NASA CR-178,709

NASA CONTRACTOR
REPORT

NASA-CR-178709
19860015036

NASA CR-178709

RESEARCH REPORTS - 1985 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

The University of Alabama in Huntsville
Huntsville, Alabama
and
The University of Alabama
Tuscaloosa, Alabama

January 1986

Final Report

Prepared for
NASA, George C. Marshall Space Flight Center
Marshall Space Flight Center, Alabama 35812

RESEARCH REPORTS

1985 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

George C. Marshall Space Flight Center
The University of Alabama in Huntsville

and

The University of Alabama

EDITORS:

Dr. Gerald R. Karr
Professor of Mechanical Engineering
The University of Alabama in Huntsville

Mr. T. Leroy Osborn
Office of University Affairs
Marshall Space Flight Center

Dr. James B. Dozier
Director, Research and Technology Office
Marshall Space Flight Center

Dr. L. Michael Freeman
Assistant Professor of Aerospace Engineering
The University of Alabama

NASA CR - 178709

PREFACE

This document is a collection of technical reports on research conducted by the participants in the 1985 NASA/ASEE Summer Faculty Fellowship Program at Marshall Space Flight Center (MSFC). This was the twenty-first consecutive year the program has been conducted at MSFC. The 1985 program was administered by the University of Alabama in Huntsville (UAH) in cooperation with MSFC and the University of Alabama (UA), University, Alabama. The program was operated under the auspices of the American Society for Engineering Education (ASEE) with sponsorship and funding from the Office of External Relations, NASA Headwuarters, Washington, D.C. The MSFC program was one of eight such Aeronautics and Space Research Programs funded by NASA Headquarters in 1985. Similar programs were conducted at seven other NASA centers. The basic common objectives of the NASA/ASEE Summer Faculty Fellowship Program are:

a. To further the professional knowledge of qualified engineering and science faculty members;

b. To stimulate an exchange of ideas between participants and NASA;

c. To enrich and refresh the research and teaching activities of participants' institutions; and,

d. To contribute to the research objectives of the NASA centers.

The MSFC Faculty Fellows spent 10 weeks (May 28 through August 2, 1985) working with NASA scientists and engineers on research of mutual interest to the University faculty member and the NASA counterpart. The editors of this document were responsible for selecting appropriately qualified faculty to address some of the many problems of current interest to NASA/MSFC. A separate document (UAH Report No. 457, September 1985) reports on the administrative aspects of the 1985 program. This document contains the technical reports on research performed by the individual 1985 participants. The NASA/ASEE program is basically a two-year program to allow indepth research by the University faculty member. In some cases, a faculty member has developed a close working relationship with a particular NASA group that has provided funding beyond the two-year limit. The reports are arranged in alphabetical order.

TECHNICAL REPORT STANDARD TITLE PAGE

1. REPORT NO.	2. GOVERNMENT ACCESSION NO.	3. RECIPIENT'S CATALOG NO.
NASA CR-178709		

4. TITLE AND SUBTITLE	5. REPORT DATE
Research Reports - 1985 NASA/ASEE Summer Faculty Fellowship Program	January 1986
	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S)	8. PERFORMING ORGANIZATION REPORT #
Editors: G. Karr, J. Dozier, L. Osborn, and M. Freeman	
9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. WORK UNIT NO.
University of Alabama in Huntsville Huntsville, Alabama	11. CONTRACT OR GRANT NO. NGT 01-008-021
	13. TYPE OF REPORT & PERIOD COVERED
12. SPONSORING AGENCY NAME AND ADDRESS	Contractor Report
National Aeronautics and Space Administration Washington, D.C. 20546	14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

16. ABSTRACT

This document is a compilation of 40 technical reports on research conducted by participants in the 1985 NASA/ASEE Summer Faculty Fellowship Program at Marshall Space Flight Center (MSFC). Eight participants performed research in the Information and Electronics Systems Laboratory on power conversion, very high order language, cascade solar cells, computer vision, radar analysis, low temperature semi-conductors and flat-panel display. Six participants performed research in the Materials and Processes Laboratory on welding Aluminum Alloy, composite materials, electrochemical corrosion and filament wound case resin. In the Space Science Laboratory, eight participants performed research on holographic x-ray gratings, solar wind, rigid bodies, superfluid helium, alloy solidification models, solid solution crystals, and electrophoresis systems. The Structures and Propulsion Laboratory had seven participants performing Space Shuttle Main Engine (SSME) research, microgravity, statistical analysis and space station thermal systems. Five participants in the Systems Dynamics Laboratory performed research on low density flows, mesosphere thermosphere, computational fluid dynamics nonlinear control and SSME flow predictions. One participant performed research at the Test Laboratory on an x-ray source monitor and another participant worked on automating electromagnetic interference testing. One participant working in Program Development studied mission analysis optimization. One researcher in the Systems Analysis and Integration Laboratory studied the scheduling for the Payload Crew Training Center.

17. KEY WORDS	18. DISTRIBUTION STATEMENT
composite materials, electrophoresis, solid solution crystals, superfluid helium, electrochemical corrosion, very high order language, cascade solar cells	Unclassified-Unlimited W T Carey 2/6/86 W. T. Carey, Director Date Executive Staff

19. SECURITY CLASSIF. (of this report)	20. SECURITY CLASSIF. (of this page)	21. NO. OF PAGES	22. PRICE
Unclassified	Unclassified	974	NTIS

MSFC - Form 3292 (May 1969)

For sale by National Technical Information Service, Springfield, Virginia 22151

TABLE OF CONTENTS

1985

NASA/ASEE SUMMER FACULTY RESEARCH FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

A MONTE CARLO STUDY OF WEIBULL RELIABILITY ANALYSIS FOR SPACE SHUTTLE MAIN ENGINE COMPONENTS

Prepared By:	Kenneth Abernethy
Academic Rank:	Associate Professor
University and Department:	Furman University Computer Science
NASA/MSFC: Laboratory: Division: Branch:	 Structures & Propulsion Propulsion Propulsion Analysis
NASA Counterpart:	Jan Monk
Date:	August 2, 1985
Contract Number:	NASA-NGT-01-008-021

A MONTE CARLO STUDY OF WEIBULL RELIABILITY ANALYSIS FOR SPACE SHUTTLE MAIN ENGINE COMPONENTS

By

Kenneth C. Abernethy
Associate Professor of Computer Science
Furman University
Greenville, South Carolina

ABSTRACT

The use of the Weibull failure distribution model has proven valuable in reliability analysis within the aeronautics industry. The Weibull analysis of samples with a high percentage of non-failures, or censored observations, must be undertaken using one of several large-sample approximations to the distributions of parameter estimators, since exact methods for such distributions are mathematically intractable. It is unknown whether these approximations will produce satisfactory results when used with samples, typical of the space shuttle main engine data, which contain small (fewer than 15) numbers of failures. An objective of this study was to design and implement a Monte Carlo computer simulation to assess the usefulness of the Weibull methods for such small-failure samples. In particular, under varying assumptions concerning failure distribution parameters and number of failures in the sample being analyzed, approximate 90% confidence intervals for predicted times to failure for different percentages of components are calculated by numerical methods. The confidence coefficient for these intervals is then tested by determining what percentages of intervals trap the true values of the parameters being estimated. Also, modifications to a Weibull analysis computer program incorporating methods for calculating approximate confidence intervals and including a number of options for analyzing interval data (as opposed to point data) are described.

INTRODUCTION

The Weibull probability distribution has become a widely used lifetime distribution model since its introduction in 1951 (3). It has been found to be especially important in reliability analysis of manufactured items. Although there is a large body of literature on the distribution and its statistical properties, the distributions of many of the usual parameter estimators seem to be mathematically intractable. The mathematical difficulties encountered in analyzing these estimators are compounded when estimating with samples that include data which has been censored in nontrivial ways. The usual methods employed in this situation involve the use of some statistics whose asymptotic behaviors are understood, but whose applicability for small samples is unknown or unsure.

More particularly, Weibull analysis has proven to be quite valuable in reliability studies for aircraft engine components (1). This success has encouraged the exploration at Marshall Space Flight Center of the applicability of Weibull techniques in reliability analysis for components of the space shuttle main engine (SSME). A primary restriction for the SSME environment is a severe limitation on the test sample size. Test and flight results are collected from fewer than 30 engines. For major SSME components, data samples may contain fewer than 5 failures.

This report documents the incorporation of a number of additional capabilities into an existing Weibull analysis computer program and the results of a Monte Carlo computer simulation study to evaluate the usefulness of the Weibull methods using samples with a very small number of failures and extensive censoring. Since the censoring mechanism inherent in the SSME data is hard to analyze, it was decided to use a random censoring model, generating censoring times from a uniform probability distribution.

Section 1 of the report describes some of the statistical techniques and computer programs that are used in the SSME Weibull analysis. The methods documented in (1) were supplemented by adding computer calculations of approximate (using iterative methods) confidence intervals for several parameters of interest. These calculations are based on a likelihood ratio statistic which is asymptotically a chi-squared statistic with one degree of freedom, the basic method being taken from (2).

The assumptions built into the computer simulations are described in section 2. The simulation program and the techniques used in it are described there also. Simulation results are tabulated for various combinations of Weibull

shape parameters and the numbers of failures in the samples, in the chart of section 3. Conclusions concerning the validity of the chosen Weibull model and estimators, for the small samples characteristic of the SSME reliability analysis, are drawn from the simulation results.

In section 4, some implications of working with interval data, as opposed to point data, are explored. Modifications to the Weibull analysis computer program, with several options for handling interval data are described briefly. Finally, in the final section, some conclusions concerning the applicability of the Weibull methods to the current SSME hardware analysis are reviewed and summarized.

1. DESCRIPTION OF STATISTICAL MODEL.

A brief description of the statistical model and methods used in the SSME reliability analysis is given in this section. For a more complete discussion of these methods and tools the reader should consult references (1) and (2). The Weibull distribution used is the two parameter Weibull whose probability density function is:

$$f(t) = \eta\beta(\eta t)^{\beta-1} \exp(-(\eta t)^{\beta}) \qquad t>0 .$$

The parameter β is called the shape parameter of the distribution. For many applications β-values in the range .5 to 3.5 are reasonable. When $\beta = 1$, the Weibull reduces to the standard negative exponential distribution. The parameter η is called the scale parameter of the distribution and changes in η simply change the horizontal scale and do not alter the basic shape of f(t). It is easily shown that for any Weibull distribution, no matter what the value of β, $Pr(t < \eta)$ is equal to $1 - e^{-1}$ (approximately .632). For this reason η is sometimes called the characteristic life of the corresponding family of Weibull distributions. Some typical Weibull probability density functions are illustrated in figure 1.1. For convenience, η is taken to be 1 for each curve in that figure.

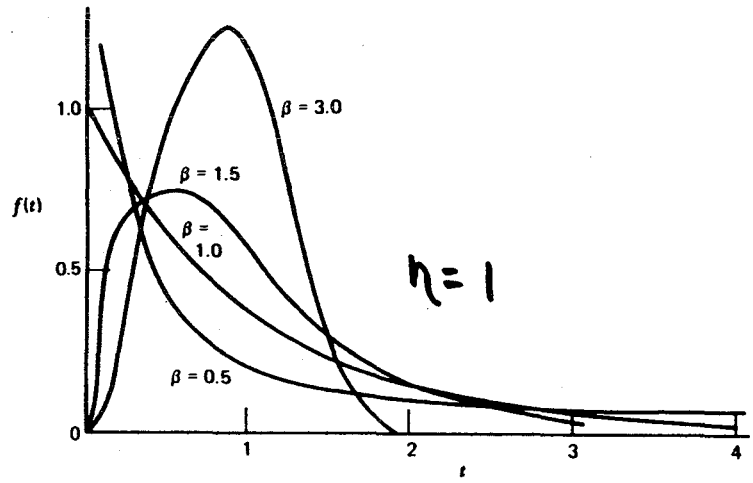


FIGURE 1.1

The basic technique in the SSME use of the Weibull distribution is to estimate the parameters β and η, and then calculate survival times for various percentages of components. For example, $t_{.1}$ (=estimated time for 10% failures) would be calculated by solving $F(t)=.1$, where $F(t)$ is the cumulative distribution function approximated by using the estimates calculated for β and η. In particular, $F(t)$ is given by:

$$F(t) = 1 - \exp(-(\eta t)^{\beta}) \quad t>0.$$

Two techniques are used to estimate β and η. A ranked regression model produces estimates for β and η and provides for a simple graphical evaluation of whether the data seems to fit the Weibull model. Graphical estimation of the t_p values can then be made. A second technique is to calculate maximum likelihood estimates, say $\hat{\beta}$ and $\hat{\eta}$, for β and η and then estimate the t_p using $F(t)$ as indicated above. Both of these techniques are included in a computer program given in (1).

It was deemed desirable to modify the above referenced program so that confidence intervals for the maximum likelihood estimators could be calculated. The type of censoring present in the data makes the calculation of exact confidence intervals impossible. Censored (or suspended) times can be observed for any values of t, and the censoring pattern is difficult to predict. The primary reason for this unpredictibility is that times may be censored for a large variety of reasons, including failures of other engine

components which would terminate a test. Additionally, there is considerable modularity built into the SSME and as a consequence, times at risk for a given type of component are likely to be very variable from particular component to component. In the statistical analysis and simulation studies, type I (see (2) for a discussion of various types of censoring) censoring was assumed. Basically this assumption independently assigns to each component both a failure and a censoring time. The time observed is then the smaller of the two times.

There are two standard methods which can be used to calculate approximate confidence intervals for the Weibull paramters mentioned above. For large samples (30 or more failures), $(\hat{\beta}, \hat{\eta})$ has an approximate bivariate normal distribution whose covariance matrix can be estimated in a straightforward manner. The approximations involved in this method are not generally adequate for samples with fewer than 30 failures (see (2)), and thus the method is not appropriate for the SSME analysis.

A second method which seems to be a better choice for moderately sized samples (around 20 failures) uses a likelihood ratio statistic. This method has been incorporated into the computer program being used at MSFC. Suppose the sample consists of observations $x_1, x_2, x_3, \ldots, x_m$, F= set of indices j, for which x_j is a failure (as opposed to a censored time) and r= number of failures in the sample. The method for calculating confidence intervals for t_p is based on the fact that under the assumption that $t_p = t^*$, the statistic $S_1(t^*)$ has approximately a chi-squared distribution with one degree of freedom. $S_1(t^*)$ is given by:

$$S_1(t^*) = -2 \log L(\bar{\beta}, \bar{\eta}) + 2 \log L(\hat{\beta}, \hat{\eta}).$$

Here log L is the log likelihood function given by:

$$\log L(\beta, \eta) = r\log\beta - r\beta\log\eta + (\beta-1)\sum_{j\in F} \log x_j - \sum_{i=1}^{m} (x_i/\eta)^{\beta}$$

Also $\hat{\beta}$ and $\hat{\eta}$ are the maximum likelihood estimators calculated from the given sample, and $\bar{\beta}$ is calculated (approximated) by solving the following equation iteratively:

$$r/\bar{\beta} - r\log t^* + \sum_{j\in F} \log x_j + \log(1-p) \sum_{i=1}^{m} (x_i/t^*)^{\bar{\beta}} \log(x_i/t^*) = 0.$$

After $\bar{\beta}$ is found then $\bar{\eta} = t^*/(-\log(1-p))^{1/\bar{\beta}}$

In the computer program, $\bar{\beta}$ is calculated using a hybrid secant/false-position method. A γ-confidence interval for t_p is found by finding the set of values t^* for which $S_1 < \chi^2_{1,\gamma}$. This is accomplished in the computer program by starting with the point estimate for t_p (based on the estimates $\hat{\beta}$ and $\hat{\eta}$), and testing values for t^* diverging from t_p in both directions until values are found on either side of t_p for

which S_1 exceeds the appropriate critical value.

In a similar manner, a γ-confidence interval for β can be calculated by finding all values β^* for which $S_2 < \chi^2_{1,\gamma}$. Here

$$S_2(\beta^*) = -2 \log L(\beta^*, \eta^*) + 2 \log L(\hat{\beta}, \hat{\eta})$$

where log L, $\hat{\beta}$ and $\hat{\eta}$ are as before and

$$\eta^* = \left(\sum_{i=1}^{m} x_i^{\beta^*} / n \right)^{1/\beta^*}$$

Representative output from the computer program is given in figure 1.2.

```
THE FOLLOWING ESTIMATES ARE RANKED REGRESSIONESTIMATES

BETA=        1.1397   ETA=                    448.33
R=   0.96470R**2=   0.93065
DO YOU WISH TO DO MAXIMUM LIKELIHOOD ESTIMATION?
ANSWER Y OR N
Y

DO YOU WISH 80% OR 90% CONFIDENCE INTERVALS?
TYPE IN 80 OR 90.
90
 MAXIMUM LIKELIHOOD ESTIMATES FOR THIS CASE FOLLOW
 BETA=        1.3264   ETA=                    440.80

90% BETA CONFIDENCE LIMITS        0.83   ...        1.93

PERCENTAGE OF FAILURES      ESTIMATED TIME      90% CONF. INTERVAL
          0.1                      2.          0.   ---        12.
          1.0                     14.          2.   ---        41.
         10.0                     81.         32.   ---       154.
         20.0                    142.         71.   ---       249.
         50.0                    334.        201.   ---       602.
         63.2 ETA                441.        286.   ---       859.
         90.0                    827.        496.   ---      2025.
```

FIGURE 1.2

A primary problem with the above analysis in the SSME case is that the methods for calculating the approximate confidence intervals are based on asymptotic distribution approximations, whose validities are in question for such small samples. The problem is a serious one since the small size of the samples make us intuitively skeptical of the point estimates calculated from them and hence anxious to have some reasonably accurate confidence measure for those estimates. In order to address this problem, Monte Carlo experiments were designed and conducted to assess the accuracy, and hence usefulness, of the confidence intervals being calculated.

2. SIMULATION SAMPLE GENERATION.

Failure times were calculated by generating a random number r from the uniform distribution on [0,1] (this was done using RAN, the built in random number function in the VAX FORTRAN), and then computing $F^{-1}(r)$, where F(t) is the Weibull cumulative distribution function for a given choice of parameters β and η. In fact, η was always taken to be 100 and β was varied. Sample failure times thus generated were tested, using the standard chi-squared goodness-of-fit test, against the given Weibull distribution and found to be representative of that distribution. Censoring times were generated to follow a uniform distribution.

To generate a sample of times, the number of failure times to be present in the sample was fixed. Failure times and censoring times were generated in pairs and the corresponding observed time was taken to be the minimum of the pair. Since the number of failures was predetermined for a given simulation run, the overall sample sizes varied. The censoring distribution range was varied (ranges to be used were determined by simulation) so as to achieve (approximately) the desired average sample size. An average sample size of 50 was chosen to be representative of a number of projected SSME sample sizes. Additional Monte Carlo studies, in which both sample sizes and censoring distributions will be varied, are planned.

For an occasional sample one or more of the iterative numerical methods employed in the confidence interval calculations failed to converge properly. Such samples were omitted in the subsequent Monte Carlo analysis. The number of samples for which this occured was small (on the order of 1%), and so it is clear that ignoring them could not unduly bias the simulation results. At any rate, when such samples occur in practice, they would not be evaluated using the confidence interval calculations.

3. SIMULATION RESULTS.

Most of the Monte Carlo simulations that have been run are summarized in Table 3.1. Some additional isolated runs with varying sample sizes and censoring distributions have been made, but these were not systematic enough to include here. It might be noted, however, that these runs showed results consistent with the results reported here. A more thorough collection of simulation results will be reported on later.

For each of the 12 simulation runs (each column represents a run) summarized in Table 3.1, 100 samples of approximate size 50 each were used. Each run of 100 simulated samples took approximately an hour of computing time on a VAX 11/780, which explains why more extensive runs were not made.

The upper part of Table 3.1 contains the number (out of 100) of calculated approximate 90% confidence intervals which actually trapped the true parameter values being estimated. Lawless (2) has predicted that these confidence intervals would tend to be too large (and hence trap true values more than 90% of the time) for samples with small numbers of failures. Our results seem to support that prediction when the number of failures is 10 or 20, but such an effect is not obvious in the 3 or 5 failure cases. More extensive simulations will be needed to gain insight into this matter.

Eta = 100

	Beta=1.5				Beta=2.0				Beta=3.0			
Failures=	3	5	10	20	3	5	10	20	3	5	10	20
β	85	91	90	99	88	92	93	89	95	87	90	84
$t_{.001}$	93	89	93	99	88	90	93	88	88	90	91	92
$t_{.01}$	93	89	93	98	87	92	90	89	91	90	91	90
$t_{.1}$	87	91	91	96	92	92	94	91	91	95	90	92
$t_{.2}$	86	91	88	95	91	91	95	89	90	96	96	95
$t_{.5}$	84	91	84	95	89	86	93	95	90	92	97	97
$\eta = t_{.632}$	85	90	85	96	88	86	95	94	90	91	96	96
$t_{.9}$	88	91	90	95	88	90	94	92	91	91	94	96
Average $\hat{\beta}$	2.03	1.75	1.59	1.58	2.76	2.33	2.18	2.10	3.78	3.42	3.25	3.07
$\hat{\beta}$ too large	59	55	48	56	64	58	61	52	64	57	52	50
Average $\hat{\eta}$	134	107	101	101	107	101	99	99	103	101	99	100
$\hat{\eta}$ too large	41	34	39	48	35	36	38	51	35	37	45	42

Number of Confidence Intervals Trapping Actual Values

Each column based on 100 samples of (approximate) size 50.

TABLE 3.1

In the bottom part of Table 3.1, the bias in the estimators $\hat{\beta}$ and $\hat{\eta}$ is examined, and we see that for smaller samples the bias in each estimator grows larger. Both $\hat{\beta}$ and $\hat{\eta}$ are biased high (in average), but in the case of $\hat{\eta}$, median values are below the actual parameter values. The possibility of constructing (by simulation) a table of bias-correcting multipliers as functions of sample size, $\hat{\beta}$, $\hat{\eta}$, and the number of failures may be worth considering.

Although these simulation results are not extensive enough to provide a definitive conclusion, the tentative conclusion is that the approximate confidence intervals calculated using the likelihood ratio statistic are reasonably accurate even for samples with very few failures. Thus, the calculation of these intervals would seem to be a useful addition to the SSME Weibull analysis methodology.

We should note here that the calculations of the maximum likelihood estimators and the associated approximate confidence intervals are based on the assumption that our failure times are true point data, i.e. that we know the failure times exactly. Whenever failure times are known only to the extent of falling in some time interval (interval data), the above calculations do not apply. To use the above methods with interval data, we must make some assumption about the placement of failure times within failure intervals. Since much of the SSME failure data is interval data, we will pursue this issue in the next section.

4. INTERVAL DATA CONSIDERATIONS.

The most conservative approach in working with interval data is to make no additional assumptions about the true failure times and use only the failure intervals in any statistical analysis. It is not difficult to modify the likelihood function to accomodate interval data, and calculate the maximum likelihood estimates for β and η based on this function. The numerical procedures are a bit more complicated and sensitive than in the point data case; hence reasonably good beginning guesses for β and η are required in order for the iterative methods to converge properly. The capability to calculate interval data maximum likelihood estimates for β and η has been added to the Weibull analysis computer program being used at MSFC.

One approach which has been used in analyzing SSME interval data is to make the optimistic assumption that all failures in a failure interval occur at the right hand boundary of the interval. As might be expected, the estimates for β and η that are calculated using the interval data method and the point data method (with the above assumption)

are quite different for most sets of data. Of course the differences in these estimates will decrease as the intervals involved get smaller, but for the interval sizes typical of current SSME data, these differences can be expected to be significant.

Chart 4.1 illustrates a range of possible assumptions about failure time placement for interval data. All of these choices have been included as options in the Weibull analysis computer program. This capability allows an exploration of the implications arising from various assumptions about the unknown failure times. In the computer calculation for choice A (interval data method), the initial "guesses" for β and η are gotten by calculating estimates using choice D.

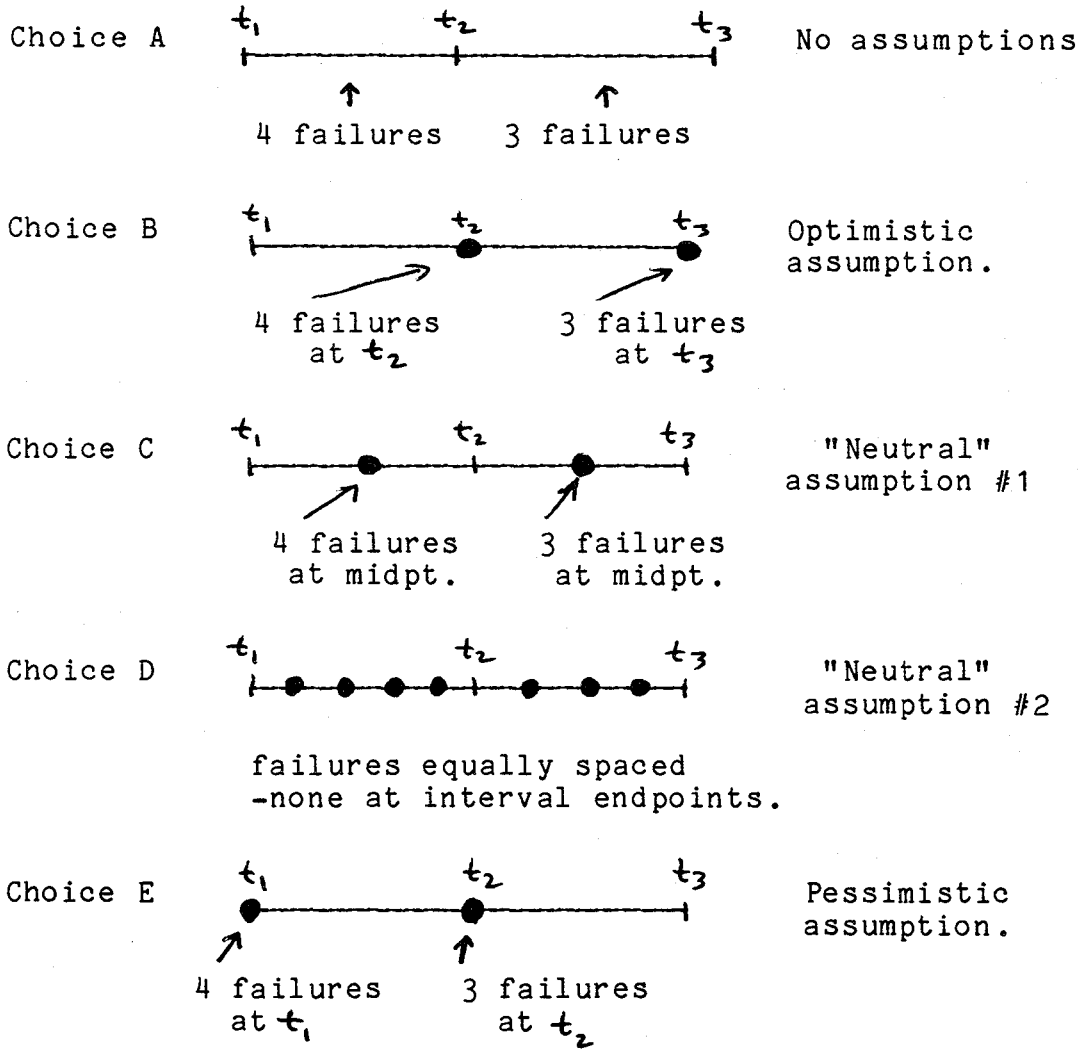


CHART 4.1

CONCLUSIONS

The primary conclusion of this study is that Weibull methods can be of positive value in SSME hardware reliability analysis. However, caution should be exercised in the use of these methods with small-failure samples and interval data.

When using point data, it is important to consider confidence intervals because point estimates can be quite erroneous and misleading, especially when samples contain very small numbers of failures. One of the primary results of this study was to implement computer methods for calculating approximate confidence intervals from such samples and to verify that these approximate confidence intervals are reasonably accurate.

The use of the Weibayes analysis method discussed in (1) is especially risky for current SSME use because of the lack of a significantly large SSME failure data base. In fact, making a priori estimates of betas from the existing SSME failure data base is not much better than simple guessing.

When analyzing interval data, it is important to realize that any assumptions made about the placement of the failure times within failure intervals is likely to have more impact upon the statistical predictions made from that data, than the choice of a particular statistical method, or even model, is likely to have. This is not to suggest that such assumptions should never be made, but rather to point out that when they are made, careful consideration and study should be given the choice of an assumption. Additions to the Weibull analysis computer program allow explorations of the implications of various of these assumptions.

Finally, no matter what assumptions are chosen, it is imperative to remember that the predictive power of any model is limited by the integrity and informational content of the data used as input to that model. The SSME data is typically interval data with a small number of failures, and no statistical model or method will be capable of extracting extremely dependable and accurate predictions from it.

In summary, the modified Weibull analysis computer program now provides a range of capabilities and options for treating estimating methods and data assumptions. These capabilities enhance the chances that the Weibull model can be used to advantage in SSME hardware reliability analysis. When applied and interpreted with the caution dictated by the considerations outlined here, the model should provide an important additional tool for SSME data analysis.

REFERENCES

1. Abernethy, R.B., Breneman, J.E., Medlin, C.H., Reinman, G.L., Weibull Analysis Handbook, Air Force Wright Aeronautical Laboratories (AFSC), Wright Patterson Air Force Base, Ohio, 1983.

2. Lawless, G.F., Statistical Models and Methods for Lifetime Data, J. Wiley, New York, NY, 1982 (see especially chapter 4).

3. Weibull, W., "A Statistical Distribution Function of Wide Applicability," Journal of Applied Mechanics, 1951, pp 293-297.

Note: Computer program listings (both Weibull analysis and simulation programs) are available from author.

1985

NASA/ASEE SUMMER FACULTY RESEARCH FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA AT HUNTSVILLE

CONTAINERLESS PROCESSING OF HYPERMONOTECTIC AND GLASS FORMING ALLOYS USING THE MARSHALL SPACE FLIGHT CENTER 100 METER DROP TUBE FACILITY

Prepared By:	J. Barry Andrews, Ph.D.
Academic Rank:	Associate Professor
University and Department:	University of Alabama at Birmingham
NASA/MSFC:	
Division:	Low Gravity Science
Branch:	Solid-State and Solidification
MSFC Counterpart:	Michael B. Robinson

Date:	August 18, 1985
Contract No.:	NGT-01-008-021 The University of Alabama at Huntsville

CONTAINERLESS PROCESSING OF HYPERMONOTECTIC AND GLASS FORMING ALLOYS USING THE MARSHALL SPACE FLIGHT CENTER 100 METER DROP TUBE FACILITY

J. Barry Andrdews
University of Alabama at Birmingham

ABSTRACT

Two separate projects were carried out to study alloys whose solidification structures can be strongly influenced by the presence of a container during melting and solidification.

One project involved containerless solidification of hypermonotectic $Au_{35}Rh_{65}$ alloys. This alloy exhibits liquid immiscibility over a temperature range. In previous studies on similar alloy systems, attempts to obtain dispersions of the immiscible liquid phases by using microgravity processing to prevent sedimentation have often failed. Surface tension driven flows occur when the minority liquid phase completely wets the walls of the container, resulting in massive segregation even under microgravity conditions. It has been suggested that containerless melting might be one solution to this problem. However, surface tension driven flows could also lead to accumulation of the minority liquid phase at the external surface of a containerlessly melted alloy. The research underway is a first step in determining the influence of containerless, microgravity processing on immiscible alloys.

Nickel-niobium alloys are also being studied using the drop tube facility. One alloy in this system, a $Ni_{60}Nb_{40}$ alloy, is a good candidate for the formation of a bulk metallic glass. Amorphous alloys of this composition have been produced using thin film and mechanical alloying techniques. However, theory indicates that if heterogeneous nucleation can be avoided, it should be possible to produce an amorphous structure in this system using a moderate cooling rate from the melt. The containerless melting and solidification capabilities of the drop tube facility provide ideal conditions for a study of this type. To date, several $Ni_{60}Nb_{40}$ samples have been levitated, melted and cooled during 4.6 seconds of free fall in the 100 meter drop tube. The structures obtained are discussed in the following report.

ACKNOWLEDDGEMENTS

I would like to thank Dr. Gerry Karr, Leroy Osborn, Jim Dozier and Mike Freeman for coordinating a program that was not only informative and beneficial, but also most enjoyable. I would also like to thank Dr. Bob Naumann and Dr. Alex Lehoczky for providing me with the opportunity to participate in this program by providing laboratory sponsorship. I am most indebted to my NASA counterpart, Mike Robinson, for the unselfish contribution of his time and resources. His enthusiasm and expertise were a great asset to the success of the project. Thanks also go to Alice Dorries for her help with SEM analyses and to Ron Harris whose expert help in preparing metallographic and SEM samples is most appreciated.

LIST OF FIGURES

"LIST OF FIGURES CONTINUED"

FORWARD

The possibility of experimentation involving low gravity, containerless processing provides new opportunities for research in many intriguing areas. Among these areas are studies concerning immiscible alloys (hypermonotectic alloys) and metallic glasses. This report will discuss research in these two areas.

A convenient way to carry out experiments requiring relatively short duration containerless, microgravity conditions is through the use of a drop tube [1-4]. The following report describes two projects carried out using the 100 meter Drop Tube Facility located at the Marshall Space Flight Center in Huntsville, Alabama. This facility permits containerless melting followed by containerless solidification during 4.6 seconds of free fall [5]. The report is divided into two parts in order to maintain continuity in the discussion of each project.

PART ONE - GLASSY METALS

INTRODUCTION

Amorphous metallic alloys were first produced by quenching materials from a melt in 1960 [6]. Very high cooling rates were utilized in this original work to rapidly "solidify" the molten alloy and quench in the amorphous, liquid like structure. Of course, in order to produce an amorphous metal it is necessary to avoid nucleation of the crystalline phase during cooling from the melt. If a high enough cooling rate is utilized (10^6 °C/sec in most systems) the glass transition temperature, Tg, is reached before nucleation can take place [7]. Once the sample is below the glass transition temperature, atomic movement is restricted and nucleation does not readily occur. The amorphous structures that result are obviously somewhat unstable and in several systems can decompose at room temperature.

After the initial break through in 1960, it was soon found that amorphous metals could be formed from many alloy systems. However, it was also apparent that production was limited to very thin films (approximately 40um) in order to allow rapid heat extraction. Commercialization of amorphous alloys eventually took place in the late 1970's with the production of thin ribbons of ferromagnetic "glassy metals". These alloys had no magnetic anisotrophy since they had no crystal structure. This feature, along with other attractive magnetic properties, made the alloys preferable to their crystalline counterparts for applications where soft magnetic materials were needed, such as transformers.

To date, the majority of the work carried out in the production of amorphous metals has been involved with thin films. Very little has been done with bulk samples because the moderate cooling rates attainable are insufficient to suppress heterogeneous nucleation. However, if heterogeneous nucleation can be prevented in bulk samples, it appears that the achievable cooling rates are sufficient to suppress homogeneous nucleation in at least a few alloy systems.

Since the main heterogeneous nucleation site in most experiments is the container wall, a process is needed where this contact can be eliminated. The Marshall Space Flight Center 100m Drop Tube Facility provides the containerless melting and solidification capabilities that are needed to avoid this contact. As a result, it is reasonable to assume that if high purity, inclusion free samples can be produced and then containerlessly melted and solidified using the drop tube, heterogeneous nucleation can be avoided.

Once the problem of heterogeous nucleation is overcome, the next difficulty is that of homogeneous nucleation. There are several alloy systems where it may be possible to avoid homogeneous nucleation using

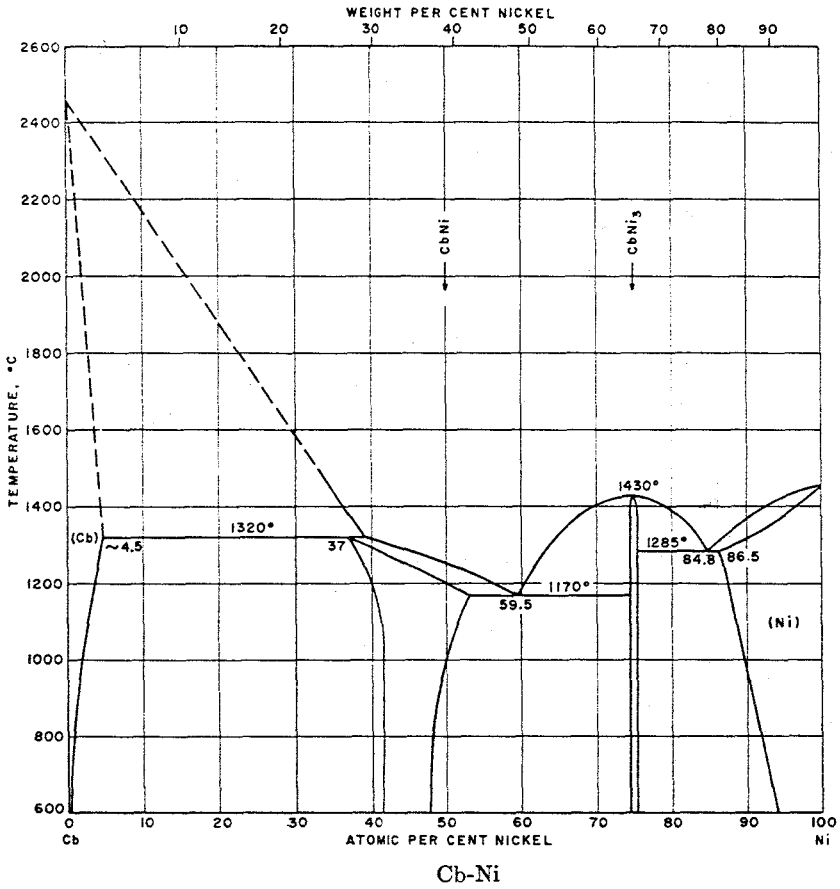


Figure 1. Phase diagram of the nickel-niobium system.

only moderate cooling rates. What is necessary is an alloy that has a high glass transition temperature, T_g, in relation to its liquidus temperature, T_m. Alloys in this category would be expected to have a relatively low homogeneous nucleation rate. The ratio $T_{rg} = T_g/T_m$ is used to quantify this relationship and is called the reduced glass transition temperature. The highest reduced glass transition temperature found in the literature is 0.67 and is for a $Ni_{60}Nb_{40}$ alloy [8]. The phase diagram for this system is shown in Figure 1. Note that the $Ni_{60}Nb_{40}$ alloy has the lowest liquidus temperature in this system.

Work by Drehman, Greer and Turnbull [9] has shown that for $Pd_{40}Ni_{40}P_{20}$ alloys(T_{rg}=0.66), if heterogeneous nucleation can be avoided,

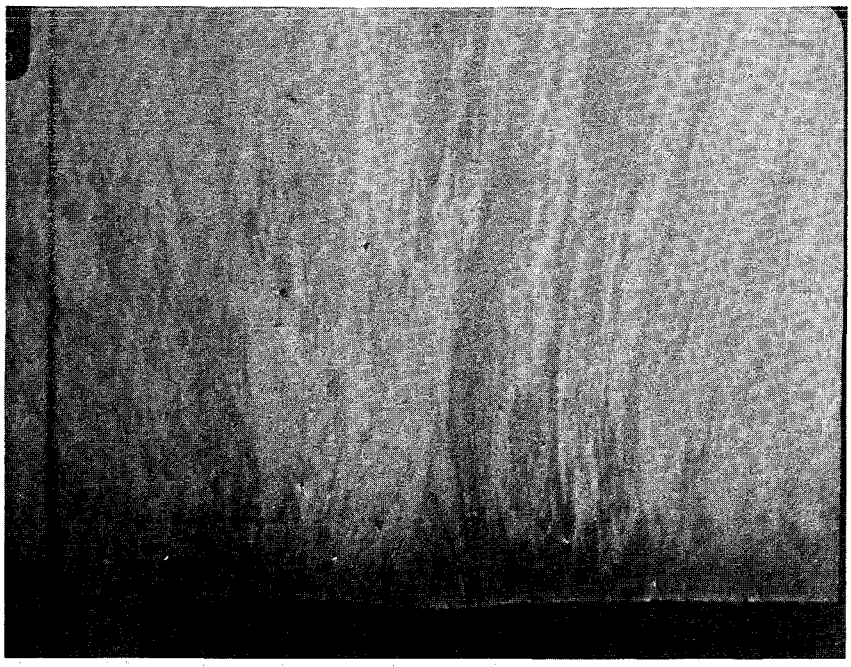


Figure 2. Microstructure of an arc melted $Ni_{60}Nb_{40}$ alloy. (60X)

cooling rates of only several degrees per second are adequate for the formation of an amorphous structure. Because of the similar T_{rg} for $Ni_{60}Nb_{40}$ alloys, it shoud also be possible to form a glassy metal using moderate cooling rates, if heterogeneous nucleation is avoided. The containerless processing capability of the Drop Tube Facility should be most usefull in this experimentation.

EXPERIMENTAL PROCEDURE

$Ni_{60}Nb_{40}$ alloys were prepared from Marz grade nickel and niobium. Each of the constituents was etched for 60 seconds in concentrated HF and washed with acetone before alloying. Melting and alloying were carried out using a small tungsten electrode arc melting unit with a water cooled copper hearth. Melting was carried out under an argon atmosphere.

One of the arc melted samples was sectioned and prepared for microscopic analysis. Analysis revealed a uniform structure consisting of columnar grains aligned parallel to the direction of heat flow during solidification. See Figure 2.

Figure 3. Microstructure of a $Ni_{60}Nb_{40}$ sample melted and solidified under containerless conditions. (60X)

After arc melting, samples were etched for one minute in concentrated HF in order to remove any oxides or other contaminants from the surface. A slight change in surface color (from silvery to a bluish tint) was noted as a result of the etching treatment. The etched samples were then utilized for drop tube experiments.

An absolute pressure of 500 Torr of argon was used in the belljar assembly of the drop tube which houses the electromagnetic levitator. 500 Torr of helium was used in the remainder of the tube in order to maximize the cooling rate of the sample.

RESULTS AND DISCUSSION

Samples levitated easily and were dropped at temperatures ranging from 2420K to 2700K. The dropped samples had a relatively dark appearance indicative of an oxidized surface.

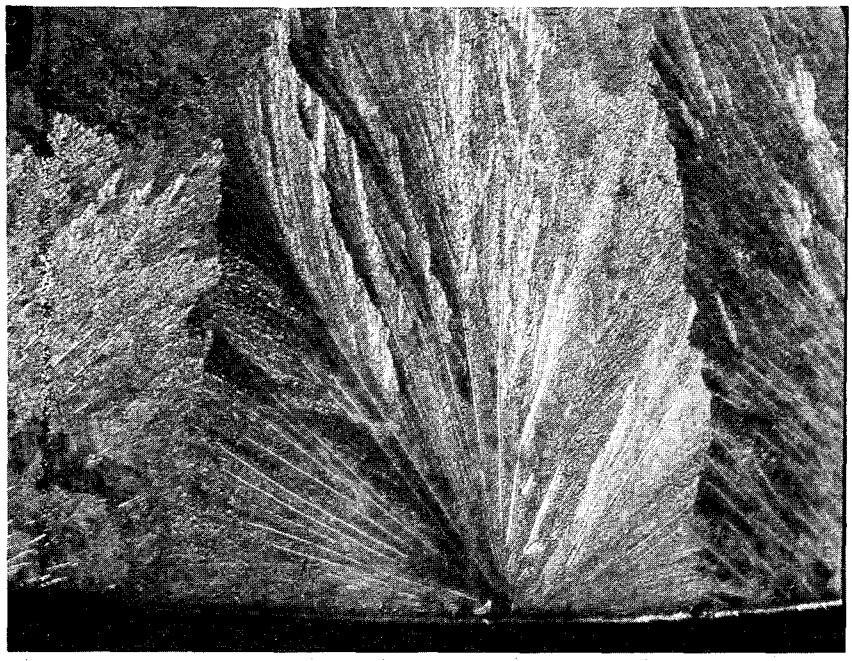


Figure 4. Needle like structure radiating from nucleation site on a dropped $Ni_{60}Nb_{40}$ sample. (240X)

Two of the samples were sectioned, mounted and polished for metallographic examination. Etching with a solution of 4 parts H_2O, 3 parts HNO_3 and 1 part HF for ten seconds provided microstructural contrast under sensitive tint. Analysis indicated that nucleation of the crystalline phases had occured at relatively few sites on the surfaces of the samples and had led to the formation of several columnar grains. See Figure 3. Ususual, needle like structures were evident which appeared to radiate out from the nucleation sites. See Figure 4. Energy dispersive x-ray analysis is planned in order to determine the composition of these needles.

A second series of samples of the same composition was produced using a similar procedure. However, in this case the arc cast samples were not subjected to etching in HF before being dropped. This was because of concern over the discoloration experienced in the first series of alloys. A helium - 6% hydrogen gas mixture at an absolute pressure of 500 Torr was utilized in the bell jar assembly of the drop tube when levitating and melting these samples. The reducing gas mixture was used to minimize the chances of the formation of oxides which could serve as heterogeneous nucleation sites.

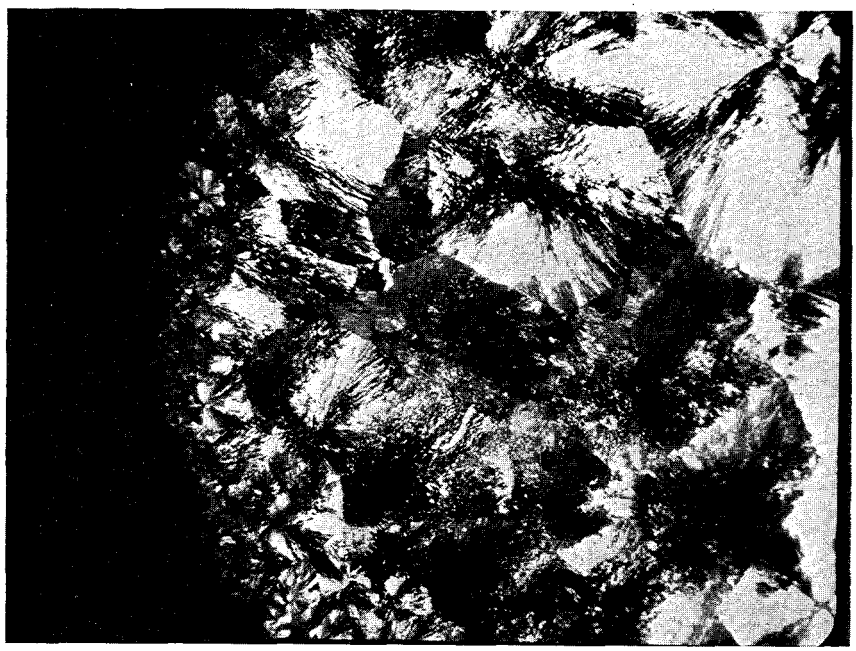


Figure 5. Equiaxed grain structure obtained from a $Ni_{60}Nb_{40}$ sample dropped without prior surface etching. (37.5X)

The samples were again easily levitated and melted. An attempt was made to drop these samples at temperatures closer to their liquidus than those dropped previously. Drop temperatures ranged from 1850K to 2340K. All of the dropped samples appeared to have some surface oxidation. Metallographic analysis again revealed that nucleation of the crystalline phase had occured. However, in this case a much larger number of nucleation sites was apparent. The grains were smaller and more equiaxed, with the grain size ranging from a minimum near the surface to a maximum near the center of the samples. See Figure 5.

While amorphous structures have not yet been obtained in these bulk Ni-Nb samples, there is good reason to believe this will eventually be accomplished in the drop tube. Current plans are to experiment with various surface etchant and bell jar atmosphere combinations as a first step. Based on the results, additional experiments may be carried out using an amorphous flux coating on the samples in order to minimize surface oxidation and remove any impurities. If the flux maintains an amorphous structure it should not contribute significantly to nucleation of the crystalline phase.

CONCLUSIONS

Additional work is needed in order to determine the feasibility of producing bulk amorphous samples using the containerless melting and solidification capabilities of the drop tube. At this time, one difficulty appears to be surface oxidation of the samples, either during melting in the levitator or during free fall down the length of the tube. The use of a flux is being considered for future experiments.

Results indicated that samples which were etched in HF, melted in an argon atmosphere and dropped through helium, contained a columnar microstruture resulting from nucleation occuring at relatively few sites on the surface. Samples melted in the unetched condition, but in a reducing helium-hydrogen gas mixture, and then dropped through a helium atmosphere, contained a much more equiaxed grain structure with a relatively fine grain size at the surface varying to a much larger grain size near the center of the sample. Obviously, nucleation events were not limited to the surface of the samples under the latter conditions. Continued work on this project is planned.

PART TWO - IMMISCIBLE ALLOYS

INTRODUCTION

In immiscible alloy systems, an example being systems that contain a monotectic reaction, two liquids are found over some temperature and composition range which will not mix. When immiscible alloys are solidified under 1g conditions, segregation usually occurs due to sedimentation of the more dense liquid phase. Alloys in these systems are interesting to study using microgravity conditions since solidification in the absence of a gravitational field removes the driving force for sedimentation. As a result, microgravity processing should permit the formation of a dispersion of the two liquid phases instead of the gross segregation that normally occurs during solidification of immiscible alloys [10-13].

Because of their grossly segregated structures, most immiscible alloys have been only of academic interest. However, if fine dispersions of the immiscible phases can be obtained in some of these alloy systems, it is quite possible that materials with unique properties can be produced. Applications could include electrical contact materials, catalytic materials, and even fine particle magnets and superconductors [14].

For these reasons, there is considerable interest in processing immiscible materials under microgravity conditions. The low gravity, containerless solidification capability of the drop tube can be most usefull in this area. Therefore, the objective of this work was to determine the effect of solidification under the microgravity conditions obtained during free fall of a sample in the drop tube. Of particular interest was the distribution of phases in an immiscible alloy.

Several factors had to be considered in selecting an appropriate alloy system for study. It was desirable to use a system with a known phase diagram and with sufficient differences in the densities of the immiscible phases to permit them to easily segregate when solidified under 1g conditions. A system was also needed that possessed a relatively high monotectic temperature. This was needed for two reasons. First, a high monotectic temperature would result in fairly rapid cooling rates and help ensure solidification during free fall. Second, high temperatures would result in sufficient radiation to permit tracking of the samples during free fall using the infrared detectors in the tube. It was also desirable to use an alloy system in which oxidation would not pose a major problem.

One system that met all of these requirements was the gold-rhodium system. See Figure 6. This system provided an additional benefit since the color differences between the phases would greatly facilitate their identification during analysis.

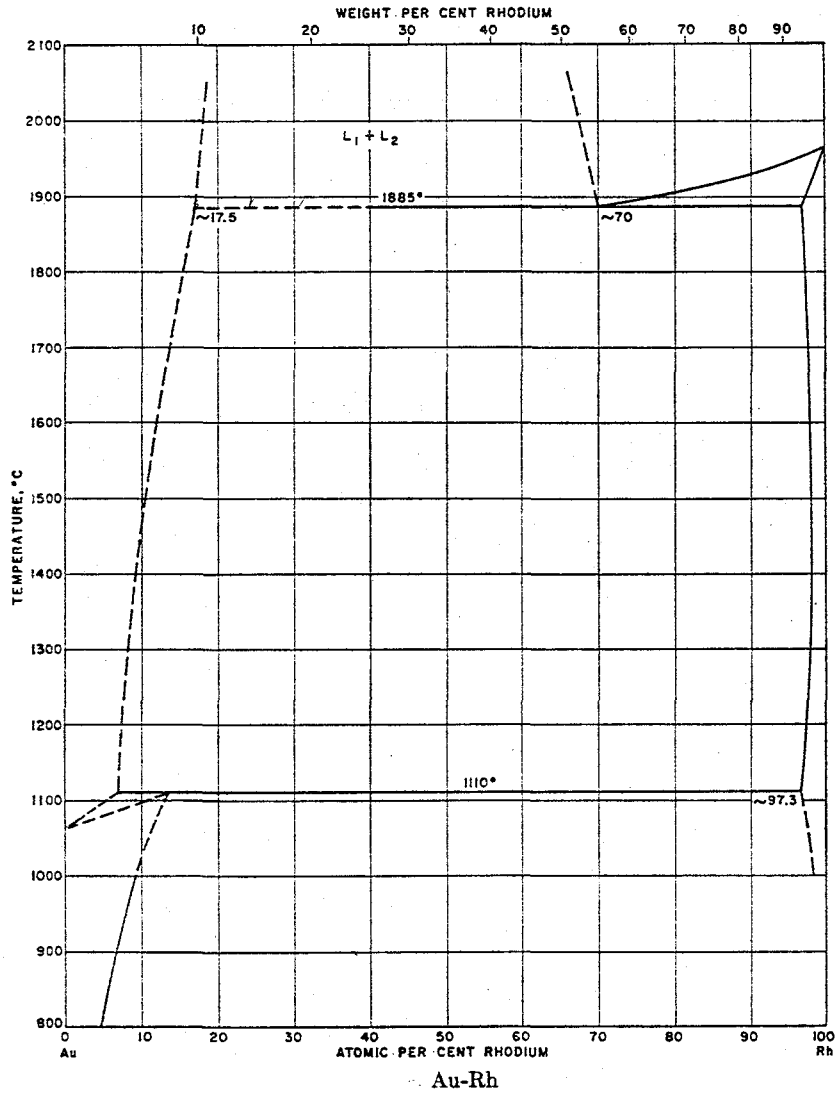


Figure 6. Phase diagram of the gold-rhodium alloy system.

EXPERIMENTAL PROCEDURE

Hypermonotectic gold-rhodium alloys ($Au_{35}Rh_{65}$) were produced by melting/alloying Marz purity elements together in a small arc melter. A tungsten electrode was utilized and melting and alloying were carried out under an argon atmosphere. The constituents were first melted together and then allowed to solidify. The solidified samples were then inverted and remelted. This was done several times. The structures obtained were roughly as expected; the more dense, gold rich immiscible phase tended to settle to the bottom of the sample. This was apparent from a simple analysis of the exterior of the samples. See Figure 7.

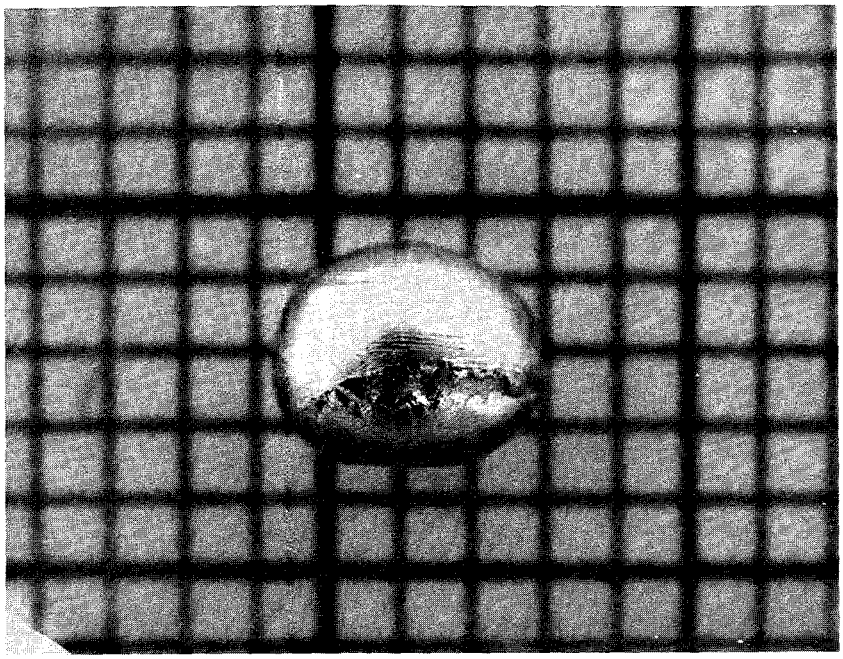


Figure 7. General appearance of an arc melted $Au_{35}Rh_{65}$ alloy. (10X)

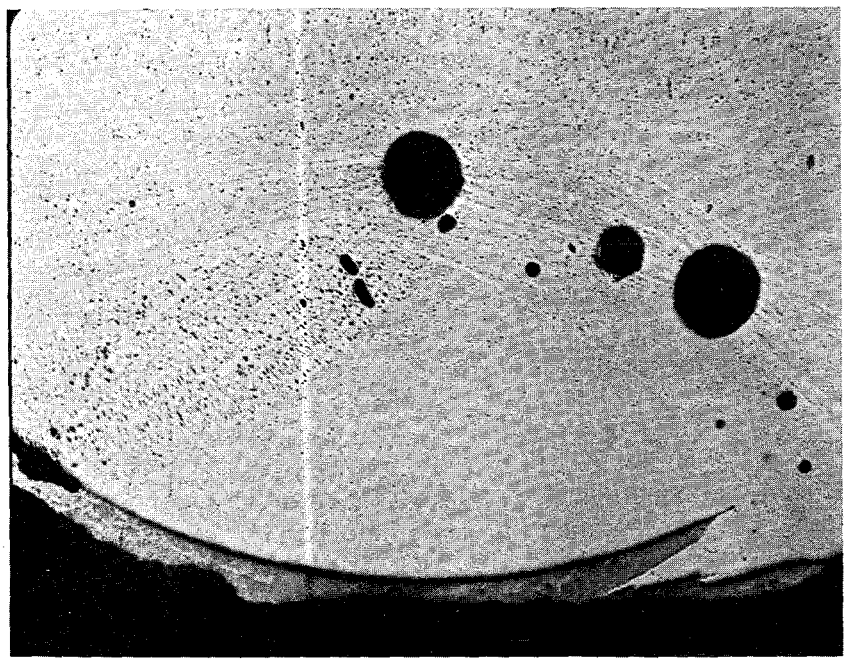


Figure 8. Section through an arc melted $Au_{35}Rh_{65}$ alloy. Note segregation. Grey areas are gold rich. (60X)

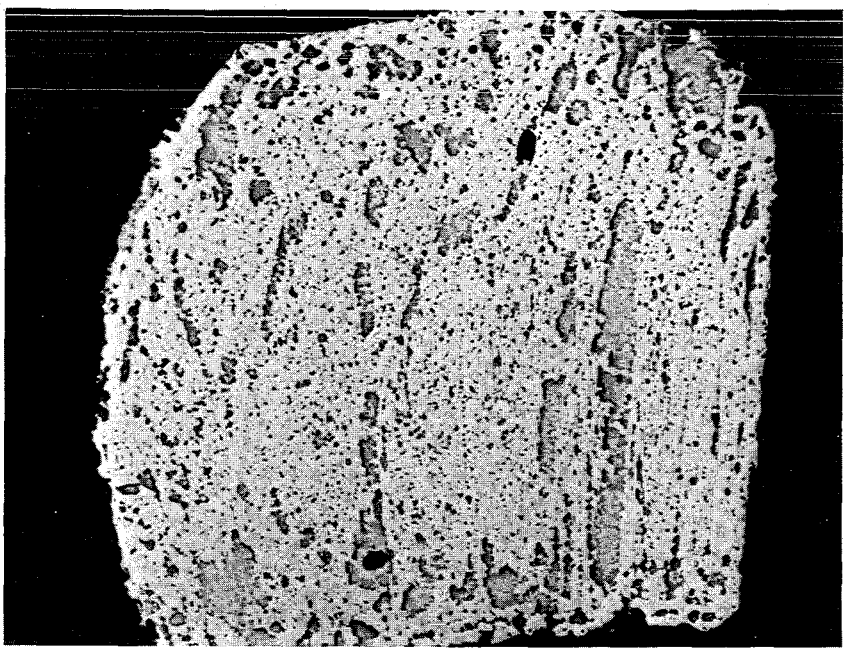


Figure 9. Fragment from dropped $Au_{35}Rh_{65}$ sample showing partially dispersed structure. Grey areas represent location of the gold rich minority liquid phase. (60X)

An arc melted sample was sectioned revealing a gold rich region near the bottom with a rhodium rich region of roughly monotectic composition being present near the top. See Figure 8.

Samples were then taken to the drop tube for levitation and melting using the electromagnetic levitator/furnace. An atmosphere of 500 Torr of argon was used in the bell jar assembly while 500 Torr of helium was utilized in the tube. Helium was used in order to aid in cooling during free fall.

RESULTS AND DISCUSSION

When taken to the drop tube for containerless, low gravity processing, the $Au_{35}Rh_{65}$ samples levitated nicely when power was initially applied to the levitator coil, but then fell unexpectedly from the coil after reaching a temperature of approximately 2200K. As a result, the first sample processed hit the tube isolation valve which is normally left closed until just before the samples are dropped. The fragmented sample was retrieved and later sectioned and polished using normal metallographic procedures. Etching was not required. Microscopic examination revealed the partially dispersed structure shown in Figure 9. From this micrograph it appears as if some melting of the sample occured before it fell from the coil. However, it is difficult

to tell if the sample reached a sufficient temperature to get into the single phase liquid region. Optical pyrometer data was obtained, but is somewhat suspect due to uncertainty in the emissivity of the samples.

It is not known at this time why these samples fell from the levitation coil during heating. It is speculated that they fell due to a change in their electrical characteristics which occured either at the monotectic temperature or at the temperature where a single phase liquid would from. Significant changes in the compositions and amounts of the phases will take place at both of these temperatures. As mentioned previously, due to the uncertainty in the emmisivity of the alloy it is difficult to determine the actual temperature at which the samples fell from the coil. However, data obtained in subsequent tests indicated only a 25°C spread between the temperatures at which four samples fell.

Bench tests were carried out to determine if the levitation coil power supply could be tuned to better levitate $Au_{35}Rh_{65}$ samples. These test were inconclusive but indicated that levitation of even the solid samples was considerably more difficult than for most alloys.

An additional set of $Au_{35}Rh_{65}$ alloys was prepared using the same procedure discussed previously. Four 300mg samples were prepared and one 650mg sample. One of the 300mg samples was saved for sectioning without dropping.

It was decided to attempt additonal drops by opening the tube isolation valve before power was applied to the levitation coil. With this approach, if the samples fell from the coil permaturely, they would fall the entire length of the tube instead of hitting the tube isolation valve. An atmosphere of 500 Torr of helium was used in the drop tube, while a mixture of helium with 6% hydrogen was used in the bell jar assembly. The samples again levitated nicely, at first, but then fell from the coil after reaching an elevated temperature. The samples fell the length of the tube and, with one exception, were intact when retrieved from the catch pan.

All of the dropped samples had a gold colored surface as opposed to the partially gold colored, partically silver colored surfaces on the arc melted samples. This was not totally unexpected. In several of the flight experiments that have been carried out on hypermonotectic alloys, surface tension driven segregation has been encountered due to one phase wetting the walls of the container [10]. It has been observed that if the minority liquid phase wets the container wall, flows occur in these alloys, even under microgravity conditions, that can lead to almost complete segregation. The resulting structure shows a central region which consists of the majority liquid phase, with the minority liquid being in the outer regions, in contact with the container. Similarly, it is possible for the minority liquid phase to wet the outer surface of a containerlessly melted and solidified sample. The gold colored outer surface is an indication that this may be taking place in $Au_{35}Rh_{65}$ alloys.

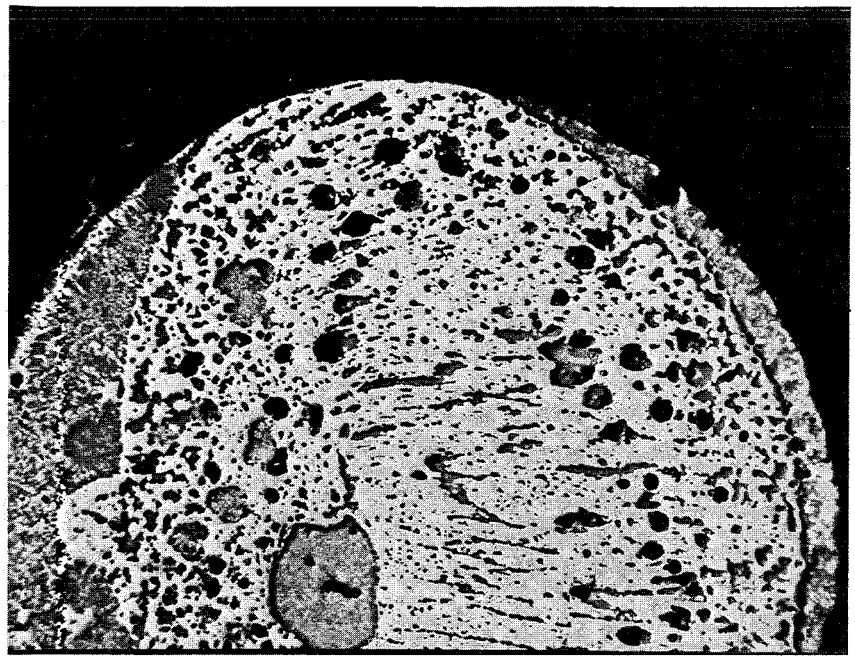


Figure 10. Dropped $Au_{35}Rh_{65}$ sample. Note wetting tendency of the gold rich minority liquid phase. (37.5X)

A section through one of the dropped $Au_{35}Rh_{65}$ samples is shown in Figure 10. The darker colored regions on each side of the spherical sample and dispersed throughout represent the location of the gold rich minority liquid phase when solidification took place. The tendency for this minority liquid to wet the outer surface is evident from the low contact angle the interface makes with the surface.

Several additional $Au_{35}Rh_{65}$ samples were levitated and dropped. The structures that resulted from two of these samples are shown in Figures 11 and 12. Note that in both figures, regions of unalloyed pure rhodium (white areas) are present. More segregation is present in the sample shown in Figure 11 than in Figure 10. However, there are regions where relatively fine dispersions of the phases are present. Figure 12 shows a more uniform dispersion of phases than in either of the two previous samples. While not easily discernible from the photomicrographs, the surfaces of the latter two samples were also coated by the gold rich, minority liquid phase.

Obviously, for samples of hypermonotectic composition, primary rhodium regions should not be present. The rhodium regions present in

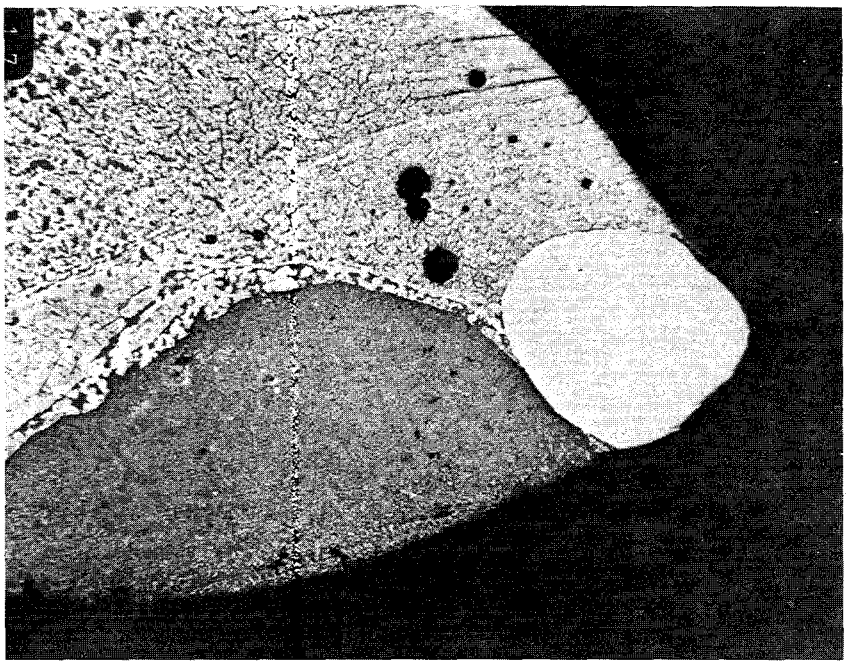


Figure 11. Dropped $Au_{35}Rh_{65}$ sample. The white area is unalloyed rhodium. (37.5X)

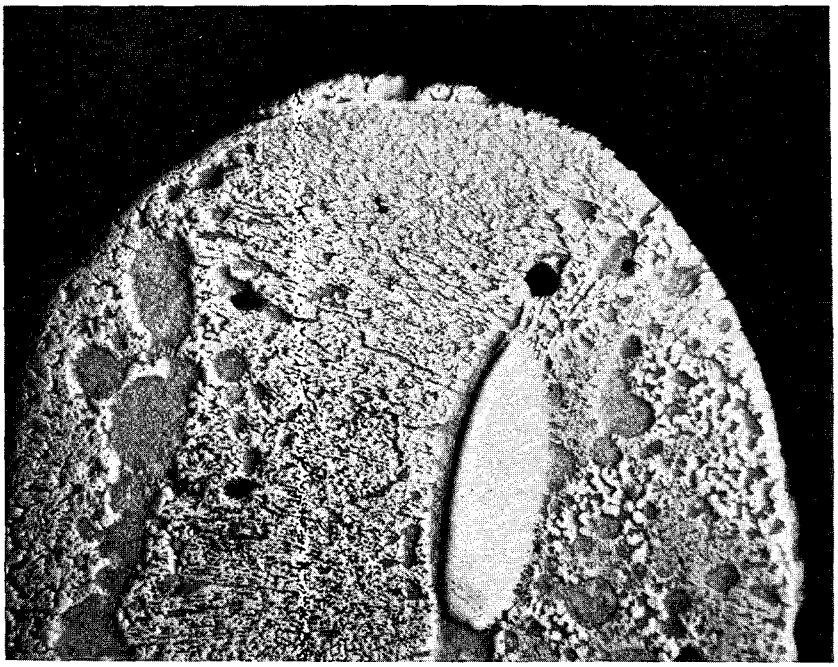


Figure 12. Dropped $Au_{35}Rh_{65}$ sample. The white area is unalloyed rhodium. (37.5X)

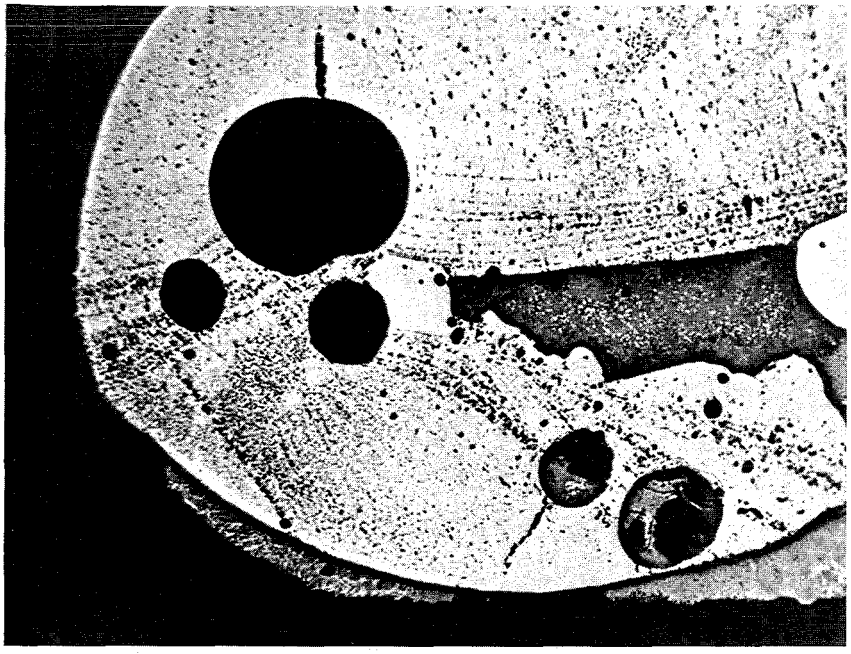


Figure 13. Arc melted $Au_{35}Rh_{65}$ sample. The white areas indicate unalloyed rhodium was present after arc melting. (37.5X)

Figures 11 and 12 were essentially pure, indicating they probably resulted from incomplete alloying in the initial preparation of the samples in the arc melter. To test this, an additional arc melted sample was sectioned and prepared for metallographic analysis. Analysis revealed areas of pure rhodium (See Figure 13) indicating greater care is needed in preparation of these alloys. The normal procedure was to repeatedly invert and remelt samples until an approximately spherically shaped sample was obtained (four to five attempts were usually required.) After obtaining a spherical shape, samples were usually inverted and remelted one or two additional times. This should be expanded to at least five times to ensure complete alloying of the high melting rhodium component.

While the previous samples had a mass of 300 mg each, one much larger sample of approximately 650mg mass was also produced. This sample was prepared to determine what effect, if any, sample size would have on the tendency of the samples to fall out of the levitation coil when heated.

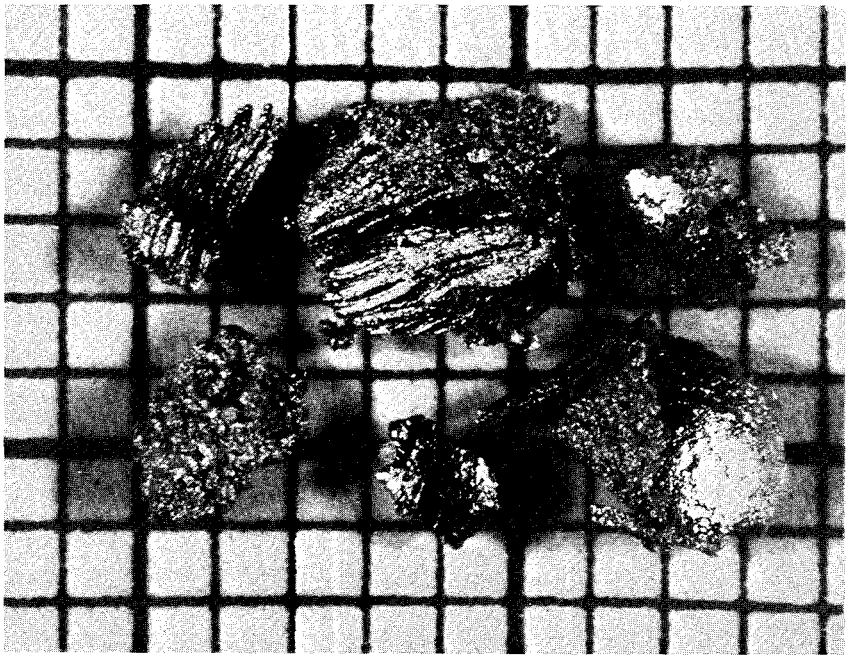


Figure 14. Fragmented 650mg $Au_{35}Rh_{65}$ sample. Note the fiborous appearance of several of the fragments. (10X)

In previous attempts, 300 mg $Au_{35}Rh_{65}$ samples had fallen from the levitator after approximately five seconds of full power. In the case of the 650mg sample, levitation took place for at least 15 seconds. In addition, a definite shape change was noted in the sample before it fell. When retrieved, the sample was found to have fractured upon impact at the bottom of the tube. Several of the fragments had a rather unusual fiborous appearance. See Figure 14. This structure was not expected.

Portions of the sample were sectioned perpendicular and parallel to the fiber axis and the resulting surfaces prepared for metallographic examination. Figure 15 shows the microstructure revealed by the section parallel to the fiber axis. (Note the higher magnification.) Filaments of the gold rich phase are apparent in the sturcture. A transverse view is shown in Figure 16. It can be seen from Figure 16 that the gold rich filaments are not rod like. This micrograph indicates there may be a relatively high connectivity in the structure.

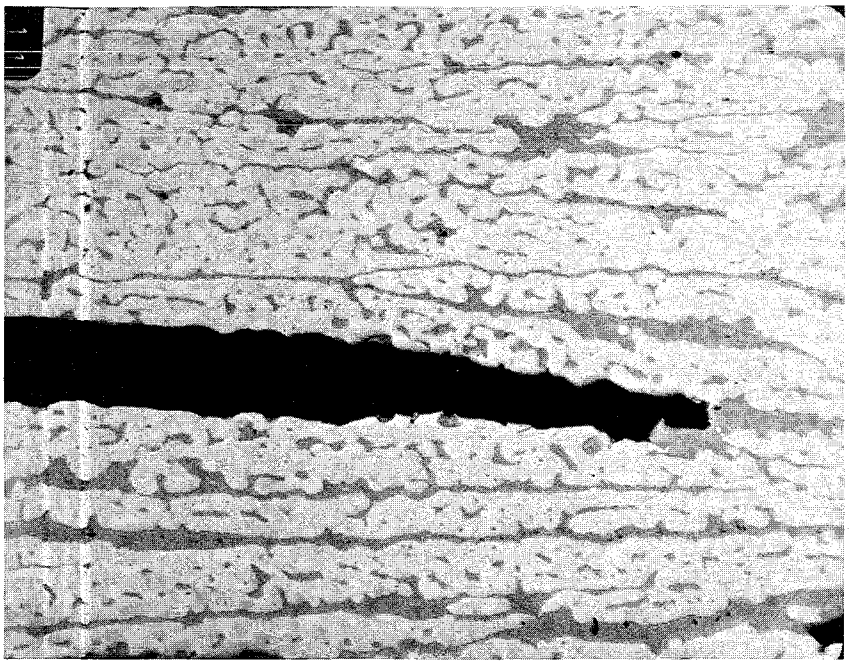


Figure 15. Section parallel to the fiber axis in an aligned structure obtained in a $Au_{35}Rh_{65}$ drop tube sample. (240X)

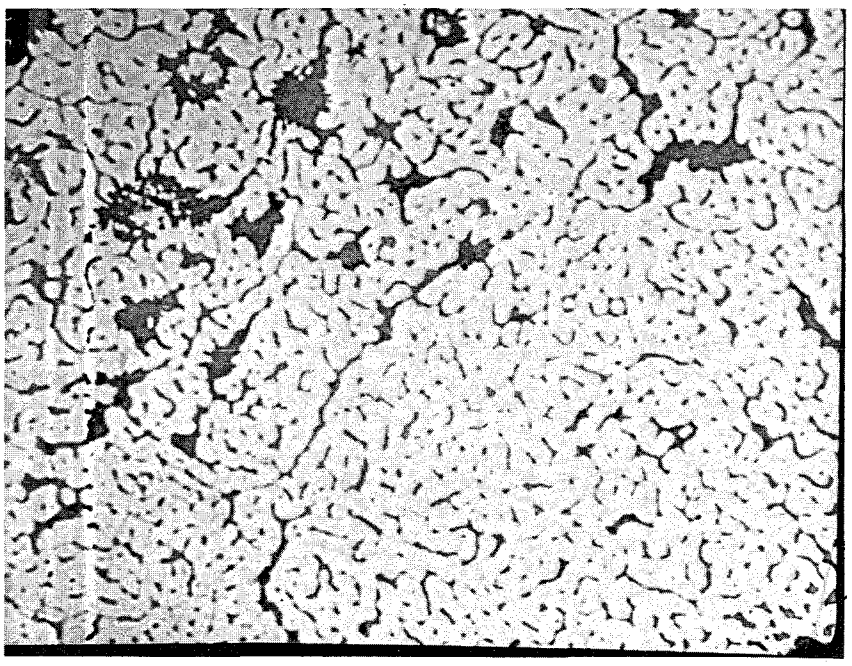


Figure 16. Section perpendicular to the fiber axis in an aligned structure obtained in a $Au_{35}Rh_{65}$ drop tube sample. (240X)

To date, the cause of the aligned structure is uncertain. Aligned microstructures are normally found in directionally solidified samples. It is possible that some degree of directionality of heat flow occured in the dropped samples due to convective heat transfer. But this is purely speculative. Additional testing is required to determine the cause of these structures.

Additional research is planned in this area. The structures obtained to date are quite unusual and warrant further investigation. As a first step, it will be necessary to improve the ability to levitate these samples in order to make certain they are heated into the single phase liquid region before they are dropped. It is also desired to melt and drop Au-Rh samples of varying compositions in a vacuum instead of an inert atmosphere in order to eliminate any influence of aerodynamic drag on the solidification process. Many exciting and informative experiments are possible.

CONCLUSIONS

It appears it may be possible to obtain disperisons of immiscible phases in metallic systems through use of the drop tube. Preliminary tests indicate a tendency towards the formation of a dispersion in $Au_{35}Rh_{65}$ samples. However, experimental difficulties have been encountered in levitating samples for a sufficient length of time to ensure the attainment of a single phase liquid before samples are dropped.

From research carried out to date, the following conclusions can be drawn.

1. Hypermonotectic $Au_{35}Rh_{65}$ samples processed under 1 g conditions exhibit gross segregation due to sedimentation of the more dense, gold rich, minority liquid phase. This segregation can be easily seen by eye due to the color differences between the upper and lower regions in arc melted samples.

2. The majority of the levitated and dropped samples contain a more dispersed microstructure than samples processed under lg conditions.

3. There is a definite tendency for the gold rich, minority liquid phase in $Au_{35}Rh_{65}$ alloys to wet the outer surface of dropped samples. This is readily apparent due to their gold colored outer surface.

Additional tests are necessary to explain some of the unusual microstructures obtained in this alloy system. Plans are currently being made to continue this investigation.

REFERENCES

1. L. L. Lacy, M. B. Robinson, T. J. Rathz, "Containerless Undercooling and Solidification in Drop Tubes". Journal of Crystal Growth, vol. 51, page 47 (1981)

2. L. L. Lacy, T. J. Rathz, M. B. Robinson, "Containerless Undercooling and Solidification of Bulk Metastable Nb_3Ge Alloys," J. Appl. Phys., vol 53, page 682 (1982)

3. L. L. Lacy, M. B. Robinson, T. J. Rathz, N. D. Evans, R. J. Bayuzick, "Solidification Studies of Nb-Ge Alloys at Large Degrees of Supercooling", Presented at the Materials Research Society Symposium, Boston, Massachusetts, November 1981.

4. M. B. Robinson, "Undercooling Measurements in a Low-Gravity Containerless Environment", NASA Report TM-82 423, May 1981.

5. R. J. Bayuzick, N. D. Evans, W. F. Hofmeister, K. R. Johnson, M. B. Robinson, "A Review of Long Drop Tubes as a Supplement/Alternative to Space Experiments", Adv. Space Res., vol 4, page 85, (1984)

6. W. Klement, Jun., R. H. Willens, Pol DuWez, "Non-crystalline Structure in Solidified Gold-Silicon Alloys", Nature, vol. 187, page 869, (1960)

7. P. Chaudhari, D. Turnbull, "Structure and Properties of Metallic Glasses", Science, vol. 199, page 11, (1978)

8. I. W. Donald, H. A. Davies, "Prediction of Glass-Forming Ability for Metallic Systems", J. Non-Crystalline Solids, vol. 30, page 77, (1978)

9. A. J. Drehman, A. L. Greer, D. Turnbull, "Bulk Formation of a Metallic Glass: $Pd_{40}Ni_{40}P_{20}$", Appl. Phys. Lett., vol. 41, page 716, (1982)

10. S. H. Gelles, A. J. Markworth, "Microgravity Studies in the Liquid-Phase Immiscible System: Aluninum-Indium", AIAA Journal, vol 16, page 431, (1978)

11. T. Carlberg, H. Fridriksson, "The Influence of Microgravity on the Solidification of Zn-Bi Immiscible Alloys", Met. Trans. A, vol. 11A, page 1665, (1980)

12. C. Potard, "Solidification of Hypermonotectic Al-In Alloys Under Microgravity Conditions", Materials Processing in the Reduced Gravity Environment of Space, G. E. Rindone, ed., Elsevier Science Publishing Company, Inc. page 543, (1982)

1985

NASA/ASEE SUMMER FACULTY RESEARCH FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER

THE UNIVERSITY OF ALABAMA AT HUNTSVILLE

THE EFFECT OF GROWTH RATE ON THE COMPOSITIONAL VARIATIONS IN DIRECTIONALLY SOLIDIFIED $Hg_{1-x}Cd_xSe$ ALLOYS

Prepared By:	Rosalia N. Andrews, Ph.D.
Academic Rank:	Associate Professor
University and Department:	Department of Materials Engineering University of Alabama at Birmingham
NASA/MSFC: Division: Branch:	 Low Gravity Science Solid State and Solidification Physics
MSFC Counterpart:	Sandor L. Lehoczky Frank R. Szofran
Date:	August 14, 1985
Contract No.:	NGT-01-008-021 The University of Alabama at Huntsville

THE EFFECT OF GROWTH RATE ON THE COMPOSITIONAL VARIATIONS IN DIRECTIONALLY SOLIDIFIED $Hg_{1-x}Cd_xSe$ ALLOYS

BY

Rosalia N. Andrews
Associate Professor of Materials Engineering
University of Alabama at Birmingham
Birmingham, Alabama 35294

ABSTRACT

Several $Hg_{1-x}Cd_xSe$ crystals of composition x = 0.2 were grown in a Bridgman-type directional solidification furnace at varying translation rates. The influence of growth rate on both the longitudinal and radial compositional uniformity for the crystals was determined using density measurements and infrared transmission-edge mapping.

ACKNOWLEDGEMENTS

I would like to express my appreciation to Gerald Karr, Leroy Osborn, Jim Dozier, and Larry Freeman for putting together a summer program which was very informative and enjoyable. I would also like to acknowledge the financial support of the Space Science Lab for my position on the ASEE/NASA Summer Faculty Fellowship Program.

Thanks go to Alice Dorries for performing the precision density measurements and to Ron Harris for his expert assistance in the slicing of crystals and in metallographic sample preparation. Thanks are also extended to Shirley Buford for her friendship and secretarial support this summer, and to Tony Xudis and Ching Hua Su, my labmates, for their assistance in various aspects of the laboratory work. Special thanks go to Frank Szofran for his day to day help and guidance in the conduction of the laboratory experiments. Also, his ability to work with and modify the IR transmission analysis system used in this project was invaluable to the successful conduction of the work.

Most of all I would like to express my sincere appreciation to both my NASA counterparts, Alex Lehoczky and Frank Szofran. Their interest in the project and involvement on a daily basis made my ten weeks a most enjoyable and professionally rewarding experience. Their patience in explaining the theory and guiding me through the many various operations in the laboratory was critical to my developing a working understanding of the project in such a short amount of time. Frank and Alex were my friends and teachers and I feel very fortunate to have been given the opportunity to get to know them and work with them this summer.

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

With an increasing emphasis on the growth of semiconducting crystals which can be used for the detection of infrared radiation and for use in the fabrication of junction devices, considerable attention has been focused on attaining a basic understanding of the mechanisms involved in the growth of homogeneous, bulk single crystals of these materials. Of particular interest is the ability to control both the longitudinal and radial compositional uniformity of directionally solidified crystals by controlling such factors as growth rate, interface shape, and temperature distributions during the growth process.

A considerable amount of work has been done to date on the HgCdTe system. There are some excellent reviews available[1,2] as well as a vast number of publications in the technical literature on this system. To cite just a few, which are concerned primarily with the influence of growth parameters on the solidification process of these crystals, Szofran et al.[3] have determined that the growth rate has a significant influence on the longitudinal compositional profile of HgCdTe crystals and that the measured radial compositional variations in these crystals also shows a strong correlation to both the growth rate as well as the temperature profile used. Bartlett, et al.[4] also studied the effects of growth rate on the compositional variations in HgCdTe crystals and found that again the uniformity is highly growth rate dependent. In several other works [5-7], it has also been shown that the shape of the melt-crystal interface is an important factor in controlling the quality of directionally solidified HgCdTe crystals. In turn, it has been shown

that the shape of the melt-solid interface is related to the relative thermal conductivities of the solid and liquid.[8,9] Lehoczky and Szofran,[10] in a summary article, discussed the influence of all these parameters, the growth rate, the interface shape and the solid and liquid alloy conductivities, on the compositional uniformity of HgCdTe crystals. Therefore, since it has been shown that all these factors influence compositional uniformity and hence crystal quality, it is extremely important in the growth of compositionally homogeneous semiconducting crystals to have a complete understanding of both the thermophysical properties of the materials involved as well as an understanding of the influence of various growth parameters on the solidification process.

In an attempt to characterize and thus better understand the growth process of semiconducting crystals, one procedure which has been used by several investigators is to compare the experimentally measured longitudinal compositional profiles of as grown crystals to theoretical compositional profiles assuming various models. One model which appears to fit the data quite well is the one-dimensional diffusion model. In this fitting procedure, it is possible to determine an effective diffusion coefficient in the liquid. In looking at the results of this modeling by Szofran et al.[3] on a Bridgman grown ingot of $Hg_{1-x}Cd_xTe$ of composition x=0.202, (Figure 1) it is seen that indeed the measured longitudinal composition profile is well modeled by the one-dimensional diffusion case. This result indicates that during the growth of HgCdTe crystals there will be a region of the crystal over which steady state solidification is proceeding and in which the average axial composition of the crystal is constant.

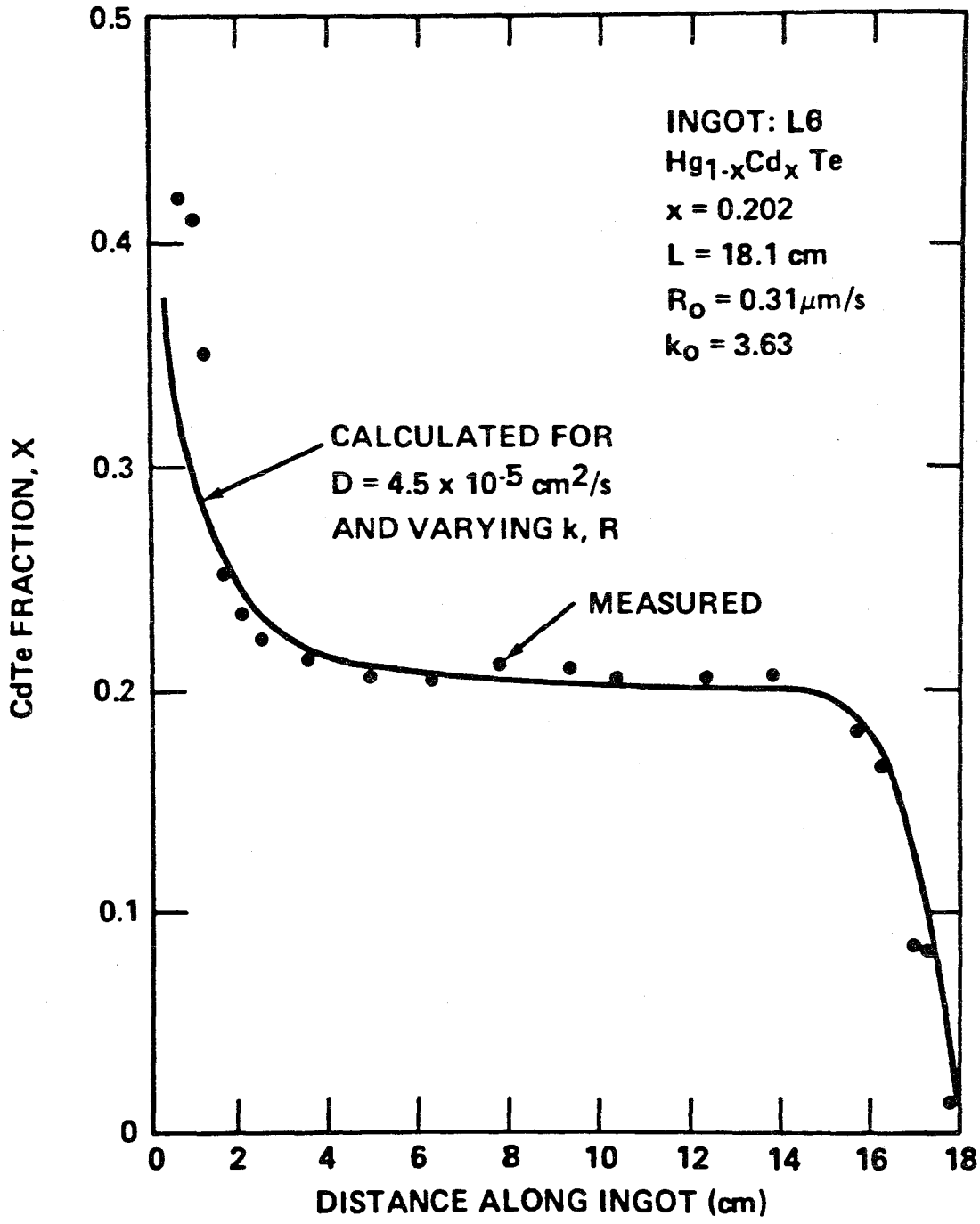


Figure 1. Theoretical and experimental profiles for a $Hg_{1-x}Cd_xTe$ crystal of composition X = 0.202 grown at 0.31 um/s.

In looking through the literature on semiconducting alloys, it was observed that in work done by Summers and Nelson[11] on a similar system, the HgCdSe system, the agreement between the measured and calculated compositional profiles, assuming a one dimensional model, was not quite as good as in the work previously quoted on HgCdTe. This lack of agreement can be seen in Figure 2. It was suspected that perhaps this disagreement, in particular the initial dip in composition followed by the continuously increasing cadmium composition throughout the normally observed steady state growth region was due to the use of a relatively fast growth rate of lum/sec.

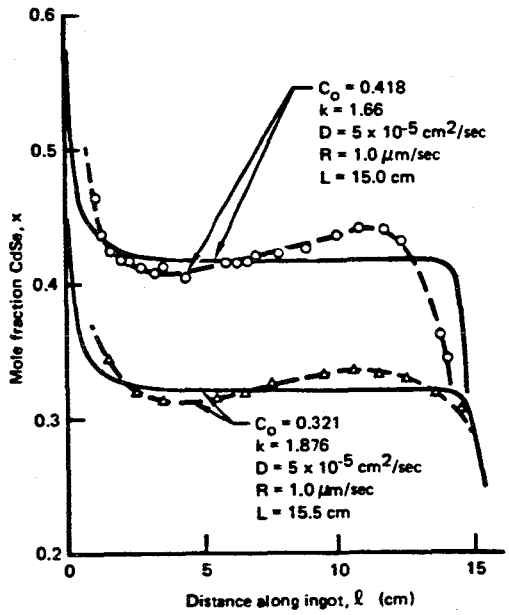


Figure 2. Theoretical and experimental compositional profiles of $Hg_{1-x}Cd_xSe$ ingots grown by the fast Bridgman method,-----, theory;----,experiment.

In looking into this system further, it was noticed that although a considerable amount of work has been done on the HgCdTe system, a lesser amount of attention has been paid to the HgCdSe system.

Cruceanu and Niculescu[12], in 1965 were the first to report in the literature the preparation of $Hg_{1-x}Cd_xSe$ alloys. Kalb and Leute[13] used x-ray techniques to determine the miscibility gap limits in the HgSe-CdSe system. They determined that for $x<0.77$, $Hg_{1-x}Cd_xSe$ alloys possess the cubic zincblende structure and that for $x>0.81$, they possess the hexagonal wurtzite structure. At room temperature, the two crystallographic phases are immiscible for $0.77<x<0.81$.

Nelson, Summers, and Whitsett[14] determined the phase diagram for the HgSe-CdSe system using differential thermal analysis measurements. Nelson et al.[15] performed an experimental and theoretical study of the electron mobility in the $Hg_{1-x}Cd_xSe$ alloy system. They grew some crystals by the Bridgman method and in their work, they discussed primarily the intrinsic and defect-scattering processes dominant in HgCdSe. A fairly comprehensive study of HgCdSe alloys was done by Whitsett, et al.[16] in which the preparation and characterization of HgCdSe alloys was emphasized. Several different growth methods were used in this investigation. There are a few additional studies on HgSe and HgCdSe, however, this brief review accounts for the majority of the published work on the growth of HgCdSe crystals.

There are several reasons for the rather limited work on this system in comparison to the HgCdTe system. One primary reason is that HgCdSe is an n-type semiconductor as grown and remains so under all types of subsequent treatment. This, coupled with the fact that the carrier

concentration is quite high (10^{18}) in the as grown state, poses several problems in the utilization of HgCdSe for many applications. In contrast, HgCdTe is p-type as grown and can be made n-type by annealing in a mercury vapor. This capability adds greater flexibility in the area of device production.

Also, HgCdSe appears to be quite unstable electrically, thus making its commercial utilization at this time unattractive. However, the lattice characteristics of HgCdSe are quite favorable for growing strain free crystals and if a better understanding of the system can be obtained, it may prove to be a feasible system for further investigation and development.

Therefore, because of the limited published results on the HgCdSe system, coupled with the interesting results of Summers and Nelson,[11] and the apparent potential usefulness of HgCdSe as a semiconducting material provided certain instabilities can be overcome, it was decided to further investigate this system.

OBJECTIVE

The objective of this work was to gain a better understanding of the influence of growth parameters on the compositional uniformity of Bridgman grown HgCdSe crystals. In particular, it was decided to determine the influence of growth rate on both the longitudinal and radial compositional uniformity of HgCdSe crystals of fixed composition. These crystals were directionally solidified in a Bridgman-type crystal growth furnace and their compositional profiles, both axially and

radially, were determined using precision density measurements and infrared transmission edge mapping.

An ultimate objective of this project will be to compare the measured longitudinal profiles with the theoretical model for a one-dimensional diffusion case and to determine an effective diffusion coefficient. Also, it is desired to utilize the radial compositional variations in conjunction with the particular growth parameters to infer the thermophysical properties of the material.

HgSe-CdSe SYSTEM

HgSe crystallizes in the zincblende structure and CdSe crystallizes in the wurtzite structure. These two compounds combine to form what is referred to as a pseudobinary system which is shown in Figure 3. As noted on this diagram, and determined by Kalb and Leute,[13] this system contains a small miscibility gap. The existence of this peritectic reaction at 947° does not appear to disturb significantly the solidus and liquidus curves which vary smoothly between HgSe and CdSe.

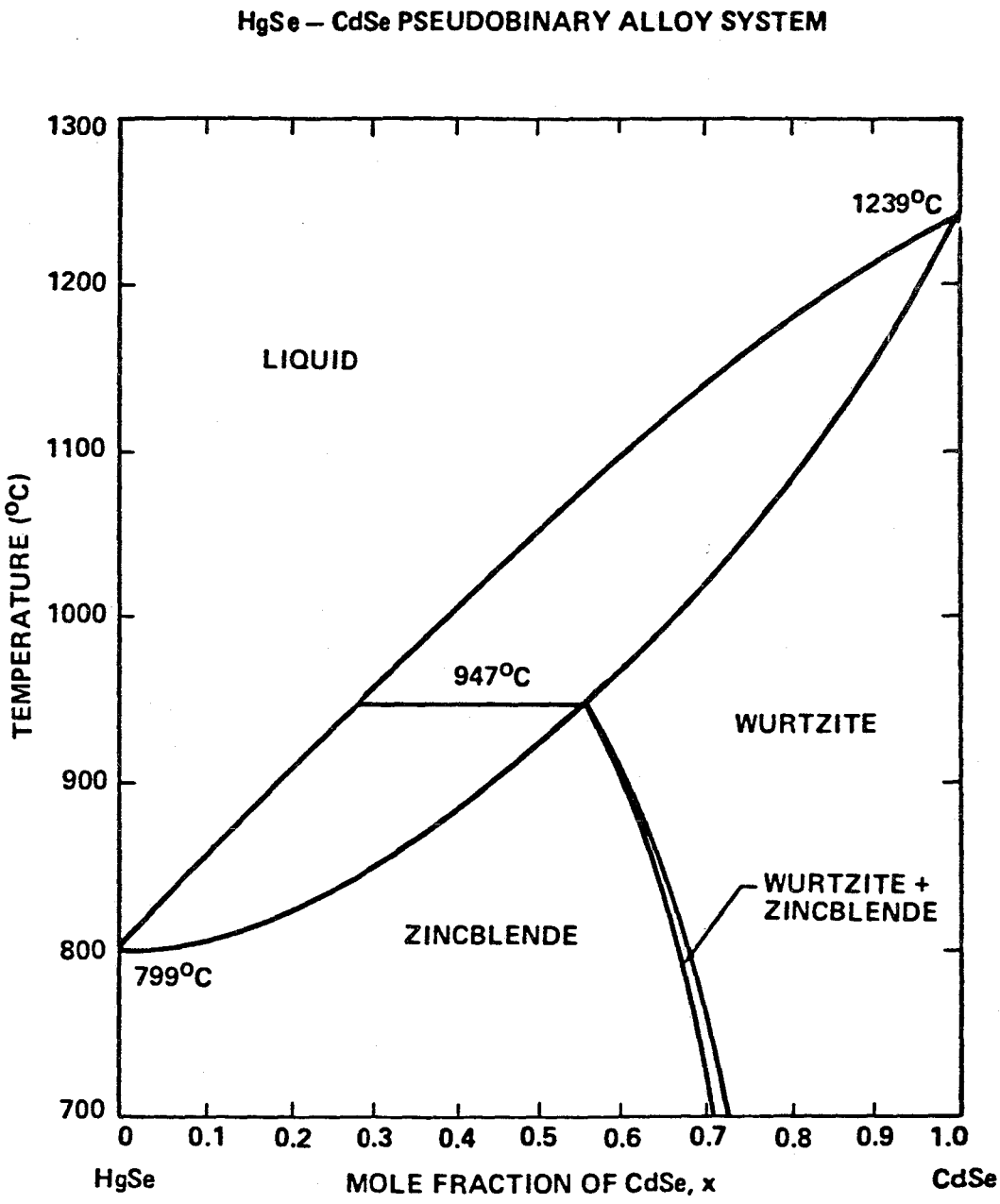


Figure 3. $Hg_{1-x}Cd_xSe$ phase diagram

Pseudobinary alloys of HgSe and CdSe are commonly represented by the formula $Hg_{1-x}Cd_xSe$ where x is the mole fraction of CdSe. With increasing values of x, it is seen in Figure 4 that the variation in lattice constant is quite small, with the variation being smaller than that for the HgCdTe system. This would be advantageous in the growth of HgCdSe crystals because slight variations in composition would not result in significant strains being introduced in the crystals due to lattice constant changes.

Mercury selenide is classified as a perfect semimetal or symmetry-induced zero-gap semiconductor. On the other hand, cadmium selenide is a wide gap semiconductor with a fundamental direct band gap of 1.8eV. When HgSe and CdSe combine to form $Hg_{1-x}Cd_xSe$, the energy gap increases continuously from a small negative value for HgSe to a larger positive value as a function of x. This transition is represented schematically in Figure 5. Thus, it is possible by varying the alloy composition to tailor the band gap of HgCdSe crystals and hence the wavelength at which they become transparent to infrared radiation. Figure 6 from Whitsett, et al.[16] shows this relationship between the band gap and x at 300°K.

However, the electrical properties of HgCdSe are quite unstable thus making its use commercially unattractive at this time. Nevertheless, as mentioned earlier, the lattice characteristics are quite desirable as far as growing strain free crystals and if the system can be better understood and controlled, it may prove to be a feasible system for further investigation and development.

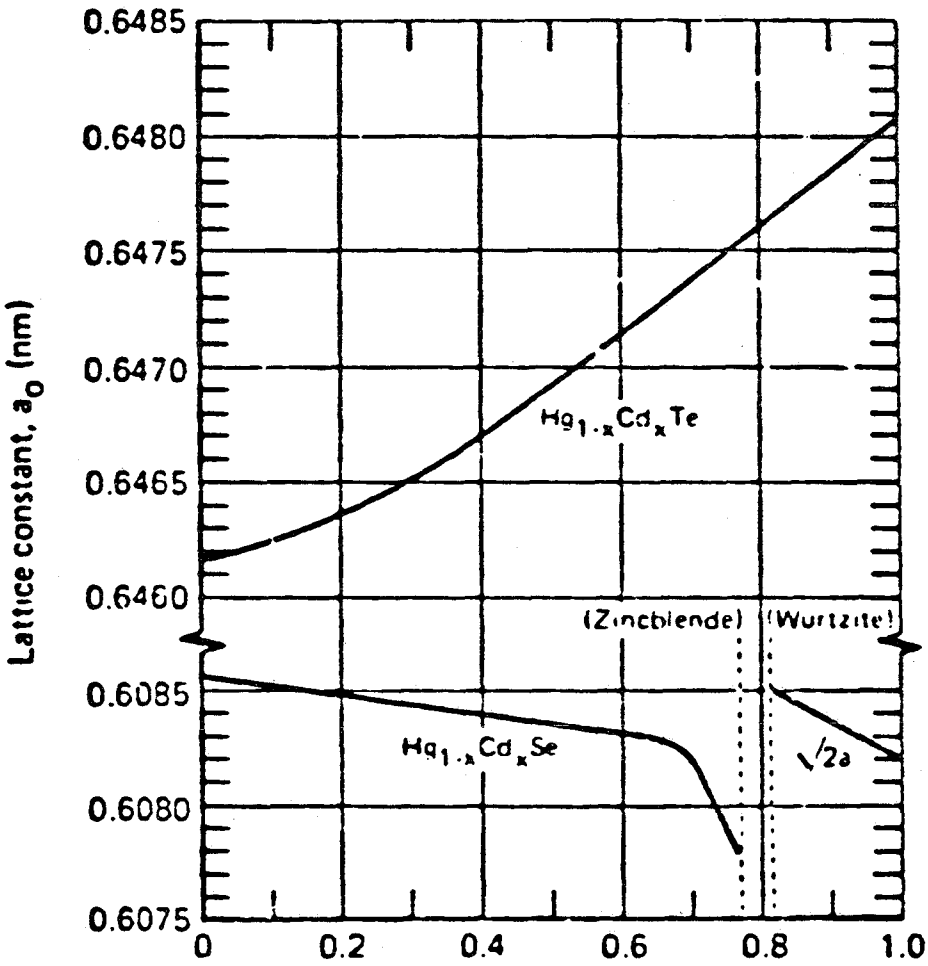


Figure 4. Variations of cubic unit-cell lattice constant with x in the $Hg_{1-x}Cd_xTe$ and $Hg_{1-x}Cd_xSe$ systems

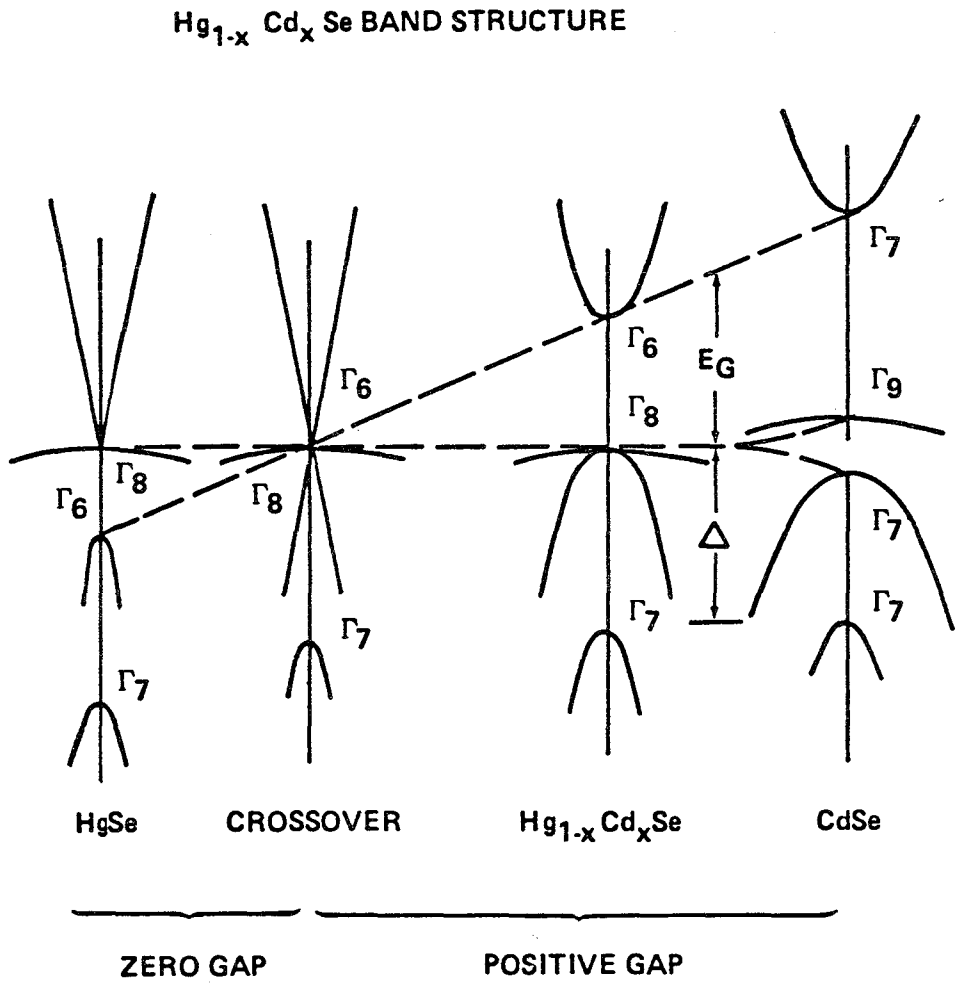


Figure 5. Energy-band model for $Hg_{1-x}Cd_xSe$ alloys

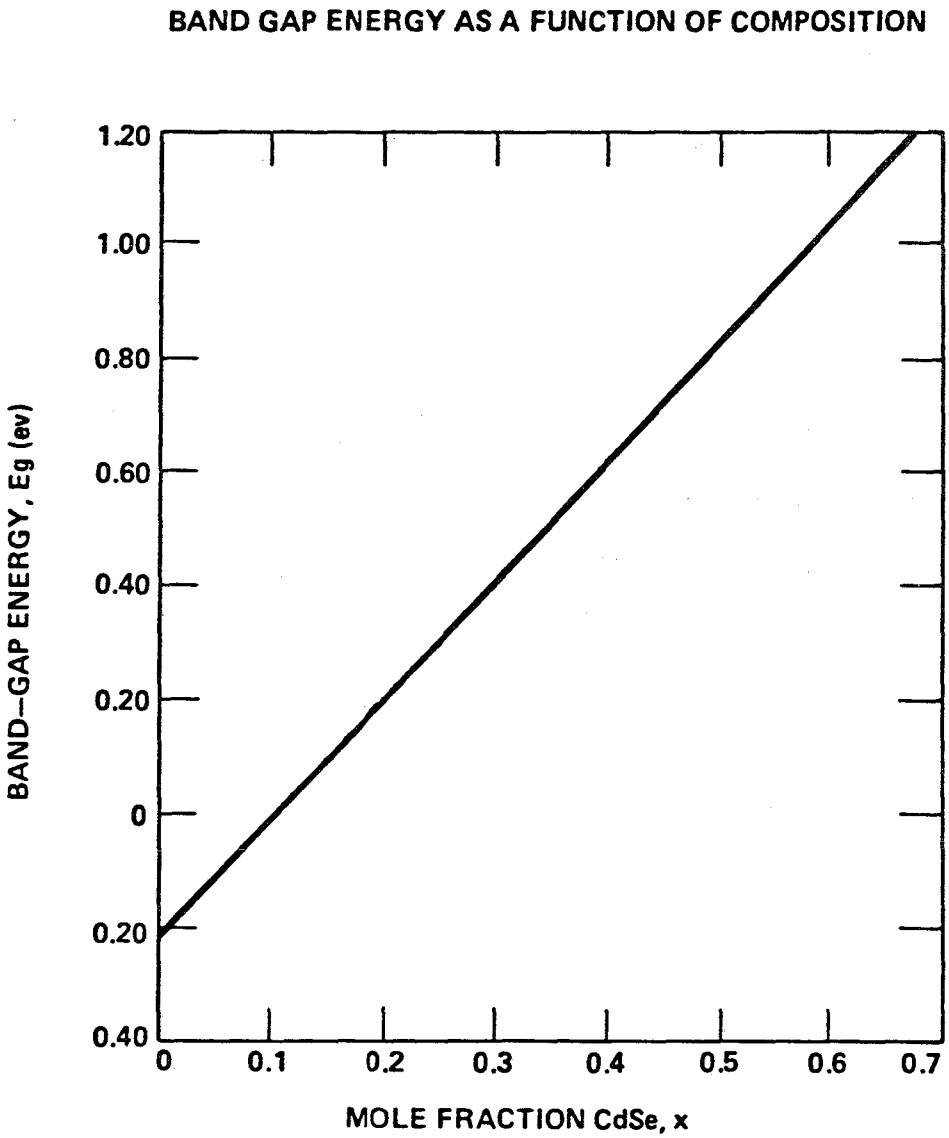


Figure 6. Compositional dependence of the fundamental energy gap of $Hg_{1-x}Cd_xSe$ at 300°K

EXPERIMENTAL PROCEDURE

All alloys used in this investigation were prepared by reacting 5-9's pure selenium(Se) 6-9's pure cadmium(Cd) and 7-9's pure mercury(Hg) in sealed, evacuated quartz ampoules of 5mm I.D, 10mm O.D., and 40 cm in length. The quartz ampoules were tapered to a point on one end to facilitate single crystal growth during subsequent directional solidification. The ampoules were first cleaned thoroughly with methanol and water and annealed at 1140°C for approximately three hours to relieve any strains induced during previous glass working processes.

Stoichiometric amounts of Hg, Cd, and Se were precisely weighed out for an alloy composition of x=0.200. The Cd and Se were inserted into the tube and the Hg was weighed into a glass container for subsequent insertion into the ampoule. The ampoule and the glass container holding the mercury were assembled and attached to a vacuum system through an O-ring fitting. The system was evacuated to a pressure of approximately 10 millitorr. The mercury was then inserted into the ampoule and the ampoule was sealed off.

The ampoule containing the Hg, Cd, and Se was then wrapped in high temperature wool and placed in a quartz lined inconel tube inside a rocking furnace. Care was taken to ensure that the ampoule was in the center of the tube furnace. The furnace was heated up slowly, particularly through the 700°C range. After the solidus temperature was reached, the furnace was rocked in an attempt to homogenize the alloy. After reaching the maximum temperature, which ranged from 80°C to 160°C above the liquidus temperature, the furnace was rocked for an additional

fifteen hours. At the end of this time, the furnace was oriented in a slightly inclined position from horizontal and the tip of the ampoule was lowered out the bottom of the furnace. This was done to initiate solidification at the ampoule tip.

After solidification, the ampoule was placed in a Bridgman-type directional solidification furnace. The alloy was remelted, and then directionally solidified at a particular growth rate and under specific thermal conditions.

A diagram of the furnace is shown in Figure 7. In this configuration, the quartz ampoule is placed on a quartz pedestal, and at the beginning of the growth process is positioned in the upper furnace. The ampoule remains stationary throughout the process and the furnace translates upwards. The furnace assembly consists of two resistively heated tubular furnaces with an upper hot zone and a lower cold zone. The two heat zones are well defined and isothermal due to the use of a sodium heat pipe in the upper zone and a potassium heat pipe in the lower zone. The two heat pipes are separated by a thermal barrier, the thickness of which is variable depending on the desired temperature profile for the particular growth process. Barrier thicknesses used in this study were 2.4cm and 0.64cm. The circular opening in the barrier was 1.19cm.

Before each crystal growth experiment, a temperature profile of the furnace was recorded to establish the thermal conditions present in the furnace. The profile is dependent on the particular combination of upper and lower zone temperatures as well as the barrier thickness. The particular profile which exists in the furnace determines the relative position of the solid-liquid interface with respect to the barrier.

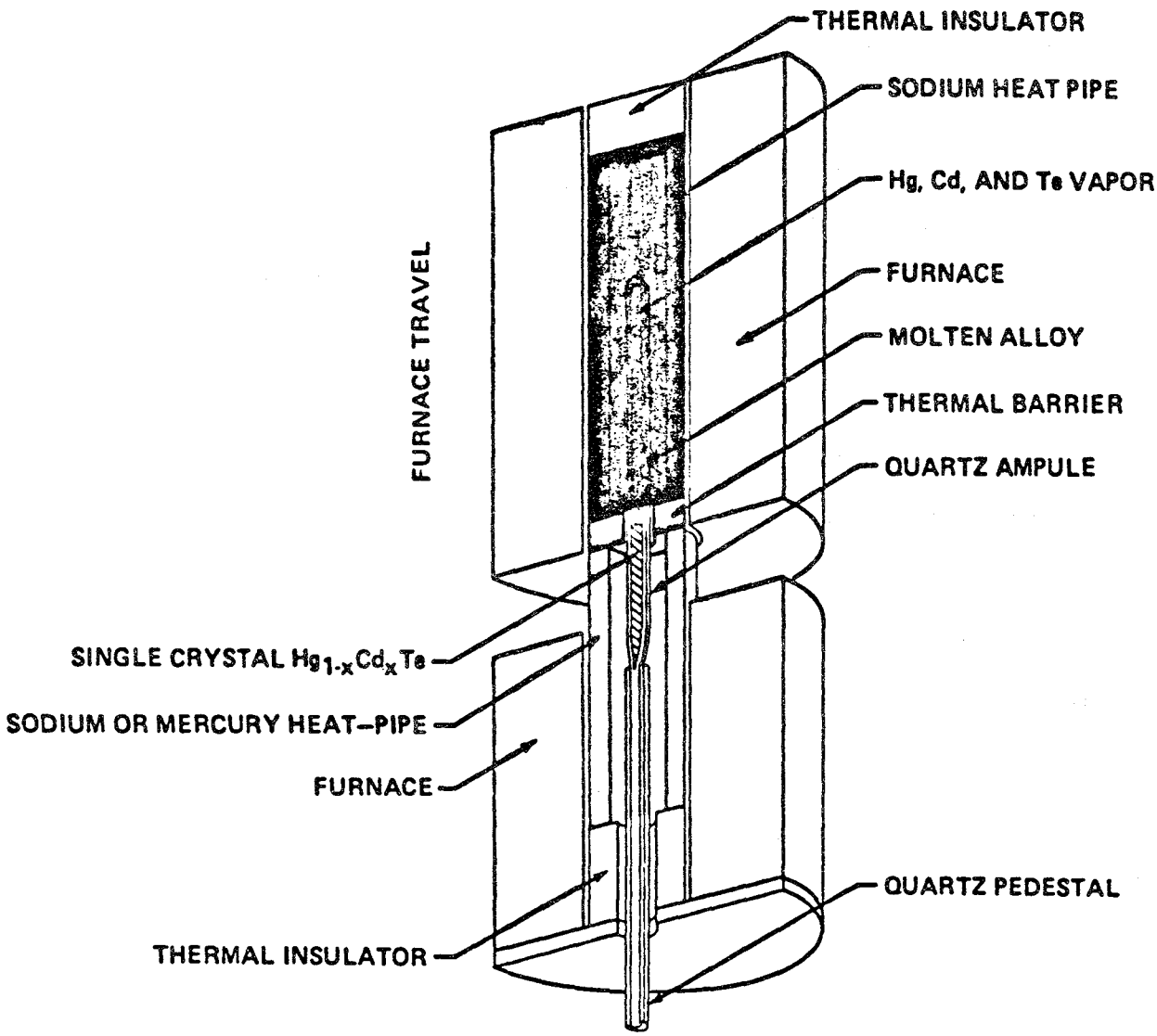


Figure 7. Bridgman-Stockbarger crystal-growth furnace assembly

The proper positioning of this interface is quite important in that it may directly affect the shape of the interface and hence the radial compositional uniformity.

The thermal gradient at the interface also is an important consideration in the growth of semiconducting crystals in that it must be steep enough to prevent constitutional supercooling. Of primary importance is the gradient at the solidus temperature, corresponding to the steady state growth interface. However, care must also be taken to ensure that the thermal gradient is steep enough at the liquidus temperature where initial solidification begins. From a knowledge of the initial segregation coefficient, k, and the thermal gradient G obtained from a temperature profile, the maximum permissible growth rate which should avoid constitutional supercooling was determined using the followingrelationship:

$$G/R=\frac{mC_o(k-1)}{Dk} \text{ where}$$

m = the slope of the liquidus curve at the composition of interest

C_o = mole fraction CdSe

D = the diffusion coefficient

R = growth rate

k = segregation coefficient

RESULTS AND DISCUSSION

At the time of this writing, three crystals, designated A-2, A-4, and A-5 have been successfully grown. The growth parameters for each are listed in Table 1. Ingot A-2 has been almost fully characterized at this time and the results of the analysis will be the primary focus of this report. As seen in Table 1, ingot A-4 was grown at a faster growth rate than ingot A-2, but with a thinner barrier in an attempt to increase the thermal gradient present at the beginning of solidification to prevent constitutional supercooling. Ingot A-5 was grown with a relatively steep thermal gradient, this time with a thick barrier in an attempt to ensure that the steady state position of the growth interface was well within the barrier. The intent was to look at the shape of the solid-liquid interface resulting from these growth conditions in an attempt to obtain some information about the thermal properties of the liquid relative to the solid.

Three primary characterization techniques were utilized in this study. Optical microscopy, precision density measurements, and infrared transmission-edge mapping were used to obtain information about the microstructure and the compositional uniformity of the alloys. These results will be presented and discussed in the following sections.

TABLE 1
GROWTH PARAMETERS FOR
SAMPLES SUCCESSFULLY GROWN TO DATE

INGOT A-2

$Hg_{1-x}Cd_xSe$

x = 0.2

L = 16.7 cm

GROWTH RATE (R) = 0.3 μM/S

THERMAL GRADIENT (G) = 97° C/CM

K = 2.25

BARRIER: 2.4 cm

T_U = 960°C

T_L = 486°C

INGOT A-4

$Hg_{1-x}Cd_xSe$

x = 0.2

L = 16.8 cm

R = 0.4 μM/S

G = 187°C/CM

BARRIER: 0.64 cm

ANNEALED AT 250°C FOR 2 DAYS

T_U = 1015°C

T_L = 417°C

INGOT A-5

$Hg_{1-x}Cd_xSe$

x = 0.2

L = 16.4 cm

R = 0.3 μM/S

G = 90°C/CM

BARRIER: 2.4 cm

T_U = 1000°C

T_L = 670°C

MICROSCOPY

All samples were ground and polished using standard metallographic procedures down to a polish on 0.25um diamond. They were etched for 15 seconds in Straughn's reageant at 60°C and then rinsed in a dilute solution of bromine in methanol, followed by rinses in methanol.

The microstructure of a sample taken at 3.95cm from the tip of ingot A-2 is shown in Figure 8. As can be seen, a fine cell structure is evident. This indicates an interface instability and the possibility that the initial thermal gradient was not steep enough and as a result, a slight amount of constitutional supercooling occurred.

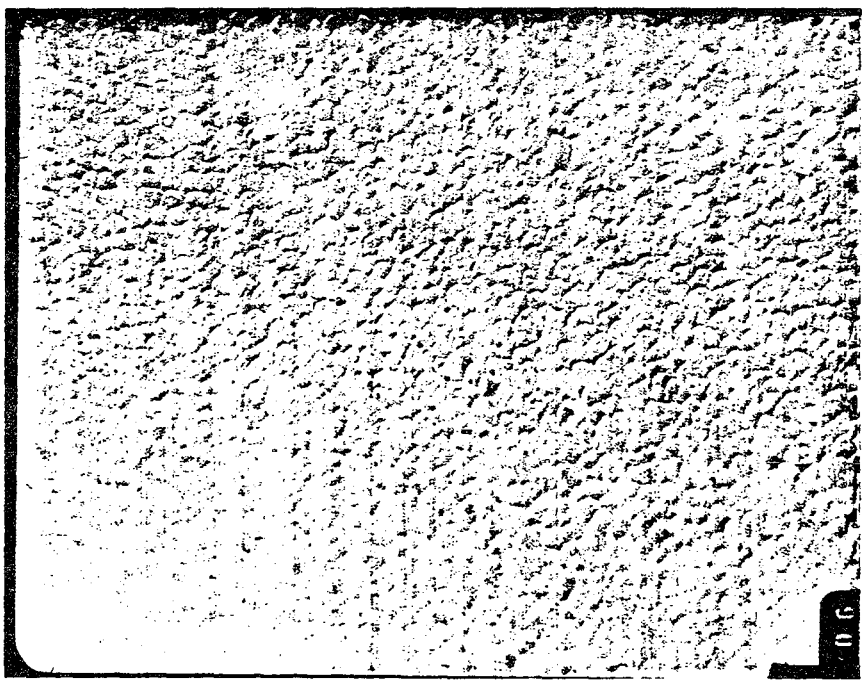


Figure 8. Optical micrograph of a sample taken 3.95cm from the tip of ingot A-2 (240x)

However, in analyzing a slice taken at 14.91cm from the tip of A-2, no apparent microstructure was observed. This seemed to indicate that once steady state growth was achieved, constitutional supercooling was eliminated. This would be expected, since the thermal gradient at the solid-liquid steady state interface was steeper than that at the liquidus. No grain boundaries were observed in the slice examined, however, x-ray analysis is needed to confirm the presence of a single crystal.

PRECISION DENSITY MEASUREMENTS

Iwanowski[17], in 1975, showed that the mass density of $Hg_{1-x}Cd_xSe$ decreases linearly with increasing x value, through both the zincblende and wurtzite structures. Therefore, precision density measurements can be used to accurately determine alloy composition. Precision density measurements were madein accordance with theprocedure described by Bowman and Schooner[18]. Table 2 summarizes the results of the density measurements for slices taken at various distances along the length of the ingot A-2. Figure 9 graphically shows the relationship between the measured axial density and the distance from the ampoule tip. As can be seen, the overall trend appears to be well modeled by the one-dimensional diffusion case, that is, an initial transient is observed, followed by a steady state growth region, ending up with a final transient.

At a distance of 6.41cm from the ampoule tip, it is seen that the average composition of the solid increased. This rise in composition is well correlated with an actual stopping of the furnace translation due to a motor failure.

Table 2

Composition As Determined From Precision Density Measurements As A Function Of Axial Position For Ingot A-2

Distance From Tip (cm)	Composition
1.00	0.289
1.55	0.247
2.00	0.228
3.05	0.204
3.55	0.213
3.95	0.202
5.55	0.213
6.41	0.245
8.52	0.206
9.25	0.204
9.95	0.182
10.98	0.182
11.91	0.204
12.91	0.182
13.71	0.175
14.91	0.165
15.15	0.155
15.35	0.138
15.50	0.103
15.75	0.054
15.90	0.022

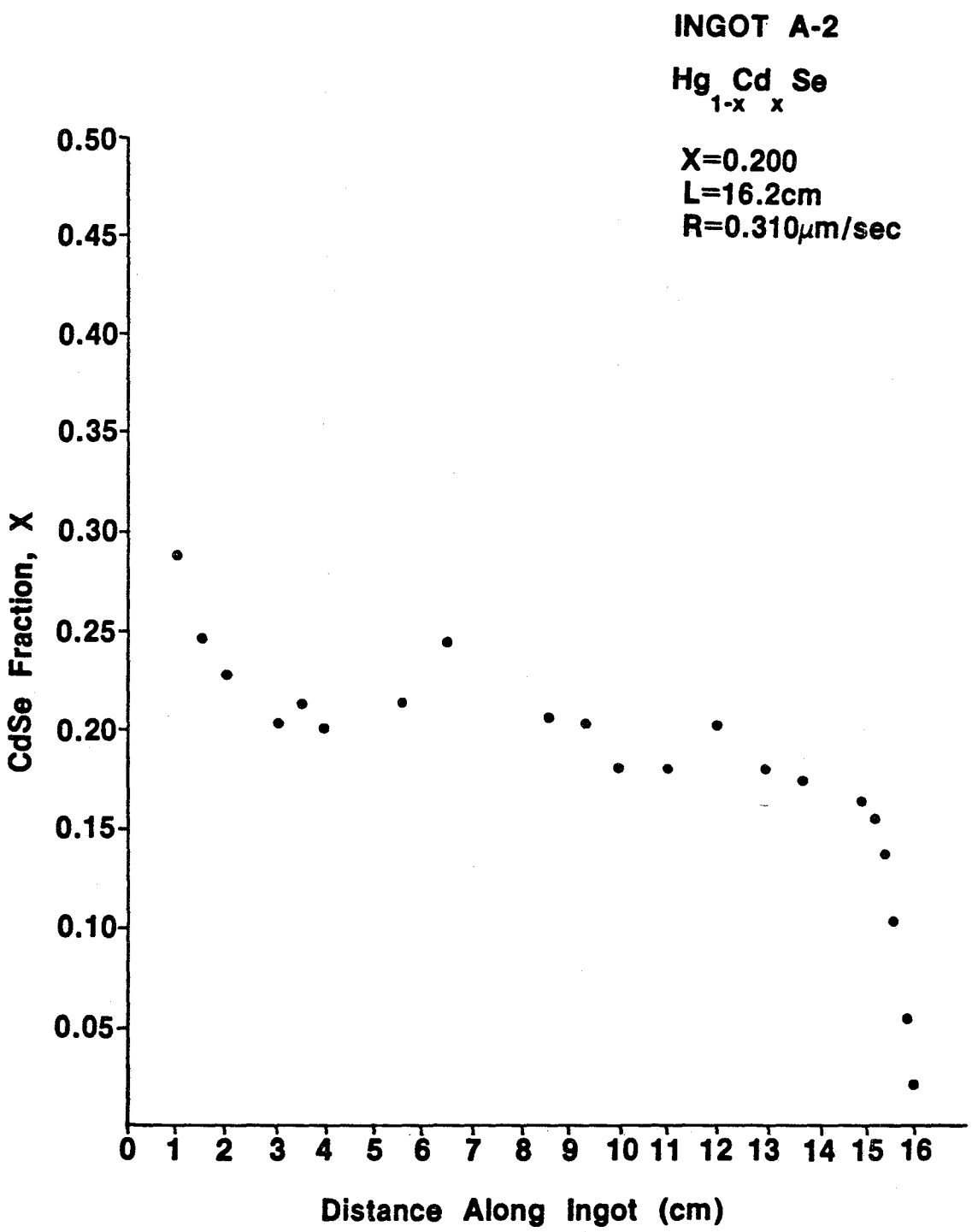


Figure 9. Measured compositional profile for ingot A-2 (From precision density measurements)

It is apparent that the composition along the ingot was approaching a steady state value when the translation stopped. However, as the alloy remained stationary with respect to the furnace, the system began to return to an equilibrium state, and thus, upon resuming translation, another transient is observed, followed by a subsequent movement into a steady state region.

Further analysis is currently underway on material sectioned from this interrupted region and is showing some interesting results. The results have prompted additional experiments which could provide some very exciting insights into the interface stability and diffusion in this system.

IR TRANSMISSION MAPPING

Infrared transmission mapping is a technique which can be used to determine the radial compositional uniformity of thin slices taken along the length of a crystal. It also is possible to utilize this technique to get an average x value for the slice. However, the outer most edges of the sample, which in this experiment are lowest in x value, are not included in the IR analysis. As a result, the average x value obtained by this method will be slightly higher than the x value obtained by precision density measurements, however density trends should be fairly similar.

Slices of approximately 1mm thickness were cut along the length of ingot A-2, ground to about 0.75mm thickness and etched with a methanol-5% bromine solution. All transmission measurements were made at room temperature at regularly spaced intervals on each slice. To improve spatial resolution, an aperture 100 um in diameter was used.

Typical results for a spectra obtained from a slice 2.00cm from the tip of ingot A-2 is shown in Figure 10. Pairs of curves correspond to symmetrical points on the sample, with the outermost curve representing the center of the slice. As can be seen in this plot of transmission vs wavenumber, a wavenumber corresponding to the cut-on where the crystal goes from being transparent to opaque can be obtained, and this wavenumber which corresponds to a particular gap energy can be used to obtain an x value, since, as discussed earlier, the energy band gap of a $Hg_{1-x}Cd_xSe$ crystal can be diretly related to the alloy composition. This relationship is also temperature dependent and can best be expressed by the following equation:[11]

$$Eg = -0.209(1-7.172x-2.17x^2)$$
$$+7.37 \times 10^{-4}(1-1.277 \times -0.151x^2)T$$
$$+2.00 \times 10^{-9}(1+23.45x-599.4x^2)T^2$$

Typical results of radial compositional profiles are shown in Figures 11 and 12. As can be seen here, the ingot, as grown, is cadmium rich in the center and cadmium deficient on the edges. This finding was quite surprising since it is directly opposite to that commonly observed in the $Hg_{1-x}Cd_xTe$ system. Such variations suggest a convex interface for the entire growth length. Such an interface shape can have definite implications as to the thermophysical properties of this system and this area is currently being investigated further as a continuation of this project.

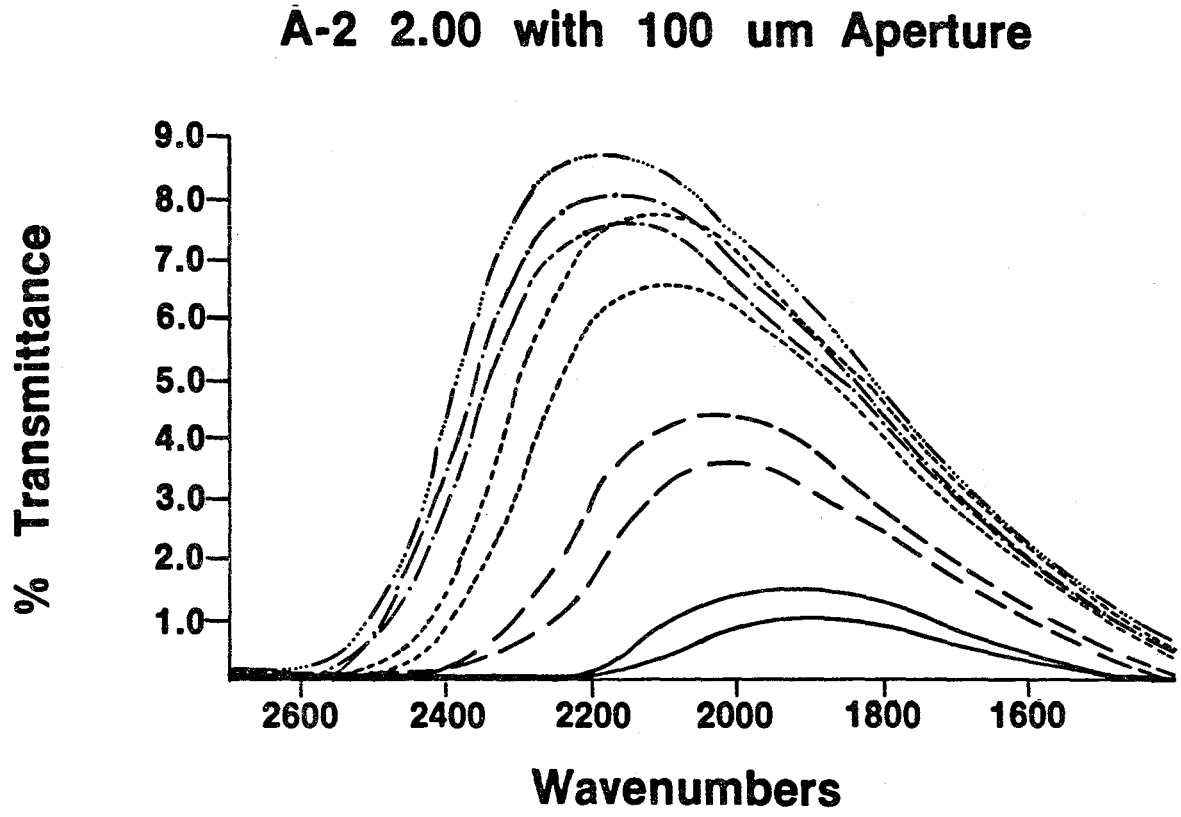


Figure 10. Infrared transmission spectra taken on a sample 2.00cm from the tip of ingot A-2

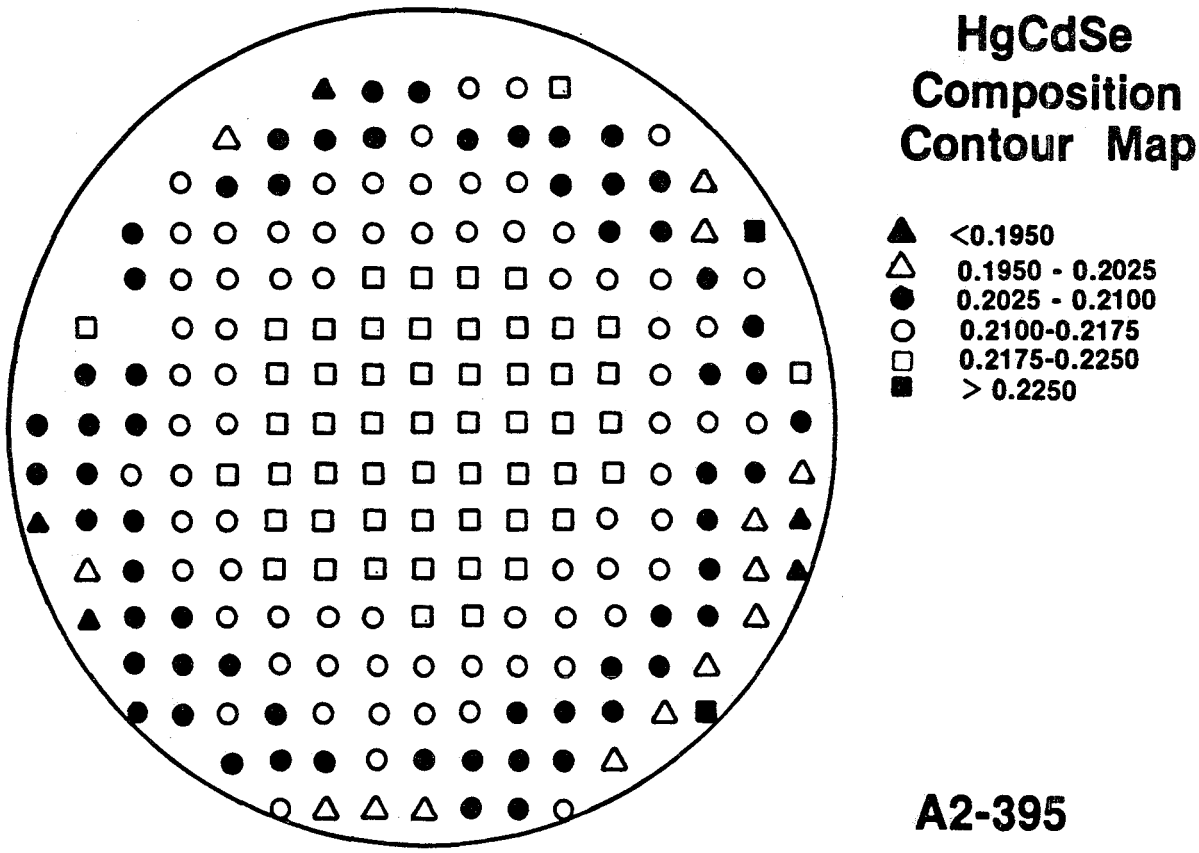


Figure 11. Compositional map obtained from infrared transmission-edge mapping of a sample taken 3.95cm from the tip of ingot A-2

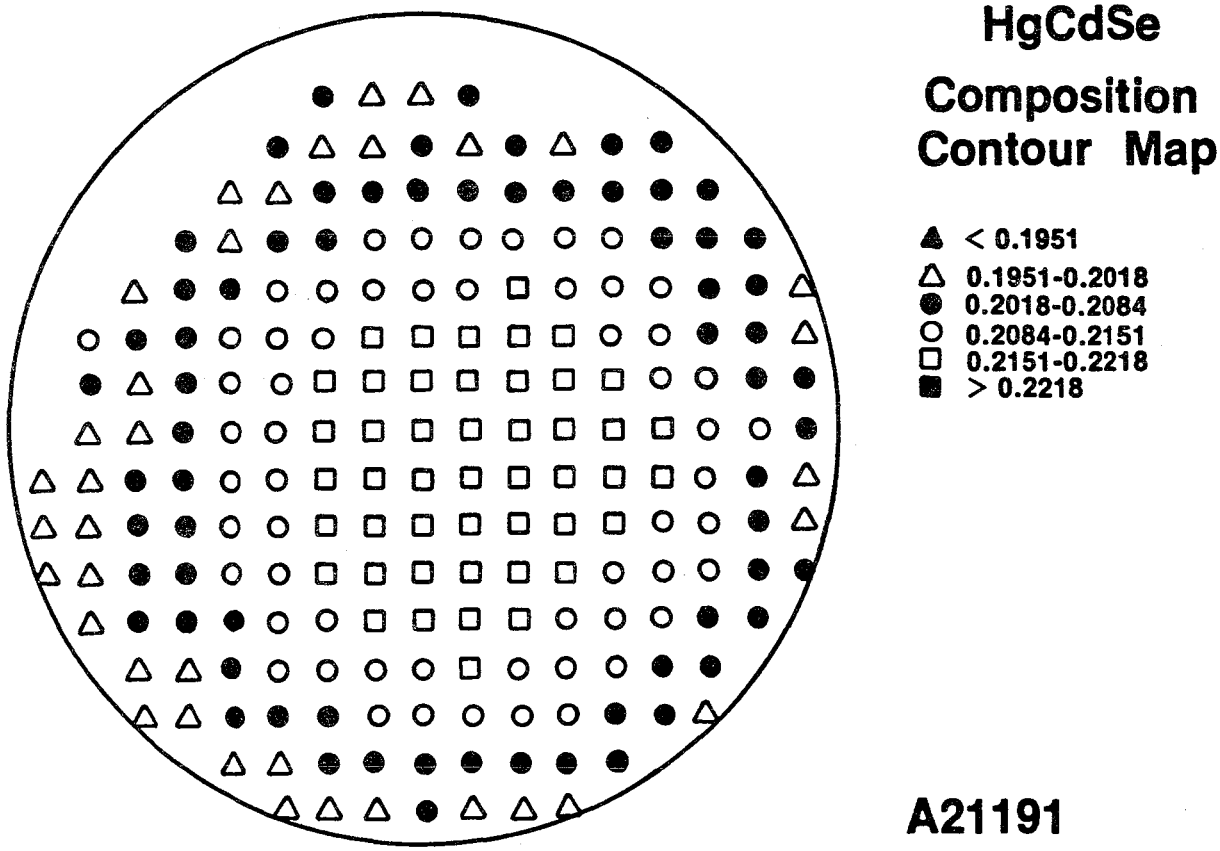


Figure 12. Compositional map obtained from infrared transmission-edge mapping of a sample taken 11.91cm from the tip of ingot A-2

Table 3 is a tabulation of the average x value obtained by IR mapping for each slice taken from ingot A-2. A plot of this average composition (x) of each radial slice as a function of distance from the tip is shown in Figure 13. As can be seen from this figure, the axial compositional profile appears at a first approximation to be well modeled by the one-dimensional diffusion model. The length of the steady state region of growth also indicates that the diffusion coefficient, D, in this system may in fact be lower than that in the HgCdTe system. This is also a very interesting finding, since it was initially assumed that they would be approximately the same.

TABLE 3

AVERAGE COMPOSITION AS DETERMINED FROM IR TRANSMISSION MAPPING AS A FUNCTION OF AXIAL POSITION FOR INGOT A-2

DISTANCE FROM TIP	COMPOSITION
1.55	.235
2.00	.218
3.05	.211
3.55	.209
3.95	.212
4.51	.207
5.55	.203
6.41	.236
8.52	.199
9.25	.204
9.95	.206
10.98	.206
11.91	.209
12.91	.206
13.71	.202
14.91	.191

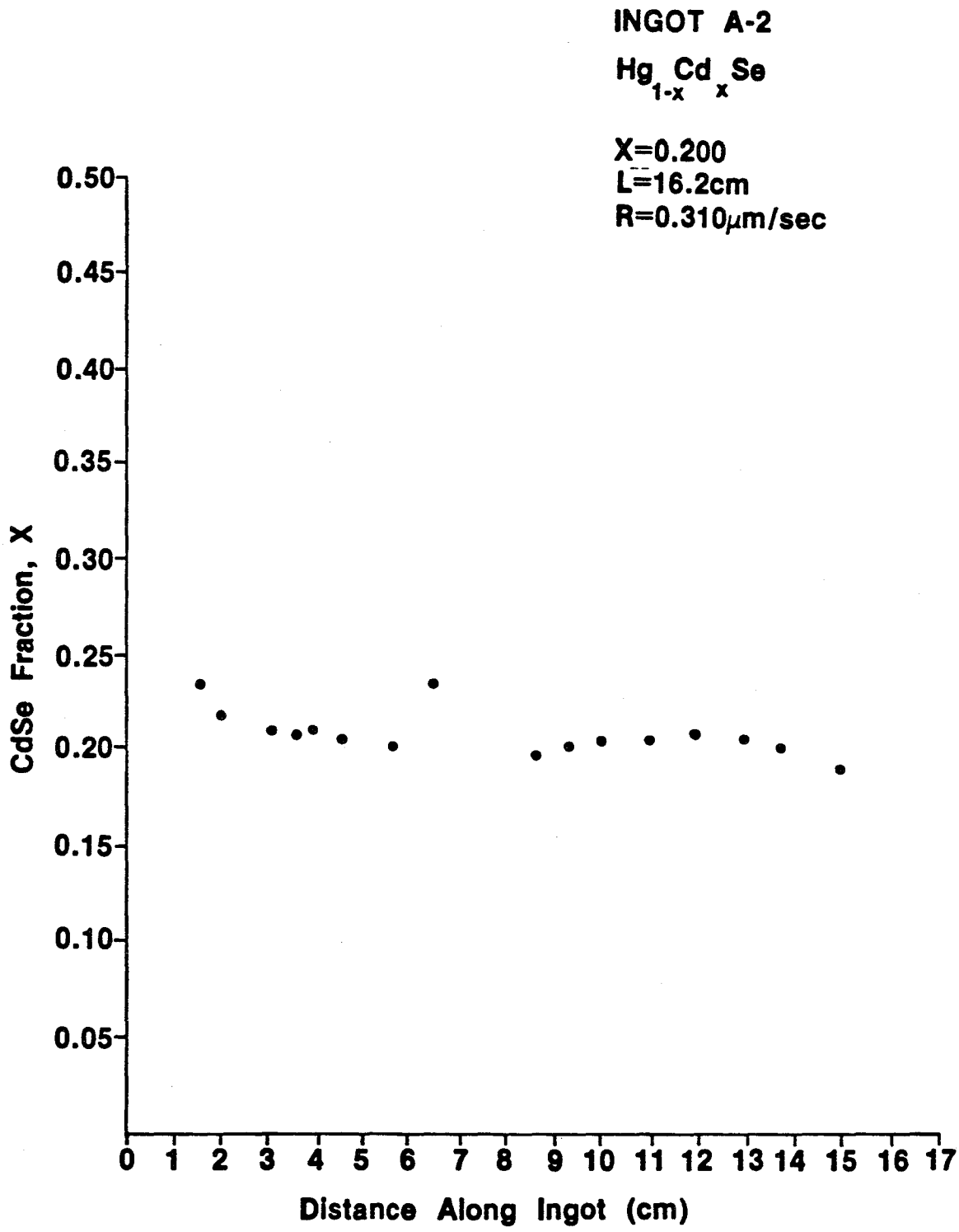


Figure 13. Measured compositional profile for ingot A-2 (from IR transmission edge-mapping data)

TENTATIVE CONCLUSION

At this time it is difficult to make definite conclusions. However, based upon the preliminary results of this investigation, the following tentative conclusions can be made:

1. The results of the longitudinal composition profile on a Bridgman grown crystal of $Hg_{1-x}Cd_xSe$ of composition x = 0.2, grown at 0.3um/sec,(ingot A-2) appears to be well modeled by the one-dimensional diffusion model. This suggests that the results of Summers and Nelson [11] were influenced by constitutional super cooling.
2. The radial composition profile of ingot A-2 consists of a cadmium rich center and hence implies a convex growth interface. This is opposite to that of the $Hg_{1-x}Cd_xTe$ system and this in turn implies differences in the thermophysical properties of the two systems.
3. Due to the length of the steady state growth region in the axial composition profile, the diffusion coefficient in the liquid in the HgCdSe system appears to be less than in the HgCdTe system.

Further experimentation is currently underway and planned. Evaluation of the results will be necessary to confirm or modify the above conclusions. Also, crystal growth experiments are planned which will be used to evaluate the influence of the growth rate on compositional uniformity.

As mentioned previously, ingot A-4, grown at 0.4um/sec, is currently being evaluated, as well as ingot A-5, grown at 0.3um/sec utilizing different thermal conditions in the growth furnace. These results will provide a better understanding between thermal conditions and interface shape.

REFERENCES

1. R. K. Willardson, A. C. Beer, Ed., Semiconductors and Semimetals, Vol 18., (Academic Press, New York, 1981).
2. J. L. Schmit, J. of Crystal Growth 65 (1983) 249
3. F. R. Szofran, D. Chandra, J. C. Wang, E. K. Cothran, and S. L. Lehoczky, J. of Crystal Growth 70 (1984) 343
4. B. E. Bartlett, P. Capper, J. F. Harris, and MJ. T. Quelch, J. of Crystal Growth, 46 (1979) 623.
5. C. L. Jones, P. Capper, J. J. Gosney, and I. Kenworthy, J. of Crystal Growth, 69 (1984) 281.
6. F. R. Szofran and S. L. Lehoczky, J. of Crystal Growth, 70 (1984) 349.
7. C. E. Chang, and W. R. Wilcox, J. of Crystal Growth, 21, (1974) 135
8. C. E. Huang, D. Elwell, and R. S. Feigelson, J. Crysstal Growth 64 (1983) 441
9. R. J. Naumann and S. L. Lehoczky, J. Crystal Growth 61 (1983) 707
10. S. L. Lehoczky and F. R. Szofran, in: Materials Processing in the Reduced Gravity Environment of Space, Vol. 9, Ed. G. E. Rindone (North-Holland, New York, 1982) p. 409
11. C. R. Whitsett, J. G. Broerman, and C. J. Summers, in: Semiconductors and Semimetals, Vol. 16, Ed. R. K. Willardson and A. C. Beer, (Academic Press, New York, 1981)
12. E. Cruceana and D. Niculescu, Compt. Rend. 261 (1965) 931
13. A. Kalb and V. Leute, Phys. Stat. Sol. (a), 5 (1971) K199

14. D. A. Nelson, C. J. Summers, and C. R. Whitsett, J. Electron. Mater. 6 (1977) 507

15. D. A. Nelson, J. G. Broerman, C. J. Summers, and C. R. Whitsett, Physical Review B, 4 Vol 18, (1987) 1658.

16. C. R. Whitsett, J. G. Broerman, D. A. Nelson, and C. J. Summers, Final Report MDRL Contract N00014-74-C-0318 (ONRC) (1976)

17. R. J. Iwanowski, Acta Physica Polonica, Vol. A 47, 5, (1975)

18. H. A. Bowman and R. M. Schooner, J. Res. Natl. Bur Std., 71C(1967)179

1985

NASA/ASEE Summer Faculty Research Fellowship Program

Marshall Space Flight Center

The University of Alabama

Space Station Common Module Thermal Management: Design and Construction of a Test Bed

Prepared By:	Ronald G. Barile, Ph.D.
Academic Rank:	Professor and Program Chairman
University and Department:	Florida Institute of Technology Department of Chemical Engineering
NASA/MSFC: Division: Branch:	Structures and Propulsion Lab Engineering Analysis Thermal Engineering
MSFC Counterpart:	James W. Owen
Date:	August 23, 1985
Contract No.:	NGT-01-008-021 The University of Alabama in Huntsville

ABSTRACT

The space station habitats, laboratories, and logistics modules will be uniformly designed as a large cylinder (approximately 15 x 35 feet) with a pressure vessel wall and high-efficiency thermal insulation. The interior and exterior environment will be such that a net quantity of heat must be rejected to space to prevent temperature buildup beyond a habitable range. Interior electrical equipment, refrigeration, pumps, experiments, and the workers' metabolism give rise to thermal energy loads, as well as the incoming solar energy.

In this project, a thermal test bed was designed, simulated, and planned for construction. The thermal system features interior and exterior thermal loads and interfacing with the central-radiator thermal bus.

Components of the test bed include: (a) Body mounted radiator loop with interface heat exchangers (6000 Btu/hr), (b) Internal loop with cabin air-conditioning and cold plates (3400 Btu/hr), (c) Interface heat exchangers to the central bus (13,000 Btu/hr), (d) Provisions for new technology including advanced radiators, thermal storage, and refrigeration. The apparatus will be mounted in a chamber, heated with lamps, and tested in a vacuum chamber with LN2-cooled walls (Sunspot I).

Simulation of the test bed was accomplished using a DEC PRO 350 computer and the software package TK!Solver. Key input variables were absorbed solar radiation and cold plate loads. The results indicate temperatures on the two loops will be nominal when the radiation and cold plate loads are in the range of 25% to 75% of peak loads. If all loads fall to zero, except the cabin air system which was fixed, the radiator fluid will drop below -100 F and may cause excessive pressure drop. If all loads reach 100%, the cabin air temperature could rise to 96 F. The mismatch between heat loads and heat removal capability is likely to be desirable when new technology is tested.

ACKNOWLEDGEMENTS

Because of my lifelong interest in rockets, space travel, and science fiction, the opportunity to work at Marshall Space Flight Center has been an overwhelming fulfillment of old dreams. The program hosts, Gerald R. Carr and Dina Conrad of UAH, and Leroy Osborn and Dr. Jim Dozier of NASA, by their kindness, patience, and consideration have made these ten weeks rewarding and enjoyable. Our tours and seminars, conducted by Gerald, Leroy, and Jim, were especially appreciated.

Day-to-day work was guided by Jim Owen, Dick Wegrich, Joe Sims, Bob Fisher, and Rick Bachtel. To them I offer heartfelt thanks. Friendly and informative discussions were had with Bill Till, Tim Page, and Farouk Huneidi. Laboratory Director Alex McCool and his staff, Paul Frederick, Bob Richmond, Randy Humphries, Harold Smyly, and Doug Lamb were helpful and kind beyond expectation. Finally, special thanks is due to Betty Golden, Jeanette Hall, and Bernice Bowling for helping with all the "little things" which kept my work running smoothly each day, and to Brenda Wade and Sue Madole who typed this report.

INTRODUCTION

Pressurized compartments in the space station, where astronauts live and work, will be subjected to various thermal loads. Internally, heat is dissipated to the atmosphere from all electrical and mechanical equipment, lights, experiments, and even human metabolism. Externally, the Sun and Earth will radiate energy to the station, while the outer walls will radiate to deep space. The external heat sources and sinks will vary on orbit and seasonally, as well as differences between individual compartments due to shading and view factor variability. Life support tolerances are such that an active thermal management system will be designed into each habitated unit or common module. Heat from the physical plant and various experiments will be transported by fluid (or liquid-vapor) loops to body-mounted radiators and to a central thermal bus. The central bus is an external loop which will transport heat to the central radiators where heat is rejected to space.

As phase B space station work is now commencing, there is a need for NASA in-house thermal management test bed facilities. The present work was initiated by J.W. Owen in 1984 to design and construct a common module test bed at Marshall Space Flight Center. The objective was to provide an early in-house test capability for thermal systems which support the MSFC work package. His approach was to integrate existing hardware, e.g., from skylab spare parts, into a test unit which simulates the essential functions of the common module thermal transport and heat rejection loops. Then, the system would be updated with new or advanced technology elements as they become available. The latter include body-mounted radiators, thermal storage, refrigeration, thermal bus, long-life fluid systems, heat transport across structural boundaries, advanced heat pipes, self-maintaining thermal surfaces, electrochromic panels, conjugating binary systems, microencapsulated phase-change material, metal hydrides, and conformal heat exchangers.

The test bed elements will be situated outside the Sunspot I thermal vacuum chamber. These elements include cabin air heat exchangers, cold plates, pumps, controls, data acquisition, and a central-bus interface heat exchanger, simulated by a facility water loop. Equipment and other interior loads will be simulated by electrical heaters. Inside the chamber, there will be two body-mounted radiators, radiator bypass and mixer control valve, and an interior-exterior interface heat exchanger. The radiators will be

exposed to high vacuum, liquid nitrogen-cooled walls, and heater lamps. Plumbing provisions will be made for both interior and exterior advanced-technology elements. Specific plans, equipment details, and test procedures are summarized in an internal NASA report (1).

OBJECTIVES

The ultimate objectives of this project are to prepare a test bed and evaluate advanced thermal systems technologies for the space station common module. The immediate objectives of the present work are:

1. Design a thermal test bed representing the essential elements of the common module.

2. Specify and select equipment for the test bed.

3. Simulate the performance under various loads with calculations.

4. Prepare for construction and testing by making arrangements with supporting groups.

CONSTRUCTION AND SIMULATION

A. CONSTRUCTION AND TEST PLAN

Much of the equipment for the test bed is already in storage at MSFC. Leftover from ATM and skylab are pumps, valves, sensors, filters, and heat exchangers (1). Design specifications of these elements are available (2). Two radiators, double sided, 101.4 ft total area and several cold plates are also available. These were constructed by MDAC and TRW in earlier contracts. The general configuration of these elements is shown in figure 1. Infrared lamps are to be situated around the radiators to simulate variable orbital conditions. The lamps are to provide 0-17.3 w/sq ft controllable, evenly distributed heating. This is equivalent to 0-6000 Btu/hr insulation for the specified radiators. Inside the radiators a Coolanol 25 fluid will be used (Monsanto). The other loops will contain water. Interface heat transfer from the internal water to the thermal bus will be simulated by a facility water loop. General specifications and parameters for the test bed elements are shown in table I.

Requests for removing equipment from storage to building 4619 have been initiated. Elements will be selected and prepared for shipment to Lockheed, Huntsville. Miscellaneous items will be procured by Lockheed, then they will construct the apparatus and perform preliminary tests. Then, the test bed will be shipped to building 4619 to prepare for tests in the Sunspot facility when it is available.

A description of the test facility requirements and of the tests is found in reference (1). As indicated in figure 1, the radiators and radiator-to-water interface (HX1) will be situated inside the vacuum chamber. Cold plate loads and lamp power will be varied to simulate a range of internal and orbital conditions. As new technology concepts and hardware become available, they will be incorporated into the tests.

B. TEST BED FLOW SHEET

Nominal conditions, fluid flows, heat loads, and heat flows are shown in figure 2. Coolanol 25 is recommended for the radiator fluid. It has kinematic viscosity of 12 Centistokes and density of 0.91 g/cc at 0 F. Half-inch ID tubing is adequate for this service. However, at -50 F, it

TABLE I

Equipment Specifications

Radiators: Two units, 11 ft by 2 2/3 ft and 5 ft by 2 2/3 ft, double sided. Total area = 101.4 sq ft. Approximate heat rejections up to 7000 Btu/hr. Coolanol 25 fluid.

Heat Exchanger 1: HX1 consists of two units in series, originally designed for ground service to Airlock, ATM and suit cooling. Original rating was 17,700 Btu/hr each, at 183 lb/hr each (UA = 258 Btu/hr. - sq ft - F each). Present application: 1,700 Btu/hr at 183 lb/hr for each of two units. Part no. 52-83700-1202, ref. (2), pg. 25.0.

Heat Exchanger 2: HX2 consists of two cabin air heat exchangers in parallel water flow. Original rating was 680 Btu/hr each at 88 CFM each, 5 psia oxygen (UA = 48.1 Btu/hr - sq ft - F each). Present application: 1025 Btu/hr each. Part no.: 52-83700-1227, ref. (2), pg. 32.0.

Heat Exchanger 3: HX3 consists of 5 units connected in series. The units are the same as HX1 units above. The present application is 13,000 Btu/hr transfer from loop water (520 lb/hr) to facility water (550 lb/hr, 65 F). Facility water represents the central bus.

Cold Plates: Cold plate 1 is nominally 0.4 kw. Cold plates 2 through 5 are nominally 0.95 kw each, or a total of 3.8 kw. Several devices are potentially available, e.g., see reference (3) for a candidate. The cold plates must be fitted with electrical heaters which simulate equipment heat loads.

Pump Systems: This pump system provides pump, accumulator, low fluid and power indicators, and a fill port. The pumps are rated for "coolant" at 183 lb/hr with 175 psi pressure rise. Part nos. 52-83700-831, 833, and 869, ref. (2), pp. 19.0-21.3.

Radiator Mixer Valve: Details of the valve selection were included in a separate communication (5). Two options were considered: (a) Proportional-integral control using a proportional valve driven by a dc gear wound motor, and (b) stepper control using 5 parallel tubes fit with orifices of different openings and 4 on-off solenoid valves to select flow paths.

will thicken to 200 Cs and 0.95 g/cc which may make it unpumpable. The original fluid on skylab, Coolanol 15, is less viscous but it is no longer made. Water flow at 520 lb/hr and 70 F is handled by 1/4 or 3/8 inch ID tubing (6.8 and 3.0 ft/s, respectively).

The lamps surrounding the radiators are designed for 0-17.3 w/sq ft, which is 0-6,000 Btu/hr total incident radiation on the double surfaces. The lamps have a narrow profile and will not greatly affect outward radiation. An estimate of the total radiator rejection power is 2 kw or 7,000 Btu/hr. Part of the Coolanol will bypass the radiator, a fraction f, and mix with the cooled fluid as shown in figure 1. The mixed outlet design temperature is 40 F. This will be accomplished by a controlled mixer valve to be discussed below. The rejection heat load on the radiator is that absorbed from the lamps and the heat picked up in heat exchanger 1 (HX1). The cabin air cooling load is roughly 2050 Btu/hr. This plus cold plate (CP-1) load totals 3420 Btu/hr, the nominal load for HX1 which is transferred to and rejected by the radiators. Cold plate 1 has nominally 0.4 kw load, while cold plates 2-5 have 0.95 kw each; i.e., the total load for cold plates 2-5 is 3.8 kw or 13,000 Btu/hr. The heat taken up in cold plates 2-5 is primarily rejected in HX3 to facility water. This heat exchanger represents the space station central bus interface.

An extra heat exchanger, HX on figure 1, is tentatively planned for the situation where heat loads are greater than the radiator rejection capability. Up to 3500 Btu/hr may be rejected in HX, in which coolant service at 20 F is required. This heat exchanger can also function as a ground service exchanger.

C. SIMULATION OF TEST BED PERFORMANCE

Equations describing heat transfer in the bed are given in table II. The 20 equations are listed in order of solution by TK!Solver:

Heat exchanger 2, Eqns. HX2 1-4

Radiator and control valve, Eqns. RAD 1-5

Heat exchanger 1, Eqns. HX1 1-4

Heat exchanger 3, Eqns. HX3 1-3

Cold plate 1, Eqns. CP1 1

TABLE II
Standard Model

	S	Rule
HX2	1	q2=114*(tw2-tw1)
	2	q2=fa*2*pa*0.6489*60*(ta1-ta2)/(ta1+460)
	3	q2=2*48.1*dt2
	4	dt2=((ta1-tw2)-(ta2-tw1))/ln((ta1-tw2)/(ta2-tw1))
RAD	1	qrn=wc*cpc*(tc2-tc3)
	2	qrr=qrn+qra
	3	tr=((qrr/ar/3.413-17.55)/0.17+35)
	4	tc1=2*tr-tc2
	5	f=(tc3-tc1)/(tc2-tc1)
HX1	1	q1=wc*cpc*(tc2-tc3)
	2	dt1=q1/516
	3	tw4=q1/114+tw1
	4	dt1=((tw4-tc2)-(tw1-tc3))/ln((tw4-tc2)/(tw1-tc3))
HX3	1	q3=ww5*cpw*(tw6-tw4)
	2	tf2=q3/wf+tf1
	3	q3=1290*((tw6-tf2)-(tw4-tf1))/ln((tw6-tf2)/(tw4-tf1))
CP1	1	tw3=qcp1/ww1+tw2
TEE	1	tw5=(ww1*tw3+ww2*tw4)/ww5
CP2-5	1	qcp=qcp2+qcp3+qcp4+qcp5
	2	qcp=ww5*cpw*(tw6-tw5)

Tee, Eqn. TEE 1

Cold plates 2-5, Eqns. CP2-5 1-2

The calculation scheme is to march through the equations in order and solve for one unknown in each equation. Since the equations are highly coupled, initial guesses of certain output variables are required. Also, input variables are required so that the problem is uniquely specified, i.e., total variables (39) minus input variables (19) equals equations (rules). A typical output list from TK!Solver is shown in table III. This case is for radiator and cold plate loads at 50% level, including:

qra	radiation absorbed by radiator
qcp 1	heat absorbed from cold plate 1
qcp 2, 3, 4, 5	heat absorbed from cold plate 2, 3, etc.

These variables are marked as "INPUT" on the computer screen and on the output list ("Comment Column"). Similarly marked is the variable tc3 which is the temperature set point for the control valve outlet downstream from the radiator, normally 40 F. Variables which are marked "Guess" require a G in the "St" column and a new guess in the "Input" column before each run. The "Comment" and "Unit" columns contain the nomenclature list. (HX1, 2, and 3 represent the heat transferred in the respective heat exchangers.)

Several types of relationships are represented in table II:

1. qi = w x cp x (tj - tk)

   Stream heat balance: Eqns. HX2-1,2, HX1-1,

   HX1-3, HX3-1,2, CP1-1, CP2-5-2

2. qi = U x A x dti

   (Where dti is log mean temperature difference)

   Heat exchanger design equation: Eqns. HX2-3,

   HX1-2, HX3-3

3. dti = ((ta1-tw2)-(ta2-tw1))/ln((ta1-tw2)/(ta2-tw1))

TABLE III
Standard Model Results
50% Heat Load

St	Input	Name	Output	Unit	Comment
	2050	q2		Btu/hr	cabin air load HX2
		tw2	60.702381	F	water out HX2
		tw1	42.719925	F	GUESS water out HX1
	88	fa		cfm	cabin air (1 unit)
	16	pa		psia	air inlet to HX2
		ta1	78.298013	F	GUESS air inlet to HX1
		ta2	68.232985	F	air out HX2
		dt2	21.309771	F	log mean del t, HX2
		qrn	3050.3357	Btu/hr	net rad ht (+ = out)
	366	wc		lb/hr	Coolanol flow
	.45	cpc		Btu/lb-F	Coolanol ht cap
		tc2	58.520557	F	GUESS inlet Cool. temp
	40	tc3		F	INPUT contr valve out
		qrr	6050.3357	Btu/hr	ht rej by rad
	3000	qra		Btu/hr	INPUT ht abs by rad
		tr	34.603342	F	avg rad temp
	101.4	ar		sq ft	total rad area
		tc1	10.686128	F	out rad Cool. temp
		f	.61281953	none	fract of flow bypass rad
		q1	3050.3357	Btu/hr	HX1
		dt1	5.9115034	F	log mean del t, HX1
		tw4	69.477256	F	water out HX3
		q3	6184.6643	Btu/hr	HX3
	520	ww5		lb/hr	water flow
	1	cpw		Btu/lb-F	ht cap water
		tw6	81.370841	F	GUESS water in HX3
		tf2	76.244844	F	facility outlet temp
	550	wf		lb/hr	facility water flow
	65	tf1		F	facil inlet temp
		tw3	66.711153	F	water out cp1
	685	qcp1		Btu/hr	INPUT cold plate 1
	114	ww1		lb/hr	water flow to hx1
		tw5	68.870841	F	water flow to cp2-5
	406	ww2		lb/hr	water flow to cp2-5
		qcp	6500	Btu/hr	cp2-5 total load
	1625	qcp2		Btu/hr	INPUT cp4 load
	1625	qcp3		"	INPUT cp3 load
	1625	qcp4		"	INPUT cp4 load
	1625	qcp5		"	INPUT cp5 load

Log mean temperature difference: Eqns. HX2-4, and HX1-4

4. qrr = qrr + qra

   Radiation balance: Eqn. RAD-2

5. tr = ((qrr/(3.413/ar) -17.55)/0.17+35)

   Radiator model for heat rejection (based on calculations in ref. (4)): Eqn. RAD-3

6. tr = 2 x tr - tc2

   Average radiator temperature: Eqn. RAD-4

7. f = (tc3 - tc1)/(tc2 - tc1)

   Stream mixing point balance: Eqn. RAD-5

   and

   tw5 = (ww1 x tw3 + ww2 x tw4)/ww5

   Eqn. TEE-1

8. qcp = qcp2 + qcp3 + qcp4 + qcp5

   Cold plate heat input sum: Eqn. CP2-5-1

The radiator model (Eqn. RAD-3) was obtained by curve fitting the results of a detailed calculation in ref. (4). The heat rejected, qrr, is simply a function of tr, the average radiator fluid temperature (Eqn. RAD-3). Emissivity and efficiency are included in the constants. This model is not directly sensitive to changes in coolant flow rate. But, it will predict very cold outlet temperature when the coolant bypass rate is around 90%, i.e., when loads are small. A better control policy for small loads may be to increase the radiator coolant flow and apply strip heaters at the inlet.

Design equations (Eqns. HX2-3, HX1-2, and HX3-3) contain an overall heat transfer parameter, UA, Btu/hr-F. This value was computed from performance temperature data (2). Since the test bed conditions differ from the original services for these exchangers, there should be some variability in UA values. No attempt was made to correct for flow and tempera-

ture level. These parameters will be more accurately determined from tests. Values for UA's in table II are:

HX2 UA = 2 x 48.1 Btu/hr -F

HX1 UA = 2 x 258 Btu/hr -F

HX3 UA = 5 x 258 Btu/hr -F

The simulation scheme was to allow heat exchangers in HX1 and HX3, q1 and q3, to vary with the level of radiation and cold plate loads. However, the cabin air load was assumed fixed, q2 = 2050 Btu/hr. The model will solve for cabin air temperature, ta1, and air outlet temperature, ta2. In this way, the value ta1 indicates whether or not the ECL standard is being met. Facility water input to HX3 is assumed constant at 550 lb/hr, 65 F. This will reasonably approximate the test situation but will not simulate a two-phase central bus. The model will require slight modification of the HX3 equations in the latter case.

Once a model is run on the computer, there are three ways to check for consistency:

1. qrn = q1

2. $0 \leq f \leq 1$

3. q2 + qcp1 + qcp = q1 + q3

If any of these equations are violated, then the simulation is in error.

During high-load situations, the radiator will not dissipate enough heat to retain tc3 at 40 F. A signal for this condition is a value of f outside its range of 0 to 1. When this happens, the model and the variables must be changed slightly. The model Eqn. RAD-5 is replaced by:

tc3 = tc1

The variable change is:

Input 0.0 for f

As a result of these changes, all Coolanol will flow through the radiator and the outlet temperatures (tc1 = tc3) will be calculated and will be above the set point of 40 F.

Several case studies were run to encompass a range of conditions to be tested. Figures 3, 4, and 5 show the results of loop simulation with TK!Solver for radiation and cold plate loads of 0, 25, 50, 75, and 100%. Temperatures at various positions are listed, as well as the Q1, Q2, and Q3 (heat transferred in HX1, HX2, and HX3) values. Also, the calculated or specified value of f is given at the radiator bypass line. Loads of 75 and 100% were calculated as overloads with f specified as 0.0. Three closer-mesh runs, 55, 60, and 65% load, were run to converge on the load value which just drives f to zero (figure 5 shows 55% case, only). The effect of load on f is plotted in figure 6. The standard model is valid up to 63% load where the curve crosses f=0. Note on the bottom of figure 6, the 65% case causes f=-0.39 using the standard model. The negative result is physically impossible, indicating that the overload model must be used instead.

Both the standard and overload models were used to produce figure 7, which represents various loop temperatures vs. % load. TC1 and TC3 show radiator outlet and mixer valve outlets. These two converge at 63% load. Beyond this point, they physically must converge because the radiator bypass is closed. In the lower portion of figure 7 are plotted TW1, the lowest water temperature in the loop (inlet to cabin heat exchanger), TW6, the highest water loop temperature (outlet from cold plates 2-5), and TA1, the cabin air temperature inlet to HX2. TW1 and TW6 show no problem, but TA1 indicates an excessive cabin air temperature as loads increase beyond 63%.

Another problem, at low load, is indicated by the very low temperatures of TC1. The Coolanol fluid will become very viscous below -50 F and probably become a gel-like fluid. This problem might be avoided by operating the mixer valve to permit no more than 75% bypass, i.e., $f \leq 0.75$. A "lowload" model was constructed by simply setting f=0.75 (input) and removing 40 F as input for TC3 (i.e., remove 40 in table for TC3; and, input 0.75 for f in table III). This results in TC1 rising to -49 F at zero load and -27 F at 25% load, but water may freeze in parts of the loop and the cabin air outlet falls as low as 54 F (TA2). These difficulties could all be avoided by (a) heating the radiator inlet line with a strip heater when TC1 falls below about -25 F, or (b) providing another heat exchanger at the radiator inlet to use the high side of the central bus (facility inlet) to preheat the Coolanol.

The mixer valve control algorithm naturally results from loop simulation. A feasible policy is to use proportional-integral modes to control the valve position with the set point being TC3=40F. During overload, the set point cannot be reached, and the valve will be closed. At low loads, when TC1 drops below -25 F, turn on an electrical heater at a fixed power, e.g., 1 kw and position the valve at the maximum bypass setting of f=0.75. This will both prevent the Coolanol and the water loops from freezing, as well as keep the cabin temperature in a reasonable zone. Alternatively, the loop conditions probably should not be set below 25% radiation load since this is not a realistic orbital situation.

CONCLUSIONS AND RECOMMENDATIONS

Conclusions: The thermal test bed has been designed, planned, and preparations are in progress to construct and test it. Simulation results indicate that the loop is well behaved and existing skylab equipment should be adequate. The heater lamps and equipment simulators (attached to cold plates) are capable of overloading the loop on the hot side. Also, the radiator loop may freeze at very low lamp power.

Recommendations: Parameters used in the simulation should be substantiated in single unit tests, especially UA values for the heat exchangers, and the radiator performance. Measured parameters should be inserted in the loop simulation for more accurate prediction of overall performance. Details of mixer valve design, reported separately, present two alternatives, with the simplest approach being a proportional valve and P-I controller. Other test situations may be simulated later if desired. A transient model of the loop should be formulated and solved before final detailed design is completed.

REFERENCES

1. J.W. Owen, "Space Station Advanced Development Thermal Test Requirements Document", MSFC internal report, 1984.

2. J.F. Schuessler, "Airlock-Environmental Control System Component Data", SCD Group Tech. Note No. E453-81, McDonnell Douglas Corp., Revision N, January 4, 1974.

3. G.L. Brown, "Test Plan - Cold Plate ORU - Thermal Vacuum Test", TRW Report No. L015.16-83-02, February 18, 1983.

4. J.R. Schuster et al., "Space Station Thermal Integration Study," Final Report, NAS8-35624, Apr., 1985, pgs. 2-10 to 2-12.

5. Private communication to J.W. Owen, EP44, Marshall Space Flight Center, AL

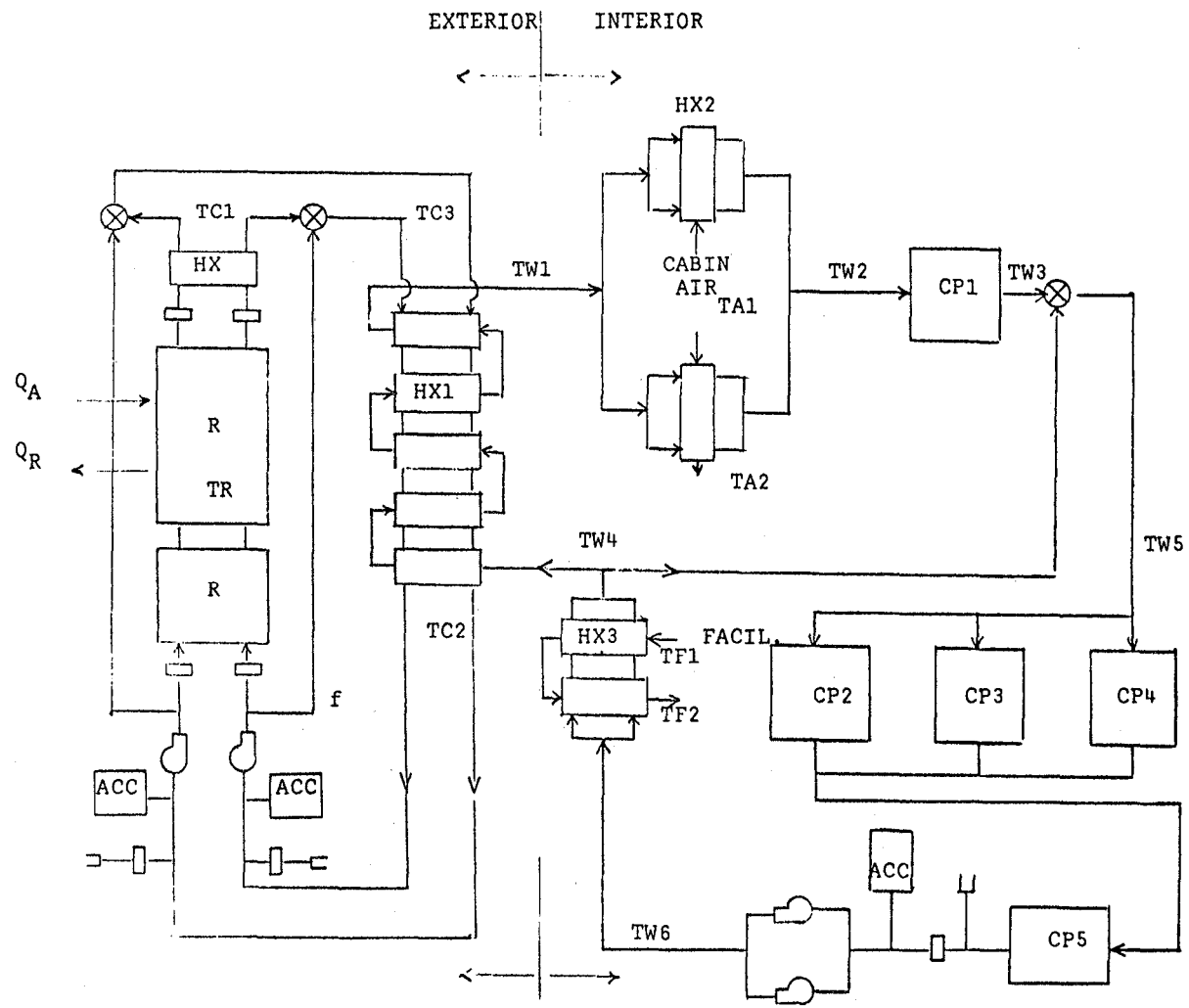


Figure 1. Test bed loop. Symbols correspond to Standard Model, table II.

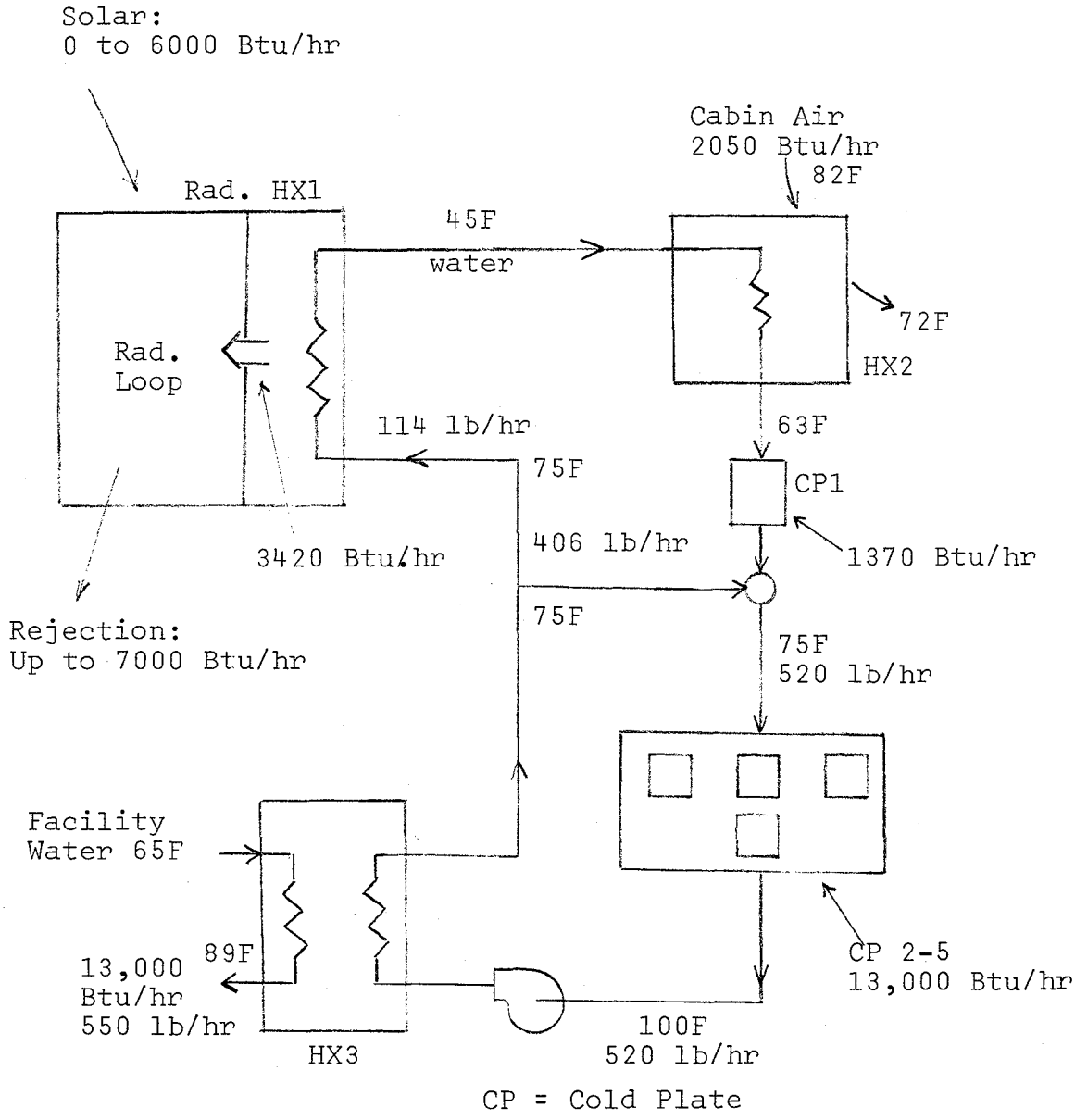


Figure 2. Example of conditions and heat flows in test bed.

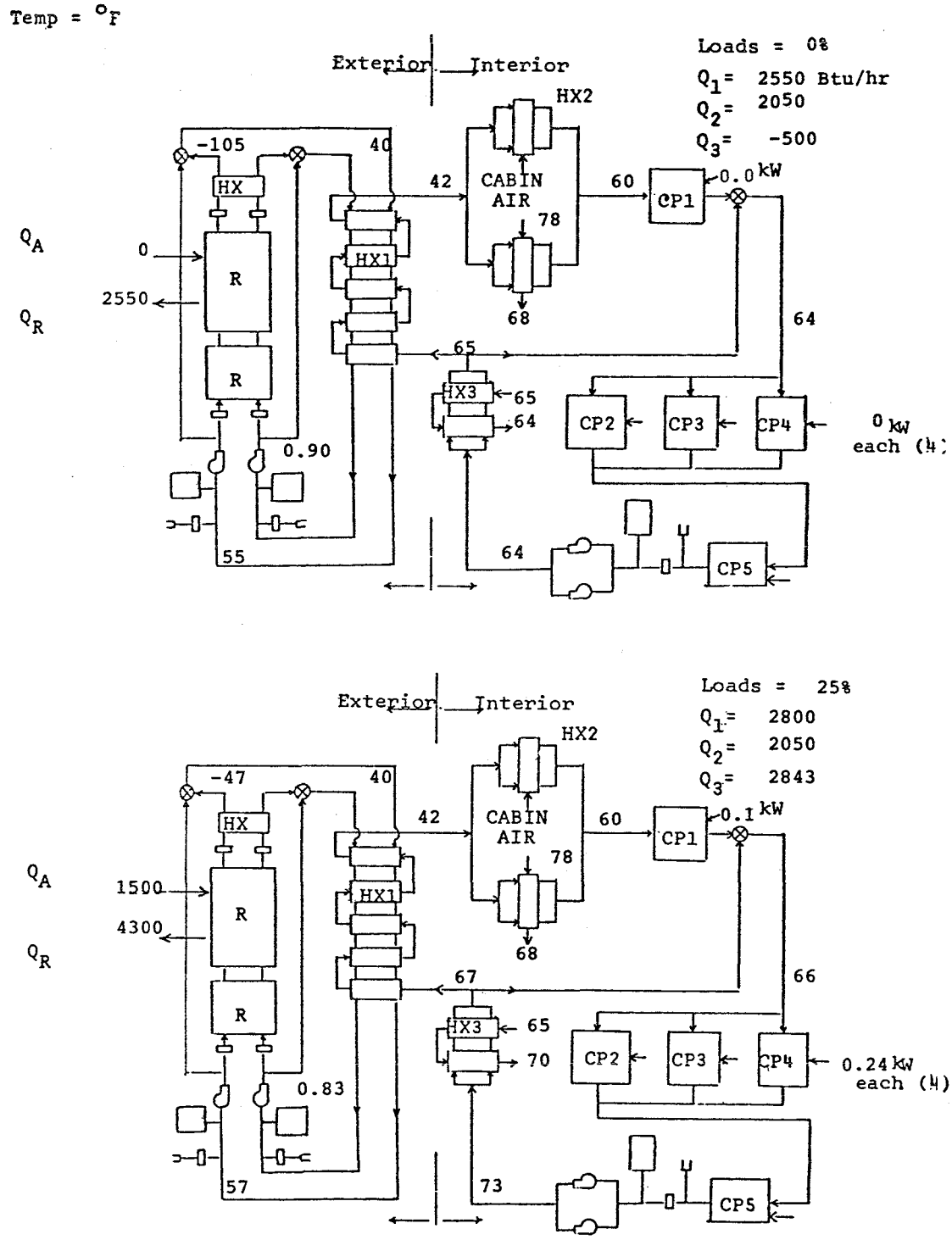


Figure 3. Results of loop simulation. Upper: 0% load; lower: 25% load.

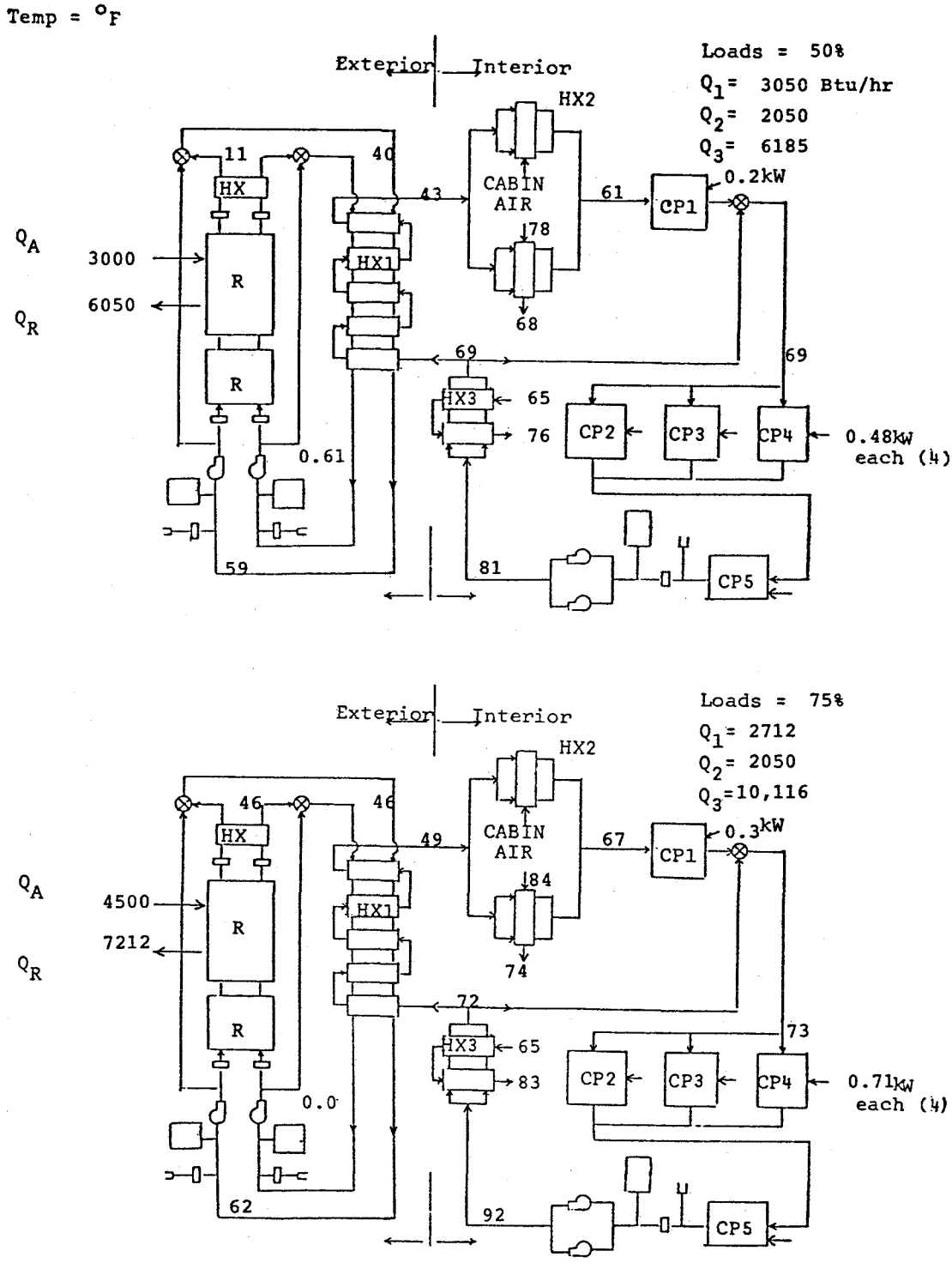


Figure 4. Results of loop simulation. Upper: 50% load; lower: 75% load.

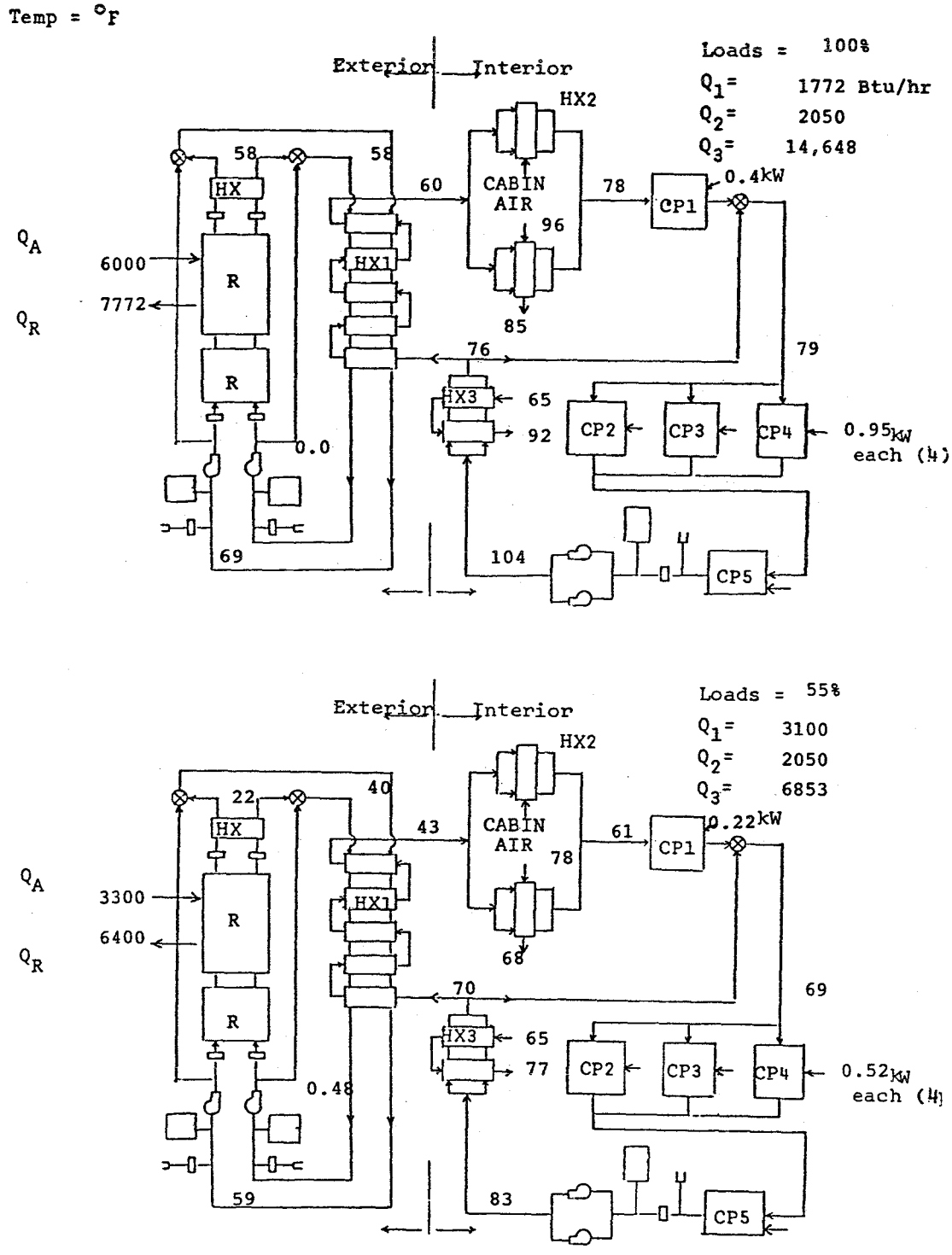


Figure 5. Results of loop simulation. Upper: 100% load; lower: 55% load.

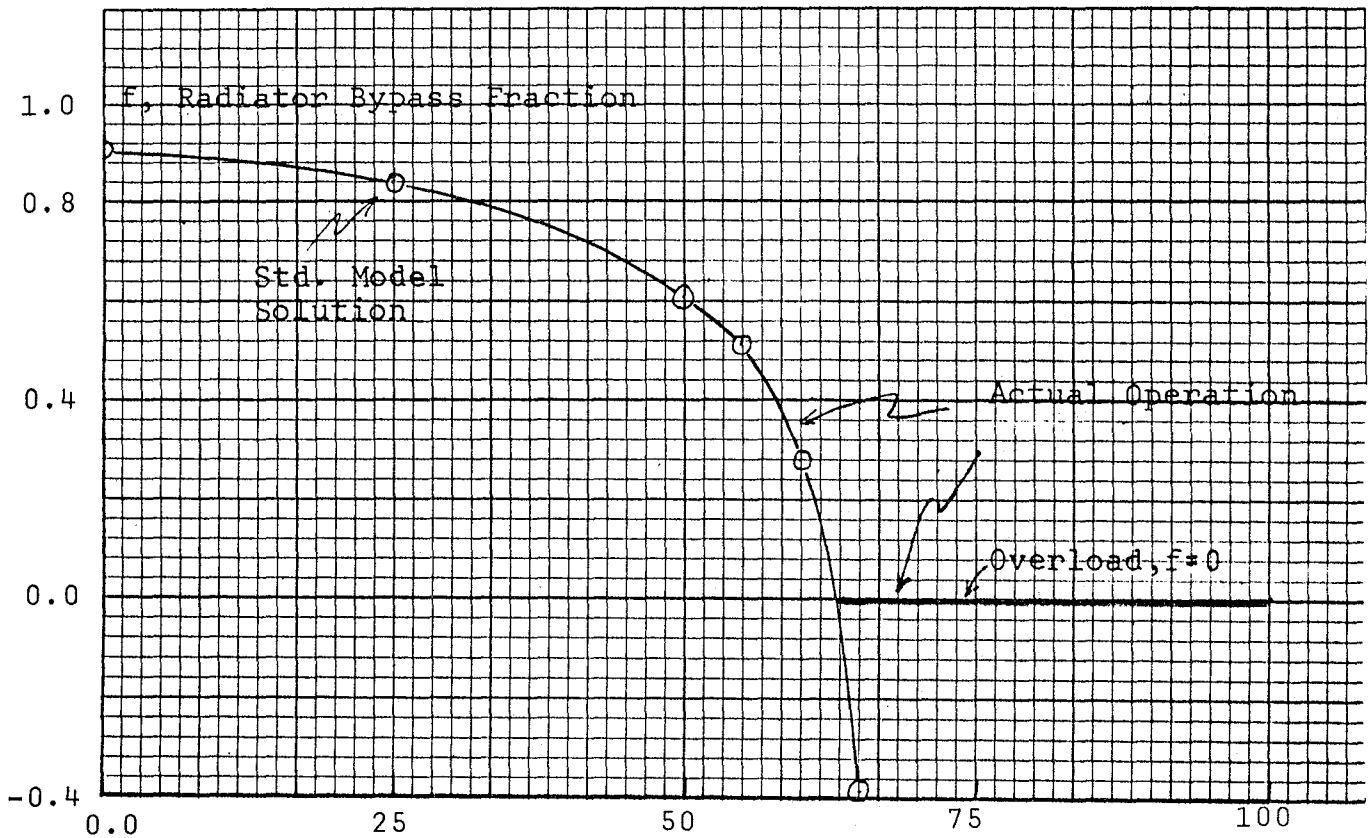


Figure 6. Standard Model solutions for f vs. % heat load.

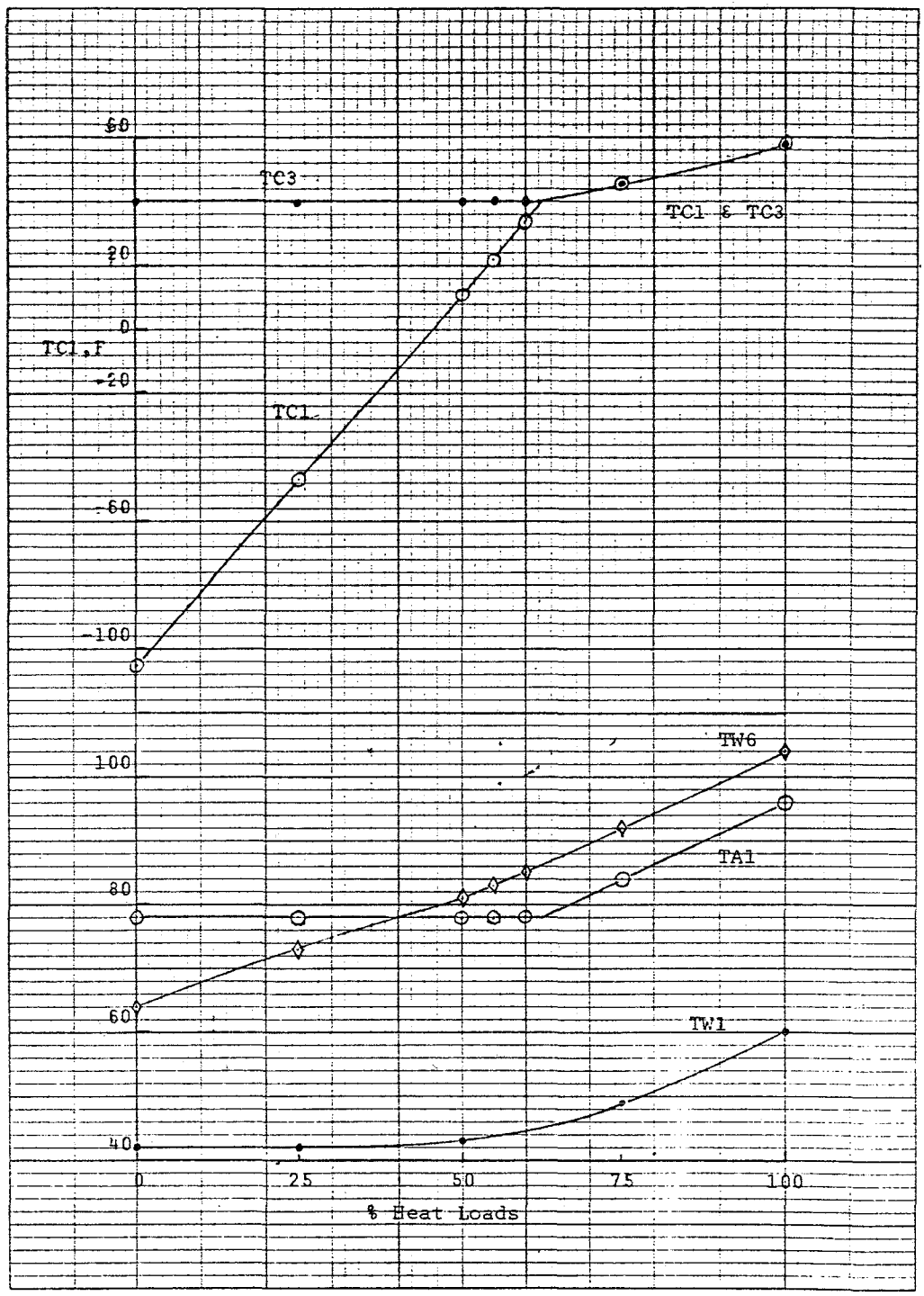


Figure 7. Typical loop temperatures vs % heat load.

IA-31

1985

NASA/ASEE SUMMER FACULTY RESEARCH FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

METHODOLOGY OF DESIGN AND ANALYSIS OF EXTERNAL WALLS OF SPACE STATION FOR HYPERVELOCITY IMPACTS BY METEOROIDS AND SPACE DEBRIS

Prepared by:	Fahim A. Batla, Ph.D., P.E.
Academic Rank:	Visiting Faculty
University and Department:	Purdue University - Indianapolis Mechanical Engineering Department
NASA/MSFC: Division: Branch:	Structures and Propulsion Lab Structures Division Structural Development Branch EP-13
MSFC Counterpart:	Erich E. Engler
Date:	August 9, 1985
Contract No.:	NGT 01-008-021 The University of Alabama in Huntsville

METHODOLOGY OF DESIGN AND ANALYSIS OF EXTERNAL WALLS OF SPACE STATION FOR HYPERVELOCITY IMPACTS BY METEOROIDS AND SPACE DEBRIS

BY

Dr. Fahim A. Batla
Mechanical Engineering Department
Purdue University - Indianapolis, Indiana

ABSTRACT

The study is directed towards the development of criteria and methodology for the design and analysis of Space Station wall elements for collisions with meteoroids and space debris at hypervelocities. These collisions will occur at velocities of 10 km/s or more and can be damaging to the external wall elements of the Space Station. The wall elements need to be designed to protect the pressurized modules of the Space Station from functional or structural failure due to these collisions at hypervelocities for a given environment and population of meteoroids and space debris. The design and analysis approach and the associated computer program presented herein is to achieve this objective, including the optimization of the design for a required overall probability of no penetration. The approach is based on the presently available experimental and actual data on meteoroids and space debris flux and damage assessments and the empirical relationships resulting from the hypervelocity impact studies in laboratories. Validation of this approach by further experimental work, specially using various structural and debris materials, is strongly recommended.

ACKNOWLEDGEMENTS

The author appreciates the opportunity of this research at NASA Marshall Space Flight Center and for the professional development and knowledge achieved through this project.

The successful conclusion of this research project is greatly due to the contributory efforts of author's NASA associates and he wishes to express his gratitude to everyone related to this project, especially to Mr. Erich E. Engler and Mr. Juan Maldonado of the Structural Development Branch of Structures and Propulsion Laboratory at MSFC.

A very special word of thanks to Dr. Jim Dosier and Mr. Leroy Osborn, the NASA Directors, and to Dr. Gerald Karr, the UAH Director of the NASA/ASEE Summer Faculty Research Fellowship Program, for making author's visit to MSFC a highly pleasant and memorable experience.

1. INTRODUCTION

In the past two decades a substantial amount of engineering and scientific research and collection of data has been done to determine the environment and population of meteoroids and man-made debris in space and the damage that may occur to the space structures by the collisions with these meteoroids and space debris at hypervelocities of 10 km/s or more. At present, there is a need to design the external wall elements of the Space Station for protection from functional and structural failures due to hypervelocity impacts for a given environment and population of meteoroids and space debris and the overall probability of no penetration. Although a large number of research papers, reports and other publications are presently avaiable on this subject, the basic phenomena and mechanisms involved in the hypervelocity impacts and the related empirical equations for structural design and analysis are still far from being well-defined or reliable.

The purpose of this report is to present the most "reasonable" methodology of structural design and analysis of the external wall elements of the Space Station for hypervelocity impacts by meteoroids and space debris and the associated computer program to be used as a preliminary working tool by the designers and analysts. The approach is based on the presently available actual data and the extrapolated assessments of meteoroid and space debris flux and the empirical relationships of the structural design parameters resulting from the hypervelocity impact studies in laboratories.

The report first presents the basic experimental and theoretical information that provides the basis for the development of copmuter program. The later part of this report presents illustrative numerical examples using this program. It is strongly suggested that the users of the computer program presented herein be familiar with the entire report, including its references, and the detailed program manual in order to clearly understand the theories and the experimental data used in the development of this program, and the extent of its usefulness, accuracy and limitations.

A listing of the computer program is provided in the Appendix.

2. METHODOLOGY

Meteoroid and Space Debris Flux and Probability of No Penetration

If N is the flux of meteoroids or space debris defined as particles of mass `m' or diameter `d' and greater per unit area per unit time impacting the spacecraft, and P_x is the probability of impacts by `n' particles or less, the flux and the probability are related as follows (1).

$$P_{x<n} = \sum_{r=0}^{r=n} \frac{e^{-NAT}(NAT)^r}{r!} \qquad (1)$$

Where:

A = exposed surface area for meteoroids or projected area for space debris

T = time of exposure in space

The probability of exactly `n' impacts is

$$P_n = \frac{e^{-NAT}(NAT)^n}{n!} \qquad (2)$$

and the probability of zero impact, P_o , by any particle of mass `m' or diameter `d' and greater is then given by the following equation.

$$P_o = e^{-NAT} \qquad (3)$$

If the external wall elements of the Space Station are designed for no penetration by any particle of mass `m' or diameter `d' and smaller, the probability of no penetration is also P_o. Equation 3 is therefore used to determine the flux for the required probability of no penetration and the flux-mass or flux-diameter models described in the following sections are used to determine the mass and the diameter of the particle that should accordingly be used for the design and analysis of the wall elements.

It should be noted that the overall probability of no penetration is expressed in terms of the individual probabilities of no penetration as follows.

$$\text{Overall } P_o = P_{o1} \times P_{o2} \times P_{o3} \times \text{- - - -} \quad (4)$$

Meteoroid Flux-Mass Model

The meteoroid flux-mass model is described in detail in Reference 2 and is illustrated in Figures 1 and 2. The model states that the average annual comulative total flux, N, in impacts per square meter per second of meteroids of mass `m' or greater in grams impacting on a spacecraft is

$$\log N = - 14.37 - 1.213 \log m + \log G + \log S \quad (5)$$

$$\text{for } 10^{-6} \leqslant m \leqslant 10^{0}$$

Where:

G (gravitational defocusing factor) = $0.568 + 0.432 \sin\theta$

$$S \text{ (earth shielding factor)} = \frac{1 + \cos\theta}{2}$$

$$\sin\theta = \frac{R}{R + H}$$

and R = radius of earth

H = altitude above the earth surface in km

Equation 5 is used to determine the mass of meteoroid particle corresponding to the previously calculated flux.

Space Debris Flux-Diameter Models

The space debris flux-diameter models are taken from Reference 3 and are further illustrated in Figure 3. The cumulative flux, N, of debris particles between 1 mm and 1 cm diameter and defined as impacts per square meter per year of debris particles of diameter `d' or greater in cm impacting on a spacecraft is given by the following equations.

a) At 400 km altitude, $\log N = - 5.82 - 2.42 \log d$

b) At 500 km altitude, $\log N = - 5.46 - 2.52 \log d$ (6)

AVERAGE TOTAL MASS – FLUX MODEL (INCLUDES A 10% STREAM CONTRIBUTION AT 1 gm, VANISHING AT 10^{-6} gm)

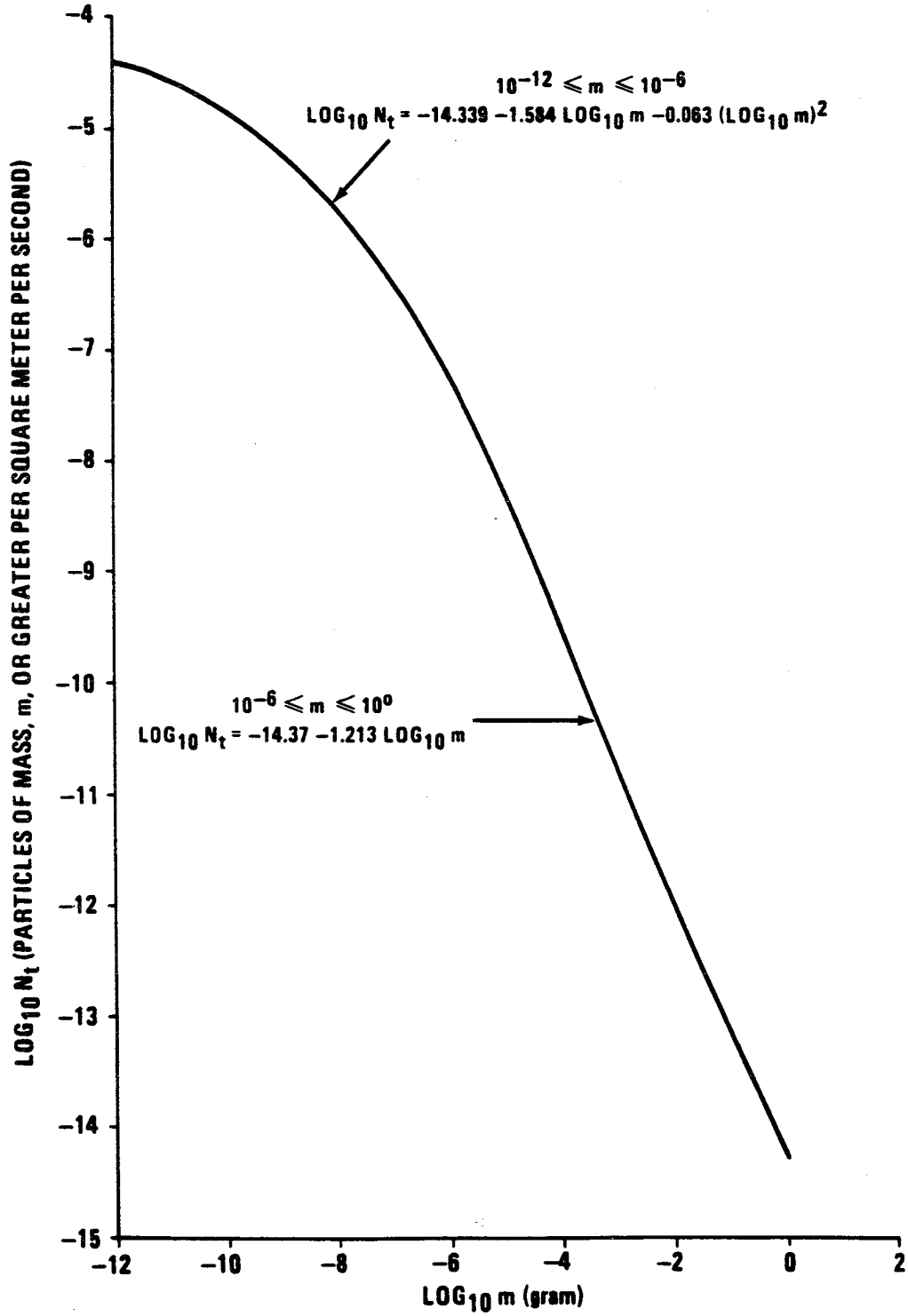


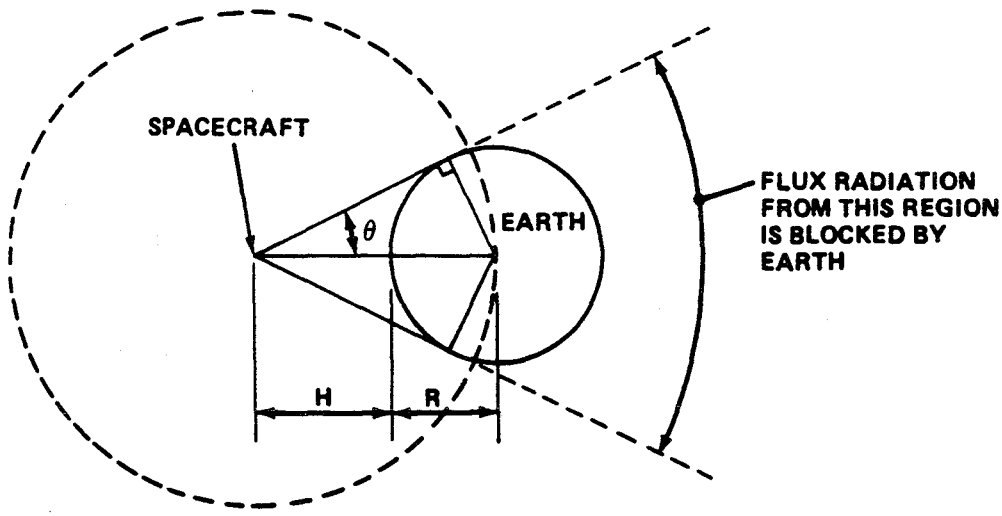
Figure 1

GRAVITATIONAL DEFOCUSING FACTOR

$$G = 0.568 + \frac{0.432}{r}$$

r ⟶ DISTANCE FROM EARTH CENTER IN EARTH RADII

EARTH SHIELDING FACTOR, S
(RANDOMLY ORIENTED SPACECRAFT)



$$\text{SIN}\,\theta = \frac{R}{R+H}$$

$$S = \frac{1+\text{COS}\,\theta}{2}$$

Figure 2

FUTURE DEBRIS FLUX

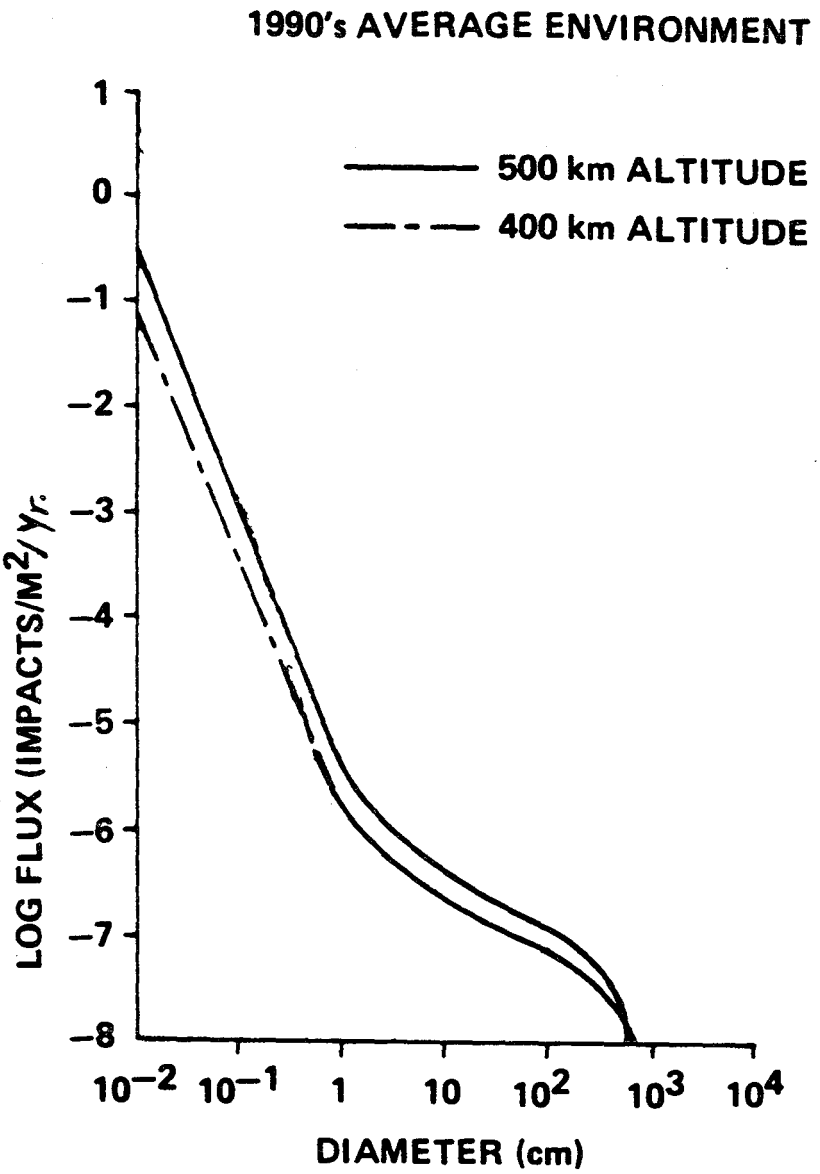


Figure 3

The debris flux for particles smaller than 1 cm is highly uncertain and is expected to change as new data on man-made space debris becomes available. The current uncertainty in the projected 1 cm flux is estimated to be a factor of 3, and for the 1 mm flux, a factor of 10 (3).

References 3 and 4 (Figure 4) are used to approximate the cumulative flux for debris particles between 1 cm and 4 cm diameters yielding the following equations.

a) At 400 km altitude, $\log N = - 5.82 - 1.13 \log d$

b) At 500 km altitude, $\log N = - 5.46 - 0.90 \log d$ (7)

Equations 6 and 7 are used to determine the diameter of the debris particle corresponding to the previously calculated magnitude of flux.

Single Wall Design Equations

Based on the hypervelocity impact mechanics and experimental studies (5,6,7), the following equations are used for the single wall design (1,2).

For thick plates:

$$P_\infty = k_\infty \, m^{0.352} \, \rho_m^{1/6} \, V^{2/3} \qquad (8)$$

For thin plates:

$$t = k_1 \, m^{0.352} \, \rho_m^{1/6} \, V^{0.875} \qquad (9)$$

Where:

P_∞ = penetration depth in thick plate, cm

t = thickness of thin plate, cm

k_∞ and k_1 = material constants

m = mass of particle, gm

ρ_m = density of particle material, gm/cu. cm

SPACE DEBRIS

OBSERVED FLUX

NORAD CURRENTLY TRACKS 5000 OBJECTS > 4 cm DIA,
AVERAGE VELOCITY ≈ 10 km/s, VARIOUS DIRECTIONS.

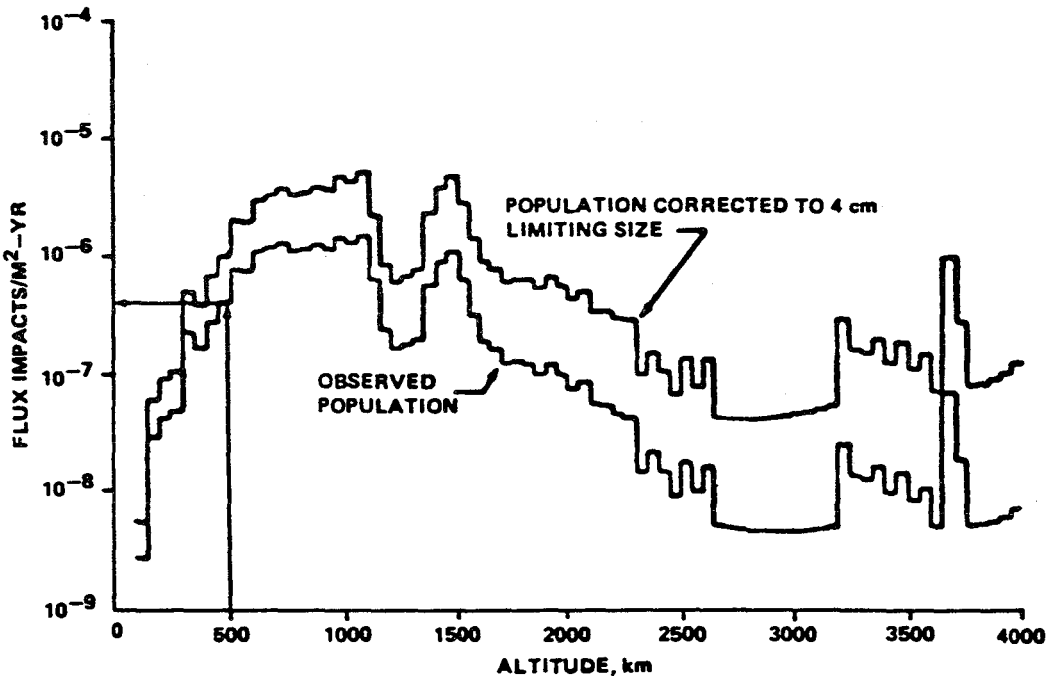


Figure 4

and V = velocity of particle, km/s

Equations 8 and 9 are further illustrated in Figure 5.

Double Wall Design Equations

The following paragraphs present two alternate equations for the the design of double wall system which consists of an exterior shield and a backup wall called bumper and hull, respectively (Figure 6).

The first equation

$$t_2 = \frac{5.08\ V^{0.278}\ d}{(t_1/d)^{0.528}\ (h/d)^{1.39}} \qquad (10)$$

is suggested by Nysmith (12).

Where:

t_1 = bumper thickness, cm

t_2 = hull thickness, cm

d = diameter of particle, cm

h = spacing between walls, cm

Equation 10 was developed using the data from tests on the ballistic limit of aluminum double wall structures impacted by pyrex-glass spherical particles . Nysmith (12) suggests the following optimum ratios if Equation 10 is to be used for the double wall design.

$t_1/d \leqslant 0.5$

$t_1/d \leqslant 1.0$

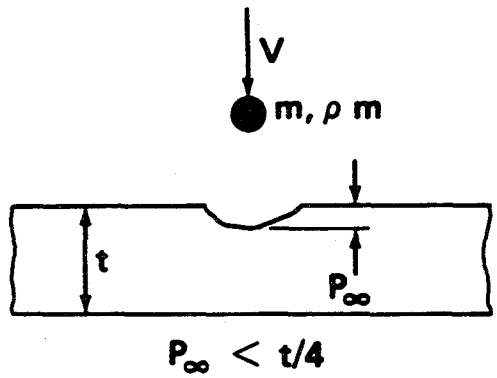
$h/d \leqslant 100$

$t_1/t_T = o.35$

CHARACTERISTICS:

HIGH PRESSURE (1 TO 10 MEGABARS)
COMPLETE PULVERIZATION OF TARGET AND PROJECTILE LOWER BOUND, 1–10 km/s

THICK PLATES

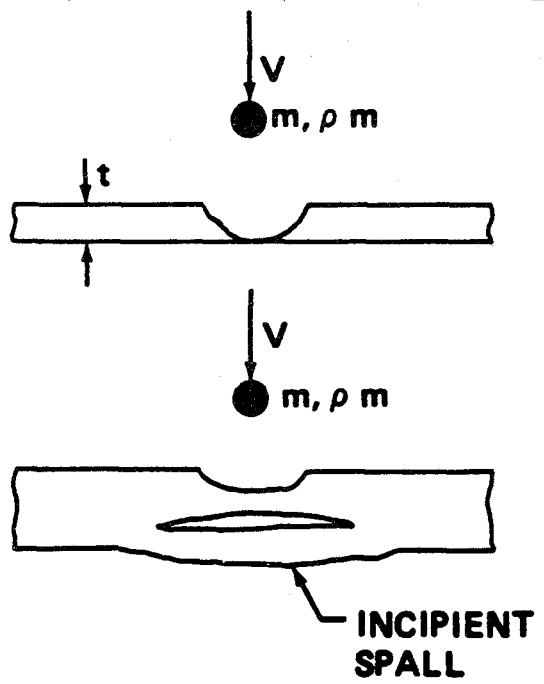


$$P_\infty = K_\infty m^{0.352} \rho_m^{1/6} V^{2/3}$$

K ⟶ CONST. ≈ 0.42 (Al)

THIN PLATES

"BALLISTIC LIMIT" OR "THRESHOLD THICKNESS"



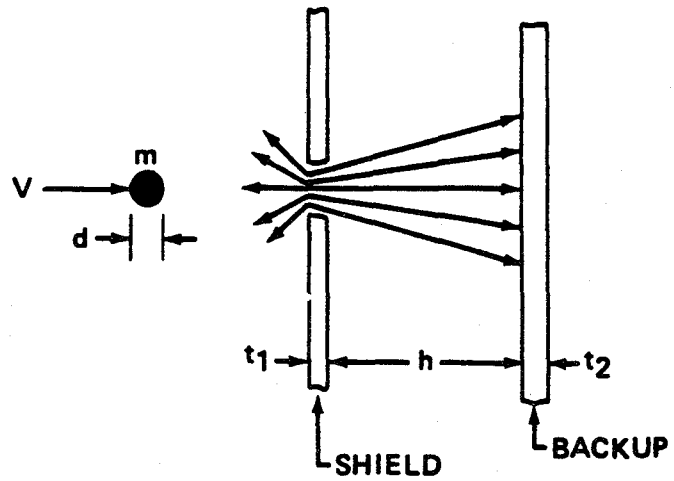
$$t = K_1 m^{0.352} \rho_m^{1/6} V^{0.875}$$

K_1 ⟶ CONST. ≈ 0.57 (Al)

$t_{SPALL} \approx 1.5\,t$

Figure 5

METEOROID SHIELDS (BUMPERS)



BALLISTIC LIMIT OF DOUBLE–WALLED STRUCTURES

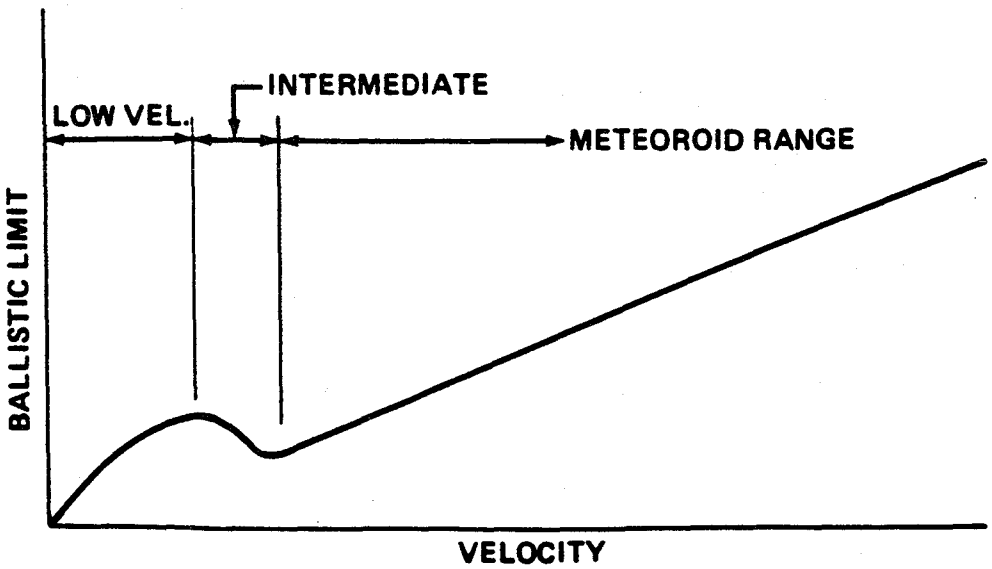


Figure 6

Where:

t_T = total thickness of both walls

The second equation

$$t_2 = \frac{0.055\,(\rho_m\,\rho_t)^{1/6}\; m^{1/3}\; V^{1/2}}{h^{1/2}}\,(70000/F_y) \qquad (11)$$

is developed by Cour-Palais (13,14).

Where:

ρ_t = density of bumper material, gm/cu. cm

F_y = 0.2 percent yield stress for hull material, lb/in^2

Cour-Palais (13) suggests the following optimum ratios if Equation 11 is to be used for the design of double walls.

$t_1/d \cong 0.16$

$t_2/d \cong 0.51$

$h/d \leqslant 25$

$t_1/t_T \cong 0.25$

The bumper/total thickness ratios mentioned above are used to determine the design of the double wall system for the given overall probability of no penetration, protecting the hull from metroids and, as well as, from space debris.

3. NUMERICAL EXAMPLES

Design Data

The design data used for the illustrative examples presented herein is as follows. It should be noted that this data does not necessarily represent the actual data that is to be used for the design of Space Station.

Exposed projected area (for debris)	= 228.41 sq. m
Exposed surface area (for meteoroid)	= 831.64 sq. m
Time of exposure in orbit	= 10 years
Orbital altitude	= 500 km
Velocity of debris	= 10 km/s
Velocity of meteoroids	= 20 km/s
Density of debris material	= 2.81 gm/cu. cm
Density of meteroid material	= 0.5 gm/cu. cm
Density of bumper material	= 2.81 gm/cu. cm
0.2 percent yield stress for hull material	= 70000 lb/sq. in

Example 1

For a 0.985 probability of no penetration for debris, the debris particle mass and diameter are as follows.

Mass = 0.682 gm

(Equations 3 and 6)

Diameter = 0.774 cm

For a 0.995 probability of no penetration for meteoroid, the meteoroid particle mass and diameter are as follows.

Mass = 0.208 gm

(Equations 3 and 5)

Diameter = 0.926 cm

Example 2

Using the double wall system with 10.16 cm spacing between walls and a bumper/hull thickness ratio of 0.5, and

the data from Example 1, following are the required bumper and hull thicknesses according to Equation 11 (13).

a) For debris protection:

Bumper thickness = 0.107 cm

Hull thickness = 0.214 cm

b) For meteoroid protection:

Bumper thickness = 0.108 cm

Hull thickness = 0.216 cm

Example 3

In this example the design of the double walls is optimized for a required overall probability of 0.982.

a) Using Equation 10 (12):

	Bumper thickness	= 0.284 cm
	Hull thickness	= 0.527 cm
and	Maximum debris mass	= 1.028 gm
	Maximum meteoroid mass	= 0.150 gm
	Maximum debris diameter	= 0.887 cm
	Maximum meteoroid diameter	= 0.831 cm
	Probability of no penetration by debris	= 0.989
	Probability of no penetration by meteoroid	= 0.993

b) Using Equation 11 (13):

	Bumper thickness	= 0.071 cm
	Hull thickness	= 0.223 cm
and	Maximum debris mass	= 0.770 gm
	Maximum meteoroid mass	= 0.228 gm
	Maximum debris diameter	= 0.806 cm
	Maximum meteoroid diameter	= 0.955 cm

Probability of no penetration by debris = 0.986

Probability of no penetration by meteoroid = 0.996

4. SUMMARY AND CONCLUSIONS

The methodology and an associated computer program for the design and analysis of the external wall elements of the Space Station has been presented in this report. The program can be used with the following options.

i) Single or double-wall design.
ii) For a given individual probability of no penetration, design for protection from space debris or meteoroid.
iii) Design optimization for a given overall probability of no penetration.
iv) Thick or thin plate design for the single wall.
v) Design of double wall system by Equation 10 (12) or by Equation 11 (13).
vi) Analysis of a given design for all the options mentioned above.

The program is based on the presently available data on meteoroid and space debris environment and population and hypervelocity impact studies. This data is mostly experimental, empirical or extrapolation of the actual data. The resulting empirical design relationships are based on a very limited number or narrow-ranged theoretical and experimental studies leading to very highly unreliable designs. For example, the design results of the double wall system of Example 3 using Equation 10 (12) and Equation 11 (13) differ by a wide margin.

The computer program presented in this report should therefore be used with good judgement and caution and only as a preliminary working tool and should be updated as soon as any new and more reliable information on this subject becomes avaiable.

In conclusion, there is a need of further and extensive study of the phenomena and mechanisms involved in the design of the external wall elements of the Space Station for protection from hypervelocity impacts by meteoroids and space debris, including experimental verifications, in order to achieve a more reliable criteria and methodology of design and analysis. Use of a wide range of materials in these tests and studies, including composites, is highly recommended for the purpose of providing a higher degree of accuracy in the resulting empirical relationships and for the fact that these tests may lead to the most suitable structural material for the design of the Space Station.

REFERENCES

1. Frost, V. C., "Meteoroid Damage Assessment," NASA SP-8042, May 1970.

2. Susko, M., "A Review of Micrometeoroids Flux Measurements and Models for Low Orbital Altitudes of the Space Station," NASA TM-86466, September 1984.

3. Cour-Palais, B. G., Private Communication consisting of a section from the Space Station Customer Services Handbook being compiled at the Goddard Space Center, July 9, 1984.

4. Kessler, D. J., "Sources of Orbital Debris and the Projected Environment for Future Spacecraft," J. Spacecraft, pp. 357-360, Vol. 18, No. 4, July-August 1981.

5. Kinslow, R., Ed., "High-Velocity Impact Phenomena," Academic Press, 1970.

6. Fu, T., "Analytical Methods for Calculating Vulnerability of Meteoroid Impacts," Teledyne Brown Engineering, ASD-ASTN-1490, March 1972.

7. Zukas, J. A., Ed., "Impact Dynamics," John Wiley and Sons, 1982.

8. Frost, V. C., "Aerospace Meteoroid Environment and Penetration Criterion," Aerospace Corporation, TOR-269 (4560-40)-2, August 1964.

9. Nevins, C. D., "Analysis of Meteoroid Impact on Spacecraft Structures," unpublished report, Structures and Propulsion Laboratory, MSFC, April 1977.

10. Swift, H. F., and Hopkins, A. K., "The Effects of Bumper Material Properties on the Operation of Space Hypervelocity Particle Shields," AFML-TR-68-257, September 1968.

11. Madden, R., "Ballistic Limit of Double-Walled Meteoroid Bumper Systems," NASA TN D-3916, April 1967.

12. Nysmith, C. R., "An Experimental Impact Investigation of Aluminum Double-Sheet Structures," AIAA Hypervelocity Impact Conference, Cincinnati, Ohio, Paper No. 69-375, April 30 - May 2, 1969.

13. Cour-Palais, B. G., "Meteoroid Protection by Multiwall Structures," AIAA Hypervelocity Impact Conference, Cincinnati, Ohio, Paper No. 69-372, April 30 - May 2, 1969.

14. Cour-Palais, B. G., "Space Vehicle Meteoroid Shielding Design," Comet Halley Micrometeoroid Hazard Workshop, Proceedings of the International Workshop sponsored by the European Space Agency held in ESTEC, Noordwijk, Netherlands, pp. 85-92, April 18-19, 1979.

APPENDIX

COMPUTER PROGRAM LISTING

```
10 DECLARE DOUBLE PR(2),PD(2),PM(2),MASS(2),DIA(2),PROB(2),AR(2),PO(2),VP(2)
20 DECLARE DOUBLE XX,T1,KM,TP,DD,FLUX,GE,SH,YY,OP,T2,S1,S2,SP,TE
30 PRINT "* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *"
40 PRINT "* Welcome to program 'NONAME'                                    *"
50 PRINT "* This program is developed by Fahim A. Batla                    *"
60 PRINT "* in summer 1985 for design/analysis of space                    *"
70 PRINT "* station exterior wall for hypervelocity impact                 *"
80 PRINT "* by metreoroid/debris.                                          *"
90 PRINT "* Consult Program Manual for development and usage               *"
100 PRINT "* of this program.                                               *"
110 PRINT "* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *"
120 PRINT \ PRINT "Enter 'yes' if a hardcopy of output is desired,"
130 PRINT "otherwise enter 'no'."
140 PRINT \ INPUT ZZ$
150 IF ZZ$='no' GOTO 190
160 IF ZZ$='yes' GOTO 180
170 PRINT \ PRINT '*****Input Error'\ PRINT \ GOTO 120
180 OPEN 'lp:' FOR OUTPUT AS FILE #9
190 PRINT
200 REM design input
210 PRINT "Enter exposed projected area in square meters"
220 INPUT AR(2)\ PRINT
230 PRINT 'Enter exposed surface area in square meters'
240 INPUT AR(1)\ PRINT
250 PRINT "Enter time of exposure in years"
260 INPUT TE\ PRINT
270 PRINT "Enter altitude in km - use 400 or 500"
280 INPUT AL\ PRINT
290 REM PRINT "Enter meteoroid density"
300 PR(1)=.5
310 REM PRINT "Enter debris density"
320 PR(2)=2.81
330 LO=-LOG10(365.25*24*3600/(10^14.37))
340 IA=0
350 PRINT " Enter 'design' or 'analysis' "
360 INPUT CC$\ PRINT
370 IF CC$='design' GOTO 420
380 IF CC$='analysis' THEN IA=1
390 IF CC$='analysis' GOTO 700
400 PRINT \ PRINT "*****Input error"\ PRINT
410 GOTO 340
420 PRINT "Enter probability of no penetration for debris in decimal form"
430 INPUT PO(2)\ PRINT
440 PRINT 'Enter probability of no penetration for meteoroid in decimal form'
450 INPUT PO(1)\ PRINT
460 PRINT "***** Check input data"\ PRINT
470 PRINT 'Projected Area=';AR(2)\ PRINT 'Surface Area=';AR(1)
480 PRINT "Time=";TE\ PRINT "Altitude=";AL
490 PRINT "Meteoroid Density=";PR(1)\ PRINT "Debris Density=";PR(2)
500 PRINT "Meteoroid Probability=";PO(1)
```

```
510 PRINT 'Debris Probability=';P0(2)\ PRINT \ PRINT \ PRINT
520 PRINT "Enter 'next' if all the input data is correct,"
530 PRINT "otherwise enter 'repeat'."\ PRINT \ PRINT
540 INPUT CC$\ PRINT
550 IF CC$='repeat' GOTO 190
560 IF CC$='next' GOTO 590
570 PRINT '*****Input Error'\ PRINT
580 GOTO 520
590 FL=-1*LOG(P0(1))/AR(1)/TE
600 FLD=-1*LOG(P0(2))/AR(2)/TE
610 REM For Debris
620 IF AL=500 GOTO 660
630 PD(2)=(FLD*10^5.82)^(-1/2.42)
640 IF PD(2)>1 THEN PD(2)=(FLD*10^5.82)^(-1/1.13)
650 GOTO 680
660 PD(2)=(FLD*10^5.46)^(-1/2.52)
670 IF PD(2)>1 THEN PD(2)=(FLD*10^5.46)^(-1/.9)
680 PM(2)=PI*PR(2)*PD(2)^3/6
690 REM For Meteoroid
700 STH=6371/(6371+AL)
710 CTH=SQR(1-STH^2)
720 GE=.568+.432*STH
730 SH=(1+CTH)/2
740 IF IA=1 GOTO 960
750 FLM=FL/GE/SH
760 PM(1)=(FLM*10^LO)^(-1/1.213)
770 PD(1)=(6*PM(1)/PI/PR(1))^(1/3)
780 PRINT \ PRINT
790 PRINT TAB(15%);'For Debris',TAB(30%);'For Meteoroid'\ PRINT
800 PRINT 'Flux',FLD;TAB(10%),FLM
810 PRINT 'Projectile'\ PRINT 'Diameter',PD(2);TAB(10%),PD(1)
820 PRINT 'Projectile'\ PRINT 'Mass',PM(2);TAB(10%),PM(1)
830 PRINT \ PRINT
840 IF PD(2)>4 THEN PRINT "***** IMPORTANT NOTE"\ PRINT
850 IF PD(2)>4 THEN PRINT "      If debris diameter is greater than 4 cm,"
860 IF PD(2)>4 THEN PRINT "      the computed diameter and mass for debris"
870 IF PD(2)>4 THEN PRINT "      is incorrect."
880 IF PD(2)>4 THEN PRINT \ PRINT
890 PRINT "If you desire to change any input data enter 'repeat',"
900 PRINT "otherwise enter 'next' to continue."
910 INPUT CC$\ PRINT
920 IF CC$='repeat' GOTO 200
930 IF CC$='next' GOTO 950
940 PRINT "*****Input Error"\ PRINT \ GOTO 890
950 PRINT
960 PRINT 'Enter debris projectile velocity in km/sec.'
970 INPUT VP(2)\ PRINT
980 PRINT 'Enter meteoroid projectile velocity in km/sec.'
990 INPUT VP(1)\ PRINT
1000 PRINT "Enter 'single' for a single wall design"
```

```
1010 PRINT "or 'double' for a double wall design."
1020 INPUT CC$\ PRINT
1030 IF CC$='single' GOTO 1060
1040 IF CC$='double' GOTO 2700
1050 PRINT '*****Input Error'\ PRINT \ GOTO 1000
1060 PRINT "Enter 'thick' for thick plate design"
1070 PRINT "or 'thin' for thin plate (threshold thickness) design."
1080 INPUT CC$\ PRINT
1090 IF CC$='thick' GOTO 1120
1100 IF CC$='thin' GOTO 1190
1110 PRINT '*****Input Error'\ PRINT \ GOTO 1060
1120 DD=2/3
1130 PRINT "Enter K (material constant) for thick plate design."
1140 INPUT KM\ PRINT
1150 PRINT 'Enter thickness/penetration ratio (>=4.0)'
1160 PRINT 'for thick plate design.'
1170 INPUT TP\ PRINT
1180 GOTO 1270
1190 DD=.875
1200 PRINT 'Enter K1 (material constant) for thin plate design.'
1210 INPUT KM\ PRINT
1220 TP=1
1230 IF IA=0 GOTO 1270
1240 PRINT 'Enter thickness of the plate in cm.'
1250 INPUT T1\ PRINT
1260 GOTO 2080
1270 PRINT "Enter 'optimize' if, for a given overall probability and"
1280 PRINT "the existing design data , the single wall thickness is"
1290 PRINT "to be minimized. Otherwise enter 'next'."\ PRINT
1300 INPUT DD$\ PRINT
1310 IF DD$='optimize' GOTO 2010
1320 IF DD$='next' GOTO 1340
1330 PRINT '*****Input Error'\ GOTO 1270
1340 PRINT 'For the design data to be used,'
1350 PRINT "enter 'debris' or 'meteoroid'."
1360 INPUT BB$\ PRINT
1370 I=0
1380 IF BB$='meteoroid' THEN I=1
1390 IF BB$='debris' THEN I=2
1400 IF I>0 GOTO 1420
1410 PRINT '*****Input Error'\ PRINT \ GOTO 1340
1420 IF I=1 THEN J=2
1430 IF I=2 THEN J=1
1440 T=KM*TP*(PM(I)^.352)*(PR(I)^(1/6))*(VP(I)^DD)
1450 PRINT 'Single ';CC$;' plate thickness for ';BB$;' = ';T;' cm'
1460 PRINT \ PRINT
1470 MASS(I)=PM(I)
1480 DIA(I)=PD(I)
1490 IF CC$='thick' THEN DD=2/3
1500 IF CC$='thin' THEN DD=.875
```

```
1510 MASS(J)=PM(I)*((PR(I)/PR(J))^(1/2.112))*((VP(I)/VP(J))^(DD/.352))
1520 DIA(J)=(6*MASS(J)/PI/PR(J))^(1/3)
1530 XX=MASS(1)^(-1.213)/(10^LO)
1540 FLUX=XX*GE*SH
1550 PROB(1)=EXP(-1*FLUX*AR(1)*TE)
1560 IF AL=500 GOTO 1600
1570 IF DIA(2)>1 THEN YY=-1.13
1580 IF DIA(2)<=1 THEN YY=-2.42
1590 FLUX=DIA(2)^YY/(10^5.82)\ GOTO 1630
1600 IF DIA(2)>1 THEN YY=-.9
1610 IF DIA(2)<=1 THEN YY=-2.52
1620 FLUX=DIA(2)^YY/(10^5.46)
1630 PROB(2)=EXP(-1*FLUX*AR(2)*TE)
1640 PRINT \ PRINT
1650 PRINT '***** For plate thickness=';T;'cm,'
1660 PRINT '      designed for ';BB$;' protection,'
1670 PRINT '      following are the comparisons between'
1680 PRINT '      debris and meteoroid.'
1690 PRINT \ PRINT
1700 PRINT TAB(15%);'For Debris',TAB(30%);'For Meteoroid'\ PRINT
1710 PRINT 'Projectile'\ PRINT 'Velocity',VP(2);TAB(10%),VP(1)
1720 PRINT 'Projectile'\ PRINT 'Diameter',DIA(2);TAB(10%),DIA(1)
1730 PRINT 'Projectile'\ PRINT 'Mass',MASS(2);TAB(10%),MASS(1)
1740 PRINT 'Probability',PROB(2);TAB(10%),PROB(1)
1750 PRINT \ PRINT \ IF DIA(2)>4 THEN PRINT '*****IMPORTANT NOTE'\ PRINT
1760 IF DIA(2)>4 THEN PRINT 'Computed probability for debris is incorrect'
1770 IF DIA(2)>4 THEN PRINT 'as the diameter of projectile is greater'
1780 IF DIA(2)>4 THEN PRINT 'than 4 cm.'
1790 PRINT \ PRINT
1800 PRINT "*****Enter 'repeat' if you want to compute the thickness"
1810 IF I=1 THEN PRINT "for debris or for different velocities etc.,"
1820 IF I=2 THEN PRINT "for meteoroid or for different velocities etc.,"
1830 PRINT "or enter 'start' to start all over,"
1840 PRINT "or enter 'next' to continue,"
1850 PRINT "or enter 'end' to stop."
1860 INPUT DD$\ PRINT
1870 IF DD$='repeat' GOTO 950
1880 IF DD$='next' GOTO 1920
1890 IF DD$='start' GOTO 120
1900 IF DD$='end' GOTO 4900
1910 PRINT '*****Input Error'\ PRINT \ GOTO 1800
1920 PRINT "Enter 'optimize' if, for a given overall probability and"
1930 PRINT "the existing design data , the single wall thickness is"
1940 PRINT "to be minimized. Otherwise enter 'start' to start"
1950 PRINT "all over, or enter 'end' to stop."
1960 INPUT DD$\ PRINT
1970 IF DD$='optimize' GOTO 2010
1980 IF DD$='end' GOTO 4900
1990 IF DD$='start' GOTO 120
2000 PRINT '*****Input Error'\ PRINT \ GOTO 1920
```

```
2010 T1=0
2020 PRINT 'Enter overall probability'
2030 INPUT OP\ PRINT
2040 INC=1
2050 K=0
2060 K=K+1
2070 T1=T1+INC
2080 XX=T1/KM/TP/(PR(1)^(1/6))/(VP(1)^DD)
2090 MASS(1)=XX^(1/.352)
2100 XX=T1/KM/TP/(PR(2)^(1/6))/(VP(2)^DD)
2110 MASS(2)=XX^(1/.352)
2120 DIA(1)=(6*MASS(1)/PI/PR(1))^(1/3)
2130 DIA(2)=(6*MASS(2)/PI/PR(2))^(1/3)
2140 XX=MASS(1)^(-1.213)/(10^LO)
2150 FLUX=XX*GE*SH
2160 PROB(1)=EXP(-1*FLUX*AR(1)*TE)
2170 IF AL=500 GOTO 2210
2180 IF DIA(2)>1 THEN YY=-1.13
2190 IF DIA(2)<=1 THEN YY=-2.42
2200 FLUX=DIA(2)^YY/(10^5.82)\ GOTO 2240
2210 IF DIA(2)>1 THEN YY=-.9
2220 IF DIA(2)<=1 THEN YY=-2.52
2230 FLUX=DIA(2)^YY/(10^5.46)
2240 PROB(2)=EXP(-1*FLUX*AR(2)*TE)
2250 IF IA=1 GOTO 2400
2260 IF K>25 GOTO 2290
2270 IF (PROB(1)*PROB(2))>OP GOTO 2320
2280 GOTO 2060
2290 PRINT \ PRINT \ PRINT '***** No optimization as thickness is getting'
2300 PRINT '      very large or very small.'
2310 GOTO 2520
2320 IF ABS(INC)=.001 GOTO 2360
2330 T1=T1-INC
2340 INC=INC/10
2350 GOTO 2050
2360 PRINT \ PRINT
2370 PRINT '***** For overall probability of';OP;', the required'
2380 PRINT '      minimum thickness is';T1;'cm. The comparisons'
2390 PRINT '      between meteoroid and debris are as follows.'\ GOTO 2420
2400 PRINT '***** For single plate thickness =';T1;'cm, the comparisons'
2410 PRINT '      between meteoroid and debris are as follows.'
2420 PRINT \ PRINT
2430 PRINT TAB(15%);'For Debris',TAB(30%);'For Meteoroid'\ PRINT
2440 PRINT 'Projectile'\ PRINT 'Velocity',VP(2);TAB(10%),VP(1)
2450 PRINT 'Projectile'\ PRINT 'Diameter',DIA(2);TAB(10%),DIA(1)
2460 PRINT 'Projectile'\ PRINT 'Mass',MASS(2);TAB(10%),MASS(1)
2470 PRINT 'Probability',PROB(2);TAB(10%),PROB(1)
2480 PRINT \ PRINT \ IF DIA(2)>4 THEN PRINT '*****IMPORTANT NOTE'\ PRINT
2490 IF DIA(2)>4 THEN PRINT 'Computed probability for debris is incorrect'
2500 IF DIA(2)>4 THEN PRINT 'as the diameter of projectile is greater'
```

```
2510 IF DIA(2)>4 THEN PRINT 'than 4 cm.
2520 PRINT \ PRINT
2530 IF IA=0 GOTO 2610
2540 PRINT "Enter 'repeat' if analysis for different thickness is desired."
2550 PRINT "otherwise enter 'start'  to start all over or 'end' to stop."
2560 INPUT CC$\ PRINT
2570 IF CC$='repeat' GOTO 1240
2580 IF CC$='start' GOTO 120
2590 IF CC$='end' GOTO 4900
2600 PRINT '*****Input Error'\ PRINT \ GOTO 2540
2610 PRINT "Enter 'repeat' if you want to minimize the thickness"
2620 PRINT "for different overall probability,"
2630 PRINT "or enter 'start' to start all over,"
2640 PRINT "or enter 'end' to stop."
2650 INPUT EE$\ PRINT
2660 IF EE$='repeat' GOTO 2010
2670 IF EE$='start' GOTO 120
2680 IF EE$='end' GOTO 4900
2690 PRINT '*****Input Error'\ PRINT \ GOTO 2610
2700 IF IA=1 GOTO 2800
2710 PRINT "Enter 'optimize' if, for a given overall probability"
2720 PRINT 'and the existing design data, the double wall'
2730 PRINT "design is to be optimized, otherwise enter 'next'."
2740 INPUT CC$\ PRINT
2750 IF CC$='next' GOTO 4020
2760 IF CC$='optimize' GOTO 2780
2770 PRINT '*****Input Error'\ PRINT \ GOTO 2710
2780 PRINT 'Enter overall probability.'
2790 INPUT OP\ PRINT
2800 PRINT "Enter 'Cour-Palais' or 'Nysmith' for design method to be used."
2810 INPUT CC$\ PRINT
2820 IF CC$='Cour-Palais' GOTO 2850
2830 IF CC$='Nysmith' GOTO 2910
2840 PRINT '*****Input Error'\ PRINT \ GOTO 2800
2850 IO=1
2860 PRINT 'Enter bumper material density in gm/cu. cm.'
2870 INPUT RB\ PRINT
2880 PRINT 'Enter 0.2 percent yield stress for hull material in PSI.'
2890 INPUT FY\ PRINT
2900 GOTO 2920
2910 IO=2
2920 PRINT 'Enter spacing between walls in cm.'
2930 INPUT SP\ PRINT
2940 IF IA=0 GOTO 3000
2950 PRINT 'Enter bumper thickness in cm.'
2960 INPUT T1\ PRINT
2970 PRINT 'Enter hull thickness in cm.'
2980 INPUT T2\ PRINT
2990 GOTO 3100
3000 T2=0
```

```
3010 INC=1
3020 IF IO=2 GOTO 3070
3030 XX=.0075*(RB^(1/3))*(70000/FY)
3040 S1=XX*PR(1)*(VP(1)^2)
3050 S2=XX*PR(2)*(VP(2)^2)
3060 IF SP<=0 THEN SP=(S1+S2)/2
3070 K=0
3080 K=K+1
3090 T2=T2+INC
3100 IF IO=2 GOTO 3150
3110 XX=T2*(SP^(1/2))/(RB^(1/6))/((70000/FY)^(1/2))/.0443
3120 DIA(1)=XX/(PR(1)^(1/2))/VP(1)
3130 DIA(2)=XX/(PR(2)^(1/2))/VP(2)
3140 GOTO 3170
3150 DIA(1)=.5122*(T2^.5236)*(SP^.4764)/(VP(1)^.0953)
3160 DIA(2)=.5122*(T2^.5236)*(SP^.4764)/(VP(2)^.0953)
3170 MASS(1)=PI*PR(1)*DIA(1)^3/6
3180 MASS(2)=PI*PR(2)*DIA(2)^3/6
3190 XX=MASS(1)^(-1.213)/(10^LO)
3200 FLUX=XX*GE*SH
3210 PROB(1)=EXP(-1*FLUX*AR(1)*TE)
3220 IF AL=500 GOTO 3260
3230 IF DIA(2)>1 THEN YY=-1.13
3240 IF DIA(2)<=1 THEN YY=-2.42
3250 FLUX=DIA(2)^YY/(10^5.82)\ GOTO 3290
3260 IF DIA(2)>1 THEN YY=-.9
3270 IF DIA(2)<=1 THEN YY=-2.52
3280 FLUX=DIA(2)^YY/(10^5.46)
3290 PROB(2)=EXP(-1*FLUX*AR(2)*TE)
3300 IF IA=1 GOTO 3420
3310 IF K>25 GOTO 3340
3320 IF (PROB(1)*PROB(2))>OP GOTO 3360
3330 GOTO 3080
3340 PRINT \ PRINT \ PRINT '***** No optimization as thickness is getting'
3350 PRINT '      very large or very small.'\ GOTO 3830
3360 IF ABS(INC)=.001 GOTO 3400
3370 T2=T2-INC
3380 INC=INC/10
3390 GOTO 3070
3400 IF IO=1 THEN T1=.32*T2
3410 IF IO=2 THEN T1=.35*T2/.65
3420 T1D1=T1/DIA(1)
3430 T1D2=T1/DIA(2)
3440 T2D1=T2/DIA(1)
3450 T2D2=T2/DIA(2)
3460 TTD1=T1D1+T2D1
3470 TTD2=T1D2+T2D2
3480 SPD1=SP/DIA(1)
3490 SPD2=SP/DIA(2)
3500 TT=T1+T2
```

```
3510 T1TT=T1/TT
3520 T2TT=T2/TT
3530 PRINT \ PRINT
3540 IF IA=0 GOTO 3600
3550 PRINT '*****  For bumper thickness=';T1;'cm.,'
3560 PRINT '             hull thickness=';T2;'cm., and'
3570 PRINT '       spacing between walls=';SP;'cm.,'
3580 PRINT '       following are comparisons between meteoroid and debris.'
3590 GOTO 3650
3600 PRINT '***** For overall probability of';OP;', the required'
3610 PRINT '       minimum bumper thickness is';T1;'cm.,'
3620 PRINT '       hull thickness is';T2;'cm, with spacing of';SP;'cm.'
3630 PRINT '       The comparisons between meteoroid and debris'
3640 PRINT '       are as follows.'
3650 PRINT \ PRINT
3660 PRINT TAB(15%);'For Debris',TAB(30%);'For Meteoroid'\ PRINT
3670 PRINT 'Velocity',VP(2);TAB(10%),VP(1)
3680 PRINT 'Diameter',DIA(2);TAB(10%),DIA(1)
3690 PRINT 'Mass',MASS(2);TAB(10%),MASS(1)
3700 PRINT 'Probability',PROB(2);TAB(10%),PROB(1)
3710 PRINT 'Bumper/Dia.',T1D2;TAB(10%),T1D1
3720 PRINT 'Hull/Dia.',T2D2;TAB(10%),T2D1
3730 PRINT 'Total T/Dia.',TTD2;TAB(10%),TTD1
3740 PRINT 'Spacing/Dia.',SPD2;TAB(10%),SPD1
3750 PRINT \ PRINT '          Bumper/Total T=';T1TT
3760         PRINT '            Hull/Total T=';T2TT
3770         PRINT '    Overall Probability=';PROB(1)*PROB(2)
3780 STOP
3790 PRINT \ PRINT \ IF DIA(2)>4 THEN PRINT '*****IMPORTANT NOTE'\ PRINT
3800 IF DIA(2)>4 THEN PRINT 'Computed probability for debris is incorrect'
3810 IF DIA(2)>4 THEN PRINT 'as the diameter of projectile is greater'
3820 IF DIA(2)>4 THEN PRINT 'than 4 cm.'
3830 PRINT \ PRINT
3840 IF IA=0 GOTO 3930
3850 PRINT "Enter 'repeat' if analysis for different thicknesses and"
3860 PRINT "spacing is desired. "
3870 PRINT "Otherwise enter 'start' to start all over or 'end' to stop."
3880 INPUT CC$\ PRINT
3890 IF CC$='repeat' GOTO 2700
3900 IF CC$='start' GOTO 120
3910 IF CC$='end' GOTO 4900
3920 PRINT '*****Input Error'\ PRINT \ GOTO 3850
3930 PRINT "Enter 'repeat' if you want to optimize the double wall design"
3940 PRINT "for different overall probability,"
3950 PRINT "or enter 'start' to start all over,"
3960 PRINT "or enter 'end' to stop."
3970 INPUT EE$\ PRINT
3980 IF EE$='repeat' GOTO 2710
3990 IF EE$='start' GOTO 120
4000 IF EE$='end' GOTO 4900
```

```
4010 PRINT '*****Input Error'\ PRINT \ GOTO 3930
4020 PRINT 'For the design data to be used,'
4030 PRINT "enter 'debris' or 'meteoroid'."
4040 INPUT BB$\ PRINT
4050 I=0
4060 IF BB$='meteoroid' THEN I=1
4070 IF BB$='debris' THEN I=2
4080 IF I>0 GOTO 4100
4090 PRINT '*****Input Error'\ PRINT \ GOTO 4020
4100 IF I=1 THEN J=2
4110 IF I=2 THEN J=1
4120 PRINT 'Enter spacing between walls in cm.'
4130 INPUT SP\ PRINT
4140 PRINT 'Enter bumper/hull thickness ratio.'
4150 INPUT T1T2\ PRINT
4160 PRINT "Enter 'Cour-Palais' or 'Nysmith' for design method to be used."
4170 INPUT CC$\ PRINT
4180 IF CC$='Cour-Palais' GOTO 4210
4190 IF CC$='Nysmith' GOTO 4310
4200 PRINT '*****Input Error'\ PRINT \ GOTO 4160
4210 PRINT 'Enter bumper material density in gm/cu. cm.'
4220 INPUT RB\ PRINT
4230 PRINT 'Enter 0.2 percent yield stress for hull material in PSI.'
4240 INPUT FY\ PRINT
4250 ZZ=.0443*(PR(I)^.5)*(RB^(1/6))*VP(I)*((70000/FY)^.5)
4260 T2=ZZ*PD(I)/(SP^.5)
4270 XX=T2*(SP^(1/2))/(RB^(1/6))/((70000/FY)^(1/2))/.0443
4280 DIA(1)=XX/(PR(1)^(1/2))/VP(1)
4290 DIA(2)=XX/(PR(2)^(1/2))/VP(2)
4300 GOTO 4350
4310 ZZ=5.08*(VP(I)^.278)*(PD(I)^2.918)/(T1T2^.528)
4320 T2=(ZZ/(SP^1.39))^(1/1.528)
4330 DIA(1)=.5122*(T2^.5236)*(SP^.4764)/(VP(1)^.0953)
4340 DIA(2)=.5122*(T2^.5236)*(SP^.4764)/(VP(2)^.0953)
4350 MASS(1)=PI*PR(1)*DIA(1)^3/6
4360 MASS(2)=PI*PR(2)*DIA(2)^3/6
4370 XX=MASS(1)^(-1.213)/(10^LO)
4380 FLUX=XX*GE*SH
4390 PROB(1)=EXP(-1*FLUX*AR(1)*TE)
4400 IF AL=500 GOTO 4440
4410 IF DIA(2)>1 THEN YY=-1.13
4420 IF DIA(2)<=1 THEN YY=-2.42
4430 FLUX=DIA(2)^YY/(10^5.82)\ GOTO 4470
4440 IF DIA(2)>1 THEN YY=-.9
4450 IF DIA(2)<=1 THEN YY=-2.52
4460 FLUX=DIA(2)^YY/(10^5.46)
4470 PROB(2)=EXP(-1*FLUX*AR(2)*TE)
4480 T1=T1T2*T2
4490 T1D1=T1/DIA(1)
4500 T1D2=T1/DIA(2)
```

```
4510 T2D1=T2/DIA(1)
4520 T2D2=T2/DIA(2)
4530 TTD1=T1D1+T2D1
4540 TTD2=T1D2+T2D2
4550 SPD1=SP/DIA(1)
4560 SPD2=SP/DIA(2)
4570 PRINT \ PRINT
4580 PRINT '***** For bumper thickness=';T1;'cm. and hull thickness=';T2;'cm.,
4590 PRINT '      designed for ';BB$;' protection,'
4600 PRINT '      following are the comparisons between'
4610 PRINT '      debris and meteoroid.'
4620 PRINT \ PRINT
4630 PRINT TAB(15%);'For Debris',TAB(30%);'For Meteoroid'\ PRINT
4640 PRINT 'Velocity',VP(2);TAB(10%),VP(1)
4650 PRINT 'Diameter',DIA(2);TAB(10%),DIA(1)
4660 PRINT 'Mass',MASS(2);TAB(10%),MASS(1)
4670 PRINT 'Probability',PROB(2);TAB(10%),PROB(1)
4680 PRINT \ PRINT 'Bumper/Dia.',T1D2;TAB(10%),T1D1
4690 PRINT 'Hull/Dia.',T2D2;TAB(10%),T2D1
4700 PRINT 'Total T/Dia.',TTD2;TAB(10%),TTD1
4710 PRINT 'Spacing/Dia.',SPD2;TAB(10%),SPD1
4720 STOP
4730 PRINT \ PRINT \ IF DIA(2)>4 THEN PRINT '*****IMPORTANT NOTE'\ PRINT
4740 IF DIA(2)>4 THEN PRINT 'Computed probability for debris is incorrect'
4750 IF DIA(2)>4 THEN PRINT 'as the diameter of projectile is greater'
4760 IF DIA(2)>4 THEN PRINT 'than 4 cm.'
4770 PRINT \ PRINT
4780 PRINT "*****Enter 'repeat' if you want the double wall design"
4790 IF I=1 THEN PRINT "for debris or for different velocities etc.,"
4800 IF I=2 THEN PRINT "for meteoroid or for different velocities etc.,"
4810 PRINT "or enter 'start' to start all over,"
4820 PRINT "or enter 'next' to continue,"
4830 PRINT "or enter 'end' to stop."
4840 INPUT DD$\ PRINT
4850 IF DD$='repeat' GOTO 4020
4860 IF DD$='next' GOTO 4900
4870 IF DD$='start' GOTO 120
4880 IF DD$='end' GOTO 4900
4890 PRINT '*****Input Error'\ PRINT \ GOTO 4780
4900 END
```

NASA/ASEE SUMMER FACULTY RESEARCH FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

DEVELOPMENT OF COMPOSITE MATERIAL TEST METHODOLOGY FOR FRACTURE TOUGHNESS/DAMAGE TOLERANCE

Prepared By:	Mark V. Bower, Ph.D.
Academic Rank:	Assistant Professor
University and Department:	The University of Alabama in Huntsville, Department of Mechanical Engineering
NASA/MSFC:	
Division:	Non-Metallic Materials
Branch:	Polymers & Composites
MSFC Counterpart:	W. J. Patterson, Ph.D.
Date:	August 22, 1985
Contract No.:	NGT 01-008-021 The Univeristy of Alabama in Huntsville

DEVELOPMENT OF COMPOSITE MATERIAL TEST METHODOLOGY FOR FRACTURE TOUGHNESS/DAMAGE TOLERANCE

BY

Mark V. Bower
Assistant Professor of Mechanical Engineering
The University of Alabama in Huntsville
Huntsville, Alabama

ABSTRACT

Fracture toughness testing techniques for composite materials are reviewed for simplicity, applicability to filament wound systems, and correlation with fracture mechanics principles. Toughness test results are presented for homogeneous epoxy and graphite reinforced epoxy.

ACKNOWLEDGEMENT

I would like to express my appreciation to all the directors and personnel of the NASA/ASEE SFFP at MSFC for their efforts to produce an effective summer of research. I particularly want to thank my NASA counterparts: Dr. W. J. Patterson, Mr. Frank Ledbetter, and Mr. Mike Clemons for their valuable time and assistance in the discussion and execution of this research.

LIST OF FIGURES

LIST OF TABLES

DEVELOPMENT OF COMPOSITE MATERIAL TEST METHODOLOGY FOR FRACTURE TOUGHNESS/DAMAGE TOLERANCE

INTRODUCTION

The Space Shuttle Orbiter is the primary vehicle in the NASA space transportation system. It will remain so for the remainder of this decade. The Orbiter can carry a crew of seven and a sixty five thousand pound load into an equatorial orbit and back to earth. The load the Orbiter can carry into a polar orbit is reduced due to the lower available energy at launch. To increase the available payload to a polar orbit NASA is using composite materials in the Solid Rocket Boosters (SRB) to reduce the launch weight of the vehicle.

Composite materials are some the most advanced materials available today. Composite materials are particularly effective in aerospace applications due to their high stiffness to weight ratio and due to the fact that they can be tailored to provide stiffness in particular directions. Composite materials are also particularly sensitive to damage. Seemingly small impacts or loads can introduce flaws into the matrix of the composite, there by reducing the overall or ultimate strength of the composite.

Fracture mechanics is the field within solid and/or structural mechanics which deals with the behavior of a solid in the presence of flaws or cracks. The early work in the field of fracture mechanics has led to the realization that the resistance of a material to particular modes of fracture is a material property. This material property is the critical stress intensity factor (K), or expressed in different terms, the critical strain energy release rate (G). This study deals with methods of determining these material properties for composite materials.

OBJECTIVES

The objectives of this study are:

(1) Review existing fracture toughness testing techniques for composite materials and metals to determine the best available methods.

(2) Determine the fracture toughness of the matrix material and the composite material used in the SRB filament wound case.

REVIEW OF FRACTURE TOUGHNESS TESTING TECHNIQUES

There are a large number of fracture toughness testing techniques presented in the literature of fracture mechanics. The review was directed primarily at methods which are used to find the mode I, or opening mode, critical stress intensity factor for composite materials and methods which use test specimens which are easily produced.

NASA Standard Tests

The ACEE Composites Project Office at NASA Langley Research Center has contributed to the development of five standardized fracture toughness tests for composite materials. These tests are designed for application to aircraft structures. The test procedures are presented in reference [1]. The tests are: (1) compression after impact test, (2) edge delamination tension test, (3) open hole tension test, (4) open hole compression test, and (5) hinged double cantilever beam test.

Compression After Impact Test

In this test a specimen is subjected to an impact load after the method of the falling weight test (described later) and then loaded in compression. The compression after impact test does not lead to the determination of a value of the critical stress intensity factor or strain energy release rate. The test produces a value of the failure stress and strain in the presence of flaws.

This test is not directly applicable to the current study because it does not lead to determination of a material property. However, the principle of the test is applicable to methods for studying the damage tolerance, or life, of composite materials. Filament wound systems are used primarily in pressure vessel applications. As such, they are usually loaded in tension. Thus, for the study of filament wound systems it is natural that the test be modified to a tension after impact test.

Edge Delamination Tension Test

The edge delamination tension test is used to determine the critical strain energy release rate for interlaminar fracture. In this test a composite laminate layup with free edges is subjected to a uniaxial tensile load. The G_c is

found from the value of strain where the stress-strain plot departs from linearity.

This test is not directly applicable to the current study because it does not lead to a mode I critical strain energy release rate. Inaddition, it is not possible to determine the mode of crack propagation. Thus, the material property found is some unknown combination of the three primary material properties, G_{Ic}, G_{IIc}, and G_{IIIc}.

Open Hole Tension Test

The open hole tension test is used to determine the ultimate tensile strength of a composite laminate in the presence of a 1/4 inch diameter hole. The test does not produce a material property. Rather, it yields an estimate of the strength of the composite in the presence of a stress riser. This test is not applicable to the present study.

Open Hole Compression Test

As with the open hole tension test this test does not lead to the determination of a material property. The result is an estimate of the compressive strength of a composite in the presence of a stress riser. This test is not applicable to the present study.

Hinged Double Cantilever Beam Test

This test is used to determine the critical strain energy release rate for delamination of a composite material. The specimen, shown in Figure 1, is loaded via hinges bonded to the composite laminate. An initial delamination is introduced by the use of a Teflon separator included in the layup of the laminate. The G_{Ic} value is found from:

$$G_{Ic} = \frac{P\delta\ (3a - 4a_o)}{2ab(a - a_o)}$$

where P is the force applied to the specimen, a is the crack length, b is the width, δ is the opening of the crack at the hinges, and a_o is a parameter found from the flexure approximation.

Note that this critical strain energy release rate is for delamination of the composite laminate. Strictly speaking this property is related to manufacturing process for the laminate and the quality of the laminate, not to the

composite itself. However, in the fracture of composites the direction of crack propagation is not what one expects based experience with fracture of homogeneous materials. For example, a crack introduced across the ply of a laminate and the laminate is loaded in tension in the ply direction, the crack propagates along the ply. Rather than across the ply as is expected with homogeneous materials. Therefore, this test with some modifications is important to this study.

This test has one very important benefit and one large detraction. The benefit is that the initial flaw or crack has a very small radius at the crack tip, as is found in nature. The detraction is that to achieve this small crack tip radius the test specimen must be layed-up explicitly for the test. It cannot be a piece cut from an existing system.

ASTM Test Methods

The American Society for Testing Methods (ASTM) has established numerous test procedures for fracture toughness testing of metals and plastics. The society has not, at this time, established test procedures for composite materials. Never the less, the procedures established for homogeneous materials are important to review for their applicability to composite materials.

The test methods prescribed by the ASTM are directed primarily at determining the mode I fracture toughness. The tests differ primarily in the geometry of the test specimen. For tension tests the geometries used are: the Compact Test specimen (ASTM E399, E813) [2,3], an arc-shaped test specimen (ASTM E399) [2], and a disk-shaped compact test specimen (ASTM E399) [2]. There are two bending test geometries: the three-point bending test (ASTM E399, E813) [2,3], and the four-point bending test (ASTM E399). An additional test prescribed by the ASTM for fracture toughness testing is the impact resistance test (ASTM D3029) [4].

ASTM Tension Tests

The specimen geometries prescribed by the ASTM for tensile fracture toughness testing are not particularly applicable to composite materials. The Compact Test specimen requires a large amount of material. The arc-shaped specimen and the disk-shaped compact test specimen are designed for application to systems with small (less than 12 inches (30.4 cm)) radii. These test geometries are not considered important to this study.

ASTM Bending Tests

In these tests a precracked specimen is loaded as a simply supported beam. The three-point bending test is shown in Figure 2. The three- and four-point bending tests differ in the loading condition at the crack. In the three-point specimen the stress at the crack tip is a combination of normal and shear stresses. In the four-point specimen the stress at the crack tip is pure normal stress. The geometry of the test specimen is the same for either method of loading. This geometry is easily produced from existing composite systems. Difficulty arises in introducing a crack with an infinitesmal radius at the crack tip. In metals testing an initial crack is introduced by fatigue. This is not possible for composite materials. Hence, an error is introduced in the procedure: the critical stress intensity factor is found for a blunt crack. This value is greater than that determined for a sharp crack. The three-point bending test is used for toughness testing in this study.

ASTM Impact Resistance Test

This test, also referred to as the falling weight test, is used to determine the energy necessary to fracture the test specimen. At this time there is no correlation between the energy to fracture the specimen and the conventional fracture toughness parameters, the critical strain energy release rate and the critical stress intensity factor. However, it is expected that the parameters are related.

In this test a weight is raised to a specified height and then released to fall on to the test specimen. If cracks are observable on the side of the specimen opposite the impact, then the weight is reduced and the test repeated on another specimen. If cracks are not observed, then the weight is increased and the test repeated on another specimen. The fracture energy is found from the average of the energies used in the tests. This test was used in this study.

Other Methods

As stated previously, there are numerous methods for fracture toughness testing described in the literature. Four of these methods were also considered in this study. They are the double cantilever beam test, the tapered double cantilever beam test, the roof-top three-point bending test specimen, and the thru-cracked plate.

Double Cantilever Beam Test

This test is similar to the hinged double cantilever beam test described earlier. Figure 3 shows a schematic of the double cantilever beam specimen. The test specimen is machined out of an existing system. As with the hinged double cantilever beam the force and opening of the crack are measured to determine the fracture toughness property of the material. This test has the advantage over the hinged double cantilever test since it can be used for an existing system. There is an additional advantage for this method, due to the length of the specimen, the crack does not propagate thru the specimen instantaneously. Instead, the crack propagates in a series of steps thru the specimen. Thus, a single specimen can yield a number of results. This test is important to this study.

Tapered Double Cantilever Beam Test

This test is nearly identical to the double cantilever beam test discussed previously. The exception is that the height of the cantilever is tapered to lead to a uniform value of the force necessary to propagate the crack. This test requires a considerable amount of material and yields results identical to those found with the double cantilever beam test. This test was used in this study.

Roof-Top Three-Point Bending Test

The test specimen used in this test has geometry identical to the specimens used in the three- and four-point bending test. The difference is the shape of the notch used. In the three-point bending test the notch used is as shown in Figure 4. The roof-top three-point bending test specimen has a notch as shown in Figure 5. This notch was originally used by Tattersall and Tappin [5] and applied to composite materials by Beaumont and Phillips [6]. This test determines the work of fracture, a parameter which is not directly related to the conventional fracture toughness parameters. There is a serious difficulty in relating the work of fracture determined by this test to the critical strain energy release rate: due to the shape of the notch the stress state at the notch is three dimensional thus eliminating the simplifying assumptions of plane strain and plane stress. These assumptions are essential to determining the strain energy release rate.

Thru-Cracked Plate Test

This test geometry was used by Griffith in his founding work in fracture mechanics. It was used by Poe and Sova [7]

to determine the fracture toughness of composite materials. In this test a plate with a crack thru the thickness of the plate is loaded axially in tension. The test lead Poe to propose a fracture toughness parameter which is independent of the laminate directions. This test was not used in this study.

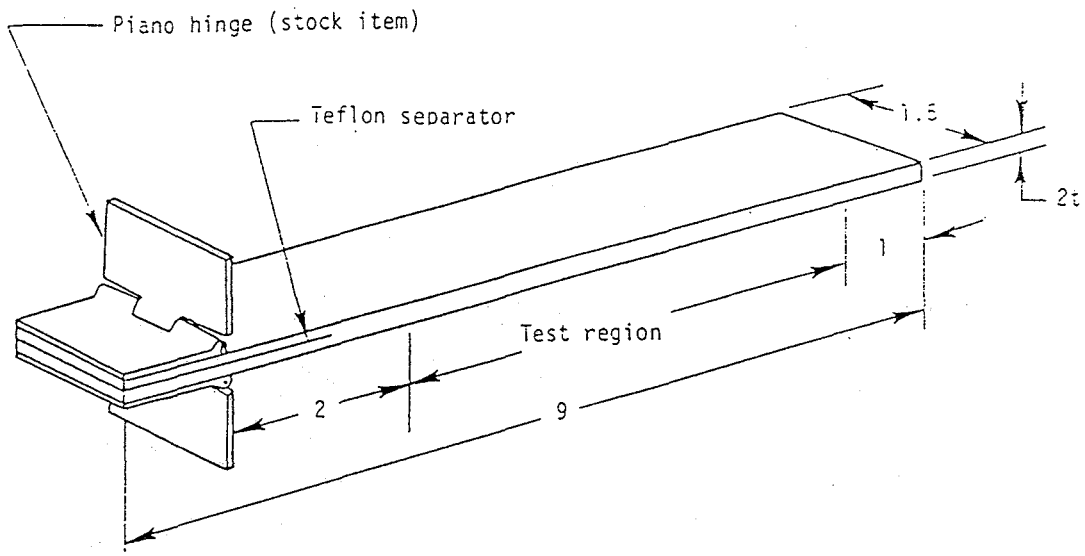


Figure 1. The Hinged Double Cantilever Beam.

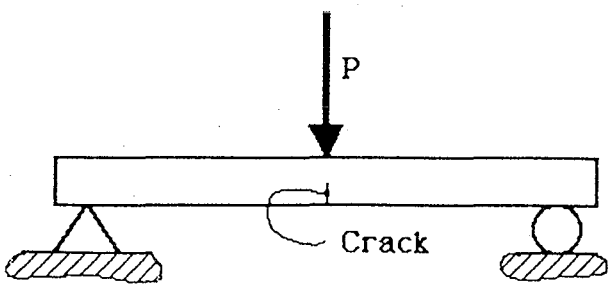


Figure 2. The Three-Point Bending Test Specimen.

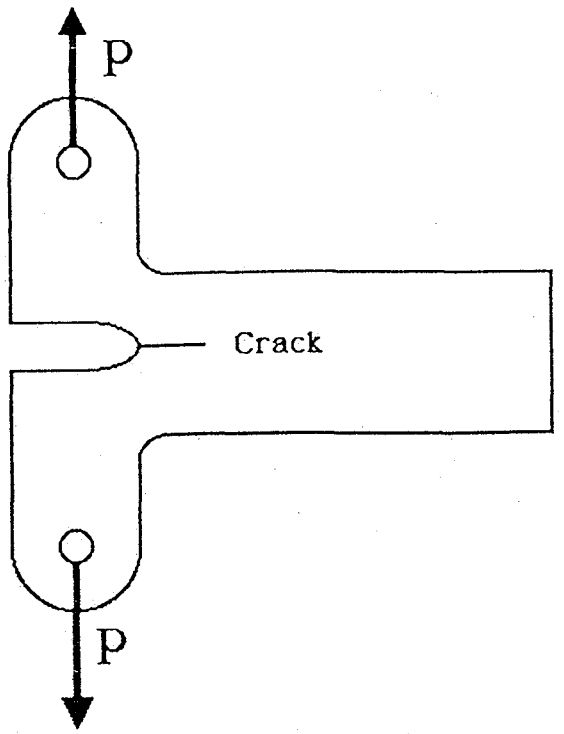


Figure 3. The Double Cantilever Beam Test Specimen.

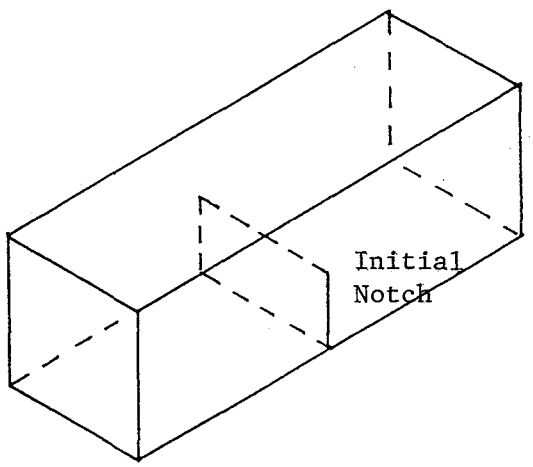


Figure 4. A Schematic of the Three-Point Bending Test specimen Showing the initial notch geometry.

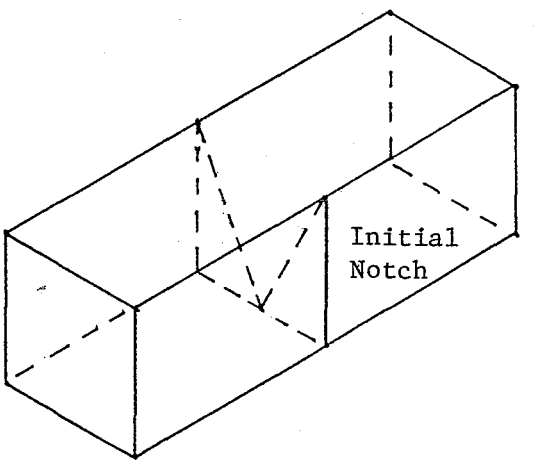


Figure 5. A Schematic of the Roof-top Three-Point Bending Test specimen showing the initial notch geometry.

FRACTURE TOUGHNESS TESTING

Fracture toughness tests were performed on homogeneous samples of the SRB filament wound case (FWC) resin (Hercules Epoxy 55A) and on the FWC composite material (graphite reinforced epoxy). Several of the different fracture toughness tests discussed above were used.

Neat Resin Tests

For testing of the "neat" or homogeneous resin the following tests were used: the three-point bending test, the double cantilever beam test, the tapered double cantilever beam test, and the roof-top three-point bending test. The thicknesses of the test specimens ranged from 0.148 inch (3.76 mm) to 0.5 inch (12.7 mm). Results are listed in Table 1 for the three groups of specimens used. Note that the average value of the stress intensity factor for the group of specimens 0.148 inch thick is significantly higher than the other averages. These pieces were cast at the thickness used in the test. Due to variations in the cure thru the thickness it was obvious that the material was not homogeneous. This tends to explain the higher values of K_{Ic} found.

In conducting the three-point bending tests the test set-up was used to measure the elastic modulus of the material. Table 2 lists the results of these measurements. The values found do not agree with the handbook values for the material. To determine if the test was in error measurements were made using the Rheometrics Dynamic Spectrometer. Figure 6 shows the results from one set of tests. These tests demonstrated that the results found from the three-point bending test were accurate. They also demonstrated quite clearly that the epoxy resin is a nonlinear viscoelastic material, with as much as a 20 percent reduction in the modulus value over 1000 sec. In order to avoid the effect of viscoelastic response on the fracture toughness tests the tests were conducted at a rate above the relaxation rate of the material.

Table 3 lists the critical values of the strain energy release rate determined from the double cantilever beam tests. As indicated in the table the corresponding values of the critical stress intensity factor are considerably above the values determined by the three-point bending tests. An explaination for this result is not readily

available. It is conceivable that the different loading rates effects the critical fracture toughness parameter.

The values of the work of fracture determined from the roof-top three-point bending tests are listed in Table 4. There is a factor of 5 difference in the values found with two different geometries. Again, no explaination is offered other than to note that work is used in the viscoelastic response of a material which other wise is not accounted for.

FWC Composite Material Tests

Tests on the FWC material were performed using the three-point bending test and the roof-top three-point bending test. These tests highlighted the fact that composite materials do not behave in the same manner as homogeneous materials. The tests did not lead to the determination of a fracture toughness parameter. This is due to the fact that the cracks did not propagate in an opening mode manner. The initial cracks either extended in an interlaminar shear mode, or sections of the material pulled out of the specimen. The later is an edge effect. This leads to the conclusion that other test procedures must be used to determine the fracture toughness of composite materials.

Table 1. Critical Stress Intensity Factors for Neat Epoxy Determined From Three-Point Bending Tests.

Specimen Thickness inch	Average K_{Ic} psi$\sqrt{in}$	Standard Deviation %
0.148	1839	11.8
0.252	1175	9.1
0.503	1601	3.9

Table 2. Elastic Modulus Values for Neat Epoxy Determined From Three-Point Bending Tests.

Specimen Thickness inch	Average E ksi	Standard Deviation %
0.252	56	2.2
0.503	52.5	4.7

Table 3. Critical Strain Energy Release Rates for Neat Epoxy Determined From Double Cantilever Tests.

Specimen Thickness inch	Average G_{Ic} lb/in	Standard Deviation %	Average K_{Ic} psi$\sqrt{in}$
0.125	9.23	13.3	2201

Table 4. The Work of Fracture for Neat Epoxy.

Specimen Thickness inch	Average Work of Fracture lb/in	Standard Deviation %
0.252	16.21	10.6
0.503	98.72	16.9

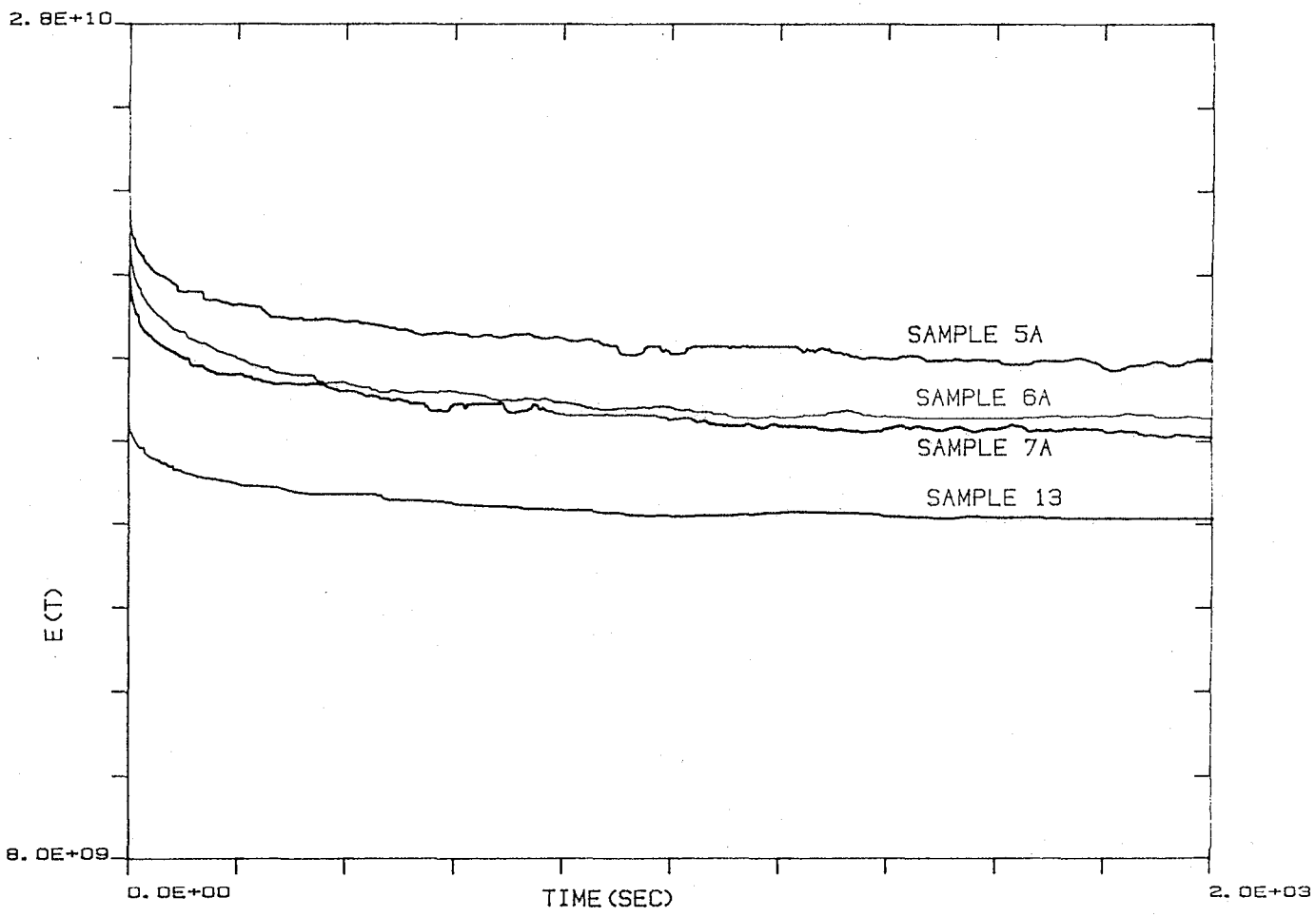


Figure 6. The Stress Relaxation Plot for Hercules 55A Epoxy Resin. The measurements were performed at 0.15% strain, loading a three-point bending specimen. E(t) is in dyne/cm^2.

CONCLUSIONS AND RECOMMENDATIONS

The conclusions of this study are:

(1) Fracture toughness parameters for homogeneous materials can be easily determined by either the three-point bending test or the double cantilever test.

(2) The three-point bending test is not effective for determining the fracture toughness of composite materials.

(3) Epoxy is a nonlinear viscoelastic material. This type of behavior may effect the fracture toughness measurements for the neat resin and for the composite material.

The recommendations of this study are:

(1) The hinged double cantilever beam test, the double cantilever beam test, and the thru-cracked plate test methods should be investigated for their applicability to composite material systems.

(2) The use of a tension after impact test to determine the damage tolerance and damage accumulation rate of a composite material should be studied.

(3) A study of the effect of viscoelastic behavior on fracture toughness and filament wound systems should be performed.

REFERENCES

1. ACEE Composites Project Office, NASA Langley Research Center, Standard Tests for Toughened Resin Composites, NASA Ref. Pub. 1092, 1983.

2. ASTM, Annual Book of ASTM Standards, Part 10, E399 - 81.

3. ASTM, Annual Book of ASTM Standards, Part 10, E813 - 81.

4. ASTM, Annual Book of ASTM Standards, Part 35, D3029 - 78.

5. Tattersall, H. G. and Tappin, G., "The Work of Fracture and its Measurement in Metals, Ceramics and other Materials", J. Mat. Sci. 1, (1966) 296-301.

6. Beaumont, P. W. R. and Phillips, D. C., "Tensile Strengths of Notched Composites", Carbon Reinforced Epoxy Systems, Vol. 1, Ed. Hilado, C. J., Technomic Pub., Westport, CT, 1974.

7. Poe, C. C. and Sova, J. A., "Fracture Toughness of Boron/Aluminum Laminates with Various Proportions of 0° and $\pm 45^\circ$ Plies", NASA TP-1707, November 1980.

1985

NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

A TECHNIQUE FOR INCORPORATING THE NASA SPACELAB PAYLOAD DEDICATED EXPERIMENT PROCESSOR SOFTWARE INTO THE SIMULATION SYSTEM FOR THE PAYLO CREW TRAINING COMPLEX

Prepared By:	Dale A. Bremmer, Ph.D.
Academic Rank:	Adjunct Associate Professor
University and Department:	Oklahoma State University Education Research
NASA/MSFC:	
Division:	Operations Development
Branch:	Man/Systems Integration
MSFC Counterpart:	Charles M. Lewis
Date:	August 5, 1985
Contract No:	NGT 01-008-021 The University of Alabama in Huntsville

A TECHNIQUE FOR INCORPORATING THE NASA SPACELAB PAYLOAD DEDICATED EXPERIMENT PROCESSOR SOFTWARE INTO THE SIMULATION SYSTEM FOR THE PAYLOAD CREW TRAINING COMPLEX

BY

DR. DALE A. BREMMER
Adjunct Associate Professor
OKLAHOMA STATE UNIVERSITY
Stillwater, Oklahoma

ABSTRACT:

The NASA space shuttle missions offer an opportunity for a principle investigator (PI) to design and operate many experiments that were impossible to conduct prior to the advent of manned space flight. With a projected shuttle schedule of two launches per month, numerous PIs will be able to take advantage of the opportunities and challenges provided by the spacelab. Many of the onboard spacelab experiments are controlled by an experiment computer. In addition, the experiment computer serves as a focal point for much of the data acquisition and transmission activities of the experiments. The result of the high demand for payload management activities and the geriatics of the antiquated technology of the spacelab's experiment computer is a limited ability to control the operation of the PI's experiment. Hence, it is often necessary to incorporate a dedicated experiment processor (DEP) in the experiment design. Shuttle mission flight experiments that gather data or require control information on an uplink from the Payload Operations Control Center (POCC) may need a (DEP) as an integral part of the flight hardware. The DEP (1) offers some capability for immediate analysis, (2) provides additional data storage and transmission capabilities, (3) allows the PI to control and/or alter the experiment operations while in flight, (4) provides for feedback control, (5) permits some real-time data analysis required during the mission. (6) simplifies the design of bulky and weighty hardware, and (7) if autonomous, allows relatively complete checkout and verification prior to integration with the onboard experiment computer.

This study proposes to assess the feasibility of some off-the-shelf microprocessors and state-of-art software: (1) as a development system for the PI in the design of the experiment model, (2) as an example of available technology application for future PI's experiments, (3) as a system capable of being interactive in the PCTC's simulation of the DEP, preferably by bringing the PI's DEP software directly into the simulation model, (4) as a system having bus compatibility with host VAX simulation computers, (5) as a system readily interfaced with mock-up panels and information displays, (6) as a functional system for post mission data analysis.

ACKNOWLEDGEMENTS:

Acknowledgements are extended to the National Aeronautics and Space Administration (NASA), the American Society of Engineering Educators (ASEE), the George C. Marshall Space Flight Center (MSFC), and in particular the MSFC Man/Systems Integration Laboratory that funded and provided the facilities needed to make the NASA/ASEE Summer Faculty Fellowship Program feasible.

Thanks are extended to Charles M. Lewis for listening to the philosophy of this project and for offering guidance in its undertaking; to Tina Melton for assisting in my understanding of the operation and scope of the Payload Crew Training Complex (PCTC) and for the techinical editing of this paper; and to Rose Ann Kafer for reading and commenting on choice of words used to make this paper more readable and for adding a little sanity at the proper time.

INTRODUCTION:

The NASA space shuttle missions offer an opportunity for a principle investigator (PI) to design and operate many experiments that were impossible to conduct prior to the advent of manned space flight. With a projected shuttle schedule of two launches per month, numerous PIs will be able to take advantage of the opportunities and challenges provided by the spacelab. Many of the onboard spacelab experiments are controlled by an experiment computer. In addition, the experiment computer serves as a focal point for much of the data acquisition and transmission activities of the experiments. The result of the high demand for payload management activities and the geriatics of the antiquated technology of the spacelab's experiment computer is a limited ability to control the operation of the PI's experiment. Hence, it is often necessary to incorporate a dedicated experiment processor (DEP) in the experiment design. Shuttle mission flight experiments that gather data or require control information on an uplink from the Payload Operations Control Center (POCC) may need a (DEP) as an integral part of the flight hardware. The DEP (1) offers some capability for immediate analysis, (2) provides additional data storage and transmission capabilities, (3) allows the PI to control and/or alter the experiment operations while in flight, (4) provides for feedback control, (5) permits some real-time data analysis required during the mission. (6) simplifies the design of bulky and weighty hardware, and (7) if autonomous, allows relatively complete checkout and verification prior to integration with the onboard experiment computer.

Each shuttle mission is an aggregation of many tasks and potentially many DEPs, each from a different PI. Once the mission's tasks have been integrated into a performance timeline, the Payload Crew Training Complex (PCTC) utilizes simulators to provide operations training for the payload and mission specialists and the ground operation teams. To produce a well trained flight crew, this simulation system must simulate faults and equipment failures, accommodate alternate mission flight profiles and model the PI's experiment operations. These software models must realistically simulate the experiment's hardware as well as the DEP and experiment computer software. This results in a complex computer programming task requiring significant data base modification with any variation in the timeline of the mission. As changes are made to the actual experiment hardware and software, the existing simulation code has to be significantly modified to accurately reflect the experiment operations.

Currently the following software simulators (models) of major flight subsystems are required to be developed for the complete simulation and training capability;

- Experiment Computer Operating Systems (ECOS) Model
  - relatively constant from mission-to-mission
- Experiment Computer Applications Software (ECAS) Model
  - special processing software which runs on the spacelab experiment computer under ECOS
  - unique to each experiment
- Experiment Hardware Model
  - simulates experiment hardware functions, feedbacks, telemetry, science data, etc.
- Experiment DEP Software (where implemented) Model
  - independent of ECOS/ECAS
  - unique to each experiment
  - interacts with the Experiment Hardware Model

As a result of the freedom offered the PIs in the configuration of their DEP software, (this is feasible because the software is the farthest removed from the interfacing spacelab systems) there are, typically, some last-minute changes to the flight DEP up to the final stages of the flight experiment integration and verification. These late changes, while often benign relative to flight software, frequently require massive changes in the corresponding DEP simulator software with consequent deleterious effects on training and simulation schedules and validity.

Neither the training function nor any other echelon should succinctly dictate the DEP utilized by the PI. However, the scientific and engineering objectives of the shuttle missions can hopefully be more near optimal in flight and be more easily simulated at the PCTC by establishing some relevant computer hardware architecture and software attributes in line with the aforementioned conditions.

It would be impractical to propose system homogeneity throughout the entire hardware and software systems utilized by the PIs, the onboard experiment computer, the Payload Crew Training Complex, the simulation computer, and any system for post mission data analysis. However, the hardware and software entities, where practical, should be relevant to the state-of-the-art, off-the-self items, compatible (on the back plane) with the PI's experiment sensors, and compatible with the host simulation computers via buses or communication protocols.

Attempts to improve the incorporation of the DEP model, hardware and software, into the PCTC simulation models should encompass the PI's modeling and development techinques and the hardware and software utilized in the experiment model. Thus, there is a need to study some currently available hardware and software relevant to the DEP's function, the shuttle mission objectives, and the PI's freedom to alter the DEP software as required by the experiment, and the inherent limitations in the PCTC simulation.

It would be advantageous to the PI and the mission objectives, if systems employed by the PI during experiment development were also interactive with the PCTC simulation computers during the training of payload and mission specialists. There would be a reduction the setup time and mock-up initializations if the PI's development system were interactive and bus compatible with processors used in conjunction with mock-up panels and displays. There would be additional benefits if the same system could be used in any post mission data analysis.

If the type of development hardware and software used by the PI is also functional in the programming and the manipulation of the simulation model by the PCTC, then there will be much higher integrality between the PI's experiment model and the PCTC's simulation activities. If the PI's software is in an appropriate language, such a C-language, then incorporating the DEP into the simulation model may be little more than doing a cross-assembly of the PI's software under the operating system of the PCTC's host simulation computers.

This study proposes to assess the feasibility of some off-the-shelf microprocessors and state-of-art software: (1) as a development system for the PI in the design of the experiment model, (2) as an example of available technology application for future PI's experiments, (3) as a system capable of being interactive in the PCTC's simulation of the DEP, preferably by bringing the PI's DEP software directly into the simulation model, (4) as a system having bus compatibility with host VAX simulation computers, (5) as a system readily interfaced with mock-up panels and information displays, (6) as a functional system for post mission data analysis.

STATEMENT OF THE PROBLEM:

The PI needs a DEP (hardware and software) that is easily obtained, somewhat universal in its application, readily adaptable and reliable to the experiment. It should also be functional and/or compatible with computers used for post mission data analysis. It is important that the PI be able to make changes to the DEP software late into the pre-flight period. The hardware and softwarte support systems, used in the development of the PI's experiment, should have a high level of technology integrity and integrality with the DEP, with the simulation computers, and with computers used in post mission analysis. Additionally, the configuration should support commonality between mission objectives.

The PCTC needs both a DEP hardware and software model that can be readily incorporated into the simulator as a module that is compatible with the host VAX computer. This DEP software should be easily transportable or readily assembled under the PCTC simulator environment. Thus, ideally the hardware and software support system used by the PI should be functionally compatible, in an on-line mode, with the simulation host computer.

Efforts, to assist in the above problems, are intended to lead toward a more efficient experiment design for the PI and better use of hardware and software expenditures. The efforts should produce a simpler, more efficient, and more relevant simulation of the mission in the PCTC. The efforts will detail a current technological example for future PIs and study systems that will also assist the PI in post mission data analysis.

OBJECTIVES OF THE PROPOSAL:

The objectives of this proposal are:

(1) To study the INTEL Model 310 series, utilizing a 80286 miroprocessor chip and C-language operating under a RMX-86 operating system (Such a system would serve as a functional DEP development system capable of assembling the necessary software for the PI's DEP and converting or transferring it into VAX simulation compatible code for use in the PCTC.)

(2) To study the feasibility of interfacing the Intel Model 310 series MULTIBUS to a VAX 785 UNIBUS via a commerically available plug-in, self-contained boards

(3) To assess the functionality of the Intel Model 310 as a system to be an interactive system in the PCTC with host VAX simulation computers and the transportability of the DEP software model into the VAX computer via cross-assembly techniques

(4) To assess the value of using the Model 310, and the DEP, or DEP compatible processors as a tool for post mission data analysis

(5) To assess the feasibility of using a 80286 single board computer (compatible with the Model 310, the DEP, and the simulation computer) as an intelligent package in a behind-the-panel arrangement within the training mock-up, functioning as an interface to the experiment control and display panels

(6) To study the feasibility of using the Model 310 in a local area network or in a bit bus communication arrangement to initialize and manage 80286 single boards in the mock-up panels during simulation operation

GOALS OF THE OBJECTIVES:

Goals to be reached in this study are:

(1) To employ the 80286 microprocessor (as a Intel Model 310 series) to serve as a development tool for the PI in the development of his software for the flight experiment

(2) To demonstrate the compatibility and ease of using a 80286 microprocessor as a DEP with the Intel Model 310 and VAX computers and associated software as support systems

(3) To use the C-language in the design and programming of a DEP

(4) To use the Intel Model 310 remotely, at the PI's parent facility, in the experiment design and the DEP software development

(5) To use the Intel Model 310 as an interactive system in the PCTC simulation operations

(6) To interface the MULTIBUS of the Intel Model 310 with the UNIBUS of the VAX simulation computers

(7) To transport the DEP software, developed under C-language, into the VAX simulation computers via a cross-assembly program operating under the VAX operating system

(8) To use the Intel Model 310 series, located at a PI's parent facility, as a post mission data analysis support computer

(9) To set precedents for future PIs designing an experiment model and developing the associated DEP hardware and software systems.

(10) To simplify and standardize development of simulator control and display panels used in the PCTC by employing DEP compatible processors behind the mock-up panels and using networking techniques to link the various panels to the host system (This would additionally reduce the current panel cabling scheme to a single co-axial or optical fiber cable.)

HARDWARE & SOFTWARE:

Two Intel 286/310-17 Real-time Starter Kits (a $42,000 market value) a 80286 microprocessor based system with an 80287 80 bit numeric processor, 1/2 MByte RAM, 320 KByte Floppy, 19 MByte Winchester, RMX operating system, Assembler 86, Link 86, Locate 86, PL/M 86, AEDIT full screen editor, PSCOPE (debugger), FORTRAN.....	$9,995.00	
C-86 Compiler.....	300.00	
C-86 Cross Compiler for VAX.....	300.00	
Interface boards (MTC Model 86/11W, Mesa Technology Corp.) for Intel 310 to VAX 785	6,000.00	
HARDWARE & SOFTWARE TOTAL.....		$16,595.00

PERSONNEL SUPPORT:

On site (Huntsville) housing.(Nov 2 to Feb 3).....	$1,000.00	
Travel expenses.....	500.00	
SUPPORT TOTAL.....		$1,500.00
TOTAL.....		$18,095.00

NOTE::

THE ABOVE IS ONLY AN ESTIMATE AND DOES NOT INCLUDE ANY OVERHEAD COSTS.

ANY COMPUTER OPERATING COST RELATIVE TO TEST OR CHECK OUT ARE NOT INCLUDED.

TIMELINE OF THE PROPOSED PROJECT:

July 20, 1985 to August 10, 1985, - on going.
November 2, 1985 to February 3, 1986.
May 10, 1985 to August 30, 1985 - (Under NASA/ASEE support)

ACTIVITIES LIST:

List of activities for the research project under this proposal to the MAN/SYSTEMS LABORATORY.

A. Phase I - Early Activities:

(1) Identify the PI's for the laboratory missions:

MISSION	LAUNCH DATE
ASTRO-1 61-E	3/6/85
EOM-1/2 61-K	9/3/86
ASTRO-2 71-B	10/30/86
IML-1 71-K	5/7/87
ASTRO-3 71-N	7/15/87
SUNLAB-1 71-Q	9/21/87
EOM-3 81-C	11/23/87

(2) Identify the type of processor the PI will use for any DEPs designed for the above missions

(3) Identify how many present, past, and future PI's are using an Intel system in the development of the DEP flight software or in the development of their original experiment model.

(4) Identify the software employed and, if used, the host mainframe systems.

(5) Determine the experiments that were, and those that could have been, coded in C-Language.

(6) Determine the limitations, if any, of using C-Language in the DEPs employed in the missions.

(7) If a PI's DEP is other than an Intel and uses (or could use) C-Language, determine what is the possibility of doing a cross-assembly of the DEP flight software under the PCTC VAX simulation computers.

(8) Identify those experiments suitable for study in this proposal.

(9) Investigate the feasibility of the PI doing his DEP modeling using an Intel Series 310-17 system employing iRMX™ 86 Real Time Operating System.

(10) Investigate the possibility of rolling the PI's DEP into the simulation model via a cross-assembler under VAX control.

(11) Investigate the possiblility of doing the above (10) for DEPs not compatible with C-Language.

(12) Investigate the feasibility of, and identify the constraints against the provisions of fault insertion techniques (for simulations and crew training) despite the use of flight configuration DEP software.

B. Phase 2 - Intermediate activities:

(1) Obtain two Intel Model 310-17 systems with 80286 micoprocessors.

(2) Construct one model relevant to the selected mission DEP.

(3) Transport the constructed DEP model into the PCTC simulation.

(4) Obtain a hardware board, bus coupler, to make the Intel Model 310's MULTIBUS function in a compatible bus extension mode with the UNIBUS of the VAX 785 of the PCTC.

C. Phase 3 - Follow-up activities:

(1) Demonstrate that the payload and mission specialists can undertake crew traing at a remote site as well as at the PCTC.

(2) Demonstrate that the 80286 system with the C-Language forms a viable structure for the flight DEP, the original modeling of the DEP, and for post mission data analysis.

(3) Verify the existing techniques of accessing NASA data banks, after the mission, for PI's post mission data analysis.

(4) Demonstrate the feasibility of using the 80286 microprocessor as an interactive interface with the simulator mock-up panels and information displays

(5) Demonstrate the feasibility of using a video disk player under a DEP control to generate the fault display scenes for simulation training activities.

(6) Demonstrate incorporating the DEP behind the panels of the mock-up and communicating to the PCTC via an optical fiber, coaxial cable, or an appropriate high speed bit bus using software controlled ports.

(7) Demonstrate using the Intel Model 310 system to load and control the processors behind the panels of the mock-up.

TIMELINE - SCHEDULING CHART:

AUG 85 NOV 85-FEB 86 MAY 86-SEPT 86

PHASE 1

PHASE 2

PHASE 3

REFERENCES:

1. Intel Corporation, C-86 Compiler User's Guide, Intel Corporation, Santa Clara, California, 1983.

2. Intel Corporation, iRMX™ 86 Programming Techniques, Intel Corporation, Santa Clara, California, 1983.

3. Plum, Thomas, Learning to Program in C, Prentice-Hall Englewood Cliffs, New Jersey, 1983.

These references are not cited in this paper. These references were supportive in the development of the philosophy implied and the hardware and software denoted.

1985

NASA/ASEE SUMMER FACULTY RESEARCH FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

INVESTIGATION OF POTENTIAL DRIVER MODULES AND TRANSMISSION LINES FOR A HIGH FREQUENCY POWER SYSTEM ON THE SPACE STATION

Prepared by:	Harold T. Brush
Academic Rank:	Assistant Professor
University and Department:	Tuskegee University Department of Electrical Engineering
NASA/MSFC:	
Laboratory:	Information and Electronic Systems
Division:	Electrical
Branch:	Electrical Power
MSFC Counterpart:	J.R. Lanier,Jr.
Date:	August 2, 1985
Contract No:	NGT 01-008-021

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to NASA/ASEE for providing me the opportunity to participate in the Summer Faculty Fellowship Program at Marshall Space Flight Center. My first summer at MSFC has been very fulfilling and intellectually stimulating, due to the generous hospitality and assistance of my NASA counterpart, Mr. John R. Lanier and his team. Special thanks goes to Mr. David Weeks for his valuable information and suggestions. Discussions with Mr. Robert Kapustka were very valuable and helpful. I learned much from these discussions and I am very appreciative to him for taking time out to share with me his expertise.

Finally, I wish to thank Mrs. Shirley Butler for typing and correcting this report.

LIST OF TABLES

ABSTRACT

The objective of this investigation is to assess the feasibility of using the Series Resonant inverter as the driver module for the high frequency power system on the Space Station. This study evaluates the performance of the Series Resonant driver when it was operated with a DC input voltage and run through a series of tests to determine its start-up performance, response to load changes, load regulation, and efficiency. Also, this study compares the Series Resonant driver to another kind of driver that uses a Power Transistor snubber.

An investigation of the various types of transmission lines is initiated. In particular, a simplified approach is utilized to describe the optimal transmission line.

INTRODUCTION

The original Space Station, Skylab, didn't require power at the levels of hundreds of kilowatts. In fact, Skylab used an 8 kilowatt bus for power distribution. In contrast, the Initial Operational Configuration (IOC) Space Station will require 75 kilowatts for primary power distribution while the growth Space Station will require 300 kilowatts. The spacecraft being designed for new NASA missions are projected to have demands for power at orders of magnitude greater than current Spacecraft. The evolving spacecraft power systems for these missions will require increased efficiency and versatility to meet load requirements of greater power, multiple users and increased life. Therefore, Space Station must have the ability to be expanded to cover expected power needs of hundreds of kilowatts.

NASA/Lewis Research Center has proposed to use a high-frequency, high-voltage AC power system to fulfill the needs of the Space Station. LeRC, the NASA center responsible for primary power generation and conditioning on the Space Station proposes to generate the AC power using a resonant AC power system. Furthermore, they are proposing a primary power distribution of 440V AC RMS single phase at 20 kilohertz, and part of this is to be distributed in the common modules.

Marshall Space Flight Center who is responsible for distributing power in the common modules have expressed deep concerns about this AC power system efficiency, transmission reliability, and corona effects inside the module. They have a special concern about using the series resonant inverter as the driver for the power system, and feel that this matter should be resolved before the final decision is made to build the system.

The resonant inverter was tested in a resonant AC power system test program that was developed by General Dynamics, a contractor for LeRC. The test results are utilized to evaluate the performance of the resonant inverter in a resonant AC power system.

1.0 SYSTEM TEST CONFIGURATION

1.1 AC System Operation

The basic circuit for the inverter is a simple series resonant circuit as shown in Figure 1-1a. The operation is as follows. When switch S1 is closed, the resonant circuit is excited and "rings" at the natural resonant frequency of the circuit determined by the values of L, R, and C. The voltage across the load will appear as in 1b. The load resistance dissipates energy from the resonant circuit and causes the voltage and current waveforms of the circuit to be damped. If the resonant circuit is now excited from a pair of opposite polarity sources through a pair of toggled switches operating at natural frequency (Figure 2a), a sustained AC wave can be developed as shown in Figure 2b.

This circuit can be implemented in the usual bridge arrangement depicted in Figure 3. Alternatively, the load can be placed in parallel with the resonant capacitor as proposed by Neville Mapham (Figure 4).

The current in the resonant circuit of the inverter is primarily determined by the L and C of the tank circuit. Therefore, the Mapham type inverter is a voltage source device because the load is in parallel with the resonant capacitor. The Mapham inverter is a voltage source but becomes overdamped and begins to operate irregularly for heavy loads. Because the AC breadboard is a utility-type power system the Mapham configuration was chosen as the driver module.

If the resonant frequency of the inverter is increased above the switching frequency then Silicon Controlled Rectifiers (SCRs) can be used as switches. Replacing the load resistor of Figure 4 with a transformer allows the inverter to be used as the driver to a high-voltage bus capable of providing power to a variety of loads. Larger power systems are possible by combining multiple inverter modules.

To test the Mapham topology, a resonant AC series inverter was operated with a DC input voltage and ran through a series of test to determine its start-up performance, response to load changes, load regulation and efficiency. The test results are discussed in the following sections.

1.2 Single Inverter (Configuration)

The test configuration was a 1.0 kw inverter module with a resistive load (Figure 5). The inverter schematic is in Figure 6. This module was operated with a 90-volt input and run through a series of tests to determine its start-up performance, response to loadchanges, loadregulation, and efficiency.

2.0 TEST RESULTS

2.1 Start Up of a Single Inverter

The test circuit for the single inverter is shown in Figure 7 and the inverter schematic in Figure 6. For this test configuration, the SCR firing signals were enabled prior to the closing of the relay. The oscilloscope was triggered as the relay coil was energized. The inverter was able to start up in this manner under all load situations except full load (1.1kw). When the 1.0kw inverter was step-started under full load conditions (1140w), the resonant circuit responds as if it were overdamped. This caused one pair of SCRs to be triggered while the current in the other pair of SCRs has not yet returned to zero (See Figure 6). The result was a shorting of the input.

In brief, the test results indicated that a single inverter in a resonant power system configuration exhibited no problems starting up with a step application of voltage while under no load to 50% resistive-load conditions.

2.2 Steady-State Operation of a Single Inverter Module

A schematic of the 1.0kw-inverter module is shown in Figure 6. The currents and voltages referred to in this section are labeled in the figure. Figure 8 shows the leg current of the inverter and its output voltage for the 0.0kw case and the 1.1kw case. The positive portion of the leg current is the current in the SCR and the negative portion is the diode current (reverse current).

As shown in Figure 8, it is apparent that the inverter output voltage amplitude does not change much with load. In fact, the graph of the output voltage versus the load power (Figure 9) indicates that the inverter is a stiff voltage

source. Over the range of 0.0 to 1.14kw, the output voltage of the inverter decreased from 124V RMS to 117V RMS. This yields a load regulation of 6% for the unregulated inverter. At 1220w, the voltage was still at 114V RMS.

It is also apparent in Figure 8 that high frequency oscillations are present on the leg currents, and they changed with load.

As expected, since the triggering of the (SCRs) was controlled by a clock, the frequency of the inverter output changed very little with load. It was measured at 20.01 KHz while unloaded and 20.00 KHz with a load of 1220w. This was not the case for efficiency (Figure 10). With a 90 Vdc input the inverter was most efficient (96.8%) under-full load conditions. In conclusion, the single phase power system breadboard test demon stated that the resonant inverter was efficient and versatile as a system building block. The 1.0kw inverter breadboard was 96.9% efficient. The inverter supplied power over long distances (50 meters) to active load modules.

The development of inverters must also be continued. Since the majority of the power loss in an inverter was attributable to the SCRs, alternatives must be explored.

2.3 Transient Load Response of a Single Inverter

The inverter circuit is shown in Figure 6 and the circuit used to abruptly change the inverter load is shown in Figure 11. Three load changes were tested:

a. 0.0w to 580w and reverse
b. 580w to 1110w and reverse
c. 130w to 1110w and reverse

The most dramatic power change took place in the 130w to 1110-w case. Yet, the inverter experienced only a short and smooth transition period as shown in Figure 12 and 13, which show the inverter output voltage and current for the 130 to 1110-w and 1110 to 130-w case respectively. Discounting the switch bounce in Figure 12 these figures showed that the overshoot of the inverter was small and the entire transient response lasted only about 150 microseconds. The other inverter parameters such as

the leg current (Figure 14) also showed a smooth and brief transition for abrupt load changes. The transient response characteristics in the other two cases (0.0w to 580w and reverse and 580w to 1110w and reverse) lasted for a shorter amount of time because the load variation was not as great.

2.4 Power Supply Sensitivity of a Single Inverter

The test circuit of Figure 15 was used togather both the transient and steady-state data on this configuration. The steady-state measurements of the power sensitivity are listed Table 1. Since the inverter is unregulated, its output voltage changes the same percentage as the input voltage, and the sensitivity is 1.0 as expected. In Table 1, the efficiency of the driver appears low because the inverter is only about 90% loaded.

In the transient case, the relays switch fast but when the input voltage was decreased, the settling time was determined by the time constant of the input capacitor (700 microforads) and the load resistance. Indeed there was a period of about 0.3 milliseconds when neither relay of Figure 15 was closed, and the input capacitor supplies all the power to the load. Figure 16 shows the input current to the input capacitor going to zero for 0.3 milliseconds.

In brief, the test results indicated that every parameter in the system showed a smooth transition. The shift to a higher input voltage was also a smooth transition, but its time constant was much longer and was determined by the response time of the power supplies.

2.5 Power Turn Off a Single Inverter

The relay used to remove the power from the inverter was placed after the input capacitor as shown in Figure 17. After the relay was opened, the input voltage became a square wave because there was no input capacitor to hold the input voltage constant. This voltage alternated between the resonant capacitor peak voltage (initially twice the input voltage) and zero at 40KHz as shown in the no-load case of Figure 18. The voltage decreased exponentially until the energy of the tank circuit was

dissipated. The time constant of this decay became shorter as the inverter was loaded as shown in Figure 19, which shows the inverter output voltage for the no-load and the 580-w case. The remainder of the inverter parameters showed this same exponential decay.

In brief, the test results indicated that the output voltage of the inverter merely decayed to zero with a time constant based on the filter components and the load.

2.6 Conclusion

The testing of the single-phase, 2.0-kw resonant AC power system breadboard demonstrated that the resonant inverter was efficient and versatile as a system building block, but it needs to be developed more to get rid of the power losses attributable to the SCRs. Furthermore, the inverter must be more efficient for the AC power system on the Space Station, because it will have to drive high frequency AC power down the boom to the common modules. Therefore, Marshall Space Flight Center, who is responsible for development of power in the common modules, is concerned about using this resonant inverter as the driver for the AC power system.

3.0 MSFC CONCERNS

3.1 Propagation of Fault to Series Resonant Inverter

If a fault in the system propagates back to the SRI the bus collapses and the system goes dead. MSFC feels that a utility-type power system should be built to provide power to remove the fault.

3.2 Efficiency at Partial Loading Due to Redundancy Schemes

LeRC efficiency projections involving SRI are based on full loading. MSFC feels that the projections should be based on 50% loading.

3.3 Corona Effect

For the Space Station LeRC is proposing a primary power distribution of 440Vrms single phase at 20 kilohertzs. MSFC, who is responsible for the development of the electrical power system in the common modules is concerned about the power distribution in the modules, because the modules are suceptible to corona at high voltage and twenty kilohertz frequency. Therefore, they are planning to reduce the voltage outside the common modules by using a step down transformer.

3.4 The AC Power System Efficiency Dependence

The test results from the AC power system test program indicated that the AC power system efficiency depended much on the resonant inverter efficiency. The test results indicated the efficiency of the power system depended on the frequency of the resonant inverter. In a resonant inverter, the resonant frequency of the reactive components is higher than the switching or line frequency. It was shown that when the two frequencies were brought closer together, the stoed energy in the reactive component increased. This increased energy increased the current in the inverter causing the resistive power losses to rise as the square of the current while the output power rose linearly with the current. Thus, the efficiency of the system was reduced.

By tracing system losses, it was learned that significant power was being dissipated in the resonant capacitor. A closer look revealed that much of the losses were caused by Eddy currents being induced in the cases of these capacitors. While desiging and testing sample transmission lines it was learned that adding the shield to the transmission line increased the line resistance and thus the power dissipation of the line. It turned out that this additional power loss was attributed to Eddy currents being induced in the shield from the current in the line. The power in these Eddy currents had dissipated in the resistance of the shield.

In brief, reduction in efficiency was attributed to the power losses in the transmission line.

Conclusions and Recommendations

Based on these concerns MSFC feels that alternate approaches to the SRI should be investigated, and that a thorough investigation of transmission lines should be initiated, before the AC power system is built. Therefore, Bob Kapustka, a power electronics expert at MSFC, has proposed alternate approaches to SRI. Also, he has been measuring various wires to determine their transmission characteristics.

Part of my research time has been spent measuring and testing wires. The results are shown in Table 1 and Table 3. The test configuration is shown in Figure 20. The other part of my research time has been spent analyzing and evaluating the test results from the resonant AC power system test program. Based on these research experiences, I agree with MSFC that the AC system should be made more efficient before it is built. Therefore, I make the following recommendations:

a. More wires should be tested.

b. Time and money should be given to Kapustka to build and test his topology in AC power system.

c. The MSFC AC power breadboard should be used as test facility for power converters and transmission lines.

TABLE 1

POWER SUPPLY SENSITIVITY OF THE INVERTER

V_{in} (v_{dc})	I_{in} (A_{dc})	V_{out} (V_{rms})	I_{out} (A_{rms})	f(khz)
69.6 −20%	10.20	93.0	6.9	20.0
87.0	12.76	117.0	8.6	20.0
104.4 +20%	15.24	140.6	10.2	20.0

TABLE 2

WIRE TEST RESULTS AT 20 khz

WIRE #	NUMBER OF FEET	ESL (uH)	ESR (mΩ)
20	2.5	216.9	58.9
16	1.5	.4538	6.02

ESL/ ESR/\Z\ FOR 10 FT.

	LOOSE (3 in SPACING)	TOGETHER NOT TWISTED	TWISTED(2-3 TURNS/FT.)
	ESL(uH)/ESR(mΩ)/\Z\(mΩ)	ESL(uH)/ESR(mΩ)/\Z\(mΩ)	ESL(uH)/ESR(mΩ)/\Z\(mΩ)
12	5.4 /52 /681	2.4 /52 /306	1.6 /59 /209
SINGLE SHIELDED	7.2 /172.05 /472	4.2 /172 /	3.4 /179.07 /468.63

TABLE 3

TEST RESULTS OF VOLTAGE LOSSES
FOR 10 FT. OF #12 WIRE

WIRE ARRANGEMENT	INPUT VOLTAGE	OUTPUT VOLTAGE	LOSSES
PARALLEL	143.0	142.6	.4
	144.4	144.0	.4
	143.2	142.4	.8
	144.4	144.2	.2
	147.4	144.6	.8
TWISTED TOGETHER (2-3 TURNS/FT.)	148.2	148.0	.2
	147.4	146.2	1.2
	146.8	143.6	3.2
	149.4	147.6	1.8
	152.0	150.2	1.8

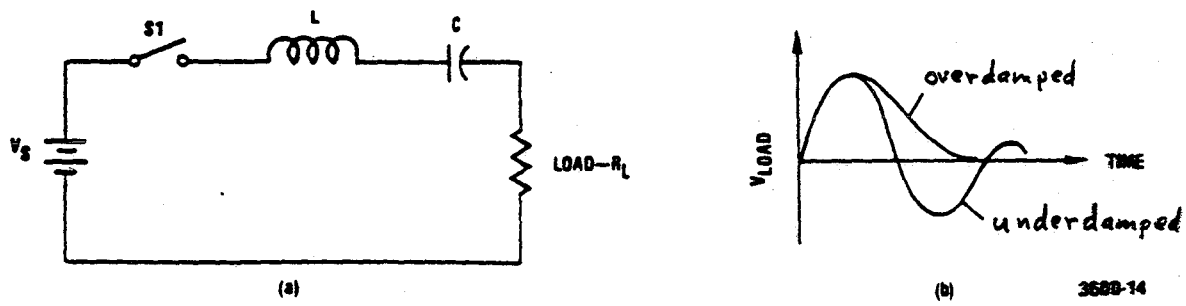


Figure 1. Basic Series Resonant Circuit

Figure A-1. Basic resonant circuit.

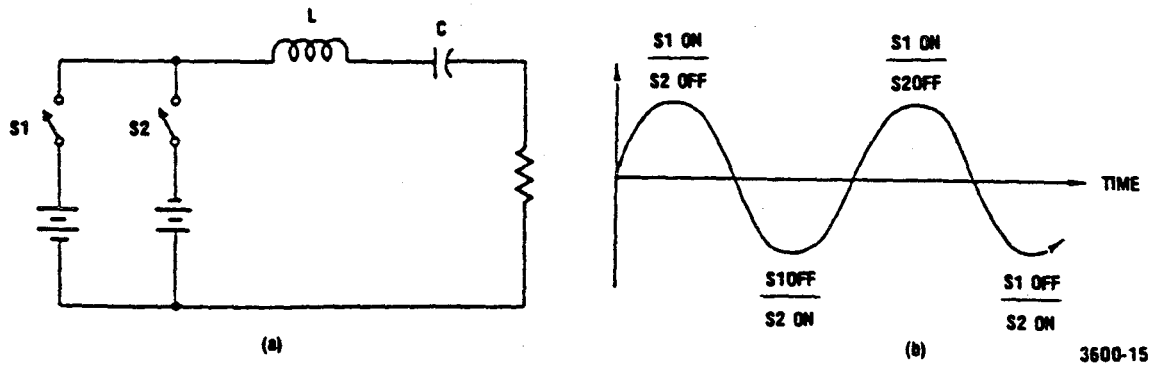


Figure 2. Dual-Polarity Series Resonant Circuit

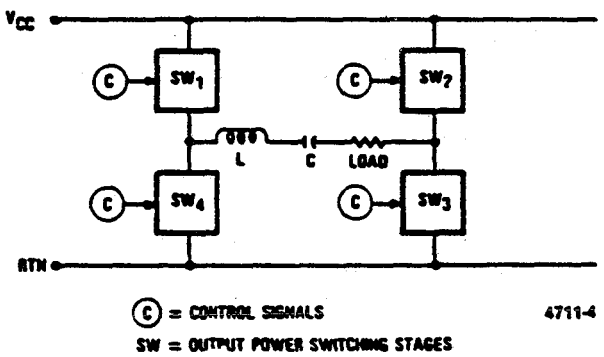


Figure 3. Series - Output Type Series Bridge Circuit

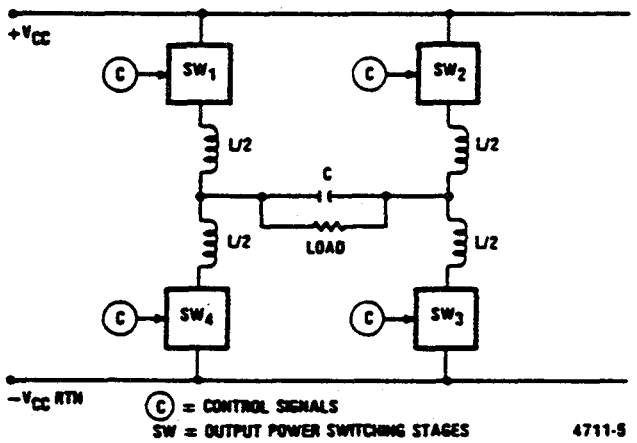


Figure 4. Parallel - Output Type Series Resonant Bridge Circuit

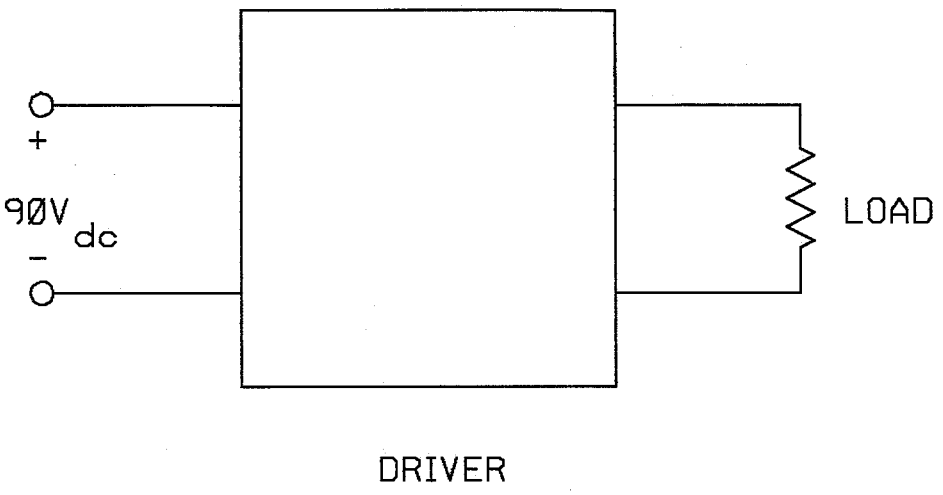
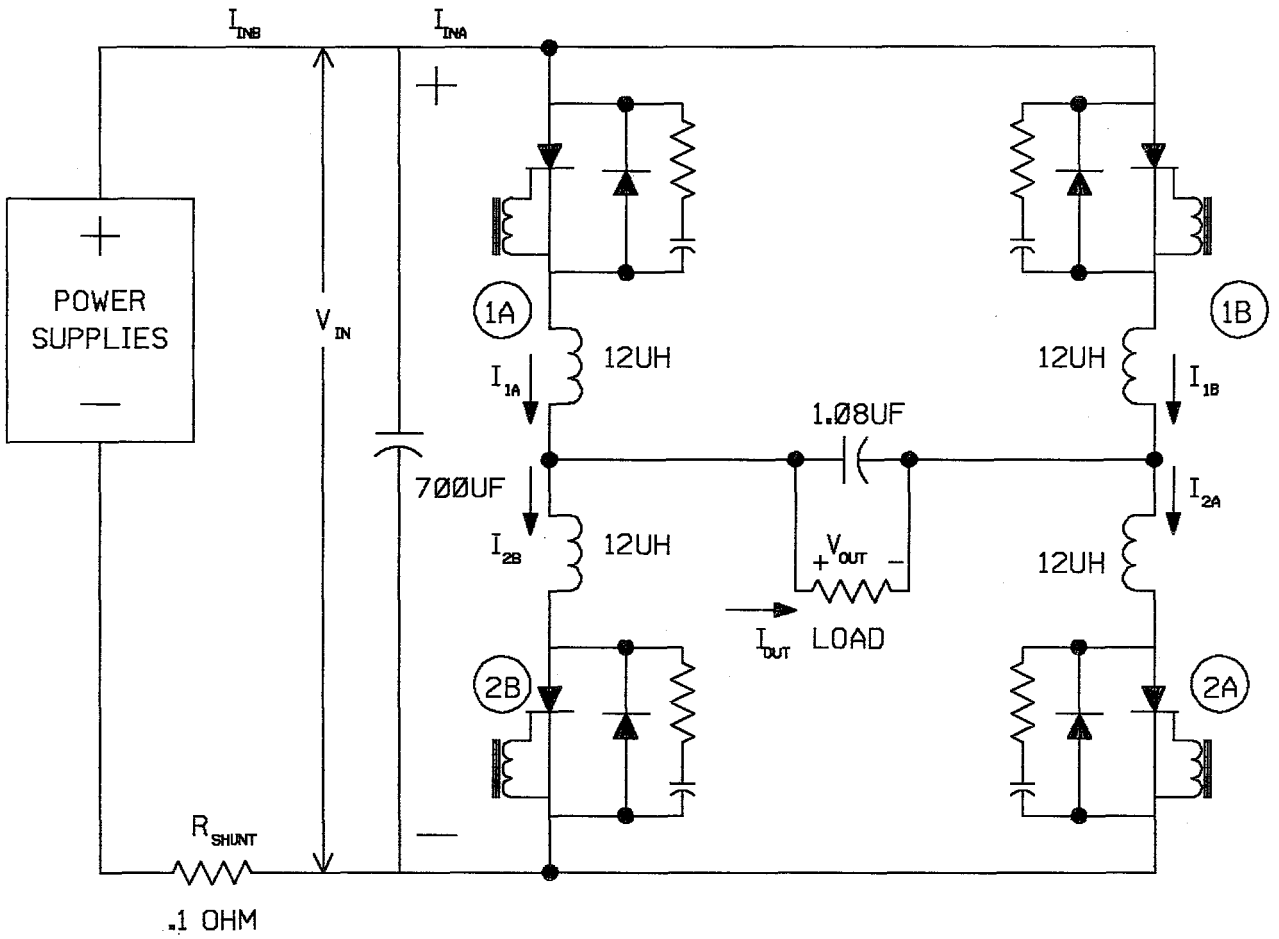


Figure 5. Test Configuration Driver with Resistive Load

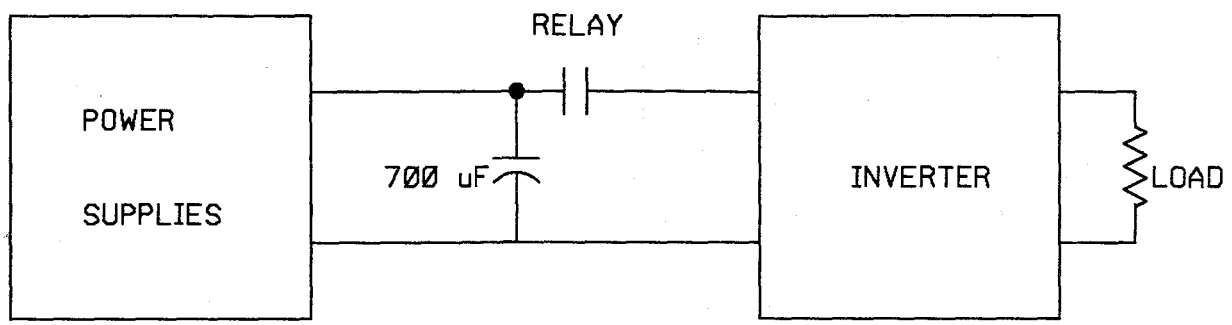


SCR'S =2N3658

DIODES =A139M

SNUBBERS =11ØOHM/.ØIUF

Figure 6. 1.0kw Inverter Schematic



LOCATION OF POWER RELAY

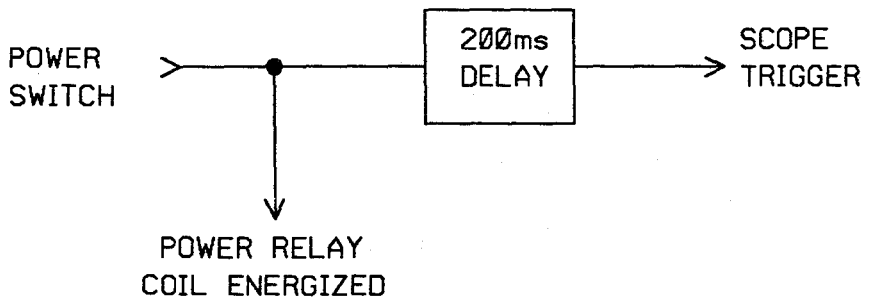


Figure 7. Start Up Circuit for Single Inverter Configuration

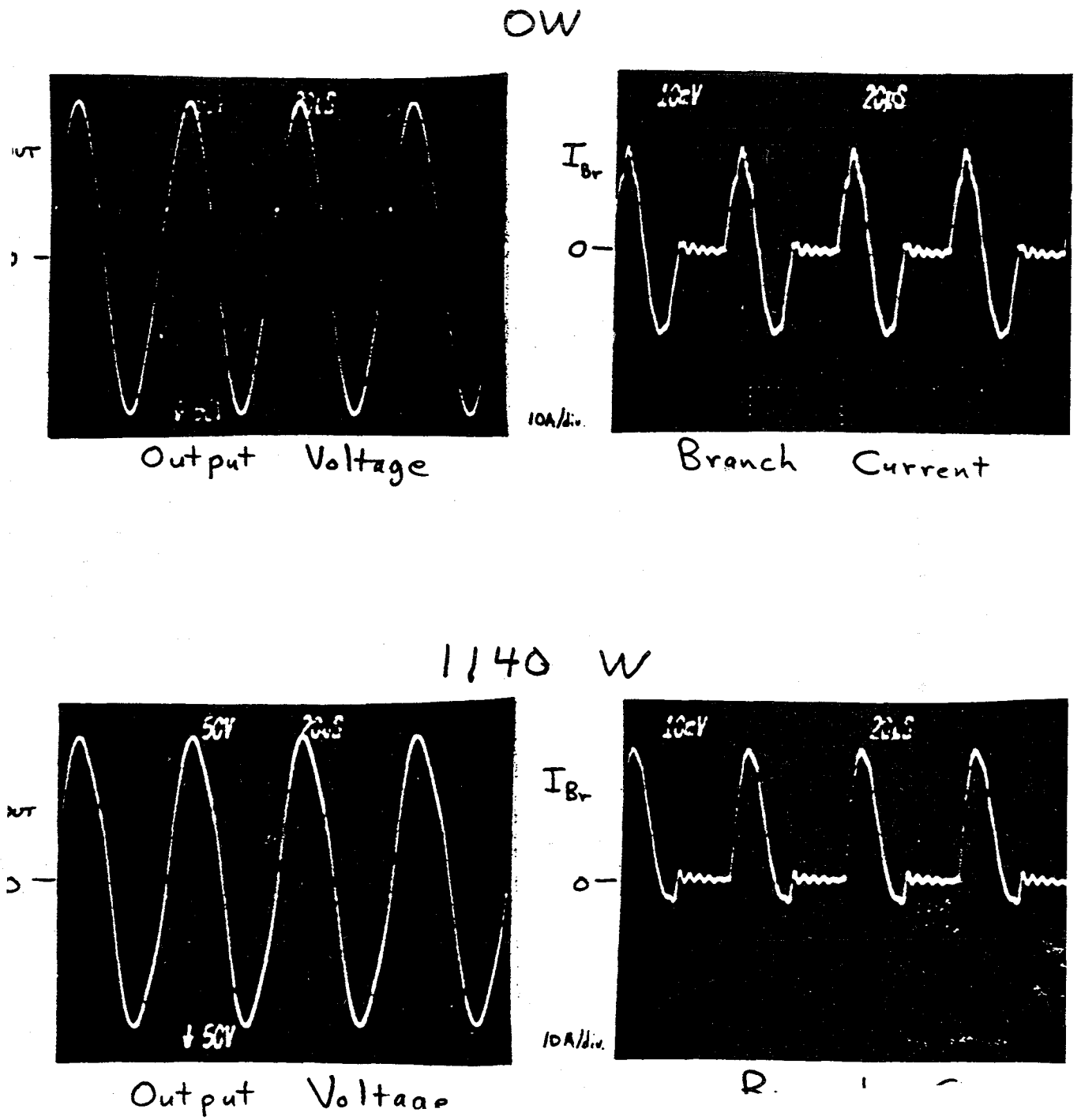


Figure 8. Inverter Leg Current, with no load and a 1.1kw load

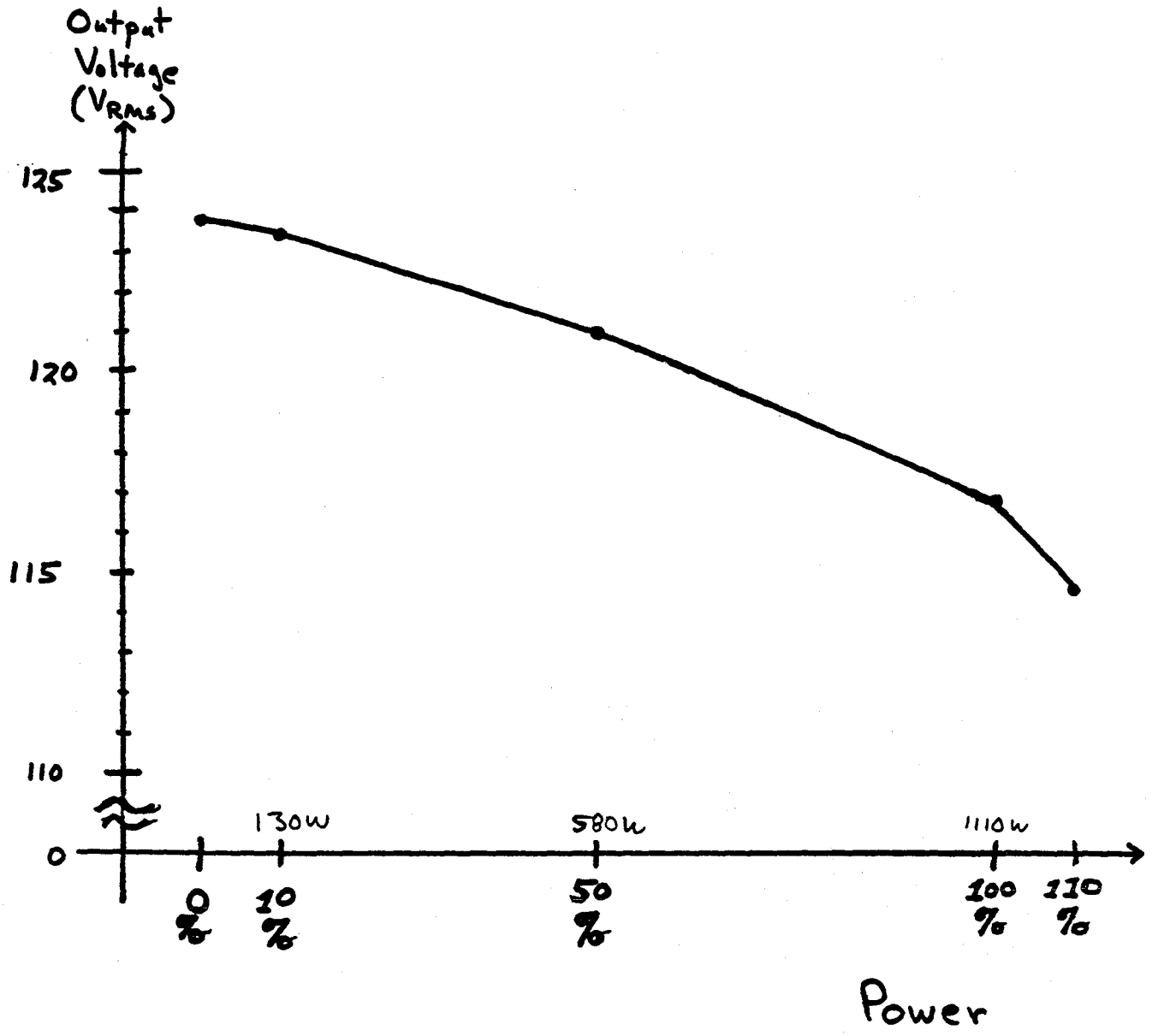


Figure 9. Output Voltage Versus Load Power

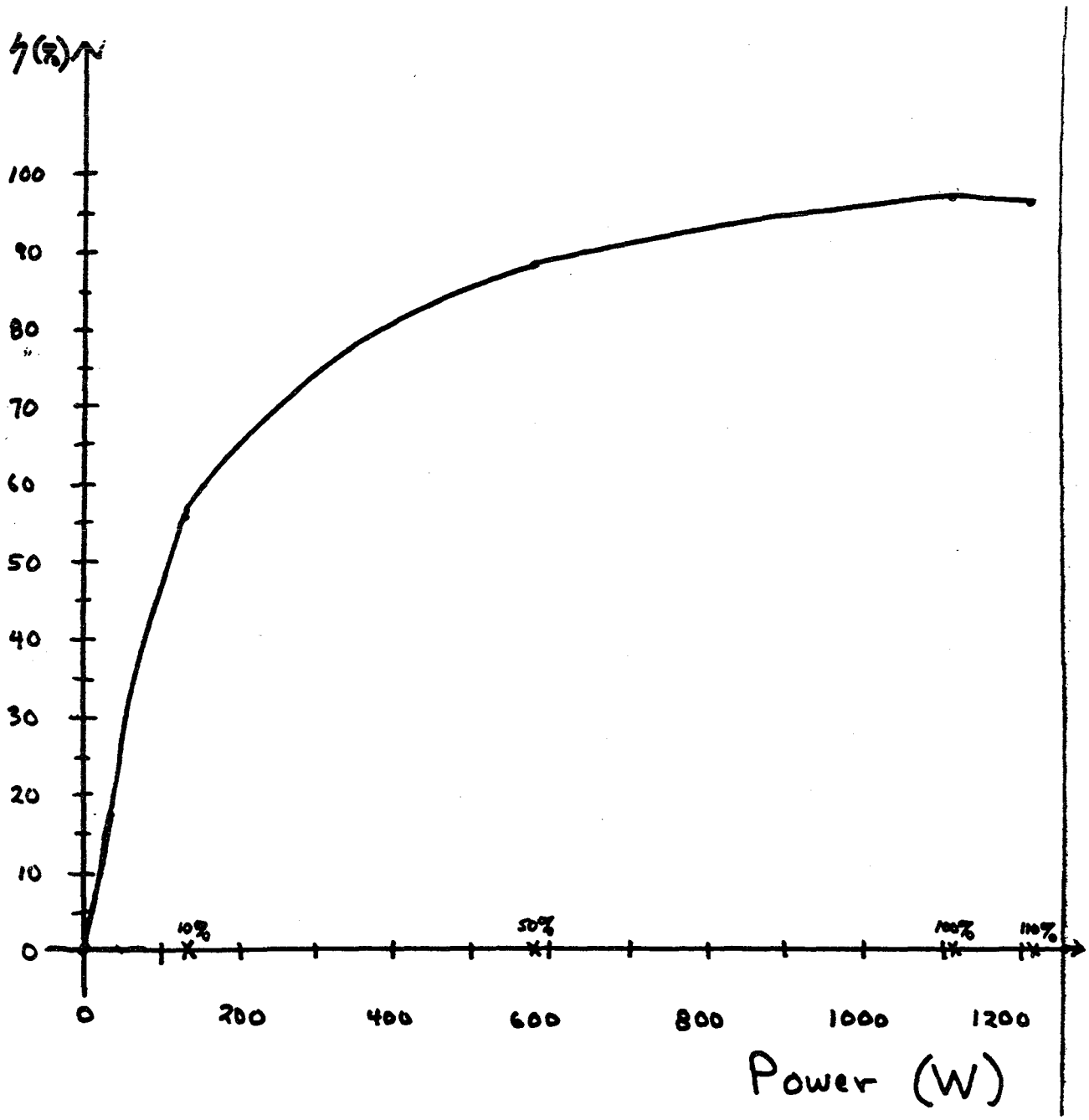


Figure 10. Efficiency Versus Load Power

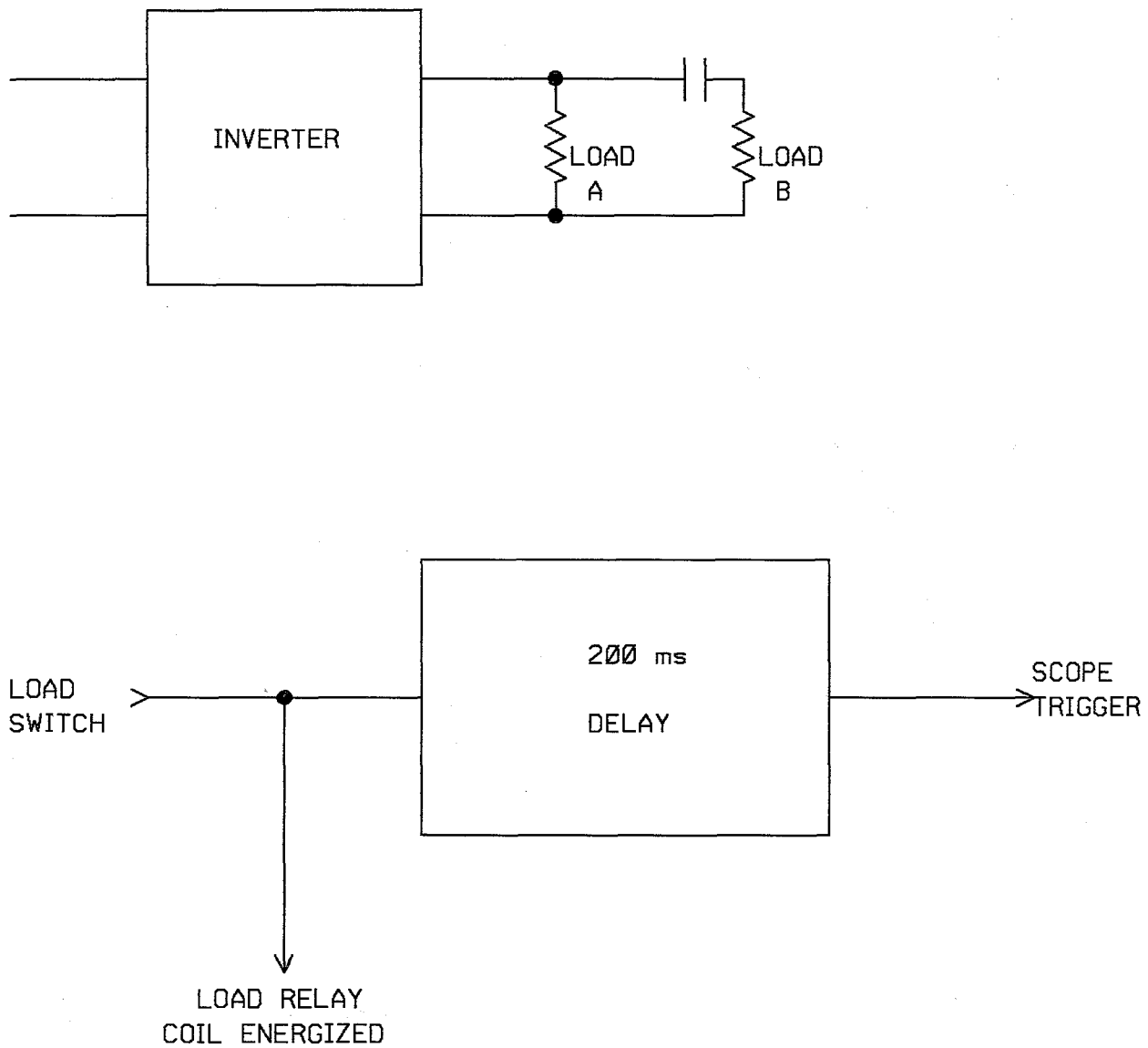


Figure 11. Test Circuit to Measure Transient Load Response of a Single Inverter Module

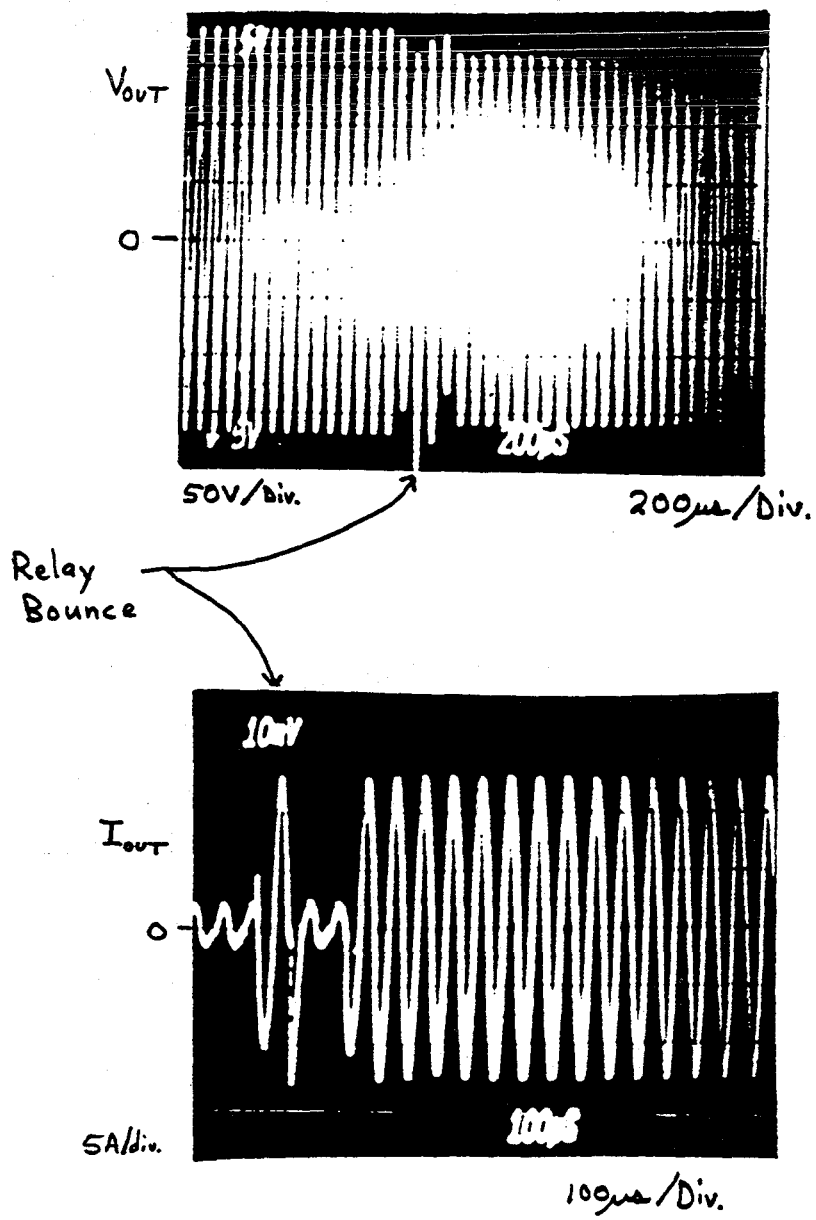


Figure 12. Inverter Output Voltage and Current as the Load is Switched from 130w to 1110w

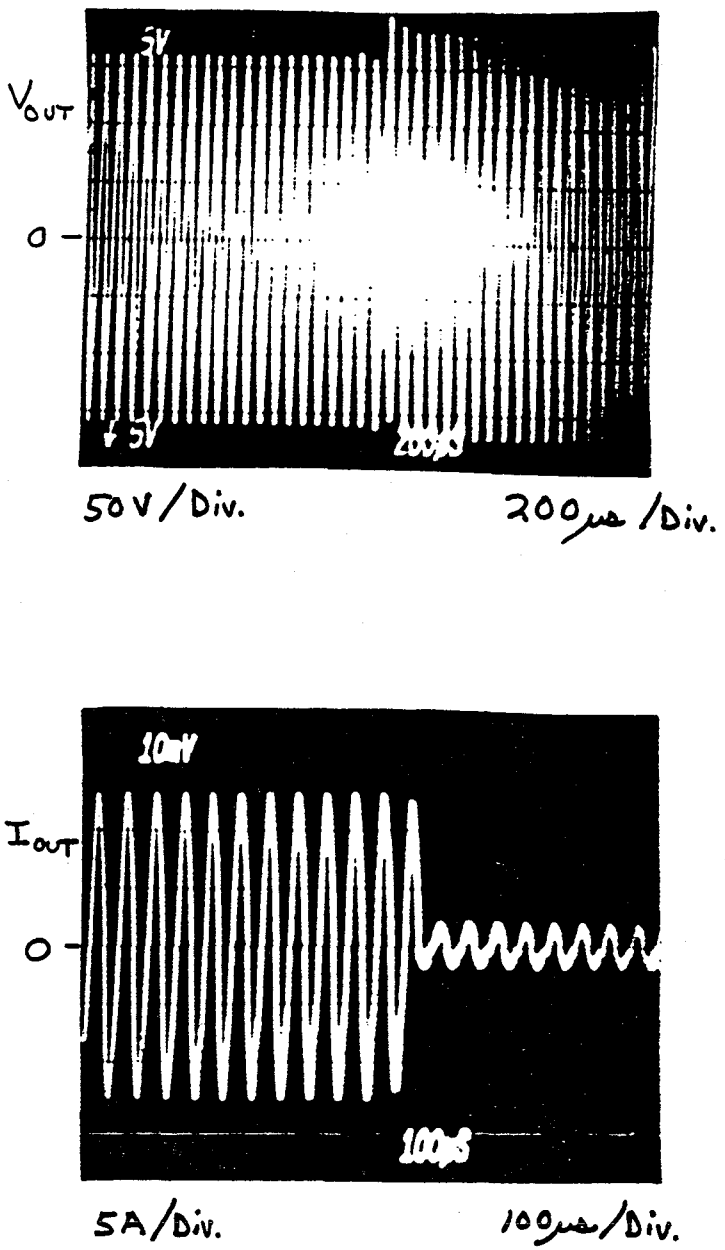


Figure 13. Inverter Output Voltage and Current as the Load is Switched from 110w to 130w

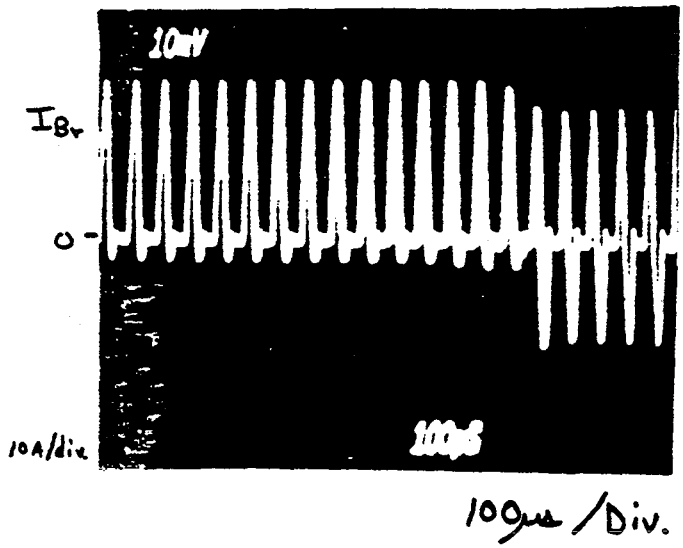


Figure 14. Inverter Leg Current as the Load is Switched from 110w to 130w

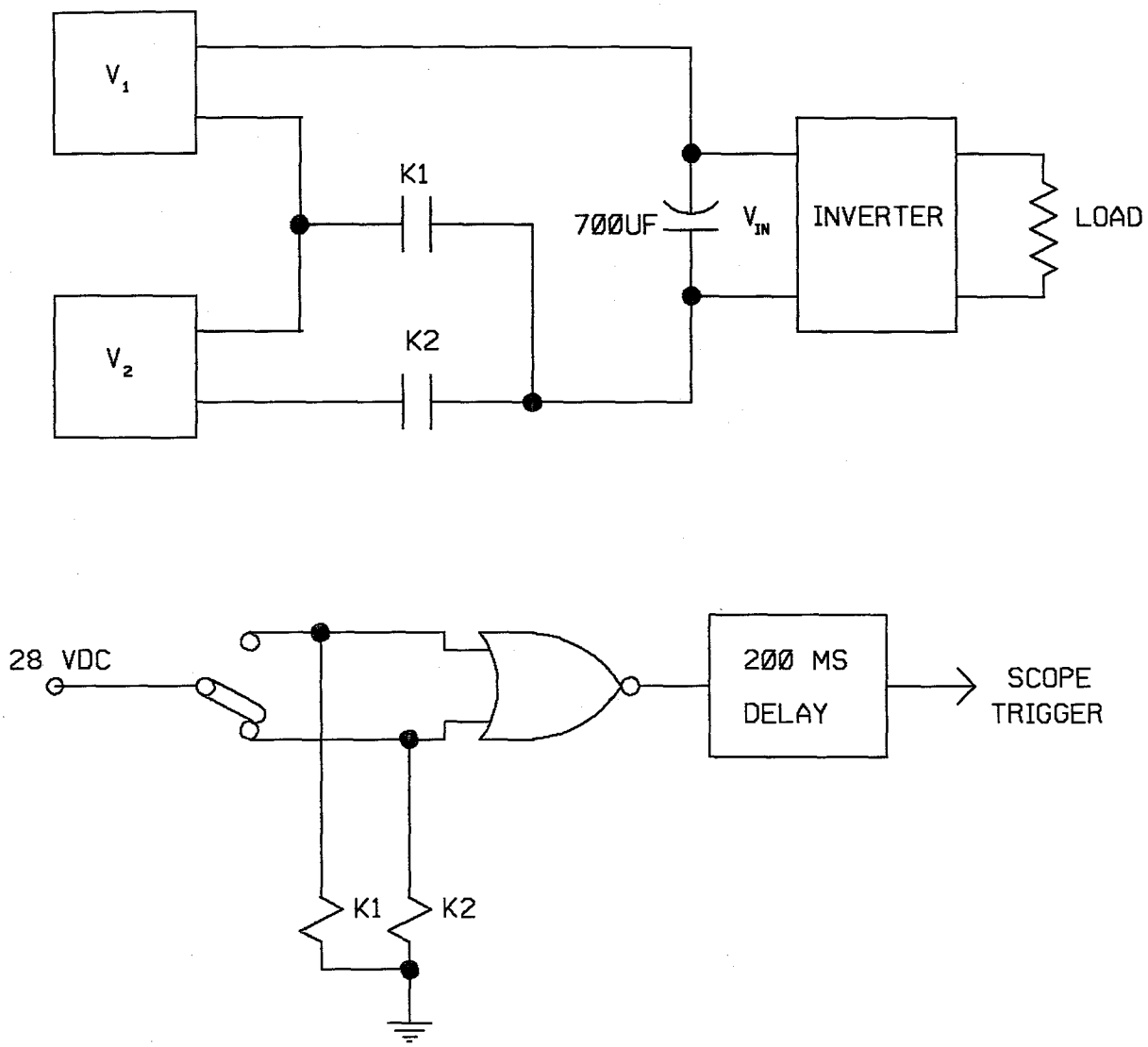


Figure 15. Test Circuits for Power Supply Sensitivity Measurements of the Inverter

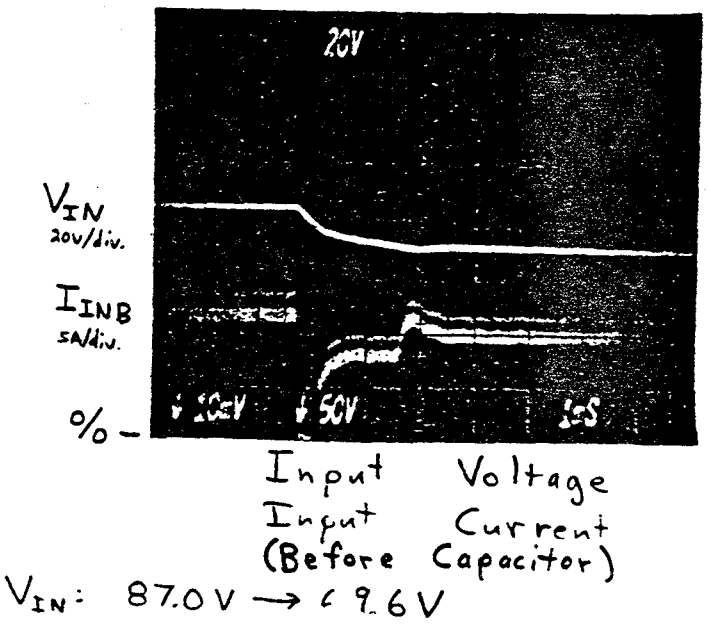


Figure 16. Input Parameters of the Inverter as the Input Voltage is Switched from 87.0Vdc to 69.Vdc

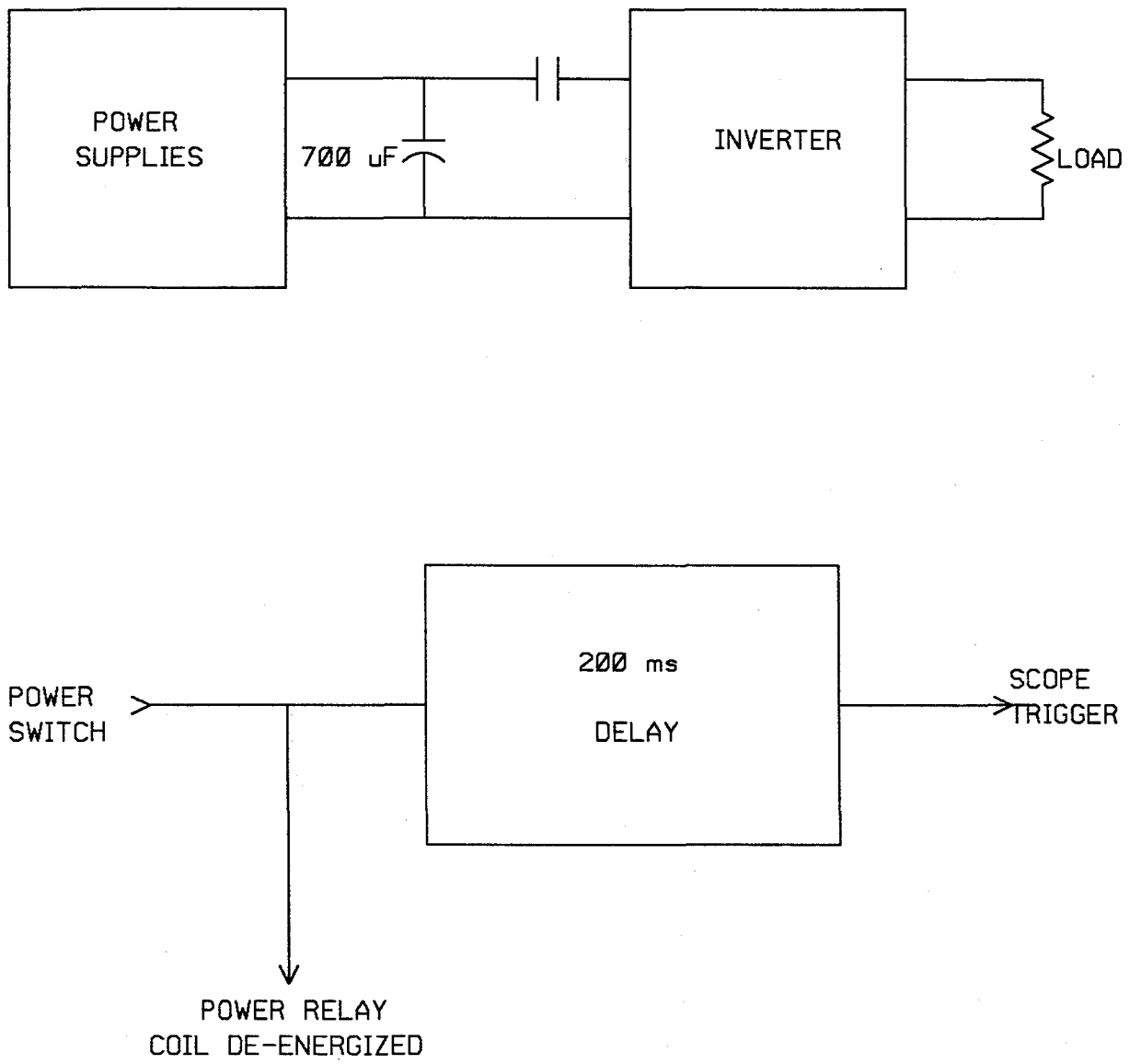


Figure 17. Circuit Used to Gather Data for the Single - Inverter Case of Power Turn Off

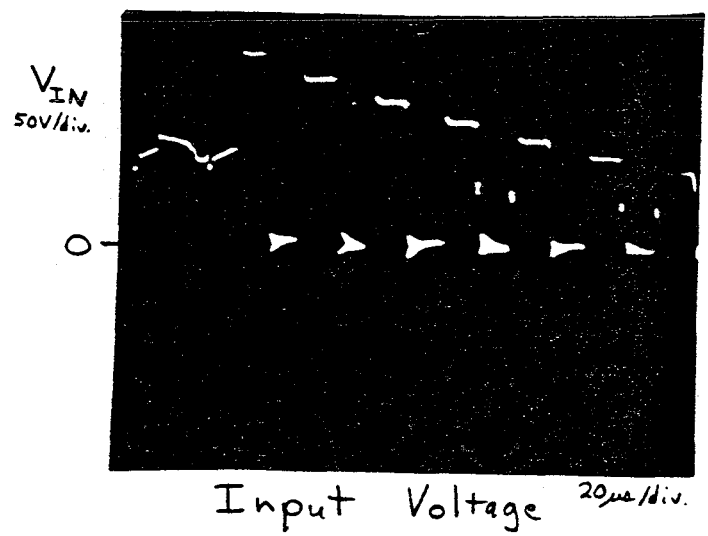
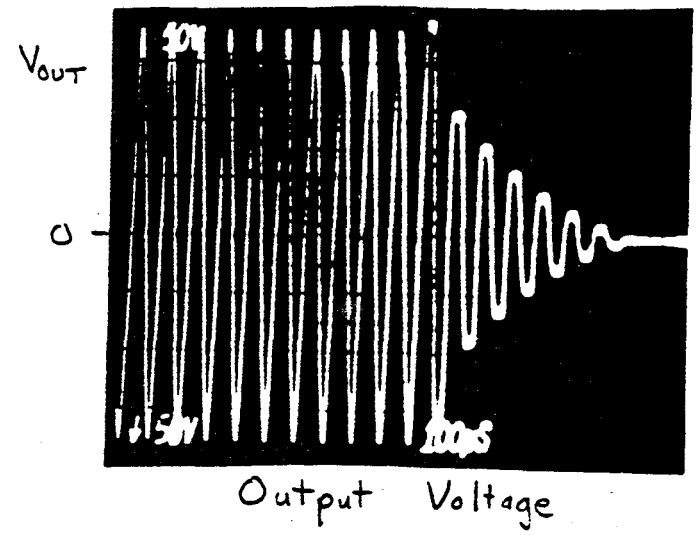


Figure 18. Input Voltage as the Power is Turned Off



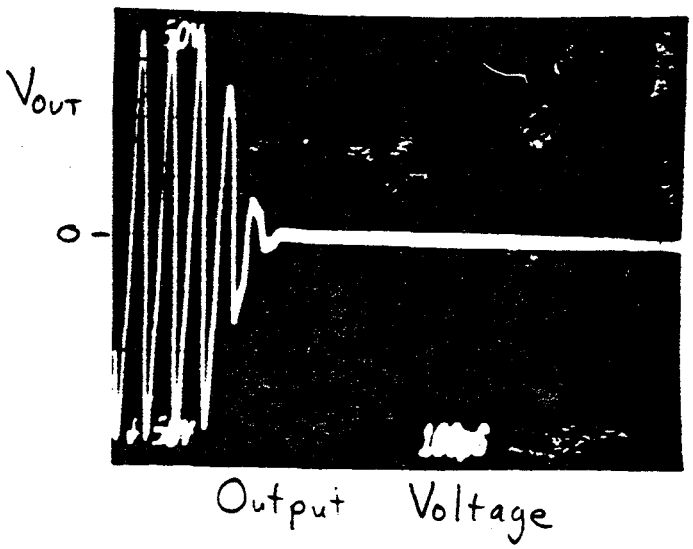


Figure 19. Power Turn Off of a Single Inverter for 0.0 and 580-w Cases

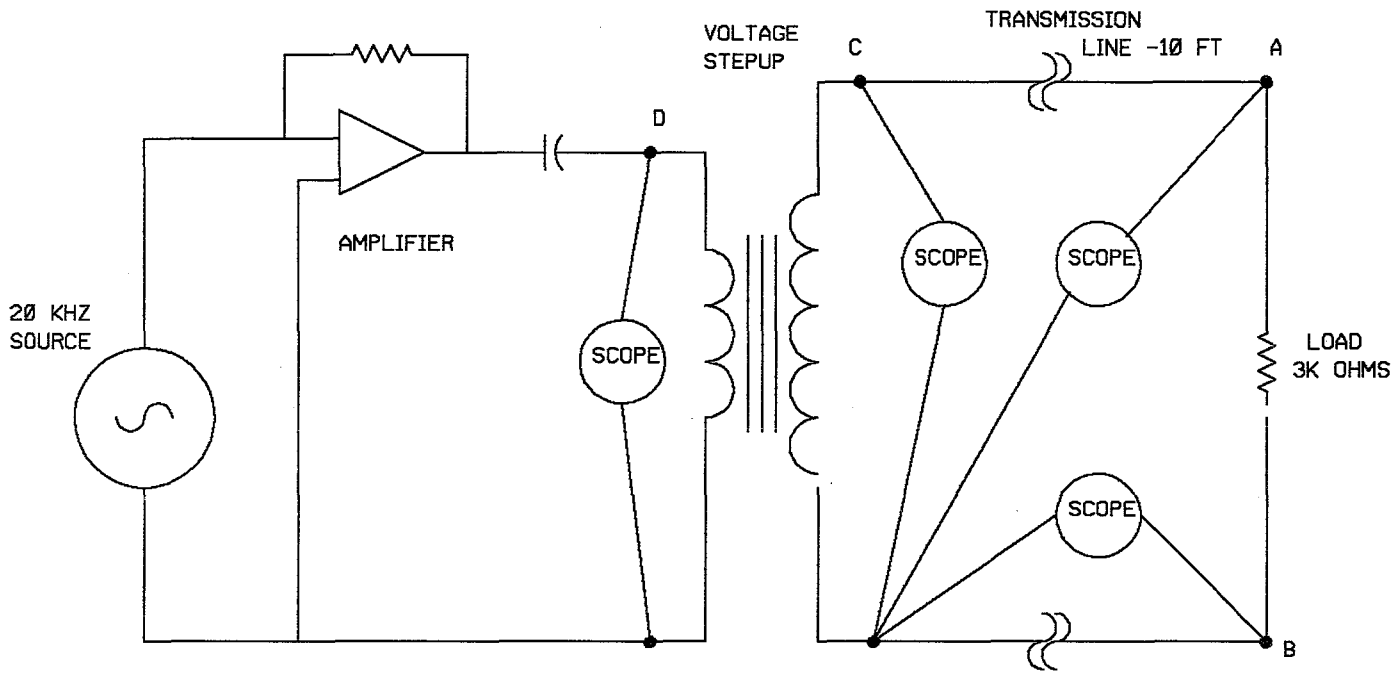


Figure 20. Wire Testing Configuration

REFERENCES

1. S. B. Dewan and A. Straughen, "Power Semiconductor Circuits", John Wiley and Sons, Inc., New York (1975), pp. 426-431

2. John R. Lanier, Jr., James R. Graves,Robert E. Kapustka, and John R. Bush, Jr., "A Programmable Power Processor for High Power Space Applications", PESC 82 Record, pp. 331-340, 1982.

3. G. E. Schwarze, "Development of High Frequency Low Weight Power Magnetics for Aerospace Power Systems", 19th Energy Conversion Engineering Conference San Francisco, California, 19-24 August 1984.

1985

NASA/ASEE SUMMER FACULTY RESEARCH FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER

THE UNIVERSITY OF ALABAMA

LOW DENSITY GAS DYNAMIC WALL BOUNDARY CONDITIONS

Prepared By:	Frank G. Collins, Ph.D.
Academic Rank:	Professor
University and Department	The University of Tennessee Space Institute
NASA/MSFC:	
Division:	Aerophysics
Branch:	Unsteady Aerodynamic and Thermal Environment
MSFC Counterpart:	Lynn C. Chou
Date:	September 13, 1985
Contract No.:	NGT 01-008-021 The University of Alabama in Huntsville

Low Density Gas Dynamic Wall Boundary Conditions

By

Frank G. Collins
Professor of Aerospace Engineering
The University of Tennessee Space Institute
Tullahoma, Tennessee

ABSTRACT

Low density nozzles or large expansion ratio nozzles used in space experience rarefaction effects near their exit in the form of velocity slip and temperature jump at the walls. In addition, the boundary layers become very thick and there is a very strong viscous/inviscid interaction. For these reasons no existing design technique has been found to accurately predict the nozzle flow properties up to the nozzle exit. The objective of this investigation was to examine the slip boundary conditions and formulate them in a form appropriate for use with a full Navier Stokes numerical code. The viscous/inviscid interaction would automatically be accounted for by using a compressible Navier Stokes code. Through examination of the interaction of molecules with solid surfaces a model for the distribution function of the reflected molecules has been determined and this distribution function has been used to develop a new slip boundary condition that can be shown to yield more realistic surface boundary conditions.

INTRODUCTION

As the density in a flow field diminishes the continuum description of that flow field begins to break down, first at the wall, and later in the interior of the flow field. The relevant parameter which describes the breakdown of the continuum description is the Knudsen number, defined as the ratio of a characteristic molecular mean free path to a relevant length that is descriptive of the flow field. The Knudsen number can be related to the Mach number and the Reynolds number, the simplest argument indicating that

$$Kn \sim \frac{M}{Re} .$$

The relevant mean free path and length scale depend upon the physical problem (see ref. 1–6).

If the Knudsen number is not considerably less than one, the flow is no longer a continuum. The most significant feature of the flow when there is slight rarefaction is a change in the boundary conditions at the wall. This regime is characterized by the fact that the rarefaction effects (i.e., changed wall boundary conditions) are small corrections to the continuum flow behavior. However, viscous effects become more important as Kn increases. Then the skin friction becomes comparable to the wall pressure and the heat transfer coefficient increases. The displacement effect of the thick boundary layer changes the surface pressure from its inviscid value. These are the dominant fluid mechanical effects in the slip flow regime, rather than the relatively small corrections introduced by the modified wall boundary conditions.

The transition regime occurs when the Knudsen number is of the order of one. In this regime molecular collisions with the surface and with the free-stream molecules are of more or less equal importance. The flow is no longer a slight correction to the continuum flow. At still larger values of the Knudsen number only collisions between the incoming molecules and the body surface need to be considered. This latter regime is called free-molecule. The delineation between the various regimes, i.e., continuum, slip, transition and free-molecule, is not sharp and depends upon the Mach number and the body geometry. References 1–6 should be consulted for more details.

Rarefied gas dynamics is the field that describes these flow regimes. NASA obviously has many applications for rarefied gas dynamics. These include the design of low density nozzles or large expansion ratio nozzles for use in space, the determination of contamination to spacecraft solar panels, optical sensor windows or temperature control surfaces, the estimation of drag and torque on satellites or the space station and of the heating and aerodynamics of vehicles entering or skip-

ping along the edge of the atmosphere. The present investigation concentrated on those applications that required slip flow boundary conditions.

The objectives of the present work are presented in the following section. This is followed by a brief review of kinetic theory, which is the tool used to solve rarefied gas dynamics problems, and of gas-surface interactions. The latter is required to understand the new formulation of the slip boundary condition that was accomplished for this study. The new slip condition is then presented and compared with previous slip boundary conditions. Evidence for the superiority of the present formulation is provided, but only incompletely. Finally, recommendations for additional work are made.

OBJECTIVES

NASA has a continuing need to improve nozzle contour design. Many of these nozzles are used under low density conditions but no design technique that is presently available can predict the flow field near the exit plane under these conditions. Accurate exit plane prediction is required to estimate thrust, as an input to contamination prediction codes and to accurately determine the back pressure on end plates of ultra-low thrust satellite control rockets.

With these applications in mind, the objective of this investigation was to devise an accurate means for the design of low density nozzles. To accommodate the strong viscous/inviscid interaction that will exist when the boundary layer becomes very thick near the nozzle exit, it will be necessary to use a full compressible Navier Stokes code and to add slip boundary conditions when the Knudsen number becomes sufficiently large. A second objective was to utilize recent gas surface interaction studies to develop a model for the distribution function of the molecules reflected from the surface that could be used to determine a slip boundary condition that can more accurately predict phenomena that are observed in slip flow, including some anomalous effects which have been observed in capillary flow.

PRINCIPLES OF KINETIC THEORY

Kinetic theory is the discipline that uses the molecular description of gases to determine the flow properties of a gas that is not in equilibrium. Because the number of molecules in any finite but small volume is very large, a statistical approach can be taken. The fundamental entity in kinetic theory is the distribution function, f, which is a function of position (x, y, z) , velocity (u, v, w) and time, t, such that

$$f(x, y, z, u, v, w, t)\, dx\, dy\, dz\, du\, dv\, dw \tag{1}$$

is the number of particles in the volume $dx\, dy\, dz$ about position (x, y, z) with velocity in the range $du\, dv\, dw$ about (u, v, w) at time t. If the distribution function is known then all flow properties can be obtained as integrals (moments) of the distribution function. For example, letting $d\boldsymbol{r} = dx\, dy\, dz$ and $d\boldsymbol{u} = du\, dv\, dw$

$$n(\boldsymbol{r}, t) = \int_{-\infty}^{\infty}\int\int f(\boldsymbol{r}, \boldsymbol{u}, t) d\boldsymbol{u}, \text{ number density} \tag{2}$$

$$\overline{\boldsymbol{u}} = \frac{1}{n}\int_{-\infty}^{\infty}\int\int f(\boldsymbol{r}, \boldsymbol{u}, t) d\boldsymbol{u}, \text{ mean (gas dynamics) velocity} \tag{3}$$

$$P_{ij} = m\int_{-\infty}^{\infty}\int\int U_i U_j f(\boldsymbol{r}, \boldsymbol{u}, t) d\boldsymbol{u} \tag{4}$$

P_{ij} is the flux of i-th momentum in the j-th direction. $\boldsymbol{U}$ is the peculiar or random velocity, defined as

$$\boldsymbol{U} = \boldsymbol{u} - \overline{\boldsymbol{u}}\,. \tag{5}$$

Other properties can be similarly defined; see reference 8.

To proceed further a molecular model must be developed. The simplest model is that of molecules composed of rigid elastic spheres of diameter σ which interact with one another only upon impact. This is the model that will be assumed in this work while more realistic models are described in reference 8. Classical mechanics can then be used to show that the distribution function varies with time according to the equation

$$\frac{Df}{Dt} \equiv \frac{\partial f}{\partial t} + \boldsymbol{u} \cdot \frac{\partial f}{\partial \boldsymbol{r}} = C^{(+)} - C^{(-)} \equiv J(f, f_1) \tag{6}$$

where $C^{(-)}\ d\boldsymbol{r}\ d\boldsymbol{u}\ dt$ is the total number of molecules having velocities in the range $\boldsymbol{u}$ to $\boldsymbol{u} + d\boldsymbol{u}$ which are lost in $d\boldsymbol{r}\ d\boldsymbol{u}$ in time dt due to collisions with molecules having other velocities and $C^{(+)}\ d\boldsymbol{r}\ d\boldsymbol{u}\ dt$ is the total number of molecules gained to the range $\boldsymbol{u}$ to $\boldsymbol{u} + d\boldsymbol{u}$ in $d\boldsymbol{r}\ d\boldsymbol{u}$ in time dt due to collisions between molecules having those initial velocities that lead to a final velocity of one of the molecules in the range $\boldsymbol{u}$ to$\boldsymbol{u} + d\boldsymbol{u}$. The specific form of the quantities $C^{(+)}$ and $C^{(-)}$ is given in reference 8. It has been assumed that the molecules are not acted upon by an external force field.

The above equation is called the Boltzmann equation and is considered to be the fundamental equation of rarefied gas dynamics. Many assumptions must be made to derive this equation, the two most important being that the equation as written applies only to a pure gas with no internal degrees of freedom (monatomic molecule). The equation can be generalized to eliminate these limitations (ref. 9) but even in this simplest case illustrated the functions $C^{(+)}$ and $C^{(-)}$ are so complex that the equation can be solved only for a few special circumstances.

Moments of the Boltzmann equation, i.e., equations of the form

$$\int_{-\infty}^{\infty}\int\int\left\{\frac{Df}{Dt} - J(f, f_1)\right\} u^n\, d\boldsymbol{u} = 0\,, \tag{7}$$

play a special role in kinetic theory. These are the Maxwell equations of change. Since mass, momentum and kinetic energy are conserved in each molecular collision (monatomic molecules) the moments of the collision integral for $n = 0$, 1 and 2 are zero. The moments of the left hand convective term lead to the familiar continuity, momentum and energy equations of continuum mechanics. Each equation, however, contains a higher order moment and the system is not closed. The system can be closed for systems close to equilbrium in the limit as Kn goes to zero, leading to the Chapman-Enskog expressions for the stress tensor and heat flux vector.

A useful exact solution of the Boltzmann equation exists for a gas in a state of local equilibrium, called a locally Maxwellian distribution (ref. 8). It can be written as

$$f_0 = n\left(\frac{m}{2\pi kT}\right)^{3/2} \exp\left\{-\frac{m}{2kT}\,U^2\right\};\boldsymbol{U} = \boldsymbol{u} - \bar{\boldsymbol{u}} \tag{8}$$

where n and $\bar{\boldsymbol{u}}$ depend upon $\boldsymbol{r}$ and t and T depends upon t. The collision integral is zero for the locally Maxwellian distribution, i.e., $J(f_0 f_{0_1}) = 0$, and the stress tensor and heat flux vector are zero. The conservation equations become the Euler equations of fluid mechanics.

A useful property of a gas is the mean free path, which is defined as the mean distance that a molecule travels between collisions. This concept has meaning only for a molecular model of hard elastic spheres. If the gas is described by a Maxwellian distribution then the mean free path is given by the expression

$$\lambda = \frac{1}{\sqrt{2}\, n\, \pi \sigma^2} \,. \tag{9}$$

More complicated situations are discussed in reference 8. It was assumed in the Introduction that a single mean free path described the flow but that is usually not the case.

The purpose of this brief introduction has been to indicate that boundary conditions must be provided on the distribution function itself to solve problems in rarefied gas dynamics. For problems of interest the most difficult boundary condition to specify is that at the wall. The interaction of molecules with solid surfaces will be reviewed in the next section and it will be shown that the experimental results can be described by a particularly useful model of the distribution function of the reemitted molecules. This model will then be used to develop a new boundary condition in the following section and this will yield a new slip regime boundary condition.

GAS-SURFACE INTERACTIONS

The study of gas-surface interactions provides the boundary conditions for the calculation of low density flows past wall boundaries. Our understanding of the molecule/wall interaction has evolved with time and is still presently incomplete. The range of relative energies between the molecule and the surface for applications of interest is very broad. Satellites operate in a free-molecule environment with a relative velocity between the satellite and the surrounding gas of around 26,000 ft/s (approximately 8000 m/s). This translates into a relative energy of $9.1eV$ for a nitrogen molecule or $5.2eV$ for an oxygen atom. The flow in low thrust or low density nozzles offer examples of the other extremes. The flow near the exit of a low density nozzle which is operating in space is in the slip regime on the wall. In this case the energy of the gas molecule relative to the wall is considerably lower, frequently below $0.1eV$, even though the wall and gas may be at a high temperature (the average molecular energy at $3000°K$ is $0.39eV$). This wide range of relative energy of interest has caused experimental problems with the study of the interaction of gases and surfaces because so many physical processes can occur over such a broad energy range (ref. 10) and because of the difficulty of producing relative energies greater than $0.3eV$; in fact, the satellite energy range has been reached only comparatively recently in laboratory experiments.

The study of gas-surface interactions has a long history that can only be highlighted here. The purpose of the discussion will be to justify the form of the distribution function of the reemitted molecules that will be used to obtain the slip flow boundary conditions.

Early experiments indicated that there was some accommodation between the incident molecules and the wall properties. If molecules reflect specularly then there is no accommodation to the wall conditions and Maxwell reasoned that a portion must be reemitted diffusely after having come to complete accommodation with the wall. He introduced the reflection coefficient, f, which is the fraction of the molecules which reflect diffusely while the fraction $(1 - f)$ reflect specularly (ref. 11). Later it was recognized that different properties of the molecule-wall collision accommodate to a different degree and accommodation coefficients were introduced. The accommodation coefficients for normal momentum, tangential momentum and energy are defined by the equations (refs. 1 and 3).

$$\sigma^1 = \frac{p_i - p_r}{p_i - p_w} = \text{normal accommodation coefficient} \tag{10}$$

$$\sigma = \frac{\tau_i - \tau_r}{\tau_i} = \text{tangential accommodation coefficient} \tag{11}$$

$$\alpha = \frac{e_i - e_r}{e_i - e_w} = \text{thermal accommodation coefficient} \tag{12}$$

where p_i, τ_i and e_i are incident fluxes of normal and tangential momentum and energy, p_r, τ_r and e_r are the reemitted fluxes of these quantities from the surface while p_w and e_w ($\tau_w = 0$) are the fluxes that would be reemitted by a gas in complete Maxwellian equilibrium with the surface. These expressions have significance only for conditions close to equilibrium.

Almost all rarefied gas dynamic problems have been solved using these accommodation coefficients with values that have been obtained from laboratory measurements. The measurements that correspond closest to the defined quantity are those of the thermal accommodation coefficient (refs. 12 to 14). Recently σ and σ' have been measured by measuring the forces exerted on a small plane surface, that is connected to a vacuum microbalance, by a stream of molecules (molecular beam) of known energy. Both of these quantities have been measured as a function of the energy of the incident molecules and of the angle between the surface normal and the incident molecular beam direction (E_i, θ_i).

The low energy and satellite energy behavior of σ and σ' have been found to be quite different (refs. 15–18). The low energy behavior was examined very completely by Seidl and Steinheil (ref. 17). In the ``as received''condition plus an ambient temperature desorption, the surface of all materials was found to be covered with adsorbed gases, primarily water vapor and hydrocarbons. However, σ was found to depend greatly upon the bulk material even though the surface layers were the same. The tangential accommodation coefficient was greatly diminished for all values of θ_i by removing the adsorbed gases if the surface was atomically smooth but σ was greater than one at small θ_i for atomically rough surfaces due to back scattering.

The tangential accommodation coefficient was found to fall with increasing θ_i at low energies (ref. 17) but the opposite behavior was observed at satellite energies (ref. 18). Also, the integrated values of σ' (over all θ_i) decreased from their low energy value near unity to a minimum at an energy somewhere below satellite energy and then increased to values comparable to the low energy value at higher energies (refs. 10, 15 and 16).

The measurements of σ and σ' at satellite energies by Knechtel and Pitts (ref. 18) are of interest. They measured both coefficients as a function of θ_i and E_i for nitrogen on aluminum and then applied them to determine the torque and drag on a surface element in free molecule flow at satellite energy ($E_i = 9eV$). A very large effect was shown to occur for the torque when compared to the assumption of full accommodation (see Fig. 16 in ref. 15). The drag was also calculated for a spherical satellite as a function of E_i. For free-molecule flow in the limit $S \longrightarrow \infty$

it can be shown that the drag coefficient on a sphere reduces to (ref. 1)

$$C_D \longrightarrow (2 - \sigma' + \sigma). \tag{13}$$

In either the limit of full accommodation ($\sigma = \sigma' = 1$) or no accommodation ($\sigma - \sigma' = 0$) this gives the value $C_D = 2$, and in general $C_D > 2$. However, integration of the forces using the measured values of σ and σ' yielded values of C_D less that 2 at E_i less than $40eV$. This significant finding will be compared with calculations that do not use accommodation coefficients later.

The momentum accommodation coefficients were historically defined only for free molecule flow. But they are also important for the understanding of gas-surface boundary conditions in transition and slip flow (ref. 19). The velocity slip and the onset of rarefaction effects can be significantly influenced by the degree of accommodation (ref. 20).

The accommodation coefficients only allow integrated properties of the reflected molecules to be determined in terms of the incident properties (fluxes are integrals over the distribution function - see the next section). However, what is desired is a joint distribution function (ref. 1)

$$g(\boldsymbol{u}_r, \boldsymbol{u}_i), \tag{14}$$

which is the probability of a molecule incident upon a wall with velocity $\boldsymbol{u}_i$ being reflected with velocity $\boldsymbol{u}_r$. This distribution function would be a function of numerous parameters, such as the type of surface material, the microscopic surface state, the wall temperature and the presence or absence of adsorbed gases. Present knowledge of the gas-surface interaction process is insufficient to yield knowledge of the joint distribution function. However, molecular beam scattering experiments have given considerable information about the interaction process. The most useful experiments are those that impinge a surface of known state with a monoenergetic molecular beam at various θ_i and measure the reflected number density and velocity distribution as a function of reflected angle (ref. 21). These experiments have not yet produced enough information to give g, however they have assisted in the development of a parametric representation of the distribution function of the reflected molecules. This representation will be used in the next section to develop a new slip condition while the experiments themselves will be examined to estimate the values of the parameters in the representation.

Accommodation coefficients have significance only for conditions close to equilibrium (ref. 13). Frequently flows of interest are far from equilibrium and the results supplied by the accommodation coefficients are not adequate (ref. 21). The molecular beam simulates conditions far from equilibrium and should thus model

such conditions better than standard accommodation coefficients.

Considerable progress has been made toward determining the joint probability distribution. Recent molecular beam scattering experiments are described in references 22 to 29 while the earlier experiments are summarized in reference 13. Only certain aspects of these measurements are relevant to the present discussion and they can be summarized as follows.

1. Maxwellian diffusive reflection, with the molecules accommodated to the wall temperature, occurs at low beam energies if the energy of adsorption of the molecules is high or if the surface is covered with adsorbed gases.

2. Lobal rather than diffusive (cosine) reflection with incomplete thermal accommodation is observed at low energies if the surfaces are properly degassed or if the incident beam energy is increased such that the incident kinetic energy is greater than the wall temperature.

3. Lobal reflected distributions are always observed at higher energies whether the surfaces are degassed or not and from polyecrystalline or single-crystal materials.

4. The peak in the lobal distribution is at an angle θ_m from the surface normal which is less than the corresponding incident angle of the molecular beam, θ_i, at low energies. As the energy of the incident beam increases (for fixed θ_i), θ_m increases so that the lobe peak approaches the surface tangent, but at some energy less than $10eV$, θ_m reaches a maximum value and decreases toward the specular angle with further increases in the beam energy. This pattern is approximately the same for all θ_i.

5. The speed distribution of the reflected molecules is analyzed by fitting the data to a ``drifting´´Maxwellian with parameters U_r and T_r. U_r is the mean velocity of the distribution. U_r, U_i and the surface normal are in the same plane. T_r is a measure of the spread of the distribution and is not usually equal to the wall temperature T_w. The parameters are usually combined to give a speed ratio $S_r = U_r/(2kT_r/m)^{1/2}$. Earlier experiments (ref. 30) indicated that a one-parameter distribution was insufficient to characterize the reflected speed distribution but most data can be adequately characterized using these two parameters.

6. At low energies, as $\theta_r \longrightarrow 90°$ (tangential reflection), S_r increases and T_r decreases for given S_i and θ_i (refs. 26 and 28).

Some years ago Nocilla (ref. 31) suggested that the reflected molecular distribution could be assumed to have the form of a ``drifting´´Maxwellian of the

form

$$f_r(\boldsymbol{u}) = n_r \left(\frac{m}{2\pi k T_r} \right)^{3/2} \exp \left\{ \frac{-m(\boldsymbol{u} - \boldsymbol{U}_r)^2}{2kT_r} \right\} \tag{15}$$

where $\boldsymbol{u}$ is the molecular velocity, U_r is the macroscopic drift velocity attributed to the outgoing stream which is assumed to lie in the plane in the incident velocity and the surface normal, T_r is the effective temperature of the distribution (not necessarily the wall temperature T_w) and n_r must be found by balancing the incoming and outgoing number fluxes. The previous discussion indicates that this form of distribution function has been found to describe most of the molecular beam data.

Nocilla made a comparison between some of the scant angular flux data that were available at the time and determined a best fit using this distribution. The fit produced the parameters θ_r and S_r (the distributions were normalized to a unit value of $(2kT_r/m)^{1/2}$). The only measurements that existed were for the scattering of low energy nitrogen from lithium fluoride crystals. Polished crystal surfaces tests produced $\theta_r > 90°$ and small values of S_r. Cleaved crystal surface tests produced θ_r slightly greater than θ_i and S_r of the order of 1.0; an exception occurred for θ_i approaching 90°, for which $\theta_r > 90°$. All the measurements were adequately described by this function except the tests with θ_i near the surface tangent. In this case the data was adequately represented except in the quadrant that contained U_i, where the distribution underpredicted the scattering flux. In this case a greater number of molecules were rejected toward the source than the distribution function predicted.

This distribution function was used by Hurlbut and Sherman (ref. 32) to determine the drag on a spherical satellite. The value of T_r was determined from an angle-dependent thermal accommodation coefficient. Both S_r and α were assumed to vary linearly with θ_i ; U_r was assumed to be directed at θ_r, which was taken to be equal to θ_i. The results yielded the classical diffusive or specular reflection results for appropriate values of the parameters but notable among the results was a prediction that C_D could be less that 2.0 for $S_r \longrightarrow \infty$ for some values of the parameters. Recall that this result was also predicted by Knechtel and Pitts (ref. 18).

The Nocilla distributions will be assumed to be an accurate representation of the reflected distribution function for slip flow and a new slip boundary condition will be presented. The problem of the estimation of the parameters U_r, S_r and θ_r, will still remain, but the experimental results described in this section can be used as a guide.

SLIP WALL BOUNDARY CONDITIONS

Slip flow occurs when there is only slight rarefaction and the departure from the continuum behavior described by the Chapman-Enskog solution is small. The most rigorous way of solving this problem near the wall would be to solve the Boltzmann equation using an appropriate kinetic boundary condition (ref. 8). Instead an approximation due to Patterson will be used (refs. 6 and 33).

The molecular motion near a wall results from the interaction of the incident and reflected molecules. The incident molecules within a mean free path of the wall have a motion consistent with the external flow while the reflected molecules have come to some partial accommodation to the state of the wall, as discussed previously. The interaction of the incident and reflected molecules results in a deviation in the distribution of molecular velocities from a local Maxwellian. This distribution function next to the wall must be obtained by solving the Boltzmann equation, which is a very difficult task.

A few mean free paths from the surface the gradients of the macroscopic velocity and temperature can be considered to be constant. These gradients cease to be constant only in the immediate vicinity of the wall where the collisions between the incident and reflected molecules have their effect (see figure 1). The extrapolation of the gradient to the surface yields the slip velocity or extrapolated wall temperature. Note that this does not yield either the gas velocity or the gas temperature at the surface. But as far as the velocity and temperature fields in the bulk of the flow are concerned, these quantities can be used as boundary conditions for the Navier Stokes equations (second term of the Chapman-Enskog expansion) to yield accurate solutions everywhere except within a few mean free path lengths of the wall.

Patterson assumed that the gas near the wall was not in equilibrium but was close to equilibrium. The distribution function was assumed to be of the form

$$f_i = f_0 \left[1 + F(H)\right] \tag{16}$$

$$H_1 = \left(\frac{m}{kT}\right)^{1/2} U_x\,;\; H_2 = \left(\frac{m}{kT}\right)^{1/2} U_y\,;\; H_3 = \left(\frac{m}{kT}\right)^{1/2} U_z$$

f_0 is the local Maxwellian.

$$f_0\, dU_x\, dU_y\, dU_z = (2\pi)^{-3/2}\, e^{-\frac{1}{2}\sum_1^3 H_i^2}\, dH_1\, dH_2\, dH_3 \tag{17}$$

$$F(H) = \sum_1^3 a_i H_i + \frac{1}{2}\sum_1^3\sum_1^3 a_{ij} H_i H_j + \frac{1}{6}\sum_1^3\sum_1^3\sum_1^3 a_{ijk} H_i H_j H_k \tag{18}$$

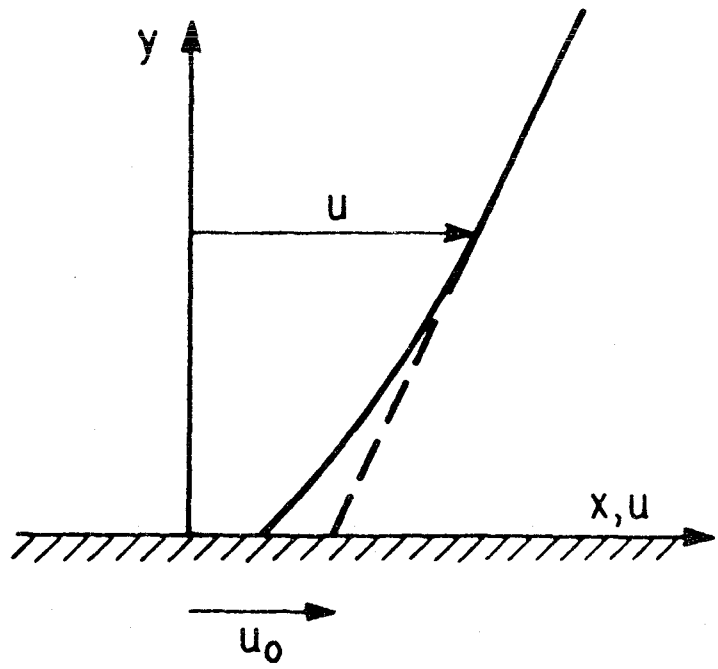


Figure 1. Coordinate geometry for slip flow boundary condition. $u_0 = u_s$.

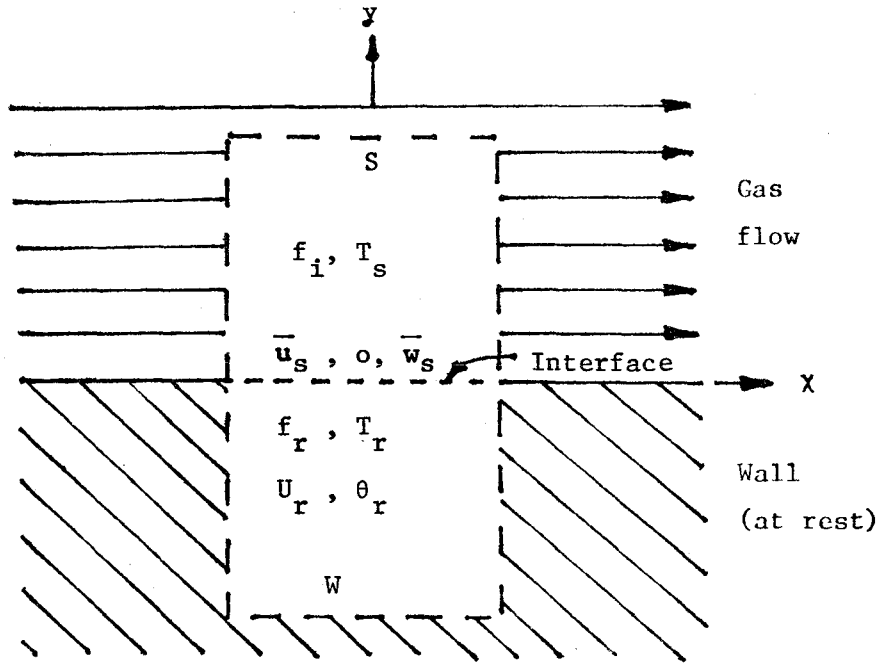


Figure 2. Integration volume for determining slip conditions

where the order of the subscripts is immaterial, e.g., $a_{12} = a_{21}$ and $a_{122} = a_{212} = a_{221}$. n, T and $\overline{U^2}$ are defined by integrals of f_0.

Certain conditions must be satisfied by the coefficients, namely (ref. 33).

$$a_{11} + a_{22} + a_{33} = 0 \tag{19}$$

and

$$2a_i + a_{i\,11} + a_{i\,22} + a_{i\,33} = 0 \ (i = 1, 2 \text{ or } 3) \tag{20}$$

Patterson demonstrated how to relate the remaining coefficients to integrals involving properties of the intermolecular collisions which were evaluated for hard elastic spheres.

Now consider the interaction of the incoming and reflected molecular streams in the vicinity of the wall. The motion of the gas on the surface depends on the transfer of mass, momentum and energy by the gas molecules to and from the surface. Consider figure 2 with the axis fixed relative to the body. The volume S in the flowing gas contains gas molecules with mean velocity components $\overline{u}_s$, 0, $\overline{\omega}_s$ and a random motion described by f_i. The reflecting molecules are assumed to act as though they emerged from an adjacent volume W with a Nocilla distribution with U_r lying in the $(x - y)$ plane. The molecular motions in the volumes S and W are mutually compatible if the total mass, momentum and energy transferred upward by the molecules from W to S are respectively equal to the total mass, momentum and energy which would be transported upward if the molecules in W moved with the same mass velocity and random motion as the molecules in S.

Compatibility of the mass flux yields

$$m\,n_s \int_{-\infty}^{\infty}\int_{0}^{\infty}\int_{-\infty}^{\infty} v\, f_i(T_s)\,du\,dv\,dw = m\,n_r \int_{-\infty}^{\infty}\int_{0}^{\infty}\int_{-\infty}^{\infty} v\, f_r(T_r)\,du\,dv\,dw \tag{21}$$

with

$$f_r = \left(\frac{2\pi\, k\, T_r}{m}\right)^{-3/2} \exp\left[-\frac{m(\boldsymbol{u} - \boldsymbol{U}_r)^2}{2\,k\,T_r}\right] \tag{22}$$

Integration and the use of the conditions on the a coefficients yields

$$\frac{m\,n_s}{4\sqrt{\pi\beta_s}}\left[2 + a_{22}\right]_s = \frac{m\,n_r}{2\sqrt{\pi\beta_r}}\,\chi\,(\sigma_r) \tag{23}$$

where $\beta = \frac{m}{2kT}$

$$\chi(\sigma_r) = \exp\left(-\sigma_r^2\right) + \sqrt{\pi}\,\sigma_r\left[1 + \operatorname{erf}\left(\sigma_r\right)\right] \tag{24}$$

$$\sigma_r = S_r \cos\theta_r\,;\ S_r = U_r\left(\frac{m}{2k\,T_r}\right)^{1/2} \tag{25}$$

This result can be written as

$$\frac{n_r}{n_s} = \sqrt{\frac{T_s}{T_r}}\ \frac{[1 + a_{22}/2]_s}{\chi(\sigma_r)} \tag{26}$$

In the hard sphere approximation used by Patterson,

$$a_{22} = \frac{5}{6}\left[\frac{\lambda}{\overline{U}}\left(\frac{\partial \overline{u}}{\partial x} + \frac{\partial \overline{w}}{\partial z} - 2\frac{\partial \overline{v}}{\partial y}\right)\right]_s \tag{27}$$

The balance of the transport of the tangential component of momentum at the interface yields an equation for the slip velocity.

$$m\, n_s \int_{-\infty}^{\infty}\int_{0}^{\infty}\int_{-\infty}^{\infty} uv f_i(T_s)\,du\,dv\,dw = m\, n_w \int_{-\infty}^{\infty}\int_{0}^{\infty}\int_{-\infty}^{\infty} u\, v f_r(T_r)\,du\,dv\,dw \tag{28}$$

$$\frac{m\, n_s}{(2\pi)^{3/2}(2\beta_s)}\left\{\pi\left[\sqrt{2\pi}\ a_{12} + a_{122}\right]\right\}_s + \frac{m\, n_s}{2^{3/2}\sqrt{2\pi\beta_s}}\left[\overline{u}(2 + a_{22})\right]_s \tag{29}$$

$$= -\frac{n_r m}{2\sqrt{\pi}}\ \frac{U_r^2}{S_r}\chi(\sigma_r)\sin\theta_r$$

Using the previous result for n_r/n_s and dropping higher order terms yields

$$\overline{u}_s = -2\, S_r\sqrt{\frac{kT_r}{m}}\sin\theta_r - \frac{\sqrt{\frac{kT_s}{m}}}{2}\left[\sqrt{2\pi}\ a_{12} + a_{122}\right]_s \tag{30}$$

$$a_{12} = -\frac{5}{4}\frac{\lambda}{\overline{U}}\left(\frac{\partial \overline{u}}{\partial y} + \frac{\partial \overline{v}}{\partial x}\right) \tag{31}$$

$$a_{122} = -\frac{15}{16}\left(\frac{\pi}{2}\right)^{1/2}\frac{\lambda}{T}\frac{\partial T}{\partial x} \tag{32}$$

Similarly,

$$\overline{w_s} = -\frac{\sqrt{\frac{kT_s}{m}}}{2}\left[\sqrt{2\pi}\ a_{32} + a_{322}\right]_s \tag{33}$$

$$a_{32} = -\frac{5}{4}\frac{\lambda}{\overline{U}}\left(\frac{\partial \overline{w}}{\partial y} + \frac{\partial \overline{v}}{\partial z}\right) \tag{34}$$

$$a_{322} = -\frac{15}{16}\left(\frac{\pi}{2}\right)^{1/2}\frac{\lambda}{T}\frac{\partial T}{\partial z} \tag{35}$$

The balance of the transport of the normal component of momentum at the interface yields a relation between p_s and p_w.

$$m\,n_s \int_{-\infty}^{\infty}\int_{0}^{\infty}\int_{-\infty}^{\infty} v^2 f_i(T_s)\,du\,dv\,dw = m\,n_r \int_{-\infty}^{\infty}\int_{0}^{\infty}\int_{-\infty}^{\infty} v^2 f_r(T_r)\,du\,dv\,dw \tag{36}$$

Integration yields

$$\frac{p_s}{2}\left[1 + a_{22} + \frac{2}{3\sqrt{2\pi}}\right]_s = \frac{p_r}{\sqrt{\pi}}\left\{\sigma_r\,\chi\,(\sigma_r) + \frac{\sqrt{\pi}}{2}\,(1 + \operatorname{erf}\sigma_r)\right\} \tag{37}$$

where the perfect gas law

$$p = n\,kT$$

has been used. This can be written as

$$\frac{p_s}{p_r} = \left[1 - a_{22} - \frac{2a_{222}}{3\sqrt{2\pi}}\right]_s \left[1 + \operatorname{erf}\sigma_r + \frac{2}{\sqrt{\pi}}\,\sigma_r\,\chi\,(\sigma_r)\right] \tag{38}$$

$$a_{222} = -\frac{45}{16}\left(\frac{\pi}{2}\right)^{1/2}\frac{\lambda}{T}\frac{\partial T}{\partial y} \tag{39}$$

The temperature jump between the gas and the wall can be determined by the flow of translational energy of the molecules across the interface.

$$\begin{aligned}&\frac{1}{2}\,m\,n_s \int_{-\infty}^{\infty}\int_{0}^{\infty}\int_{-\infty}^{\infty} v(u^2 + v^2 + w^2) f_i(T_s)\,du\,dv\,dw \\ &= \frac{1}{2} m\,n_r \int_{-\infty}^{\infty}\int_{0}^{\infty}\int_{-\infty}^{\infty} v(u^2 + v^2 + w^2) f_r(T_r)\,du\,dv\,dw\end{aligned} \tag{40}$$

Integration yields the following result

$$\begin{aligned}&\frac{m\,n_s}{4\sqrt{\pi\beta_s}}\left[1 + \frac{a_{22}}{2}\right]_s \left(\overline{u}^2 + \overline{w}^2\right)_s + \frac{m\,n_s}{4\sqrt{2\pi}\,\beta_s}\left[\left(\sqrt{2\pi}\,a_{12} + a_{122}\right)\overline{u}\right. \\ &\left. + \left(\sqrt{2\pi}\,a_{23} + a_{322}\right)\overline{w}\right]_s + \frac{m\,n_s}{\sqrt{\pi}\,(4\beta_s)^{3/2}}\left[4 + 3a_{22} - (2\pi)^{1/2}\,2\,a_2\right]_s\end{aligned} \tag{41}$$

$$= m n_r \frac{kT_r}{m} \left(\frac{kT_r}{2\pi m}\right)^{\frac{7}{2}} \left\{(S_r^2 + 2)\chi(\sigma_r) + \frac{\sqrt{\pi}}{2}\sigma_r(1 + \text{erf}\,\sigma_r)\right\}$$

Only keeping higher order terms, assuming S_r is small, gives

$$\frac{T_s}{T_r} = \left[1 - \frac{1}{4}a_{22} + \frac{1}{2}\sqrt{\frac{\pi}{2}}\,a_2\right]_s \left\{\left(1 + \frac{S_r^2}{2}\right) + \sqrt{\frac{\pi}{2}}\,\frac{\sigma_r\,(1 + \text{erf}\,\sigma_r)}{\chi(\sigma_r)}\right\} \tag{42}$$

$$a_2 = \frac{75}{32}\left(\frac{\pi}{2}\right)^{1/2}\frac{\lambda}{T}\frac{\partial T}{\partial y} \tag{43}$$

These results can be compared with the results of other investigators by writing the expressions in the form used by Kogan (ref. 34.). He wrote, for diffuse reflection

$$\overline{u_s} = A\frac{\mu}{n}\sqrt{\frac{2}{mkT}}\frac{\partial \overline{u}}{\partial y} + B\frac{2\mu}{n\,m}\frac{1}{T}\frac{\partial T}{\partial x} \tag{44}$$

$$\Delta T_s = C\frac{K}{2kn}\sqrt{\frac{\pi m}{2kT}}\frac{\partial T}{\partial y} \tag{45}$$

where μ is the coefficient of viscosity and K is the coefficient of thermal conductivity. The magnitude of the coefficients A, B and C depend upon the equation used for their determination and the molecular model assumed.

The present results can be written in this form by using the transport coefficient relations derived by Patterson for a hard sphere molecular model, namely,

$$\begin{aligned} \lambda &= \frac{1}{\sqrt{2}\,n\pi\sigma^2} \quad ; \qquad \mu = \frac{5m}{16\sigma^2}\left(\frac{kT}{\pi m}\right)^{1/2} \\ \overline{U} &= \left(\frac{8}{\pi}\frac{kT}{m}\right)^{1/2} \quad ; \qquad K = \frac{15}{4}\mu\frac{k}{m} \end{aligned} \tag{46}$$

Then, the results can be compared as follows:

1. Patterson, hard spheres, diffuse reflection.

$$\overline{u}_s = \left\{0.8862\,\frac{\mu}{n}\sqrt{\frac{2}{mkT}}\left(\frac{\partial \overline{u}}{\partial y} + \frac{\partial \overline{v}}{\partial x}\right) + 0.3750\,\frac{2\mu}{nm}\frac{1}{T}\frac{\partial T}{\partial x}\right\}_s \tag{47}$$

$$\frac{T_s}{T_w} = \left\{1 + \frac{K}{2kn}\sqrt{\frac{\pi m}{2kT}}\frac{\partial T}{\partial y} - 0.1667\,\frac{\mu}{p}\left(\frac{\partial \overline{u}}{\partial x} + \frac{\partial \overline{w}}{\partial z} - 2\frac{\partial \overline{v}}{\partial y}\right)\right\}_s \tag{48}$$

or $A = 0.8862$, $B = 0.3750$ and $C = 1.000$.

2. Present results, using Nocilla reflection model and hard spheres.

$$\bar{u}_s = -2S_r\sqrt{\frac{kT_r}{m}}\sin\theta_r + \left\{0.8862\,\frac{\mu}{n}\sqrt{\frac{2}{mkT}}\left(\frac{\partial\bar{u}}{\partial y}+\frac{\partial\bar{v}}{\partial x}\right)\right.$$

$$\left. +0.3750\,\frac{2\mu}{nm}\frac{1}{T}\frac{\partial T}{\partial x}\right\}_s \tag{49}$$

$$\frac{T_s}{T_w} = \left\{1+\frac{K}{2km}\sqrt{\frac{\pi m}{2kT}}\frac{\partial T}{\partial y} - 0.1667\,\frac{\mu}{p}\left(\frac{\partial\bar{u}}{\partial x}+\frac{\partial\bar{w}}{\partial z}-2\frac{\partial\bar{v}}{\partial y}\right)\right\}_s \times$$

$$\left\{\left(1+\frac{S_r^2}{2}\right)+\sqrt{\frac{\pi}{2}}\,\frac{\sigma_r(1+\operatorname{erf}\sigma_r)}{\chi(\sigma_r)}\right\} \tag{50}$$

3. The traditional approach using accommodation coefficients and Patterson's technique is given by Shidlovskiy (ref. 6). θ = percentage of diffusely reflected molecules and α is the thermal accommodation coefficient.

$$\bar{u}_s = \left\{0.8862\left(\frac{2-\theta}{\theta}\right)\frac{\mu}{n}\sqrt{\frac{2}{mkT}}\left(\frac{\partial\bar{u}}{\partial y}+\frac{\partial\bar{v}}{\partial x}\right)+0.3750\,\frac{2\mu}{nm}\frac{1}{T}\frac{\partial T}{\partial x}\right\}_s \tag{51}$$

$$\frac{T_s}{T_w} = \left\{1+\frac{2-\alpha}{\alpha}\,\frac{K}{2kn}\sqrt{\frac{\pi m}{2kT}}\frac{\partial T}{\partial y} - 0.1667\,\frac{\mu}{p}\left(\frac{\partial\bar{u}}{\partial x}+\frac{\partial\bar{w}}{\partial z}-2\frac{\partial\bar{v}}{\partial y}\right)\right\}_s \tag{52}$$

These results are identical to those of Chrusciel, et al. (ref. 18) except that their results were nondimensionalized.

Kogan summarized other determinations of the coefficients A, B and C. This is repeated in Table I.

The new slip formulation (equations (38), (39), (27), (49) and (50)) contains the parameters S_r, U_r or T_r and θ_r that must be specified before a problem can be solved. The parameters used will depend upon the problem being solved. That is, flow in a nozzle has the free stream flow tangent to the surface at all locations and it would be expected that θ_r would be close to 90° or greater, S_r would be large and T_r would be less than T_w. On the other hand, for the flow over a blunt body, the parameters would be a function of θ_i and various forms could be assumed, much as

Hurlbut and Sherman did (ref. 32) for sphere drag, and ranges of the values could be examined for comparison with experiments. Notice that the form of the boundary conditions is somewhat different from those that include the accommodation coefficients. The slip velocity contains an subtractive term which is of the order of U_r. The scattering experiments have shown that U_r is never larger than the mean thermal velocity of the reemitted gas. For tangential flow over a surface this is small compared to the mean thermal velocity of the gas at the wall temperature. Since this term is negative it diminishes the magnitude of the usual slip boundary condition. This was also found by Chrusciel, et al. (refs. 21 and 35) using Monte Carlo methods to solve the flow over a blunt-nosed body in hypersonic flow. The temperature jump expression is multiplied by an expression that goes between the values 0.82 at $\sigma_r = -0.20$ to 1.44 for $\sigma_r = 0.50$. Values less than 1.0 would be required to match the results of Chrusciel, et al.

Table I

Coefficient	Value	Method of Determination
A	1.012	Krook equation
B	0.383	Krook equation
C	1.173	Krook equation
A	1.047	Boltzmann equation Maxwell molecules
A	1.015	``
B	0.383	``
B	0.402	``
A	1.103	Boltzmann equation, hard spheres
B	0.4456	``
B	0.3292	``
C	1.113	``
A	0.8862	Present results
B	0.3750	Present results
C	1.000	Present results

CONCLUSIONS AND RECOMMENDATIONS

A new form of the slip flow boundary conditions has been developed which appears to have characteristics that can improve the agreement between the computations and measurements. However, the boundary conditions have been obtained in an approximate manner and they contain unknown parameters. Therefore, the following steps should be taken to complete the task of obtaining useful slip flow boundary conditions.

1. The analysis of the previous section should be repeated using the Chapman-Enskog expansion rather than the approximate solution of Patterson.

2. Since the Chapman-Enskog expansion assumes that the flow is close to equilibrium, the analysis should be repeated by solving the Boltzmann equation close to the wall. It is not anticipated that this could be done exactly but a series expansion would be required. The boundary condition would then apply for flows arbitrarily far from equilibrium (ref. 21).

3. The boundary condition should be used on a simple flow, such as Poiseville flow through a circular capillary. This problem is well understood theoretically (ref. 36) and has an extensive data base (refs. 37 and 38). Therefore, it would be a good test case for the new boundary conditions and would help develop a rational for determining the parameters in the equations.

4. Once the parameters have been obtained then the equations can be applied to the calculation of low density nozzle flows. For low throat Reynolds numbers ($Re^* < 100$) the interaction of the wall viscous region with the nozzle core is so great that only the full Navier Stokes equations can be used to adequately solve the flow field. Therefore, an axisymmetric Navier Stokes code, such as that of Lombard (ref. 39), will be used with the new slip boundary conditions to calculate nozzle flow fields and compare the properties with any existing measurements that can be found (ref. 38).

REFERENCES

1. Schaaf, S. A., "Mechanics of Rarefied Gases," Handbuch der Physik VIII/2, 1963, p. 591.
2. Probstein, R. F., "Continuum Theory and Rarefied Hypersonic Aerodynamics," Rarefied Gas Dynamics, Proc. First Int. Symp., Pergamon Press, New York, 1960, p. 416.
3. Schaaf, S. A. and Chambre, P. L., "Flow of Rarefied Gases," Fundamentals of Gas Dynamics, Emmons, H. W., Editor, Princeton University Press, Princeton, NJ, 1958, p. 687.
4. Potter, J. L., "The Transition Rarefied-Flow Regime," Rarefied Gas Dynamics, Proc. Fifth Int. Symp., Academic Press, New York, 1967, p. 881.
5. Potter, J. L., "Transitional, Hypervelocity Aerodynamic Simulation and Scaling in Light of Recent Flight Data," AIAA-85-1028, presented at AIAA 20th Thermophysics Conference, June 1985.
6. Shidlovskiy, V. P. Introduction to the Dynamics of Rarefied Gases, American Elsevier Pub. Co., New York, 1967.
7. Bailey, A. B., et al., "Flow Field Mapping of Gas/Particle Nozzle Expansion Into Vacuum," AEDC-TR-84-38, July, 1985.
8. Kogan, M.N., Rarefied Gas Dynamics, Plenum Press, New York, 1969.
9. Clarke, J.F. and McChesney, M., The Dynamics of Real Gases, Butterworths, London, 1964.
10. Hurlbut, F. C., "Aerospace Applications of Molecular Beams," Entropie 18, Nov-Dec, 1967, pp. 98–110.
11. Kennard, E. H., Kinetic Theory of Gases, McGraw-Hill Book Co., New York, 1938.
12. Wachman, H. Y., "The Thermal Accommodation Coefficient: A Critical Survey," Am. Rocket Soc. J. 32, 2 (1962).
13. Hurlbut, F. C., "Current Developments in the Study of Gas-Surface Interactions," Rarefied Gas Dynamics, Proc. Fifth Int. Symp., Academic Press, New York, 1967, p. 1.
14. Thomas, L. B., "Thermal Accommodation of Gases on Solids," in Fundamentals of Gas-Surface Interactions, H. Saltsburg, et al., Editors, Academic Press, New York 1967, p. 346.

15. Mair, W. N., Viney, B. W. and Colligon, J. S., ``Experiments on the Accommodation of Normal Momentum,'' Rarefied Gas Dynamics, Proc. Fifth Int. Symp., Academic Press, New York, 1967, p. 187.

16. Abuaf, A. and Marsden, D. G. H. ``Momentum Accommodation of Argon in the 0.06 to $5eV$ Range,'' Rarefied Gas Dynamics, Proc. Fifth Int. Symp., Academic Press, New York, 1967, p. 199.

17. Seidl, M. and Steinheil, E. ``Measurement of Momentum Accommodation Coefficients on Surfaces Characterized by Auger Spectroscopy, SIMS and LEED,'' Rarefied Gas Dynamics, Proc. Ninth Int. Symp., DFVLR Press, Germany 1974, p. E.9.

18. Knechtel, E. D. and Pitts, W. C. ``Normal and Tangential Momentum Accommodation for Earth Satellite Conditions,'' Astro. Acta 18, 171 (1973)

19. Fan, C. and Warr, J. W., ``Gas-Surface Interactions and Orbital Aerodynamic Calculations,'' Lockheed Rept. HREC D 162–228 (1970).

20. Epstein, M., ``Effect of Incomplete Accommodation on the Slip Coefficient,'' Aerospace Corp. Tech. Rept. TR-0158 (3240-20)-12 (1968).

21. Chrusciel, G. T., Lewis, C. H. and Sugimura, T., ``Slip Effects in Hypersonic Rarefied Flow,'' Rarefied Gas Dynamics, Proc. Twelvth Int. Symp., AIAA, New York, 1981, p. 1040.

22. Miller, D. R. and Subbarao, R. B., ``Scattering of $0.06-2.5eV$ Neon and Argon Atoms from a Silver (111) Crystal,'' J. Chem. Phys. 52, 425 (1970).

23. Bishara, M. N. and Fisher, S. S., ``Observed Intensity and Speed Distributions of Thermal-Energy Argon Atoms Scattered from the (111) Face of Silver,'' J. Chem. Phys. 52, 5661 (1970).

24. Subbarao, R. B. and Miller, D. R., ``Accommodation Coefficients and the Gas-Surface Boundary Conditions,'' Rarefied Gas Dynamics, Proc. Seventh Int. Symp., Editrice Tecnico Scientifica, Posta, Italy, 1971, p. 223.

25. Hayes, W. J., Rogers, W. E. and Knuth, E. L., ``Scattering of Argon Beams with Incident Energies up to 20eV from a (111) Silver Surface,'' J. Chem. Phys. 56, 1652 (1972).

26. Weinberg, W. H. and Merrill, R. P., ``Scattering of Helium, Neon, Argon, Krypton, Xenon and Deuterium from a Tungsten (110) Surface Characterized by LEED,'' J. Chem. Phys. 56, 2881 (1972).

27. Subbarao, R. B. and Miller, D. R., ``Velocity Distribution Measurements of $0.06 - 1.4eV$ Argon and Neon Atoms Scattered from the (111) Plane of a Silver Crystal,'' J. Chem. Phys. 58, 5247 (1973).

28. Sau, R. and Merrill, R. P., ``The Scattering of Hydrogen, Deuterium, and Rare Gases from Silver (111) Single Crystals,'' Surface Sci. 34, 268 (1973).

29. Jih, C.T.R., and Hurlbut F. C., ``Time-of-Flight Studies of Argon Beams Scattered from a Silver (111) Crystal Surface,'' Rarefied Gas Dynamics, Proc. Tenth Int. Symp., AIAA, New York, 1977, p. 539. (1976).

30. Jakus, K. and Hurlbut, F. C., ``Gas Surface Scattering Studies Using Nozzle Beams and Time-of-Flight Techniques,'' Rarefied Gas Dynamics, Proc. Sixth Int. Symp., Academic Press, New York, 1969, p. 1171.

31. Nocilla, S. ``The Surface Re-emission Law in Free Molecule Flow,'' Rarefied Gas Dynamics, Proc. Third Int. Symp. pp. 327–346 (1963).

32. Hurlbut, F. C., and Sherman, F. S., ``Application of the Nocilla Wall Reflection Model to Free-Molecule Kinetic Theory,'' Phys. Fluids 11, 486 (1968).

33. Patterson, G. N., Molecular Flow of Gases, John Wiley and Sons, New York, 1956.

34. Kogan, M. N., ``Molecular Gas Dynamics,'' Ann. Rev. Fluid Mech. 5, 383 (1973).

35. G. T. Chrusciel, and L. A. Pool, ``Knudsen Layer Characteristics for a Highly Cooled Blunt Body in Hypersonic Rarefied Flow,'' AIAA J. 23, 826 (1985).

36. Shidlovskiy, V. P., ``Special Case of Viscous Gas Motion in Cylindrical Tube in Slip Flow Regime,'' Rarefied Gas Dynamics, Proc. Sixth Int. Symp., Academic Press, New York, 1969, p. 215.

37. Lund, L. M. and Berman, A. S., ``Flow and Self-Diffusion of Gases in Capillaries,'' J. Appl. Phys. 37, 2489, 2496 (1966).

38. R. H. Edwards, ``Low Density Flows Through Tubes and Nozzles,'' Rarefied Gas Dynamics, Proc. Tenth Int. Symp., (1976), pp. 199-223.

39. Lombard, C. K., et al., ``Multi-Dimensional Formulation of CSCM-An Upwind Flux Difference Eigenvector Split Method for the Compressible Navier-Stokes Equation,'' AIAA-83-1985, presented at AIAA Sixth Computational Fluid Dynamics Conference, July 13–15, 1983.

1985

NASA/ASEE SUMMER FACULTY RESEARCH FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

PRELIMINARY DESIGNS FOR X-RAY SOURCE MODIFICATIONS FOR THE MARSHALL SPACE FLIGHT CENTER'S X-RAY CALIBRATION FACILITY

Prepared By:	W. L. Croft, Ph.D.
Academic Rank:	Professor
University and Department:	Mississippi State University Department of Physics
NASA/MSFC:	
Laboratory:	Test
Division:	Systems and Components Test
Branch:	Environmental Test
MSFC Counterpart:	J. C. Reily, Jr.
Date:	August 22, 1985
Contract No.:	NGT 01-008-021 The University of Alabama in Huntsville

PRELIMINARY DESIGNS FOR X-RAY SOURCE MODIFICATIONS FOR THE MARSHALL SPACE FLIGHT CENTER'S X-RAY CALIBRATION FACILITY

BY

W. L. Croft
Professor of Physics
Mississippi State University
Mississippi State, Mississippi

ABSTRACT

The objective of this investigation is to develop preliminary designs for modifications to the X-ray source of the MSFC X-Ray Calibration Facility. Recommendations are made regarding: (1) The production of an unpolarized X-ray beam, (2) Modification of the source to provide characteristic X-rays with energies up to 40 keV, and (3) Addition of the capability to calibrate instruments in the extreme ultraviolet wavelength region.

ACKNOWLEDGEMENTS

I would like to express my appreciation to the NASA/ASEE Summer Faculty Research Fellowship Program for the opportunity to participate. Thanks are extended to Dr. Gerald Karr, Dr. Mike Freeman, Dr. Jim Dozier and Mr. Leroy Osborn for providing the outstanding program at MSFC and for their cooperation which made my participation possible. I appreciate the hospitality extended me by Messrs. H. Coldwater, J. H. Newton and W. E. Dickson, and their staffs. Thanks are due Dave Watson for helpful discussions and assistance, and Dr. Richard Hoover of the Space Science Laboratory for valuable discussions. Special thanks go to my NASA counterpart, Mr. Cary Reily, for his friendship, hospitality and helpful discussions regarding this work. Special thanks are also extended to Mrs. Pat Blackmon and Mrs. Glynda Thomas who typed this report.

UNPOLARIZED X-RAY SOURCE

Requirements for a source to produce an X-ray beam which is 100% unpolarized were investigated by J. H. Kitterman.[1] He made an extensive survey of commercial X-ray tubes and found that an end-window toroidal cathode tube manufactured by Machlett Laboratories should have the desired properties if electrons are emitted from the cathode uniformly and are focused onto the anode symmetrically. Important features of the tube are illustrated in Figure 1.

This tube was designed as a source for use in X-ray fluorescence spectrometry and is said to provide an intense broad-spectrum X-ray beam. The theory of production of continuous (bremsstrahlung) X-rays is given by Compton[2] and Dyson[3] and is summarized by Kitterman. According to this theory, the plane of polarization of X-rays produced at points 180° apart on the annular shaped area of the flat target which is bombarded by electrons will be in the same direction but will be perpendicular to that of X-rays produced at the 90° and 270° positions. Thus when contributions to the X-ray beam are summed around the annular area of electron bombardment, the result should be an unpolarized beam if the source is axially symmetric.

While the Machlett tube should produce an unpolarized X-ray beam it is not without problems. The tube was designed to operate between 20 and 75 kV and Kitterman reports that the designer of the tube is uncertain if the intensity will be adequate at much lower voltages. Since the operating characteristics of this tube at low voltages have not been determined, the spot-size (area of X-ray emission) is unknown. The spot-size for normal operating voltages is reported to be 1.0 to 1.5 Cm o.d. A spot this size would subtend an angle at the optical bench pivot which is 10 to 15 times as large as the ~0.7 arc seconds of the present source. This spot size would appear to be too large for some applications. Targets are not easily interchangeable in this tube so a series of tubes with targets of different elements would be required to provide characteristic X-rays of the desired energies. Since tubes cost approximately $12,000 each, this series of tubes would be expensive.

Kitterman pointed out that another possible X-ray source is a modified electron-beam-evaporator and he obtained literature on one produced by Vacuum Generators. This model has a toroidal cathode and its geometry is basically the same as that of the Machlett tube. However, it is designed for easy disassembly and has a crucible for holding the different materials which are to be evaporated. A brief search of the literature revealed only one example of a modified electron beam evaporator being used as an X-ray source and it was this model by Vacuum Generators.[4] No details of modifications or operating characteristics were reported. However, it should be a straight-forward procedure to modify this evaporator to accept interchangeable flat faced target anodes which would screw onto the water-cooled stem where the crucible is normally located.

This evaporator is designed to operate at voltages up to 10 Kv but the literature indicates that in practice these evaporators are usually operated below 5 kV to reduce arcing in the vapors which are produced when materials are being evaporated. In applications as an X-ray source where target evaporation is avoided, operation at the 10 kV upper limit of the power supply should be possible. Information from the factory indicates that it might be possible to push this evaporator to a maximum voltage of 15 kV which is the rating of the high voltage electrical feedthroughs.[5] This low maximum operating voltage is a serious limitation on the useful energy range of this X-ray source. It does have a focus adjustment which allows a spot size of 1 mm diameter. This evaporator would be a much more economical X-ray source since the evaporator and power supply cost less than one Machlett tube.

Inquiries to several suppliers of electron beam evaporators led to one other one for possible use as an X-ray source. Dr. R. Bakish of Bakish Materials Corporation indicates that his line of evaporators is often used as X-ray sources.[6] An axial electron-gun with magnetic focusing and magnetic beam turning through 90° will result in a geometry which produces unpolarized X-rays. This is due to the fact that when a parallel beam of electrons strikes a flat target perpendicular to the plane of the target-face that there is no preferred direction of scatter of the electrons. The X-rays, produced by these randomly scattered electrons, which are observed at 0° and 180° with respect to the preimpact velocity of the electrons will be unpolarized due to symmetry.[7] Figure 2 shows a sketch of such an arrangement. An electron beam focused by a magnetic lens in the electron gun is bent through 90° by a uniform magnetic field external to the gun and perpendicular to the original beam direction.

The main advantage of the Bakish system is that electron guns that operate up to 60 kV are available. A discussion of electron-guns and electron beam apparatus is contained in books by Bakish[8] and Schiller, et.al.[9]

The 90° rotation of the unpolarized source which is required to check for asymmetries should be obtained by mounting the source on a special rotatable vacuum flange such as one manufactured by Ferrofluidics.

The X-ray Calibration Facility provides primarily characteristic X-rays which are known to be unpolarized.[10] These characteristic X-rays are superimposed on the bremsstrahlung continuum which is partially polarized. Thus it is only the continuum X-rays which contribute to the polarization of the present X-ray beam. Kitterman estimates that polarization of the present beam from an Al target is only 0.3%. Estimates of the polarized X-ray component in the beams from other targets should be made since Reily reports that the Al line is one of the cleanest.[11]

In the opinion of the author of this report, it is highly unlikely that the "unpolarized sources" discussed by him or Kitterman will be known to be 100% unpolarized with an accuracy greater than 0.3%. While the author of this report is not an expert on polarization measurements of celestial X-ray sources, he is not aware of measurements with an uncertainty as small as 0.3%. Therefore, I suggest that careful discussions with principal investigators be held to be sure that it is really

necessary to go to the considerable trouble and expense to try to improve on the present source as a generator of an unpolarized X-ray beam. In my opinion, the polarization characteristics of the present beam should be measured with a polarimeter before proceeding to other alternatives.

Since the unpolarized characteristic X-rays are superimposed on the partially polarized continuum, the relative amount of polarized X-rays accepted will be a function of counter resolution and "window-width." These can be varied in the experiment to study the polarized contribution to the X-ray beam.

40-KEV X-RAY MODIFICATION

In order to extend the energy capability of the X-ray facility to 40 keV, modifications will have to be made in the X-ray source and beam monitor. The efficiency of Production of Characteristic K X-rays is proportional to: $(E_o/E_k-1)^{1.63}$, where E_o is the energy of the projectile electrons and E_k is the binding energy of the K-shell orbital electrons.[12] From this, it is clear that for the efficient production of K X-rays at 40 keV, accelerating voltages of at least 80 kV will be necessary. Since the high voltage power supply for the present source has a maximum value of 60 kV it will be necessary to obtain a new power supply with perhaps a 100 kV maximum output.

The energies of the Characteristic K X-rays increase with the atomic number, Z, of the target material. Thus targets of higher Z-values than those used in the past will have to be obtained. To help select promising target and filter materials, I have surveyed the elements in the periodic table of the elements starting from where the HEAO-2 targets stopped out through samarium which has a K X-ray energy of about 40 keV. Table I includes a list of these elements with comments regarding their possible usefulness as targets.

These higher X-ray energies will require that a different beam monitor be used. The Xenon proportional counter designed to operate at energies up to 10 keV for the AXAF calibration will not be adequate.[21] Extrapolation of the efficiency vs. energy curve for this Xe counter indicates that its efficiency is essentially zero at approximately 22 keV. Two possible detectors for use up to 40 keV are NaI(Tl) crystals coupled to photomultiplier tubes, and solid state detectors. Thin NaI(Tl) scintillation crystals coupled to photomultiplier tubes are available from Tennelec and Harshaw. These detectors are rugged, reliable, have moderate energy resolution and are relatively inexpensive. High purity germanium (HPGe) solid state detectors can provide excellent efficiency from approximately 3 keV to energies well above the 40 keV of interest. These HPGe detectors can be shipped and stored at room temperature without damage but are operated at the temperature of LN_2. These HPGe detectors have a great advantage over the older lithium drifted germanium detectors which were ruined when brought up to room temperature. These detectors

TABLE I SOME INFORMATION RELATIVE TO ELEMENTS SUITABLE AS X-RAY TARGETS

CHEMICAL SYMBOL	ATOMIC NO.	ENERGY $K_{\alpha 1}$ (kev)[13]	USEFUL AS TARGET?	ELECTRO-PLATE?	REFERENCES	COMMENTS[14]
Co	27	6.93	yes	yes	15(p.328)	
Ni	28	7.48	yes	yes	15(p.317)	have filter from HEAO-2
Cu	29	8.05	yes	yes		have from HEAO-2
Zn	30	8.64	yes	yes	16(p.198-207)	
Ga	31	9.25	no	no	20(p.71)	(Alloys with metals!) liquid near room temperature
Ge	32	9.89	maybe	?	20(p.71)	Vacuum evaporate
AS	33	10.54				
Se	34	11.22			20(p.74)	may contaminate vacuum system
Br	35	11.92	no	no		liquid at room temperature
kr	36	12.65	no	no		gas
Rb	37	13.40	no	no		liquid near room temperature
Sr	38	14.17	no	no		keep under kerosene, oxidizes rapidly
Y	39	14.96				relatively stable in air
Zr	40	15.77	yes			have from HEAO-2
Nb	41	16.62		?	20(p.73)	vacuum evaporate
Mo	42	17.48	yes			known to have been used as X-ray target
TC	43	18.37	no			not found in nature
Ru	44	19.28	yes	yes	18(p.227)	in platnium group
Rh	45	20.22	yes	yes	18,21(p.167)	proposed source at 2.6-kev
Pd	46	21.18	yes	yes	18(p.213)	platnium group
Ag	47	22.12	yes	yes		have from HEAO-2
Cd	48	23.17	yes	yes	16(p.95-102)	
In	49	24.21	yes	yes	17(p.161)	
Sn	50	25.27	yes	yes		have from HEAO-2
Sb	51	26.36	yes	yes	17(p.150)	stable in air at room temperature
Te	52	27.47	?		20(p.75)	P-type semiconductor
I	53	28.61	no	no		volatizes at room temperature
Xe	54	29.78	no	no		gas
CS	55	30.97	no	no		liquid near room temperature
Ba	56	32.19	no	no	19(p.512)	(vacuum evaporate BaO) oxidizes easily, keep under oil
La	57	33.03	no	no		oxidizes rapidly in air
Ce	58	34.28	no	no		oxidizes readily at room temperature
Pr	59	36.03	no	no		keep in oil or sealed in plastic
Nd	60	37.36	no	no		keep in oil, quickly oxidizes in air
Pm	61	38.72	no	no		not naturally occuring
Sm	62	40.12	maybe			reasonably stable in air, used as a neutron absorber in nuclear reactors

are rugged, reliable, have excellent energy resolution and are very expensive. Two sources of these detectors are ORTEC and Canberra. An efficiency curve for a Canberra planar HPGe detector is shown in Figure 3.

Experimenters need to be aware that electrical shock hazards and radiation hazards may be greater at these higher voltages and thus should exercise appropriate precautions.

ULTRAVIOLET SOURCE

There is growing interest by astronomers in studying the universe in the extreme ultraviolet wavelengths. Measurements must be made from above the earth's atmosphere which is opaque at these energies and orbiting observatories are the most effective means for long term studies. Figure 4, adapted from J.A.R. Samson's classic work on vacuum ultraviolet spectroscopy indicates the relationship between the Vacuum UV, Extreme UV and SOFT X-Ray regions of the electromagnetic spectrum.[22] It should be noted, however, that these wavelength boundaries are not universally accepted.

Wavelength specifications of some Vacuum UV instruments on several orbiting observatories are shown in Table II. Results of International Ultraviolet Explorer investigations of comets, satellites, planets, the sun, stars of all types, supernova remnants, the interstellar medium, nebulae, clusters, galaxies, quasars, etc. in the wavelength range 1150-3200 Å have been reported.[16] Astronomers now want to study these and other interesting objects in the extreme UV region. At this time the Extreme Ultraviolet Explorer (EUVE) is being developed to survey the entire celestial sphere for astronomical sources in the extreme UV region (100-1000 Å). An interest has been expressed in extending the capability of the MSFC X-Ray Calibration Facility for instrument calibration into the extreme UV region.[25] This extension would be reasonable and the possibility deserves consideration since the EUV region overlaps the soft X-ray region where a capability already exists.

Although facility requirements in this energy range have not been well defined, it is clear that a source of EUV radiation, a grazing incidence monochromator to select energies, and a detector to measure the flux will be ncesssary. As a modest first step in investigating the possibilities of developing a capability in the EUV region, I have surveyed suppliers of UV equipment to find what items are commercially available and what companies can supply them.

Several Samson designed UV sources are available from Minuteman Laboratories. However, it appears that perhaps the best source for this application is a continuous discharge Penning source with emission lines between 50 and 300 Å which is described in the literature but is not available commercially.[26] Grazing incidence monochromators for the EUV region are available from McPherson, Acton Research, Minuteman Laboratories and Instruments S.A. If off-the-shelf models are not satisfactory, these companies have considerable experience in supplying custom monochromators.

Detectors of several types including channel electron multipliers, multichannel plates, photodiodes, photomultiplier tubes, thin window proportional counters and ionization chambers are available from Galileo Electro-Optics and Minuteman. Appendix A contains a list of company names, addresses, telephone numbers and sales representatives of suppliers of VUV equipment. A number of companies listed in "Research and Development's Telephone Directory" and "Physics Today's Annual Buyers Guide" were contacted for descriptive literature. Only those companies with products that looked promising are listed.

Since some personnel of the X-Ray Calibration Facility are not familiar with EUV instruments and terminology, a list of definitions from manufacturer's literature and other sources are listed in Appendix B.

Table III lists some useful filter materials for the extreme ultraviolet region.[27]

RECOMMENDATIONS

1. I recommend that the polarization of X-rays from the current source be studied to see if they might meet the needs for an unpolarized source. These X-rays are primarily characteristic X-rays which are inherently unpolarized and may satisfy the requirements within the uncertainty limits of some measurements.

2. If it becomes necessary to acquire an additional X-ray source for the production of low energy unpolarized X-rays, I recommend that a vacuum generators electron beam evaporator be purchased and modified as an X-ray source. It would be worthwhile to determine if this apparatus can be modified to operate at significantly higher voltages than the 15-kv upper limit currently set by the rating of the electrical feed throughs. An axial electron gun such as that of Bakish Materials Corporation is worth consideration as a source of high energy unpolarized X-rays.

3. If there is continued interest in the calibration of instruments in the extreme ultraviolet region, I recommend that serious consideration be given to developing the capability to perform calibration in this region. This would appear to be a natural extension of the present capability. A first step should probably be to consult experts in optics and UV spectroscopy here at MSFC for recommendations regarding sources, monochromators and detectors.

TABLE II
WAVELENGTH AND ENERGY RANGES FOR SOME UV AND X-RAY EXPERIMENTS

EXPERIMENT	LAUNCH DATE	WAVELENGTH RANGE	ENERGY RANGE
Apollo-Soyuz[26]	July 1975	50-1000Å	12.4 - 248ev
IUE[23]	Jan 26,1978	1150-3200Å	3.9 - 10.8ev
HEAO-2	Nov. 1978	2-67Å	0.185 - 6.2kev
Space Telescope	1986	1100-3200Å	3.9 - 11.3ev
EUVE[24]	198_	100-1000Å	12.4 - 124ev
AXAF	199_	1.24 - 100Å	0.1 - 10kev

TABLE III
FILTERS FOR THE EXTREME ULTRAVIOLET REGION[26]

MATERIAL	BAND ev	Pass nm
Parylene N	83 - 225	5.5 - 15.0
Be/Parylene N	83 - 109	11.4 - 15.0
Aluminum Plus Carbon	20 - 73	17 - 62
Tin	16 - 25	50 - 78
Barium Fluoride	8.0 - 9.2	135 - 154

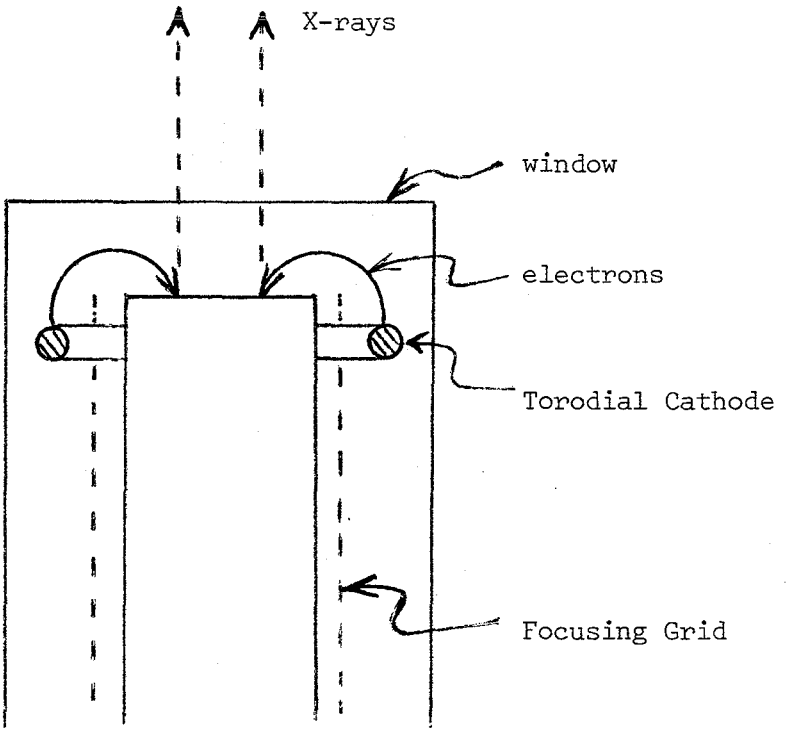


Figure 1 FEATURES OF AN END WINDOW TORODIAL X-RAY TUBE

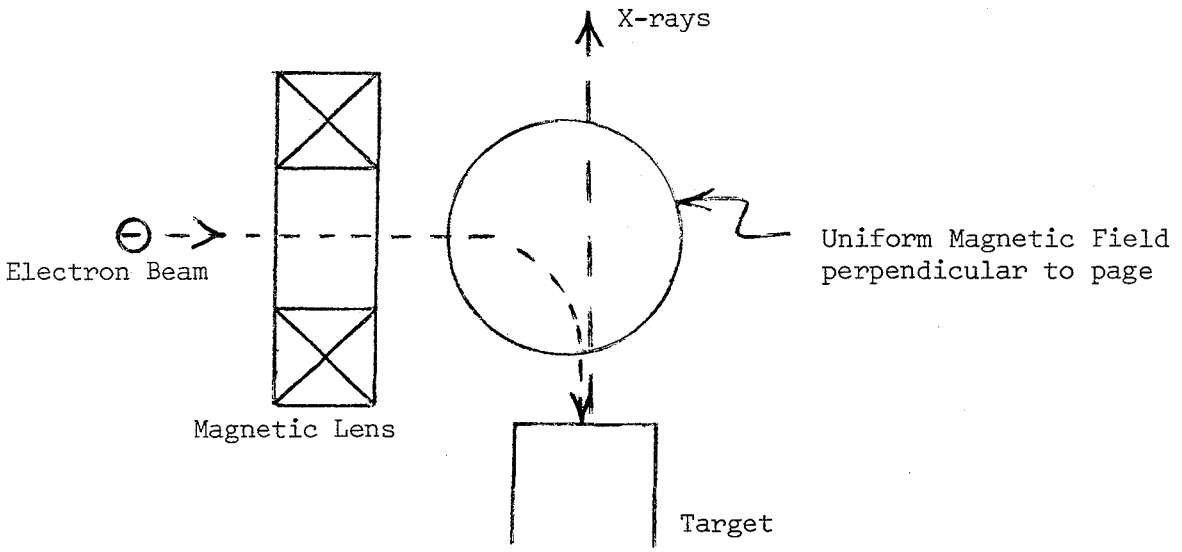


Figure 2 Axial Gun with Magnetic Focusing and Magnetic Beam Turning by 90°

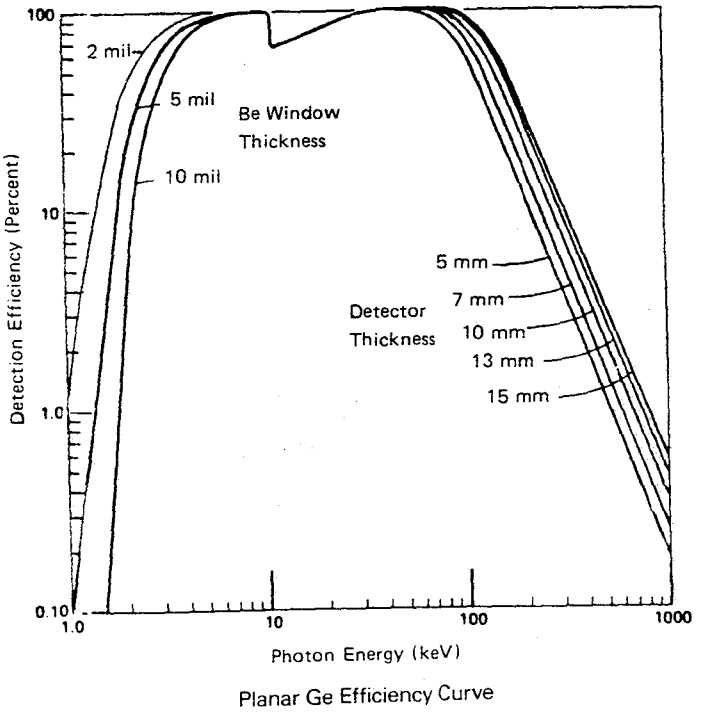


Figure 3 Planar Ge Efficiency Curve

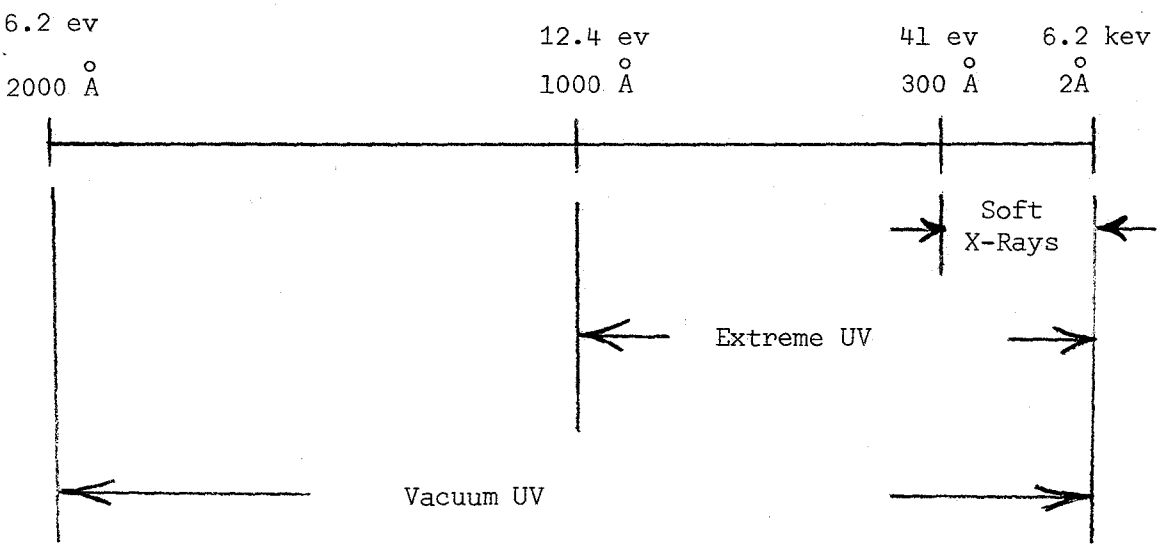


Figure 4 Energy and Wavelength Relationships in the Vacuum Ultraviolet

APPENDIX A

SOURCES OF SOME EQUIPMENT OF INTEREST

Electron Guns

Dr. R. Bakish
Bakish Materials Corporation
171 Sherwood Place
P.O. Box 148
Englewood, N.J. 07631
(201-567-5873)

Mr. Mike Flinko, Product Manager
UHV Components
Kurt J. Lesker Company
5635 Horning Road
Pittsburg, PA 15236
(412-655-9500)
(800-245-1656)
(Exclusive Distributor for Vacuum Generators, LTD)

Grazing Incidence Monochromators

Mr. Richard W. Merk
Operations Manager
Acton Reserach Corporation
Box 215
525 Main Street
Acton, Mass. 01720
(617-263-3584)

Mr. John Gilmore
Vice President
Minuteman Laboratories, Inc.
916 Main Street
Acton, Mass. 01702
(617-263-2632)

Mr. Jack Parmley
Sales Representative
S.I./McPherson
530 Main Street
Acton, Mass. 01720
(617-263-7733)
(800-255-1055)

Thin Window NaI(T1) X-Ray Detectors

Mr. J. E. Bradley
Sales Engineer
Tennelec, Inc.
601 Oak Ridge Turnpike
Oak Ridge, TN 37830-2560
(615-483-8405)

Rotatable Vacuum Flange

Ferrofluidics Corporation
40 Simon Street
Nashua, NH 03061
(603-883-9800)

Detectors for Extreme Ultra Violet Radiation

Mr. Richard E. Shepardson
Contracts Administrator
Galileo Electro-Optics Corp.
Galileo Park
Sturbridge, Mass. 01518
(617-347-9191)

APPENDIX B

DEFINITIONS RELATED TO MONOCHROMATORS

Aberrations are imperfections in slit image formation resulting from optical design, limitations of a configuration, or imaging at high aperature.

Aperture or f/Number is the ratio of the focal length of the monochromator to the grating diameter. The smaller the f/number, the larger the aperture and vice versa. A large aperture features a wide collection angle - making it more effective in gathering light while less effective in resolving it. A small aperture gathers only a small amount of light but finely resolves it.

Bandpass in nm is the actual resolution of a monochromator as a function of its slit width. It is the product of linear dispersion and the slit width.

Blaze Wavelength is the wavelength at which the grating is at its maximum efficiency. The useful range of a grating can be described by the "2/3-3/2 Rule" which gives the range of a grating to be: lower limit = $2/3\ \lambda_{blaze}$; upper limit = $3/2\ \lambda_{blaze}$. It is sometimes possible to operate the grating with reasonable efficiency above the 3/2 value, but operation below the 2/3 value is not recommended.

Dispersion is classically defined as the amount of the focal plane, expressed in mm, taken by one nm of light. Now it is more common to refer to dispersion as how well the monochromator spreads the light spectrum over the focal plane of the exit plane, this is expressed in the amount of spectrum (in nm) over a single mm of the focal plane.

Focal Length is the distance between the slit and the focusing component of the monochromator. Generally, the longer the focal length, the greater the linear dispersion.

Holographic grating is a grating produced by interference fringes from a laser beam. The holographic process produces gratings which are virtually free of spacing errors. Holographic gratings have been produced with efficiencies very near those of classically ruled gratings. Holographic gratings are most useful when a more dense groove spacing is required and efficiency is not of great concern.

Order number is an integer representing multiples of a given wavelength. When a monochromator is set to pass a wavelength of 800 nm, for example, integral multiples of other wavelengths (such as the second order of 400 nm) will be allowed to pass. In cases where transmission of higher orders causes a problem, bandpass filters can remove unwanted radiation.

Resolution is a measure of how finely a given monochromaton differentiates between spectral lines - the minimum detectable difference between peaks. While theoretically resolution can be approximated by multiplying slit width (in mm) by dispersion (in nm/mm), actual resolution rarely equals this due to aberations inherent in monochromators.

Rowland Circle is a circle which contains the concave grating and the entrance and exit slits in most monochromators. The circle has a diameter equal to the radius of the grating which is mounted tanget to the circle.

Throughput is the overall effectiveness with which a monochromator transmits light. It is defined as the ratio of the amount of light passing into the entrance slit to the amount of light passing out of the exit slit. Mathematically this is defined as:

$$T = \frac{\pi D h}{4(f/number)^2 R_1 R_2 R_3 \cdots R_n R_x}$$ where

D = dispersion (expressed classically as mm/nm), h= slit height in mm, $R_1 \ldots R_n$ = reflectivities of various mirror surfaces and R_x = the spectral efficiency of the grating.

REFERENCES

1. Kitterman, J. H., NASA/ASEE Summer Faculty Research Fellowship Program, Contract No. NASA-NGT-01-002-099, The University of Alabama, August 1984.

2. Compton, A. H. and Allison, S.K., X-rays in Theory and Experiment 97-115, D. Van Nostrand, New York, 1935.

3. Dyson, N.A., X-rays in Atomic and Nuclear Physics, 7-61, Longman Group Ltd., 1973.

4. E. Spiller, R. Feder and J. Topalian, "Proceedings of the V International Conference on Vacuum Ultraviolet Radiation Physics," Montpellier, September 1977 (eds. M. C. Castex, R. Pinchaux and M. Pouey, Journal DE Physique Volume 39), 1978, C4-205.

5. Flinko, Mike: Private Communication, July 1985.

6. Bakish, R.: Private Communication, June 1985.

7. Dyson, N.A., X-rays in Atomic and Nuclear Physics, 48, Longman Group Ltd, 1973.

8. Bakish, Robert, Editor, Introduction to Electron Beam Technology, John Wiley and Sons, New York, 1962.

9. Siegfried Schiller, Ullrich Heisig and Siegfried Pauzer, Electron Beam Technology, John Wiley and Sons, New York, 1982

10. Sproull, W.T., X-Rays in Practice, 49, McGraw Hill, New York, 1946

11. Reily, J. C. Jr.: Private Communication, August 1985.

12. Green, M., in X-ray Optics and X-Ray Microanalysis, (H. H. Pattee, V. E. Crosslett and Arne Engstrom, Eds.), 192, Academic Press, New York, 1963.

13. Burr, Alex, Handbook of Spectroscopy Volume I, 6-14, J. W. Robinson, CRC Press, Cleveland, 1975.

14. Hammond, C. R., Handbook of Chemistry and Physics, 56th Edition (Robert C. Weast, Ed.) B:6-42, CRC Press Cleveland, 1975.

15. Brugger, Robert, Nickel Plating, Robert Draper LTd., Teddington, 1970.

16. Mohler, J. B., Electroplating and Related Processes, Chemical Publishing Co., New York, 1969.

17. Ollard, E. A. and Smith, E. B., Handbook of Industrial Electroplating, American Elsevier Publishing Co., New York, 1964.

18. Fischer, J. and Weimer, D. E., Precious Metal Plating, Robert Draper Ltd., Teddington, 1964.

19. Holland, L., Vacuum Deposition of Thin Films, Chapman and Hall Ltd., London, 1963.

20. Chopra, K. L., Thin Film Phenomena, McGraw & Hill, New York, 1969.

21. Croft, W. L., NASA/ASEE Summer Faculty Research Fellowship Program, Contract No. NGT-01-008-21, The University of Alabama in Huntsville, August 1983.

22. Samson, J. A. R., Techniques of Vacuum Ultraviolet Spectroscopy, 2, John Wiley and Sons, New York, 1967.

23. Future of Ultraviolet Astronomy Based on Six Years of IUE Research, ed. J. M. Mead, R. D. Chapman and Yoji Kondo, NASA Conf. Publ. No. 2349, 1984.

24. S. Boyer, R. Malina, M. Lampton, R. Paresce, and G. Penegor, "The Extreme Ultraviolet Explorer," Ultraviolet and Vacuum Ultraviolet Systems, Proc. SPIE 279, 176-182 (1981).

25. Hoover, Richard; Private Communication, June 1985.

26. D. S. Finley, S. Bowyer, F. Paresce, and R. F. Malina, Appl. Opt. 18, 649 (1979).

27. S. Bowyer, B. Margon, M. Lampton, F. Paresce and R. Stern, "Extreme Ultraviolet Survey," Apollo-Soyuz Test Project, Vol. 1, NASA Science Report SP-412, 49-70 (1977).

1985

NASA/ASEE SUMMER FACULTY RESEARCH FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

CALCULATIONS ON THE PRODUCTION AND USE OF SUPERFINE HOLOGRAPHIC X-RAY GRATINGS FOR ASTROPHYSICAL OBSERVATIONS

Prepared by:	Paul L. Csonka, Ph.D.
Academic Rank:	Professor
University and Department:	University of Oregon Institute of Theoretical Science
NASA/MSFC:	
Laboratory:	Space Science Laboratory
Division:	Astrophysics
Branch:	High Energy Physics
NASA Counterpart:	Martin C. Weisskopf
Date:	August 30, 1985
Contract No.:	NGT 01-008-021 The University of Alabama

CALCULATIONS ON THE PRODUCTION AND USE OF SUPERFINE HOLOGAPHIC X-RAY GRATINGS FOR ASTROPHYSICAL OBSERVATIONS

by

Paul L. Csonka
Professor of Physics
Institute of Theoretical Science
University of Oregon
Eugene, Oregon 97403

ABSTRACT

Superfine holographic x-ray gratings may be produced by transferring onto metal an inteference pattern generated by two branches of a sufficiently coherent x-ray beam, emitted in the form of synchrotron radiation from high energy electron storage rings. Generation of the coherent beam requires restrictions on the beam size. A calculation is presented which allows an exact evaluation of this restriction. The effect of defocusing optics on the expected resolution is also studied.

1. INTRODUCTION/OBJECTIVES

Superfine holographic x-ray gratings are to be produced by transferring onto metal an interference pattern generated by two branches of a sufficiently monochromatic and coherent x-ray beam[1]. Such an x-ray beam can be obtained by selecting an appropriately monochromatized and coherent portion of a synchrotron radiaton beam emitted by circulating electrons in a high energy electron storage ring, such as SSRL at Stanford.

Selection of the appropriate beam portion requires first of all adequate monochromatization. Denoting by λ the wavelength of the recording x-ray radiation, and by $\Delta\ell$ the maximum pathlength difference between the two branches of the imprinting beam, one has to have

$$\frac{\Delta\lambda}{\lambda} \lesssim \frac{1}{2} f_1 \frac{\lambda}{\Delta\ell} . \qquad (1)$$

Here f_1 is a constant between zero and unity, to be specified below. Second, the angular divergence, $\Delta\theta$, needs to be restricted as

$$\Delta\theta \lesssim f_2^{1/2} \left(\frac{\lambda}{\Delta\ell}\right)^{1/2} , \qquad (2)$$

where $0 \leq f_2 \leq 1$ is a constant. Third, the cross-section of the beam has to be limited to achieve adequate coherence, which, in turn requires that the effective size of the radiation source (here circulating electrons) be constrained depending on the size of the area over which the interference pattern is to be established. That, of course, will reduce the number of photons which can be used to produce interference, and increase the time which has to elapse while the pattern is imprinted on a photosensitive material (such as PMMA). It is, therefore, important to accurately evaluate the necessary beam size, to be able to restrict the source area as required, but no more. A calculation will be described here which enables one to evalute the necessary restriction in beam size under quite general conditions, as illustrated in Fig. 1. Superfine holographic x-ray gratings are expected to achieve line densities $\gtrsim 10^5$ lines/mm. For an incident wavelength of λ = 100 Å at normal incidence on the grating, when the radiation is collected by an optical element with 1 arcsec accuracy, the intrinsic grating resolution is of the order of $\Delta\lambda/\lambda = 10^{-5}$. The question will be investigated to what extent this resolution can be approached in an actual instrument.

2. COHERENCE REQUIREMENTS

The notation is explained in Figs. 1 and 2. For any vector $\bar{v}$, we will write $|\bar{v}| = v$.

Referring to Fig. 2, one observes

$$L^2_{ik} = \left[-(\bar{L} + \bar{P}_{sk}) + \bar{P}_i\right]^2 \; ; \; i, k = 1, 2 \tag{1}$$

$$L_{ik} = \left\{(\bar{L} + \bar{P}_{sk})^2 - 2(\bar{L} + \bar{P}_{sk})\,\bar{P}_i + \bar{P}_i{}^2\right\}^{1/2}. \tag{2}$$

Introduce the notion

$$\bar{a}_k = \bar{L} + \bar{P}_{sk}, \tag{3}$$

and write a_{ky} and a_{kz} for the y and z component of $\bar{a}_k$. Since the coorindate axes are so chosen that the z component of P_1 and P_2 vanishes, one can rewrite (2) as

$$L_{ik} = \left\{a_{kz}{}^2 + (a_{ky} - P_i)^2\right\}^{1/2}. \tag{4}$$

The

$$\Delta L_1 = L_{21} - L_{11} = |a_{1z}| \left\{\left[1 + \frac{(a_{1y} - P_2)^2}{a_{1z}{}^2}\right]^{1/2} - \left[1 + \frac{(a_{1y} - P_1)^2}{a_{1z}{}^2}\right]^{1/2}\right\} \tag{5}$$

$$\xrightarrow[|a_{1z}| \to \infty]{} \frac{(a_{1y} - P_2)^2 - (a_{1y} - P_1)^2}{2|a_{1z}|}.$$

Next we find the maximum value of ΔL_1. Clearly ΔL_1 will reach its highest value for any chosen point P_{s1},

when the point P_2 is such that for it L_{21}^2 is maximum, and at the same time P_1 is such that L_{11}^2 is minimum.

If P_{s1} is located in Region 1, then the minimum value of L_{11}^2 occurs for that P_1, for which $P_1 = a_{1y}$, and then $L_{11} = a_{1z}$. To find that P_2 for which L_{21}^2 is maximum, note that for the choice i = 2, Eq. (4) reaches extremum as a function of P_2, when

$$P_2 = -a_{1y} , \tag{6}$$

and this extremum is a minimum. Since L_{21}^2 is a second order polynomial in P_2, L_{21}^2 can have no other extremum. Its maximum value as a function P_2 is therefore reached at the endpoints of the allowed range of variation for P_2, i.e., at $P_2 = \pm D_y/2$. By direct inspection, the correct choice is

$$P_2 = -\frac{a_{1y}}{|a_{1y}|}\frac{D_y}{2} , \tag{7}$$

so that the maximum of L_{21}^2 is

$$L_{21}^2 = \{a_1^2 + (|a_{1y}| + \frac{D_y}{2})\, D_y\}^{1/2} , \tag{8}$$

and the maximum over P_1 and P_2 of ΔL_1, for a given P_{s1} is

$$\max [P_1, P_2]\, \Delta L_1 = \{a_{1z}^2 + a_{1y}^2 + (2|a_{1y}| + \frac{D_y}{2})\frac{D_y}{2}\}^{1/2} - |a_{1z}|$$

$$\xrightarrow[|a_{1z}| \to +\infty]{} \frac{1}{2|a_{1z}|}(|a_{1y}| + \frac{D_y}{2})^2 . \tag{9}$$

Next, consider the case when P_{s1} is located in Region 2. We already know that L_{11} can have no extremum of any type in Region 2, nor can L_{21}. Therefore, that P_1 for which L_{11} is smallest, once P_{s1} is fixed, must lie on the boundaries of its allowed region of variation, i.e., at $\pm D_y/2$. Similarly, that P_2, for which L_{21} is maximum, must also lie there. By directly comparing the two possible choices, one finds

$$P_1 = (a_{1y}/|a_{1y}|)\, D_y/2$$

$$P_2 = -P_1 \qquad (10)$$

so that the maximum over P_1 and P_2 of ΔL_1 for a given P_{s1} is

$$\max [P_1, P_2]\Delta L_1 = \{a_{1z}^2 + (|a_{1y}| + \frac{D_y}{2})^2\}^{1/2}$$

$$-\{a_{1z}^2 + (|a_{1y}| - \frac{D_y}{2})^2\}^{1/2} \xrightarrow[|a_{1z}|\to\infty]{} \frac{1}{2|a_{1z}|}\, 4|a_{1y}|\, \frac{D_y}{2}\,. \qquad (11)$$

From Eqs. (9) and (11)

$$\max[P_1, P_2]\Delta L_1 \xrightarrow[|a_{1z}|\to\infty]{} \frac{1}{2|a_{1z}|} \begin{cases} (|a_{1y}| + \frac{1}{2} D_y)^2 & \text{in Reg. 1} \\ 2|a_{1y}|\, D_y & \text{in Reg. 2} \end{cases}$$

$$= \frac{1}{2|a_{1z}|}\, f_y^2 \qquad (12)$$

where f^2 is defined by the above.

The maximum of ΔL_{11} over all possible P_1, P_2 and P_{s1} can now be easily evaluated. Denoting the maximum value of $|a_{1y}|$ by $|a_{1y}|_{max}$, one finds

$$\max[P_1, P_2, P_{s1}]\Delta L_{11} \xrightarrow[|a_{1z}|\to\infty]{} \frac{1}{2|a_{1z}|} \begin{cases} (|a_{1y}|_{max} + \frac{1}{2} D_y)^2 ; & \text{in Reg. 1} \\ 2|a_{1y}|_{max}\, D_y ; & \text{in Reg. 2} \end{cases} \qquad (13)$$

So far we dealt with the source and interference planes as if they were one dimensional. The second surface dimension, along the x axis, can be easily incorporated, by allowing $\bar{P}_i$ and $\bar{a}_k$ to have x components. Then by similar reasoning

$$\max[\bar{P}_1, \bar{P}_2]\Delta L_1 = \{a_{1z}^2 + \sum_{n=x,y} (|a_{1n}| + \frac{1}{2} D_n)^2\}^{1/2} -$$

$$-\{a_{1z}^2 + \sum_{n=z,y} (|a_{1n}| - \frac{1}{2} D_n)^2\}^{1/2} \qquad (14)$$

$$\xrightarrow[|a_{1z}|\to\infty]{} \frac{1}{2|a_{1z}|} \sum_{n=x,y} f_n^2$$

and the definition of f_x is obtained from that of f_y by changing all subscripts y to x.

In the special case when $\bar{L}$ is parallel to z and both the source plane and the interference plane are perpendicular to L, then Eq. (13) reduces to the known formula

$$\max[P_1, P_2]\Delta L_{11} \xrightarrow[|a_{1z}|\to\infty]{} \frac{1}{2|a_{1z}|} \begin{cases} \frac{1}{4} (|D_s| + |D|)^2 ; & \text{if } D_s \leq D \\ |D_s|\,|D_y| ; & \text{if } D_s > D \end{cases} \qquad (14a)$$

and f_n^2 goes over into the known[1] function Δ_n^2.

III. DEFOCUSING ELEMENTS

Assume that the colletor has an intrinsic angular inaccuracy of ε. For AXAF ε is of the order of 2×10^{-5} radians. It will also be assumed that the detector size can be made arbitrarily small, so that the instrument resolution is not limited by detector size. For AXAF, that assumption is not automatically valid, but for superfine holographic x-ray gratings, in many cases, instrument resolution is limited by the location of the grating, and that is the topic to be studied here.

The geometry and notation is explained in Fig. 3. The angular deviation introduced by the grating will be detectable over the "noise" caused by finite ε and ℓ, if

$$\tan \psi \gtrsim \frac{F}{L} \tan \varepsilon + \frac{\ell}{L} \tan \theta \ . \tag{15}$$

The deviation in the detector plane induced by the grating in m^{th} order near normal incidence for wavelength λ will differ from that for $\lambda + \Delta\lambda$ by

$$L\Delta\lambda \frac{d\tan\psi}{d\cos\psi} \frac{d\cos\psi}{d\lambda} = \Delta\lambda L \frac{\pm m}{d} \frac{d\tan\psi}{d\cos\psi} = \frac{\Delta\lambda}{\lambda} Lf(\psi) \tag{16}$$

where f is defined as

$$f(\psi) = -(\cos^2\psi - \tan^2\psi)/\sin\psi \ .$$

Therefore, the image of a single point source in m^{th} order, near normal incidence, formed by two different wavelengths, λ and $\lambda + \Delta\lambda$, will be distinguishable, provided that

$$\frac{\Delta\lambda}{\lambda} \gtrsim \frac{F}{L} \frac{\tan \varepsilon}{f(\psi)} + \frac{\ell}{L} \frac{\tan \theta}{f(\psi)} \ . \tag{17}$$

For AXAF, one has $F = 10^3$ cm, $\varepsilon \simeq 2 \times 10^{-5}$ radians, and $\theta = 0.1$. Then

$$\frac{\Delta\lambda}{\lambda} \gtrsim \left[\frac{2 \times 10^{-2}}{L} + \frac{\ell}{L} 0.1003\right] \frac{1}{f(\psi)} \ , \tag{18}$$

if L and ℓ are measured in centimeters. Now clearly the best choice is $\ell = 0$, and then

$$\frac{\Delta\lambda}{\lambda} \gtrsim \frac{2 \times 10^{-2}}{L \text{ (in cm)}} \quad \frac{1}{f(\psi)} \; . \tag{19}$$

For superfine gratings with $d = 10^{-6}$ cm, and $\lambda = 100$ Å, one has $f(\psi) = 1/\sqrt{2}$, and

$$\frac{\Delta\lambda}{\lambda} \gtrsim \frac{1.41 \times 10^{-2}}{L \text{ (in cm)}} \; ; \text{ for } \lambda = 100 \text{ Å} \; . \tag{20}$$

For gratings with diameter $\lesssim 20$ cm in AXAF, one has $L \lesssim 100$ cm, which would give $\Delta\lambda/\lambda \gtrsim 1.4 \times 10^{-4}$ for $\lambda = 100$ Å. This resolution is clearly far from the intrinsic resolution of the grating, which is about 5×10^{-6} for this wavelength.

To increase the instrument resolution, one can either increase the grating diameter, D_g, or increase L.

The increase in resolution with D_g is linear, and we will not concern ourselves further with that variation.

To increase the resolution while keeping D_g fixed, one may introduce a defocusing element at a distance ℓ_d before the original focal plane. If that element changes θ by a factor h^{-1} $(h > 1)$, then the new focal plane will be farther from the defocusing element than the original focal plane was, by a factor h. Then the wavelength resolution of the new instrument will be given by Eq. (17), provided that in it one substitutes for F, L, ℓ, and θ the new effective values: F', L', ℓ', and θ'. Assuming $F, L > \ell_g > \ell$, one has

$$F' = (F - \ell_g) + h\ell_g = F + \ell_g \, (h - 1)$$

$$L' = L + \ell_g \, (h - 1)$$

$$\ell' = h \, \ell$$

$$\theta = h^{-1} \, \theta$$

and the new resolution is

$$\frac{\Delta'\lambda}{\lambda} \geq \frac{F+\ell_g(h-1)}{L+\ell_g(h-1)} \frac{\tan \varepsilon}{f(\psi)} + \frac{h\ell}{L+\ell_g(h-1)} \frac{\tan (h^{-1}\theta)}{f(\psi)} . \qquad (21)$$

When $\ell = 0$, and the defocusing element is located close behind the grating, $\ell_g \approx L$, then

$$\frac{\Delta'\lambda}{\lambda} \geq \frac{1}{h} \frac{F}{L} \left(1 + \frac{h-1}{F/L}\right) \frac{\tan \varepsilon}{f(\psi)} . \qquad (22)$$

Clearly, for F/L large, the resolution will improve essentially linearly with h, up to values $h \approx F/L$. Beyond that, the resolution tapers off, and approaches its asymptotic value for large h:

$$\frac{\Delta'\lambda}{\lambda} \xrightarrow[h\to\infty]{} \frac{\tan \varepsilon}{f(\psi)} ,$$

which is, of course, the intrinsic resolution of the grating.

For AXAF, it appears possible to reflect the beam at least once, maybe twice (once backward and once forward) along the axis of the focusing telescope. In that way one could increase the wavelength resolution for F - L = 900 cm, h = 21, to 66% of its intrinsic value; i.e., to $\Delta\lambda/\lambda = 7.3 \times 10^{-6}$ at $\lambda = 100$ Å. Reflecting the beam in this manner causes no difficulty in this wavelength range, since the reflectivity of multilayer mirrors can approach 66%. However, the use of multilayer reflection optics will reduce the bandwidth of the instrument.

An alternative option would be to place the same instrument into Spacelab. Defocusing could be done with grazing incidence mirrors, which would alleviate the problem of wavelength reduction. Additional reflections would not be required if the dimensions of Spacelab are to be as envisioned today.

ACKNOWLEDGEMENTS

I wish to thank Mike Hettrick of BNL and Martin Weisskopf of MSFC for interesting discussions on this topic.

REFERENCES

1. Csonka, Paul L., "Holographic X-Ray Gratings Generated by Synchrotron Radiation", J. Appl. Phys., 52, 2692 (1981).

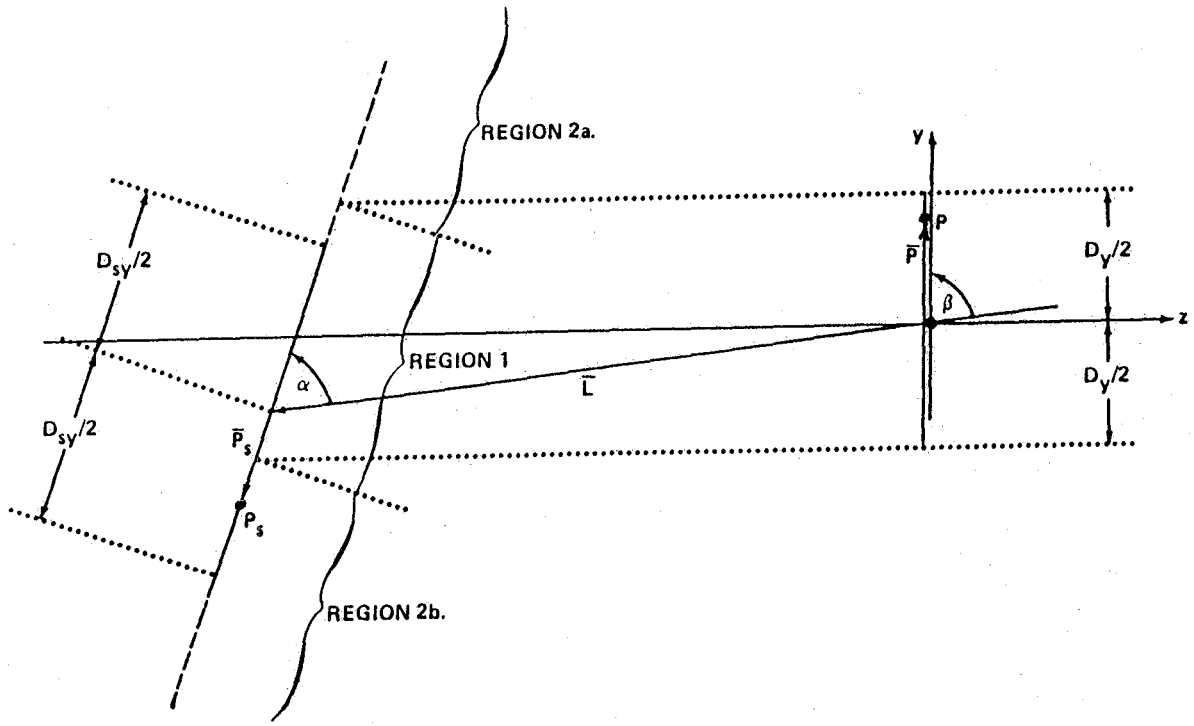


Figure 1. The interference pattern is generated on the "interference surface." It has diameter (in the y,z plane) D_y. The point P lies on this surface. The source of the beam is a plane whose diameter (in the y,z plane) is D_{sy}. The point P_s lies in the source plane. The center of the beam is along the vector $\bar{L}$, and the radius vector of P and P_s is $\bar{P}$ and $(\bar{L} + \bar{P}_s)$, respectively. The plane in which the interference takes place is divided into three regions: Regions 1, 2a, and 2b, which together cover the whole plane. Region 1 is defined as the projection along y onto this plane of the area on which the interference pattern is generated. There are no restrictions on the orientation of the source plane and that of the interference plane with respect to $\bar{L}$.

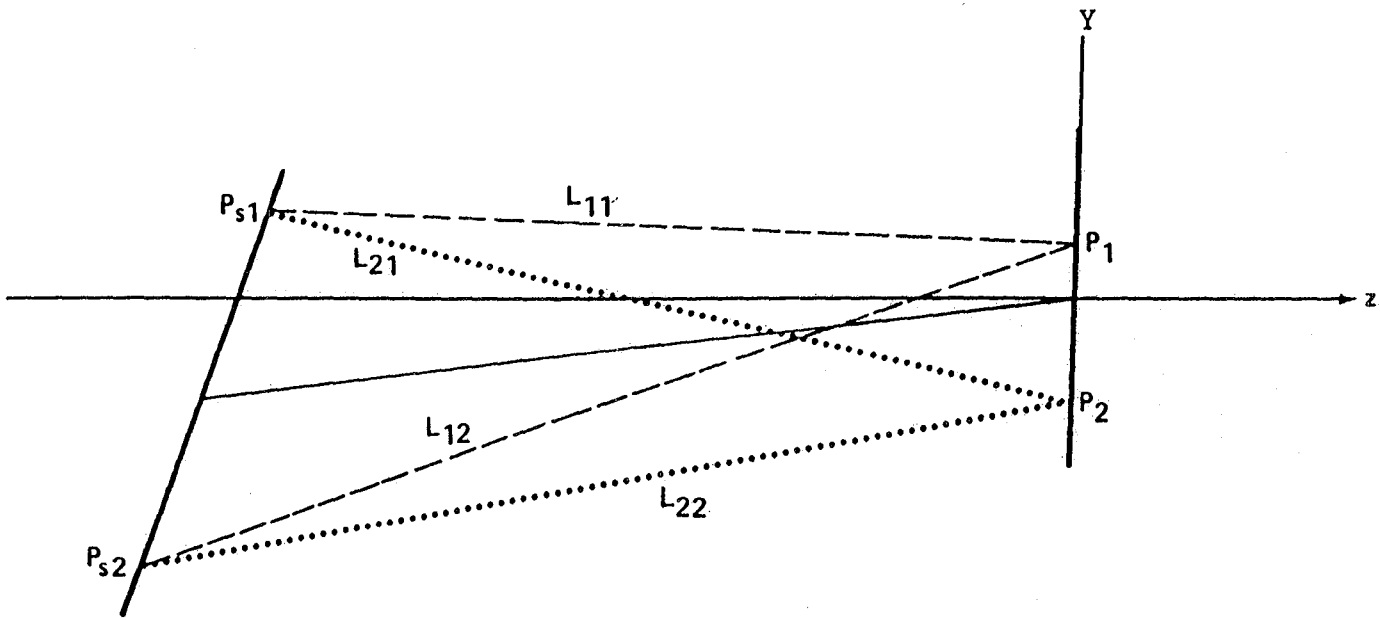


Figure 2. The optical pathlength to P_1 from P_{s1} and P_{s2} is L_{11} and L_{12}, respectively. The optical pathlength to P_2 from P_{s1} and P_{s2} is L_{21} and L_{22}, respectively. The optical path difference $\Delta L_i = (L_{2i} - L_{1i})$ measures the extent to which the interference plane deviates from a wavefront for a wave emitted at P_{si}.

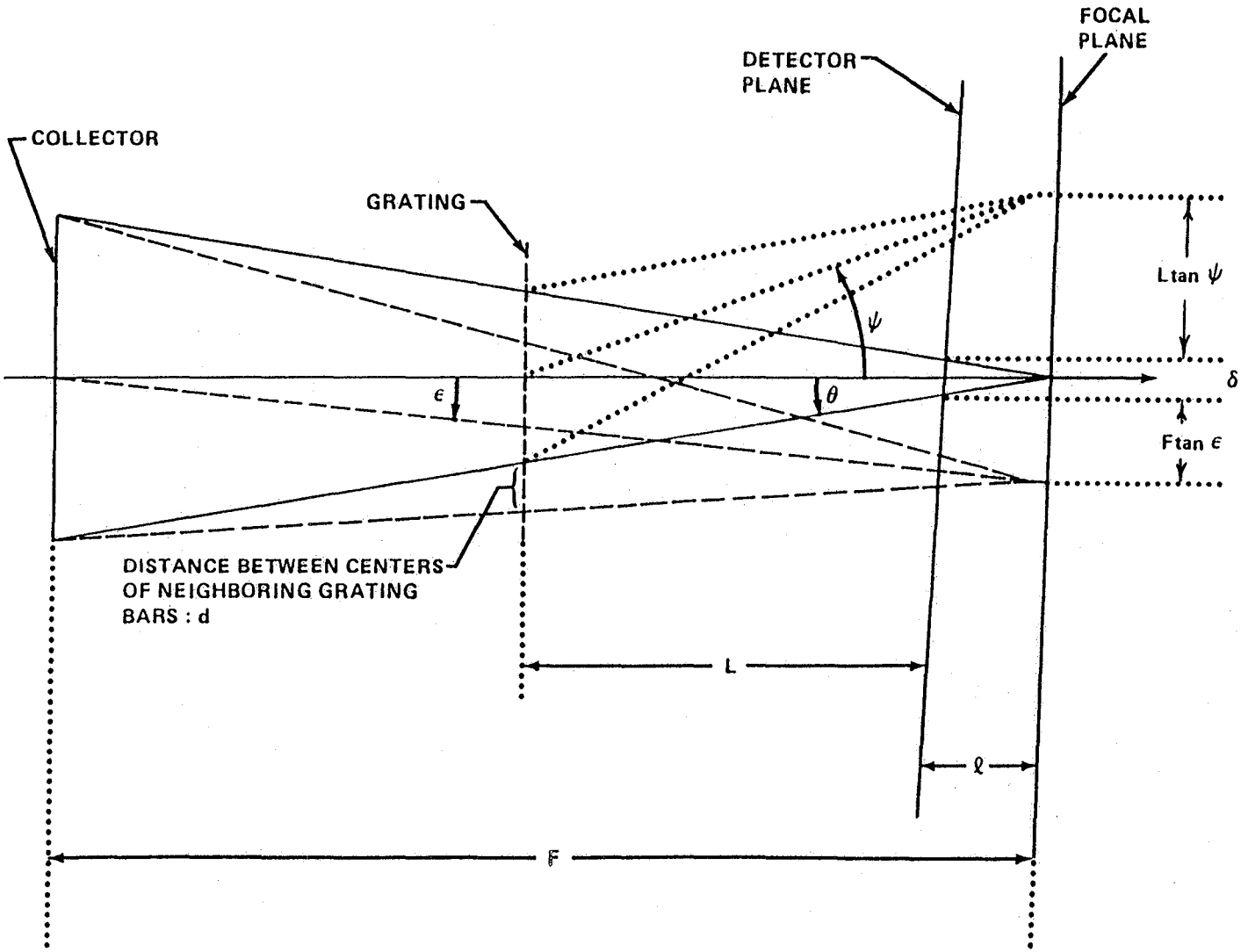


Figure 3. The collector has focal length F, and focuses radiation with an angular divergence θ in the focal plane, with an angular inaccuracy ϵ. As a result of this inaccuracy, the position of the image on the focal plane is undetermined to within a distance F tan ϵ. The size at the detector plane of an image focused on a point at the focal plane is $\delta = \ell \tan \theta$. The grating introduces an angular deviation ψ, causing the image point in the focal plane to be displaced by a distance L tan ψ. This displacement will be detectable if L tan ψ F tan ϵ + ℓ tan θ.

1985

NASA/ASEE SUMMER FACULTY RESEARCH FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

FOAM FLOTATION AS A SEPARATION PROCESS

Prepared by:	Ben L. Currin, Ph. D.
Academic Rank:	Instructor
University and Department:	Calhoun Community College Department of Natural Sciences
NASA/MSFC: (Laboratory) (Division) (Branch)	 Space Science Laboratory Low Gravity Science Fluid and Transport Processes
MSFC Counterpart:	Robert S. Snyder
Date:	August 12, 1985
Contract No.:	NGT-01-008-021

FOAM FLOTATION AS A SEPARATION PROCESS

By

Ben L. Currin, Ph. D.
Department of Natural Sciences
Calhoun Community College
Decatur, Alabama

ABSTRACT

The basic principles of foam separation techniques are discussed. A review of the research concerning bubble-particle interaction and its role in the kinetics of the flotation process is given. Most of the research in this area deals with the use of theoretical models to predict the effects of bubble and particle sizes, of liquid flow, and of various forces on the capture and retention of particles by bubbles. A discussion of fluid mechanical aspects of particle flotation is given.

INTRODUCTION

Foam flotation is a separation method that has been applied to a wide variety of substances and has been utilized for many years in the field of mineral processing. Recently, many new variations of these techniques have been developed in which removals of species such as organic compounds and colloidal substances have been achieved.

In order to maximize the efficiency of foam flotation techniques, the attachment of particles to bubbles must be made more likely. An understanding of the factors involved in particle-bubble encounters and interactions is essential. Another factor that must be investigated is the effect of viscous drag forces on a particle attached to a bubble.

GENERAL DESCRIPTION OF FOAM FLOTATION TECHNIQUES

Foam flotation is a subdivision of adsorptive bubble separation techniques, a group of separation techniques based on differences in surface activity. Material of various types and composition is selectively adsorbed at the surfaces of bubbles rising through a liquid and is thereby concentrated or separated. A substance which is not itself surface active can often be made effectively surface active through interaction with a surface active species. A classification scheme for adsorptive bubble separation techniques is shown in Figure 1 [1].

In practice, foam separation involves the passage of a gas through the solution containing the species to be removed and a surfactant. The adsorbed components are separated by removal of the foam. Figure 2 shows two modes of foam column operation. In the batch system, shown in Figure 2(a), gas bubbles are generated in the liquid and the foam is collected. This method is impracticle for large-scale processes, so the continuous mode of operation, shown in Figure 2(b), is used. The bulk solution may be fed into the column continuously and the bulk residue collected continuously. Part of the foamate may be fed back to the top of the column to increase the separation factor. The column may be operated in the stripping mode by introducing the feed into the foam so that separation takes place while it is descending through the foam.

THEORETICAL ASPECTS OF PARTICLE FLOTATION

Two models have been used to investigate adsorption of particles at an air-water interface. In the Coulombic model [2], the binding force between particle and the air-water interface is due to coulombic attraction between the charged interface and the oppositely charged particles. The charge at the interface is due to the adsorption of ionic surfactant.

In the second model, the non-Coulombic model [3], adsorption of surfactant molecules to the surface of the solid particles, results in a hydrophobic surface. Nonzero contact angles then allow bubble attachment and flotation. The binding energy of the particle to the air-water interface may be estimated as follows: Assume the floc particles are spherical and that the radius of the air bubble is much larger than that of the floc particle. The free energy change during attachment is given by

$$\Delta G = -\gamma_{AW}\pi r^2(1 - \cos\theta)^2 \qquad (1)$$

where γ_{AW} is the interfacial free energy at the air-water interface, r is the particle radius,and θ is the contact angle as shown in Figure 3.

In Table 1 [4], the magnitude of the floc particle-air bubble

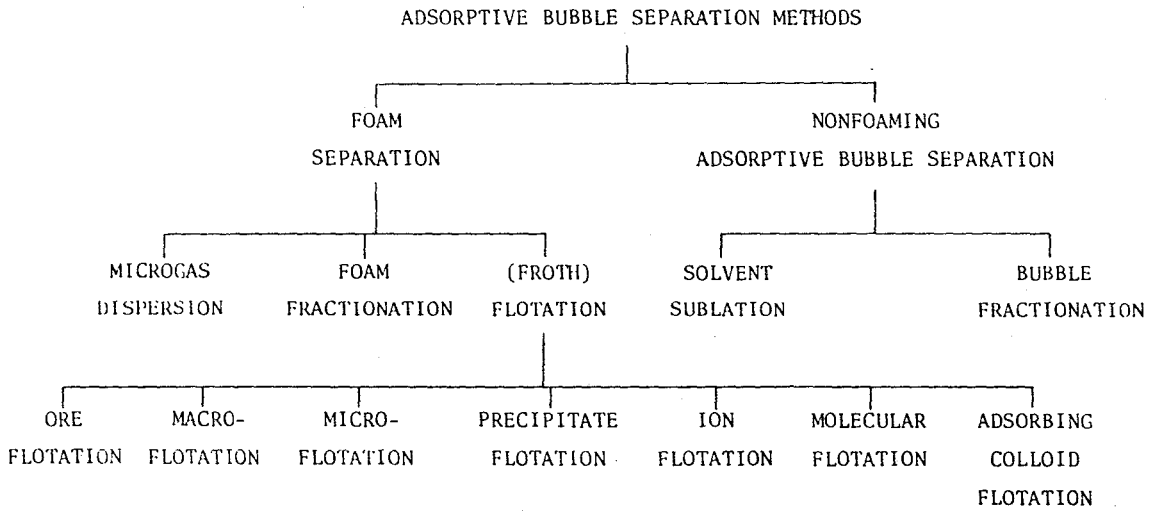


Figure 1. Classification of Adsorptive Bubble Separation Techniques

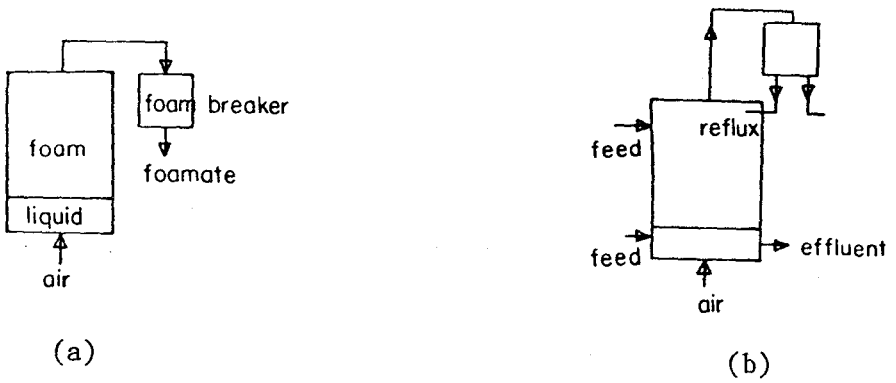


Figure 2. Modes of Foam Column Operation

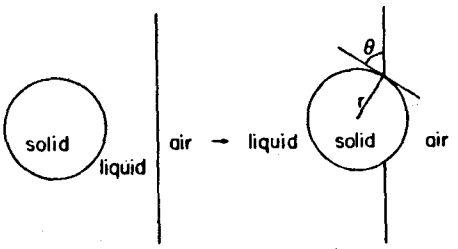


Figure 3. Model of Floc-Bubble Attachment [4]

cosθ	-ΔG (erg)
1.0	0.0×10^{-10}
0.9	1.257
0.8	5.027
0.6	20.11
0.4	45.24
0.2	80.42
0	125.66
-1.0	502.64

$kT = 4.14 \times 10^{-14}$ erg at 298 K.

Table 1. Floc-Bubble Binding Energy as a Function of Contact Angle

interaction energy is given as a function of θ. For these calculations, 40 dyne/cm was chosen for γ and a value of 0.1 mm was used for r. It is apparent that even for small particles, the floc-bubble binding energy is several orders of magnitude larger than kT unless the contact angle is almost zero.

REVIEW OF RESEARCH ON BUBBLE-PARTICLE ATTACHMENT

Gaudin [5] analyzed collision between an air bubble and a mineral particle with the assumptions that water is non-viscous and incompressible, the mineral particles are large compared with the mean free path of the water molecules, bubble and particle are rigid spheres, bubble and particle are the only factors that affect flow of liquid, and there is streamline flow around the spheres. From his analysis, he determined that collision between bubble and particle was impossible unless the center of the particle was located on the central line of motion of the bubble. Sutherland [6] points out that Gaudin's analysis was based on an erroneous assumption that two bodies move independently in a fluid. He also defines the "collision radius" of a bubble, D, which is given by

$$D = \sqrt{3rR} \qquad (2)$$

where r is the particle radius and R is the bubble radius. Particles lying within this distance from the line of motion of the bubble will collide with it.

Trajectories for particles in the path of a spherical bubble rising in an infinite pool of liquid were calculated by Flint and Howarth [7]. They defined a collision efficiency, which is the ratio of the number of particles that actually collide with the bubble to the number that would collide if the fluid streamlines were not diverted by the bubble. Whether or not a particle collides with the bubble depends on the balance of viscous, inertial, and gravitational forces acting on it, and the form of the streamlines around the bubble. Their theoretical analysis predicted two regions of particle-bubble behavior. For larger particles, collision efficiency depends most strongly upon inertial forces. In this coarse particle region, collision efficiency is increased by increasing the bubble size. For smaller particles, inertial effects of the particle may be neglected and collision efficiency is independent of whether Stokes or potential flow is assumed. They also showed that the collision efficiency of particles with bubbles in a swarm can be several times as large as those calculated from a single sphere model.

The concept of two types of bubble-particle interaction is also given by Reay and Ratcliff [8]. They define a collision regime which applies to particles larger than about 3 microns diameter. In this regime, particles will come into contact with the bubble only if their hydrodynamically-determined trajectories come within one particle radius, r , of the bubble surface. The collection efficiency should increase with increasing r since larger particles have a better chance

of intercepting the bubble. For this regime, they assume that the flow pattern around the front of the bubble is given by the Stokes equation for creeping flow around a rigid sphere, that electrical interactions between particle and bubble have a negligible effect on the particle trajectory, that the motion of the bubble is not affected by the presence of the particle, and that the fluid velocity used in computing the drag on a particle is the velocity which would exist at the point occupied by the center of the particle if the particle were absent. In the diffusion regime, submicron particles reach the bubble mainly by Brownian diffusion. In this regime, the collection efficiency should decrease with increasing r since larger particles diffuse more slowly.

The analysis of Reay and Ratcliff gave the following results: (1) In the collision regime, gravity is the only factor causing the particle's trajectory to deviate from the fluid streamlines. (2) In the collision regime, the number of particles picked up by a bubble is independent of bubble size and proportional to bubble frequency. (3) In the collision regime, flotation rate should increase with the square of the particle diameter. (4) In the diffusion regime, the flotation rate should be inversely proportional to particle diameter. Experimental results were in general agreement with theory.

Schulze and Gottschalk [9] performed experiments on hydrodynamic interaction between a single immobile air bubble streamlined by a flow of liquid and solid particles. The trajectories of individual particles were recorded stroboscopically. They found that the experimental particle trajectories followed the streamline given by potential flow around the bubble. The particle radius in their experiments was 0.080 mm and the bubble had a radius of 1.53 mm.

The coordinate system of particle-bubble attachment used by Schulze and Gottschalk is shown in Figure 4. Kinetic energy of particles flowing near the symmetry axis of the bubble is used for the elastic deformation of the bubble surface. These particles are repelled from the bubble once or several times. The time of contact on collision is less than 4 ms and the energy loss on the first collision is more than 70%. If $\phi < 30^{\circ}$, there are no collisions - the particles slide over the bubble. The attachment appears to occur mainly in the transition region between colliding and sliding. The fraction of particles that attach to the bubble varies from 0.2 to 0.5.

Conclusions of Schulze and Gottschalk are: (1) Particle attachment takes place within a time interval longer than that of collision shorter than that of sliding. Collision times range from 1 to 4 ms and sliding times range from 30 to 50 ms. (2) The time of contact has a maximum at angles of about $\phi = 30^{\circ}$ and there is an optimum distance for attachment.

Anfruns and Kitchener [10] measured the rate of capture of quartz particles of 0.012 to 0.040 mm diameter by single rising bubbles of 0.5 to 1.1 mm diameter. They described a model of particle capture in which the particle must first make a hydrodynamic collision with the bubble.

V_B, V_r, V_ϕ - radial components of particle velocity,

E_{st} - kinetic energy on collision of particles and the bubble,

τ_{st} - time of contact on collision,

ϕ_{Gr} - polar angle representing a transition from collision to sliding,

τ_{Gl} - sliding time,

ϕ_B - polar angle at which the limiting particle trajectory is intercepted by the bubble

E - hydrodynamic efficiency of collision.

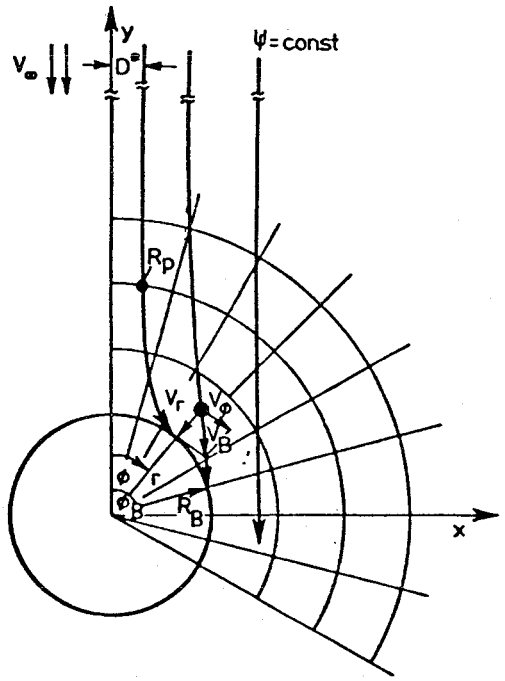


Figure 4. Coordinate System of Particle-Bubble Interaction [9]

During the collision period of a few milliseconds, the thin liquid film between particle and bubble ruptures. The contact meniscus then expands rapidly over the particle.

Derjaguin, Dukhin, and Rulyov [11] analyzed the kinetic theory of flotation of small particles. For comparison purposes, their analysis included some aspects of the theory of flotation of large particles. They divided the flotation act into two stages. The first is the approach of the particle to the bubble and the second is the fixation of the particle to the bubble. In agreement with Reay and Ratcliff [8], they postulate different mechanisms for small and large particle flotation. In the case of large particles, formation of a three-phase wetting perimeter is able to resist detachment forces. They refer to this phenomenon as contact flotation. For small particles a three-phase wetting perimeter is not formed but the detachment forces are several orders of magnitude smaller than for large particles and can be overcome by London-van der Waals attractions. This is called contactless flotation. In most cases the particle and the bubble have like charges and overlapping of the double layers results in repulsion. In the case of large particles, this repulsion can be overcome by an inertial impact on the bubble surface. Small particles do not undergo such an impact but hydrodynamic forces may press the particle into the bubble surface.

FLUID MECHANICAL ASPECTS OF PARTICLE FLOTATION

It has been shown that the binding energies of particles to the air-water interface are much greater than the thermal energies of the particles [2]. It is possible that viscous drag forces could detach adsorbed particles from bubbles rising through a liquid. An analysis of these drag forces is helpful in the determination of the efficiency of bubble-particle attachment and in the efficiency of foam flotation techniques.

Several models that vary in sophistication have been used to estimate the viscous drag force on a floc particle attached to a rising bubble [12]. The viscous force seems to be several orders of magnitude smaller than the force that binds the particle to the bubble. The results of calculations based on these models show that if the bubble size is small, the viscous forces are much too small to detach particles from the bubbles and as bubble size increases, viscous drag forces also increase. It should be noted that these models are approximate and qualitative because of the use of presuppositions and simplifying assumptions.

A much more rigorous analysis of the fluid mechanical aspects of particle-bubble attachment has been made by French and Wilson [13]. The floc particle is subject to three forces; the binding force F_B, a drag force F_D, and a lift force F_L. Each of the three forces was calculated by numerical methods and it was found that the sign of the lift force was negative, indicating that the lift force pressed the particle into the bubble.

Also, the nature of the binding force is such that the particle is free to move about on the bubble surface. The effect of F_B, F_D, and F_L is to roll the floc particle toward $\theta' = 0^\circ$. As shown in Figure 5, an axisymmetric cap of particles would be formed on the bubble about $\theta' = 0^\circ$.

Kiefer and Wilson [14] investigated a "squeeze-out" mechanism in which drag forces on the cap of floc particles cause development of a surface pressure which, if large enough, will "pop" a floc particle from the cap. Their model is shown in Figure 6. The assumption is made that the viscous drag force on the floc particle is independent of its position on the bubble surface. It is shown that the fraction of the area of the bubble that is covered by particles is a function of bubble size. As the radius of the bubble increases, the fraction of its surface that is covered decreases. There is a critical bubble radius, r_c, which represents the maximum radius of a bubble that can be completely covered. The critical bubble radius is given by

$$r_c = \left(\frac{3\Delta G}{\pi \alpha \rho g}\right)^{1/2} \tag{3}$$

where ΔG is the energy of attachment, $\alpha = \pi a^2$ (a is the particle radius), ρ is the density of the solution, and g is the gravitational constant. If a is chosen as 0.1 μm and $G \cong 2 \times 10^{-9}$ erg, the value of r_c is approximately 0.080 cm, meaning that bubbles of this radius and smaller may be completely covered. If the radius of the bubble increases to 0.112 cm, its coverage drops to 50 %. It should be noted that these calculations are quite rough.

This "squeeze-out" model was reexamined [15] by an approach that took into account the variation of the surface pressure with position and the separation of the boundary layer, and was also applicable to larger bubble sizes. Figure 7(a) shows the geometry that was used in the development of the model. In this diagram, represents the top of the particle cap and represents the point at which boundary layer separation occurs. In Figure 7(b) a plot is given of the void (uncovered) fraction of bubble surface vs. bubble radius. Figure 7(c) gives a plot of total surface/volume and loaded surface/volume vs. bubble radius. It can be seen that in agreement with the simpler model, if bubbles are of radius > 0.080 cm, the fraction of covered surface and, therefore, the efficiency of flotation will decrease.

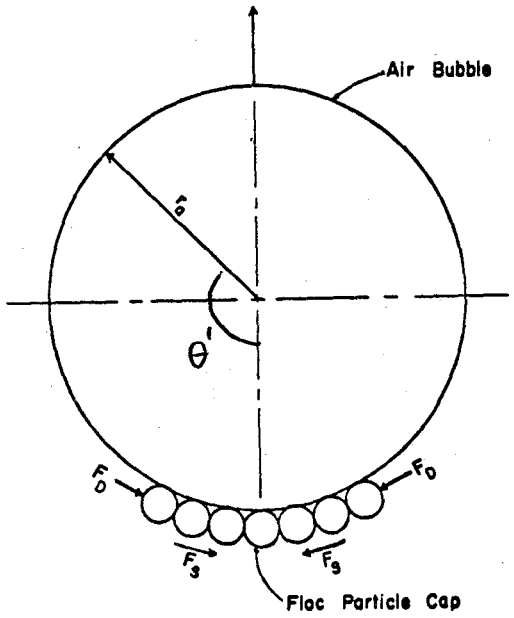


Figure 5. Effect of Various Forces on Particle Attachment [13]

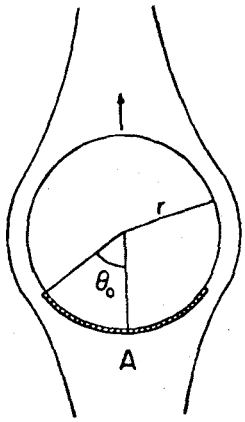
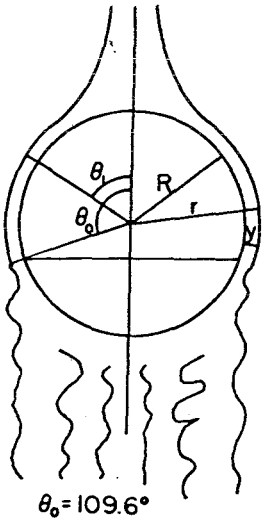
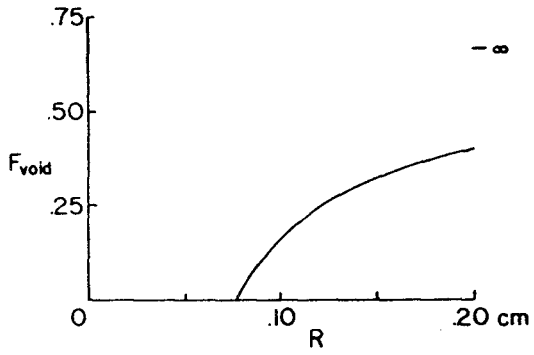


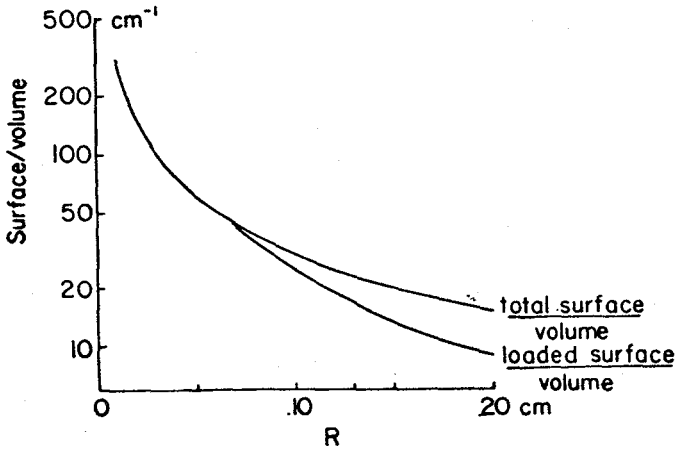
Figure 6. Model of a Spherical Cap of Floc Particles on the Bottom of a Rising Bubble [14]



(a) Boundary Layer Separation Around a Rising Bubble



(b) Void Fraction of Bubble Surface as a Function of Bubble Radius



(c) Total Surface/Volume and Loaded Surface/Volume as Functions of Bubble Radius

Figure 7

CONCLUSIONS AND RECOMMENDATIONS

The interaction of bubbles and particles is a complex process which has been investigated by many researchers from different approaches. It is evident that there is general agreement on several important points, and disagreement on others.

The attachment of particles to bubbles depends upon the degree of hydrophobicity of the particle surface and in most cases viscous drag forces are far too small to detach a single particle from the bubble surface. There are two different mechanisms for bubble-particle attachment, depending on the particle size. There is disagreement as to the role of various forces for the attachment of large and small particles. Particles may slide to the back of the bubble after attachment, creating a cap in which excess pressure may cause detachment.

It would be helpful to investigate some of the aspects of particle-bubble attachment that are not understood clearly. Experimental work in this area could be carried out by means of a captive bubble technique which would allow better control of conditions by eliminating problems associated with uncontrollable bubble rise. A flow of liquid past a captive bubble can be used to simulate the rise of a bubble in a liquid. Some suggested experiments are:

1. In the absence of bubble rise, determine the extent of particle capture as a function of particle size, bubble size, and flow rate.

2. Investigate the extent of particle capture as a function of particle hydrophobicity. This could give insight into the mechanism of particle capture. The particle hydrophobicity can be varied by addition of surfactant.

3. Determine the effect of added electrolytes on particle capture.

4. Determine bubble coverage as a function of bubble size and flow rate.

5. There is disagreement concerning the role of gravity in particle-bubble attachment. A possible experiment for low gravity conditions is to observe what effect the absence of gravity has on particle-bubble interaction.

REFERENCES

1. B. L. Karger, R. B. Grieves, R. Lemlich, A. J. Rubin, and F. Sebba, Separ. Sci., 2: 401 (1967).

2. D. J. Wilson, Separ. Sci., 11: 391 (1976).

3. P. Somasundaran, T. W. Healy, and D. W. Fuerstenau, J. Phys. Chem., 68:3562 (1964).

4. D. J. Wilson, Sep. Sci. Technol., 13: 107 (1978).

5. A. M. Gaudin, Flotation, 2nd Ed., McGraw-Hill, New York, 1956, pp. 340-355.

6. K. L. Sutherland, J. Phys. Chem., 58: 394 (1948).

7. L. R. Flint and A. Howarth, Chem. Eng. Sci., 26:1155 (1971).

8. D. Reay and G. A. Ratcliff, Can. J. Chem. Eng., 51: 178 (1973).

9. H. J. Schulze and G. Gottschalk, in Developments in Mineral Processing, J. Laskowski, ed., 1979, pp. 63-85.

10. J. D. Anfruns and J. A. Kitchener, in Flotation, A. M. Gaudin Memorial Volume, V. 2, M. C. Fuerstenau, ed., American Institute of Mining, Metallurgical, and Petroleum Engineers, Inc., New York, 1976, pp. 625-637.

11. B. V. Derjaguin, S. S. Dukhin, and N. N. Rulyov, in Surface and Colloid Science, V. 13, E. Matijevic and R. J. Good, eds., Plenum Press, New York, 1983, pp. 71-113.

12. B. L. Currin, F. J. Potter, D. J. Wilson, and R. H. French, Sep. Sci. Technol., 13: 285 (1978).

13. R. H. French and D. J. Wilson, Sep. Sci. Technol., 15:1213 (1980).

14. J. E. Kiefer and D. J. Wilson, Sep. Sci. Technol., 15: 57 (1980).

15. A. N. Clarke and D. J. Wilson, Foam Flotation, Theory and Applications, Marcel Dekker, Inc., New York, 1983, pp. 110-119.

1985

NASA/ASEE SUMMER FACULTY RESEARCH FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

DENSITY AND PRESSURE VARIABILITY IN THE MESOSPHERE AND THERMOSPHERE

Prepared by:	T. Michael Davis, Ph.D.
Academic Rank:	Associate Professor
University and Department:	North Georgia College Department of Physics
NASA/MSFC:	
Laboratory:	Systems Dynamics
Division:	Atmospheric Sciences
Branch:	Atmospheric Physics
MSFC Counterpart:	Robert E. Smith
Date:	August 16, 1985
Contract No.:	NGT 01-008-021 The University of Alabama in Huntsville

DENSITY AND PRESSURE VARIABILITY IN THE MESOSPHERE AND THERMOSPHERE

by

T. Michael Davis
Associate Professor of Physics
North Georgia College
Dahlonega, Georgia

ABSTRACT

In an effort to isolate the essential physics of the mesosphere and the thermosphere, a steady one-dimensional density and pressure model has been developed in support of related NASA activities, i.e., projects such as the AOTV and the Space Station. The model incorporates a zeroth order basic state including both the three-dimensional wind field and its associated shear structure, etc. A first order wave field is also incorporated in period bands ranging from about one second to one day. Both basic state and perturbation quantities satisfy the combined constraints of mass, linear momentum and energy conservation on the midlatitude beta plane. A numerical (iterative) technique is used to solve for the vertical wind which is coupled to the density and pressure fields. The temperature structure from 1 to 1000 km and the lower boundary conditions are specified using the U. S. Standard Atmosphere 1976. Vertical winds are initialized at the top of the Planetary Boundary Layer using Ekman pumping values over flat terrain. The model also allows for the generation of waves during the geostrophic adjustment process and incorporates wave nonlinearity effects.

Preliminary results indicate that lower atmosphere wave processes can account for much of the observed variability of the density and pressure in the mesosphere and lower thermosphere. Basic state processes, especially vertical gradients of vertical winds are also influential under certain conditions. At very high heights (greater than 150 km) the model predicts that air density and pressure can deviate by as much as 20 % from Standard Atmosphere conditions in the presence of strong updrafts. This kind of effect would be likely to occur in the high latitude thermosphere where Joule heating is a factor leading to the development of strong vertical winds.

ACKNOWLEDGEMENTS

I am extremely honored by being chosen by Dr. Robert Smith and Dr. Gerald Karr to participate a second time in the Summer Faculty Fellowship Program. I acknowledge with sincere gratitude the "re-boost" that working with Dr. Smith has given to my career as a professor and researcher.

Dr. Douglas ReVelle has been a continual source of enlightenment. It has been a pleasure to work with him on his atmospheric model. Dr. ReVelle is an outstanding researcher, possessing a quite impressive understanding of so many areas of research. I know that our talks took a lot of valuable time from him this summer; I thank him for his kindness and patience.

LIST OF FIGURES

LIST OF FIGURES

(concluded)

INTRODUCTION

The mesosphere extends from an altitude of approximately 50 km to approximately 90 km. The temperature in this region drops from about 270 K to around 180 K, and the density ranges from 10^{-3} kg/m^3 to 10^{-5} kg/m^3. The mesopause separates the mesosphere from the thermosphere, where the temperature rises to more than 1000 K at altitudes of hundreds of kilometers. The density of the thermosphere decreases to values near 10^{-13} kg/m^3.

It is important for scientists and engineers working on the Space Shuttle, Space Station, and Aeroassisted Orbital Transfer Vehicle (AOTV) projects to be able to predict the air densities which these vehicles would encounter in the upper atmosphere, since large changes in the density can drastically affect their status while on orbit or during (Shuttle and AOTV) re-entry. In an effort to improve understanding of conditions in the upper atmosphere which can result in these changes, Dr. Douglas ReVelle has developed a model (ReVelle, 1985) which contains some of the atmospheric drivers of these density variations: winds and waves. The computer model, called "AIR", has performed admirably at generating reasonable values of physical quantities.

OBJECTIVES

The objectives of this work were to:

1) transfer the program AIR from HP-75 BASIC to Applesoft BASIC so that

   a) editing would be easy,
   b) variables could designated with letters and numbers more closly resembling their names (HP-75 allows at most one letter and one number),
   c) graphing could be done,
   d) copies of the program could be distributed to others,

2) run the program for a large number of different situations to

   a) determine whether or not the predictions of the model are reasonable,
   b) narrow choices of input parameters,

3) modify the program to allow the input of latitude, making the model more "global" (calculations are now done only at 40^{o} latitude),

4) rewrite the code, separating the various modes and repeated blocks of calculations into subroutines to make the program run faster,

5) translate the program into FORTRAN.

THEORY

Consider an atmosphere which is assumed to obey the following relationships:

1) perfect stratification, i.e., density and pressure are functions of altitude only,

2) steady state, i.e., $\partial/\partial t$ of any variable vanishes,

3) inviscid flow,

4) ideal gas law.

The mass conservation equation is

$$\vec{\nabla}\cdot(\rho\vec{v}) = 0 \quad , \tag{1}$$

and the momentum equations are

$$(\vec{v}\cdot\vec{\nabla})\vec{v} + 2\vec{\Omega}\times\vec{v} = -(1/\rho)\vec{\nabla}p - g\hat{k} \quad , \tag{2}$$

where ρ is the density, $\vec{v}$ is the air flow velocity, $\vec{\Omega}$ is the angular velocity of the earth, p is the pressure, and g is the acceleration due to gravity. The unit vector in the vertical direction is $\hat{k}$; $\hat{j}$ points north, and $\hat{i}$ points east. The wind velocity components are

$v_x = u$ = zonal wind,

$v_y = v$ = meridional wind,

$v_z = w$ = vertical wind.

The theoretical model developed by ReVelle is along the lines of the work done by Ghosh (Ghosh, 1970). Ghosh considered a steady state atmosphere with a vertical wind increasing in strength with height, z, and no zonal or meridional wind. Effects of viscosity and rotation of the earth were neglected. The solution of the continuity and momentum equations in this case yields a

density of

$$\rho = \rho_o \exp\{-\int (mg/kT + (1/T)\partial T/\partial z + (m/kT)(mg/kT + (1/T)\partial T/\partial z)w^2) dz\} \quad , \qquad (3)$$

where m is the mass of a molecule, k is Boltzmann's constant, and T is the temperature. The first term in the integral represents the gravity dependence of the density, the second term the temperature variation, and the third term the contribution of the vertical wind.

Substitution of typical values at 100 km yields (Ghosh, 1970)

$$\text{g-term} \sim 1.6(10)^{-4}\ m^{-1} \quad ,$$

$$\text{T-term} \sim 1.4(10)^{-5}\ m^{-1} \quad ,$$

$$\text{total of non-wind terms} \sim 1.7(10)^{-4}\ m^{-1} \quad ,$$

$$\text{w-term} \sim 2.8(10)^{-9} w^2\ m^{-1} \quad .$$

Therefore, if the vertical wind w = 100 m/sec,

$$\frac{\text{w-terms}}{\text{non w-terms}} \sim 10\%.$$

ReVelle considers a rotating earth with non-zero zonal wind and meridional winds and non-zero spatial derivatives of the winds. The continuity equation becomes

$$\rho\partial u/\partial x + \rho\partial v/\partial y + \rho\partial w/\partial z + w\partial\rho/\partial z = 0 \quad , \qquad (4)$$

or

$$\rho D_h + \rho\partial w/\partial z + w\partial\rho/\partial z = 0 \quad , \qquad (5)$$

where D_h, the horizontal divergence, is assumed to be constant for all z.

The atmosphere is broken up into layers of thickness Δz ($\sim$.1 km), and the conservation equation and momentum equations are numerically integrated from a lower boundary of altitude z = 1 km. The initial value of the vertical wind is taken from Ekman pumping theory in which

$$w_o \sim \zeta_z h_{PBL}/6 \quad , \tag{6}$$

where ζ_z is the vorticity (in sec^{-1}) of some rotating body of air at the surface and h_{PBL} is the height of the planetary boundary layer, 1 km.

Note that if w is assumed a constant, w_o, over some layer of thickness Δz, then equation (5) can be written as

$$\rho w = \rho_o w_o \exp\{-(D_h/w_o)\Delta z\} \quad . \tag{7}$$

The case in which $D_h = 0$ and the case in which w is very large should both act similarly with $\rho w \sim$ constant. Also note that if w becomes negative (a downdraft), the exponent becomes positive, causing the density to increase with height. This sign in the exponent prevents a downdraft from becoming an updraft again; therefore, the solutions which were taken to be acceptable in this study were the ones in which the vertical wind was always upward.

It would be desirable to also utilize the conservation of energy equation in the model. However, due to the complexity of the equation and the uncertainty in values of some variables, it was decided to stick with only the mechanical relationships. A rough calculation does reveal that the winds that would be expected from an energy point of view are compatible with those being calculated by the model. The heating due to vertical lifting is calculated in the program (variable name Q7), and the numbers obtained are consistent with calculations which show heating and cooling rates of 1 - 10 K/day in the lower 100 km of atmosphere and up to 1000's K/day in the thermosphere.

The details of using the program are given in the next section. The rest of this section is devoted to a discussion of a few items essential to the understanding of the operation of the program.

There are a number of program input requirements; the first is that of mode. The program operates in one of four modes:

1) barotropic atmosphere with vertical winds (the only mode run in this study)

2) sub/super geostrophic with either constant or height variable vertical winds)

3) sub/super geostrophic with waves (includes nonlinear saturation limit)

4) hydrostatic with waves (includes nonlinear saturation limit).

A response of "yes" to the input concerning heating constraint forces the vertical wind to become constant when the heating due to vertical lifting becomes greater than 100 K/day. A response of "no" allows the vertical wind and its associated heating to increase without bound.

The value of vorticity input is used to calculate the initial vertical wind in equation (6). The number most often used in this study is $0.606433553(10)^{-5}$ sec^{-1}; it generates an initial vertical wind of 1.00687771 mm/sec. This vorticity was found by ReVelle prior to this study to yield results (winds, densities, etc.) which are realistic and well behaved.

The DEL:H DOT V and DELH.V terms in the program refer to the horizontal divergence, $D_h = \partial u/\partial x + \partial v/\partial y$, in equation (5). A response of "yes" to the "NOMINAL DELH.V CASE?" input causes the value of D_h to be set at $1.21915(10)^{-7}$ sec^{-1}, which was also previously determined by ReVelle to yield reasonable results. An input of "no" prompts a request for the desired value of D_h. Whatever the value of D_h used in the program, it is held constant everywhere.

The zonal wind can be either constant with altitude or variable according to the season. A constant zonal wind is obtained by answering "no" to the "VARIABLE CORIOLIS TERM?" input and then typing in the desired wind. Otherwise, the zonal wind is variable according to whether winter or summer is chosen. In this case the wind is assumed geostrophic such that

$$u = -(R/fM)\partial T/\partial y \quad , \tag{8}$$

where f is the Coriolis term ($2\Omega\sin\Theta$), R is the gas constant, M is molecular mass per mole, and $\partial T/\partial y$ is the horizontal temperature gradient. The zonal wind may in addition be made subgeostrophic,

geostrophic, or supergeostrophic via the u geostrophic factor (<1, 1, >1, respectively) which simply multiplies the value calculated in equation (8).

OPERATING INSTRUCTIONS

Step-by-step instructions on how to use the Apple II+ or IIe (or III with II emulator) to run AIR and its derivatives are given in this section.

1) Insert the disk into the disk drive and close it.

2) Turn on the computer (switch is on your left, at rear of computer) and the monitor.

3) The menu of the disk contents will appear on the screen.

4) Type RUN AIR (or RUN AIR PRINT or RUN AIR LARGE SCALE) and press the "return" key.

5) Respond to each input request by typing the number and pressing "return".

6) When the run is complete, remove the disk and turn off the computer and monitor.

Notes on AIR:

* Input

  MODE (1, 2, 3, 4) =
  (Mode 1 was used for all runs.)

  HEATING CONSTRAINT: 1(Y), 0(N) =
  (0 used for all runs.)

  PBL Z VORTICITY: MEAN (1*E-5/SEC)
  (0.606433553 used for most runs.)

  NOMINAL DELH.V CASE? (1/YES, 0/NO)
  (YES chooses 1.21915.)

  VARIABLE CORIOLIS TERM? (1/YES, 0/NO)
  (Variable or constant zonal wind.)

ZONAL WIND =
(Appears only if Variable Coriolis Term? = NO.)

WINTER/SUMMER CASE? (1/WIN, 0/SUM)
(Appears only if Variable Coriolis Term? = YES.)

U GEOSTROPHIC FACTOR =
(1 used for all runs.)

DEL:H DOT V(1*E-7/SEC) =
Horizontal divergence in units of $(10)^{-7}$ sec^{-1}; usually in the range of 1.2. (Appears only if Nominal DEL.V Case? = NO.)

* Output

Any number of variables can be listed on the screen. Which variables are to be listed is set in line number 18700 (and 18710 if the list is very long). The altitude, L8, is already part of the PRINT statement. If, for example, the variable Y4 is also needed, then this sequence of steps should be followed:

1) Instead of typing RUN AIR, type LOAD AIR.

2) Type LIST 18700-18800 (press "return") to see what the PRINT statement looks like.

3) Retype line 18700. In this example it would be
18700 PRINT:PRINT "L8 = ";LR:PRINT "Y4 = ";Y4

4) Type RUN and press "return".

The program may be stopped at any time by pressing the CTRL ("control") and C keys simultaneously. The current value of a variable, for example R6, is listed by typing ?R6 (press "return"). The program operation is continued with CONT.

* Some useful variables and their code names:

LR - altitude (rounded to the nearest 0.01 km)
T1 - vertical temperature gradient
T2 - temperature
M - density
M9 - standard atmosphere density
P1 - pressure
P9 - standard atmosphere pressure
Y4 - % difference between M and M9
Y9 - % difference between P1 and P9

R6 - Reynolds number
G1 - density scale height
S2 - pressure scale height
W - vertical wind
M0 - Mach number
Q7 - heating due to vertical lifting
U2 - zonal wind
H0 - horizontal divergence

Notes on AIR LARGE SCALE:

This program was created to eliminate the need to modify AIR every time a run up to 300 km was desired. The main difference between AIR LARGE SCALE and AIR is in the definition of the x- and y-coordinates on the graph. The altitude scale goes to 300 km (as opposed to 150 km in AIR) and the wind scale goes to 14 mm/sec (2.8 mm/sec in AIR).

Other variables may be graphed by redefining the array variable W(Z) used in the graphics. W(Z) is dimensioned in line 50 and defined in line 18630.

Notes on AIR PRINT:

This program was created to give a paper printout of the variables being listed in line 18700. The printer is accessed through slot #1 (PR#1 in line 3310), and it must be turned on before the program is run. The program is set to run until the altitude reaches 500 km, but it can be stopped at any point with CTRL-C. Type PR#0 after this to return to use of the monitor.

RESULTS

This section comprises a summary of the results of the runs of AIR and their significance. All runs were made with the vorticity $\zeta_z = 0.606433553(10)^{-5}$ sec^{-1} unless otherwise stated.

1) u = 10 m/sec, D_h nominal

Figure 1 shows a typical behavior of the vertical wind. A later run of this case to higher heights shows that the vertical wind increases to 10 m/sec at 155 km, where the density is 0.2 % greater than the 1976 Standard Atmosphere (National Oceanic and Atmospheric Administration, 1976) density, to 100 m/sec at 214 km,

where the density is 1.3 % less than standard, to almost 400 m/sec at 257 km, where the density drops to 16.6 % less than standard.

It is noted here that the shape of the curve in Figure 1 is very similar to the temperature profile, Figure 2. Because of this similarity, it was decided to attempt to find values of D_h and u which would cause the w vs. z graph to follow the temperature curve. The results of this search are given later in this section.

2) u = 0, D_h nominal

Figure 3 shows how the vertical wind profile changes with the zonal wind reduced to zero. A negative (toward the west) zonal wind drives the vertical wind to large values at even lower heights. See Figure 4.

3) u = 10 m/sec, $D_h = 1.2(10)^{-7}$ sec^{-1} and
u = 10 m/sec, $D_h = 1.24(10)^{-7}$ sec^{-1}

These graphs demonstrate how sensitive the w vs. z graphs are to the value of the horizontal divergence. A value of D_h only slightly smaller than $1.21915(10)^{-7}$ sec^{-1} causes the wind to increase very rapidly (Figure 5). In fact a zero divergence yields a vertical wind of 200 m/sec at only 70 km. A value of D_h slightly larger than nominal drives the wind negative very quickly. See Figure 6.

A comparison of the u = 0 and D_h = 0 case with Ghosh's results might be in order, although they will still differ somewhat due to Ghosh's assumption of a non-rotating earth. A run with u = 0 and D_h = 0 yields a vertical wind of 100 m/sec at 83 km ($\sim$ 100 km), which produces a density 9 % ($\sim$ 10 %) smaller than Standard Atmosphere density, which includes the gravity and temperature variations.

4) u = 11.723013 m/sec, D_h nominal

Figure 7 shows that this value of a constant zonal wind constrains the wind to follow the temperature profile for a good distance and limits it to a few cm/sec near 300 km.

Figure 8 demonstrates how the winds vary as a result of a change in only the last digit of the zonal wind. The vertical wind is just beginning to decrease at 300 km and will eventually go negative.

5) Winter and summer zonal winds

Figures 9 and 10 show the zonal (geostrophic) wind profile for winter and summer, respectively.

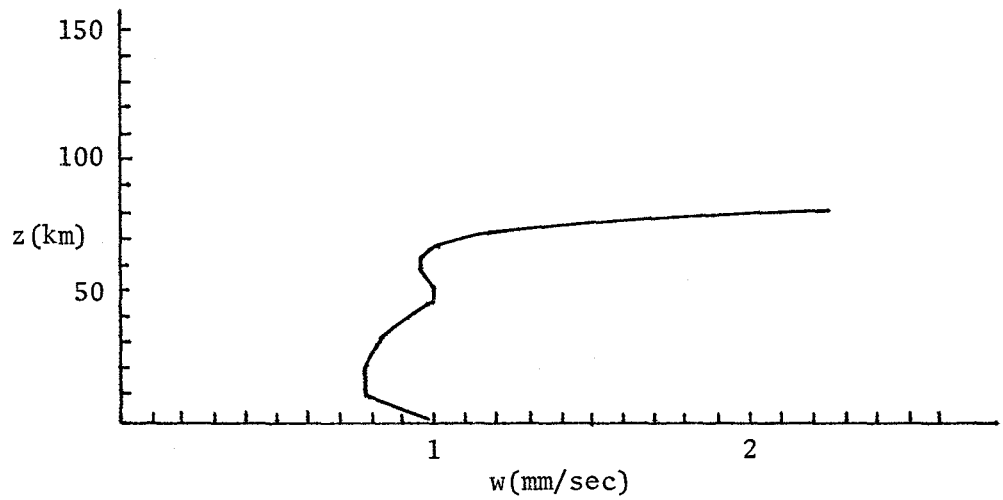


Figure 1. Vertical wind vs. altitude for u = 10 m/sec and $D_h = 1.21915(10)^{-7}$ sec^{-1}.

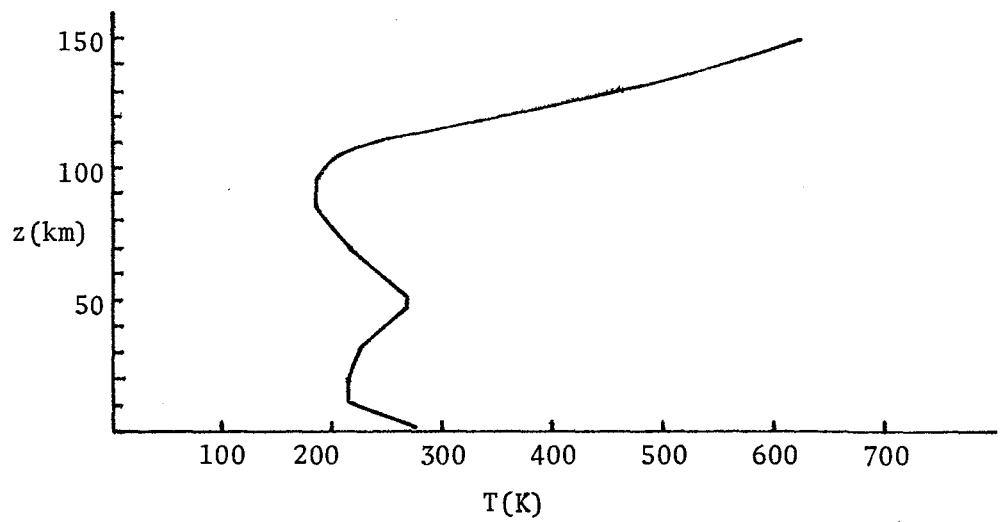


Figure 2. Temperature profile.

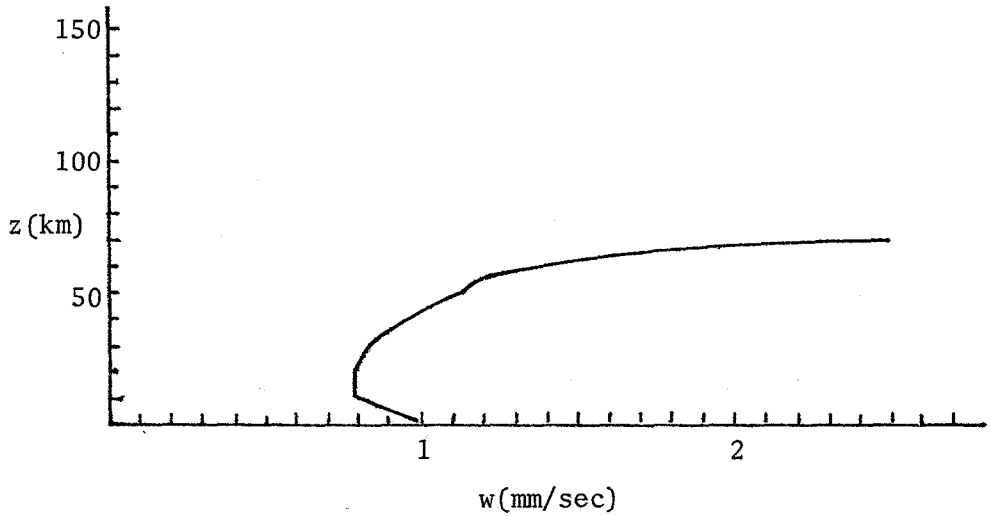


Figure 3. Vertical wind vs. altitude for u = 0 m/sec and $D_h = 1.21915(10)^{-7}$ sec^{-1}.*

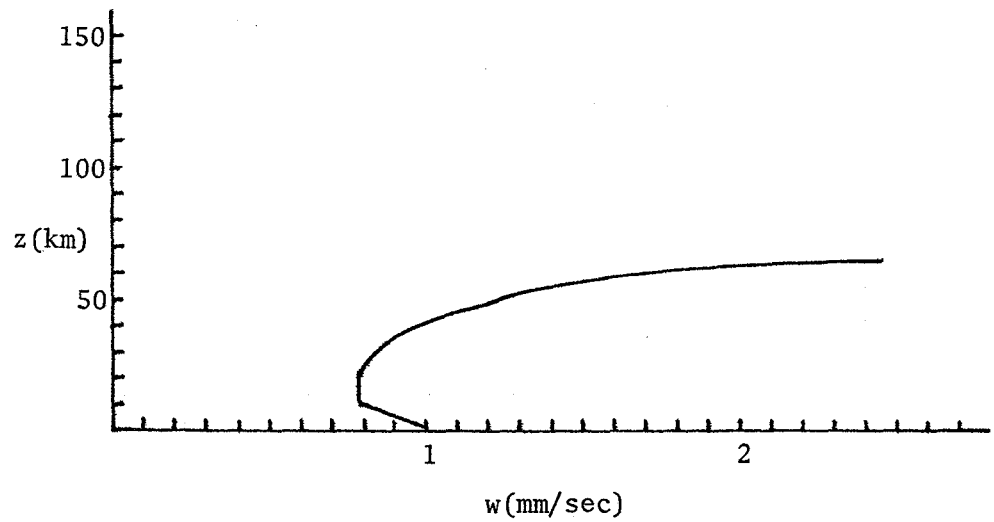


Figure 4. Vertical wind vs. altitude for u = -10 m/sec and $D_h = 1.21915(10)^{-7}$ sec^{-1}.

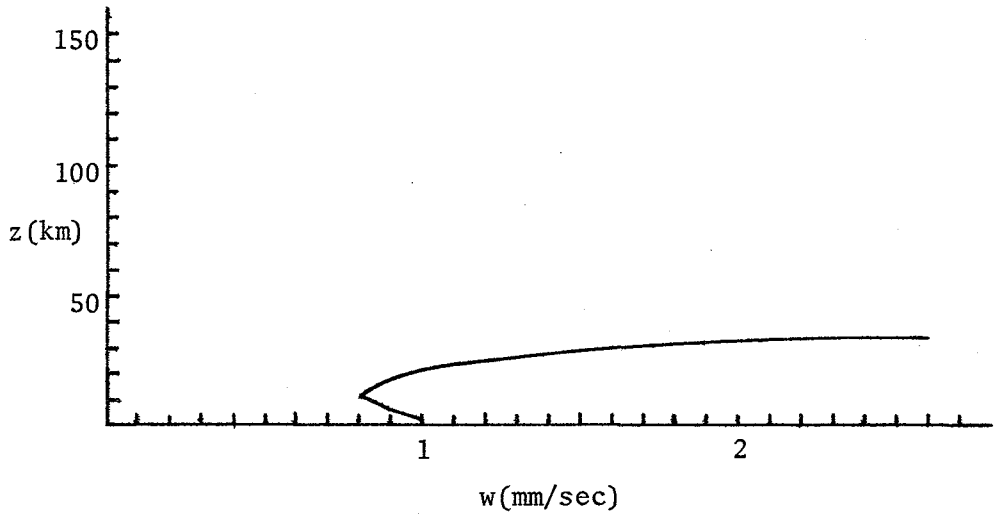


Figure 5. Vertical wind vs. altitude for u = 10 m/sec and $D_h = 1.2(10)^{-7}$ sec^{-1}.

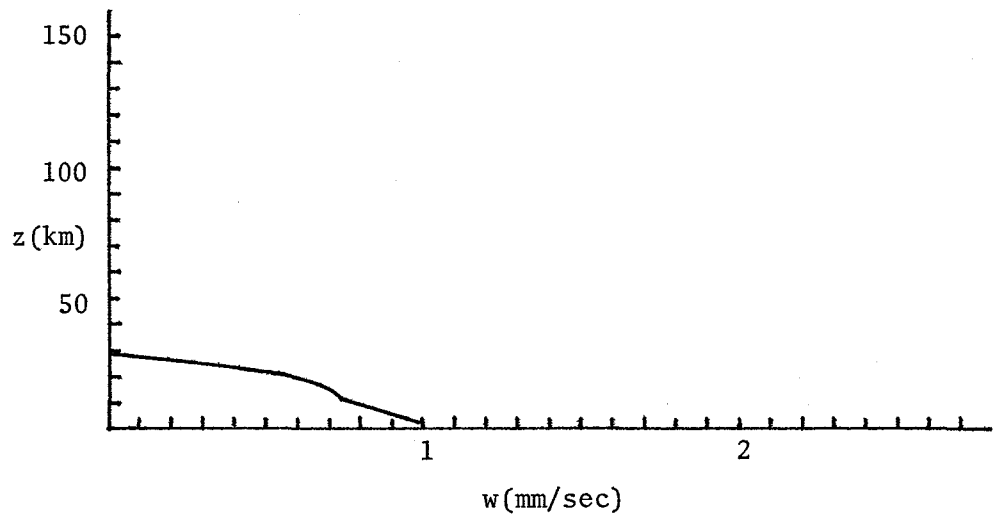


Figure 6. Vertical wind vs. altitude for u = 10 m/sec and $D_h = 1.24(10)^{-7}$ sec^{-1}.

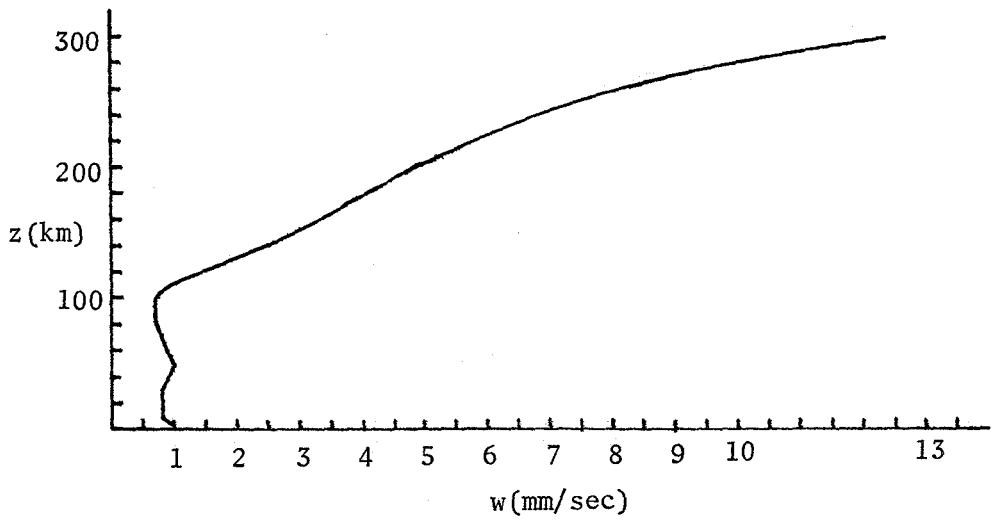


Figure 7. Vertical wind vs. altitude for u = 11.723013 m/sec and $D_h = 1.21915(10)^{-7}$ sec^{-1}.

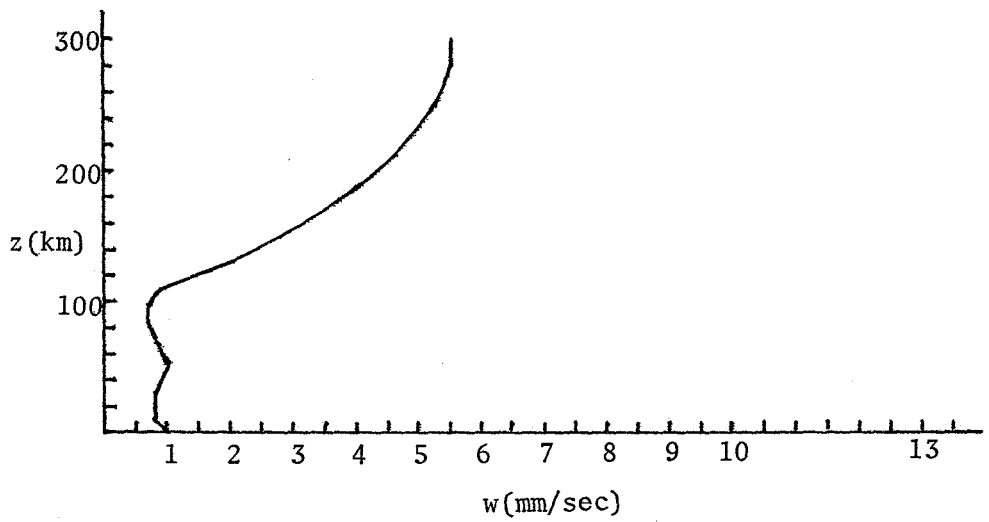


Figure 8. Vertical wind vs. altitude for u = 11.723014 m/sec and $D_h = 1.21915(10)^{-7}$ sec^{-1}.

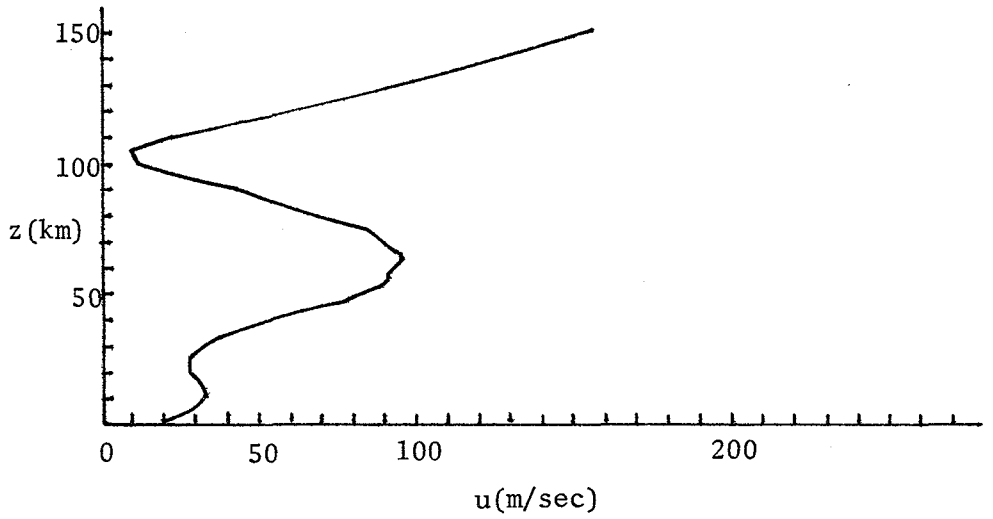


Figure 9. Winter zonal wind profile.

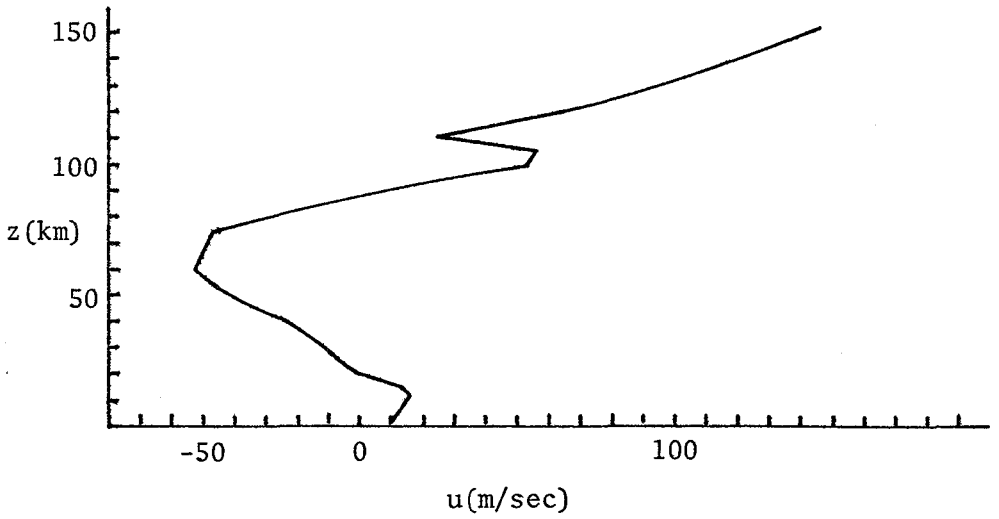


Figure 10. Summer zonal wind profile.

6) u = summer profile, $D_h = 1.21917253(10)^{-7}$ sec^{-1}

This value of the horizontal divergence was found to be the one which constrained the vertical wind to follow the temperature profile as closely as possible in the case of summer zonal winds. The vertical wind (see Figure 11) is only about 1 cm/sec at 170 km, and rises to 10.4 m/sec at 500 km, where the density is 3.2 % above the Standard Atmosphere density.

7) u = winter profile, $D_h = 1.21892828(10)^{-7}$ sec^{-1}

Similarly, this constraining value of D_h was found for the winter case. See Figure 12. The vertical wind increases to 1 cm/sec at 155 km, where the density is 1.3 % greater than standard, to 21.5 m/sec at 500 km, where the density is 4.9 % greater than standard.

8) u = 10.105 m/sec, $D_h = 1.21917253(10)^{-7}$ sec^{-1}

It was noted early in the study that for a nominal D_h the w vs. z graphs for u = 10 m/sec and u = summer profile were almost identical and that the u = 25 m/sec and u = winter profile were almost identical. The ratio of these two constant zonal winds is 2.5, which is roughly equal to the ratio of the altitude-averaged zonal winds for winter and summer. Therefore, once the constraining values of the horizontal divergence were found for winter and summer, the corresponding constant zonal winds for these divergences were sought. Figure 13 shows how the vertical wind follows the temperature profile for u = 10.105 m/sec.

9) u = 27.6 m/sec, $D_h = 1.21892828(10)^{-7}$ sec^{-1}

The constrained vertical wind for the case of a constant zonal wind and the winter divergence is graphed in Figure 14. Again, note that the ratio of these two constant zonal winds in the winter and summer cases is 2.73 ($\sim$ 2.5). Time did not permit the determination of the zonal winds to more decimal places.

10) Note on heating

As discussed earlier, if the theory is to be reasonable, the heating due to the vertical winds must be within the limits mentioned. A run of the constrained summer case produced vertical wind heating which a) followed the temperature profile, b) was less than or approximately equal to 1 K/day up to 108 km, c) hit 100 K/day at 279 km, and d) rose to 3700 K/day at 500 km.

The heating due to vertical lifting in the constrained winter case behaved much like the summer case except that the heating hit 100 K/day at 242 km and 7700 K/day at 500 km.

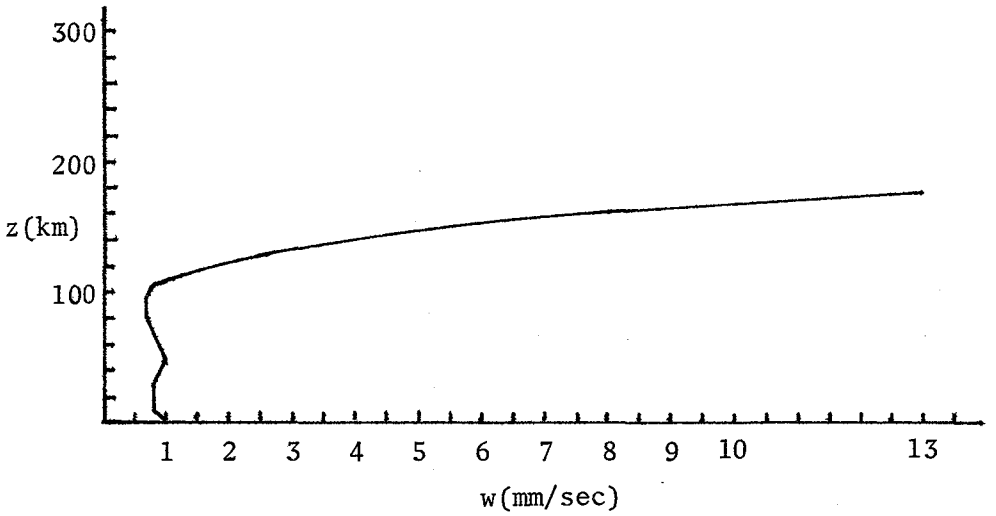


Figure 11. Vertical wind vs. altitude for u = summer and $D_h = 1.21917253(10)^{-7}$ sec^{-1}.

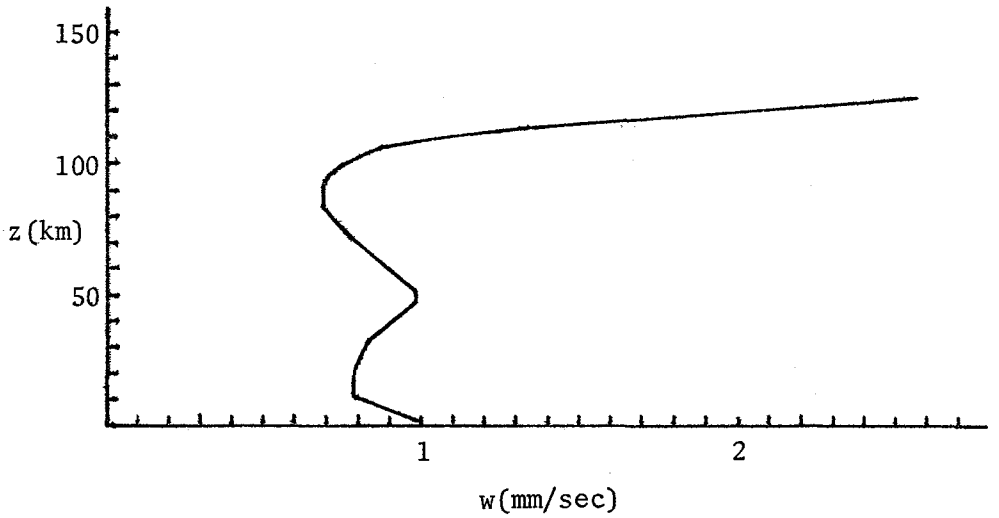


Figure 12. Vertical wind vs. altitude for u = winter and $D_h = 1.21892828(10)^{-7}$ sec^{-1}.

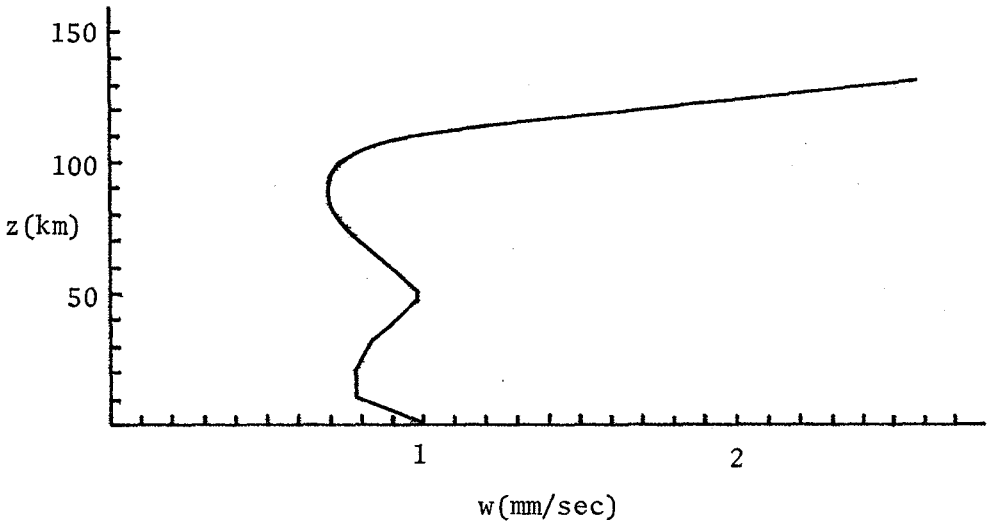


Figure 13. Vertical wind vs. altitude for u = 10.105 m/sec and $D_h = 1.21917253(10)^{-7}$ sec^{-1}.

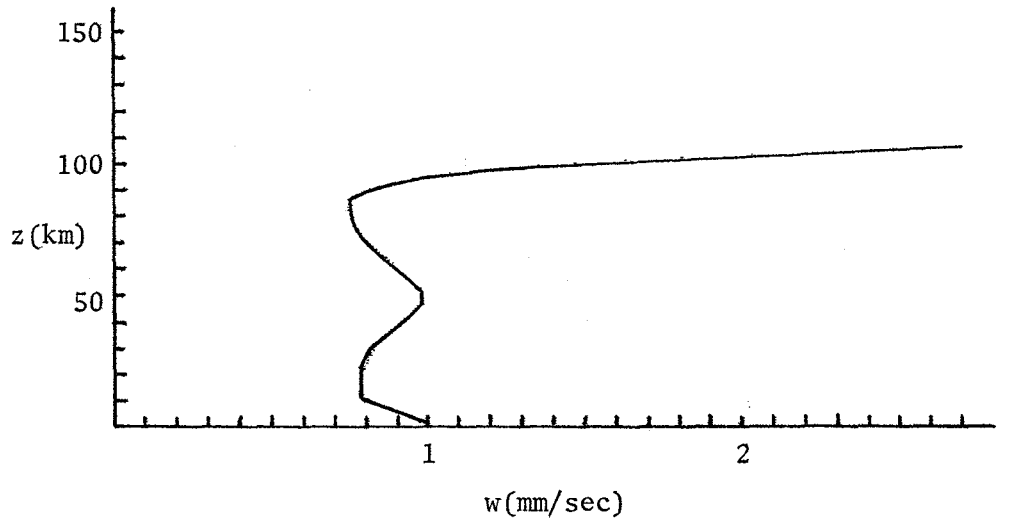


Figure 14. Vertical wind vs. altitude for u = 27.6 m/sec and $D_h = 1.21892828(10)^{-7}$ sec^{-1}.

11) Note on mass flux

A run was made to test the earlier assertion that as w gets very big, ρw, the mass flux in $kg \cdot m^{-2} \cdot sec^{-1}$, should become constant. Figures 15 and 16 give the wind vs. altitude and mass flux vs. altitude graphs, respectively. The vertical wind is 435 m/sec at 263 km, and the density is 21 % smaller than standard. The mass flux reaches a constant value of $1.5(10)^{-8}$ $kg \cdot m^{-2} \cdot sec^{-1}$ at around 120 km.

At this rate of mass flux the earth's atmosphere would be lost in

$$t \sim \frac{5.3(10)^{18} \text{ kg}}{\{1.5(10)^{-8} \text{ kg/m}^2 \cdot \text{sec}\}\{4\pi(6.7(10)^6 \text{m})^2\}} \sim 5.2(10)^{11} \text{ sec}$$

or about 20,000 years. Even if it is only the upper 1 % that is lost, the time is reduced to 200 years. These numbers are not meant to show that the atmosphere is actually dissipating at this rate, rather that the quantities produced by AIR are reasonable. That is, the time is not as short as 2 seconds, nor is it as long as 2 billion years.

12) Note on pressure variations

All runs yielded a pressure variation (variable Y9) very close to the density variation (variable Y4). So, when the density was a certain percent above (below) the standard, the pressure was also that percent above (below) the standard.

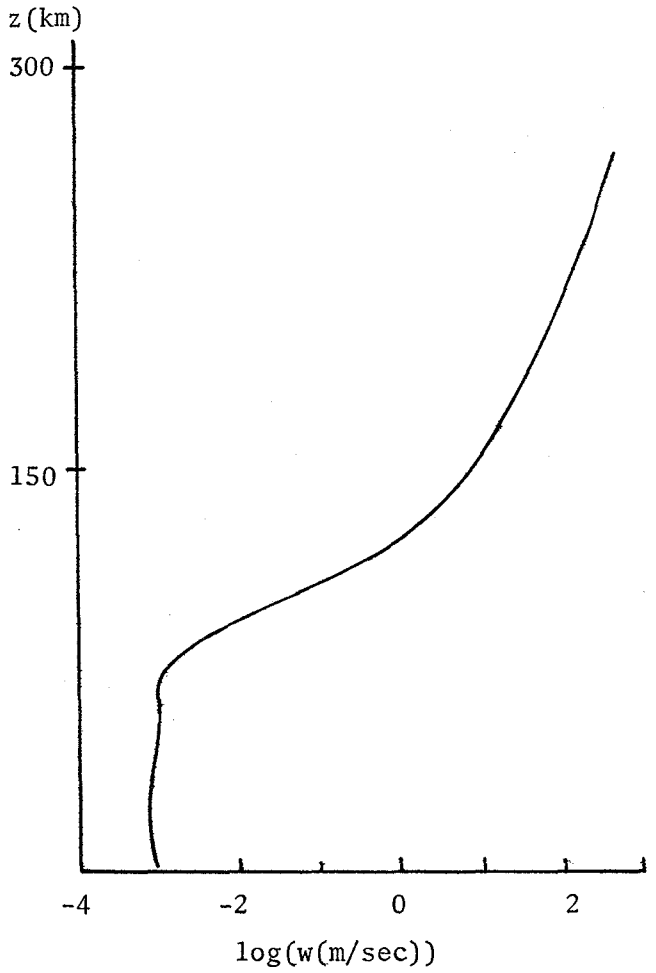


Figure 15. Vertical wind vs. altitude for $\zeta_z = 0.57(10)^{-5}$ sec^{-1}, u = 10 m/sec, $D_h = 1.21915(10)^{-7}$ sec^{-1}.

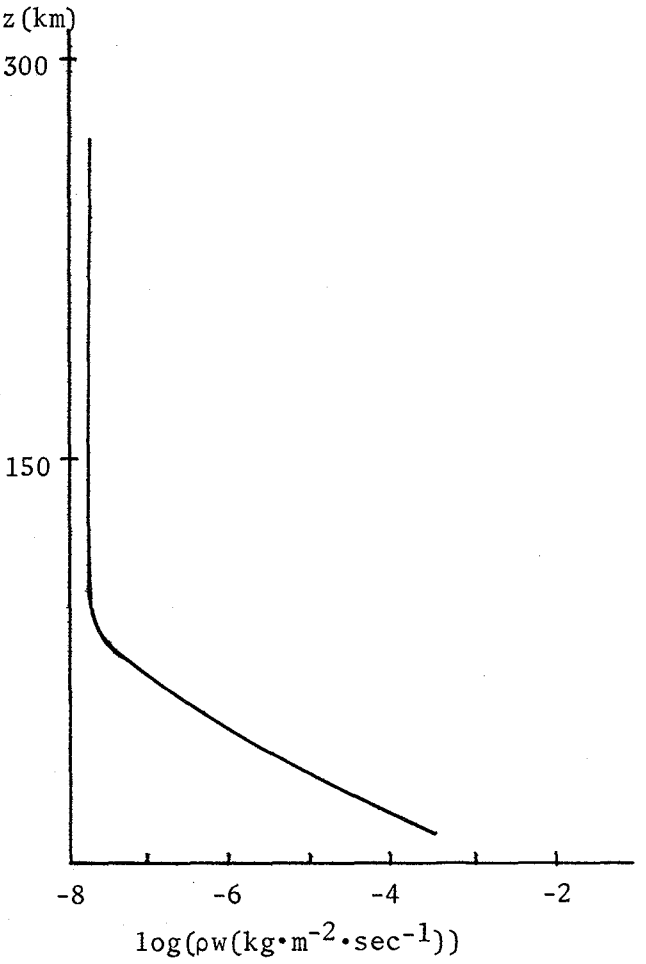


Figure 16. Mass flux vs. altitude for $\zeta_z = 0.57(10)^{-5}$ sec^{-1}, u = 10 m/sec, $D_h = 1.21915(10)^{-7}$ sec^{-1}.

CONCLUSIONS AND RECOMMENDATIONS

Objectives 1) and 2) were achieved with the possible exception of 1b); some variables have been renamed, but the changes have not been made in the code. Due to the fact that the code was still in a state of flux throughout most of the summer, time was short and objectives 3) and 4) were not achieved. They will remain as long term objectives. Objective 5) was never really attempted because a) the Apple microcomputer is so convenient in terms of accessibility and editing and graphing capabilities, b) time on the bigger computers is very limited, c) a decrease in run time by a factor of four on the Hewlett-Pachard 1000F is, although desirable, not significant, and d) time, as mentioned above, was limited. Perhaps, once the present code is brought to its final form, the translation to FORTRAN can take place.

As a result of this study, the following recommendations are put forth:

1) continue work on the code to achieve objectives 3) and 4),

2) extend the work done on objective 2) to the other modes in the program,

3) study the possibility of allowing for a variable horizontal divergence, rather than keeping it constant for all altitudes.

REFERENCES

Ghosh, S. N., "Effect of motion on the altitude distribution of atmospheric density", Annales de Geophysique 26, 795-799 (1970).

National Oceanic and Atmospheric Administration, U. S. Standard Atmosphere, U. S. Government Printing Office, Washington (1976).

ReVelle, D. O., (report in preparation, 1985).

1985

NASA/ASEE SUMMER FACULTY RESEARCH FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

OBSERVATIONAL AND NUMERICAL STUDIES OF THE LOW-LEVEL JET STREAM IN THE ATMOSPHERE

Prepared By:	Dušan Djurić, Dr. Sc.
Academic Rank:	Professor
University and Department:	Texas A&M University Department of Meteorology
NASA/MSFC: Division: Branch:	 Atmospheric Sciences Systems Dynamics Laboratory
MSFC Counterpart:	Michael Kalb (USRA)
Date:	12 August 1985
Contract No.:	NGT 01-008-021 The University of Alabama in Huntsville

OBSERVATIONAL AND NUMERICAL STUDIES OF THE LOW-LEVEL JET STREAM IN THE ATMOSPHERE

BY

Dušan Djurić
Professor of Meteorology
Texas A&M University
College Station, Texas

ABSTRACT

Two cases of low-level jet (LLJ), studied with the Atmospheric Sciences Division MCIDAS facility, support the hypothesis that the LLJ is an integral part of synoptic-scale cyclogenesis. The appearance of an arch with a row of cumulonimbus is discovered at the front of the LLJ. The inertia oscillations in the LLJ are pointed out as part of the quasi-diurnal variation. A plan for further research by numerical modelling is proposed, which will bring more clarification of the mechanics of the LLJ.

LIST OF FIGURES

1. Introduction

The low-level jet is the appearance of a wind maximum in the lowest 2 km of the atmosphere. For the purpose of selection of cases, it is practical to use Bonner's (1968) requirement that only a maximum wind of over 12 m s^{-1} is considered, followed with the decrease of wind speed by at least of 2.5 m s^{-1}, all under 2 km elevation above the ground level. The LLJ is recognized as the most important mechanism for transport of water vapor into the storms of middle latitudes (Palmen and Newton, 1969, p.385).

The scientific crew and the technical facilities at the Atmospheric Sciences Division at MSFC/NASA offer unique possibilities to conduct an intensive research of the LLJ. The personnel at this division enjoys high reputation in meteorological research through its publications. The technical facilities at MSFC encompass the McIDAS computerized system for analysis of weather charts and satellite photographs and a battery of computers for modelling of atmospheric processes.

Previous research of the LLJ has led to two different points of view about the dynamics of this jet: (a) theories that the LLJ is a local consequence of momentum transfer in the boundary layer (Blackadar, 1957; Wippermann, 1973), and (b) theories that the LLJ is part of the large-scale dynamical processes in the atmosphere (Uccellini and Johnson, 1979; Djurić, 1981; McNider and Pielke, 1981). Clarification of the controversy between these types of theories is of significant importance, since only with sound theoretical basis we can expect successful modelling and forecasting of the large increase of humidity that occurs before the storms of middle latitudes over North America.

This research opens the way to develop a systematic program for research of the LLJ. The diagnostic part utilizes the most modern analysis tools for proper assesment of physical circumstances that accompany the LLJ. The findings from the diagnostic studies will then be used as initial conditions in the LAMPS model for simulation of the atmosphere. Pairs of computations with altered physics will show the dynamics of the LLJ, when the crucial parameters are altered between the experiments.

2. Objectives

The task of this research is to attempt to identify the mechanism of the LLJ using observations and numerical experimentation. This general objective leads to three sub-tasks:

(a) Observational: to use the McIDAS facility for construction of weather charts and satellite photographs that give insight in the mechanism of the

LLJ.

(b) Numerical: to reproduce LLJ under controlled physical circumstances, when the numerical results can be compared to the observed states.

(c) Physical: to follow a definite physical theory, so that the analysis of observations and numerical experimentation can be scientifically guided.

3. Procedures

The exposed objectives were approached when a number of cases with LLJ was analysed using the advanced facilities of McIDAS. After an investigation of days with LLJ and data available on the McIDAS system, it was deemed practical to investigate two cases of LLJ over the Great Plains: of 5-6 May and of 2-3 December, both in 1983. The data needed for this study encompassed the conventional surface and upper-air meteorological observations, and the satellite photographs. This part of research was significantly aided by Mr. Paul Meyer of NASA/MSFC.

The numerical simulation of the LLJ in this project was started using the LAMPS model for simulation of mesoscale atmospheric processes. This model, originally developed by D. Perkey of Drexel University, is operational on the Perkin-Elmer computer at the MSFC. A test run of the model, using the SESAME data (of 24 April 1982) showed that the vertical resolution in the model is adequate for simulation of the LLJ. This part of the research was conducted in close cooperation with Dr. Michael Kalb of USRA/MSFC.

The theoretical basis for the mechanism of the LLJ, as tentatively adopted for this study, is based on my previous work (Djurić, 1981). The geostrophic adjustment, in combination with inertia oscillations, so far seems capable to explain the observed development of the LLJ.

4. The case study of 5-6 May 1983

Radiosonde soundings, as available on the McIDAS, made it possible to extract the maximum wind in the lowest 2 km of the atmosphere, in all stations that showed such a maximum. These results are summarized in Fig. 1 which shows the wind at three subsequent synoptic hours. The first frame shows the first occurence of the LLJ in this sequence, this was at 0000 GMT over Dodge City, Kansas. In the next observation hour, at 1200 GMT, the LLJ is widespread from West Texas to Nebraska. The intensification of the LLJ can be understood in terms of the boundary-layer theory, which calls for maximum wind speed at the time of day (local dawn) when the stability is strongest. The third observation hour, 0000 GMT 6 May 1983, shows that the

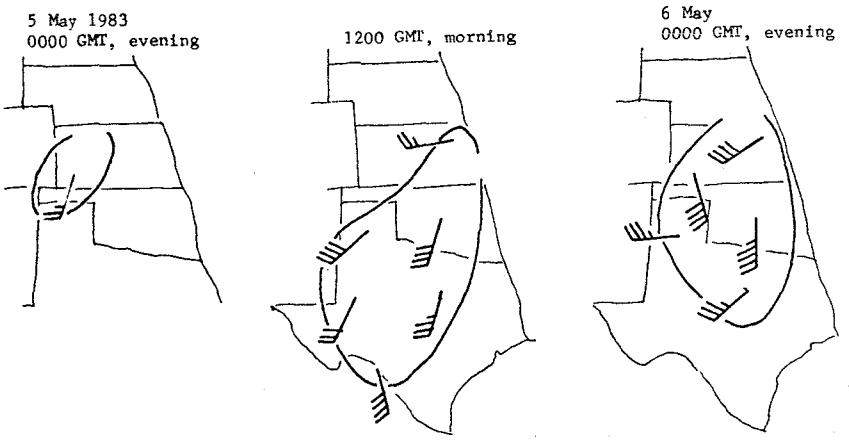


Fig. 1. Development of the LLJ on 5 May 1983. The wind speed is in knots, at the LLJ level. The isotach of 23 kt circles the region with the LLJ. The data outside this region are not plotted.

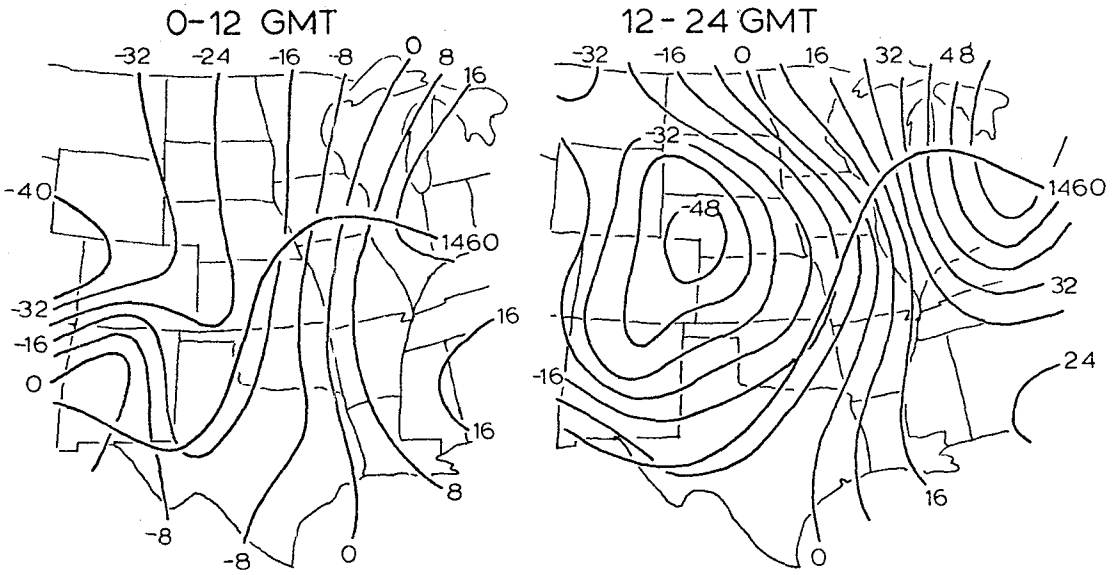


Fig. 2. 12-h height change at 850 mb 0-12 and 12-24 GMT on 5 May 1983, in gpm.

wind speed in the LLJ has further intensified. At this hour the wind reached 44 kt over Oklahoma City. A physical explanation of this wind intensification must be sought in other processes other than the strengthening of the nocturnal inversion. This hour is the local evening. Using the assumption of Uccellini and Johnson (1979), and of Djurić (1981), the general redistribution of pressure in the area can explain the development of the LLJ. The height change at 850 mb between the observation hours every 12 h is shown in Fig. 2. The isallobaric wind of over 1 m s^{-1} can be evaluated over Oklahoma and Kansas. (The expression for the isallobaric wind is shown below in Sec. 7.) Since this wind points across the contours, the air motion is accelerated, thus the LLJ increases with time, irrespectively from the boundary-layer processes.

The other physical problem in this situation is: why does the wind not accelerate in the whole troposphere, instead only in the lowest layer with the LLJ? Now we can recall the premise of Djurić (1981), based on the theory of geostrophic adjustment, that the air can be set in motion only in stable layers. Unstable and neutral layers do not transmit the signal away from the point of disturbance. The existence of the stable layer in the area of the LLJ enables the acceleration of the air. Vertical sections through the atmosphere are shown in Fig. 3 for the period of this study. The location of stations for the sections is shown in Fig. 4.

In the section of 0000 GMT (Fig. 3, top frame) a LLJ can be recognized only at Dodge City, Kansas. The isentropes in this section are quite sparse in the low troposphere, showing low stability, which is understandable for this is the local evening. Next observing hour (1200 GMT, Fig. 3, middle frame) illustrates higher stability, actually an inversion in the surface layer. Now the LLJ is very prominent, maximum speed is over 20 m s^{-1}. The third section, 0000 GMT 6 May 1983 (Fig. 3, bottom frame), is of the evening hour, but now the wind over Oklahoma City (OKC) reached a speed faster than at the previous morning hour.

For the sake of completeness, the environmental condition are documented here with several synoptic charts. At 850 mb (Fig. 4), the isotherms show a warm ridge over the Rocky Mountains and high plains. The warm trough in the contours at 850 mb is quite typical for the circumstances in which the LLJ develops, as it was shown previously by Djurić and Damiani (1980) and Djurić and Ladwig (1983). It is also very interesting to notice that this weather situation is conducive to severe storms. The typical thunderstorm situation, as shown by Doswell (1980), is very similar to the case described here.

The layers from 700 to 200 mb show an almost straight zonal flow over the Great Plains. Out of these, only the 200-mb chart is shown here. The contours on this chart show this zonal flow (Fig. 5). The subtropical jet stream is located zonally over Texas. The region of Oklahoma-Kansas,

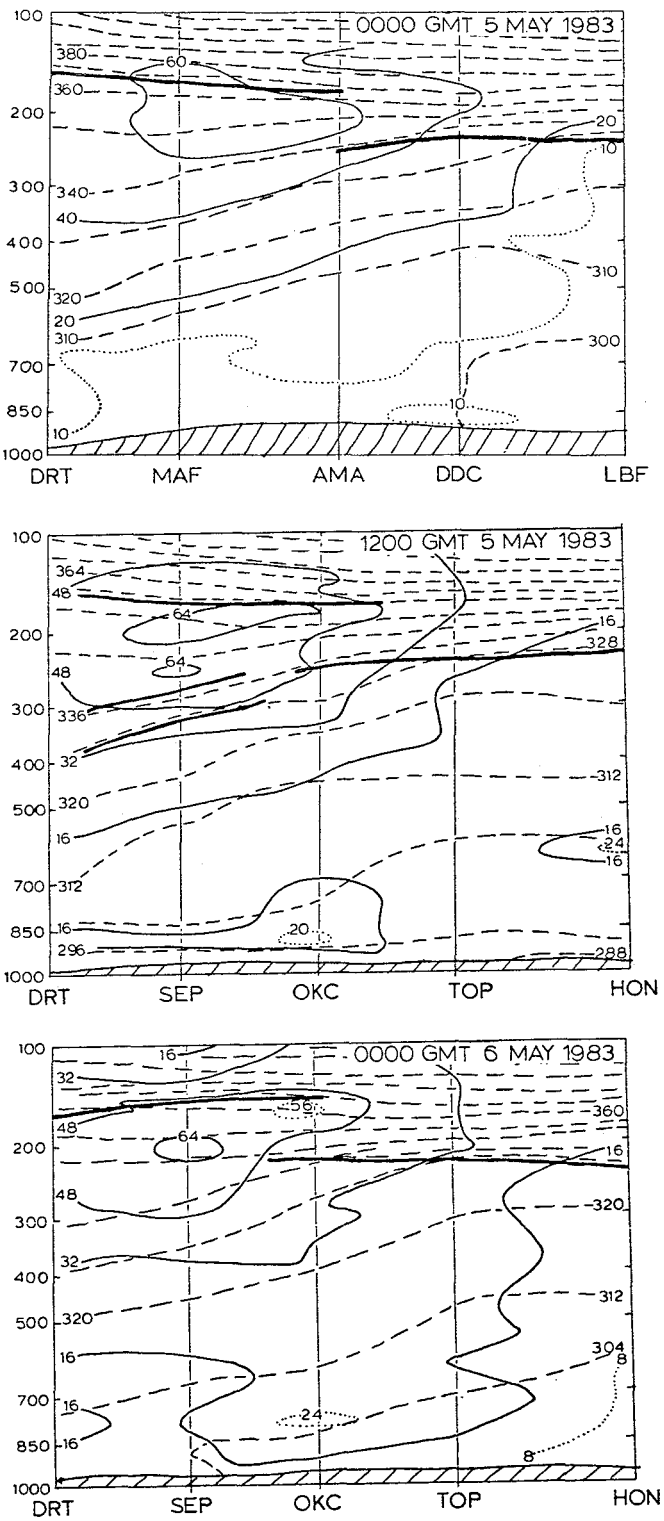


Fig.3. Vertical sections through the atmosphere at 0000 (top) 1200 (middle) and 2400 GMT (bottom) on 5 May 1983. Full lines are isotachs in m s-1, with intermediate values dotted. Dashed lines are isentropes in K of potential temperature. Heavy lines indicate the tropopause and the subtropical front.

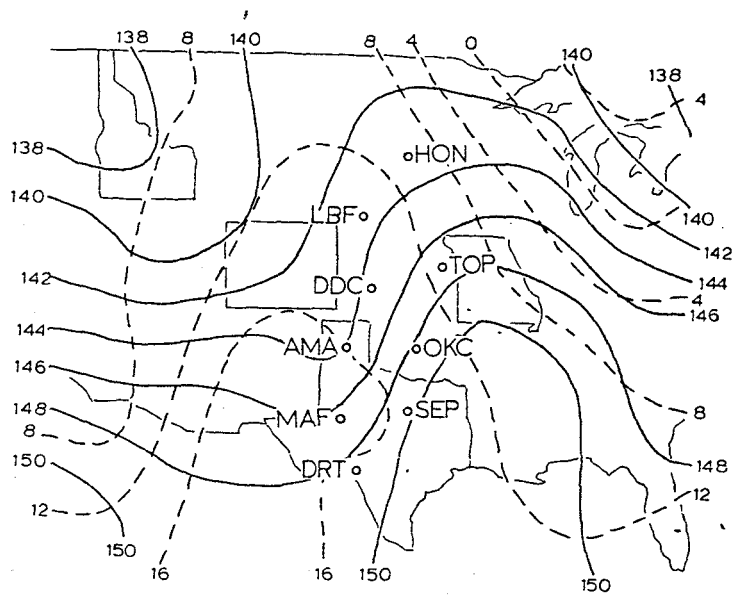


Fig. 4. 850-mb contours (full lines, 10 gpm) and isotherms (dashed, °C) at 1200 GMT 5 May 1983. The stations used in Fig. 3 are shown by the 3-letter call names.

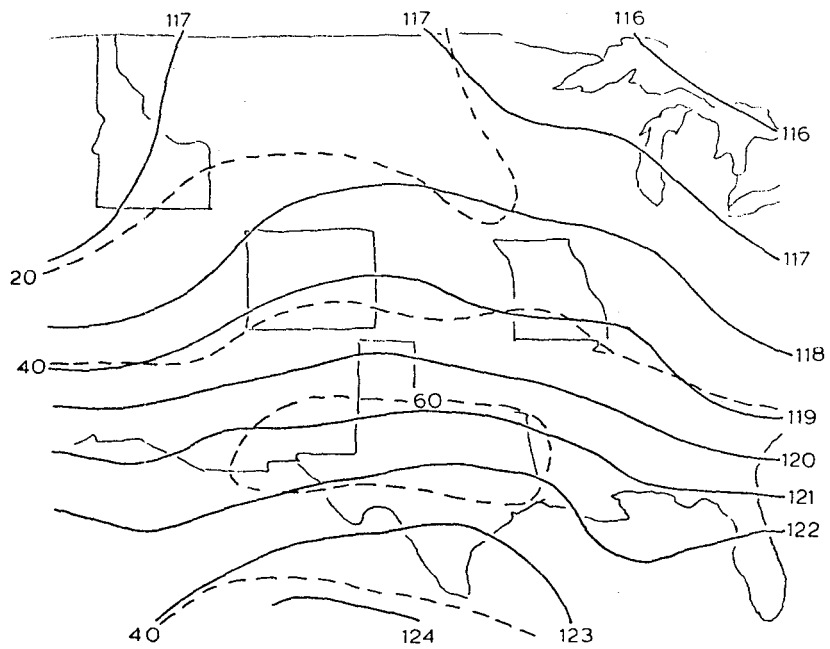


Fig. 5. 200-mb contours (full, intervals of 100 gpm) and isotachs (dashed, 20 m s-1) 12 GMT 5 May 1983.

where the LLJ develops is not associated with large curvature of the flow or with an exit from a jet streak. The polar jet stream is located near Great Lakes. There is a trough in the polar jet stream over the West Coast, this trough is related to pressure fall over the western third of United States. This pressure fall has been illustrated in Fig. 2. It may be pointed out that the trough in the lee of the Rocky Mountains is not in the immediate vicinity of the polar jet stream or the polar front.

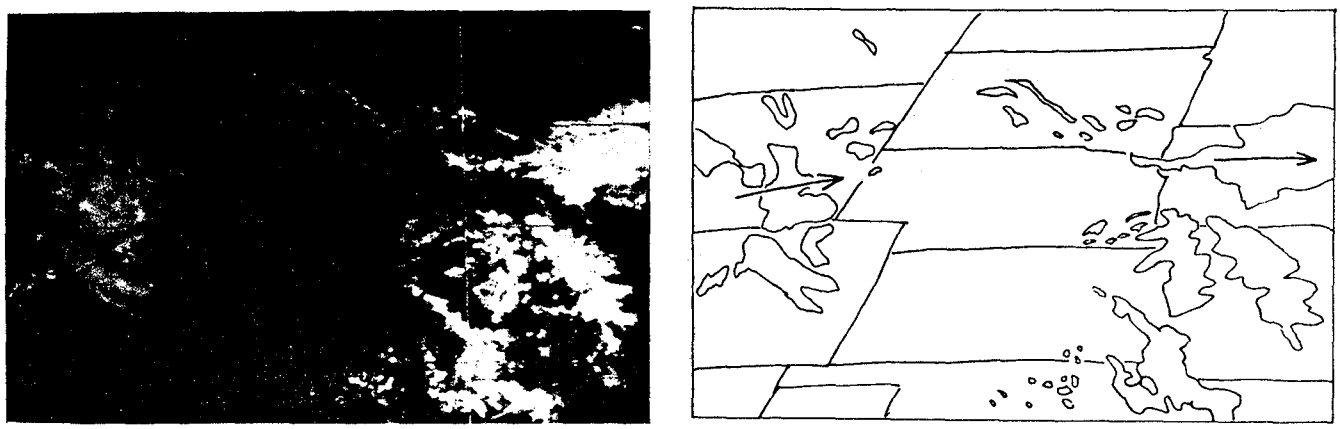
In conclusion, the data from this case may be interpretted to support, but in any case not to contradict, the premise that the LLJ develops in a strong isallobaric field in a stable layer near the earth's surface.

5. The "outflow boundary" with the LLJ

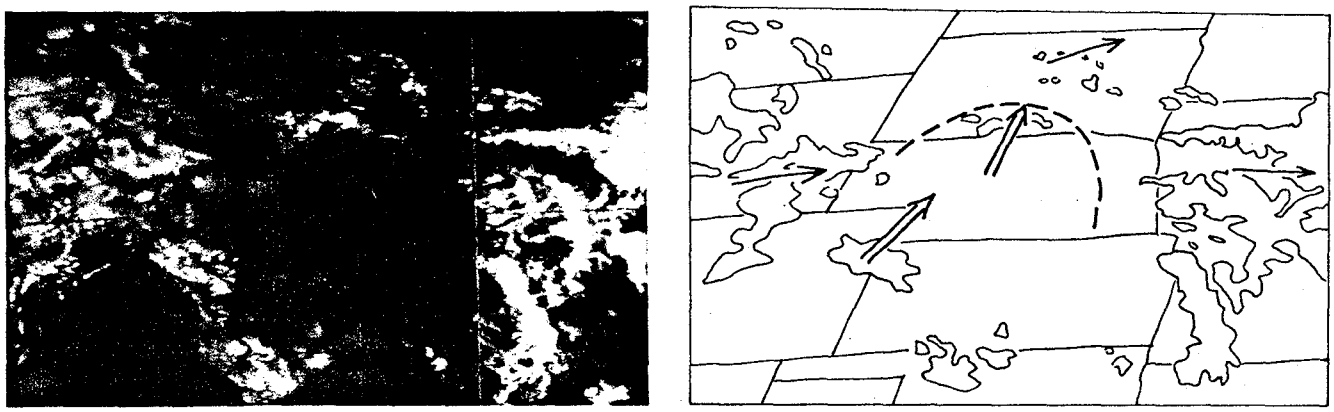
The satellite photographs of 5 May 1983 show an interesting case of formation of several cumulonimbus in a ring on the mesoscale that looks similar to the thunderstorm outflow boundary, but in this case there were no thunderstorms in the area to serve as a source for this outflow. In the multitude of clouds on this day, no still photograph reveals clearly a semicircle of cumulonimbus as the looping of hourly photographs does. The phenomenon in question is illustrated in Fig. 6, where three selected photographs are shown, at intervals of 3 h. These photographs are taken in visible light, however, most of analysis was done in hourly infrared photographs that are not shown here. For the purpose of reporting, the photos in the visible spectrum show a better contrast.

All clouds in the first frame (1500 GMT) move with the wind in the middle and higher troposphere, that is zonally, as shown by arrows in the sketches that accompany the photos. The region covered by the LLJ (as seen in Fig. 1) covers a cloud-free area over Kansas and Nebraska. By 1800 GMT, second frame in Fig. 6, the northern extent of the LLJ can be identified in the cloud photograph by a number of clouds that move toward the north. These clouds are quite noticeable when the loops of hourly photographs are examined. In sketches in Fig. 6 the northward-moving clouds are shown within a dashed arc. In the sketches, the motion of the air within the LLJ is shown by double arrows. The circularly arranged clouds along the arc are strikingly similar to the arches of new convection cells that occur in the outflow from thunderstorms (e.g. Maddox *et al.*, 1981). In the last frame of 2100 GMT, the arc with convective clouds penetrated deeper to South Dakota and to Iowa. The western end of the arc in Colorado already assumed the general zonal flow, so that this part of the arc started retreating into the area occupied by the southerly LLJ.

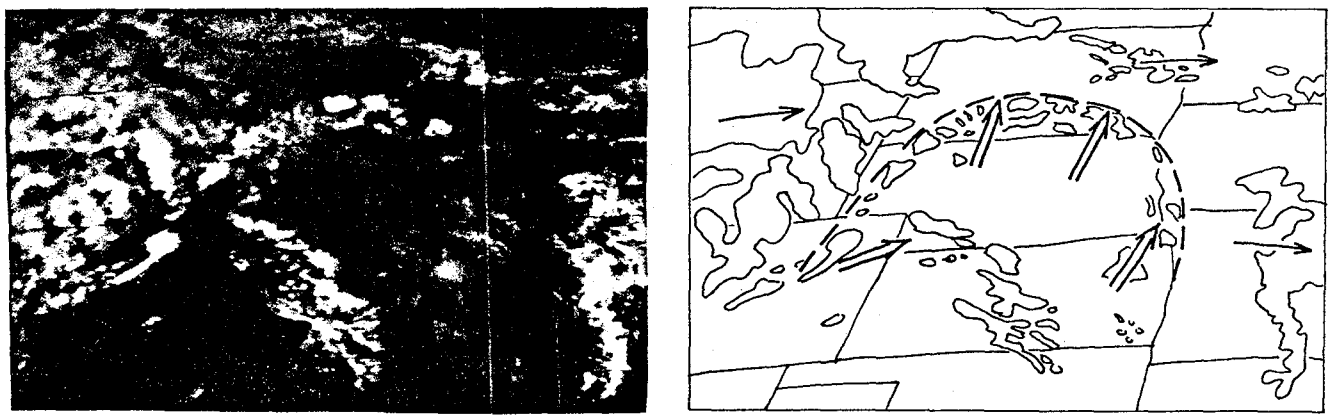
The progression of the arc with clouds can be interpretted as an appearance of a shallow warm front where all lifting results in convection since the atmosphere is conditionally unstable. The clouds are in the middle-latitude air in a raised trade wind inversion with which the LLJ is



15 GMT



18 GMT



21 GMT

Fig.6. Satellite photographs of the Great Plains in visible light, at intervals of 3 h, on 5 May 1983, and sketches with interpretation of the photographs side-by-side.

associated.

6. The case of 2-3 December 1983

The development of the LLJ on 2 December is shown in Fig. 7. This figure contains the LLJ observations and the positions of the polar front at 850 mb and the upper tropospheric jet streams. It may be seen again that the LLJ develops in the warm sector, between two loops of the polar front. This warm sector was prominent at 850 and 700 mb, but was almost absent in the temperature field at the surface. The LLJ in Fig. 7 is shown at 0000 GMT, which is the local evening hour. This makes it difficult to interpret this jet as an inertia wave that started with decoupling of stable layers at night in the boundary layer. However, it is unfortunate that the soundings of the morning hour of 1200 GMT were not available on the McIDAS system during my short stay at MSFC.

There were several lows and highs in the area, but the LLJ occured nearest to the low that developed in northeast Texas on the 3 December. There was a prominent pressure fall over Texas and Louisiana, as shown in Fig. 8. In the same figure the location of the center of the low of 3 December is shown at two times: at 0600 and 1800 GMT. The LLJ retreated during the 2 December until it was confined to the area southand east of theplace where the low in Texas developed on 3 December. The development of the LLJ was again related to the isallobaric field (Fig. 8), but the northern part of the chart was under an isallobaric field in which the gradient of height tendency ($-\nabla(\partial z/\partial t)$, with minus) pointed south. In that part the LLJ and any southerly flow did not develop. The LLJ appeared only in the southern part of the region where the gradient $-\nabla(\partial z/\partial t)$ pointed north.

The isotherms at the earth's surface, which are not commonly drawn in routine weather analysis, are shown in Fig. 9 for 0000 GMT 2 December 1983. The distribution of temperature shows a remarkably zonal form, with the only larger exception of a cold trough in the lee of the Rocky mountains from West Texas to western Nebraska. A hardly noticeable thermal ridge can be detected over northeastern Texas and Oklahoma. This ridge moved to the Mississippi river between Missouri and Illinois in 24 h, with a small increase in amplitude.

The development of the LLJ during 2 December 1983 confirms the synoptic-scale character of this jet. Also, its development further southeast than the usual position in West Oklahoma gives indication that the sloping terrain of the high planes is not a necessary element. It is, however, possible that the effects of sloping terrain contribute to the dynamics of the LLJ (McNider and Pielke, 1981). The climatological frequency of the LLJ over high plains can be interpretted as a consequence of the sloping terrain (Bonner, 1968). On the other hand, the frequency of cyclogenesis in that area ("Colorado cyclone") is a more likely direct cause

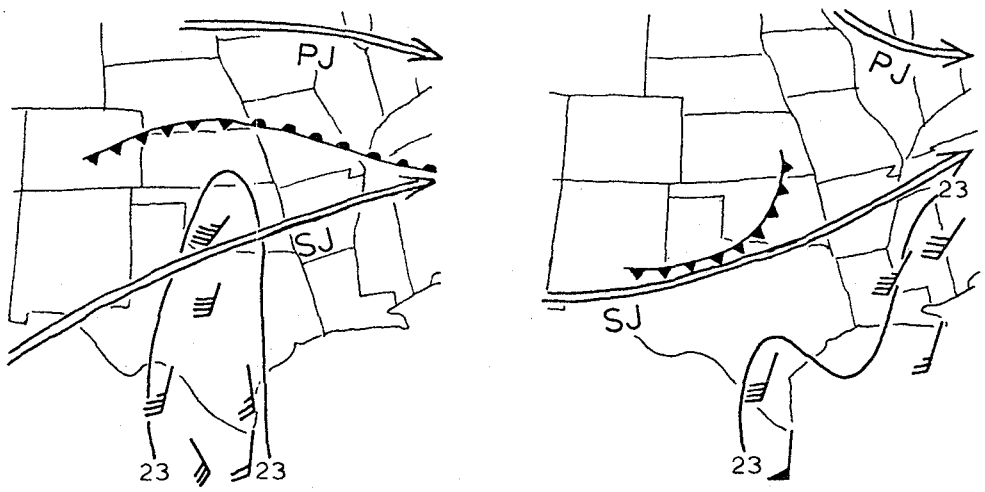


Fig. 7. LLJ on 2 December 1983. Wind speed in knots at the LLJ level. The isotach of 23 kt circles the LLJ, other data are not plotted. The front at 850 mb is shown by the sawtooth line. The polar jet (PJ) and the subtropical jet (SJ) are shown by the double arrows.

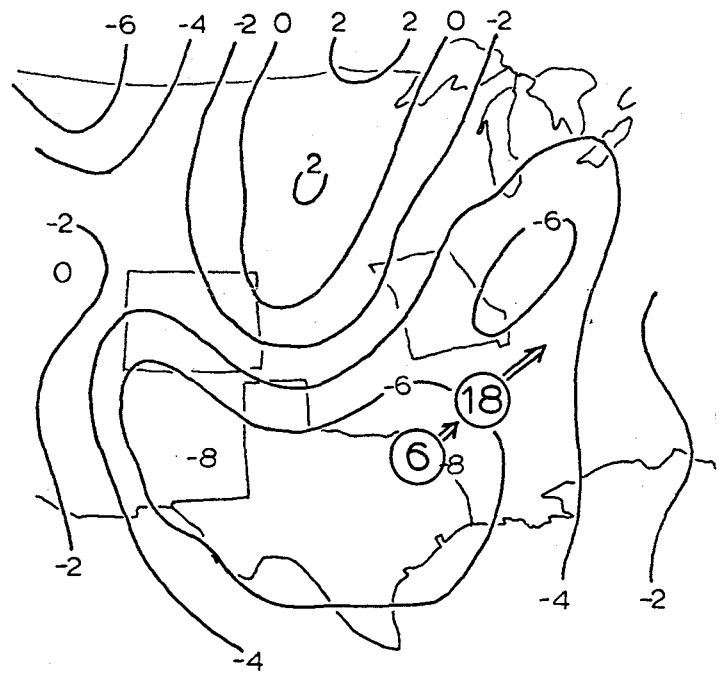


Fig. 8. 24-h height change between 0000 and 2400 GMT on 2 December 1983 at 850 mb, in gpm. The positions of a surface low in Texas and Arkansas at 0600 and 1800 GMT on 3 December are shown by circled numbers.

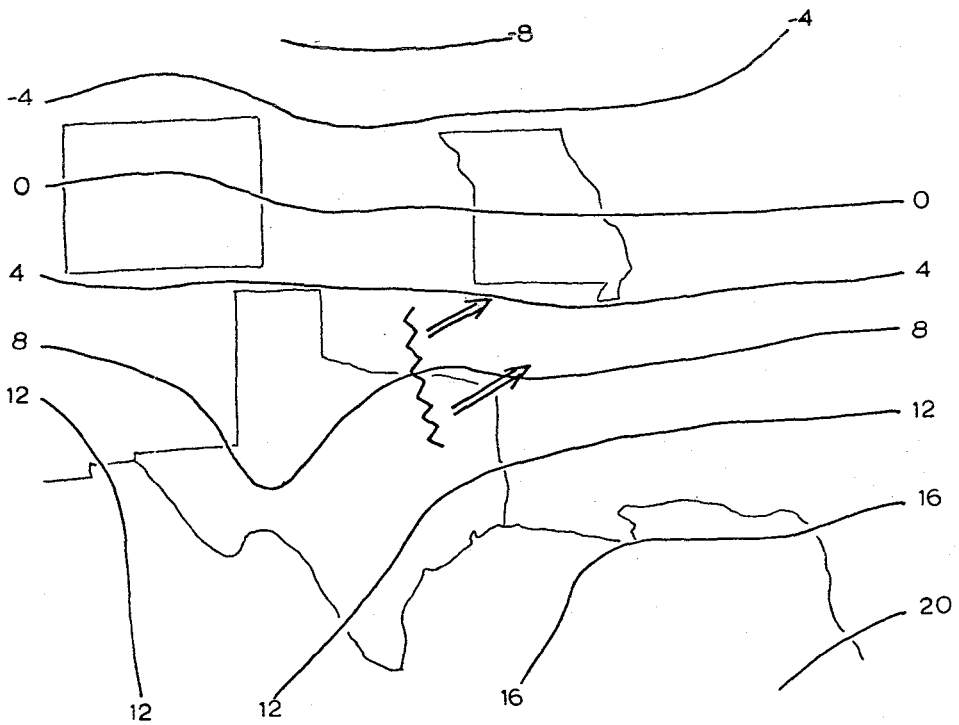


Fig.9. Isotherms at the earth's surface, 1200 GMT 5 May 1983.

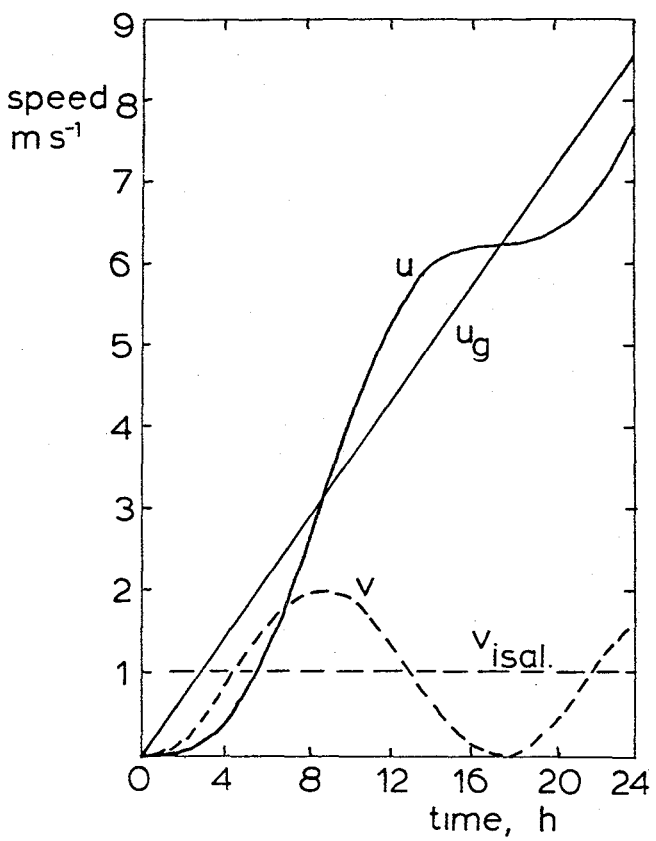


Fig. 10. Theoretical variation of wind and isallobaric wind. The wind components are u, v. The isallobaric wind is $v_{isal.}$ and u_g is the geostrophic wind.

of frequent LLJ.

7. The LLJ as an inertia oscillation

Several theories of LLJ show that a sudden change of forces may cause the LLJ. These theories assume that the LLJ is formed by nocturnal decoupling of layers (Blackadar, 1957), by heating of the sloping boundary layer (McNider and Pielke, 1981) or by onset of cyclogenesis (Djurić, 1981). The LLJ under these circumstances should show inertia oscillations. In order to clarify this possibility, here an example is computed that demonstrates some of the basic properties of acceleration of the air in a stable layer.

If we consider a layer of air where the horizontal variation is comparatively small and the influence of friction is negligible, the air movement may be approximated by the following equations of motion:

$$\frac{du}{dt} = f v \qquad (1)$$

$$\frac{dv}{dt} = f v_g - f u$$

Here u and v are the x and y components of wind, v_g is the component of geostrophic wind under consideration. The other component of the geostrophic wind is set equal to zero by the turning of the coordinate system. The Coriolis parameter f is considered constant for this discussion. Therefore, the case of a steadily increasing pressure field is considered, defined by

$$\frac{du_g}{dt} = \text{const.} = 10^{-4}\ \text{m s}^{-2}$$

The general solution of (1) is

$$u = u_0 \cos ft + v_0 \sin ft + u_{go} + t\,\frac{du_g}{dt}$$

$$v = -u_0 \sin ft + v_0 \cos ft + \frac{1}{f}\,\frac{du_g}{dt}$$

where u_0 and v_0 are the arbitrary constants of integration.

We are interested in the simple particular solution that shows the initial state of rest, when a pressure tendency suddenly appears. Then the particular solution is

$$u = u_{go} + \frac{du_g}{dt}\left(t - \frac{1}{f}\sin ft\right)$$

$$v = -\frac{du_g}{dt}(1 - \cos ft)$$

The trigonometric terms show the inertia oscillations, while the other terms

represent the part analogous to the normal mode:

$$u_n = u_{go} + t\frac{du_g}{dt} = u_g$$

$$v_n = -\frac{du_g}{dt}^0 = v_{isal.} = \text{const.}$$

This shows that the isallobaric part of wind $v_{isal.}$ is constant and it suddenly appears at the time of onset of the pressure tendency.

This theoretical situation is represented in Fig. 10. The slanted straight line is the increasing geostrophic wind and the dashed straight line is the isallobaric wind. The motion of air oscillates around the "normal", nonoscillating part (u_g, $v_{isal.}$).

The implications of this theoretical example are that similar oscillations occur in nature, albeit more complicated. We cannot know whether the observation comes from the peak or valley in the oscillations. However, it can be seen that the ageostrophic component of flow is never too large. It is possible that the ageostrophic component in the atmosphere appears larger if the value of du_g/dt is larger.

The flow in this situation is accelerating, even if it is not much different from the geostrophic balance. This may explain the origination of the LLJ in the synoptic-scale isallobaric field. A decline of the LLJ is not illustrated here. It is assumed that an opposite isallobaric gradient or friction in absence of pressure force will influence the LLJ to weaken and eventually dissappear.

8. Modeling studies

The LAMPS (Limited Area Mesoscale Prediction System) model is one of the most advanced mesoscale models in the world. It has been demonstrated that it can succesfully simulate a number of weather processes (e.g. Kalb and Perkey, 1985). Recently, this model was made operational at the Atmospheric Sciences Division of NASA/MSFC, using the Perkin-Elmer computer. The usefulness of this model for simulation of the LLJ can be demonstated by the wind profile taken from an experiment which was not specially tailored for the study of the LLJ (Fig. 11). This wind profile is from a grid point close to Oklahoma City where the LLJ developed. For comparison, two observed wind profiles with LLJ are shown in the same figure. Those two are not from the same weather situation, but from the similar situation in the case of 5 May 1983, described above. Similarity of the wind profile in the lower layers is apparent. This shows that the LAMPS model is suitable for simulation of the LLJ. There are differences between

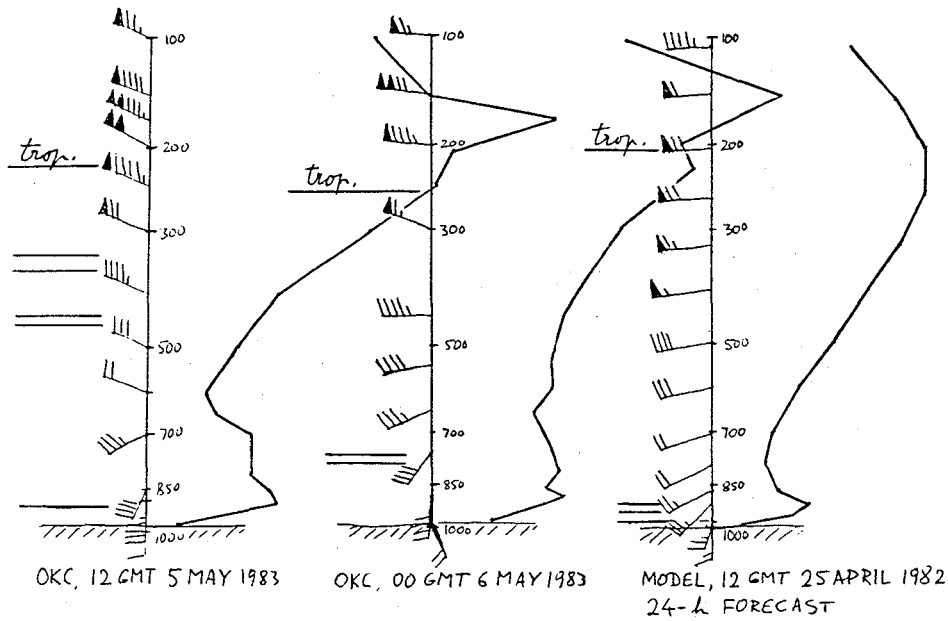


Fig. 11. Wind profiles from the LAMPS model and observed at OKC 0000 GMT 5 May 1983.

the observed and simulated wind profiles in upper layers, these can be attributed to the difference in the structure of the atmosphere: the case of 5 May 1983 had the subtropical jet in the upper troposphere, and the LAMPS case was under the polar jet.

After discussing several possibilities for further work on the problem of LLJ development with Dr. M. Kalb of MSFC, we decided to run several parallel experiments with the model in which a schematic weather situation will be used as the initial condition. In this way it will become clear which factors influence the development of the LLJ, since all the "irregularities" of observations will not be present. We designed a model initial state that contains the baroclinic zone of the polar front, with waves such that the cyclogenetic region is situated in the middle of the computational domain. This scheme is illustrated in Figs. 12 and 13. In Fig. 12 there is a plane view of the computational domain. Between the two sinusoidas labeled FN and FS is the frontal zone, 600 km wide. The amplitude of the frontal zone is 1200 km. The meridional vertical section through the frontal zone is shown in Fig. 13, with several isotherms and discontinuity surfaces at the front and tropopause. Cyclogenesis is expected in the model in the region of positive vorticity advection in the middle of the computational domain. One other experiment will be computed with addition of an inversion south of the polar front, simulating the existence of the trade wind inversion which is typical for the warm sector of the cyclones. This inversion is typical for the generation and development of the LLJ. Experiments of this type will clearly single out the crucial factors in the development of the LLJ.

9. Conclusions

The research done during my short stay at MSFC yielded several results that show the possibilities for the study of the LLJ, as are

(a) The diagnostic studies can be done efficiently using the McIDAS facility for handling weather data. There are still some difficulties in such work, e.g. unavailability of some data, but as the time progresses, such difficulties will be less present.

(b) Numerical modeling can be done at MSFC with great efficiency, mainly due to the available model and adequate computer facility.

As the examples in previous sections show, further work on the LLJ can and will be continued, since this task is well defined, useful for future application and suitable for the facilities at MSFC.

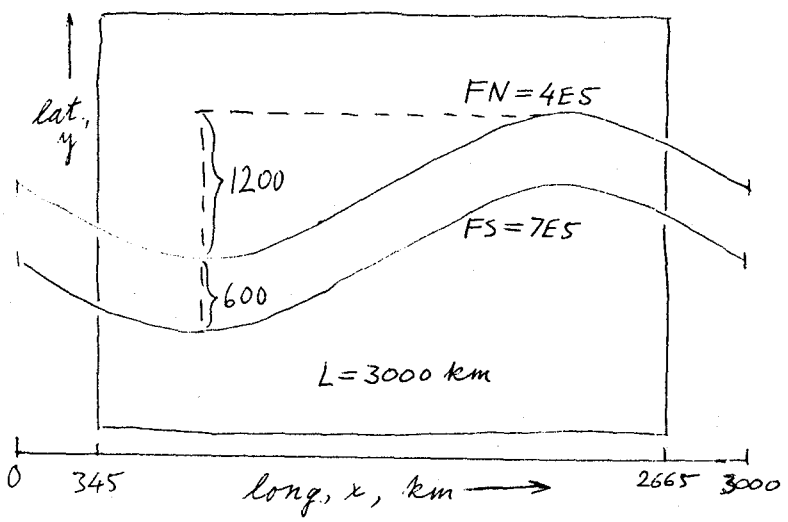


Fig. 12. Proposed initial conditions for the experiment with LLJ, plane view. The frontal zone is between the two sinusoidas.

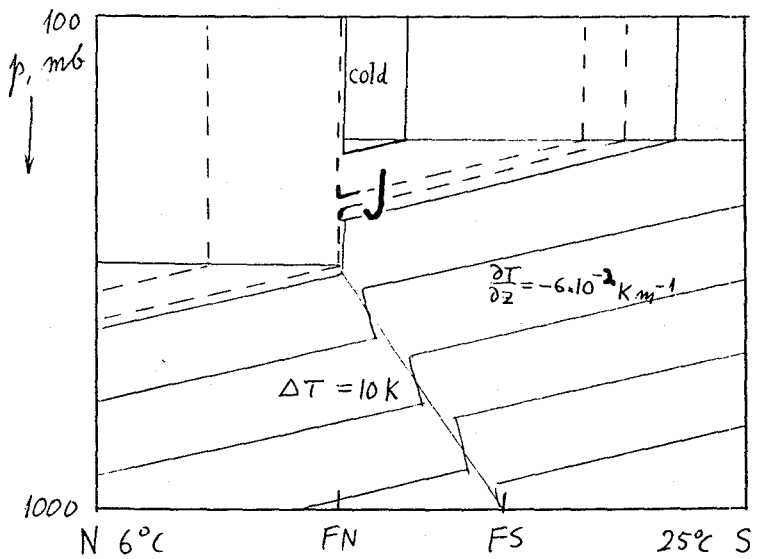


Fig. 13. Proposed initial conditions for the experiment with LLJ, vertical section. Isotherms, front and tropopause are shown.

REFERENCES

Bonner, W.D., 1968: Climatology of the low-level jet. *Mon. Wea. Rev.*, **96**, 833-850.

Blackadar, A.K., 1957: Boundary layer wind maxima and their significance for the growth of the nocturnal inversions. *Bull. Amer. Meteor. Soc.*, **38**, 283-290.

Djurić, D., 1981: A numerical model of the formation and evolution of a low-level jet. *Mon. Wea. Rev.*, **109**, 384-390.

Djurić, D., and M.S. Damiani, Jr., 1980: On the formation of the low-level jet over Texas. *Mon. Wea. Rev.*, **108**, 1854-1865.

Djurić, D., and D.S. Ladwig, 1983: Southerly low-level jet in the winter cyclones of the southwestern Great Plains. *Mon. Wea. Rev.* **111**,2275-2281.

Doswell, C.A.,III, 1980: Synoptic-scale environments associated with High Plains severe thunderstorms. *Bull. Amer. Meteorol. Soc.*, **61**, 1388-1400.

Kalb, M., and D.J. Perkey, 1985: Experiments with LAMPS model. *Proceedings, Conference on Numerical Weather Prediction, Toronto.* Amer. Meteor. Soc.

Maddox, R.A., D.J. Perkey and J.M. Fritch, 1981: Evolution of upper tropospheric features during the development of a mesoscale convective complex. *J. Atmos. Sci.*, **38**, 1664-1674.

McNider, R., and R.A. Pielke, 1981: Diurnal boundary-layer development over sloping Terrain. *J. Atmos. Sci.*, **38**, 2198-2212.

Palmen, E., and C. W. Newton, 1969: *Atmospheric Circulation Systems*, Acad. Press, 603 pp.

Shen, T.-C., and D. Djurić, 1981: Formation of the low-level jet under an inversion. *Preprints, Fifth Conference on Numerical Weather Prediction*, Monterey, Amer. Meteorol. Soc., 301-302.

Uccelini, L.W., and D. R. Johnson, 1979: The coupling of upper and lower tropospheric jet streaks and implications for the development of severe convective storms. *Mon. Wea. Rev.*, **107**, 682-703.

Wippermann, F., 1973: *The planetary boundary-layer of the atmosphere.* Deutsch. Wetterdienst, 346 pp.

1985
NASA/ASEE SUMMER FACULTY RESEARCH FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

PRELIMINARY DESIGN METHODS FOR FIBER REINFORCED COMPOSITE STRUCTURES EMPLOYING A PERSONAL COMPUTER

Prepared by:	Charles N. Eastlake, P.E.
Academic Rank:	Associate Professor
University and Department:	Embry-Riddle Aeronautical Univ. Aeronautical Engineering Dept.
NASA/MSFC: Division: Branch:	Structures and Propulsion Lab Structures Division Structural Development Branch EP-13
MSFC Counterpart:	Erich E. Engler
Date:	August 2, 1985
Contract No.:	NGT 01-008-021 The University of Alabama in Huntsville

PRELIMINARY DESIGN METHODS FOR FIBER REINFORCED COMPOSITE STRUCTURES EMPLOYING A PERSONAL COMPUTER

BY

Charles N. Eastlake
Associate Professor of Aeronautical Engineering
Embry-Riddle Aeronautical University
Daytona Beach, Florida

ABSTRACT

The objective of this project was to develop a user-friendly interactive computer program to be used as an analytical tool by structural designers. Its intent was to do preliminary, approximate stress analysis to help select or verify sizing choices for composite structural members. The approach to the project was to provide a subroutine which uses classical lamination theory to predict an effective elastic modulus for a laminate of arbitrary material and ply orientation. This effective elastic modulus can then be used in a family of other subroutines which employ the familiar basic structural analysis methods for isotropic materials. This method is simple and convenient to use but only approximate, as is appropriate for a preliminary design tool which will be subsequently verified by more sophisticated analysis.

Additional subroutines have been provided to calculate laminate coefficient of thermal expansion and to calculate ply-by-ply strains within a laminate.

1. INTRODUCTION

The NASA Marshall Space Flight Center has both research and hardware development within its mission. The responsibility for developing a variety of items of space flight hardware brings with it the necessity of doing structural design in-house. In the beginning stages of design concept definition it is helpful to have quick and convenient approximate stress analysis tools available to the designers so that a wide variety of structural configurations can be sketched out and checked for feasibility. At this stage of the design process the stress/strain analysis does not need to be particularly accurate because any configurations tentatively defined as feasible will later be analyzed in detail by stress analysis specialists. Besides, the sophisticated finite element methods such as NASTRAN which are industry analysis standards are simply too time-consuming and complex for practical use in preliminary design. The emphasis in preliminary design must be on fast, user-friendly methods so that rough but technically sound evaluation of a broad variety of conceptual designs can be accomplished. The development of a computer program to function as such a tool was the goal of this project.

2. OBJECTIVES: The broad concept of the computer program COMPSIZE

To maximize user convenience the desire of the sponsoring branch was from the very beginning to employ a personal computer located in the designers' offices. The preliminary designer must usually make frequent reference to drawings, manuals, handbooks, etc. as he progresses through his calculations. If the computer is located remote from his "home" office the author can attest from experience that the designer is almost certain to waste frustrating time going back after needed data which is not included in the bundle of information that he has carried to the computer area with him. To meet this goal, both a Digital Equipment Corp. DEC 350 and an IBM PC were possible candidate hardware. The IBM PC was chosen because a greater number of machines appear to be likely to be available in the user's facility for the near term future.

Several aspects of the practicality of the project were not certain at the outset and were to be clarified during the writing of this program.

1. Does a personal computer have adequate memory?
2. Is BASIC language adaptable enough and fast enough to handle problems of the type required?
3. Is it possible to build in sufficient user-friendliness to encourage widespread use of the program by the design personnel?

The final conclusions to these questions are presented in the Conclusions and Recommendations section. The rationale for the questions and some of the pertinent information are discussed here.

The IBM PC used for this project has dual 5.25" floppy disk drives employing double side, double density disks each having a storage capacity of 366 K bytes. The processor has a random access memory (RAM) of 512 K. This seemed likely to be adequate and turned out to be so. The disk storage space required to save the main program and six subprograms is approximately 40 K. Less than 1 K additional RAM is required to run the most memory-consuming portion of the program due to the fact that the number of variables used in the program is not particularly large. IBM Advanced BASIC requires 27 K, so the bare bones 64 K PC may not be adequate but 128 K versions should run the program without difficulty.

There was some speculation about the suitability of the BASIC language, a common discussion among the engineering community. BASIC was severely limited in capability in its early days, dating back to the mid-sixties. However, that

limitation does not appear to be a problem any longer. The IBM Advanced BASIC used in this project has 152 program commands, 15 built-in mathematical operations, user definable mathematical functions, color graphics, and sound production ability. It does not have for some reason the matrix manipulation functions frequently found in BASIC, which would have been useful for the composite laminate analysis portion of this program. At any rate the author did not feel any serious constraints due to the language employed.

User-friendliness goals are centered around the sad but true fact that many users will not read a user's manual until they get into some mess that they can't find their way out of. In addition, many users are unfamiliar with the operation of PC hardware. With that in mind, this program was written in such a way that if the user knows how to turn on the computer and initiate the program he should be able to solve his problem with no further guesswork even if he has not read the manual. The program uses a main menu format listing 15 major problem types which can be selected. Several of the major problem selections have sub-menus of up to 5 further specialized problems. In each problem subroutine all units are specified for both input and output. At the conclusion of each problem the user is asked whether he wants a printed copy of the results, then whether he wants another problem of the same type or to go back to the main menu. To speed operations, the user can answer "no" to any yes/no input requested by simply pressing the carriage return. The video game fascination is assuaged by a color garphics NASA logo during the initial sign-on. The author feels that the user-friendliness has worked out well and will be of value as well for classroom use of the program at his home institution. The programming of these prompts, messages, and fault-tolerant branches did, however, take possibly twice as much storage space and programming time as the solution of the equations.

3. A BRIEF REVIEW OF COMPOSITE ANALYSIS METHODOLOGY

Finite element methods (FEM) have become firmly established as the standard structural analysis method in the aerospace industry. The most widely known is NASTRAN, a contraction for NASA Structural Analysis. FEM methods when correctly applied are relatively reliable and accurate. However, the creation of the geometric model of the structure to be analyzed is very time consuming and something of an art for complex structural shapes. Automatic generation of FEM models directly from data bases assembled during the creation of drawings on Computer-aided Design (CAD) systems is close to operational availability and will broaden the use of FEM. But for now the tedium of model generation makes FEM analysis impractical for preliminary design. Even after automatic model generation is a reality the method will still require a significant amount of time expenditure in creation of a fairly detailed CAD drawing of the structure being studied. So the hand-drawn sketch method with quick-look calculations will probably endure forever in the early concept definition stages of structural preliminary design.

It is necessary even as part of FEM analysis to specify a layer-by-layer description of the stackup of plies in a laminated composite part. In its most accurate form this composite analysis technique calculates the mechanical strain separately in each ply of the composite part, which is typically .005 to .010 inches thick. The analysis is done by matrix algebra methods which the computer can do very quickly, but evaluation of the data can be a slow and painstaking process. The description of COMPSIZE main menu problem 2 in Section 4.2 gives a more detailed picture of the geometry and material description conventions used for that process.

A much faster though less accurate method widely used for preliminary design (Ref. 3,4,and 8) is to use the stackup of plies to calculate an effective stiffness of the laminated material and then analyze it using the much simpler equations which describe homogeneous and isotropic materials. This calculation of effective stiffness can be done using classical lamination theory which is employed in COMPSIZE problems 1, 2, and 3. Even this can be quickened considerably by solving for stiffness of a large variety of physical orientations of ply stackups and plotting the results on an easy-to-use figure known as a carpet plot. However, the designer must have the carpet plots on hand for the material combinations which he is considering using. A simple mathematical representation of a carpet plot (a single equation) developed by the author during this project is

presented in COMPSIZE problem 1 sub-menu 1. Another simple approximation in lieu of a carpet plot, from Reference 3, is presented in problem 1 sub-menu 3.

Estimation of coefficient of thermal expansion (CTE) of laminated composites is of great significance in space structures because of large size parts and large temperature differences between sunlight and shade on orbit. CTE can be estimated using an extension of stiffness estimation methods which is presented as problem 3.

4. SPECIFIC DESCRIPTION OF THE "COMPSIZE" PROGRAM

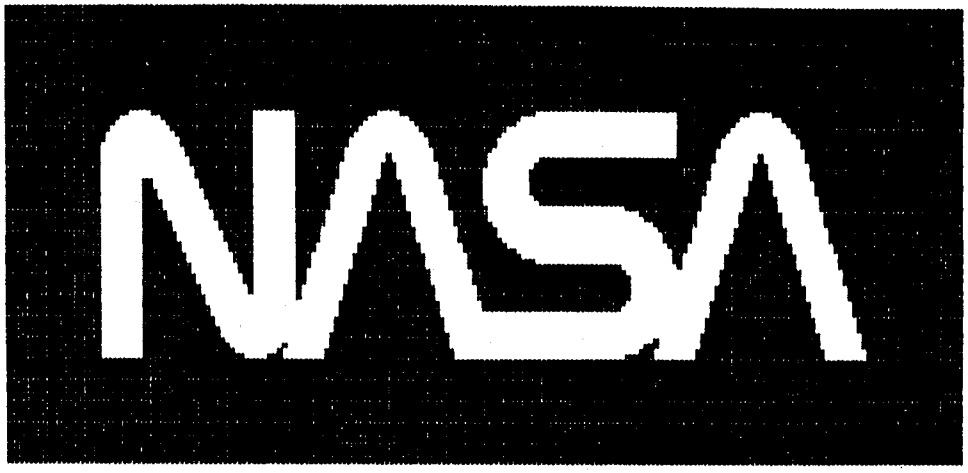
It would be easiest to describe COMPSIZE by referring to a listing of the program. However, the listing would require about 30 pages, which is well beyond the acceptable limit for this report. Instead, replicas of the first several monitor screens will be presented to give the flavor of the program.

A detailed User's Manual was written as the culmination of the project. It includes a listing, description of the calculation procedures, and sample problem for each problem (subroutine). Readers seriously interested in using the program should contact the sponsoring branch listed on the title page to request copies of the program disk and manual.

Figure 1 shows the first screen to appear when the program is initiated. It displays the NASA logo and program name and asks if the user wants to see the program introduction. This introductory description, shown in Figure 2, is similar to the abstract of this report. The next screen, shown in Figure 3, lists the default values used in the program for safety factor. It also lists stiffness and allowable stress for a typical intermediate strength graphite/epoxy laminate in case the user does not have other data at hand and simply wants to input something to get ballpark numbers. Default material properties are only written into the actual calculation process for problem 1 sub-menu 1. Creation of files of default or selectable material properties would be a helpful next task in the continued development of this program.

Figure 4 shows the main menu, which appears next. Once a user is familiar with the program he can go directly to this point by entering RUN 298 instead of simply RUN. There are 15 problems listed, seven of which have sub-menus for further specialized problem selection. Each main menu item is described in detail in a separate section of the user's manual. Because this report is more descriptive in nature, Figure 5 shows the screen exactly as it would appear to a user solving problem 1 sub-menu 3. The choice is an approximate stiffness calculation using a set of equations from Reference 3. The exact layup of ply orientations is not required for this problem, only the percent of plies to be laid up at each angular orientation. As is common industry design practice 0,90, or +/-45 orientation is presumed. The stiffness properties of the individual layer material with all layers assumed to be the same are input and the output is the stiffness properties of the entire laminate.

The program asks if another problem of the same type is



COMPSIZE
7/19/85

COMPOSITE STRUCTURE PRELIMINARY SIZING

WANT AN INTRODUCTION? (Y/N) ? ■

FIGURE 1

COMPSIZE

This program was written by Prof. Chuck Eastlake of Embry-Riddle Aeronautical University under the sponsorship of the 1985 NASA/ASEE Summer Faculty Fellowship Program. Work was performed at the NASA Marshall Space Flight Center, Structures and Propulsion Laboratory, Structural Development Branch, EP-13.

It is intended to be a user-friendly, interactive analytical tool for preliminary,approximate stress analysis by structural designers to select or verify sizing choices for composite structural members. Homogeneous materials can be analyzed as simplified cases. A flat laminate analysis routine is included for analysis of arbitrary layups. However, it should be sufficient for most users to presume a quasi-isotropic layup and use the familiar basic structural analysis methods for isotropic materials, after estimating an appropriate elastic modulus. All the equations used in this program assume elastic deformation. Intermediate strength graphite/epoxy is used as the default material selection. Example problems are included in the program manual.

PRESS ANY KEY TO CONTINUE ?

FIGURE 2

```
DEFAULT VALUES USED IN THE PROGRAM:
Safety factor=2 for pressure vessels.
Safety factor=1.4 for all other problems.
Material properties for quasi-isotropic laminate of intermediate
strength graphite/epoxy:
    E=7.2E6 psi,  G=2.8E6 psi,  Poissons ratio=.32
            ultimate tensile stress=54.6 ksi
        ultimate compressive stress=52 ksi
     ultimate in-plane shear stress=34 ksi
      ultimate net tension in joint=27 ksi
    ultimate bearing stress (e/D=2)=65 ksi

To speed up running of the program, you may press the RETURN/ENTER key to answer
 NO for any question requiring a (Y/N) answer.

PRESS ANY KEY TO CONTINUE ?
```

FIGURE 3

```
MAIN MENU: PROBLEM TYPES
     1. COMPOSITE MATERIAL PROPERTY ESTIMATION
     2. LAMINATE ANALYSIS
     3. LAMINATE COEFFICIENT OF THERMAL EXPANSION
     4. AXIAL TENSION OR COMPRESSION (NO BUCKLING)
     5. DIRECT SHEAR OR PUNCH-OUT
     6. SYMMETRIC BENDING
     7. COMBINED AXIAL AND BENDING
     8. SHEAR INDUCED BY BENDING~
     9. TORSIONAL SHEAR
    10. COLUMN BUCKLING (EULER)
    11. PRESSURE VESSELS
    12. STRESS INTERACTIONS FOR COMBINED LOADS
    13. FITTING STRESSES
    14. SECTION AREA AND MOMENT OF INERTIA
    15. SANDWICH PANEL ANALYSIS
    20. QUIT
TYPE THE NUMBER YOU WANT ?
```

FIGURE 4

```
COMPOSITE MATERIAL PROPERTY ESTIMATION

MATERIAL PROPERTY ESTIMATION SUB-MENU:
    1. CARPET PLOT READ-OFF
    2. CALCULATED FROM A MATRIX
    3. APPROXIMATE SOLUTION BASED ON LAYER ORIENTATION

ENTER NUMBER FOR THE PROCEDURE YOU WANT ? 3

ENTER PERCENT OF LAYERS TO BE SET AT 0 DEGREES ? 25
ENTER PERCENT OF LAYERS TO BE SET AT 90 DEGREES ? 25
ENTER PERCENT OF LAYERS TO BE SET AT +/-45 DEGREES ? 50

ENTER MAJOR ELASTIC MODULUS, E11 (MSI) ? 22
ENTER MINOR ELASTIC MODULUS, E22 (MSI) ? 1.5
ENTER SHEAR MODULUS, G12 (MSI) ? .52
ENTER MAJOR POISSONS RATIO, V12 ? .25

Ex (MSI)= 8.333333          Ey (MSI)= 8.333333    Gxy (MSI)
Vxy= .3197279               Vyx= .3197279          = 3.0825

ANOTHER CALCULATION OF THIS TYPE? Y/N ? N

MAIN MENU: PROBLEM TYPES
    1. COMPOSITE MATERIAL PROPERTY ESTIMATION
    2. LAMINATE ANALYSIS
    3. LAMINATE COEFFICIENT OF THERMAL EXPANSION
    4. AXIAL TENSION OR COMPRESSION (NO BUCKLING)
    5. DIRECT SHEAR OR PUNCH-OUT
    6. SYMMETRIC BENDING
    7. COMBINED AXIAL AND BENDING
    8. SHEAR INDUCED BY BENDING
    9. TORSIONAL SHEAR
   10. COLUMN BUCKLING (EULER)
   11. PRESSURE VESSELS
   12. STRESS INTERACTIONS FOR COMBINED LOADS
   13. FITTING STRESSES
   14. SECTION AREA AND MOMENT OF INERTIA
   15. SANDWICH PANEL ANALYSIS
   20. QUIT
TYPE THE NUMBER YOU WANT ? 20

THE END. HOPE YOU GOT GOOD NUMBERS!

Ok
```

FIGURE 5

desired. If the response is Y the program returns to the beginning of the same subroutine. If the response is anything else it returns to the main menu, from which the user may select another problem type or terminate the program.

Problems 1,2 and 3 are specifically devoted to design with composites. Problem 11 sub-menu 1 calculates the Tsai-Hill failure criterion which is specific to composites. Problem 15 is not necessarily a composite design tool, but sandwich structures are most frequently composite. Problem 15 is also the only subroutine which has an optimization process for minimum weight written in as a sub-menu choice. That is because this optimization is represented by a fairly simple equation having a closed form solution.

The remaining 10 problems are standard stress analysis methods for homogeneous and isotropic materials. They can be found in many basic texts on strength of materials like References 5 and 6. These should be familiar to almost any engineer or engineering student and thus do not warrant extensive explanation.

The first 3 subroutines are the heart of the composite application of this program and will be described in modest detail. Again, a more detailed explanation can be found in the User's Manual.

4.1 SUBROUTINE 1 Composite Material Property Estimation

There are three methods known to the author for estimation of stiffness (elasticity) properties of laminated composites, represented by the three sub-menu options available in this subroutine. The calculated results of the three methods are compared in Table 1. The method which should be most accurate is that involving the calculation of the terms in the so-called A-matrix, so this is used as the standard of comparison. It is used in the subroutine named AMATRIX which is chained into the program by selection of problem 1 sub-menu 2.

The equations presented by Lubin (Ref. 3) are the simplest and are in reasonably good agreement except for the symmetric angle ply case. For that case, which unfortunately is typical of structures designed primarily for torsional loads, the agreement is quite bad. The difference is over 50 percent and is due to the fact that the equations inherently presume that the modulus of elasticity for a 45 degree angle ply laminate is identical to the transverse direction modulus of the ply material. For some materials this is a reasonable assumption but for many others it is not. This set of equations is available as problem 1 sub-menu 3.

The author felt is was appropriate because of this error to attempt to formulate another more accurate method which would still be simple enough to use on programmable calculators. The method which resulted is problem 1 sub-menu 1. This subroutine named CARPLOT solves a single equation, albeit a moderately long one, which represents a math model of a carpet plot. The equation, shown below, has a closed form solution. It requires no iterations or matrix algebra and could thus be solved on nearly any programmable hand calculator. Figure 6 and 7 show the accuracy achieved.

$$E_x = E_{11}P_0 + E_{22}(1-P_0) + \left(\frac{4QG_{12}}{Q+G_{12}} - E_{22}\right)P_{45} + .06(1-P_0)E_{11}\,\text{SIN}\left[\pi\left(\frac{P_{45}}{1.001-P_0}\right)\right]$$

where $Q = \frac{1}{4}\left[\left(\frac{E_{11}}{E_{11}-\nu_{12}^2 E_{22}}\right)(E_{11}+E_{22}+2\nu_{12}E_{22})\right]$

P_0 =percent of 0 degree plies
P_{45}=percent of +/-45 degree plies
E_{11}, E_{22}, G_{12}, ν_{12}=ply material properties

LAMINATE GEOMETRY	PERCENT 0/90/±45	"AMATRIX" E_x (MSI)	"CARPLOT" E_x / %VAR.	REF. 3 E_x / %VAR.
QUASI-ISOTROPIC $[0/45/-45/90]_S$	25/25/50	7.362	7.284/-1.06%	7.067/-4.17%
SYM. ANGLE PLY $[45/-45/45/-45]_S$	0/0/100	3.362	3.366/-0.12%	1.6/-52.4%
CROSS PLY $[0/90/0/90]_S$	50/50/0	9.889	9.80/-0.90%	9.80/-0.90%
GENERAL $[0_8/90/45/-45]_S$	73/9/18	14.18	14.143/-0.26%	14.064/-0.82%

STIFFNESS ESTIMATES
GRAPHITE/EPOXY

TABLE 1

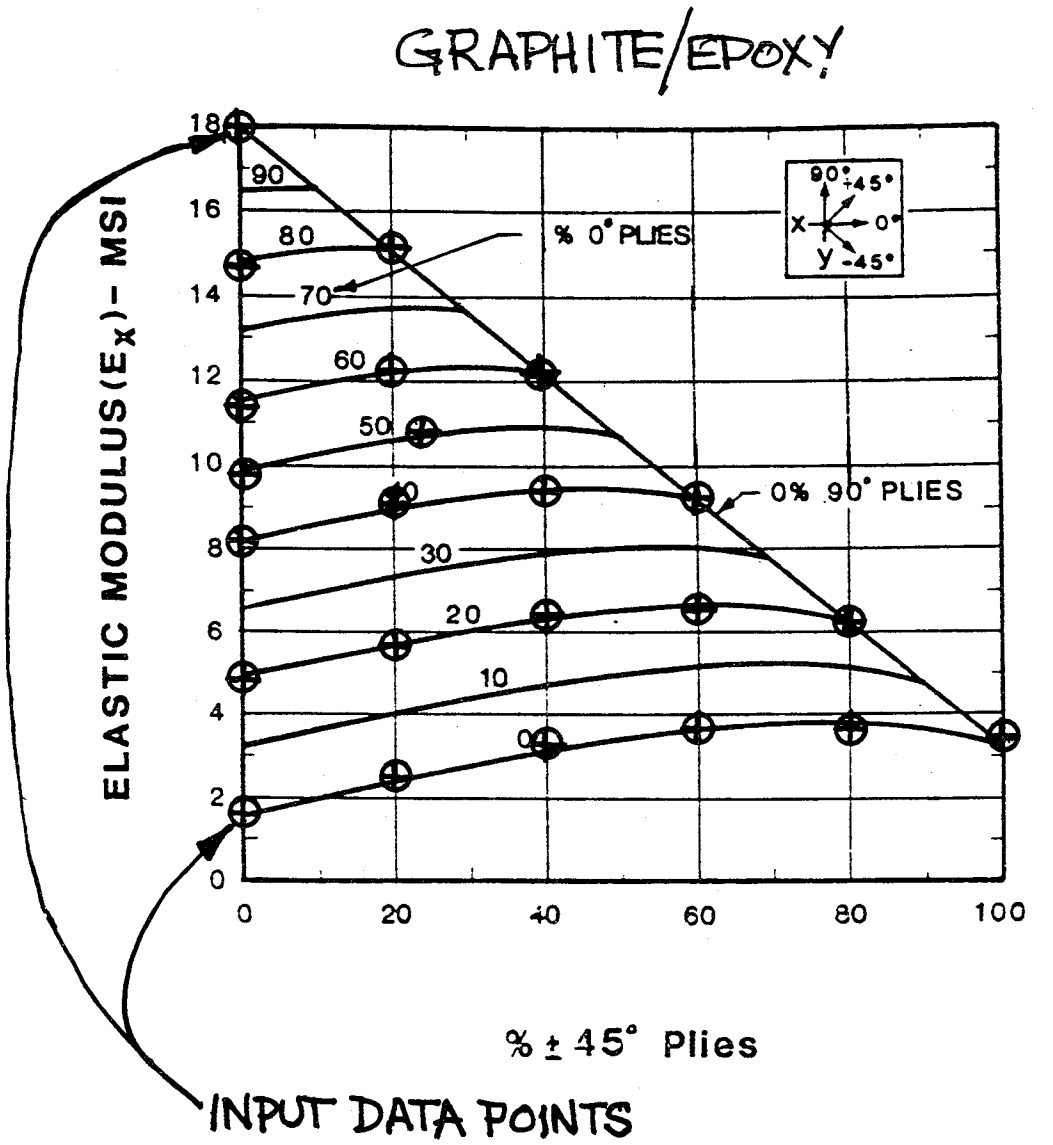
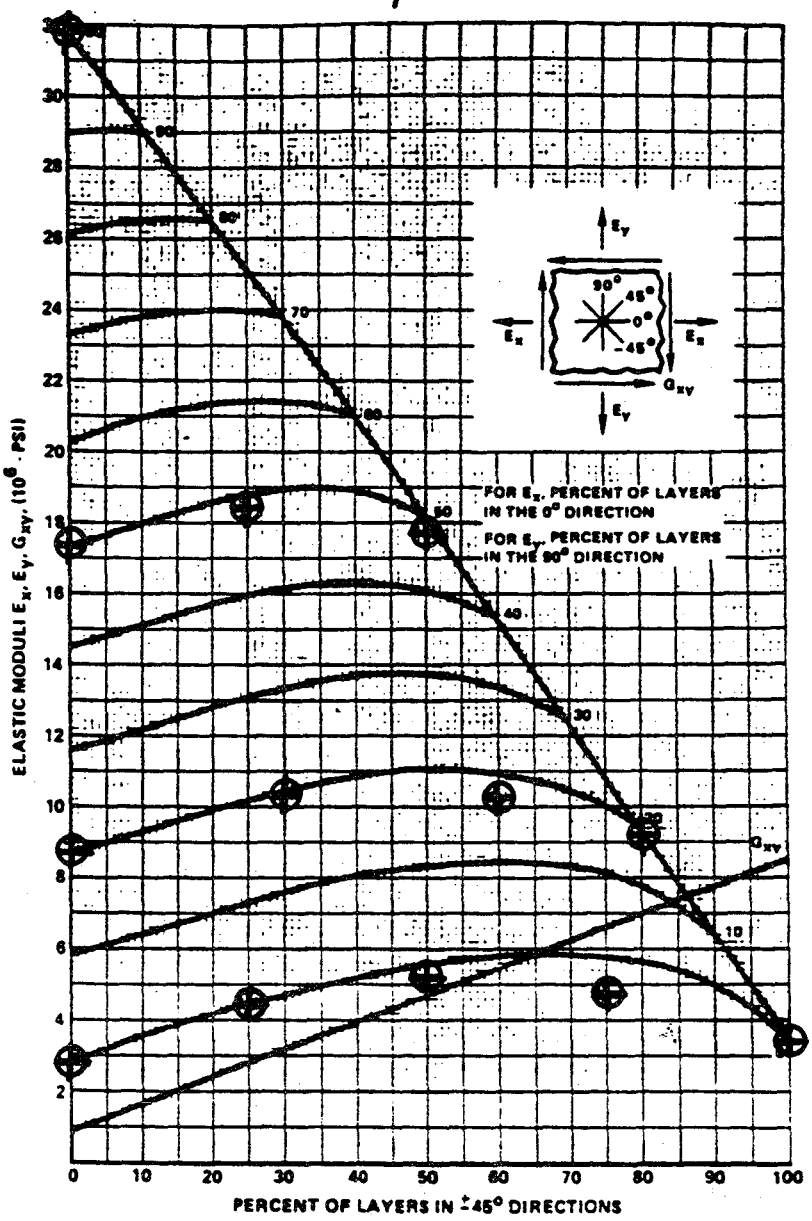


FIGURE 6 (FROM REF. 9)

BORON/EPOXY



ELASTIC MODULI FOR $[0_i/\pm 45_j/90_k]$ FAMILY, RT

⊕ OUTPUT OF "CARPLOT" SUBROUTINE

FIGURE 7 (FROM REF. 8)

4.2 SUBROUTINE 2 Laminate Analysis

The solution of classical laminate theory as performed in this subroutine is described in detail in References 1-4. The notation of Ref. 1 is used in this description. It begins with the formulation of the 3-dimensional Hooke's Law in matrix notation:

$$(1)\quad \begin{Bmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \tau_{23} \\ \tau_{31} \\ \tau_{12} \end{Bmatrix} = \begin{bmatrix} C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16} \\ C_{12} & C_{22} & C_{23} & C_{24} & C_{25} & C_{26} \\ C_{13} & C_{23} & C_{33} & C_{34} & C_{35} & C_{36} \\ C_{14} & C_{24} & C_{34} & C_{44} & C_{45} & C_{46} \\ C_{15} & C_{25} & C_{35} & C_{45} & C_{55} & C_{56} \\ C_{16} & C_{26} & C_{36} & C_{46} & C_{56} & C_{66} \end{bmatrix} \begin{Bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \gamma_{23} \\ \gamma_{31} \\ \gamma_{12} \end{Bmatrix}$$

The 6x6 matrix of C_{ij} terms is known as the stiffness matrix. When we limit our analysis to planar laminates it reduces to a 3x3 matrix because row and column 3,4, and 5 become zero by definition. The equation is then written:

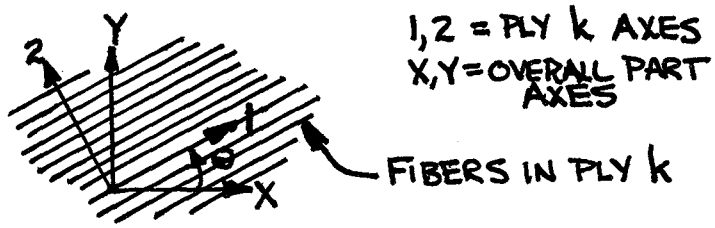
$$(2)\quad \begin{Bmatrix} \sigma_1 \\ \sigma_2 \\ \tau_{12} \end{Bmatrix} = \begin{bmatrix} Q_{11} & Q_{12} & 0 \\ Q_{12} & Q_{22} & 0 \\ 0 & 0 & Q_{66} \end{bmatrix} \begin{Bmatrix} \epsilon_1 \\ \epsilon_2 \\ \gamma_{12} \end{Bmatrix}$$

Stiffness terms are determinable from engineering constants which define the basic material properties of the plies:

$$(3)\quad Q_{11} = \frac{E_{11}}{1 - \nu_{12}\nu_{21}} \qquad Q_{22} = \frac{E_{22}}{1 - \nu_{12}\nu_{21}}$$

$$Q_{12} = \frac{\nu_{12}E_{22}}{1 - \nu_{12}\nu_{21}} \qquad Q_{66} = G_{12}$$

A further generalization of Hooke's Law is required, however, to allow for the possibility that the fibers in any individual layer, or ply, may not be aligned with the XY axes which are used to describe the overall size and shape of the finished composite part and which are also used to specify the applied loads.



The coordinate rotation is described by:

$$(4)\quad \begin{Bmatrix} \sigma_x \\ \sigma_y \\ \tau_{xy} \end{Bmatrix} = [T]^{-1} \begin{Bmatrix} \sigma_1 \\ \sigma_2 \\ \tau_{12} \end{Bmatrix}$$

$$(5)\quad [T] = \begin{bmatrix} \cos^2\theta & \sin^2\theta & 2\sin\theta\cos\theta \\ \sin^2\theta & \cos^2\theta & -2\sin\theta\cos\theta \\ -\sin\theta\cos\theta & \sin\theta\cos\theta & \cos^2\theta - \sin^2\theta \end{bmatrix}$$

Figure 8 shows the stackup of a typical laminate to be analyzed and the axis and sign conventions used in the mathematical description of the laminate.

The concept of tensor strain, which is simply half of the actual strain, is introduced so that stress and strain can be tansformed using the same matrix:

$$(6)\quad \begin{Bmatrix} \epsilon_1 \\ \epsilon_2 \\ \gamma_{12} \end{Bmatrix} = [R] \begin{Bmatrix} \epsilon_1 \\ \epsilon_2 \\ \frac{\gamma_{12}}{2} \end{Bmatrix} \quad \text{WHERE} \quad [R] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

Hooke's Law can now be written as

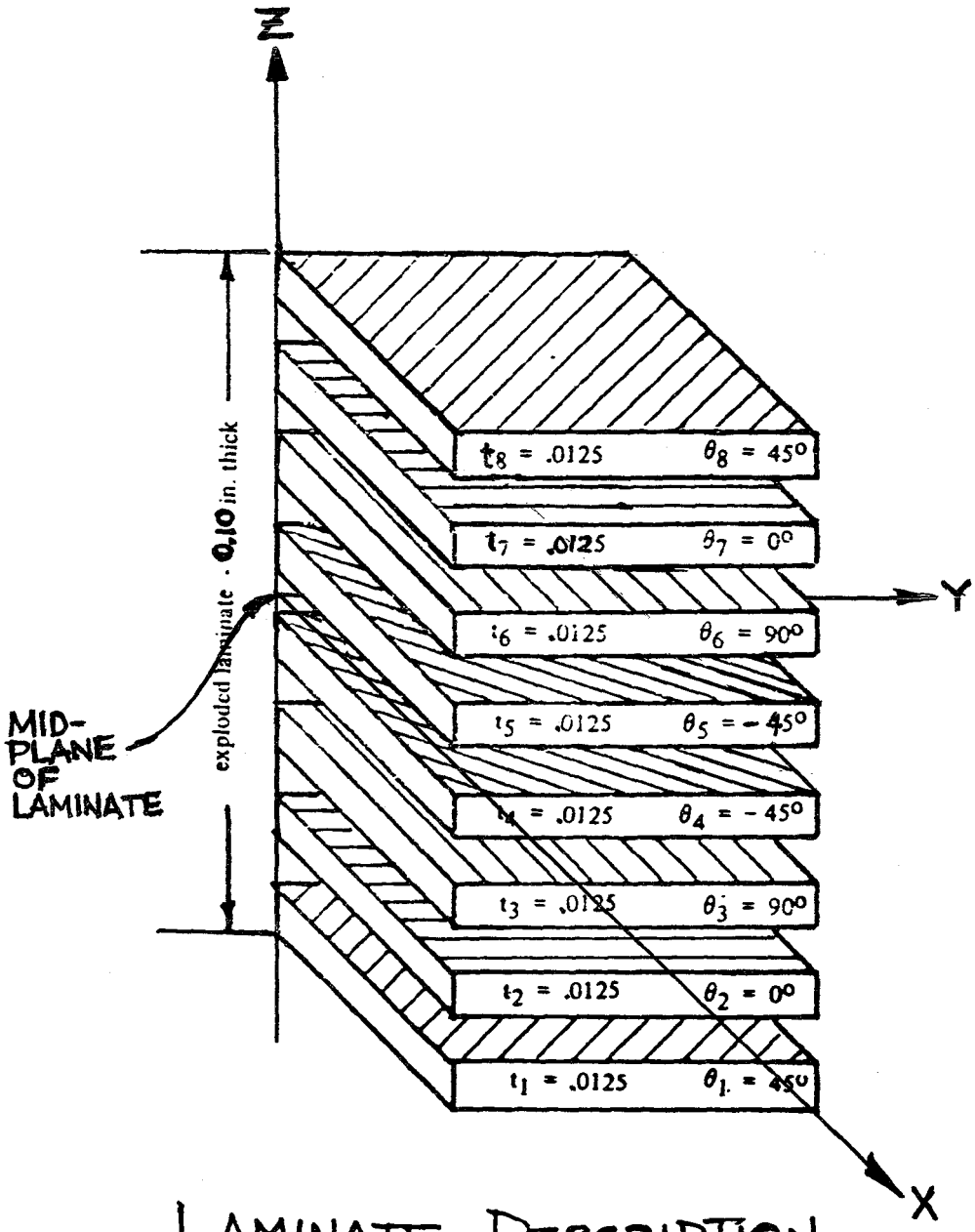
$$(7)\quad \begin{Bmatrix} \sigma_x \\ \sigma_y \\ \tau_{xy} \end{Bmatrix} = [T]^{-1}[Q_{ij}][R][T][R]^{-1} \begin{Bmatrix} \epsilon_x \\ \epsilon_y \\ \gamma_{xy} \end{Bmatrix}$$

or, in shorthand notation

$$(8)\quad \begin{Bmatrix} \sigma_x \\ \sigma_y \\ \tau_{xy} \end{Bmatrix} = [\bar{Q}_{ij}] \begin{Bmatrix} \epsilon_x \\ \epsilon_y \\ \gamma_{xy} \end{Bmatrix}$$

where

$$\bar{Q}_{11} = Q_{11}\cos^4\theta + 2(Q_{12}+2Q_{66})\sin^2\theta\cos^2\theta + Q_{22}\sin^4\theta$$
$$\bar{Q}_{12} = (Q_{11}+Q_{22}-4Q_{66})\sin^2\theta\cos^2\theta + Q_{12}(\sin^4\theta+\cos^4\theta)$$
$$\bar{Q}_{22} = Q_{11}\sin^4\theta + 2(Q_{12}+2Q_{66})\sin^2\theta\cos^2\theta + Q_{22}\cos^4\theta$$
$$\bar{Q}_{16} = (Q_{11}-Q_{12}-2Q_{66})\sin\theta\cos^3\theta + (Q_{12}-Q_{22}+2Q_{66})\sin^3\theta\cos\theta$$
$$\bar{Q}_{26} = (Q_{11}-Q_{12}-2Q_{66})\sin^3\theta\cos\theta + (Q_{12}-Q_{22}+2Q_{66})\sin\theta\cos^3\theta$$
$$\bar{Q}_{66} = (Q_{11}+Q_{22}-2Q_{12}-2Q_{66})\sin^2\theta\cos^2\theta + Q_{66}(\sin^4\theta+\cos^4\theta)$$



LAMINATE DESCRIPTION

$t, \theta; E_{11}, E_{22}, G_{12}, \nu_{12}$

FIGURE 8

Keep in mind that this relationship is only valid for a single layer. The terms in these equations must be recalculated for each ply.

The manner in which these relations are most frequently used employs the equations which result from integrating the stress distribution over the thickness of the built-up laminate. The equations thus derived relate inplane forces, N, and out of plane moments, M, to the extension strains, ϵ and γ, and curvatures of the laminate center plane, κ.

$$(9)\quad \begin{Bmatrix} N_x \\ N_y \\ N_{xy} \end{Bmatrix} = \begin{bmatrix} A_{11} & A_{12} & A_{16} \\ A_{12} & A_{22} & A_{26} \\ A_{16} & A_{26} & A_{66} \end{bmatrix} \begin{Bmatrix} \epsilon_x \\ \epsilon_y \\ \gamma_{xy} \end{Bmatrix} + \begin{bmatrix} B_{11} & B_{12} & B_{16} \\ B_{12} & B_{22} & B_{26} \\ B_{16} & B_{26} & B_{66} \end{bmatrix} \begin{Bmatrix} K_x \\ K_y \\ K_{xy} \end{Bmatrix}$$

$$(10)\quad \begin{Bmatrix} M_x \\ M_y \\ M_{xy} \end{Bmatrix} = \begin{bmatrix} B_{11} & B_{12} & B_{16} \\ B_{12} & B_{22} & B_{26} \\ B_{16} & B_{26} & B_{66} \end{bmatrix} \begin{Bmatrix} \epsilon_x \\ \epsilon_y \\ \gamma_{xy} \end{Bmatrix} + \begin{bmatrix} D_{11} & D_{12} & D_{16} \\ D_{12} & D_{22} & D_{26} \\ D_{16} & D_{26} & D_{66} \end{bmatrix} \begin{Bmatrix} K_x \\ K_y \\ K_{xy} \end{Bmatrix}$$

$$(11)\quad A_{ij} = \sum_{k=1}^{N} (\bar{Q}_{ij})_k t_k \;,\quad B_{ij} = \frac{1}{2} \sum_{k=1}^{N} (\bar{Q}_{ij})_k (z_k^2 - z_{k-1}^2) \;,\quad D_{ij} = \frac{1}{3} \sum_{k=1}^{N} (\bar{Q}_{ij})_k (z_k^3 - z_{k-1}^3)$$

The A,B and D-matrices contained in these equations are widely used descriptions of laminated composite behavior. In shorthand notation, the above equations are written as:

$$(12)\quad \begin{Bmatrix} N \\ \hline M \end{Bmatrix} = \left[\begin{array}{c|c} A & B \\ \hline B & D \end{array}\right] \begin{Bmatrix} \epsilon \\ \hline K \end{Bmatrix}$$

Since applied forces and moments are usually the known quantities, the above equation is frequently inverted to:

$$(13)\quad \begin{Bmatrix} \epsilon \\ \hline K \end{Bmatrix} = \left[\begin{array}{c|c} A' & B' \\ \hline B' & D' \end{array}\right] \begin{Bmatrix} N \\ \hline M \end{Bmatrix}$$

Using this equation, strains in each individual ply are calculated. A judgement of likelihood of failure is based upon comparison of strain in each ply to the allowable strain for the type of material used in that ply. If any ply fails, the part is considered to have failed.

A special note is necessary here. It is only true that $A'=A^{-1}$ and $D'=D^{-1}$ if $B=B'=0$. The designer can cause the B-matrix to be identically zero by specifying a balanced and symmetric laminate. This means that any $+\theta$ layer orientation is balanced by a layer with $-\theta$ orientation. In other words, the number of $+\theta$ layers equals the number of $-\theta$

layers. Also, the distribution of layer orientations is symmetric above and below the center plane of the laminate. This does not occur automatically. It is up to the designer to be aware of this and to specify a proper layup. All aerospace companies known to the author instruct designers to use a balanced and symmetric layup except in very specialized designs.

The LAMINAT2 subroutine, which is problem 2 in the main menu, calculates the A, B, and D-matrix and inverts the A and D-matrix. It solves for the strains and curvatures presuming B=0, and prints a warning to that effect immediately after the B-matrix in the output. The designer can thus verify that he has input an appropriate layup.

4.3 SUBROUTINE 3 Laminate Coefficient of Thermal Expansion

Thermal expansion is typically of far greater concern in the space part of aerospace design than it is in the aeronautical part. Space structures under development are quite large in some cases and temperature variations from sunlight to shade on orbit are typically 200 degrees Farenheit. Both of these factors can contribute to unacceptably large thermal deformations in spite of the fact that CTE's for commonly used graphite/epoxy are considerably smaller than CTE's for aluminum or steel. Proper latching of Space Shuttle payload bay doors and pointing accuracy and/or focus of the Space Telescope are examples of potential problems of this nature. For this reason a subroutine named CTE using the equations of Ref. 4 is included as problem 3. It calculates CTE in both longitudinal and transverse direction for a laminate of any arrangement of ply material and ply orientation. Stiffness properties and CTE's for individual plies must be known for input. This CTE calculation is probably inseparably coupled with the stiffness properties calculated using problem 1 or 2 in most design projects. Varying ply orientation in the stackup to minimize CTE must, of course, be done with care since it also changes the stiffness of the laminate. It may be impossible to optimize both CTE and stiffness for a particular structural part.

5. CONCLUSIONS AND RECOMMENDATIONS

The following conclusions and recommendations must be considered tentative in some cases because the COMPSIZE program is only becoming available to MSFC designers at the time of writing of this report. The author has demonstrated the program to most of the Structural Development Branch engineers for their initial impressions and suggestions but it has not yet been used for "production".

Conclusions:

1. Based on the reactions of the branch design engineers and upon the author's experience as both designer and design teacher it seems highly likely that COMPSIZE will be a useful tool and will function as intended. Users adjust quickly to the program and appear satisfied with the type and quantity of answers they can get. The accuracy of calculated results in determining the practicality of preliminary design concepts can only be determined after it has been used for a period of time. It is currently in use on a Space Station project.

2. The employment of a personal computer for composite design seems to be a workable idea. The IBM PC on which the program was developed is located in a 4-man office, and its proximity noticeably improves the feelings of the branch personnel about its utility. The PC has more than adequate memory and its solution speed, as far as the user can tell, is instantaneous. BASIC language was adequate to produce a highly user-friendly program format and presented no difficulty in setting up solutions of the required equations.

Recommendations:

Recommendations for further development of the program to increase its utility are as follows.

1. Add selectable files of material properties for composite and homogeneous materials typically used in space structures. This would significantly reduce the time required for data input, particularly for laminate analysis.

2. Develop methods for structural optimization, probably to minimize weight, maximize stiffness, and minimize CTE.

3. Add further problem types to the main menu. These might include:

buckling of stiffened panels
ring frames with inplane loads

thermal stresses
expansion of composites due to moisture
absorption
unsymmetric bending
normal loads on plates
stress concentrations

4. Make COMPSIZE available to other users. The sponsoring branch listed on the title page is the appropriate contact.

6. REFERENCES

1. Jones, Robert M., Mechanics of Composite Materials, McGraw-Hill Book Company, New York, 1975.

2. Tsai, Stephen W. and Hahn, H. Thomas, Introduction to Composite Materials, Technomic Publishing Co., Westport, CT, 1980.

3. Lubin, George, Handbook of Composites, Van Nostrand Reinhold co., New York, 1982.

4. Whitney, James M., Isaac, Daniel M., and Pipes, Byron R., Experimental Mechanics of Fiber Reinforced Composite Materials, Society for Experimental Stress Analysis, Brookfield Center, CT, 1982.

5. Beer, Ferdinand P., and Johnston, E. Russell, Mechanics of Materials, McGraw-Hill Book Company, New York, 1981.

6. Higdon, Archie, et al., Mechanics of Materials, John Wiley and Sons, New York, 1978.

7. Military Handbook 17A, Plastics for Aerospace Vehicles, Department of Defense, Washington, D.C., 1971.

8. DOD/NASA Advanced Composites Design Guide, Flight Dynamics Laboratory, Wright-Patterson Air Force Base, Ohio, 1983.

9. Advanced Composites Design Handbook, Boeing Commercial Airplane Co. Report D6-44714 (Proprietary), 1978.

10. Verderaime, V., Development of In Situ Stiffness Properties for Shuttle Booster Filament Wound Case, NASA Technical Paper 2377, Marshall Space Flight Center, 1984.

11. Development of Engineering Data on Advanced Composite Materials, Air Force Wright Aeronautical Laboratories Report AFWAL-TR-81-4172, University of Dayton Research Institute, Dayton, Ohio, 1982.

1985

NASA/ ASEE SUMMER FACULTY RESEARCH FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

A PHOENICS MODEL FOR THE TRIAXIAL LOADING OF AN INITIALLY CYLINDRICAL MASS OF RATE-TYPE MATERIAL WITH PROVISIONS FOR BULGING AND YIELD

Prepared By:	Kenneth W. French, Jr. PhD
Academic Rank:	Professor
University & Department	John Brown University Mechanical Engineering
NASA/ MSFC Laboratory: Division: Branch:	 Systems Dynamics Atmospheric Sciences Fluid Dynamics
NASA Counterpart:	N. C. Costes, PhD.
Date:	August 9, 1985
Contract Number:	NASA-NGT-01-008-021 University of Alabama in Huntsville

A PHOENICS MODEL FOR THE TRIAXIAL LOADING OF AN INITIALLY CYLINDRICAL MASS OF RATE-TYPE MATERIAL WITH PROVISIONS FOR BULGING AND YIELD

BY

Kenneth W. French, Jr.
Professor of Mechanical Engineering
John Brown University
Siloam Springs, Arkansas

ABSTRACT

This work traces the response of a granular material via the Ten Coefficient Truesdell rate-type constituitive model into the simplest meaningful loading: the triaxial test configuration. A functional relation has been posed for computing the rather peculiar relation between average applied stress and average porosity. Using that relation an attack has been mounted on the dilemna that exists between dynamic and constitutive use of the pressure variable; that is relating dynamic pressure, thermodynamic pressure, stress deviator and higher stress invariants. The resolution was as a linear superposition with a one-way feedback, in that while the dynamic component could not effect the constituitive component, the converse was not true since density appears in the momentum transport relation.

There were two stages of preparation for this effort. The first was to use the PHOENICS Satellite to create the boundary conditions simulating the triaxial test. Two opposed parallel disks are joined with a thin cylindrical membrane with the granular material inside and a confining pressure outside. The membrane condition is modeled as a circular tension ring active on slices parallel to the disks; ie "slabs". Use was made of the PHOENICS provision for porosity greater than unity to represent buldging. The second preparation was for the granular constitution, which was coded into the Ground portion of the PHOENICS package. Provision was made by an ancillary computation to test for yield condition attempt to locate shear banding. Several non-granular cases were built to test the boundary conditions independent of the constitution and the low velocity response of PHOENICS and to disarm inherent simple fluid prejudice in the code. Some results were obtained and a covey of suggestions for parameter and computational control were generated.

ACKNOWLEDGEMENT

The efforts of my counterpart, Dr. N.C. Costes have insured that this second interaction with NASA at MSFC has been as productive and mutually rewarding as the first. All the members of the Fluid Dynamics Branch from Dr. Fichtl to the USRA and NRC Fellows have shown real enthusiasm for cooperative work and have lent me physical insight and CFD consultation.

I would like to acknowledge the ASEE/ SFF management team as well. Dr. Gerald Karr has administered the program with great dispatch, having held to a minimum the required paperwork and shown a considerable flexibility in dealing with the special requirements of university faculty on short term assignment. And on the NASA side my compliments to Dr. Dozier and especially to Leroy Osborn for hospitality and an excellent series of overview presentations on the status and future of NASA.

LIST OF FIGURES

CONTENTS

INTRODUCTION

This study has been directed at supporting an established and ongoing preparation for a triaxial test of spherical glass beads to be operated in the low-gravity environment of the Space Shuttle on orbit. The accronym MGM, mechanics of granular media, has been coined by the principal investigators, Dr. N.C. Costes and Prof. S. Sture, as the descriptor of this activity, and it is into that general arena that this effort has been cast.

The first phases of this work were performed in the summer of 1984 and reported in the image of this volume as Section XIV; "The Compatibility of an Existing CFD Code with a Broader Class of Constitutions". The results and nomenclature of that report are to some extent assumed as background, especially the portions directed at the flexibilities built into the PHOENICS CFD Code. Reference will be made on occasion to groupings of variables unique to that report and to the general forms of transport density and pressure.

To some extent the preliminary planning has been preoccupied with some of the aspects of getting the experiment into position for loading. These are all fairly extensive problems in their own right. Some of them are: 1)consolidation by launch acceleration with vibration, 2) membrane leakage during drawdown during confinement, 3) consequences of granule friction on the end platens during preload cycling, 4) the visibility of shear-bands with non-orientable granules (spheres), and 5) the prejudice of a failure surface to a noncylindrical initial boundary.

Granular mechanics with its more complicated constitutive behavoir has had to wait longer for analytic treatment than simple fluids and simple solids. The general realm of constitutive types is shown as Figure 1. In looking to associate a complicated behavoir with "fluid" or "solid", the back and forth coupling lines on the figure are very illuminating. The point being that a step up the tree toward generality from either simple fluid or simple solid branches is coupled to both. Hence it is perhaps not surprising that a flexible scheme for CFD might be ammenable to computational MGM.

The governing parameter is the extent of intergranular stresses. In gravity currents of dust laden air or silt laden water, the granular material supplements the density but can be ignored in the constitution. In a fluidized bed the granular matrix remains stationary or slowly circulating, but once levitated the intergranular stresses are low, likewise in the liquid counterpart percolation. Alternately in the very similar "pore-water" problem the active element is the granular matrix with the liquid adding a deviator hence the conditions; drained/ undrained. Recently there has been a flurry of activity using kinetic

theory to model high-speed granular flows, obviously a class of flows in which the effects of the environmental fluid are ignored. A CFD package like PHOENICS includes features for multiphase flows to make modeling of granular-laden fluids possible, but of course allow no automatic mechanism for intergranular stress computation. Further Phoenics provides a constant named DARCY which builds a linear pressure gradient on volume flow rate. Herein quite a different approach is taken in which the constitution is coded into the exchange coefficients and functionally related to gradient of velocity and the density as well as a "pressure" adjustment which adds intergranular stresses to the dynamic component.

This study is then at the intersection of a contemporary CFD package taken to the outer limit of its flexibility and numerical MGM takenas its most comprehensive constitutive description. The emphasis is on matching feature for feature rather than creating a massive supplement to the PHOENICS code.

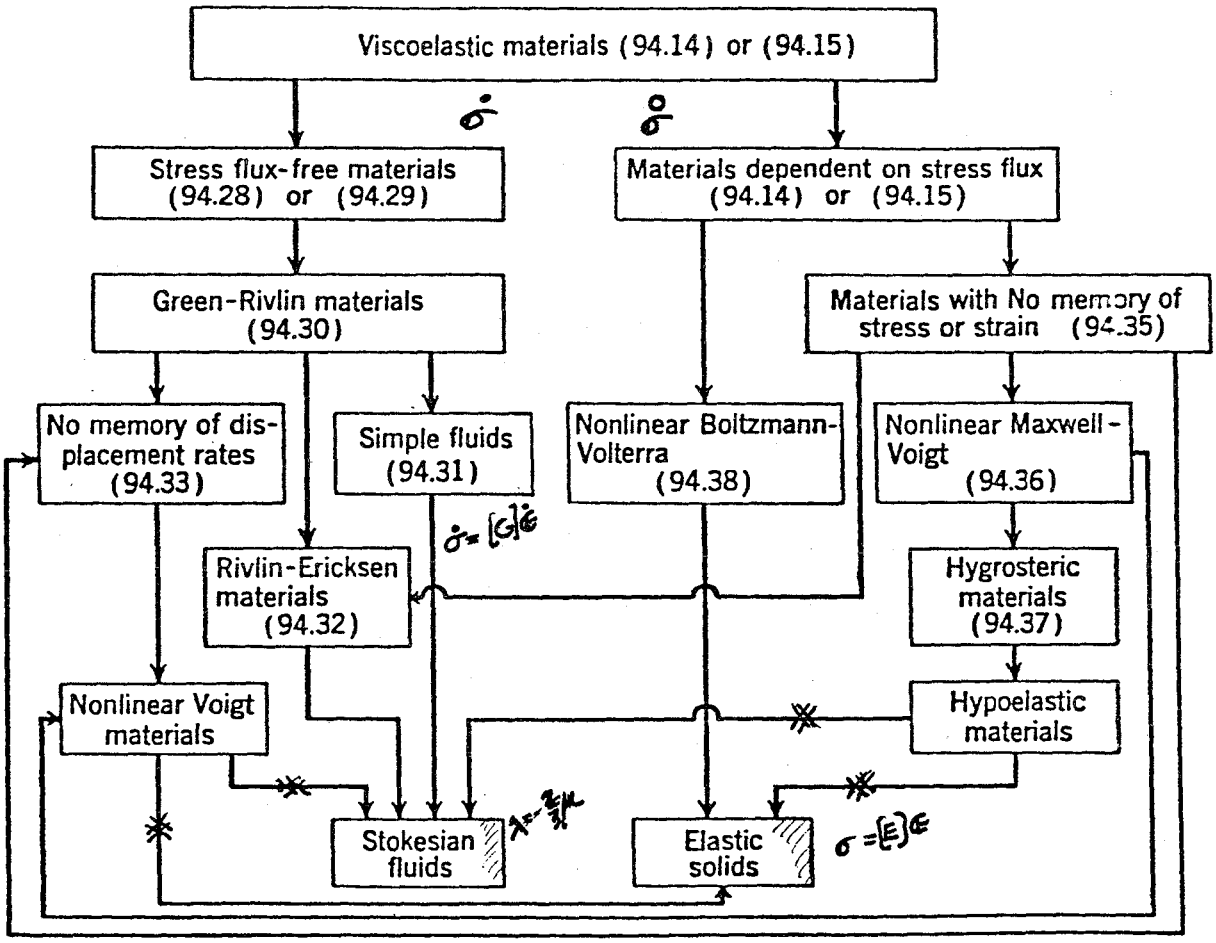


FIGURE 1 Constitutive Realms

OBJECTIVES

The primary objective of this study was to produce a cataloging of constitutive and boundary features that must be included in numerical models wether the destination code be PHOENICS or some alternative. While "PHOENICS thought" has been used to give some explicit definitions to the features, the features themselves are generic. Hence the object sought was the simplest model that retained the full features of the rationally derived MSD procedure while applying only to the triaxial configuration. Perhaps the greatest accomplishment in this regard is to arrive at specific conclusions for handling the boundary conditions in a way that is ammenable to the PHOENICS code, MICROFEM, or a custom code and more importantly to include the discoveries of preliminary testing. Further objectivity was sought by making explicit the offset terms in the Jaumann stress flux for this geometry and loading.

Specifically the intention herein was to approach the goal of predicting specific loading responses from the rational direction, carefully highlighting the simplifications to permit backtracking should comparison of the model with experimental results ultimately show it to be inadequate due to over simplification. Since this study is preliminary to extensive testing it was not desireable to over simplify just to obtain "instant gratification" as analytic solutions.

Another objective has been to emphasize the need for preparation of a numerical approach that explicitly generates fields of "signed" deviatoric stress power and makes them available to the evolving calculation. It is also desireable to determine the inhibition or enhancement parameter for shear bands based on granule size and shape, viz large spherical granules.

THE PRESSURE-DENSITY DILEMMA

The goal is to relate pressure in the PHOENICS sense with the elements of tr σ and to trace the flow of influence of density into the PHOENICS pressure both inherently and user coded. The goal is extended to include the realm of the triaxial test and to show what measures are necessary to circumvent PHOENICS provisions and to implement the full constitution of a granular media.

Seveness to the ancient Hebrews indicated completion of a cycle and so it is with PHOENICS; six adjacent neighbors in space and one in time. Figure 2 shows PHOENICS pressure as the pressure component with a dynamic origin, P_{DYN}, and in obedience to Pascal's Principle it has

equal components and lies along the hydrostatic line. The source term for momentum transport equations ($\phi_i = V_i$) at dynamic equilibrium is the gradient of pressure; hence:

$$S_{\nabla p} = -\rho\beta\frac{D\phi}{Dt} + \frac{1}{R\beta}\nabla\cdot\left(X\frac{\Gamma}{\rho}\nabla\phi\right) + \frac{\phi X}{R\beta}\,tr\,D$$

The appearance of density, ρ, and transport density $X=\rho\beta R$, are the coupling between the constitutive equation and the computation of pressure. Figure 4 illustrates a "Ring Model" of computation to highlight the numerical symbiosis of pressure, momentum and mass. If thermodynamic pressure is defined in terms of the stress tensor as;

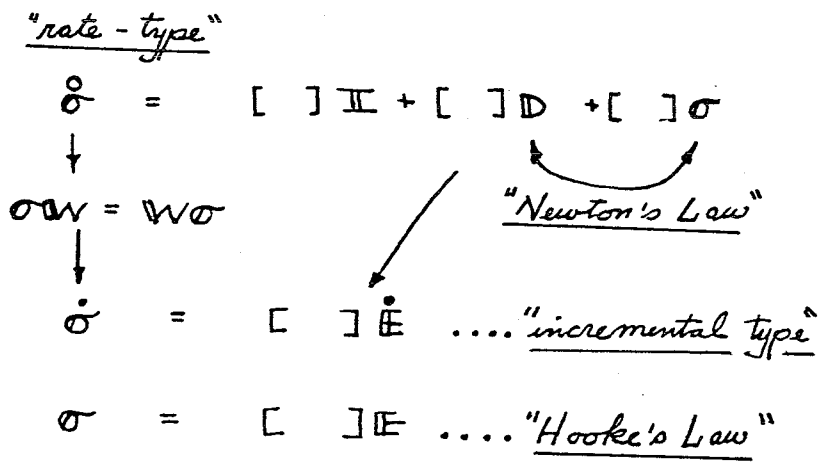
$$P_{TH} = 1/3\ tr\ \sigma$$

and further if thermodynamic pressure is considered to be a linear superposition of a constitutive and a dynamic part;

$$P_{TH} = P_{CON} + P_{DYN}$$

which would still appear along the hydrostatic axis for some types of constitutions viz; ideal gas. The linear assumption is of course not correct in general since the components have separate functional dependencies on density. In fact with a granular media behavoir changes during loading must be included and so tests have to be made for yield or rupture to determine the local functional relation between pressure and density. As a further complication, granular materials require a rate-type constitutive model,(all the references but #2 deal with this necessity). Otherwise the time dependency in the thermodynamic pressure could be isolated in the dynamic part, which would be very convenient in the CFD environment. The lure of incremental models is that time can be "eliminated" by multiplying the constitutive equation by Dt. But at best the incremental formulation introduces coordinate

system problems more difficult for CFD than the gradient of velocity components which are necessary for rate-type models with stress flux.



To see what reductions can be made with a triaxial configuration, a glance at Figure 2 shows the profile of the triaxial test in normal stress space ($\sigma_2 = \sigma_3$), the hydrostatic condition ($\sigma_1 = \sigma_2 = \sigma_3$) which is a subset, and the stress deviator ($\sigma_1 + \sigma_2 + \sigma_3 = C$) which has the same direction, but less restriction. Consider first the normal stress-normal strain relationship as shown in Figure 5; an accepted shape in soils testing. The very important "knee region" shows the difficult dependence on initial density; double-valued for initially dense conditions and for large axial strain approaching the critical state asymptotically; for which it is impossible to distinquish between initially loose and initially dense. It is analytically convenient to introduce without any rational justification (and apparently without precedence) the three parameter relation:

$$SR / SR_C = \tanh(ae) + b(1 - [\tanh(ae - c)])$$

The triaxial case gives: $p = 1/3 \text{ tr } \sigma = 1/3 \, \sigma_3 (SR + 2)$

with: $SR = \sigma_1 / \sigma_3$. The slope of the lower portion of Figure 5 is a sort of 3-D Poisson's Ratio: $\frac{2}{25}(\int d \ln v)$ it is notable that the zero of the ordinate represents: tr D = 0.

While the loose extreme compresses monotonically, the dense extreme exhibits an initial compression followed by expansion to critical state. This change of behavoir can prove to be very destabilizing to a numerical scheme.

The classical description of the critical state is given as;

$$P_c / P_o = \exp\left(\frac{v_o - v_c}{K} \right)$$ in which specific volume, the inverse of density, has been used and $P = \sigma_1 - \sigma_3$

so that tr σ = P -1/3 σ_3(1 + 2*SR)....σ_3 - confining pressure. In the more comprehensive rate-type model, this behavoir is nested in the first term of the denominator of the pressure expression:

$$P = - \frac{DP}{Dt} / \left(\alpha \dot{V} + \frac{2b\mu}{M^2} \left[\left(\frac{V}{V_c} \right)^{\gamma} - 1 \right] \mathrm{tr}\, \bar{\sigma}_* D_* \right)$$

and if all the thermodynamic time dependence were in the dynamic component of pressure;

$$P = - \dot{P}_{TH} / (\ . \ . \ . \ )$$

The exponent γ represents the nonclassical coupling between pressure and density. One of the difficulties of using this pressure model is that the deviatoric stress power can change sign in some regions of the test and the resulting loading/unloading behavoir must be traced by a supplementary computation in a CFD environment.

Isochoric plastic flow is one ostensibly at constant volume in which tr D = 0 hence $\nabla \cdot V = 0$, a flow with identically constant density. For this case the full constitutive relation reduces to:

$$\mathring{\bar{\sigma}}_* = \dot{\bar{\sigma}}_* - W\bar{\sigma}_* + \bar{\sigma}_* W$$

and by definition the deviatoric stress rate tensor is:

$$\mathring{\bar{\sigma}}_* = \frac{2\mu}{M^2} (P_c \mathbb{I} - \bar{\sigma}_*) \mathrm{tr}\, \bar{\sigma}_* \mathbb{D} + M^2 \mathbb{D}$$

The velocity vector has its gradient distributed in the standard way into deformation rate and rotation tensors:

$$\nabla \underline{V} = \mathbb{D} + W$$

the assistance of the latter has already been invoked in making an objective quantity from the substantial derivative of stress. The deformation rate is inherent in the rate-type constitution and of course appears there in the stress power with deviatoric and normal components and in the yield condition also in the stress power.

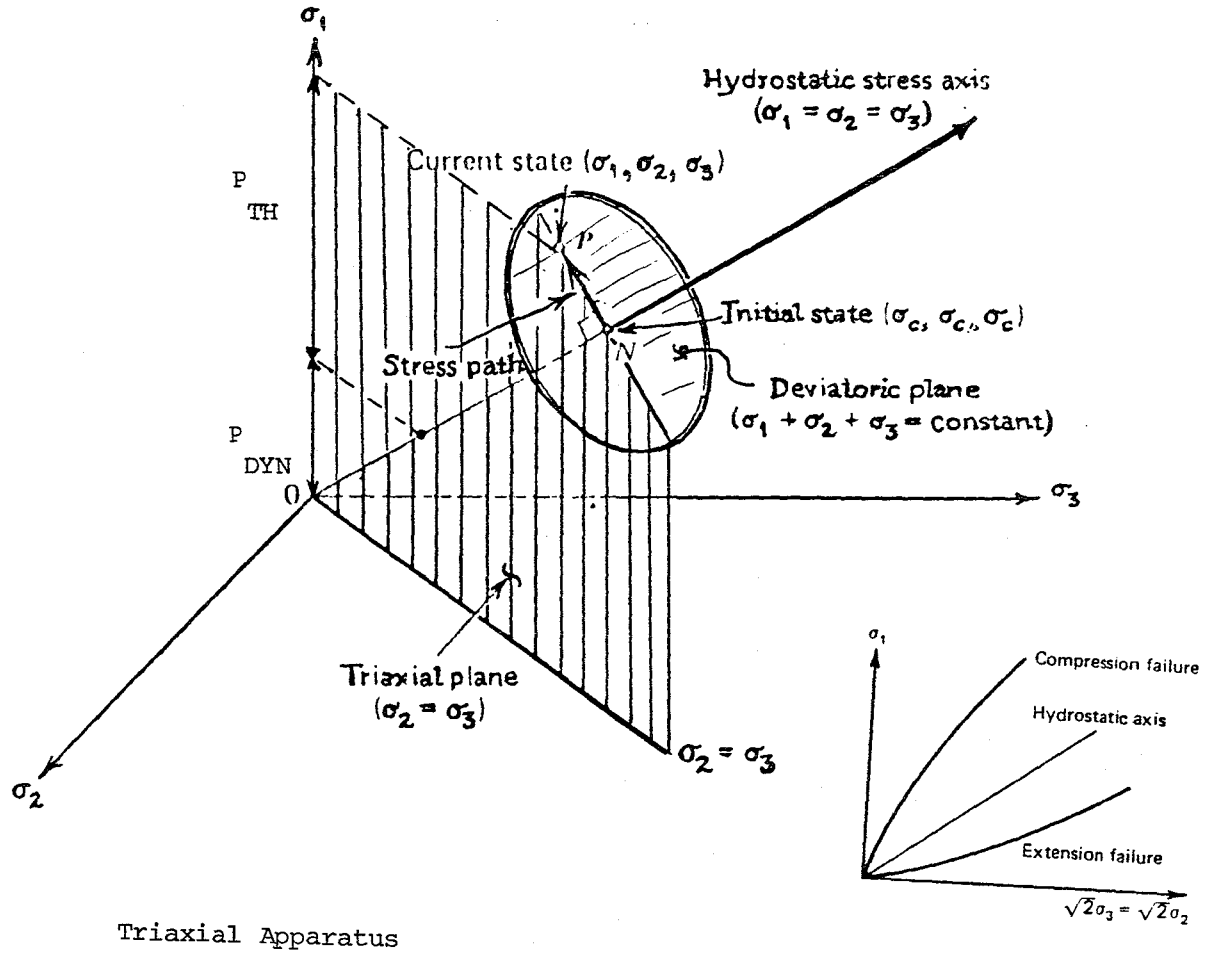
$$\mathrm{tr}\, \sigma \mathbb{D} = \mathrm{tr}\, \bar{\sigma}_* \mathbb{D}_* + P \mathrm{tr}\, \mathbb{D}$$

Macroscopically thermodynamics will place a condition on the integral

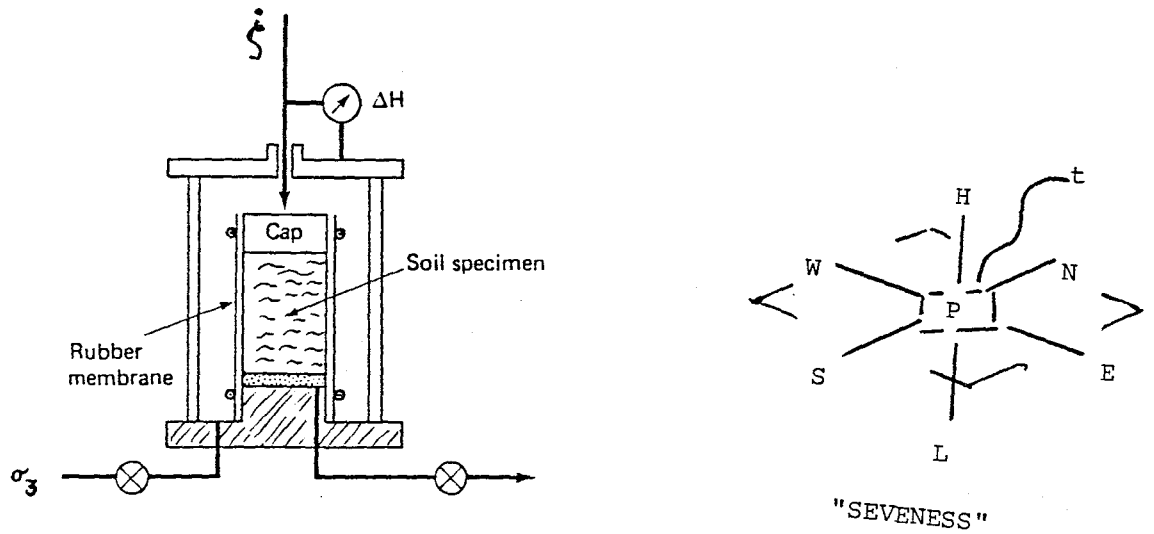
$$\int \left[\frac{D}{Dt} \int \left[\int \sigma \, dE \right] dV \right] dt$$

which represents the total energy expenditure on the test material.

$$\mathbb{D} = \mathbb{D}_* + \left(\tfrac{1}{3} \mathrm{tr}\, \mathbb{D} \right) \mathbb{I}$$



Triaxial Apparatus



"SEVENESS"

FIGURE 2 NORMAL STRESS SPACE

Failure comes in the forms of rupture and yield. Figure 2 shows on a triaxial plane the lack of symmetry of the rupture boundaries about the hydrostatic axis ("compression/ extension failure"). The other surface forming the failure boundary is given by the yield cap. Following the procedure of MSD (78 - 84) the yield surface is established by a set of Q vectors which render the array G, singular

$$\dot{Q} = [G(Q, \rho)]\, \mathbb{P} \quad \ldots \quad Q = \begin{bmatrix} p^2 \\ q^2 \end{bmatrix}$$

p - stress deviator...1/3 tr O

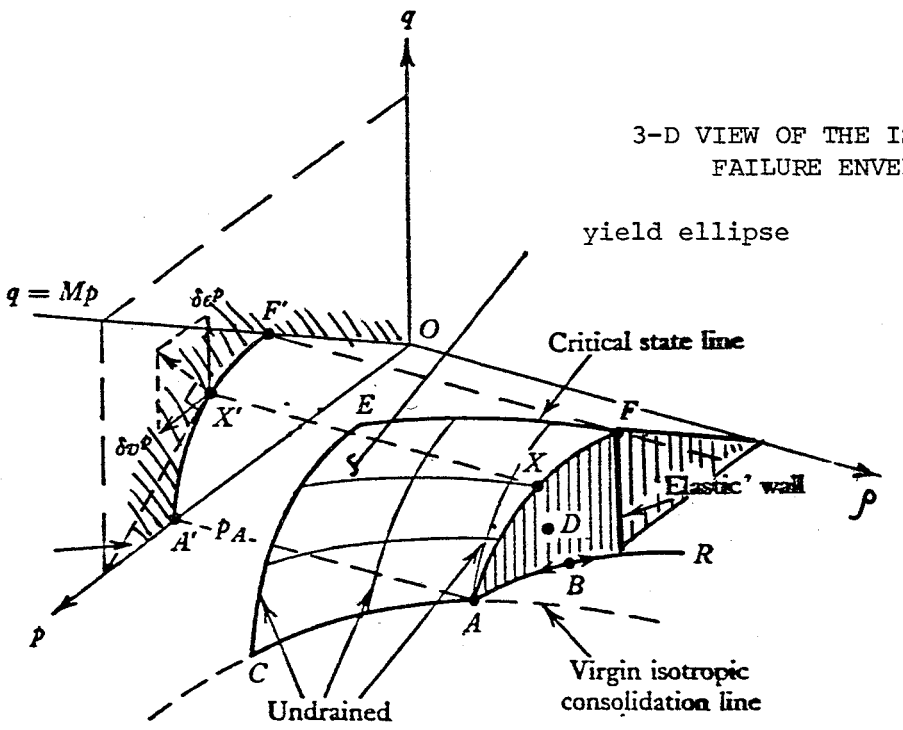
q - square root of the second invariant $\sqrt{\sigma^* \sigma^*}$ $\quad q_c = M_o p_c$

$\mathbb{P}$ - power tensor - $\begin{bmatrix} p \text{ tr } D \\ \text{tr } \sigma^* D^* \end{bmatrix}$

The density variation coming from the critical pressure-specific volume relation given above. An overly simple verbal description of this equation might run; the stress invariants are related to exponentially decaying transients scaled in the components of stress power. Aside from that description an algebraic result can be obtained if the reduced constitution (isochoric) above is used;

$$q^2 = M_o^2 (2 p_c p + p^2)$$

M_o and p_c are the classical parameters of critical state. This relation plots as an ellipse in q-p space and hence the failure envelop looks like a somewhat distorted icecream cone; as it appears in sections of the q-p-ρ space at constant density.



3-D VIEW OF THE ISOCHORIC FAILURE ENVELOPE

yield ellipse

Digressing to the specifics of a triaxial configuration with full Θ symmetry; the stress deviator is:

$$1/3 \mathrm{tr}\ \sigma = \sigma_1 + 2\sigma_3$$

and hence the deviatoric stress is:

$$\sigma^* = \begin{bmatrix} 2/3(\sigma_1-\sigma_3) & & \\ & 1/3(\sigma_3-\sigma_1) & \\ & & 1/3(\sigma_3-\sigma_1) \end{bmatrix}$$

and the deformation rate is:

$$D = \tfrac{1}{2}\begin{vmatrix} 2\frac{\partial v_z}{\partial z} & \frac{\partial v_z}{\partial r}+\frac{\partial v_r}{\partial z} & 0 \\ \frac{\partial v_r}{\partial z}+\frac{\partial v_z}{\partial r} & 2\frac{\partial v_r}{\partial r} & 0 \\ 0 & 0 & 0 \end{vmatrix}$$

and the deformation rate deviator is ;

$$\tfrac{1}{3}\mathrm{tr}\, D = \tfrac{1}{3}\nabla\cdot\underline{V} = \tfrac{1}{3}\left(\frac{\partial v_z}{\partial z} + \frac{\partial v_r}{\partial r}\right)$$

$$D^* = \begin{vmatrix} \frac{2}{3}\frac{\partial v_z}{\partial z} - \frac{1}{3}\frac{\partial v_r}{\partial r} & — & 0 \\ — & \frac{2}{3}\frac{\partial v_r}{\partial r} - \frac{1}{3}\frac{\partial v_z}{\partial z} & 0 \\ 0 & 0 & 0 \end{vmatrix}$$

and the spin rate is:

$$W = \begin{vmatrix} 0 & \frac{\partial v_z}{\partial r} - \frac{\partial v_r}{\partial z} & 0 \\ \frac{\partial v_r}{\partial z} - \frac{\partial v_z}{\partial r} & 0 & 0 \\ 0 & 0 & 0 \end{vmatrix}$$

All these components then can be used with the exchange coefficients for the PHOENICS momentum transport equations; for the ith :

$$\dot{\sigma}^* \rightarrow \begin{vmatrix} 2/3\,\dot{\sigma}_1 & & \\ & -\frac{1}{3}\dot{\sigma}_1 & \\ & & -\frac{1}{3}\dot{\sigma}_1 \end{vmatrix}$$

$$(\dot{\sigma}_3 = 0)\,\&\,(W\sigma^* = \sigma^* W)$$

$$\Gamma_{MOM_i} = \frac{\left(\dfrac{\dot{\sigma}^*}{b\left(\frac{\rho_c}{\rho}\right)^{\gamma} 2\mu \frac{\mathrm{tr}\,\sigma^* D^*}{M^2}} - 2\mu D^*\right)}{\nabla V_i}$$

$$\nabla V_i \rightarrow D_i + W_i$$

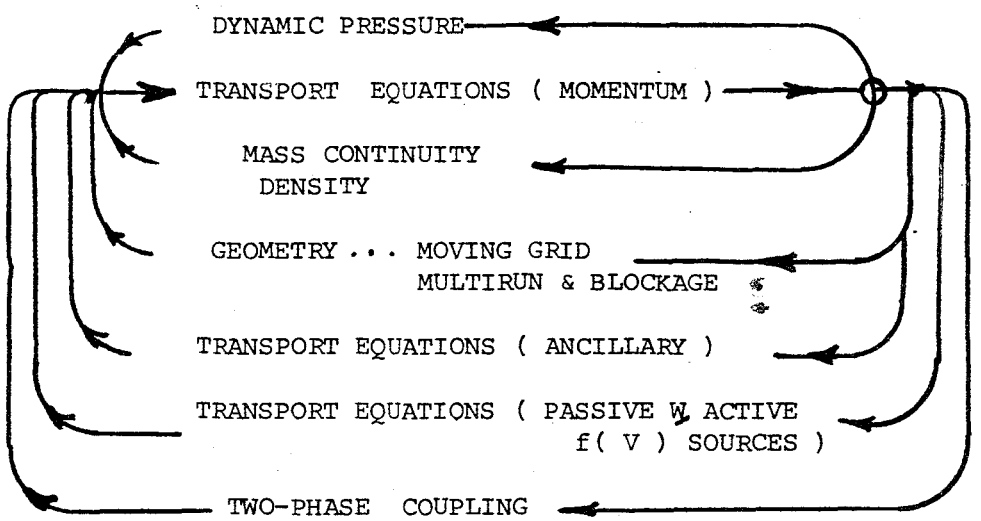
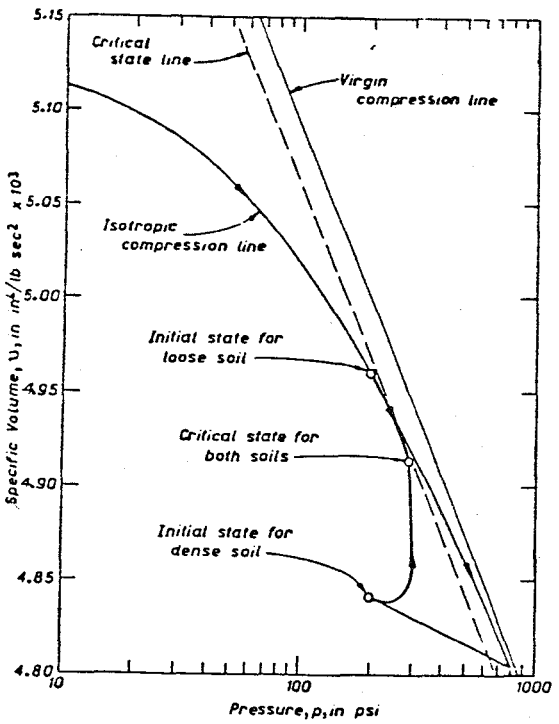
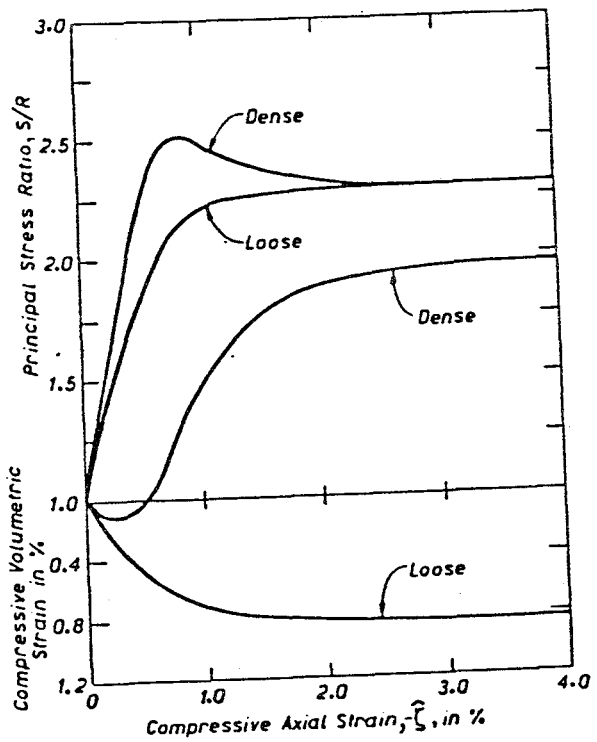


FIGURE 5 A RING MODEL FOR PHOENICS



PRESSURE-DENSITY STATES

FIGURE 3



GRANULAR STRESS-STRAIN & POISSON

FIGURE 4

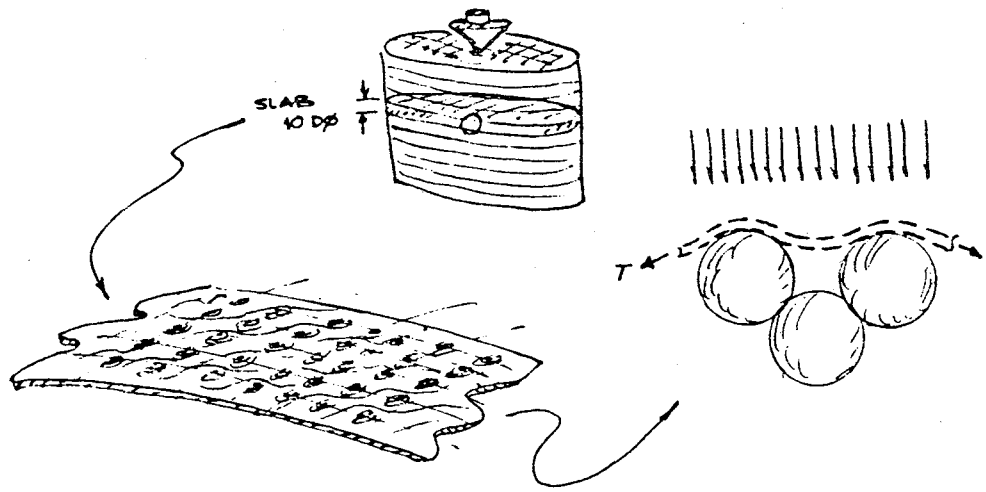


FIGURE 9 SLABS OF SPHERES

* PLATEN FRICTION
  (pseudo-bulging)

* LOCAL CURVATURE

* SHEAR BAND
  LOCATION/ DETECTION

* CONSOLIDATION
  HISTORY

* CONFINEMENT
  vs
  MEMBRANE ELASTICITY

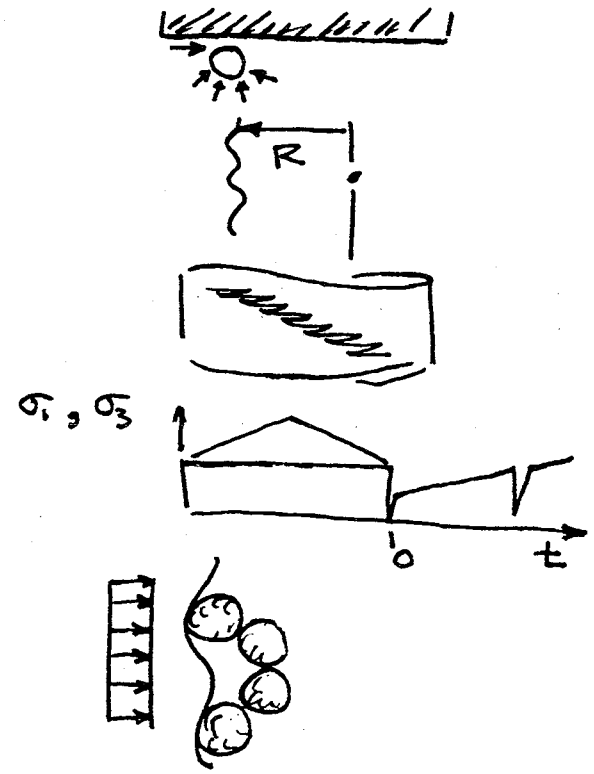


FIGURE 10 SPECIAL CONSIDERATIONS

FIGURE 7

.....List of PHOENICS FILES (KENFILES.TXT).....xxxSAT.FTN

#	xxx	Implementations
1	KEN	2D POLAR PORUS PLUG 10X20X1 UNIFORM INFLOW INTO ANNULUS FREE REGION AROUND PLUG
2	FRE	3D CARTESIAN DIAMOND FIELD 20X20X20 OBSTRUCTIONS UNIFORM INFLOW
3	KWF	DEL SQR T = DT/Dt..CONDUCTION TRANSIENT 20X20X1 2D NOFLOW
4	NCH	DEL SQR H1 = KC1 & DEL SQR H2 = KC2
5	NET	BUILD FILE..LOW V IN KWF
6	TAL	TRIAXIAL LOADING
7	TAG	TRIAXIAL GEOMETRY WITH IDEAL GAS
8	TAG1	ENCHROACHING END CAP
9	TAG2	DRAINING UNDER VACUUM
10	TAG3	MASS REDISTRIBUTED OVER MULTIRUN
11	TAG4	MEMBRANE CONDITION/ LATERAL
12	TRI	TRIAXIAL LOADING: HYPERBOLIC TANH PREYIELD: "EXCO":GRANULAR
13	TRY	(TRI AND TRY AGAIN !) YIELD DET G SINGULAR PORTION OF FIELD
14	BEN	BENARD: LOW VELOCITY TEST WITH ONSET OF CONVECTION:PLANE LAYER 10X10X10 BENFLD INITIAL FIELD FOR INSTABILITY
15	KEN (.JOB)	USER NAME KWF
16	TAL (.JOB)	USER NAME KWF..RUNS TALSAT.FTN, .EAR,.LOG

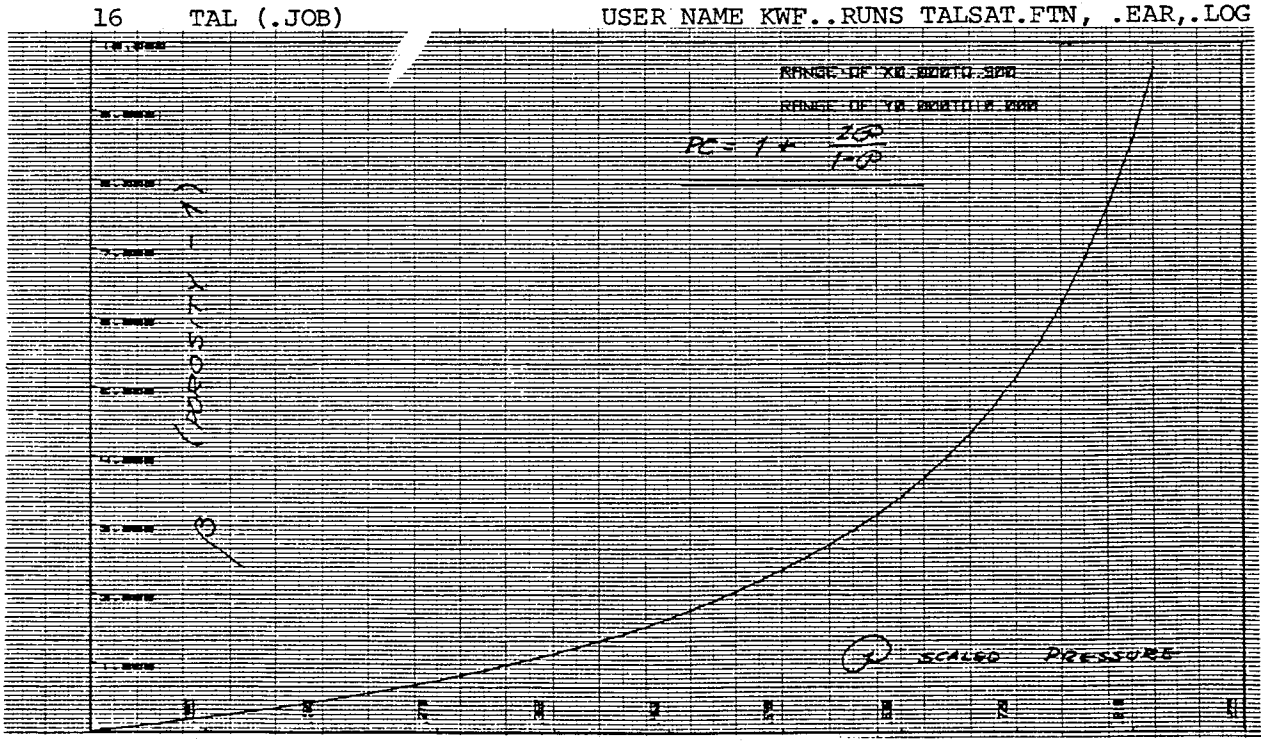
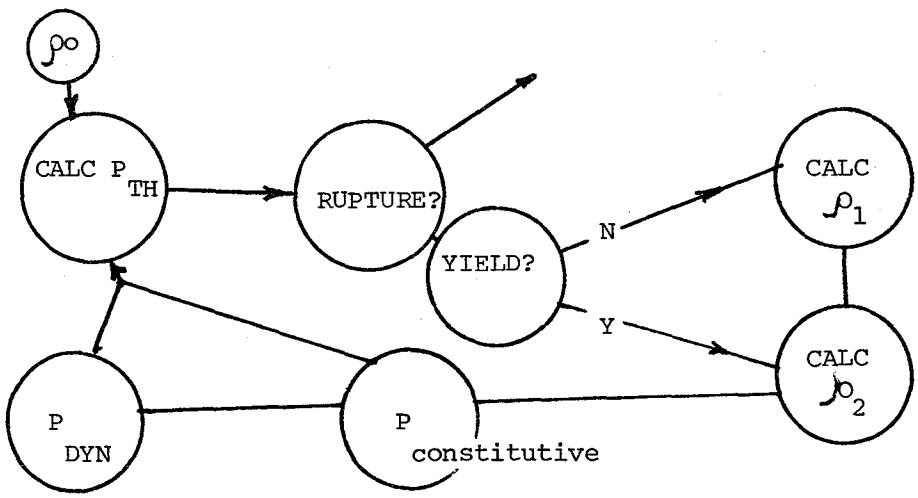


FIGURE 8 POROSITY VARYING WITH SCALED PRESSURE

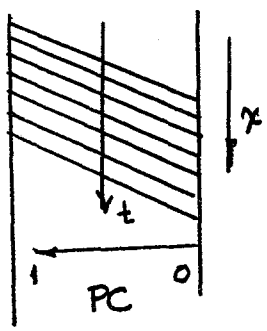
LATERAL & LONGITUDINAL BOUNDARIES

Aside from dealing with a very difficult constitution internal to the cylindrical test region, the boundary conditions present formidable difficulties to a CFD package; Figures 2,9&10. The end platens are rigid and nominally frictionless; however preliminary testing has shown that if the granular material is glass spheres(a necessity for a SS based experiment), the friction is very tenuous. The consequences of friction to the boundary conditions and to numerical computation are that an additional confinement is present on the upper and lower regions (1/3 platen diameter ?) of the cylinder of test material. Typically this over-confinement causes premature, asymmetrical bulging. Further there will certainly be a disruption to the proper growth of the failure surfaces; viz. shear bands.

The steadily advancing (constant c Figure 4) upper platen must be treated with caution numerically, since a step displacement might cause a jump in local density and hence an irreversible catastrophic change in the local constitution. That is the parallel running computations for testing realms of behavoir might be triggered, before the density wave could diffuse...."wave and diffuse referring to the numerical and not physical uses of the terms".



This advancing platen problem has been tackled in the same way as the confining membrane condition. In both a use of the porosity in the transport density, $\rho\beta R$, was implemented. The platen is replaced by an advancing wave of increasing porosity; ie the "fuzzy platen". For the bulging of the membrane it will be treated by allowing the circumferencial cells to have porosity in excess of one. Figure 8 shows a simple model of the change of porosity with which is a pressure scaled to membrane stress, P/2MK. Since friction over the spherical surfaces changes the membrane tension exponentially ($T/T_0 = \exp(-f\Theta)$), the lateral membrane effects



are small and a narrow slab (say 10 D_o) can be isolated for analysis. This will give an artificial "aliasing" as the resolution jumps from D_o to $10*D_o$ and since shear bands will be suppressed to vertical orientation on each slab.

RESULTS & CONCLUSIONS

The preliminary specifics for the numerical treatment of triaxial loading of a granular material in the PHOENICS environment have been laid to rest; constitution, boundary conditions and loading simulation; with the details given herein for slabwise, O independent modeling. Figure 6 shows some of the forays into the numerical realm to test various schemes for handling this configuration and loading. Special interest was placed on the treatment of dynamic and thermodynamic pressure and the correct decomposition of stress tensor into that pressure and the momentum exchange coefficient. The O dependent terms in Grad V, O, etc, if the shear bands are to be predicted. That inclusion will be necessary but advisable only after the much simpler version is compared with test results. Those O dependent terms make a considerable change in the accounting and in the "objectivity offset", but the changes are straightforward in the current framework.

A continuing doubt of the validity of PHOENICS at very low velocities and especially at small displacements,(V dt) persists. However the preparations and the lessons in coding will be useful in general and to any numerical approach to the triaxial loading of a granular material. A considerable portion of the summer's effort was spent pursuing minor avenues (draining, percolation, preloading, etc) that are not included in this report but which provide support to the project as a whole and all delt with various aspects of MGM.

REFERENCES

1. Sture, Stein; "Simularities between Hypoelastic and Rate-type Models";Private Communications; Civil and Environmental Engineering, Univ. Colorado, Boulder; 7/85.

2. Biot, M.A.; Mechanics of Incremental Deformation John-Wiley, 1965.

3. Chen, W.F. & A.F. Saleeb; Constitutive Equations For Engineering Materials, Volume I, John-Wiley, 1982

4. Davis, R.O., Mullenger, G.& R.F. Scott; Int. J. For Numerical and Analytic Methods in Geomechanics; John-Wiley

   1978 v.2 pp 255 "A Rate-type Constitutive Model for Soil with a Critical State".

   1979 v.3 pp 279 "Derived Failure Criteria for a Granular Media".

   1981 v.5 pp 155 " A unified Yield Criterion for Cohesionless Granular Material"

   1984(a) v.8n.2 ppl25 "Rapid Expansion in a Rate-type Soil".

   1984(b) v.8n.2 ppl44 "Rapid Shearing in a Rate-type Soil".

5. Eringen, A.C.; Nonlinear Theories of Continuous Media ; McGraw-Hill; 1962

   The Mechanics of Continua; John-Wiley;1967.

   Continuum Physics; Academic Press; 1976.

6. Fung, Y.G.; The Foundations of Solid Mechanics ; Prentice-Hall; 1965.

7. Guton, M.C., H.I. Rosten, D.B. Spaulding & D.G. Tatchell; PHOENICS- An Instruction Manual; CHAM;London; 1983.

   Supplements for AUTOPLOT & BFC; 1985

6. Hill, R. & J.W. Hutchinson;" Bifurcation Phenomena in the Plane Tension Test"; J. Mech. Phys. Solids; v.23; 1975.

7. Spalding, D.B.; " A General Purpose Computer Program for Multidimensional One and Two- Phase Flow"; Mathematics and Computers in Simulation; North Holland Press;v 23;1981.

8. Truesdell, C.; "The Simplest Rate Theory of Pure Elasticity"; Comm. of Pure and Applied Mathematics; v.8; 1955.

9. Truesdell, C.; Continuum Mechanics I - The Foundations of Elasticity and Fluid Dynamics; Gordon & Breach; 1966.

1985

NASA/ASEE SUMMER FACULTY RESEARCH FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

INVESTIGATION AND MODELING OF SPACE SHUTTLE MAIN ENGINE SHUTDOWN TRANSIENT CHUGGING

Prepared by:	Paul E. George, II
Academic Rank:	Asst. Professor
University and Department:	The University of Tennessee, Knoxville Mechanical and Aerospace Engineering
NASA/MSFC:	
Laboratory:	Structures and Propulsion
Division:	Propulsion
Branch:	Propulsion Analysis
MSFC Counterpart:	Dr. T. Dwayne McCay

Date:	September 6, 1985
Contract No.:	NASA-NGT-01-008-021 (The University of Alabama in Huntsville)

ACKNOWLEDGEMENTS

The assistance and friendship of Carol Dexter, Sam Lowry and Glenn Wilmer are gratefully acknowledged; I learned a lot from discussing the problems they were working on. Rex Bailey answered many of my questions and allowed me to witness two small engine LNG-Lox tests.

Special thanks to Dr. Dwayne McCay, my NASA counterpart, for providing an interesting problem and guidance for its solution.

The financial support of the National Aeronautics and Space Administration (NASA) and the American Society for Engineering Education is gratefully acknowledged.

INVESTIGATION AND MODELING OF SPACE SHUTTLE MAIN ENGINE SHUTDOWN TRANSIENT CHUGGING

BY

PAUL E. GEORGE, II
ASSISTANT PROFESSOR OF MECHANICAL AND AEROSPACE ENGINEERING
THE UNIVERSITY OF TENNESSEE, KNOXVILLE
KNOXVILLE, TENNESSEE

ABSTRACT

The space shuttle main engines experience a low frequency pressure pulsation in both the fuel and oxidizer preburners during the shutdown transient. This pressure pulsation, called chugging, has been linked to undesirable bearing loads and possible damage to the spark ignitor supply piping for the fuel preburner. This report briefly describes the problem then proposes a model which includes: 1) a transient stirred tank reactor model for the combustion chamber, 2) a resistance capacitance model for the supply piping and 3) purge gas/liquid oxygen interface tracking.

INVESTIGATION AND MODELING OF SPACE SHUTTLE MAIN ENGINE SHUTDOWN TRANSIENT CHUGGING

INTRODUCTION

The Space shuttle main engines have successfully completed 20 flights and several hundred test stand firings. They are stable for steady-state and programmed load change firing from minimum power level (~360,000 lbf thrust) to full power level (512,000 lbf). Upon shutdown, the engine oxidizer system is purged with helium prior to cutoff of the fuel. During this purge, the engines experience a low amplitude pressure pulsation in the preburners which is reflected as a slight variation in main combustion chamber pressure. Since the thrust is intentionally reduced to zero, this pulsation has no effect on space shuttle performance; however, the pulsations have been linked to undesirable bearing loads and damage to the augmented spark ignitor oxidizer supply piping.

This report begins with a brief description of main engine operation with emphasis on the shutdown transient. The chugging problem is then discussed along with a summary of the results from a previous investigation of the problem. Finally, a model for the transient behavior which can follow low frequency pressure pulsations in the time domain is proposed along with a discussion of the difficulties which were encountered in attempting to numerically solve the governing equations.

SSME DESCRIPTION

Figure 1 is a propellant flow schematic of the space shuttle main engines showing the major equipment and the propellant flow rates for full power level. Liquid hydrogen enters the engine from the external tank (flight configuration) via a low pressure pump which supplies a high pressure pump with sufficient head to prevent cavitation. The hydrogen leaves the pump at approximately 8000 psia and cools the nozzle, combustion chamber and throat prior to entering the preburners. The preburners partially oxidize the hydrogen to provide power to the turbines which drive the high pressure pumps. Almost all of the hydrogen enters the engine via the preburners. The oxidizer follows a similar path through a low then a high pressure pump; however, oxygen is not used for component cooling and most of the oxygen enters the main combustion chamber directly.

Engine power is controlled by throttling the oxidizer via the preburner oxidizer valves (FPOV and OPOV). The preburners operate at an equivalence ratio of about 8.0; the power available to the turbopumps and hence the reactant flow rate is controlled by the availability of oxygen to the preburners. Under steady conditions the preburners are operating very fuel rich. Oxidizer flow is halted prior to the fuel flow; therefore, combustion in the preburner and in the main chamber is extinguished by shifting the equivalence ratio beyond the rich combustion limit.

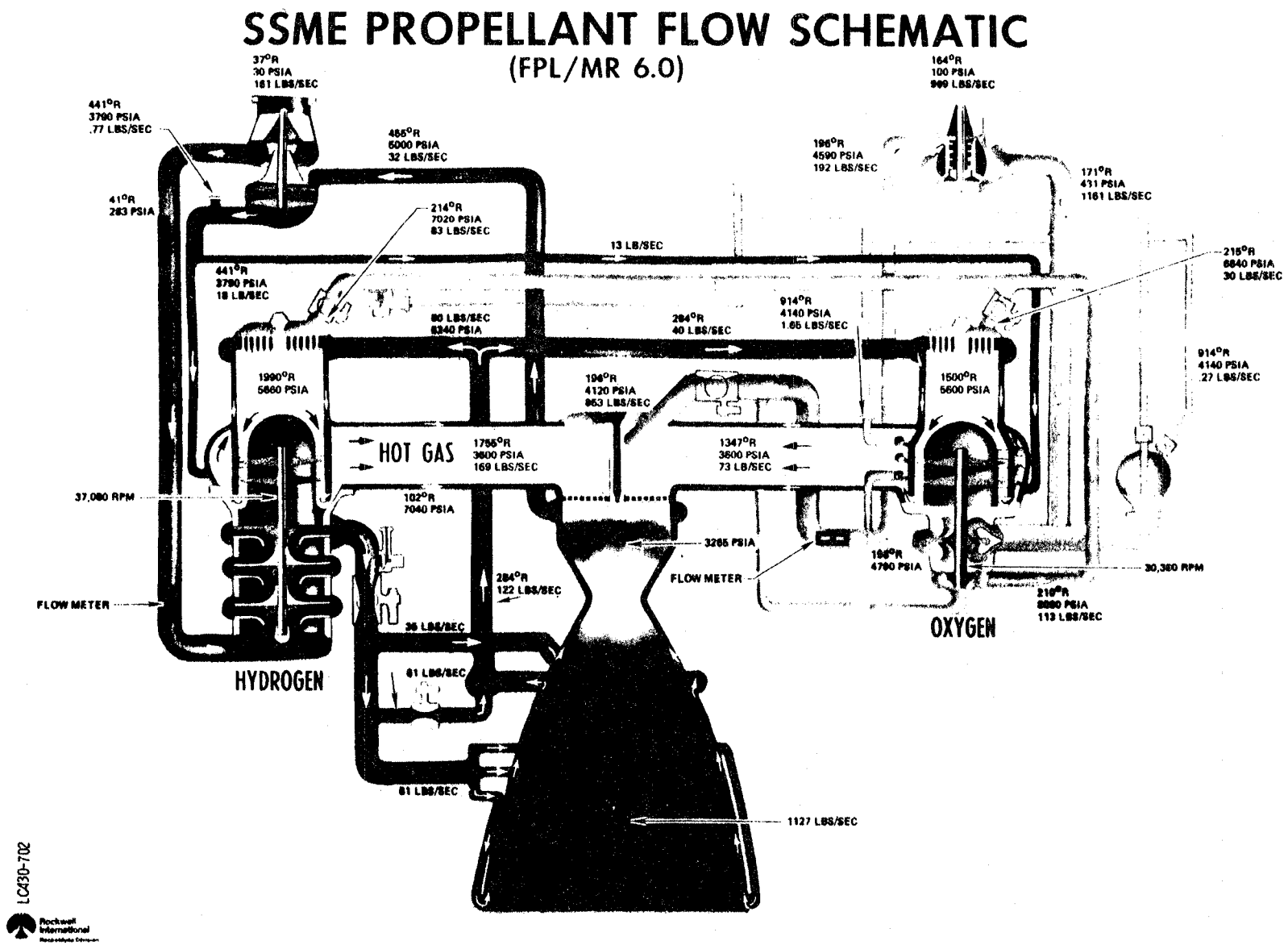


Figure 1: Propellant Flow Schematic

For flight conditions, the engines are typically throttled back prior to shutdown. Throttling is dictated by shuttle maximum acceleration limitations, not engine operation, and ground tests are frequently shutdown from 100% of rated power level or more. Figure 2 shows the valve sequence for a typical shutdown from near 100% of rated power. The oxidizer preburner oxidizer valve (OPOV) is ramped closed first followed shortly by the fuel preburner oxidizer valve (FPOV). Once the two preburner oxidizer valves are closed, the preburners are isolated from the oxygen system; the only oxidizer available to the preburners is the residual trapped between the oxidizer valve and the combustion chamber. This oxygen is expelled into the preburner combustion chamber by a helium purge.

Figure 3 is a detail of the purge and ignitor piping for the fuel preburner. A similar arrangement exists for the oxidizer preburner. As Figure 3 shows, there are check valves in the two helium purge supply lines which remain closed until pressure downstream of the valves has dropped to about 750 psia. These are poppet type valves and remain open even when the pressure drop across the valves is less than the cracking pressure. This inherent hysteresis avoids valve chatter while preventing any reverse flow. Oxygen is cleared from the ASI line and the preburner oxidizer valve into the preburner chamber where it combusts with the fuel. Although the flow rate varies, substantial fuel flow is maintained until after purge is complete.

Pressure pulsations in the fuel preburner, called chugging based on their relatively low frequency ($\frac{1}{4}$200 Hz), begin about 2.3-2.5 seconds after the cutoff command on ground tests. Flight data for chugging were not available; however, conditions are expected to be similar though perhaps more severe. Figure 3 shows the region where chugging is experienced. Although the start of chugging appears to correspond to the closing of the MOV, recall that the preburners are completely isolated from the oxidizer system at the time the valve closes; hence, the closing of the MOV is ruled out as a triggering mechanism.

A previous investigation (George, 1984) concentrated on identifying any possible triggering mechanisms and on determining operating procedures and hardware changes which had affected the chugging characteristics. That study concluded that there was probably not a simple "event" trigger for the chugging; rather, that the system passed through a region where chugging was possible and that natural fluctuations in the system were amplified into the observed chug. The rate of helium flow was found to affect the chug. Changing the helium purge orifice from 0.068" to 0.291" (Figure 3) reduced the amplitude but lengthened the duration of the chug. Although helium flow was increased by this change, the chug ending time was not affected by the increased flow, suggesting that some mechanism other than oxygen depletion is responsible for terminating the chug. Cut-off power was also shown to be significant with lower cutoff levels tending to exacerbate the chugging. For this reason, flight shutdowns are expected to experience more severe chugging.

Typical fuel preburner combustor pressure traces are shown for two tests in Figure 4. Figure 4a is for the small purge orifice. There are three significant factors which should be noted here. First, there is a low

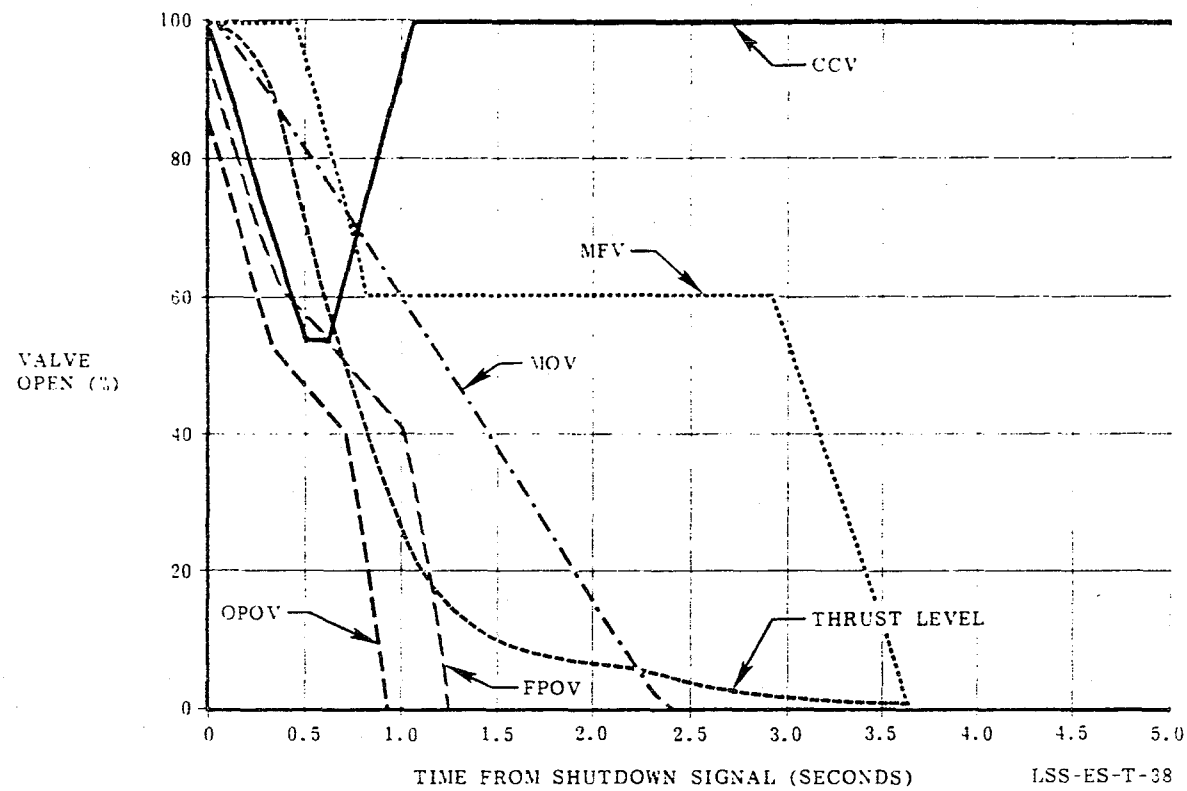


Figure 2. Typical SSME Shutdown Valve Sequence

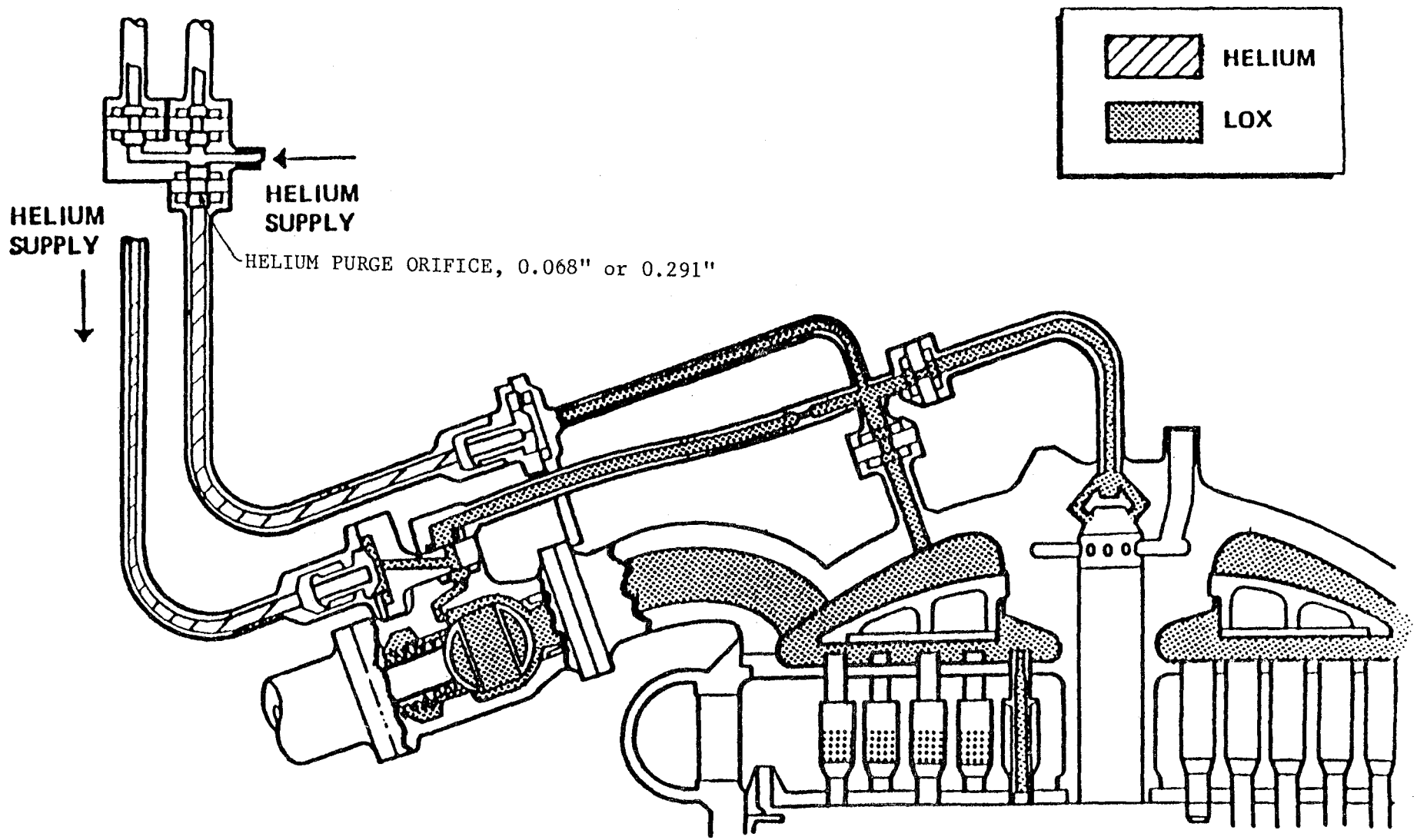
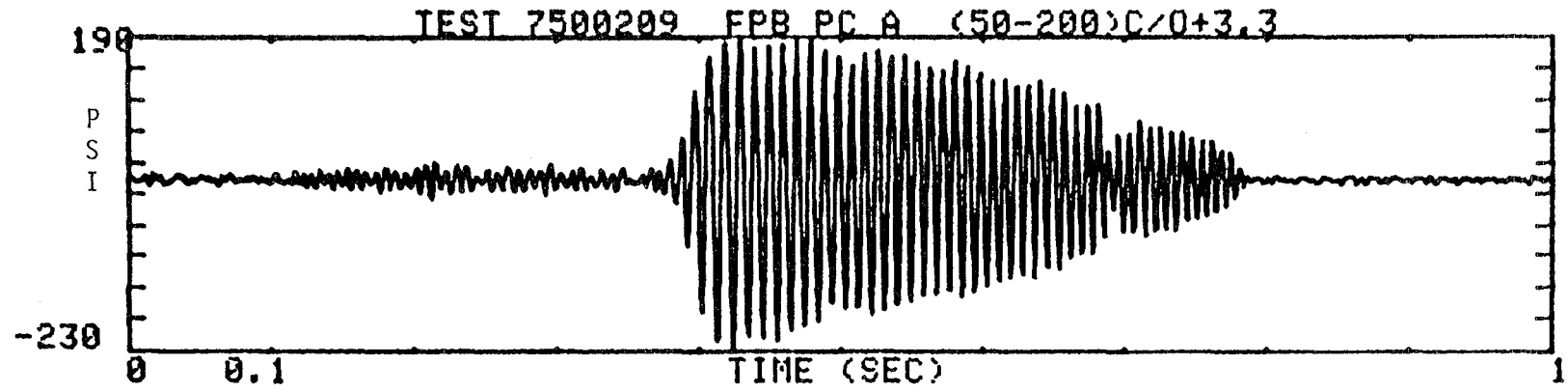
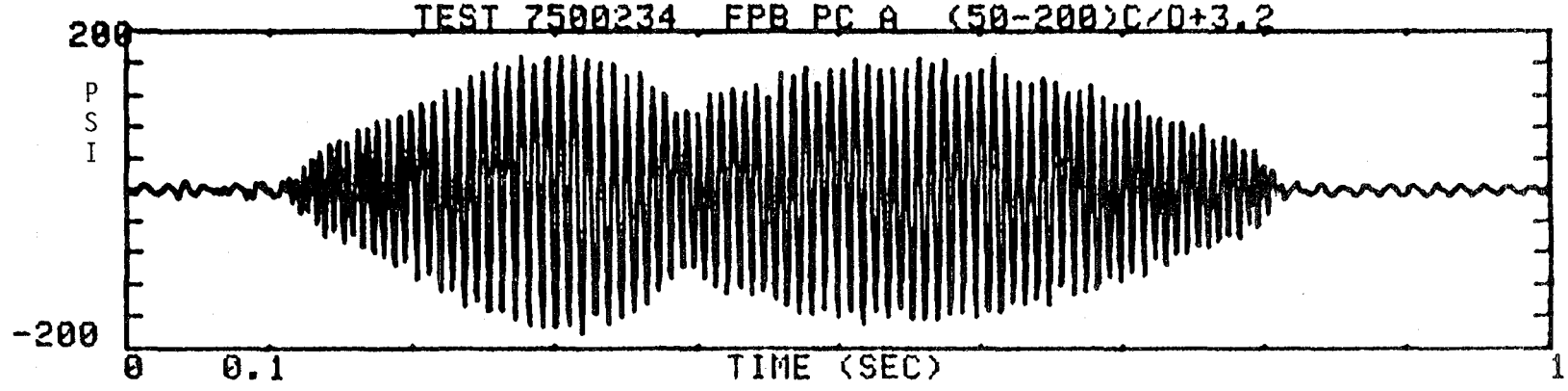


Figure 3. Fuel Preburner Manifold and Purge System Schematic



a. Test 209 using a small purge orifice



b. Test 234 using a large purge orifice

Figure 4: Typical Fuel Preburner Pressure Traces

amplitude chugging which occurs prior to the main chug in Figure 4a. In Figure 4b that chug has apparently been exacerabated by the higher purge flow rates available with the large orifice. Second, the maximum amplitude of the chug was reduced by the large orifice and the shape became somewhat smoother with an amplitude waist occuring at about the same time as the start of chugging with the small orifice. Finally, the chug frequency appears to decrease with amplitude. This is somewhat subjective when viewing traces in Figure 4 but has been confirmed by Fourier transform analysis of small sections of the trace. Clearly there is a non-linear coupling between frequency and amplitude and further, there is apparently an amplitude limiting mechanism since no chug has been observed to grow catastrophically.

MODELING OBJECTIVES

Combustion instability, particularly as applied to rocket engines, has been the subject of several previous investigations. Two of the more comprehensive are Crocco and Cheng (1956) and Harrje and Reardon (1972). The objectives of the previous studies were to identify conditions which should be avoided in rocket engine construction. The models developed were linearized and depended on a classical Eigenvalue analysis to identify any characteristic roots with non-negative real parts. These models have shown that chugging is critically sensitive to combustion time delay and to low injector pressure drop. Compressibility in the feed system is also considered to be a contributor to chugging; however, the actual requirement is that reactant feed rate be proportional in some way to the combustion chamber pressure. Linearized models are extremely useful but limited to defining conditions where chugging can occur while providing no information on amplitude limitations or non-linear instability. There is frequently significant uncertainty in the values to be used for such critical factors as combustion time delay.

The objectives of this study are 1) to provide a tool for the study of purge condition changes, most notably changes in the purge system orifices. 2) to identify any conditions traversed by the SSME during the shutdown purge which are conducive to chugging and to relate these to the onset or end of chugging. 3) to study amplitude/frequency coupling and amplitude limitations with the hope of utilizing these phenomena to limit the scope or intensity of the chug. The first two of these objectives can be met with existing models with the limitation that these models are subject to uncertainty in the fitting parameters used, most notably combustion time delays. The final objective can only be met with a non-linear model solved in the time (as opposed to the frequency) domain. There are secondary considerations which make development of a new time domain model desirable. First, in the process of defining the governing equations and seeking solutions, physical understanding of the problem can be gained. Second, a more fundamental model which does not depend as explicitly on the fitting parameters can be developed. These considerations are offset by the time required for model, particularly computer program, development.

Both modeling approaches are being followed. A chug analysis program has been requested from COSMIC (Szuch, 1971) and will be implemented when received. The program had not been received as of the end of the

summer program. A new model capable of satisfying all objectives was formulated and will be presented below.

MODEL DESCRIPTION

The combustion chamber and the feed system are coupled in a critical way when chugging occurs, however, it is desirable to separate the two when devising a model since different governing equations are required for each. The prime phenomena of interest in the combustion chamber is the release of energy with an attendant pressure rise. The primary interest in the feed system is the mass flow of reactants into the combustion chamber. The combustion chamber pressure serves as the coupling interface between the two systems.

Combustion Chamber

The fuel and oxidizer are admitted to the combustion chamber through an array of closely spaced concentric orifices. In the SSME fuel preburner, 264 injectors are used with fuel occupying an annular region surrounding the oxygen. Because the reactant jets are closely spaced, there is a dominant flow direction from the injector face to the nozzle; however, there is also a significant amount of turbulent mixing, both parallel and perpendicular to the flow direction. In addition, there are recirculation zones between the jets in the upper portion of the chamber.

Previous models have considerd the combustor to be a series of parallel stream tubes each behaving like an independent plug flow reactor. That approach ignores the importance of cross stream mixing and recirculation. At the other extreme of the simple chemical reactor models lies the continuous stirred tank reactor (CSTR) which assumes complete and random mixing of reactants with the entire contents of the vessel; any dominant flow direction is ignored. CSTR's are not usually applied to rocket combustion due to the perceived importance of the dominant flow direction; however, they offer the advantage that the behavior of the combustion chamber can be represented by a single volume without the necessity of spacial integration. There is also a mathematical advantage in that the governing equations are, at least at steady state, algebraic rather than differential equations. In view of the objective of determining the chemically controlled time delay between reactant introduction and pressure increase, a transient CSTR (TSTR) was chosen for this model. While the details of the combustor flow are obviously ignored, this model is still substantially superior to a constant time delay parameter in that the appropriate kinetic parameters may be determined from laboratory experiments and are in fact relatively well known for the H2 - O2 system. A substantial uncertainty enters through the kinetics of droplet dispersion. Since both oxidizer and fuel enter at supercritical conditions, no fundamental droplet dispersion model is available.

There are NS plus two governing equations for a CSTR undergoing transients where NS represents the number of species in the reactor. Conservation of species provides NS equations for the species

concentrations or mole numbers (kgmoles i/kg total) while the energy equation constrains the temperature and conservation of total mass constrains the pressure (or equivalently the density). The species conservation equations are of the form:

$$\dot{m}_f \sigma_{if} + \dot{m}_o \sigma_{io} + \dot{m}_p \sigma_{ip} - \dot{m} \sigma_i + R_i V = \frac{d}{dt}(\sigma_i \rho V) \qquad i = 1, NS \tag{1}$$

where σ_i represents the mole number of species i, R the net reaction rate, V the reactor volume, m the mass flow rate and ρ the density. Subscripts f, o and p refer to the fuel, oxidizer and purge respectively. Unsubscripted variables refer to conditions in the reactor which are the same as the exit conditions. The energy equation is:

$$\sum_{i=1}^{NS} [\dot{m}_f \sigma_{if} h_{if} + \dot{m}_o \sigma_{io} h_{io} + \dot{m}_p \sigma_{ip} h_{ip} - \dot{m} \sigma_i h_i] - Q =$$

$$\frac{d}{dt}\left[\sum_{i=1}^{NS} (\sigma_i u_i \rho V)\right] \tag{2}$$

where we have introduced the additional variables h and u representing enthalpy and internal energy respectively. Q is the net heat loss from the reactor and is currently assumed to be a function of temperature (T) only. Conservation of mass is given by:

$$\dot{m}_f + \dot{m}_o + \dot{m}_p - \dot{m} = \frac{d}{dt}(\rho V) \tag{3}$$

The combustor pressure is determined via the equation of state (ideal gas assumed) and the computed temperature and density.

By expanding the multiple derivatives and combining equations the primary variables may be isolated as simple derivatives with respect to time. This yields a usable form of the equations:

$$\frac{d\sigma_i}{dt} = \frac{1}{\rho}\left(m_{vf}\sigma_{if} + m_{vo}\sigma_{io} + m_{vp}\sigma_{ip} - (m_{vf} + m_{vo} + m_{vp})\sigma_i + R_i\right) \qquad i = 1, NS \tag{4}$$

$$\frac{dT}{dt} = \Big\{ \sum_{i=1}^{NS} (m_{vf}\sigma_{if}h_{if} + m_{vo}\sigma_{io}h_{io} + m_{vp}\sigma_{ip}h_{ip} - m_v\sigma_i h_i) - \frac{Q}{V} - (m_{vf} + m_{vo} + m_{vp} - m_v)\left(\sum_{i=1}^{NS}\sigma_i u_i\right) + \sum_{i=1}^{NS} u_i (m_{vf}\sigma_{if} + m_{vo}\sigma_{io} + m_{vp}\sigma_{ip} - (m_{vf} + m_{vo} + m_{vp})\sigma_i + R_i)\Big\} \Big/ \Big\{\rho \sum_{i=1}^{NS}\sigma_i (C_p - Ru)\Big\} \tag{5}$$

$$\frac{d\rho}{dt} = m_{vf} + m_{vo} + m_{vp} - m_v \tag{6}$$

where Mv is the throughput (mass flow rate/unit volume), Ru the universal gas constant and Cp the constant pressure specific heat.

Equations 4 through 6, along with appropriate expressions for the chemical kinetics and properties can be coupled directly to the feed system equations which supply the required mass flow rates. The

combustor also requires an expression for the exit flow as a function of combustor temperature and pressure. Neglecting turbine pumping contributions, this is given by:

$$\dot{m} = m_v V = C\,(T_{cc}\,P_{cc})\left[\left(\frac{P_{HG}}{P_{cc}}\right)^{1.4286} - \left(\frac{P_{HG}}{P_{cc}}\right)^{1.7143}\right]^{1/2} \qquad (7)$$

where C is an empirically determined constant; P and T are pressure and temperature. Subscripts CC and HG represent the combustion chamber and hot gas manifold respectively (Nguyen, 1981).

Equations 4 through 6 were not used as shown but were converted to integrate the natural logarithm of the variables. Since none of the variables may be negative, but species mole numbers may be very small, use of the log variables ensures that no negatives will be encountered while normalizing the magnitudes of the derivatives. This procedure has been found to be very beneficial in the solution of equilibrium and steady state kinetic problems (Gordon and McBride, 1976; Pratt and Wormeck, 1976). Case and Pratt (1977) recommend it for spacial and time integrations. Conversion to logarithimic variables is easily accomplished by noting that:

$$\frac{d\ln x}{dt} = \frac{1}{x}\frac{dx}{dt} \qquad (8)$$

e.g.

$$\frac{d\ln\rho}{dt} = \frac{1}{\rho}\left(m_{vf} + m_{vo} + m_{vp} - m_v\right) \qquad (9)$$

Supply System Equations

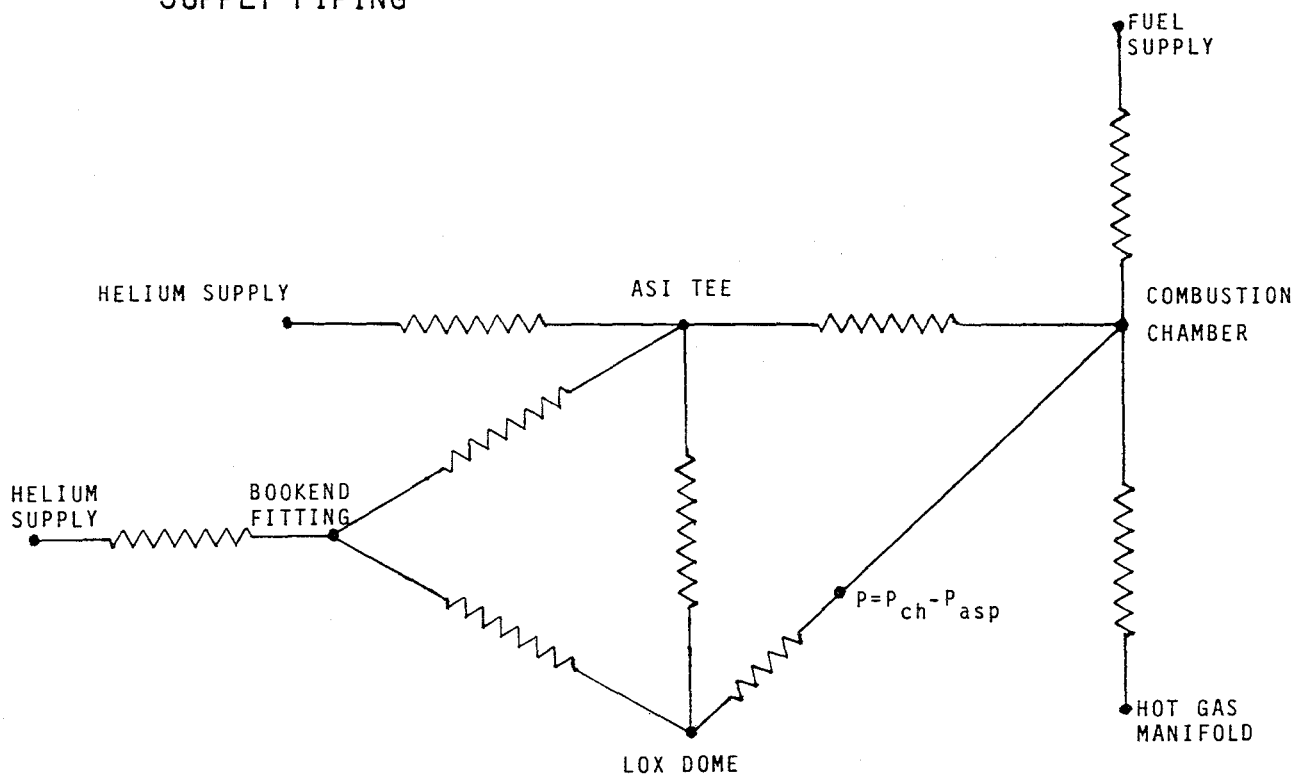
The supply system to be modeled consists of the Lox lines, including the ASI lines shown in Figure 3. Figure 5 shows the reduced resistance network used to represent the system for this model. The boundaries of the system are assumed to be constant pressure helium at the check valves and the combustion chamber pressure calculated above. Conservation of momentum is required in each pipe and conservation of mass is required for each node, i.e., no storage at the node. The gas/fluid interface is followed through the system since line pressure drop depends strongly on the fluid properties and it is necessary to know when the oxidizer has been exhausted. The gas/fluid interface is assumed to be perpendicular to the pipe centerline at all times.

Conservation of momentum for a general one dimensional case can be written:

$$\Sigma F = \frac{d}{dt}\iiint_{CV} \vec{U}\,\rho\,dV + \iint_{CS} \vec{U}\,\rho\,(\vec{U}\cdot\vec{n})\,dA \qquad (10)$$

where U is the velocity, A the cross sectional area and F the external forces acting on the fluid. For this case, the external forces are considered to consist of only the pressure drop and the pipe frictional loss. For a single inlet, single outlet system with uniform velocity profiles this reduces to:

Supply Piping



Bookend volume attributed to Lox Valve line

Lox Dome Volume attributed to injector

A.S.I. Tee - Volume neglected

Figure 5: Supply Piping Resistance Diagram

$$\left[P_1 - P_2 - f\left(\frac{\dot{m}^2}{\rho}\right)\right]\frac{1}{A} = \frac{d}{dt}\left(\rho_1 V U + \rho_2 V U\right) + \rho_2 U^2 A - \rho_1 U^2 A \tag{11}$$

where f is the friction factor and subscripts 1 and 2 refer to the two ends of the pipe. If a single phase, incompressible flow is considered, this further simplifies to:

$$P_1 - P_2 - f\frac{\dot{m}^2}{\rho} = \frac{1}{A}\frac{d}{dt}(\rho V U) \tag{12}$$

which is equivalent to the form proposed by Harrje and Reardon (1972). Such a reduction is not appropriate here since the density of the purge gas will be significantly different from that of the Lox; hence the two ends of the pipe must be retained. Note that the second term in Equation 11 is equivalent to using a weighted average density and a single characteristic velocity.

In order to track the interface location, one additional equation is needed for each pipe with an interface. Conservation of mass for each pipe provides this equation. Conservation of mass at each node provides closure and ensures proper coupling of the feed pipes. If we assume a single characteristic velocity for the pipe equal to the interface velocity (equivalent to assuming that each fluid is incompressible), then conservation of mass requires:

$$\frac{dV_g}{dt} = -\frac{dV_e}{dt} = \frac{\dot{m}_g}{\rho_g} = -\frac{\dot{m}_e}{\rho_e} = A\frac{d\ell_g}{dt} = -A\frac{d\ell_e}{dt} \tag{13}$$

At each node we also have:

$$\frac{d}{dt}\left(\sum_{j=11}^{NP}\dot{m}_{kj}\right) = 0 \qquad k = 1, NN \tag{14}$$

where V again represents volume, A is a characteristic area for each pipe, l the length along the pipe axis, NP the total number of pipes and NN the number of nodes. Subscripts g and l refer to the gas and liquid volumes respectively. As before, these equations may be combined to yield working equations; however, it is not possible to completely isolate the derivatives as was done for the combustion chamber. Non-dimensionalizing lg with the total pipe length and combining equations where possible yield:

$$\frac{d\ell_{gj}}{dt} = U_j \qquad j = 1, NP \tag{15}$$

$$U_j\left(\ell_{gj}\frac{d\rho_{1j}}{dt} + \ell_{\ell j}\frac{d\rho_{2j}}{dt}\right) + \left(\rho_{1j}\ell_{gj} + \rho_2\ell_{\ell j}\right)\frac{dU_j}{dt} = P_{1j} - P_{2j} - f\bar{\rho}_j U_j^2$$

$$j = 1, NP \tag{16}$$

where $\bar{\rho} = \frac{1}{V_j}\left(V_{gj}\rho_1 + V_{\ell j}\rho_2\right)$

$$\frac{d\rho_k}{dt}\sum_{j=1}^{NP}A_j U_{kj} + \sum_{j=1}^{NP}\rho_k A_j\frac{dU_j}{dt} = 0 \tag{17}$$

These are the final forms of the working equations. No conversion to logarithmic variables was undertaken since U may be positive or negative. Equations 16 and 17 are coupled and must be solved simultaneously for values of the first derivatives. In the computer program this is accomplished using Gauss iteration.

NUMERICAL METHODS

The governing equations derived above must be solved numerically in the time domain to determine the various species mole numbers, temperatures, pressures and flow rates as functions of time. Flow reversals and pressure oscillations are allowed. Auxilliary relations are used as needed to convert from primary variables to secondary variables, e.g. from temperature and density to pressure.

The solution of these equations is not straightfoward. The chemical kinetic equations are stiff; that is, they have widely varying characteristic times. When coupled with the inlet feed system, the problem is exacerbated. First order numerical techniques will not work because extremely small time steps are required to avoid instability and roundoff error becomes too large when small time steps are used. Higher order single step methods such as Runge-Kutta offer some hope for success but are generally prohibitively expensive due to the time required for multiple evaluations of the derivatives. Based on experience with this program, it appears that Runge-Kutta integration schemes are not appropriate for this equation set due to instability associated with single step explicit methods.

Because of the stiffness, these equations should be solved with a multi-step implicit integrator (Gerald, 1969). The most popular method is that proposed by Gear (1971) and modified by Hindmarsh (1974). A commercial version of the Hindmarsh subroutine is available through the Internation Mathematical and Scientific Library (IMSL). Unfortunately, that library was not implemented on the Sperry, Perkin-Elmer or VAX computers available for solving the problem at NASA, Marshall. A fourth order Runge-Kutta integrator was written to assist in debugging the program at NASA with hopes that the program could be transported to a machine with IMSL available. Success with the Runge-Kutta routine was not achieved due to apparent numerical difficulties, i.e. excessive rates of change of the primary variables, usually the chemical species mole numbers. Plans were made to transport the program to the University of Alabama, Huntsville computer where IMSL is available; however, transfer was not accomplished during the 10-week summer program.

PROGRAM DESCRIPTION

Table 1 provides a summary of the computer program (TRNCHG) and its associated subroutines. The equations presented above are programmed in subroutines CHGSLP and RATES. Subroutine DGEARX is the Runge-Kutta

TABLE 1

PROGRAM ROUTINE DESCRIPTION

ROUTINE	DESCRIPTION
TRNCHG	Main Program. Controls program flow and all I/O except debugging output which is local to the various subroutines.
BLKDTA	Block Data subroutine to set default values.
CHGSLP	Computes the first derivatives of all primary variables for use by the integrating subroutine. Contains the programming for the governing equations. Calls RATES to set the net rate of species production by chemical reaction.
DGEARX	Runge-Kutta integration subroutine which answers a call mimicking a call to IMSL subroutine DGEAR.
HCPS	Computes Enthalpy, Internal Energy and specific heat using the NASA "ODE" polynomial form: $H = R*T*(Z1 + \frac{Z2*T}{2} + \frac{Z3*T**2}{3} + \frac{Z4*T**3}{4} + \frac{Z5*T**4}{5})$.
NPRNT	Prints Reactant Streams.
NPT	Main input subroutine. Reads all thermochemical, kinetic and reactants data.
OUTPT	Prints final or intermediate output.
RATES	Computes net rate of species production due to chemical reaction and derivatives with respect to species mole numbers and temperature.

integrator subroutine. The remaining subroutines are associated with data input and solution output or thermodynamic properties. The program listing (>2000 lines) is not reproduced here for obvious reasons, not the least of which is that the program does not work.

As noted above, successful operation of the model was not achieved in the time available. A substantial effort was expended in setting up the proper input files and in I/O overhead. All of the I/O-setup problems seem to have been solved so that the work remaining is to solve the numerical difficulties. This may only require use of a Gear integration subroutine or it may be necessary to recouch the equations in a different form less susceptable to instability. Work on this problem will be continued at the University of Tennessee, Knoxville with support from NASA. Once operational, the program will be implemented on the NASA computers. The IMSL package will hopefully be available by that time.

CONCLUSIONS

Although a successful model was not achieved, the problem has been defined and much of the set-up work has been completed. A study of the shutdown chugging has shown that non-linear effects are important to the chugging amplitude and that equipment modifications can have an effect on chugging. A literature survey concerning methods which could be used for numerically solving the governing equations has suggested that implicit methods, most notably Gear's method, are available and that the equations can be solved without linearizing.

This summer has been most educational and enjoyable in spite of the frustration of not completing the model. I was afforded the opportunity to work with several outstanding engineers on other problems not mentioned here and to exchange ideas with NASA personnel and other summer faculty. The summer faculty fellowship program has been an outstanding opportunity to broaden my engineering background.

1985

NASA/ASEE SUMMER FACULTY RESEARCH FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

A STUDY OF THE VERY HIGH ORDER NATURAL USER LANGUAGE (WITH AI CAPABILITIES) FOR THE NASA SPACE STATION COMMON MODULE

Prepared By:	Esther N. Gill, Ed.D.
Academic Rank:	Associate Professor
University and Department:	Oakwood College Business and Information Systems Management
NASA/MSFC: Division: Branch:	 Software & Data Management Systems Software
MSFC Counterparts:	John W. Wolfsberger and Robert L. Stevens
Date:	August 2, 1985
Contract No.:	NGT 01-008-021 The University of Alabama in Huntsville

A STUDY OF THE VERY HIGH ORDER NATURAL USER LANGUAGE (WITH AI CAPABILITIES) FOR THE NASA SPACE STATION COMMON MODULE

BY

Esther N. Gill, Ed.D.
Associate Professor of Computer Science
Oakwood College
Huntsville, Alabama 35896

ABSTRACT

This study will identify the requirements for a very high order natural language to be used by crew members on board the Space Station and will be a part of each module's common computer core operating system. The study will take into consideration the hardware facilities, databases, real-time processes, software support, etc. available or selected and will evaluate the operations and capabilities that will be required in both normal (routine) and abnormal (nonroutine) situations. The study will recommend a structure and syntax for an interface (front-end) language to satisfy the above requirements.

ACKNOWLEDGEMENTS

This researcher would like to acknowledge the contributions of her counterparts, Mr. John W. Wolfsberger and Mr. R. L. Stevens, whose aide was invaluable in obtaining source documents and other resources on the subject explored. Mr. Wolfsberger's enthusiasm for his work also motivated and stimulated this effort. I would also like to thank Mr. Dave Aichele, Mr. William Bradford, Mr. Jack Lucas, and Mr. Gabe Wallace for the privilege of entrance afforded me to the Software Evaluation Lab which made this research project possible. Thank you Mr. Jimmy Watkins for making my working surroundings more comfortable and for proofreading this paper.

NASA/ASEE Program Chairpersons, Mr. Thomas L. Osborn, Dr. James Dozier, and Dr. Gerald R. Karr, also are commended for their time and effort expended coordinating and administering the rewarding research experiences of the summer fellows. Thank you also Mrs. Dina Conrad for your personal helpful flare.

To my research colleague, Dr. Arthur Knoebel, gracias for making my first experience as a NASA/ASEE fellow a shared office pleasure.

I have gained a wealth of knowledge which will be taken back to my classroom and used for enriching the experiences of my students, for future proposal presentations to NASA for possible approval and funding, and for future NASA-employment, college-recruitment efforts.

Dr. Esther Naomi Gill

A STUDY OF THE VERY HIGH ORDER NATURAL USER LANGUAGE
(WITH AI CAPABILITIES)
FOR THE NASA SPACE STATION COMMON MODULE

T A B L E O F C O N T E N T S

GLOSSARY OF TERMS USED

AI HANDBOOK: The Handbook of Artificial Intelligence, E.A. Feigenbaum, A. Barr and P.R. Cohen (Eds.). Published by W. Kaufmann, Los Altos, CA in 1981 and 1982.

ARTIFICIAL INTELLIGENCE (AI) APPROACH: An approach that has its emphasis on symbolic processes for representing and manipulating knowledge in a problem solving mode.

BACKWARD CHAINING: A form of reasoning starting with a goal and recursively chaining backwards to its antecedent goals or states by applying applicable operators until an appropriate earlier state is reached or the system backtracks. This is a form of depth-first search. When the application of operators changes a single goal or state into multiple goals or states, the approach is referred to as problem reduction.

COMMON SENSE: The ability to act appropriately in everyday situations based on one's lifetime accumulation of experiential knowledge.

COMMON SENSE REASONING: Low-level reasoning based on a wealth of experience.

COMPUTER GRAPHICS: Visual representations generated by a computer (usually observed on a monitoring screen).

COMPUTER VISION (COMPUTATIONAL OR MACHINE VISION): Perception by a computer, based on visual sernsory input, in which a symbolic description is developed of a scene depicted in an image. It is often a knowledge-based expectation-guided process that uses models to interpret sensory data. Used somewhat synonymously with image understanding and scene analysis.

CONCEPTUAL DEPENDENCY: An approach to natural language understanding in which sentences are translated into basic concepts expressed as a small set of semantic primatives.

DATA BASE: An organized collection of data about some subject.

DATA BASE MANAGEMENT SYSTEM: A computer system for the storage and retrieval of information about some domain.

DATA STRUCTURE: The form in which data is stored in a computer.

DOMAIN: The problem area of interest.

EXPERT SYSTEM: A computer program that uses knowledge and reasoning techniques to solve problems normally requiring the abilities of human experts. Can capture knowledge of experienced engineers before they leave a firm, reclaiming years of "learning by doing" that can then be passed on to new engineers through the system.

FIFTH GENERATION COMPUTER: A non-Von Neumann, intelligent, parallel processing form of computer now being pursued by Japan.

FORWARD CHAINING: Event-driven or data-driven reasoning.

HEURISTICS: Rules of thumb or empirical knowledge used to help guide a problem solution.

HIGHER ORDER LANGUAGE (HOL): A computer language (such as FORTRAN or LISP) requiring fewer steatements than machine language and usually substantially easier to use and read.

IMAGE UNDERSTANDING (IU): Visual perception by a computer employing geometric modeling and the AI techniques of knowledge representation and cognitive processing to develop scene interpretations from image data. IU has dealt extensively with three-dimensional objects.

INFERENCE: The process of reaching a conclusion based on an initial set of propositions, the truths of which are known or assumed.

INFERENCE ENGINE: Another name given to the control structure of an AI problem solver in which the control is separate from the knowledge.

INTELLIGENT ASSISTANT: An AI computer program (usually an expert system) that aids a person in the performance of a task.

INTERACTIVE ENVIRONMENTS: A computational system in which the user interacts (dialogues) with the system (in real time) during the process of developing or running a computer program.

INTERFACE: The system by which the user interacts with the computer. In general, the junction between two components.

KNOWLEDGE BASE: AI data bases that are not merely files of uniform content, but are collections of facts, inferences, and procedures, corresponding to the types of information needed for problem solution.

KNOWLEDGE ENGINEERING: The AI approach focusing on the use of knowledge (e.g., as in expert systems) to solve problems.

NATURAL LANGUAGE INFERENCE (NLI): A system for communicating with a computer by using a natural language.

NATURAL LANGUAGE PROCESSING (NLP): Processing of natural language (e.g., English) by a computer to facilitate communication with the computer or for other purposes such as language translation.

NATURAL LANGAUGE UNDERSTANDING (NLU): Response by a computer based on the meaning of a natural language input.

PERSONAL AI COMPUTER: New, small, interactive, stand-alone computers for use by an AI researcher in developing AI programs. Usually specifically designed to run an AI language such as LISP.

PROGRAMMING IN LOGIC (PROLOG): A logic-oriented AI language developed in France and popular in Europe and Japan.

SEMANTIC: Of or relating to meaning.

SOFTWARE: A computer program.

SPEECH RECOGNITION: Recognition by a computer (primarily by pattern matching) of spoken words or sentences.

SPEECH SYNTHESIS: Developing spoken speech from text or other representations.

SPEECH UNDERSTANDING: Speech perception by a computer.

SYMBOLICS: Relating to the substitution of abstract representations (symbols) for concrete objects.

SYNTAX: The order or arrangement (e,.g., the grammar of a language).

VON NEUMANN ARCHITECTURE: The current standard computer architecture that uses sequential processing.

LANGUAGES

CATEGORIES	CANDIDATES?
Requirements and specification	PSL/PSA, SREM, SADT, CADSAT
Design	PDL, SDDL
Development	HAL/S, Fortran, PL/I, Jovial, Ada, C, Modula-2, Pascal
User interface	GOAL, ATLAS, SCOL,STOL, Ada
AI/expert systems	LISP, PROLOG

ILLUSTRATION 2

PROGRAM SUCCESS DEPENDS ON SOFTWARE

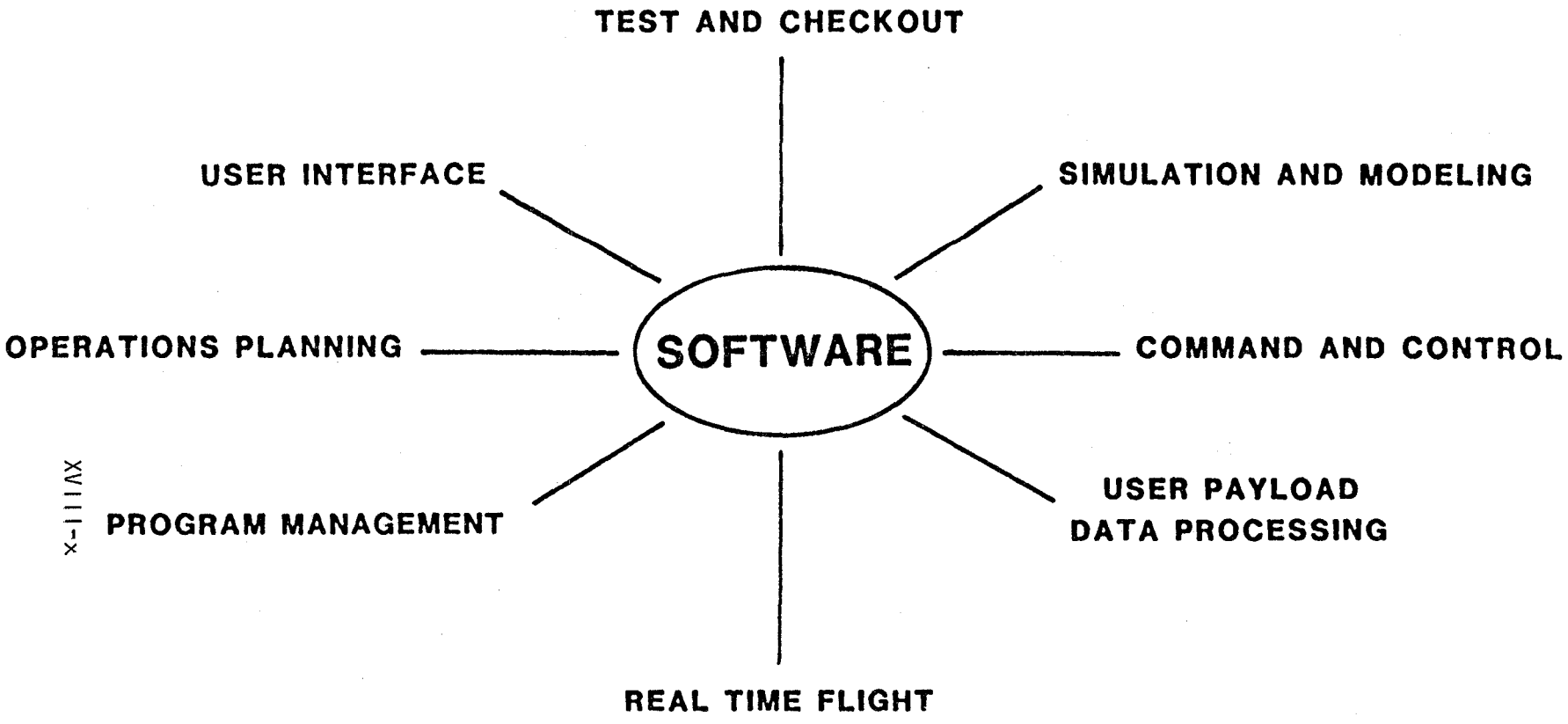


ILLUSTRATION 3

NASA-S-84-01364

TECHNOLOGY

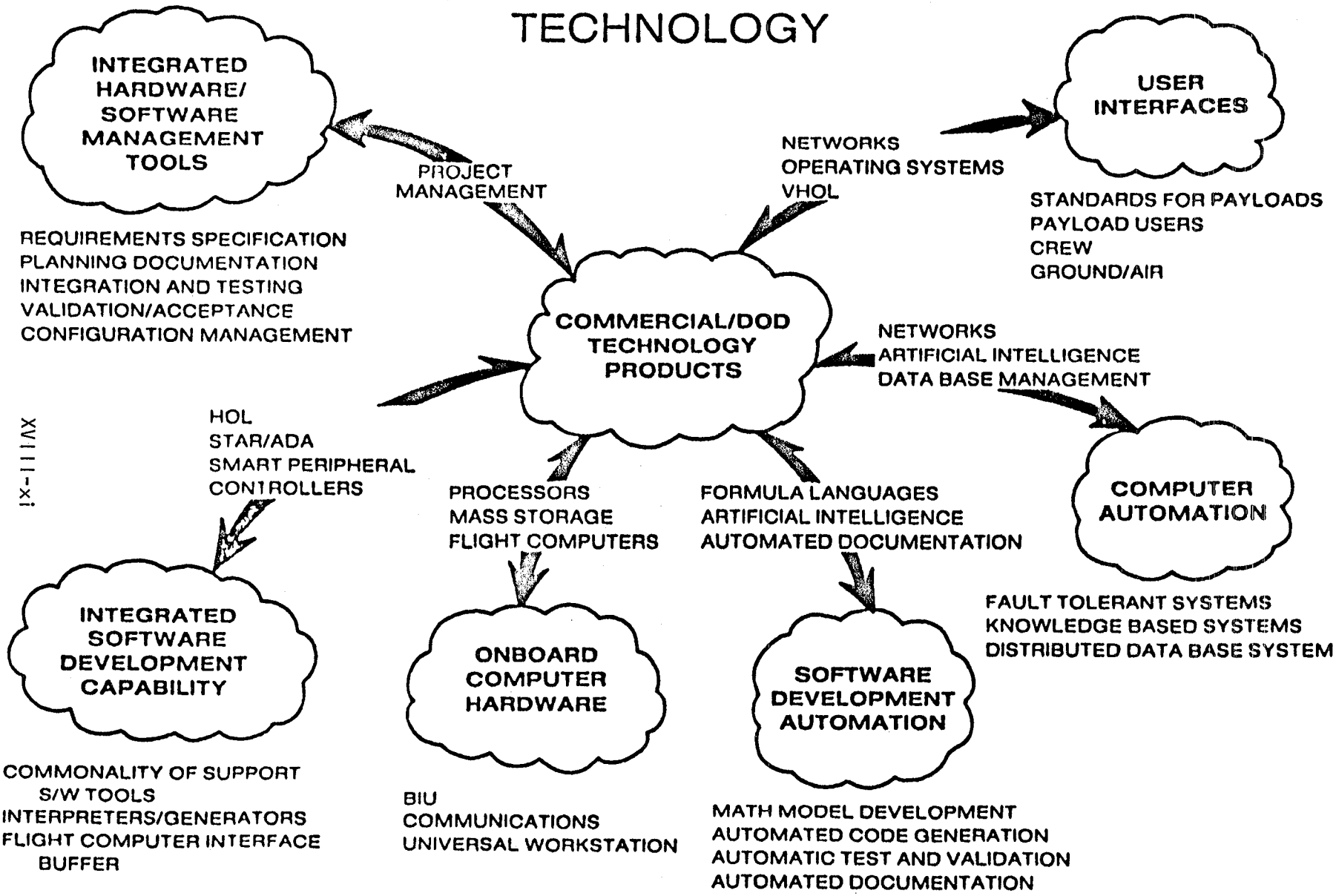


ILLUSTRATION 4

5.0 INTRODUCTION

5.1 Purpose Statement

The SPACE STATION's design technology (deployment) is the next giant step in the coming decade of the United States' space program. Within the design and technology of the SPACE STATION will be computer capabilities which will operate with AI (artificially intelligent) expert systems which could utilize robotics, visual perception, voice synthesis; such as, speech synthesis and speech recognition.

These AI technologies have been predicted by industry analysts to account for 50 percent of all EDP (Electronic Data Processing) by the end of the 1990s--a forward direction for the computer hardware/software industry as well.

5.2 Natural Language Justification

The power of the hardware and computation facilities provided on board the SPACE STATION cannot be realized if there is not a user language available which gives the crew members free and flexible access to those resources.

The language must make full use of the on board facilities (which include the high technology development for on board software capabilities over the Space Station's life cycle) for the anticipated applications, while giving consideration to the possibility of commonality of the user interface with other user languages.

The crew members (previous programming experience not needed) must be able to interrogate the system to obtain answers, to monitor the system to determine performance, to create visuals to display screen dialogue and to trial check simulations, to obtain readouts to indicate status and diagnose failures, to examine databases for update, retrieval, and management, to exert manual control over applications for emergency management or for overriding systems software rationale, etc.

All of the above-mentioned crew-member tasks must be performed with ease without prior preparation, not to mention the other prepared-for and rehearsed crew member tasks.

What is needed is a natural; that is, English, French or Hebrew as opposed to FORTRAN, COBOL, Assembler, or BASIC, language that permits the untrained crew members to interact (access, query, update, etc.) with the underlying system in an extremely productive user-friendly manner. By user friendly is meant conversational, everyday English, no described syntax, etc.

The idea behind AI or logical processing is to make people more productive, and a natural language is perhaps the strickest and most difficult area of AI development.

5.3 Objectives

The SPACE STATION, on ground and in orbit, will require a Very High Order Language (VHOL) to query, produce and monitor performances of ground and on orbit tests.

The VHOL shall be structured with English type requests. The objectives of this research task are

1. Establish MSFC Mission Requirements to determine user language needs.

2. Review with other centers (KSC, JSC, GSFC) their activities in these areas.

3. Determine and design an analysis plan to develop MSFC requirements for VHOL.

4. Survey present VHOL commercial language.

5. Identify in-house capabilities presently available and determine future needs for on-going evaluations.

This research is related to the following MSFC Mission Assignments: SPACE STATION user language for use in test beds and inclusion in module development.

6.0 IDENTIFIED COMMON MODULE COMPUTER TASKS

The very high order natural language to be used by Space Station crew members must be able to interface with prewritten ADA, C, FORTH, FORTRAN, HAL/S, LISP, PL/1, Pascal, PROLOG, ASSEMBLER, etc. program applications.

This user-friendly, interface language will have built in AI applications.

These programs will be required to support such user operations as:

-Monitoring and Reporting Observations
-Examining and Updating System Databases
-Creating and Describing Graphic Displays of Picture Images' Innerconnection
-Monitoring On Going Systems and Making Adjustments Where Needed
-Troubleshooting On Going Processes
-Repairing and Recovering Satellites
-Identifying Extremely Dangerous Power Failures
-Detecting and Diagnosing Equipment Malfunctions
-Examining Realistic and Complex Situations and Making Judgments Based on these Situations
-Interpreting Data
-Testing Hypotheses
-Validating System's Performance
-Examining Programs' Performance
-Analyzing Stress on Internal and External Structures
-Avoiding Space Accidents and Correcting Problems in the Process
-Assessing Problems to Demonstrate Prototypes (Simulations) and Feasibility
-Controlling and Monitoring SPACE STATION's Atmosphere and Life-Support Systems
-Automatically Correcting Present and/or Predictable Failures
-Intervening or Concurring with Automated Machine Actions
-Interacting with Real Time Processes
-Forecasting Potential Conditions and Making Contingency Analyses

7.0 RECOMMENDED USER LANGUAGE STRUCTURES AND SYNTAX

There is justification and needful purposes for the development and/or incorporation of a Natural User Language (NUL) to be utilized by SPACE STATION crew members in accessing, querying, and retrieving information from each module's common computer databases, operating systems, expert systems or conventional applications. Illustration 1 (SPACE STATION Labs Initial Configuration) is Boeing's concept of the location of the computer in the common shell of each module. Note specifically the Workstation (in mockup) area of the illustration. Also of interest is Boeing's commitment to AI (Speech Recognition and Speech Synthesis) as well as Screen and Keyboard Applications.

7.1 Syntax

In order to accomplish the purpose of easy access and retrieval of information thru the system, syntax or order of arrangement (grammar) should be a secondary consideration.

7.2 Semantics

Semantics (meanings of the words identified for use in the user's natural language) would have primary emphasis. The natural user language would concern itself only in a transparent way with syntax by identifying either the spoken, written, or selected part or parts of speech utilized with the parser, interpreter, or translator; i.e., an interrogation and/or readout from one of the on board instruments is made. He/She might say, "I need to know what the attitude control requirements are for a given experimental configuration."

The parser would use its language interpreter to identify the nouns (attitude control), action verbs (need to know), etc. to interface (activate) a program and output the results.

7.3 Structure

SPACE STATION

Space Station Labs Initial Configuration

BOEING

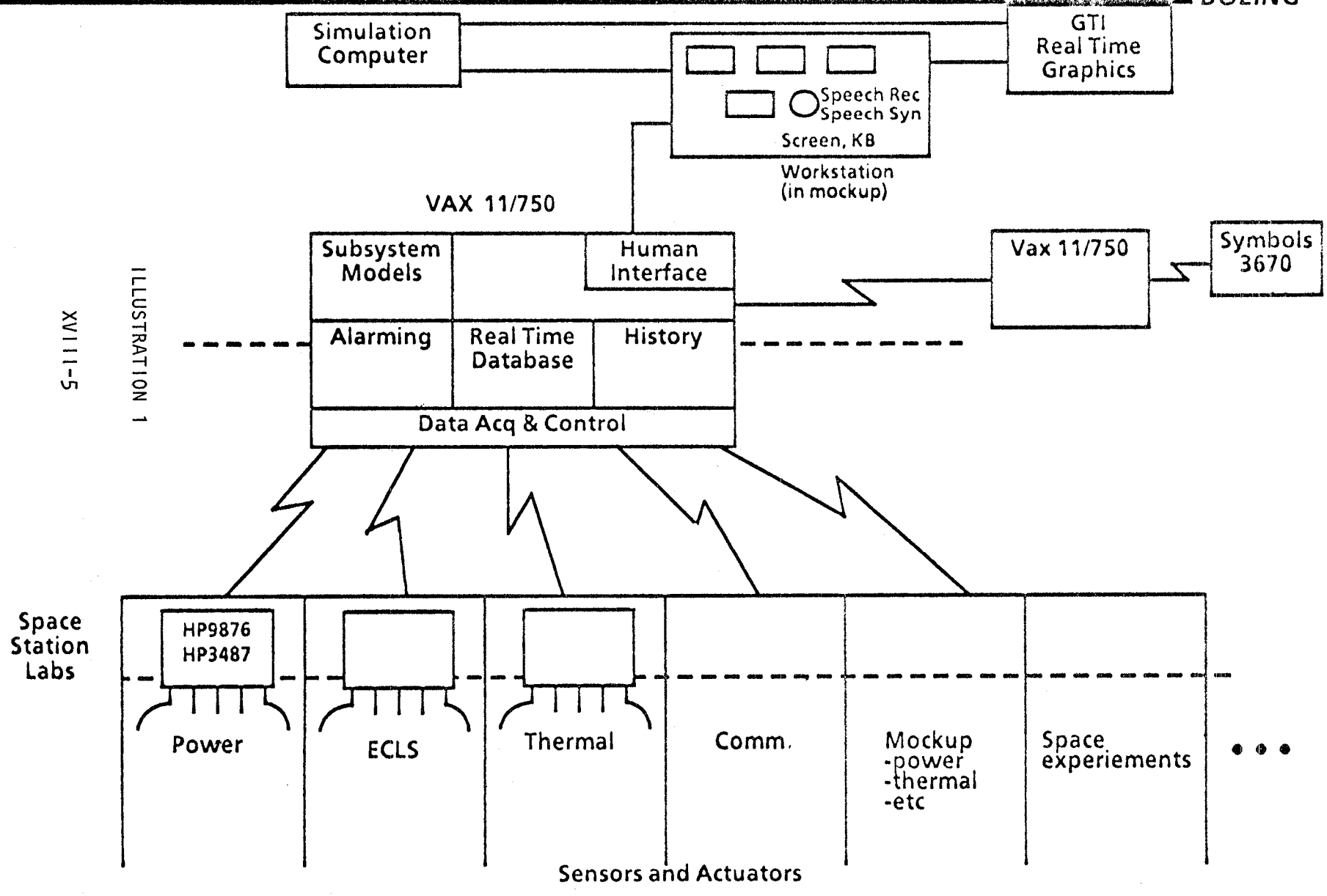


ILLUSTRATION 1

XVIII-5

The structure or form of this natural user language would be made up of prompts (statements from the system describing a specific action one needs to take in order for a task to continue).

7.3.1 Prompts

Prompts give the operator choices from which an appropriate response for the task one wishes to perform is selected.

Menus could be incorporated for use as displays of tasks to be performed or for initial user access to interface with ADA. Menus could display on the screen a list of tasks that the user can perform by choosing an ID letter and/or item description. Letters are used to identify the choices and descriptions or give details about each item.

By choosing from menus which help guide the user, the user tells the system what he/she wants to be done or indicates whether his/her speech synthesis and the computer's speech recognition are synchronized (which varifies and assures these AI processes).

7.3.2 Menus

Menus could be set apart as help menus, command menus, task menus, and procedure menus.

7.3.2.1 Help

Whatever user need arrises, there should be a help, command, or procedure menu to aid the accomplishment of the task or provide a message or statement which tells you about what has or is happening on the system and what response to make to get the appropriate action.

7.3.2.2 Tasks

The possible tasks that could be performed would be limited only by the firmware of the system and would be expandable depending on previous SPACE STATION experiences, payload applications, and scientific and technological advances.

Better still the very high order language (VHOL) Expert System could utilize real-time processing with menus and prompts to interact and execute interrogations, creations, indications, diagnoses, observations, adjustments, controls, description, etc.

7.3.2.3 Commands & Procedures

Task commands could also be placed on a menu with a letter or number code beside each listed task so that a selection could be made by the crew member which would in turn display a procedural menu with specific non-technical directions and/or prompted data entry requests responded to again by the crew member which would be interfaced with the task program for executing, reporting, updating, monitoring, adjusting, etc.

The task command menus would list all possible tasks common to all modules as well as tasks unique to the particular module's purpose. Programs which interface with these tasks could be written in any high order language acceptable or compatible to ADA or with whatever SPACE STATION language NASA selects.

7.3.3 Dialogue

The interaction between the crew members (astronauts) and the common module computer via the VHOL would result in a dialogue between human and computer understandable to both man and machine minus human knowledge base other than native language usage. In other words, it is anticipated that this natural user language be translated into the languages of the world in order that universal users of the SPACE STATION's technology and experimental capabilities might indeed be a cooperative (international) effort thereby advancing our goal of the commercialization of space.

7.4 ARTIFICIAL INTELLIGENCE

Artificial Intelligence (AI) has been defined in many ways. Notable among these definitions is the computer program that is knowledge based (logical) rather than digital in its handling of the data (numeric or text).

7.4.1 Expert Systems

Several programming tasks could be written in AI structure and format. These programs if they use knowledge and reasoning techniques to solve problems which normally require the abilities of human experts would be described as expert systems.

7.4.2 Robotics

Computers which have programmed applications which make manipulative (physical) motions (tasks) would fall into the AI Robotics category. Robotics could be invoked by crew members at night when they are sleeping instead of placing all their systems in the hands of ground control. If human intervention were required, it could alarm the crew members to take appropriate actions. Robotics could also be used on the arm which retrieves the satellites for repairs and maintenance as well as for the repairs made, thus relieving the crew members of this often frustrating, (crew members make human errors which consume time, energy, and fuel and make critical their second try), dangerous (crew members are assured of life for only eight hours when suited up in their 400-pound space suits), and human-expert (required) endeavor. With a working robotics expert system, crew members could successfully invoke the system with the VHOL.

Robotics would be used in the NUL's interface to make adjustments, to interact with real time processing, to intervene and to correct failures, etc.

7.4.3 Speech Recognition

Speech Recognition is described as the ability of the computer to understand the spoken word. Via the NUL (natural user language), the SPACE STATION crew members could give

verbal commands to cause the computer to translate the commands into program activation using speech recognition techniques. Before activating any programs, however, there would have to be varifying techniques built in so that the crew members could varify the computer's understanding of the voice commands.

7.4.4 Speech Synthesis

The computer's ability to develop spoken speech from text or other representations is an idea whose time to be tested has come.

If this AI approach were used for the NUL interface language, the computer's feedback would have to be monitored closely to insure reliability.

7.4.5 Computer Vision

Computers that reason, plan, and perceive based on visual sensory input, in which a symbolic description is developed of a scene depicted in an image, describes computer vision.

On the common module's computer, machine vision, or image understanding might be used in forecasting, tracking, creating and describing various SPACE STATION phenomena with the NUL being the interface to these knowledge-based, expectation-guided processes.

7.5 Functional Requirements

Depending upon the amount of AI utilization by the programmers who've written programs needing solutions as well as the yet-to-be-determined (YTBD) direction (AI or conventional) of the standardization and selection of one user language or the acceptance of a few user languages, specific functional requirements for the NUL will be defined.

If the direction is going toward AI, then certainly the functional requirements of the hardware would change from conventional hardware to hardware with AI capabilities, and the software interface NUL would undoubtedly be upgraded from CRT and keyboard input (menus, commands, prompts and feedback) to AI languages; such as, PROLOG and LISP and AI

programming tools; such as, ART, KEE and/or Language Craft which permit expert system programming to make use of speech recognition, speech synthesis, robotics, and/or visual perception for inputs/outputs.

7.6 Vocabulary (Lexicon) Dictionary for AI Applications

Here again AI input systems capabilities (parsers, interpreters, and language translators) would have to be incorporated as a required part of speech recognition expert systems which would handle the following:

7.6.1 Paraphrases--to give the meaning of a passage in other form or language.

7.6.2 Relative Clauses--a clause adjunct introduced by a relative pronoun referring back to a noun also linking a subordinate clause to a main clause.

7.6.3 Synonyms--one of two or more words of the language having the same or nearly the same meaning.

7.6.4 Pronouns--a word used instead of a noun; one of a small group of words referring to persons or things either named, asked for, or understood in the context.

7.6.5 Abbreviations--a shortened or contracted form of a word.

7.6.6 Antecedents--nouns or noun equivalents (whether word, phrase, or clause) referred to by a relative personal pronoun.

7.6.7 Acronyms--words formed from the initial letters or syllables of the successive parts of a compound term.

7.6.8 Idioms--expressions in the usage of a language that are peculiar to themselves either in grammatical construction or in having a meaning which cannot be derived as a whole from the conjoined meanings of their elements.

The lexicon dictionary would be limited to the paraphrases, synonyms, pronouns, abbreviations, acronyms, and idioms placed in the NUL dictionary and would be geared to prompt for another (other) word(s) by displaying a message indicating its level of understanding or would automatically display a tasks' menu it could perform from its understanding of the command given, so that the crewman could verify the computer's understanding with the correct response.

8.0 REVIEW OF POSSIBLE USER LANGUAGE INTERFACES WITH ADA

8.1 GOAL (Ground Operations Aerospace Language)

GOAL is a KSC (Kennedy Space Center) language design for shuttle integrated test checkout and launch operations, for payload integration with shuttle, and ESA (European Space Administration) Spacelab operations.

--Host/Target Machines: Honeywell 6680/Modified Modcomp IV and special purpose control console, keyboard, and graphics display

--Potential Applicability to Space Station NUL include: Language syntax and functional design meets requirements for English-like readable language for text and routine operational procedures; additional design work needed to develop short form of language for real-time, interactive operations; and re-implementation from language specificiation would be needed for new host hardware.

--Strengths: Language well matched to the process it is designed to control. Easy to learn English-like syntax provides readability and self-documenting procedures which aid in verification, validation, and maintenance.

--Weaknesses: High level of dependence on Shuttle ground computer hardware, not implemented with an interpreted/on-line interactive mode (compiled), limited application program interface and display hardware dependent.

8.2 JSC Prototype Crew Workstation for Space Station

Environment:

Serves three classes of users: command and control (station maintenance), proximity operations, and potentially payload operations:

Primary Graphic Interface: Menu selection of predefined procedures and subsystem parameter/status displays selected by cursor position on system schematic,

includes concept that user must provide graphic display sequences and control software for all payload sequences in conventional programming langauges.

Limited prototype coded in assembler on 16-bit minicomputer.

Potential Applicability to Space Station NUL:

1. Mock-up is a potential user demonstration and evaluation facility for both graphical and command interfaces.

2. Mock-up can be used to study transition between and combined use of displays and command language.

3. NUL could be used to provide command interface mode and develop additional pre-defined procedures for mock-up.

4. Would require:

   a. Redevelopment on larger computer in high level language.
   b. Design of interface between NUL and display control software.
   c. Development of utility software for user display development.

8.3 Language Craft

Language Craft is the first and so far only commercial environment for constructing NUL interfaces that exploits the principle of caseframe instantiation and its methods of integrating semantics and symbol knowledge.

A natural language interface frees users from the cognitive burden of determining how to formulate commands or desires in artificial, unfamiliar, command or query languages. Thus, users can focus their attention on the underlying task, rather than on the medium of communication. Natural language interfaces enables users to make productive use of a system from the outset, rather than going through lengthy and frustrating familiarization rituals with each new system encountered.

From the end user's point of view, Language Craft prompts for natural language input, the application replies, Language Craft issues another prompt, and so on. Sometimes

Language Craft may find a user's input ambiguous. In this case, it paraphrases its alternative understandings back to the user and asks for a choice between them. Other characteristics of Language Craft the user may notice include spelling correction and ellipsis (filling in information missing from the user's current input from the context provided by previous input (relative clauses)). In addition, Language Craft handles yes/no and WH (who, what, where, when, and which) questions.

Language Craft supports two quite differnt types of user:

--The end-user of a Language Craft interface to an application.
--The developer of a Language Craft interface to an application system; i.e., the person who specifies the information that Language Craft needs to provide an interface to a specific application system. This person is called a grammar writer.

Natural language interfaces provide friendly and effective communication to a wide variety of software systems and utilities including databases, expert systems, electronic mail, operating systems, computer-aided design, and office automation software.

Given the desirability and widespread utility of Natural Language Interfaces (NLI), the question arises of why all software systems do not already incorporate NL front ends. The language comprehension task is far more complex than originally envisioned. Only in the past decade have NL researchers been able to build automated language analyzers. And very few of them are sufficiently robust and efficient to form the basis of a commercial quality product. To date, there exists no universal general-purpose language comprehension system capable of reading a textbook and conducting unconstrained conversation with humans. Nevertheless, language interfaces for well-defined tasks have been built and have proven quite successful.

Constructing a NLI has especially required several person-years of effort by highly-skilled computational linguists and AI researchers. In order to reduce the human resource requirements, the idea emerged of creating an environment to support the development of NLIs--an environment that would minimize the need for specialized AI skills. Perhaps more importantly, such an environment would cut down the development time required to produce a functional

natural language interface from several person-years to a few person-months.

Language Craft (LC) is implemented in Common LISP and currently runs on Symbolics 3600 LISP machines and DEC VAX operating systems. The LC environment is portable to several standard AI workstations and a variety of LISP machines.

8.4 LISP (List Processing Language)

LISP (List Processing) (1957) is the most widely used AI language in the United States. AI languages differ from conventional languages; such as FORTRAN or C, in that they concentrate on manipulating symbols and defining relationships. Moreover these tasks can be accomplished easily without concern for data handling and memory allocation. Because of this ease and flexibility in symbol handling, AI languages are much better at tackling unpredictable situations. And AI languages can be easily altered or expanded.

LISP is made up of thousands of functions and users can add their own. These new functions are then treated as those already defined.

AI languages such as LISP also come with many functions for writing, editing, and debugging programs. Windows divide the screen into several areas to show the program, results, and data simultaneously. As a result, LISP programs can be easily altered. Moreover, LISP can be interpreted to speed programming or compiled to increase performance.

Many versions of LISP (called DIALECTS) have been created. Some DIALECTS are written on top of standard DIALECTS, optimizing the use of the language for specific applications. But the disadvantage is that a program written in one DIALECT may not run on a machine supporting another DIALECT. Standards have been created in an effort to overcome the drawbacks associated with many DIALECTS, with COMMON LISP receiving much attention due to the backing by the Federal Government. There is word currently that Common LISP will try to set a more uniform standard for LISP programs.

Recent improvements in computer technology, together with the availability of some of the finest software development environments in existence have made LISP a

viable language. The protocol support Symbolics, IBM PC XT, and DEC VAX link LISP to their machines. Language Craft, ART, PROLOG, MACSYMA, ESDT, etc. are implemented in LISP.

LISP offers a complete set of tools for productive and creative software development. This environment makes exploratory programming and rapid prototyping practical and allows for the easy creation and maintenance of large software systems.

The LISP (or LISP Subsets) language offers extensible data structure for maximum flexibility in data representation and manipulation. Text and data structures are represented in identical manner so programs that generate other programs can be easily written. The programmer can write new constructs or even completely new languages and easily build them on top of LISP to handle a particular problem domain.

LISP has capabilities lacking in other computer languages in that it permits manipulation of complex structures and symbolic information. Because of its ability to manipulate symbols, it is used for expert systems, visual recognition, and other subcategories of AI. Its major drawback, however, is that its large size and flexibility can easily swamp a conventional mainframe.

New computers have been designed specifically to provide for an interactive environment in which both data and functions coexist and can be inspected or modified easily. Functions can be tested as they are written, and problems found quickly. These features have made LISP the preferred tool of AI researchers.

Symbolic processing allows computers to deal with complex knowledge and data in such a way that it appears to mimic human intelligence. The LISP computer language evolved to handle the constructs of symbolic processing. It now enables computers to be easily programmed to represent objects and the realationship among them.

8.5 PROLOG (Programming in Logic)

The Japanese announced that they had chosen logic programming and PROLOG for their ambitious long-term nationwide effort known as the Fifth Generation Computer System (FGCS) Project thus increasing the interest in the language.

Among the expressed objectives of the FGCS project is the development of fast, intelligent computer systems with the following capabilities: human-like decision making and learning, natural language and voice I/O (input/output), automatic program generation, and distributed processing.

In the FGCS project, logic programming is envisioned as the link between the fields of software engineering, data-base systems, computer architecture & knowledge engineering. It is to be used for problem specification and transform-ation, unifying functional programming and relational databases, developing single-assignment languages, and constructing rule-based expert systems and natural-language processors.

8.6 SCOL (System Control and Operations Language)

A need existed for a standard high-level interactive, user-oriented language to control system level testing and operation. The system test and operation tasks, written in a language other than SCOL (e.g., ADA, ATLAS, FORTRAN), would be responsible for performing complex and time-critical actions and analysis, system monitoring and test instrument interface.

SCOL shall be used in a spacecraft system test environment. Precompiled software may be in Assembly, FORTRAN, ATLAS, PASCAL or any language a user selects. SCOL could reside in a real-time computer system used for control and monitoring of spacecraft system operations. Precompiled software performs telemetry processing, command processing, and display processing.

The language is to be defined using a formal syntax notation and syntax diagrams. The syntax definition shall be supplemented with semantic rules and constraints. All definitions shall be consistent with internationally recognized standards. The language and its implementation could provide generality only to the extent necessary to satisfy the needs of system level testing and operation.

The language should be designed to avoid error prone features and to maximize automatic detection of programming errors. The language should be easy to learn. Implementations should be easy for experienced and inexperienced personnel to use. The language and its implementations should be designed to facilitate rapid interactive use of terminology familiar to system control

and operation personnel. The language will provide a user-friendly tool for complete control and monitoring of the system.

8.7 SMALLTALK

SMALLTALK created by XEROX's Palo 'Alto Research Center (PARC) has gained prominence recently because of its ability to handle objects rather than functions. In this approach, a string not only has a value but a range of characteristics as well. These characteristics are "inherited" from the class that an object belongs to unless an exception is defined. Messages are used to perform operations on objects. Object-oriented programming can also be implemented in LISP by a set of functions; known as FLAVORS.

8.8 SSOL (Space Station Operating Language)

SSOL is an automated environment friendly to users, in which SPACE STATION Integration and Tests (I&T) Activities can be designed, developed, tested and performed. The SSOL System is being developed by KSC (Kennedy Space Center) and will be responsive to users independent of their location. Also, it will promote standardization and transportability of user-developed procedures from site to site, capitalizing on the functional commonality of the I&T activities at these sites. Advances in the user-friendliness of commercial real-time operating systems, data base management systems, supporting software development tools, processing speed, and memory capacity are only a few of the anticipated technological improvements which may enhance the SSOL System.

An analysis has been performed to determine the validity, integrity, and completeness of the initial SSOL System concept. The results of this analysis, and the emphasis placed on the need for an I&T concept of automation, were used to derive and refine SSOL System requirements and concepts.

The SSOL System is both a ground and onboard SPACE STATION Data System (SSDS) service.

8.9 STOL (Standard Test & Operations Languages)

This language is used extensively at GSFC (Goddard Space Flight Center) and Laboratory for atmospheric and space

physics (LASP) and implemented on DEC Systems PDP 11/70, PDP 11/34 and VAX machines.

Strengths:

1. Transportable--90% of code is FORTRAN IV Plus
2. Designed for interactive use in real-time operations
   a. Processed via language interpreter
   b. Responsive to user needs in real-time environment
   c. Easy to learn
   d. Can generate command procedure lists callable by name
3. STOL was designed to be and has been successfully extended
4. Data query commands for selected parameters, limits, state variables, and conversions allow little user modification

Weaknesses:

1. Slower than compiler languages
2. Readability compromised somewhat by concise command words
3. STOL is not under strick configuration control
4. "Single User" mode disallows multiple command stream generation required by KSC OPS (except for LASP SME-VAX version)
5. Terminal device I/O is operating system dependent
6. No graphics capability

Potential Applicability to SPACE STATION VHOL:

Storing "pre-interpreted" procedure lists appears feasible and would provide speed advantage. Readability can be improved by incorporating front-end translators. Extensions required to provide additional data query functions. Performance good for real-time operations, not necessarily for the critical or concurrent process control functions.

8.10 TAE (Transportable Applications Executive)

TAE was developed at Goddard Space Flight Center (GSFC).

Environment:

--Payload/instrument data analysis

--Provides standard interface:
  -to user for application program control and parameter definition
  -to application program for user interaction and operating system services
--In wide use in NASA and universities
--Originally developed for VAX under VMS. Posted to: VAX under UNIX, PDP 11 under RSX-11M and Data General under RDOS.

TAE - Interface Modes:

--Menus - Tutored Input Mode
--Command Mode
--User Created Procedure with
  -Global and local variables
  -Conditional statements
  -Looping

Strengths:

--Effective user interface for infrequent users (menu mode) as well as expert users (commands and procedures).
--Provides common interactive user interface to applications programs. Support System extendability.
--Highly protable - 87% of code.

Weaknesses:

--Interactive but not designed for real-time - some speed penalty.
--No graphics interface.
--Not in use for either of primary UIF (User Interface) functions - I&T or real-time operations.

Potential Applicability to SPACE STATION UIL:

--Use as prototype for user feedback on limited subset of language features (e.g., TAE tutored Input and Command Modes, approach to procedure capability).
--Design concept of common user interface software with application program interface routines for parameter acquisition.

8.11 Third-Party Software: such as, ART, KEE, & MACSYMA

ESDT (Expert System Development Tools) are programs that aid the user or knowledge engineer extract and code information required to build and test and expert system.

ART (Automated Reasoning Tool) is a comprehensive software tool that basically turns engineers into builders of full-scale expert systems. ART was designed as a tool for developing expert systems, programs that combine human expertise with artificial intelligence (AI) to produce results that have been beyond either computers or humans in the past. It contains an inteference engine, which resembles an operating system of conventional systems; a knowledge base, which replaces a data base; an editor, which helps create the knowledge base, and a monitor, which lets users visually follow the tool's reasoning steps.

Users can interact with the ART system quickly and easily--the novice, through hierarchial menus, and the experienced engineer, directly through the keyboard. Software mechanisms automatically correct errors, and graphic representations of the run-time knowledge base let the user visualize ART's approach to a problem immediately.

The world of AI deals with reasoning through forward and backward chaining. The first uses goal patterns to cull new facts from existing declarative knowledge. The latter depends on strategy patterns to draw simpler goals from more complex goals. Forward chaining moves from the data toward the overall goal; backward chaining attempts to successively reduce the overall goal until it finds matches in the known data. Both methods in chaining are valuable procedural organizations for expert systems.

ART reasons through both types of chaining; rules may involve both goal and strategy patterns. ART automatically constructs and maintains a viewpoint structure that represents all hypothetical possibilities an expert system is considering. Rules can explicitly compare and contrast the possibilities to decide on the best possible solution.

The expert systems that ART produces are written in LISP.

KEE (Knowledge Engineering Environment) (IntelliCorp - 1983) was the first ESDT commercially available "Tell and Ask" language which helps define and retrieve facts from the knowledge base in sentences. Active images permit users to create graphical displays for viewing and controlling the system. KEE is run on all AI machines and is being applied to a number of engineering and manufacturing situations, including simulation of production facilities.

MACSYMA, developed at the MIT Artificial Intelligence Laboratory and marketed by Symbolics, Inc. of Cambridge, MA is one of the earliest and still most impressive AI based software packages. It is an intelligent system that uses the flexibility and power of symbolic processing to solve equations of algebra and calculus so complex that they are generally beyond the skill and patience of human mathematicians. MACSYMA produces important analytical solutions that could otherwise only be approximated by numerical methods. Moreover, computation results can be printed out as more easily comprehended, and often as beautiful graphic presentations. Indicative of the system's power is the fact that a single line of MACSYMA code is typically equivalent to eleven lines of conventional FORTRAN.

MACSYMA's ability to compress entire mathematical entities like variables, symbols and operations into symbolic objects points out one of the major commercial advantages symbolic processing has over numeric processing. Applications software can be created faster and less expensively using symbolic programming methods. Because data and commands are separate, programs can be developed, prototyped, tested, and edited in more efficient increments.

MACSYMA is an "expert system." It provides logical rules which use facts in the data base to infer other facts about how to solve an equation. As in all expert systems, the secret of MACSYMA's power is in its data base.

8.12 UIL (User Interface Language)

UIL is a set of software tools for a flexible but standard user interface to the SPACE STATION System, SPACE STATION Payloads, and Platforms which supports on-orbit plus ground integration, tests, and operations, for use by engineers, crew and scientists. It provides English-like language for familiarity and readability; short form for real-time interactive control procedure capability; includes software utility support for graphics, display, data, and dialogue interface and control. Users would include: Development, test and checkout engineers, launch operations engineers, ground mission operators, flight-crew, flight and ground payload/instrument scientists and engineers.

As a result of the UIL study, it was recommended that a standard, high-level language be developed to make phase and location differences transparent to users & user procedures.

9.0 VERY HIGH ORDER SPACE STATION NUIL SUMMARY

After having reviewed the President Ronald Reagan State of the Union Message of January 25, 1985 which directed NASA to develop a permanently manned SPACE STATION within the next decade, NASA began to divide the task into digestible parts. The NASA/MSFC EB Lab's part evolves around the software evaluation of the common module computer cores.

Because of the various trained- and practiced-for tasks of the crew members, and because of the different levels of expertise of the crew, it was felt that the development of a user-friendly interface language would increase productivity of the crew members as well as encourage and increase commercial ground users. This NUL would be able to either prompt the users for responses or would accept their natural language for commands. Crew members or ground users could utilize the computer with or without having prior programming knowledge; however, users with programming knowledge would not be limited to the applications (expert systems) firmware either.

The system should be expandable over the 30-year life of the SPACE STATION. To envision what will be 30 years in the future is exhaustive, but by "sky is the limit" brainstorming, I can foresee expert systems that have language translators in the popular languages of the world (i.e., English, French, German, Spanish, etc.) which would accommodate Americans and our friends around the world in free enterprise, accept each users natural language input, and return each users natural language output by using speech recognition, speech synthesis with audible, display and/or hardcopy outputs.

This expertise would be evolutionary and would initially require tutored display, keyboard responses, or touch screen inputs for testing, monitoring, controlling, and maintaining the system and who knows what else this system will be expected to do in the future.

The next problem involved a review (evaluation) of what has been done in this area to see whether the expert-system program had to be started from scratch using a newly created language or whether an already existing interface language could be used or modified for use in this situation.

GSFC's, KSC's, JSC's, LRC's, and MSFC's current reports and expertise on the subject were reviewed, demonstrations and three workshops were attended, two long-distance calls were made to JSC (Sandy Richardson), and Prime Contractors, Clive A. Arlington - Martin Marietta Denver Aerospace and Dr. Byron Purvis - BOEING, Corp. were consulted, a class in LISP was taken, and John Wolfsberger, my boss, added to my knowledge base and contributed greatly to this study.

Some existing interface languages and AI programming tools were reviewed: GOAL, JSC, Language Craft, LISP, PROLOG, SCOL, SMALLTALK, SSOL, STOL, TAE, ART, KEE, and MACSYMA. From my review, I found that there were two, one already developed language and another AI programming tool which I felt needed further study for possible adoption for use in the SPACE STATION NUIL development.

Recommendation 1--TAE originally developed for DEC VAX under VMS by GSFC for the initial SPACE STATION Common Module UIL which utilizes all the user-friendly requirements, I envisioned as being needed initially (tutored inputs with prompted responses, menus, helps, and commands) for user program development, and program application execution.

Modifications and expansions would, however, be needed to increase its real-time speed and interactive abilities, to add graphic interface, and to include I&T and real-time operations which are primary interface functions.

Recommendation 2--Language Craft, an integrated natural language processing environment for developing NLI for operating systems, databases, and knowledge-based systems (which has a language parser which handles synonyms, paraphrases, pronouns, verbs, etc. needed for speech recognition and synthesis) be used to eventually convert a TAE Subset interface to a speech synthesis, speech recognition expert interface system and retain parts of TAE Subset for verification of spoken commands via prompted responses.

In the researcher's view these were the only two languages reviewed which would not require prior user programming knowledge. Also, previously written conventional language programs could use this NUL as transparent front-end software for easy user access.

10.0 REFERENCES

1. NASA Advanced Technology Committee (ATAC), NASA Executive Overview, NASA Technical Memorandum 87566, Advanced Automation & Robotics Technology for the SPACE STATION and for the U.S. Economy, Volume 1, March, 1985.

2. Mike Argabright, Chairman, SCOL Subcommittee of IEEE Atlas Committee, "SCOL (System Control & Operation Language) Requirements Specification," 1 October 1983.

3. Jaime Carbonell, "Language Craft Reference Manual (Limited Edition Release)," Version 2.0, Carnegie Group, Inc., 1985.

4. Clara Y. and John L. Cuadrado, "Prolog Goes to Work - What Prolog Is, Who's Using it, and Why?" BYTE, August, 1985, Volume 10, No. 8, pp. 151-158.

5. John Dalton (GSFC), Chairperson, "Report of the Ad Hoc Committee for SPACE STATION User Interface Language," July 9, 1984.

6. A. Dorofee and L. Dickison, "Space Station Operations Language (SSOL) System Requirements and Concept Definitions," August 1, 1985.

7. E. A. Feigenbraum, A. Barr, and P. R. Cohen (Eds.), The Handbook of Artificial Intelligence, W. Kaufmann, Publisher, Los Altos, CA, 1982.

8. William K. Hartmann, Ron Miller, and Pamela Lee, Out of the Cradle--Exploring the Frontiers Beyond Earth, Workman Publishing Co., NYC, 1984.

9. Abraham Hirsch, Program Marketing Manager, Symbolics, Inc., "Artificial Intelligence Comes of Age," COMPUTERS AND ELECTRONICS, March, 1984.

10. Abraham Hirsch, "Tagged Architecture Supports Symbolic Processing," COMPUTER DESIGN, June, 1984.

11. Robert B. Mills, "Building Your Own Expert System," COMPUTER AIDED ENGINEERING, June, 1985, pp. 76-78, 80, 84, 86, and 90.

12. Susan Voigt, Editor, NASA Language Research Center, "Report on an Open Forum with Industy and Academia," Held at NASA/MSFC, Huntsville, AL, April 24 and 25, 1985, NASA Scientific and Technical Information Branch.

13. Robert Wilensky, LispCraft, W. W. Norton and Co., New York, 1985.

14. Chuck Williams, "Software Tool Packages the Expertise Needed to Build Expert Systems," ELECTRONIC DESIGN, August 9, 1984.

SUMMER 1985

NASA/ASEE SUMMER FACULTY RESEARCH
FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

"APPLICATION OF WEIBULL ANALYSIS TO SSME HARDWARE"

BY

LOU ALLEN BELL GRAY
ASSOCIATE PROFESSOR OF MATHEMATICS
JACKSON STATE UNIVERSITY

NASA/MSFC

LABORATORY:	STRUCTURES & PROPULSION
DIVISION:	PROPULSION
BRANCH:	LIQUID PROPULSION SYSTEMS
NASA COUNTERPART:	JAMES S. RICHARDS
DATE:	AUGUST 16, 1985

Contract No: NGT 01-008-021
The University of
Alabama in Huntsville

A STUDY OF WEIBULL ANALYSIS TECHNIQUES AND ITS UTILIZATION TO STUDY PARTS/ASSEMBLIES FAILURE OF THE SPACE SHUTTLE MAIN ENGINE

BY

Lou Allen Bell Gray
Associate Professor of Mathematics
Jackson State University
Jackson, Mississippi

ABSTRACT

Generally, it has been documented that the wearing of engine parts forms a failure distribution which can be approximated by a function developed by Weibull. The purpose of the above mentioned study is to examine to what extent does the Weibull distribution approximate failure data for designated engine parts of the Space Shuttle Main Engine (SSME). The current testing certification requirements will be examined inorder to establish confidence levels. An examination of the failure history of SSME parts/assemblies (turbine blades, main combustion chamber, or high pressure fuel pump first stage impellers) which are limited in usage by time or starts will be done by using updated Weibull techniques. Efforts will be made by the investigator to predict failure trends by using Weibull techniques for SSME parts (turbine temperature sensors, chamber pressure transducers, actuators, and controllers) which are not severely limited by time or starts.

INTRODUCTION

During the early thirties, Fisher and other statisticians developed a new probability distribution. The distribution developed was not utilized significantly in application until the late forties by Waloddi Weibull. In 1951, Weibull presented his argument internationally for the extensive usage of the new distribution to describe mathematically the behavior of fatigue data. As time progressed, Weibull distributions were found to be very good mathematical approximations of failure data related to engine parts analysis, thermo cycles analysis (large numbers of thermo cycles), and electronics analysis. The Weibull distribution is currently being applied extensively to predict the reliability of Space Shuttle Main Engine (SSME) parts or units. The mathematical model for the Weibull distribution is:

$$F(t) = 1 - e^{-\left(\frac{t-to}{\eta}\right)^{\beta}}$$ where

$F(t)$ = Fraction Failing Up to Time t

t = Failure time

$*t_o$ = Starting Point of the Distribution

η = Characteristic Life

β = Slope on Weibull Paper

e = Exponential

$$R(t) = e^{-\left(\frac{t-t_o}{\eta}\right)^{\beta}}$$ where

$R(t)$ = Reliability Up to Time t

$F(t) = 1 - R(t)$

$*t_o$ is generally assumed to be zero.

The Weibull distribution's shape is determined by the value of beta. Beta is the slope of the line graphed on 1 - to -1 Weibull paper. The probability density function is:

$$\frac{dF(t)}{dt} = f(t)$$

$$f(t) = \frac{\beta t^{\beta-1}}{\eta^{\beta}} e^{-\left(\frac{t}{\eta}\right)^{\beta}}$$

The hazard function describes the instantaneous failure rate. The hazard function is:

$$h(t = \frac{f(t)}{R(t)}$$

$$h(t) = \frac{\beta t^{\beta-1}}{\eta^{\beta}}$$

For the engineer, beta gives a cue as to what type of failure conditions exist, i.e. rapid-wear, green-run etc. Eta does not provide extensive information. Eta is mathematically derived as always being the time where approximately 63.2% of the sample fails. Eta is used in calculations related to risk analysis and substantiation work. The engineer should examine B life points, i.e. B1 is the time where 1% of the sample fails, B.05 is the time where 0.05% of the sample fails etc. For the eta, which is called the characteristic life, can be large but due to the slope the B1 or B10 life may be small. This paradox occurred in a study by Rheinfurth (1985) related to chamfered and unchamfered blades. The unchamfered blade had eta equal to 1,142,538 seconds and a B1-life of 413 seconds. The chamfered blades had an eta equal to 69,324 seconds and a B1-life of 5,429 seconds. The difference in B1-life was created by the difference in the slope, beta. The usage of the Weibull distribution can also help to unmask multiple failure modes. A significant increase in the slope can indicate lot (batch) problems.

GENERAL FINDINGS

In most instances, there is an assumption made pertaining to the distribution type. Carter, Bompas-Smith, and Nelson describes the Weibull distribution as being one of the most accomodating distribution available. For, the distribution can represent increasing and decreasing failure rate. Particularly, skewed distributions can be represented. Skewed distributions can not be represented by log normal distributions. Brook asserts that the

Weibull distribution allows the making of simple statistical judgements with very little labour or efforts. Documentation by reliability engineers and statisticians in industry and government exist with respect to the high level of predictability of Weibull distributions. The U.S. Air Force and Pratt-Whitney in particular use Weibull Analysis extensively to predict the failure of jet engine parts. Breneman of Pratt-Whitney Company has 15 years of documentation that the physics of wear for a particular part of a jet engine does not significantly alter with modifications. This means the slope or failure mode doe not significantly change. This finding can be used in a process to predict failure on new parts with material changes or design modifications where the new parts have not experienced failure. The methodology used is Weibayes analysis. The confidence intervals for β or the Weibull line can assist the engineer in reaching a decision as to whether or not a given part's failure mode has been modified.

WEIBULL ANALYSIS FOR SSME PARTS

The establishment of a slope (beta) file for SSME parts should be the initial task of engineers. Due to the small number of failures allowed by NASA engineers for most SSME parts, documentation of the initially derived slopes over an extended period of time will not be possible. Yet, Breneman asserts that 3 failures in a sample are sufficient to give a good approximation of the failure mode.

The likelihood function for the Weibull distribution is a useful tool. The likelihood function is more mathematical in nature since reliability is a part of the computational process. Waloddi Weibull (1969) described the mathematical traits of the likelihood function for three different conditions. The models are:

Type 1: r units fail (L_1) for some t_i failure time

$$L_1 = \prod_{i=1}^{r} f(t_i) \text{ where } f(t_i) \text{ is the density function}$$

Type 2: S units still running (L_2) for some Z_j time

$$L_2 = \prod_{j=1}^{S} [1 - F(Zj)]$$

or

$$L_2 = \prod_{j=1}^{S} R(Zj) \text{ where } R(Zj) \text{ is the reliability}$$

Type 3: d_K units fail for some K number of inspection intervals, where the interval size is $W_K - W_{K-1}$ of time.

$$L_3 = \prod_{k=1}^{K} [F(W_K) - F(W_{K-1})]^{d_K}$$

or

$$L_3 = \prod_{k=1}^{K} [R(W_{K-1}) - R(W_K)]^{d_K}$$

where: w_k = the end of the interval

w_{k-1} = the beginning of the interval

So for each given condition, the final likelihood function for a total sample is

$$L = L_1 \cdot L_2 \cdot L_3$$

$$\ln L = \ln L_1 + \ln L_2 + \ln L_3$$

For the Weibull distribution in logarithmic form

$$\ln L = \sum_{i=1}^{r} \ln \left[\frac{\beta t_i^{\beta-1}}{\eta^{\beta}} e^{-\left(\frac{t_i}{\eta}\right)^{\beta}} \right] + \sum_{j=1}^{s} \ln \left[e^{-\left(\frac{z_i}{\eta}\right)^{\beta}} \right] + \sum_{k=1}^{K} d_k \ln \left[e^{-\left(\frac{\omega_{k-1}}{\eta}\right)^{\beta}} - e^{-\left(\frac{\omega_k}{\eta}\right)^{\beta}} \right]$$

The Weibull likelihood function can be further expanded by the properties of logarithims. Other models can be generated when all three conditions do not exist.

$L_{12} = L_1 \cdot L_2$ Model which gives the description of η, Weibayes analysis

$$L_{12} = \prod_{i=1}^{r} f(t_i) \prod_{j=1}^{s} R(Z_j)$$

$$L_{12} = \prod_{i=1}^{r} \frac{\beta t^{\beta-1}}{\eta^{\beta}} e^{-\left(\frac{t_i}{\eta}\right)^{\beta}} \prod_{j=1}^{s} e^{-\left(\frac{Z_j}{\eta}\right)^{\beta}}$$

r = number of units run to failure

s = number of unfailed units

t_i = known failure time

Z_j = operating time on unit

Take the logarithm of both sides and differentiate. We get:

$$\frac{\partial \ln L}{\partial \beta} = 0, \quad \frac{\partial \ln L}{\partial \eta} = 0$$ solve simultaneously

Generally the Newton-Raphson method can be used to solve the equations. But, caution is needed when using this technique since under certain conditions the process may not converge (Campbell, 1985). Initializing a beta value is best done by obtaining some estimation from a Weibull plot.

The model L_{23} should be utilized when the failure data is interval of inspection.

$$L_{23} = L_2 \cdot L_3$$

$$L_{23} = \prod_{k=1}^{K} [R(\omega_{k-1}) - R(\omega_k)]^{d_k} \prod_{j=1}^{s} R(Z_j)$$

$$\ln L_{23} = \sum_{k=1}^{K} d_k \ln \left[e^{-\left(\frac{\omega_{k-1}}{\eta}\right)^{\beta}} - e^{-\left(\frac{\omega_k}{\eta}\right)^{\beta}} \right] + \sum_{j=1}^{s} \ln \left[e^{-\left(\frac{Z_j}{\eta}\right)^{\beta}} \right]$$

The L_{23} model can be modified to coincide descriptively with test runs. Rheinfurth (1985) at MSFC, ED01, developed a Weibull likelihood for the observation interval data base on N trial runs. Using the L_{23} model

$$\ln L = \sum_{i=1}^{N} d_k \ln \left[e^{-\left(\frac{\omega_{k-1}}{\eta}\right)^{\beta}} - e^{-\left(\frac{\omega_k}{\eta}\right)^{\beta}} \right] - \left(\frac{1}{\eta^{\beta}}\right) \sum_{i=1}^{N} (M - d_k) \omega_k^{\beta}$$

d_K = number failing in the interval test

$m-d_K$ = number still working at the end of test

n = number of test runs

It should be noted that the first term goes to zero when there are no failures for all test runs. The likelihood function is best for interval data since it makes no assumption as to the time of failure. For Weibayes analysis where some units run to failure and others donot, we get a derived mathematical description of eta which is

$$\eta = \left(\frac{\sum_{i=1}^{N} t_i^{\beta}}{r} \right)^{\beta}$$

where N is the number of suspensions or unfailed units

t_i is the operating time accumulated

r is the number of failures (no failures, set r=1)

β is the assumed value based on previous experience (records, etc.)

Mathematically, the assumption of at least one failure ensures that η is definable. When there are no failures, the formula becomes

$$\eta = \left[\sum_{i=1}^{N} (t_i^{\beta}) \right]^{1/\beta}$$

Since Weibayes analysis estimates the possible value of the slope, engineers may wish to try a variety of slopes over some range. Weibayes tests will not indicate the existance of a new failure mode. If the parts being analyzed are nonserialized, then assumptions pertaining to success or failure times must be made. Mathematically, the confidence level of the Weibayes lower bound is unknown because it depends on the actual times to the first failure.

SUMMARY

The mathematical flexibility of the Weibull distribution makes it the best choice for failure analysis. The extension of the fundamental concepts of the Weibull distribution allows it to be highly applicable to perform risk and substantiation analysis. Not previously mentioned is the problem with identifying outliers. An engineer most decide when such points exists and why. For sometimes outliers at the beginning of Weibull line may actually be the results of a t_o which is not initially zero. There is a basic technique for determining t described by Abernathy (1983). A procedure developed at General Motors also describes solving for t_o (Bompas-Smith, 1973). This value is subtracted or added (depending on conditions) to each point of the Weibull line. Brook suggests a trial and error process to determine t_o. The t_o which has a none zero value may be the results of storage deteriation, testing procedures etc. When using the Weibayes techniques one must continuously keep in mind the underlying assumptions of the procedure. The most important caution for the engineer when using Weibayes is that new failure mode must not exist. With the fundamental likelihoods developed by Weibull, the engineer or statistician can by algebraic means describe their particular data situation i.e. trial runs, interval data, no failures. Confidence bands allow the engineer to visually examine the possibillty of two or more slopes being produced by the same failure mode. Mathematically Weibull distributions are very pliable. Yet, one must keep in mind real life does not generate perfect solutions. Weibull distributions provide a descriptive idea as to failure behavior for a wide variety of failure rates.

Currently, a problem is the form of the failure data. With the actuators in particular, failure by cycles or hot-firings time can occur. Also, failure is partly a decision of the engineer. Modifications to data collection procedure is recommended to reflect the actual point of such failure. The cycles must be sufficiently large numbers in order to be represented by Weibull distributions. Also, multiple failure times on each machine is a problem. For my analysis, I used only the first time to failure.

DATA RESULTS

For the FPOVA, POPVA, and CCVA, the beta is approximately .75 when the OPOVA with 1.5 seconds is ignored. The beta of .75 indicates infant mortality. The Weibull of the CCVA indicates the possibility of two failure modes. The beta of 1 indicates a failure mode of random failure (green- run, etc.) The beta of 3.5 indicates normal wear. The initial graph of FPOVA failures indicate a beta of .75. A modified graph indicates a beta of .375. The modified graph approximates a straight line better. The engineer needs to examine all aspects of usage of the FPOVA prior to hot-firing. For the MFVA and MOVA, two failure modes seem to be indicated. Upon further examination, two failure modes for the MFVA in particular seems to exist. The MOVA has an infant mortality failure mode. Modification of the time does not alter the value of beta.

TABLE 1

Actuator	Number	Time (Sec.)	Rank	Median Rank
OPOVA	086-524	1.5	1	.53
CCVA	019-003	622	2	1.29
FPOVA	050-501	622	3	2.05
FPOVA	032-323	805	4	2.86
CCVA	016-003	1222	5	3.58
FPOVA	078-552	1566	6	4.34
FPOVA	045-588	1848	7	5.10
CCVA	034-573	2059	8	5,86
CCVA	028-255	2596	9	6.62
CCVA	027-154	2597	10	7.38
CCVA	020-001	2864	11	8.14
FPOVA	020-005	4056	12	8.90
CCVA	023-003	5133	13	9.67
FPOVA	022-006	5460	14	10.43
CCVA	013-003	6170	15	11.19
CCVA	021-004	6521	16	11.95
CCVA	015-002	3553	17	12.71
OPOVA	027-004	12569	18	13.47

Sample Size = 131 Failures = 18

FPOVA, OPOVA, CCVA
Weibull

OPOVA Point 1.5 sec
not graphed

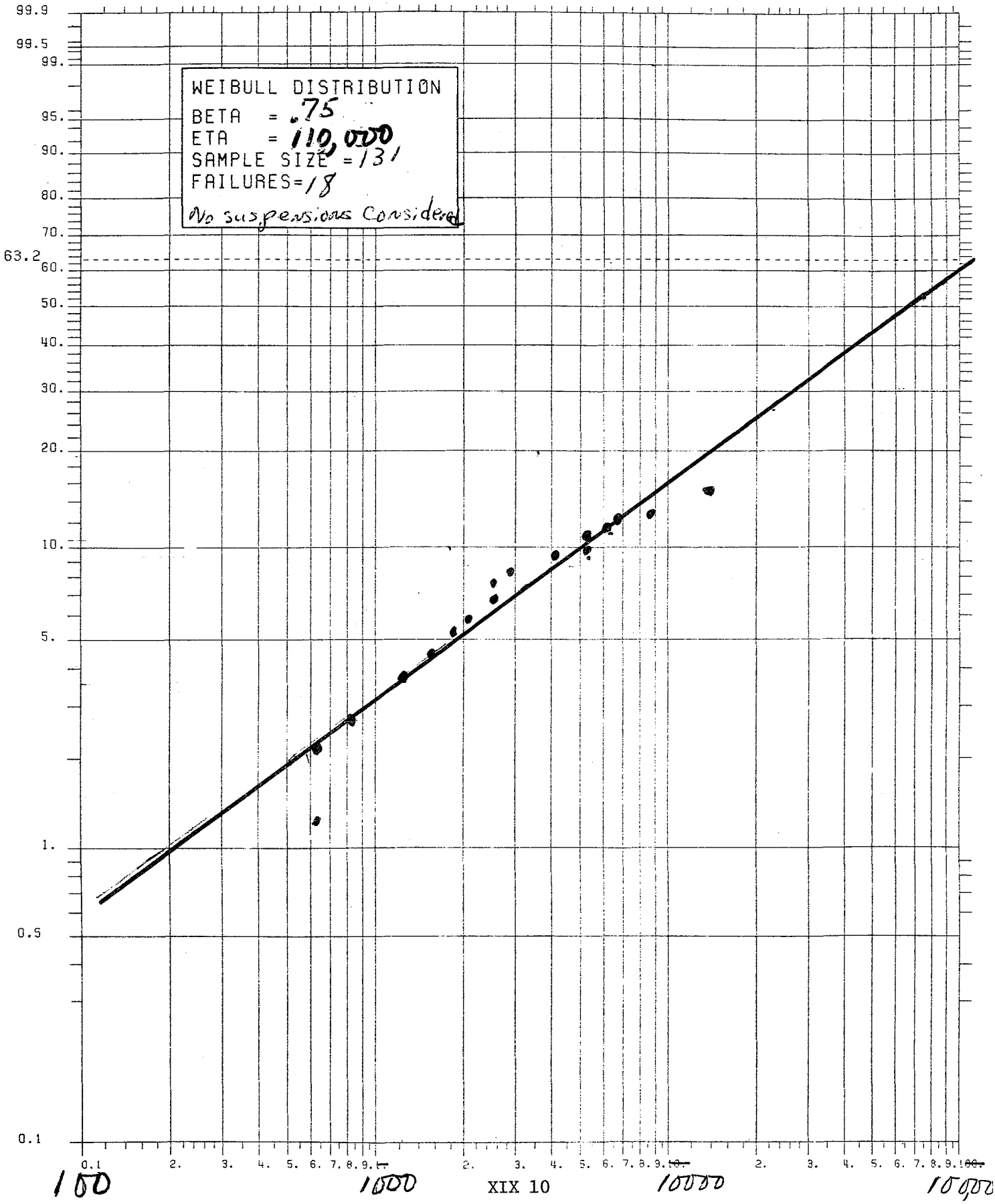


TABLE 2

Actuator	Number	Time (Sec.)	Rank	Median Rank
CCVA	019-003	622	1	1.5
CCVA	016-003	1222	2	3.8
CCVA	034-573	2059	3	6.1
CCVA	028-255	2596	4	8.4
CCVA	027-154	2597	5	10.7
CCVA	020-001	2864	6	13
CCVA	023-003	5133	7	15.3
CCVA	013-003	6170	8	17.6
CCVA	021-004	6521	9	20
CCVA	015-002	8553	10	22.3

Sample Size = 43 Number of Failures = 10

CCVA
Weibull

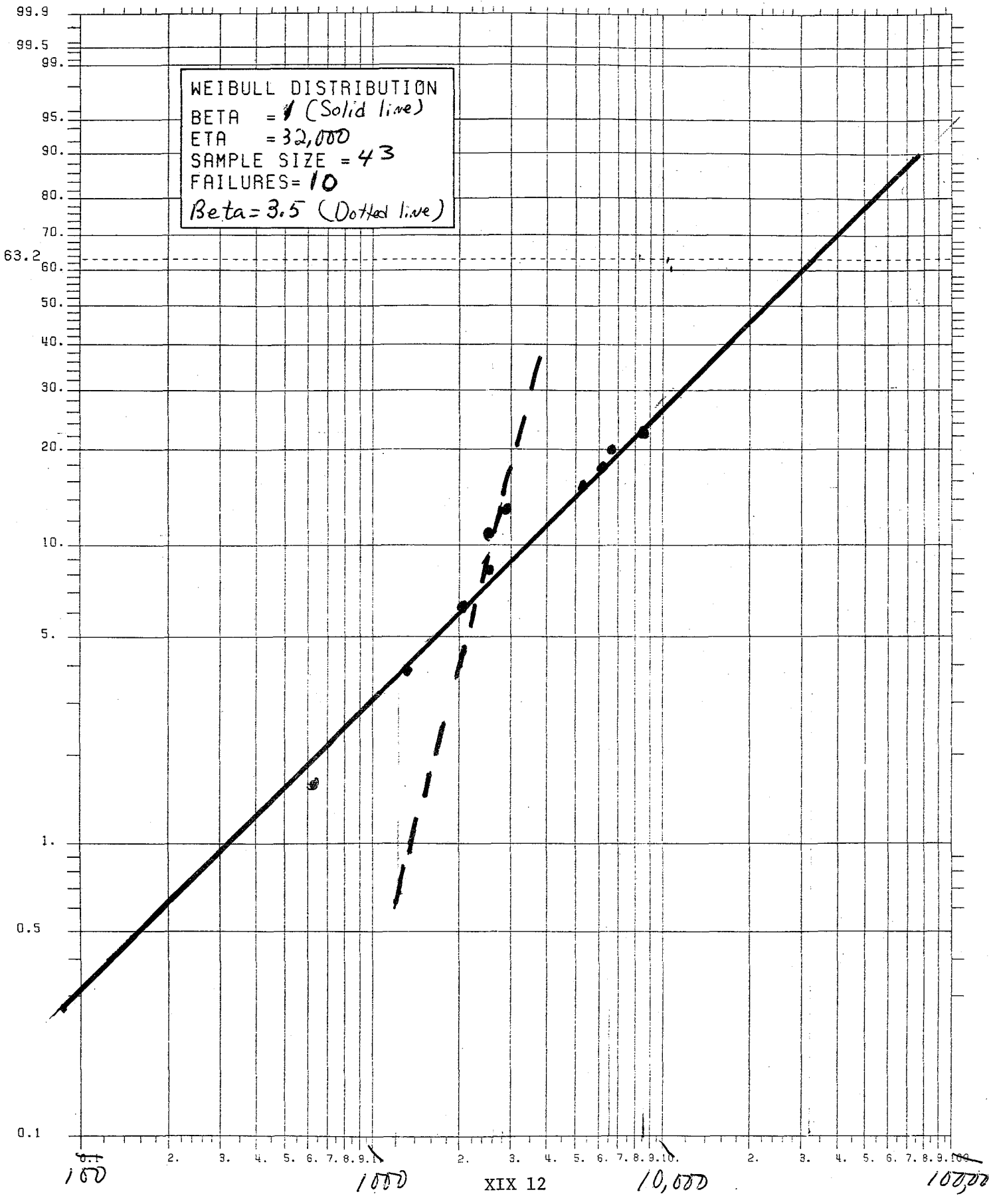


TABLE 3

Actuator	Number	Time (Sec.)	Rank	Median Rank
FPOVA	050-501	622	1	1.4
FPOVA	032-323	805	2	3.5
FPOVA	078-552	1566	3	5.6
FPOVA	045-588	1848	4	7.7
FPOVA	020-005	4056	5	9.8
FPOVA	022-006	5460	6	11.9

Sample size = 47 Number of Failures = 6

Using the formula for t_o (Abernethy, 1983), we t_o = 610 (subtract)

Time Changes: 12, 195, 956, 1238, 3446, 4850

$$t_o = t_2 - \frac{(t_3 - t_2)(t_2 - t_1)}{(t_3 - t_2) - (t_2 - t_1)}$$

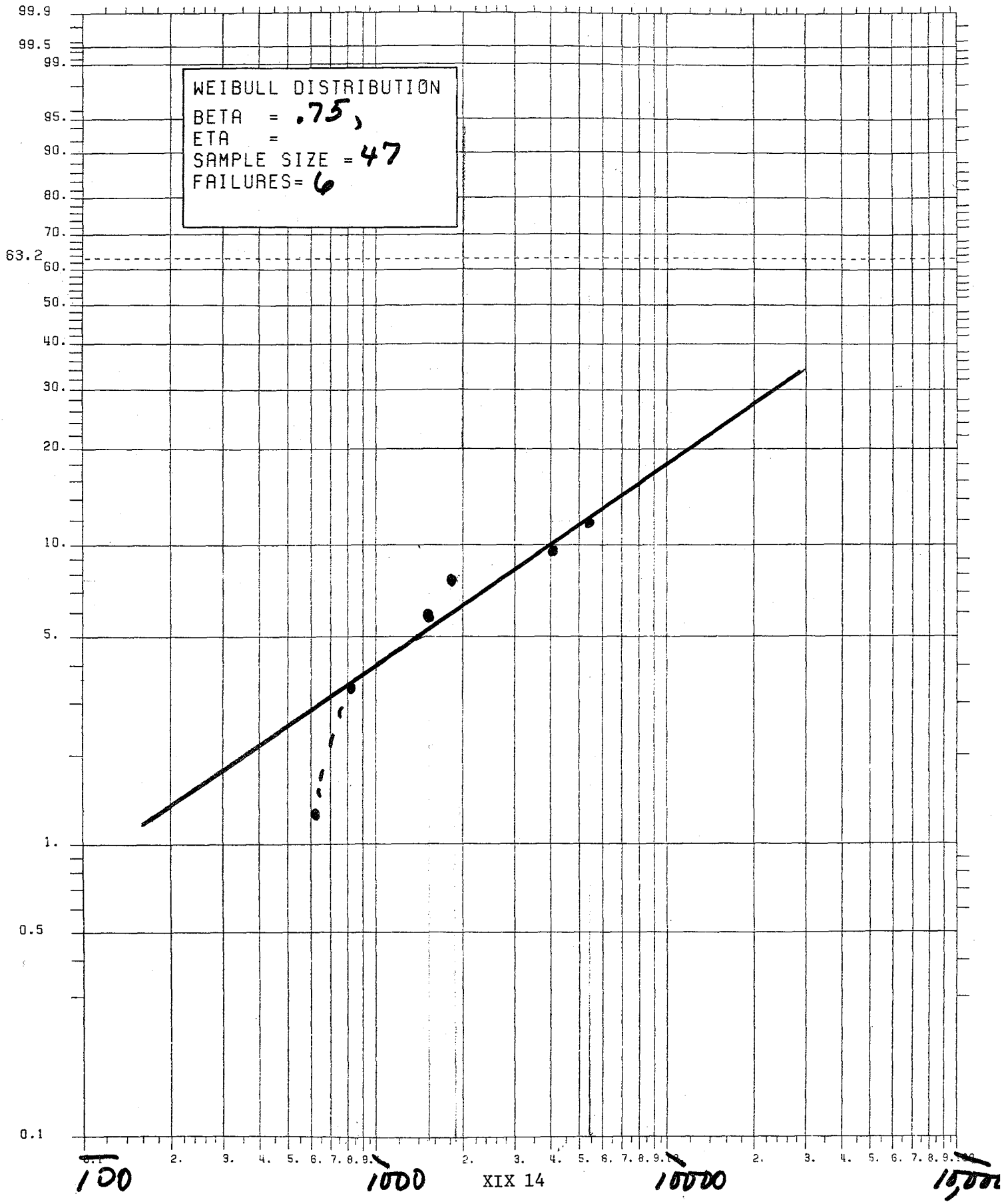
where each t is determined by measuring or its location.

t_1 = 622

t_2 = 850 (by measuring)

t_3 = 5460

FPOVA
needs to correction



FPOV A correction to Subtract 610

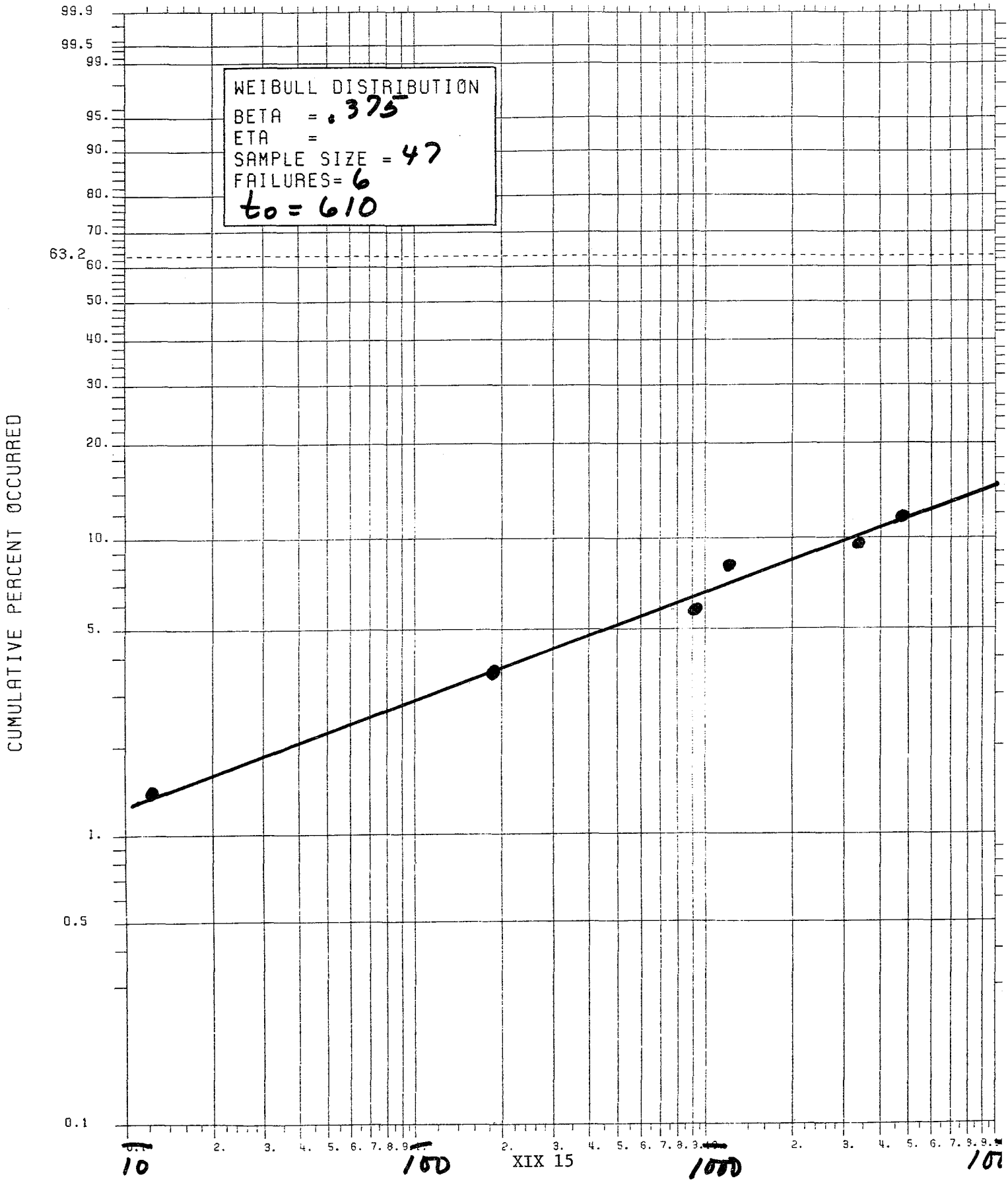


TABLE 4

Actuator	Number	Time(Sec)	Rank	Median Rank
MFVA	068-452	20	1	.774
MFVA	046-002	622	2	1.88
MFVA	070-490	932	3	2.99
MFVA	088-591	1012	4	4.90
MFVA	036-272	1593	5	5.20
MOVA	075-252	1778	6	6.31
MFVA	011-007	2035	7	7.41
MFVA	050-445	2698	8	8.52
MFVA	064-049	2946	9	9.62
MFVA	030-006	2968	10	10.73
MFVA	032-001	3325	11	11.84
MOVA	029-001	3703	12	12.94
MOVA	022-002	6725	13	14.05
MFVA	034-004	9866	14	15.15
MFVA	054-114	9882	15	16.26
MOVA	063-079	11590	16	17.37
MFVA	066-142	11590	17	18.47
MOVA	051-415	16643	18	19.58
MOVA	027-002	23834	19	20.69
MFVA	028-004	28094	20	21.79

Sample Size = 90 Number of Failures = 20

MFVA
MOVA
Together

20 sec not graphed
Two failure modes

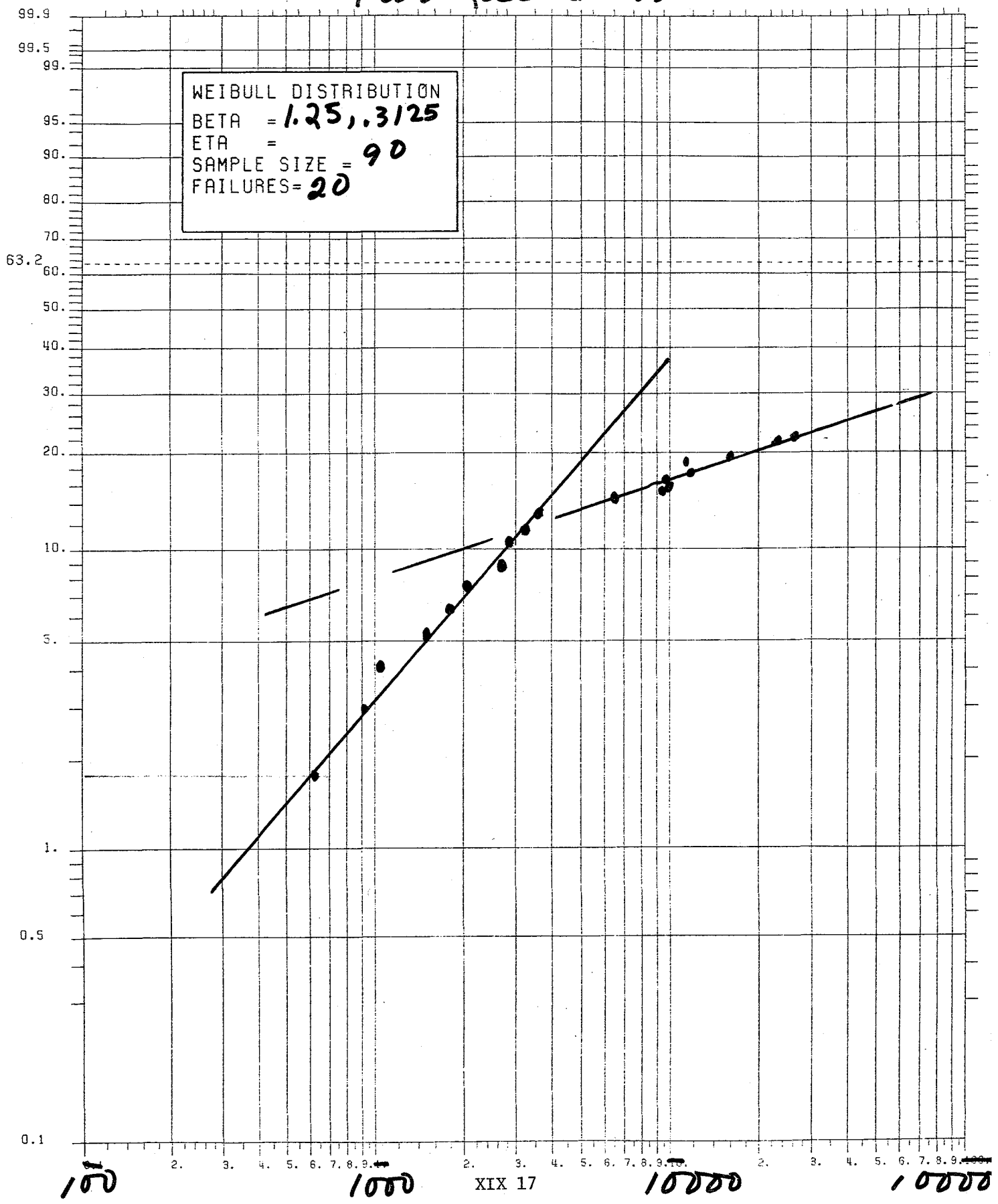


TABLE 5

Actuator	Number	Time (Sec)	Rank	Median Rank
MFVA	068-453	20	1	1.6
MFVA	046-002	622	2	3.9
MFVA	070-490	932	3	6.3
MFVA	088-591	1012	4	8.6
MFVA	036-272	1593	5	11.0
MFVA	011-007	2035	6	13.3
MFVA	050-445	2698	7	15.7
MFVA	064-049	2946	8	18.1
MFVA	030-006	2968	9	20.4
MFVA	032-001	3325	10	22.8
MFVA	034-004	9866	11	25.7
MFVA	054-114	9882	12	27.5
MFVA	066-142	11590	13	29.9
MFVA	028-004	28094	14	32.2

Sample Size = 42 Number of Failures = 14

MFVA
Sample Size 42
Two failure modes

20 sec is not graphed
Maybe outlier

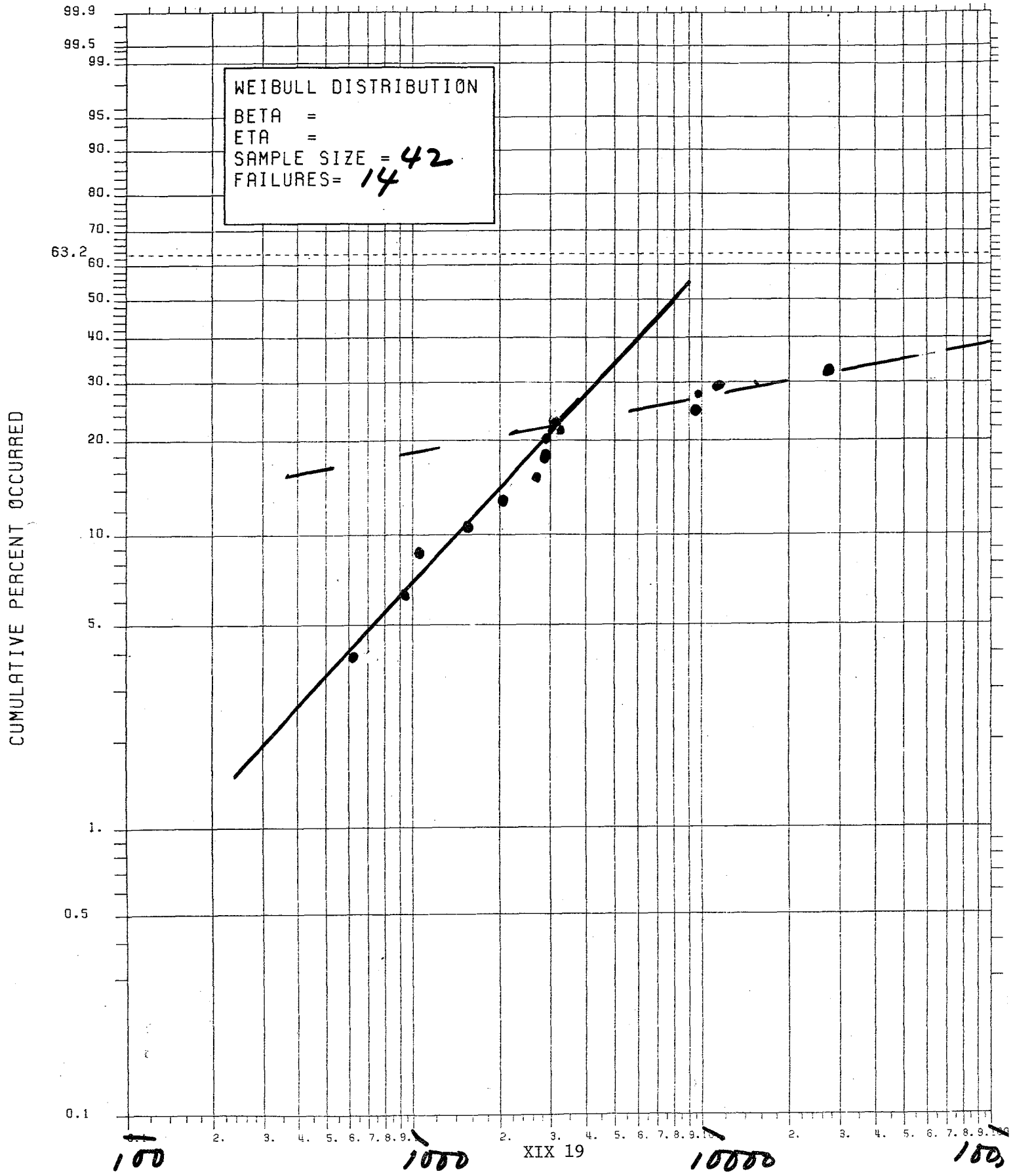


TABLE 6

Actuator	Number	Time (Sec)	Rank	Median Rank
MOVA	075-252	1778	1	1.4
MOVA	029-001	3703	2	3.4
MOVA	022-002	6725	3	5.5
MOVA	063-079	11590	4	7.5
MOVA	051-415	16643	5	9.6
MOVA	027-002	23834	6	11.7

Sample Size = 48 Number of Failures = 6

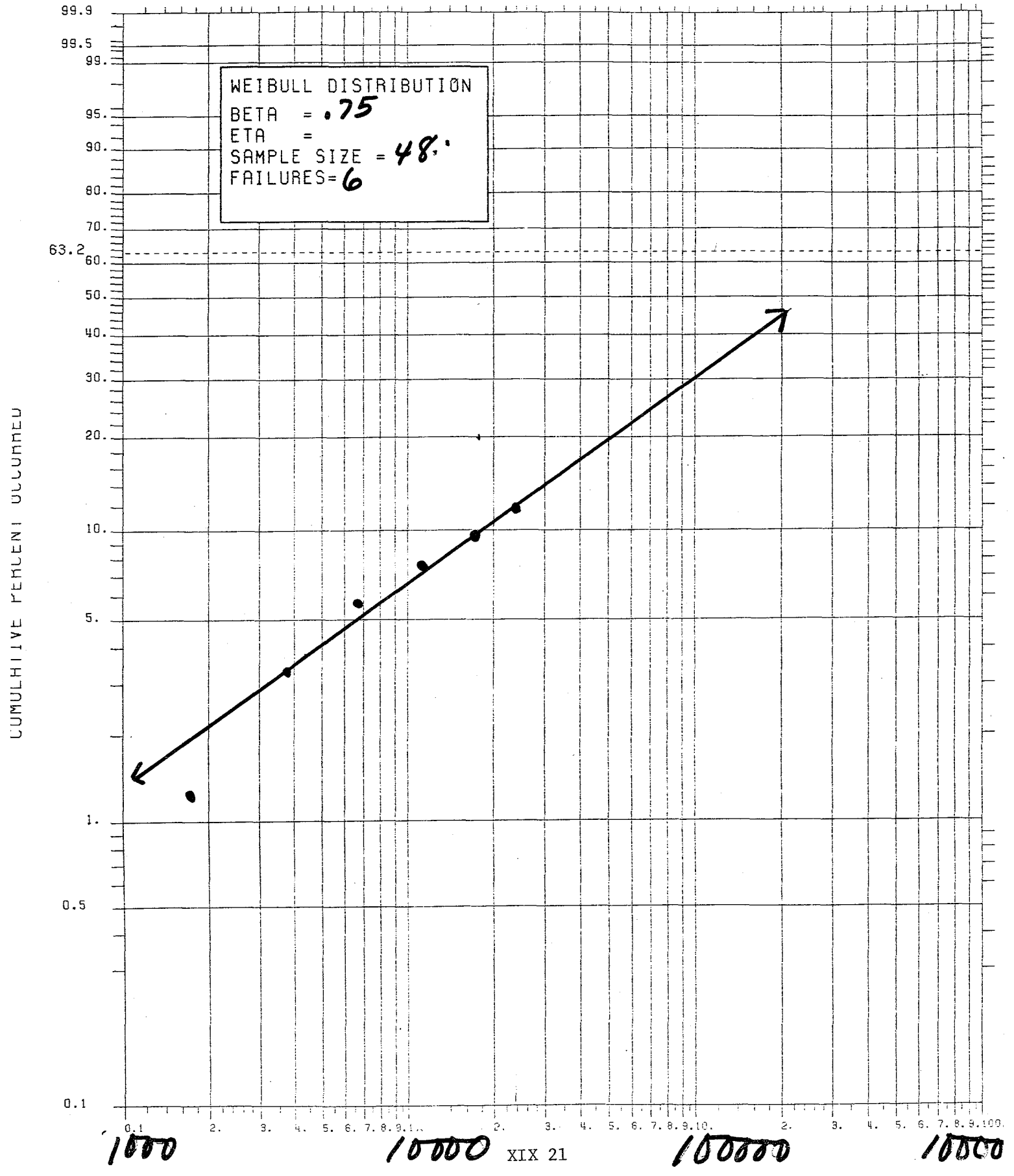
For Correction of t_o :

$t_1 = 1778$ $t_2 = 4500$ $t_3 = 19334$

Subtract $t_o = 1332$

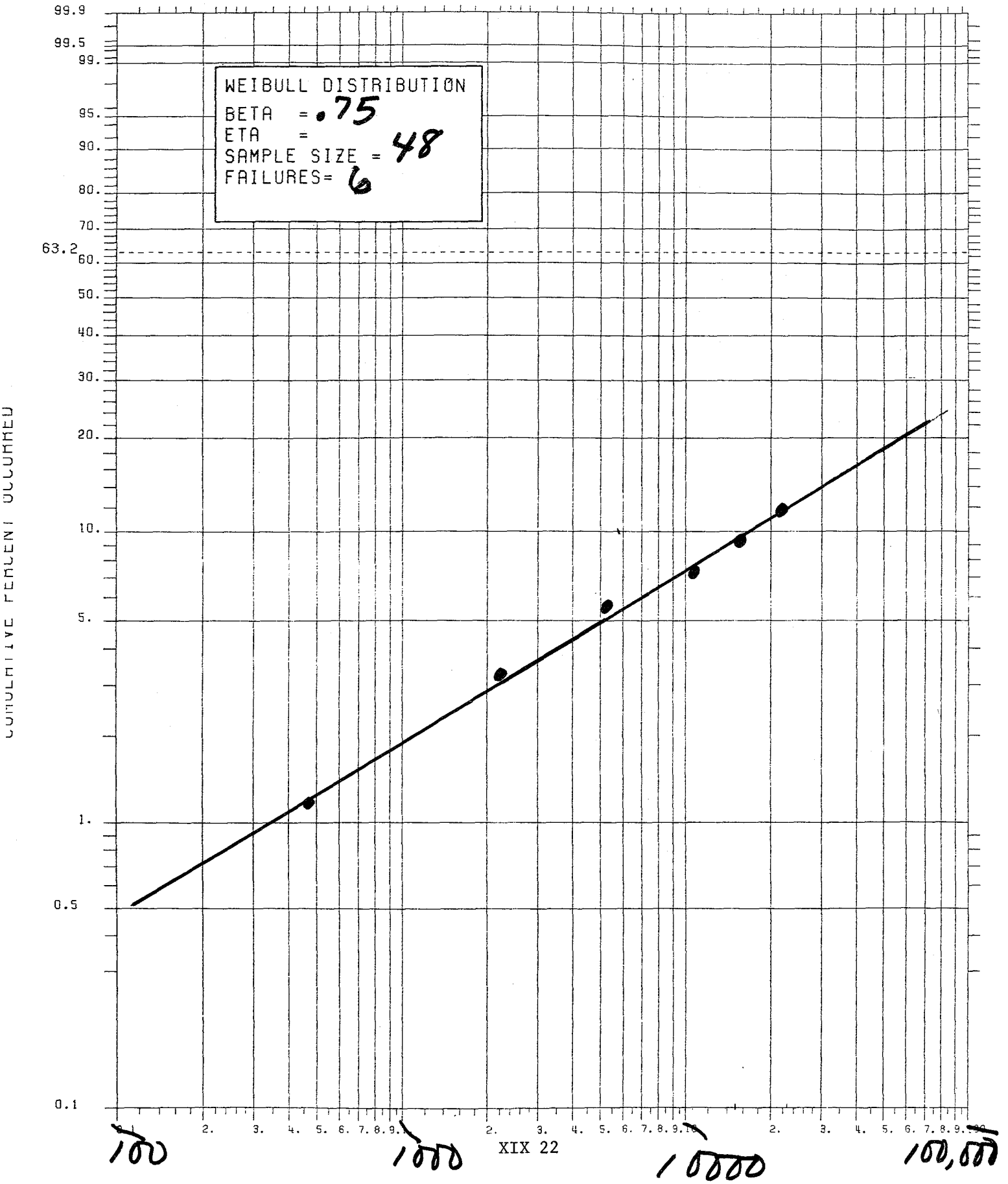
Time: 466, 2371, 5393, 10258, 15311, 22502

MOVA


WEIBULL DISTRIBUTION
BETA = .75
ETA =
SAMPLE SIZE = 48.
FAILURES= 6
CUMULATIVE PERCENT OCCURRED
99.9
99.5
99.
95.
90.
80.
70.
63.2
60.
50.
40.
30.
20.
10.
5.
1.
0.5
0.1
1000
10000
100000
1000000

NOVA
to correction

$t_o = 1332$
Subtract



BIBLIOGRAPHY

Books

Abernethy, R. B. et al. Weibull Analysis Handbook, United States Air Force, Wright-Patterson AFB, Ohio, 1983

Bain, Lee. Statistical Analysis of Reliability and Life-Testing Models Theory and Method, Marcel Dekker, N.Y. 1978

Bompas-Smith, J. H. Mechanical Survival, McGraw-Hill, St Louis, 1973

Carter, A. D. Mechanical Reliability, John Wiley & Sons, New York, 1972

Cramer, Harald. Mathematical Methods of Statistics. Princeton University Press, Princeton, 1946

Miller Irvin and Freund, John. Probability and Statistics for Engineers, Prentice-Hall, Englewood Cliffs, N.J., 1965

Nelson, Wayne. Applied Life Data Analysis John Wiley & Sons, N.Y. 1982

Report

Campbell, C. Warren. Convergence of Newton's Method for a Single Real Equation, NASA, June 1985

Rheinfurth, M. H. "Oxidizer Turbine (HPOTP) first Stage Blade Reliability Analysis," ED01 Marshall Space Flight Center, July 1985.

Weibull, Waloddi. "Estimationof Distribution Parameters by a Combination of the Best Linear Order Statistic Method and Maximum Likelihood," Wright-Patterson AFB, April 1967.

NASA/ASEE SUMMER FACULTY RESEARCH FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

STANDARDIZATION OF CARBON-PHENOLIC COMPOSITE TEST METHODOLOGY

Prepared by:	William B. Hall, Ph.D.
Academic Rank:	Professor
University and Department:	Mississippi State University Department of Chemical Engineering
NASA/MSFC:	
Division:	Non-Metallic
Branch:	Ceramics & Coatings
MSFC Counterpart:	Ron L. Nichols
Date:	August 23, 1985
Contract No:	NASA-NGT 010-008-021 The University of Alabama In Huntsville

STANDARDIZATION OF CARBON-PHENOLIC COMPOSITE TEST METHODOLOGY

BY

William B. Hall
Professor, Chemical Engineering
Mississippi State University
ASEE-NASA Fellow

Ron L. Nichols
Ceramics Unit, EH34
Marshall Space Flight Center
ASEE-NASA Counterpart

ABSTRACT

The objective of this study was to evaluate the residual volatiles, filler content, and resin flow test procedures for carbon-phenolic prepreg materials. The residual volatile test procedure was rewritten with tighter procedure control which was then evaluated by round robin testing by four laboratories on the same rolls of prepreg. Results indicated that the residual volatiles test was too operator and equipment dependent to be reliable, and it was recommended that the test be discontinued. The resin flow test procedures were rewritten with tighter procedure control, and it is now considered to be an acceptable test. It was recommended that the filler content determination be made prior to prepregging.

INTRODUCTION

Carbon-carbon and carbon-phenolic composites are the most common nozzle materials utilized in space vehicles. These materials have been developed and placed into service based on empirical information and evaluation. In an attempt to place this critical field on a more technical basis, Marshall Space Flight Center (MSFC) established a program entitled Solid Rocket Motor (SRM) Enhancement and Technology Development Program. The objectives of this program included the following; to implement a carbon-phenolic processing variable study, to expand the design and evaluation data bases, to enhance SRM quality and productivity, to solve current problems in SRM nozzle performance, and to improve SRM technology in general. The accomplishment of these objectives would certainly establish MSFC as the NASA lead center in SRM technology. If the improvements are made in the nozzle materials field, then new test methods and specifications are neeeded to measure the improvements and to define the necessary materials and processing to routinely obtain and keep these improved materials.

Objective

The objective of this study was to look at three test methods and specifications currently in use in SRM technology. The tests were to be evaluated on two basis; first, does the test method yield the desired results, and secondly, do these results give useful information in evaluating the products. The three tests were percent resin flow, percent residual volatiles and percent filler content. These tests were chosen as the first tests to evaluate because at the present time there are 45,000 lbs. of SRM prepreg out of one of these specifications and cannot be used. The value of this prepreg exceeds one million dollars, and in addition, this problem has created a shortage of acceptable prepreg. The problem is not critical at this time because procurement of prepreg followed original scheduling while actual prepreg usage fell below the predicted rate because of delays in nozzle fabrication and shuttle flight launch schedules.

INTRODUCTION

Carbon-carbon and carbon-phenolic composites are the most common nozzle materials utilized in space vehicles. These materials have been developed and placed into service based on empirical information and evaluation. In an attempt to place this critical field on a more technical basis, Marshall Space Flight Center (MSFC) established a program entitled Solid Rocket Motor (SRM) Enhancement and Technology Development Program. The objectives of this program included the following; to implement a carbon-phenolic processing variable study, to expand the design and evaluation data bases, to enhance SRM quality and productivity, to solve current problems in SRM nozzle performance, and to improve SRM technology in general. The accomplishment of these objectives would certainly establish MSFC as the NASA lead center in SRM technology. If the improvements are made in the nozzle materials field, then new test methods and specifications are neeeded to measure the improvements and to define the necessary materials and processing to routinely obtain and keep these improved materials.

Objective

The objective of this study was to look at three test methods and specifications currently in use in SRM technology. The tests were to be evaluated on two basis; first, does the test method yield the desired results, and secondly, do these results give useful information in evaluating the products. The three tests were percent resin flow, percent residual volatiles and percent filler content. These tests were chosen as the first tests to evaluate because at the present time there are 45,000 lbs. of SRM prepreg out of one of these specifications and cannot be used. The value of this prepreg exceeds one million dollars, and in addition, this problem has created a shortage of acceptable prepreg. The problem is not critical at this time because procurement of prepreg followed original scheduling while actual prepreg usage fell below the predicted rate because of delays in nozzle fabrication and shuttle flight launch schedules.

Test Methodology Studies

Filler Content: At the present time, there is no accurate method to measure carbon filler content in carbon-phenolic prepreg material. Two methods currently utilized are the Soxhlet extraction method and the DMF wash method. Both systems remove some of the fabric and do not remove all of the carbon filler. The amount of filler left depends on the type, size, and shape of the carbon filler material. In general, the DMF wash method results in values approximately 2% higher than those obtained by the Soxhlet extraction method, as shown in the data given below:

Percent Filler in C.C. Phenolic Lot C01656*

Roll #1	DMF	Soxhlet	Roll #2	DMF	Soxhlet
Sample 1	16.6	15.2	Sample 1	16.2	14.1
Sample 2	16.8	15.3	Sample 2	15.4	14.1
Sample 3	16.0	15.7	Sample 3	14.8	13.9
Sample 4	17.2	14.5	Sample 4	16.6	14.2
Sample 5	15.8	14.2	Sample 5	15.9	13.9
Sample 6	15.7	14.3	Sample 6	15.6	13.9
Avg. X =	16.5	14.9		15.8	14.0

* MTI data

It should be possible to prepare small batches of accurately measured filler/resin ratio, and add this mix to a accurately weighed cloth and prepare calibration curves for both the DMF wash and Soxhlet extraction methods. However, a better control could be acheived by constant monitoring the filler/resin ratio just prior to its being applied to the carbon cloth. Since the filler content does not change with aging, the initial determination should be sufficient for the life of the prepreg.

Percent Resin Flow: The test procedure for percent resin flow is given in Appendix D. This test is important because it assures that the prepreg can be tape wrapped and properly cured. The percent resin flow not only is a measure of staging of the resin, but of total resin content. The content is a function of the cloth itself in that the amount retained is a function of several characteristics of the cloth and the filler/resin mixture. However, there is a wide variance in the percent resin flow data, especially when tested in different laboratories. The initial explanation for this variance was the delay in closing the press, as step c of appendix D just specifies immediately. Closing time versus percent resin flow was determined for three different rolls of prepreg with the following results:

		Precent Resin Flow		
Delay in Closure	Roll	1	2	3
0 sec		29.6	18.7	31.5
20 sec		31.1	19.1	30.0
40 sec		30.0	20.1	--
60 sec		29.3	22.1	30.7
120 sec		29.9	19.4	24.8

The closure time with 0 seconds delay was 15 seconds, so the total closure time was delay time plus 15 seconds. It can be seen that total closure times of 75 seconds and less had a negibile effect on the percent resin flow.

Percent Residual Volatiles: The major difficulty with the original residual volatiles test method was the variance in results between the two prepreg suppplier, U.S. Poly and Fiberite, and the user of the prepreg, Thiokol. The problem was compounded when one lab obtained results that were high but still within specs and the other lab obtained results that were too high, i.e., out of specification. During a meeting at MSFC with U.S. Poly, Fiberite, Thiokal and MSFC represented, the original test method (Appendix A) was rewritten to tighten the test procedure. The reason that it was felt that it was necessary to rewrite the test method is shown in Figure 1. Although each laboratory was following the same test procedure, there was significant difference in the actual time-temperature-pressure sequence in the prepartion of test specimen. This would give different amounts of staging and resin flow yielding significant different results. In an attempt to solve this problem, the test procedure was modified to tighten up the test method. The agreed upon new test method is shown graphically in Figure I and written in Appendix B. The four labs then agreed to a round robin testing of the same prepreg rolls to evaulte the precision and the ability of each lab to duplicate each others results. The four labs, U.S. Poly, Thiokal, Fiberate and MSFC then tested the same seven rolls of prepreg utilizing the rewritten test method. The results of the test are given below:

ROUND ROBIN TESTING OF RESIDUAL VOLATILES IN SELECTED MATERIALS

MATERIAL USP -2115-

40	46	56
F-1.21	F-1.35	F-1.52
M-2.23	M-1.85	M-1.89
P-2.47	P-2.08	P-2.21
T-2.50	T-2.14	T-2.63

FIBERITE 1029-

4A	19A	22A	26A
F-0.99	F-1.43	F-1.04	F-1.44
M-1.31	T-1.69	M-1.52	M-1.51
T-1.44	M-1.76	P-1.88	T-1.94
P-1.79	P-2.29	T-1.90	P-2.21

LAB KEY, F=FIBERITE. M=MARSHALL. P=USP. T=THIOKOL

Data from MSFC testing is shown in Figure 2, indicating that the precision was excellent, but as shown in the above data , the ability of the labs to reproduce the same results did not improve. It is the opinion of the author that the major reason for this difference in lab agreement is difference in the equipment utilized in the testing. Different presses, drying ovens, and bleeder cloth could result in different amount a resin left in the specimen. In addition, the last time any of the presses and/or ovens were calibrated is also an unknown.

After the round robin test were completed, the general consensus was that it appeared that it was not possible under the current laboratory procedures

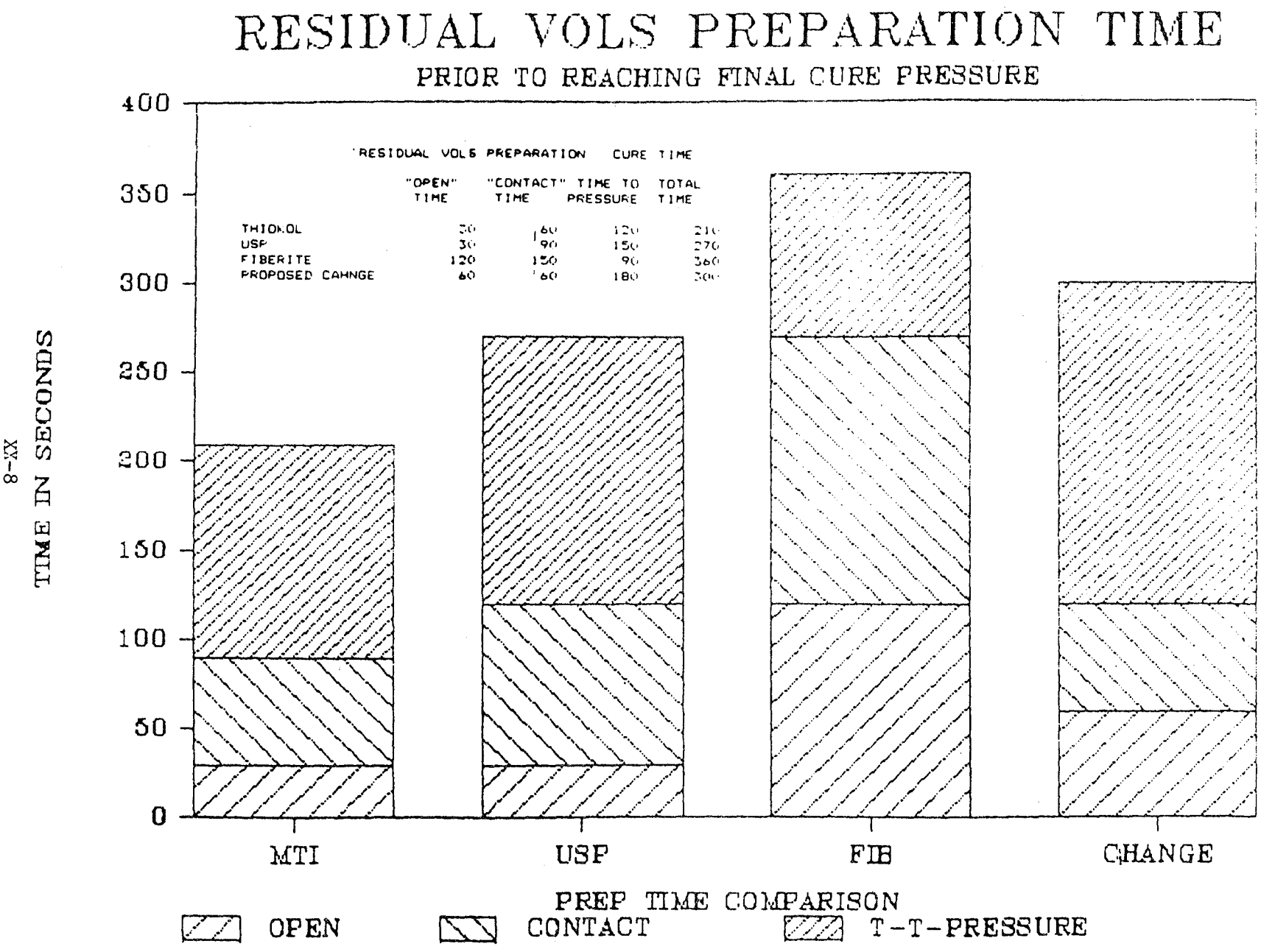


Figure 1. Preparation Time for Residual Volatiles Test Specimen

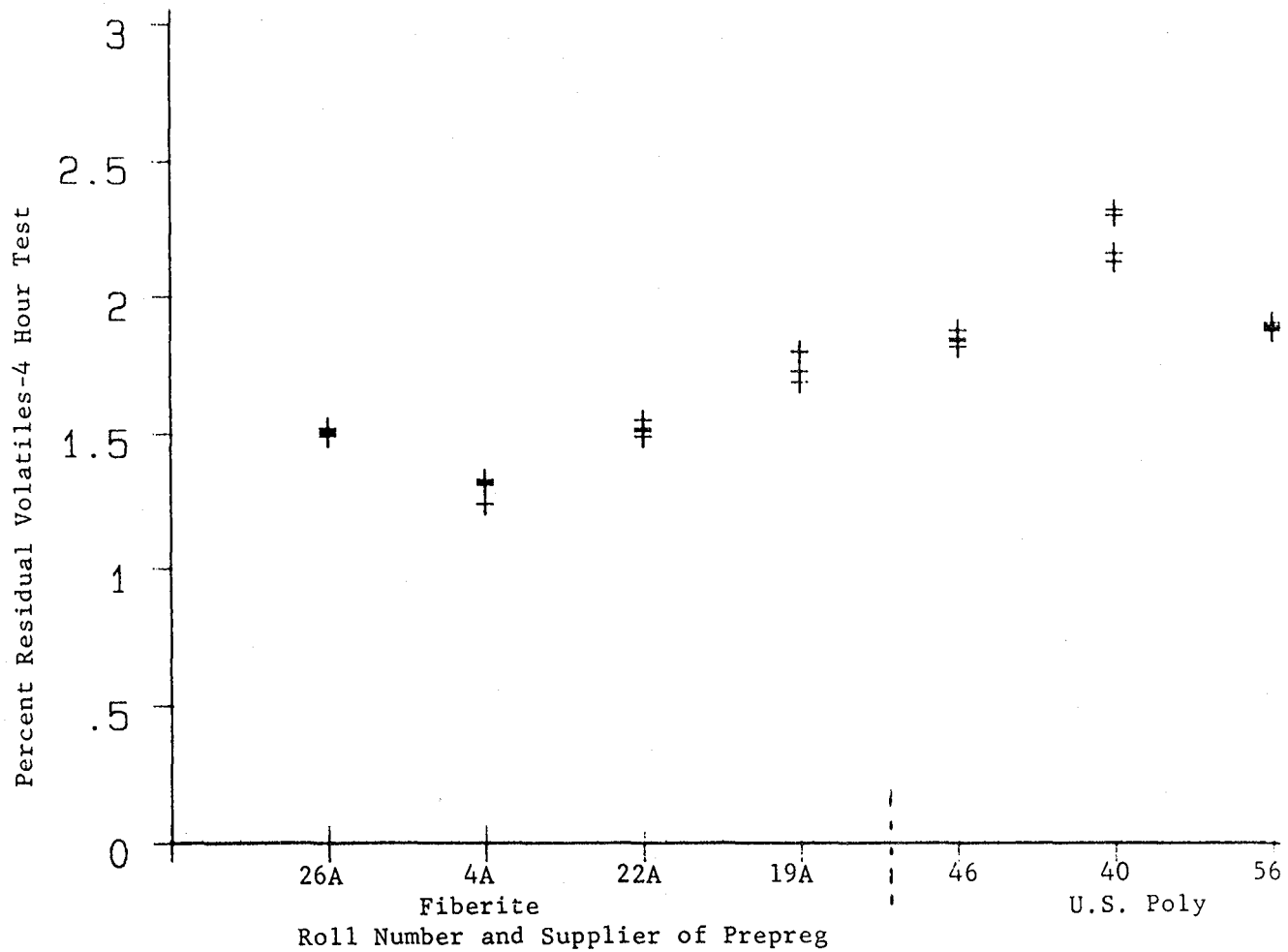


Figure 2. MSFC Residual Volatiles Test Values for Round Robin Test

to obtain reproducible results utilizing the residual volatiles test procedure. Therefore, since the first objective of the basic policy was not met, the second overall objective was not addressed. It is the authors opinion that the second basic policy objective was also not met. The minimum additional information needed to meet the second objective would be to identify and quantify the residual volatiles.

A new volatile content procedure was established to circumvent the above problems, and this test procedure is given in Appendix C. One change in the procedure was to change the phase "place specimen in oven" to "suspend specimen in oven" since laying the specimen on racks at 325 degree F causes sticking and excessive resin loss. A round robin evaluation between the same four labs for one roll of prepreg gave the following results:

Lab	Percent Volatiles (Average of 12 Specimens)	Variance
Fiberite	4.3	0.20
MTI	4.8	0.21
U. S. Poly	4.8	0.41
MSFC	4.9	0.30

The data is also shown graphically in Figure 3. It is noted that Fiberite was the lab with the lowest value as it was with all the rolls evaluated in the percent residual volatiles round robin, eight out of eight rolls tested. This would indicate some constant function such as lower than indicated temperature causing the difference in values between Fiberite and the other labs. The other three labs were extremly close so it does appear that the percent volatiles test can be duplicated in the various labs. It is still considered necessary, for the second objective or the basic policy for this study to be satisfied, to identify and quantify the volatiles coming from the specimen.

It is known that carbon-phenolic composite pick up moisture when exposed to humid conditions. Both cured and uncured (prepreg) samples were exposed to 100 percent relative humidity for 30 days. The weight gain versus exposure time curves for both of these materials are shown in Figure 4 (uncured) and Figure 5 (cured). The uncured curve is interesting in that it seems to indicate a weight gain for the first week followed by a weight loss the second week and then a gradual weight gain. The high value specimen on the 21st day is not the same high value specimen on the 29th day. This curve indicates varying reactivities and could be the area of a mare complete study. The curve for the cured and specimen indicates a gradual increase in weight for the entire exposure time of 30 days. The values remain approximately the same for the 36 day exposure time, so it would appear that a constant weight was achieved at 29-30 days of exposure time to 100 percent humidity conditions. The normal testing for residual volatiles that had picked up extra moisture was to dessicate for a minimum of 24 hours or a minimum of 72 hours on retest. The weight loss versus time in desicator is shown in Figure 6 and indicates an increase in weight loss continuing up to the 72 hours. Therefore, time in desicattor should be specified in a plus or minus frame instead of a minimum value. The weight loss versus exposure time in humidity champer is shown in Figure 7 indicating fairly constant loss of weight in the desiccator regardless of the time in the humidity chamber.

After the 72 hours in the desicator, percent residual volatiles were

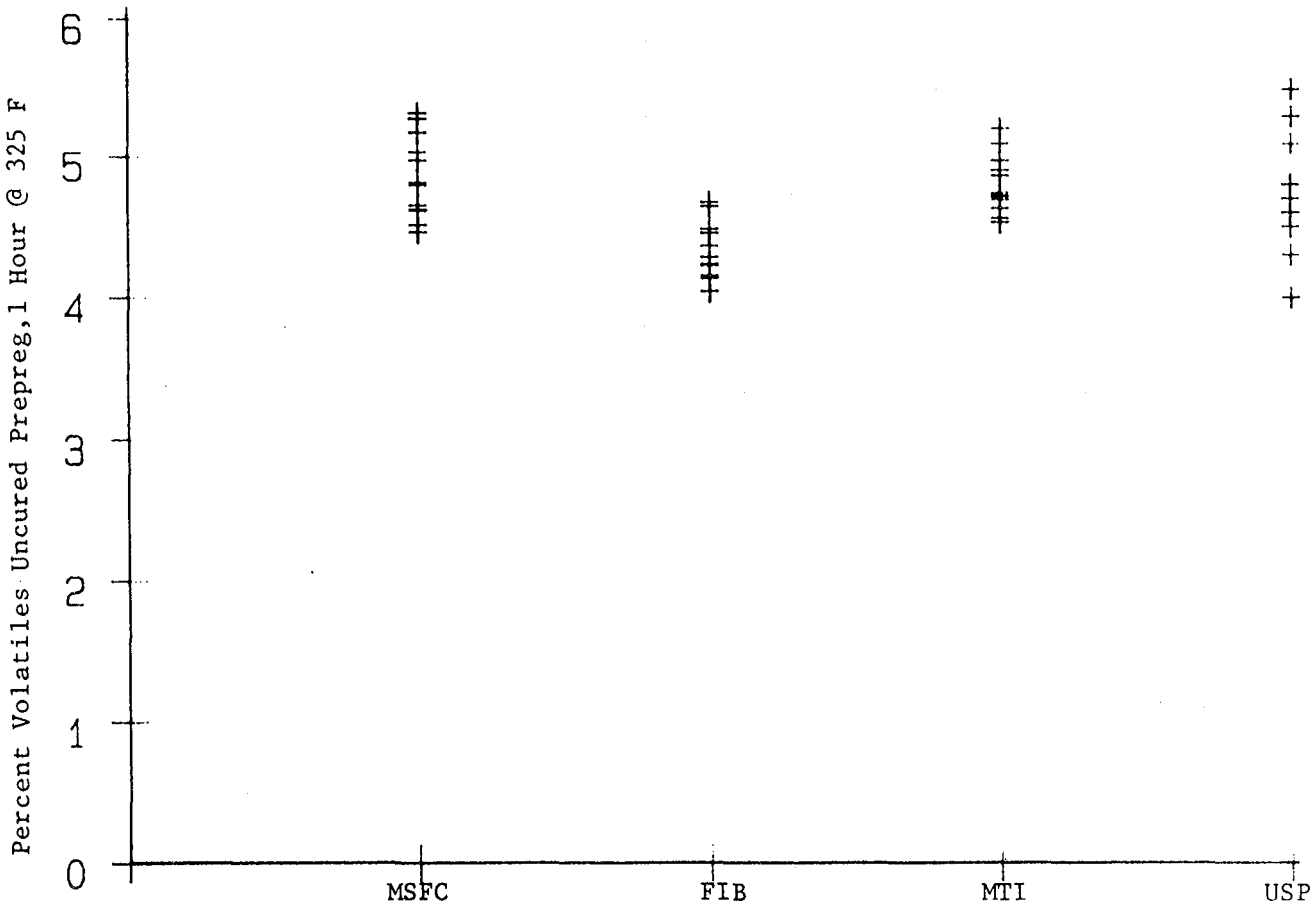


Figure 3. Values for Percent Volatiles from Round Robin Test

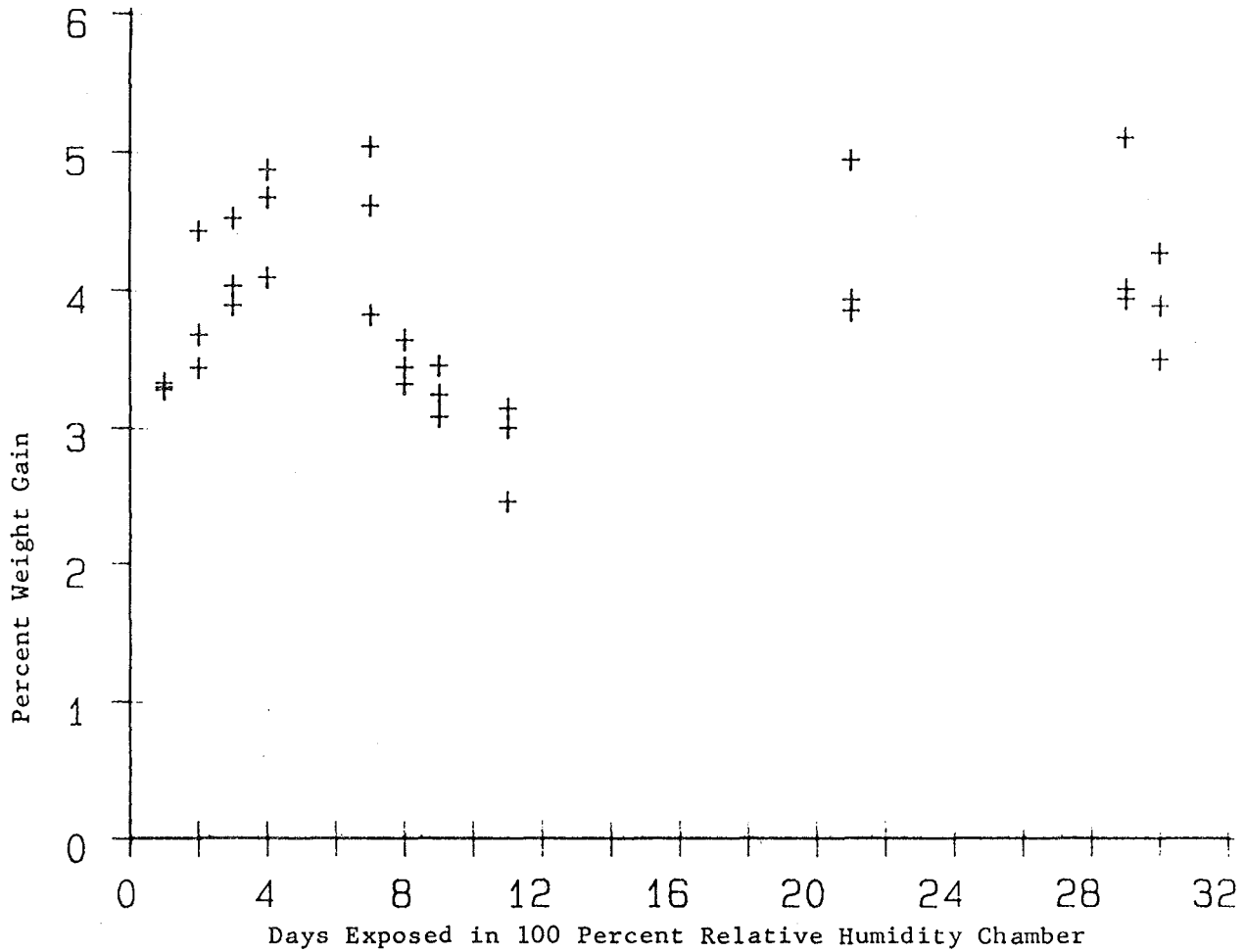


Figure 4. Weight Gain versus Time in Humidity Chamber for Uncured Prepreg

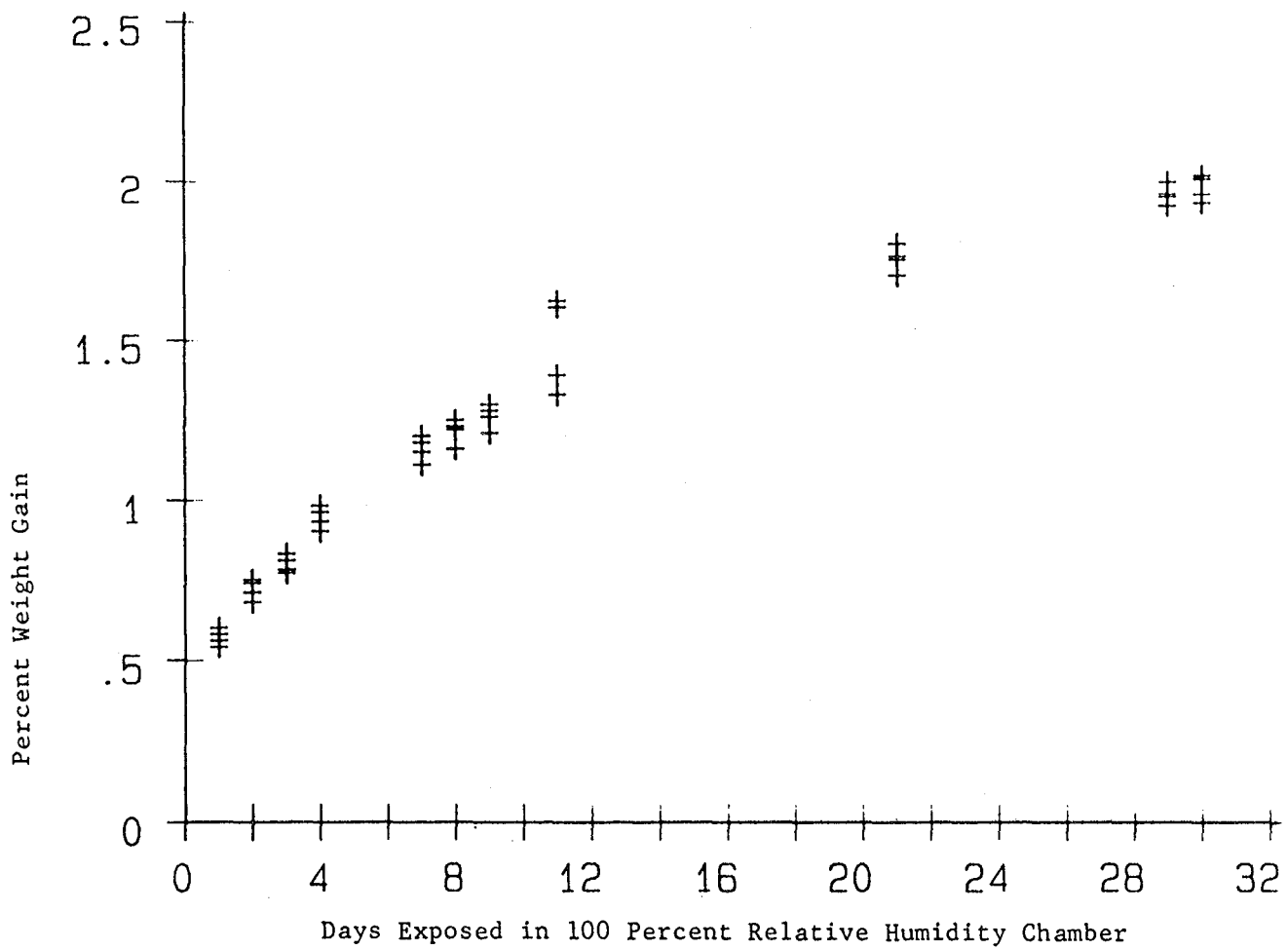


Figure 5. Weight Gain versus Time in Humidity Chamber for Cured Specimen

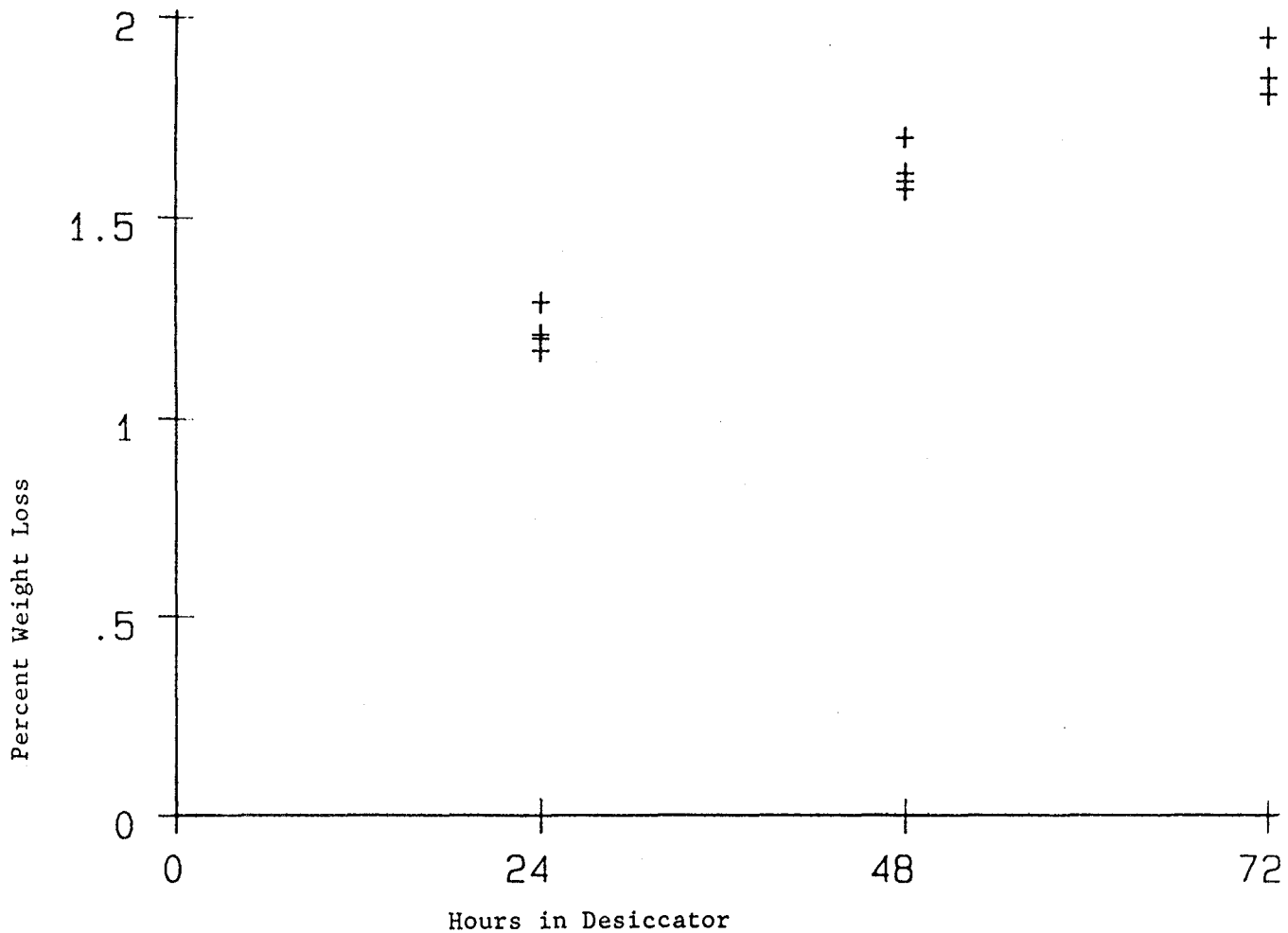


Figure 6. Weight Loss versus Time in Desiccator

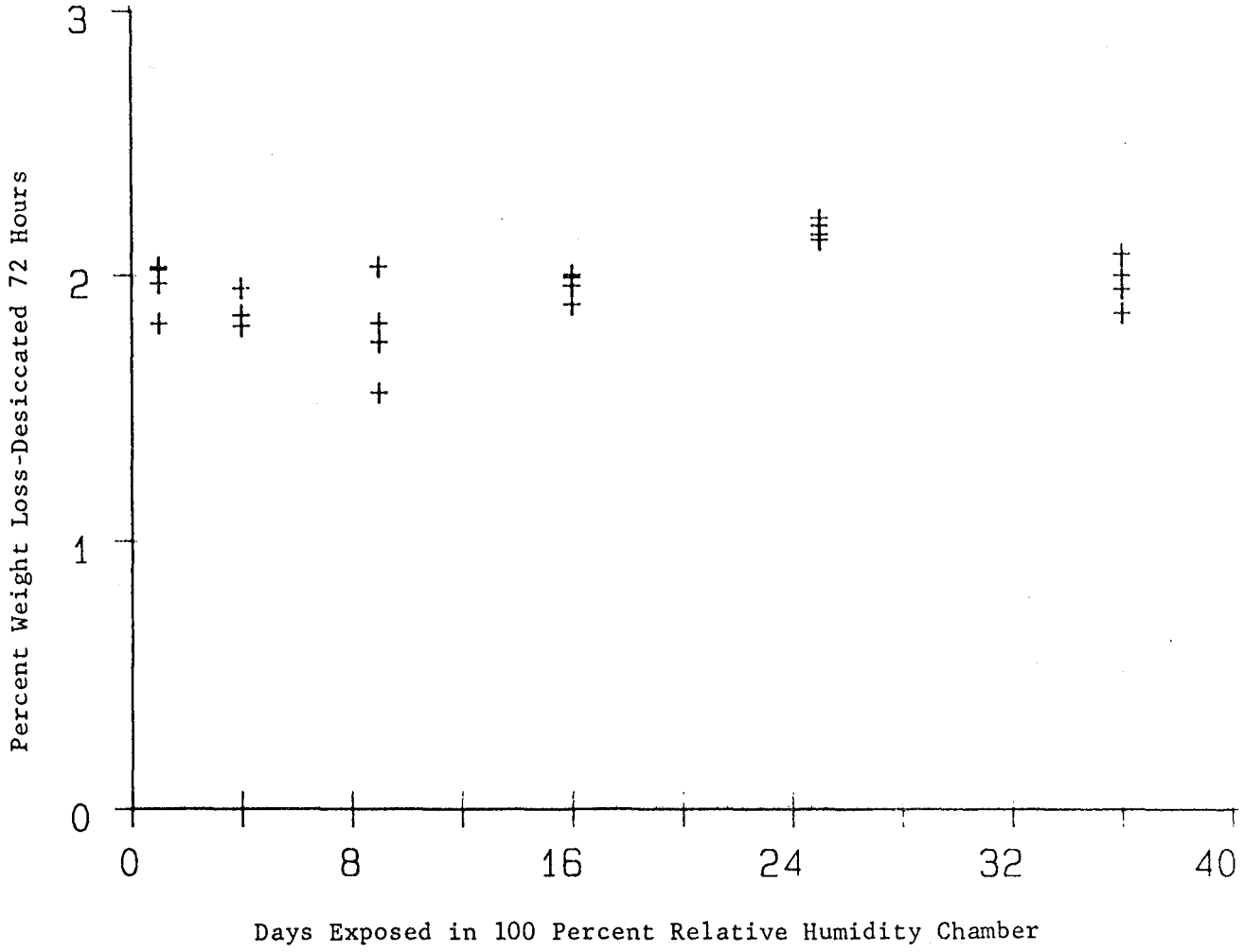


Figure 7. Weight Loss after 72 Hours in Desiccator versus Time in Humidity Chamber

determined for the exposed specimens. The results of this study is shown in Figure 8, indicating a slight increase in percent residual volatiles with an increase in exposure time in the humidity chamber. This was expected since Figure 5 indicates increase in weight versus exposure time while Figure 7 indicates essentially constant weight loss in the desiccator.

The last area studied was percent residual volatiles versus oven curing time. Figure 9 shows the percent residual volatiles for a oven period up to 48 hours, while Figure 10 extends the time to 360 hours. The important points in these two curves are that there is about one percent higher values when tested at 24 hours instead of 4 hours, and that constant values are achieved after 100 hours. It is the authors opinion that the specifications hould require the 24 hours oven time and adjust the percent residual volatiles value as required.

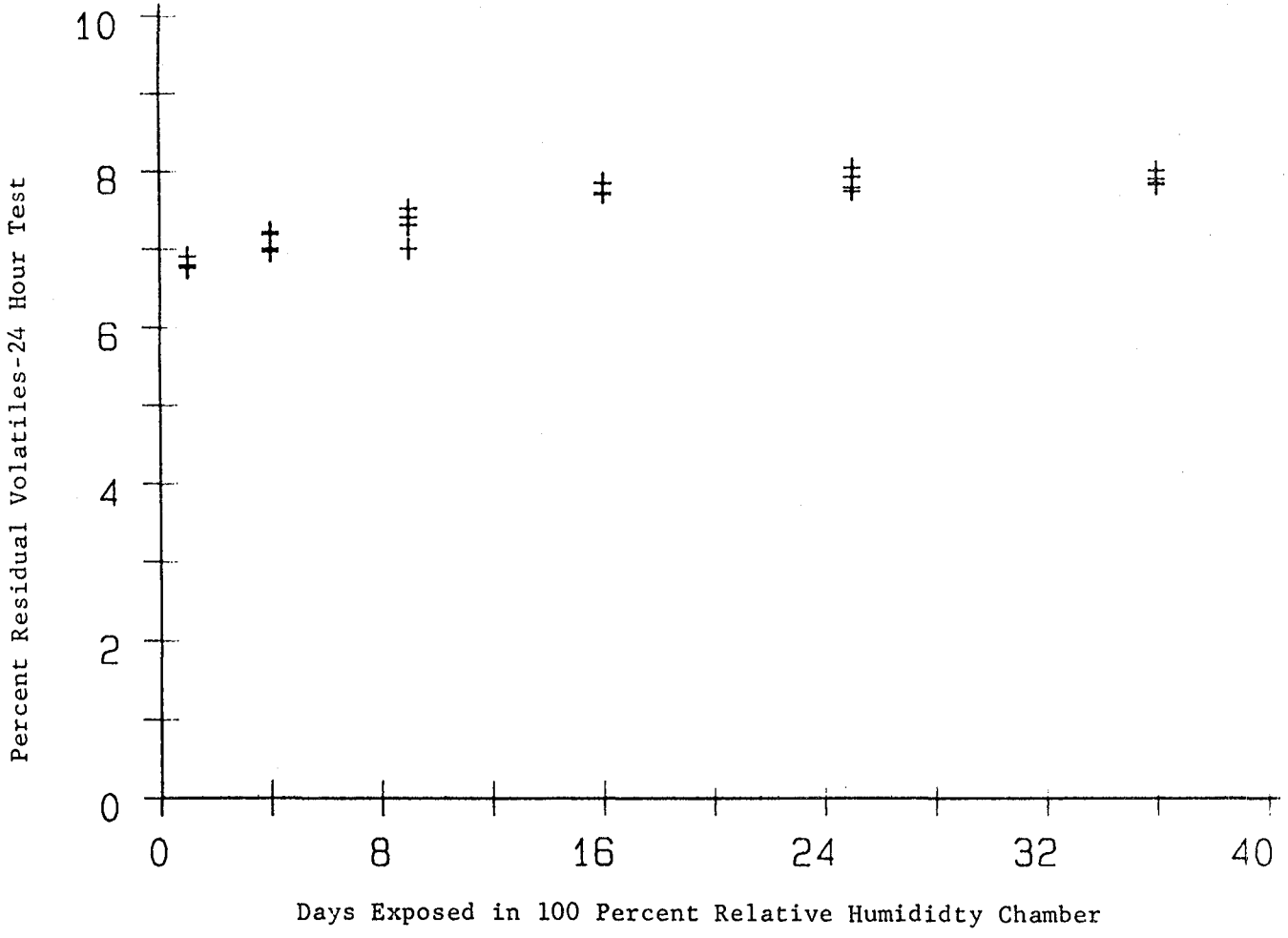


Figure 8. Percent Residual Volatiles versus Time in Humidity Chamber

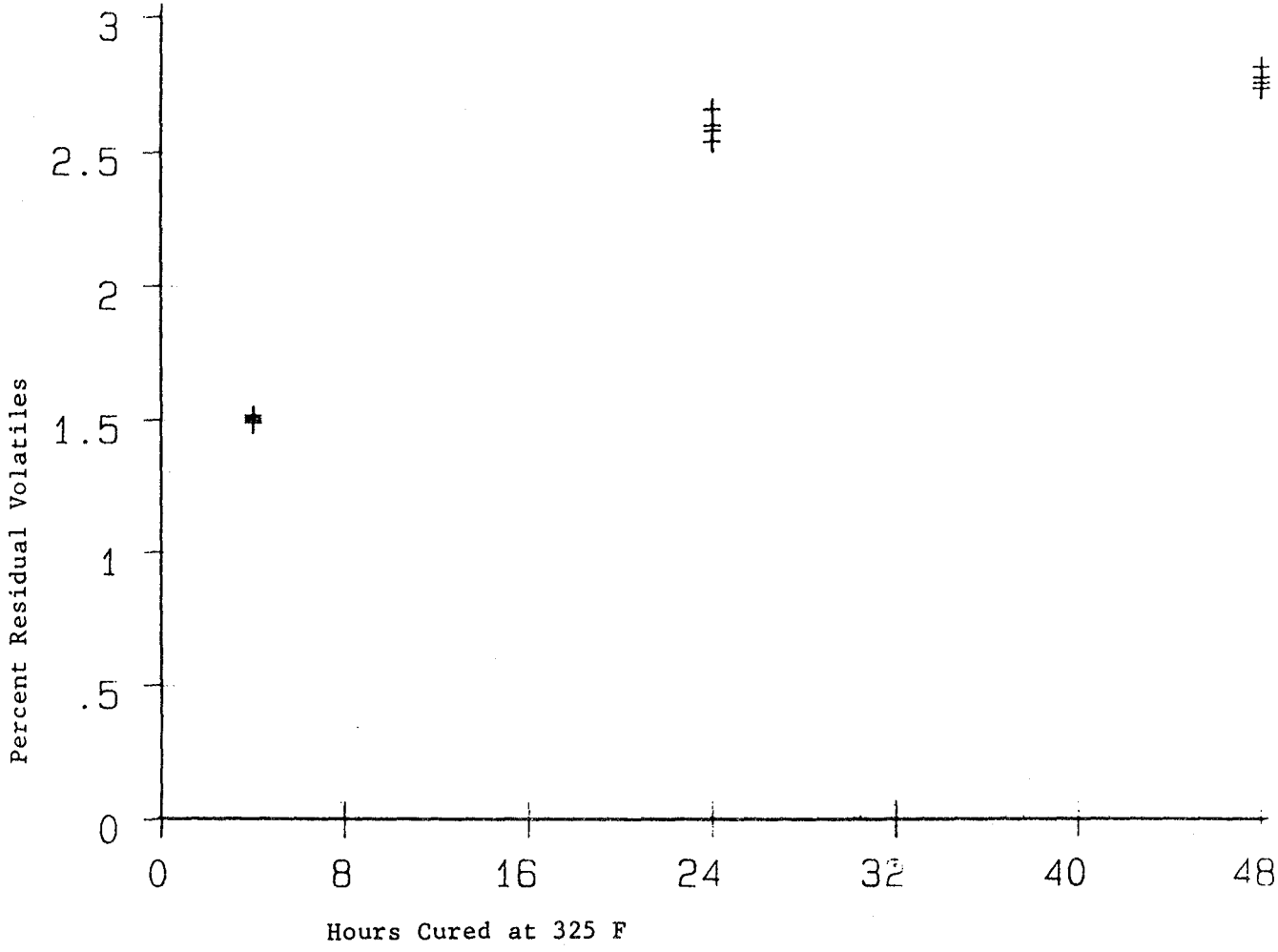


Figure 9. Percent Residual Volatiles versus Hours Cured at 325 F

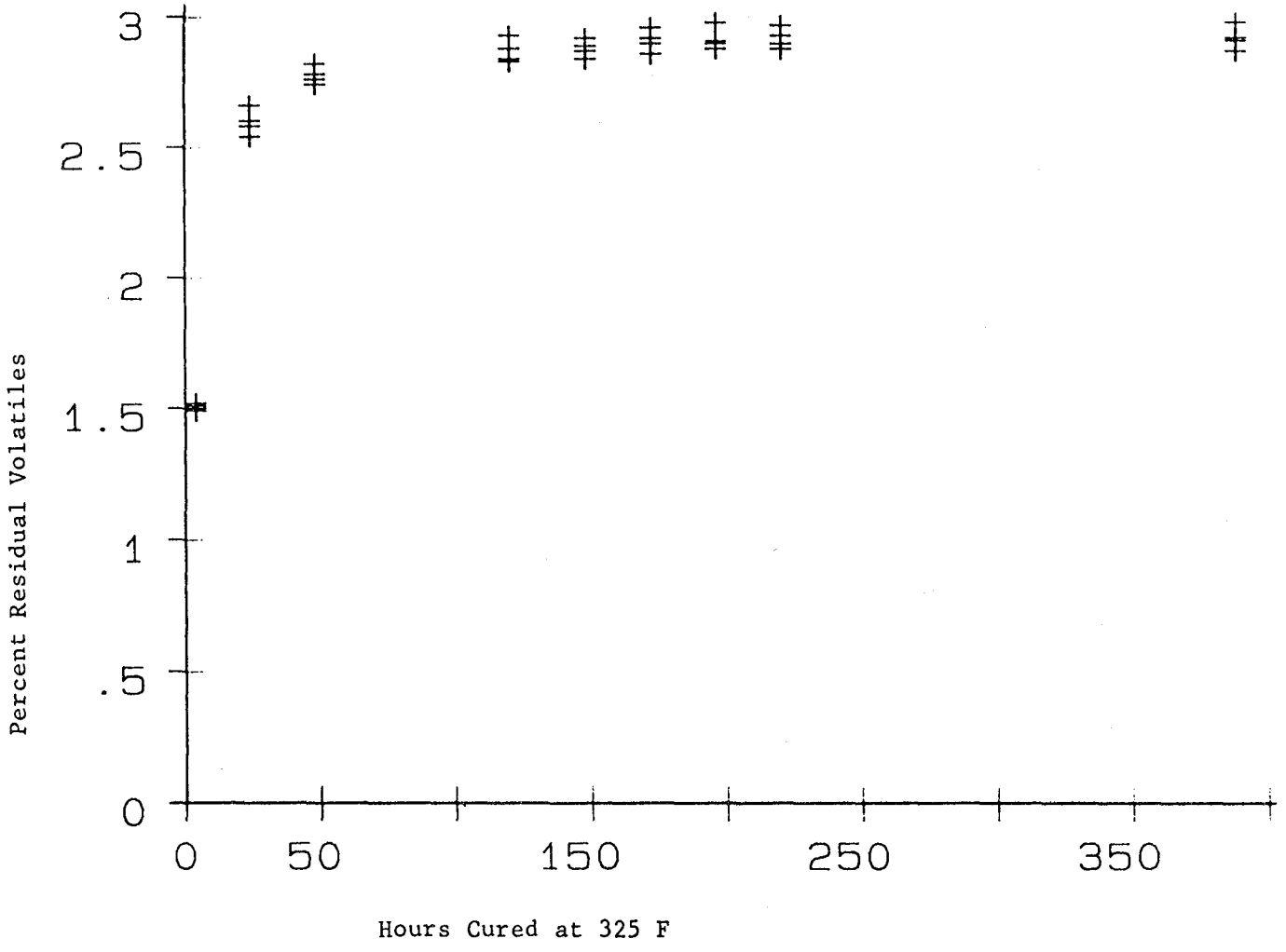


Figure 10. Percent Residual Volatiles versus Hours Cured at 325 F

CONCLUSIONS

Conclusions: Three standard test utilized in the carbon-phenolic composite industry were studied. These three test were percent residual volatiles, percent filler, and percent resin flow. The percent resin flow test was found to be acceptable in the present form, with the addition of one change. The change was to substitute for "apply press load immediately" to the following "apply press load in less than 30 seconds". The percent filler is determined at present by two alternative procedures, both yeilding different and incorrect results. It is recommended that this test be discontinued at the prepreg state and be conducted on the resin-filler mixture prior to prepregging. The percent residual volatiles test was determined to be to dependent on lab procedure for reproducibility and it is recommended this test be dropped and a now test for percent volatiles be utilized.

Appendix A

Orginal Residual Volatiles Test Procedure

Residual Volatile Content: The residual volatile content of the cured material shall be determined in accordance with the following:

a. Cut sufficient plies to produce a test panel 0.250 ± 0.050 thickness by 4 ± 0.125 by 4 ± 0.125 inches. Cure the panel as follows:

1. Place the stack of plies into a press preset at 325 ± 10 degrees F.

2. Apply contact pressure for 30 ± 5 seconds and dump pressure for 2 cycles before slowly applying maximum pressure. (The material shall not be prestaged.)

3. Increase the pressure slowly to 1000 ± 50 psi, allowing the resin to stage (see 6.3.1) to minimize flash at ply edges.

4. Hold the pressure and temperature for 120 ± 15 minutes.

5. Decrease the temperature and pressure to ambient.

The outer 1/2 inch of the test panel shall not be used for preparation of specimens.

b. Cut a specimen 1.000 ± 0.050 inch by 1.000 + 0.050 inch panel thickness from the center 2 inch by 2 inch section of the test panel. Top and bottom molded surfaces shall not be machined. Wipe the specimen clean using MEK and allow to air dry 20 minutes minimum before testing.

c. Place the specimen in a desiccator and desiccate for 18 hours minimum.

d. Weigh the specimen to the nearest 0.01 gram and record as W1.

e. Place the specimen in an air circulating oven, preheated and stabilized at 325 ± 10 degrees F for approximately 30 minutes; condition the specimen at a temperature of 325 + 10 degrees F for 24 hours minimum.

f. Remove the specimen from the oven and cool in the desiccator for approximately 30 minutes or until the specimen reaches room temperature.

g. Reweigh the specimen to the nearest 0.01 gram. Record as W2.

h. Calculate the percent residual volatiles as follows:

$$\text{Percent volatiles} = \frac{W1 - W2}{W1} \; 100$$

Where: W1 = original weight of specimen in grams

W2 = devolatilized weight of specimen in grams

i. Report residual volatiles to the nearest 0.01 percent.

Appendix B

Proposed Residual Volatiles Test Procedure

Residual Volatile Content: The residual volatile content of the cured material shall be determined in accordance with the following:

a. Cut 18 plies to produce a test panel 4 ± 0.125 by 4 ± 0.125 inches. Cure the panel as follows:

a. Cover top and bottom of lay-up with one layer of non-porous release film

b. Place the stack of plies into a press preset at 325 ± 10 degrees F.

c. Insert into the press and immediately close to contact pressure (close time ten seconds or less)

d. Hold for 30 plus or minus 5 seconds (at "contact" pressure)

e. Open the press to 1 - 2 inches and allow to "dwell" for 20 plus or minus 5 seconds

f. Immediately close to contact pressure and hold for 30 plus or minus 5 seconds.

g. Open the press to 1 - 2 inches and allow to dwell for 20 plus or minus 5 seconds.

h. Immediately close to contact pressure and slowly increase the pressure to 1000 - 1200 psi over a period of 180 plus or minus 30 seconds.

i. Total prep time is approx 300 seconds (assuming 10 seconds close times)

j. Cure time will be 120 - 135 minutes

k. Decrease the temperature and pressure to ambient.

l. The outer 1/2 inch of the test panel shall not be used for preparation of specimens.

m. Cut a specimen 1.000 + 0.050 inch by 1.000 + 0.050 inch by panel thickness from the center 2 inch by 2 inch section of the test panel. Top and bottom molded surfaces shall not be machined.

n. Place the specimen in a desiccator and desiccate for hours 18 hours minimum.

o. Weigh the specimen to the nearest 0.01 gram and record as W1.

p. Place the specimen in an air circulating oven, preheated and stabilized at 325 ± 10 degrees F for approximately 30 minutes; condition the specimen at a temperature of 325 ± 10 degrees F for 4 hours minimum.

q. Remove the specimen from the oven and cool in the desiccator for

approximately 30 minutes or until the specimen reaches room temperature.

r. Reweigh the specimen to the nearest 0.01 gram. Record as W2.

s. Calculate the percent residual volatiles as follows:

$$\text{Percent volatiles} = \frac{W1 - W2}{W1} 100$$

where W1 = original weight of specimen in grams

W2 = devolatilized weight of specimen in grams

t. Report residual volatiles to the nearest 0.01 percent.

Appendix C

Proposed Volatile Content Test

The volatile content of 12 specimens taken randomly from the uncured material shall be determined in accordance with the following:

a. Cut a 16 ± 2 square inch specimen.

b. Weigh the specimen to the nearest 0.01 gram (W1).

c. Place specimen in a recirculating oven preheated and stablized to 325 ± 10 degrees F for 60 ± 1 minutes. Specimen should be placed to get adequate air circulation.

d. Remove specimen and place in a desicator and cool to room temperature.

e. Remove from the desicator and weigh the specimen to the nearest 0.01 gram (W2).

f. Calculate percent volatiles as follows:

$$\text{Percent volatiles} = \frac{W1 - W2}{W1} \times 100$$

Where: W1 = uncured weight of specimen, g
W2 = final weight of specimen, g

g. Report volatile content of specimen to the nearest 0.1 percent.

Appendix D

Percent Resin Flow Test Procedure

Resin Flow: Resin flow of each sample of the uncured material shall be determined in accordance with the following:

a. Cut a 4-inch by 4-inch ± 1/8-inch squares across the width of the fabric. All squares shall be bias cut to eliminate fiber loss in testing. Stack the squares uniformly on each other to make a specimen. Where the width of the tape is less than 4-inches wide, cut the tape in 4-inch lengths and place side by side to fabricate a 4-inch by 4-inch specimen ply. Stack four plies alternately placed 90 degrees to the previous ply roll direction. Place each ply uniformly on each other to make up a specimen.

b. Weigh the specimen of 4 plies to the nearest 0.01 gram and record as W1.

c. Place the specimen between release film. Preheat the press to 325 ± 10 degrees F, position the specimen in the middle of the press plate and apply the press lead of 150 ± 10 psig immediately. Press load the specimen for a minimum of 10 minutes at 325 ± 10 degrees F.

d. Remove the specimen from the press and cool to ambient temperature.

e. Using a knife, scrape off the resin flash to the original size of the specimen. Do not remove any reinforcement from the original dimensions.

Note: Any fibers which may be displaced and scraped off during the process shall be included in the weighed back specimen weight (W2).

f. Reweigh the specimen to the nearest 0.01 gram and record as W2.

g. Calculate the percent resin flow as follows:

$$\text{Percent flow} = \frac{W1 - W2}{W1} \times 100$$

where: W1 = uncured weight of specimen, g

W2 = final weight of specimen, g

h. Report resin flow to the nearest 0.1 percent.

1985

NASA/ASEE SUMMER FACULTY RESEARCH FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

AUTOMATION OF ELECTROMAGNETIC COMPATIBILITY (EMC) TEST FACILITIES

Prepared By:	Cecil A. Harrison
Academic Rank:	Assistant Professor
University and Department:	University of Southern Mississippi Department of Engineering Technology
NASA/MSFC:	
Division:	Systems and Components Test
Branch:	Experiments and Components Test
MSFC Counterpart:	Jimmy W. Rees
Date:	August 9, 1985
Contract No:	NGT 01-008-021 The University of Alabama in Huntsville

AUTOMATION OF ELECTROMAGNETIC COMPATIBILITY (EMC) TEST FACILITIES

BY

Cecil A. Harrison
Assistant Professor of Engineering Technology
University of Southern Mississippi
Hattiesburg, Mississippi

ABSTRACT

Efforts to automate electromagnetic compatibility (EMC) test facilities at MSFC are discussed. The present facility is used to accomplish a battery of nine standard tests (with limited variations) designed to certify EMC of Shuttle payload equipment. Prior to this project, some EMC tests were partially automated, but others were performed "manually."

Software has been developed to integrate all testing by means of a desk-top computer-controller. Near real-time data reduction and onboard graphics capabilities permit immediate assessment of test results. Provisions for disk storage of test data permit computer production of the test engineer's certification report. Software flexibility permits variation in the test procedure, the ability to examine more closely those frequency bands which indicate compatibility problems, and the capability to incorporate additional test procedures.

ACKNOWLEDGEMENTS

The author expresses appreciation to Jim Rees of the Experiments and Components Test Branch of the George C. Marshall Space Flight Center whose suggestions, insights and preliminary work were of much value to this project. He also thanks Norm Pabon of the Experiments and Components Test Branch for technical assistance during the project and Glynda Thomas of Systems and Components Test Division for clerical assistance in preparing this report.

I. INTRODUCTION

Two principal considerations determine the electromagnetic compatibility (EMC) of a system or system component:

(1) emissions from the system or system component which may degrade performance of another system or of another component of the same system,

(2) susceptibility of the system or system component to emissions from another system or from another component of the same system.

MSFC-SPEC-521A [1], the document which serves to integrate EMC requirements for Shuttle and Spacelab payload equipment, reflects the dual nature of EMC in its broad statement of the EMC requirement:

> "Payload equipment shall not generate levels of interference which would degrade performance or cause malfunction in the Orbiter, Spacelab or other payload systems. Payload equipment shall not malfunction due to susceptibility to system emissions."

The EMC requirements set forth in [1] are derived from the "Spacelab Payload Accommodation Handbook" [2], the "Orbiter/Spacelab Interface Control Document" [3], and Space Shuttle System Accommodation Handbook, Vol. XIV [4]. The methods for EMC test and certification which are specified in [1] are consistent with MIL-STD-461B [5] and MIL-STD-462 [6]. An outline of these test specifications may be found in Table 1.

In FY84, MSFC initiated procurement of a computer/controller and the programmable test instruments required to automate EMC test and certification procedures. The goals were:

(1) to realize the increased accuracy and precision possible with digital instrumentation,

(2) to achieve a facility in which the entire battery of EMC tests could be performed with a minimum of intervention by the test engineer,

(3) to realize a significant reduction in the time required to perform the tests, and

(4) to provide for computer production of the test engineer's certification report.

By mid-1985, the computer/controller, with plotter and printer, and aprroximately 75% of the instruments had been delivered, and interim programs had been written to run five of the nine standard EMC tests specified by [1].

These interim programs provided only limited flexibility, were not integrated, and did not permit mass storage of test data for later use in certification report production. This paper describes the efforts to increase software flexibility, consolidate the entire test battery under a menu-driven main program, and develop a format for mass (flexible disk) storage of test data consistent with certification report requirements.

II. EMC TEST PROCEDURES

A diagram of the EMC test facility at MSFC may be found in Figure 1. This facility permits what is essentially an open-site measurement technique, with the exception that the screen room serves to eliminate environmental factors and to attenuate background electromagnetic fields. Some questions arise concerning reverberations within the screen room and their effect on measurements; however, the purpose of this project is to automate the existing facility and the test methods specified in [1]. These questions concerning the test facility and test methods will be addressed later.

Conducted Emissions

Conducted emissions are those emissions produced by the unit under test (UUT) and conducted on the power lines. Those conducted emissions which occur on the DC power line during steady state operation induce a voltage across a current probe (loop) (Table 2) placed around the power line. The voltage is measured with a spectrum analyzer, and the spectrum analyzer data is adjusted for probe factors to provide the rms current level in terms of dB relative to 1 uA. For RF emissions (20 kHz to 50 MHz), both narrow band analysis and broad band analysis (dB(uA/MHz)) are performed.

Those conducted emissions which occur during turnon, turnoff or other load changes within the UUT are identified as transient emissions. To measure transient emissions, passive R-L-C networks designed to simulate the Orbiter load impedance are placed in series with the power line, and an oscilloscope is used to monitor the voltage developed between the DC power leads, or between any AC lead and the neutral lead.

Radiated Emissions

Electric field emissions (14 kHz to 10 GHz) radiated by the UUT are detected by antennas (Table 2) located 1 m from the UUT, amplified as required, and measured with a spectrum analyzer. The spectrum analyzer data is adjusted for antenna factors to provide the rms field intensity in terms of dB relative to 1 V/m. Both narrow band analysis and broad band analysis (dB (V/m/MHz)) are performed.

Magnetic field emissions are measured in a similar manner, except that a magnetic pickup coil (Table 2) is used and the rms field intensity is referred to 1 picotesla. Only narrow band analysis is performed.

Conducted Susceptibility

Conducted susceptibility refers to the ability of the UUT to operate without degradation when AC or transient signals are impressed on the power lines. AC signals from a signal generator (Table 3), amplified as required, are applied through an isolation transformer with its secondary in the power lead (30 Hz to 50 kHz), or injected through a capacitor (50 kHz to 400 MHz). A digital voltmeter (rf voltmeter above 100 KHz) is used to monitor the applied voltage. Performance of the UUT is observed as the applied voltage is raised to the required level and swept through the required frequency range.

Susceptibility to conducted transients is determined in a similar manner, except that the applied pulse is monitored with an oscilloscope.

Radiated Susceptibility

Radiated susceptibility refers to the ability of the UUT to operate in the presence of an electromagnetic field. AC signals from a signal generator (Table 3) are amplified as required, and radiated by an antenna (Table 2) located 1 m from the UUT. The radiated electric field intensity is monitored with an E-field meter (14 kHz to 220 MHz) or a power meter (above 220 MHz).

III. AUTOMATION OF TEST PROCEDURES

The prinicpal test instruments listed in Table 3 are all fully programmable and are compatible with GPIB. The GPIB is controlled by a technical computer (HP 9826) with integral VDT and flexible disk drive. High-speed programming language (HPL) is used for the controller program. A thermal dot-matrix printer and an X-Y plotter are also attached to the GPIB (Figure 2).

Program Concept

The several factors considered in arriving at a concept for the controller program were:

(1) continuity in transition from "manual" to automated procedures,

(2) functional organization of the procedures with regard to use of test equipment and facilities,

(3) on-line data reduction,

(4) flexibility to allow modification of standard test procedures,

(5) convenient and near-real-time display of test results,

(6) mass storage of reduced data for later use in producing the EMC certification report.

Data Display

Appropriate VDT cues are provided to keep the test engineer aware of the progress of the test procedure. All test options, such as frequency bands, signal levels, etc., are clearly annunicated so that the test engineer can modify standard test procedures.

All data is available for optional tabulation and/or plotting on the VDT immediately after reduction. At that time the tables or plots may be printed on the thermal printer and the plots may be reproduced on the X-Y plotter. The computer-produced plots are also a major advantage to the test engineer who previously drafted these plots for the certification report.

Data Storage

After execution of a major subprogram, the reduced data may be stored in coordinate pairs on flexible disk. The format of the storage file includes test date, UUT designation, number of data points and an optional comment (up to 350 characters) by the test engineer. A separate file is created on the data disk for each execution of a major subprogram.

IV. PROGRESS SUMMARY

Main Program

The main program is fully operable as described in Section III. No major revisions are planned.

Steady State Emission Tests

Steady state conducted and radiated emission tests are fully operable as described in Section III. Revisions related to the installation of programmable relays and delivery of a rotatable antenna platform are discussed below.

Steady State Susceptibility Tests

Steady state conducted susceptibility tests are operable for frequencies below 1 MHz. The program will be revised to extend the frequency to 400 MHz upon receipt of a programmable RF digital voltmeter. Other revisions related to the installation of programmable relays are discussed below.

The steady state radiated susceptibility test is operable for frequencies above 220 MHz. The program will be revised to extend the frequency to 14 kHz upon receipt of antenna correction factors for the IFI EFG-2 antenna. Other revisions related to the installation of programmable relays are discussed below.

Controller Program Organization

The controller program is organized into nine major subprograms (Figure 3), one to implement each of the nine standard EMC test procedures from [1]. These subprograms are executed under the control of a supervisory program, called the main program, which provides a menu for selection of a standard test procedure. Once the main program was developed, the major subprograms were developed and incorporated. This scheme provided continuity in the transition to automated testing and permitted some automated tests to be performed before all programmable test equipment was delivered and before all major subprograms were developed.

In addition to the menu facility, the main program contains a number of global utility routines such as array and string dimensioning, GPIB device designation, VDT operator cues, plotting, and disk storage. The main program and each of the nine major subprograms are stored in separate files on flexibl disk. Loading and execution of the main program causes the menu of standard test procedures to be displayed. When a standard test procedure is selected from the menu, the appropriate major subprogram is loaded from disk and append to the main program in RAM. The combination of main program and a major subprogram is a complete controller program for execution of a standard test procedure. After execution of the test procedure, the menu is displayed again Selection of a different test procedure causes a different major subprogram to be loaded in place of the previously selected major subprogram. This scheme minimizes use of RAM for storage of the controller program and leaves adequate RAM for data collection and processing.

There are some functional shortcomings in the controller program organization because most of the standard test procedures cover wide frequency bands which necessitates changing transducers, amplifiers and even the principal instruments during execution of a major subprogram. Execution of the program must be interrupted while the test engineer completes the equipment changes. VDT cues to the test engineer incorporated into the major subprograms are required to facilitate the equipment changes. A better functional organization of the controller program could be realized if each test procedure was associated with a suite of test equipment. This would permit the use of more, but simpler major subprograms which would execute without interruption for equipment change. Plans to incorporate programmable relays and a programmable antenna turntable will eliminate some of these problems. These plans will, however, require even more complex programming in order to make the automated equipment changes transparent to the test engineer.

Data Reduction

The data reduction required in EMC testing includes sampling, averaging, elimination of narrow band noise (filtering) when appropriate, and correction for transducer factors, cable losses and equipment errors. The sampling, averaging and filtering are straight forward and execute quickly. Transducer factors, cable loss data and equipment errors are stored in look-up tables. This real-time data reduction is a major advantage to the test engineer who formerly performed these computations.

Transient Emission and Susceptibility Tests

No work has been done on the tests for transient emissions or susceptibility pending delivery of a programmable storage oscilloscope and a programmable pulse generator.

Programmable Relays

Construction of the programmable relay arrays for selection of transducers and amplifiers is in progress. When these relays are installed, all presently operable tests will be modified to provide programmed relay actuation. At that time, the VDT operator prompts which are presently used to establish the proper interconnection of test equipment will be eliminated. The transient tests will be written to incorporate programmed relay actuation.

Rotatable Antenna Platform

Procurement of a 4-foot diameter table, which is rotatable under program control, is planned. When the table is delivered, some of the antennas used in radiation tests will be placed on the table and the appropriate antenna will be rotated into position by the test program. At that time, the VDT operator prompts which are presently used to establish proper antenna placement will be eliminated.

Production of Certification Report

At this point, only a rudimentary program is available for recovery of the data stored on flexible disk. A more sophisticated program which will recover the data and produce a certification report is planned. This program will use the stored data to produce tables, graphs, etc., and will generate the appropriate text for a certification report ready for the test engineer's signature.

Controller Program Code

The controller program, which consists of more than 3000 lines of code, has not been reproduced here. Inquiries concerning the controller program code may be directed to Mr. Jimmy W. Rees, MSFC, Code ET45, Huntsville, AL 35812.

V. CONCLUSIONS AND RECOMMENDATIONS

The equipment used in the EMC test facility is of highest quality, and careful planning has gone into the interconnection and programming of the equipment. The result should be a first-class EMC test facility capable of being operated with a minimum of intervention by the test engineer. Completion of the plans for computer-generation of the EMC certification report will give this facility a unique capability.

The controller program has been written for use with MSFC-SPEC-521A [1]. Some modification of frequency bands and of signal levels by the test engineer is possible. Consideration should be given to permit other modifications by the test engineer, such as keyboard entry of different limit specifications than those given in [1], or execution of an ad hoc subprogram for a non-standard test.

The requirements of [1] appear to be a compromise between the desired degree of EMC certification for payload equipment and the degree which was possible with "manual" test procedures and analog instruments. The present automated EMC test facility provides better accuracy and precision than was previously available. Accordingly, [1] should be reviewed, and the EMC requirements revised to take advantage of this facility.

As was mentioned in Section II, there are some questions which should be addressed concerning the measurement technique which is possible with this EMC facility.

(1) The test geometry does not conform to standard test geometries recognized by National Bureau of Standards researches for EMI/ECM testing [7].

(2) The ground plane, which has an extent of less than 1 wavelength in some directions at some frequencies, may be inadequate.

(3) The screen room may produce reverberations which are not accounted for in data processing.

(4) The UUT, positioned on a 1-m wide copper bench, 0.5 m from a vertical metal wall, is essentially in a "corner reflector" environment.

MSFC-SPEC-521A
EMC TEST REQUIREMENTS

EMISSION	SUSCEPTIBILITY
CONDUCTED (CE)	CONDUCTED (CS)
* DC Power Bus Ripple (30Hz-20kHz)	* DC Power Bus Ripple (30Hz-50kHz)
* DC Power Bus RF (20kHz-50MHz)	* DC Power Bus RF (50kHz-400MHz)
* Power Bus Transients	* Power Bus Transients
RADIATED (RE)	RADIATED (RS)
* Electric Field (14kHz-10GHz)	* Electric Field (14kHz-10GHz)
* AC Magnetic Field (20Hz-50kHz)	

Table 1. EMC Test Requirements for Space Shuttle Payload Equipment

CONDUCTED EMISSIONS		
30 Hz - 20 kHz	Current Probe	(Electrometrics PCL-10)
20 kHz - 50 MHz	Current Probe	(Empire Devices CP-105)
Transients	Orbiter Power Impedance Simulation Network	
RADIATED EMISSIONS		
14 kHz - 30 MHz	E-field Rod Antenna	(EMCO 3301)
30 MHz - 200 MHz	Biconical Antenna	(EMCO 3104)
200 MHz - 1 GHz	Conical Log Spiral Antenna	(Singer 93490-1)
1 GHz - 10 GHz	Conical Log Spiral Antenna	(Singer 93491-1)
20 Hz - 50 kHz	Magnetic Field Pick-up Coil	(EMCO 7604)
CONDUCTED SUSCEPTIBILITY		
30 Hz - 50 kHz	Isolation Transformer	
50 kHz - 400 MHz	Capacitor	
Transient	Direct Connection to Pulse Generator	
RADIATED SUSCEPTIBILITY		
14 kHz - 220 MHz	Antenna	(IFI EFG-2)
220 MHz - 1 GHz	Conical Log Spiral Antenna	(Singer 93490-1)
1 GHz - 10 GHz	Conical Log Spiral Antenna	(Singer 93491-2)
13 GHz - 15 GHz	Standard Gain Horn	(Scientific-Atlanta SGH 12.4)

Table 2. Transducers

CONDUCTED EMISSIONS		
30 Hz - 50 MHz	Spectrum Analyzer	(HP8566A)
Transients	Digital Storage Oscilloscope	(Teletronix 5223)*
RADIATED EMISSIONS		
All Tests	Spectrum Analyzer	(HP8566A)
CONDUCTED SUSCEPTIBILITY		
30 Hz - 50 kHz	Synthesizer/Function Generator	(HP3325A)
50 kHz - 400 MHz	Synthesized Signal Generator	(HP8662A)
30 Hz - 100 kHz	Digital Voltmeter (HP3455A)	
200 kHz - 400 MHz	RF Voltmeter **	
Transients	Digital Storage Oscilloscope	(Teletronix 5223)*
Transients	Pulse Generator **	
RADIATED SUSCEPTIBILITY		
14 kHz - 1GHz	Synthesized Signal Generator	(HP8662A)
1 GHz - 2 GHz	Doubler	(HP)
2 GHz - 15 GHz	Synthesized Signal Generator	(HP8672A)
14 kHz - 220 MHz		
220 MHz - 15 GHz	Power Meter (HP436A)	

* Delivery Pending
** Delivery pending; exact model not determined

Table 3. Principal Instruments

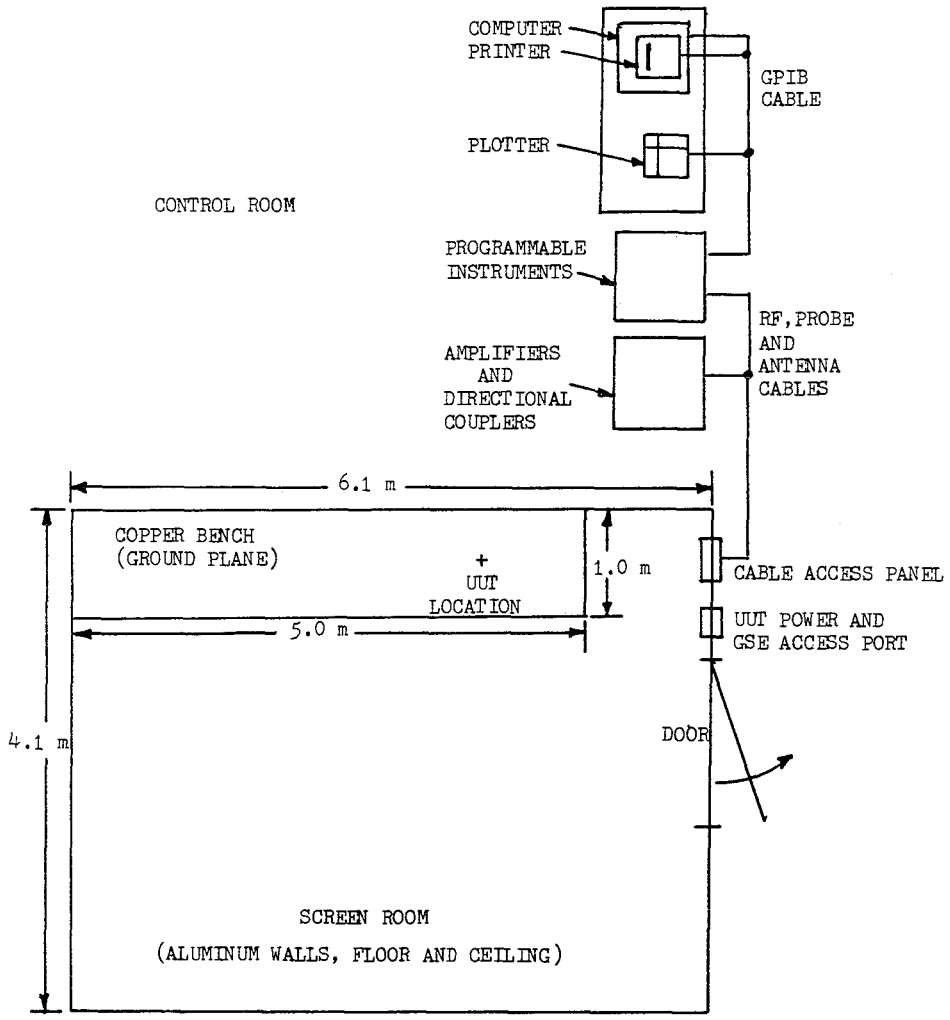


Figure 1. Physical Layout of EMC Test Facility at MSFC

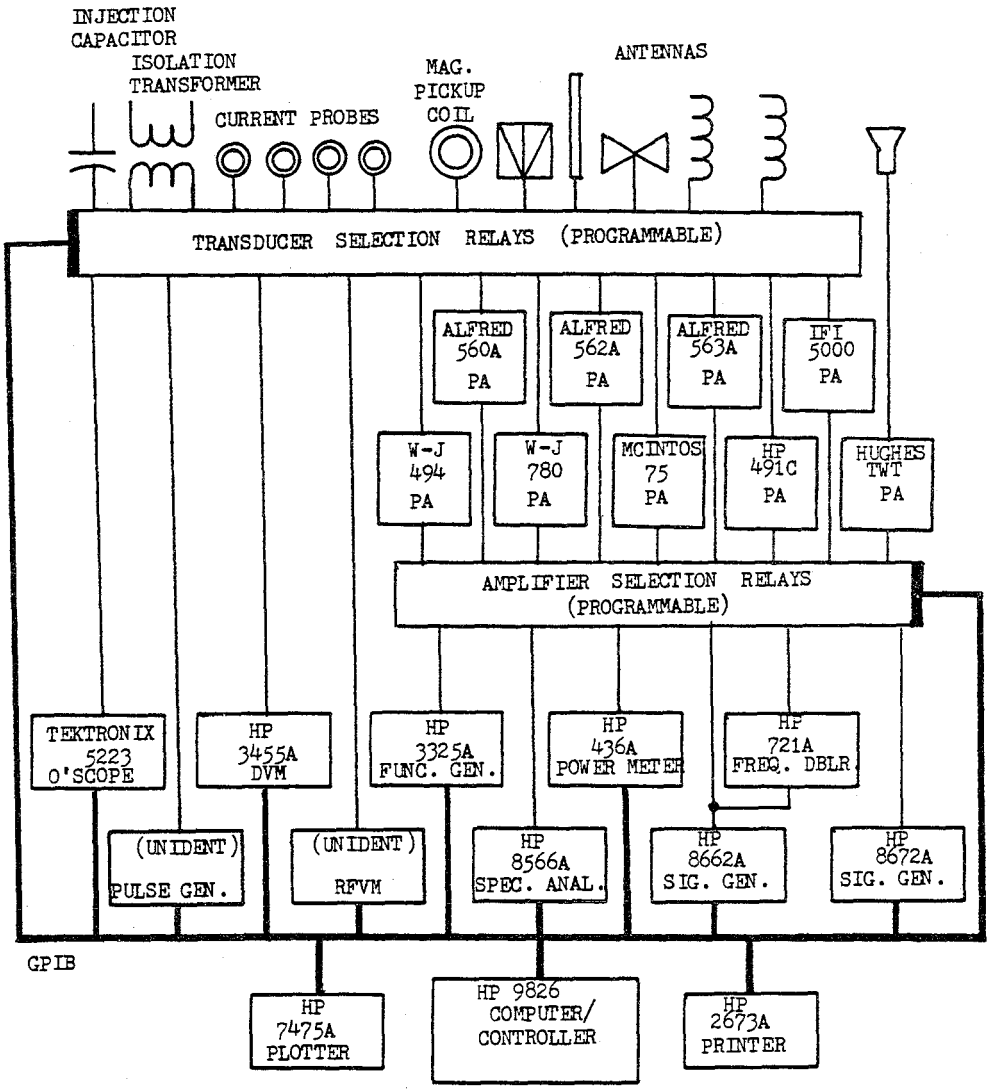


Figure 2. Instrument Suite, EMC Test Facility at MSFC

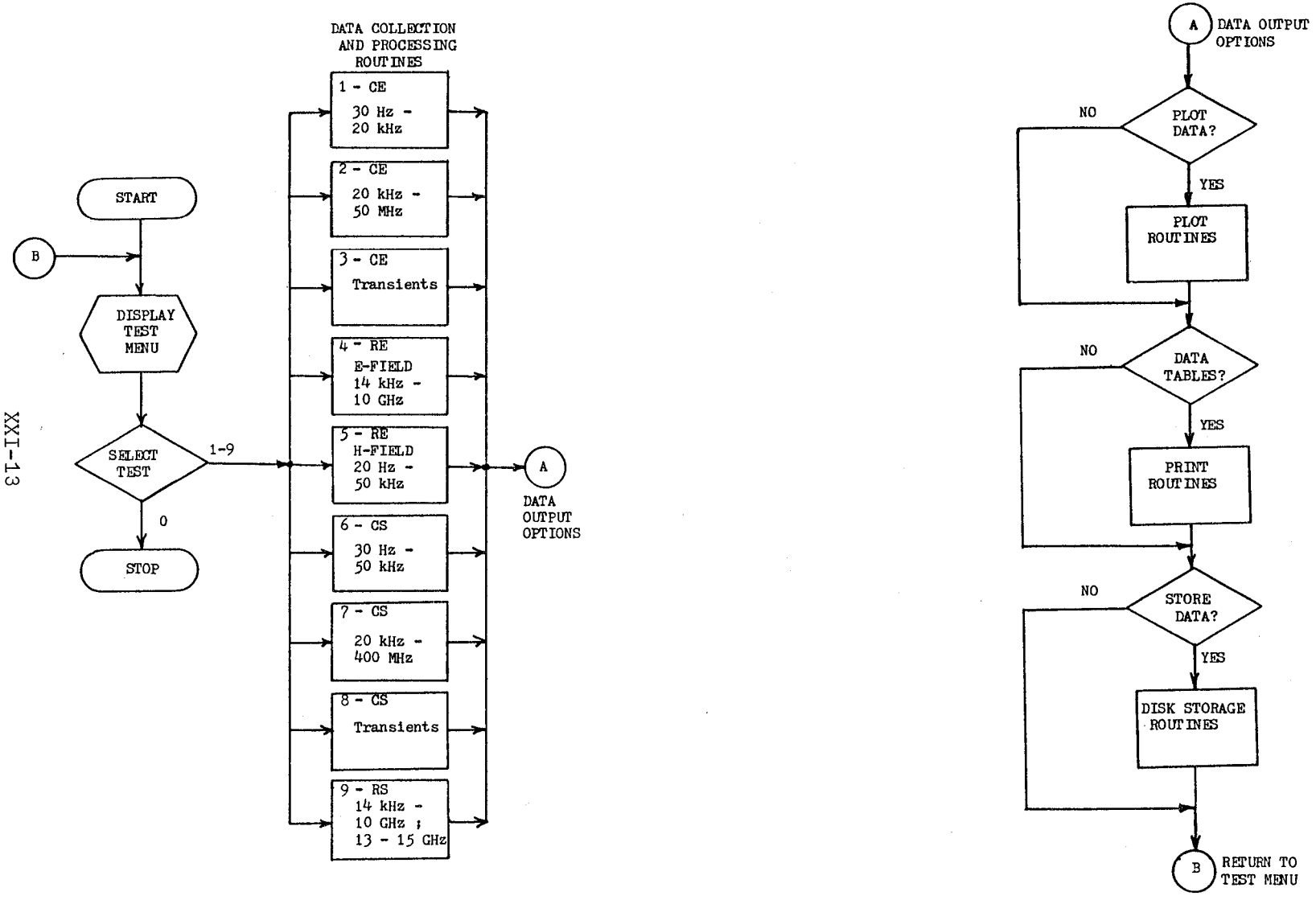


Figure 3. Flow Chart for EMC Test Program

REFERENCES

1. MSFC-SPEC-521A, "Electromagentic Compatibility Requirements on Spacelab Payload Equipment," February 11, 1983.

2. NASA SLP/2104, "Spacelab Payload Accommodation Handbook," January 31, 1981.

3. Space Shuttle System Payload Accommodation Handbook, Vol. XIV, Revision G, NASA, September 26, 1980

4. NACA, ICD-2-05301, "Orbiter/Spacelab Interface Control Document."

5. MIL-STD-461B, "Electromagnetic Emission and Susceptibility Requirements for the Control of Electromagnetic Interference," April 1980.

6. MIL-STD-462, "Measurement of Electromagnetic Interference Characteristics," July 31, 1967.

7. Ma, M.T., et al, "A Review of Electromagnetic Compatibility/Interference Measurement Methodologies," IEEE Proceedings, vol. 73, no. 3, pp. 388-411, March 1985.

1985

NASA/ASEE SUMMER FACULTY RESEARCH FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

MODELING AND ANALYSIS OF CASCADE SOLAR CELLS

Prepared by:	Fat Duen Ho, Ph.D
Academic Rank:	Associate Professor
University and Department:	The University of Alabama in Huntsville Department of Electrical and Computer Engineering
NASA/MSFC:	
Laboratory:	Information and Electronic Systems
Division:	Electrical
Branch:	Electrical/Electronic Parts
MSFC Counterpart:	Teddy M. Edge
Date:	August 9, 1985
Contract No.:	NASA-NGT-01-008-021 The University of Alabama in Huntsville

CASCADE SOLAR CELL MODELING

by

Fat Duen Ho
Associate Professor of Electrical and Computer Engineering
The University of Alabama in Huntsville

ABSTRACT

The report begins with a brief review of the present status of the development of cascade solar cells. It is known that photovoltaic efficiencies can be improved through this development. The designs and calculations of the multijunction cells, however, are quite complicated. The main goal of this study is to find a method which is a compromise between accuracy and simplicity for modeling a cascade solar cell. Three approaches are presently under way, among them (1) equivalent circuit approach, (2) numerical approach, and (3) analytical approach. In this report, we only present the discussion concerning the first and the second approaches. The equivalent circuit approach using SPICE (Simulation Program, Integrated Circuit Emphasis) to the cascade cells and the cascade-cell array highlights this report. The methods of extracting parameters for modeling are discussed.

ACKNOWLEDGEMENTS

The author wishes to express appreciation to Mr. Teddy M. Edge, his counterpart at the Marshall Space Flight Center (MSFC), for his help and encouragement. He would also like to express his gratitude to Mr. Michael D. Martin, Chief of the Electrical/Electronic Parts Branch, for his support and generous hospitality during his stay at MSFC. Thanks are extended to Mr. J. L. Miller, Chief of the Electrical Division, for his encouragement.

The author is grateful to Mrs. Martha S. McCartney, an efficient and good-humored secretary of the Branch, who typed this report with skill and patience. His thanks also go to Mr. P. D. McManus and the other members of the Electrical/Electronic Parts Branch for their help and friendship.

I also wish to extend appreciation to Dr. Gerald Karr and other members of the 1985-NASA/ASEE Coordinating Team for selecting me to participate in this program.

1. INTRODUCTION

Although the solar cell is a rather simple semiconductor device, the design of cells for space applications can be difficult. One of the most important reasons is that the cells used in space must be designed for "end-of-life" operation. The radiation damage can affect and degrade minority carrier diffusion length. In order to experimentally study the performance of a cell, it must be irradiated in the laboratory to simulate the radiation damage in the space environment. This is quite a complicated process. The problems are compounded when a cascade solar cell with its multijunctions is under study.

The analytical method developed for cascade solar cells has been discussed elsewhere[1]. This analysis obtains a solution for minority carrier concentration for the steady-state integral-differential continuity equation in each distinct zone of the cascade cell. Once the minority carrier solutions satisfy the imposed boundary conditions in each distinct region, the voltage-current reltionship for the cascade structure may be obtained, in principle, in closed form. In doing so, we need to solve 14 simultaneous equations to get the V-I relationship even for the simplest cases. Because of this reason, the closed form equation is never explicitly formulated[2].

In order to solve this formidable problem, we should pursue vigorously the study of numerical analysis approach and equivalent circuit approach to model a cascade cell. The equivalent circuit approach incorporated with the SPICE (Simulation Program, Integrated Circuit Emphasis) computer program appears to be a method which is a reasonable compromise between simplicity and accuracy and is thus a more likely candidate for extension to the modeling of cascade solar cell arrays for space application.

2. OBJECTIVES

The objective of this work has been (1) to review the present status of the development of cascade solar cells, (2) to identify the problems of modeling cascade solar cells, (3) to develop theoretical models for two-junction cascade structures or three-junction cascade structures for space application, and (4) to specify the experimental work needed for extracting the parameters for modeling.

3. PRINCIPLES OF OPERATION OF CASCADE SOLAR CELL SYSTEMS

3.1 The Concept of Cascade Cells

The maximum theoretical efficiencies for single-crystal solar cells for Air Mass 1, one sun operation are 20%-21% for silicon cells and 21%-23% for gallium arsenide[3]. Experimental values, however, have remained below 18% at irradiation level of one sun for silicon cells. There are intrinsic limits imposed on the materials due to (a) the finite lifetime of minority carriers and (b) the intrinsic energy bandgap of a given semiconductor material.

There are also losses of potential output. One of them occurs because the photons absorbed in the semiconductor have energy $h\nu$ in excess of the forbidden energy gap E_g; the excess energy $(h\nu-E_g)$ is degraded into heat. The efficiency of solar energy conversion by the photovoltaic effect could be increased significantly by constructing a system of cascade (multijunction, multibandgap) solar cells.

The cascade structure consists of multiple layers of different bandgap materials. The cascade cell concept aims at reducing the losses associated with photons whose energy is not completely utilized, i.e., those associated with the $h\nu - E_g$ losses.

Substantial efficiency increases are expected for tandem cell structures in comparison with single-junction cells[4]. To illustrate the principles underlying the cascade cell concept, we assume that all photovoltaic p/n homojunction cells used in a tandem structure show the same value of the maximum efficiency, η_{max}. For example, we assume that the tandem system consists of three p/n homojunction cells made from semiconductors with energy gaps $E_{g2} > E_{g1} > E_{g0}$. A rough estimation of the efficiency of tandem cells can be done as follows. Assume that solar illumination of 100 mW is incident on the cell made from E_{g2} in the system. If each of the three cells has an efficiency of 20%, the cell with energy gap E_{g2} will produce 20 mW; the cell with E_{g1} will produce 10 mW due to its short circuit current whose value is one-half what it would have been if the top cell were not present. The cell with E_{g0} will produce a 6.67 mW because its current is one-third of its potential value. The total output of the stack would therefore be 36.7 mW and hence the efficiency of the cascading system of three cells would be 36.7%.

3.2 Fundamental Requirements of Materials for a Cascade Solar Cell

The two fundamental requirements of materials for a monolithic cascade solar cell are (a) a close lattice match between the materials and (b) appropriate bandgaps for the cells.

A number of material combinations in the III-V system meet these criteria[5]. Choices for the narrow-bandgap cell include: the ternaries GaInAs, GaAsSb, InPAs, AlGaSb, and AlInSb. Choices for the wide-bandgap cell include: GaPAs, AlGaAs, GaInP, and AlAsSb.

The $Al_{1-x}Ga_xAs$-GaAs material combination ("x" represents alloy ratio), for all practical purposes, is lattice matched over the complete ternary alloy range. In general, GaAs served as a reasonably well-characterized base material for comparison with other III-V compounds.

In addition to the proper bandgaps and lattice-matched materials, a cascade cell must have other desirable electrical properties. The minority-carrier lifetime must be long enough, which means a diffusion length of several micrometers, for direct band-gap III-V semiconductors. The technology of obtaining a degenerately doped tunneling junction is not trivial. One of the main problems is the doping requirements.

3.3 The Major Components of a Cascade Solar Cell

The major components of a complete cascade solar cell can be summarized as:

a. Substrate
b. Low-bandgap cell
c. Connecting (or tunnel) junction
d. Wide-bandgap cell
e. Window layer
f. Antireflecting layer and ohmic contacts

The connecting (or tunnel) diode is one of the most critical components in the cascade-cell structure[6,7]. It should have a low impedance to current flow in both directions, and a low voltage drop across the diode. It should also present a very low absorption for photons which pass through the top cell. The tunnel junction must have a bandgap as large as that of the top cell.

4. CALCULATIONS AND DESIGN OF A CASCADE SOLAR CELL

4.1 Calculations of the Efficiency of a Cascade Solar Cell

The efficiency of a cascade cell can be more exactly calculated by using the following assumptions:

(a) The collection efficiency is 100%, i.e., every photon with energy $h\nu > E_g$ contributes one electron charge to the short circuit current, I_{SC}.

(b) If two p/n homojunction cells having E_{g0} and E_{g1}, respectively, form a cascade system, the former is the bottom cell while the latter is the top cell. We also assume that $E_{g1} > E_{g0}$. Then the short circuit current of the form is reduced to

$$I_{SC}\ (E_{g0},\ E_{g1}) = I_{SC}\ (E_{g0}) - I_{SC}\ (E_{g1})$$

(c) The reverse saturation current I_0 of the junction can be expressed as

$$I_0\ (E_G) = K_1 \exp\ (-E_G/BKT)$$

Where B can be assumed to be equal to 1 and we determine the value of K_1 by choosing the value $I_0 = 2 \times 10^{-12}$ A/cm^2 for silicon at room temperature.

(d) The voltage at maximum power V_{mp} is given by

$$(1 + \Lambda V_{mp}) \exp\ (\Lambda V_{mp}) = I_{SC}\ (E_{g0},\ E_{g1})/I_0\ (E_{g0})$$

(e) The current at maximum power I_{mp} is given by

$$I_{mp}\ (E_{g0}, E_{g1}) = \frac{\Lambda V_{mp}}{1 + \Lambda V_{mp}}\ I_{SC}\ (E_{g0},\ E_{g1})$$

(f) The parameter $\Lambda \equiv q/AKT$. It is assumed that A = 1.

(g) The maximum efficiency is calculated by

$$\eta_{max} = \frac{I_{mp}\ V_{mp}}{P_{in}}$$

Where P_{in} is the input power.

(h) Note that $I_{SC}(E_g)$ can be approximated by

$$I_{SC}(E_g) = I_{SC}(0) \exp(-\alpha E_g)$$

Where $I_{SC}(0)$ is the intercept of the extrapolation of I_{SC} from the semilog region to the ordinate $E_g = 0$ and α is an empirical constant whose value is determined by

$$I_{SC}(E_g) = q \int_{h\nu = E_g}^{\infty} n_{ph}(h\nu)\, d(h\nu)$$

Where $n_{ph}(h\nu)$ is the number of photons of energy $h\nu$ in the solar spectrum under consideration in the interval $d(h\nu)$.

(i) Also note that for the AMO spectrum, we can approximately have

$$I_{SC}(E_g) = \{202 \exp(-1.2\, E_g)\} \text{ mA}$$

The values of α range from 1.2 to 2.1 for solar spectra under different atmospheric conditions and solar azimuth.

4.2 Another Method of Calculations

Another approach of calculations of short circuit densities, open-circuit voltages, fill factors, and combined efficiencies has been reported in reference 8.

The short-circuit current densities J_{SC1} and J_{SC2} for cell 1 and cell 2, respectively, are calculated as

$$J_{SC1} = \int_{\lambda_{g1(\mu m)}}^{\lambda_{g1(\mu m)}} qF(\lambda)d\lambda$$

and

$$J_{SC2} = \int_{\lambda_{g1}(\mu m)}^{\lambda_{g2}(\mu m)} qF(\lambda)d\lambda$$

Where $F(\lambda)$ represents the solar photon flux density at λ, which varies with air mass.

The open-circuit voltage V_{OC} for each of the cells can be determined by

$$V_{OC} = \frac{kT}{q} \ell n \left(\frac{XJ_{SC}}{J_{oo}} + 1 \right)$$

Where X is the concentration ratio, and

$$J_{oo} = K'T^3 \exp \left(- \frac{E_g}{kT} \right)$$

if a simple diffusion current is assumed. K' may be different for different material.

The fill factor ff for each of the cells is expressed as

$$ff = \frac{V_m}{V_{OC}} \left[1 - \frac{\exp (qV_m/kT)-1}{\exp (qV_{OC}/kT)-1} \right]$$

Where V_m is given by

$$\exp \left(\frac{qV_m}{kT} \right) \left(1 + \frac{qV_m}{kT} \right) = \frac{XJ_{SC}}{J_{oo}} + 1$$

The combined efficiency for the two-terminal case is calculated by

$$\eta_{tot} = J_{SC} \frac{V_{OC1} (ff)_1 + V_{OC2} (ff)_2}{P_{in}}$$

Where P_{in} is the incident solar power density.

For four-terminal case

$$\eta_{tot} = \frac{J_{SC1} V_{OC1} (ff)_1 + J_{SC2} V_{OC2} (ff)_2}{P_{in}}$$

The discussion concerning the two-terminal structure and the four-terminal structure will be presented in Section 5.

4.3 Tandem Cell Design

The requirements of designing a cascade cell are summarized as follows:

a. Each cell must transmit efficiently the photons with less than its bandgap energy.

b. The contacts on the backs of the upper cells must be transparent to these photons.

c. A good first-order approximation for optimum designs requires the short-circuit currents of the two cells (or three cells) to be equal. This design constraint can be used to obtain a required relationship between the bandgaps of the two cells (or three cells).

5. STUDY OF TWO-JUNCTION CASCADE SOLAR CELL STRUCTURE

The simplest tandem structure consists of two cells, which can be connected to form either two-terminal, three-terminal or four-terminal devices. In a two-terminal or four-terminal device, the cells are connected

in series. In this case, the photocurrent for both junctions of the cascade cells must be equal. On the other hand, in the three- and four-terminal devices, the photocurrents of the two cells do not have to be equal.

Let us now compare the characteristics of two- and four-terminal tandem cells. There is an essential difference between these two structures. A two-terminal device only needs one external circuit load. But the requirements for an equal photocurrent for the whole structure impose a limitation to the choice of energy gaps of the two cells. Furthermore, the calculated maximum photocurrent densities do not only depend on energy gaps, but also depend on air masses. For different air masses, we have different spectral distribution. Therefore, the optimal design for a two-terminal cell in terms of energy gaps for all air masses is much more difficult.

A four-terminal device uses two separate external circuit loads. Because photocurrent matching is not required for this structure, the selection of optimal energy gap combinations for each air mass is much wider. Since the energy gap ranges are greater, the control of energy gaps and film thickness can be much less stringent than that required for two-terminal structures. In four-terminal structures, the variation in overall conversion efficiency is more insensitive to air mass for greater air masses. This differs from that of two-terminal cells, which are more sensitive to it.

A two-terminal tandem structure can be formed by using a transparent conducting bonding material, while the insulating bonding material can be applied to obtain a four-terminal structure.

Further comparison between the two-terminal structure and the four-terminal structure is shown in Table 1.

Table 1. Further Comparison Between the Two-Terminal Structures and the Four-Terminal Structures

	Two-Terminals	Four-Terminals
Bonding	Transparent Conducting Bonding	Insulating Bonding Material
Lattice Matching	Critical	Not Critical
Antireflection Coatings	Simpler (Only the front surface of the top cell requires an antireflection coating.)	More Complicated (AR coatings are needed on the front and back surfaces of the top cell and on the surface of the bottom cell.)
Flexibility	Lower	Higher

Based on the results and progress achieved in multibandgap research, it has been determined that the stacked four-terminal approaches have advantages over two-terminal devices[8].

6. CASCADE SOLAR CELL MODELING

To date no single model can accurately represent the cascade cells over all ranges of temperature, illumination intensity, and radiation damage. In solar cell modeling, we must meet the following criteria:

a. It must be able to provide sufficiently accurate simulation of I-V curves over the range of interest of temperature, illumination level and radiation damage.

b. It must, with sufficient accuracy, be able to manipulate the I-V curves, as required for predicting the cell performance under certain specified operating conditions.

c. It must be able to extend to the design of the solar array without introducing too much complexity.

Three approaches are presently under way, among them (1) equivalent circuit approach, (2) numerical approach, and (3) analytical approach. In this report, we only present the discussion concerning the first and the second approaches.

6.1 Equivalent Circuit Approach

The equivalent circuit representing a two-junction cascade solar cell is shown in Figure 1. The solar cell diode equations are applied. Terms for the light generated currents, diffusion currents, space charge recombination currents, series and shunt resistance, the resistance for the window layer and the substrate, and the equivalent resistance for the tunnel diode are included. The current sources in the equivalent circuit are as follows:

I_{ph1} = light-generated current for the top cell

$$I_{D1} = I_{01} \left[\exp\left(\frac{qV_{01}}{kT}\right) - 1\right]$$

$$I_{D2} = I_{02} \left[\exp\left(\frac{qV_{01}}{2kT}\right) - 1\right]$$

I_{ph2} = light-generated current for the bottom cell

$$I_{D3} = I_{03} \left[\exp\left(\frac{qV_{02}}{kT}\right) - 1\right]$$

$$I_{D4} = I_{04} \left[\exp\left(\frac{qV_{02}}{2kT}\right) - 1\right]$$

Where I_{01} = coefficient of diffusion current term for the top cell

I_{02} = coefficient of space charge region recombination current term for the top cell

I_{03} = coefficient of diffusion current term for the bottom cell

I_{04} = coefficient of space charge region recombination current term for the botton cell

V_{01} = the diode voltage inside the top cell

V_{02} = the diode voltage inside the bottom cell

The resistance in the equivalent circuit is defined as follows:

R_{S1} = the series resistance of the top cell

R_{Sh1} = the shunt resistance of the top cell

R_{S2} = the series resistance of the bottom cell

R_{Sh2} = the shunt resistance of the bottom cell

R_{w} = the resistance of the window layer

R_{sub} = the shunt resistance of the bottom cell

R_{tun} = the equivalent resistance of the tunnel diode

The computer program SPICE (Simulation Program, Integrated Circuit Emphasis) is used to simulate cascade solar cells.

For doing the simulation, we need to determine the values of parameters more accurately. The parameter extraction needs to be done through the experimental work if those values are not available. After we obtain more accurate parameters, we are going to use SPICE to do the following simulations:

a. The cascade solar cell V-I curve with temperature as a parameter.

b. Cascade efficiency and top- and bottom-cell efficiencies versus temperature.

c. Voltage at the maximum power point of the cascade, top and bottom cell V-I curves versus temperature.

d. The effects of radiation damage.

e. The effects of the various values of R_{Sh} and R_S.

For studying the radiation effects, we will alter carrier diffusion lengths.

The example we have been using to carry out our modeling study is an AlGaAs/GaAs cascade solar cell. The cross section of this cell has been presented elsewhere[9], and some useful information can also be obtained in other publications[10].

Note that the equivalent circuit of a cascade solar cell structure is not self-evident. It comes from the solution to the continuity equation. Therefore, the equivalent circuit approach should be studied in conjunction with the numerical or analytical approach or both.

6.2 Numerical Approach

We plan to develop a computer program for modeling cascade solar cells. The equations associated with p/n junctions are given in a normalized form by:

$$J_n = -D_n (n\psi' - n')$$

$$J_p = -D_p (p\psi' + p')$$

$$J_{n'} = U\text{-}G$$

$$J_{p'} = (U\text{-}G)$$

$$\psi'' = n\text{-}p\text{-}N$$

The above equations are normalized by the following:

Position: $L_D = (\varepsilon_s \, kT/q^2 \, n_i)^{\frac{1}{2}}$

Potential:

Carrier/Impurity Concentration: n_i

Mobility: $M_0 = 1 \text{ cm}^2/\text{V-sec}$

Diffusion Constant: $D_0 = 1 \text{ cm}^2/\text{sec}$

Current Density: $q \, n_i \, D_0/L_D$

Time: L_D^2/D_0

Our program is planned to solve the three-coupled, nonlinear partial differential equations, Poisson's equation, and the hole and electron continuity equations in both one-dimension and two-dimensions in a cascade cell subject to appropriate boundary conditions. Finite difference methods will be used to discretize the equations. The basic solutions of the equations will consist of the electric field - or equivalently, the electrical potential -and the hole and electron concentrations.

7. PHOTOVOLTAIC MEASUREMENTS

The performance of solar cells is critically related with the properties of the semiconductors, as well as the nature of the photovoltaic barrier interface. Therefore, it is important to measure the basic material and interface parameters. Some of the important photovoltaic measurements[11] are very briefly discussed as follows.

7.1 Current-Voltage Characteristics

The essential performance parameters of a solar cell, i.e., the open-circuit voltage V_{OC}, the short-circuit current I_{SC}, the fill factor ff, and the cell efficiency can be determined by using the I-V characteristics under illumination. Besides these external performance parameters of the solar cell, the internal parameters such as series resistance R_S, shunt resistance R_{Sh}, and reverse saturation current I_0 are also important. I_0 and n (the ideality or quality factor) can be evaluated by the dark I-V characteristics, while R_S and R_{Sh} may be obtained from the illuminated I-V curves.

The dark I-V characteristics can be very useful in identifying the recombination mechanisms such as bulk diffusion - recombination, space charge region generation-recombination, thermionic emission, thermionic-assisted tunneling, and direct or trap-assisted tunneling.

7.2 Spectral Response and Quantum Efficiency

The spectral response indicates the relative contribution of photons of different energy to the short-circuit photocurrent of the solar cell. The external quantum efficiency η_{ext} is defined as

$$\eta_{ext}(\lambda) = \frac{I_{SC}(\lambda)}{q\phi(\lambda)}$$

Where $\phi(\lambda)$ represents the photon flux per second at λ incident on the cell. A plot of η_{ext} versus λ gives the absolute spectral response, which is determined by the optical absorption coefficient $\alpha(\lambda)$, the minority carrier diffusion length ℓ, the surface recombination velocity S, the type of, photovoltaic barrier, and the presence of antireflection coatings. So, the measurement on spectral response can be helpful to evaluate the cell's performance.

7.3 Diffusion Length

The minority carrier diffusion length is one of the most important material parameter of a solar cell. It determines the photocurrent, the limiting value of the dark current and, therefore, the open-circuit voltage. The three most common methods for measuring this parameter are:

a. Spectral response measurements

b. Surface photo voltage measurements

c. Beam-induced current measurements.

The more detailed discussion can be found in Reference 11.

7.4 Minority Carrier Lifetime

There are many methods for measuring the minority carrier lifetime in the base of p/n junction, both in dark and in illuminated conditions[11]. Some of the lifetime measurement techniques are: (a) the open-circuit voltage decay, (b) diode reverse recovery, (c) small signal impedance measurement, (d) light induced photovoltaic decay, (e) open circuit to short circuit switching, and (f) the decay of light induced short circuit current.

There are many articles concerning mathematical formulations for these methods. The transient behavior of the p/n junction devices during experiments of minority lifetime measurements has been explained successfully with the help of some models [12,13]. The author of this report has been studying an equivalent circuit approach incorporating the SPICE computer program to study this transient behavior. It is expected that some positive results will be obtained from this investigation.

8. CONCLUSION AND RECOMMENDATIONS

The cascade solar cell is one of the most promising devices for increasing the conversion efficiency. The main problems of the two-terminal cascade structures are the growth of high-quality semiconductor layers under lattice-mismatching conditions and the development of low-impedance connection between the top cell and the bottom cell. Because photocurrent matching is not necessary for the cells, four-terminal devices have some important advantages over two-terminal devices. Progress has been made in improving and in studying four-terminal structure, but a lot of room for improvement remains. More study should be done on these four-terminal devices.

This study also suggests the importance of modeling the cascade solar cells. It is concluded that we should pursue vigorously the study of numerical analysis approach and equivalent circuit approach to model a cascade cell.

The latter approach incorporated with the SPICE computer program appears to be a method which is a compromise between simplicity and accuracy and is thus a more likely candidate for extension to the modeling of cascade solar cell arrays for space applications. Therefore, it deserves much more attention.

We also recommend that much more vigorous efforts should be done in the experimental work concerning extraction of the parameters of the cells.

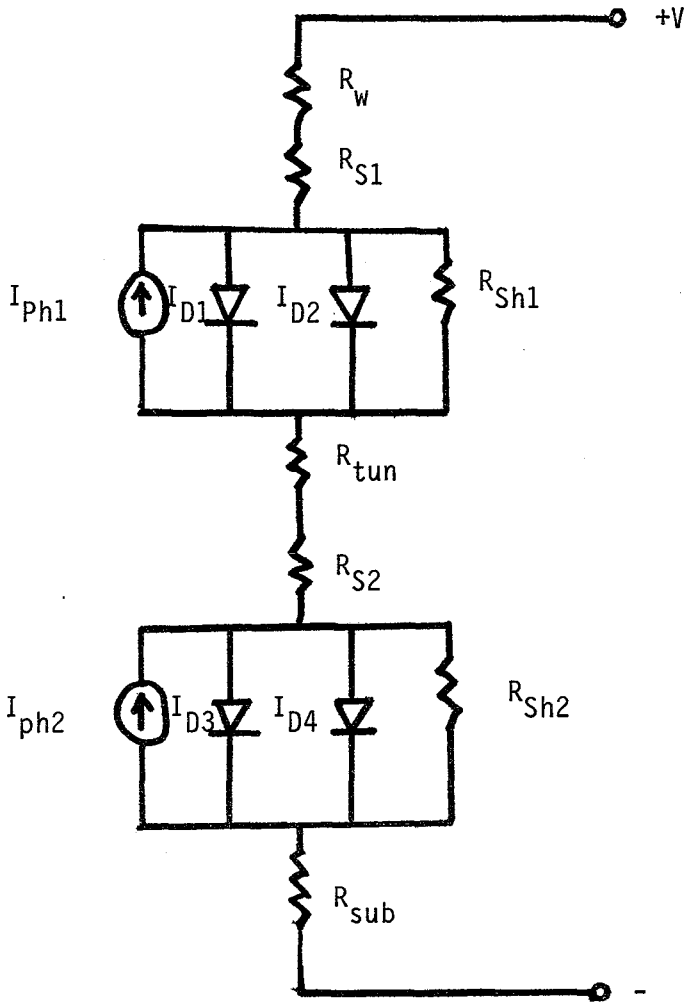


Figure 1. Equivalent Circuit Model of a Cascade Solar Cell for SPICE Simulation

REFERENCES

1. M. F. Lamorte and D. H. Abbott, IEEE Trans. Electron Devices, Vol. ED-27, 231-249, 1980.

2. M. F. Lamorte and D. H. Abbott, Solar Cells, Vol. 9, 311-326, 1983.

3. S. M. Bedair, Solar Cells, Vol. 12, 11-16, 1984.

4. J. J. Loferski, Twelth IEEE Photovoltaic Specialists Conf. Proc., Baton Rouge, LA, 957-961, 1976.

5. S. M. Bedair, S. B. Phatak, and J. R. Hauser, IEEE Trans. Electron Devices, Vol. ED-27, 822-831, 1980.

6. S. M. Bedair, J. Appl. Phys. (50(11), 7267-7268, 1979.

7. J. P. Contour and A. M'Baye, Solar Cells, Vol. 13, 67-76, 1984.

8. J.C.C. Fan and B. J. Palm, Solar Cells, Vol. 12, 401-420, 1984.

9. M. F. Lamorte and D. H. Abbott, Solar Cells, Vol. 9, 311-326, 1983.

10. R. C. Knechtli, R. Y. Loo, and G. S. Kamath, Electron Devices, Vol. ED-31, 577-588, 1984.

11. S. Ashok and K. P. Pande, Solar Cells, Vol. 14, 61-81, 1985.

12. S. R. Dhariwal and R. Gadre, Solid State Electronics, Vol. 27, 837-847, 1984.

13. S. R. Dhariwal and N. K. Vasu, Solid State Electronics, Vol. 24, 915-927, 1981.

1985

NASA/ASEE SUMMER FACULTY RESEARCH FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

A METHOD FOR LINEARIZING A NONLINEAR SYSTEM WITH SIX STATE VARIABLES AND THREE CONTROL VARIABLES

Prepared by:	Wei-Shen Hsia, Ph.D.
Academic Rank:	Professor
University and Department:	University of Alabama - Tuscaloosa Department of Mathematics
NASA/MSFC:	
Laboratory:	Systems Dynamics Laboratory
Division:	Control Systems Division
Branch:	Pointing Control Systems Branch
MSFC Counterpart:	Henry B. Waites John P. Sharkey

Date:	August 2, 1985
Contract No.:	NGT-01-008-021 The University of Alabama in Huntsville

ACKNOWLEDGEMENT

My thanks and apprecations to Dr. Gerald Karr, Dr. James B. Dozier, and Mr. Leroy Osborn for their efforts in conducting the NASA/ASEE Summer Faculty Fellowship Program. I appreciate the exciting and challenging problem shared by Dr. Henry B. Waites. Most importantly, I appreciate the efforts made by Mr. John B. Sharkey, who improved the algorithm and verified it through computer simulation. Without his efforts the report could not be in the present form.

A METHOD FOR LINEARIZING A NONLINEAR SYSTEM WITH SIX STATE VARIABLES AND THREE CONTROL VARIABLES

by

Wei-Shen Hsia
Professor of Mathematics
The University of Alabama
University, Alabama 35486

John P. Sharkey
Henry B. Waites, Ph.D.
Marshall Space Flight Center
Control Systems Division

ABSTRACT

A nonlinear system governed by $\dot{x} = f(x,u)$ with six state variables and three control variables is considered in this project. A set of transformation from (x,u) - space to (z,v) - space is defined such that the linear tangent model is independent of the operating point in the z-space. Therefore, it is possible to design a control law satisfying all operating points in the transformed space. An algorithm to construct the above transformations and to obtain the associated linearized system is described in this report.

This method is applied to a rigid body using pole placement for the control law. Results are verified by numerical simulation. Closed loop poles in x-space using traditional local linearization are compared with those pole placements in the z-space.

LIST OF FIGURES

INTRODUCTION

In general, there are two commonly used approaches to linearize a nonlinear control system: one is to linearize locally about some operating points, the other is to linearize globally. However, difficulties will arise when these approaches are applied. For the former one, it is hard to design a unified control law over the entire set of operating points; on the other hand, for the later approach, usually, the control does not permit an acceptable closed-loop system behavior over a wide operating range. But, if a nonlinear system has a tangent model independent of the operating point, the above stated difficulties disappear. Using this concept, Reboulet and Champetier [1] developed a procedure which transforms a nonlinear system with one control variable to another space such that the desired independent property is satisfied. In this project, we generalize their procedure to the case where a nonlinear system consists of six state variables and three control variables. This method is then applied to a rigid body dynamic model using pole placement for the control law.

OBJECTIVES

The objectives of this project are:

1. Develop a pseudo-linearization technique to analyze a nonlinear control system with six state variables and three control variables.

2. Apply this technique to a rigid body dynamic model.

3. Verify the feasibility of this technique by computer simulation.

BODY OF REPORT

Concept

Consider the class of 6-state and 3-input nonlinear systems governed by

$$\dot{x} = f(x,u) \tag{1}$$

where $x \in R^6$, $u \in R^3$, and $f = (f_1, \ldots, f_6)^t : R^9 \to R^6$ of class C^1. The set of operating points is defined by

$$M_{x,u} \underline{\Delta} \{(x_o,u_o) : f(x_o,u_o) = 0\}$$

and its projection in the state space is defined by

$$M_x \underline{\Delta} \{x_o : \exists\, u_o \text{ s.t. } f(x_o,u_o) = 0\}.$$

In the neighborhood of an operating point (x_o,u_o), the dynamic behaviour of the system may be considered as linear. It is then described by

$$\delta\dot{x} = F(x_o,u_o)\delta x + G(x_o,u_o)\delta u \tag{2}$$

where $\delta x \in R^6$, $\delta u \in R^3$ and $F \underline{\Delta} \partial f/\partial x$, $G \underline{\Delta} \partial f/\partial u$.

Our aim is to find mappings

$$\begin{aligned} z_i &= T_i(x) && , i = 1, \ldots, 6 \\ v_j &= T_{6+j}(x,u) && , j = 1, 2, 3 \end{aligned} \tag{3}$$

($T_1, \ldots, T_6$ being functionally independent and $\partial T_{6+j}/\partial u \neq 0$, $j = 1, 2, 3$) such that, in the z-state space, the linear tangent model is independent of the operating point and can be written under the controllability canonical form

$$\begin{aligned} \delta\dot{z}_1 &= \delta z_4 \\ \delta\dot{z}_2 &= \delta z_5 \\ \delta\dot{z}_3 &= \delta z_6 \\ \delta\dot{z}_4 &= \delta v_1 \\ \delta\dot{z}_5 &= \delta v_2 \\ \delta\dot{z}_6 &= \delta v_3 \end{aligned} \tag{4}$$

Note that the controllability of the linear tangent model is invariant under the mappings (3). Hence, to obtain (4) the following condition must hold:

For any $(x_o,u_o) \in M_{x,u}$ the pair $(F(x_o,u_o), G(x_o,u_o))$ is controllable. (5)

In practice, the solution provided by our method is effective along $M_{x,u}$, except at the points (x_o,u_o) for which (5) is not satisfied.

Now, the variations in the z-state space δz will be rewritten in terms of δx and δu. From eqn. (1) and (3) we have

$$\dot{z}_i = (\partial T_i/\partial x)f(x,u) \qquad , i = 1, \ldots, 6.$$

Hence, at any points (x_o,u_o) of $M_{x,u}$

$$\delta\dot{z}_i = \alpha_i F(x_o,u_o)\delta x + \alpha_i G(x_o,u_o)\delta u, \qquad i = 1, \ldots, 6$$

where α_i are the 1-forms over M_x

$$\alpha_i \triangleq d\, T_i\Big|_{M_x} \qquad , i = 1, \ldots, 6. \tag{6}$$

For convenience, reference to the operating point will be omitted. To obtain eqn. (4) it is easy to show the following equations must be satisfied on $M_{x,u}$

$$\alpha_i G = 0 \qquad , i = 1, 2, 3 \tag{7a}$$

$$\alpha_i F = \alpha_{3+i} \qquad , i = 1, 2, 3 \tag{7b}$$

$$\alpha_{3+i} = \alpha_i(F\ G) \qquad , i = 4, 5, 6 \tag{7c}$$

with $\alpha_j \triangleq d\, T_j\Big|_{M_{x,u}} \qquad , j = 7, 8, 9.$

With this result, the problem then becomes that of finding 1-forms $\alpha_1, \ldots, \alpha_6$ (resp. $\alpha_7, \alpha_8, \alpha_9$) satisfying eqns. (7) at any point of M_x (resp. $M_{x,u}$) and such that there exist mappings $T_i(x)$ $(i = 1, \ldots, 6)$ and $T_j(x,u)$, $(j = 7, 8, 9)$ such that

$$\alpha_i = d\,T_i\Big|_{M_x} \qquad , i = 1, \ldots, 6$$

$$\alpha_j = d\,T_j\Big|_{M_{x,u}} \qquad , j = 7, 8, 9.$$

Analysis

Recall that $M_{x,u}$ is a set of points described by six nonlinear equations with nine unknowns.

Let $\dot{M}_{x,u}$ be the set of operating points in the neighborhood of which the linear tangent model of system (1) is controllable. Since the uncontrollability of the tangent model (2) is equivalent to the nullity of all 5-dimension minors (Kalman rank condition). Therefore, the complement of $\dot{M}_{x,u}$ is a closed subset in $M_{x,u}$ which implies $\dot{M}_{x,u}$ is an open subset in $M_{x,u}$.

Furthermore, another controllability condition is given by

$$\text{rank}(F-\lambda I \quad G) = 6 \qquad ,\forall\, \lambda\ .$$

In particular, for $\lambda = 0$

$$\text{rank}(F \quad G) = 6.$$

This means that the mappings f_i, $i = 1, \ldots, 6$ are functionally independent at the point of interest. Hence, $M_{x,u}$ is a 3-dimensional submanifold of R^9. Which also implies that $\dot{M}_{x,u}$ is a 3-dimensional submanifold of R^9 since $\dim M_{x,u} = \dim \dot{M}_{x,u}$ if $\dot{M}_{x,u} \neq \phi$.

Let $\dot{M}_x$ be the projection of $\dot{M}_{x,u}$ on the x-state space. In order to develop global results we will suppose that the following condition holds:

There exists a C^1-diffeomorphism $\phi = (\phi_1, \ldots, \phi_6)^t$ from R^6 to R^6 such that $\dot{M}_x$ is given by

$$\phi_4(x) = \phi_5(x) = \phi_6(x) = 0 \tag{8}$$

Under this assumption, if we define

$$z_i = \phi_i(x) \qquad , i = 1, \ldots, 6$$

then the surface M_x is transformed in the z-space into a 3-dimensional manifold D_z : $z_4 = z_5 = z_6 = 0$. Let $\bar{\alpha}$ be the image of a 1-form $\alpha = d\,T\Big|_{M_x}$ under ϕ, then

$$\bar{\alpha}(z) \triangleq \alpha(\phi^{-1}(z))(\partial x/\partial z)_{D_z}, \quad z \in D_z. \tag{9}$$

Note that if we post-multiply $\alpha = d\,T\Big|_{M_x}$ by $(\partial x/\partial x)\Big|_{D_z}$,

we obtain

$$\alpha(\phi^{-1}(z))(\partial x/\partial z)\Big|_{D_z} = d\,T(\Phi^{-1}(z))\Big|_{D_z}(\partial x/\partial z)_{D_z}, \ \forall\ z \in D_z, \tag{10}$$

i.e.,

$$\bar{\alpha} = d\,\bar{T}\Big|_{D_z} \tag{11}$$

with

$$\bar{T} = T \circ \phi^{-1} \tag{12}$$

Therefore, the integrability of α over M_x in the x-space can be transposed in terms of integrability of $\bar{\alpha}$ over D_z.

Note that $\bar{\alpha}$ only depends on z_1, z_2, z_3 over D_z, i.e.,

$$\bar{\alpha}(z) = \sum_{i=1}^{6} \bar{\alpha}_i(z_1, z_2, z_3)dz_i \qquad , \quad z \in D_z.$$

If we choose the 1-form α such that $\bar{\alpha}_1$, $\bar{\alpha}_2$, and $\bar{\alpha}_3$ depend on z_1, z_2, and z_3 only, respectively. We can define

$$\bar{T}(z) = \sum_{i=1}^{6} \int \bar{\alpha}_i(z_1, z_2, z_3)dz_i \tag{13}$$

and we have

$$\begin{aligned} d\bar{T}(z) &= \sum_{i=1}^{6} \bar{\alpha}_i(z_i, z_2, z_3)dz_i + \sum_{i=4}^{6} z_i\Big[\sum_{j=1}^{3}(\partial\bar{\alpha}_i/\partial z_j)dz_j\Big] \\ &= \sum_{i=1}^{6} \bar{\alpha}_i(z_i, z_2, z_3)dz_i \\ &= \bar{\alpha} \end{aligned}$$

Since the second term of the right hand side vanishes over D_z. Therefore, $\bar{\alpha}$ is integrable over D_z and so is α over M_x.

Method

The procedure for searching for these transformations can be decomposed into three steps.

Step 1: Find three independent directions α_1, α_2, α_3 perpendicular to G.

Step 2: From eqns. (7b) (resp. (7c)), α_4, α_5, α_6 (resp. α_7, α_8, α_9) can be computed.

Step 3: Integration of $\alpha_1, \ldots, \alpha_6$ (resp. α_7, α_8, α_9) along M_x (resp. $M_{x,u}$) provides the desired mapping $T_1, \ldots, T_9$.

Application

Consider a rigid body in which the angular velocities are measured along a body fixed axis, and with an Euler angle sequence (space - three 1-2-3) defining the attitudes in inertial space:

$$\dot{\omega} + \omega \times I\omega = u$$

$$\begin{aligned} \dot{\theta}_1 &= \omega_x + (\omega_y s_1 + \omega_z c_1) s_2/c_2 \\ \dot{\theta}_2 &= \omega_y c_1 - \omega_z s_1 \\ \dot{\theta}_3 &= (\omega_y s_1 + \omega_z c_1)/c_2 \end{aligned} \qquad (14)$$

where $\omega = (\omega_x, \omega_y, \omega_z)$ is the angular velocity vector

$\theta = (\theta_1, \theta_2, \theta_3)$ is the Euler angle vector

$u = (u_x, u_y, u_z)$ is the external torque acting on the rigid body

$s_1 = \sin\theta_1$, $c_1 = \cos\theta_1$, $s_2 = \sin\theta_2$, $c_2 = \cos\theta_2$

and I, a 3x3 matrix, is the inertial tensor.

Without loss of generality, principle axes are used to describe this system. Therefore, we can assume the matrix I is a diagonal matrix, i.e.,

$$I = \begin{bmatrix} I_x & 0 & 0 \\ 0 & I_y & 0 \\ 0 & 0 & I_z \end{bmatrix}$$

Using principle axes, system (14) become:

$$\begin{aligned}
\dot{\omega}_x &= [u_2 - (I_z - I_y)\omega_y\omega_z]\frac{1}{I_x} \\
\dot{\omega}_y &= [u_y - (I_x - I_z)\omega_x\omega_z]\frac{1}{I_y} \\
\dot{\omega}_z &= [u_z - (I_y - I_x)\omega_x\omega_y]\frac{1}{I_z} \\
\dot{\theta}_1 &= \omega_x + (\omega_y s_1 + \omega_z c_1)s_z/c_z \\
\dot{\theta}_2 &= \omega_y c_1 - \omega_z s_1 \\
\dot{\theta}_3 &= (\omega_y s_1 + \omega_z c_1)/c_2
\end{aligned} \tag{15}$$

The set of operating points is

$$M_{x,u} = \{(0, 0, 0, \theta_1, \theta_2, \theta_3, 0, 0, 0) \mid \frac{-\pi}{2} < \theta_2 < \frac{\pi}{2}\} \tag{16}$$

and its projection in the state space is

$$M_x = \{(0, 0, 0, \theta_1, \theta_2, \theta_3) \mid \frac{-\pi}{2} < \theta_2 < \frac{\pi}{2}\} \tag{17}$$

On $M_{x,u}$,

$$F(x_o,u_o) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & s_1s_2/c_2 & c_1s_2/c_2 & 0 & 0 & 0 \\ 0 & s_1 & -s_1 & 0 & 0 & 0 \\ 0 & s_1/c_2 & c_1/c_2 & 0 & 0 & 0 \end{bmatrix}$$

and

$$G(x_o,u_o) = \begin{Bmatrix} 1/I_x & 0 & 0 \\ 0 & 1/I_y & 0 \\ 0 & 0 & 1/I_z \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{Bmatrix}$$

It is clear the system is controllable at every point $(x_o,u_o) \in M_{x,u}$.

Now, we shall choose three independent vectors α_1, α_2, α_3 all orthogonal to G.

Let $\alpha_1 = (0, 0, 0, 1, 0, 0)$, $\alpha_2 = (0, 0, 0, 0, 1, 0)$ and $\alpha_3 = (0, 0, 0, 0, 0, 1)$.

From (7b) and (7c), we obtain

$$\alpha_4 = (1, s_1 s_2/c_2, c_1 s_2/c_2, 0, 0, 0).$$

$$\alpha_5 = (0, c_1, -s_1, 0, 0, 0)$$

$$\alpha_6 = (0, s_1/c_2, c_1/c_2, 0, 0, 0).$$

$$\alpha_7 = (0, 0, 0, 0, 0, 0, 1/I_x, s_1 s_2/I_y c_2, c_1 s_2/I_z c_2),$$

$$\alpha_8 = (0, 0, 0, 0, 0, 0, 0, c_1/I_y, -s_1/I_z),$$

$$\alpha_9 = (0, 0, 0, 0, 0, 0, 0, s_1/I_y c_2, c_1/I_z c_2).$$

After integrating α_i, $i = 1, \ldots, 6$, (resp. α_7, α_8, α_9) over M_x (resp. $M_{x,u}$), the desired transformations are:

$$z_1 = \theta_1$$

$$z_2 = \theta_2$$

$$z_3 = \theta_3$$

$$z_4 = \omega_x + (\omega_y s_1 + \omega_z c_1)s_2/c_2 = \dot{\theta}_1$$

$$z_5 = (\omega_y c_1 - \omega_z s_1) = \dot{\theta}_2 \qquad (18)$$

$$z_6 = (\omega_y s_1 + \omega_z c_1)/c_2 = \dot{\theta}_3$$

$$v_1 = v_{1a} + v_{1b}$$

$$v_2 = v_{2a} + v_{2b}$$

$$v_3 = v_{3a} + v_{3b}$$

where

$$v_{1a} = u_z/I_x + (u_y s_1/I_y + u_z c_1/I_z)s_2/c_2,$$

$$v_{1b} = \{[(I_z - I_x)\omega_z s_1/I_y + (I_x - I_y)\omega_y c_1/I_z]\omega_x s_2/c_2 + (I_y - I_z)\omega_y\omega_z/I_x\} + (\omega_y c_1 - \omega_z s_1)[\omega_x s_2/c_2 + (\omega_y s_1 + \omega_z c_1) \times (1 + 2s_2^2/c_2^2)\ ,$$

$$v_{2a} = u_y c_1/I_y - u_z s_1/I_z,$$

$$v_{2b} = [(I_z - I_x)\omega_z c_1/I_y + (I_x - I_y)\omega_y s_1/I_z]\omega_x - (\omega_y s_1 + \omega_z c_1)[\omega_x + (\omega_y s_1 + \omega_z c_1)s_2/c_2],$$

$$v_{3a} = (u_y s_1/I_y + u_z c_1/I_z)/c_1,$$

$$v_{3b} = [(I_z - I_x)\omega_z s_1/I_y + (I_x - I_y)\omega_y c_1/I_{zz}]\omega_x/c_z + (\omega_y c_1 - \omega_z s_1)[\omega_x + (\omega_y s_1 + \omega_z c_1)(1 + 1/c_2)s_2/c_2]/c_2$$

Note that each transformed input v_i ($i_1 = 1, 2, 3$) is partitioned into v_{ia} and v_{ib} terms. The v_{ia} term arises directly from integration of the corresponding α_{6+i} vector. The v_{ib} term can be identified with the nonlinear terms which appear in the expression of $\dot{z}_{3+i}$. According to [1], these high order terms may be neglected in the neighborhood of the operating point, in which case the equations of motion are said to be psuedo-linearized. However, in this particular case, since the operating points require $\omega = 0$, the transformation from v_i can be augmented to include the v_{ib} term without affecting the requirement that $dv_{ib}\Big|_{M_{x,u}} = \alpha_j$, $j = i+6$. Therefore, as shown before, the equation of motion for a rigid body become extirely linear when expressed in terms of Euler angles and Euler angular rates.

Taking account of (18), (15) can be rewritten as a linear systems:

$$\dot{z}_1 = z_4, \quad \dot{z}_2 = z_5, \quad \dot{z}_2 = z_6, \quad \dot{z}_4 = v_1, \quad \dot{z}_5 = v_2, \quad \dot{z}_6 = v_3,$$

or, in a matrix form:

$$\begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \\ \dot{z}_3 \\ \dot{z}_4 \\ \dot{z}_5 \\ \dot{z}_6 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \\ z_5 \\ z_6 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \qquad (19)$$

Furthermore, the inverse transformations also can be found as:

$$\Theta_1 = z_1$$

$$\Theta_2 = z_2$$

$$\Theta_3 = z_3$$

$$\omega_y = - z_5 s_1 + z_6 c_1 c_2$$

$$\omega_y = (z_5 + \omega_z s_1)/c_2 \qquad , \quad -\frac{\pi}{2} < \Theta_1 < \frac{\pi}{2} \tag{20}$$

$$\omega_x = z_4 - (\omega_y s_1 + \omega_z c_1) s_2/c_2 \qquad , \quad -\frac{\pi}{2} < \Theta_2 < \frac{\pi}{2}$$

$$u_x = I_x[(v_1 - v_{1b}) - (v_3 - v_{3b})s_2]$$

$$u_y = I_y[(v_2 - v_{2b})c_1 + (v_3 = v_{3b})c_2 s_1]$$

$$u_z = I_z[c_1 c_2 (v_3 - v_{3b}) - s_1 (v_2 - v_{2b})]$$

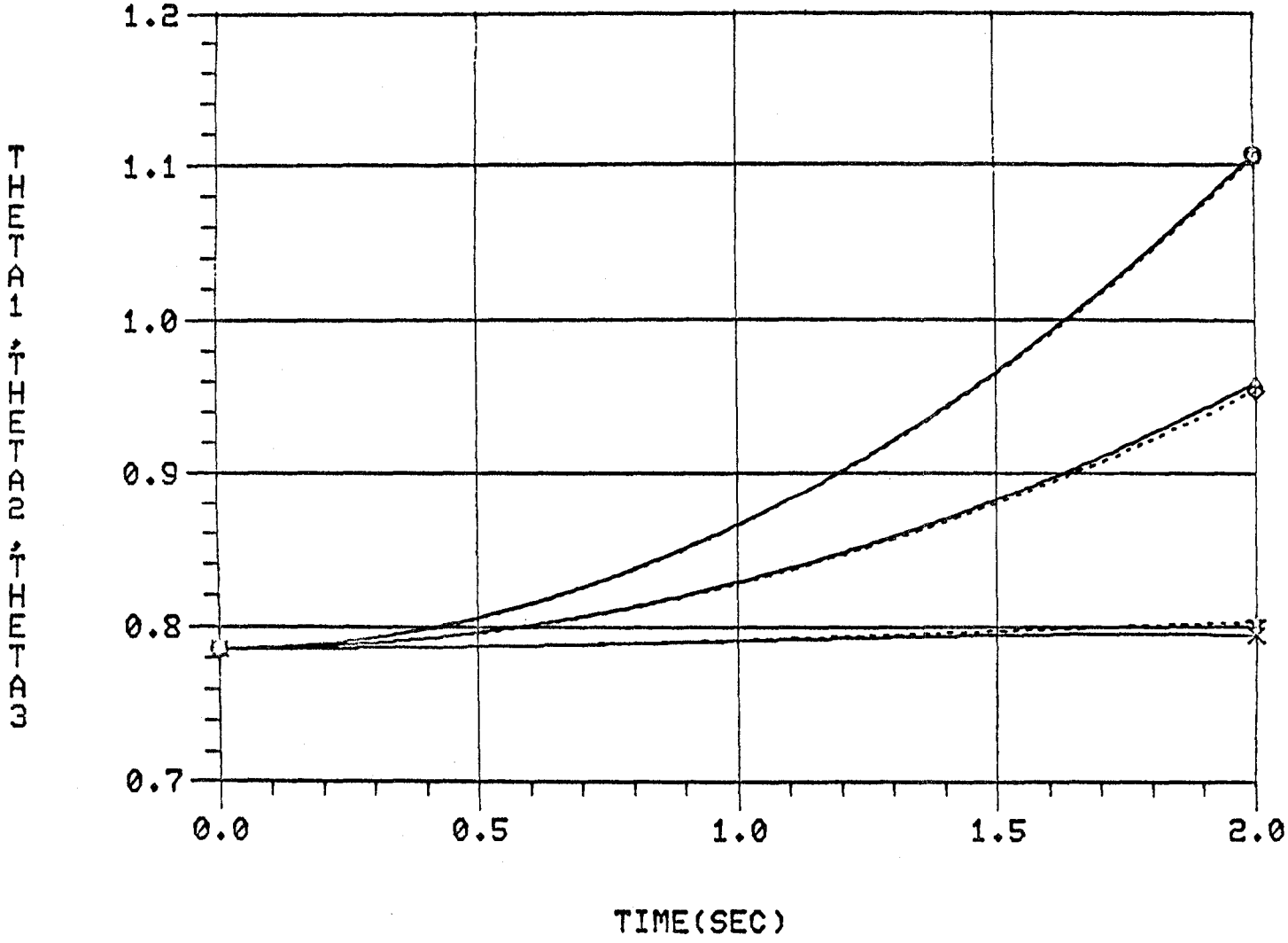
The above transformations were verified by numerical integration in both the x-space and z-space for a prescribed input u. Trajectories obtained by integrating in x-space were virtually identical to those obtained by integrating in z-space then transforming to x-space via equation (20). Figure 1 shows the effect of psuedo-linearization at an operating point of $\Theta_1 = \Theta_2 = \Theta_3 = \pi/4$ with a unit step input for u. The solid lines are trajectories obtained by directly integrating equation (15), while the dashed curves represent the transformation from z-space to x-space neglecting the v_{ib} terms.

For control studies, a constant gain, full state feedback control $v = -Kz$ was employed. Working in the z-space with $\dot{z} = Az + Bv$, the following gain matrix K allows for arbitrary, complex conjugate pole placement,

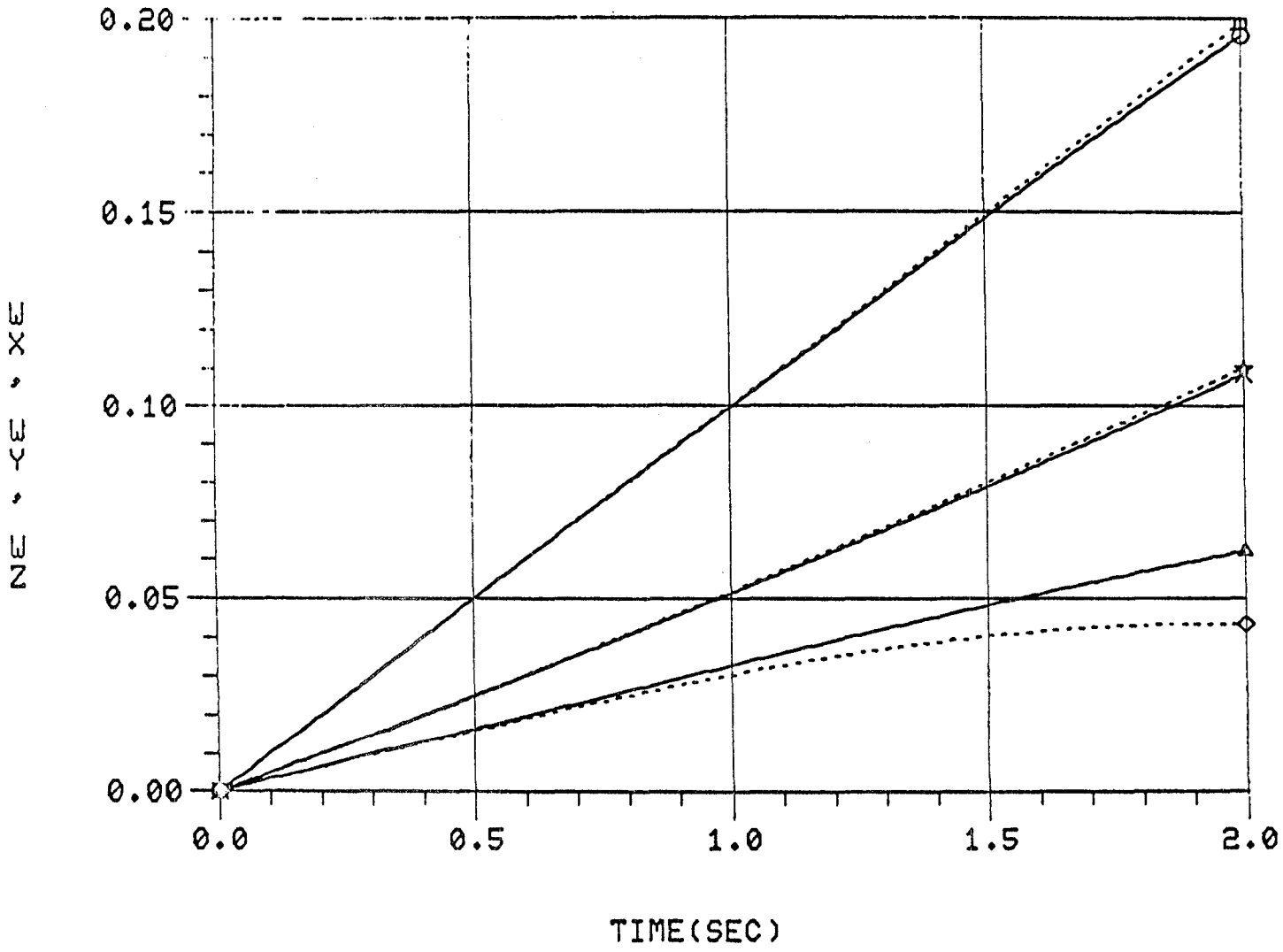
$$K = \begin{bmatrix} \omega_1^2 & 0 & 0 & 2\zeta_1\omega_1 & 0 & 0 \\ 0 & \omega_2^2 & 0 & 0 & 2\zeta_2\omega_2 & 0 \\ 0 & 0 & \omega_3^2 & 0 & 0 & 2\zeta_3\omega_3 \end{bmatrix} \qquad (21)$$

where ω_i and ζ_i prescribe the closed loop frequency and damping ratio of the ith axis. The actual control to be used in x-space comes from equations (18), (20) and (21), where both ω and Θ are assumed to be available as plant measurements. Figures 2, 3 show the system response to the arbitrary initial conditions $\Theta_1 = \Theta_2 = \Theta_3 = \pi/4$ and $\omega_x = \omega_y = \omega_z = 0.5$ using control frequencies and dampings of $\omega_1 = \omega_2 = \omega_3 = 1.0$ and $\zeta_1 = \zeta_2 = \zeta_3 = 0.707$. Once again, the solid lines represent the response using the exact transformation of equation (20) while the dashed lines show the effect of psuedo-linearization.

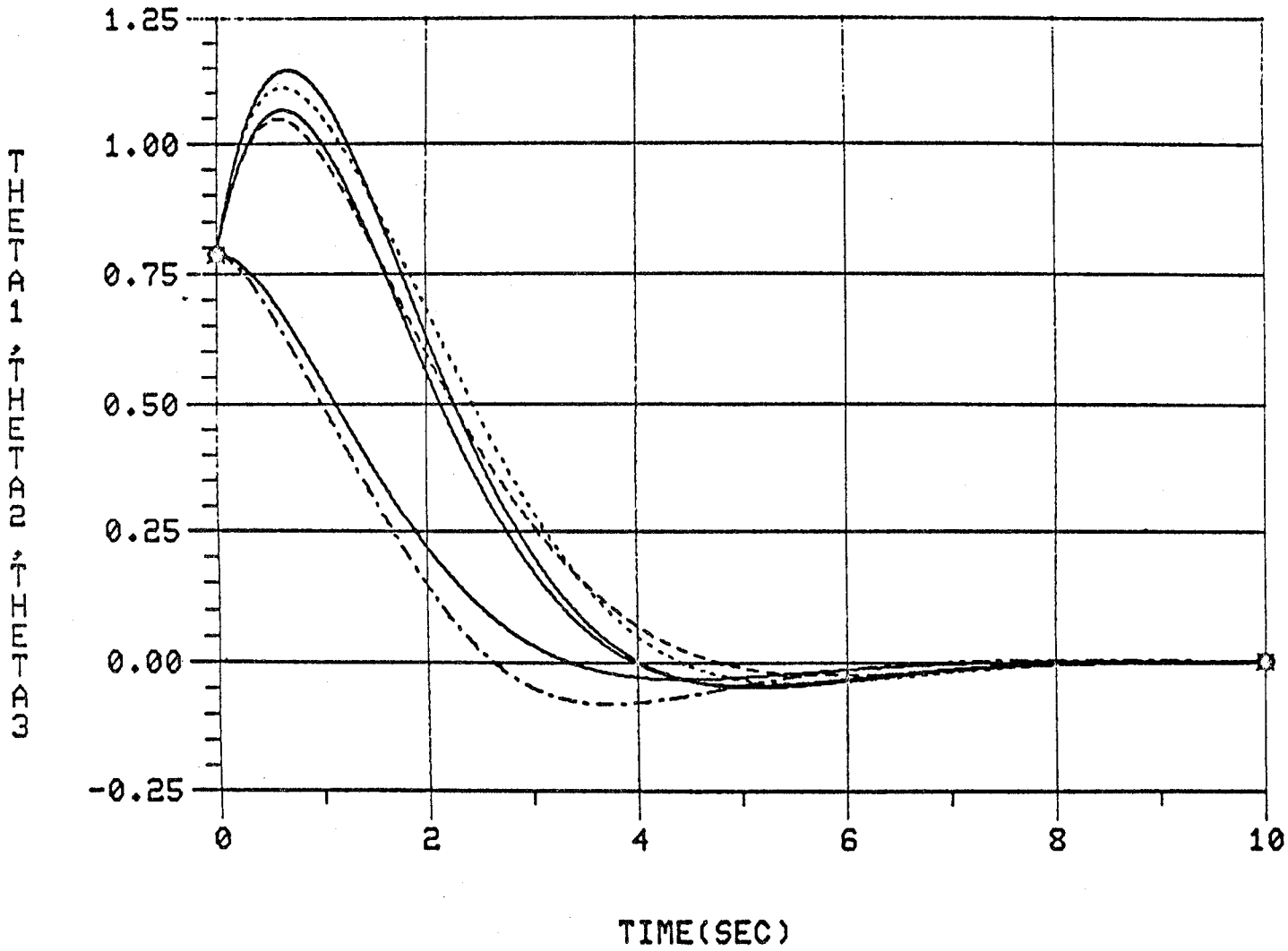
OPEN LOOP EFFECTS OF PSUEDO-LINEARIZATION



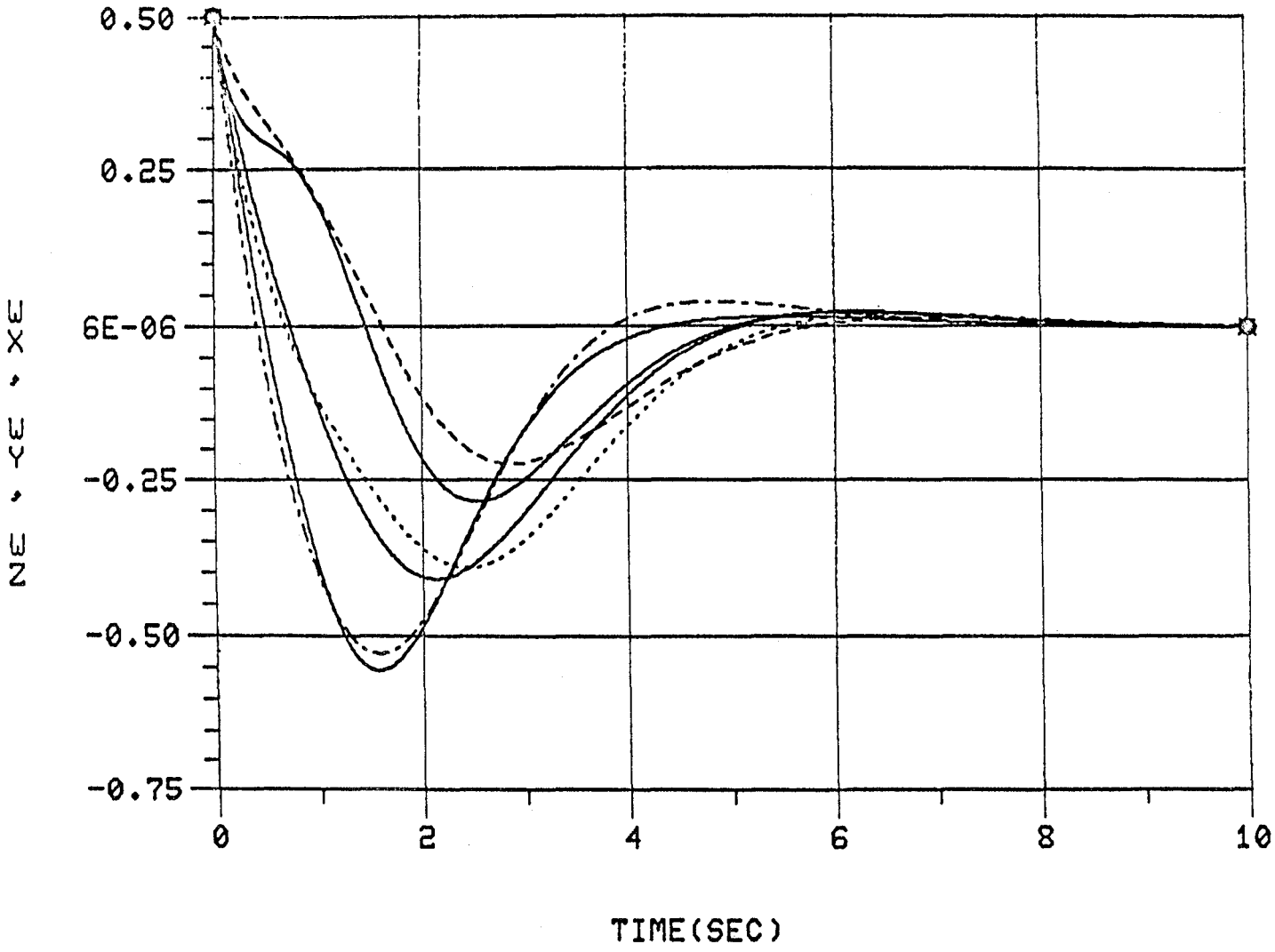
OPEN LOOP EFFECTS OF PSUEDO-LINEARIZATION



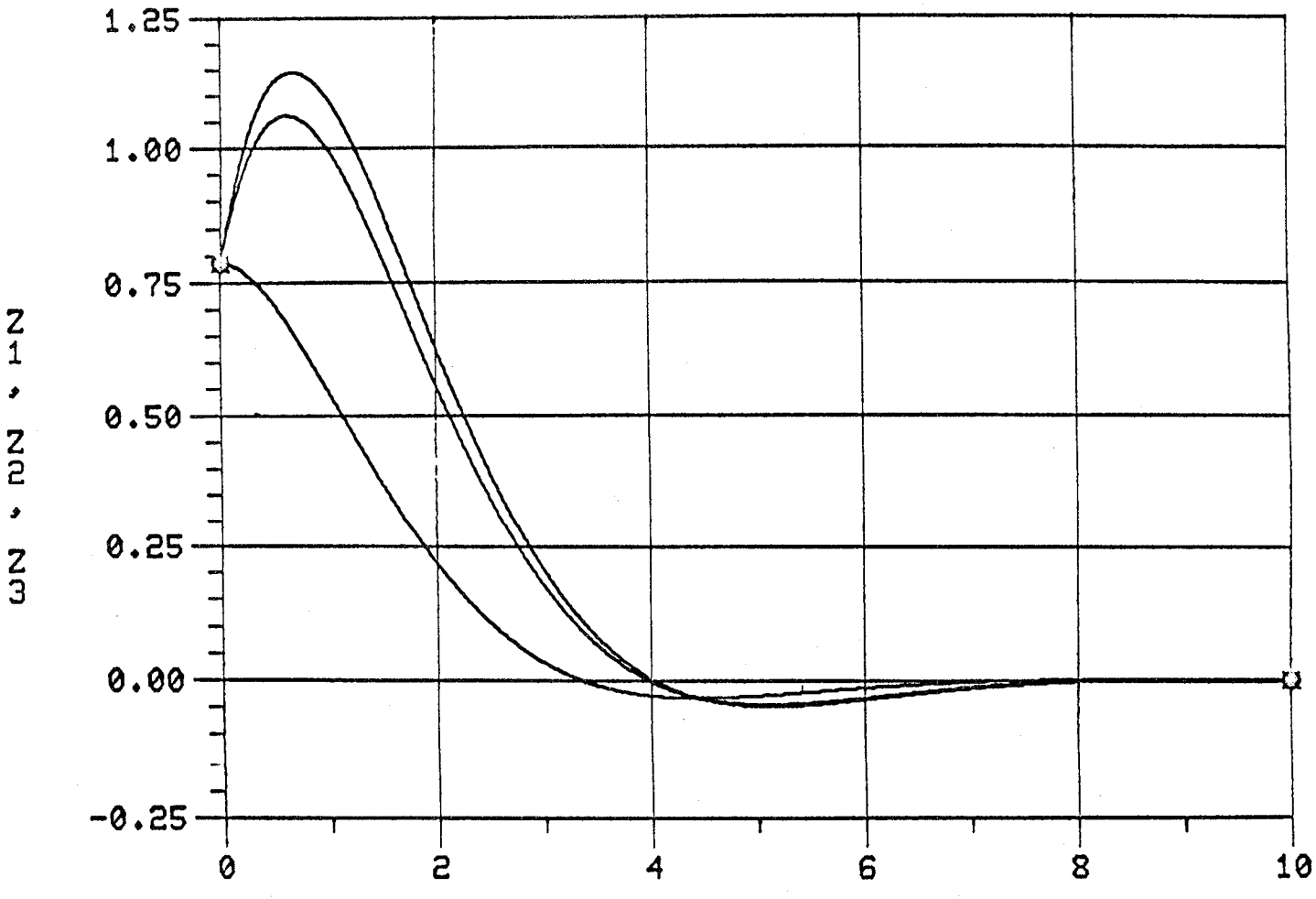
CLOSED LOOP EFFECTS OF PSUEDO-LINEARIZATION



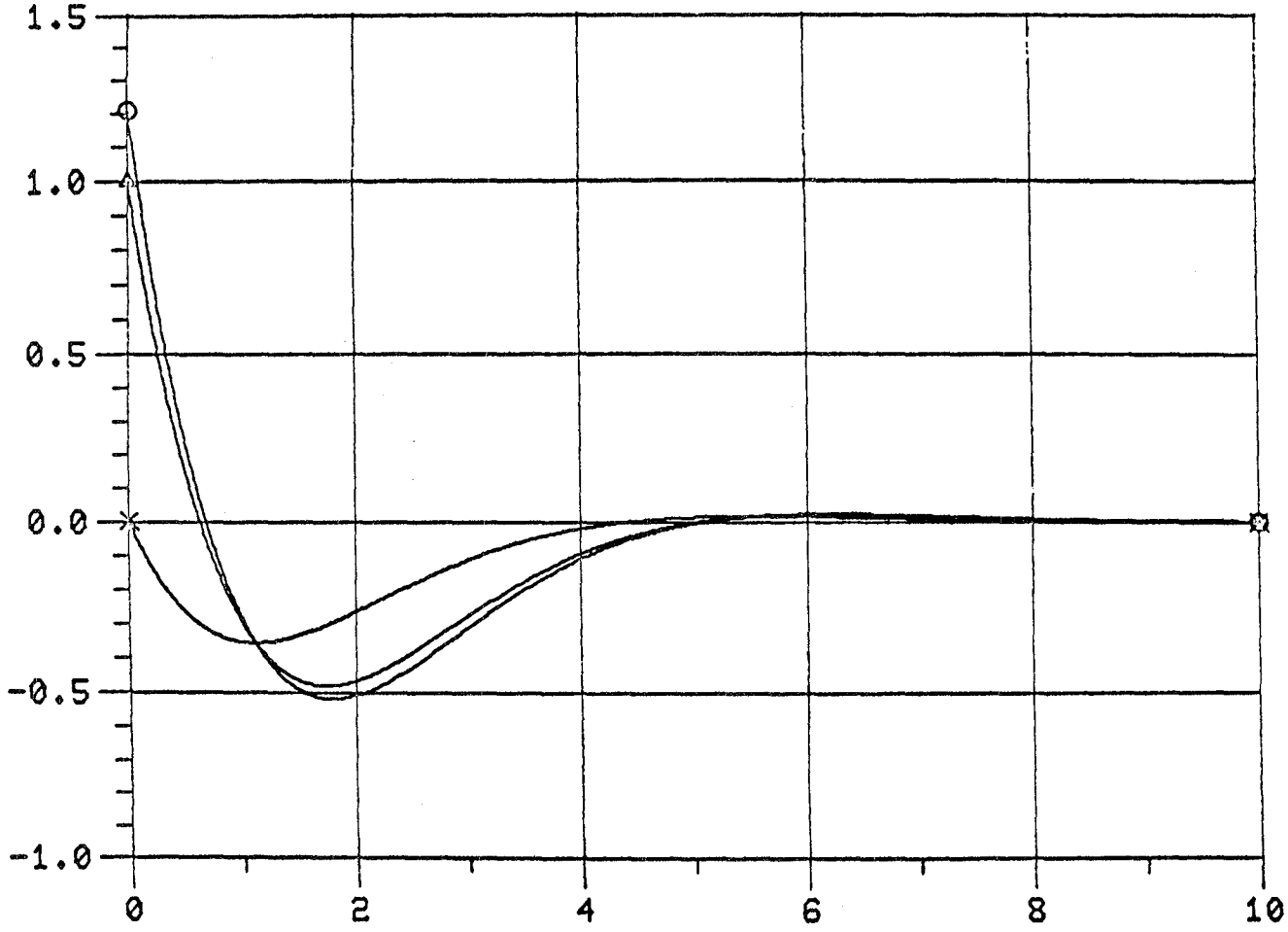
CLOSED LOOP EFFECTS OF PSUEDO-LINEARIZATION



CLOSED LOOP CONTROL IN THE Z - PLANE



CLOSED LOOP CONTROL IN THE Z - PLANE



CONCLUSION AND RECOMMENDATION

A global linearization method for a nonlinear control system with 6 states and 3 inputs is successfully developed in this project. This method is an extension of a pseudo-linearization technique proposed by Reboulet and Champetier [1] in which only one input is considered. After applying this method to a rigid body in which the angular velocities are measured along a body fixed axis, and with an Euler angle sequence defining the altitudes in inertial space, the original non-linear system is described by a linear system in a transformed space. Therefore, a global control law can be established accordingly.

For further application of this method, we will turn our attention to the case of control systems of multiple bodies. Applying this method, Mr. Sharkey has obtained very positive numerical results on a two-body model using pole placement for a control law. We suggest a mathematical research to support this new application which is necessary and should start immediately after this program to pave the road for the next summer's project.

REFERENCE

1. Reboulet, A. and Champetier, C., "A New Method for Linearizing Non-linear Systems: The Pseudo-Linearization," International Journal of Control, 1984, Vol. 40, No. 4, 631-638.

1985

NASA/ASEE SUMMER FACULTY RESEARCH FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

ON THE DETERMINATION OF THE ORIGIN OF LINEAR ANOMALY IN THE MACROSTRUCTURE OF VPPA WELDED 2219-T87 ALUMINUM ALLOY

Prepared by:	Wartan A. Jemian
Academic Rank:	Professor
University: Department:	Auburn University Mechanical and Materials Engineering
NASA/MSFC:	
Laboratory:	Materials and Processes Lab
Division:	Processes Engineering
Branch:	Metal Processes
MSFC Counterpart:	Arthur Nunes
Date:	August 16, 1985
Contract Number:	NGT 01-008-021 The University of Alabama in Huntsville

FINAL REPORT

ON THE DETERMINATION OF THE ORIGIN

OF LINEAR ANOMALY IN THE MACROSTRUCTURE OF

VPPA WELDED 2219-T87 ALUMINUM ALLOY

by Wartan A. Jemian, Ph. D.
Professor of Mechanical Engineering and Materials Engineering
Auburn University
Alabama

16 August 1985

ABSTRACT

The objective of this summer research was to determine the cause and significance of the weld radiograph enigma, which is a linear anomaly in the features of the x-ray film. By observing features on available radiographs and in studying published reports of similar features it was possible to conclude that there are many manifestations of the enigma, and that they are all specific features of fine structure in radiographs due to natural processes connected with welding and to specific X-ray absorption and diffraction phenomena. These processes include the thermal distribution and liquid metal flow in welding, the development of microstructure, morphology, second phase particles and porosity due to the solidification process and to the pattern of residual stresses after the weld metal has cooled to the ambient temperature. Microdensitometer traces were made across weld radiographs of standard and enigmatic types. Similar patterns were produced by computer simulation. These show that the enigma is a relatively low contrast feature compared to real weld defects, such as undercuts or centerline cracks. The enigma can be distinguished from weld defects by these microdensitometer traces. The enigma effect on weld properties is not known but is expected to be minor.

ACKNOWLEDGEMENTS

This work was primarily of a survey nature but is dependent on the experimental work of Mr. E. O. Bayless and his staff in experimentally simulating these features. It also benefitted from the information and inspiration of Dr. Arthur C. Nunes. Both are my counterparts. Any success is directly due to their guidance and cooperation.

It is my pleasure to acknowledge the assistance of several other individuals who helped me directly. They are Mr. Carl Wood, who was totally sympathetic and supportive of my needs with respect to basic engineering information. Mr. Paul H. Schuerer inherited me into his branch with good grace and was very encouraging in his support. Messrs. Paul Munafo, Bobby Rowe and Joe Montano offered constructive criticism. Claud Williamon and I worked together collecting x-ray data on fluorescence effects and Jim Coston was very helpful in analogous studies using the SEM. Dr. Walter Fountain provided the critical bit of experimental evidence to confirm the basic ideas concerning the enigmas. Mr. C. Lovoy, USBI, retired NASA staff member has taken an interest in my acitivities and provided helpful background and information on several occassions. Several individuals of the Martin Marietta staff at Michoud were helpful in providing information. I am especially grateful to Messrs. Monty McAndrews and W. Mattheessen and to Ms. Linda Johnston.

Introduction

There is an anomalous, generally linear feature in line with the weld direction that appears intermittently in weld x-ray radiographs. It bears an appearance similar to that of a defect, but for which there is no defect or discontinuity in the structure of the weld metal and no apparent effect on the weld mechanical properties. It is commonly called an *enigma* or *ghost defect*. Duren and Risch reported in 1970 (1) that: *The enigma is the most difficult discontinuity to determine, often being mistaken for incomplete fusion, and many times for a crack, or even a diffraction pattern. Identification of an enigma demands long experience in interpreting radiographs. The interpreter should notice that the dark line is always accompanied by a parallel light line. Many destructive tests for strength have been made of this type of discontinuity, but so far there has been no noticeable effect on the weld's ultimate strength.* This represents the understanding 15 years ago and indicates that the enigma must have been considered a problem for some time before that period. Welds of that time were not plasma arc, therefore the phenomenon is general.

Present indications are that there are two or more distinct manifestations of the enigma that occur separately. These are the dark form, frequently occurring along the weld centerline, and the gray or white line at the side of the weld, just within the fusion boundary. Real weld

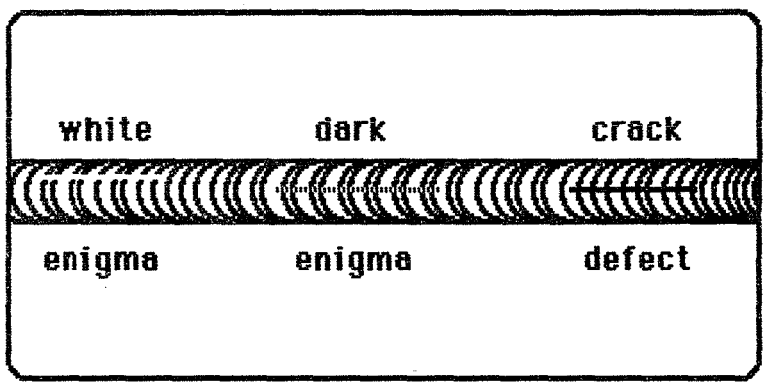


Figure 1. The typical appearance of enigmas and crack indications in a weld radiograph.

defects are not a manifestation of the enigma. Both features, which have

been observed separately, have the appearance of a serious weld defect, but on close examination, by metallographic sectioning, have been found to be free of any defect. The problem is that the condition is misleading and disconcerting. The appearance is similar to that shown in Figure 1.

Recently, L. Johnston (2) reported that *the . . .enigma occurrence was characterized by a grey line running along the toe side of the weld. The line ran continuously, only occurring on one side at a time. No violations of weld parameters were found. The trim area was large enough for one tensile bar and one macro specimen. The UTS of the tensile bar was greater than 40 KSI. Visual examination of the weld revealed no cracks, undercut, suckback, open porosities or tunnels, or other evidences of inconsistent welding processes. No anomalies were found in ultrasonic inspection. Post-proof x-ray showed no change of character, position or intensity.* This report was corroborated by Kinchen and Brown (3).

The suggestion was made that the phenomenon is a diffraction effect, similar to Kikuchi lines in electron diffraction or Kossel lines in Xray diffraction. Rummell and Gregory (4) reported a similar occurrance in radiographs of TIG welded 2014 aluminum alloy. *Recent studies on welding defects have revealed a phenomenon which appears as "lack of fusion" or "lack of penetration" in a weld radiograph, but does not affect the performance of the weld. . . . indications are generally not as sharp as true lack of fusion, and are curved at the ends. . . . Macro and micro examination of the weld cross section shows a dendritic grain structure in the fusion zone . . . . Grain orientations are at right angles to each other. . . . The unique grain orientation is due to a particular cooling and freezing process in the weld. . . . Now, if a series of crystals are properly oriented with respect to an X-ray beam, a "focusing" effect will be observed on the radiograph, in the form of a dark band. . . . The white line is caused by deviation of the transmitted beam by the diffraction process, thus decreasing the total transmitted radiation in a particular line or band. . . . The diffraction process is not limited to the weld structure illustrated, but will occur in any exposure when crystalline planes are properly oriented with respect to the X-ray beam. Likewise, this process is not limited to aluminum welds.*

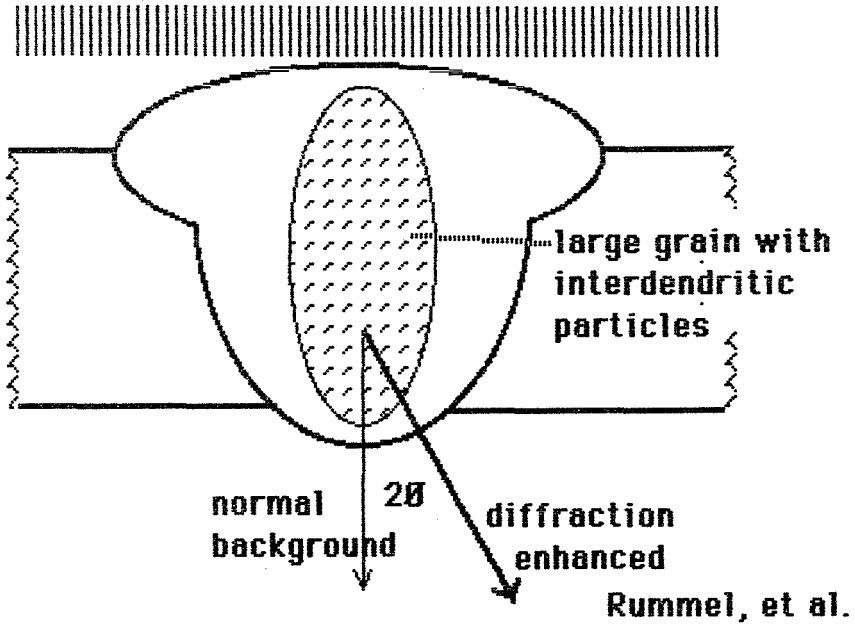


Figure 2. X-ray diffraction as a cause of the dark line weld radiograph enigma.

Figure 2 illustrates the general features of the Rummell model. The interdendritic particles shield the film producing the light line and the diffracted beam produces the dark band. Large grains of this type are encountered occassionally. Since the solidification mode is generally dendritic and the interdendritic constituent is of an eutectic type the second phase particles are likely to be coherent, or at least highly aligned, with the dendritic branches. Finally, since the film is generally placed very close to the sample, the physical separation of the direct and diffracted beams on the film is also close.

Hirosawa et al. (5) reviewed the literature, including the work of Rummell et al. and Tucker et al., who reported similar findings and conclusions to those of Rummell (6). Rabkin et al. also reported dark band enigma formation in Al-Mg alloy welds (7). Issiki reported similar effects in aluminum alloy castings (8), and Irie, et al. reported similar features in stainless steel welds (9). Hirosawa et al. reported an analytical procedure related to magnesium content in aluminum alloys (5).

It is evident that the enigma, ghost defect or linear anomaly has been encountered for a number of years and in a number of alloy systems and welding processes. It appears to consist of dark or light lines, separately or together, primarily straight, in alignment with the welding direction, but also may have a curvature. The curvature is due to changing

conditions along the weld path. All reports place these features in the fusion zone. The dominant questions to be answered in finding the solution to the weld radiograph enigma problem are:

1. What is the enigma?

2. What produces the effect?

3. How can it be recognized?

4. How are weld properties affected?

The question about the effect on properties is fundamental. However, it must be realized that even if there is a visible feature in the sample there are established size tolerances. Perhaps the greatest problem is the inability to distinguish the enigma from a weld defect. A method was evaluated for this purpose and is described in a later section. There are many manifestations of the enigma. It will be shown to be a fine structure in the weld radiograph that is related to a real condition in the weld. The occurrance of this fine structure will be related to such factors as concentration gradients, shape effects and possibly features as severe as porosity. Even these latter, at levels to produce a visible feature, are probably within the established size tolerances. The correct interpretation of these features requires an understanding of the principles of welding physical metallurgy and image formation in x-ray radiography.

Experimental

The space shuttle external tank is fabricated, primarily, of aluminum alloy 2219-T87. The initial joining process was TIG welding which is being replaced by VPPA welding. The filler material is 2319 aluminum alloy wire. Weld radiograph enigmas were reported for TIG welds as well as the VPPA process. TIG weld radiographs show a higher level of density variation than VPPA radiographs, therefore fine structure is not as easily detected.

In most thicknesses, two weld passes are used, both from the same side. The root pass is in the keyhole mode and the cover pass in partial penetration, frequently referred to as "in the TIG mode." The plasma has a remarkable cleaning action. The heavy ions break up the tenacious oxide layer and the fragments are swept away by the plasma jet. If the

torch is properly aligned and the parameters properly set, a flawless weld results. Even the cover pass maintains the high level of structural integrity due to the action of the keyhole pass which leaves a clean system. The shape of the liquid puddle during the vertical up keyhole pass is shown in Figure 3.

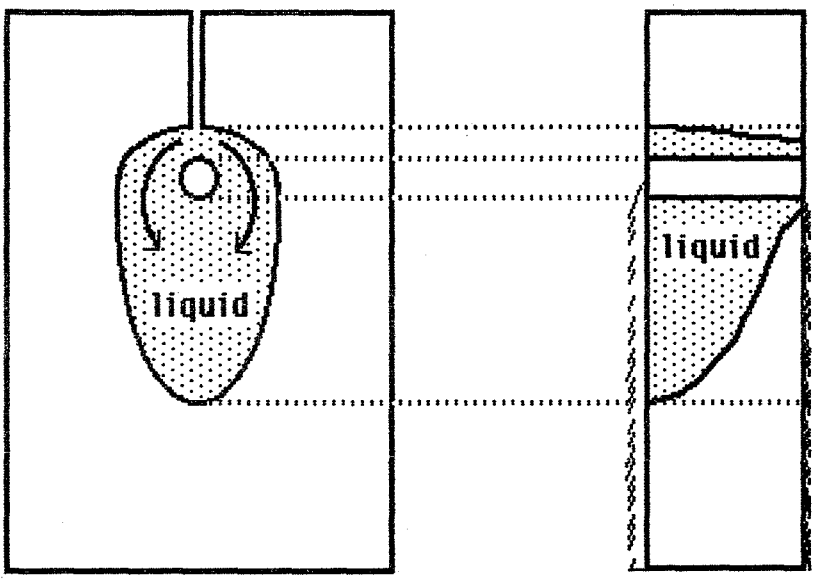


Figure 3. General shape of the keyhole mode weld pool. The shaded portion is the liquid.

The final weld structure is characterized by crown and root reinforcements making a smooth transition to the base metal plate on both surfaces. Figure 4 describes the general appearance of the weld after the root and cover passes. The root reinforcement is produced in the keyhole pass and the top surface of the weld is left approximately flush. Thinner plates are welded in only a single pass leaving reinforcements on both surfaces. The cover pass remelts a portion of the fusion zone of the previous keyhole pass and part of the base metal. The symmetry shown in the sketches of Figure 4 can be expected as the result of symmetric torch alignment. The heat affected zone exhibits only a minor change of grain size. The hardness profile across the weld from one base metal plate to the other shows a generally symmetrical decrease in hardness towards the weld centerline with a moderate increase close to the fusion boundary in the heat affected zone. This small increase is due to an age hardening effect.

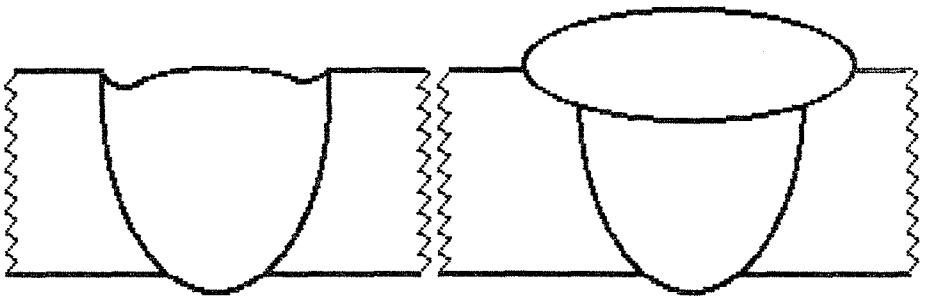


Figure 4. Shape of the weld after the root pass (left) and the cover pass (right) in VPPA welding.

X-ray radiography uses a tungsten target tube operated at a tube voltage of 80 Kv and a tube current of 15 ma. The typical standoff distance is 40 inches. In most cases the film is place in direct contact at the back of the sample. However, in one of the fixtures this space is restricted so that the film is approximately 1.25 inches away. The exposed film is processed automatically and is read and interpreted visually. Other NDE procedures are employed to clarify and decide specific conditions. Present practice requires 100% radiographic inspection.

Three densitometers were used to obtain line traces across the weld pattern. These are the MacBeth TD 102 densitometer with an aperture of approximately 3 mm diameter, an instrument which is designed for analyzing spectroscopic plates, and the Perkin Elmer PDS Microdensitometer developed for analyzing stellar plates. The Perkin Elmer instrument was operated with a square aperture of 125 μm side dimension at contiguous fields. An offset density of 1.5 was uniformly used for all measurements. The data was printed out and manually transferred to a personal computer for processing. The second instrument could not be used due to its inability to work with high density radiographs.

The scanning electron microscope with an EDAX chemical analysis attachment was used to map the concentration of copper and iron over transverse sections of the weld.

Radiography, based on the fluorescence of alloying elements, was used to map concentration distribution. In another investigation of fluorescence effects, the standard x-ray facility was operated at a standoff distance of 40 inches over a range of tube voltages. The film (DuPont

Cronex 45) was placed over blocks of pure copper and lead, placed side by side. The exposure was controlled by setting tube current, exposure time and using various thicknesses of 5052 aluminum alloy plate.

Contributing Factors

Welding:

Welding exerts its influence on the condition of the final solid through all of its operating parameters. Conditions in the liquid pool control weld geometry, grain shapes and final distortion and residual stresses. The hot, just solidified alloy behind the pool, contracts as it cools, whereas the base metal away from the weld may not have any tendency to contract. If the effective center of this initially hot metal is out of line with the center of mass of the plate, the contraction produces a net bend and leave a serious pattern of residual stresses as shown in Figure 5, which illustrates the directions of these distortions. This is explained in detail by Masubuchi (10). These stresses, in concert with applied stresses can lead to premature service failure. If a stress riser, or other mechanically sensitive feature is in the field of influence of the stress pattern, cracking can occur.

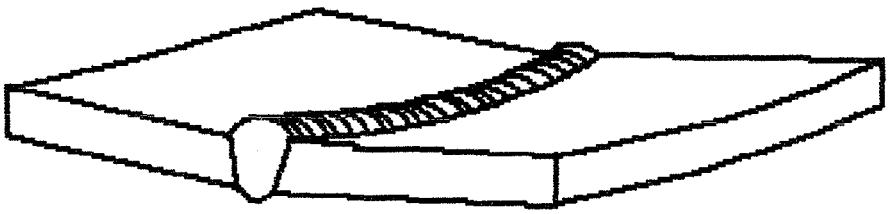


Figure 5. Shrinkage distortion in a butt weld. This distortion includes peaking and longitudinal curvature.

Fixturing can be used to control or reduce distortion. It also exerts an effect on the final microstructure by alteration of the heat sink configuration. This modifies the pattern of grain growth and affects

properties. The effect on grain growth direction is illustrated in Figure 6. In a previous study (11) the application of cooling blocks along a short region at the side of the weld produced a finer microstructure and increased the mechanical properties in that section.

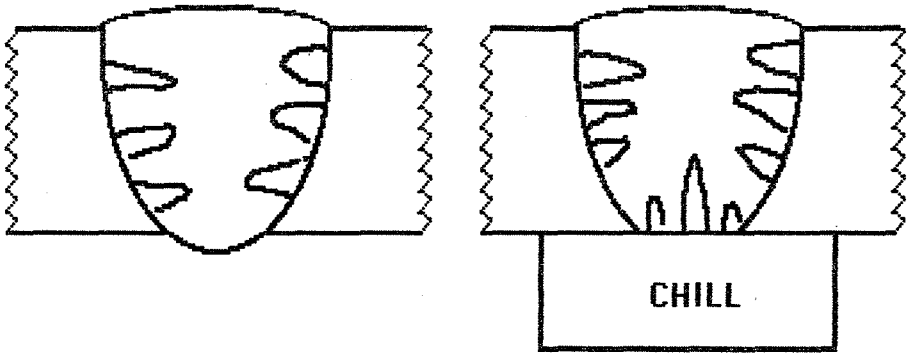


Figure 6. The effect of heat sink configuration on the weld metal grain structure.

Welding power and speed affect the final weld structure through their action on final cooling rate. Power must, of course, be set to control penetration and width. The higher the speed the more power is required to penetrate to the proper depth. At higher welding speeds the base metal is not heated to the same extent and has a greater cooling effect after the torch passes. This changes weld shape and affects the direction of heat flow. The desired ellipsoidal shape of the weld pool becomes pointed at the back as illustrated in Figure 7. The consequence is that the direction of columnar grain growth in the fusion zone is affected. If the condition is sufficiently pronounced a plane of weakness is formed along the weld centerline. Cooling stresses may then result in centerline cracking.

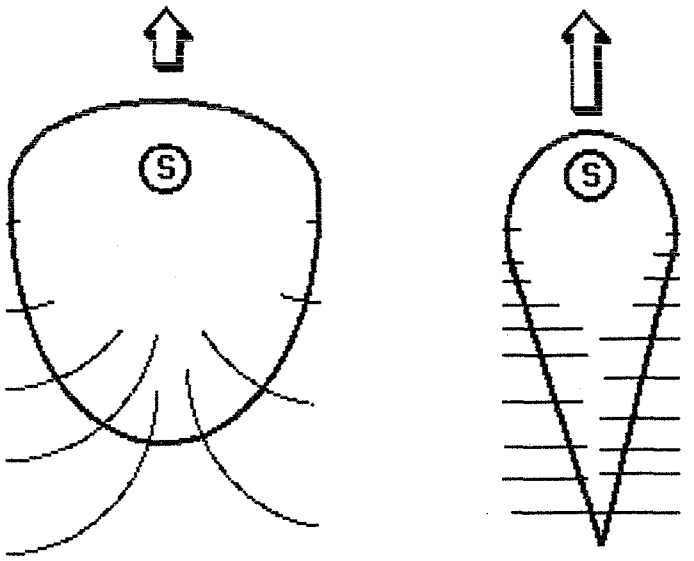


Figure 7. The effect of welding speed on shape of the fusion zone and weld metal microstructure. The relative lengths of the arrows represent welding speed.

At the other extreme of welding speed the final fusion zone structure will be too coarse due to the slow cooling conditions supported by the well heated base plate. Coarse microstructure is associated with lower mechanical properties. In a previous research it was shown that an increase in interdendritic particle size from 10 to 150 μm correlates with a 10 ksi decrease in weld tensile strength (11). The effect of welding speed in electron beam welding is dramatically illustrated in a brief review note by Komizo, et al. (12). These same effects are also characteristic of any moving source fusion welding process, such as VPPA welding.

Weld metal morphology, which is primarily dendritic, is controlled by the solidification process. Grains have columnar shapes directed along the line of heat transfer. For a given materials system welding speed has the greatest influence on the grain shapes and directional pattern in the fusion zone as well as the size of the structural features. The secondary dendrite branch dimension is the critical size parameter in the fusion zone microstructure. The product of welding speed and the square of the size parameter is a constant for a given materials system and solidification process.

Stresses are developed by differential cooling from a

condition where pronounced temperature differences exist within the system. Under these conditons any possible stresses are released due to the enhanced plasticity at the elevated temperature. As the hot metal cools, its strength increases and its contraction, relative to the condition of the initially cool metal causes locked-in stresses to develop. The magnitude and direction of these stresses is influenced to a lesser degree by variations in solidification conditions.

Solidification:

Solidification is the most significant process in weld metal structure and properties. The basic weld characteristics that are directly affected include:

1. chemical segregation
2. second phase production
3. porosity

It is not possible to remove solidification factors from welding process factors since they are totally interdependent. They must be considered together in any given application. However, selected topics in the physical metallurgy of weld metal solidification will now be considered.

Chemical segregation is a dominant factor and can be readily explained in relation to the aluminum-copper phase diagram, which applies to the 2219 alloy. The nominal copper composition is 6.3% by weight. Reference to the diagram shows that the composition line is to the right of the maximum solid solubility point of 5.62% copper (548 °C). Therefore it can be expected that there will be free particles of $CuAl_2$ intermetallic compound in the microstructure, regardless of the treatment after solidification. Under steady state welding, the weld pool composition corresponds to this same 6.3% Cu value.

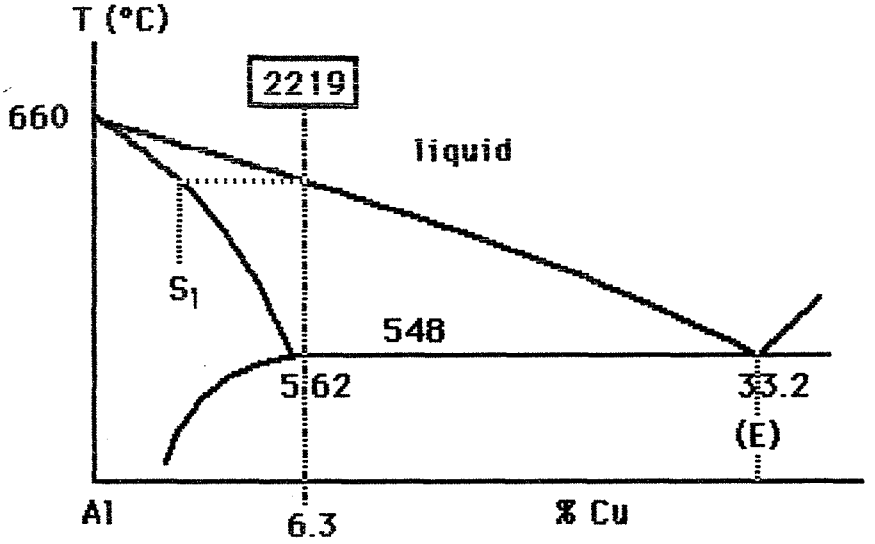


Figure 8. Aluminum rich end of the Al-Cu phase diagram. The composition of 2219 aluminum alloy and the solid that forms initially are shown. The eutectic constituent, (E), is a coherent mixture of aluminum solid solution and $CuAl_2$ intermetallic compound.

The composition of the first solid to form on cooling is S_1 (marked on the diagram) and is formed at a temperature of 640 °C. As solidification continues, the copper, that is not incorporated in the first solid, enriches the liquid so that it freezes out at a lower temperature, corresponding to a further position along the solidus line. The last solid to form is an eutectic mixture, with a composition of 33.2% copper. It freezes at 548 °C. It is this liquid that contains the highest copper content and is most likely to flow against the coolest surfaces of the cooling weld. It has the greatest capacity to fill any cracks or other openings in the weld metal.

The solid-liquid boundary surface of the weld pool is not isothermal but varies from a temperature corresponding to the freezing temperature of S_1 at the positions where solidification commences and drops to the eutectic solidification temperature at positions directly behind the moving source.

Second phase production is an important aspect of solidification. In this alloy system the second phase can be formed in two constituents. The intermetallic compound phase (53% Cu) is formed in either the fusion zone or heat affected zone and is in the parent metal

microstructure. The second constituent, the eutectic mixture is only in the fusion zone - a product of solidification. It is this constituent that is important in some forms of enigma and this is a major controlling factor on the weld tensile strength. It does not influence the yield strength very strongly since this property is dominated by the solid solution that comprises most of the microstructure.

Bayless (13) produced weld structures to demonstrate the light line enigma. His method involved the intentional misalignment of the torch for the root pass, thus producing a deep undercut to one side of the weld and an exaggerated buildup of metal to the other side. The normal and misaligned final profiles are shown in Figure 9.

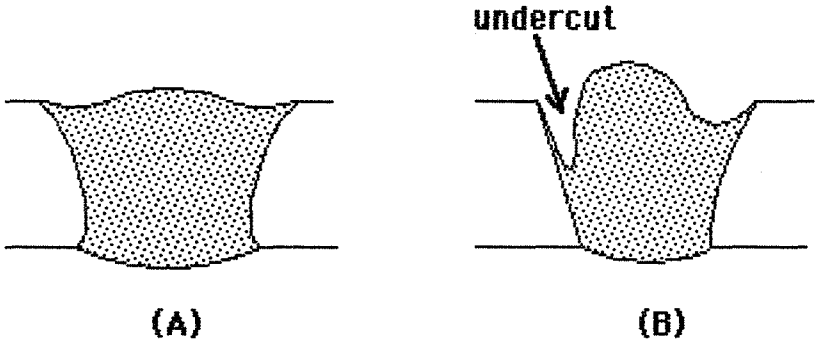


Figure 9. The shapes of the desired (left) and misaligned (right) shapes of the root pass weld metal.

When the cover pass was applied, the final structure showed the enigma due to the concentration of eutectic mixture in the undercut. The cover pass only partly repaired the initial improper shape. The deepest part of the undercut, which was not remelted was filled with eutectic liquid during welding. This left a residual copper-rich band of the form described above in the enigma simulations. The first three microdensitometer traces were made on the radiographs of the Bayless simulation samples. Lancaster discusses other methods to cause these effects (14).

Porosity is a type of second phase that results from solidification. There are several causes including:

1. The release of gases dissolved in the liquid.

2. The entrapment and release of hydrogen that is

produced by the decomposition of water or hydrocarbons left on the surface of the metal or filler wire prior to welding.

3. The failure to fill the interdendritic regions due to insufficient eutectic liquid.

4. The freezing over of interdendritic regions, thus preventing access by the available eutectic liquid.

The first two sources produce a more uniform distribution of porosity throughout the new solid but the latter two develop concentrated openings in the vicinity of the center line. None of these are typical of VPPA welding of 2219-T87 aluminum alloy. See Easterling (15) for more information about porosity in welding.

Radiography:

The initial x-ray beam is directed normal to the principal plane of the sample and is uniform in intensity. This uniformity is achieved by employing a long distance between source and film.

Passage through the sample reduces the intensity of the entire spectrum in proportion to the sample thickness and its absorption characteristics, which are a function of wavelength. The beam is further attenuated as it passes through the film. The energy lost in the sample produces the pattern seen in the film after it is developed. The radiation absorbed in the film causes the darkening. The thinner the sample, the more radiation penetrates to reach the film, and the more particles of silver that are developed in the emulsion to darken the film.

These features are represented in Figure 10. The thickness of the lines representing the x-ray beam, that is moving downwards, indicate the relative intensities. The points in the film cross section represent particles of silver, after develment. The effect of the reinforcement is to increase the absorption and reduce the final silver content in the corresponding parts of the film.

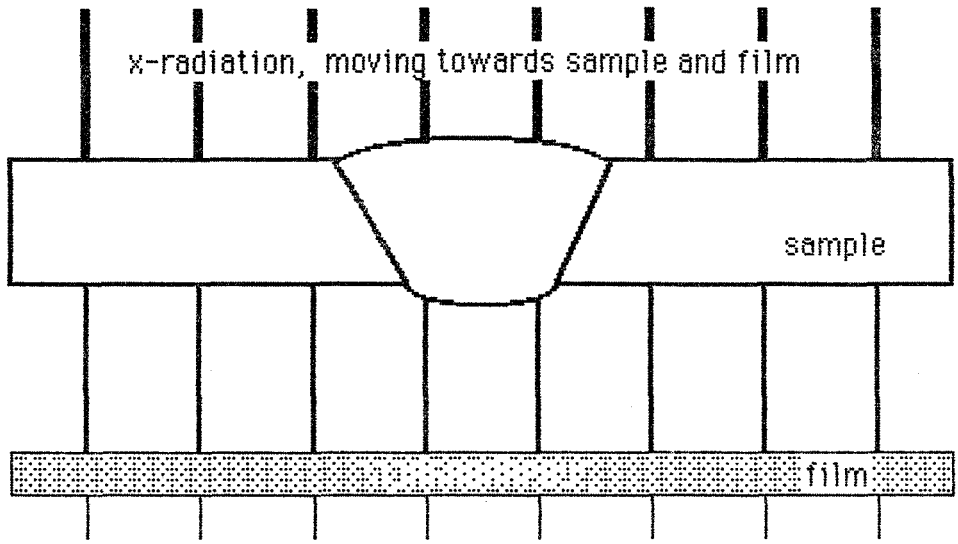


Figure 10. Representation of the general configuration for the preparation of an x-ray weld radiograph. The distance between the sample and film is exaggerated in this figure.

The physical processes involved are outlined in figures 11 and 12. See Cullitty for a comprehensive presentation (16). The spectrum of the initial beam is shown in the first block at the left of figure 11. The spectrum of the final beam is shown at the right. Under the typical 80 Kv tube operating conditions, spectral content of the initial radiation is primarily a continuous distribution of x-rays, generally identified as *continuous, white* or *brehmsstrahlung*. A characteristic of tungsten and other high atomic number target material tubes is the suppression of characteristic radiation peaks. There probably is a small component of characteristic K_{α} or K_{β} radiation. Such a peak is shown schematically. There is only a small energy content in these characteristic radiation components.

After passage through the sample the intensity of radiation is generally diminished, as shown in the block diagram to the right. This attentuation is most pronounced in the vicinity of the absorption edge of the important chemical components in the sample. The relationship between thickness, materials characteristics and initial and reduced intensities is written below the sample.

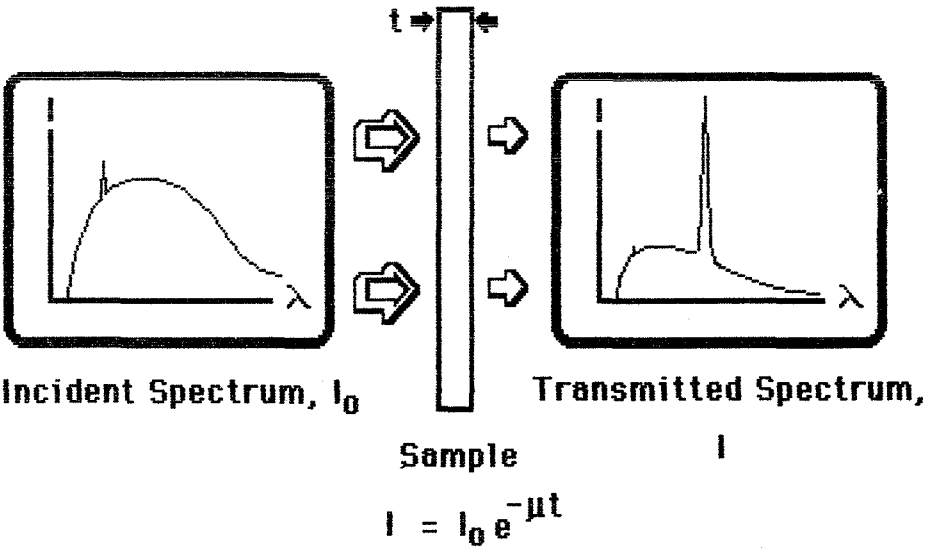


Figure 11. The spectrums of the initial and transmitted x-ray beams in radiography.

Each element has characteristic absorption characteristics in the form shown in Figure 12. Energy is absorbed when specific electronic processes are induced within the elements of the sample. The energy invested in these processes can be recovered in different ways. Copper, for example, has an absorption edge at 1.38059 Å and uses this energy to produce fluorescent radiation of 1.54 Å wavelength. In this case, the characteristic radiation is added to the spectrum. These processes are most active when there is energy at the lower wavelength side of the absorption edge. In this case, the sample acts as a source of *secondary radiation*. This usually has the undesireable effect of increasing the general background density of the film and tending to obscure the features of the image.

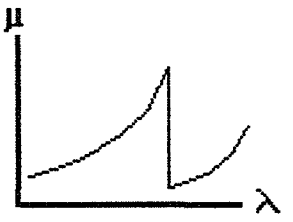


Figure 12. Typical absorption curve.

Diffraction is a possible contributing factor because diffraction can cause extra radiant energy to be concentrated along specific directions. Rummell referred to one specific diffraction effect. Another, that has been mentioned, and should be considered is the development of Kossell lines, which are a phenomenon in the class of wide angle scattering (17-19). If the linear anomaly in the radiograph is actually due to a diffraction effect, the features will be characteristic of that effect and can be recognized by appearance.

In one application, the Kossell pattern is developed by placing a thin layer of fluorescent material, such as copper, against the side of the single crystal sample that is towards the source. When the narrow beam of x-rays strikes the sample this layer then provides an x-ray source at that close location to produce divergent diffracted beams. These rays are contained within a conical surface, intersecting the film in arcs. This type of enigma structure has not been reported.

Anomalous transmission, known as the Borrmann effect (20-23), can also make a contribution to the intensity of x-rays delivered to alloying elements concentrated close to the film within the weld. The necessary conditions are a large, favorably oriented grain of a single crystal phase. Such large grains have been observed in weld metal. They are formed due to the maintainance of conditions favoring stable growth, as opposed to nucleation. These conditions include a favorable alignment of a high symmetry crystal direction with the vertical, along which the x-ray beam propagates. In aluminum, as is typical of most alloy systems, the favored growth direction is <100> which has the highest symmetry. There is a greater likelihood that for a grain that grows in a <100> direction parallel to the weld surface that it will also have another <100> type direction aligned vertically. Such an alignment would be naturally favored and would tend to prevent the growth of grains of other alignments. The formation of large, highly aligned grains should be no surprise in a solidification process as controlled as that in a VPPA fusion zone.

The natural conclusion is that a number of different factors contribute to the variation of radiographic film density across the weld. A number of these are summarized in Table 1 with respect to the direction of the darkening effect of each.

Table 1
ABSORPTION

FACTOR	DARKENING
Shape and Thickness	**↓**
Macro Openings	**↑**
Chemical Accumulation	**↓ (Cu in 2219 Al)**
Porosity	**↑**
Fluorescence	**↑**
Diffraction	**↑ (possibly with ↓)**
Anomalous Transmission	**↑**

It was found that persons can unambiguously distinguish adjacent light and dark areas when the density difference is only 1% (e.g. 2.98 and 3.01). The significance is that many features are observed that are only of secondary significance.

Results and Discussion

The above factors can produce a complicated visual pattern when combined. The following are microdensitometer traces made across welded panels of 2219-T87 aluminum plate. Figure 13 is a trace across the VPPA root pass made under conditions representing one set of torch alignment conditions marking the limit of acceptable practice. These traces are plots of data taken at a spacing of 125 µm along a line across the weld radiograph. There are 204 points per inch. The sample surface conforms to the desired shape, but the effect of the misalignment is visually obvious as is the lack of symmetry in the trace about the weld centerline. There is a minor degree of undercut along both edges of the weld (see 1 and 2). The undercut at 2 is more pronounced and wider. The peak at 3 corresponds to a

surface marking that runs along the side of the root reinforcement.

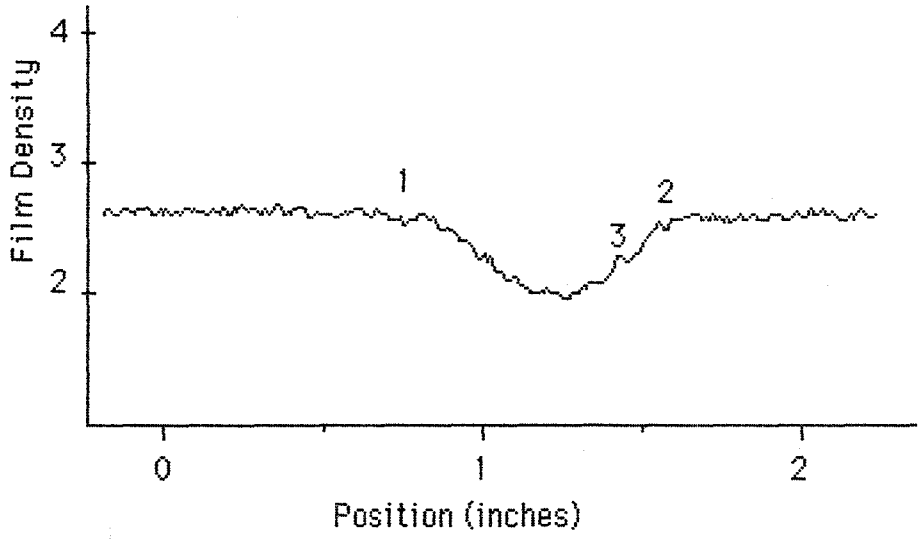


Figure 13. Microdensitometer trace across radiograph of nearly symmetric root pass in a VPPA weld.

Figure 14 is a similar trace across another unfinished weld (root pass only) made in the same plate. However, the torch misalignment

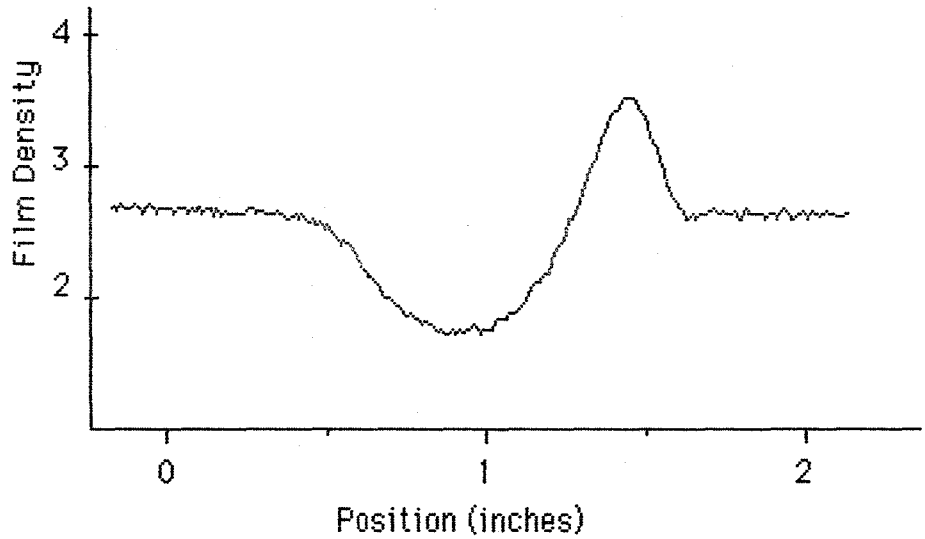


Figure 14. Microdensitometer trace across radiograph of root pass made under conditions of severe torch misalignment.

was intentionally set to produce an extreme undercut along one edge. This is represented by the pronounced peak at the right. The deep valley at the left represents the region where the weld metal accumulated. This trace provides a good representation of the weld profile.

Figure 15 is across another weld made in the same plate, under the same conditions of misalignment, but with the cover pass applied. Even here the misalignment is detectable in the microdensitometer profile, although it is not visually detectable on the finished weld. Features 4 and 5 correspond with a weld root marking (see 1 in the first trace) and a surface scratch on the underside of the plate.

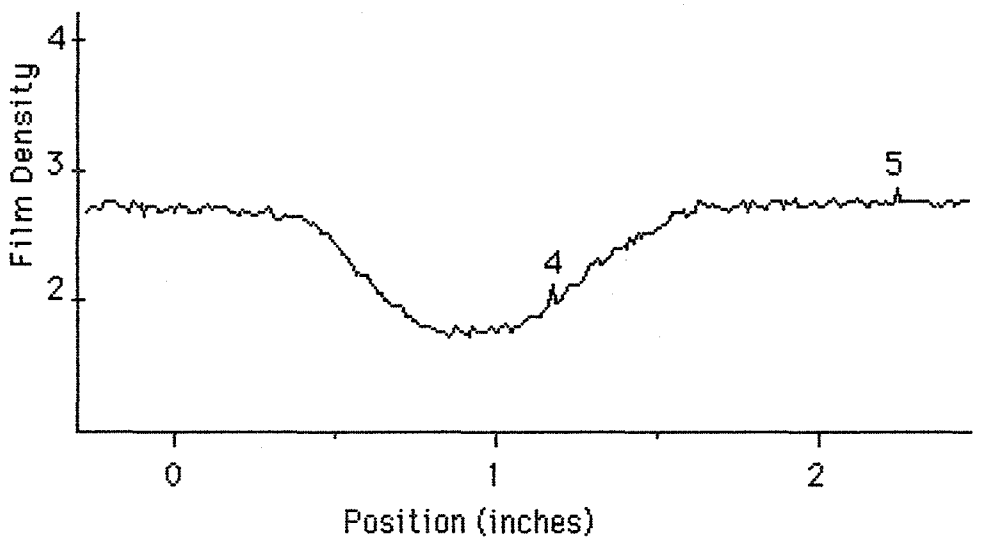


Figure 15. Microdensitometer trace of weld radiograph with cover pass over a misaligned root pass.

Figure 16 is a trace across a production radiograph showing two light line enigmas (marked 6 and 7). The appearance of the radiograph matches the appearance of the radiograph of Figure 15. The sample is not available.

Figure 16. Microdensitometer trace across production weld radiograph.

There is considerable detail, or *fine structure*, in the trace. These features can be related to the factors listed in Table 1.

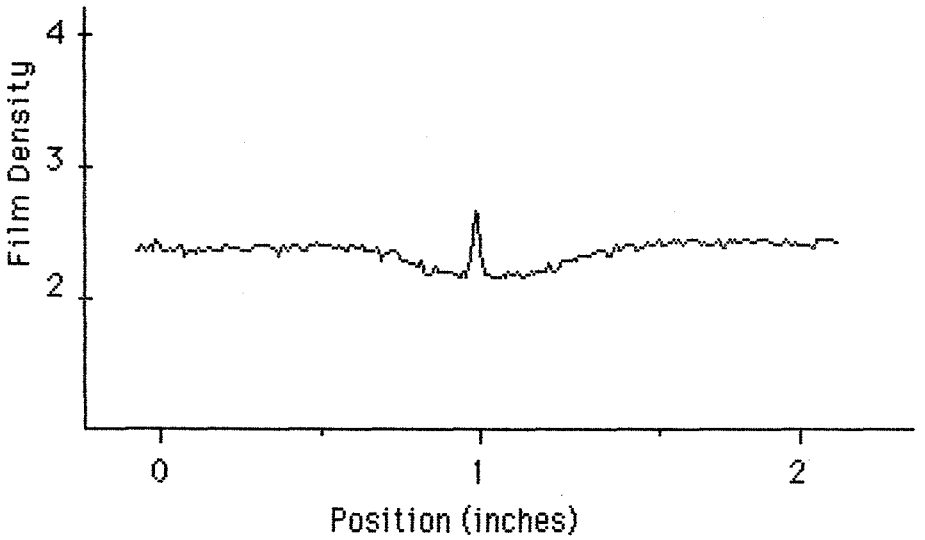


Figure 17. Microdensitometer trace across center line crack. TIG welded 2219 alloy.

Figure 17 displays a trace made across the radiograph of a

TIG welded 2219-T87 plate with a definite centerline end crack. The appearance of the crack feature in the radiograph is very similar to the dark line enigma with the difference that the darkening associated with the real crack is much stronger in intensity.

Table 2 summarizes the density differences of the features noted in the microdensitometer traces.

TABLE 2
THE EFFECT OF ENIGMAS AND CRACKS
ON RADIOGRAPHIC FILM DENSITY

SAMPLE AND FEATURE	TYPE	DENSITY INCREMENT
Fig. 13, feature 1	light	0.09
Fig. 13, feature 2	light	0.05
Fig. 13, feature 3	dark	0.05
Fig. 14, total	shape	1.81
Fig. 15, feature 4	dark	0.18
Fig. 15, feature 5	dark	0.10
Fig. 16, feature 6	light	0.13
Fig. 16, feature 7	dark	0.18
Fig. 17, crack	defect	0.50

A series of measurements were made to evaluate the possible contribution of fluorescence. The film was placed over a composite background, with copper to one side and lead the other. The density of the resulting film was measured at five points within each field and averaged. The results are plotted in Figure 18 as darkening ratio versus tube voltage. At low voltages the copper produces a greater darkening, but this changes pronouncedly at a voltage between 80 and 120 Kv. The L absorption edge for lead is 0.95073 Å, which indicates a critical wavelength just longer than 130 Kv. The nearest edge for copper is activated most effectively at tube voltages of 8900 v and slightly higher.

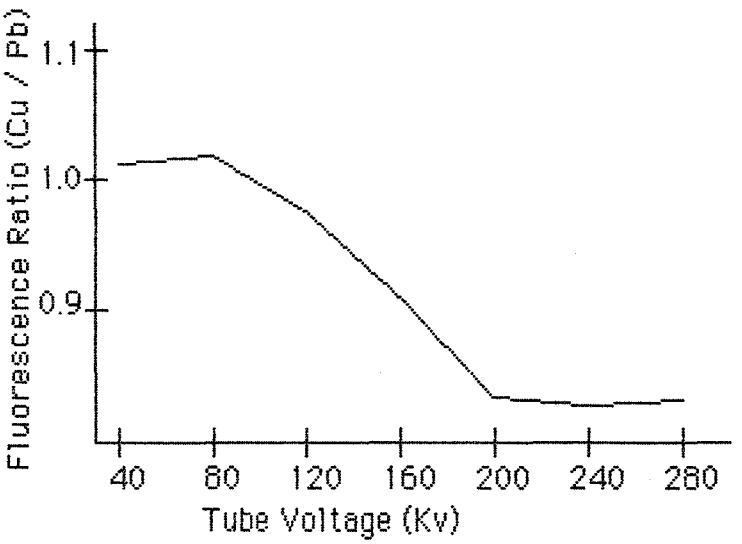


Figure 18. Ratio of backscatter over Cu to backscatter over Pb as a function of tube voltage.

Along with these measurements, a radiograph was made of a large grained pure aluminum sample. This showed a general pattern of detail corresponding to the grain structure, although none of the grains is separately identifiable. This is evidence of the possible contribution of diffraction or anomalous transmission.

Enigma Simulation:

It is possible to simulate enigma formation. This is done by calculating the transmission ratio, I/I_0, through the plate at each position across the weld. The basis of the calculation is Lambert's Law, $I = I_0 e^{-\mu x}$. In these equations, I is the intensity of the transmitted radiation and I_0 its initial intensity on the front side of the sample. e is the base of the Naperian logarithms, μ the absorption factor and x the distance of transmission through the plate. Since human sensitivity relates to ln|, this response is directly represented by the magnitude of the exponential quantity, μx. Absorption is a characteristic of each chemical species and a function of wavelength. The values are generally available, usually listed as the mass absorption coefficient, μ/ρ, where ρ is the density. Thus response is measured by the product, $(\mu/\rho)\rho x$. In any mixture, the net mass

absorption coefficient is the linear average of the values for the elements, in proportion to their weight fractions in the mixture.

Figures 19-21 are density traces generated in this way through the sample. In each figure, the sample is shown above and the film density representation directly below. It can be seen how the pattern of complexity in the radiographic film develops with the inclusion of complexity in sample shape and chemical segregation. This illustrates the origin of the enigma as well as the feasibility of simulating and interpreting these effects.

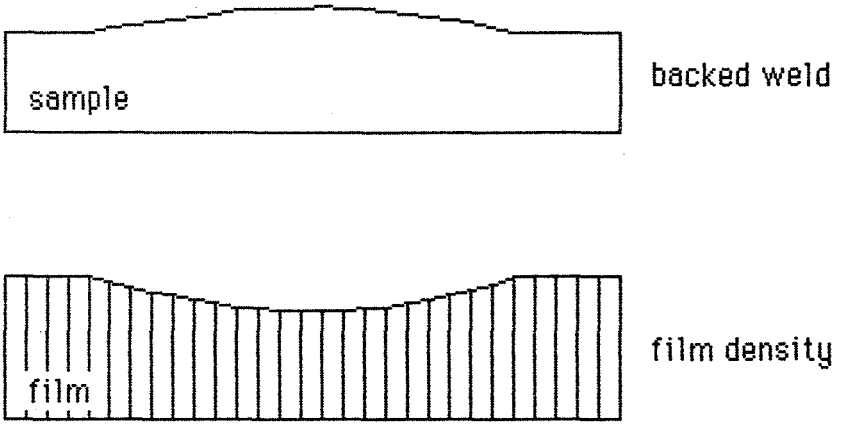


Figure 19. Computed film density for the sample shape shown.

Note that in this plot the density of the radiograph does not vary across the ends and has a curvature that matches the sample curvature. The sample shape corresponds closely to that of an actual weld, but the outline was drawn freehand on graph paper. The dimensions were read directly from this drawing and entered into the computer for the calculation. The result of the calculation was plotted and processed in a similar manner to that used for Figures 13-17. The variation in film density represented by the thickness of the cross hatched section is exaggerated several times.

The addition of the root reinforcement adds a bit of detail to the film density profile. The width of each reinforcement is clearly displayed in Figure 20.

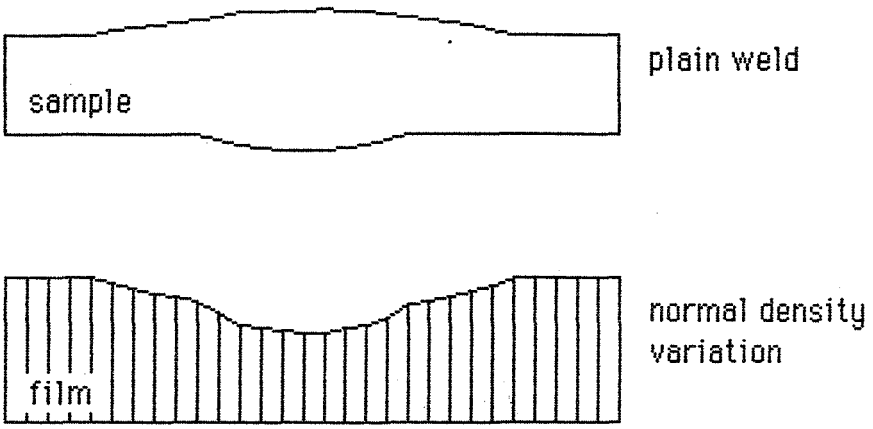


Figure 20. Film density variation across weld with both crown and root reinforcements.

Figure 21 was generated from the same weld configuration with an arbitrary addition of eutectic mixture. The angle of the deposit cross section, shown in the weld cross section, its thickness (0.5 mm) and the effective mass absorption coefficient, are reasonably close to experimental values, but are used primarily for illustration. The result is exaggerated.

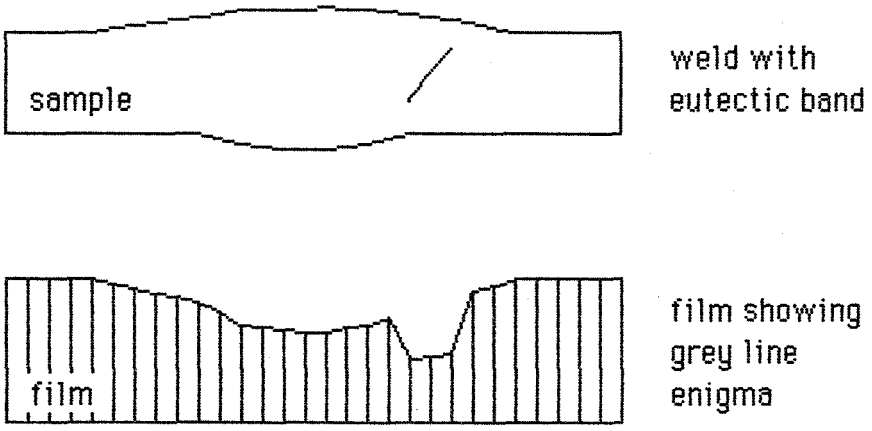


Figure 21. Computer simulated radiograph density across a weld with both reinforcements and a longitudinal band of eutectic mixture.

The gray line enigma of Figure 21 corresponds to that demonstrated by Bayless

Conclusions

The enigma is fine structure in the x-ray radiograph representing a real condition in the weld metal. There are many types of weld radiograph enigmas. The microdensitometer traces show that some features of radiograph fine structure can be related to visible features on the weld. The correspondence between radiograph fine structure and weld metal condition awaits direct verification, although the present indications are promising. The microdensitometer trace method can be used to reconcile these features of weld radiographs. Initial measurements indicate that the enigma represents a density difference of the order of 0.1 and that the crack defect represents a difference of the order of 0.5. It is suggested that this method be developed for weld evaluation and possible in-process control.

Weld radiograph fine structure is generally aligned in the welding direction, hence when it is sufficiently pronounced it is classed as a linear anomaly, or enigma. Changing conditions along the weld can produce a general curvature, which can explain curved enigma features. The full effect on properties is not known. The enigma cannot be eliminated and is not a feature of VPPA welding alone - it is a general phenomenon.

Bibliography

1. P. C. Duren and E. R. Risch, "Radiographic Interpretation Guide for Aluminum Alloy Welds," *NASA Technical Memorandum NASA TM X-53939*, (May 19, 1970), page 8.

2. L. Johnston, "VPPA 'black line' Anomaly Indication on X-Ray Films," Memo for the Record, August 21, 1984.

3. Kinchen and Brown, "Evaluation of VPPA Radiographic Anomalies in 2219 Aluminun Weldments," Laboratory Test Report, Metallurgical Analysis, 85A205, (6-3-85).

4. W. D. Rummell and B. E. Gregory, "'Ghost Lack of Fusion' in Aluminum Alloy Butt Fusion Welds," *Materials Evaluation, XXIII*, (1965), 586.

5. E. Hirosawa, M. Naoe and T. Fukui, "'Ghost Defects' In Radiographs of

Aluminum Alloy Welds," *Materials Evaluation, XXIX*, (1971), 99.

6. M. S. Tucker and P. A. Larssen, "Markings in Radiographs of 2014 Aluminum Alloy Gas Tungsten-Arc Welds," *Welding Journal, 47*, (1968), 223.

7. D. M. Rabkin, L. A. Bukalo, V. Ta. Kirzhova and A. S. Deniyanchuck, "Weld Heterogeneity in Aluminium-Magnesium Alloys," *Avt. Svarka*, No. 5, (1966), 74.

8. S. Issiki, J. Ko, K. Kataoka and T. Yamazawa, "The abnormal Pattern in Radiographs of Aluminum Alloy Castings," *Hihakai-Kensa, 15*, (1966), 257.

9. M. Irie, M. Fujii and S. Yamashita, "The Influence of Structure on the X-ray Radiograph of Stainless Steel Welds," *Hihakai-Kensa, 12*, (1963), 263.

10. K. Masubuchi, "Residual Stresses and Distortion, Chapter 6," *Welding Handbook, 1*, 7th Edition, American Welding Society, (1981), 223.

11. W. A. Jemian, "The Strength and Characteristics of VPPA Welded 2219-T87 Aluminum Alloy," Final Report, NASA-ASEE Summer Research, (1984).

12. Y. Komizo, C. S. Punshon, T. G. Gooch and P. J. Blakely, "Effect of process parameters on centre line solidification structures of electron beam welds in steel," Welding Institute Members Report No. 252, (1984).

13. E. O. Bayless, personal communication.

14. J. F. Lancaster, *The Physics of Welding*, Pergamon, (1984).

15. K. Easterling, *Introduction to the Physical Metallurgy of Welding*, Butterworths, (1983).

16. B. D. Cullitty, *Elements of X-Ray Diffraction*, Addison-Wesley, (1978).

17. A. Taylor, *X-ray Metallography*, Wiley, (1961), 770.

18. L. S. Birks, *Electron Probe Microanalysis*, 2nd Edition, (1971), 140.

19. H. Yakowitz, "The Divergent Beam X-Ray Technique," *Electron Probe Microanalysis*, Adademic Press, (1969), 361.

20. G. Borrmann, "Röntgenwellenfelder," *Trends in Atomic Physics*, Risch, Paneth, Roosbaud, Editors, Intersciecnce, (1959), 262.

21. L. P. Hunter, "X-Ray Measurement of Microstrains in Germanium Single Crystals," *J. Appl. Phys.*, *30*, No. 6, (1959), 874.

22. B. W. Batterman, "Imaginary Part of X-Ray Scattering Facotr for Germanium. Comp[arison of Theory and Experiment," *J. Appl. Phys.*, *30*, No. 5, (1959), 998.

23. A. Guinier, "Small Angle X-Ray Conference Proceedings," *J. Appl. Phys.*, *30*, No. 5, (1959), 601.

1985

NASA/ASEE SUMMER FACULTY RESEARCH FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

ELECTROCHEMICAL CORROSION STUDIES

Prepared By:	Ward W. Knockemus, Ph.D.
Academic Rank:	Professor
University and Department:	Huntingdon College Department of Chemistry
NASA/MSFC: Division: Branch:	 Metallic Materials Corrosion Research
MSFC Counterpart:	Dr. M. D. Danford
Date:	August 2, 1985
Contract No.:	NGT 01-008-021 #12 The University of Alabama in Huntsville

ELECTROCHEMICAL CORROSION STUDIES

BY

Ward W. Knockemus
Professor of Chemistry
Huntingdon College
Montgomery, Alabama

ABSTRACT

Objectives of these investigations are to gain familiarity with the Model 350 Corrosion Measurement Console, to determine if metal protection by grease coatings can be measured by the polarization-resistance method, and to compare corrosion rates of 4130 steel coated with various greases. Results show that grease protection of steel may be determined electrochemically. Studies were also conducted to determine the effectiveness of certain corrosion inhibitors on aluminum and steel.

ACKNOWLEDGEMENTS

I wish to express my deep appreciation to NASA and ASEE for the opportunity to do summer research for NASA. Since my regular position is in a teaching institution where little research is conducted, it was a treat to do research again.

Special thanks to Dr. M. D. Danford my NASA counterpart in the Corrosion Research Branch for his patient instruction on the Model 350 Corrosion Measurement Console. His considerable knowledge of computer processing of polarization-resistance data was especially helpful to the objectives of the research.

The efforts of Dr. Gerald R. Karr, the 1985 Summer Faculty Fellowship Program Director, made the program even more valuable. Thanks to him and Dr. James Dozier along with Mr. Leroy Osborn for their tremendous contributions to an outstanding program.

INTRODUCTION

General Theory of Electrochemical Corrosion and Corrosion Rate Measurement

Pure and impure, homogeneous or heterogeneous metallic materials corrode to lower the Gibb's free energy of the system. Corrosion returns the metal to its natural oxidation state, usually an oxide or hydrated oxide. Free, "active" metals (those with negative half-cell potentials in a Table of Standard Reduction Potentials) are in a metastable state; they wait only for an oxidant, and sometimes an electrolyte solution to return them to a stable oxidized form.

The electrochemical theory of corrosion rests on the assumption that any metal surface can have areas where metal atoms are oxidizable and other areas where electron acceptor atoms or groups can be reduced. Oxidation (metal dissolution or corrosion) areas are named anodic sites and areas on the metal surface where reduction can occur are named cathodic sites. Figure 1 illustrates this.

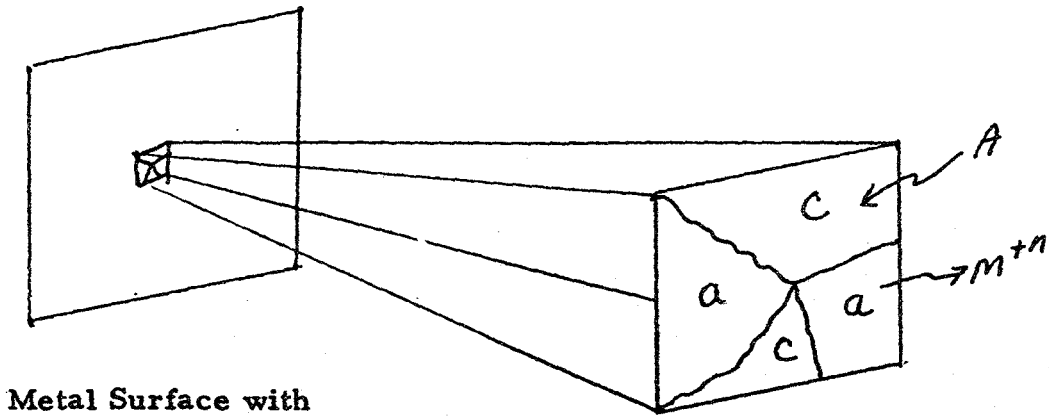
The general reactions:
at the anodic sites: $M \longrightarrow M^{n+} + n\ e^-$

at the cathodic sites: $A + n\ e^- \longrightarrow$ Product species

For iron or steel on the atomic level the reactions in an oxygen-water environment might be as shown by Figure 2.

Although there is only one anodic reaction (metal atom oxidation and dissolution), the several possible reactions that might occur on the cathode sites are the following:

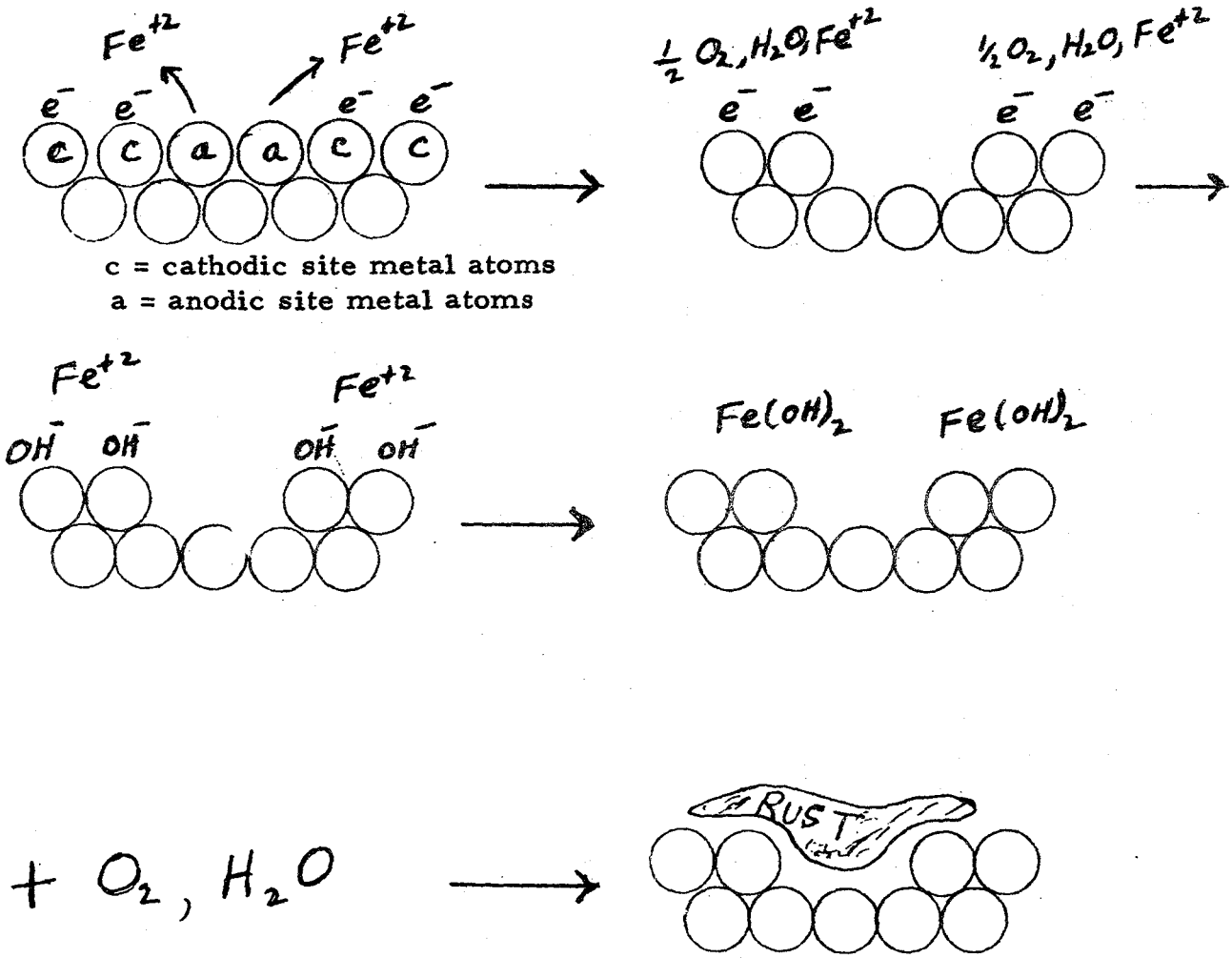
1. Hydrogen ion reduction: $2\ H^+ + 2\ e^- = H_2$ (pH less than 5)
2. Oxygen reduction: $O_2 + 4\ H^+ + 4\ e^- = 2\ H_2O$ (low pH)
3. Oxygen reduction: $O_2 + 2\ H_2O + 4\ e^- = 4\ OH^-$ (high pH)
4. Water reduction: $2\ H_2O + 2\ e^- = H_2 + 2\ OH^-$ (neutral, high pH)
5. Metal ion reduction: $Cu^{++} + 2\ e^- = Cu$ (typical)
   $Fe^{+3} + e^- = Fe^{++}$
6. Oxy-anion reduction: $NO_3^- + e^- + 3\ H^+ = HNO_2 + H_2O$ (low pH)



Metal Surface with Local Corrosion

a = anodic site
c = cathodic site
A = reducible species migrating to cathodic site
M^{+n} = metal ion departing anodic site

Figure 1 Anodic and Cathodic Sites on Metal Surface



c = cathodic site metal atoms
a = anodic site metal atoms

RUST = $Fe_2O_3 \cdot x\,H_2O$, $Fe_3O_4 \cdot y\,H_2O$, FeOOH, etc.

Figure 2 A Model of the Rusting Mechanism

The cathode reaction occurring is the one that produces the highest cathodic current (or electron flow to the cathodic sites) for the given conditions. Usually in the presence of air a solution above pH 5 contains sufficient dissolved oxygen to support reaction (2) or (3) above. In the vicinity of pH 7, hydrogen ion from water reduces at -0.41 volts while O_2 has a reduction potential of -0.83 volts (neglecting overvoltage). Therefore, the oxygen is reduced producing OH^- ions. (The reduction of water also forms OH^- ions.) Below pH 5 the cathode reaction is mainly hydrogen evolution.

Electrochemical Theory

For a spontaneously corroding metal surface not connected to an external electrical device, the anodic current, i_a, must equal the cathodic current, i_c. Since the anodic current is from the dissolution of the metal atoms it is also called the corrosion current, Icorr. The corrosion current sometimes is referred to as the exchange current, i_{ex}.

The corrosion current is an important quantity because it yields the actual rate at which the metal dissolves or forms corrosion product. The following relationships show the uses of Icorr: A corrosion current of 1 amp corrodes about 22 pounds of iron in one year; an actual corrosion rate calculated from data for aluminum corroding in corrosive water in the presence of Mobay inhibitor with an Icorr of 0.00785 microamps yields a corrosion rate of 0.00345 mils per year. Equations 6, 7, and 8 show how Icorr can give a corrosion rate.

6. $\frac{\text{Icorr (amps)}}{\text{F (96500 coulombs/equiv.)}}$ = gram equivalents of metal corroding per second

7. Weight (grams/second) of corroding metal = $\frac{\text{Icorr (amps)}}{\text{F (96500 C)}}$ X gram equivalent weight of metal corroding

The commonly used equation that gives mils per year (mpy) of metal corroding is as follows:

8. Corrosion Rate (mpy) = $0.1288 \times \text{Icorr} \times \frac{\text{gram equivalent weight of metal}}{\text{density of metal}}$

--Icorr in microamps/cm^2
--gram equivalent weight of metal in grams
--density of metal in grams/cm^3

The electrical potential of any corroding metal surface (without any external connections) is the potential at which the corrosion current is produced. It is called the corrosion potential, Ecorr. Ecorr always lies between the equilibrium potential for the metal/solution half-cell and the equilibrium potential for the electron acceptor/solution half-cell. (These equilibrium potentials are not the Standard Reduction or Oxidation Potentials commonly found in references and textbooks. They depend on the metal, its surface condition, and the solution contacting the metal.) Unlike Icorr, Ecorr can be directly measured and can provide limited information about the corrosion system. Generally relatively positive Ecorr's (-0.2 to -0.4 volts) indicate low rates of corrosion while more cathodic values of Ecorr (-.6 to 1.0 volts) indicate higher corrosion rates for the metal surface. In an oft-quoated paper, Wormwell and Brasher' reported for painted steel that Ecorr initially dropped, then rose, to a maximum before finally dropping to permanent low values after fifty days (Ref. 1).

A more negative or cathodic potential applied externally accelerates the cathodic reaction. Likewise causing the potential to be more positive than Ecorr increases the anodic current. Figure 3a shows the current-over-potential relation for the general metal dissolution equilibrium. As shown on the left-hand side of Figure 3b (left of the zero current ordinate), an increasingly negative voltage departing from the equilibrium value for the cathodic site, produces an increase in the cathodic current.

If a metal/solution half-cell or electron-acceptor/solution half-cell are at some potential other than their equilibrium potential, they are said to be polarized. Therefore the two half-cells (one the cathodic site and the other the anodic site) are always polarized during the corrosion process. Since the cathodic and anodic sites are probably close on the metal surface, they are said to be at the same potential which is called the corrosion potential, Ecorr.

For a short-circuited non-connected corrosion cell

$$i_c = i_a = \text{Icorr}$$

and net current $= |i_c = i_a| = \Delta i$ and at Ecorr, $\Delta i = 0$.

If an external potential is applied to the metal surface to make it more positive or negative than the Ecorr of the corrosion, the polarization of the cell is altered and a net anodic or cathodic current can be observed. Overpotentials more positive than Ecorr increase the anodic current, while those less positive increase the cathodic current.

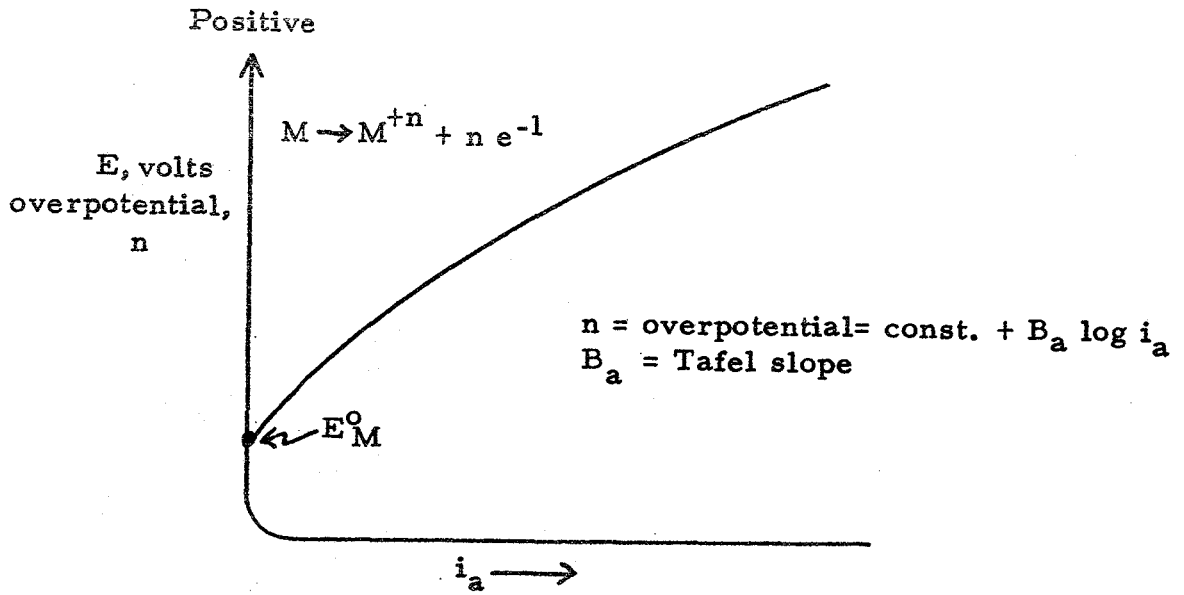


Figure 3a. The Anodic Current-Overpotential Diagram

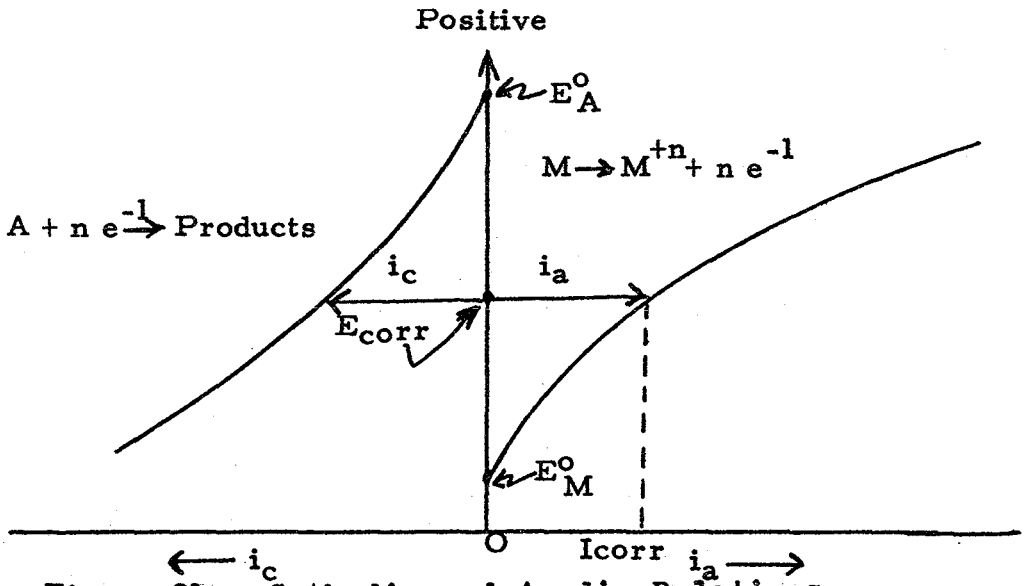


Figure 3b. Cathodic and Anodic Relations

Summarizing, if $E \neq Ecorr$, $\Delta i \neq 0$ (E = applied or imposed potential.)

for (E-Ecorr) >0, $i_a > i_c$ (anodic polarization)

for (E-Ecorr) <0, $i_c > i_a$ (cathodic polarization)

The difference between the imposed potential, E, and Ecorr is called overpotential, n, also labeled ΔE.

So, Eapplied - Ecorr = n = ΔE = overpotential

Development of the Resistance Polarization, Rp, Equation

The challenge of electrochemists studying corrosion is to obtain values for Icorr or io, the corrosion current density. Once obtained the values are easily converted to a corrosion rate by equation 8. About eighty years ago Tafel in studying hydrogen overvoltages on various metals formulated empirical equations (since then derived on theoretical bases) relating the overvoltage to current and exchange current. At Ecorr the exchange current density equals the corrosion current density, i_o.

Tafel's equation's:

$\Delta E = + B_c \log i_c/i_o$ (cathodic overpotential)

$\Delta E = - B_a \log ia/io$ (anodic overpotential)

B_a, B_c = Tafel slopes of ΔE vs. log i / io plots

$i_c/i_o = e^{2.3\ \Delta E/B_c}$, $i_a/i_o = e^{-2.3\ \Delta E/B_a}$

Δi = net current caused by overpotential, ΔE.

$$\Delta i = |i_c - i_a| = i_o\ (e^{2.3\Delta E/B_c} - e^{-2,3\Delta E/B_a})$$

Since $e^x = 1 + x + x^2/2! + \ldots$

$$\Delta i = i_o\ (1 + 2.3\Delta E/B_c - 1 + 2.3\Delta E/B_a)$$

$$= i_o\ (2.3\Delta E)\ (B_a + B_c\ /\ B_cB_a)$$

$$\text{So, } i_o = \frac{\Delta i}{2.3\Delta E} \left[\frac{B_c\ B_a}{B_c+B_a}\right]$$

This last equation is the Stern-Geary Equation and applies only when Activation Polarization exists. Concentration polarization must not be important. This is usually the case when the applied overpotentials and resultant currents are small enough to not affect species concentrations at the metal surface.

Dividing the overpotential by the net current gives the polarization resistance, R_p: $\Delta E/\Delta i = R_p$

Thus, $R_p = B_c B_a / 2.3 i_o (B_a + B_c)$. ($i_o$ = Icorr/cm^2.)

So, Rp provides a pathway to the important parameter, i_o. It is seen that R_p is the derivative of the potential at Ecorr with respect to the observed current. For this derivative or slope to be accurate, a straight line relationship at Ecorr should exist at small overpotentials which is usually the case. R_p is inversely proportional to Icorr and the corrosion rate which means the more polarized the metal surface is at Ecorr the lower the rate of metal dissolution and corrosion.

Danford and Higgins (Ref. 2) and Mansfeld (Ref. 3) point out that a computer program named POLCUR can carry out effective non-linear least squares evaluation of over-potential-current data and produce accurate values for R_p, the Tafel slopes, Icorr and Ecorr. In this research it was assumed that the POLCUR treatment of the data and the subsequent computer calculations produced valid and reliable results. Occasionally data would not compute; this occurred usually at high corrosion currents. Sometimes this problem was remedied by adjusting the overvoltage range scanned from ±20 mV to ±10 mV.

The instrument used in the research would determine R_p and calculate Icorr and corrosion rate. Tafel slopes were keyed in at assumed values of 100 mV/decade along with area, equivalent weight and density of the metal being studied. These instrument calculations were disregarded in nearly case in deference to the POLCUR results.

The measuring instrument was an EG&G Princeton Applied Research Model 350 Corrosion Measurement Console. This instrument and related information are described in Danford and Higgins paper (Ref. 2).

OBJECTIVES

The main objectives of the research were to determine if the Model 350 could be used to study the corrosion protection afforded by a thin layer of grease on a metal surface and if so, to evaluate and compare protection provided by two versions of Conoco grease on 4130 steel. If the electrochemical method proved successful in the initial investigations of grease-coated metal surfaces, then protection by a Texaco grease was to be determined.

Finally, an electrochemical study of the rate of corrosion inhibition of aluminum was to be investigated.

RESULTS AND DISCUSSION

1. Conoco grease studies. The purpose of this particular research was to compare the corrosion protection given by two Conoco greases used on metal joint surfaces of the SRB. The two greases differ slightly in their formulation; the "old" grease was used before January, 1983, and the "new" grease after this date. It was known that the "old" grease was effective and the question was can any difference in protective capability be detected electrochemically. The "new" grease is Conoco HD Calcium Grease No. 2.

To prepare the samples used in the corrosion measurement cell, 9/16 inch diameter by 1/16 inch thick 4230 steel discs were punched out, fine wet-sanded, degreased, and sprayed with a solution or suspension of the grease in trichloroethane. Sufficient spray passes deposited two mils of grease layer on the metal surface.

Following air drying three days, four discs were prepared, two with the "old" grease and two with the "new" grease formulation. They were placed in the corrosion measurement cell that contained 3.5% NaCl at pH 8.2. Alternate day measurements of Ecorr and ohmic resistance (Runc) were made plus the polarization-resistance scan if a measureable cathodic current was obtained. Only one sample of the "old" grease and one sample of the "new" grease provided a corrosion current. The other samples of each grease gave no corrosion current for the one month duration of the experiment. The corrosion rates calculated from these corrosion currents are shown in Figure 4.

Also graphed in this figure is the change of the ohmic resistance of the samples that occured with time. Of particular interest was the increase in resistance observed after about

CORROSION RATES AND RESISTANCES OF 4130 STEEL
COATED WITH 2 MIL GREASE

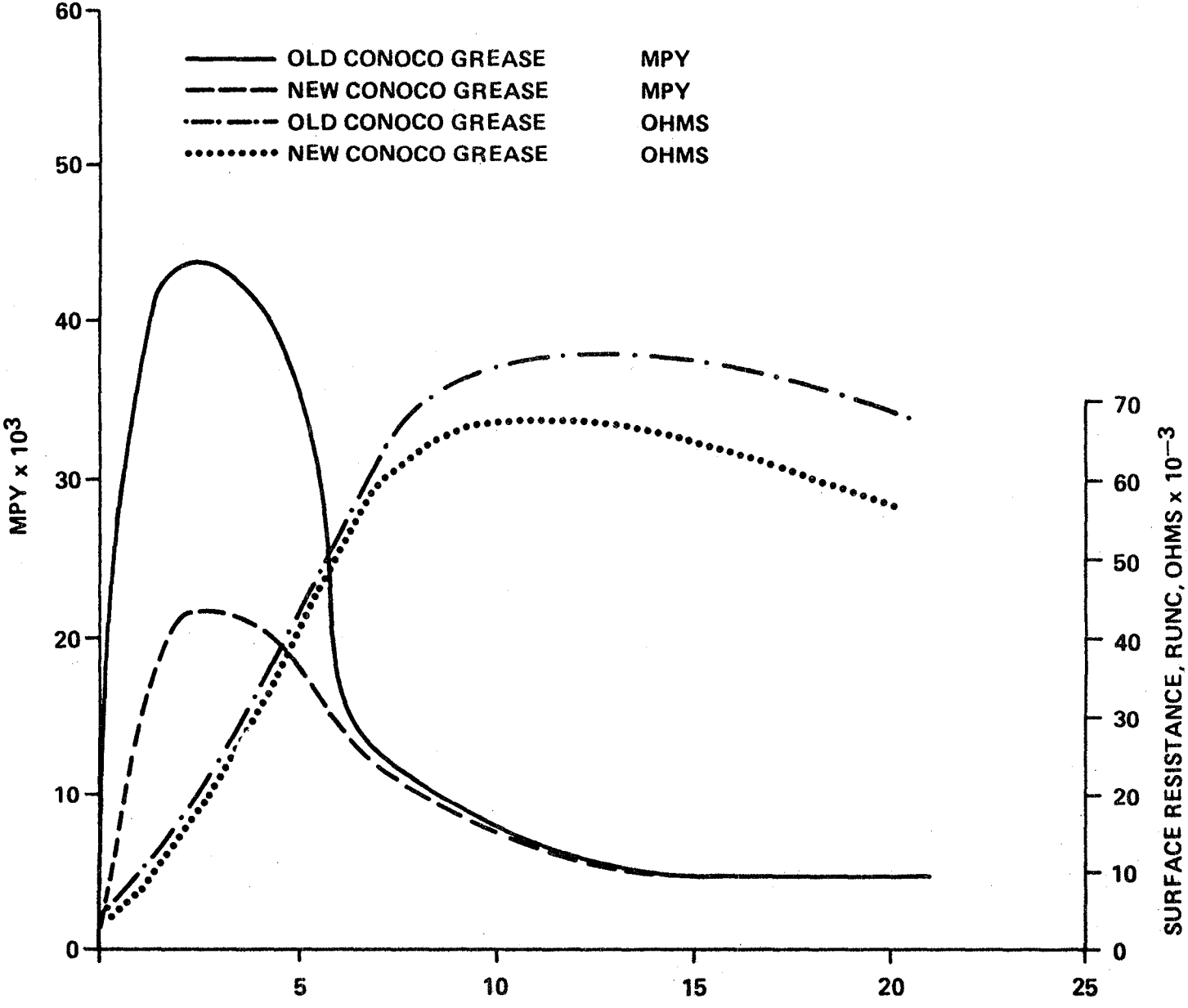


Figure 4

	OLD GREASE	NEW GREASE*
Ca^{++}(mg/g) (deionized water)	1.22	0.32
Ca^{++}(mg/g) 3.5% NaCl	2.94	2.60
Na^{+} (mg/g) (Deionized water)	0.017	0.249

*Conoco HD Calcium Grease No. 2 currently used on SRB

Notes:

1. No dedectable lead in any grease samples.
2. No zinc analysis performed.
3. Anion analysis complete at this time, but more chloride being found in New grease than in Old grease. Also, the sulfonate content of each grease reported to be similar.

Table 1. Cations Leached from Conoco Grease in Two Weeks

COMPARATIVE CORROSION RATES OF 4130 STEEL WITH 2 MILS
CONOCO GREASE① AND TEXACO GREASE②

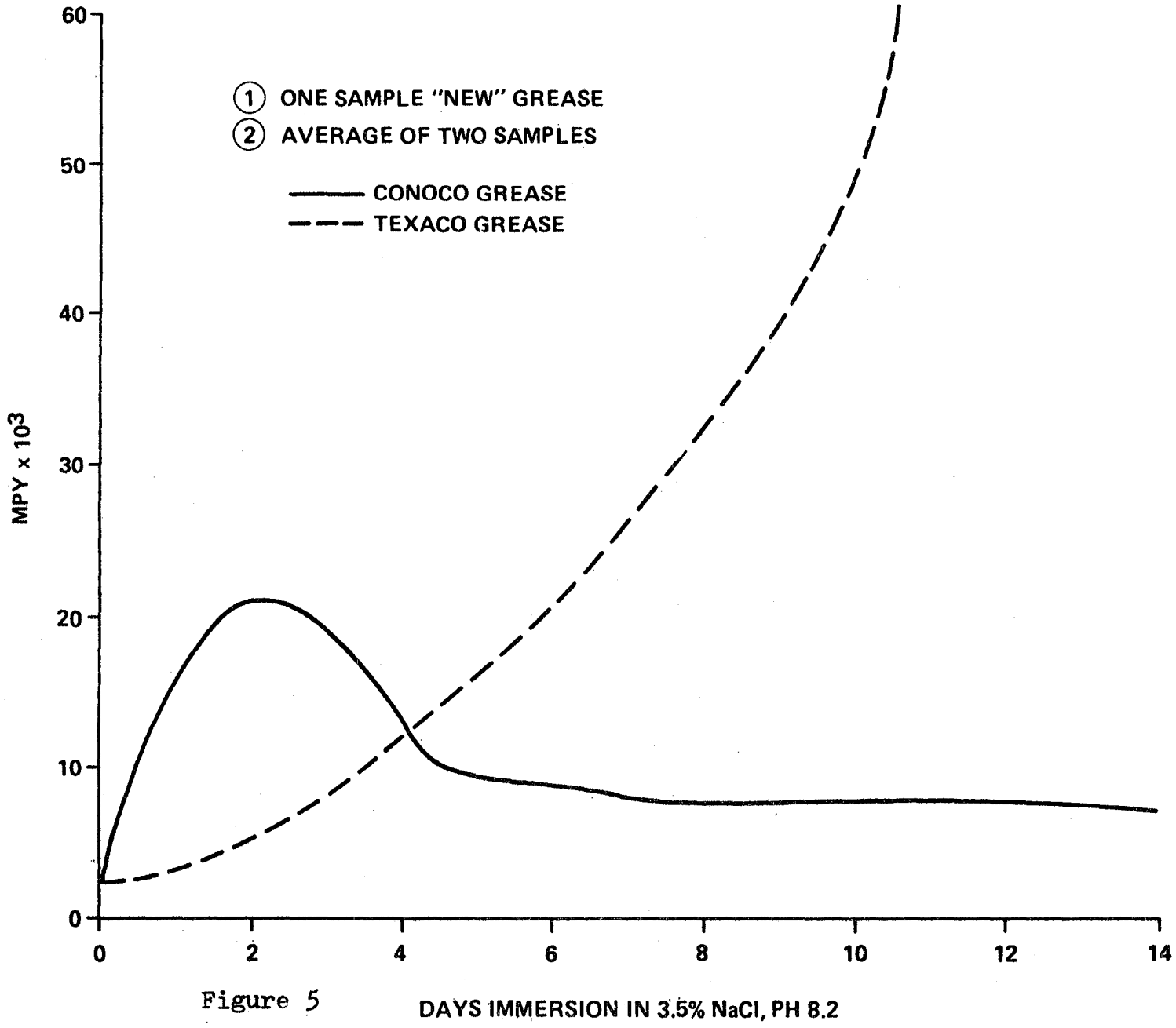


Figure 5

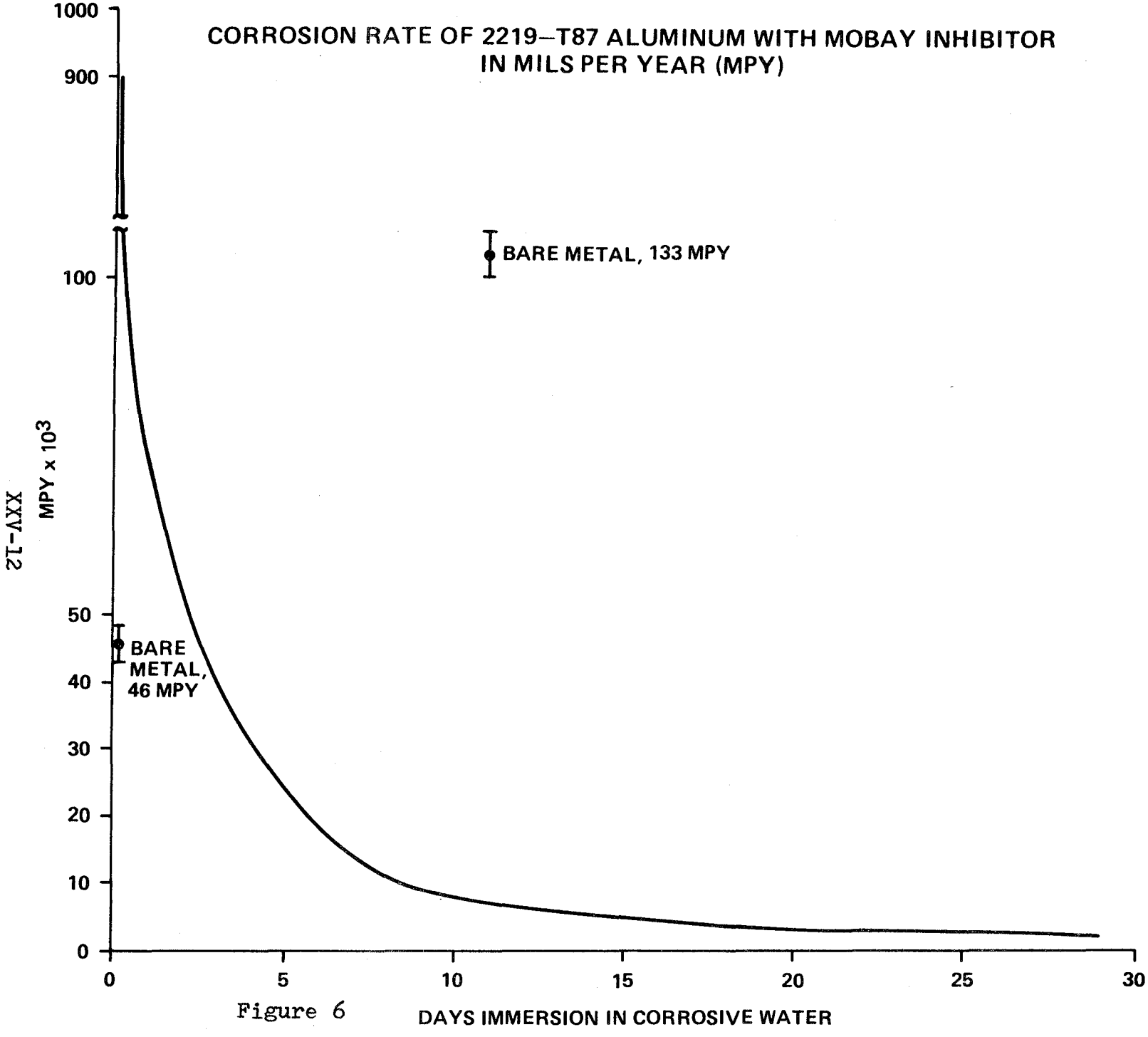

CORROSION RATE OF 2219–T87 ALUMINUM WITH MOBAY INHIBITOR
IN MILS PER YEAR (MPY)
BARE METAL, 133 MPY
BARE METAL, 46 MPY
MPY x 10³
DAYS IMMERSION IN CORROSIVE WATER
1000
900
100
50
40
30
20
10
0
0
5
10
15
20
25
30

Figure 6

one week immersion in the 3.5% salt water. Analysis of solutions in contact with the greases for two weeks indicated significant concentrations of calcium ions as shown in Table 1. Most likely the loss of ions from the greases increased their electrical resistance. In the salt solution an ion-exchange mechanism probably occurs; the solubility of the resulting calcium hydroxide increases because of the high ionic strength (of the salt solution) according to the Debye-Huckel Theory.

2. Texaco grease study. Because Texaco Regal AFB-2 grease was known to be comparatively ineffective in preventing steel corrosion, it was thought useful to examine this grease electrochemically. Two samples of the grease on 4130 steel were prepared as described above. The results indicated that one sample showed fairly good resistance to salt water corrosion, while the other sample produced relatively high corrosion currents after three days. An average corrosion rate calculated from the data from both Texaco samples is compared to Conoco grease coated samples in Figure 5.

3. Aluminum corrosion inhibition study. Research reported in 1978 showed that Mobay OC 2002 Inhibitor gave excellent aluminum protection (Ref. 4). A subsequent electrochemical study showed that the same inhibitor decreased the corrosion rate very little (Ref 5).

A thirty day experiment showed a high initial rate of corrosion that rapidly decreased in about one day, reaching very low rates in about five days. Figure 6 pictures this change. Obviously an "induction period" is needed for the protection mechanism to occur on the metal surface.

4. Seawater inhibition of steel and aluminum corrosion. Various reports over many years point out that seawater contains inhibitors that lessen immersed steel and aluminum corrosion. A brief electrochemical investigation to verify (at least superficially) these reports was conducted using artificial sea water prepared from ASTM Sea-Salt (42.0 grams/liter solution, pH 9.0).

Aluminum (2219-T87) showed significant protection in the seawater with a corrosion rate of about 0.01 mpy. In 3.5% NaCl solution at the same pH the aluminum corroded at a rate of 2 mpy.

Similar experiments with 1010 and 4130 steels indicated no difference of corrosion rate in the artificial seawater and 3.5% salt solutions.

5. Effect of magnesium ions on steel and aluminum corrosion. A brief study to determine the difference in corrosion rates for 4130 steel and 2219-T87 aluminum in 3.5% NaCl and 2.86% NaCl/0.52 $MgCl_2$ solutions was conducted.

At pH's 4.9, 6.9, and 8.9 no inhibition of steel corrosion by the magnesium was observed. (A sample of 1010 steel, however, did show about a 50% reduction in corrosion at pH 8.9 with magnesium ions present).

Although the large variation of corrosion currents with the aluminum makes definite conclusions impossible, a three to four-fold decrease in corrosion at pH 4.9 was observed.

6. Effect of two chelating species on steel corrosion rates. In as much as oxalate and ethylenediammine tetraacetic acid (EDTA) are excellent chelating species for Fe^{2+} ions, it was hypotehsized that these two species would accelerate corrosion by solubilizing the dissolving iron and prevent a protective rust coating from forming.

This was verified by 1010 steel showing a 40% increase in corrosion rate in corrosive water 0.01 M in oxalic acid and a 90% increase in the same medium containing 0.01 M EDTA.

CONCLUSIONS

1. The electrochemical polarization-resistance method utilizing the Model 350 Corrosion Measurement Console is an effective means of determining corrosion rates of grease coated steel surfaces.

2. Conoco HD Calcium Grease #2 was shown to be as effective as the Conoco grease used before January, 1983 on the SRB joints in limiting corrosion rates on 4130 steel surfaces immersed in 3.5% NaCl, pH 8.2.

3. Conoco HD Calcium Grease #2 was shown to provide greater corrosion prtotection than Texaco Regal AFB-2 Grease on 4130 steel surfaces immersed in 3.5% NaCl, pH 8.2.

4. Mobay OC 2002 aluminum corrosion inhibitor provides excellent corrosion protection for 2219-T87 aluminum.

5. The corrosion of aluminum is inhibited to a significant degree in an aqueous environment that contains Mg^{++} ions.

6. Oxalic acid and ethylenediamminetetraacetic acid both accelerate the corrosion rate of 1010 steel in an aqueous environment.

RECOMMENDATIONS

1. The corrosion rates of nickel and nickel alloys in ammonia and hydrazine solutions be examined electrochemically by the polarization-resistance method.

2. At least two more corrosion measurement cells should be purchased by EH 24 in order that more concurrent investigations can be conducted.

3. The POLCUR computer program should be examined to determine the reason(s) that it does not process some polarization-resistance data effectively.

4. An extended investigation should be conducted on the effect of dissolved chelating agents on the corrosion rates of various metals.

REFERENCES

1. Wormwell, F. and Brasher, D. M., Surface Potential Changes of Painted Steel, J. Iron Steel Institute, Vol. 162, 1949, p. 129.

2. Danford, M. D. and Higgins, R. H., An Electrochemical Study of the Corrosion Behavior of Primer Coated 2219-T87 Aluminum, NASA Technical Paper 2459, April 1985.

3. Mansfeld, F., The Effect of Uncompensated IR-Drop on Polarization Resistance Measurements, Corrosion, Vol. 32, 1976, p. 143.

4. Humphreys, T. S., Corrosion Inhibitors for Solar Heating and Cooling Systems, DOE/NASA TM-78180, August, 1978.

5. Danford, M. D., A Brief Examination of the Possible Application of the Model 350 Corrosion Measurement System to the Study of Inhibitors, NASA Corrosion Research Branch Note, Nov.24,1981.

1985

NASA/ASEE SUMMER FACULTY RESEARCH FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

ANALYSIS OF HIGH-ORDER LANGUAGES FOR USE ON SPACE STATION APPLICATION SOFTWARE

Prepared by:	Arthur Knoebel
Academic Rank:	Professor
University and Department:	New Mexico State University Department of Mathematical Sciences
NASA/MSFC:	
Division:	Software and Data Management
Branch:	Systems Software
MSFC Counterpart:	John W. Wolfsberger
Date:	August 15, 1985
Contract No.:	NGT 01-008-21 (The University of Alabama)

Analysis of High-order Languages for Use on Space Station Application Software

by

Arthur Knoebel
Professor of Mathematical Sciences
New Mexico State University
Las Cruces, New Mexico

ABSTRACT

Considered in this study is the general and not easily resolved problem of how to choose the right programming language for a particular task. This is specialized to the question of which versions of what languages should be chosen for the multitude of tasks that the Marshall Space Flight Center will be responsible for in the Space Station. Four criteria are presented: theoretical considerations, quantitative matrices, qualitative benchmarks, and the monitoring of programmers. The report concludes with specific recommendations for future studies to resolve these questions for the Space Station.

ACKNOWLEDGEMENTS

John Wolfsberger, with his incredible energy, is creating a language laboratory whose rare climate is extraordinarily stimulating to work in. Bob Stevens, Dave Aichele and Gabe Wallace have been very supportive and sensitive to the needs of a summer fellow. Gerry Karr, with his dry sense of humor, has administered the summer program with aplomb and grace. Jim Dozier and Leroy Osborn planned and executed an informative round of talks and tours. Dina Conrad, with much patience, has thoughtfully tended to the day-to-day needs of the summer faculty. Jimmy Watkins and Ester Gill, my office mates, have been quite helpful in breaking down the resistance of recalcitrant computer programs. All of these deserve a round of applause and a bouquet of flowers for making my summer a profitable busman's holiday away from academia.

TABLE OF CONTENTS

INTRODUCTION

Problem As one of the future major projects of NASA with a long lead time, the Space Station is a good place to reconsider software and hardware decisions made in the past. This in fact is being done: it appears that Ada will be chosen as the principal or preferred programming language for the Space Station. Given that this happens, we still have some significant questions to answer. What should the backup language be, in case the current doubts about Ada materialize as experience with this new, large and untried language is gained. What about special purposes? Certainly we should allow other languages specially designed for such areas as robotics and artificial intelligence. And if we allow alternates, then we must select the right dialect. Clearly there is much to resolve in the language arena.

History of Languages. A brief history of programming must start with assempbly languages forty years ago. To this FORTRAN added algebraic expressions. Control abstraction came with Algol. Pascal is notable for a variety of data types and the ability to check their conformity at compile time. Pascal's biggest drawback, the inability to compile parts of a program separately, was taken care of in a variety of languages.

At Marshall Space Flight Center, the four most popular languges by far have been assembly , FORTRAN, HAL/S and C. Assembly has been essential from the first for real-time applications, where until recently there have been no high-order languages with adequate input-output facilities for communicating with sensors and servos. FORTRAN, the perenial bellwhether, is now used principally for calculations in scientific research and engineering design . The space shuttle gave birth to HAL/S, and no doubt HAL/S will die with it. C is the least used at MSFC.

Language Choice. Does it make a difference what language is chosen? Yes, it certainly does. For surprisingly not all so-called high order languages are even equivalent computationally to a primitive Turing machine; we discuss this point in more detail in the body of this report. For Turing-machine equivalent languages, the selection of primitive operations, data types and control structures can make a big difference in the ease of programming and the avoidance of errors. Also, a variety of features may provide for more extensive error checking at compile time.

The need for an intelligent decision on a programming language comes about for a variety of reasons. The kind of task is all important -- is it numerical or symbolic, real-time or batch, for research or production? How big is the task -- is it for one programmer over a month or two, or at the other extreme, for more that a hundred programmers over a several years? How many programmers -- one programmer can be somewhat sloppy and informal about direction and management, in a multi-programming some kind of configuration controls are essential, and modularity of the language can help here. The complexity of the task makes a difference too -- here we don't mean size; rather are the parts of the problem to be programmed similar in style or are there many different concepts to be integrated?

Objectives. The original objectives were five in number:

1) Review Space Station mission requirements to establish variant needs;
2) Identify logical language candidates to satisfy needs;
3) Establish benchmarks to compare different languages;
4) Set up a plan for the analysis;
5) Identify in-house capabilities to perform this analysis.

This is not all that is covered in this report. Because of the importance of being able to fruitfully compare programming languages and the lack of surveys of this area, several additional topics are presented. In particular, we consider several other language criteia: theoretical, matrix methods and the direct observation of programmers at work. Because of the effort spent in the laying of this extensive groundwork, there was not time to complete the third objective of establishing benchmarks.

SPACE STATION REQUIREMENTS

General Mission Requirements. Some of the key features specified for the Space Station which are pertinent to this discussion are one language, commonality of software, evolutionary growth, flexibility for many different tasks such as data handling, scientific computation and real-time control, and, most unusual for an endeavor of this kind, a 30 year lifetime. In more detail, from the Space Station Program Description Document [SSPDD, book 1, p. 3-2], we can pick out those particular features which will be of concern

to programmers. It is mandated that the Space Station:

> Shall assume a phase C/D start by or before FY 1987 to support a flight in the early 1990's;

> Will be shuttle-compatible for delivery, assembly, and disassembly;

> Will be supported by the Shuttle, initially on 90 day cycles;

> Shall have a design goal for indefinite life through on-orbit maintenance, repair, or replacement;

> Shall have modular-evolutionary design that permits growth and accepts new technology;

> Shall consider both the initial development cost and life-cycle costs as a design driver;

> Shall be user-oriented to the maximum extent possible implying flexibility and simplicity of user interface systems and documentation;

> Shall have a design goal of commonality for hardware and software of identical or similar functions in terms of systems, subsystems, and interfaces;

> Shall incorporate on-orbit autonomous operations to minimize crew and ground involvement as a design driver."

MSFC Tasks. Marshall Space Flight Center is responsible for Work Package 1, which includes the common module, the logistics module, the laboratory module, the orbital maneuvering vehicle, and the orbital transfer vehicle. The common module will contain the following systems: data management, electrical power, thermal communications, environmental control and life-support. By outfitting it in various ways the common module can be turned into the logistics, laboratory or other kind of module. The logistics module is the supply depot and will hold fuel and fluids, racks of solid consumables, payloads, and equipment to support transport and transfer. The laboratory module is for technology and manufacturing and is to be outfitted so as to be as simple as possible for the users. Marshall is also responsible for accomodating and servicing the OMV and OTV. For more details, see [SSDPD].

Their Programming Needs. From the contractors' reports([Int], Table 1 and [Liu], Table 1), we find what is needed in the way of language characteristics to fullfill Work Package 1. Since we will be talking more about these kinds of matrices in a later section, for now we summarize their contents. Needed for Work Package 1, in addition to the usual primitive numerical, character, logical and array data types, and scientific functions built out of these, are machine addressibility, priority control, graphics capability and variable-length records, the last for operating systems. In each these reports, these language characteristics are correlated with the particular systems given above, where

they are needed. Interestingly enough, nothing in these lists mentions the need for languages capable of supporting more exotic systems such as robotics or expert systems; we will now address this.

Robotics. Automation and robotics are mandated to play a significant role in the Space Station. No doubt, this will drive the invention of new computing devices, and consequently, new computer languages. For this reason, we briefly review the future of robotics in the Space Station.

The best reference for this purpose is [AART]. This report, prepared by the Advanced Technology Advisory Committee of NASA, makes thirteen principal recommendations to NASA, of which four are pertinent to us:

> Automation and robotics should be a significant element of the Space Station program;

> The initial Space Station should be designed to accomodate evolution and growth in automation and robotics;

> The initial Space Station should utilize significant elements of automation and robotics technology;

> NASA should provide the measures and assessments to verify the inclusion of automation and robotics in the Space Station.

Then the report on p.11 goes on to propose specific goals in each of the subsytems listed in the a previous chapter; these goals recommend the development of robots, intelligent controllers and expert systems for a variety of specific tasks.

In anticipation of such acitivity, it would seem prudent to identify appropriate programming languages that will be needed or will enhance such projects. Projects now underway at NASA in automation and robotics are outlined in [NAR].

Artificial Intelligence. Expert systems[McA], such as the Liquid Oxygen Expert System at Kennedy Space Center and others[CA], appear as the most visible and successful subfield of that amorphous body of assorted knowledge known as artificial intelligence. Whether or not other research now being done in A.I. will evolve over the lifetime of the Space Station into equally useful applications remains to be seen. LISP and Prolog are advocated and used for research and some applications because of the lack of explicit control This is often a drawback since the resulting large search space causes slow performance. Some, therefore, advocate using these functional logical languages for research and, when the research proves successful, redoing the algorithm in a procedural high-order language. This goes along with the findings of Schwartz and Melliar-Smith[SM, p.1] that Ada, as presently standardized, is unsuitable for artificial

intelligence applications, but that, "with relatively modest extensions to Ada within the spirit of the language, it would be possible to translate a substantial proportion of AI algorithms ... into Ada."

CANDIDATE LANGUAGES.

Desideratum. A time scale of thirty years into the future is extraordinary. The opposite, thirty years into the past, by way of comparison, would take us back to 1955, the time of the first Univacs and IBM's. Enormous changes have occured since then. Programming has advanced from assembly language to high-order languages, with major concepts emerging at least every decade: algebraic expressions, control abstraction, data types and modularity (Cf. Ng and Falkenstein in [SSSI]). How can we intelligently plan in the next three decades for breakthroughs yet to come when we probably now have no inkling of what some of them will be? Truly a Herculean task requiring the wisdom of Solomon.

Given that design and development of the space station is to begin in 1987[SSPDD], and that the software design environment should presumably be in place by about then, together with the fact that significant software innovations typically take around ten years to develop from first concept to finished product, then we see that there is not much time for extended experimentation. What is to be done? We should identify those seeds of contemporary software research which are likely to blossom and bear fruit in future languages. Examples might be: equational inference for data types as in OBJ; more generally, relational inference as in PROLOG; and further advances in generic definitions. Whatever languages are chosen must be extensible and leave room for such innovations in the decades to come.

Candidates. We should evaluate and compare these languages: HAL/S, Forth, C, Pascal, Modula-2, Ada, FORTRAN, APL, LISP, and Prolog. The first seven are procedural languages; the last three are functional. FORTRAN and APL are useful in mathematics, engineering and applied mathematics. LISP and Prolog are often seen in artificial intelligence.

We briefly review the principle features of these candidates.

> HAL/s. Block-oriented and structured with many modern featues but also some unsafe ones. Good for real-time applications; used only within NASA and its contractors.

> FORTH. A small but powerful, extensible

language originally designed for real-time control of telescopes.

> C. Designed for systems software; structured but considered lower-level by many. Some programmers find C, as well as Forth, hard to read.

> Pascal. Highly popular but with design deficiencies -- less powerful than a Turing machine. Structured but no sparate compilation of program units.

> Modula-2. Wirth's worthy successor of Pascal. Capable of multitasking, intended for dedicated computers. Not widely used yet. See [MG] and [Wi].

> Ada. For embedded systems; with multitasking; separately compilable and reusable packages; generics.

> FORTRAN. Second high-order language to be designed. Lacks many features now available on languages designed later, but very widely-used among in engineering circles.

> APL. A mathematically oriented interactive language with many powerful primitives. Limited data types make it difficult to use for purposes for which it was not intended.

> LISP. The first high-order language to be designed. Useful for logically -oriented problems. Many versions, the later ones of which are quite powerful.

> Prolog. Intended for logic programming, which means working in essentially first-order predicate calculus, as opposed to working with functions in LISP. Quite special purpose. See [CM].

ADVANCES IN HARDWARE

As with software, hardware has also seen significant breakthroughs, again about once a decade. It has advanced from vacuum tubes through discrete transistor-diode logic to small scale intregration and finally to very large scale integration today. What further advances can be seen in a crystal ball? Clearly, further miniaturization of circuits and consequent parallel processing[And] are in store for us. Possible, but not certain, are wafer-size chips and truly three-dimensional layouts. What effect these developments will have on language design has been little explored[Brol], but some conjectures will have to be made. We could chose to ignore these questions, but in as much as room and power are at a premium in the Space Station, it behooves us to incorporate hardware improvements as they become available. As an aside, CAD advances are making it easier and easier for small groups to design their own support chips (e.g., the Berkeley toolkit and silicon compilers for microcomputers). Witness the current efforts of JPL in this regard. Thus,

hardware design no longer entirely depends on what chips are commercially available.

CRITERIA

Theoretical Considerations. It is often said that all programming languages are equivalent in the sense that anything which can be computed in one language can be computed in any other. According to this view, any perceived differences are strictly a matter of performance, speed, memory, or ease of programming. When pressed, an advocate of this view will claim that it is obvious that all programming languages are computationally equivalent to a Turing machine. To the contrary, we claim this is not obvious, in fact, it is false for some well known languages, for example, Pascal as original defined by Wirth.

We would also claim that one must clarify what is meant by a programming language being equivalent to a Turing machine. The simplest assertion to try to prove is that all the functions which a Turing machine can compute, namely, the partial recursive funcions, are precisely the ones which the language in question can compute. But this is patently false for Pascal since it does not allow dynamic arrays, and hence cannot even compute the concatenation of an arbitrary array with itself.

But one can pursue this line of inquiry further, and ask what is the common alphabet over which these functions are being computed. This is necessary in order to talk about recursiveness. When input to hardware was a paper tape, the analogy with a Turing machine with its hypothetically infinite tape was easy to see. What is it today: a screen, a disc, a printer? Since interactive computing has become so popular, the screen would seem to be the obvious medium. If any two languages are equivalent, and one can do whatever the other can, then it would follow that it should be possible to write a program in the first language which will make the screen behave in such a way that an unsuspecting programmer would think he was dealing exclusively with the second (this is reminiscent of Turing's test for artificial intelligence). But does anyone really think they could write a Basic program which would present to the unwary programmer sitting at his terminal what would appear to be a bona fide Ada editor, compiler and executor?

So far this is accepting the definition of a Turing machine at face value. But in its definition, there is no provision for formulating concepts having to do with input-

output, multitasking, and real-time behaviour. Thus, it really makes no sense to talk of recent high-order languages such as Ada and FORTH as being equivalent to a Turing machine.

Quantitative Benchmarks. In general, a benchmark is a standard or reference point for measurement. In computer science, it refers to a program which has been so written as to compare quantitatively the performance of different computers, compilers or languages. That a benchmark can do all three is a strong point. It is also a weak point in that it is often difficult to isolate what one is really measuring. Let us look at several benchmarks.

An early one is the Gibson mix, which "is a weighted average of instruction times , reflecting usage expected of scientific work."[CW] Because the performance of this mix reflects peculiarities of particular computers, it has been criticized for unduly penalizing computers without pipelining and primitive floating point operations.

The Whetstone benchmark[CW], developed to meet these objections by the Computer Agency of the United Kingdom, "does not solve a real problem; it is a mix of higher order language statements that reflect the frequency of such statements as measured in over 900 actual scientific programs.[Car]" As such, it is widely used. The result is measured in Whetstone instructions per second(WIPS). Martin-Marietta of Denver, over the last few years, has run this test on a variety of machines, large and small, using an assortment of operating systems and high-order languages. But note: the Whetstone does not test multitasking, interrupt handling, or imput and output.

Caveats. Designing benchmarks is tricky; we want sensitivity but a particular benchmark may measure something different from what we want. For example , the benchmarks of Pulcini[Pul] are quite sensitive to the degree of optimizaton of the FORTRAN compiler. If this is what we're after, fine; otherwise we are in trouble. Consult Pulcini's article to see how other factors may skew the metric.

What are the implications of these studies for choosing high-order languages? First, there do not appear to be any benchmarks that thoroughly test all the features of the more recently designed languages, such as HAL/S, Modula-2 and Ada. Secondly, different benchmarks will have to be designed to measure the performance of languages dedicated to different tasks, e.g., robotics and artificial intelligence. Here again, in general, there are no ready-made benchmarks. As an exception, however, here at Marshall Yancy[Yan] has used a

database benchmark to compare two intelligent database machines.

Also, we mention the paper of Walters[Walt], which discusses what he calls 'efficacy factors' of benchmarks. These are accuracy, sensitivity, effort, and portability. While Walters's comments are addressed mainly to how these affect the evaluation of architectures, much of what he says is transportable to other domains.

See [Lev] for a critical account of other papers discussing benchmarks.

Qualitative Matrices. This is the most popular way to evaluate languages for a particular task. Roughly, one draws up a table with the language candidates at the top, the desired tasks at the side, and the evaluation in the body. This is done in a number of contractors' reports(e.g.,[Int], [Liu])

In practice, the process is more complicated. As an intermediary, one must postulate the programming features which the task needs. A complex project, such as the Space Station, has many systems, i.e., tasks. This leads to two matrices: one with tasks vs. features, and the other with features vs. languages. As an example of this process we reproduce small portions of two matrices, the first from Table 1 on p.5 of [Liu] and the second from Table 2 on p.8 of [Int].

Table 1 -- Requirement Levels for Various HOL Features in Space Station Applications

FEATURE/CRITERIA	GNC	ECLSS	ELEC	THRML	C&T	PL	OMS
Float. Pt. Arith.	m	m	m	m	m	m	n
Fixed Pt. Arith.	n	n	n	n	n	u	n
Integer Arith	m	m	m	m	m	m	m
Bit Manipulation	u	m	m	m	m	u	m
Matrix Algebra	m	n	n	n	m	m	n
.						.	
.						.	
.						.	
Data Abstraction	u	u	u	u	u	u	u
Interfacing with other Languages	u	n	n	n	u	u	n

In the preceding table, "n" stands for nil, "u" for useful, and "m" for mandatory.

Table 2 -- Language Features Comparison.

FEATURE/LANGUAGE	HAL/S	Ada	JOVIAL	conc.Pascal
Float. Pt. Arith.	D	D	D	D
Fixed Pt. Arith.	D	D	-	-
Integer Arith.	D	D	D	D
Bit Manipulation	D	D	D	-
Matrix Algebra	D	L	L	L
Data Abstraction	-	D	-	D
.			.	
.			.	
.			.	

In this table, "D" means that the feature is supported by the language, "L" that the feature can be added as a library of subprograms, and "-" that the feature is effectively unavailable. (We have taken the liberty of rearranging this excerpt of Table 2 somewhat so that the same language features appear in both tables.)

As remarked before, these are the easiest kind of comparative studies to do on languages, particularly if one has personnel who are familiar with the tasks and the languages under consideration. Unfortunately, it is very easy for the entries in these matrices, and hence the conclusions of the studies themselves, to be skewed by subjective biases, as has been demonstrated. See [Cug], [Knos] and [TRW] for other matrix studies.

Direct Evaluation by Programming. The idea here is to set programmers to coding pertinent tasks in the candidate languages. Then observe them at work , time them, and interview them afterwards. On the basis of all this evidence, choose the right language.

One reason for observing programmers at work is that the newer languages are more than just that -- they also encompass programming environments. Thus one aspect of the worth of a new language is how well it can be integrated into a productive system.

The method should be the following. Pick tasks typical of Space Station needs. Perhaps consider recoding some of the programs already done on the Shuttle or Skylab. Have programmers of comparable skill and experience put on teams spread among the candidate languages to do a selected sample of tasks typical for the Space Station. Then time the programmers at their work, obtain compile and run times, check for readability of the resulting code, and ask them their comments, opinions, and recommendations about various

aspects of the languages. One can go further and test for maintainability of the languguages by slightly changing the specifications of the tasks which were just programmed to see how easily the coding can be modified.

This method takes some care to get reliable and reproducible comparisons. As remarked already, the programmer teams chosen must be compararble in experience and background. We can't use the same programmers to code the same task in two different languages since, in recoding the task in the second language, the programmers would already know the task and be ahead. A statistical sample of sufficient size would probably mean too many programmers. Nevertheless, we could probably learn much from such an experiment of small size. For example, see [Kle].

PLANS AND COST

Plan for Analysis. In view of the facts that no backup language for Ada has been identified, and that no languages have been singled out for work in robotics, automation, and artificial intelligence, and that only one of the four methods of testing comparing languages has been done in depth for more than a few of the candidate languages, we offer the following advice.

> Do theoretical work on the lack of equivalence between various high-order languages. Try to find measures of complexity which actually separate well-known languages according to our intuitive feeling for their relative power.

> Develop and establish benchmarks which evaluate advanced programming concepts such as user-defined data types, generics, multi-tasking, modularity and input-output capabilities.

> Develop and establish benchmarks for tasks typical of robotics, automation and artificial intelligence.

> Perform these benchmark tests on our candidate languages.

> Identify language characteristics suitable for robotics and artificial intelligence.

> Carefully pick a set of programmers comparable in ability, pick a subset of the candidate languages which have done well in benchmark and matrix tests as well as scoring high on theoretical grounds, and let the programmers try their respective hands at coding typical tasks in the various languages.

This should be an ongoing study. There are several tasks appropriate for future work.

> Do the previous tests on languages likely to gain

maturity in the near furture, such as Smalltalk and OBJ, which are object-oriented languages suitable for work in artificial intelligence.

> Do a study anticipating future advances in VLSI impinging on languages and programming.

Cost. We make the following ball-park estimates only for personnel required since most of the hardware and software will soon be in place in the language laboratory at Marshall. We do this for each of the four methods for comparison described earlier.

Theoretical Criteria. One-half a man-year should produce some worthwhile results. Not all ramifications could be followed up in that time, but one should know better then what to pursue.

Quantitative Benchmarks. We distinguish benchmarks with only the usual features, those with advanced features such as user-defined data types, and those for artificial intelligence. The cost is given tabularly in man-months.

	usual	advanced	A.I.
design	2	4	3
coding	4	8	6
executing and follow-up	1	2	1 1/2

Qualitative matrices. Much has already been done with conventional procedural languages. The area in which new work should be done is the languages for robotics and artificial intelligence. Identifying features for systems of the Space Station has already been done[AART]. Identifying which special purpose languages have which features should take on the order of a quarter of a man-year.

Direct Evaluation by Programming. A modest effort would be six programmers split into teams for six months, for a total of three man-years.

In-house Capability. We give brief opinions as to what extent the Systems Software Group at Marshall could perform each of the preceding tasks, both as regards equipment and personnel. With the coming installation of several IBM PC/XT's and a VAX, the Software Evalution Laboratory at Marshall, as well as compilers for a wide variety of languages, the Software Evaluation Laboratory at Marshall will soon be have excellent equipment for performing this research. As to personnel, the situation is less clear, and we break these estimates down again by the four kinds of

comparisons.

Theoretical Criteria. There appears to be no one at Marshall with these skills. Perhaps this could be done in academia.

Quantitative Benchmarks. As this would involve mimicking tasks similar to those already done in previous missions, the personnel at Marshall could certainly do this.

Qualitative Matrices. Clearly there are sufficient staff at MSFC in A.I. and robotics to draw up a matrix of desirable language features in these areas. Thus the remaining matrix of features vs. language can be drawn up in house.

Direct Evaluation by Programming. There are more than enough programmers at Marshall to form some matched teams. The only question is whether Marshall can spare so many.

SUMMARY

We have explored how to compare and evaluate a variety of programmer languages for the many-faceted space Station. Out of this study came four tests: theoretical criteria, quantitative benchmarks, qualitative matrices, and direct evaluation by programming. Marshall should select two or three of these methods to pursue in depth. Some should be done at Marshall to develop in-house expertise. Others can be contracted out.

REFERENCES

[AART] Advanced Technology Advisory Committee. **Advancing Automation and Robotics Technmology for the Space Station and for the U.S. Economy. Vol. 1 -- Executive Overview.** NASA Technical Memorandum 87566. Submitted to the U.S. Congress April 1, 1985.

[And] Warren Andrews. Silicon solutions begin surfacing for high-resolution display problems. **Electronic Engineering Times,** 337 (July 8, 1985), pp. 35-38.

[BEU] Erwin Book; Paul Eggert; Robert Uzgalis. **A Methodology for Evaluating Languages and Their Compilers for Secure Applications.** Command and Control Technical Center TM 171-78, Jan. 31, 1978. AD A056933.

[Bro] Chappell Brown. Ada system rivals supercomputer speeds. **Electronic Engineering Times,** 337, (July 8, 1985).

[Car] Phillip C. Carney. Selecting on-board satellite computer systems. **Computer** (April 1983), pp.35-41.

[CM] W. F. Clocksin; C. S. Mellish. **Programming in Prolog.** Springer-Verlag, 1981.

[Cug] John V. Cugini. **Selection and Use of General-purpose Programming Languages -- Overview, vol.1, Programming Examples, vol.2.** National Bureau of Standards Special Publication 500-117.

[CW] H. J. Curnow; B. A. Wichmann. A synthetic benchmark. **The Computer Journal** 19:1 (1976), pp.43-49.

[FG] Alan Feuer, Narain Gehani, eds. **Comparing and Assessing Programming Languages -- Ada, C and Pascal.** Prentice-Hall, Inc. 1984.

[Hor1] Ellis Horowitz. **Fundamentals of Programming Languages,** 2nd Edition. Computer Science Press, 1984.

[Hor2] Ellis Horowitz. **Programming Languages: A Grand Tour.** Computer Science Press, 1983.

[Int] Intermetrics, Inc. **Justification for Selection of a Standard Language for Space Station Applications.** Final Report, 28, June 1985. (Prepared for Rockwell International)

[Kle] Melvin Klerer. **Software Modeling Studies -- Experimental Study of a Two Dimensional Language vs. FORTRAN for First Course Programmers.** Rome Air Development Center, July 1981. RADC-TR-81-183, vol. III.

[Klu] Allan R. Klumpp. Space station flight software: HAL/S or Ada? **Computer** 18:3 (March 1985), pp. 20-28.

[Knos] Alton A. Knosp, Jr. **AIPS Language Trade Study.** Charles Stark Draper Laboratory, Cambridge, Mass. March 1984. CSDL-C-5694.

[Lev] Bruce Leverett. **Performance Evaluation of High-level Language Systems.** Air Force Office of Scientific Research, Nov. 77. AFOSR TR-78-1088.

[Liu] O. Liu. **Assessment of Ada for Space Station Applications Software.** McDonnell Douglas Astronautics Co. July 1, 1985.

[McA] Michael A. McAnulty. Evaluationof Expert Systems. **Research Reports -- NASA/ASEE Summer Faculty Research Fellowship Program,** Marshall Space Flight Center, Jan. 1985. NASA CR-171317.

[MG] Joel McCormack; Richard Gleaves. Modula-2 -- a worthy successor to Pascal. **Byte** (April 1983). pp.385-395.

[MM] Martin-Marietta Corp. **Whetstone Benchmark Test.** (Tests performed 1981-1984).

[NAR] Advanced Technology Advisory Committee. **NASA Automation and Robotics -- Information Exchange Workshop - Proceedings. Vol. 1 - Executive Summary.** Held at Lyndon B. Johnson Space Center, Houston, Texas, May 13-17, 1985. JSC Artificial and Information Sciences Office, June 5,1985.

[Per] Alan Perlis; Frederick Sayward; Mary Shaw (editors). **Software Metrics: An Analysis and Evaluation.** MIT Press, 1981.

[Pul] Jeffrey Pulcini. Constructing benchmarks that measure up. **Computer Design** 22:11 (Oct. 1983). pp.161-168.

[SM] Richard L. Schwartz; P. M. Melliar-Smith. **On the Suitability of Ada For Artificial Intelligence Applications.** SRI Int. July 1980. AD A090790.

[SMB] C. H. Smedema, P. Medema, M. Boasson. **The Programming Languages: Pascal, Modula, CHILL, Ada.** Prentice-Hall, 1983.

[SSDPD] **Space Station Definition and Preliminary Design -- Request for Proposal.** NASA, Sept 15,1984.

[SSFDD] **Space Station Program Description Document.** NASA, Space Station Task Force, March 1984. NAS 1.15:86652/bk.1-4,6-7.

[SSSI] **Space Station Software Issues.** Report of a workshop held at NASA Langley Research Center, Hampton, Virginia, August 20-21, 1984. NASA Conference Pub. 2361, 1985.

[SSSR] **Space Station Software Recommendations.** Report of an Open Forum with Industry and Academia held at NASA Marshall Space Flight Center, Huntsville, Alabama, April 24-25, 1985.

[TRW] **Language Task 3 Trade Study.** TRW Defense Systems Group, April 1985.

[Walt] R. E. Walters. Benchmark techniques: a constructive approach. **The Computer Journal** 19:1 (Feb.1976), pp.50-55.

[WS] B. A. Wichmann; A. H. J. Sale. **A Pascal Processor Validation Suite.** National Physical Laboratory Report CSU 7, Teddington, Middlesex, U.k. march 1980. ISSN o143/733X.

[Wi] Niklaus Wirth. **Programming in Modula-2,** 2nd edition. Springer-Verlag, 1982.

[Yan] K. E. Yancey. The intelligent database machine. **Research Reports -- The 1984 NASA/ASEE Summer Faculty Research Fellowship Program.** Marshall Space Flight Center, Jan. 1985. NASA CR-171317.

1985

NASA/ASEE SUMMER FACULTY

RESEARCH FELLOWSHIP

PROGRAM

MARSHALL SPACE FLIGHT CENTER

THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

INVESTIGATIONS OF TOUGHENING MECHANISMS OF EPOXY RESINS

Prepared by: T. Koenig

Academic Rank: Professor

University and Department: University of Oregon
Department of Chemistry

NASA/MSFC:

Division: Non-Metallic Materials

Branch: Polymers and Coatings

MSFC Counterpart: Dr. W. J. Patterson
M. L. Semmel

Date: August 1, 1985

Contract No.: NASA-NGT-002-099

The University of Alabama in Huntsville

INVESTIGATION OF TOUGHENING MECHANISMS OF EPOXY RESINS

by

T. KOENIG

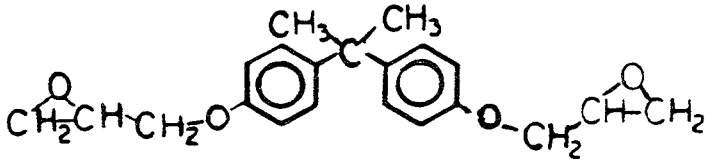
Professor of Chemistry
University of Oregon
Eugene, Oregon 97403

ABSTRACT

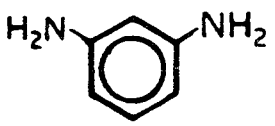
Composite material technology has been applied to the solid rocket booster by the development of a carbon filament-epoxy resin case which yields a net increase of 4000 lbs. in payload in the shuttle. The question of reusability of the new composite tanks has not yet been answered and will depend on the toughness of the matrix resin. The present study was aimed at providing conditions whereby test specimens of the epoxy resin (EPON/825) and curing agents of systematically varied structures could be produced in a controlled manner. Three sets of conditions were found that might allow the isolation of the structural effects on toughness from the cure effects. The kinetic methods leading to the determination of these conditions are described.

I. Introduction and Objective

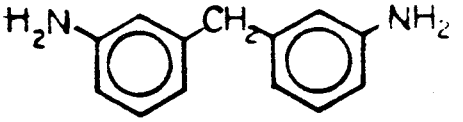
A composite material tank for the solid rocket booster of the space shuttle is being developed. This development will result in a net increase of 4000 lbs. in payload in the shuttle. The composite material consists of a carbon filament in an epoxy-diamine matrix resin. One of the unanswered questions about the new filament wound tanks concerns their reusability. The answer to this question will depend on the impact strength and fracture toughness of the matrix resin. The present project involved a study of the kinetics of polymerization of epoxy[1] EPON/825 (E)[2] with curing agents of systematic variation in structure (M, D, T)[2].



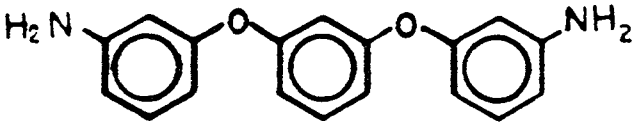
E



M



D

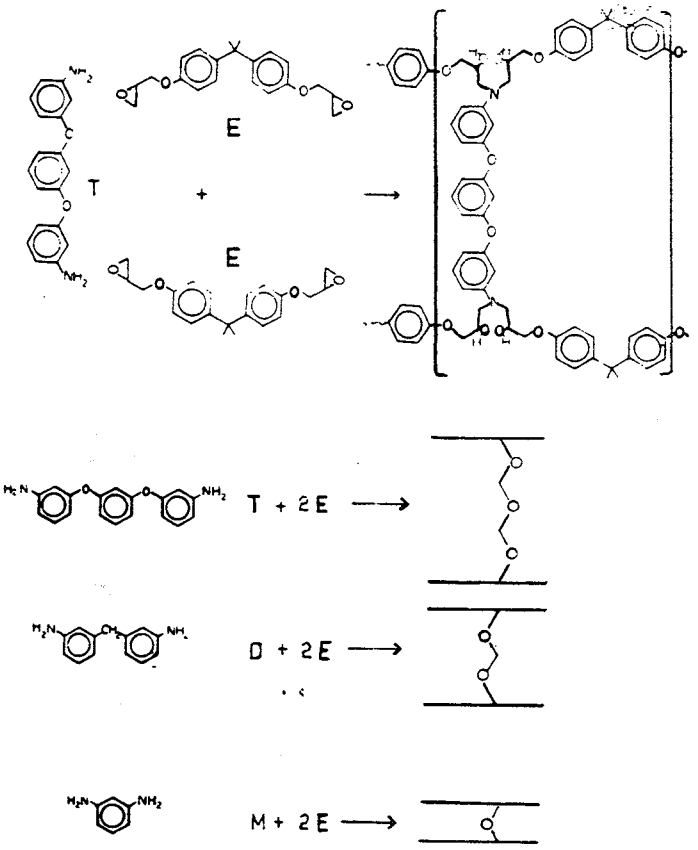


T

The important features of the structural variation of the diamines are the number of aryl groups (M, monaryl; D, diaryl and T, triaryl) and the meta arrangement of all arene linkages. Scheme I shows the overall reactions[1] responsible for formation of the matrix resin. As indicated in the top reaction between T and E, each of the four H atoms bound to the nitrogen atoms of the starting diamine is transformed to an OH group in the resinous product through reaction with one epoxy unit on E. The chemical stoichiometry thus requires 2 moles E for each mole diamine (M, D or T). The lower three equations are simplified representations of the reaction of 2E with the three respective diamines. There

is a good theoretical basis to expect that the present variation of meta-aryl diamines should produce materials of variable impact strength as suggested by the simplified representation of the three products in Scheme I. An increase in the number of meta-aryl groups between the cross-linked chains should provide a mechanism for dissipation of the energy of impact through the increased number of random coil conformational possibilities associated with the increase in number of meta-aryl subgroups.

SCHEME 1



The objective of the present work was to find reaction conditions that will allow the production of test specimens of matrix resins with variable diamine structure and common (E) epoxy structure. The challenge in this objective is to be able to produce specimens that will isolate the effects of the diamine structural variation from effects simply due to the extent of polymerization. Meeting this challenge requires a fairly detailed study of the kinetics of the polymerization reaction with each of the three diamines. The results of these studies suggest several conditions where specimens can be produced that will allow the desired isolation of the structural variation effects.

II. Materials and Methods

Some attention was paid to the purity of the starting materials. Appendix I contains the proton magnetic resonance spectra of each of the four compounds (E, M, D and T) at low (60 MHz) and high (300 on 360 MHz) field which confirm the structure and purity in each case. The monoaryl system (M) is easily oxidized in air and pure material required three distillations at 0.5 torr/80-90°C, m.p. 63-65°C; lit.[3], 64.5°. Two sources of the diaryl diamine (D) were available. The Fairfield Chemical Co. material was found to be of excellent purity (m.p. 85-86°; lit.[4], 82-83°). The Synmet Inc. batch of D was grossly the correct material but contained colored impurities and had a low melting point (80-82°). The spectral data on T indicated it was of high purity (m.p. 106-108°) but a previous chromatographic analysis showed minor impurities were present. This material was used without further purification.

The infrared spectra were also examined for each of the compounds and found to confirm the structures. Figure 1 contains a portion of the infrared spectrum and the gel permeation chromatogram of EPON/825 (E). The infrared in this case shows the presence of a small amount of OH group which is characteristic of a small amount of the dimer of E usually[5] present in this product. This impurity is important since it affects the epoxy equivalents of the resin. The gel permeation chromatogram indicates a relative area of 0.9% for the dimer. This translates into an epoxy equivalent weight of 170.2±.2 vs. 170.0 for the pure material. All experiments used 170 as the epoxide equivalent weight of the EPON/825 batch which was the same for all runs.

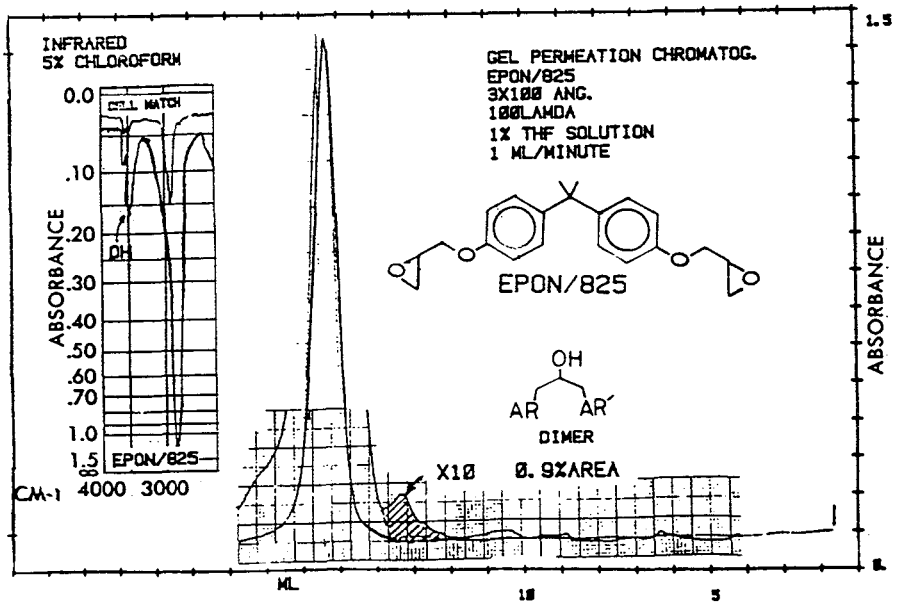


FIGURE 1.

Kinetic determinations were by differential scanning calorimetry (DSC) with a DuPont 1090 system. Figure 2 shows the principal observables with this equipment. The polymerization reaction is exothermic and the heat released by the reaction can be measured by integration (Figure 1, upper) of the DSC peak (Figure 1, lower). Melting transitions and glass transition temperatures are also revealed by this method.

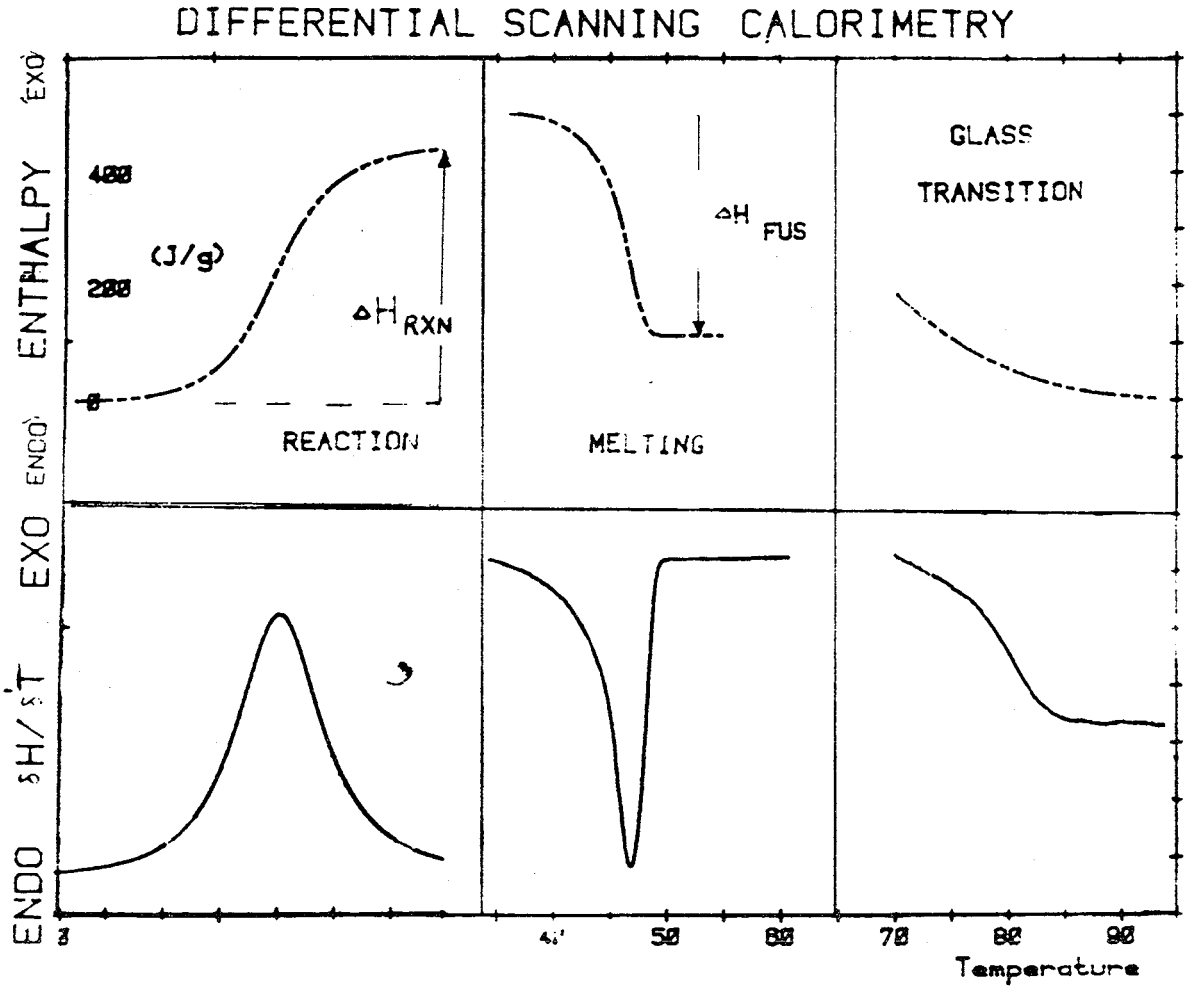


FIGURE 2.

Kinetic samples were prepared by dissolving 2 mmolar equivalents (4 epoxy mequivalents) of E and 1 mmolar equivalent (4-H mequivalents) of diamine (M, D or T) in I to 1.5 g of methylene chloride. Only T presented any solubility difficulties and at least 1.5 g of solvent was required to obtain a homogeneous solution in this case. Constant aliquots (usually 25 microliters) were added to tared DSC pans. Most (90%) of the solvent evaporated in 1 hour at room temperature and 3 of a particular sample set were heated at 103° for ca. 1 hour to complete evaporation. The residual weights of resin mixture obtained in this way were generally within a 1% range. The average residual weights were taken as constants for the remaining samples prepared from the same original solution. The rate of reaction at room temperature was shown to be slow during the first two hours for all three systems. However, the monoaryl system (M) does begin to react at an appreciable rate after a few hours at 25°. Kinetic

samples, prepared in this way, were heated for variable times in ovens regulated to ± 1° C. Temperature calibration was to the DSC through a common thermocouple also to ± 1° C.

Table 1 gives the average heats of reaction at zero reaction time for each of the variable diamines used in this study. The average deviations (± A) are within the 2.5% typically reported[6] in the literature for similar systems. The range (R) is, however, large enough to make small extents of reaction (short reaction times) difficult to determine. All DSC runs were heated at 10°/min unless otherwise specified.

Table 1

Heats of Reaction (Q_o), Average Deviations (A) and Deviation Ranges (R)
N Determination of 2E:1M, 2E:1D, 2E:1T

Resin [a]	Q_o (J/g) [b]	A	R	N
2E:1M	434	10	26	4
2E:1D	371	8	32	10
2E:1T	321	8	32	8

a) 2 moles E to 1 mole diamine. b) At zero time, DSC heat rate 10°/min.

III. Results

Figure 3 represents the simplest method of kinetic determination in these systems which is the batch version. In this method, a sample of the resin mixture of known weight is simply heated for a time, quenched and subjected to a standard DSC scan. The integrated area (ΔQ) divided by the appropriate Q_o (Table 1) can be taken to measure the Fraction cure (F) or percent cure (Fx100) as shown.

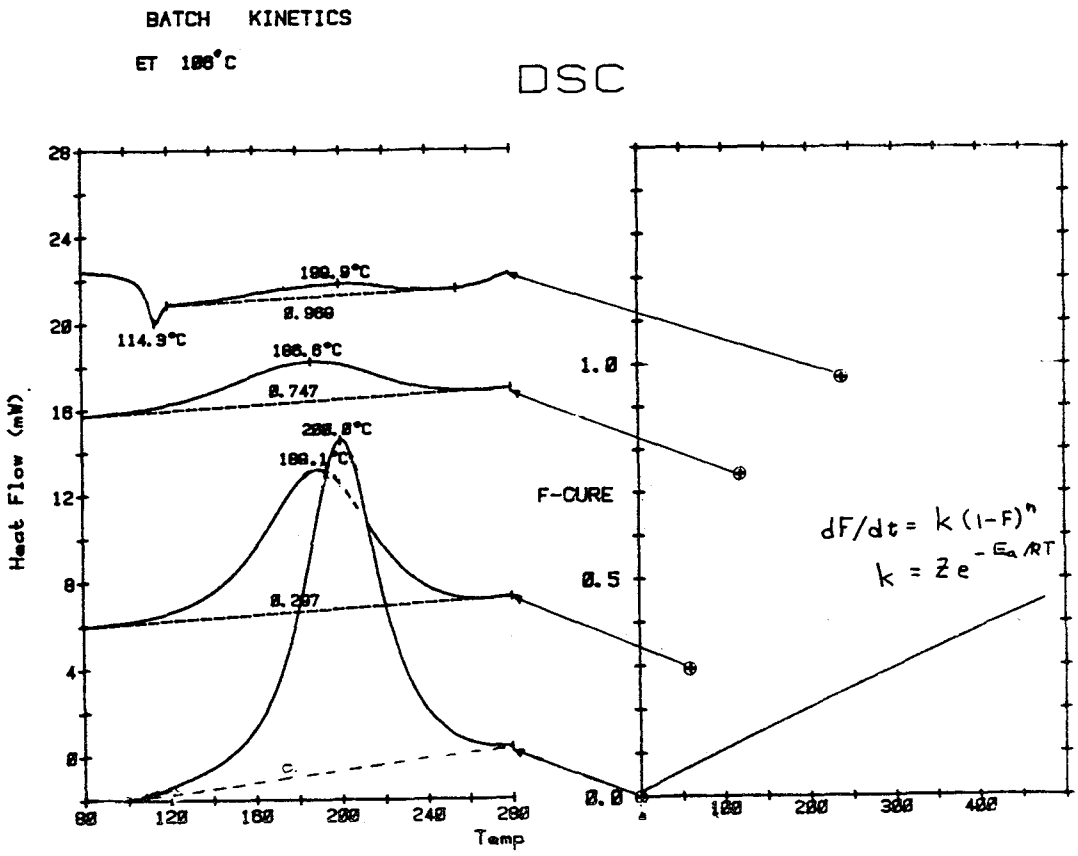


FIGURE 3.

A second method of analysis of a DSC scan is a least squares fit of the shape of the DSC exotherm to equation (1) (Borchardt-Daniels analysis[7]).

$$(1) \quad dF/dt = (1-F)^n \cdot Z \cdot \exp\left[-Ea/RT\right].$$

Here (equation 1), dF/dt is the normalized DSC scan measurement, F is fraction cure (conversion), Z is an Arrhenius frequency factor and Ea is the corresponding activation energy. The inset of Figure 3 shows the predicted F - time curve using this (1) analysis of the zero time DSC scan.

The assumption of the Borchardt-Daniels method is that no change in mechanism occurs during the cure reaction. The very large discrepancy between the predicted F - time and the pointwise batch cure is suggestive of an autocatalytic change in the mechanism as the reaction proceeds. The same implication is suggested by the decrease in the temperature of the maximum rate of the batch samples of Figure 3.

Figure 4 shows the variation in the predicted cure rates by the Borchardt-Daniels fits to each of the first three batch samples of Figure 3. As expected, the activation energy is reduced with the temperature of maximum heat flow. The reaction order (n) also changes significantly, further indications of autocatalysis during the first 75% in apparent conversion. A plausible explanation of the apparent autocatalysis is shown in Scheme II.

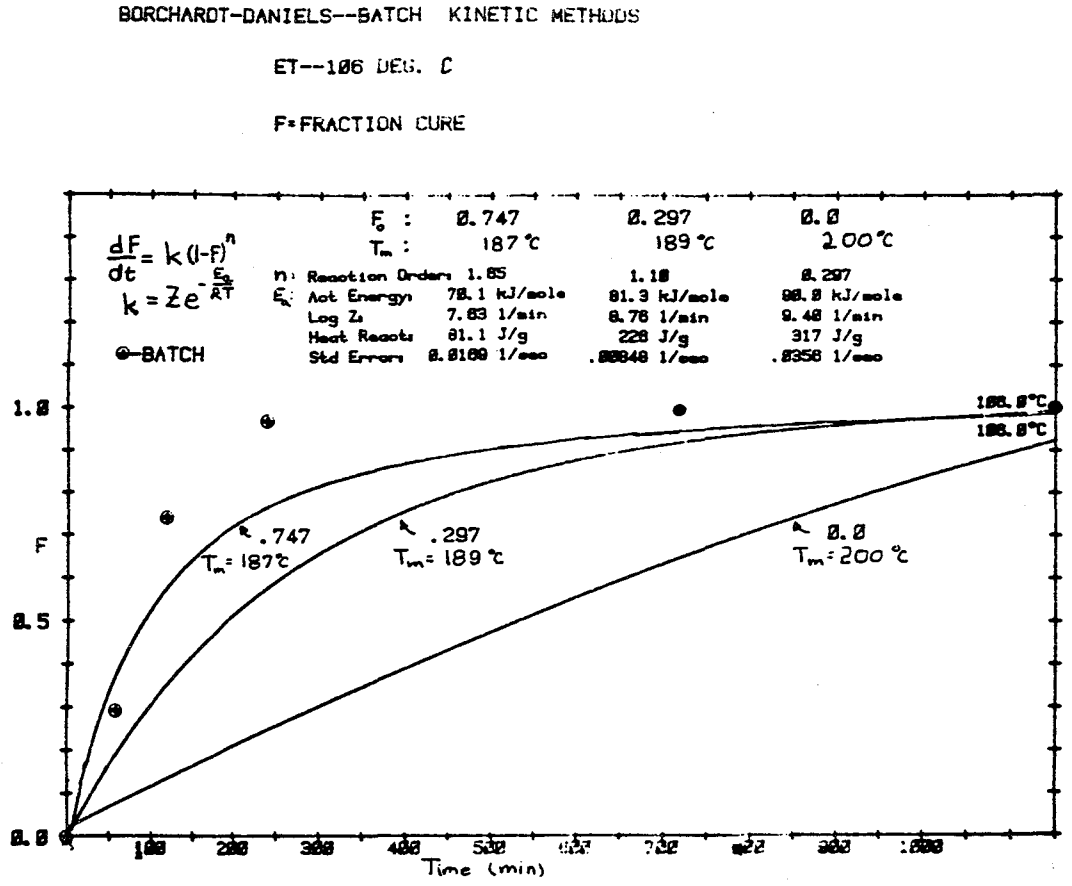


FIGURE 4.

Scheme II

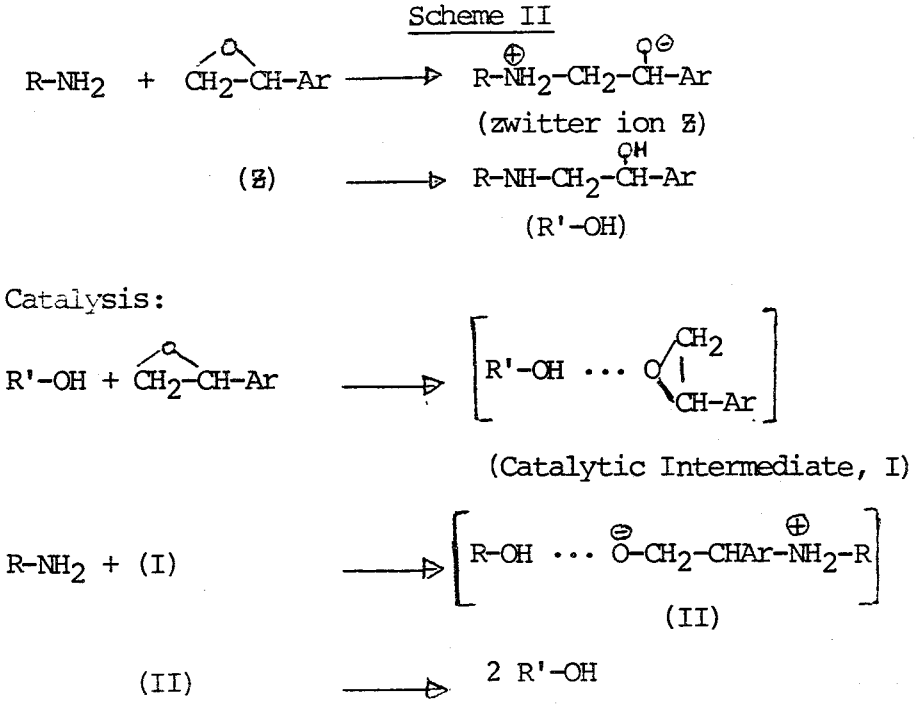


Figure 5 shows an example of the isothermal heat flow method for following the cure reaction. In this method, the kinetic sample is quickly brought to a temperature of appreciable reaction rate and held constant while the heat flow is recorded. The shape of the heat flow (reaction rate) curve at the top is indicative of an autocatalysis in that the reaction rate increases for a time before it begins to decrease. The lower portion shows the agreement between the batch data and the normalized integration of the peak at the top. As is evident, the agreement is good for the integrated curve at 107° and the batch data at 106°. The 101.7° curve gives a calibration of the sensitivity of this system to change in oven or isothermal temperature.

Figure 6 shows isothermal heat flow measurements which confirm the gross features of the catalysis that are depicted in Scheme II. The heat flow peak in the upper right is a portion of that shown in Figure 5. A system undergoing simple reaction should show a monotonous decrease in rate as starting material is consumed, like the peak shape shown for the ET system after pretreatment at 106° to 75% apparent cure (Figure 5, middle, right). The lower pair of curves of Figure 6 were obtained for a comparable sample of the ET system with 1.1 equivalents of triethylene glycol as an additive. The temperature of maximum rate of the diol sample is reduced to 186° without prior heat treatment and the form of the isothermal rate curve shows almost no increase after reaching reaction temperature (107°).

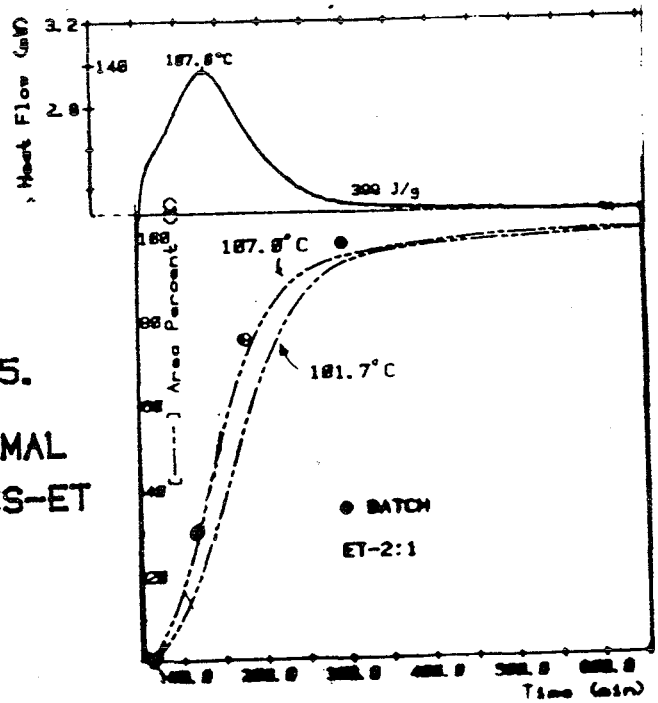


FIGURE 5.

ISOTHERMAL KINETICS-ET

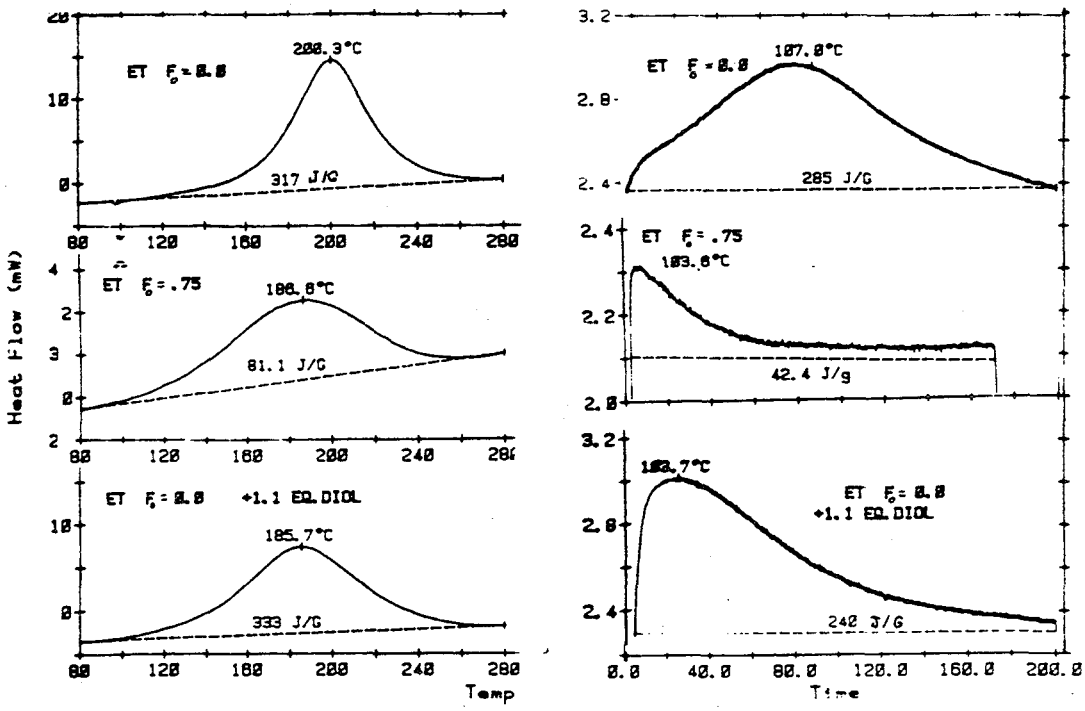


FIGURE 6. AUTOCATALYSIS

Figure 7 (middle portion) allows a comparison of the integrated rates of the three present systems (EM, ED and ET) at comparable temperatures (103-107° C). The upper and lower curves are similar to that for ET (Figure 5) and indicate autocatalysis is common to all three systems. These data, together with literature precedent[8], make it clear that all three of the present epoxy systems are subject to an autocatalytic stage like that depicted as Scheme II. The catalysis is turned off when the mobility of the subgroups in the resin matrix becomes slow. This accounts for the increase in the temperature of maximum rate during the last observable stages of the ~100° cure (e.g. Fig. 3).

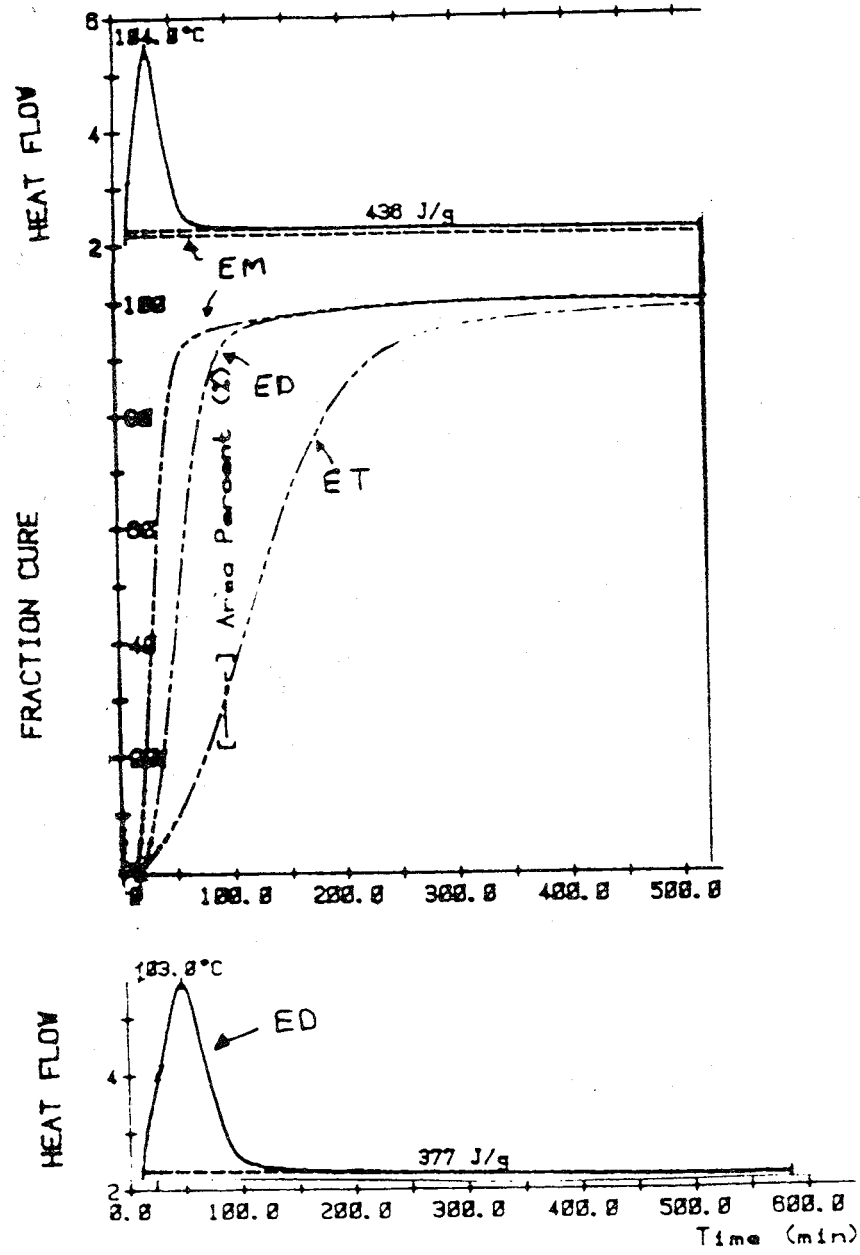


FIGURE 7.

Figure 8 shows the DSC curves of each of the three resin systems obtained from a 103-107° cure temperature and at relatively long cure times after the reaction exotherm appears to have ceased. The sharply peaked endothermic transitions are like melting transitions but are ascribed to enthalpy relaxation[9] which differs only in the nature of initial state. That is, these sharply peaked transitions do correspond to the change from an ordered to a disordered state but do not require crystallinity in the ordered state. For our purposes, the results of this figure suggest the first important point whereby tests responding to structure variation, separated from cure, could be made. The slowly changing nature of the cure process in this temperature range and the similarity in transition temperatures suggest that it should be possible to prepare specimens of the three structurally different resins with identical transition temperature. Tests of toughness of these specimens would at least be on resins of constant transition temperature and "complete" cure insofar as exothermal behavior is concerned. Results of such tests could be very interesting.

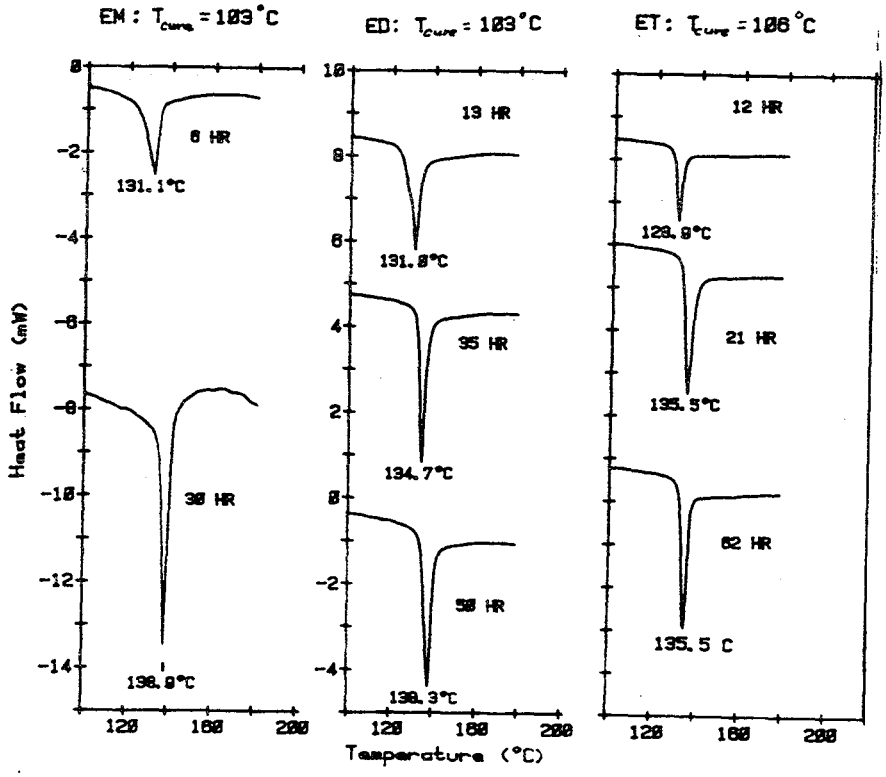


FIGURE 8.

Figure 9 is a composite of the types of measurements which were made on the EM resin system. It shows that the features already discussed for the ET system are present in the EM system as well. It shows, in addition, the effects of subjecting the resins to a second higher cure temperature after the cure at ca. 103° C is apparently complete. The isothermal method reveals an appreciable amount of additional heat flow at the higher second temperature which can be used to further quantify the systems.

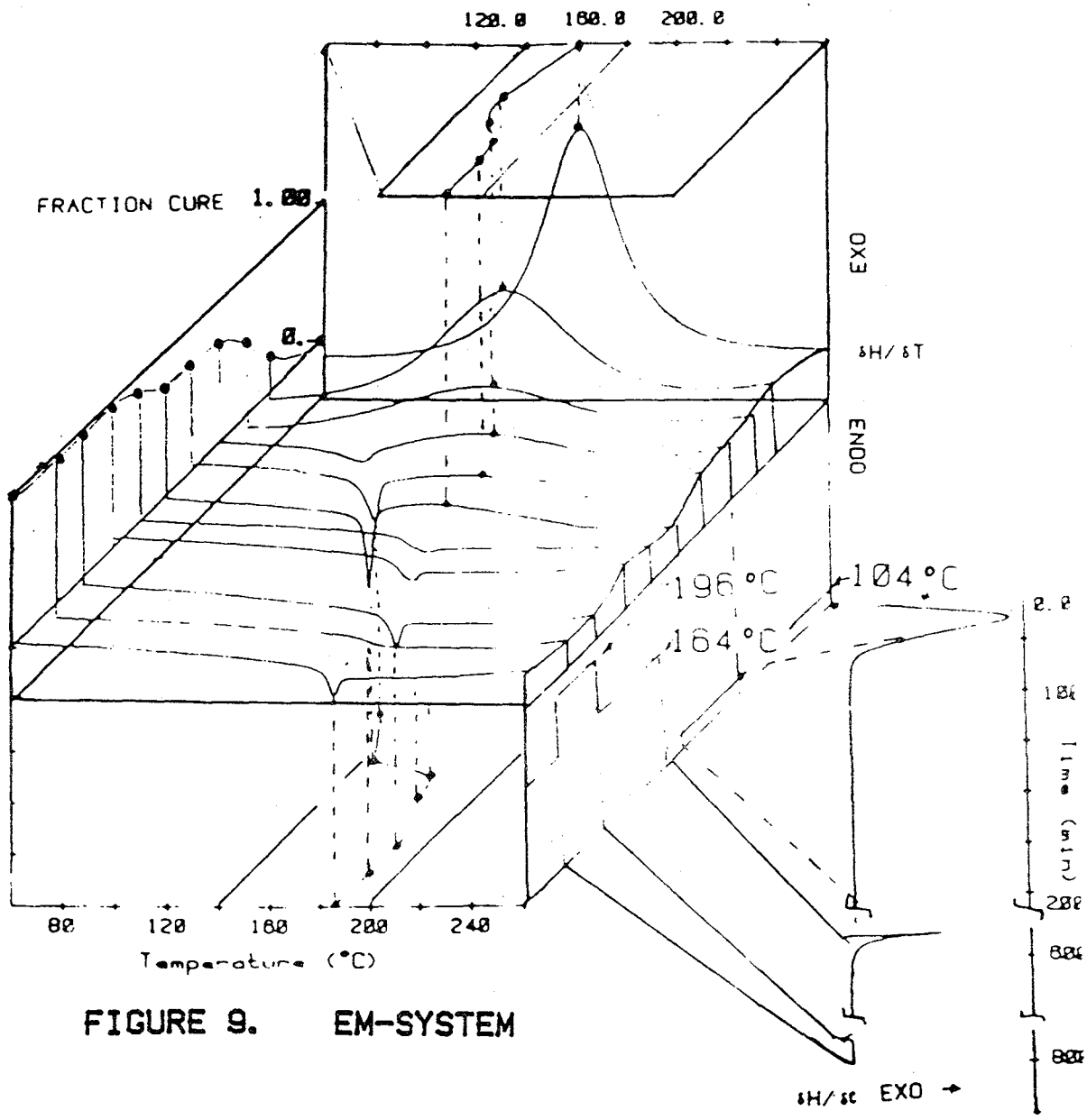


FIGURE 9. EM-SYSTEM

Figure 10 gives a more detailed view of the measureable heat flow for the EM system. The upper portions show the isothermal exotherm measured at temperatures of 103° C for the first 9 hours, 164.4° C for the next 200 minutes and 196.7° C for the final hour. The lower portion shows the integral of the 103° exotherm normalized for the first 9 hours and compared to batch samples at the same temperature. The agreement between the integral and batch methods is quite good. However, the observed exotherms at the higher temperatures clearly show that the batch method gives an apparent cure which is too high. The (+) points mark the F (fraction cure) values after correction for the higher temperature exotherm measurements. The dotted curves below the 164.4° isotherm and the 196.7° isotherm show the integrals of the corresponding exotherms after approximate normalization to the total measureable heat evolution. Corresponding determinations with the ET and ED systems show corresponding exotherms at 150-164° but no measureable heat evolution at a higher temperature (175°).

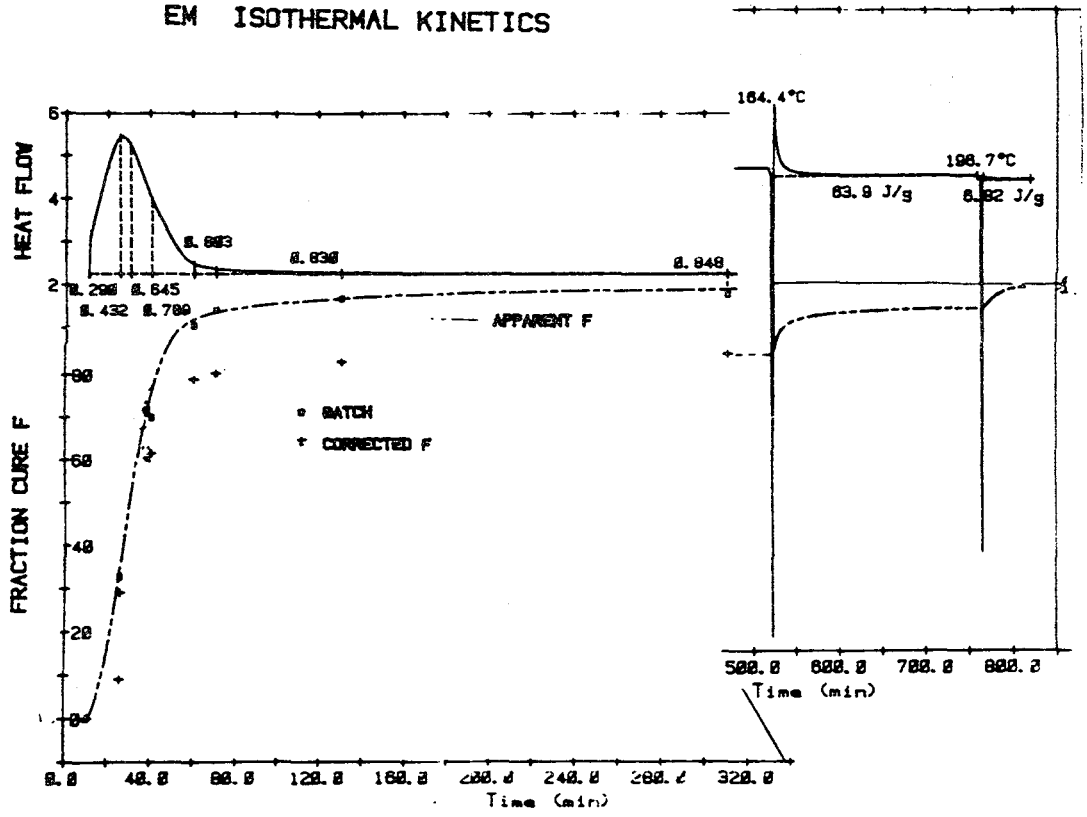
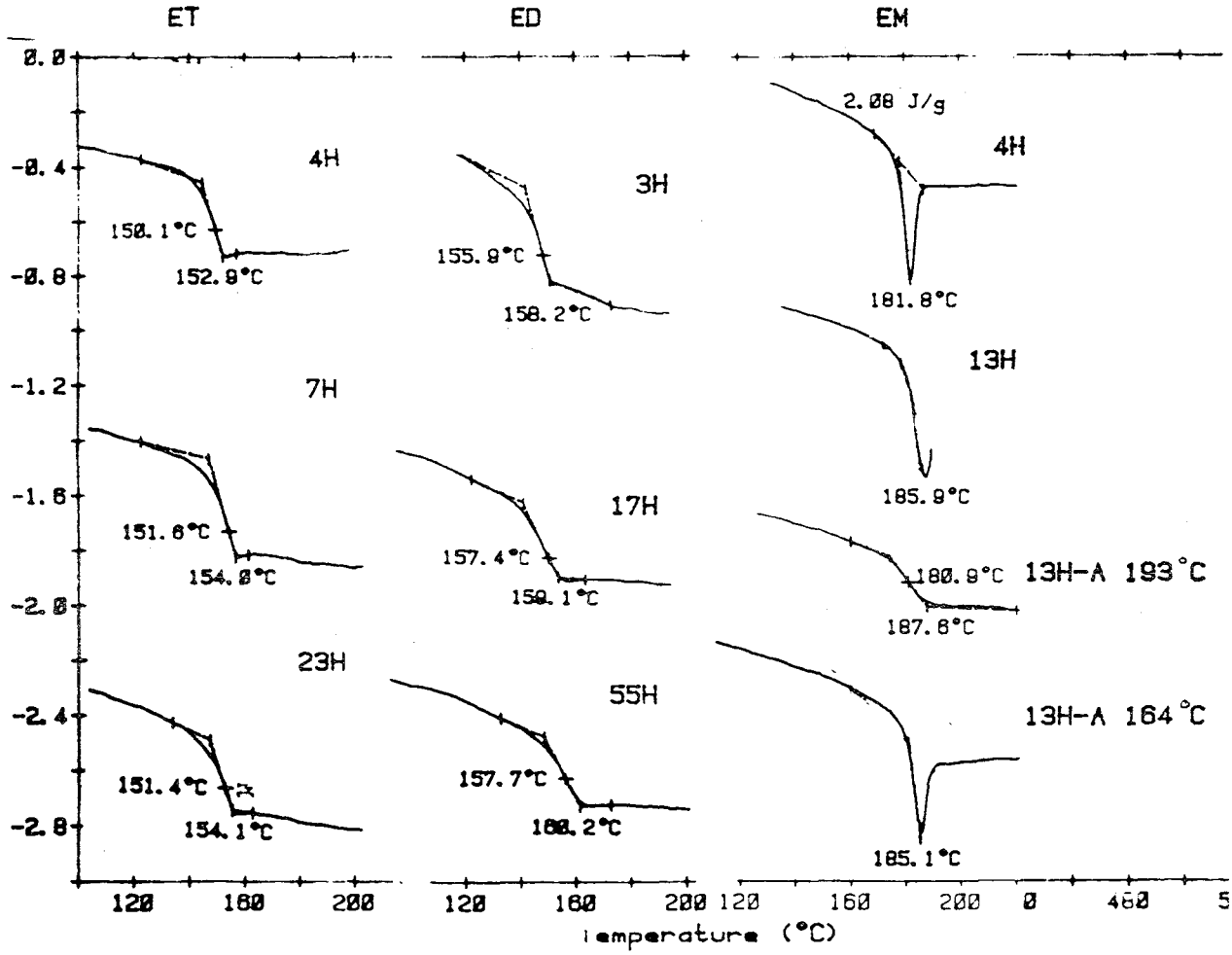


FIGURE 10.

Figure 11 shows the glass transitions of the resins obtained from the three systems after a long-term cure at 164° C. The ED and ET systems show an apparently constant glass transition temperature after 7 hours and 17 hours respectively. The ultimate glass transition temperatures (151° C and 157° respectively) for these two sytems are well below the final cure temperature (164°). The results for the ED and ET systems indicate that it should be possible to prepare specimens of these two matrix resins at "complete" cure as defined by the absence of a measureable exotherm at a higher temperature and no change in glass temperature on extended reaction time at the final (164°) cure temperature.

FIGURE 11.

FINAL CURE 164°C



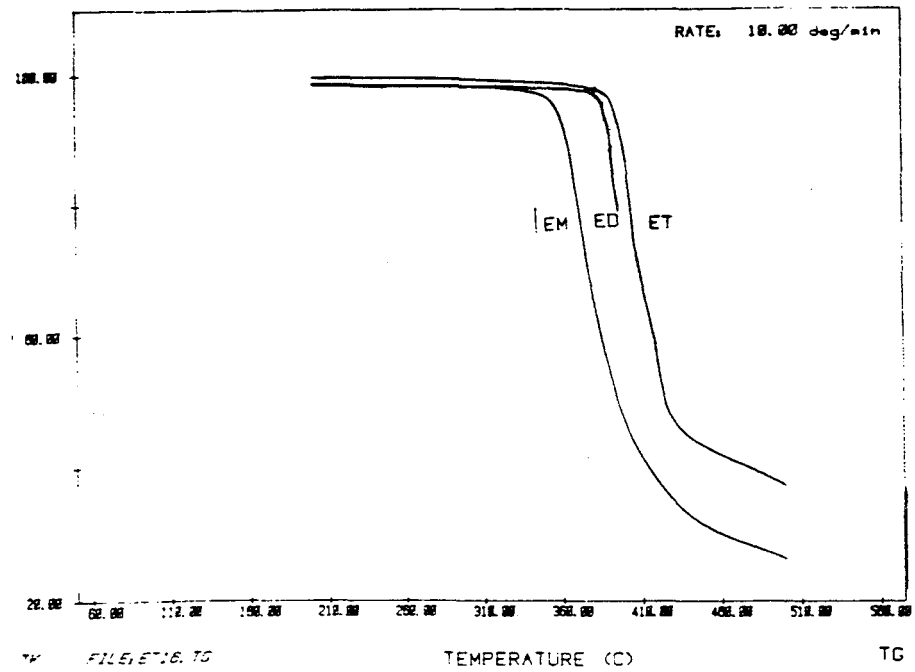
The results of Figure 11 for the EM system are more involved. The transition temperature in this case is above the final cure temperature (164° C). The sharply peaked exotherm is again indicative of the enthalpy relaxation[9] observed for all three systems at the lower cure temperature (103-107°, Figure 8). The curve marked 13H-A is from the identical sample that gave the 13H curve except that the sample was annealed for 10 minutes at 193° and cooled quickly. The curve marked 13H-A2 was treated in an identical fashion to 13H-A and then re-annealed at 164° for 24 hours. These observations are a demonstration of the enthalpy relaxation phenomenon which is generally known[9] to be a characteristic of such epoxy systems.

The enthalpy relaxation phenomenon is especially important for the present purposes because it demonstrates that these systems exist in two states which may have different mechanical properties. Furthermore the present results demonstrate the possibility of producing specimens at comparably "complete" cures in each of the two (ordered and disordered) states for all three resin systems.

Finally, Figure 12 shows the results of thermogravimetric analysis of each of the three resin systems in air with a heat rate of 10°/min. These results show a modest increase in thermal stability on going from the monoaryl (M) to the triaryl (T) system. This observation is mildly surprising since thermal degradation would be expected to begin in the side chain groups derived from the epoxy unit which was held constant in these studies. The increased thermal stability might offer advantages for rocket applications.

FIGURE 12.

THERMOGRAVIMETRIC ANALYSIS



IV. Discussion

The present results provide the suggestion of three sets of conditions under which toughness test specimens might be prepared such that the diamine structural variation effects can be isolated. They are, first, after complete apparent cure at a temperature near 103° C. These specimens could possibly be made so that they show identical enthalpy relaxation peak temperatures. Second, the three-resin system could be prepared at complete cure at a final-cure temperature near 164° C. Complete cure in this case would be defined as no additional exotherm at higher temperature and no change in observed glass transition temperature on extended cure times. Finally, the latter specimens could be induced to populate their more stable states by annealing them for 24 hours at temperatures 10° below their glass transition temperatures. These conditions are summarized in Table II.

Table II

Conditions for Preparation of Comparable Specimens of EM, ED and ET Matrix Resins

Resin	T_1 °C	Time(hr.)	T_g °C	T_2 °C	Time(hr.)	T_g °C	T_A °C[c]
EM	103	30	138.9	164	23± 6	180[a]	170
ED	103	50	138.3	164	23± 6	157[b]	147
ET	106	40	135.5	164	17	151[b]	141

a) Specimens must be annealed at 190° for 10-15 min. to be free of enthalpy relaxation.

b) Specimens without enthalpy relaxation.

c) Temperature of annealment for 24 hours producing the stable states of the resins that will show the characteristic sharp enthalpy relaxation transition peak.

It should be mentioned in closing that the present results apply to the milligram scale of sampling. Technical problems remain in translating these results to the 100-gram scale necessary for mechanical specimens. Some of these problems relate to the fact that all of the materials here are solids and must be melted in order to obtain the resin mixture. Some attention to cure rates in the 60-70° C range is needed. Also, the degree of overheating by the exothermic reaction in a large sample requires some attention.

V. References

1. Martin T. Goosey, "Plastics for Electronics", Elsevier Applied Science Publishers, New York, 1985, pp. 99-135.

2. The chemical nomenclature of E, M, D and T is problematic, being consistent in neither time nor convention. The epoxide (E) is usually referred to as the bis-glycidyl diether of bisphenol A. Diamine M is m-phenylenediamine or 1,3-benzenediamine. Diamine D is 3-[(3-aminophenyl)methyl] benzenamine or 3,3'-diaminodiphenylmethane. Diamine T is listed in Chemical Abstracts under 3,3'-[1,3-phenylenebis(oxy)] bis-benzenamine.

3. Syst. No. 1756, "Beilsteins Handbuch Der Organishen Chemie", Edwards Brothers Inc., Ann Arbor, Michigan, 1944.

4. V. Bell, B. L. Stump and H. Gager, J. Polymer Sci. (Polymer Chem. Sec.), 14, 2275-2292 (1976).

5. S. C. Misra, J. A. Manson and L. H. Sperling in "Epoxy Resin Chemistry", ACS Symposium Series 114, R. S. Bauer Ed., American Chemical Society, Washington, D.C., 1979, pp. 137-155.

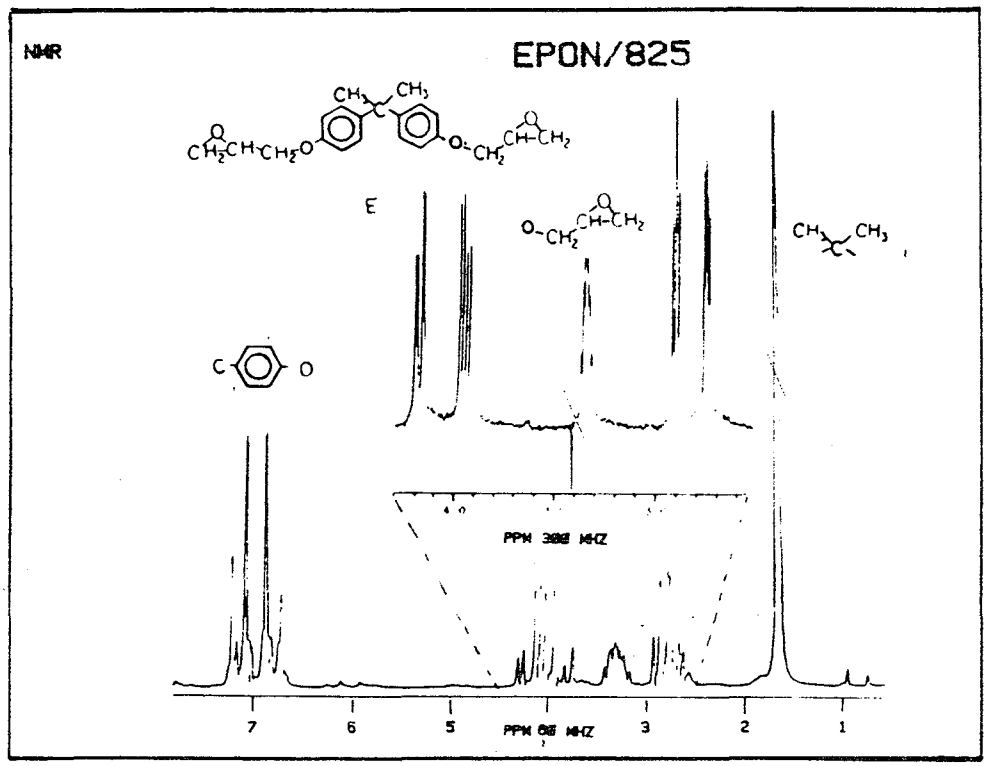
6. R. Riesen and H. Sommerauer, Amer. Lab., Jan. 1983, pp. 30-37.

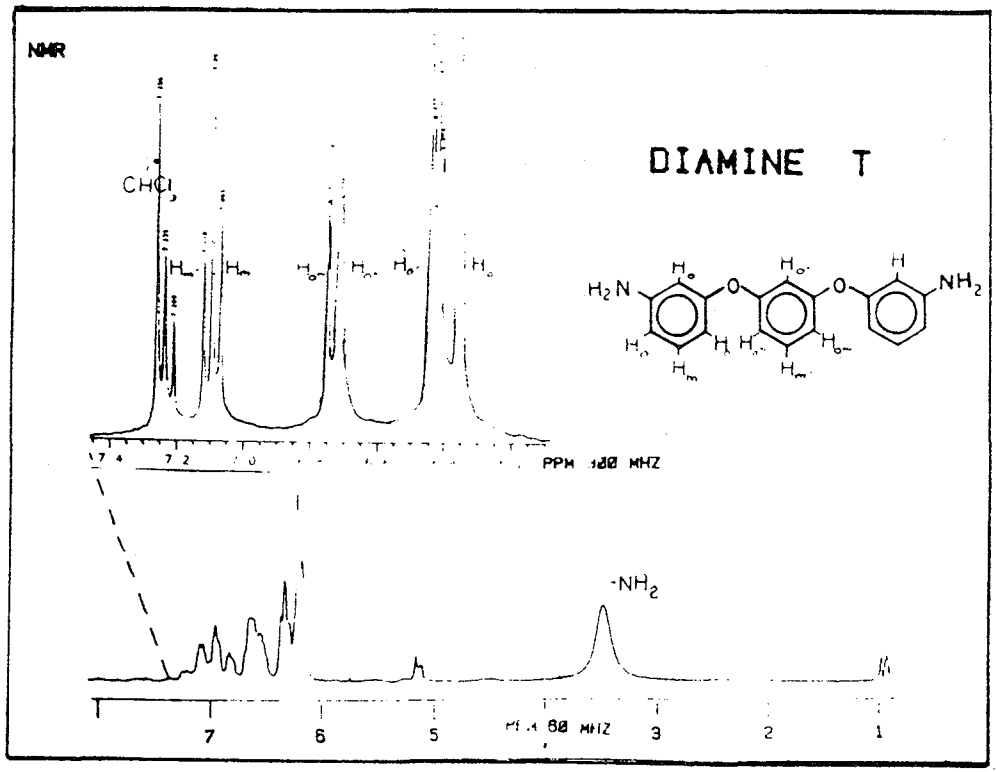
7. H. F. Borchardt and F. Daniels, J. Am. Chem. Soc., 79, 41 (1957). More sophisticated methods are forthcoming. H. Hsich, R. Zurn and R. Ambrose, Polymer, 20, 203(1979); C. Douglas and D. Roylance in "Chemorheology of Thermosetting Polymers", ACS Symposium Series 227, Clayton May Ed., American Chemical Society, Washington D.C., 1985, pp. 263-279.

8. S. Sourour and M. R. Kamal, Thermochimica Acta, 14, 41 (1976); T. Oleese and O. Spelta, J. Polymer Sci., Symposium No. 53, 113-126 (1975); J. Barton, Die Makromolekulare Chemie, 171, 247 (1973); R. A. Fava, Polymer, 9, 137 (1968); L. Shechter, J. Wynstra and R. Kurlejy, Ind. and Eng. Chem., 48, 94 (1956). For a dissenting opinion see M. Acitelli, R. B. Prime and E. Sacher, Polymer, 12, 335 (1971).

9. R. Riessen and Horst Wyden, Mettler Application No. 3408, 1984; S. Petrie, J. Macromol. Sci.-Phys., B12(2), 225-247 (1976).

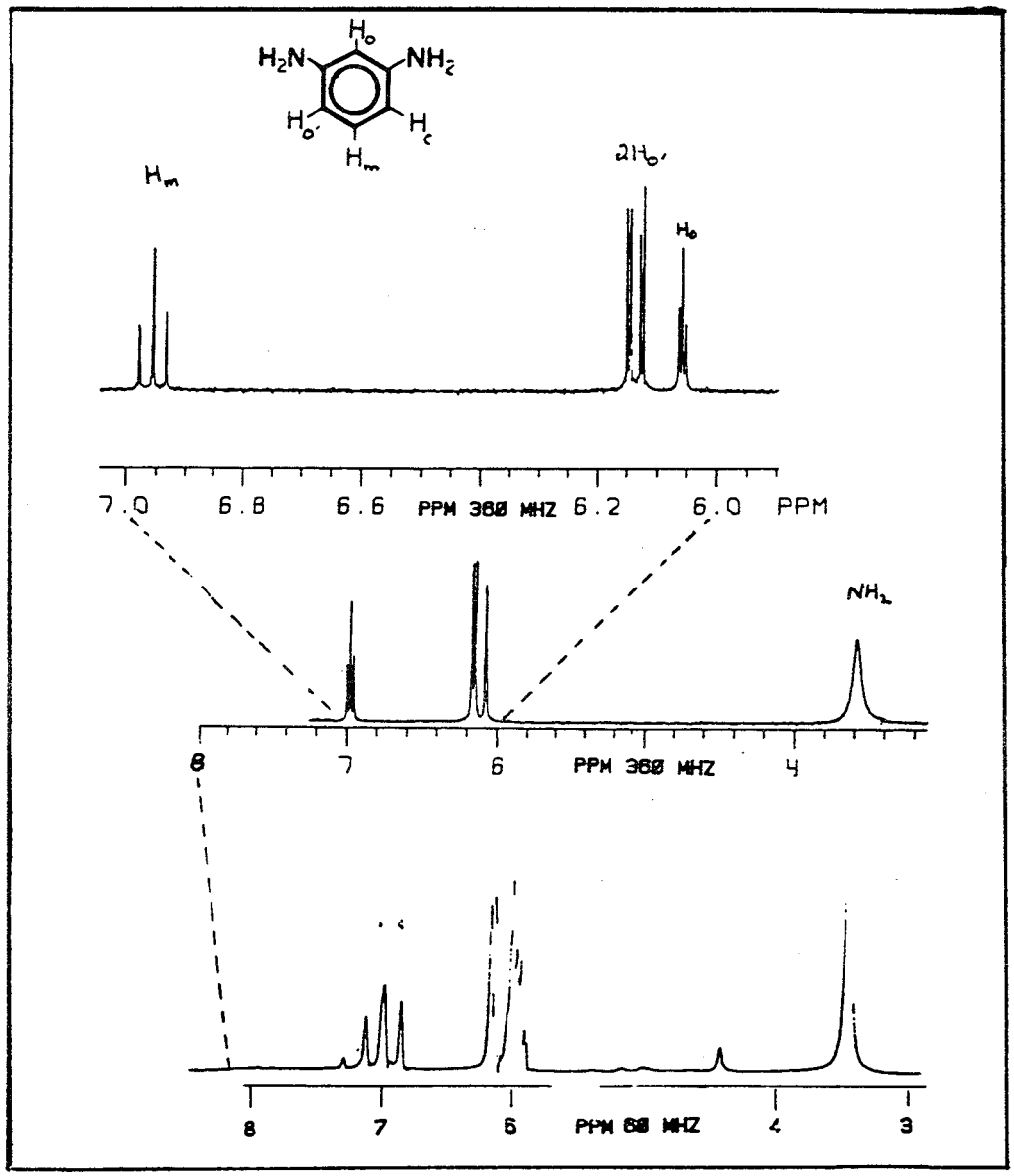
APPENDIX I PROTON MAGNETIC RESONANCE SPECTRA





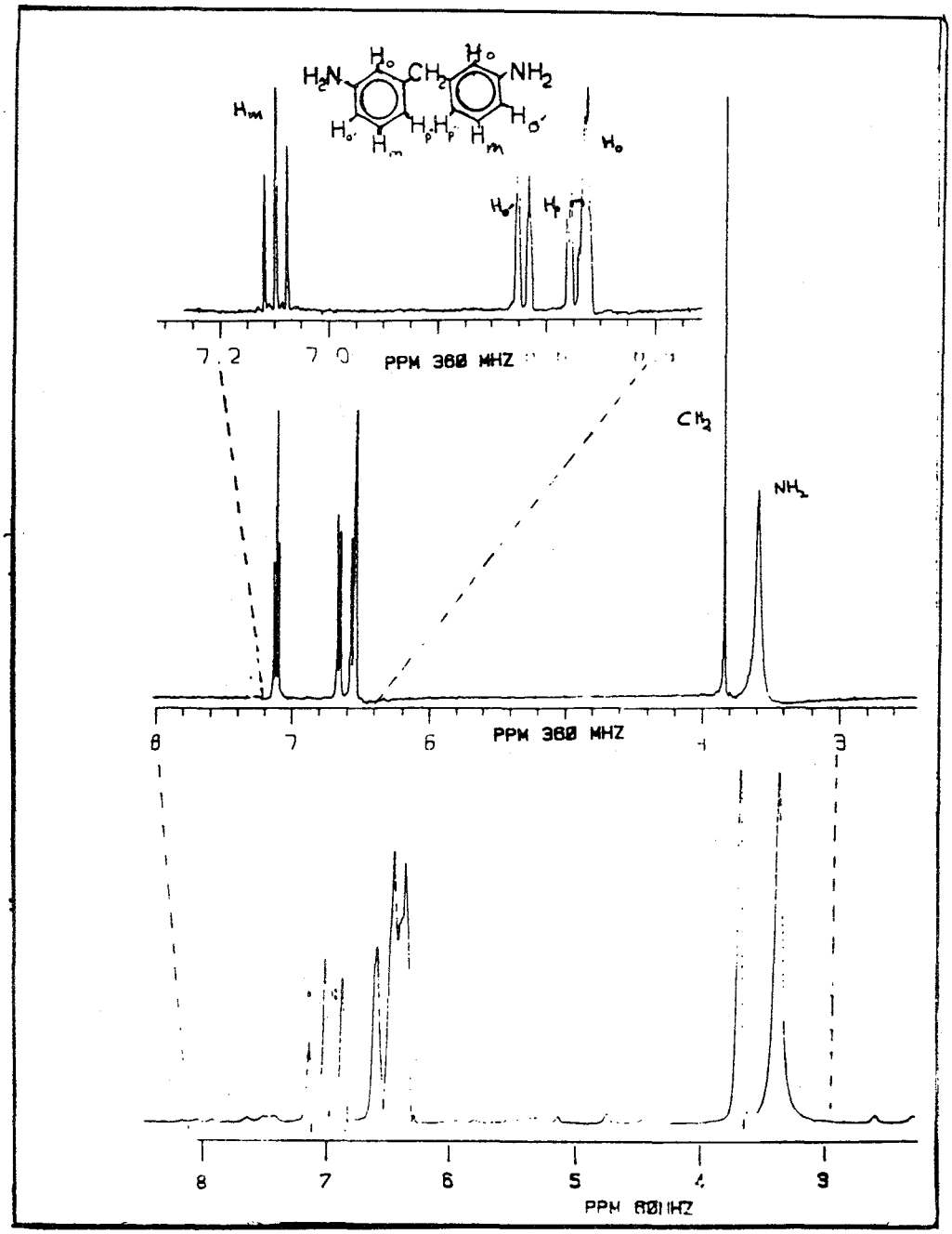
PROTON MAGNETIC RESONANCE SPECTRA

DIAMINE M



PROTON MAGNETIC RESONANCE SPECTRA

DIAMINE D



1985

NASA/ASEE SUMMER FACULTY RESEARCH FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER

THE UNIVERSITY OF ALABAMA

ANALYTICAL DETERMINATION OF SPACE STATION RESPONSE TO CREW MOTION AND DESIGN OF SUSPENSION SYSTEM FOR MICROGRAVITY EXPERIMENTS

Prepared By:	Frank C. Liu, Ph.D., PE
Academic Rank:	Professor
University and Department:	The University of Alabama in Huntsville Department of Mechanical Engineering
NASA/MSFC	
Division:	Mechanical
Branch:	Mechanical System Development
MSFC Counterpart:	Harold Smyly
Date:	August 1, 1985
Contract No.:	NGT-01-008-021 The University of Alabama in Huntsville

ANALYTICAL DETERMINATION OF SPACE STATION RESPONSE TO CREW MOTION AND DESIGN OF SUSPENSION SYSTEM FOR MICROGRAVITY EXPERIMENTS

by

Frank C. Liu

Professor of Mechanical Engineering
The University of Alabama in Huntsville
Huntsville, Alabama

ABSTRACT

The objective of this investigation is to make analytical determination of the acceleration produced by crew motion in an orbiting space station and define design parameters for the suspension system of microgravity experiments. A simple structural model for simulation of the IOC space station is proposed. Mathematical formulation of this model provides the engineers a simple and direct tool for designing an effective suspension system.

ACKNOWLEDGMENT

It has been an exciting experience for me to work on this interesting project. I am grateful to Mr. Doug Lamb, Chief of Mechanical Division, Structures and Propulsion Laboratory, and his staff Mr. Charles Cornelius, Mr. Harold Smyly, and Mr. Charles Miller for giving me an enjoyable summer to carry out this investigation. I express appreciation to Mr. Mike Moore, my office mate, for making all the computer graphs. Special thanks are due to Ms. Margarete Landers for typing this report.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. Introduction

Some microgravity experiments to be performed on the IOC space station require an environment in which the acceleration level must be below 10^{-5} g. Among the various sources of disturbances, crew motion is the severe one which can produce acceleration to a magnitude of 10^{-4} g. The objectives of this investigation are to define design characteristics for the suspension system and to find effective means for isolation of the experiment packages from disturbance. To achieve these goals, the following have been accomplished.

1. Analytical formulation of acceleration response of IOC to crew motion

Assumed modes method and modal transformation are used to obtain mathematical solution for the normal coordinates to an input function due to crew motion. Acceleration response of any point in the space station can be analytically formulated.

2. Structural modeling for IOC space station

The finite element model for IOC space station, shown in Figure 1, used by JFC [1] for vibration analysis generates a wide range and closely spaced spectrum of vibration frequencies (see Table 4.3.3.3-3, reference 1). It makes it extremely difficult for mechanical design engineers to identify a particular mode which has dominant effect on the experiment package. A simple structural model which is made of a rigid main body (extension of keel frame which supports all the massive modules), a cantilever beam (the keel frame), and a transverse cantilever beam (the solar boom) is proposed for simulation of the IOC. A complete analytical formulation for this model is made. This formulation provides the design engineers a simple and direct tool for determination of acceleration and disturbing frequencies acting on the experiment package. Computations can be carried out by using a pocket calculator. This model can be easily improved by increasing the degrees-of-freedom. A simple computer program will do this work.

3. Derivation of design characteristics for suspension system

Consider that a microgravity experiment package is supported by a spring and a mass-spring-damper vibration absorber is attached to it. The acceleration magnification factor and the ratio of acceleration response of package to acceleration input of the support is then derived. The plots of this factor versus frequency ratio (IOC to mass-spring) provide engineers design parameters of the suspension system.

4. Significant findings for effective isolation of disturbance

The microgravity laboratory is 23.3/83.2 feet from the center of mass of IOC with/without the orbiter attached. It is shown by

numerical study that the acceleration disturbance is greatly reduced when the orbiter is not present (i.e., microgravity lab is actually 59.9 feet farther away from the C.M. of IOC). This shows that the most favorable location beside the C.M. of IOC is the node of the fundamental elastic vibration mode of the IOC.

2. Analytical Formulation

2.1 Modal analysis of a structure system

A brief presentation of modal analysis of a structure is given here. Let P be the coordinate of a generic mass point P in a structure and u(P,t) be its displacement which is expressed in the form [2]

$$u(P,t) = \sum_{i=1}^{N} \psi_i(P) q_i(t) \tag{2.1}$$

where $\psi_i(P)$ is an admissible function and $q_i(t)$ is the ith generalized coordinate of a set N. The equation of motion of the structure system in matrix form is

$$[m]\ \{\ddot{q}\} + [k]\ \{q\} = \{Q\} \tag{2.2}$$

where [m], [k], and {q} are the generalized mass, stiffness, and force matrices, respectively.

The kinetic and bending strain energies of the system are formulated from the following integrals:

$$T = \tfrac{1}{2}\int_B m_P[\dot{u}(P,t)]^2\, dP = \tfrac{1}{2}\sum\int_B m_P\psi_i^2(P)dP\dot{q}_i^2 \tag{2.3}$$

$$V = \tfrac{1}{2}\int_B EI[u''(P,t)]^2 dP = \tfrac{1}{2}\sum\int_B m_P[\psi_i''(P)]^2\, dP q_i^2 \tag{2.4}$$

where m_P is the mass density at P and EI is the bending stiffness of the structure at P, the prime denotes partial differentiation with respect to spatial coordinates and the symbol B with the integral sign means integration over the entire body B. From the above integrals, the elements of the [m] and [k] matrices are obtained respectively,

$$m_{ij} = \frac{\partial^2 T}{\partial\dot{q}_i\partial\dot{q}_j} = \int_B m_P\psi_i(P)\psi_j(P)dP \tag{2.5}$$

$$k_{ij} = \frac{\partial^2 V}{\partial q_i \partial q_j} = \int_B EI\psi_i''(P)\psi_j''(P)dP \tag{2.6}$$

Denoting the force acting at P in the direction of u(P,t) by $f_P(t)$, then the virtual work is

$$\delta W = \int_B f_P(t)\delta u dP = \sum \int_B f_P \psi_i(P)\delta q_i dP \tag{2.7}$$

It follows that

$$Q_i = \frac{\partial W}{\partial q_i} = \int_B f_P \psi_i(P)dP \tag{2.8}$$

Next, the natural frequencies of the system are determined from the determinant

$$\left|[k] - \omega^2[m]\right| = 0 \tag{2.9}$$

and the eigenvectors (or modal columns) $\{\phi_i\}$ are the solution of the matrix equations,

$$([k] - \omega_i^2[m])\ \{\phi_i\} = \{0\} \qquad i = 1,2,\ldots,N \tag{2.10}$$

The normal coordinates $\{\eta\}$ and the generalized coordinates are related by the transformation

$$\{q\} = [\Phi]\ \{\eta\} \tag{2.11}$$

where the modal matrix

$$[\Phi] = [\{\phi_1\}, \{\phi_2\}, \ldots, \{\phi_N\}]$$

Applying Eq. (2.11) to Eq. (2.2) and making use of the property of generalized orthogonality of the eigenvector with respect to [m] and [k], the equation of motion in normal coordinates is

$$\ddot{\eta}_j + \omega_j^2\ \eta_j = N_j(t)/M_{jj} \qquad j = 1,2,\ldots,N \tag{2.12}$$

where

$$M_{jj} = \{\phi_j\}^T[m]\{\phi_j\} \tag{2.12a}$$

$$\omega_j^2 = \{\phi_j\}^T[k]\{\phi_j\}/M_{jj} \tag{2.12b}$$

$$N_j(t) = \{\phi_j\}^T\{Q\} \quad . \tag{2.12c}$$

2.2 Structure response to disturbing force

Consider that a disturbing force $f_a(t)$ is applied to the structure at P_a, from Eqs. (2.8) and (2.12c) one obtains

$$N_j(t) = \sum_{k=1} \phi_{kj}\psi_k(P_a)f_a(t) \tag{2.13}$$

where ϕ_{kj} is the kth element of the jth eigenvector. The response at point P_e in the structure, $u(P_e,t)$, can be obtained by using Eqs. (2.1), (2.11), and (2.13). It results in

$$u(P_e,t) = \sum_i \psi_i(P_e)q_i(t) = \sum_i \sum_j \psi_i(P_e)\phi_{ij}\eta_j(t)$$

$$= \sum_j \left[\left(\sum_i \phi_{ij}\psi_i(P_e)\right)\left(\sum_k \phi_{kj}\psi_k(P_a)\right)\right] \bar{\eta}_j(t)/M_{jj} \tag{2.14}$$

where $\bar{\eta}_j(t)$ denotes the solution of the differential equation,

$$\ddot{\eta}_j + \omega_j^2\eta_j = f_a(t) \qquad \text{with} \qquad \eta_j(0) = \dot{\eta}_j(0) = 0 \quad . \tag{2.15}$$

2.3 Model of crew motion and response function

The motion of an astronaut inside a space module is started by pushing one wall and motion is stopped by pushing the opposite wall. A simple mathematical model is suggested [1] as shown in Figure 2.

$$f_a(t) = \begin{cases} (f_o/t_1)t & 0 < t \leq t_1 \\ 0 & t_1 < t \leq t_{13} \\ (f_o/t_1)(t - t_{14}) & t_{13} < t \leq t_{14} \\ 0 & t_{14} < t \end{cases} \tag{2.16}$$

The magnitude of f_o is 25 lbs and t_1, t_{13}, and t_{14} are 1, 13, and 14 seconds, respectively. These notations for time will be kept through the formulation for the sake that one may wish to change their magnitudes and secondly, that the equation involved will have appropriate units (use $\sin\omega t_{13}$ rather than $\sin 13\omega$ where ω is in rad/s). The solution of equation (2.15) is readily obtained. Denoting that

$$\eta(t) = (f_o/\omega^3 t_1)U(t) \tag{2.17}$$

the dimensionless displacement function U(t) is given by

$0 < t \leq t_1$ $\quad U(t) = \omega t - \sin\omega t$ (2.17a)

$t_1 < t \leq t_{13}$ $\quad U(t) = -\sin\omega t + \sin\omega(t - t_1) + \omega t_1 \cos\omega(t - t_1)$ (2.17b)

$t_{13} < t \leq t_{14}$ $\quad U(t) = \omega(t - t_{14}) - \sin\omega t + \sin\omega(t - t_1) - \sin\omega(t - t_{13}) + \omega t_1[\cos\omega(t - t_1) + \cos\omega(t - t_{13})]$ (2.17c)

$t_{14} < t$ $\quad U(t) = -\sin\omega t + \sin\omega(t - t_1) - \sin\omega(t - t_{13}) + \sin\omega(t - t_{14}) + \omega t_1[\cos\omega(t - t_1) + \cos\omega(t - t_{13})]$ (2.17d)

The function U(t) is continuous and has continuous first derivative (velocity); its second derivative is only piecewise continuous. Denoting

$$\ddot{\eta}(t) = (f_o/\omega t_1)A(t) \tag{2.18}$$

and differentiating Eq. (2.17) twice, one obtains the dimensionless acceleration A(t),

$$0 < t \leq t_1 \qquad A(t) = \sin\omega t \tag{2.19a}$$

$$t_1 < t \leq t_{13} \qquad A(t) = \sin\omega t - \sin\omega(t - t_1) - \omega t_1 \cos\omega(t - t_1) \tag{2.19b}$$

$$t_{13} < t \leq t_{14} \qquad A(t) = \sin\omega t - \sin\omega(t - t_1) + \sin\omega(t - t_{13}) - \omega t_1 [\cos\omega(t - t_1) + \cos\omega(t - t_{13})] \tag{2.19c}$$

$$t_{14} < t \qquad A(t) = \sin\omega t - \sin\omega(t - t_1) + \sin\omega(t - t_{13}) - \sin\omega(t - t_{14}) - \omega t_1 [\cos\omega(t - t_1) + \cos\omega(t - t_{13})] \tag{2.19d}$$

The functions of U(t) and A(t) are plotted versus time for frequencies ranging from 0.1 to 0.4 as shown in Figures 3 and 4 respectively. Now, the displacement and acceleration response of point P_e in the space station can be expressed in the form

$$u(P_e,t) = \sum_j \left(\sum_i \phi_{ij}\psi_i(P_e) \right) \left(\sum_k \phi_{kj}\psi_k(P_a) \right) U_j(t)(f_o/\omega_j{}^3 t_1 M_{jj}) \tag{2.20a}$$

$$a(P_e,t) = \sum_j \left(\sum_i \phi_{ij}\psi_i(P_e) \right) \left(\sum_k \phi_{kj}\psi_k(P_a) \right) A_j(t)(f_o/\omega_j t_1 M_{jj}) \tag{2.20b}$$

The subscript "j" with U and A denotes that these functions are corresponding to $\omega = \omega_j$.

2.4 Modeling of IOC space station

The structure of the IOC space station may be treated as a structural system having three elements. The main body is a frame structure which supports all the massive members, the vertical and horizontal HAB modules, the vertical and horizontal LAB modules, logistic and common modules, and above all, the orbiter. Attached to the main body is a keel frame structure 296 feet in length which supports an antenna system at its other end and a transverse boom at a distance 165.5 feet from the main body. The third member of the system is a transverse boom which is a frame structure 264 feet long. Its main purpose is to carry eight solar arrays and power system radiators.

A simple structural model for the IOC proposed consists of a rigid main body which supports a cantilever beam (keel frame) and a transverse cantilever beam (solar boom) mounted on the keel. As shown in Table 1 in Appendix A, the modules, equipment, solar arrays, fuel tanks, orbiter, etc., are treated as concentrated masses. However, the rotational moments of inertia of the orbiter and solar arrays must be included in forming the mass matrix.

Motions of the IOC in X-Z and Y-Z planes will be treated separately. In each plane the proposed IOC has four degrees-of-freedom, namely, rigid-body translation, rigid-body rotation, and one bending mode for each cantilever beam. Thus, the corresponding admissible functions are

$$\psi_1(P) = 1 \tag{2.21a}$$

$$\psi_2(P) = \bar{z} \tag{2.21b}$$

$$\psi_3(P) = \bar{z}^2 - \bar{z}^3/3 \tag{2.21c}$$

$$\psi_4(P) = \bar{y}^2 - \bar{y}^3/3 \tag{2.21d}$$

where $\bar{z} = z/\ell_k$ and $\bar{y} = y/\ell_s$ in which ℓ_k and ℓ_s are the length of the keel (296 ft) and solar boom (132 ft), respectively.

2.5 Modal analysis of IOC model

Based on Eq. (2.21) the mass and stiffness matrices for the model are formulated in Appendices A and B. The matrix equation of free motion of the IOC is

$$\begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \\ m_{41} & m_{42} & m_{43} & m_{44} \end{bmatrix} \begin{bmatrix} \ddot{q}_1 \\ \ddot{q}_2 \\ \ddot{q}_3 \\ \ddot{q}_4 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & k_{22} & 0 & 0 \\ 0 & 0 & k_{33} & 0 \\ 0 & 0 & 0 & k_{44} \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{2.22}$$

One may eliminate the rigid-body translation q_1 from the system by solving q_1 in terms of the rest of q's from the first equation of Eq. (2.22) and substituting it into the remaining equations. A further simplification can be made by disregarding the small coupling effect of the rigid-body rotation q_2 and the elastic modes q_3 and q_4 due to gravity gradient torque. Thus, one may eliminate both q_1 and q_2 from the system and obtain the following:

$$\begin{bmatrix} q_1 \\ q_2 \end{bmatrix} = [T] \begin{bmatrix} q_3 \\ q_4 \end{bmatrix} \tag{2.23}$$

$$\begin{bmatrix} m_{33} & m_{34} \\ m_{43} & m_{44} \end{bmatrix} \begin{bmatrix} \ddot{q}_3 \\ \ddot{q}_4 \end{bmatrix} + \begin{bmatrix} k_{33} & 0 \\ 0 & k_{44} \end{bmatrix} \begin{bmatrix} q_3 \\ q_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{2.24}$$

where

$$[T] = -\begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix}^{-1} \begin{bmatrix} m_{13} & m_{14} \\ m_{23} & m_{24} \end{bmatrix}$$

$$\begin{bmatrix} \bar{m}_{33} & \bar{m}_{34} \\ \bar{m}_{43} & \bar{m}_{44} \end{bmatrix} = \begin{bmatrix} m_{33} & m_{34} \\ m_{43} & m_{44} \end{bmatrix} + \begin{bmatrix} m_{31} & m_{32} \\ m_{41} & m_{42} \end{bmatrix} [T]$$

The two elastic frequencies of the IOC can be written out directly in the form

$$\omega_{1,2}^2 = \frac{\bar{m}_{33}\, k_{44} + \bar{m}_{44}\, k_{33} \mp \sqrt{(\bar{m}_{33}\, k_{44} - \bar{m}_{44}\, k_{33})^2 + 4\bar{m}_{34}\, \bar{m}_{43}\, k_{33}\, k_{44}}}{2(\bar{m}_{33}\, \bar{m}_{44} - \bar{m}_{34}\, \bar{m}_{43})} \tag{2.25}$$

and the eigenvector is

$$\{\phi_i\} = \begin{bmatrix} T_{11}(\bar{m}_{34}\, \omega_i^2) + T_{12}(k_{33} - \bar{m}_{33}\, \omega_i^2) \\ T_{21}(\bar{m}_{34}\, \omega_i^2) + T_{22}(k_{33} - \bar{m}_{33}\, \omega_i^2) \\ \bar{m}_{33}\, \omega_i^2 \\ k_{33} - \bar{m}_{33}\, \omega_i^2 \end{bmatrix} \quad i = 1,2 \tag{2.26}$$

2.6 Rigid-body librational motion of IOC

The equations of librational motion of an orbiting body which has its principal axes parallel to the orbital axes subjected to disturbing torque M can be written directly in the form [3]

$$I_x\ddot{\theta}_x + 3\,\omega_o^2(I_y - I_z)\theta_x = M_x \tag{2.27}$$

$$I_y\ddot{\theta}_y + 3\,\omega_o^2(I_x - I_z)\theta_y = M_y \tag{2.28}$$

where the I's are the moments of inertia about their respective principal axes through the center of mass. The orbital frequency of a circular orbit is

$$\omega_o = \sqrt{\mu/R^3} \quad \text{rad./s} \tag{2.29}$$

where R is the orbital radius and the gravitational constant

$$\mu = 1.407 \times 10^{16}\ \text{ft}^3/\text{s}^2$$

$$= 3.986 \times 10^{14}\ \text{m}^3/\text{s}^2$$

Thus, Eqs. (2.27) and (2.28) yield the librational frequencies

$$\omega_x = \omega_o \sqrt{3(I_y - I_z)/I_x} \tag{2.30a}$$

$$\omega_y = \omega_o \sqrt{3(I_x - I_z)/I_y} \tag{2.30b}$$

2.7 Response of IOC to crew motion due to rigid-body modes

As shown in Figure 6, due to the arrangement of the modules, crew motion will create a disturbing force either in x-direction (motion in horizontal modules) or z-direction (motion in vertical modules). Hence, the torque M_x is negligible in comparison with M_y which has the magnitudes,

$$M_y = \begin{cases} f_x Z_a & \text{(motion in horizontal modules)} \\ f_z X_a & \text{(motion in vertical modules)} \end{cases} \tag{2.31}$$

As indicated by Eq. (2.30), the librational frequency is approximately equal to $\sqrt{3}$ times the orbital frequency ($I_x \cong I_y$, $I_z << I_x$, see Table 1). The crew kicking motion is completed in a time interval of 1 second which is very short in comparison with the period of librational motion of 3,300 second. This means that the torque can be treated as an impulse torque. The action of an impulse torque is equivalent to give the space station an initial angular velocity, i.e.,

$$\hat{M}_y = \int_o^t M_y(t)dt = \dot{\theta}_y(0)\, I_y \tag{2.32}$$

The solution of Eq. (2.28b) is simply

$$\theta_y = (\hat{M}_y/\omega_y)\sin\omega_y t \tag{2.33}$$

Using the mathematical model for crew motion given by Eq. (2.16), one obtains the response at P_e

$$u(P_e,t) = Z_e\theta_y$$

$$= \begin{cases} \frac{1}{2}\ Z_e Z_a f_o t_1/\omega_y I_y)\sin\omega_y t & \text{(motion in horizontal module)} \\ \frac{1}{2}\ Z_e X_a f_o t_1/\omega_y I_y)\sin\omega_y t & \text{(motion in vertical module)} \end{cases} \tag{2.34}$$

The magnitude of acceleration produced by crew motion at P_e is

$$a(P_e) = \begin{cases} Z_e Z_a f_o \omega_y t_1/2I_y & \text{(motion in horizontal module)} \\ Z_e X_a f_o \omega_y t_1/2I_y & \text{(motion in vertical module)} \end{cases} \tag{2.35}$$

3. Numerical Results of Acceleration Due to Crew Motion

Using data given in Table 4.3.3.4-2 to 4.3.3.4-4 of Reference 1, Table 3 is formed for the formulation of the mass matrix for IOC space station. Note that the total weight given by Tables 4.3.3.4-2 and 4.3.3.4-4 [1] is 77,600 lbs heavier than that given by Table 2. In an effort to match the total weight given in Table 2, some of the weights are not included. There is a significant difference on the location of the center of mass between the present model to that given in Table 2, as shown in the following:

	Table 2	Present Model	Difference
With Orbiter	68.3 ft	113.8 ft	14.4 ft
Without Orbiter	128.2	59.5	8.8

Consider that the distance from the reference point to the microgravity experiment is equal that of the location of crew motion, i.e., $z_e = z_c = -55$ ft. The magnitude of acceleration on the microgravity experiment due to crew motion as given by Eq. (2.20b) is

$$a_e = [\phi_{1j} + \phi_{2j}(z_e/\ell_k)]^2 (f_o/M_{jj}\omega_j t_1) [A_j(t)]_{max} \quad j = 1,2 \qquad (3.1)$$

The value of $[A_j(t)]_{max}$ can be estimated from curves given in Figure 4 for a given value of ω. The summation is omitted so that a_e is calculated for each mode. Based on the librational motion approach, one has from Eq. (2.35)

$$a_e = Z_e^2 f_o \omega_y t_1 / 2I_y \qquad (3.2)$$

The numerical resulted obtained are summarized in Table 4.

It is important to note the following:

(1) No direct comparison can be made on the frequencies obtained to that given in Table 4.3.3.3-3 [1] due to the difference of inertia properties of the models, and furthermore, the JSC model has no distinct fundamental mode that can be singled out.

(2) The acceleration level obtained here is about one order smaller than that given by Table 4.3.3.5-15 [1]. This is due to the fact that the inertia data given by Table 4.3.3.5-6 [1] is about 1/3 of that given by Table 4.3.3.5-6 [1] (without orbiter). In addition, the disturbance torque given by Table 4.3.3.5-7 [1] is more than 2 times the value used here. Thus, the magnitude of accelerations presented in Table 5 are reasonable.

(3) Motion of the space station in Y-Z plane will occur if the disturbing force is in the direction parallel to the solar boom.

(4) The acceleration given by the librational motion is 10^{-3} of that given by elastic motion. This, due to the frequency of librational motion, is only 10^{-3} of the frequency of elastic motion.

4. Design Characteristics of Suspension System

4.1 Acceleration response of a vibration absorber system

Consider that a microgravity experiment package is mounted on the laboratory module structure which has motion $u_o(t)$ as a result of crew motion or other disturbance. As shown in Figure 3, $u_o(t)$ is approximately a harmonic. It is required to design a suspension system which can effectively reduce this disturbance over some frequency range. Vibration can be effectively reduced by using a vibration absorber [4] which is a mass-spring-damper system attached to the main mass-spring system as shown in Figure 7.

Denoting u_o, u_1, u_2, the absolute displacement of the structure, main mass, and absorber mass, respectively, the equations of motion of the system are

$$m_1\ddot{u}_1 + c\dot{u}_1 + (k_1 + k_2)u_1 - c\dot{u}_2 - k_2u_2 = k_1u_o(t) \tag{4.1a}$$

$$m_2\ddot{u}_2 + c\dot{u}_2 + k_2u_2 - c\dot{u}_1 - k_2u_1 = 0 \tag{4.1b}$$

Letting the input $u_o(t)$ be a harmonic disturbance, one may put

$$u_o(t) = U_o e^{i\omega t} \quad , \quad u_1 = U_1 e^{i\omega t} \quad , \quad \text{and} \quad u_2 = U_2 e^{i\omega t} \quad i = \sqrt{-1} \tag{4.2}$$

where U_1 and U_2 are complex quantities that can be determined from the matric equation

$$\begin{bmatrix} (k_1 + k_2 - m_1\omega^2 + ic\omega) & -(k_2 + ic\omega) \\ -(k_2 + ic\omega) & (k_2 - m_2\omega^2 + ic\omega) \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} = \begin{bmatrix} k_1U_o \\ 0 \end{bmatrix} \tag{4.3}$$

If one wishes to determine acceleration response rather than displacement, set

$$\ddot{u}_1 = A_1 e^{i\omega t} \quad \text{with} \quad A_1 = -\omega^2 B_1 \tag{4.4}$$

Now, let M_a denote the magnification factor of acceleration, the ratio of A_1 to input acceleration,

$$M_a = |A_1| / U_o \omega^2 \tag{4.5}$$

First, introduce the following dimensionless parameters:

$$k = k_1/k_2 \quad \mu = m_2/m_1 \quad \zeta = c/2\sqrt{m_2 k_2} \quad r = \omega/\omega_n \tag{4.6}$$

It can be shown that

$$M_a = \left\{ [(1 - \mu k r^2)^2 + 4\mu k \zeta^2 r^2]/D \right\}^{1/2} \tag{4.7}$$

where

$$D = \left\{ 1 - [1 + (1+k)\mu] r^2 + \mu k r^2 \right\}^2 + 4\mu k \zeta^2 r^2 \, [1 + (1+\mu) r^2]^2$$

4.2 Design considerations

The magnitude of the magnification factor of acceleration depends on four parameters:

k, the spring ratio (main spring/absorber spring),

μ, the mass ratio (absorber mass/main mass),

ζ, damping factor (damping coefficient/critical damping),

r, frequency ratio (space station frequency/natural frequency of main mass-spring system).

To plot M_a versus frequency ratio squared as shown in Figure 8, six sets of curves are illustrated for $\zeta = 0.2$. The first three sets are for fixed values of $\mu = 0.01$, 0.025, and 0.05, respectively, with various values of k. The next three sets are for fixed values of k = 10, 15, and 20, respectively, with various values of μ. All these curves have one common characteristic, M_a which can be effectively reduced for $r^2 > 2$. Since the frequency of the IOC space station in elastic vibration is about 1 rad/s, it requires that $k_1 > 2W_1/g$. For a microgravity experiment package of 1,000 lbs, the spring constant of the suspension must be less than 5 lb/in. For the frequency range $0.5 < r^2 < 2$, M_a can be made less than 0.5 by a proper combination of μ and k.

5. Conclusions

A simple structural model for simulation of IOC space station has been presented and formulation of this simple model provides engineers a simple and direct method for computing the fundamental frequencies of the space station and determining the magnitude of acceleration at any point produced by crew motion. Acceleration response of a mass-spring-absorber system to a moving support is also formulated. Design engineers can use plots of acceleration magnification factor versus frequency ratio squared to determine design parameters for the suspension system of the microgravity experiment package.

The following are some significant findings:

(1) The acceleration due to librational rigid-body motion is 3 orders smaller than that due to elastic bending motion.

(2) The frequency of librational motion is 1.8×10^{-3} rad/s, which is approximately $\sqrt{3}$ times the orbital frequency.

(3) The frequencies of elastic motion of the simple model are in the range 0.6 to 1.5 rad/s.

(4) Only the fundamental bending mode has dominant contribution to the acceleration, therefore, a simple model is adequate.

(5) An effective suspension system can reduce the acceleration to 1/4 of its magnitude, at most.

(6) The ideal location of the microgravity lab does not have to be near the center of mass of the space station.

(7) The most effective means to eliminate acceleration is to have the experiment module near the nodal point of the fundamental bending mode, as illustrated in Figure 9. This means that the factor $(\phi_{11} + \phi_{21}\, z_e/\ell_k)$ in Eq. (4.1) becomes very small. Example: The IOC without orbiter has moved the lab module 59.9 ft further away from the center of mass, but the acceleration is reduced 3 orders smaller than the IOC with orbiter.

(8) It is favorable to perform the microgravity experiments when the orbiter is not present.

(9) It is possible that by rearranging some massive elements, a minimum value of $(\phi_{11} + \phi_{21}\, z_e/\ell_k)/M_{11}$ can be reached. However, this has to be done by trial and error method.

Appendix A. Formulation of Mass Matrix

Notations:

I_{ox}, I_{oy}	moment of inertia of orbiter about x- and y-axis through C.M. of orbiter, respectively
I_{sx}, I_{sy}	moment of inertia of solar arrays about x- and y-axis through solar boom at the attachment, respectively
I_{UB}	moment of inertia of upper boom about x-axis through end of keel
I_{LB}	moment of inertia of lower boom about x-axis through attachment
ℓ_k	length of keel frame
ℓ_s	length of solar boom
M_k	mass of keel frame alone
m_k	kth concentrated mass attached to keel at distance z_k
M_s	mass of solar boom frame alone
m_s	sth concentrated mass attached to solar boom at y_s
m_r	rth concentrated mass attached to rigid main body at z_r
m_o	mass of orbiter
u_x, u_y	displacement in x and y direction respectively
$\bar{y}_s$	$= y_s/\ell_s$, coordinate of concentrated mass attached to solar boom
$\bar{z}_k$	$= z_k/\ell_k$, coordinate of concentrated mass attached to keel
z_o	coordinate of orbiter
subscripts & summation index	k(keel), r(rigid main body), s(solar boom)

A-1. Motion of IOC in X-Z Plane

The kinetic energy of rigid main body is

$$T_R = 1/2 \sum_r m_r \dot{u}_x^2(z_r)$$

$$= 1/2 \sum_r m_r[\psi_1(z_r)\dot{q}_1 + \psi_2(z_r)\dot{q}_2]^2 + 1/2\ I_{oy}[\psi_2'(z_o)\ \dot{q}_2]^2 \qquad \text{(A-1)}$$

The kinetic energy of keel structure is

$$T_K = 1/2 \sum_k m_k[\psi_1(z_k)\dot{q}_1 + \psi_2(z_k)\dot{q}_2 + \psi_3(z_k)\dot{q}_3]^2$$

$$+ 1/2\ M_k \int [\psi_1\dot{q}_1 + \psi_2(z)\dot{q}_2 + \psi_3(z)\dot{q}_3]^2\ dz/\ell_k$$

$$+ 1/2 \sum_s I_{sy}[\psi_2(z_s)\dot{q}_2 + \psi_3'(z_s)\dot{q}_3]^2 \qquad \text{(A-2)}$$

The kinetic energy of solar boom is

$$T_S = 1/2\ M_s \int [\psi_1(z_s)\dot{q}_1 + \psi_2(z_s)\dot{q}_2 + \psi_4(y_s)\dot{q}_4]^2 dy_s/\ell_s$$

$$+ 1/2 \sum_s m_s[\psi_1(z_s)\dot{q}_1 + \psi_2(z_s)\dot{q}_2 + \psi_4(y_s)\dot{q}_4]^2$$

$$+ 1/2 \sum_s I_{sx}[\psi_4'(y_s)\dot{q}_4]^2 \qquad \text{(A-3)}$$

By using Eq. (2.5) the elements of the mass matrix are obtained as follows:

$$m_{11} = M_{total} = M_r + M_k + \sum_k m_k + M_s + \sum_s m_s$$

$$m_{12} = \sum_r m_r\bar{z}_r + 1/2\ M_k + \sum_k m_k\bar{z}_k + (M_s + \sum_s m_s)\bar{z}_s$$

$$m_{13} = 1/4\ M_k + \sum_k m_k(\bar{z}_k^2 - 1/3\ \bar{z}_k^3) + (M_s + \sum_s m_s)(\bar{z}_s^2 - 1/3\ \bar{z}_s^3)$$

$$m_{14} = 1/4\ M_s + \sum_s m_s(y_s^2 - 1/3\ y_s^3)$$

$$m_{22} = \sum_r m_r\bar{z}_r^2 + 1/3\ M_k + \sum_k m_k\bar{z}_k^2 + (M_s + \sum_s m_s)\bar{z}_s^2 + \sum_s I_{sy}/\ell_k^2 + I_{oy}(\psi_2')^2$$

$$m_{23} = \frac{11}{60} M_k + \sum_k m_k\bar{z}_k^3(1 - 1/3\ \bar{z}_k) + (M_s + \sum_s m_s)(1 - 1/3\ \bar{z}_s)\ \bar{z}_s^3$$

$$+ (\sum_s I_{sy})\ \bar{z}_s(2 - \bar{z}_s)/\ell_k^2$$

$$m_{24} = \bar{z}_s m_{14}$$

$$m_{34} = (\bar{z}_s^2 - 1/3\ \bar{z}_s^3)\ m_{14}$$

$$m_{33} = (11/105)M_k + \sum_k m_k(\bar{z}_k^2 - 1/3\ \bar{z}_k^3)^2 + (M_s + \sum_s m_s)(\bar{z}_s^2 - 1/3\ \bar{z}_s^3)^2$$

$$+ (\sum_s I_{sy})\ (2\bar{z}_s - \bar{z}_s^2)^2/\ell_k^2$$

$$m_{44} = (11/105)M_s + \sum_s m_s(\bar{y}_s^2 - 1/3\ \bar{y}_s^3)^2 + 4I_{sx}\{[\bar{y}_{s1}(2 - y_{s1})^2$$

$$+ [\bar{y}_{s2}(2 - y_{s2})^2\}/\ell_s^2$$

A-2. Motion in Y-Z Plane

The kinetic energy of rigid main body is

$$T_R = 1/2 \sum_r m_r[\psi_1(z_r)\dot{q}_1 + \psi_2(z_r)\dot{q}_2]^2 + 1/2\ I_{ox}[\psi_2'(z_o)\dot{q}_2]^2$$

$$+ 1/2\ I_{UB}[\psi_2'(z_{UB})\dot{q}_2]^2 \qquad \text{(A-4)}$$

The kinetic energy of the keel structure is

$$T_K = 1/2 \sum_k m_k[\psi_1\dot{q}_1 + \psi_2(z_k)\dot{q}_2 + \psi_3(z_k)\dot{q}_3]^2$$

$$+ 1/2\ M_k \int_k [\psi_1\dot{q}_1 + \psi_2(z_k)\dot{q}_2 + \psi_3(z_k)\dot{q}_3]^2 dz_k/\ell_k$$

$$+ 1/2\ I_{UB}[\psi_2'(z_s)\dot{q}_2 + \psi_3'(z_s)\dot{q}_2]^2 \qquad \text{(A-5)}$$

The kinetic energy of the solar boom structure is

$$T_S = 1/2\ M_s \int_s [\psi_1\dot{q}_1 + \psi_2(z_s)\dot{q}_2 + \psi_3(z_s)\dot{q}_3]^2 dy_s/\ell_s$$

$$+ 1/2\ M_s \int_s \{[\psi_2'\dot{q}_2 + \psi_3'(z_s)\dot{q}_3]y_s - \psi_4(y_s)\dot{q}_4\}^2 dy_s/\ell_s$$

$$+ 1/2 \sum_s \{m_s[\psi_1\dot{q}_1 + \psi_2(z_s)\dot{q}_2 + \psi_3(z_s)\dot{q}_3]^2$$

$$+ [\psi_2'\dot{q}_2 + \psi_3'(z_s)\dot{q}_3]y_s - \psi_4(y_s)\dot{q}_4]^2\}$$

$$+ 1/2 \sum_s I_{sx}[\psi_2'\dot{q}_2 + \psi_3'(z_s)\dot{q}_3 - \psi_4'(y_s)\dot{q}_4]^2 \qquad \text{(A-6)}$$

Applying Eq. (2.5) results in the mass matrix for motion in Y-Z plane:

$$m_{11} = M_T = (m_{11})_{xz}$$

$$m_{12} = (m_{12})_{xz}$$

$$m_{13} = (m_{13})_{xz}$$

$$m_{14} = 0$$

$$m_{22}=(m_{22})_{x\bar{z}} + (\psi_2')^2 \left[\sum_s (I_{sx} - I_{sy}) + 1/3\ M_s \ell_s^2 + \sum_s m_s y_s^2 + I_{ox} - I_{oy} + I_{UB} + I_{LB}\right]$$

$$m_{23} = (m_{23})_{xz} + \psi_2'\psi_3'(z_s)\left[\sum_s (_{sx} - I_{sy}) + \sum_s m_s y_s^2 + 1/3\ M_s \ell_s^2\right] + I_{UB}\psi_2'\psi_3'(\ell_k)$$

$$m_{24} = -\left[\frac{13}{30} M_s \ell_s + \sum_s m_s y_s \psi_4(y_s) + \sum_s I_{sx}\psi_4'(y_s)\right]\psi_2'$$

$$m_{33} = (m_{33})_{xz} + \left[\sum_s m_s y_s^2 + \sum_s (I_{sx} - I_{sy})\right][\psi_3'(z_s)]^2 + I_{UB}[\psi_3'(\ell_k)]^2$$

$$m_{34} = -\left[\frac{13}{30} M_s \ell_s + \sum_s m_s \psi_4(y_s) y_s + \sum_s I_{sx}\psi_4'(y_s)\right]\psi_3'(z_s)$$

$$m_{44} = (m_{44})xz$$

The subscript "xz" denotes element of mass matrix of motion in X-Z plane.

A-3. Mass and inertia properties for IOC model

Table 3 is formed based on data given by Table 4.3.3.4-2 to 4.3.3.4-4 [1] for the purpose of formulation of the mass matrix.

Appendix B. Determination of Stiffness Matrix

B-1. Bending strain energy

The bending strain energy of keel and solar boom structure is given by

$$V = 1/2 \int_0^{\ell_k} (EI)_k [\psi_3'' q_3]^2 dz_k + 1/2 \int_0^{\ell_s} (EI)_s [\psi_4''(y_s) q_4]^2 dy_s$$

Applying Eq. (2.6), one obtains the following non-zero elements of the stiffness matrix

$$k_{33} = 4(EI)_k/3\ell_k^3 \qquad k_{44} = 4(EI)_s/3\ell_s^3$$

B-2. Moment due to gravity gradient

Using the formula given by Reference 3, the following are obtained:

Motion in X-Z Plane

$$M_y = -3(\mu/R^3)(I_x - I_z)q_2 \ , \qquad k_{22} = 3(\mu/R^3)(I_x - I_z)$$

Motion in Y-Z Plane

$$M_x = -3(\mu/R^3)(I_y - I_z)q_2 \ , \qquad k_{22} = 3(\mu/R^3)(I_y - I_z)$$

where the moments of inertia are about the axes through the c.m. of the IOC space station.

Appendix C. Numerical Results

C-1. Motion in X-Z plane (IOC with orbiter attached)

$$[m] = \begin{bmatrix} 18900 & -2758 & 1096 & 171.6 \\ -2758 & 3168 & 831.9 & 96.0 \\ 1096 & 831.9 & 483.7 & 43.7 \\ 171.6 & 96.0 & 43.7 & 122.9 \end{bmatrix}$$

$$\omega_1^2 = 1.0431$$

$$\omega_2^2 = 6.64$$

$$\{\phi_1\} = \begin{bmatrix} 1.1135 \\ .9546 \\ -10.118 \\ .1639 \end{bmatrix} \qquad \{\phi_2\} = \begin{bmatrix} .2770 \\ .1165 \\ -6.49 \\ -411.87 \end{bmatrix}$$

$$M_{11} = 29157$$

$$M_{22} = 1488$$

C-2. Motion in X-Z plane (IOC without orbiter attached)

$$[m] = \begin{bmatrix} 11590 & 376.2 & 1096 & 171.6 \\ 376.2 & 1829.6 & 831.9 & 96.0 \\ 1096 & 831.9 & 483.7 & 43.7 \\ 171.6 & 96.0 & 43.7 & 115.2 \end{bmatrix}$$

$$\omega_1^2 = 2.0243$$

$$\omega_2^2 = 7.436$$

$$\{\phi_1\} = \begin{bmatrix} .1936 \\ 1.054 \\ -2.429 \\ .1095 \end{bmatrix} \qquad \{\phi_2\} = \begin{bmatrix} .9485 \\ 4.769 \\ -8.923 \\ -17.603 \end{bmatrix}$$

$$M_{11} = 191.5$$

$$M_{22} = 32169$$

C-3. Motion in Y-Z plane (IOC with orbiter attached)

$$[m] = \begin{bmatrix} 18900 & -2758 & 1096 & 0 \\ -2758 & 3298 & 872.3 & -92.5 \\ 1096 & 872.3 & 513.5 & -74.5 \\ 0 & -92.5 & -74.5 & 122.9 \end{bmatrix}$$

$$\omega_1^2 = .8022$$

$$\omega_2^2 = 7.881$$

$$\{\phi_1\} = \begin{bmatrix} .3673 \\ 1.194 \\ -3.332 \\ .1674 \end{bmatrix} \qquad \{\phi_2\} = \begin{bmatrix} .3332 \\ .9829 \\ -3.274 \\ -5.779 \end{bmatrix}$$

$$M_{11} = 959.8$$

$$M_{22} = 3312$$

C-4. Motion in Y-Z plane (IOC without orbiter attached)

$$[m] = \begin{bmatrix} 11590 & 369.8 & 1096 & 0 \\ 369.8 & 3298 & 872.3 & -92.5 \\ 1096 & 872.3 & 513.5 & -74.5 \\ 0 & -92.5 & -74.5 & 11 \end{bmatrix} \qquad \omega_1^2 = .4301 \qquad \omega_2^2 = 7.046$$

$$\{\phi_1\} = \begin{bmatrix} .1906 \\ .510 \\ -2.191 \\ -1.716 \end{bmatrix} \qquad \{\phi_2\} = \begin{bmatrix} .3221 \\ .5438 \\ -3.589 \\ -13.17 \end{bmatrix} \qquad M_{11} = 927.5 \qquad M_{22} = 18582$$

C-5 Computation of accelerations

Using the data given above, Eq. (3.1), and with the aid of Figure 4 for the value of $[A(t)]_{max}$, acceleration of the experiment package can be calculated. The results are shown in Table 4.

References

1. "Space Station Configuration Description," Systems Engineering Integration Space Station Program Office, L.B.J. Space Center, August, 1984.

2. Craig, R. R., Jr.: "Structural Dynamics," Wiley, 1981.

3. Kane, T. R.; Likins, P. W.; and Levinson, D. A.: "Spacecraft Dynamics," McGraw-Hill, 1983, p. 177.

4. Tong, K. N.: "Theory of Mechanical Vibrations," Wiley, 1960.

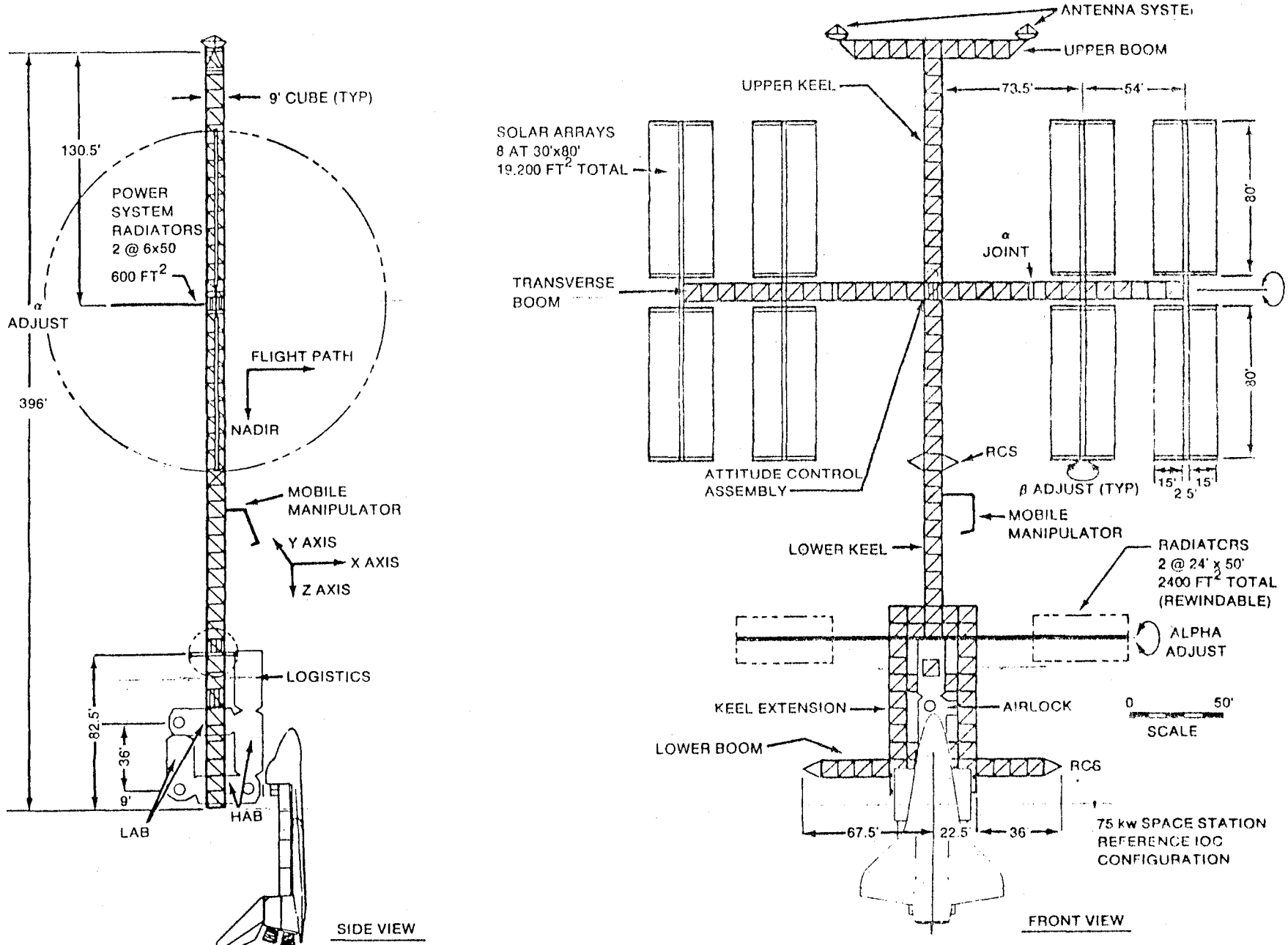


Figure 1 (Figure 4.2.1-1 [1])IOC REFERENCE CONFIGURATION

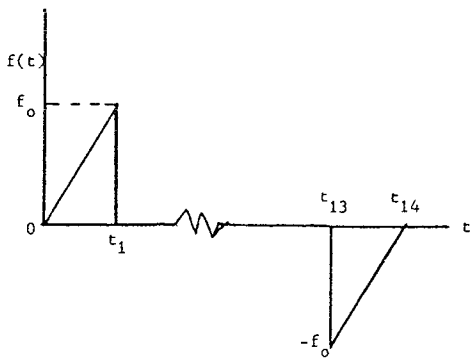


Figure 2 Mathematical Model of Crew Motion

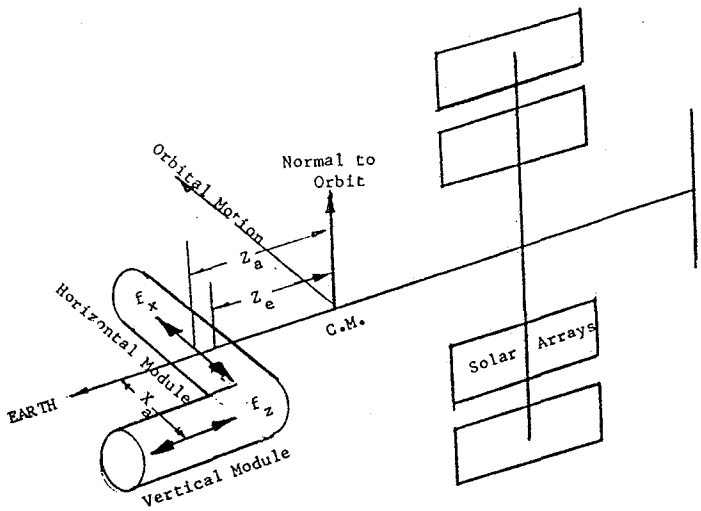
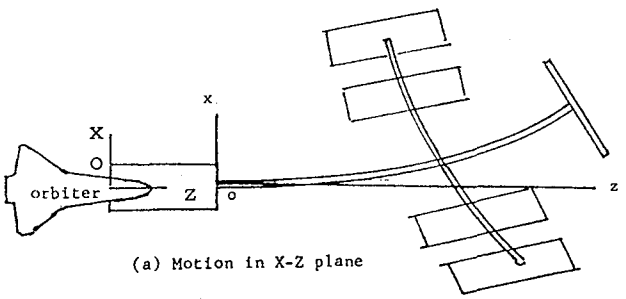
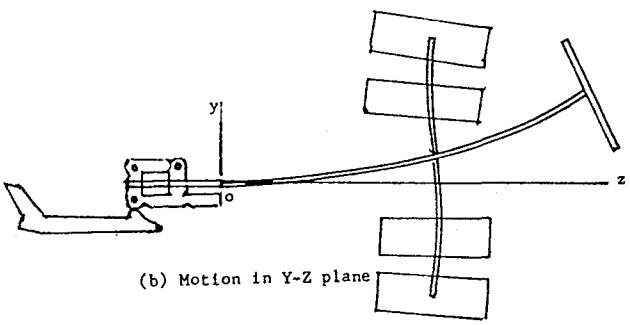


Figure 6 Reference of Crew Motion



(a) Motion in X-Z plane



(b) Motion in Y-Z plane

Figure 5 Model of IOC Space Station

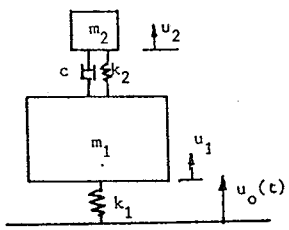


Figure 7 Model of SuspensionSystem

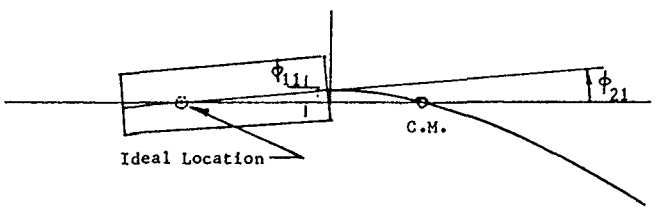


Figure 9 Ideal Location of Microgravity Experiment

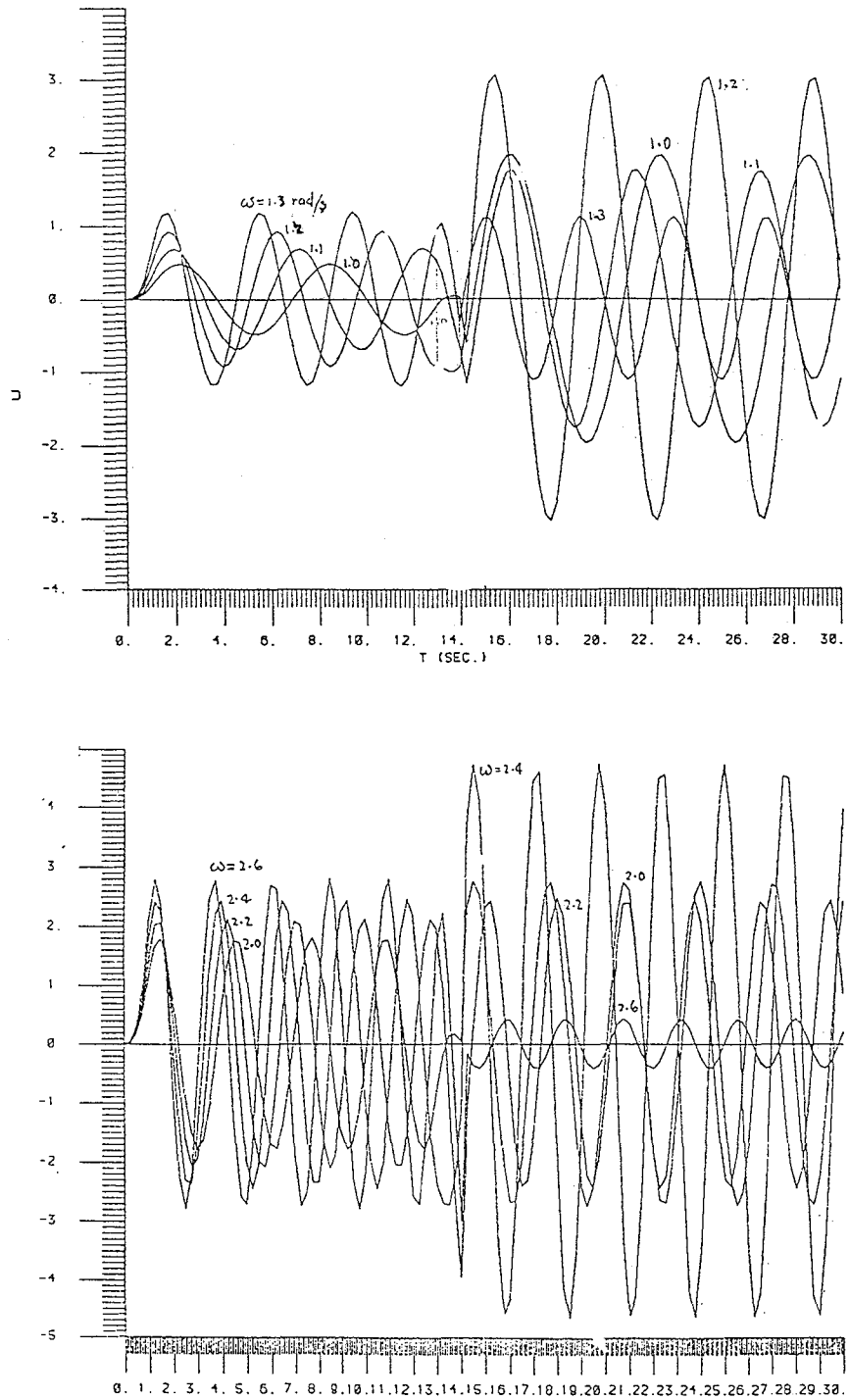


Figure 3 Displacement Functions

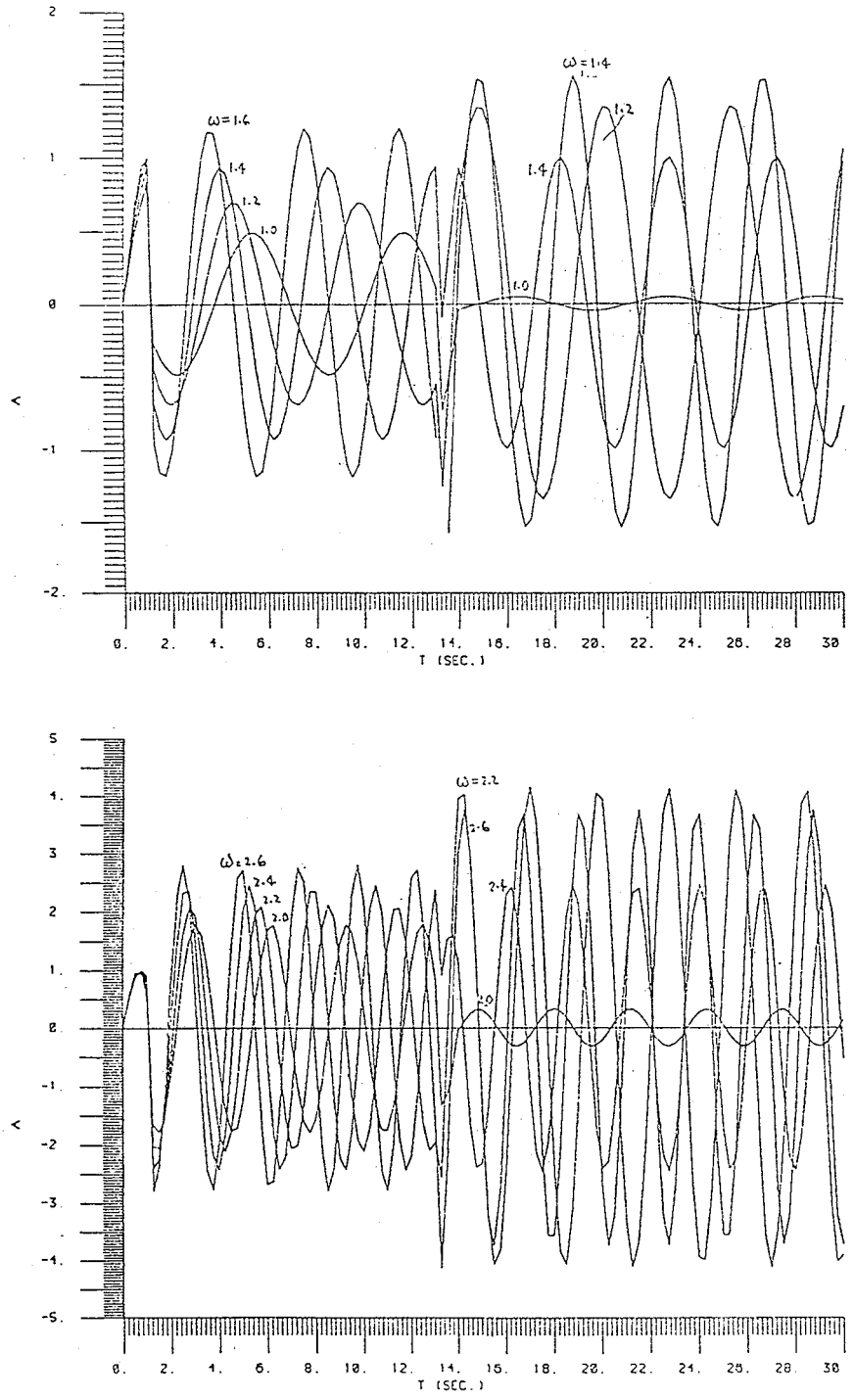


Figure 4 Acceleration Functions

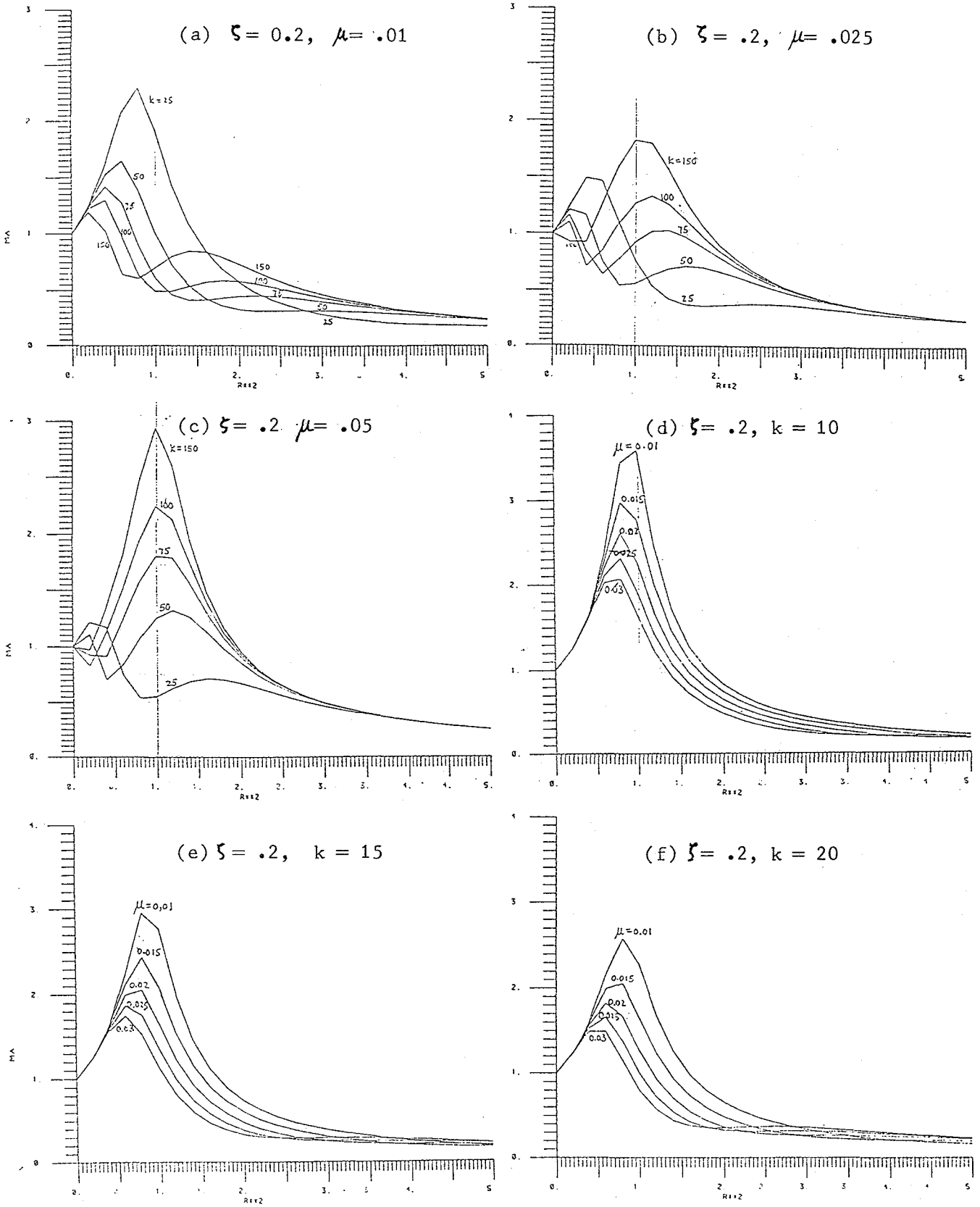


Figure 8 Acceleration Magnification Factor

Table 1. Properties of Analytical Reference Configuration Space Station Model (Table 4.3.3.3-1 [1])

Component	Bending Stiffness (ft-lbs-ft)	Torsional Stiffness (ft-lbs-ft)	Mass Length (Slugs/ft)	Bending Strength (ft-lbs)	Torsional Strength (ft-lbs)
Booms and Keels	1.31 E+9	3.18 E+8	0.25	35,000	15,000
30-Inch Astro Mast	3.13 E+6	2.08 E+5	2.3	3,480	208

Table 2. Inertial Properties of Analytical Model (Table 4.3.3.3-2 [1])

Case	Weight (lbs)	C. G. Coordinates (ft)	Moments of Inertia (lb-ft-sec^2)
Without Payloads and Orbiter	269,000	(1.1,0,84.2)	(8.63 E+7, 7.82 E+7, 1.20 E+7)
With Payloads Only	373,200	(-1.1,0,128.2)	(2.06 E+8, 1.98 E+8, 1.45 E+7)
With Orbiter Only	508,800	(5.7,0,34.4)	(1.45 E+8, 2.37 E+8, 1.35 E+7)
With Payloads and Orbiter	608,600	(3.59,0,68.3)	(3.21 E+8, 3.14 E+8, 1.62 E+7)

Table 3. Masses and Inertia Properties Used for IOC Model

r	Masses attached to main body	Z_r ft	z_r ft	W_r lbs
1	Fuel tank and gases	94.5	- 5.5	17,002
2	Module radiators	82.5	- 17.5	3,000
3	Five-bay platform	76.5	- 23.5	1,403
4	COM 1203	64.0	- 26.0	11,000
5	Logistics module	71.5	- 28.5	37,823
6	OMV and kits	67.5	- 32.5	23,750
7	Horizontal lab	45.0	- 55.0	54,295
8	Vertical lab	34.5	- 65.5	27,067
9	Keel extension	27.0	- 73.0	970
10	Vertical lab	19.5	- 80.5	47,709
11	Lower boom	13.5	- 86.5	728
12	Horizontal lab, SAA0207 and SAA0201	9.0	- 91.0	37,089
13	Orbiter	-26.66	-126.66	235,400

s	Masses attached to solar boom	Z_s ft	z_s ft	y_s ft	W_s lbs
	Solar boom structure	265.5	165.5	0-132	2,345
1	Power system radiators	265.5	165.5	54.0	750
2	TDM 2010	265.5	165.5	63.4	1,540
3	4 inboard solar arrays	265.5	165.5	78.0	4,787
4	4 outboard solar arrays	265.5	165.5	132	4,787

k	Mass attached to keel		Z_k ft	z_k ft	W_k lbs
	Upper and lower keel structure		100-396	0-296	2,504
1	Remote manipulator		162.0	62.0	2,000
2	Refuel attachment, tanks and tools		107.5	7.5	4,625
3	TDM 2570		210.5	110.5	2,000
4	Instruments and storage shelter		212.5	112.5	4,625
5	Storage boxes and tools		272.5	172.5	9.850
6	Service attachments		290.5	190.5	3,750
7	TDM 2560		295.5	195.5	7,055
8	Satellite		324.5	224.5	20,000
9	Upper boom and antenna system		396.0	296.0	17,734

	Rotational Moment of Inertia	I_x ft-lb-s^2	I_y ft-lb-s^2
	Orbiter (about c.m. of orbiter)	7×10^6	8×10^6
	Solar array (about attachment) each		
	Parallel to Nadir	101,000	107,300
	Normal	2,790	107,300
	Upper boom with antenna system	158,800	small
	Module radiators (about attachment)	583,500	4,470
	Power system radiators (about attachment)	750	25,120

Note: Z is measured from the bottom end of the keel extension and z is measured from the joint of keel and keel extension

Table 4. Numerical Results for IOC Model

Motion	ω and a_e	Motion in X-Z Plane		Motion in Y-Z Plane	
		With Orbiter	Without Orbiter	With Orbiter	Without Orbiter
Bending	ω rad/s	1.021	1.423	0.896	0.656
	a_e	$1.14\ g \times 10^{-5}$	$1.44\ g \times 10^{-8}$	$0.86\ g \times 10^{-5}$	$1.17\ g \times 10^{-5}$
	ω rad/s	2.577	2.730	2.810	2.654
	a_e	$2.26\ g \times 10^{-6}$	$1.38\ g \times 10^{-7}$	$0.66\ g \times 10^{-5}$	$2.9\ g \times 10^{-6}$
Libra. Motion	ω rad/s	1.89×10^{-3}	1.89×10^{-3}	1.85×10^{-3}	1.81×10^{-3}
	a_e	$0.254\ g \times 10^{-8}$	$5.13\ g \times 10^{-8}$	$0.24\ g \times 10^{-8}$	4.72×10^{-8}

1985

NASA/ASEE SUMMER FACULTY RESEARCH FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA AT HUNTSVILLE

MACHINE VISION AND THE OMV

Prepared By:	Michael A. McAnulty
Academic Rank:	Assistant Professor
University and Department:	University of Alabama in Birmingham Department of Computer and Information Science
NASA/MSFC:	
Laboratory:	Information and Electronic Systems
Division:	Software and Data Management
Branch:	Data Management
NASA Counterpart:	Frank Vinz
Date:	30 August 1985

NASA-NGT-01-008-021
(The University of Alabama in Huntsville)

Machine Vision And The OMV

M. A. McAnulty, NASA/ASEE Fellow
NASA Contact : Frank Vinz
Marshall Space Flight Center

The orbital Maneuvering Vehicle (OMV) is intended to close with orbiting targets for relocation or servicing. It will be controlled via video signals and thruster activation based upon earth or space station directives. A human operator is squarely in the middle of the control loop for close work. Without directly addressing future, more autonomous versions of a remote servicer, several techniques that will doubtless be important in a future increase of autonomy also have some direct application to the current situation, particularly in the area of image enhancement and predictive analysis.

Several techniques are presented, and some few have been implemented, which support a machine vision capability proposed to be adequate for detection, recognition, and tracking. Once feasibly implemented, they must then be further modified to operate together in real time. This may be achieved by two courses, the use of an array processor and some initial steps toward data reduction. The methodology for adapting to a vector architecture is discussed in preliminary form, and a highly tentative rationale for data reduction at the front end is also discussed. As a by-product, a working implementation of the most advanced graphic display technique, ray-casting, is described.

Acknowledgements

Appreciation comes in many flavors. For overall mission orientation and administrative support we owe James Dozier, Leroy Osborn, and Gerry Karr a great deal. For further elucidation of the NASA situation and requirements, as well as technical support, thanks are due to Caroline Wang, Steve Purinton, Tom Bryan, Debbie Graham, and Audie Anderson. My counterparts, Frank Vinz and Ken Fernandez, were liberal with both information and resources, including an excellent technical report which I am still trying to catch up to. [14]

I. Introduction

The orbiting Maneuvering Vehicle (OMV) has ben described as an orbiting tugboat, and is intended to reposition other orbiting vehicles and, eventually, perform simple maintenance procedures. It is expected to be the first in a long line of unmanned remote vehicles, much in the way Mariner was but more than a passive presence. Although the planned version of the OMV is to be entirely ground- or space station-controlled, that is, a telepresence or tele-operated vehicle, the case is made here that some considerable testing and verification of more autonomous versions of the OMV can be performed within the teleoperation framework. The proposed techniques do not in any way propose to modify how the OMV would be operated, but fit well within the scope of enhancements to a ground based operation, primarily in the area of predictive graphics, to be described below. This report describes some preliminary procedures that pertain primarily to on-board vision and scene interpretation, but which would necessarily be well-integrated with other sensory and effector modalities in a working prototype.

The following sections describe, in turn, the planned mode of operation of the OMV, aspects of machine vision that relate to it, a global description of necessary tasks, and a report upon work accomplished so far and planned for the near future.

II. Current Planned OMV Operations

A considerable amount of ground simulation of the teleoperation of a remote vehicle has been performed, utilizing a compressed air controlled vehicle that rides on air bearings on the 'flat floor' at MSFC. Included in the simulations is a very low bandwidth communication channel for video signals from the vehicle, which allows for no more than approximately one 256 by 256 pixel frame per second. This mimics the expected performance of the TDRSS satellite channel that would be used for orbital operation. A further feature of this channel is that significant time delays occur, on the order of a second each way in low orbit.

The vehicle operator controls a six-degree-of-freedon joystick to 'fly' the vehicle, and observes the vehicle's surroundings via one or more video cameras mounted, with lights, on the vehicle. Four cameras are currently planned for far and near work, and for different views of a docking operation, although only two are mounted at present. Thus, practically all control of the vehicle is exercised by an attentive human operator.

The communication channel delay means that an operator does not see the result of an action for approximately two seconds. This is approximate because there are also abrupt changes in the signal delay, on the order of a half-second or so, as the routing changes due to earth position. Such considerations are critical, since docking and close work cannot be done at great speed. Collisions at too great a speed will either be near-elastic or, if kinetic energy is absorbed, damaging. Even if contact can be achieved and maintained, several internal systems on some targets may be impaired by even moderate shock. In fact, the necessity for slow maneuvering tends to lessen the time delay problem as the operator will necessarily work in finer increments.

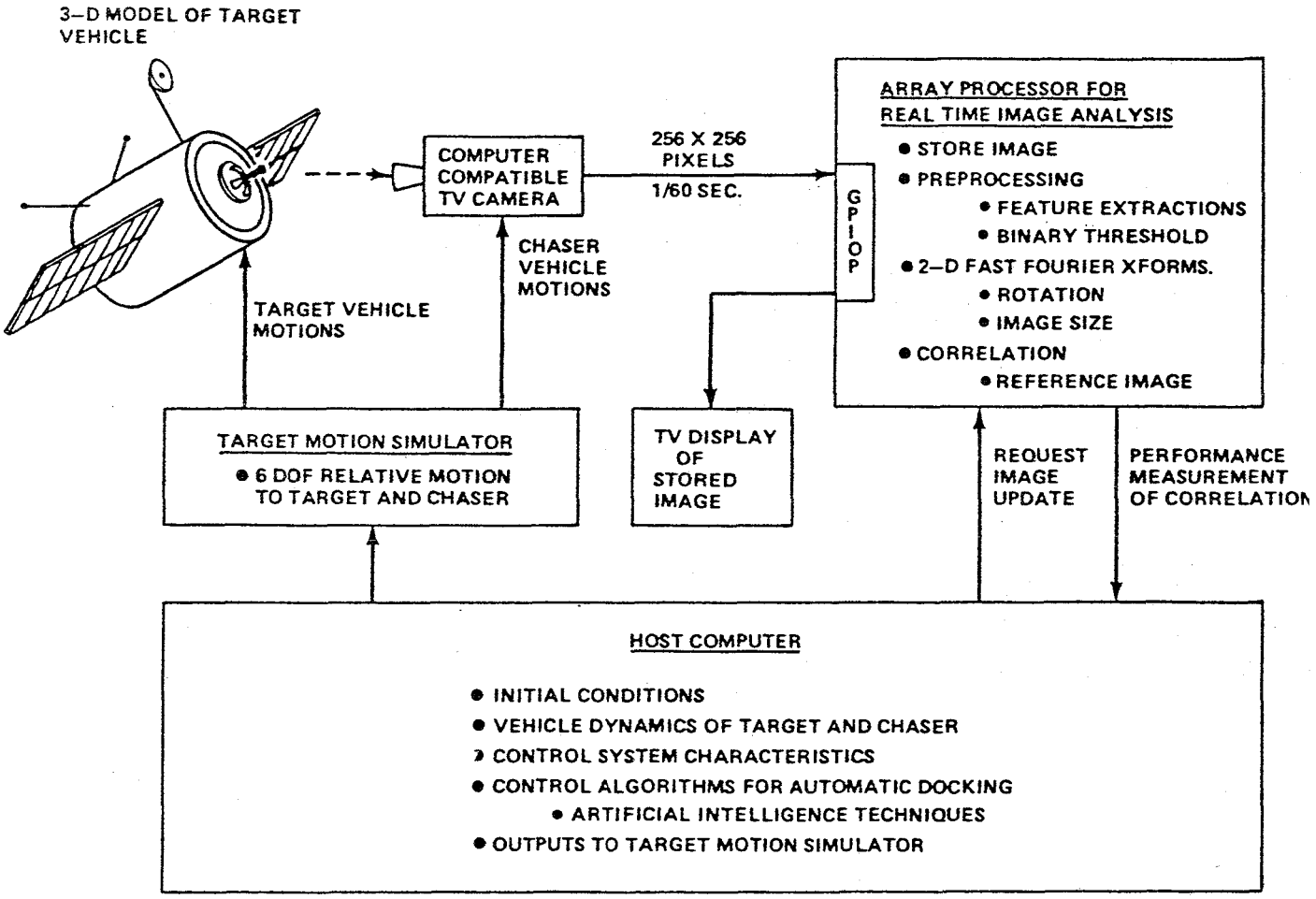


Exhibit 1 - The simulation environment at MSFC. Note that the vector processor, tightly coupled to a host computer, is squarely in the loop as well, and may be used electively to either perform operations on the images, or simply pass them through unaltered.

Although a straight video signal is sufficient for earth simulations, it is anticipated that some cleaning and enhancement of the video picture might be desirable in actual operations. To this end, a signal matching interface between the video signal and the General Purpose Input-Output Processor (GPIOP) of an FPS-5210 floating point array processor is under construction. The FPS can perform simple filtering operations in under half a second (to be described in a later section) upon the entire image, which matches the input rate anticipated via TDRSS, but which also will introduce more delay into the loop.

It has been suggested [1] that the time delay problem might be addressed by a technique termed 'predictive graphics'. Rather than showing the operator the original video signal, a graphic rendering from a formal model, such as an engineering description or a constructive solid geometric model, is shown instead. The model would be so presented as to appear like the object would after a suitable delay. This requires two things which are at present entirely absent from the simulation environment.

- Matching procedures for fitting the formal model to the video picture, in order to determine the target's current orientation and position, and
- Means for predicting where the target will be after a suitable delay, which would utilize both measured trajectories of the pieces of the target, including transforming these to geometric operations on the formal model, and knowledge of the vehicle's motion relative to the target, to be derived from joystick or thruster signals.

This topic will be returned to after a generic discussion of what is generally agreed to constitute machine vision.

III. Aspects of Machine Vision

Two properties of machine vision relevant to this project. The first of these is that machine vision is not an end in itself, rather a means for achieving some end. The second is that current thinking about machine vision separates it into a number of steps that deal with quite different kinds and amounts of data. The flavor of artificial intelligence is, in both aspects, difficult to ignore.

Purely passive vision, with no goal orientation, is rather difficult to contemplate meaningfully. At the very least, a seeing entity is constructing a story. Some situations are rather clear cut. Standard bin-picking problems use a vision capability to aid a robot in the task of picking an object from among other objects, lying at an arbitrary orientation, and move it to another location to be placed at a specified orientation. Surveillance systems, in their simplest form, are required to detect unexpected motion, or more generally unexpected phenomena. This raises the problem of specifying what expected phenomena look like, and requires a rather more complicated world model. Further, human vision works explicitly or implicitly with not only a well-developed world model, but in concert with several other senses as well, of which locomotion and touching are possibly most important.

The basic functions planned for the OMV would appear to lie within a restrictive world model consisting of a specification of the target vehicle and of planned procedures, although even this may become more complicated in the event that damage assessment or contingency actions are required. For the near term, having a person in the control loop bypasses the more sophisticated requirements entirely.

The second aspect of machine vision, its multistep nature, is of some considerable relevance. Consider that a system begins with a very simple data structure, the sampled intensity function in a rectangular array (the raster), and from it should derive a symbolic description of the essential features of the scene, sufficiently informative to enable actions to be taken. A representative, although by no means authoritative, list of steps follows. It should be noted that each of them involves both data reduction and data transformation.

- *Segmentation* is the process of classifying regions of the picture solely in terms of their distinctness from each other. A usual picture of 65,000 or 250,000 pixels will generally produce a number of regions two or more orders of magnitude fewer than the number of pixels. The description of each segment will include a boundary encoding and perhaps a list of central moments, or some other topological description. Naive segmentation is probably the most over-investigated and best understood component of computer vision.
- *Articulation* , or the analysis of the relationships between segments, analyzes adjacencies and connections. Since a segment represents a surface, at this point possible and probable orientations of the surface and its curvature are essayed. Relevant data structures include adjacency graphs and preliminary parametric descriptions.
- *Naive Construction* attempts to interpret the surfaces as belonging to one or more physically possible objects, and attempts to build a three-dimensional model consistent with the scene. At this point components of the scene may be labelled as generic, domain-independent geometric primitives.
- *Labelling* ,or attaching domain-dependent names to the objects, relates the world of possible objects to that of known, named objects. By this stage a preliminary recognition may be said to have occurred.
- *Updating* the current world model, in the case of a time sequence of scenes, is important. If a single scene were being analyzed, this would not be a critical step, but most useful applications not only involve changes in time, but make considerable use of these changes in interpreting the field of view. In the case of predictive graphics, both quantified trajectories and logical data (visible/occluded) would be used.
- *Domain Consistency* is analogous to naive construction, in that the interpretation is tested for possibility as a world situation. The plausibility and consistency of the domain labelling is evaluated in domain terms. It is at this level that symbolic programming is most likely to be employed, as well as sophisticated concept structures. Further, considerable control of all previous steps is likely to be exercised here, perhaps redoing some of them in different fashion to provide a more consistent interpretation. Thus, a system 'thinks' not only about the application domain, but about its own operation as well. [8]

Although these steps are presented as discrete phases, there is likely to be considerable crosstalk between them. It is desirable, however, to construct and maintain a system in separate components, since the objects manipulated differ widely between phases, and in fact different phases may be best implemented by different programming paradigms.

Situations which make explicit use of time variance can work in a much more efficient mode, making use of the property that in the real world things should not change radically from one view to the next. In a sequence of views, components may be 'tracked' rather than recognized from scratch every time.

This property is termed time coherence. Its relative, spatial coherence, may be stated that if some pixel is determined to be on a particular surface, then neighboring pixels have a high probability of being on the same surface.

On the proscriptive side, it is useful to discuss briefly what machine vision is not. That is, while the following techniques may play a modified part in certain portions of a system, none is in itself sufficient to achieve the desired objectives on its own.

Image processing projects a picture function onto another functional basis, such as the Fourier, Hadamard, or Walsh functions, and examines the projection for various clusters of coefficients which may relate to image features. The weakness of this approach is that one is confronted with as large a data space after the transform as before, no reduction in the problem size has been accomplished. Further, features in a function space relate only occasionally to useful scene properties. Segments of a scene may be usefully transformed for surface properties such as texture.

Classical *pattern recognition* assigns a data cluster to one of a finite number of classes based upon a training sample. There are both statistical and syntactic forms of pattern recognition, but as traditionally employed they result in extremely difficult program control structures. This is partly the result of trying to push domain reasoning into procedural code, and why the multicomponent model prescribed above has evolved. Another weakness is that part of the recognition process, related to model fitting, involves a continuum of interpretations rather than a finite number of names.

Model fitting appears especially appropriate to the predictive graphics situation, in that certain parameters (the position and velocity coordinates of the target) must be optimally determined. There is, however, no global closed form solution that will fit a structural model to the two-dimensional projection in the scene. A trial and error method, using an expensive correlation as criterion function, is necessary if one wishes to take a global approach. It appears that an analysis of the scene components, essentially all the prescriptive steps described above, is much more likely to provide an initial parameter setting that may then be tested at specific points or regions rather than over the entire model. For one thing, a model must necessarily be modified to reflect that certain of its components will not even be visible in the projection, hence cannot be explicitly fit.

IV. Proposed Tasks

There are two broad classes of necessary activity. The first is to develop procedures that will enable recognizing and tracking a typical target. The second is to then modify the procedures so that they will work in near real time, an objective that is initially to be met by using the vector processor, and which can be further approached by reducing data handling requirements.

Recognizing and tracking appear to require at least three different areas of work, segmentation, labelling, and motion interpretation. Before discussing any of these in depth, an overall strategy should be explained, which is an essentially top-down, or global, approach. Under this, no single component of the objective is to receive special attention or optimization in the first pass. Rather, a minimally competent module for each required function should be implemented as quickly as possible, and all the prototypical modules be made to work together. This is primarily because the least well understood aspect of a modular vision system is the communication between modules, the types of data

structures and control messages that inform and direct the performance of each of the modules. To put it another way, it is not possible to evaluate a single component of a system, an edge detector as a trivial example, in isolation. There are at present no criteria for determining that a particular module is optimal for overall performance of the system, we simply do not understand the coupling between modules well enough.

Related to this lack of organizational knowledge, it is expected that a complete effort will, in succeeding passes, experiment with a number of different techniques fit within the overall structure. There exist several documented techniques for segmentation, for example, in both the edge-detection form and its dual, region growing. There are a great number of region description techniques suitable for fitting to a two-space model, although fitting to three-dimensional models is less well understood. Specific algorithms are proposed in the following for two reasons, they represent the state of the art at the component level, and they appear suitable for adaptation to parallel processing. As a result, the proposed inventory is weighted heavily toward the class of relaxation algorithms, those which appear to rely upon local, semi-cooperative processes. That these happen to coincide with our concept of biological vision is merely fortuitous, their main property is that they may be adapted, in some fashion, to near simultaneous processors. Although each will be referenced individually, they are all drawn from the same source [2] and will be given page numbers in the references.

Segmentation Using Edges

The Prager edge relaxation algorithm [10] essentially amplifies consistent edge phenomena, initially detected as a nontrivial intensity gradient across any pixel. Even weak gradients will be amplified providing that a spatial series of them across many pixels exists. Thus, pure gradient magnitude is a secondary determinant, rather the repetition of a gradient on neighboring pixels determines which edges are amplified and which suppressed.

Although the detection and amplification of edges by this process is indeed highly parallel, the result of it still requires considerable work in order to define regions. Each edge must be traversed, an essentially serial process, to determine closed curves (which thus define a segment), and it will often happen that curves do not close, so that some higher order processing is eventually necessary.

Segmentation By Growing Regions

The line-handling problems of the edge detector may be quite elegantly side-stepped by the device of merging picture regions, which must always maintain well defined boundaries, though this is expensive in terms of storage. An older method [4], suitable for monochrome images, initally treats each pixel as a separate region (segment), and then dissolves boundaries between segments if there is sufficient similarity between regions at the boundary. Regions are maintained in tabular form, recording various statistics such as central moments and intensity distribution statistics, as well as suitable boundary encodings. This is a formidably large data-base, and it tends to remain somewhat constant in size, since as the number of regions declines the length of the boundaries increases.

The adaptation of this procedure to vectorized form has not yet been attempted, even on paper, since adjusting to the unsystematically changing granularity of the dynamic pictorial data base appears to be unsuited to the systematic gridding of a vector processor. However, as described in Appendix B, sufficient other techniques have proven feasible that this may eventually yield as

well, and a proof that it will not, if demonstrated, is in itself a highly informative and useful result.

Segmentation By Thresholding

This is perhaps the most ancient of research techniques, highly useful because the resulting objects, binary images, are by far the most tractable. Further, the camera controller currently being used will threshold automatically, "on the fly", at any arbitrary intensity value, which eliminates a processing step. This has not been considered for two reasons. First, although initial approaches to an orbiting target in fact present an object-versus-background situation, the background in question may be light (earth), dark(space), or composite (horizon in view). This complicates the background definition. Further, it has been argued previously that simple interpretation of an entire target is not likely to be as useful as an analytic, piecewise interpretation of its components. Second, thresholding does not deal at all well with curved surfaces which will, under anticipated lighting conditions, exhibit a range of intensities. Both edge detection and region growing handle such surfaces quite gracefully.

It should be noted that the result of a segmentation, however achieved, is a binary image, so that all the algorithmic advantages remain, and the various regions are more descriptive than would be the case with pure thresholding.

Optical Flow

We anticipate that the process of connecting image pieces (segments) and analysis of scene changes, often termed optical flow, will be intimately connected, in that segments which move together are connected. Two algorithms for optical flow are proposed, and one is the object of careful scrutiny with respect to vectorization. Each method constrains neighboring flow (motion) vectors to be similar in orientation and magnitude, thus enforcing a motion coherence heuristic.

The first, due to Horn and Schunk [6], relates object motion to changes in intensity, using the deceptively simple relationship between time, intensity and position

$$f(x + dx, y + dy, t + dt) = f(x, y, t).$$

A system of constraints results in a problem soluble by Gauss-Seidel relaxation, a well-understood iterative technique quite suited to vector processors. This is one instance of the application of regularization, esssentially the introduction of a stabilizing functional (the coherence assumption) into a minimization calculation, and is widely applied in computer models of early vision.

The second method, due to Barnard and Thompson [3], is more constructive and involves a large series of local cross-correlations at 'interesting' image points. It attempts to relate, subject to a smooth motion constraint, interesting points in two images. An 'interesting' region may simply be a large intensity gradient (an edge or corner, by assumption) or may instead correspond to specified phenomena in a segmentation description. This method is considerably more challenging than the first as a candidate for vectorization but has the advantage that processing is not global over the picture, but rather only in selected portions of the images.

A velocity map is useful in two ways. First, it provides a lot of information prior to recognition and model mapping, independent of the problem domain, concerning range and timing. Avoidanc maneuvers, if necessary, can work from restricted 'objects world' consideration of velocities. Thus, the rate and direction of 'looming' provides quite accurate information about the relative motion of chaser and target. Second, since non-trivial velocities generally correspond

to edges and other detectable features, rather a lot of segmentation work can be achieved in the process of deriving velocities. It appears that relatively competent velocity (flow) derivations can be achieved at considerably reduced resolution, which may in fortunate cases speed up this process by a significant amount.

Data Reduction

The two alternatives sketched in optical flow illustrate a powerful choice, that between global techniques, analogous to image processing transforms, and local, semantically informed techniques. The major challenge of the second sort is to efficiently represent and describe interesting regions, so that the overhead required in restricting computation to those areas is not unduly great. As an initial guess, this is expected to involve heavy use of list structures quite unsuited to the systematic coordinate scheme within vector processors.

Another, highly systematic, form of data reduction involves a variable resolution scheme. This term is generally applied to pyramid schemes, which involve calculating lower resolution pseudo-pixels at a level of the pyramid as an average of several (usually four) pixels at the next lower level. The overhead involved in calculating a full resolution pyramid is substantial and unacceptable for current purposes, but the advantage of it is that considerable work can be done at lower-resolution levels, which cuts processing time considerably. We suggest that, rather than calculate a full pyramid, some small percentage of the image points that samples the scene may perform as well as a true averaged pyramid level. This is the object of an independent study.

Model Matching

The process of matching image segments to model components has not been directly addressed yet, although a candidate model has been developed somewhat fortuitously. The constructive solid geometry (CSG) used to generate test objects is quite adequate for man-made objects, constructed of relatively simple primitive shapes. Further, the absence of streamlining, necessary for aeronautical objects but not astronautical ones, means that very few complex surfaces are likely to be present. In addition to this correspondence, a CSG description is more concise and complete than most surface-oriented models, and it is well structured (a binary tree) for algorithmic manipulation. The model is discussed briefly in Appendix A, and the reader is referred to Roth [12] for a highly readable and competent discussion.

V. Summer 1985 Achievements

Activity during this period concentrated, not by design, upon building an infrastructure. This consisted of the generation of test images, the establishment of data paths between the three involved processors, and acclimatization to the computing environment. One segmentation procedure was successfully implemented in serial form, and analyzed on paper in vector form. These topics are discussed in turn.

Test Images

It is desirable to begin with images that incorporate properties that will exist in the real environment, such as perspective distortion, shadowing, shading, hidden surfaces/objects, and noise. Of all the graphical rendering

techniques, only ray-casting can elegantly incorporate all these phenomena, and transparency, mass properties, and translucency as well. (Noise generation is a different, though routine, matter, and hasn't been necessary yet.) The geometric model that underlies this implementation is constructive solid geometry, and is described, along with ray-casting proper, in Roth [12]. A brief orientation to the implemented software is presented in Appendix A. Ray-casting has been adapted to a vector processor [9], but to do so in this project was felt to be of minor significance, since implementing the serial version (10-30 minutes per image) used up an unexpected amount of calendar time.

Data Paths, Environment

The vector processor, an FPS-5210, is hosted by a PDP-11/570 running RSX-11/M. This is connected via Decnet to a VAX-11/780 running VMS. The connection turns out to be fortunate. Evaluation of performance, particularly of the segmenting procedures, can only be done visually, using pictures, so that some graphic capability is essential. A small half-toning routine enables shaded raster images to be sufficiently well-displayed on the available vector displays, which are various Tektronix terminals, to enable interpretation. This is described in Exhibit 4. Raw pixel intensities are logarithmically adjusted by the half-toning routine so that a satisfying spread of shading levels is achieved. Because of a noisy connection to the 11/570 terminal, all graphics work must be done on the VAX. Raster images generated in the VAX cannot be directly transmitted and used by the 570, so must be put into character form and sent to the 570, where they can then be directly loaded into the FPS. Returning images to the VAX for display requires the inverse process.

Although anticipated image sizes will be 256 x 256, many procedures take long enough during testing that smaller images are desirable. To this end, all raster files (*.ras) are preceded with two numbers describing the dimensionality of the image, and most programs adjust themselves to this information.

While this seems to be an inordinate amount of procedural overhead, it does not directly relate to the target situation at all, where images will go directly from the video controller into FPS memory, and it is certainly acceptable for development and testing purposes.

A further accomodation is necessary in that many algorithms are expressed most concisely in a language that supports dynamic allocation (for lists, queues, etc), data structuring (so that different types may be included in a single object), and recursion (for concise management of list and tree structures). These capabilities exist in Pascal, Ada, PL-1, and C, but of these only Pascal is available, and its treatment of libraries and modularity is extremely awkward. As a result, all implementations are in Fortran.

The Prager Edge Relaxor

Of all the proposed procedures, only this was fully implemented in serial form. Although the serial implementation was quite straightforward, the time performance is disappointing, requiring about five minutes for each cycle, or thirty or so minutes for reasonable convergence on a 256 x 256 image. An illustration of its operation is included as Exhibit 6.

As suggested in a previous section, an edge-based segmentor is inconvenient, in that the results do not directly define segments. This is because a segment must include a closed and unambiguous boundary, and the edges that are produced by this algorithm need further attention before closed boundaries can be assured. The choice of this procedure was due to its always local nature, so that it looked like a good practice candidate compared with region growing.

As an exercise for learning the existing capabilities of the FPS it worked well.

Using The Vector Processor

There are two considerations, how to vectorize algorithms, and how to use the specific processor. The first issue is addressed in Appendix B, the second is treated here.

A first rule, which will not vary, is that once one is into the vector processor one must stay there until finished. A single vector operation on a full picture (65,536 pixels) takes about a tenth of a second. Loading or unloading a picture to host disk (the 570 can't hold an entire picture) requires thirty seconds. This is a general rule for any peripheral accelerator.

A second principle, for development purposes only, is to use only the FPS Math Library routines, and forgo various optimizing possibilities such as the vector function chainer, the arithmetic coprocessor, and direct assembly coding until the need for these becomes apparent. The repertoire of routines is quite rich, if one stays within a single page of the system memory. Routines which work across page boundaries include only a few arithmetic operations. There is the option of coding some of the logical vector routines to work across page boundaries as well, thus avoiding the necessity of working with images in smaller pieces (the 'mosaic' problem), but it appears that in many cases temporary data requirements will be so extensive that subdivision cannot be avoided in any case, so that a direct confrontation with the subdivision-reconstruction problem must be made.

The FPS will perform a simple vector operation such as addition or comparison in about .1 seconds for a 65,536 element vector, which is the size of a picture (256 x 256). A very simple four-point averaging filter can thus execute in under a second (see Appendix B), although more complex operations will, if the full resolution of the picture is used, take on the order of several seconds or more. The speed of the FPS is nearly the same as that of the VAX for floating point operations, so that if the co-processor is used along with the main processor a significant amount of routine work can be offloaded to it.

Further Proposed Work

The summer's effort has barely scratched the surface of the tasks necessary to enable, even in non-real-time, a primitive matching and prediction facility. This was occasioned partly by the amount of effort to generate raster test images (what a video camera 'sees'), and partly by extensive upgrading of equipment, rendering it unavailable.

The tasks necessary to establish a working, though creaky, prototype are essentially of two kinds. The first is to implement a better segmentation algorithm using region growing, and to implement a motion analysis algorithm, essentially a tabulation of differences between two pictures. This set of tasks, while nontrivial, can be expected to require two months of effort. The second kind is considerably more challenging, and deals with matching picture segments to a geometric model. This is considerably eased if we assume that the object in view is already known, so that we know which model specification to work from, but it is still a very challenging opportunity.

Tasks of the first kind appear to adapt well to a parallel processor, those of the second appear not to. These latter will probably, in a first cut, consist of some variation upon (partial) graph matching. Rather little work of any kind has been reported concerning graph matching on a parallel processor, it is more intuitively suited to a symbolic or list-processing environment and perhaps best

implemented there for development purposes. This is not to say that graph matching can't be done 'in parallel', merely that to do so in a general way requires breaking new ground.

Finally, the integration of the pieces discussed is in itself not very well defined, and a control structure that applies the pieces correctly is also likely to be better specified in a linear, symbolic environment. Further, we suggest that analysis would proceed in two rather different modes. One of these would bypass labelling entirely, analyzing a motion sequence entirely in terms of object continuity and apparent motion, a task that can be done at considerably reduced resolution and can cut the cost of a vector operation by a factor of 100 or more, and provide sufficient information for obstacle avoidance and tracking.

Motion analysis itself can provide considerable segmentation information, thus aiding recognition of target components and their labeling. Matching and tracking also can make do with considerably less data than an entire picture, but the data is localized in 'interesting' areas, generally object boundaries the specification of which is a primary objective of segmentation. While motion analysis should provide rapid and continuing results for avoidance and tracking, recognition and labeling may proceed in a piecemeal fashion. A graphical rendition might thus be a simple and coarse outline, tracked by local trajectories. Assuming there is leftover bandwidth, it may be used to identify and correlate scene pieces with the known model, which eventually will generate the displayed scene entirely, and tracking of the target would switch over into a model-oriented scheme. Stability of the tracking can be maintained using only a few regions of later pictures.

REFERENCES

[1] Akin,D.L., M. L. Minsky, E.D. Thiel, and C. R. Kurtzmann, NASA Contractor Report 3734, *Space Applications of Automation,Robotics and Machine Intelligence* , Contract NAS8-34381, October 1983

[2] Ballard, Dana H. and Christopher M. Brown, *Computer Vision* , Prentice-Hall, Englewood Cliffs NJ, 1982

[3] Barnard. S. T. and W. B. Thompson, "Disparity Analysis Of Images", Technical Report 79-1, Computer Science Department, Univ. Minnesota, January 1979 ([2],208-210)

[4] Brice, C. and C. Fennema, "Scene Analysis Using Regions", *Artificial Intelligence 1* (3):205-226, Fall 1970 (described in [2], pp. 158-159)

[5] Freitas, R.A. and W. P Gilbreath (eds), *Advanced Automation for Space Missions* , NASA Conference Publication 2255, 1982

[6] Horn, B. K. P., and B. G. Schunk, "Determining Optical Flow", AI Memo 572, AI Lab, MIT, April 1980 (described in [2] pp. 102-105)

[7] Iverson, Kenneth E., *A Programming Language* , John Wiley 1962

[8] Nazif, A. M. and M. D. Levine, "Low Level Image Segmentation: An Expert System", *IEEE Transaction on Pattern Analysis and Machine Intelligence 6* (5):555-574, September 1984

[9] Plunkett, David J. and Michael J. Bailey, "The Vectorization Of Ray-Tracing", IEEE Computer Graphics and Applications *5* (8):52-60, August 1985

[10] Prager, J. M.,"Extracting and Labeling Boundary Segments In Natural Scenes", IEEE Transactions on Pattern Analysis and Machine Intelligence *2* (1):16-27, January 1980 ([2], 85-88)

[11] Requicha, A. A. G.,"Representations For Rigid Solids: Theory, Methods, and Systems", *Computing Surveys 12* (4), December 1980

[12] Roth, Sott D.,"Ray Casting For Modelling Solids", *Computer Graphics and Image Processing 18* ,109-144 1982

[13] Tanimoto, S. and T. Pavlidis, "A Hierarchical Data Structure For Picture Processing", *Computer Graphics And Image Processing 4* (2):104-119, June 1975

[14] Vinz, Frank L., Linda L. Brewster, and L. Dale Thomas, *Computer Vision for Real-Time Orbital Operations* , NASA Technical Memorandum TM-86457, August 1984

APPENDIX A - Raycasting And Constructive Solid Geometry

Raycasting is a deceptively simple computer graphic rendering technique, which consists of mathematically constructing a line from the viewer through each of the viewscreen coordinate positions, and determining what part of a scene or model each hits. If the object so hit is solid, then if its surface normal can be determined at that point its shading (knowing the direction from which light is coming) may be determined, thus determining the desired intensity for that pixel on the rendering. This may be considerably expanded. Another ray may be cast from the point of intersection toward the light source, and if it should hit some object in the scene then the pixel through which the original ray was cast should be darkened to represent shadowing at that point. Further, if the object should be transparent, and possibly refractive, further rays, appropriately bent, may be cast from the point of intersection to determine just what light the viewer should see at that point.

Raycasting in and of itself is rather straightforward and lends itself to practically all forms of light modelling. Of more challenging interest is the determination of intersection points, or how we know where some ray actually hits. For arbitrary objects represented by some volumetric technique such as oct-trees this can be quite exhaustive. We have chosen a more tractable means of constructing objects from simple geometric solid primitives which is known as constructive solid geometry (CSG). This and alternative models are discussed in an extensive review by Requicha. [11] Transformed primitives may be combined using the set operations of intersection, union, and differencing, and quite complex objects may be so constructed. This appears to be a useful model for orbiting manmade objects, since they tend to comprise relatively simple forms in combination. Some moderately simple examples are shown below. They are copied from Roth [12], to which the reader is referred for a highly competent treatment of raycasting and CSG.

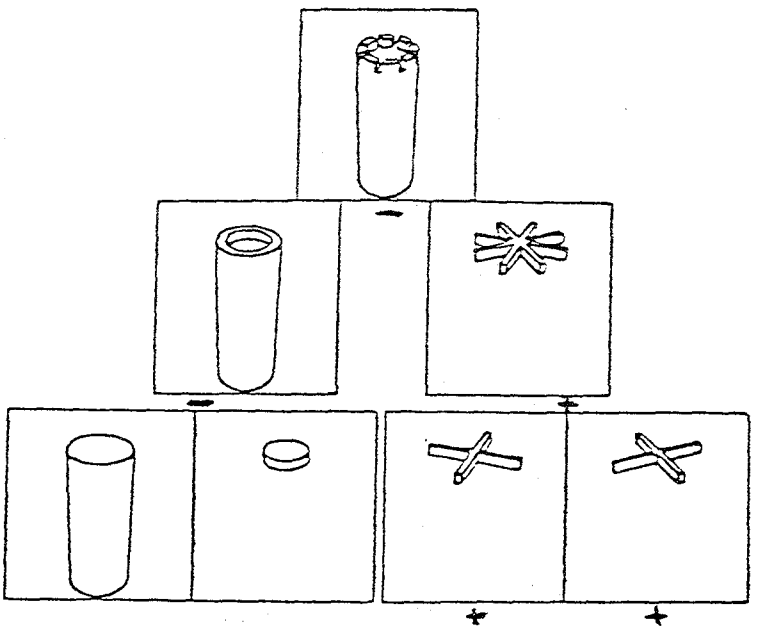


FIG. 2. Example of composition tree.

Current Implementation

A reasonably competent CSG model works with five primitives, the cube, the sphere, the cylinder, the cone, and the torus. This effort has implemented only the first three, although addition of the other two (and, indeed, any other formally tractable object) is reasonably straightforward and truly additive to the existing software. Only matte lighting (single light source) and shadowing are currently implemented. As well, monochrome (colorless) objects alone are treated. These properties are sufficient to represent a useful range of targets.

An object is constructed as a binary tree, specified in pre-order (root, left child, right child), where every node is of degree two or zero. All interior nodes are combinations (intersect, union, difference), and are of degree two. All leaf nodes are primitives and are of degree zero. At any node one or more geometric transformations may be applied which will affect that node and all its children, if any. These include translation, scaling (stretching), and rotation.

The current program reads a file of extension .csg, which consists of comment lines (begin with semicolon), node names (in quotes), or transformations (again, in quotes) to be applied to the most recently encountered node (and its children). Each transformation specifies three parameters. A primitive telescope file specification is shown in Exhibit 3. The object is shown in three different views in Exhibit 5, in which object outlining and shadowing are selectively illustrated. The output of the raycaster is a raster file which is fully competent for display on a raster graphic terminal. Since none is available locally, the single intensity dot-drawing capability of the vector terminal is utilized to make a primitive half-tone rendering, described in Exhibit 4.

The raycasting technique competently renders smoothly varying shading, shadowing, and horizon effects, all of which will present challenges to any scene interpretation procedure, and it does so in a concise and unified manner. Specification of an object is considerably more concise and intuitive in CSG than in a surface patch model, the other widely-used alternative. Further, it is a truly three-dimensional model, entirely formal, and thus a highly suitable candidate for matching to two-dimensional scenes.

The current program reads an object specification from FOR010.DAT, into which one has previously copied a *.csg file, accepts specification for boundary lines, shadowing, and display size (default of 256 x 256), and in 5 to 20 minutes creates a file in FOR008.DAT. If the file is acceptable, it should be copied to a *.ras file. The program rplay will render the contents of FOR008.DAT to a Tektronix using the ddot routine.

APPENDIX B - Vectorization - An Initial Cut

At the outset specific recognition should be given to existing vector languages, of which only APL is truly deserving of the name. The techniques and operations described in its original presentation (Iverson [7]) have been embodied substantially in the design of certain vector architectures such as the Cyber 205, and to an unfortunately lesser extent in the FPS Math Library.

All vector processors, and parallel Pascal and PL-1 as well, concisely express element-by-element operations. Thus, the statement

C = A + B (in PL-1) or c <- a + b (in APL)

will assign the sum of the i'th elements of A and B to the i'th element of C. It is assumed, implicitly in APL and by specification in PL-1, that A,B and C are all of the same magnitude, otherwise the operation is undefined.

Aggregate operations, such as sum, maximum, or mean of vector elements are treated in PL-1 with library functions. In APL basic reduction operators are used. Thus,

X = SUM(A) (in PL-1, X a scalar) is x <- +/a (in APL)

where / is the reduction operator.

A capability of APL (and the FPS as well) is the logical vector. While IF A > B THEN... is undefined in PL-1 for vector operands, the operation z <- a > b in APL creates a vector z of elements zero and one, of same magnitude as a and b, where z(i) is one if a(i) > b(i), zero otherwise.

To this point the FPS library embodies all these functions. The departure is selection/compression. In APL, a logical vector may be used to select elements from another arbitrary vector (of same length), producing a new and generally shorter vector consisting only of those elements of the operand vector that corresponded to 1's in the selection vector. This sort of arbitrary selection is endemic to a lot of vision routines, and is so far the only routine that we would implement outside the math library.

The image must in all cases be processed in pieces, because most procedures create too much temporary data to be held by the four pages of the FPS. A 256 by 256 image occupies exactly one page. The Prager edge amplifier is a global/local technique, in that operations on a pixel depend only upon neighboring pixels (local), and all pixels may process independently and in parallel. This routine works, in fact, upon the edges between pixels, of which there are twice as many, so that two pages are required to hold edges. One step in the Prager amplifier is, for each edge, to determine at each end whether 0, 1, 2, or 3 other edges abut the edge (recall that each edge has three neighbors at each end), based on their strengths. From this information the strength of the edge will be raised, lowered, or left alone. A part of this operation requires ordering the strengths of the three neighbors, a small sort, and simply to do the sort requires temporarily maintaining over six numbers for each edge, which clearly outstrips memory capacity.

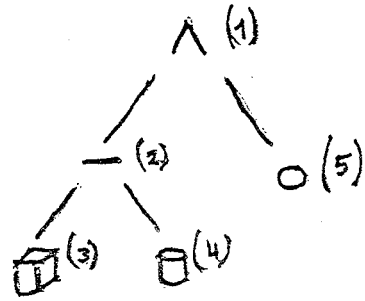
Thus, operations must necessarily be performed on subdivided images (the mosaic problem) if one wishes to retain the entire picture. This applies as well to the Horn and Schunk optical flow routine, which requires a permanent data structure of six to eight pages exclusive of temporary memory. While the mosaic problem is a major irritant, the more fundamental objection to global processing is its speed. A single step of an iterative process on the whole image will take many seconds, which is clearly too slow.

The most satisfying route through the timing impasse is some form of data reduction, already discussed in the body of this report. Two forms are considered. The first is to work at reduced resolution, either by aggregating (averaging) into a resolution pyramid (Tanimoto [13]) or simply selecting systematically spaced pixels into a smaller picture. Working at reduced resolution serves to identify areas of the picture which require further, high-resolution analysis.

The second is similar, in that it begins by identifying interesting areas (cf. the Barnard-Thompson optical flow procedure), and then works only within them. The major challenge in either case is that procedures at this level work by updating tables and lists, a form of computing that departs substantially from traditional vector problems. This is currently under intense scrutiny.

```
ty simp.csg
; A simple object-- that part of a thick board that lies
; within a sphere. The board has a hole in it. The calculus
; is intersect big sphere with (board minus cylinder).
1 { 'int'
    'sca' 100 100 100     blow up to "world" size
    'rot' 20 20 0          and tilt some so we can see hole
  ;
2 [ 'dif'  left subtree, board(flat cube) minus thin cylinder
3 { 'cube'
    'scale' 1 .2 1      flatten it in y axis
  { 'cylinder'          this makes hole in board
4 { 'scale' .3 1 .3
    'translate' .25 0 -.2  not in center of board
  ;
5 { 'sphere'     right subtree, big sphere
    'sca' 1.1 1.1 1.1      leave some straight edge on board
    'end'
```

(a) A file (simp.csg) that describes a simple, three-primitive object, with two connectives.
The tree at the right is numbered to show correspondence.



(b) Running the ray-caster. Boundary lines are only for easier visualization, not test images.
Shadowing should be specified for test images.

```
cop simp.csg for010.dat
$ run rayc
enter non-zero if display run
0
type nonzero if boundary lines explicit
1
 type non-zero for shadowing
0
  enter size (square) if not 256
130
```

(c) The file (simp.ras) may be fed to program rplay, which makes this picture.

What simp.csg makes

Exhibit 2 - A Simple CSG Specification, and The Result Of 'Playing' It On A Vector Graphic Terminal.

```
;       a rudimentary space telescope, based on a picture
;       in the hall.
;       Three parts to this, main barrel, funny little flap
;       at end, and the solar panels.
;
;       mcan 26 Aug 85
;
'uni'
'sca' 50 50 50
'rot' 0 -50 0
'rot' 40 0 0
'rot' 0 -10 0
'tra' 0 0 30
;************ you fly this thing using previous transform
;
;       that was the main global transform (whole object)
;       all the following are relative to origin
;
;       main barrel
;
'uni'
'cyl'
'sca' 1.5 3 1.5
'uni'
;
;       these are two end collars
;
'cyl'
'sca' 2 1.5 2
'tra' 0 -1.5 0
;       second collar is hollow
'union'
'tra' 0 1.5 0
'cyl'
'sca' 2 1.5 2
'cyl'
'sca' 1.7 2 1.7
;
;       funny little flap next
;
'union'
;       well actually union of flap and panels
'union'
'rot' 90 0 0
'tra' 0 3 .5
'cyl'
'sca' 2 .1 2
'cub'
'sca' 2,.1,1
'tra' 0 0 -.5
;
;       ok, here's the panels
;
'union'   of two slabs and a stick
'rot' 45 0 0
'cyl'
'sca' .05 4 .05
'rot' 0 0 90
;       now two slabs
'uni'
'cub'
'sca' 2 5 .05
'tra' 2 0 0
'cub'
'sca' 2 5 .05
'tra' -2 0 0
'end'
```

Exhibit 3 - A Constructive Solid Geometry Specification of Primitive Telescope

The tree is specified in pre-order form. Lines beginning with ; are comments. Leaf nodes are 'cube', 'sphere', and 'cylinder' ('cub','sph','cyl'). Internal nodes are 'union','intersection', and 'difference'.

Any node may be followed with unlimited number of transformations, which are applied in order. The line 'end' is necessary at present to terminate tree.

```
c
c       ddot(ix,iy,den) - mcan 3 jul 85
c
c
c       fix 19 jul - use log map
c
c       half-tone dot drawer- draw only if supplied
c       density is below next random no.
c
c       the sense would be reversed if we were
c       drawing white on black
c
      subroutine ddot(ix,iy,den)
      integer*4 seed
      real e
      data seed,e/38517243,2.71828/
      d = log10 ( (den*(49) + 1) )/ log10(50.)
      if ( d .lt. ran(seed)) call pntabs(ix,iy)
      return
      end
```

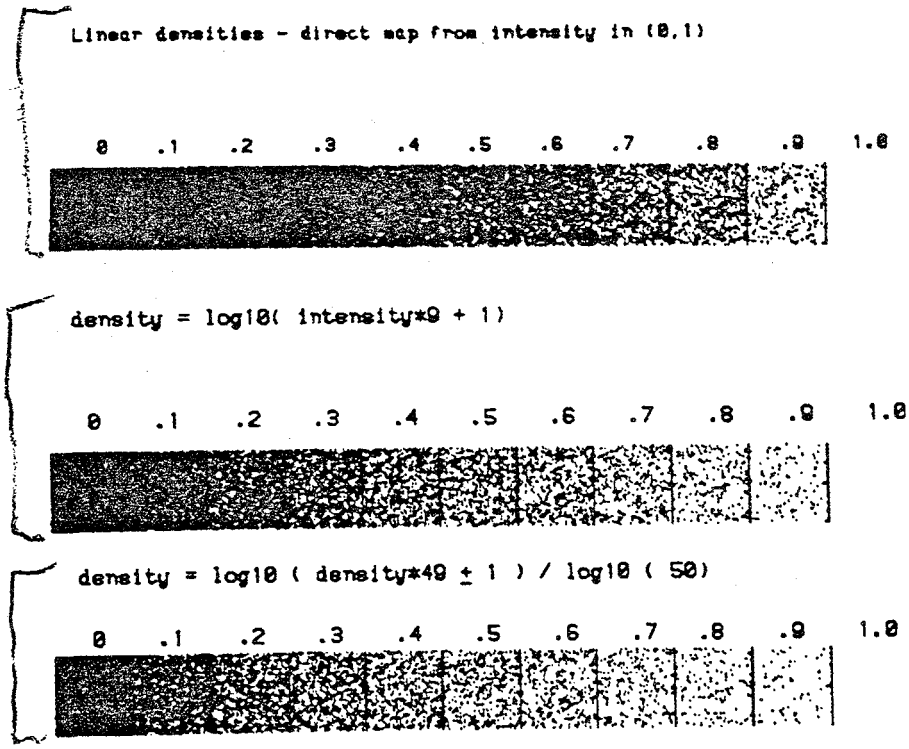
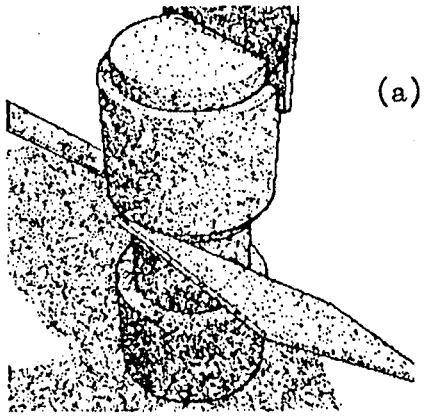


Exhibit 4 - Raster Graphics On A Vector Terminal

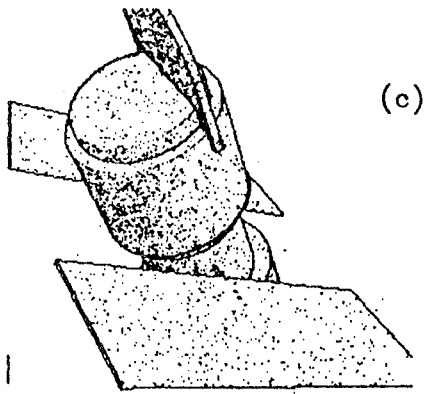
It is much easier to check a picture program by eye than by data listing. The routine ddot() listed at the top illustrates how pixel-independent half-toning works. pntabs() will draw a single dot, based on the supplied intensity and the next random number.
Several logarithmic transformations were tried, the last judged by eye to provide best separation on plotter.



(a)

View of Primitive Telescope, Lines Added

Same telescope, twisted to its right and omoving away from you. Also, boundary lines not added.



(c)

A third view, tilted toward and twisted a bit more, and father away. All this flying is done with the 'rot' and 'tra' lines at top node, nothing else. Just six numbers.

Exhibit 5 - Three Views Of A First-Draft Impressionistic Telescope

The raycasting program offers the options of boundary lines between different primitives, and the casting of a feeler for shadows. The three views here have the target moving away from you, turning to its right, and dipping toward you. The first view uses shadows and bound lines, the second shadows only, and the third bound lines without shadows. On a raster display the boundary lines would not be necessary.

Same telescope, twisted to its right and omoving away from you. Al;so, boundary lines not added.

Zeroith Pass - Raw Differences

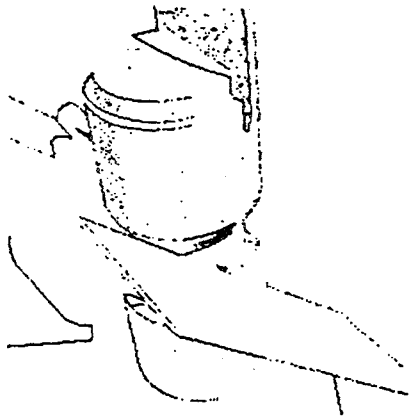
2828 148 3768

First Adjustment Pass

577 270 3699

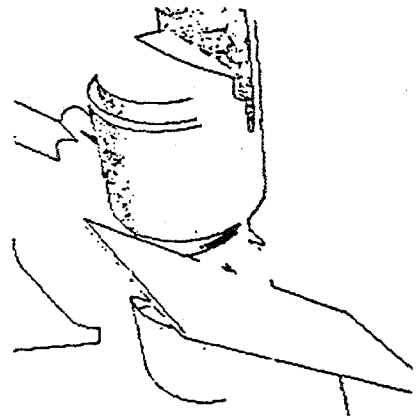
Second Adjuistment Pass. Shadow lines show best, since mostly on light background, hence began with strongest gradient

771 266 3312



Third Adjustment Pass

201 306 2280



Fourth Adjustment. Note right hand corner of distal panel visible for first time.

Exhibit 6 - The Prager Edge Amplifier At Work

This begins with the same view of the previous Exhibit, (b), with shadows but no boundary lines. The three numbers are single edges that have been downgraded, left alone, and upgraded during the pass.

1985

NASA/ASEE SUMMER FACULTY RESEARCH FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER

THE UNIVERSITY OF ALABAMA AT HUNTSVILLE

A STUDY OF 35-GHZ RADAR-ASSISTED ORBITAL MANEUVERING VEHICLE/SPACE TELESCOPE DOCKING

Prepared by:	Malcolm W. McDonald, Ph.D.
Academic Rank:	Associate Professor
University/Department:	Berry College Department of Physics
NASA/MSFC: (Laboratory) (Division) (Branch)	 Information & Electronic Systems Computers and Communication Communication Systems
MSFC Counterpart:	E. H. Gleason
Date:	August 16, 1985
Contract Number:	NGT-01-008-021 The University of Alabama at Huntsville

A STUDY OF 35-GHZ RADAR-ASSISTED ORBITAL MANEUVERING SYSTEM/SPACE TELESCOPE DOCKING

By

Malcolm W. McDonald
Associate Professor of Physics
Berry College
Mount Berry, Georgia

ABSTRACT

An experiment was conducted to study the effects of measuring range and range rate information from a complex radar target (a one-third scale model of the Edwin P. Hubble Space Telescope). The radar ranging system was a 35-GHz frequency-modulated continuous wave unit developed in the Communication Systems Branch of the Information and Electronic Systems Laboratory at Marshall Space Flight Center. Measurements were made over radar-to-target distances of 5 meters to 15 meters to simulate the close distance realized in the final stages of space vehicle docking.

The Space Telescope model target was driven by an antenna positioner through a range of azimuth and elevation (pitch) angles to present a variety of visual aspects of the aft end to the radar. Measurements were obtained with and without a cube corner reflector mounted in the center of the aft end of the model. The results indicate that range and range rate measurements are performed significantly more accurately with the cooperative radar reflector affixed. The results further reveal that range rate (velocity) can be measured accurately enough to support the required "soft" docking with the Space Telescope.

ACKNOWLEDGEMENTS

I am grateful to Marshall Space Flight Center, NASA, ASEE, and The University of Alabama at Huntsville for the opportunities provided by the Summer Faculty Research Fellowship Program. The privilege of working with NASA personnel on problems of mutual interest is stimulating, challenging, and gratifying. The summer program was made valuable by the capable leadership and coordination of Dr. Gerald Karr, Dr. James Dozier, and Mr Leroy Osborne.

Special thanks is given to my NASA counterpart, Mr. Ed Gleason, and my associate NASA counterpart, Mr. Lee Malone, for their full encouragement and support at all times. I salute them both for their refreshing latitude of ideas from which I was able to focus on this chosen topic of study.

My indebtedness to the personnel of the Communications Branch who have exercised roles directly supporting my work on this project is so great as to make me humble. Those individuals include, in addition to those already mentioned, Mr. Grover Tucker, Mr. Steve Jones, Mr. Ed Martin, Mr. Dave Harris, Mr. Reggie Inman, and Mr. Allen Duncan. I thank you all.

Indispensable assistance was rendered by a very capable and clever computer programmer, Mr. Randy Kirby. His skills enabled automation of the collection, analysis, and graphical plotting of the experimental data. This resulted in an incalculable savings of time and effort for me. I extend my sincere thanks.

LIST OF FIGURES AND TABLE

I. INTRODUCTION

The Edwin P. Hubble Space Telescope (ST), illustrated in Fig. 1, is due to be launched into orbit from the shuttle in approximately one year. After approximately ten years in service it is scheduled to be retrieved and returned to the shuttle, the space station, or earth for refurbishing. Then it will be re-boosted into orbit for another period of use.

The vehicle planned for use in rendezvous, docking, and retrieval of the ST is the Orbital Maneuvering Vehicle (OMV). A concept drawing of the OMV is shown in Fig. 2. The OMV will be carried aloft and released by the shuttle to participate in the rendezvous and docking procedure.

Studies have led to the establishment of maximum dynamic parameters for the OMV/ST docking encounter, to prevent excessive agitation of the ST (Ref. 3). Pertinent docking limitations are listed in Table I. The work reported in this paper was designed to test the ability of a 35 GHz radar system, assumed mounted on the OMV, to measure range and range rate (velocity) information with an accuracy sufficient to support a "soft" docking between the OMV and the ST. The study focuses on the evaluation of a need for a cooperative radar target aid (cube corner or other high radar cross section reflector) to serve as a well-defined range reference on the ST aft end.

Docking simulation studies recently reported at Marshall Space Flight Center revealed that the most capable human pilot in the study performed the most accurate docking of the OMV with the ST when armed with range rate information in addition to the visual information provided by a video camera (Ref. 4). This seems especially necessary because of the low limiting closing velocity of slightly over 3 cm/sec at the point of contact.

II. RADAR SYSTEM

The radar system used in this work has been adequately described in previous papers (Refs. 1 and 2). It operates in a frequency-modulated continuous wave mode at a center frequency of 35 gigahertz with a transmitted power level of 5 milliwatts. The radar uses a single transmit/receive parabolic dish antenna with a 15 cm diameter and a 3dB-to-3dB beamwidth of 4 degrees. As a result, the ST model target subtended an angle larger than the beamwidth at all the ranges studied (5 to 15 meters).

Previous experimentation in the laboratory with cube corner targets had revealed the radar range measuring accuracy on the

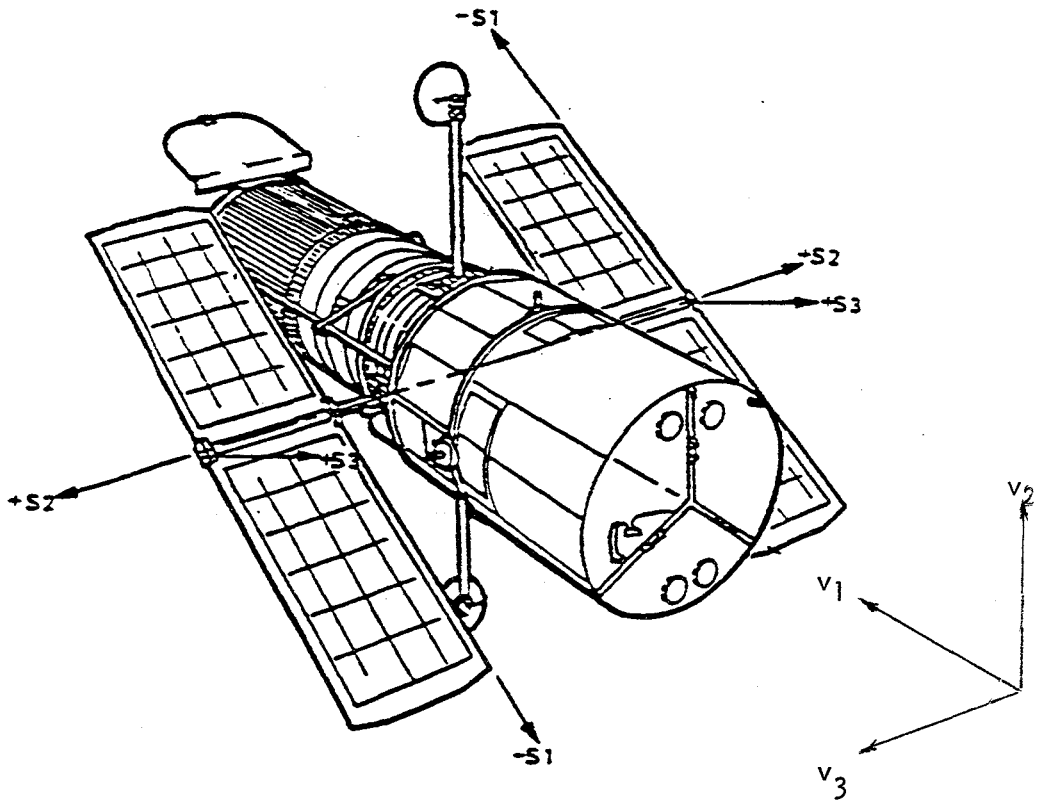


Figure 1. The Edwin P. Hubble Space Telescope. The vehicle coordinate system (v_1,v_2,v_3) is referred to in the text. The v_2 axis is called the "azimuth" axis and the v_3 axis is called the "pitch" axis.

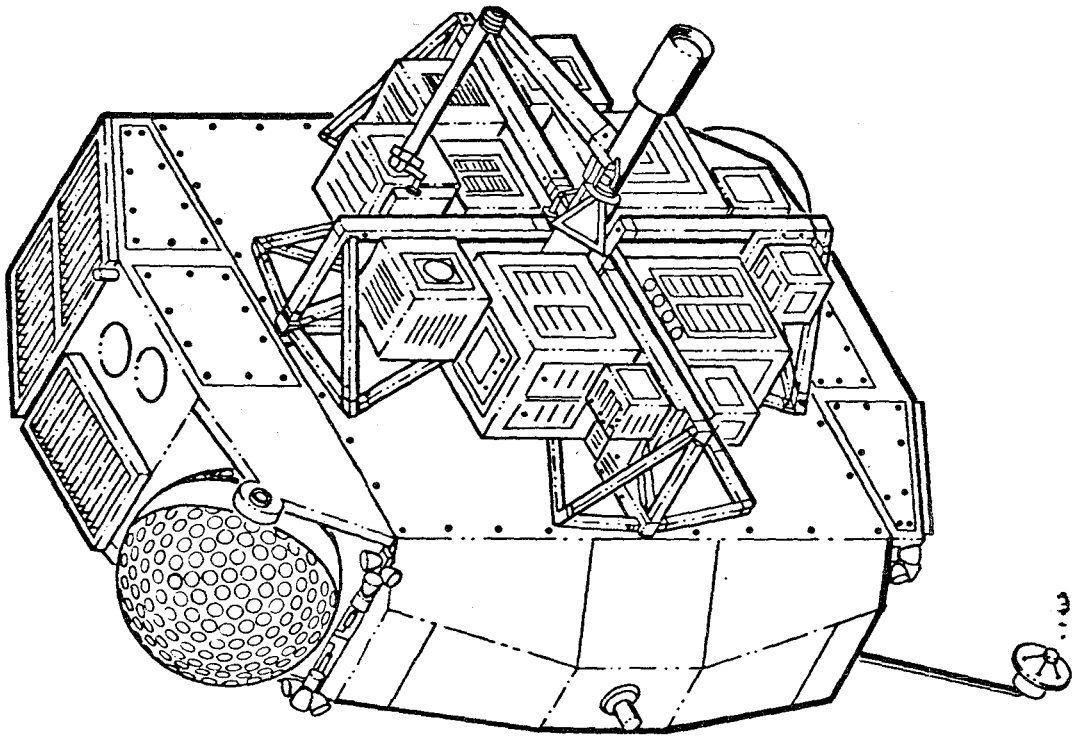


Figure 2. A Concept of the Orbital Maneuvering Vehicle (OMV).

OMV-ST MASSES AND LOAD LIMITS

ST MASS	10800 Kg
OMV MASS (NOMINAL).....	4000 Kg
MAX. IMPACT FORCE ON *ST*	2320 N
MAX. *ST* ACCELERATION ..	0.023g
MAX. SOLAR ARRAY ACCEL.	
(a) Sustained	0.002g
(b) Impulse	0.005g
MAX. *OMV-ST* IMPACT VEL.	3.05 cm/s

TABLE I. Parameters Pertinent to OMV-ST Docking. Note that the key limiting parameter is the maximum OMV-ST closing velocity of slightly more than three centimeters per second.

order of Ø.1% to Ø.2% of range for ranges up to 5 meters. The current work looked at the ability of the radar system to range off a complex target, the aft end of a one-third scale model of the ST. That will be the docking location for the OMV.

III. SPACE TELESCOPE MODEL TARGET

A one-third scale model of the ST, available at the antenna range at Marshall Space Flight Center, was used in this experiment. The model measured 4.3 meters in length and the aft end measured 1.4 meters in diameter.

The model was mounted on an antenna positioner at one end of the 4ØØ-foot antenna range. The model could be rotated by the positioner about a horizontal axis (S2 or V3 in Fig. 1) extending through the two solar array support arms. That angular rotation is referred to as the "pitch" angle. The positioner was able to rotate the model about a vertical axis. That rotational displacement is referred to as "azimuth".

The term "aspect" is taken to mean the prevailing combination of azimuth and pitch as the radar views the aft end of the ST model. For reference, the normal aspect was defined as Ø degrees in both azimuth and pitch when the radar viewed along the V1 axis of the model (See Fig. 1).

The aspects of the model viewed by the radar were limited to azimuth variations from -2Ø degrees to +2Ø degrees coupled with pitch angles from Ø to +2Ø degrees. Those pitch angles represent the aft end elevated above the horizontal. Negative pitch angles were not studied as they were assumed to image approximately what was observed at the positive angles. Also, pitching the model up tended to reflect radar energy up into the sky rather than toward the ground. The negative pitches would have allowed some multipath radar returns, not typical of the free space environment in which the docking will take place, which were eliminated by the use of the positive pitch angles.

IV. EXPERIMENTAL PROCEDURE

The radar system was mounted in a movable gantry at approximately the same height as the pitch axis of the model. The gantry was mounted on flat rails and was driveable either toward or away from the target at a constant speed in either direction. The average speed of the gantry, and thus the radar, was measured accurately over a ten-meter distance to be 13.3 cm/sec as it moved toward the target. In fact, the radar moved along at the top of the gantry elevated some 6 meters above the rails with a jerking

or oscillatory motion, producing instantaneous speeds above and below the measured average speed.

The radar unit was mounted in the gantry on a pan and tilt turret which allowed the antenna boresight to be guided during the course of the experiment. A video camera mounted alongside the radar and pointed in the same direction provided a view of the target on a monitor. The radar was steered manually to keep the boresight pointed toward a reference point at the center of the aft end of the ST model.

The experiment consisted of range measurements with the gantry at rest at nominal target distances of 5, 10, and 15 meters while the target was swept from -20 degrees to +20 degrees with a fixed pitch of 0, 5, 10, 15, or 20 degrees. The "nominal" distances were the actual distances from a reference point at the center of the target aft end to the center of the radar dish for normal aspect. The actual range varied somewhat throughout an azimuthal sweep.

All the range measurement runs were performed first without, then with, a cube corner reflector affixed at the center reference point of the ST model aft end. The cube corner measured 15.2 cm along each of its mutually perpendicular edges and presented a maximum radar cross section of 14.7 dB above one square meter.

Velocity measurements were performed by moving the gantry along its track toward the target both with and without the cube reflector in place. Additionally those runs were performed both with, in one instance, the aspect of the model randomly varying as the azimuth and pitch were manually driven by a human operator in a range of values + or - 5 degrees on either side of the normal aspect during the run, and secondly, without the model executing any change in position, rather maintaining a constant attitude. In the latter runs the target was in the normal aspect position.

The experimental runs were performed under the control of a cleverly written program executed on a Hewlett-Packard model 9836 computer. The program controlled the azimuthal sweep of the model during the range measurement runs. It acquired radar data at a rapid sampling rate and at one-degree azimuth increments calculated and stored values of range, range rate, theoretical range, theoretical range rate, and real time. The so-called "theoretical" range and range rates were computed from equations based upon the prevailing geometry of the target-radar system and the trigonometric relations of the aspect angles and their rates of change.

During the velocity measurement runs the computer sampled the range and real time clock at a rapid rate. It computed the

velocity from the changes in those data and stored the results in a sliding array. The average value in the sliding array was computed and stored as the velocity value representative of each one second time increment to provide a second-by-second history of the velocity of the radar relative to the ST model. The range data was handled in a sliding array in a similar manner.

V. DATA ANALYSIS AND RESULTS

The results of the nominal 5-meter range measurements are illustrated in Figs. 3-7. Figure 3 demonstrates large fluctuations in measured range at pitch angles up to 1Ø degrees when the radar was viewing the ST model without the cooperative target aid affixed. The wild fluctuations were determined to be due to enhanced multiple reflections (radar antenna to target, back to gantry structure, back to target, and back into the radar antenna) at certain viewing aspects, effectively doubling the perceived range in those instances.

Although the multiple reflection spurious behavior resulted from an unintended interaction of the radar signal with the gantry structure, it was immediately realized that a similar interaction could be expected to occur at close range between the ST and the OMV. Thus it simulated a very real condition to be anticipated.

In Figs. 3 and 4 the range scale was chosen to illustrate the gross results at all five pitch angles studied. As indicated, the range scale is noncontinuous, and each unit on the scale is representative of 2.5 meters. In all the range graphs presented in Figs. 3-17 the dotted curves are "theoretical" range as explained earlier. The dotted curves serve as handy references in spite of their minor inaccuracies stemming from differences between the actual geometry of the experiment and the assumed geometry in the equation for calculating the "theoretical" range.

Figure 4 illustrates the improved ranging capability when the cube corner reflector is affixed to the target. Figures 5-7 highlight comparisons with and without the target aid at pitch angles of Ø, 1Ø, and 2Ø degrees, respectively. Note the range scale is expanded by a factor of ten to emphasize the comparison.

Figures 8-12 display similar kinds of data for a nominal range of 1Ø meters. At this increased distance the spurious range measurements were observed over a smaller set of pitch and azimuth angles, as might be expected. Again the measurements are improved by the presence of the cube corner reflector.

The fifteen-meter nominal ranging results are shown in Figs. 13-17. Note in the interpretation of the last three graphs

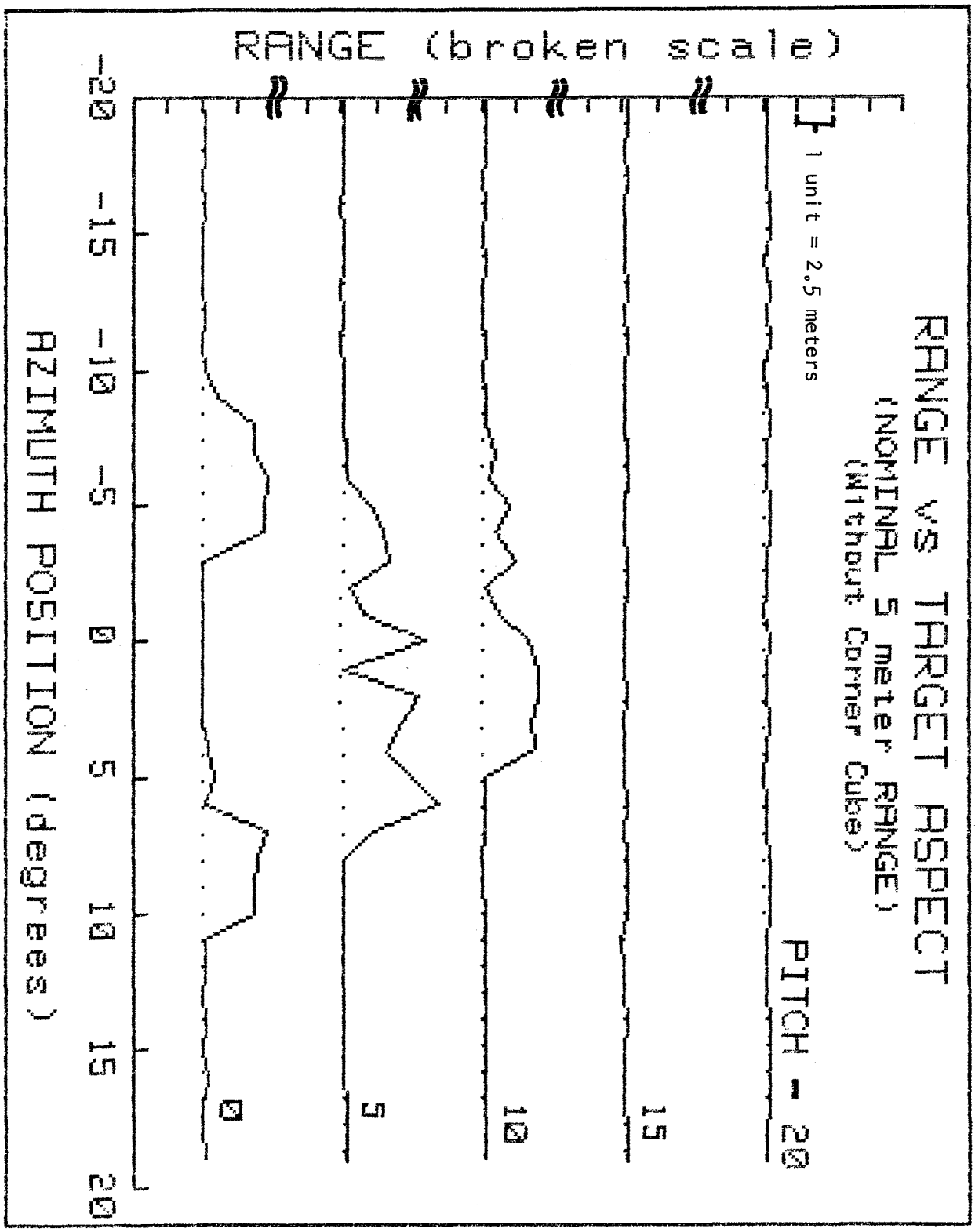


Figure 3. Range vs. Target Aspect (Nominal Range of 5 meters; WITHOUT a Cube Corner Reflector)

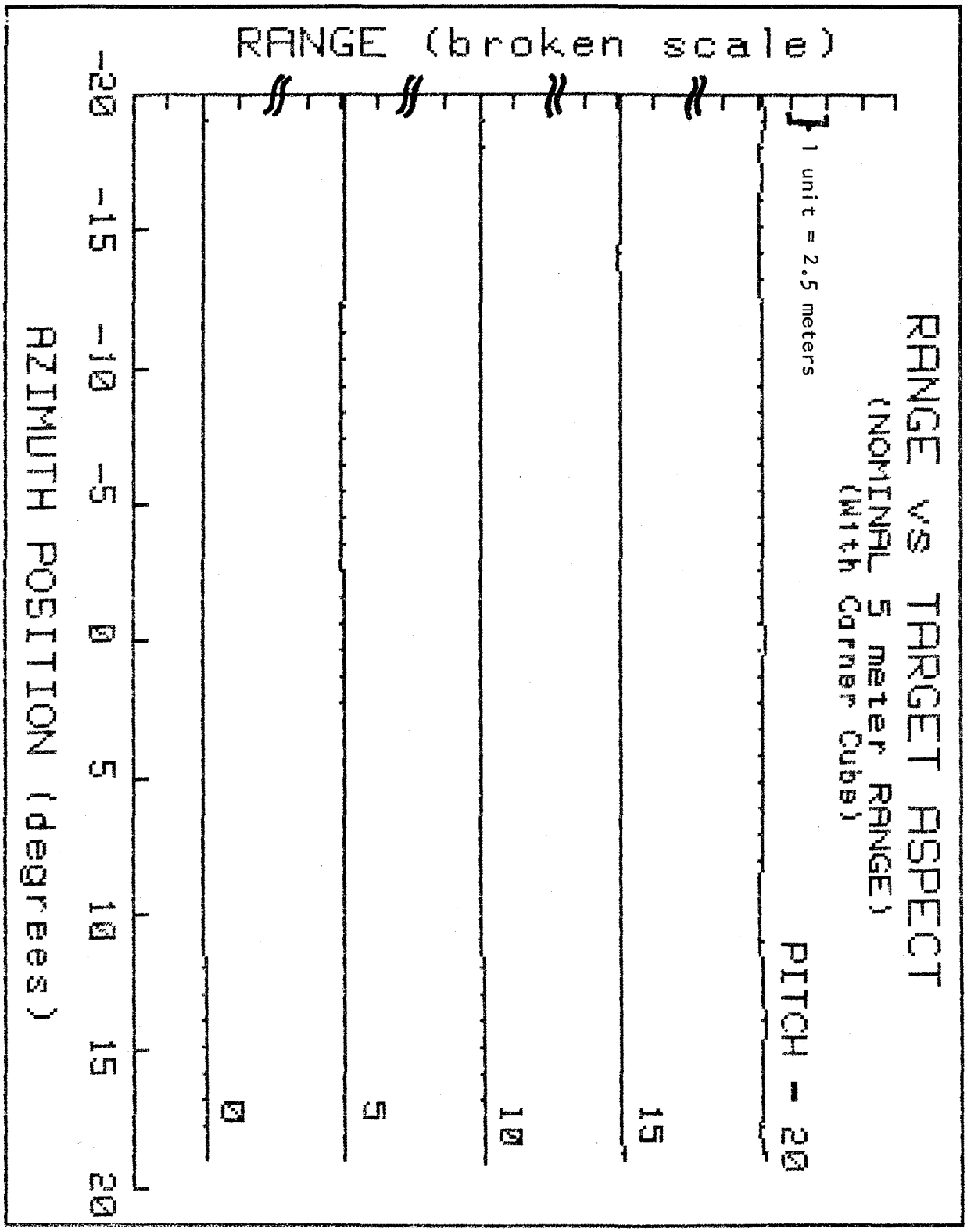


Figure 4. Range vs. Target Aspect (Nominal Range of 5 meters; WITH Cube Corner Reflector)

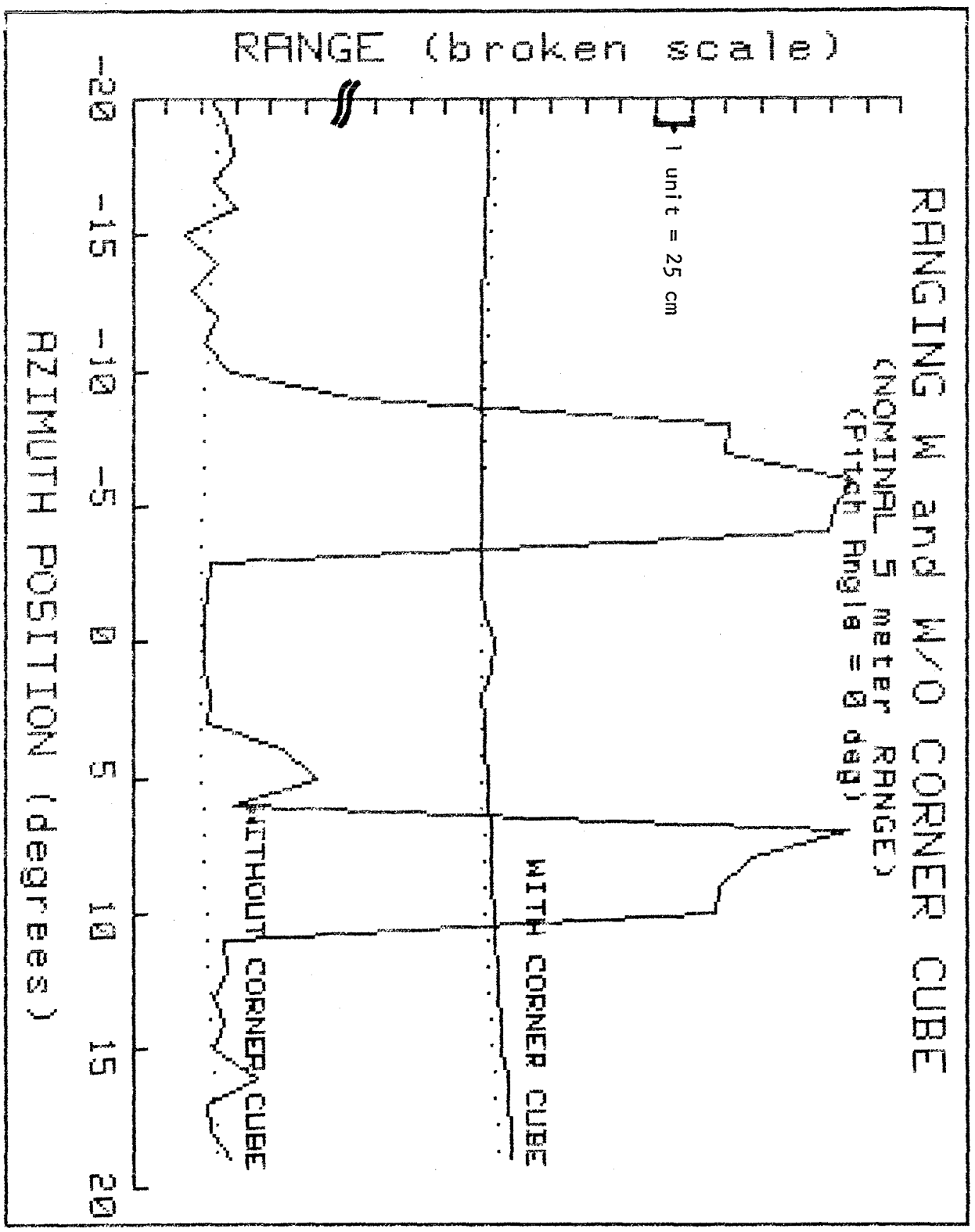


Figure 5. Five-Meter Ranging With and Without Cube Corner Reflector. (Target Pitch Angle = 0 Degrees)

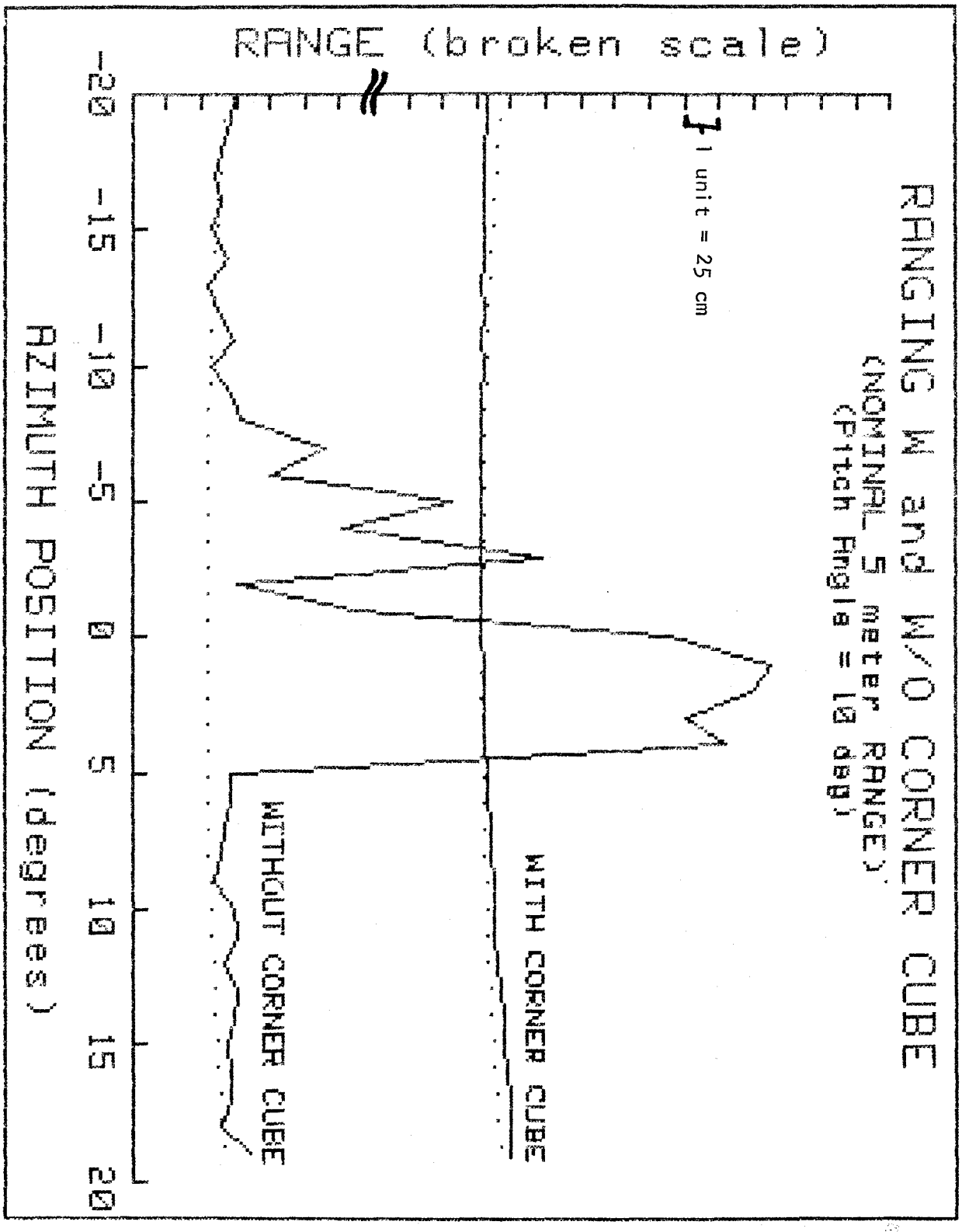


Figure 6. Five-Meter Ranging With and Without Cube Corner Reflector. (Target Pitch Angle = 10 Degrees)

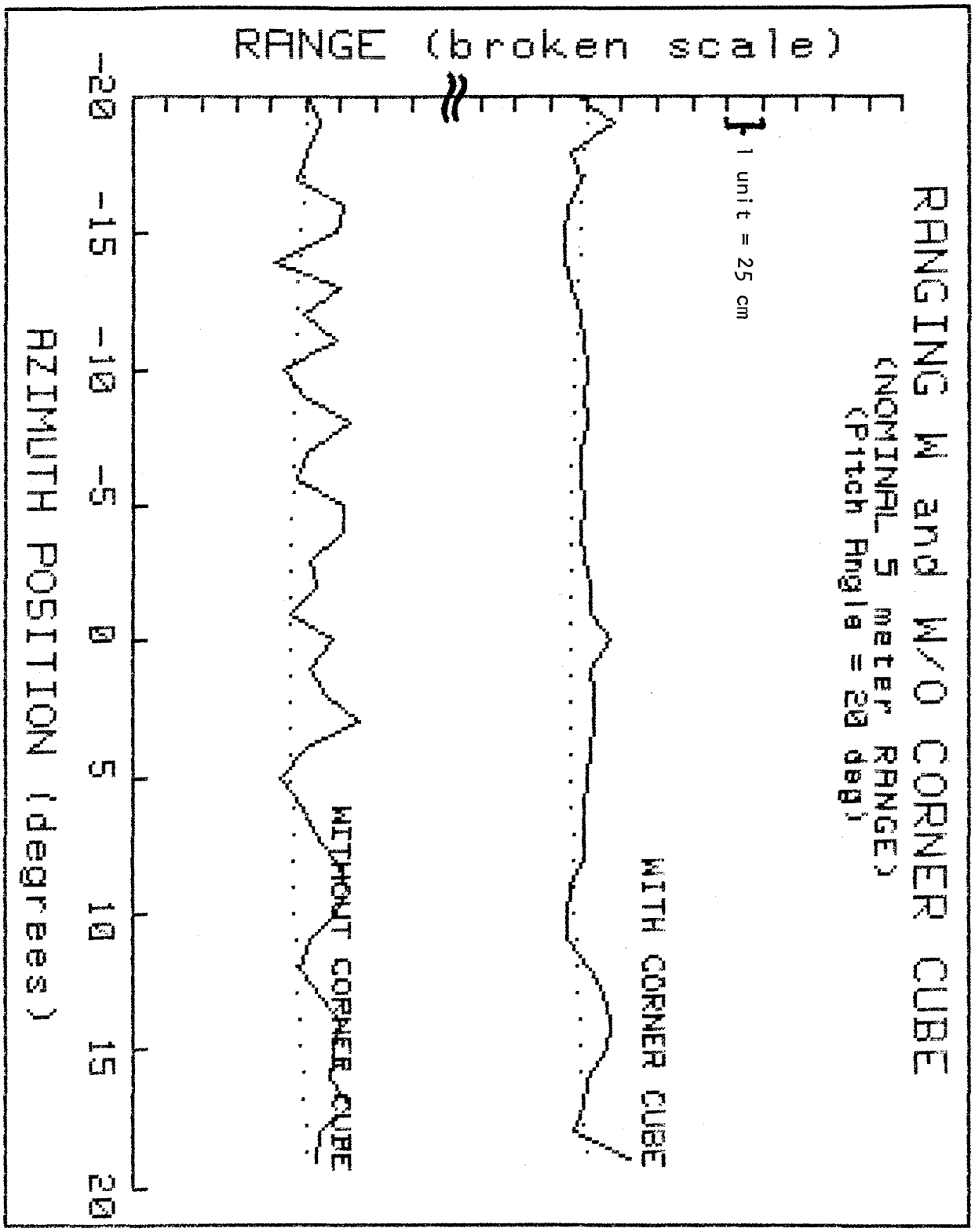


Figure 7. Five-Meter Ranging With and Without Cube Corner Reflector. (Target Pitch Angle = 20 Degrees)

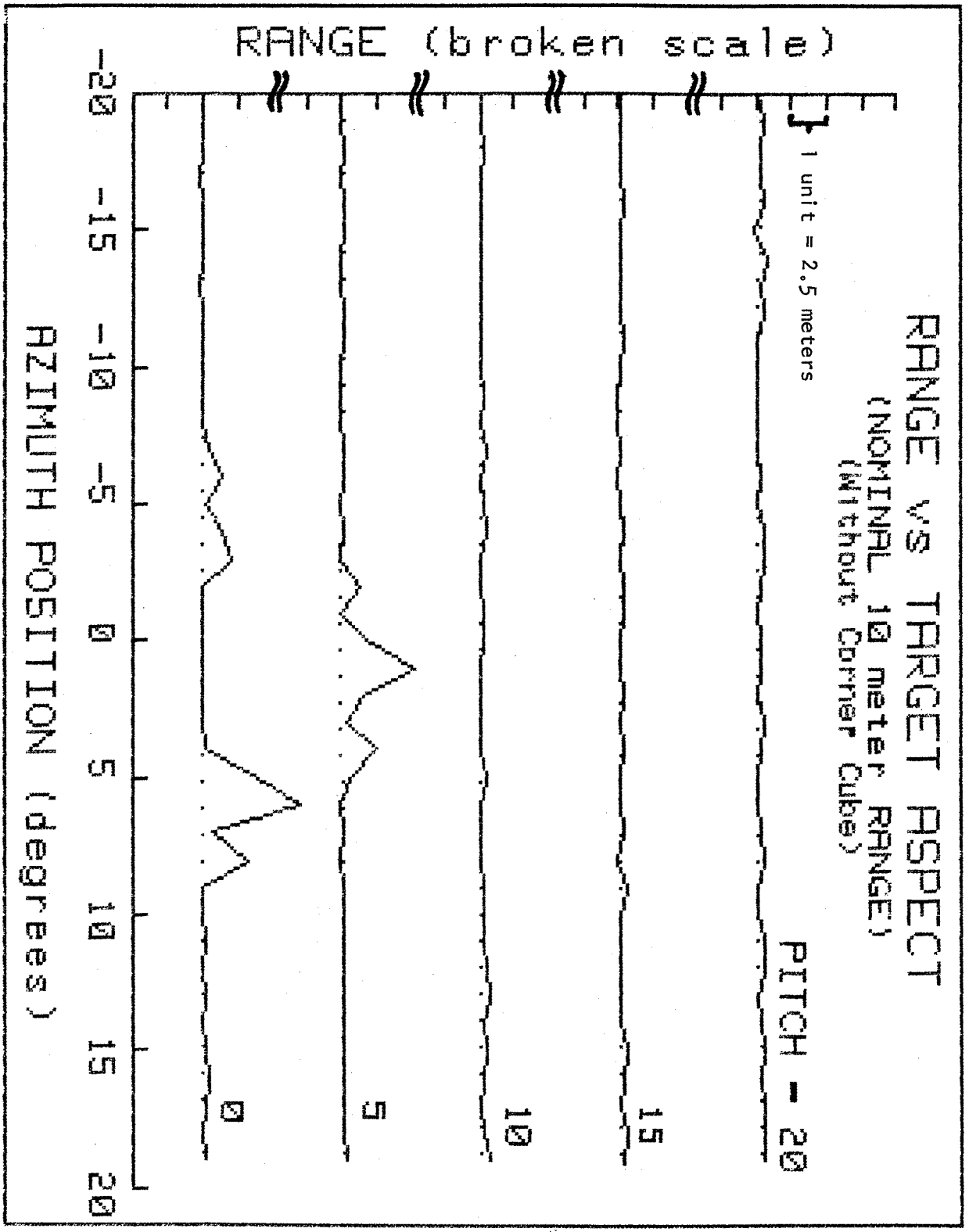


Figure 8. Range vs. Target Aspect (Nominal Range of 10 meters; WITHOUT Cube Corner Reflector).

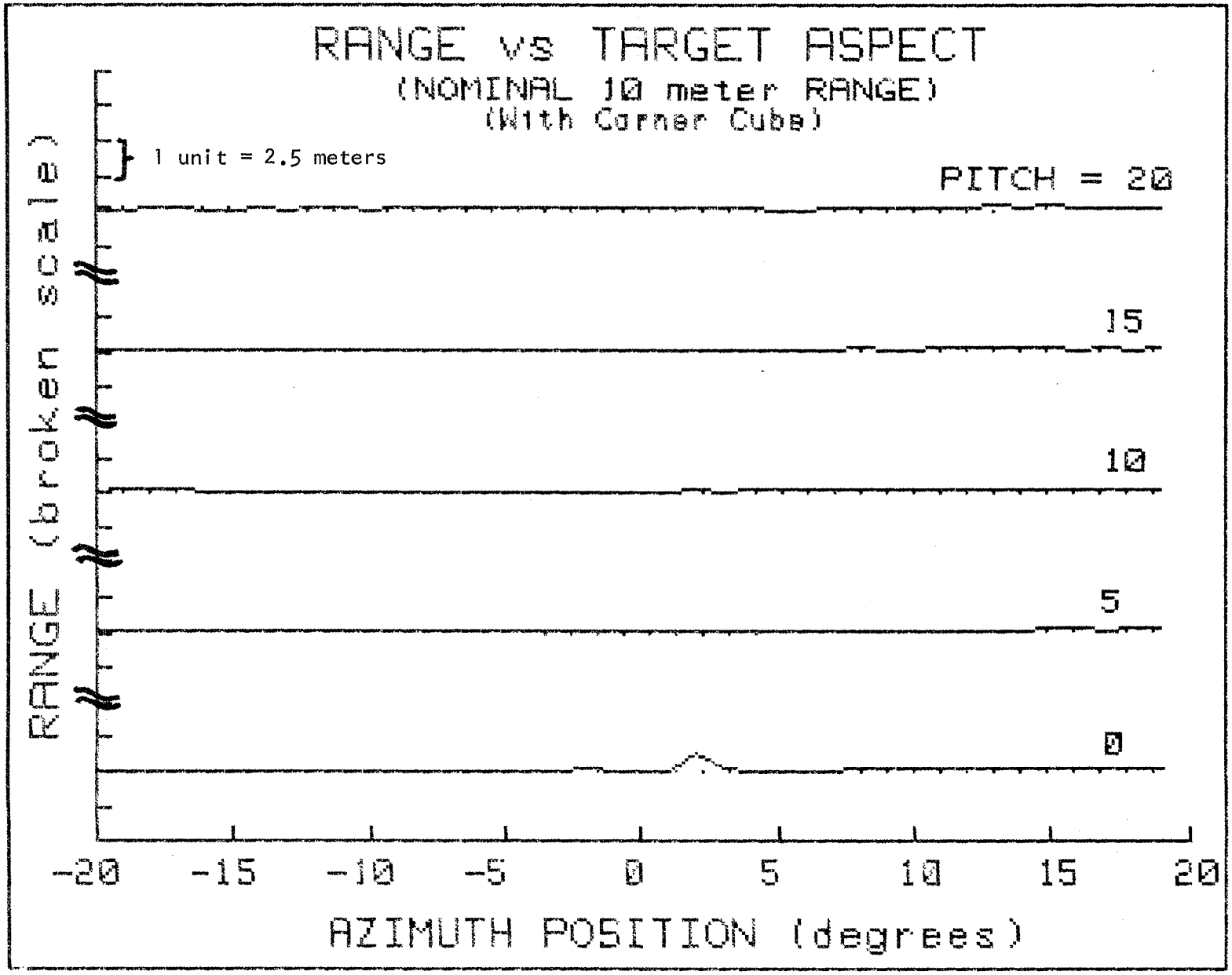


Figure 9. Range vs. Target Aspect (Nominal Range of 10 meters; WITH Cube Corner Reflector).

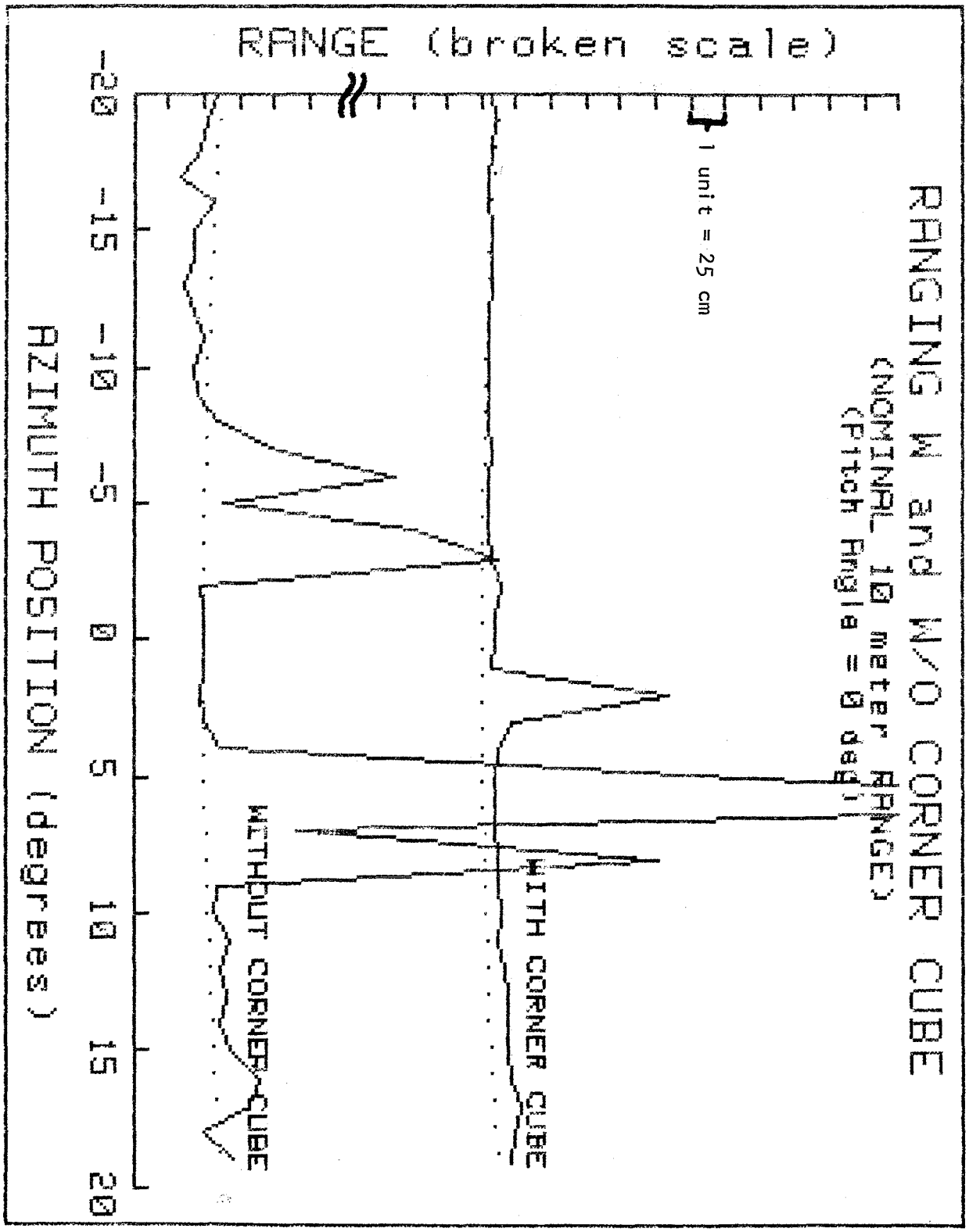


Figure 10. Ten-Meter Ranging With and Without Cube Corner Reflector. (Target Pitch Angle = 0 Degrees)

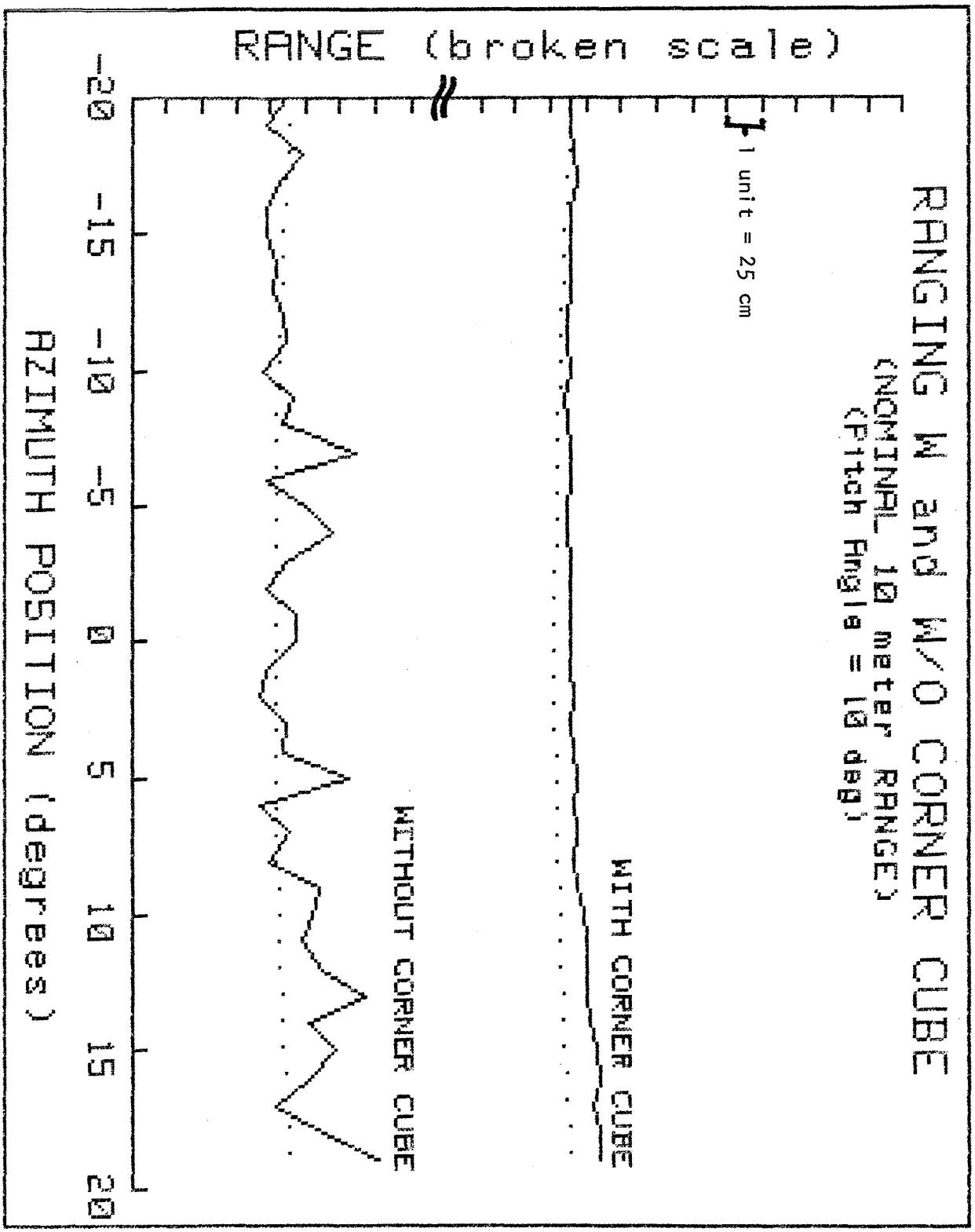


Figure 11. Ten-Meter Ranging With and Without Cube Corner Reflector. (Target Pitch Angle = 10 Degrees)

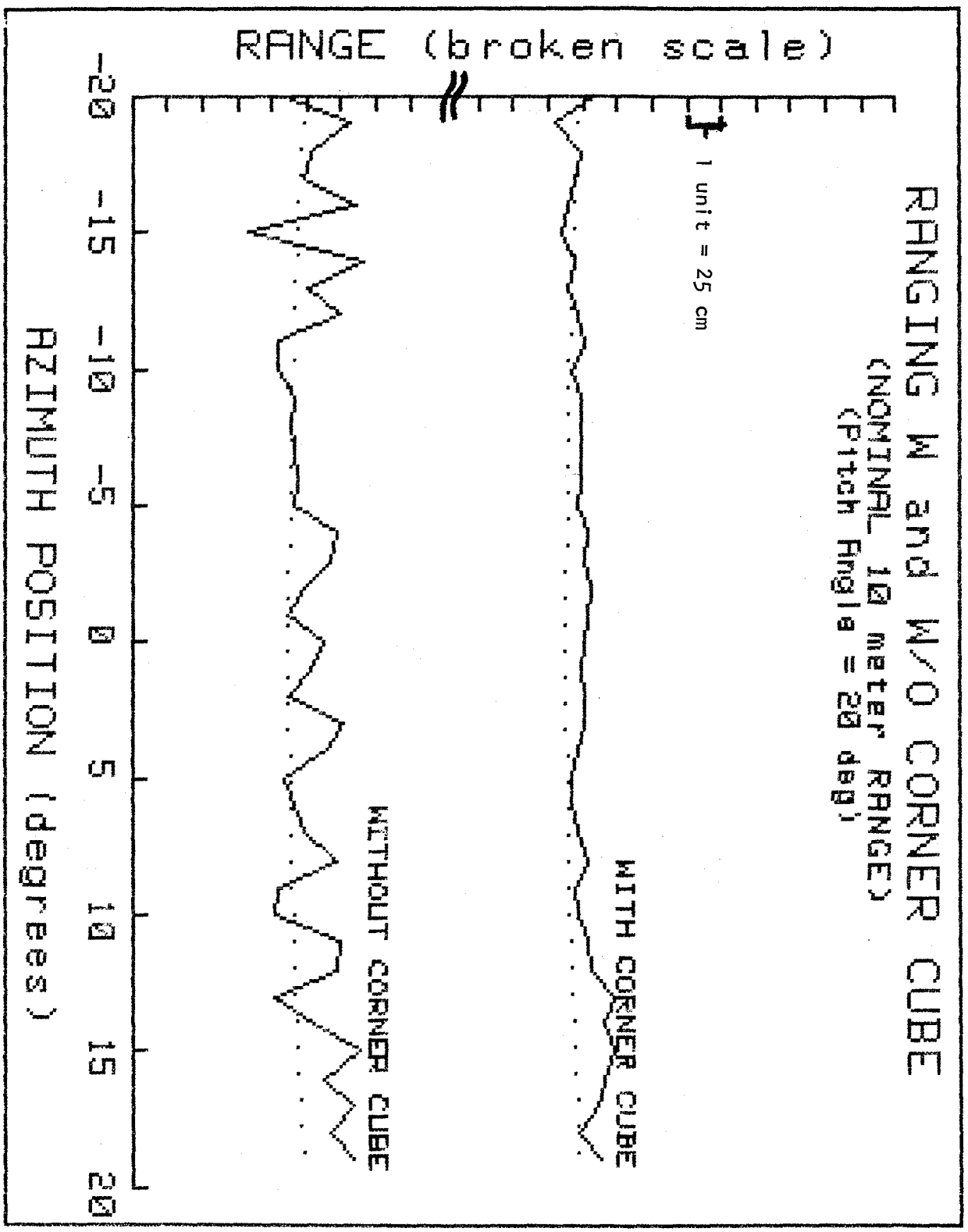


Figure 12. Ten-Meter Ranging With and Without Cube Corner Reflector. (Target Pitch Angle = 20 Degrees)

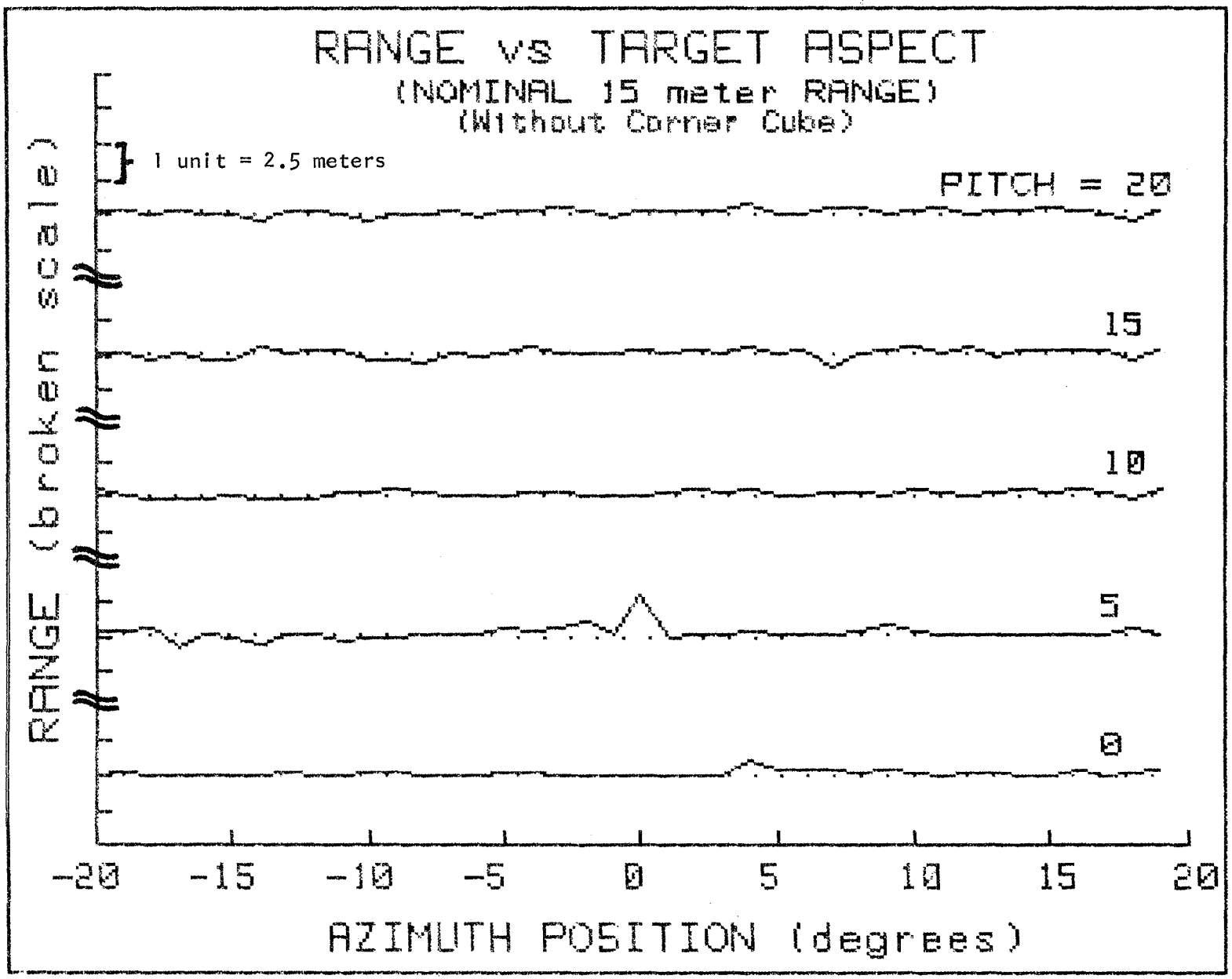
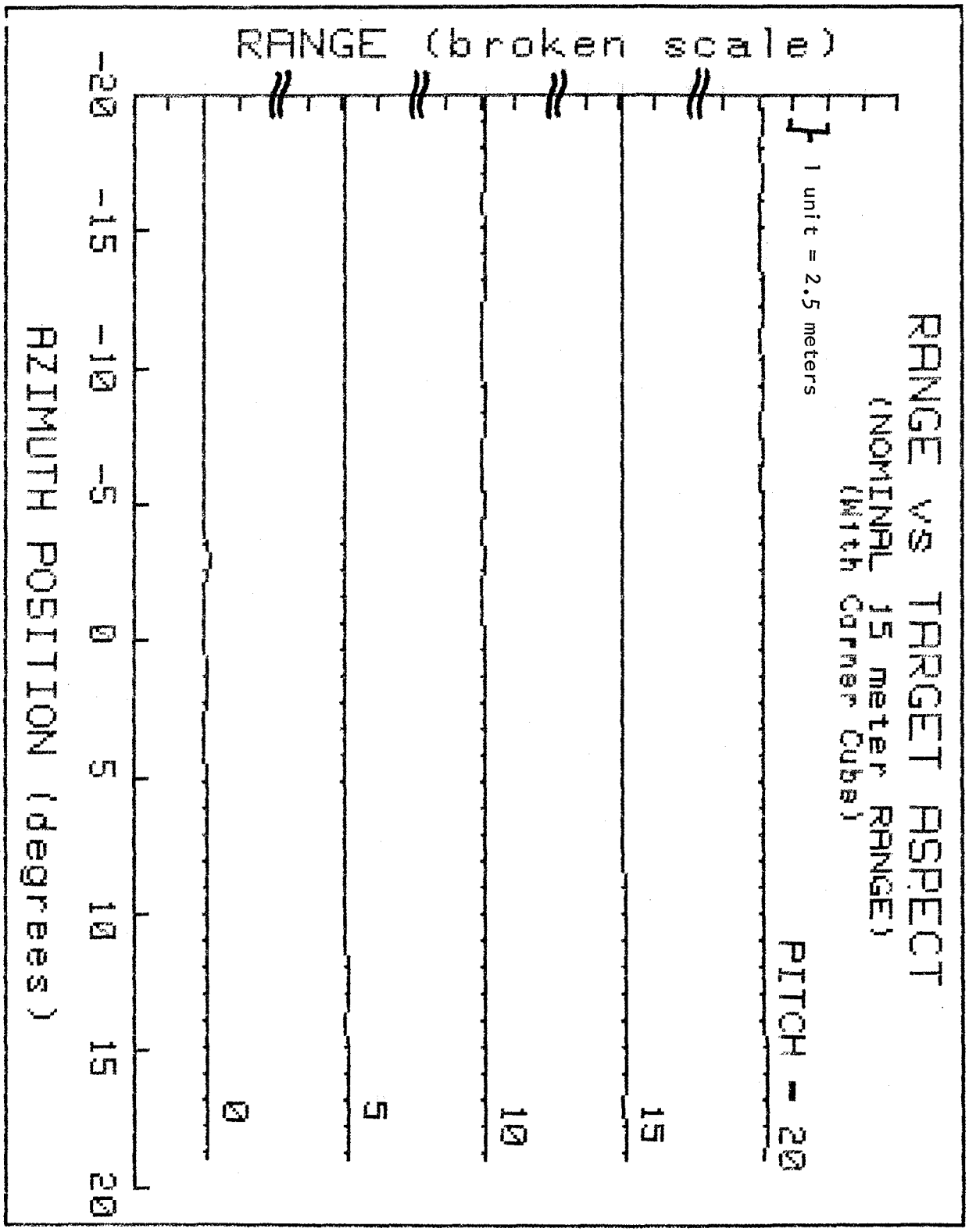


Figure 13. Range vs. Target Aspect (Nominal Range of 15 meters; WITHOUT Cube Corner Reflector).



Figute 14. Range vs. Target Aspect (Nominal Range of 15 meters; WITH Cube Corner Reflector).

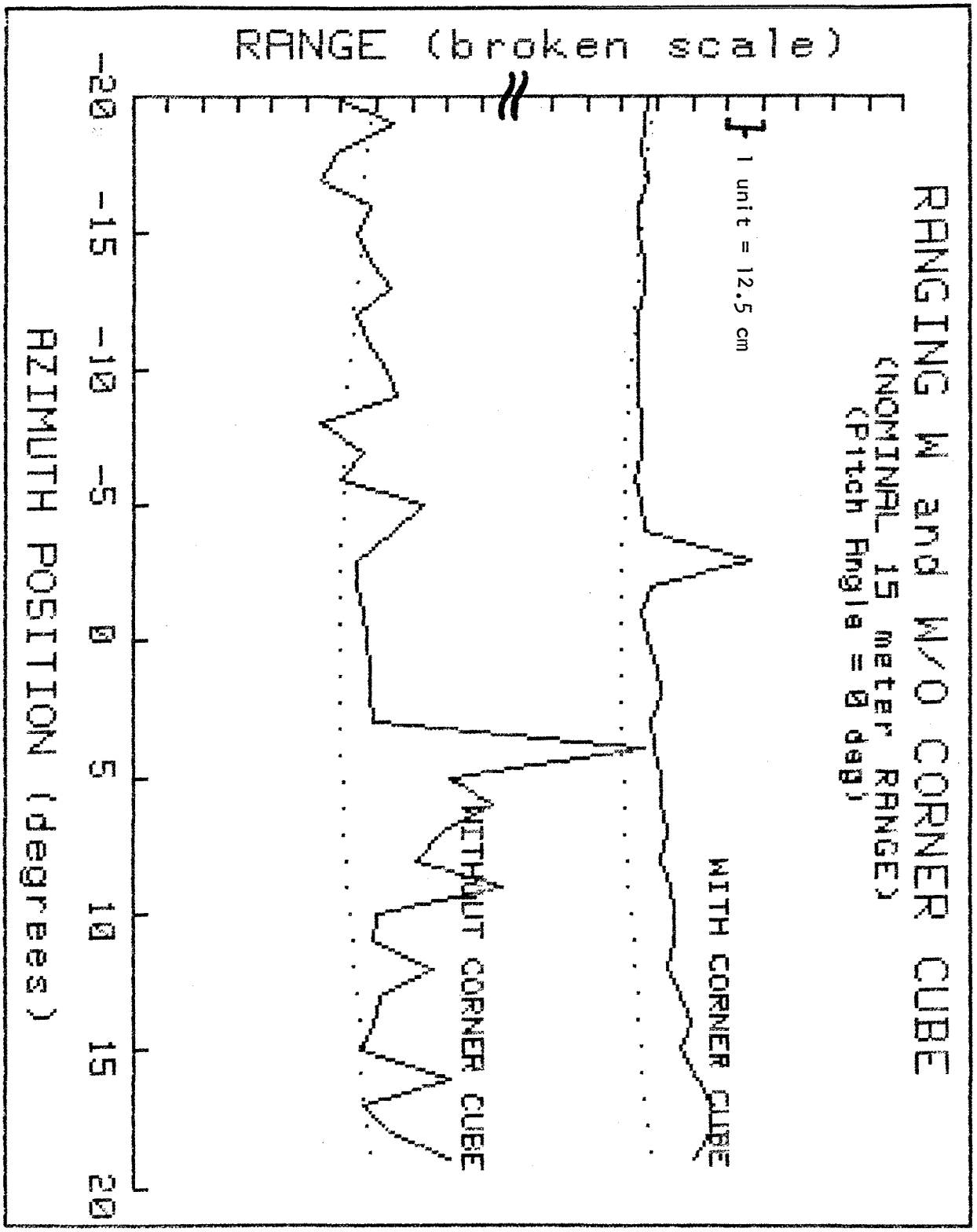


Figure 15. Fifteen-Meter Ranging With and Without Cube Corner Reflector. (Target Pitch Angle = 0 Degrees).

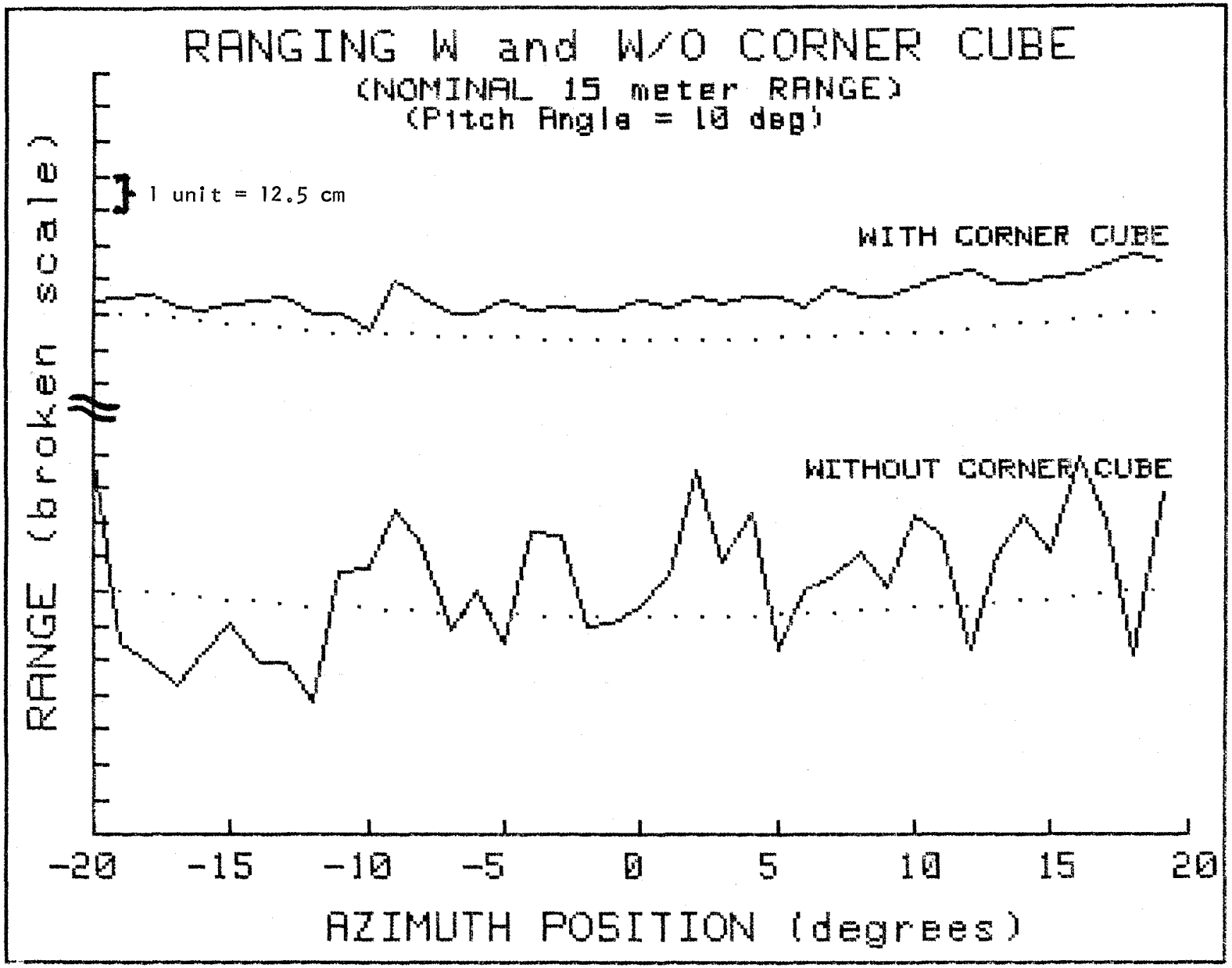


Figure 16. Fifteen-Meter Ranging With and Without Cube Corner Reflector. (Target Pitch Angle = 10 Degrees)

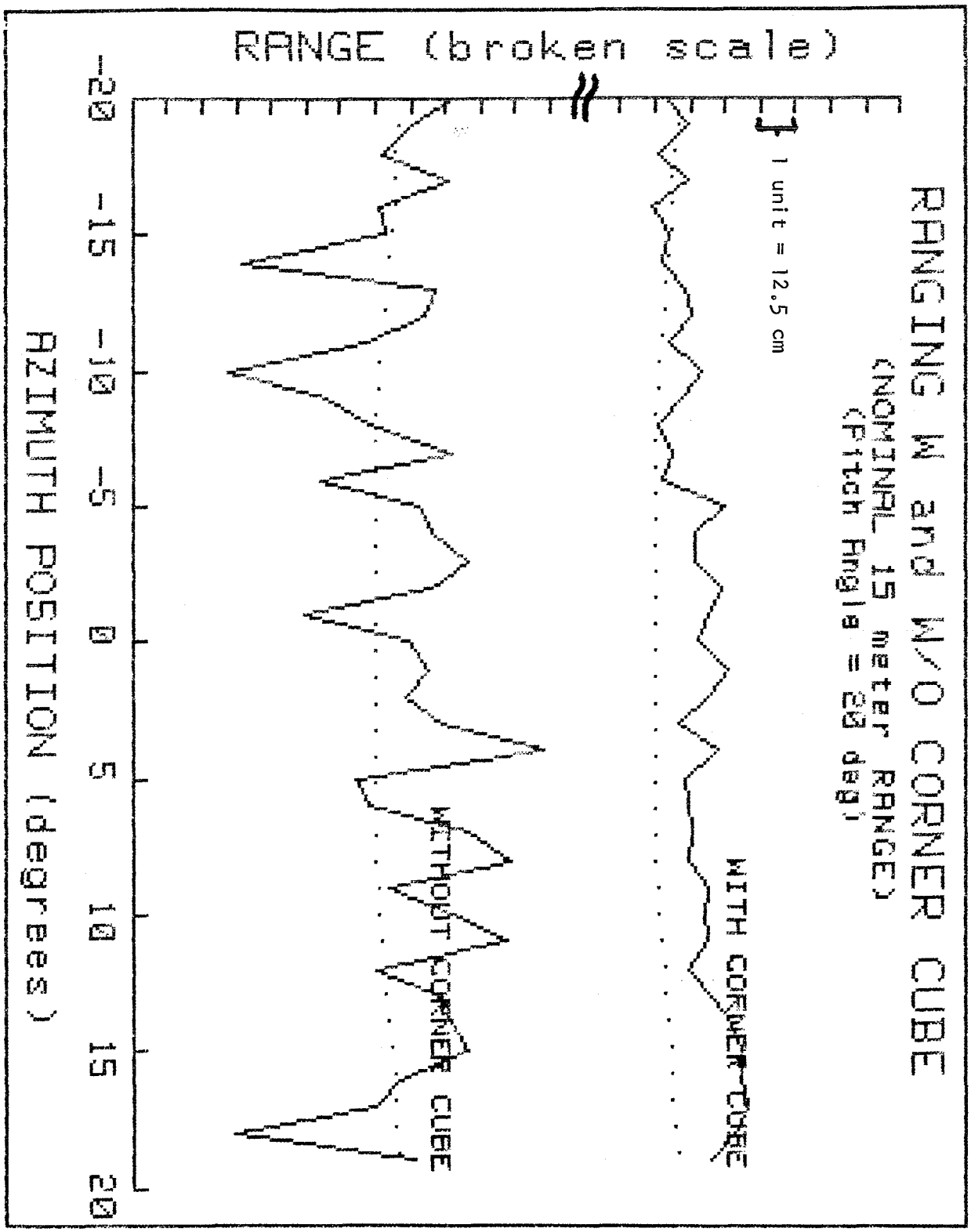


Figure 17. Fifteen-Meter Ranging With and Without Cube Corner Reflector. (Target Pitch Angle = 20 Degrees).

in this series (Figs. 15-17) the range scale is expanded by a factor of twenty over the gross representations of Figs. 13 and 14. One unit on the expanded scales is equivalent to 12.5 cm.

The results of the velocity measurements are shown in Figs. 18 and 19. Figure 18 documents graphically the difficulty of measuring the closing velocity under conditions of a target rotating in position to present varying aspects to the closing vehicle. During these runs a human operator was instructed to produce randomly varying aspects by sweeping both pitch and azimuth in a range of angles within 5 degrees either side of the normal aspect. The same random variation of aspect was not repeated in the two runs depicted in Fig. 18. Note the extreme range of velocity values on the velocity scale.

To present a stark contrast, the measured velocity values in Fig. 19 indicate an accuracy with which velocity can be measured by the radar system when the target maintains a fairly stable attitude. Note that the velocity oscillations shown in each of the two separate runs over the same stretch of gantry track show peak-to-peak fluctuations of about two cm/sec, following almost identical periodicity. The periodicity is believed due to the minor lurching motion observed at the top of the gantry as the gantry is driven along on the rails below. The 20 second and 50 second time points in both Figs. 18 and 19 correspond to nominal ranges of 12 meters and 8 meters, respectively, in the middle of longer runs from 15 meters to 5 meters nominal range.

The velocity vs. time results in Fig. 19 seem to depict with a high degree of fidelity the actual relative motion between the radar and the ST model as a second-by-second function of time. This suggests a capability for resolving closing velocities with accuracies on the order of centimeters/second, remembering that the maximum allowable closing velocity for the engagement of the OMV with the ST is slightly over 3.0 cm/sec.

VI. CONCLUSION

In conclusion, this experiment has demonstrated that a cooperative target aid with high radar cross section helps immensely in measuring range and velocity. A choice of reflector superior in its properties to the cube corner is a Luneberg lens reflector, with its much wider acceptance angle. To illustrate, the 3dB-to-3dB response of the cube corner subtends only about 45 degrees whereas the similar response of a properly constructed Luneberg lens can exceed 140 degrees of acceptance angle.

Finally, in reflecting on the fact that the OMV/ST docking simulation studies with human pilots operating the docking controls, cited earlier, showed best pilot results occurred with the aid of closing velocity information provided in addition to the video information, this study has shown that the radar system provides accurate enough velocity information to support "soft" docking if the ST attitude remains relatively stable. It need not be pointed out that attitude stability is a prime design feature of the Space Telescope.

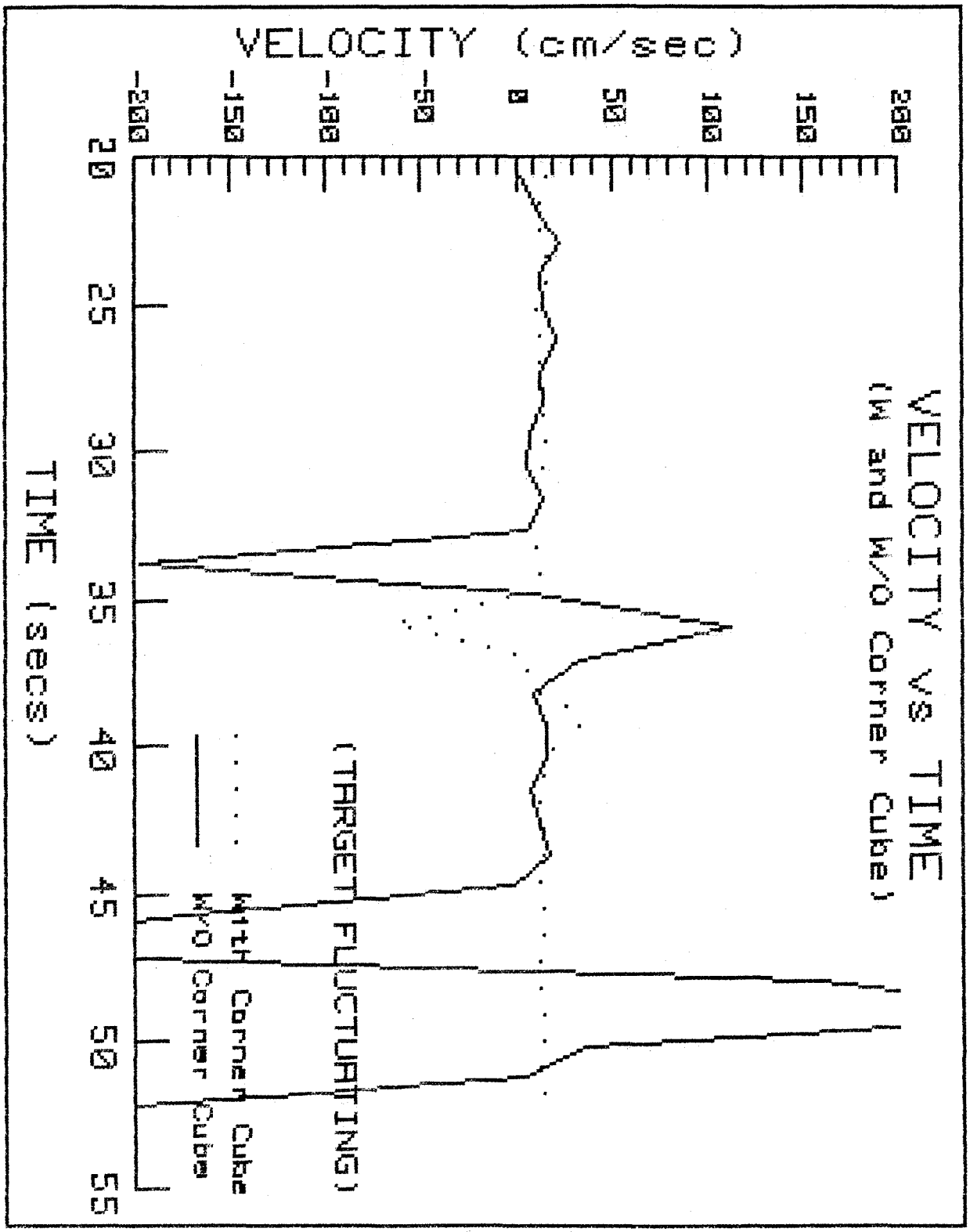


Figure 18. Velocity vs. Time With and Without Cube Corner Reflector. (Target Aspect Fluctuating Randomly).

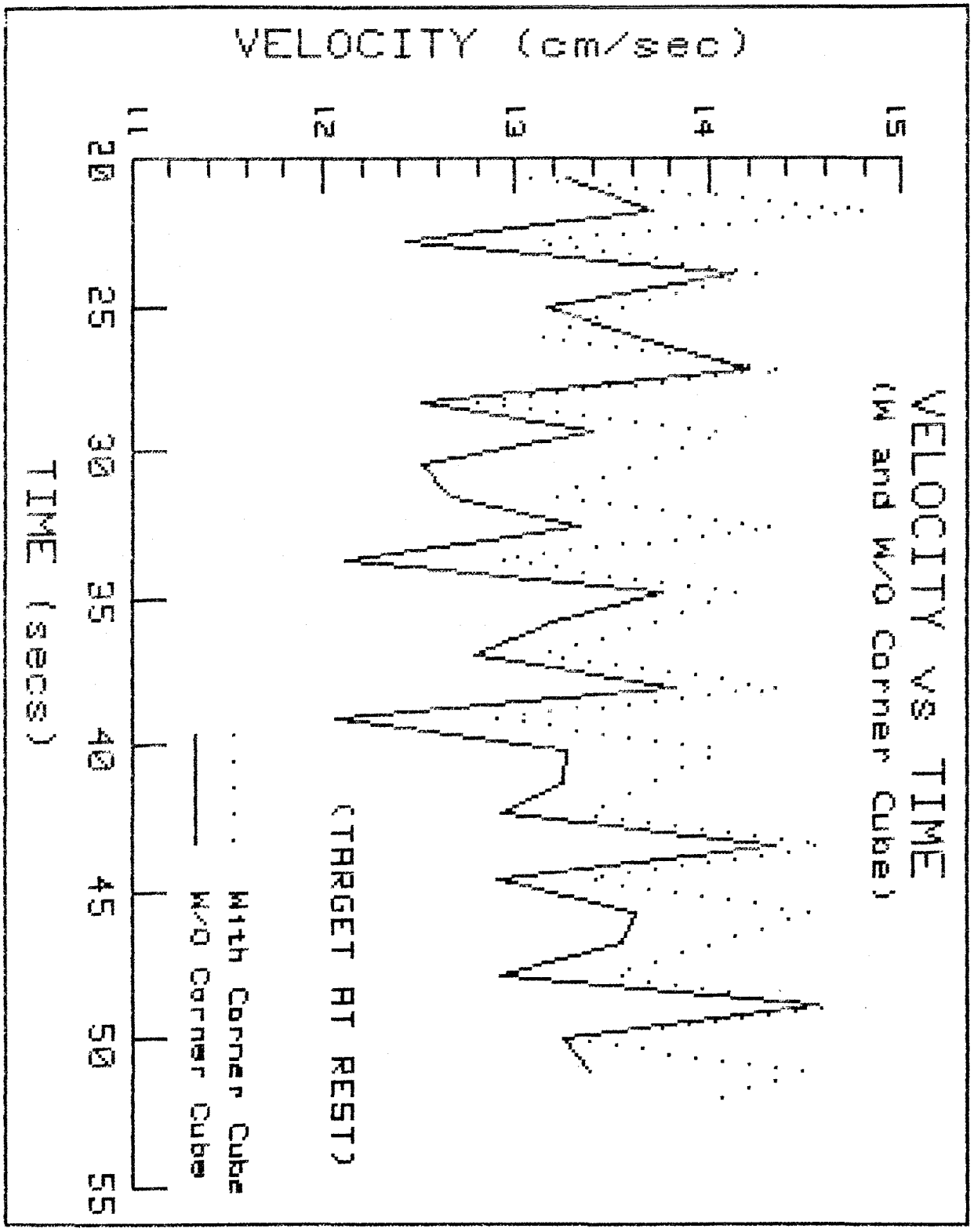


Figure 19. Velocity vs. Time With and Without Cube Corner Reflector. (Target Aspect Stable).

REFERENCES

1. McDonald, M.W., "Study of Proposals for the Correction of Errors in a Radar Ranging Device Designed to Facilitate Docking of the Teleoperator Maneuvering System", NASA Contractor Report - 1982 NASA/ASEE Summer Faculty Fellowship Program, edited by B.F. Barfield, et al, NASA CR-162051, August, 1982.

2. McDonald, M.W., "A FMCW Radar Ranging Device for the Teleoperator Maneuvering System", NASA Contractor Report - 1983 NASA/ASEE Summer Faculty Fellowship Program, edited by G. Karr, et al, NASA CR-162051, August, 1983.

3. Bailey, J.C., "OMV/PAYLOADS INTERFACE REQUIREMENTS DOCUMENT I - SPACE TELESCOPE", Marshall Space Flight Center, NASA Internal Distribution Document: OMV IRD, Date 12/4/84, December, 1984.

4. Micheal, J.D., and F.L. Vinz, "Docking Simulation Analysis of Range Data Requirements for the Orbital Maneuvering Vehicle", NASA Technical Memorandum, NASA TM-86510, April, 1985.

1985

NASA/ASEE SUMMER FACULTY RESEARCH FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

DIELECTRIC CURE MONITORING OF THE 55A FILAMENT WOUND CASE RESIN

Prepared by:	Ronald C. McNutt, Ph.D.
Academic Rank:	Professor
University and Department	Athens State College Chemistry Department
NASA/MSFC Division: Branch:	 Nonmetallic Materials Ceramics and Coatings
MSFC Counterpart	Benjamin E. Goldberg Marie L. Semmel
Date:	August 2, 1985
Contract No.:	NGT 01-008-021 The University of Alabama in Huntsville

DIELECTRIC CURE MONITORING OF THE 55A FILAMENT WOUND CASE RESIN

By

Ronald C. McNutt
Professor of Chemistry
Athens State College
Athens, Alabama

ABSTRACT

Studies have been conducted to correlate data from dielectric cure monitoring of 55A filament would case resin using the Micromet System II with differential scanning calorimetry and resin strength tests. Strength tests for resin bonding at various time intervals and DSC are standard accepted techniques for monitoring resin cure cycles. They are time consuming, destructive, and non-continuous. The Micromet System II will yield reliable information on degree of cure and other parameters, which should allow better quality control on filament wound case resin cure.

Acknowledgements

The author would like to thank Mr. Ben Goldberg and Mrs. Louise Semmel for their guidance and support. It is appropriate to express appreciation to all the people in the Non-Metallic Materials Division of the Materials and Processing Laboratory for their unselfish lending of time and effort.

LIST OF FIGURES

LIST OF TABLES

Introduction

Preliminary studies on dielectric cure monitoring of the Hercules 55A resin have been carried out by Goldberg and Semmel using the Micromet II system developed by Senturia, Day and others. This study continues the work of Goldberg and Semmel with the intention of correlating the dielectric properties of the various resin mixes with data collected from differential scanning calorimetry, bonding strength tests, and solubility considerations. The dielectric parameters of primary concern are loss factor (ϵ'') and permittivity (ϵ'). Loss factor appears to be the more informative property. The permittivity, $\epsilon'(\omega)$, is defined as the frequency dependent dielectric constant. The loss factor, $\epsilon''(\omega)$, is a measure of the amplitude of the component of the dielectric displacement, D(t), 90 degrees out of phase with E(t). E(t) is a cosine term while , the loss factor, is a sine term.

$$D(t) = \epsilon'(\omega)\, E^{\circ} \cos(\omega t) + \epsilon''(\omega)\, E^{\circ} \sin(\omega t)$$

$$E(t) = E^{\circ} \cos(\omega t)$$

(Where E° is the amplitude of the imposed field and $\omega = 2\pi$ frequency)

A much more complete description of the Micromet system and explanation of time dependent dielectric measurements may be found in the study by Goldberg and Semmel.

Most of the literature pertinent to microdielectrometry used in the monitoring of epoxy resin cure cycles is found in govermnent contract reports (Ref 2 thru 8). This work resulted in the availability of the Micromet system. The principal advantage of this system is that the tranducer is very small and independent of geometric considerations. Dielectric measurements have been made on resins and composite materials for quite some time. There were usually in conjunction with other techniques and with rather cumbersome equipment compared to microdielectrometry. (Ref 9 - 17)

Procedures

The general procedure for these measurements is to prepare a sample by weighing an amount of resin and catalyst prescribed by specifications. They are mixed thoroughly and divided into aliquots (depending on how many concurrent analyses are to be done). In this work three aliquots are usually sufficient. The aliquot for dielectric analysis is immediately set up and the computer controlled measurement begun. This determination runs continuously for 50 hours with one parameter displayed on the computer screen and data for two parameters plus resin temperature is printed each 15 minutes on a paper roll. All parameters and varibles are stored on data tape. An initial DSC (differential scanning calorimetry) determination is made as quickly as possible after starting the micromet run. This involves weighing a small sample, 5 to 20 milligrams, for the DuPont 1090 scanning calorimeter available in the Materials and Processing Laboratory. The initial determination for the heat of reaction is used as the base value for degree of cure. Additional determinations are then obtained as often as possible in keeping with instrument time available and other analyses required. Usually one can do about four to six DSC measurements in an eight hour day. It is desirable to take some data at night while the Micromet run is underway. Once the initial DSC measurement is made the resin mix is prepared for strength tests. This is done by lightly spreading the resin mix on one square inch surface area aluminum blocks. Six samples are prepared with as many replicates as one has blocks. We could do a maximum of two replicates (at least five replicates are needed for statistically significant results). The initial sample is prepared with the original mix, the others require a fresh mix for each sample. The samples are allowed to cure for at least fifty hours and then pulled apart on the Instron instrument and the force recorded. Attempts to do solubility studies by dissolving the resin mix in methyl ethyl ketone and flocking with methyl alcohol were unsuccessful because a satisfactory filtering technique was not found.

Results and Discussions

This work was confined to the Hercules 55A resin system. Several fifty hour runs were made using the Micromet II microelectrometer system. The runs displayed in this report (Fig 3 - 14) show minimum nominal, and maximum specification resin to catalyst ratios in addition to an excess resin to catalyst run. The figures showing loss factor versus time are overlaid with percent cure and exotherm temperatures from DSC measurements (Fig 3 - 10). There are figures of permittivity versus time for each of the above resin to catalyst ratios. Figure 2 shows a typical DSC run. Table I is compilation of the strength test measurements.

The Micromet II runs of loss factor and permittivity versus time are explained, by Goldberg and Semmel, in terms of gelation and vitrification. Loss factor appears to be the more informative parameter. Several frequencies were monitored, i.e. 100, 500, 1000, 5000, and 10,000 Hz. We have shown only 100, 1000, and 10,000 Hz primarily in the interest of clarity. An explanation for the loss factor curve at 100 Hz (Figures 3 - 10) could be that dipole functions are polymerizing molecules which results in a lower loss factor. The loss factor at 100 Hz declines until about 25 or 30 hours when gelation begins. The loss factor for each frequency then goes through a maximum in this region with the higher frequencies going through the maximum earlier. It is possible that gelation causes the moites being formed to expend energy in the stiffening medium until solidification stop them from following the changing polarity at all. This earlier maximum for higher frequencies is observed for each resin to catalyst ratio. A general shift of the maxima will occur at different catalyst ratios and temperature. They will occur earlier for higher catalyst concentration and at higher temperatures. The behavior of these maxima at each frequency and the general behavior of the loss factor is a primary reason for the contention that this technique is one of the better ones available for continuous cure monitoring of the 55A resin system. Our work to correlate the system, primarily to DSC measurements, has also tended to support this contention. The dielectric probe is very small, independent of geometry, and nonintrusive, therefore, it is quite possible that in addition to continuous monitoring one could undertake nondestructive spot monitoring of a cure cycle if the proper combination of frequency values for each degree of cure could be determined.

The DSC data for percent cure and exotherm temperature have been overlain on the loss factor plots. Figure 2 shows a typical DSC run is done for each point on the loss factor and permittivity curves. All resin mix cures were done at room temperature for

fifty hours to duplicate the actual situation before oven cure on the filament wound case itself. A small inset on each figure shows a temperature plot for that particular run. Therefore, one can make a mental adjustment for temperature in considering the other variables. The plots for percent cure (Fig. 3, 4, 5, 6) show primarily S-shaped curves indicating that the cure rate tends to increase somewhat until the beginning of gelation and tends to slow after vitrification. This is in agreement with an interpretation of the temperature shift of the exotherm temperature shown in figures 7, 8, 9, and 10. This shift from higher temperatures indicates a lowering of the activation energy until the solidification process. This could be explained as autocatalysis followed by a change in mechanism at vitrification after which the exotherm temperature begins to increase. This type behavior was noted for other similiar epoxy systems being synthesized and studied in this laboratory.

Strength test measurements (Table I), as suggested in the procedural section, required more replicates to be considered good statistical data. However, there are definite trends in bond type and strength. Typically the bonds are cohesive until the beginning of gelation becoming mostly cohesive until vitrification when they become primarily adhesive. The bond strength appears to become generally stronger until vitrification then becoming weaker with more adhesive type bonding, following a rather inverted u-shaped curve from initial mix to fifty hours. A more statistically correct study could be helpful when considering the point at which a mix should be discarded in the event of a breakdown during winding, etc.

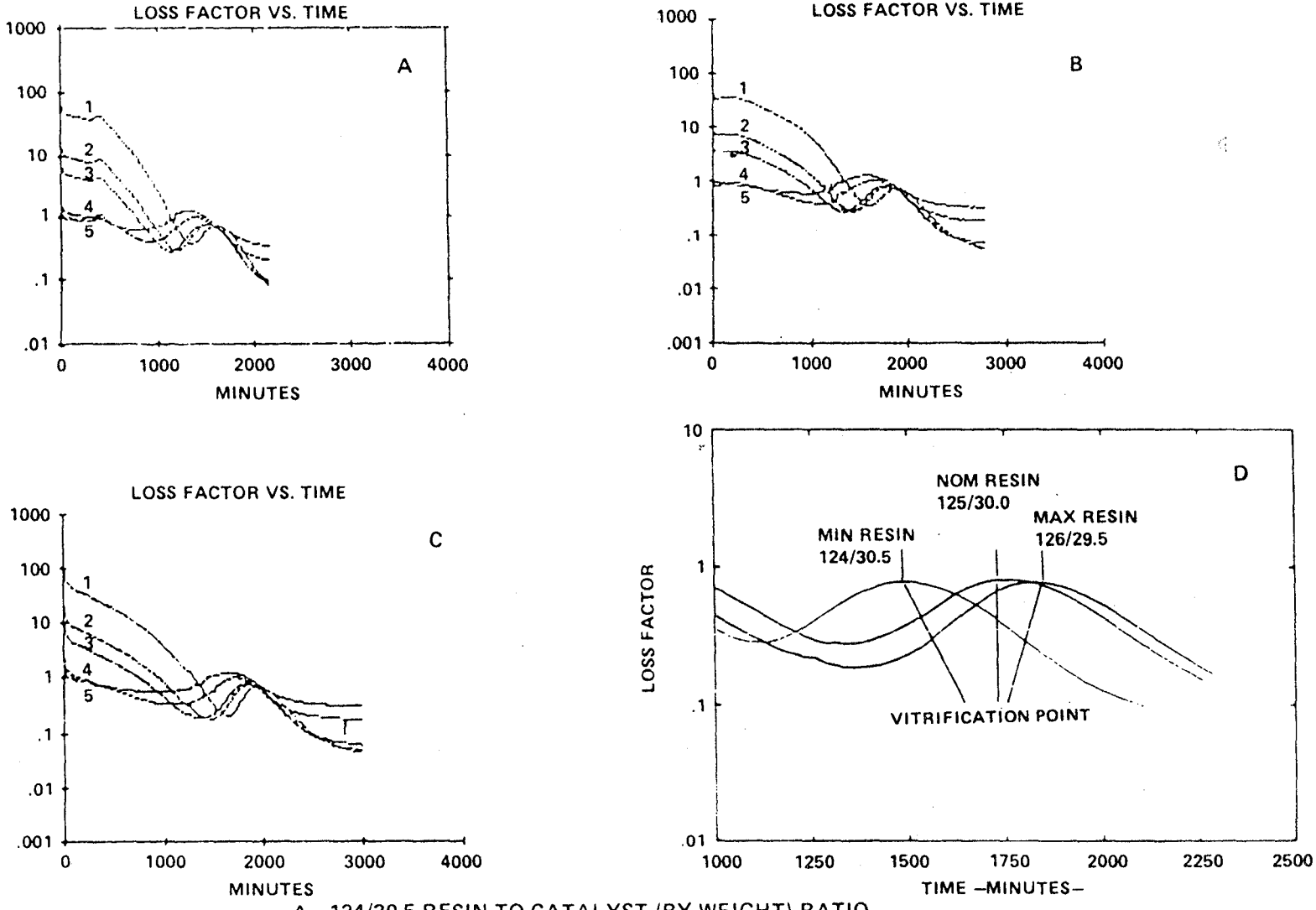
Recommendations

The Micromet II dielectric cure monitoring system should be given serious consideration for the continuous monitoring of the Hercules 55A Filament Would Case resin. It should be determined if this system could be used as a nondestructive technique for spot monitoring of the resin cure. Arrangements should be made in any further work to take concurrent data at regular intervals throughout the fifty hour cure cycle not just the regular eight hour day period.

TABLE I

Strength Test Data from 55A Resin Cure Cycles

Sample	Time (Minutes)	Bond Strength lbs/in^2	Bond Type	Remarks
Nominal Resin 55A-8	Initial	2850	cohesive	
"	379	3450	cohesive	
"	1354	4800	mostly cohesive	
"	1819	4250	cohesive-adhesive	
"	2793	6350	mostly adhesive	anomalous
"	2974	3850	adhesive	
Minimum Resin 55A-10	Initial	1800	cohesive	
"	1118	3600	cohesive	
"	1328	3550	cohesive	
"	1578	3200	mostly cohesive	
"	2623	1600	mostly adhesive	anomalous
"	2958	1900	adhesive	
Maximum Resin 55A-11	Initial	1740	cohesive	
"	1165	3145	cohesive	
"	1505	4480	cohesive	
"	1620	5730	mostly cohesive	
"	1975	2640	adhesive	
"	2935	2535	adhesive	
Excess Resin 55A-7	Initial	1340	cohesive	
"	265	4900	cohesive	
"	1302	4750	cohesive	
"	1688	2840	mostly adhesive	
"	2663	1950	adhesive	
"	2943	—	———	Broken in test



A. 124/30.5 RESIN TO CATALYST (BY WEIGHT) RATIO
B. 125/30
C. 126/29.5
D. GEL POINT COMPARISON OF A, B, AND C AT ω = 1 K Hz.
(1 = 100 Hz, 2 = 500, 3 = 1000 Hz, 4 = 5000 Hz, 5 = 10000 Hz.)

FIGURE 1: MICROMET II SCANS OF MINIMUM, NOMINAL, AND MAXIMUM RESIN TO CATALYST RATIOS

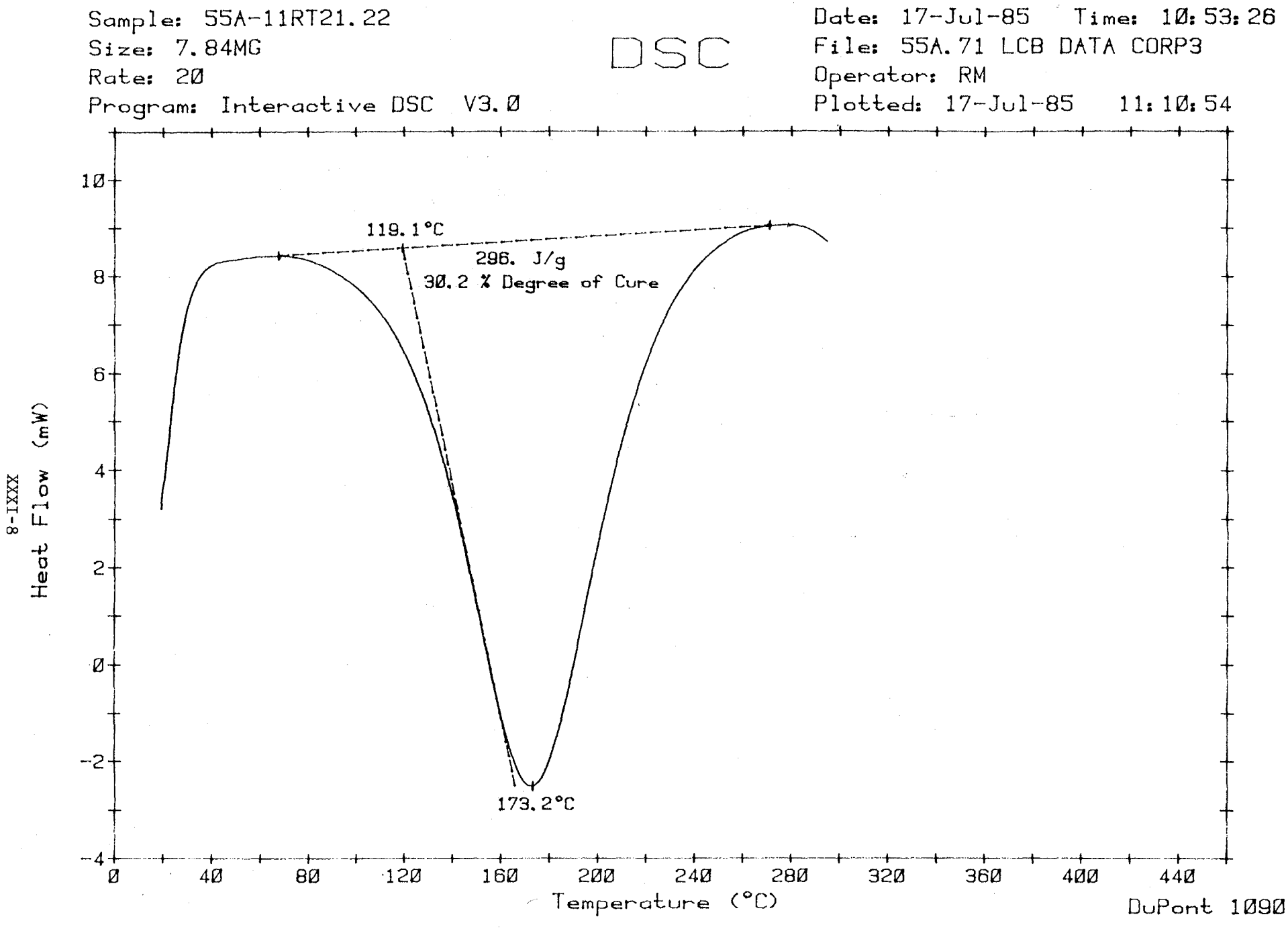
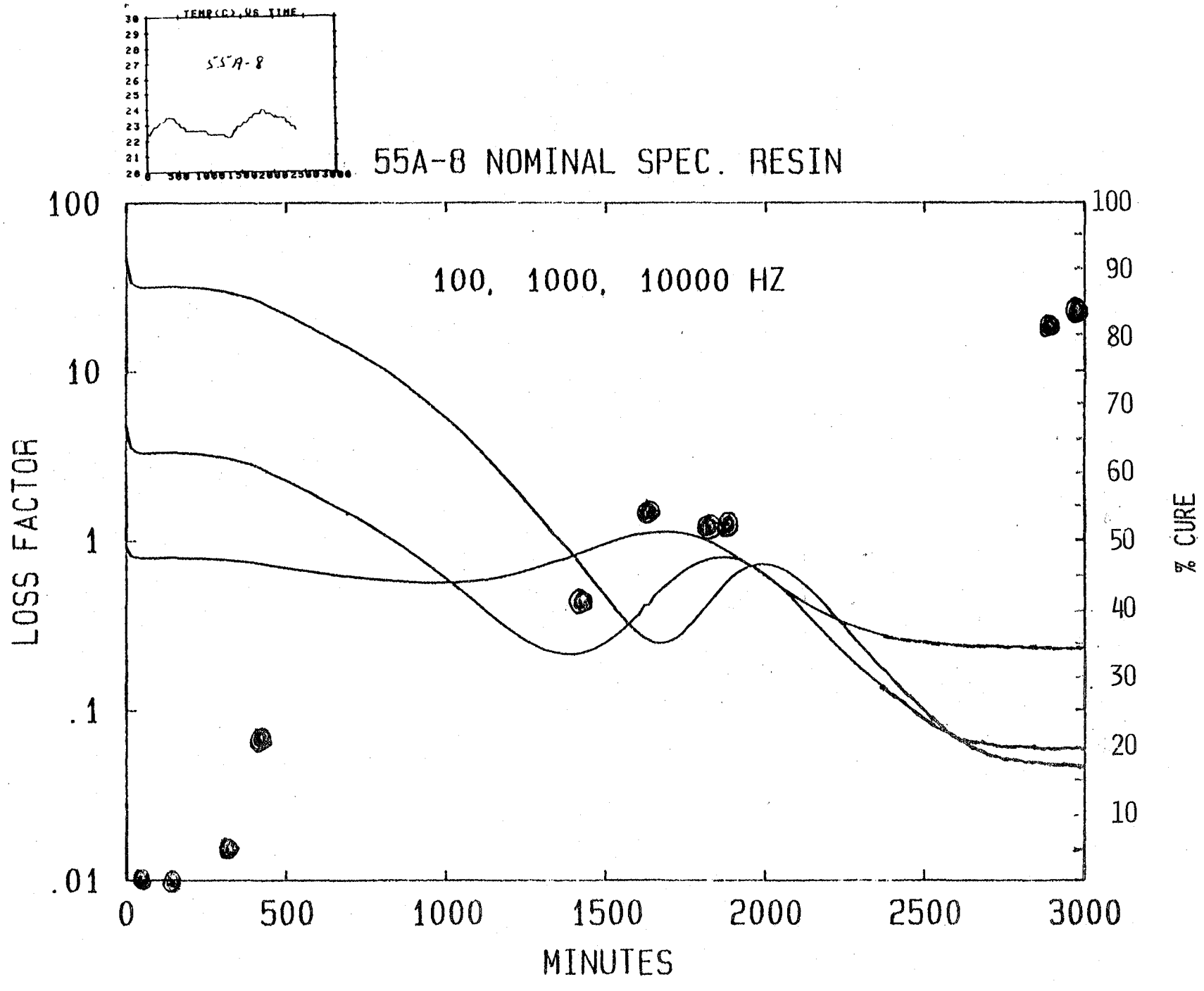
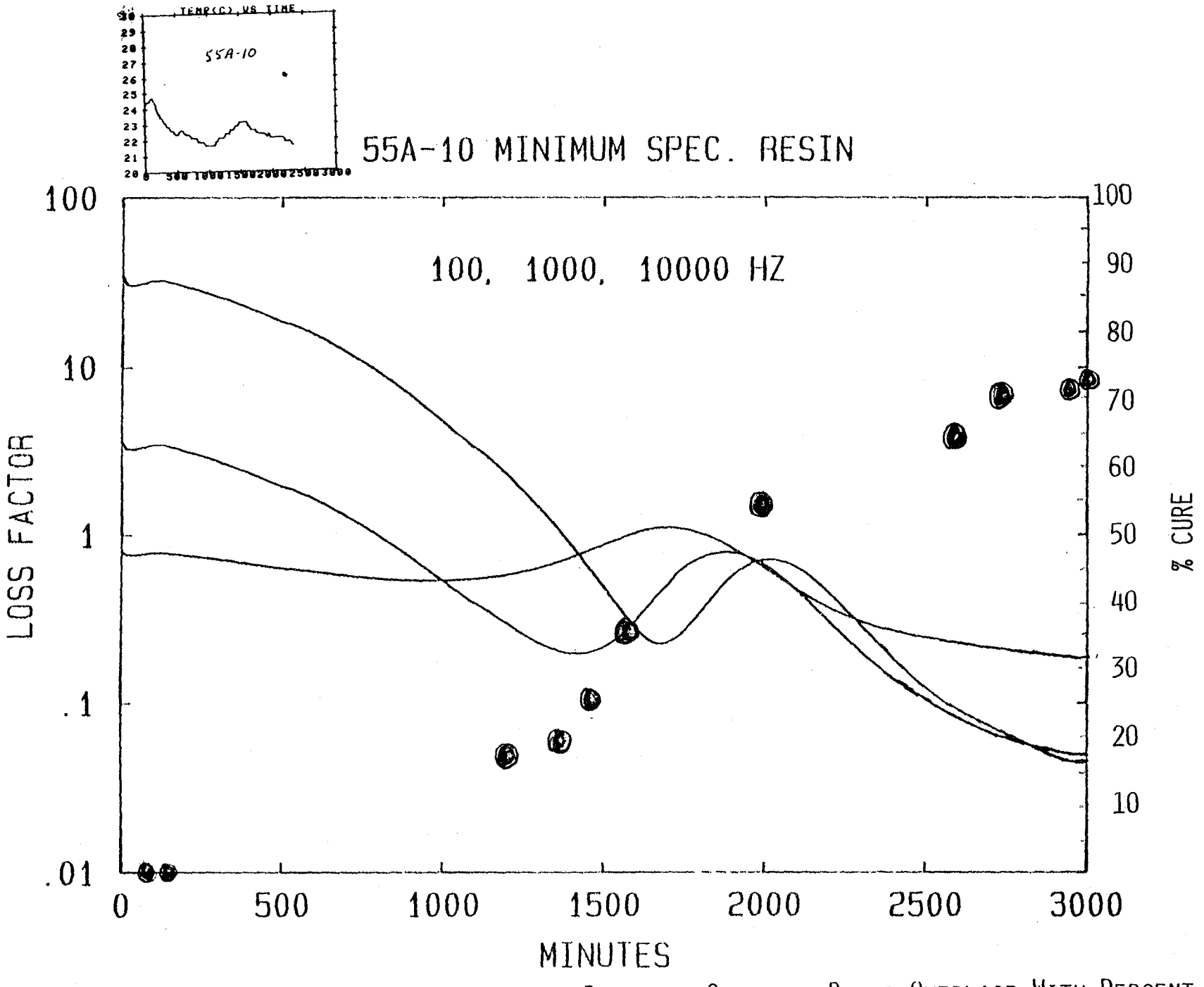


FIGURE 2: TYPICAL DIFFERENTIAL SCANNING CALORIMETER PLOT OF 55A FILAMENT WOUND CASE RESIN DURING CURE CYCLE



6-IXXX

FIGURE 3: Loss Factor vs Time Plot of Nominal Resin to Catalyst Ratio Overlaid With Percent Cure



XXXI-10

FIGURE 4: Loss Factor vs Time Plot of Minimum Resin to Catalyst Ratio Overlaid With Percent Cure

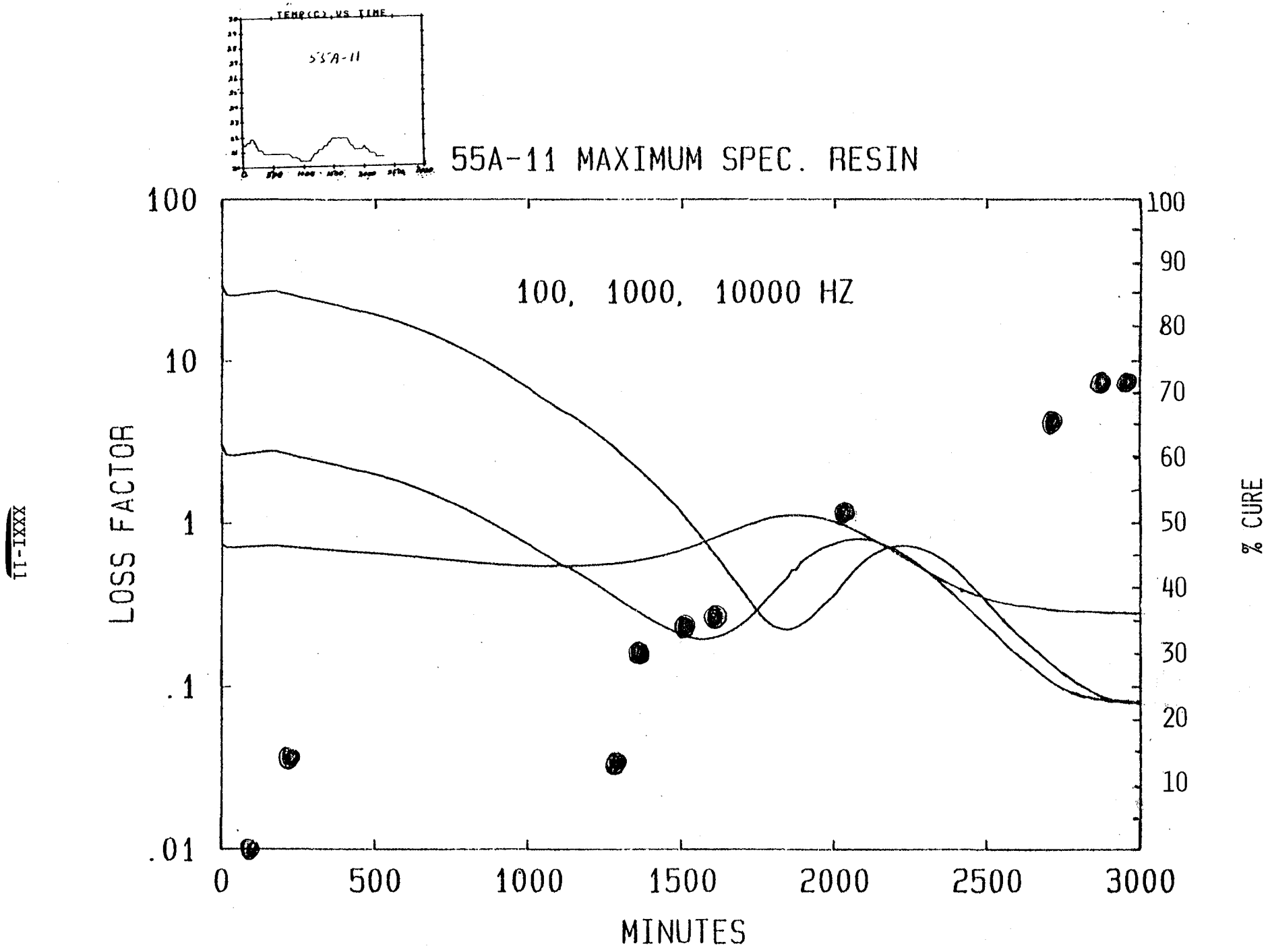


FIGURE 5: Loss Factor vs Time Plot of Maximum Resin to Catalyst Ratio Overlaid With Percent Cure

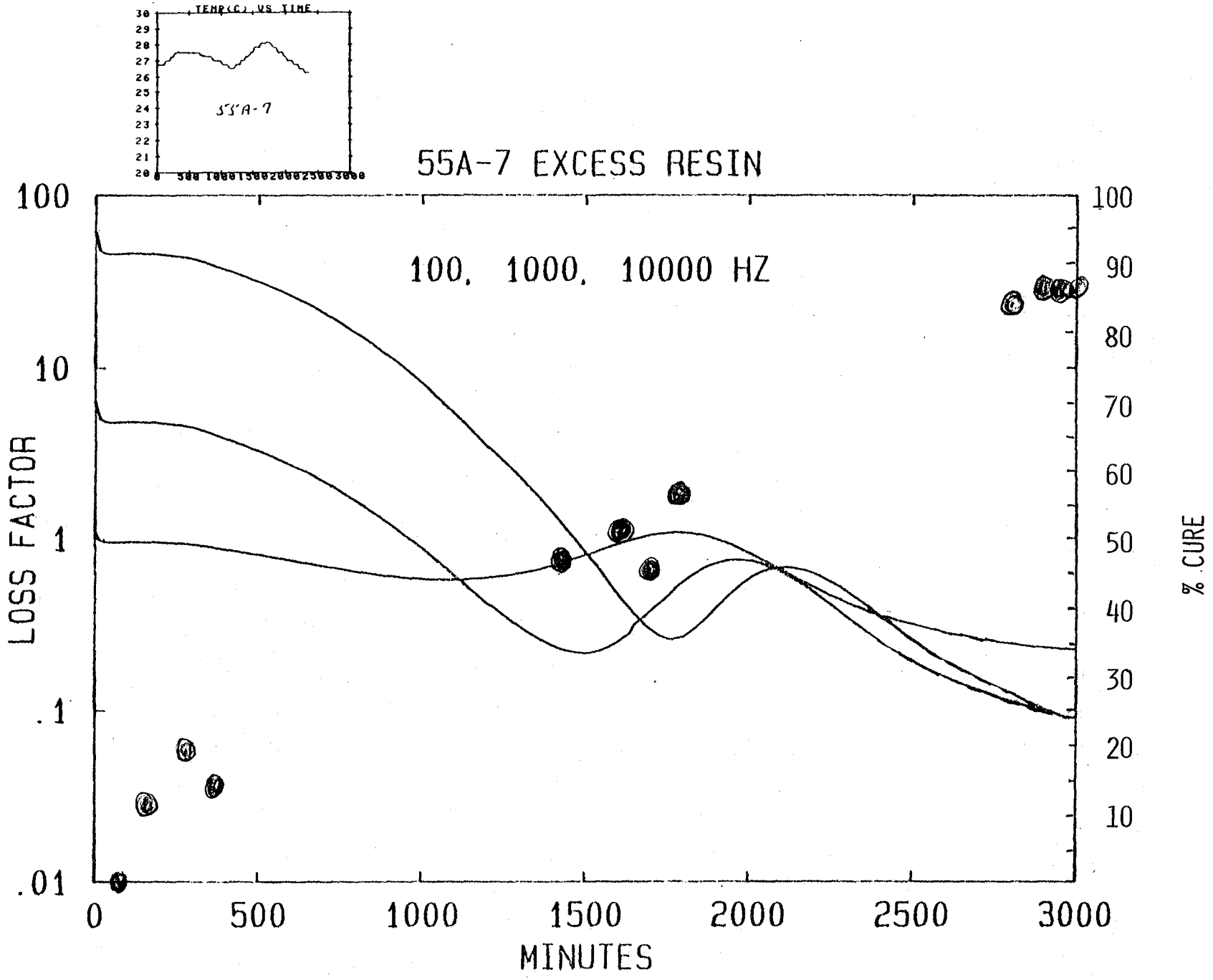


FIGURE 6: Loss Factor vs Time Plot of Excess Resin to Catalyst Ratio Overlaid With Percent Cure

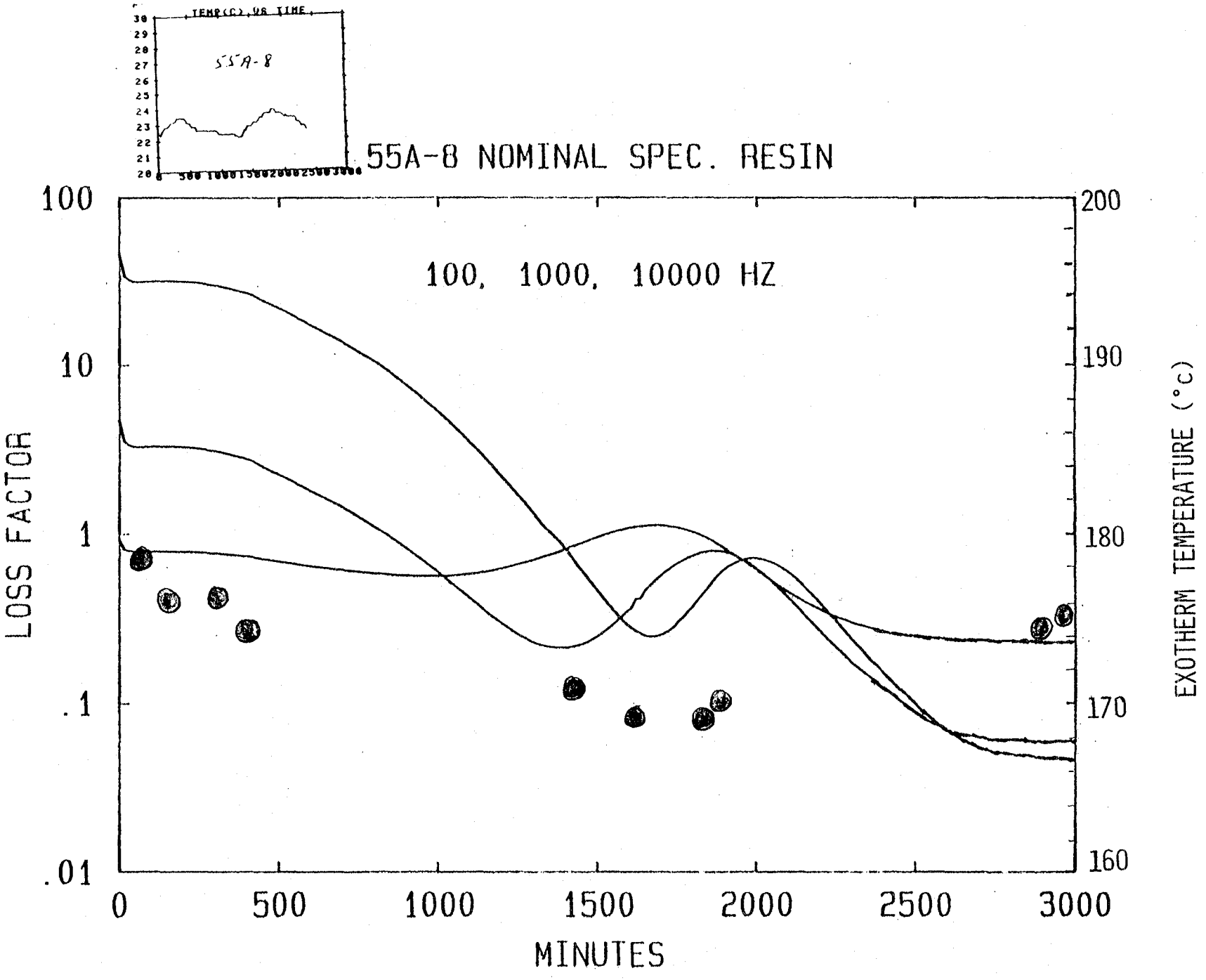


FIGURE 7: LOSS FACTOR VS TIME PLOT OF NOMINAL RESIN TO CATALYST RATIO OVERLAID WITH EXOTHERM

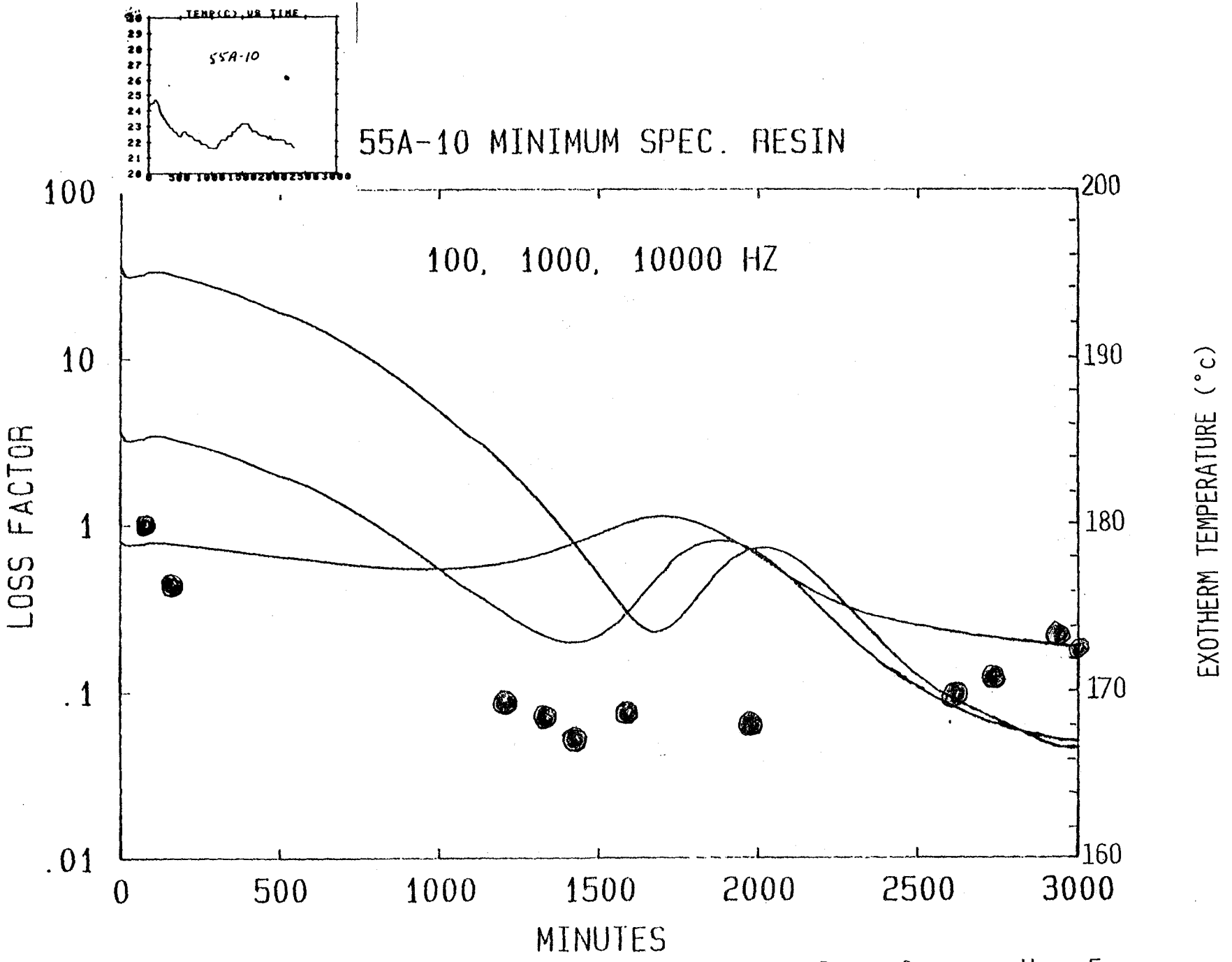


FIGURE 8: Loss Factor vs Time Plot of Minimum Resin to Catalyst Ratio Overlaid With Exotherm Temperature

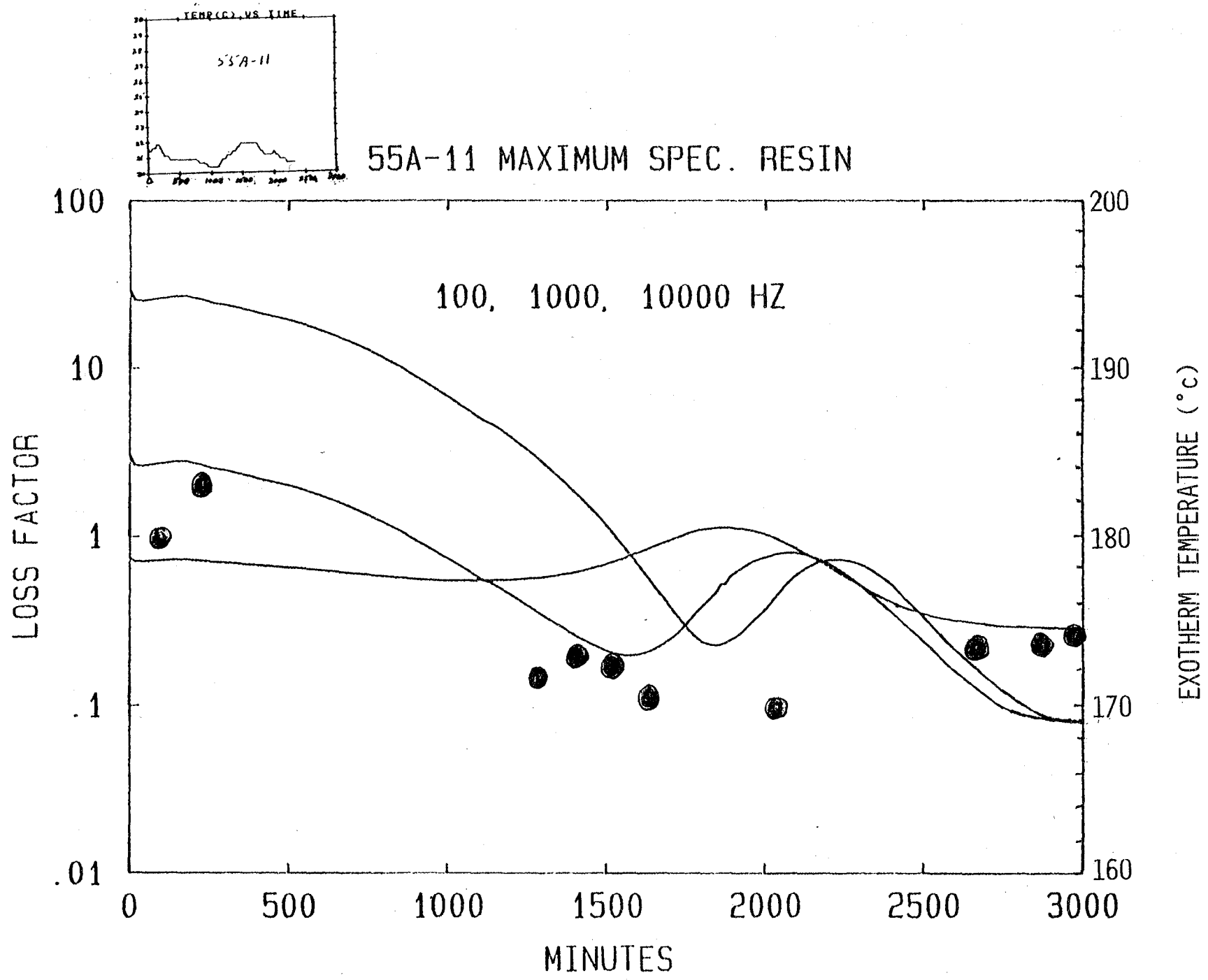
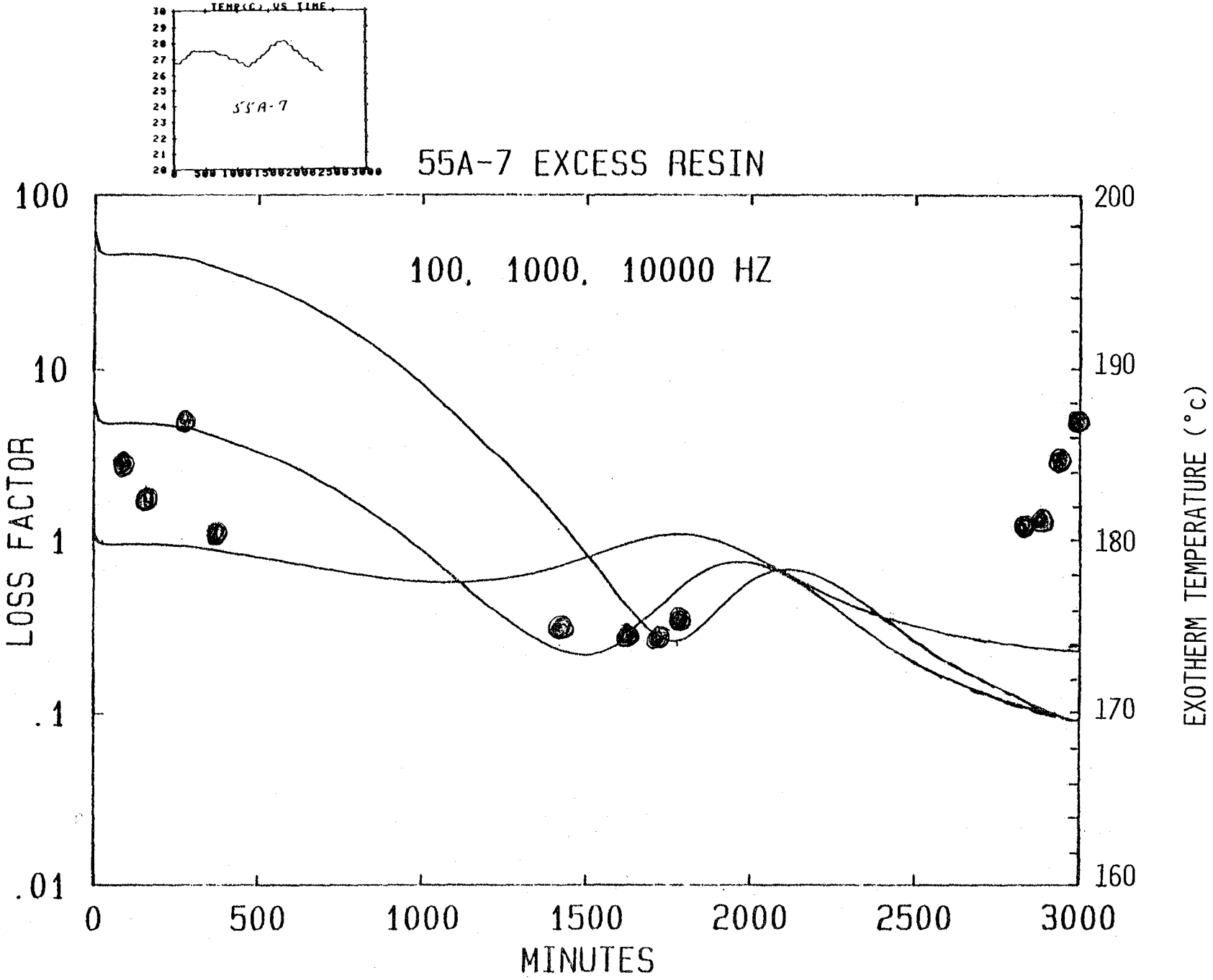


FIGURE 9: Loss Factor vs Time Plot of Maximum Resin to Catalyst Ratio Overlaid With Exotherm



91-IXXX-16

FIGURE 10: Loss Factor vs Time Plot of Excess Resin to Catalyst Ratio Overlaid With Exotherm Temperature

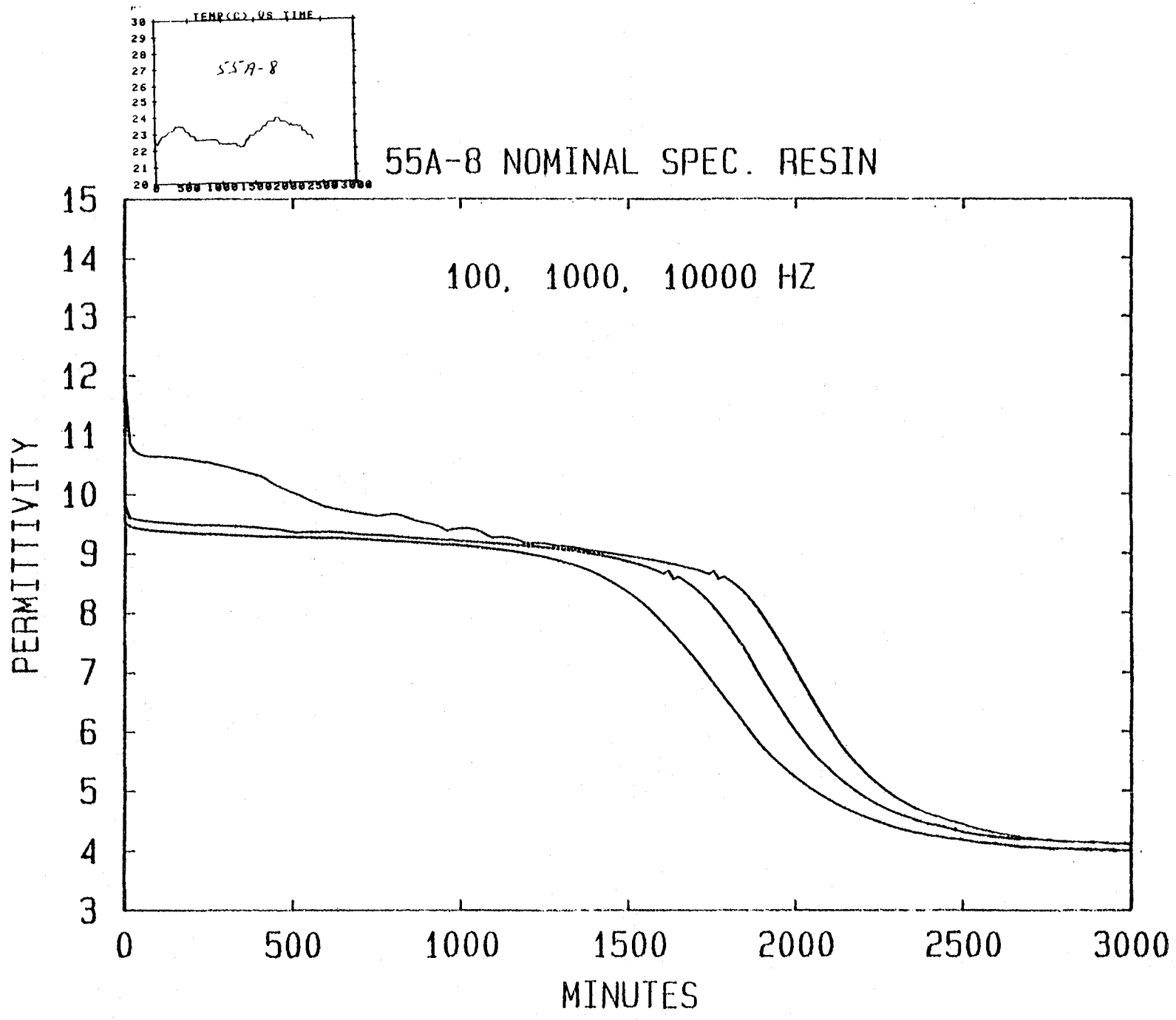


FIGURE 11: PERMITTIVITY VS TIME PLOT OF NOMINAL RESIN TO CATALYST RATIO CURE CYCLE

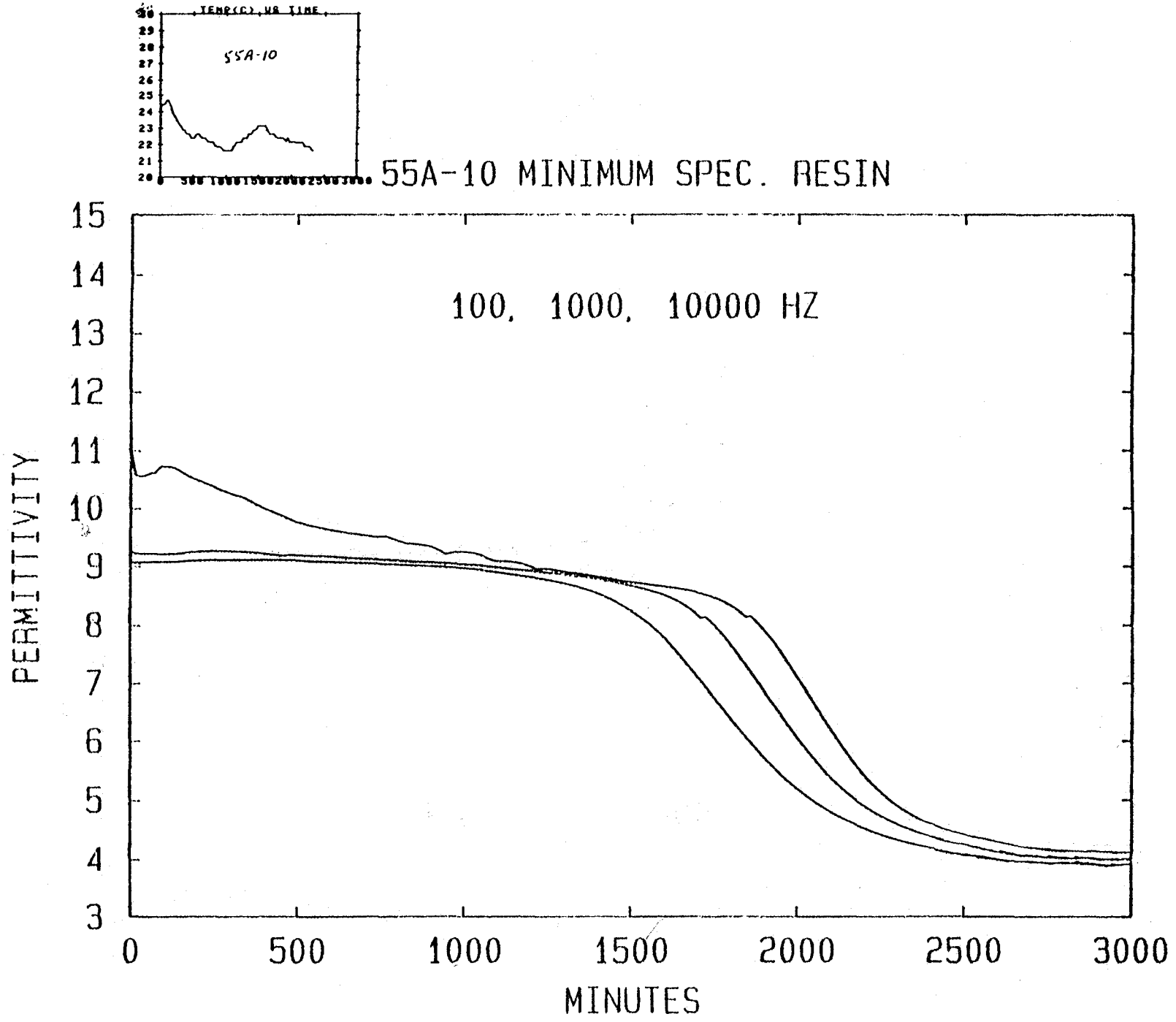


FIGURE 12: PERMITTIVITY VS TIME PLOT OF MINIMUM RESIN TO CATALYST RATIO CURE CYCLE

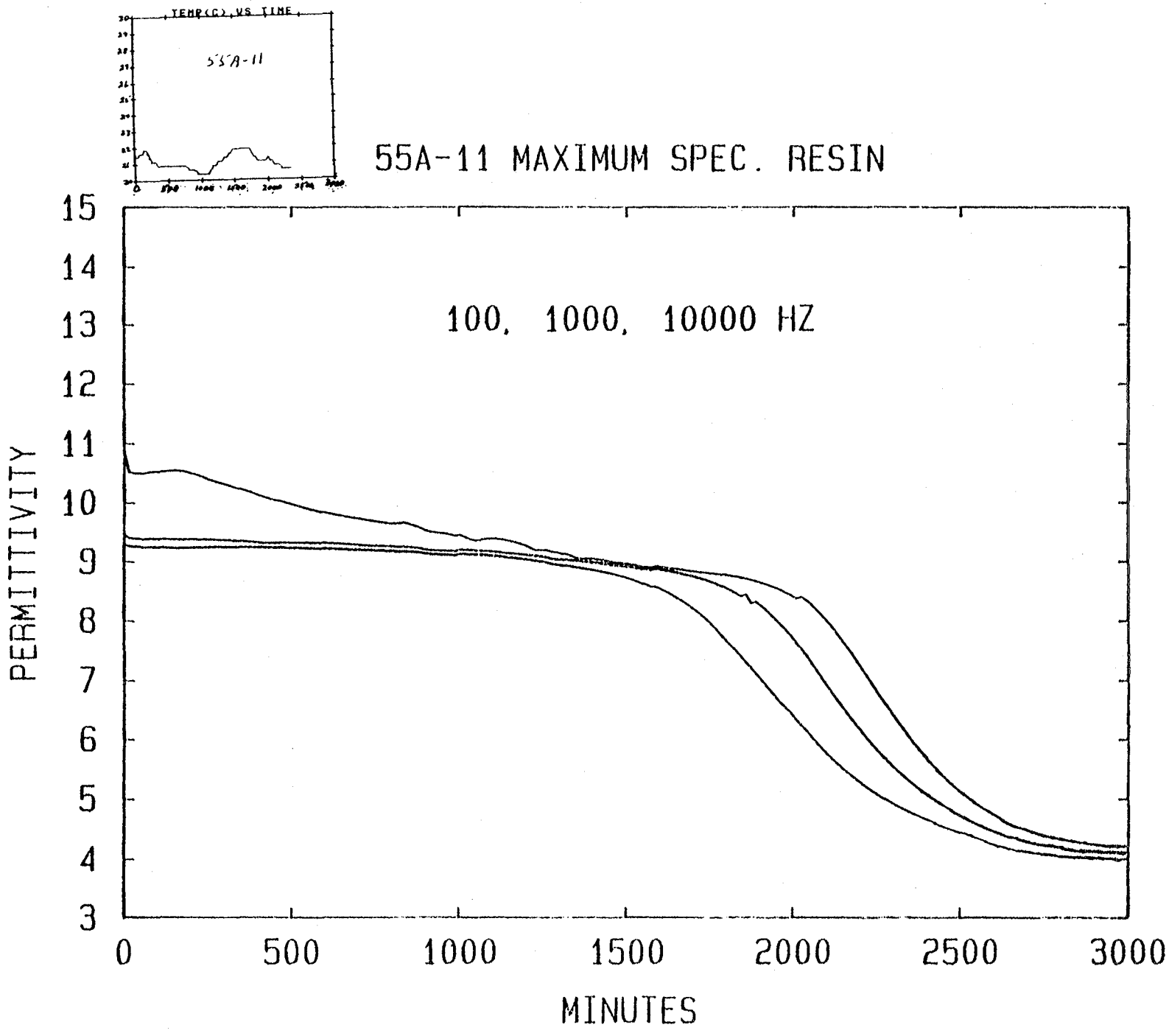


FIGURE 13: PERMITTIVITY VS TIME PLOT OF MAXIMUM RESIN TO CATALYST RATIO CURE CYCLE

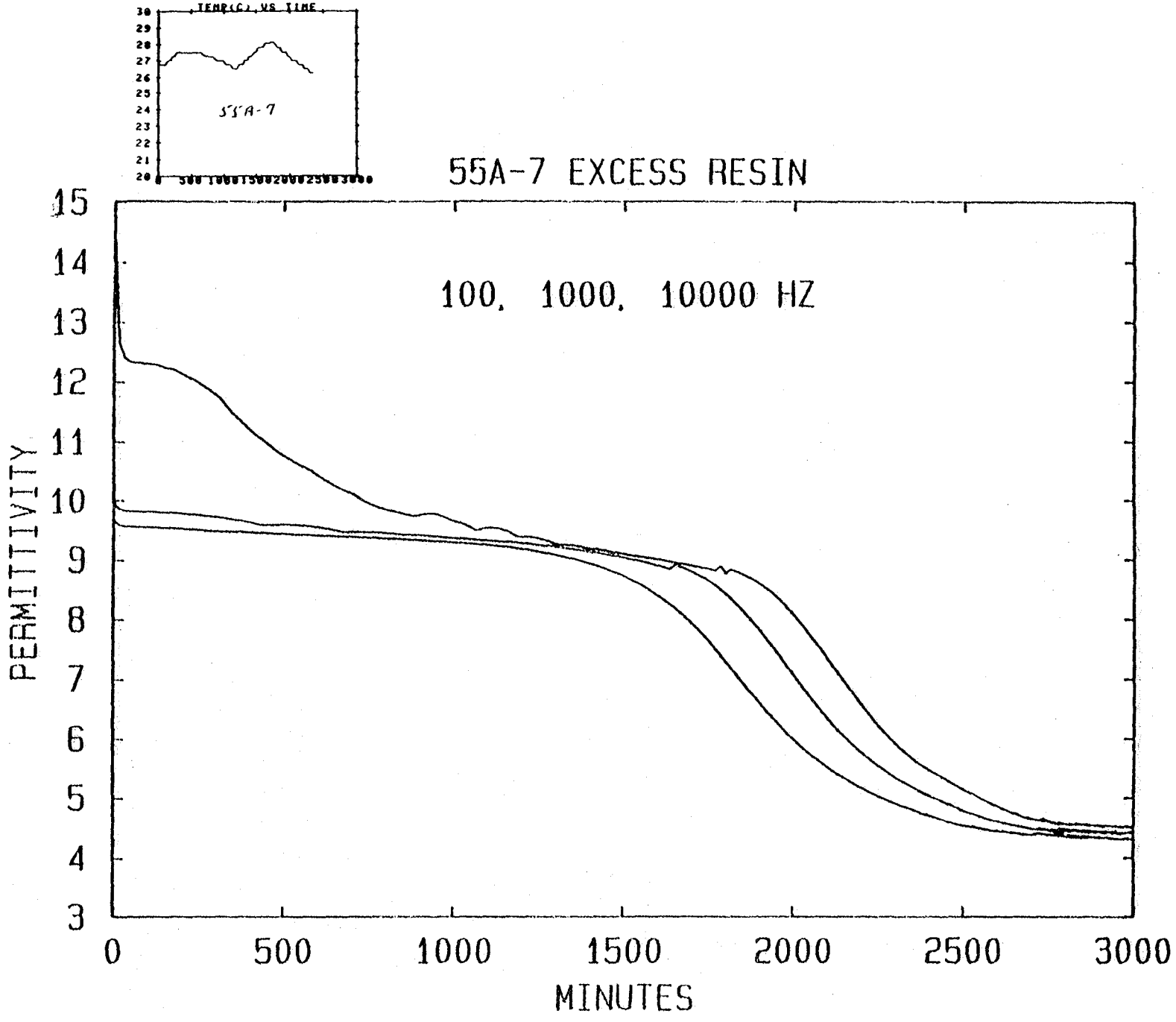


FIGURE 14: PERMITTIVITY VS TIME PLOT OF EXCESS RESIN TO CATALYST RATIO

REFERENCES

1. B.E. Goldberg and M.L. Semmel: Dilelectric Cure Monitoring: Preliminary Studies, NASA TM-86452, May 1984.

2. N.F. Sheppard Jr. and S.D. Senturia: Chemical Interpretation of the Relaxed Permittivity During Epoxy Resin Cure, TLSP Technical Report, 1 May - 30 November 1984, Massachusetts Institute Technology, Cambridge, Ma.

3. Z.N. Sanjana: Use of Microdielectrometry in Monitoring the Cure of Resins and Composites, TLSP Technical Report, January - July, 1984, Westinghouse Research and Development Center, Pittsburgh, Pa.

4. Z.N. Sanjana: Monitoring Cure of Epoxy REsins Using a Microdielectrometer, TLSP Technical Report, February 1982 - January 1984, Westinghouse Research and Development Center, Pittsburgh, Pa.

5. S.D. Senturia: Electrical Monitoring of Polymerization Reactions with the Charge-Flow Transitor, TLSP Final Report, July 1978 - May 1984, Massachussetts Institute Technology, Cambridge, Ma.

6. N.F. Sheppard Jr., M.C. Coln, and S.D. Senturia: A Dielectric Study of the Time-Temperature-Transformation (TTT) Diagram of DGEBA Epoxy Resins Cured with DDS (Diaminodiphenylsulfone), TLSP Technical Report, June - December 1983, Massachusetts Institute Technology, Cambridge, Ma.

7. S.D. Senturia, N.S. Sheppard, S.L. Garverick, H.L. Lee and D.R. Day: Microdielectrometry, TLSP Technical Report, April 1980 - June 1981, Massachusetts Institute Technology, Cambridge, Ma.

8. N.F. Sheppard, S.L. Garverick, D.R. Day, and S.D. Senturia Microdielectrometry: A New Method for in situ Cure Monitoring. TLSP: Technical Report, September 1979 - December 1980, Massachusetts Institute Technology, Cambridge, Ma.

9. J.D.B. Smith: The Dielectric Properties of Anhydride-Cured Epoxy Potting Compounds, Society for the Advancement of Material and Process Engineering, 1972, p. 476-488.

10. G.W. Lawless: High Temperature Dielectric Study of Epoxy Resins, Polymer Engineering and Science, V. 20, No. 8, May 1980, p. 546-550.

11. L.S. Penn: Physiochemical Characterization of Composites and Quality Control of Raw Materials, Americal Society for Testing and Materials, 1979, p. 519-532.

12. G.W. Lawless: Use of Automatic Dielectrometry to Study Composite/Resin Cure Cycles, Society for the Advancement of Material and Process Engineering, 1978, p. 610-621.

13. C.A. May, D.K. Hadad, and C.E. Browning: Physiochemical Quality Assurance Methods for Composite Matrix Resins, New York Society of the Plastics Industry, Inc., 1978, p. 15-D1 - 15-D6.

14. M.J. Yokota: In-Process Controlled Curing of Resin Matrix Composites, Society for the Advancement of Material and Process Engineering, 1977, p. 416-429.

15. C.A. May, J.S. Fritzen, and G.R. Brown: Analysis of Resin Matrix Changes During Cure, Metallurgical Society of AIME, 1976, p. 264-284.

16. B.G. Martin: Monitoring Cure Cycle by Dielectric Analysis, Materials Evaluation, V. 34, March 1976, p. 49-54.

17. R.A. Pike, F.C. Douglas, and G.R. Wisner: Electrical Dissipation Factor-Guide to Molding Quality Graphite/Resin Composites, Polymer Engineering and Science, V. 11, November 1971.

1985

NASA/ASEE SUMMER FACULTY RESEARCH FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

RADIATIVELY DRIVEN WINDS FROM MAGNETIC, FAST-ROTATING STARS

Prepared By:	Steven Nerney, Ph.D.
Academic Rank:	Associate Professor
University and Department:	Naval Postgraduate School Physics Department
NASA/MSFC: Division: Branch:	Space Science Laboratory Solar-Terrestrial Physics Solar Physics
MSFC Counterpart:	S. T. Suess
Date:	August 28, 1985
Contract No.:	NGT 01-008-021 The University of Alabama in Huntsville

RADIATIVELY DRIVEN WINDS FROM MAGNETIC, FAST-ROTATING STARS

BY

Steven Nerney

ABSTRACT

An analytical procedure is developed to solve the magnetohydrodynamic equations for the stellar wind problem in the strong-magnetic field, optically thick limit for hot stars. The slow-mode, Alfven, and fast-mode critical points are modified by the radiation terms in the force equation but in a manner that can be treated relatively easily. Once the velocities at the critical points and the distances to the points are known, the streamline constants are determined in a straight-forward manner. This allows the structure of the wind to be elucidated without recourse to complicated computational schemes.

I. INTRODUCTION

Since the study of Weber and Davis (1967, WD) showed that the sun could be spun down by the solar wind over its main sequence lifetime, work has continued on angular momentum loss in the solar wind and in stellar winds (Pizzo, et al., 1983; Nerney and Suess, 1975; Nerney, 1980; Goldreich and Julian, 1970; and Friend and MacGregor, 1984).

Early-type stars present additional problems in understanding angular momentum loss as stellar winds from O and B stars are primarily driven by their intense flux of ultraviolet radiation. The radiation has an associated outward force due to Thomson scattering by photospheric electrons as well as a line radiation force that acts on metals in the wind. Unfortunately, analytical solutions are not possible as the basic stream-line constants are not known until a full numerical solution is generated. The complexity arises because the force due to line radiation depends nonlinearly on the acceleration in the wind. The critical point analysis shows that there is a focus of singular points which the numerical solution must be tangent to. This is accomplished through an iteration scheme that produces an acceptable physical solution.

The purpose of this study is to attempt to analytically model radiatively driven stellar winds with strong magnetic fields from fast-rotating stars in the optically-thick limit. This will allow a simplification of the problem so that it becomes readily apparent how the streamline constants determine the solution. Also, the modified magnetohydrodynamic critical points are treated and a procedure is developed for constructing wind models which does not require a complicated interaction scheme.

II. EQUATIONS OF MOTION

The steady, spherically-symmetric, dissipationless, magnetohydrodynamic (MHD) equations, together with terms that give rise to radiative bulk forces may be written as follows:

(1) $4\pi r^2 \rho v_r = \dot{M}$, which is the continuity equation where $\dot{M}$ is the mass-loss rate in kg/sec, r is radial distance from the center of the star in the usual r,θ,ϕ spherical polar coordinate system, ρ is the mass density in kg/m^3, and v_r is the radial velocity of the wind.

$$(2)\quad v_r \frac{dv_r}{dr} = -\frac{1}{\rho}\frac{dP}{dr} - \frac{GM(1-\Gamma)}{r^2} + \frac{v_\phi^2}{r} - \frac{1}{8\pi\rho r^2}\frac{d}{dr}(rB_\phi)^2 + f_L$$

which is the radial momentum equation where P is the pressure, G is the gravitational constant, M is the mass of the star, Γ is a constant (see below) related to the radiative force due to Thomson scattering of continuum photons by electrons, v_ϕ is the rotational (azumuthal) velocity of the wind, B_ϕ is the azimuthal component of the magnetic field, and f_L is the force/kg due to line radiation. The term in B_ϕ is the radial component of the $\vec{j} \times \vec{B}$ force.

(3) $\Gamma = \frac{\sigma_e L}{4\pi GMC}$ where L is the stellar luminosity, c is the speed of light, and σ_e is the electron scattering opacity per unit mass.

$$(4)\quad f_L = \frac{\Gamma GM}{r^2} kt^{-\alpha}$$

which is an approximate law derived in the original theory of Castor, Abbott, and Klein (1975, CAK) where k is a constant which is a measure of the mixture of the number of strong radiation lines in the wind, t is an optical depth parameter, and α is a constant which depends on the importance of optically thick and thin radiation lines and is model-dependent.

(5) $t = \sigma_e \rho v_{th} \left(\frac{dv_r}{dr}\right)^{-1}$ and v_{th} is the thermal velocity of the ions which absorb and scatter radiation. CAK used values of $k = 1/30$ and $\alpha = .7$, but I will use $\alpha = 1$ and choose various values for the constant that k appears in. This Sobolev high-velocity gradient approximation to the line radiation force is discussed in Mihalas, 1978, p.561.

The azimuthal equation of motion and the induction equation (a form of Faraday's law) are the same as in Weber and Davis (1967, WD) and Friend and MacGregor (1984), namely:

$$(6) \quad \frac{d}{dr}(rv_\phi) = \frac{B_r}{4\pi\rho v_r} \frac{d}{dr}(rB_\phi)$$

$$(7) \quad (\nabla \times E)_\phi = 0 = \frac{1}{r} \frac{d}{dr}[r(v_r B_\phi - v_\phi B_r)]$$

as $\nabla \cdot \vec{B} = 0$ gives

(8) $B_r = f_b/r^2$ for spherical symmetry and as $B_r/\rho v_r$ is a constant, (6) integrates to

$$(9) \quad rv_\phi - \frac{B_r}{4\pi\rho v_r} rB_\phi = L$$ where L is the stream constant total angular momentum per unit mass which is the sum of the angular momentum carried in particles, rv_ϕ, as well as that in the electromagnetic field. Faraday's law integrates to

(10) $r(v_r B_\phi - v_\phi B_r) = \text{constant} = -\Omega^2 f_b$ and (9) and (10) together can be solved for v_ϕ and B_ϕ as in Weber and Davis (1967).

$$(11) \quad v_\phi = \frac{\Omega r}{v_{rA}} \frac{v_{rA} - v_r}{1 - M_A^2 r}$$

$$(12)\quad B_\phi = - B_r \frac{\Omega r}{v_{rA}} \frac{r_A^2 - r^2}{r_A^2 (1 - M_A^2 r)}$$

where Ω is the rotation rate of the region where the field lines are rooted (usually taken to be the photospheric value), f_b is the conserved radial magnetic flux, v_{rA} is the radial velocity at the Alfven radius (r_A) where v_r equals the speed of an Alfven wave in the medium, and M_{Ar} is the radial Alfven mach number.

$$(13)\quad M_A^2 r = 4\pi\rho \frac{v_r^2}{B_r^2} = \rho_a/\rho \ .$$

Closure is obtained by using the polytropic approximation.

(14) $P = P_o (\rho/\rho_o)^\gamma$ where P_o and ρ_o refer to values at the base of the wind (whether the base is in a corona or photosphere), and γ is the polytropic index.

This approximation is often used when the heating sources in the wind are not well-known. Generally, modellers replace this with an isothermal approximation for hot stars and use the stellar effective temperature. This is formally incorrect as it leads to inappropriate asymptotic states of the wind, Hundhausen (1972, p.9). However, the use of effective temperatures of even 50,000 K as for O stars does not lead to significant problems as it would for stars with coronae. Most early-type stars show x-ray emission. Therefore, I will examine a hot-wind solution and the polytropic approximation is appropriate.

The radial momentum equation, (2), can be simplified using $a^2 = \gamma p/\rho$ (a is the sound speed) together with

(15) $f_L = \frac{\Lambda}{\rho^{\alpha-1}} \left(\frac{udu}{dr}\right)^\alpha$. Using $\alpha = 1$, (2) eventually reduces to

$$(16)\quad \frac{r}{v_r}\frac{dv_r}{dr} = \frac{(v_r^2 - A_r^2)\left(2a^2 + v_\phi^2 - \frac{v_e^2}{2}(1-\Gamma)\right) + 2\, v_r\, v_\phi\, A_r\, A_\phi}{(v_r^2 - A_r^2)\,[v_r^2(1-\Lambda) - a^2] - v_r^2\, A_\phi^2}$$

where $A_r^2 = B_r^2/4\pi\rho$, $A_\phi^2 = B_\phi^2/4\pi\rho$, , the radial and azimuthal Alfven speeds, $v_e^2 = \frac{2\cdot GM}{r^2}$, the escape velocity at r, and

$$(17)\quad \Lambda = \left(\frac{4\pi}{\dot{M}\,\sigma_e\,v_{th}}\right)^{\alpha} \Gamma\, GMk \text{ or, for } \alpha = 1,$$

$$(18)\quad \Lambda = \frac{kL}{c\dot{M}\,v_{th}}$$

Λ is the essential radiation parameter in this wind model and can be estimated for a particular star. For instance, Friend and MacGregor (1984) model the 06ef star λ Cephei with $\alpha = .7$, $\dot{M} \approx 5.2 \times 10^{-6} M_\odot$/year, $T = 6.5\times10^4$K, $L = 6.76\times10^5 L_\odot$, and k was specified by using the CAK value of 1/30. These numbers give $\Lambda = .66$ in the current model. It is difficult to estimate Λ because k is model dependent, v_{th} depends on both the base temperature as well as the mass of the ions that absorb and scatter radiation, and mass-loss rates are not well-known for many stars.

Therefore Λ will be specified for a range of values and a variety of models will be calculated.

The general results of this study are based on the strong magnetic field limit of the MHD equations. This allows the explicit calculation of the position of the critical points based upon a similar prescription for non-radiative strong-field winds in the study of Hartmann and MacGregor (1982, HM).

III. THE CRITICAL POINTS

A. Critical Velocities

The WD theory gives rise to three critical points in the radial momentum equation. These points correspond to distances from the star where the speed of the stellar wind equals the speed of the characteristic MHD disturbances that occur in the fluid approximation; namely, the slow-mode wave, Alfven wave, and fast-mode wave. Setting $\Lambda = \Gamma = 0$ for the moment, the denominator of equation (16) goes to zero for three different wind speeds and, hence, the numerator must be zero at these critical points. These boundary conditions partially specify the physical characteristics of the wind. In the current study, radiation changes the nature of the critical points in a manner that can be explicitly shown.

The strong-field limit requires that $v_r^2 << A_r^2$ at the modified slow-mode point, r_s. This is equivalent to $m_{A\,r}^2 << 1$ at r_s.

This will be true for sufficiently strong magnetic fields and is justified a posteriori. Setting the denominator of (16) equal to zero:

$$A_r^2 \{v_r^2 (1 - \Lambda) - a^2 - v_r^2 A_\phi^2/A_r^2\} = 0$$

Faraday's law can be written as

$$\frac{v_r^2 A_\phi^2}{A_r^2} = (v_\phi - \Omega r)^2$$

and as v_ϕ is equal to the corotation value at r_s to $0(M^2_{Ar})$ we can neglect $v_r^2 A_\phi^2/A_r^2$ in the strong-field limit. Therefore,

$$(19) \quad v_{rs} = \frac{a_s}{1 - \Lambda}$$

which reduces to a well-known result when $\Lambda = 0$, $v_{rs} = a_s$.

At the Alfven point, the radial velocity is assumed to be much greater than its value at r_s so that

$$v_{rA}^2 (1 - \Lambda) >> a^2_A$$

Setting the denominator of (16) equal to zero.

(20) $v_r^2 = A_r^2 + \frac{A_\phi^2}{1-\Lambda}$ at r_a at which reduces to the well-known result $v_r = A$ at r_A when $\Lambda = 0$.

The modified fast-mode velocity is found by assumming $v_r^2 > A_r^2$ ($M_{Ar}^2 > 1$) at r_f. In a similar manner,

$$(21) \quad v_{rf}^2 = \frac{A^2_{\phi f}}{1 - \Lambda}$$

B. Optically-Thick Slow-Mode Point

Expanding Equation (11) for v_ϕ

$$(22) \quad v_\phi = \Omega r(1 - M_{Ar} + \frac{r^2}{r_A^2} M_{Ar} + \ldots)$$

For clarity, let $v_\phi = \Omega r(1 - \delta)$ where $\delta < 1$. Setting the numerator equal to zero at r_s

$$2a^2 - v_e^2/2 + v_\phi (v_\phi - 2V_r A_\phi/A_r) = 0 .$$

Using equation (10) gives

$$(23) \quad 2a^2 - \frac{v_e^2}{2} (1-\Gamma) + \Omega^2 r^2 = 0 \text{ to } 0(\delta^2)$$

as terms of $O(\delta)$ exactly cancel.

Using the non-dimensional parameters in HM, $Z_s = r_s/r_o$, $\Omega^2 = \alpha^2\, GM/r_o^3$, and the non-dimensional Parker critical point $Z_p \cong \frac{GM}{2r_o\, a_o^2}$, we find

$$(24)\quad Z_s \frac{a^2}{a_o^2} + \alpha^2\, Z_p\, Z_s^3 - Z_p\,(1-\Gamma) = 0$$

which reduces to equation (9) in HM for the isothermal limit. Equation (24) does not explicity depend on the radiation parameter, Λ. The distance to the slow-mode point is the same in the radiative wind (except for the hottest stars where Γ is not negligible) as in the non-radiative MHD wind, except that the radial velocity in the radiative case is greater by the factor $1/\sqrt{1-\Lambda}$.

Another point is that the calculation of Z_s assumes v_ϕ is given by the corolation value at r_s. This is accurate to $O(M_{Ar})$, yet due to a fortuitous cancellation, the equation for Z_s is accurate to $O(M^2_{Ar})$ so that these results and HM's are more accurate than anticipated.

To solve (24) for Z_s, we must know a^2/a_o^2 which depends on v_{ro}. In the isothermal limit, $a^2/a_o^2 = 1$ and this problem does not arise. However,

$$(25)\quad \frac{a^2}{a_o^2} = \left(\frac{v_{ro}\sqrt{1-\Lambda}}{a_o\, Z_s^2}\right)^{\beta} \text{ where}$$

$$(26)\quad \beta = \frac{2\,(\gamma-1)}{\gamma+1}$$

A second equation in Z_s and v_{ro} can be found by examining the energy/kg in the flow.

$$(27)\quad E = \frac{1}{2}(1-\Lambda)\, v_r^2 + \frac{1}{2}\, v_\phi^2 + \frac{a^2}{\gamma-1} - \frac{v_e^2}{2}(1-\Gamma) - \frac{\Omega r\, A_r\, A_\phi}{v_r}$$

The Poynting flux term can be calculated from equation (9) so that

$$(28)\quad E = \frac{1}{2}(1-\Lambda)\,v^2 + \frac{1}{2}v_\phi^2 + \frac{a^2}{\gamma-1} - \frac{v_e^2}{2} + \Omega L - \Omega r\, v_\phi$$

Consider the two terms in v_ϕ using $v_\phi = \Omega r(1-\delta)$

$$\frac{1}{2}v_\phi^2 - \Omega r\, v_\phi = -\frac{\Omega^2 r^2}{2}(1-\delta^2)$$

so that terms of $O(M_{Ar})$ have again cancelled.

Setting $E_o = E_s$ we find after some work

$$(29)\quad \left(\frac{1-\Lambda}{2}\right)\frac{v_{ro}^2}{a_o^2} - \left(\frac{1}{2}+\frac{1}{\gamma-1}\right)\left(\frac{v_{ro}\sqrt{1-\Lambda}}{a_o z_s^2}\right)^\beta =$$

$$2z_p\left\{\left(1-\frac{1}{z_s}\right)(1-\Gamma) - \frac{\alpha^2}{2}(z_s^2-1)\right\} - \frac{1}{\gamma-1}$$

This reduces to equation (13) in HM with $\gamma = 1$ and $\Lambda = 0$ provided we replace $\frac{a^2}{\gamma-1}$ by the isothermal term $a^2 \ell n\rho$.

Equations (24) and (29) can be self-consistently iterated between to find v_{ro} and z_s using HM's initial guess, modified for radiation

$$z_s,\ z_s = \frac{3z_p(1-\Gamma)}{1+3\,\alpha^{2/3}\,z_p(1-\Gamma)^{2/3}}$$

which assumes $a^2/ao^2 \approx 1$. This value for z_s allows the calculation of v_{ro} in (29). The iteration procedure rapidly converges to a solution.

C. Optically Thick Fast-Mode Point

We have shown that for $M^2_{ar} >> 1$ $v_{rf} \approx -\frac{A_\phi}{1-\Lambda}$ which together with equation (12) gives

$$v_{rf} = \frac{v_m^{3/2}}{\sqrt{1-\Lambda}\, v_{rf}^{1/2}} \left(1 - \frac{r_A^2}{r_f^2}\right)$$

where v_m is the usual Michel velocity, Michel (1969).

(30) $v_m^3 = \frac{\Omega^2 f_b^2}{\dot{m}}$ so that

(31) $v_{rf} \approx \frac{v_m}{(1-\Lambda)^{1/3}}$

In order to calcuate z_f, we need to know $v_{\phi f}$ which can be calcuated from (11):

$$v_{\phi f} \frac{L}{r} \{1 - \varepsilon\} \approx \frac{L}{r} \left(1 - \frac{v_{rA}}{v_{rf}}\right\} \approx \varepsilon < 1$$

which is formally correct with or without radiation.

Setting the numerator of (16) equal to zero at z_f:

(32) $2a^2 - v_e^2/2\ (1-\Gamma) + v_\phi\ (v_\phi + 2\ (v_\phi - L/r)) = 0$

The term in v_ϕ reduces to $\frac{L^2}{2}$ $(1 - 4\delta)$ to $O(M_{Ar}^{-1})$. to Equation (32) now reduces to

$$(33)\quad \frac{a_f^2}{a_o^2} Z_f^2 - Z_p Z_f (1-\Gamma) + \alpha^2 Z_p Z_A^4 - 2Z_A^2 \frac{v_m^2}{a_o^2} (1-\Lambda)^{1/3} = 0$$

where

$$(34)\quad \frac{a_f^2}{a_o^2} = \left(\frac{v_{ro}(1-\Lambda)^{1/3}}{Z_f^2 v_m} \right)^{\beta}$$

Equation (33) reduces to (17) in HM in the isothermal limit with $\Lambda = 0$ except for the term in $\frac{v_m^4}{a^4}$. This is $O(M_{Ar}^{-2})$ and should not appear as other terms of the same size have been neglected in calculating Z_f.

As Z_A appears in the calculation of Z_f we must generate another equation for Z_A and Z_f. The Alfven radius is not important except insofar as it is necessary in the calculation of Z_f. If the wind passes through Z_s and Z_f, it must of necessity pass through Z_A, Goldreich and Julian (1970).

Setting $E_o = E_f$

$$(35)\quad Z_A^2 = \frac{1}{2\alpha^2 Z_p} \left\{ \eta \frac{v_m^2}{a_o^2} + \frac{a_f^2/a_o^2 - 1}{\gamma - 1} - \frac{v_{ro}^2}{2a_o^2} \right\}$$

$$+ \frac{1}{\gamma^2} \left(1 - \frac{1}{Z_f}\right) (1-\Gamma) + \frac{1}{2}$$

with

$$(36)\quad \eta \equiv \frac{1}{2} (1-\Lambda)^{1/3} + 1$$

A good first approximation to Z_A neglects $1/Z_f$ and sets $a_f^2/a_o^2 \approx 0$. This allows the calculation of Z_f in (33) and the iteration proceeds as usual.

III. CONCLUSIONS

While the results of this study are preliminary, several important conclusions can be drawn from the analytical techniques that have been developed. The strong-magnetic field, optically thick limit of the MHD equations produces physical solutions through a tractable technique that shows both how the critical points are to be treated and gives the values of the velocities at these points as well as the distances to the critical points.

The streamline constants are determined once the solution has been generated at the critical points. In particular, the distance to the slow-mode point and the radial velocity at the base must be iterated on to produce physical solutions. Once this is done, the mass-loss rate can be determined from the value of the density in the photosphere together with the radius of the star. Additionally, the total energy/kg is also determined. The angular momentum/kg is determined after the iteration procedure to determine the Alfven radius and fast-mode radius is completed.

Preliminary numerical solutions have been generated and are being compared to the work of Friend and MacGregor, 1984, although it is still too early to report these results. These solutions have been found by using an IBM-PC and, in principle, could have been done on a hand calculator.

REFERENCES

1. Castor, J., D. Abott, and R. Klein, Astrophys. J. 195, 157.

2. Friend, D. and K. MacGregor, 1984, Astrophys. J. 282, 591.

3. Goldreich, P. G. and W. H. Julian, Astrophys. J., 1970 160, 971.

4. Hartmann, L. and K. B. MacGregor, 1982, Astrophys. J. 259, 180.

5. Hundhausen, A. J., 1972, Coronal Expansion and Solar Wind, Springer-Verlag.

6. Michel, F. C., 1969, Astrophys. J. 158, 727.

7. Mihalas, O., 1978, Stellar Atmospheres, second edition, University of Chicago Press.

8. Nerney, S. F. and S. T. Suess, 1975, Astrophys. J. 196, 837.

9. Nerney, S. F., 1980, Astrophys. J. 242, 723.

10. Pizzo, V., R. Schwenn, E. Marsch, H. Rosenbauer, K. Muhlhauser, and F. Neubauer, 1983, Astrophys. J. 271, 335.

11. Weber, E. J. and L. Davis, 1967, 148, 217.

1985

NASA/ASEE SUMMER FACULTY RESEARCH FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

INVESTIGATION OF LOW-TEMPERATURE SEMICONDUCTOR DEVICES

Prepared by:	D. R. Overcash, Ph.D.
Academic Rank:	Associate Professor, Chairman
University and Department:	Department of Math, Physics, and Computer Science Claflin College
NASA/MSFC: Division: Branch:	 Electrical Division Electrical/Electronic Parts Branch
MSFC Counterpart:	R. F. DeHaye
Date:	August 1, 1985
Contract No.:	The University of Alabama in Huntsville

INVESTIGATION OF LOW-TEMPERATURE SEMICONDUCTOR DEVICES

by

D. R. Overcash, Chairman
Department of Mathematics, Physics and Computer Sciences
Claflin College
Orangeburg, South Carolina

ABSTRACT

The objectives of this investigation are to determine the characteristics of semiconductor devices at low temperatures.

The output characteristics of several currently available semiconductor devices and devices fabricated at MSFC were measured over a range of gate voltages. Large variations in low-temperature performance, not only from type to type, but from FET to FET of the same type were obtained.

By increasing the carrier concentrations at low temperatures through extra heavy doping, a MOSFET device was fabricated that operated at 4.2K. To verify the low-temperature operation of a heavily doped device, this procedure was repeated. Their noise spectrum was analyzed at 4.2K.

Suggestions were made as to possible fabrication methods, techniques, and other semiconductor materials that might improve the device characteristics at low temperatures.

A procedure for using an electric field to drift dopants through the insulating oxide to produce a very thin, heavily doped region under the gate of a MOSFET was initiated.

Capacitance as a function of voltage plots was suggested as a method for determining the level of dopant drift.

LIST OF FIGURES

Free carriers necessary for a semiconductor devices operation can be generated by thermal excitation.[1] The objective of this investigation was to determine the characteristics of several semiconductor devices at low temperatures.

Carrier freezeout occurs in most semiconductor devices as they are cooled to liquid helium temperatures. The infrared telescope detector, to reduce noise and increase sensitivity, will be cooled to 1.5 K. Semiconductor devices will be needed to operate at or near this 1.5 K and to amplify and record the signal received by the infrared telescope. Figure 1 is the current and voltage characteristics of a J230 transistor at 300 K. The typical transistor response was observed, and as the transistor is cooled to 77 K in liquid nitrogen, figure 2 shows a decrease in aplification.

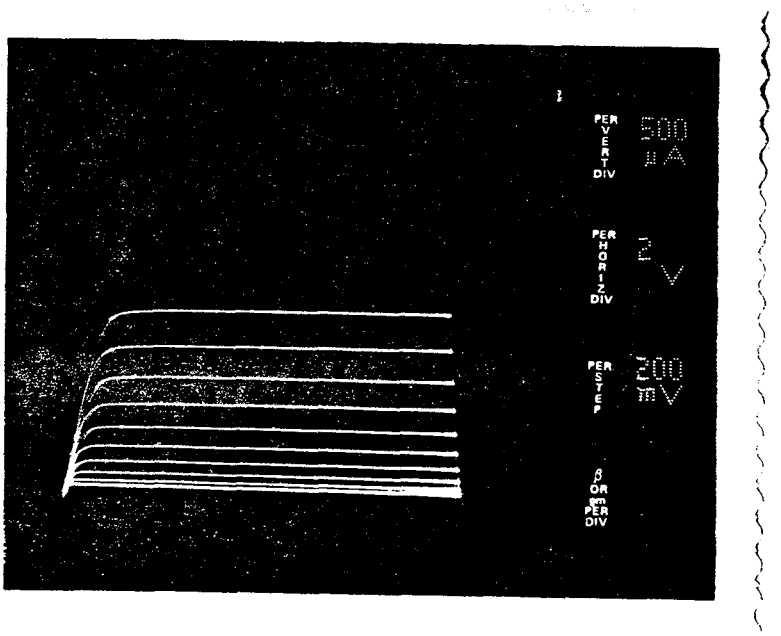


Figure 1. Current and voltage for a J230 at 300 K.

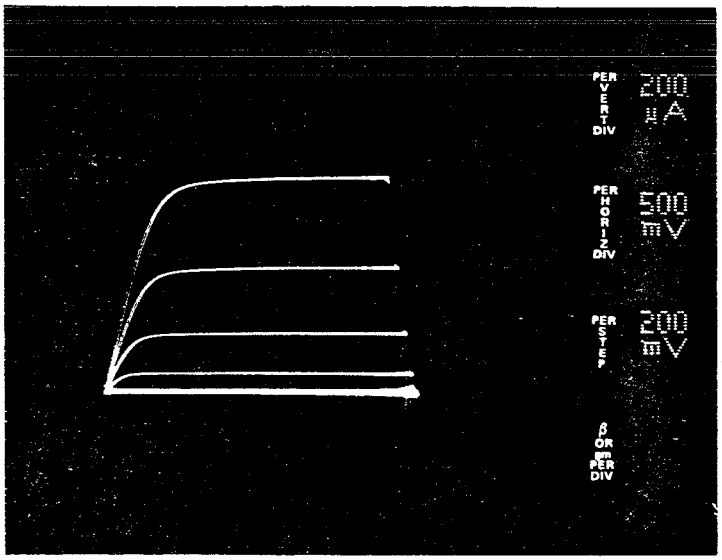


Figure 2. Current and voltage for a J230 at 77 K.

At 4.2 K, carrier freezeout has occurred and shows zero gain as shown in figure 3. The device returns to normal response as it is warmed to room temperature.

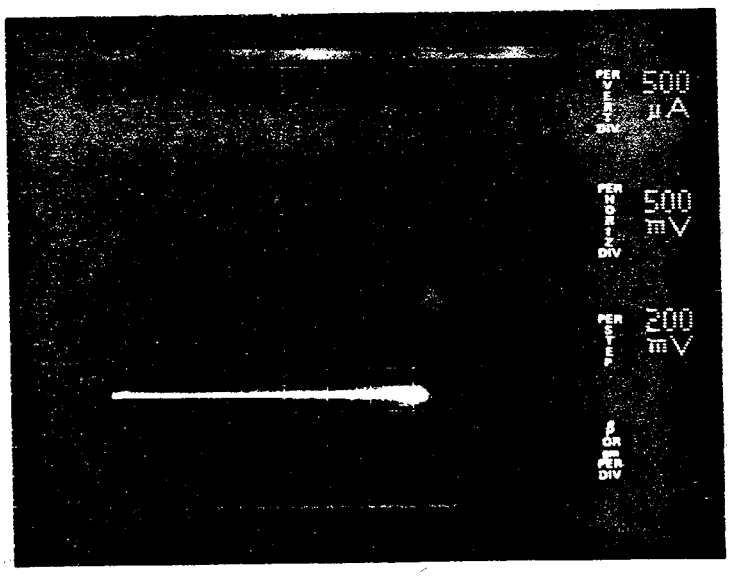


Figure 3. Current and voltage characteristics of a J230 at 4.2 K.

Five of these devices were measured at 4.2 K, and they all had zero gain with the detector at the same sensitivity as was used at 300K.

A 2N6483 FET semiuconductor was operated at 4.2 K at the University of Arizona.[2] As the temperature of the 2N6483 was lowered, its gain increased and then decreased. At the current sensitivity of 5 mA, no gain was observed. By increasing the current sensitivity to 0.5 mA, and by applying an offset voltage of 10 volts, the semiconductor operated as shown in figure 4.

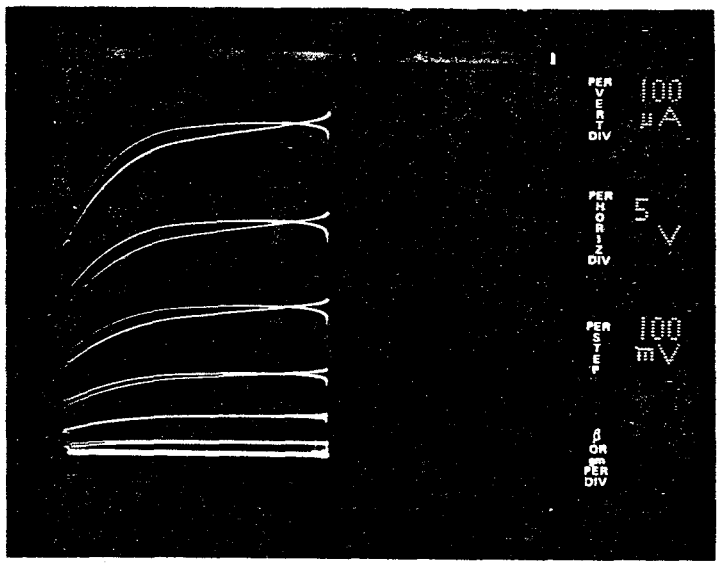


Figure 4. Current and voltage characteristics of a 2N6483 at 4.2 K.

The hysteresis is a measuring instrument problem and not a characteristic of the semiconductor. This offset voltage increases the risk of the device failure and is a property that is not recommended.

Heavy doping can generate carriers in semiconductors[3] at low temperatures. A MOSFET semiconductor was built here at the Marshall Space Flight Center and labeled 3-2. Gate 4 was measured, and the results are shown in figure 5, figure 6, and figure 7.

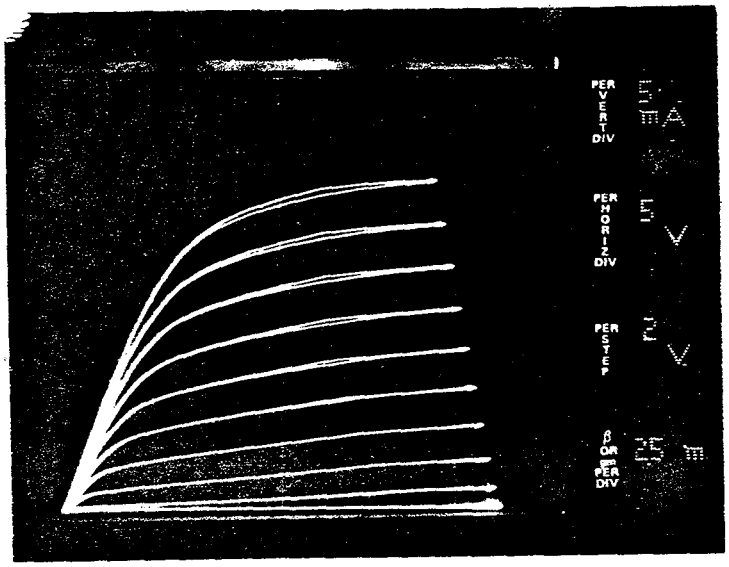


Figure 5. Current and voltage characteristics of gate 4 at 300 K.

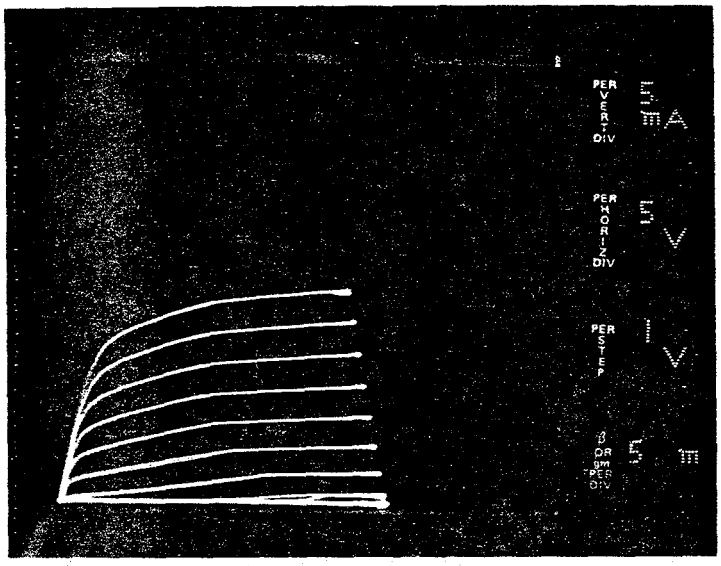


Figure 6. Current and voltage characteristics of gate 4 at 77 K.

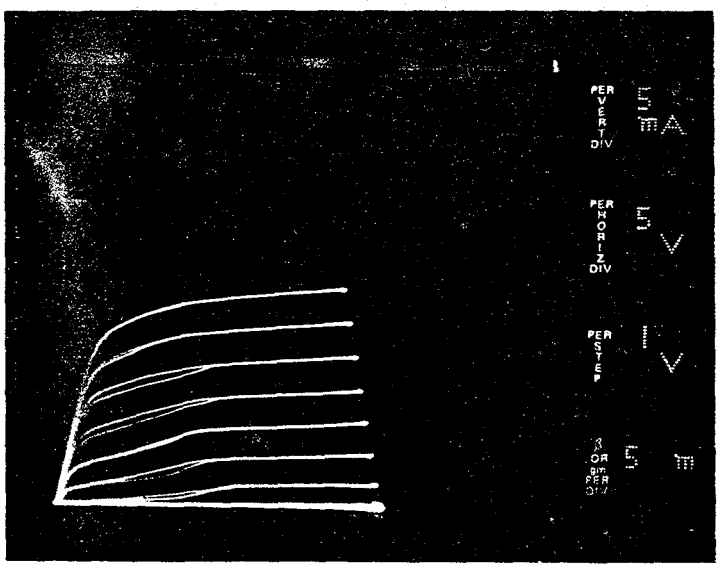


Figure 7. Current and voltage characteristics of gate 4 at 4.2 K.

The heavily doped p-channel semiconductor continues to operate at 4.2K.

Figure 8 shows the curves produced by a 3N165 at 77 K.

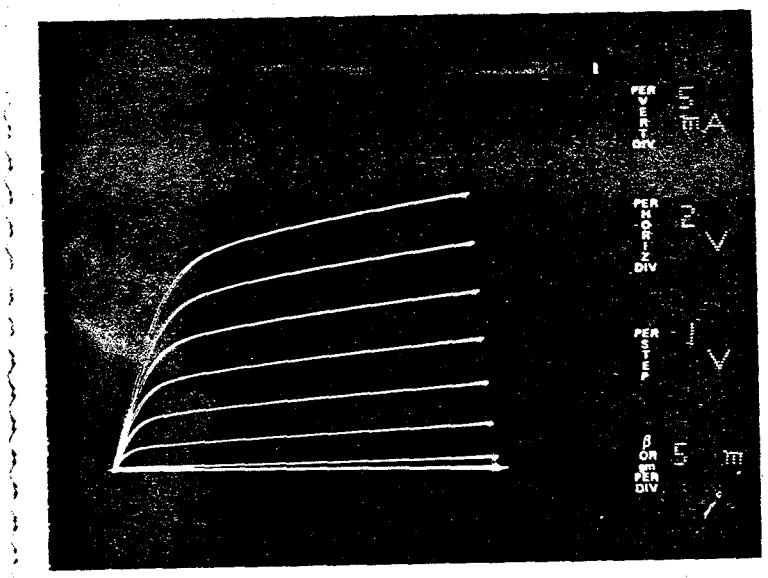


Figure 8. Current and voltage characteristics of a 3N165 at 77 K.

The calculated thermal noise for a TIXM12 is .16 nV/Hz and .26 nV/Hz^4 for the 2N6483 at 4.2 K, so thermal noise will be much smaller than the 1/f noise only at frequencies well above the infrared detector frequencies.

Cooling the 3N165 results in a gain amplification t 77 K. At 4.2 K, the device continues to operate as shown in figure 9.

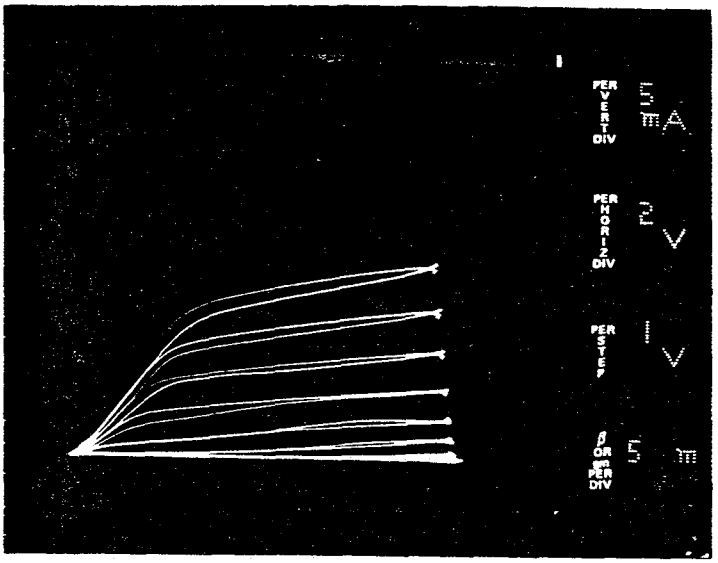


Figure 9. Current and voltage characteristics of a 3N165 at 4.2 K.

The 3N165 continues to operate without an offset voltage and has a gain equivalent to its gain at 300 K.

My recommendation is to:

Repeat the fabrication procedures to build another heavily doped semiconductor to check the reproducibility of the process.

Fick's law gives the relationship between the diffusion distance of an ion and time as an electric field is applied to a solid that contains ions.[5] I recommend the construction of an apparatus that may be used to apply an electric field to dopant ions placed on the gate oxide of a MOSFET. This electric field will diffuse the dopant ions into the gate channel and form a p-channel that is shallow and will not require an offset voltage to operate at low temperatures.

The use of voltage vs. capacitance measurement before and after heating be used to measure the dopant drift.

Measurements be continued on devices that have been found to operate, at low temperatures, by others such as the RCA 3N139,[5] SiAs JFET's,[6] and TIXM12.

The semiconductor devices that operate should have their noise spectrum measured.

The noise spectrum should be measured for the semiconductor devices that operate at 4.2K.

REFERENCES

1. Lengeler, S., Cryogenics, August (1974).
2. Rieke, G. H., Applied Optics, 20, No. 5, 814 (1981).
3. Johnson, Christina, Infrared Physics, 19, 617 (1979).
4. Hall, N.B., Applied Optics, 14, 450 (1975).
5. Hamilton, C.A., Cryogenics, November, 235 (1974).
6. Kelm, E. C., Review of Scientific Instruments, 39, 775 (1968).

1985

NASA/ASEE SUMMER FACULTY RESEARCH FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

RELIABILITY MODELS APPLICABLE TO SPACE TELESCOPE SOLAR ARRAY ASSEMBLY SYSTEM

Prepared By:	S. A. Patil, Ph.D.
Academic Rank:	Professor
University and Department:	Tennessee Technological University Mathematics and Computer Science
NASA/MSFC:	
Division:	Reliability and Engineering Office
Branch:	Reliability Engineering Branch, EG23
MSFC Counterpart:	Frank Pizzano
Date:	August 2, 1985
Contract No.:	NASA-NGT 01-008-021 (The University of Alabama)

LIST OF CONTENTS

RELIABILITY MODELS APPLICABLE TO
SPACE TELESCOPE
SOLAR ARRAY ASSEMBLY SYSTEM

ACKNOWLEDGEMENTS

My summer at the Marshall Space Flight Center has added greatly to my knowledge of space flights, in particular, my understanding of the Space Telescope and the Space Telescope Solar Array System. The understanding of the subject is mainly due to Mr. Frank Pizzano. Author is very grateful for this help from Mr. Pizzano. Ms. Julie Scott has cheerfully typed this mathematical manuscript, I am very thankful for her help. Also, thanks are due to Ms. Deborah Stephens for drawing the graphs.

Finally, the author is thankful to the Reliability Engineering Branch and MSFC for giving this opportunity.

RELIABILITY MODELS APPLICABLE TO SPACE TELESCOPE SOLAR ARRAY ASSEMBLY SYSTEM

By

S. A. Patil
Professor of Mathematics
Tennessee Technological University
Cookeville, Tennessee

ABSTRACT

A complex system may consist of a number of subsystems with several components in series, parallel, or combination of both series and parallel. In order to predict how well the system will perform, it is necessary to know the reliabilities of the subsystems and the reliability of the whole system. The objective of the present study is to develop mathematical models of the reliability which are applicable to complex systems. The models are determined by assuming k failures out of n components in a subsystem. By taking k = 1 and k = n, these models reduce to parallel and series models; hence, the models can be specialized to parallel, series combination systems. The models are developed by assuming the failure rates of the components as functions of time and as such, can be applied to processes with or without aging affects. The reliability models are further specialized to Space Telescope Solar Array (STSA) System. The STSA consists of 20 identical solar panel assemblies (SPA's). The reliabilities of the SPA's are determined by the reliabilities of solar cell strings, interconnects, and diodes. The estimates of the reliability of the system for one to five years are calculated by using the reliability estimates of solar cells and interconnects given in ESA documents. Aging effects in relation to breaks in interconnects are discussed.

1. Introduction

If events occur in time and the outcomes are countable numbers, the outcomes of the process can be described by a continuous time discrete stochastic process $N(t)$. The failures of various components in a complex system are countable events occurring in time, and hence, form a continuous time discrete stochastic process. Let $N_1(t)$, $N_2(t)$ ..., $N_n(t)$ be the outcomes of various failure event processes of various components in a system, and T_i be the random variable denoting the failure time of the ith component, then the reliability of this component is denoted by $R_i(t)$ and is defined as the probability that the component has not failed during time t. The reliability of the overall system $R(t)$ is the probability that the system has not failed during time t. This reliability $R(t)$ can be determined from the various mathematical models for the reliabilities of the components $R_i(t)$ and possible assembly of the components in parallel and series combination.

In this report, we shall develop mathematical models for systems which form continuous time discrete stochastic processes. We shall obtain the reliability models for the components of subsystems in which the failures occur randomly. Using these models, the reliability model for the system is obtained. The models are obtained for the failure rate $\lambda(t)$, which is a function of time t, and therefore, can be applied to a process in which failure rate is constant as well as time dependent failure rate which takes into account any aging effect. The models are formulated for series and parallel combination of components for which the reliabilities may be independent or nested.

The mathematical models developed will be applied in determining the reliability of the Space Telescope Solar Array (STSA) System. STSA System consists of two wings with two blankets in each wing. Each blanket is made up of five Solar Panel Assemblies (SPA's). The SPA consists of several strings. Each string consists of an array of solar cells in series and parallel, which are connected by interconnects and diodes. Assuming a reliability model for the solar cells and diodes, a reliability model will be built up for the whole STSA System. The reliability estimates for the STSA will be calculated for some models using the failure rates of solar cells and diodes given in European Space Agency (ESA)

documents. The estimates for the reliability of STSA System are obtained under different power losses in the STSA System. Also, the effects of breaks in solar cell interconnects will be discussed.

2. Stochastic Process, Failure Time and Reliability

A stochastic process is defined in Karlin and Taylor (1975) as a family of random variables determined by a process. A realization of a stochastic process is denoted by $N(t)$, $t \in (0, \infty)$. If t denotes the time and $N(t)$ corresponds to some outcome of the process, then $N(t)$ is called the time dependent stochastic process. If the random variable takes on countable values, 0, 1, 2, 3, ... the $N(t)$ process is called a discrete stochastic process. Many physical processes are time dependent discrete stochastic processes. If a complex system is working in time, then the failures of the components of the system occur randomly and the number of failures form a time dependent stochastic process $N(t)$. If $N(t_1)$, $N(t_2)$, ... denote the outcomes of a time dependent stochastic process $N(t)$, corresponding to a complex system at times t_1, t_2, ... t_n... then $T = t_n - t_{n-1}$, is a random variable and T is called the interevent or interarrival time of the process. If the outcomes of the stochastic process $N(t)$ corresponds to failure events of the process then T is called the failure time of the process and is a random measure of the time between two successive failures. Since T is random, it has a distribution function denoted by $F(t)$ and is defined by

$$F(t) = Pr \{ T \leq t \} . \qquad (1)$$

Here pr { } is abbreviated for probability. The reliability of the system is the probability that the system will not fail at least until time t, and is denoted by $R(t)$ and can be expressed by

$$R(t) = Pr \{ T > t \} = 1 - F(t) . \qquad (2)$$

Since time T is a continuous random variable, its distribution function $F(t)$ is continuous and also, differentiable. The derivative of $F(t)$ is called the probability density function (p.d.f.) of T and is denoted by $f(t)$. The p.d.f. $f(t)$ can be written as

$$f(t) = \frac{d}{dt} F(t) \qquad (3)$$

For a given distribution of a process $N(t)$, there is an unique p.d.f. $f(t)$, of T. Hence, the process $N(t)$ can be characterized by the p.d.f. of T.

The reliability $R(t)$ can be expressed in terms of $f(t)$ as

$$R(t) = \int_t^{\infty} f(t)\, d\, t \quad . \qquad (4)$$

and conversely,

$$f(t) = - \frac{d\, R\,(t)}{d\, t} \quad . \qquad (5)$$

From (4) the reliability can be obtained from the p.d.f of T. The theory of reliability and historical prospective is discussed by Barlow (1984), also in the paper, a large list of references on the topic is given.

3. Failure Rates

The failure rate $\lambda(t)$ is the rate of failure at which the components fail and the failure rate can be defined by

$$\lambda(t) = \lim_{dt \to o} Pr\,(t < T < t + \Delta\, t \mid T > t) \,.$$

$$= \lim_{dt \to o} P\{(t < T < t + \Delta\, t, T > t)\}/P(T > t)$$

$$= f\,(t)/(1 - F(t)\} = - \frac{d\, R(t)/}{R(t)}\, d\,(t) \qquad (6)$$

From the relation (6), the failure rate can also be expressed in terms of $R(t)$ as

$$\int_0^t \lambda\,(u)\, d\, u = -\ln\ R(t) \qquad (7)$$

and

$$R(t) = e^{-\int_0^t \lambda(u)\, du} \qquad (8)$$

The failure rate is also called the hazard function. The hazard function may be increasing function, constant function or decreasing function of t.

3.1 Constant Failure Rate

If the events occur in time randomly and in a short interval of time, at most one event can occur and the events in nonoverlapping intervals occur independently, then it can be shown that the resulting process is a poisson process and the number of events occurring in the process in given time, t can be shown to be

$$\Pr\{N(t) \leq n\} = \sum_{j=o}^{n} e^{-\lambda t} \frac{(\lambda t)^j}{j!}, \quad n=0,1,2... \qquad (9)$$

In this case, it can be shown that the waiting time for the process has a gamma distribution and the distribution of interevent time, T has the exponential distribution with p.d.f. f(t) given by

$$f(t) = \lambda e^{-\lambda t} \qquad t > o,\ \lambda > o \quad . \qquad (10)$$

The corresponding reliability function R(t) is given by

$$R(t) = e^{-\lambda t} \quad . \qquad (11)$$

and the failure rate function $\lambda(t)$ reduces to

$$\lambda(t) = \frac{d}{dt} \ln \{R(t)\} = \lambda e^{-\lambda t} / e^{-\lambda t} = \lambda.$$

In this case, the failure rate $\lambda(t)$ is a constant function.

3.2 Variable Failure Rate

The rate of failures in a small interval may not be constant, and the failure rate may be either increasing or decreasing. In the case of system components, the

components may have an aging effect and the hazard may increase with time, in this case $\lambda(t)$ is an increasing function of time and failure may occur more often such a system is called by Asher (1983) as a "sad" system. On the other hand, the process may be a learning process, in this case the hazard rate $\lambda(t)$ is a decreasing function of t, and the system in this case is called a "happy" system. There can be many different hazard functions associated with different phenomena. The failure rate which varies directly with a power of t can be defined by

$$\lambda(t) = \lambda\beta t^{\beta-1}, \quad t > o. \tag{12}$$

For $\beta=1$ the hazard rate is constant, for $\beta<1$ the hazard rate is decreasing and for $\beta>1$ the hazard rate is increasing. The model with the failure rate given in (12) is called Weibull model. Using equation (8) the reliability can be written as

$$R(t) = e^{-\lambda t^{\beta}}. \tag{13}$$

and the p.d.f. of T is given by

$$f(t) = \lambda\beta\, t^{\beta-1} e^{-\lambda t^{\beta}}, \quad t>o, \ \lambda>o. \tag{14}$$

It should be noted that the failure rate contains two parameters λ, β. Often $\beta=2$ is used, in this case failure rate is linear function of t.

Other failure rate models could be used in different situations. If the failure rate decreases sharply with time, then

$$\lambda(t)= e^{-\lambda t}. \tag{15}$$

is suitable.

If the failure rate decreases initially, then increases, the failure rate in such situation is discussed in Shooman (1968) is given by

$$\lambda(t) = \begin{cases} K_o - K_1 t, & o < t \leq \frac{K_o}{K_1} \\ o & \frac{K_o}{K_1} < t \leq t_o \\ K(t-to), & to < t < \infty \end{cases} \quad . \qquad (16)$$

4. Reliability Models

Suppose a system S consists of number of subsystems and each subsystem consists of number of components. The components may work in series, parallel or combination, also components may work independently or dependently. Different models would be used for different situations. Two types of model are discussed below.

4.1 Independent Components

Suppose a subsystem has m components and the components work independently and the system works satisfactory if at least k of the m components work. If all the components are different then from Basu and Mawaziny (1978) if $R_{\alpha i}(t)$ is the reliability of αith component then reliability of the subsystem $R_{si}(t)$ can be written as

$$R_{si}(t) = \sum_{j=k}^{m} \sum_{\alpha i} \prod_{i=1}^{j} (R_{\alpha i}(t)) \prod_{i=j+1}^{m} (1-R_{\alpha i}(t)) \quad . \qquad (17)$$

Where $\sum_{\alpha i}$ is taken over $\binom{m}{j}$ distinct values of αi.

If all the components are identical then R(t) reduces to

$$R_{si}(t) = \sum_{j=k}^{m} \binom{m}{j} (R(t))^{j} (1-R(t))^{m-j} \quad . \qquad (18)$$

The system could be called k redundant system, for k=1 the system is called completely redundant or parallel system, for k=n the system is completely nonredundant and system works if each component works and the system reduces to a series system.

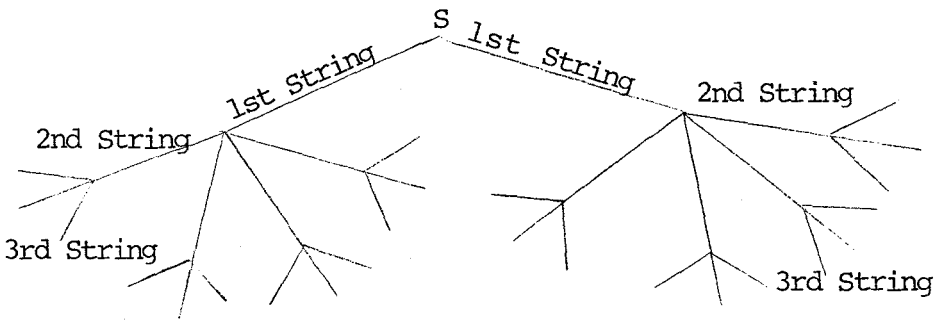
If there are s subsystems in a total system and the $R_i(t)$ is the reliability of components in ith subsystem and the subsystems are independent then the reliability of the system using (18) can be obtained as

$$R_S = \prod_{i=1}^{s} R_{si} = \prod_{i=1}^{s} \{\sum_{j=k}^{m} \binom{m}{j} (R_i(t))^j (1-R_i(t))^{m-j}\}. \quad (19)$$

The subsystems in a system may not be completely independent and the system might work with one or more subsystems working, in this case, the nested model may be useful.

4.2 Nested Model

In the nested models we consider the subsystems which work independently and the system may work with one or more subsystems, further, the probability of success of a subsystem may dependent on sub-subsystems and so on. Suppose there are y stages of a subsystem connected in strings, the system could be represented with a tree diagram. As an example, we consider a system with 3 stages with 2 components in 1st string, 4 components in 2nd string and 3 components in 3rd string. The system can be represented by the tree diagram as follows:



Suppose the components in each string are identical and the reliabilities of the upper string depend on the reliabilities of the lower branches. Suppose the system works if at least 2 components work in 3rd string, at least 2 in the 2nd string, and at least 1 in 1st string.

Then the reliability of the system is determined as

$$R_s = \sum_{j=1}^{2} \binom{2}{j} R_1^{j} (1-R_1)^{2-j} , \quad . \tag{20}$$

$$R_1 = \sum_{j=2}^{4} \binom{4}{j} R_{12}^{j} (1-R_{12})^{4-j} . \tag{21}$$

$$R_{12} = \sum_{j=2}^{3} \binom{3}{j} R_{123}^{j} (1-R_{123})^{3-j} . \tag{22}$$

The reliability R_{123} is either estimated or given then R_s can be obtained successive using (22), (21), and (20).

More generally, if the system consists of y stages and there are n_i identical components in ith stage and ith stage works if at least k_i components work and the system is completely nested then

$$R_s = \sum_{j=k_1}^{n_1} \binom{n_1}{j} R_1^{j} (1-R_1)^{n_1-j} . \tag{23}$$

$$R_1 = \sum_{j=k_2}^{n_2} \binom{n_2}{j} (R_{12})^{j} (1-R_{12})^{n_2-j} . \tag{24}$$

$$R_{12\ldots y-1} = \sum_{j=k_y}^{n_y} \binom{n_y}{j} (R_{12\ldots y})^{j} (1-R_{12\ldots y})^{n-j} . \tag{25}$$

The reliability of the system can be completely determined knowing the reliability of $R_{12\ldots y}$ and using equations (25), (24), and (23). Such models are used often in complex systems. The Space Telescope Solar Array could be based on such a model. Also, it is noted that if $k_i = n_i$, i=1, 2, y then each subsystem becomes series model and the whole system reduces model with no loss allowance.

5. Estimation of Parameters

In this section we discuss estimation of parameters which appear in the reliability expressions. We shall restrict to exponential and Weibull models as these are the models which are used often and we have given expressions for the reliabilities involving these models. We consider the maximum likelihood estimators (m.l.e), these are most widely used and have good statistical properties like asymptotic normality and consistency.

5.1 The Exponential Model

If t_1 t_2, ...t_n is a sample from the exponential distribution with p.d.f. in (10). The m.l.e. for λ is denoted by $\hat{\lambda}_t$ and is given by

$$\hat{\lambda}_T = n/\sum_{i=1}^{n} t_i \quad . \tag{26}$$

The estimator is biased but is consistent.

However, if the observations are made on the counting process, if $n_1 \ldots n_k$ are failures in k independent trials in given time t from the homogenous poisson distribution, then the maximum likelihood estimator for λ can be written as

$$\hat{\lambda}_N = \sum_{i=1}^{k} n_i/kt \quad . \tag{27}$$

This estimator is the minimum variance unbiased estimator for λ and its variance can be found. Both estimators can be used to find interval estimators for λ.

5.2 Weibull Model

If $t_1, t_2 \ldots t_n$ is a sample from the distribution with p.d.f given in (14), if β is known, then taking the likelihood of the sample the m.l.e of λ can be found as

$$\hat{\lambda}_T = n/\sum_{i=1}^{n} t_i^{\beta} \quad . \tag{28}$$

If the observations are made on the number of failures, for given time $t, n_1, n_2 \ldots n_k$ are failures in k independent trials from the nonhomogenous poisson process with intensity function $\lambda(t) = \lambda\beta t^{\beta-1}$ then the m.l.e of λ can be obtained as

$$\hat{\lambda}_N = \sum_{i=1}^{k} n_i / (kt^{\beta}). \tag{29}$$

If the shape parameter β is not known, then the maximum likelihood estimators of λ and β can not be obtained explicitly, however, the estimating equations for λ and β can be obtained. If $t_1, t_2 \ldots t_n$ are n independent failure time observations from the Weibull p.d.f in (14) then the m.l.e for $\hat{\beta}$ is obtained by solving

$$1/\hat{\beta} - (\sum_{i=1}^{n} t_i^{\hat{\beta}} \ln t_i)/(\sum_{i=1}^{n} t_i^{\hat{\beta}}) + (\sum_{i=1}^{n} \ln t_i)/n = o. \tag{30}$$

and the m.l.e of λ is obtained as

$$\hat{\lambda} = n/(\sum_{i=1}^{n} t_i^{\hat{\beta}}). \tag{31}$$

The m.l.e's of $\hat{\lambda}$, $\hat{\beta}$ and their properties are discussed by Cohen (1965).

The confidence interval estimators of λ can be obtained by using distributions of $\hat{\lambda}$. Large sample confidence intervals are discussed by Abernethy et al. (1983).

Crow (1974) has obtained explicit expressions for m.l.e's of β and λ based ordered observations for a repairable system.

6. Application to Space Telescope Solar Array System

The methodology developed in section 2-5 is now applied to Space Telescope Solar Array (STSA) System. First, we describe the STSA System.

6.1 Description of STSA System

Space Telescope is made up two identical wings. Each wing consists of two identical blankets. Each blanket is made up of five identical Solar Panel Assembly (SPA). There are 20 identical SPA's in STSA System. These SPA's are connected by 40 connecting diodes with 2 on each SPA. Each SPA is built up of three strings of solar cells. Each string has a length of 106 solar cells. Two end strings have eight cells in parallel and the middle string has seven cells in parallel. Strings 1 and 3 are built by 848 cells each whereas the middle string is made up of 742 cells. Each string is further broken into 7 substrings of 14, 15, or 16 cells so the substring consists of either 14, 15, and 17 long and 8 or 7 cells wide. There are seven shunting diodes. Each SPA has 21 shunting diodes and 2,438 solar cells. The cells are connected by Cell Interconnects (CIC's). There are the same number of CIC's as the cells. The STSA System consists of 48,760 cells, same number of CIC's, 420 shunting diodes, 40 connecting diodes and same number of solders. The description of STSA System is given in a number of ESA documents. Two of these are GL-SA-B002 and AN-1367-108.

Each individual cell is made up of silicon and 20 mms. wide, 40 mms. long, and generates .349 volts and carries current .300 amp. at 55^{o}. However, total voltage at SA/SSM interface is approximately 34 volts and working current is .27 amps. Approximate power without any loss of cells diodes and CIC's is 4.18 watts per string. Since 460 parallel strings in STSA System, STSA produces approximately 4,222.8 watts. The losses due to solder and other connectors in current and voltages are not clear. The estimates mentioned are taken from GL-SA-B002.

6.2 The Reliability of STSA System

The reliabilities of photovoltaic devices and system are discussed in a special issue of IEEE Transaction (1982) Vol. R-31. The papers of some interest on the topic are, "Photovoltaic Module Reliability Improvement Through Application Testing and Failure Analysis" by Dumas and Shumk (1982). "Reliability Terminology and Formulae for Photovoltaic Power System" by Lauftenburger and Anderson (1982), and "A Methodology for Photovoltaic System Reliability and Economic Analysis" by Stember Huss and Bridgman (1982).

The reliability of STSA System depends on the reliabilities of the various components in the system, the design of the components, the definition of the failure,

and the mathematical models used in determining the reliabilities of the components. Most of the estimates for failure rate are based on the constant failure rate, which arises out of the exponential models as these models have been used by ESA. However, estimates based on Weibull model could be obtained if some estimates or knowledge of shape parameter is assumed.

First, we discuss the reliabilities of blocking and shunted diodes. For connecting diodes, since there are 40 blocking diodes, if all of them have to work, then we find

$$R_0 = (R_{BD})^{40}, \tag{32}$$

R_1 is the reliability of individual connecting diode.

If the success is defined with k (k is an integer close to 40) diodes working the R_1 reduces to

$$R_1 = \sum_{j=k}^{40} \binom{40}{k} R_{BD}^{j} (1-R_{BD})^{40-j}. \tag{33}$$

For shunting diodes, the reliability, when each diode is working, is

$$R_0 = (R_{SD})^{420}. \tag{34}$$

For k diodes working, the reliability reduces to

$$R_1 = \sum_{j=k}^{420} \binom{420}{j} (R_{SD})^{j} (1-R_{SD})^{420-j}. \tag{35}$$

The reliabilities R_{SD} for the exponential for given time t can be evaluated by

$$R_{SD} = e^{-\lambda_1 t},$$

Where λ_1 is the failure rate of the diodes to be evaluated from the data.

Next, we consider the reliabilities associated with CIC. Since there are 20 SPA's and each SPA has 3 strings with width 8 and 1 with width 7. Here, for calculating

reliability, we assume width 8. Each string has 106 CIC's in series. The reliabilities of CIC's by nested design can be written as

$$R_{CIC} = \sum_{j=k_1}^{60} \binom{60}{j} R_1^j (1-R_1)^{620-j} , \tag{36}$$

$$R_1 = \sum_{j=k_2}^{8} \binom{8}{j} R_{12}^j (1-R_{12})^{8-j}, \tag{37}$$

and

$$R_{12} = R_{123}^{106} . \tag{38}$$

Here k_1, k_2 are integers close to upper limits and R_{123} is the reliability of interconnect can be obtained from

$$R_{123} = e^{-\lambda_c t},$$

λc being failure rate of an interconnect.

We now discuss the reliability models associated with the solar cell arrays. In the first model, we assume that a substring consists of an average of 15 cells in series and each string has 7 substrings and there are 8 parallel strings associated with each string. Since there are 3 strings per SPA, out of 21 subgroups any number of subgroups might fail, a nested model could be used. Let $R_c = R_{1234}$ be the reliability of a cell, then

$R_{123} = R_{1234}^{15}$ is the reliability of a substring, and

$$R_{12} = \sum_{j=k_3}^{21} \binom{21}{j} (R_{123})^j (1-R_{123})^{21-j} . \tag{39}$$

R_{12} gives the reliability with k_3 substrings working. Since there are 8 strings in parallel, if some of these fail, then the power of the system would be affected. Hence, using k_2 parallel strings working the reliability of SPA is determined by

$$R_1 = R_{SPA} = \sum_{j=k_2}^{8} \binom{8}{j} (R_{12})^j (1-R_{12})^{8-j} . \tag{40}$$

Since there are 20 SPA's, the reliability of STSA System from cell failures can be reduced to

$$R_S = \sum_{j=k_1}^{20} \binom{20}{j} R_1^j (1-R_1)^{20-j} . \tag{41}$$

If there is no loss of power due to any cells in the entire system, then the reliability of the system reduces to

$$R_S = (R_c)^{48760}. \tag{42}$$

There are other models considered for the reliability of solar cell system. In AN-1367-108, it is suggested that each string should be considered as one unit. Since there are 460 strings of 106 cells long in the STSA System, the reliability R_S can be written as

$$R_S = \sum_{j=k}^{460} \binom{460}{j} R_1^j (1-R_1)^{460-j} , \tag{43}$$

Where R_1 is the reliability of the string and is given by

$$R_1 = R_c^{106} . \tag{44}$$

R_c is the reliability of a single cell and is determined from

$$R_c = e^{-\lambda_c t}$$

in the exponential case and for Weibull case R_c is given by

$$R_c = e^{-\lambda_c t^{\beta}} ,$$

$\lambda_c t^{\beta}$ could be determined from a set of data.

7. Estimates of Reliabilities

In this section, we calculate the reliabilities associated with diodes, CIC's and solar panels. These reliabilities are then used to calculate the reliabilities of STSA System. In calculating the reliabilities, we are using various models proposed in section 6. The estimates are based on failure rate estimates given in ESA documents.

Since these estimates are based on the exponential models, we have to resort to these models. The reliabilities are calculated from one to five years and presented in a tabular form as well as in graphical form. First, we calculate the reliabilities of the diodes.

7.1 Reliability Estimates of Diodes

For blocking diodes, the failure rate given in TN-SA-B147 is $\lambda = 1.2 \times 10^{-9}$/hr., hence, the reliabilities are calculated from $R_{BD}(t) = e^{-\lambda t}$. Since there are 40 diodes, if all of them have to work reliabilities are calculated from

$$R_o(t) = (R_{BD})^{40} . \tag{45}$$

If we allow one of them to fail then

$$R_1(t) = (R_{BD})^{40} + 40\ (R_{BD})^{39}\ (1-R_{BD}) \tag{46}$$

Using these formulae, we find the reliabilities assuming one year is 8,760 hours. The results are presented in table I.

Table I. Reliabilities of Blocking Diodes

Time t (in years)	$R_{BD}(t)$	$R_o(t)$	$R_1(t)$
1	.99998948	.99957961	.99999991
2	.99997897	.99915939	.99999965
3	.99996846	.99873935	.99999922
4	.99995795	.99831949	.95999862
5	.99994744	.99789981	.99999785

Next, we find the reliabilities for the shunting diodes. Since there are 420 shunting diodes and failure rate of each diode is $\lambda = 1 \times 10^{-9}$, the reliabilities for all diodes working are obtained from

$$R_D(t) = e^{-\lambda t}$$

$$R_o(t) = (R_{SD})^{420} . \tag{47}$$

If we allow in our success criteria one diode to fail, then the reliabilities are calculated from

$$R_1(t) = (R_{SD})^{420} + 420\ (R_{SD})^{419}(1-R_{SD}). \qquad (48)$$

The reliabilities for different years are given in table II.

Table II. Reliabilities of Shunting Diode

Time t (years)	$R_{SD}(t)$	$R_o(t)$	$R_1(t)$
1	.99999124	.99632760	.99999328
2	.99998248	.99266860	.99997314
3	.99997372	.98902309	.99993969
4	.99996496	.98539096	.99989302
5	.99995620	.98177217	.99983327

7.2 Reliability Estimates of Solar Cells

We now discuss the reliabilities of solar cells. First, we use the nested model discussed in section 6. For the series model

$$R_o(t) = R_S = (R_C)^{48760}. \qquad (49)$$

Also, taking $K_3 = 21$, $K_2 = 8$ and $K_1 = 20$ in equations for solar cells in section 6, we find

$$R_{123} = R_C^{15}. \qquad (50)$$

$$R_{12} = R_{123}^{21} + (21)\ (R_{123})^{20}\ (1-R_{123}). \qquad (51)$$

$$R_{SPA} = R_{12}^{8} + 8\ R_{12}^{7}\ (1-R_{12}). \qquad (52)$$

$$R_1(t) = R_S = R_1^{20}. \qquad (53)$$

This model allows a loss of up to 1 substring and up to 1 parallel string to fail in each of the 20 SPA's. This amounts to 4.9% power loss in STSA System. The loss figure is taken from TN-SA-B151.

Alternative model with .22% of power loss is considered. This is based on 460 total strings (counting parallel strings) in the STSA System. The loss corresponds to a loss of one string of 106 cells in series. The reliabilities are obtained by taking K = 459 in equation (43). The reliability of STSA System is calculated from

$$R_1 = R_C^{106}. \qquad (54)$$

$$R_1'(t) = R_1^{460} + 460\, R_1^{459}\,(1-R_1) \qquad (55)$$

For the cell, the failure rate from TR-STSA-42 is 1×10^{-9}/hr. Using this estimate and the above equations for various t's, the reliabilities $R_o(t)$, $R_1(t)$, and $R_1'(t)$ are calculated and presented in table III.

Table III. Reliabilities of Solar Cells in STSA System

Time t	$R_C(t)$	$R_o(t)$ (Zero Loss)		$R_1(t)$ (1 String Loss)	$R_1'(t)$ (.22% Loss)
1	.99999124	.65237696	=	.99999999	.93115794
2	.99998248	.42559154	=	.99999989	.78950169
3	.99997372	.27764476	=	.99999941	.63391835
4	.99996496	.18112816	=	.95999815	.49117021
5	.99995620	.11816326	=	.99999550	.37110988

The reliabilities of the CIC's could be calculated from the formulae for CIC's. However, the reliabilities for a CIC as functions of time are not available. Since there are the same number of CIC's as the cells and their configuration is similar to those of the cells, it is possible to combine the reliabilities of CIC's and cell and find the reliabilities of CIC's with cells in STSA System. The estimate of reliability for a CIC from AN-1367-108 is R= .9999976714. The reliability of CIC and cell is given by

$$R_C'(t) = (.9999976714)\, e^{-\lambda t} . \qquad (56)$$

We use this formula to calculate R_C' and the formulae $R_o(t)$, $R_1(t)$, and $R_1'(t)$ given in equation (49) to equation (55). The reliabilities for different t are given in table IV.

Table IV. Reliabilities of Solar Cells With CIC in STSA

Time t (in yrs.)	$R_C'(t)$	$R_o(t)$ (Zero Loss)	$R_1(t)$ (1 String Loss)	$R_1'(t)$ (.22% Loss)
1 yr.	.99998891	.58234352	.99999998	.89739874
2 yrs.	.99998015	.37990379	.99999982	.74797591
3 yrs.	.99997139	.24784048	.99999917	.59409527
4 yrs.	.99996263	.16168463	.99999760	.45687567
5 yrs.	.99995387	.10547829	.99999445	.34330636

7.3 The Reliabilities of STSA System

The STSA System consists of diodes, CIC's and solar cells. The reliability of STSA System R_S is calculated from

$$R_S(t) = R_D^{(t)} \; R_{Cell}^{(t)} \tag{57}$$

For zero loss $R_{S_o}(t)$ is calculated by multiplying $R_o(t)$ from tables I, II, and IV. For .22% power loss $R_{S_1}(t)$ is tabulated from $R_1(t)$ in tables I and II and R_1 in table IV, and these are tabulated in table V. Also, at the beginning there is no loss, hence, at time 0, the reliability is 1.

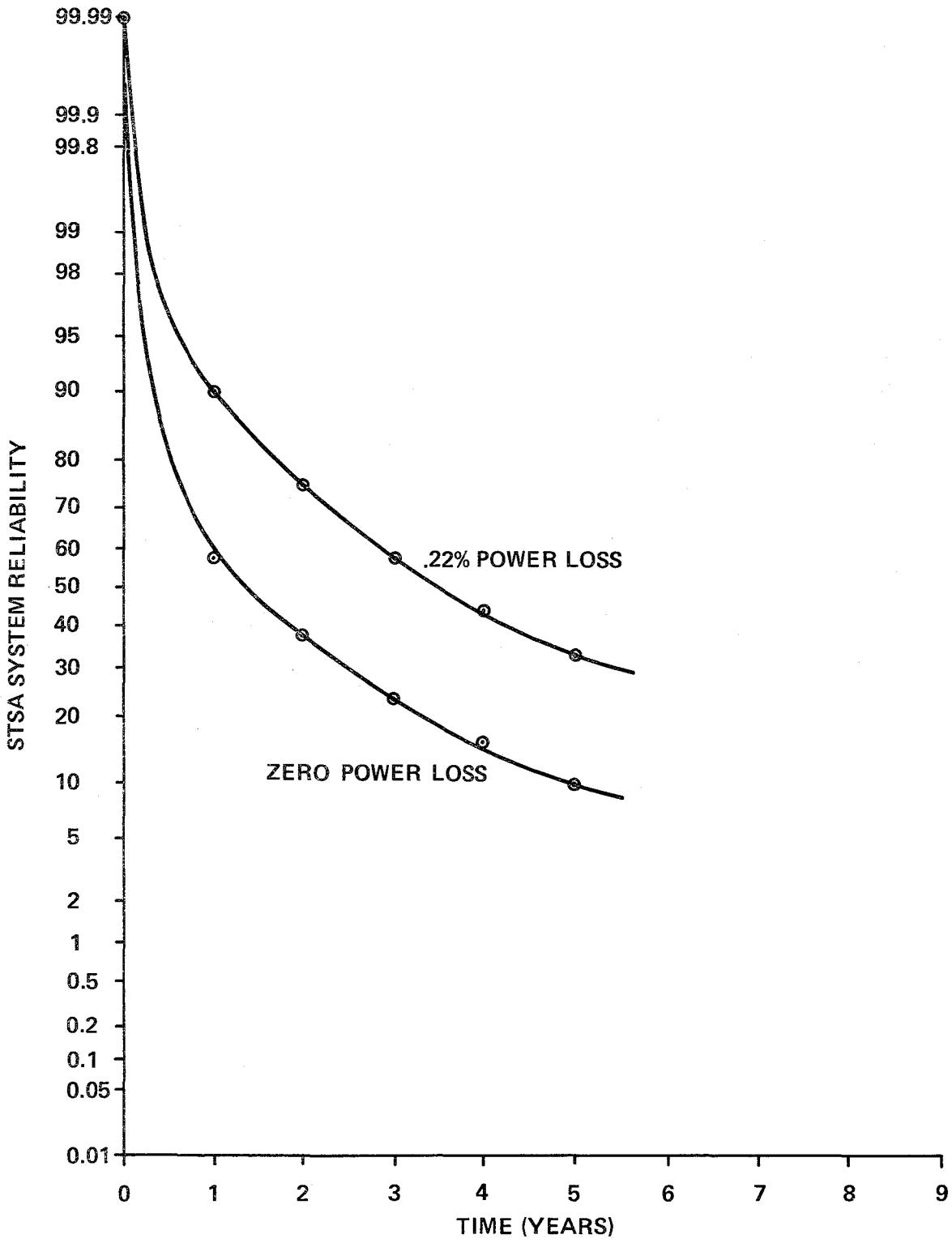
Table V. STSA System Reliabilities

Time t	$R_{S_o}(t)$ (No Power Loss)	$R_{S_1}(t)$ (.22% Power Loss)
0	1	1
1	.57996100	.89739874
2	.37680155	.74795555
3	.24481094	.59405897
4	.15905482	.45682516
5	.10333816	.34324838

The graphs of the reliabilities of STSA System are presented for zero power loss and .22% power loss.

8. Discussion and Conclusion

After a search of literature on the failure rate estimates and reliabilities of solar cells, most of the identifiable sources on the subject were found in the ESA documents. However, none of the documents contained any original data. The failure rate of the solar cells was determined from the failure rate of the solar cell diodes. The failure rate of solar cells is taken as the same as that of the diode (STSA B151) because both have the same base material, doping, contact system, cover slides, and adhesive. The data on failure rate of diodes is also not available. It appears that the failure rate of diode is determined from the estimates of the components in the diode. The failure rate on CIC's is estimated from a confidence procedure on the binomial distribution, which is independent of time.


99.99
99.9
99.8
99
98
95
90
80
70
60
50
40
30
20
10
5
2
1
0.5
0.2
0.1
0.05
0.01
STSA SYSTEM RELIABILITY
.22% POWER LOSS
ZERO POWER LOSS
0
1
2
3
4
5
6
7
8
9
TIME (YEARS)

There are no explicit results available on the effect on the failure rate due to the temperature change of the solar cells, although the cells go through a change of temperature from -80^{o} to 180^{o} centigrade. The effect of temperature on the cell voltage is discussed by Rajeswaren et al. (1982). Also, Anderson and Kim (1978) give the relation between open voltage and temperature. There may not be an immediate impact on the performance of the solar cell, but long term effect is not known. In the (1978) paper Anderson and Kim state that there was a degradation after the crack appeared in the cell. This suggests that the constant failure rate models may not be suitable for systems working for a long time.

Also recently, Alexander (1985) has run some tests on the solar cells. The tests were run on only 12 cells. In the study, he found some breaks in the strands on interconnects. The total number of unbroken strands on the interconnects is not clear. Also, there were not any complete breaks in interconnects. It is difficult to determine failure rate on the interconnects.

Although general models applicable to STSA System are developed, the models are not tested for any other model except the constant failure rate model. There is a need to test aging models. From the available failure rate estimates, the reliability estimate of STSA System is obtained based on the maximum likelihood estimators of λ. The maximum likelihood estimators of the reliability are biased, hence, the biases and variances need to be investigated. And the confidence intervals on the reliabilities need to be developed to gain further information on the reliability of STSA System.

REFERENCES

1. Abernethy, R. B; Breneman, J. E.; Medlin, C. H.; and Reinman, G. L.; Weibull Analysis Handbook (1983), Pratt Whitney Aircraft, Govt. Products Division.

2. Alexander, D., "Space Telescope Solar Cell Module Thermal Cycle Test," (1985), Marshall Space Flight Center NASA Report.

3. Anderson, W. A. and Kim, J. K., "Reliability Studies on MIS Solar Cells," (1978), Applied Physics, Vol. 17, p. 401-404.

4. Asher, H., "Discussion," On Statistical Methods in Reliability, (1983), Technometrics, Vol. 25, p. 320-326.

5. Barlow, R., "Mathematical Theory of Reliability: A Historical Prospective," (1984), IEEE Tran., Vol. R-33, p. 16-20.

6. Basu, A. P. and Mawaziny, A. H. FL, "Estimates of Reliability of K-Out-Of-M Structures in the Independent Exponential Case," (1978), Jour. Amer. Stat. Asso., Vol. 73, p. 850-854.

7. Cohen, A. C., "Maximum Likelihood Estimation in the Weibull Distribution Based on Complete and on Censored Samples," (1965), Technometric, Vol. 7, p. 579-588.

8. Crow, H. L., "Reliability Analysis For Complex Repairable System," (1974), Reliability and Biometry Statistical Analysis of Life Length SIAM Philadelphia.

9. Dumas, L. N. and Shuma, A., "Photovoltaic Module Reliability Improvement Through Application Testing and Failure Analysis," (1982), IEEE Trans., Vol. R-31, p. 228-233.

10. Garlach, "Reliability Prediction For STSA-BSFR Solar Array Blanket," (1984), AN-1367-108 (ESA Document).

11. Garner, P. D., "Solar Array Power Generation Reliability Analysis," (1981), TN-SA-B151 (ESA Document).

12. Hazell, H. H. C., "Solar Array System Description Handbook," GL-SA-B002 (ESA Document).

13. Karlin, S. and Taylor, H. M., A First Course in Stochastic Processes (1975), Acadamic Press, New York.

14. Lauftenburger, H. A. and Anderson, R. T., "Reliability Terminology and Formulae For Photovoltaic Power System," (1982), IEEE Trans., Vol. R-31, p. 289-295.

15. Rajeswaren, G.; Anderson, W. A.; Thayer, M.; and Lee, B. W.; "Statistical Analysis of Cr-Mis Solar Cells," (1982), IEEE Tran., Vol. R-31, p. 276-280.

16. Shooman, M. L., Probabilistic Reliability: An Engineering Approach (1968), McGraw-Hill

17. Stember, L. H.; Huss, W. R.; and Bridgeman, M. S., "A Methodology For Photovoltaic System Reliability and Economic Analysis," (1982) IEEE Tran., Vol. R-31, p. 296-303.

1985

NASA/ASEE SUMMER FACULTY RESEARCH FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER

THE UNIVERSITY OF ALABAMA AT HUNTSVILLE

Prepared By:	L. Michael Santi, Ph.D.
Academic Rank:	Assistant Professor
University and Department:	Memphis State University Mechanical Engineering Department
NASA/MSFC:	
Laboratory:	Systems Dynamics
Division:	Atmospheric Science
Branch:	Computational Fluid Dynamics
MSFC Counterpart:	C. Warren Campbell
Date:	August 9, 1985
Contract No:	NGT 01-008-021 The University of Alabama in Huntsville

TURBULENT FLOW FIELD PREDICTIONS IN SHARPLY CURVED TURN AROUND DUCTS

BY

L. Michael Santi
Assistant Professor of Mechanical Engineering
Memphis State University
Memphis, Tennessee

ABSTRACT

In this investigation, two-dimensional turbulent flow of incompressible Newtonian fluids in sharply curved 180° turn around ducts is studied. Results of an approximate numerical flow field analysis utilizing an orthogonal, body-fitted, curvilinear coordinate system are compared to results based on a traditional cylindrical reference frame. Qualitative indication of general streamfield characteristics as well as quantitative benchmarks for the planning of future experimentation are provided. In addition, preliminary results of an augmented $k-\varepsilon$ turbulence model analysis, which explicitly accounts for the effects of streamline curvature and pressure strain in internal turbulent flows, are presented. Specific model difficulties are discussed and comparisons with standard $k-\varepsilon$ model predictions are included.

LIST OF FIGURES

LIST OF TABLES

NOMEMCLATURE

Symbol	Definition
$C_{\mu 0}$	constant (=0.09) in traditional k-ε model turbulent viscosity expression
C_{μ}	spatially variant function in modified k-ε model turbulent viscosity expression ($\mu_t=C_{\mu}\rho k^2/\varepsilon$)
C_w	constant in wall function correction ($=\kappa/C_{\mu 0}^{3/4}$), see equation (13)
D	duct width
D_{ij}	dissipation term in the Reynolds stress transport equations
E	east side cell, see Figure 12
F	cubic polynomial in $\sqrt{C_{\mu}}$, see equation (12)
k	turbulent kinetic energy
ℓ	turbulence length scale
ℓ_o	turbulent length scale for flow over a plane
N	north side cell, see Figure 12
P	time averaged pressure
P_d	turbulent energy production rate
P_e	exit plane pressure assumed uniform
P_{ij}	production term in the Reynolds stress transport equations
r	radial direction coordinate
r_i	duct inner convex wall radius of curvature
r_o	duct outer concave wall radius of curvature
R_c	duct centerline radius of curvature ($=[r_i+r_o]/2$)
S	south side cell, see Figure 12
S_r, S_θ	momentum source terms, see equations (5) and (6)
$\overline{u_r u_\theta}$, $\overline{u_r^2}$, $\overline{u_\theta^2}$	Reynolds stress components in cylindrical coordinates
U	time averaged streamwise velocity
$\overline{U}$	cross-stream averaged streamwise velocity
U_r, U_θ	time averaged velocity components in cylindrical coordinates
W	west side cell, see Figure 12
y	distance from the duct inner convex surface measured in the radial direction

Symbol	Definition
β	empirical constant in curvature modified length scale, see equation (1)
ε	energy dissipation rate
ε_{ij}	dissipation term in the Reynolds stress transport equations
κ	von Karman constant
μ	absolute viscosity
μ_t	turbulent viscosity
μ_{eff}	effective viscosity ($=\mu_t+\mu$)
Π_{ij}	pressure strain term in the Reynolds stress transport equations
ρ	mass density
θ	angular position measured from the duct entrance plane
BFC	body-fitted coordinates
FDP	fully developed pipe or duct
SSME	space shuttle main engine
TAD	turn around duct

INTRODUCTION

Performance demands of modern machinery require designs exhibiting both dimensional and operational economy. In applications which require the motion of a working fluid, these economic considerations have fueled a proliferation of devices containing complex flow passages through which fluids travel at extreme rates. Examples include flows in turbomachinery, nozzles, diffusers, and curved ducts. Of primary interest in this investigation are flows in sharply curved turn around ducts (TAD's) common in rocket engine design, including current and projected versions of the space shuttle main engine (SSME).

Despite the design impetus provided by increased performance requirements, a detailed understanding of turbulent fluid motion with significant streamline curvature is inhibited by formidable obstacles to both experimental measurement and computational study. Since the system of motion governing equations for turbulent flows is essentially incomplete, advances in the computational prediction of turbulent field characteristics are tied to turbulence models requiring a sound empirical basis. In the case of sharply curved passages and extreme flow rates prevalent in modern equipment, the experimental basis required for detailed flow prediction is difficult to obtain with confidence.

An example of a high rate, strongly curved flow field with little experimental basis on which to develop adequate turbulence models occurs in the SSME fuel side turbopump TAD. A simplified two-dimensional version of this TAD configuration is depicted in Figure 1. Although several studies of turbulent curved duct flow have been reported [1-7], none deals with curvatures or rates of the magnitude occurring in the SSME. The state of turbulence modeling in general is such that little confidence can be placed in extrapolation of the results of these investigations to a broader class of curved flows.

The sensitivity of turbulent flow characteristics to even mild streamline curvatures has been discussed by Bradshaw [8]. In this extensive review, a curvature modified length scale of the form

$$\ell = \left(1 \pm \beta \frac{U_\theta / r}{U_{\theta,r}}\right) \ell_o \qquad (1)$$

was proposed. In this relation β is an empirical constant of order 10, ℓ_o is a length scale appropriate for plane flow, and subscripts following a comma indicate differentiation with respect to the specified variable (e.g., $U_{\theta,r} = \partial U_\theta / \partial r$). Inclusion of this curvature modified scale into a

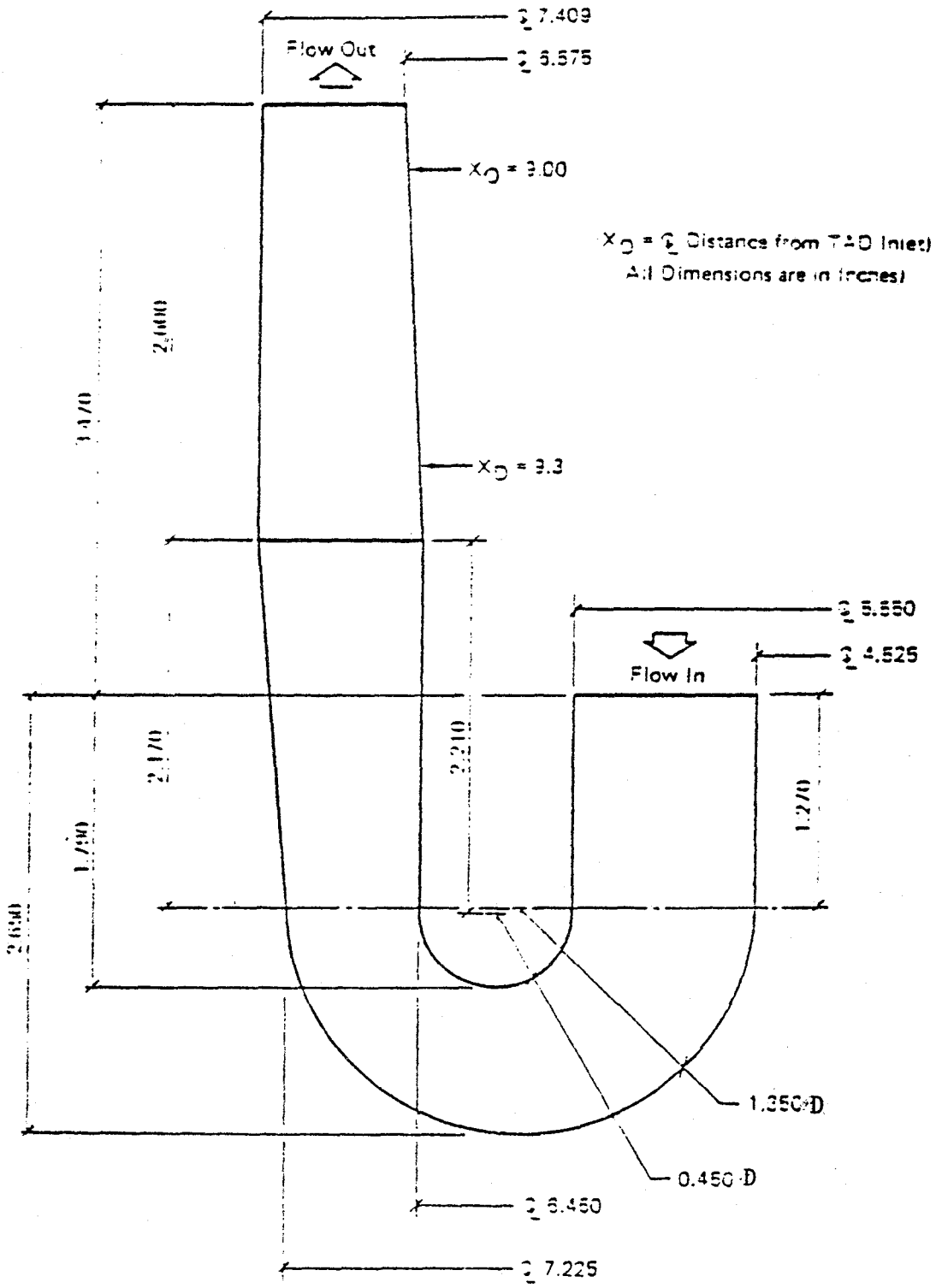


Figure 1. Early SSME fuel side turbopump TAD design.

general mixing length model of turbulent motion is hampered by the flow dependence of both β and ℓ_o. Lack of extensive applicability is characteristic of all mixing length modeling strategies.

Launder et al. [4] have proposed a two equation closure model with the length scale dependence obtained from solution of the equations for turbulent kinetic energy and the energy dissipation rate. The direct effect of curvature is introduced as a single term in the dissipation equation which is proportional to the turbulent Richardson number. The required proportionality constant appears as a new empiricism.

Recently, an extended k-ε turbulence model, which uses algebraic approximations for the Reynolds stress transport equations in order to develop an explicit curvature dependent expression for C_μ in $\nu_t = C_\mu k^2/\varepsilon$, has been reported by Pourahmadi and Humphrey [9,10] and applied to curved duct flows. The new expression for C_μ accounts simultaneously for effects of streamline curvature and pressure strain with wall dampening. The development is a formidable task, calling upon a variety of traditional turbulence modeling approximations [11-15]. The approach taken by Pourahmadi and Humphrey presents a promising explicit curvature model with roots, albeit approximate, in the Reynolds stress transport equations. Unfortunately, previous results based on this computational procedure appear to be based on a rather fundamental algebraic error which will be addressed in subsequent sections of this report.

Another approach which has been pursued is to include curvature effects in a modified wall law [3,16]. This method is well suited for near wall treatment in coarse grid numerical calculations of turbulent boundary layers under adverse pressure gradients. Incorporation into computational procedures involving strongly curved, fully elliptic flow fields is certainly plausible.

Because of the scarcity of reliable experimental data involving high rate flows in sharply curved passages, and the difficulty in extrapolating turbulent model predictions, none of the above described curvature modified turbulence models can be objectively recommended. Each, however, presents an opportunity for computational investigation, refinement, and comparison with results of standard turbulence models as well as experimentation.

OBJECTIVES

The objectives of this investigation have a dual character. On one hand, computational predictions of turbulent flow properties in sharply curved TAD's are sought in order to provide additional information on the operational characteristics of devices incorporating such flows. Of specific interest are TAD flows in the SSME fuel side turbopump. As such, results of this investigation fall into the category of computational analysis. A second view is more appropriate, however, with results of the investigation taken as presenting contributions to the base of computational experimentation needed to improve general models of turbulent flow with significant streamline curvature. The data base building aspect of this study is considered to be the primary goal.

In order to achieve this primary goal, three specific objectives were established.

1. To estimate the general characteristics of turbulent flows in sharply curved TAD's using a standard k-ε turbulence model on an abbreviated geometry as displayed in Figure 2.

2. To implement and test a specific turbulence model which explicitly incorporates the effects of streamline curvature and pressure strain.

3. To examine the elliptic character of the governing motion equations by implementing an orthogonal, body-fitted coordinate reference frame analysis on an extended duct geometry shown in Figure 3.

Background information concerning each of the three stated objectives is provided in the following section of this report.

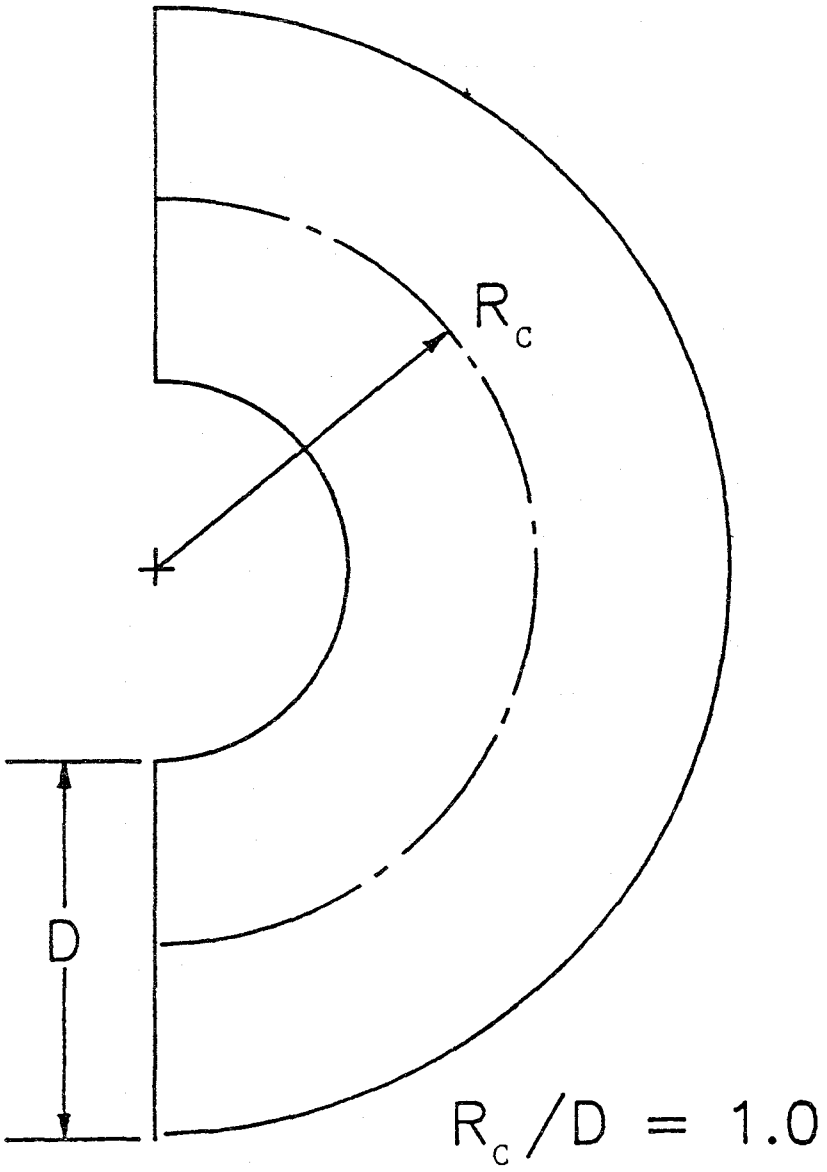


Figure 2. Simple TAD analysis geometry.

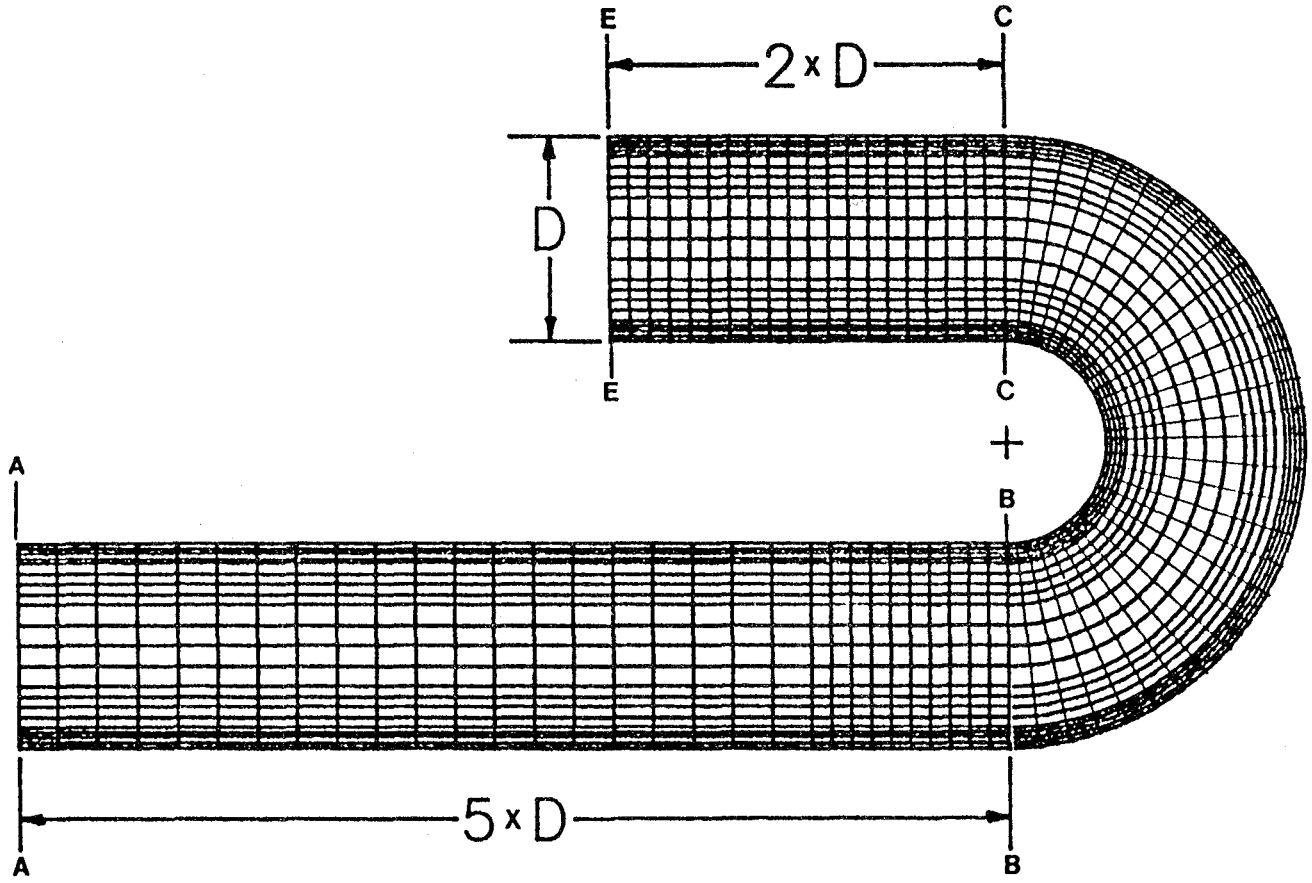


Figure 3. Geometry and grid pattern for BFC TAD analyses: 20 X 80 grid.

BACKGROUND INFORMATION

Equations governing the motion of steady, two-dimensional, turbulent flow of an incompressible, Newtonian fluid in cylindrical coordinates are given below [10].

Continuity:
$$U_{r,r} + U_{\theta,\theta}/r + U_{,r}/r = 0 \tag{2}$$

r-Momentum:
$$\rho[U_{r,r} + U_\theta U_{r,\theta}/r - U_\theta^2/r] = -P_{,r} + (\mu_{eff} r U_{r,r})_{,r}/r + (\mu_{eff} U_{r,\theta}/r)_{,\theta}/r - \mu_{eff} U_r/r^2 - 2\mu_{eff} U_{\theta,\theta}/r^2 + S_r \tag{3}$$

θ-Momentum:
$$\rho[U_r U_{\theta,r} + U_\theta U_{\theta,\theta}/r + U_r U_\theta/r] = -P_{,\theta}/r + (\mu_{eff} r U_{\theta,r})_{,r}/r + (\mu_{eff} U_{\theta,\theta}/r)_{,\theta}/r - \mu_{eff} U_\theta/r^2 + 2\mu_{eff} U_{r,\theta}/r^2 + S_\theta \tag{4}$$

In equations (3) and (4) above, the Boussinesq assumption has been employed to model the Reynolds stresses in terms of mean velocity gradients and a turbulent viscosity component. The momentum source terms S_r and S_θ are given by the relations below.

$$S_r = [\mu_t r(U_\theta/r)_{,r}]_{,\theta}/r + (\mu_t r U_{r,r})_{,r}/r - \mu_t U_r/r^2 \tag{5}$$

$$S_\theta = [\mu_t(2U_r/r + U_{\theta,\theta}/r)]_{,\theta}/r + [\mu_t(U_{r,\theta} - U_\theta)/r]_{,r} + \mu_t(U_{\theta,r} - U_\theta/r)/r \tag{6}$$

In order to provide and solve for spatial variation of the turbulent viscosity μ_t, the k-ε turbulence model as described by Launder and Spalding [17] was employed. This two equation model has been compared favorably to a variety of turbulence modeling schemes [18]. It has been tested on a wide range of flow configurations, exhibiting some degree of universality with an acceptable degree of complexity. Following this approach, μ_t is determined by the relation

$$\mu_t = C_\mu \rho k^2/\varepsilon \ , \tag{7}$$

where the turbulent kinetic energy and dissipation rate are determined upon solving the differential transport equations given below [10,19].

turbulent kinetic energy k:

$$\rho[U_r k_{,r} + U_\theta k_{,\theta}/r] = (\mu_{eff} k_{,r}/\sigma_k)_{,r}/r + (\mu_{eff} k_{,\theta}/\sigma_k)_{,\theta}/r^2 - \rho\varepsilon + P_d \qquad (8)$$

energy dissipation rate ε:

$$\rho[U_r \varepsilon_{,r} + U_\theta \varepsilon_{,\theta}/r] = (\mu_{eff} r\varepsilon_{,r}/\sigma_\varepsilon)_{,r}/r + (\mu_{eff} \varepsilon_{,\theta})_{,\theta}/r^2 + C_{\varepsilon 1}\varepsilon P_d/k - C_{\varepsilon 2}\rho\varepsilon^2/k \qquad (9)$$

The production term P_d in equations (8) and (9) is given by the relation

$$P_d = \mu_t\{2[(U_{r,r})^2 + (U_{\theta,\theta}/r)^2 - U_\theta(U_{r,\theta}/r + U_{\theta,r})/r + U_r(U_r + 2U_{\theta,\theta})/r^2 + U_{r,\theta}U_{\theta,r}/r] + (U_\theta/r)^2 + (U_{\theta,r})^2 + (U_{r,\theta}/r)^2\} . \qquad (10)$$

The numerical procedure used to solve the discretized form of equations (2), (3), (4), (8) and (9) was provided by the proprietary PHOENICS computer code supplied and supported by CHAM of North America. A discussion of the philosophical and mathematical basis of this code is provided in reference [19]. A good presentation of the numerical procedures associated with a large family of CFD codes including the PHOENICS code is available in the text by Patanker [20].

The standard k-ε turbulence model, the basis of which is presented above for cylindrical coordinate systems, was employed in computational analyses of simple geometry (Figure 2) TAD flows. A 20 x 30 grid was employed in these analyses as displayed between planes BB and CC of Figure 3. In all simple TAD configuration analyses, inlet plane velocity, turbulent kinetic energy, and dissipation rate conditions were specified. A constant exit plane (CC in Figure 3) pressure was prescribed and boundary condition closure was achieved by assuming zero velocity field gradients in the exit plane.

The Pourahmadi and Humphrey [9,10] extension of the standard model was also implemented. As discussed above, this modified turbulence model incorporates the effects of streamline curvature and pressure strain in the fluid by using a variety of approximations to reduce the Reynolds stress transport equations to a set of algebraic equations. Because of the complexity of the algebraic manipulations employed, only a brief outline of the model derivation and implementation is presented below. The reader is referred to the original work in reference [9] for a detailed development.

The derivation starts from a high Reynolds number form of the $\overline{u_i u_j}$ transport equation [9] characterized by the general form shown below.

$$\frac{D\,\overline{u_i u_j}}{Dt} = P_{ij} + \varepsilon_{ij} + \Pi_{ij} + D_{ij} \tag{11}$$

The development then proceeds in the following stepwise manner.

1) The $\overline{u_i u_j}$ transport equations are reduced to a system of algebraic equations by invoking a variety of approximations including
   A) Rotta's [12] return to isotropy principle in Π_{ij}
   B) Launder, Reece and Rodi's [13] expression for the contribution to Π_{ij} from interaction between mean strain and fluctuating velocities
   C) Daly and Harlow's [14] correction to Π_{ij} due to wall effects
   D) Gibson and Launder's [15] correction to Π_{ij} due to wall effects
   E) Rodi's [11] assumption that $\overline{u_i u_j}/k$ is constant throughout the flow field.
2) The resulting system of algebraic equations is solved for the turbulent stresses.
3) The Boussinesq approximation is then employed, transforming the solution into an algebraic relation for C_μ.
4) The turbulent energy production term is rearranged into the form
   $$(k/\varepsilon)U_{\theta,r} = g(P_d/\varepsilon,\ U_\theta/r,\ U_{i,j})/\sqrt{C_\mu}$$
   where g is a complex function of the quantities indicated, including the various partials of the mean velocity, $U_{i,j}$.
5) The relation for $(k/\varepsilon)U_{\theta,r}$ from step 4 is substituted into the algebraic relation for C_μ from step 3. Simplification of the resulting expression leads to the cubic polynomial relation
   $$F(\sqrt{C_\mu}) = C_\mu^{3/2} + a_1 C_\mu + a_2 C_\mu^{1/2} + a_3 = 0 \tag{12}$$
   where $a_1, a_2, a_3 = a_1, a_2, a_3(P_d/\varepsilon,\ U_\theta/r,\ f,\ U_{i,j})$
6) The wall correction function f presented in reference [13] and appearing in step 5 is modified to the form
   $$f = \frac{k^{3/2}}{C_w \varepsilon}\left[\frac{1}{y} + \frac{(y/D)^m}{D-y}\right] \tag{13}$$

thereby permitting asymmetric wall corrections to the pressure strain. In this expression, the asymmetry measure m appears as a new empirical constant.

7) Following Bradshaw [8], the expression for dissipation at near wall nodes is modified by a multiplicative factor of $1/[1\pm\beta U_\theta/(U_{\theta,r}r)]$ in order to introduce the influence of extra strain curvature on the magnitude of the near wall turbulence length scale. This expression also contains a new empirical constant, β.

Suggested values for constants m and β are 7.95 and the order of 10 respectively.

Although the complex development terminating at equation (12) is quite impressive, an unfortunate algebraic error is made in establishing the roots of this equation. Instead of a single positive real root for $\sqrt{C_\mu}$ as predicted in references [9] and [10], multiple real roots, not necessarily positive, can (and did) occur in practice. The implemented version of the Pourahmadi and Humphrey modification included a corrected method of solving for the curvature dependent values of $\sqrt{C_\mu}$. Adding to the difficulties associated with this implementation was the need to prescribe a selection criterion for regions in which multiple roots occur. This problem was never resolved with complete success as will be discussed more fully in the next section of this report.

In order to more fully understand the elliptic character of developing TAD flows, analyses utilizing an orthogonal body-fitted coordinate (BFC) reference system on an expanded duct geometry were performed. Two such analyses were performed; one on the full geometry, 20 x 80 grid system depicted in Figure 3, and a second on a reduced geometry 20 x 60 grid depicted between planes AA and CC of the same figure. In each case the 20 x 30 grid configuration of the simple TAD analyses was preserved in the region bounded by surfaces BB and CC in Figure 3. In this manner comparison of results using both simple TAD configurations and expanded BFC geometries was facilitated.

Numerical computations employing the grid configuration of Figure 3 were performed utilizing the BFC option of the PHOENICS computer code described in reference [21]. Although potentially more accurate non-orthogonal grid specifications could have been implemented within the framework of the PHOENICS code, direct comparison with results of simple cylindrical system analyses would have then become more difficult. An excellent comprehensive review of methods to numerically generate curvilinear coordinate systems with coordinate lines coincident with boundary segments is given in reference [22].

RESULTS

Results of this investigation can be separated into three categories dependent upon coordinate reference frame and turbulence model employed. Analyses 1, 2, and 6 described in Table I were performed on simple TAD geometries using a cylindrical coordinate system and a standard k-ε turbulence model [17]. In analysis 3 the modified k-ε model of Pourahmadi and Humphrey [9,10], with the correction discussed in the previous section, was employed. The simple TAD geometry displayed in Figure 2 was maintained. Analyses 4 and 5 employed orthogonal BFC reference systems with a standard k-ε turbulence model. The grid system depicted in Figure 3 was used in analysis 4. The same grid pattern, restricted to the corresponding analysis geometry, was employed in all cases. In addition, the flow Reynolds number was fixed at 10^6 and a curvature ratio $R_c/D = 1$ was prescribed. These values obviously refer to a high rate flow with extremely sharp curvature. A complete summary of the computational analyses of this study is presented in Table I.

In Figures 4, 5, and 6 are displayed developing profiles of velocity, pressure, and turbulent kinetic energy respectively, for the case of a simple TAD geometry with inlet plug flow. In response to the strong cross stream pressure gradients evident throughout the flow domain, the velocity profiles are observed to become increasingly asymmetric. The most substantial pressure adjustments occur within the final 30° of duct travel, and in response there is a significant acceleration of flow near the outer wall with a corresponding deceleration near the inner, convex, duct wall. The net effect is a nearly linear velocity profile very near the exit plane reminiscent of a solid body rotation. Quantitatively, the near exit plane velocity varies from just under 25 % to somewhat over 140 % of the average flow velocity. In the entrance region of the duct, the plug flow characteristic is observed to persist virtually unchanged past the midplane ($\theta = 90°$) of the duct. The large cross stream pressure gradient predicted for the TAD inlet region together with the specified plug flow entrance velocity profile is suggestive of a very large swirl chamber feed to the TAD. This type of configuration is of course not prevalent in rocket engine or turbomachinery applications.

The turbulent kinetic energy profiles displayed in Figure 6 exhibit rapid decay in the plug flow core region as expected. An initial decay in the near wall turbulent kinetic energy is followed by a significant rise in the accelerating concave wall region, and a continuing gradual decay in the decelerating convex wall region. Near the exit plane, the turbulent energy profile develops the knee observed [1,2] and predicted [10] in previous curved duct investigations conducted at smaller values of Re and R_c/D.

Table I

Analyses Summary

Analysis No.	Reference Frame	Figure 2 I/O Plane	Grid	Inlet U Profile	Turbulence Model
1	Cyl.	BB/CC	20 x 30	Plug	k-ε
2	Cyl.	BB/CC	20 x 30	FDP	k-ε
3	Cyl.	BB/CC	20 x 30	Plug	Modified k-ε
4	BFC	AA/EE	20 x 80	Plug	k-ε
5	BFC	AA/CC	20 x 60	Plug	k-ε
6	Cyl.	BB/CC	20 x 30	Plane BB from 5	k-ε

Analyses Constants: $Re = 10^6$ $R_c/D = 1.0$

Turbulence Inlet Conditions For Plug Inlet U Profile :

$k/\bar{U}^2 = 0.005$

$\varepsilon/(\bar{U}^3/D) = (0.005)^{3/2}/0.01$

Exit Plane Pressure: Uniform across exit plane

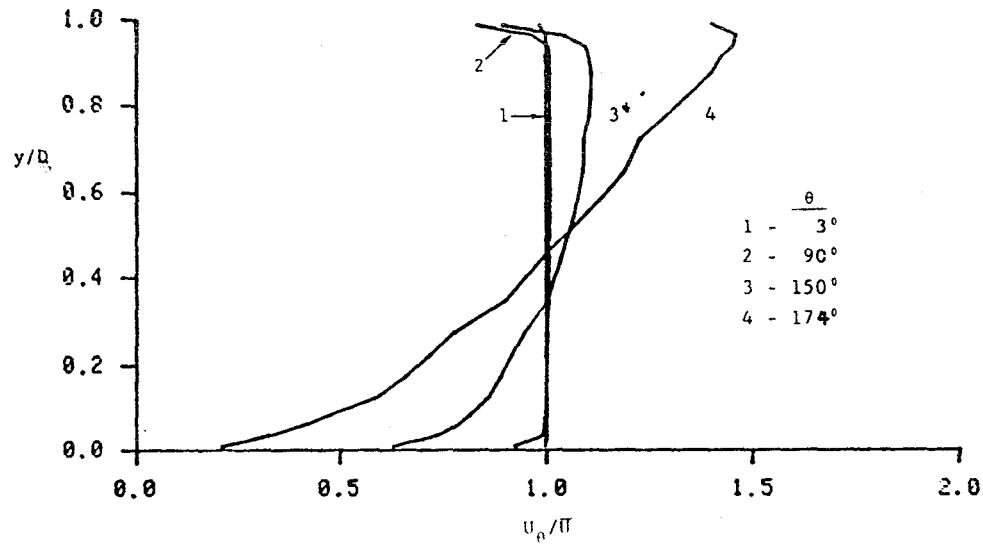


Figure 4. Streamwise velocity profiles for simple TAD geometry with plug flow inlet: Analysis 1.

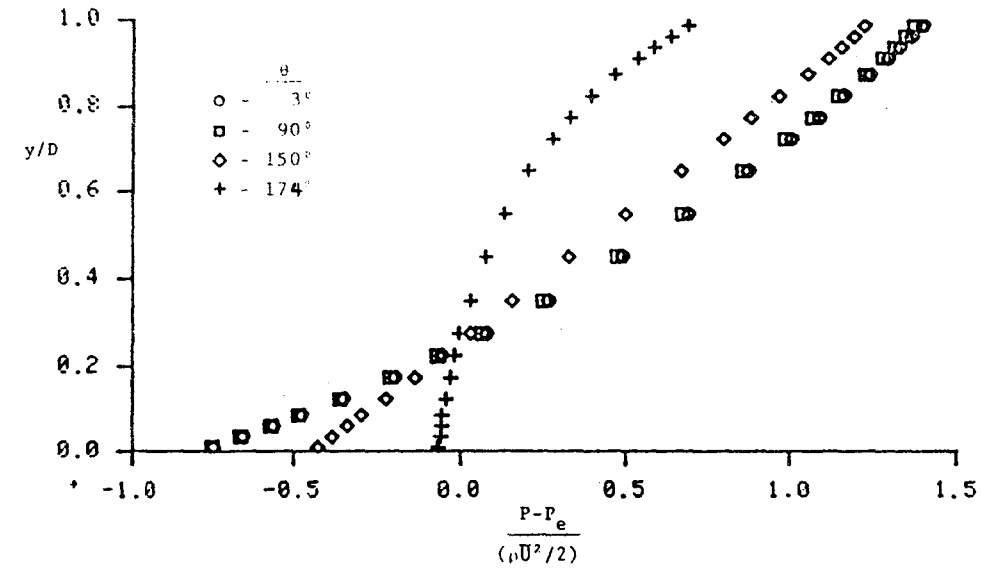


Figure 5. Pressure profiles for simple TAD geometry with plug flow inlet: Analysis 1.

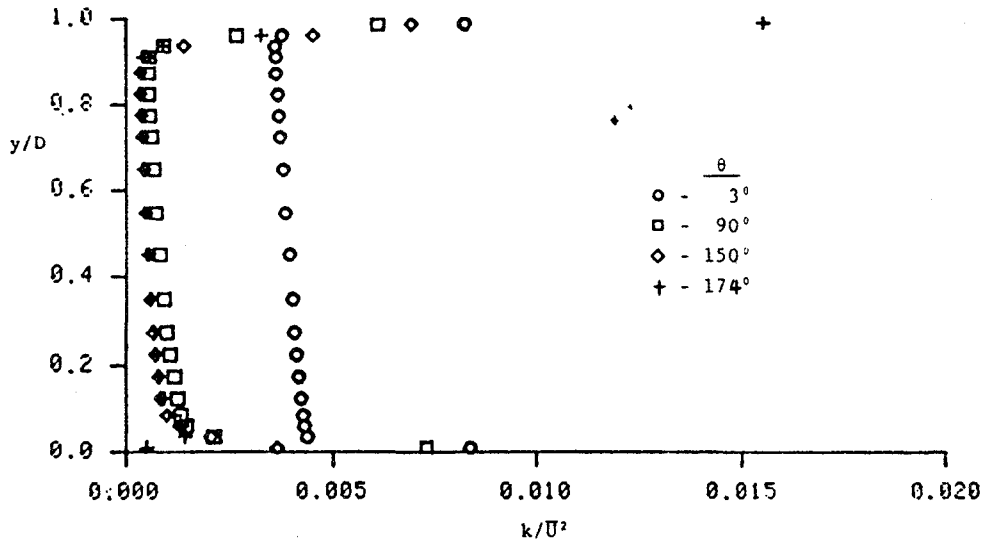


Figure 6. Turbulent kinetic energy profiles for simple TAD geometry with plug flow inlet: Analysis 1.

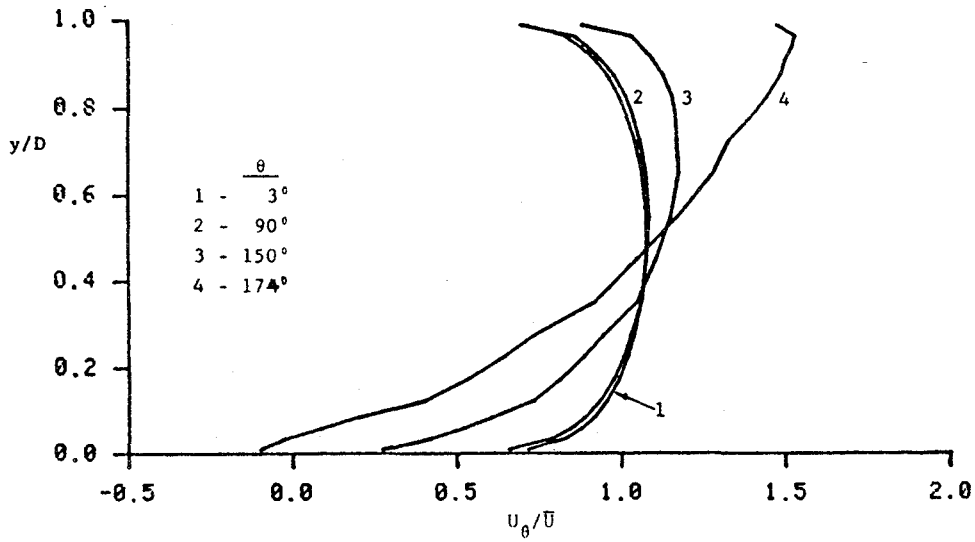


Figure 7. Streamwise velocity profiles for simple TAD geometry with FDP flow inlet: Analysis 2.

Results of a simple TAD geometry analysis with a fully developed turbulent duct flow inlet condition are displayed in Figures 7, 8, and 9. The general shape characteristics and scale of the pressure profiles in Figure 8 are similar to the previous plug flow inlet case. The development sequence for the velocity profiles is also similar to the analysis 1 case with little profile adjustment prior to the 90° midplane of the TAD, and significant velocity gradient increases and profile linearization in the rapid adjustment exit region. The salient feature displayed in Figure 7 is the appearance of a recirculation region near the inner wall of the duct. This region extends back somewhat over 6° from the exit plane and occupies approximately 4 % of the duct cross section at the exit plane. The occurrence of a negative velocity region in the exit plane presents certain difficulties involving computational error as discussed in reference [21]. Despite this problem, flow separation from the inner convex wall is predicted to occur between $\theta = 171°$ and $\theta = 174°$. This phenomena is not observed to occur in less sharply curved channels with smaller values of the flow Reynolds number [1,2,6].

The turbulent kinetic energy profiles depicted in Figure 9 differ substantially from the plug flow inlet case both in shape and development history. The initial profile shape is symmetric and of the same magnitude as displayed for the analysis 1 entrance profile. Because of the entrance cross stream velocity gradients present in the FDP entrance case, however, the turbulent kinetic energy is predicted to increase significantly in the flow direction away from the convex wall. This is in sharp contrast to the rapid post entrance decay of k exhibited in Figure 6 for the plug flow inlet case. Near the exit plane, cross stream velocity gradients and the energy dissipation rate away from the convex wall increase rapidly in the flow direction, and the increase in turbulent energy is arrested. In the vicinity of the convex wall there is a substantial adverse pressure gradient leading to deceleration and, as the exit plane is approached, flow separation. This is accompanied by a sharp decrease in the inner wall region values of k.

Direct comparison of the downstream velocity and pressure profiles predicted in analyses 1 and 2 are displayed in Figures 10 and 11. The similarities in each of the profile pairs is obvious. The near exit plane pressure adjustment is predicted to occur more suddenly in the case of plug flow inlet conditions. It is evident that the inner wall characteristics of the near exit plane flow are substantially affected by changes in the inflow boundary conditions.

In the third analysis of the computational sequence, the modified $k-\varepsilon$ model of Pourahmadi and Humphrey [9,10] was implemented in the user definition section of the PHOENICS routine. Severe difficulties were initially encountered in attempts to obtain reasonable solutions. Further investigation revealed the equation (12) rooting error described previously.

In the context of the PHOENICS implementation, the Pourahmadi and

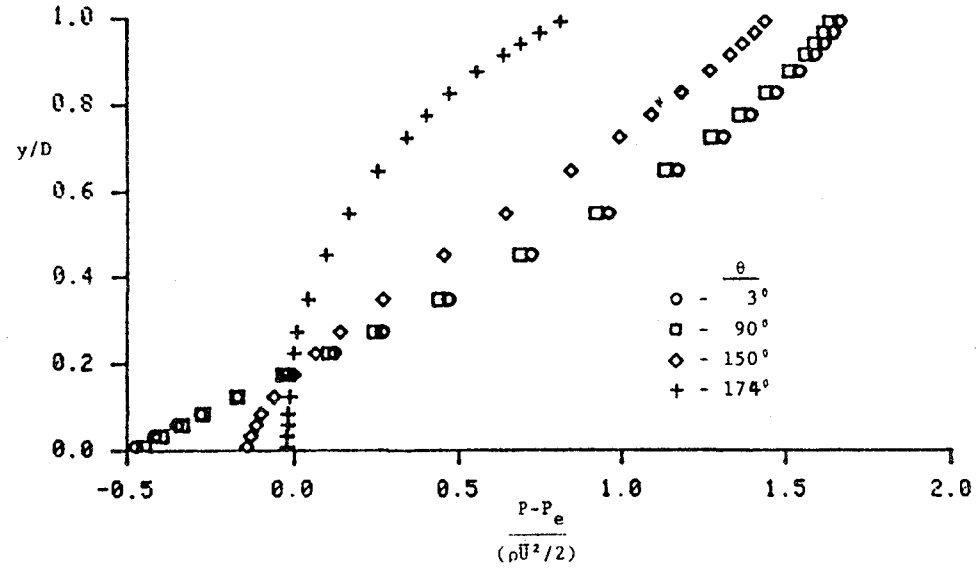


Figure 8. Pressure profiles for simple TAD geometry with FDP flow inlet: Analysis 2.

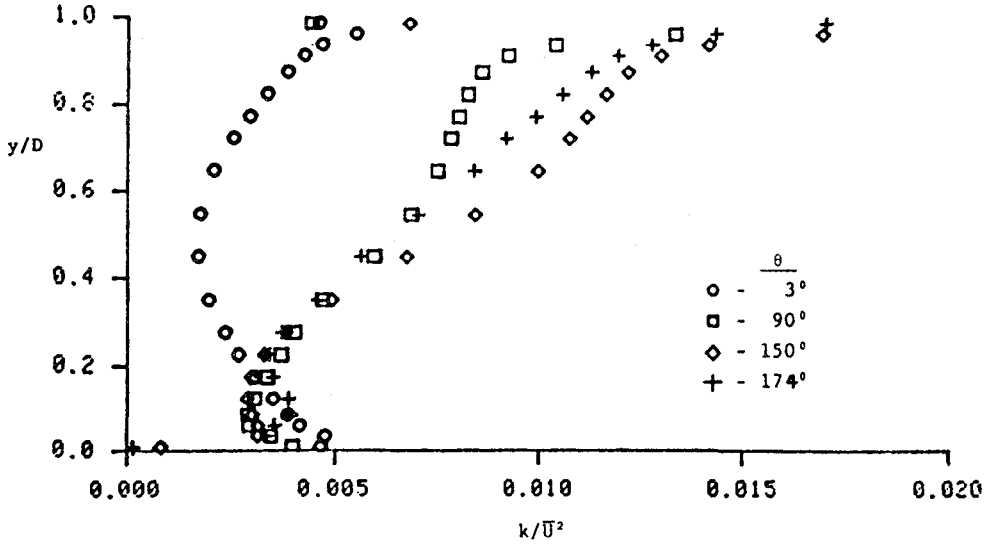


Figure 9. Turbulent kinetic energy profiles for simple TAD geometry with FDP flow inlet: Analysis 2.

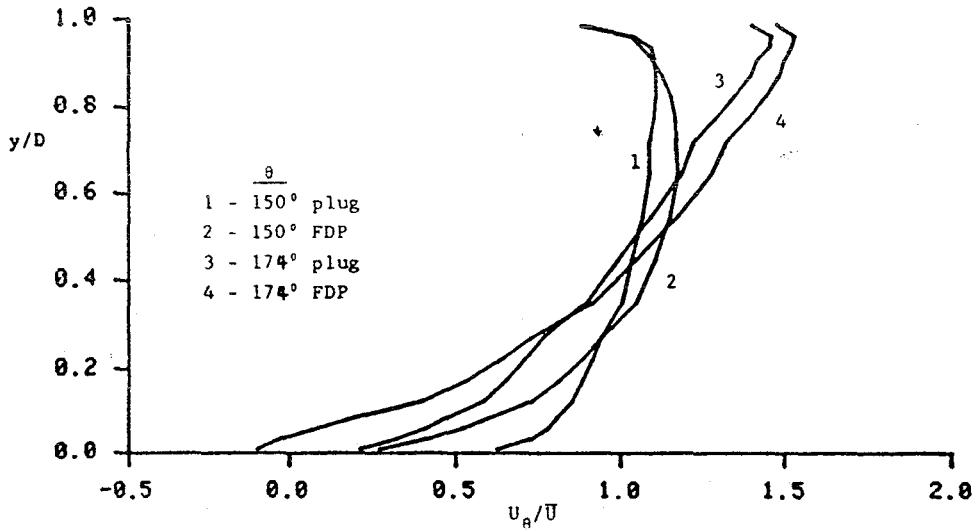


Figure 10. Streamwise velocity profile comparisons for simple TAD geometry.

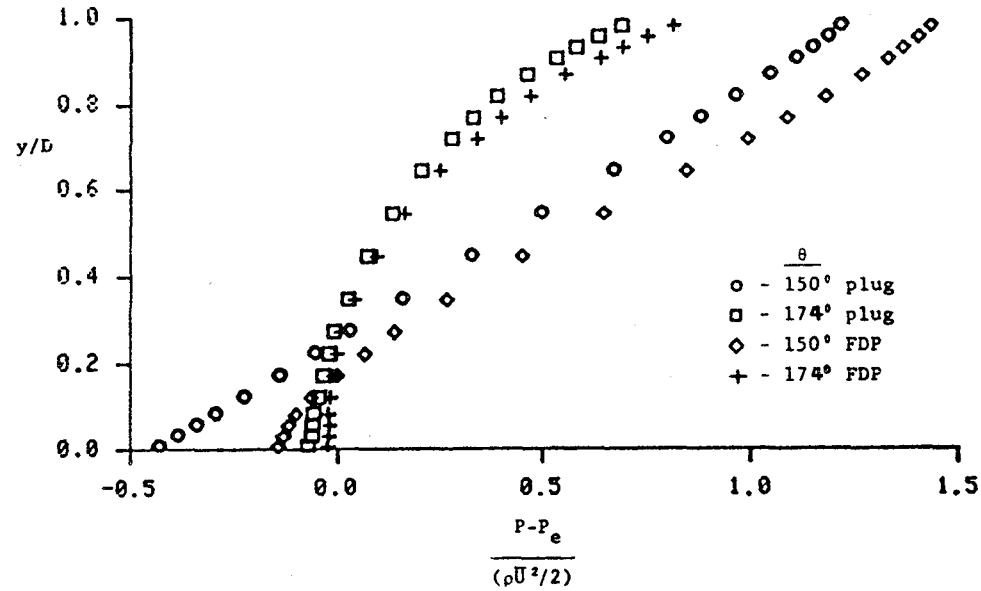


Figure 11. Pressure profile comparisons for simple TAD geometry.

Humphrey [8,10] modification for a specific finite difference cell P requires manipulation of the staggered grid velocity vectors indicated in Figure 12. In addition, grid distances from each of the indicated neighboring cells to the P cell boundaries are needed together with estimates of k and ε at P. The entire manipulation is computationally intense, often tripling the clock time of analysis.

Despite the increased computational effort, convergence difficulties were repeatedly encountered. The source of the problem was the behavior of the polynomial $F(\sqrt{C_\mu})$, the roots of which indicate candidate values of C_μ. Typical $F(\sqrt{C_\mu})$ loci are depicted in Figure 13. These polynomial traces were arrived at by computationally testing the modified k-ε turbulence model on a simple TAD geometry, plug inlet flow problem. Analysis 3 as described in Table I was initiated at the standard k-ε model solution value of analysis 1. An intermediate flow field result was determined by arbitrarily selecting the principal theta transformation root of the cubic polynomial $F(\sqrt{C_\mu})$ whenever multiple roots occurred. The trigonometric or theta transformation method for determining roots of a cubic polynomial is described in most mathmatical handbooks (see e.g. [23]).

It is evident from Figure 13 that for $\theta = 168°$, the near concave wall behavior of F was that of a cubic polynomial with multiple real roots. This behavior was not isolated, occurring irregularly throughout the flow field and frequently in near wall regions. Unfortunately it is in these near wall regions that the modified k-ε model was expected to yield its most meaningful results [9,10]. The behavior of the $F(\sqrt{C_\mu})$ polynomial at y/D = 0.9875 (the node closest to the concave wall) was particularly disturbing. As indicated in Figure 13, the trace of F was virtually colinear with the C_μ axis for $-0.2<C_\mu<0.2$. Three distinct real roots were determined at this nodal position, however, the ill natured character of F is apparent, making the computational process extremely susceptible to even small numerical errors.

Near exit plane values of C_μ were determined as shown in Figure 14. It should be noted that the line of $C_\mu = 0.09$, the standard k-ε model value, occurring at $\theta = 150°$ was an assigned default value selected whenever the cross stream velocity gradient became small. Small values of $U_{\theta,r}$ led to extreme variation of the $F(\sqrt{C_\mu})$ polynomial, hence the default selection in these regions. Pourahmadi and Humphrey [9,10] also note the ill behavior of C_μ in the core region of the duct. They attribute this behavior to a breakdown in the assumption that $u_i u_j/k$ is constant in this region. Their reasoning is less than compelling. It is entirely possible that the problem of determining reasonable values for C_μ is completely numerical, and that consideration of a more primitive implicit form of the C_μ relation might yield stable results across the entire duct cross section.

Recognizing that the analysis 3 results are only intermediate, lacking a completely acceptable method of selecting C_μ, it is useful and

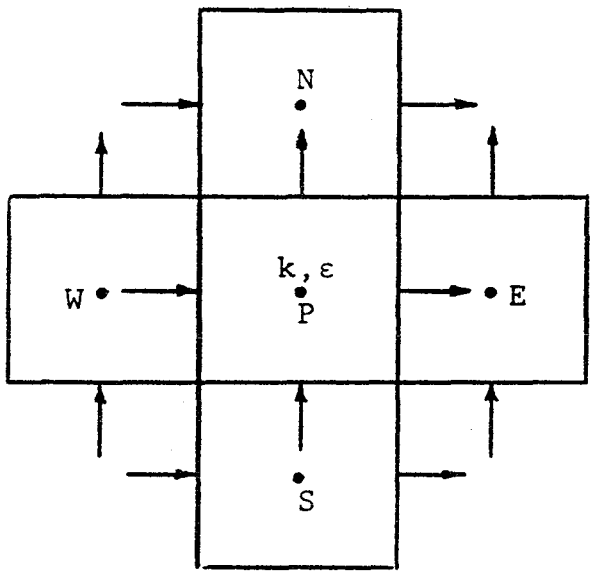


Figure 12. Pourahmadi and Humphrey model 9,10 staggered grid basis cell.

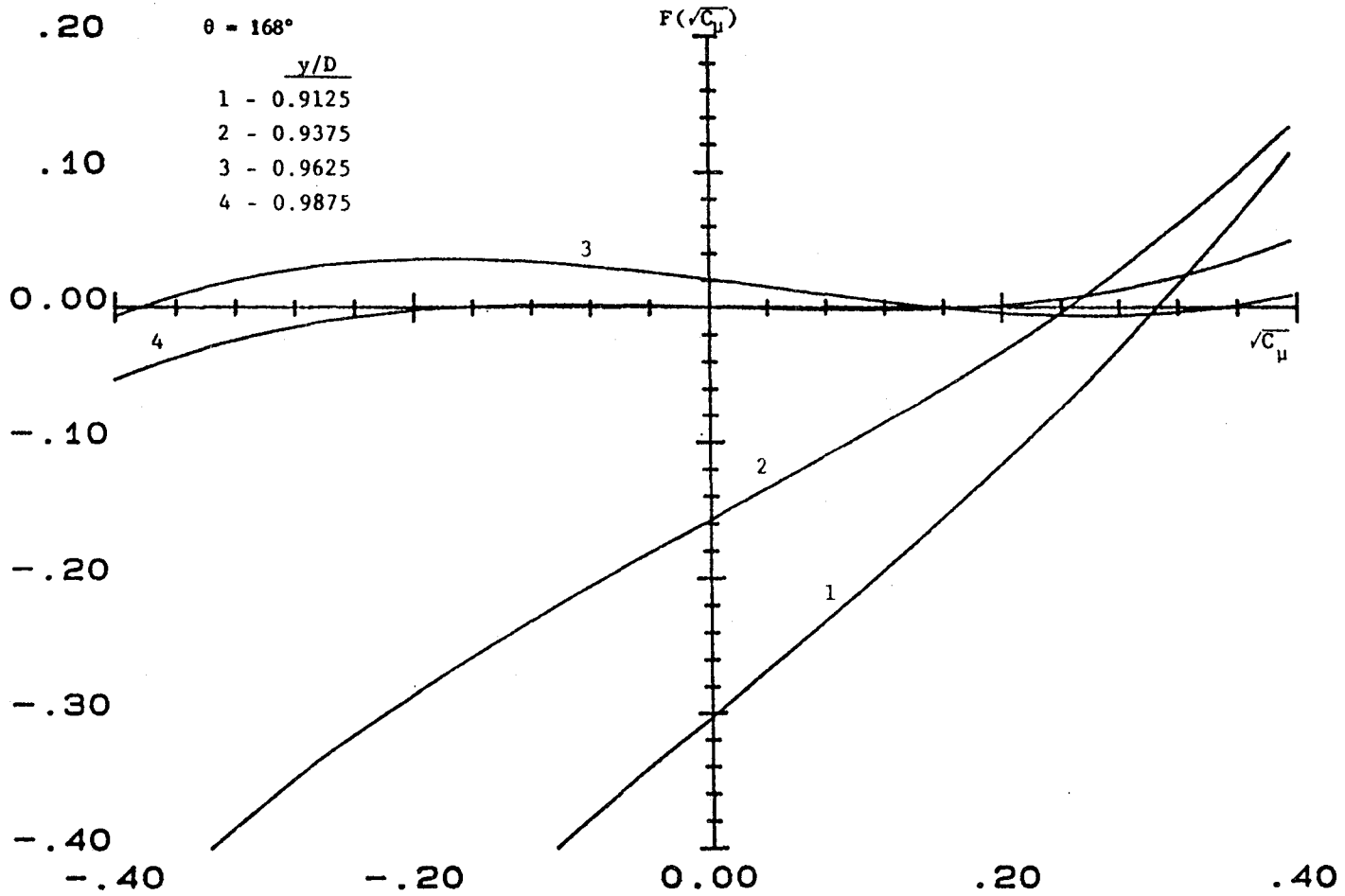


Figure 13. Behavior of $F(\sqrt{C_\mu})$ polynomial at various cross stream locations for θ=168°: Analysis 3.

interesting to observe position dependent characteristics of the variant C_μ . In Figure 14, for $\theta = 174°$, C_μ is observed to take on values substantially less than the standard value of 0.09 in near wall regions, rising to nearly 0.09 just outside the convex wall region and then falling to near zero in the flow core, rising again to near the standard value and well beyond just outside the concave wall region. Figure 15 displays the near concave wall behavior of C_μ throughout the flow domain. It is evident that the near wall node value of C_μ remains essentially constant at around 0.03 throughout while the adjacent node value of C_μ took on default values near the TAD entrance, rising dramatically to a value of nearly 0.4, and then falling steadily to a value in excess of 0.1 before rising dramatically near the duct exit. The turbulent kinetic energy profiles displayed in Figure 16 exhibit the same qualitative character as the unmodified k-ε model profiles of Figure 6 except that near outer wall k values are greatly magnified and exhibit a knee as the wall is approached. Details of the near convex wall comparisons of turbulent kinetic energy are shown in Figure 17. Both the standard and modified k-ε models predict the same k values at the adjacent wall node, however, modified model kinetic energy predictions were somewhat smaller than standard at the neighboring y node near the duct exit. Estimated velocity and pressure profiles obtained using the modified k-ε model were virtually identical to standard model results.

Two separate BFC computational analyses were performed in order to explore the elliptic character of the TAD flow field. These are described in Table I, analyses 4 and 5. Before describing results of these studies, several general observations are warranted. First, the early version of the SSME TAD geometry depicted in Figure 1 suggests a computational model with a short entrance region and a relatively long straight exit region. In Figure 3 this would indicate flow entrance at plane EE and exit at plane AA. Several attempts at such an analysis using the BFC option of the PHOENICS computer code were unsuccessful. This does not necessarily indicate a shortcoming in the PHOENICS/BFC solution technique. It is likely that more extensive experience with judicious selection of the relaxation factors and other heuristic tuning parameters inherent in the PHOENICS code, and indeed any CFD code with like capabilities, would have led to solution convergence with the flow orientation described above. Because of the limited time available to gain computational experience, and because it was found that extension of the straight duct inlet zone greatly enhanced convergence characteristics, the suggested SSME TAD flow orientation was reversed, with the inlet prescribed at plane AA and exit at plane EE.

A second observation concerning the BFC computations involves the degree of solution convergence at analysis termination. Using the absolute residuals sum criterion described more fully in the PHOENICS documentation, the convergence rate at termination had slowed to approximately 0.1 % of total absolute residuals sum per solution sweep. This rate was achieved after approximately 30 clock hours of computation on the Perkin-Elmer 3220 computer system utilized during this study. Although clock run time is a poor absolute measure of computational

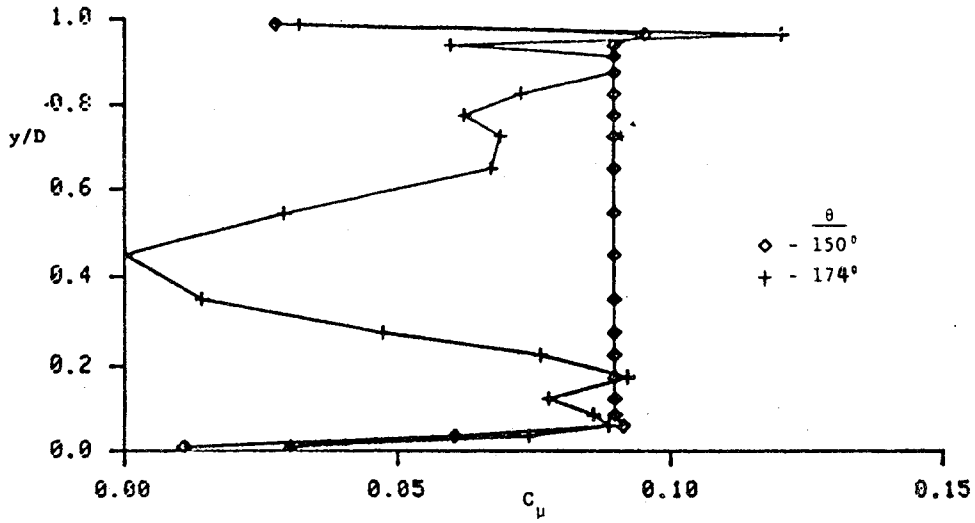


Figure 14. Characteristic C_μ profiles for simple TAD geometry with plug flow inlet condition using the modified k-ε model: Analysis 3.

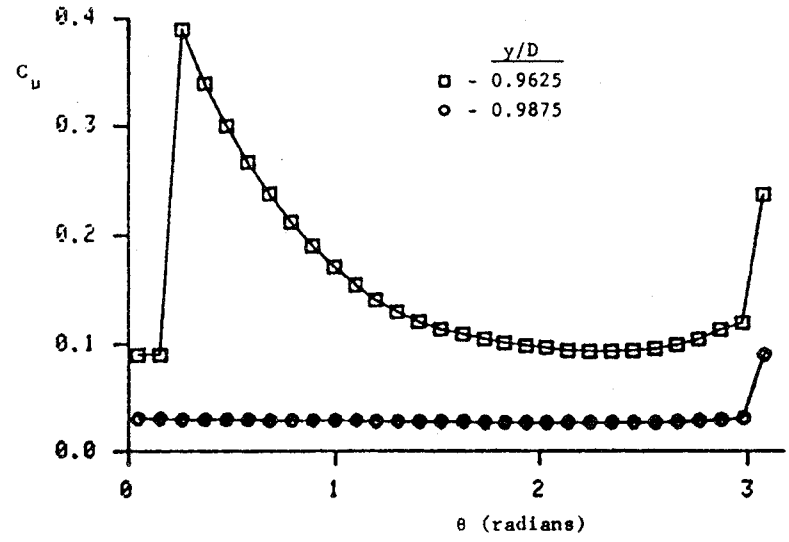


Figure 15. Streamwise variation of C_μ close to the concave wall: Analysis 3.

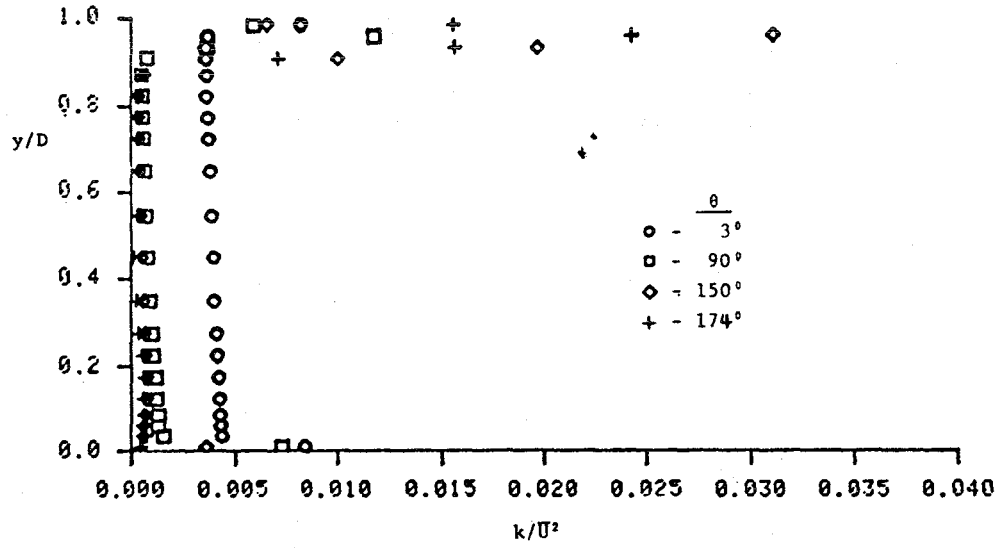


Figure 16. Turbulent kinetic energy profiles for simple TAD geometry using the modified k-ε model: Analysis 3.

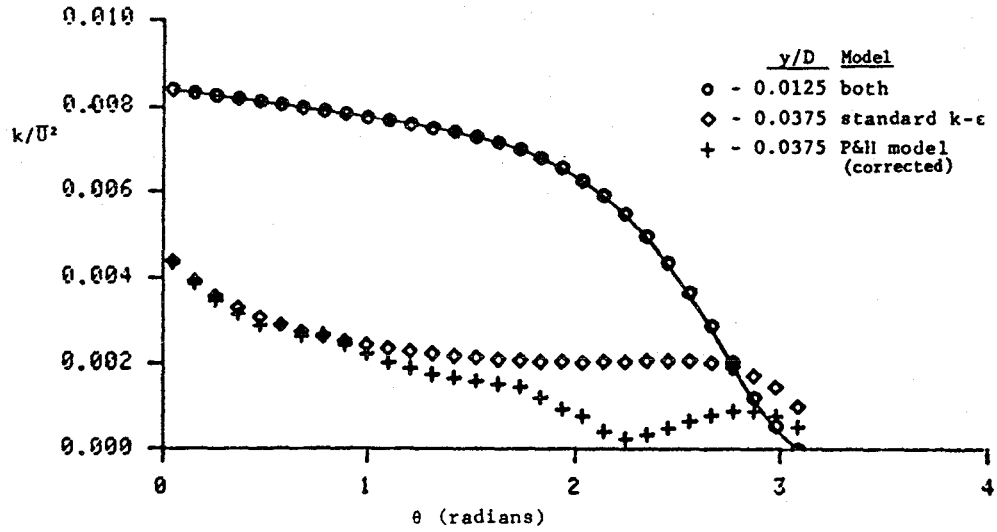


Figure 17. Comparison of near wall turbulent kinetic energy predictions.

efficiency, relative to simple TAD configurations, the BFC analyses required an order of magnitude more clock run time and at best achieved an absolute residuals sum convergence measure a full order of magnitude larger than the simple TAD analyses. Noting that the BFC grid patterns were approximately three times the size of the simple cylindrical system TAD grids, it is obvious that the BFC analyses were relatively high overhead computations. In addition, because calculation was interrupted before achieving a proportionately small convergence measure, results of the BFC analyses are considered to yield only qualitative predictions of flow field characteristics.

Figures 18, 19, and 20 display results of the analysis 4 BFC computations for TAD inlet/outlet velocity, pressure, and turbulent kinetic energy profiles respectively. These calculations were performed using the full grid pattern displayed in Figure 3 with flow inlet at plane AA. It is evident in Figures 18 and 19 that BFC calculated TAD inlet flow field features are radically different from the inlet boundary assumptions of the simple duct analyses. The inlet velocity profile resembles the 1/r proportionality of ideal fluid flow except in the near wall regions while the outlet profile appears very much like a fully developed turbulent duct flow. Obviously the antisymmetric character of the predicted TAD outlet pressure profile is quite different from the assumed uniform exit plane distribution of the simple geometry analyses. Because of the centrifugally induced cross-stream pressure gradient at the TAD exit, the duct like flow profile at plane CC developed into the stretched, r proportional profile reminiscent of rigid body rotation in the 2 x D straight exit duct. Predicted developing velocity profiles in the exit duct are displayed in Figure 21. In the exit section, the pressure adjusts to the assumed uniform condition at exit plane EE. The core of the TAD exit plane k profile shown in Figure 20 resembles the core suppressed profile of the simple geometry analysis with plug flow inlet conditions displayed in Figure 6. No inner wall knee is observed in the BFC k predictions, however, outer wall turbulent kinetic energies are reduced in the BFC analysis.

In order to reinforce the extended geometry predictions of analysis 4, a comparison test was devised. BFC analysis 5 was run on the reduced configuration between planes AA and CC. Plug flow inlet conditions were prescribed at plane AA and uniform exit plane pressure was assumed. Results of this computational procedure at the TAD entrance plane BB were used as boundary conditions for a simple geometry analysis using a cylindrical system with the same uniform exit plane pressure assumed. Predictions of analysis 5 and 6 should have been identical, at least theoretically. Unfortunately, there was a substantial discrepancy in the predicted flow field characteristics as demonstrated in Figures 22 and 23. Even considering the qualitative nature of the BFC analysis, the comparisons were disturbing. Cylindrical system predictions of cross-stream velocity gradients at the exit plane were generally much larger than results using BFC analysis as is evident in Figure 22. In addition, a rather large recirculation zone was predicted using the traditional cylindrical coordinate system analysis. Inner wall flow

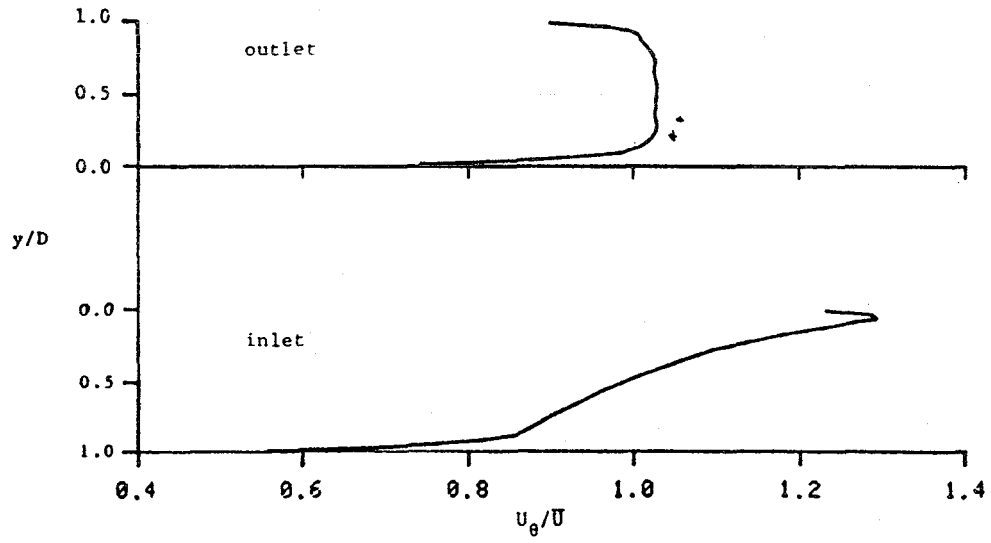


Figure 18. TAD I/O velocity profiles for BFC computation: Analysis 4.

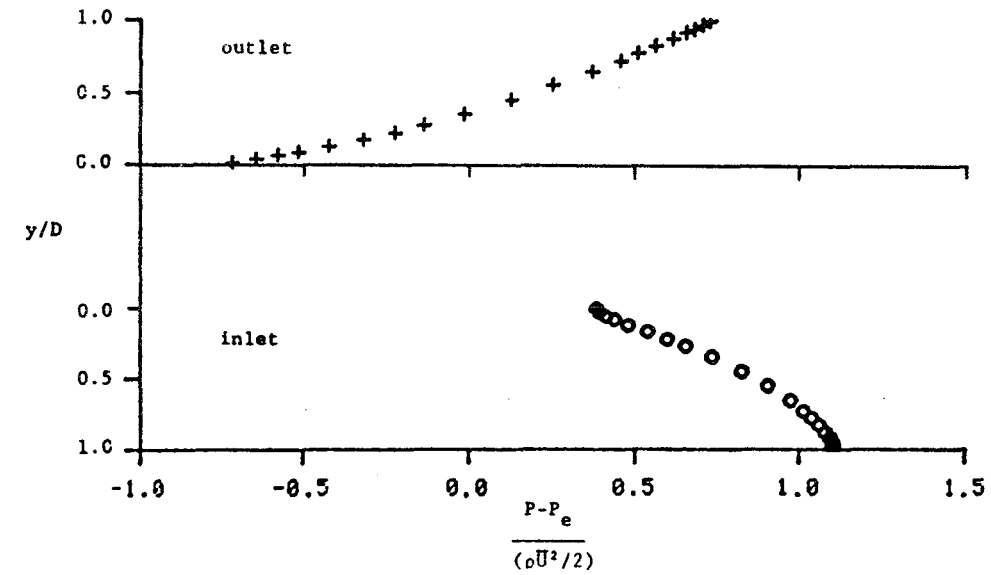


Figure 19. TAD I/O pressure profiles for BFC computation: Analysis 4.

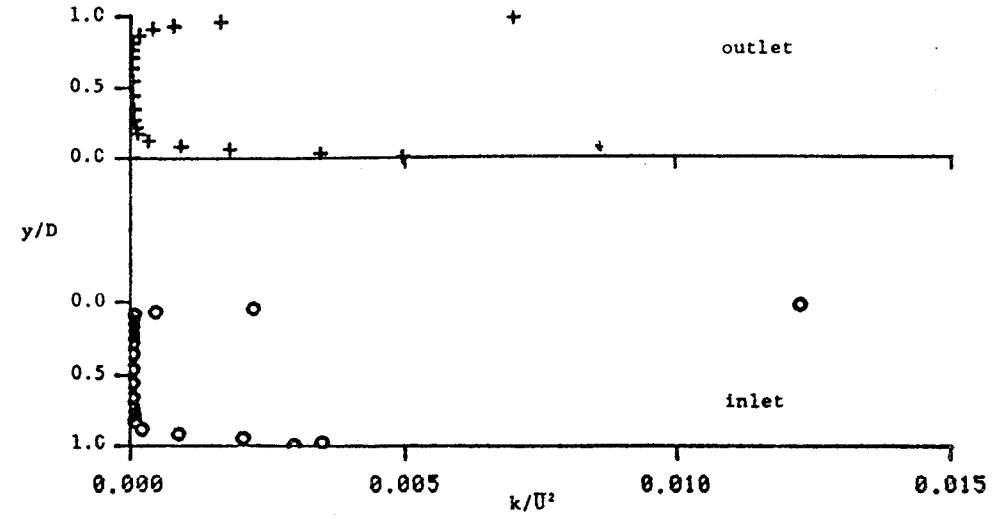


Figure 20. TAD I/0 turbulent kinetic energy profiles for BFC computation: Analysis 4.

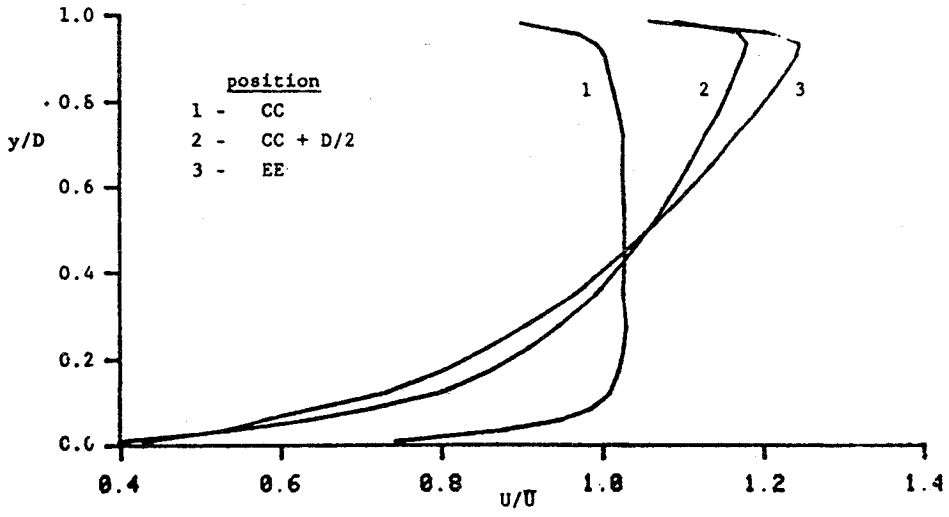


Figure 21. Developing velocity profiles in the straight 2 X D exit section of Analysis 4.

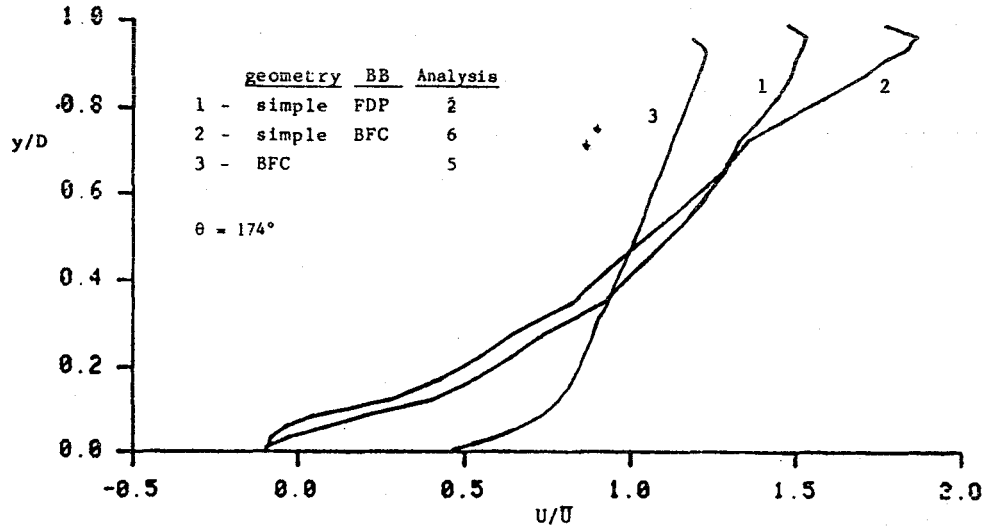


Figure 22. Comparison of TAD outlet velocity profiles.

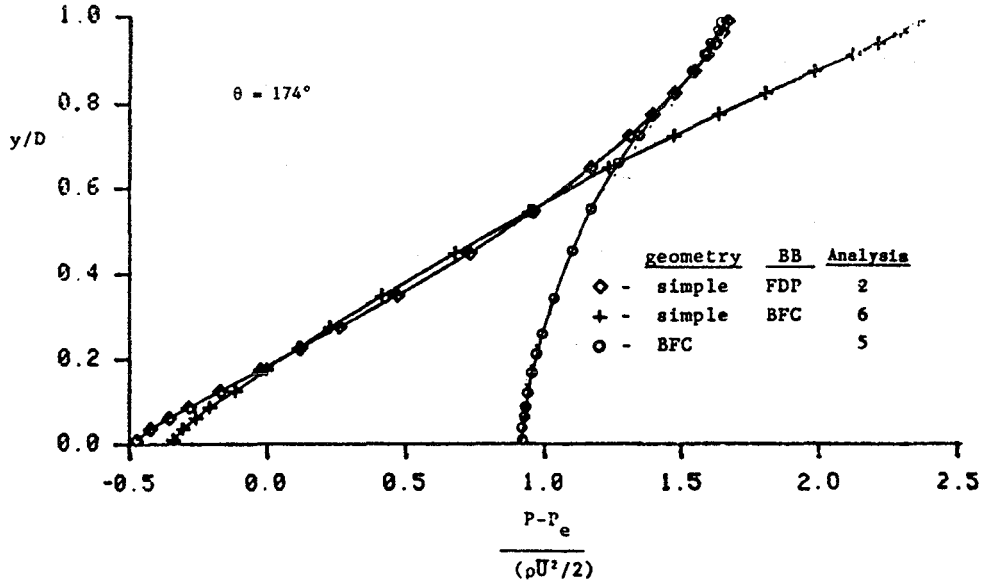


Figure 23. Comparison of TAD inlet pressure profiles.

separation was predicted approximately 24° from the exit plane, with the counterflow velocity zone extending into the fluid over 7.5% of the duct width at the exit plane. By contrast, the near inner wall node velocity at the exit plane, obtained using BFC conputations was over 40% of the average velocity. Pressure profiles from the two analyses at the TAD entrance plane were not even similar in character, as displayed in Figure 23.

If the standard cylindrical system analysis is presumed correct, based on an extensive history of computational success and better convergence characteristics, then the reliability of the current PHOENICS BFC option must be suspect. This conclusion must, however, be tempered by the fact that acceptable convergence was attained with cylindrical system analysis while only run time constrained convergence was achieved using BFC's. In addition, in order to prevent divergence of the PHOENICS based computational procedure involving BFC's, it was necessary to neglect certain curvature contributions in the finite difference approximations. The rationale for neglect of these terms is more fully described in the PHOENICS BFC documentation [21], however, for curvature dominated flows, such as that occurring in the SSME TAD, neglect of such contributions is dissatisfying. Only by further computational study and experimental comparison can deficiencies in the approximation procedure be adequately quantified.

CONCLUSIONS

1. The curvature modified k-ε turbulence model developed by Pourahmadi and Humphrey [9,10] cannot be utilized as presented due to algebraic errors in the $F(\sqrt{C_\mu})$ rooting scheme.

2. A corrected version of the Pourahmadi and Humphrey [9,10] curvature modified turbulence model has potential if an adequate method of selecting the multivalued root $\sqrt{C_\mu}$ can be determined.

3. Neglect of curvature related terms within the BFC option of the general purpose CFD code PHOENICS yields results inconsistent with the predictions of standard orthogonal system analyses for sharply curved internal flows.

4. Construction of an acceptable turbulence model for use in the analysis of sharply curved internal flows must await the availability of adequate experimental data, and will require a systematic program of computational verification.

REFERENCES

1. Eskinazi, S. and Yeh, H., "An Investigation of Fully Developed Turbulent Flows in a Curved Channel," Journal of Aeronautical Sciences, 1956, pp 23-24.

2. Ellis, L. B. and Joubert, P. N., "Turbulent Shear Flow in a Curved Duct," Journal of Fluid Mechanics, Vol.62, 1974, pp 65-84.

3. Meroney, R. N. and Bradshaw, P., "Turbulent Boundary-Layer Growth Over a Longitudinally Curved Surface," AIAA Journal, Vol. 13, 1975, pp 1446-1453.

4. Launder, B. E., Priddin, C. H., and Sharma, B. I., "The Calculation of Turbulent Boundary Layers on Spinning and Curved Surfaces," Trans. ASME, Journal of Fluids Engineering, 1977, pp 231-239.

5. Brinich, P. F. and Graham, R. W., "Flow and Heat Transfer in a Curved Channel," NASA TN D-8464, 1977.

6. Hunt, I. A. and Joubert, P. N., "Effects of Small Streamline Curvature on Turbulent Duct Flow," Journal of Fluid Mechanics, Vol. 91, 1979, pp 633-659.

7. Humphrey, J. A. C., Whitelaw, J. H., and Yee, G., "Turbulent Flow in a Square Duct with Strong Curvature," Journal of Fluid Mechanics, Vol. 103, 1981, pp 443-463.

8. Bradshaw, P., "Effects of Streamline Curvature on Turbulent Flow," AGARDograph No. 169, 1973.

9. Humphrey, J. A. C. and Pourahmadi, F., "A Generalized Algebraic Relation for Predicting Developing Curved Channel Flow with a k-ε Model of Turbulence," University of California, LBL Rept. 12009 Rev., 1981.

10. Pourahmadi, F. and Humphrey, J. A. C., "Prediction of Curved Channel Flow with an Extended k-ε Model of Turbulence," AIAA Journal, Vol. 21, 1983, pp 1365-1373.

11. Rodi, W., "A New Algebraic Relation for Calculating the Reynolds Stresses," ZAMM, Vol. 56, 1976, pp T219-T221.

12. Rotta, J. C., Turbulente Stromumgen, B. G. Teubner, Stuttgart, 1972.

13. Launder, B. E., Reece, G. J., and Rodi, W., "Progress in the Development of a Reynolds-Stress Turbulence Closure," Journal of Fluid Mechanics, Vol. 68, 1975, pp 537-566.

14. Daly, B. J. and Harlow, F. H., "Transport Equations of Turbulence," The Physics of Fluids, Vol. 13, 1970, pp 2634-2649.

15. Gibson, N. M. and Launder, B. E., "Ground Effects on Pressure Fluctuations in the Atmospheric Boundary," Journal of Fluid Mechanics, Vol. 86, 1978, pp 491-507.

16. Nakayama, A. and Koyama, H., "A Wall Law for Turbulent Boundary Layers in Adverse Pressure Gradients," AIAA Journal, Vol. 22, 1984, pp 1386-1389.

17. Launder, B. E. and Spalding, D. B., "The Numerical Computation of Turbulent Flows," Computer Methods in Applied Mechanics and Engineering, Vol. 3, 1974, pp 269-289.

18. Launder, B. E., Morse, A., Rodi, W., and Spalding, D. B., "The Prediction of Free Shear Flows - A Comparison of the Performance of Six Turbulence Models," Proceedings of NASA Conference on Free Shear Flows, Langley, 1972.

19. Spalding, D. B., "A General Purpose Computer Program for Multi-Dimensional One and Two-Phase Flow," Mathematics and Computers in Simulation, Vol. 23, 1981, pp 267-276.

20. Patankar, S. V., Numerical Heat Transfer and Fluid Flow, Hemisphere Publishing Corp., McGraw-Hill Book Co., New York, 1980.

21. Gunton, M. C., Malin, M. R., Rosten, H. I., Spalding, D. B., and Tatchell, D. G., "Use of Body-Fitted-Coordinate Scheme in Phoenics," CHAM TR/97, 1984.

22. Thompson, J. F. and Warsi, Z. U. A., "Boundary-Fitted Coordinate Systems for Numerical Solution of Partial Differential Equations - A Review," Journal of Computational Physics, Vol. 47, 1982, pp 1-108.

23. Korn, G. A. and Korn, T. M., Mathematical Handbook for Scientists and Engineers, Second Edition, McGraw-Hill Book Co., New York, 1968.

1985

NASA/ASEE SUMMER FACULTY RESEARCH FELLOWSHIP PROGRAM

Marshall Space Flight Center
The University of Alabama

MISSION ANALYSIS FLOW SEQUENCING OPTIMIZATION

Prepared By:	Meckinley Scott, Ph.D.
Academic Rank:	Professor
University and Department:	The University of Alabama Department of Mathematics
NASA/MSFC:	
Laboratory:	Program Development, Preliminary Design
Division:	Mission Analysis & Integration
Branch:	Operations Analysis
Date:	July 12, 1985
Contract No:	NGT 01-008-021 The University of Alabama in Huntsville

MISSION ANALYSIS FLOW SEQUENCING OPTIMIZATION

BY

Meckinley Scott
Professor of Mathematics
Unuversity of Alabama
Tuscaloosa, AL

ABSTRACT

This investigation is an extension of last year's project dealing with the problem of optimal use of ground resources for future space missions. This problem was formulated as a linear programming problem using an indirect approach. Instead of minimizing the inventory level of needed ground resources, we minimize the overlapping periods during which the same types of resources are used by various flights. The model was built upon the assumption that during the time interval under consideration, the costs of various needed resources remain constant. Under other assumptions concerning costs of resources, the objective function, in general, assumes a non-linear form. In this study, one case where the form of the objective function turns out to be quadratic is considered. Also, disadvantages and limitations of the approach used are briefly discussed.

1. Introduction

The problem of optimal utilization of ground resources for scheduling future space missions has been one of continuing interest and concern in the Program Development Division of NASA. Currently, planning for efficient use of ground resources is carried out using "GROPE" (Ground Resources Operations Program Executive). This is a series of computer programs that works with a traffic model and ground processing time lines as basic inputs. The traffic model is the specific Shuttle flights in a given year. Requirements for different ground resources and equipment are determined by the specific type of each flight, e.g., a Spacelab (pallet or module), deployed satellites with or without upper stages or Department of Defense flight. All Shuttle flights require: an Obiter Processing Facility where Orbiter refurbishment is done and also where horizontally installed payloads are integrated; and a Vertical Assembly Building where the solid rockets and external tank are stacked on the mobile launcher platform, then the Orbiter is attached to this stack, and a launch pad. Many flights also use the Vertical Processing Facility for vertically installed payload processing.

GROPE can be constrained or unconstrained. Constrained means limited resources are available on limited dates. Here it may not be possible for all flights to be scheduled within a given year. If this is the case, those flights that are not placed in the schedule are moved to the following year and scheduled first. In the unconstrained case, the entire traffic model is scheduled in the

proper year and it includes a complete specification of the various resource requirements in terms of "quantity" and "need" dates. Our interest in this project corresponds to the unconstrained case of GROPE. A precise description of the actual development of the mechanics of GROPE is, unfortunately, unavailable. However, it is known that its development is based mostly on heuristic grounds and lacks complete mathematical justification, rigor and formality. Due to the presence of a large number of variables, a complete mathematical formulation of this problem is no doubt very complex. The problem is further complicated by the fact that some of these variables are stochastic in nature. In our last year's report [5], we proposed an approximate mathematical model to formulate the problem. This formulation was based on assumptions which describe the actual situation fairly closely. It must, however, be pointed out that we have ignored a number of variables which have no direct bearing on the problem and all variables considered are assumed to be non-stochastic. Under the assumption that the costs of resources remain constant during the period under consideration, using an indirect approach, the problem was formulated as a linear programming problem. Under other assumptions for the costs of the resources, in general, the objective function is non-linear. In this study, one case where the form of the objective function turns out to be quadratic is considered. Also, in general, disadvantages and limitations of the approach used in this study are discussed.

2. The Model

The model was built and dealt with in an indirect manner.

Instead of analyzing the problem by minimizing the inventory level of needed resources, we use an approach whereby the flights are scheduled in such a way so as to minimize the overlapping periods during which the same types of resources are used by the various flights. Associated with each overlap for the use of the same type of resource by any pair of flights is a penalty cost which depends upon the number of units of that resource needed by these flights. Assuming that costs of resources remain unchanged during the whole period of the traffic model under consideration and that penalty costs are directly proportional to the lengths of the corresponding overlaps, the over-all objective function, which is the sum of such costs, is linear. Associated with the objective function is a number of sequencing and resource constraints which are also linear in form. Thus the problem under consideration falls within the domain of linear programming.

One program involvig use of linear programming for scheduling flights concerning space mission was prepared by Lockheed Electronics Co. in 1976 in the form of a technical report [1]. However, use of this program is basically restricted to cases dealing only with a fixed set of resources and the problem is one of selecting a traffic model from among various flight candidates which satisfy certain objectives. In our investigation, we have elaborated and refined the approach proposed in [6]. Stated below are the basic assumptions that concern our investigation.

(1) There are n flights to be launched during some given interval of time, [0, T].

(2) The order in which these flights are to be launched is predetermined.

(3) Of the n flights, there are p ($\leq$ n) specific flights with fixed launch dates. Each such flight may, however, have a launch window of a certain specified length.

(4) All flights utilize at least one type of ground resource from a collection of M different types.

(5) Any flight that utilizes say, a type 'k' ground resource may require n_k (≥ 1) units of that resource.

3. Notation and Formulation

Let, t_i, (i = 1,2,...,n) be a variable denoting the launch time of the ith flight.

For any flight i that utilizes a type 'k' resource before its launch time, s_i^{kl} denotes the length of time in which this resource is seized before t_i and d_i^{kl} denotes the corresponding duration for its use.

For $1 \leq i < j \leq n$, the non-negative overlap variables are denoted by O_{ij}^{kl}, (k = 1,2,...,M; r = 1,2 3,4).

The variable O_{ij}^{kl} measures the amount of the overlapping period for use of a type 'k' resource by the ith and jth flights when the seize times for the resource occur before their launch times. The variable O_{ij}^{k2} measures the amount of the overlapping period for use of a type 'k' resource by the ith and jth flights when the seize times for the resource occur before t_i for the ith flight and after t_j for the jth flight. The variables O_{ij}^{k3} and O_{ij}^{k4} are similarly interpreted. In the case of O_{ij}^{k3}, the

seize times for the resource occurs after t_i for the ith flight and before t_j for the jth flight. For O_{ij}^{k4}, the seize times for the resource by both flights occur after their launch times.

Clearly, for a traffic model consisting of n flights and M types of resources, the launch time variables $t_1, t_2, \ldots, t_n$ generate $2Mn(n-1)$ non-negative overlap variables.

The assumption that no two flights are to be scheduled at the same instant of time leads to constraints of the type

$$t_{i+1} - t_i \geq d > 0, \quad (i = 1,2,\ldots,n). \tag{1}$$

The provision that there are p specific flights with fixed launch dates at times, say, t_{n_i}, $(i = 1,2,\ldots,n)$ may have launch windows yields constraints of the type

$$t_{n_i} \geq \underline{l}_i, \; t_n \leq \bar{l}_i, \quad (i = 1,2,\ldots,p). \tag{2}$$

Also note that

$$t_n \leq T. \tag{3}$$

The relationships between the variables O_{ij}^{kr} and t_i, $(1 \leq i < j \leq n, \; r = 1,2,3,4; \; k = 1,2,\ldots,M)$ are given by the equality constraints

$$-t_i + t_j + O_{ij}^{kr} = d_{ij}^{kr}, \tag{4}$$

where the constants d_{ij}^{kr} are defined by

$$d_{ij}^{k1} = d_i^{k1} - s_i^{k1} + s_j^{k1},$$

$$d_{ij}^{k2} = d_i^{k1} - s_i^{k1} - s_j^{k2},$$

$$d_{ij}^{k3} = d_i^{k2} + s_i^{k2} + s_j^{k1},$$

$$d_{ij}^{k4} = d_i^{k2} + s_i^{k2} - s_j^{k2}.$$

Let n_i^{k1} and n_i^{k2} denote, respectively, the number of units of a type 'k' resource utilized by the ith flight when the seize times for this resource occur before and after t_i, n_j^{k1}

and n_j^{k2} for the jth flight are similarly defined. These quantities are used in the formulation of the cost functions. Under the assumption that costs of resources remain constant during the period under consideration, we consider the case where the cost, C_{ij}^{kr} associated with the overlap variable O_{ij}^{kr} is given by

$$C_{ij}^{kr} = n_{ij}^{kr}\, O_{ij}^{kr}\, C^{(k)},$$

$(1 \leq i < j \leq n;\ r = 1,2,3,4;\ k = 1,2,\ldots,M).$

Here,

$$n_{ij}^{k1} = \text{Min}(n_i^{k1},\ n_j^{k1}),$$

$$n_{ij}^{k2} = \text{Min}(n_i^{k1},\ n_j^{k2}),$$

and $C^{(k)}$ is the weight given to a type 'k' resource based on cost considerations relative to other types of resources in the collection. The over-all objective function T, say, is the sum of costs C^{kr} over all values of k, r, i and j, $(i < j)$.

After a considerable amount of straight forward algebra, it can be shown that mininizing T is equivalent to minimizing T_1 given by

$$T_1 = \sum_{i=1}^{n} c_i t_i,$$

where

$$c_i = \sum_{k=1}^{M} C^{(k)} \left(\sum_{i<r\leq n} m_{ir}^{k} - \sum_{1\leq s<i} m_{si}^{k} \right),\ (i = 1,2,\ldots,n)$$

$$m_{ij}^{k} = \sum_{r=1}^{4} n_{ij}^{kr},\ (1 \leq i < j \leq n).$$

Recall that T_1 which is linear in t_i's is to be minimized

subject to linear constraints given by (1), (2), (3), and (4).

4. Some Remarks

In the model developed above, we have assumed that there are p specific flights out of n with launch dates fixed at times t_{n_i}, $(i = 1,2,3,\ldots,p)$; each of which has a launch window of a certain specified length. Since the order of the flights for any given traffic model is predetermined, the launch times t_i, $(i = 1,2,\ldots,n)$ satisfy

$$0 \leq t_1 < t_2 \ldots < t_{n_1} < t_{n_1+1} \ldots < t_{n_2-1} < t_{n_2} < t_{n_2+1} \ldots < t_{n_p-1} < t_{n_p} < t_{n_p+1} \ldots < t_{n-1} < t_n.$$

In the case where $n_p = n$, that is, the last flight is one with a fixed launch date, the constraint $t_n \leq T$ will be replaced by one of the form

$$t_n \leq \bar{l}_p.$$

Since all flights are to take place during the interval $[0,T]$, $\bar{l}_p$ should satisfy the condition

$$\bar{l}_p \leq T.$$

Similarly, in the case where $n_1 = 1$, we have the condition

$$t_1 \geq \underline{l}_1$$

with $\underline{l}_1 \geq 0$.

5. Other Cost Functions

In the model developed above, due to the assumption that costs of various resources remain constant, the objective function turned out to be linear. If this assumption is relaxed, in general, the objective function is non-linear and this would considerably increase the degree of difficulty of the problem. Below we consider the case where the objective function assumes a quadratic form.

Here, we assume that the cost of each type of resource increases linearly with increasing time. Specifically, we assume that the cost $C^{(k)}$ for the kth resource at time t_i is

$$C^{(k)} = a_k + b_k t_i,$$

and the penalty cost C_{ij}^{kr} associated with the overlap variable O_{ij}^{kr} is given by

$$C^{kr} = (C_j^{(k)} - C_i^{(k)})\, n_{ij}^{kr}\, O_{ij}^{kr}$$
$$= b_k\, (t_j - t_i)\, n_{ij}^{kr}\, O_{ij}^{kr},$$

$(r = 1,2,3,4;\ k = 1,2,\ldots,M;\ 1 \leq i < j \leq n).$

The overall objective function is thus

$$T = \sum_{} C_{ij}^{kr},$$

where the sum is to be performed over all values of r, k, i and j, $(1 \leq i < j \leq n)$.

After a lengthy but straight forward algebra, it can be shown that

$$T = \sum_{} l_{ij}(t_i - t_j)^2 + \sum_{} r_{ij}(t_i - t_j)$$

where each sum is to be performed over the set

$\{(i,j)/1 \leq i < j \leq n\}$, and

$$l_{ij} = - \sum_{k=1}^{M} \sum_{r=1}^{4} n_{ij}^{kr}\, b_k,$$

$$r_{ij} = - \sum_{k=1}^{M} \sum_{r=1}^{4} n_{ij}^{kr}\, d_{ij}^{kr}\, b_k.$$

Further simplification reduces the objective function to the form

$$T = \sum_{i=1}^{n} p_i t_i + \sum_{1 \le i < j \le n} 2(l_{ij} t_i t_j) + \sum_{i=1}^{n} l_{ii} t^2,$$

where

$$p_i = \sum_{1 \le j < i} r_{ji} - \sum_{i < j \le n} r_{ij},$$

$$l_{ii} = - \sum_{1 \le r < i} l_{ri} - \sum_{i < r \le n} l_{ir}, \quad (i = 1,2,\ldots,n).$$

The form of the objective function T does not in any way alter the form of the constraints and hence T is to be minimized subject to the same constraints given by (1), (2), (3), and (4). In terms of matrix notation, the objective function can be written as

$T = \underline{p}t' + \underline{t}Lt'$,

where

$L = (l_{ij})$ is a symmetric n X n matrix,

$\underline{t} = (t_1, t_2, \ldots, t_n)$,

$\underline{p} = (p_1, p_2, \ldots, p_n)$,

and t' and p' are column vectors corresponding to $\underline{t}$ and $\underline{p}$.

6. Discussion and Conclusion

At the present time, there are some convincing arguments that if we are dealing with a relatively short interval of time, (e.g., one year), the costs of resources could very well remain constant. Thus, in such a case, the formulation of the problem as a linear programming problem is well justified. Furthermore, this assumption

does lend itself to easier treatment both from the mathematical and computational aspects of the problem. Other assumptions for the cost function, in general, considerably increase the complexity of the problem especially from the computational point of view.

As mentioned earlier, for a traffic model with n flights, in addition to the n launch times, variables, in general, there are $2Mn(n-1)$ overlap variables and the number of constraints is $n + 2p + 2Mn(n-1)$. Thus, if there are no flights with fixed launch dates, the number of variables is the same as the number of constraints. An increase in the number of resources by one may cause an increase of as many as $2n(n-1)$ variables. In actual practice, the number of variables actually needed may be reduced if, for example, we know that the duration for use of any resource is short relative to the interval under consideration. On the other hand, for any flight type, there would be a minimum of about 15 ground resources where cost considerations should not be ignored. Thus, even if we have as few as 20 flights per year, the number of variables involved may be as many as $20 + 2(15)(20)(19) = 11420$, a very large number indeed. The recent findings of Karmarkar [2] concerning an algorithm for large scale linear programming problems when available will, hopefully, prove useful to our problem. So far his approach exists only in what has been described as rougher computer code and, perhaps, it has not been tested on a wide range of problems.

In the analysis of the problem, we have used an indirect approach to optimize the use of needed ground resources. This approach provides us with launch times for scheduling flights of any

given traffic model. However, note that it does not directly lend itself to the computation of the total cost of resources needed. In order to find the total cost corresponding to any solution $(t_1,t_2,\ldots,t_n)$ of the launch times, the peak requirement for each resource will have to be determined in each case.

REFERENCES

1. Chang, A., "Set Covering Algorithm - A Subprogram of the Scheduling Algorithm for Mission Planning and Logistic Evaluation", Lockheed Electronics Co. Publication, NASA-CR-147551, March 1976.

2. Karmarkar, N., Report in the New York Times, November 20, 1984, (Other sources, e.g., article in Combinatorica, are unavailable at the present time.)

3. Mitra, G., Theory and Application of Mathematical Programming, Academic Press, New York, 1976.

4. Rao, S.S., Optimization - Theory and Applications, Halsted Press, New Delhi, 1984.

5. Scott, M., "A Mathematical Formulation of the Problem of Optimal Use of Ground Resources for Future Space Missions", Research Reports NASA-ASEE Summer Faculty Fellowship Program, 1977.

6. Traver, A.E., "Investigation of the Application of Linear Programming Techniques to Optimal Project Scheduling", Research Reports NASA-ASE Summer Faculty Fellowship Program, 1984.

SIMULATION AND STUDY OF SMALL NUMBERS OF RANDOM EVENTS

BY

RUSSELL D. SHELTON

INSTRUCTOR IN PHYSICS ELECTRONICS AND DATA PROCESSING
J. C. CALHOUN COMMUNITY COLLEGE
DECATUR, ALABAMA

ABSTRACT

RANDOM EVENTS WERE SIMULATED BY COMPUTER AND SUBJECTED TO VARIOUS STATISTICAL METHODS TO EXTRACT IMPORTANT PARAMETERS. VARIOUS FORMS OF CURVE FITTING WERE EXPLORED, SUCH AS LEAST SQUARES, LEAST DISTANCE FROM A LINE, MAXIMUM LIKELIHOOD. PROBLEMS CONSIDERED WERE DEAD TIME, EXPONENTIAL DECAY, AND SPECTRUM EXTRACTION FROM COSMIC RAY DATA USING BINNED DATA AND DATA FROM INDIVIDUAL EVENTS. COMPUTER PROGRAMS, MOSTLY OF AN ITERATIVE NATURE, WERE DEVELOPED TO DO THESE SIMULATIONS AND EXTRACTIONS AND ARE PARTIALLY LISTED AS APPENDICES. THE MATHEMATICAL BASIS FOR THE COMPUTER PROGRAMS IS GIVEN IN THE TEXT OF THE REPORT.

ACKNOWLEDGEMENTS

DR. P. B. EBY SUGGESTED THE POSSIBLE INTEREST IN EXTRACTING DECAY PARAMETERS FROM RARE EVENTS. A NOTE BY DR. R. F. ELSNER STIMULATED A LOOK AT DEAD TIME CORRECTIONS. CONVERSATIONS WITH DRS. R. DECHER, E. URBAN, P. PETERS, AND T. PARNELL HELPED WITH THE CHOICE OF WORK TO BE DONE. MR. JOHN WATTS PROVIDED COSMIC RAY DATA AND HELPFUL INFORMATION ON HOW IT HAD BEEN PROCESSED AND INTERPRETED.

TABLE OF CONTENTS

SIMULATION AND STUDY OF SMALL NUMBERS OF RANDOM EVENTS

2. RELATIONSHIP OF BINOMIAL, POISSON, AND NORMAL DISTRIBUTIONS

IF WE DIVIDE A TIME INTERVAL T INTO n EQUAL SMALL SUBINTERVALS OF LENGTH Dt, A RANDOM EVENT IS EQUALLY LIKELY TO OCCUR DURING ANY SUBINTERVAL OF DURATION Dt. THE PROBABILITY Dp OF AN EVENT DURING ANY GIVEN SUBINTERVAL Dt IS GIVEN BY THE EXPRESSION,

$$Dp = L\ Dt \ , \qquad 2.1$$

WHERE L IS THE AVERAGE NUMBER OF EVENTS PER UNIT TIME AND Dt IS SO SMALL THAT TWO COUNTS DURING THE INTERVAL Dt IS HIGHLY UNLIKELY. THE PRODABILITY THAT AN EVENT WILL NOT OCCUR IN ANY GIVEN INTERVAL Dt IS GIVEN BY THE EXPRESSION,

$$1 - Dp = 1 - L\ Dt \qquad 2.2$$

THE PROBALITY P(N,T) OF N EVENTS IN TIME T IS GIVEN BY THE BINOMIAL DISTRIBUTION,

$$P(N,T) = [L*Dt\ ]^N * [(1-Dt]^(n-N)]* n!/[(n-N)!\ N!] \ . \qquad 2.3$$

IN EQ 2.3 AND FOLLOWING, THE SYMBOL * DENOTES MULTIPLICATION AND THE SYMBOL ^ PRECEDES AN EXPONENT. P(N,T) IS THE Nth TERM OF THE POLYNOMIAL [L*Dt+(1-L*Dt)]^n. THIS DISTRIBUTION IS THEREFORE NORMALIZED, AND ANY DISTRIBUTION DERIVED FROM IT IN A LIMITING PROCESS SHOULD BE NORMALIZED.

THE POISSON DISTRIBUTION IS EASILY OBTAINED FROM THE BINOMIAL DISTRIBUTION BY USING THE STERLING APPROXIMATION,

$$LOG\ n! = (1/2)\ LOG\ (2\ PI) + (\ n + 1/2\ )\ LOG\ n - n \ , \qquad 2.4$$

AND APPROXIMATIONS LIKE

$$LOG\ (n-N) = LOG\ [\ n\ (1 - N\ /\ n\ )\ ] = LOG\ n - n/N \ . \qquad 2.5$$

THESE APPROXIMATIONS ARE QUITE GOOD WHEN n > > N. IF WE TAKE THE NATURAL LOGARITHM OF EQ 2.3 AND USE EQUATIONS 2.4 AND 2.5, WE OBTAIN THE POISSON DISTRIBUTION,

$$P(N,T) = [(\ L\ T\ )^N]*[EXP\ (-\ L*T)]/N! \qquad 2.6$$

THE AVERAGE NUMBER OF EVENTS N IN TIME T IS (L*T) AND THE STANDARD DEVIATION FOR THE POISSON DISTRIBUTION IS SQR (L*T).

AS THE NUMBER OF EVENTS N BECOMES LARGE, THE POISSON BECOMES EQUIVALENT TO THE NORMAL DISTRIBUTION. IF WE TAKE THE NATURAL LOGARITHM OF THE POISSON DISTRIBUTION AND USE THE STIRLING APPROXIMATION AND THE APPROXIMATION,

$$LOG\ [\ 1+X/(L*T)\ ] = X\ /\ L\ T - [(X/L*T)^2] \ /\ 2 \ , \qquad 2.7$$

WHERE

$$X = N - L\,T\,, \qquad 2.8$$

WE FIND THAT

$$P(X,T) = SQR\,[\,1\,/\,(2\,PI\,L*T\,]\,*\,EXP\,[-X\hat{}2/(\,2\,L*T\,)\,], \qquad 1.9$$

WHICH IS THE NORMAL DISTRIBUTION WITH AVERAGE LT (X = 0) AND STANDARD DEVIATION SQR (L T). TABLE 2.1 COMPARES THE BINOMIAL, POISSON AND NORMAL DISTRIBUTIONS FOR N = 30 , T = 5 , AND L = 3 .

3 DEAD TIME CORRECTIONS FOR RANDOM EVENTS.

SUPPOSE THAT AN EVENT DETECTOR (COUNTER) HAS A DEAD TIME t AFTER EACH COUNT. IF N' RANDOM EVENTS ARE OBSERVED DURING TIME T, WE KNOW THAT THE COUNTER WAS DEAD (INOPERATIVE) FOR A TIME N'*t. THE AVERAGE NUMBER OF EVENTS EXPECTED DURING THIS DEAD TIME WOULD BE L*N'*t AND WOULD ALSO OBEY A POISSON DISTRIBUTION. THE NUMBER OF EVENTS N IS ESTIMATED TO BE GIVEN BY THE EQUATION,

$$N = N\,' + L\,N\,'\,t \qquad . \qquad 3.1$$

IT DOES NOT MATTER HOW OFTEN OUR COUNTER IS TURNED ON OR OFF. IF THE EVENTS ARE RANDOM AND L IS A CONSTANT, EQ 3.1 IS VALID WHEN N ' REPRESENTS ALL THE OBSERVED COUNTS AND T REPRESENTS THE TOTAL TIME THAT THE COUNTER IS ON. IF WE MAKE THE ASSUMTION THAT N IS EQUAL TO L*T , EQ 3.1 BECOMES

$$L*T = N' + L*N'*t \qquad 3.2$$

OR

$$L = N'/\,(\,T - N'*t\,) \qquad . \qquad 3.3$$

EQ 3.3 IS THE SIMPLE STATEMENT THAT THE BEST ESTIMATE OF THE NATURAL COUNT RATE L IS THE TOTAL NUMBER OF COUNTS DIVIDED BY THE TOTAL TIME THE COUNTER IS ON. IT IS POOR PRACTICE TO REPLACE L IN EQ 3.1 BY N / T AND USE THE RESULT,

$$N = N\,'\,/\,(\,1 - N\,'\,t\,/\,T\,), \qquad 3.4$$

AS A CORRECTED COUNT. THIS PRACTICE OVERCORRECTS WHEN N ' IS LARGE AND UNDERCORRECTS WHEN N ' IS SMALL.

THERE ARE A NUMBER OF TECHNIQUES WHICH MAY BE USED TO STUDY COUNTER DEAD TIME. A FIRST APPROACH IS TO OVERWHELM THE COUNTER WITH MANY MORE EVENTS THAN IT CAN POSSIBLY COUNT. IF THE EVENT RATE L IS RELATED TO THE COUNT RATE R BY THE EQUATION,

$$R = L\,/\,(\,1 + L*t\,) \qquad , \qquad 3.5$$

WHERE t IS THE DEAD TIME AFTER EACH EVENT, THE MAXIMUM POSSIBLE COUNT RATE R(MAX) IS SEEN TO BE GIVEN AS 1 / t, SO THAT THE DEAD TIME t = 1 / R(MAX). IF THE COUNTER DETECTS RADIATION, A MORE

TABLE 2.1

A comparison of binomial, Poisson, and Gaussian distributions for n = 30, T = 5, and λ = 3 corresponding to an average count of 15

N	B(N)	P(N)	G(N)
1.000	0.000	0.000	0.000
2.000	0.000	0.000	0.000
3.000	0.000	0.000	0.001
4.000	0.000	0.001	0.002
5.000	0.000	0.002	0.004
6.000	0.001	0.005	0.007
7.000	0.002	0.010	0.012
8.000	0.005	0.019	0.020
9.000	0.013	0.032	0.031
10.000	0.028	0.049	0.045
11.000	0.051	0.066	0.060
12.000	0.081	0.083	0.076
13.000	0.112	0.096	0.090
14.000	0.135	0.102	0.100
15.000	0.144	0.102	0.103
16.000	0.135	0.096	0.100
17.000	0.112	0.085	0.090
18.000	0.081	0.071	0.076
19.000	0.051	0.056	0.060
20.000	0.028	0.042	0.045
21.000	0.013	0.030	0.031
22.000	0.005	0.020	0.020
23.000	0.002	0.013	0.012
24.000	0.001	0.008	0.007
25.000	0.000	0.005	0.004
26.000	0.000	0.003	0.002
27.000	0.000	0.002	0.001
28.000	0.000	0.001	0.000
29.000	0.000	0.000	0.000
30.000	0.000	0.000	0.000

SUITABLE APPROACH INITIALLY IS TO USE A RADIATION SOURCE AND THE INVERSE SQUARE LAW FOR CALIBRATION. THE EVENT RATE AT THE COUNTER IS GIVEN AS

2 R = K / r^2 , 3.6

WHERE K IS PROPORTIONAL TO THE SOURCE INTENSITY. FROM EQS 3.5 AND 3.6,

R = L / (1 = L*T) = 1 / [(r^2 / k) + t], 2.7

SO THAT

R*r^2 = - R*K*t + K . 2.9

A PLOT OF R*r^2 VERSUS R SHOULD BE A STRAIGHT LINE WITH SLOPE -K * AND INTERCEPTS,

R*r^2 = K ; R = 1 / t . 3.10

IF THIS IS NOT SO, DEAD TIME t IS NOT INDEPENDENT OF COUNT RATE.

4. COMMON STATISTICS FOR EXPONENTIAL DECAY

THE PROBABILITY Dp OF A RADIOISOTOPE DECAY DURING THE TIME INTERVAL DT IS GIVEN BY THE EQUATION,

Dp = L*N*DT , 4.1

WHERE N IS THE NUMBER OF ATOMS AVAILABLE FOR DECAY AND L IS THE DECAY CONSTANT. IF WE DIVIDE TIME T INTO n EQUAL INTERVALS DT, THE PROBABILITY OF NO DECAY IN TIME T FOLLOWED BY A DECAY IN TIME DT IS GIVEN AS

P(T)*Dt = (1 - L*N*DT)^n *L*N*DT . 4.2

AS n BECOMES LARGE THIS REDUCES TO

P(T)*dT = [EXP (-L*N*T)]*L*N*dT . 4.3

THIS DISTRIBUTION OF THE TIME BETWEEN DECAYS HAS AN AVERAGE OF L*N AND A STANDARD DEVIATION OF L*N. GIVEN THE TIMES T(I) ASSOCIATED WITH A NUMBER OF DECAYS N(I), A NUMBER OF TECHNIQUES HAVE BEEN USED TO COMPUTE THE ORIGINAL NUMBER OF ATOMS N(O) AND THE DECAY CONSTANT L. IF THERE ARE MANY DECAYS PER UNIT TIME, IT IS CONVENIENT TO WRITE THE DIFFERENTIAL EQUATION,

dN / dT = - L N , 4.4

WHICH HAS THE SOLUTION,

N(T) = N(O) EXP (- L T) . 4.5

SINCE THE COUNT RATE IS PROPORTIONAL TO N(T), A PLOT OF LOG COUNT RATE VERSUS TIME HAS A SLOPE OF -L AND AN INTERCEPT L*N(0).

IF EVENTS ARE FEW, WE NEED TO DEVELOP TECHNIQUES TO EXTRACT AS MUCH INFORMATION AS POSSIBLE FROM THE DATA AT HAND. TO TEST THE VARIOUS TECHNIQUES, DECAY DATA WAS GENERATED BY COMPUTER AND THE DATA WAS PROCESSED BY EACH TECHNIQUE TO SEE HOW WELL IT WOULD DO.

A COMPUTER GENERATES RANDOM NUMBERS X(I) DISTRIBUTED UNIFORMLY IN THE INTERVAL BETWEEN ZERO AND UNITY. TO PRODUCE NUMBERS CORRESPONDING TO RANDOM DECAY TIMES, WE EQUATE THE DISTRIBUTIONS,

$$P(T)\,dT = F(X)\,dX, \qquad 4.6$$

WHERE P(T) COMES FROM EQ 4.3 AND F(X) IS UNITY. THE INDEFINITE INTEGRAL OF BOTH SIDES OF EQ 4.6 YIELDS

$$T = -\,(\,1\,/\,N*L\,)\,\mathrm{LOG}\,(\,1 - X\,) \quad . \qquad 4.7$$

BECAUSE X IS A RANDOM VARIABLE, WE CAN WRITE

$$T(I) = -\,(\,1\,/\,N(I)*L\,)\,\mathrm{LOG}\,(\,1 - X(I)\,) \quad , \qquad 4.8$$

WHERE

$$N(I) = N(0) - I + 1 \quad . \qquad 4.9$$

IF EVENTS ARE FEW, WE CAN RECORD THE TIME T(I) FOR EACH DECAY AND FIT EQ 4.5 TO THE DATA. THE RESULTS ARE GENERALLY POOR IF WE USE LEAST SQUARES FITS. USING THE DISTRIBUTION OF EQ 4.3, WE CAN CALCULATE THE AVERAGE TIME t(I) BETWEEN DECAYS TO BE

$$t(I) = 1\,/\,[\,(\,N(0) - I + 1)*L\,] \quad . \qquad 4.10$$

AFTER REARRANGING,

$$I*t(I) = (\,N(0) + 1\,)*t(I) - 1\,/\,L \quad . \qquad 4.11$$

A LEAST SQUARES CURVE FIT OF I*t(I) VERSUS t(I) SHOULD YIELD A SLOPE OF N(0)+1 AND AN INTERCEPT OF -1 / L . THIS OFFERS LITTLE IMPROVEMENT. WE COULD REARRANGE EQ 4.5 TO THE FORM,

$$T(I) = (1/L)\,\mathrm{LOG}\,N(0) - (1/L)\,\mathrm{LOG}\,N(I) \quad . \qquad 4.12$$

A LEAST SQUARES FIT OF T(I) VERSUS LOG N(I) YIELDS A SLOPE OF -1/L AND AN INTERCEPT OF (1/L) LOG N(0).

WE MIGHT THINK THAT THE METHOD OF LEAST SQUARES GIVES TOO MUCH WEIGHT TO ERRATIC (RARE) EVENTS AND THAT SOME SCHEME TO GIVE LESS WEIGHT TO POINTS DISTANT FROM THE CURVE FIT MIGHT BE BETTER. A SQUARE ROOT AND FOURTH ROOT OF THE SUM OF SQUARES RESULTED IN SMALL IMPROVEMENTS.

AS A FINAL EXERCISE, THE METHOD OF MAXIMUM LIKELIHOOD WAS USED TO DERIVE L AND N(O). IT IS ASSUMED THAT A SEQUENCE OF DECAY EVENTS, DESCRIBED BY I AND t(I), THE TIME BETWEEN EVENTS I AND I-1, HAS A PROBABILITY CLOSE TO THE MAXIMUM. THE PROBABILITY FOR SUCH A SEQUENCE CAN BE WRITTEN AS

$$P[t(1),t(2),..] = P(1)*P(2)*P(3)*..... \qquad 4.13$$

WHERE K IS A CONSTANT OF PROPORTIONALITY AND

$$P(I) = K*L*N(I)*EXP\ [-N(I)*L*t(I)\ ] \quad . \qquad 4.14$$

THIS PRODUCT IS MORE MANAGABLE AS A LOGARITHM IN THE FORM,

$$LOG\ P[t(1),t(2),..]=SUM(I)(LOG[(K*L*N(I)]-N(I)*L*t(I)) \qquad 4.15$$

USING AN ITERATION SCHEME WE CHOOSE VALUES OF N(O) AND L WHICH MAXIMIZE THIS FUNCTION.

RESULTS USING THE VARIOUS TECHNIQUES ARE SHOWN IN TABLE 4.1. IT IS SEEN THAT THE METHOD OF MAXIMUM LIKELIHOOD IS CONSISTENTLY SUPERIOR FOR N(O) > 10. PLACING LESS EMPHASIS ON DISTANT POINTS IS WORTHWHILE IF THE METHOD OF MAXIMUM LIKELIHOOD IS NOT USED. TABLE 4.2 SHOWS THE RESULTS OF USING THE DIFFERENT METHODS ON THE SAME SET OF DATA.

AS AN ASIDE, IT IS NOTED THAT EQ 4.15 COULD BE USED FOR A POISSON DISTRIBUTION IF L*N(I) IS REPLACED BY L. THE RESULT IS GIVEN AS

$$\begin{aligned} LOG\ P[t(1),t(2),..] &= SUM(I)\ LOG\ K*L - L*t(I) \quad , \\ &= N*LOG\ K*L - L*T \quad , \qquad 4.16 \end{aligned}$$

WHERE T IS THE TOTAL TIME FOR N COUNTS. THE MAXIMIZATION OF THIS FUNCTION OCCURS WHEN L=N/T, SO THAT THE DISTRIBUTION OF TIME BETWEEN EVENTS IS NOT USEFUL IN STUDYING A POISSON DISTRIBUTION. ONLY THE TOTAL TIME T AND THE TOTAL COUNTS N IS IMPORTANT.

5. TREATMENT OF POWER SPECTRUM DATA

EVENTS OBEYING THE DISTRIBUTION,

$$P(E)*dE = A*E\ \hat{}\ B*dE\ , \qquad 5.1$$

ARE CHARACTERISTIC OF COSMIC RAYS AND RADIATION BELTS, WHERE E IS THE PARTICLE ENERGY AND P(E)*dE IS THE PROBABILITY OF OBSERVING A PARTICLE WITH ENERGY BETWEEN E AND E+dE. THE NORMALIZATION OF THIS FUNCTION RESULTS IN THE EQUALITY,

$$A = (B-1)/(LL\hat{}(-B+1)-UL\hat{}(-B+1)) \quad , \qquad 5.2$$

WHERE LL IS THE LOWER ENERGY LIMIT AND UL IS THE UPPER ENERGY LIMIT. USING THE METHOD OF SECTION 4, THIS SPECTRUM IS SIMULATED

TABLE 4.1

SUMMARY OF FINDINGS ON CURVE FITTING

N(O) = 5 LAMBDA = 1 SEQUENCES = 7

METHOD*	LAMBDA	S. D.	N(O)	S. D.
1	2.45	5.23	3.68	1.14
2	0.91	0.58	5.46	1.62
3	1.01	0.62	6.13	2.66
4	1.24	1.08	7.49	6.61

N(O) = 10 LAMBDA = 1 SEQUENCES = 7

METHOD*	LAMBDA	S. D.	N(O)	S. D.
1	2.03	1.08	8.56	0.69
2	0.93	0.50	10.65	2.27
3	0.99	0.52	11.32	2.50
4	0.97	0.18	10.46	1.19
5	1.04	0.28	10.06	1.96

N(O) = 30 LAMBDA = 1 SEQUENCES = 7

METHOD*	LAMBDA	S. D.	N(O)	S. D.
1	2.03	0.45	27.92	1.44
2	0.93	0.07	28.21	4.05
3	0.99	0.08	29.65	4.30
4	1.00	0.12	30.11	0.29

N(O) = 50 LAMBDA = 1 SEQUENCES = 7

METHOD*	LAMBDA	S. D.	N(O)	S. D.
1	1.39	0.22	47.02	1.67
2	0.91	0.14	47.12	3.01
3	0.93	0.14	48.06	2.87
4	1.00	0.11	50.51	0.81
5	1.01	0.11	50.46	1.00
6				

METHODS

1. L.S.F. TO I*TAU(J) = (N(O) + 1)*TAU(I) - 1/L
2. L.S.F. TO LOG (N(O) - I) = -L*T(I) + LOG (N(O))
3. L.S.F. TO T(I) = (-1/L)*LOG (N(O) - I) + (1/L) LOG (N(O))
4. MAXIMUM LIKELIHOOD ITERATION
5. LEAST DISTANCE ITERATION
6. ROOT MEAN SQUARE ITERATION

L.S.F. - LEAST SQUARES FIT ; L - DECAY RATE CONSTANT TAU(I) - TIME BETWEEN EVENTS I-1 AND I ; N(O) - ORIGINAL NUMBER OF EVENTS ; T(I) - TIME WHEN EVENT I OCCURS.

TABLE 4.2

COMPARISON OF SOME METHODS USING THE SAME SET OF DATA

SEED = 11111 EVENTS = 30 LAMBDA = 1

N(O)	LAMBDA	METHOD*	EQUATION**
28.9	1.27	1	1
27.5	0.88	1	2
28.8	0.93	1	3
29.3	2.05	2	1
29.7	0.97	2	2
29.8	0.98	2	3
29.1	2.99	3	1
28.9	0.96	3	2
33.9	1.17	3	3
29.2	1.99	4	1
27.5	0.88	4	2
30.6	1.01	4	3
29.5	1.14	5	4

* METHODS
1. LEAST SUM OF SQUARES OF (Y(I)-M*X(I)-B)
2. LEAST SUM OF ABS(Y(I)-M*X(I)-B)
3. LEAST SUM OF SQR[ABS(Y(I)-M*X(I)+B))
4. LEAST SUM OF DISTANCES OF POINTS FROM LINE
5. MAXIMUM LIKELIHOOD

** EQUATIONS
1. I*t(I) = (N(O)+1)*t(I)-1/L
2. LOG N(I) = -L*T(J) + LOG N(O)
3. T(I) = -[LOG N(I)]/L +[LOG N(O)]/L

N(I) = EVENTS REMAINING
T(I) = TIME WHEN Ith EVENT OCCURS
t(I) = T(I)-T(I-1)

BY THE RANDOM NUMBER STATEMENT,

$$E(I) = (LL^{(-B+1)} - [(B-1)*X(I)/A])^{(1/(B-1))} , \qquad 5.3$$

WHERE X(I) IS A RANDOM NUMBER DISTRIBUTED UNIFORMLY IN THE INTERVAL [0,1]. THE MAXIMUM LIKELIHOOD STATEMENT IS WRITTEN AS

$$LOG\ P(1,2,..) = SUM(I)\ [\ LOG\ A - B*LOG\ E(I) \qquad . \qquad 5.4$$

TABLE 5.1 SHOWS THE EFFECTIVENESS OF THE MAXIMUM LIKELIHOOD METHOD OF EXTRACTING THE EXPONENT B FOR THE DISTRIBUTION.

IF DATA IS BINNED, THE PROBABILTY P [E(I),E(I+1)] OF AN AN OBSERVED PARTICLE HAVING ENERGY IN THE "BIN" BETWEEN E(I) AND E(I+1) IS THE INTEGRAL OF THE NORMALIZED DISTRIBUTION FUNCTION BETWEEN THESE ENERGIES OR

$$P[E(I),E(I+1)] = (A/(B-1))[E(I)^{(-B+1)}-E(I+1)^{(-B+1)}]. \qquad 5.5$$

THE MAXIMUM LIKELIHOOD STATEMENT FOR A BINNED DATA IS

$$LOG\ P\ [1,2,3,...] = SUM(I)\ B(I)\ LOG\ (\ Q\ ) \qquad 5.6$$

WHERE

$$Q=([\ E(I)^{(-B+1)}-E(I+1)^{(-B+1)}]/[\ LL^{(-B+1)}-UL^{(-B+1)}]. \qquad 5.7$$

THE RESULTS OF SOME COMPUTER EXTRACTION OF EXPONENTS FOR POWER LAW SPECTRA ARE GIVEN IN TABLE 5.1. IT SHOULD BE NOTED THAT BINNING IS A WAY OF INTRODUCING UNCERTAINTIES IN THE DATA OR A WAY OF THROWING AWAY INFORMATION EITHER BECAUSE OF CONVENIENCE OR NECESSITY. THE RESULTS OF TABLE 5.2 ARE FOR BINS OF EQUAL SIZE. THE ENERGIES E(I) DEFINING BINS COULD HAVE BEEN CHOSEN BY GEOMETRIC PROGRESSION OR INTERVALS OF EQUAL PROBABILITY OR ANY OTHER WAY DESIRED.

AS A SUMMARY, WE CAN SAY THAT DATA SHOULD NOT BE BINNED UNLESS IT IS INESCAPABLE.

6. RANDOM ERROR AND SPECTRUM EXTRACTION

FOR VARIOUS REASONS, THE RESPONSE OF THE DETECTING INSTRUMENTS IS DIFFERENT FROM THE REAL ENERGY SPECTRUM. THE LEAST TROUBLESOME ERRORS ARE THOSE WHICH RESULT IN AN ENERGY UNCERTAINTY OVER A LIMITED RANGE. BINNING, FOR EXAMPLE, INTRODUCES AN UNCERTAINTY EQUAL TO THE BIN WIDTH, AND WE CAN SIMULATE QUITE EASILY WHAT BINNING DOES TO THE SPECTRUM EXTRACTION. GIVEN A BIN DISTRIBUTION AND THE ABILITY TO SIMULATE THE EFFECT OF BINNING ON A GIVEN SPECTRUM, WE CAN ALWAYS FIND A SPECTRUM WHICH REPRODUCES, WITHIN THE VAGARIES OF STATISTICS, THE BIN DISTRIBUTION. SOME INSTRUMENTS WILL DETECT PARTICLES OF A GIVEN ENERGY E(I) AND INDICATE A DISTRIBUTION OF ENERGIES E(J) SUCH THAT $0 < E(J) < E(I)$. THE PROBLEM OF EXTRACTING THE ORIGINAL SPECTRUM THEN BECOMES DIFFICULT OR IMPOSSIBLE.

TABLE 5.1

MAXIMUM LIKELIHOOD TREATMENT OF POWER DISTRIBUTIONS

SEED = 21212 LL=1 UL=5 EXP = -2.7		SEED = 21212 LL=1 UL=100 EXP = -2.7		SEED = 21212 LL=1 UL=1000 EXP = -2.7	
EVENTS	EXP	EVENTS	EXP	EVENTS	EXP
10	-2.3	10	-2.7	10	-3.0
20	-2.9	20	-3.0	20	-3.1
30	-2.9	30	-2.9	30	-2.9
40	-2.7	50	-2.8	50	-2.7
50	-2.8	100	-2.8	100	-2.6
100	-2.7	200	-2.9	200	-2.6
		400	-2.9	500	-2.7

TABLE 5.2

RESULTS OF MAXIMUM LIKELIHOOD TREATMENT OF BINNING ERRORS

EVENTS	LL	UL	BINS	EXP(GIVEN)	EXP(FIT)
30	1	2	4	-2.7	-3.1
50	1	3	5	-2.7	-3.2
100	1	3	10	-2.7	-3.15
200	1	3	20	-2.7	-2.87
500	1	5	20	-2.7	-2.82
500	1	5	40	-2.7	-2.81
500	1	10	20	-2.7	-2.79
100	1	20	40	-2.7	-2.52

AS A FIRST EXERCISE WE WILL LOOK AT THE EFFECT OF A MEASURING UNCERTAINTY WHICH IS PROPORTIONAL TO ENERGY, NAMELY,

$$E(F) = E(I)*(1 + f*(0.5-RND)) \quad , \qquad 6.1$$

WHERE E(I) IS THE INCOMING ENERGY, E(F) IS THE READING FROM THE INSTRUMENT, f REPRESENTS A RELATIVE SPREAD AND RND IS A RANDOM NUMBER UNIFORMLY DISTRIBUTED BETWEEN ZERO AND ONE. FOR EXAMPLE, IF f = 0.1 THEN 0.95*E(I) < E(F) < 1.05*E(I). TABLE 6.1 SHOWS THE EFFECT OF INTRODUCING A RANDOM ERROR SUCH AS INDICATED IN EQ 6.1. IT IS SEEN THAT A RATHER LARGE ERROR OF THIS TYPE CAN BE TOLERATED WITHOUT SERIOUS DEGRADATION OF OUR ABILITY TO EXTRACT A SPECTRUM.

TO ILLUSTRATE THE SECOND KIND OF ERROR, WE IMAGINE THAT THE INCOMING ENERGY E IS DEGRADED TO A READING E' ACCORDING TO THE DISTRIBUTION,

$$P(E,E') = A*(E^(-B))*A'*EXP(-(E-E')/(h*E) \quad , \qquad 6.2$$

WHERE A*E(-B) REPRESENTS THE INCOMING SPECTRUM AND THE REMAINING FACTOR REPRESENTS THE DEGRADATION OF THE INSTRUMENT. THE PRODUCT h*E IS THE AVERAGE DEGRADATION E-E'. THE RESULT OF USING THIS DISTRIBUTION IS SHOWN IN TABLE 6.2. IT IS SEEN THAT THIS TYPE OF ERROR MAKES THE SPECTRUM LOOK STEEPER BY TRANSFERRING EVENTS FROM HIGHER TO LOWER ENERGIES.

7. COSMIC RAY DATA FROM CERENKOV COUNTERS

WHEN A HIGH ENERGY NUCLEUS WITH SPEED V AND ATOMIC NUMBER Z TRAVERSES A REFRACTIVE MEDIUM WITH REFRACTIVE INDEX N, A LIGHT PULSE IS GENERATED ACCORDING TO THE FORMULA,

$$L = K*(Z^2)*(1 - (C/(V*N))^2) \quad , \qquad 7.1$$

WHERE C IS THE SPEED OF LIGHT IN VACUUM AND K IS A CONSTANT. IT IS SEEN FROM EQ 7.1 THAT L(MAX), THE MAXIMUM LIGHT PULSE AVAILABLE WHEN V/C = 1, IS GIVEN AS

$$L(MAX) = K*(Z^2)/(1 - 1/N^2) \quad . \qquad 7.2$$

THE INDEX OF REFRACTION N OF THE GAS IN THE COUNTER WAS 1.00115 SO THAT L(MAX)/(K*(Z^2)) = 435.53 .

THE CERENKOV COUNTER ACTS TO CONVERT THE ENERGY SPECTRUM,

$$N(E) = A * E^(-B) \quad , \qquad 7.3$$

TO THE PULSE HEIGHT SPECTRUM P(L) WHERE

$$N(E)\, dE = P(L)\, dL \quad , \qquad 7.4$$

TABLE 6.1

EFFECT OF SIMPLE ENERGY UNCERTAINTY ON SPECTRUM EXTRACTION *

$P(E) = A\ E^{(-B)}$ B = 2.7 LL = 1 UL =5

$E(F) = E(I)*(1 + h*(0.5 - RND))$

EVENTS	ERROR PARAMETER h	EXTRACTED B
10	0.1	2.62
10	0.2	2.59
10	0.3	2.57
10	0.8	2.57
10	0.9	2.63
10	1.9	3.84

TABLE 6.2

EFFECT OF EXPONENTIAL DEGRADATION ON SPECTRUM EXTRACTION

$P(E) = A\ E^{(-B)}$ B = 2.7 LL = 1 UL = 5

$P(E,E') = A*E^{(-B)}*A'EXP\ [-(E-E')/(h*E)]$

EVENTS	ERROR PARAMETER h	EXTRACTED B
10	0.01	2.31
10	0.10	3.04
10	0.20	4.53
10	0.30	9.90
10	0.50	****
30	0.001	3.00
30	0.01	3.09
30	0.05	3.52
30	0.10	4.36
30	0.20	****

**** PROGRAM FAILED TO RUN

OR

$$P(L) = N(E)\ (dE/dL) \quad . \qquad 7.5$$

THE KINETIC ENERGY E OF A PARTICLE OF MASS M IS GIVEN BY THE EXPRESSION,

$$E = (GAMMA - 1)*M*C^2 \quad , \qquad 7.6$$

WHERE

$$GAMMA = SQR(\ 1\ /\ (1-BETA^2)\ )\ ;\ BETA = V/C \quad , \qquad 7.7$$

AND C IS THE SPEED OF LIGHT. IT IS CONVENIENT TO EXPRESS ENERGY IN UNITS OF MC^2 SO THAT E = GAMMA-1. AFTER A LITTLE ALGEBRA WE FIND THAT

$$E = [\ X/(X-1)\ ] - 1 \quad , \qquad 7.8$$

WHERE

$$X = (N^2)*(\ 1 - (L/K)\ ) \quad . \qquad 7.9$$

USING THESE DEFINITIONS,

$$dE/dL = [(N^2)/(2*K)]*[\ X*(X-1)^3\ ]^(-1/2) \qquad 7.10$$

THE NORMALIZATION CONSTANT A IN EQ 7.3 IS FOUND BY DIRECT INTEGRATION OF EQ 7.3 TO BE

$$A = (B-1)/[\ LL^(-B+1) - UL^(-B+1)\ ] \quad , \qquad 7.11$$

WHERE LL IS THE LOWER ENERGY LIMIT CONSIDERED FOR THE DISTRIBUTION AND UL IS THE UPPER LIMIT. PUTTING EQS 7.3, 7.8, 7.10, AND 7.11 TOGETHER, WE FIND AFTER A LITTLE ALGEBRA THAT

$$P(L) = [(A*N^2)/(2*K)]*[1/(SQR(X)-SQR(1-X))]^B * SQR(1/X) * (X-1)^((B-3)/2) \qquad 7.12$$

THIS DISTRIBUTION IN PULSE HEIGHT IS RATHER FLAT WHEN B=2.9, IS MONOTONIC UPWARD FOR B < 2.9 AND IS MONATONIC DOWNWARD FOR B > 2.9. THE (B-3) EXPONENT IN THE LAST FACTOR OF EQ 7.12 IS LARGELY RESPONSIBLE FOR THIS BEHAVIOIR. AN IMPORTANT FEATURE OF THESE CURVES IS THAT, AS B DECREASES FROM 3, THE STACKING OF THE DATA POINTS JUST BELOW L(MAX) BECOMES MORE PRONOUNCED. FOR B > 3 , P(L) IS RATHER EVENLY MONATONIC DOWNWARD FROM BEGINNING TO END.

THE PROBABILITY OF ONE COUNT WITH L(I) < L < L(I+1) IS P(L(I)) SO THAT THE PROBABILITY OF N(I) COUNTS IN THE INTERVAL IS P(L(I))^N(I). A GOOD MAXIMUM LIKELIHOOD FUNCTION P IS GIVEN AS

$$P = PRODUCT(I)\ P(L(I))^N(I) \quad , \qquad 7.13$$

OR

$$\text{LOG } P = \text{SUM(I) } N(I) * \text{LOG[P(L(I))]} \quad . \qquad 7.14$$

THIS FUNCTION IS EASILY MAXIMIZED BY THE METHODS DEVELOPED IN THE ATTACHED COMPUTER PROGRAMS. IF WE DO NOT BIN THE DATA, A CONVENIENT MAXIMUM LIKELIHOOD FUNCTION IS GIVEN AS

$$\text{LOG } P = \text{SUM(I) LOG P[L(I)]} \qquad 7.15$$

IF WE LET B AND L(MAX) BE VARIABLES, WHERE L(MAX) MUST BE GREATER THAN ANY PULSE HEIGHT L(I) USED, WE CAN FIND B AND L(MAX) WHICH MAXIMIZE LOG P FOR THE DATA. ANY ANSWER $2 < B < 4$ IS POSSIBLE FOR SMALL VARIATIONS IN L(MAX). THE BEST AND MOST BELIEVABLE RESULTS WERE OBTAINED BY MAXIMIZING [LOG(P)]/N WITH B AND L(MAX) AS VARIABLES, WHERE N IS THE TOTAL NUMBER OF PARTICLES USED IN THE MAXIMIZATION. IT SHOULD BE NOTED THAT P(L) IS RATHER FLAT AND THAT ANY FUNCTION WHICH RANDOMLY CHANGES THE SIZE OF THE LIGHT PULSES WILL NOT CHANGE TO ANY APPRECIABLE EXTENT THE SHAPE OF P(L) PREDICTED BY EQUATION 7.12, EXCEPT AT THE ENDS. DATA WITH $L(I) > L(MAX)$ IS REFLECTED THROUGH L(MAX) TO A LOWER ENERGY $2*L(MAX)-L(I)$. THE ARGUMENT IN FAVOR OF THIS IS AS FOLLOWS. IF WE IMPOSE TO THE RIGHT OF P(L) A MIRROR IMAGE P'(L), THE PROBABILITIES EXCHANGED BETWEEN THE TWO DISTRIBUTIONS ARE LIKEWISE MIRROR IMAGES. SINCE EXCHANGES ARE EQUAL, THE LEAKAGES AT THE BOUNDARIES ARE SUFFICIENT TO PRESERVE P(L) IF THEY ARE REFLECTED BACK IN.

THE RESULTS OF TREATING CARBON AND IRON COSMIC RAYS BY THE ABOVE MAXIMUM LIKELIHOOD TREATMENT ARE SHOWN IN FIG 7.1 AND TABLE 7.1.

8 SUMMARY

SEVERAL USEFUL TECHNIQUES WERE DEVELOPED TO EXTRACT PARAMETERS FROM DATA ON RANDOM EVENTS. THE BEST OF THEM WERE BASED ON MAXIMUM LIKELIHOOD STATEMENTS. THESE METHODS PROVED SUCCESSFUL IN TREATING DATA FROM COSMIC RAYS FOR THE CASES OF CARBON AND IRON NUCLEI.

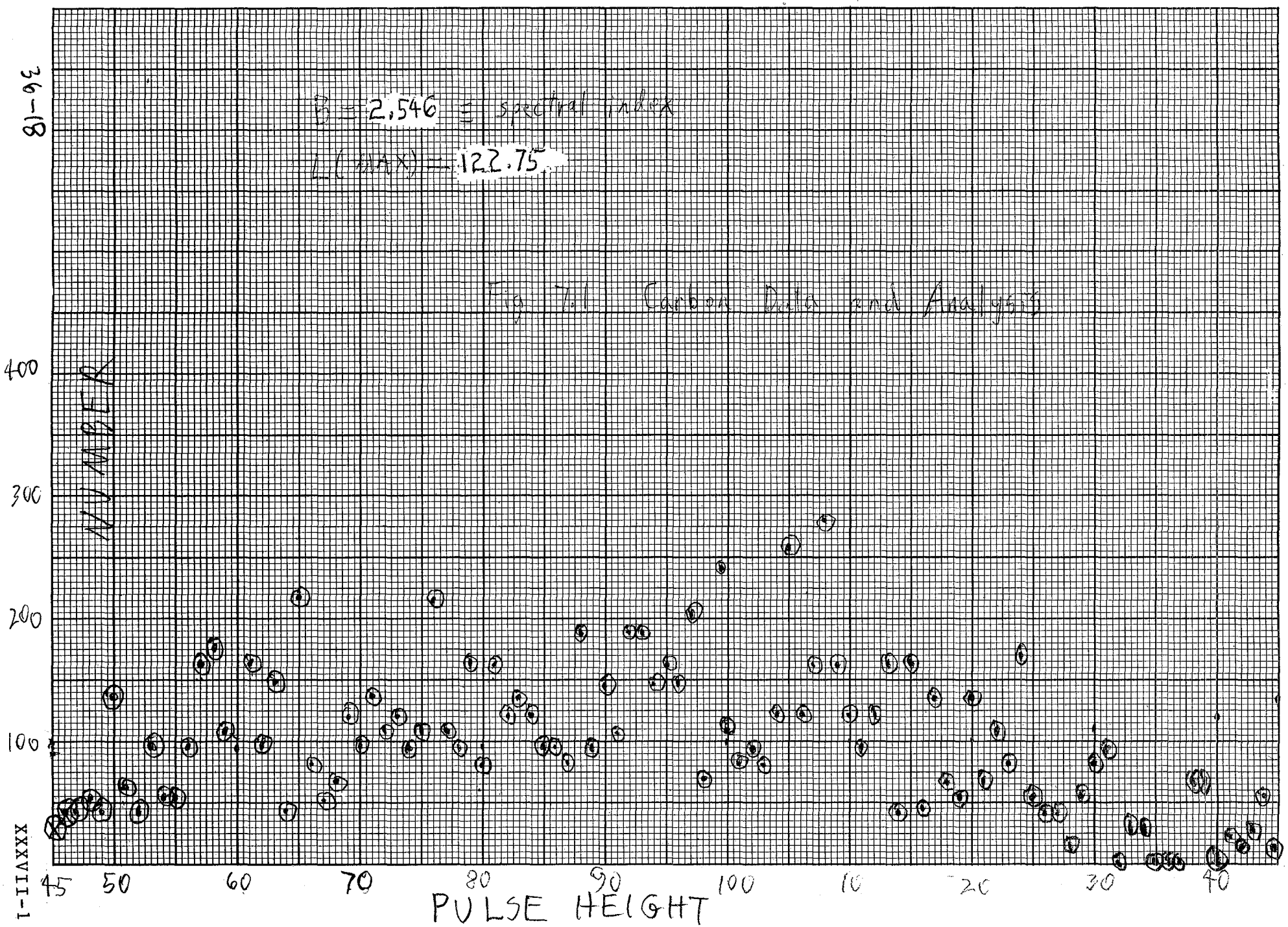

B = 2,546 = spectral index
L (MAX) = 122.75
Fig. 7.1 Carbon Data and Analysis
81-93
400
300
200
100
45
50
60
70
80
90
100
10
20
30
40
PULSE HEIGHT
XXXVII-1

TABLE 7.1

ANALYSIS OF DATA ON IRON NUCLEI

NUMBER OF DATA POINTS 82

STARTING SPECTRAL INDEX FOR THE MAXIMUM LIKELIHOOD ITERATION 2.7

STARTING L(MAX) 2114

FINAL SPECTRAL INDEX 2.0938

FINAL L(MAX) 2022.5

```
10 ' PROGRAM TO FIT P(L) TO IRON DATA UL=L(MAX)-1 L(I) > L(MAX) REFLECTED
20 ' B:NASA48D.BAS   R. D. SHELTON 29 JULY 85 09:32
30 CLS
40 DIM L(200),LL(200),P#(200),E#(200)
50 DATA 1230,767,1937,1950,1986,1628,2466,1367,2079,2243
60 DATA 1974,1895,2107,1968,2195,1965,468,2023,1602,1958
70 DATA 632,1217,1018,2034,1494,911,1207,1948,1880,1947
80 DATA 1828,2003,944,901,1990,1965,1805,1070,808,1463
90 DATA 1647,1370,625,640,1419,484,1951,662,613
100 DATA 1663,1264,1330,2103,1726,1125,2324,630,2468,1133
110 DATA 642,1822,2176,1958,1154,754,1841,1738,2259,977
120 DATA 1627,1538,1226,1730,2034,1874,2438,1448,812,2242
130 DATA 1590,852,2068
140 SUM#=0
150 P$="#####.####" :PP$="### "
160 FOR I=1 TO 82
170       READ LL(I)
180       ' SUM#=SUM#+LL#(I)
190       ' PRINT I,LL(I):IF I MOD 20 = 0 THEN INPUT Z$
200 NEXT I
210 FOR I=1 TO 82
220      FOR J=I TO 82
230               IF LL(I)>LL(J) THEN SWAP LL(I),LL(J)
240      NEXT J
250      IF I MOD 10 =0 THEN PRINT I;
260 NEXT I
270 ' FOR I=1 TO 82
280 '     LPRINT I;"  ";LL(I)
290 ' NEXT I
300      ' PREPARE FOR ITERATION
310 CLS:PRINT
320 PRINT
330    MM=1:BB=1:IT=1:PRINT:INPUT " ENTER STARTING EXPONENT    ",M#
340 PRINT
350    INPUT " ENTER L(MAX)    ",B#
370    ET#=1.00115:ET2#=ET#^2
380    PRINT:INPUT " ENTER STEP FOR EXPONENT  ",Z$
390    IF Z$<>"" THEN MM#=VAL(Z$)
400    PRINT:INPUT " ENTER STEP FOR L(MAX)   ",Z$:PRINT
410    IF Z$<>"" THEN HB#=VAL(Z$)
420    IF IT>1 THEN GOTO 460
430    GOSUB 610
440    S1#=S#
450    'ITERATION SCHEME
460       GOSUB 810
470       GOSUB 940
480       LOOK=IT MOD 10:PRINT
490       IF LOOK<>0 THEN GOTO 460 ELSE GOTO 500
500 PRINT
510 INPUT "CHOICE ? CR-GO ;C-CHANGE STEP ;P-LPRINT LL(I),L(I);QQ-QUIT ",Z$
520       PRINT
530       IF Z$="QQ" THEN PRINT I;" ";IT;" EXP =  ";M#;"   L(MAX) =   ";B#
540       IF Z$="C" THEN GOTO 380
550       IF Z$="P" THEN GOTO 970
560       IF Z$="QQ" THEN GOTO 600
580       IF Z$="QQ" THEN INPUT Z$
590    GOTO 460                                          XXXVII-17
600 END
610  'SBR TO COMPUTE SUMS FOR ITERATION
620 S#=0:LL=468:K#=B#/(1-1/ET2#)
630 FOR JJ=1 TO 82
640      IF LL(JJ)<=B#-.5 THEN L(JJ)=LL(JJ) ELSE L(JJ)=2*B#-LL(JJ)
650 NEXT JJ
660 X#=ET2#*(1-LL/K#):ELL#=SQR(X#/(X#-1))-1
670 X#=ET2#*(1-(B#-.5)/K#):EUL#=SQR(X#/(X#-1))-1
```

```
680 A#=(M#-1)/(ELL#^(-M#+1)-EUL#^(-M#+1))
690 ' PRINT ELL#,EUL#,A#:INPUT Z$
700 ' PRINT B#,IND,K#,LL,UL,ELL#,EUL#:INPUT Z$
710 FOR JJ=1 TO 82
720     IF L(JJ)<=(B#-.5) THEN X#=ET2#*(1-L(JJ)/K#) ELSE X#=ET2#*(1-(B#-1)/K#)
730     E#=SQR(X#/(X#-1))-1
740     DE#=(ET2#/(2*K#))*(X#*(X#-1)^3)^(-1/2)
750     F#=A#*(E#^-M#)*DE#
760     IF P#<=.000001 THEN PRINT "JJ,L(JJ),A#,B#,E#,M#,DE#  ";JJ,L(JJ),A#,B#,E#
,M#,DE#
770     S#=S#+LOG(F#)/82
780     ' LPRINT USING PP$;JJ;:LPRINT USING P$;P#,L#(JJ)   ,S#
790 NEXT JJ
800 RETURN
810 ' SBR TO ITERATE FOR FINDING MINIMUM OF S
820 IT=IT+1:ZM#=M#:M#=M#+MM*HM#:ZZM#=M#:SS1#=S1#
830 GOSUB 610
840 IF S#<S1# THEN MM=-MM
850 IF S#<S1# THEN HM#=HM#/2
860 M#=ZM#:ZB#=B#:B#=B#+BB*HB#:ZZB#=B#
870 GOSUB 610
880 IF S#<S1# THEN BB=-BB
890 IF S#<S1# THEN HB#=HB#/2
900 M#=ZZM#:B#=ZZB#:S1#=S#
910 GOSUB 610
920 S1#=S#
930 RETURN
940 'PRINT SUBROUTINE FOR ITERATION
950 PRINT USING PP$;IT;:PRINT USING P$;100*SS1#;100*S1#;M#;B#;HM#;HB#;:PRINT USI
NG PP$;MM;BB
960 RETURN
970 FOR I=1 TO 82
980 LPRINT I,LL(I),L(I)
990 NEXT I
1000 GOTO 510
```

1985

NASA/ASEE SUMMER FACULTY RESEARCH FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

AUTOMATED CALIBRATION OF A FLIGHT PARTICLE SPECTROMETER

Prepared By:	Roy B. Torbert, Ph.D.
Academic Rank:	Associate Research Physicist
University and Department:	University of California San Diego, California Center for Astrophysics and Space Science
NASA/MSFC: Division: Branch:	Solar-Terrestrial Magnetospheric
MSFC Counterpart:	Thomas Moore
Date:	December 3, 1985
Contract No.:	NGT 01-008-021 The University of Alabama in Huntsville

AUTOMATED CALIBRATION OF A FLIGHT PARTICLE SPECTROMETER

BY

Roy B. Torbert
Associate Research Physicist
University of California
San Diego, California

ABSTRACT

A system for calibrating both electron and ion imaging particle spectrometers was devised to calibrate flight instruments in a large vacuum facility in the Space Science Laboratory at MSFC. An IBM-compatible computer was used to control, via an IEEE 488 buss protocol, a two-axis gimbled table, constructed to fit inside the tank. Test settings of various diagnostic voltages were also acquired via the buss. These spectrometers constructed by the author at UCSD were calibrated in an automatic procedure programmed on the small computer. Data was up-loaded to the SSL VAX where a program was developed to plot the results.

Introduction

During the summer of 1985, the proposed automated calibration system was designed, constructed, and tested in the vacuum facility at MSFC. Due to failures of the vacuum vessel itself, no actual calibration was completed of the sounding rocket instrument. Fortunately, NASA delayed sounding rocket flight 35.014 for totally different reasons associated with payload weight and apogee requirements. The instrument will go back into the chamber in December for final calibration.

While waiting for chamber repairs, I devoted considerable time to a computer model of the Critical Velocity Effect , which the sounding rocket experiment will investigate. The first results of this work will be described.

Objectives

The objective of this summer's study was to construct an automated calibration system for instruments with many simultaneous channels. The instrument that I developed at UCSD can sample twenty different pitch angles at one time and do this for up to 64 different energies in 10 milliseconds. The manual calibration of such an instrument would exhaust at least 20 graduate students. We proposed to utilize the automatic gimbal table at MSFC for orienting the detector under computer control within the ion beam from a local source. Data would be collected by a ground station computer and transmitted to the SSL VAX 11/780 computer for analysis and display.

Calibration System

Figure 1 presents a block diagram of the automated calibration system that was constructed. Figure 2 is a photograph of the system being tested on the workbench before check-out in the chamber. The system host is a Compaq portable computer, which also serves as the ground station computer for the rocket instrument computer. A RS-232 link was used to transfer data files accumulated during calibration runs. This link was tested with dummy data and a student programmer produced line printer plots of the dummy data. I purchased a QuaTech MXI-241 plug-in board for the Compaq, which provide both a versatile 8255 peripheral interface and the required IEEE-488 port. Software was written that drove the stepper motors on the three-axis gimbal table, on which the detector was mounted in the vacuum chamber. An ion source provides a known ion species with a precise energy center and width, directed at the detector. The software listed in Figure 3, was written to control the angular orientation of the table. Other software was created to write controlled DC voltage sources that we would use to construct a analyzing high voltage for the detector itself and to provide diagnostic and labeling data for the particular calibration mode.

A custom bi-directional buss was designed to be controlled by the 8255 port on the QuaTech board. This allowed us to use the actual flight accumulator board to sample the count rate from up to 32 channels of the detector simultaneously. It also greatly simplifies the interface between the calibration system

and the detector since this is exactly the flight configuration. Actual counter data was collected on the single pump down cycle that we were able to make before chamber failure. However, this data was later determined to be in error due to an electronic problem in our buss set-up. This problem was corrected and remains to be tested when we have the chamber again.

In summary, when the chamber is available again in December, the true test of this system will begin. Since the new launch date for 35.014 is May of 1986, the pressure still exists for a rapid calibration and we hope to be finished by March of that year.

Computer Simulation of Critical Velocity Effect

The project which is motivating the designed instrument and its calibration is fully described in the appendix. Alfven proposed in 1954 that, if a neutral gas flows at a velocity, perpendicular to the magnetic field, greater than a critical velocity,Vc,

$$V_c = \sqrt{2e\Phi_{ion}/M}$$

then the gas would be anomalously ionized. The sounding rocket experiment will attempt to create these conditions in space. I undertook work to simulate the release of these neutrals from barium shaped charge releases. The following are the essential new features of this work.

First, the barium neutral density at a distance R from an instantaneous injection of particles with a velocity distribution, dN/dV, and angular distribution around a central polar angle, , is given by :

$$n(R) = \left(V \frac{dN}{dV}\right)_{V=R/t} \frac{1}{R^3} f(\theta)$$

where:

$$\int f(\theta)\, d\Omega = 1$$

The rapid fall-off of this density with R is a prime factor limiting the discharge, since the ionization rate and thereby the available free energy density are proportional to it. This later is given by:

$$\dot{N}_e (\Delta V)^2 \frac{M}{2}$$

where (ΔV) is the relative perpendicular velocity of the neutrals and the ambient plasma. This velocity, ΔV , may be deduced from equating the current that results from ionization(i.e. due to dNe/dt) to that which can be carried off in Alfven waves by the background plasma. If more current were generated by the burning process that this, then the cloud of neutrals and plasma would electrically polarize and then decrease the relative drift of the neutrals and plasma, thereby

shutting off the effect. This provides a natural limit to the amount that can be ionized. Using this approach, we derived :

$$\Delta U = V_{Neutral} / (1 + \dot{N}_e M \ell_{II} / 2 \rho_A V_A)$$

Now, the ionization rate was computed using the following form of the impact ionization cross-section, which can be derived using the Born approximation:

$$\sigma(U) \propto \frac{1}{U^2} \ln(U^2)$$

Assuming an electron Maxwellian distribution with density,Ne, and temperature, Te, we derive the ionization rate:

$$\dot{N}_e = \sigma_{MAX} U_e N_n N_e Z(\phi_i / T_e) + Q \cdot N_n$$

where the function $Z(\phi_i / Te)$ is given by:

$$Z(\lambda) = \frac{8}{\sqrt{2\pi}} \lambda \left[Ei(\lambda) + \emptyset.307 e^{-\lambda} \right]$$

The factor, Q, allows ionization by photoionization or some other 'free' source such as impact ionization from a high energy auroral electron beam. It could also model charge exchange. This is a critical factor since it provides the `seed' for the discharge to occur.

Finally, the total energy balance for this electron population may be computed as follows:

$$\frac{d}{dt}(3/2 N_e \bar{T}_e) = \eta \left[\dot{N}_e (\Delta U)^2 M/2 \right] - \dot{N}_e \phi_{ion} - \alpha_{Ba} N_e N_n \phi_B - \alpha_O N_e N_O \phi_O$$

where the positive term is the above energy input and the negative terms are due to energy loss from the impact ionization itself and from electron excitation of barium and the background oxygen.

All the ingredients of the simulation are now in place and

the computer was used to compute the evolution of Ne and Te at successive R's along the beam path. A typical profile that resulted is given in Figure 4, where these parameters are plotted versus time for an injection distance of 2 km. Clearly, the electrons are heated and the population grows. However, we never found the large yield that Haerendel and colleagues claim in their earlier work (Haerendel,1982). We are presently working to revise the model to include convection effects and a better charge exchange model.

Conclusions and Recommendations

The calibration project should and will continue and, perhaps during the second year of this project, will be functional as a reliable and useful facility for calibrating space-flight particle spectrometers.

References

1. Alfven,H.,On the Origin of the Solar System,Oxford at the Clarendon Press,London,1954.

2. Haerendel,G.,Alfven's Critical Velocity Effect Tested in Space,Z.Naturforsch,A.37,728 ,1982.

Automated Calibration System

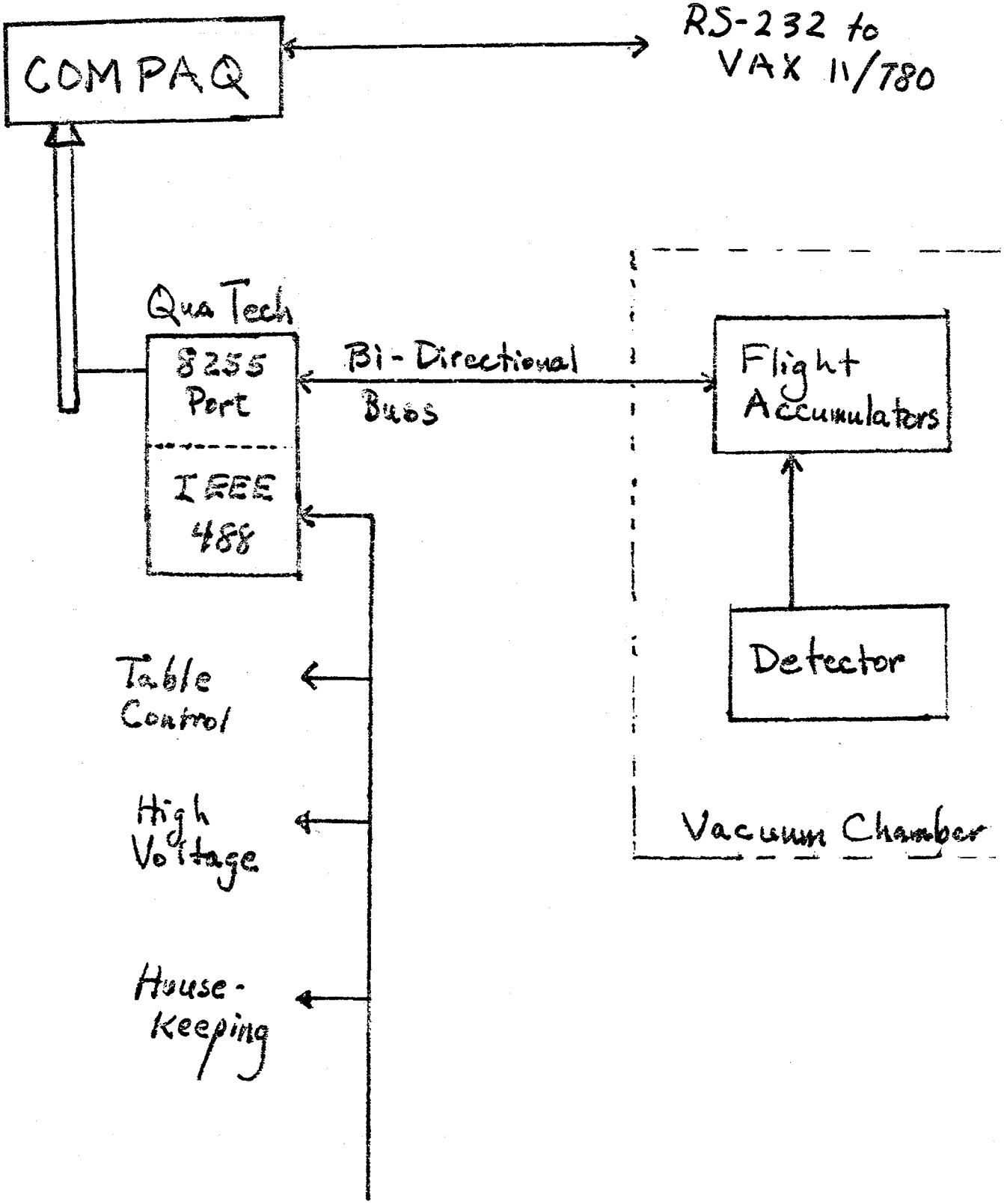


Figure 1

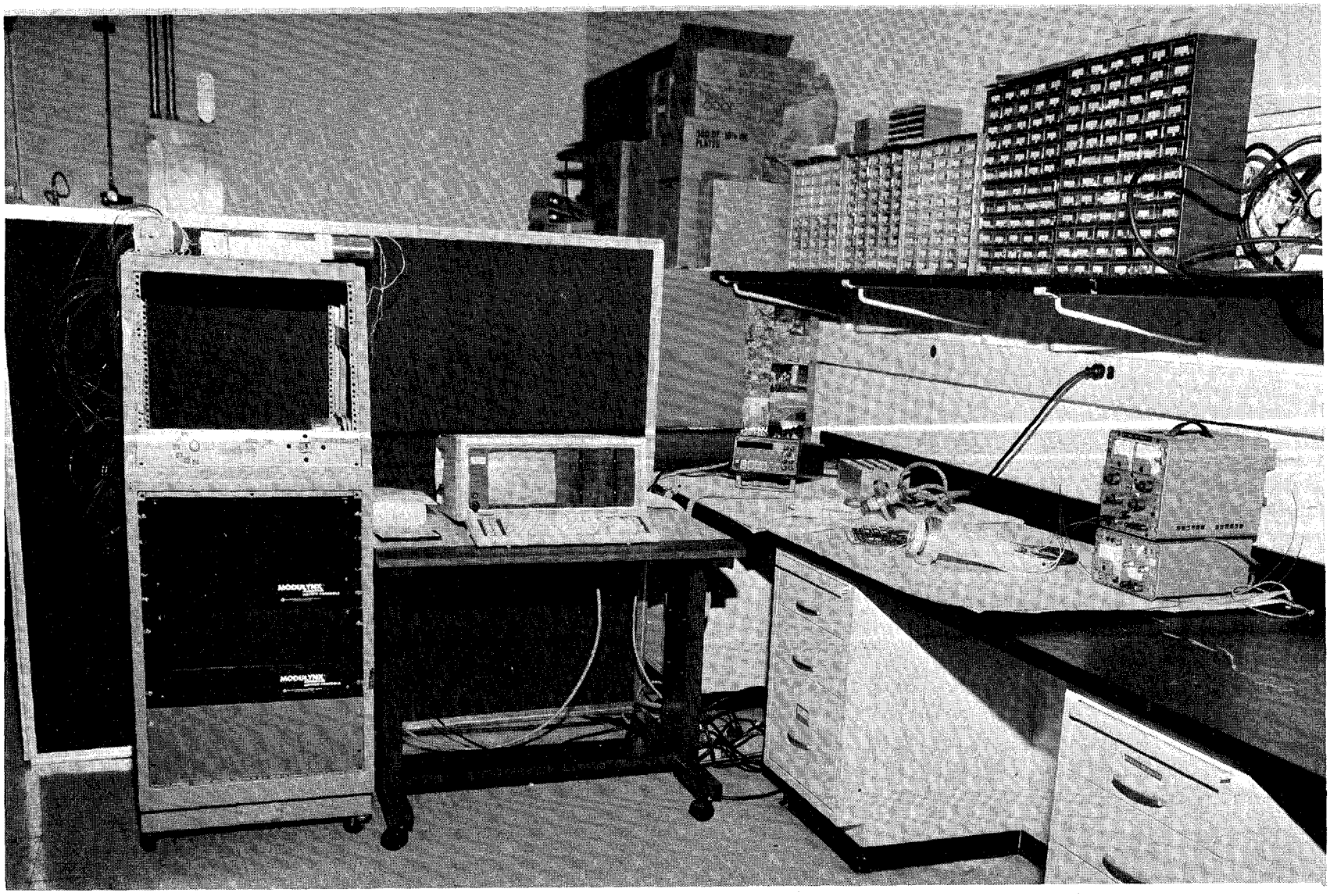


Figure 2

```
10 BUSSTA=&H4A5:CLR=&H9F0:DCL=&H388:EOI=&H2A5:GTL=&H482
20 GTS=&H295:IFC=&H2B9:INIT=&H308:LISTEN=&H6C7
30 LLO=&H3A8:LON=&H2A9:PASSCONT=&H5DB:PPDS=&H4B7
40 PPEN=&H3D8:PPOL=&H56E:PPU=&H3B8:READS=&H7CE
50 READB=&H7B0:RECECONT=&H617:REMOTE=&H6A0:SDC=&H398
60 SPOL=&H4D9:TALK=&H775:TCA=&H235:TCS=&H285
70 TON=&H2B1:TRIG=&H493:ULON=&H2AD:UNL=&H34D
80 UNT=&H378:UTON=&H2B5:WRITES=&H89D:WRITEB=&H88D
90 XFER=&H969
100 DEF SEG = 0
110 CSEG=(256*PEEK(&H3C7))+PEEK(&H3C6)
120 DEF SEG=CSEG
130 RDATA$="+2000,P,C,X49406                    "
140 SYSCON%=1
150 I488ADDR%=&H338
160 DEVADDR%=1
170 CALL INIT(SYSCON%,I488ADDR%,DEVADDR%)
180 CALL REMOTE
190 CALL LLO
200 COMM$="XD"        'TURN ON POWER TO X-MOTOR
210 EOI$="10"
220 LISTEN$="24"
230 CALL WRITES(LISTEN$,EOI$,COMM$)
240 COMM$="<CPT>"        'GIVE STATUS OF INTIAL X-AXIS POSITION
250 CALL WRITES(LISTEN$,EOI$,COMM$)
260 TALK$="24"
270 LISTEN$=""
280 CALL READS(TALK$,LISTEN$,EOI$,RDATA$,STATUS%)
290 PRINT "RDATA",RDATA$
300 LISTEN$="24"
310 TALK$=""
320 COMM$="XM1000G;"         'MOVE MOTOR
330 CALL WRITES(LISTEN$,EOI$,COMM$)
340 COMM$="<CPT>"        'GIVE STATUS OF X-AXIS AFTER MOVEMENT
350 CALL WRITES(LISTEN$,EOI$,COMM$)
360 TALK$="24"
370 LISTEN$=""
380 CALL READS(TALK$,LISTEN$,EOI$,RDATA$,STATUS%)
390 PRINT "RDATA",RDATA$
400 COMM$="XE;YE;ZE;UE;"         'TURN OFF POWER FOR ALL AXES
410 LISTEN$="24"
420 TALK$=""
430 CALL WRITES(LISTEN$,EOI$,COMM$)
440 CALL GTL
450 CALL UNL
460 STOP
```

Figure 3

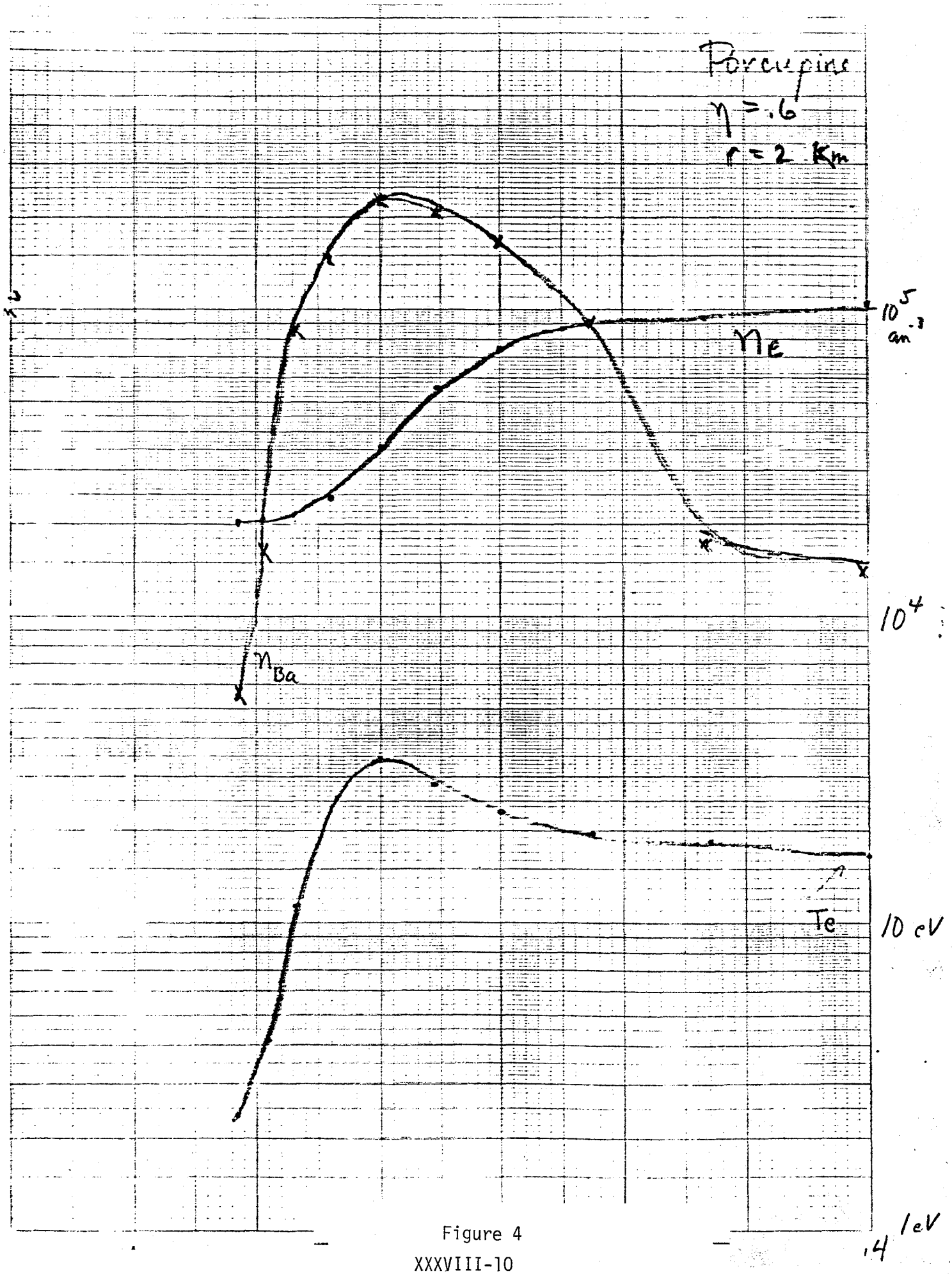


Figure 4

Appendix

Critical Velocity Experiments

Two complete sounding rocket experiments are planned for flight from Wallops Flight Facility in October, 1985, to study the Critical Velocity Effect in space plasmas. The international research team consists of investigators from University of California, San Diego (R. B. Torbert, P.I.), Cornell University (M. C. Kelley, P.I.), University of Alaska (E. M. Wescott, P.I.), Max Planck Institute for Extraterrestrial Physics, Garching, West Germany (G. Haerendel, P.I.), Royal Institute of Technology (C.-G. Falthammar, P.I.), and Utah State University (C. Howlett and J. Foster , P.Is.).

In 1954, Alfven proposed that, if an element in a nearly neutral plasma was ionized when it attained a flow velocity which depended on its ionization potential, then several facets of the structure of the solar system could be explained. With remarkable conviction, he maintained that there should therefore be a 'Critical Velocity effect,' whereby a neutral atom of mass M drifting perpendicular to the ambient magnetic field at a velocity greater than

$$Vcr = \mathrm{sqrt}(2e\ \phi_{ion}/M)$$

would be anamolously ionized.

At the time, this speculation was greeted with some skepticism, since the classical cross-section for impact ionization at this velocity was exactly zero and only reached some reasonable probability at very large energies. However, a series of laboratory experiments showed that just such an effect occurs in plasma fusion devices (for a review, see Danielsson, 1973). The best explanations of what occurs all go something like this. Imagine, for some reason, that a few neutrals are ionized. This could be due to photoionization or some residual thermal ionization. The resulting ions, traveling with a large velocity perpendicular to B, can easily be shown to create lower hybrid waves, a type of plasma wave with a large phase velocity parallel to B. Electrons are resonantly accelerated along B in this wave to energies of the order of the ionization potential of the original neutral. These electrons have a large cross-section for ionizing the remaining neutrals. The effect snowballs into a discharge as more and more ions produce more and more hot electrons and so on. This result aroused the interest not only of astrophysicists, but also fusion physicists who needed to heat plasmas in their machines.

In applying these results to the case of the solar system, two difficulties arise. Namely, although its clever that somehow the energy of the ion (poor at ionizing) gets transfered to the electron (great at ionizing), what happened to the momentum? The standard answer in the laboratory is that the walls of the device absorb the momentum. There are no walls in space, however. Secondly, what keeps the hot electrons from leaving the interaction region? Again, the sheath fields near the walls accomplish this in the laboratory. Since there are many other possible applications in space besides Alfven's theory of the solar system (the formation of the inner coma cometary plasma [Formisano et. al., 1981], and the ionization of the plasma torus of Io

in Jupiter's magnetosphere, [Galeev and Chabibrachmanov, 1983]) it is of great interest to determine the exact conditions under which the effect might occur in space. The best resolution to these problems was to directly test the hypothesis in space, using chemical release experiments where the ejection velocities of around nearly 10 km/sec exceed the critical velocity of barium (2.7 km/sec) and strontium (3.5 km/sec).

Although there had been tantalizing hints of some sort of prompt ionization in previous barium release experiments (Deehr, et. al., 1982), the first conclusive evidence was reported from one of the Porcupine flights by G. Haerendel (1982). In this experiment, the shaped-charge barium beam traveled upwards at an angle of 28 degrees to the magnetic field at a height of 450km, which was below ionizing UV radiation. Any barium ionized while still below the uv terminator would travel up along field lines and appear above the terminator at the horizontal distance that ionization occured. Figure 1 shows a densitometer trace taken just after the cloud emerged into sunlight. A heavy ionization in the first few seconds after the explosion (while the cloud was still in darkness) is indicated by the ionization peak extending out to 10 or 15km from the explosion. Haerendel estimates approximately 30% of the barium cloud which had $V_{perp} > Vc$, was ionized. Haerendel also showed how the background plasma was capable of absorbing the lost momentum of the ions by carrying away Alfven waves from the beam region. Electric field obsevations from a remote payload confirmed his hypothesis.

To directly measure the plasma properties of the discharge, a new set of experiments, flown from Peru, were designed to both provide a stringent test of the critical velocity hypothesis, and to determine the range of pitch angles over which the critical velocity operated. In the Star of Condor (34.009) experiment ,a strontium radial shaped-charge was placed in the plane of the magnetic field lines, so that all ranges of pitch angles, from parallel to B to perpendicular to B were covered. The limiting value of perpendicular velocity for which the effect was observed, could have been used as a measure of the efficiency with which the kinetic energy of the newly created ions is transferred to the electrons. Strontium does not ionize in sunlight; hence any appreciable ionization observed (the initial explosion ionizes less than 10E-4) could be attributed to the critical velocity effect. However during the flight in March, 1983, essentially no ionization whatsoever was observed by E. M. Wescott, the principal investigator. Figure 2 is a photograph of the strontium neutrals from that observation.

The companion experiment, Star of Lima (34.010), placed an instrumented payload directly in a barium neutral beam directed almost directly downwards (hence almost perpendicular to the magnetic field lines) from an altitude of 450km. Figure 3 shows data from the UCSD electron analyzer near the time of the ignition. Quasi-dc electric field results from the Cornell University are included for comparison.

The percentage ionization observed in this experiment as determined either by ground based optical observations by E.M. Wescott, the principal investigator, or from UCSD ion detectors was quite low : about 10E-4 of the released cloud ionized. Even this could plausibly be accounted for by a slight exposure to photoionizing uv (about 1/40 of full exposure). However the instrumented payload detected many of the phenomena associated with the critical velocity effect, including broad band electric field waves around the

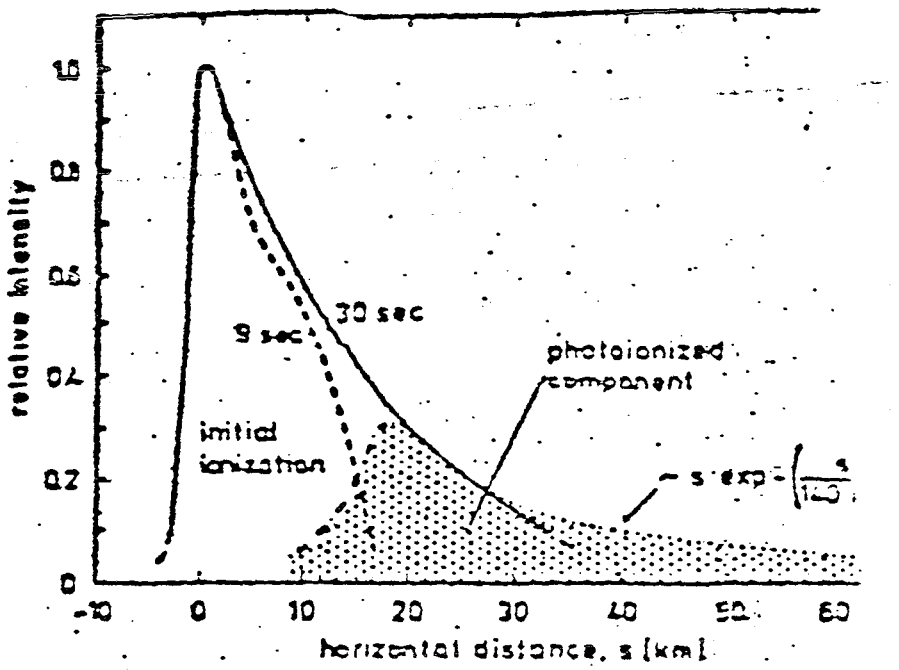


Fig 1 Densitometer traces of the ionized barium appearing above the terminator. The dotted part i of the data to the expected contribution from photo- tion. Nearly all ions to the left of $s = 15$ km are the of an initial ionization in darkness.

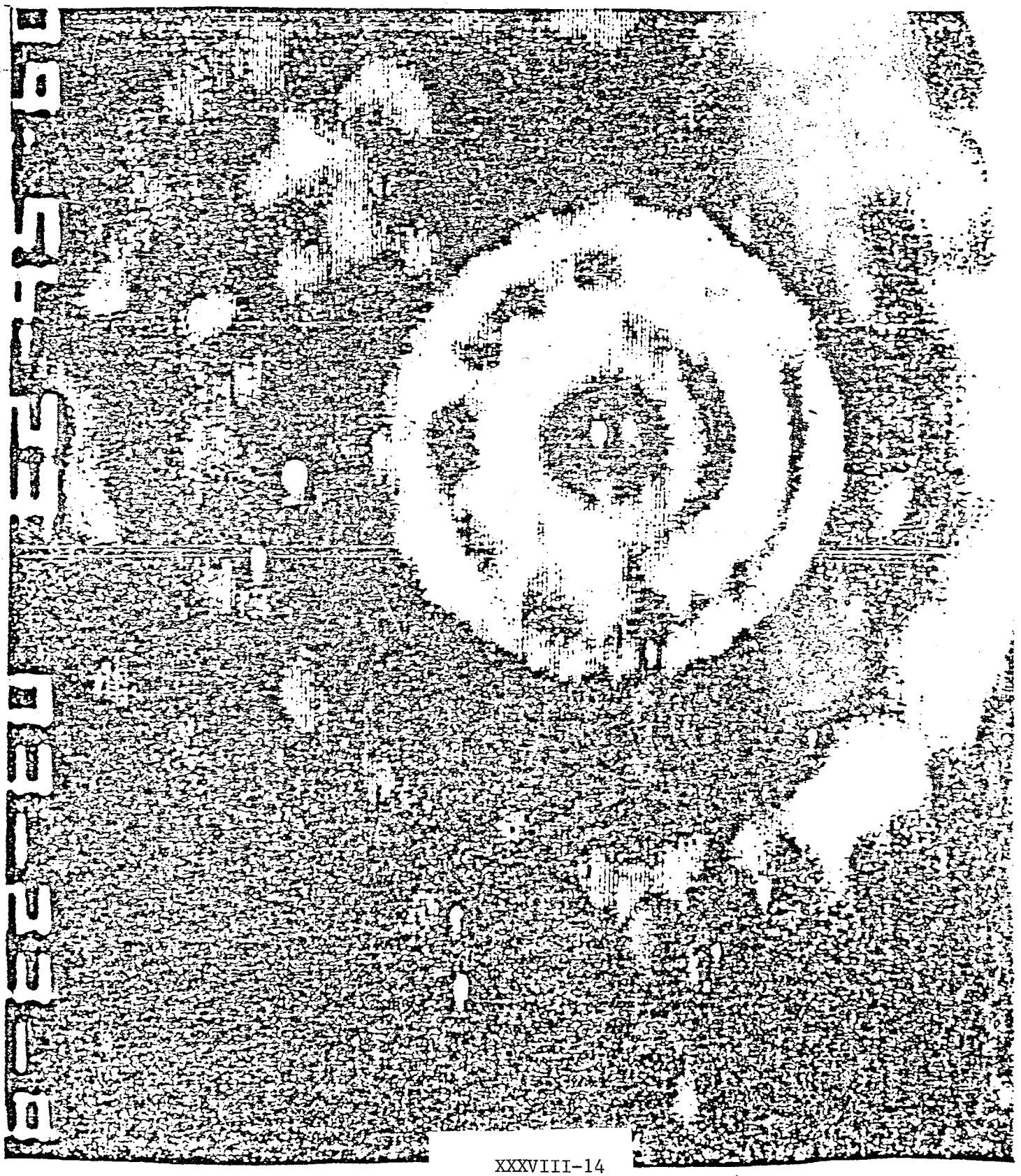


Figure 2. TV picture of neutral strontium vapor at +15 seconds, viewed from Paracas, Peru. Magnetic field line runs from center of

barium lower hybrid frequency and associated heating of electrons (the energetic electrons in Figure 3 could not have been produced directly from photoionization). It is clear that the processes believed to occur in critical ionization velocity experiments were in fact observed. The question arises: why was discharge not reached in Peru, yet was on Porcupine? Certain obvious candidate explanations can be ruled out; for example, the suggestion that the electrons escape from the system along field lines before they can have an ionizing collision can be ruled out, because of the electron pitch-angle isotropy observed. At present, the most likely explanation for the difference in the two barium release experiments is that the stronger magnetic field and higher plasma density in the Porcupine experiment allowed the background plasma to more effectively take up the momentum of the newly ionized barium atoms (enabling them to transfer their kinetic energy to heating electrons). The question of when the critical velocity effect will occur in space plasmas remains very open however.

The present set of experiments will be conducted with a complete set of instrumentation capable of resolving some of these questions. In one of the two Wallops flights, we will use barium as the neutral jet, and in the other, strontium, thereby exploring the differences between the two Condor experiments. As shown in Figure 4, there will be two instrumented payloads, so that both the electron thermal flux escaping from the beam and the size of the Alfven wave radiating from the interaction region (related to the efficiency of momentum transfer) can be directly measured. Each rocket will consist of a main payload, a daughter payload attached to the ejected nose cone, and the chemical release module. The daughter payload will be ejected along the magnetic field by orienting the vehicle with an Attitude Control System (ACS). The ACS will then reorient the main payload to eject the chemical canister nearly perpendicular to the local magnetic field. There will be a slight upward cant to the injection velocity so that ions and neutrals preferentially move upward toward the terminator for 320 nanometer solar ionizing radiation. Thus, the conditions for both Porcupine and Condor will simultaneously be realized. The main payload will measure directly the ionizing front. However, now the daughter payload will be able to determine not only the amplitude of the Alfven wave created by the interaction, but also the amount of escaping superthermal electrons from the interaction the region. Thus, the two largest imponderables in both theoretical and experimental discussions of the effect will be directly measured. An extra bonus can be realized from the mother-daughter configuration in this case. Namely, if the lower-hybrid plasma waves are observed on both electric field detectors, the parallel phase velocity of the waves can be measured by correlating the two signals. The parallel wavelenth of 10 km is ideal for this sort of measurement. The high phase velocity along B is the feature of these waves that results in the large electron resonant heating responsible for maintaining the ionization process. Thus, we believe that some detailed plasma physics can be learned from this experiment.

The instrumentation includes 3-axis electric field and electron density measurements on both the mother and daughter (Cornell and Royal Institute), a fast ion and electron spectrometer measuring complete ion and electron energy and pitch angle distributions 500 times per second during the iqnition period by means of a burst-capture on board (UCSD), a precise plasma density measurement (Utah State), and search coil magnetometers (RIT). The chemical release modules are constructed by the University of Alaska and Max

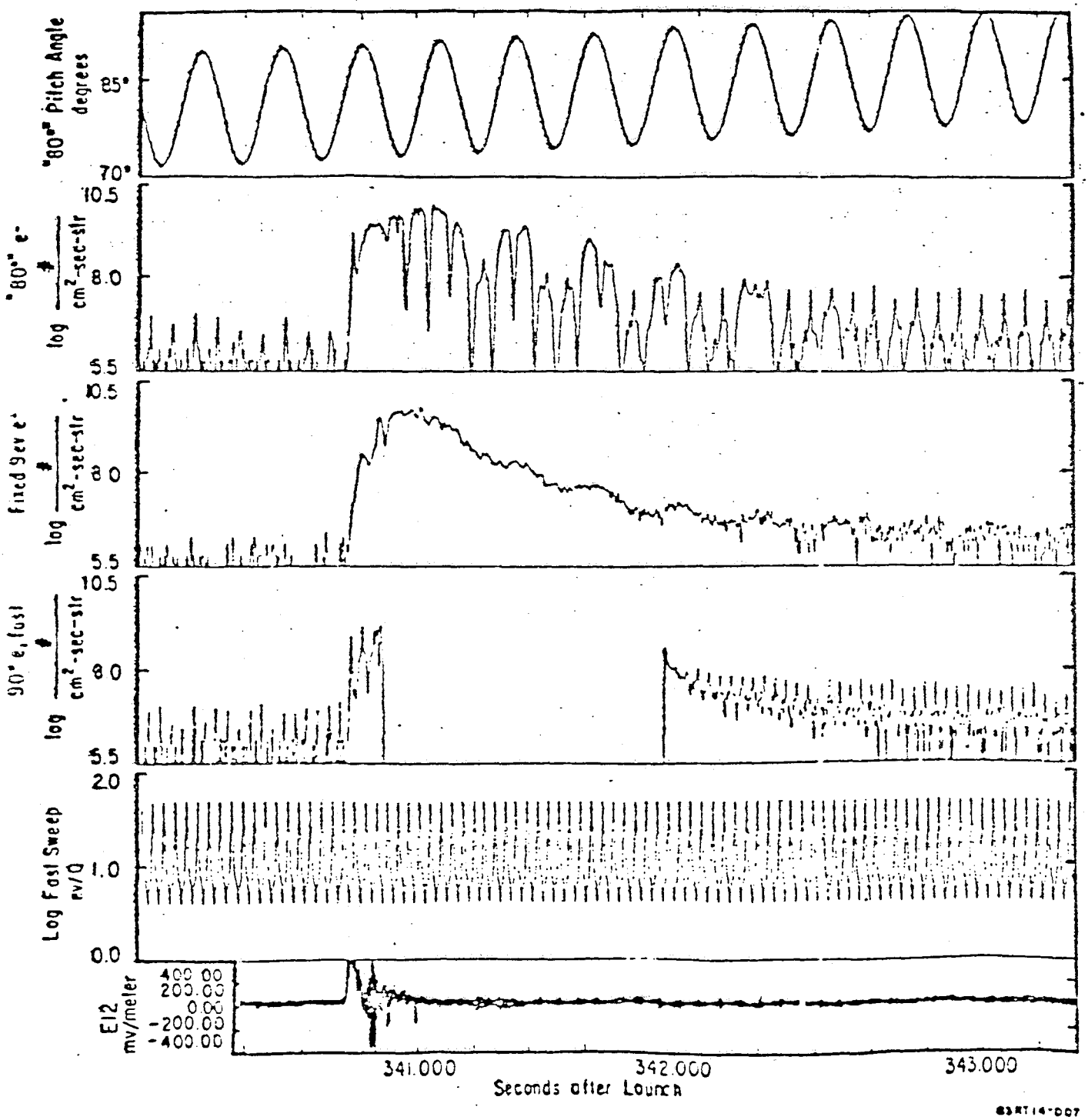


Fig. 3 Energetic electron fluxes measured by UCSD RPAS near the Ba release event at 340.658 flight time.

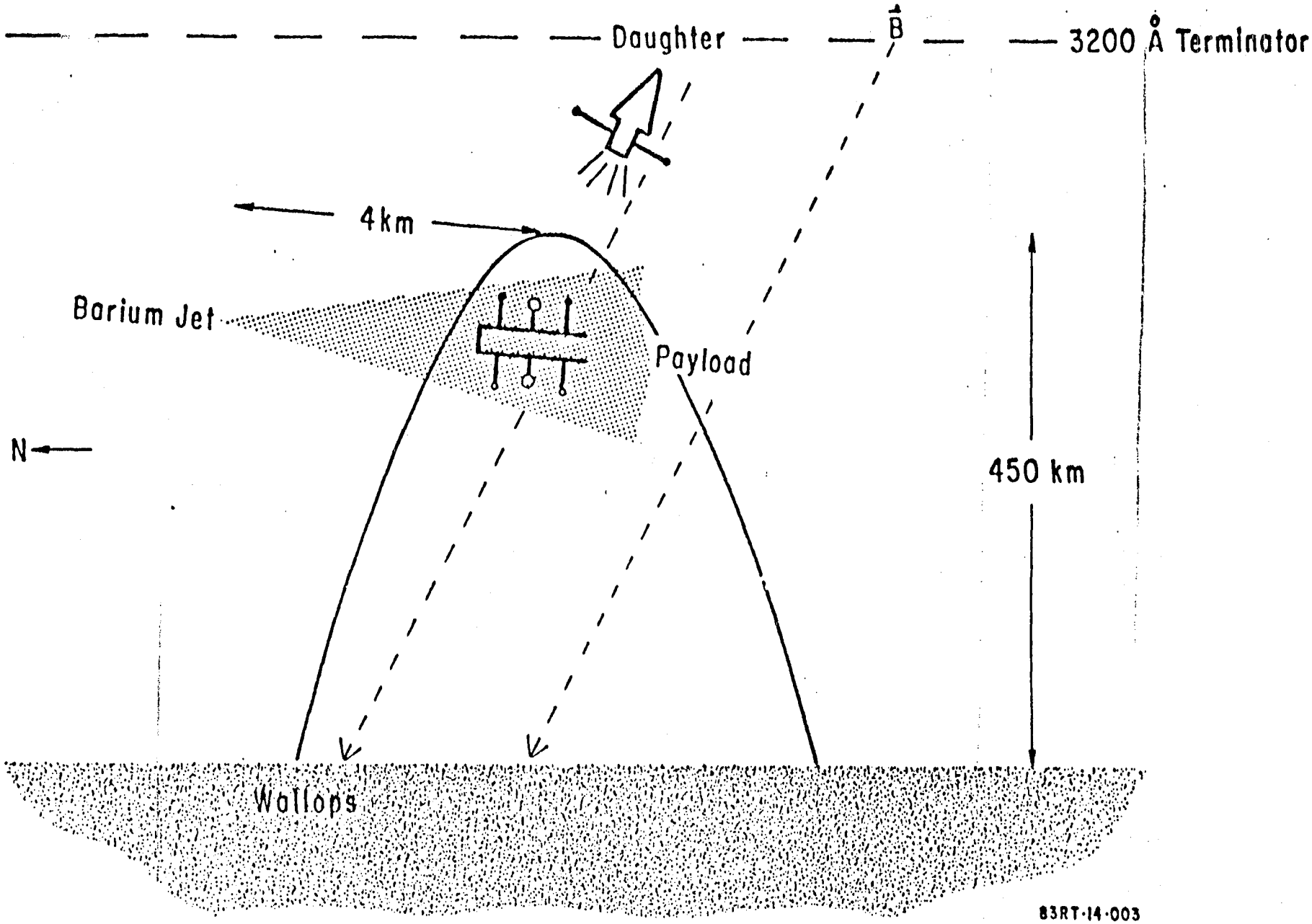

Daughter
B
3200 Å Terminator
4km
Barium Jet
Payload
N
450 km
Wallops
83RT-14-003

Fig. 4

1985

NASA/ASEE SUMMER FACULTY RESEARCH FELLOWSHIP PROGRAM

Marshall Space Flight Center
The University of Alabama

A Preliminary Study of Flat-Panel Displays

Prepared by:	Kenneth E. Yancey, Ph. D.
Academic Rank:	Professor
School and Department:	Lambuth College Department of Physics, Math, and Computer Science
NASA/MSFC: Division: Branch:	 Computers and Communications Mission Computer Systems
MSFC Counterpart:	Douglas T. Thomas
Date:	July 26, 1985
Contract Number:	NGT 01-008-021 The University of Alabama in Huntsville

A Preliminary Study of Flat-Panel Displays

by

Kenneth E. Yancey
Professor of Physics
Lambuth College
Jackson, Tennessee

ABSTRACT

This paper presents six display technologies that might be of future value in a spacelab workstation. Some have been developed to the point where they could be used as a computer display while others have not. The display technologies studied are electroluminescents, light-emitting diodes, gas plasma, liquid crystal, electrochromic, and electrophoretic. An explanation of each mechanism is provided along with the state-of-the-art development.

LIST OF TABLES

Introduction and Objective

The display that is portable or that requires more than a 26 inch diagonal screen will require something other than a cathode-ray tube. Firstly, the CRT's are heavy because of the thick glass needed to maintain the vacuum. Another problem with CRT's is the high voltage power source required to accelerate the electron beam. And, lastly, they are awkward and take up space due to the length of the electron gun.

The six flat-panel technologies in production now that might replace the CRT are plasma, liquid crystal, electroluminescent, vacuum fluorescent (VF), electrophoretic, and electrochromic display panels. VF has not been able to keep up with improvements in CRT's. And it will be another year before the last two mentioned are in the running. This paper will look at the first three mentioned. The objective is to collect information about displays that will be useful in making a display selection for spacelab.

Electroluminescent Displays (EL)

The basic means of operation of EL is not as well understood as is that of other display technologies. The AC or DC powder and AC or DC thin film configurations are based as much on experimentation as on theory. In the displays industry, single-crystal EL is called LED. Work by Professor R. Evrard at the University of Liege in 1980 and recorded by Professor F. Williams in the Journal of Luminescence volume 23, in 1981 made strides toward understanding high-voltage AC thin-film ZmS: Mn EL configurations. Theoretical work by Professor A. G. Fischer still holds for the ZmS: CuS powder configuration.

Three major types of mechanisms used in displays are:

TYPE I - Bipolar alternating double-injection luminescense, demonstratable with AC powder ZmS: CuS, or the Fischer Model.

TYPE II - Hot electron impact ionization of activator, demonstratable with AC thin-film ZmS: Mn and DC powder ZnS: CuS phosphors or the Chen and Krupka Model.

TYPE III - Memory effect in AC thin-film ZnS: Mn phosphors first reported by Yamauchi et. al. (1974) and described by Onton and Marrello.

AC Thin Film EL

The following configuration has been proven to have long life, high luminance, and video display capabilities: phosphor (ZmS: Mn) sandwiched between dielectrics (Y_2O_3, Si_O_2, and Al_2O_3), then these are sandwiched between the row electrode (Al), and the column electrode (Indium-tin oxide) on the viewing side, and then these are sandwiched between a seal material and glass on the front viewing side. When an electric field is set up between the electrodes, the thin-film emits light.

From -269°C to about 77°C only minor changes are observed in the luminance and electrical characteristics of ZmS: Mn AC thin-films. One problem with thin-films is the occurance of pinholes. The pinhole breakdown failure rate of weak thin-film sandwiches is accelerated at higher temperatures. Weak thin-film dielectrics will conduct and cause the panel to heat. High operating temperatures due to internal heating causes rapid panel deterioration. On the other hand, the insensitivity of AC thin-film ZmS: Mn panel enhances its potential range of applications.

When properly fabricated a luminance vs voltage hystersis has been demonstrated with the AC thin-film EL using ZmS: Mn phosphor. The hysteresis can be used to freeze a frame on the panel. And images frozen on the panel can be read to a memory device.

Failure mechanisms that have slowed the development of AC thin-film EL displays are electrical breakdown, film delamination, and electric field-induced chemical reactions. Electrical breakdowns in the form of pinhole electrical arcs are unavoidable. These burnouts, less than 50 micrometers in diameter and stable, are to be expected and tolerated. Two pinhole breakdowns per square inch is achievable, and if they do not destroy the pixel or line, this is allowable.

The Coulomb forces can tend to separate the thin-films at the time of external field polarity reversal. This can occur when air is included as a background gas during dielectric dispositions. Water entrapped in the film ionizes causing delamination. This delamination can be controlled by using state-of-the-art techniques for improved adhesion.

The field-induced chemical reaction is oxidation of the aluminum electrode. It can be stopped by blocking the source of oxygen with an oxygen-free material.

Sharp has reported the life of ACTFEL at over 20,000 hours in a controlled environment. They have also reported a typical 30% reduction in luminance in large panels after 10,000 hours. Aerojet Electrosystems report no pixel failures out of 900 pixels during 12,000 hours of operations.

AC powder EL is made from ZmS and ZmS phosphor, and is activated with supersaturated copper sulfide and other coactivators such as Mn, Cl, and Ag for color. AC powder EL devices are primarily used for transillumination of panels, keyboards, and other displays such as AC plasma panels and twisted-nematic liquid crystals. They can be used anywhere a continuous low-illuminance film is needed.

Of the powder EL sandwiches, the DC is more readily matrix addressable than is the AC. The DC powder EL using ZmS: Mn has been used commercially in 80- and 256-character readout displays and automative panels.

The technical maturity of AC thin-film EL is behind that of gas plasma and of the CRT in accumulated production volume and production experience, user experience, and associated test data. Also, the basic knowledge of the EL physical phonemena is immature compared with that of gas plasma and CRT. However, these are being overcome due to advantages of TFEL.

Although not as bright as gas plasma (GP) displays, EL display panels find their place between liquid crystal display (LCD) panels, which cannot rapidly display large amounts of data, and very expensive plasma displays, which show as much data as a CRT, but cannot produce a gray scale for video.

LCD's have difficulty providing a high-quality six-inch-diagonal screen while EL panels run into difficulty around 15 inches. Planar systems predict that EL will dominate the 6-inch to 15-inch flat-panel market.

The integration of an EL display with a portable computer ($8,000 mostly for the panel) demonstrated that a good flat panel display could be produced for personal use and that manufacturing processors need to be developed to reduce the cost of the EL panel.

The thin-film EL panels consist ZmS doped with Mn placed between two insulating dielectric layers. This is then sandwiched between vertically oriented and horizontally oriented electrodes. A black layer provides high contrast for the panel, and the entire thing is sealed in glass.

Only three companies are engaged in large scale production of thin-film EL -- Sharp of Japan and Lohja of Finland (U. S. subsidiary: Finlux, Saratoga, CA) and Planar Systems. Planar is a spin-off from the company Tektronix, Inc. when Tektronix decided to close down flat panel R&D, three engineers continued the work, backed by the company as a silent partner with a major interest. In October of 1984 Planar was awarded a $1.5 million contract by the Army for the development and manufacture of multicolor EL flat panel displays. These will be used in military ground vehicles.

In first quarter of 1985 the first 640 x 440 pixel EL display monitor, the largest of its kind was demonstrated at the Society for Information Display conference. This full-page panel, offering a resolution of 60 lines per inch on an active display area of 6 x 10 inches came from Planar Systems, Inc.

Takashi, Ito, and Hikasa reported this year in the Journal of Applied Physics a new multicolor ACTFEL device. A gate layer is deposited between layers of ZnS: ErF_3 and ZnS: SmF_3. This is placed between layers of dielectrics, Y_2O_3. And this is placed between the electrodes. The emission color of the device varies continuously between red and green with the driving voltage or frequency.

Gas Plasma Displays (GP)

Why have plasma displays been so successful? To achieve intrinsic matrix addressability, the display element must have a nonlineas relationship between luminance and applied voltage. Practically no light is emitted from the gas discharge element for voltages below the firing voltage. A lifetime of 50,000 hours is not uncommon in plasma displays. And the failure is not catastrophic but is a slow degradation in display quality. The pixels that have been discharged the most become the weakest. It is common for a plasma display to

outlast the entire life of the product. GP has good brightness and luminance, although not as good as in some other display technologies. Plasma displays compete favorably since they can achieve high contrast ratios because the display media is transparent. The basic structure of GP is only slightly more complex than that necessary for an array of X and Y electrodes on glass. Plasma displays are usually as rugged as the sheet of glass of which they are made. This rugged structure makes large-area plasma displays practical. They are available on the market with resolutions as high as 100 lines per inch and as large as one meter diagonal panels.

Another plus for GP is its tolerance to harsh environments and temperature extremes. To exhibit a reasonable lifetime, plasma displays must be hermetically sealed with a high-temperature glass frit seal. Hence GP will operate in high-humidity or in the presence of reactive gases. For many plasma displays the limiting factor on the operating temperature is determined by the drive circuits and not the display. The characteristics of most AC plasma displays can be considered invariant with temperature.

The capacitance of GP displays is low compared to thin-film electroluminescence. Therefore, the drive circuits require less current capability, a distinct advantage when using thin-film conductors.

A full range of color can be achieved by depositing phosphors in the GP display. Flat color TV displays with sixteen inch diagonals and excellent picture quality, have been made in laboratories. The brightness is not as good as that of a CRT. Plasma displays, as do all flat-panel displays, have low volume, slim profile, and freedom from distraction.

The most important reaction of a gas discharge is ionization. When an electric field is placed across the gas, electrons from a primer cell are accelerated by the field. If after reaching an energy of 21.6 ev the electron collides with a neutral neon atom, an electron can be ejected from the atom. Characteristic light is emitted from the ion and there are now two electrons being accelerated to a collision. These two electrons can then go on to create two more ionizations and so on, resulting in an avalanche. The number of ionizations occuring in an avalanche is in the range from ten to three-hundred.

Some electron will not have enough energy to knock an electron out of the neon atom but will have enough energy to knock an electron up to a higher energy level (excited level). This excited electron will soon fall back down to the ground state of the atom and emit a characteristic photon of light.

Not all excited states of the neon atom will allow the electron to drop to a lower state and emit a photon of light. One

that does not is called a metastable state. Thus themetastable neon atom wanders around carrying extra energy. If the metastable neon collides with an argon atom the extra energy is enough to ionize the argon causing its characteristic light to be emitted. The mixture that gives the largest amount of ionization per applied volt was found by F. M. Penning in 1937 to be Ne plus 0.1% AR. This is called a Penning mixture. Other Penning mixtures include neon plus small amounts of Xenon, Krypton, or even certain noninert gases. Argon-neon mixtures are commonly used in plasma displays.

Gas Plasma (GP) is the only flat-panel technology of the top six that currently offers performance equal to or better than that of the CRT. There are three types: AC, DC, and hybred. The DC plasma shows aging at temperatures below $0^{o}C$ and above $50^{o}C$. The AC-plasma panels operate over a wider temperature range, from $-40^{o}C$ to $+85^{o}C$. The DC-plasma has a typical life of 30,000 hours while that for AC-plasma is longer. The GP displays, also, have a 120^{o} viewing cone.

The DC-plasma is less expensive because it is easier to drive but, at the same time, it has no inherent memory and must be refreshed. The AC-plasma need not be refreshed. IBM now leads the AC-plasma technology with its 22 x 20 inch and 25,000 pixel display.

The major manufacturers of DC-plasma displays are Fuji Electronics America (Torrance, CA), NEC Corporation, OKI, and Matusushita.

The hybrid methoid combines the best features of both AC- and DC-plasma technologies. It is being developed by Plasma Graphics Corporation (Warren, NJ). The reduced number of drivers limits the use to handling static data, like word processing. Dynamic displays are not possible.

Liquid Crystal Displays (LCD)

For a century now some organic materials have been known to melt into a stable, ordered liquid state before melting again at a higher temperature into an isotropic liquid state. The molecules that make up the ordered liquid are rigid-crystal state can be indicated by an array of cigar-shaped rigid molecules. Customarily the direction the molecules point is indicated by a unit vector called the director.

In another phase of liquid crystal the elongated molecules lie in planes, parallel to the plane, and with the director of each plane rotated with respect to adjacent planes. This circular-staircase structure is characteristic of the chalerleric phase, that derives its name from the fact that some derivatives of cholesterol molecules exhibit this phase.

Another liquid-crystal phase is called smectic. The characteristics of the smectic phase is elongated molecules in layers usually with the molecular axis perpendicular to the plane. A typical melting sequence of a smectic could be from crystalline to smectic, to nematic, to isotropic liquid.

The dominant technology tends to be the twisted nematic mode. The mechanism for this mode is described later in the report.

Some researchers claim that by 1986 the liquid crystal displays will overtake plasma displays as the unit popular flat-panel graphics display. There are three types of LCD's: Twisted-Nematic Field Effect (TNFE), Active Matrix, and Smectic LCD's.

Commonly used in calculators TNFE LCD's consist to two glass plates with rows and columns of transparent conductor lines. The LC mixture is sandwiched between the two plates. The front plate has a transparent polarizing filter. The back plate has a reflecting polarizer, oriented at 90^{o} to the polarizer on the front plate.

When no current is applied to the conductors, the long LC molecules form spiral staircase shaped structures, joining the two glass surfaces. These molecules give light passing through them a 90^{o} twist. However, when current is applied, the crystals line up, and let pass through unaffected.

The result is that these cells with no current applied reflect ambient light, and appear light-colored. This is because light passing through the face plate is polarized in one direction, get twisted 90^{o} by the LC, allowing it to pass through the back plate polarized, which reflects it forward again. As it passes through the LC again, the light gets a second 90^{o} twist, reorienting it to its original position, and thus allowing it to pass out through the polarizing face plate.

Conversely, cells with current applied absorb ambient light, and appear dark. In this case, light is polarized in one direction by the front face plate, and passes through the aligned LC unaffected. When the polarized light reaches the back plate, the reflective polarizer of which is oriented 90^{o} opposite to that of the front plate, it is blocked. Therefore, light is not reflected back, and the cell appears significantly darker than surrounding cells.

One difficulty with LCD's is the slow response time of the LC's as they reorient themselves. Another is the viewing cone angle is about 40^{o}. Low brightness and poor contrast are other problems. One solution to the latter two problems is the Active Matrix LCD's. Manufacturers have deposited active, non-linear elements at each pixel. This creates a matrix with

a driver at each display point. Two devices used to provide the active matrix are thin-film-transistors and metal-insulator-metal diodes. This causes sharp rise in contrast and the viewing cone jumps from 40° to at least 90°.

Crystal vision (Sunnyvale, CA) has developed a smectic LC that is solid at room temperature, but liquid at slightly higher temperatures. In this case the data are literally frozen in place and the LCD does not need to be refreshed. This simplifies the electronics, and provides a much larger display. However, display time is slow and drivers needed to handle the heating requirements are expensive.

Crystal vision, Inc. has developed a LCD that has 640 cols x 256 rows and can be used in a 25 line x 80-char format. Although the writing speed is about 2,000 chars/sec, it special cases it can approach 1,000 chars/sec.

The display consists of two parallel glass plates separated by about 145 micro-meters which is filled with a smectic liquid crystal and dye mixture. The front plate has transparent indium-tin oxide column traces running vertically. The back glass has textured mirror-finished metalic row heating traces running horizontally. A pixel is created at the intersection of each of the 640 cols and 250 rows, for a total of 16,000 pixels.

In this dichroic display, the cigar-shaped, liquid crystal molecules take up dyes along their length. When the molecules lie parallel to the viewing surface, the dyes are visible. When the molecules are perpendicular to that surface, they are invisible. In the non-excited state, the molecules are parallel to the viewing surface and the color appears (the background). When excited, they stand up on end and become transparent, and a negative image is produced.

The metalic electrode on the back glass plate serves both as an electric-field source and as a heater for the dichroic LC; it also serves as a reflector. To generate a display, the electrode first melts the liquid crystal. The the electrodes field aligns the dichroic molecules. Finally, when the electrode is turned off the LC refreezes, capturing the display.

Vacuum Fluorescent Display (VF)

VF displays a number of advantages over other emissive technologies, partially for small alphanumeric displays. VF displays are five times brighter than any plasma type display. Brightness for VF is between 150 and 200 footlamberts compared to between 30 and 45 for gas plasma. The display characters can be in a wider range of sizes for VF than for GP displays. VF displays also operate over a much wider temperature range than GP; from -40°C to +85°C for VF. For GP, the best range

will operate from 0°C to +55°C. As with GP, the viewing cone for VF display is 120°.

In formation content flat panel displays are usually either limited by electronic or physical/manufacturing constraints. VFD's are limited by the latter, while multiplexing large arrays is not difficult, manufacturing large grids is. Consequently, the largest VFD's available are 16 square inches.

For low information displays, VFD's offer low cost and low power requirements. As a result they are competing successfully in the market with LCD's and GPD's. There is a large market developing for VF displays in the automative industry for clocks, radios, and dashboard displays. They are attractive and can be produced in various colors.

Electrophoresis

Electrophoresis has been around for some time, but is only in the last several years that interest in its use as a display technology has grown. In electrophoretic displays, particles migrate from a cathode to an anode, or vise versa. Titanium-dioxide particles, that are white, are microencapulated in an organic material that can be changed. These white particles are placed in a black suspension. If a voltage is placed across the electrodes, the positively-charged titanium dioxide will migrate toward the negatively charged anode producing a white image. The anode must be optically transparent, since it is the front plate and the viewing surface. A problem about to be solved is that of keeping the particles in the suspension stable.

Two companies working on electrophoretic displays are Exxon Enterprises (San Jose, CA) and North American Philips (Briarcliff Manor, NY). However, the product probably won't be ready for another year. When ready the product will be a passive display similar in performance to LCD's. But it will not have polarizers as LCD's do and so will have greater contract and viewing angle.

Electrochromic Display

Electrochromic displays are material that change colors when exposed to an electric field. When the field is removed, the color remains, applying a field of reversed polarity changes the color back again.

Electrochromic displays based on tungsten trioxide change from colorless to deep blue. Daini Seikasha (Tokyo, Japan) has developed a display based on this material, which was used in Seiko digital watch display. More recently, Rockwell

International reported on an electrochromic display using diphthalocyanines that can display a variety of colors - including shades of blue, orange, and purple - depending on the voltage applied.

Early use of the electrochromic displays will be in instruments and aircraft cockpits readouts. Although every material responds differently, response time tends to be a problem for most of them, limiting the technology to lower information content applications. Most companies involved in electrochromic displays stress the need for more research.

Conclusion and Recommendation

If black and white will suffice for the display, then electroluminenscent (EL) has some definite advantages like weight, size, and voltage required. If color is needed, the shadow mask CRT is needed.

DISPLAY UNITS

PARAMETER	(SHUTTLE)	E.L.	PLASMA	L.C.D.	SHADOW MASK
Viewable size	5 x 7	5 x 6 1/4	8 x 8	4 x 9	11" x 14"
Power	90	20	40	.15	130
Temperature	-28° to 54°C	-55° to 71°C	-30° to 63°C	0° to 50°C	-30° to 55°C
Pressure	8.3 PSIA	0 PSIA	Problem	Problem	8.3 PSIA
Weight	27	10	55	3	35
Highest Voltage	14 KV	160 V	150 V	5 V	25 KV
Readability	Sunlight	Sunlight	Room	Sunlight	Sunlight
Brightness (F1-L)	55	30	30	N/A	30
Viewing Angle	100° Cone	140° Cone	120° Cone	40° Cone	140° Cone
Resolution (Characters)	51 x 26	80 x 58	85 x 51		132 x 55
Spec	Space (Norden)	Battlefield (Hycom)	Submarine (Interstate)	None	Aircraft

TABLE I -- IBM SUMMARY DATA

References

1. Bindra, Ashok, "Active Matrix Addressing Brightens LCD-Panel Picture", Electronic Engineering Times, Page 43, May 13, 1985.

2. Bindra, Ashok, "Planar Systems Claims First Full-Page-Size EL Monitor", Electronic Engineering Times, Page 4, May 13, 1985.

3. Bloom, Michael, "Flat-Panel Displays", Electronic Engineering Times, Page 29, November 26, 1984.

4. Hirate, Ito, and Hikasa, "A New Multicolor Electroluminescent Device", Journal of Applied Physics 57 (3), February 1985.

5. Hirchon, Robert, "Flat-Panel Display Technologies", Electronic Imaging, Page 40, January 1984.

6. IBM, "Space Workstation: Space Shuttle Programs"

7. Markhoff, Nicolas, "Flat Panel Technologies vie to displace CRT in Terminals", Computer Design, Page 42, September 1984.

8. Tannas, Lawrence, Flat Panel Displays and CRT's, Von Nostrand Reinhold co., New York, 1985.

1985

NASA/ASEE SUMMER FACULTY RESEARCH FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA
IN HUNTSVILLE

SOME ENERGY CONSIDERATIONS IN GAMMA RAY BURST LOCATION DETERMINATIONS BY AN ANISOTROPIC ARRAY OF DETECTORS

Prepared by:	John H. Young
Academic Rank:	Associate Professor
University:	University of Alabama at Birmingham
Department:	Physics
NASA/MSFC:	
Laboratory:	Space Science
Division:	Astrophysics
Branch:	High Energy Physics
MSFC Counterpart:	Gerald J. Fishman

Date:	August 16, 1985
Contract Number:	NGT 01-008-021 The University of Alabama in Huntsville

SOME ENERGY CONSIDERATIONS IN GAMMA RAY BURST LOCATION DETERMINATIONS BY AN ANISOTROPIC ARRAY OF DETECTORS

BY

John H. Young
Associate Professor of Physics
University of Alabama at Birmingham
Birmingham, Alabama

ABSTRACT

The anisotropic array of detectors to be used in the Burst and Transient Source Experiment (BATSE) for locating gamma ray burst sources is examined with respect to its ability to locate those sources by means of the relative responses of its eight detectors. It has been shown that the energy-dependent attenuation effects of the aluminum window covering each detector has a significant effect on source location determinations. Location formulas have been derived as a function of detector counts and gamma ray energies in the range 50 - 150 keV. Deviation formulas have also been derived and serve to indicate the location error that would be caused by ignoring the influence of the passive absorber.

ACKNOWLEDGEMENTS

I am genuinely grateful for having been given the honor and opportunity to participate in the NASA/ASEE Summer Faculty Fellowship Program. The hospitality, facilities, and exposure to the operations of the MSFC have afforded a valuable experience and appreciation of the enviable record of achievement of the Center.

My counterpart, Dr. Gerald Fishman, and his BATSE colleagues warrant special thanks for accepting my participation in a research area which was completely new to me. I hope that our collaboration will continue with mutual benefit.

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

One of the four experiment packages aboard the Gamma Ray Observatory (GRO), scheduled for release into orbit from a shuttle flight in May 1988, is the Burst and Transient Source Experiment (BATSE). Utilizing an array of eight NaI(Tℓ) detectors BATSE will, as one of its major objectives, attempt to determine the angular locations of gamma ray burst sources by means of analyzing the relative count responses of the detector array. Figure 1 depicts the GRO configuration[1]. If the planes of the detector faces, each having an area of about 2000 cm^2, were extended so as to form a closed surface, that surface would be the regular octohedron also depicted in Figure 1.

The 1.27 cm thick detector crystals will be overlaid by 2mm of aluminum. The gamma ray flux incident on the aluminum windows will be attenuated by that material and a diminished flux will consequently be incident on the detectors. The relative orientation of the detectors, dictated by the octohedral geometry, necessitates that gamma rays incident on one detector will have transversed a different thickness of aluminum than those incident on that detector's neighbors. Thus, in addition to being energy dependent due to the material's attenuation coefficient, the attenuation will be orientation dependent.

The work described here is the result of an effort to take the energy dependence of the relative detector responses to a burst into account in the angular location determination of the source. The energy-dependent attenuation effects are discussed and location formulas containing these effects are presented.

THE BATSE GEOMETRY

The BATSE detector array conforms to the geometry of a regular octohedron shown in Figure 2. The planes, corresponding to the eight detector faces, are identified by the numbering scheme indicated. The outward normals from the eight planes are given by the following, in which $\hat{i}$, $\hat{j}$, and $\hat{k}$ are unit vectors along the positive x, y, and z axes, respectively:

$$\begin{aligned} \hat{u}_1 &= \frac{1}{\sqrt{3}}(\hat{i} + \hat{j} + \hat{k}) & \hat{u}_5 &= \frac{1}{\sqrt{3}}(-\hat{i} - \hat{j} + \hat{k}) \\ \hat{u}_2 &= \frac{1}{\sqrt{3}}(\hat{i} + \hat{j} - \hat{k}) & \hat{u}_5 &= \frac{1}{\sqrt{3}}(-\hat{i} - \hat{j} - \hat{k}) \\ \hat{u}_3 &= \frac{1}{\sqrt{3}}(-\hat{i} + \hat{j} + \hat{k}) & \hat{u}_7 &= \frac{1}{\sqrt{3}}(\hat{i} - \hat{j} + \hat{k}) \\ \hat{u}_4 &= \frac{1}{\sqrt{3}}(-\hat{i} + \hat{j} - \hat{k}) & \hat{u}_8 &= \frac{1}{\sqrt{3}}(\hat{i} - \hat{j} - \hat{k}) \end{aligned} \qquad (1)$$

The following relationships between these normals are worthy of note and will be used later:

$$\hat{u}_5 = -\hat{u}_2,\ \hat{u}_6 = -\hat{u}_1,\ \hat{u}_7 = -\hat{u}_4,\ \hat{u}_8 = -\hat{u}_3,$$

$$\hat{u}_4 = \hat{u}_3 + \hat{u}_2 - \hat{u}_1 \ . \qquad (2)$$

If the octohedron is imagined to be centered at the origin of a large sphere of radius R (Figure 3) from whose surface a burst originates with angular coordinates θ and ϕ, the detected

portion of the burst flux will be that which propagates along the direction $-\vec{R}/R$. Let $\vec{I}$ represent the flux density (photons per unit area) in this direction, with magnitude I_o; it can be simply represented as

$$\vec{I} = I_o(-\frac{\vec{R}}{R}) = -I_o(\sin\theta\cos\phi\hat{i}+\sin\theta\sin\phi\,\hat{j}+\cos\theta\hat{k}) \quad . \qquad (3)$$

BATSE data of interest here are derived from the incidence of $\vec{I}$ onto any combination of the eight planes of the octohedron. The number of photons passing through the i-th plane whose area is A will be

$$N_i = \hat{I}\cdot(-\hat{u}_i A) = I_o A\cos\alpha_i \quad , \qquad (4)$$

where α_i is the angle between the burst direction and $-\hat{u}_i$.

DETECTOR RESPONSE

Only under ideal, and thus rarely attainable, conditions would a detector whose outward normal is $\hat{u}_i$ actually record the number of photons given by Eqn. (4). That expression must be multiplied by the detector efficiency, resulting in a decrease from Eqn. (4). The efficiency is an energy-dependent factor which accounts for attenuation losses in the passive absorber covering the detector crystal as well as the probability that a photon will interact with the crystal upon entering it. If t and D represent, respectively, the window and crystal thicknesses, the efficiency for the case of normal incidence is given by [2]

$$\eta(E) = e^{-\mu_A(E)t}\left(1 - e^{-\mu_N(E)D}\right), \qquad (5)$$

where $\mu_A(E)$ and $\mu_N(E)$ are the energy-dependent attenuation coefficients of the aluminum and sodium iodide. If, however, a photon were to enter the system at some angle α relative to the system axis, it would transverse a thickness $t/\cos\alpha$ of absorber before getting into the crystal. Further, propagation at that angle gives the photon an increased effective thickness of crystal with which to interact and thus be recorded. The latter effect amounts to replacing D by $D/\cos\alpha$ in Eqn. (5). The detector efficiency will be used here in the form

$$\eta(E) = e^{-\mu_A(E)t/\cos\alpha}\left(1 - e^{-\mu_N(E)D/\cos\alpha}\right). \qquad (6)$$

The number of photons, N_i, recorded by the i-th detector will then be given by the product of Eqns. (4) and (6); that is

$$N_i(E,\theta,\phi) = I_o A \cos\alpha_i e^{-\mu_A(E)t/\cos\alpha_i}\left(1 - e^{-\mu_N(E)D/\cos\alpha_i}\right). \qquad (7)$$

The energy dependence of the N_i considered here will be the result of the attenuation coefficients' dependence on photon energy. The relationship of N_i to the angular location of the burst is dictated by the functional relationship between α_i and (θ,ϕ); these relationships are easily determined from Eqns. (1), (3), and (4). The objective of the present work is to develop the means of determining θ and ϕ in terms of the recorded

detector counts, N_i. It is apparent from Eqn. (7) that the computation of the angular coordinates of the burst will give results dependent on photon energy. The following discussion will, in particular, be focussed on the sort of corrections required in location determinations mandated by the inclusion of energy considerations above 50 keV.

The exponential function $e^{-\mu t/\cos\alpha}$ has maximum value $(0<\alpha<90^0)e^{-\mu t}$, corresponding to normal incidence. Table 1 lists the attenuation coefficient values for aluminum[3] over the energy range of interest (50 - 200 keV) and compares the value of the exponential $e^{-\mu t}$ for t = 0.2 cm with its linear approximation, $1 - \mu t$. The largest discrepancy between the exact and approximate values is seen to occur at 50 keV, at which the difference is about 2%. At 200 keV, on the other hand, the linear approximation of the exponential function is seen to be accurate to within 0.2%. It will be shown below that the most probable energy of a photon within a typical burst is about 100 keV. The linear approximation agrees with the exponential at that energy to with 0.5%. It would appear reasonable then to approximate the first exponential appearing in Eqn. (7) by its linear representation, and thereby obtain angular locations which will be "first order," a refinement over the "zero-th order" attainment based on totally ignoring attenuation effects.

Table II displays values of attenuation coefficients of sodium iodide[4] over the energy range of interest, along with values of the second exponential factor appearing in Eqn. (7). It is seen that for a 1.27 cm-thick crystal, this factor is

essentially unity up to about 125 keV. It is important to recognize, however, that the values displayed in the fourth column correspond to normal incidence and thus correspond to minimum values of $(1 - e^{-\mu D/\cos\alpha})$. Alternatively, that column can be looked upon as exhibiting the proximity of this factor to unity. More typically an angle of incidence $\alpha \sim 45°$ might be encountered, thus requiring photons to traverse a greater thickness of the crystal, while increasing their probability of detection. The latter values are tabulated in the fifth column. It should be noted here that if these same considerations regarding the angle of incidence were given to Table I that the exponential factor and its linear approximation would be in agreement to within 0.8%. The burst location discussion presented here will assume validity of the linear approximation of attenuation in the aluminum window and that the crystal will respond to any photon which enters it.

Gamma rays within the 50 - 150 keV then can be expected to give rise to the following approximate forms of Eqn. (7):

$$N_i(E),\theta,\phi) \simeq I_o A(\cos\alpha_i - \mu_A(E)t) \qquad (8)$$

The fact that the attenuation coefficient of sodium iodide no longer appears in this approximation allows the subscript to be dropped from μ_A; hereafter $\mu(E)$ will refer to aluminum.

BURST LOCATION DETERMINATION

Use of Eqn. (8) in determining the angular coordinates of a burst source will now be described. The energy dependence of the results will be emphasized and location formulas including the energy dependent term in Eqn. (8) will be derived.

Consider a burst activiting detectors 1 through 4. (This means that $55° < \theta < 125°$ and $45° < \phi < 135°$. Generalization of the results to other detector combinations will be discussed later.) Using the definition of $\cos\alpha_i = (\vec{I}/I_o)\cdot(-\hat{u}_i)$ along with Eqns. (1) and (3) gives the following expressions for the $\cos\alpha_i$ in terms of θ and ϕ:

$$\begin{aligned}
\cos\alpha_1 &= \frac{1}{\sqrt{3}}(\sin\theta\cos\phi + \sin\theta\sin\phi + \cos\theta) ,\\
\cos\alpha_2 &= \frac{1}{\sqrt{3}}(-\sin\theta\cos\phi + \sin\theta\sin\phi - \cos\theta) ,\\
\cos\alpha_3 &= \frac{1}{\sqrt{3}}(-\sin\theta\cos\phi + \sin\theta\sin\phi + \cos\theta) ,\\
\cos\alpha_4 &= \frac{1}{\sqrt{3}}(-\sin\theta\cos\phi + \sin\theta\sin\phi - \cos\theta) .
\end{aligned} \tag{9}$$

At this point two important identities between the $\cos\alpha_i$ should be noted. It can be immediately seen from inspection of the statements given in Eqn. (9) that

$$\cos\alpha_4 = \cos\alpha_3 + \cos\alpha_2 - \cos\alpha_1 . \tag{10}$$

This relationship simply restates the last connection between the $\hat{u}_i$ noted in Eqn. (2) and indicates the sufficiency for the use of three detectors only. A less obvious but, as will be seen below, very important additional identify which can be easily verified is the following:

$$(\cos\alpha_1 - \cos\alpha_2)^2 = \frac{4}{3} - [(\cos\alpha_2 + \cos\alpha_3)^2 + (\cos\alpha_1 - \cos\alpha_3)^2] \quad .(11)$$

It is reasonable to expect that burst location determinations could be made with increased confidence for the case of the more highly activated detectors. Eqn. (10) will allow those determinations to be made in terms of the three highest detector responses, assumed here to be N_1, N_2, and N_3. The number of photons incident from a given burst, I_oA, appears in each detector statement of the form given in Eqn. (8). This quantity is unknown, however, and Eqn. (11) serves the purpose of allowing it to be determined.

Combining Eqns. (9) and (8) gives the following system:

$$\begin{aligned} N_1 &= \frac{1}{\sqrt{3}} I_o A[\sin\theta\cos\phi + \sin\theta\sin\phi + \cos\theta - \mu(E)t] \\ N_2 &= \frac{1}{\sqrt{3}} I_o A[\sin\theta\cos\phi + \sin\theta\sin\phi - \cos\theta - \mu(E)t] \qquad (12) \\ N_3 &= \frac{1}{\sqrt{3}} I_o A[-\sin\theta\cos\phi + \sin\theta\sin\phi + \cos\theta - \mu(E)t] \end{aligned}$$

The explicit statement for N_4 will be suppressed since Eqn. (10) implies that $N_4 = N_3 + N_2 - N_1$. It is more convenient to express

the angular functions of Eqn. (12) in terms of a "corrected count" expressed as

$$N_i' = N_i + \frac{1}{3} I_o A\mu(E)t \ ; \tag{13}$$

Then

$$N_1' = \frac{1}{\sqrt{3}} I_o A(\sin\theta \cos\phi + \sin\theta \sin\phi + \cos\theta) \ ,$$

$$N_2' = \frac{1}{\sqrt{3}} I_o A(\sin\theta \cos\phi + \sin\theta \sin\phi - \cos\theta) \ , \tag{14}$$

$$N_3' = \frac{1}{\sqrt{3}} I_o A(-\sin\theta \cos\phi + \sin\theta \sin\phi + \cos\theta) \ .$$

Eqns. (14) are easily combined to yield expressions which can be solved for θ and ϕ. It is apparent that

$$N_2' + N_3' = \frac{2}{\sqrt{3}} I_o A \sin\theta \sin\phi \ , \tag{15}$$

and

$$N_1' - N_3' = \frac{2}{\sqrt{3}} I_o A \sin\theta \cos\phi \ . \tag{16}$$

Division of these and use of Eqn. (13) gives

$$\tan\phi = \frac{N_2' + N_3'}{N_1' - N_3'} = \frac{N_2 + N_3}{N_1 - N_3} + \frac{\frac{2}{\sqrt{3}} I_o A\mu(E)t}{N_1 - N_3} \tag{17}$$

The second term on the right of Eqn. (17) will be viewed as a connection to the azimuthal location of the burst, and is due to the finite thickness of the detectors' aluminum window. Its significance will later be discussed in detail.

The expression for $\tan\phi$ given in Eqn. (17) can be used to construct $\sin\phi$; the result is, in terms of the N_i',

$$\sin\phi = \frac{N_2' + N_3'}{(N_1' - N_3')^2 + (N_2' + N_3')^2} . \tag{18}$$

The statements in Eqn. (14) further give

$$N_1' - N_2' = \frac{2}{\sqrt{3}} I_o A \cos\theta , \tag{19}$$

and

$$N_2' + N_3' = \frac{2}{\sqrt{3}} I_o A \sin\theta \sin\phi . \tag{20}$$

Division of these and use of Eqns. (13) and (18) gives

$$\tan\theta = [N_1'^2 + N_2'^2 + N_3'^2 + 2N_3'(N_2 - N_1)]^{\frac{1}{2}} \tag{21}$$

The expressions for determining θ and ϕ in terms of N_1, N_2, and N_3 are seen to be energy dependent and contain the unknown flux density I_o. The latter quantity can be eliminated from Eqns. (17) and (21) by expressing it in terms of the detector counts. This can be achieved by use of the $\cos\alpha_i$ identity given in Eqn. (11). In terms of the N_i',

$$\left(\frac{N_1'}{I_o A} - \frac{N_2'}{I_o A}\right)^2 = \frac{4}{3} - \left[\left(\frac{N_2'}{I_o A} + \frac{N_3'}{I_o A}\right)^2 + \left(\frac{N_1'}{I_o A} - \frac{N_3'}{I_o A}\right)^2\right] \tag{22}$$

If the N_i' are expressed according to Eqn. (13) a quadratic in $I_o A$ results which has solution

$$I_oA = \frac{2\mu(E)t(N_2 + N_3)}{\left[\frac{4}{3} - 4(\mu(E)t)^2\right]} +$$

$$\frac{\left\{\left[2\mu(E)t(N_2+N_3)\right]^2+\left[\frac{4}{3} - 4(\mu(E)t)^2\right]\left[(N_1- N_2)^2+(N_1- N_3)^2+(N_2+N_3)^2\right]\right\}^{1/2}}{\left[\frac{4}{3} - 4(\mu(E)t)^2\right]} \quad (23)$$

The negative root has been discarded by the requirement $I_oA > 0$.

Equations (17) and (21), along with (23), are sufficient to determine the angular location of the burst source for photons of energy between 50 keV and, perhaps, 150 keV. The expressions to which these quantities reduce in the limit $t \to 0$ are those which can easily be derived by ignoring attenuation (and thus energy) effects from the outset. The latter expressions, as previously mentioned, can be thought of as "zero-th order expressions" and will hereafter be superscripted "(o)". Explicitly,

$$\tan\phi^{(o)} = \frac{N_2 + N_3}{N_1 - N_3}, \quad (24)$$

$$\tan\theta^{(o)} = \frac{\left[N_1^2 + N_2^2 + 2N_3^2 + 2N_3(N_2 - N_1)\right]^{1/2}}{N_1 - N_2} \quad (25)$$

and

$$(I_oA)^{(o)} = \frac{\sqrt{3}}{2}\left[(N_1 - N_2)^2 + (N_1 - N_3)^2 + (N_2 + N_3)^2\right]^{1/2}. \quad (26)$$

That neither Eqn. (24) nor (25) contains the flux density explicitly means that Eqn. (26) need not be evaluated for locations in this order.

The effect of the aluminum window on burst location determination is seen, by comparing Eqns. (17) with (24) and (21) with (25), to give rise to an abberation. The role of the energy dependence in this abberation will now be examined. Computations of the source location and flux density must be made in terms of detector counts. The energy influence on these quantities can perhaps best be illustrated by choosing reasonable values of what might typify data registered by the three detectors. For this purpose let $N_1 = 3$, $N_3 = 2$, and $N_3 = 1$. (These numbers have, of course, been scaled down by $\sim 10^3$, which is permissible since only the relative counts really matter.) These values give, with t = 0.2 cm and Eqns. (23), (17), and (21) the results displayed in Table III. The deviations, $\delta\phi = \phi - \phi^{(o)}$ and $\delta\theta = \theta - \theta^{(o)}$, are seen to diminish with increasing energy. This result is what would be expected, as the attenuation would be reduced for the more energetic photons. It would appear, however, that in no case displayed there is the deviation negligably small.

Although not small with respect to burst location determinations $\delta\phi$ and $\delta\theta$ are small in the sense that their tangents can be set equal to these angles (in radians) themselves; e.g., $\tan\delta\phi \simeq \delta\phi$. These approximations are known to hold for angles considerably larger than those displayed in Table III and allow general deviation formulas to be derived.

According to Eqn. (17)

$$\tan\phi - \tan\phi^{(o)} = \frac{\frac{2}{\sqrt{3}} I_o A\phi(E)t}{N_1 - N_3} .$$

Setting $\phi = \phi^{(o)} + \delta\phi$, using the addition formula to express $\tan(\phi^{(o)} + \delta\phi)$, with $\tan\delta\phi \simeq \delta\phi$ gives

$$\frac{\tan\phi^{(o)} + \delta\phi}{1-\delta\phi\,\tan\phi^{(o)}} - \tan\phi^{(o)} = \frac{\frac{2}{\sqrt{3}} I_o A\mu(E)t}{N_1 - N_3} .$$

Solving for $\delta\phi$ and expressing $\tan\phi^{(o)}$ according to Eqn. (24) gives the result

$$\delta\phi = \frac{\frac{2}{\sqrt{3}} I_o A\mu(E)t(N_1-N_2)}{(N_1-N_3)^2 + (N_2+N_3)^2 + \frac{2}{\sqrt{3}} I_o A\mu(E)t(N_2+N_3)} . \qquad (27)$$

As might be anticipated from Eqn. (21) the corresponding calculation of $\delta\theta$ is more involved. The fact that $\tan\theta$ is expressed in terms of a square root necessitates solving a quadratic equation in $\delta\theta$. The somewhat laborous details of the calculation will be omitted here in favor of getting on with the result:

$$\delta\theta_{\pm} = \frac{-\tan\theta^{(o)}[1+C+\tan^2\theta^{(o)}]}{[1 - C\tan^2\theta^{(o)} - \tan^4\theta^{(o)}]} \qquad (28)$$

$$\pm \frac{\{\tan^2\theta^{(o)}[1+C+\tan^2\theta^{(o)}]^2 + C[1-C\tan^2\theta^{(o)} - \tan^4\theta^{(o)}]\}^{1/2}}{[1 - C\tan^2\theta^{(o)} - \tan^4\theta^{(o)}]}$$

where $\tan\theta^{(o)}$ is given by Eqn. (25) and

$$C = \frac{\frac{4}{\sqrt{3}} I_o A\mu(E)t(N_2+N_3 + \frac{1}{\sqrt{3}} I_o A\mu(E)t)}{(N_1-N_2)^2} . \qquad (29)$$

The choice of sign is made according to the requirement that $\delta\theta_{\pm} \to 0$ as $t \to 0$ (i.e., as $C \to 0$). This means that if $\tan\theta^{(o)} > 0$, choose $\delta\theta_{+}$; if $\tan\theta^{(o)} < 0$, choose $\delta\theta_{-}$. It might appear that $\delta\theta$ would become infinite for that value of $\theta^{(o)}$ which causes the denominator to vanish. It must be kept in mind that the above was derived under the premise of the activation of detectors 1 through 4. This means that $55^0 \lesssim \theta \lesssim 125^0$; it is easily shown that the value of $\theta^{(o)}$ which would cause the denominator of Eqn. (28) to vanish is not within this range.

The fact that the deviation formulas are energy dependent makes their expressions, Eqns. (27) and (28), somewhat difficult to interpret. (It should be kept in mind that the N_i appearing in those expressions are themselves effected by gamma ray energy: the fact that μ decreases with increasing energy means, according to the relations given in Eqn. (12), that the N_i will correspondingly increase and thereby additionally influence $\delta\phi$ and $\delta\theta$.) The significance of these formulas might best be appreciated by expressing the deviations in terms of the most probable energy of a photon in a burst over the 50 - 150 keV range.

Burst spectra so far obtained appear to obey the following empirical relation[5]:

$$\frac{dn}{dE} \propto \frac{1}{E^{1.4}} .$$

The most probable energy, $\langle E \rangle$, is then calculated as

$$\langle E \rangle = \frac{\int n(E)EdE}{\int n(E)dE} , \qquad (30)$$

integrated over the afore-mentioned limits. The result for the above power law gives $\langle E \rangle$ = 96.4 keV.

The deviations given by Eqns. (27) and (28) can now be assessed at this value of energy for various detector counts. Table IV displays a variety of detector response examples along with the corresponding deviations at 96.4 keV. (The value of μ used here was 0.513 cm^{-1}.) The column headed "$(\delta\theta_{\pm})$" has been included to indicate the choice of sign used in Eqn. (28). The zero values in $\delta\phi$ and $\delta\theta$ result from the equality of N_1 and N_3, and N_1 and N_2, respectively.

DISCUSSION AND CONCLUSIONS

Burst location determinations based on detector responses have been discussed in a way that includes gamma ray energies. Equations (17) and (21), along with (23) can be used for energies in the approximate range 50 - 150 keV. Although those expressions were developed under the assumption of a burst activating detectors 1 through 4, so that at least two of these have $N_i > 0$, it is a straight-forward matter to generate the appropriate expressions in the event of activation of some other detector combination. The unit vector relationships given in Eqn. (2) indicate, e.g., that $N_7 = -N_4$ and $N_8 = -N_3$. Thus, if a burst activated detectors 1, 2, 7, and 8 it would only be necessary to replace N_3 in the expressions given by $-N_8$. N_2 need

not appear in the final expression, as was the case with N_4 in the situation discussed, since $N_2 = N_1 + N_8 - N_7$. The deviation formulas can be similarly altered.

The examples of deviations associated with the detector responses considered in Table IV assumed a particular power-law energy spectrum of the burst in order to compute the most probable energy, <E>. This quantity can easily be calculated from Eqn. (30) for any burst spectrum. If, for example, the spectrum[6] $dn/dE \propto \exp(-E/E_o)$, where E_o = 150 keV, were used then <E> = 94.5 keV. This value is only slightly different from that used in generating Table IV.

It has been demonstrated that energy considerations connected with the role of the passive absorber shielding each detector are significant if locations are to be determined to within a few degrees. A major objective of the BATSE is to make those determinations to within less than 1°. It would therefore seem necessary to incorporate the effects discussed in the present work into the overall scheme of location determinations.

The difficulties involved in obtaining angular (θ,ϕ) solutions from a set of statements of the form given by Eqn. (7) are apparent. The efforts of the present study were restricted to obtaining first order corrections to the locations given by Eqns. (24) and (25). It has been found possible to obtain the second order correction corresponding to the retention of an additional term in the expansion of the aluminum exponential expansion. The details of that calculation have not yet been fully explored, but can and should be done.

The need to impose the upper limit on the energy range as discussed here to 150 keV resulted from the detector efficiency decreasing too rapidly above this energy (as well as somewhat below it). Further investigation is clearly necessary so that the upper limit can approach the desired 200 keV value. It is not entirely clear at this point that those calculations can be performed analytically.

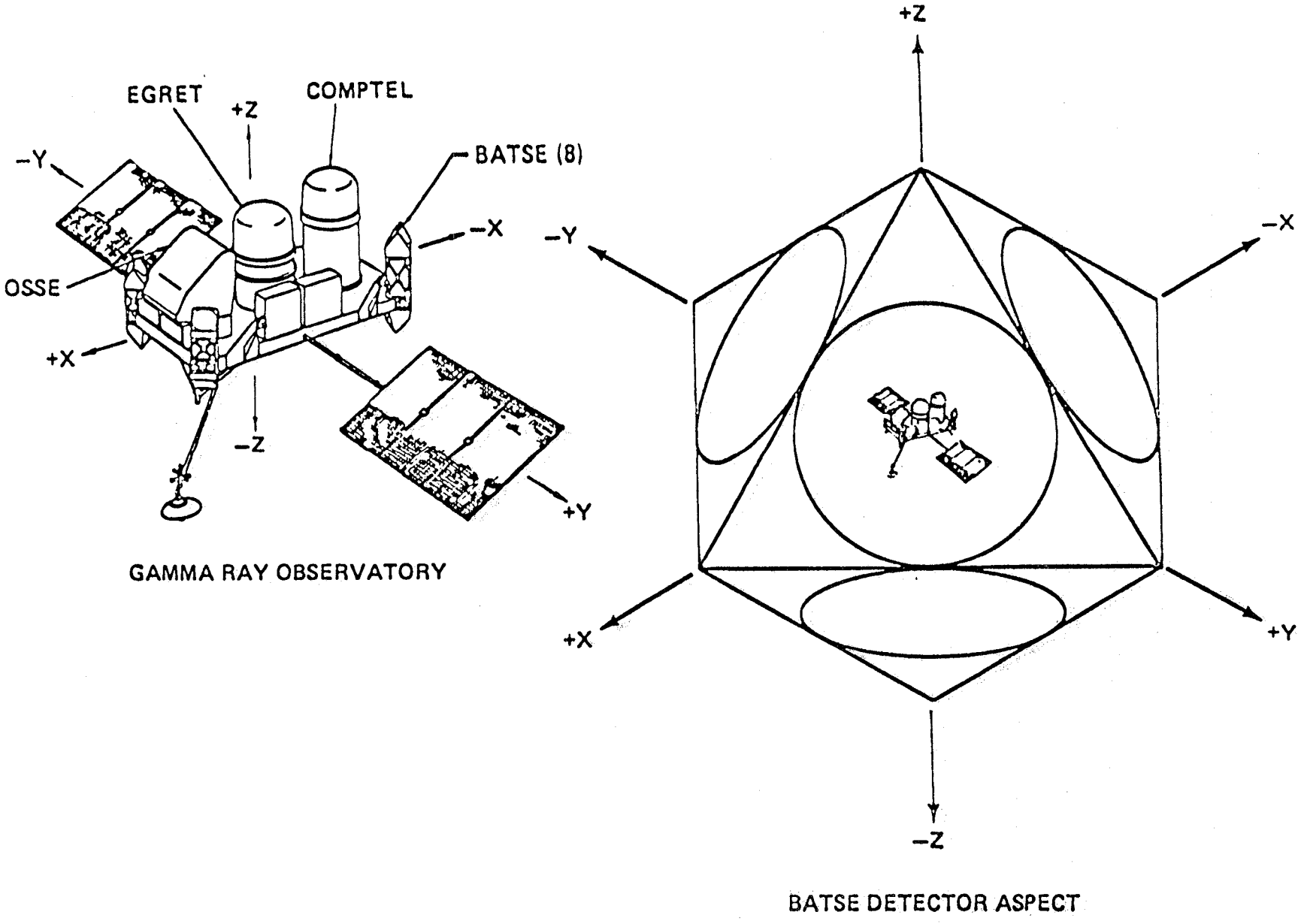


Figure 1. Gamma Ray Observatory Configuration; BATSE Octohedral Geometry

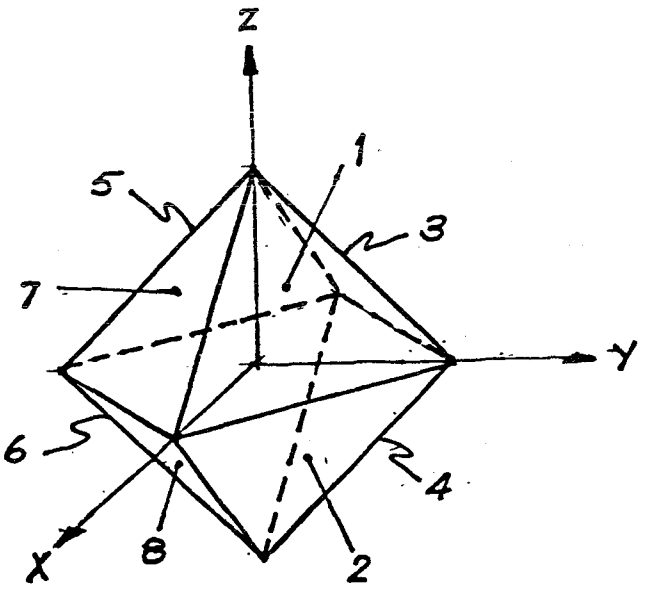


Figure 2. Plane Identification Scheme in the Octohedral Geometry

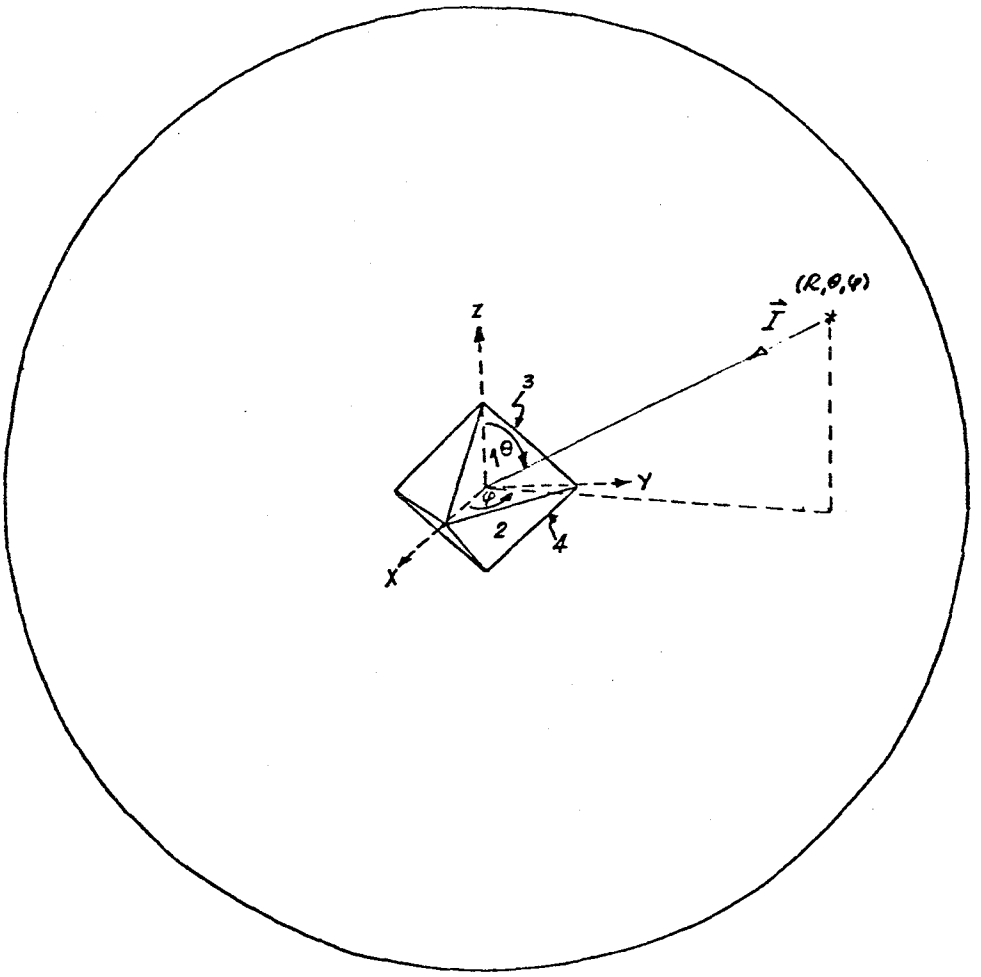


Figure 3. Burst Propagation Toward the BATSE Geometry

TABLE I

Attenuation Coefficients for Aluminum and Evaluation of Their Influence

E(keV)	$\mu(cm^{-1})$	$e^{-\mu t}$	$1-\mu t$
50	0.972	0.823	0.806
75	0.621	0.883	0.876
100	0.473	0.910	0.905
125	0.405	0.922	0.919
150	0.378	0.927	0.924
175	0.351	0.932	0.930
200	0.324	0.937	0.935

TABLE II

Attenuation Coefficients for Sodium Iodide and Evaluation of Their Influence

E(keV)	$\mu(cm^{-1})$	$e^{-\mu D}$	$(1-e^{-\mu D})$	$(1-e^{-\mu D/\cos 45^0})$
50	36.70	5.7×10^{-21}	~1	~1
75	12.85	2.6×10^{-6}	~1	~1
100	6.97	1.4×10^{-4}	~1	~1
125	3.67	9.8×10^{-3}	0.99	~1
150	2.20	6.1×10^{-2}	0.94	0.98
175	1.47	0.155	0.85	0.93
200	1.28	0.278	0.72	0.90

TABLE III

Computed Angular Deviations of Source as a Function of Energy and Fixed N_1, N_2, N_3

E(keV)	I_oA	ϕ	θ	$\phi^{(o)}$	$\theta^{(o)}$	$\delta\phi$	$\delta\theta$
50	4.57	63.58°	77.44°	56.31°	74.50°	7.27°	2.94°
75	3.96	60.73°	76.26°	56.31°	74.50°	4.42°	1.76°
100	3.75	59.61°	75.80°	56.31°	74.50°	3.30°	1.30°
125	3.67	59.11°	75.60°	56.31°	74.50°	2.80°	1.10°
150	3.64	58.92°	75.53°	56.31°	74.50°	2.61°	1.03°

TABLE IV

Computed Angular Deviations of Source for <E> and Various N_1, N_2, $N_3(x10^n)$

N_1	N_2	N_3	I_oA	$\delta\phi$	$\delta\theta$	$(\delta\theta_+)$
3	2	1	3.80	+3.59°	+1.42°	(+)
1	3	2	5.68	-1.31°	-2.27°	(-)
2	1	3	4.42	-1.57°	+1.45°	(+)
2	2	3	5.35	-1.31°	0°	
3	2	2	4.42	+1.57°	+1.45°	(+)
2	3	2	5.35	0°	-1.25°	(-)

REFERENCES

1. G. J. Fishman, et al.: BATSE/GRO Observational Capabilities, Space Science Laboratory Preprint Series No. 84-103.

2. S. S. Holt in Introduction to Experimental Techniques of High-Energy Astrophysics, H. Ogleman and J. R. Wayland, ed., NASA Office of Technology Utilization, 1970.

3. J. H. Hubbell: Photon Cross Sections, Attenuation Coefficients, and Energy Absorption Coefficients from 10 keV to 100 GeV, NBS-29, National Bureau of Standards, 1969.

4. R. D. Evans: The Atomic Nucleus, McGraw-Hill Book Company, New York, p. 717, 1955.

5. G. J. Fishman, private communication.

6. E. C. Chupp: Gamma Ray Astronomy, D. Reidel Publishing Company, Dordrecht-Holland/Boston, p. 196, 1976.

☆U.S. GOVERNMENT PRINTING OFFICE 1986—631-058/20081

End of Document