhance the detection of delaminations and other subsurface defects in materials and also to aid in data compression. Difficulties arise in interpreting the wavelet coefficients and correlating them to spatial deformities due to the time-variancy of the transform. A better understanding of wavelet decomposition was gained by analyzing fundamental signals such as impulses and pure sinusoids. An optimal class of wavelets still needs to be determined for the non-destructive evaluation of materials.

WAVELET ANALYSIS OF ULTRASONIC SIGNALS

Theresa A. Tuthill

Introduction

Ultrasound scanning in the non-destructive evaluation (NDE) of materials and composites has become increasingly useful in detecting subsurface delaminations. Research efforts continue to analyze echo scan lines using both time and frequency signal processing techniques. This study looks at a new analysis process, wavelet theory, for extracting and interpreting information from ultrasonic signals. The goal is to improve detection of localized signal fluctuations due to defects and also to reduce data storage requirements.

Wavelet theory, a recently developed mathematical framework, actually encompasses a variety of signal processing techniques [1]. Signals are modelled as a composition of dilated and translated versions of a single waveform or wavelet. As the envelope of the wavelet is shortened, the frequency increases. Thus, increased resolution is provided in both the time and frequency domains. Portions of wavelet theory are related to multi-resolution analysis, split spectrum processing with constant frequency-to-bandwidth ratio, sub-band coding, and other signal compression techniques.

An appreciation for wavelet transforms can be found in a comparison to the commonly used Fourier Transform. Fourier analysis interprets a signal as a summation of cosines with varying amplitude and frequency. Any abrupt change in the time signal is broadly distributed in the frequency range for a particular windowed segment. Frequency analysis of localized variations is therefore ill-suited for long segments. Reducing the window length leads to use a "short-time Fourier transform" which diminishes frequency resolution due to the Heisenberg principle. Wavelet theory, however, employs varying time windows and frequency bands, thus allowing for increased frequency resolution within narrow time gates.

The original objective of this research was to gain an understanding of the wavelet transform and to determine the usefulness of wavelets over standard techniques in flaw detection and data compression.

Theory

A typical wavelet, h(t), is shown in Figure 1. The set of scaled and shifted wavelets forming a wavelet family [1] can be described by

$$h_{a,b}(t) = \frac{1}{\sqrt{a}} h\left(\frac{t-b}{a}\right) \qquad (1)$$

where a represents the time scale factor, and b is the time shift. For a<1, the waveform is compressed, and the frequency is correspondingly increased. The resulting family of wavelets forms a basis function for the construction of a given time signal. The contribution of each wavelet is determined by the associated coefficient which is dependent on the wavelet's scale and time shift.

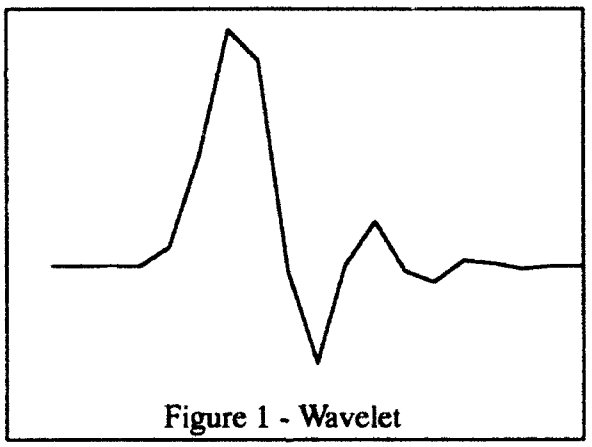


Figure 1 - Wavelet

The wavelet coefficients, Wt(a,b) are obtained from the correlation of the wavelet with the signal, f(t), by

$$Wf(a,b) = \langle f, h_{a,b} \rangle = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t)\, h\left(\frac{t-b}{a}\right) \partial t \qquad (2)$$

The reconstruction of the signal is then determined by the equation

$$f(t) = \frac{1}{C}\int_{0}^{\infty}\int_{-\infty}^{\infty} Wf(a,b)\, h\, \frac{\partial a\, \partial b}{a^2} \qquad (3)$$

where C is a constant dependent on the choice of wavelet.

The discretized wavelet takes the form [3]

$$h_{j,k}(n) = a_0^{-j/2} h(a_0^{-j} n - kT) \qquad (4)$$

where a_0 and T are arbitrary constants. However, if $a_0=2$ and T = 1, then the resulting dyadic wavelets form an orthonormal basis.

In the frequency domain, the wavelet transform behaves like a multi-frequency channel decomposition [4]. Each wavelet scale has a frequency bandwidth proportional to its center frequency (constant frequency-to-bandwidth ratio.) One of the advantages for a dyadic wavelet is the simple implementation of the discrete wavelet transform using digital filters.

The discrete transform coefficients can be plotted as sample points in the time-frequency domain. Along the time axis, the points are evenly distributed, but along the frequency axis, the points become closer as frequency increases. Higher frequencies have increased resolution. The time-frequency plot for the dyadic wavelet is shown in Figure 2.

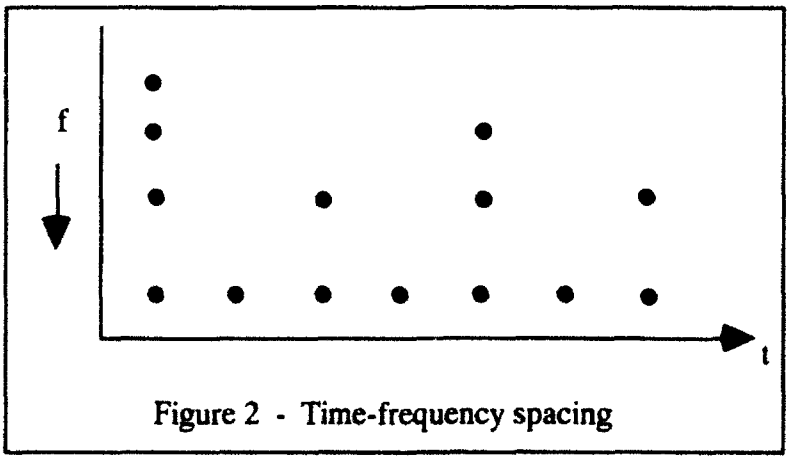


Figure 2 - Time-frequency spacing

The choice of a specific wavelet function is dependent on the application. For a lossless reconstruction, the wavelet must adhere to the following criteria [5]:

- admissibililty - the wavelet has zero-mean
- finite energy
- bandpass
- compactly supported
- orthonormal

Wavelets forming an orthonormal basis eliminate redundancy in the coefficients and can be useful in data compression. For a specific application, the scaling and shift parameters, a and T, and the wavelet function, h(t) must be optimized.

In essence, the wavelet transform is a multi-resolution analysis [6]; a given signal can be approximated using increasing resolution. For compactly supported orthonormal bases, each of these approximations can be written as a linear combination of a scale function, $\Phi_O(x)$,

$$[P_m f](x) = \sum_l c_{m,l}\Phi_{m,l}(x) \tag{5}$$

where c_m are the scale function coefficients. The difference between each approximation

$$Q_{m+1} = P_m - P_{m+1} \tag{6}$$

provides the high frequency detail, and can be written as a linear combination of wavelets, $\Psi(x)$,

$$[Q_m f](x) = \sum_l d_{m,l}\psi_{m,l}(x) \tag{7}$$

where d_m are the wavelet coefficients. The decomposition of a signal into low frequency scale functions and high frequency wavelets is the basis of wavelet transforms implemented through digital filters.

Discussion of the Problem

Ultrasound has long been used in the non-destructive evaluation of materials, especially in detecting sub-surface cracks, flaws, and delaminations. The effectiveness and accuracy of this technique is limited by the inability to detect flaws less than a wavelength and weak signal amplitude. Current practice relies on C-scan images which are derived from the peak amplitude of

the returned echo at a specific depth. Other analyses include examination of the frequency content based on the Fourier transform of windowed segments. The wavelet transform offers another analytic tool for two areas of continued research - material characterization and data compression

• Material Characterization

In theory, the wavelet transform should provide a useful representation of a time signal such that transient components related to material flaws are identifiable. The implementation of the transform and the choice of wavelet parameters will determine which signal properties are enhanced. The effect of a particular wavelet, its form, and the number of points used, must by examined.

Once a specific wavelet is chosen for the analysis, an understanding of the resulting transform coefficients must be developed. The transform images can then be interpreted to determine correlations with a sample.

• Data Compression

The objective of data compression is to reduce the number of points and/or the number of bits representing each point for ease of storage requirements. Transformations convert a data sequence of length N, a single point in N-dimensional space, to a point in M-dimensional space. An orthogonal transform inherently uses an equal number of points to represent a signal. For example, the Fourier Transform of an N-point sequence is and N-point complex sequence, yet because the real part is symmetric and the imaginary part is anti-symmetric, there are only N unique values. However, if there is a redundancy in the data, i.e. the signal is oversampled, many of the Fourier coefficients will be zero and can be eliminated. Thus, data compression occurs if transform coefficients can be discarded without creating a large error in the reconstruction process.

The wavelet transform, being orthogonal, retains the same number of data points in its representation of a signal. Yet, the decompostion of the signal at various frequency scales corresponds to sub-band coding [2] which is useful in data compression. Thus it should be possible to determine a reduced number of coefficients to adequately represent a signal.

Methodology

The multi-resolution technique for determining the wavelet transform [Eqs. 5-7] is amenable to implementation through digital filtering. Two filters are used to decompose the signal. The lowpass filter with impulse resonse, h(n), corresponds to the scale function, and the high pass filter with impulse response, g(n), corresponds to the wavelet function. After sub-sampling by two, the highpass output points, d_n, are the wavelet coefficients for that resolution level; the lowpass output points, c_n, are the scale function coefficients. The filtering process is then repeated on the lowpass output as shown in Figure 3. Note that at each increasing level the number of coefficients is halved.

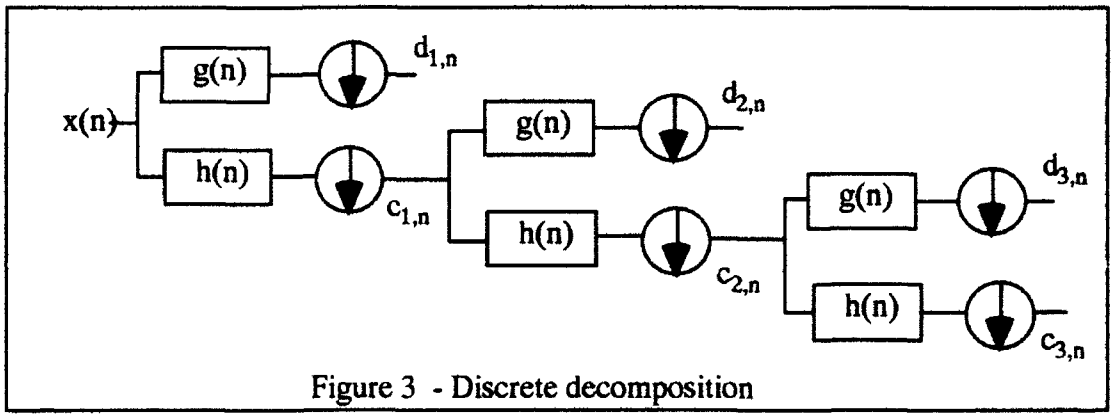


Figure 3 - Discrete decomposition

The choice of a specifc type of filter is determined by the wavelet and must meet specific conditions:

1) The filters must be quadrature mirror filters (QMF) for the wavelet coefficients to provide an exact reconstruction of the signal. By definition the QMFs satisfy the equation

$$|H(f)|^2 + |G(f)|^2 = 1 \tag{8}$$

2) The corresponding impulse responses are given by

$$\sum g(n) = 0 \tag{9}$$

$$\sum h(n) = \sqrt{2} \tag{10}$$

where the $\sqrt{2}$ normalization factor is due to the decimation factor in the filter.

3) The filters must be "regular" such that the filtering iterations converge.

In general, for a given lowpass filter, the corresponding highpass QMF is given by

$$g(n) = (-1)^n h(L-1-n) \tag{11}$$

and the frequency response is

$$G(f) = e^{-j2\pi f} H\left(f + \frac{1}{2}\right) \tag{12}$$

Based on maximally-flat low-pass filters, a variety of wavelets that meet the above criteria were computed by Daubechies for various filter lengths.

To test the wavelet transform, a graphite epoxy, quasi-isotropic sample was scanned ultrasonically using a 3.5 MHz (2" focused, 0.5" diameter) Aerotech transducer. The sample had a known 5^{lb} impact defect which was visible on the surface. After placing the sample in a water tank, the transducer was focused on the top surface. Using 100MHz sampling, a B-scan image was formed from 500 A-lines spaced 5 mil apart and containing 512 points each. A reflection from a flat metal plate was also taken as a reference. The data was then stored for later analysis.

Displaying the wavelet transform is difficult because it is two dimensional, a function of both time and frequency. By defining the squared modulus of the transform as a scalogram, the energy of the signal can be represented using gray scale on a time-frequency plot. Time or distance is shown on the horizontal axis and increasing frequency is shown on the vertical axis going from top to bottom. Because the number of wavelet coefficients decreases with decreasing frequency, the lateral resolution decreases at each successive level, and more pixels are assigned to a single coefficient. The highest resolution occurs for the highest frequency and is displayed on the bottom row. Any localized or transient phenomena should appear here.

After creating scalograms from scan lines through actual samples, it became apparent that a basic understanding of resulting image patterns was necessary to aid in interpretation. Wavelet transforms were made of some basic functions (i.e. impulse and sinusoids) and comparisons were made on the resulting scalograms.

Results

• Material Characterization

An ultrasonic scan line through the delamination of the graphite epoxy sample is shown in Fig. 4. The pulse propagated from right to left, and the front echo is visible on the right. The back echo is lost in the noise.

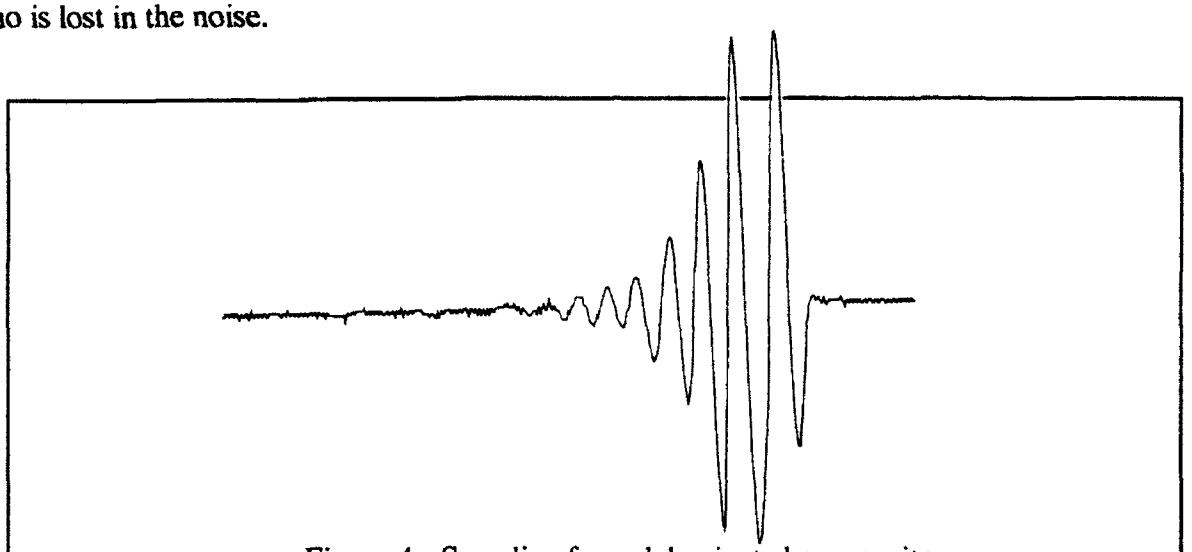


Figure 4 - Scan line from delaminated composite

Scalograms of the above scan line were computed using two different Daubechies wavelets. Figure 5 was configured using 6 filter coefficients (referred to as the D3 filter), and Figure 6 is derived from the 8 coefficient filter (D4).

Figure 5 - D3 scalogram of composite scan line.

Figure 6 - D4 scalogram of composite.

The higher frequencies have the most resolution and are located at the bottom of the scalogram. The D4 scalogram has improved resolution because the longer filter length increases the filter sharpness. Note in both the darker bands occuring at the 4th and 5th level. These correspond to the center frequency of the insonifying pulse, 3.5 MHz. The scalogram can be interpreted by referring to the center frequencies for the individual wavelets (which serve as bandpass filters) at each level (Table 2).

Level	1	2	3	4	5	6	7	8
Freq. (MHz)	38	19	9.5	4.7	2.3	1.2	0.6	0.3

Table 1 - Center frequencies for dyadic decomposition.

To further examine the effect of the choice of wavelet, the bandpass reference waveform was sampled at a low rate, shifting it to a high pass wavelet. The coefficients were adjusted slightly to meet the impluse response criteria (Eqs. 9 and 10). The resulting scalogram is shown in Figure 7. Little difference is noted from that of the D3 wavelet.

Figure 7 - Scalogram using the reference waveform

To further understand the wavelet transformation, scalograms were made of an impulse function (Figs. 8 and 9). An impulse has all frequencies present, and the scalogram was expected to spread out symmetrically with decreasing frequency, appearing as an upside down triangle. However, due to the causality of the filters, the frequencies are only shown spreading to the left and are not symmetric. Figure 10 shows a D4 scalogram of a shifted impulse. The pattern is noticeably different from Figure 9, demonstrating the time-variancy of the wavelet transform.

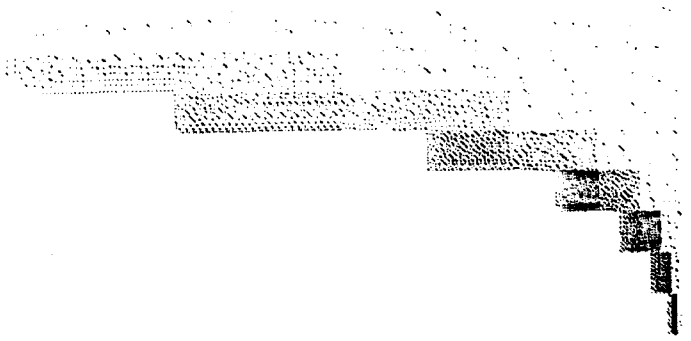


Figure 8 - D3 scalogram of an impulse.

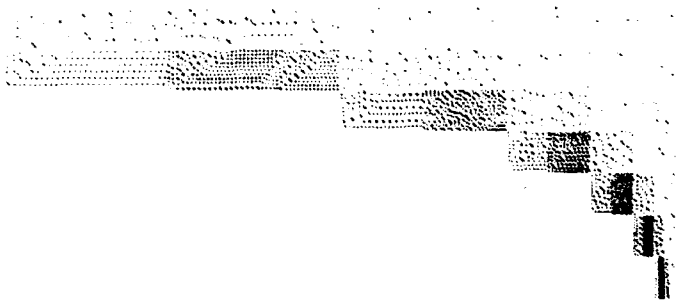


Figure 9 - D4 scalogram of an impulse.

Figure 10 - D4 scalogram of a shifted impulse.

The effect of two impulses was also examined. Non-zero points were placed side by side and gradually moved apart until they were resolvable in the scalogram. Figure 11 shows the D3 scalogram for two impulses separated by 5 sample points. The rightmost sample has twice the magnitude as the left. Figure 12 is the corresponding D4 scalogram, and the points are just bearly resolvable.

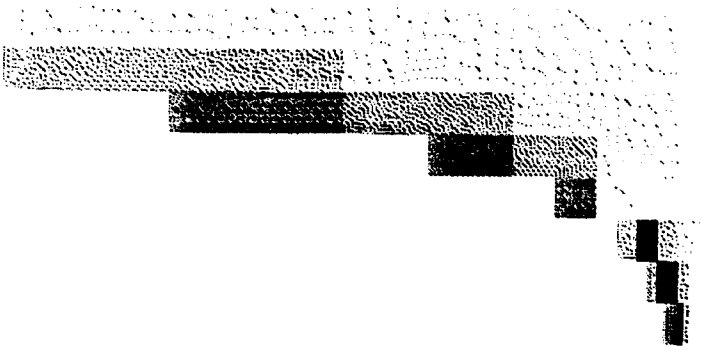


Figure 11 - D3 scalogram of two impulses.

Figure 12 - D4 scalogram of two impulses.

Further separation shows the two impulses readily resolvable in a D3 scalogram (Figure 13, separation = 15 points). By switching the magnitudes ot the two impulses (FIgure 14), such that the leftmost is double in magnitude, a new pattern is produced. The left side of the "plume" is visibly darker.

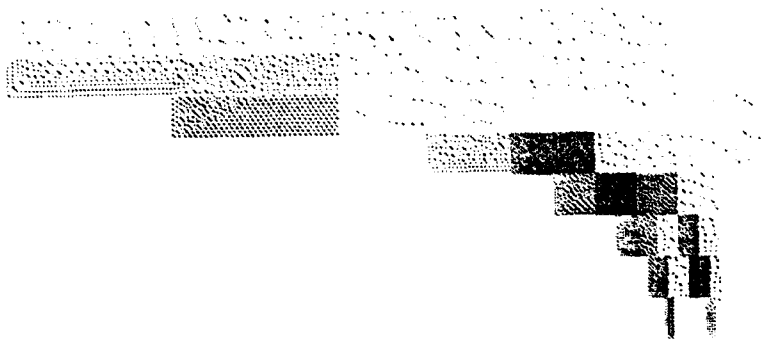


Figure 13 - D3 scalogram of two impulses (d=15).

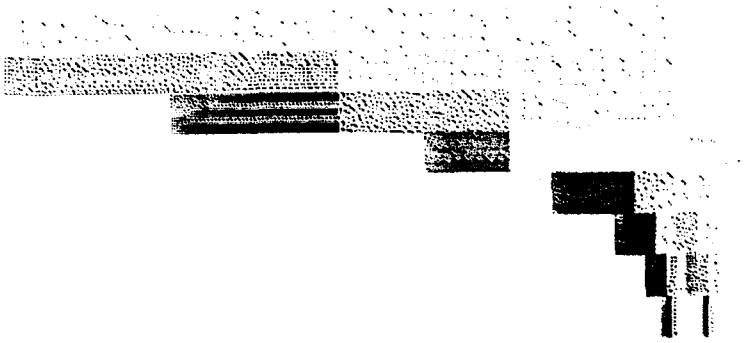


Figure 14 - D3 scalogram of two impulses (d=-15).

Finally, the scalograms of pure sinusoids were also analyzed. A 3.5 MHz signal is shown in Figure 15. Again, the 4th and 5th levels are the dominant bands. Referring to Table 1, 3.5 MHz falls between the two center frequencies for these levels. The scalogram's periodicity along the horizontal axis demonstrates that the time variancy of the transform is in fact phase information.

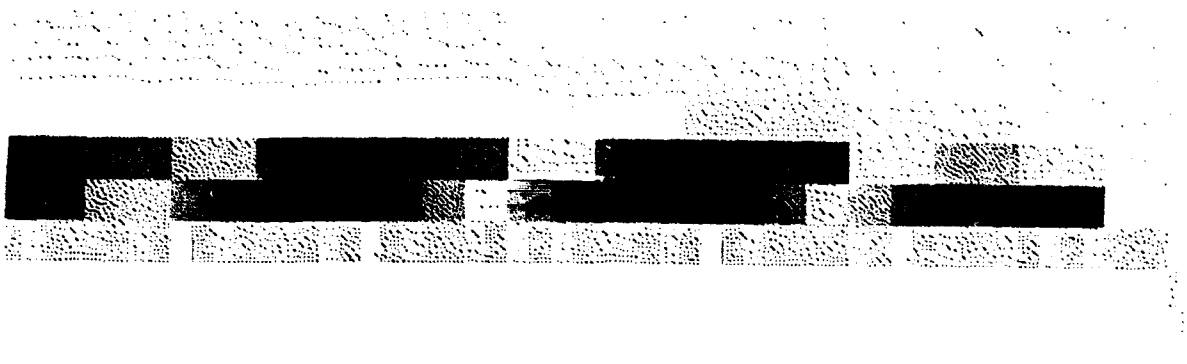


Figure 15 - D3 scalogram of 3.5 MHz.

Scalograms were done for a variety of frequencies. With frequencies related directly to the center frequency of a level, a single band stands out. Figure 16 shows a D3 scalogram for a 2.3 MHz signal, the center frequency for level 4. Figures 17 and 18 give the corresponding D4 scalograms for 3.5 and 2.3 MHz, respectively. Here, no significant difference between the wavelets is noted.

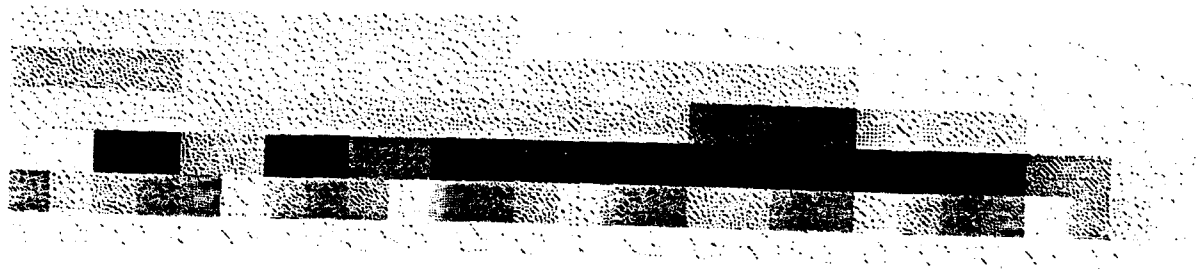


Figure 16 - D3 scalogram of 2.3 MHz.

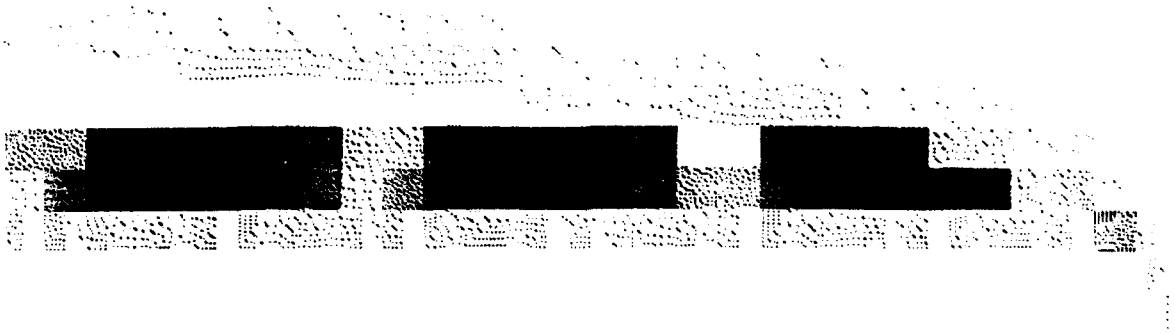


Figure 17 - D3 scalogram of 3.5 MHz.

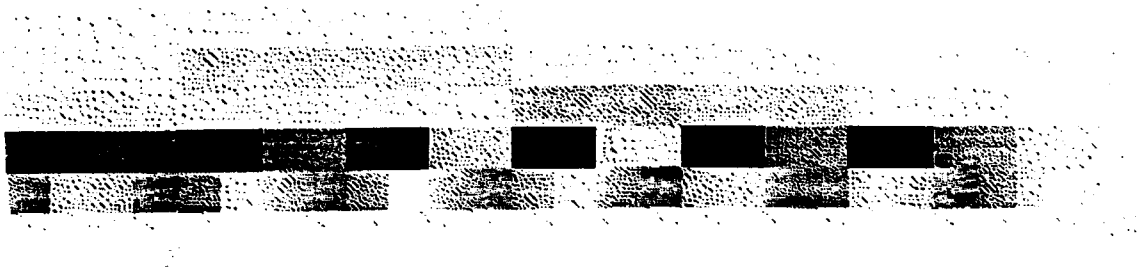


Figure 16 - D4 scalogram of 2.3 MHz.

• Data Compression

The wavelet transform using an orthogonal basis also employs N coefficients to represent an N-point sequence. However, in the case of the dyadic wavelet, the removal of the high frequency coefficients is equivalent to lowpass filtering or subsampling. High frequency coefficients were removed from both the wavelet (WT) and Fourier (FFT) transform coefficients of the ultrasonic signals from the composite sample. The signals were then resynthesized, and the mean square-error computed (Table 2). The wavelet transform showed no improvement over the Fourier technique unless more than 7/8 of the coefficients were zero-padded.

Data pts	FFT	WT
1:2	0.24	0.31
1:4	0.34	0.89
1:8	0.81	13.4
1:16	151.6	91.7

Table 2 - MSE comparison of FFT and WT after reducing the number of coefficients.

Conclusions

The wavelet transformation provides another useful means of analyzing signals. Its effectiveness, however, is limited by the interpretation of the wavelet transform coefficients. The transform incorporates both magnitude and phase information into a single image, and correspondingly is time-variant. The choice of wavelet also effects the transform and should be optimized for a specific use. With ultrasound scanning, the insonifying pulse does resemble a wavelet and can be used to detect flaws and delaminations. Use of longer filters in the decomposition of a signal will also improve the overall resolution.

As a data compression technique, the wavelet transform alone does not offer any advantage over the Fourier transform. However, it does offer potential if combined with a coding technique, such as Huffman coding.

Further examination of wavelet coefficients from ultrasound scans and their correlation with the insonified medium is necessary. Techniques such as the application of neural networks, could enhance the wavelet transforms utility.

Acknowledgements

The author would like to thank Dr. P. Kapur, B. Frock, Dr. T. Moran, Dr. C. Bhagat, and Dr. G. Warhola for their assistance in this study.

References

[1] Chui, C.K. An Introduction to Wavelets, (Academic Press, San Diego, 1992)

[2] Rioul,O. and Vetterli, M.,"Wavelets and signal processing," IEEE Signal Processing Magazine, Oct. 1991, pp. 14-38

[3] Daubechies, I., "Orthonormal bases of compactly supported wavelets," Comm. Pure Appl. Math., V. 41, pp. 909-996, 1988

[4] Mallat, S.G., "Multifrequency channel decompositions of images and wavelet models", IEEE Trans. Acoust., Speech, Signal Processing, V. 37, pp. 2091-2110, 1989

[5] Bliven, L.F. and Chapron, B., "Wavelet analysis and radar scattering from water waves," Naval Research Reviews, V. 41, pp. 11-16, 1989

[6] Mallat, S.G., "A theory for multiresolution signal decomposition: the wavelet representation," IEEE Trans. Pattern Anal. Machine Intell., V. 11, pp. 674-693, 1989

MBE SURFACE KINETICS OF SEMICONDUCTORS- A STOCHASTIC MODEL STUDY

R. Venkatasubramanian
Assistant Professor
Department of Electrical and Computer Engineering

University of Nevada - Las vegas
4505, Maryland Parkway
Las Vegas, Nevada 89154

Final Report for:
Summer Research Program
MLPO/WL
Wright Patterson Air Force Base

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

August, 1992

MBE Surface Kinetics of Semiconductors- A Stochastic Model Study

R. Venkatasubramanian

Abstract

A stochastic model for the MBE growth of Ge is developed based on the master equation approach with solid-on-solid restriction and quasi-chemical approximation. The surface kinetic processes included are: adsorption, evaporation and intralayer and interlayer surface migrations. The growth rate, the time averaged surface roughness and the time averaged intensity of reflection high energy electron diffraction (RHEED) (using kinematical theory of electron diffraction) were obtained for the MBE growth of Ge with the temperature in the range 423 – 698°K and a flux of 1 Å/sec. The growth rate was observed to be 1 Å/sec implying unity sticking coefficient. The time averaged surface roughness and isolated terrace adatom concentration which are good indicators of the average surface roughness, are found to be independent of the substrate temperature below 473°K and above 648°K. In the intermediate temperature range, the isolated terrace adatom concentration and the surface roughness decrease with increasing temperature. The kinetic roughening temperature above which a smooth surface remains smooth, is identified from the temperature dependence of the time averaged surface roughness, terrace adatom concentration and the RHEED intensity, as 648°K. The temperature dependence of the time averaged RHEED intensity and the kinetic roughening temperature obtained from this study compare favorably with the experimental results. The phenomenon of kinetic surface roughening is explained in terms of competition between the incorporation of Ge atoms on the surface which is a surface roughening process and the migration of Ge atoms to energetically favorable island edge sites which is a surface smoothing process.

1. INTRODUCTION

The evolution of surface morphology of epilayers during the MBE growth depends critically on the relative rates of the incorporation of atoms and the surface migration of these atoms to stable surface sites. The growth surface morphology of these epilayers determine the roughness of the hetero-interface in a heterostructure device. A rough interface degrades the optical and transport properties of the structure and hence the performance of the device. Understanding the mechanisms behind the surface roughening and their relation to the growth parameters is vital to grow high quality device structures reproducibly.

Due to the potential use of Ge in Si-Ge heterostructure devices, the MBE growth of Ge has been studied experimentally by various groups [1-12]. Chason et al in their experimental study [1], have focussed on the time evolution of the surface morphology of Ge (001) surface during the growth as a function of the substrate temperature in the temperature range of 100-550°C using the steady state intensity of RHEED beam as a measure of the growth surface roughness. It was observed that at a growth rate of $0.035 - 0.055 \mathring{A}$/sec., the kinetic surface roughening temperature is 375°C. Above 375°C, the surface remains smooth during growth. This observation was interpreted in terms of a phenomenological model of crystal growth based on the competition between two surface kinetic processes, the roughening process of incorporation of atoms and the smoothing process of migration of atoms to stable sites. In a later study [2], they employed ion bombardment of the growth surface during growth and showed that low energy Argon ions cause smoothing of the growth surface.

Venkatasubramanian [13,14] has developed a stochastic model for the MBE growth kinetic studies of compound semiconductors based on the master equation approach with solid-on-solid restriction and quasi-chemical approximation. The diamond cubic and two sublattice nature of the crystal were taken in to account in the modeling. He has considered two applications of the stochastic model - kinetics of MBE growth of GaAs like system and alloy clustering kinetics in a hypothetical Si-Ge like alloy. In the former application, he showed that the temperature dependence of growth rate and surface roughening can be explained in terms of the temperature dependence of surface migration. In the latter application, he showed that for the hypothetical alloy system chosen, the growth conditions are incompatible to grow both a smooth growth surface and an homogeneous alloy.

In this work, the stochastic model for the MBE growth described in Ref.[13] is modified to suit the elemental diamond cubic crystal and em-

ployed the model for a theoretical study of the MBE growth surface roughening of Ge (001). This manuscript is organized as follows. In section 2, the modifications necessary to the stochastic model to make it suitable for the Ge system is presented. In section 3, evaluation of the model parameters using bulk properties of Ge and the available experimental data is discussed. The results of this study and a comparison to the experimental work of Ref.[1] are presented along with a discussion in section 4. Finally, in section 5, the conclusions of this study are presented.

2. MODIFICATIONS TO THE STOCHASTIC MODEL

The time evolution kinetic equations of MBE growth are described in terms of the primary macrovariables, the concentration of atoms in the n^{th} layer. C(n), and the atom-vacancy bond density in the n^{th} layer. Q(n) and is given by Eqs 8a and 8b of Ref.[13]. The secondary variables are: the vacancy concentration, $C_V(n)$, the second neighbor atom-atom bond density, $\tilde{N}_{GaGa2}(2n)$, and $\tilde{N}_{AsAs2}(2n+1)$, the second neighbor vacancy-vacancy bond density, $\tilde{N}_{VV2}(2n)$, and $\tilde{N}_{VV2}(2n+1)$, the total fraction of atoms available for interlayer migration, e^{Lo}, and the total fraction of sites available for the interlayer migration process, e^{Lp}.

Firstly, the equations given in Ref.[13] are for a two sublattice diamond cubic crystal (zinc blende). But, Ge is a single element diamond cubic crystal. Therefore, the number of primary and secondary macrovariables reduce by half as discussed below.

$$C_{Ga}(2n),\ C_{As}(2n+1) \longrightarrow C_{Ge}(n)$$

$$Q_{Ga}(2n),\ Q_{As}(2n+1) \longrightarrow Q_{Ge}(n)$$

$$\tilde{N}_{GaGa2}(2n),\ \tilde{N}_{AsAs2}(2n+1) \longrightarrow \tilde{N}_{GeGe2}(n)$$
$$\tilde{N}_{VV2}(2n),\ \tilde{N}_{VV2}(2n+1) \longrightarrow \tilde{N}_{VV2}(n) \qquad (1)$$

The primary and secondary variables are related as:

$$\tilde{N}_{GeGe2}(n) = 2C_{Ge}(n) - (1/2)Q_{Ge}(n)$$

and

$$\tilde{N}_{VV2}(n) = 2(1 - C_{Ge}(n)) - (1/2)Q_{Ge}(n) \qquad (2)$$

Secondly, the number of model parameters employed in the description of the kinetic processes decreases by a factor of two. The model parameters employed are: chemical potential, L_{Ge}, atom pair interaction energies, k_{Ge}, relaxation time constant, τ_{rGe}, intralayer migration time constant, $\tau_d^{intraGe}$, and interlayer migration time constant, $\tau_d^{interGe}$.

Thirdly, the process of interlayer migration of atoms between layers in the diamond cubic structure is different from that in the zinc blende structure as discussed below. In the diamond cubic crystal, atoms migrate between two (001) layers, because all the (001) layers belong to the same sublattice and such migrations do not result in any anti-site defects. In the case of the zinc blende crystal, there are two distinct sublattices, one for the anions and the other for the cations. If anti-site defects are not permitted, then the cations and anions migrate between any two (001) layers which belong to their sublattice. Such a difference in the migration process due to the structural difference results in modifications to the term, e^{L_p}, which is described by Eq 6 of Ref.[13] and corresponds to the total number of available sites for the interlayer migration process and is given by:

$$e^{L_p} = \sum_{-\infty}^{\infty} \left[\frac{N_{GeGe2}(n-1)}{2} - C_{Ge}(n) \right] \neq 1 \qquad (3)$$

Due to changes discussed in Eqs 1-2, the term e^{L_o} which is described by Eq 7 in Ref.[13], modifies to:

$$e^{L_o} = \sum_{-\infty}^{\infty} \left[C_{Ge}(n) - 2C_{Ge}(n+1) + \frac{\tilde{N}_{GeGe2}(n+1)}{2} \right] \left[\frac{\tilde{N}_{GeGe2}(n)e^{-k_{Ge}} + \frac{1}{2}Q_{Ge}(n)e^{k_{Ge}}}{2C_{Ge}(n)} \right]^4 \qquad (4)$$

Even though, the incorporation process is complex, it is assumed in this model that the sticking coefficient of an atom arriving at a site having both its covalent bonds satisfied (Modified Solid on Solid restriction which is discussed in section IIA of Ref.[13]) is unity. Those Ge atoms arriving at a site where only one of the surface covalent bonds is satisfied, are allowed to get into a physisorbed state in which they are allowed to migrate to a proper nearest neighbor site where the MSOS is satisfied. Thus, all the Ge atoms arriving on the surface get incorporated resulting in unity sticking coefficient.

With the modifications given by Eqs 1-4, the time evolution equations for the primary macrovariables, $C_{Ge}(n)$ and $Q_{Ge}(n)$, are given by:

$$\frac{dC_{Ge}(n)}{dt} = \frac{1}{\tau_{rGe}}\left(\left[\frac{\hat{N}_{GeGe2}(n-1)}{2} - C_{Ge}(n)\right] e^{L_{Ge}} \quad \textbf{(A)}\right.$$

$$- \left[C_{Ge}(n) - 2C_{Ge}(n+1) + \frac{\hat{N}_{GeGe2}(n+1)}{2}\right]\left[\frac{\hat{N}_{GeGe2}(n)e^{-k_{Ge}} + \frac{1}{2}Q_{Ge}(n)e^{k_{Ge}}}{2C_{Ge}(n)}\right]^4\Bigg) \quad \textbf{(B)}$$

$$+ \frac{4}{\tau_d^{interGe}}\left(\left[\frac{\hat{N}_{GeGe2}(n-1)}{2} - C_{Ge}(n)\right] e^{L_o} \quad \textbf{(C)}\right.$$

$$- e^{L_p}\left[C_{Ge}(n) - 2C_{Ge}(n+1) + \frac{\hat{N}_{GeGe2}(n+1)}{2}\right]\left[\frac{\hat{N}_{GeGe2}(n)e^{-k_{Ge}} + 1/2Q_{Ge}(n+1)e^{k_{Ge}}}{2C_{Ge}(n)}\right]^4\Bigg) \textbf{(D)} \quad (5)$$

$$\frac{dQ_{Ge}(n)}{dt} = \frac{4}{\tau_{rGe}}\left(\left[\frac{\hat{N}_{GeGe2}(n-1)}{2} - C_{Ge}(n)\right] e^{L_{Ge}}\left[\frac{\hat{N}_{VV2}(n) - Q_{Ge}(n)}{2C_V(n)}\right] \quad \textbf{(E)}\right.$$

$$- \left[C_{Ge}(n) - 2C_{Ge}(n+1) + \frac{\hat{N}_{GeGe2}(n+1)}{2}\right]\left[\frac{\hat{N}_{GeGe2}(n)e^{-k_{Ge}} + 1/2Q_{Ge}(n)e^{k_{Ge}}}{2C_{Ge}(n)}\right]^4$$

$$\times \left[\frac{\hat{N}_{GeGe2}(n)e^{-k_{Ge}} - 1/2Q_{Ge}(n)e^{k_{Ge}}}{\hat{N}_{GeGe2}(n)e^{-k_{Ge}} + \frac{1}{2}Q_{Ge}(n)e^{k_{Ge}}}\right]\Bigg) \quad \textbf{(F)}$$

$$+ \frac{3}{\tau_d^{intraGe}}\left(\left[\frac{\frac{\hat{N}_{GeGe2}(n-1)}{2} - C_{Ge}(n)}{C_V(n)}\right] e^{k_{Ge}Q(n)}\right.$$

$$\times \left[\frac{C_{Ge}(n) - 2C_{Ge}(n+1) + \frac{\hat{N}_{GeGe2}(n+1)}{2}}{C_{Ge}(n)}\right]\left[\frac{\hat{N}_{GeGe2}(n)e^{-k_{Ge}} + \frac{1}{2}Q_{Ge}(n)e^{k_{Ge}}}{2C_{Ge}(n)}\right]^3$$

$$\times \left[\frac{\hat{N}_{GeGe2}(n)e^{-k_{Ge}} - \frac{1}{2}Q_{Ge}(n)e^{k_{Ge}}}{\hat{N}_{GeGe2}(n)e^{-k_{Ge}} + \frac{1}{2}Q(n)e^{k_{Ge}}} + \frac{\hat{N}_{VV2}(n) - Q_{Ge}(n)}{2C_V(n)}\right]\Bigg) \quad \textbf{(G)}$$

$$+ \frac{16}{\tau_d^{interGe}}\left(\left[\frac{\hat{N}_{GeGe2}(n-1)}{2} - C_{Ge}(n)\right] e^{L_o}\left[\frac{\hat{N}_{VV2}(n) - Q_{Ge}(n)}{2C_V(n)}\right]\right.$$

$$+ e^{L_p}\left[C_{Ge}(n) - 2C_{Ge}(n+1) + \frac{\hat{N}_{GeGe2}(n+1)}{2}\right]\left[\frac{\hat{N}_{GeGe2}(n)e^{-k_{Ge}} + \frac{1}{2}Q_{Ge}(n)e^{k_{Ge}}}{2C_{Ge}(n)}\right]^4$$

$$\times \left[\frac{\hat{N}_{GeGe2}(n)e^{-k_{Ge}} - \frac{1}{2}Q_{Ge}(n)e^{k_{Ge}}}{\hat{N}_{GeGe2}(n)e^{-k_{Ge}} + \frac{1}{2}Q_{Ge}(n)e^{k_{Ge}}}\right]\Bigg) \quad \textbf{(H)}$$

$$- \frac{16}{\tau_d^{interGe}}\left(\left[\frac{\hat{N}_{GeGe2}(n-1)}{2} - C_{Ge}(n)\right]\right.$$

$$\times \left[C_{Ge}(n) - 2C_{Ge}(n+1) + \frac{\hat{N}_{GeGe2}(n+1)}{2}\right]\left[\frac{\hat{N}_{GeGe2}(n)e^{-k_{Ge}} + \frac{1}{2}Q_{Ge}(n)e^{k_{Ge}}}{2C_{Ge}(n)}\right]^4$$

$$\times \left[\frac{\hat{N}_{GeGe2}(n)e^{-k_{Ge}} - \frac{1}{2}Q_{Ge}(n)e^{k_{Ge}}}{\hat{N}_{GeGe2}(n)e^{-K-Ge} + \frac{1}{2}Q_{Ge}(n)e^{k_{Ge}}} + \frac{\hat{N}_{VV2}(n) - Q_{Ge}(n)}{2C_V(n)}\right]\Bigg) \quad \textbf{(I)} \quad (6)$$

respectively. Terms A-D correspond to the change of $C_{Ge}(n)$ due to the incorporation, evaporation, and interlayer migration processes, respectively. It is noted that the intralayer migration process does not alter the atom concentration in the layer as the atom stays within the layer. Terms E-H correspond to change of $Q_{Ge}(n)$ due to incorporation, evaporation, intralayer migration and interlayer migration processes, respectively. Term I takes care of the double counting involved in the description of interlayer migration in term H.

3. EVALUATION OF THE MODEL PARAMETERS

The model parameters for the present study of the MBE growth of Ge were obtained from the bulk properties of Ge and the MBE growth parameters of Ref.[1]. The atom pair interaction energies for the first and second nearest neighbors were chosen as 1.5 eV and 0.25 eV, respectively, based on the available bulk cohesive energy data. The frequency factors for the evaporation and migration processes were chosen as 1.0 x 10^{13}/sec. The activation energy for surface migration of Ge atoms was chosen as 1.3 eV which is a typical number reported for diamond cubic semiconductor surfaces [16]. In the temperature range chosen for this model, the above frequency factor results in negligible evaporation.

The MBE growth parameters for this study were obtained from the experimental data given in Ref.[1]. The growth temperature was chosen in the range 423 – 698°K and the flux rate was set at 1 Å/sec. The flux rate chosen is about 25 times larger than that employed for the experimental studies of Ref.[1]. Such a scaling up of the flux rate was performed to minimize the computer time for the growth of a typical epilayer of 6 monolayers. Since the growth rate was scaled up by a factor of 25, the frequency factors for the evaporation and migration processes were also scaled up by the same factor in order to maintain the kinetics of growth the same. It is noted that the substrate surface employed in this study is flat without any steps unlike the substrates with steps on the surface that were employed in the experimental work of Ref.[1].

The material and growth data discussed above were employed to

calculate the model parameters according to the procedure detailed in section IVB of Ref.[13]. The model parameters were obtained as a function of growth temperature and are listed in Table I.

4. RESULTS AND DISCUSSION

The time evolution equations given by Eqs 5a and 5b and the boundary conditions corresponding a flat substrate described by Eq 14 of Ref.[13] were solved numerically on a Silicon Graphics work station. The CPU time for a typical growth of 10 $\AA$ of Ge was about 2 hours. It should be pointed out that for the growths in the high temperature regime, the steady state was attained within the growth time of 10 seconds. For the low temperature growths, steady state was not reached within the growth time of 6 monolayers, but the time change of growth results with time was not significant.

Concentration profiles were obtained as a function of time for various growth temperatures. The growth rate, G, was obtained from the concentration profiles for various growth temperatures. The growth rate was observed to be constant with temperature at 1 Å/sec. In other words, the sticking coefficient of Ge is unity for temperatures in the range $423 - 698^{o}$K which is in agreement with the experimental results of Ref.[1].

The instantaneous growth surface roughness can be described in terms of the the atom-vacancy bond density, $Q_{Ge}(\mathrm{n})$, as [15]:

$$IRP(t) = \sum_{i=1}^{i=\infty} Q_{Ge}(i)(t) \tag{7}$$

where $IRP(t)$ is the instantaneous surface roughness parameter with arbitrary units. The time averaged surface roughness parameter, $TRP(\mathrm{T})$, is given by:

$$TRP(T) = \frac{1}{t_{gro}} \int_{t=0}^{t=t_{gro}} IRP(t)dt \tag{8}$$

where t_{gro} is the growth time. The $TRP(T)$ parameter was obtained for various growth temperatures. Another useful parameter which describes the roughness of the surface is the instantaneous isolated terrace adatom concentration, $ITAC(\mathrm{t})$. The $ITAC(\mathrm{t})$ is obtained from the growth data as

follows:

$$ITAC(t) = \sum_{n=1}^{n=\infty} C(n) \begin{pmatrix} 4 \\ 0 \end{pmatrix} \left(\frac{\tilde{N}_{GeGe}(n)}{2.0C_{Ge}(n)}\right)^{0} \left(\frac{Q_{Ge}(n)}{2.0C_{Ge}(n)}\right)^{4} \quad (9)$$

The time averaged isolated terrace adatom concentration, $TITACP$(T), for various temperatures can be obtained as:

$$TITACP(T) = \frac{1}{t_{gro}} \int_{t=0}^{t=t_{gro}} ITAC(t)dt \quad (10)$$

where t_{gro} is the growth time.

Plots of $TRP(T)$ and $TIT.ACP$(T) versus growth temperature are shown in Figures 1 and 2, respectively. The $TRP(T)$ and $TITACP$(T) parameters are independent of the growth temperature below 473°K and above 648°K. In the intermediate temperature range, both parameters decrease with increasing temperature.

These observation can be explained as follows. At temperatures below 473°K, intralayer and interlayer migration and evaporation rates are negligible. The Ge atoms incorporate at random sites and do not migrate to island edges due to low migration rates and therefore, the surface is very rough. In the temperature range below 473°K, the $TITACP(T)$ is constant and very high implying that there is a large number of isolated adatoms on the growing surface. In the intermediate temperature range, the intralayer and interlayer migration rates increase with increasing temperature, aiding in the formation of larger surface islands. Hence the smoothing out of the growth surface occurs with decreasing surface roughness. In the high temperature range above 648°K, the intralayer and interlayer migration processes attain the saturation (maximum possible) value owing to the finite number of available sites for migration process such as the island edges. Therefore, the surface roughness becomes independent of temperature and attains its minimum value.

The intensity of a specular spot (1° off Bragg) of reflection high energy electron diffraction system with 10 kV electron beam was calculated using kinematical theory of electron diffraction as a function of growth time

for various growth temperatures. The RHEED intensity as a function of time for high temperature growth was oscillatory with period corresponding to the growth of a monolayer. The reason for oscillatory RHEED intensity is that the starting surface was flat and stepped unlike in the work of Ref.[1]. The time averaged RHEED intensities, TRI(T) were calculated for various growth temperatures. A plot of TRI(T) versus growth temperature is shown in Figure 3. The TRI(T) is independent of temperature above 648°K and below 473°K, because the surface roughness is a constant due to the reasons already discussed. In the intermediate temperature range, since the surface roughness decreases with the growth temperature, the amount of destructive inference of the electron beams from various layers decreases, resulting in an increase in the average RHEED intensity.

The plot of TRI(T) versus temperature obtained in this study was compared with that of the experimental study of Chason et al [1]. The results compare favorably well. The kinetic surface roughening temperature above which a surface remains smooth, is obtained from the temperature dependence of TRP(T) and TRI(T) shown in Figures 1 and 2, respectively. The kinetic roughening temperature for the MBE growth of Ge thus obtained, is 648°K. In the experimental study [1] with a similar flux rate to migration rate ratio, the kinetic surface roughening temperature is obtained as 648°K also. Considering the simplicity of the model and the origin of the model parameters Which is the bulk properties, the agreement between the present study and the experimental work is excellent.

The experimental observations of steady state RHEED intensity were explained in Ref.[1] in terms of the the competition between the incorporation of Ge atoms which is a surface roughening process and the Ge atoms migrating to the step edges which is a surface smoothing process. This explanation is similar to the one given in this paper as discussed below. Surface roughening occurs at all temperatures due to random incorporation of atoms. At low temperatures, the surface migration is less and hence the surface smoothing effect is less. But, at high temperatures, the interlayer and intralayer surface migration, aid in atoms reaching energetically favor-

able sites. In the case of Ref.[1], the energetically favorable site is kink sites located at the step edges since the starting surface was a stepped surface. In this study, since the starting surface is a flat smooth surface, there are no step edges and therefore, the energetically favorable sites are island edges where more than one of the inplane second nearest neighbor bonds are satisfied. Thus, the stochastic model developed in Ref.[13,14] and used in this study for the MBE growth studies of Ge, accurately describes the kinetics of surface roughening of Ge and also aids in understanding the details of the mechanisms underlying the surface roughening phenomenon.

Here, it is appropriate to discus the advantages and limitations of the stochastic model in comparison to the existing theoretical models such as Monte Carlo simulations, Molecular Dynamics simulations and other models. The advantages are: Firstly, it is not size limited unlike Monte Carlo and Molecular Dynamics simulations and therefore, is suitable for substitutional doping kinetics studies. Secondly, for the same growth simulation, it takes much less CPU time compared to Monte Carlo and Molecular Dynamics simulations. Thirdly, the author has developed the stochastic model for growth of tetranary alloys in either 100 or 111 directions with either monoatomic or diatomic or tetratomic anionic molecular beam [17] which shows the versatility of the approach. Overall, the stochastic model can be employed for the macroscopic simulation of MBE growth of compound semiconductors to investigate such surface phenomena as ordering and clustering in alloys and surface roughening and incorporation of dopants. The disadvantages are as follows. Firstly, it is based on a rigid lattice gas model like Monte Carlo simulations and therefore it is difficult to develop a stochastic model for defect creation studies. But, the author has, recently, developed a stochastic model incorporating possibilities of stacking faults in the epilayer [18]. Secondly, the effect of surface reconstruction is completely ignored as is the case in all other models. Thirdly, it is based on a random distribution approximation weighted by energy of a surface configuration (like second approximation or Bethe-Pierles approximation) and does not provide step density information directly unlike the Monte Carlo simulations. But, by

making certain physically reasonable assumption, such growth information can be obtained.

5. CONCLUSION

A stochastic model of the MBE growth of elemental semiconductors such as Ge is developed. It is based on the master equation approach with solid-on-solid restriction and quasi-chemical approximation. The growth rate, the time averaged surface roughness and the time averaged RHEED intensity were obtained for various growth temperatures. The kinetic surface roughening temperature for the MBE growth of Ge at a flux rate of 1 Å/sec. is identified as 648°K from the temperature dependence of the time averaged surface roughness and the time averaged RHEED intensity which compare favorably with experimental results obtained under similar growth conditions[1]. The phenomenon of kinetic surface roughening is explained in terms competition between the incorporation of Ge atoms which roughens the surface and the surface migration of Ge atoms to energetically favorable site which is a smoothing process.

REFERENCES

1. E.Chason, J.Y.Tsao, K.M.Horn and S.T.Picraux, J. Vac. Sci. Tech., **B7(2)**, 332, (1989).

2. E.Chason, K.M.Horn, J.Y.Tsao and S.T.Picraux, Proceedings of MRS, Boston, December, (1988).

3. E.Chason, J.Y.Tsao, K.M.Horn, S.T.Picraux and H.A.Atwater, J. Vac. Sci. Tech., **A8(3)**, 2507, (1990).

4. S.S.Iyer, P.R.Pukite, J.C.Tsang and M.W.Copel, Proceedings of Fifth International MBE Conference, Sapporo, Japan, August, (1988).

5. S.S.Iyer, R.A.Metzger and F.G.Allen, J. Appl. Phys., **31**, 5608, (1981).

6. H.J.Gossman and L.C.Feldman, Phys. Rev., **B 32**, 6, (1985).

7. H.J.Gossman and L.C.Feldman, Appl. Phys., **A 38**, 171, (1985).

8. H.J.Gossman, L.C.Feldman and W.M.Gibso, Surf Sci., **155**, 413, (1985).

9. H.J.Gossman, J.C.Bean, L.C.Feldman, E.G.McRea and I.K. Robinson, Phys. Rev., **55**, 1106, (1985).

10. H.J.Gossman, J.C.Bean, L.C.Feldman, E.G.McRea and I.K. Robinson, J. Vac. Sci. Tech., **A 3**, 1633, (1985).

11. A.V.Rzhanov, S.I.Stenin, O.P.Pchelyakov, and B.Z.Kanter, Thin Solid Films, **139**, 169, (1986).

12. The above list on Ge MBE growth is not exhaustive, but provides a good starting point.

13. R.Venkatasubramanian, J. Matl. Research., **7**, 1222, (1992).

14. R.Venkatasubramanian, J. Matl. Research., **7**, 1236, (1992).

15. H.J.Leamy and K.A.Jackson, J. Appl. Phys., **42**, 2121, (1971).

16. J.H.Neave, P.J.Dobson, B.A.Joyce and J.Zhang, Appl. Phys. Lett., 47, 100, (1985).

17. R.Venkatasubramanian, (unpublished work)

18. R.Venkatasubramanian, D.L.Dorsey and S.G.Das, "A Theoretical Study of Amorphous-Crystalline Transition in Si 111 MBE Growth", to be presented at MRS, Boston, November, 1992.

Table I. The model parameters, flux parameter, L_{Ge} and adsorption and evaporation time constant, τ_{rGe} were kept equal to 69.0 and 1.4 x 10^30 so that the growth rate is 1 Å/sec. The surface diffusion parameter, τ_d, and the second nearest neighbor pair interaction energy parameters, k_{Ge}, for various substrate temperatures. It is assumed that $\tau_d^{intra} = \tau_d^{inter} = \tau_d$.

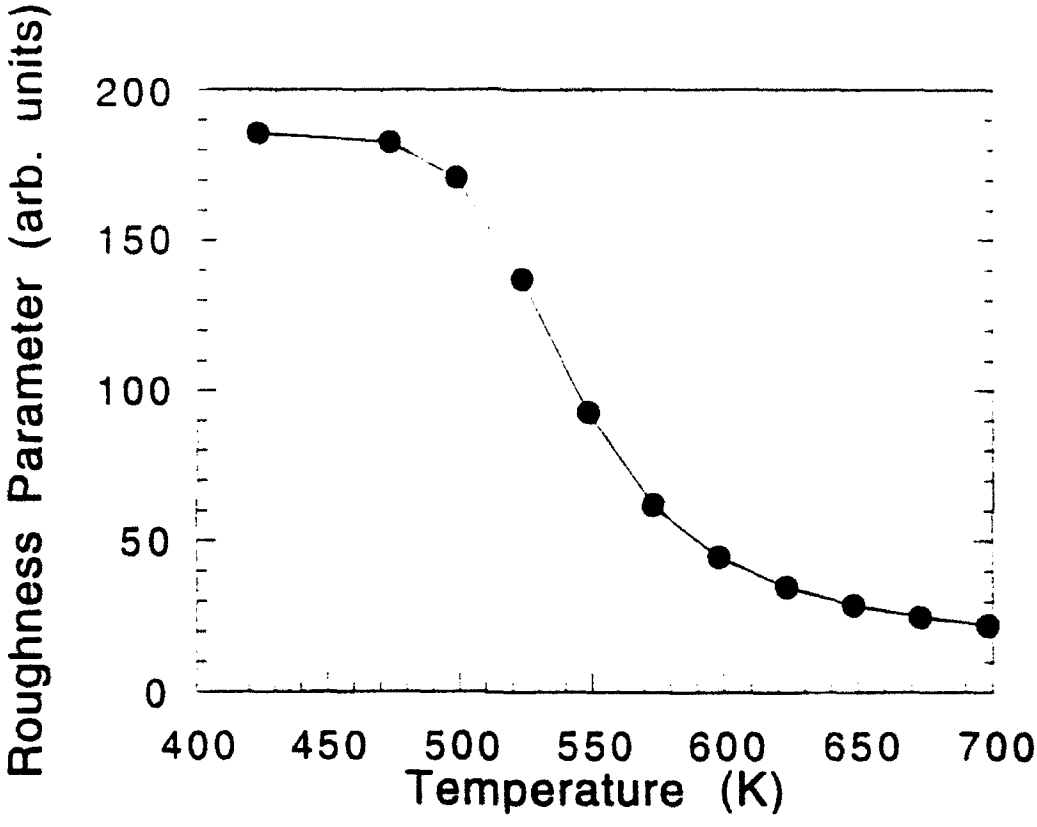
$T(^oK)$	$\tau_d(sec.)$	k_{Ge}
423	724.0	6.85
473	12.6	6.12
498	2.2	5.82
523	0.47	5.54
548	0.12	5.29
573	3.16e-2	5.06
598	9.67e-3	4.84
623	3.26e-7	4.65
648	1.19e-3	4.47
673	4.71e-4	4.30
698	1.98e-4	4.15

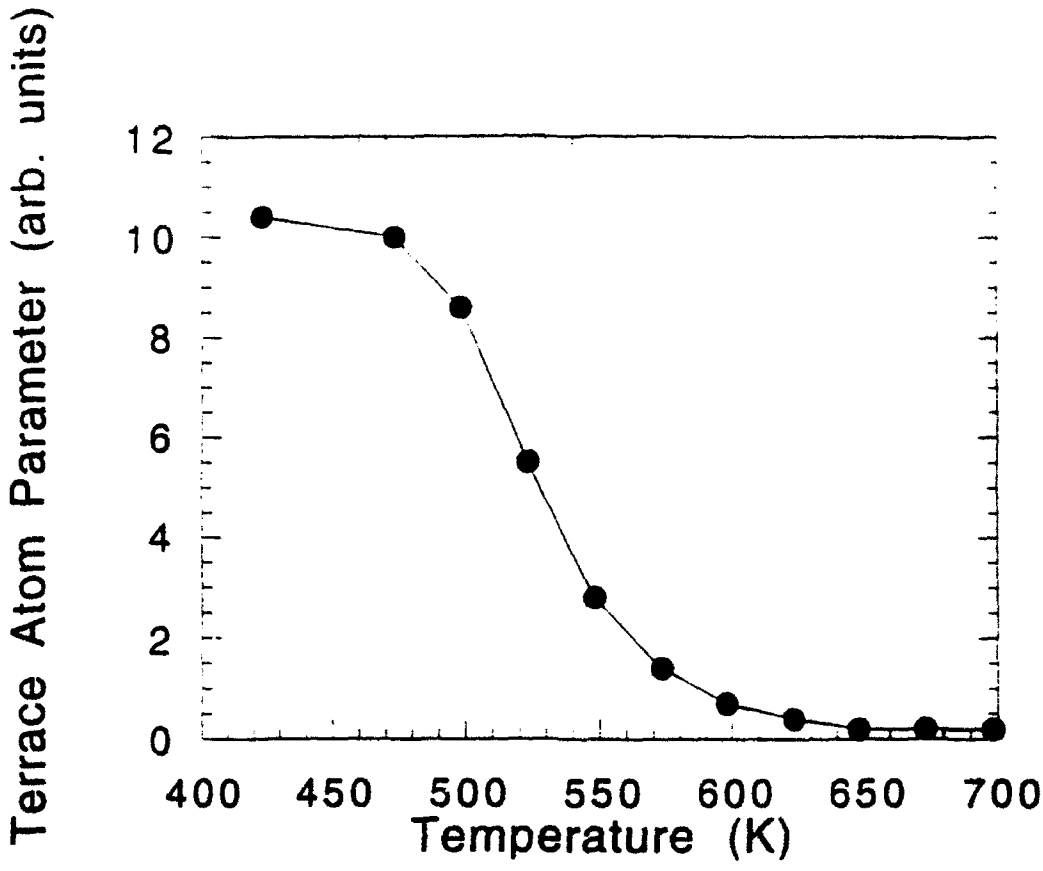
Figure Captions

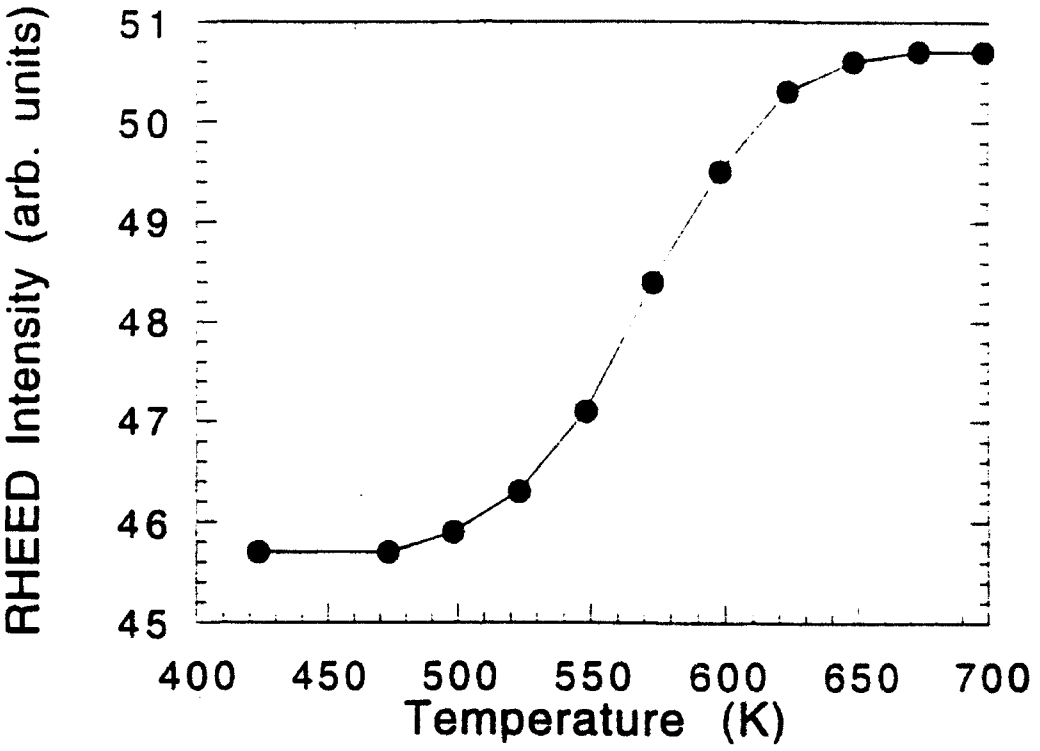
Figure 1. Plot of the surface roughness parameter, $TRT(T)$ vs temperature in the range of423 – 698°K.

Figure 2. Plot of the isolated terrace adatom concentration parameter, $TITACP(T)$ vs temperature in the range of423 – 698°K.

Figure 3. Plot of time averaged RHEED intensity, TRI(T) vs temperature in the range of423 – 698°K.


Roughness Parameter (arb. units)
200
150
100
50
0
400
450
500
550
600
650
700
Temperature (K)


Terrace Atom Parameter (arb. units)
12
10
8
6
4
2
0
400
450
500
550
600
650
700
Temperature (K)


RHEED Intensity (arb. units)
51
50
49
48
47
46
45
400
450
500
550
600
650
700
Temperature (K)

DEVELOPMENT OF A RESONANT DC LINK INVERTER FOR INDUCTION MOTOR DRIVE

Subbaraya Yuvarajan
Associate Professor
Department of Electrical Engineering

North Dakota State University
Fargo, ND 58105

Final Report for:
Summer Research Program
Wright Laboratory at Wright Patterson AFB

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

August 1992

DEVELOPMENT OF A RESONANT DC LINK INVERTER FOR INDUCTION MOTOR DRIVE

Subbaraya Yuvarajan
Associate Professor
Department of Electrical Engineering
North Dakota State University

Abstract

Resonant DC Link Inverters have negligible switching losses compared to the conventional hard-switched inverters. A three phase dc link inverter suitable for feeding an induction motor was developed. The complete operation of the resonant dc link inverter system was studied using the simulation package PSpice. The elements of the inverter like the link inductor and the link switch were designed and built. The control circuit for the three phase inverter was then developed. The experimental system with the inverter feeding the induction motor was tested. The principle of the resonant link was later extended to a single phase Pulse Width Modulated (PWM) inverter.

DEVELOPMENT OF A RESONANT DC LINK INVERTER FOR INDUCTION MOTOR DRIVE

Subbaraya Yuvarajan

INTRODUCTION

Variable speed induction motor drives are used in a number of industrial applications. High power density induction motor drives are likely to be used in aircraft control systems. The efficiency of the inverter feeding the induction motors can be improved through soft-switching by employing a Resonant DC Link (RDCL) [1] - [4]. The inverters operating on a resonant dc link can be made to switch at zero voltage. With the result, there are no switching losses. Also, the semiconductor devices realizing the switches need not be protected by snubbers. The RDCL inverters will be suitable for space power and aircraft applications. While the concept of RDCL is very attractive with a lot of application potential, there are some specific problems to be solved before building a reliable inverter. The major problem is to establish a stable dc link whose amplitude is independent of load fluctuations. The Electrical Laboratory at the Wright Patterson Air Force Base is interested in developing an RDCL inverter-fed induction motor drive for aircraft applications.

OBJECTIVE OF THE RESEARCH WORK

The aim of the summer research at Wright Patterson AFB was to understand the operation of the RDCL inverter through

PSpice simulation, to build an experimental inverter system, and to suggest possible modifications to improve its performance. Since the operation of RDCL is basically nonlinear, the analysis is to be carried out using a simulation package like PSpice [5]. The preliminary results obtained from simulation were to be used as a basis for the choice of switching devices and the design of the other components for the inverter system. The control circuit for a three phase full-bridge inverter was to be developed using digital logic gates and drivers. The operation of the experimental inverter feeding the induction motor was to be tested. Certain limitations of the circuit like the link-voltage overshoot and resonant frequency change with load current were to be studied. Some of the problems observed will be later studied in detail under the Air Force Office of Scientific Research Initiation Program (RIP) award.

BLOCK DIAGRAM OF INDUCTION MOTOR DRIVE

The block diagram of the complete induction motor speed control system is shown in Fig. 1. It is the same as that of a conventional inverter-fed system except for the resonant inverter in the place of a hard-switched inverter. The system is shown to contain standard speed and current controllers. However, induction motor drives tend to use field oriented controllers whose operation depends very much on digital signal processors (DSPs). In both the control schemes, the actual speed of the induction motor is compared

with the reference speed and the speed error is used to drive the inner current loop. A Proportional-plus-Integral (PI) controller and a function generator are used to derive the reference input for the current loop.

The frequency information from the speed loop is used to operate the inverter. The error in the current loop is used to operate the inverter switches in the correct sequence. The drive also includes the 'constant volt/Hz' implementation. The harmonic content in the output current is reduced by using Pulse Width Modulation (PWM).

RESONANT DC LINK

The input voltage to a conventional inverter is a smooth dc voltage obtained from a three phase rectifier and filter. The switching of the inverter on a stiff dc bus results in considerable power loss which increases with an increase in the switching frequency. In addition, the devices are to be protected by adequate snubber circuitry. The above problems are overcome by allowing the dc link to resonate at a chosen frequency and turning-on and off the inverter switches only when the link voltage goes through zero [1]. The arrangement of the resonant link is shown in Fig. 2. It consists of a series inductor and a shunt capacitor. The resonant frequency of the inverter is chosen to be much higher than the inverter output frequency with the result the harmonics introduced have negligible effect on the load.

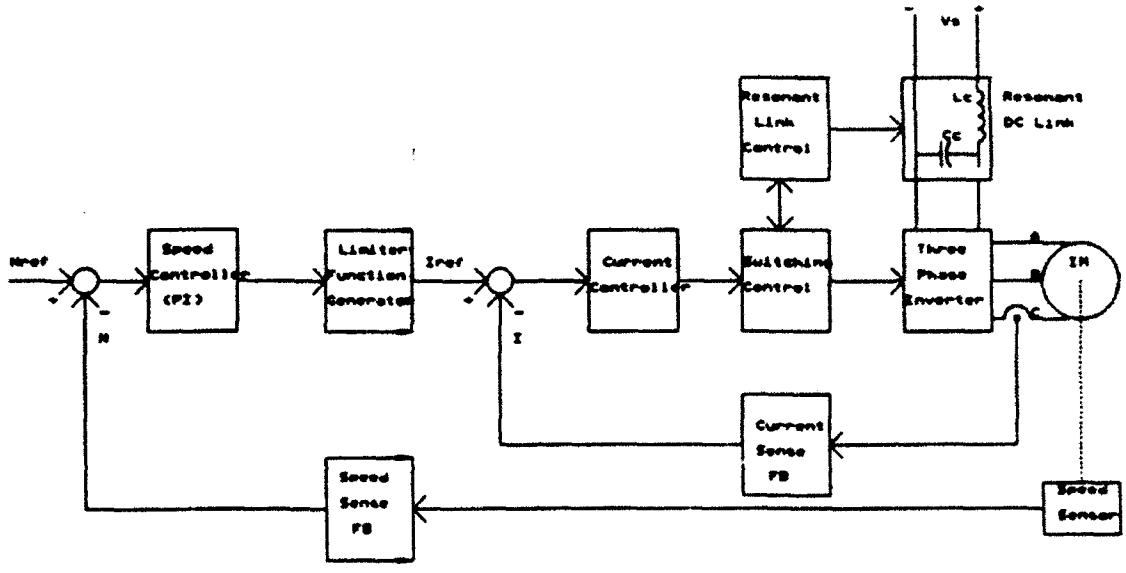


Fig. 1 Block diagram of induction motor drive

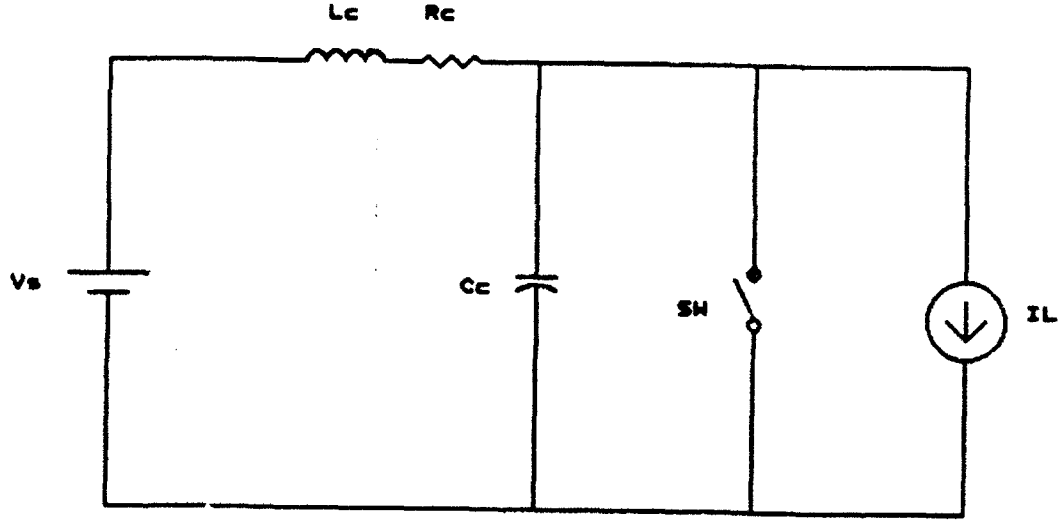


Fig. 2 Circuit diagram of resonant link

Before building the resonant link in the laboratory, it was simulated using PSpice. In the first phase of the simulation, the inverter was assumed to be a constant current source as shown in Fig. 2. The waveform of the link capacitor was obtained. It was noted that the capacitor voltage oscillates in the form of a damped sinusoid. In order to sustain the oscillations, the current through the inductor is to be varied as a function of the load current. A second order relation between the two currents will result in an undamped dc link.

An ideal switch was added across the link capacitor to provide the inductor with adequate initial current that helps to sustain the oscillations. The switch was kept closed for a fraction of the resonant period (small duty cycle). A duty cycle whose value lies between 10% to 15% was found to give a satisfactory performance. It was also found out that the resonant period varies with the load current in a nonlinear form.

The final simulation was carried out using single phase and three phase inverters connected to the resonant link. The load on the inverter was chosen to be of R-L type as well as a periodic current representing an induction motor. Several useful results were obtained from simulation. When the inverter switches are not synchronized with the link switch, irregular load voltage spikes were found to occur. It was also observed that relatively higher voltage spikes

occur when the load current undergoes a sharp transition. Therefore, the inverter switches are to be closed or opened only when the link voltage goes through zero. Some of the simulation waveforms are shown in Fig. 3.

The main problem with building the resonant link is to design an inductor with a very high 'Q' factor. A high 'Q' means that the L-C oscillations are fairly undamped, a feature essential in sustaining a resonant link. Several designs were tried before arriving at a satisfactory one. The air-cored inductor selected uses a one-inch well insulated copper plate [6]. The inductance value is 25uH with a Q of 45. The capacitor used is of polypropylene type with a value of 0.33uF. Together, the L-C circuit has a resonant frequency of about 38kHz.

The oscillations in the link have to be stabilized by using a switch across the capacitor. The switch is periodically closed thereby allowing the inductor current to build up. When opened after a small a time duration, the inductor current charges up the capacitor compensating the lost charge. The switch is realized using a Power MOSFET which is capable of operating at very high frequencies. The gate is supplied with a low duty cycle square wave generated by a 555 timer and a MOSFET driver SG 3626. The duty cycle is low and is adjustable depending on the load conditions.

THREE PHASE INVERTER

The complete control circuit of the Resonant DC Link

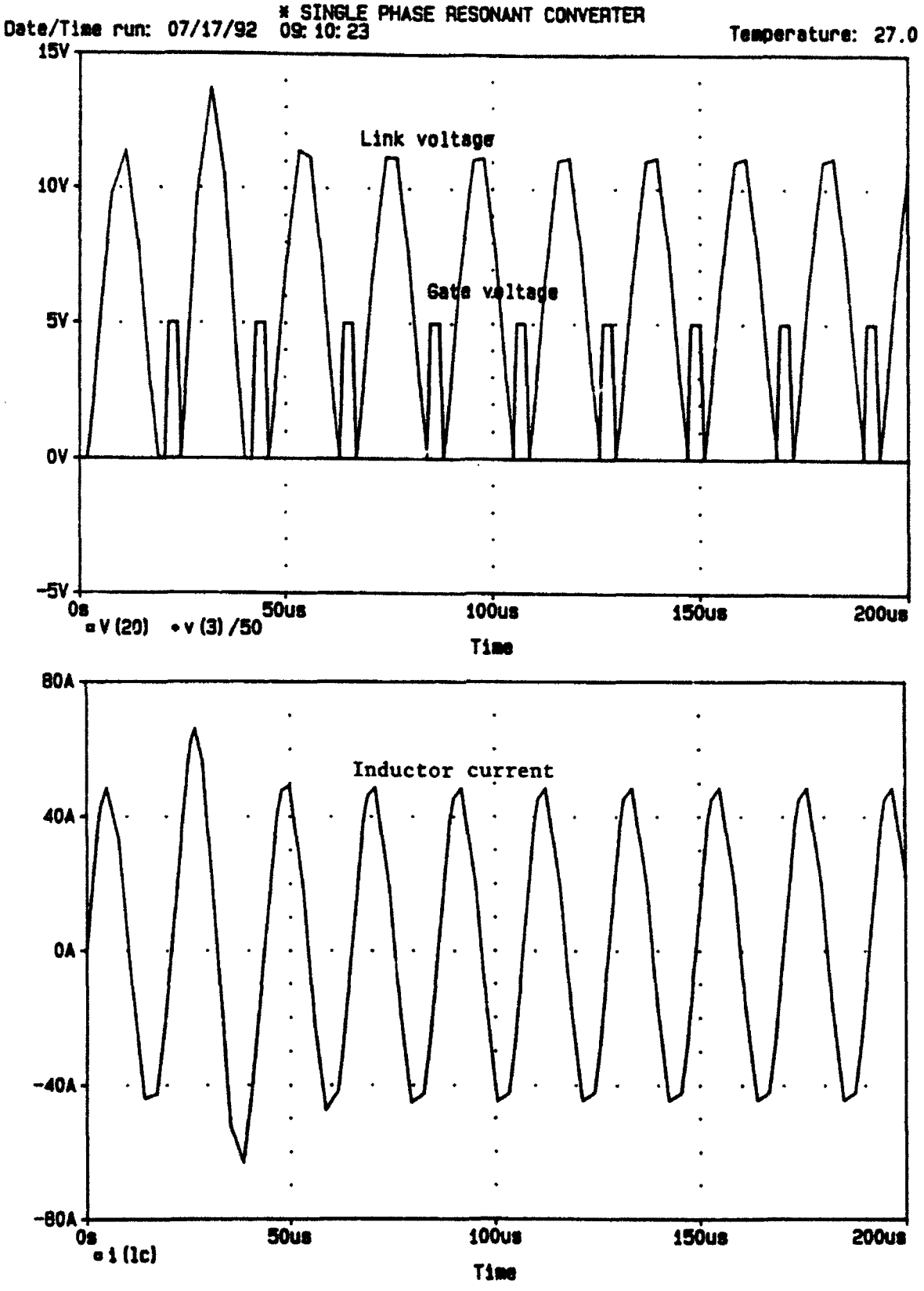


Fig. 3 Simulated waveforms of RDCL

inverter is shown in Fig. 4. The high frequency clock signal for the link switching MOSFET is generated by a 555 timer and applied through the driver SG 3626. The signal is also used to synchronize the inverter control signal. Since the MOSFET is switched on when the capacitor voltage is zero, it does not need any snubber.

The low frequency control signal with frequency f_o for driving the inverter switches is generated using a 555 timer and three J-K Flip-flops. The timer generates a clock signal whose frequency is triple the frequency ($6f_o$) of the desired output. The frequency is continuously variable using the potentiometer Rx. The flip-flops are interconnected to produce a balanced three phase square wave signal. Each of the three signals is passed through a D-type flip-flop for synchronization with the link switching signal to ensure zero-voltage switching. The output from the D flip-flops are passed through a set of logic inverters to get the complementary signals for the low and high side switches in each of the three phase groups.

The power circuit of the three phase inverter is shown in Fig. 5. The high side switches are to be isolated through special drivers like IR 2110. The operation of the MOSFET driver IR 2110 is sensitive to the amplitude of the input signal. It requires a signal level of around 15V for a driver supply voltage ($V_{CC} = V_{DD}$) of 20V. A 15V signal is obtained using open-collector inverters and pull up

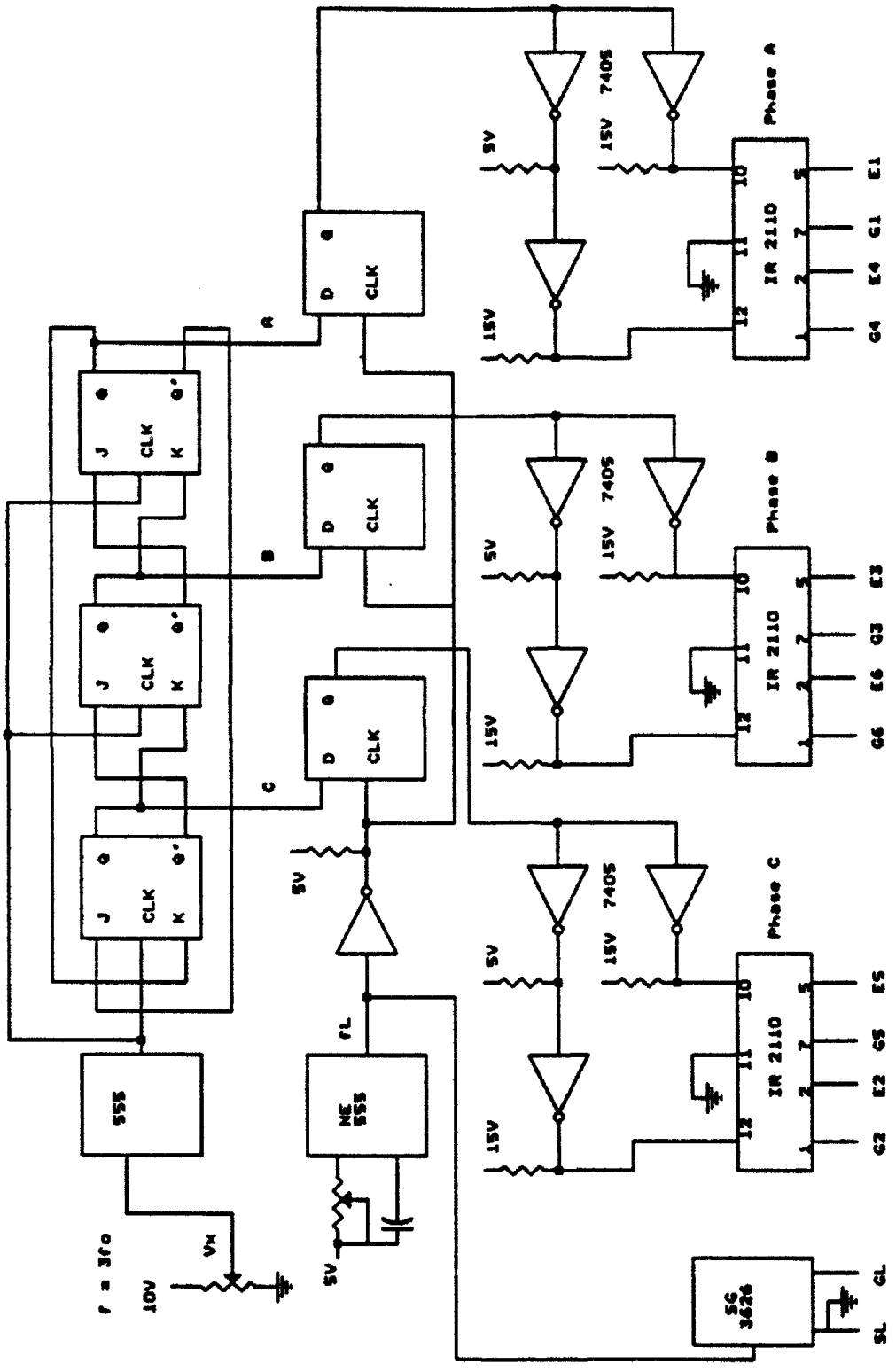


Fig. 4 Control circuit of RDCL inverter

resistors.

The complete inverter was built and tested in the Electrical Laboratory at the Wright Patterson AFB. First, the resonant dc link was tested without the inverter. After adjusting for the correct resonant frequency and link-switch duty cycle, the inverter was added. The inverter feeds an induction motor coupled to a blower. The experimental waveforms of the link-switch drive signal and the capacitor voltage recorded on a digitizing oscilloscope are shown in Fig. 6. The inverter drive signal, the load current, and the load voltage are shown in Fig. 7.

PULSE WIDTH MODULATION

In an inverter with Sine Pulse Width Modulation (SPWM) having P pulses per half cycle of the reference sine wave, all harmonics below (2P - 1) will be eliminated. The SPWM was implemented on a single phase resonant dc link inverter. The sine and triangular waveforms required for the SPWM were generated using two ICL 8038 waveform generator ICs [7] as shown in Fig. 8. Each IC uses a fixed capacitor and a potentiometer to make fine adjustments in the output frequency. The number of pulses per half cycle, P is controlled by the ratio of capacitors as

$$P = C_s/(2C_t).$$

Both the waveform generators are fed with the same dc voltage at the frequency modulation pin from the buffer OA1. This ensures a constant value of P. The triangular waveform

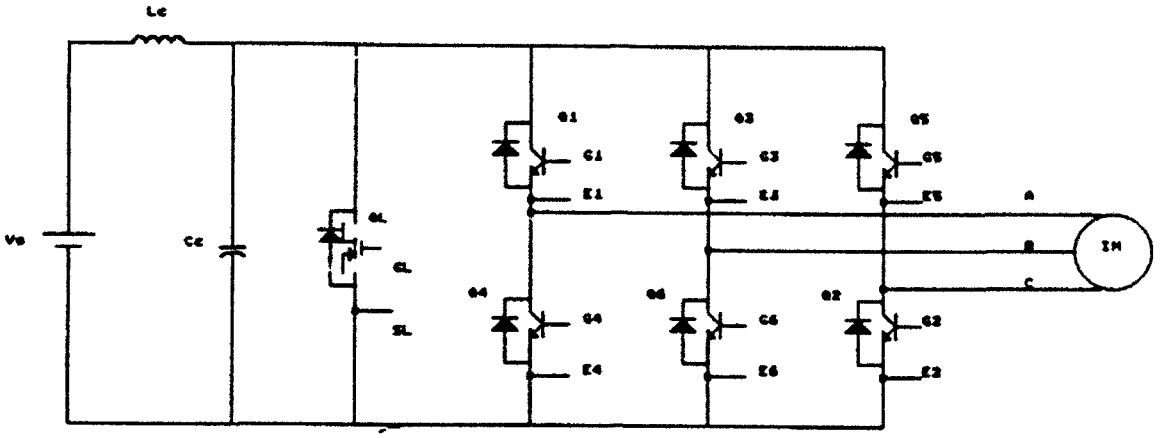


Fig. 5 Power circuit of three phase inverter

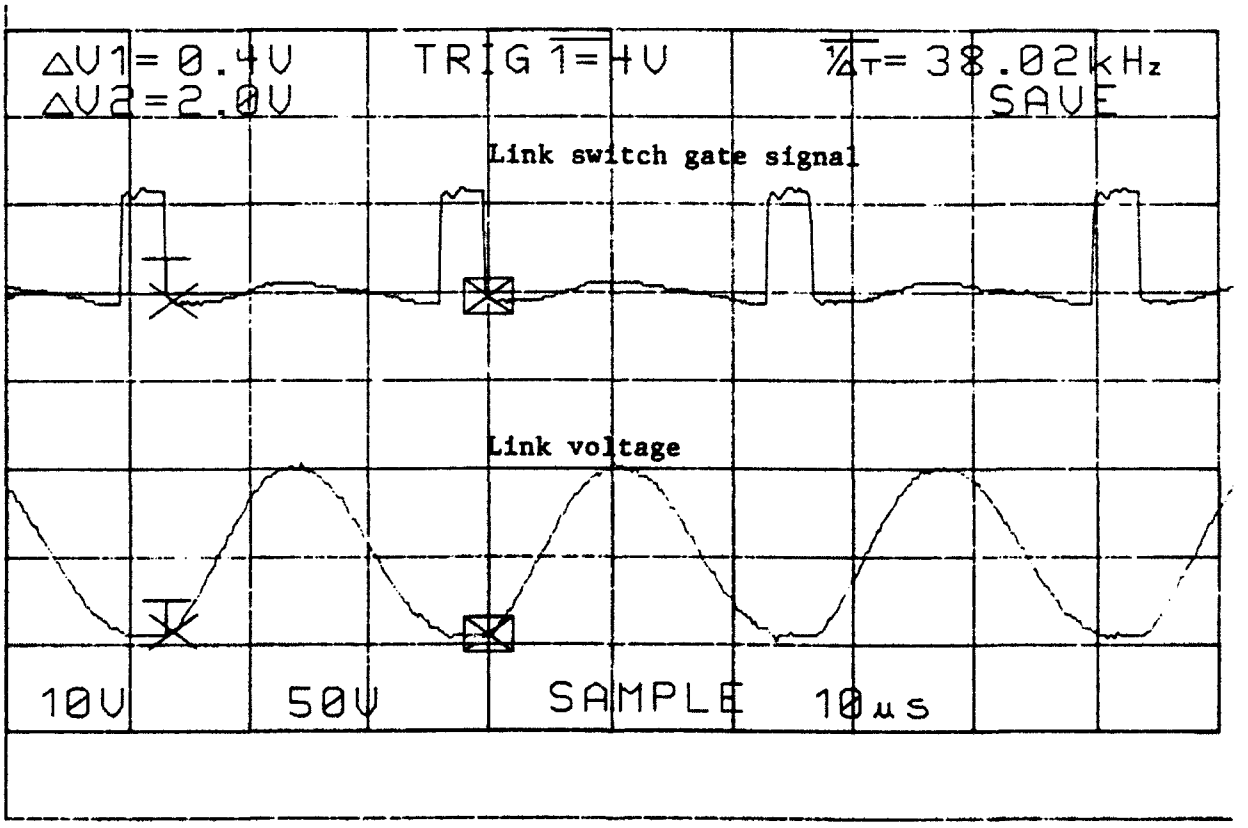
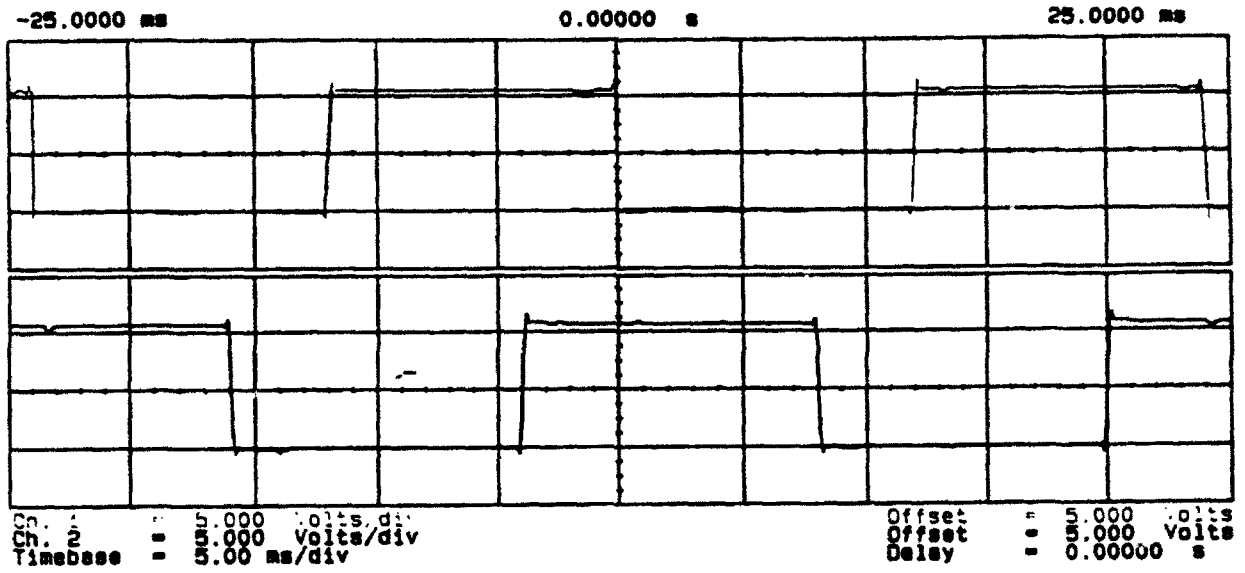
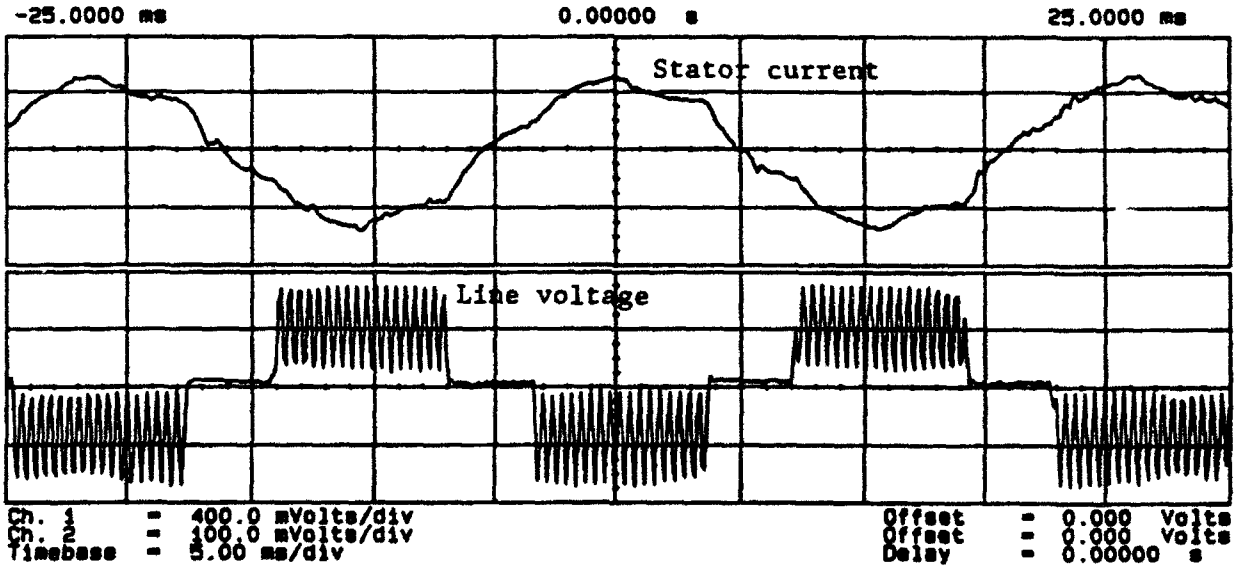


Fig. 6 Experimental waveforms of RDCL



a) Gate drive signals for Phase A and Phase B



b) Stator current and load voltage waveforms

Fig. 7 Experimental waveforms of inverter feeding an induction motor

is synchronized with the sine waveform with the help of a JFET which shorts the capacitor at the end of each sinusoidal cycle. The gate pulse for the JFET is generated using a comparator (OA2) and a monostable operating on the sine wave.

The sine and triangular signals are compared using the comparator OA3 to get the SPWM signal. The resulting signal is synchronized with the link-switch signal using a D-type flip-flop. The gate drive signal for the diagonal IGBTs Q1-Q2 are obtained by carrying out an AND operation with the output of OA2. The gate signal for the pair Q3-Q4 is obtained by inverting and ANDing as shown in Fig. 8. The IGBTs are driven by two IR 2110 drivers. The link-switch MOSFET QL is driven by an SG 3626 driver. The complete circuit diagram is shown in Fig. 8. The experimental waveforms of the various signals are shown in Fig. 9.

The gate signal for the IGBTs can also be obtained using the dedicated driver EXB 841 supplied by Fuji Electric [8]. Two drivers supply the gates of the high-side IGBTs Q1 and Q3 and an SG 3626 driver supplies the gates of the low-side switches Q2 and Q4. The driver EXB 841 provide the isolation for the high side switches Q1 and Q3. They also provide the necessary overcurrent protection through a feedback of the collector voltage through a fast recovery diode MUR 1100E. The drivers along with the power circuit of a single phase inverter are shown in Fig. 10.

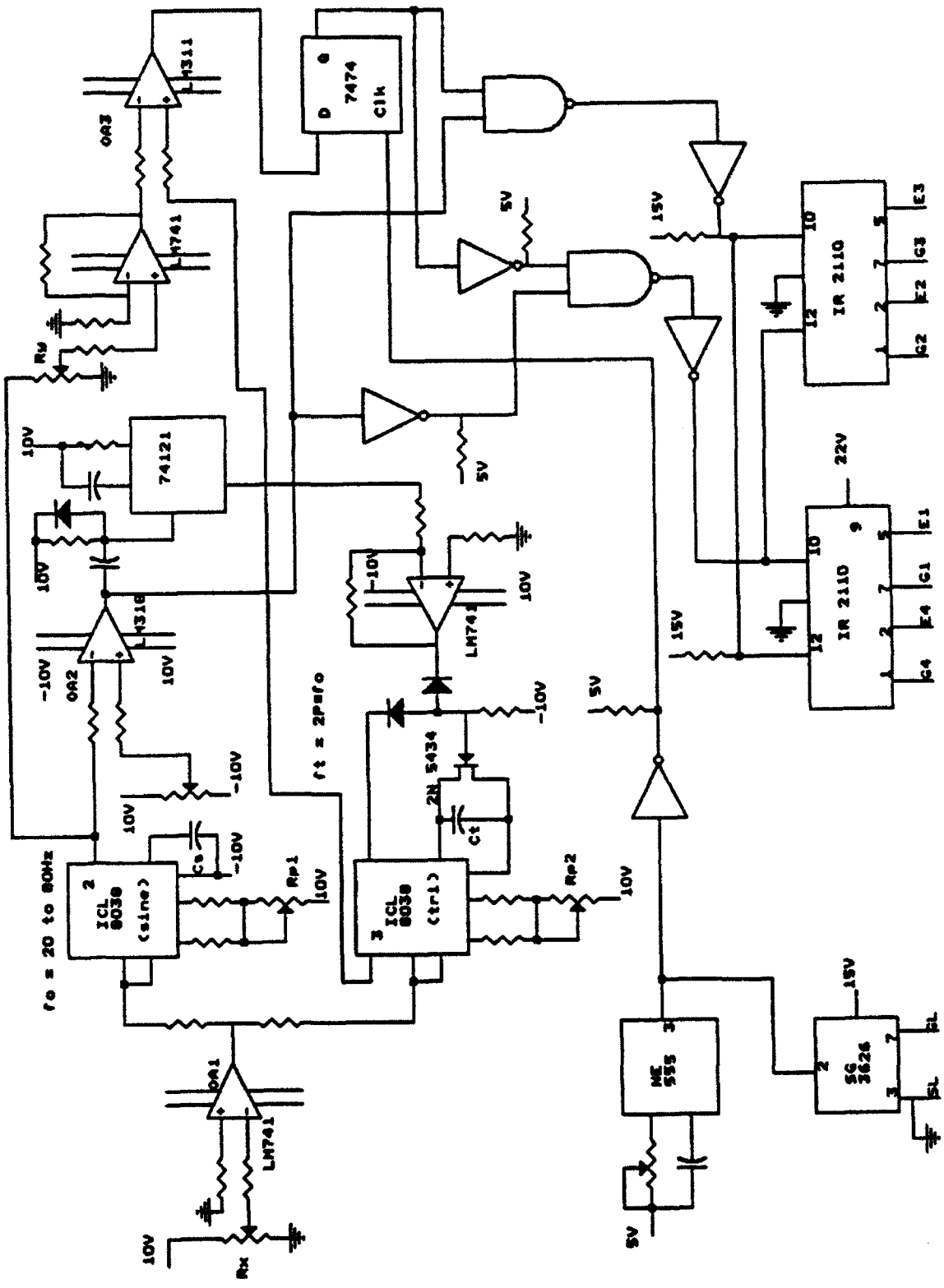


Fig. 8 Complete circuit diagram of PWM inverter (Control circuit)

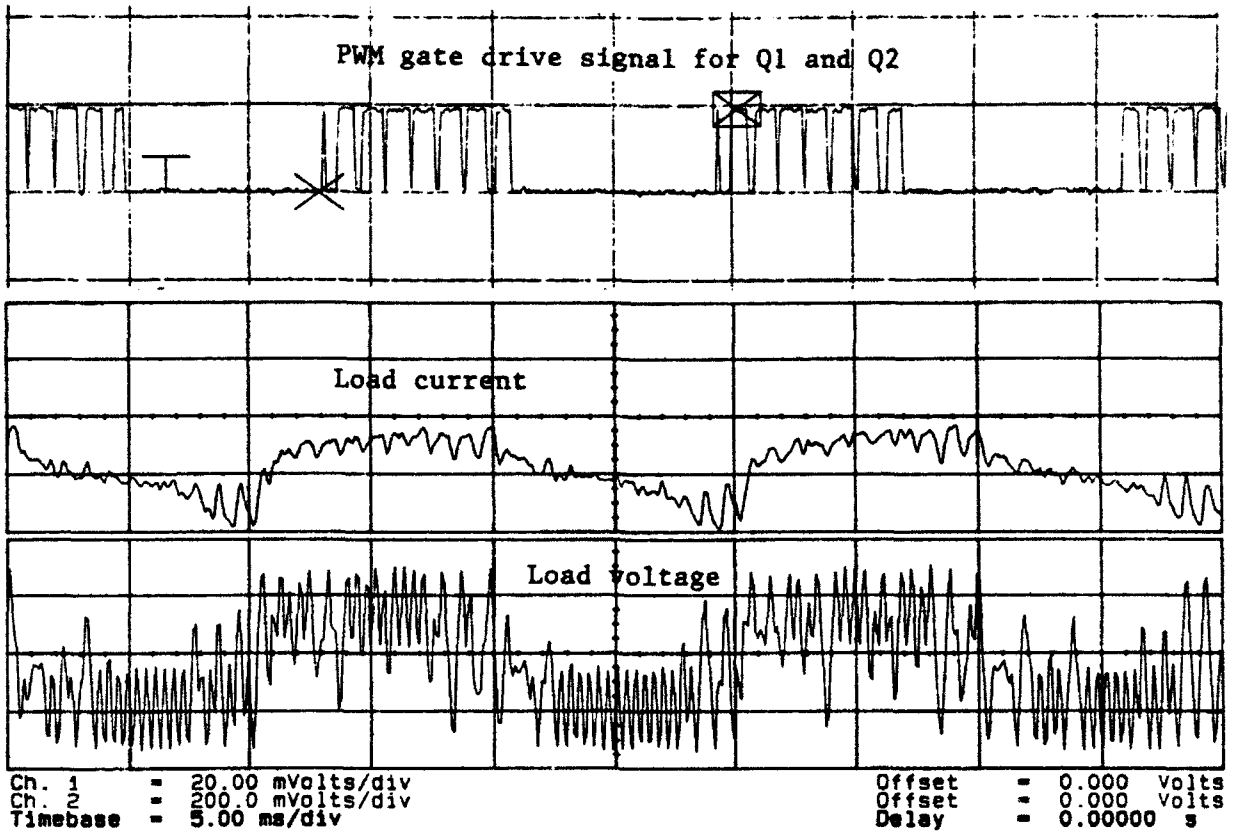


Fig. 9 Experimental waveforms of single phase PWM inverter

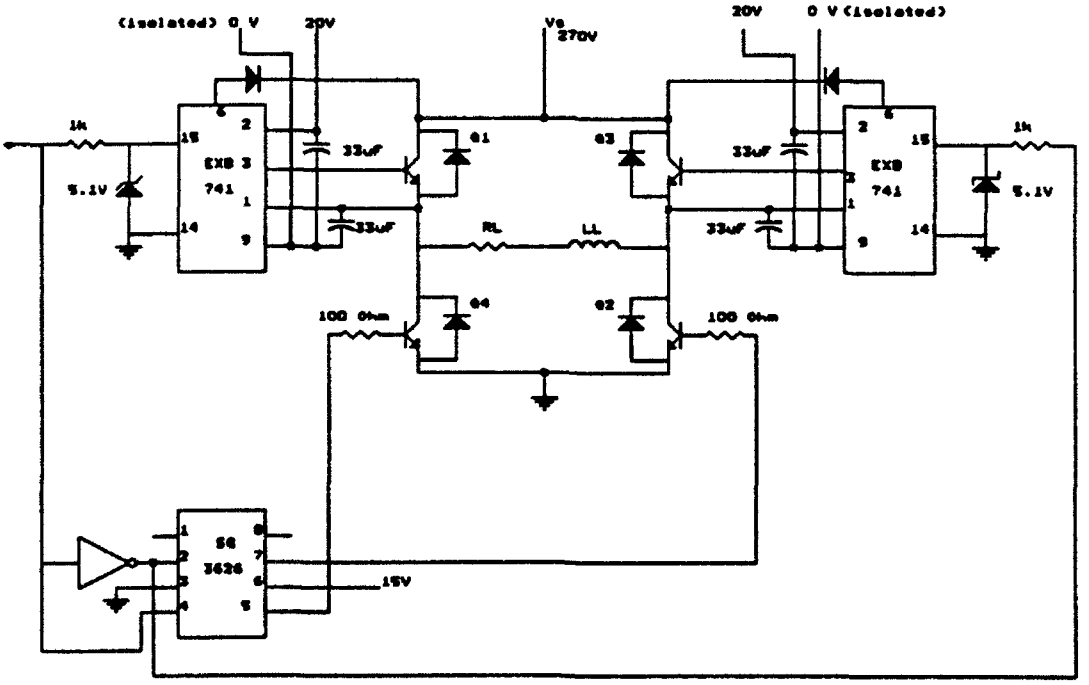


Fig. 10 Power circuit of inverter with EXB 841 driver

RESULTS

The operation of the RDCL inverter was demonstrated by first simulating the inverter using PSpice. Simulation provided several useful information that helped in the design and development of an experimental inverter. Both single phase and three phase inverters operating on an RDCL were built and tested. The experimental waveforms of the experimental inverter feeding a three phase induction motor were obtained on a digitizing oscilloscope. The waveform of the motor current is smooth and is free from any high frequency harmonics introduced by link-switch. A Sine PWM inverter was also developed and tested on the RDCL.

The following are some of the observations made in the course of the development and testing of the RDCL inverters:

a. The link voltage in an RDCL inverter with a limited value of Q-factor does not go to zero for loads with a high power factor. The link-switch turns on at a small dc voltage which introduces some power loss.
b. The link voltage shows an overshoot at inverter switching instants. Both simulation and experiment confirmed this phenomenon. The overshoot, however, decays subsequently.
c. In the case of PWM inverters operating on RDCL, some additional overshoot was observed for inductive load.
d. The gate drivers used for the RDCL inverter failed on some occasions. The source of the failure is to be identified.

Some of the problems identified in the course of the summer research will be studied in detail under the RIP.

CONCLUSION

The concept of RDCL inverter is simple and highly efficient. Induction motor drives using RDCL inverters will be ideally suited for aircraft applications. Computer simulation with PSpice provided several data which were useful in designing an experiental inverter. The experience in building the complete inverter system including the control circuit and the power circuit was highly rewarding. Some of the problems identified in the course of the summer research will be solved under the subsequent RIP grant.

ACKNOWLEDGEMENTS

The author wishes to thank the Air Force Office of Scientific Research for awarding the Faculty Research Associateship and the Research and Development Laboratories for the help and concern in all the administrative and financial aspects of the program. The excellent working environment provided in the Electrical Laboratory at the Wright Patterson Air Force Base during the course of the research period is highly appreciated. In particular, the author is thankful to Mr. Joseph Weimer for all the support and encouragement provided. He is also thankful to Mr. Greg Fronista for arranging the necessary facilities and helping with the various experiments.

REFERENCES

[1] D.M. Divan, "The resonant dc link converter - A new concept in static power conversion," Conference Record of 1986 IEEE IAS Annual Meeting, pp. 648-656.

[2] D.M. Divan et. al., "Design methodologies for soft switched inverters," Conference Record of 1987 IEEE IAS Annual Meeting, pp. 758-766.

[3] K.S. Rajasekhara et. al., "Resonant dc link inverter-fed ac machines control," 1987 Power Electronics Specialists Conference, pp. 491-496.

[4] J. Lai and B.K. Bose, "An induction motor drive using an improved high frequency resonant dc link inverter," IEEE Trans. Power Electronics, Vol. 6, No. 3, pp. 504-513, July 1991.

[5] PSpice User's Manual, MicroSim Corporation, 1991.

[6] I. Smit et. al., "Investigation of limitations in large converters with resonant input link above 30kHz using G.T.O.'s," 1989 Power Electronics Specialists Conference, pp. 1003-1009.

[7] ICL 8038 Precision Waveform Generator, INTERSIL User's Manual, 1990.

[8] MBT (IGBT) Driver EXB 841, Fuji Electric, 1990.