UNITED STATES AIR FORCE

SUMMER RESEARCH PROGRAM -- 1991

SUMMER FACULTY RESEARCH PROGRAM (SFRP) REPORTS

VOLUME 5A

WRIGHT LABORATORY

RESEARCH & DEVELOPMENT LABORATORIES

5800 UPLANDER WAY

CULVER CITY, CA 90230-6608

SUBMITTED TO:

LT. COL. CLAUDE CAVENDER
PROGRAM MANAGER

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

BOLLING AIR FORCE BASE

WASHINGTON, D.C.

DECEMBER 1991

UNITED STATES AIR FORCE

SUMMER RESEARCH PROGRAM -- 1991

SUMMER FACULTY RESEARCH PROGRAM (SFRP) REPORTS

VOLUME 5A

WRIGHT LABORATORY

RESEARCH & DEVELOPMENT LABORATORIES

5800 Uplander Way

Culver City, CA 90230-6608

Program Director, RDL
Gary Moore

Program Manager, AFOSR
Lt. Col. Claude Cavender

Program Manager, RDL
Claude Baum

Program Administrator, RDL
Gwendolyn Smith

Submitted to:

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

Bolling Air Force Base

Washington, D.C.

December 1991

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3 REPORT TYPE AND DATES COVERED
	9 January 1992	30 Sep 90-30 Sep 91

4. TITLE AND SUBTITLE
1991 Summer Faculty Research Program (SFRP)
Volumes 2-5b
Vol. 5A

5. FUNDING NUMBERS
F49620-90-C-0076

6. AUTHOR(S)
Mr Gary Moore

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Research Development Laboratories (RDL)
5800 Uplander Way
Culver City CA 90230-6608

8. PERFORMING ORGANIZATION REPORT NUMBER
AFOSR-TR- 92 0171

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR/NI
Bldg 410
Boll ng AFB DC 20332-6448
Lt Col V. Claude Cavender

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
UNLIMITED

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

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

During the summer of 1991 170 university faculty conducted research at Air Force laboratories for a period of 10 weeks . Each participant provided a report of their research, and these reports are consolidated into this annual report.

14. SUBJECT TERMS			15. NUMBER OF PAGES
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

PREFACE

Reports in this volume are numbered consecutively beginning with number 1. Each report is paginated with the report number followed by consecutive page numbers, e.g., 1-1, 1-2, 1-3; 2-1, 2-2, 2-3.

Due to its length, Volume 5 is bound in two parts, 5A and 5B. Volume 5A contains reports #1-30. Volume 5B contains reports #31-55. The Table of Contents for Volume 5 is included in both parts.

This document is one of a set of 13 volumes describing the 1991 AFOSR Summer Research Program. The following volumes comprise the set:

VOLUME	TITLE
1	Program Management Report
	Summer Faculty Research Program (SFRP) Reports
2	Armstrong Laboratory, Wilford Hall Medical Center
3	Phillips Laboratory, Civil Engineering Laboratory
4	Rome Laboratory, Arnold Engineering Development Center, Frank J. Seiler Research Laboratory
5A&B	Wright Laboratory
	Graduate Student Research Program (GSRP) Reports
6	Armstrong Laboratory, Wilford Hall Medical Center
7	Phillips Laboratory, Civil Engineering Laboratory
8	Rome Laboratory, Arnold Engineering Development Center, Frank J. Seiler Research Laboratory
9	Wright Laboratory
	High School Apprenticeship Program (HSAP) Reports
10	Armstrong Laboratory
11	Phillips Laboratory, Civil Engineering Laboratory
12	Rome Laboratory, Arnold Engineering Development Center
13	Wright Laboratory

1991 FACULTY RESEARCH REPORTS

Wright Laboratory

Report Number	Report Title	Author
	Wright Laboratory (cont.)	
35	(Not Used)	
36	Evaluation of the Analytical Design Package (ADP) for Frameless Transparency Program	Dr. Joe Chow
37	A Massively Parallel Algorithm for Large-Scale Nonlinear Computation	Dr. Shurit Dey
38	Decentralized Control of Flexible Structures with Uncertain Interconnections	Dr. George Flowers
39	Computational Prediction of Static and Rolling of an Aircraft Tire on a Rigid Surface as Prelude to Wear Studies	Dr. Manjriker Gunaratne
40	Deply of Laminated Panels with Perforation Due to Impact	Dr. David Hui
41	Monitoring of Damage Accumulation for the Prediction of Fatigue Lifetime of Cord-Rubber Composites	Dr. Byung-Lip Lee
42	In-Flight Structural Combat Damage Detection and Evaluation for Enhanced Survivability of Military Aircraft	Dr. Vernon Matzen
43	A Report on Robust Control Design for Structured Uncertainties	Dr. Jenny Rawson
44	Experimental Investigation of the Influence of Constrained-Layer Damping Treatment on Parametric and Autoparametric Resonances in Nonlinear Structural Systems	Dr. Lawrence Zavodney
	Materials Laboratory (MAT)	
45	Semi-Empirical Self-Consistent Calculations of GAAS Surface Restructuring	Dr. Shashikala Das
46	X-ray Crystallographic Studies of Model Systems for Rigid Rod Polymers and Molecules with Potential Nonlinear Optical Properties	Dr. Albert Fratini
47	Computer-Aided Process Planning for Grinding Operations	Dr. Rakesh Govind
48	Transmission Electron Microscopy of Deformation at the Interface of Ti-6-4//SCS6 SiC Fiber Composites	Dr. Warren Moberly
49	Ultrasonic Beam Propagation: Diffractionless Beams and Beams in Anisotropic Media	Dr. Byron Newberry
50	Geometric Reasoning for Process Planning	Dr. Joseph Nurre
51	Synthesis and Characterization of Chiral Mesogens for Use in Cyclic Siloxane Liquid Crystalline Materials	Dr. Steven Pollack

Report Number	Report Title	Author
	Wright Laboratory (cont.)	
52	Ab Initio Computational and NMR Relaxation Time Investigations of Rotational Barriers and Chain Dynamics in Perfluoropolyalkylethers	Dr. Martin Schwartz
53	On Photoreflectance Spectra from Two Dimensional Electron Gas in GaAs/AlGaAs Heterojunctions	Dr. Michael Sydor
54	Creep Behavior of a Fine-Grained $Y_3Al_5O_{12}$+$YAlO_3$ (18 Vol. %) Material	Dr. Jeffrey Wolfenstine
55	Elastic Moduli of Fiber Reinforced Brittle Matrix Composites with Interfacial Debonding	Dr. Fuh-Gwo Yuan

INFLUENCE OF OPERATING TEMPERATURE ON QUENCH AND STABILITY OF OXIDE HIGH-T_C SUPERCONDUCTORS

M.K. Chyu
Department of Mechanical Engineering
Carnegie Mellon University
Pittsburgh, PA 15213

C.E. Oberly
Aero Propulsion and Power Directorate
Wright Laboratory
Wright Patterson Air Force Base, OH 45433

ABSTRACT

This paper examines the influence of operating temperature, ranging from 20K to 80K, on the stability and normal zone propagation in a silver sheathed, YBCO superconductor tape. The distributions of temperature and heat generation are obtained numerically by solving a transient, two-dimensional energy equation with temperature-dependent properties and a current-sharing model. The present results suggest that a 20K operation is considerably more stable than its 80K counterpart. In addition, during a pulse-induced quench zone propagation, most of the ohmic heating is generated in the YBCO superconductor for a 20K operation. On the other hand, the silver sheath generates most of the heat for an 80K operation. Imposing transverse cooling significantly promotes stability and reduces normal zone propagation velocity. However, it has little influence on the instantaneous rise in local temperature during a disturbance. Such a temperature spike, largely caused by the low thermal diffusivity of YBCO, may exceed YBCO melting temperature.

NOMENCLATURE

A	cross-section area of superconductor
AG	Silver Sheath
As	cross-section area of stabilizer
C	specific heat
D	tape thickness, 1mm
E	volumetric energy of disturbance
g	volumetric heat generation
h	heat conductance or heat transfer coefficient
I	current
I_c	critical current at given temperature
I_{op}	operating current
J	current density
J_c	critical current density of superconductor at given temperature
J_{co}	critical current density at operating temperature
J_{op}	operating current density
J_s	current density in stabilizer
k	thermal conductivity
L	conductor length
SC	superconductor
T	temperature
T_b	ambient coolant temperature and tape initial temperature
T_c	critical temperature at given current density
T_{co}	critical temperature at zero current, 92K
t	time
v	normal zone propagation velocity
x	coordinate in principal current flowing direction, see Fig. 1
y	coordinate in transverse direction, see Fig.1

Greek Symbol

ε	electrical field strength
λ	volume fraction of superconductor in a composite
γ	density
θ	dimensionless temperature, $(T-T_b) / (T_{co} - T_b)$.
ρ	electric resistivity

INTRODUCTION

The discovery of high-temperature, ceramics based, superconductors (HTSC) at 90K has a significant implication that superconducting magnets can be operated in a liquid-nitrogen cooled environment. Despite enormous research progress since the HTSC discovery, this speculation has evolved to be a long-range goal. This is due mainly to difficulties encountered in material processing and thermally assisted flux flow at elevated temperatures. The HTSC has been demonstrated at higher critical current and very high upper critical fields at 20K with liquid hydrogen cooling. This has led to the suggestions that HTSC may be useful in high-field, high-current applications of low temperatures (Collings, 1989; Oberly et al., 1990). For advanced aeronautical and space propulsion, an effective superconducting power system cooled by liquid hydrogen, which is typically the fuel, is particularly desirable (Oberly et al., 1991).

One of the critical issues relating to the selection of operating temperature for a power system employing HTSC is quench protection. Partly because of large thermal margins, HTSC operated in a low temperature range are far more stable against external disturbance than their low-temperature counterparts. Nevertheless, once a disturbance is sufficiently strong to trigger a quench, the normal zone propagation is extremely slow (Laquer et al., 1989; Chyu and Oberly, 1991), as a large amount of resistive heating is generated and retained in a small region. This detrimental effect has posed a great challenge to the design of protection systems for coils employing HTSC.

The primary objective of this study is to examine the effects of operating temperature on the nature of normal zone growth in a HTSC coil tape. The tape configuration of present interest is a YBCO superconductor sheathed by a layer of silver on each side, as shown in Figure 1. The purpose of using a silver sheath is two-fold: first, it strengthens the overall device structure; and second, it is a passive quench protector (stabilizer) which shares excessive current when the superconductor becomes intrinsically unstable. This latter aspect near liquid-nitrogen temperature has recently been studied by Chyu and Oberly (1991) using a two-dimensional numerical model. These results have demonstrated that a viable quench analysis must consider the details in composite geometry with a proper current sharing model, rather than the conventional, volumetric-averaging approach. Such a numerical methodology in conjunction with temperature varying properties will be used for the present analysis.

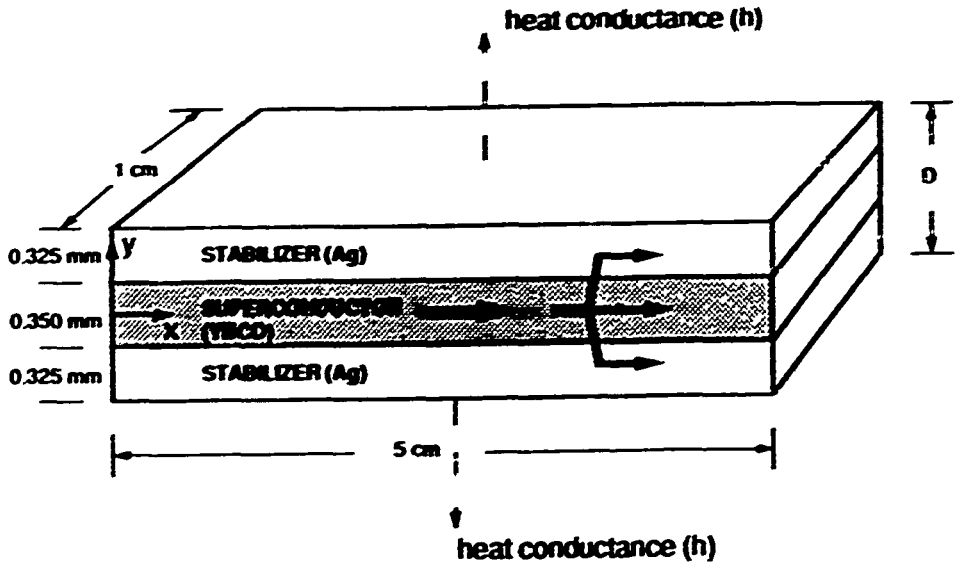


Figure 1. Tape Geometry

CRYOPHYSICAL PROPERTIES

The cryophysical properties for YBCO superconductors are strongly dependent on temperature. Table 1 lists sample data at 20K, 40K and 80K based on several different reports: density (Heremans et al., 1988), specific heat (Collocott et al., 1987; Van Miltenburg et al., 1987), thermal conductivity (Uher and Kaiser, 1987; Morelli et al., 1987) and normal electric resistivity (Collins, 1989). The values of physical properties generally increase with temperature, except for thermal conductivity which possesses a local maximum at about one-half T_c or near 40K. Nevertheless the value of this maximum thermal conductivity varies with different measurements; Jenzowski et al (1987) have reported a peak around 40K nearly doubled the value of that listed in Table 1. The cryoproperties of silver are taken from several standard references: density and thermal conductivity (Powell et al., 1966), specific heat (White, 1979) and electric resistivity (Hultgren et al. 1963). In the present computation, properties for both YBCO and silver at intermediate temperatures are evaluated based on linear interpolations between two adjacent values listed in Table 1. The nominal variations of density due to volumetric thermal expansion are neglected. This approach not only preserves the primary feature of temperature dependency but also significantly reduces computing time and programming complexity. Besides, a literature review reveals that complete properties correlations for YBCO in the range of 20K - 80K remain non-existent.

	YBCO			AG		
	20K	40K	80K	20K	40K	80K
γ [kg/m3] x 10^{-3}	6.37	6.37	6.37	10.5	10.5	10.5
C [J/kg-K]	8.4	51.2	162	15.8	76.1	166.2
k [W/m-K]	2.8	4.2	3.9	5100	1050	471
ρ [ohm-m] x 10^{6}	3.6	4.1	5.0	2.7 x 10^{-6}	6.0 x 10^{-5}	2.0 x 10^{-4}
α [m2/s] x 10^{6}	52	12.9	3.78	30700	1310	270

Table 1 Cryophysical Properties

NUMERICAL MODELLING AND CURRENT SHARING MODEL

The temperature distribution in the tape including both the superconductor and stabilizer is governed by the transient heat equation in two-dimensional form; i.e.,

$$\gamma C \frac{\partial T}{\partial t} = \frac{\partial}{\partial x}(k\frac{\partial T}{\partial x}) + \frac{\partial}{\partial y}(k\frac{\partial T}{\partial y}) + \rho J^2 \qquad (1)$$

with initial and boundary conditions

$T(x,y,0) = T_b$

$\partial T(0,y,t)/\partial x = 0$

$\partial T(L,y,t)/\partial x = 0$

$\partial T(x,0,t)/\partial y = 0$

$\partial T(x,D/2,t)/\partial y = -(h/k)(T(x,D/2,t) - T_b)$

Note that all properties are temperature dependent, which represents major nonlinearity to the problem and thus requires numerical iteration in the spatial domain for a given time. The representation of h, an effective thermal conductance, is either a convective heat transfer coefficient for a face-cooled tape or the inverse of the total thermal resistance between the tape located generally inside a multiple winding device and ambient coolant. Due to symmetry, only one-half the tape is discretized in a 50 by 10 grid and solved by the finite-

difference scheme of Patankar (1980). The grid independence as well as computational optimization is verified prior to the actual calculation. To initiate a normal zone, an energy pulse is supplied to a small region near the left edge of superconductor; i.e., $0 < x/D < 1$ and $0 < y/D < 0.175$. A converged solution over a 200 ms duration with a 10 ms time-increment needs about 5-minute CPU time of a MicroVax III work station.

For problem closure, a current sharing model is required to link the temperature and amount of current flowing in both the superconductor and the stabilizer. As described earlier by Chyu and Oberly (1991), this is done by assuming (1) the superconductor carries the critical current corresponding to the local temperature T, and (2) a linearly inverse relation between J_c and T; i.e., for $T_c < T < T_{co}$ over an entire cross-section of the superconductor,

$$J_c(T)/J_{co} = (T_{co} - T) / (T_{co} - T_b) \qquad (2)$$

where $T_{co} = 92K$, $J_{co} = 1.0 \times 10^5$ A/m^2 are assumed for the present case. Further assuming that the excessive current branched out of the superconductor is uniformly distributed over the stabilizer's cross section, thus the current density in the stabilizer can be expressed by

$$J_S = (I_{op} - \int_{sc} J_c(T)\, dA) / A_S = \int_{sc} (J_{op} - J_c(T))\, dA / A_S \qquad (3)$$

Since the currents in both superconductor or stabilizer are driven by the same electrical field strength ε along the principal direction of current flow (see Fig.1, x-axis),

$$\varepsilon = \rho_S J_S = \rho J_c \qquad (4)$$

This equation then gives the resistivity of the superconductor in the current sharing state, ρ.

When temperature of the whole superconductor is lower than the critical temperature corresponding to a given operating current density with a given field strength; i.e, $T \le T_c(J_{op})$, the conductor is perfectly superconducting and carries the entirety of the operating current I_{op}. This implies no heat generation and $\rho = 0$ over the entire composite domain. As a disturbance is sufficiently strong to create a normal zone primarily via magnetic flux instability in the superconductor, both the magnitude of ohmic heating and occurrence of current sharing depend on the overall current carrying capability in the

superconductor. If $T_c < T$ implying $J_c(T) < J_{op}$ locally, but the operating current remains smaller than the critical current ($I_{op} < I_c$) over the whole cross-sectional area, the superconductor redistributes its current flowing completely through the superconducting portion without heat generation.

When $T > T_{co}$ over an entire cross section, the superconductor becomes fully normal, and the current distribution in the magnet composite follows the general Kirchhoff Law. Since the normal resistivity of YBCO is about three orders of magnitude higher than that of the silver stabilizer, most of the operating current is carried by the stabilizer.

RESULTS AND DISCUSSION

Although other cases have been modeled, the sample results presented here consist of three different initial (bath) temperatures; i.e. T_b = 20K, 40K and 80K, for an operating current density equal to 90% of J_{co}. The influence of operating current density on both quench initiation and normal zone propagation reveals virtually the same characteristics as those reported by Chyu and Oberly (1991) with temperature independent properties at 77K. Three different levels of initial disturbance has been imposed, E = 3.5 mJ, 35 mJ and 350 mJ, which respectively corresponds to 1.0×10^8, 10^9 and 10^{10} J/m^3 over the disturbed region. Despite that, based on the theory with volumetric averaging properties (Wipf, 1978; Wilson, 1983), any of these disturbances is sufficiently strong to induce an adiabatic MPZ growth, two cases (T_b = 20K and 40K for E = 3.5 mJ), however, recover their full superconductivity in less than 50 ms. For cases with 350 mJ (or 10^{10} J/m^3) initial disturbance, temperature spikes in the perturbed region are found to exceed the melting temperature of YBCO, signifying permanent device damage. This overall implies that, for the present tape configuration and operating-current level, rigorous normal zone propagation can only take place when the triggering energy is on the order of 10 to 100 mJ. Lower than this range, the tape is remarkably stable especially for low-temperature operation; on the other hand, however, a strong energy pulse can result in a nearly spontaneous local melting inside the superconductor disregarding a large cooling-margin imposed externally by convection. The relatively low thermal diffusivity in YBCO, though increases nearly ten-fold from 80K to 20K, is responsible for this phenomenon.

Figures 2 to 6 present the results concerning the superconductor temperature and resistive heating for the three different operating temperatures. All cases are subjected to the same initial disturbance of 35 mJ and without transverse cooling (h = 0). The particular

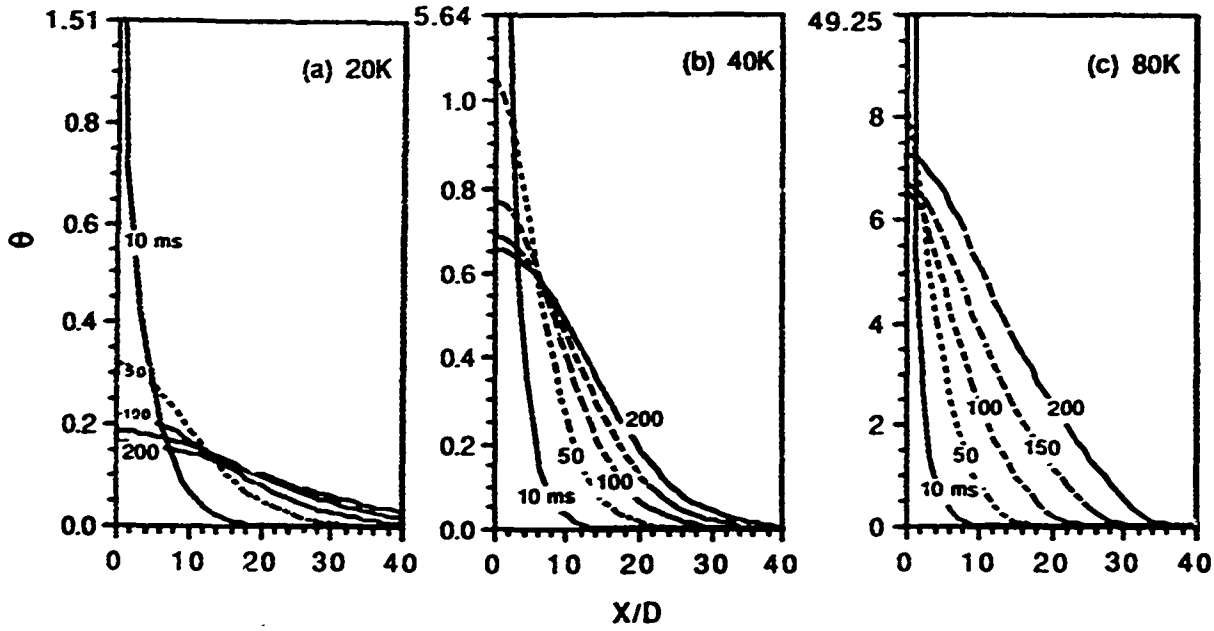


Figure 2. Distribution of Temperature in YBCO, h = 0, E = 35 mJ

time-resolved temperature distribution shown in Figure 2 is that along the tape center-line (y = 0); however, this is typical for the entire tape composite as the transverse temperature variation is virtually negligible when external cooling is absent. To facilitate a sensible comparison, the temperature is scaled to a dimensionless form, $\theta = (T-T_b) / (T_{co} - T_b)$. Note that, based on the present model for $J_{op}/J_{co} = 0.9$, the current sharing occurs only when $0.1 < \theta < 1.0$. For all cases, a sharp rise in temperature exists near the disturbed region at the end of energy-pulse duration, t = 10 ms. Because of different cooling margin, such a rise increases with the operating temperature; i.e, 129K, 333K and 671K for T_b = 20K, 40K and 80K, respectively. During the next 50 to 100 ms, the high temperature smears out and hot zone expands due to a combined effect of heat diffusion and excessive ohmic heating in the superconductor. While the smearing phenomenon continues to prevail for the other two cases, the temperature over the entire conductor for the case of T_b = 80K rises again after 100 ms, signifying an eventual quench being underway. For T_b = 20K and 40K, it is not conclusive if a complete quench will indeed occur based solely on the temperature history during the first 200 ms period. To resolve this issue, it requires additional information concerning energy balance, as delineated later.

Figures 3 to 5 display the results of volumetric heat generation in both superconductor (SC, sub-figure a) and stabilizer (Ag, sub-figure b), which are distinctly different among the three cases studied. One notable feature, for T_b = 80K, is the wave-

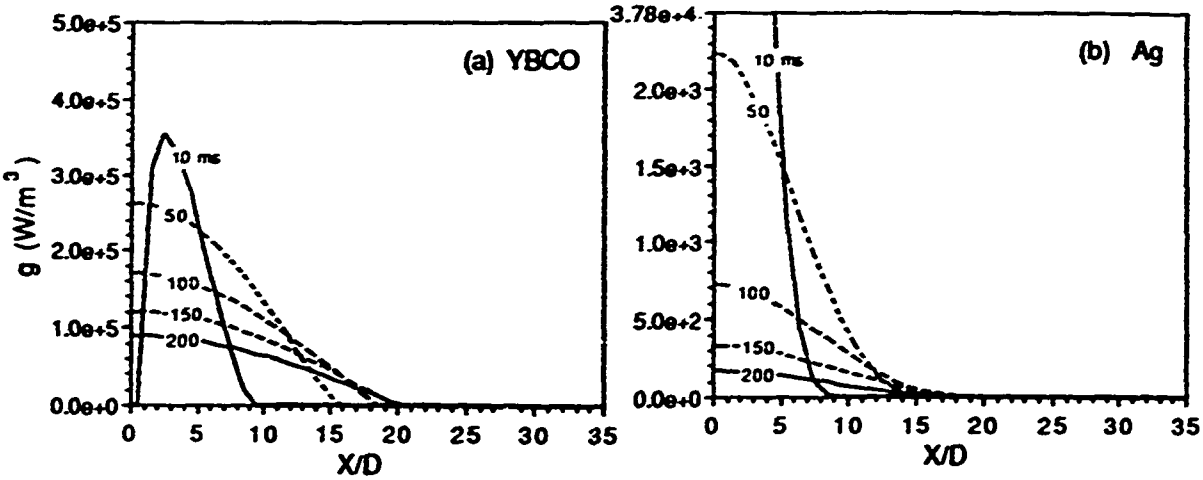


Figure 3. Distribution of Heat Generation for T_b = 20K, h = 0, E = 35 mJ

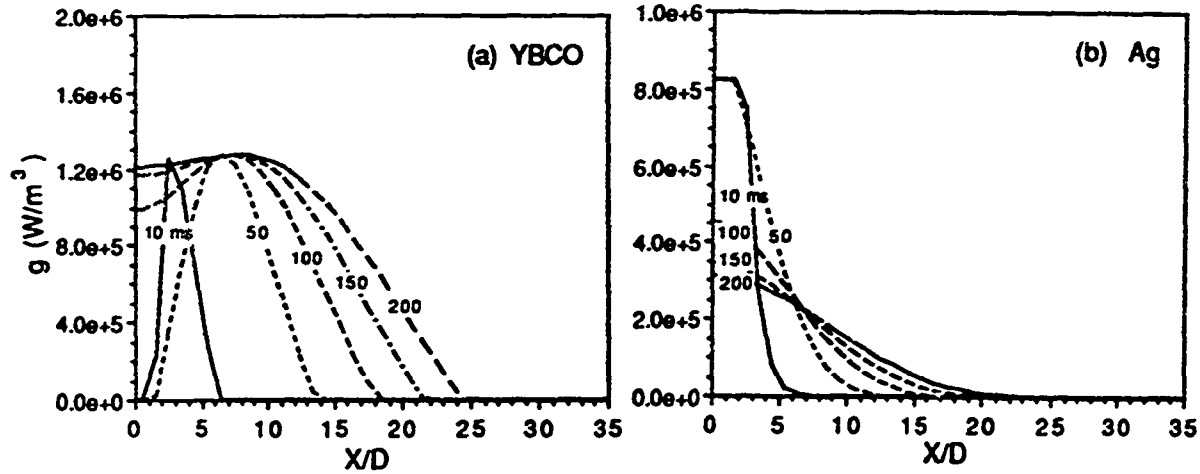


Figure 4. Distribution of Heat Generation for T_b = 40K, h = 0, E = 35 mJ

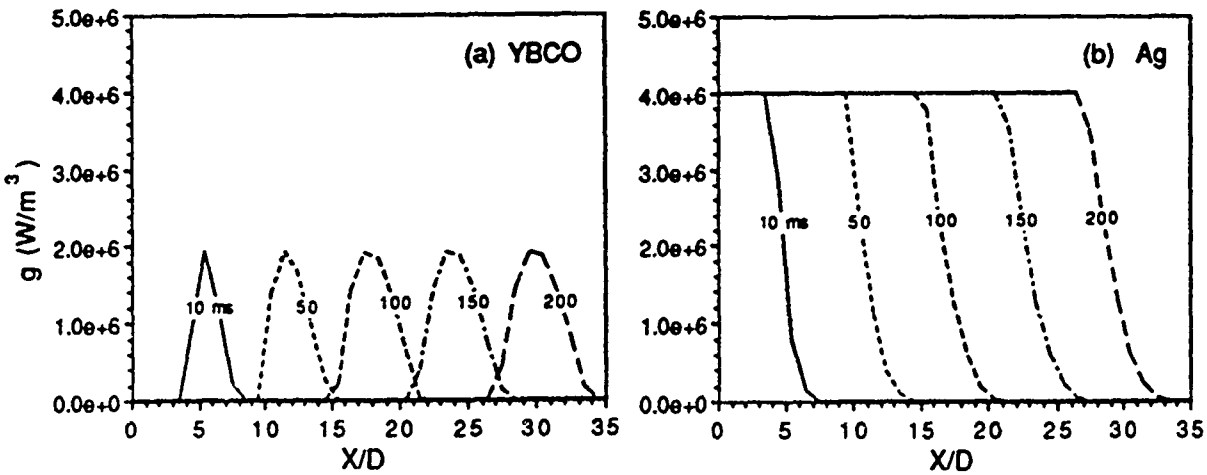


Figure 5. Distribution of Heat Generation for T_b = 80K, h = 0, E = 35 mJ

alike heat generation near the front edge of a normal zone (Figure 5a), and each wave front is preceded by a much higher ohmic heating in the silver stabilizer. A similar finding has been reported at 77K operating range by Chyu and Oberly (1991). Since heat generation in the superconductor occurs in the region of current sharing; i.e., $0 1 < \theta < 1.0$, as previously mentioned, the temperature distribution shown in Figure 2c implies such a condition only exists near the edge of propagating normal zone. Inside the normal zone, virtually all the operating current is flowing through the stabilizer, accordingly, the ohmic heating there becomes very substantial. As a sharp contrast, the value of θ shown in Figure 2a and 2b for the 20K and 40K operating temperatures lies primarily in the range of 0.1 and 1, and heating in the superconductor is predominant, as disclosed in Figures 3 and 4. The envelope of a propagating normal zone is centered around the location of initial disturbance. To a limited extent, the normal zone appears to propagate slower with a decrease in operating temperature.

The important feature concerning the heat partition can also be depicted from Figure 6, which gives the time-evolved, component-resolved total heat generation for different cases. It is interesting to note that ohmic heating in the silver sheath accounts for more than 90% of heat generation for a pulse-disturbed YBCO tape operated near 80K; a completely opposite trend exists when it is operated at 20K. For the intermediate temperature at 40K, while both heating components are comparable in the initial stage, the superconductor heating will be overwhelming eventually. This overall presc.. ... ...rtant insight toward quench protection for HTSC low-temperature operation, where energy removal from the superconductor and its sheath must be targeted accurately and effectively. From the

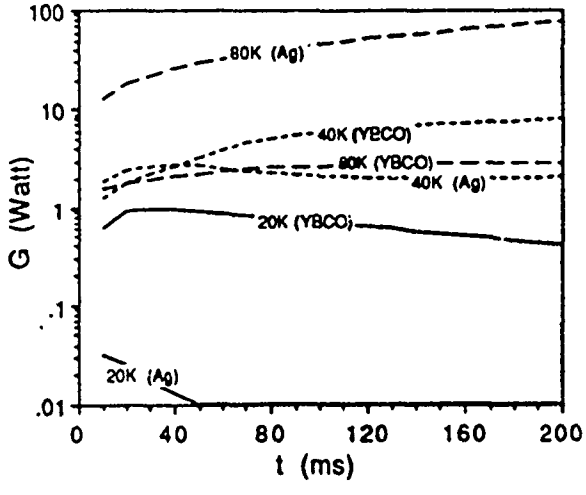


Figure 6. Total Heat Generation; h = 0, E = 35 mJ

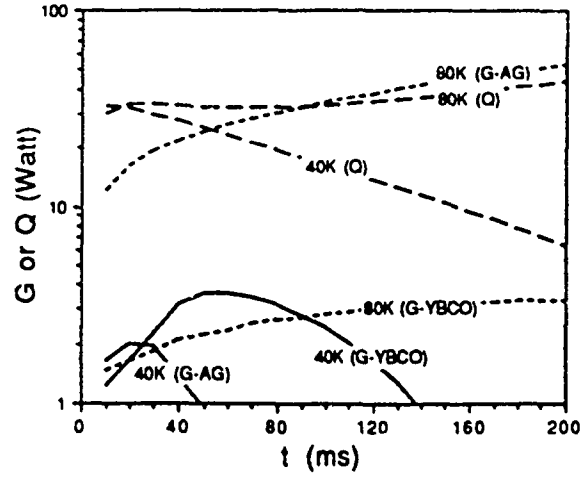


Figure 7. Total Heat Generation and Cooling; h = 50 W/m^2-K, E = 35 mJ

standpoint of energy balance, quench will eventually prevail if the total heat generation over the entire conductor , which is the sum of heating from the superconductor and the stabilizer, never ceases. Apparently the case of 80K is under a stage of thermal run-away, since the total amount of heat generation always increases with time. On the other hand, for the 20K operation, the negative slope of G vs t implies that a complete recovery of superconductivity is underway. As mentioned earlier, this information is somewhat concealed if viewed from temperature distribution alone (Figure 2a). Although it is not very obvious based on the plotted data, the actual sum of heat generation for the 40K case also has an increasing trend as time progresses. This will lead to an eventual quench, but with a much longer time than that of the 80K case.

Imposing a transverse cooling ($h \neq 0$) generally induces two primary effects: (1) promotes stability, and (2) reduces normal zone propagation speed. These features have been clearly observed in the present calculation. As shown in Figure 7, a nominal level of cooling at $h = 50$ W/m^2-K reverts the case of 40K from adiabatically unstable to fully recovered superconductivity after about 140 ms. This is evidenced by the fact that the total amount of heat removal (Q) induced by transverse cooling is always exceeding the overall heat generation in the tape composite. Except for heat generation during the first 60 ms, both magnitudes decrease with time as the tape is regaining a thermal equilibrium with the liquid hydrogen bulk. For the 20K case, which is adiabatically stable as discussed earlier, the present transverse cooling further stabilizes the tape without any sensible heating even in the first 20 ms.

With the same cooling condition at $h = 50$ W/m^2-K, the 80K case, however, remains unstable according to the results shown in Figure 7. Although the amounts of heat generation in both superconductor and stabilizer are reduced to about one-tenth of their adiabatic counterparts, such a nominal level of cooling, after all, is insufficient to overcome the resistive heating. In fact, the amount of energy generated in the silver sheath alone starts to surpass that removed by cooling at approximately 100 ms. Accompanying with less amount of energy involvement, the normal zone expands about 30% slower as compared to the adiabatic case, 0.13 m/s vs. 0.18 m/s averaged over the entire 200 ms period. It is worthy of mentioning that the feature of temperature spike at the end of pulse disturbance ($t = 10$ ms, shown in Figure 2c) is virtually unaltered by the transverse cooling imposed. In fact, with a cooling level at $h = 100$ W/m^2-K, such a peak temperature is reduced by less than 3% compared to the adiabatic case. Hence the possibility of local melting induced by a

strong disturbance still exists, and to enhance the external cooling solely may not be a viable remedy.

CONCLUSIONS

The effects of operating temperature, ranging from 20K to 80K, on the normal zone propagation induced by a pulse disturbance in a silver-sheathed, YBCO tape are examined using a transient, two-dimensional, finite-difference computation. The present numerical model which incorporates temperature-dependent cryophysical properties reveals detailed information on the distributions of heat generation and removal over the entire tape. Although actual phenomena may vary with different current carried and field strength, a decrease in operating temperature generally promotes conductor stability with an additional thermal margin. The present results show that the tape composite is remarkably stable in a liquid hydrogen environment at 20K. However, all cases are susceptible to large disturbances ($> 10^{10}$ W/m^3), as local melting may occur due mainly to the low thermal diffusivity of YBCO. For a 20K operation without transverse cooling, when a pulse disturbance is sufficiently strong to induce a normal zone propagation, the propagation velocity is extremely low (~ 0.1 m/s) and the YBCO superconductor generates most of the heat. If the tape operated at 80K, near the liquid nitrogen temperature, the normal zone propagates about 50% faster with most of the heating generated in the silver stabilizer. Imposing transverse cooling significantly promotes conductor stability and reduces normal zone expansion. However, even with an enhanced cooling, the potential danger of local melting due to temperature spike near the disturbed region remains unaltered. These findings overall may lead to different design criteria for protection of superconductors under different operating temperatures.

REFERENCES

Chyu, M.K. and Oberly, C.E., 1991, "Numerical Modeling of Normal Zone Propagation and Heat Transfer in a Superconducting Composite Tape," *IEEE Trans. MAG.*, Vol. 27, pp. 2100-2103.

Collings, E.W., 1989, "Conductor Design with High-Tc Ceramics - A Review," *Proc. Int Symposium on Superconductivity*, Tsukuba-Ibaraki, Japan, Nov. 14-17, 1989.

Collocott, S.J., White, G.K., Dou, S.X., Williams, R.K., 1987, "Thermal Properties of the High-Tc Superconductors $La_{1.85}Sr_{0.15}CuO_4$ and $YBa_2Cu_3O_7$," *Physical Review B*, Vol. 36, No. 10, pp. 5684-5685.

Heremans, J., Morelli, D.T., Smith, G.W., and Strite III, S.C., 1988, "Thermal and Electronic Properties of Rare-Earth Ba2Cu3Ox Superconductors," *Physical Review B*, Vol. 37, No. 4, pp. 1604-1610.

Hultgren, R., Orr, R.L., Anderson, P.D. and Kelley, K.K., 1963, *Selected Values of Thermodynamic-Properties of Metals and Alloys*, John Wiley and Sons, New York.

Jezowski, A., Mucha, J., Rogacki, K., Horyn, R., Bukowski, Z., Horobiowski, M., Rafalowicz, J., Stepien-Damm, J. Sulkowski, C., Trojnar, E., Zaleski, A.J., and Klamut, J., 1987, " Thermal Conductivity and Electrical Resistivity of the High-T_c Superconductor $YBa_2Cu_3O_{9-\delta}$," *Physics Letters A*, Vol. 122, No. 8, pp. 431-433.

Laquer, H.L., Edeskuty, F.J., Hassenzahl, W.V. and Wipf, S.L., 1989, "Stability Projections for High Temperature Superconductors, " *IEEE Trans. MAG.* Vol. 25, No. 2, pp. 1516-1519.

Morelli, D.T., Heremans, J. and Swets, D.E., 1987, "Thermal Conductivity of Superconductive Y-Ba-Cu-O," *Physical Review B*, Vol. 36, No. 7, pp. 3917-3919.

Oberly, C.E., Gooden, C.E., Lenard, R.X., Sarkar, A.K., Maartense, I., and Ho, J.C., 1990, "Principles of Application of High Temperature Superconductors to Electromagnetic Launch Technology," *IEEE Trans. MAG.*, Vol. 27, No. 1, pp. 509-514

Oberly, C.E., 1991, "Implication of High Temperature Superconductor for Power Generation," Presented in 1991 CEC/ICMC conference, Huntsville, June 11-14, 1991.

Patankar, S.V., 1980, *Numerical Heat Transfer and Fluid Flow*, Hemisphere.

Powell, R.W., Ho, C.Y. and Liley, P.E., 1966, *Thermal Conductivity of Materials*, National Standard Reference Data Series - 8, National Bureau of Standards.

Uher, C. and Kaiser, A.B., 1987, "Thermal Transport Properties of $YBa_2Cu_3O_7$ Superconductors," *Physical Review B*, Vol. 36, No. 10, pp. 5680-5683.

Van Miltenburg, J.C., Schuijff, A., Kadowaki, K., Van Sprang, M., Koster, J.Q.A., Huang, Y.K., Menovsky, A.A., and Barten, H., 1987, "Specific Heat Measurements of High-T_C Superconductor $YBa_2Cu_3O_7$ Between 78K and 260K," *Physica*, Vol. 146B, pp. 319-323.

White, G. K., 1979, Experimental Techniques in Low Temperature Physics, Oxford: Clarendon Publication, Oxford, UK.

Wilson, M.N., 1983, *Superconducting Magnets*, Clarendon Press, Oxford.

Wipf, S.L., 1978, " Stability and Degradation of Superconducting Current-Carrying Devices," *LASL Report, LA-7275*.

TURBULENT LENGTH SCALE MEASUREMENTS IN AXISYMMETRIC SUDDEN EXPANSION USING LDV

Dr. Richard D. Gould

Mechanical and Aerospace Engineering
North Carolina State University
Raleigh, NC 27695

ABSTRACT

Successful two-point velocity correlation measurements were made in the anisotropic flow field of an axisymmetric sudden expansion. Both longitudinal and lateral spatial correlations were measured. The integral length scales were estimated and compared with those obtained from autocorrelation measurements in conjunction with Taylor's hypothesis. The agreement between these two methods was poor and it is believed that the spatial correlation measurements give more reliable results.

INTRODUCTION

Very few two-point velocity correlation measurements have been made using laser Doppler velocimetry(LDV). This is primarily due to the difficulty in obtaining suitable optical access when the laser probes are separated by large distances and the requirement that two independent single component LDV systems must be available to make general two point measurements. One LDV probe volume must be movable relative to the other in a very precise fashion. In addition, collection of scattered light from each probe volume must occur at all probe locations which can be difficult due to space limitations, possible signal cross -talk and stray reflections from windows and lens. Autocorrelation measurements using LDV, however, are numerous due to the fact that a single component LDV system with time recording capability is all that is required. Theoretical discussions of turbulent correlation functions(spatial and auto) and their physical interpretation can be found in the texts by Batchelor (1953),

Bradshaw (1971), Tennekes and Lumley (1972), Hinze (1975), and Townsend (1976) to name a few.

Pfeifer (1986) has written a rather complete review paper on the topic of correlation measurements using LDV and thus is a good general reference. Besides this paper the author knows of only two other papers where spatial correlation measurements using LDV are made. This is not to say that others do not exist. In the first, lateral spatial correlation measurements at one point on the centerline of fully developed pipe flow ($Re_D = 11800$ based on centerline velocity) using LDV were made by Fraser et al. (1986). An elongated probe volume($\lambda = 514.5\mu m$) was used so that velocity measurements separated by as much as 9 mm in the radial direction could be made. Two photo detectors oriented 90° from the forward scattering direction were used to collect the signals. The apertures of the two photo detectors were mounted on a traversing mechanism and thus determined the separation distance of the measurement points. Correlation measurements at separation distances closer than 1 mm were not possible with this system due to signal cross-talk problems. In the second, Absil (1988) made lateral spatial correlation measurements using a single LDV ($\lambda = 514.5\mu m$) at three radial positions at a plane 125 diameters downstream of a circular cylinder. The probe volume was 600 μm in diameter and 31 mm in length. The signals were detected using a setup similar to the one mentioned above. Autocorrelation measurements were also made in this study.

EXPERIMENTAL APPARATUS

An axisymmetric sudden expansion flow geometry was produced by joining a 3.5 m long entry pipe having a 101.6 mm (4 in.) inside diameter to a 152.4 mm(6 in.) inside diameter clear acrylic test section. The step height for this geometry was 25.4 mm(1 in.). The entry pipe and sudden expansion face were mounted on a movable table and thus could be positioned at various axial locations in the rigidly fixed test section as shown in Figure 1. This arrangement allowed for measurements at various downstream positions within the sudden expansion flow field without having to move the LDV probe volume location in the axial direction. The face of the sudden

expansion is moved to place the probe volume at a new x/H location. High quality optical access can be limited to a small window with this apparatus. A felt gasket was used to seal the small gap between the sudden expansion face and the inside diameter of the test section. A fully developed turbulent velocity profile was present at the inlet to the sudden expansion. Air was provided by well regulated shop air compressors and was monitored using a calibrated orifice plate located upstream of a large settling chamber which precedes the entrance pipe. Flat quartz windows 50 mm $\times$ 152 mm $\times$ 3.2 mm(2 $\times$ 6 $\times$ .125 in) where mounted in flanges on both sides of the 152.4 mm diameter test section such that the inner flat surfaces where flush with the inside diameter of the test section.

Two TSI single component dual-beam LDV systems, both operating in backscatter mode, where used in this study. Both systems were oriented to measure the axial velocity component on the diameter of the test section as shown in Figure 2. The stationary LDV system was adjusted so that the probe volume was located on the diameter of the test section($z = 0$) and at the required axial, x, and radial, r, measurement location. Once this position was found the LDV system was locked in place. The 514.5μm laser line from a Model 2025 Spectra Physics argon ion laser was used in this system. A Bragg cell shifted the frequency of one beam by 40 MHz causing the fringes to move in the downstream direction. Fringe spacing and half-angle were measured and found to be 1.886 μm $\pm$.006 and 7.838° $\pm$ .025, respectively. A second LDV system(fiber optic head), mounted on a precision xyz positioning table with resolution of $\pm$2.5μm in each axis, was located on the opposite side of the test section. The 488μm laser line from a Model 165 Spectra Physics argon ion laser was used in this system. A frequency shift of 40 MHz was used causing the fringes of this system to move in the upstream direction. Fringe spacing and half-angle were measured and found to be 1.728 μm $\pm$.006 and 8.117° $\pm$ .025, respectively. Both LDV systems employed $\times$ 3.75 beam expansion optics and gave probe volumes approximately 60μm in diameter and 450μm in length. A 20μm diameter pinhole mounted on a fixture supported on a spare test section window was used to find the position where both

laser beam probe volumes overlap. This fixture was used to ensure that both beams overlapped before each testing sequence. Specially designed beam blocks were fabricated to block reflections from the LDV focusing lens(they face one another) and from test section windows. Narrow bandpass filters were placed in front of each photomultiplier tube to eliminate cross-talk between the two channels. Two TSI Model 1990 counter processors interfaced to a custom built coincidence timing unit were used in the data collection and processing system. High and low pass filters were set to 10 MHz and 50 MHz, respectively, for the stationary LDV system, and 20 MHz and 100 MHz, respectively, for the fiber optic LDV system. Both processors were set to make a single measurement per burst, count 16 fringes and use a 1 % comparator. A hardware coincident window was set at 20μs for all of the tests. Data(two velocities and the running time for each realization) were transferred through two DMA ports to a MicroVax minicomputer and later uploaded to a VAX 8650 for analysis.

The flow field was seeded using titanium dioxide (TiO_2) particles generated by reacting dry titanium tetrachloride ($TiCl_4$) with the moist shop air. Craig et al. (1984) measured the particle sizes generated by this device and found that they were fairly uniform and in the $0.2 - 1\mu$m diameter range. Data validation rates varied between 5000 and 500 per second on each counter processor and depended mainly on how well the chemical reaction proceeded. This seemed to be very sensitive to shop air temperature and relative humidity. Coincident data validation rates ranged from 1000 to 50 measurements per second. Velocity bias was not considered in this study.

EXPERIMENTAL PROCEDURE

All flow conditions were maintained at near constant values throughout the testing procedure. The inlet centerline velocity, U_{cl}, was maintained at 18.0 m/s $\pm$ 0.1 m/s(59 ft/s $\pm$ 0.3 ft/s). Spatial correlation statistics and histograms were formed by using 5000 individual realizations for each velocity channel at each measurement point. Autocorrelations were formed by using 50000 individual realizations from the stationary LDV system. In computing statistical parameters a two step process was used to eliminate noise from the data. In the first step, a 5% threshold level was

applied to the raw velocity data in an effort to estimate the standard deviation of the valid data. This estimate was made by putting the raw velocity data in 100 equally spaced bins bounded by the actual maximum and minimum velocity found in the raw data sample. Next, the bin with the maximum number of samples in it is found. Finally, all bins having at least 5% of the number of samples found in the "maximum" bin are located. The width of the data which meets this threshold criteria is then used to estimate the standard deviation of the "good" data. Upper and lower cutoff limits are set by adding and subtracting, respectively, 2.5 times the half-width of the data which meets this 5% threshold. Applying this technique to a Gaussian distribution is equivalent to setting cutoff limits which correspond to ± 4.1 standard deviations and thus this first step is used only to remove spurious data. This method is a variation of the method suggested by Meyers (1988) and is used to eliminate spurious data which if not removed would give an abnormally large value for the standard deviation and thus wider cutoff limits. In the second step, the mean and standard deviation of the remaining data(spurious data removed) are calculated. This data is then further filtered to remove data which deviates more than 3 standard deviations from this new mean. Finally, revised statistics are calculated once these additional outliers are discarded. For a properly operating LDV system very few points are removed during the first step(typically less than 10 out of 5000) and less than 1% of the data should be discarded after both steps.

EXPERIMENTAL RESULTS

Two-point velocity correlation measurements were made at three locations in the axisymmetric sudden expansion flow field as shown in Figure 3. The first two spatial correlation measurements were made at an axial location of ten step heights($x/H = 10$) downstream of the sudden expansion plane, one on the centerline of the flow($r/H = 0$) and the other at the same radial location as the step($r/H = 2$). The third spatial correlation measurement was made at an axial location of six step heights and at the same radial location as the step($x/H = 6,\ r/H = 2$). These are the locations where the stationary LDV probe volume remains. Spatial correlations are obtained by po-

sitioning the movable LDV probe volume(fiber optic system) at various separation distances from the stationary probe volume. Longitudinal spatial correlations defined by,

$$R_{11}(r,0,0) = f(r) = R_{11}(\Delta x) = \frac{\overline{u'(x)u'(x+\Delta x)}}{\sqrt{\overline{u'^2(x)}}\sqrt{\overline{u'^2(x+\Delta x)}}}$$

were made at the three locations in both the plus and minus Δx directions. Lateral spatial correlations defined by,

$$R_{11}(0,r,0) = g(r) = R_{11}(\Delta y) = \frac{\overline{u'(y)u'(y+\Delta y)}}{\sqrt{\overline{u'^2(y)}}\sqrt{\overline{u'^2(y+\Delta y)}}}$$

were made at two of the locations($x/H = 6.$ $r/H = 2$ and $x/H = 10.$ $r/H = 2$) in both the plus and minus Δy(same as Δr) directions. Both directions (plus and minus) were considered to determine the homogeneity of the flow. It should be noted that turbulent flows are three-dimensional and that these measurements give only the one-dimensional correlation coefficient. In addition to these spatial correlation measurements, autocorrelation measurements were also made at these three locations.

Before correlation measurements were made axial velocity measurements at three axial planes($x/H = 1,$ 6, 10) were made using the fiber optic LDV system. This was done in an effort to validate that the flow was symmetric and to ensure that the flow was what was expected. Turbulence statistics were calculated using 5000 samples for each measurement point and the filtering procedure described above. Figure 4 shows the measurements of the mean axial velocity at these three planes while Figure 5 shows the axial standard deviation profiles. The inlet mean velocity profile was found to be similar to that of a fully developed turbulent pipe profile. These figures indicate that the spatial correlation measurements made at $r/H = 2$ were in regions of large velocity gradient and high turbulence. The turbulence intensity ($TI = \sqrt{\overline{u'^2}}/\bar{U}$) at both these points was found to be approximately 50%. The turbulence intensity at $x/H = 10.$ $r/H = 0$ was found to be approximately 8%.

Figures 6 through 10 show the measured spatial correlation coefficients as a function of separation distance at the three measurement points mentioned above. Figures 6, 7 and 9 show the longitudinal spatial correlations while Figures 8 and 10

show the lateral spatial correlations. Longitudinal spatial correlation measurements were made with a minimum separation distance equal to 254 μm(0.010 in) and a maximum separation distance equal to 101.6 mm(4.0 in). The same minimum separation distance was used for lateral spatial correlation measurements, but different maximum separation distances, depending upon whether the direction was toward the wall (where Δy_{max} = 22.86 mm(0.9 in)) or toward the centerline (where Δy_{max} = -50.8 mm(2.0 in)), were used. The spatial correlation data were fit with a "best fit" curve having the form $R(r) = C\exp(r/\Lambda)$ and are shown on each figure by a solid line. The integral length scales defined as,

$$\Lambda_f = \int_0^\infty f(x)dx \qquad or \qquad \Lambda_g = \int_0^\infty g(y)dy$$

give a measure of the longest connection, or correlation distance, between velocities at two points in the flow field. This is because for a given separation distance, r, only eddies larger than r will contribute to the correlation function while eddies smaller than r will not. If the data can be fitted with the simple exponential function given above the integral of this function (which is the integral length scale) is simply equal to the coefficient, Λ, in the exponential function. Values for the estimated integral length scales are listed on Figures 6 through 10 and are summarized in Table 1. The integral length scales were found to vary between 25 and 30 mm in both separation directions for the longitudinal spatial scale and were found to vary between approximately 8 and 14 mm for the lateral spatial scale. These results indicate that the flow appears reasonably homogeneous in the axial flow direction but is non-homogeneous in the radial direction where the wall influences the flow field. Also, the lateral integral length scales, Λ_g, are approximately one-half the value of the longitudinal integral length scales, Λ_f, indicating strong anisotropy. By definition all correlation coefficients should equal one when the separation distance is zero. The data presented here show that this is not the case. It is not clear at this time why the correlation coefficients do not equal one at zero separation distance, but noise in the data or probe volume length effects(*i.e.* LDV probe length larger than Kolmogoroff scale) are suspected. Further data analysis may reveal this discrepancy. In addition,

further data analysis is required to determine why there is a large degree of scatter in these correlation measurements. There are many explanations for why low values of correlation coefficient might result(*i.e.* noise in the data giving large standard deviations,...) which need to be investigated.

Dissipation(micro) length scales have not been determined yet, but could be estimated by performing a Taylor series expansion on the correlation coefficient curve near zero separation distance(see Hinze, 1975). The dissipation length scales, λ_f and λ_g, which result from fitting a parabola to the appropriate correlation functions near Δx or $\Delta y = 0$ are obtained from.

$$R(\Delta x) \approx 1 - \frac{\Delta x^2}{\lambda_f} \qquad or \qquad R(\Delta y) \approx 1 - \frac{\Delta y^2}{\lambda_g}$$

These length scales give an estimate of the average dimension of the smallest eddies in the flow which are responsible for viscous dissipation. An estimate of the turbulent viscous dissipation rate can be made once these scales and the turbulence intensity are known(see Hinze, 1975).

Discrete autocorrelation measurements were made using the slotting technique described by Jones (1972) and Mayo, et al. (1974). The time lag axis was divided into 1000 bins of equal width, $\Delta\tau = 200\mu$s, and the exact lag produces of all points up to the maximum lag time, $\Delta\tau_{max} = 0.2$ s, were accumulated in appropriate bins. The average of all the auto-products falling in each bin was assumed to be the value of the discrete autocorrelation function, *i.e.*

$$R_E(\Delta\tau) = \frac{\overline{u'(t)u'(t+\Delta\tau)}}{\overline{u'^2(t)}}$$

at the midpoint of the bin. The data was filtered first using the previously described method to eliminate noise before the autocorrelation was estimated. In addition, the zero-lag autoproducts were not included in the first bin in order to minimize the ambiguity spectrum due to uncorrelated data(Gaster and Roberts, 1975, Srikantaiah and Coleman, 1985). A Hamming window was applied to each autocorrelation function. Figures 11 through 13 show the windowed autocorrelation functions calculated from

the data measured at the three locations in the flow field. The Eulerian integral scale can be estimated by finding the area under the autocorrelation curve as given by,

$$\tau_E = \int_0^\infty \mathbf{R}_E(t)dt$$

The Eulerian dissipation(micro) time scale can be estimated by applying a Taylor series expansion to the autocorrelation function near $t = 0$. The equation for the osculating parabola in the vertex of the $R_E(t)$ curve is,

$$R_E(t) \approx 1 - \frac{t^2}{\tau_E}$$

Here the Eulerian micro time scale, τ_E, is a measure of the most rapid changes that occur in the fluctuations of $u(t)$. Taylor's hypothesis, which is valid only if the flow field has uniform mean velocity, $\bar{U}$, and small turbulence intensity, gives a relationship between temporal and spatial quantities(i.e. $x = \bar{U}t$). If Taylor's hypothesis applies we get the relations: $\Lambda_f = \bar{U}\tau_E$ and $\lambda_f = \bar{U}\tau_E$. Note that only the longitudinal length scales can be estimated with Taylor's hypothesis. Table 1 gives a summary of the results obtained from this study. Comparisons of the integral length scales obtained from autocorrelation measurements with those obtained directly from spatial correlation measurements show that the autocorrelation method gives a length scale 50 % too large at the low turbulence location and gives a length scale a factor of two too small at the locations in the shear layer. Considering the limitations of Taylor's hypothesis the spatial correlation length scale estimates are believed to be more reliable. Numbers appearing in brackets in this table refer to length scales obtained with negative separation distances. The turbulent kinetic energy(TKE) was estimated by assuming that TKE $\approx \sqrt{\overline{u'^2}}$ which is a reasonable approximation for this flow field (see Gould, et al. 1990). The Kolmogoroff length scale, $\eta = (\nu^3/\varepsilon)^{1/4}$, was estimated by assuming that the turbulent viscous dissipation rate, ε, equaled three-fourths the production of TKE. The production of TKE was assumed to equal $\overline{u'v'}\partial\bar{U}/\partial r$, where Bradshaw et al.'s model (*i.e.* $\overline{u'v'} \approx 0.35$TKE) was used to estimate the shear stress and data from Figure 4 was used to estimate the mean velocity gradient.

Table 1a. Autocorrelation Results.

Location	x/H=10, r/H=0	x/H=10, r/H=2	x/H=6, r/H=2
$\bar{U}$, (m/s)	15.69	6.81	7.99
$\sqrt{\overline{u'^2}}$, (m/s)	1.35	3.64	3.93
TI	0.08	0.54	0.49
TKE, (m^2/s^2)	1.82	13.26	15.42
τ_{E_f} (s)	.0024	.0025	.0017
τ_E, (s)	.00064	.00036	.00032
Λ_f, (mm)	37.7	17.0	13.6
λ_f, (mm)	10.0	2.5	2.6
η, (μm)	200	45	40

Table 1b. Spatial Correlation Results.

Location	x/H=10, r/H=0	x/H=10, r/H=2	x/H=6, r/H=2
Λ_f, (mm)	27.7(28.8)	35.1(31.0)	28.7(23.4)
λ_f, (mm)	-	-	-
Λ_g, (mm)	-	11.9(14.8)	7.6(13.6)
λ_g, (mm)	-	-	-

Finally, power spectrum estimates were made by performing the discrete cosine transform(see Jones, 1972, Mayo, et al., 1974, Bell, 1983, Srikantaiah and Coleman, 1985) on the windowed discrete autocorrelation functions given above. Figures 14 through 16 show these results and indicate that there is significant energy at frequencies below 100 Hz. This result is consistent with the measured Eulerian integral time scales (*i.e.* $\tau_E \approx 0.002$) as expected.

CONCLUSIONS

Successful two-point velocity correlation measurements were made in the anisotropic flow field of an axisymmetric sudden expansion. Both longitudinal and lateral spatial correlations were measured. The integral length scales were estimated and compared with those obtained from autocorrelation measurements in conjunction with Taylor's

hypothesis. The agreement between these two methods was poor and it is believed that the spatial correlation measurements give more reliable results. The spatial correlations did not equal one at zero spatial separation distance and there was much scatter in the spatial correlation data. Spurious noise is believed responsible for this behavior. Further data analysis will be required to determine if this is the case. Once this is performed estimates of the dissipation length scales and the turbulent dissipation rate can be made. In addition, an estimate of the one-dimensional energy spectrum as a function of wavenumber can be calculated by performing the Fourier transform of the spatial correlation functions. Future work should include a study of the effects of probe volume length on the measured correlation coefficient. Photo detectors could be placed 90° off-axis in order to decrease the effective probe volume size such that the length is approximately equal to the diameter.

ACKNOWLEDGMENTS

The author would like to thank Dr. A. S. Nejad for use of his laboratory and equipment and for all the help he provided. Thanks are also due to Dr. S. A. Ahmed for help in operating the LDV data acquisition system, to K. G. Schwartzkopf for computer programming and C. Smith for all his technical support in the laboratory. This investigation was performed at Aeropropulsion and Power Directorate, Wright Laboratory (WL/POPT) under the Summer Faculty Research Program supported by AFOSR.

REFERENCES

1. Absil, L. H. J., (1988),"Laser Doppler Measurements of Mea- Turbulence Quantities, Time and Spatial-Correlation Coefficients in the Wake of a Circular Cylinder," Proceedings of the 4th International Symposium on Applications of Laser Anemometry to Fluid Mechanics, Lisbon, Portugal, p. 1.1.

2. Batchelor, G. K., (1953), *The Theory of Homogeneous Turbulence*. Cambridge Univ. Press, New York.

3. Bell, W. A., (1983), "Spectral Analysis Algorithms for the Laser Velocimeter: A Comparative Study," *AIAA Journal*, **21**, pp. 714-719.

4. Bradshaw, P., Ferriss, D. H. and Atwell, N. P., (1967) ,"Calculation of Boundary Layer Development using the Turbulent Kinetic Energy Equation," *J. Fluid Mech.*, **28**, pp. 593-616.

5. Bradshaw, P.,(1971), *An Introduction to Turbulence and its Measurement*, Pergamon Press, Oxford.

6. Craig, R. R., Nejad, A. S., Hahn, E. Y. and Schwartzkopf, K. G. 1984 "A General Approach for Obtaining Unbiased LDV Data in Highly Turbulent Non-Reacting and Reacting Flows." AIAA Paper. No. 84-0366.

7. Fraser, R., Pack, C. J. and Santavicca, D. A., (1986), "An LDV System for Turbulence Length Scale Measurements." *Experiments in Fluids*, **4**, pp. 150-152.

8. Gaster, M. and Roberts, J. B., (1975), "Spectral analysis of Randomly Sampled Signals." *J. Inst. Math Appl.*, **15**, pp. 195-216.

9. Gould, R. D., Stevenson, W. H. and Thompson, H. D., (1990), "Investigation of Turbulent Transport in an Axisymmetric Sudden Expansion," *AIAA Journal*, **28**, pp. 276-283.

10. Hinze, J. O., (1975), *Turbulence*, McGraw-Hill.

11. Jones, R. H., (1972), "Aliasing with Unequally Spaced Observations," it J. Appl. Meteorology, **11**, pp. 245-254.

12. Mayo, W. T., Shay, M. T. and Ritter, S., (1974), "The Development of New Digital Data Processing Techniques for Turbulence Measurements with Laser Velocimetry," Final Report AEDC-TR-74-53.

13. Meyers, J. F., (1988) "Laser Velocimeter Data Acquisition and Real Time Processing using a Microcomputer." Proceedings of the 4th International Symposium on Applications of Laser Anemometry to Fluid Mechanics, Lisbon, Portugal, p. 7.20.

14. Pfeifer, H. J., (1986), "Correlation and Spectral Density Measurements by LDA," Proceedings of the 5th International Congress on Applications of Lasers and Electro-Optics, Arlington, Va., **58**, pp. 59-77.

15. Srikantaiah, D. V. and Coleman, H. W., (1985), "Turbulence Spectra from Individual Realization Laser Velocimetry Data," *Experiments in Fluids*, **3**, pp. 35-44.

16. Tennekes, H. and Lumley, J. L. (1972), *A First Course in Turbulence*, Clarendon Press, Oxford.

17. Townsend, A. A., (1976), *The Structure of Turbulent Shear Flow*, Cambridge Univ. Press, London.

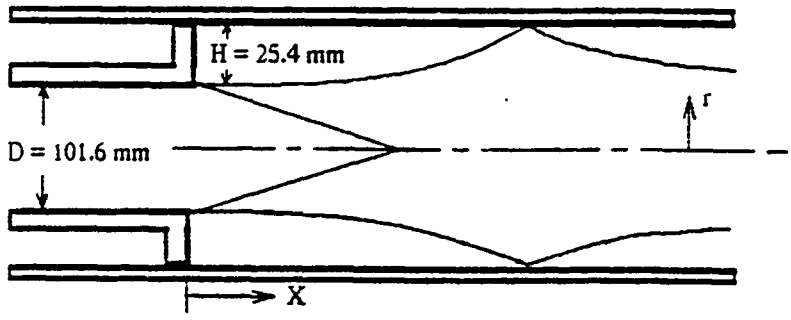
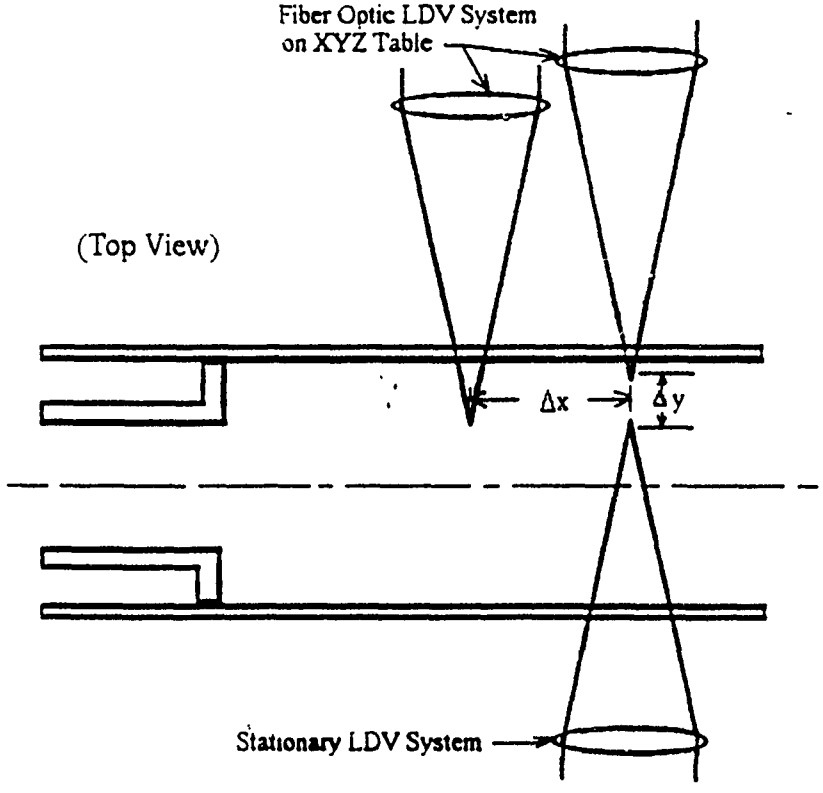


Figure 1. Axisymmetric sudden expansion geometry.



2. LDV orientation.

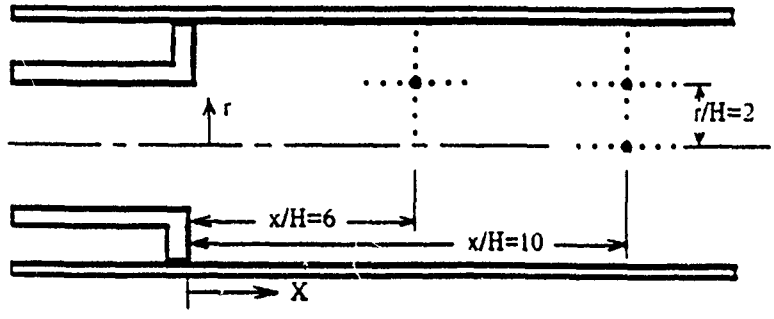


Figure 3. Measurement locations(top view).

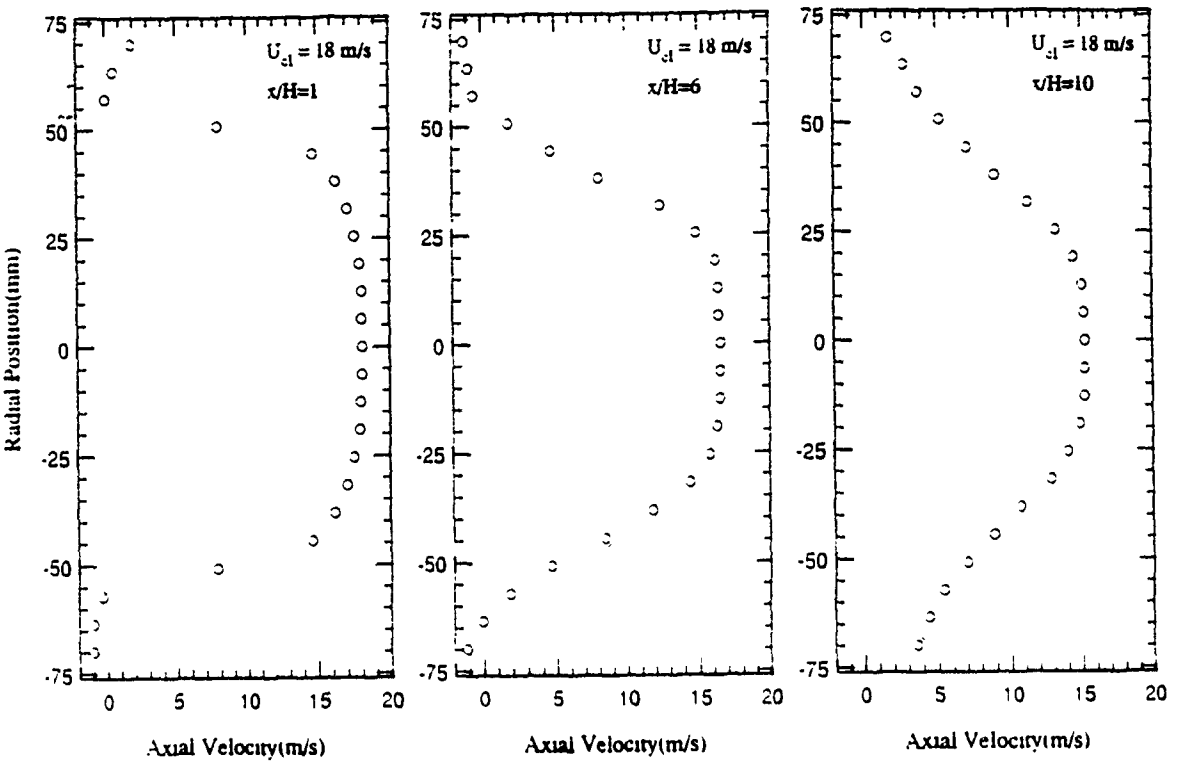


Figure 4. Mean axial velocity profiles.

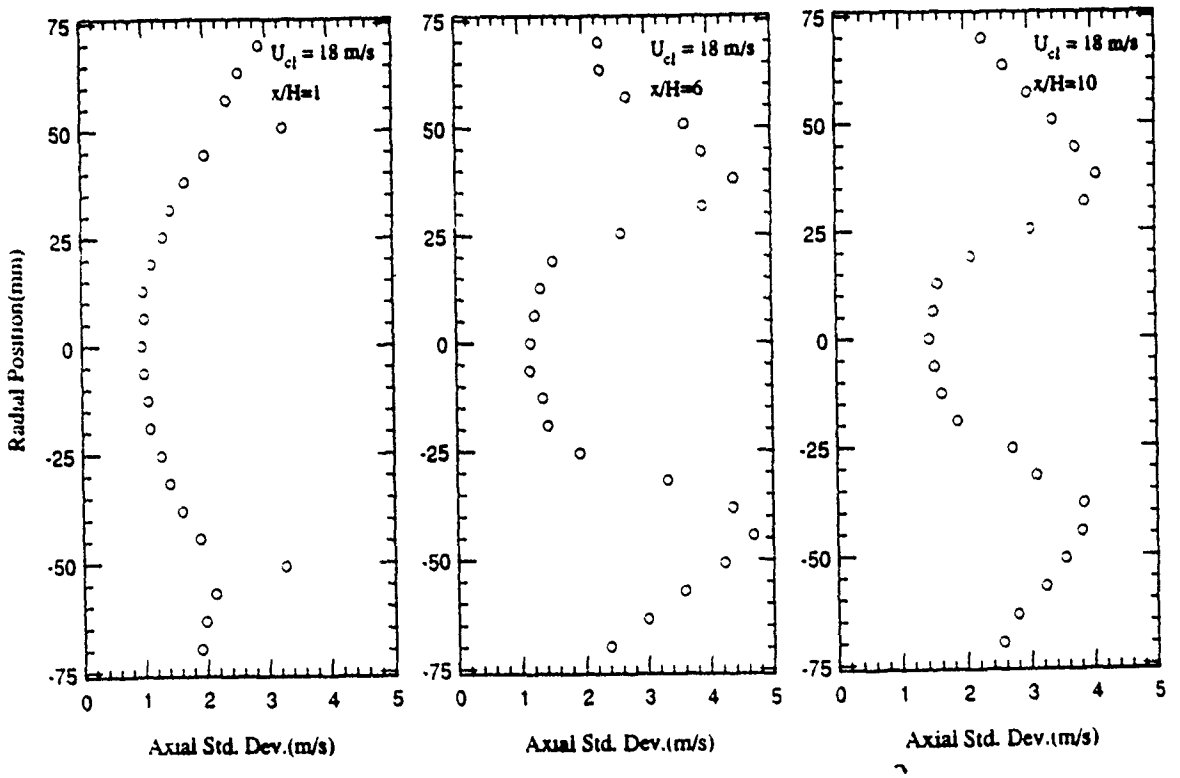


Figure 5. Axial standard deviation profiles.

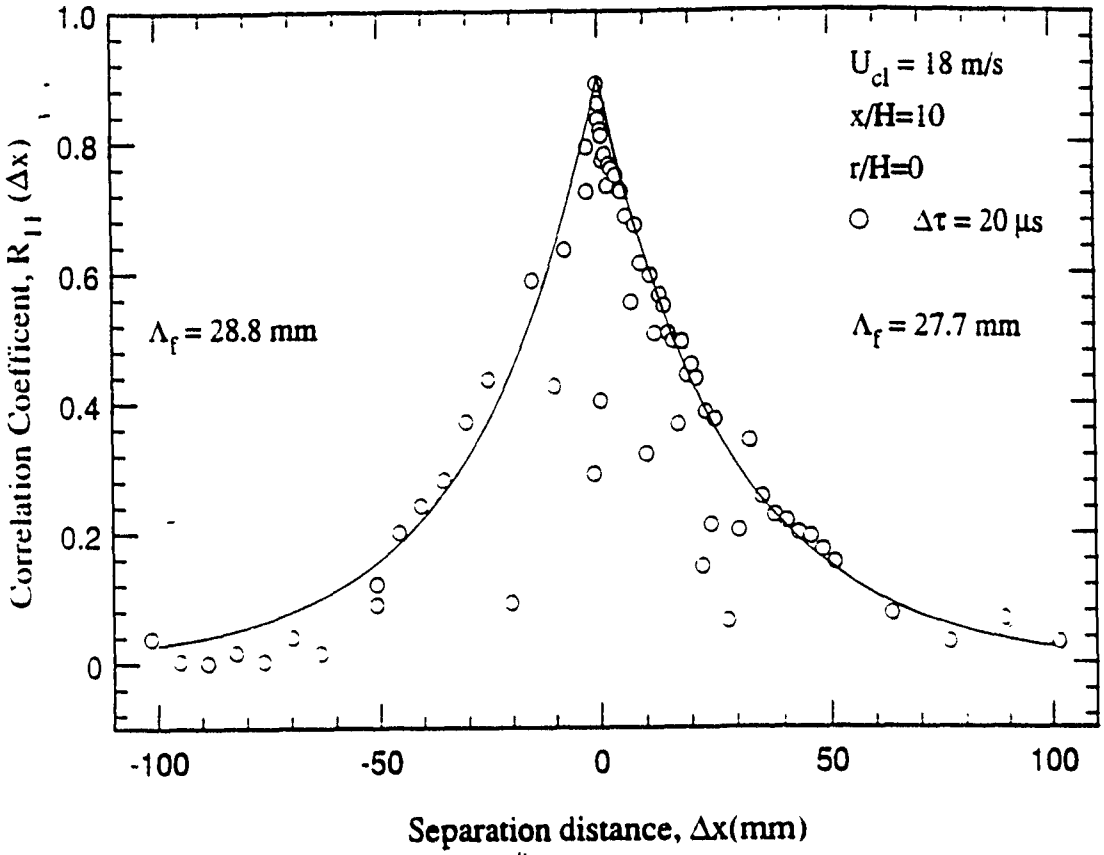


Figure 6. Longitudinal spatial correlation measurements (x/H=10, r/H=0).

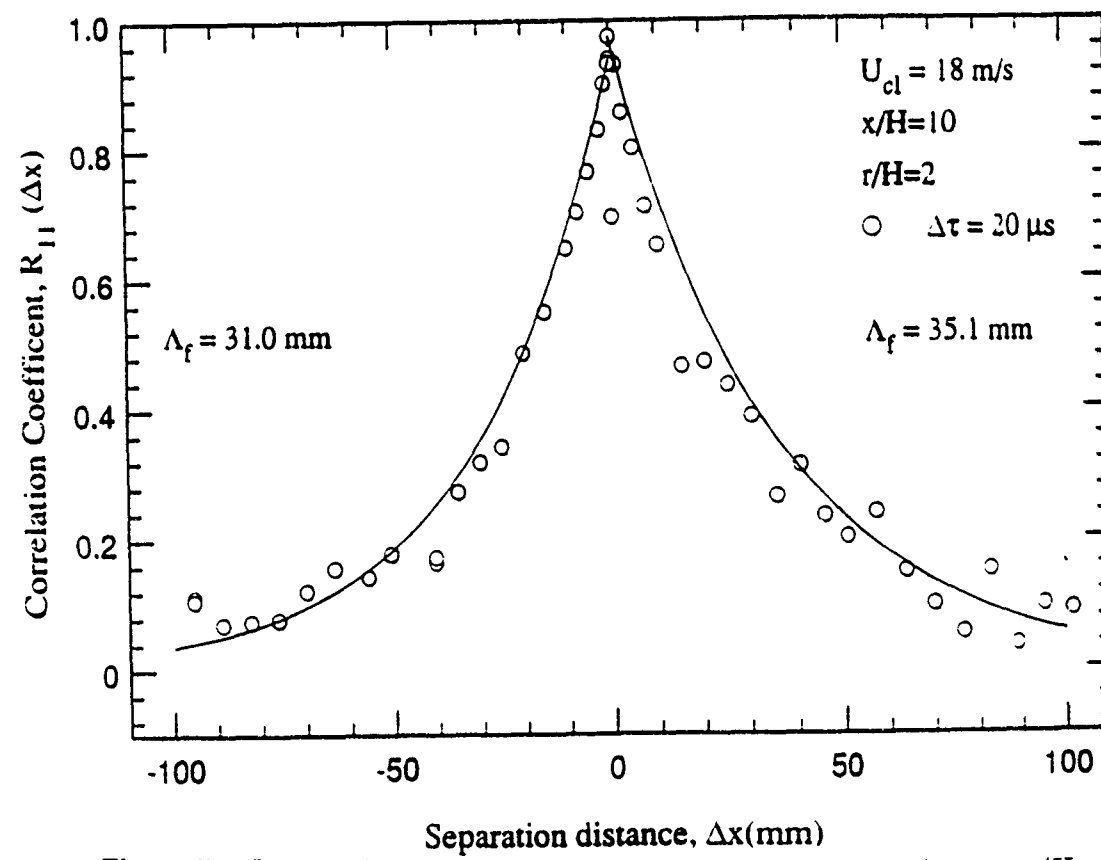


Figure 7. Longitudinal spatial correlation measurements (x/H=10, r/H=2).

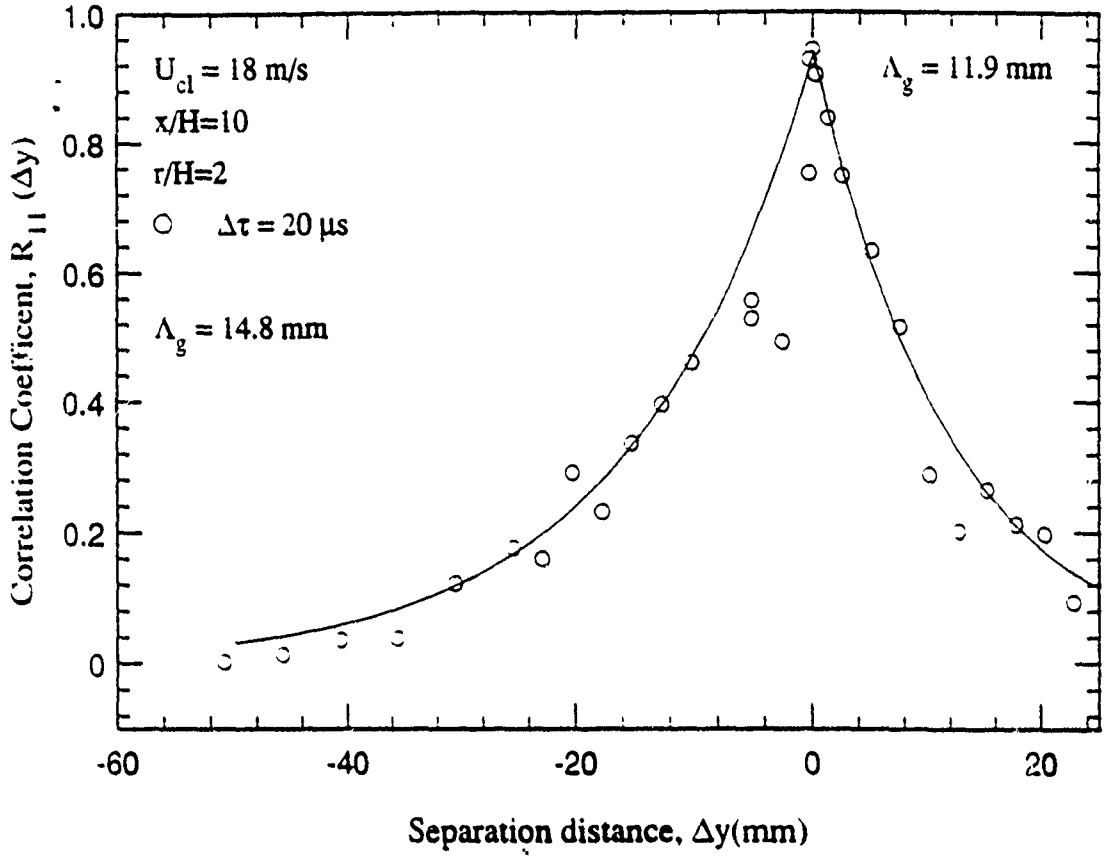


Figure 8. Lateral spatial correlation measurements (x/H=10, r/H=2).

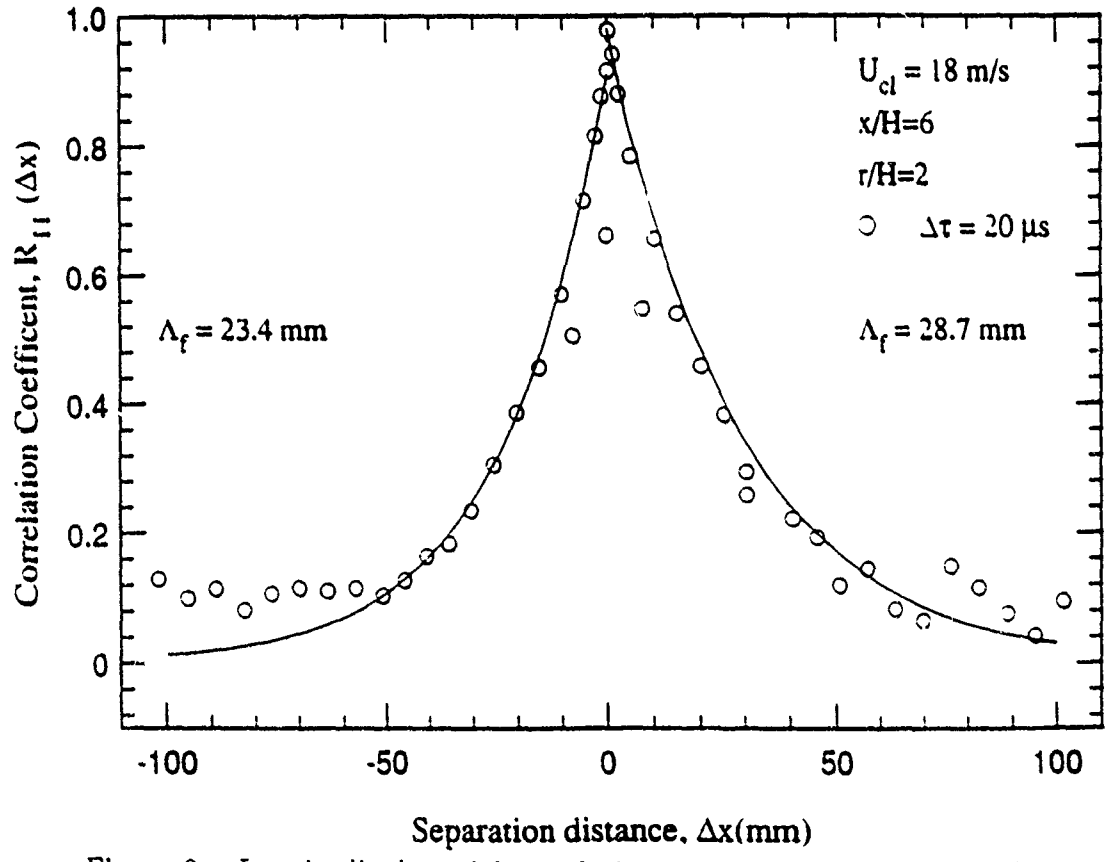


Figure 9. Longitudinal spatial correlation measurements (x/H=6, r/H=2).

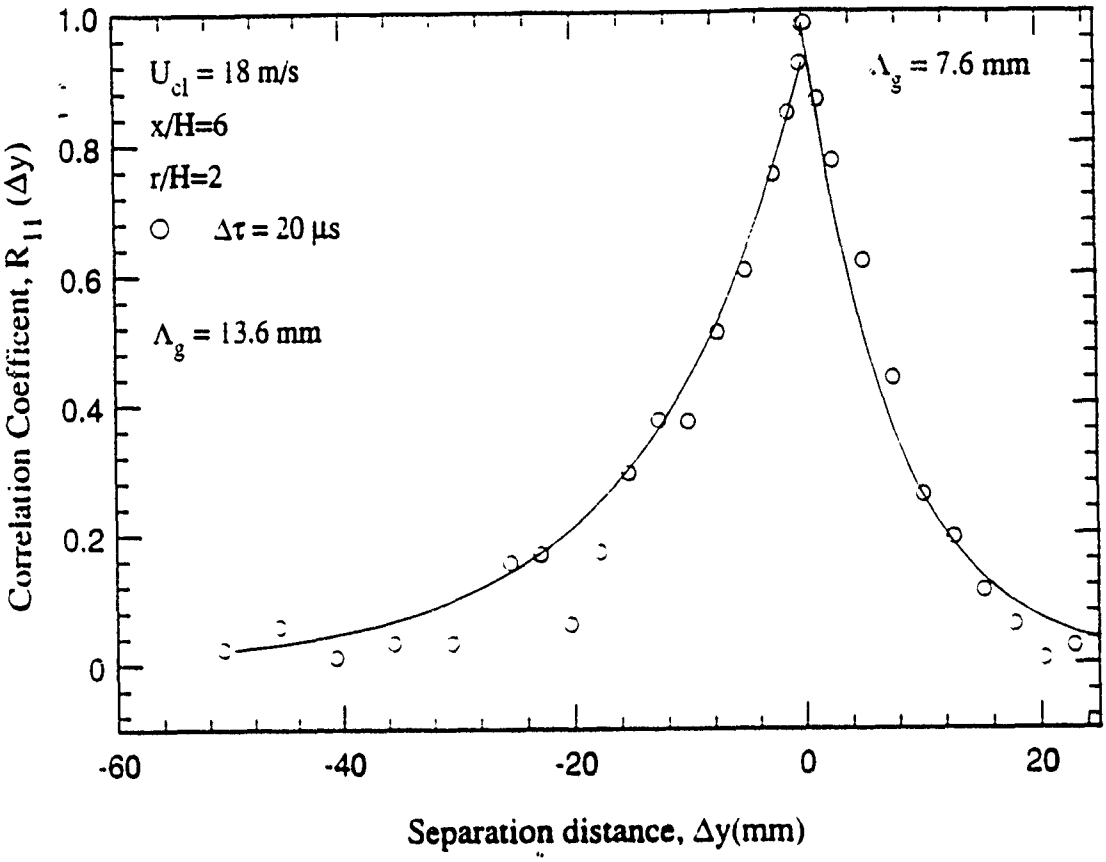


Figure 10. Lateral spatial correlation measurements (x/H=6, r/H=2).

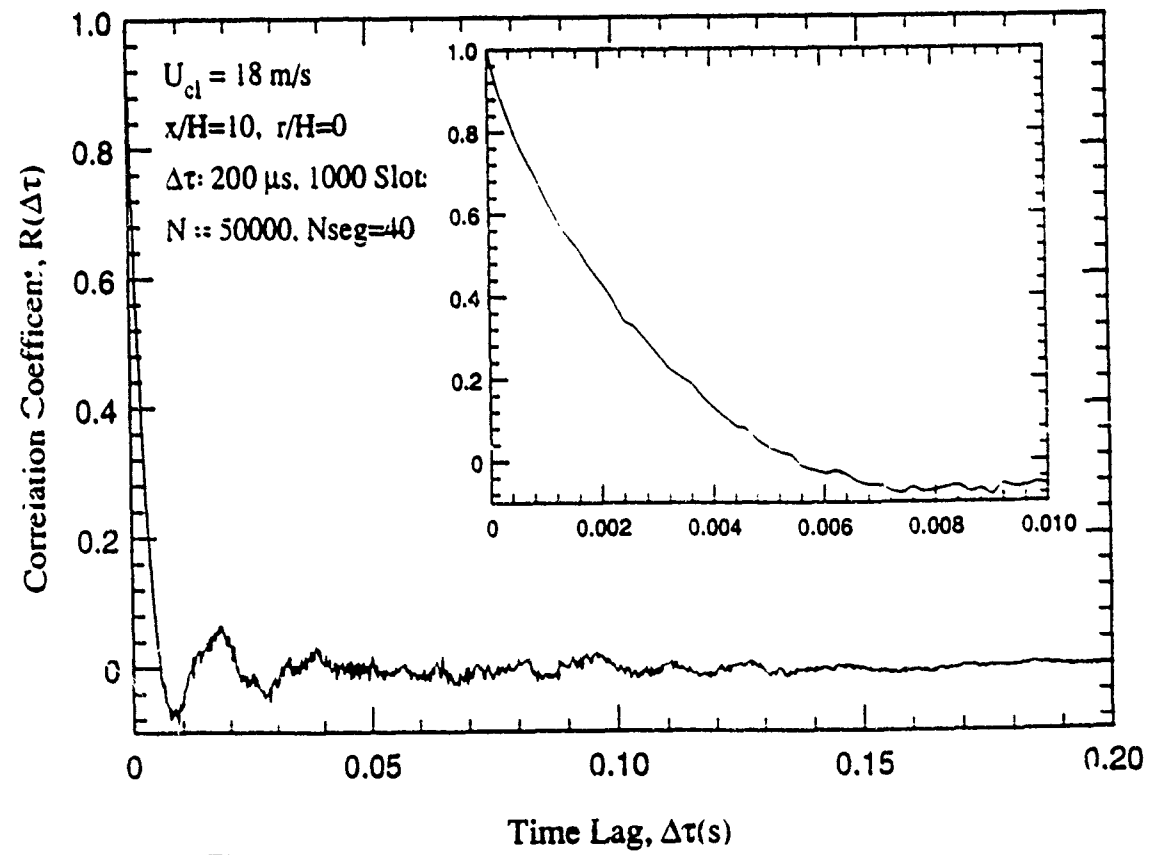


Figure 11. Autocorrelation measurements (x/H=10, r/H=0).

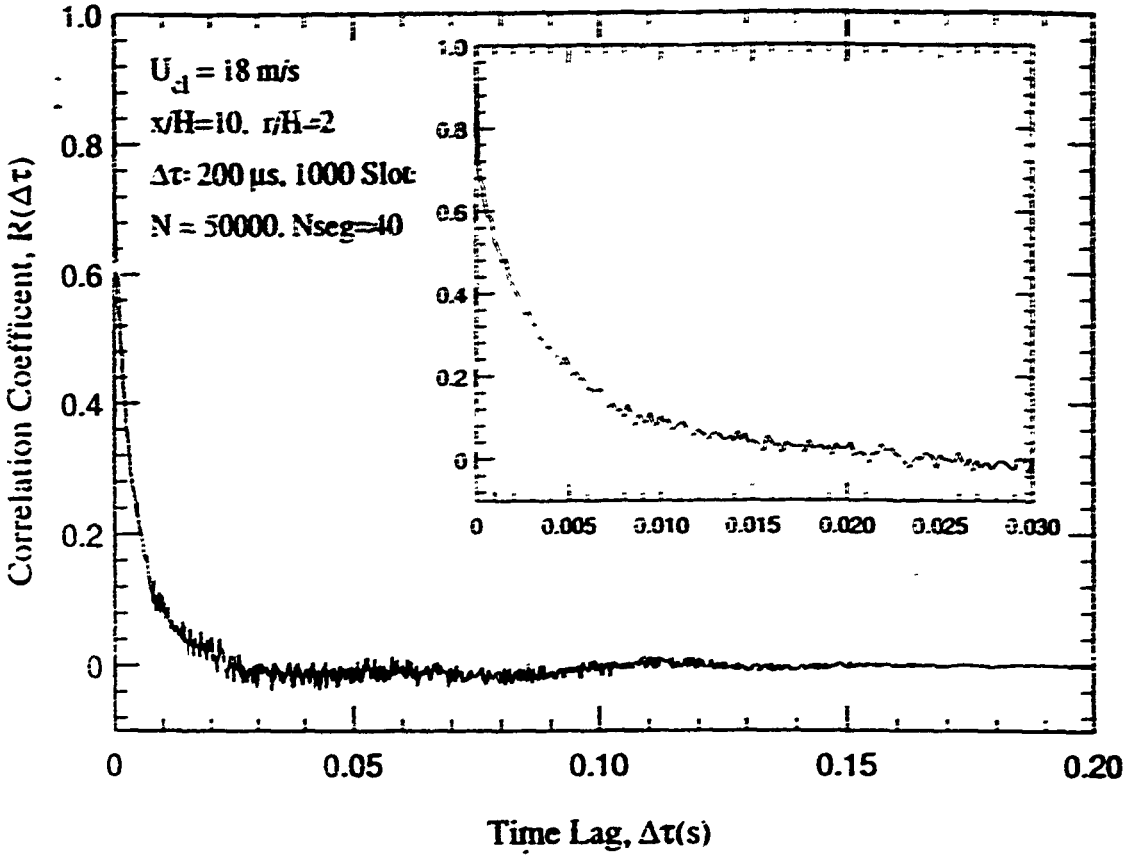


Figure 12. Autocorrelation measurements (x/H=10. r/H=2).

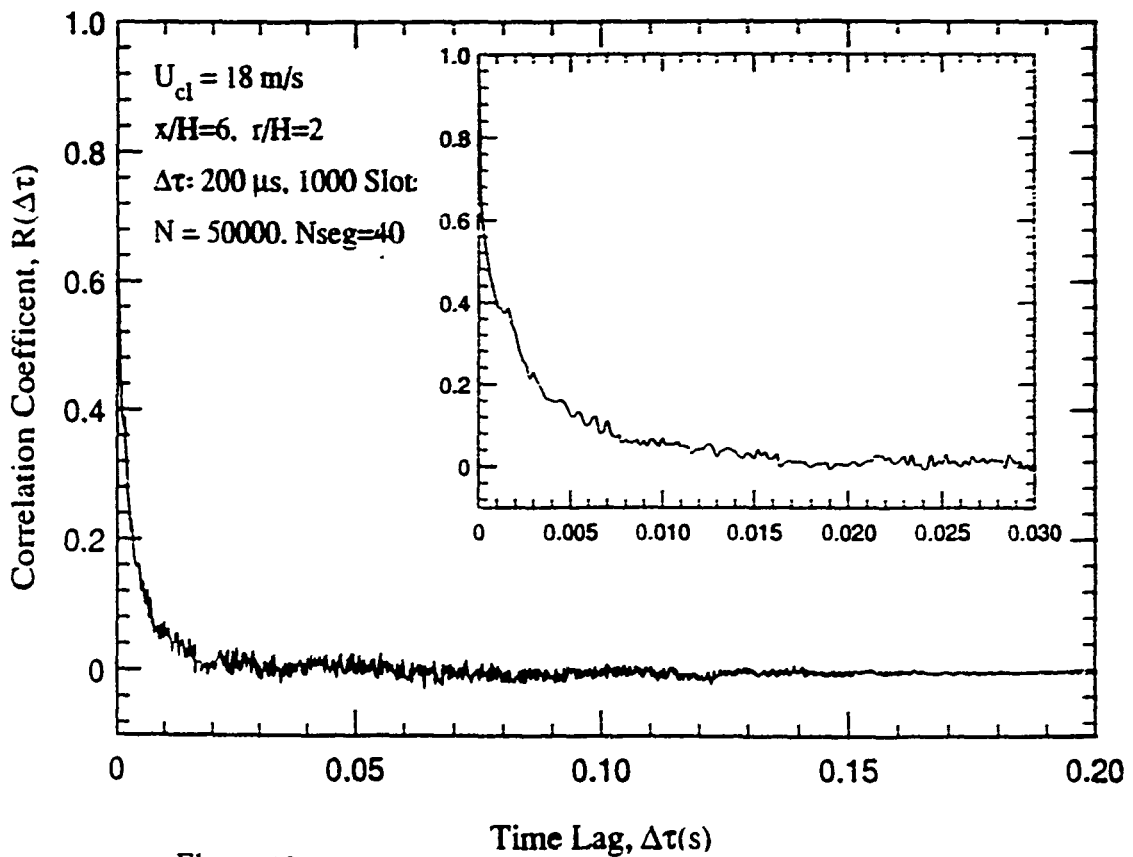


Figure 13. Autocorrelation measurements (x/H=6. r/H=2).

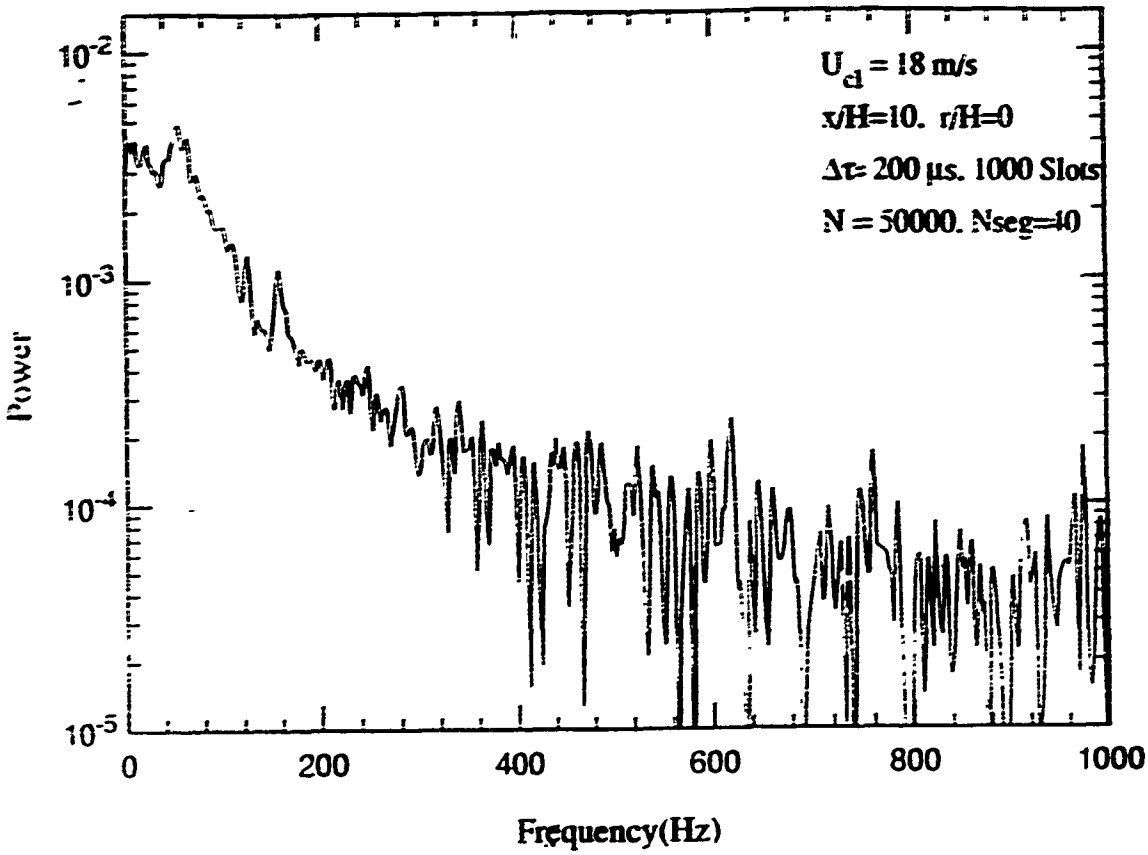


Figure 14. Power spectrum measurements (x/H=10, r/H=0).

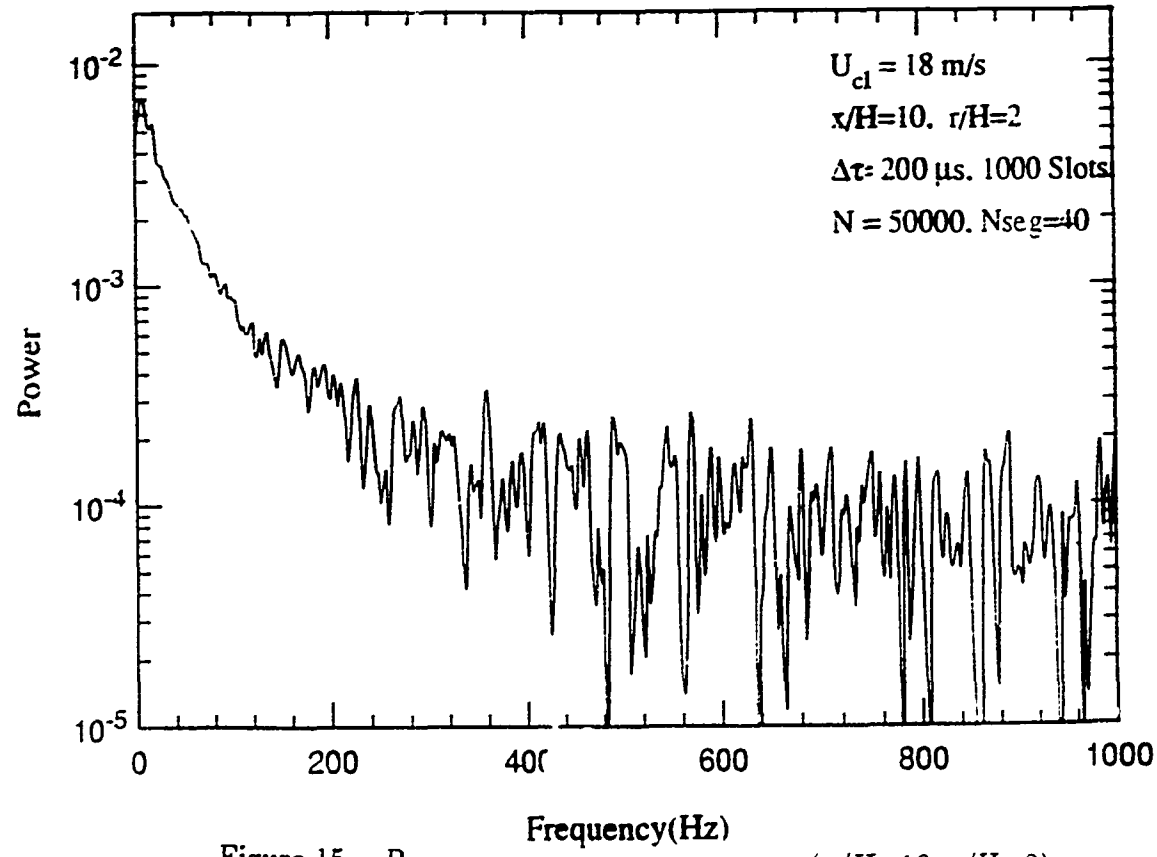


Figure 15. Power spectrum measurements (x/H=10, r/H=2).

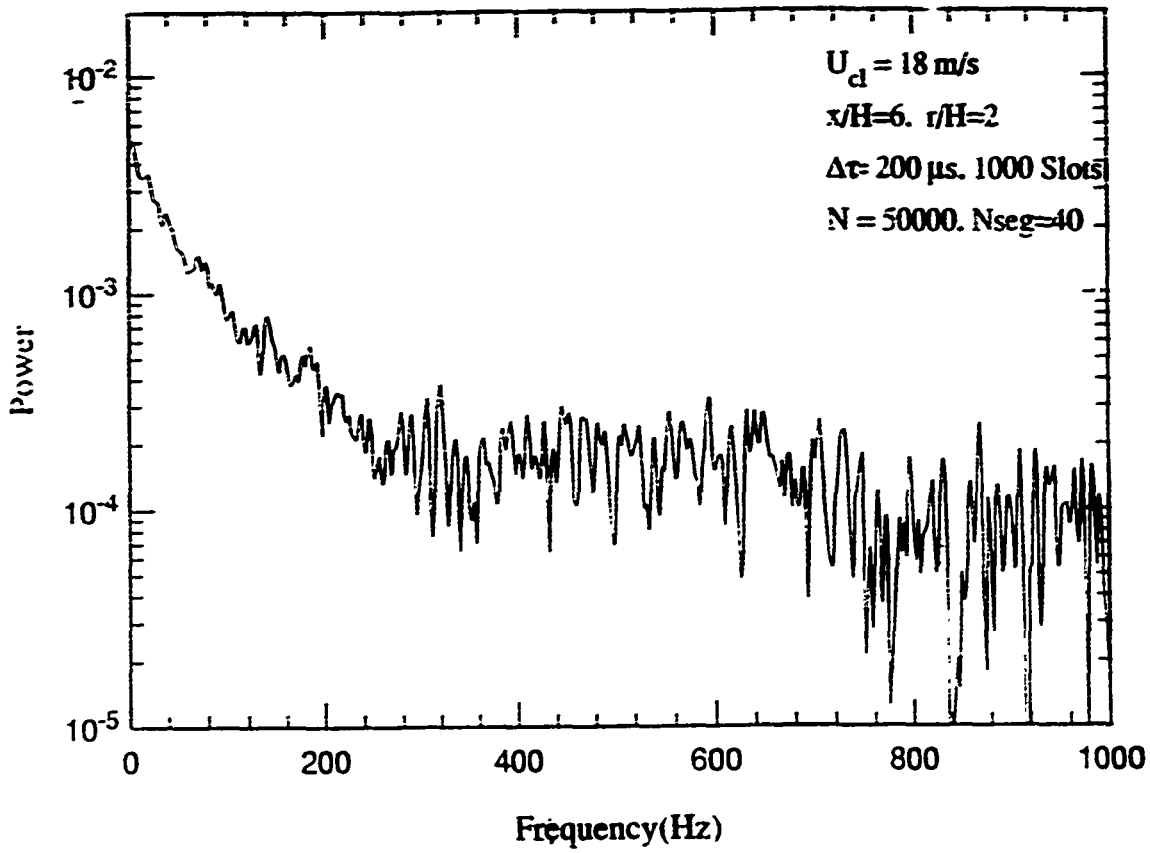


Figure 16. Power spectrum measurements (x/H=6, r/H=2).

ELECTRIC FIELD EFFECTS ON PROPANE/AIR FLAMES

Harold H. Harris
Department of Chemistry
University of Missouri-St. Louis
St. Louis, MO 63121

Report of investigations performed
in collaboratoration with Dr. Bishwa Ganguly
at the Aero Propulsion Laboratory of
Wright-Patterson Air Force Base,
Summer, 1991

Abstract

The effect of fields of several hundred to a few thousand volts per centimeter on propane/air flames burning at atmospheric pressure have been determined by measurements of the changes in the spatial distribution of the flames as revealed by emission spectra from such flames, and the approximate potentials in the flames caused by the interaction of the flame plasma and an applied field. It is shown that the effects are very pronounced in slightly rich flames (air slightly lower than the stoichiometric ratio) and almost nonexistent in slightly lean flames. Flames susceptible to deflection by the field coincide almost always to those which exhibit emission from C_2 ($A\ ^3\Pi_g \rightarrow X\ ^3\Pi_u$), but there are exceptions to this rule when flames are seeded with alkali metals. The observations are suggestive of a mechanism involving particular ionic or molecular species, rather than the more general "ionic wind" hypothesis.

Introduction

The effect of electric potentials on flames has been the subject of numerous experimental and theoretical studies over many years, including a classic book on the subject published in 1969[1]. Nevertheless, there does not seem to exist a satisfactory molecular-level understanding of these phenomena[2]. This is not surprising, given the complexity of a flame, which involves interactions between stable reactant and product species, neutral and charged, ground-state and excited reactive intermediates and electrons, all undergoing collisions in a turbulent medium with substantial thermal gradients and fields generated by the flame itself. Such an understanding might, however, be of great practical significance because of these interactions have been reported to change the temperature of flames[3], to modify their heat transfer[4], to stabilize flames at low pressures and low fuel/air ratios[5] and in the absence of gravity[6], to extinguish flames under certain conditions[7], and to alter the yield of soot[8]. It is especially important to note that the electrical energy input required for many of these effects is only a very small fraction of the flame power. Thus, a small electric field perturbation, designed on the

basis of understanding of the detailed mechanism, could lead to extremely important practical consequences.

Ions in Flames

It goes without saying that electric fields will interact substantially only with the charged species in a flame. This was the reason that electric field effects were proof that ionic species existed in flames. The main source of ionization in hydrocarbon flames is believed to be the chemi-ionization reactions[9]:

$$CH + O \rightarrow CHO^+ + e$$

$$CH^* + C_2H_2 \rightarrow C_3H_3^+ + e,$$

while the dominant positive ion, H_3O^+, is presumably the product of a fast ion-molecule reaction (such as):

$$CHO^+ + H_2O \rightarrow CO + H_3O^+$$

Ion destruction is due primarily to the fast dissociative recombination of electrons with molecular ions; for example:

$$H_3O^+ + e \rightarrow H_2O + H$$

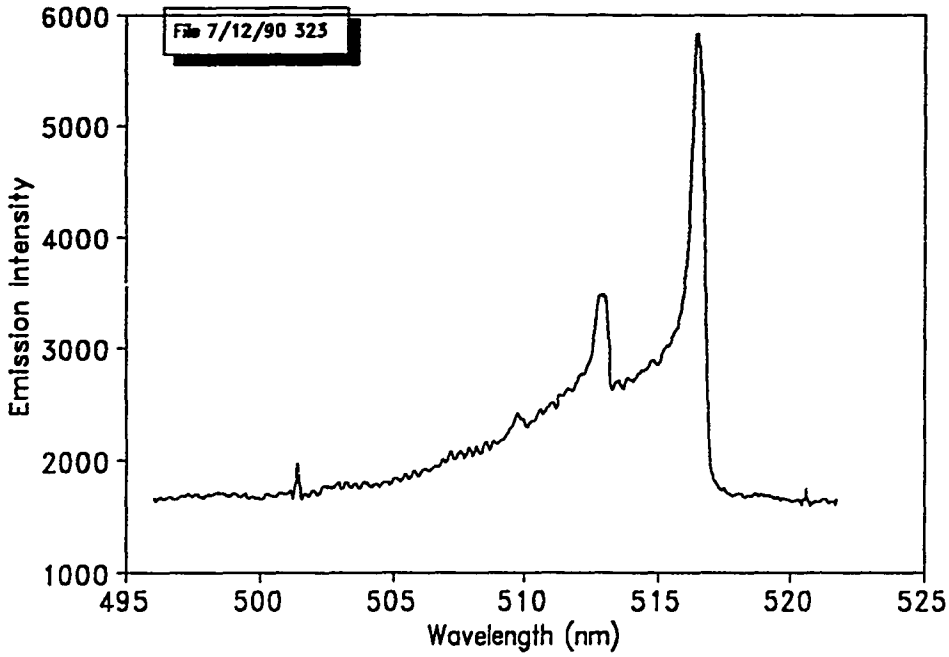
The complexity of both positive and negative ion flame chemistry is suggested by the observation by Green and Sugden[10] of 83 positive ion peaks in the mass spectrum of an $H_2/O_2/N_2$ + 1% C_2H_2 flame, and Feugier and Van Tiggelen's detection of 12 negative ions in a stoichiometric neopentane-oxygen/N_2 flame.

Emission from Hydrocarbon Flames

The emission observable visually from most hydrocarbon flames is due principally to the excited radicals C_2 ("Swan bands", $A\ ^3\Pi_g \rightarrow X\ ^3\Pi_u$) and CH ($A\ ^2\Delta \rightarrow X\ ^2\Pi$). A weaker CH emission is the so-called "3900 Angstrom System", due to a $^2\Sigma - ^2\Pi$ transition. In addition to these systems, the transition ($A\ ^2\Sigma^+ \rightarrow X\ ^2\Pi$) of OH is often a strong emission in the ultraviolet, and can be observed with an appropriate spectroscopic system. Examples of these emissions can be shown from our own work.

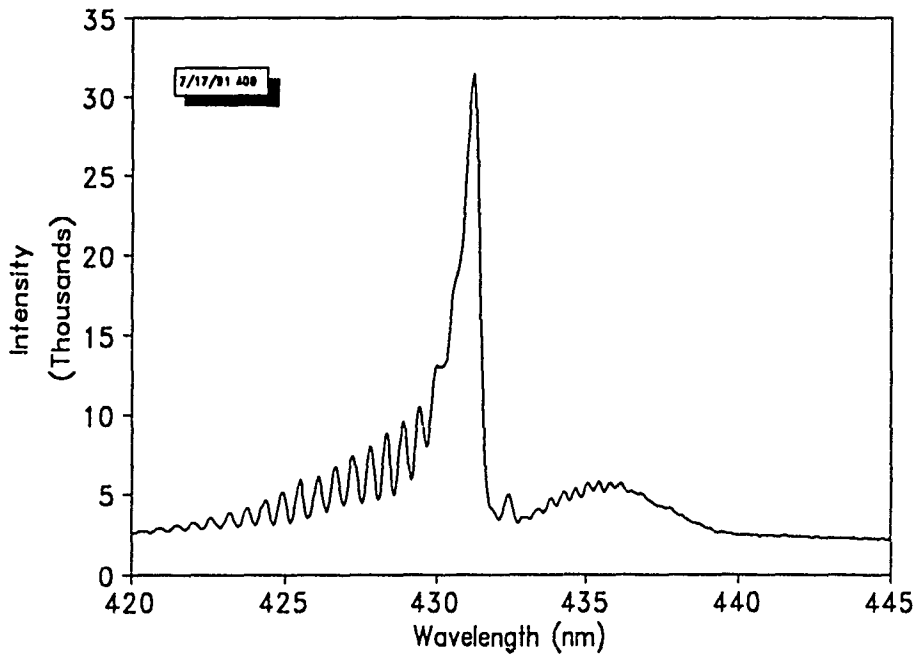
Figure 1 displays a portion of the Swan bands emitted by a slightly "rich" propane/air flame. The portion displayed in the figure includes the (0-0) band head at 516.52 nm and the (1-1) band head at 512.93 nm.

Figure 1
Portion of C2 "Swan Band"



This emission is the principal reason why such a flame exhibits a green reaction zone. Another strong feature of hydrocarbon flames is the CH emission displayed in Figure 2. Here, the Q (0,0) band head occurs at 431.50 nm, and the narrow, line-like feature at 432.4 nm is the Q(2 ,2) band. Further to the red is found the spread-out rotational structure of P(0 ,0) band, whose origin is at 438.4 nm. It would appear particularly promising that one could obtain useful information

Figure 2
Portion of CH Emission



about the flame temperature from this band, since rotational structure is so easily resolved (and even better resolution is easily achieved by recording the second-order monochromator reflection). But, because these species are produced in processes which result in non-Boltzmann distributions, and they do not achieve temperature equilibrium with the gases of the flame before they emit, they are not generally useful for determinations of flame temperature. Since the nascent populations are partially relaxed by the many collisions which occur before emission in a flame at atmospheric pressure, one cannot directly obtain information about the excitation processes in this way, either. For these reasons, we do not make any serious attempt to deconvolute the spectra of excited diatomic radicals. Significant differences in population distributions were noticed, however, and there may later be useful information obtainable from such an approach.

A portion of the "3900 Angstrom System" of CH is displayed in Figure 3. The portion shown includes the (1,1) band head at 402.53 nm and includes partially resolved rotational structure which is degraded to the red.

Figure 3
CH "3900 Angstrom System"

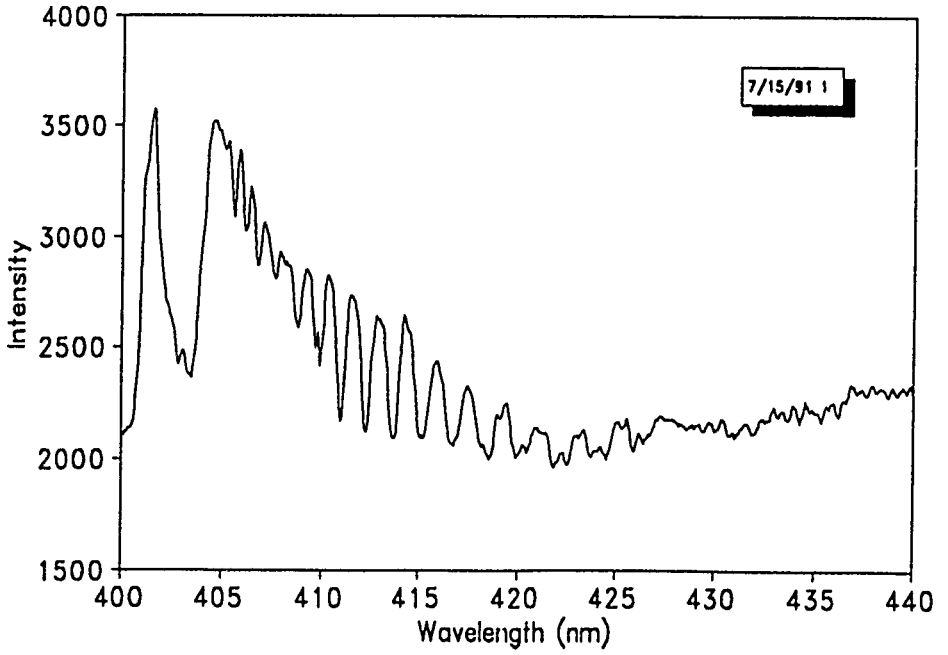
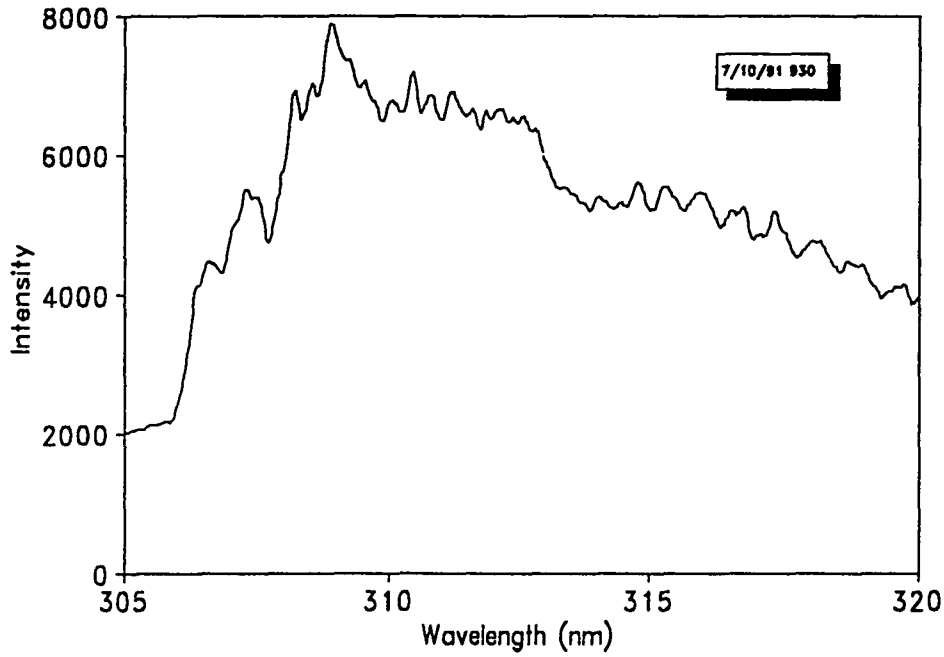


Figure 4 displays the complicated, $^2\Pi$–$^2\Sigma$ OH emission bands from a propane/air flame. Each band has four heads, with only the first, second, and fourth usually observable. In the present spectrum, none of the features is clearly distinguishable. The (0,0) (R_2) head occurs at the left of the spectrum shown, at 306.36 nm. The R_1 head is next to it, at 306.72 nm, and the Q_2 and Q_1 heads are 307.8 and 308.9 nm, respectively. It is clear that this band can be useful for determination of internal energy distributions only when high-resolution spectrometry allows the separation of the rotational bands.

Figure 4
OH Emission



Experimental

The spectra shown above were obtained using a modified burner of a type designed for atomic absorption spectroscopy by the Perkin-Elmer Corporation. This type of burner is intended to premix air with hydrocarbon fuel in the burner throat. Because atomic absorption requires injection of the material It is possible to aspirate a solution (originally the analyte) into the air flow stream. In our case, solutions of alkali metal salts, such as cesium or potassium hydroxide were sometimes so injected into the flame. In experiments in which no such metals were desired, pure deionized water replaced the alkali metal solution, so as to eliminate the seed ions without modifying the fuel/air ratio or the flow characteristics of the burner. The slot-type original burner head was replaced with a stainless steel burner w[illegible] fuel/air holes drilled in a pattern approximately 2.5 cm across.

A microscreen was installed below the effusion holes, so as to help mix the reagents without precluding the injection of metallic salts as a porous plug would have done. In addition, a second (and sometimes third) layer of stainless steel microscreen was placed on top of the burner head. These greatly improved the flame stability and eliminated the spatial variations which the holes tended to impose on the flame structure. With two screens atop the burner, the unperturbed flame was nearly conical in shape, very similar to the constant-velocity profile nozzle flame illustrated in *Flame Structure* by Fristom and Westenberg[11]. The burner was mounted on a specially-made X-Z translation stage so as to simplify the study of spatial differences by spectroscopy or potential measurements.

The spectrum of the flame or of seed atoms was obtained by focusing the emission with a lens onto the entrance slit of a Spex model 1870, 0.5 m monochromator. The optical system was arranged so that magnification of approximately 3x was obtained, allowing spatial resolution of somewhat better than 1 mm^2. A variable aperture (0.2 - 3 cm) was used to improve the spectral resolution of intense emissions, yet allow increased sensitivity (at the expense of some resolution) for weak features. The monochromator slit was continuously adjustable, and values from 10-100 μ were used at various times in this work. Spectra were recorded with a Tracor Northern Corp., TN 6134, 1024-channel photodiode array and associated TN 6500 electronics. All of the spectra displayed in this report were obtained by integrating emission intensities over a temporal window of 1.0 - 30.0 seconds, depending on the intensity of the band and the monochromator inlet slit setting. It was also possible with this system to record intensities integrated over a molecular band, so as to gain a measure of the relative concentration of a given excited species. Since the electronic system used did not include an output devices such as a printer or plotter, emission spectra of interest were permanently recorded (in a quaint TN format) on floppy disks, which were subsequently processed so as to be input to other programs which could produce acceptable graphical results. QuattroPro 3.0 and Sigmaplot 4.1 produced the graphs shown in this report.

Two different geometries were used for imposing the electrical field which perturbed the flame. At first, we used flat copper sheet electrodes mounted transversely, 2.0 cm apart and 0.2 cm above the burner head. One of these electrodes and the burner head itself were grounded, and the perturbing potential was applied to the second electrode. The electrodes were oriented parallel to the optical path, so that emission could be detected anywhere between the plates. The second geometry employed a stainless-steel cylindrical ring electrode, 2.5 cm in diameter located in an axially-symmetric position 2.0 cm above the burner head. The burner head was again grounded in this arrangement, and potentials up to 3000 volts were applied to the active electrode. Current between ground and the active electrode were measured with a conventional microammeter.

Electrical potentials in the flame were made using a tungsten rod mounted on a fixed stand, so that the flame could be moved through the X-Z plane with respect to the rod, in exactly the same way and with the same precision as is the case with the spatially-resolved spectroscopic measurements. The rod potential was measured through a Tektronix HV probe, model P6015 (3 pF, 100 MΩ) with a Tektronix oscilloscope. There was no convenient means to record these data electronically, so the results reported later in this report are the result of manual logging of the potentials and the associated approximate variations.

Attempts to Obtain Temperatures from Seeded Flames

One of the first measurements we attempted was to repeat and improve the observation that an electric field can significantly change the temperature of a hydrocarbon/air flame[3]. Since the emission spectra of atoms seeded into flames had been used for this purpose by others, it was expected that this would be a straightforward, preliminary measurement. However, several constraints complicate the application of this method[12]. First, the flame temperature must be high enough to populate the levels whose emission will be monitored. Propane/air flames at atmospheric pressure typically achieve just over 1900 Celsius maximum, and thus limit the usable upper levels to below 3 eV, if areas at temperatures below the maximum are to be surveyed. The emission must occur in an experimentally accessible region; because of the characteristics of our detector system, this meant approximately 300-910 nm. Measurement of the absolute intensity of a single spectral feature can, theoretically, yield the temperature of the system, but this requires careful absolute calibration of the optical system sensitivity, as well as knowledge of the number density of emitters and the active path length, and is applied to systems as complicated as a flame only with difficulty. For these reasons, the relative intensities of emissions which originate in atomic states separated by a few kT in energy provides a more reliable method.

While it is preferable to have several emission intensities so as to ensure that equilibrium is being achieved, as few as two lines can be used in this so-called "radiance ratio method". One needs the transition probabilities for the measured transitions, and errors in the ratios of the transition probabilities propagate into large errors in absolute temperature, so the method is usually applied only with atoms whose transition probabilities are well-known from theory and/or experimental measurements (for temperature differences this factor is not so important). Iron has been used for measurements of this type[13,14], and we considered it for the present experiments. However, the spectral resolution required (.035 nm) for measurement of the iron line intensities was far beyond that possible with a 0.5 m monochromator.

Cesium has reportedly been used for temperature measurements of hydrocarbon/air flames[15] and plasmas[16]. While we observed a number of Cs transitions (See Table I), none of them proved useful for our purpose. The intense, resonance lines at 852.11 and 894.35 nm, which had been reported by Hollander and Broida[16] as being useful for measurements of flame temperature, were found by us to be seriously radiation-trapped, even at quite low concentrations. Perhaps some additional study using these lines could be undertaken, but linear relationships between number density and emission intensity must occur below the lowest number densities we tested, which was the result of injecting solutions in the 10^{-5} M CsOH range.

Table 1

Cesium line emissions observed in propane/air flames

Transition	λ (nm)	g_n	A_{mn}[17]	E_n (eV)	Comments
$6s\ ^2S_{1/2}$- $6p\ ^2P_{1/2}$	894.3	2	3.51×10^{-1}	1.38	Strong resonance line
$6s\ ^2S_{1/2}$- $6p\ ^2P_{3/2}$	852.1	4	7.22×10^{-1}	1.45	Radiation-trapped
$6s\ ^2S_{1/2}$- $7p\ ^2P_{1/2}$	459.3	2	2.53×10^{-3}	2.70	
$6s\ ^2S_{1/2}$- $7p\ ^2P_{3/2}$	455.5	4	1.48×10^{-2}	2.72	
$6p\ ^2P_{1/2}$- $8s\ ^2S_{1/2}$	760.8	2	1.86×10^{-2}	3.01	K: 766.4, 769.8 nm
$6p\ ^2P_{1/2}$- $8s\ ^2P_{1/2}$	794.3	2	1.86×10^{-2}	3.01	Rb interference-794.7
$6p\ ^2P_{3/2}$- $6d\ ^2D_{5/2}$	917.2	6	3.35×10^{-1}	2.80	
$6p\ ^2P_{3/2}$- $6d\ ^2D_{3/2}$	920.8	4	4.14×10^{-2}	2.80	
$6p\ ^2P_{1/2}$- $6d\ ^2D_{3/2}$	876.1	4	3.21×10^{-1}	2.80	
$6p\ ^2P_{3/2}$- $7d\ ^2D_{5/2}$	697.3	6	9.51×10^{-2}	3.23	
$6p\ ^2P_{1/2}$- $7d\ ^2D_{3/2}$	672.3	4	9.63×10^{-2}	3.23	
$5d\ ^2D_{3/2}$- $7f\ ^2F_{5/2}$	682.4	6	3.20×10^{-2}	3.61	
$5d\ ^2D_{5/2}$- $7f\ ^2F_{7/2}$	687.0	8	3.35×10^{-2}	3.61	
$5d\ ^2D_{3/2}$- $7f\ ^2F_{5/2}$	722.8	4	5.55×10^{-2}	3.61	
$5d\ ^2D_{5/2}$- $6f\ ^2F_{5/2}$	727.9	6	5.86×10^{-2}	3.51	Possible Rb interfere
$5d\ ^2D_{5/2}$- $6f\ ^2F_{5/2}$	727.9	8	5.86×10^{-2}	3.51	Rb:780.0, 794.7 nm

Field-Induced Changes in the Spatial Distribution of Flame Emission

The change in the appearance of a susceptible flame as a modest field is applied is dramatic and remarkable. While many experiments were performed with a field applied transverse to the flow of reagents, the largest, most consistent and reproducible results were obtained with the axial electrode described above. Results for only that geometry are reported here. Surveys of the spatial variation of several of the excited radicals accessible to us were recorded as functions of electric field, fuel/air ratio, and total flow rate. The results illustrated below were chosen because the conditions produced a particularly dramatic effect. In this case, a potential of 2000 volts was applied to the axially-symmetical electrode 2.0 cm above the burner top, producing a mean field of 1000 volts/cm. The propane/air ratio was 1:12.8 by volume, so that this flame was "rich"; that is, there was more fuel in the premixture than would be required by the 1:5 propane/O_2 stoichiometry of the reaction:

$$C_3H_8\,(g) + 5\,O_2(g) = 3\,CO_2\,(g) + 4\,H_2O\,(g)$$

(although the effective ratio in the experiment might also include some participation by environmental air). This flame was quite green, due to the C_2 Swan band emission.

Figure 5
Change in Flame Due to Applied Field

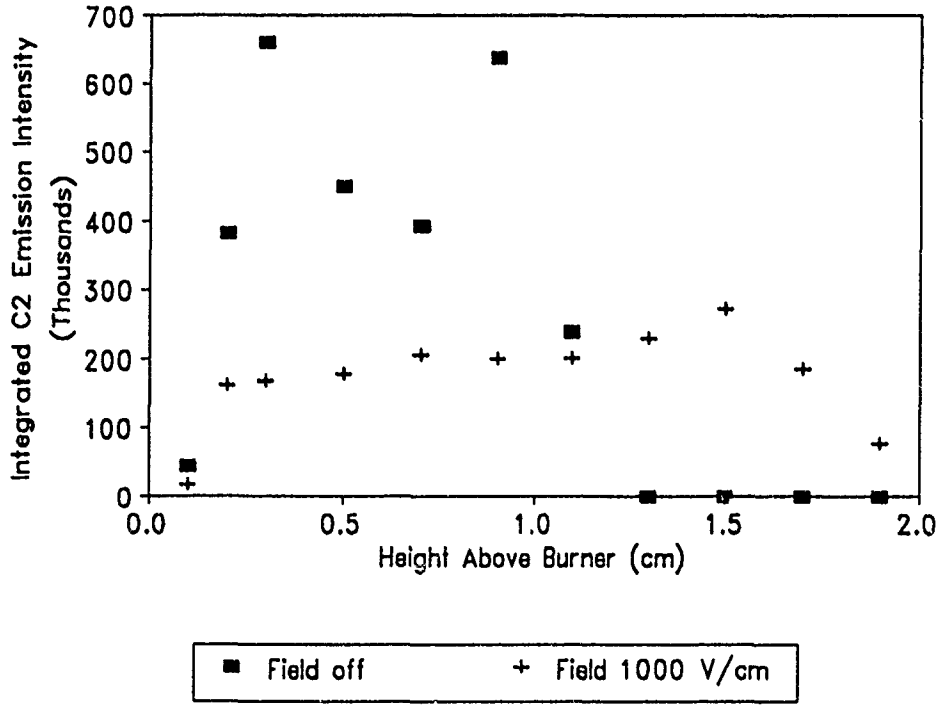


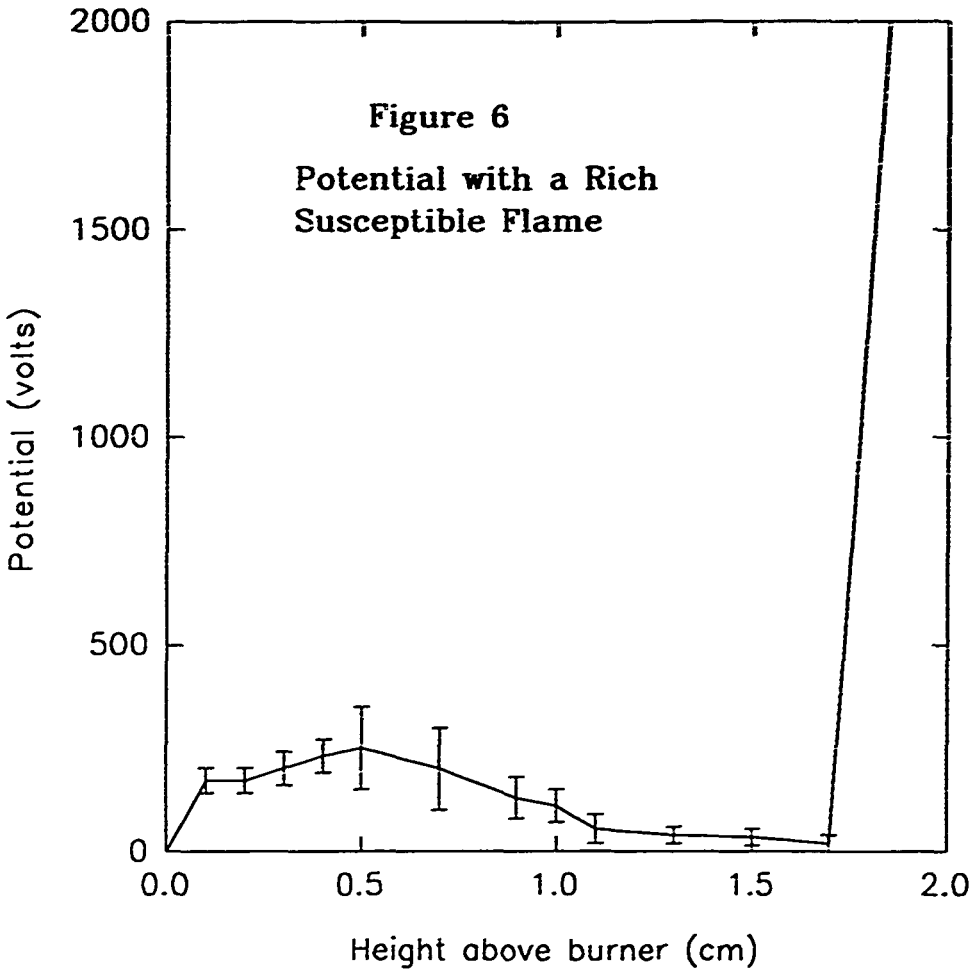
Figure 5 shows the result of measuring the integrated intensity of the C_2^* emission near 510 nm, as the flame was moved vertically, with a horizontal position such that light was collected from the center of the flame. As may be seen, the electric field effectively diminishes the height of the reaction zone by nearly half. The intensity of the emission in the field-perturbed flame also appears to increase, but this could be partially due to the fact that the effective optical path is increased when the flame height is decreased. While it is believed that the sensitivity of the detection system was the same in both cases, quantitative treatment of the spatially-integrated intensity would require consideration of the source path length.

An applied field affects flames at any other fuel/air ratio less than it does at this one. As the mixture becomes even more rich, the temperature (presumably) decreases, and hence fewer ions and electrons are produced. For this reason, it is not surprising that the influence of electric fields decreases when the air flow is decreased. However, since the maximum temperature of hydrocarbon flames is always found near the stoichiometric ratio, it might be expected that, if an "ion-wind" (a nonspecific ion-induced mixing) mechanism were most important, that such flames would be more susceptible to deflection by the field, and such is not found to be the case. At the stoichiometric ratio, the flame height and morphology are little changed by the field. (Such a flame exhibits less C_2^* emission and more CH^* emission, as described above.) One can also burn propane/air mixtures well beyond the stoichiometric ratio, toward the "lean", or excess-oxidant side. At approximately the ratio at which C_2^* emission becomes instrumentally undetectable, the influence of the electric field also disappears.

Measurements of Potentials Within the Flame

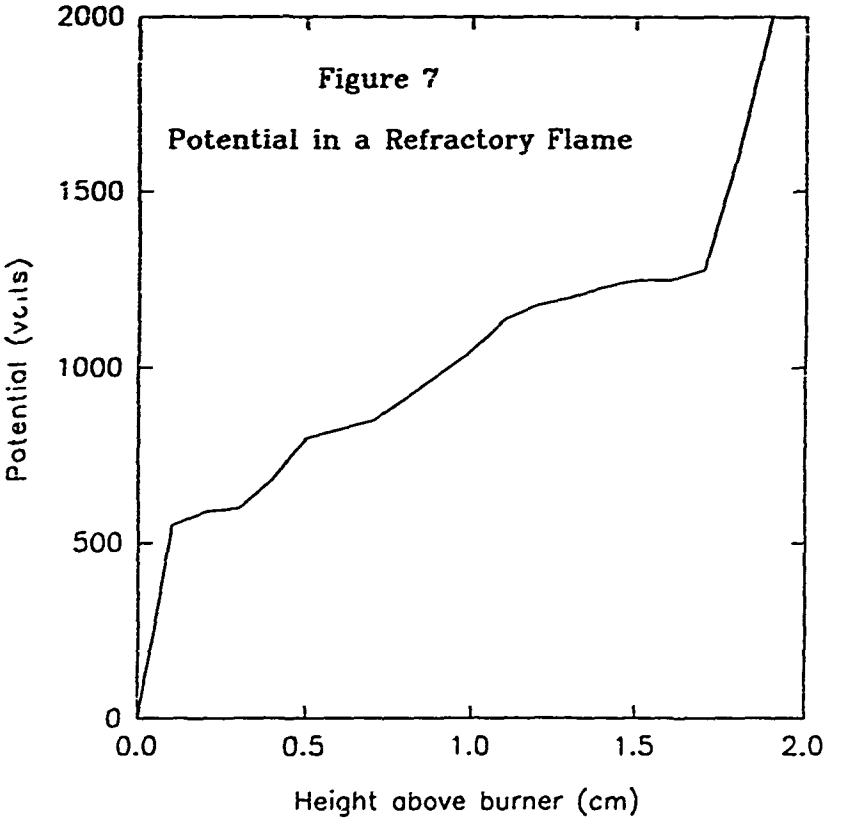
Electrical measurements of the flame were also undertaken, using equipment as described earlier. These potential and current measurements were undertaken primarily for qualitative rather than quantitative purposes, because the quantitative characterization of the electrical properties of a flame generally require more sophisticated methods than those available for this project[18]. We used the current through the flame only as a very approximate measure of the amount of ionization, and the changes in potential distribution only as an indication of changes in the distribution of ionized species in the flame. Despite the high-impedance probe used to make the measurements, there is no doubt that the tungsten wire perturbed the flame, both electrically and chemically, since such a hot metal surface can catalyze certain chemical reactions which might not occur in a purely gas-phase system. In some very rich flames, solid graphite (presumably) had a tendency to collect on the probe when it was inside the reaction zone, whereas soot was not produced in noticable amounts, even when the fuel/air ratio was so rich as to constitute a diffusion flame.

Figure 6, below, shows the apparent potential within a flame similar to the one whose large susceptibility to distortion by an electric field was illustrated in the previous figure. For this measurement, the upper electrode potential was +2000V, and the current was 100 μA. The green flame very shortened, and exhibited a visible high-frequency spatial fluctuation. The probe was positioned in the center of the flame, so that its tip visited the center of the reaction zone.



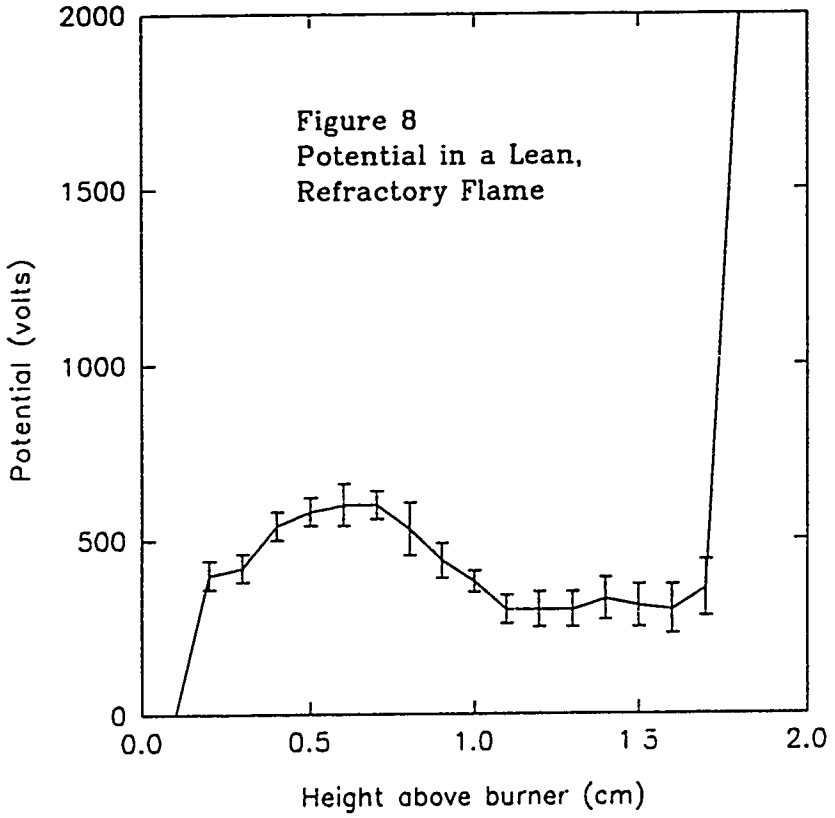
We believe that the large potential gradient at the bottom of the flame may be important, as this characteristic seemed to be common in susceptible flames. Although we are not showing the potential distribution data in this report, when the sign of the applied potential is reversed, the flame is completely unaffected by the field. In fact, the reversed-potential arrangement produces no measurable effect even when the applied field is 1500 V/cm. The apparent potential distribution is dramatically different, with all of the potential drop occuring within one or two millimeters of the electrode. Therefore, the entire reaction zone in such a case is within an equipotential, and it is not surprising that no deflection occurs.

The rate of fuel flow in Figure 7 was the same as in Figure 6, but the air flow rate was increased beyond the stoichiometric ratio, where the current is maximum, so as to produce a "lean" flame with the same current as was the case for the "rich" flame. We presume that these two flames had roughly the same number of charge carriers and, if the "ionic wind" were the principle cause of the deflection, neglecting the fact that the total flow was larger, a similar magnitude of the deflection could be expected. However, this flame exhibited only a small amount of fluttering at the top of the reaction cone, and no significant shortening. Notice that the potential gradient across the bottom of the flame is much smaller than it was in a susceptible flame.



Continuing in the same series, we next produced a lean flame with the same total flow rate as that which produced Figure 6, and with the same 100 μA current as was measured for that flame. Since the air flow is a large fraction of the total, these conditions were very near what one would obtain by, while maintaining the air flow as produced Figure 6, reducing the fuel flow until the current again fell to 100 μA. Naturally, this flame was much smaller (a total height of about 5 mm), even before the electric field was turned on, than was the original flame. The resulting potential distribution, shown in Figure 8, cannot be explained within the simple

assumptions of an ohmic theory. Clearly, something unusual is likely to be required to rationalize the maximum in the apparent potential. This flame also exhibited some periodic fluctuations in the audible range of frequencies (although no sound was audible) in the potential near the top of the flame. Phenomena like this have been previously studied in so-called "singing flames", although the mechanism has not been elucidated[19]. This experiment proves that the differences between the susceptible flame in Figure 6 and the refractory one in Figure 7 is not just because the flow rates are different, since the flame in Figure 8 was not measurably deflected by the field.



Of course, seeding a flame with an alkali metal can very sharply change the potential distributions by providing a source of facile ionization, and therefore producing both ions and electrons in the flame. We have not found in the literature (although such measurements may exist), evidence that the effect of seeding may be qualitatively different in rich as compared to lean flames. Figures 9 and 10 address this point. Figure 9 displays the potential distribution in a refractory, lean flame seeded with KOH solution, while Figure 10 shows the distribution in a similar system, with the same air flow rate, but with the fuel flow increased so as to produce the same, 180 μA, current. It is interesting to note that this flame is not perturbed by the field, even though the apparent potential distribution indicates as large a potential gradient at the burner as was found for other, susceptible flames. This tends to indicate that a specific kinds of ion(s) responsible for altering the reaction kinetics, and that K^+ does not do so.

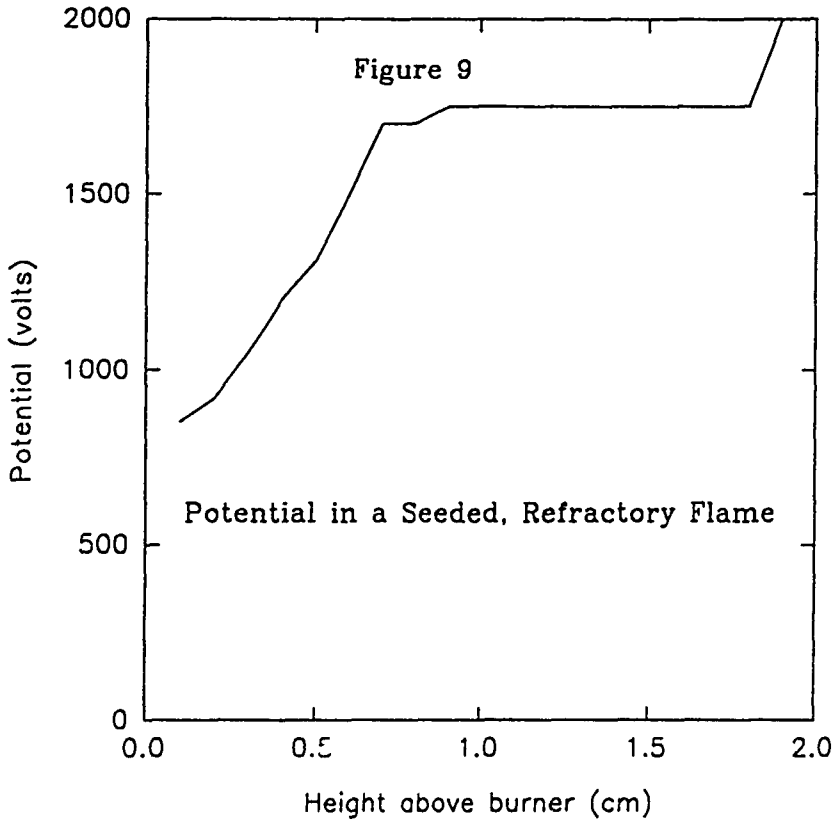


Figure 9
2000
1500
1000
500
0
Potential (volts)
Potential in a Seeded, Refractory Flame
0.0
0.5
1.0
1.5
2.0
Height above burner (cm)

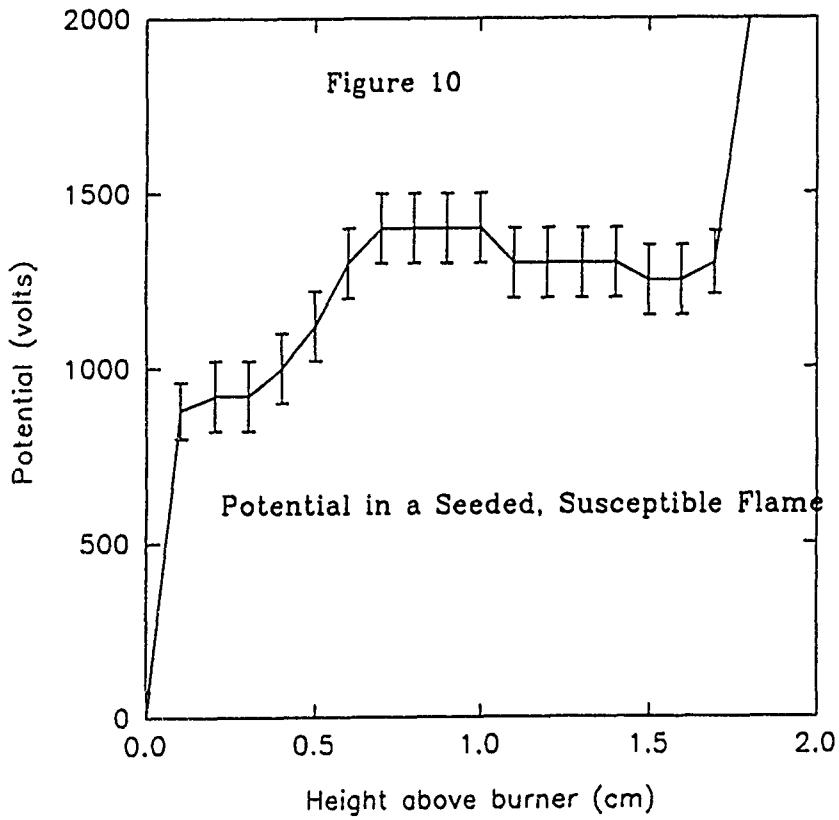


Figure 10
2000
1500
1000
500
0
Potential (volts)
Potential in a Seeded, Susceptible Flame
0.0
0.5
1.0
1.5
2.0
Height above burner (cm)

Conclusions

These studies have shown that there is a large difference in the susceptibility of propane/air flames to deflection or distortion by electric fields, depending on the ratio of fuel to air in the premixed reagents. While slightly "rich" flames are highly susceptible to such perturbation, slightly "lean" flames with similar concentrations of charged species are slightly affected, if at all. The influence of modest fields on susceptible flames can be very large. Our spectroscopic surveys of such flames indicate that the height of the reaction zone can be caused to decrease by approximately half, by a field of less than 1 KV/cm. Our results suggest that some chemical species which exist in rich flames, and which may be associated with the production of excited C_2 radicals, are necessary for high flame sensitivity to electric fields. Crude measurements of the apparent potentials within susceptible and refractory flames show that susceptible flames tend to have a large potential gradient at the bottom of the flame, (the gradient likely extends across the reaction zone). However somewhat lean flames which are not sensitive can be made so by seeding with potassium salts. Since this sensitivity is not accompanied by C_2 emission, we believe that it is more likely a result of a change in potential gradients than a gross change in flame chemistry.

References

[1]James Lawton and Felix Weinberg, *Electrical Aspects of Combustion*, Clarendon Press (Oxford), 1969

[2]G. A. Gulyaev, G. A. Popkov, and Yu. N. Shebeko, "Effect of a constant electrical field on combustion of a propane-butane mixture with air", *Fizika Goreniya i Vzryva*, **21**, 23-25 (1985).

[3]B. Lewis and C. D. Kreutz," The Effect of an electric field on the flame temperature of combustible gas mixtures",*J. Am. Chem. Soc.*, **55**, 934 (1933).

[4]S. S. Sandhu and F. J. Weinberg, "Laser interferometric studies of the control of heat transfer from flame gases by electric fields", *Combust. Flame*, **25**, 321-334 (1975).

[5]H. F. Calcote and C. H. Berman, "Increased methane-air stability limits by a DC electric field", *Fossil Fuels Combustion Symposium PD- Vol. 25* and references therein.

[6]F. B. Carleton and F. J. Weinberg, Electric field-induced flame convection in the absence of gravity", *Nature*, **330**, 635-6 (1987).

[7]G. A. Gulyaev, G. A. Popkov, and YU. N. Shebeko, "Synergistic effects in the simultaneous actio of an electric field and an inert diluent on gas-phase flames",*Fizika Goreniya i Vzryva*, **23**, 57-9 (1987).

[8]M. Kono, F. B. Carleton, A. R. Jones, and F. J. Weinberg, "The effect of nonsteady electric fields on sooting flames", *Combust. Flame*, **78**, 357-64 (1989) and references therein.

[9]H. F. Calcote and D. E. Jensen, "Ion-molecule reactions in flames", in *Ion-Molecule Reactions in the Gas Phase*,,"Advances in Chemistry Series, No.58", American Chemical Society, Washington, D. C. (1966); G. B. Kistiakowsky and J. V. Michael, "Mechanism of Chemi-Ionization in Hydrocarbon Oxidations", *J. Chem. Phys*, **40**, 1447 (1964).

[10]J. A. Green and T. M. Sugden, *Ninth Symposium (International) on Combustion*, , p. 607, Academic Press, New York (1963).

[11]Figure II-8 (a) and II-10 of *Flame Structure*by R. M. Fristrom and A. A. Westenberg, McGraw-Hill (New York), 1965.

[12]Richard H. Tourin, *Spectroscopic Gas Temperature Measurement*,Elsevier Publishing Company (Amsterdam, New York, London) 1966.

[13]H. P. Broida and K. E. Shuler, "Spectroscopic study of electronic flame temperatures and energy distributions", *J. Chem. Phys*. **27**, 933 (1957).

[14]G. F. Kirkbright, M. Sargent, and S. Vetter, "The selection of line-pairs for the iron 'two-line' method of flame temperature measurement", *Spectrochimica Acta*, **25B**, 465 (1970).

[15]Tj. Hollander and H. P. Broida, "Zeeman scanning of absorption line profiles in flames", *J. Quant. Spectrosc. Radiat. Transfer*,**7**, 965 (1967).

[16]M. Fabry, J. R. Cussenot, and M. Numano, "Interpretation of excitation temperature measurements in a cesium plasma", *J. Quant. Spectrosc. Radiat. Transfer*,**15**, 811 (1975).

[17]Oscillator strengths from M. Fabry, "Theoretical and experimental determinations of cesium oscillator strengths", *J. Quant. Spectrosc. Radiat. Transfer*, **16**, 127 (1976).

[18]See reference 1, especially Chapter 5, "Experimental Methods for the Study of Flame Ionization".

[19]S. A. Brukov, V. V. Kurzhunov, and V. N. Mezdrikov, "Effect of an electric field on the vibratory combustion of propane", *Fizika Goreniya i Vzryva*,**2**, 68 (1966).

1991 USAF-RDL, INC. SUMMER FACULTY AND GRADUATE STUDENT RESEARCH PROGRAM/

Sponsored by the

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

Conducted by the

Research Development Laboratories, Inc.

FINAL REPORT

Investigation of the Combustion Characteristics of Swirled Injectors in a Confined Coannular System with a Sudden Expansion

Prepared by:	Paul O. Hedman, Professor David L. Warren, Master Candidate
Departments and	Chemical and Mechanical Engineering
University:	Brigham Young University Provo, Utah 64602
Research Location:	Aero Propulsion and Power Laboratory Wright-Patterson AFB, OH 45433
USAF Researcher:	W. M. Roquemore, Ph.D.
Date:	September 11, 1991

ABSTRACT

This report contains a brief summary of the work done to investigate the operational characteristics of a burner that was designed to "specifically reproduce recirculation patterns and LBO processes that occur in a real gas turbine combustor." The burner, refered to as the Pratt & Whitney Task 150 Combustor, uses a swirling fuel injector from an actual Pratt & Whitney turbojet engine installed in a sudden expansion combustor that closely simulates the geometry of a combustor from an actual jet engine. The Task 150 configuration has been configured so that the geometry around the injector is nearly axi-symmetric, but the combustor incorporates quartz windows so that optical (laser based) instruments can be used to make measurements in the flame. The Task 150 configuration uses a swirling injector similar to those used in the Task 200 combustor, and the inlet diffuser sections and inconel chimney of the Pratt & Whitney Task 100 burner. This unique configuration allows complex diagnostic measurements to be measured in a simpler geometry than the Task 200 combustor, but embodies most of the features of an actual jet engine combustor in an axi-symmetric configuration that is easier to mathematically model.

The primary effort during this summers AFOSR sponsored research program for faculty and graduate students was to assemble the hardware, and investigate the basic operational behavior of the burner. It was found that the flame would exhibit very different operating characteristics over the operating stoichiometric range. The flame would be attached to the burner or lifted from the burner as the fuel equivalence ratio was changed. Film images and video tape of the various operating modes as a function of fuel equivalence ratio were obtained.

Measurements of the fuel equivalence ratio at lean blow out as a function of air flow rate were also obtained. At high air flow rates, the flame would blow out for both injectors tested from a well lifted flame that was being stabilized on a downstream recirculation zone. At low air flow rates, the low swirl injector would still blow out from the flame stabilized on the down stream recirculation zone. However, at low air flow rates, the flame on the high swirl nozzle would remain attached to the center core of the fuel injector right up to lean blow out. As a consequence, the fuel equivalence ratios (based on total fuel and total air flow) where lean blow out occurred with the high swirl nozzle were very much lower (ca 0.3) than that observed for the low swirl injector or the high swirl nozzle at high air flow rates (ca 0.5).

The pressure drop across the low swirl and high swirl injectors were determined as a function of total air flow rate through the combustor. In a separate investigation, the partitioning of the flows through the dome jets, the insert jets, the primary air swirler and the secondary air swirler was determined. The effect of fuel flow on the pressure drop across the injectors and its effect on the flow partitioning was also investigated by introducing CO2 into the fuel passage at varying flow rates. How the air flows partition between the various flow passages in an injector is expected to have a major impact on how well that injector operates.

Measurements of wall pressures and wall temperatures were made in an attempt to better understand the locations of the various flame zones in the combustor. These measurements have indicated that the location of the recirculation zones moves very little as operating condition is changed. They also indicated a second recirculation zone in the dome region that is caused by the dome cooling jets.

I. INTRODUCTION

This report presents a brief summary of results of an initial investigation to determine the flame characteristics when swirling fuel injectors from actual Pratt-Whitney jet engines are installed in a burner with a sudden expansion (Pratt and Whitney Task 150 Combustor). This work provides a bridge between the work where the combustion characteristics of confined, coannular fuel and air jets are discharged into a sudden expansion (Pratt and Whitney Task 100 Combustor), and the Task 200 combustor where four swirling fuel injectors from actual Pratt-Whitney jet engines are installed in a a rectangular combustion chamber that simulates a segment of a real jet engine combustor. The advantage of the Pratt and Whitney Task 150 Combustor is that it allows the actual combustion characteristics of a real injector to be

investigated in a simpler geometry where various diagnostic measurements (primarily laser based optical measurements) can be more easily made. The Task 100 and 150 combustor configurations have been specifically developed to study the phenomenon of lean blowout (LBO) in modern annular aircraft gas turbine combustors. The combustor has been carefully designed (Sturgess, et al. 1990) to "specifically reproduce recirculation patterns and LBO processes that occur in a real gas turbine combustor."

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

The Task 150 combustor configuration utilizes the basic Task 100 hardware, but replaces the confined, coannular jets with an insert and an actual swirling fuel injector from a Pratt and Whitney jet engine. A schematic drawing of the Task 150 Combustor is shown in Figure 1. A drawing that shows the installation of the fuel injector in greater detail is presented in Figure 2. Two different fuel injectors were used for this study, a high swirl injector, and a low swirl injector. The high swirl injector is refered to by representatives of Pratt and Whitney as a "bill of materials injector" used in production engines. The low swirl injector was reported to match the characteristics of the injectors supplied by Pratt and Whitney for use in the Task 200 combustor.

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

The results of this study relate to a flame blowout modeling study being conducted by other investigators from Pratt and Whitney (Sturgess, et al. 1990). A secondary purpose of this study has been to collect data that could be used in validating a computer code that predicts the flame phenomena and blowout limits (Sturgess, et al. 1990).

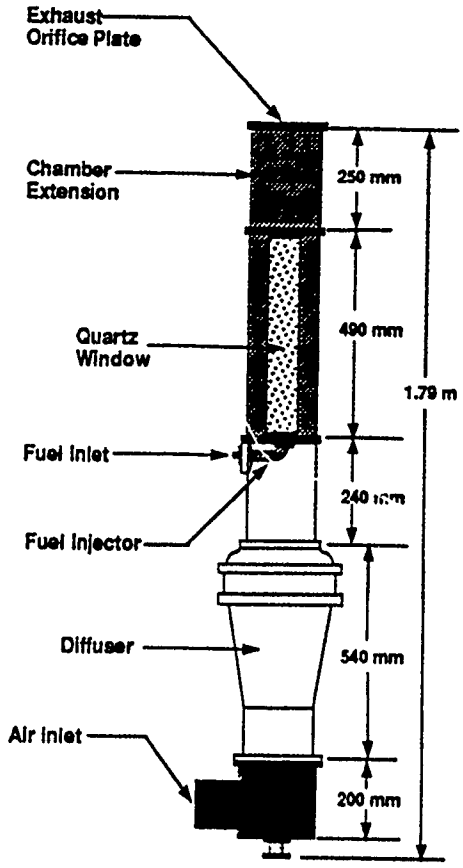


Figure 1. Schematic of the Task 150 Pratt and Whitney Combustor

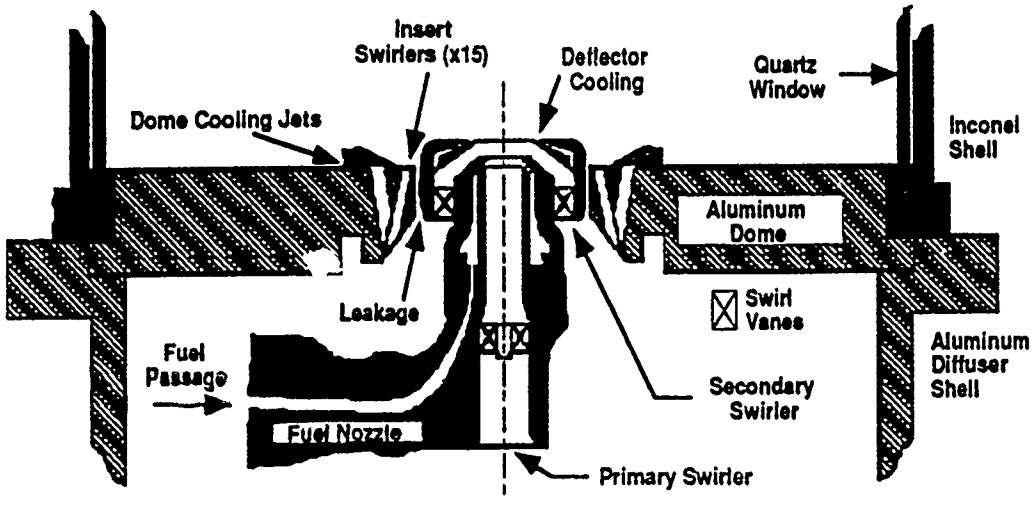


Figure 2 - Task 150 Burner Configuration

The page constraints of this report prevent a detailed summary and discussion of all of the experimental results obtained. Consequently, only example results are presented. A more complete compilation of the data has been prepared in the form of an appendix which has been supplied to the Air Force sponsor at Wright Patterson Air Force Base.

II. RESULTS

Calibration of Flow Meters

The main instruments used in these test series were four flow meters. These four included one for the air (6000 slpm air capacity), another for the nitrogen (1000 slpm air capacity), and two (300 and 45 slpm air capacity) for the fuel. The largest reference standard available was a laminar flow cell with approximately a 300 slpm limitation. This arrangement worked satisfactorily for the fuel meters, and showed the nitrogen meter to be linear up to the flow cell's limit. However, the air meter, not being designed to run at this low of flow rate, showed serious but predictable inaccuracies. Careful calibration against the laminar flow cell at low flow rates and against the linear nitrogen flow meter at higher air flow rates allowed the air flow meter to be calibrated over the flow rate range of interest. Figure 3 presents the final calibration of the air flow meter used in this study.

The calibrations of the fuel flowmeters were found to be very close to linear, and to match the factory calibrations very closely. The calibration equations obtained for the two fuel meters used are shown below.

0 to 300 slpm air (0 to 108 slpm propane) meter:

$$Y \text{ (actual flow)} = 0.99158 + 0.988 * X \text{ (indicated flow)}; \qquad R^2 = 1.000$$

0 to 45 slpm air (0 to 16.2 slpm propane) meter:

$$Y \text{ (actual flow)} = -0.018719 + 0.938 * X \text{ (indicated flow)}; \qquad R^2 = 0.999$$

Flame Characterization

One of the traits of the Task 150 burner is the wide variety of flame structures that are observed. Each flame shape indicates a different mode of operation, which differ from one another in the location of the flame fronts, or by some structure such as thickness or intensity. These different structures might arise from changes in the flow fields, mixing patterns, or fuel equivalence ratio. The differences and similarities of the flame structures for both nozzles together with the Task 100 configuration may provide significant

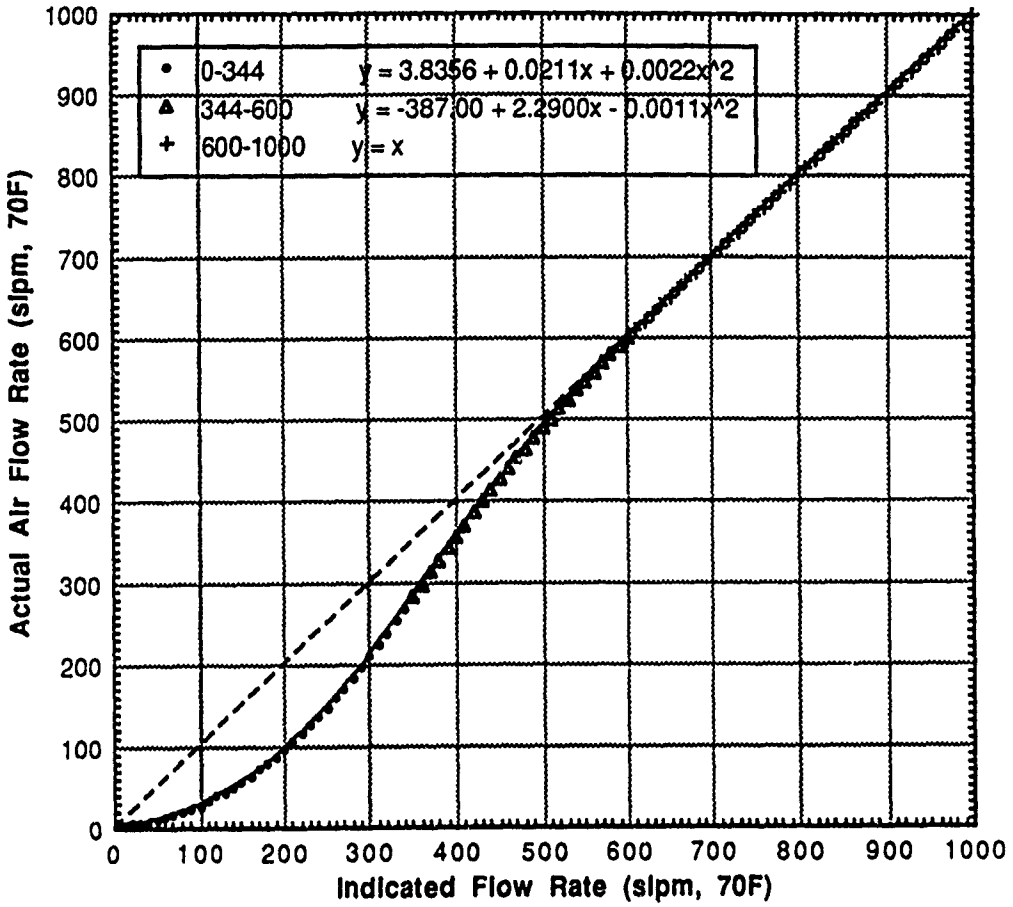


Figure 3 - Final Air Flow Meter Flow Calibration

insights to the combustion processes. Characterization simply defines the modes and the differences that differentiate the structures.

The shape of the flame, at the minimum, provides qualitative information on the mixing process and location of flame fronts. Such information can yield precious insights into what processes are present. Flames fronts exist because fuel and oxidizer have been transported to a point where combustion can be supported. The location of these fronts relative to the outlet orifices of fuel and air are of obvious interest.

Flame structure is also important in the macroscopic effects it can have. In the high swirl burner near lean blow out, for example, the flame is very small, with much of the air bypassing the actual combustion region. This small flame structure supported much lower fuel equivalence ratios than the comparable low swirl injector, which had a much larger flame structure to maintain. These structures not only have an effect in

laboratory experiments but also affect the real-life operation of a turbojet engine. Understanding of the causes and effects of flame structures may be of significant help to the injector designer.

The fuel equivalence ratios where the transitions from one flame structure to another were determined as a function of air flow rate. The flames for both injectors were attached to the outside of the insert air jets when the burner was operated very fuel rich. The flame would then lift, reattach, and lift again as the fuel equivalence ratio was progressively reduced depending on the injector (high swirl versus low swirl) and the air flow rate. During the reattachment phase, the flame would take on many of the characteristics of a tornado. Consequently, the terminology of funnel cloud, tornado, and debris cloud were adopted to describe some of the observed flame structure.

The flame characterization of the Task 150 combustor was in large part carried out by visual observations. Still film photographs were taken of the different structures. These images were digitized and manipulated using various computer programs into the isochromatic contour plots found in Figures 4 and 5 for the low swirl and high swirl injectors respectively. As can be seen, the flame for both nozzles look alike at very fuel rich conditions. The flame is attached to the insert jets, in a manner similar to the Task 100 burner at rich conditions. Unlike the Task 100, these flames are very short, presumably because the oxygen in the air was being depleted much faster due to the swirling motion mixing the gases. As the amount of fuel was reduced, the flames lifted and stabilized on some downstream recirculation zone that appeared to be associated with the injector. The primary combustion zone lengthened as relatively more oxygen entered the chamber. At this point the two nozzles diverged in flame characteristics. The low swirl nozzle would abruptly switch to an attached cone flame. Although this cone seemed similar in appearance to the high swirl flame near lean blow out, it was much different. The attached cone was much more vigorous, and produced a howling roar, giving no indication of being near blow out. As fuel was further reduced, the attached cone would once again lifted would stabilize on a downstream recirculation zone associated with the injector. At this point, however, a flame structure could still be perceived, as illustrated in Figure 4. As the fuel equivalence ratio continued to drop, the cone dissipated and became a shapeless, separated flame that appeared to be stabilized on the larger recirculation zone associated with the chamber itself. It is believed in this structure, that the nozzle no longer exerted an influence on the flame, and the flame was being stabilized on conditions sustained by the chamber itself. With the lean lifted flame structures, flames with the low swirl nozzle once again took on the appearance of flames seen in the the Task 100 combustor.

With the high swirl nozzle, after the rich lifted condition, the swirling air began to stabilize a flame in the center of the combustor, with a result that behaved much like a tornado. As seen in Figure 5, a funnel cloud formed within the the rich lifted flame and gradually descended as the fuel flow was decreased. Eventually, the funnel appeared to disappear into the inside of the primary swirler passage in the injector.

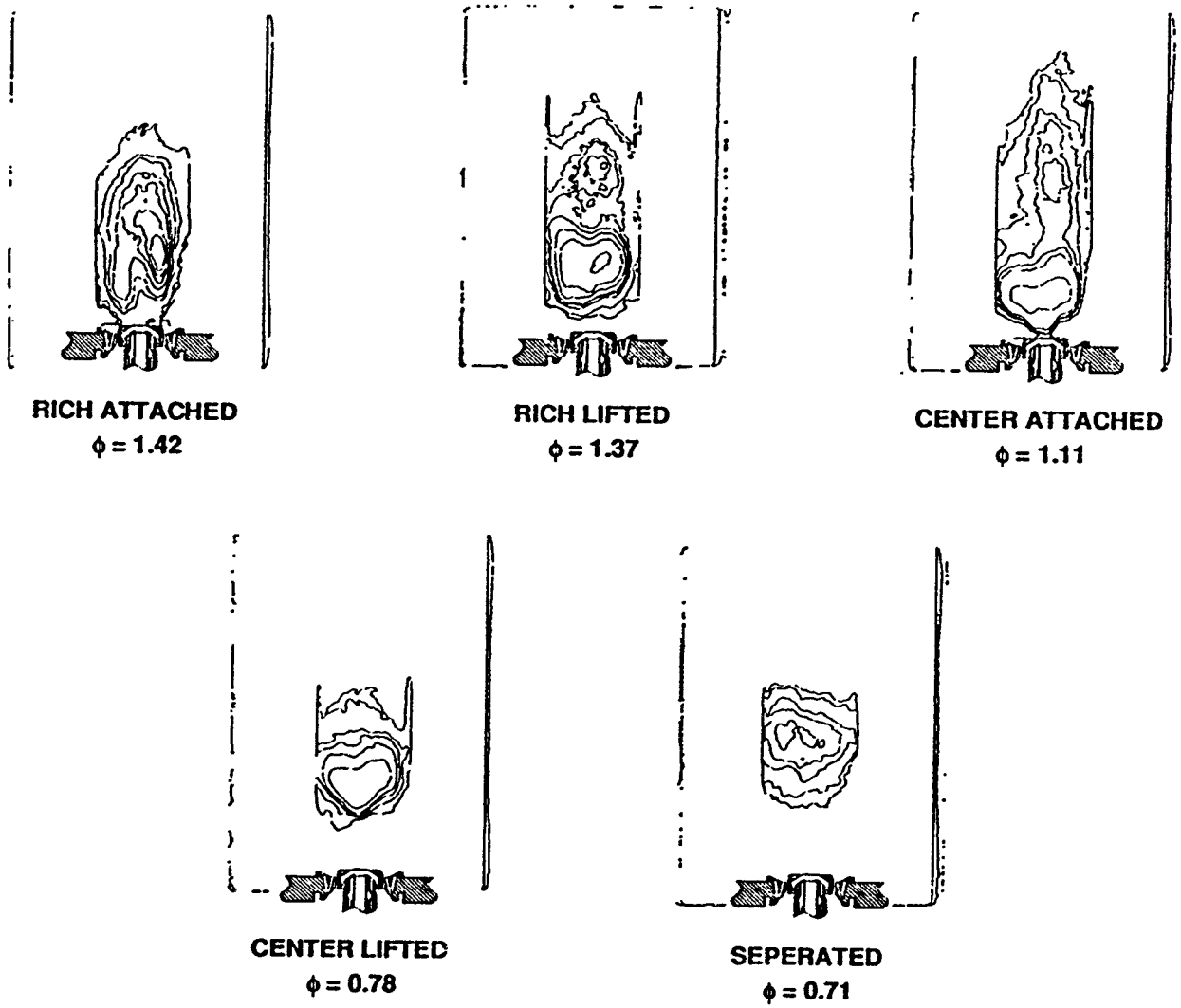


Figure 4. The five different flame structures observed with the low swirl nozzle. (Air flow = 310 slpm, 70 F).

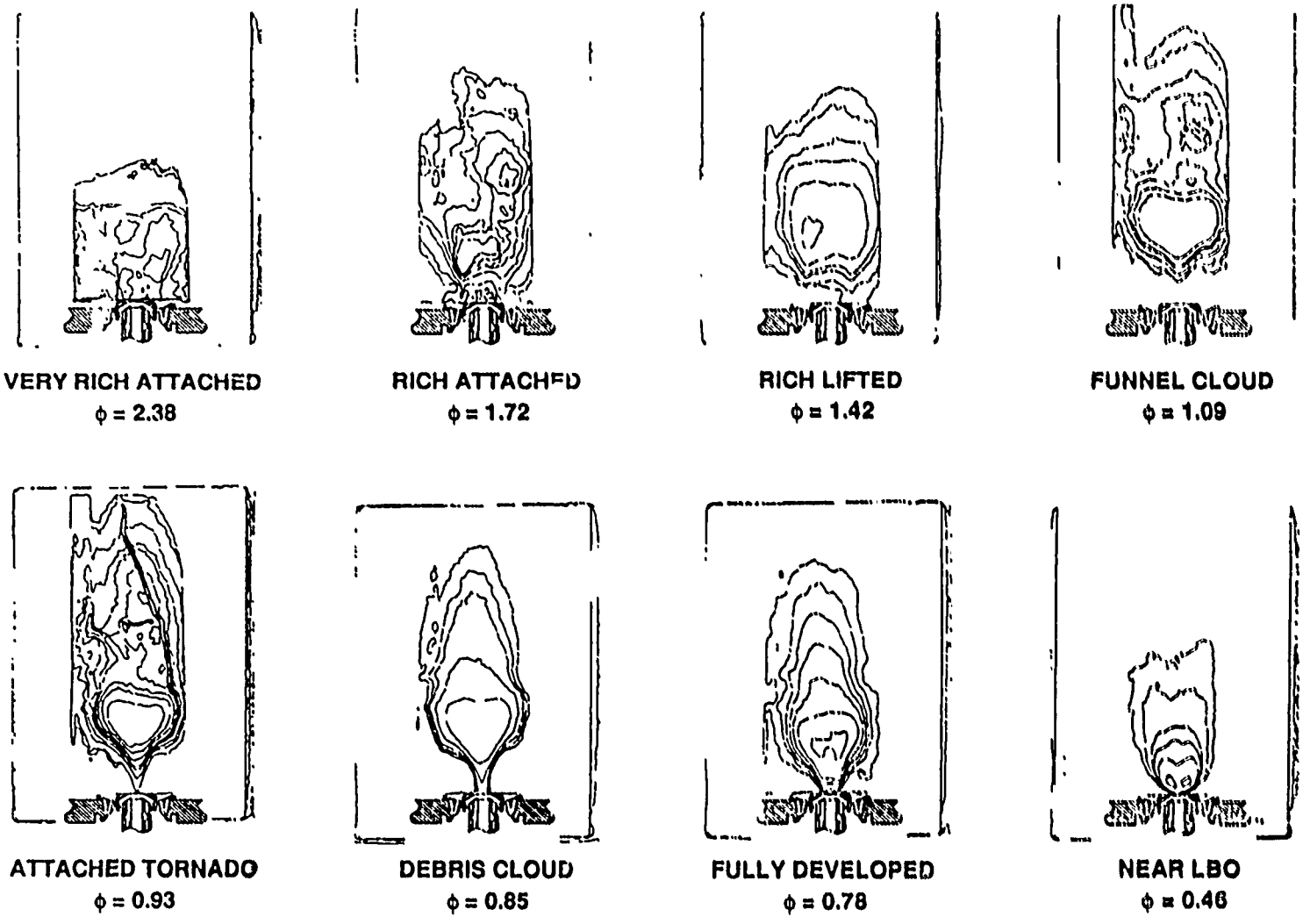


Figure 5. The eight different flame structures observed with the high swirl nozzle. (Air flow = 310 slpm, 70 F)

A flame in the shape of a little bowl, which looked much like the debris cloud of a tornado, attached to the nozzle, formed on the outside of the tornado cloud. This detail was lost in the process of converting from still photograph to Figure 5. Continued reduction of the fuel flow resulted in the growth in size and intensity of this cloud while simultaneously decreasing the funnel cloud. The disappearance of the funnel cloud marked a transition to a fully developed flame that was strongly attached to the center of the injector. At low air flow rates, this strongly attached flame would weaken until the lean blow out limit was reached. At high air flow rates, the flame would once again lift, attach to down stream recirculation zones, and eventually blow out from the separated flame structure, much as did the Task 100 Combustor, and the LS Task 150 Combustor.

One of the major factors in flame structure was fuel equivalence ratio. At a given air flow, the fuel would be reduced until a transition was judged to have been reached. These observations were not easily made. With the low swirl nozzle, most transitions between modes were quite abrupt, but the flame would oscillate from one mode to another (without obvious cause) over a narrow range of fuel equivalence ratios. In these cases, the practice was to wait until a complete (no jumping between modes) transition was reached. With the high swirl nozzle, however, the behavior was quite different. Instead of abrupt changes, the flame flowed smoothly from one mode to another. While eliminating the problem of flicker, these smooth transitions also left no sharp break point in flame behavior. Easily recognizable events, such as when the tornado appeared to touch the nozzle, were chosen to differentiate operating modes. Even still, events such as the onset of the funnel cloud were in large part a matter of judgment. With these understandings, Figures 6 and 7 quantify the different fuel equivalence ratio regimes in which the various modes operate.

Lean Blow Out Experiments

Lean blow out tests were conducted at various times during the course of the summer research effort. These measurements were conducted with the high swirl and low swirl nozzles, but were conducted with the 10 inch extension, and the 45% exhaust orifice plate installed. A more systematic set of measurements was made as part of the tests to measure wall pressure and temperature with various operating configurations, and over a range of both air and nitrogen flow rates. While making these tests, the burner was operated near LBO while the wall pressure and temperature measurements were made. Once this data had been collected, the burner would be further leaned until lean blow out was achieved and recorded. Frequently, the burner would be relighted, and the lean blow out measurement repeated.

It is beyond the scope of this document to report all of the lean blow out data taken during the course of the summer research program. The additional data is available on request. An example set of representative blow out data is contained in Figure 8 for both the low swirl and high swirl nozzles. For reasons not yet fully understood, the fuel equivalence ratios at lean blow out for air flow rates below about

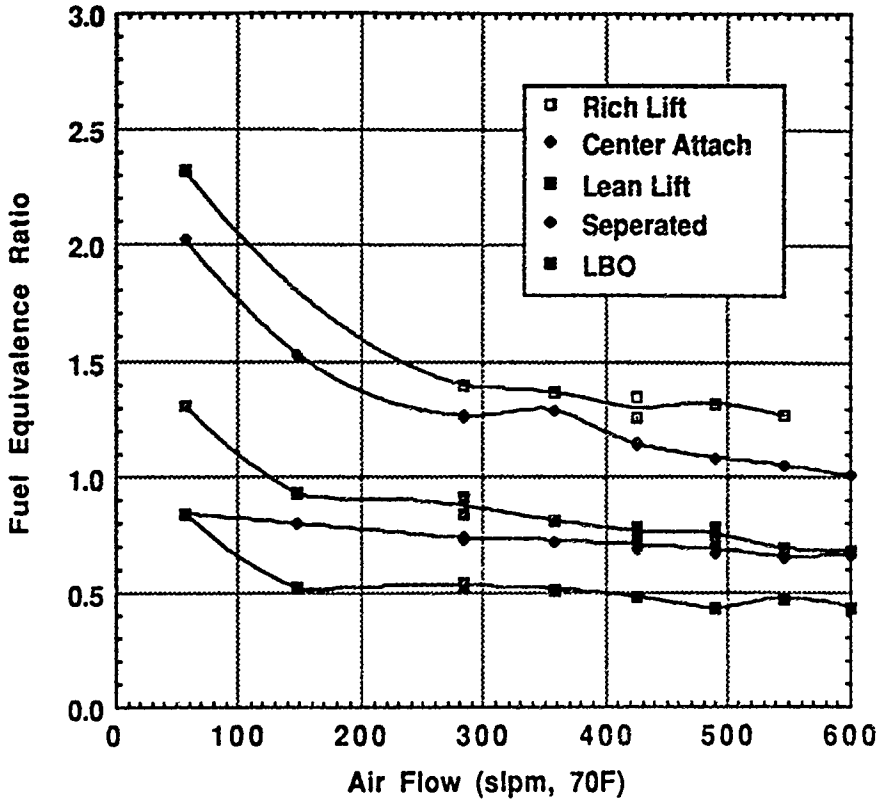


Figure 6 - Operating modes of the Task 150 Combustor with the Low Swirl Nozzle.

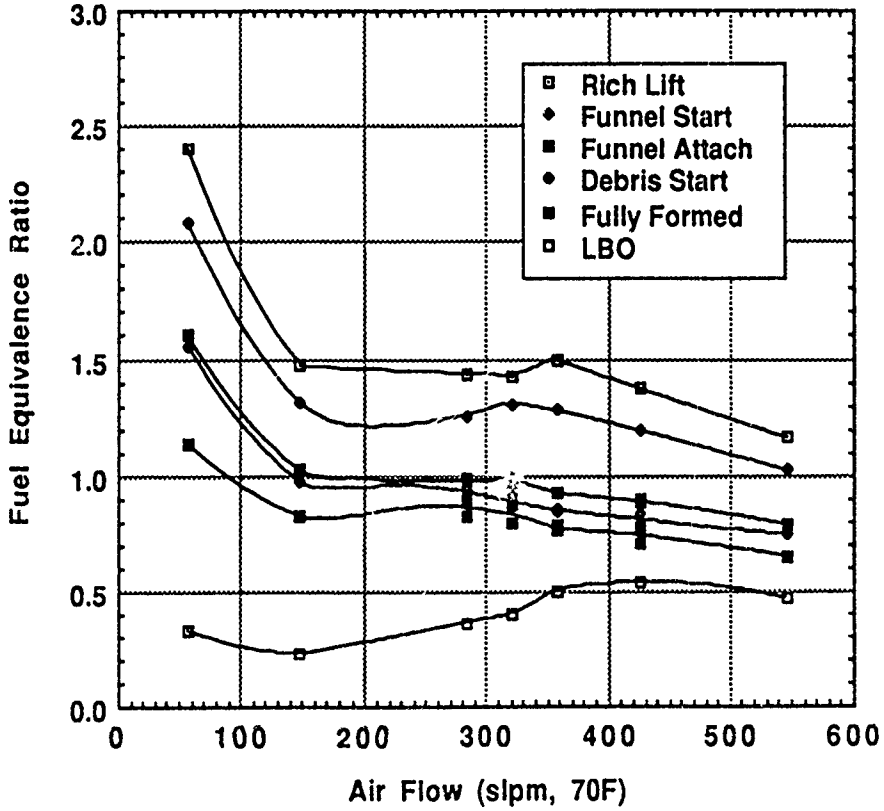


Figure 7 . Operating modes of the Task 150 Combustor with the High Swirl Nozzle.

400 slpm depart from the expected value of about 0.5. As Figure 8 shows, the ϕ at LBO for the low swirl fuel injector increases significantly as air flow rate is reduced below about 400 slpm (70 F). Conversely,

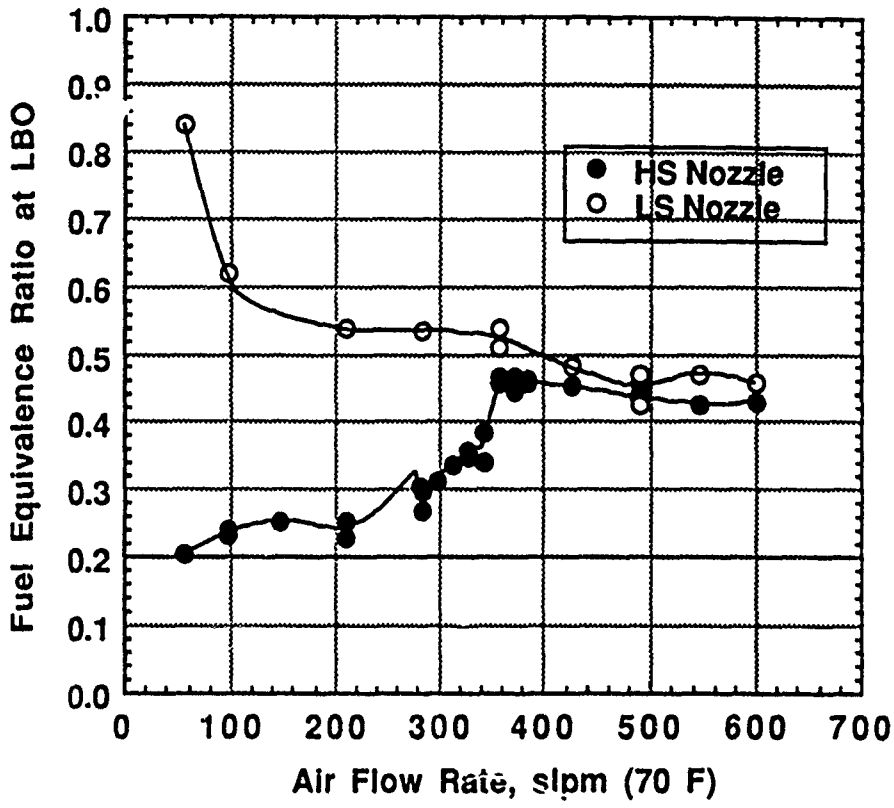


Figure 8 - Example Set of Lean Blow Out Data for Task 150 Combi (10 inch Extension, 45% Orifice)

the ϕ at LBO for the high swirl fuel injector decreases markedly as air flow rate is reduced below about 800 slpm (70 F).

Time has not permitted a complete evaluation of this lean blow out data, but a few observations have been made. At high air flow rates, the flame just prior to blow out for both injectors tested was a well lifted flame that was being stabilized on a downstream recirculation zone. At low air flow rates, the low swirl injector would still blow out from the flame stabilized on the down stream recirculation zone. However, at low air flow rates (less than ca 600 slpm), the flame with the high swirl nozzle would remain attached to the center core of the fuel injector right up to lean blow out. As a consequence, the fuel equivalence ratios (based on total fuel and total air flow) where lean blow out occurred with the high swirl nozzle were very much lower (ca 0.3) than that observed for the low swirl injector or the high swirl nozzle at high air flow rates (ca 0.5).

It was evident from observations of the flame associated with the high swirl burner at the lower air flow rates, that the volume of the flame zone was very small compared to the volume of the combustion

chamber, and that some of the combustion air was bypassing the combustion zone. Thus analysis of the flame in terms of a "loading parameter" must be done very carefully.

It is suspected that the flame zone is still behaving like a well stirred reactor, but that the local stoichiometry is not well represented by the overall fuel equivalence ratio, the total air mass flow, and the total volume of the reactor. It is difficult to understand how a flame can be sustained at a fuel equivalence ratio well below the lean flammability limit (f = 0.5). It is suspected that the f at lean blow out would be much closer to 0.5 if the volume of the flame zone could be accurately estimated, and if some estimate of the actual mass of air and fuel entering the flame zone could be made.

Flow Splits

Measurements of pressure drop across the various flow passages in the fuel injectors installed in the Pratt & Whitney Task 150 combustor were used to estimate the partitioning of the flows between the primary swirler passage, the secondary swirl passage, the insert jets, and the dome jets. The various passages were plugged in sequence so that the flow rate through a single passage could be correlated with the pressure drop across the dome of the combustor.

Each set of air flow, pressure drop data was curve correlated using a quadratic least squares function. It was assumed that the flow correlations for the insert and dome jets would be the same for both the HS nozzle and the LS nozzle sets. In making the measurements through the insert and dome jets, the flow through the combined insert and dome jets was correlated with pressure, and the separate flow through the insert jets was correlated with pressure. The separate flow through the dome jets was then obtained by subtracting the insert flow correlation from the combined correlation. The separate equations for each of the four passages for both the HS and LS injectors were then used to determine the flow rate as a function of pressure drop for each of the passages, and the sum was used to determine the correlation between the total air flow rate and the pressure drop across the injector. With the total air flow determined, the percentage of flow through each injector passage was determined as a function of air flow rate. Figures 9 and 10 summarize the percentage of flows through the primary swirler passage, the secondary swirler passage, and the combined insert and dome jets for the high and low swirl nozzles respectively. The higher pressure drop through the HS primary flow passage and the attendant lower flow rate is apparent.

A question on the effect of fuel injection on the pressure drop across the fuel injector, and its effect on the partitioning of the flow between the various passages was raised near the end of this summer's faculty and graduate student research program. There was insufficient time to fully explore this effect. Nevertheless, a brief investigation into this effect was undertaken. CO2 was introduced into the fuel orifice of both the HS and LS injectors. The insert and dome jets were completely blocked, since it was felt that fuel injection could not have any effect there. The flows in the primary swirler or the secondary

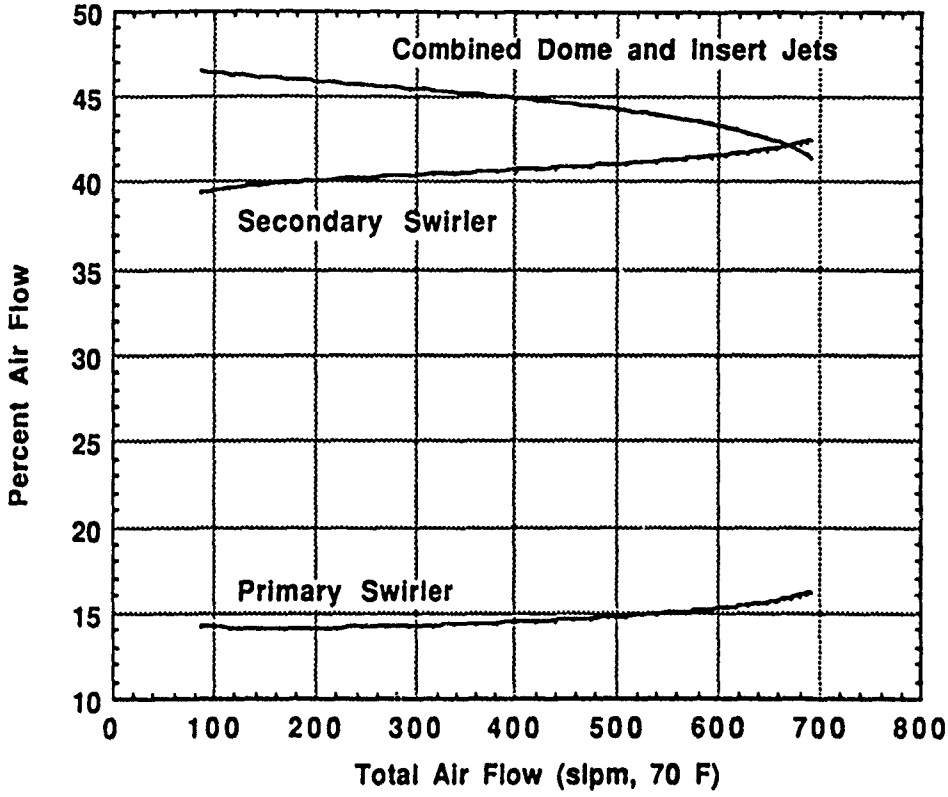


Figure 9 - Flow Splits for Task 150 Combustor with High Swirl Nozzle.

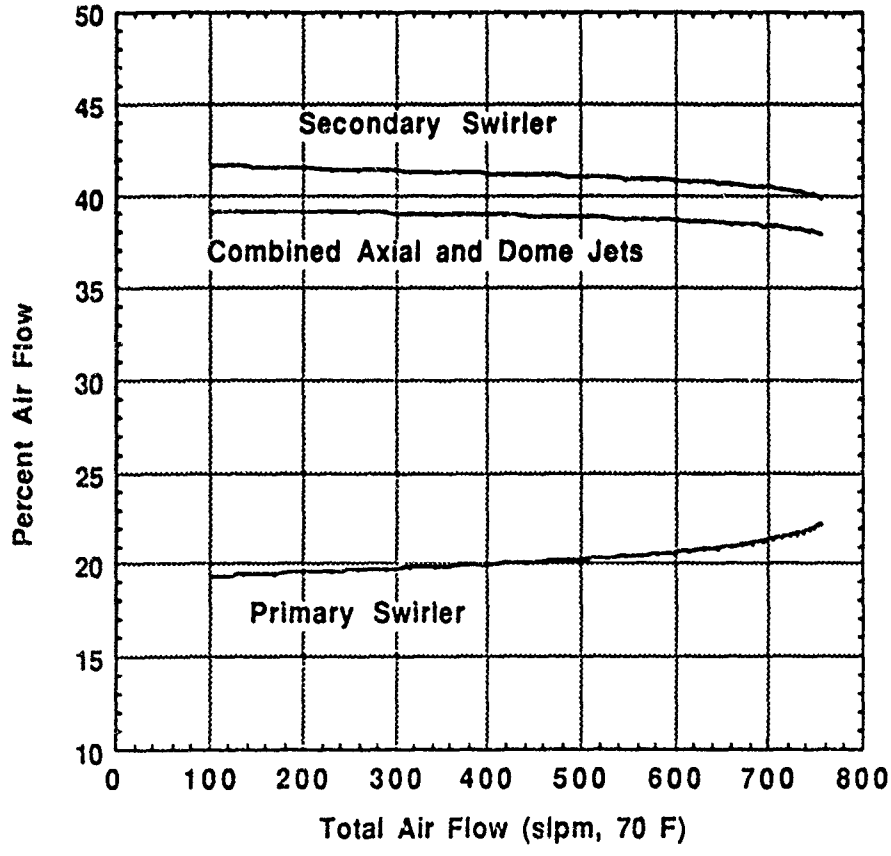


Figure 10 - Flow splits for Task 150 Combustor with Low Swirl Nozzle.

swirler were successively blocked so that the change in pressure across the injector could be monitored as a function of CO2 flow rate through the fuel orifice. Four separate conditions were used, two with the HS nozzle, and two with the LS nozzle. The air flow rate required to achieve a given pressure drop (LS, 1.67 and 4.35 in. H2O; and HS, 4.35 and 10.15 in H2O) was set and held constant while the CO2 flow rate was changed. The pressure change (referenced to the set pressure drop) as a function of CO2 flow rate for the four cases is summarized in Figures 11 and 12. It can be seen that the introduction of CO2 (fuel) has a very significant effect on the pressure drop across the passage, and by implication, on the flow rate through the passage for a given pressure drop. The effect is much more pronounced for the primary swirler passage that the secondary swirler passage. The effect can be either positive or negative depending on the injector (HS or LS) and air flow rate. That is, it can impede the flow causing an increase in the pressure drop, or it can act like an ejector causing a reduced pressure and consequent higher flow rate. Examination of the figures shows that the effect may be quite nonlinear and can vary from positive to negative back to positive depending on condition. It is evident the the addition of fuel will have a significant effect on the way that the flows split as they pass through the various passages of the injector. Further work in this area is warranted.

It should be noted that the percentage flow splits are thought to be quite accurate over the range of flow rates and delta pressure measured. This work needs to be extended to higher flow rates and higher delta pressures. Some of the actual combustion experiments conducted used air flow rates as high as 2500 slpm. This gave a delta pressure across the dome of the burner of about 4 psi. There were also some discrepancies when the total air flow rate as determined by the summation of the separate equations was compared to that measured directly with all passages open. There were some leaks in the facility that were identified at a later time that might have been affecting the results. Also, the flow through the very small gap between the injector and the dome insert may have been affecting the results. The introduction of CO2 to simulate the effect of fuel injection was found, as was discussed above, to be significant and very nonlinear. It is recommended that additional work be undertaken to further investigate the flow partitioning through the various passages over a wider range of flow and pressure conditions, and with additional study of the effect of fuel addition on the flow splits.

Wall Pressure and Wall Temperature Measurements

Wall temperature and pressure measurements were taken with thermocouples and pressure taps located in wall plates at selected axial locations. Measurements were made at or near the lean flammability limit and for selected cases near the maximum heat release. Test configurations included combinations of the HS and LS nozzles, 0 in and 10 in chamber extensions, and 0%, 45%, and 62% exhaust orifice. Much of the data were collected without any supplemental nitrogen being added to the combustion air stream. There were some sets where a low levels (ca. 10%) and high levels (ca. 20%) of N2 were added to the

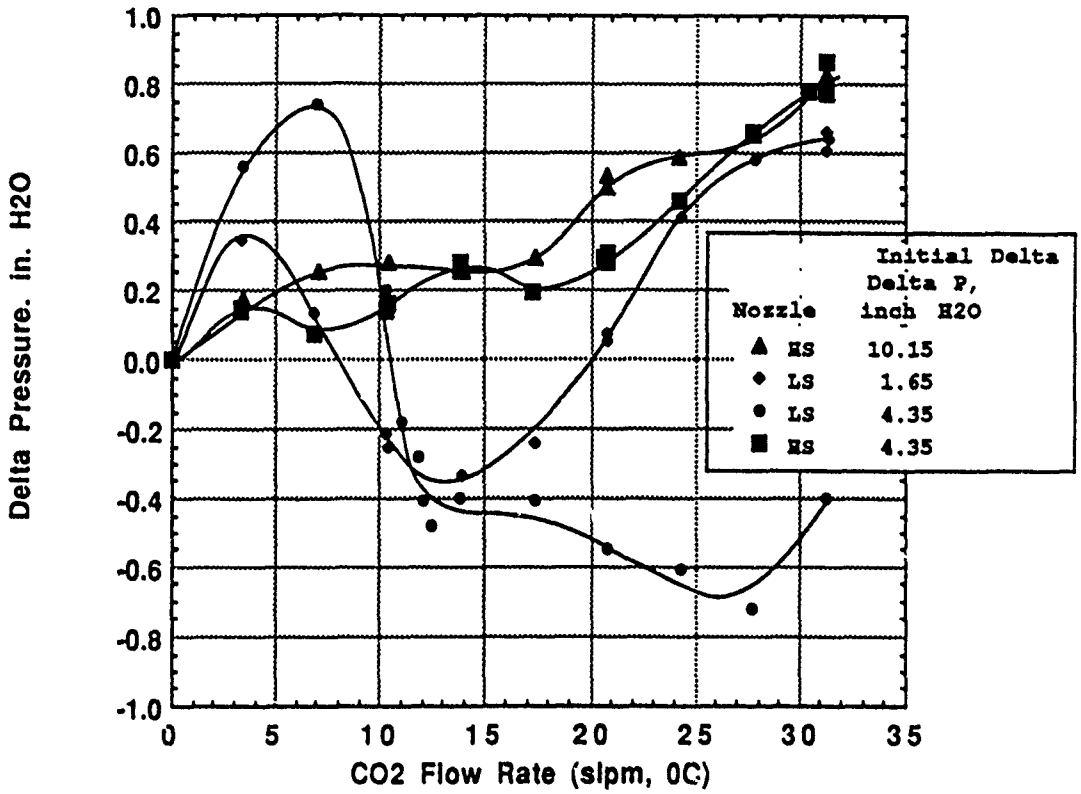


Figure 11 - Pressure Drop across Primary Swirler with Simulated Fuel (CO_2) Injection

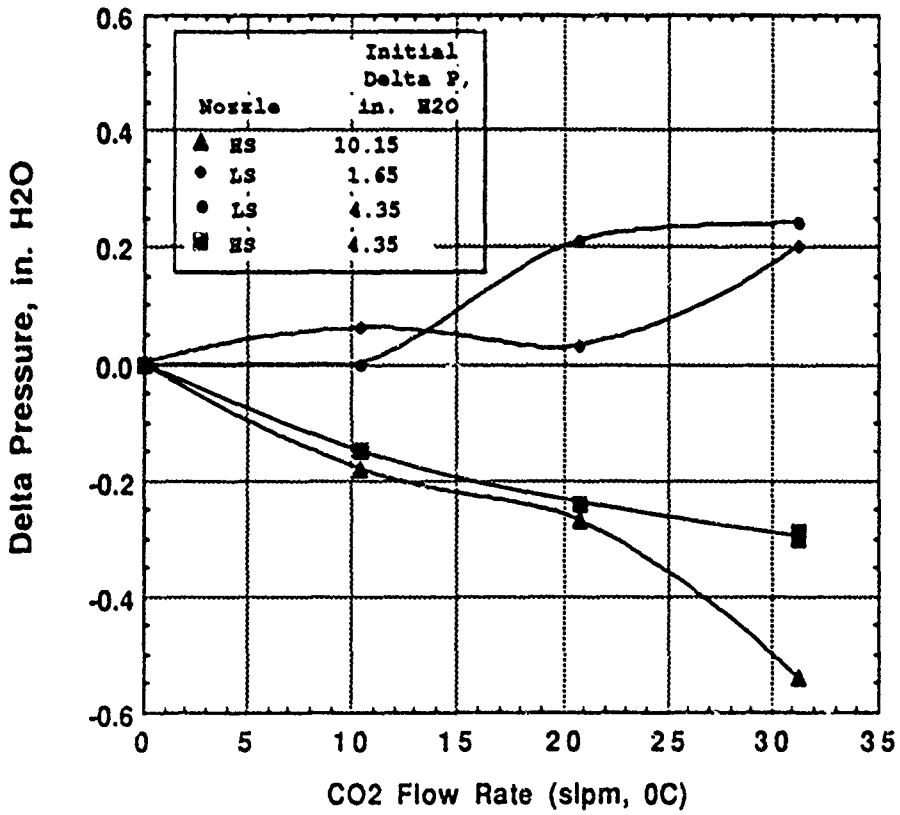


Figure 12 - Pressure Drop across Secondary Swirler with Simulated Fuel (CO_2) Injection

combustion air. In as much as time was limiting the latter test phases, A sparse test matrix was employed, i.e. only three levels of air flow, and zero or a high level of nitrogen were added to the combustion air.

With the burner exit open to the atmosphere, it was expected that the gage pressure at the wall would be indicative of the large-scale recirculation zone flow patterns. At points where the flow had a significant radial velocity, the pressure would change according to the stagnation pressure. In other regions, the pressure would be dominated by Bernoulli's equation--the higher the velocity, the lower the pressure. The most obvious result of the information collected was the invariability of the flow structure. This consistency can be see in Figure 13, which shows all the data points taken. There were three zones in evidence. The first recirculation zone, occurring in the first 50 mm, was dominated by the dome jets. The second zone, located between 50 and 200 mm, consisted of a lower recirculation zone caused by the flow in the primary combustion volume. The last zone was the combustion products flowing through the chamber. A few data sets also showed evidence of another recirculation zone near the exit, set up by the orifice blockage plate. This structure remained invariant for a wide range of test conditions. The main effect of the extension and orifice plate was to shift the curve up or down (Figure 13). The air flow changed only the magnitudes of the pressures , but did not appreciably change the location of the zones. The principle nozzle effect was on the strength of the first (dome) zone, with the high swirl nozzle creating a stronger flow due to the change in the flow splits.

Figure 14 contains an example plot of the Metal Temperature Factor (MTF) as a function of location. In general, the MTF data show a peak value at about the 200 mm location. However, the variation in location of the peak MTF does range from about 120 mm to about 250 mm depending on the flame structure which is being affected by the hardware configuration, air flow rate, and nitrogen flow rate. The location of the peak MTF probably represents a stagnation point associated with the major recirculation zone.

III. RECOMMENDATIONS:

Suggestions for Follow-on Research.

There have been considerable insights into the operational characterikstics gained form this summer's faculty and graduate student research program. There is much that yet needs to be done before a full understanding of the burner is achieved. There are several investigations that might be suggested to further these insights. Additional work needs to be done on the flow partitioning in the low and high swirl nozzles. The flow and pressure drop ranges need to be carried to higher values. Additional lean blow out measurements also need to be done, particularily at higher flow ranges. Work needs to be done to quantify the volume of a flame that is attached to the center of the burner and the amount of air and fuel that is actually entering the flame zone. This data is necessary inorder to better determine the actual fuel equivalence ratio in the "well mixed" flame zone. Laser sheet lighted images (either MIE scattering from a

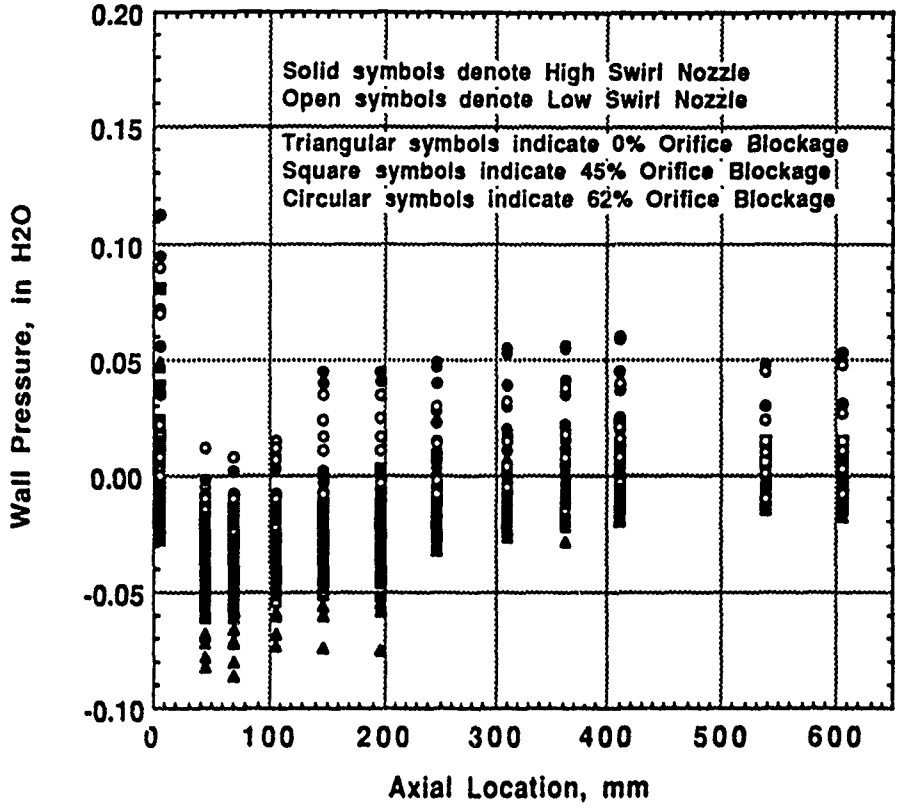


Figure 13 - Summary Graph Showing All Wall Pressure Data Taken.

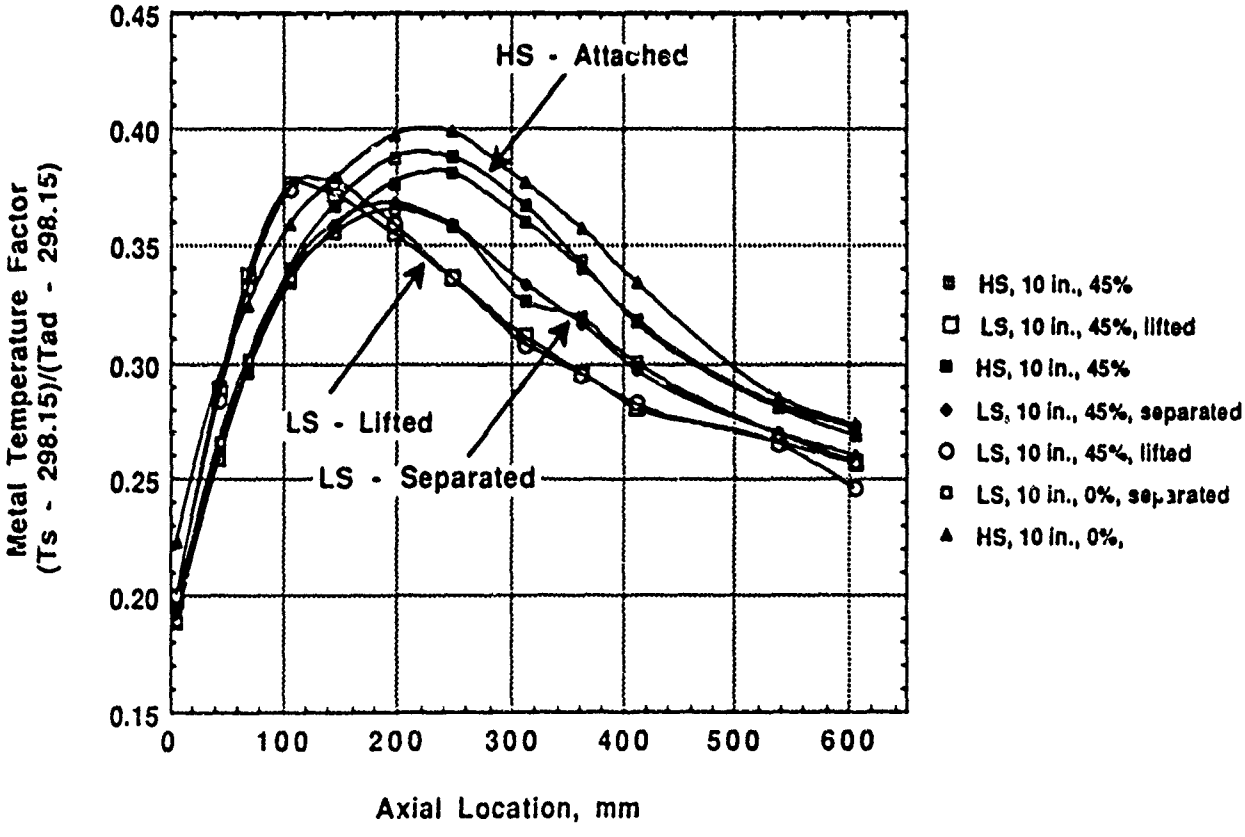


Figure 14 - Example of Metal Temperature Factors at 368 slpm (70 F) Air Flow

seeded flow, or LIF images of OH or NO) would be very useful to provide mappings of the flow fields. This information is essential in order to better understand where the flame fronts are located, and the characteristics of the eddies and other turbulent structures. Local gas temperature data and mixture fraction measurements would be very useful information for comparison to model predictions and similiar data from either the Task 100 or Task 200 Combustors. The temperature data could be obtained using a CARS diagnostic system. Gas mixture fraction is a more difficult diagnostic, but there may be other diagnostics that could be used to measure the concentration of a seed gas (e.g. CO2) such as degenerate four wave mixing. This area needs further investigation. The air flow rates to the primary and secondary swirlers were thought to have a major impact on the operation of the combustor. Much needed data on this characteristic could be obtained if the facility were modified so that the air flow rates to the various injector passages could be independently controlled.

Application for a Mini Research Initiation Grant.

An application for a mini research grant will be prepared as a result of the summer fellow research program. The proposed project would include the measurements on the Task 150 provided to BYU and installed under a research initiation grant during this past year. This burner is identical to the one tested in this summer at the Aero Propulsion and Power Laboratory, Wright Patterson AFB during this fellowship project. The burner incorporates the high swirl injector supplied by Pratt and Whitney Aircraft Co, East Hartford, Connecticut. The proposed study will include additional measurements of lean blow out, additional flame characterization measurments using film photographs and video images, CARS gas temperaturemeasurements, and sheet lighted film photographs from MIE scattering from a seeded flow. The Aero Propulsion and Power Laboratory, Wright Patterson AFB has companion programs, the Task 100 burner program, and the Task 200 burner program. The Task 100 burner was the burner investigated during the previous (1990) summer program. The Task 200 burner incorporates four swirling burners in a linear array. The proposed project would bridge the Task 100 and Task 200 programs and would provide much needed data that can not be obtained in any other burner. The proposed BYU project would fill a need to investigate the swirling burner in a single burner configuration. In addition, to basic combustion measurements, a future program at BYU could begin to investigate the formation of NO_x pollutants at the higher fuel equivalence ratios associated with higher performance gas turbines.

ACKNOWLEDGEMENTS

I wish to thank the Aero Propulsion and Power Laboratory, Wright Patterson Air Force Base, and the Air Force Office of Scientific Research for sponsorship of this research. The timely help of Mr. Gary Moore and other individuals at Research and Development Laboratories, Inc. is greatly appreciated. Special recognition of the help of Mr. Milton H. Danishik (RDL) is acknowledged. Not only did he provide

extraordinary help in providing information about living accommodations and base information, but he became a good friend as we shared the same office.

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

IV. REFERENCES

Lewis, B., and G. von Elbe, (1987), Combustion, Flames and Explosions of Gases, 3rd Edition, Academic Press, Inc.

Longwell, J.P., E.E. Frost, and M.A. Weiss, (August 1953), "Flame Stability in Bluff Body Recirculation Zones", Industrial and Engineering Chemistry, 45(8), pp 1629-1633

Longwell, J.P., and M.A. Weiss, (August 1955), "High Temperature Reaction Rates in Hydrocarbon Combustion", Industrial and Engineering Chemistry, 47(8), pp 1634-1643

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

CRITICAL ASSESSMENT OF RESEARCH IN ELASTOHYDRODYNAMIC LUBRICATION

by

Michael M. Khonsari
Department of Mechanical Engineering
University of Pittsburgh
Pittsburgh, PA 15261

Report Submitted To:

The United State Air Force

SUMMARY

Presented in this report is a portion of the work performed by the author as a Visiting Summer Faculty in the Lubrication Branch at Wright Patterson Air Force Base.

Two major tasks both dealing with the subject of elastohydrodynamic lubrication (EHL) of bearings were undertaken. First, a critical assessment of published research in EHL was made with particular attention to those sponsored by the United States Air Force. To this end, a number of important key subjects were investigated and critical problem areas were identified. This report provides a brief discussion of the effort.

The second portion of the work dealt with developing a computer program which can accurately calculate the pressure distribution and film thickness in concentrated contacts. Particular attention was given to the execution time and the efficiency of the algorithm as it is well known that EHL calculations are very delicate and time-consuming.

Due to the page limitation imposed on this report (maximum: 20 pages), the derivation of equations and the working of the program is not discussed herein. This program is to be used as a research tool. It is anticipated that it will be expanded upon in an applied research which is to be proposed as follow-up work.

INTRODUCTION

The archives of published research in tribology contain large volumes of work devoted to the subject of elastohydrodynamic lubrication (EHL). Extensive work in EHL began in the late 1960's whereas, in contrast, research in the field of hydrodynamic lubrication (HL) was well underway at that time.

Active research in EHL began when it was realized that the theories of hydrodynamic lubrication failed to provide a realistic assessment of the fluid film thickness in concentrated contacts. For such applications Dowson & Higginson [1] showed quantitatively that the film thickness, as predicted by one-dimensional HL theory, was much less than what would be required to maintain separation between solids. In fact, the film thickness was found to be even smaller than the surface roughness. Thus a refinement to the classical HL theory, was found to be essential for concentrated contact problems.

Initially, attempts were made by relaxing the basic assumptions in the HL theory, one by one. These included the consideration of thermal effects, side-leakage, non-Newtonian aspects of the lubricating fluid and pressure-dependent lubricant properties.

The relaxation of thermal effects and side-leakage both yielded a still lower film thickness than the isothermal solutions in one-dimension and therefore had an adverse effect on the film thickness predictions. The consideration of non-Newtonian and pressure-dependent viscosity effects were somewhat more promising in the sense that they were thought to have the right characteristics in order to account for an enhancement in the lubricating film thickness.

To account for non-Newtonian characteristics, one must start by redefining an appropriate stress-strain relationship for the fluid that may, in fact, deviate from the linearly viscous Newtonian equation. It was thought that time-dependence, or transient effects, as characterized by visco-elasticity effects in lubricant could be responsible for altering the film thickness. The so-called Maxwell fluid model was a first candidate for the non-Newtonian fluid model. Various other early attempts were made to look into the problem, but no noticeable

improvement in the film thickness was found. In particular, Burton's results for a rolling contact [2] yielded a reduction in the load-carrying capacity, and hence a lower film thickness.

Property variation due to high pressures developed in concentrated contacts also required careful consideration. Both the density and viscosity (particularly the latter) increase remarkably with pressure. But even the inclusion of such variations in hydrodynamic theories did not yield film thickness values much higher than the surface roughness. Therefore, it was concluded that the hydrodynamic lubrication theory, by itself, fails to fully describe the lubrication characteristics of concentrated contacts.

The key element which gave birth to the elastohydrodynamic theory is the elastic deformation of the solids. This was precisely the missing link needed to augment with the fluid mechanics of lubrication in order to make realistic predictions of the film thickness as well as many of the other bearing performance parameters.

Since the discovery of the EHL regime, significant progress has been made toward the understanding of the lubrication characteristics of both line and elliptical contact problems. However, film thickness remains as the key parameter to which successful operation of bearings depend upon.

While the mechanics of deformation are now well understood, interestingly enough the hydrodynamic components of the problem, particularly those of thermal effects and the non-Newtonian behavior of lubricant, require further research.

The purpose of this report is to provide insight to the research that has been conducted and published in the form of reports and papers particularly those which were sponsored by the United States Air Force. The report will also address some of the important problem areas that warrant future investigation.

THERMAL EFFECTS

The results of a major effort on EHL Lubrication was reported by Mechanical Technology Incorporated in 1970 [3]. The report contains an impressive volume of technical information

including an extensive reference list. Some of the work described in that report and publications that resulted from it are extensively used in sophisticated analysis packages even today.

Perhaps the most important portion of the work by the MTI group is the modeling of thermal effects. There are several tacit assumptions and restrictions on their formulation which one should keep in mind before utilizing the results.

The thermal analysis is based upon the Grubin-type formulation for the film thickness in a line-contact configuration. The Grubin formulation does **not** solve the elasticity and the Reynolds equation for the entire contact region. Rather, it determines the pressure generated in the inlet region which would be required to separate the bounding surfaces within the Hertzian contact zone. Therefore, it is assumed that a significant amount of pressure is generated in the inlet region and that the deformation of the solids in the contact region is identical to the dry case, as given by Hertz's classical solution.

Grubin arrived at the following film thickness equation for the line-contact problem (cf. ref. [1])

$$H = 1.95(GU)^{0.73} W^{-0.091} \qquad (1)$$

where

$H = h/R$ (dimensionless film thickness)

$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2}$ (effective radius)

$G = \alpha E$

$\frac{1}{E} = \frac{1}{2}\left[\frac{1-\upsilon_1^2}{E_1} + \frac{1-\upsilon_2^2}{E_2}\right]$ (effective modulus of elasticity)

$W = \frac{w}{ER}$ (dimensionless load)

Grubin's analysis is, of course, restricted to the assumption that the lubricant remains isothermal throughout the film. According to Cheng [3], when the surface speed of the rolling element, made of steel, exceeds 250 in/s (≈ 1 m/s), thermal effects become significant. The reduction in film thickness at high speeds has been observed experimentally by researchers at Battelle Memorial Institute in a major contract work sponsored by the United States Air Force.

Cheng made a significant contribution to the EHL theory by solving the energy equation with the Grubin-type film thickness analysis formulation [4]. However, in addition to this approximation, the compressibility of the lubricant was neglected. Consideration of compressibility is known to lead to a reduction in the magnitude of the pressure spike.

Cheng introduced a new parameter called, Q_m, as the most influential factor in thermal analysis. This parameter essentially relates the heat generation in the lubricant to the heat conduction across the film. It is given as defined below

$$Q_m = \frac{\mu_0 (U_1 + U_2)^2}{2 K_f T_0}$$

where μ_0 and T_0 are the viscosity and temperature under the ambient conditions and K_f is the fluid thermal conductivity. The validity of representing the entire heat generation by a single parameter must be examined. This parameter, Q_m, contains three other parameters used for categorizing the lubricant properties: μ_0, T_0, and K_f. It also includes the effect of surface velocities. (It would seem that the factor 2 in the denominator should have been replaced by 4, however, that would merely scale the Q_m factor). A "heat generation parameter" as such naturally arises when one non-dimensionalizes the energy equation and examines the dominant parameters. In a recent study which dealt with hydrodynamic journal bearings, this procedure led to the derivation of a similar factor which we called temperature-rise parameter, κ. It was

shown that κ has most of the pertinent bearing parameters needed for prediction of the maximum surface temperature [5].

Cheng [4] performed extensive simulations for two types of oils. Polyphenyl ether and a super refined mineral oil. In order to make a comparison with isothermal results, he used Dowson-Higginson's film thickness equation for line-contact as a benchmark. This equation is given below

$$H_{isothermal} = 1.6G^{.6}U^{.7}W^{-.13}. \qquad (2)$$

(It is to be noted that this equation actually gives the minimum film thickness).

Cheng proceeded by defining a so-called "thermal reduction factor, φ_T parameter as

$$\varphi_T = \frac{H}{H_{isothermal}} \qquad 0 < \varphi_T < 1$$

where H is Cheng's results which account for thermal effects.

The parameter Q_m lacks an important property which is known to play an important role in thermal effects. This has to do with the relative difference between the surface velocity of the solids or the so-called slip ratio parameter, $S = \frac{U_2 - U_1}{U_2}$. It probably would have been appropriate to consider the product of S by Q_m as the thermal generation factor. However, Cheng chose to examine this parameter separately through extensive simulations for the range of S = 0. - 0.4. Later in reference [3], the results were curve-fitted and presented in the form of a so-called thermal reduction factor given below:

$$\varphi_T = f_1(1 - 0.1s)\left(1 + f_3\frac{P_{HZ}}{E}\right)$$

where f_1 and f_3 are functions of Q_m, $\alpha^* = \alpha \frac{R}{2} 10^5$ and $\beta^1 = \frac{\beta}{T_0}$. The values for these functions are given in Table XIII in reference [3]. The parameter $\frac{P_{HZ}}{E}$ is the dimensionless Hertzian pressure defined as

$$\frac{P_{HZ}}{E} = \frac{1}{\pi}\left(\frac{W}{ER}\right)^{0.5} \qquad \textit{for line contact}$$

and

$$\frac{P_{HZ}}{E} = \frac{3}{2}\frac{P}{\pi abE} \qquad \textit{for elliptical contact (with } R = R_x\textit{)}.$$

It should be noted that when applying the theory to elliptical contacts one must keep in mind that the theory was developed for one-dimensional line-contact and therefore a side leakage factor φ_L must be introduced so that

$$H_{min} = (H_{min})_{isothermal}\ \varphi_T \varphi_L \qquad (4)$$

There are many other forms of thermal reduction factors. For example in a more recent article [6] Cheng presents Figure 1 for evaluating the thermal reduction factor. Still another form of φ is given below [7]:

$$\varphi_T = \frac{1 - 13.2\left(\frac{P_{max}}{E}\right)L^{0.42}}{1 + 0.213\left(1 + 2.23S^{0.83}\right)L^{0.640}} \qquad (5)$$

where S = slide to roll ratio = $\frac{(U_2 - U_1)}{(U_1 + U_2)/2}$;

$$L = \frac{\beta \mu_0 U^2}{K_f}, \qquad U = \frac{U_1 + U_2}{2};$$

$$P_{max} = \frac{3F}{2\pi ab}, \quad F = normal\ load.$$

Gupta [8] compares the experimental results of Wilson [9] to those predicted using equation (5) and reports good agreement between the two. This form of φ_T is very efficient for implementation on the computer. Furthermore, it appears to be well suited for use under high rolling speeds, nowever, further investigation is needed to verify its appropriateness for high slide-to-roll ratios.

Gupta [8] uses equation (5) for elliptical contacts although the original development was based on line contacts. His argument is that the contact ellipticity ratio has little influence on φ_T.

Surprisingly, little comparison between independent film thickness measurements and theoretical predictions have been reported in the literature. Coy and Zaretsky [10] expressed their concern on this matter particularly for high speed bearings. They conducted their own experimental measurements and compared their results with the film thickness predicted by formulas due to Grubin, Hamrock & Dowson, along with Cheng's results with inlet shear heating. Reproduced from [10] in Figure 2, experimental film thickness values are shown as a function of the "contact lubrication flow parameter, GU". It appears that the theoretical results deviate from experimental measurements for higher speed cases and that Cheng's results offer a slight improvement to those predicted by Hamrock & Dowson [11]. The experimental results exhibit a much stronger non-linear trend with speed than do the theoretical simulations. Not mentioned in the paper, or in discussions that followed it, is the non-Newtonian aspect of the problem neglect ⁴ in theoretical formulation. The stress-strain relationship for most non-Newtonian fluids exhibit a deviation from linearity. This could be, in part, responsible for the "flattening" behavior of the experimental results for high speeds.

At this point, a number of important remarks can be made. First, the film thickness equations derived by Hamrock & Dowson, and the Grubin's formula, are all based on regression analyses which includes distinct limitations on load, speed and material properties. More importantly, they are derived based upon the assumption of Newtonian fluid. Second, Cheng's [4] thermal reduction factor has several limitations: (a) it is based on the line-contact solution with Grubin-type formulation, and (b) it is restricted to a certain upper limit on the slip-ratio. Given that the non-Newtonian characteristic of fluids can play a crucial role in EHL, a **consistent** analysis is needed to describe the film thickness and thermal reduction factor from the first principles. The analysis must include **lubricant** and **surface** property variation with temperature. Third, the flattening of the film thickness at high speed deserves attention from the point of view of "stabilization of operation". That is to say, since the film thickness does not vary much with speed, a smoother behavior is expected. Thus physically, when the speed increases, the EHL pressure distribution approaches to that of a hydrodynamic bearing, as if the load gets distributed over a larger area. This may be considered to be a beneficial affect. Lastly, it will be re-emphasized that at high operating speeds the behavior of bearings may be dominated by hydrodynamic effects, hence, signifying the importance of the rheology of the lubricant. Experimental results for film thickness and traction attest to this fact.

NON-NEWTONIAN FLUIDS

There is an abundance of non-Newtonian fluid models in the literature. For EHL applications, however, it appears that an appropriate rheological model is one which characterizes the lubricant by a limiting shear stress, τ_0. Initially the shear stress increases linearly with strain at a given slope. However after a certain point, further increase in deformation of the fluid does not appreciably influence the shear stress. This is generally referred to as the limiting shear stress.

While the classical Reynolds equation is appropriate for the linear region of the stress-strain curve, it must be rederived for the second portion of the curve where it begins to exhibit a non-linear variation.

There is evidence to support the hypothesis that for certain values of the limiting shear stress (4-6 MPa), the film thickness may be considerably less than what the Newtonian fluid model would predict [12]. The values of the limiting shear stress for the viscoelastic fluid of type II considered by Gupta [8] appears to be within this critical range.

For military applications that use MIL-L-7808 Gupta [8, 13] and Forster et al. [14] considered two variations of viscoelastic-type fluid. The constitutive equations are shown below for completeness. In dimensionless form, following Gupta's notation [8]:

$$\frac{\partial \bar{\tau}}{\partial y} = S = \frac{1}{D} f(\bar{\tau}) \tag{6}$$

where

$$S = \frac{U_s G b}{h U \tau_0} \quad and \quad D = \frac{\mu U}{G b}.$$

The functional form of $f(\bar{\tau})$, which is of primary interest to us, is classified as type I and II:

$$Type\ I: \quad f(\bar{\tau}) = sinh(\bar{\tau})\ (based\ on\ Johnson\ \&\ Tevaarwerk\ [15])$$

$$Type\ II: \quad f(\bar{\tau}) = tanh^{-1}(\bar{\tau})\ (based\ on\ Bair\ \&\ Winer\ [16]).$$

These two very similar models essentially curve-fit experimental data. The series expansion of these two functions show why they are similar:

$$sinh(\bar{\tau}) = \bar{\tau} + \frac{1}{3!}\bar{\tau}^3 + \frac{1}{5!}\bar{\tau}^5 + \ldots \tag{7}$$

$$\tanh^{-1}(\bar{\tau}) = \bar{\tau} - \frac{1}{3}\bar{\tau}^3 + \frac{2}{15}\bar{\tau}^5 - \ldots \tag{8}$$

The denominator of terms in equation (7) grow much more rapidly than their counterpart in (8). Therefore, the higher order terms in type I became negligible much faster than that of type II. Both of these models under limiting cases approach the classical Maxwell's equation for a viscoelastic fluid:

$$\dot{s} = \frac{1}{G}\frac{\partial \tau}{\partial t} + \frac{\tau_0}{\mu}. \tag{9}$$

In Gupta's [8] notation:

$$\frac{U_s}{h} = \frac{1}{G}U\frac{\partial}{\partial y}(\tau_0\bar{\tau}) + \frac{\tau_0}{\mu}$$

$$\frac{\partial \bar{\tau}}{\partial y} = \frac{U_s}{h}\frac{G_b}{U\tau_0} - \frac{G_b}{U\mu}$$

or

$$\frac{\partial \bar{\tau}}{\partial y} = S - \frac{1}{D}, \tag{10}$$

i.e.,

$$f(\bar{\tau}) = 1.$$

The key behind the understanding the relevance and applicability of these viscoelastic models is the relative comparison between the relaxation time of the fluid and the time it takes for the fluid to go through the contact. This stems from the fact that, in general, most liquids when put under very rapid stress exhibit elastic behavior, more like an elastic body. Of course, viscous properties are still present and therefore the fluid is said to be viscoelastic.

In EHL applications since shear rates are very large and the time spent by the fluid in the contact is short, the viscoelasticity becomes an important consideration. A dimensionless

parameter known as the Deborah number defined as the ratio of the relaxation time to the characteristic process time is a measure of the importance of viscoelasticity. The higher the Deborah number the more important is the viscoelasticity effect.

Before bringing this section to an end, it will be pointed out that Cheng et al. also considered the non-Newtonian behavior of fluids with a formulation similar to the equations shown above [17]. They introduced a new constitutive equation for the lubricant written below. In Gupta's [8] notation

$$\tau = \frac{\mu \dot{s}}{1 + \mu \frac{\dot{s}}{\tau_0}} \tag{11}$$

where

$$\dot{s} = strain\ rate = \frac{\partial u}{\partial y}.$$

The analysis presented in [17] compared favorably with various experimental measurements. It is surprising, however, that no attempt was made by the author to come up with a thermal reduction factor based on this work.

STARVATION EFFECTS

Another factor that could play a significant role in the determination of the film thickness is that of starvation effects. The importance of such considerations in theoretical studies could be crucial as one may see by examining Figure 2 As is evident, the theoretical predictions of Chiu which include starvation effects, show a reasonable trend for the variation of film thickness as a function of the flow parameter. In these experiments, the flow rates were very small (2-3 drops/min), and starvation is very likely.

Hamrock [18] treated the problem as a new inlet boundary condition and developed appropriate expressions for lubricant film thickness under the starved conditions. This he did by revising his computer program for elliptical contacts. He let the distance from the inlet edge

to the center of the Hertzian contact be denoted by $\overline{m} = \frac{b}{a}$. Starting from the fully flooded case, Hamrock gradually reduced $\overline{m}$ and monitored the variation of H_{min}. The values of $\overline{m}$ at which H_{min} shows a reduction (when compared to the finely flooded case) is called the starved inlet condition, labeled m*.

Regression analyses performed by Hamrock led to the following results for the hard EHL problems. For the central film thickness

$$H_{c,s} = H_c\left(\frac{\overline{m}-1}{m^*-1}\right)^{0.29} \tag{12}$$

where

$$m^* = 1 + 3.09\left[\left(\frac{R_x}{b}\right)^2 H_c\right]^{0.58}$$

For the minimum film thickness

$$H_{min,s} = H_{min}\left(\frac{\overline{m}-1}{m^*-1}\right)^{0.25}$$

where

$$m^* = 1 + 3.34\left[\left(\frac{R_x}{a}\right)^2 H_{min}\right]^{0.56}.$$

To implement, one would evaluate m* from (12) or (13) and compare to $\overline{m}$. If $m^* < \overline{m}$, then fully flooded results must be modified by the factors given above.

The fully flooded central film thickness was corrected in this fashion by McCool [19] in a contract work for the United States Air Force.

Although the validity of Hamrock & Dowson's fully flooded film thickness equations have been verified by Kaye & Winer [20], to the best of the author's knowledge, the starvation correction factors derived in [18] have not been experimentally investigated. Nevertheless,

extensive measurements by Wedeven [21] compared to his own theoretical correction factor yielded good agreement. Interestingly, Wedeven's formula is very similar to that of Hamrock [18]. Therefore, Hamrock's starvation correction factor can be considered to be accurate, as far as the central and minimum film thickness calculations are concerned.

There are, of course, other theoretical models available for implementation of starvation effects. One such model is described in the SKF Bearing Design Manual [22] supposedly developed Y.P. Chiu. In this model the film thickness and m* are related to the upstream (ambient) film layer thickness, $h_{1,1}$ and $h_{1,2}$ see Figure 3. Although no direct mention of experimental verification is made, the formulation appears to be based on good physical insight and vast experience in this field. The adherence of fluid to the surfaces at the inlet and film replenishment model for a full bearing are very worthwhile for further research. Surprisingly very little published research is reported on this subject in the open literature.

At this point it is worthwhile to consider the effect of starvation on parameters other than the film thickness. The classical Hertzian analysis pertains to the dry contact. The Hertzian pressure distribution has a normal parabolic shape. In EHL applications when a layer of lubricant is present, however, a pressure spike appears toward the exit. One would, therefore, expect that starving the film at the inlet should, in limit, approach the Hertzian pressure distribution by suppressing the pressure spike. This is precisely what Hamrock's numerical solutions show.

It would be interesting to examine the phenomenon using optical film thickness contours experimentally particularly, the traction behavior of starved film under high speed and high shear conditions is much worth further research. Wedeven's measurements [21] show that a starved film leads to a higher traction than the fully flooded contact unless the shear rate -- defined as $\frac{U_1 - U_2}{h_c}$ is very low.

At low shear rates, the film is considered to be severely starved and the traction results turn out to be surprisingly lower than expected. This should be further investigated particularly with close attention to thermal effects and non-Newtonian characteristics of the fluid. To the author's best knowledge, there are no theoretical analyses available that take both thermal effects and starvation into consideration. The widely used thermal analysis performed by Cheng [4] and the thermal reduction factor φ_T derived from it appear to be based on the fully flooded case.

Described by Cheng [4] is the formation of backflow in the velocity profile at the inlet zone. According to Wedeven [21] backflow is essential for the establishment of a flooded condition. As with the non-Newtonian characteristics of the lubricant, the starvation may very well affect the φ_T calculations. The validity of representing separating these interrelated elements by simple correction factors needs to be closely scrutinized.

ROUGHNESS EFFECTS

As all engineering surfaces are rough on a microscopic scale, one must investigate the influence of surface characteristics on EHL performance, particularly for partial EHL. Unlike the EHL theories that are well-accepted, there is no universally accepted model for surface roughness. Furthermore, there appears to be a lack of experimentally verified analyses in this area. Also, most of the papers published in this field necessarily resort to the principles of statistics. In doing so, they require a number of input variables which are not readily available or at least are very hard to relate to the laboratory measurements.

Recently McCool [23, 24] made an outstanding contribution to this field which deserves attention. McCool developed a method to relate the output of profile measurement equipment to the input needed for the Greenwood-Williamson's microcontact model [25]. Two main parameters are needed: R_q and Δq. R_q is the rms deviation of the surface height from its mean value and Δq is a slope sensitive parameter. These two parameters allow one to calculate the

mean square curvature parameter which is needed in the Greenwood-Williamson's model. The model can be applied to both isotropic and anisotropic surface roughness [23]. Furthermore, McCool [24] showed that it is possible to estimate the value of Δq from the R_q measurements. In this fashion only the height sensitive parameter, R_q, needs to be measured.

McCool incorporated his roughness model in a theoretical study to evaluate the traction behavior of several ball bearings with various lubricants [19]. It appears, however, that roughness consideration was not one of the major concerns in that study since no parametric study was presented.

Now that a practical model for surface roughness is available, we are in a position to shed some light on the relationship between the fatigue life and surface roughness in the partial EHL regime. A recent contribution by Averbach & Bamberger [26] showed that under partial EHL conditions, failure occurs right at the surface rather than at the sub-surface, where it was generally thought to start. This finding is based on an extensive examination of bearings returned from the field. This has significant practical implications since as Averbach & Bamberger pointed-out most aircraft gas turbine bearings operate in partial EHD mode even under ideal conditions.

The initiation of fatigue depends very strongly on the specific film thickness defined as

$\Lambda = \frac{h_{min}}{\sqrt{\sigma^2 + \sigma^2}}$ where σ_1 and σ_2 are the rms surface roughness of the interacting surfaces. It is believed that for low Λ values ($\Lambda < 1$) the fatigue initiates at the surface while for high values of Λ ($\Lambda > 5$), the initiation of fatigue occurs below the surface [27].

It would be valuable to conduct a co mplete theoretical study of the surface characteristics which includes asperity interaction, thermal effects, and surface/subsurface stresses. To determine the stress field in the solid, one may have to develop a computer program that determines the EHL pressure distribution. Therefore, it may not be possible to

directly use McCool's program [19], since he relies on Hamrock's equation for film thickness and does not compute the pressure distribution.

For practical purposes, there are two parameters which should be related to bearing life: one is the surface characteristic and the other is the lubricant properties, i.e., the viscosity. A high viscosity lubricant would, of course, result in a thicker oil, but this has not been quantitatively translated to the bearing life enhancement. Certainly there exists a trade-off between excessive power loss, starting torque and the bearing life.

To meet the stringent requirements of Integrated High Performance Turbine Engine Technology (IHPTET), in addition to the lubricant properties, the surface properties and its characteristics should be included in theoretical analyses of ball bearings.

Quantitative assessment of bearing life based on the lubricant viscosity, surface characteristics, and operating conditions requires a thorough study of the elastohydrodynamic lubrication. Significant savings can result through improvement of performance, reliability and maintenance schedule.

ACKNOWLEDGEMENT

The author wishes to thank Nelson Forster of the Lubrication Branch of the Wright Patterson Air Force Base for making available many technical reports used in this study. I also wish to thank Ron Dayton, the Lubrication Branch Chief, for his encouragement during the course of this study.

REFERENCES

[1] Dowson, D. and Higginson, G. *Elastohydrodynamic Lubrication*, Pergamon Press, 1977.

[2] Burton, R. "An Analytical Investigation of Viscoelastic Effects in the Lubrication of Rolling Contact," *ASLE*, V. 3, N.1, p. 1, 1960.

[3] McGrew, J., Gu, A. Cheng, H., Murray. *Elastohydrodynamic Lubrication*, Technical Report AFAPL-TR-70-27, 1970.

[4] Cheng, H. "Calculation of EHD Film Thickness in High Speed Rolling and Sliding Contacts," MTI-67TR24, 1967.

[5] Khonsari, M. and Wang, S. "On the Maximum Temperature in Double-Layered Journal Bearings," *ASME J. of Tribology*, To appear 1991.

[6] Cheng, H. "EHD Lubrication." *CRC Handbook of Lubrication*, V.II, p. 139.

[7] Rosado, L. "Analysis of an Elastohydrodynamic Lubricated Point Contact," WRDC-TR-89-2106, 1989.

[8] Gupta, P. "Traction Modeling of Military Lubricants," WRDC-TR-89-2064, 1989.

[9] Wilson, A. "An Experimental Thermal Correction for Predicted Oil Film Thickness on EHD Contacts," Proc. Leeds-Lyon Sumposium, p. 179, 1979.

[10] Coy, J and Zaretsky, E. "Some Limitation in Applying Classical EHD Film Thickness Formulas to a High-Speed Bearing," *ASME J. of Tribology*, Vol. 103, p. 295, 1981.

[11] Hamrock, B. and Dowson, D., *Ball Bearing Lubrication*, Wiley, 1981.

[12] Bakashvivili, D., Chlehaidze, V., Shvartsman, V., and Shockhet, K. "Influence of Lubricant Limit Shear Stress on EHD Contact Film Thickness," Trenie i Iznos, V. 5, N. 2, pp. 295-304, 1984.

[13] Gupta, P., Cheng, H., Zhu, D., Forster, N., and Schrand, J. "Visco-Elastic Effects in MIL-L-7808 Type Lubricant, Part I: Analytical Formulation," To appear in *STLE Tribology Transactions*.

[14] Forster, N., Schrand, J. and Gupta, P. "Visco-Elastic Effects in MIL-L-7808 Type Lubricant, Part II: Experimental Data Correlations," to be published in *STLE Tribology Transactions*.

[15] Johnson, K., and Tevaarwerk, J. "Shear Behavior of EHD Oil Films," *Proc. Royal Soc. London*, A. 356, p. 215, 1977.

[16] Bair, S. and Winer, W. "A Rheological Model for EHD Contacts Based on Primary Laboratory Data," *J. Lub. Tech.*, V. 101, p. 258, 1979.

[17] Trachman, E. and Cheng, H. "Thermal and Non-Newtonian Effects on Traction in EHL Contacts," *Proc. Inst. Mech. Engr.*, C37/72, p. 142, 1972.

[18] Hamrock, B. "Starved EHD Lubricated Elliptical Contacts," NASA TM #81549, 1980.

[19] McCool, J. "Traction Model Development," AFWAL-TR-87-407, 1987.

[20] Kaye, K. and Winer, W. "An Experimental Evaluation of the Hamrock and Davison Minimum Film Thickness Equation for Fully Flooded EHD Point Contacts," *ASME J. of Tribology*, V. 103, p. 284, 1981.

[21] Wedeven, L. "Effect of Starvation on Film Thickness and Traction under EHD Rolling and Sliding Conditions," NASA TN D-8087, 1975.

[22] McCool, J., Chiu, Y., Crecelius, W., Liu, J. and Rosenlieb, J. "Influence of Hydrodynamic Lubrication on Life and Operation of Turbine Engine Ball Bearings -- Bearing Design Manual," AFAPL-TR-75-25.

[23] McCool, J. "Comparison of Models for the Contact of Rough Surfaces," *Wear*, V. 107, p. 37, 1986.

[24] McCool, J. "Relating Profile Instrument Measurements to the Functional Performance of Rough Surfaces," *ASME J. of Tribology*, V. 109, p. 264, 1987.

[25] Greenwood, J. and Williamson, J. "Contact of Nominally Flat Surfaces," *Proc. Royal Soc. London*, Series A, V. 295, p. 300, 1966.

[26] Averbach, B. and Bamberger, E. "Analysis of Bearing Incidents in Aircraft Gas Turbine Mainshaft Bearings," *STLE Tribology Transc.*, V. 34, p. 241, 1991.

[27] Andereason, S. and Lund, T. "Ball Bearing Endurance Testing Considering Elastohydrodynamic Lubrication," *Proc. Inst. Mech. Engr.*, C36/72, p. 138, 1972.

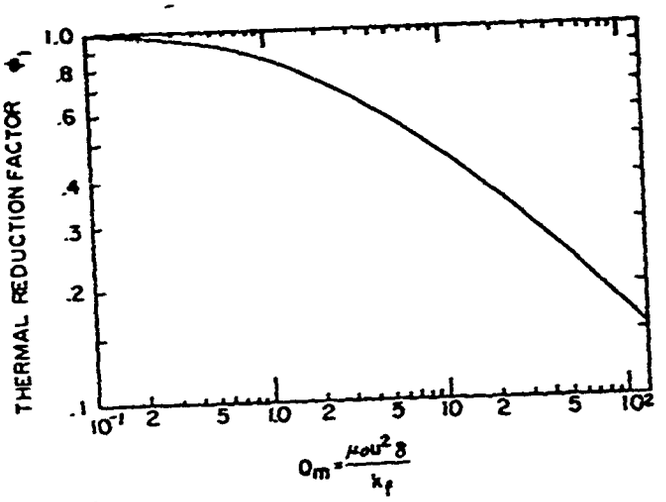


FIGURE 1 Thermal reduction factor, where Q_m = thermal parameter, μ_0 = ambient lubricant viscosity, u = rolling velocity, T_0 = ambient temperature, k_f = lubricant thermal conductivity, and β = temperature viscosity coefficient in the empirical relation $\mu = \mu_0 e^{\beta(T - T_0)}$. [ref. 6]

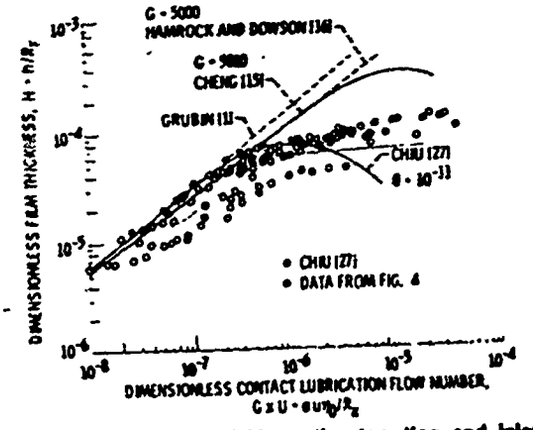


Fig. 2 Theoretical effect of kinematic starvation and inlet shear heating on film thickness and comparison with experiment [ref. 10]

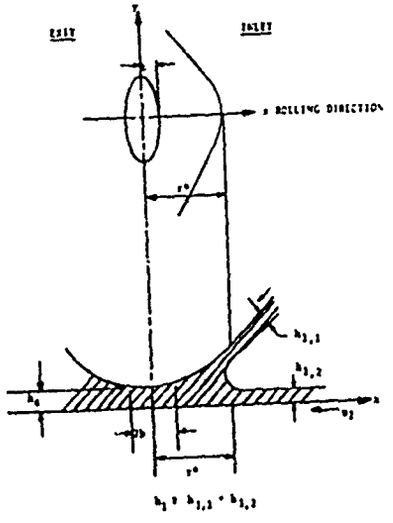


FIGURE 3 FILM GEOMETRY [ref. 22]

Effects of Riblets on Turbine Blade Heat Transfer

and

Velocity and Heat Transfer Measurements in a Ribbed Channel

Paul K. Maciejewski
University of Pittsburgh
Mechanical Engineering Department
Pittsburgh, PA

August 14, 1991

Abstract

This report describes two experiments in the general area of gas turbine heat transfer. The first is a study of the effects of riblets on the average heat transfer rate on a turbine blade in a cascade for cases with and without free stream turbulence. The second is a study of the velocity and heat transfer in a ribbed channel for six different flow geometries. Each experiment will be presented separately.

Effects of Riblets on Turbine Blade Heat Transfer

Introduction

Ribleted surfaces are believed to reduce surface heat transfer rates, under certain conditions. A reduction in local heat transfer rates on a flat plate of up to 6% is reported to occur when the riblet height is approximately 5 to 15 wall units. However, riblet heights greater than 25 wall units increase heat transfer rates on a flat plate by up to 20%. Presumably, for a fixed physical riblet size, riblets may increase, decrease, or have no effect on the average heat transfer rate from a turbine blade, depending on the Reynolds number of the flow.

The present experiments were designed to answer the following two questions: (1) Do riblets reduce the average heat transfer on a turbine blade when the approaching flow has no free stream turbulence?, and (2) Does the presence of free stream turbulence in the approaching flow alter the effects of riblets on the average heat transfer from a turbine blade?

Discussion and Results

An instrumented heat transfer surface was designed for testing in the cascade facility in Building 19 (The Five Foot Wind Tunnel Building). Since the maximum expected reduction in the heat transfer rate on the test turbine blade is only 6%, the experiment was designed to minimize the uncertainty in measured average Stanton number. The entire blade was 4.5 inches wide, but only the center 2 inches were used for heat transfer measurements in order to eliminate end wall effects. The outer portions of the blade were heated and served to guard the center portion of the blade against conduction losses. The same blade was to be used both for baseline testing with a smooth surface, then for further testing with a ribleted surface.

The blade was tested prior to the engraving of the riblets, both with and without free stream turbulence, at three different Reynolds numbers. The results of these baseline tests are presented in Table 1. The velocities reported in Table 1 were measured approximately 0.5 inches upstream of the stagnation point on the blade. The cases run with free stream turbulence, which was produced by a grid upstream of the cascade, have heat transfer rates that are about 25%-35% higher than the corresponding runs without the turbulence generating grids in place.

The testing of the blade after the etching of the riblets was to be completed by Capt. Steve Meschwitz. At the highest Reynolds number, one would expect the riblets to augment heat transfer, since the riblet height relative to the boundary layer thickest will be greater than y+=20 over most if the turbulent section of the boundary layer on the blade. At the lowest Reynolds number, one might expect a possible reduction in the average heat transfer rate from the blade.

Recommendations

(1) Redesign test blade to create more uniform boundary conditions.

(2) Redesign test blade to further minimize uncertainty in Stanton number.

(3) Develop an analysis to estimate the expected maximum reduction in heat transfer on the blade due to the presence of riblets and determine the optimum Reynolds number for possible heat transfer reduction.

Velocity and Heat Transfer Measurements in a Ribbed Channel

Introduction

The design of cooling passages for turbine blades involve complex geometries. The purpose of the present study is to determine which of six geometries is most effective and to explore the possibility of correlating cooling effectiveness with local measures of the turbulence inside a passage.

Discussion and Results

The experiments were performed in a pre-existing test channel that was 60 inches in length and which was 6 inches by 6 inches in cross section. Each configuration of the channel consisted of five ribs, each cut from a 1 inch by 1 inch aluminum bar, being mounted on each of two opposing walls of the channel, with the fourth rib on one side of the channel instrumented for heat transfer measurements. Velocity measurements were taken just upstream of the heated rib, at locations of 1, 1.5, 2, 2.5 and 3 inches from the wall on which the heated rib was mounted. The six configurations tested were the

following: (1) perpendicular ribs, in line, (2) perpendicular ribs, staggered, (3) slanted ribs, parallel and in line, (4) slanted ribs, parallel and staggered, (5) slanted ribs, crossed and in line, and (6) slanted ribs, crossed and staggered. Table 2 presents the velocity, heat transfer, and pressure drop measurements for each of these configurations, which are labelled geometries 1 through 6, respectively. While the velocity data and heat transfer data are accurate to within a few percent, the pressure drop data is more uncertain and difficult to interpret. Pressure drop data for Geometry #5 and Geometry #6 are particularly suspect, since these data where taken at a point in the experiment where the pressure transducer used was behaving unreliably. Geometry #4 offers the highest heat transfer rates on average. The velocity data indicate that each configuration establishes its own flow pattern.

Recommendations

(1) Instrument channel to provide more accurate pressure drop data.

(2) Instrument channel to measure total flow rate through channel. Perhaps an orifice in a pipe somewhere downstream of the test section would suffice. In combination with the pressure drop data, this would allow one to estimate the pumping power required to achieve a desired heat transfer rate.

(3) Take measurements of the spanwise variation in the velocity, particularly for cases with asymmetric geometry.

Table 1: Results from cascade flow with and without free stream turbulence

Without Free Stream Turbulence

U	rms U	TUu	V	rms V	TUv	Re	h	St
6.59	0.09	1.4	0.15	0.07	1.1	44446	66.10	0.00868
17.11	0.13	0.8	0.06	0.15	0.9	115472	111.87	0.00566
49.55	0.71	1.4	-1.23	0.76	1.5	334327	210.95	0.00368

With Free Stream Turbulence

U	rms U	TUu	V	rms V	TUv	Re	h	St
5.73	0.94	16.5	1.63	0.67	11.6	38658	77.23	0.01166
14.87	1.82	12.3	-0.27	1.84	12.4	100348	121.28	0.00706
44.59	4.66	10.5	0.35	5.07	11.4	300822	254.12	0.00493

Table 2: Results from channel flow with transverse ribs

Geometry #1

	U	rms U	V	rms V	h	dP/Dx
Run #1	14.39	3.25	0.61	4.68	137.53	-0.252
	16.78	3.87	-0.08	4.72		
	20.89	3.81	-0.29	4.29		
	22.34	3.23	-0.03	3.52		
	22.91	3.11	-0.31	3.04		
Run #2	6.71	1.62	0.04	2.20	90.23	-0.072
	8.32	2.08	0.15	2.21		
	9.98	1.95	-0.41	1.82		
	11.19	1.68	0.02	1.63		
	10.93	1.64	-0.18	1.56		
Run #3	2.96	0.81	0.68	0.93	58.47	-0.018
	4.30	0.85	0.40	0.94		
	4.81	0.88	0.16	0.79		
	4.49	0.88	-0.04	0.69		
	4.63	0.82	-0.01	0.66		

Geometry #2

	U	rms U	V	rms V	h	dP/Dx
Run #1	13.55	3.17	0.99	3.97	125.73	-0.129
	17.89	3.36	0.17	4.16		
	20.09	3.35	0.31	4.14		
	21.97	3.05	0.45	3.65		
	21.88	3.20	0.04	3.24		
Run #2	6.80	1.62	0.85	2.35	84.26	-0.063
	8.44	1.79	0.72	2.11		
	9.61	1.70	0.30	2.05		
	10.11	1.81	0.45	1.86		
	10.75	1.75	0.44	1.65		
Run #3	2.81	0.61	0.38	0.81	56.31	-0.023
	3.72	0.79	0.58	1.10		
	4.05	0.71	0.32	0.88		
	4.46	0.67	0.37	0.86		
	4.48	0.64	0.44	0.79		

Table 2: Results from channel flow with transverse ribs (continued)

Geometry #3

	U	rms U	V	rms V	h	dP/Dx
Run #1	16.83	3.16	3.39	3.62	141.83	-0.230
	15.89	3.40	0.19	4.69		
	14.80	3.31	-4.07	4.06		
	13.90	3.11	-6.75	3.14		
	13.76	2.99	-6.30	3.37		
Run #2	8.57	1.62	2.04	1.72	90.15	-0.067
	7.71	1.51	0.65	2.38		
	7.56	1.62	-1.87	1.81		
	7.02	1.57	-3.47	1.52		
	6.64	1.48	-2.90	1.56		
Run #3	4.62	0.89	1.52	0.83	59.87	-0.029
	4.40	0.72	0.07	0.99		
	3.83	0.87	-1.32	0.86		
	3.47	0.74	-1.94	0.76		
	3.54	0.66	-1.29	0.74		

Geometry #4

	U	rms U	V	rms V	h	dP/Dx
Run #1	21.38	2.49	2.34	2.57	152.41	-0.304
	18.78	2.82	-4.42	3.68		
	15.22	3.31	-9.66	4.33		
	12.17	3.65	-10.86	4.31		
	13.91	3.94	-7.42	4.67		
Run #2	9.45	0.77	0.87	0.94	93.08	-0.048
	8.24	1.33	-2.32	1.55		
	6.39	1.47	-4.62	1.58		
	5.00	1.44	-5.25	1.46		
	5.90	1.77	-4.24	1.62		
Run #3	3.73	0.81	1.77	0.67	54.48	-0.033
	3.91	0.61	-0.37	0.47		
	2.48	0.64	-1.75	0.50		
	1.57	0.58	-1.89	0.43		
	1.84	0.65	-1.73	0.52		

Table 2: Results from channel flow with transverse ribs (continued)

Geometry #5

	U	rms U	V	rms V	h	dP/Dx
Run #1	15.17	2.29	4.79	3.06	119.76	-0.117
	18.01	2.10	5.91	3.07		
	20.72	1.89	4.94	2.46		
	22.00	1.50	2.57	2.15		
	22.14	1.47	-0.14	1.90		
Run #2	7.24	0.92	2.15	1.09	75.74	-0.024
	8.11	0.88	2.53	1.23		
	8.88	0.94	2.14	1.16		
	9.43	0.95	1.37	0.94		
	9.82	0.81	0.33	0.86		
Run #3	2.74	0.42	0.84	0.56	50.52	-0.057
	3.02	0.45	1.40	0.57		
	3.51	0.44	1.21	0.49		
	3.95	0.47	0.87	0.46		
	4.01	0.42	0.24	0.39		

Geometry #6

	U	rms U	V	rms V	h	dP/Dx
Run #1	13.70	2.23	2.57	3.12	116.76	-0.085
	15.65	2.31	4.19	3.50		
	18.88	2.17	4.70	2.80		
	21.03	1.55	3.30	1.98		
	22.03	1.44	0.47	1.49		
Run #2	6.47	1.12	1.36	1.51	72.84	-0.038
	7.75	1.08	2.56	1.47		
	8.96	0.73	2.66	1.13		
	9.62	0.76	1.53	0.90		
	9.60	0.66	0.29	0.70		
Run #3	2.44	0.49	0.56	0.55	46.71	0.021
	2.81	0.49	1.12	0.61		
	3.41	0.55	1.44	0.64		
	3.90	0.46	1.01	0.46		
	3.90	0.41	0.22	0.38		

ELECTRON DENSITY MEASUREMENTS IN THERMIONICALLY-ASSISTED DISCHARGES IN CESIUM-ARGON PLASMAS

S. Douglas Marcum, Associate Professor of Physics
Miami University, Oxford, OH 45056
RDL Associate Employee #100

Abstract

Electron excitation temperature and density were measured in a low pressure (0.05 Torr Cs, 2 Torr total pressure) cesium-argon discharge that uses a heated cathode (1100 K). The excitation temperature determination is based upon a model that allows for the calculation of cesium excited state densities and resulting absolute emission intensities for low electron density, thermionically-assisted Cs-Ar discharges. The model assumes that the dominant creation processes for Cs excited states are electron impact excitation and radiative cascade from higher levels. De-excitation is assumed to be by spontaneous emission only. Electron excitation temperature is determined by comparison of measured and calculated emission spectra. Measured absolute emission intensities then allow for electron density determination. For the low current (0.035 mA/cm^2) discharge studied, an excitation temperature of 4900 K and an electron density of 2.5×10^7 cm^{-3} was determined. The measured density was, as expected for such non-equilibrium plasmas, considerably lower than the roughly 10^9 cm^{-3} value obtained from current continuity.

INTRODUCTION

Several theoretical and experimental investigations have been reported[1-7] that focus on either excited state density distributions, electron (excitation) temperatures and electron densities in low pressure (several Torr and below) argon-cesium and pure cesium plasmas. The most complete theoretical treatment of pure cesium plasmas published to date is that of Norcross and Stone,[2] the results of which apply directly to ignited mode thermionic converter plasmas with electron densities in excess of 10^{12} cm^{-3}. Their treatment of cesium level populations was, in turn, based upon the classic work on hydrogen plasmas by Bates, Kingston and McWhirter.[8] In both cases, single Maxwellian electron energy distribution functions (EEDF) were used in the solution (steady state) to the transition rate equations for a large number of excited states.

However, with low pressure argon-cesium discharges that employ an ambient temperature ("cold") cathode and have lower electron densities, large deviations from equilibrium (Maxwellian electron energy distributions) occur[1,6,7]. The reason for this is that at low degrees of ionization, elastic and inelastic electron-atom collisions dominate over Coulomb collisions. As an alternative to the complex problem of solving the Boltzmann equation[7] for the distribution function, Vriens[6] has used two- and three-electron group (temperature) models for the bulk and high energy tail electrons in energy balance calculations in such cases. More recently, Wani[9] has reported some success with a simpler two-temperature model for similar (Ar-Hg) plasmas that uses different temperature Maxwellian

distributions for the bulk and tail electrons to describe the overall EEDF. Although the purpose of this work is not to evaluate the EEDF of our plasmas in detail, the results have a direct bearing on the validity of Wani's approach, as discussed below.

In this work, we studied low pressure, thermally-assisted glow discharges in an argon-cesium mixture. In such discharges, large numbers of electrons are emitted from the hot cathode surface (1100 K) and a discharge is sustainable by potentials less than 20 V. At the low discharge currents used in this measurement (5 mA, 0.35 mA/cm^2), the bulk electron density is of the order of 10^9 cm^{-3} (inferred from current continuity) and the electron temperature is relatively low (<5000 K). Under these conditions, it is possible to construct a simple model[1] that allows calculation of cesium excited state densities. The model is based on the assumptions that the excited state populations are sustained by electron impact excitation from the ground state along with radiative cascade from higher levels, and de-populated only by spontaneous emission. Spontaneous emission intensities are easily computed from the resulting excited state densities and the calculated spectra fit to measured spectra to yield electron excitation temperature[1]. The observed spectra sample only the tail of the overall EEDF. As suggested by Wani[9], we took the tail of the distribution to be described by a Maxwellian, the temperature of which defines the excitation temperature of cesium excited states. To our knowledge, the effective electron density in the tail of such a distribution has not been measured prior to this study.

Such plasmas are similar to those studied in the theoretical work of Norcross and Stone[2], but, due to the low electron density in the present case, spontaneous emission intensities can be predicted by a simpler model. The surprisingly large role played by radiative cascade in the population of excited states in such plasmas was recently reported[1].

DISCUSSION: EXPERIMENT AND THEORY

The primary aim of the present study was to perform spectral measurements sufficient to allow for an electron density measurement based upon the high-energy tail of the non-equilibrium EEDF. Experimentally, this involves essentially repetition of the emission intensity measurements carried out in Reference 1, using a spectroscopic system that has been carefully calibrated for absolute intensity.

The apparatus used to generate and confine the Ar-Cs plasmas and the spectroscopic measurement system used is shown in the block diagram of Figure 1. Since the experimental system is essentially identical to that used in a previous study[1], a detailed description is omitted. The essential features of the experimental apparatus and operating conditions are as follows: The discharge electrodes were planar nickel disks 4.3 cm in diameter and the anode-cathode gap was 1 cm. Cathode temperature was obtained from blackbody radiation spectrum measurement.

Discharge current and voltage were monitored at all times and the measurements reported here were obtained at 5 mA current (0.35 mA/cm^2). Discharge voltage ranged from 70 V with a cathode

temperature of 520 K, to 19 V at the normal operating temperature of 1100 K. The gas mixture used was 0.05 Torr cesium in 2 Torr argon. Spontaneous emission and cathode blackbody spectra were measured using an intensified-diode-array based optical multichannel analyzer that was absolutely calibrated for intensity using an NIST-traceable standard lamp. The optical arrangement was such that spatial resolution was roughly 1 mm.

The theory supporting the kinetics model used to predict absolute emission intensities for the type of plasma studied in this work is fully developed in Reference 1. The essential features of the theory are reiterated here for the sake of completeness. The mechanisms[1,2] that determine cesium excited state densities under the plasma conditions described above are all electron-atom collisions and radiative decays (spontaneous emission and radiative recombination). The collision processes are electron impact excitation and ionization, super-elastic relaxation and three-body [illegible]tion. Excitation by photoabsorption is not considered since the plasmas are optically thin to all but a few lines for typical plasma dimensions. Partial self absorption of the 6S - nP resonance lines is found for low n, but does not alter the conclusions of this work. Similarly, molecular processes are not considered since the concentrations of molecular ions and neutrals in equilibrium vapor at typical conditions are low[10-12]. Atom-atom and atom-ion collisions are neglected[1,2,7] due to their low temperatures.

Justification for only considering excited state production by electron impact excitation from the ground state and radiative cascade from higher levels, and destruction by spontaneous emission, is as follows:

Electron impact excitation of cesium from the ground state to excited state j is written,

$$e + Cs(1) \rightarrow Cs^*(j) + e, \quad (1)$$

which has a typical cross section[13-19], $\sigma_{exj} = 10^{-17}$ cm^2. Electron impact excitation from only the ground state is considered because at low current densities the ground state population is many orders of magnitude larger than any excited state population. Electron impact ionization contributes to excited state populations only indirectly through subsequent radiative or three-body recombination, i.e.

$$e + Cs \rightarrow Cs^+ + e + e, \quad (2)$$

followed by,

$$Cs^+ + e \rightarrow Cs^*(j) + h\nu \quad (3)$$

or

$$Cs^+ + e + e \rightarrow Cs^*(j) + e. \quad (4)$$

Typical cross sections for reactions (2) and (3) are $\sigma_{ion} = 10^{-16}$ cm^2,[18] and $\sigma_{recj} = 10^{-20}$ cm^2,[11,20], respectively, while reaction 4 has a rate coefficient[1,2,21] of 10^{-23} cm^{6}s^{-1}.

Taking the electron density approximately equal to the ion density in the plasma region ($n_e \simeq n_i = 10^9$ cm^{-3}), assuming a ground state density of 10^{15} cm^{-3} (0.05 Torr) and using an average electron velocity[7] of 5 x 10^6 cm/s gives order of magnitude production rates for processes 1-4 of 10^{13} cm^{-3}s^{-1}, 10^{13} cm^{-3}s^{-1}, 10^4 cm^{-3}s^{-1} and 10^4 cm^{-3}s^{-1}, respectively. Under these conditions, electron impact excitation from the ground state dominates the production of cesium excited states. For higher electron densities, three-body recombination becomes an important

process,[1,2,8] thus the model should be valid for electron densities up to $\sim 10^{11}$ cm^{-3}. For example, with $n_e = 10^{12}$ cm^{-3} and at the typical electron temperatures encountered in this work, the excited state production rate by three-body recombination[2] is roughly 10^{13} $cm^{-3}s^{-1}$.

Turning to destruction mechanisms, spontaneous emission dominates, as is shown below. Spontaneous emission,

$$Cs^*(j) \rightarrow Cs^*(k) + h\nu, \qquad (5)$$

has a rate determined by a given excited state's density and Einstein coefficient, A_{jk}. Estimating a typical excited state density to be of the order of 10^7 cm^{-3} and using a typical A coefficient of 10^7 s^{-1} (100 ns lifetime), gives a destruction rate by spontaneous emission of 10^{14} $cm^{-3}s^{-1}$. Of course, in reaction 5, loss of state j means production of state k and spontaneous emission in the form of radiative cascade from higher levels must also be taken as a non-negligible production mechanism for excited states.

Super-elastic collision or electron impact de-excitation is written,

$$Cs^*(j) + e \rightarrow Cs^*(k) + e, \quad j>k. \qquad (6)$$

This is essentially the reverse of reaction (1). Detailed balancing gives a typical cross section, $\sigma_{dexj} = 10^{-15}$ cm^2. Assuming conditions already cited provides an order of magnitude destruction rate for (6) of 10^6 $cm^{-3}s^{-1}$, which is extremely small compared to the spontaneous emission decay rate.

Penning ionization of cesium by argon metastables cannot contribute substantially to the Cs^+ population under the plasma conditions specified above, and thus significantly populate cesium

excited states by subsequent recombination; the low electron temperatures of hot cathode discharges do not provide significant numbers of energetic electrons to populate the metastable states of argon. This is evidenced by the absence of significant argon emissions in spectra measured using a heated cathode. However, at ambient cathode temperature (higher electron energy distribution), significant argon emissions are apparent and the assumption breaks down. The cathode temperature dependent emission spectra, shown in Figure 2, clearly show this effect.

Following the arguments above and using only processes (1) and (5), the populations of excited states are found by the steady state solution of the set of equations of the form[1,2,8]:

$$\frac{dn(j)}{dt} = <\sigma_{exj} v> n_e n(1) + \sum_{i>j} n(i) A_{ij} - n(j) \sum_{j>k} A_{jk} \qquad 7$$

The first term on the right represents electron impact excitation from the ground state, with the ground state density taken as, n(1) = 1.67 x 10^{15} cm^{-3}, corresponding to the vapor pressure of cesium used in the experiments (0.05 Torr) and v is electron velocity. The second term accounts for radiative cascade from higher levels, while the third gives the loss due to spontaneous emission to all possible lower levels (dipole allowed transitions only).

It is necessary from a practical standpoint to reduce the number of equations (7) that must be solved. For the purposes of this work, the infinite set of equations (7) was reduced to a finite set that included only excited states up through n=10 (all L and J) for a total of 53 cesium levels including the ground state. This number of levels has previously been shown sufficient[1] to model the plasma of interest to this work.

Solution of equations (7) in steady state requires extensive knowledge of electron impact excitation cross sections and A coefficients of the cesium levels involved. A consistent set of numerical electron impact excitation cross sections were used that were calculated by first order many body theory[19]. They agree well with measurements and other calculations[2,13-17,23] and are available for all levels through n = 10 and are in more suitable (numerical) form than the measured cross sections. The cross sections were numerically integrated over assumed Maxwellian distributions for a range of temperatures (3000 - 6000 K in 100 K increments) to provide an extensive set of electron impact excitation rate coefficients for each state at each temperature. The A coefficients used were calculated by the same code[19] as the cross sections and also agree well with available calculated and measured values[24-27].

The resulting set of differential equations was solved by using Saha populations as initial conditions and then employing an iterative (in time) Runge-Kutta method to evolve to steady state. Calculated level densities for $n_e = 10^9$ cm^{-3} and T_e ranging from 4000 K to 5500 K are shown in Table 1 of Reference 1.

Calculation of absolute emission intensities from the excited state densities derived from solution of equations (7) is by the relation,

$$I_{jk} = A_{jk}\, n(j)\, (E_j - E_k) \text{ watts/cm}^3, \qquad (8)$$

where E_j and E_k are the energies of the upper and lower levels of the transition, respectively. Table I shows predicted absolute emission intensities calculated using the same electron density and range of electron temperatures cited above. The radiative cascade

contribution to emission intensity (last column in Table I) is quite large for states with low n and l.

RESULTS

The visible and near UV emission spectrum measured at a location 5.9 mm from the cathode (plasma region) for 1100 K cathode temperature is shown in Figure 3. The calculated best fit normalized intensities are also shown in that Figure for lines in the sharp, principal and diffuse series. The intensity error bars shown are calculated at plus and minus 500 K from the best fit temperature. The electron excitation temperature corresponding to the best fit in the case presented here is found to be 4900 $\pm$ 500 K.

As mentioned above, the calculation ignores radiation trapping and, therefore, the predicted intensities for 6S to low nP transitions are expected to be greater than the measured intensities; of course the calculated intensities of 6S to high nP transitions agree well with the measured spectra.

Due to the low electron density and the large excitation cross sections of the resonance levels, the EEDF should be significantly depressed[1,6,7] above the resonance excitation threshold (1.4 eV). This is shown schematically in Figure 4. However, our observations sample only the tail of the distribution (above 3 eV). With cascade properly taken into account, the excellent fit of Figure 3 indicates that the the cesium excited states sampled in our experiments can be represented by a Maxwellian. That is to say, <u>at least</u> the tail of the distribution is well characterized by a Maxwellian. The results of this work strongly support the notion that the EEDF for such plasmas

can be reasonably approximated by the simple two-temperature Maxwellian distribution suggested by Wani[9].

The bulk electron density inferred from current continuity is $1.0 \pm 0.5 \times 10^9$ cm^{-3}. As suggested by Figure 4, the non-equilibrium nature of our plasma would tend to cause the number density of electrons in the high energy tail of the distribution to be smaller than that derived from any measurement of the bulk electron density. Simply, Coulomb collisions cannot keep pace with the selective consumption of hot electrons by inelastic collisions with cesium ground states which populate the resonance levels. Since our spectroscopic measurements sample only the high energy tail of the distribution, the depressed tail should result in measured absolute emission intensities which are substantially than the predicted intensities, since the predictions were based on the bulk electron density, $n_e = 10^9$ cm^{-3}. The measured tail electron density is found in our case by simply taking the average ratio of measured to calculated absolute intensity and multiplying the resulting value by the electron density used for the calculations. Complete equilibrium would yield an average ratio of 1, while non-equilibrium should produce a ratio less than one, with smaller ratios representing larger departures from equilibrium. In the case of interest here, that ratio was roughly 1/40, giving a tail electron density of roughly $2.5 \pm 0.5 \times 10^7$ cm^{-3}.

The terms electron excitation density or tail electron density as used here should be clarified somewhat. The term electron excitation temperature is used to indicate an electron temperature inferred from the observed atomic excitation. The electron density that has been

measured here is similarly inferred from the observed atomic excitation - hence the term electron excitation density.

CONCLUSIONS

A simple but accurate model has been applied to measure electron excitation temperature and density in a non-equilibrium hot cathode discharge. The accuracy of the calculation is evident not only by the excellent agreement with the measured spectra, but also by the predictive variation from the measured self-absorbed emission lines. The model is applicable to both quasi-equilibrium and thermal equilibrium plasmas. Although the present calculation is applicable to discharges with current densities below 0.1 A/cm^2, inclusion of the three-body recombination term in the rate equation will permit extension of applicability to high current density thermionic diodes. For high current density (large electron density) plasmas which are much closer to equilibrium, the temperature and density measured in the manner described above should reflect the true electron temperature and density. Even in such cases, radiative cascade would be an important contributor to excited state densities and resulting emission intensities.

REFERENCES

1. S.D. Marcum, J.L. Myers, M.A. Gieske, M. Tackett and B.N. Ganguly, J. Appl. Phys. 69, 27 (1991).
2. D.W. Norcross, and P.M. Stone, J. Quant. Spectrosc. Radiat. Transfer 8, 1655-1656 (1968).
3. P.E. Oettinger, J. Appl. Phys. 55, 3411 (1984).
4. W.H. Reichelt, Appl. Phys. Lett. 14, 382 (1969).
5. R.J. Donahue, and R.F. Majkowski, J. Appl. Phys. 33, 3 (1962).
6. L. Vriens, J. Appl. Phys. 45, 1191 (1974); J. Appl. Phys 44, 3980 (1973).
7. A.J. Postma, Physica 43, 229 (1969); Physica 43, 465 (1969); Physica 48, 447 (1970).
8. D.R. Bates, A.E. Kingston, and R.W.P. McWhirter, Proc. Roy. Soc. A267, 297 (1962); Proc. Roy. Soc. A270, 155 (1962).
9. K. Wani, J. Appl. Phys., 58, 2968 (1985); J. Appl. Phys., 63, 5683 (1988).
10. D.G. Clifton, Los Alamos Report, LA - 2419 (1960).
11. L. Agnew, and W.H. Reichelt, J. Appl. Phys. 39, 3149 (1968).
12. A.G.F. Kniazzeh, and E.N. Carabateas, IEEE Thermionic Conversion Specialist Conf. Proc., San Diego, CA October 1965.
13. I.P. Zapesochnyi, and L.L. Shimon, Opt. Spect. 20, 291 (1966).
14. I.P. Zapesochnyi, High Temp. 5, 6, (1967).
15. I.P. Zapesochnyi and L.L. Shimon, Opt. Spect. 20, 421 (1965).
16. P.S. Ganas, J. Chem. Phys. 76, 2103 (1982).
17. I.P. Zapeschnyi, E.N. Postol, and A. Aleksakhin, Sov. Phys. JETP, 41, 865 (1976).

18. R.E.H. Clark, Radiation Transport Group, LASL - Private communication.

19. J. Abdallah Jr., R.E.H. Clark and R.D. Cowlan, Los Alamos Manual #LA-11436-M Vol. 1, 1988; R.E.H. Clark, J. Abdallah Jr., G. Csanak, J.B. Mann and R.D. Cowlan, Los Alamos Manual #LA-11436-M Vol. 2, 1988; R.E.H. Clark, J. Abdallah Jr., G. Csanak and S.T. Kramer, Phys. Rev. A 40, 2935 (1989).

20. L. Agnew and C. Summers, Rev. Sci. Instr. 37, 1224 (1966).

21. D.L. Book, NRL Plasma Formulary, NRL Memorandum Report #3332, 46 (1970).

22. D.R. Bates and A.E. Kingston, Planet. Space Sci. 11, 1 (1963).

23. I.P. Zapeschnyi and L.L. Shimon, Sov. Phys. Dokl. 11, 44 (1966).

24. M. Fabry and J.R. Cussenot, Can. Jour. Phys. 54, 836 (1976).

25. M. Fabry, J. Quant. Spectrosc. Radiat. Transfer, 16, 127 (1976).

26. P.M. Stone, Phys. Rev. 127, 1151 (1962).

27. J. Lahiri and S.T. Manson, Phys. Rev. Let., 33, 3151 (1986).

CAPTIONS

Figure 1: Experimental setup.

Figure 2: Cathode temperature dependent spectra of the argon-cesium discharge. With a cathode temperature of 1100 K only cesium emission lines are observed. Note the large number of emission lines due to argon excitation in the spectrum taken using a cathode temperature of 520 K. With argon excitation present, the electron temperature is obviously higher than in the hot cathode case.

Figure 3: Measured versus calculated (denoted by x) cesium emission spectrum. The measured spectrum was taken at a point 5.9 mm away from the cathode (plasma region). The best fit between calculated and measured intensities indicates an electron excitation temperature of 4900 $\pm$ 500 K. The numbers appearing near each calculated intensity point refer to the transition number found in the left-most column of Table I. Transitions marked with letters A-E are lines from the fundamental series in cesium.

Figure 4: EEDF. a) Equilibrium case, b) Non-equilibrium case.

Table I: Calculated absolute intensities of 31 Cs emission lines based on solution of equations (7). The next to last column gives intensities ignoring the cascade contribution to excited state densities and the last column shows the effect of cascade on emission intensities for each transition. Index refers to the level indices used in Table I of Reference 1 and UL and LL represent upper and lower levels of transitions, respectively.

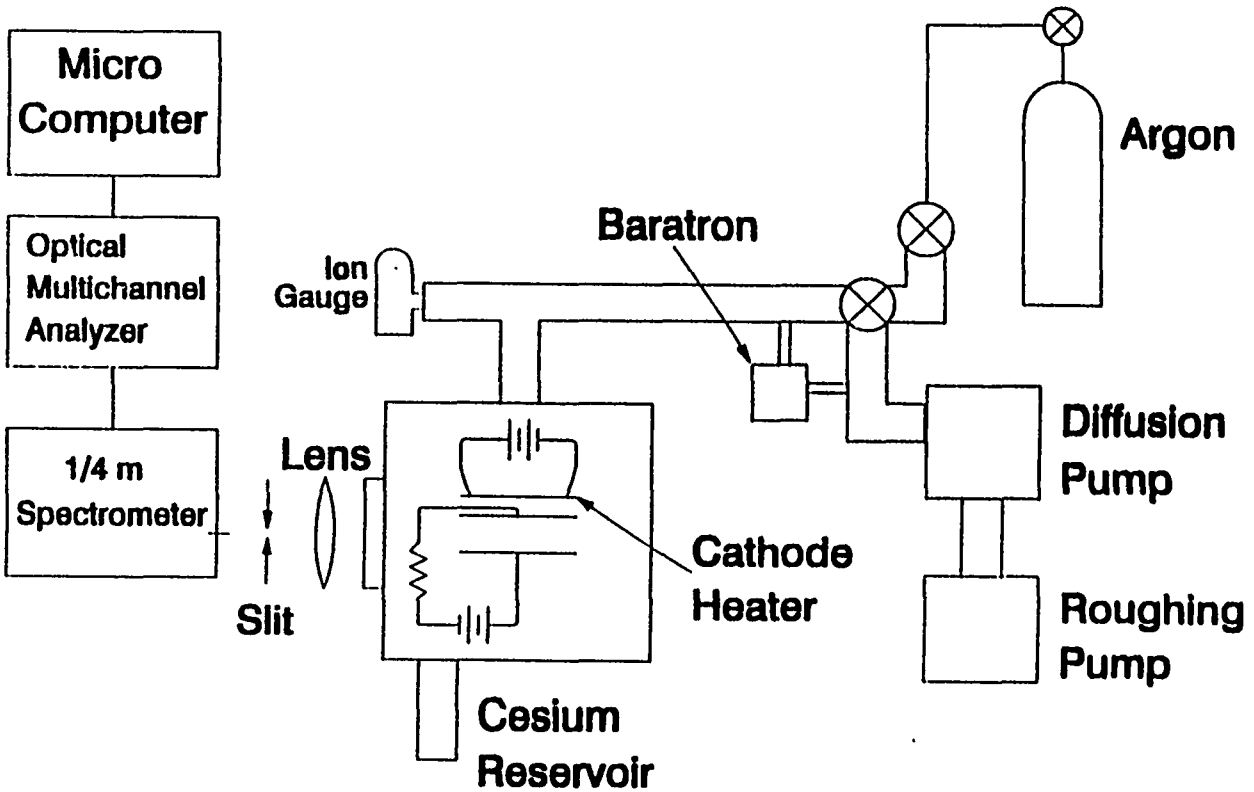
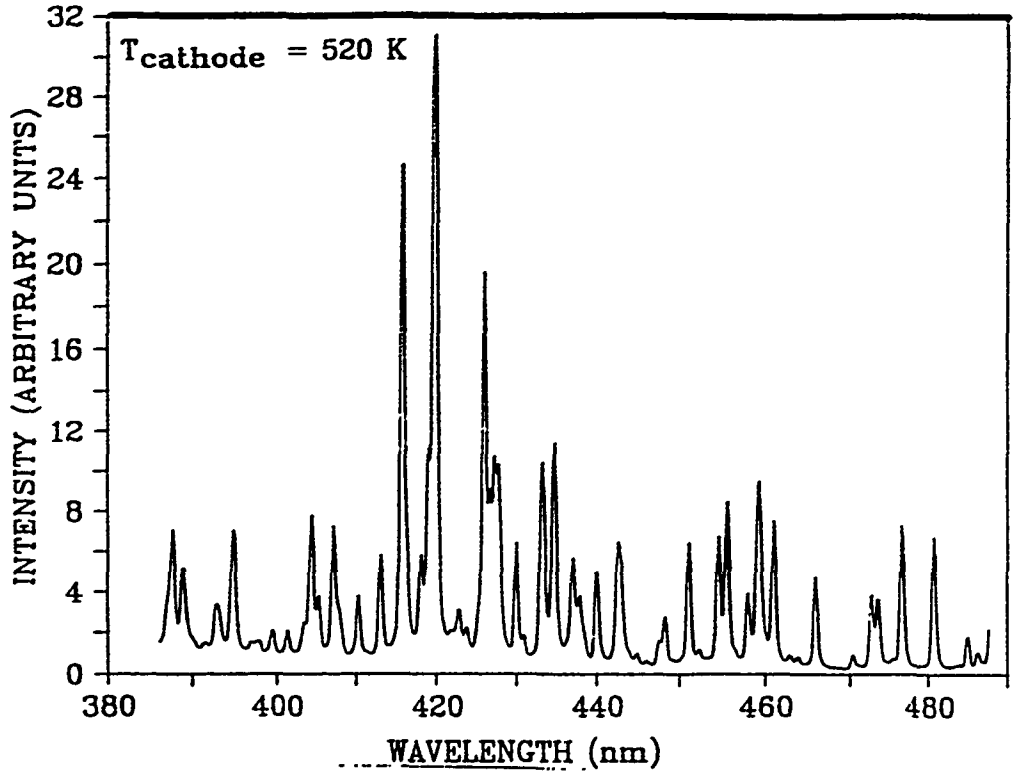

Micro Computer
Optical Multichannel Analyzer
1/4 m Spectrometer
Slit
Lens
Ion Gauge
Baratron
Argon
Diffusion Pump
Cathode Heater
Roughing Pump
Cesium Reservoir

Figure 1:


$T_{cathode}$ = 520 K
INTENSITY (ARBITRARY UNITS)
0
4
8
12
16
20
24
28
32
380
400
420
440
460
480
WAVELENGTH (nm)

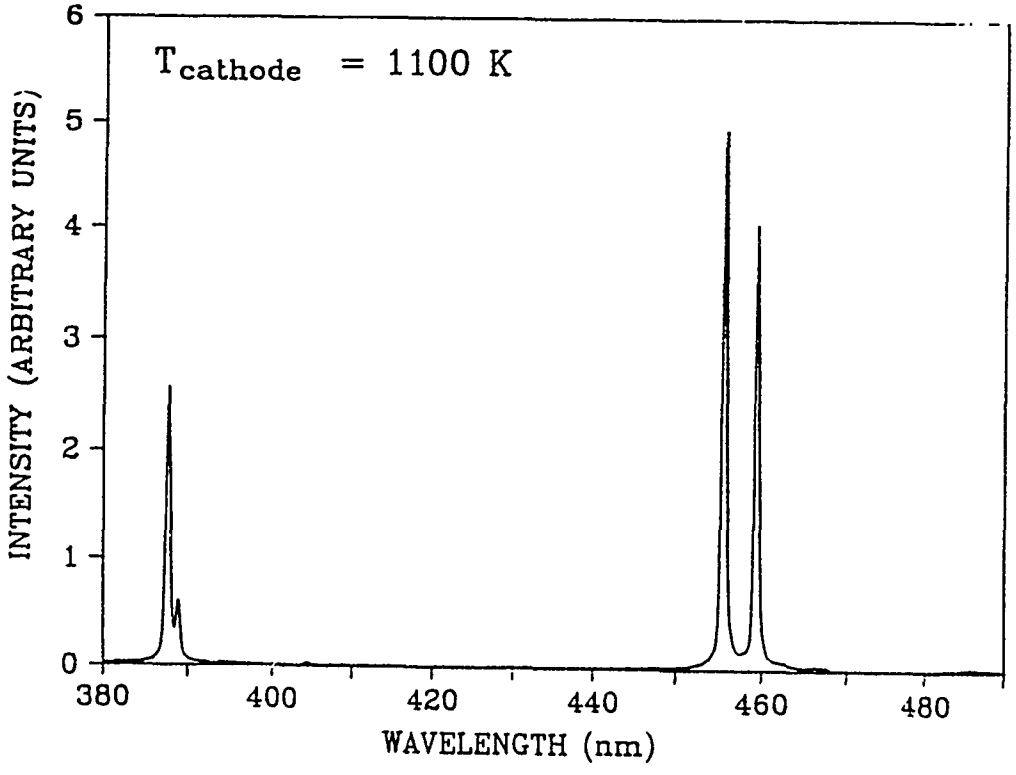
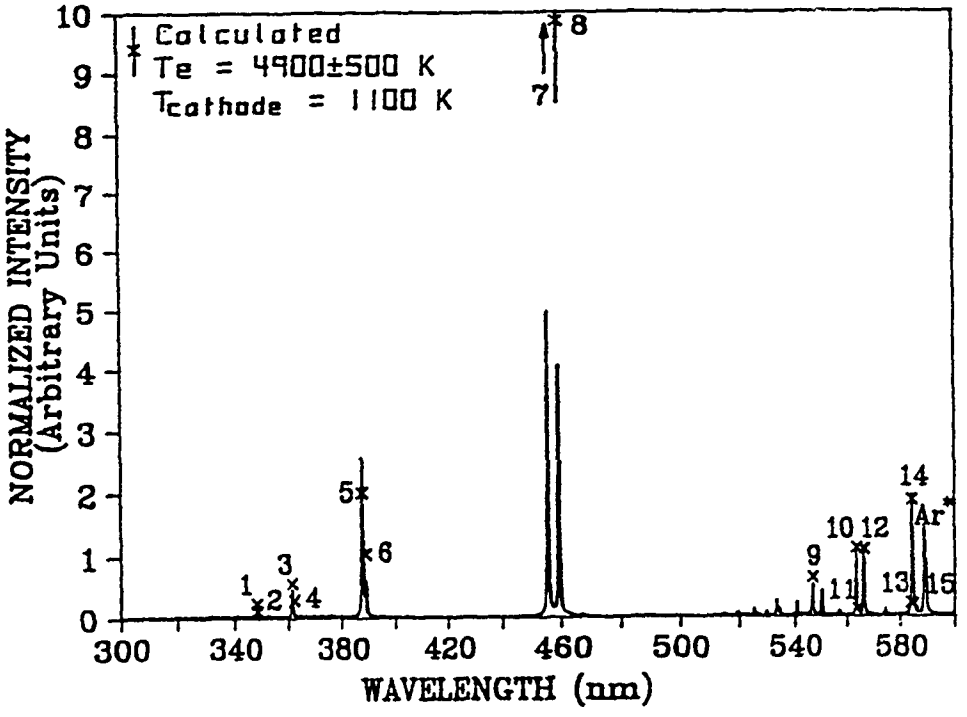

$T_{cathode}$ = 1100 K
INTENSITY (ARBITRARY UNITS)
0
1
2
3
4
5
6
380
400
420
440
460
480
WAVELENGTH (nm)

Figure 2:


Calculated
Te = 4900±500 K
Tcathode = 1100 K
NORMALIZED INTENSITY
(Arbitrary Units)
WAVELENGTH (nm)
Ar*

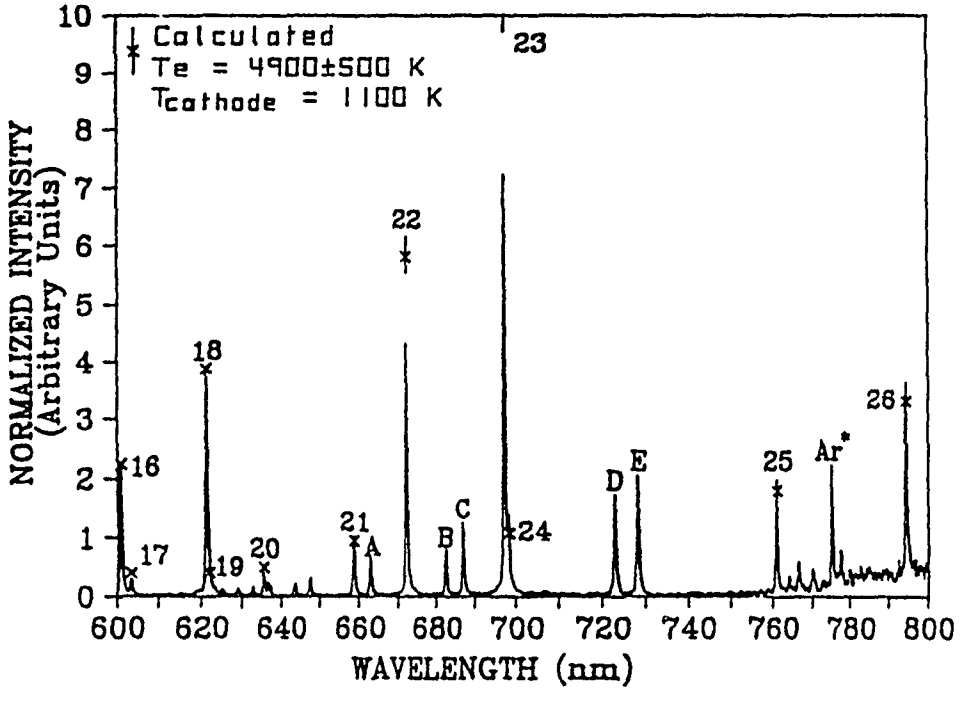

Calculated
Te = 4900±500 K
Tcathode = 1100 K
NORMALIZED INTENSITY
(Arbitrary Units)
WAVELENGTH (nm)
Ar*

Figure 3:

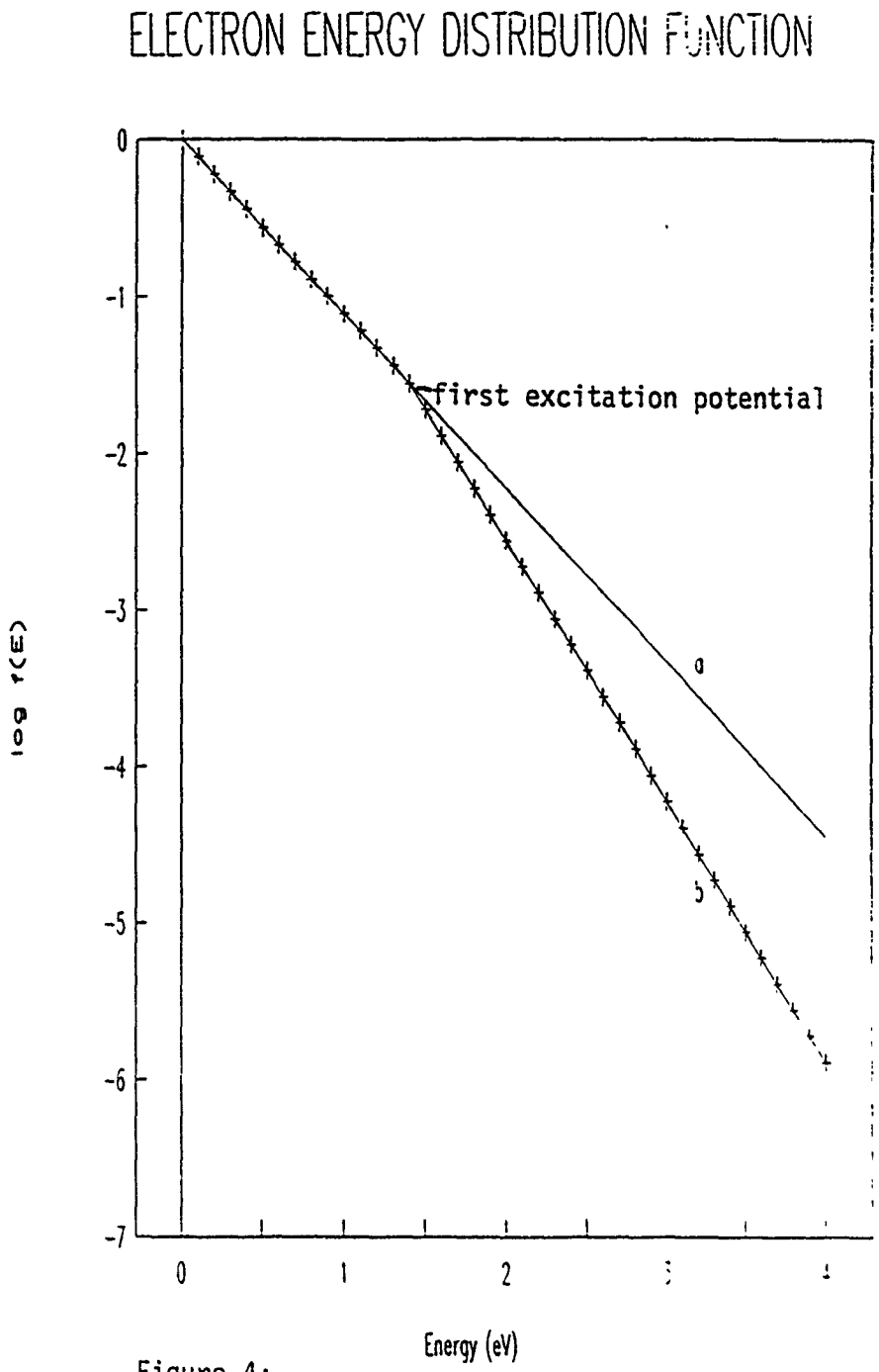

ELECTRON ENERGY DISTRIBUTION FUNCTION
0
-1
-2
-3
-4
-5
-6
-7
log f(E)
first excitation potential
a
b
0
1
2
Energy (eV)

Figure 4:

Table I:

Electron temperature =						4000 K	4500 K	5000 K	5500 K	5500 K	
	Transition		Index		Lambda			Intensity		No	
No.	UL	LL	UL	LL	(nm)			(W/cm^3)		cascade	% DIF
1	$10P_{1/2}$	$6S_{1/2}$	33	1	347.7	4.63(− 7)	1.27(− 6)	2.82(− 6)	5.37(− 6)	5.29(− 6)	1.5
2	$10P_{1/2}$	$6S_{1/2}$	32	1	348 0	2.33(− 7)	6.40(− 7)	1.42(− 6)	2.70(− 6)	2.66(− 6)	1.5
3	$9P_{3/2}$	$6S_{1/2}$	24	1	361.1	1.17(− 6)	3.09(− 6)	6.60(− 6)	1.22(− 5)	1.13(− 5)	8.0
4	$9P_{1/2}$	$6S_{1/2}$	23	1	361.7	5.99(− 7)	1.57(− 6)	3.36(− 6)	6.22(− 6)	5.71(− 6)	8.9
5	$8P_{3/2}$	$6S_{1/2}$	15	1	387.6	4.72(− 6)	1.14(− 5)	2.30(− 5)	4.05(− 5)	3.49(− 5)	16.0
6	$8P_{1/2}$	$6S_{1/2}$	14	1	388.9	2.43(− 6)	5.90(− 6)	1.18(− 5)	2.08(− 5)	1.78(− 5)	16.9
7	$7P_{3/2}$	$6S_{1/2}$	8	1	455.5	5.60(− 5)	1.16(− 4)	2.07(− 4)	3.30(− 4)	2.64(− 4)	25.0
8	$7P_{1/2}$	$6S_{1/2}$	7	1	459.3	2.96(− 5)	6.15(− 5)	1.10(− 4)	1.75(− 4)	1.37(− 4)	27.7
9	$10D_{3/2}$	$6P_{1/2}$	40	2	546.6	1.19(− 6)	3.30(− 6)	7.38(− 6)	1.41(− 5)	1.40(− 5)	0.7
10	$10D_{5/2}$	$6P_{3/2}$	41	3	563.5	2.09(− 6)	5.81(− 6)	1.30(− 5)	2.49(− 5)	2.47(− 5)	0.8
11	$10D_{3/2}$	$6P_{3/2}$	40	3	563.7	2.21(− 7)	6.15(− 7)	1.38(− 6)	2.63(− 6)	2.61(− 6)	0.8
12	$9D_{3/2}$	$6P_{1/2}$	34	2	566.4	2.11(− 6)	5.72(− 6)	1.25(− 5)	2.36(− 5)	2.32(− 5)	1.7
13	$10S_{1/2}$	$6P_{1/2}$	27	2	583.9	4.25(− 7)	1.15(− 6)	2.52(− 6)	4.73(− 6)	3.94(− 6)	20.1
14	$9D_{5/2}$	$6P_{3/2}$	35	3	584.5	3.65(− 6)	9.86(− 6)	2.17(− 5)	4.08(− 5)	4.02(− 5)	1.5
15	$9D_{3/2}$	$6P_{3/2}$	34	3	584.8	3.94(− 7)	1.07(− 6)	2.33(− 6)	4.40(− 6)	4.33(− 6)	1.6
16	$8D_{3/2}$	$6P_{1/2}$	25	2	601.0	4.66(− 6)	1.21(− 5)	2.57(− 5)	4.70(− 5)	4.34(− 5)	8.3
17	$10S_{1/2}$	$6P_{3/2}$	27	3	603.4	7.92(− 7)	2.14(− 6)	4.70(− 6)	8.81(− 6)	7.33(− 6)	20.2
18	$8D_{5/2}$	$6P_{3/2}$	26	3	621.3	8.07(− 6)	2.10(− 5)	4.46(− 5)	8.17(− 5)	7.55(− 5)	8.2
19	$8D_{3/2}$	$6P_{3/2}$	25	3	621.8	8.65(− 7)	2.24(− 6)	4.76(− 6)	8.72(− 6)	8.05(− 6)	8.3
20	$9S_{1/2}$	$P_{1/2}$	18	2	635.5	1.10(− 6)	2.80(− 6)	5.89(− 6)	1.07(− 5)	7.99(− 6)	33.9
21	$9S_{1/2}$	$P_{3/2}$	18	3	658.7	2.02(− 6)	5.17(− 6)	1.09(− 5)	1.98(− 5)	1.48(− 5)	33.8
22	$7D_{3/2}$	$6P_{1/2}$	16	2	672.3	1.36(− 5)	3.29(− 5)	6.61(− 5)	1.16(− 4)	1.02(− 4)	13.7
23	$7D_{5/2}$	$6P_{3/2}$	17	3	697.3	2.38(− 5)	5.76(− 5)	1.16(− 4)	2.03(− 4)	1.78(− 4)	14.0
24	$7D_{3/2}$	$6P_{3/2}$	16	3	698.3	2.49(− 6)	6.05(− 6)	1.21(− 5)	2.13(− 5)	1.87(− 5)	13.9
25	$8S_{1/2}$	$6P_{1/2}$	11	2	760.9	4.55(− 6)	1.05(− 5)	2.04(− 5)	3.50(− 5)	2.34(− 5)	49.5
26	$8S_{1/2}$	$6P_{3/2}$	11	3	794.4	8.29(− 6)	1.91(− 5)	3.72(− 5)	6.37(− 5)	4.25(− 5)	49.9
27	$6P_{3/2}$	$6S_{1/2}$	3	1	852.1	9.18(− 3)	1.41(− 2)	1.98(− 2)	2.63(− 2)	2.06(− 2)	27.7
28	$6D_{3/2}$	$6P_{1/2}$	9	2	876.1	6.49(− 5)	1.37(− 4)	2.47(− 4)	3.98(− 4)	3.42(− 4)	16.4
29	$6P_{1/2}$	$6S_{1/2}$	2	1	894.3	4.47(− 3)	6.82(− 3)	9.63(− 3)	1.28(− 2)	1.00(− 2)	28.0
30	$6D_{5/2}$	$6P_{3/2}$	10	3	917.2	1.16(− 4)	2.44(− 4)	4.41(− 4)	7.09(− 4)	6.16(− 4)	15.1
31	$6D_{3/2}$	$6P_{3/2}$	9	3	920.8	1.16(− 5)	2.46(− 5)	4.41(− 5)	7.11(− 5)	6.13(− 5)	16.0

Progress Report for 1991 Summer Faculty Research
(5/20/91 - 7/26/91)

A Numerical Method for Time-Dependent Incompressible and Compressible Navier-Stokes Flows

Summer Faculty Member:
Tien-Mo Shih (emplyee number #132)
Department of Mechanical Engineering
The University of Maryland
College Park, MD 20742

Air Force Focal Point:
Abdollah Nejad
WL/POPT
Wright-Patterson AFB, OH 45433

ABSTRACT

During the 10-week stay at the WPAFB, several tasks have been finished:

1. The AMDEK computer code was studied, run, and shortened cosmetically. It was further found that the code cannot converge as is. See Appendix A.
2. The subject of developing a numerical scheme capable of handling both the incompressible flow regime and the compressible regime was explored. See the main text.
3. A one-dimensional code was written and applied to computing isentropic flows in convergent-divergent nozzles. See Appendix B for the FORTRAN listing.
4. A two-dimensional code was written and applied to computing flows in planar combustors. See Appendix C for the FORTRAN listing.

1. Introduction

In the area of Computational Fluid Dynamics (CFD), several difficult subjects are involved in the computations of practical problems in both incompressible and compressible flow regimes. These, in the authors' opinion, include: (1) pressure and density link through the continuity equation and the equation of state, (2) diagonal dominance of matrices for high-speed flows, (3) grid staggeredness, (4) accuracy associated with steep gradients in boundary layers or shock waves, (4) irregular geometries, (5) three dimensional zonal matching, and (6) improvement of convergence rates.

In this report, we will concentrate on subject (1). Conventionally, the continuity equation is designated to solve for the density in the compressible flow regime (CFR), and to solve for the pressure in the incompressible flow regime (IFR). If a flow problem enters both IFR and CFR, naturally it is desirable to use the continuity equation to solve for the same variable throughtout the entire domain. However, if the density is chosen and solved using the continuity equation in IFR, slight error in the density will lead to large errors in the pressure and subsequently will lead to grossly inaccurate solutions in the momentum equations. On the other hand, if the pressure is selected and solved using the continuity equation in CFR, errors in the pressure will also lead to errors in the density. The later does not appear in IFR, but will appear in CFR making the solution inaccurate.

Some numerical schemes for IFR computations are designed to eliminate the pressure, such as streamfunction-vorticity [1,2], biharmonic functions [3,4], vorticity-velocity [5,6].

In IFR, most researchers who adopt the primitive formulation use pressure Poisson equation [7,8], penalty function method [9,10], artificial compressibility scheme [11-13], and pressure correction algorithm [14-17]. Two noteworthy points here are:

(1) The pressue, being computed through the continuity equation [11-13] or combined momentum equations and the continuity equation [14-17], is not used to find the density. The density is either constant (hence does not enter the problem) or is found by the isobaric relation,

$$T\rho = c_1. \tag{1}$$

In this manner, p and ρ are dissociated completely in IFR.

(2) In the continuum governing equations in IFR, only pressure gradients appear, not the absolute values of the pressure. Consequently, the pressure values can be corrected freely, as long as the nodal gradients remain the same.

In CFR, the continuity equation is used to govern the density [18-22]. The pressure is then computed through the equation of state.

Most numerical methods valid in IFR fail to work in CFR. Since the continuity equation is reserved for the relative pressure, the absolute pressure is recovered from the boundary condition. The density is then computed through the equation of state, leading to instability because of absence of the spatial linkage with neighboring ρ. Furthermore, physically, the continuity equation is derived under the law of mass conservation, and therefore should be reserved for computation of ρ.

The following paragraphs will be verified and revised in the research proposal:

Verify the capabilities of incompressible-flow schemes which have entered CFR [23-27].

Verify the capabilities of compressible-flow schemes which have entered IFR [11,28-33],

It appears that methods available in the literature rarely are applicable to both IFR and CFR. For example, a method is used to solve a compressible flow problem. Upon completion of the task, the method is then modified to solve an incompressible flow problem. We have not found a method applied to a flow problem covering a wide range from, say, M=.01 to M=1.2 in the free stream. It is also worth noting that a scheme can be claimed to cover both IFR and CFR regimes only when the freestream velocity (for external flows) or the centerline velocity (for internal flows) varies appreciably from, say, 1 m/s to 100 m/ s. Under such conditions, the magnitude of the absolute static pressure becomes comparable with the change of the static pressure, i.e.,

$$p \sim \delta p.$$

A scheme should not be considered capable of solving both IFR and CFR simply because the Mach number distribution covers from nearly zero near the wall and 1.2 at the free stream (such as the boundary-layer flows). The reason is that the flow velocity changes under the influence of the viscosity, but not due to the "Bernoulli effect". It is seen that in the boundary layer flow,

$$p \gg \delta p.$$

We will verify this assertion. If our assertion is incorrect, we then will find out how our scheme is different from others.

2. Continuum Governing Equations

The time-dependent, compressible, nonisothermal, 2-D Navier-Stokes flows are governed by the following continuum equations in the conservative form:

the continuity equation,

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \rho \mathbf{u} = 0, \tag{2a}$$

the momentum equation,

$$\frac{\partial(\rho \mathbf{u})}{\partial t} + (\nabla \cdot \rho \mathbf{u})\mathbf{u} + \nabla p = \frac{1}{Re}\nabla^2 \mathbf{u}, \tag{2b}$$

and the energy equation,

$$\frac{\partial(\rho e_1)}{\partial t} + \gamma(\nabla \cdot \rho \mathbf{u})e = \frac{1}{Pe}\nabla^2 \theta, \tag{2c}$$

where $\mathbf{u} = (u, v)^T$, Pe is the Peclect number defined as $PrRe$, $e_1 = \theta + u^2/c_v$, and $e = \theta + u^2/c_p$.

3. Discretized Governing Equations

Figure 1 shows the 2-D grid system with nodal unknowns located at the grid points. For a typical grid point (i,j), the finite difference approximations are:

$$\partial p/\partial x = (p(i,j) - p(i-1,j))/dx,$$

$$\partial u/\partial x = (u(i,j) - u(i-1,j))/dx,$$

$$\partial v/\partial y = (v(i,j) - v(i,j-1))/dy,$$

$$\partial p/\partial y = (p(i,j) - p(i,j-1))/dy.$$

to be strengthened.

4. Basics of the Present Method

If an iterative scheme is capable of yielding a diagonally-dominant coefficient matrix, and the resulting algebraic equations are consistent (we do not mean "consistent with the differential equations", we mean "consistent algebraically"), then most likely the scheme should work. If the equations are nonlinear, they need to be linearized or under-relaxed.

For easy presentation, we will describe the basics of the present method on a one-D nonstaggered grid. Application of this concept has been extended to a 2-D nonstaggered grid. There seems to be no obvious reason that the same concept cannot be extended to the staggered grid.

* Fig. 2, steady state, isothermal, pt 1 is the left boundary.
* make the analysis fansier and more sophiscated.
* method of least squares
* algebraically consistent

to be strenghened here.

5. Computational Procedure

The point-by-point Gauss-Seidel iteration (convergence rate is not the issue here) is used. The discretized equations are written in the residual forms, which simplifies the coding greatly.

6. Test Problems

To test the validity of the proposed computer code, we have used it to compute several benchmark flow problems, and made comparisons of the result computed by us and that reported in the literature. These benchmark problems are: (a) isentropic flows in 1 one-dimensional convergent- divergent nozzle, (b) lid-driven cavity flows, and (c) 2-D combustor.

(a) 1-D converg. diverg. nozzle.
Fig. 3 shape of the 1-D CD nozzle and its grid.
Fig. 4a pressure distribution vs. x compared with the literature table (no shock)
Fig. 4b. with shock

(b) lid-driven cavity flows (Re=2,000)

Fig. 5 Center-location U as function of y, compared with Harwell report.

(c) 2-D planar combustor

Fig. 6 u(y) at various locations of x.
Fig. 7 p(y) at various locations of x. compared with experimental data.

7. Results and Discussions

The limitations and merits of the computer code are summarized as follows:

Limitations:

1. The code has not been applied to flows where shock waves take place.
2. No effort has been directed towards improving the convergence rate such as implementing approximate factorization, multigrids, etc.
3. The code has not been extensively tested against various conditions or/and cases.

Merits:

1. The code is applicable to both IFR and CFR at least for the 1-D case.
2. The structure of the code is simple and the length of the code is short, making it easy for users to learn and understand.

8. Concluding Remarks

There is much room for improvement in the current version of 2-D Navier- Stokes code. However, judging from the result of 1-D isentropic-flow problems and that of planar combustor flows, we feel hopeful that both the knowledge gained so far and the codes themselves will serve as a foundation for further 2-D code development in the near future. To be precise and realistic, we expect to produce by (or prior to) May 1991 a 2-D code that is time-dependent, turbulent, applicable to both incompressible and compressible flow regimes, and applicable to irregular geometries.

9. Acknowledgement

The first author wishes to express his gratitude towards the fellowship and the opportunity offered by Wright-Patterson Air Force Base. Technical communication with Mr. Kevin Cope has proved enlightening.

10. References

1. P. J. Roache, Computational Fluid Dynamics, Hermosa, 1972.
2. T. M. Shih, A literature survey on numerical heat transfer (1988- 1989), Numerical Heat Transfer, 18, part A, 387-426 (1990).
3. biharmonic function. to be found.
4. biharmonic function. to be found.
5. vorticity-velocity. to be found.
6. vorticity-velocity. to be found.
7. F. H. Harlow and J. E. Welch, Numerical calculation of time- dependent viscous incompressible flow of fluid with free surface, Phys. Fluids, 8, 2182-2189 (1965).
8. W. R. Briley, A numerical study of time-dependent rotation flow in a cylindrical container at low and moderate Reynolds numbers, J. Compu. Phys., 14, 20-35 (1974).
9. T. J. R. Hughes, W. K. Liu, and A. Brooks, Finite element analysis of incompressible viscous flows by the penalty function formulation, J. Compu. Phys., 30, 1-60 (1979).

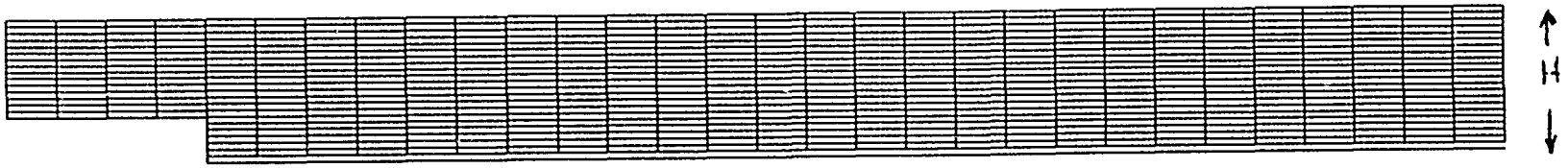
10. M. Bercovier and M. Engelman, A finite element for the numerical solution of viscous incompressible flows, J. compu. Phys., 30, 181-201 (1979).
11. A. J. Chorin, A numerical method for solving incompressible viscous flow problems, J. Compu. Phys., 2, 12-26 (1967).
12. D. Kwak, J. L. C. Chang, S. P. Shanks, and S. R. Chakravarthy, A three-dimensional incompressible Navier-Stokes flow solver using primitive variables, AIAA J., 24, 390-396 (1986).
13. D. Choi and C. L. Merkle, Application of time-iterative schemes to incompressible flow, AIAA J., 23, 1518-1524 (1985).
14. S. V. Patankar, Numerical Heat Transfer and Fluid Flow, Hemisphere, 1980.
15. TEACH
16. Raithby
17. Rhie and Chow
18. MacCormack
19. Beam and Warming
20. J. Steger
21. Harteen
22. van Lear
23. Issa and Lockwood
24. J. P. van Doormaal, G. D. Raithby, and B. H. McDonald, The segregated approach to predicting viscous compressible fluid flows, ASME J. Turbomachinery, 109, 268-277, 1987.
25. Rhie Rhie
26. Rhie Stowers
27. McGuirk Page
28. Tackel
29. Sharmath
30. Pletcher
31. Jameson
32. Steger

iter	RescM	ResuM	RestM
1	1.09329	5.16549	0.04318
2	1.19603	4.92249	0.02902
3	1.22849	4.60184	0.02340
4	1.25219	4.44695	0.02500
5	1.26836	4.23544	0.02493
1000	0.57280	0.14886	0.00232
2000	0.38836	0.29591	0.00091
3000	0.28985	0.32602	0.00061
4000	0.22560	0.28876	0.00078
5000	0.18544	0.24509	0.00093
6000	0.15717	0.21200	0.00103
7000	0.13344	0.18290	0.00109
8000	0.11271	0.15570	0.00114
9000	0.09489	0.13076	0.00117
10000	0.07978	0.10866	0.00119
11000	0.06702	0.08969	0.00118
12000	0.05633	0.07367	0.00115
13000	0.04742	0.06035	0.00110
14000	0.03995	0.04945	0.00105
15000	0.03364	0.04054	0.00099
16000	0.02837	0.03330	0.00092
17000	0.02394	0.02740	0.00085
18000	0.02023	0.02261	0.00078
19000	0.01709	0.01869	0.00072
20000	0.01446	0.01549	0.00065
21000	0.01226	0.01286	0.00059
22000	0.01040	0.01070	0.00054
23000	0.00883	0.00892	0.00048
24000	0.00750	0.00745	0.00043
25000	0.00637	0.00623	0.00039
26000	0.00542	0.00522	0.00035
27000	0.00461	0.00438	0.00031
28000	0.00393	0.00368	0.00028
29000	0.00335	0.00309	0.00025
30000	0.00286	0.00260	0.00022
31000	0.00244	0.00219	0.00019
32000	0.00208	0.00184	0.00017
33000	0.00178	0.00155	0.00015
34000	0.00153	0.00131	0.00013
35000	0.00131	0.00110	0.00012
36000	0.00112	0.00092	0.00010
37000	0.00096	0.00078	0.00009
38000	0.00083	0.00065	0.00008
39000	0.00072	0.00054	0.00007
40000	0.00062	0.00045	0.00006
41000	0.00053	0.00038	0.00005
42000	0.00046	0.00031	0.00005
43000	0.00040	0.00026	0.00004
44000	0.00035	0.00021	0.00004
45000	0.00031	0.00017	0.00003
46000	0.00027	0.00013	0.00003
47000	0.00024	0.00011	0.00003
48000	0.00021	0.00009	0.00002
49000	0.00019	0.00007	0.00002
50000	0.00017	0.00006	0.00002
51000	0.00015	0.00005	0.00002
52000	0.00014	0.00004	0.00001
53000	0.00012	0.00003	0.00001
53359	0.00012	0.00003	0.00001

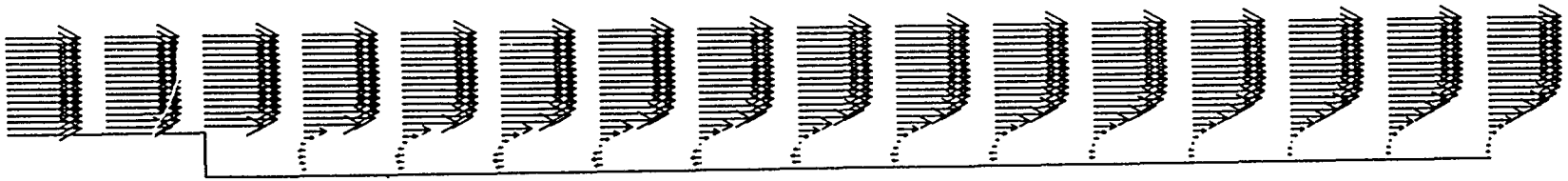
1-D isentropic flows in CD nozzle

Mach #	p	density	t	1/a	p/(ρRT)	u
0.0500	0.9983	0.9988	0.9995	0.0863	1.0000	0.2509
0.0831	0.9954	0.9968	0.9986	0.1429	1.0000	0.4166
0.1001	0.9933	0.9953	0.9980	0.1720	1.0000	0.5022
0.1125	0.9914	0.9940	0.9974	0.1930	1.0000	0.5642
0.1232	0.9897	0.9928	0.9969	0.2109	1.0000	0.6173
0.1327	0.9880	0.9916	0.9964	0.2269	1.0000	0.6648

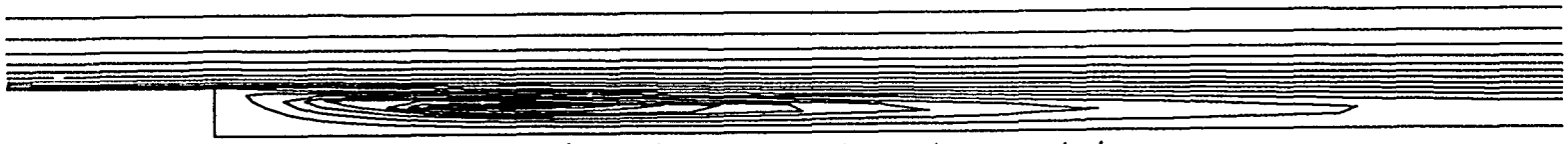
0.1415	0.9864	0.9904	0.9959	0.2417	1.0000	0.7090
0.1501	0.9846	0.9892	0.9954	0.2560	1.0000	0.7520
0.1581	0.9830	0.9880	0.9949	0.2692	1.0000	0.7918
0.1656	0.9813	0.9868	0.9944	0.2815	1.0000	0.8292
0.1730	0.9796	0.9856	0.9939	0.2935	1.0000	0.8656
0.1800	0.9779	0.9844	0.9934	0.3051	1.0000	0.9007
0.1874	0.9761	0.9830	0.9929	0.3170	1.0000	0.9374
0.1946	0.9742	0.9817	0.9923	0.3286	1.0000	0.9731
0.2018	0.9723	0.9803	0.9918	0.3403	1.0000	1.0090
0.2091	0.9702	0.9789	0.9912	0.3519	1.0000	1.0451
0.2165	0.9681	0.9774	0.9906	0.3636	1.0000	1.0815
0.2239	0.9659	0.9758	0.9899	0.3754	1.0000	1.1183
0.2314	0.9636	0.9741	0.9892	0.3872	1.0000	1.1554
0.2391	0.9612	0.9724	0.9885	0.3991	1.0000	1.1933
0.2469	0.9587	0.9706	0.9878	0.4112	1.0000	1.2317
0.2548	0.9561	0.9687	0.9870	0.4234	1.0000	1.2708
0.2630	0.9533	0.9667	0.9861	0.4359	1.0000	1.3109
0.2713	0.9504	0.9646	0.9853	0.4485	1.0000	1.3519
0.2799	0.9473	0.9623	0.9843	0.4614	1.0000	1.3940
0.2887	0.9440	0.9600	0.9834	0.4745	1.0000	1.4373
0.2979	0.9405	0.9575	0.9823	0.4879	1.0000	1.4819
0.3073	0.9368	0.9548	0.9812	0.5017	1.0000	1.5280
0.3171	0.9329	0.9519	0.9800	0.5159	1.0000	1.5759
0.3273	0.9287	0.9489	0.9788	0.5304	1.0000	1.6256
0.3380	0.9242	0.9456	0.9774	0.5455	1.0000	1.6775
0.3492	0.9193	0.9420	0.9759	0.5610	1.0000	1.7319
0.3610	0.9141	0.9382	0.9743	0.5771	1.0000	1.7890
0.3735	0.9084	0.9340	0.9726	0.5939	1.0000	1.8492
0.3868	0.9022	0.9294	0.9707	0.6114	1.0000	1.9130
0.4010	0.8954	0.9244	0.9685	0.6297	1.0000	1.9811
0.4161	0.8879	0.9189	0.9662	0.6487	1.0000	2.0532
0.4326	0.8796	0.9128	0.9636	0.6690	1.0000	2.1316
0.4505	0.8702	0.9058	0.9607	0.6904	1.0000	2.2166
0.4702	0.8597	0.8980	0.9573	0.7131	1.0000	2.3094
0.4920	0.8477	0.8890	0.9535	0.7372	1.0000	2.4117
0.5165	0.8339	0.8787	0.9490	0.7631	1.0000	2.5258
0.5443	0.8178	0.8665	0.9438	0.7908	1.0000	2.6542
0.5763	0.7987	0.8520	0.9374	0.8205	1.0000	2.8009
0.6139	0.7757	0.8344	0.9296	0.8524	1.0000	2.9713
0.6592	0.7473	0.8125	0.9197	0.8865	1.0000	3.1735
0.7152	0.7114	0.7844	0.9069	0.9221	1.0000	3.4190
0.7873	0.6646	0.7473	0.8894	0.9576	1.0000	3.7271
0.8853	0.6011	0.6956	0.8642	0.9880	1.0000	4.1312
1.0271	0.5123	0.6206	0.8255	0.9995	1.0000	4.6843
1.2009	0.4124	0.5316	0.7757	0.9710	1.0000	5.3119



(a) numerical grid (not to the true scale)
true scale: $L = 100$, $H = 1$.



(b) velocity vector plot.



(c) streamline at the steady state

2-D Navier-Stokes Flows in the Combustor

i=1	i=4	i=7	i=10	i=13	i=16	i=19	i=22	i=25	i=28
u(i,j)									
0.5004	0.5053	0.5042	0.5022	0.5009	0.4996	0.4984	0.4972	0.4960	0.4946
0.5004	0.5053	0.5042	0.5022	0.5009	0.4996	0.4984	0.4972	0.4960	0.4946
0.5004	0.5055	0.5044	0.5023	0.5009	0.4997	0.4985	0.4973	0.4960	0.4946
0.5004	0.5057	0.5046	0.5025	0.5010	0.4997	0.4985	0.4972	0.4960	0.4946
0.5004	0.5060	0.5049	0.5027	0.5011	0.4998	0.4985	0.4972	0.4958	0.4944
0.5004	0.5063	0.5052	0.5028	0.5012	0.4997	0.4983	0.4970	0.4956	0.4942
0.5004	0.5067	0.5056	0.5030	0.5012	0.4996	0.4981	0.4967	0.4952	0.4938
0.5004	0.5071	0.5059	0.5031	0.5012	0.4994	0.4978	0.4962	0.4947	0.4932
0.5004	0.5074	0.5062	0.5032	0.5011	0.4991	0.4973	0.4956	0.4939	0.4923
0.5004	0.5078	0.5065	0.5033	0.5008	0.4986	0.4965	0.4946	0.4927	0.4909
0.5004	0.5081	0.5067	0.5030	0.5002	0.4975	0.4950	0.4925	0.4902	0.4879
0.5004	0.5080	0.5060	0.5016	0.4979	0.4942	0.4905	0.4868	0.4832	0.4796
0.5004	0.5059	0.5012	0.4946	0.4884	0.4820	0.4755	0.4691	0.4628	0.4568
0.5004	0.4926	0.4777	0.4650	0.4529	0.4413	0.4305	0.4206	0.4118	0.4041
0.5004	0.4220	0.3830	0.3652	0.3522	0.3422	0.3343	0.3280	0.3230	0.3192
0.0000	0.0000	0.1451	0.1897	0.2083	0.2180	0.2235	0.2269	0.2293	0.2314
0.0000	0.0000	0.0607	0.1041	0.1227	0.1332	0.1404	0.1460	0.1509	0.1558
0.0000	0.0000	-0.0200	0.0296	0.0481	0.0596	0.0686	0.0764	0.0836	0.0906
0.0000	0.0000	-0.0354	-0.0274	-0.0103	0.0018	0.0120	0.0214	0.0302	0.0385
0.0000	0.0000	-0.0352	-0.0514	-0.0477	-0.0374	-0.0270	-0.0169	-0.0075	0.0013
0.0000	0.0000	-0.0270	-0.0546	-0.0597	-0.0543	-0.0455	-0.0361	-0.0274	-0.0196
0.0000	0.0000	-0.0139	-0.0391	-0.0452	-0.0441	-0.0390	-0.0327	-0.0265	-0.0213
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

u at the inlet = .5004, jstp=8, and istp=5.

U distribution in the 2-D combustor

Appendix A

Communication with Amtek and the Status of Debugging the Dumpster Code

1. I have studied carefully their report, which is technically sound.

2. I have obtained the printout of the FORTRAN listing and have studied it carefully.

3. The Dumpster code was run at the University of Maryland. It ran up to 10,000 time steps. The maximum continuity residual decreased from 8 to 6 in log scale.

4. Regarding to 3., Mr. Peery suggested that the code be run at higher number of time steps. Some knowledge is required to stop at somewhere around 5000th time step to manually change the flux index. If not, the convergence rate will be extremely slow.

5. The Dumpster code was run on Pohost VAX machine without changing a single statement. The machine stopped at 40 some time steps because of either overflow or underflow.

6. Both 4. and 5. are for zero swirl case.

7. Mr. Peery agreed early in June to send POP™ two versions: one can be run as is and converge on POPT's computers, and the other is to run a benchmark problem of the 1-D CD nozzle. He was made aware that my stay is only very brief, so he agreed to send them within two weeks.

8. When the day arrived, he called to tell us that he was not ready. Kevin answered the call, I did not call Mr. Peery back.

Appendix B 1-D Isentropic Flows in CD Nozzle

```
      Program CDnozzle
      dimension r(105),u(105),t(105),p(105),a(105),c(105)
     &, resc(105),rsum(105),rese(105),eqs(105),avg(105),aMach(101)
     &, pdp(105),ada(105)
      nter=70000
      kd=1000
      nx=50
      nx1=nx+1
      dx=1./nx
      dx2=dx*2
      ithro=nx1-5
      ithrm=ithro-1
c...under-relaxation parameters.
      gar=.3
      gau=.3
      gat=.3
      gap=.3
      convc=.0001
      convu=.0001
      convt=.001
c...set Mach=.05 at the inlet.
      aMach(1)=.05
      pdp(1)=.9983
      rdri=.9988
      tdti=.9995
      ada(1)=.08627
c.....................Mach=.3  entering the compressible regime.
c.....................Mach=1.  entering supersonic regime.
      ada(ithro)=1.
c...Mach=1.2
      aMach(nx1)=1.2
      pdp(nx1)=.4124
      rdrf=.5311
      tdtf=.7764
      ada(nx1)=.9705
c...relevant thermal properties.  Reference state is at Mach=.2.
      Tref=.9921*300.
      gamma=1.4
      ui=.2*sqrt(gamma*287.*Tref)
      uis=ui*ui
      rs=287*300./uis
      cv=rs/.4
      cp=cv+rs
c...setup the boundary conditions.
      a(1)=1./ada(1)
      r(1)=a(1)*rdri
      t(1)=tdti
      p(1)=rs*r(1)*t(1)/a(1)
      c(1)=sqrt(rs*1.4*t(1))
      u(1)=aMach(1)*c(1)
      a(ithro)=1.
c...Eventually, all but p(nx1) will be overwritten and computed.
      a(nx1)=1./ada(nx1)
      r(nx1)=a(nx1)*rdrf
      t(nx1)=tdtf
                                  p(nx1)=pdp(nx1)*rs
      c(nx1)=sqrt(rs*1.4*t(nx1))
      u(nx1)=aMach(nx1)*c(nx1)
c...setup the initial guess and define the nozzle area.
      do 935 i=2,nx
      u(i)=u(1)+(i-1)*(u(nx1)-u(1))/nx
      r(i)=r(1)+(i-1)*(r(nx1)-r(1))/nx
      t(i)=t(1)+(i-1)*(t(nx1)-t(1))/nx
      p(i)=p(1)+(i-1)*(p(nx1)-p(1))/nx
      if(i.le.ithro)a(i)=a(1)+(i-1)*(1.-a(1))/ithrm
      if(i.gt.ithro)a(i)=1.+(i-ithro)*(a(nx1)-1.)/(nx1-ithro
```

```
935   continue
c....
      do 2 iter=1,nter
      if(iter.eq.3)then
      epsc=convc*rescM
      epsu=convu*rsumM
      epst=convt*reseM
      endif
      rescM=0
      rsumM=0
      reseM=0
c...compute the density.
      do 490 i=1,nx
      ip=i+1
      resc(i)=r(ip)*u(ip)-r(i)*u(i)
      rescM=amax1(rescM,abs(resc(i)))
490   continue
      do 4590 i=2,nx
      resr=(resc(i-1)-resc(i))/u(i)-eqs(i)*a(i)/(rs*t(i))
      r(i)=r(i)-gar*resr/3
4590   continue
      r(nx1)=r(nx1)-gar*(resc(nx)/u(nx1)-eqs(nx1)*a(nx1)/(rs*t(nx1)))/2
c...compute u.
      do 35 i=1,nx
      ip=i+1
      avg(i)=.5*(a(i)+a(ip))
      rsum(i)=r(ip)*u(ip)*u(ip)-r(i)*u(i)*u(i)+avg(i)*(p(ip)-p(i))
      rsumM=amax1(rsumM,abs(rsum(i)))
35    continue
      do 550 i=2,nx
      u(i)=u(i)-gau*(rsum(i-1)-rsum(i))/(2*r(i)*u(i))
550   continue
      u(nx1)=u(nx1)-gau*rsum(nx)/(r(nx1)*u(nx1))
c...compute temperature.
      do 45 i=1,nx
      ip=i+1
      rese(i)=t(ip)-t(i)+.5*(u(ip)*u(ip)-u(i)*u(i))/cp
      reseM=amax1(reseM,abs(rese(i)))
45    continue
      do 650 i=2,nx
      t(i)=t(i)-gat*(rese(i-1)-rese(i))/2
650   continue
      t(nx1)=t(nx1)-gat*rese(nx)
c...compute pressure.
      do 655 i=2,nx
      im=i-1
      resp=u(i)*(resc(i)-resc(im))+rsum(im)-rsum(i)
      p(i)=p(i)-gap*resp/(avg(im)+avg(i))
655   continue
c...compute the nozzle area.
      do 665 i=2,nx
      density=p(i)/(rs*t(i))
      a(i)=r(i)/density
      c(i)=sqrt(rs*1.4*t(i))
665   continue
      density=p(nx1)/(rs*t(nx1))
      a(nx1)=r(nx1)/density
c...print out.
      if(iter.le.5.or.iter.eq.iter/kd*kd)
     &  print 102,iter,rescM,rsumM,reseM
      if(rescM.le.epsc.and.resumM.le.epsu.and.reseM.le.epst)goto 333
2     continue
c...finish the iteration loop.
333   print 102,iter,rescM,rsumM,reseM
      do 545 i=1,nx1
      aMach(i)=u(i)/c(i)
```

```
545   continue
      print *,'   Mach #      p      density       t
     &1/a    p/(ro)RT    u'
      do 5695 i=1,nx1
      print 101,aMach(i),p(i)/rs,r(i)/a(i),t(i),1/a(i),p(i)*a(i)
     &  /(rs*r(i)*t(i)),u(i)
5695  continue
101   format(8f10.4)
102   format(I5,5f12.5)
      end
```

Appendix C 2-D Combustor Flows

```
      program NS2D1
      dimension u(51,51),up(51,51),v(51,51),vp(51,51)
     &, r(51,51),rp(51,51),t(51,51),tp(51,51),p(51,51)
     &, rsum(51,51),rsvm(51,51),resc(51,51),eqs(51,51)
     &, cfu(51,51),pp(51), py(51,51)
      nx=30
      ny=22
      nxp=nx+1
      nyp=ny+1
      nxm=nx-1
      nym=ny-1
      dx=100./nx
      dy=1./ny
      istp=5
      jstp=8
      istq=istp+1
      jstq=jstp+1
c...
      gar=.4
      gau=.4
      gav=.5
      gpx=.4
c...For time accurate computations, set ntime=6, dt=7*dx, and nter=200.
      nter=2000
      kd=20
      ntime=1
      dt=100000*dx
c...set the reference state at Mach=.1.
      aMar=.1
      tdtr=.9980
      ur=aMar*sqrt(1.4*287*tdtr*300.)
      rs=287*300./ur**2
      re=ur*.02/1.5e-5
c...set the inlet condition according to the isentropic nozzle table.
      aMai=.05
      pdpi=.9983
      rdri=.9988
      tdti=.9995
c...define auxiliary or convenient variables.
      inc=nx/10
      dx2=dx*2
      dy2=dy*2
      rdx=re*dx
      rdy=re*dy
      cofd=2/dx/rdx+2/dy/rdy
c...specify the inlet condition.
      do 205 j=1,nyp
      if(j.gt.jstp) u(1,j)=aMai*sqrt(1.4*287*tdti*300)/ur
      r(1,j)=rdri
      p(1,j)=pdpi*rs
      t(1,j)=tdti
205   continue
c...guess the initial field distributions.
      do 310 i=2,nxp
      do 310 j=1,nyp
      if(j.gt.jstp) u(i,j)=u(1,j)+(i-1)*(.3-u(1,j))/nx
      if(j.lt.jstp.and.i.lt.istp)goto 310
      r(i,j)=r(1,j)
      p(i,j)=p(1,j)
      t(i,j)=t(1,j)
310   continue
      pref=p(1,nyp)
      do 320 i=1,nxp
      do 320 j=1,nyp
      up(i,j)=u(i,j)
      vp(i,j)=v(i,j)
```

```
      rp(i,j)=r(i,j)
      tp(i,j)=t(i,j)
      p(i,j)=p(i,j)-pref
320   continue
      epsc=.0002
      epsu=.0005
      epsv=.0001
      uflow1=r(1,ny)*u(1,ny)*dy
c...start the computation here.
      do 1 itime=1,ntime
      print *,'itime=',itime
      do 2 iter=1,nter
      rescM=0
      resuM=0
      resvM=0
      eqsM=0
c...specify some boundary conditions.
      do 395 j=2,nyp
      u(nxp,j)=u(nx,j)
      v(nxp,j)=v(nx,j)
      if(j.ge.jstp)p(1,j)=2*p(2,j)-p(3,j)
      if(j.lt.jstp)p(istp,j)=2*p(istp+1,j)-p(istp+2,j)
395   continue
      do 30 i=2,nx
      im=i-1
      ip=i+1
      if(i.le.istp)then
      p(i,jstp)=p(i,jstq)-py(i,jstq)*dy
      r(i,jstp)=(p(i,jstp)+pref)/(rs*t(i,jstp))
      else
      p(i,1)=p(i,2)-py(i,2)*dy
      r(i,1)=(p(i,1)+pref)/(rs*t(i,1))
      endif
      p(i,nyp)=p(i,ny)
      r(i,nyp)=(p(i,nyp)+pref)/(rs*t(i,nyp))
      u(i,nyp)=u(i,ny)
      resc(i,nyp)=(-r(im,nyp)*u(im,nyp)+r(i,nyp)*u(i,nyp))/dx
     & +(-r(i,ny)*v(i,ny))/dy
      do 20 j=2,ny
      jm=j-1
      if(i.le.istp.and.j.le.jstp)goto 20
      jp=j+1
c...compute the fluxes and residuals.
      rcnx=(-r(im,j)*u(im,j)+r(i,j)*u(i,j))/dx
      rcny=(-r(i,jm)*v(i,jm)+r(i,j)*v(i,j))/dy
c...diffusion fluxes.
      udw=(-u(im,j)+ u(i,j))/dx
      vdw=(-v(im,j)+ v(i,j))/dx
      ude=(-u(i,j)+ u(ip,j))/dx
      vde=(-v(i,j)+ v(ip,j))/dx
      uds=(-u(i,jm)+u(i,j))/dy
      vds=(-v(i,jm)+v(i,j))/dy
      udn=(-u(i,j)+ u(i,jp))/dy
      vdn=(-v(i,j)+ v(i,jp))/dy
c...convection fluxes.
      if(u(i,j).ge.0.)then
      ucnx=(-r(im,j)*u(im,j)*u(im,j)+r(i,j)*u(i,j)*u(i,j))/dx
      vcnx=(-r(im,j)*u(im,j)*v(im,j)+r(i,j)*u(i,j)*v(i,j))/dx
      else
      ucnx=(-r(i,j)*u(i,j)*u(i,j)+r(ip,j)*u(ip,j)*u(ip,j))/dx
      vcnx=(-r(i,j)*u(i,j)*v(i,j)+r(ip,j)*u(ip,j)*v(ip,j))/dx
      endif
      if(v(i,j).ge.0.)then
      ucny=(-r(i,jm)*v(i,jm)*u(i,jm)+r(i,j)*v(i,j)*u(i,j))/dy
      vcny=(-r(i,jm)*v(i,jm)*v(i,jm)+r(i,j)*v(i,j)*v(i,j))/dy
      else
```

```
      ucny=(-r(i,j)*v(i,j)*u(i,j)+r(i,jp)*v(i,jp)*u(i,jp))/dy
      vcny=(-r(i,j)*v(i,j)*v(i,j)+r(i,jp)*v(i,jp)*v(i,jp))/dy
      endif
c...
      dpdx=(-p(im,j)+p(i,j))/dx
      py(i,j)=py(i,j)-gav*rsvm(i,j)
      tir=(r(i,j)-rp(i,j))/dt
      tiu=(r(i,j)*u(i,j)-rp(i,j)*up(i,j))/dt
      tiv=(r(i,j)*v(i,j)-rp(i,j)*vp(i,j))/dt
      resc(i,j)=tir+rcnx+rcny
      rsum(i,j)=tiu-(-udw+ude)/rdx-(-uds+udn)/rdy+ucnx+ucny+dpdx
      rsvm(i,j)=tiv-(-vdw+vde)/rdx-(-vds+vdn)/rdy+vcnx+vcny+py(i,j)
      eqs(i,j)=r(i,j)-(p(i,j)+pref)/(rs*t(i,j))
c...
      abu=abs(u(i,j))
      abv=abs(v(i,j))
      uxpvy=1/dt+u(i,j)/dx+v(i,j)/dx
      cfrl=uxpvy*uxpvy+1
      cfu(i,j)=amax1(r(i,j)/dt+cofd+r(i,j)*(abu/dx+abv/dy),.1)
      cfv=2*r(i,j)/dy
      cful=(r(i,j)/dx)**2+cfu(i,j)**2
      resr=uxpvy*resc(i,j)+eqs(i,j)
      resu=r(i,j)/dx*resc(i,j)+cfu(i,j)*rsum(i,j)
c...
      p(i,j)=p(i,jp)-py(i,jp)*dy
      r(i,j)=r(i,j)-gar*resr/cfrl
      u(i,j)=u(i,j)-gau*resu/cful
      v(i,j)=v(i,j)-gav*(resc(i,j)-resc(i,jp))/cfv
c...
      rescM=amax1(rescM,abs(resc(i,j)))
      resuM=amax1(resuM,abs(rsum(i,j)))
      resvM=amax1(resvM,abs(rsvm(i,j)))
      eqsM =amax1(eqsM ,abs(eqs(i,j)))
20    continue
30    continue
c...compute the pressure gradient correction.
      vout=0
      drdt=0
      do 345 i=2,nx
      vout=vout+(-r(i,nym)*v(i,nym)+r(i,ny)*v(i,ny))*dx
      drdt=drdt+(r(i,j)-rp(i,j))/dt*dx*dy
      uflow=r(i,ny)*u(i,ny)*dy
      def=gpx*(uflow1-uflow-vout-drdt)/dy
      ppx=-cfu(i,ny)*def
      pp(i)=ppx*dx+pp(i-1)
      p(i,ny)=p(i,ny)+pp(i)
345   continue
      if(iter.le.3.or.iter.eq.iter/kd*kd)
     & print 101,iter,rescM,resuM,resvM*10,eqsM,def
      if(rescM.le.epsc.and.resuM.le.epsu.and.resvM.le.epsv)
     & goto 333
2     continue
333   print 101,iter,rescM,resuM,resvM*10,eqsM,def
      do 70 i=1,nxp
      do 70 j=1,nyp
      if(i.lt.istp.and.j.lt.jstp)goto 70
      rp(i,j)=r(i,j)
      up(i,j)=u(i,j)
      vp(i,j)=v(i,j)
      tp(i,j)=t(i,j)
70    continue
1     continue
      write(13,*)'ro'
      do 354 j=ny,1,-1
354   write(13,102)(r(i,j),i=1,nx,inc)
      print *,'u'
```

```
      write(13,*)'u'
      do 355 j=nyp,1,-1
      write(13,102)(u(i,j),i=1,nx,inc)
355   print 102,(u(i,j),i=1,nx,inc)
      write(13,*)'p/rs'
      do 375 j=nyp,1,-1
375   write(13,102)((p(i,j)+pref)/rs,i=1,nx,inc)
      write(13,*)'v'
      do 385 j=nyp,1,-1
      write(13,102)(v(i,j),i=1,nx,inc)
385   continue
      write(13,*)'resc'
      do 3905 j=nyp,2,-1
      write(13,102)(resc(i,j),i=1,nx,inc)
3905  continue
      write(13,*)'rsum'
      do 3906 j=nyp,2,-1
3906  write(13,102)(rsum(i,j),i=1,nx,inc)
101   format(I5,5f13.5)
102   format(10f8.4)
      end
```

Experimental and Analytical Investigation of Effects of Noncondensible Gases On On-Axis Rotating Heat Pipes

Kaveh A. Tagavi & Bryan T. Martin
Department of Mechanical Engineering
University of Kentucky
Lexington, Kentucky

Air Force Mentors:

Kirk L. Yerkes & Won S. Chang
Wright Laboratory
Dayton, Ohio

ABSTRACT

During the ten week summer appointment, we conducted an investigation of the effects of noncondensible gases on the performance of on-axis rotating heat pipes. Our efforts were divided in two fronts: experimental and analytical.

On the experimental front we designed and manufactured a rotating heat pipe made principally of oxygen-free-hard-copper with water as the working fluid. The assembly was isolated from the working table and electrical motor by using two air bearings. To study the effects of noncondensible gases some predetermined amount of nitrogen will be added to the working liquid. The heat pipe's inner wall was tapered at about two degree slope so that the centrifugal force would help the transfer of the working liquid from the condenser end to the evaporator end. Several thermocouples were installed both throughout the vapor passage and at the inner and outer sides of the heat pipe's wall. This will facilitate the measurement of temperatures of the heat pipe wall and the vapor inside. Accordingly, heat flux through the wall and the amount

of noncondensible gases may be calculated from the temperature measurements. The thermocouples are connected to a data acquisition system by a slip-ring. The condenser end, which is covered by radial fins, is cooled by the aid of a Vortex tube while the evaporator end is heated radiantly via a set of element heaters.

Although no systematic set of experiments was performed during this period, several sets of experiments are planned for the follow-up study during the next academic year.

On the analytical front, a dual approach was adopted. The vapor region, including the effect of noncondensible gases, was solved in an exact manner with a minimal number of simplifying assumptions. The liquid film, consisting of the working liquid in the condenser and evaporative regions, was analyzed in a manner similar to pervious reports published by Marto et al. Then these two solutions were coupled through the interface condition. Of course, the coupling cannot be satisfied analytically in a way that would result in closed form solutions; rather, the coupling act has to be carried numerically. Again, the numerical solution of the analytical approach is planned for the upcoming academic year.

Finally, it is planned that the experimental and analytical results be compared with each other and with those available in the literature.

NOMENCLATURE

a	aspect ratio, L/r_1
C	total moles
c	molar concentration
D	diffusion coefficient
h	heat transfer coefficient in surrounding environment
h_{fg}	latent heat of vaporization
I_0	Bessel function of the first kind and zero order
I_1	Bessel function of the first kind and first order

K	property group, $(3k\Delta T\mu)/(h_{fg}\rho^2 r_1 \omega^2)$
k	liquid film thermal conductivity
L	axial length of heat pipe
N	molar flux
n	any integer
P_0	constant determining pressure field
p	pressure
R	dimensionless radial coordinate, r/r_1
R_c	universal gas constant
r	radial coordinate
r_1	inner radius of the heat pipe
T	temperature
T_i	interface temperature between liquid film and gas/vapor region
T_∞	temperature of surroundings
t	time coordinate
U	liquid velocity at condensation and evaporative interfaces
u	dimensionless liquid film velocity, Ur_1/D
V	molar velocity
x	molar fraction
y	$\ln(x_2)$
Z	dimensionless axial coordinate, z/L
z	axial coordinate

Greek Letters

β	variable, δ^4
δ	liquid film thickness
δ_0	liquid film thickness at z=0
λ	eigen value
μ	viscosity
ξ	variable, $d\beta/dz$

ρ mass density

ϕ angle of tapered inner wall

ω angular velocity

Subscripts and Superscripts

1 vapor

2 noncondensible gas

c condenser section

e evaporator section

i interface

∞ surrounding environment

INTRODUCTION

A rotating heat pipe (thermosyphon) utilizes the gravity field generated by rotation as a driving force for the movement of the working fluid from the condenser end to the evaporator end. This movement is due to the difference in hydrostatic pressure between the condenser and the evaporator (liquid film in the condenser side is generally thicker than that in the evaporator side). The difference in the hydrostatic pressure can further be enhanced by tapering the inside of the heat pipe.

Noncondensible gases exist in heat pipes either intentionally or unintentionally. Gases are generated in most heat pipes due to chemical processes such as corrosion. Noncondensible gases are also introduced in the heat pipe as a means of controlling the heat load ("Gas-Loaded Heat Pipes"). At any rate, even a small amount of noncondensible gases could affect the operation of the heat pipe considerably. As the working vapor moves from the evaporator end to the condenser end it also carries the noncondensible gases. Unlike the vapor, the noncondensible gases cannot condense and therefore will pile up at the condensation front. Thereafter, the pile-up of noncondensible gases will hinder the condensation process. For these reasons, it is

desirable to understand the behavior of the noncondensible gases, and how and where these gases build up (find the gas' distribution).

BACKGROUND

Since its invention in 1944, the heat pipe has been used in many situations that require transfer of heat with a minimum drop in temperature. A special king of heat pipe (i.e. rotating heat pipe) is one that utilizes rotation to transport the working liquid from the condenser end to the evaporator end. An extensive survey of applications of rotating heat pipes is icluded in the paper by Yerkes (1990). Early on, it was realized that a greater thermal control of the heat pipe can be achieved by adding some predetermined amount of noncondensible gases. These gases will pile up in the condenser section, thereby reducing the effective area available for condensation. Gas-loaded heat pipes have been the subject of many studies. The simplest treatment of the noncondensible gases is reported by Dunn and Reay (1982). It simply assumes that there is a sharp wall separating the noncondensible gases from the vapor. This resulted in a quick calculation of the temperature profile along the heat pipe. Next, Edwards and Marcus (1972) treated the problem as one-dimensional and obtained the axial dependence of the noncondensible gas distribution. Later, Peterson and Tien (1989) used an integral technique to obtain an approximate, yet two-dimensional, profile for the noncondensible gas distribution. This was followed by Chang and Yu (1990) who showed how to get yet higher degree solutions.

Recently some attention has been devoted to rotating heat pipes. The idea is that the rotation can be used to transport the fluid from the condenser side to the evaporator side. This is achieved by tapering the inner wall. The liquid condensate flows for two reasons: hydrostatic pressure induced by rotation and the fact that the inner wall is tapered. Marto (1973) set out the foundation of treating the liquid flow in rotating heat pipes. He used the technique introduced by Leppert and Nimmo (1968). They dealt with laminar film condensation on horizontal surfaces where the only driving force is the hydrodynamic pressure differences.

Daniels and Al-Jumaily (1975) used a Nusselt-type approach for modeling the liquid film in a rotating heat pipe with no noncondensible gases. Later, Daniels and Williams (1978)

Baharnah (1980) studied rotating heat pipes with noncondensible gases. In their work they limited the effect to that of higher pressure due to the added gas and did not offer a rigorous analysis of the gas concentration profile.

EXPERIMENTAL EFFORTS

The most effective way to define the noncondensible gas front is to find the distribution of temperatures within the heat pipe. The reason for this is that as the working vapor becomes mixed with the noncondensible gases (in increasing concentrations toward the condenser end) the partial pressure of the working vapor will decrease, therefore changing the saturation temperature of the vapor. If we know how the temperature varies within the heat pipe, we will know how the partial pressure of the vapor changes. From this, we can then extrapolate the concentration of the noncondensible gas at various positions within the heat pipe. The experimental apparatus was designed with the use of this method in mind.

It is important, however, to keep in mind the effects of hydrostatic pressure, due to rotation, on the vapor pressure. In order to get some idea of the importance of this factor, a short analysis was done to approximate the increase in hydrostatic pressure at any given radius from the canter of the heat pipe. The relation was found to be $p(\omega)/p(\omega=0)=exp[(\omega^2 r_1^2)/(2R_cT)]$ using ideal gas assumptions, which were substantiated by the values this relation gives. Figure 1 shows the results of this relation for water vapor. As can be seen, this effect is not very large, even at very high rotational speeds and large radius. With this knowledge, we were satisfied that the saturation temperature of the vapor would be within reasonable limits. This gives us a means of correcting our data for the effects of rotation. Figure 2 is a plot of how this pressure difference will effect the temperature difference. It was composed with data from standard thermodynamic tables.

The test rig which we will use is needed by other researchers, so our apparatus had to be designed in modular form so it could be easily installed or removed when the time came for data collection. The test module is built around a foot-long thermosyphon cylinder. Figure 3a depicts the heat pipe while Figure 3b shows the overall test assembly. This cylinder has a two degree

internal taper to drive the working liquid, which is methanol. The condenser end has radial external fins to increase the wattage the condenser can handle. We decided against axial fins for several important reasons, which will not be detailed here. There are also several slots and holes machined into the exterior of the cylinder; these will accommodate thermocouple wires.

Each end has a series of flanges to facilitate easy dismantling for such things as refilling, cleaning, and servicing internal components. These are parts 2-7. Many are sealed with O-rings. Parts 2 and 4 have axial grooves, which are part of the thermocouple wire system. Part 6 is unique, with a diametral groove to channel the wires from the exterior of the module to the interior of the shaft, to be taken out of the module. Part 8 is the bearing interface, and is the piece which allows the apparatus to be truly modular.

Parts 9-11 connect together inside of the chamber to lend structural support to the "thermocouple tree" (part 11). Part 10 is the fill tube, and is welded to part 5. Part 9 is merely a coupler which allows 10 and 11 to connect with one another. The thermocouple tree is the means of getting temperature data from the inside of the thermosyphon. It is a straight copper tube with two rows of branches along its length at 180 degrees. This serves as a conduit for thermocouple wires inside the thermosyphon. Various configurations of branches will be used in order to accomplish the goal of mapping the temperature distribution within the heat pipe.

Temperature data is taken out of the rotating module by means of a slip ring assembly. The wires are channelled into the shaft (part 14), through the bearing, through the slip ring adaptor (13) and finally to the slip ring itself. The readings are then recorded on a Fluke data acquisition unit.

Heat is supplied to the system by two clam-shell heaters, which will be contained in an insulated heater housing. They have a maximum heating capacity of 800 W, which is high enough for methanol.

Cooling at the condenser end will be accomplished with cold air. A Vortex tube forces cold air into a small chamber, which is within a plexiglass cooling box (similar in design to the heater housing).

There are three ways to estimate the heat load of the system. The first is simply to note the amount of power supplied to the heater. The second is to measure the total heat flux at the evaporator wall. The thermocouples are embedded at different depths in the evaporator wall in order to establish a temperature gradient. This will give an estimate of the conductive heat flux. Lastly, and probably the most accurate, is the measure of heat convected out by the air flow in the condenser. Flow-rate will be measured and air temperatures monitored to get an overall rate of heat removal. We are in the process of calibrating the slip ring for inaccuracies due to temperature and rotation speeds of the slip ring itself.

The test will be driven by a 2 hp electric motor through a tuned shaft. We are therefore limited, at the moment, to conducting tests at fairly low speeds (up to 3000 rpm). It is our hope to eventually acquire the means to run our test at much higher speeds (20,000) to better simulate aircraft operating conditions. It was for this reason that our apparatus was made much stronger than is necessary for our tests. Data will be taken while varying several different parameters; these include rotational speed, amount of noncondensible gases, and heat load.

Due to a great deal of out-of-house machining, and the failure to get the air bearings shipped in time, the test is not yet assembled. This will be accomplished over the course of the coming academic year. Bryan Martin will work on this project as the topic of his masters thesis and will maintain close contact with this laboratory, making return trips whenever possible.

ANALYTICAL EFFORTS

As mentioned before, the analytical approach consists of modeling the vapor/gas region and the liquid film separately. Later the two solutions are coupled through the interfacial matching condition. Figure 4 shows the two different regions.

VAPOR/GAS SOLUTION

In the absence of chemical reactions, the equation governing the molar concentration is written as

$$\frac{\partial c_2}{\partial t} + \nabla(N_2) = 0 \tag{1}$$

where c is molar concentration, t is time, N is the molar flux, and subscripts 1 and 2 represent vapor and gas. Assuming steady state, the governing equation reduces to

$$\nabla(N_2) = 0 \tag{2}$$

Since N_2 is zero at all boundaries, it can be shown that the only acceptable solution is

$$N_2 = 0 \qquad \text{everywhere} \tag{3}$$

The molar flux N_2, however, could be related to the mixture velocity as

$$N_2 = x_2 cV - cD\nabla(x_2) \tag{4}$$

where x is molar fraction, V is velocity, D is diffusion coefficient, and variables with no subscript represent those for the vapor/gas mixture. Equations (3) and (4) may be combined to solve for V as

$$V = D\nabla(\ln x_2) \tag{5}$$

This equation and the continuity equation

$$\nabla(V) = 0 \tag{6}$$

result in a conduction-type equation that governs the molar fraction for the gas.

$$\nabla^2(y) = 0 \tag{7}$$

where a new variable has been defined, mainly for convenience, as

$$y = \ln x_2 \tag{8}$$

Finally, the molar fractions for the gas and vapor are related as

$$x_1 + x_2 = 1 \tag{9}$$

Equation (8) may be solved to obtain molar fractions for the vapor and gas. Having the molar concentrations, then equation (9) may be used to obtain the molar velocity of the mixture at any point. Next, assuming inviscid flow, the pressure of the mixture is obtained from potential flow considerations as

$$p = P_0 - \frac{\rho}{2}V^2 \tag{10}$$

where P_0 is an arbitrary number. The determination of the constant P_0 will be discussed later.

BOUNDARY CONDITIONS

Figure 5 shows a schematic of the vapor/gas region including the coordinates. The vapor molar flux may be written as

$$N_1 = x_1 cV - cD\nabla(x_1) \tag{11}$$

The molar velocity is related to the molar fluxes as

$$cV = N_1 + N_2 \tag{12}$$

Substitution of equation (12) into equation (11), considering that N_2=0, and rearranging results in

$$N_1 = -\frac{cD\nabla x_1}{1 - x_1} \tag{13}$$

Or in terms of x_2

$$N_1 = \frac{cD\nabla x_2}{x_2} = cD\nabla y \tag{14}$$

At the evaporative boundary, we can write

$$N_1 = cD\nabla y = -cU_e \tag{15}$$

or

$$\nabla y = \left(\frac{\partial y}{\partial r}\right)_{r=r_1} = -\frac{U_e}{D} \tag{16}$$

Similarly at the condensation boundary,

$$\left(\frac{\partial y}{\partial r}\right)_{r=r_1} = \frac{U_c}{D} \tag{17}$$

At all other boundaries the condition is that of no mass transfer, that is

$$(\frac{\partial y}{\partial r})_{r=r_1} = 0 \tag{18}$$

The symmetrical boundary condition at r=0 is written as

$$(\frac{\partial y}{\partial r})_{r=0} = 0 \tag{19}$$

The velocities U_c and U_e are related to the rate of cooling and heating at the condenser and evaporative sections. Figure 6 shows the schematic of a simple model that describes these sections. The rate of heat transfer can be related to the velocities U as

$$\rho_c U_c h_{fg} = \frac{T_{i,c} - T_{\infty,c}}{\frac{1}{h_c} + \frac{\delta_c}{k_c}} \tag{20}$$

$$\rho_e U_e h_{fg} = \frac{T_{\infty,e} - T_{i,e}}{\frac{1}{h_e} + \frac{\delta_e}{k_e}} \tag{21}$$

where h_{fg} is the latent heat of vaporization, T_i is the temperature of the liquid film interface, T_∞ is the surrounding temperature, h is the outside heat transfer coefficient, δ is the liquid film thickness, and k is the liquid thermal conductivity.

SOLUTION FOR y

The governing equation (7) and boundary conditions (16-19) can, in principle, be solved by separation technique. Assuming circumferencial symmetry, the governing equation reduces to:

$$\frac{\partial^2 y}{\partial z^2} + \frac{1}{r}(r\frac{\partial y}{\partial r}) = 0 \tag{22}$$

We may switch to a dimensionless set of variables such that

$$Z = z/L \qquad R = r/r_1 \qquad a = L/r_1 \tag{23}$$

The dimensionless governing equation can be written as:

$$\frac{\partial^2 y}{\partial Z^2} + a^2 \frac{1}{R}\frac{\partial}{\partial R}(R\frac{\partial y}{\partial R}) = 0 \tag{24}$$

The variable y can be separated such that

$$y(R,Z) = y_R(R)\, y_Z(Z) \tag{25}$$

The governing equation may now be separated as

$$\frac{1}{y_Z}\frac{d^2 y_Z}{dZ^2} = -\frac{a^2}{y_R}\frac{1}{R}\frac{d}{dR}(R\frac{dy_R}{dR}) = -\lambda^2 \tag{26}$$

with the following homogeneous boundary conditions

$$\frac{\partial y_Z}{\partial Z} = 0 \qquad \text{at} \quad Z=0 \text{ and } 1 \tag{27}$$

and

$$\frac{dy_R}{dR} = 0 \qquad \text{at} \quad R=0 \tag{28}$$

The separate solutions are

$$y_Z = (\text{const.})\cos(\lambda Z) \tag{29}$$

where the eigenvalue λ is obtained from

$$\lambda = n\pi \qquad n = \text{any integer} \tag{30}$$

and

$$y_R = (\text{const.})\, I_0(\frac{\lambda}{a}R) \tag{31}$$

Therefore

$$y = \sum_{n=1}^{\infty} A_n I_0(\frac{n\pi}{a}R)\cos(n\pi Z) \tag{32}$$

The constant A is calculated by satisfying the interface condition at the condenser and evaporative sections. The interface condition is written as

$$(\frac{\partial y_R}{\partial R})_{R=1} = u(Z) \tag{33}$$

where

$$u(Z) = -\frac{U_e r_1}{D} \qquad \text{for } 0<Z<\frac{L_e}{L} \tag{34}$$

$$u(Z) = 0 \qquad \text{for } \frac{L_e}{L}<Z<1-\frac{L_c}{L} \tag{35}$$

$$u(Z) = \frac{U_c r_1}{D} \qquad \text{for } 1-\frac{L_c}{L}<Z<1 \tag{36}$$

The remaining task is to expand u in terms of a series of Bessel functions.

$$u(Z) = \sum_{n=1}^{\infty} 2\left\{\int_0^1 u(\zeta)\cos(n\pi\zeta)\mathrm{d}\zeta\right\}\cos(n\pi Z) \tag{37}$$

The constant A can be found by comparing equations (37) and (32) and substituting in equation (32) to get the final result for y as

$$y(R,Z) = -\sum_{n=1}^{\infty} \frac{2\int_0^1 u(\eta)\cos(n\pi\eta)\mathrm{d}\eta}{\frac{n\pi}{a}\mathrm{I}_1(\frac{n\pi}{a})}\mathrm{I}_0(\frac{n\pi}{a}R)\cos(n\pi Z) \tag{38}$$

Since all the boundary conditions resulting in equation (38) were of the second kind, an arbitrary constant could be added to the solution. Using equation (8), the molar ratio for the gas may be written as

$$x_2(R,Z) = (\text{const.})\ e^{y(R,Z)} \tag{39}$$

This constant may be obtained by checking the inventory of the noncondensible gases

$$2\pi r_1^2 L\int_0^1\int_0^1(\text{const.})e^{y(R,Z)}c\mathrm{d}R\mathrm{d}Z = C_2 \tag{40}$$

where C_2 is the total number of moles of the noncondensible gas in the heat pipe. Calculating the constant from equation (39), replacing c by p/R_cT where R_c is the universal gas constant, and substituting it in equation (38) yields

$$x_2(R,Z)=\frac{C_2 e^{y(R,Z)}}{2\pi r_1^2 L\int_0^1\int_0^1 e^{y(R,Z)}(\frac{p}{R_cT})\mathrm{d}R\mathrm{d}Z} \tag{41}$$

where y is calculated from equation (38).

SOLUTION FOR LIQUID FILM

The governing equation relating the rate of heat transfer to the rate of change in the liquid film thickness is taken from Marto et al., and is written as

$$\frac{k(r_1+z\sin\phi)\Delta T}{\delta h_{fg}} = \frac{d}{dz}\left\{\frac{\rho^2\omega^2}{\mu}(r_1+z\sin\phi)(\sin\phi-\cos\phi\frac{d\delta}{dz})\frac{\delta^3}{3}(r_1+z\sin\phi)\right\} \quad (42)$$

where variables without subscripts are now those of the liquid, ϕ is the angle of the tapered inner wall, ω is the angular velocity of the heat pipe, and μ is the viscosity. Assuming $r_1 >> z\sin\phi$, equation (42) may be rearranged to

$$\delta\frac{d}{dz}\left\{(\sin\phi-\cos\phi\frac{d\delta}{dz})\delta^3\right\} = K \quad (43)$$

where K is a dimensionless group defined as

$$K=\frac{3k\Delta T\mu}{h_{fg}\rho^2 r_1\omega^2} \quad (44)$$

This equation (42) may be rearranged as

$$\frac{3}{4}\sin\phi\frac{d\beta}{dz} - \frac{1}{4}\cos\phi\ \beta^{1/4}\ \frac{d^2\beta}{dz^2} = K \quad (45)$$

where $\beta=\delta^{-1}$. This equation cannot be solved in its present form. However, a change of variable such as

$$\xi = \frac{d\beta}{dz} \quad (46)$$

will make a solution viable. The new form of equation (45) is written as

$$\sin\phi\ \xi - \frac{1}{8}\cos\phi\frac{d\xi^2}{d\beta^{3/4}} = K \quad (47)$$

Rearranging equation (47) results in

$$d(\beta^{3/4}) = -\frac{\cos\phi}{4}\,\frac{\xi d\xi}{\frac{4}{3}K-\sin\phi\,\xi} \tag{48}$$

This equation can be integrated to solve for the liquid film thickness. The boundary conditions for δ are

$$\frac{d\delta}{dz} = \tan\phi, \qquad \delta = \delta_0 \qquad \text{at} \qquad z=0 \tag{49}$$

The first condition is derived from the fact that at z=0 there would be no liquid flow. These two conditions in terms of ξ and β become a single equation, which is

$$\beta = \delta_0^{-1} \qquad \text{at} \quad \xi=0 \tag{50}$$

Equation (48) now can be integrated using boundary condition

$$\delta^3 = \delta_0^3 + \frac{\cos\phi}{4}\left\{\frac{\xi}{\sin\phi} + \frac{4K}{3\sin^2\phi}\ln\left[\frac{4/3K-\sin\phi\,\xi}{4/3K}\right]\right\} \tag{51}$$

A closed from solution in terms of δ and z cannot be obtained. Once δ vs ξ is obtained from equation (51), ξ may be related to δ and z from

$$\xi = 4\delta^3\frac{d\delta}{dz} \tag{52}$$

This is a versatile equation: it can be used with heating (evaporative free surface), cooling (condensing free surface), or insulated wall (no mass transfer at the free surface), by using a positive, negative, or 0 value for ΔT.

SOLUTION METHODOLOGY

In the previous sections, analytical solutions for the vapor/gas region and liquid film have been obtained. These two solutions are related through the thermodynamics of the interface. An overall closed form and analytical solution is not possible due to the complex nature of the analytical results and the thermodynamic relation that relates the local molar fraction of the vapor to the vapor/liquid film interface temperature. Numerical results, however, can easily be obtained.

An iterative procedure is unavoidable, due to the nature of the nonlinearities in the problem.

One possible iterative scheme is listed below:

1) Guess an initial U_c and U_e, realizing that to achieve steady state the mass balance dictates that the vaporization mass should be equal to the condensing mass.

$$\int_0^{L_e} \rho_e U_e dz = \int_{L-L_c}^{L} \rho_c U_c dz \tag{53}$$

2) Calculate u(Z) from equations (33)-(37)

3) Calculate y(R.Z) from equation (38).

4) Calculate x_2 from equation (41).

5) Calculate x_1 from equation (9).

6) Having x_2, the partial pressure of the vapor may be calculated from $p_2=x_2p$.

7) The liquid film-vapor interface temperatures $T_{i,c}$ and $T_{i,e}$ may now be calculated from thermodynamic tables.

8) The temperature difference may then be calculated as

$$\Delta T = T_{\infty,e} - T_{i,e} \quad \text{or} \quad \Delta T = T_{i,c} - T_{\infty,c} \tag{54}$$

9) The liquid film thickness δ is calculated from equation (51).

10) The velocities U_c and U_e can now be calculated from equations (20) and (21).

11) Check if calculated values of U_c and U_e are close to the guessed values? If yes, stop the iterations; if not, go back to step 1.

FUTURE DIRECTIONS

This subject has barely been touched. The combination of noncondensible gases and rotation has not been fully investigated before. The same combination is also responsible for the complex and interesting phenomenon at hand. Since no extensive result has been obtained here due to limited time and the scope of the project, more work will need to be planned for the future. Among the activities proposed for the future are:

1) Perform analysis for the case of no tapering. Because the heat pipe might be rotating very fast (up to 30,000 rpm) it is suspected that tapering is unnecessary for liquid transport. A thicker liquid film at the condensate section and a high centrifugal force might generate enough hydrostatic head to supply the necessary liquid flow.

2) Following the analysis in (1) above, numerical results should be obtained and presented in terms of plots and tables.

3) Numerical results will be obtained for the concentration of the noncondensible gas. These results will be coupled with those obtained by Edwards, Tien.

4) Overall results will be obtained in which the solutions of the vapor\gas region and the liquid film are coupled. These results will be compared with those available in the literature -- possibly those of Al-Jumaily (1973) and Williams (1976).

5) Experiments will be conducted. The data will be compared with those results obtained here and those by Edwards and Marcus (1972) and Peterson and Tien (1989).

6) A comprehensive computer program which accepts a few input parameters such as angular revolution, system pressure, and amount of noncondensible gases and solves for rate of heat transferred and the temperature profile in the heat pipe is desirable.

7) Upon examination of the experimental data and analytical results, recommendations will be made for better designs and more efficient analytical treatment.

8) The possible effect of ultra high revolutions on the performance of the gas-loaded heat pipes is not that clear at this point. Future experimental efforts aimed at investigating this are recommended.

REFERENCES

Al-Jumaily, F. K., "Investigation of the Factors Affecting Performance of a Rotating Heat Pipe," Ph.D. Thesis, University College of Swansea, Wales, 1973.

Chang, W. S., and Yu, J. S., "A note on the Gas Distribution in a Cylindrical Gas-Loaded Heat Pipe," *Journal of Heat Transfer*, Volume 12, 1990, pp. 779-781.

Daniels, T. C., and Al-Baharnah, N. S., "Temperature and Heat Load Distribution in Rotating Heat Pipes," *AIAA Journal*, Volume 18, Number 2, 1980, pp. 202-207.

Daniels, T. C., and Al-Jumaily, F. K., "Investigations of the Factors Affecting the Performance of a Rotating Heat Pipe," *International Journal of Heat and Mass Transfer*, Volume 18, 1975, pp. 961-973.

Daniels, T. C., and Williams, R. J., "Experimental Temperature Distribution and Heat Load Characteristics of Rotating Heat Pipes," *International Journal of Heat and Mass Transfer*, Volume 21, 1978, pp. 193-201.

Daniels, T. C., and Williams, R. J., "The Effect of External Boundary Conditions on Condensation Heat Transfer in Rotating Heat Pipes," *International Journal of Heat and Mass Transfer*, Volume 22, 1979, pp. 1237-1241.

Edwards, D. K., and Marcus, B. D., "Heat and Mass Transfer in the Vicinity of the Vapor-Gas Front in a Gas-Loaded Heat Pipe," *Journal of Heat Transfer*, Volume 94, 1972, pp. 155-162.

Leppert G, and Nimmo, B., "Laminar Film Condensation on Surfaces Normal to Body or Inertial Forces," *Journal of Heat Transfer*, Volume xx, 1968, pp. 178-179.

Marto, P. J., "Laminar Film Condensation on the Inside of Slender, Rotating Truncated Cones," *Journal of Heat Transfer*, Volume xx, 1973, pp. 270-272.

Petreson, P. F., and Tien, C. L., "Numerical and Analytical Solutions for Two-Dimensional Gas Distribution in Gas-Loaded Heat Pipes," *Journal of Heat Transfer*, Volume 111, 1989, pp. 598-604.

Williams, R. J., "Influence of Inert Gases on the Performance of Rotating Heat Pipes," Ph.D. Thesis, University of College Swansea, Wales, (1976).

Yerkes, K. L., "Technology Review: Utilizing Rotating Thermosyphon Technology in Aircraft Thermal Management and Control," *Aerospace Technology and Exposition*, Long Beach, California, October 1-4, 1990.

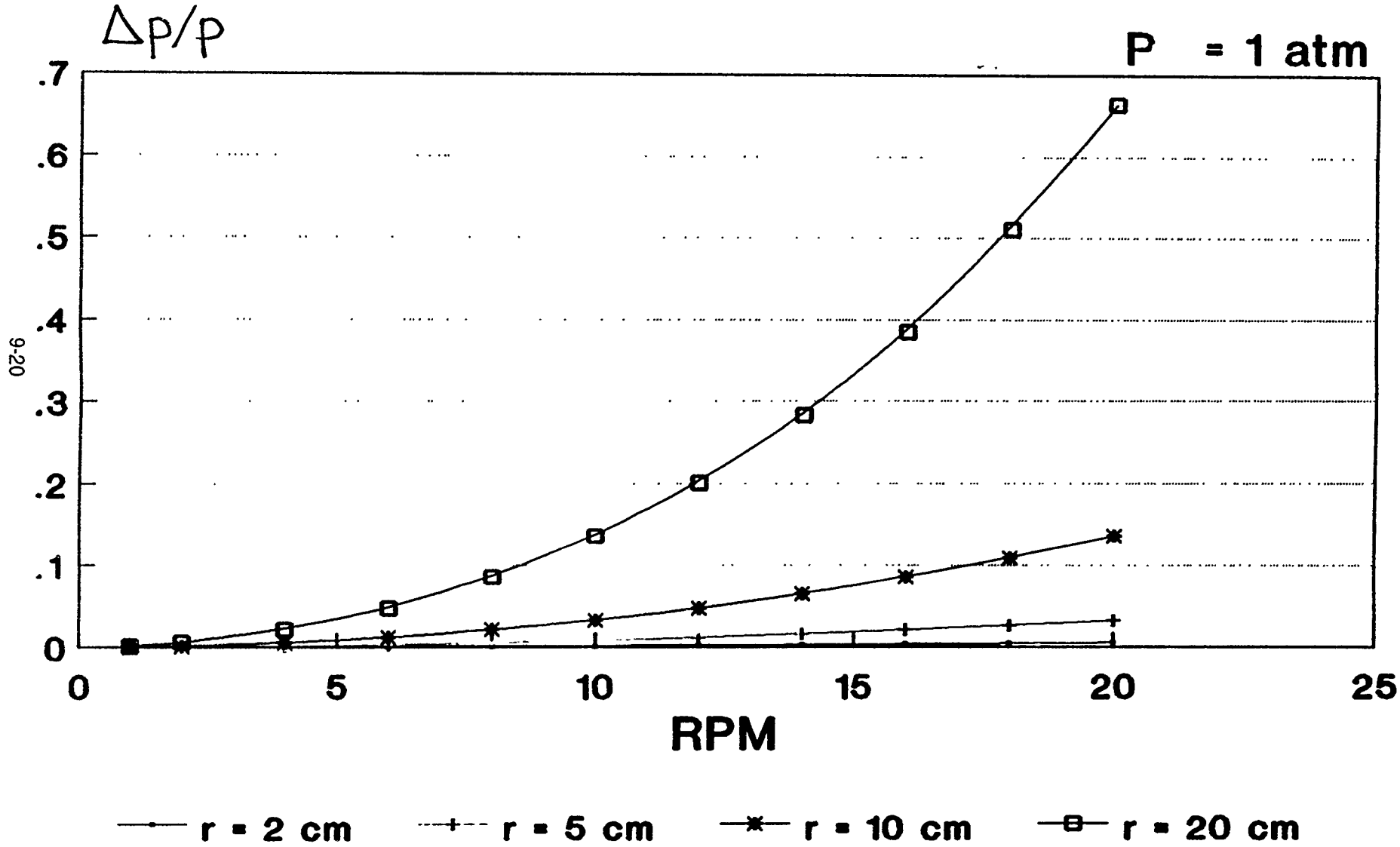


Figure 1.

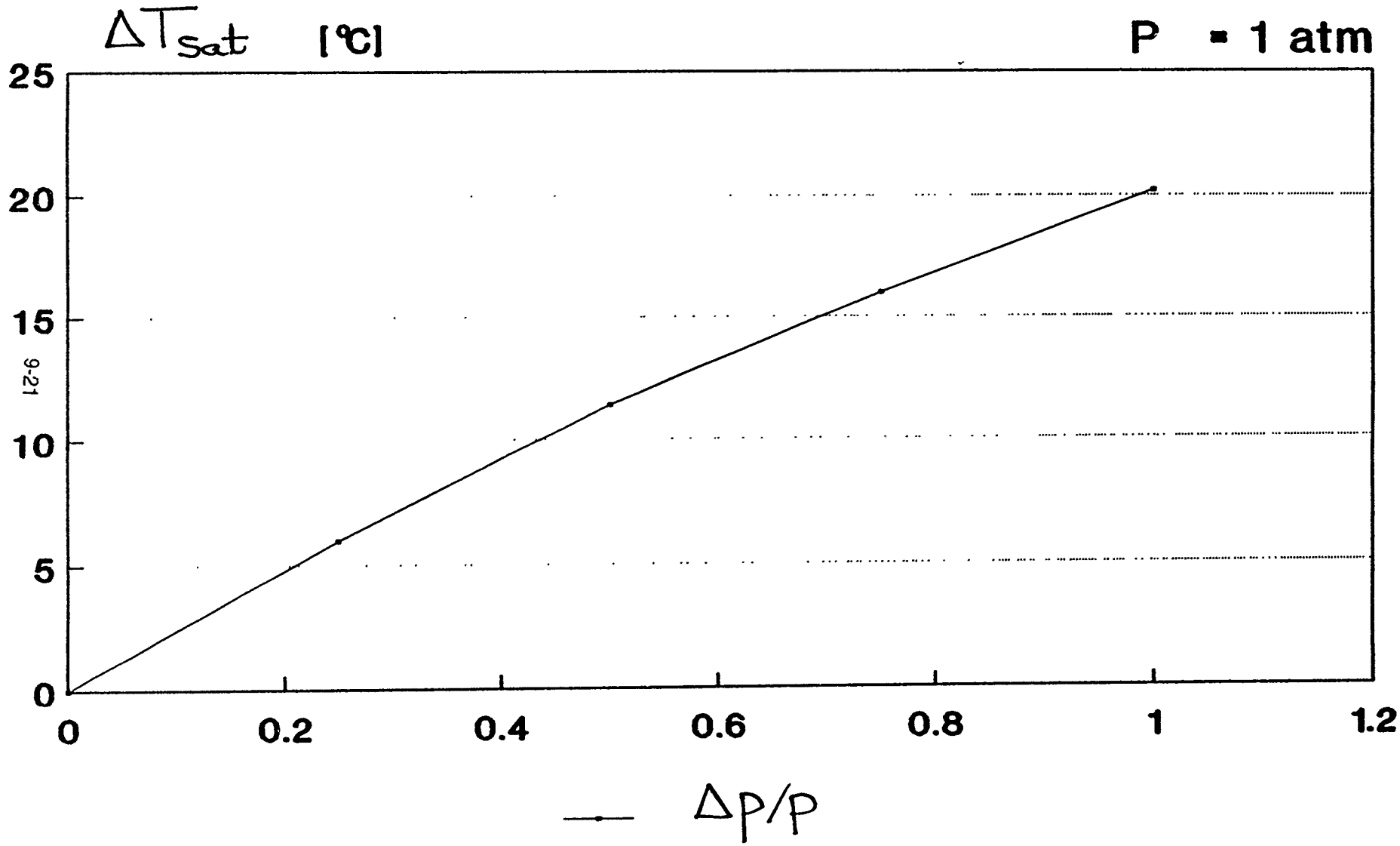
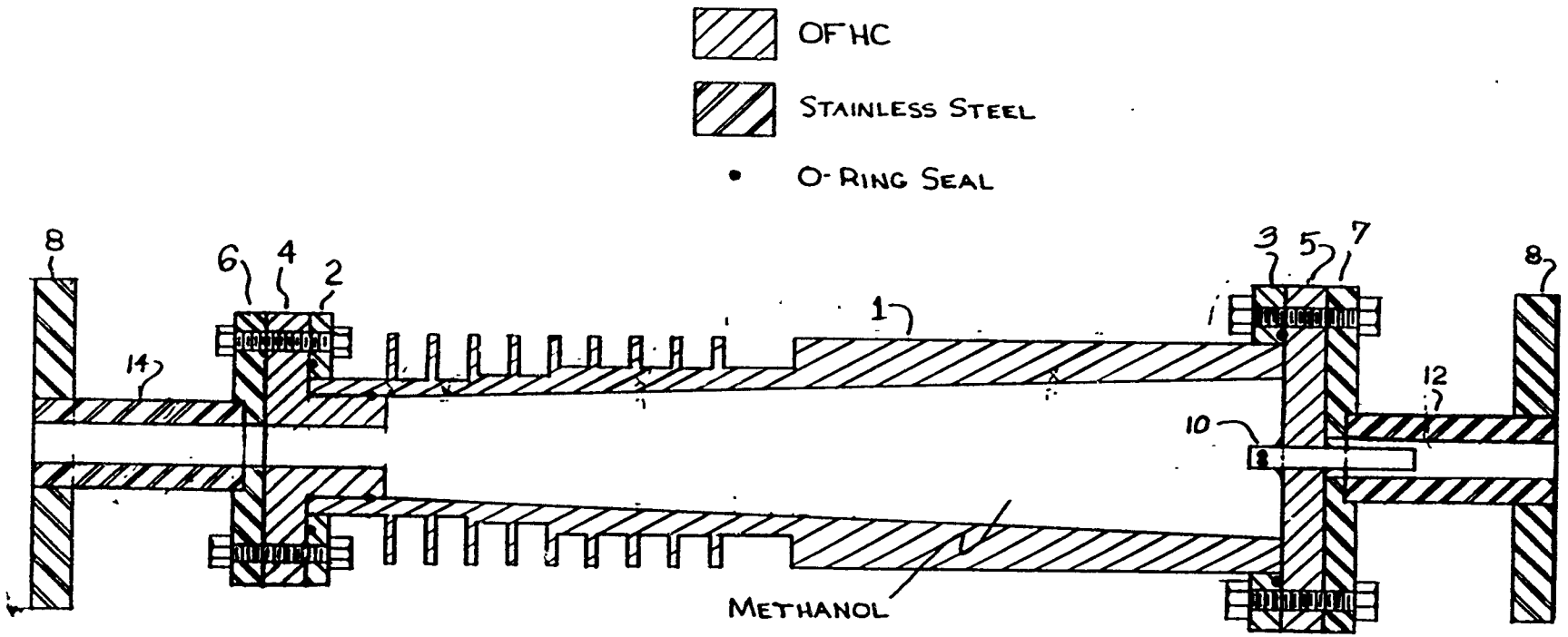
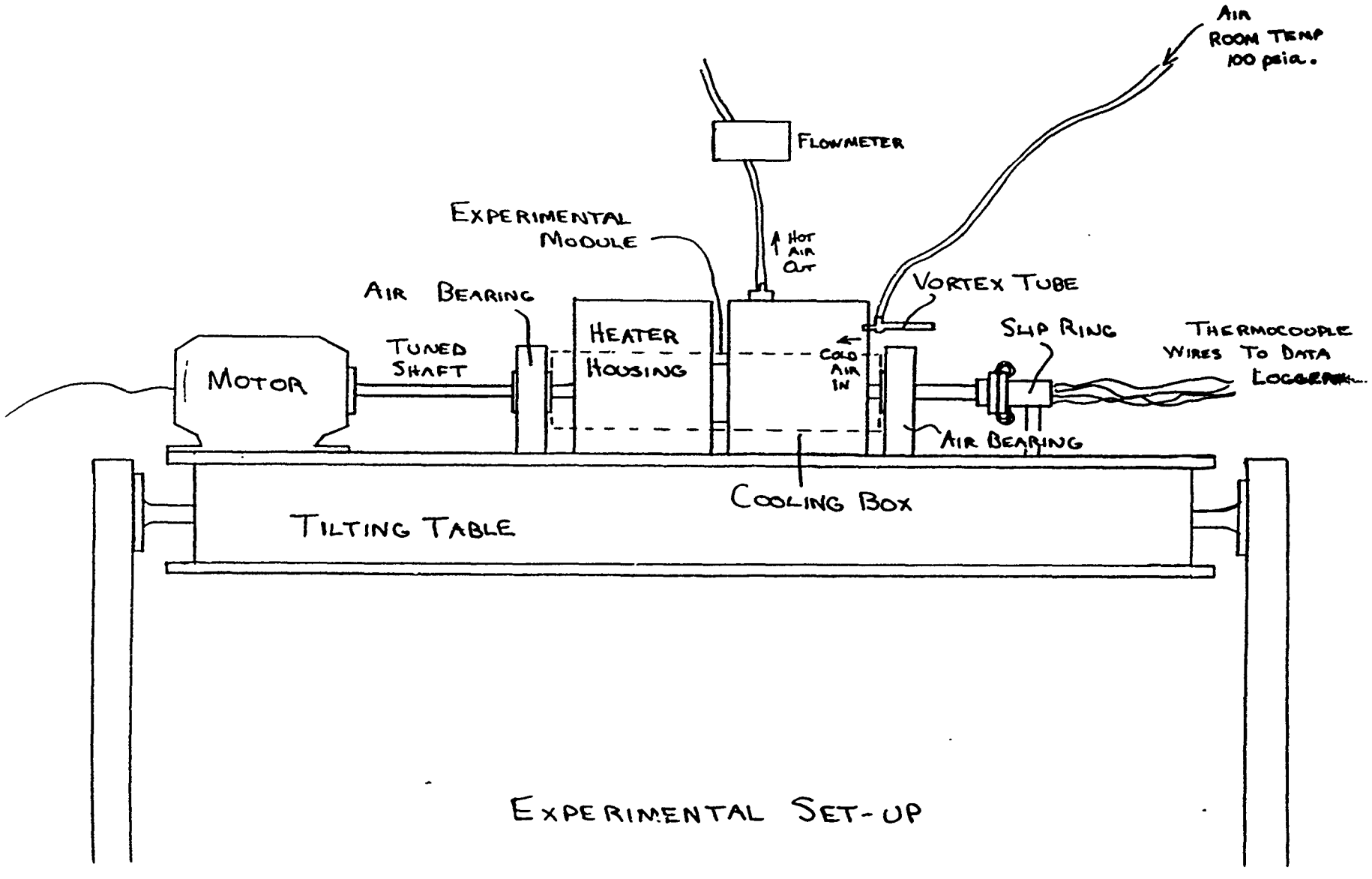


Figure 2.



TEST MODULE
ROTATING THERMOSYPHON

Figure 3a



EXPERIMENTAL SET-UP

Figure 3b

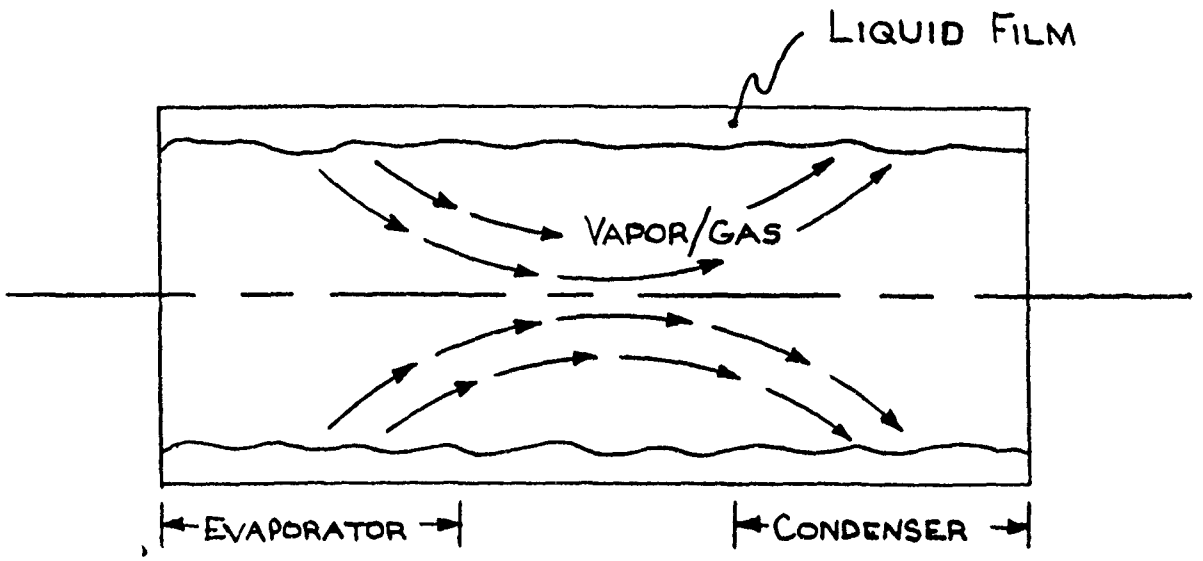


FIGURE 4

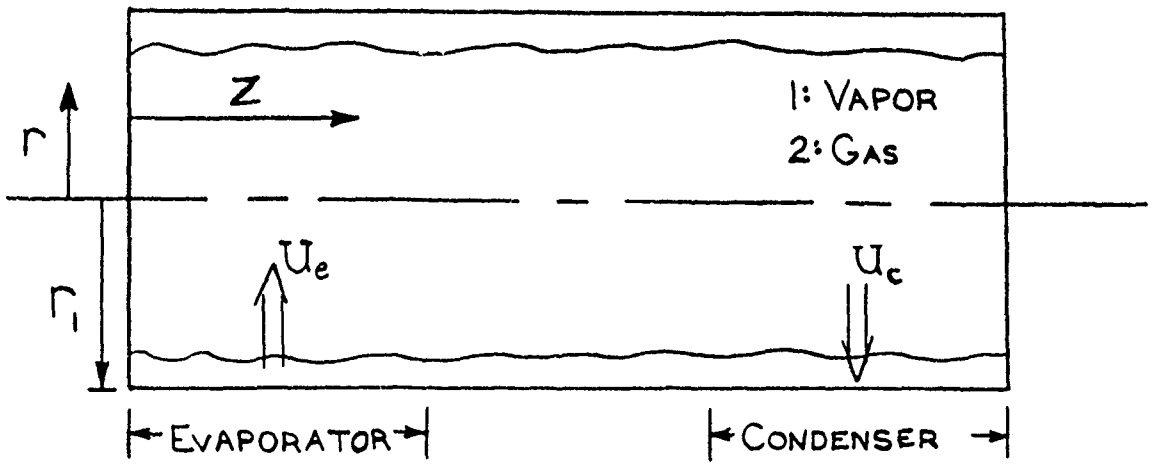


FIGURE 5

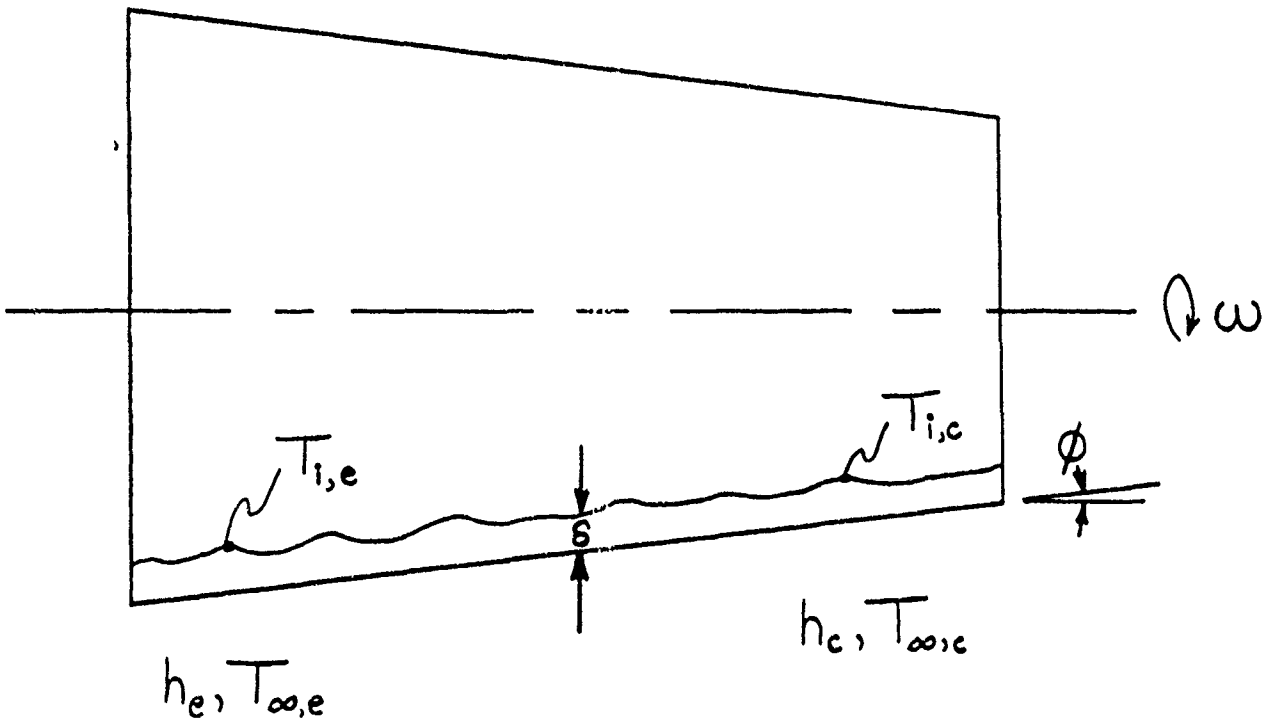


FIGURE 6

1991 USAF - RDL SUMMER FACULTY RESEARCH PROGRAM

GRADUATE STUDENT RESEARCH PROGRAM

Sponsored by the

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

Conducted by the

Research and Development Laboratories

FINAL REPORT

MEASUREMENTS OF DROPLET VELOCITIES AND SIZE DISTRIBUTIONS IN PRESSURE/AIR BLAST ATOMIZER

Prepared by:	Richard S. Tankin
Academic Rank:	Professor
Department and University:	Mechanical Engineering Department Northwestern University
Research Location:	WL/POSF Wright Patterson Air Force Base Dayton, Ohio 454336563
USAF Researcher:	W. M. Roquemore
Date:	6 September 1991

CONTRACT NO. F49620-90-C-0076

Acknowledgements

I wish to thank WL/POSF at Wright Patterson Air Force Base and the Air Force Office of Scientific Research for the sponsorship of this research. I also want to thank Research and Development Laboratories for their efficient handling of all administrative aspects of the program.

This summer's research was beneficial to me and hopefully will lead to joint research activities with Wright Patterson. Having worked before with Dr. W. M. Roquemore, it was no surprise to find him very cooperative and encouraging. The experimental work could not have been accomplished without the expertise and experience of Dr. G. Switzer. The ease and readiness with which Jeffrey Stutrud handled complex computer problems was invaluable. I found the staff at Wright Patterson to be friendly, stimulating, and helpful.

MEASUREMENTS OF DROPLET VELOCITIES AND SIZE DISTRIBUTIONS IN PRESSURE/AIR BLAST ATOMIZER

by

Richard S. Tankin

Abstract

A phase doppler instrument was used to measure droplet sizes and velocities in a water spray. This nozzle consisted of a hollow cone water spray and two swirling, concentric air channels. The water flow was maintained at 2.75ml/sec; and air flow was either zero or 4089 ml/sec. Horizontal traverses were made across the spray near the sheet break-up region. With no air flow, the traverse was made at 10mm from nozzle; with air flow, the traverse was made at 5mm from the nozzle. More than 140,000 samples were taken in each traverse. The results show that the spray is axially symmetric which is important for the planned theoretical analysis. A new procedure is being developed to analyze the data. This required collecting the raw data and then placing them in velocity (two components) and diameter bins. When using this procedure, there will be no restriction on the size of windows used during the collection of the data. This is an improvement on the procedure that was developed two years ago to analyze spray data from a pressure atomizer. Measurements were also made on a spray that was driven by a piezoelectric unit.

I. INTRODUCTION

Gas turbines in nearly all cases utilize liquid fuels. These fuels must be delivered as small droplets to the combustion zone if efficient combustion is to occur. The process by which these small drops are formed is called atomization. The method of achieving atomization is varied and not a topic for consideration in this study. In this study the droplet distributions from a pressure/air blast atomizer are examined. The liquid spray discharges from the nozzle as a liquid sheet which breaks up downstream to form droplets. Surrounding this liquid sheet are two concentric, swirling air flows.

It is important to determine the spray characteristic if one hopes to correlate different sets of experimental data, compute numerical simulations, determine droplet trajectories, etc. In the past, these characteristics were limited to drop size distribution, patternation, cone angle, dispersion, and penetration. For example various techniques have been used to determine the drop size distribution - photographic, optical, collection devices, etc. Each of these techniques have very limited capabilities. The Malvern technique obtains a size distribution that is integrated over the optical path length. To determine radial distribution requires the use of Abel inversion. Dense sprays, or asymmetry can complicate this technique. Recently, a new instrument (Phase Doppler/Particle Analyzer) has been developed which simultaneously measures the droplet size and two velocity components. Also a new method for analyzing the droplet distributions - using the principle of maximum entropy - has been proposed.

Ia. INSTRUMENTATION

A highly sophisticated instrument was developed by Aerometrics which is capable of optically measuring the sizes and velocity of droplets. This method utilizes light scattering techniques, and consists of a transmitter, receiver, signal processor, computer and laser light source. The transmitter has a beam expander which reduces the size of the measuring point - which in our case is about 1 x $10^{-4}cm^2$. Thus excellent spatial resolution is achieved. Since the detectors in the receiver unit are photomultipliers (three), the response time of

this instrument is very short. Signals from individual drops can be processed and the data transferred to computer memory in 20 μsec. For the operator, an important aspect of this system is the software program associated with the signal processor. As data are being collected and stored by the computer, they are presented in bar graph form for the operator to observe. After a selected number of droplets are collected, a listing of pertinent data - such as attempts, validations, corrected count, probe area, mean velocities, D_{32} etc. - are displayed. The bar graphs consist of particle size counts, and velocity distributions. A typical example of such a display is shown in Figure 1. Although this instrument is a self contained unit, it requires a skilled operator to accurately collect data. The operator has to set the photomultiplier voltages, windows for the size and velocity distributions, velocity off-set, etc.

Last summer at Wright Patterson Air Force Base , I collected similar data. However a major portion of that summer program was to analyze the size and velocity distributions from a single column of drops (Rayleigh problem). This summer we carefully measured the size and two components of velocity (radial and axial) of a pressure atomizer including the effects of high velocity, concentric, swirling air. This is a very complicated flow and if we can successfully predict the size and velocity distributions for this nozzle in a swirling air environment, we have made a major step in analyzing an air blast nozzle.

Ib. THEORY

The concept of information entropy was developed by Claude Shannon (1948), and Jaynes (1957) later extended this concept into the well-known method of maximum entropy formalism. This formalism can be applied to problems involving probability, i.e., where insufficient information is available to obtain exact solutions. We have applied the maximum entropy formalism to liquid sprays in which we predicted the droplet size and velocity distributions. Since the application to this problem has been adequately discussed by several researchers - Kelly (1976), Sellens and Brzustowski (1985, 1986), Sellens (1989), and Li and Tankin (1987, 1988, 1989), it will not be necessary to develop the background material once again (for details, see Li 1989). The data collected in this study will be examined using the maximum entropy principle. There have been four papers in which comparisons between maximum entropy theory and measurements exist. In one, Li and Tankin (1988) used data that

were obtained from holographic and photographic methods. Thus these results were limited to size distributions. Sellens (1989) used a phase doppler instrument to obtain data and make comparisons with theory. However in this study there are some inconsistencies and the experiments are questionable (see Li and Tankin, 1989). A recent paper by Li, Chin, Jackson, et al (1991) contains measurements of size and velocity (one component) distributions. These data were taken over the entire spray cross section and the experimental data were compared with theory. Agreement between theory and experiments is reasonable good. In another study by Chin, LaRose, Jackson at al (1991) comparisons were made between theory and experiments for an array of drops formed from the breakup of a liquid column. Again there is good agreement between experiments and theory. Recently it was determined that when conservation of momentum in the axial and radial directions are uesed in the maximum entropy principle, the appearance of a bimodal size distribution is predicted (Chin and Tankin, 1991). The occurence of this bimodal distribution depends on the values of the source terms in the theory. It would be a significant step in the development of the theory if bimodal distributions are observed experimentally.

II. OBJECTIVES OF THE RESEARCH EFFORT

In the present study we examined the following:

1. A new computer program needed to be developed which can process the data which is obtained at many spatial locations in the spray. Prior to this new program, we were restricted to collecting data at specific window settings of the PD/PA instrument.

2. Determine if a bimodal size distribution is observed experimentally for a pressure atomizer for particular flow rates which correspond to the predicted values.

3. Operate the pressure atomizing nozzle in the presence of high swirling air flow. This has the effect of converting the pressure atomizer into an air assist atomizer.

4. Determine the effects of a piezoelectric unit which imposes a disturbance on the spray of a pressure atomizer.

The nozzle used in this study was provided by Allison Corp.

III. RESULTS OF THE RESEARCH EFFORT:

A. Computer Program to Handle the PDPA Data

In earlier tests Li, Chin, Jackson, et al (1991) the window for the droplet size measurement was restricted to integral multiples; the maximum diameter equal to either 300 microns or 150 microns. Those droplets in the bins of the 150 micron window were then resorted in the bins of the 300 micron window. With intregral multiple windows it is relatively easy to determine the relationship between bin_{300} and bin_{150}. However the requirement that the window be integral multiples puts restrictions on the PD/PA measurements. To circumvent this problem a computer program is being developed at Wright Patterson Air Force Base which can handle any variety of window sizes.

At the beginning of this summer study, a computer program was written by J. Stutrud that takes the data (in bin form) from the PD/PA (short form) and sorts the data in bins created by the programer. These bins are a three dimensional array (diameter, velocity1 and velocity 2). A problem occurred when the PD/PA data is resorted in this manner. Since the bins in the PD/PA short form do not match the bins in the new program, an oscillation appears in the final distribution plots. A typical example is shown in Figure 2. To eliminate this problem, raw data from the PD/PA, which is not in a bin format, is collected and stored on a floppy disk. Then this raw data is read into the new computer program. The results obtained using this program are smooth and more accurate plots than prior plots (see Figure 3.).

B. Experiments

We have had success in comparing experimental data with predictions for a pressure atomizer operating at normal operating conditions (Li, Chin, Jackson et al, 1991) and for a column of liquid which breaks up into a single column of droplets (Chin, Jackson, et al, 1991) This summer, we were interested in determining whether a bimodal size distribution could be obtained experimentally that would match our predictions (see Chin and Tankin, 1991). Another aspect of this summer study is measure the effects of high velocity air flow on the spray. These measurements were carried out where the air flow rate (4089ml/sec) is

approximately 2,000 times greater than the water flow (2.75ml/sec). The air flow has a very significant effect on the Sauter mean diameter, radial and axial velocities. This is a hybrid air assist pressure nozzle. We are interested to see if the maximum entropy principle will predict these results. If we are successful in predicting these distributions, this will be a major step toward examining an air blast nozzle.

In Figure 4 are the normalized profiles for the Sauter mean diameter (D_{32}) and axial velocity for water discharging from the Allison nozzle at 2.75ml/sec into still air. The sheet region, where D_{32} and the axial velocity are maximum, is relatively thin. The negative axial velocities are due to the recirculation that forms within the hollow cone spray. This region extends from 0 to 7.5mm. We have excluded this region when the velocity and size distributions are computed. As an aside, if the entire profile had been used, the effects of the recirculation zone on the size distribution is negligible. This is because the ring areas for these interior regions are small and the time for collecting the 5,000 drops are relatively long. The resulting size distribution exhibits a bimodal distribution as seen in Figure 3.

In Figure 5 are the normalized profiles for the Sauter mean diameter (D_{32}) and axial velocity for water discharging from the Allison nozzle at 2.75ml/sec into a swirling air environment (4089 ml/sec). In comparing Figure 5 with Figure 3, there is a significant difference. The spray in the presence of swirling air is spread over a much wider area (the sheet region is not nearly as well defined), the axial velocity is significantly greater, and the Sauter mean diameter is significantly smaller (not shown on normalized plot). If should be added that the measurements with air were made 5mm from the nozzle exit, and those into still air were made at 10mm from the nozzle exit. In the analysis of the air assist data, we excluded the region from the centerline to 2.5mm because we considered this to be a recirculating zone. Again, if included, it would have insignificant effects on the size distribution. The resulting size distribution is shown in Figure 6.

Figure 7 is a traverse of the spray in which D_{32} is compared for a spray that is piezoelectrially driven and one that is not driven. In Figure 8 are the measured axial velocities for the driven and undriven sprays. One can readily see that the distributions for the driven spray are much broader than those of the undriven spray. D_{32} is significantly smaller in the case of the driven flow; and the axial velocity is somewhat larger. Before drawing conclusions, it is necessary to

further analyze this flow - determining the probability distribution functions. It may be that the piezoelectrically driven flow may have the drops redistributed without significantly effecting the probability distribution function.

IV RECOMMENDATIONS

The necessary experimental data have been collected and stored on floppy disks. Most of the reduction of the experimental data has been completed.Since this nozzle was supplied to us by Allison Gas Turbine Division; we will report our results to them, through Wright Patterson. If we are successful, which I think we will be, in modeling this nozzle forlow flow rates, then we will extend the data to higher water flow rates and various air flow rates. Having gained experience in the taking the experimental data and modelling the flow, we will be in a position to study an air blast nozzle.

VI. REFERENCES:

1. Shannon, C. E., "A Mathematical Theory of Communication ", The Bell System Technical Journal, 27, 1948, p.379.

2. Jaynes, E.T., "Information Theory and Statistical Mechanics", Physical Review, 106, 1957, p. 620.

3. Kelly, A. J., "Electrostatic Metallic Spray Theory", J. of Applied Physics, 47, 1976, p. 5264.

4. Sellens, R. W. and Brzustowski, T. A., "A Prediction of Drop Size Distribution in a Spray from First Principle", Atomization and Spray, 1, 1985, p. 89.

5. Sellens, R. W. and Brzustowski, T. A., "A Simplified Prediction of Droplet Velocity Distributions in a Spray", Combustion and Flame, 65, 1986, p. 273.

6. Li, X. and Tankin, R.S., "Droplet Size Distribution:A Derivation of a Nukiyama-Tanasawa Type Distribution Function", Combustion Science and Flame, 56, 1987, p. 65.

7. Li, X. and Tankin, R.S., "Derivation of Droplet Size Distribution in Sprays Using Information Theory", Combustion Science and Flame, 60, 1988, p. 345.

8. Li, X. and Tankin, R.S., "On Prediction of Droplet Size and Velocity Distributions in Sprays Through Maximum Entropy Formalism", submitted for publication, 1989.

9 Li, X., PhD Thesis, Department of Mechanical Engineering, Northwestern University, June, 1989.

10. Li, X., L.P. Chin, R.S.Tankin, T. Jackson, J. Stutrud, and G Switzer, "Comparison Between Experiments and Predictions Based on Maximum Entropy for Sprays from a Pressure Atomizer", Combustion and Flame, 86, 1991.

11. Chin, L.P., P.G. LaRose, R.S. Tankin, T. Jackson, J. Stutrud, and G Switzer, "Droplet Distributions from the Breakup of a Cylindrical Liquid Jet", Physics of Fluids A, 3. 1991.

12. Chin, L.P. and R. S.Tankin, "Theoretical Prediction of Droplet Distributions in Sprays Based on Maximum Entropy Theory", to be presented at 1991 Winter Annual ASME Meeting.

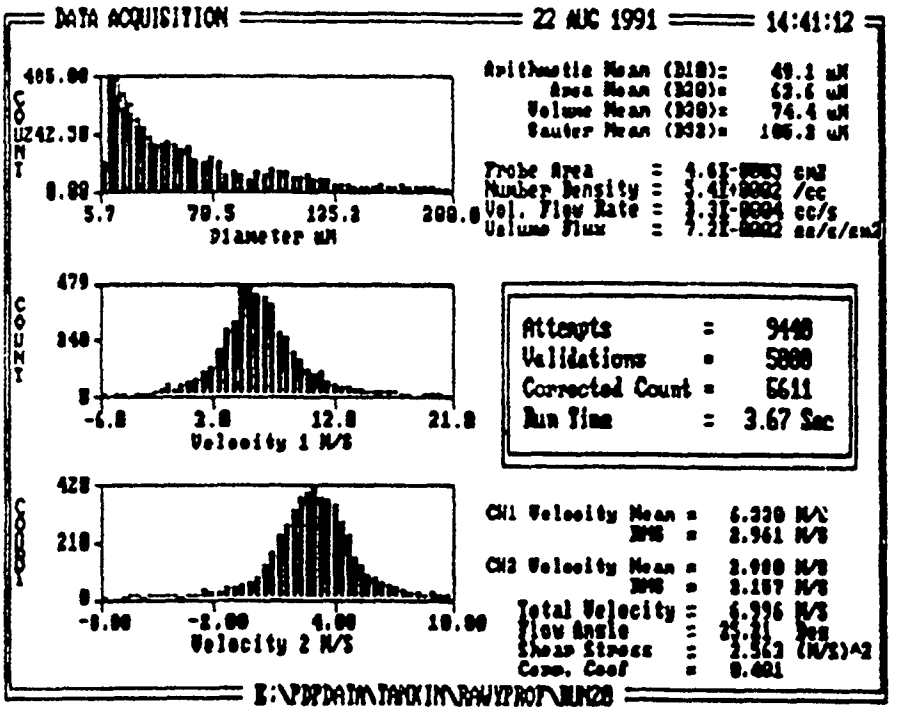


Figure 1. Typical display of print out from Aerometrics PD/PA.

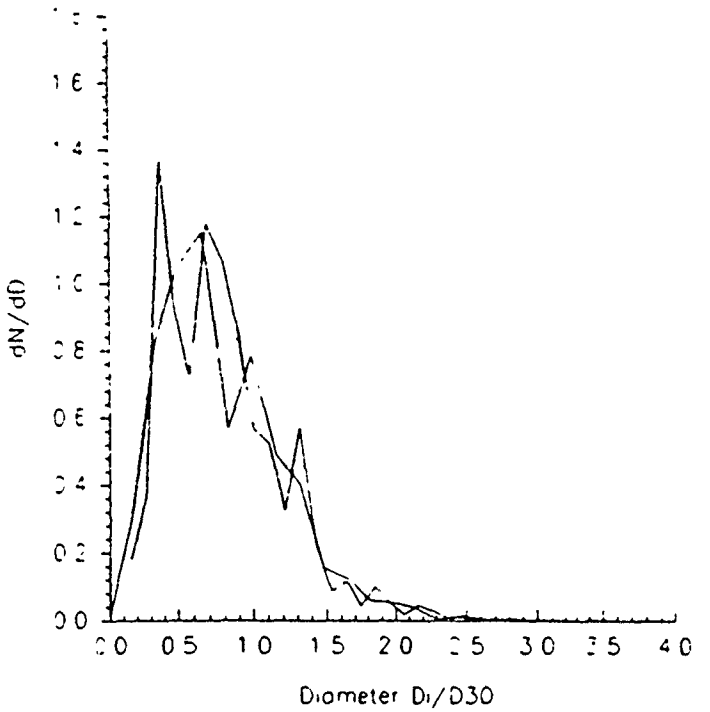


Figure 2. Oscillations in a typical size distribution due to unaccounted for variations in the windows used in the PD/PA.

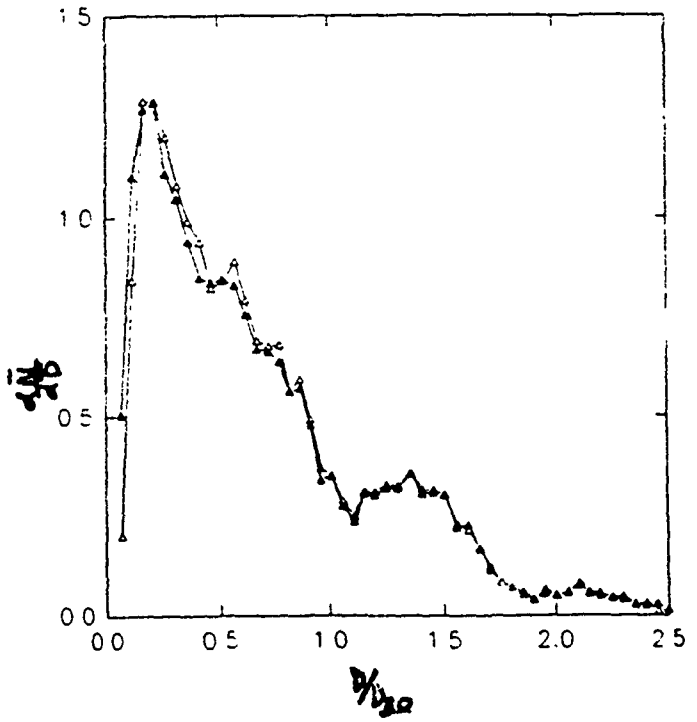


Figure 3. Typical plot of the size distribution when raw data is used.

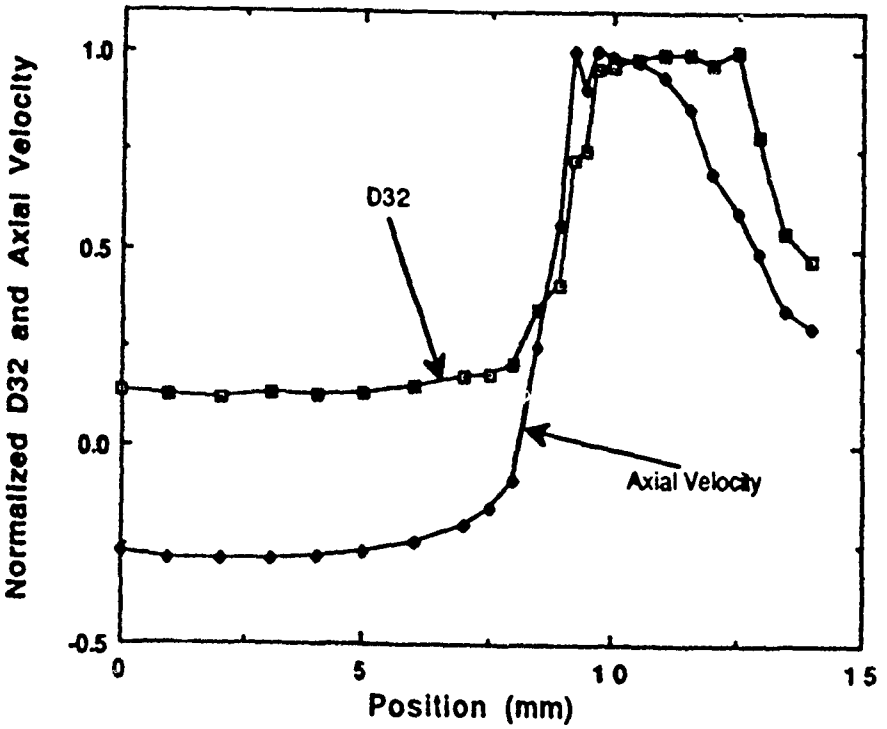


Figure 4. Normalized profiles of Sauter mean diameter and axial velocity. Spray is discharging into ambient air.

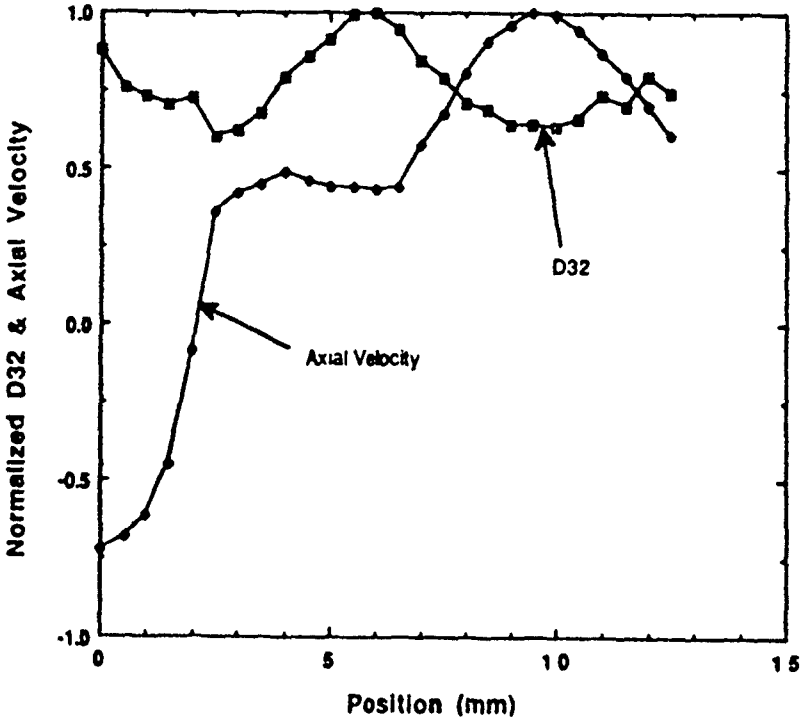


Figure 5. Normalized profiles of Sauter mean diameter and axial velocity. Spray is discharging into swirling air.

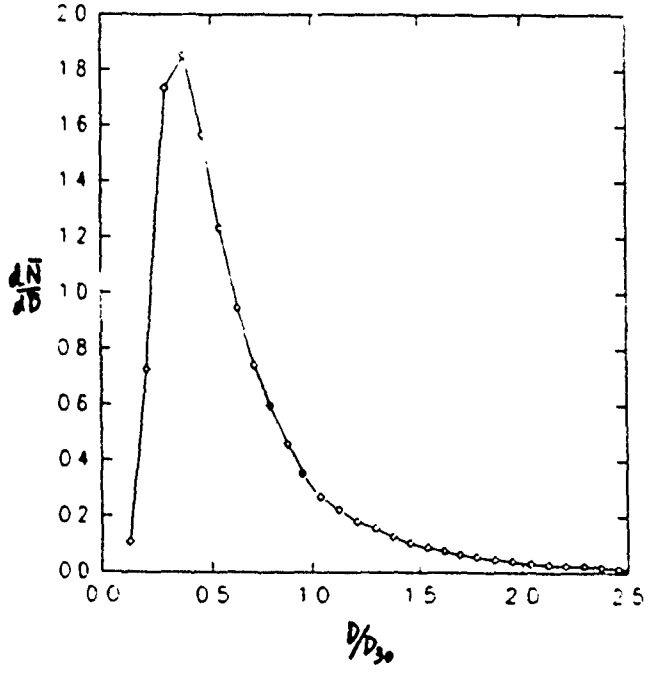


Figure 6. Size distribution of spray in the presence of swirling air.

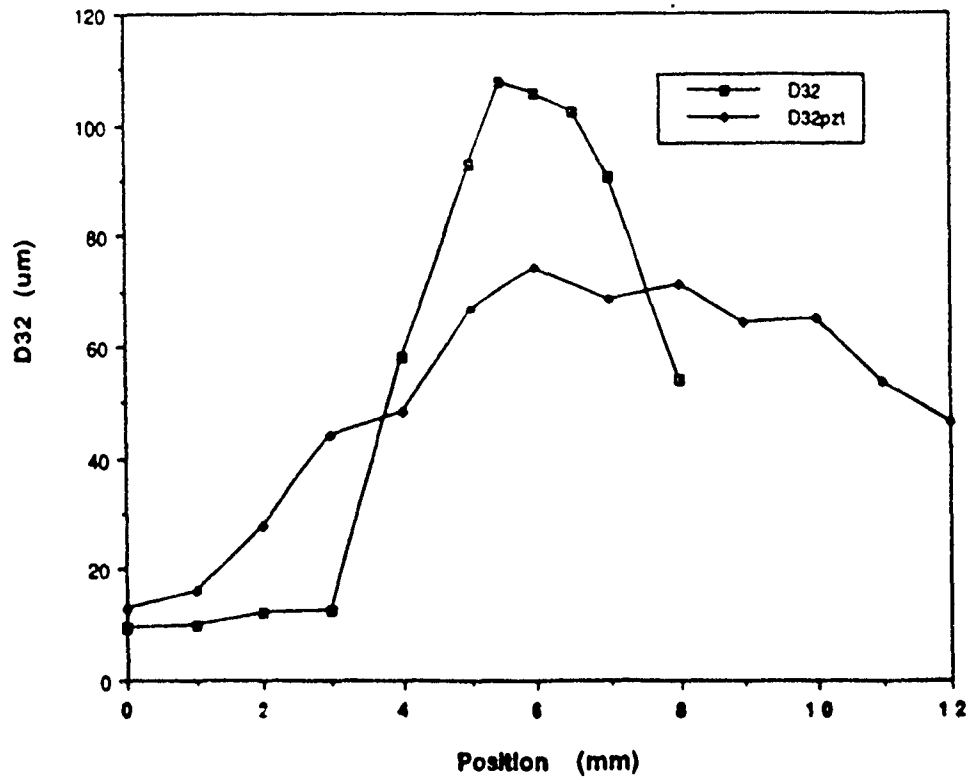


Figure 7. Traverse of the spray plane showing the effect of the piezoelectric driver on D_{32}.

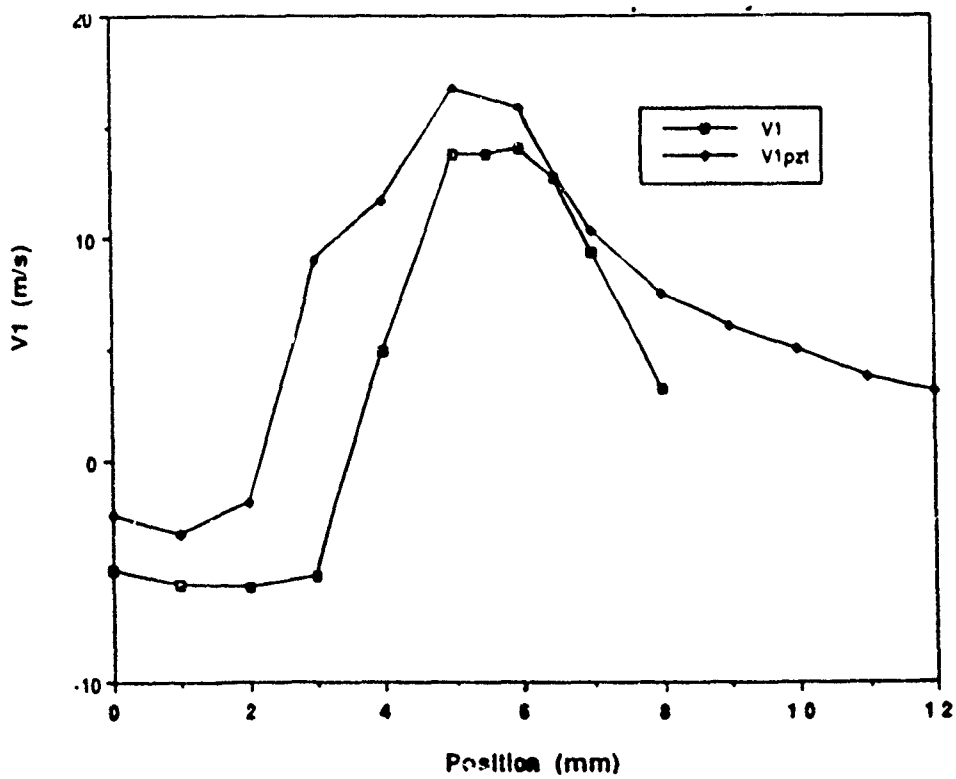


Figure 8. Traverse of the spray plane showing the effect of the piezoelectric driver on axial velocity.

TRANSIENT EFFECTS IN GLOW DISCHARGES USING THE GEC REFERENCE REACTOR

Dr. Fred V. Wells
Associate Professor
Idaho State University
Pocatello, ID 83209

Abstract

Transient effects in pulsed rf (radio frequency) excited Ar, He and N_2 plasmas were investigated in the GEC (Gaseous Electronics Conference) reference reactor. The build-up time constant for the DC bias on the order of 50 to 300 μs for Ar and He and 1 to 3.5 ms for N_2. This time constant was a function of the coupling capacitor which provides a measure of the time scale for breakdown and establishment of the resistive component of the discharge for the gases studied. On switch-off, a fast decay depending on the pressure and power was observed in addition to a slow decay that is due to the circuit time constant was observed for all the gases.

Introduction

The phenomenon of breakdown and conduction of electricity through gases has long been of interest to mankind. Perhaps the "modern" era in the study of gas discharges began in the early 1700's with Van Musschenbroek's work on electric sparks and Franklin's interest in the

natural phenomenon of lightning. Investigations of gaseous discharges during the 1800's by Plucker, Crookes, J. J. Thomson, Lorentz and Zeeman to mention only a few led to fundamental advances in electromagnetic theory and modern atomic theory. It is noteworthy that in many of the experiments gas discharges were not the subject of investigation, but rather a means of producing and studying other phenomena. Many of the fundamental mechanism involved in the gas discharge remained unexplained.

In the early part of the 20th century, gas discharges were a significant and necessary component in development of electronic devices that characterize contemporary society. The advent of the computer revolution beginning roughly about 1970 has only given increasing importance to understanding fundamental processes involved in gas breakdown and discharge. In our contemporary high-tech society, gas discharges are utilized more extensively than ever before in the area of materials preparation and processing. Continued research is essential if we are to remain competitive in the world market place and maintain our strategic position in the world economic and political arena.

It is interesting that while the study of gas discharges has contributed much to our body of scientific knowledge, only recently has technology advanced to the point where instruments have the capability to measure many phenomena of interest. Indeed, it is only recently that advances in the micro-electronics industry have led to the production of instruments that are ideally suited for the investigation of transient phenomena in rf excited discharges.

In addition to investigation of fundamentally important phenomena involved in gas break down and sheath formation, modulated or pulsed discharges have been shown to have a profound effect on processing and material deposition characteristics (1). For example, the band gap of amorphous silicon deposited from silane decreases when a pulsed discharge is used. This appears to be due to decreasing "dust" formation and the resulting decrease in pinholes (2, 3). Both electron density and the plasma chemistry are a function of the modulation frequency. The electron density between 10 and 5000 Hz increases to a maximum of 3 times that of the continuous wave discharge. This results in an increase in negative ion density. Alteration in the microscopic and macroscopic characteristics indicates that further investigation in modulated rf excited discharges is warranted.

Experimental Methodology

All of the experiments described in this report were conducted using the GEC reference cell which has been described in great detail elsewhere (4). Briefly, the stainless steel chamber contains two 4 inch diameter aluminum electrodes that are mounted on alumnia insulators. The electrodes have a 1 inch space between them. The chamber is kept at low pressure by a cryo pump when not in use. When in use the chamber is pumped with a molecular pump which has a membrane fore pump. The pressure is monitored with MKS baratron (0-2 torr).

An ENI LPI-10 rf amplifier is capacitively coupled to the upper electrode with no impedance matching network. The rf amplifier signal

input is provided by an HP 8657B frequency synthesizer which has the capability to modulated by an external TTL pulse. Typically, the modulation was at 10 Hz. Current and voltage waveforms were obtained using a Pearson 2878 probe and a Tektronix P6015 voltage probe (1000:1). Both waveforms were stored in either a LeCroy 9400 or a LeCroy 7200 digital oscilloscope. The current and voltage probes were placed at the working electrode after the coupling capacitor.

Results

A considerable amount of data was obtained on three gases, Ar, He and N_2. Complete data analysis is still underway and will continue for some time. The results presented here will focus upon the methods of analysis for a limited amount of the data.

The onset of gas breakdown is readily observed in the oscilloscope traces (Fig. 1). Figure 1 is digitized oscilloscope waveform for current and voltage for a 13.56 MHz rf input. It covers a time period of 200 μs (about 2700 periods) and consists of about 40,000 data points for each of the current and voltage waveforms. While the individual sine waves have been compressed to the point where they are not resolved, these waveforms show the overall process. Upon gas breakdown a decrease in the voltage amplitude with concurrent increase in the current and the development of the negative DC bias are observed. In the case of an argon plasma at 1.0 torr with an applied potential of 100 V, the current increases from 6.95 to 7.90 ampere from the time of break down to about 30 μs into the fully established discharge. The current

input is provided by an HP 8657B frequency synthesizer which has the capability to modulated by an external TTL pulse. Typically, the modulation was at 10 Hz. Current and voltage waveforms were obtained using a Pearson 2878 probe and a Tektronix P6015 voltage probe (1000:1). Both waveforms were stored in either a LeCroy 9400 or a LeCroy 7200 digital oscilloscope. The current and voltage probes were placed at the working electrode after the coupling capacitor.

Results

A considerable amount of data was obtained on three gases, Ar, He and N_2. Complete data analysis is still underway and will continue for some time. The results presented here will focus upon the methods of analysis for a limited amount of the data.

The onset of gas breakdown is readily observed in the oscilloscope traces (Fig. 1). Figure 1 is digitized oscilloscope waveform for current and voltage for a 13.56 MHz rf input. It covers a time period of 200 μs (about 2700 periods) and consists of about 40,000 data points for each of the current and voltage waveforms. While the individual sine waves have been compressed to the point where they are not resolved, these waveforms show the overall process. Upon gas breakdown a decrease in the voltage amplitude with concurrent increase in the current and the development of the negative DC bias are observed. In the case of an argon plasma at 1.0 torr with an applied potential of 100 V, the current increases from 6.95 to 7.90 ampere from the time of break down to about 30 μs into the fully established discharge. The current

appears to remain essentially constant after the first 30 μs beginning at the onset of discharge. The voltage waveform exhibits a simultaneous decrease in amplitude from 363 volts to approximately 290 volts during the same 30 μs period of time. In contrast the voltage undergoes slight changes in amplitude during the period 30 to 200 μs. The plasma appears to begin establishing a negative DC bias at the onset of discharge with the full DC bias established only after a period of in excess of 200 μs has elapsed.

All three gases exhibited the same general phenomenon. The "time constant" for establishing the DC bias on the order of 50 to 300 μs for Ar and He and 1 to 3.5 ms for N_2. This "time constant" was a function of the coupling capacitor which provides a measure of the time scale for breakdown and establishment of the resistive component of the discharge for the gases studied. It should be noted that these "time constants" were obtained by setting cursers on the oscilloscope to measure the time required to establish 70% of the DC bias. While these initial "time constants" are useful for a quick comparison, a more in depth analysis is necessary.

In order to determine if the curves that describe the current-voltage characteristics and the development of the DC bias are true RC time constants, it is necessary to show that the curves can be described by exponential functions. Curve fitting was accomplished using a variation of the nonlinear least squares algorithm by Marquet. Fitting the oscilloscope traces in a piece wise manner to a Fourier series allows one to determine the I-V amplitudes and the magnitude of

the DC bias as a function of the extent of discharge. This analysis completed for a several plasmas as a function of gas, power and pressure.

The DC bias as a function of the extent of discharge for argon and helium plasmas at 0.5 torr with 300 volts excitation applied are shown in Figs. 2 & 3. These plots show characteristics that are common to most of the conditions examined. In general, the entire curve can not be described by a single exponential function. This becomes more pronounced at higher power levels which required in some cases a 3 term exponential function to adequately describe the data. The DC bias curve for Ar plasmas at higher powers and higher pressures (0.5 torr and above) always fall rapidly to a minimum then slowly rise to the final DC bias potential. In contrast, none of the helium plasmas exhibited a minimum in the DC bias curve even at the highest powers and pressures examined. While one might suggest that the behavior exhibited by argon is possibly due to nonlinear behavior of the rf amplifier at higher powers, it remains to be explained why this does not occur for helium at high powers as well.

An approach which might provide meaningful results would be to do a piece wise fit throughout the establishment of the discharge. This approach would adequately describe a situation where as the gas breaks down conductivity is constantly increasing during the establishment of the discharge and would be characterized by a constantly changing time constant. However, this approach may not adequately describe plasmas that exhibit a minimum in the DC bias vs extent of discharge curve since

one would not expect to observe a discontinuity in this constantly changing time constant.

Since transient effects in helium plasmas appear to be less complicated than in argon plasmas, the more extensive examination of this data was chosen as the starting point. At low pressures and powers, the DC bias vs extent of discharge curves were adequately described by a single time constant. As pressure and power increased, two and the three term exponential functions appear to adequately describe the data (Fig. 2-4). To examine the effects of pressure and power on the development of a glow discharge, the initial time constants vs applied excitation voltage were plotted (Fig. 5). The general characteristics of these curves, that the time constant decreases as pressure decreases and as power increases, appears to be valid for all three gases.

The voltage and current amplitudes for the early part of the gas break down has been examined for the argon discharge at 0.5 torr with an excitation voltage of 300 volts (Fig. 6). The initial current is 17.42 ampere and the initial voltage is 695. In order to over lay the plots, a constant value was subtracted from both current and voltage. Both curves can be fit to a time constant of about 7 μs, though it should be noted that this constraint was imposed in the fitting process.

When the rf excitation is switched off, the DC bias follows an exponential recovery. The time constant for this recovery is largely governed by the external circuitry, i.e. the coupling capacitor and any external resistance to ground. For example, connecting the working

electrode to ground through a 20 kΩ resistor provides a time constant consistent with the product of the coupling capacitance with the external resistance. Pulsed rf excited plasmas may be very useful in semiconductor processing where the magnitude of the DC bias impacts upon device characteristics. The effect of modulation frequency vs DC Bias voltage is shown in Figure 7.

References

1. J.T. Verdeyen, J. Beberman, & L. Overzet, J. Vac. Sci. Technol. A, 8(3), 1851, 1990.

2. Y. Watanabe, M. Shiratani, Y. Kubo, I. Ogowa, & S. Ogi, Appl. Phys. Lett. 53, 1263, 1988.

3. G. S. Selwyn, J. Singh, & R. S. Bennett, J. Vac. Sci. Technol. A 7, 2758, 1989.

4. J. R. Roberts, Ed., GEC RF Reference Cell Newsletter, 3(1), 1991.

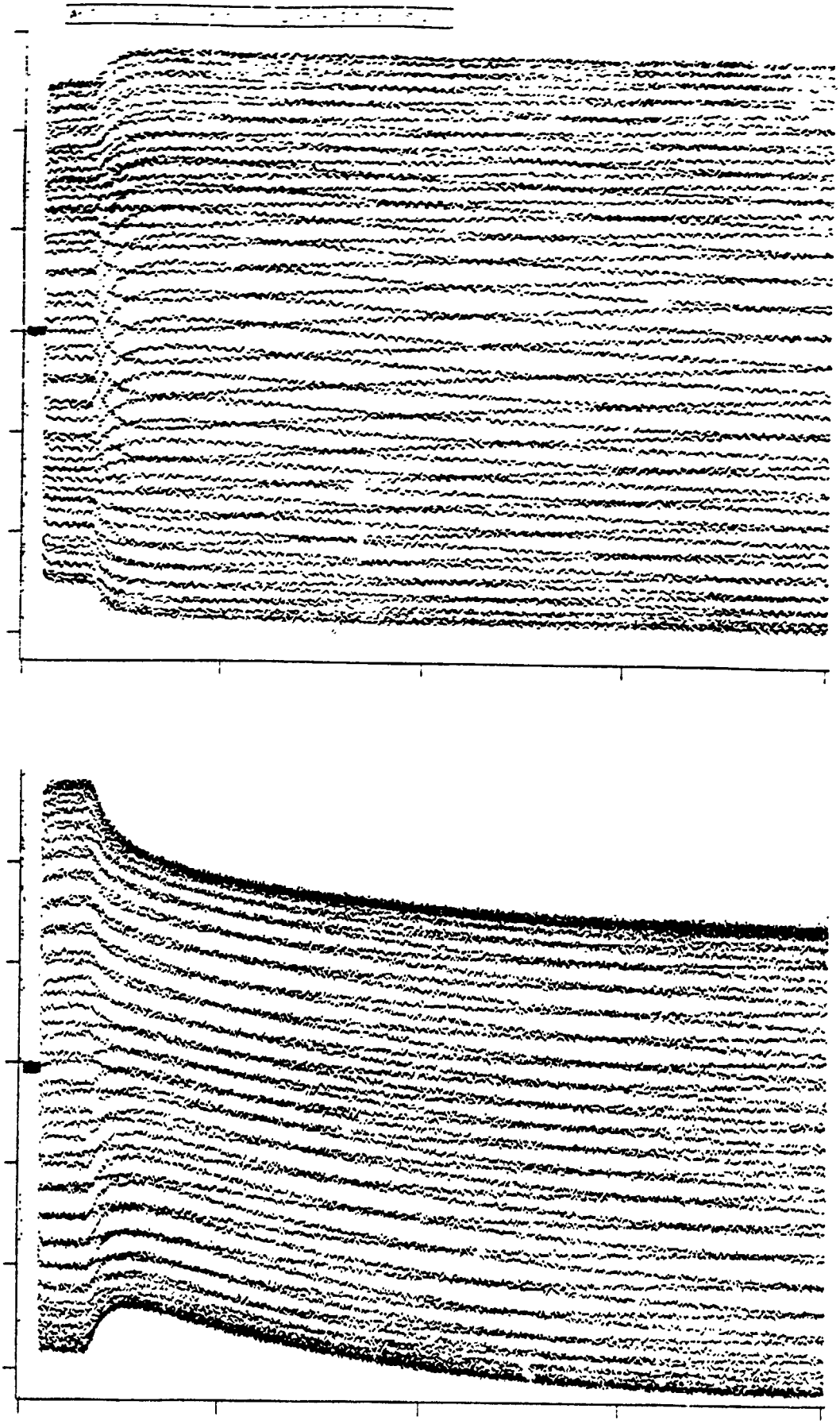


Figure 1. (Top) Current waveform, (Bottom) Voltage waveform.

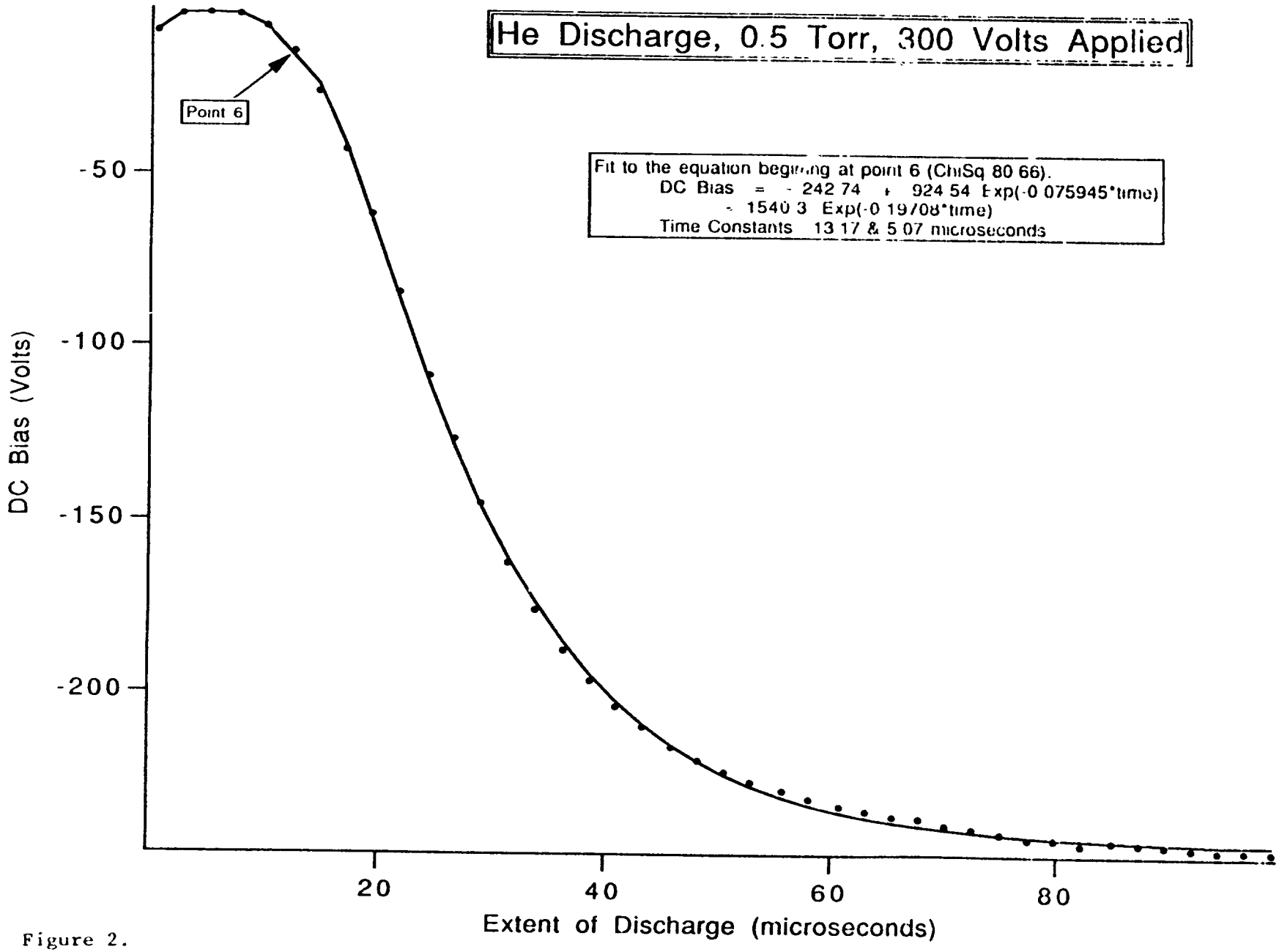

He Discharge, 0.5 Torr, 300 Volts Applied
Point 6
Fit to the equation beginning at point 6 (ChiSq 80 66).
DC Bias = - 242 74 + 924 54 Exp(-0 075945*time)
- 1540 3 Exp(-0 19708*time)
Time Constants 13 17 & 5 07 microseconds
DC Bias (Volts)
-50
-100
-150
-200
20
40
60
80
Extent of Discharge (microseconds)

Figure 2.

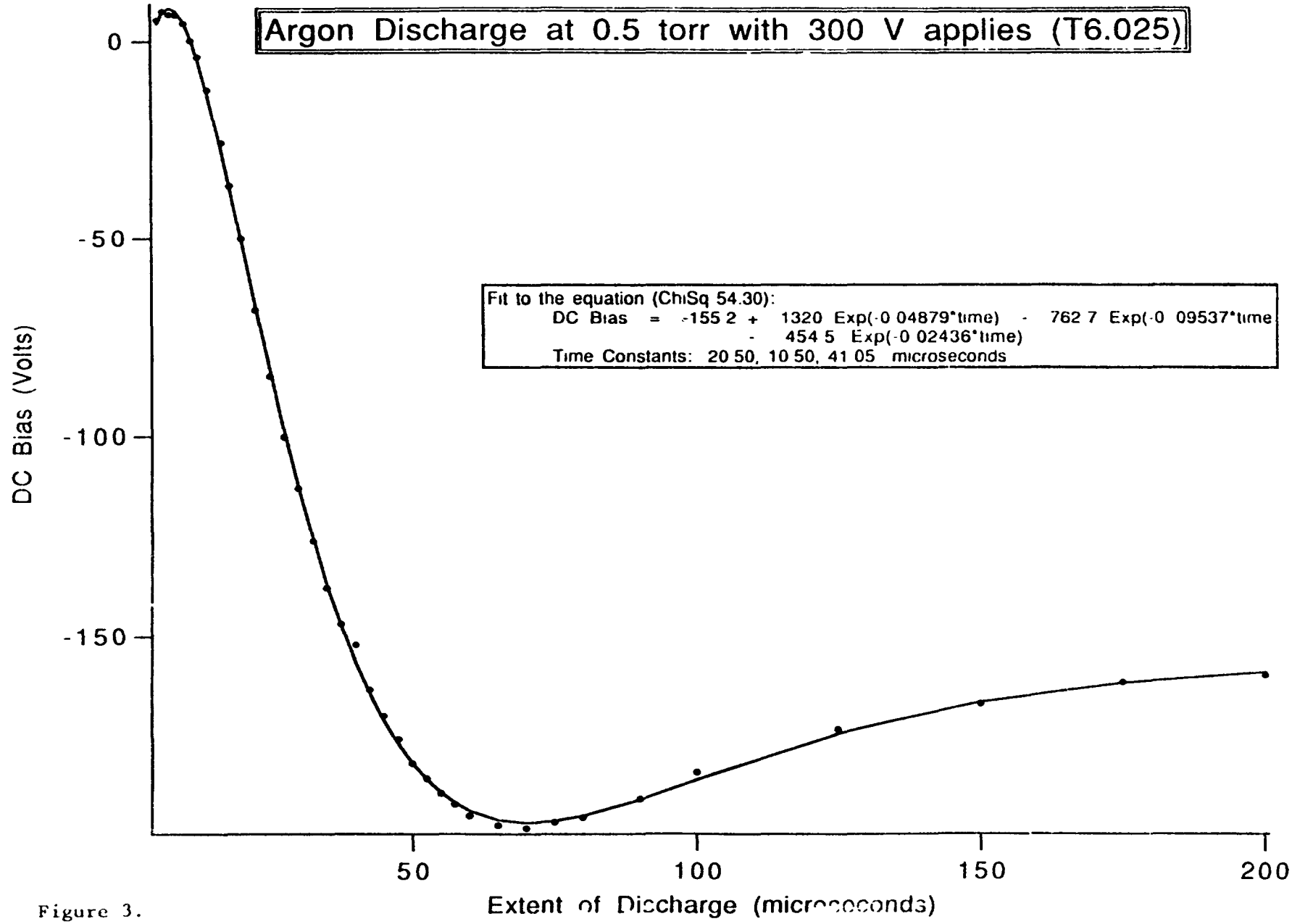

Argon Discharge at 0.5 torr with 300 V applies (T6.025)
Fit to the equation (ChiSq 54.30):
DC Bias = -155 2 + 1320 Exp(-0 04879*time) - 762 7 Exp(-0 09537*time
- 454 5 Exp(-0 02436*time)
Time Constants: 20 50, 10 50, 41 05 microseconds
0
-50
-100
-150
DC Bias (Volts)
50
100
150
200
Extent of Discharge (microseconds)

Figure 3.

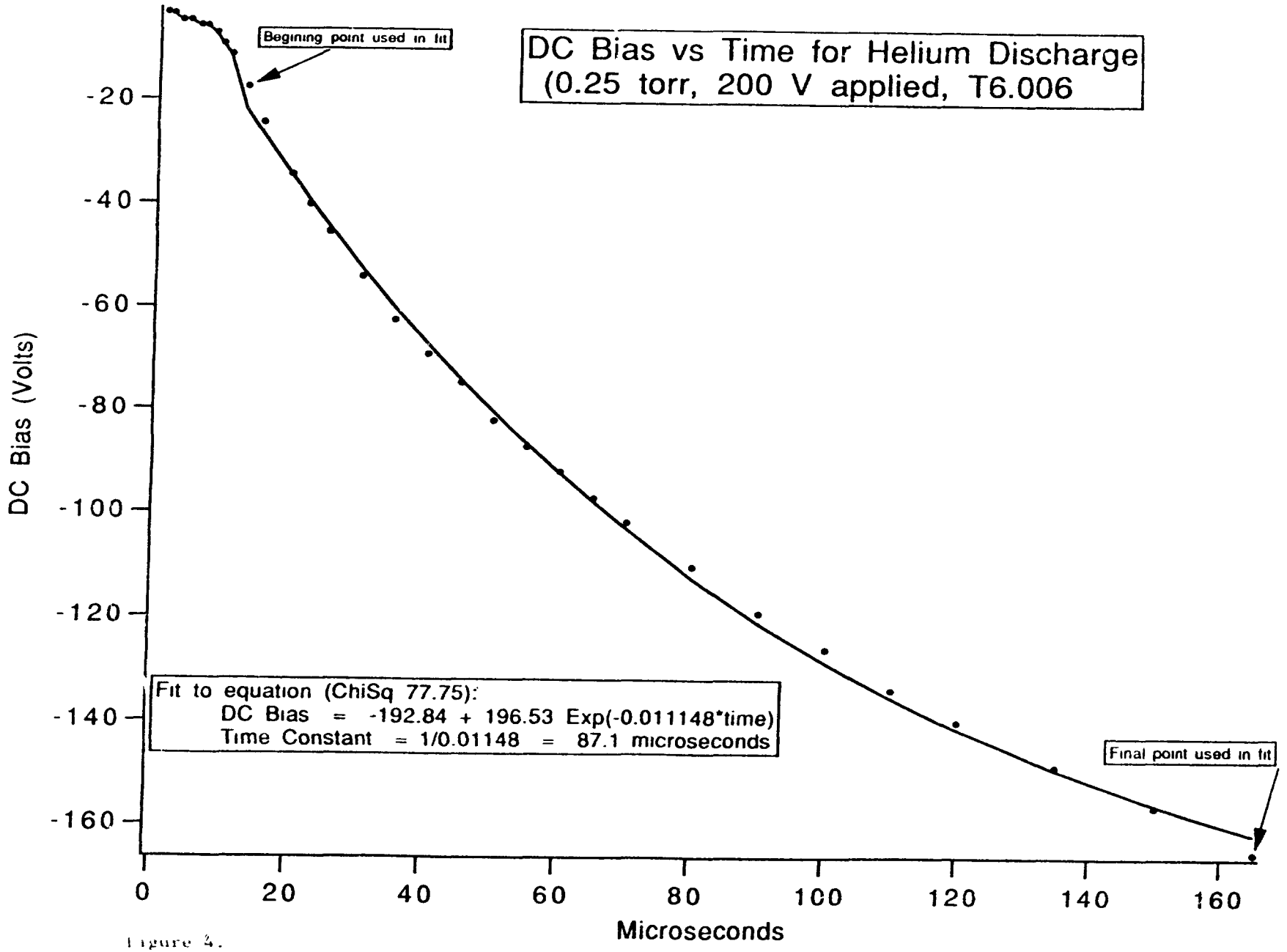

DC Bias vs Time for Helium Discharge
(0.25 torr, 200 V applied, T6.006
Begining point used in fit
Final point used in fit
Fit to equation (ChiSq 77.75):
DC Bias = -192.84 + 196.53 Exp(-0.011148*time)
Time Constant = 1/0.01148 = 87.1 microseconds
DC Bias (Volts)
-20
-40
-60
-80
-100
-120
-140
-160
0
20
40
60
80
100
120
140
160
Microseconds

Figure 4.

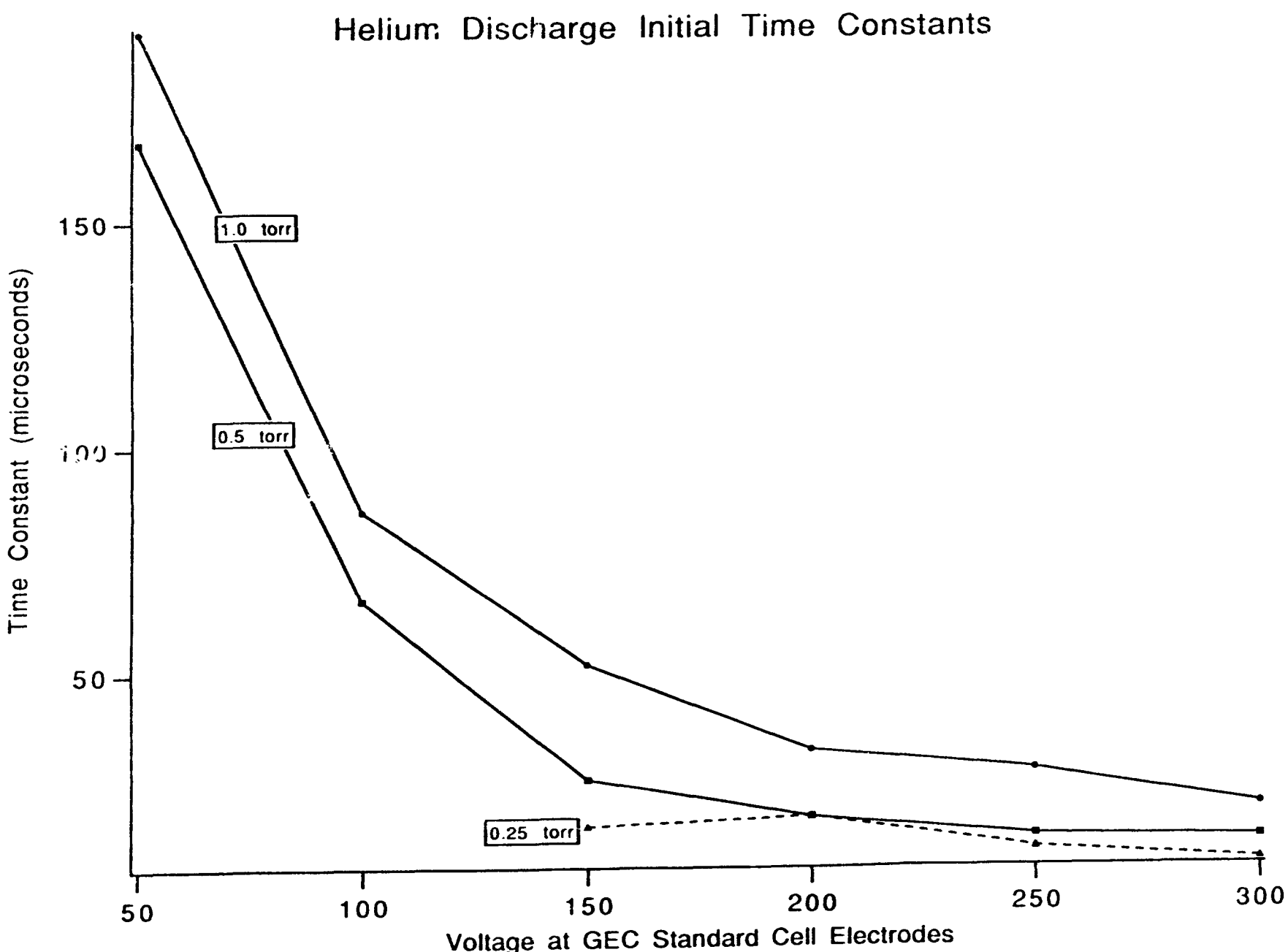

Helium Discharge Initial Time Constants
Time Constant (microseconds)
150
100
50
1.0 torr
0.5 torr
0.25 torr
50
100
150
200
250
300
Voltage at GEC Standard Cell Electrodes

Figure 5.

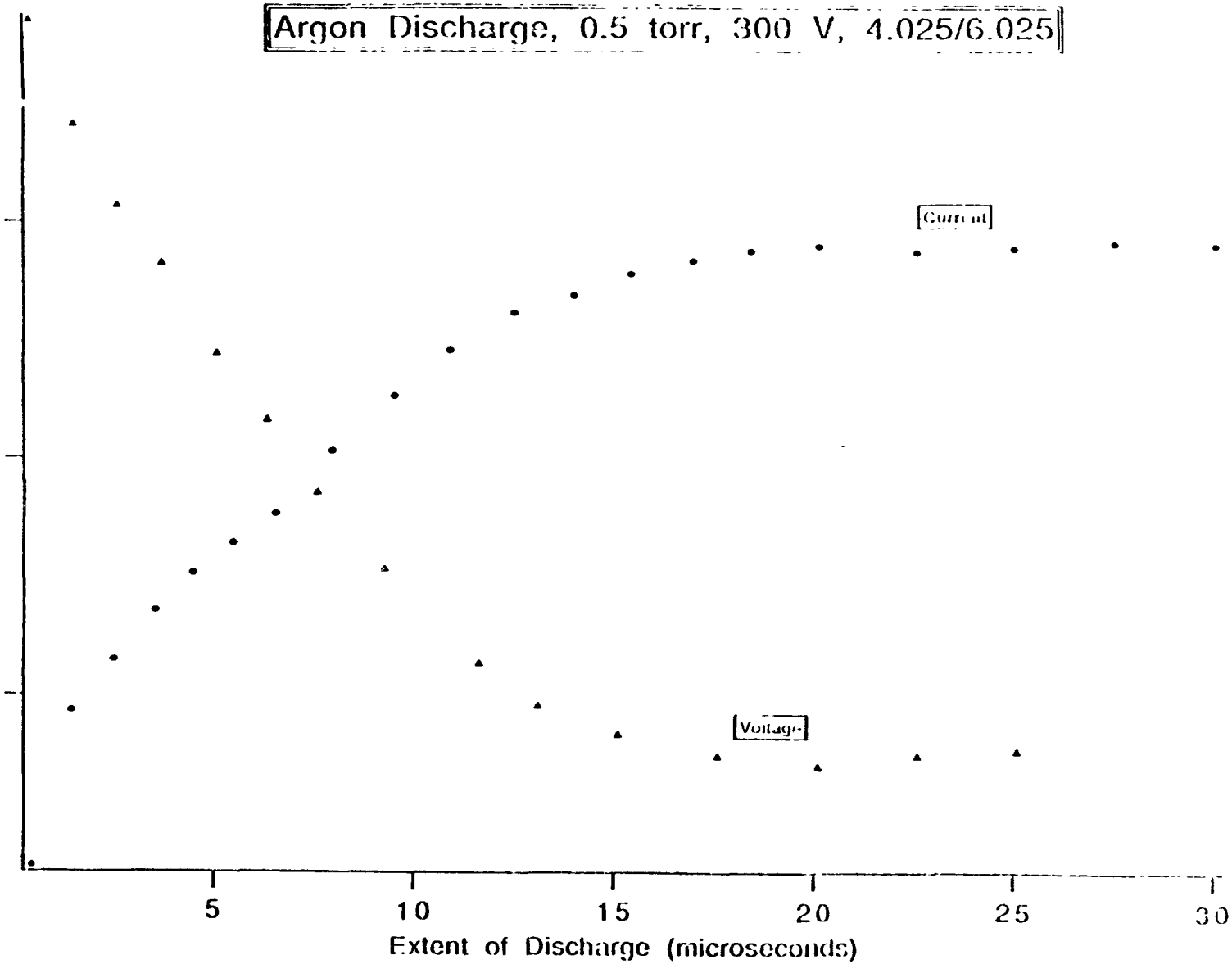

Argon Discharge, 0.5 torr, 300 V, 4.025/6.025
Current
Voltage
5
10
15
20
25
30
Extent of Discharge (microseconds)

Figure 6.

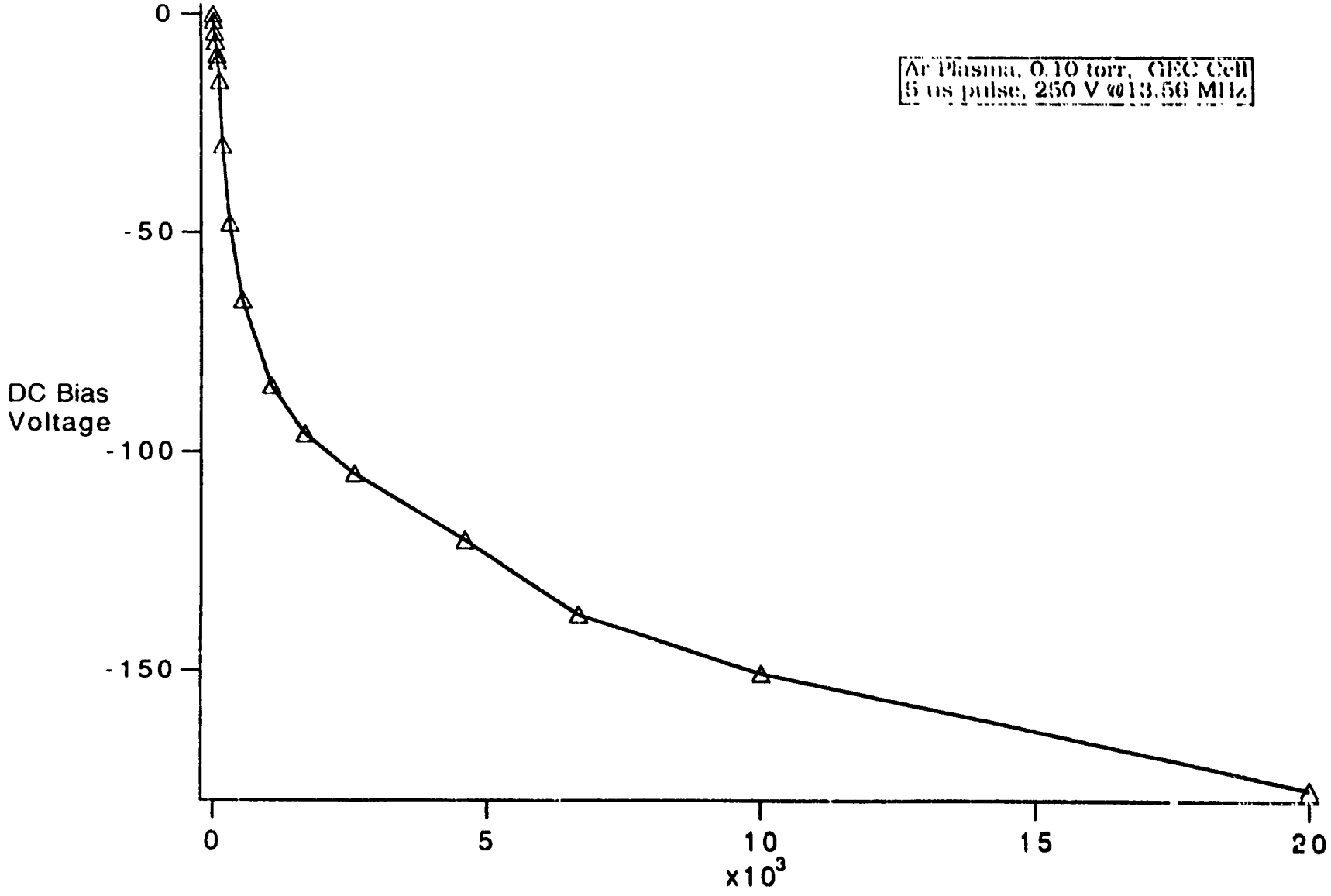

Ar Plasma, 0.10 torr, GEC Cell
5 us pulse, 250 V @13.56 MHz
DC Bias Voltage
0
-50
-100
-150
0
5
10
15
20
x10^3
Modulation Pulse Frequency

Figure 7.

SIGNAL PROCESSING FOR HIGH SPEED VIDEO TECHNOLOGY

Eugene R. Chenette, Professor
University of Florida
Graduate Engineering and Research Center
Eglin Air Force Base, Florida

Abstract:

This report presents the results of several experiments on the characteristics of video signals and of the performance of electronic systems used to process those signals. High Speed Video Technology has the potential for providing new and innovative tools for scientific and engineering research. Electronics components and systems capable of performing at the rate required for HSVT are only now becoming available. The need for large area high resolution imagers was one of the early prime movers motivating intensive research to build silicon with dimensions precisely controlled at the micrometer and submicrometer level. And the development of imaging arrays with multi-millions of pixels, with every pixel operating, and the dimensions of every pixel sensibly identical is one of the great achievements. The bottom line is that an HSVT camera meeting all of the requirements defined by WL/MNGI can be built. Building it will develop an engineering knowledge base that will be of value for many years.

Introduction:

This report presents the results of recent experiments related to the problem of dealing with the tremendous amount of data which will be produced by an operating High Speed Video Technology (HSVT) camera head. The definition of HSVT, in the context of this report, is a video system capable of megapixel resolution with frame rates on the order of 1000 frames per second. To help explain the difficulty of the technical problems involved, this is a system with video resolution comparable to that of commercial High Definition Television (HDTV) with data rates more than 30 times faster than those of HDTV.

One of the important applications proposed for HSVT has to do with

measurement of motion during the first few seconds following release of a weapon from a fast moving airplane. Workers at Wright Laboratories/MNGI at Eglin Air Force Base have established that this application requires video resolution of at least 1024(H) by 512(V) picture elements (pixels); a frame rate of at least 920 frames per second; and exposure time for each frame of not more than 220 microseconds. [1,2]

This is a realistic goal. The system cannot be assembled with off-the-shelf components available today, but the technologies required to do what needs to be done have been demonstrated. Indeed, several laboratories have recently announced the development of high performance imaging arrays with much higher resolution than mentioned above; and they are operating with clocking signals on the chip much faster, higher in frequency, than those needed for HSTV. For example, the Toshiba Corporation has announced a two million pixel CCD image sensor, typical of those being developed for use in professional cameras for high definition television (HDTV) broadcasting. Toshiba reports this imaging array has achieved a sensitivity of 210 nanoAmperes/lux and a dynamic range of 72 decibels. [3,4]

The idea of using parallel data outputs has been implemented in other high speed focal plane arrays. [5,6, 7] A quick calculation shows the results of dividing an array of 1024(H) by 1024(V) picture elements (pixels) into 64 subarrays, each with only 32(H) by 512(V) pixels. The data rate corresponding to a frame rate of 1000 frames per second is reduced to only 16,384,000 pixels per second. The actual data rate will be slightly higher during active data periods. One simple model, under the assumption that all pixel to pixel transfers occur at the same rate yields the frequency of 17,408,000 pixels per second. It is not important to know the exact number at this time. What is

important is to recognize that parallel architectures make data rate problems manageable. The data rate per port required to build an HSVT system is not much faster than data rates of standard NTSC and monochrome television, and they are much less than those required for single port HDTV imagers.

However, the use of parallel outputs does not solve the problem of dealing with the enormous amount of data which will be produced. The performance required of the HSVT system dictates that data at each of the 64 output channels be converted to digital format. Data accumulates at the rate of 1,048,576,000 words per second. Much of the discussion which follows has to do with design of the analog to digital conversion system and a system for lossless data compression.

The real problem is to package the entire system in a cylindrical volume with a diameter of 12.75 centimeters and length 91 centimeters. Success is going to require dense three dimensional packaging of very low power dissipation electronics.

The most important result of the work related to this report is a long and growing list of things that need to be done, problems that need to be investigated in greater detail. Much of the work requires more and better computer-aided test and simulation equipment. Some requires the design and fabrication of new and more versatile test fixtures for operation and testing of micromodule systems.

Experimental Results

Figure 1 shows a block diagram of a proposed "Test Bed for High Speed Video." Some have called this idea a "cathedral for the development of high speed video technology. It is a cathedral in the sense that it it will never be finished. It is a place where ideas, components, and algorithms can be evaluated. It must evolve as technology continues to evolve. But it must also be

a smoothly operating high performance test facility where engineers can find answers to questions about the next generation of HSVT.

Little more need be said about the test bed at this time. The "test bed" for most of the experiments of this report was an industrial monochrome video camera. (PULNIX TM-7). For most experiments the test scene was a precision optical test pattern. The test scene and optics were used to control the signal at the output of the Focal Plane Array. Illumination of the scene under test was measured with a Minolta Model LS-110 Luminance Meter.

The video output signal of the PULNIX TM-7 is very similar to that which will be available at each of the 64 outputs of the HSVT focal plane array. Data formats are different but the shape and speed of the signal produced by each pixel must be very similar. The pixel clock rates (14.381818 and 17.408 Megahertz) close enough together that an A/D system demonstrated at one frequency has a high probability of operating with the same precision at the other.

The analog signal processing function for the SONY TM-7 is handled by a small 32 pin surface-mount integrated circuit, the SONY CXA1310AQ. The signal from the output of the ICX038AL image sensor in the PULNIX TM-7 goes first to a low noise JFET source follower. The output from that source follower is a faithful reproduction of the signal from the CCD array; the video signal is buried in clock noise. The CXA1310AQ provides the circuits which performs correlated double sampling, it provides for automatic gain control and gamma correction, and produces, at its output, a complete video waveform.

Figure 2 is a set of graphs showing waveforms of the signal at the video output of the Model TM-7 camera. These waveforms were obtained with a Tektronix Model RTD 710A Digitizer. Illumination was the same for both sets of data and varied slowly from one side of the test scene to the other. Data for

Figure 2a were collected with a sampling period of 10 nanosconds per point. The graph of the 8192 sample output shows an 81.92 microsecond segment of the video signal, slightly more than the length of one complete horizontal line. The sampling period for Figure 2b was increased to 100 nanoseconds. The result is a graph of an 8.192 millisecond segment of the video signal, ten times the length of Fig 2a.

The spectral density of both signals was calculated with the help of Tektronix signal processing software. Spectral resolution for Fig 2a is about 12 kiloHertz per point and for Figure 2b is about 1.22 kHz per point.

Figure 3 shows data that are similar to that of Figure 2a. The difference is that the test scene here is a uniformly illuminated precision sinusoidal test pattern. Data shown here could be used to prepare a graph of the modulation transfer function (MTF) of the system. The period or frequency of signals in the video signal can be expressed in either spatial or temporal units. The lowest frequency sine wave on Figure 3a has a period of about 4.5 microseconds or 64 pixels. The lowest frequency signal is Fig. 3b has a period of about one microsecond or a frequency of one MegaHertz. The frequency of every second segment doubles. Hence the frequency of the third segment is about two MegaHertz, the frequency of the fifth about four MegaHertz. The frequency corresponding to the Nyquist limit (7.19 MegaHertz) is in the region between the frequencies of segments 7 and 8. Figure 4 shows similar data but with the test scene a well illuminated square wave test pattern. These results could be used to determine an experimental contrast transfer function (CTF).

Figure 5 shows a comparison of data of low frequency spectra obtained by using two different methods. Figure 5a was obtained with the help of the Tektronix Signal Processing software operating on data collected with the Tektronix RTD 710A digitizer. Data of Figure 5b were obtained using a tunable

narow band analog wave analyzer. Figure 5 b shows two sets of data. The set with higher amplitude was obtained with the camera illuminated with a flat field and and illumination adjusted so that the video level was about one half the value required to saturate the array. The other set of data shows the noise level of each of the spectral components of the low frequency signal with no illumination.

EG&G has provided an evaluation board (RC0504ANN) complete with a Reticon RA2568 N array for review. This evaluation board includes analog signal processing for all eight channels. This is another version of the "test bed." All that was required to make this unit work was to connect power supplies. Operation appears to be exactly as described in the instructions provided by EG&G. To date the system has been operated only with diffused light falling on the photoactive area. No lens has been installed and no attempt has been made to produce data corresponding to an image. Figure 6 shows a comparison of the FFT spectra obtained for both a "dim" and a "well-illuminated" array.

Figure 7b shows the result of an "old-fashioned" test of the performance of an Analog Devices AD9060 10-Bit, 75 MSPS A/D Converter. Figure 7a shows a block diagram of the system. The unit under test was mounted in AD9060/PCB Evaluation Board provided by Analog Devices, Inc. The idea of the measurement is to consider the A/D converter and the companion D/A converter as a complete analog system. Evaluation of the performance if the converter is reduced to searching for errors in the output analog waveform, making measurement of the frequency dependence of analog small signal gain and measuring distortion and noise.

A laboratory pulse generator was used to supply the ECL level signals

required to enable the unit. The sinusoidal input signal applied to the input of the A/D converter was supplied by a precision waveform generator. The output signal was analyzed using the Tektronix RTD 710A digitizer, and the results were as shown in Figure 7. This performance is consistent with specifications listed in the manufacturers data sheets.

Much useful information about the performance of an A/D converter can be obtained with this straightforward approach, but more complete and thorough testing requires the use of a computer based measurement system. such as is shown in the block diagram of Figure 9.

Signal Processing

It's appropriate at this point to discuss the HSVT signal processing problem. Consider a focal plane array with a total of I columns and J rows of picture elements. $p_{i,j}$ is the pixel located at the intersection of the ith column and the jth row. $P_{i,j}$ photons incident upon the photoactive area of this pixel produce $n_{i,j}(T_e)$ electrons during the exposure interval T_e. The details of the process for measuring the charge in each pixel are different for different focal plane array architectures and for different video formats. What's important is that at some time during the readout process the pulse train will include an output voltage pulse $v_{i,j}(T_e)$ which is an accurate measure of $n_{i,j}(T_e)$ and hence of the illumination of $p_{i,j}$.

The video output signal, for each frame, of an I column J row focal plane is a time dependent sequence of voltage pulses of length I x J. The magnitude of the jth pulse in the ith line is proportional to the number of photons incident upon pixel $p_{i,j}$ in the time interval T_e. The magnitude of each pulse in the video signal train must be converted to a digital word which is an adequate representation of v(i,j) and hence of n(i,j) and P(i,j). The factors which limit the dynamic range of a focal plane array are well known. The largest magnitude

video pulse allowed is determined by the maximum charge storage capacity of the pixels. The minimum signal which can be detected is limited by dark current and fluctuations inherent in the operation of the FPA. Noise of the output charge to voltage converter may limit the performance in the case of a very well designed array.

For the purpose of this discussion it is sufficient to note that several laboratories have announced high resolution focal plane arrays with the dynamic range in excess of 70 db. The dynamic range of an N bit analog to digital converter with full scale sine wave input is given in decibels by the expression

$$SNR = 6.02\,N + 1.76 + 10\log(f(s)/2f(a)) \quad (1)$$

Here f(s) is the sampling rate and f(a) is the analog bandwidth of the signal being sampled. This expression is subject to the requirement $f(s) \geq 2f(a)$. Note that SNR can be improved by limiting the analog bandwidth. A 3 db improvement can result when $f(s) = 4f(a)$.

Eq (1) leads immediately to the idea that a focal plane array with a dynamic range of 70 db justifies the requirement for use of 12 bit analog to digital converters. Unfortunately this requirement exceeds the performance of all known A to D converters which are physically small enough to fit in a practical HSVT camera head.

If it were necessary today to make final decisions about the components to be used in the HSVT camera head the choice would be between eight bit and ten bit converters. Other things being equal, the converter which dissipates the lowest power would be the one chosen.

The HSVT system is a sampled data system. The image incident on the focal plane array is sampled spatially. The signal at the output of the array is buried in clock noise. Correlated double sampling is often used to recover the

true video signal . With proper processing the signal available at the output of the analog processing block of the test bed is a good approximation of the classic sample and hold waveform. This is the signal which must be converted to a stream of digital words by a well designed A/D converter.

Both the spatial sampling process on the focal plane array and the temporal process of the electrical A/D converter are subject to the fundamental requirement

$$f(s) \geq 2f(a).$$

This problem, and the solution, are well known [8,9,10,11.12] The solution is to place an analog low pass filter ahead of the input to the sampled data system, and to design that low pass filter to insure that energy in the frequency range which will be aliased into the active signal is limited so that the results of aliasing do not degrade the performance (dynamic range) of the system. Fortunately this is relatively simple for an electrical signal path. Unfortunately it's impossible inside the imaging array.

Figure 10 shows an example of a circuit suggested by TRW, Inc. on the data sheet for the TMC digital decimating filter. Two things are important here. One is the five pole passive analog filter shown between the video source and the input to the TDC 1049 A/D converter. The second important idea is the use of oversampling. The SNR of an oversampled system is given by

$$SNR = 6.02N + 1.76dB + 10 \log[f_s/2f_a]$$

Doubling the sampling rate, from 14.38 to 28.636 MHz has the potential of increasing the dynamic range of the A/D conversion system by 3 dB.

The only tool available to the designer of a video system to optimize the performance in the dynamic range/resolution domain is the use of optical prefilters. The decision must be application dependent and can only be made

on the basis of empirical data.

The need for and results of processing video data are application specific. Criteria for the design of a video system to be used for (computer-aided) study of motion may be much different than those for the design of a video system intended primarily to produce images for viewing by humans for entertainment. The test bed, when complete, will provide a facility for evaluation of HSVT components and algorithms in the context of the WL/MNGI mission.

Data Compression

Workers at the NASA Space Engineering Research Center at the University of Idaho, [13] working in cooperation with people from JPL and the Goddard Space Flight Center have developed a chip set which may be exactly what is needed for the WL/MNGI HSVT system. This chip set is a CMOS implementation of the Rice Algorithm for lossless compression and decompressions of digital data. Figures 8a, 8b, and 8c are a set of block diagrams to help explain this encoder/decoder system. Everything known about the performance of this chip set is based on the results of computer simulations. Fabrication of the first chips is being completed while this report is being written. The encoder chip is designed to operate at a 20 Megaword per second data rate with words ranging up to 14 bits. Power dissipation is expected to be less than 1/4 watts per chip. The decoder is not as fast the encoder, but it is designed so that two decoders can be operated in parallel on a single channel if an application requires real time decoding of data at the maximum encoder rate.

Figure 8d is a block diagram showing the position of the encoder and decoder in a typical system. What's important here that the operation of the encoder and decoder require that the compressed data be placed in packets. The design of the packetizer must be tailored to the system.

Conclusions:

This report has presented the results of several experiments investigating the characteristics of typical video signals and of the performance of electronics systems used to process those signals.

The major problems that stand in the way of success in building a practical HSVT camera have hardly been mentioned. The engineering challenge is that of packaging the unit so that it will fit in a cylinder 12.5 centimeters in diameter and 91 centimeters long. In principle doing electronics with components in three dimensional packages with high density component count is no more difficult than it is to work on two-dimensional printed circuit boards. In practice, this requires new tools and new skills. To succeed in doing this kind of electronics means that new tools must be bought and designed and built, and that people must learn to use these new tools.

As this work continues the thing that must be done first is to build or buy a system similar to that shown in Figure 9. We must be able to evaluate the performance of the A/D converters; to verify the precision of operation of data compression hardware and software; and to verify with digital precision the effect of changing the optical scene or installing an optical prefilter. We must also be able to simulate the performance of large mixed-mode (analog/digital) systems before they are built.

Acknowledgements:

Thanks to AFOSR and RDL for the support which made is possible to work on the engineering problems of High Speed Video Technology. Thanks to Don Snyder and the people of WL/MNGI for providing an interesting problem and making all the facilities of the Special Projects Laboratory available. Thanks to my young associates, Laurence Flath and Martin Sensiper for help with measurements and always being a source of encouragement.

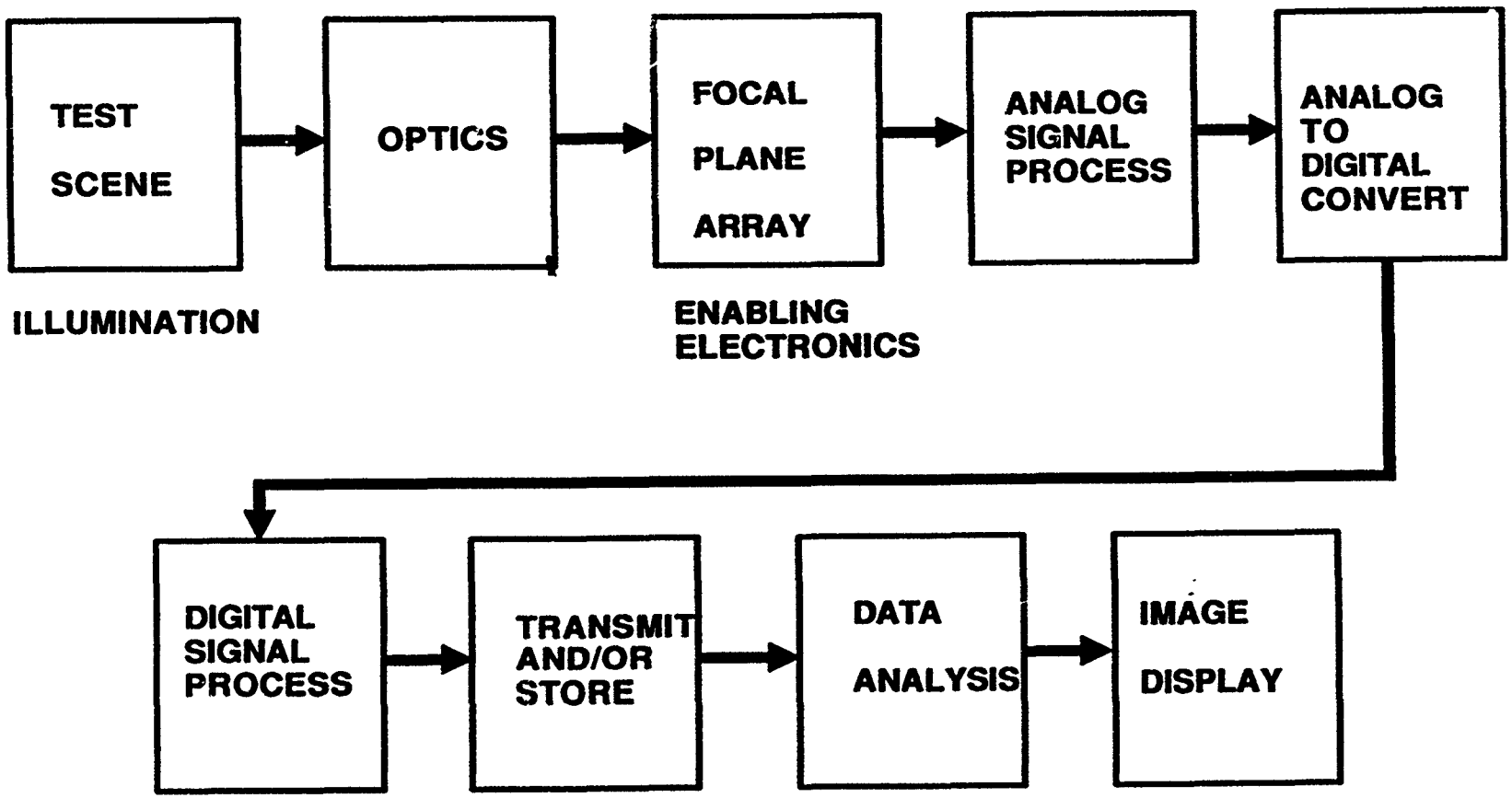
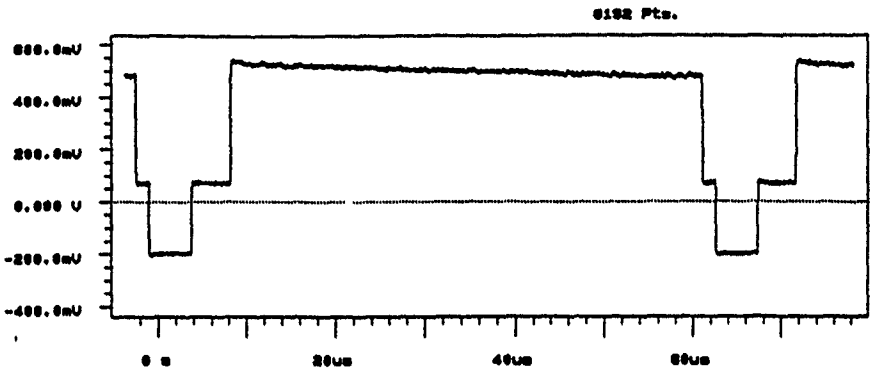
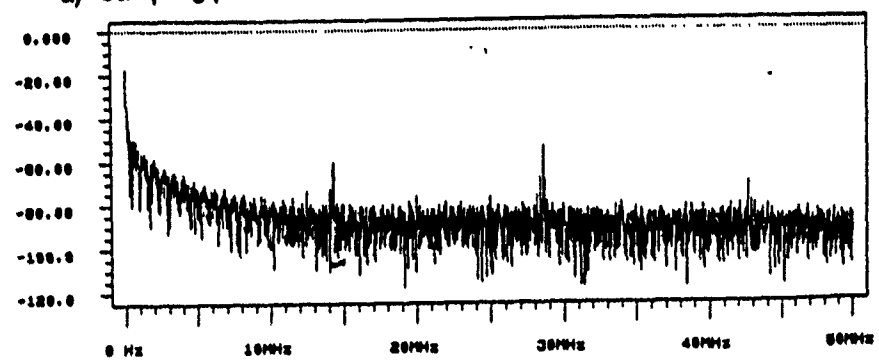


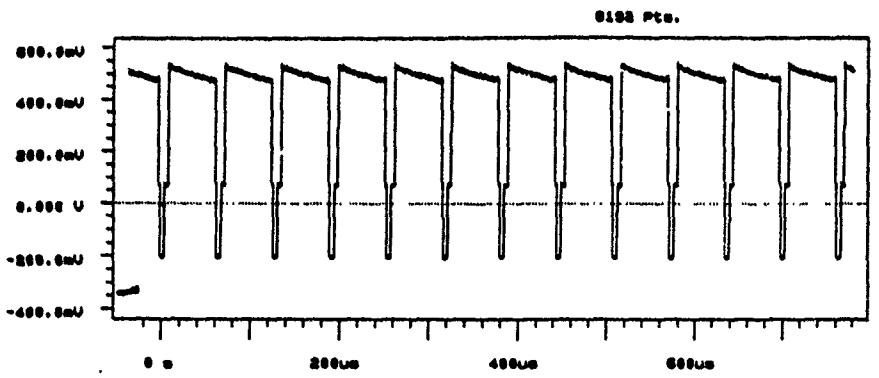
Figure 1

TEST BED FOR HIGH SPEED VIDEO TECHNOLOGY



a) Sampling period of 10 nanoseconds, 8192 points.





b) Sampling period of 100 nanoseconds, 8192 points.

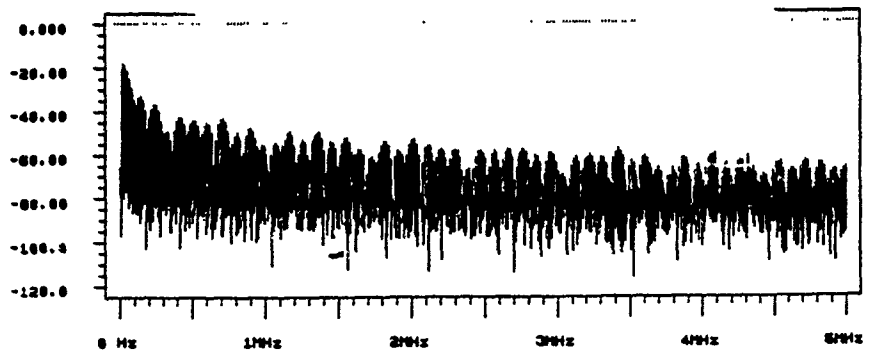


Figure 2. The voltage waveform of a video signal and the spectrum of that signal for samples of two different sampling periods.

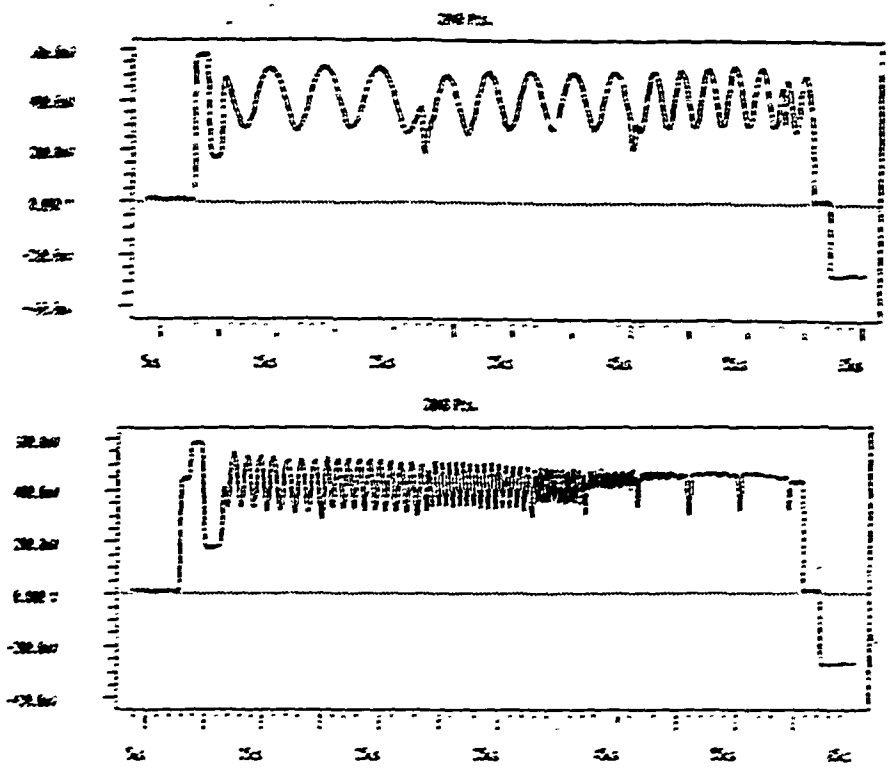


Figure 3. The video waveform of a single line when the scene viewed by the camera is precisely sinusoidal in the horizontal direction

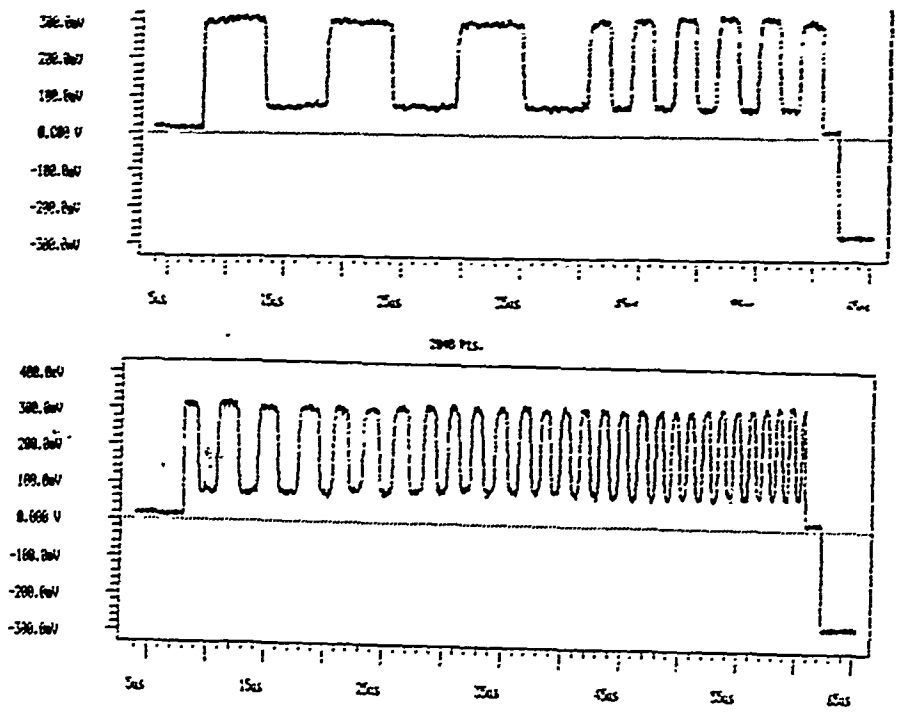
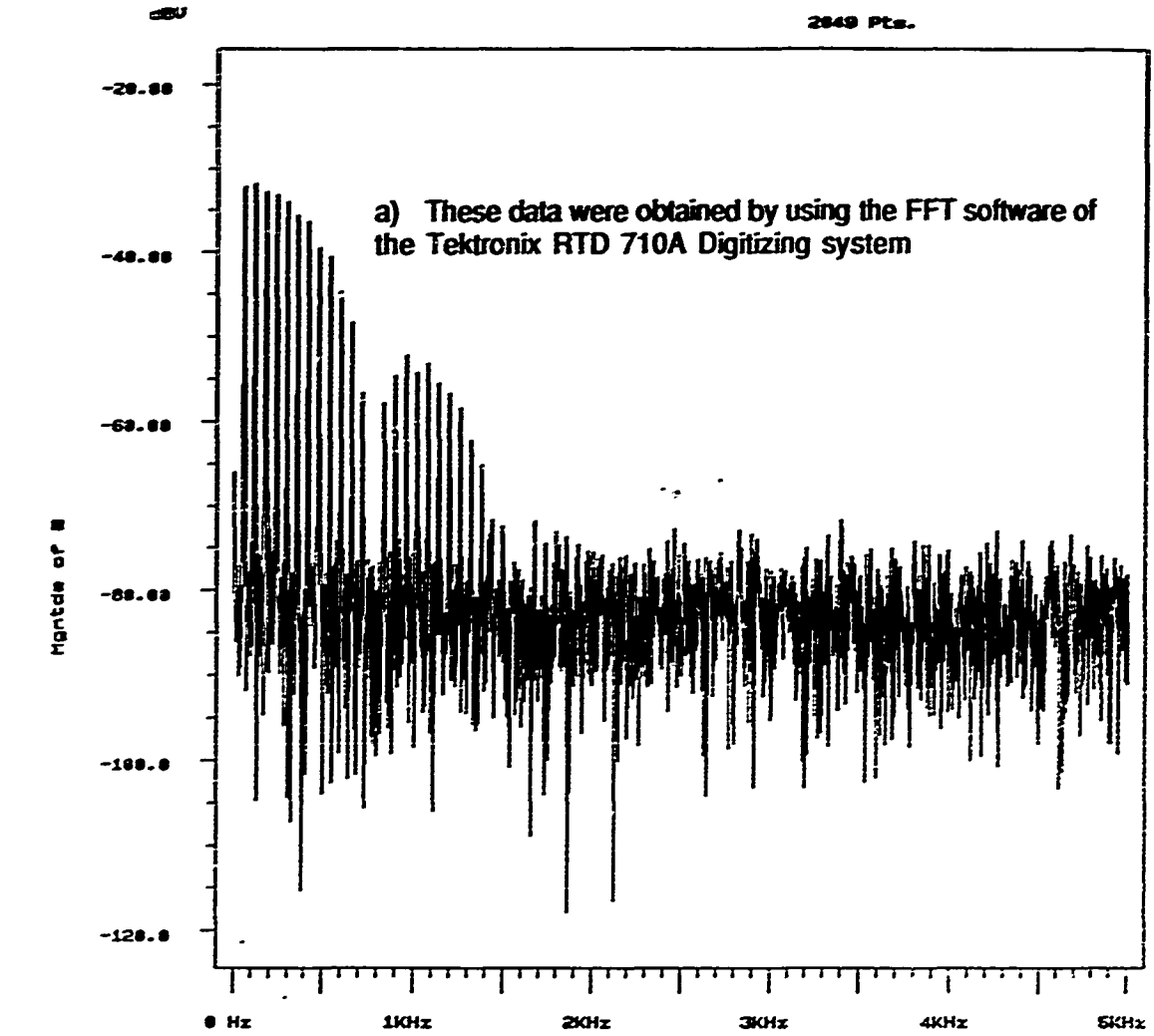


Figure 4. The video waveform when the scene has precise square-law spatial dependence.



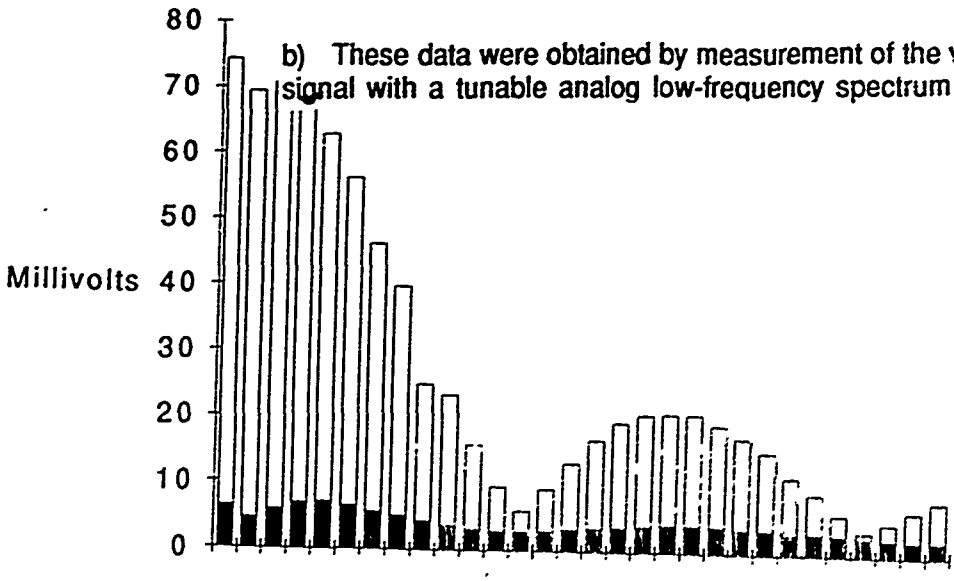


Figure 5. Comparison of low frequency spectra measured by two different methods.

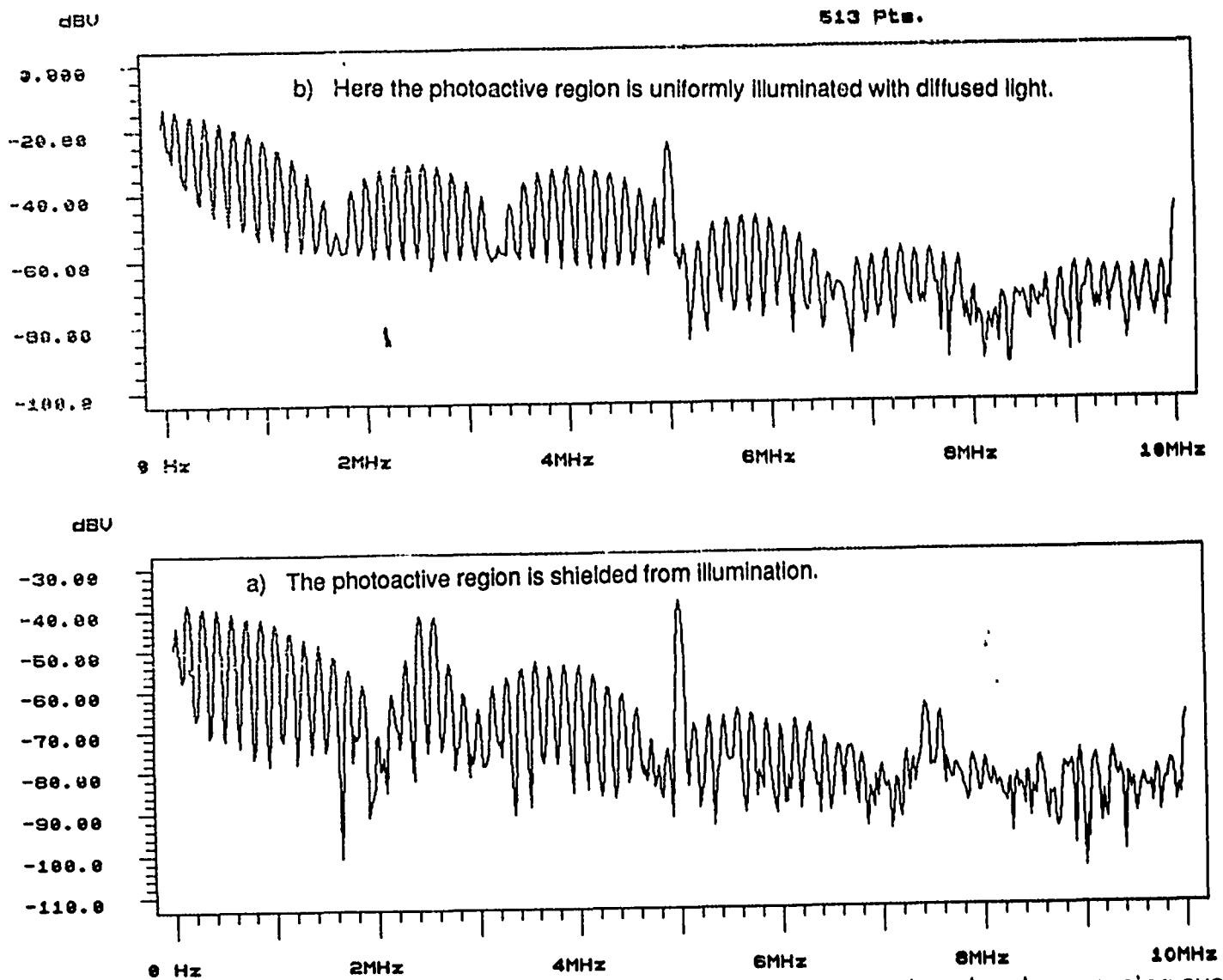


Figure 6 Spectrum of the video signal at the output of the analog signal processing system of the EG&G RC00504ANN Evaluation Board. The pixel clock rate is five Megahertz.

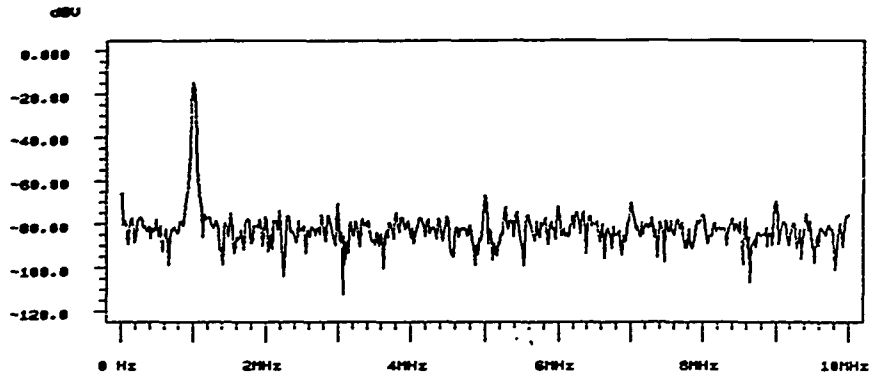


Figure 7b. The spectrum of the signal at the output of the Analog Devices AD6090 "10 bit 75MSPS A/D converter" Here the clock rate was set to 20 MHz. The input signal was a 1 MHz sinusoid.

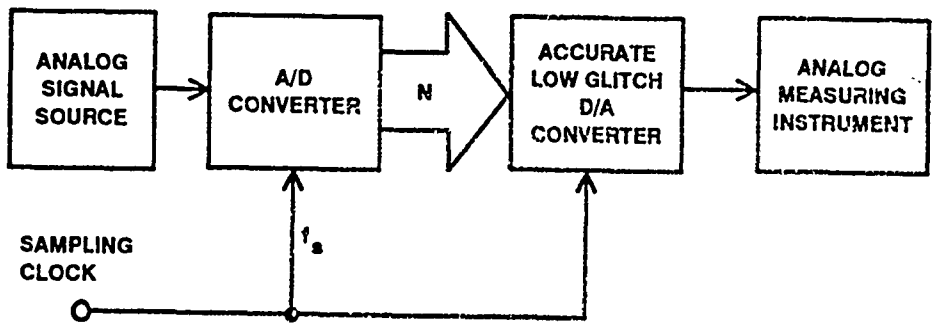


Figure 7. The system used to measure the performance of the A/D converter.

The evaluation board , including the accurate low glitch D/A converter was provided by Analog Devices, Inc.

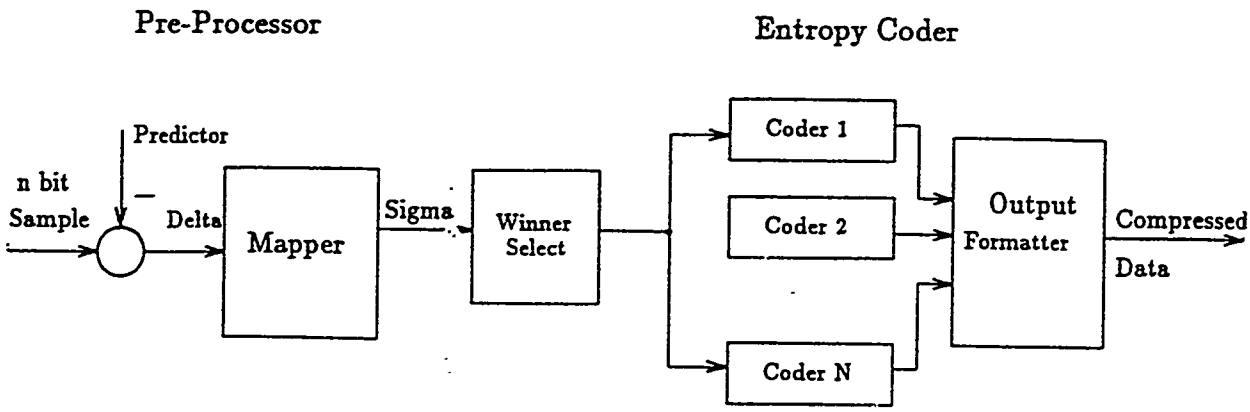


Figure 8a. Block diagram of the architecture of the Rice Algorithm.

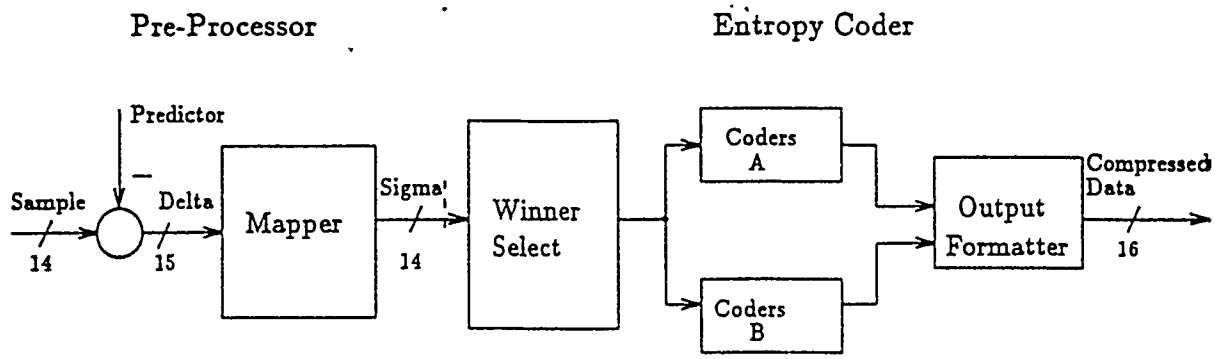


Figure 8b. Block diagram of the NASA /University of Idaho Silicon Data Compression Chip.

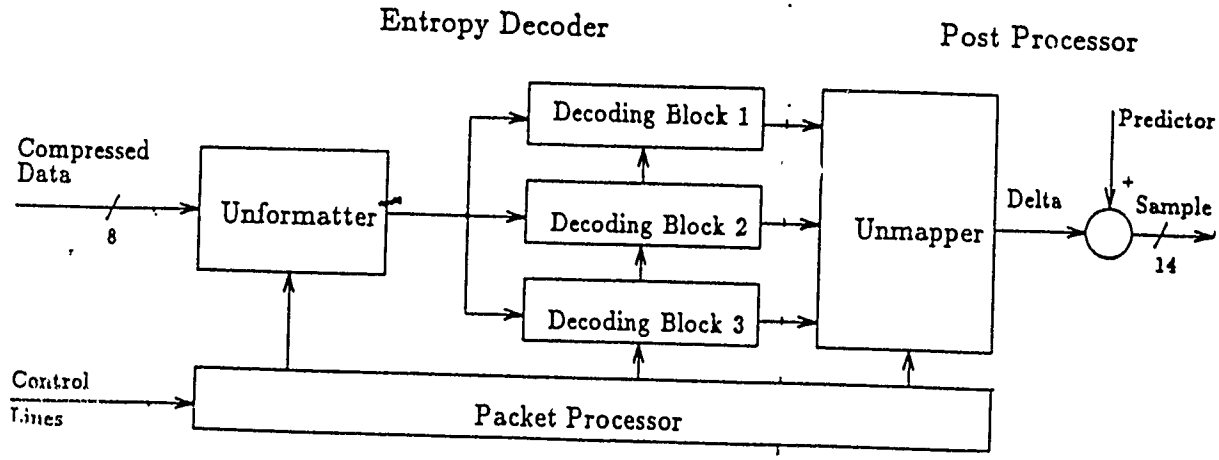


Figure 8c. Block diagram of the NASA/University of Idaho Silicon Decoder Chip.

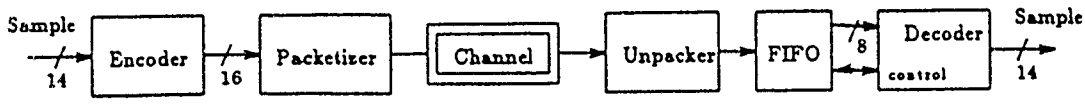


Figure 8d. Use of the encoder and decoder requires a packetizer and depacketizer tailored to the system.

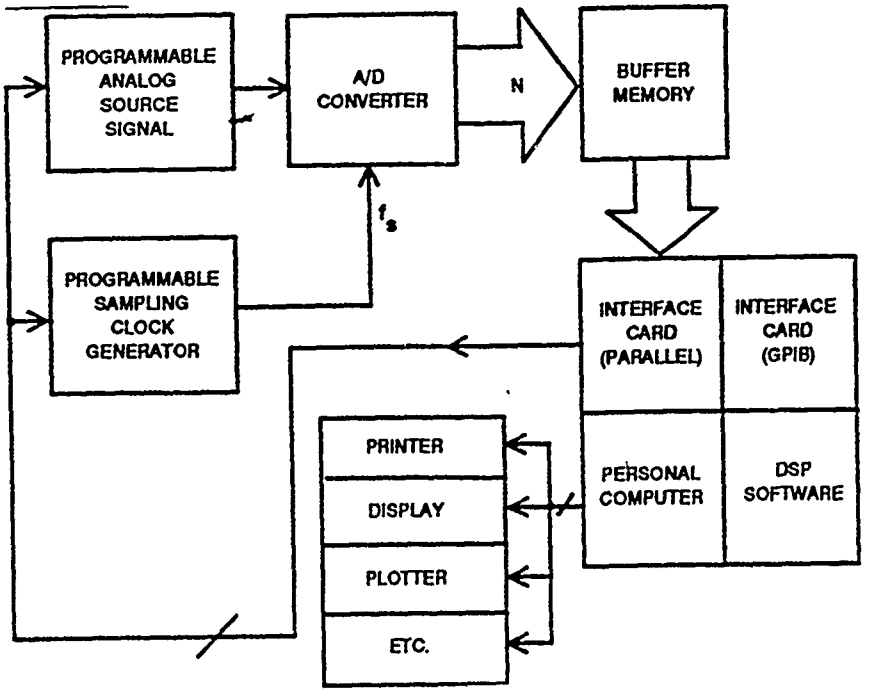


Figure 9. Block diagram of the system recommended for more accurate and speedier evaluations of the total analog/digital system. This system will be useful for evaluation of data compression/decompression as well as evaluation of the A/D converters.

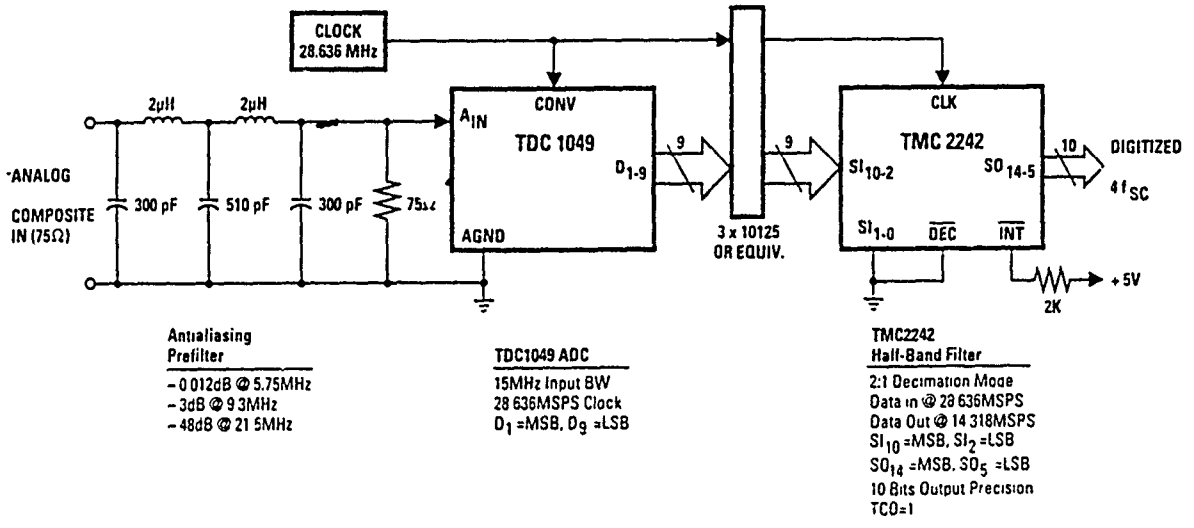


Figure 10. The design of the analog low-pass filter at the input of the A/D converter is critical to achieving an optimum compromise between resolution and dynamic range. Oversampling can increase the dynamic range.

REFERENCES

1. Donald. R. Snyder and W. Jeffrey Rowe, "System Analysis and Design for Next Generation High Speed Video Systems," *Proc.SPIE,* 1346 ,1990.

2. Patrick N. Carlton, Eugene R. Chenette, and W. Jeffrey Rowe, "Characterization of high speed video systems: tests and analyses," *Proc. SPIE,* 1539 (in press)

3. Sohei Manabe and others, "A 2-Million-Pixel CCD Image Sensor Overlaid with an Amorphous Silicon Photoconversion Layer," *IEEE Trans. Electron Devices*, vol. 38, p. 1765, August 1991.

4. Machio Yamagishi and others, " A 2-Million Pixel FIT-CCD Image Sensor for HDT Camera Systems," *IEEE Trans. Electron Devices*, vol 38, p. 926, 1991

5. Richard Bredthauer, "64 Output 1024x1024 pixel imager for high frame rate applications, " *SPIE Proc* 1155, 1989

6. James A. Bixby, "High-speed television camera and video tape recording system for motion analysis,"*Proc. SPIE*, vol. 301, 1981

7. See the data sheet for the Kodak KHS-0016 "ULTRA HIGH FRAME RATE INTERLINE CCD IMAGER."

8. John E. Grievenkamp, "Color dependent optical prefilter for the suppression of aliasing artifacts,: *Applied Optics,* vol 29, p. 676, 1990.

9. P. Mertz and F. Fray, "A Theory of Scanning and Its Relation to the Characteristics of the Transmitted Signal in Telephotography and Television,: *Bell Syst. Tech. J.*, vol. 13, pp 464-515 , 1934.

10. R. Legault, "The Aliasing Problems in Two-Dimensional Sampled Imagery," in *Perception of Displayed Information,* L.M. Biberman, Ed. , Plenum Press. New York, 1973, pp. 233-278.

11. D. F. Barbe and S. B. Campana, "Imaging Arrays Using the Charge Coupled Concept," in *Advances in Image Pickup and Display*," Vol 3, B. Kazan, Ed. Academic Press, New Your, 1977, pp 171-296.

12. F. O. Huck, N. Haylo, and S. K. Park, "Aliasing and Blurring in 2-D Sampled Imaagery," *Applied Optics,* vol 19, pp 2174-2181, 1980

13. Jack Venbrux and others, "A Very High Speed Lossless Compression/Decompression Chip Set," preprint of JPL Publication 91-13

Take the expected values of the system given by equation (1):

$$E[X_k] = E[a X_{k-1}] + E[\Gamma_{k-1} W_{k-1}], \tag{4}$$

$$\text{call } E[X_k] = \mu_x$$

$$\text{and } E[W_{k-1}] = \mu_w,$$

$$\text{Substitute } \mu_x = a\,\mu_x + \Gamma_k\ \mu_w \text{ and obtain the result,}$$

$$(1 - a)\,\mu_x = \Gamma_{k-1}\ \mu_w\,. \tag{5}$$

Here μ_x is the mean.

Subtract the mean from both sides of the system equation and obtain,

$$X_k - \mu_x = a X_{k-1} + \Gamma_k\ W_{k-1} - \mu_x \tag{6}$$

$$X_k - \mu_x = a X_{k-1} + \Gamma_{k-1} W_{k-1} - (a\,\mu_x + \Gamma_{k-1}\mu_w)\,. \tag{7}$$

Expand equation (7), square the result and calculate the expected value as follows,

$$(X_k - \mu_x) = a\,(X_{k-1} - \mu_x) + \Gamma_{k-1}\,(W_{k-1} - \mu_w) \tag{8}$$

$$(X_k - \mu_x)^2 = [a\,(X_{k-1} - \mu_x) + \Gamma_{k-1}\,(W_{k-1} - \mu_w)]^2 \tag{9}$$

$$E[(X_k - \mu_x)^2] = E[a\,(X_{k-1} - \mu_x) + \Gamma_{k-1}\,(W_{k-1} - \mu_w)]^2\,. \tag{10}$$

$$\text{Here, } E\,(X_k - \mu_x)^2 = \sigma_x^2 \text{ and } E\,(W_{k-1} - \mu_w)^2 = \sigma_w^2\,.$$

$$\text{Now, } \sigma^2_x = a^2\sigma^2_x + \Gamma_{k-1}^2\,\sigma^2_w + 2a\,\Gamma_{k-1}\,E((X_k - \mu_x)(W_{k-1} - \mu_w))\,. \tag{11}$$

It is assumed that W_{k-1} is zero mean. The cross term in equation (11) disappears because X_k depends on both X_{k-1}, W_{k-1}. But, X_{k-1} depends only on W_{k-1}, W_{k-2}, ...W_{k-3} and these are uncorrelated. This is because for this model, it was assumed that W_{k-1} had white noise characteristics.

$$\sigma^2_x = \Gamma_k{}^2 \, \sigma^2_w / 1\text{-}a^2 . \tag{12}$$

The classical Kalman filter defines $\sigma^2_x = 1$ [5],

$$1 = \Gamma_k{}^2 \, \sigma^2_w / 1\text{-}a^2 .$$

σ^2_w is the variance of the disturbances, for this model the variances of the disturbances will be equal to the variance of the system; $\sigma^2_x = \sigma^2_w$. (13)

Call: $1 = \sigma^2_w$.

$$\text{Approximate: } \Gamma = \sqrt{1 - a^2} . \tag{14}$$

The Filter algorithm defines $Q = E[W_{k-1} W_{k-1}{}^T]$. (15)

In my model, Q is a scalar.

Q will represent a measure of uncertainty to be expressed as [5]:

$$Q = 1 - a^2. \tag{16}$$

The Measurement Model

The data collected are subject to errors. Assume that the data can be described by a normal distribution. The equation for normal distribution can be expressed as follows [6, 7]:

$$p(z) = \frac{1}{\sqrt{2\pi}\ \sigma} \exp - (z_j - \mu)^2 / 2\,\sigma^2 \tag{17}$$

where μ = mean and σ = variance.

Linear regression was used to establish a relationship in the data that allows predictions of one variable to be made in terms of the others. Define,

$$z = \alpha + \beta x. \tag{18}$$

This is the equation of a straight line with y intercept = α and slope = β. For any given value of x, assume the values of z are normally distributed about the regression line, $z = \alpha + \beta x$. Then, the conditional probability density function of z can be written as:

$$f(z/x;\, \alpha, \beta, \sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp - \frac{1}{2\sigma^2} [z_j - (\alpha + \beta x_j)]^2 . \tag{19}$$

The method of maximum likelihood can be used to generate the estimates of α, β, and σ. The likelihood of a random sample of independent observations can be written as the joint density of the variates of the samples evaluated at $x_1, \ldots x_n$ [6]:

$$L(\phi_1, \ldots \phi_k) = \Pi f(X_i, \phi_i, \ldots \phi_k). \tag{20}$$

Thus, the likelihood function is a relative measure of the likelihood of a particular sample. The method of maximum likelihood [illegible] that the estimates of parameter ϕ_i be selected so as to maximize the likelihood function for a given sample. In this case, the parameters that will be estimated are α, β, and σ so that a value for x, for an observed z, has the greatest chance of being correct.

If (x_j, z_j) j=1,2,,..... n are n sample points, the likelihood function (from the conditional probability density function) is then [6]:

$$L = \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} \exp - \frac{1}{2\sigma^2} [z_j - (\alpha + \beta x_j)]^2 . \qquad (21)$$

Since the majority of likelihoods encountered usually contain exponential terms, it is more convenient to work with the natural log of the likelihood,

$$\ln L = -n/2 \ln(2\Pi) - n \ln\sigma - \frac{1}{2\sigma^2} \Sigma [z_j - (\alpha + \beta x_j)]^2 . \qquad (22)$$

Estimates of the values of α, β, and σ required to maximize L can be obtained by calculating the partial derivatives of L with respect to α, β, and σ:

$$\delta[\ln L]/\delta\alpha = 1/\sigma^2 \quad \Sigma [z_j - (\alpha + \beta x_j)] = 0 \qquad (23)$$

$$\delta[\ln L]/\delta\beta = 1/\sigma^2 \quad \Sigma x_j [z_j - (\alpha + \beta x_j)] = 0 \qquad (24)$$

$$\delta[\ln L]/\delta\sigma = -n/\sigma + 1/\sigma^3 \Sigma [z_j - (\alpha + \beta x_j)]^2 = 0. \qquad (25)$$

Equations (23) and (24) will be used to evaluate the coefficients of α and β in the regression equation and equation (25) represents the variance about the regression line. Solving equations (23) and (24) simultaneously yields:

$$\beta = \frac{\sum x_j z_j - \sum x_j \sum z_j}{\sum x_j^2 - (\sum x_j)^2} \qquad (26)$$

$$\alpha = \frac{\sum x_j^2 \sum z_j - (\sum x_j)(\sum x_j z_j)}{\sum x_j^2 - (\sum x_j)^2} \qquad (27)$$

These are the values that are then used in equation (18):

$$z = \frac{\sum x_j^2 \sum z_j - (\sum x_j)(\sum x_j z_j)}{\sum x_j^2 - (\sum x_j)^2} + \frac{\sum x_j z_j - (\sum x_j)(\sum z_j)}{\sum x_j^2 - (\sum x_j)^2} X \,. \qquad (28)$$

This procedure will be applied to a given set of data to establish a "best-fit" of the data to allow the prediction of one variable given another. The accuracy of results can be improved by implementing a Kalman filter [5, 8, 9, 10]. As will be seen in the discussion which follows, the Kalman filter is used to process the data and obtain improved estimates of range and miss distance.

This process requires knowledge of the characteristics of the ECM system and measurements of the system parameters including initial conditions and a statistical description of the system error. The goal is to be able to extend the estimates of parameter performance (miss distance) to regions (range) where no data was collected.

LIGHT-GAS GUN FIRING-CYCLE DESIGN FOR HIGH VELOCITY AND LOW PROJECTILE LOADING

Robert W. Courter
Associate Professor of Mechanical Engineering
Louisiana State University

ABSTRACT

A study is made of a novel light-gas gun firing cycle that will produce higher muzzle velocity at lower model loading than can be achieved with a conventional cycle. The study centers on the application of the new cycle to the existing light-gas gun in the Aeroballistic Research Facility at Eglin AFB, Florida. A one-dimensional, unsteady, compressible-flow computer code obtained from the Arnold Engineering Development Center is adapted to permit simulation of the Eglin gun. Taguchi optimization techniques are employed to design a sequence of gun firing cycle simulations that isolate important effects. The results of the simulations indicate that for a given shot condition, modulation of internal wave interactions through judicious selection of the spacing between the piston face and projectile base can result in the multiple benefits of increased muzzle velocity and reduced projectile accelerations and system pressures. It is concluded that the Eglin gun must be provided with a shorter pump tube to adapt it for operation with the new cycle. Further, it is recommended that the gun be instrumented to provide data for more accurate simulation and that the numerical

studies be continued to determine the sensitivity of the optimal cycle to projectile weight and desired launch velocity.

ACKNOWLEDGEMENTS

I would like to acknowledge the contributions of individuals who have made this research activity enjoyable, educational and productive. First, I thank Mr. Gerry Winchenbach, Chief of the Aeroballistics Section of the Armament Directorate at Eglin AFB, for suggesting the research topic and providing excellent technical support, and for his friendship and continuing good humor in the face of countless questions.

Next, I offer thanks to Lt. Rico Vitale, who expertly guided me through the intricacies of the Eglin computer system and provided me with countless gems of wisdom regarding various systems within the Air Force. Lt. Russ Adelgren is also due a fair share of gratitude for joining with Lt. Vitale in helping me understand the Taguchi methodology as it applies to experiment design.

I am also indebted to Chief Technician R. J. Kelley for providing a wealth of information regarding testing procedures, gun geometry, projectile and sabot technology and propellant characteristics. The other members of the section staff, Ms. Ruth Hunt, Capt. Mike Valentino and Lt. Mike Stephens, were also most helpful throughout my tenure in the lab.

I received sound advice on the operation of the AEDC computer code from Mr. Jack DeWitt of the CALSPAN Corporation of Tullahoma, Tennessee. He was the author of many enhancements to the original code, and his guidance was graciously offered and gratefully accepted.

Finally, I would like to thank Dr. Dennis Goldstein, Director of Summer Research Programs at Eglin AFB, the Air Force Office of Scientific Research and the Research Development Laboratories, Inc., for making this experience possible.

INTRODUCTION

The aeroballistic range is a unique and important tool for aerodynamic research. In the operation of an aeroballistic range, a model is propelled by a launcher into the enclosed range which is filled with a quiescent atmosphere of some desired composition and is equipped with a series of data stations, each of which allows determination of model position, orientation and speed. Aerodynamic characteristics of the model can be determined by regressive techniques through comparison of the experimental data with an analytically determined trajectory. The launcher is obviously a vital part of the facility. An aeroballistic range launcher must permit launching of unusual shapes at flight attitude and speed while maintaining the structural integrity of the projectile. The recent national interest in hypervelocity flight, as manifested in the Aerospace Plane and Strategic Defense Initiative programs, for example, has placed new demands on the capabilities of ballistic

range launchers. The emphasis is in achieving higher model launch velocities without imposing excessive loads on the model, while keeping the launcher operational pressures and temperatures within acceptable limits.

The light-gas gun has traditionally been used to achieve high velocity in the aeroballistic range. The typical firing cycle of the light-gas gun uses combustion of a conventional propellant to drive a piston which compresses a low-molecular-weight gas (hydrogen or helium) in a chamber which is separated from the projectile and gun barrel by an area-reduction section and a frangible diaphragm. The compression process continues until the diaphragm ruptures at some design pressure, and the high pressure gas drives the projectile out of the barrel. By adjusting various physical parameters of the launcher, the firing cycle (time history of launcher operation) can be adjusted to achieve different launch conditions.

About ten years ago, a novel light-gas gun firing cycle was suggested and developed by engineers at the Astron Corporation in connection with a program to develop a rapid fire weapon [1]. They dubbed the concept a "Wave Gun" because its success depended on adjusting gun and shot parameters so that multiple shock compressions could occur while the projectile was still traveling down the barrel. They were able to show through simulations that the cycle would produce higher velocities at lower accelerations than were possible with a conventional gun cycle under the same firing

conditions. Subsequently, they verified the concept through experiments with a specially-designed 20 mm gun.

The work reported in this document concerns an investigation of the application of the Astron firing cycle to the light-gas gun at Eglin AFB. The purpose of the investigation is to assess the feasibility of using the cycle without modification of the existing gun. If direct application of the new cycle is not feasible, suggestions of how the gun might be modified to effect use of the new cycle will be offered.

METHOD OF ANALYSIS

Simulation Technique - The evaluation of the new cycle is carried out via numerical simulation of the internal gas dynamics and kinematics of the gun. The operation of a light-gas gun is very simple. However, the analytical simulation of the gun cycle can be very challenging. Figure 1 is a sketch of a typical light-gas gun firing cycle showing various flow regimes. Part "a" represents an "isentropic" compression system in which a heavy piston moves slowly in compressing the light gas. With this cycle the projectile is not loaded by shock wave impingement, but it is, nevertheless, subjected to high pressures. Part "b" shows a shock compression cycle. In this case the piston is lighter and moves more rapidly, generating a shock wave ahead of it. Note, in particular, the wave interactions between piston and projectile.

The Aeroballistics Section at Eglin AFB uses a computer code, written by Patin and Courter [2], to provide predictions of muzzle velocity for a given set of launch parameters. However, this program is essentially "zero-dimensional" in that it does not provide spatial variations of gas properties within the gun. Because the success of the proposed firing cycle depends on the timing of internal wave interactions, it is essential for the present work that a gun-cycle simulation code include the capability for computing spatial property variations at each time. Consequently, a computer code, originally developed at the Naval Ordnance Laboratory and subsequently modified by the CALSPAN Corporation at AEDC, was acquired for use in the present project [3]. The code solves the one-dimensional, unsteady, compressible flow equations using a von Neumann - Richtmeyer artificial viscosity parameter, "q". Friction and heat transfer effects are included.

It has been the experience of past users of the AEDC program that some "tailoring" of the code, including an appropriate propellant combustion routine, is necessary to achieve excellent correlation of simulated results with experimental data for different launcher facilities. In fact, some internal diagnostic data (propellant pressure and piston velocity, for example) are necessary to perform a good match of the code to a particular facility. In the present study, because of time limitations and the lack of experimental data for the Eglin gun, the AEDC program was used "as received", including the propellant combustion subroutine. Thus, while the simulations

performed do not show good correlation with experiment, they do show the effects of firing cycle alterations on the simulated gun performance. Figure 2 compares the geometry of the Eglin light-gas gun with the geometry which was used in the AEDC code.

Firing Cycle Design Parameters - The design of a firing cycle depends on the properties of the gun, the propellant and the shot conditions. These can be listed as follows:

Gun properties:	Combustion chamber volume
	Pump tube bore and length
	Area reduction bores and length
	Launch tube bore and length
	Piston length, weight and material
	Piston release pressure
Propellant properties:	Type, density and impetus
	Burn rate coefficient and exponent
	Combustion gas properties
Shot conditions:	Projectile (sabot and model) mass
	Propellant mass
	Pump tube initial pressure
	Launch tube initial pressure
	Diaphragm burst pressure

The existing Eglin light-gas gun has four possible barrel configurations [20 mm, 30 mm and 40 mm (305 cm length) and 30 mm (366 cm length)] and two pump tube lengths (297 cm and 601 cm). Pistons and projectiles can be manufactured in a variety of weights and lengths. The above list constitutes a formidable array of parameters for the study. Systematic variation of each of the parameters through its probable range of values would require a large amount of time and resources. Two things have been done to reduce the labor of the study: (1) the number of parameters has

been reduced from past experience with interior ballistics modeling , and (2) the schedule of simulation runs has been optimized through the use of Taguchi quality control techniques [4], [5].

In the present study the following parameters have been held constant:

Projectile mass (69.22 gm to match an experimental item)
Barrel geometry (30 mm by 304.8 cm)
Powder chamber geometry
Area reduction geometry
Combustion model and powder (AEDC provided values) *
Barrel initial pressure (atmospheric)
Piston mass (100 gm - light piston required for cycle)

* A limited study of the effect of powder burning characteristics was performed in connection with the present work.

Later work can focus on the subtle cycle changes required to effect efficient gun performance for various projectiles, launch tubes and propellants. The labor involved in the investigation of the remaining parameters is reduced by application of Taguchi techniques.

Taguchi Simulation Optimization - The Taguchi techniques described in References 4 and 5 actually represent a systematic and mathematically simple method of optimizing the choice of parametric values with which to begin the search for the best possible gun cycle. It is beyond the scope of the present report to give rigorous details of the methodology. However, the results of a typical analysis as applied to the gun firing cycle can be characterized by the simple interaction graphs shown in Figure 3. This particular set, for example, shows that an interaction between pump tube pressure and

piston length has an effect on muzzle velocity but has little effect on system pressure and projectile acceleration. Here, the piston length controls the distance between piston face and projectile base, with the longer piston representing a shorter inter-surface distance. The figures indicate that to maximize muzzle velocity, a large value of pressure and a long piston should be selected. The acceleration and pressure graphs indicate that this selection also tends to yield low levels of projectile loading and system pressure. These results serve as guidelines for the selection of parameter values with which to start the search for the optimum.

RESULTS

Time limitations preclude a thorough parametric study of the Astron gun cycle. However, it is possible tc demonstrate, through isolated examples, that improved performance can be achieved with the multiple-compression technique.

All of the calculations performed were for the existing gun geometry in order to assess the feasibility of adapting the gun to the new firing cycle. Early studies indicated that the long, high-volume pump tube configuration would not produce the highest velocities for a given shot configuration because the large distance between piston face and projectile base did not permit optimum shock reflection characteristics. Thus, most of the calculations were carried out with the "short" pump-tube configuration.

A large number of simulations for the Eglin gun configurations was performed during this study. The results of these simulations do not make the Eglin gun a clear-cut choice for conversion to the multiple-compression cycle. However, it appears that the gun can be converted to the new cycle if a shorter pump tube is used. The simulations show that for given shot conditions, the wave interactions developed in the light-piston cycle can be modulated to give a range of muzzle velocities. In addition, with this methodology the projectile loading and system pressure levels can be kept well below those incurred during a conventional heavy-piston cycle.

The highly non-linear nature of the flow field within the gun precludes the formulation of any general rules for peak performance. However, the essential features of the cycle can be demonstrated by letting the gun geometry, projectile mass, piston mass and charge pressure remain constant in the simulations. Subsequent variation of the piston length and propellant weight effectively alter the timing of wave propagation in the pump tube, and variation of the diaphragm burst pressure has a similar effect on the timing of shock compressions reaching the projectile. This is demonstrated in Figures 4 through 8. Figure 4 shows the variation of muzzle velocity and projectile acceleration with piston length. In the simulations of the Eglin gun, the piston length is used to control the spacing between piston face and projectile base. Note the dramatic improvement in gun performance that occurs when the piston length is changed from 3 cm to 30 cm. This is indicative of the important role that inter-surface spacing plays in performance. Figure 5 is a three-

dimensional plot showing pressure distributions between piston face and projectile base for the maximum-velocity shot as a function of time. The propagation of wave fronts between the surfaces is apparent following the onset of projectile motion. Figure 6 is a comparison of the projectile base pressure time histories for the maximum-velocity shot and another shot. The higher sustained base pressures produced in the long-piston shot is clearly superior to the pressure distribution from the other shot, even though the initial pressure spike of the latter is much larger. Figures 7 and 8 show the corresponding trajectories of piston and projectile, respectively, for these cases. While these results do not show that the multiple-compression cycle can produce the highest possible muzzle velocity at the lowest possible projectile acceleration, they do indicate that tailoring of the new cycle offers advantages that are not available from conventional gun cycles.

CONCLUSIONS AND RECOMMENDATIONS

Because of the non-linear nature of the gas dynamics of the light-gas gun and the interactions among pertinent parameters, it is impossible to develop a closed-form analytical model of the gun firing cycle. Consequently, this study has involved using numerical simulation to predict the cycle behavior. A large number of simulation calculations indicates that for given shot conditions, performance benefits in the form of high muzzle velocity at reduced projectile loading and system pressure levels are possible with the existing Eglin gun configuration only if the gun is supplied with a

shorter pump tube which will optimize internal wave interactions. The use of the light-piston cycle has obvious benefits from the standpoint of loading on the gun itself. However, the value of the above results from the standpoint of model loading and high muzzle velocity cannot be evaluated for the Eglin gun at this time because that gun has no instrumentation, and the simulation model cannot be matched to the gun without some experimental data for comparison.

As a consequence of the above conclusions, the following recommendations are offered:

1. The Eglin gun should be provided with sufficient instrumentation to permit the development of an accurate numerical model which simulates the internal gas dynamics and kinematics of the gun.

2. Numerical simulation studies should be continued to determine the sensitivity of the optimal cycle to projectile weight and desired muzzle velocity.

3. A new pump tube should be designed for the Eglin gun which will permit implementation of the new firing cycle for a range of projectile weights and launch velocities.

REFERENCES

1. Dahm, T. J. and D. S. Randall, "The Wave Gun Concept for a Rapid-Fire Hypervelocity Weapon," Astron Research and Engineering paper, January, 1984.
2. Patin, R. and R. W. Courter, "A One-dimensional Simulation Model for a Two-stage Light-Gas Gun with Deformable Piston," AIAA Preprint No. 86-0046, presented at the AIAA Aerospace Sciences Conference, Reno, Nevada, January, 1986.
3. Piacesi, R., D. F. Gates and A. E. Seigel, "Computer Analysis of Two-Stage Hypervelocity Model Launchers," Naval Ordnance Laboratory Report NOLTR 62-87, White Oak, Maryland, August, 1963.
4. Ross, P. J., Taguchi Techniques for Quality Engineering, McGraw-Hill Book Co., New York, 1988.
5. Kacker, R. N. and K. Tsui, "Interaction Graphs: Graphical Aids for Planning Experiments," Journal of Quality Technology, v. 22, n. 1, January, 1990, pp. 1-14.

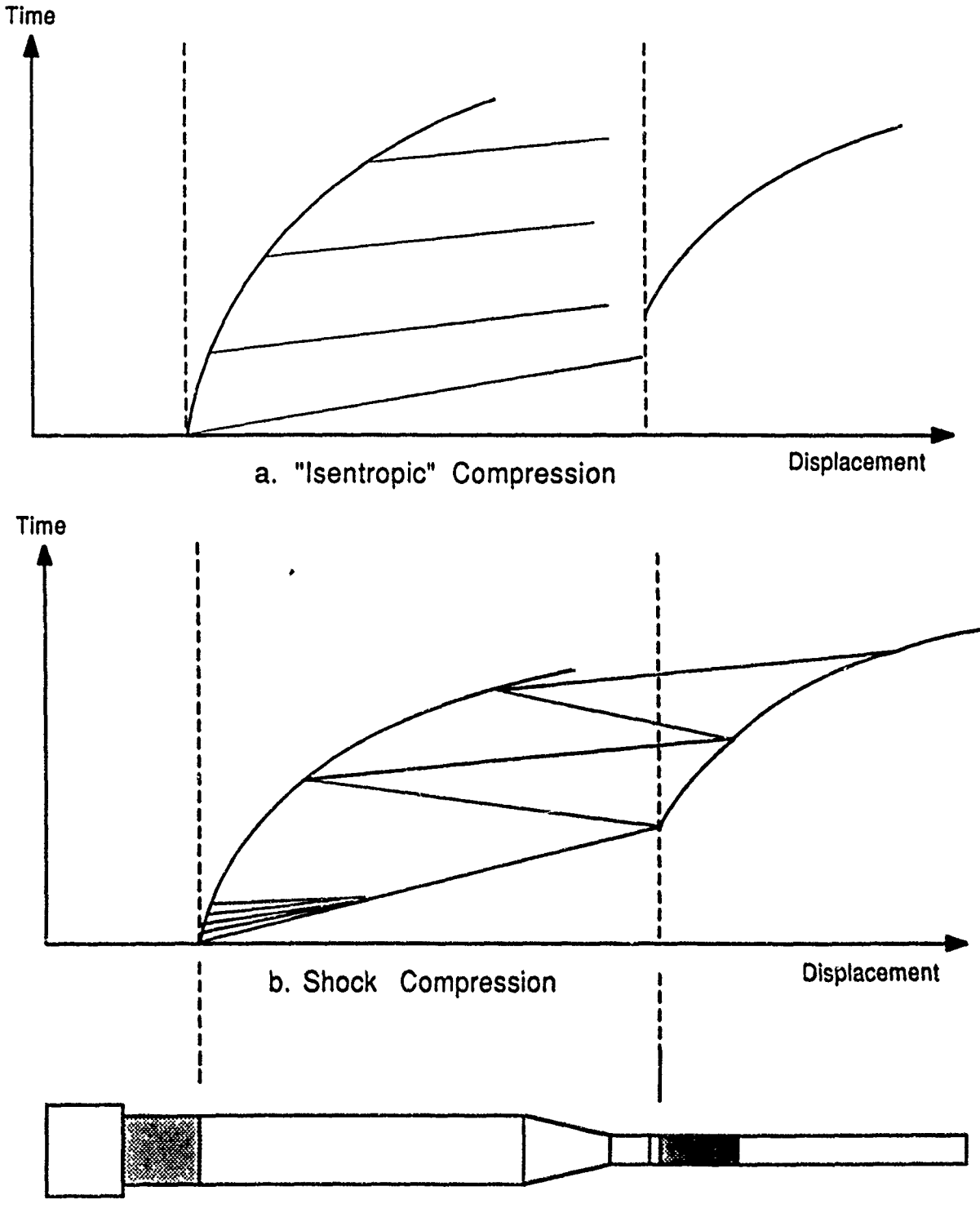
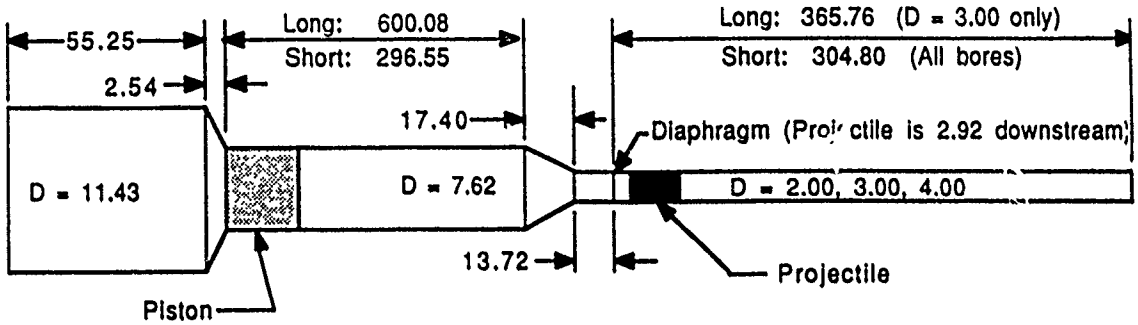
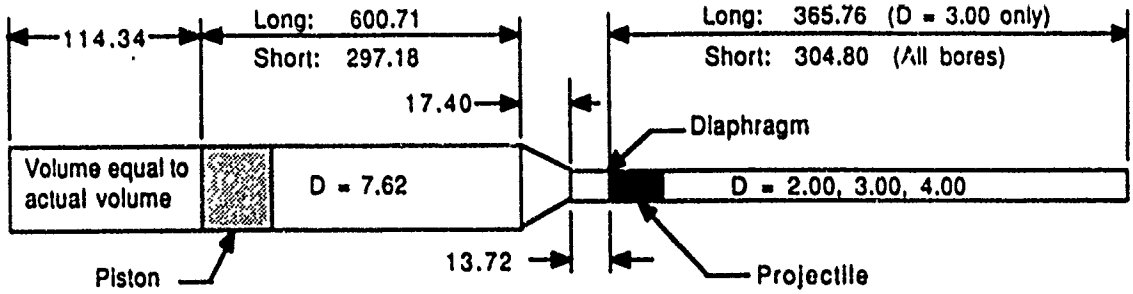


Figure 1. Light-Gas Gun Firing Cycles



ACTUAL GUN GEOMETRY



MODELED GUN GEOMETRY

Figure 2. Eglin AFB Light-Gas Gun Geometry
All dimensions are in centimeters.

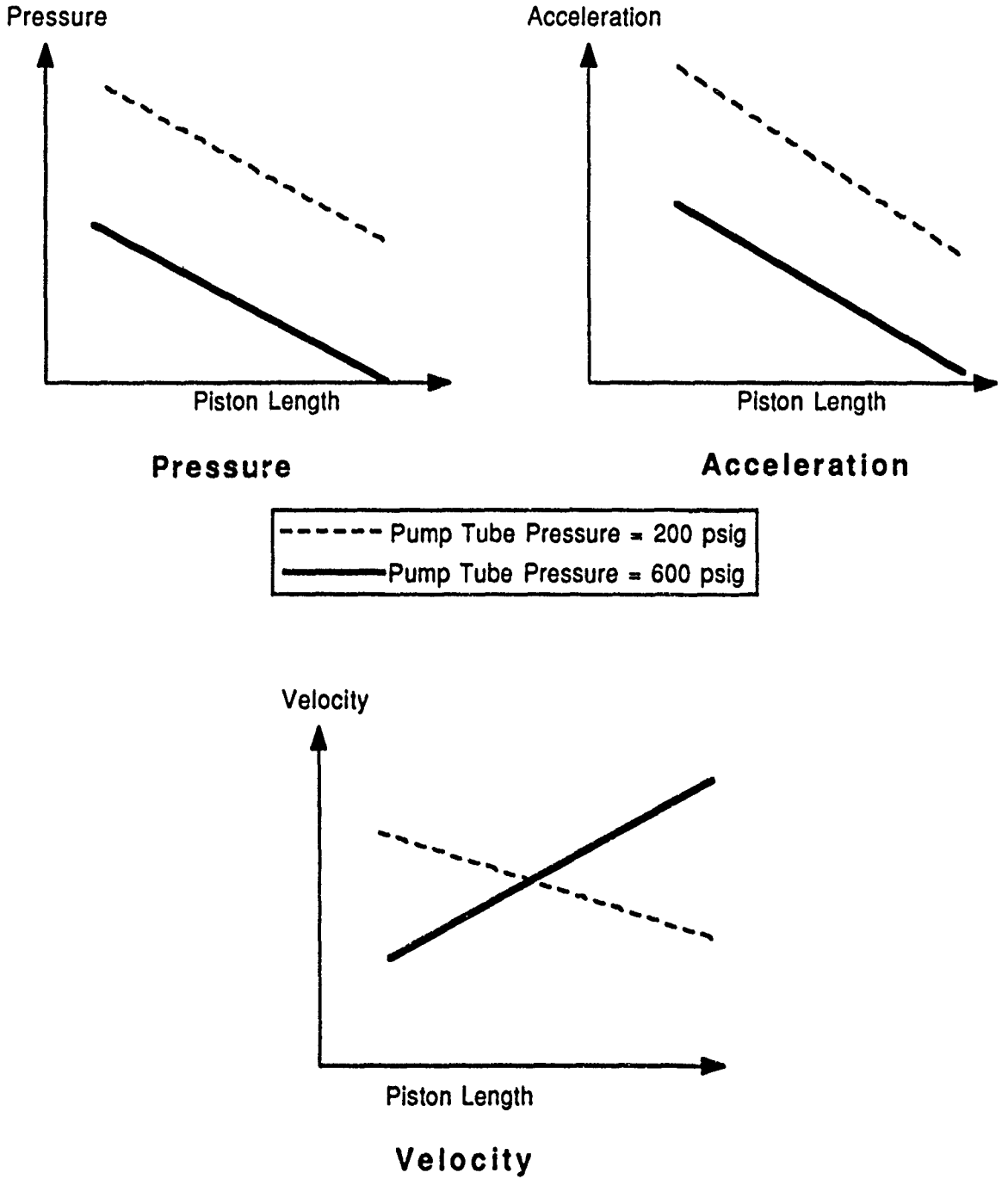


Figure 3. Typical Interaction Graphs

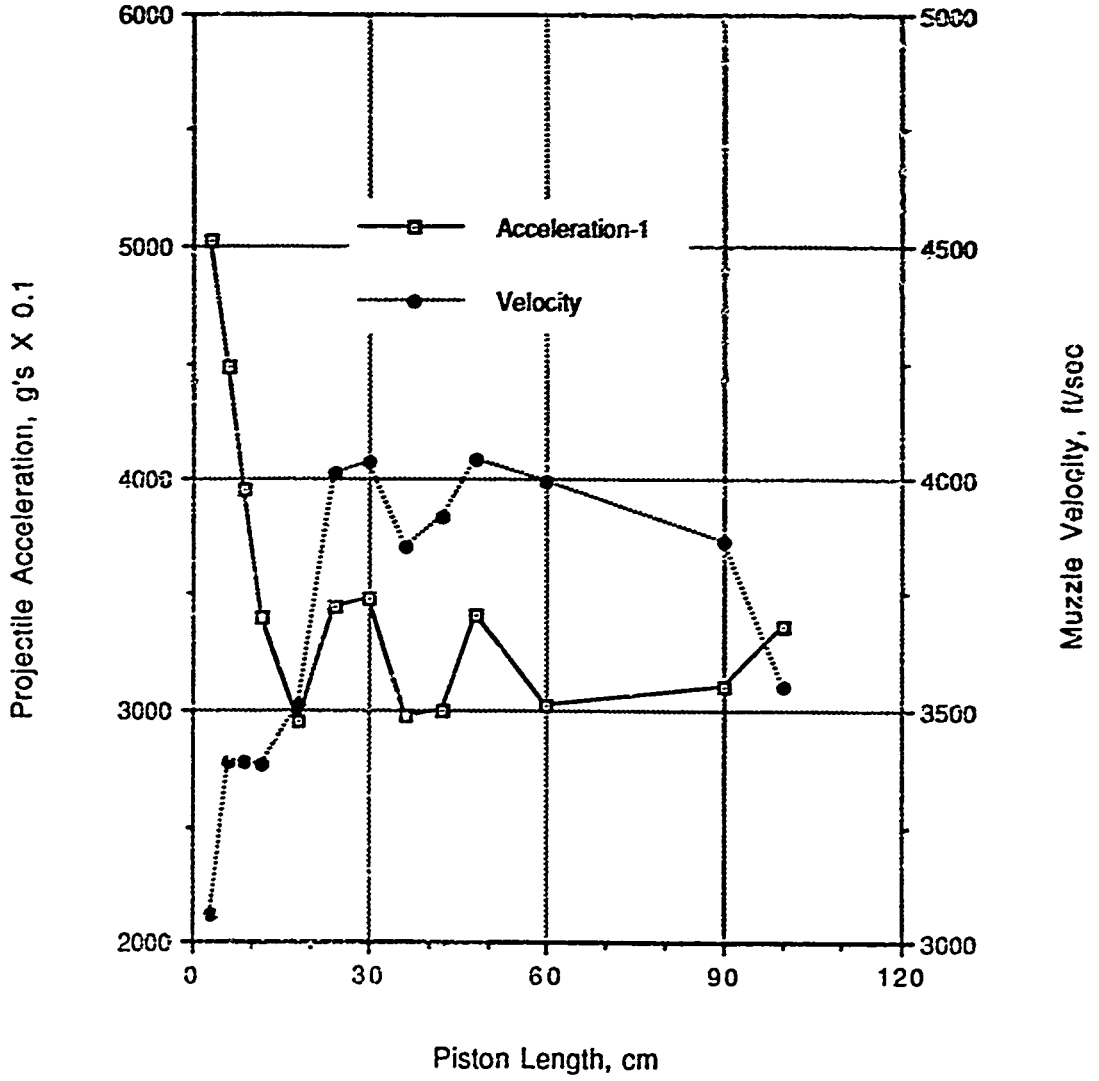


Figure 4. Projectile Velocity and Acceleration for Fixed Shot Conditions

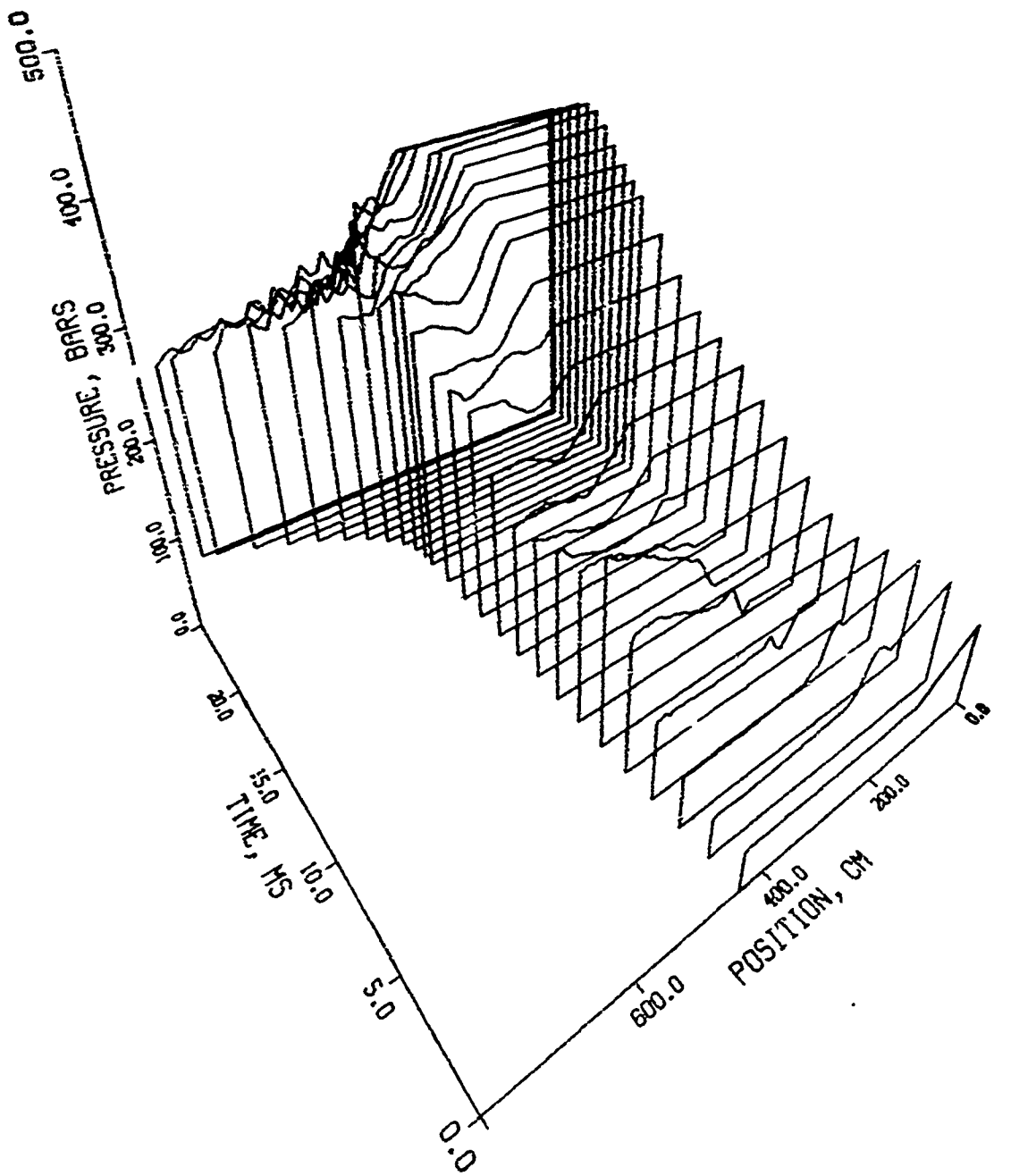


Figure 5. Pressure Distribution History for Light-Gas Gun Firing Cycle

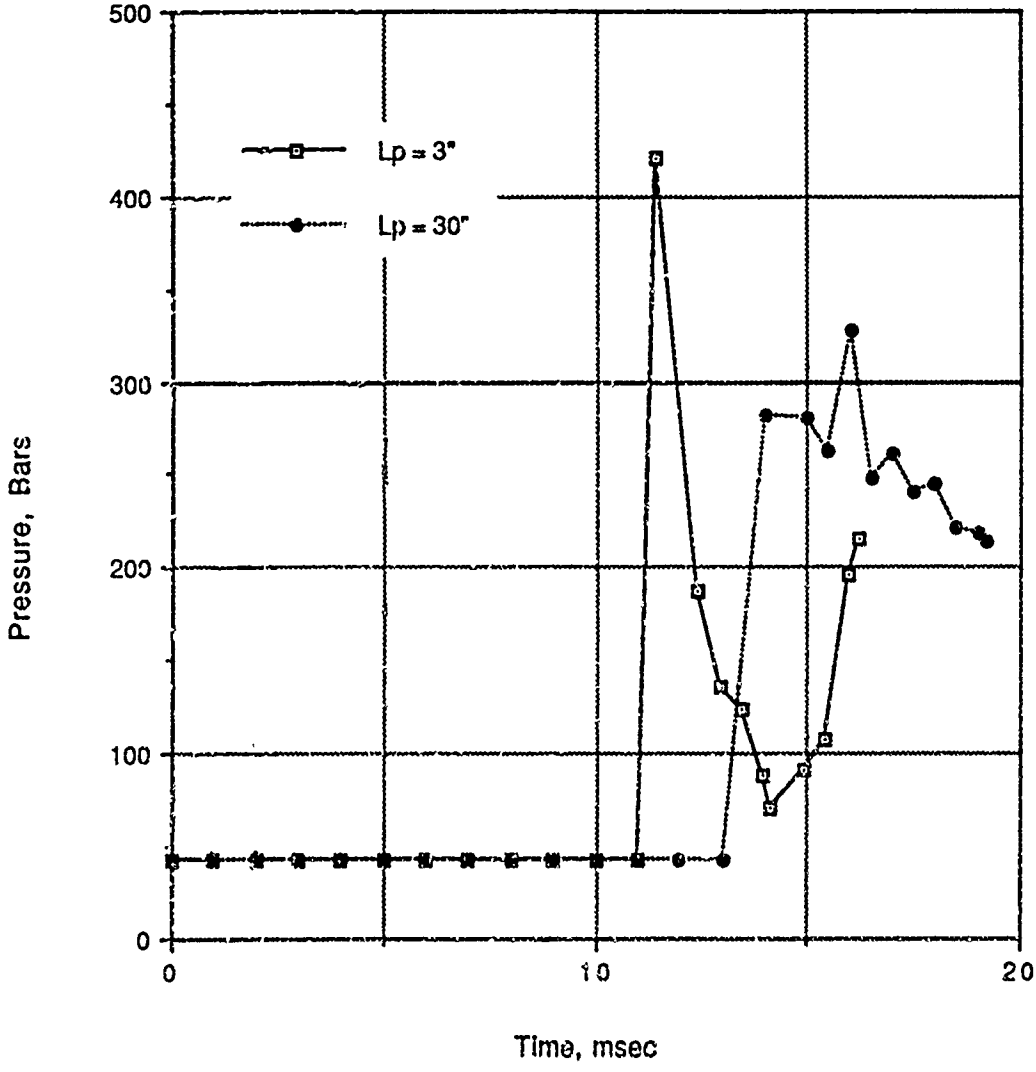


Figure 6. Base Pressure Histories

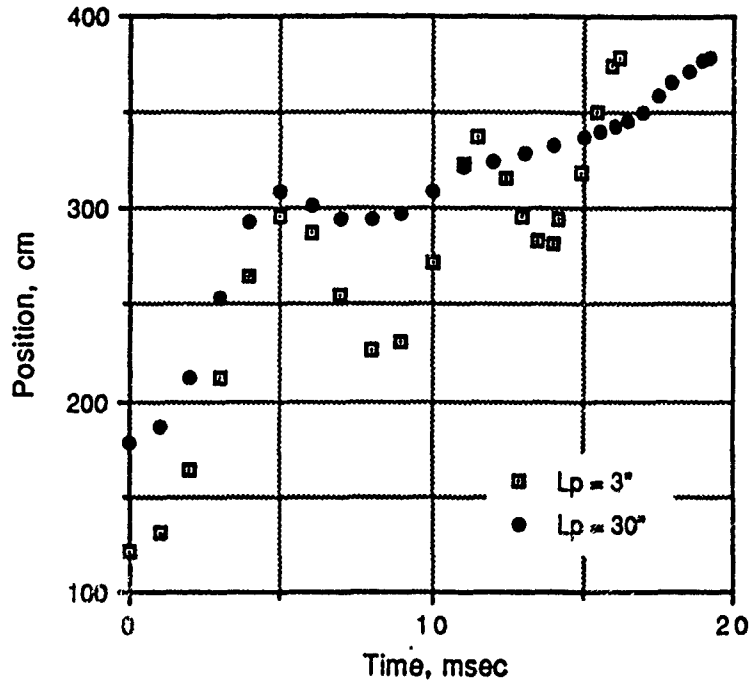


Figure 7. History of Piston Motion

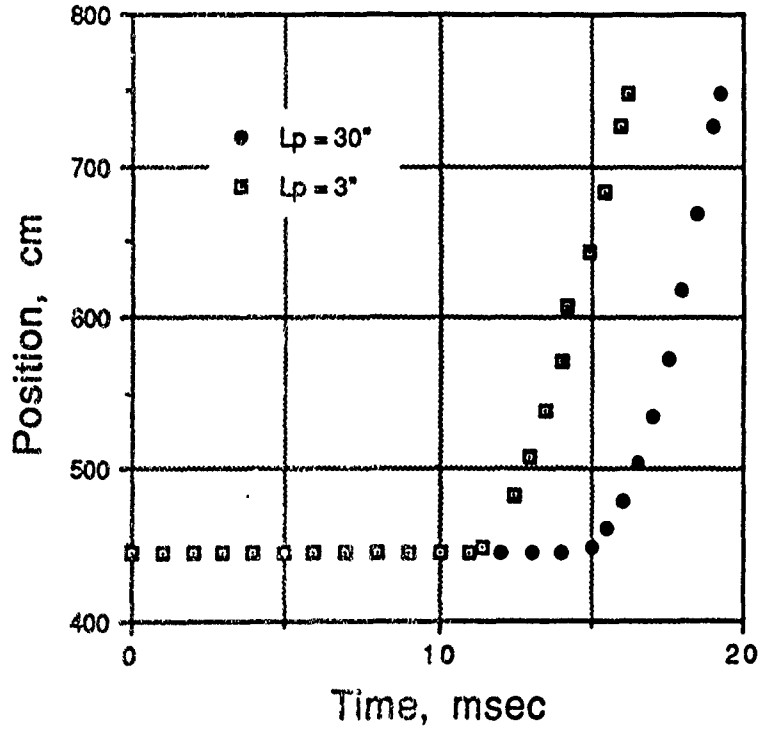


Figure 8. Projectile Position History

FINAL REPORT

TWO DIMENSIONAL SIMULATION OF RAILGUN PLASMA ARCS

Submitted to

1991 US AFOSR - RDL SUMMER FACULTY/
GRADUATE STUDENT RESEARCH PROGRAM

Research Period: June 20, 1991 to August 28, 1991

Faculty Research Associate: Dr. Manuel A. Huerta
Professor of Physics
University of Miami
Physics Department
Coral Gables, FL. 33124

Soc. Sec. No. 262-70-9459
Telephone No. (305) 284-2323, Ext.8

Graduate Student Research Associate:
Mr. G. Christopher Boynton
University of Miami
Physics Department
Coral Gables, FL. 33124
Telephone No. (305) 284-2326, Ext.6

Research Location: WL/MNSH
Eglin AFB, FL 32542-5000

Effort Focal Point: Mr. Kenneth K. Cobb
WL/MNSH

Signed: Kenneth K. Cobb

Signed: Manuel A. Huerta

Manuel A. Huerta
Faculty Research Associate

G. C. Boynton

G. Christopher Boynton
Graduate Student Research Associate

Date: August 27, 1991

TWO DIMENSIONAL SIMULATION OF RAILGUN PLASMA ARMATURES

by

Dr. Manuel A. Huerta, Professor of Physics and Mr. George C. Boynton

ABSTRACT

Our code uses the equations of two-dimensional resistive MHD with Ohmic heating and radiation heat transport to simulate the internal dynamics of a railgun plasma armature. All quantities are advanced in time using an explicit Flux Corrected Transport scheme. We have done some theory to describe the initial fuse explosion and have modified the simulation to have a more realistic initial state. We now allow the driving current to be input from a data file. We also have developed a version that computes only one half of the rail to rail distance and forces mirror symmetry for the other half to save running time. We have done the background work to use an electrical conductivity that accounts for nonideal effects and to include turbulent viscous drag. We also studied the problem of the lubrication and drag of the projectile against the rail walls and found that it could be made to have little effect. A good deal time was spent doing a review of a new scheme for pulse radiation.

I. INTRODUCTION

I have spent the summer of 1991 working at WL/MNSH site A15, together with my graduate student G. C. Boynton who was supported under the 1991 USAF Graduate Student Research Program. We have been working on a two dimensional simulation of armature plasmas. Our main objective this summer was to improve our model by incorporating a better initial state, allowing for a time varying current from a data file, including nonideal electrical conductivity, and field augmentation, and examining the problem of viscosity.

We enjoyed abundant access to the VAX 8650 at site A15. The Cray time that we were provided came a little late in the research period and was not fully utilized. An impressive railgun facility has been built on Okaloosa Island at site A15 and there is quite a bit of experimental activity on diagnostics of the arc plasma armature. We had very useful discussions with Mr. Kenneth Cobb and Dr. Eugene Clothiaux about their experiments. They provided us with the data files needed to include an experimental current profile in our simulation. We also profited from discussions with Dr. Glen Rolader of SAIC regarding many aspects of the problem.

II. WORK ACCOMPLISHED

In our previous simulations we took the total current to be a constant in time and the initial state of the plasma was closely related to the equilibrium calculated by Powell and Batteh[1]. The new simulation has an initial state in which the plasma starts in a much more dense state that is related to the thin exploding foil that produces the initial plasma. We did a good bit of theoretical modeling of the different stages as the foil melts, vaporizes, and ionizes. This would be useful if we decide to pursue the matter of foil initiation more fully. As it is the details are not important right now because the simulation takes the initial plasma to occupy only one computational cell. This is due to running time limitations and leads to large gradients, as shown in the current profile of Fig. 1. FCT is good at handling this sort of thing, however, and the code runs fine. Figs. 2 and 3 show how the plasma expands nicely.

The time varying current profile that we have used is derived from the experiments of Clothiaux and Cobb on their capacitor driven railgun. The profile is shown

in Fig. 4. The current drops as a capacitor module discharges and rises again as another module cuts in. As expected the plasma grows toward the rear when the current begins dips. The reason for this is that the current drop is felt mainly at the rear of the plasma due to the finite time required for magnetic field diffusion. The rear region begins to fall behind the main arc because a lower current means a lower magnetic force on it, so its acceleration becomes less than that of the main arc.

The rapid growth of the arc toward the rear has caused a problem in our simulation, but we can easily fix it. Our initial conducting arc is only one cell long but our computational region is much longer and mostly nonconducting initially. We do this so we can monitor the growth of the conducting region of the arc. The problem is that the way our code runs now the information of the true value of the total current is lost if the entire computational region becomes conducting. In effect we not only lose current out the rear of the computational region, but we lose contact with the true value of the total current and of the rear magnetic field. A good fix of this problem will be to change our boundary condition at the rear of the computational region so the magnetic field there is made to always follow the true magnetic field at the breech. This would tend to introduce a magnetic field gradient (read current density) right at the rear of the computational region, but it would diffuse rapidly because this region is not a good conductor. This kind of trouble highlights the importance of any drop in the total current, as expected.

Due to the rear boundary condition problem the acceleration starts to drop off as shown in Fig. 5. This drop off is not valid. One of our motivating factors in introducing the time varying current was to see if we could reproduce the buffer zone results of Cobb and Clothiaux, so far we have not been able to.

Our simulations so far have used the well known Spitzer[2] conductivity

$$\sigma = \frac{2}{Z\sqrt{m_e}}\left(\frac{4\pi\epsilon_0}{e}\right)^2\left(\frac{2k_BT}{\pi}\right)^{3/2}\sigma^*_{SP} \qquad \text{with} \qquad \sigma^*_{SP} = \frac{\gamma_E}{\ln\Lambda}$$

where γ_E is of order unity and

$$\Lambda = \frac{12\pi}{Z_f Z\sqrt{n_e}}\left(\frac{\epsilon_0 k_B T}{e^2}\right)^{3/2}$$

is the Coulomb logarithm. We have decided to use the nonideal conductivity where σ^*_{SP} is replaced by σ^*_{KV} as discussed by Rolader and Batteh[3]. This can reduce the

conductivity by half at low temperatures which could have a significant effect on the current distribution in the plasma.

The problem of viscous drag was given a good deal of attention. We only have a small number (10–20) of cells in the rail to rail direction due to the need to reduce the running time. It is hopeless to have enough resolution to truly do the viscous problem and have the boundary layer come out correctly. It becomes necessary to do some sort of matching of the boundary layer flow to the exterior flow. We use the Spitzer[2] viscosity for the plasma

$$\mu = 2.21 \times 10^{-15} \frac{T^{5/2} A_i^{1/2}}{Z^4 \ln\Lambda} \frac{\text{gm}}{\text{cm} \cdot \text{sec}},$$

where A_i is the ion atomic weight. For a typical velocity of 1 km/sec and length of 10 cm the Reynolds number

$$R = \frac{\rho U l}{\mu} \approx 2.5 \times 10^8$$

which is a very large value. This puts us in the regime where there is a turbulent boundary layer. For this reason we use the empirical equations for the shear stress due to viscous drag in turbulent pipe flow taken from Schlichting[4]

$$\tau_0 = \frac{1}{2} \rho U_\infty^2 c_f' \qquad \text{where} \qquad c_f' = 0.0592 \left(\frac{U_\infty x}{\nu}\right)^{-1/5}.$$

We integrate this stress along the rails and subtract the total drag force from the forward force that acts on the projectile.

We also considered the problem of the viscous flow that lubricates the motion of the projectile along the rail. This is a generalization of lubrication flow[4] because here the pressure is different behind and in front of the projectile. The lubrication problem is coupled to the projectile deformation due to the stresses caused by the force that gives the projectile its acceleration a_P. Say we have rigid rails and a projectile of square cross-section w^2, length H and density ρ_P. Let the gap between the rail and the projectile have a width h_1 at the front of the projectile and h_2 at its rear. We simplify by letting the pressure in the lubrication flow have an average value P_L. The quantities

$$h_1 - h_2 \approx \frac{\nu_P \rho_P g H}{E - 2\nu_P G} \frac{w}{2} \qquad \text{and} \qquad h_1 \approx \frac{w}{2} \frac{p_L}{E}, \qquad (1)_{h1h2}$$

where E is the Young's modulus, G the shear modulus and ν_P the Poisson ratio of the projectile, are crucial in determining the flow in the gap. Let μ be the viscosity of the fluid in the gap and U be the speed of the projectile. The average pressure in the gap is

$$p_L \approx \frac{6\mu U H}{(h_1 - h_2)^2}(\ln(\frac{h_1}{h_2}) - 2\frac{h_1 - h_2}{h_1 + h_2})$$

From this and Eq. (1) we are able to get h_1 and h_2, which come out quite small, of the order of micrometers. The drag force due to this flow turns out to be quite small. The volume rate of fluid flow toward the rear of the projectile is

$$Q = \frac{w h_1 h_2}{h_1 + h_2}(U - \Delta p \frac{h_1 h_2}{6\mu H})$$

where Δp is the pressure difference between the back and the front of the projectile. A negative Q would indicate a plasma blow by. Typical values of Q can be made small enough to be unimportant.

Finally a good deal of time was spent preparing a review of a proposal for a new type of electromagnetic pulse generator called an impulse gun.

IV. CONCLUSIONS

Our code has shown the potential to describe the main experimental facts and is now approaching a stage where it can be compared with experiments. This should give valuable insights into the workings of plasma armatures.

V. ACKNOWLEDGEMENTS

We would like to thank the Air Force Systems Command and the Air Force Office of Scientific Re. rch for sponsorship of this research. We would also like to thank Research & Development Laboratories for the efficient way in which they administered the USAF - RDL Summer Faculty and Graduate Student Research Programs.

Our experience was productive and enjoyable thanks to the cooperation we received from many people at WL/MNSH. We would specially like to thank Mr. Kenneth K. Cobb who helped us so much in getting access to the computer resources that we needed. We also thank him for useful discussions of work in rail guns. We had fruitful discussions with several people, specially Dr. Glen Rolader of SAIC,

Dr. Eugene Clothiaux, and Dr. Chia Thio. We thank Mr. Ed Taconis for his support and Mr. Andy Marino for setting us up with abundant disk space and other resources on the VAX 8650 at site A15.

REFERENCES

1. Powell, J. D., and Batteh, J. H., "Plasma Dynamics of an Arc-Driven Electromagnetic Projectile Accelerator," J. Appl. Phys., vol. 52, 2717 (1981)
2. Spitzer, L., "Physics of Fully Ionized Gases", Interscience Publishers, 1962.
3. Rolader, G. E., and Batteh, J. H., "Thermodynamic and Electrical Properties of Railgun Plasma Armatures", IEEE Trans. Plasma Sci., vol. 17, pp. 439–445, June 1989.
4. Schlichting, H., "Boundary Layer Theory", McGraw–Hill, 1960, Chapter XXI.

JY: Y-CMP. CURRENT DENSITY - AMPERES/M^2
Run: Elmo
Max = 4.8E+09
Min = 0.0E+00
SFRM = 0.0E+00m
VFRM = 4.5E+02m/s
GFRM = 9.530E+05m/s^2
ETIME = 1.000000E-04s
Cells: X: 1-060 Y:1-5. Steps: 0000000

FIGURE 1

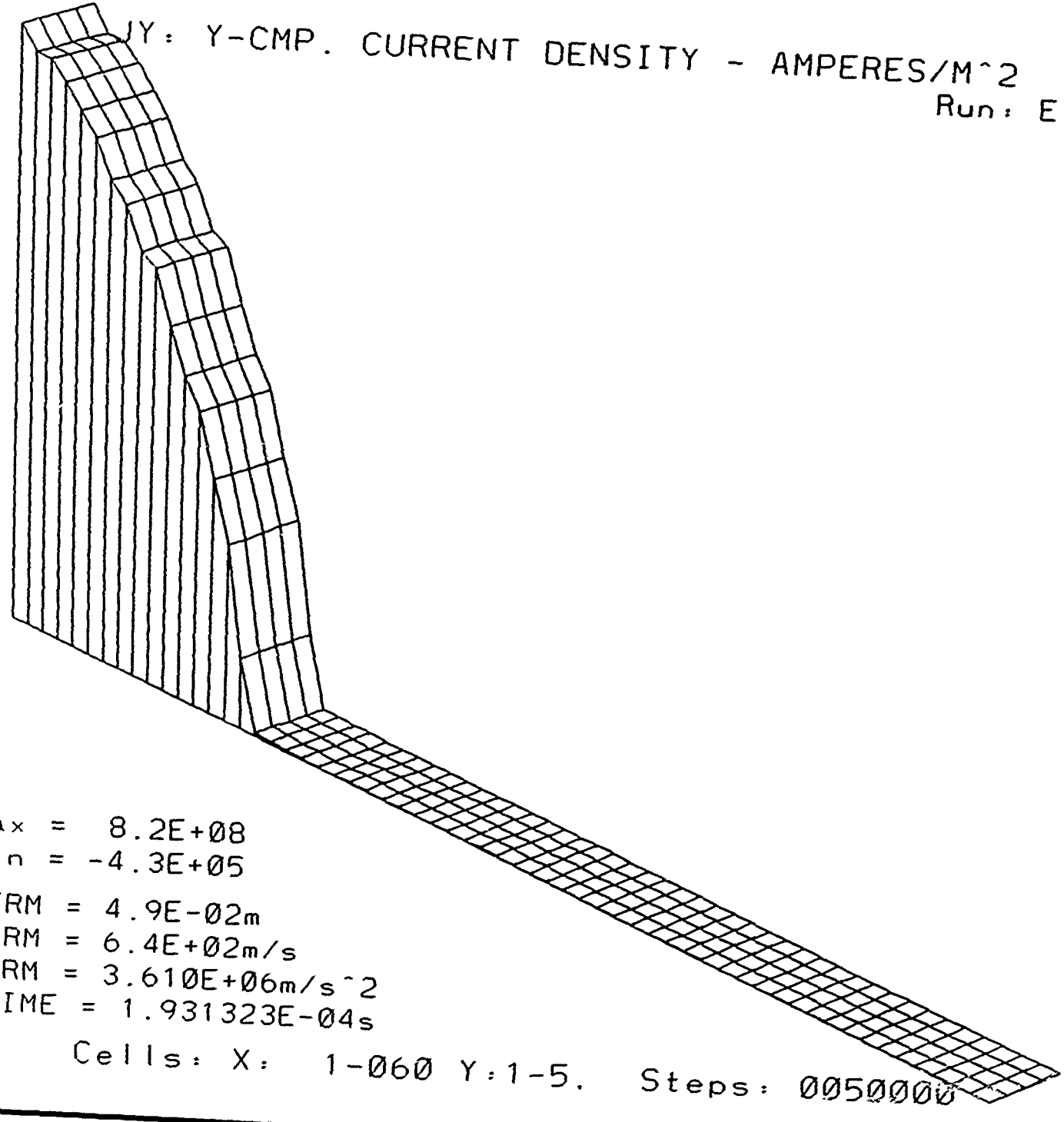

JY: Y-CMP. CURRENT DENSITY - AMPERES/M^2
Run: Elmo
Max = 8.2E+08
Min = -4.3E+05
SFRM = 4.9E-02m
VFRM = 6.4E+02m/s
GFRM = 3.610E+06m/s^2
ETIME = 1.931323E-04s
Cells: X: 1-060 Y:1-5. Steps: 005000

FIGURE 2

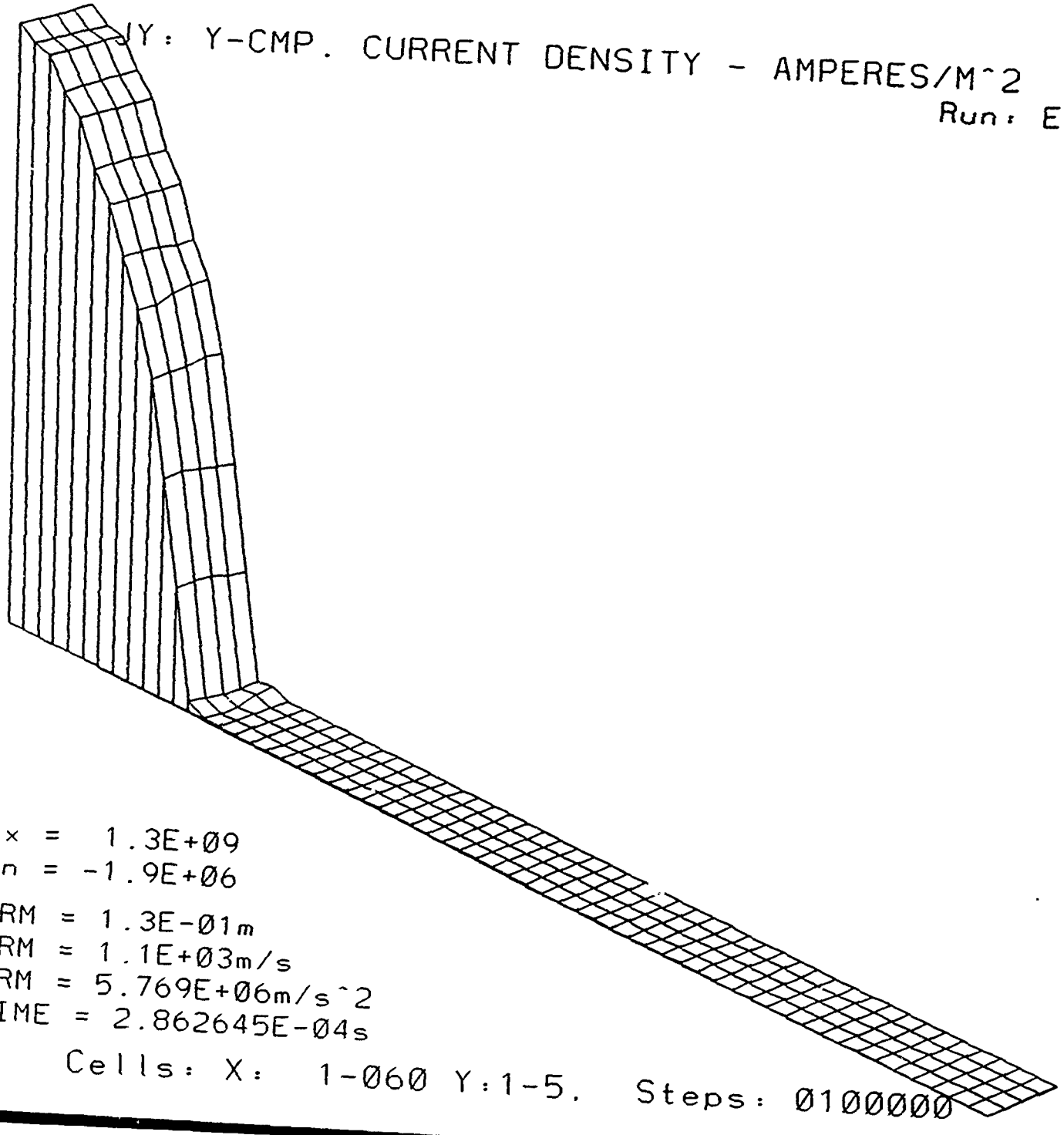

JY: Y-CMP. CURRENT DENSITY - AMPERES/M^2
Run: Elmo
Max = 1.3E+09
Min = -1.9E+06
SFRM = 1.3E-01m
VFRM = 1.1E+03m/s
GFRM = 5.769E+06m/s^2
ETIME = 2.862645E-04s
Cells: X: 1-060 Y:1-5. Steps: 0100000

FIGURE 3

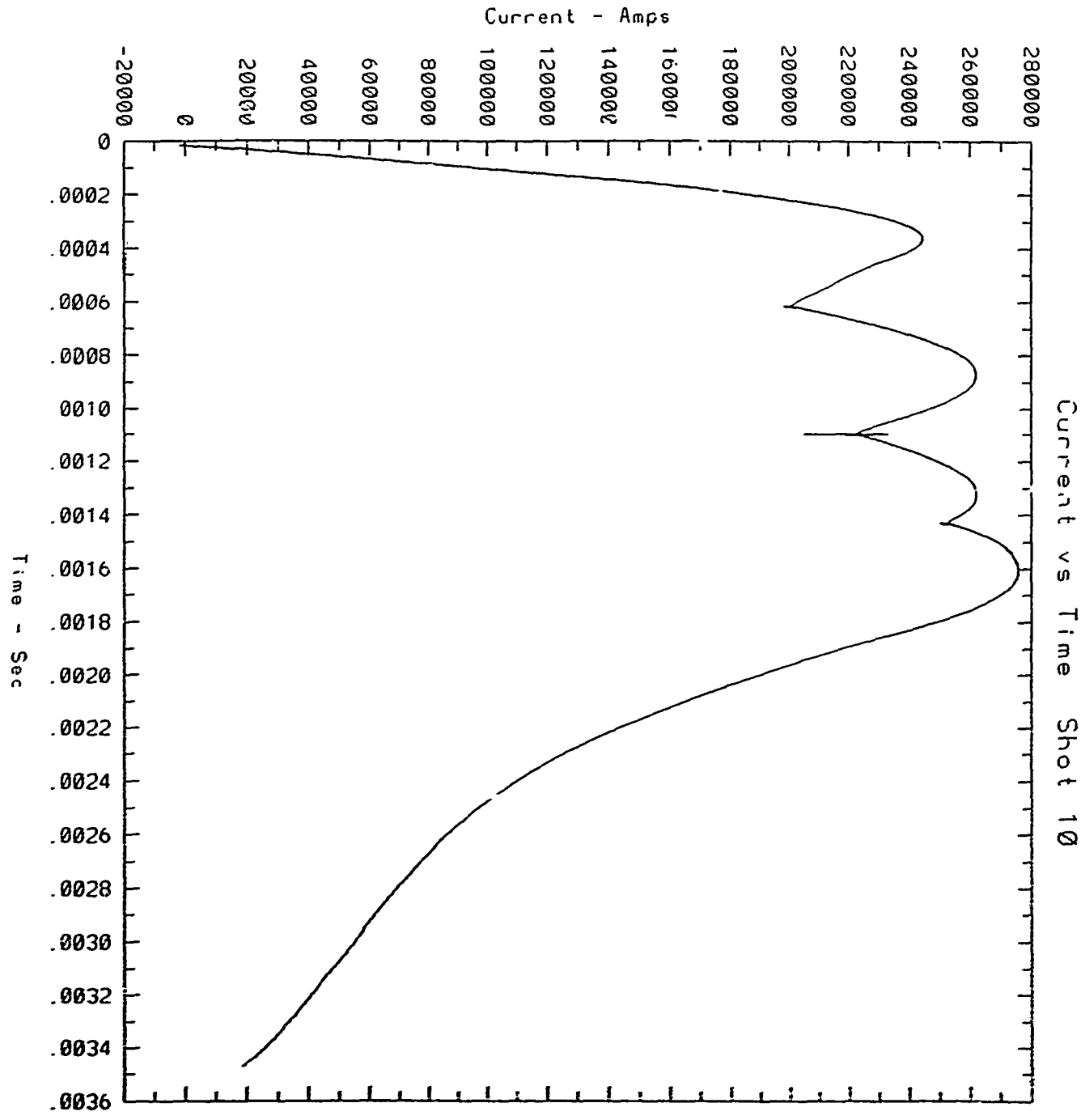

Current - Amps
-20000
0
20000
40000
60000
80000
100000
120000
140000
160000
180000
200000
220000
240000
260000
280000
Time - Sec
0
.0002
.0004
.0006
.0008
.0010
.0012
.0014
.0016
.0018
.0020
.0022
.0024
.0026
.0028
.0030
.0032
.0034
.0036
Current vs Time Shot 10

FIGURE 4

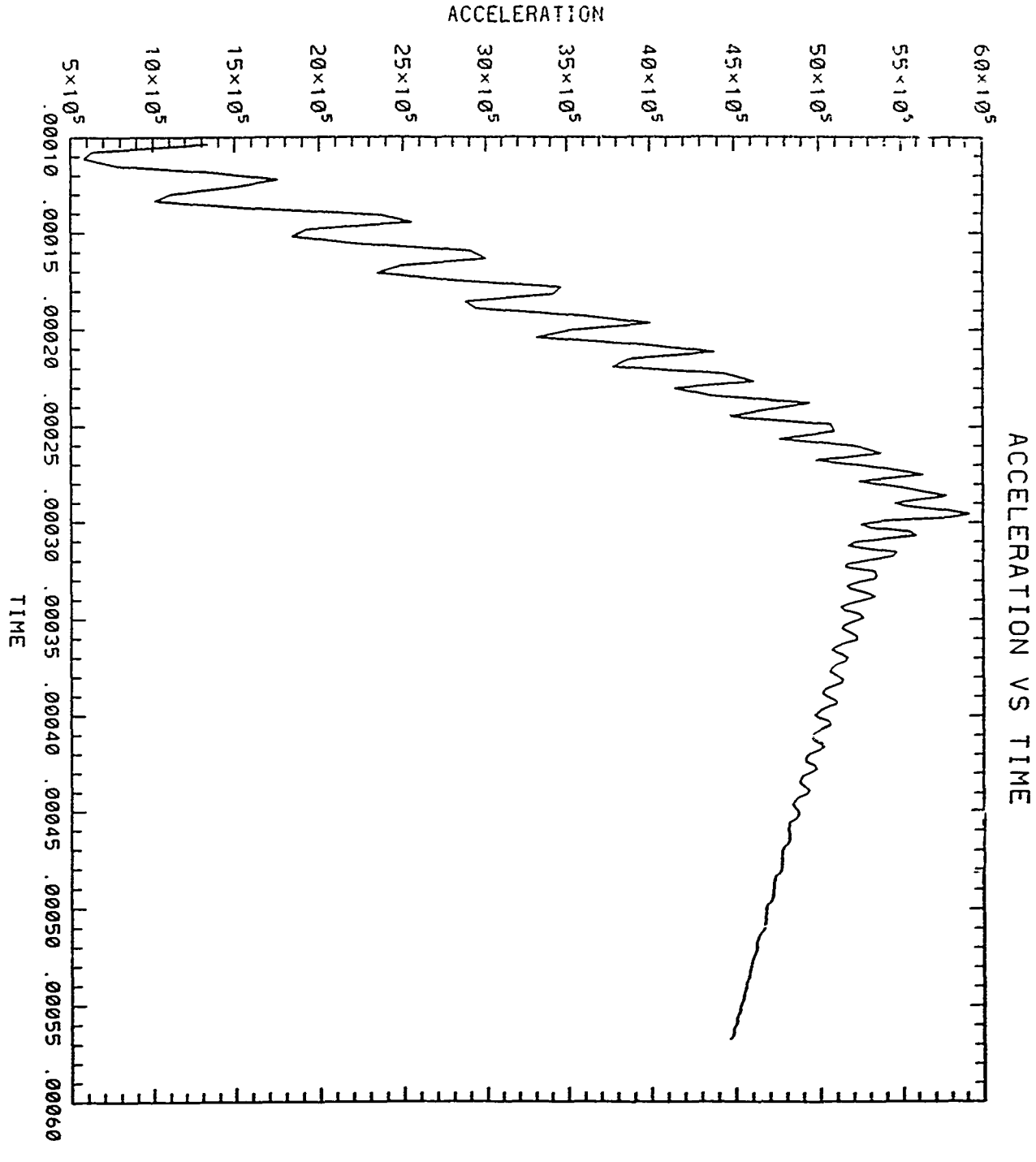

ACCELERATION
5×10⁵
10×10⁵
15×10⁵
20×10⁵
25×10⁵
30×10⁵
35×10⁵
40×10⁵
45×10⁵
50×10⁵
55×10⁵
60×10⁵
.00010
.00015
.00020
.00025
.00030
.00035
.00040
.00045
.00050
.00055
.00060
TIME
ACCELERATION VS TIME

FIGURE 5

NONLINEAR ESTIMATION FOR EXOATMOSPHERIC TRAJECTORIES: THE DAUM-BASED FILTER

Dr. Antonio A. Magliaro

Abstract

A sequential nonlinear filter based on the work of Daum and applied to the problem of estimating the location and motion of a boosting ICBM was evaluated in terms of its performance relative to the current state-of-the art estimator, the extended Kalman Filter. The Daum-based filter was modified by tuning critical parameters involving process noise, maneuver detection and the guidance law associated with the estimator. Stability of the filter, significant increase in the probability of hit, and reduction of the total miss to the target was achieved as a result. Recommendations for further testing and enhancements of the filter are discussed.

Acknowledgement

I wish to thank the Air Force Office of Scientific Research and Mr. John Provine and the Guided Interceptor Technology Branch of the Air Force Armament Directorate for providing me the opportunity to participate in this research program. I am grateful to Mr. Craig Ewing and Mr. Michael Couvillon for their assistance, and I wish to express my special thanks to Ms. Anne Carstens for serving as my focal point in this endeavor, and for her invaluable support.

I. Introduction

The proolem of estimating the location and analyzing the motion of a boosting ICBM vehicle is both challenging and critical to the SDI program. This is coupled with the requirement of guiding an exoatmospheric kinetic kill vehicle to intercept the booster. Existing methods make use of an extended Kalman Filter that utilizes linearization about its latest filtering estimates[1,2,3]. This design typically models target acceleration as a first order Gauss-Markov process. However, the assumption of linear acceleration runs counter to the expected scenario of a boost vehicle engaging in nonlinear and unpredictable acceleration due to energy management maneuvers or evasive strategies. Hence, there is a natural interest in nonlinear estimators which would accommodate these nonlinear maneuvers. Several such estimators are currently being developed and tested to meet the requirements of the SDI program. These include a nonlinear filter based on the work of Daum[4] and applied to the problem at hand by the Boeing Company[5], a "moment propagating filter" developed by Boeing[5], and a nonlinear estimator based on the work of Kolmogorov[3] and developed by the General Electric Company[6].

This report analyzes the Daum-based filter and its associated maneuver detection algorithm and guidance law as developed by Boeing. A brief description of the filter equations is presented. Modifications to the theoretical

formulation of the filter are noted, including the scheme used for injecting plant noise in the system. The maneuver detection algorithm is discussed and evaluated.

Significant improvements to the filter have been realized by changing various input noise parameters and maneuver detection and guidance law parameters. These efforts have resulted in stabilizing the filter, increasing the "probability of hit" and reducing the "total miss" to the target. Results are documented in this report through Monte Carlo simulation on various engagement scenarios. Comparisons are made with the Extended Kalman Filter and the relative strengths and weaknesses of the two estimation processes are discussed. Recommendations for further study and possible enhancements are presented.

II. The Daum-Based Filter

The general estimation problem involves estimating the state vector $x=(x_1,\ldots,x_n)$ which satisfies the differential equation

$$\dot{x} = f(x,t) + w,$$

where w represents a stochastic process. Observations are made at discrete time intervals

$$z = h(x) + v,$$

where v represents noise in the measurement process. It has been shown that the probability density function p(x,t)

satisfies the Fokker-Planck equation[3]

$$\frac{\partial p}{\partial t} = -\frac{\partial p}{\partial x} f - p\,\mathrm{tr}\left(\frac{\partial f}{\partial x}\right) + \frac{1}{2}\,\mathrm{tr}\left(Q\frac{\partial^2 p}{\partial x^2}\right),$$

where Q(t) is the process noise matrix associated with the stochastic process w(t). Assuming that p(x,t) has the (unnormalized) form

$$p(x,t_0) = \psi(x,t_0)^{\frac{1}{2}}\, e^{-\frac{1}{2}(x-m_0)^T P_0^{-1}(x-m_0)},$$

where m_0 is an n dimensional vector and P_0 is an nxn positive-definite matrix, the following conditions result:

D1: $\psi(x,t)$ itself satisfies the Fokker-Plank equation

D2: $\mathrm{tr}(\frac{\partial f}{\partial x}) + \frac{1}{4} rQr^T = x^T A x + b^T x + c$

D3: $f - \frac{1}{2}Qr^T = Dx + E,$

where $r = \frac{\partial}{\partial x}[\ln \psi(x,t)]$ and A=A(t) is a symmetric semidefinite nxn matrix, b(t) is an n-dimensional vector, c=c(t) is a scalar function, D=D(t) is an nxn matrix and E=E(t) is also an n-dimensional vector. Under these conditions, Daum proves that the unnormalized pdf is given by

$$p(x,t \mid z_1, z_2, \ldots, z_k) = \psi(x,t)^{\frac{1}{2}}\, e^{-\frac{1}{2}(x-m)^T P^{-1}(x-m)}.$$

The parameters m and P are related to the covariance M and mean $\hat{x}$ of the pdf.

Making further simplifying assumptions that f does not

explicitly depend upon t, f has the form

$$f(x) = \hat{n} + Bx + U + (\text{h.o.t.}),$$

where $$\hat{n} = f(\hat{x}), \quad B = \left.\frac{\partial f}{\partial X}\right|_{X=\hat{X}}, \quad U = \begin{bmatrix} x^T O_1 x \\ \vdots \\ x^T O_n x \end{bmatrix},$$

where O_i is a symmetric matrix and $x^T O_i x$ corresponds to the quadratic term for component f_i. This gives a simplifying second order approximation for f=f(x). Similarly, assume

$$\text{tr}\left(\frac{\partial \bar{f}}{\partial x}\right) = d + S^T x + x^T L x + (\text{h.o.t.}),$$

where d is a scalar, S is a vector, and L is a symmetric matrix, none of which depend upon t. Finally, assume that

$$r = \frac{\partial}{\partial X}[\ln \psi(x,t)] = x^T V x + g^T x + \alpha(t)$$

for some symmetric matrix V, vector g and function $\alpha(t)$. Using the above approximations, it can be shown that Daum's conditions reduce to

$$2(VQV-VB) = L + (g_1 O_1 + \ldots + g_n O_n),$$
$$g^T(2QV-B) = S^T + 2\hat{n}^T V.$$

The effect of these conditions upon the recursive filter is realized in the following generalizations of the classical Kalman Filter equations for propagating the covariance matrix M and estimate $\hat{x}$:

$$\dot{M} = BM + MB^T - M(2L+g_1 O_1 + \ldots + g_n O_n)M + Q$$
$$\dot{\hat{x}} = \hat{n} + B\hat{x} - M(2L+g_1 O_1 + \ldots + g_n O_n)\hat{x} - MS.$$

If L=[0], S=[0], $g=\vec{0}$, then the approximations are linear and the equations reduce to the Kalman Filter.

Although Daum's conditions have only been approximately satisfied, the resulting filter should outperform the Kalman Filter through the addition of the nonlinear terms $M(2L+g_1O_1+\ldots+g_nO_n)M$, $M(g_1O_1+\ldots+g_nO_n)\hat{x}$ and MS.

III. Application to Estimating Exoatmospheric Trajectories

The azimuth and elevation angles of the target in the terminal homing coordinate system (THCS) are taken as measurements for the nonlinear estimator. For the azimuth plane the filter states are defined as

$$y_1 = \psi \quad \text{(azimuth)}$$
$$y_2 = \dot{\psi} \quad \text{(azimuth rate)}$$
$$y_3 = c_y(T/M) \quad \text{(target acceleration)}$$
$$y_4 = T/gIspM \ ,$$

where T denotes the thrust, M the mass, and Isp the specific impulse of the ICBM, and c_y is the direction cosine of the target acceleration with respect to the y-axis. A parallel filter for the elevation plane employs the similarly defined states. These filters are formulated to take advantage of the direct use of the azimuth and elevation measurements.

Range axis states are also formulated in terms of the range and range rate, R and $\dot{R}$, respectively, as

$$x_1 = \dot{R}/R$$

$$x_2 = 1/R$$

$$x_3 = c_x(T/M)$$

$$x_4 = T/gIspM \ .$$

Using the above configuration, and small angle approximations, the azimuth filter state equations have the form

$$\dot{y}_1 = y_2$$

$$\dot{y}_2 = -x_2x_3y_1 - 2x_1y_2 + x_2(y_3 - A_{Iy})$$

$$\dot{y}_3 = y_3y_4 + w_1$$

$$\dot{y}_4 = y_4^2 + w_2 \ ,$$

where A_{Iy} is the y component of the interceptor's acceleration and w_1, w_2 are stochastic processes which reflect the uncertainty in acceleration due to target maneuvering and uncertainty in typing the booster, with associated process noise matrix Q(t), given by

$$Q(t) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & q_1 & 0 \\ 0 & 0 & 0 & q_2 \end{bmatrix} \qquad q_1 = q_1(t), \quad q_2 = q_2(t).$$

similar form. Note that nonlinear term appear in the last two state equations.

Applying Daum's conditions leads to the stipulation that there exist a vector all of whose components are zero except for the fourth element which is denoted by g, and a matrix V

whose entries are all zero except for the (4,4) element designated by v, satisfying the equations

$$g = 2(q_2 v^2 - 2y_4 v)$$

$$2(q_2 v - y_4)g = 3 + 2y_4{}^2 v.$$

If v=v(t) is allowed to vary, then

$$q_2 = 2y_4/v + g/2v^2$$

and

$$g = -y_4 v - \sqrt{3v(1+vy_4{}^2)}.$$

For stability purposes, the parameter v was given two different formulations as exponentially decreasing functions of time: v_1 to be used for all propagation equations except for the one associated with the fourth state, which makes use of v_2. Hence, two corresponding values of g result,

$$g_1 = -y_4 v_1 - \sqrt{3v_1(1+v_1 y_4{}^2)}$$

$$g_2 = -y_4 v_2 - \sqrt{3v_2(1+v_2 y_4{}^2)}.$$

IV. Maneuver Detection Algorithm and Guidance Law

The angle state Daum-based filter has been equipped with a maneuver detection scheme which monitors a shifting set of weighted azimuth and elevation residuals given by

$$Y_r = Az_r{}^2/(yP_{11} + v_f)$$

$$Z_r = El_r{}^2/(zP_{11} + v_f),$$

where Az_r, El_r are the azimuth and elevation measurement

residuals, and yP_{11}, zP_{11} are the corresponding covariance matrix entries, and v_f is the variance of the measurement noise distributions. After a minimum of 240 samples are processed, enabling the filter to "settle down", a χ^2-test is performed, whereby a maneuver is said to have taken place if either Az_r or El_r exceed a value of .004 for more than m=8 times out of the last n=12 samples. A detection results in the refiltering of the past 100 measurements. The process noise $q_1(t)$, which is defined as an exponentially decaying function of time, is reset to its original initialization value, and the estimates $\hat{y}_1, \hat{y}_2, \hat{y}_3, \hat{z}_1, \hat{z}_2, \hat{z}_3$ are set to their previous values at the start of the refiltering time. After a first detection, 80 sample measurements are processed before a second detection is deemed justified.

The angle state filter uses an augmented proportional navigation guidance law in steering the interceptor to its target. This consists of proportional navigation[7] augmented by a correction term involving the current estimate of the target's acceleration, and is given by

$$U_{cy} = (1+a(tg)^2)\Delta(v_c\dot{\psi} + \tfrac{1}{2}[b+\tfrac{1}{3}(1-b)t]A_{Ty})$$

$$U_{cz} = (1+a(tg)^2)\Delta(v_c\dot{\theta} + \tfrac{1}{2}[b+\tfrac{1}{3}(1-b)t]A_{Tz}).$$

Here, U_{cy} and U_{cz} are the y and z components of the commanded acceleration, v_c is the estimated closing velocity, t is time, tg is the "time to go" to the target, A_{Ty} and A_{Tz} are the y and z components of the target's acceleration and a, Δ, b are constant gain parameters.

V. Performance and Modifications to the Filter

The nonlinear filter was evaluated using the GSP-EXO simulation program originally developed by the General Electric Company and modified by the Air Force and the Boeing Company. Testing involved conducting thirty-one Monte Carlo noise sequence runs for scenarios which varied according to the orientation of the plane (relative to the THCS) in which the target is rotating, and the initializations of the target's azimuth and elevation angles in that plane. The filter performance was measured by recording the total miss distance to the center of gravity of the ICBM, computing the probability of "hit", where a hit is defined as a miss distance less than or equal to a function of the ICBM's dimensions and aspect angle, and the fuel expenditure of the interceptor in reaching the target.

A preliminary set of runs were made in which the boosting ICBM made a continuous turn with an azimuth angle rate of 5^0/second for the duration of the ten-second engagement. The filter experienced instability and failed to converge for several such runs. It was found that by reducing the original value of the input parameter DVE2I from 920,000 to 420,000 the filter achieved stability for all the subsequent scenarios tested. This parameter is used in the initialization of $v_2(t)$:

$$\dot{v}_2 = -(v_2-VE2SS), \quad v_2(0)=VE2SS+DVE2I,$$

where VE2SS is the steady-state value of $v_2(t)$. The

nonlinear filter with the above modification will be denoted as Filter 1.

The experiment was expanded to include thirteen different initialization scenarios. These were used for the non-maneuvering target case, and for a more challenging scenario in which the target rotated in the azimuthal plane with a rate of 10^{o}/sec. These trajectories are defined and referred to as NM1-NM13 and M1-M13 in Table 1 and Table 2. Results for the various modified versions of the Daum-based filter and the extended Kalman Filter (EKF) are summarized in these tables. It should be noted that the EKF was configured as a Cartesian (translational) state filter and was not provided with any maneuver detection capability.

Filter 1 exhibited irregular performance for the maneuvering target cases and was conspicuously outperformed by the EKF except in scenarios M1 and M11 in which it did exceptionally better, as can be seen from Table 2. Efforts were focused on tuning the maneuver detection parameters, making the filter more sensitive to possible acceleration maneuvers. A systematic parameter analysis was made, the end result of which (Filter 2) was to reduce the χ^2-test threshold value to .002 and to reduce the number of samples to be processed before a second detection from the value of NSSD2=80 to the value of 50. This had the effect of significantly improving the probability of hit and reducing the total miss distance in all the scenarios tested for which Filter 1 displayed poor performance (M3,M4,M9,M10). These

changes enabled the estimator to make a second correction to its acceleration estimates of the target through a second detection, which allowed the acceleration covariance entries to open up, steering the estimates back to more accurate values as can be seen in Figure 1.

A further upgrade (Filter 3) involved redefining the process noise $q_1(t)$ from a decaying exponential function to a constant equal to its initial value of 95.0. This was done by setting the decay rate parameter QP1TC to zero. Improvements in performance can be seen in significant reductions of the total miss distance for all the maneuvering trajectories except M6, and corresponding increases in the probability of hit. Improved accuracy in estimating the states is evident in Figure 2, which displays reduction in angle state errors. This upgraded filter clearly outperforms the Kalman Filter in the non-maneuvering target case, and in the head-on and tail-chase scenarios for maneuvering targets (M1,M5,M6,M11). It should be noted that the EKF consistently realized a smaller total miss distance in those trajectories for which it did hit, for both the non-maneuvering and maneuvering targets. This may be due to the EKF's more stable nature and accuracy in estimating acceleration (see Figure 3). The Daum-based filter does exhibit greater sensitivity to high rotation rates, perhaps because it processes the measurements directly. Both filters experienced difficulty in scenario M5 in which the booster is initially accelerating vertically upwards.

Several runs were undertaken in which the nonlinear terms in the propagation equations of the Daum-based estimator were suppressed. The resulting performance was consistently inferior to that of the Kalman Filter, suggesting that the latter's associated guidance law was more effective. The gain parameters of the nonlinear estimator's guidance law were varied in an effort to make it more closely resemble that of the EKF. Optimal performance occurred with the constant gain factor of the augmentation term reduced from the value of 0.5 to 0.25 to give

$$U_{cv} = (1+a(tg)^2)\Delta(v_c\dot{\psi} + \frac{1}{4}[b+\frac{1}{3}(1-b)t]A_{Ty})$$

$$U_{cz} = (1+a(tg)^2)\Delta(v_c\dot{\theta} + \frac{1}{4}[b+\frac{1}{3}(1-b)t]A_{Tz}).$$

The resulting Filter 4 showed significant improvement in performance for the non-maneuvering scenarios, with a probability of hit of 100% throughout, and comparable performance for the maneuvering cases. It did compare unfavorably with Filter 3 and the EKF in scenarios M3 and M8, both characterized by an initial azimuth angle of 135° and a maneuver which results in a high angle rate across the line of sight. Further tuning and modification of the guidance laws may realize additional improved performance in minimizing the total miss distance and reducing the expenditure of fuel in reaching the target.

VI. Conclusion and Recommendations

The angle-state Daum-based filter with maneuver detection capability has been assessed with respect to the state-of-the-art extended Kalman Filter, of which it is a generalization. This report has presented evidence that, through modifications in process noise inputs and maneuver detection and guidance law parameters, a stable filter results which in most cases exceeds the performance of the EKF in the scenarios tested.

Further testing and tuning of the filter is recommended. Specifically, a parameter study of the various gain parameters of the guidance law should result in greater effectiveness of the estimator. Stability properties of the filter should be investigated further, especially with respect to the covariance terms associated with the acceleration-related third and fourth states of the filter. Alternate maneuver detection schemes can also be implemented (see Reference 8, for example) which may provide smoother adaptation to detected anomalies in residuals. Testing should be expanded to include a larger range of engagement scenarios, and a greater azimuth or elevation rotation rate of the maneuvering target.

NON-MANEUVERING TARGET

TRAJECTORY (angles in degrees)	PERFORMANCE PARAMETERS	FILTER 1	FILTER 2	FILTER 3	FILTER 4	EKF
NM1	Probability of hit	1.00	1.00	1.00	1.00	
ϕ = 0						
ψ = 0	MISS CEP(m)	0.24	0.18	0.21	0.39	
θ = 0						
	FUEL CEP(kg)	0.71	0.68	0.62	0.36	
NM2	Probability of hit	0.87	0.94	1.00	1.00	
ϕ = 0						
ψ = 45	MISS CEP(m)	0.78	0.81	0.395	0.42	
θ = 0						
	FUEL CEP(kg)	0.74	0.72	0.95	0.89	
NM3	Probability of hit	1.00	1.00	1.00	1.00	
ϕ = 0						
ψ =135	MISS CEP(m)	0.63	0.54	0.25	0.16	
θ = 45						
	FUEL CEP(kg)	0.72	0.70	1.05	1.07	
NM4	Probability of hit	0.94	0.90	1.00	0.03	
ϕ = 0						
ψ = 90	MISS CEP(m)	0.87	0.80	0.41	22.37	
θ = 0						
	FUEL CEP(kg)	0.77	0.72	1.08	1.04	
NM5	Probability of hit	0.94	0.97	1.00	0.00	
ϕ = 0						
ψ = 0	MISS CEP(m)	0.81	0.77	0.31	20.66	
θ = 90						
	FUEL CEP(kg)	0.81	0.76	1.07	1.03	
NM6	Probability of hit	1.00	1.00	1.00	1.00	
ϕ = 0						
ψ =180	MISS CEP(m)	0.28	0.29	0.31	0.20	
θ = 0						
	FUEL CEP(kg)	0.63	0.63	0.59	0.32	
NM7	Probability of hit	0.90	0.94	1.00	0.03	
ϕ = 0						
ψ = 90	MISS CEP(m)	0.80	0.76	0.41	22.37	
θ = 45						
	FUEL CEP(kg)	0.72	0.73	1.08	1.04	

ϕ = plane rotation angle
ψ = azimuth angle
θ = elevation angle

Table 1

NON-MANEUVERING TARGET

TRAJECTORY (angles in degrees)	PERFORMANCE PARAMETERS	FILTER 1	FILTER 2	FILTER 3	FILTER 4	EKF
NM8	Probability of hit	1.00	1.00	1.00	1.00	1.00
ϕ = 0						
ψ =135	MISS CEP(m)	0.72	0.59	0.32	0.21	
θ = 0						
	FUEL CEP(kg)	0.66	0.67	0.87	0.80	
NM9	Probability of hit	0.90	0.90	1.00	1.00	
ϕ = 45						
ψ = 45	MISS CEP(m)	0.85	0.93	0.33	0.25	
θ = 45						
	FUEL CEP(kg)	0.81	0.76	1.19	1.23	
NM10	Probability of hit	1.00	1.00	1.00	1.00	
ϕ = 45						
ψ = 45	MISS CEP(m)	0.45	0.49	0.27	0.17	
θ =135						
	FUEL CEP(kg)	0.73	0.71	0.96	0.92	
NM11	Probability of hit	1.00	1.00	1.00	1.00	
ϕ = 90						
ψ = 0	MISS CEP(m)	0.24	0.18	0.21	0.39	
θ = 0						
	FUEL CEP(kg)	0.71	0.68	0.62	0.36	
NM12	Probability of hit	0.90	0.87	1.00	1.00	
ϕ = 90						
ψ = 45	MISS CEP(m)	0.76	0.99	0.43	0.44	
θ = 0						
	FUEL CEP(kg)	0.77	0.76	0.96	0.90	
NM13	Probability of hit	0.29	0.16	1.00	0.84	
ϕ = 90						
ψ = 0	MISS CEP(m)	3.83	4.12	1.07	1.30	
θ = 45						
	FUEL CEP(kg)	0.80	0.76	0.98	0.90	

ϕ = plane rotation angle
ψ = azimuth angle
θ = elevation angle

Table 1 (continued)

MANEUVERING TARGET ($\dot{\psi}$ = 10° /sec)

TRAJECTORY (angles in degrees)	PERFORMANCE PARAMETERS	FILTER 1	FILTER 2	FILTER 3	FILTER 4	EKF
M1	Probability of hit	1.00	1.00	1.00	1.00	0.00
ϕ = 0	MISS CEP(m)	0.58	0.96	0.59	0.43	93.49
ψ = 0	FUEL CEP(kg)	1.14	1.14	1.18	1.24	1.11
θ = 0						
M2	Probability of hit	0.87	0.97	1.00	1.00	1.00
ϕ = 0	MISS CEP(m)	1.41	1.79	0.99	1.11	0.20
ψ = 45	FUEL CEP(kg)	0.94	0.95	1.00	1.01	0.99
θ = 0						
M3	Probability of hit	0.00	0.61	0.81	0.35	1.00
ϕ = 0	MISS CEP(m)	7.06	2.52	0.78	4.30	0.27
ψ =135	FUEL CEP(kg)	1.25	1.22	1.31	1.33	1.21
θ = 45						
M4	Probability of hit	0.23	1.00	0.94	1.00	1.00
ϕ = 0	MISS CEP(m)	7.37	2.10	0.72	1.16	0.15
ψ = 90	FUEL CEP(kg)	0.92	0.90	1.00	0.92	0.56
θ = 0						
M5	Probability of hit		0.00	0.48	0.32	0.00
ϕ = 0	MISS CEP(m)		2.95	1.80	1.81	31.57
ψ = 0	FUEL CEP(kg)		1.42	1.50	1.49	1.28
θ = 90						
M6	Probability of hit		0.87	0.77	1.00	0.00
ϕ = 0	MISS CEP(m)		1.12	2.18	0.95	67.29
ψ =180	FUEL CEP(kg)		1.09	1.10	1.20	1.01
θ = 0						
M7	Probability of hit		0.45	1.00	1.00	1.00
ϕ = 0	MISS CEP(m)		2.53	0.67	1.13	0.20
ψ = 90	FUEL CEP(kg)		1.28	1.29	1.30	1.04
θ = 45						

ϕ = plane rotation angle
ψ = azimuth angle
θ = elevation angle

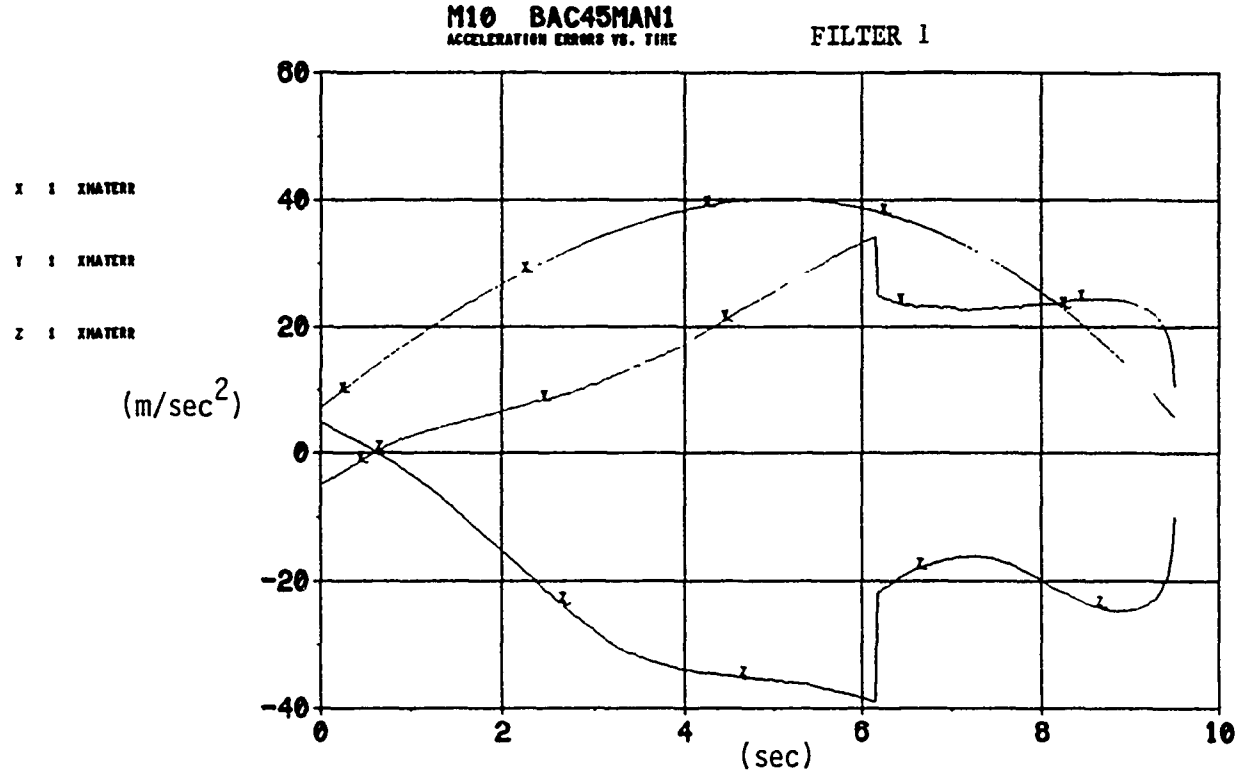
Table 2

MANEUVERING TARGET ($\dot{\psi}$ = 10°/sec)

TRAJECTORY (angles in degrees)	PERFORMANCE PARAMETERS	FILTER 1	FILTER 2	FILTER 3	FILTER 4	EKF
M8	Probability of hit		0.71	1.00	0.03	0.97
ϕ = 0						
ψ =135	MISS CEP(m)		0.98	0.41	19.96	0.73
θ = 0						
	FUEL CEP(kg)		1.15	1.16	1.03	0.85
M9	Probability of hit	0.29	0.90	1.00	0.97	1.00
ϕ = 45						
ψ = 45	MISS CEP(m)	9.17	2.34	1.85	2.03	0.14
θ = 45						
	FUEL CEP(kg)	1.12	1.11	1.30	1.11	0.92
M10	Probability of hit	0.00	0.94	1.00	1.00	1.00
ϕ = 45						
ψ = 45	MISS CEP(m)	5.62	1.08	0.38	0.37	0.16
θ =135						
	FUEL CEP(kg)	1.30	1.28	1.33	1.26	0.99
M11	Probability of hit	1.00	1.00	1.00	1.00	0.00
ϕ = 90						
ψ = 0	MISS CEP(m)	0.65	0.86	0.61	0.45	100.32
θ = 0						
	FUEL CEP(kg)	1.13	1.13	1.16	1.24	1.06
M12	Probability of hit		1.00	1.00	1.00	1.00
ϕ = 90						
ψ = 45	MISS CEP(m)		1.58	0.97	1.17	0.14
θ = 0						
	FUEL CEP(kg)		0.93	0.99	1.01	0.95
M13	Probability of hit		0.97	1.00	1.00	1.00
ϕ = 90						
ψ = 0	MISS CEP(m)		3.54	2.07	1.76	0.18
θ = 45						
	FUEL CEP(kg)		0.93	0.98	1.03	0.97

ϕ = plane rotation angle
ψ = azimuth angle
θ = elevation angle

Table 2 (continued)



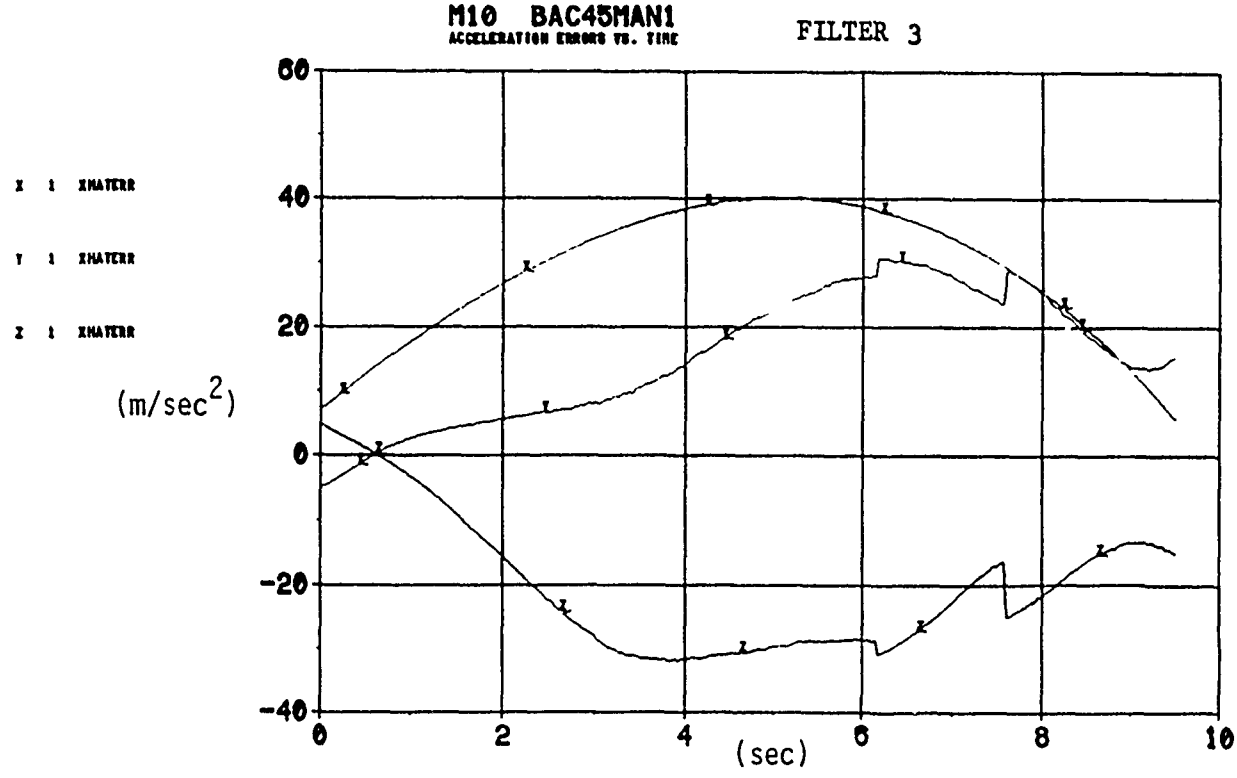
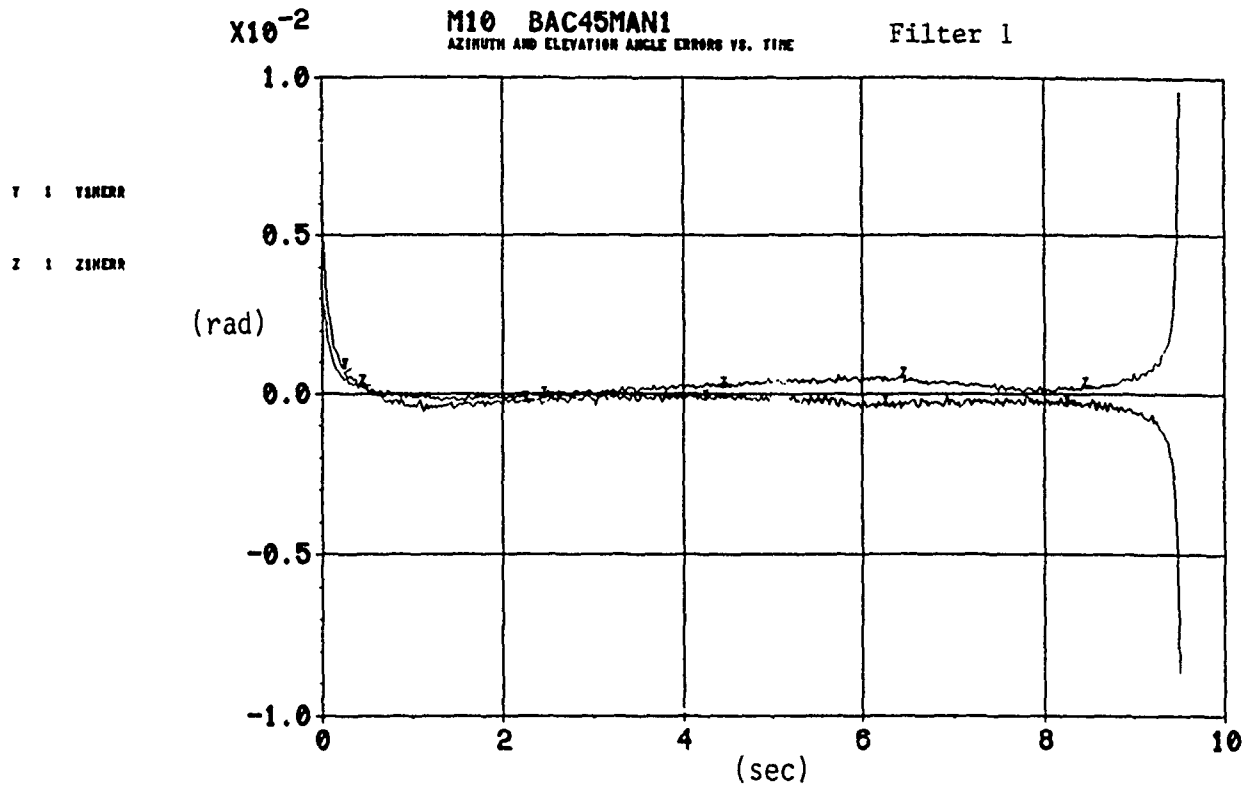


Figure 1



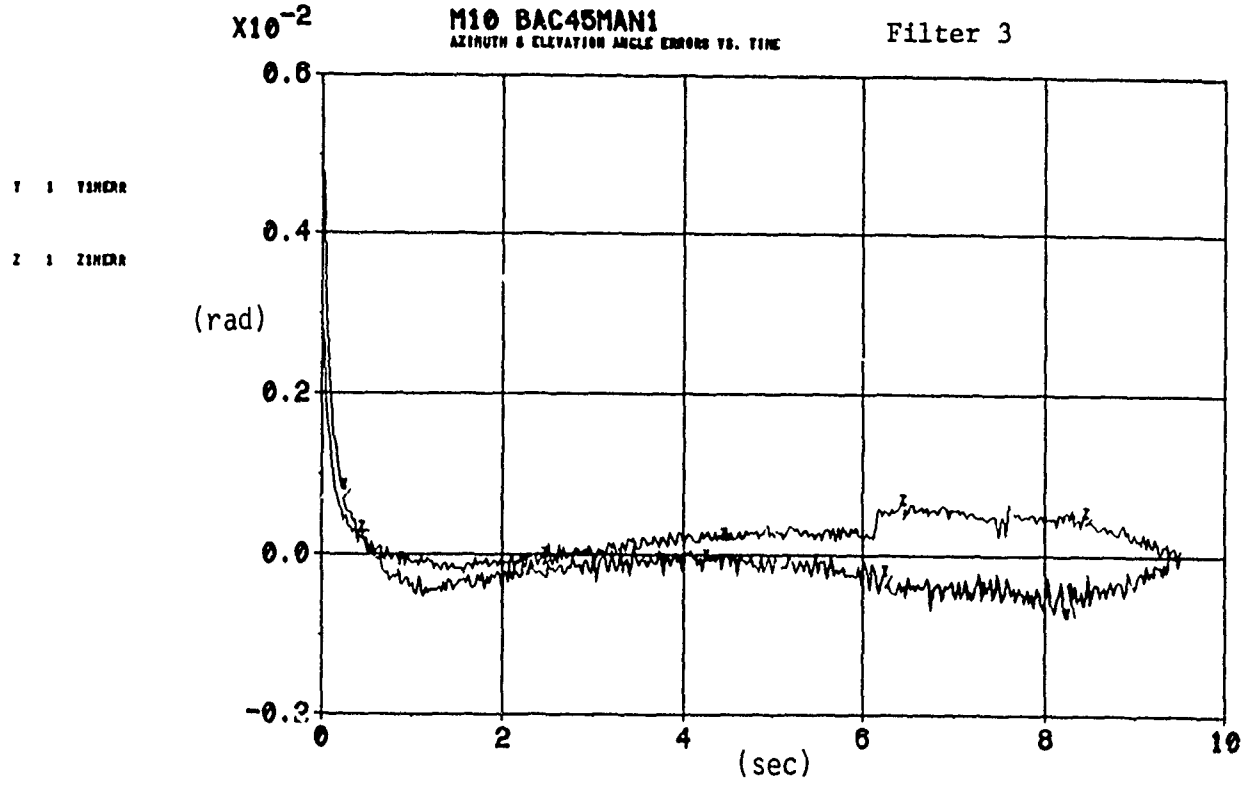


Figure 2

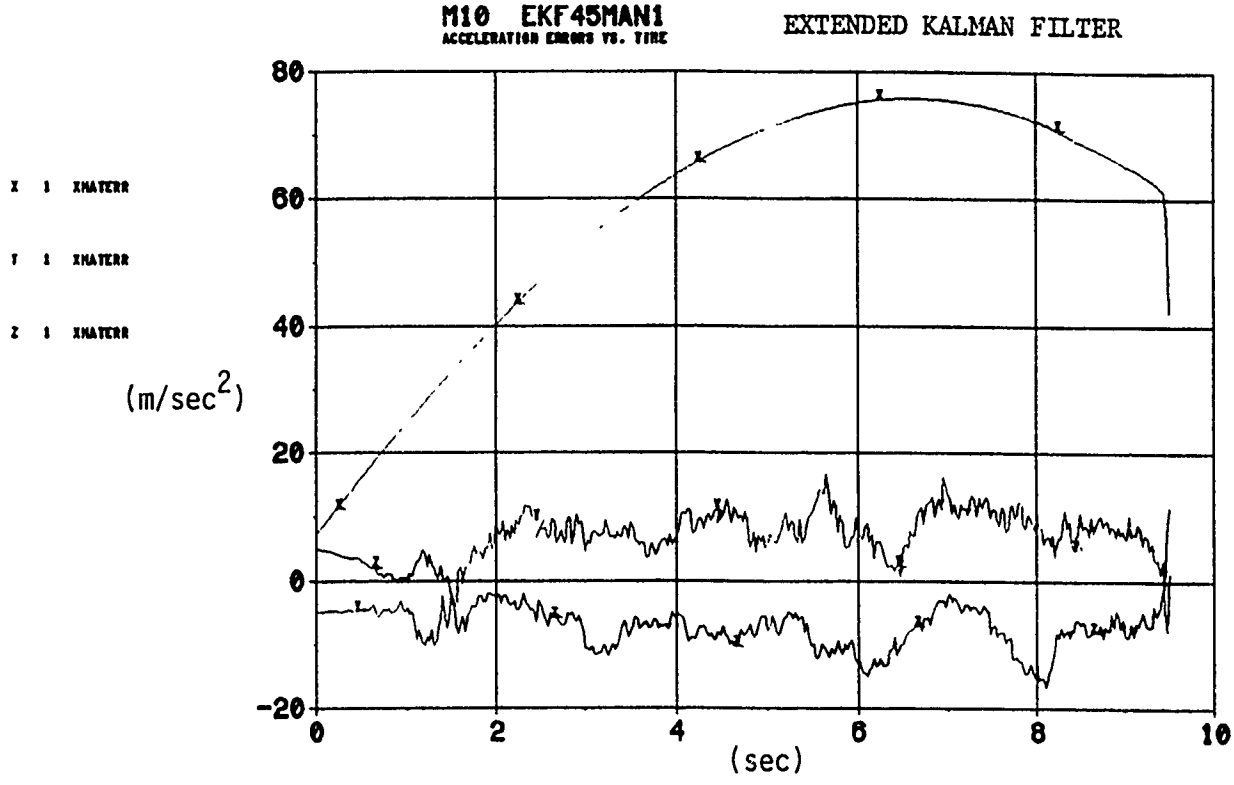

M10 EKF45MAN1
ACCELERATION ERRORS VS. TIME
X 1 XMATERR
Y 1 XMATERR
Z 1 XMATERR
(m/sec^2)
80
60
40
20
0
-20
0
2
4
6
8
10
(sec)

Figure 3


CREATED BY:
MAKE:MA/1AR0
ORG :SCSN
DATE: 0-AUG-01
TIME:00:14:07

References

1. Kalman,R.E.,Bucy,R.S., "A New Approach to Linear Filtering and Prediction Problems, Journal of Basic Engineering, March 1961.

2. Gelb,A.,Applied Optimal Estimation, M.I.T. Press, Cambridge, Massachusetts, 1974.

3. Jazwinski,A.H.,Stochastic Processes and Filtering Theory, Academic Press, New York, 1970.

4. Daum, F.E., "Exact Finite-Dimensional Nonlinear Filters", IEEE Transactions on Automatic Control, Vol. AC-31, No.7, July 1976.

5. "Final Report for Nonlinear Estimation of Atmospheric Trajectories", The Boeing Company, Boeing Aerospace and Electronics Division, Contract No. F08635-89-C0096, Seattle, Washington, February, 1991.

6. Gido,J.,Stuckwisch,S.,Trailie,J.,Zenor,P., "Non-Linear Estimation for Exoatmospheric Trajectories (NEET)", General Electric Company, WL/MN-TR-91-09, May 1991.

7. Locke,A.S., Guidance, D. Van Nostrand, Co., 1955.

8. Murphy,D.J.,Guimond,B.W.,Moore,D.W., "Linear Array Tracking of a Maneuvering Target", Proceedings of the Fifteenth Asilomar Conference on Circuits, Systems and Computers, Pacific Grove, California, November 1981.

1991 USAF SUMMER FACULTY RESEARCH/
GRADUATE STUDENT RESEARCH PROGRAM

Sponsored by the

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

Conducted by

RDL-SUMMER RESEARCH PROGRAM

FINAL REPORT

PRACTICAL CONSIDERATIONS FOR A FIRST CUT MULTI-SENSOR FUSION SEEKER

Prepared by:	Charlesworth R. Martin
Academic Rank:	Associate Professor
Department and University:	Norfolk State University
Research Location:	AFAL/MNGA Eglin AFB Florida 32542-5434
USAF Focal Point:	Martin Moorman
Date:	20 September 1991

PRACTICAL CONSIDERATIONS FOR A FIRST CUT MULTI-SENSOR FUSION SEEKER

by

Charlesworth R. Martin

ABSTRACT

The significant technical issues impacting on design are explored. The selection of an appropriate sensor suite. The physical constraints imposed on sensor geometry, side-by-side, or common aperture. The challenges of maintaining an appropriate radome geometry to achieve low aerodynamic drag. Decision on what point in the implementation should fusion actually occur for the data from the selected sensor suite and the most practical algorithm for doing so. Multi-sensor fusion seeker performance trade off with hardware complexity, cost associated with the fusion process, packaging challenges and increase computational load.

INTRODUCTION

Sensor fusion is the process of merging the information for multiple sensors into a unified representation that is in some way better than the interpretations of any of the individual sensors. Broadly, the fusion process can occur on three levels. The first is the fusion of raw data which leads to the interpretation of "raw data fusion". The second is the fusion of processed data, referred to in the literature as feature level fusion. The third is the fusion of decisions referred to as decision level fusion.

All existing missile seekers use either radar or infrared sensors for the purpose of guiding a missile to its target. Multi-sensor fusion for seeker applications is motivated by the failure of a single sensor system to perform robust target detection, tracking and classification under conditions of adverse weather, clutter and countermeasures. As the technology of single sensor systems approaches a point of diminishing return, the present focus is towards the incorporation of multiple sensors such as radar, IR, LADAR, etc., to exploit the synergism in a maximum way.

Multi-sensor fusion is a relatively young discipline. However, the quantity of research papers published in conference proceedings and technical journals is quite astounding.

The Systems Branch (MNGA) of the Advanced Guidance Division of the USAF Armament Laboratory at Eglin Air Force Base in its efforts to provide focus for the development of a multisensor seeker proposed this research effort to ascertain what is viable from amongst the myriad of theoretical possibilities.

OBJECTIVES

The discipline of multi-sensor fusion has not yet matured. As a result, in spite of the volume of literature that has been published in this area, no clear consensus has emerged amongst researchers as to the most feasible way to achieve multi-sensor fusion. A clear perspective is needed in this area if precious time and valuable resources are not to be wasted in moving from fusion concepts to fusion systems.

The objectives of this research are: (A) Development of a rationale for the selection of a given sensor suite based on the scenario of a medium range air-to-air missile. (B) Ascertain at what point in the multi-sensor implementation fusion should actually occur. (C) Investigate the suitability of the various ways to integrate the data from the various sensors so as to achieve a performance superior to that of a single sensor. (D) Put forth a set of recommendations for a near-term multi-sensor seeker design.

RATIONAL FOR SELECTION OF A GIVEN SENSOR SUITE

The available sensors for missile applications fall into three basic categories: (a) Radar sensors (whose precise classification is determined by the operating frequency range), (b) Infrared sensors which can be either passive or active, and (c) A digital uplink data receiver.

To fully appreciate the selection of a given sensor suite a brief synopsis of the advantages and disadvantages of each sensor is given.

RADAR SENSORS

(a) ADVANTAGES

1. Able to achieve accurate range and range-rate easily
2. Performs well in inclement weather, i.e., good fog, cloud and smoke penetration

(b) DISADVANTAGES

1. Susceptibility to electronic countermeasures.
2. Radar imaging is impractical.
3. Adequate resolution is difficult to obtain with practical apertures.

PASSIVE INFRARED (IR) SENSORS:

(a) ADVANTAGES

1. Able to produce high resolution images with a relatively small aperture.

2. Passive IR technology is mature.

(b) DISADVANTAGES

1. Passive IR has poor bad weather penetration capability.

2. Passive IR does not provide direct range and range-rate information.

ACTIVE INFRARED (LASER RADARS) SENSORS:

(a) ADVANTAGES

1. Provide good range and angular resolution information.

2. Good countermeasure resistance.

3. Provide a stable target profile.

4. Better bad weather penetration than passive IR.

(b) DISADVANTAGES

1. Produce speckled images and are sensitive to target glints.

2. LADAR technology is not yet mature and production experience is limited.

DIGITAL UPLINK DATA RECEIVER

(a) ADVANTAGE

1. Provides continuous target information update.

(b) DISADVANTAGES

1. Requires fighter to remain in threat environment.

2. Contributes to processing overload.

3. At variance with the requirement of an autonomous missile i.e., fire and forget.

Sensor suite selection is also impacted by design considerations, such as: the practicality of implementing a common aperture for the radar and LADAR sensors, the feasibility of providing optical windows(s) for IR Sensor(s) without adversely affecting the performance capability of either sensor. In view of these difficulties and in light of the real senergism to be had from dissimilar sensors, the picture that emerges is that the two most promising sensor suite candidates are: (a) passive IR and a MMV radar and (b) a passive IR and a LADAR sensor.

FUSION LEVEL DETERMINATION

The fusion of multiple sensors reduces the reliance on any single sensor or sensor type. In a general way, this can improve system performance even with the loss of individual sensor performance. Thus making the overall system fault-tolerant and robust. The real-time demand of multi-sensor fusion seekers requires the use of automation for processing the data, and making decisions based on the extracted information. The efficiency of the combined utilization of the sensors hinges on the manner in which the different sensor

information is combined. The fusion process may be serial or parallel or a combination of the two. Fusion may be carried out at the data level, feature level or decision level. The determination of the appropriate fusion level determines the nature and complexity of the fusion task.

DATA LEVEL FUSION

Intuitively, data or measurement fusion should provide the highest degree of confidence since no data is as yet lost. This is borne out by a filtering and tracking example [1]. A form of data fusion applicable to high resolution imaging sensors like FLIR and LADAR is known as pixel-level fusion. This is the lowest level of fusion and requires perfectly registered sensors and high communication bandwidth between the sensors. Pixel-level fusion has the potential benefit of increasing the number of pixels on target thus improving detection performance, feature extraction and leads to improved target classification or recognition. There are real hindrances to the use of pixel level fusion. The data must be taken from the same location with commensurate sensors i.e., sensors whose pixel-level data values (intensities etc.,) relate to the same target signature characteristic [2,3]. The difficulty of implementation in real time hardware, and limited applicability.

FEATURE FUSION

Feature level fusion requires some pre-processing to first extract from each sensor, features relating to target attribute. The features are then combined and subsequently classified. One way to combine the features is to concatenate the feature vectors to form a longer vector that would then be channeled to the classifier.

Feature level fusion is not without difficulty. Features and segments from different sensors may not match. This may be due to the inherent nature of what each sensor responds to, or to differences in the way the segmentation algorithms work [2,13].

DECISION LEVEL FUSION

Decision level fusion is generally considered the easiest method of fusion. It can be accomplished by series, parallel, and a mixture of series and parallel methods. In a series implementation the decision from one sensor is used by the next sensor to obtain it's own measurements and decision. In parallel decision fusion, the weighted average for the decisions of individual sensors are used to devise an optimum decision rule [14]. The three main issues challenging the implementation of decision level fusion are: (a) Fusion logic is critical. (b) Considerable loss of information due to more processing and (c) Local optimization for each sensor does not necessarily imply global optimality.

To ascertain the level at which fusion should occur demands the selection of an appropriate sensor suite. If it is assumed that the selected sensor suite consists of MMW and FLIR sensors, then the choice is narrowed to feature level fusion and decision level fusion. The reason is that pixel level fusion has no meaning for these two widely dissimilar sensor mode of operation.

MULTI-SENSOR INFORMATION INTEGRATION METHODS

For the sensor suite considered it is assumed that each sensor has its own built-in processing capability and provides its decision as well as the feature set extracted from its observations. This information provided to the fusion processor is rarely known with complete confidence because of lack of sensor coverage, inaccuracies in measured data or ambiguities in the sensor reports. Various uncertainty calculi have been developed that can handle the uncertain associated with sensor information. The fusion algorithms used must provide more accurate decisions thus enhancing detection and tracking capability [6,7,8,10].

The more prominent calculi probability theory being focused upon in the context of fusion research are: The Bayesian approach, Dempster-Shafer theory, and fuzzy set theory. The central feature of all probabilistic methods is the requirement of having prior probabilities to describe the

population. Below are listed the advantages and disadvantages of each approach.

BAYESIAN APPROACH

(a) ADVANTAGES:

1. Computational methods based on Baye's rule are axiomatic and have well-understood mathematical properties.

2. Generally require only a modest amount of computation.

(b) DISADVANTAGES

1. The statistical analyses to determine the prior probabilities tend to require large amount of data.

2. An exponential amount of prior probabilities are required.

DEMPSTER-SHAFER THEORY

(a) ADVANTAGES

1. The information and time complexity can be quite low if certain conditions are met.

2. The notion that evidence can be relevant to subsets, rather than just singletons is intuitively pleasing.

(b) DISADVANTAGES

1. The assumption that pieces of evidence are independent.

2. No theoretical justification for the combination rule [6,7].

FUZZY-SET THEORY

(a) ADVANTAGES

1. Fuzzy set theory techniques are quite flexible.

2. Have low information and time complexity (depending on the definitions of the operators and the particular methods used).

(b) DISADVANTAGES

1. At times it is unclear how to construct reasonable membership functions.

2. Its inherent flexibility can be a disadvantage in determining which method to use to solve a problem.

The number of sensors used in an air-to-air sensor fusion seeker is likely to be two and perhaps occasionally three. Probabilities relating to these sensors can be ascertained from captive flight data or other test-bed arrangements. Thus the Bayesian and Dempster-Shafer fusion methods are likely to be the two leading contenders.

RECOMMENDATIONS

The real motivation for the implementation of multisensor fusion in a medium range air-to-air missile is for optimization target detection and tracking brought about by the failure of single sensor systems to do robust targeting under conditions of heavy clutter, adverse weather, and countermeasures.

The development of hardware systems to meet this challenge requires that answers be given to the following questions:

(a) What sensor suite can best provide the broadest coverage of target information and yet be feasible?

(b) At what level should fusion occur?

(c) What methodology can best accomplish fusion?

Based on this investigation the following selection is offered as a tentative solution.

Sensor suite - MMV and FLIR
Fusion level - feature
Fusion algorithm - Bayes estimator (for classification) and Neyman-Pearson Test (for detection).

CONCLUSION

This paper provides the basis for an understanding of the issues pertinent to the implementation of sensor fusion for the case of an air-to-air multi-sensor seeker. Implementation of any of the three levels of fusion, pixel, feature, and decision levels should only be accomplished after an adequate trade-off analysis between, sensor suite, fusion level, sensor co-location difficulties, aerodynamic penalty and cost of implementation.

From my survey, none of the research papers reviewed provided any information on quantifying the improvement of multi-sensor fusion over single sensor performance. This should be investigated.

Acknowledgements

I wish to express my deep indebtedness to Martin Moorman, Larry E. Lewis, Lt. Col. Franklin, Ellis J. Boudoeaux Jr., and Dr. Dennis Goldstein for the discussions we have had and the enjoyable working atmosphere they fostered at The Air Force Armament Laboratory's Advanced Guidance Division, Eglin AFB.

Finally, I gratefully acknowledge the financial support I received from the USAF Summer Research Program in the form of a Summer Faculty Award.

REFERENCES

1. J.A. Roecker and C.D. MeGillem, Comparison of two-sensor tracking methods based on state vector fusion and measurement fusion, IEEE Trans., AES-24, pp. 447-449, (1988).

2. "A Hierarchical Structure Approach To Multisensor Information Fusion" By Alianna J. Maren, Robert M. Pap and Craig T. Harston. GACIAC PR-89-01 pp. 7-19, (1989).

3. "Sensor Fusion For Airborne Platform", By D. Rai, GACIAC PR-89-01 pp. 71-81, (1989).

4. "Optimal Decision Fusion in Multiple Sensor Systems" By Stellios C.A. Thomopoulos et.al., IEEE Trans. On Aerospace And Electronic Systems. VOL. AES-23, NO. 5 pp. 644-653, Sept. 1987.

5. "Optimal Data Fusion in Multiple Sensor Detection Systems", By Z. Chair and P.K. Varshney, IEEE Trans. On Aerospace and Electronic Systems. VOL. AES-22, NO. pp. 98-101 Jan. 1986.

6. "Shafer-Dempster Reasoning with Applications to Multisensor Target Identification Systems" By Philip L. Bogler, IEEE Trans. On Systems, Man, and Cybernetics, VOL., SMC-17, NO. 6 pp. 968-977 Nov./Dec. 1987.

7. "An Analysis of Four Uncertainty Calculi, By Steven J. Henkind And Malcolm C. Harrison, IEEE Trans. On Systems, Man, And Cybernetics, VOL. 18, NO. 5, pp. 700-714, Sept./Oct. 1988.

8. "Multisensor Data Fusion", by Edward Waltz and James Llinas, Artech House, Inc., Norwood, MA, 1990.

9. "Introduction To Sensor Systems", by S.A. Hovanessioan, Artech House, Inc. Norwood, MA, 1988.

10. Samuel S. Blackman, Multiple-Target Tracking with Radar Applications, Artech House, Inc., Norwood, MA, 1986.

11. R.C. Horney, Dual Active/Passive Imaging Systems, Optical Engineering, VOL. 20, NO. 6, pp. 976-980, Nov./Dec. 1981.

12. IR and MMW Sensor Fusion for precision Guided Munitions, By J.A. Hoschette and C.R. Seashore, SPIE VOL. 931 Sensor Fusion pp. 124-130, 1988.

13. Mitiche, A., Aggarwal, J.K. "Multiple sensor integration fusion through image processing: A review," Optical Engineering, pp. 380-386 March 1986.

14. Z. Chair and P.K. Varshney, Optimal data fusion in multiple sensor detection systems, IEEE Trans., AES-22, pp. 98-101 (1986).

NEWTON'S METHOD SOLVERS FOR THE NAVIER-STOKES EQUATIONS

Dr. Paul D. Orkwis

Assistant Professor

University of Cincinnati

Abstract

Several issues involving improvements to the performance of Orkwis and McRae's exact Jacobian Newton's method solver for the Navier-Stokes equations have been explored. Development and storage of the sparse Jacobian matrix has been greatly simplified so that expensive sorting routines can be avoided. An evaluation of local timestepping versus global timestepping was performed which indicates that global timestepping is superior for a variety of reasons. The conjugate gradient squared (CGS) scheme was added to the method as an alternative to a direct LU decomposition solution of the system $A\mathbf{x} = \mathbf{b}$. Preconditioners based on incomplete LU (ILU) factorizations were tested. Comparisons were made between exact and approximate inversion routines and between exact and approximate Jacobian matrices. Results indicate that the approximate Newton schemes work well for cases with weak shock waves but perform poorly as shock strength increases.

Introduction

The goal of the current research is to improve the Newton's method solver for the Navier-Stokes equations originally developed by Orkwis and McRae [1, 2]. This

scheme has exhibited the ability to obtain quadratic convergence rates for complex discretizations of the Navier-Stokes equations such as Roe's flux difference splitting (FDS). Improvements were sought which would reduce computational and memory requirements while still permitting quadratic or near quadratic convergence rates. Several improvements were made to the original scheme to increase the efficiency of the Jacobian matrix inversion, improve the timestep variation procedure, and test several basic variations on the general solution procedure.

The following sections describe this work in detail by first presenting the governing Navier-Stokes equations and the original Orkwis and McRae Newton's method procedure. A discussion of global versus local timestepping is included. The initial *LU* decomposition routine and the basic conjugate gradient squared (CGS) algorithm for inverting the sparse Jacobian matrix are then listed. Several variations on the basic CGS scheme and matrix preconditioning ideas are discussed. Finally, the results obtained from testing these new ideas numerically are presented and summarized.

Governing Equations

The Newton's method solver has been developed for the solution of the steady, two-dimensional, laminar Navier-Stokes equations, shown below:

$$\frac{\partial \mathbf{F}}{\partial x} + \frac{\partial \mathbf{G}}{\partial y} = 0 \tag{1}$$

where

$$\mathbf{F} = \begin{bmatrix} \rho u \\ \rho u^2 + p - \tau_{xx} \\ \rho uv - \tau_{xy} \\ (e+p)u - b_x \end{bmatrix} \qquad \mathbf{G} = \begin{bmatrix} \rho v \\ \rho uv - \tau_{xy} \\ \rho v^2 + p - \tau_{yy} \\ (e+p)v - b_y \end{bmatrix}$$

$$\begin{aligned}
\tau_{xx} &= \frac{2}{3}\frac{\mu}{R_e}\left(2\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right) \\
\tau_{xy} &= \frac{\mu}{R_e}\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right) \\
\tau_{yy} &= \frac{2}{3}\frac{\mu}{R_e}\left(2\frac{\partial v}{\partial y} - \frac{\partial u}{\partial x}\right) \\
b_x &= \frac{\gamma\mu}{R_e P_r}\frac{\partial e_i}{\partial x} + u\tau_{xx} + v\tau_{xy} \\
b_y &= \frac{\gamma\mu}{R_e P_r}\frac{\partial e_i}{\partial y} + u\tau_{xy} + v\tau_{yy} \\
e_i &= \frac{e}{\rho} - \frac{1}{2}\left(u^2 + v^2\right) \\
p &= (\gamma - 1)\left(e - \frac{1}{2}\rho\left(u^2 + v^2\right)\right)
\end{aligned}$$

ρ is the density, u and v are the velocity components, p is the pressure and e is the total energy. The viscosity, μ, is determined from Sutherland's law, R_e is the Reynolds number, and P_r is the Prandtl number, a constant equal to 0.72.

Equation 1 is transformed into generalized coordinates and discretized using finite differences. The resulting set of nonlinear equations is then solved via Newton's method for systems of equations.

Numerical Method

Orkwis and McRae's [1, 2] solution procedure is based on Newton's method for systems of nonlinear equations. Systems of this type have the form

$$\mathcal{F}\left(\bar{U}\right) = 0 \tag{2}$$

The general Newton's method is

$$\left(\frac{\partial \mathcal{F}}{\partial \bar{U}}\right)^n \Delta^n \bar{U} = -\mathcal{F}^n\left(\bar{U}\right) \tag{3}$$

A solution is obtained by forming $\mathcal{F}\left(\bar{U}\right)$ and the Jacobian $\frac{\partial \mathcal{F}}{\partial \bar{U}}$ at the known nth iterate. The increment $\Delta^n \bar{U}$ is then found by inverting the Jacobian matrix. The value of $\bar{U}$ at the new iterate is given by

$$\bar{U}^{n+1} = \bar{U}^n + \Delta^n \bar{U}$$

Equation 3 requires a "close" initial guess. The system is modified to obtain this guess from some arbitrary initial condition in the following way

$$\left[\frac{I}{\Delta t} + \frac{\partial \mathcal{F}}{\partial \bar{U}}\right]^n \Delta^n \bar{U} = -\mathcal{F}^n(\bar{U}) \tag{4}$$

The pseudo-timestep is changed based on the value of the residual using the equation

$$\Delta t^n = \Delta t^0 \frac{\| \mathcal{F}(\bar{U}) \|^0}{\| \mathcal{F}(\bar{U}) \|^n} \tag{5}$$

The effect of the modification is to make the scheme behave like a backward Euler solver during startup and like Newton's method as the residual is reduced.

This scheme has been shown [1, 2] to exhibit quadratic convergence once the interim solutions approach the final result.

Several modifications of the basic scheme are possible. One idea is to use a Jacobian from a less complicated equation set discretization. These might include one of the flux vector splitting ideas or a lower order FDS. However, Liou and Van Leer [3] have shown that use of Jacobians that are inconsistent with the equation set discretization may result in linear convergence rates.

Another modification of the Newton solver is to allow Δt to vary differently for each equation. The $\frac{I}{\Delta t}$ term is then replaced by a diagonal matrix composed of the timestep from each equation. The individual timesteps can be varied by using equation 5 with the local values of the RHS rather than a global norm.

Matrix Inversion Technique

One of the more difficult problems that one faces in using a full Newton method is the inversion of a large sparse matrix. For viscous flow problems a stencil of at least 9 points is required. This stencil produces 9 block entries in the matrix which results in a block tridiagonal structure. If a higher order discretization is used 13 blocks are produced and a block pentadiagonal matrix results.

Inversion of this matrix using an LU decomposition direct solver results in significant amounts of fill-in. This fill-in can be minimized [4] by node reorderings, but generally remains significant. To eliminate fill-in completely one needs something like an iterative scheme. However, these schemes usually require a significant diago-

nal terms [5] due to limitations placed on the spectral radius of the iterative system. Unfortunately these terms decrease in equation 4 as the timestep increases.

A promising alternative to the direct and iterative schemes are the conjugate gradient type methods [6, 7, 8, 9]. These schemes are iterative in nature and are related to the method of steepest descent. Diagonal dominance is not required and memory needs are significantly reduced. Several new solvers for aerospace applications [10, 11, 12, 13, 14] have recently been presented that use conjugate gradient type inversion routines based on Saad and Schultz' Generalized Minimum Residual (GMRES) algorithm. An even more recent development is the Conjugate Gradient Squared (CGS) [9] routine which is similar to the Bi Conjugate Gradient procedure. This scheme eliminates the need for a functional minimization operation, is more efficient per iteration and has a faster convergence rate.

The CGS algorithm for solving $A\mathbf{x} - \mathbf{b} = \emptyset$ with left and right preconditioners, P_L and P_R is listed below:

$$\begin{aligned} \mathbf{r} &= P_L(\mathbf{b} - A\mathbf{x}_0) \\ \mathbf{q} &= \emptyset \\ \mathbf{p} &= \emptyset \\ \rho_{-1} &= 1 \\ n &= 0 \end{aligned}$$

while $\| \mathbf{r} \| >$ *tolerance* do

begin

$$\rho = \mathbf{r}^T\mathbf{r}$$

$$\beta = \frac{\rho}{\rho_{-1}}$$

$$\mathbf{u} = \mathbf{r} + \beta\mathbf{q}$$

$$\mathbf{p} = \mathbf{u} + \beta(\mathbf{q} + \beta\mathbf{p})$$

$$\mathbf{v} = P_L A P_R \mathbf{p}$$

$$\sigma = \mathbf{r}^T\mathbf{v}$$

$$\alpha = \frac{\rho}{\sigma}$$

$$\mathbf{q} = \mathbf{u} - \alpha\mathbf{v}$$

$$\mathbf{v} = \alpha P_R(\mathbf{u} + \mathbf{q})$$

$$\mathbf{x} = \mathbf{x} + \mathbf{v}$$

$$\mathbf{r} = \mathbf{r} - P_L A\mathbf{v}$$

$$n = n + 1$$

$$\rho_{-1} = \rho$$

end

The above scheme is quaranteed to converge (in exact arithmetic) in at most N iterations, where N is the number of equations. However, in general the solver converges before this point. Additional storage is required only for the preconditioning

matrix and approximately four vectors of length N. A proper preconditioning matrix as well as a good initial guess play a big role in determining the actual convergence rate of the CGS scheme.

Equation 4 together with a CGS matrix inversion routine lends itself to several approximate variations. Just as the exact Jacobian matrix can be approximated by a simpler or lower order approximation, the CGS inverse routine can be halted before convergence to produce an approximate inversion of the Jacobian matrix. The basic variations are to either stop the CGS process after a selected number of subiterations, regardless of the residual level, or to stop the iteration after some residual level greater than machine zero has been reached. In a limiting sense the latter will result in a solution that is equal to the initial guess sent to the CGS routine. Therefore, the preconditioner effectively approximates the inverse of the original system. Other initial guesses as possible, such as $\mathbf{x}_o = \emptyset$ or $\mathbf{x}_o{}^n = \mathbf{x}_o{}^{n-1}$, but they themselves do not lead to approximate solutions of the original system.

Preconditioners

A large part of the success of the CGS algorithm is played by the soundness of the preconditioner. The reason one needs a preconditioning matrix is to reduce the overall operation count of the CGS routine by solving an equivalent system with more suitable eigenvalues. To accomplish this many different matrices are available. The $ILU(0)$ scheme and a modification will be discussed.

Since Orkwis and McRae's original Newton's method solver employed an LU

decomposition inversion routine, the most easily coded preconditioners were of the LU type. $ILU(0)$ is one of a general class of incomplete LU decompositions referred to as $ILU(n)$ approximate factorizations. The values of n denote the degree to which fill-in is allowed. $ILU(0)$ refers to a scheme which allows zero fill-in. The greater n is, the more closely the decomposition approximates the complete LU decomposition. However, the notation $ILU(n)$ refers to a specific set of locations for fill-in; other incomplete LU decomposition options exist for fill-in.

Another ILU idea is to allow fill-in only within the original matrix bands. That is, to allow fill-in only in the tri or pentadiagonal banded structure even if the original matrix entry was zero. This becomes only a slight modification of the original $ILU(0)$ process but results in a considerably different approximation. It follows naturally if all possible matrix contributions are input, including zeros within the predominately nonzero bands, and if the matrix storage procedure does not eliminate zero entries. This new preconditioner will be referred to as $ILUB(0)$.

It should be noted that a dilemma surrounds the use of preconditioned CGS schemes. That is, the better a preconditioner is, the less work the CGS algorithm needs to do. Hence CGS iterations might not be needed if a "good enough" preconditioner is used. The following sections will attempt to explain the tradeoffs inherent in these schemes by discussing the numerical results obtained by applying the options presented above.

Results

This section describes what occurred when the previously discussed ideas were applied to Orkwis and McRae's Newton method solver. It begins by explaining the efficiency improvements made by eliminating some of the RSLIB sort and check routines from the inversion routine. The application of the local timestepping idea is then presented. Experience obtained while testing the CGS scheme with the $ILUB(0)$ preconditioner is then explained. Finally the results of numerically testing four different versions of the Newton scheme are presented and discussed.

To alleviate the problem of memory size limitations for Newton's method solvers Orkwis and McRae [1, 2] chose the low in-core memory Boeing Real Sparse Library (RSLIB) routines for the inversion of the Jacobian matrices. However, these routines are inefficient in some respects because they allow the matrix to be input in any order. They also allow contributions to a single matrix location to be made by two or more separate entries. Expensive sorting and checking routines are then employed to put the matrix into the necessary form for sparse storage. The first task of this research was to simplify the matrix storage process.

After extensive investigation it was determined that the RSLIB routines are a front-end to the more basic LU decomposition code of Forrestor Johnson. The RSLIB routines take whatever form of the matrix is given and rearrange it into sequential row oriented storage buffers. The process adds entries with like indices and then sorts the resulting terms. The matrix is transformed into a series of vector buffers and is written to tape. The buffers have the form

Entry	Description
1	# of rows in buffer
2	# of entries in row 1
3	diagonal element location
4	1
5	column locations for the matrix entries of row 1
⋮	⋮
NE_1+5	matrix entries row 1
⋮	⋮
$2NE_1$+5	# of entries in row 2
⋮	⋮
NBUF-3	$2NE_1$+5 + $2NE_2$+3
NBUF-2	$2NE_1$+5
NBUF-1	2
NBUF	matrix location of row 1

where NE_i is the number of matrix entries in row i and NBUF is the size of the buffer.

The point of the initial research was to code the solver to automatically write the Jacobian matrix buffers in the above form without using the RSLIB routines. This proved to be extremely successful since the automatic formation of the Jacobian

matrix buffers resulted in a 60% reduction in required CPU time as compared with using the full set of RSLIB routines.

The next task was to test the local timestep modification idea. The original method used equation 4 and varied Δt as per equation 5. In this way, as the norm of the residual was reduced Δt would increase and the diagonal term addition would decrease. The idea is to use the added diagonal term to temper the size of the early Newton iterates and then to remove this term from later iterates. This is done because the scheme tends to overpredict the early conserved variable changes since the solution is still far from the final result. However, once the residual decreases, the diagonal term disappears and the scheme returns to a quadratic convergence Newton scheme.

It was felt that using a local timestep for each equation would result in better convergence properties since the method would then be governed by what occurs locally. However this idea proved to be somewhat cumbersome as additional "if statement" checks were required at each point because of zero residuals in the freestream areas. For example, in viscous flat plate problems a shock wave is formed at the plate which eventually becomes the bow shock wave. When this initial shock moves into the freestream region the local residual increases from zero to some number. Additional logic is required to produce a finite timestep from equation 5, since the initial timestep is zero. The local timestep also tends to oscillate greatly from iteration to iteration because the local residual is not as smooth as the global residual. Therefore, using a global timestep is a much better approach than using a local timestep.

Next an evaluation of the conjugate gradient squared (CGS) scheme for inverting $A\mathbf{x} = \mathbf{b}$ was performed. The first question was whether or not the CGS scheme could match the accuracy of the full LU decomposition. In this regard CGS was exemplary. Convergence of the subiterate (i.e. inversion of the $A\mathbf{x} = \mathbf{b}$ system) to machine zero produced virtually the same answer as the exact LU decomposition. Tests of the scheme were performed by calculating the supersonic flat plate test case from Orkwis and McRae's [1] earlier work on a 40x40 stretched viscous grid. This case uses slug flow initial conditions with $M = 2$ and $R_e = 1.65 \cdot 10^{+6}$. However, the CGS scheme needed more than 5 times as much CPU time as the exact LU decomposition to achieve the required accuracy.

It should also be noted that the CGS scheme used here was preconditioned by $ILUB(0)$ not the standard $ILU(0)$ approach because no success was attained with the latter approach. This is in contrast to Venkatakrishnan [13] who reported success with $ILU(0)$ as a preconditioner for the GMRES algorithm in his transonic airfoil calculations. Possible explanations of this discrepancy are the different flow field conditions and the different conjugate gradient approaches.

Three initial guesses for the $A\mathbf{x} = \mathbf{b}$ system were examined. They were; $\mathbf{x}_o = \emptyset$, $\mathbf{x}_o = \mathbf{x}^n$ (from the previous global iterate), and $\mathbf{x}_o$ from the approximate system $\tilde{L}\tilde{U}\mathbf{x}_o = \mathbf{b}$ ($\tilde{L}$,$\tilde{U}$ from $ILUB(0)$). The results indicated that the last worked best in terms of total number of subiterations and total CPU time required for convergence. This can be attributed to the fact that $ILUB(0)$ forms a fairly good approximate system for $A\mathbf{x} = \mathbf{b}$.

Two major approximations were then tested. Both involved changes to how the subiterates were to be halted. The first idea was to stop after a prescribed number of subiterations. This was effective provided that the number of subiterates was large. Global convergence was attained in 95.2% of the exact LU decomposition time when the maximum number of subiterates was set to 50, and in 65.2% when set to 25. However, for maximum subiteration levels less than 25 the scheme always diverged at some subiterate. It was observed that the number of subiterations varied widely from step to step. Hence when the maximum subiterate count was below 25 there would invariably be one subiterate that would not converge in the allotted number of steps, which would cause the global system to diverge.

To alleviate this problem it was decided to halt convergence at a given level of the subiterate residual rather than after a given subiteration. The drawback to this idea is that the only guarantee on CGS iteration time is that the number of subiterates required for convergence will be less than or equal to the number of equations in the system. This gives a CPU time upperbound that is extremely conservative. Tests of CPU time versus subiterate tolerance level showed that CPU time significantly decreases with increases in the tolerance, until the tolerance reaches 10^{-3}. Convergence to a global residual of 10^{-6} with subiterate residuals of 10^{-12}, 10^{-6}, 10^{-4}, and 10^{-3} required respectively 44, 45, 50 and 58 global iterations, and 533.0%, 66.2%, 47.9%, and 39.1% of the exact LU decomposition CPU time. However the global convergence rate of the system becomes more and more linear as the tolerance of the CGS residual decreased. Subiterate residual reduction could require nearly as many

global iterations to converge from 10^{-6} to 10^{-12} as it would initially to converge to 10^{-6}. This should be contrasted with the global quadratic convergence obtained from the CGS scheme with 10^{-12} subiterate tolerance, which produced the 10^{-12} global residual result in as little as 3 additional iterations after the 10^{-6} global residual had been reached.

The final approximations to the Newton scheme involved using lower order Jacobian matrices. In other words, Jacobians formed by using derivatives of lower order discretizations. Convergence histories drawn down to 10^{-12} global residual were made using both first and second order Jacobians with both exact LU and CGS/$ILUB(0)$ 10^{-3} subiterate tolerance inversion routines. Results are summarized in table 1 for flat plate and flat plate/15° wedge test cases with the same initial and boundary conditions as the earlier flat plate test. The flat plate cases used a $40x40$ grid while the wedge cases used an $80x40$ grid (Note that the $40x40$ grid is different than the one used earlier).

The results of the study indicate that the approximate CGS results were by far superior for problems with weak shock waves. However, the presence of large discontinuities drastically altered the results. Observations of the convergence properties of these schemes showed that the second order approximate Jacobian was least stable, followed by the first order approximate, the first order exact, and finally the second order exact. The results also show that for problems with strong shock waves, the complete Newton's method Jacobian with the exact LU inversion is needed. Therefore the type of problem one solves is a determining factor in the choice of Newton

	Flat Plate			Wedge		
Method	CPU Time (seconds)	Iterations	Speed-up	CPU Time (seconds)	Iterations	Speed-up
2nd/Exact	750.489	46	1	1051.753	23	1
1st/Exact	559.623	79	1.341	7413.960	372	0.14
2nd/Approx	348.377	74	2.154	Did Not Converge		
1st/Approx	275.642	85	2.723	Did Not Converge		

Table 1. Comparison of Convergence Times and Speed-ups for Several Newton and Approximate Newton Schemes Applied to Flat Plate and Wedge Problems.

or Newton-like scheme.

Conclusions

The following conclusions have resulted from the efforts to improve the efficiency of Orkwis and McRae's Newton's method solver.

- Removal of the Boeing RSLIB front-end to Forrestor Johnson's LU decomposition routines resulted in a 60% savings in CPU time.

- Local timestepping is not an adequate substitute for global timestepping in Newton's method solvers.

- Quadratic convergence can be obtained with Newton's method solvers that employ CGS schemes for the solution of $A\mathbf{x} = \mathbf{b}$.

- Machine zero convergence of CGS schemes with $ILUB(0)$ preconditioning can take considerably longer than an exact LU decomposition.

- $ILUB(0)$ is an easily coded alternative to $ILU(0)$ that performed better for the cases tested and required little additional effort.

- Approximate CGS schemes performed better than exact matrix inversion routines for problems with weak shock waves.

- Approximate Jacobian schemes performed better than exact Jacobian schemes for problems with weak shock waves.

- Exact *LU* decomposition of an exact Jacobian matrix is needed for problems with strong shock waves.

Acknowledgements

The author wishes to thank Drs. Dave Belk and Jack George, and Mr. Kirk Vanden for stimulating discussions on Newton method and CGS strategies. Support for this research was provided by the U.S. Air Force Office of Scientific Research Summer Faculty Research Program. Computational support was provided by resources at Eglin AFB, Florida.

References

[1] Orkwis, P.D. and McRae, D.S., "A Newton's Method Solver for the Navier-Stokes Equations," AIAA 90-1524, Seattle, Washington, June 1990, see also, "A Newton's Method Solver for High-Speed Viscous Separated Flow Fields," to appear in *AIAA Journal*, November 1991 .

[2] Orkwis, P.D. and McRae, D.S., "A Newton's Method Solver for the Axisymmetric Navier-Stokes Equations," AIAA Paper 91-1554, Honolulu, Hawaii, June 1991, also to appear in *AIAA Journal*.

[3] Liou, M-S. and Van Leer, B., "Choice of Implicit and Explicit Operators for the Upwind Differencing Method," AIAA Paper 88-0624, Reno, Nevada, January 1988.

[4] George, A., "Nested Dissection of a Regular Finite Element Mesh", *SIAM Journal of Numerical Analysis*, Volume 10, Number 2, April 1973.

[5] Golub, G.H. and Van Loan, C.F., Matrix Computations 2nd Edition, The Johns Hopkins University Press, Baltimore, Maryland, 1989, pg 506-509.

[6] Saad, Y., "Krylov Subspace Methods for Solving Large Unsymmetric Linear Systems," *Mathematics of Computation*, Volume 37, Number 155, July 1981, pg 105-126.

[7] Saad, Y. and Schultz, M.H., "GMRES: A Generalized Minimum Residual Algorithm for Solving Nonsymmetric Linear Systems," *SIAM Journal on Scientific and Statistical Computing*," Volume 7, 1986, pg 856-853.

[8] Hageman, L.A. and Young, D.M., Computer Science and Applied Mathematics, Applied Iterative Methods, Academic Press, New York, 1981, pg 138-161.

[9] Sonneveld, P., "CGS, A Fast Lanczos-Type Solver for Nonlinear Systems," *SIAM Journal on Scientific and Statistical Computing*, Volume 10, 1989, pg 350-356.

[10] Wigton, Laurence B., "Application of MACSYMA and Sparse Matrix Technology to Multielement Airfoil Calculations", AIAA Paper 87-1142, Honolulu, Hawaii, June 1987.

[11] Venkatakrishnan, V., "Newton Solution of Inviscid and Viscous Problems", AIAA Paper 88-0413, Reno, Nevada, January 1988, see also, "Newton Solu-

tion of Inviscid and Viscous Problems", *AIAA Journal*, Vol. 27, No. 7, July 1989, pp 885-891.

[12] Venkatakrishnan, V., "Preconditioned Conjugate Gradient Methods for the Compressible Navier-Stokes Equations", AIAA Paper 90-0586, Reno, Nevada, January 1990, see also, "Preconditioned Conjugate Gradient Methods for the Compressible Navier-Stokes Equations", *AIAA Journal*, Volume 29, Number 7, July 1991.

[13] Venkatakrishnan, V., "Implicit Solvers for Unstructured Meshes," AIAA Paper 91-1537, Honolulu, Hawaii, June 1991.

[14] Ajmani, K., Ng, W. and Liou, M-S., "Generalized Conjugate-Gradient Methods for the Navier-Stokes Equations," AIAA Paper 91-1556, Honolulu, Hawaii, June 1991.

AFOSR SUMMER RESEARCH PROGRAM—FINAL REPORT
ANALYSIS AND DESIGN OF NONLINEAR MISSILE AUTOPILOTS

Jeff S. Shamma
Department of Electrical Engineering
University of Minnesota
Minneapolis, MN 55455

September 23, 1991

Abstract

This report discusses the following two problems in missile autopilot design: (1) nonlinear stability analysis in the presence of dynamic modelling errors and (2) nonlinear autopilot design via gain-scheduling. Traditionally, these problems are addressed by linearizing the missile dynamics and applying linear theory. In this report, we present an alternate approach which does not resort to linearizations. First, we derive necessary and sufficient conditions for the robust stability of nonlinear dynamics subject to unstructured nonlinear dynamics uncertainty. Second, we present a novel approach to gain-scheduled autopilot design which relies on an alternate representation of the missile dynamics rather than linearizations.

1 Introduction

The analysis and design of missile autopilots typically involve a linearization of nonlinear dynamics and an application of linear techniques. Since both the missile dynamics and autopilot dynamics are nonlinear, it is worthwhile to develop methods for analysis and design of nonlinear missile autopilots without recourse to linearizations and the associated limitations. This report presents results in two problems regarding nonlinear missile autopilot design: (1) nonlinear stability analysis in the presence of dynamic modelling errors and (2) nonlinear autopilot design via gain-scheduling. With both problems, the objective is to develop techniques which address nonlinearities directly.

The first problem considered is robust stability of nonlinear systems, more precisely, the derivation of conditions under which stability is maintained in the presence of certain classes of model perturbations. A common paradigm in robust stability is that of unstructured dynamic uncertainy. In this framework, the deviation from the nominal model is characterized by a perturbation representing an unknown but bounded dynamical system. One example is the so called "additive plant uncertainty" family given by $\mathcal{P}_{\text{add}} = \{P : P = P_o + \Delta W\}$. This family of systems may be given the following interpretation. The nominal system is represented by P_o. The "unstructured uncertainty", Δ, is a dynamical system which is known only to be stable and satisfy a given norm bound. Finally the weighting, W, is a known stable dynamical system which "shapes and normalizes" the effects of Δ. The question of robust stabilization then takes the form "under what conditions does a compensator which stabilizes P_o also stabilize every P in $\mathcal{P}_{\text{add}}$?"

The robust stabilization of families of plants characterized by unstructured uncertainty typically reduces to the stable invertibility of a certain operator for all admissible perturbations (cf., [4] and references contained therein). Once reduced to this problem of stable invertibility, the small-gain theorem [16] then may be used to give a sufficient condition for invertibility, and hence for robust stability. We show that for a broad class of systems—namely, systems with fading memory—the small gain theorem is also *necessary* for robust stability.

The second problem considered is nonlinear missile autopilot design. Here, we present a gain-scheduled design for a missile longitudinal autopilot which is novel in that it *does not* involve linearizations about trim conditions of the missile dynamics. Rather, the missile dynamics are brought to a quasi-linear parameter varying (LPV) form via a state transformation. An LPV system is defined as a linear system whose dynamics depend on an exogenous variable whose values are unknown *a priori* but can be measured upon system operation. In this case, the variable is the

angle-of-attack. This is actually an endogenous variable, hence the expression "quasi-LPV". Once in a quasi-LPV form, a robust controller using μ-synthesis [1] is designed to achieve angle-of-attack control via fin deflections.

The angle-of-attack control constitutes the inner feedback loop. The effect of the inner feedback loop is to effectively linearize the dynamics from angle-of-attack commands to angle-of-attack. However, it is the *linearizing effect* of a robust servo-loop which makes the angle-of-attack dynamics linear—rather than a geometric condition on the state dynamics as in geometric feedback linearization.

The regulated variable of interest is actually normal acceleration. Thus, an outer loop robust controller using μ-synthesis is designed to generate angle-of-attack commands to achieve the desired normal acceleration. The outer loop is designed to generate angle-of-attack commands within the bandwidth of the inner loop, thereby assuring the essentially linear behavior of the inner loop, and hence the overall performance of the autopilot.

The remainder of this report is organized as follows. Section 2 discusses the problem of nonlinear robust stability analysis. Section 3 presents the missile autopilot design. Finally, concluding remarks are given in Section 4.

This report summarizes the research performed during the authors AFOSR Summer Research Fellowship, 1991. These results have been documented in [14,15,12,13].

2 Robust Stability of Nonlinear Systems

In the design of a missile autopilot, one typically makes various simplifying assumptions and approximations of the missile dynamics. A controller is then designed on the basis of the simplified model. A natural question is then that of robust stability; i.e., under what conditions is stability maintained in the presence of modelling errors?

First, we establish the following notation (e.g., [2]). For a time-signal, g, we define $\|g\|$ as

$$\|g\| \stackrel{\text{def}}{=} \left(\int_0^\infty g^T(t) g(t) dt \right)^{1/2}.$$

For a stable dynamical system, H, we define $\|H\|$ as

$$\|H\| \stackrel{\text{def}}{=} \sup_{\substack{g \neq 0 \\ \|g\| < \infty}} \frac{\|Hg\|}{\|g\|}.$$

In words, $\|H\|$ represents the maximum amplification of energy achievable by the system H.

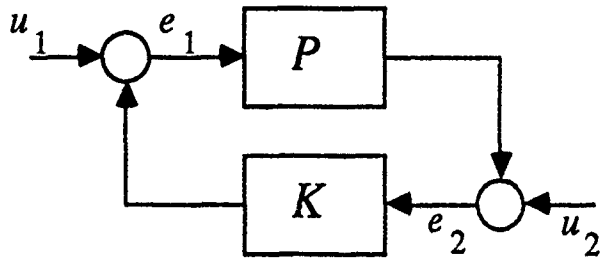


Figure 1: Block Diagram for Robust Stabilization

In this section, we consider the problem of robust stability in the context of Figure 1. In this figure, the block P represents the plant dynamics, and the block K represents the compensator. Now define the following *family* of plants:

$$\mathcal{P}_{\text{add}} \stackrel{\text{def}}{=} \{P : P = P_o + \Delta W\}$$

where Δ is a nonlinear system with $\|\Delta\| < 1$, and W, W^{-1} are stable nonlinear systems. The problem of robust stabilization is under what conditions does a compensator, K, which stabilizes P_o also stabilize every $P \in \mathcal{P}_{\text{add}}$ with

$$\sup_{P \in \mathcal{P}_{\text{add}}} \|\Phi(P, K)\| < \infty,$$

where $\Phi(P, K)$ represents the closed-loop dynamics from (u_1, u_2) to (e_1, e_2). In this case, the compensator, K, is said to *uniformly robustly stabilize* the family $\mathcal{P}_{\text{add}}$.

The main result is the following. See [14,15] for further technical assumptions and discussion.

Theorem 2.1 *Consider the plant family $\mathcal{P}_{\text{add}}$. Let the compensator, K, pointwise fading memory incrementally stabilize P_o. Then K uniformly robustly stabilizes the family $\mathcal{P}_{\text{add}}$ if and only if*

$$\left\| WK(I - P_o K)^{-1} \right\| < 1.$$

The stated condition for robust stability is well known to be sufficient. The main contribution of Theorem 2.1 is that it shows this condition is also *necessary* for robust stability with the added assumption that the closed-loop dynamics have *fading memory*. Informally, the fading memory property means that the current output is primarily a function of recent inputs and not the remote past.

The question of robust stability and fading memory are pursued further in [14,15]. Since the results are quite technical, they are briefly summarized here. In these papers, the authors define two notions of fading memory: uniform and pointwise. They then provide conditions under

which linear or nonlinear systems exhibit uniform or pointwise fading memory. In particular, they show that (1) all discrete-time linear time-varying (LTV) systems have uniform fading-memory, (2) continuous-time LTV systems need not have uniform fading memory but always have pointwise fading-memory, and (3) finite-dimensional continuous-time LTV systems have uniform fading-memory. They then show that a version of the small-gain theorem which employs the asymptotic gain of a fading-memory system is necessary and sufficient for the stable invertibility of certain feedback operators. These results are presented for both continuous-time and discrete-time systems using general ℓ^p or $\mathcal{L}^p$ notions of input/output stability and generalize existing results for ℓ^2 stability. They further investigate fading memory in a closed-loop context. For linear plants, they parameterize all nonlinear controllers which lead to closed-loop pointwise fading memory. As a corollary, they provide an alternate derivation of the result that nonlinear compensation offers no advantage for robust stabilization of linear systems subject to unstructured nonlinear perturbations.

3 Nonlinear Missile Autopilot Design

Future tactical missiles will be required to operate over an expanded flight envelope in order to meet the challenge of highly maneuverable tactical aircraft. In such a scenario, an autopilot derived from linearization about a single flight condition will be unable to achieve suitable performance over all envisioned operating conditions. A particularly challenge, however, is that of the missile endgame. This involves the final few seconds before delivery of ordnance. During this phase, a missile autopilot can expect large and rapidly time-varying acceleration commands from the guidance law. In turn, the missile is operating at a high and rapidly changing angle-of-attack.

Traditionally, satisfactory performance across the flight envelope can be attained by gain scheduling local autopilot controllers to yield a global controller. Often the angle-of-attack is used as a scheduling variable. However during the rapid transitions in the missile endgame, a fundmental guideline of gain-scheduling to "schedule on a slow variable" is violated. Given the existing track record of gain-scheduling, any improvement in the gain-scheduling design procedure—especially in the endgame—could have an important impact on future missile autopilot designs.

In this section, we present a novel approach to gain-scheduled missile autopilot design. The missile control problem under consideration is normal acceleration control of the longitudinal dynamics during the missile endgame. In standard gain-scheduling, the design plants consist of a collection of linearizations about equilibrium conditions indexed by the scheduling variable, in this case the angle-of-attack, α (cf., [8,10]). In the present approach, the design plants also consist of a family of linear plants indexed by the angle-of-attack. A key difference between the present approach

and standard gain-scheduling is that this family is *not* the result of linearizations. Rather, it is derived via a state-transformation of the original missile dynamics (i.e., an alternate selection of state variables). Since no linearization is involved, the approach is not limited by the local nature of standard gain-scheduled designs.

Since gain-scheduling generally encounters families of linear plants indexed by a scheduling variable, we shall refer to such a family as a Linear Parameter Varying (LPV) plant. LPV plants *differ* from linear time-varying plants in that the time-variations (i.e., the scheduling variable) is unknown *a priori* but may be measured/estimated upon operation of the feedback system. We shall call such a family quasi-LPV in case the scheduling variable is actually endogenous to the state dynamics (as in the missile problem). In [8,10], it was shown that LPV and quasi-LPV dynamics form the underlying structure of gain-scheduled designs.

The design for the resulting quasi-LPV system is performed via μ-synthesis [1]. Briefly, μ-synthesis exploits the structure of performance requirements and robustness considerations in order to achieve robust performance in a non-conservative manner. Thus, the present approach makes use of gain-scheduling's ability to incorporate modern linear synthesis techniques into a nonlinear design.

Another feature in the present approach is its interpretation of an inner/outer loop approach to nonlinear control design. In standard gain-scheduling (as well as geometric nonlinear control [5]), one often applies an inner-loop feedback. In gain-scheduling, this feedback is an update of the current trim condition. In geometric nonlinear control, this feedback serves to invert certain system dynamics to yield linear behavior in the modified plant. In either case, unless the inner-loop robustly performs its task, the outer-loop performance and even stability can be destroyed. In other words, any inner/outer loop approach must be built from the inside out. Reference [8] presents a more detailed discussion of this possibility in the context of standard gain-scheduling.

The present approach also takes an inner/outer loop approach to the autopilot design. The inner-loop consists of a robust angle-of-attack servo. The reason for the inner loop is that nonlinear gain-scheduling techniques prefer to directly control the scheduling variable. The actual regulated variable of interest is the normal acceleration. Thus, the outer-loop serves to generate angle-of-attack commands, α_c, to obtain the desired normal acceleration. A *consequence* of the inner-loop design is that the dynamics from the commanded angle-of-attack, α_c, to the angle-of-attack, α, exhibits a linear behavior *within the bandwidth* of the inner-loop design. Thus, as in geometric nonlinear control, the inner-loop linearizes certain dynamics. However, the approximate linearization stems from the natural *linearizing effect of feedback* (cf., [3]) as opposed to an exact linear geometric

condition on the plant state dynamics. Thus, the outer-loop design is essentially a linear design. However, in the design process, it is acknowledged that the linear behavior due to the inner-loop is *approximate* and *bandlimited.*

3.1 Missile Dynamics

The missile dynamics considered here are taken from [6]. These dynamics are representative of a missile travelling at Mach 3 at an altitude of 20,000 feet. However, they do not correspond to any particular missile airframe.

The nonlinear dynamics are as follows:

$$\dot{\alpha} = f\frac{g\cos(\alpha/f)}{WV}Z + q,$$

$$\dot{q} = fm/I_{yy},$$

where

α = angle of attack (deg)

q = pitch rate (deg/s)

g = acceleration of gravity (32.2 ft/sec^2)

W = weight (450 lbs)

V = speed (3109.3 ft/sec)

I_{yy} = pitch moment of inertia (182.5 slug-ft^2)

f = radians-to-degrees conversion ($180/\pi$)

$Z = C_ZQS$ = normal force (lbs)

$m = C_mQSd$ = pitch moment (ft-lb)

Q = dynamic pressure (6132.8 lb-ft^2)

S = reference area (0.44 ft^2)

d = reference diameter (0.75 ft)

The normal force and pitch moment aerodynamic coefficients are approximated by

$$C_Z = \phi_Z(\alpha) + b_Z\delta,$$

$$C_m = \phi_m(\alpha) + b_m\delta,$$

where

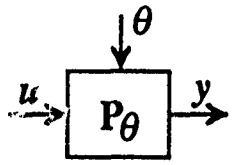


Figure 2: A Linear Parameter Varying (LPV) Systems

δ = fin deflection (deg)

$\phi_Z(\alpha) = 0.000103\alpha^3 - 0.00945\alpha|\alpha| - 0.170\alpha$

$b_Z = -0.034$

$\phi_m(\alpha) = 0.000215\alpha^3 - 0.0195\alpha|\alpha| + 0.051\alpha$

$b_m = -0.206$

These approximations are accurate for α in the range of ± 20 degrees.

Finally, the missile tailfin actuator is modeled as the second order system with transfer function

$$\delta(s) = \frac{\omega_a^2}{s^2 + 1.4\omega_a s + \omega_a^2}\delta_c(s),$$

where

δ_c = commanded fin deflection (deg)

ω_a = actuator bandwidth (150 rad/s)

The autopilot will be required to control the normal acceleration (expressed in g's),

$$\eta_Z = Z/W,$$

via commanded fin deflections, δ_c. The general performance objective is to track acceleration step commands with a steady state accuracy of less than 0.5% and a time constant of 0.2 seconds. Of course, the controller is bandlimited by flexible mode dynamics and actuator/sensor nonlinearities (e.g., rate saturations) [6].

3.2 LPV Systems

An LPV system [8,9,10] is defined as a linear system whose coefficients depend on an exogenous time-varying parameter. Let $y = P_\theta u$ be an LPV system as in Figure 2. A possible realization for P_θ is

$$\dot{x} = A(\theta)x + B(\theta)u,$$

$$y = C(\theta)x.$$

The exogenous parameter, θ, is unknown *a priori*, however can be measured/estimated upon operation of the system. The reason for the special nomenclature is to distinguish LPV systems from linear time-varying systems for which the time-variations are known beforehand (as in periodic systems). Typical *a priori* assumptions on θ are bounds on its magnitude and rate-of-change.

A gain-scheduled approach to controling an LPV system is to design a collection of controllers based on frozen parameter values. This leads to an LPV controller, K_θ. It was shown in [8,10] that this approach has guaranteed robustness and performance properties provided that the parameter time-variations are "sufficiently slow." Quantitative statements qualifying "sufficiently slow" are provided in [8,10]. Of course, sufficiently slow is with regards to the closed-loop system dynamics. Work on modifying gain-scheduling to accomodate arbitrarily fast parameter time-variations is in progress (e.g., [7,11]).

In [8,9,10], it was shown that LPV systems provide the underlying framework for nonlinear gain-scheduled systems. To see this relationship, consider the nonlinear square plant

$$\frac{d}{dt}\begin{pmatrix} z \\ w \end{pmatrix} = f(z) + A(z)\begin{pmatrix} z \\ w \end{pmatrix} + B(z)u,$$

where u is the control input and z is the controlled output. For this system, the nonlinearities depend only on the controlled output. Such systems are subsequently called "output-nonlinear" systems. Note that the missile dynamics are output-nonlinear with the angle-of-attack, α, as the controlled output as follows:

$$\frac{d}{dt}\begin{pmatrix} \alpha \\ q \end{pmatrix} = \begin{pmatrix} \frac{fgQS\cos(\alpha/f)}{WV}\phi_Z(\alpha) \\ \frac{fQSd}{I_{yy}}\phi_m(\alpha) \end{pmatrix} + \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} \alpha \\ q \end{pmatrix} + \begin{pmatrix} \frac{fgQSb_Z\cos(\alpha/f)}{WV} \\ \frac{fQSdb_m}{I_{yy}} \end{pmatrix}\delta.$$

Section 3.3.2 discusses how to accomodate normal acceleration, η_Z, as the controlled output even though the dynamics are output-nonlinear in α.

We assume that there exist continuously differentible functions, $w_{eq}(z)$ and $u_{eq}(z)$, such that

$$0 = f(z) + A(z)\begin{pmatrix} z \\ w_{eq}(z) \end{pmatrix} + B(z)u_{eq}(z).$$

In other words, we have a *family* of equilibrium states parameterized by the controlled output z. For the missile problem, we have

$$\delta_{eq}(\alpha) = -\phi_m(\alpha)/b_m,$$

$$q_{eq}(\alpha) = -\frac{fgQS\cos(\alpha/f)}{WV}\left(\phi_Z(\alpha) - \frac{b_Z}{b_m}\phi_m(\alpha)\right).$$

Let $A(z)$ and $B(z)$ be partitioned

$$A(z) = \begin{pmatrix} A_{11}(z) & A_{12}(z) \\ A_{21}(z) & A_{22}(z) \end{pmatrix}, \quad B(z) = \begin{pmatrix} B_1(z) \\ B_2(z) \end{pmatrix},$$

to conform with $(z \quad w)^T$. Then it is easy to show that the state dynamics may be written as

$$\frac{d}{dt}\begin{pmatrix} z \\ w - w_{eq}(z) \end{pmatrix} = \begin{pmatrix} 0 & A_{12}(z) \\ 0 & A_{22}(z) - Dw_{eq}(z)A_{12}(z) \end{pmatrix}\begin{pmatrix} z \\ w - w_{eq}(z) \end{pmatrix} + \begin{pmatrix} B_1(z) \\ B_2(z) - Dw_{eq}(z)B_1(z) \end{pmatrix}(u - u_{eq}(z)).$$

Thus, we have *transformed* the original dyanamics into a quasi-LPV form, with the variable z as the "exogenous" parameter. In case all nonlinearities are not contained in the output, the above transformation will be approximate up to first order terms in $w - w_{eq}(z)$ [8]. It is interesting to note that this quasi-LPV family *is not* the same family we would obtain by performing linearizations about equilibrium conditions.

Now although we may use the above quasi-LPV plant as the design plant, a possible drawback is the inner-loop feedback term $u_{eq}(z)$. More precisely, if one were to design a controller for the above quasi-LPV plant, the actual applied control signal would be

$$u = u_{eq}(z) + \tilde{u},$$

where $\tilde{u}$ is the controller output. Even though the outer-loop may have guaranteed robustness properties, the inner-loop feedback $u_{eq}(z)$ can destroy these properties by adversly exciting flexible mode dynamics [8,9]. For example in the missile problem, we have $\delta_{eq}(\alpha) = -\phi_m(\alpha)/b_m$. Thus fast angle-of-attack variations (as in the end game) could excite neglected flexible mode dynamics.

This problem can be avoided by augmenting integrators at the plant input. Let

$$u = \int v.$$

Then the system dynamics take the form

$$\frac{d}{dt}\begin{pmatrix} z \\ w - w_{eq}(z) \\ u - u_{eq}(z) \end{pmatrix} = \begin{pmatrix} 0 & A_{12}(z) & B_1(z) \\ 0 & A_{22}(z) - Dw_{eq}(z)A_{12}(z) & B_2(z) - Dw_{eq}(z)B_1(z) \\ 0 & -Du_{eq}(z)A_{12}(z) & -Du_{eq}(z)B_1(z) \end{pmatrix}\begin{pmatrix} z \\ w - w_{eq}(z) \\ u - u_{eq}(z) \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} v.$$

Now if we design a controller *without* using the state $u - u_{eq}(z)$ for feedback, no inner-loop feedback of a trim condition is applied. Thus, any robustness properties of the quasi-LPV design remain in tact.

The forthcoming missile design uses the above representation of the missile dynamics for the autopilot design. This representation is the result of a *state transformation* only. That is, no approximation/linearization of the original dynamics has occurred. As mentioned earlier, in case

the system dynamics are nonlinear in w as well, then the above representation is accurate up to first order in $w - w_{eq}(z)$. Systems for which higher-order-terms in $w - w_{eq}(z)$ are large are not well suited for gain-scheduling in the first place. In other words, gain-scheduling seeks to exploit predominantly output-nonlinear dynamics.

A gain-scheduled approach to control design for quasi-LPV systems resembles that for LPV systems. Namely, a series of designs are performed for frozen z values of the state-space matrices. This leads to a quasi-LPV controller with z as the external parameter.

3.3 Missile Auto Pilot Design

In this section, we discuss the missile autopilot design. The objective is to control normal acceleration, η, via commanded fin deflections, δ_c. We assume that the angle-of-attack, α, the pitch rate, q, and the normal acceleration, η, are available for feedback. The pitch rate and normal acceleration measurements are obtained via rate gyros and accelerometers, respectively [6]. However, in practice the angle-of-attack, α, must be estimated.

As mentioned in Section 3.2, gain-scheduling seeks to control the output variable of an output-nonlinear system. The longitudinal missile dynamics are output-nonlinear with the angle-of-attack, α, as the output variable. Since normal acceleration, η_Z, is the regulated variable of interest, the gain-scheduled design is modified as follows. First, we design a controller for angle-of-attack. This constitutes the inner-feedback loop. We then design an outer feedback loop to generate angle-of-attack commands to achieve desired normal accelerations.

3.3.1 Inner-Loop Angle-of-Attack Control

The inner feedback loop consists of angle-of-attack control. First, the missile dynamics are augmented with an integrator and transformed to a quasi-LPV form as in Section 3.2. The actuator dynamics are then augmented onto the quasi-LPV missile dynamics. This process leads to a design plant, P_{des}, whose states are $(\alpha \quad q - q_{eq}(\alpha) \quad \delta - \delta_{eq}(\alpha) \quad x_a \quad \dot{x}_a)^T$, where x_a and $\dot{x}_a$ are the actuator state variables.

Figure 3 shows the block diagram used for the μ-synthesis design. See [1] for a complete discussion of μ-synthesis. The block P_{des} denotes the quasi-LPV design plant. The input, u, to P_{des} is actually the time-derivative of the commanded fin deflection. That is,

$$\delta_c = \int u.$$

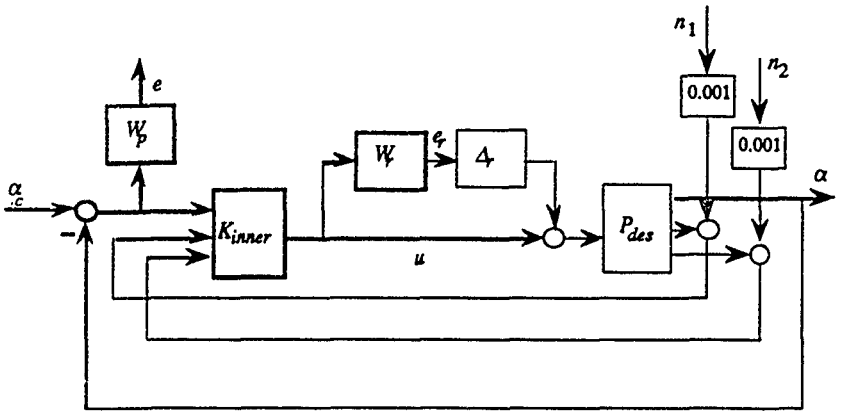


Figure 3: Inner-Loop Angle-of-Attack Control

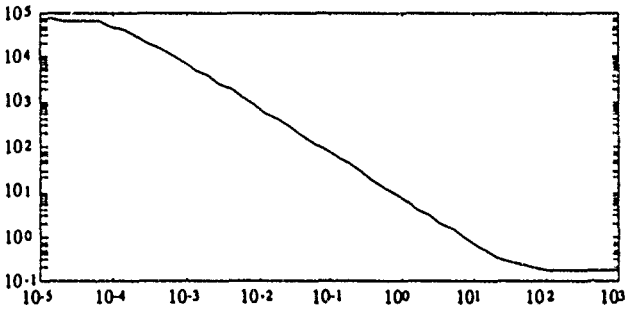


Figure 4: Performance Weight $W_p(j\omega)$ for Angle-of-Attack Control

The measurements from P_{des} are the angle-of-attack, α, the pitch rate trim deviation, $q - q_{eq}(\alpha)$, and the normal acceleration trim deviation, $\eta - \eta z_{,eq}(\alpha)$. Note that

$$\eta z - \eta z_{,eq}(\alpha) = \frac{QSb_Z}{W}(\delta - \delta_{eq}(\alpha)).$$

In the actual implementation, the values of $q - q_{eq}(\alpha)$ and $\eta - \eta_{eq}(\alpha)$ would be constructed from α, q, and η measurements.

The robustness and performance objectives are described as follows. The block Δ_r represents linear time-varying multiplicative perturbation weighted by W_r. This uncertainty reflects actuator phase/gain uncertainty and flexible mode dynamics. Figure 5 shows the frequency response of W_r, where

$$W_r(s) = 2\frac{(s+100)(s+200)}{(s+1000)(s+2000)}.$$

The performance objective is to keep $\|\alpha_c \mapsto e\| \leq 1$. The performance weight,

$$W_p(s) = \frac{7(s+40)}{40(s+0.0001)},$$

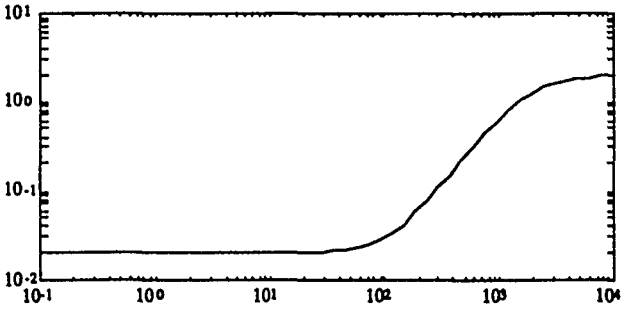


Figure 5: Robustness Weight $W_r(j\omega)$

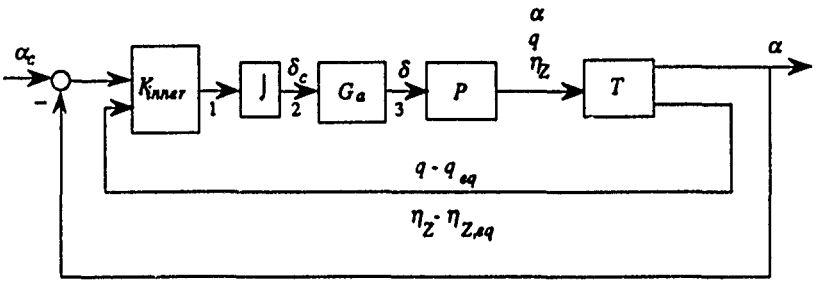


Figure 6: Implementation of K_{inner}

has a frequency response shown in Figure 4. Finally, the signals n_1 and n_2 are small noises injected to satisfy certain rank conditions in the μ-synthesis design.

A μ-synthesis design procedure was performed with this interconnection structure at the set point $\alpha = 0$. That is, the α-dependent coefficient matrices of the quasi-LPV plant P_{des} were evaluated at $\alpha = 0$ for the design. The first pass led to a frozen α robust performance level of 1.09. After six iterations, this value was reduced to 0.5232.

Now a gain-scheduled design procedure would typically involve repeating the fixed-α designs for several α set points. However, it turns out the $\alpha = 0$ controller delivered robust performance for all α in the range ± 20 degrees. Thus, *no controller gain-scheduling* was required. For this particular airframe, the missile dynamics at $\alpha = 0$ are statically unstable and become stable at higher values of α. Thus, the dynamics at $\alpha = 0$ are, in some sense, the most difficult to control. Note that even though K_{inner} stems from only one linear design, it is still a nonlinear controller in that it uses $q - q_{eq}(\alpha)$ and $\eta - \eta_{Z,eq}(\alpha)$ as inputs.

Figure 6 shows the actual implementation of K_{inner}. The block P denotes the nonlinear missile dynamics. The block G_a denotes the actuator dynamics. The block T denotes a transformation of the actual measurements into their "deviation from trim" form. Note that the inner-loop was

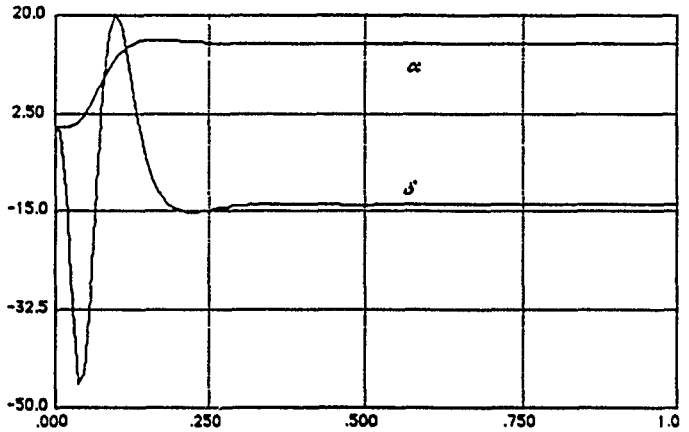


Figure 7: Angle-of-Attack Step Response

designed to give guaranteed robustness properties at loop breaking point 1 (Figure 6). That is, the design was to deliver robust performance for all admissible linear time-varying perturbations. However, we see from Figure 6 that similar robustness properties for linear time-invariant perturbations are obtained at loop breaking points 2 or 3, where actual model deviations are likely to occur. As mentioned in Section 3.2, this is one of the benefits of augmenting integrators and *not* having an inner-loop which updates the trim control input [8].

Regarding stability properties of the feedback system, it is shown in [8,9] that robust stability and robust performance is maintained provided "sufficiently slow" time-variations in the scheduling variable. Although references [8,9] provide quantitative statements regarding "sufficiently slow," an application of these inequalities is likely to lead to conservative conclusions. Rather, the following qualitative interpretation is more suggestive. Namely, a sufficiently slow requirement is with regards to the closed-loop system dynamics. In this case, the closed-loop system has a bandwidth of about 120 rad/s. Thus, one may expect robust performance for angle-of-attack time variations of the same order. Of course, only extensive simulations, non-conservative stability criteria, or an alternate scheduling procedure entirely can really confirm/refute closed-loop robust performance [10].

Finally, Figure 7 shows the response to a 15 degree α step command. Note that this step command leads to excessively large fin deflections. This will not be the case for the overall controller, since the outer feedback loop will be designed to discourage drastic angle-of-attack commands, such as a significantly large step.

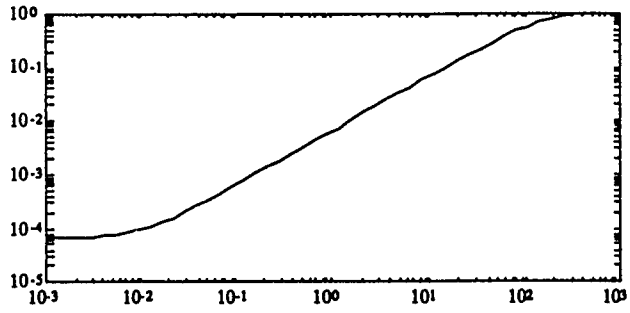


Figure 8: Robustness Weight $W_a(j\omega)$

3.3.2 Outer-Loop Normal Acceleration Control

The outer-loop feedback is designed to generate angle-of-attack commands, α_c, to achieve desired normal accelerations. As mentioned in Section 3.1, the performance requirements are to track acceleration step commands with a steady state accuracy of less than 0.5% and a time constant of 0.2 seconds. However, the outer-loop is bandlimited by the inner feedback loop. That is, the outer-loop should only generate angle-of-attack commands, α_c, within the bandwidth of the inner-loop. It is over this range that we have a reasonable model of the inner-loop behavior.

The first part of this design is to obtain the appropriate quasi-LPV system dynamics from α_c to η_Z. The inner-loop leads to quasi-LPV dynamics from α_c to the states $(\alpha \quad q - \dot{q}_{eq}(\alpha) \quad \delta - \delta_{eq}(\alpha))^T$. However, the normal acceleration is given by

$$\begin{aligned} \eta_Z &= \eta_{Z,eq} + (\eta_Z - \eta_{Z,eq}(\alpha)) \\ &= \frac{QS}{W}(\phi_Z(\alpha) - \frac{b_Z}{b_m}\phi_m(\alpha)) + \frac{QSb_Z}{W}(\delta - \delta_{eq}(\alpha)). \end{aligned}$$

Thus the normal acceleration, η_Z, is a nonlinear function of the quasi-LPV states. Performing a linear approximation of the term $(\phi_Z(\alpha) - \frac{b_Z}{b_m}\phi_m(\alpha))$ leads to

$$\eta_Z \doteq -1.54\alpha + \frac{QSb_Z}{W}(\delta - \delta_{eq}(\alpha)).$$

This leads to an output coefficient matrix which approximates the normal acceleration by a linear function of the quasi-LPV states.

Figure 10 shows the block diagram used for the μ-synthesis design. The block T_{des} denotes the closed-loop dynamics from α_c to η_Z. The measurement is $\eta_c - \eta_Z$, where η_c is the commanded acceleration. The performance objective is to keep $\|\eta_c \mapsto e\| \leq 1$, where

$$W_p(s) = \frac{15s + 200}{40s + 1}.$$

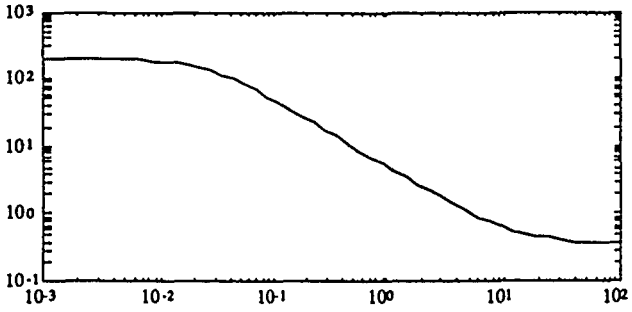


Figure 9: Performance Weight $W_p(j\omega)$ for Acceleration Control

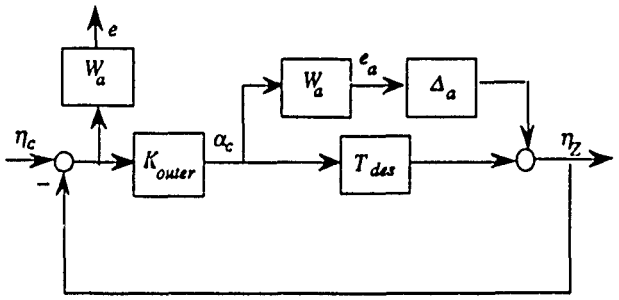


Figure 10: Outer-Loop Acceleration Control

Figure 9 shows the frequency response for W_p. The block Δ_a represents an additive perturbation on the closed inner-loop dynamics. The weighting (cf., Figure 8)

$$W_a(s) = \frac{s + 0.01}{s + 150}$$

reflects that the inner-loop model is fairly accurate at low frequencies. However, the model is less accurate for high frequency angle-of-attack commands.

A μ-sythesis design procedure was performed with this interconnection structure at the set point $\alpha = 0$. The first pass led to a frozen α robust performance level of 3.125. After two iterations, this value was reduced to 0.732.

Once again, the controller for $\alpha = 0$ proved adequate for the entire α range of ± 20 degrees. Thus, no gain-scheduling is required. Recall that this was the case in the inner-loop α control. This was due to the airframe dynamics at $\alpha = 0$ being the most difficult to control. However, the reasoning for the outer-loop not requiring gain-scheduling is different. Recall that the inner-loop was the result of a robust servo design. This feedback in itself has a linearizing effect on the missile dynamics (cf., [3]). For example, one has that $\alpha \doteq I\alpha_c$, for a class of α_c *within the inner-loop bandwidth*. Outside of this bandwidth, the linear behavior will deteriorate—hence the additive uncertainty weighting W_a. Thus, while not needing gain-scheduling in the inner-loop is not typical,

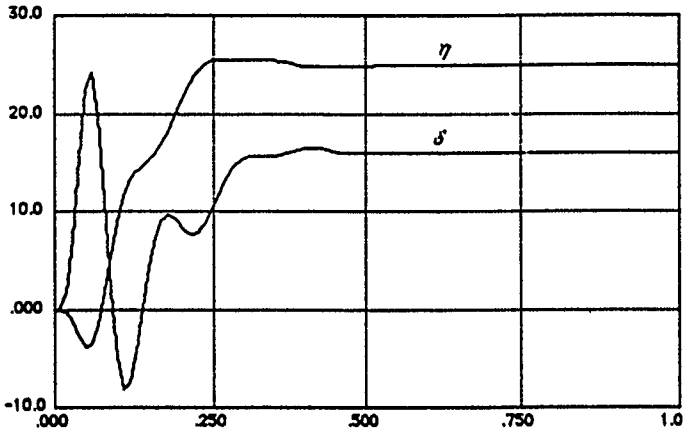


Figure 11: Acceleration Step Response

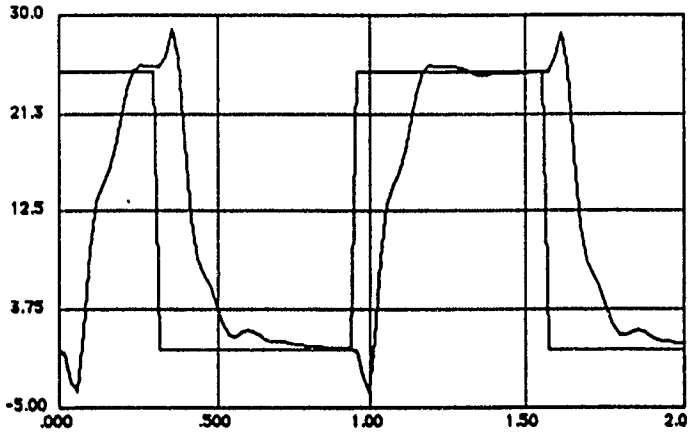


Figure 12: Acceleration Square Wave Response

it seems reasonable in the outer-loop. Note that the linearization through feedback *differs* from that of geometric feedback linearization (e.g., [5]). In this case, the linearization is due to a robust servo design. Hence it is approximate and bandlimited. Furthermore, the design of the outer-loop takes this into account.

Figure 11 shows the response to a $25g$ step command. Note that large fin deflections do not occur as in the angle-of-attack step response (Figure 7). As mentioned earlier, this is due to the outer-loop acknowledging the band-limited performance of the inner-loop. During the missile endgame, the guidance law typically generates large rapidly-varying acceleration commands. To illustrate the performance in such a scenario, Figure 12 shows the response to a square-wave command oscillating between $25g$ and $0g$. Figure 13 shows the fin deflections and angle-of-attack response.

4 Conclusion

In this report, we have considered two problems in nonlinear missile autopilot design.

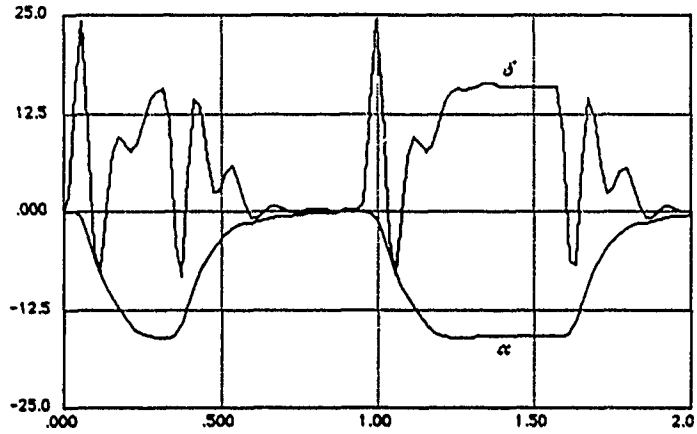


Figure 13: Fin and Angle-of-Attack Square Wave Responses

First, we have investigated the fading memory property primarily in the context of robust stability. Some possible directions are the following. One direction is the extension of these results to the case of structured dynamic uncertainty. Another direction is determining what classes of nonlinear systems have fading memory. Finally, there is the issue of norm-computation of nonlinear systems.

Second, we presented a novel approach to gain-scheduled missile autopilot design for longitudinal missile dynamics. Some key features of this approach are as follows:

- The missile dynamics are brought to an LPV form via a *state-transformation* rather than the usual linearization.
- An integrator is augmented so that no "update of trim control" feedback loop is present.
- An inner/outer-loop decomposition is applied. An effect of the inner loop is to linearize the missile dynamics in an approximate and bandlimited manner. This leads to a simplified outer-loop design with guaranteed inner-loop robustness properties.

Work is in progress on applying these concepts to a 6DOF missile autopilot design.

References

[1] G. Balas, et al., "μ-Analysis and Synthesis Toolbox: μ-Tools," MUSYN Inc. and The Mathworks Inc., December 1990.

[2] C.A. Desoer and M. Vidyasagar, *Feedback Systems: Input-Output Properties*, Academic Press, New York, 1975.

[3] C.A. Desoer and Y.-T. Wang, "Foundations of Feedback Theory for Nonlinear Dynamical Systems ," *IEEE Transactions on Circuits and Systems*, 1980.

[4] P. Dorato, ed., *Robust Control*, IEEE Press, New York, 1987.

[5] A. Isidori, *Nonlinear Control Systems: An Introduction*, Lecture Notes in Control and Information Science, Vol. 72, Springer-Verlag, Berlin, 1985.

[6] R. Reichert, "Modern Robust control for Missile Autopilot Design," *Proceedings of the American Control Conference*, San Diego, CA, June 1990.

[7] S.M. Shahruz and S. Behtash, "Design of Controllers for Linear Parameter Varying Systems by the Gain Scheduling Technique," Berkeley Engineering Research Institute, Memorandum No. BERI M90/1, January 1990.

[8] J.S. Shamma, *Analysis and Design of Gain Scheduled Control Systems*, Doctoral Thesis, Dept. of Mechanical Engineering, MIT, 1988.

[9] J.S. Shamma and M. Athans, "Guaranteed Properties of Gain Scheduled Control of Linear Parameter-Varying Plants," *Automatica*, May 1991.

[10] J.S. Shamma and M. Athans, "Analysis of Nonlinear Gain Scheduled Control Systems," *IEEE Transactions on Automatic Control*, August 1990.

[11] J.S. Shamma and M. Athans, "Gain Scheduling: Potential Hazards and Possible Remedies," submitted to *American Control Conference*, Boston, MA, June 1991.

[12] J.S. Shamma and J.R. Cloutier, "A Linear Parameter Varying Approach to Gain Scheduled Missile Autopilot Design," submitted to *1992 American Control Conference*, Chicago, IL, 1991.

[13] J.S. Shamma and J.R. Cloutier, "Gain Scheduled Missile Autopilot Design using Linear Parameter Varying Transformations and μ-synthesis," submitted to *AIAA Journal of Guidance, Control, and Dynamics*, 1991.

[14] J.S. Shamma and R. Zhao, "Fading Memory and Necessity of the Small Gain Theorem," *Proceedings of the 30th IEEE Conference on Decision and Control*, Brighton, UK, December 1991.

[15] J.S. Shamma and R. Zhao, "Fading Memory Feedback Systems and Robust Stability," submitted to *Automatica*, 1991.

[16] G. Zames, "On the Input-Output Stability of Time-Varying Nonlinear Feedback Systems. Part I: Conditions Using Concepts of Loop Gain, Conicity, and Positivity," *IEEE Transactions on Automatic Control*, Vol. AC-11, No. 2, 1966, pp. 228–238.

QUALITATIVE EFFECTS OF KKV IMPACT LOCATIONS ON HYDRAULIC RAM IN FUEL TANKS AT FIFTY PERCENT ULLAGE (Hydrocode Analysis)

Steven A. Trogdon, Associate Professor
Department of Mathematics and Statistics
University of Minnesota - Duluth
Duluth, MN 55803

and

Research Associate
AFOSR Summer Faculty Research Program
Wright Laboratory
WL/MNSA
Eglin AFB, FL 32542

August 1991

ABSTRACT

The impact of a KKV traveling at 7 km/sec and interacting with a non-pressurized fuel tank at fifty percent ullage had been modeled with the CTH hydrocode. It was determined that proximity of the KKV to that portion of the fuel tank which is empty can produce significantly different responses of the tank to hydraulic ram. Two different impact scenarios were analyzed to exhibit that different responses are possible. From a lethality point of view one would like to know which factors will produce what distribution of impulse loading. Hydrocode calculations provide a useful tool for not only giving a qualitative assessment of which factors are important but for also giving the desired distribution of impulse loading for these factors. This research effort forms a starting point for future efforts to characterize lethality effects associated with hydraulic ram in fuel tanks.

INTRODUCTION

It is well-known that hydraulic ram pressures generated by a projectile penetrating a fuel tank can be a major contributor to aircraft attrition in combat [1]. Damage to such fluid-filled tanks at low velocities (less than 2 km/sec) has been studied extensively [2,3]. During penetration of the tank, intense transient pressures are generated within the fluid and are subsequently transmitted to the walls of the tank. Due to excessive deformation, the tank structure can fail (rupture) catastrophically in response to the hydraulic ram pressure loading. In addition, breakup of the tank can be augmented by potential fracture failures emanating from ballistic damage [4,5,6].

All the particulars mentioned above become accentuated in space-based applications where the penetrating projectile or kinetic energy kill vehicle (KKV) travels at speeds in excess of 5 km/sec (hyper-velocities). At such speeds it is not uncommon for the KKV, even though it may eventually be depleted, to make initial penetration of the tank and fluid and yet remain relatively intact. Upon contacting the fluid the KKV is rapidly decelerated as it transfers its kinetic energy to the fluid in the form of a strong shock wave (hydraulic ram). It is during this deceleration of the KKV that it breaks up and is eventually depleted, i.e. deposits its energy and momentum in the fluid. Since the KKV quickly transfers its energy to the fluid, it acts as if a point source of energy were deposited in the fluid at the point of impact; the result being a hemispherical shock wave that emanates from the point of impact. It has been observed that at the higher impact velocities the location of the shock wave as a function of time depends only on the initial kinetic energy of the KKV and not on its material and size.

Because the resulting shock wave is hemispherical it has been proposed that the impact event could be simulated by an explosive charge located at the point of impact and producing a hemispherical blast front.

It has been demonstrated, both through experimentation and through hydrocode calculations, that the above mentioned simulation procedure has validity, provided the KKV has been depleted [7,8,9]. This assumption will undoubtedly depend upon the relative size of the KKV to that of the tank. In situations where size is not an issue the depletion assumption appears to be a good one provided the tank is entirely full, i.e. the tank has zero percent ullage. It is not clear, a priori, that the KKV will be depleted in situations where the tank is not entirely full. If the KKV is not depleted then it is not clear if a charge could be designed which, when exploded, would simulate the impact event since the potential would exist for fracture failures from fragments of the KKV. If the tank is not entirely full then the effect of hydraulic ram on the breakup of the tank could strongly be dependent on the proximity of the KKV to that portion of the tank which is empty. In this report we give a qualitative assessment of how proximity of the KKV to that portion of the tank which is empty affects the response of the tank to hydraulic ram when the tank ullage is fifty percent.

PROBLEM DISCUSSION

The impact of a kinetic energy kill vehicle (KKV) traveling at 7 km/sec and interacting with a non-pressurized spherical aluminum shell filled with water at fifty percent ullage has been modeled with the CTH hydrocode. Water was used in the modeling since only hydrodynamic and not explosive effects were of interest. The proximity of the KKV to that

portion of the tank which is empty can produce different responses of the tank to hydraulic ram. Two different impact scenarios were analyzed in order to exhibit that different responses are possible. In both scenarios the direction of impact was perpendicular to the water free surface and along a radius of the spherical tank. The first scenario was one in which the KKV first impacted that portion of the tank which was filled with water. The second was one in which the KKV first impacted that portion of the tank which was empty. Both impact scenarios were run out to 150 sec. The configuration for the two different problems is given respectively in the upper left frames of Figures 1 and 5. The aluminum tank was modeled using approximately five computational zones of 0.05 cm and weighed (empty) approximately 171 gm. The thickness of the tank was 0.254 cm. Considerable detail was used in modeling the KKV which weighed approximately 85 gm. Materials used in modeling the KKV were aluminum, brass and nylon (see APPENDIX).

The calculation was, in the terminolgy of CTH, two-dimensional cylindrical and used approximately 400 and 850 zones in the respective x (radial) and y (vertical) directions. The mesh and material insertion records sections of the input file to CTHGEN, the CTH mesh generator, are given in the APPENDIX. These records are sufficient to define the discretization and geometry of the impact problem. The total problem time was 150 sec and dumps were performed at 10 sec intervals. The calculations were performed on Eglin AFB's Cray Y-MP8/2128 supercomputer. Each problem took something slightly in excess of 60 CPU hours to run to completion. Needless to say, the analysis reported herein would be impossible without this computational resource.

RESULTS

The qualitative results of the two impact scenarios are given in Figures 1 through 8. Four frames have been plotted in each figure, each frame giving the response of the tank and water to impact at different dump times. The interval between dump times was 10 sec. The left half of each frame shows the location of the different materials at the corresponding dump time. The right half of each frame shows the interfacial lines between the different materials. In addition, lagrangian tracer particles were embedded in the center of the tank wall at five degree intervals. Although these tracer particles were not used for this presentation, they were included in the two problems to enable one to recover, as a function of time, how hydraulic ram effects impulsive loading on the tank wall.

The response of the tank and its eventual breakup is very different for the two impact scenarios. Figures 1 through 4 show plots of the impact scenario when the filled portion of the tank is closer to the KKV. For this impact scenario the KKV is entirely depleted. Upon impact, the KKV generates a hemispherical shock wave in the water. Between 30 and 40 sec after impact, the shock wave reaches the water free surface and starts to deform it. Between 90 and 100 sec after impact, the water free surface slams into the opposite side of the tank from that of impact. Immediately, from this location, a flexural wave propagates in the latitudinal direction in the tank wall. During the time it takes for the water free surface to reach the opposite tank wall, the shock wave that was generated upon impact interacts with the tank wall which, at time zero, was in contact with water. Again, a flexural wave is generated in the tank wall and propagates from the

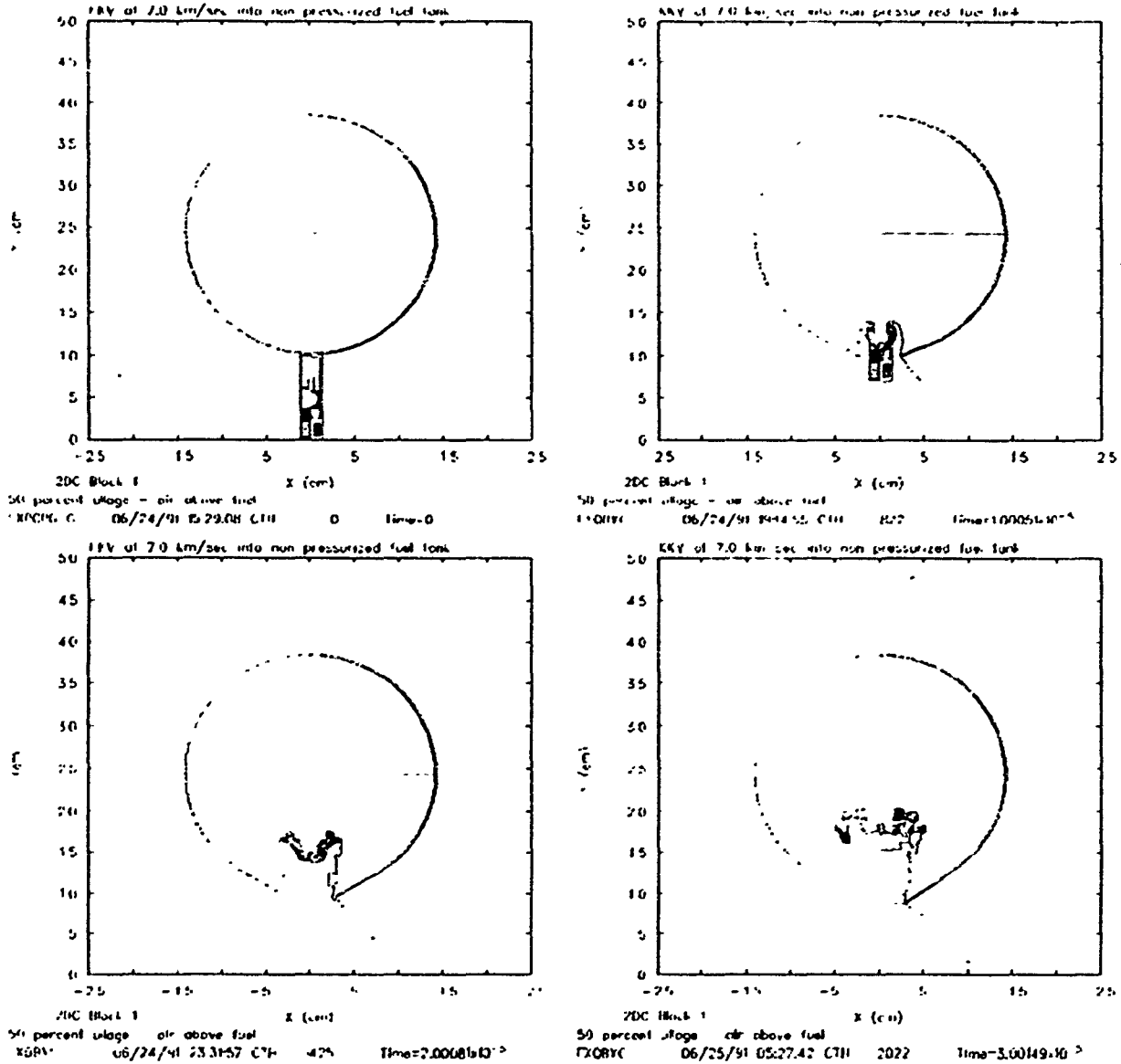


Fig. 1.

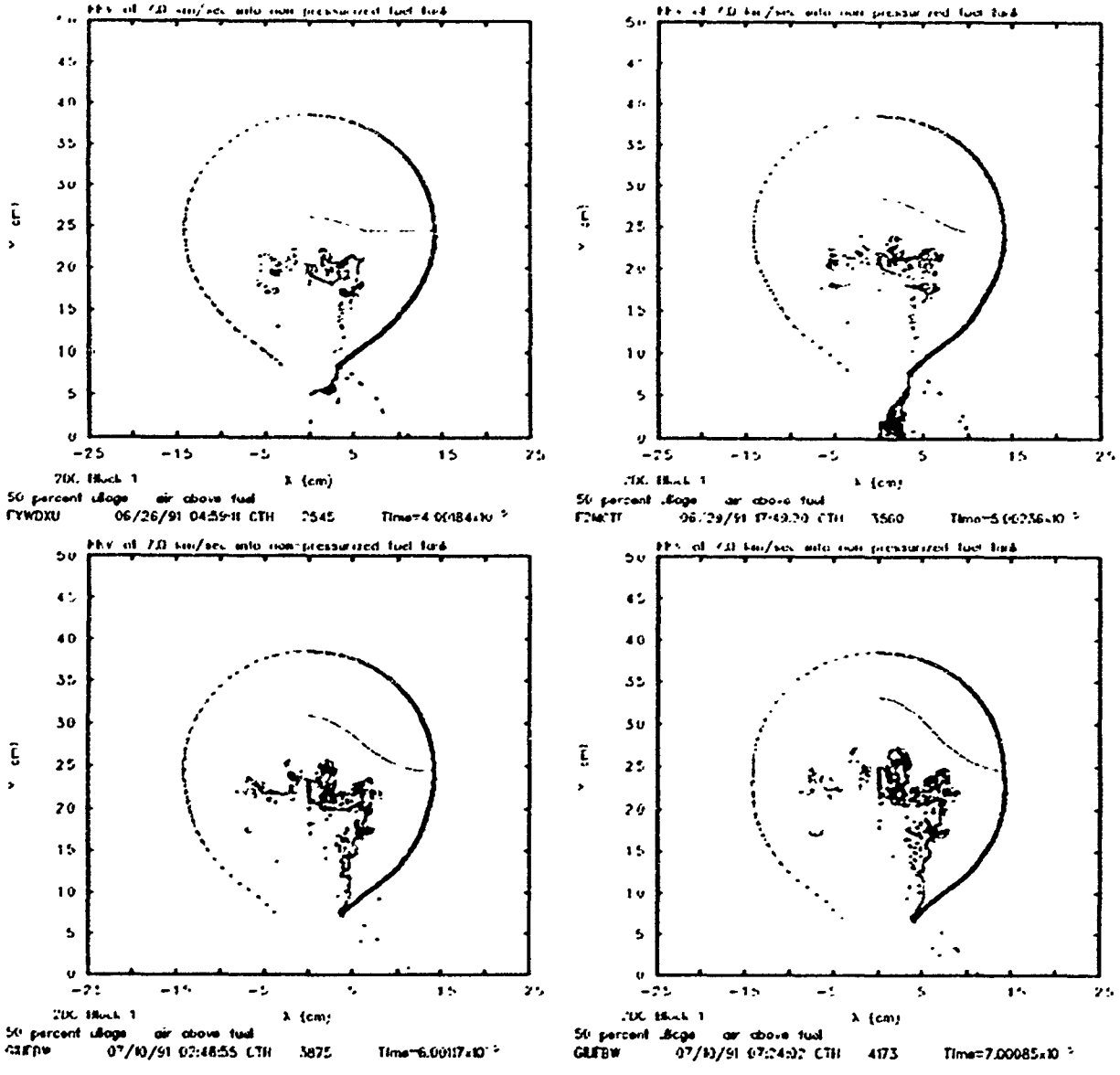

z (cm)
r (cm)
50 percent ullage air above fuel
50 percent ullage air above fuel
50 percent ullage air above fuel
50 percent ullage air above fuel

Fig. 2.

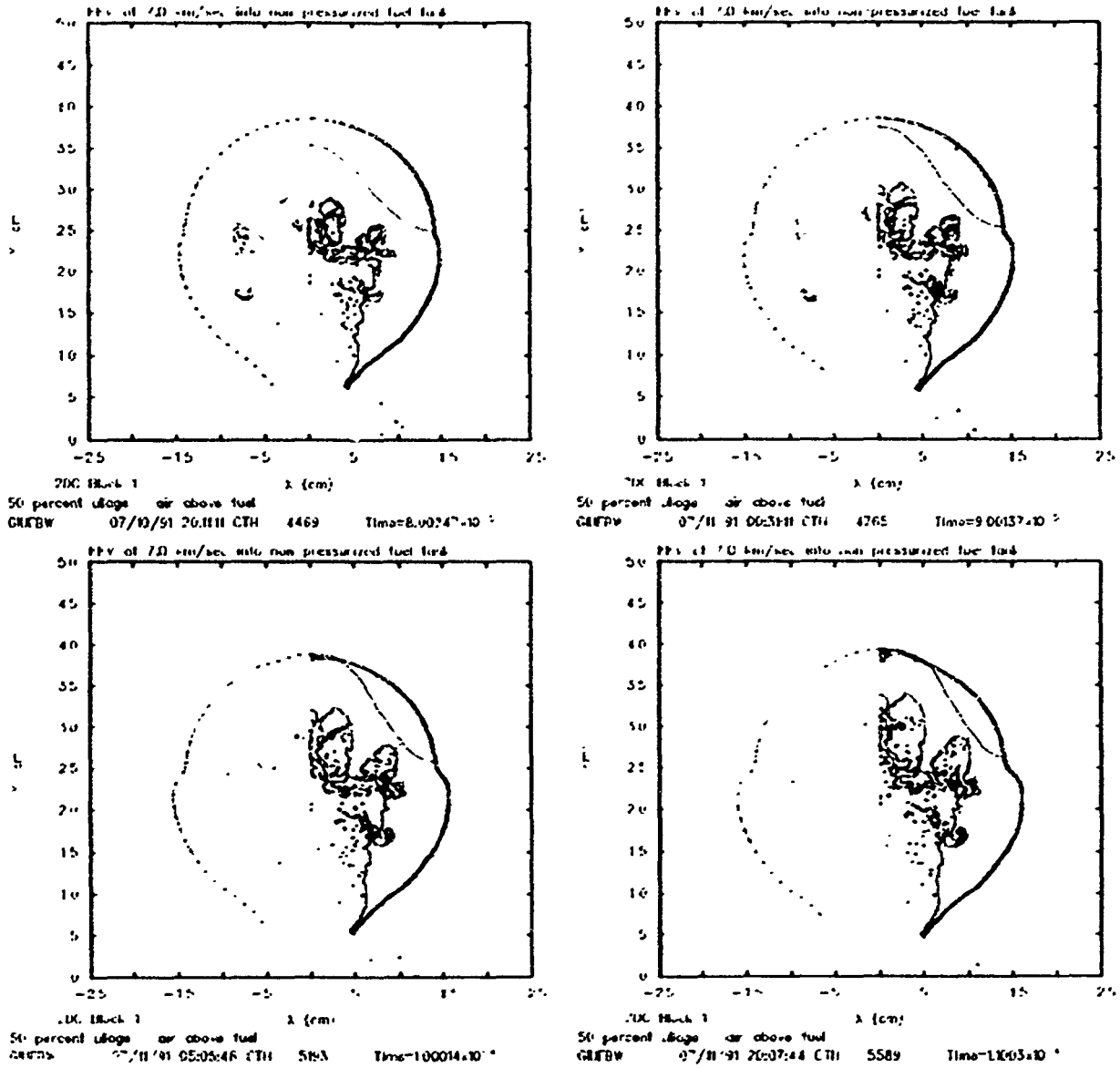


Fig. 3.

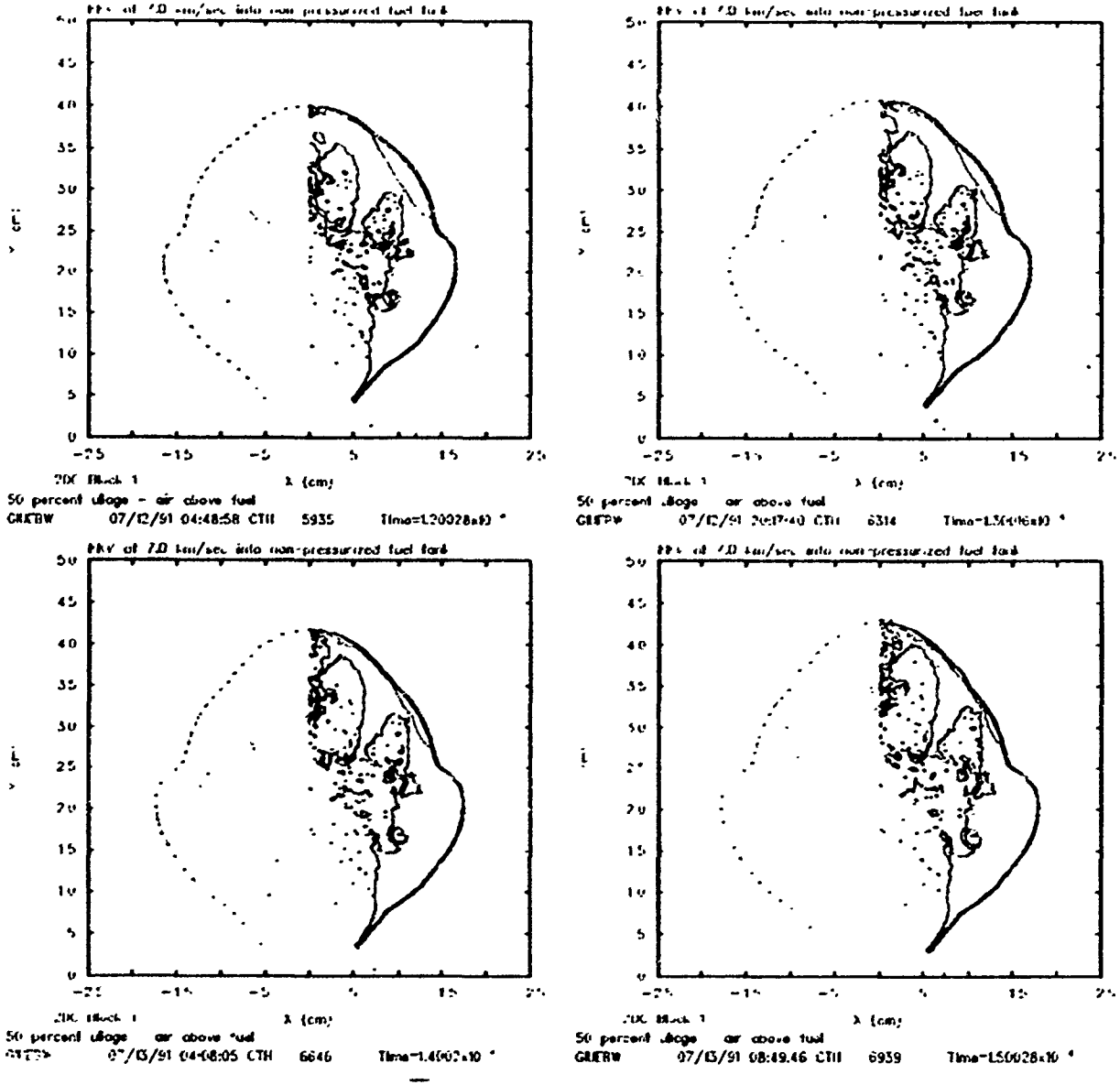

PRV at 7.0 km/sec into non-pressurized fuel tank
Y (cm)
X (cm)
50 percent ullage - air above fuel
07/12/91 04:48:58 CTH 5935
50 percent ullage air above fuel
07/13/91 04:08:05 CTH 6646
50 percent ullage air above fuel
07/13/91 08:49:46 CTH 6939

Fig. 4.

location of impact around the tank wall. The net effect of this flexural wave is to open up the tank and to cause significant deformation of the tank wall near the contact line where the water initially met the tank. Thus, the effect of the impact process is to spawn, through hydraulic ram, two flexural waves and it is these two flexural waves that are responsible for the response and breakup of the tank. The effect of impacting the tank near its filled portion is that the momentum and energy of the KKV is distributed in a nearly uniform fashion along a hemispherical shock in the water. Since the KKV was completely depleted during impact, it would appear that one could accurately simulate this impact scenario by replacing the impact process with an appropriately designed explosive charge.

Figures 5 through 8 show plots of the impact scenario when the empty portion of the tank is closer to the KKV. For this impact scenario the KKV is not entirely depleted. Upon impact, the KKV slices through the tank wall and remains relatively intact. The portion of the tank that was sliced out appears to have been vaporized and it is this material that first reaches the water free surface at around 20 sec. The KKV starts to break up significantly when it impacts the water free surface; however, even after 40 sec the lower portion of the KKV is recognizable. As in the discussion of the previous impact scenario a shock wave is generated in the water at the location of impact; however the shock here is not as strong as the one discussed previously. The present shock moves at a speed close to that of the debris from the KKV, while in the previous impact scenario the shock wave out ran the debris from the KKV. Even though the shock is not as strong as in the previously discussed case, it is nevertheless more concentrated and focused. This is perhaps due to the impulse from the KKV interacting with a free

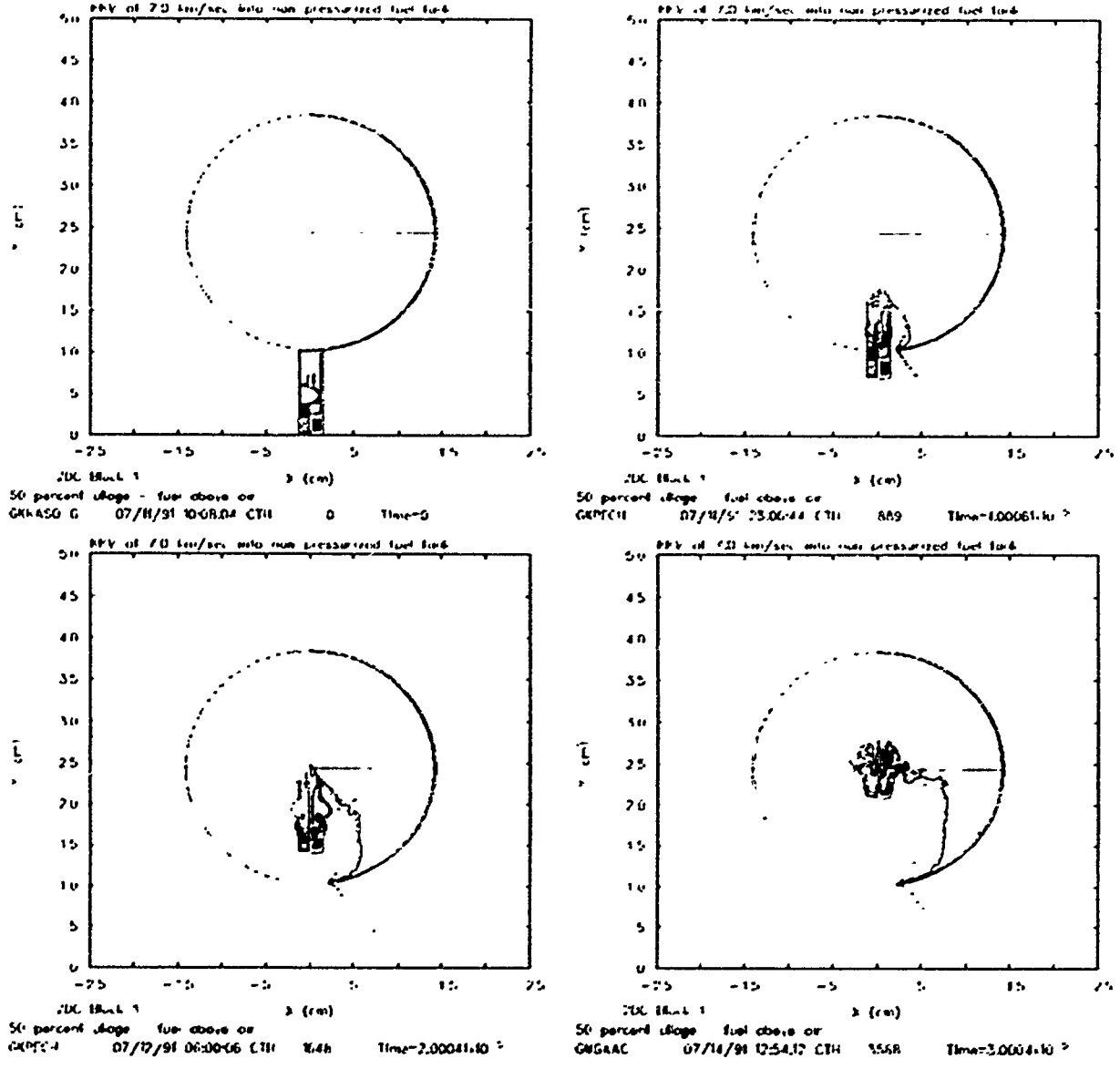


Fig. 5.

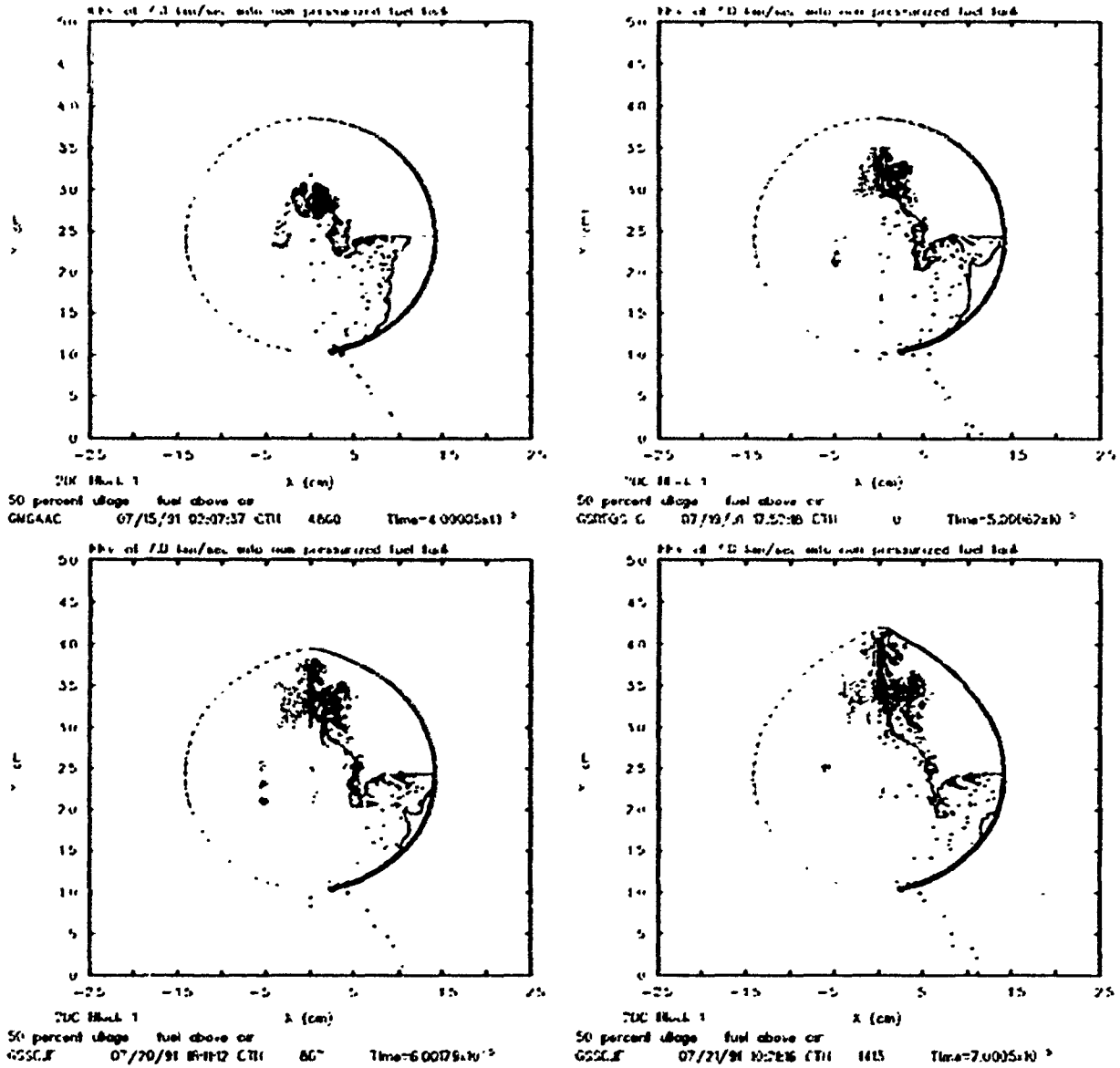


Fig. 6.

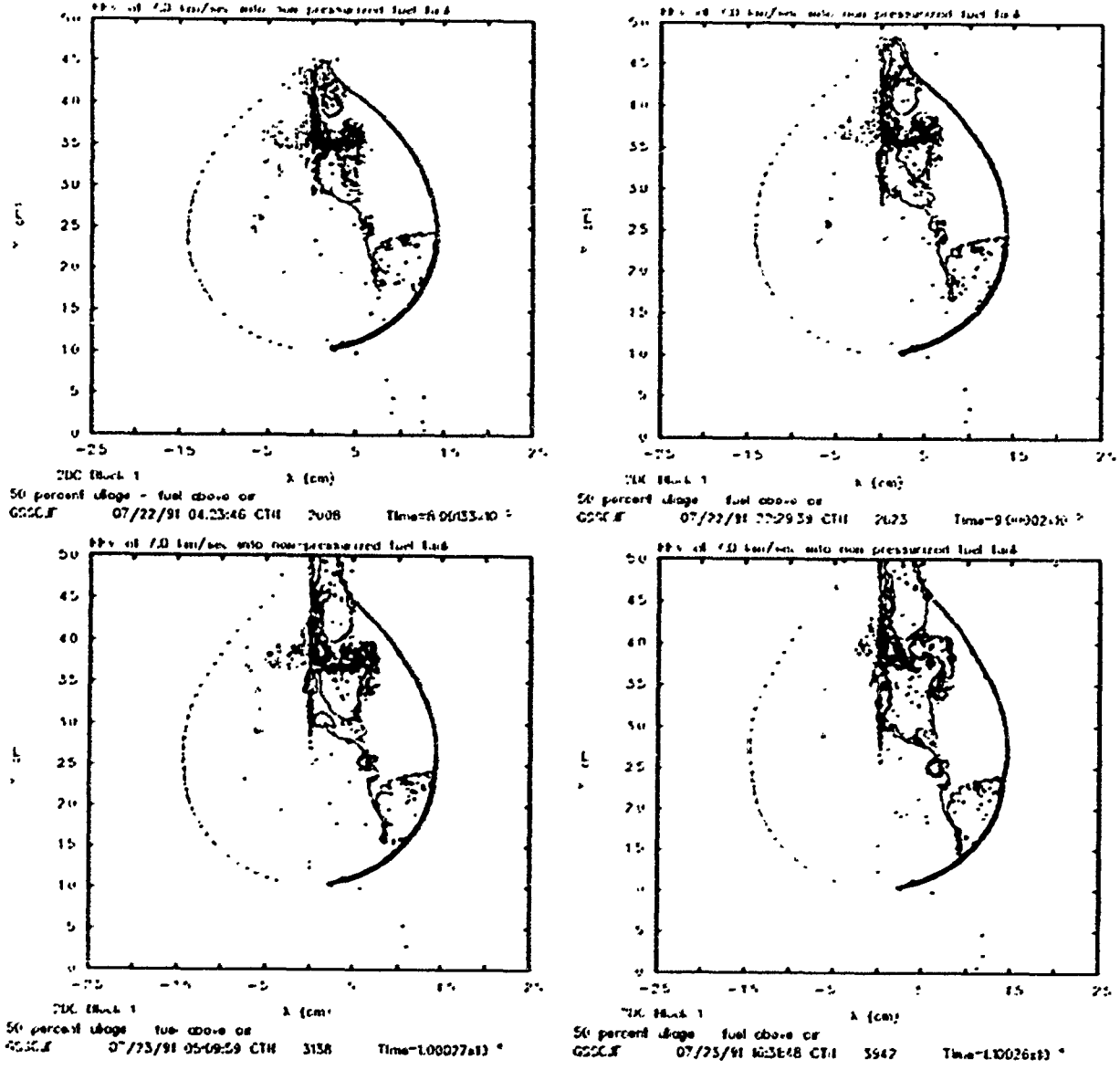


Fig. 7.

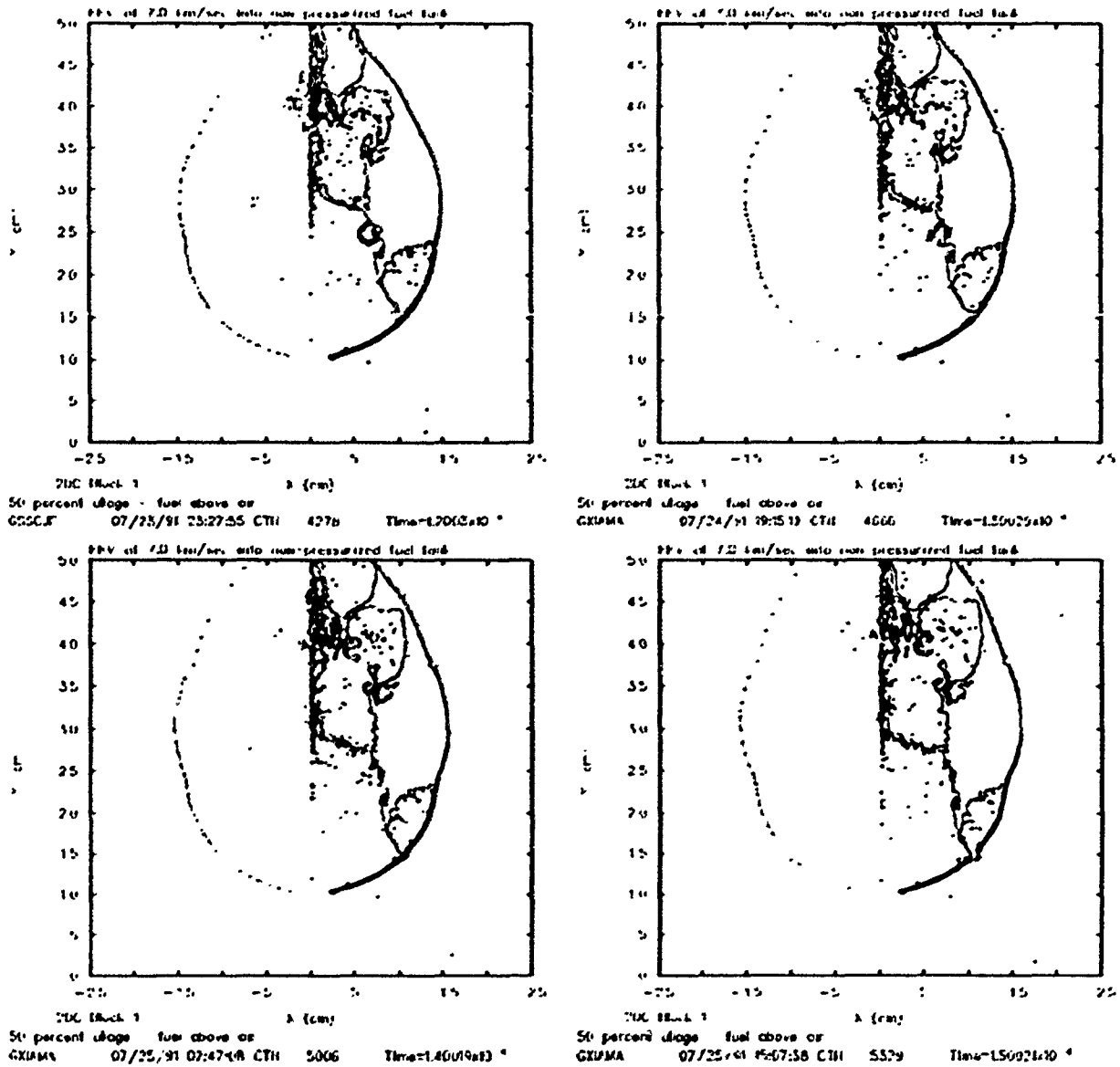


Fig. 8.

surface boundary and not a no-slip boundary as in the first impact scenario. The focused nature of the shock is a major factor in contributing to tank breakup. However, the shock alone is not the sole reason for eventual tank breakup. Since the shock and the debris from the KKV are traveling at nearly the same speed, they both contribute to excessive deformation of the tank wall. Like the previous impact scenario, a flexural wave is generated in the tank wall which eminates from the location where the shock wave in the water and debris from the KKV impact the tank wall. This flexural wave propagates around the tank causing a modest amount of deformation at 150 sec. However, it is the excessive deformation of the tank due to shock pressures and KKV debris that are the primary reason for tank breakup. Since the KKV was not completely depleted during impact, it is doubtful whether this impact scenario could be simulated by means of an appropriately designed explosive charge.

CONCLUSIONS

The impact of a KKV traveling at 7 km/sec and interacting with a non-pressurized spherical shell filled with water at fifty percent ullage has been modeled with the CTH hydrocode. The purpose of the modeling was to give a qualitative assessment of how impact locations affect the response of fuel tanks to hydraulic ram. It was determined that proximity of the KKV to that portion of the fuel tank which is empty can produce different responses of the tank. In particular the mechanisms which control breakup of the tank are strongly dependent on where the KKV impacts the fuel tank relative to location of fuel in the tank. In all cases hydraulic ram is the source of the mechanisms; however,

other factors such as ullage or initial conditions effect which mechanisms will be present. The closer the KKV is to the filled portion of the tank the more likely breakup of the tank will be due to propagation of large amplitude flexural waves in the tank. The closer the KKV is to the empty portion of the tank the more likely breakup of the tank will be due to excessive deformation from a focused shock wave and KKV debris.

From a lethality point of view it is imperative to understand which factors have what effects on fuel tank breakup. It has been demonstrated that not only is ullage such a factor, but so is proximity of the KKV to fuel in a tank. A statement that hydraulic ram is important is, of itself, insufficient to describe effects hydraulic ram have on fuel tank breakup since, depending on which factors are present, hydraulic ram can produce different effects. Ultimately, from a lethality point of view, one would like to know which factors will produce what distribution of impulsive loading. Hydrocode calculations for different impact scenarios provide a useful tool for not only giving a qualitative assessment of which factors are important but for also giving the desired distribution of impulse loading for these factors. The hydrocode calculations provided as part of this research effort form a starting point for future efforts at defining criteria and extracting information relative to characterizing lethality effects associated with hydraulic ram in fuel tanks.

REFERENCES

1. W. M. Herlin and J. G. Avery, "Hydraulic Ram Structural Response Computer Program (HRSR) Volume II: Final Report," September 1981.

2. Francis S. Stepka and C. Robert Morse., "Preliminary Investigation of Catastropohic Fractrure of Liquid-Filled Tanks Impacted by High-Velocity Particles," NASA TN D-1537, 1963.

3. F.S. Stepka, C.R. Morse and R.P. Dengle, "1ntroduction of Characteristics of Pressure Waves Generated in Water-Filled Tanks Impacted by High- Velocity Projectiles," Lewis Research Center report, 1965.

4. Naval Weapons Center, "Fluid Dynamic Analysis of Hydraulic Ram" by E. A. Lundstrom, China Lake, CA, NWC, July 1971, (NWC TP 5227, publication UNCLASSIFIED.)

5. Naval Weapons Center, "Fluid Dynamic Analysis of Hydraulic Ram III" by E. A. Lindstrom and W. K. Fung, China Lake, CA, NWC, October 1974, (JTCG/AS-74-T-015, publication, UNCLASSFIED.)

6. G. Birchoff and E. H. Zarantonello, "Jets, Wakes and Cavities," New York, Academic Press, 1957.

7. M. A. Fry, "Hypervelocity Impact on Pressurized Fuel Tanks (Hydrodynamic Calculations)," SAIC Report 88/1431, January 15, 1988.

8. M. A. Fry, "Hypervelocity Impact on Pressurized Fuel Tanks with 0.5 Ullage," SAIC Report 88/1677, April 26, 1988.

9. M. A. Fry, "Definition of the Full Scale PBV Propellant Tank Response at 15 km/sec," SAIC Report 88/1880, September, 26, 1988.

APPENDIX

```
*************************************************************************
*
*  mesh records
*
mesh
  block 1  geom=2dc  type=e    * 2-D rectangular geometry, Eulerian mesh
    x0  0.                     * X starts at 0, (req'd for axisym. case)
      x1  dxf=.05 number=300 width=15.0
      x2  dxfirst=.05 dxlast=.15 width=10.0
    endx
    y0=-.15
      y1  dyf=.05 number=776 width=38.8
      y3  dyfirst=.05 dylast=.15 width=6.2
    endy
    xact = 0.  1.60 * upper and lower bounds of X mesh (initial activity)
    yact = 0.  10.5 * upper and lower bounds of Y mesh (initial activity)
  endb
endm
```

```
*************************************************************************
*
*  material insertion records
*
insertion of material
*
  block 1
*
    package aluminum (part 1)
      material 1
      numsub 5
      yvel 7.0e5
      insert box
        x1=1.016 y1=0.0 x2=1.27 y2=10.16
      endi
    endp
*
    package aluminum (part 2)
      material 1
      numsub 5
      yvel 7.0e5
      insert box
        x1=0.0 y1=0.0 x2=1.016 y2=0.508
      endi
    endp
*
    package aluminum (part 8)
      material 1
      numsub 5
      yvel 7.0e5
      insert box
        x1=0.0 y1=9.906 x2=1.016 y2=10.16
      endi
    endp
*
    package nylon (part 3a)
      material 2
      numsub 5
      yvel 7.0e5
      insert box
        x1=0.254 y1=6.35 x2=.381 y2=7.366
      endi
    endp
*
    package nylon (part 3b)
      material 2
      numsub 5
      yvel 7.0e5
      insert box
        x1=0.254 y1=5.842 x2=1.016 y2=6.35
      endi
    endp
*
    package aluminum (spherical shell)
      material 3
      numsub 50
      yvel 7.0e5
      insert circle
        center=0.0,4.826 radius=1.016
      endi
      delete circle
        center=0.0,4.826 radius=.889
      endd
    endp
*
    package nylon (part 5)
      material 2
      numsub 5
      yvel 7.0e5
      insert box
        x1=0.0 y1=2.794 x2=1.016 y2=3.81
      endi
    endp
*
    package brass (part 5)
      material 4
      numsub 5
      yvel 7.0e5
      insert box
        x1=0.0 y1=2.54 x2=1.016 y2=2.794
      endi
    endp
*
    package aluminum (part 6a)
      material 1
      numsub 5
```

```
    yvel 7.0e5
    insert box
      x1=0.0 y1=2.032 x2=1.016 y2=2.54
    endi
  endp
*
  package aluminum (part 6b)
    material 1
    numsub 5
    yvel 7.0e5
    insert box
      x1=0.0 y1=.508 x2=0.254 y2=2.032
    endi
  endp
*
  package aluminum (upper torus)
    material 3
    numsub 75
    yvel 7.0e5
    insert circle
      center=0.635,1.651 radius=.381
    endi
    delete circle
      center=0.635,1.651 radius=.254
    endd
  endp
*
  package aluminum (lower torus)
    material 3
    numsub 75
    yvel 7.0e5
    insert circle
      center=0.635,0.889 radius=.381
    endi
    delete circle
      center=0.635,0.889 radius=.254
    endd
  endp
*
  package water
    material 6
    numsub 5
    yvel=0.0e5
*     pressure=41368560.0
    insert circle
      center=0.0,24.384 radius=13.97
    endi
    delete box
      x1=0.0 y1=24.384 x2=13.97 y2=38.354
    endd
  endp
*
  package air
    material 7
    numsub 5
    yvel=0.0e5
*     pressure=41368560
    insert circle
      center=0.0,24.384 radius=13.97
    endi
    delete box
      x1=0.0 y1=24.384 x2=13.97 y2=10.414
    endd
  endp
*
  package aluminum (tank)
    material 5
    numsub 5
    yvel=0.0e5
    insert circle
      center=0.0,24.384 radius=14.224
    endi
    delete circle
      center=0.0,24.384 radius=13.97
    endd
*
  endp
*
endb
*
endi
```

MOUNTING TECHNIQUES FOR HIGH G IMPACT SENSORS

Wayne J. Zimmermann, PhD

INTRODUCTION.

The design of a smart, real-time, earth penetrating device capable of identifying the enveloping strata requires a complex system having the following characteristics: highly shock resistant, quick response, completely self contained, reliable sensors and low self generated noise. Figure 1 presents a functional view of the primary components of such a system.

For the remainder of the report such a device will frequently be referred to by the phrase: EP device.

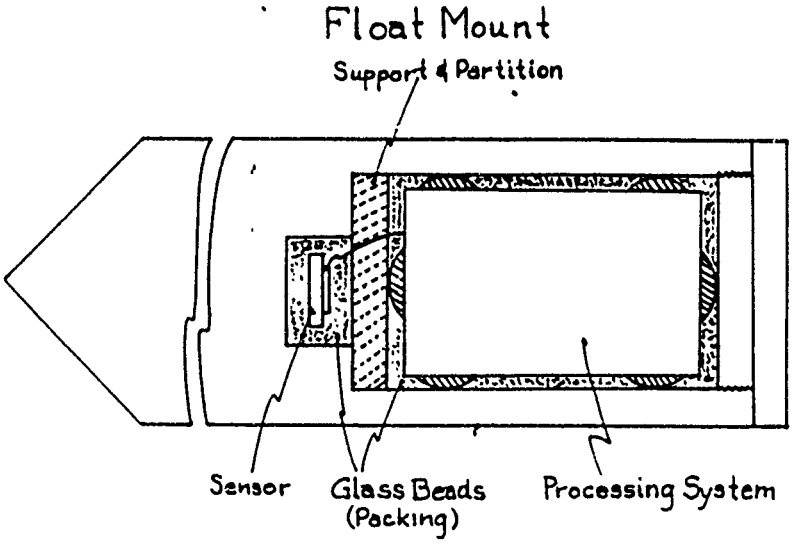


Figure 1. A functional view of a smart EP device.

Research sponsored by the Air Force Office of Scientific Research, United States Air Force, Wright Laboratory (AFSC), Eglin Air Force Base, Fl.

Time limitation required that a particular aspect of the overall design be addressed. Hence, this report is restricted to a functional review and the problems of interest, namely those associated with the sensors.

FUNCTIONAL DESCRIPTION OF A EARTH PENETRATING DEVICE.

Assuming the penetrating system is given enough velocity to significantly penetrate the earth the system must perform as follows:

1: Activate on impact.

2: The sensor should convert its information to proportional voltage levels.

3: Sense the strata and variations.

4: Perform the identification in real-time, i.e., all calculations must be done in less than 300 micro-seconds.

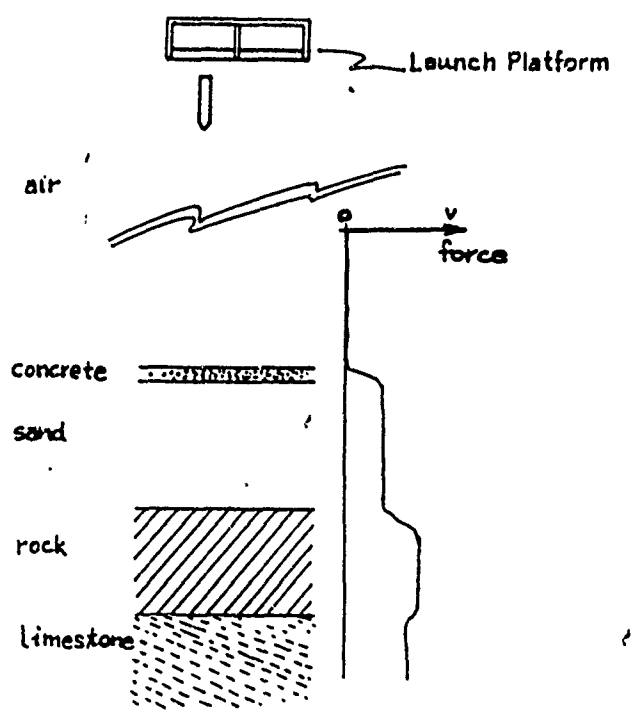


Figure 2. Cross-section of selected earth.

To understand the functional behavior of a EP device consider the following theoretical experiment: Let the structure of the earth be defined by: a thin strata of hard concrete, several feet of compressed strata such a sand supported by several feet of hard uniform rock which in turn overlays a very thick limestone strata, see Figure 2.

Generally the structure of the earth is not known, hence a penetrating device can be dropped from some height, thereby providing a significant penetration which provides the sensing system with the necessary information needed to identify the structure.

Throughout the following discussion we will assume that the EP device strikes the earth at a normal and proceeds along a straight line. Further, all EP devices are assumed to have a conic nose-cone. These are strong assumptions since the device almost never strikes the earth at a normal and shape of the nose has a definite effect on the penetration has shown by C. W. Young [6].

During free fall the sensor should record zero since the atmospheric resistance is small (relative). On impact with the earth the device should begin recording. The resistant force induced by the strata should cause the sensor to react in the following theoretical manner: The initial output voltage of the sensor should be zero. As the device penetrates to its full cross-section the resistant forces increase to their maximum. If the strata's thickness is significantly larger than the diameter of the penetrating device then the device will detect a constant resistant force. Once the device is sufficiently close to exiting from the strata the sensor will detect a variation in the resisting force. If the supporting strata has a higher PSI then the resisting forces will increase; if it has a lower PSI then the resisting

forces will decrease. Hence the voltage output function, vo[t], should take the form:

vo[0] = 0

vo[t] = a sharp increasing function. 0 < t < t1

vo[t] = a flat constant function. t1 < t < t2

vo[t] = a sharp change. t2 < t < t3

. . . . . .

as illustrated in Figure 3.

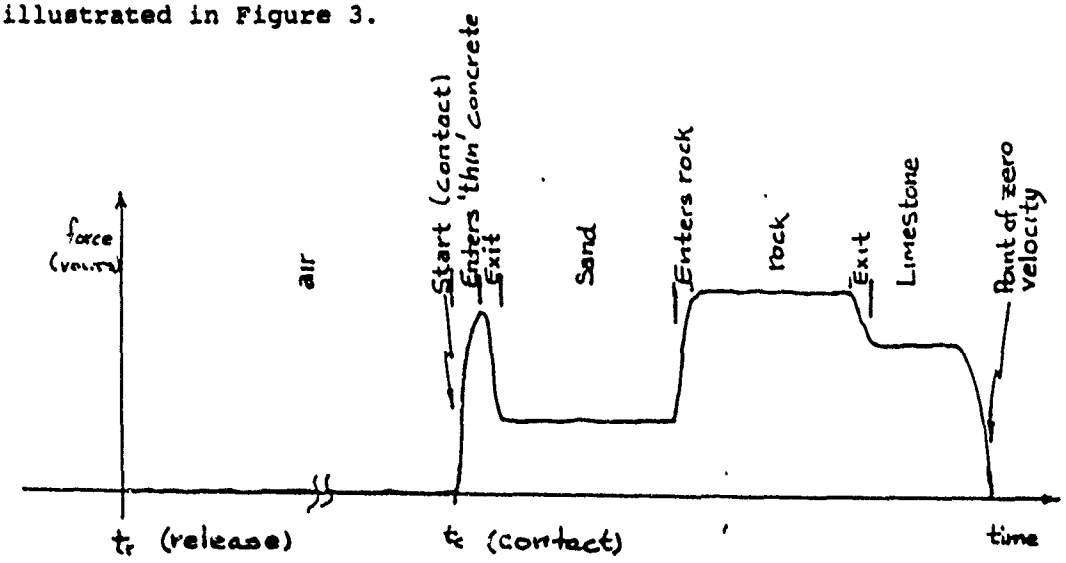


Figure 3. A force plot for a EP device.

Similarly the velocity function will decrease in a step-like function beginning with Vo and ending with zero. An associate velocity curve is given in Figure 4.

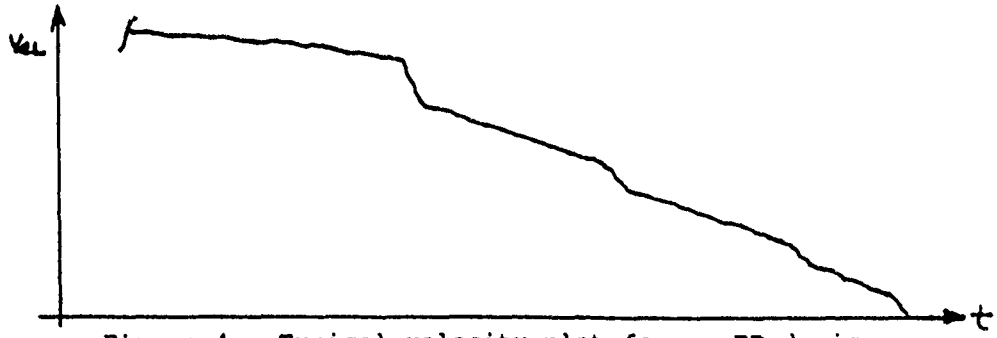


Figure 4. Typical velocity plot for an EP device.

PROBLEM DEFINITION.

The focus of this report is the sub-problem: for a EP device investigate various mounting designs for a sensor capable of responding to the varied strata.

A generic description of a sensor is best described by the beam bending problem. Figure 5 illustrates such a sensing device.

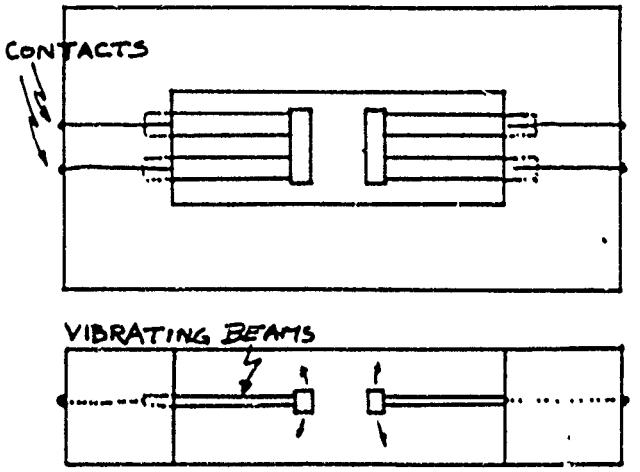


Figure 5. A strata sensor based on beam bending.

The principle behind the use of such a sensor is simply that the amount of bending is proportional to the force encountered. Also the beam will remain bent as long as the force is present. By running a current through the beams of the sensor can be configured as a bridge capable of measuring the resistant force by generating a voltage output proportional to the force.

An important aspect of any system is sensitivity, its ability to differentiate between various strata. If it is capable of making a distinction between two dissimilar strata but not too dissimilar then it is said to be highly sensitive. If the strata must be radically different then it will have a low sensitivity. Much of the study

assumed that the system possessed a low sensitivity.

The reliability of any system is dependent upon the sensors used to gather data hence any sensor used by the system must be consistent, reliable and simple. The Endveco accelerometer employed in this study is a sensor capable of withstanding the high g loads hence it should be reliable. It also respond to high frequency information.

Because the sensor frequently failed when configured with a hard mount and high g loads the experiment was designed to test various mounting configurations. Each configuration must still provide accuracy, reliability, minimal system noise and durability.

EXPERIMENTAL STUDY.

Two experiments were performed, both based on reverse ballistics. One experiment used a pendulum to accelerate the medium; the other use a vacuum based acceleration device.

The pendulum based experiment provided low velocity impacts, hence an acceleration configured in a hard mount never failed. The experiment using the vacuum based acceleration device caused sensor failure.

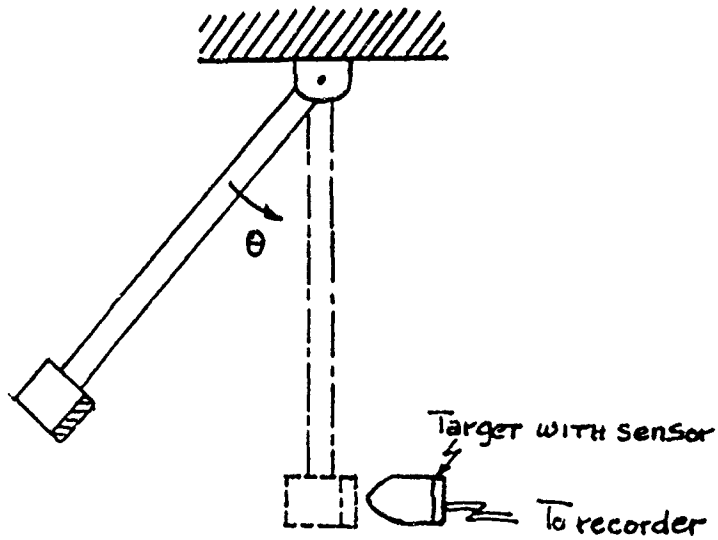


Figure 6. Setup using a pendulum configuration.

Further, for this experiment the contact time between the medium and the target is short and is therefore comparable to Figure 3 over the range tr to tc plus a small time delta. But this is enough to determine the amount of induced ringing.

In testing three distinct mounts it was found that the sensitivity of the system and the accuracy of the system varied. This is seen in the following argument: the hard mount produced large amounts of ringing with the output not at all looking like the Figure 3. It is expected that no system will produce a signal similar to that of Figure 3, but in an attempt to design such a system the accelerometer mount was configured a number of ways. Thr configurations are:

Type 1: Hard mount -- the accelerometer is directly mounted to the EP device.

Type 2: Lose small mass float mount -- the accelerometer is mounted on a small disk. The glass micro beads are poured into a cavity in the EP device while vibrating. The mounted accelerometer is supported in the center of the cavity by the beads.

Type 3: Lose large mass float mount -- the accelerometer is mounted on a cylinder weighting twenty times that of the disk. The glass micro beads are poured into a cavity in the EP device while vibrating. The mounted accelerometer is supported in the center of the cavity by the beads.

Type 4: Tight small mass float mount -- the accelerometer is mounted on a small disk. The glass micro beads are poured into a cavity in the EP device while vibrating. The mounted accelerometer is supported in the center of the cavity by the beads. While the EP device is vibrating the beads are then compressed.

Type 5: Tight large mass float mount -- the accelerometer is mounted on a large cylinder. The glass micro beads are poured into a cavity in the EP device while vibrating. The mounted accelerometer is supported in the center of the cavity by the beads. While the EP device is vibrating the beads are then compressed.

Based on these configuration the following conclusion were drawn. The hard mount responded to the high frequency ringing of the EP device itself. By floating the sensor in a cavity filled with beads this ringing is reduced. But if the beads are too lose the sensor does not respond well to the forces. If the beads are compressed too tight they will form react more like a rigid body and the sensor response is similar to the hard mount.

With respect to the vacuum accelerator on two trial runs were performed. This provided enough information to plan further experiments but no conclusion were drawn from these trials.

RESULTS.

Base on the representative signals given in Figure A.1 through A.4 the following conjecture explaining the physical processes was formulated. The signal acquired during the time interval B is generated by the bending of the of beam while that occurring during period C is due to the forces of resistance of the medium. The third component of the signal is associated with decompression of the material and the last component represents the beam moving to its normal position in response to diminishing external forces and its internal forces. Figure 7 is a typical response.

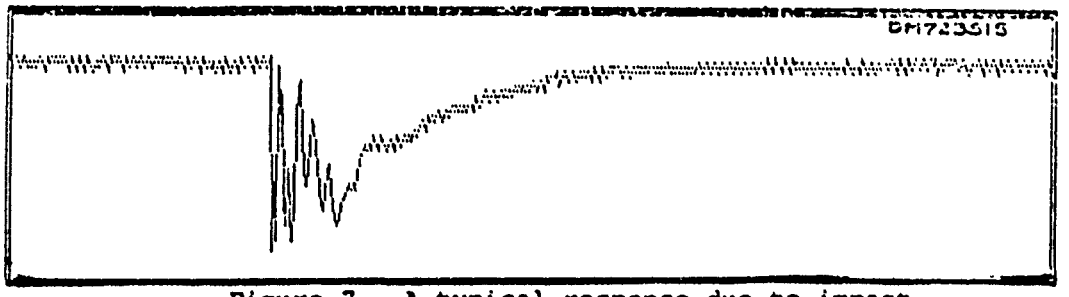


Figure 7. A typical response due to impact.

CONCLUSIONS AND FUTURE DIRECTIONS.

The experiment has indicated that varied mounting schemes must be tested. Further, additional study concerning the type of accelerometer must continue with the direction being focused on the reduction of the induced ringing. Since theoretically, the force is axial and opposite in direction to the motion of the EP device some directional damping device might serve to reduce the ringing. Figure 8. details a possible configuration.

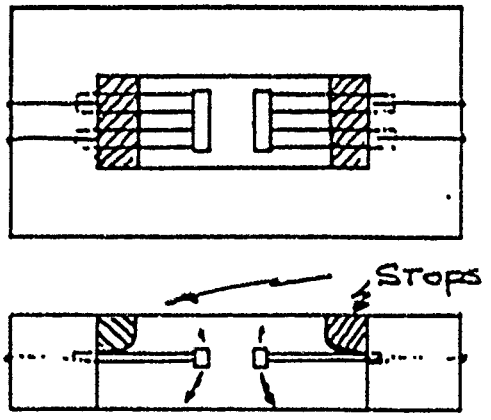


Figure 8. A directional accelerometer.

Any enhancement in the design of the sensing device will require a greater understanding of the problem. Hence answers to a number of questions such as: what activities take place when an EP device moves through a medium like concrete or hard uniform rock? What equations govern the phenomenon?

Forthcoming is a short paper containing a mathematical description of the interaction between the EP device and the earth. Further, work on the packaging of the accelerometer will continue. Work will also continue in the characterizing of the accelerometer itself. In their paper, Yasukawa, et. al., [5], discuss the problem of simulation of circular silicon pressure sensors with a center boss. An investigation of the various designs of existing accelerometer would be of value and might generate an improved sensing device. Improved signal processing might be produced by the use of neural network, adaptive filters, time frequency analysis, hidden Markov processes and similar techniques. These techniques can be found in: Boashash and O'Shea [1]; Wilpon, Rabiner, Lee, and Goldman, [4] and Widrow and Stearns [3]. Further in the future it is planned to look at the physics of penetration in a uniform concrete-like medium.

REFERENCES.

1. Boashash, B., O'Shea, P., "A Methodology for Detection and Classification of Some Underwater Acoustic Signals Using Time-Frequency Analysis Techniques", IEEE Trans. on Acoustics, Speech and Signal Processing, Vol 38, No 11, 1990

2. Reid, S. R., Bell, W. W., Barr, R. A., "Structural Plastic Shock Model for One-Dimensional Ring Systems", Int. J. Impact Engng., Vol 1, No 2, 1983

3. Widrow, B., Stearns, S., Adaptive Signal Processing, Prentice Hall, 1985

4. Wilpon, J., Rabiner, L., Lee, C., Goldman, E.R., "Automatic Recognition of Keywords in Unconstrained Speech Using Hidden Markov Models", IEEE Trans. on Acoustics, Speech and Signal Processing, Vol 38, No 11, 1990

5. Yasukawa, A., Shimazoe, M., Matsuoka, Y., "Simulation of Circular Silicon Pressure Sensors with a Center Boss for Very Low Pressure Measurement", IEEE Trans. Electron Devices, Vol 36, No. 7, 1989

6. Young, C.W., "The Development of Empirical Equations for Predicting Depth of an Earth-Penetrating Projectile", SC-DR-67-60, Sandia Laboratories, Albuquerque, NM, 1967

APPENDIX.

Typical response curves of the Endve- accelerator using a compressed (tight) large mass float mount.

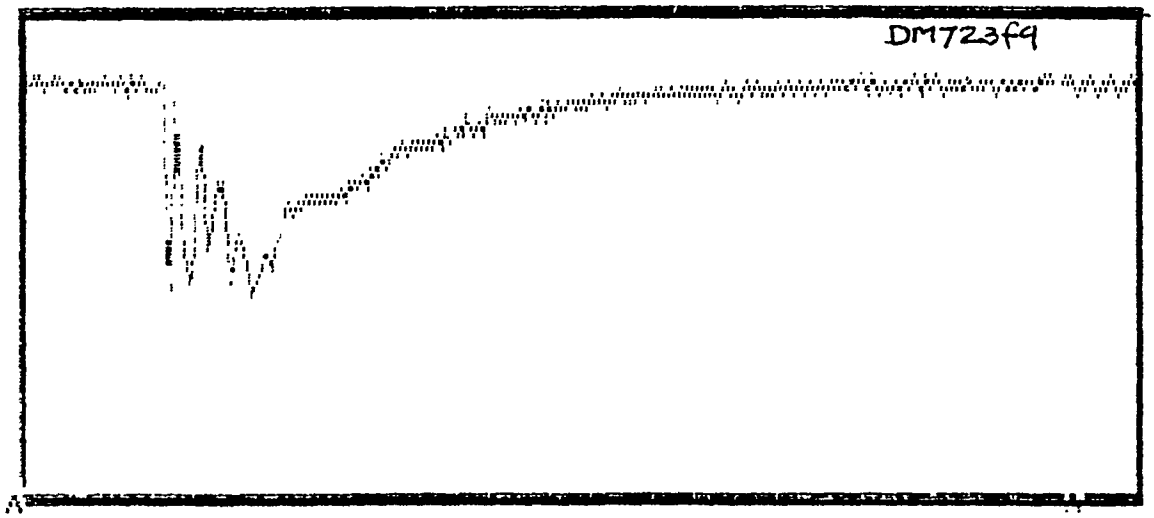


Figure A.1. Steel on foam response.

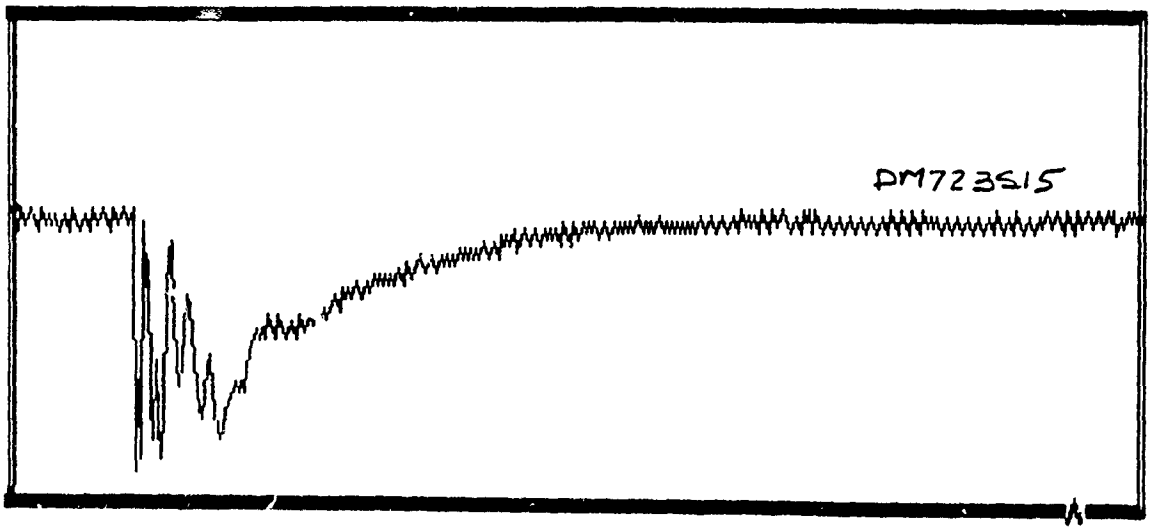


Figure A.2. Steel on steel response.

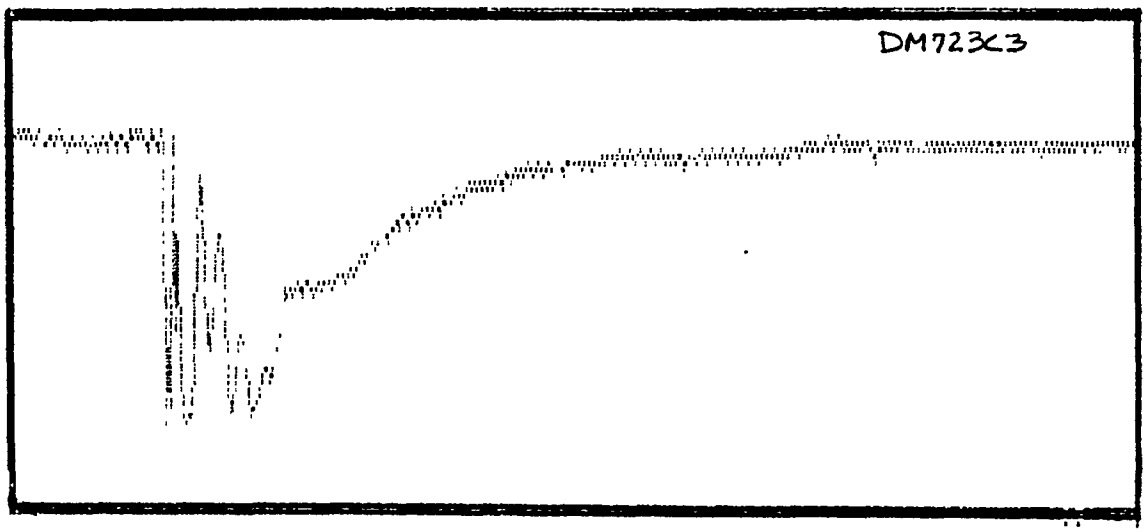


Figure A.3. Steel on concrete response.

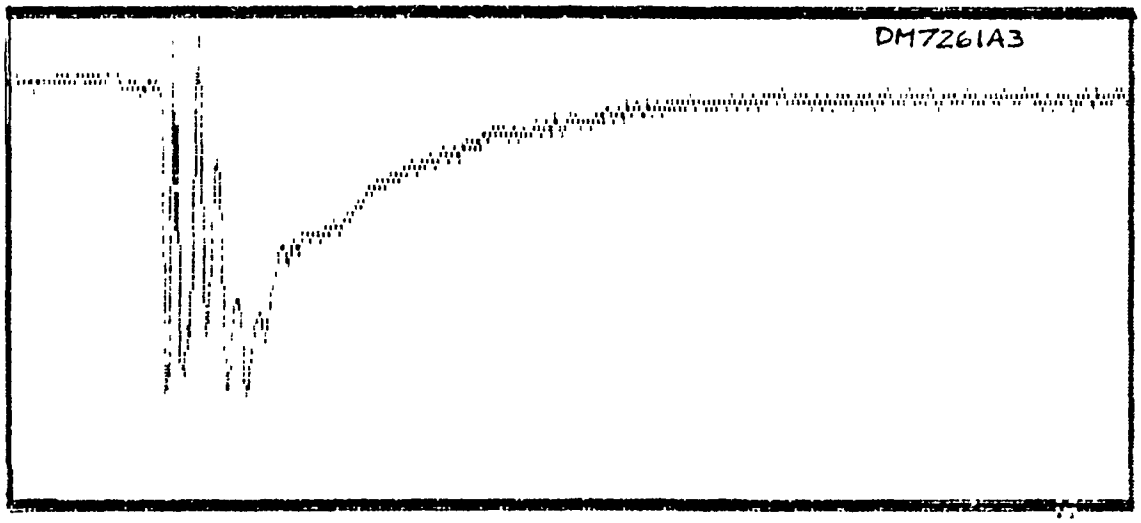


Figure A.4. Steel on aluminum response.

Multiresolution FLIR Image Analysis

Final Report

Dr. Raj Acharya

Electrical and Computer Engineering

State University of New York at Buffalo

Buffalo, NY 14260

Multiresolution FLIR Image Analysis

In this report, we have accomplished two specific tasks:

(1) **Development of Multiresolution Representations for FLIR Image Analysis using Nonlinear Scale Space.**

(2) **FLIR Image Analysis using a Nonlinear Multiresolution Representation.**

Hypothesis/Rationale

1. Multiresolution analysis of FLIR images is not only only optimal but also practical.
2. Nonlinear Scalespace (unlike Wavelet representation) provides a multiresolution representation without blurring the features.

Introduction

In this progress report we present some initial results on FLIR image analysis using nonlinear scalespace. We have proposed a nonlinear multiresolution approach to image analysis(1). We implement the multiresolution pyramid within the framework of Mathematical Morphology (2,4). We propose the use of ASF(Alternating Sequential Filters) for nonlinear scale space generation.

In the following section, we briefly review some important concepts of Mathematical Morphology.

Mathematical Morphology

Mathematical Morphology (2,3,5,6,7) is primarily a Set theory and uses set transformations for Image Analysis. It extracts the impact of a particular shape on images via the concept of Structuring Element (SE). The SE encodes the primitive shape information. In a discrete approach, the shape is described as a set of vectors referenced to a particular point, the Center, which does not necessarily belong to the SE. During Morphological transformation, the Center scans the whole image and matching shape information is used to define the transformation. The transformed image thus, is a function of the SE distribution in the original image.

The two most fundamental transforms of Mathematical Morphology are *EROSION* and *DILATION*. Let $P(E)$ be the Set of all subsets $X \in E$. Consider an arbitrary space(or Set) E. With each point X of the space E, a spatially varying set $B(X)$ called the SE is associated. The set $X \in P(E)$ can be modified based on set transformation of X by E. Let B_x denote the translate of B by the vector x.

EROSION : $\{X : B_x \subset X\}$

The eroded Set of X is the locus of centers x of translate B_x included in the set X. This transformation looks like the classical Minkowski Subtraction $X \ominus B$ of set X by set B:

$$X \ominus B = \bigcap_{b \in B} X_b$$

DILATION

Dilation is a dual transform of erosion and can be expressed as:

$$X^C \oplus B = (X \ominus B)^C$$

where $\bigoplus$ denotes dilation.

Let $\check{B}$ be the transposed set of B, i.e., the reflected set of B with respect to origin. Then the dilate $X \bigoplus \check{B}$ is the locus of the centers of B_x which hit (denoted by $\hat{\|}$) the set X.

$$X \oplus \check{B} = \{x : B_x \bigcap X \neq \phi\} = \{x : B_x \hat{\|} X\}$$

Dilation is related to Minkowski addition of Integral Geometry,

$$X \oplus \check{B} = \bigcup_{y \in \check{B}} X_y$$

The next two important operators are obtained by a suitable combination of erosion and dilation. The opening X with respect to B is defined as:

$$X_B = (X \ominus B) \oplus \check{B}$$

The dual transform closing of X by B is similarly defined as:

$$X^B = \left[(X^C)_B\right]^C = (X \oplus B) \ominus B$$

Grey level operators

Morphological operators have been extended for use with grey scale images (8). The concept of the *surface* of a set and its *umbra* allows us to solve the grey level problem via the binary one. A greyscale 2-D image can be viewed as a 3-D entity; two of the dimensions are the spatial co-ordinates of the image, and the third dimension is the intensity of the image. The intensity can be interpreted as the surface over the E^2 plane. The Umbra provides us

with the means to express the Surface as a Set, so that morphological operations can be performed with this set.

Let $B = \{B_r\}$, denote the structuring element. B is a set of vectors, with a grey value associated with each vector. Let $B(r)$ denote the grey value of the r^{th} vector. The Erosion $E_i(X, B)$ and Dilation $D_i(X, B)$ respectively of X by B can now be defined as:

$$E_i(X, B) = inf(X(i + r) - B(r)), r \in B$$

$$D_i(X, B) = sup(X(i + r) + B(r)), r \in B$$

Signed Morphological Operators for Edge Detection

The Morphological Operators proposed until now use positive elements in the definition of the structuring element. This has been due to the fact that, these operations have been conventionally used with pictures which are non-negative. We now propose the Signed Morphological Operators, as edge/surface detection operators.

Let the set of structuring elements be partitioned into $B = \{B_P, B_N\}$, where B_P represents nonnegative vectors and B_N represents negative vectors. Then, the Signed operators can be defined as:

$$SE(X, B) = inf(X(i + r) - B_P(r)) - sup(X(i + r) + B_N(r))$$

$$SD(X, B) = sup(X(i + r) + B_P(r)) - inf(X(i + r) - B_N(r))$$

where, SE and SD represent the Signed Erosion and Signed Dilation respectively. These Signed operators can be implemented via generalized operators for fast implementation.

Nonlinear Scalespace

Theorem Let C_λ (closing) and O_λ (opening) be a pair of primitives defined on a complete lattice ρ. The indexed sequence of filters

$$M_k(\lambda, \lambda') = m_{\lambda'} \cdots m_{\lambda+i2^{-k}(\lambda'-\lambda)} \cdots m_\lambda$$

with $\lambda' \geq \lambda \geq 0$, $m_\lambda = C_\lambda O_\lambda$ and $0 \leq i \leq 2k$ allows to define an operation

$$M = M_\lambda^{\lambda'} = \bigwedge_k M_k(\lambda, \lambda')$$

which is a morphological filter called an alternating sequential filter of primitives C (closing) and O (opening) and with bounds λ and λ'.

Nonlinear Scalespace Edges

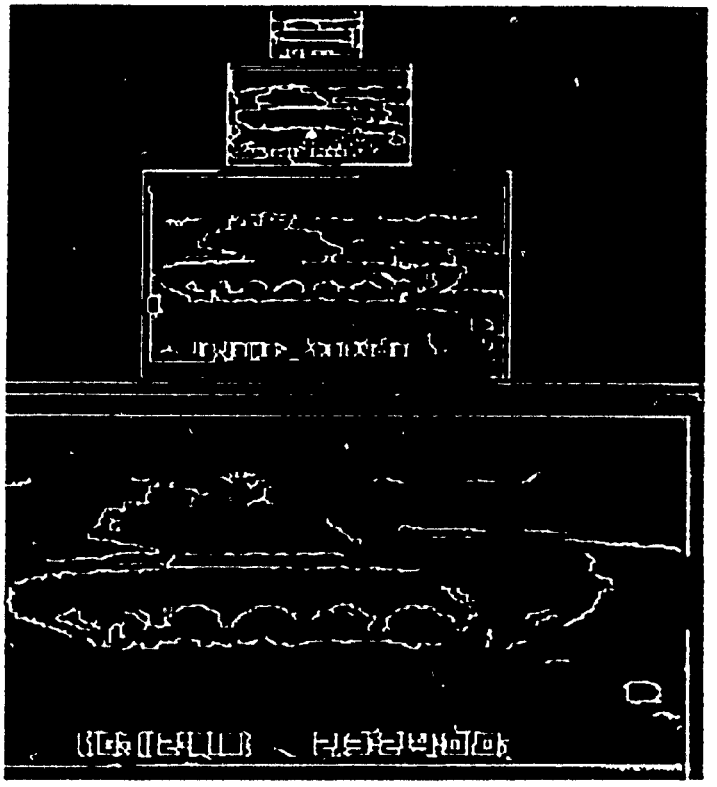


Fig 1

Linear Scalespace Edges

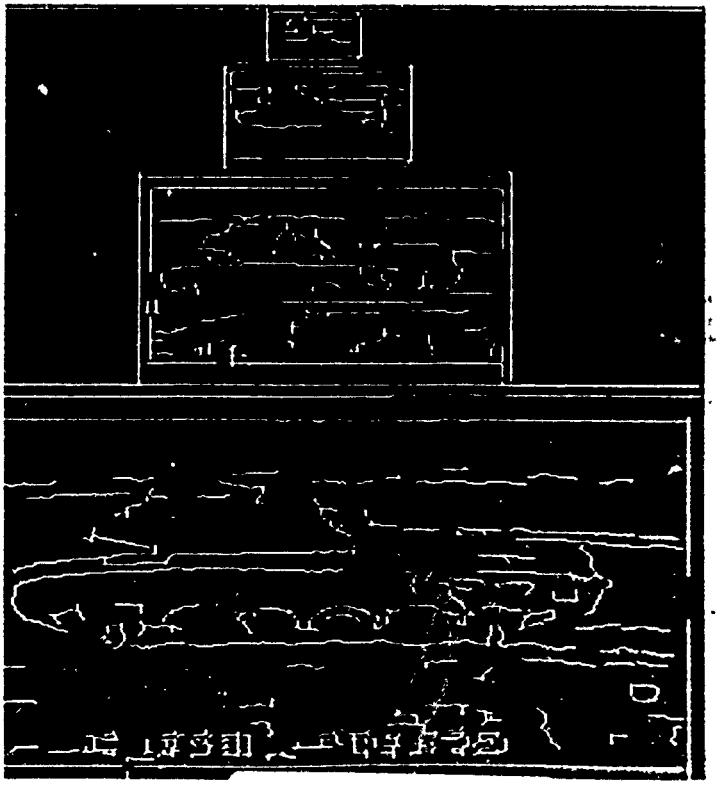
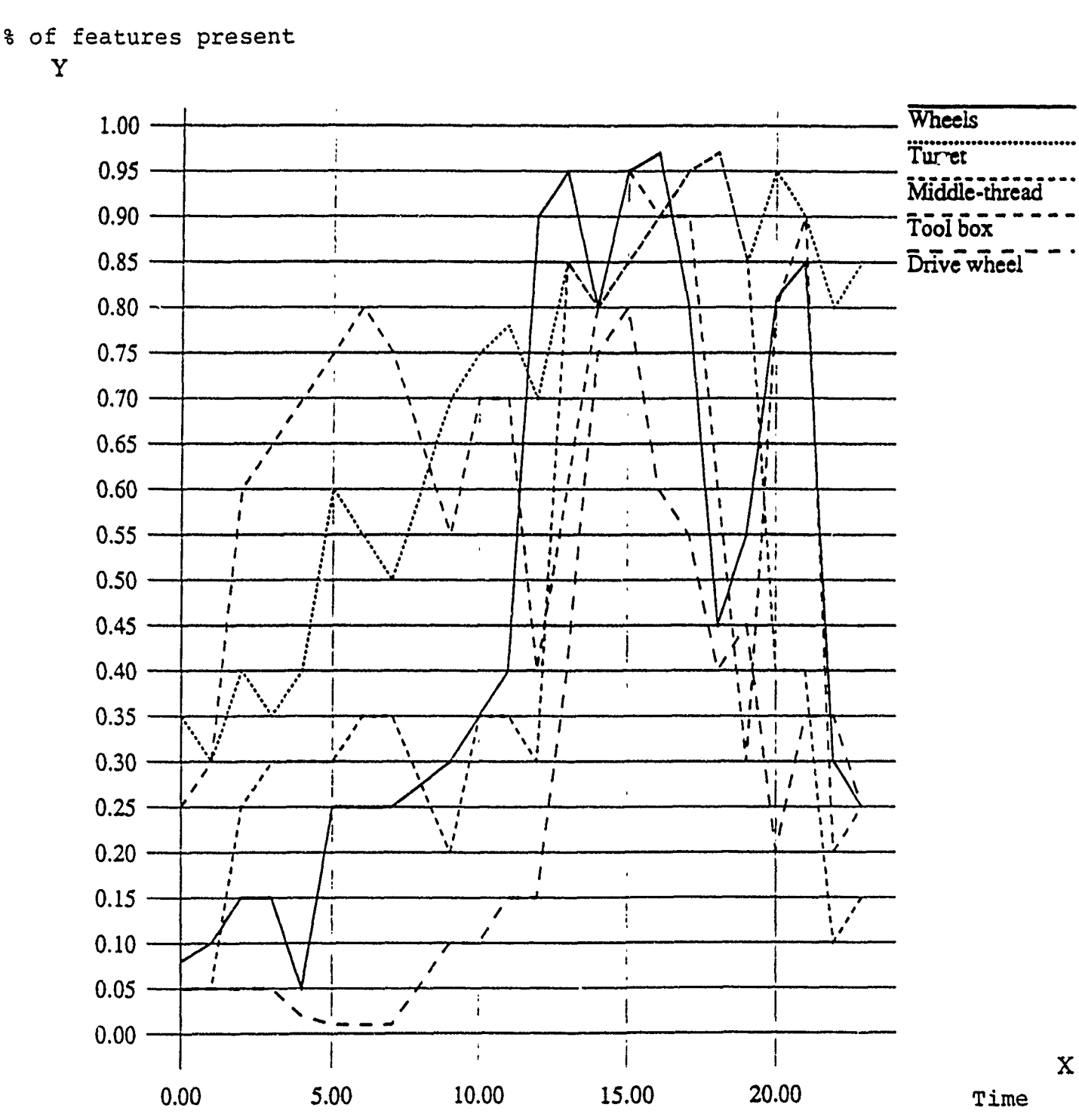


Fig 2

Scale = 0



Features vs Time_Scale 0

Fig 3

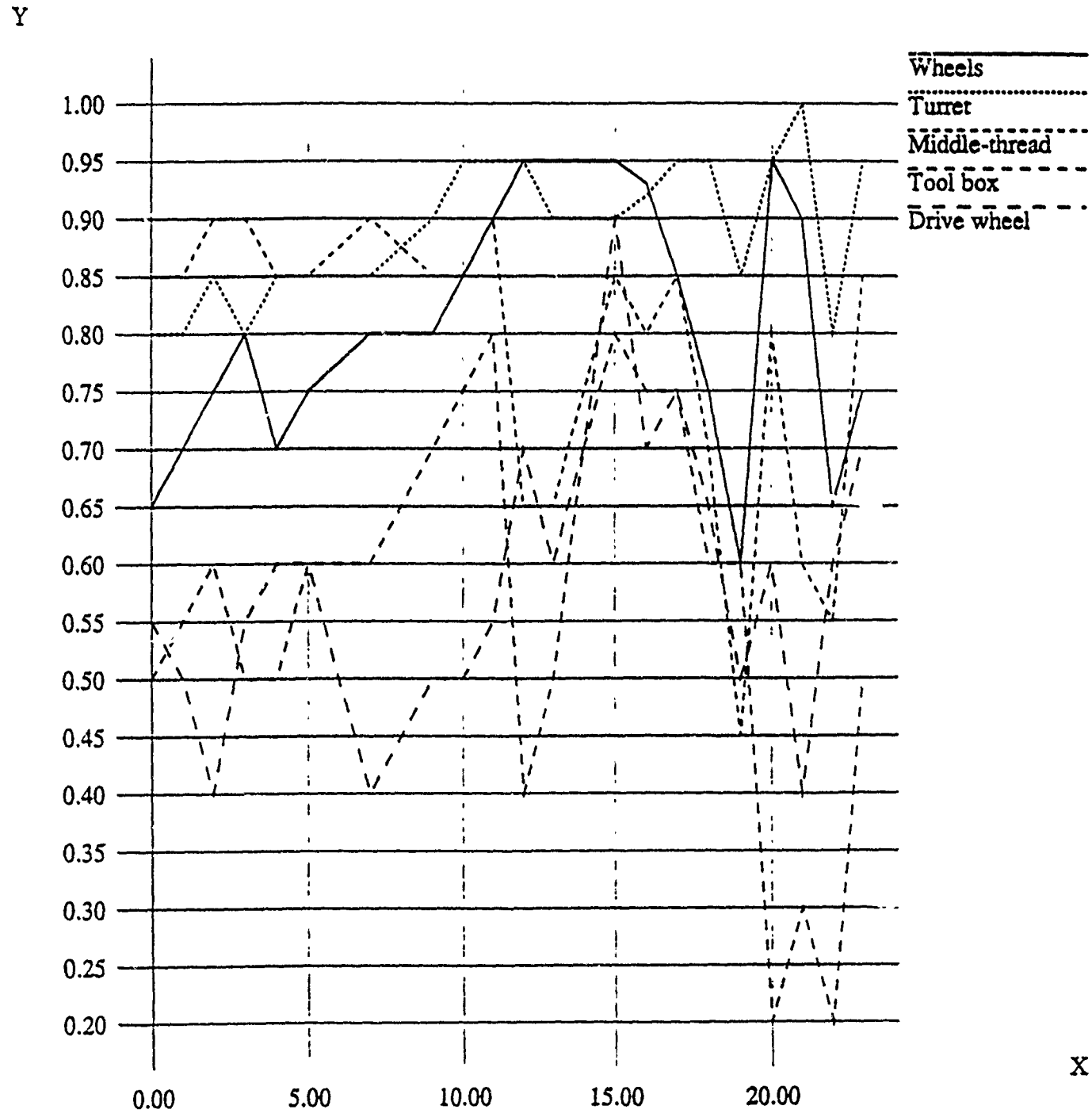

Features vs Time_scale 1
Y
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.00
5.00
10.00
15.00
20.00
X
Wheels
Turret
Middle-thread
Tool box
Drive wheel

Fig 4

Features vs Time-scale 2

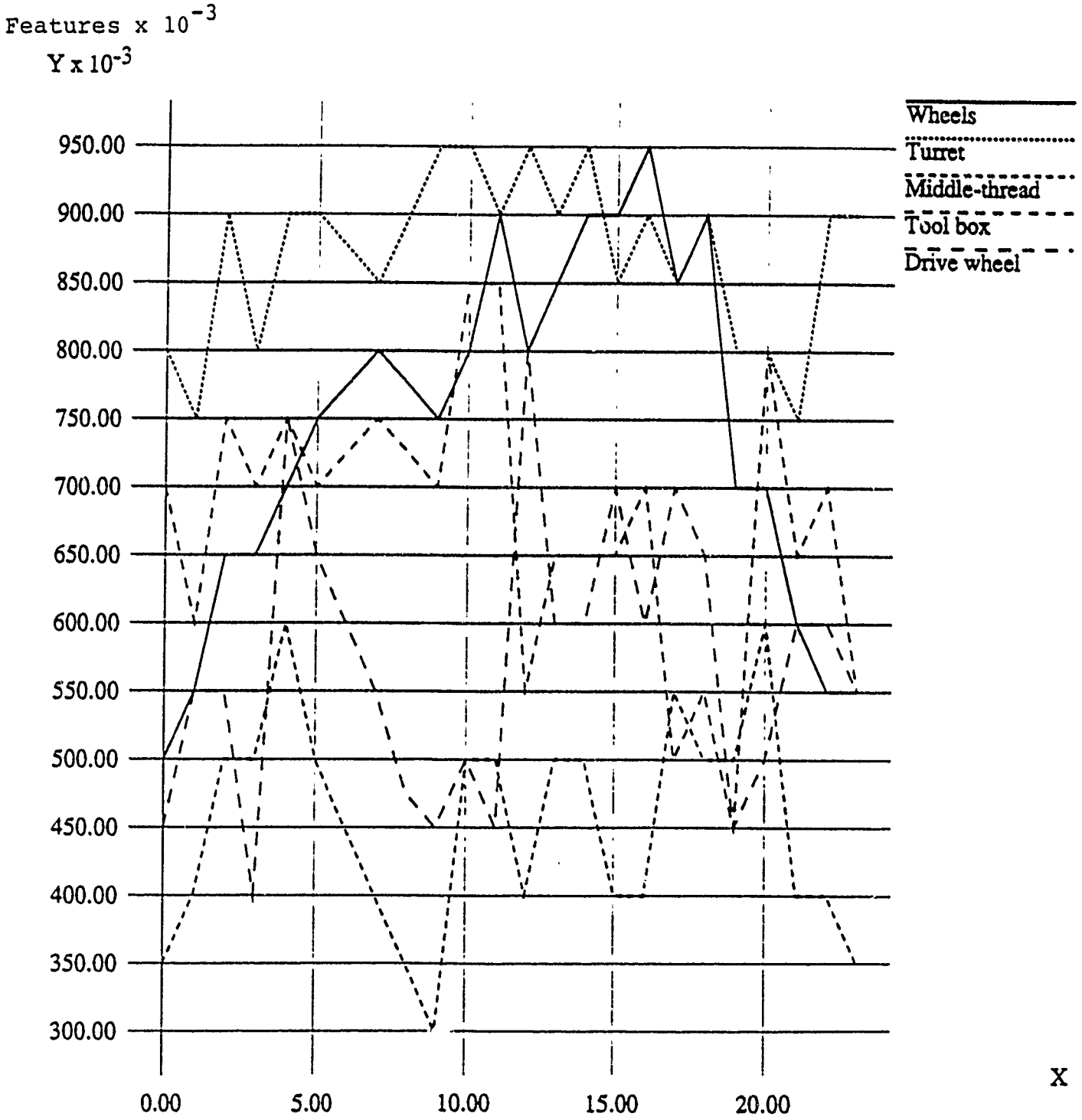


Fig 5

Alternating sequential filters are less severe and introduce less distortions when a small structuring element is used before a larger one. Generally speaking, these filters can be categorized as low pass filters. The properties of alternating sequential filters which include progressivity, semigroup, measure of symmetry "shape sense" are excellent and can be found in [7].

Multiresolution Feature Analysis

We study the diurnal changes of detectable features of tactile targets in infrared images at different resolutions. Twenty four images of tank T3, used in the experiments have been obtained at each hour starting at 8.24pm, 12th October, 1988.

The images were preprocessed using histogram equalization technique [11]. Images at reduced resolutions were derived using morphological scale space[1].

Fig 1 shows Nonlinear Scalespace edges while, Fig 2 shows the Linear (Wavelet based) Scalespace edges.

The desired features in the Tank image were wheels, turret, middle thread, tool box, and drive wheel. These features were identified by detecting their boundary edge contours. After thinning, edges were partitioned into edge segments at the intersection points and the corner points. Points where the curvature of the contour changes considerably were marked as corner points[13].

The uncertainty about the features were determined as a fractal measure of boundary edges [14]. Fractal measure of a boundary contour is the percentage of the edge points detected in the contour. The graphs(Figs 3,4,5) show this measure for different features against time at three different resolutions : 1/2, 1/4 and 1/8, of the original resolution. Although these percentage values do not express an absolute measure of uncertainty of features, they are useful in comparing uncertainties of different features and their variations with time.

The graphs of edge fractal measures, or *diurnal thermal signatures*, show the thermal behavior of the materials of the features. Most of the features observed have shown similar thermal properties and have appeared in a constant manner over the day. The invariant behavior of these features are due to their high thermal mass and can be used to distinguish the target from other surrounding objects and terrain patterns. *We believe that diurnal thermal signatures may be successfully used in the detection of tactile targets.*

Since high resolution signals are often affected with noise and redundant information, it is important to examine thermal signatures at different resolutions. We adhere to morphological transforms to obtain the multi scale representation to avoid feature blurring in linear scale space. At low resolution, the signatures display similar variations, while some of the features, such as wheels and the turret, are more prominent than others. **We conclude that confidence of the target detection process can be improved by utilizing the thermal signatures of features that remain significant in scale space.**

References

[1] A. Morales and R. Acharya, An Image Pyramid with Morphological Operators ". *Transactions on PAMI* (In press). (Short version also in CVPR-91).

[2] C. Giardina and E. Dougherty, *Morphological Methods in Image and Signal Processing*, New Jersey, Prentice Hall 1988.

[3] R. M. Haralick, S. R. Sternberg, and X. Zhuang, " Image Analysis Using Mathematical Morphology", *IEEE Transactions on Pattern Analysis and Machine Intelligence* , vol. PAMI-9, July 1987.

[4] R. Haralick, J. Lee, C. Lin, and X. Zhuang, " Multi-resolution Morphology" . *Proc. IEEE First conference on Computer Vision.* London 1987.

[5] P. Maragos and R. W. Schaffer, " Morphological Filters-Part II: Their Relations to Median, Order Statistics, and Stack Filters", *IEEE Transactions on Acoustics, Speech and Signal Processing*", Vol. ASSP-35 , August 1987.

[6] J. Serra, *Image Analysis and Mathematical Morphology*, New York, Academic Press, 1982.

[7] J. Serra, *Image Analysis and Mathematical Morphology*, Vol. 2, New York, Academic Press, 1988.

[8] S. Sternberg, " Grey Scale Morphology", *Computer Vision, Graphics and Image Processing*, Vol. 35, 1986.

[10] A. Toet. " A Morphological Pyramid Image Decomposition". *Pattern Recognition Letters 9.* May 1989.

[11]. Jain, Anil K., *Fundermentals of Image Processing,* Prentice Hall, Englewood Cliffs, NJ 07632.

[12]. Mallat, Stephane G, " A Theory of Multiresolution Signal Decomposition, " *IEEE Transaction of Pattern Analysis and Machine Intelligence,* Vol 11, No. 7, July 1989.

[13]. Freeman, H., and Davis, L., " A corner Finding Algorithm For Chain-Coded Curves, " *IEEE Transaction of Computers,* March, 1977.

[14]. Stephanou, H. E., and Erkman, A. M., " Evidential Classification of Dexterous Graphs for the Integration of Perception and Action, " *Journal of Robotics Systems,* vol. 5, no. 4, pp. 309-336, 1988.

APPENDIX

(ON FILE AT RDL)

AUTOMATIC SEGMENTATION OF INFRARED IMAGES

Dr. D.V. Satish Chandra

ABSTRACT

Image segmentation is an important step in any image processing or target recognition system. The motivation behind this research is to investigate if the performance of the segmentation algorithms can be improved by setting the parameters of the segmentation algorithms depending on the global image and target characteristics such as contextual scene information, time of the day, weather conditions etc. This report presents the performance of a segmentation algorithm based on probabilistic relaxation on several infrared tank images taken at various times of the day.

I. INTRODUCTION

In the early 1980's Low Altitude Navigation and Targeting Infrared for Night (LANTRIN) program was initiated to develop a capability for an airborne forward-looking infrared (FLIR) system which would automatically detect and locate targets in an area on the ground known to contain targets. The segmentation operation, which extracts the target from the background, is an important step in the target recognition process. The main problem with segmentation algorithms is the degradation in the performance on a set of imagery different from the ones used in the initial development of the algorithms.

Infrared (IR) images are obtained by sensing the thermal radiations emitted or reflected by the target in the infrared spectrum. Time of day and weather conditions greatly affect the quality of the FLIR images. The atmospheric scatter and absorption of thermal radiations result in the loss of contrast between target boundary and its background. Therefore, in order to achieve good segmentation performance it is essential that the

segmentation algorithms should take into account the specific characteristics of the FLIR images which depend on the time of day, weather and atmospheric conditions. In the following sections a segmentation algorithm based on probabilistic relaxation and several variations of the basic algorithm are presented. The results of the segmentation algorithm whose parameters are adjusted based on the gray level and gradient models of the image are also given.

II. RELAXATION LABELING

Relaxation is an iterative approach for using contextual information to reduce local ambiguities. Suppose that there exists a set of objects A = $\{a_1, a_2, \ldots, a_N\}$, each of which can belong to one of M classes $w_1, w_2, \ldots, w_M$. In image segmentation problem, the set of objects in general is the set of pixels in the image and classes correspond to target and background classes. In probabilistic relaxation, the probabilities $P_i(T)$ and $P_i(B)$ that pixel i belongs to the target class and the background class respectively, are estimated on the basis of the gray level or a set of local measurements. These probabilities satisfy the condition

$$P_i(T) + P_i(B) = 1 \qquad (1)$$

The pixel probabilities are updated iteratively based upon initial probabilities and a set of measurements made on a local neighborhood. When the process is iterated several times the contrast between the object and background increases, so that segmentation becomes trivial. However, the performance of the relaxation method, in general, depends on the initial labeling of pixel probabilities and the iterative updating process. These steps are discussed in detail in the following sections.

Assignment of Pixel Probabilities:

The initial assignment of probabilities to every pixel is very important since it affects the convergence rate and the results of segmentation. The simplest way to assign initial probabilities is to

define

$$P_i(T) = \frac{I(i)}{(G-1)} \tag{2}$$

Where $P_i(T)$ denotes the probability that the ith pixel belongs to the target class, I(i) is the gray level of the ith pixel, and G is the number of possible gray levels.

Updating Rule:

There are many possible iterative schemes that can be used to update the pixel probabilities based upon initial probabilities and compatibility coefficients. A simple updating rule used by Rosenfeld et al. [ ] has the form

$$P_i^*(T) = \frac{P_i(T)\,[1+q_i(T)]}{P_i(T)\,[1+q_i(T)] + P_i(B)\,[1+q_i(B)]} \tag{3}$$

Where the update factor $q_i(T)$ is given by

$$q_i(T) = \frac{1}{N_i} \sum_{j \in N_i} [c(i,T;j,T)\ P_j(T) + c(i,T;j,B)\ P_j(B)] \tag{4}$$

Here the compatibility coefficient c(i,T;j,B) represents the quantitative measure of the class assignment I(i) $\in$ T and I(j) $\in$ B and N_i is some local neighborhood of pixel i. The compatibility coefficients for the updating rule (3) always lie in the fixed range [-1,1], with positive values representing compatibility and negative values representing incompatibility of the class assignment. Nonnegative compatibility coefficients, with 0 representing high incompatibility and high values representing high compatibility, can be used instead if the pixel probabilities are updated by the following rule:

$$P_i^* = \frac{P_i(T)\ q_i(T)}{P_i(T)\ q_i(T) + P_i(B)\ q_i(B)} \tag{5}$$

Pixel Neighborhood:

The pixel neighborhood, N_i, used in the updating process generally

consists of the ith pixel and its eight nearest neighbors. In this research, an adaptive neighborhood consisting of the five nearest pixels which are closest in gray value to the pixel under consideration is also used in computing the update factor. Fig. 1 shows the basic structure of the automatic segmentation system.

III. RESULTS

The test images used in the segmentation experiment consist of twelve FLIR images of a military tank taken at two-hour intervals of the day. The mean and the standard deviation of the gray values of these images plotted as a function of the time of the day are shown in Fig. 2. Even though the histograms of most of the test images are bimodal, the valley between the two peaks in any histogram does not correspond to the border between the target and background. The reason for this is that the contrast between the ground and sky regions is higher than that between the target and background regions. However, thresholds can be selected manually from the histograms to obtain the best possible segmentation of the targets and are shown in Fig. 2. Note that the large variation in threshold values makes it difficult to automate the threshold selection process. The threshold variation can be reduced to some extent by stretching the histogram after centering the mean pixel value on the gray scale and is shown in Fig. 3. However, it is not always possible to segment the target from the background by selecting the threshold in this manner without additional processing.

Threshold selection process can be automated by using a reliable segmentation scheme such as relaxation labeling described in the previous section. In relaxation labeling approach the initially ambiguous "light" and "dark" probabilities converge to virtual certainties as the process is iterated making the threshold selection trivial. The target and background pixels of the test images lie predominantly in the dark half, and as a result the gray levels of both target and background pixels would be

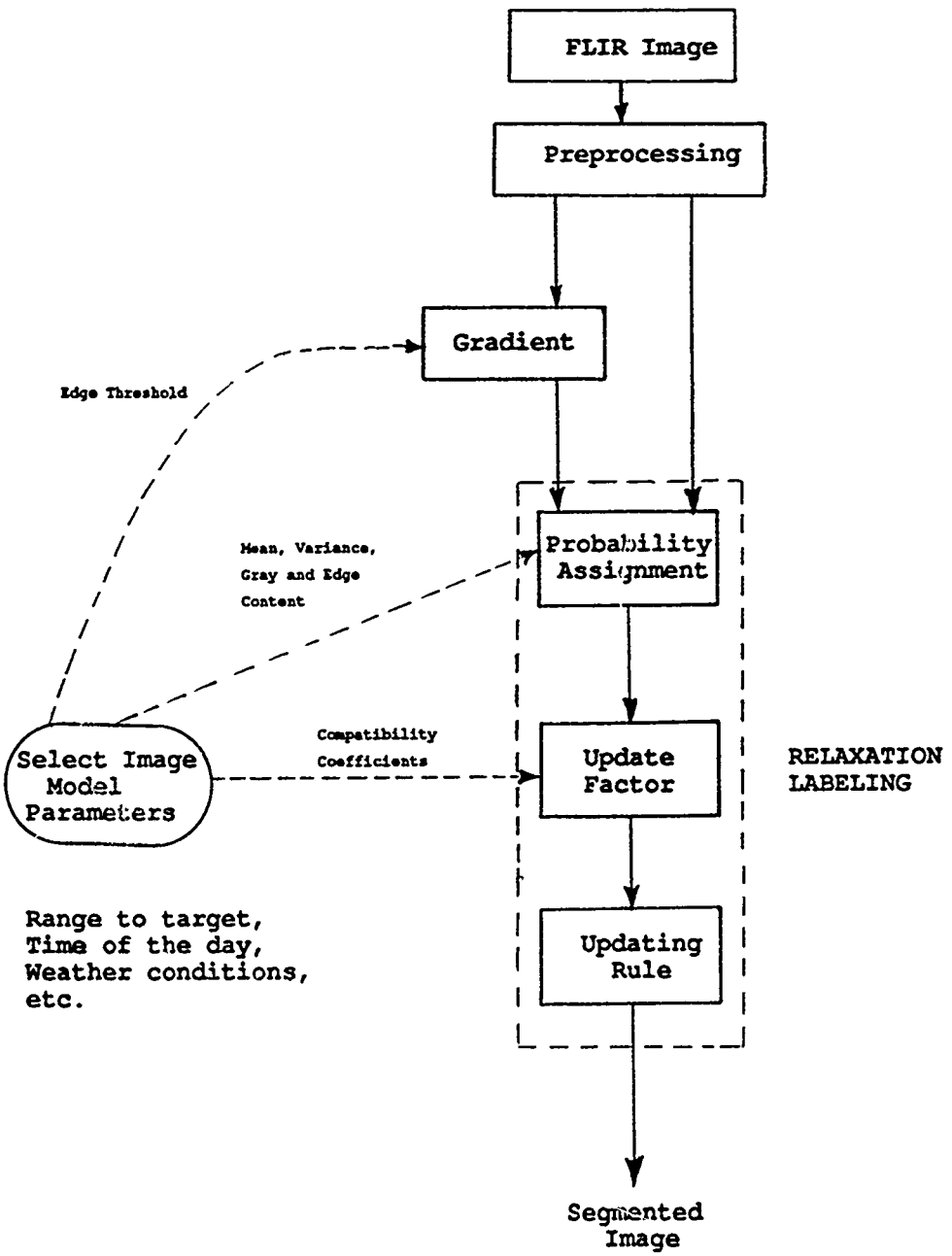


Figure 1. Automatic segmentation system.

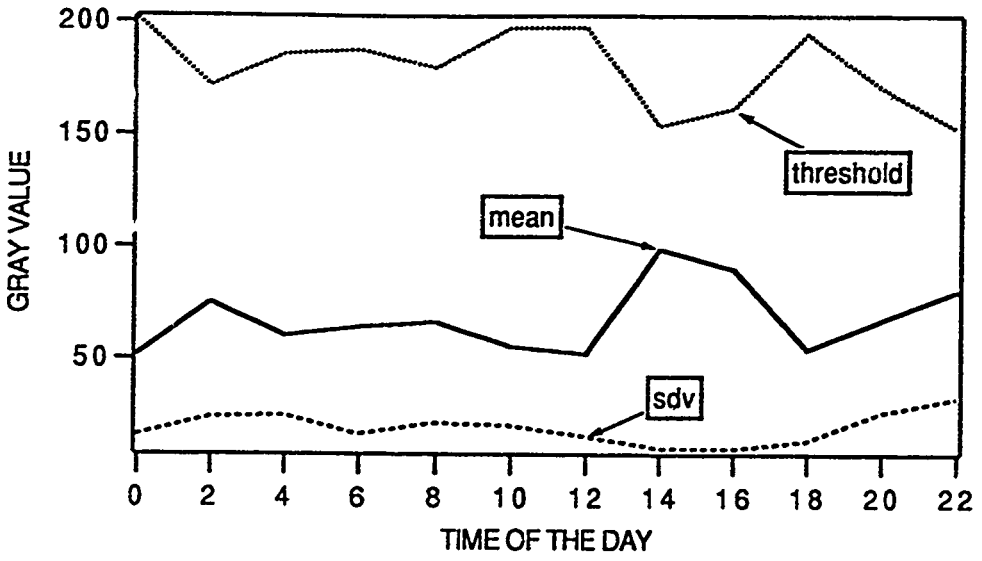


Figure 2. Mean and standard deviation of pixel values of test images.

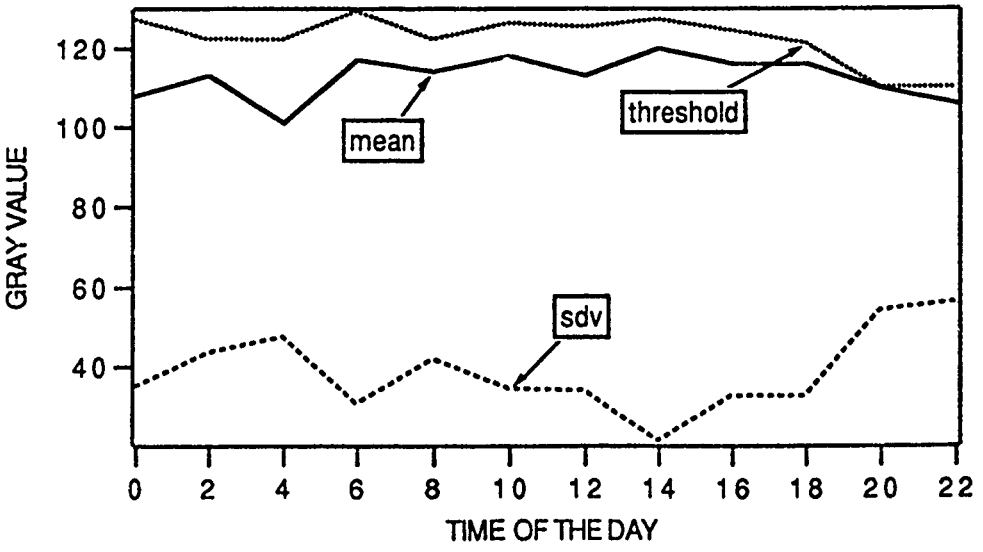


Figure 3. Mean and standard deviation of pixel values of test images after histogram stretching.

driven toward the dark end of the gray scale by the iterative process if the initial probability assignment is made according to (2). In order to avoid this, an alternate initialization scheme that stretches the histogram after centering the mean gray value on the gray scale as given by the following equation can be used.

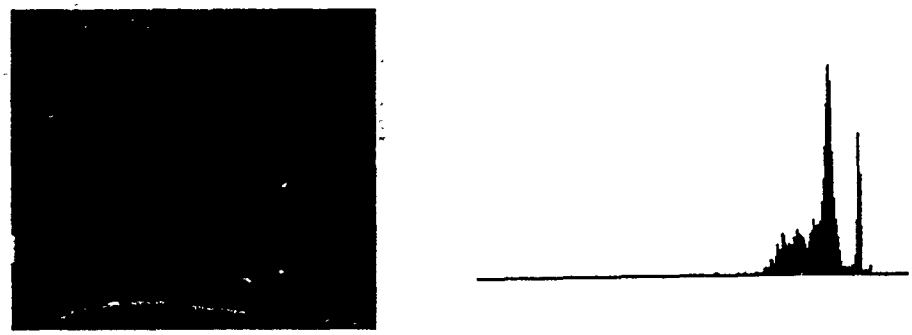
$$P_i(T) = \frac{1}{2} + \frac{I(i) - \mu_I}{X} \tag{6}$$

where

$$X = \begin{cases} 2(I_{high} - \mu_I), & I(i) \geq \mu_I \\ 2(\mu_I - I_{low}), & I(i) < \mu_I \end{cases} \tag{7}$$

and I_{high}, I_{low} and μ_I are the high, low and mean intensity values respectively, computed over the entire image. It was observed that the variations in the choice of the compatibility coefficients affect the convergence rate, but has very little effect on the quality of the segmentation result. The entire simulation experiment was carried out using rule (5) to update the probabilities, with the compatibility coefficients c(i,T;j,T) and c(i,T;j,B) set at 1 and 0, respectively. The performance of the relaxation algorithm described above is evaluated using the test images shown in Fig. 4. In order to improve the segmentation result at the target boundary, an adaptive neighborhood, consisting of five neighborhood pixels which are closest in gray value to the pixel under consideration, is used in the computation of the update factor. The segmentation results on the tank image of Fig. 4(a) are shown in Fig. 5. It can be seen from the histogram plots that after a few iterations, histogram becomes bimodal separating the two classes making the threshold selection (chosen at the center of the gray scale) trivial.

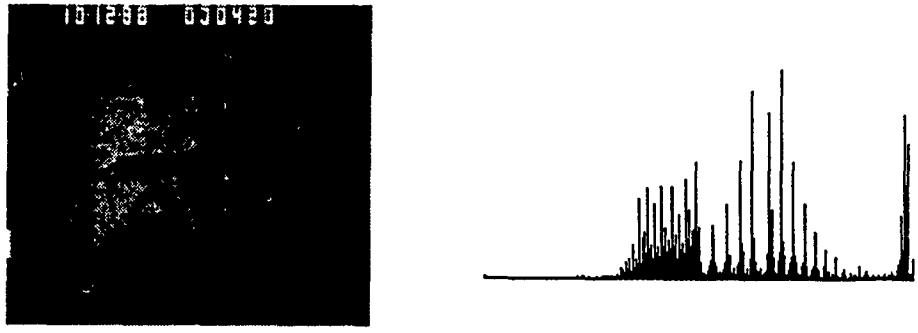
Another method of improving the segmentation performance at the target boundary is to use adaptive smoothing [11] in the computation of the update factor. The adaptive smoothing operation adjusts the weights of the convolution mask based on the gradient so as to avoid smoothing at the



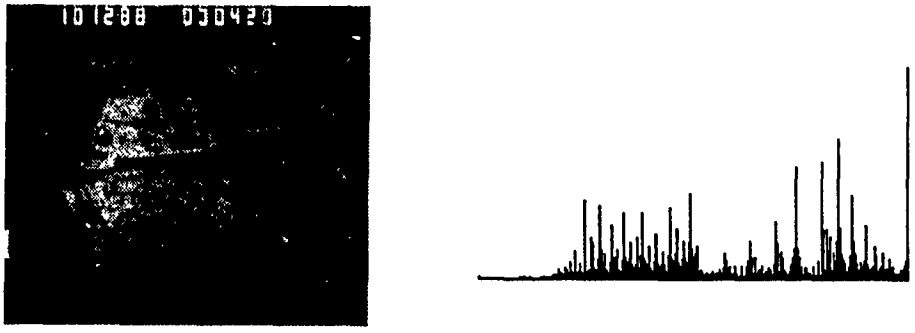
(a) Tank image taken at the 0th hour of the day.

(b) Tank image taken at the 20th hour of the day.

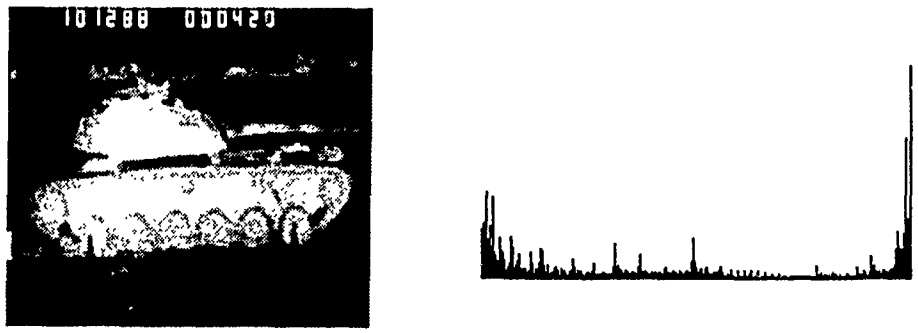
Figure 4. Test images and their histograms.



(a) Iteration 1



(b) Iteration 2



(c) Iteration 4

Figure 5. Results of applying the relaxation processs to test image of Fig. 4(a).

discontinuities. Segmentation results on the tank image of Fig. 4(a) after four iterations of the relaxation process when a fixed eight-pixel neighborhood, five-pixel adaptive neighborhood, and adaptive smoothing are used in the computation of the update factor are shown in Fig. 6. The results indicate that the boundary of the segmented target is sharper when either adaptive neighborhood or adaptive smoothing is used.

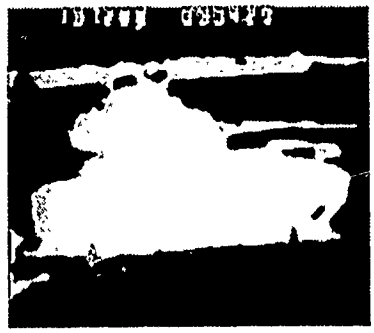
The performance of the relaxation algorithm can be further improved by incorporating the standard deviation of the gray values in the assignment of initial pixel probabilities. Segmentation results for the image of Fig. 4(a), when the initial probability assignment is made according to (6) with X replaced by $X\sigma_I$, where σ_I is the standard deviation of the gray values, are shown in Fig. 7. The results indicate that the incorporation of standard deviation information increases the rate of convergence of the algorithm.

The results of segmentation on the tank image of Fig. 4(b) using the procedure described above are shown in Fig. 8. Observe that it is not possible to segment the target from the background even after the convergence of the algorithm. This is due to the fact that the contrast between the target and the background is very low at this time of the day. However, if the edge information is incorporated in addition to the gray level information into the relaxation process the performance can be improved. This can easily be done by modifying the initial assignment of pixel probabilities as shown below.

$$P_i(T) = \frac{1}{2} + A\left(\frac{I(i)-\mu_I}{X}\right) + B\left(\frac{(E(i)-\mu_E}{Y}\right) \tag{8}$$

Where

$$X = \begin{cases} I_{high}-\mu_I, & I(i) \geq \mu_I \\ \mu_I - I_{low}, & I(i) < \mu_I \end{cases} \tag{9}$$

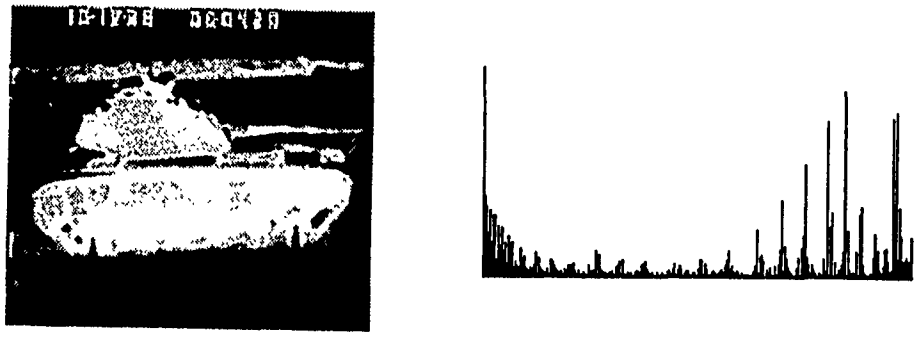


(a) Fixed eight-pixel neighborhood

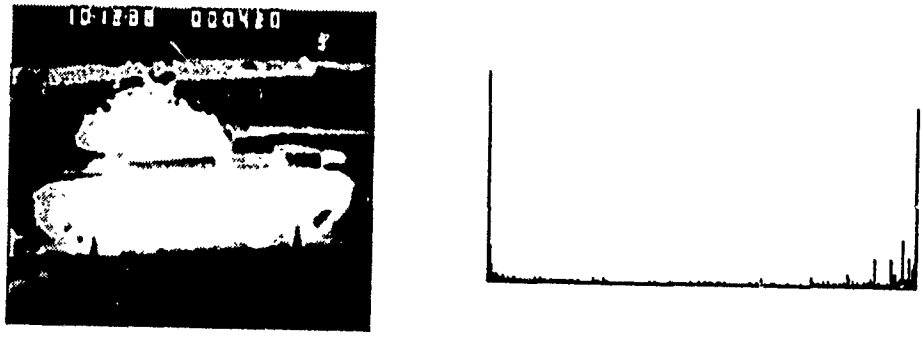
(b) Eight-pixel adaptive smoothing

(c) Five-pixel adaptive neighborhood

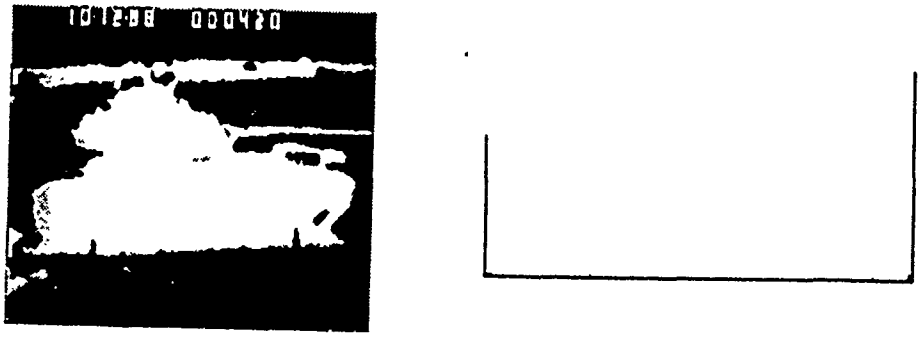
Figure 6. Results (after four iterations) of applying the relaxation processs, with different update factors, to test image of Fig. 4(a).



(a) Iteration 1

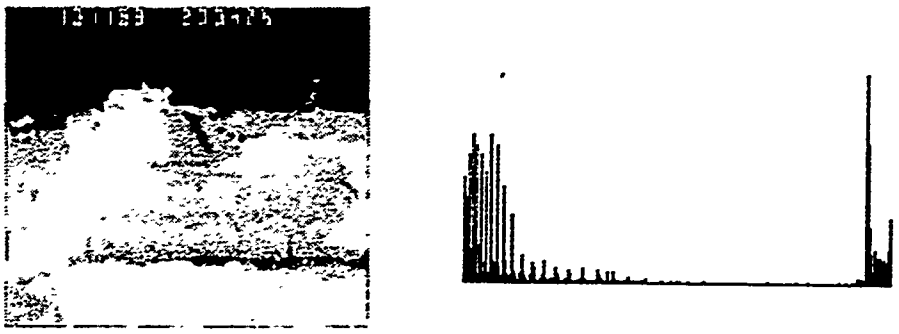


(b) Iteration 2

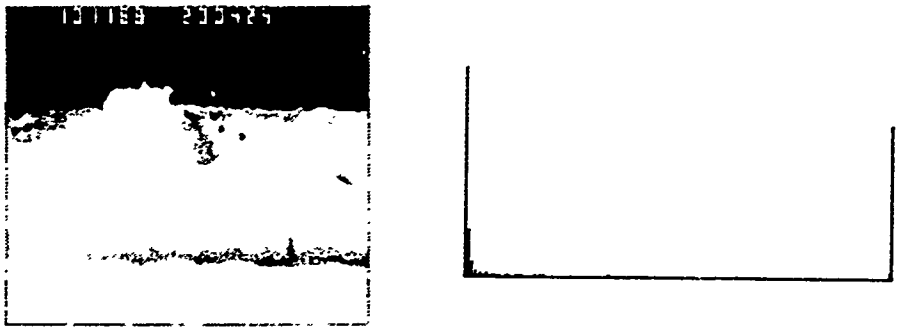


(c) Iteration 4

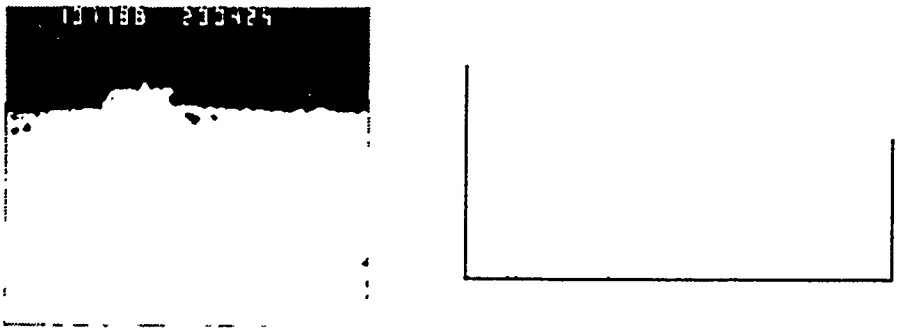
Figure 7. Results of applying the relaxation processs to test image of Fig. 4(a), when the initial assignment of pixel probabilities is a function of the standard deviation of pixel values.



(a) Iteration 1



(b) Iteration 2



(c) Iteration 4

Figure 8. Results of applying the relaxation processs tc test image of Fig. 4(b).

$$Y = \begin{cases} E_{high} - \mu_E, & E(i) \geq \mu_E \\ \mu_E - E_{low}, & E(i) < \mu_E \end{cases} \quad (10)$$

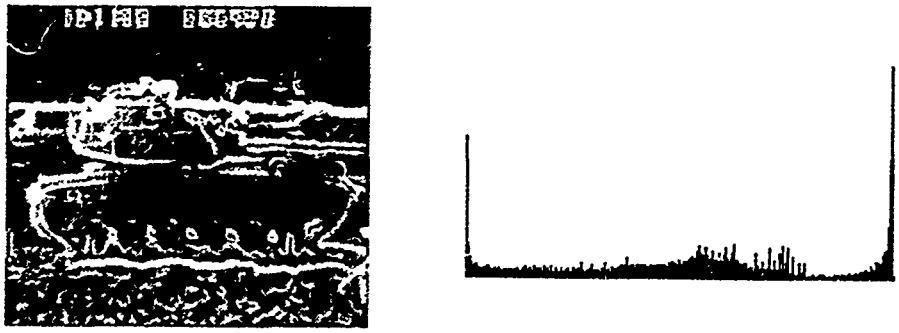
E(i) is the Sobel edge magnitude of the ith pixel, E_{high}, E_{low}, and μ_E represent the high, low, and mean edge values over the entire image, and A and B are factors which control the gray and edge value components in the initial assignment of pixel probabilities. The values for A and B are determined based on the target/background contrast, variance of gray and edge values, time of the day, etc. The results of segmentation on image of Fig. 4(b) with initial probabilities assigned according to (8) are shown in Fig. 9. Here, the parameters A and B are taken as a constant multiplied by the inverse of the standard deviation of gray value and edge magnitude, respectively.

IV. SUMMARY

The simulation results indicate that the performance of segmentation algorithms can be improved by tuning a set of algorithm parameters based on image models. The segmentation algorithm described in this report uses the gray level and the Sobel-gradient models of the image. The contrast between the target and background in a FLIR image is greatly dependent on the weather conditions and time of the day at which the image is taken. It is believed that the segmentation performance can be improved further if time of day information can be used to adjust some of the algorithm parameters. For example, time of day information can be used to determine the gray value and edge content in the initial assignment of pixel probabilities or to control the degree of smoothing at each iteration of the relaxation process.

It was also observed that the effectiveness and accuracy of the relaxation technique in segmenting images depend on the choice of the neighborhood and the method used in computing the updating factor. Incorporation of adaptive smoothing into the computation of the updating factor resulted in an improved segmentation performance at the target

(a) Iteration 1



(b) Iteration 2

(c) Iteration 4

Figure 9. Results of applying the relaxation processs, using both gray and edge values in the assignment of initial probabilities, to test image of Fig. 4(a).

boundary.

Also, due to significant differences in target surface temperatures and the statistical changes the target goes through during the course of a day, the target surface may have more than one homogeneous region depending on the time of the day. It may therefore be more appropriate to regard the segmentation problem as a multiclass problem instead of a two class problem and to segment the target into several constituent parts. The number of classes to be used in the segmentation algorithm is a function of the time of day and can be determined by having some knowledge of the thermal characteristics of the target. Since the segmented target may have more than one region, segmentation and recognition should be treated as an integral operation and implemented as a rule based system.

ACKNOWLEDGEMENTS

I wish to thank the Air Force Systems Command and the Air Force Office of Scientific Research for sponsorship of this research and the Research Development Laboratories for their concern and help to me in all the administrative and directional aspects of this program. I am sincerely grateful to my Effort Focal Point Mr. Ed Zelnio, who suggested the research problem. I appreciate the support of Mr. Vince Velten and Mr. Mike Bryant. Finally, all the members of the Automatic Target Recognition group made my stay at Wright Patterson Air Force Base a memorable one.

REFERENCES

[1] B.J. Schachter, "A Survey and Evaluation of FLIR Target Detection/Segmentation Algorithms," Proc. DARPA Image Understanding Workshop, pp. 49-57, Sept. 1982.

[2] J.B. Jordan and G.M. Flachs, "Statistical Segmentation of FLIR Images," Proc. SPIE, vol. 754, pp. 220-228, 1987.

[3] J.P. Gambotto, "Segmentation and Interpretation of Infrared Image Sequences," in Advances in Computer Vision and Image Processing,

T.S. Huang (Ed.), Volume 3, JAI Press Ltd., Connecticut, 1988.

[4] A. Rosenfeld, R. Hummel, and S. Zucker, "Scene Labeling by Relaxation Algorithms," IEEE Trans. Syst., Man, Cybern., vol. SMC-6, pp. 420-433, 1976.

[5] S. Peleg, "A New Probabilistic Relaxation Scheme," IEEE Trans. Pattern Anal. Machine Intell., vol. PAMI-2, pp. 362-369, July 1980.

[6] J.Y. Hsiao and A.A. Sawchuk, "Supervised Textured Image Segmentation Using Feature Smoothing and Probabilistic Relaxation Techniques, "IEEE Trans. Pattern Anal. Machine Intell., vol. PAMI-11, pp. 1279-1292, Dec. 1989.

[7] B. Bhanu and R.D. Holben, "Model-Based Segmentation of FLIR Images," IEEE Trans. Aerospace Electronic Systems, vol. AES-26, pp.2-10, Jan. 1990.

[8] R.M. Haralick and L.G. Shapiro, "Image Segmentation Techniques," Comput. Vision, Graph., Image Processing, vol. 29, no. 1, pp. 100-132, 1985.

[9] W.A. Yasnoff, J.K. Mui and J.W. Bacus, "Error Measures for Scene Segmentation," Pattern Recognition, vol. 9, pp. 217-231, 1977.

[10] D.J. Shazeer, "Performance Measures for Statistical Segmentation," Proc., SPIE, vol. 359, pp. 348-360, 1982.

[11] P. Saint-Marc, J-S. Chen and G. Medioni, "Adaptive Smoothing: General Tool for Early Vision," IEEE Trans. Pattern Anal. Machine Intell., vol. PAMI-13, pp. 514-529, June 1991.

[12] N. Nandhakumar and J.K. Aggarwal, "Integrated Analysis of Thermal and Visual Images for Scene Interpretation," IEEE Trans. Pattern. Anal. Machine Intell., vol. PAMI-10, pp. 469-481, July 1988.

Performance Evaluation of Rule Grouping Algorithm Running on the Activation Framework Architecture

Ing-Ray Chen and Bryant Poole*

August 9, 1991

Abstract

The objective of this research is to design and develop rule grouping algorithms for maximizing the performance of real-time rule-based systems running on the Activation Framework (AF) architecture [GRE 87]. This research involved the development of a formal theory for quantifying the estimation of performance metrics based on probabilistic parameter modeling and the design of a rule grouping algorithm based on Kernighan-Lin (KL) heuristic graph-partitioning for a single processor system. A demonstration system based on the theory and algorithm has been developed and tested on a portion of Advanced GPS Receiver (AGR) and Manned Maneuvering Unit (MMU) knowledge bases.

1 Introduction

The SKRAM toolset is an ongoing project [GRE 91A] that is designed to automate the development and testing of real-time expert systems running on the Activation Framework (AF) architecture. The AF architecture supports a dynamically prioritized, distributed, message based, object oriented, real-time Ada run-time environment. The user is allowed to specify independently scheduled groups of rules which are automatically converted into Ada Activation Framework Objects (AFOs). Unlike a conventional real-time scheduler which uses a static time-sliced scheduling mechanism to execute AFOs in a processor, the AF architecture uses an innovative event-sliced scheduling mechanism: the AFO priorities are changed dynamically by the arrival of events (e.g., messages sent from one AFO to other AFOs or alarm messages sent from the AF to AFOs) which reflect the change of the system environment; the most important AFO is always scheduled to execute next.

*The authors are with the Department of Computer and Information Science, University of Mississippi, Weir 302, University, MS 38677.

The advantage of this approach is that it eliminates the use of interrupts to invoke the processing of critical AFOs. In the AF architecture, the importance of AFOs is always re-computed every time an event occurs, e.g., when a message carrying a new fact is delivered by the AF. Furthermore, this guarantees that the system runs in real-time because it ensures that the most important rule or AFO is executing at any time.

Despite the technical soundness of the SKRAM project for real-time expert systems, there is a perceivable system overhead for executing the event-drive scheduling mechanism. First, the processor is essentially interrupted (to determine the most important AFO) every time a message is delivered from one AFO to other AFOs, or when an alarm is triggered. Second, if the currently executing AFO indeed becomes less important than some other AFO, then the AF must perform a context switch to allow the most important AFO to run. This issue is further complicated by the rule-grouping algorithm which allocates independently scheduled rules to separate AFOs. For example, if each rule is allocated to a separate AFO, then this overhead may be quite significant because the AF has to deal with a handful of processes (AFOs) at the AFO level and the chance that this overhead will become a factor is high; on the other hand, if the rule-grouping algorithm allocates many rules to a single AFO, then, although the overhead at the AFO level is reduced (because there are fewer processes), the overhead required for scheduling different rules within a single AFO may again be significant. A similar mechanism (e.g., [GRE 91B, FOR 82]) has to be applied at the rule level for scheduling the execution of rules within a single AFO, thus introducing a lot of system maintenance overhead for keeping track of which rules are ready and which rule should be fired, etc. Both overheads, one at the AFO level and the other at the rule level, degrade the performance of the system since they defer the time required for firing productions rules. Therefore, the best rule-clustering algorithm for optimizing the system performance should balance these two overheads, thereby minimizing the total system overhead.

Current research directions toward rule-clustering algorithms for expert systems [MEH 91, JAC 90] are not tied in with the AF architecture and thus do not consider this design trade-off. The basic approach of these algorithms is to select a distance metric between each pair of rules and, after allocating each rule to a separate group, iteratively merge groups with the minimum inter-group distance until a stopping condition is met. These algorithms stop either after a pre-determined number of groups is obtained [MEH 91] or the inter-group distance between the next two groups to be combined is no longer positive [JAC 90]. These algorithms are not applicable to real-time expert

systems running on the AF architecture because there is no provision for balancing the AFO-level and rule-level system overheads.

Another class of algorithms [KER 70, LEE 89, BAR 82] deal with a more general graph-partitioning problem as follows: given a graph G with costs on its nodes and edges, partition the nodes of G into k groups of specified sizes, $s_1, s_2, \cdots, s_k$, so as to minimize the total cost of the edges cut. The problem can be related to the rule-clustering problem in the AF architecture as follows: (a) each node corresponds to a rule with its cost proportional to the size of that rule, (b) each edge going from rule i to rule j is assigned a cost of one if rule j uses a fact generated by rule i; or zero otherwise, (c) minimizing the total cost of the edges cut when the graph is partitioned into k groups of specified sizes corresponds to minimizing the AFO-level overhead with k AFOs in the system because the message flow between these k AFOs is minimized, and (d) the specified size for each of the k AFOs (AFO_i, $1 \leq i \leq k$) is related to the rule-level overhead with the overhead of AFO_i proportional to s_i. After the problem is formulated this way, the key idea then is to select the best k and $s_1, s_2, \cdots, s_k$, such that the performance of the system is optimized. Thus, these graph-partitioning algorithms (which vary in time and space requirements) are not solutions to our problem, but can be used as subroutines in the construction of the best rule-clustering algorithm for determining the best k and $s_1, s_2, \cdots, s_k$. To construct the best rule-clustering algorithm, first a model has to be constructed to describe the run-time behavior of the real-time expert system running on the AF architecture. The model will allow us to evaluate a selected $(k, s_1, s_2, \cdots, s_k)$ set by taking into account the trade-off between the AFO-level and rule-level overheads. The best k and $s_1, s_2, \cdots, s_k$ obtained this way will be the one that optimizes the performance of the system. This report concerns the case in which $s_1, s_2, \cdots$ and s_k are equal, i.e., k-way balanced partition, and is interested in knowing what k value can best optimize the performance of the system.

The rest of this report is organized as follows. Section 2 proposes a probabilistic Markov model for estimating the performance of real-time expert systems running on the AF architecture. Based on this model, a closed-form expression for the system performance is derived which serves as a basis for evaluating rule grouping algorithms. Section 3 discusses techniques for parameterizing the model parameters. Based on the model, Section 4 develops a best-k rule grouping algorithm by utilizing a graph-partitioning algorithm as a subroutine, with the goal of searching for the best k value for optimizing the performance of the resulting system. Section 5 demonstrates the feasibility of the model and algorithm by applying them to the Advanced GPS Receiver (AGR) and the

Manned Maneuvering Unit (MMU) knowledge bases and compare the theoretical prediction with the empirical result for the AGR system. Finally, Section 6 summarizes the report and outlines some future research areas.

2 Modeling of AF Architecture

In the construction of the model for describing the run-time behavior of the AF architecture, we first distinguish the following two classes of AFOs: (a) Application Code Objects (ACOs) which are created from groups of production rules, and (b) System Code Objects (SCO) which perform system-level functions such as injecting or logging data [RTI 91].

A Markov probability model is proposed as a tool for evaluating the performance of rule clustering algorithms running on the single-processor AF architecture. The model uses a finite state machine in which the system execution is modeled as a progression through the following states:

- *ACO*: in this state the CPU is allocated to an ACO. The ACO may be performing useful work, e.g., firing a rule, generating some new fact, etc., or it may be spent for the rule-level overhead. e.g., determining which rules are ready and which rules should be fired by applying a pattern matching algorithm for saving the states of the rules within a single ACO, etc.

- *AF*: in this state, the CPU is allocated to the AF's operating system. This state models part of the AFO-level overhead due to the employment of the event-driven scheduling mechanism. If the AF gets control because the currently executing ACO sent a message to a value distribution node (VDN) [RTI 91], then the CPU power is used to (a) deliver the message, (b) calculate the priming function of the receiving ACOs, and (c) compare the importance levels of the sending and receiving ACOs to determine which ACO is to get the CPU next.

- *SW*: in this state the AF uses the CPU for performing a context switch, i.e., saving the states of the sending ACO (or the ACO which just returned to AF) and transferring the control to the ACO having the highest importance level. This state also models part of the AFO-level overhead.

- *SCO*: in this state an input or output SCO takes over due to a data injection or logging event (e.g., an alarm event). The CPU power is used to service the I/O alarm event, switch the CPU from the currently executing *ACO*, perform a data injection or logging operation, and

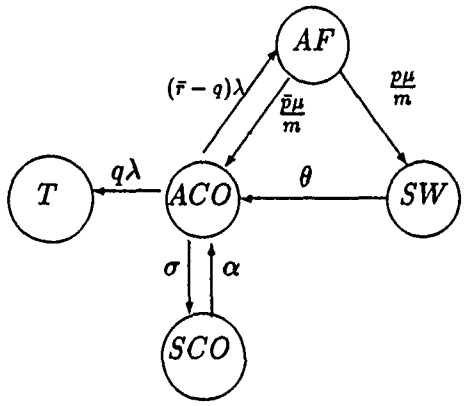


Figure 1: The Markov Model for the Single Processor AF Architecture.

switch the CPU back to an ACO when the operation is completed. This state also models part of the AFO-level overhead.

- T: the termination state.

Figure 1 shows the model which is constructed as follows:

1. When the system is in the *ACO* state, the CPU is used by an ACO for processing an arriving fact, i.e., determining which rules are ready to fire due to the arrival of the new fact, and selecting one to fire among the set of ready rules, if any. Firing a rule may generate new facts which may (a) terminate the system, (b) instantiate other rules within the ACO, thus requiring the ACO to process the new fact again, or (c) instantiate rules in other ACOs, thus requiring the ACO to send a message carrying the fact to other ACOs via a VDN. The time that is required for the ACO to process an arriving fact is assumed to be exponentially distributed with a constant rate λ. The probability of terminating the system (event (a) above), executing the same ACO again (event (b)) and returning the control to AF (event (c) or when there is no rule that can be fired) are q, r, and $1 - r - q$, respectively. In Figure 1 these correspond to the horizontal transition at rate $q\lambda$, the self-looping transition at rate $r\lambda$ (not shown in the figure) and the diagonal transition at rate $(\bar{r} - q)\lambda$, respectively.

2. When the AF takes over (i.e., in state *AF*), the time that is required for the AF to deliver a message from the sending ACO to another ACO and re-compute the priorities of these two

ACOs to determine which one has the higher priority is assumed to be exponentially and randomly distributed with a constant service rate μ. Assume that the average number of *ACO* connections for each rule having at least one ACO connection is m. Then, the overall rate for the AF to deliver a message and re-compute priorities of all involved ACOs is $\frac{\mu}{m}$. Further, it is assumed that with probability p, the control will be transferred to a new ACO. Hence, with probability $\bar{p}$ the sending ACO will retain the CPU. These events correspond to the diagonal transitions at rate $\frac{\bar{p}\mu}{m}$ if the CPU is retained by the sending ACO and at rate $\frac{p\mu}{m}$ if the CPU is allocated to a new ACO, respectively.

3. When the AF allocates the CPU to a new ACO, the system is in the *SW* state. The CPU time is used for performing a context switch. The time for performing a context switch is assumed to be exponentially and randomly distributed with a constant context-switching rate θ. After the context switch is performed, the CPU is allocated to a new ACO and the system is again in state *ACO*. This event corresponds to the horizontal transition at rate θ.

4. If the list of SCOs to be checked by AF at regular intervals exists, then state *SCO* exists. (See section 4, AFC manual [RTI 91].) In this case, whenever an ACO makes a call to a framework function, AF will compare the importance level of the executing ACO with those SCOs in the list (if sufficient time has elapsed since the last time the comparison was performed) to see if the CPU should be allocated to one of them. The time for this event to occur is assumed to be exponentially distributed with a constant alarm rate σ (e.g., the data sensor rate plus data logging rate). The event corresponds to the vertical transition at rate σ from state *ACO* to state *SCO*. When the system is in state *SCO*, the time required for the system to return to state *ACO* is assumed to be exponentially distributed with a constant rate α. This quantity includes the time required to perform a data injection or logging operation as well as the time for performing a context switch twice: one to switch from state *ACO* to *SCO* and one to switch back from *SCO* to *ACO*. This event corresponds the vertical transition at rate α from state *SCO* to state *ACO*.

5. The system continues to perform state transitions until it reaches state T.

The performance measure of interest in this model (Figure 1) is the mean time to termination (*MTTT*), i.e., the mean time required for the system to transit to state T from state *ACO*. This

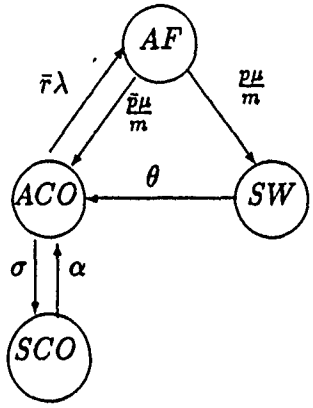


Figure 2: The Single Processor Markov Model with No Termination State.

measure reflects the average time required for a rule system to terminate. Thus, the design goal in this case is to minimize the $MTTT$.

A variation of the above model is the case in which a rule system never terminates (e.g., running an infinite loop in an embedded process-control system). This is shown in Figure 2. In this case, the performance measure of interest is the effective production rate, namely, $\lambda P_{ACO}(\infty)$. This performance metric measures the number of facts generated (or rules fired) per time unit and is obtained as follows: λ is the production rate when the system is at state ACO, but the system only stays in state ACO with probability $P_{ACO}(\infty)$; consequently, the effective production rate is $\lambda P_{ACO}(\infty)$. Note that this performance metric accounts for the trade-off between the AFO-level and rule-level overheads by considering the proportion of time the system stays at state ACO (which accounts for the AFO-level overhead) and the rate at which the system fires a rule once the system is in that state (which accounts for the rule-level overhead). An important observation is that λ decreases as more rules are grouped within one ACO because more time is spent for processing a new arriving fact over a larger sorting network.

The following table summarizes the symbols which will be used in the report.

N: number of rules in the system.

k: number of ACOs (groups) in the system.

m: the average number of external ACOs that a rule connects to.

s_j : number of rules in the jth group.

λ : the *new-fact generation* (or rule-firing) rate when the system is at state ACO.

λ_i : λ with each ACO containing i rules.

μ : the *service* rate of AF.

θ : the *context switching* rate of AF.

σ : the *alarm* rate of SCOs.

α : the service rate of an SCO, including the context switch overhead.

D : the average time to complete an I/O operation.

q : the probability that when an ACO generates a fact, the termination condition has been met.

$\bar{q}$: $1 - q$

r : the probability that when an ACO generates a fact, the fact will instantiate rules within the same ACO.

$\bar{r}$: $1 - r$

$\bar{r} - q$: the probability that when an ACO generates a fact, the fact will instantiate rules in other ACOs.

η : the average number of rules affected (that have to be inspected) in an ACO when a new fact arrives.

η_j : η with each ACO containing j rules.

p : the probability that AF transfers the CPU to a new ACO after servicing a message delivery.

$\bar{p}$: $1 - p$, the probability that the sending ACO retains the CPU.

$P_i(t)$: the probability that the system is at state i at time t.

p_o : the probability that the AF will allocate the CPU to a connecting ACO.

$P_i(\infty)$: the steady state probability that the system is at state i.

P_{system} : the effective production rate of an expert system, i.e., $\lambda P_{ACO}(\infty)$.

$P_{system,i}$: P_{system} with each ACO containing i rules.

Without loss of generality, the analysis henceforth assumes that the rule system never terminates. The performance measure of interest is the effective production rate, i.e., the number of facts generated (or rules fired) per second, $\lambda P_{ACO}(\infty)$, or, P_{system}. This measure can be obtained by solving for $P_{ACO}(\infty)$ from the following set of linear equations describing the Markov model in Figure 2 [ROS 89]:

$$P_{ACO}(\infty) + P_{SCO}(\infty) + P_{SW}(\infty) + P_{AF}(\infty) = 1;$$

$$\mu P_{AF}(\infty) = \bar{r}\lambda P_{ACO}(\infty);$$

$$\theta P_{SW}(\infty) = p\mu P_{AF}(\infty);$$

$$\alpha P_{SCO}(\infty) = \sigma P_{ACO}(\infty).$$

This yields

$$P_{ACO}(\infty) = \frac{1}{1 + \frac{\sigma}{\alpha} + \frac{\lambda\bar{r}}{\mu} + \frac{\lambda\bar{r}p}{\theta}}$$

and thus

$$P_{system} = \frac{\lambda}{1 + \frac{\sigma}{\alpha} + \frac{\lambda\bar{r}}{\mu} + \frac{\lambda\bar{r}p}{\theta}} \quad (1)$$

This performance equation has an important implication: for a given k-way partition (e.g., $k = 1$ means there is only one group), the system performance increases as λ, θ, μ or α increases, and as $\bar{r}$, p, m, or σ decreases. To improve the performance of the system, we like to decrease $\bar{r}$, p and m by selecting a lower k value (thus minimizing the AFO-level overhead); however, this would decrease λ (i.e., increasing the rule-level overhead) which will adversely degrade the system performance. Therefore, the goal is to select a best k value which can balance these two opposite effects.

3 Parameterization of Model Parameters

Equation (1) can serve as a basis for predicting the performance of the system when all parameters are quantified after a particular rule grouping algorithm has been applied to partition the production rules into groups. Some of these parameters are machine-dependent but insensitive to the use of rule grouping algorithms. These are called *statistically measurable* parameters. Others are sensitive to the utilization of rule grouping algorithms and the way a group processes a new fact. These are called *computable* parameters. This section discusses techniques for quantifying these two types of parameters.

3.1 Statistically Measurable Parameters

Statistically Measurable Parameters include θ, μ, σ and α. These are discussed in the following.

- θ denotes the average number of times the CPU is capable of performing a context switch (dedicated for that purpose) per time unit. For an 80386 machine, for example, the average time required for the AF to perform a context switch is about 2 milliseconds [RTI 91].

Therefore,

$$\theta \approx 500\ sec^{-1} \quad \text{for 80386}$$

- μ denotes the average number of times the CPU is capable of (a) delivering a message from an ACO to another ACO and (b) re-computing the priorities of these two ACOs to determine which one has the higher priority. For an 80386 machine, the average time required for the AF to perform this service is about 1.5 milliseconds [RTI 91]. Therefore, μ can be estimated by

$$\mu \approx 700\ sec^{-1} \quad \text{for 80386}$$

- σ denotes the average number of times per time unit that the execution of the rule system is interrupted by data injection or logging activities. For example, if it is measured that there are about five data injection and five data logging operations per second, then $\sigma = 10\ sec^{-1}$.

- α denotes the average number of times the CPU is capable of performing a data injection or logging operation, given that an alarm event has just interrupted the execution of the rule system. The way the system is modeled assumes that whenever the system is interrupted by an alarm event, an input or output SCO is ready to perform an I/O operation. Further, the time required to service the operation includes the overhead for the system to switch back and forth between states ACO and SCO. Based on this, α can be approximated by

$$\alpha \approx \frac{1}{\frac{2}{\mu} + \frac{2}{\theta} + D}.$$

  where D stands for the average time actually required to input/output the data from/to the I/O devices, e.g., reading a block of data from the serial port or dumping a block of data to the disk, etc. D is a statistically measurable quantity, e.g., $D \approx 20$ milliseconds for an 80386 disk read/write operation.

3.2 Computable Parameters

Computable parameters include r, m, p and λ. These parameters are sensitive to the use of rule grouping algorithms. (In fact, they are outputs of rule grouping algorithms.) In addition, λ is sensitive to the mechanism with which each group (or ACO) processes new arriving facts.

We first define some data structures used by rule grouping algorithms. Then, we explain how to compute these parameters using these data structures. Let $A[1..N, 1..N]$ be the adjacency matrix

of the input Evidence Flow Graph (EFG) that shows the connectivity of production rules, i.e., $a_{ii} = 0$ for all i and $a_{ij} = 1$ if rule j uses a fact generated by rule i; 0 otherwise. Let $B[1..N]$ be the output partition vector of a particular rule grouping algorithm which has been applied to partition the rule production rules into k groups, such that if $b[i] = j$ then rule i is allocated to the jth group. In the following we discuss techniques for computing r, p, n and λ.

- r denotes the probability that when an ACO generates a fact, the fact will instantiate rules within the same ACO, rather than instantiating rules in other ACOs. We can approximate r by

$$r = \frac{C_i}{C} = \frac{\sum_{i=1}^{N} \sum_{\substack{j=1 \\ b[j]=b[i]}}^{N} a_{ij}}{\sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij}} \tag{2}$$

  where C stands for the total number of connections in the connection matrix and C_i stands for the total number of internal connections for all k groups after the partition. When there are fewer groups, $r \rightarrow 1$ and, conversely, when there are many groups, $r \rightarrow 0$. For a given k, a rule grouping algorithm should maximize r (thereby minimizing $\bar{r}$) as much as possible to reduce the AFO-level overhead.

- m denotes the average number of external groups that a rule connects to. This can be estimated in three steps: (a) computing, for each rule having at least one external connection, the number of groups it connects to, (b) accumulating the total count for all such rules, and (c) dividing the total count by all such rules. Hence, the average number of external ACO connections per rule is an output of a partition. Obviously, m increases as the number of groups increases and, as a result, the AFO-level overhead involved for sending a message from an ACO to all its connecting ACOs increases, a behavior which we have modeled in the Markov model.

- p represents the probability that the AF will perform a context switch after it had serviced a message delivery. If we assume that the AF, when servicing a message delivery from an ACO, inspects each connecting ACO in succession to determine whether the CPU should be switched to that connecting ACO with a success probability p_o, then the probability that the

AF will perform a context switch is proportional to m as follows:

$$p = \sum_{i=1}^{m} p_o(1-p_o)^{i-1} \tag{3}$$

This approach gives a reasonable estimation of p since it can well explain why $p \approx 1$ when there are many groups.

- λ denotes the rate at which a new fact is generated by a group. The magnitude of this parameter largely depends on how many rules there are in each group. For that purpose, let λ_j denote this rate when there are j rules in each group. Furthermore, let λ_j relate to λ_1 (the new fact generation rate when each ACO contains exactly one rule) by the following equation:

$$\frac{\lambda_1}{\lambda_j} = \eta_j$$

where η_j stands for the number of rules that have to be processed by a group containing j rules when a new fact arrives. The advantage of relating λ_j with λ_1 is that λ_1 is a measurable quantity and, therefore, if η_j can be estimated properly, λ_j can be computed from λ_1. For example, for 80386 machines with 1Mflops processing capability, if each rule contains approximately 500 machine language (floating point) instructions, then

$$\lambda_1 \approx 2000\ sec^{-1}$$

To properly estimate η_j, we distinguish the following two mechanisms with which each group may process an arriving fact:

  - *Non Data-Driven Mechanism:* In this scheme, all j rules within a group are encoded as a succession of `if-then` code blocks. Therefore the processing time required to generate a new fact is about j times as large as required when there is only one rule per group because all j code blocks must be examined before one is selected to fire. Under this *non data-driven* scheme:

$$\eta_j = j \ ; \text{ and } \lambda_j = \frac{\lambda_1}{j}$$

Consequently, for a selected k-way partition, since $j = N/k$,

$$\lambda = \frac{\lambda_1 k}{N} \quad \text{for non data-driven}$$

where N is the number of rules in the rule system.

– *Data-Driven Mechanism:* In this scheme, each group maintains a sorting network (consisting of RTGO, BFO and RO data structure objects – see [GRE 91B]) with a size proportional to the number of rules in each group. When a new fact arrives, instead of inspecting all j rules, only a portion of these rules in the sorting network (that are affected by the new fact) are inspected before the most important rule is selected to run. As a result, η_j can be estimated by the average number of internal connections per rule, i.e.,

$$\eta_j = \frac{C_i}{N}$$

where C_i stands for the total number of internal connections for all groups as defined in equation (2). Recall that $r = C_i/C$ where C stands for the total number of connections of the input connectivity matrix. Therefore the average number of rules that have to be inspected, η_j, is related to r by

$$\eta_j = \frac{rC}{N}$$

Consequently, for a selected k-way partition,

$$\lambda = \frac{\lambda_1 N}{rC} \quad \text{for data-driven}$$

4 The Best-k Rule Grouping Algorithm

In this section, we develop a rule grouping algorithm, called the best-k rule grouping algorithm. The goal is to compute the best k value under which the performance of the system can be optimized.

The best-k Algorithm

input: EFG, θ, μ, σ, λ_1, D, and p_o.
output: k and the k-way partition vector ($B[1..N]$).
data structure: the adjacency matrix $A[1..N, 1..N]$.

1. determine N, and $A[1..N, 1..N]$ from the input EFG.
2. set $k = 1$.
3. while $k \leq N$ do

(a) parameterize r, p, m, and λ based on the output from the KL algorithm; keep the resulting k-way partition.

(b) compute the performance metric of the system using Equation (1).

(c) if the system performance is the greatest so far, record this value, k, and the k-way partition in the variable $B[1..N]$.

(d) $k = k + 1$.

4. output best k and $B[1..N]$.

This algorithm utilizes the KL algorithm [KER 70] for grouping rules such that the total external cost on edge cuts (which results in the k-way partition) is minimized, thereby maximizing r and the system performance for a specified k value. The algorithm runs about 10 seconds for a rule system of size 10 (e.g., AGR) and about five minutes for a rule size of 100 (e.g., MMU).

5 Case Studies

In the following, the best-k rule grouping algorithm is applied to the AGR knowledge base system consisting of seven rules (therefore $N = 7$) and the MMU system consisting of 92 rules. The objective is to identify the best k value and its associated k-way partition which can maximize the effective production rate (i.e., system performance) as a function of λ_1. For the AGR system, we also compare the theoretical prediction with empirical results to demonstrate the feasibility of our model and algorithm.

The processor on which these knowledge base systems run is an 80386 and the following performance data are observed (e.g., see [WPI 91]):

- $\mu \approx 700 \ \ sec^{-1}$;
- $\theta \approx 500 \ \ sec^{-1}$;
- $\sigma \approx 10 \ \ sec^{-1}$;
- $\lambda_1 \approx 2000 \ \ sec^{-1}$;
- $D \approx 20$ milliseconds, the average disk service time required for performing a disk read/write operation.

k	r	m	p	λ^* (data-driven)	λ^* (non data-driven)
1	1.00	0	0.00	500	286
2	0.46	1	0.80	1076	571
3	0.32	1	0.80	1555	857
4	0.21	2	0.96	2000	1143
5	0.14	3	0.99	2000	1429
6	0.07	3	0.99	2000	1714
7	0.00	4	1.00	2000	2000

Table 1: r, m, p and λ As a Function of k ($\lambda_1 = 2000$).

- $\alpha \approx \frac{1}{\frac{2}{\mu}+\frac{2}{\theta}+D} = 35.$

5.1 AGR Knowledge Base System

5.1.1 Theoretical Prediction

By applying the KL algorithm on the EFG of the AGR rule set, it is observed that r, m, p, and λ (under data-driven as well as non data-driven) are related to k ($1 \leq k \leq 7$) as shown in Table 1 when $\lambda_1 = 2000$.

Figure 3 shows that for the AGR knowledge base when the data-driven scheme is used to process arriving facts, the best value of k depends on the magnitude of λ_1 (which reflects the degree of the rule-level overhead). The Figure shows that, if λ_1 is in the same order of magnitude as θ or μ (which occurs when each rule contains many instructions), then the best partition favors a high k. Conversely, if λ_1 is an order of magnitude higher than θ or μ (which occurs when each rule contains only a few instructions), then the best partition favors a low k. The physical interpretation is as follows: in the latter case the time required for the AF to schedule ACOs is an order of magnitude longer than that required for an ACO to process an arriving fact, select a rule, and fire the rule. Consequently, the system performance is improved by shifting the time-consuming scheduling work from the AF to within an ACO. An important observation from this figure is that, for each λ_1 value, there exists a k value under which the system performance is maximized. Since it is estimated $\lambda_1 \approx 2000$ for the AGR knowledge base system, the theoretical prediction suggests that the best partition should group all rules together when the data driven scheme is used.

Figure 4 shows the theoretical prediction for the AGR system when the *non data-driven* scheme

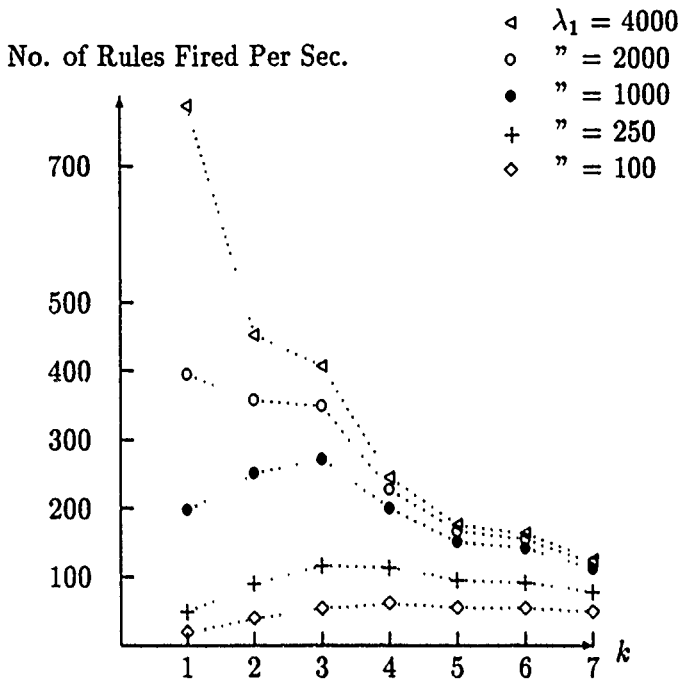


Figure 3: Theoretical Prediction for AGR: the Data-Driven Scheme.

is used to process arriving facts. It predicts the same trend as that of the data-driven scheme except that it shows that a higher value of λ_1 is required for the best partition to favor a low k value since more rules have to be inspected when a new fact arrives. Comparing Figure 3 with Figure 4, we observe the following: (a) the performance of the data-driven mechanism is always better than the non data-driven mechanism, and (b) for the same λ_1 value, the best k value selected under these two schemes may be different due to a different degree of rule-level overhead, e.g., when $\lambda_1 = 2000$, Figure 4 predicts that $k = 3$ is the best selection for the non data-driven scheme, while Figure 3 predicts that $k = 1$ is the best selection for the data-driven scheme.

5.1.2 Comparison To Experimental Results

Figure 5 compares the theoretical prediction with the experimental results for the AGR system where the y-coordinate represents the performance ratio between the one-group case and the k-group case with the x-coordinate varying k from one to seven. This experimental result is obtained by encoding the AGR rules in the form of Activation Framework Language (AFL) rule groups based on the output generated by the best-k rule grouping algorithm and having the AFL translator [GRE 91C] automatically translate them into groups of C program modules incorporating the data driven

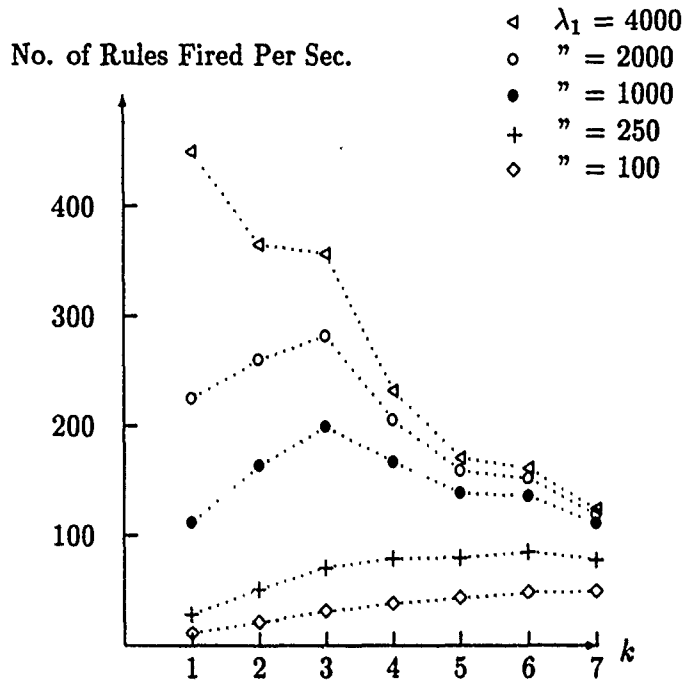


Figure 4: Theoretical Prediction for AGR: the Non Data-Driven Scheme.

mechanism as described in [GRE 91B]. These C program modules are then subsequently linked with AFO test modules and the Activation Framework For C Language (AFC) run-time libraries into one executable program. To ensure that no injected data are lost, the times at which data are injected into the system are manually adjusted (separately for each k value) in the input file (input.txt) such that the execution order of rules is repeated, e.g., rule 2 first, followed by rule 6 and then by rule 3, etc. The time required for AGR to output an expected sequence of logging records in the output file (output.txt) is measured separately for each k value and is used as the performance standard with which the theoretical prediction is compared. It was observed that, including the disk operation overhead required to inject and log the data, 7.19 seconds are needed for seven groups, 3.3 seconds are needed for two groups and 2.3 seconds are needed for one group for the expected execution order to be observed.

Figure 5 shows that, for the data-driven scheme, the theoretical prediction correlates well with the experimental results. The result that one group is the best selection is not surprising because the AGR system consists of only seven rules and therefore the rule-level overhead is not significant when compared with the AFO-level overhead, especially when the data driven mechanism is utilized. In the following we will show that for a system with moderate size, e.g., the MMU system with 92 rules, one group is not necessarily the best selection because as the the number of rules increases,

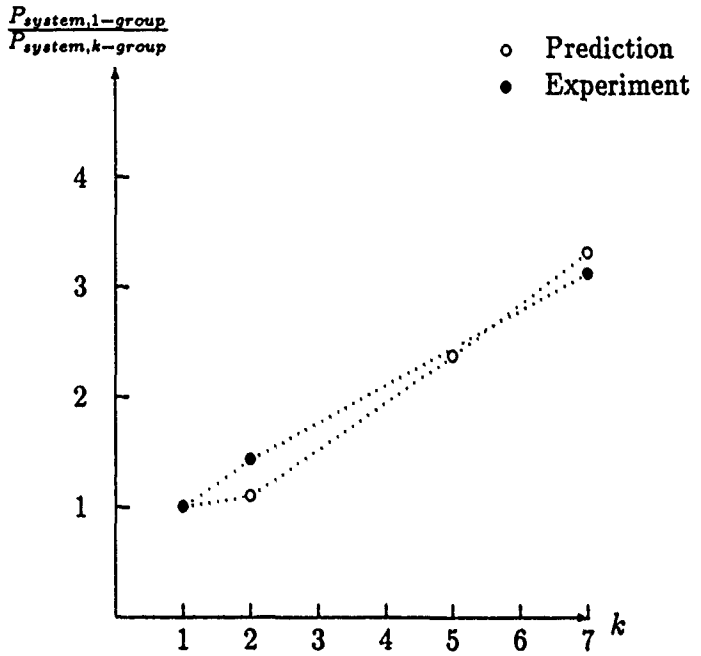


Figure 5: Theoretical Prediction vs. Experimental Result for AGR System.

the rule-level overhead also increases significantly since all rules are put into one group.

5.2 The MMU Knowledge Base

Figure 6 shows the theoretical prediction for the MMU knowledge base system (for data-driven scheme). The result suggests that four groups should be the best partition for $\lambda_1 = 2000$ although the result remains to be validated in the future.

6 Summary

In the report we have developed a theoretical basis with which the performance of rule grouping algorithms can be evaluated. The theoretical basis derives from the concept of a finite state machine by considering the execution of the rule system running on the AF architecture as a progression through states representing the rule-level overhead (the *ACO* state) and the AFO-level overhead (the remaining states), respectively. Based on the model, we derived the system performance equation. A best-k rule grouping algorithm was developed utilizing (a) the KL graph-partitioning heuristic for k-way partitioning the rule system and (b) the system equation for determining the best k value under which the system performance can be maximized. The results showed that the

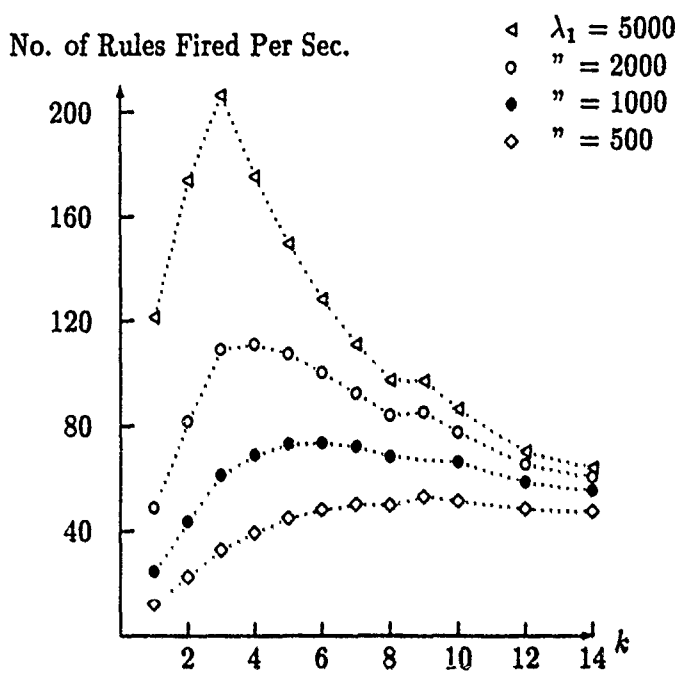


Figure 6: Theoretical Prediction for MMU: the Data-Driven Scheme.

best k-way partition is dictated by the balance between the rule-level and AFO-level overheads: if the rule-level overhead dominates the AFO-level overhead (e.g., when there are a lot of rules, or a lot of instructions per rule and/or the non data-driven mechanism is used), a high group number is favored; conversely, if the AFO-level overhead dominates the rule-level overhead (e.g., each rule is simple) a low group number is favored. In general, the best k-way partition has to be determined from the performance equation because one overhead may not always dominate the other. Finally, to demonstrate the feasibility of our approach, we compared the theoretical prediction from the theory and algorithm with empirical results for the AGR knowledge base system and a good correlation was observed.

Some possible future research directions include (a) further validating the model for the single processor AF architecture by comparing theoretical predictions with experimental results for medium (100 rules) to large (1000 rules) knowledge base systems, (b) investigating whether an unbalanced k-way partition can perform better than a balanced k-way partition by refining the model to accommodate groups of different size, (c) incorporating the notion of group importance level to more accurately estimate computable parameters, (d) refining the modeling of the data-driven mechanism by including the overhead required for updating the rule importance levels and selecting the most important rule to fire within a group, and (e) developing and validating a the-

oretical basis for evaluating the performance of rule grouping algorithms for rule systems running on a distributed multiprocessor AF architecture based on the experience learned from the single processor AF architecture.

References

[BAR 82] Barnes, E.R., "An algorithm for partitioning the nodes of a graph" *SIAM Journal on Algebraic and Discrete Methods*, Vol. 3, No. 4, Dec. 1982. pp. 541-550.

[FOR 82] Forgy, C.L., "Rete: a fast algorithm for the many pattern/many object pattern match problem," *Artificial Intelligence*, 1982, pp.17-37.

[GRE 87] P.E. Green, "AF: A framework for real-time distributed cooperative problem solving," *Distributed Artificial Intelligence*, edited by Michael N. Huhns, Morgan Kaufmann Publishers, Inc., 1987, pp. 153-176.

[GRE 91A] P.E. Green, J. Duckworth, L. Becker, and S. Cotterill, "Maintenance system for AI knowledge bases, Phase I - final report, Design of the AKRAM system", *contract number F33615-90-C-1470*, 1991.

[GRE 91B] P.E. Green, "A data driven mechanism for the execution of production rules in real-time computer based systems", *The Real-Time Intelligent Systems Cooperation*, 1991.

[GRE 91C] *AFL users manual*, Version 1.0, *The Real-Time Intelligent Systems Cooperation*, 1991.

[JAC 90] Jacob, R.J.K., and Froscher, J.N., "A software engineeering methodology for rule-based systems," *IEEE Trans. Knowledge and Data Eng.*, Vol. 2, No. 2, June 1990. *Computer-Aided Design*, Vol. 21, No. 10, Dec. 1989, pp. 611-618.

[KER 70] Kernighan, B.W., and Lin, S., "An efficient heuristic procedure for partitioning graphs," *Bell Sys. Tech. J.*, Vol. 49, 1970, pp. 291-307.

[LEE 89] Lee, C-H, Park C-I, and Kim M., "Efficient algorithm for graph-partitioning problem using a problem transformation method,"

[MEH 91] Mala Mehrotra, "Rule groupings: a software engineering approach towards verification of expert system," *NASA Contract NAS1-18585*, Final Report, Feb. 1991.

[ROS 89] S.M. Ross, *Introduction to Probability Models*, New York, Academic Press, 4th Edition, 1989.

[RTI 91] The Real-Time Intelligent Systems, *Activation Framework Operating System Environment*, Users Manual, AFC Version 2.5, 1991.

[WPI 91] Worcester Polytechnic Institute, *Knowledge Representation into Ada Methodlodgy*, Project Review Meeting, May, 1991.

1991 USAF-UES SUMMER FACULTY RESEARCH PROGRAM/

GRADUATE STUDENT SUMMER SUPPORT PROGRAM

Sponsored by the

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

Conducted by the

Research Developement Laboratories

FINAL REPORT

A NOTE ON PRONY'S METHOD

Prepared by:	David Choate and Wendy Barkman
Academic Rank:	Associate Professor
Department and	Science Division
University	Transylvania University
Research Location:	Air Force Avionics Laboratory,
	Electronic Support Measure Group
USAF Researcher:	Dr. Jim Tsui
Date:	August 12, 1991
Employee Number:	183

A NOTE ON PRONY'S METHOD

David B. Choate, Associate Professor, and Wendy Barkman

ABSTRACT- Prony's method can be used to compute the frequencies of two signals using four receivers. It is shown in this note that this can be done with only three receivers.

I. INTRODUCTION

Briefly stated Prony's algorithm [1] for the two signal case can be presented as follows:

Assume that the unknown frequencies, a and b, are roots of the quadratic equation

$$y^2 - \alpha_1 y - \alpha_2 .$$

Given the four observation equations from receivers

(i) $x_0 = c_1 + c_2$

(ii) $x_1 = c_1 a + c_2 b$

(iii) $x_2 = c_1 a^2 + c_2 b^2$

(iv) $x_3 = c_1 a^3 + c_2 b^3$

, we first multiply equation (i) by α_2, equation (ii) by α_1 and equation (iii) by -1 to obtain

$$\alpha_2 x_0 = c_1\alpha_2 + c_2\alpha_2$$

$$\alpha_1 x_1 = c_1\alpha_1 a + c_2\alpha_1 b$$

$$-x_2 = -c_1 a^2 - c_2 b^2 \ .$$

Adding these three equations gives

$$\alpha_2 x_0 + \alpha_1 x_1 - x_2 = c_1(\alpha_2 + \alpha_1 a - a^2) + c_2(\alpha_2 + \alpha_1 b - b^2)$$

$$= 0 + 0 = 0 \ , \text{ or}$$

$$\alpha_1 x_1 + \alpha_2 x_0 = x_2 \quad (1)$$

Now multiply equation (ii) by α_2, equation (iii) by α_1 and equation (iv) by -1 to obtain

$$\alpha_1 x_2 + \alpha_2 x_1 = x_3 \quad (2).$$

Equations (1) and (2) yield solutions for the α's provided

$$(x_1)^2 \neq x_0 x_2 \ . \quad (3)$$

Substituting these values into the original quadratic gives

$$a, b = \frac{\alpha_1 \pm \sqrt{\alpha_1^2 + 4\alpha_2}}{2}$$

But observe that equation (3) demonstrates that the <u>existence</u> of distinct solutions depends only on the results of the first three receivers: x_0, x_1 and x_2. So it is reasonable to ask whether their location depends only on just these x's as well.

Furthermore, a and b are typically assumed to be on the

unit circle. And yet nowhere in Prony's derivation is this property exploited. In this paper we will show that if attention is paid to this detail only three receivers are required to obtain both solutions.

II. THREE-FOUR

Consider the system

$$\begin{aligned} &(i) \quad x_0 = c_1 + c_2 \\ &(ii) \quad x_1 = c_1 e^{-j\theta_1} + c_2 e^{-j\theta_2} \\ &(iii) \quad x_2 = c_1 e^{-j2\theta_1} + c_2 e^{-j2\theta_2} \end{aligned} \qquad (I)$$

, where the c's are nonzero complex numbers and the θ's are real. Both are unknown. And the x's are given complex numbers.

Set a = $e^{-j\theta}1$ and b = $e^{-j\theta}2$ and assume that a ≠ b. System (I) is equivalent to

$$\begin{pmatrix} 1 & 1 \\ a & b \\ a^2 & b^2 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} x_0 \\ x_1 \\ x_2 \end{pmatrix}$$

which in turn is equivalent to

$$\begin{pmatrix} 1 & 1 & 1 \\ a & b & 0 \\ a^2 & b^2 & 0 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ 0 \end{pmatrix} = \begin{pmatrix} x_0 \\ x_1 \\ x_2 \end{pmatrix} \qquad (II)$$

(Note that the 1 we have added to the 1x3 position can be replaced by any nonzero constant without changing the sequel.)

Since a ≠ b, the inverse of the coefficient matrix of (II) is

$$\begin{pmatrix} 0 & -b/(a^2-ab) & 1/(a^2-ab) \\ 0 & (a^2/b)/(a^2-ab) & (-a/b)/(a^2-ab) \\ 1 & [b-(a^2/b)]/(a^2-ab) & -[1-(a/b)]/(a^2-ab) \end{pmatrix}$$

After solving the system we obtain

$$c_1 = [-b/(a^2-ab)]x_1+[1/(a^2-ab)]x_2$$
$$c_2 = [(a^2/b)/(a^2-ab)]x_1+[(-a/b)/(a^2-ab)]x_2$$

By (I)i we get

$$x_0 = c_1 + c_2 = [\frac{a+b}{ab}]x_1 + [\frac{-1}{ab}]x_2.$$

So $abx_0 = (a + b)x_1 - x_2$. (3)

Since a is on the unit circle, 1/a = a*, the complex conjugate of a. Taking conjugates of (3) gives

$$ab(x_2)^* = (a + b)(x_1)^* - (x_0)^* . \qquad (4)$$

Solving for a in both (3) and (4) yields

$$\frac{bx_1 - x_2}{bx_0 - x_1} = \frac{b(x_1)^* - (x_2)^*}{b(x_2)^* - (x_1)^*}$$

provided $b \neq x_1/x_0$.

Expanding and combining gives

$$[x_1(x2)^* - (x_1)^*x_0]b^2 + [x_0(x_0)^* - x_2(x_2)^*]b + [x_2(x_1)^* - (x_0)^*x_1] = 0.$$

Set $A = x_1(x_2)^* - (x_1)^*x_0$

$B = x_0(x_0)^* - x_2(x_2)^* = |x_0|^2 - |x_2|^2$

and $C = A^* = x_2(x_1)^* - (x_0)^*x_1$.

Therefore

$$e^{-j\theta_2} = b = \frac{B \pm \sqrt{B^2 - 4|A|^2}}{2A}$$

provided $A \neq 0$.

But (3) and (4) are symmetric in a and b. So by a similar computation

$$e^{j\theta_1} = a = \frac{-B \pm \sqrt{B^2 - 4|A|^2}}{2A}$$

provided $A \neq 0$.

But since $a \neq b$ we conclude after checking in system (I) that

$$\theta_1 = [\ln(\frac{-B - \sqrt{B^2 - 4|A|^2}}{2A})]j$$

and

$$\theta_2 = [\ln(\frac{-B + \sqrt{B^2 - 4|A|^2}}{2A})]j$$

But certainly there is something disturbing about this derivation. In completing it we have assumed that $|a| = |b| = 1$. And yet the expression

$$\frac{-B \pm \sqrt{B^2 - 4|A|^2}}{2A}$$

will have a norm of 1 only if $B^2 - 4|A|^2 \leq 0$.

So at first it would seem that three receivers will suffice only if this inequality is satisfied. But we can avoid this restriction by a slight of hand: we will simply abandon our requirement that a and b lie on the unit circle as soon as it cannot be satisfied.

When $B^2 - 4|A|^2 > 0$, set

$\theta_1 = \beta_1 + \gamma_1 j$, its complex form.

Then

$$a = e^{-j\theta_1} = e^{\gamma_1} e^{-j\beta_1}$$

Thus one frequency is the real part of θ_1 and, of course, the other is the real part of θ_2. These values appear to agree with Prony's- at least to fifteen decimal places.

But there are still two cases for which we have not obtained a solution; namely, $b = x_1/x_0$ and $A = 0$.

III. SHUT THE DOOR

To deal with each of these cases we will examine the polar form of our x's and c's:

$$x_i = |x_i| e^{j\alpha_i} \ , \ i = 0,1,2$$
$$c_i = |c_i| e^{j\phi_i} \ , \ i=1,2$$

With these equations we will find that the case $b = x_1/x_0$ is not really a possibility under the assumption that $a \neq b$.

Case 1: If the given x's are nonzero, then $a = b$ only if $b = x_1/x_0$.

Proof. ($\rightarrow$) If $a = b$, then $x_1 = c_1 a + c_2 b$

$$= (c_1 + c_2) b$$
$$= x_0 b$$

($\leftarrow$) Suppose $b = x_1/x_0$. If $a \neq b$, then by equation (3)

$$0 = a(bx_0 - x_1) = bx_1 - x_2$$

which implies that $b = x_2/x_1$. So $(x_1)^2 = x_0x_2$. Therefore

$$\alpha_1 = (\alpha_0 + \alpha_2)/2 \qquad (5)$$

Furthermore $b = (x_1)^*/(x_2)^*$. Thus

$$0 = A = x_1(x_2)^* - (x_1)^*x_0$$

$$= (c_1a + c_2b)\{[(c_1)^*/a^2] + [(c_2)^*/b^2]\}$$

$$- \{[(c_1)^*/a] + [(c_2)^*/b]\}(c_1 + c_2), \text{ or}$$

$$a^2 = \left[\frac{c_2(c_1)^*}{c_1(c_2)^*}\right]b^2$$

Thus by (I)iii

$$x_2 = \left[\frac{c_2}{(c_2)^*}\right]\left[\frac{(c_1)^* + (c_2)^*}{(c_2)^*}\right]b^2$$

$$; so \quad \frac{x_2}{(x_0)^*} = \left[\frac{c_2}{(c_2)^*}\right]b^2$$

Taking logs and using the polar forms of x and c gives

$$\alpha_1 = \phi_2 - \theta_2.$$

Similarly, $\alpha_1 = \phi_1 - \theta_1.$

So, of course,

$$\phi_2 - \phi_1 = \theta_2 - \theta_1 .$$

By (I)ii we obtain:

$$|x_1|e^{\alpha_1} = |c_1|e^{j\phi_1}e^{-j\theta_1} + |c_2|e^{j\phi_2}e^{-j\theta_2}$$

$$= |c_1|e^{j\alpha_1} + |c_2|e^{j\alpha_1}, \text{ or}$$

$$|x_1| = |c_1| + |c_2|. \qquad (6)$$

But since $b = x_1/x_0$, $|x_0| = |x_1|$ (7) So

$$|c_1 + c_2| = |c_1| + |c_2|.$$

By the triangle inequality $\phi_1 = \phi_2$, or $\theta_1 = \theta_2$. So a = b, a contradiction.

Case 2: If A = 0 and the x's are nonzero, then a solution to the system (I) is given by

$$a = \frac{x_2}{x_1} - \sqrt{\frac{x_2^2}{x_1^2} - \frac{x_2}{x_0}}$$

and

$$b = \frac{x_2}{x_1} + \sqrt{\frac{x_2^2}{x_1^2} - \frac{x_2}{x_0}}.$$

Proof. From the information gathered in the proof of Lemma 1 we have

$$x_2 = c_1a^2 + c_2b^2$$

$$= |c_1|e^{j\alpha_1}a + |c_2|e^{j\alpha_1}b, \text{ or}$$

$$\frac{|x_1|x_2}{x_1} = |c_1|a + |c_2|b \quad (8)$$

After writing (6) and (8) in matrix form we get

$$\begin{pmatrix} 1 & 1 \\ a & b \end{pmatrix} \begin{pmatrix} |c_1| \\ |c_2| \end{pmatrix} = |x_1| \begin{pmatrix} 1 \\ \frac{x_2}{x_1} \end{pmatrix}$$

from which we obtain

$$|c_1| = \frac{|x_1|}{(b - a)} [b - \frac{x_2}{x_1}] \quad (9)$$

$$|c_2| = \frac{|x_1|}{(b - a)} [-a + \frac{x_2}{x_1}] \quad (10)$$

Now suppose $|c_1| = |c_2|$.

Using (7), (9) and (10) we get

$$a + b = 2x_2/x_1 \quad (11)$$

But by (3) and (11)

$$abx_0 = 2x_2 - x_2 = x_2 ,$$

$$\text{or } ab = x_2/x_0. \quad (12)$$

Solving (11) and (12) simultaneously gives our formulas.

Our above argument certainly provides solutions, but it does not demonstrate that they are unique. However, our formulas yield frequencies which agree with Prony's method.

So it appears we can determine our frequencies in this situation unless it so happens that an x is zero. But as we

might expect deriving solutions for degenerate spectra amounts to little more than manual labor.

IV. PICK UP STICKS

If all our x's are 0, then there is only one signal and it is not unique.

If $x_1 = x_2 = 0$, then so is x_3. So, again, there is only one signal and it is not unique. This also follows if $x_1 = x_2 = 0$.

But if $x_0 = x_2 = 0$ and $x_1 \neq 0$, then $\theta_1 = 3\pi/2$ and $\theta_2 = \pi/2$.

If $x_1 = 0$, $x_0 \neq 0$ and $x_2 \neq 0$, then

$$\theta_1 = (\alpha_1 - \alpha_2)/2$$

and

$$\theta_2 = \theta_1 + \pi \ .$$

If $x_2 = 0$ and $x_0 \neq 0 \neq x_1$, then $A \neq 0$ and we can use our derived formulas.

If $x_0 = 0$ and $x_1 \neq 0 \neq x_2$, then $A \neq 0$ in this case as well.

So the solutions to system (I) have been determined.

V. LAY THEM STRAIGHT

Finally it would be convenient for the sake of programming to be able to determine at the outset whether the

signals are identical. To this end we prove:

Theorem. For nonzero x the equation a = b holds only if $|x_0|$ = $|x_1|$ = $|x_2|$.

Proof. (<-) Since a = b, $x_1 = (c_1 + c_2)a = x_0 a$, or $x_1/x_0 = a$.

Thus

$|x_0| = |x_1|$. Similarly $|x_1| = |x_2|$.

(->) If $|x_0| = |x_1|$, then $|c_1 + c_2| = |c_1| + |c_2|$. By the triangle inequality $\phi_1 = \phi_2$ which implies $\theta_1 = \theta_2$, or a = b.

REFERENCES

1. Hildebrand, F. B., Introduction to Numerical Analysis, McGraw-Hill, New York, 2nd ed., 1956.

CORRELATION DIMENSION OF CHAOTIC ATTRACTORS

Thomas K. Gearhart, Ph. D.
Associate Professor, Capital University

Abstract

The correlation dimension, introduced by Grassberger and Procaccia, provides a means of quantifying the degree of chaos in strange attractors associated with nonlinear dynamical systems. This report compares various methods of estimating the correlation dimension and discusses some of the difficulties associated with obtaining reliable estimates. Estimation algorithms are used to investigate the chaotic attractors associated with the Mackey-Glass differential delay equation.

1. Introduction

It is now well established that many nonlinear dissipative dynamical systems possess strange attractors for some values of system parameters. This means that trajectories are attracted to a zero volume fractal subset of the ambient phase space. On such a strange attractor, motion appears to be unpredictable and there is extreme sensitivity to initial conditions. The fractal nature of the attractor and the sensitivity to initial conditions result from stretching and folding operations within the system dynamics. Although motion on the attractor appears to be random, attractors often have dimensions much lower than the ambient phase space. The importance of this observation lies in its implication that the number of active degrees of freedom in the system is actually much lower than the dimension of the phase space. Measuring

the dimension of the attractor is a means of quantifying the degree of chaos in the system and distinguishing it from random noise. Thus, much research has been devoted to analytic and experimental determination of the dimension of strange attractors. One of the most effective dimension calculation techniques is based on the Correlation Dimension defined by Grassberger and Procaccia. In 1990, Nelson et. al. at the Avionics Directorate, Wright-Patterson A.F.B., began a study of strange attractors associated with a certain differential delay equation, known as the Mackey-Glass equation. Using the correlation dimension as their dimensional analysis tool, they encountered unexpected discontinuities in estimated dimension as they varied the number of sample points. In my preliminary visit in March, it was agreed that my research objective for the Summer would be to investigate the dimensional properties of the Mackey-Glass equation and, in particular, to explain the anomalies encountered in the preliminary study.

2. Correlation Dimension

Let $A \subseteq R^k$ and suppose $S = (x_1, x_2, \ldots, x_n)$ is a sample of points chosen from A. If r is a positive real number, define $\#(x_i,r)$ to be the number of points in $S\backslash(x_i)$ that are within distance r of x_i. Here distance is the usual Euclidean distance in R^k. Observe that $0 \le \#(x_i,r)/(n-1) \le 1$ and that $\#(x_i,r)/(n-1)$ represents the fraction of points in $S\backslash(x_i)$ that are within distance r of x_i. To understand how $\#(x_i,r)/(n-1)$ is related to the notion of the dimension of A, let us observe how $\#(x_i,r)/(n-1)$ scales with r for some familiar sets A. Let A be the unit square $[0,1] \times [0,1] \subseteq R^2$. Select a large random sample $S = (x_1, x_2, \ldots, x_n)$ of points in A and let r be a small positive real

number. We would expect the fraction of the points in $S\backslash\{x_i\}$ that are closer than r to x_i to be approximated by the ratio of the area of a circle of radius r centered at x_i to the area of the unit square. That is, $\#(x_i,r)/(n-1) = \pi*r^2$. So we see that for the two dimensional set A, $\#(x_i,r)/(n-1)$ scales as the second power of r. If A were instead the three dimensional set $A = [0,1] \times [0,1] \times [0,1]$ in R^3, we would have $\#(x_i,r)/(n-1) = 4*\pi*r^3/3$; i.e. the scaling of $\#(x_i,r)/(n-1)$ is as the third power of r. These examples make plausible the following generalization of the notion of dimension: If $\#(x_i,r)/(n-1)$ is directly proportional to r^d for small r, then the local dimension of A at x_i is d. To obtain a global, rather than local, measure of the dimension of A, consider the average of the expressions $\#(x_i,r)/(n-1)$ over the entire sample set $x_1, x_2, \ldots, x_n$. If the function H is defined by $H(x,y,r) = 1$ if $|x-y| < r$ and $H(x,y,r) = 0$ if $|x-y| \geq r$, we have:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\#(x_i,r)}{n-1} = \frac{1}{n(n-1)}\sum_{i=1}^{n}\#(x_i,r)$$

$$= \frac{1}{n(n-1)}\sum_{\substack{i,j \\ i\neq j}}^{n} H(x_i,x_j,r)$$

$$= \frac{2}{n(n-1)}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} H(x_i,x_j,r)$$

$$= \frac{2*(\text{Number of pairs closer than } r)}{n(n-1)}$$

Grassberger and Procaccia [1] call this expression the Correlation Integral, and they suggest that its scaling behavior relative to r be

examined to determine the dimension of A. Denoting the correlation integral associated with an n point sample by C(n,r), they define the Correlation Dimension d of A to be the real number d such that for small r and large n, C(n,r) scales as r^d.

To see how this d can be calculated, suppose that C(n,r) does indeed scale as r^d, i.e. suppose $C(n,r) = K\, r^d$ for some constant K. If r_1 and r_2 are two r values for which the correlation integral has been calculated, then $C(n,r_1) = K\, r_1^d$ and $C(n,r_2) = K\, r_2^d$. Taking natural logarithms of each side of these two equations yields the simultaneous equations $\ln C(n,r_1) = \ln K + d \ln r_1$ and $\ln C(n,r_2) = \ln K + d \ln r_2$. Thus $d = (\ln C(n,r_2) - \ln C(n,r_1)) / (\ln r_2 - \ln r_1)$. So one approach to estimating d is to plot several points on the graph of ln C(n,r) vs ln r and extract the slope of the linear part of the graph.

3. Examples

To further clarify the procedure for experimental estimation of correlation dimension and to verify its effectiveness, consider first a unit circle in R^2. Since the unit circle is intuitively a one dimensional structure, one would expect a correlation dimension of 1 to within reasonable experimental error. Five thousand randomly generated points on the unit circle were used to generate the graph of ln C(n,r) vs ln r shown in figure 1. The slope of the linear part of the graph is 1.00 to two decimal digits of accuracy.

As a second example, consider a well-known fractal, the Sierpinski triangle shown in figure 2. The fractal dimension of this set can be analytically calculated [7] and is known to be ln 3 / ln 2 = 1.585. Five thousand sample points uniformly distributed over the Sierpinski

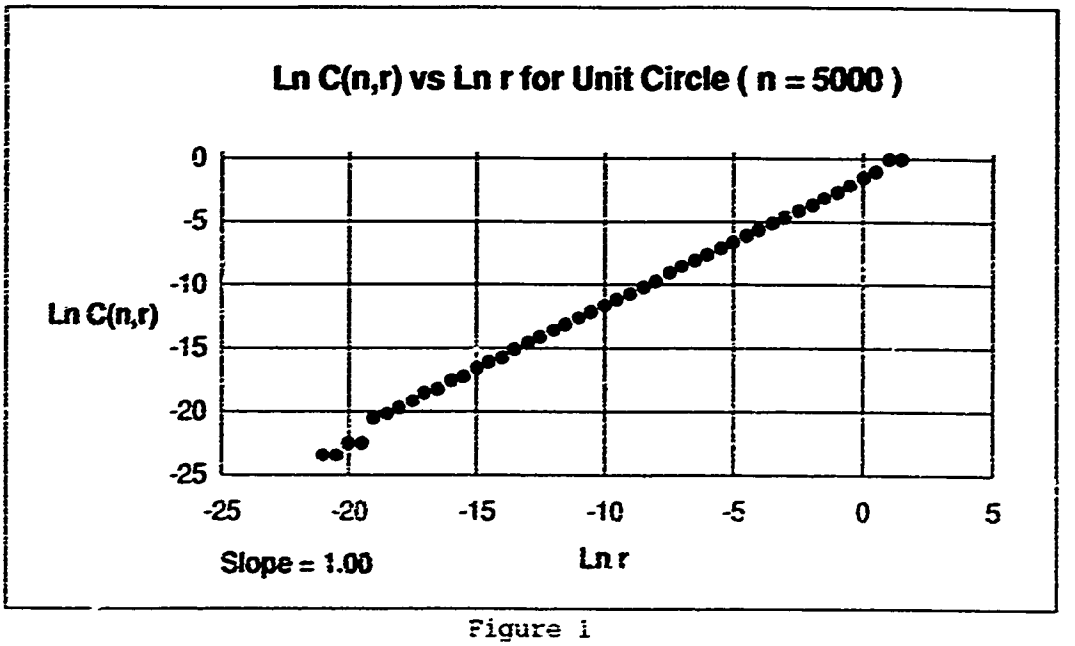


Figure 1

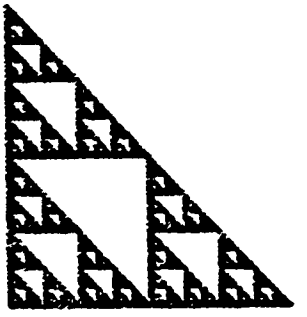


Figure 2

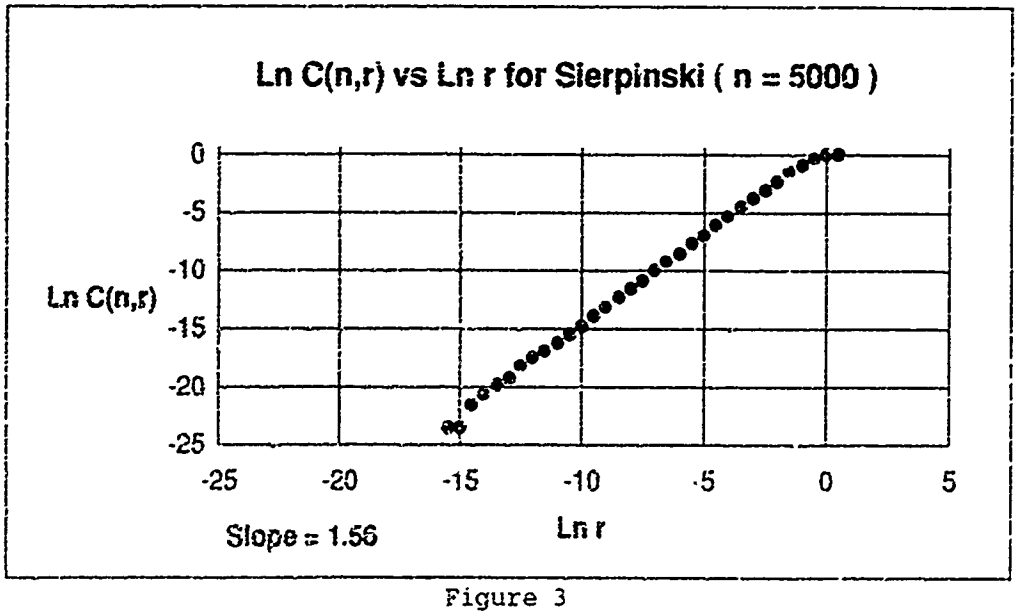


Figure 3

triangle yields the ln C(n,r) vs ln r graph shown in figure 3. The slope of the linear part of the graph is 1.56 to two decimal places of accuracy.

4. Dynamical Systems

A dynamical system with k degrees of freedom can be mathematically represented by R^k along with a family of functions $f_t: R^k \rightarrow R^k$ that map the current state, represented by a k-tuple x, into its future state t time units from now, $f_t(x)$. The functions f_t can be given explicitly or can be determined by differential equations governing the motion of the system. A dynamical system is called nonlinear if in general $f_t(c_1x_1 + c_2x_2) \neq c_1 f_t(x_1) + c_2 f_t(x_2)$. If the volume of $f_t(B)$ tends to zero as $t \rightarrow \infty$ for an arbitrary bounded subset B of R^k, then the dynamical system is called dissipative. As was noted in the introduction, many dissipative dynamical systems have strange attractors. A strange attractor A is a fractal subset of the phase space R^k, toward which the system evolves. If x is a phase space point sufficiently near A then $f_t(x)$ will be drawn toward A and then exhibit highly irregular motion remaining close to the attractor. By calculating the dimension of A, one can distinguish between irregular motion due to a low dimensional attractor and irregular motion which is essentially stochastic in a high dimensional phase space.

In the next section, we will begin to examine dynamical systems with infinite dimensional phase spaces. It is no surprise that complex, irregular motion can be observed in such systems with their infinite number of degrees of freedom. However, using the correlation dimension

as our tool, we will see that the irregular motion is not stochastic, but is rather associated with low dimensional attractors.

5. Mackey-Glass Equation

Mackey and Glass [9] have studied many physiological processes in which irregularities due to chaos are manifest. One such process is the production of white blood cells in humans. The rate of production of new cells is dependent both on the current levels of such cells and on the levels which existed T days ago. More precisely, the quantity of white blood cells is governed by the following differential delay equation:

$$\frac{dx(t)}{dt} = \frac{0.2x(t - T)}{1 + x^{10}(t - T)} - 0.1x(t)$$

To specify an initial condition for this system, one must define x(t) throughout the infinite closed interval [0,T]. Thus the system is infinite dimensional. To generate values of the function x(t), a predictor-corrector technique [8] can be used to numerically integrate the Mackey-Glass equation. Except where noted, the initial condition x(t) = 2.0 and a step size of 0.1 were used in the predictor-corrector algorithm that produced the data for this report. Selecting a starting time t_0, a sampling interval S, and a sample size N, one can generate a time series $x(t_0)$, $x(t_1)$, ..., $x(t_{N-1})$ where $t_i = t_0 + iS$ for i = 0, 1, 2, ..., N-1. Graphs of the time series for t_0 = 1000, S = 1, N = 500, and T = 10, 20, and 100 are shown in figure 4.

Packard, Crutchfield, Farmer, and Shaw [10] have developed a procedure which permits the recovery of the essential structure of a strange attractor of a dynamical system from the time series of a single

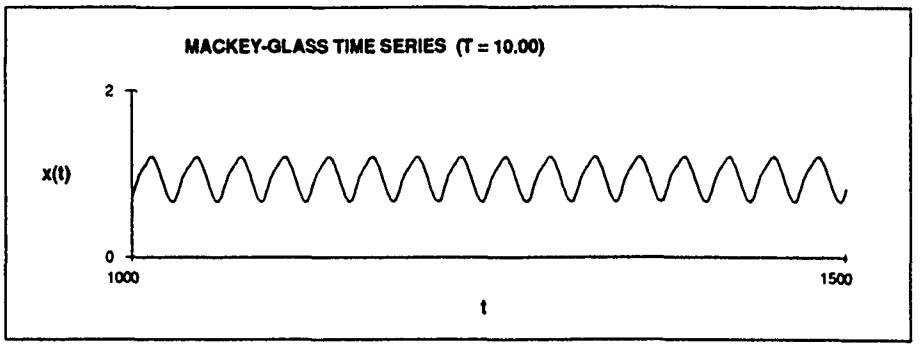


Figure 4a

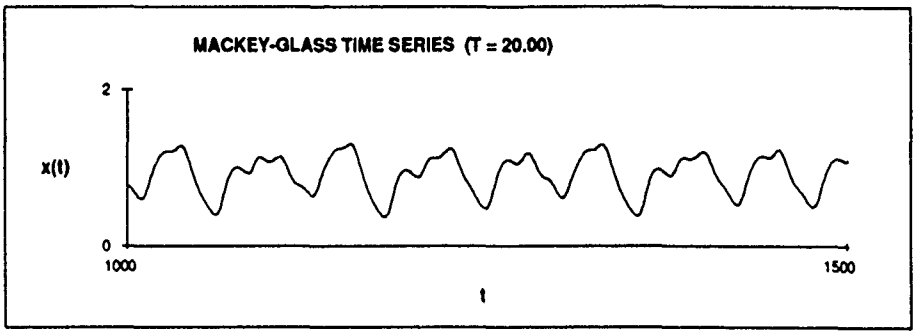


Figure 4b

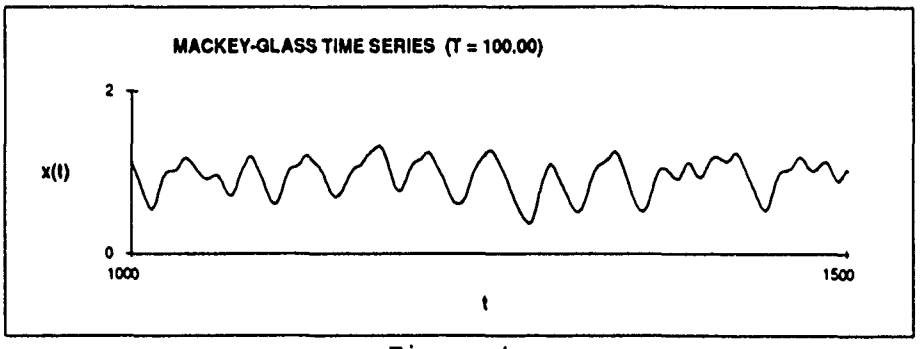


Figure 4c

state variable of the system. In the context of the Mackey-Glass time series, select an embedding dimension k and then build N-k+1 points in R^k as follows:

$$X_1 = (x(t_0), x(t_1), \ldots , x(t_{k-1}))$$

$$X_2 = (x(t_1), x(t_2), \ldots , x(t_k))$$

$$X_3 = (x(t_2), x(t_3), \ldots , x(t_{k+1}))$$

$$\vdots$$

$$X_{N-k+1} = (x(t_{N-k}), x(t_{N-k+1}), \ldots , x(t_{N-1}))$$

For sufficiently large embedding dimensions and large enough sample sizes, the structure of the finite set { x_1, x_2, ... , x_{N-k+1} } will approximate that of the strange attractor of the system. In particular the dimension of the attractor can be estimated by calculating the correlation dimension using this finite set.

6. Cautions in Estimating Correlation Dimension

An implementation of the correlation dimension estimation procedure described in section 2 requires a decision on the r values for which the correlation integrals will be calculated. Simply making an arbitrary selection of a small set of r values is problematic. The estimated diameter of the attractor needs to be considered. Also, when the embedding dimension is large, the interpoint distances tend to be proportionately large, and so an appropriate set of r values might be very different from those which work well for lower embedding dimensions.

As noted in the introduction, one objective of this research project was to find an explanation for unexpected discontinuities in estimated dimension encountered in a preliminary study of Mackey-Glass data in late 1990. In that study, a fixed set of r values, namely { 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7 } was used in calculations based on embedding dimensions of two through eleven. For each embedding dimension, the slope calculation (ln C(n,0.7)-ln C(n,0.1))/(ln 0.7-ln 0.1) was used as the correlation dimension estimate. (Here n is the number of points in the embedding space; i.e. n = N - k + 1 where N is the number of time series samples and k is the embedding dimension--see the last paragraph of section 5.) Finally, a weighted average of the

correlation dimension estimates for each of the ten embedding dimensions was computed with the estimates associated with the higher embedding dimensions receiving the greater weights.

To understand the source of the discontinuities, consider the following actual data from the preliminary study for T = 60 and embedding dimension k = 11:

case 1) N = 1860, n = 1860 - 11 + 1 = 1850.

# of pairs closer than 0.1 = 1
# of pairs closer than 0.7 = 197399

c.d. est. = 6.27 (for embedding dimension = 11)

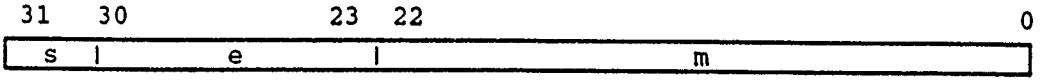
case 2) N = 1870, n = 1870 - 11 + 1 = 1860.

# of pairs closer than 0.1 = 5
# of pairs closer than 0.7 = 198460

c.d. est. = 5.44 (for embedding dimension = 11)

Thus adding a mere ten additional time series values to the sample has resulted in a dramatic drop in the correlation dimension estimate. The problem is clear: a combination of small sample size and high embedding dimension has resulted in there being too few pairs closer than 0.1. From a graphical point of view, a slope estimate has been made using a point that is not on the linear part of the log log graph. Until the number of pairs closer than 0.1 has grown to a representative level, the point corresponding to r = 0.1 is of no value in a slope estimation procedure.

To avoid the problems inherent in r value choice, the algorithms used in this project employ a suggestion of Parker and Chua [5] which exploits the exponent/mantissa format used by computers to store real numbers. In IEEE single precision floating point format (which Turbo Pascal implements with the data type SINGLE), real numbers are represented as 32-bit strings:



The value represented is $v = (-1)^s * 2^{(e-127)} * (1.m)$. If x is a variable of type SINGLE in Turbo Pascal, then the biased exponent can be extracted as follows:

```
bit_string := ( longint ) x;
biased_exponent := ( bit_string SHR 23 ) AND $000000FF;
```

If, for example, the biased exponent was 124, we could conclude that $2^{-3} \leq x < 2^{-2}$ without having to even consider the mantissa or do any comparison operations. These observations suggest that we classify interpoint distances according to their biased exponents, effectively using a range of r values from 2^{-127} to 2^{+127}. In fact, the interpoint calculations can be made even more efficient by looking at biased exponents of the squared interpoint distances, thus avoiding expensive square root evaluations. A further advantage of this approach is that since the r values are powers of two, the points on the log log graphs will have uniform horizontal separation. A detailed pseudocode description can be found in [5].

7. Computation

If a sample of N time series values is used, and all interpoint distances are computed, the time complexity of the correlation dimension algorithm will clearly be $O(N^2)$. Improvements in efficiency can be obtained by recalling that the estimation procedure is based on the assumption that the r values are small. Thus we only need to calculate those interpoint distances that are relatively small. Our algorithms exploit this strategy in two ways.

First, after translating and rescaling the time series values so that the points in the embedding space will be in a unit hypercube, we use a variation of an algorithm due to Theiler [3] to sort the embedding space points into boxes so that all points closer than some threshold distance r_0 will be in the same box or neighboring boxes. To clarify the sorting technique, consider the following 3 points in a k = 7 dimensional embedding space:

P1 = (0.71,0.23,0.95,0.89,0.36,0.53,0.43)
P2 = (0.32,0.86,0.30,0.53,0.92,0.04,0.55)
P3 = (0.89,0.32,0.64,0.74,0.64,0.21,0.80)

If a box resolution of 4 is specified, we divide a unit square into a 4x4 grid and place the points into a box based on the value of the first and seventh coordinate of the point (see figure 5). Only points in the same or neighboring boxes can possibly be within r_0 = 1/4 of each other, and thus we can avoid many interpoint distance calculations. For example, the distance between P_1 and P_3 above need not be calculated.

As a second approach to minimizing distance calculations, we associate bit-strings with each point in the embedding space so as to partially extend the above sorting strategy to the other k-2 dimensions

of the unit hypercube. Applying the function tag(x) = floor(4 * x) to each of the intermediate components (components two through six) of the point P_1 yields tag(0.23) = 00_2, tag(0.95) = 11_2, tag(0.89) = 11_2, tag(0.36) = 01_2, tag(0.53) = 10_2. Concatenating these two-bit tags produces a ten-bit string which, along with the sort described in the previous paragraph, resolves the point P_1 into a box of side length 1/4 in seven dimensional space. Before doing a full-blown floating point distance calculation the bit strings associated with two points can be processed using shifting, masking, and integer subtraction operations to see if there is actually a potential for the points to be closer than the threshold distance r_0. In this way many expensive floating point operations can be circumvented. To limit memory expenditure for the bit string storage, our algorithms use a uniform sample of the intermediate components rather than all the intermediate components when there is a combination of high embedding dimension and high box resolution.

A significant savings in computing time resulted from employing these techniques. For example, calculating correlation dimension using 10,000 time series samples required 3 hours and 19 minutes using a

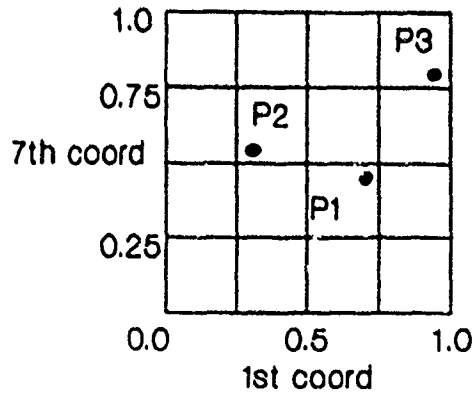


Figure 5

standard algorithm unassisted by the box sort and tags. Using the box sort and tags with box resolution = 8 (so $r_0 = 1/8$), the same calculation took 10 minutes and 53 seconds. Increasing the box resolution to 64 (so $r_0 = 1/64$) further reduced the computing time to 35 seconds. Computations were done on a 16 MHz 80386 machine with a coprocessor.

8. Processing Interpoint Distance Data

Having addressed the issue of computational efficiency in the calculation of interpoint distances, we turn now to the issue of how to extract the correlation dimension from that distance information. Although the graphical approach of section 2 is used extensively in the literature, researchers say little about how one systematically chooses the most linear part of the log log graph. Parker and Chua [5] in fact report that there is currently no robust technique for determining the useful part of the graph. Theiler [2] specifically cautions against the dangers inherent in using a least squares linear fit.

In an effort to develop a consistent strategy which does not depend on ad hoc visual analysis of graphs, two systematic estimates of correlation dimension, called the MinVar and the Pairwise estimates were developed and applied to a dimensional study of the Mackey-Glass equation, the results of which are reported in the next section.

With each point in the log log graph, except the rightmost, associate a local slope computed by using the point and its immediate neighbor to the right. Fix a small integer m, and compute the variance of the local slopes for each collection of m local slopes. For the

collection which has the minimum variance, report the average of the collection as the MinVar estimate of correlation dimension.

The Pairwise estimate is motivated by the observation that when a threshold distance r_0 is being employed, as in the box oriented algorithm described in section 7, the most linear portion of the log log graph tends to be at the rightmost portion of the graph. For a fixed integer m, compute a slope using each possible pair of points chosen from the rightmost m points. Report the average of this collection of m(m-1)/2 slopes as the Pairwise estimate of correlation dimension.

A third systematic way to extract a correlation dimension estimate from interpoint distance data is due to Takens [11]. Denoting the interpoint distance between point i and point j by r_{ij}, he suggests calculating the average of $\ln(r_{ij} / r_0)$ over all interpoint distances which are less than r_0, and then estimating the correlation dimension by the negative reciprocal of this average. Takens used the theory of best estimators to develop this method. Its connection to slope-based methods is described in [2].

Since the Takens estimate requires a natural logarithm calculation at each interpoint distance below the threshold r_0, it is computationally more expensive than the slope-based methods. In an effort to get a Takens estimate without additional computational expense, we have developed an estimate called QuickTakens, which is computed as follows. In our slope based algorithm, we obtain counts of pairs for which the logarithm of interpoint distance lie in intervals of the form $I_a = [(a-0.5) \ln 2, a \ln 2)$. Denoting the count associated with I_a by C(a), we can approximate the sum of $\ln(r_{ij} / r_0)$ for those

r_{ij} for which $\ln(r_{ij})$ lies in I_a by the expression $C(a)\ \ln(\ r^*/r_0\ )$ where r^* is a number in the interval $(2^{a-0.5}, 2^a)$. Summing such expressions over all the intervals I_a below the threshold yields an estimate of the sum of $\ln(\ r_{ij}\ /\ r_0\ )$ over all interpoint distances which are less than r_0, and hence the average required in the Takens estimate can be approximated. In the results to be reported in the next section, we obtained excellent agreement between the Takens estimate and the Quick-Takens estimate by choosing r^* so that $\log_2(r^*) = a-0.25+0.013*\text{MinVar}$.

9. Results

In [6], Farmer undertook a systematic analysis of chaotic attractors associated with the Mackey-Glass equation. His measures of degree of chaos were based on Lyapunov exponents and their conjectured relationship to the dimension of strange attractors. Farmer's work predated the introduction of the correlation dimension by Grassberger and Procaccia.

In this section we report the results of a correlation dimension analysis of attractors associated with the Mackey-Glass equation. Figure 6 shows correlation dimension estimates for integer values of T between T = 16 and T = 45 inclusive. As noted in section 7, the time series values were rescaled so that the points in the embedding space would all be in a unit hypercube. An embedding space of dimension k = 7 was used, and in all computations we chose the sampling interval S to be the same as the delay time T. 10,000 time series samples were used, with sampling beginning at time t = 1000. The box resolution for all runs was 8, making the threshold interpoint distance $r_0 = 1/8$. We find reasonable agreement with the dimension calculations based on Lyapunov

exponents reported in [6], although correlation dimension is slightly less than the Lyapunov dimension for most values of T. The correlation dimension successfully detects low-dimensional attractors near T = 19, 21, 34 and 39. Although [6] implies that these low dimensional attractors are actually limit cycles of dimension exactly 1, our studies do not confirm that assertion. Correlation dimension calculations and phase map analyses suggest rather that they are merely very low dimensional chaotic attractors.

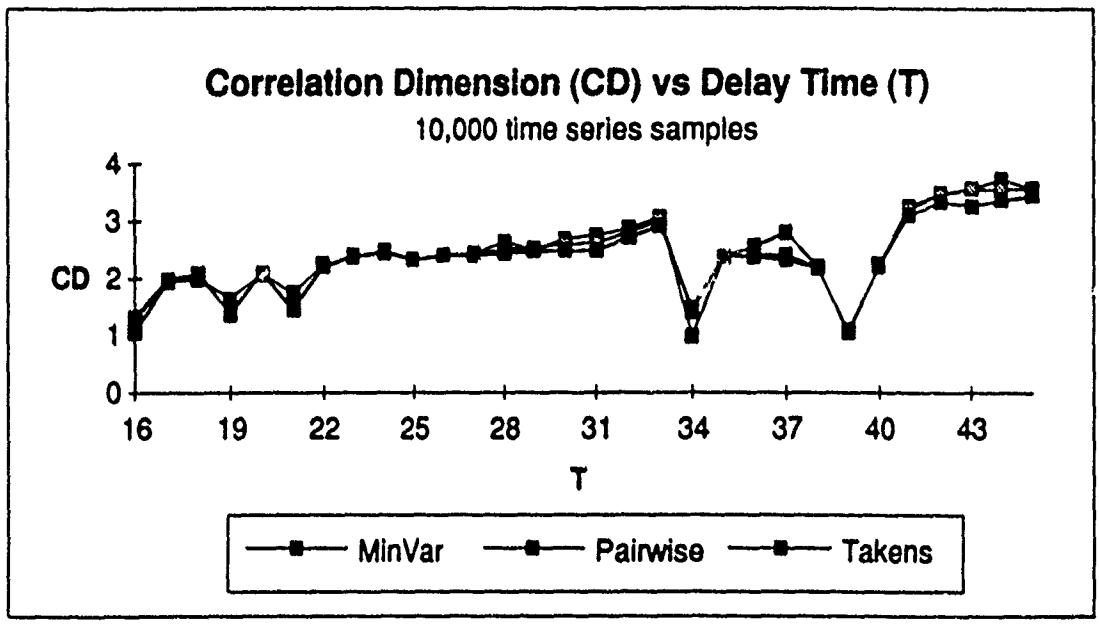


Figure 6

Figure 7 graphs the Takens estimate and the Quick-Takens estimate together to show the degree of agreement in these two estimates.

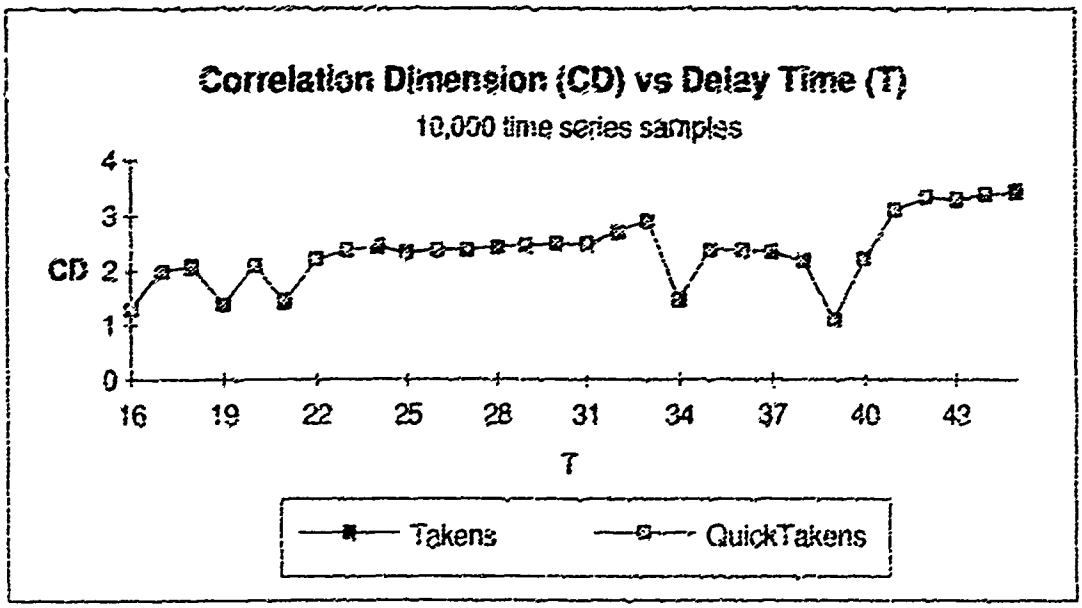


Figure 7

10. Conclusions and Suggestions for Additional Research

The results of this study confirm the effectiveness of the correlation dimension as a measure of chaos in dynamical systems. We have suggested some consistent ways of extracting correlation dimension from interpoint distance data, and have found the resulting estimates to be in reasonable agreement with each other and with findings reported in the literature.

In studying the Mackey-Glass data, we occasionally encountered log log graphs with shoulders which hampered identification of the linear part of the graph. Initially these shoulders were thought to be caused by autocorrelation in the time series data, but techniques for elimination of autocorrelation [4] were applied with no effect. Such shoulders are now thought to be a real effect, possibly indicating that the local dimension of the attractor varies significantly from point to point. Further study of local dimension may lead to a better understanding of this phenomenon.

Acknowledgments

I wish to thank the Air Force Systems Command and the Air Force Office of Scientific Research, whose sponsorship provided me this opportunity to participate in a stimulating research project. I also thank the staff of the Summer Research Program Office at Research and Development Laboratories for their careful and helpful handling of the administrative aspects of the program. Numerous individuals in the Avionics Directorate at Wright-Patterson AFB contributed to making my stay enjoyable and productive. Edward Gliatti, Maj. Robert Bennington, and their staff in the Information Processing Technology Branch were most hospitable and provided a very pleasant work environment. I am especially grateful to Dale Nelson, my effort focal point, for his assistance in identifying an interesting and challenging research problem, and for providing direction to my investigations throughout the Summer. Conversations with Dennis Buck, and his sharing of data from the earlier Mackey-Glass study were very helpful in the initial weeks of my work. I also thank Dennis Buck for many hours of meticulous work he devoted to reformatting and integrating the graphical images into this document. Dale Nelson, Eileen Lantz, Dale Van Cleave, and David Tsui also provided valuable technical assistance.

References

1. P. Grassberger and I. Procaccia. "Measuring the Strangeness of Strange Attractors," *Physica 9D*, 189 (1983).

2. J. Theiler, "Estimating the Fractal Dimension of Chaotic Time Series," *Lincoln Laboratory Journal 3*, 63 (1990).

3. J. Theiler, "Efficient Algorithm for Estimating the Correlation Dimension from a Set of Discrete Points," *Phys. Rev. A* ***36***, 4456 (1987).

4. J. Theiler, "Spurious Dimension from Correlation Algorithms Applied to Limited Time-Series Data," *Phys. Rev. A* ***34***, 2427 (1986).

5. T. Parker and L. Chua, "Chaos: A Tutorial for Engineers," *Proceedings of the IEEE* ***75***, 982 (1987).

6. J. D. Farmer, "Chaotic Attractors of an Infinite-Dimensional Dynamical System," *Physica* ***4D***, 366 (1982).

7. M. Barnsley, *Fractals Everywhere* (Academic Press, San Diego, 1988).

8. S. Conte and C. de Boor, *Elementary Numerical Analysis: An Algorithmic Approach* (McGraw-Hill, New York, 1972).

9. L. Glass and M. Mackey, *From Clocks to Chaos* (Princeton University Press, Princeton, N. J., 1988).

10. N. H. Packard, J. P. Crutchfield, J. D. Farmer and R. S. Shaw. "Geometry from a Time Series," *Phys. Rev. Lett.* ***45***, 712 (1980).

11. F. Takens, "On the Numerical Determination of the Dimension of an Attractor," in *Dynamical Systems and Bifurcations: Proceedings of a Workshop Held in Groningen, The Netherlands, 16-20 April, 1984, in Lecture Notes is Mathematics* ***1125*** (Springer-Verlag, Berlin, 1985), p. 99.

12. J. Stright, "A Neural Network Implementation of Chaotic Time Series Prediction", Thesis, Air Force Institute of Technology, December 1988.

1991 USAF-RDL SUMMER FACULTY RESEARCH PROGRAM/

GRADUATE STUDENT RESEARCH PROGRAM

Sponsored by the

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

Conducted by

Research Development Laboratories

FINAL REPORT

Fiber Laser Preamplifier for Laser Radar Detectors

Prepared by:	Richard E. Miers
Academic Rank:	Associate Professor
Department and	Physics
University:	Indiana University/Purdue University at Fort Wayne
Research Location:	WRDC/AARI-2 Wright Patterson AFB Dayton, OH 45433
USAF Researcher:	Paul F. McManamon
Date:	July 25, 1991

Acknowledgements

I wish to thank the Air Force Systems Command and the Air Force Office of Scientific Research for sponsorship of this research. The assistance in administrative and directional aspects of this program by Research Development Laboratories is acknowledged.

I appreciate the support provided by my laboratory focal point Dr. Paul F. McManamon and the personnel of AARI-2. The assistance of Lt. Scott McCracken and Mike Salisbury was extremely valuable to this project.

I would like to also express my appreciation for the provision of Nd-doped fibers by Professor Elias Snitzer of Rutgers University Fiber Optics Materials Research Program and Professor Ted Morse of Brown University Division of Engineering. The information and advice accompanying these fibers has been extremely useful to this project.

FIBER LASER PREAMPLIFIER FOR LASER RADAR DETECTORS

by

Richard E. Miers, Associate Professor of Physics

ABSTRACT

Nd-doped fiber laser amplifiers for incorporation into a laser radar test sytem were developed around two Nd-doped fibers provided by Rutgers University and Brown University. Both fibers exhibited a fluorescent band peaking at or near 1064 nm. A gain of 10 dB was measured in an amplifier incorporating the double-clad fiber provided by Rutgers University. Recommendations and designs for improved amplifiers using both fibers are given.

FIBER LASER PREAMPLIFIER FOR LASER RADAR DETECTORS

I. INTRODUCTION

The Electro-optics Division of the Avionics Laboratory at Wright-Patterson Air Force Base is involved in the development of laser radar systems. One possible method of increasing the detectablility of a returning laser radar signal might be to use a fiber optical laser preamplifier immediately before the photo detector. This type of amplifier shows promise as a means of increasing the signal to noise ratio of a laser radar system detector.

During the summer of 1990 I was a USAF Summer Faculty Research Associate working with the Electro-optics Division of the Avionics laboratory at Wright-Patterson AFB. The objective of that research was to study the feasibility of developing a fiber laser amplifier for use in their laser radar test system. Since the results of that study indicated that such an amplifier could be useful, a Research Initiation Proposal for the development of this type amplifier was funded by the Air Force Office of Scientific Research. My assignment as a particiant in the 1991 Summer Faculty Research Program (SFRP) was to continue development of this fiber laser preamplifier.

II. OBJECTIVES OF THE RESEARCH EFFORT

A Nd-doped silica fiber is a four-level laser medium. For a four-level amplifier the unsaturated single pass gain factor can be given as

$$\gamma = \frac{\sigma \tau_f P_{abs}}{h\nu_p A_p^*} \qquad (1)$$

where the gain is given by

$$G = \frac{I_{out}}{I_{in}} = e^{\gamma}. \qquad (2)$$

I_{out}/I_{in} is the ratio of amplified signal to the input signal, σ is the stimulated emission cross section for the amplified wave, τ_f is the flourescent lifetime of the upper lasing level, $h\nu_p$ is the energy per photon of the pump light, and P_{abs}/A^* is the effective intensity of the absorbed pump light in the fiber. [1]

For a Nd-doped silica fiber assuming $\sigma = 3 \times 10^{-20}$ cm^2, $\tau_f = 4.5 \times 10^{-4}$ s, $h\nu_p(800\text{ nm}) = 2.48 \times 10^{-19}$ J, $A_p^* \sim 5 \times 10^{-7}$ cm^2 gives a slope efficiency of $\gamma/P_{abs} = 0.11/\text{mW}$ or $G(\text{dB})/P_{abs} = 0.47$ dB/mW. Po, et al. reported a slope efficiency of 0.437 dB/mW in a Nd-doped silica fiber when pumped at 800 nm. [2]

My assignment as a participant in the 1991 Summer Faculty Research Program (SFRP) was to construct a fiber laser amplifier using optimum amplification parameters and minimum noise to be integrated into a test

laser radar system at the Electro-optics Division of the Avionics Laboratory at Wright-Patterson AFB. This amplifier would use a Nd-doped silica fiber with a single mode core pumped with a laser diode at around 819 nm wavelength. The signal to be amplified would be a 1064 nm wavelength from a Nd-YAG laser.

III.

Fiber laser amplifiers appear to have advantages over other optical amplifiers. Fibers have been developed to transmit a single mode signal. These fibers are polarization independant. Laser diodes in the wavelength range of 800 nm to 850 nm are readily available for end pumping. Rare-earth-doped fibers have been developed for amplification at a number of laser wavelengths.

Two Nd-doped fibers were obtained for development of this amplifier. One fiber produced at Brown University, Division of Engineering has a core of 8 μm diameter with a cladding diameter of 122 μm. The refractive index of the core is given as 1.4618. The cladding index is given as 1.4566 resulting in a numerical aperture (NA) of 0.123. An absorption profile for this fiber provided by Brown University shows a sharp absorption peak at a wavelength of 840 nm with maximum absorption at this wavelength of 3 dB/m of fiber length. This absorption drops to near 0 dB/m at 800 nm and at 860 nm. At 820 nm and 850 nm the absorpion is approximately 1 dB/m.

The second fiber was fabricated by Rutgers University Fiber Optic Materials Research Program. This is a double clad fiber similar to the one described in reference 2. The NA between the core and first cladding is given as 0.15. The NA between the first cladding and the second cladding is 0.4. The first cladding is rectangular in shape with dimensions 45 μm by 110 μm. The core is slightly elliptical with dimensions of 5 μm and 3.3 μm on the major and minor axes. The design of this fiber allows efficient coupling of diode laser pump light by focussing the collimated laser diode light into the first cladding. Efficient coupling of diode light into this fiber can be accomplished through butt-coupling or the use of simple optics. Researchers have reported up to 86 % of diode pump light coupled into the core of similar fibers by pumping into the first cladding. Absorption curves for this second fiber show peak absorption of 0.8 dB/m at a wavelength of 805 nm and 0.3 dB/m at 820 nm. There is also a loss of 17.4 dB/km at 1060 nm due to the Nd-doping.

A laser diode was purchased from Laser Diode Technologies capable of producing 140 mW maximum power output at a wavelength of 819 nm. This laser is a 6 μm by 1 μm single mode source. The output from the laser diode was collimated using a 0.14 pitch gradient index (GRIN) lens. A Newport high-precision single mode fiber coupler, F-1015LD, was used to couple the collimated laser power into a fiber. This coupler includes a steering lens AR-coated for high transmittance at 850 nm and a Newport F-L10 diode laser focusing lens with effective focal length of 12 mm.

The maximum power measured in the collimated laser beam was 90 mW. Using a 0.25 m length of the Brown University fiber and the F-1015LD coupler with this collimated laser diode resulted in a measurement of 100 μW maximum power coupled into this fiber. The same experimental arrangement using a 0.25 m length of the Rutgers University double-clad fiber resulted in a maximum of 40 mW of power coupled into the fiber.

The flourescence generated by the coupled pump light was measured using a Jarrell-Ash quarter meter monochrometer with a grating blazed at 1.0 μm wavelength. Both fibers exhibited a band of fluorescence that peaked near 1064 μm wavelength.

An amplifier was constructed using a 29 m length of the Rutgers University fiber. A block diagram of this amplifier is shown in Figure 1. A dichroic mirror purchased from CVI Laser Inc. was placed at an angle of 45° with the axis of the fiber end. This mirror has high reflectivity (> 98 %) for 820 nm wavelength and high transmitance (>85 %) for 1064 nm wavelength. The laser pump light and an attenuated signal light from a Nd-YAG laser were simultaneously coupled into the fiber using this mirror as shown. When approximately 40 mW of pump power was coupled into the fiber a gain of slightly over 10 dB in the 1064 nm light signal was measured.

The Rutgers University fiber was replaced with a 1.0 m length of the Brown University fiber in the amplifier configuration described above. The amount of pump light coupled into the fiber was not readily measurable.

An apparent high absorption resulted in the flourescent light masking the pump light when measured directly by the power meter. When measured with the Jarrell-Ash monochrometer separating the laser output signals, only a small amount of pump light was measurable compared with the fluorescent light at 1064 μm. However a signal gain of greater than 0.5 dB was measured for signals as high as several hundred μW power. This indicates that at least about 100 μW of pump power was absorbed.

An amplifier has been designed around the Brown University fiber. A block diagram of this amplifier is shown in Figure 2. A 100 mW laser from Spectra Diode Labs will be pigtailed into a wavelength division multiplexer (WDM) by Seastar Optics, Inc.. This WDM has been ordered from Amphenol Corporation. This device will be custom tuned to couple the diode laser pump light from input 1 with the 1064 nm signal light from input 2 into output 1. Output 1 will be spliced to the Nd-doped gain fiber. Seastar Optics guarantees that a minimum of 45 mW of the 100 mW diode laser light will be coupled into the single mode inputs of the WDM.

An increase in gain from the Rutgers University fiber amplifier is expected if one replaces the present pump laser with a multimode diode laser having higher power and a lasing wavelength at the maximum of the absorption curve.

IV. RECOMMENDATIONS AND CONCLUSIONS

Recommendations for continuing research on these amplifiers include the fabrication of the amplifier shown in Figure 2 using the Brown University Nd-doped fiber. An improvement of the amplifier shown in Figure 1 using the Rutgers University Nd-doped fiber should result from replacing the present diode laser with a multimode 500 mW diode laser having wavelength centered at 805 nm to allow more efficient absorption and higher gain.

These amplifiers should be tested using both direct detection and coherent detection of amplified signals and by making relative signal-to-noise measurements. These measurements should then be compared with measurements obtained using unamplified signals alone. Improvements in signal-to-noise in the amplified signal may be necessary through proper narrow band filtering for direct detection and elimination of all Rayleigh reflections through isolation and AR coatings of all unspliced connections.

Researchers appear to have been able to produce Er-doped fiber laser amplifiers with noise factors near the quantum limit. [3] [4] [5] It seems reasonable to assume that similar techniques applied to the development of Nd-doped fiber amplifiers will produce efficient noise-free gain for signals at the 1.06 μm wavelength.

References

[1] M. J. F. Digonnet and C. J. Gaita, "Theoretical Analysis of Optical Fiber Laser Amplifiers and Oscillators," Appl. Optics, vol 24, pp 333-342, 1987.

[2] H. Po, F. Hakimi, R. J. Mansfield, B. C. McCollum, R. P. Tumminelli, E. Snitzer, "Neodymium Fiber Laser at 0.905, 1.06, and 1.4 μm, "Abstracts of Annual Meeting of Optical Soc. of America, Seattle, WA (1986) Paper FD4, P 103.

[3] C. R. Giles, E. Desurvire, J. L. Zyskind, and J. R. Simpson, "Noise Performance of Erbium-doped Fiber Amplifier Pumped at 1.49 μm, and Application to Signal Preamplification at 1.8 Gbits/s," IEEE Photon. Technol. Lett., Vol 1, pp 367-369, 1989.

[4] R. Olshansky, "Noise Figure For Erbium-Doped Optical Fibre Amplifiers," Electron. Lett., Vol 24, pp 1363-1365, 1988.

[5] M. Yamada, M. Shimizu, M. Okayasu, T. Takeshita, M. Horiguchi, Y. Tachikawa, and E. Sugita, "Noise Characteristics of Er^{3+} doped Fiber Amplifiers Pumped by 0.98 and 1.48 μm Laser Diodes," IEEE Photon. Technol. Lett., Vol 2, pp 205-207, 1990.

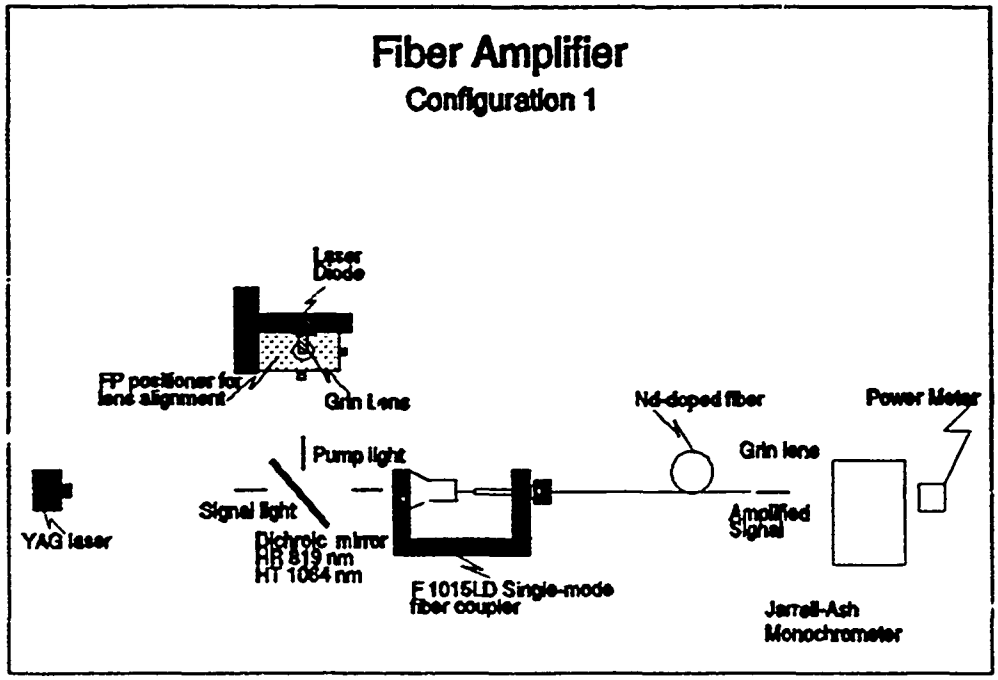


Figure 1

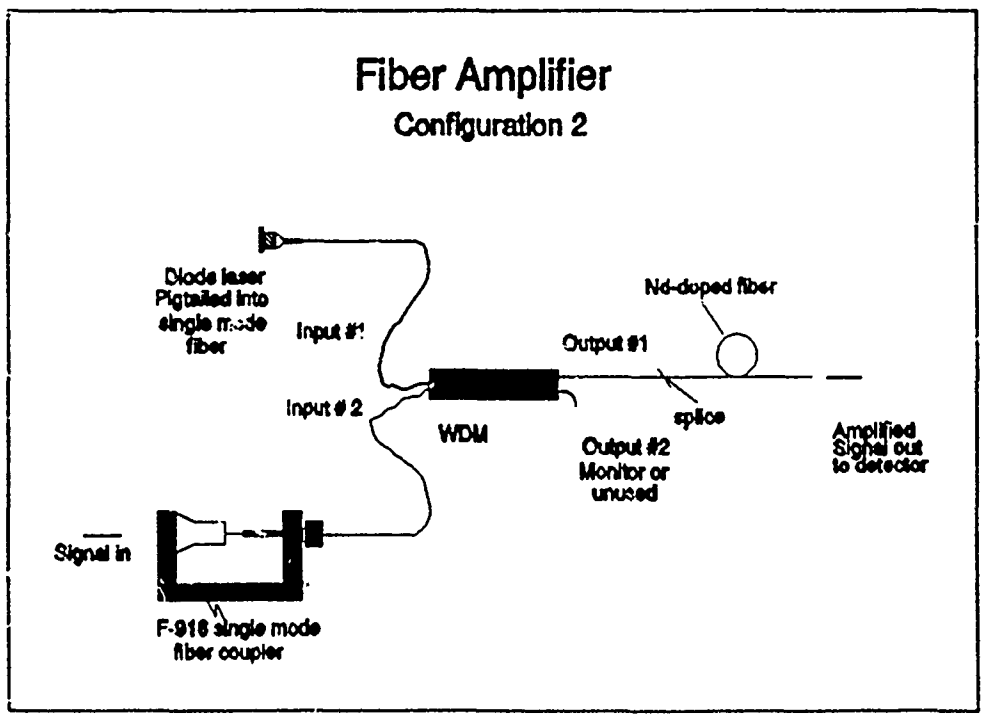


Figure 2

A METHODOLOGY FOR EMPLOYING MODULATION QUALITY FACTORS IN THE ANALYSIS OF LPI WAVEFORMS

Glenn E. Prescott*
Lawrence L. Gutman**

Abstract

LPI system quality factors were described in a previous effort [1] in order to provide a quantitative analysis tool for the system engineer to employ in evaluating the effectiveness of LPI techniques in the presence of jammers and intercept receivers. These LPI system quality factors were derived from the system link equations which describe the signal power gains and losses as a function of system link parameters. In this paper, we focus on the issue of detectability of LPI waveforms as defined by the modulation quality factor. The LPI modulation quality factor is a measure of the covertness of a particular type of modulation when detection is attempted by a particular type of intercept receiver. Intercept receiver detection models are provided for wideband nonlinear feature detectors. The utility of this quality factor is illustrated by an example and performance curves demonstrate the concept.

Introduction

A conventional communications link analysis is based on a power budget approach which typically originates at the transmitter and accounts for signal gains and losses across a propagation path. An LPI analysis, on the other hand, typically begins at the communications receiver by specifying some minimum performance requirement (usually probability of error, P_e). From this performance requirement a minimum input signal-to-noise ratio is obtained. Link computations are then used to determine the maximum range over which the transmitter and receiver can communication with the specified level of performance. In an LPI link, it may be essential that the transmitted power be adjusted to allow communications at the required level of performance with little or no margin, since excess power margin gives an advantage to the interceptor. The interceptor link is likewise analyzed beginning at the receiver. To evaluate the effectiveness of a particular intercept receiver in detecting a transmitted signal, some acceptable P_d and P_{fa} are initially specified. From these performance specifications, and the signal processing characteristics of the intercept receiver, a minimum required input signal-to-noise ratio is obtained. Given a knowledge of the scenario dependent link parameters (antenna gains, transmitter power, path loss, etc), the intercept range is then adjusted to achieve the minimum required input signal to noise ratio. The interceptor seeks to maximize the intercept range, while the communicator seeks to minimize it. The LPI quality factor (Q_{LPI}) is simply a measure of the ratio of the communication

* *University of Kansas Telecommunications and Information Sciences Laboratory.*
** Wright Research and Development Center, Wright-Patterson AFB, OH.

link range to the interceptor range. The goal of the LPI communication system is to require the interceptor to be closer to the transmitter than the communications receiver in order to detect the transmitted signal (i. e., $Q_{LPI} > 0$ dB).

The LPI scenario is illustrated in Figure 1. At the communications receiver, the required signal power to noise power spectral density is usually less than that received by some factor M, the link margin:

$$\left(\frac{S_c}{N_{oc}}\right)_{rcvd} = M\frac{S_c}{N_{oc}} \tag{1}$$

where $M \geq 1$. The transmitter power and communications link parameters which influence the received signal power can be expressed as follows:

$$M\frac{S_c}{N_{oc}} = \frac{P_t G_{tc} G_{ct}}{L_c N_{oc}}\left(\frac{\lambda}{4\pi R_c}\right)^2 \tag{2}$$

Therefore, the maximum attainable range for the communications link is:

$$R_c = \sqrt{\frac{P_t G_{tc} G_{ct}}{L_c N_{oc}}\left(\frac{\lambda}{4\pi}\right)^2 \frac{1}{M(S_c/N_{oc})}} \tag{3}$$

For the interceptor link, some specified minimum acceptable P_d and P_{fa} will require a particular input signal-to-noise ratio for each type of intercept receiver. The interceptor link can be expressed as:

$$\frac{S_i}{N_{oi}} = \frac{P_t G_{ti} G_{it}}{L_i N_{oi}}\left(\frac{\lambda}{4\pi R_i}\right)^2 \tag{4}$$

This required S_i/N_{oi} for some P_d and P_{fa} is reflected through the link equation resulting in some *maximum* intercept range (i.e., beyond this range, interception is unlikely), R_i:

$$R_i = \sqrt{\frac{P_t G_{ti} G_{it}}{L_i N_{oi}}\left(\frac{\lambda}{4\pi}\right)^2 \frac{1}{(S_i/N_{oi})}} \tag{5}$$

Taking the ratio of ranges, we can express the communications link versus the interceptor link as:

$$\left(\frac{R_c}{R_i}\right)^2 = \frac{G_{ct}G_{tc}}{G_{ti}G_{it}}\frac{L_i}{L_c}\frac{N_{oi}}{N_{oc}}\frac{S_i/N_{oi}}{S_c/N_{oc}}M \tag{6}$$

LPI System Quality Factors

The previous expression for the ratio of ranges (6) allows a comparison of the effectiveness of the LPI communication system in gaining an operating range advantage over the interceptor. Since the goal of the LPI communication system is to maximize the ratio R_C/R_i, we therefore base our definition of the LPI quality factor (expressed in decibels) on this ratio,

$$Q_{LPI} = 20 \log\left(\frac{R_c}{R_i}\right) \tag{7}$$

The scenario dependent parameters, which include antenna gains and path losses can each be defined in terms of quality factors as described in [1]. For the present analysis it is sufficient to allow these to be fixed, so we can concentrate on waveform detectability. Furthermore we will make the simplifying assumption that the noise power spectrum for both the communications receiver and the intercept receiver is gaussian. Therefore, the LPI equation reduces to:

$$\left(\frac{R_c}{R_i}\right)^2 = \beta \frac{S_i/N_{oi}}{S_c/N_{oc}} \tag{8}$$

where

$$\beta = \frac{G_{ct}G_{tc}}{G_{ti}G_{it}} \frac{L_i}{L_c} \frac{N_{oi}}{N_{oc}} M \tag{9}$$

The Modulation Quality Factor

Regardless of the scenario, the relative performance of the communications receiver compared to the intercept receiver is of critical importance. The communications receiver has the advantage of a priori knowledge of the signal waveform, which gives the communicator a significant signal processing advantage over the interceptor. On the other hand, since the interceptor only requires the detection of signal energy for a successful intercept, the interceptor requires far less signal information (and hence, less input signal-to-noise ratio) than does the communicator. The goal of the LPI system designer is to select a waveform that is effective for communication purposes but is relatively difficult for the intercept receiver to detect.

We will focus our evaluation on the modulation quality factor, which is defined as:

$$Q_{MOD} = \frac{S_i/N_{oi}}{S_c/N_{oc}} \tag{10}$$

The input signal-to-noise ratio for the communication receiver can be expressed in normalized form as,

$$\frac{S_c}{N_{oc}} = \frac{E_b}{N_{oc}} r_d \tag{11}$$

Where E_b/N_{OC} is a function of the required P_e for a given modulation type and method of detection, which is defined as

$$\frac{E_b}{N_{oc}} = \zeta_c (P_e) \tag{12}$$

Therefore,

$$\frac{S_c}{N_{oc}} = \zeta_c (P_e) r_d \tag{13}$$

For the intercept receiver, the input signal to noise ratio is a function of P_d, P_{fa} as well as the integration time and bandwidth,

$$\frac{S_i}{N_{oi}} = \zeta_i (P_d, P_{fa}, T, W) \tag{14}$$

So that

$$Q_{MOD} = \frac{\zeta_i (P_d, P_{fa}, T, W)}{\zeta_c (P_e) r_d} \tag{15}$$

In order to use the modulation quality factor in an LPI analysis, we need to have the appropriate expressions for ζ_i and ζ_c. These will be discussed in the succeeding paragraphs.

Signal Detection Models

A simple detection model for LPI signals (such as suggested in [2]) is often useful in visualizing the complexities of the waveform and providing insight in how to best intercept it. The model employed here describes the frequency bandwidth versus time duration of the transmitted signal. This model can be used to represent the transmitted waveform as a hierarchy of time-bandwidth units, proceeding from a course structure, which typically has a large degree of complexity (large time-bandwidth product), to progressively finer structures which eventually approach time-bandwidth products on the order of unity.

The LPI signal chosen for the present analysis is a hybrid spread spectrum waveform consisting of direct sequence, frequency hopping and time hopping. The detection model for this signal structure is illustrated in Figure 2.

Intercept Receiver Models

The intercept receiver models considered in this analysis are assumed to perform some nonlinear operation on the received signal for the purpose of extracting signal energy, or some other waveform feature. We will assume that the received signal to noise ratio at the intercept receiver is small. This is usually a valid assumption since our LPI quality factor is based on the maximum communication range versus the maximum interception range. At the maximum interception range the signal to noise ratio will be very small. Under these conditions, (and assuming a suitably designed waveform) the optimum receiver has been shown to be a wideband total power radiometer [5]. If feature extraction is the goal of the interceptor, a higher signal to noise ratio is required resulting a smaller intercept range. This causes the LPI quality factor to increase significantly.

For signals having large time-bandwidth products, the output statistics of the radiometer can be assumed to be gaussian, and the detector performance can be completely characterized by the detectability factor d, which has been defined as the square of the difference in the means of the output densities under noise and signal plus noise conditions [3]. The detectability factor d, is a measure of the post detection, or output signal-to-noise power ratio of the detector.

$$\delta^2 = \left[Q^{-1}(P_{fa}) - Q^{-1}(P_d)\right]^2 \tag{16}$$

Five nonlinear wideband radiometer models are assumed for use in the present LPI analysis. These models were suggested in [6] and shown in Table 1.

From these models it is apparent that the total power radiometer requires the least signal power for a successful intercept. Since the total power radiometer is the optimum receiver for detecting the presence of an unknown signal in a gaussian noise environment, we can use it as the standard against which other radiometers and feature detectors are measured.

Communication Receiver Models

For the communications receiver, the performance criterion is the probability of bit error. The probability of bit error can be related to the received E_b/N_{OC} for any particular type of

waveform modulation and detection process. This relationship is expressed in the parameter $\zeta_C(P_e)$. Several popular modulation techniques are shown in Table 2 with the corresponding $\zeta_C(P_e)$.

More complex forms of modulation cannot easily be expressed in closed form. For example, the expression for the probability of bit error for 32-ary orthogonal noncoherent signaling is not easily expressed in the form specified for $\zeta_C(P_e)$:

$$P_e = \frac{1}{62} \sum_{n}^{32} \binom{32}{n} (-1)^n \exp\left[-5 \frac{E_b}{N_o}\left(\frac{n-1}{n}\right)\right] (\qquad (17)$$

However, the expression can be inverted iteratively, or curves can be used to obtain ζ_C for a specific P_e.

Quality Factor Analysis

An LPI quality factor analysis was performed on a candidate waveform using the five wideband intercept receivers, with the communications receiver using 32-ary orthogonal noncoherent signaling. A comparative summary of the modulation quality factor for all five intercept receivers is shown in Figure 3. All curves in this figure are obtained from the modulation quality factor expression given in (5). Note that the wideband total power radiometer requires the smallest input signal-to-noise ratio for a successful intercept, as expected.

Conclusions

A methodology has been presented which demonstrates the use of LPI quality factors in assessing the detectability of a candidate LPI waveform by a variety of wideband nonlinear intercept receivers. The methodology described here focused on the modulation format of the LPI waveform, and described the quality of its covertness in terms of a modulation quality factor. A technique was described for modeling the signal, the intercept receiver and the communications receiver so that their essential characteristics could be incorporated into a modulation quality factor. This modulation quality factor promises to be a valuable standardized is valuable as a standard

References

[1]. Gutman, L. L., and Prescott, G. E. "System Quality Factors for Low Probability of Intercept Communications," Proceedings of the 1989 IEEE International Conference on Systems Engineering, pgs. 475-478, Dayton, Ohio, 24-26 August 1989.

[2]. Dillard, Robin A., and Dillard, George M., Detectability of Spread Spectrum Signals, Artech House (ISBN 0-89006-299-4), Norwood MA, 1989.

[3] Edell, John D., "Wideband, Noncoherent, Frequency-Hopped Waveforms and Their Hybrids in Low Probability of Intercept Communications," NRL Report 8025, Naval Research Lab, Washington D. C., 8 November 1976.

[4]. Chandler, Edward W., and Cooper, George R., "Development and Evaluation of an LPI Figure of Merit for Direct Sequence and Frequency Hop Systems," Proceedings of the 1985 Military Communications Conference, pp. 33.7.1 - 33.7.6, October 1985.

[5]. Urkowitz, H., "Energy Detection of Unknown Deterministic Signals," Proceedings of the IEEE, vol. 55, pp. 523 - 531, April 1967.

[6]. Nicholson, David L., Spread Spectrum Signal Design: LPE and AJ Systems, Computer Science Press, Rockville, Maryland, 1988.

Table 1

Intercept Receiver	$\zeta_i(P_d, P_{fa}, T, W)$
Total Power Radiometer	$\delta\sqrt{\frac{W_1}{T_1}}$
AC Radiometer	$\sqrt[4]{\frac{2\delta^2 W_1^2}{T_1 T_3}}$
Pulse Rate Detector	$\frac{\delta}{\rho}\sqrt{\frac{2W_1}{T_2}}$
Hop Rate Detector	$\sqrt[4]{\frac{14.36\,\pi\,\delta^2 W_1^2}{T_2 T_3\,\rho}}$
Chip Rate Detector	$\delta\sqrt{\frac{W_1}{T_2}(316.2)}$

Table 2

Modulation Type	$\zeta_c(P_e)$
Noncoherent Binary FSK	$-2Ln(2P_e)$
Differentially Coherent Binary PSK	$-Ln(2P_e)$
Coherent Binary & Quadrature PSK	$\frac{1}{2}[Q^{-1}(P_e)]^2$

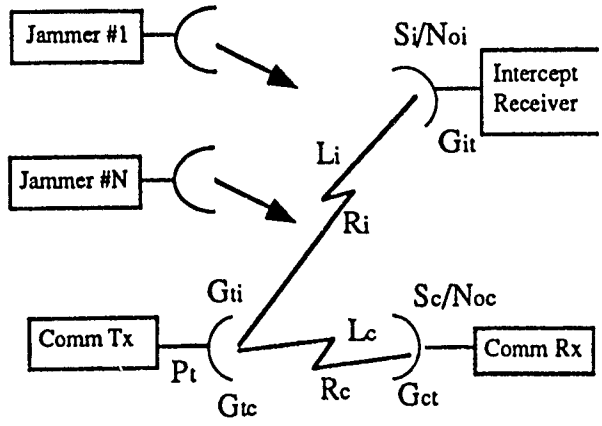


Figure 1 - LPI Scenario

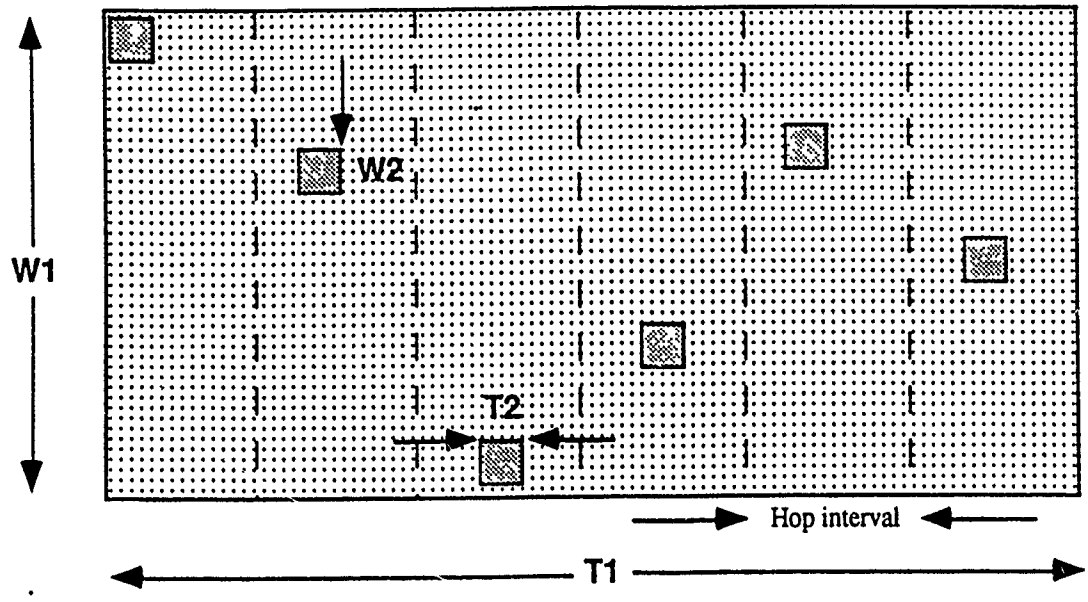


Figure 2 - Detectability Model for the Spread Spectrum Waveform

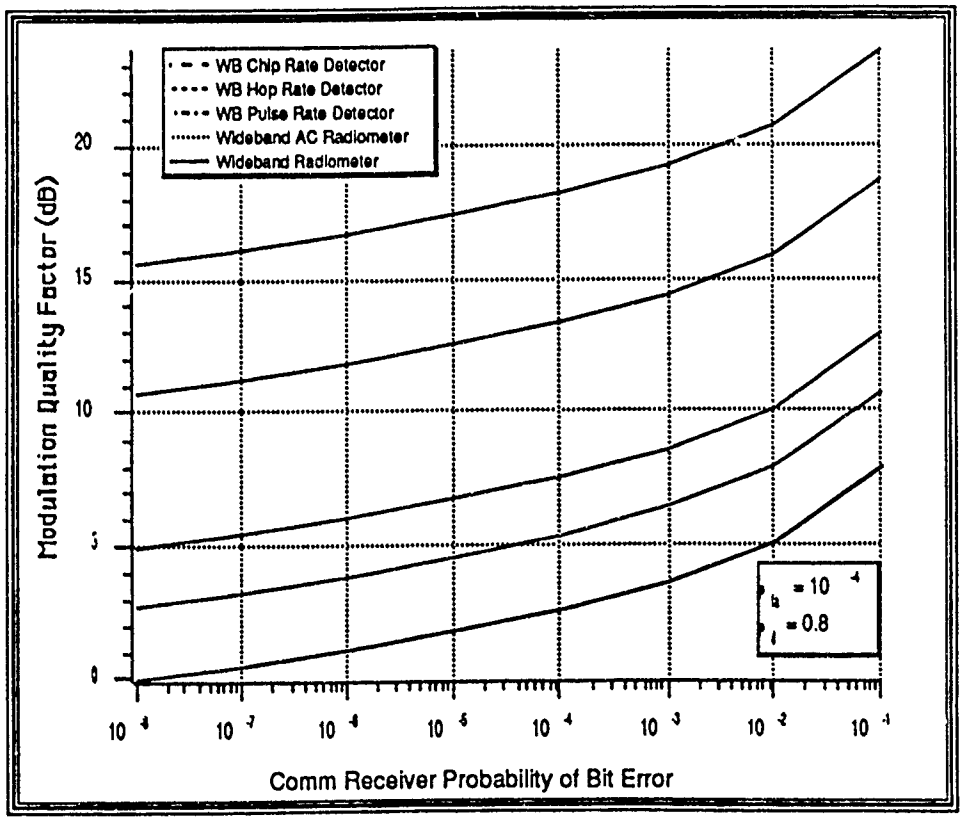


Figure 3 - Modulation Quality Factors for Nonlinear Intercept Receivers

COMPUTER AIDED ANALYSIS OF LPI SIGNAL DETECTABILITY

Scott P. Francis Glenn E. Prescott
University of Kansas Center for Research

Abstract

A low probability of intercept signal detectability analysis program is described which calculates the detectability of certain spread spectrum signals by radiometric detection systems. LPI/SDA is a PC-based computer aided analysis system which expresses the detectability of a signal in terms of the signal carrier power to one-sided noise power spectral density C/N0I required at the input to a radiometer to achieve user-specified detection and false alarm probabilities. LPI/SDA also has the ability to calculate LPI system Quality Factors, which describe the detectability of a signal separately in terms of scenario dependent and scenario independent factors. Sample detectability calculations are illustrated, and performance curves are provided.

Introduction

Military communication system designers have traditionally employed spread spectrum waveforms to achieve a particular level of transmitted signal covertness. These spread spectrum signals, in addition to permitting the use of code division multiple access (CDMA) for efficient bandwidth utilization, also incorporate significant anti-jam (AJ) and low probability of intercept (LPI) characteristics due to their low-level radiated power densities.

The receiver in an LPI communication system possesses knowledge of the code which was used at the transmitter to spread the signal, and thus can despread the received signal by mixing it coherently with the code. An unintended receiver does not typically have knowledge of this spreading code and must make signal present decisions based solely on the received energy in some frequency band over some period of time. Receivers which make binary signal present decisions based on energy detection are called radiometric systems (radiometers), and represent the most common detection threat to LPI signals.

The inherent vulnerability of an LPI (spread spectrum) signal to detection by a particular radiometric system can be quantified in terms of the required carrier signal power to one-sided noise power spectral density ratio C/N_{0I} required at the front end of the radiometer to achieve a specified probability of detection, P_D, and probability of false alarm P_{FA} performance level [1, 5]. The LPI communication system designer uses this detectability information to select the spread spectrum modulation type and parameters to yield a signal which is minimally detectable by the most likely detection threat, in this case a particular type of radiometer system.

Many researchers have developed analytical models which map the radiometer performance probabilities to the required front-end C/N_0, e.g. [1]. LPI/SDA is a PC-based computer aided analysis system which automates the use of these models. The LPI communication system designer interacts with LPI/SDA's user-friendly hierarchical interface to describe a signal and radiometer system, and quickly determines the required C/N_0. Signal and radiometer parameters can be easily modified to evaluate the effects of these changes on the required C/N_0. LPI/SDA also has the ability to calculate five different LPI system Quality Factors. These Quality Factors describe the detectability of a signal separately in terms of scenario dependent and scenario independent factors [3].

LPI Signal Models

LPI/SDA models three standard and four hybrid types of spread spectrum signals. They are:

- Direct Sequence (DS)
- Frequency Hopped (FH)
- Time Hopped (TH)
- FH/DS
- TH/DS
- FH/TH
- FH/TH/DS

From an energy detection standpoint, these spread spectrum signals can be described in terms of relatively few parameters. These parameters are listed below and illustrated in the time-frequency plane shown in Figure 1.

T1	— message duration (sec)
W1	— spread spectrum bandwidth (Hz)
T2	— pulse duration (sec)
b2	— number of pulses in T1
N	— # of frequency hop bands in W1

To describe a DS signal, for instance, one would set N = 1, b2 = 1, and T2 = T1. T1 and W1 would be set to the desired message time and spread bandwidth, respectively.

Radiometer Models

The heart of all radiometric systems which LPI/SDA models is the wideband radiometer, also known as an energy detector or total power radiometer. This system, shown in Figure 2, filters a portion of the RF spectrum, squares this filtered signal to obtain signal power, and

integrates from $t - T$ to t to yield signal energy (typically this integration is implemented as integrate and dump rather than continuous integration). This signal energy is then compared to a threshold and, if the threshold is exceeded, a signal is declared present; otherwise no signal is declared present. Assuming ideal signals and filters, the wideband radiometer can equivalently be described as a system which observes a rectangular time-frequency *cell* with bandwidth equal to the bandpass filter bandwidth and time interval equal to the integration time T. It measures the total signal plus noise energy received in this cell and compares this energy to a threshold. A signal is declared present if the cell energy exceeds the threshold.

For signals which completely occupy a single time-bandwidth cell (e.g., DS signals) and are embedded in additive white Gaussian noise, total power radiometers represent essentially the best performing detection systems which can be constructed [1]. If a signal is pulsed in time and/or frequency, the interceptor may be able to improve his detection performance significantly if he has knowledge of the pulse positions in both time and frequency and exploits this knowledge by using an appropriate radiometer system. These radiometer systems consist of one or more total power radiometers, each with a bandwidth and integration time matched to the time-bandwidth dimensions of a pulse. The binary signal present/not present decisions that each of these radiometers makes are processed in some manner to minimize the false alarm probability while maintaining a high detection probability. These systems are described further in [1].

The LPI/SDA Program

LPI/SDA contains a highly accurate analytical model for calculating the required C/N_{01} at the front end of an intercept receiver given the time-bandwidth product TW of its observation cell and the desired P_D and P_{FA}. For pulse detection radiometer systems, the overall detection and false alarm probabilities can be mapped to single cell detection and false alarm probabilities, Q_D and Q_{FA}, and thus this detectability model is used for all types of radiometer systems.

The radiometer model is based on the following. When only noise is present at the input to a wideband radiometer, the output follows a chi-square probability density function (PDF) with $2TW$ degrees of freedom. With signal present, the output has a noncentral chi-square PDF with 2TW degrees of freedom and a noncentrality parameter of $2E_S/N_{01}$, where E_S is the signal energy received during a time interval of length T [5].

For large TW products, the output statistics approach Gaussian density functions for the noise and signal plus noise cases. Simple detectability models make this approximation and often further assume that the variance of the noise is equal to the variance of the signal plus noise (i.e., the signal is very weak). After signal detectability is calculated, a correction factor is typically

applied to correct for the error introduced by making the Gaussian assumption (see for instance, [4 pg. 298]).

LPI/SDA calculates the required C/N_{01} using chi- square statistics except in the case of large TW products (> 500) where the Gaussian approximation is very good, and makes no assumption whatsoever about the signal and noise powers. A typical detectability calculation in LPI/SDA proceeds as follows. The user describes a signal and radiometer, and specifies a TW product, P_{FA} and P_D. LPI/SDA maps P_{FA} and P_D to Q_{FA} and Q_D if necessary. The radiometer output PDF, given that the input is noise alone, is now fixed, and a detection threshold can be calculated by using a chi-square tail function routine to yield the correct Q_{FA}. With knowledge of this threshold, the noncentrality parameter of the output PDF given signal plus noise at the input can be varied until the correct Q_D is obtained. Q_D is calculated using a recursive generalized Q function algorithm given in [2]. With the radiometer output PDFs determined, the noncentrality parameter of the noncentral chi-square PDF ($2E_S/N_{01}$) is multiplied by $1/2T$ to yield the required C/N_{01}. LPI/SDA modifies this number appropriately if the radiometer dimensions are not perfectly matched to the transmitted signal. This would be the case if, for example, a DS signal is specified along with a wideband radiometer whose bandwidth is something less than the spread bandwidth of the signal. For TW products greater than 500, the calculation proceeds similarly, except that the output PDFs are assumed to be Gaussian with the first and second moments equivalent to the corresponding chi-square moments, and thus Gaussian tail probabilities are used rather than chi-square.

Sample Detectability Calculations

The LPI/SDA user interface contains independent input pages for describing the spread spectrum signal and radiometer. Figure 3 illustrates the structure of the user interface and the paths which one may follow when going from one input page to another. On each of these pages, the user selects a signal or radiometer type via pop-up menus, and specifies the parameters that accompany the selected type. Another page allows the user to specify P_D and P_{FA} and calculate the required C/N_{01}. On yet another page, the user may input communications link information (path losses, antenna gains, receiver noise temperatures, etc.) which are required for Quality Factor calculations. Values for the five Quality Factors can be computed, and one of these Quality Factors, the Modulation Quality Factor, can be plotted against one of six system parameters, including P_D and P_{FA}. The user may move freely from page to page to change the signal type, radiometer type, and associated parameters and find immediately the effect of these changes on the

required C/N_{01}. Compatibility and range checking are performed on all inputs to ensure that they are tenable.

Figure 4 illustrates a plot of the calculated C/N_{01} required as a function of Q_D for several values of Q_{FA} and a TW product of 10,000. These curves correspond within .2 dB to the curves of required E_S/N_{01} presented in [1 pp. 65- 70] (note that $E_S/N_{01} = C/N_{01}$ when $T = 1$).

Woodring and Edell [5] present several examples of calculating the detectability of spread spectrum signals by radiometers. In their first example, they describe a frequency hopped signal with a message time $T1 = 4$ sec, a spread bandwidth $W1 = 2$ GHz, and a hop rate of 2000 hops/sec (which together with T1 corresponds to $b2 = 8000$ and $T2 = 500$ μsec). The desired P_D and P_{FA} are 0.1 and 10^{-6} respectively. They report a required C/N_{01} of 48.9 dB-Hz, and given the same parameters, LPI/SDA returns 48.900 dB-Hz. In their second example, they calculate the detectability of a single hop or pulse. In this case, $T1 = T2 = 500$ μsec is the pulse duration and $W1 = 2000$ Hz is the pulse bandwidth, yielding a TW product of unity. Note that $b2 = 1$. Woodring and Edell give a required C/N_{01} of 41.7 dB-Hz whereas LPI/SDA returns 41.740 dB-Hz.

Conclusions

LPI/SDA is an efficient and highly accurate computer aided analysis tool to assist the LPI communications design engineer in determining the detectability of certain spread spectrum waveforms by radiometric detectors. Using this tool, the designer can construct a signal waveform which is least vulnerable to detection by the most likely detection threat. Further, the designer can analyze the covertness of the LPI signal in terms of scenario dependent and scenario independent factors via calculation of five different Quality Factor. LPI/SDA's results have been shown to be within .2 dB of similar analytical results presented in [1] and [5].

References

[1] Dillard, George M. and Robin A., *Detectability of Spread-Spectrum Signals*, Artech House, Norwood, MA, 1989.

[2] Dillard, George M., "Recursive Computation of the Generalized Q Function," *IEEE Transactions on Aerospace and Electronic Systems*, July 1973, pp. 614-615.

[3] Gutman, Lawrence L. and Prescott, Glenn E., "System Quality Factors for LPI Communications," *Proceedings of the 1989 IEEE International Conference on Systems Engineering*, Dayton, Ohio, 1989.

[4] Simon, M., Omura, J., Scholtz, R., and Levitt, *Spread Spectrum Communications*, Vol. 3, Computer Science Press, Rockville, MA, 1985.

[5] Woodring, D. and Edell, J., "Detectability Calculation Techniques," U.S. Naval Research Laboratory, SCTN 1977-1, Sept. 1, 1977.

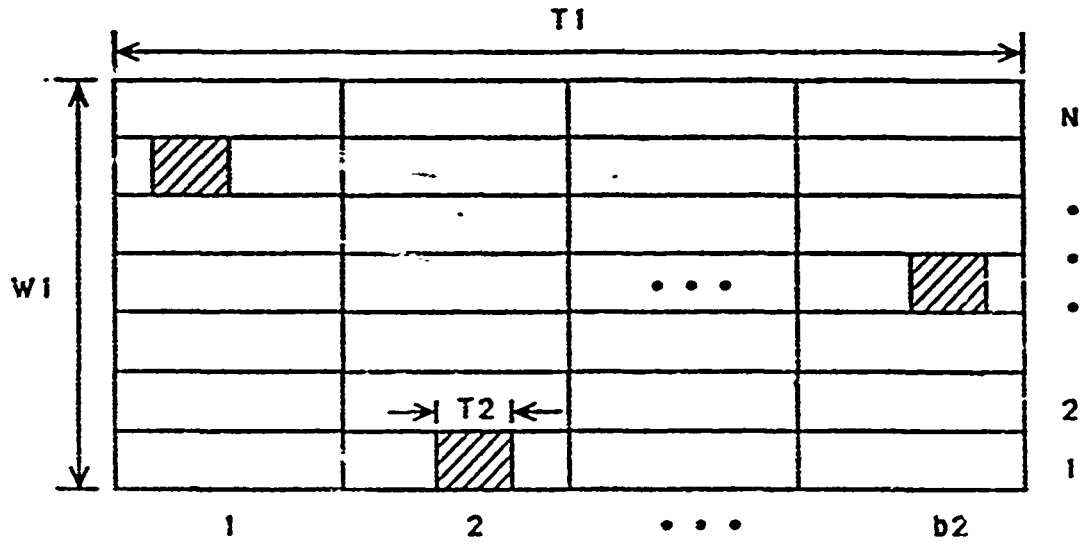


Figure 1: LPI/SDA signal parameter notation. Adapted from [1].

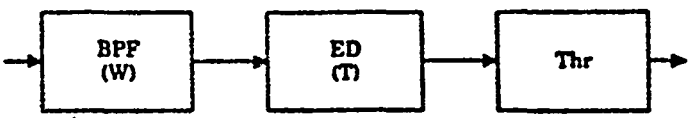


Figure 2: Wideband radiometer system (bandpass filter, energy detector, and thresholder). [1]

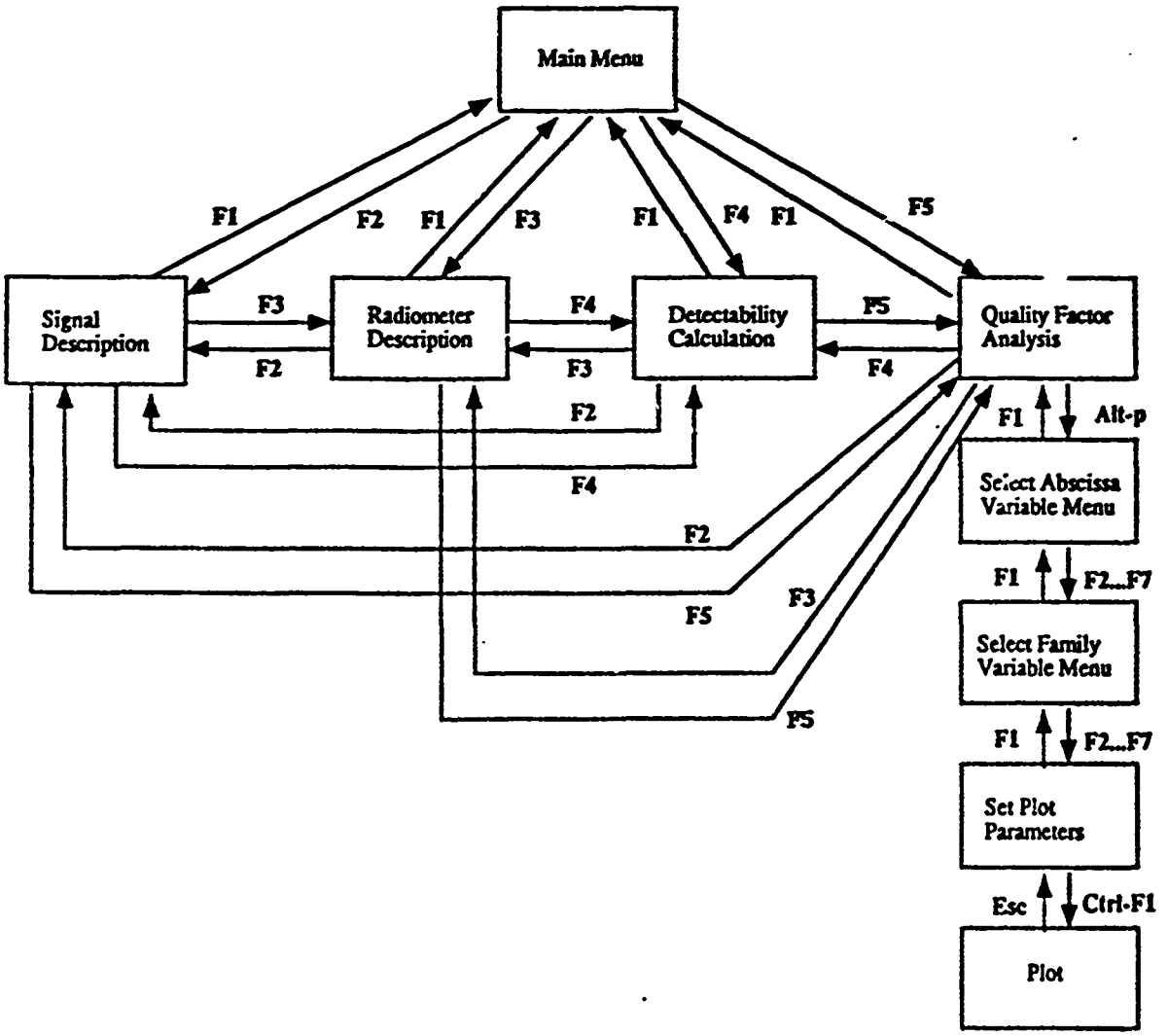


Figure 3: LPI/SDA user interface structure.

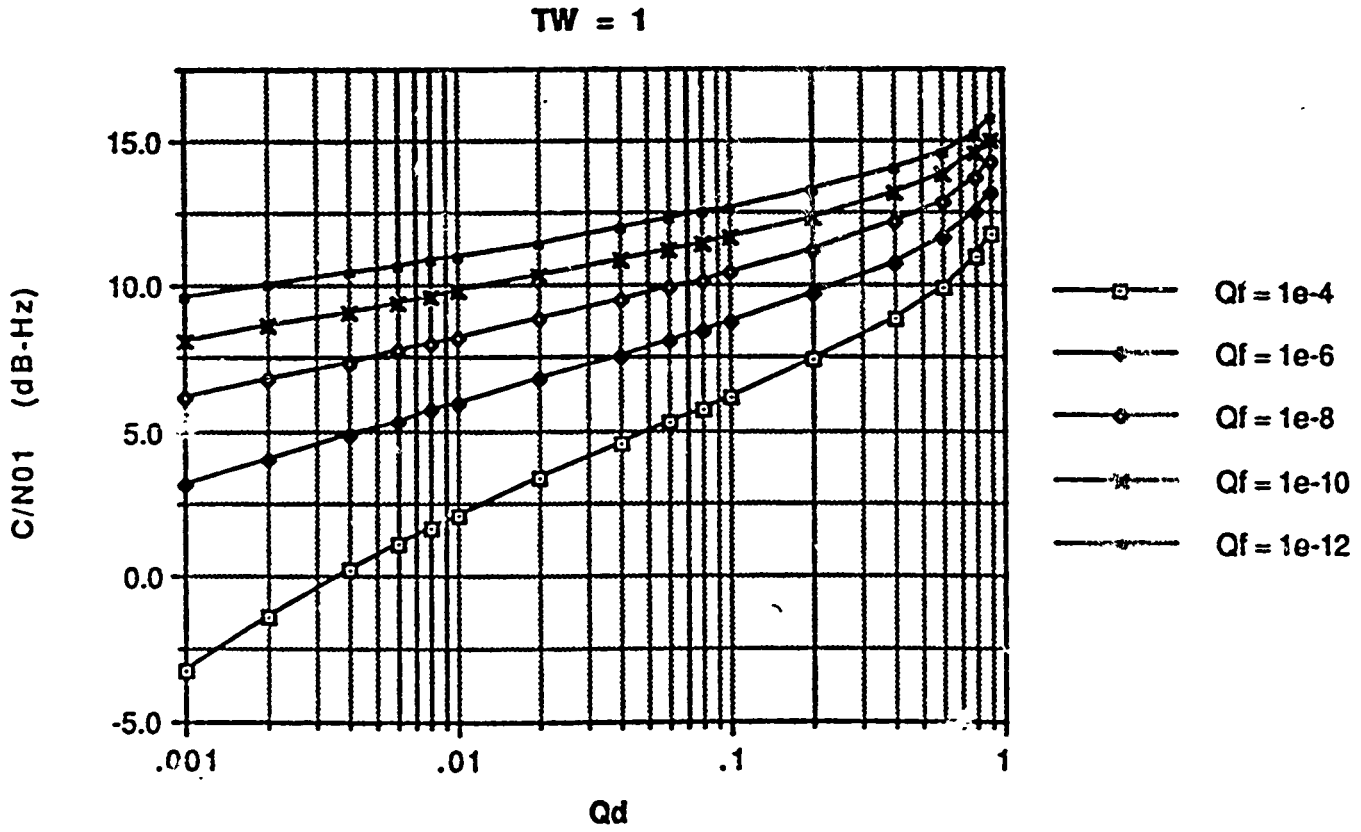


Figure 4: Required C/N_{01} for $TW = 1$.

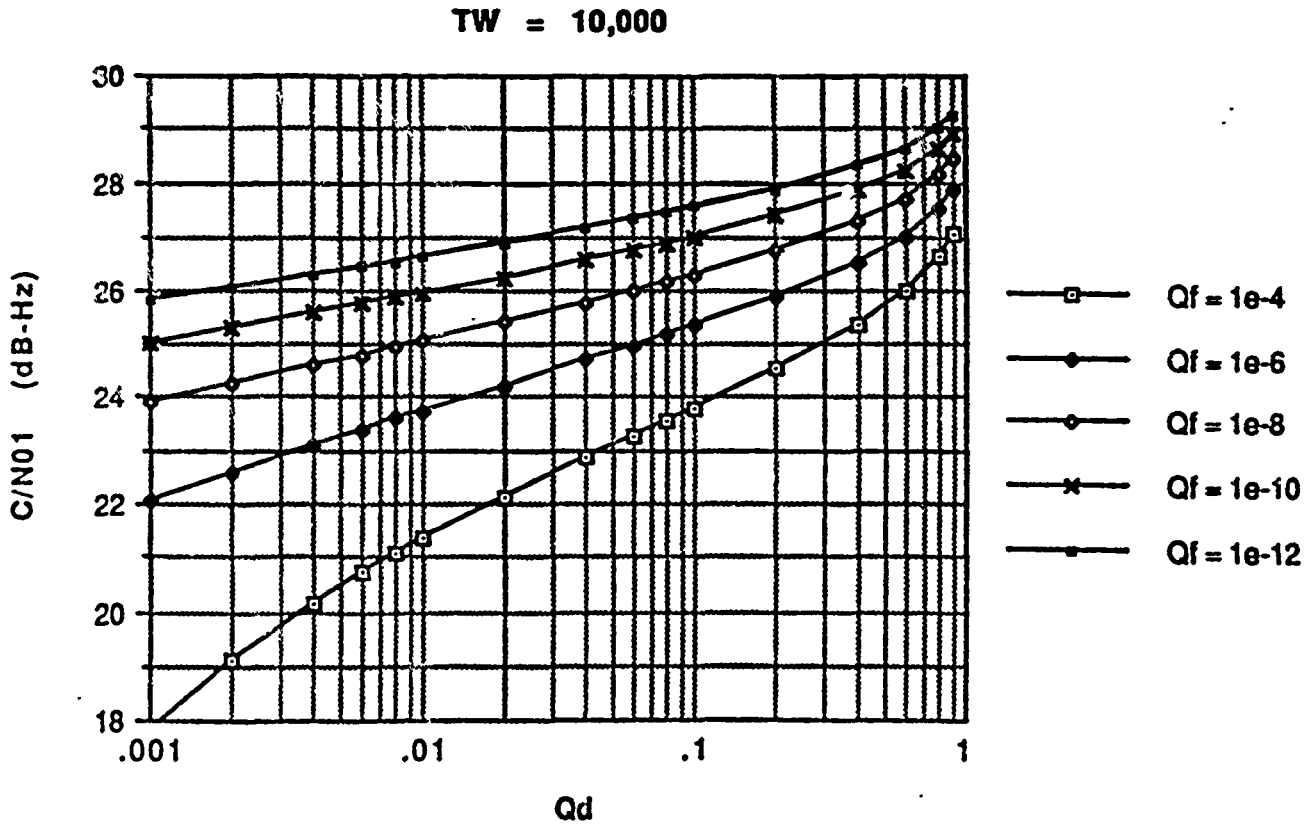


Figure 5: Required C/N_{01} for $TW = 10{,}000$.

1991 USAF-RDL SUMMER FACULTY RESEARCH PROGRAM

Sponsored by the

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
(AFOSR)

conducted by

Research & Development Laboratories

FINAL REPORT

Reusable Ada Software

Evaluating the Reusable Ada Avionics Software Packages (RAASP)

Prepared by:	Brian J. Shelburne Ph.D.
Academic Rank:	Associate Professor of Computer Science
Department and	Mathematics and Computer Science
University	Wittenberg University
Research Location:	WL/AAAF-3 Wright-Patterson AFB Dayton, Ohio 45433
USAF Researcher:	Marc J. Pitarys
Date:	August 23, 1991

REUSABLE ADA SOFTWARE : EVALUATING THE REUSABLE ADA AVIONICS SOFTWARE PACKAGES (RAASP)

Brian J. Shelburne Ph.D.

ABSTRACT

RAASP is a Air Force reusable software effort for avionics currently being performed by Westinghouse Electric Corporation. Under the effort, a library of Reusable Software Objects (RSOs) is maintained using a hypertext based library system which supports the administration, user browsing, and extraction of RSOs. This report evaluates a prototype reuse system from the RAASP contract. Although this reuse system is configured to handle avionics software, the prototype version used for this evaluation only contained RSOs dealing with numeric routines and Booch data structure parts. The system was tested by writing sample applications using the hypertext based library system to select the proper RSO. Preliminary indications are that the RAASP reuse system can be an effective software reuse system.

Acknowledgements

I wish to thank the Air Force Systems Command (Avionics Directorate) and the AFOSR for sponsorship of this research. Research & Development Laboratories (RDL) must be mentioned for their concern and help to me in all administrative and directional aspects of this program.

I wish to especially recognize Marc Pitarys for his guidance and assistance with this project and to thank him for the many productive discussions we had on the problems and pitfalls related to software reusability. I wish also to thank Kenneth Littlejohn for his technical assistance with both the Ada language and the VMS operating system.

I also wish to thank Paul Medlock of Westinghouse Corporation for his help in answering my many questions about the FarVIEW hypertext package.

I. INTRODUCTION:

RAASP, the Reusable Ada Avionics Software Packages, is a Westinghouse Electric Corporation software reuse project sponsored by the Air Force. The system "consists of an Avionics Domain, other demonstration domains and a Library System ... [whose purpose is to] promote cost savings and higher quality software through reuse of well developed software." [1 p. 1].

Since the RAASP contract was awarded last year, only a prototype version was available for evaluation. Even though the prototype version did not contain any avionics domain reusable software objects (RSOs), it did contain RSOs from other domains like numerics and data structures which were used to generate sample applications. (RSOs can be complicated objects which can include source code, documentation, and test code.) The lack of avionics software did not impede the evaluation of the library system which is a critical component of any software reuse system.

The library system used to access the RSO library was written using a hypertext package called FarVIEW. The library system, called ReuSE (Reuse Search Expert), supports library browsing, library and domain administration, and RSO submittals.

II. OBJECTIVES:

The objectives of the research were threefold : learn the RAASP ReuSE system, evaluate the use of the ReuSE system as a means of supporting reusable software, and evaluate the quality of the

RSOs.

No effort was made to evaluate the submittal and/or library/domain administration capabilities of the ReuSE system. Indeed, this feature was not implemented in the prototype system we had.

III. ReuSE - AN OVERVIEW:

FarVIEW is a commercially available hypertext system which was written for a PC and ported to a VAX. A hypertext system "stores" information in discrete chunks called "frames" which are then linked to one another, the links generally following the logical structure of the underlying database of information. Frames can contain text or graphics and their size are not limited to a single screen. Moving between frames is usually accomplished by "clicking" on "buttons" - positioning the screen cursor over a button icon and clicking. The user can also access frames using cross-references (the user "clicks" on the reference), "back-track" through previously traveled links, or jump immediately to any frame given its name. Thus a user can quickly and easily browse through any database.

In addition FarVIEW supports an "in-house" language called FarSlang which allows developers to write code frames that execute FarSlang code. FarVIEW supports a window editor that uses WordStar-like commands. It has the ability to link to ASCII files. The result is a very flexible and powerful hypertext support tool.

FarVIEW works best with a mouse although a user can use the keypad and/or function keys to point and click on buttons. ReuSE (Reuse Search Expert), the RAASP library system used to maintain a library of RSOs, was written in FarVIEW. Four library functions are supported by ReuSE. The first allows a user to browse the library for an RSO with the option of extracting it. A Library Administrator function handles the setting up of a new library. The Domain Administrator function handles the administration of individual domains within the library, and a submittal function permits new RSOs to be submitted to the library.

ReuSE uses a faceted classification scheme to catalog RSOs. In a faceted classification scheme (as opposed to an enumerative scheme like the Library of Congress classification), RSOs are classified using predefined keywords from "facet" lists. For example "language" could be a facet and "Ada", "C", "Fortran", "Pascal", etc. would be the "values" or keywords in the "language" facet list (see [5] p. 89). Initially a user selects a particular domain (facet) to search (e.g. "Data_Structures"). Next, the user selects a value from the primary facet list (primary in that it is most descriptive for the domain). For example, in the "Data_Structures" domain, the primary facet is "Structure" (Stacks, Queues, Sets, etc.). The primary facet always depends on the particular domain. Once a value for a primary facet is chosen, the user is presented with a set of secondary facets appropriate to that domain. For example, "Data_Structures" has a set of ten secondary facets. (Note that the secondary facets

are determined by the domain facet and not the primary facet.) Then the user either selects values or uses the "Don't Care" default value for each secondary facet and initiates a search of the library for RSOs that satisfy the facet values chosen. It is expected that users will choose the "Don't Care" default for many secondary facets.

To reiterate, searching for an RSO requires the selection of a domain, a primary facet, and a set of secondary facets which are specific to the chosen domain.

At this point, if one or more RSOs are found that satisfy the facet values, their frames (called "RSO Root Frames") are presented in a "book form" where the user can "page" back and forth (with the click on a button) examining each frame in turn. Each RSO Root Frame uses a standard format which displays the frame name, the RSO name, the values for the primary and secondary facets, an abstract, and a set of buttons that allows the user to access additional documentation, source code, and test code frames for the RSO. It also has an "Extract" button that allows the user to extract the source code and all files upon which it depends.

If no RSOs are found which match the facet values, a distance measuring function is employed to obtain "close" matches.

Responses to queries generally result in the return of RSO Root Frames but "Description Frames" which contain additional documentation can also be returned.

IV ReuSE - RSO DOCUMENTATION :

A typical RSO Root Frame looks like the following ([1] p. 3-5) :

```
DATASTR.stack_0001                                          2 of 2

RSO name : Stack_Package                                [  More  ]

Structure:              Stack     Abstraction:          Multiple
Element Visit:  No_Visitation     Environment:         Semaphore
Memory:             Run_Time      Exceptions:      No_Propagate
Portability: Machine_Specific     Private Type:         Private
Confidence:       Confidence1     Optimization:            Time
Sub Expansion:     No_Inline

    Basic Stack Operations

[ Extract ] [ Dependencies ] [  Metrics  ] [  Log  ]

[ Submittal ] [  Install  ] [  Package  ]
```

The first line is the frame name (DATASTR.stack_0001). The "2 of 2" indicates this RSO Root Frame is the second of two RSO Root Frames whose facets match the user's requests. The RSO name (Stack_Package) appears on the next line along with the [More] button. (Adopting a convention used by ReuSE, buttons will appear in square brackets.) Clicking on this button will link to the RSO More Frame (discussed below). Next comes the values for all facets (Structure being the primary facet, the other ten being secondary), then the abstract (in this case the one line phrase "Basic Stack Operations"). Finally there are two rows of buttons which link to other frames of this RSO.

Clicking on the [Extract] button will extract the RSO and all RSOs it depends on. Clicking on the [Dependencies] button will link to a frame listing all other RSOs the current RSO depends

on. Clicking on the [Package] button will link to a frame containing the source code for the RSO. (Some RSOs have [Spec] and [Body] buttons instead.) The other buttons link to frames containing information of lesser importance.

The [More] button brings up the RSO More Frame which contains information about the author and/or submitter of the RSO as well as twelves more buttons which link to frames containing additional documentation

V. ReuSE : AN EVALUATION:

The ReuSE system is easy to use even when a user is not familiar with hypertext systems (ReuSE has a built in tutor facility). Each query requires only three levels of selection : the domain, the primary facet, and a set of values for the secondary facets.

Once the selections are made, the library is queried for matches. RSO Root Frames of RSOs which match the selected facet values are presented in sequence which the browser can inspect. The standardized format makes them easy to scan; the user is not overwhelmed with information about an RSO so the he or she can quickly decide whether to keep or discard any selection. Clicking on a button makes it easy to obtain additional information for any RSO.

The [Package] or [Spec]/[Body] buttons are useful for examining source code. FarVIEW uses "slider bars" to manage frames larger than a single screen, so it is easy to scroll through and

examine source code.

The [Dependencies] button is particularly useful since many Ada language parts often "with" other packages. Unfortunately, this particular feature was not completely implemented in the prototype version.

The [Extract] button, used to extracted the RSO and "all needed files of the RSO, as well as all the files associated with any RSOs on which [that] RSO depends, and RSOs on which they depend, etc." [1 p. 3-16], was not completely implement in the prototype version. Often files on which the extracted RSO depended on were missing. In addition, the [Extract] button gave no indication as to what files were extracted. ReuSE documentation suggested extracting files to an empty subdirectory so the user would know which files were extracted.

Most of the other buttons revealed empty frames although in many cases, the attached frames were not germane to the RSO.

In summary the ReuSE library system is easy to use since it allows the user to freely browse through the library of RSOs without being overwhelmed with information or documentation. At the same time, the "point and click" approach of a hypertext based library system allows the user to access the details of or additional information about an RSO. The only drawback was the failure to fully implement some very useful features.

VI. APPLICATIONS

ReuSE listed fourteen different domains although only two, Numerics and Data_Structures, were populated with RSOs; none of the twelve avionics domains were populated with RSOs.

The Numerics domain was divided into seven areas according to the primary facet "function" (Primitive; Elementary; Matrices/Vectors; Linear-Algebra; Eigenvalues,/FFT/etc; Polynomials; Quaternions) many of which were not populated with RSOs. The Data_Structures domain was divided into eleven areas according to the primary facet "structure" (Stacks, Queues, Lists, Strings, Deques, Maps, Sets, Bags, Trees, and Graphs). The RSOs for this domain were Booch parts but only three (Stacks, Sets, and Bags) were populated with more than one RSO.

Despite the scarcity of RSOs, four application programs were written.

The first was a simple program to compute a square root. The RSO containing a square root function was under the facet value "elementary" in the "numerics" domain. The application to call the square root routine was straightforward and there was little problem generating the program except for the fact that the [Extract] feature did NOT extract from the RSO library all the files needed by the square root function. A scan of the specification file revealed that the package containing the square root function "withed" another package that was not extracted. The missing files had to be located and extracted by hand.

The second application, an infix to postfix expression evaluation application, required the use of three data structures - a stack, a queue, and a variable-length string, all from the Data_Structures domain. Some difficulty was encountered in obtaining the RSOs for this application. A wrong value for one particular secondary facet on a number of stack RSOs lead to a problem in locating the correct RSO. (This problem was subsequently communicated to Westinghouse.) Since exactly one queue RSO and one string RSO were cataloged (and neither was quite the one needed) the other needed RSOs had to be obtained independent of the ReuSE system.

The third application solved a system of real valued linear equations. Like the Data_Structures domain, the Numerics domain was underpopulated and the only RSO available used complex values instead of reals. The RSO was made to work without too much difficulty but the [Extract] feature failed to extract all the files the RSO depended on. The required files had to be located and extracted by hand. One additional difficulty with the third application was the failure of the RSO to detect whether a system of linear equations had a unique solution. Documentation for the RSO stated that an exception would be raised if the result could not be computed but testing revealed that singular systems were not detected. The comment was in error or at least misleading. The problem was easily worked around using a determinant procedure from the same RSO package.

The fourth application used a Sets generic package from the Data_Structures domain. Again there were the problems with

incorrect values for secondary facets which complicated the selection of the desired RSO.

VII CONCLUSIONS:

1. The ReuSE library system facilitates accessing RSOs. It is easy to use (just point and click), it does not overwhelm the user with information, yet it has the capability to allow the user to access as much documentation as necessary. The ease of use and flexibility was particularly helpful in locating some Data_Structure RSOs which had been classified incorrectly.

2. ReuSE has the capacity for many desirable features that ought to be fully implemented. For example, the [Dependencies] button should correctly show all other RSOs that an RSO is dependent on. In fact, given the capabilities of the FarVIEW hypertext system used to implement ReuSE, displaying the RSO dependencies graphically would make the dependencies clearer to the user. Currently the [Extract] facility extracts the RSOs to a user's directory but, as mentioned above, there is no way of knowing which RSOs were extracted. Extracting to an empty directory so that the RSOs might be known is awkward; a better solution would be to include as one of the extracted files, a text file which lists all extracted objects plus compilation instructions.

3. Care must be taken in assigning the correct values to secondary facets as some miss-classified RSOs were found. Also,

the significance or meaning of the secondary values should be documented in some way (a help button attached to each secondary facet?).

4. ReuSE implements the Booch parts as data structures but it does not use Booch's classification. Users familiar with Booch parts might find his classification more helpful (see [2]).

5. As mentioned above, a query would sometimes return "Description Frames" in addition to RSO Root Frames. The former are useful particularly in documenting Booch parts and should be included as part of the documentation of an RSO and not as a separate "RSO Root-type Frame".

6. Currently the ReuSE system is restricted to fourteen domains. It is not clear how difficult it would be to extend ReuSE to more domains.

7. Each domain uses for each primary facet the same set of secondary facets. While this simplifies the selection process for an RSO (as well as the submittal process), it's questionable whether a single set of facets can adequately characterize all RSOs in a domain. For example, in the Data_Structures domain, queues have two properties, balking vs non-balking and priority vs non-priority, which no other data structure has (except Deques). (A balking queue permits an item to be removed from any position in the queue [2 p. 145].) Neither of these properties is listed as a secondary facet which means that an RSO with these properties cannot be selected and can only be found by using a sequential search. On the other hand, making these properties

into secondary facets may make no sense when applied to other structures. One could either tailor the secondary facets to the individual structure (i.e. make them dependent on the primary facet and not on the domain choice) or use the same secondary facets for all RSOs in a domain but color code the buttons in such a way as to indicate the meaningful ones.

8. As noted above, if a search for an RSO does not yield any exact matches with the selected facets, a distance measuring function is employed to find a "close" match. The method used by the distance measuring function needs to be more specific. For example, in generating a query for a certain type of queue, no matches where found (not surprising since only one queue RSO was in the library). However, 14 similar matches were found of which 13 were stacks and the other was a list. The one known queue RSO was not returned.

It should be observed that the "distance" between two RSOs is not symmetric. For example, the ReuSE RSO library currently has Stack RSOs with and without "iterators". (An iterator allows the user to process each element of a stack.) All other facets being the same, a stack RSO with an iterator can be used in place of a stack RSO without an iterator. The reverse is not necessarily true. The distance from the second RSO to the first is essentially "zero" while in the reverse direction the distance is "non-zero".

VIII. FINAL OBSERVATIONS

In a previous paper [6] I addressed a number of issues dealing with software reuse. Since the prototype version of RAASP's ReuSE system is not well populated with RSOs, it is only possible to evaluate the RAASP system with respect to the three of these issues: parts engineering systems, documentation, and quality.

VIII.1. Parts Engineering System Issues

The ReuSE hypertext approach is an excellent way to access the RSO library. First, the selection process allows the user to quickly narrow down the search to a handful of parts. A query yielding too many or too few candidates can be changed by using fewer or more "don't care" defaults for secondary facets. Second, the RSO Root Frames present only the salient information about an RSO. The user is not overwhelmed by information yet if more information is needed, clicking the proper button will yield it. The user can quickly decide whether to investigate an RSO further or reject it. Third, the user can easily access and review the source code itself. Last, when properly implemented, it is easy for the user to extract an RSO plus all RSOs it depends on.

VIII.2 Documentation

Much of the documentation for an RSO is contained in the ReuSE "RSO Root frames" where it is readily accessible. It is difficult to judge the overall quality of the internal documentation (comments) in the RSO source code since the Data_Structure RSOs were Booch parts which are well documented in

Booch's text ([2]) while the Numerics RSOs were a mixed lot. However, some sort of standard should be adopted for internal documentation of source code.

One type of documentation that is lacking is the use of pre and post conditions for procedures. This type of documentation should be included (at least in the source code but possibly as a separate frame linked to the RSO Root Frame via a button).

VIII.3. Quality

There are really two issues here : the quality of the RSO parts themselves and steps taken to assure the user of their quality.

The quality of the Data_Structure domain RSOs was good which is what would be expected of the commercially available Booch parts. The quality of the Numerics domain RSOs was less so (see the exception problem with the third application discussed in section VI above) but no major problems were found.

Quality assurance of an RSO depends on the user's understanding of the RSO which in turn depends on good documentation and easy access to well commented source code. The hypertext based ReuSE system allowed easy access to frames containing documentation and/or source code for an RSO. The capability of quick and easy access to documentation and source code cannot be underemphasized in its usefulness in establishing the quality of the RSOs.

One impediment to quality assurance of software is heavy use of dependencies (via withing) between RSOs. Since the ReuSE system

is not heavily populated with RSOs, this did not seem to be a problem but as more RSOs are added, dependencies between RSOs should be minimized, well documented, and the frame accessed by the [Dependencies] button carefully implemented.

Quality can be assured through the development of applications using RSOs while the library is being populated. In this way, as the RSOs are being added to the library, they can be tested.

IX. RECOMMENDATIONS

ReuSE as it now exists is a very powerful hypertext based library system that supports RSOs from only two well understood domains, Numerics and Data_Structures. Even so, several of the features of ReuSE are not fully or properly implemented (eg. the [Extract] feature or the [Dependencies] documentation). ReuSE needs to be populated with avionics RSOs from domains which are not as well understood. Currently only three selections (domain, primary facet, and secondary facets) are needed to generate a query which returns a small selection of candidate RSOs. Can the same be done with less well understood avionics domains? Based on this summer's research, the following three recommendations are presented.

1. Implement and test the faceted classification schemes for the other twelve avionics domains. The schemes used for Data_Structures and Numerics work well because both domains are mature and well understood. However, it is not clear that the classification of avionics software will work as well although

the hypertext based ReuSE library system is very powerful and flexible.

2. Populate the ReuSE library with avionics RSOs.

3. Generate sample avionics applications with ReuSE to demonstrate that ReuSE is an effective avionics reuse system.

REFERENCES

[1] _____, "Preliminary Software User's Manual for the Library System of Reusable Ada Avionics Software Packages (RAASP)", Westinghouse Electric Corporation, Electronic Systems Group, Aerospace Software Engineering, January 30, 1991.

[2] Booch, Grady, Software Components With Ada, The Benjamin/Cummings Publishing Company Inc., Menlo Park, California, 1987.

[3] Conklin, J., "Hypertext : An Introduction and Survey", IEEE Computer Vol. 20 No. 9, Sept 1987, p. 17-41.

[4] Medlock, Paul J., farVIEW Hypertext, The Manual, farSIGHT Technologies Inc., Baltimore MD, 1991.

[5] Prieto-Diaz, Ruben, "Implementing Faceted Classification for Software Reuse", Comm of ACM Vol. 34, No. 5, May 1991, p. 88 - 97

[6] Shelburne, B. , "White Paper : Reusable Ada Software - Evaluating the Common Ada Missile Packages (CAMP-3)"; Technical Report, WL/AAAF-3, Wright-Patterson AFB, Dayton Ohio 45433, August 1990.

[7] Watt, David A., Wichmann, Brian A, and Findlay, William, ADA Language and Methodology, Prentice/Hall International, Englewood Cliffs, New Jersey, 1987.

Filtering by Similarity

Thomas Sudkamp
Department of Computer Science
Wright State University
Dayton, Ohio 45435
email:tsudkamp@cs.wright.edu

Abstract: A fuzzy filter is a mapping that employs similarity information to transform a fuzzy set to enhance the continuity of compatibility (membership) values. The objective of filtering is to discount noise that may be introduced into domain information by the knowledge elicitation process or by the limitations of mechanical sensors. The effect of filtering is determined by a similarity relation and the aggregation methods employed. Four families of fuzzy filters based on t-norms and t-conorms are presented. Iterative applications of the filters extend the influence of elements beyond their immediate similarity neighborhoods.

1. Introduction

The primary motivation for the development of fuzzy set theory [14] was to provide mathematical tools for representing and reasoning with vague and imprecise information. Fuzzy techniques represent a departure from the classical approach to approximate reasoning in that they are founded not on statistical analysis but rather on the more natural notions of similarity and possibility [15]. The similarity of two objects suggests the degree to which properties of one may be inferred from those the other. Thus similarity provides the foundation for both analogical and interpolative reasoning.

It is often the case that the degree to which objects satisfy an imprecisely defined concept is assumed to change continuously with variations in the charateristics of the objects. For example, a small change in height produces only a slight modification in the degree to which a person would be considered tall. In domains in which such continuity conditions are presupposed, the similarity of domain elements provides information that may be used to evaluate and modify compatibility values that have been obtained from unreliable sources or have been corrupted by noise. Four families of fuzzy filters, transformations that enhance the continuity of compatibility

assignments based on the similarity of the elements in the universe, are introduced and their properties examined.

The relationship between similarity relations and distance functions is reviewed in Section 2. The correspondence between these fundamental notions is based on the t-norm transitivity of similarity and satisfaction of 'triangle inequality'-like restrictions by distance functions. This is followed by the introduction of the families of fuzzy filters. Section 4 uses direct similarity assessments to determine the (indirect) influence of one domain element upon another. This notion provides the foundation for the analysis of iterative filtering.

Throughout this paper elements of a universe will be denoted by u, v, or w, with or without subscripts. Lower case letters x, y, and z are variables over the nonnegative real numbers. The membership function of a fuzzy set A over a universe $U = \{u_1, \ldots, u_n\}$ is given by $\mu_A(u_i) = x_i$ where x_i is the degree of membership of u_i in A. The infix operators $\vee$ and $\wedge$ represent the maximum and minimum functions, respectively.

2. Similarity Relations and Distance Functions

The concept of fuzzy relation [14] generalizes the classical notion of relation to permit the specification of partial relationships between elements. For example, the relation defining 'is a child of' is precise in that either u is a child of v or u is not, there is no matter of degree in this relationship. However, a description of 'is a friend of' or 'is the same size as' or 'is similar to' does not present such unambiguous alternatives. A fuzzy binary relation on a set U is a function $R : U \times U \rightarrow [0, 1]$ or, equivalently, a fuzzy subset of $U \times U$. $R(u, v) = x$ indicates that the degree to which u has relationship R to v is x. Being an extension of standard relations, $R(u, v) = 1$ specifies that u is completely in relation R to v and $R(u, v) = 0$ indicates that u has no R relationship to v.

Zadeh [16] introduced similarity relations to specify the degree of similarity between elements of a universe U. The intuitive nature of similarity requires that certain fundamental properties be satisfied by any similarity relation S. An element must be considered completely similar to itself. Formally, a relation that satisfies $S(u, u) = 1$ for every domain element $u \in U$ is said to be *reflexive*. Similarity is also *symmetric*, that is, the degree of similarity of u to v must be the same as that of v to u. These properties are formalized by conditions i) and ii) of the definition of similarity relation.

Definition 1. A *t-similarity relation* on a domain U is a function $S : U \times U \rightarrow [0,1]$ that satisfies

i) $S(u,u) = 1$

ii) $S(u,v) = S(v,u)$

iii) $S(u,v) \geq S(u,w) \otimes S(w,v), \forall w \in U$

where $\otimes$ is the t-norm t.

A t-norm is a binary function that extends the domain of logical conjunction from the set $\{0,1\}$ to the interval $[0,1]$. Formally, a t-norm is a nondecreasing, commutative, and associative function $\otimes : [0,1] \times [0,1] \rightarrow [0,1]$ that satisfies $0 \otimes x = 0$ and $1 \otimes x = x$. Several common t-norms that will be used in the sequel are given in Tabie 1. Every t-norm t assumes values bounded by t_0 and t_3. That is,

$$t_0(x,y) \leq t(x,y) \leq t_3(x,y)$$

for all $x, y \in [0,1]$. The Lukasïewicz t-norm t_1 provides the link between similarity relations and metric spaces. The relationship between these two concepts will be reviewed below. While many of the results in this paper hold for arbitrary t-norms, we shall primarily be concerned with the Lukasïewicz, product, and minimum t-norms. An examination of the general properties of t-norms can be found in [12,4,6].

Condition iii) in the definition of similarity relation is referred to as $\otimes$-*transitivity*. Transitivity places a lower bound on the inheritability of similarity. If u is extremely similar to w and w is extremely similar to v, then one would expect a nontrivial degree of similarity between u and v. Moreover, an increase in the similarity of the arguments on the right-hand side of the transitivity condition should be accompanied by an increase in $S(u,v)$. Algebraic properties of t-norms coincide with intuitive limitations on similarity transitivity—the complete dissimilarity of u and w should impose no restrictions on the similarity of u and v. Moreover, if u and w are considered completely similar $(S(u,w) = 1)$ then every element has the same degree of similarity to u as it has to w.

The original definition of similarity relation specified the minimum as the combining function in the transitivity condition. Zadeh noted, however, that a similarity relation could be defined using $*$-transitivity where $*$ is any function that is associative and nondecreasing in both of its arguments. Defining similarity with

t-norm	t-conorm
$t_0(x,y) = \begin{cases} min(x,y) & \text{if } max(x,y) = 1 \\ 0 & \text{otherwise} \end{cases}$	$s_0(x,y) = \begin{cases} max(x,y) & \text{if } min(x,y) = 0 \\ 1 & \text{otherwise} \end{cases}$
$t_1(x,y) = max(0, x+y-1)$	$s_1(x,y) = min(1, x+y)$
$t_2(x,y) = xy$	$s_2(x,y) = x+y-xy$
$t_3(x,y) = min(x,y)$	$s_3(x,y) = max(x,y)$

Table 1: Common t-norms and t-conorms

min-transitivity (t_3) provides a natural extension of the crisp set-theoretic notion of equivalence relation (see [16]). The α-level sets of a *min*-transitive relation form a family of equivalence relations. The equivalence classes partition the universe into sets containing elements that are all similar to each other to degree at least α. Rosenfeld [8] and Potoczny [7] have shown that, in the presence of symmetry, *min*-transitivity is equivalent to requiring that, for every three elements in the universe, the degree of similarity of the two less related be equal. This unintuitive constraint provided the motivation for expanding the type of transitivity operators and permitting more general families of similarity assessments [9,10,18].

Example 1: Three similarity relations are defined over the universe $U = \{u_1, \ldots, u_n\}$. The relations S_1 and S_2 are t_1-similarity relations.

$$S_1(u_i, u_j) = 1 - |i-j|/n$$

$$S_2(u_i, u_j) = \begin{cases} 1 & \text{if } i = j \\ .5 & \text{if } |i-j| = 1 \\ 0 & \text{otherwise.} \end{cases}$$

The relation S_3 over $\{u_1, \ldots, u_6\}$, taken from [16], is a t_3-similarity relation.

S_3	u_1	u_2	u_3	u_4	u_5	u_6
u_1	1	.2	1	.6	.2	.6
u_2	.2	1	.2	.2	.8	.2
u_3	1	.2	1	.6	.2	.6
u_4	.6	.2	.6	1	.2	.8
u_5	.2	.8	.2	.2	1	.2
u_6	.6	.2	.6	.8	.2	1

Note that, for every three elements u_i, u_j, and u_k, at least two of the three pairwise similarities are equal.

The concept of similarity is inextricably related to that of nearness or distance. Corollary 1 expresses the well known relationship between similarity and distance.

Definition 2. A *t_i-distance function* on a domain U is a function $d : U \times U \rightarrow [0, \infty)$ that satisfies $d(u,u) = 0$, $d(u,v) = d(v,u)$ and the inequality ti

t1: $d(u,v) \leq d(u,w) + d(w,v)$,

t2: $d(u,v) \leq d(u,w) + d(w,v) - d(u,w)d(w,v)$

t3: $d(u,v) \leq max\{d(u,w), d(w,v)\}$

for all $u, v, w \in U$.

Inequality **t1** is commonly referred to as the triangle inequality. The pair (U,d), where U is a set and d a distance function on U that satisfies the triangle inequality, is a pseudo-metric space. Note that a t_i-distance function is also a t_j-distance function for $j < i$. Thus any t_i-distance function defines a standard metric over the domain U. Inequality **t3** is known as the *ultrametric* inequality.

Definition 3. A t_i-distance function on a set U is said to be *bounded* if there is a $z \in \mathbf{R}^+$ such that $d(u,v) \leq z$, $\forall u, v \in U$.

Corollary 1. For any domain U, d is a t_i-distance function on U with bound z if, and only if,

$$S(u,v) = 1 - d(u,v)/z$$

is a t_i-similarity relation over U.

The existence of an upper bound permits a scaling so that the distance between any two elements of U lies within the interval $[0,1]$. The satisfaction of the ti inequality by a distance function is equivalent to t_i-transitivity in the corresponding similarity relation. Thus the standard distance functions (pseudo-metrics) produce t_1-similarity relations. The more restrictive similarity transitivity conditions impose additional constraints on the corresponding distance functions.

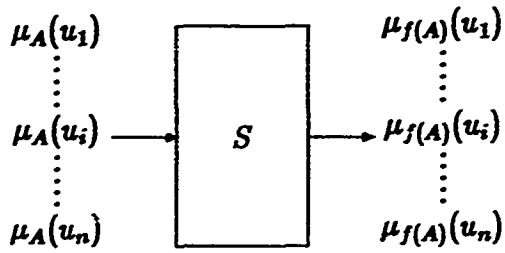


Figure 1: Similarity filter

3. Fuzzy filters

A similarity relation is said to *filter* a fuzzy set A if it is used to transform the degree of compatibility of an element based on those of similar elements. The input to a filter f is a fuzzy set A and the result a modified fuzzy set $f(A)$ over the same universe. Pictorially this is illustrated in Figure 1. Intuitively, filtering is used to 'smooth' the compability assignments. When the compatibility values of two similar elements differ significantly, filtering may be used to reduce the difference. The degree of the modification depends upon the type and sensitivity of the filter. The propriety of applying a filter depends upon the assumption that the compatibility values of the underlying set are typically continuous with respect to the similarity evaluation and the deviations from this are due to imprecision (noise) in the elicitation of the membership function.

In this section we define four classes of fuzzy filters. The designation of the classes indicates the type of smoothing to be performed by the filter. Throughout the development of fuzzy filters, A will denote a fuzzy set over $U = \{u_1, \ldots, u_n\}$ and S a similarity relation on U. Unless specifically indicated, S may be a t_i-similarity relation for any i. The operator $\otimes$ represents a t-norm and $\oplus$ a t-conorm. A t-conorm is an extension of logical disjunction from $\{0,1\}$ to $[0,1]$. t-conorms are commutative, associative, increasing in both arguments and satisfy $0 \oplus x = x$ and $1 \oplus x = 1$ (see [12,6]). The t-conorms that will be utilized in this section are given in Table 1. The t-conorms s_0 and s_3 provide upper and lower bounds for the values of an arbitrary t-conorm. Since t-norms and t-conorms are associative, they may be recursively extended to any number of arguments. $\otimes$ and $\oplus$ will be used to represent these extended operations. Following standard conventions, the multiplicative operator $\otimes$ will take precedence over $\oplus$ in the evaluation of formulae.

Definition 4. A fuzzy set transformation defined by

$$\mu_{f(A)}(u_i) = \bigoplus_{j=1}^{n} S(u_i, u_j) \otimes \mu_A(u_j)$$

is an *increasing filter* on U.

The role of a t-conorm in an increasing filter is that of an aggregator, combining the information obtained by considering the relationship of u_i to all the elements of U. Aggregating with *max* (s_3) produces the standard *max*-$\otimes$ composition [16]. Corollary 2 shows that the transformation defined by an increasing filter produces a nondecreasing updating of compatibility values, generating a fuzzy superset of the original fuzzy set.

Corollary 2. Let f be an increasing filter on U defined by similarity relation S. Then $\mu_{f(A)}(u_i) \geq \mu_A(u_i)$.

Proof:

$$\begin{aligned}
\mu_{f(A)}(u_i) &= \bigoplus_{j=1}^{n} S(u_i, u_j) \otimes \mu_A(u_j) \\
&\geq \bigvee_{j=1}^{n} S(u_i, u_j) \otimes \mu_A(u_j) \\
&\geq S(u_i, u_i) \otimes \mu_A(u_i) \\
&= \mu_A(u_i)
\end{aligned}$$

The initial inequality is a consequence of *max* being a lower bound on the t-conorms. The final step follows from the reflexivity of similarity. □

Increasing filters were employed by Zemankova and Kandel [17] to expand the focal set of a query in a fuzzy relational database system and by Zadeh [13] to transform fuzzy restrictions on the natural numbers. The product was used as the t-norm in the database system while Zadeh used the minimum to incorporate information provided by similarity assessments into the membership values. Both of these applications employed *max* as the aggregation operation.

For each element u_i, a similiarity relation S induces a fuzzy set $N(u_i, S)$ that defines the neighborhood of u_i. The degree of membership of an element u_j in the neighborhood of u_i is

$$\mu_{N(u_i,S)}(u_j) = S(u_i, u_j).$$

The symmetry of S ensures that $\mu_{N(u_i,S)}(u_j) = \mu_{N(u_j,S)}(u_i)$. Elements u_i and u_j are said to be α-neighbors whenever $S(u_i, u_j) = \alpha$. An α-neighborhood of a point u_i is the set of elements that are neighbors of degree α or more. That is, α-neighborhood $N_\alpha(u_i, S)$ is the α-level set of the fuzzy set $N(u_i, S)$.

Applying an increasing filter to a fuzzy set A places a lower bound on the membership values of all elements in the α-neighborhood of an element.

Corollary 3. Let f be an increasing filter defined by similarity relation S and let $u_i \in U$. For any element $u_j \in N_\alpha(u_i, S)$, $\mu_{f(A)}(u_j) \geq \alpha \otimes \mu_A(u_i)$.

The degree to which an element can affect an α-neighbor is fixed by α and the t-norm. For an element u_i, the membership value of every α-neighbor u_j less than $\alpha \otimes \mu_A(u_i)$ will be increased to that value regardless of its current value. Thus, the similarity relation defines the sensitivity of a filter. The coarser the distinctions in similarity, the greater the potential effect of the filter.

Interchanging the role of the t-norm and the t-conorm produces the family of decreasing filters. An argument completely analogous to that given in Corollary 2 shows that transforming a fuzzy set A with a decreasing filter produces a fuzzy subset of A.

Definition 5. A fuzzy set transformation defined by

$$\mu_{f(A)}(u_i) = \bigotimes_{j=1}^{n}((1 - S(u_i, u_j)) \oplus \mu_A(u_j))$$

is a *decreasing filter* on U.

Corollary 4. Let f be a decreasing filter on U defined by similarity relation S. Then $\mu_{f(A)}(u_i) \leq \mu_A(u_i)$.

Proof: The result follows since $\wedge$ is the maximal t-norm and 0 an identity element of the t-conorms.

$$\begin{aligned}
\mu_{f(A)}(u_i) &= \bigotimes_{j=1}^{n}((1 - S(u_i, u_j)) \oplus \mu_A(u_j)) \\
&\leq \bigwedge_{j=1}^{n}((1 - S(u_i, u_j)) \oplus \mu_A(u_j)) \\
&\leq (1 - S(u_i, u_i)) \oplus \mu_A(u_i) \\
&= \mu_A(u_i)
\end{aligned}$$

□

The "direction" of the smoothing accomplished by the preceding filters is uniform and determined by the order of the application of the t-norm and t-conorm operations. The application of a directional filter indicates a belief that the cause of the imprecision in the information has a tendency to consistently increase or decrease compatibility values. When the source of the imprecision does not satisfy these conditions, a filter that averages compatibility values may be appropriate. Two families of filters are defined to transform a fuzzy set based on an average of the elements in the neighborhood.

One approach to averaging is to combine the information provided by an increasing and a decreasing filter. The combination may be obtained by the use of a generalized mean [5,6]. A generalized mean is a function of the form

$$g_r(x_1, \ldots, x_n) = \left(\frac{x_1^r + \ldots + x_n^r}{n}\right)^{1/r}$$

where $-\infty \leq r \leq \infty$. The standard arithmetic mean is obtained by setting r to 1. The limiting values are max and min as r approaches ∞ and $-\infty$ respectively. Thus the generalized means provide combining operators that bridge the gap between the t-norms (bounded above by min) and the t-conorms (bounded below by max).

Definition 6. Let f^+ be an increasing filter, f^- a decreasing filter, and g_r a generalized mean. A fuzzy set transformation defined by

$$\mu_{f(A)}(u_i) = g_r(f^+(u_i), f^-(u_i))$$

is an *averaging filter* on U.

A second approach is to use similarity directly to reevaluate the membership value for a point based on the values assigned to its neighbors (including itself). The degree of influence of a neighbor v on the updated membership value of u is proportional to the contribution of v to the total similarity of the neighbors to u.

Definition 78. A fuzzy set transformation defined by

$$\mu_{f(A)}(u_i) = \sum_{u_j \in U} \frac{\mu_A(u_j) S(u_i, u_j)}{\sum_{u_k} S(u_i, u_k)}$$

is a *similarity weighting filter* on U.

The denominator in Definition 8 is the relative cardinality of the fuzzy set $N(u_i, S)$, that is, the total of the similarity of all the elements in the universe to u_i.

Since filters have been introduced to smooth fuzzy sets, it follows that filtering a set without variance should leave that set unchanged. Corollary 5, which follows immediately from the definitions of the four families of filters, indicates that a smooth set is a fixed point of every filter.

Corollary 5. Let f be a filter and A a fuzzy set such that $\mu_A(u_i) = c$ for all $u_i \in U$. Then $f(A) = A$.

Example 2: Five fuzzy filters are constructed to illustrate the effect of a filtering transformation. The universe is the set $U = \{u_1, \ldots, u_{12}\}$ and the t_1-similarity relation S_2 on U from Example 1. The averaging filter is constructed from the increasing and decreasing filters f_1 and f_3 using the arithmetic mean.

	type	t-norm	t-conorm	g
f_1	increasing	$min(x,y)$	$max(x,y)$	
f_2	increasing	xy	$max(x,y)$	
f_3	decreasing	$min(x,y)$	$max(x,y)$	
f_4	averaging			$(f_1 + f_3)/2$

The filter f_5 is the similarity weighting filter associated with the similarity relation S_2.

U	u_1	u_2	u_3	u_4	u_5	u_6	u_7	u_8	u_9	u_{10}	u_{11}	u_{12}
A	0	1	0	0	0	.5	.75	0	0	0	.5	0
$f_1(A)$	.5	1	.5	0	.5	.5	.75	.5	0	.5	.5	.5
$f_2(A)$	.5	1	.5	0	.25	.5	.75	.375	0	.25	.5	.25
$f_3(A)$	0	.5	0	0	0	.5	.5	0	0	0	.5	0
$f_4(A)$	.25	.75	.25	0	.25	.5	.6755	.25	0	.25	.5	.25
$f_5(A)$	.33	.5	.125	0	.125	.4375	.5	.1875	0	.1725	.25	.125

Note that f_1 increases the compatibility values of the neighbors of any point u with $\mu(u) \geq .5$ to .5. The application of the multiplicative t-norm induces a scaling of the influence based on the magnitude of the compatibility value of an α-neighbor and α. The asymmetry of the similarity weighting about points u_2 and u_{11} is the result of u_1 being the 'left endpoint' and u_{12} being the 'right endpoint' of the set U.

---

4. Similarity and Influence

A similarity relation provides a direct assessment of the resemblance of elements in the universe. The influence of an element, however, may extend (indirectly) beyond the confines of its neighborhood. To examine this we formally introduce the notion of influence. The definitions and results will be given for increasing filters. Similar results can be obtained for decreasing filters. Throughout this section $\otimes$ represents a fixed, but arbitrary, t-norm.

Definition 8. The influence of a path $v_0, v_1, \ldots, v_k$ of length k from v_0 to v_k is $S(v_0, v_1) \otimes S(v_1, v_2) \otimes \ldots \otimes S(v_{k-1}, v_k)$.

From the boundary conditions of t-norms, a path has influence 1 if, and only if, every adjacent pair of elements in the path are completely similar. A path which contains adjacent elements v_i, v_{i+1} with $S(v_i, v_{i+1}) = 0$ has 0 influence. Note, however, that the influence of a path in which each adjacent pair of elements has nonzero similarity may be zero. In [16] Zadeh referred to the influence of a path as its strength. The former term is used here because of the role to be played in filtering.

Definition 9. The k-influence of u_i and u_j is

$$max\{S(v_0, v_1) \otimes S(v_1, v_2) \otimes \ldots \otimes S(v_{k-1}, v_k) \mid v_1, \ldots, v_{k-1} \in U\}$$

where $v_0 = u_i$ and $v_k = u_j$.

The k-influence u_i asserts on u_j is the maximum influence of paths of length k from u_i to u_j. Since t-norms are commutative, the k-influence of u_i on u_j and that of u_j on u_i are identical. 1-influence is just the similarity between elements. Thus, the 1-influence of u_i on the elements of U is given by the fuzzy $N(u_i, S)$.

Example 3 illustrates how the influence of an element may be propagated throughout the entire universe.

Example 3: Let $U = \{u_1, u_2, u_3, u_4, u_5\}$ and let S_2 by the t_1-similarity relation from Example 1. The k-influence of u_2 on the elements of the universe is given for influence paths defined by the minimum, product and Lukasiewicz t-norms.

min

k	u_1	u_2	u_3	u_4	u_5
1	.5	1	.5	0	0
2	.5	1	.5	.5	0
3	.5	1	.5	.5	.5

t_2

k	u_1	u_2	u_3	u_4	u_5
1	.5	1	.5	0	0
2	.5	1	.5	.25	0
3	.5	1	.5	.25	.125

	k	u_1	u_2	u_3	u_4	u_5
	1	.5	1	.5	0	0
t_3	2	.5	1	.5	0	0
	3	.5	1	.5	0	0

Note that the uniform similarity between adjacent points and the *min* operator combine to produce an influence of u_2 that is uniform on all the other elements of the universe. Employing the product scales the influence based on the number of the arcs in the path with nonzero similarity. The influence of the Lukasïewicz t-norm does not extend beyond its similarity neighborhood. This is a consequence of a general property of *max*-t_i composition of t_i-similarity relations (Corollary 7).

The relationships specified by a similarity relation can be represented as a matrix. This is attained by designating an ordering of the elements of the universe. We will assume that the universe U is comprised of elements $\{u_1, \ldots, u_n\}$ with the ordering determined by the subscripts. With this supposition, a similarity relation S over U may be represented by a matrix $\mathbf{S}$ with $\mathbf{S}(i,j) = S(u_i, u_j)$. Throughout the remainder of this section $\circ$ will be used to denote *max*-$\otimes$ matrix multiplication. That is, the i,j'th entry of $\mathbf{R} \circ \mathbf{S}$, where $\mathbf{R}$ is an n by k matrix and $\mathbf{S}$ is a k by m matrices, is

$$\bigvee_{k=1}^{n} \mathbf{R}(i,k) \otimes \mathbf{S}(k,j).$$

Exponents are used to indicate repeated multiplications, i.e., $\mathbf{S}^2 = \mathbf{S} \circ \mathbf{S}$, $\mathbf{S}^3 = \mathbf{S}^2 \circ \mathbf{S}$, etc. The index of element v in the ordering of U is denoted $ind(v)$. Thus if $v = u_j$ then $ind(v) = j$.

Lemma 1. Let $\mathbf{S}$ be a similarity matrix over U. Then $\mathbf{S}^k(i,j)$ is the k-influence of u_i on u_j for $k > 0$.

Proof: For $k = 1$, this follows directly from the definition of influence. Now assume $\mathbf{S}^k$ provides influence information for paths of length k. By definition,

$$\mathbf{S}^{k+1}(i,j) = \bigvee_{t=1}^{n} \mathbf{S}^k(i,t) \otimes \mathbf{S}(t,j).$$

Assume that there is a path $u_i, v_1, \ldots, v_k, u_j$ from u_i to u_j with influence greater than $\mathbf{S}^{k+1}(i,j)$. The influence generated by this path is $S(u_i, v_1) \otimes S(v_1, v_2) \otimes \cdots \otimes$

$S(v_{k-1}, v_k) \otimes S(v_k, u_j)$ for some sequence $v_1, \ldots, v_k$. Let m be the index of v_k in the ordering of U, that is, $v_k = u_m$. By the inductive hypothesis,

$$\mathbf{S}^k(i,m) \geq \mathbf{S}(i, ind(v_1)) \otimes \cdots \otimes \mathbf{S}(ind(v_{k-1}), m).$$

However,

$$\mathbf{S}^{k+1}(i,j) \geq \mathbf{S}^k(i,m) \otimes \mathbf{S}(m,j) \geq \mathbf{S}(i, ind(v_1)) \otimes \cdots \otimes \mathbf{S}(ind(v_{k-1}), m) \otimes \mathbf{S}(m,j).$$

But this contradicts the assumption that $u_i, v_1, \ldots, v_k, u_j$ is a path of greater influence than $\mathbf{S}^{k+1}(i,j)$ and the proof is complete. □

Matrix $\mathbf{R}$ is said to be greater than or equal to matrix $\mathbf{S}$, written $\mathbf{R} \geq \mathbf{S}$, if $\mathbf{R}(i,j) \geq \mathbf{S}(i,j)$ for every coordinate pair (i,j). For similarity matrices, this is equivalent to $\mathbf{R}$ being a fuzzy superset of $\mathbf{S}$. Lemma 2 shows that the powers of a t_i-similarity matrix form a nested sequence of fuzzy supersets of the original similarity relation S.

Lemma 2. Let $\mathbf{S}$ be a similarity matrix over U. Then $\mathbf{S}^{k+1} \geq \mathbf{S}^k$ for all $k > 0$.

Proof:

$$\begin{aligned}
\mathbf{S}^{k+1}(i,j) &= \mathbf{S}^k \circ \mathbf{S}(i,j) \\
&= \bigvee_{t=1}^{n} \mathbf{S}^k(i,t) \otimes \mathbf{S}(t,j) \\
&\geq \mathbf{S}^k(i,j) \otimes \mathbf{S}(j,j) \\
&= \mathbf{S}^k(i,j) \otimes 1 \\
&= \mathbf{S}^k(i,j)
\end{aligned}$$

□

The intuitive justification for Lemma 2 follows from the observation that a path of length k from u_i to u_j defines a path of length $k+1$ that has the same influence. The extended path is constructed by adding an arc from u_j to itself to the end of the existing path. Since the similarity relation is reflexive and 1 is an identity for t-norms, the influence obtained by the addition of the arc is identical to that of the original path. Consequently the influence associated with paths of length k from u_i to u_j provides a lower bound for the influence of paths of length $k+1$.

Corollary 6. Let $\mathbf{S}$ be a similarity matrix over U. Then $\mathbf{S}^j \geq \mathbf{S}^k$ for all $j \geq k > 0$.

Corollary 7. Let S be a t_i-similarity relation and let $\circ$ be max-t_i composition. Then $\mathbf{S}^2 = \mathbf{S}$.

Proof: By Lemma 2,

$$\mathbf{S}^2(i,j) \geq \mathbf{S}(i,j).$$

The opposite inequality

$$\mathbf{S}^2(i,j) = \bigvee_{t=1}^{n} \mathbf{S}(i,t) \otimes \mathbf{S}(t,j) \leq \mathbf{S}(i,j)$$

follows from the t_i-transitivity of S. □

Lemma 2 establishes the general fuzzy inclusion relationship satisfied by powers of arbitrary similarity matrixes using max-t_j composition. When the similarity matrix is also t_j, Corollary 7 specializes the inclusion to equality.

Lemma 3. Let $\mathbf{S}$ be a similarity matrix over U with $|U| = n$. Then $\mathbf{S}^n = \mathbf{S}^{n+k}$ for $k > 0$.

Lemma 3 follows from the observation that there is an acyclic path that produces the maximal influence between two elements. Since t-norms satisfy $x \otimes y \leq x$, the presence of a cycle in a path from u_i to u_j produces an influence that is at most that generated by the path obtained by removing the cycle. Since every path of length greater than n must contain a cycle, no such paths can be discovered that would increase the influence.

The construction of the maximal influence matrix $\mathbf{S}^n$ is a variation of the all pairs shortest path algorithm. The algebraic system $(U, \vee, \otimes, 0, 1)$ forms a closed semi-ring and the result follows from the general theory [1,3] of labeled paths in closed semi-rings.

5. Iterative filtering

We will now consider the properties of repeated iterations of a fuzzy filter. Let $\mu_A^k(u_i)$ be the membership value of the element u_i after the k'th iteration of filtering fuzzy set A by filter f. The resulting fuzzy set is denoted $f^k(A)$. The halting condition can be activated by a preassigned number of iterations or when the process converges to a steady state. Proposition 1 and Lemma 3 combine to show that convergence will always occur for an increasing filter.

Proposition 1. Let f be an increasing fuzzy filter defined using t-norm $\otimes$ and t-conorm t_3 (max). Then the fuzzy set $f^k(A)$ is given by the product $\mathbf{A} \circ \mathbf{S}^k$.

Proof: By definition

$$\begin{aligned}\mu_{f^1(A)}(u_i) &= \bigvee_{j=1}^{n} \mu_A(u_j) \otimes S(u_j, u_i) \\ &= \bigvee_{j=1}^{n} \mathbf{A}(j) \otimes \mathbf{S}(j,i) \\ &= \mathbf{A} \circ \mathbf{S}(i).\end{aligned}$$

Now assume that $f^k(A)$ is given by $\mathbf{A} \circ \mathbf{S}^k$.

$$\begin{aligned}\mu_{f^{k+1}(A)}(u_i) &= \bigvee_{j=1}^{n} \mu_A^k(u_j) \otimes S(u_j, u_i) \\ &= \bigvee_{j=1}^{n} \left(\bigvee_{t=1}^{n} \mathbf{A}(t) \otimes \mathbf{S}^k(t,j) \right) \otimes \mathbf{S}(j,i) \\ &= \bigvee_{j=1}^{n} \bigvee_{t=1}^{n} \mathbf{A}(t) \otimes \mathbf{S}^k(t,j) \otimes \mathbf{S}(j,i) \\ &= \bigvee_{t=1}^{n} \mathbf{A}(t) \otimes \left(\bigvee_{j=1}^{n} \mathbf{S}^k(t,j) \otimes \mathbf{S}(j,i) \right) \\ &= \bigvee_{t=1}^{n} \mathbf{A}(t) \otimes \mathbf{S}^{k+1}(t,i) \\ &= \mathbf{A} \circ \mathbf{S}^k(i).\end{aligned}$$

The second step follows from the inductive hypotheses. The remainder of the argument consists of routine algebraic manipulation. □

Corollary 8. The fuzzy sets produced by an increasing or decreasing filter will converage after at most $|U|$ iterations.

The computations given in Table 2 show the effect of iterating the filtering process begun in Example 3. The similarity relation S_2 defines the neighborhood of an element as those elements immediately adjacent to it in the ordering $u_1, u_2, \ldots, u_{12}$. The tables show the transformations obtained by iterated filtering with the t-norm and t-conorm indicated. The number on the left hand side indicates the iteration. The initial membership function is given as iteration 0. The columns in row i give the value $\mu_{f^i(A)}(u_j)$. The results of the iterated filtering up to the convergence of the membership function or to the tenth iteration, whichever comes first. If the process has not completed in that time, the final converged values and the number of iterations required are given.

```
t-norm 1  t-conorm 3

0   0.000 0.000 1.000 0.000 0.000 0.500 0.750 0.000 0.000 0.500 0.000 0.000
1   0.000 0.500 1.000 0.500 0.000 0.500 0.750 0.250 0.000 0.500 0.000 0.000
2   0.000 0.500 1.000 0.500 0.000 0.500 0.750 0.250 0.000 0.500 0.000 0.000

t-norm 2  t-conorm 3

0   0.000 0.000 1.000 0.000 0.000 0.500 0.750 0.000 0.000 0.500 0.000 0.000
1   0.000 0.500 1.000 0.500 0.250 0.500 0.750 0.375 0.250 0.500 0.250 0.000
2   0.250 0.500 1.000 0.500 0.250 0.500 0.750 0.375 0.250 0.500 0.250 0.125
3   0.250 0.500 1.000 0.500 0.250 0.500 0.750 0.375 0.250 0.500 0.250 0.125

t-norm 3  t-conorm 3

0   0.000 0.000 1.000 0.000 0.000 0.500 0.750 0.000 0.000 0.500 0.000 0.000
1   0.000 0.500 1.000 0.500 0.500 0.500 0.750 0.500 0.500 0.500 0.500 0.000
2   0.500 0.500 1.000 0.500 0.500 0.500 0.750 0.500 0.500 0.500 0.500 0.500
3   0.500 0.500 1.000 0.500 0.500 0.500 0.750 0.500 0.500 0.500 0.500 0.500
```

a) Increasing Filter

```
t-norm 1  t-conorm 3

0   0.000 0.000 1.000 0.000 0.000 0.500 0.750 0.000 0.000 0.500 0.000 0.000
1   0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

t-norm 2  t-conorm 3

0   0.000 0.000 1.000 0.000 0.000 0.500 0.750 0.000 0.000 0.500 0.000 0.000
1   0.000 0.000 0.250 0.000 0.000 0.188 0.188 0.000 0.000 0.125 0.000 0.000
2   0.000 0.000 0.063 0.000 0.000 0.047 0.047 0.000 0.000 0.031 0.000 0.000
3   0.000 0.000 0.016 0.000 0.000 0.012 0.012 0.000 0.000 0.008 0.000 0.000
4   0.000 0.000 0.004 0.000 0.000 0.003 0.003 0.000 0.000 0.002 0.000 0.000
5   0.000 0.000 0.001 0.000 0.000 0.001 0.001 0.000 0.000 0.000 0.000 0.000
6   0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

t-norm 3  t-conorm 3

0   0.000 0.000 1.000 0.000 0.000 0.500 0.750 0.000 0.000 0.500 0.000 0.000
1   0.000 0.000 0.500 0.000 0.000 0.500 0.500 0.000 0.000 0.500 0.000 0.000
2   0.000 0.000 0.500 0.000 0.000 0.500 0.500 0.000 0.000 0.500 0.000 0.000
```

b) Decreasing Filter

t-norm 1 t-conorm 3

0	0.000	0.000	1.000	0.000	0.000	0.500	0.750	0.000	0.000	0.500	0.000	0.000
1	0.000	0.250	0.500	0.250	0.000	0.250	0.375	0.125	0.000	0.250	0.000	0.000
2	0.000	0.125	0.250	0.125	0.000	0.125	0.188	0.063	0.000	0.125	0.000	0.000
3	0.000	0.063	0.125	0.063	0.000	0.063	0.094	0.031	0.000	0.063	0.000	0.000
4	0.000	0.031	0.063	0.031	0.000	0.031	0.047	0.016	0.000	0.031	0.000	0.000
5	0.000	0.016	0.031	0.016	0.000	0.016	0.023	0.008	0.000	0.016	0.000	0.000
6	0.000	0.008	0.016	0.008	0.000	0.008	0.012	0.004	0.000	0.008	0.000	0.000
7	0.000	0.004	0.008	0.004	0.000	0.004	0.006	0.002	0.000	0.004	0.000	0.000
8	0.000	0.002	0.004	0.002	0.000	0.002	0.003	0.001	0.000	0.002	0.000	0.000
9	0.000	0.001	0.002	0.001	0.000	0.001	0.001	0.000	0.000	0.001	0.000	0.000
10	0.000	0.000	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
11	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

t-norm 2 t-conorm 3

0	0.000	0.000	1.000	0.000	0.000	0.500	0.750	0.000	0.000	0.500	0.000	0.000
1	0.000	0.250	0.625	0.250	0.125	0.344	0.469	0.188	0.125	0.313	0.125	0.000
2	0.063	0.195	0.391	0.195	0.102	0.215	0.293	0.141	0.094	0.195	0.094	0.031
3	0.064	0.122	0.244	0.122	0.066	0.134	0.183	0.091	0.061	0.122	0.061	0.031
4	0.048	0.076	0.153	0.076	0.042	0.084	0.114	0.057	0.038	0.076	0.038	0.023
5	0.036	0.048	0.095	0.048	0.026	0.052	0.072	0.036	0.024	0.048	0.024	0.018
6	0.027	0.030	0.060	0.030	0.016	0.033	0.045	0.022	0.015	0.030	0.015	0.013
7	0.020	0.019	0.037	0.019	0.010	0.020	0.028	0.014	0.009	0.019	0.009	0.010
8	0.015	0.012	0.023	0.012	0.006	0.013	0.017	0.009	0.006	0.012	0.006	0.007
9	0.011	0.007	0.015	0.007	0.004	0.008	0.011	0.005	0.004	0.007	0.004	0.006
10	0.009	0.005	0.009	0.005	0.003	0.005	0.007	0.003	0.002	0.005	0.002	0.004
20	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

t-norm 3 t-conorm 3

0	0.000	0.000	1.000	0.000	0.000	0.500	0.750	0.000	0.000	0.500	0.000	0.000
1	0.000	0.250	0.750	0.250	0.250	0.500	0.625	0.250	0.250	0.500	0.250	0.000
2	0.125	0.375	0.625	0.375	0.375	0.500	0.563	0.375	0.375	0.500	0.375	0.125
3	0.250	0.438	0.563	0.438	0.438	0.500	0.531	0.438	0.438	0.500	0.438	0.250
4	0.344	0.469	0.531	0.469	0.469	0.500	0.516	0.469	0.469	0.500	0.469	0.344
5	0.406	0.484	0.516	0.484	0.484	0.500	0.508	0.484	0.484	0.500	0.484	0.406
6	0.445	0.492	0.508	0.492	0.492	0.500	0.504	0.492	0.492	0.500	0.492	0.445
7	0.469	0.496	0.504	0.496	0.496	0.500	0.502	0.496	0.496	0.500	0.496	0.469
8	0.482	0.498	0.502	0.498	0.498	0.500	0.501	0.498	0.498	0.500	0.498	0.482
9	0.490	0.499	0.501	0.499	0.499	0.500	0.500	0.499	0.499	0.500	0.499	0.490
10	0.495	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.495
14	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500

c) Averaging Filter

0	0.000	0.000	1.000	0.000	0.000	0.500	0.750	0.000	0.000	0.500	0.000	0.000
1	0.000	0.250	0.500	0.250	0.125	0.438	0.500	0.188	0.125	0.250	0.125	0.000
2	0.083	0.250	0.375	0.281	0.234	0.375	0.406	0.250	0.172	0.188	0.125	0.042
3	0.139	0.240	0.320	0.293	0.281	0.348	0.359	0.270	0.195	0.168	0.120	0.069
4	0.172	0.235	0.293	0.297	0.301	0.334	0.334	0.273	0.207	0.163	0.119	0.086
5	0.193	0.234	0.280	0.297	0.308	0.326	0.319	0.272	0.213	0.163	0.122	0.097
6	0.207	0.235	0.272	0.295	0.310	0.320	0.309	0.269	0.215	0.165	0.126	0.105
7	0.216	0.237	0.269	0.293	0.309	0.314	0.302	0.265	0.216	0.168	0.131	0.112
8	0.223	0.240	0.267	0.291	0.306	0.310	0.296	0.262	0.216	0.171	0.135	0.118
9	0.229	0.243	0.266	0.289	0.303	0.305	0.291	0.259	0.216	0.173	0.140	0.124
10	0.233	0.245	0.266	0.287	0.300	0.301	0.287	0.256	0.216	0.176	0.144	0.129
280	0.239	0.239	0.239	0.239	0.239	0.239	0.239	0.239	0.239	0.239	0.239	0.239

d) Similarity weighting filter

Table 2: Iterated Filters

The first application of an increasing filter, Table 2 a), raises the membership values of the neighbors of the elements with high membership values. A second iteration shows that the effect of a filter extends beyond the neighborhood defined by the similarity relation. The *max-min* filter progates a value of .5 to all the elements that were originally assigned 0.

A decreasing filter would appear to be an appropriate choice for filtering this membership function. The preponderance of the values are 0 and the non-zero values may be attributed to noise in the elicitation process. The filters employing t_1 and t_2 eventually converge assigning 0 to all elements. Using *min* lessens, but does not remove, the difference between the aberrant points and remainder of the universe.

Similar properties are exhibited by the averaging filters, with the two stronger t-norms converging to a uniform assignment of 0 while *max-min* converges to .5.

Conclusions

Fuzzy filters enhance the continuity of partial membership evaluations based on similarity information. These techniques may be used as a knowledge engineering tool to refine the compatibility assessments that describe the imprecisely defined concepts or for removing noise from data obtained from sensors. The varying properties of the families of filters provide the flexibility to choose transformations specifically suited for the type of noise that may be anticipated in the acquistion of information. We are currently applying fuzzy filters to remove noise from gray scale images with the choice of the filter based on local variation of pixel intensities.

References

[1] A. V. Aho, J. E. Hopcroft, and J. D. Ullman. *The Design and Analysis of Computer Algorithms.* Addison-Wesley, Reading, MA, 1974.

[2] B. Bouchon and G. Cohen. On fuzzy relations and partitions. In P. P. Wang, editor, *Advances in Fuzzy Sets, Possibility Theory, and Applications*, pages 97–106, Plenum Press, 1983.

[3] T. H. Cormen, C. E. Leiserson, and R. L. Rivest. *Introduction to Algorithms.* MIT Press, Cambridge, MA, 1990.

[4] D. DuBois and H. Prade. *Possibility Theory: An Approach to Computerized Processing of Uncertainty.* Plenum, New York, 1988.

[5] H. Dykhoff and W. Pedrycz. Generalized means as a model of compensative connectives. *Fuzzy Sets and Systems*, 14:143–154, 1984.

[6] M. Mizumoto. Pictorial representation of fuzzy connectives, part I: Cases of t-norms, t-conorms and averaging operators. *Fuzzy Sets and Systems*, 32:217–242, 1989.

[7] H. B. Potoczny. On similarity relations in fuzzy relational databases. *Fuzzy Sets and Systems*, 12:231–235, 1984.

[8] A. Rosenfeld. Fuzzy graphs. In L. A. Zadeh, K. S. Fu, K. Tanaka, and M. Shimura, editors, *Fuzzy Sets and their Application to Cognitive and Decision Processes*, pages 77–95, Academic Press, New York, 1975.

[9] E. A. Rundensteiner, L. W. Hawkes, and W. Bandler. On nearness measures in fuzzy relational data models. *Internation Journal of Approximate Reasoning*, 3(3):267–298, 1989.

[10] S. Shenoi and A. Melton. Proximity relations in the fuzzy relational database model. *Fuzzy Sets and Systems*, 31:285–296, 1989.

[11] A. Tversky. Features of similarity. *Psychological Review*, 84(4):327–353, 1977.

[12] S. Weber. A general concept of fuzzy connectives, negations and implications based on t-norms and t-conorms. *Fuzzy Sets and Systems*, 11:115–134, 1983.

[13] L. A. Zadeh. Calculus of fuzzy restrictions. In L. A. Zadeh, K. S. Fu, K. Tanaka, and M. Shimura, editors, *Fuzzy Sets and Their Applications to Cognitive and Decision Processes*, pages 1–39, Academic Press, New York, 1975.

[14] L. A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[15] L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 3–28, 1978.

[16] L. A. Zadeh. Similarity relations and fuzzy orderings. *Information Sciences*, 3:177–200, 1971.

[17] M. Zemankova-Leech and Abraham Kandel. *Fuzzy Relational Databses—A Key to Expert Systems.* (Verlag TUV Rheinland), 1984.

[18] A. Zvieli. On complete fuzzy relational query languages. In *Proceedings NAFIPS 86*, pages 704–726, New Orleans, 1986.