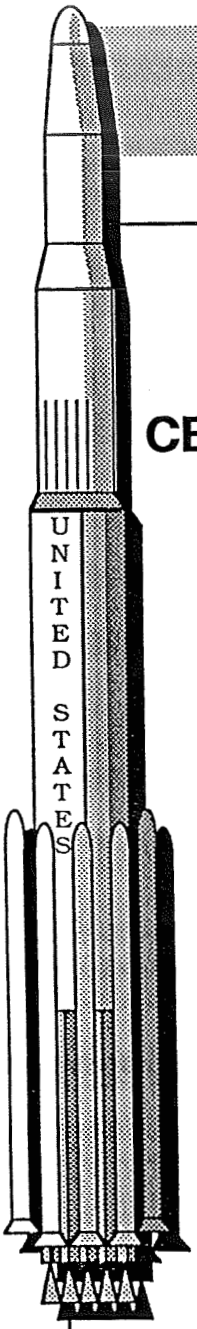
NASA-CR-199700



C A S P

Center for Advanced Space Propulsion

CENTER FOR ADVANCED SPACE PROPULSION

SECOND ANNUAL

TECHNICAL SYMPOSIUM

PROCEEDINGS

November 1st and 2nd, 1990

Center for Advanced Space Propulsion
P.O. Box 850
UTSI Research Park
Tullahoma, TN 37388

ATTACHMENT F
NAGW-1195
CSTAR-FINAL REPORT

The University of Tennessee - Calspan
Center for Aerospace Research
UTSI Research Park, Bldg 1 (615) 454-9294
Box 1385, Tullahoma, TN 37388 FAX: (615) 455-6167

CENTER FOR ADVANCED SPACE PROPULSION
SECOND ANNUAL
TECHNICAL SYMPOSIUM
PROCEEDINGS

TABLE OF CONTENTS

FINAL AGENDA

CASP TECHNICAL SYMPOSIUM

NOVEMBER 1, 1990

UTSI - AUDITORIUM

TIME		
A.M.		
7:50	REGISTRATION - LOBBY	
8:25	INTRODUCTION	DR. GEORGE GARRISON
8:30	WELCOME	DR. WESLEY HARRIS UTSI Vice President
8:35	KEYNOTE ADDRESS	MR. GLENN REYNOLDS The University of Tennessee School of Law
CHEMICAL PROPULSION, CFD		
9:00	High Area Ratio Nozzles	DR. FRANK COLLINS
9:30	OMV Variable Thrust Engine	MR. LOUIS LAROSILIERE
10:00	Spray Combustion Stability	MR. RONNIE LITCHFORD
10:30	BREAK (10 Minutes)	
10:40	SSME Leak Detection	DR. ROGER CRAWFORD
SPACE TRANSPORTATION		
11:10	COMET PROGRAM	MR. JOE PAWLICK
LUNCH -	12:15 - 1:25 P.M. - UTSI CAFETERIA	
P.M.		
ELECTRIC PROPULSION		
1:30	Electrostatic Plasma Accelerator(EPL)	MR. VERLIN FRIEDLY
2:00	Experimental Facilities	DR. WIM RUYTEN
2:30	Ion Thruster Modeling	MR. JEFF PENG
3:00	BREAK (10 Minutes)	
ARTIFICIAL INTELLIGENCE		
3:10	Identification & Interpretation of Patterns in Rocket Engine Data	DR. M. ALI
3:30	Neural Networks	DR. M. ALI
3:50	SSME Fault Monitoring & Diagnosis Expert System	DR. M. ALI
4:10	Hazardous Gas Detection Intelligent Hypertext	DR. C-F LO
4:40	Engineering Monitoring Expert Systems Developer	DR. C-F LO
5:10	Adjourn to the UTSI Recreation Area	
5:30	BARBECUE SUPPER - UTSI RECREATION AREA	

CASP TECHNICAL SYMPOSIUM

NOVEMBER 2, 1990

UTSI - AUDITORIUM

TIME A.M.		
9:00	INTRODUCTION- 2ND DAY	DR. GEORGE GARRISON
LOW-G FLUID MANAGEMENT		
9:05	Cryogenic Transfer Line Chilldown	DR. BASIL ANTAR
9:35	SOFTE Flow Visualization	DR. FRANK COLLINS
10:05	2-Phase Helical Flow	DR. EDWARD KESHOCK
ROCKET ENGINE MATERIALS		
10:35	Inconel Welding	DR. T.D. McCAY
11:05	CLOSING REMARKS	DR. GEORGE GARRISON
11:15	ADJOURN	
11:20	LUNCH - UTSI CAFETERIA	

CHEMICAL PROPULSION, CFD

Advanced High Area Ratio Nozzles

Frank G. Collins
Brian Myruski
Mark Jechura

Combustion and Flow Modelling Applied to the OMV VTE

Louis M. Larosiliere
San-Mou Jeng

Spray Combustion Stability Project

San-Mou Jeng
Ron J. Litchford

SSME Leak Detection Feasibility Investigation by Utilization of Infrared Sensor Technology

Ahmad A. Shohadaee
Roger A. Crawford

COMBUSTION AND FLOW MODELLING APPLIED TO THE OMV VTE+

Louis M. Larosiliere* and San-Mou Jeng**

Center for Advanced Space Propulsion
University of Tennessee Space Institute
Tullahoma, Tennessee 37388

ABSTRACT

A predictive tool for hypergolic bipropellant spray combustion and flow evolution in the OMV VTE is described. It encompasses a computational technique for the gas phase governing equations, a discrete particle method for liquid bipropellant sprays, and constitutive models for combustion chemistry, interphase exchanges, and unlike impinging liquid hypergolic stream interactions. Emphasis is placed on the phenomenological modelling of the hypergolic liquid bipropellant gasification processes. An application to the OMV VTE combustion chamber is given in order to show some of the capabilities and inadequacies of this tool.

Introduction

Small rocket engines are typically employed in propulsion applications requiring relatively low thrust levels and total impulse. An example of such applications is attitude control systems for vehicle orientation and orbital or plane maneuvers. TRW is currently developing a hypergolic bipropellant variable thrust engine (VTE) to be qualified specifically for the orbital maneuvering vehicle (OMV). The available analytical prediction methodology is incapable of providing physical and phenomenological insight about the complex spray combustion and flow processes occurring within thrust chambers such as the VTE. Thus, an evolutionary prediction methodology that is comprehensive enough to allow local resolution of certain observable critical processes is needed. The work to be described herein is stimulated by the desire to model some of the critical processes within a hypergolic bipropellant thruster.

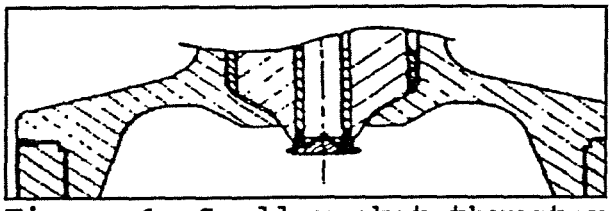


Figure 1. Small rocket thruster configuration.

Thrust chambers involving hypergolic liquid bipropellants reveal a multitude of complex chemical and physical phenomena. These phenomena include homogeneous and heterogeneous chemical reactions, liquid phase dynamics, and turbulent gas transport. A typical thrust chamber of interest is sketched in Figure 1. Injection of a hypergolic bipropellant system is accomplished via a single element coaxial flow injector with a throttling sleeve. One propellant stream flows as an axially directed jet while the other stream is turned radially just prior to injection. A complete description of the ensuing developments soon after injection is not available. It is believed that the jets impinge forming oxidizer and fuel sheets in the immediate vicinity of the injection ports. Because of the hypergolicity of the bipropellant system, chemical reactions and heat release during impingement may result in ignition within the sheets and, hence, affect the atomization, mixing and combustion of the propellants. Subsequently, the sheets disintegrate into ligaments and these ligaments eventually form a spray containing a wide spectrum of attributes. Thereafter, liquid spray droplet aerothermodynamics along with the turbulent gas transport dominate.

+ This work was jointly supported by the University of Tennessee-Calspan Center for Advanced Space Propulsion under NASA Grant NAGW-1195 and TRW, Redondo Beach, CA.

*Graduate Research Assistant, UTSI

**Assistant Professor, Mechanical and Aerospace Engineering, Principal Investigator, UTSI

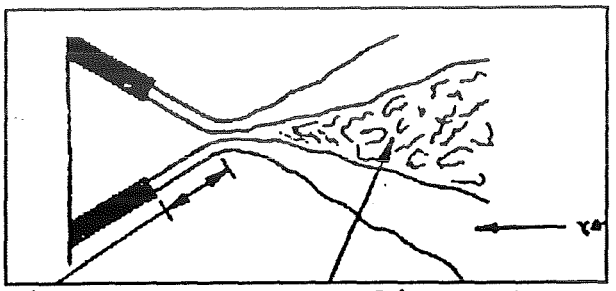


Figure 2. Hypergolic stream separation.

Previous applied research work and investigative studies have shown that a wide range of coupled and interrelated processes are manifested in a hypergolic bipropellant thruster. Elverum and Staudhammer[1] first reported that impinging hypergolic liquid streams under certain conditions have a tendency to be blown apart rather than achieve the intended degree of mixing. Burrows[2] and Breen et al[3] have made photographic observations of the impingement of two hypergolic liquid streams confirming this hypergolic stream separation tendency. Tentative criteria to predict the presence or absence of hypergolic stream separation have been advanced by Kushida and Houseman[4] on the basis of theoretical analysis. Their results showed that depending on the pressure level, stream separation originated from either liquid/liquid interfacial reactions or gas phase reactions. The physical nature of the separation process, when it did occur, was generally presumed to involve a quasi-steady lamination of the spray fans with fuel on one side and oxidizer on the other (Figure 2). Campbell et al[5] demonstrated by way of photography that the hypergolic stream separation phenomenum is a cyclic process in which explosions disrupt the spray fan and drive the jets apart, thereby producing temporary physical separation of fuel and oxidizer.

Somogyi and Feiler[6] determined the heat release rates produced by the reactions of hydrazine with nitric acid in a bomb calorimeter under conditions of forced mixing. Its heat release depends on the interfacial area between the propellants. A combustion mechanism based on bipropellant interfacial reaction was postulated for highly reactive bipropellant systems under conditions of forced mixing. This interfacial reaction process along with the usual liquid droplet combustion mechanisms emerge as the most likely modes of propellant gasification within a hypergolic thruster. Attempts at quantifying the relative importance of these two gasification modes have either not been successful or were not reported.

Most current models[7] for performance analysis of rocket combustion tend to decouple the critical processes. A zonal approach is usually taken whereby different assumed physical processes of interest are modeled and then conveniently connected. Liang et al[8] constructed a computational model for the rocket combustion of H2/LOX with coaxial injection. They described a comprehensive three-phase (liquid, gas, and spray droplets) model capable of resolving some of the atomization length and time scales. This model is computationally too intensive and not directly applicable to a liquid hypergolic bipropellant system. The state-of-the art in spray combustion modelling requires as external input an ample amount of injection system characterization in order to be useful. Recent developments[9] in spray combustion modelling aimed at the gas turbine and internal combustion engine arena have produced a wealth of simulation techniques which haven't yet been fully applied to the rocket engine combustor. Creative exploitation of these techniques for the case of a hypergolic bipropellant system remains to be done.

This paper discusses a fundamental CFD-based modelling approach which is under development for analyzing the hypergolic bipropellant spray combustion and flow evolution in the VTE thruster. An inclusive physical model is formulated and integrated within the computational framework of the KIVA-II[10] chemically reactive flow code developed for internal combustion engines. Emphasis is placed on the "sub-scale" modelling of the hypergolic liquid bipropellant gasification processes. The model is applied to the VTE thrust chamber configuration employing the bipropellant system of N2H4/N2O4. Some of the more doubtful elements of the model are identified for further refinements.

Model Formulation

Consider the thrust chamber to be a closed volume containing a reactive gas and allow a pair of continuous liquid phases to be introduced into it via an injection system. Upon entering the volume, the two distinct liquid phases will undergo mutual interactions as well as interactions with the gas phase. A tendency to disintegrate will set in and eventually the liquid will breakup into ligaments and a spectrum of different drop sizes, and through different intermediate steps, depending on the relative

magnitudes of the aerodynamic, viscous, inertia and surface forces at play. Concurrently, the motion of the gas is altered by the interfacial exchanges with the liquid. Hence, the realm of two-phase flow is entered. Although conservation equations may be formulated for both phases, it should be noted that for the liquid phases, physical conservation only has meaning over a finite portion of the volume and a finite time interval. This is due to the discontinuous breakup and turbulent dispersion processes. In order to keep the problem manageable, a model consisting of two dispersed liquid phases imbedded within a continuum gas is formulated. The dispersed liquid phases are taken to be collections of quasi-spherical liquid material volumes (i.e. "globules") capable of undergoing deformations and breakup. Details of the derivations may be found by consulting the dissertation of O'Rourke[11]. A brief description of the mathematical model follows.

Dispersed Liquid Phase

The evolution in space and time of certain average properties of a collection of entities may be obtained in a manner analogous to the method of kinetic theory. A globule number distribution function is defined for globules of composition j in such a way that

$$f_j(\vec{s}_j, t)\, dV_j \qquad j=1,$$

is the probable average number of globules in the infinitesimal globule hyper volume. The subscript j differentiates between oxidizer and fuel globules. S_j is a vector whose coordinates give the values of those properties specifying the state of a globule at a given instant.

$$dV_j = ds_{j_1} ds_{j_2} \cdots ds_{j_N}$$

is the infinitesimal globule hyper volume in the N dimensional state space. A stochastic equation giving the rate of change of the distribution function over a continuous time interval follows:

$$\frac{d_j}{dt}\int f(\vec{s}_j, t)\, dV_j = \int \dot{f}_{coll,j} dV_j + \int \dot{f}_{bu,j} dV_j \qquad (1)$$

This equation states that the rate of change of globule number along the globule trajectory in its state space is balanced by: collision rates among globules of the same composition as well as those of another composition, and breakup rates. The globule state vector is taken to be:

$$\vec{s}_j = \begin{bmatrix} \vec{x}_j \\ \vec{v}_j \\ r_j \\ T_{d,j} \\ y_j \\ \dot{y}_j \end{bmatrix}$$

where,

$\vec{x}_j$ globule position vector,
$\vec{v}_j$ globule velocity vector,
r_j globule radius,
$T_{d,j}$ globule temperature,
y_j globule spherical deviatio
$\dot{y}_j$ globule oscillation.

The evolution equations for the globule state variables may be obtained by taking appropriate moments of equation (1). Characteristic equations governing the trajectory, acceleration, vaporization, heat-up, deformation, and oscillation of a globule are thus furnished. These are Lagrangian equations describing the development of the state of a globule from some initial state.

Continuum Gas Phase

Equation (1) is directly coupled with the gas dynamics. The Navier-Stokes equations are valid for the average gas properties provided suitable source terms are added to account for the effect of the dispersed liquid phases. In the formulation, the averaging is carried over the entire spatial volume. Hence, the gas properties are viewed as being continuously distributed throughout the entire volume. This point of view eliminates the need to carry along the direct effect of the liquid phase volume displacement. If the gas is considered to be a turbulent reactive mixture of ideal gases, equations for the conservation of species mass, mixture momentum, and internal energy may be formulated. A modified k-epsilon turbulence closure along with a gradient diffusion model are employed to describe the fluctuation correlations. When the source terms are given along with initial and boundary conditions, a solution may be sought describing the average gas field variables.

The limitations of the present formulation are the source term modelling, and the numerical solution technique. These two subjects are considered next.

Sub-Scale Modelling

Sub-scale modelling attempts to describe phenomena occurring on a scale much smaller than that of the macroscale view. The formulation needs a certain degree of a priori knowledge of phenomenological processes occurring on the sub-scale level. As such, it lacks the details and also the fidelity of the macroscale model.

Dispersed Phase Interactions

The collision source term in equation (1) depends upon the details of globule interactions. Only binary collisions are considered. Following the analysis of Williams[12] under the assumption that globule collisions are similar to collisions among hard spheres, an expression may be formulated for the collision source term. The derivation is complex and will not be repeated here. Suffice to say that the result is:

$$\dot{f}_{coll,\alpha}=\Gamma_\alpha^{(+)}-\Gamma_\alpha^{(-)}$$

where,

$$\Gamma_\alpha^{(-)}=f_\alpha\int f_\alpha(\xi'_\alpha,\vec{x},t)\,|\vec{v}'_\alpha-\vec{v}_\alpha|\sigma^{(-)}(\xi_\alpha;\xi'_\alpha)\,d\xi'_\alpha$$
$$+f_\alpha\int f_\beta(\xi_\beta,\vec{x},t)\,|\vec{v}_\beta-\vec{v}_\alpha|\sigma^{(-)}(\xi_\alpha;\xi_\beta)\,d\xi_\beta$$

$$\Gamma_\alpha^{(+)}=\int f_\alpha(\xi'_\alpha,\vec{x},t)\,f_\alpha(\xi''_\alpha,\vec{x},t)\,|\vec{v}'_\alpha-\vec{v}''_\alpha|\sigma^{(+)}d\xi'_\alpha d\xi''_\alpha$$
$$+\int f_\alpha(\xi_\alpha,\vec{x},t)\,f_\beta(\xi_\beta,\vec{x},t)\,|\vec{v}_\alpha-\vec{v}_\beta|\sigma^{(+)}d\xi_\alpha d\xi_\beta$$

$$\xi=\begin{bmatrix}\vec{v}\\ x\\ y\\ \dot{y}\\ T_d\end{bmatrix}$$

The above expressions include both the effects of collisions among globules of the same composition as well as those among α and β globules. Provided aerodynamic distortions of the globules are small, an expression for the geometrical collision crossection is:

$$\sigma^{(-)}(\xi;\xi')=\pi(r+r')^2$$

The differential crossection for the production of globules, $\sigma^{(+)}$ needs to be modelled depending on the outcomes of a collision.

Collisions among globules of the same composition are rare under the present circumstances and will not be discussed further. The possible outcomes of a collision between alpha and beta globules are postulated to be: 1) an interface contact collision whereby the globules maintain their sizes and temperatures but undergo velocity changes, or 2) an interfacial mixing allowing for the possibility of an interfacial chemical reaction. The boundaries of these two possible outcomes are defined by a reaction probability function given by:

$$\eta_c=\frac{\tau_{res}}{\tau_{chem}}$$

where the interfacial residence time under the assumption that the alpha globule is the lager of the two is,

$$\tau_{res}=\frac{1}{\int\pi(r_\beta+r_\alpha)^2|\vec{v}_\alpha-\vec{v}_\beta|f_\beta(\xi_\beta,\vec{x},t)\,d\xi_\beta}$$

and the chemical conversion time is,

$$\tau_{chem}=\frac{1}{\int k_f(T_1)\,v_\beta\frac{m_\beta}{\hat{w}_\beta}f_\beta(\xi_\beta,\vec{x},t)\,d\xi_\beta}$$

The reaction rate is assumed to have an Arrhenius temperature dependence so that

$$k_f(T_1)=A\exp(-\frac{E_a}{T_1})$$

A critical impact parameter is defined as:

$$b_{cr}^2=\eta_c(r_\alpha+r_\beta)^2$$

If the impact parameter is greater than the critical impact parameter, a non-reacting contact collision is possible otherwise, an interfacial chemical reaction results. When a reaction results, the law of mass action is employed along with a chemical stoichiometry in order to calculate the rate of change of the liquid mass concentration. The globule velocities and temperature are set equal to the liquid temperature and velocity in the volume containing the globules. For a non-reacting contact reaction, the

globule velocities are assumed to be the same as the liquid velocity. The differential crossection for production of globules may be written as:

$$\sigma^{(+)}(\xi;\xi_\alpha;\xi_\beta) = \pi\frac{b_{\alpha\beta}^2}{2}\left[\delta(\xi-\xi_\beta^c)-\delta(\xi-\xi_\alpha^c)\right]$$
$$+\pi\int_{b_{\alpha\beta}}^{r_\alpha+r_\beta}\left[\delta(\xi-\xi_\alpha')+\delta(\xi-\xi_\beta')\right]$$

where the superscript c and the prime refer to appropriate values of the state variables after collision.

Dispersed Phase Breakup

The globule breakup model is based upon a spring-mass analogy developed by O'Rourke and Amsden[13]. This model was formulated for a single component spray. Nevertheless, it is being applied separately to either oxidizer or fuel sprays since the two streams are immiscible. Breakup times are adequately predicted for the "bag" and "stripping" breakup regimes. Product drop sizes and distributions are not directly predicted by the model. Energy conservation arguments are invoked in order to derive product drop average sizes which are then distributed via a chi-squared distribution.

Droplet Combustion

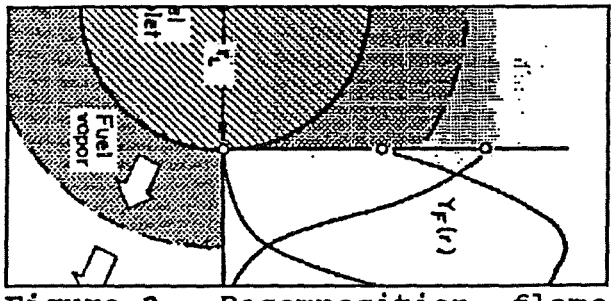


Figure 3. Decomposition flame monopropellant droplet vaporization.

Once the globules breakup into droplets, vaporization becomes dominant. Presently, only isolated droplet vaporization is considered. An extensive review of droplet vaporization has been given by Faeth[14]. For high ambient temperatures, experiments reveal the formation of an exothermic decomposition flame enveloping the hydrazine-based fuel droplets as depicted in Figure 3. The existence of the decomposition flame depends on the convective environment and the decomposition kinetics. Two modes of vaporization are considered for the fuel droplet vaporization: 1) decomposition flame within a diffusion mantle, or 2) pure diffusion flame without any decomposition. The formulation of the model is an adaptation of that given by Faeth[15] for pure monopropellant droplets.

Detail computations of the vaporization of N2O4 droplets show that an endothermic decomposition occurs rapidly very close to the droplet surface. A bipropellant droplet vaporization model given in reference 14 has been used by many other investigators for the N2O4 droplets. For lack of a better model, the same procedure is invoked here.

Gas Phase Chemistry

The N2H4-N2O4 chemistry has been investigated by Sawyer[16]. It was surmised that the N2H4-N2O4 reaction is a two step process. The initial step is the reduction of the NO2 formed from N2O4 decomposition to NO by the action of hydrazine. The second step is the decomposition of hydrazine and the reduction of NO. A five step chemical kinetic scheme (Table I) is employed herein. Equilibrium reactions are not considered since the kinetic reaction rates for this hypergolic system are sufficiently fast to allow the flame-front to propagate.

Table I Gas phase kinetic reactions.

N2H4-N2O4 System:
Decomposition
2N2H4->2NH3+H2+N2
N2O4->2NO2
Oxidation
2N2H4+2NO2->4H2O+3N2
2NH3+H2+4NO2->4H2O+N2+4NO
N2H4+2NO->2H2O+2N2

Droplet Acceleration and Turbulence Dispersion

The acceleration of a droplet depends on the external forces exerted upon it. For the present application, only skin friction and separation drag are considered. Correlations[17] which allow both for vaporization and distortion are employed for the drag coefficient.

The turbulence effects on droplet motion are simulated stochastically by superimposing upon the gas field turbulent eddies each having a length, lifetime, and fluctuation velocities. A discussion of the technique is given by Faeth[18].

Solution Procedure

The approach to solving the globule number conservation equation is to discretize the globule streams into computational parcels that follow globule characteristic paths. Each parcel represents a number of globules of identical states. In the KIVA-II code, the globule number distribution function is approximated by Dirac delta functions. Representing the globule streams by discrete parcels rather than by continuous distributions amounts to a statistical description of the problem, since the finite number of parcels used represents a sample of the total globule population. The globule characteristic equations are integrated implicitly where appropriate due to their strong coupling with the gas phase. Thus, the dispersed liquid phase solution procedure amounts to what is called a stochastic discrete particle tracking.

The gas phase solution procedure is based on an axisymmetric finite volume method called the ALE (Arbitrary Lagrangian Eulerian) method. A staggered mesh is used wherein scalar quantities are defined at cell centers and velocity components at cell vertices. The momentum equation is integrated over momentum cells centered around each grid nodes. Time marching is performed by use of operator splitting and the solution is advanced in a series of phases. A Lagrangian phase whereby the grid nodes move with the local flow is performed first, followed by a rezone phase to set the required grid motion relative to the flow, and finally a convective fluxing phase. Processes that strongly couple the globule and spray dynamics are treated implicitly. A variable implicitness is employed in the gas phase time marching algorithm along with a SIMPLE-styled pressure correction procedure. Details of the computational scheme are fully described in reference 10.

Application to the OMV VTE

The OMV VTE thrust chamber configuration with a central single element coaxial injector is modelled with the N2H4-N2O4 propellant system. In this simulation, it is assumed that spray droplets are formed right at the injection ports. Hence, the stream impingement and breakup processes are not modelled. This case is a baseline case of pure spray combustion with some details of the injection system characteristics known. Table II lists the relevant parameters for this simulation.

Table II Conditions for spray combustion simulation.

	N2H4	N2O4
SMD (cm)	4.E-3	5.2E-3
Cone ang.	20.0	20.0
Tilt ang.	40.0	40.0
v_{inj} (cm/s)	1270.	1300.
T_{inj} (K)	320.	310.

At time zero, N2H4 and N2O4 spray droplets with a Sauter Mean Diameter (SMD) as given in an Table II are introduced within a volume containing an initially inert gas (N2). A chi-squared initial drop size distribution is used. The spray droplets interact only with the gas via momentum, mass, energy, and turbulence exchanges. Liquid droplet vaporization is the only gasification mode for this simulation. The computational grid is illustrated in Figure 4. Since this is an axisymmetric configuration, only half of the geometry is discretized. After a time of about 4 ms, an almost steady state combustion is established. Figures 5 and 6 show the droplet trajectories at thistime. It may be observed that both fuel and oxidizer droplets are found near the wall especially near the cylindrical portion. The wall boundary conditions are set as free-slip with wall velocities determined by a wall function matching approach. An adiabatic wall temperature is specified along with a constant pressure outflow boundary condition. The propellant mass flow rate is 197 grams/s with an oxidizer fuel ratio of 1.25. Spray injection is accomplished by injecting 100000 parcels per second in such a way as to establish the required propellant mass flow rate. Figure 7 shows the axisymmetric velocity field at a time of 4 ms. A small recirculation region is established in the injector bowl close to the injector face. The average velocity inside the chamber is about 30 m/s. Away from the injector, a unidirectional flow field prevails. The pressure temperature fields are illustrated in Figure 8. A maximum temperature of 2200 K along with an

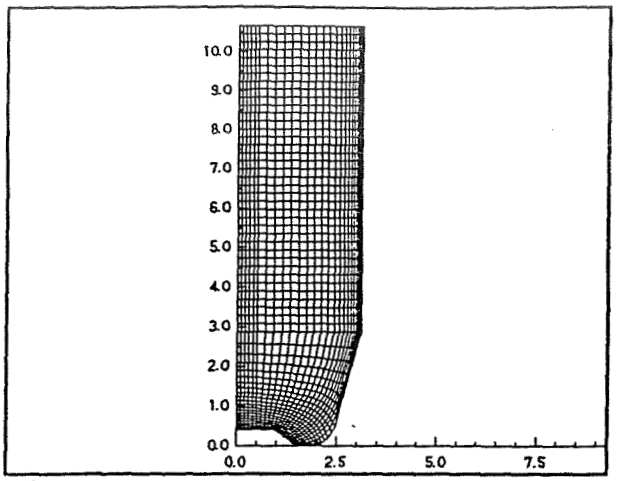


Figure 4. OMV VTE thrust chamber computational grid.

Figure 5. N2H4 spray droplet velocity field.

Figure 6. N2O4 spray droplet velocity field.

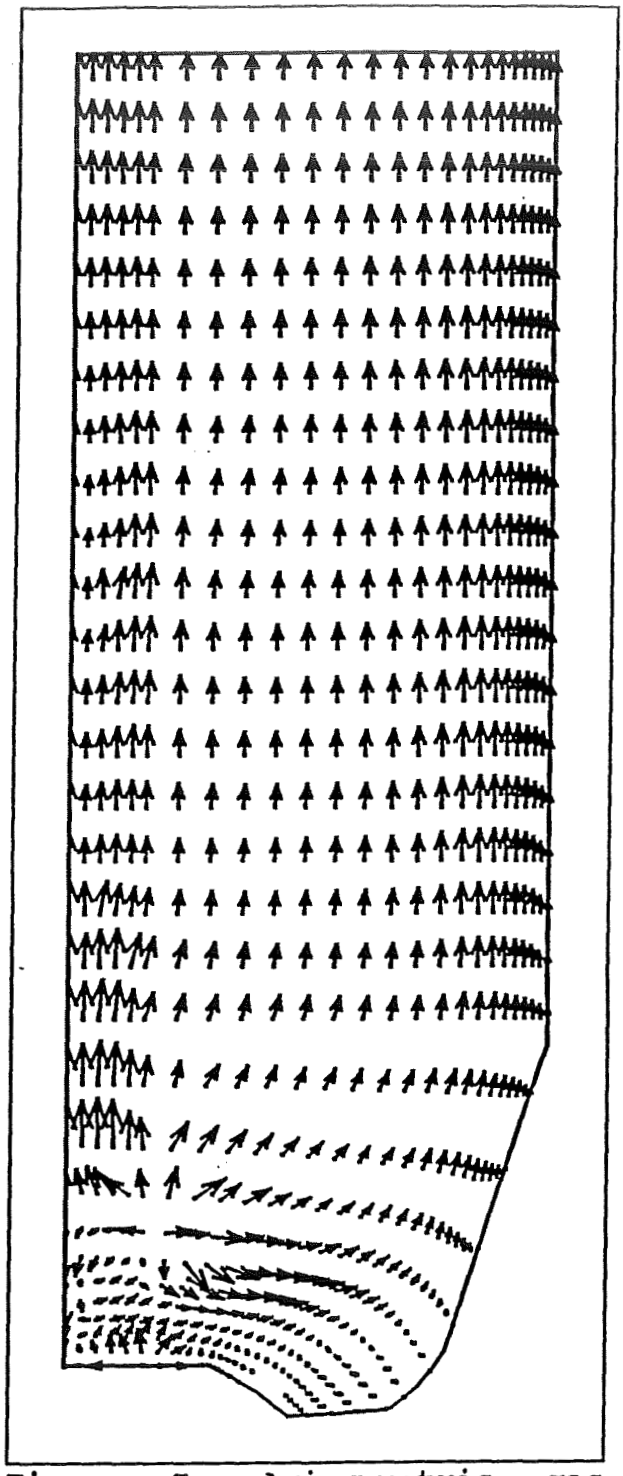


Figure 7. Axisymmetric gas velocity field.

average pressure of about 7 atm can be seen within the chamber. Mass fraction contours are presented in Figure 9. Propellant decomposition is marked by regions of NH3 for hydrazine and NO2 for nitrogen tetroxide. It may be noted that a majority of the hydrazine decomposition occurs near the chamber axis where the relative velocity between the gas and a droplet is low.

Summary and Conclusions

An inclusive modelling framework has been established for the VTE thrust chamber. The gross effects of the injection and atomization processes are included. Computational and experimental tests are required to fine tune some of the sub-models. A baseline computation of pure spray combustion without the effects of spray interactions has been given. Future computations with the complete model

are forthcoming. After further refinements, this model should prove most useful in the development of small rocket thrusters such as the VTE.

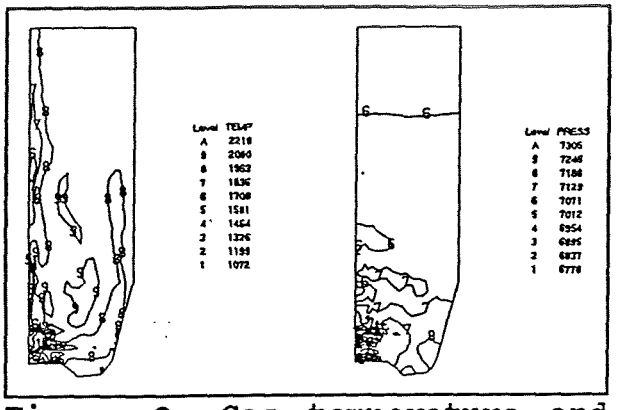


Figure 8. Gas temperature and pressure contours at a time of 4 ms.

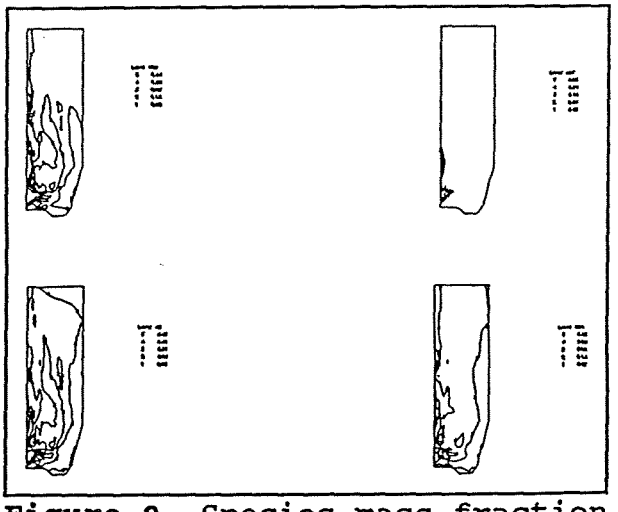


Figure 9. Species mass fraction contours: N2, H2O, NH3, NO2.

ACKNOWLEDGMENTS

This work was conducted as part of an ongoing combustion stability project supported by the University of Tennessee-Calspan Center for Advance Space Propulsion, under NASA Grant NAGW-1195.

REFERENCES

1. Elverum, G. W., Jr. and Standhammer P., "The Effect of Rapid Liquid-Phase Reactions on Injector Design and Combustion in Rocket Motors," Progress Report 30-4, Jet Propulsion Lab, August 1959.

2. Burrows, M. C., "Mixing Reaction of Hydrazine and Nitrogen Tetroxide at Elevated Pressure," AIAA Journal, Vol. 5, No. 9, Sept. 1967, pp. 1700-1701.

3. Breen, B. P. et al., "Injection and Combustion of Hypergolic Propellants," AFRPL-TR 69-48, April 1969.

4. Kushida, R. and Houseman, J., "Criteria for Separation of Impinging Streams of Hypergolic Propellants," TM 33-395, Jet Propulsion Lab, July 1968.

5. Campbell D. T. et al., "Reactive Stream Separation Photography," AIAA Paper No. 70-608.

6. Somogyi, D. and Feiler, C. E., "Liquid-Phase Heat Release Rates of the Systems Hydrazine-Nitric Acid and Unsymmetrical Dimethylhydrazine-Nitric Acid," NASA TN-D-469, Sept. 1960.

7. Schuman, M. D. et a., "Transient Performance Programs," Vol. I, AFRPL-TR 80-22, Jan. 1981.

8. Liang, P. Y., et al., "Comprehensive Modeling of a Liquid Rocket Combustion Chamber," Journal of Propulsion and Power, Vol. 2, March 1986, pp. 97-104.

9. Amsden, A. A. et al., "KIVA: A Comprehensive Model for 2D and 3D Engine Simulations," SAE Technical Paper 850554 (1985).

10. Amsden, A. A. et al., "KIVA-II: A Computer Program for Chemically Reactive Flows with Sprays," Los Alamos National Lab Report LA-11560-ms (May 1989).

11. O'Rourke, P. J., "Collective Drop Effects in Vaporizing Liquid Sprays," Ph.D. Thesis 1532-T, Princeton University (August 1981).

12. Williams, F. A., "Progress in Spray Combustion Analysis," Eighth Intern. Symp. on Combustion, pp. 50-69 (1962).

13. O'Rourke, P. J., Amsden, A. A., "The TAB Method for Numerical Calculation of Spray Droplet Breakup," SAE Technical Paper 872089 (1987).

14. Faeth, G. M., Prog. Energy Combust. Sci 3, 191 (1977).

15. Faeth, G. M., "Prediction of Pure Monopropellant Droplet Life Histories," AIAA Journal, Vol. 8, No. 7, pp. 1308-1314, 1969.

16. Sawyer, R. F., "The Homogeneous Gas-phase Kinetics of Reactions in the Hydrazine-Nitrogen Tetroxide Propellant System," AFOSR-66-0855, AD-634277 (1965).

17. Rabin, E. et al., "Displacement and Shattering of Propellant Droplets," AFOSR TR 60-75, 1960.

18. Faeth, G. M., "Prog. Energy Combust. Sci 9,pp. 1-76, 1983.

SPRAY COMBUSTION STABILITY PROJECT

N96-70670

Dr. San-Mou Jeng*, Principal Investigator
Mr. Ron J. Litchford†, Co-Investigator

Center for Advanced Space Propulsion
UTSI Research Park
Tullahoma, Tennessee 37388

ABSTRACT

This report summarizes research activity on the Spray Combustion Stability Project, characterizes accomplishments and current status, and discusses projected future work. The purpose is to provide a concise conceptual overview of the research effort so the reader can quickly assimilate the gist of the research results and place them within the context of their potential impact on liquid rocket engine design technology. Therefore, this report does not elaborate on many of the detailed technical aspects of the research program. Technical issues are addressed in substantial detail in recent AIAA papers which were presented at the 26^{th} Joint Propulsion Conference in July 1990 and are referenced herein.

I. Project Synopsis

The Spray Combustion Stability project is a research program which is cost shared by CASP and Rocketdyne Div., Rockwell Int. Corp. The principal goal of the project is to advance liquid rocket engine design technology, specifically analysis of resonant combustion instability. Project work began in fiscal year 1989 and is scheduled to continue through fiscal year 1991.

The specific objectives of the research effort as elaborated on in the Statement of Work (SOW) is to evaluate spray combustion physical submodels for application in CFD instability analyses. Submodels found deficient were to be refined or replaced by newly developed ones. In either case, the resulting submodels were to be validated with experimental data insofar as possible.

In the original SOW, Rocketdyne's ARICC spray combustion code was to serve as the testbed for submodel evaluation with improvements incorporated as second-generation code developments. During the initial part of the first year's effort, it became apparent that evaluation using the ARICC code was a tedious and lengthy procedure. Since a large amount of computer time and some degree of expertise with the code was required to obtain a converged solution, productivity was low. A decision was therefore made to evaluate and develop submodels independent of and external to ARICC. This method proved not only more efficient but more practical since physical processes could be studied independently. Final implementation of submodel improvements is being left to the discretion of code developers and users at Rocketdyne.

Based on a thorough review of the scientific literature and some contemplative reasoning, we have thus far concentrated our attention on two spray combustion physical processes which are inherently important in the proper assessment of resonant combustion instability and which are either inadequately or inefficiently simulated by current mathematical submodels. The physical processes considered are (a) high-pressure droplet gasification and (b) dispersion of the spray droplets and the associated spatial distribution of mass and energy release. These processes have been the focus for submodel development during this project. The technical progress achieved thus far is summarized in the following sections.

II. Technical Discussion

The guiding rationale in selecting spray combustion physical submodels for evaluation was the known or potential role of a particular physical process as a combustion instability mechanism or the pressing requirement to simulate a particular physical process with greater accuracy and efficiency during unstable combustion than is normally required during stable combustion. The basis of selection led to consideration of the two physical processes of high-pressure gasification and droplet dispersion. Other physical processes are certainly fundamental to combustion instability, but these seemed to have the greatest potential for significant improvement within the scope and limits of this project.

Because liquid rocket engines typically operate at high chamber pressures, generally above the critical pressure of the propellants, the process of high-pressure droplet gasification plays a fundamental role in combustor dynamics. This role can become even

---

* Assistant Professor, Mechanical and Aerospace Engineering, UTSI.

† Research Engineer, Center for Advanced Space Propulsion.

more central when considering acoustic pressure fluctuations associated with resonant instability. In instability analyses, it has become a common assumption to consider propellant vaporization as a rate-controlling feedback mechanism for Rayleigh-type driving of the acoustic waves. Thus, because the vaporization process is demonstrably sensitive to pressure, proper assessment of droplet gasification can be essential to the validity of the instability analysis. In many cases, this submodeling issue has been averted by the application of low-pressure models under high-pressure conditions with the hope that the low-pressure model sufficiently approximates the high-pressure process. Evaluation of high-pressure gasification submodeling was therefore seen to be and became a major concern of this research program.

Accurate description of droplet dispersion and the associated spatial distribution of mass and energy release was also found to be an important physical spray process that needed submodeling upgrades in obtaining an efficient technique suitable for instability analyses. The currently accepted modeling philosophy is to represent many physical droplets by a computational parcel which interacts with the flow as a single effective droplet. That is as a point source for interphase transport. To simulate dispersion due to turbulent fluctuations, a stochastic sampling procedure is conventionally utilized. However, the difficulty with this approach, particularly for instability analysis, is essentially one of computational efficiency. To obtain reasonable dispersion results using the current stochastic procedure, a statistically significant number of parcels (typically several thousand) must be evaluated. Similarly, to prevent the generation of large fluctuating interphase transport sources in space and time (computational shot noise), a large number of parcels must be evaluated to more uniformly distribute the source effects throughout the computational cells. The level of accuracy desired for instability analyses therefore imposes severe computational penalties by requiring an excessive number of parcels using current stochastic sampling techniques. Improved modeling of the dispersion process was needed and this problem has been addressed in our research effort.

High-Pressure Droplet Gasification

The high-pressure droplet gasification modeling effort has been ongoing since inception of this project. Early submodel development has been summarized in a 1989 interim report[1] with later developments included in a recent AIAA conference paper.[2] The reader is referred to these supplementary publications for mathematical details and graphical presentation of computational results. The intent of this section is to present a brief conceptual description of the formulated submodel and a summary of performance evaluation including current capabilitites and limits. Tasks accomplished in the high-pressure droplet gasification submodeling effort are summarized in Table 1.

TABLE 1
Gasification Modeling Accomplishments

(1) Formulation of transport model for subcritical vaporization including surface regression, bulk thermal expansion, and internal circulation effects.
(2) Implementation of a high-pressure thermodynamic vapor-liquid equilibria model to account for real-gas effects and ambient-gas absorption.
(3) Evaluation of submodel with experimental data for n-heptane droplets vaporizing in hot, high-pressure nitrogen gas.
(4) Calibration of thermodynamic model for O_2/H_2 phase equilibria.
(5) Analysis of LOX droplets vaporizing in high-pressure, hydrogen-rich gas.

The aim of the theoretical development was to obtain a realistic physical submodel with reasonable computational efficiency. A detailed comprehensive analysis is not our purpose. In particular, the desire was to bring high-pressure effects on thermodynamic and transport processes into consideration while maintaining mathematical simplicity. This required reformulation of the basic transport model and inclusion of a high-pressure multicomponent thermodynamic vapor-liquid equilibria model. In addition, supercritical gasification had to be considered since sufficiently high-pressures can drive the droplet surface to the thermodynamic critical state.

The reformulated subcritical transport equations for the gas-phase are based on the quasi-steady assumption with inclusion of surface regression and bulk thermal expansion effects which are neglected in low-pressure models. Transport is modeled as a spherically-symmetric, diffusion-controlled process with empirical corrections for convective effects. Liquid-phase energy transport is currently handeled using 3 alternative models: (1) infinite conductivity model, (2) pure conduction model, and (3) effective conductivity model. The effective conductivity model is the most sophisticated approach and attempts to account for internal circulation effects.

Because real-gas effects and ambient-gas absorption becomes important at higher pressures, a high-pressure thermodynamic model has been implemented to provide the species concentration boundary condition at the drop surface and to obtain a corrected enthalpy of vaporization which can devi-

ate substantially for the pure component latent heat of vaporization. The model is currently applicable to weakly polar molecules and mixtures containing quantum gases (eg. hydrogen). There are some extensions of the model that can handle aqueous mixtures which try to account for the large dipole moment of the H_2O molecule, but they are highly approximate.

Once the droplet surface reaches the critical state, a supercritical gasification model is needed. Unfortunately, there is as yet no definitive experimental description of the critical transition and supercritical gasification processes. It is therefore necessary to resort to ad hoc schemes based on plausible presuppositions. The simplest approach is to assume flash supercritical transition of the entire droplet, but this seems unrealistic since the droplet should actually have internal temperature and concentration gradients. To account for a presumed finite-rate process, two alternative schemes are utilized. In one scheme, the basic idea is to transform to a constant temperature boundary condition corresponding to the critical mixing point and apply the subcritical diffusion-controlled transport relationships. This is considered as a representative lower limit for the gasification rate. An alternative scheme is to model the process using an empirical stripping mode breakup expression. This is considered as a representative upper limit for the gasification rate.

A FORTRAN computer code was developed to evaluate the proposed submodel. Initially, validation calculations were made for comparison with experimental data on n-heptane droplets vaporizing in high-pressure nitrogen gas. Agreement was good verifying the basic validity of the code and the subcritical transport model. Because of unavailable experimental data, there was no way to evaluate the supercritical gasification schemes.

The code was further modified to consider LOX droplets vaporizing in high-pressure, hydrogen-rich gas since this propellant combination is of great interest for high-performance liquid rocket engines. The thermodynamic model was calibrated using high-pressure solid oxygen/hydrogen equilibria data. High-pressure effects on thermodynamic and thermophysical properties were also considered. Representative computations were then made to demonstrate the transition process from steady-state vaporization to fully transient heat-up to the critical state with subsequent supercritical vaporization. Presentation and discussion of the results are in an accompanying AIAA paper.[2]

Additional validation with experimental data would be desirable. It should also be stated that formulations for including an envelope flame have been constructed but not included in the code because of the additional complications which would result from the required extension of the thermodynamic model at this time. Extension of the thermodynamic model for a wider range of substances including polar molecules (eg., H_2O) will be necessary before practical implementation in a comprehensive combustor code. Nevertheless, the model is currently useful for examining the general effect of pressure on the vaporization process and the resulting influence on combustion instability.

Turbulent Droplet Dispersion

The turbulent droplet dispersion submodeling effort began in the latter part of 1989. Based on the difficulties encountered with current dispersion models as previously discussed, the conceptual basis for a novel modeling technique aimed at avoiding these deficiencies was conceived. Development of the model has proceeded at a quick pace, and technical progress has been summarized in a recent AIAA conference paper.[3] The reader is referred to this publication for mathematical details and presentation of computational results. This section will be devoted to providing a conceptual description of the model and a report of model evaluation efforts to date. Tasks accomplished in the droplet dispersion submodeling effort are summarized in Table 2.

The fundamental concept of this new model was to combine a statistical spatial distribution representation for each parcel with a statistical turbulent dispersion width transport scheme. Specifically, each parcel represents a group of physical droplets through a normal probability density function (PDF) in space. The instantaneous mean position of each parcel PDF is determined from Lagrangian tracking using either a deterministic or stochastic sampling procedure. In addition the variance of each parcel PDF is determined from a statistical, turbulence-induced mean-squared dispersion which depends upon previous parcel interactions with the characteristic turbulence

TABLE 2
Dispersion Modeling Accomplishments

(1) Conceptualization and mathematical formulation of a novel model for turbulent droplet dispersion with PDF representation for droplet parcels.
(2) Proof-of-concept computations for new model in comparison with conventional point source model for nonevaporating particles injected into nearly homogeneous turbulent flow.
(3) Proof-of-concept computations for new model in comparison with conventional point source model for nonevaporating particles injected into a round turbulent jet.

properties of the flow. This approach allows for a smooth distribution of source terms over the grid cells with reduced sampling requirements for obtaining the dispersion width. A concept schematic of the model with stochastic sampling for the mean parcel position is depicted in Figure 1.

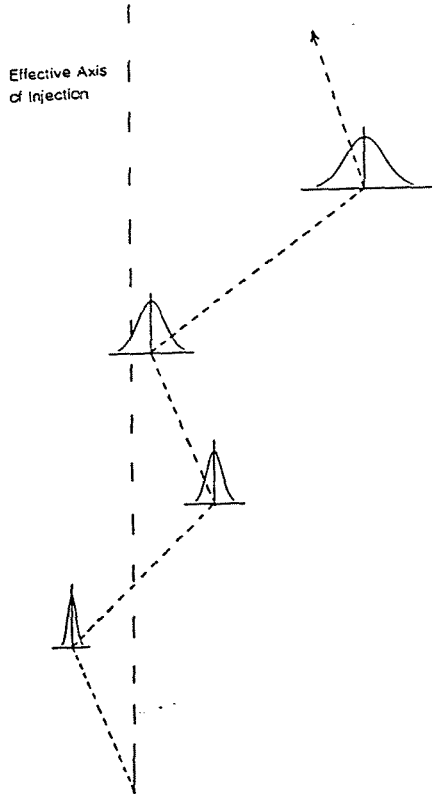


Figure 1. Concept schematic of proposed dispersion width transport model with stochastic sampling for mean parcel position.

The dispersion width transport scheme is the key to increased efficiency since it makes use of statistical generalizations allowing reduction in sampling. The idea is to retain the concept of parcels interacting with a sequence of turbulent eddies as in the stochastic point source method but to generalize it by formulating the mathematics in such a way to allow statistical treatment. This is accomplished by repeated application of the linearized particle equations of motion for each eddy interaction with successive substitution such that a single expression is obtained describing dispersion in terms of the turbulent velocity fluctuations in each eddy, parcel/eddy interaction times, and the inertial time lag for each interaction. By introduction of a characteristic turbulent velocity fluctuation, it then becomes possible to express the mean-squared dispersion as the variance of the PDF after the n^{th} eddy in terms of preceeding parcel/eddy interaction properties. With the mean and variance of each parcel PDF, it is possible to combine the PDF distributions to obtain the physical particle distribution profile.

After following through with the described mathematical formulation procedure, model evaluation was considered. Two simple test cases chosen were for nonevaporating particles injected into nearly-homogeneous grid-generated turbulence and into a round, turbulent air jet. In both cases particle loading was small and interphase momentum coupling was neglected. The new model was to be compared with the stochastic point source model which has been previously validated in the scientific literature. FORTRAN computer codes were developed for these purposes and computations carried out. Presentation and lengthy discussion of the results are in the accompanying AIAA paper.[3]

In summarizing these results, the new dispersion width transport model was found to have significantly improved computational efficiency over the stochastic point source method while producing good approximations for the physical particle spatial distribution profiles. Two schemes were considered for evaluating the mean parcel position: (a) Deterministic Dispersion Width Transport Model (DDWT) and (b) Stochastic Dispersion Width Transport Model (SDWT). For nearly-homogeneous turbulence, deterministic tracking with a single parcel (DDWT model) was found to sufficiently sample the turbulence properties and produce accurate dispersion results as shown in Figure 2. For the jet flow with nonhomogeneous turbulence, however, deterministic tracking was inadequate. More thorough sampling of the turbulence was required through stochastic tracking of the mean parcel position (SDWT model). Computational efficiency was still impressive, though, as dispersion results using no more than 200 parcels were found to produce a good approximation to dispersion results using 20,000 particles in the stochastic point source methodology. An example of such results is shown in Figure 3.

The potential impact of this new dispersion model on the accuracy and efficiency of spray combustion CFD codes, particularly with regard to instability analyses, is significant. Extension of the model to include evaporating droplets with interphase coupling of mass, momentum, and energy is obvious follow-up work for the future. It appears that additional adaptation for considering acoustic perturbations and improvements in efficiency are also possible, and model development is continuing.

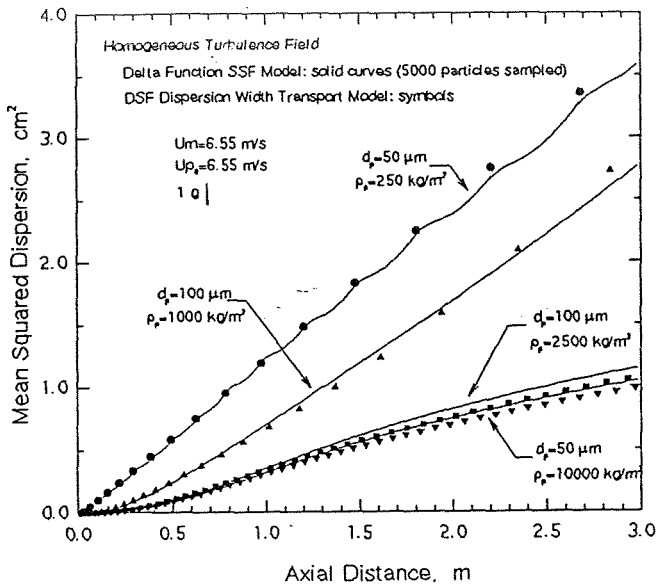


Figure 2. Proof-of-concept computations for the deterministic dispersion width transport model.

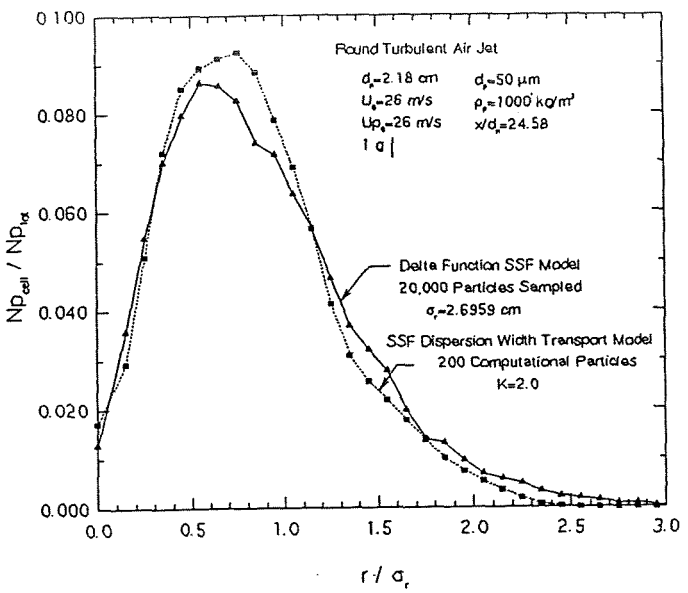


Figure 3. Proof-of-concept computations for the stochastic dispersion width transport model.

III. Project Direction

Research emphasis during the final phases of this project will be on development of the droplet dispersion model. The potential impact of this method on spray combustion modeling is sufficiently great to justify concentrated effort. Further development of the high-pressure droplet gasification model will require new experimental data which probably won't be available during the project lifetime. However, limited effort will be devoted to assessing the general influence of pressure on vaporization-controlled combustion instability using the current model.

Tentative plans for systematic development of the droplet dispersion model have been made. Work has already began on extension of the model to include evaporation effects.. Based on experimental work in the literature, evaluation of vaporizing monodispersed methanol droplets in a methane-fuelled jet diffusion flame is in progress. Eventually, interphase transport coupling will be addressed in an effective manner. Other plans include improving the efficiency of the procedure by limiting the number of interactions allowed to influence dispersion in a given eddy. Extension to recirculating flows and adaptation for acoustic perturbations are also under consideration at this time.

A simplified assessment of the high-pressure vaporization process on combustion instability is being undertaken concurrent with the droplet dispersion model development. The basic concept of this approach is to apply a simple analysis of the instability response factor using low-pressure and high-pressure vaporization models to quantify the effect of pressure on the feedback mechanism. Such an approach has practical application as a combustion response module for combustion instability design codes. The droplet vaporization code has been modified to simulate a simplified 1-D combustor with n-heptane droplets vaporizing in the stoichiometric products of combustion with oxygen. Sinusoidal oscillations characteristic of acoustic mode instability in a cylindrical chamber are superimposed on the combustor flow. The high-pressure thermodynamic model takes account of real-gas effects including H_2O molecules using an approximate engineering model. Absorption of ambient-gas into the droplet is neglected, however. The analysis is limited to subcritical vaporization since supercritical gasification models are dubious. A total of 36 droplets are injected per cycle of oscillation with the resulting vaporization histories combined to obtain the instability response factors based on the Rayleigh criteria. The response factors are to be evaluated for a given mean chamber pressure as a function of frequency. Comparison

of results for low-pressure and high-pressure vaporization models will yield a measure of importance of high-pressure vaporization effects on resonant combustion instability. The greater the response factor the greater the tendency to drive instability. Preliminary computations at moderate pressures indicate that consideration of high-pressure effects implies a substantial increase in the response factor. This increase is expected to be even larger at higher pressures.

IV. Concluding Remarks

This report has attempted to describe the accomplishments and current status of the Spray Combustion Stability Project as well as indicate the planned direction of future research. While research effort in this project has been primarily restricted to physical submodel development, it is desirable that these submodeling concepts be pursued to actual implementation into comprehensive CFD spray combustion codes. This will hopefully make full scale instability analyses more practicable and useful and improve spray combustor design technology.

ACKNOWLEDGMENTS

The Spray Combustion Stability Project is supported by the UT–Calspan Center for Advanced Space Propulsion, UTSI Research Park, Tullahoma, TN 37388 under NASA Grant NAGW–1195 and in part by Rocketdyne Division, Rockwell International, Canoga Park, CA 91303 with Dr. Pak-Yan Liang as technical monitor.

REFERENCES

1. Report Prepared by The Center for Advanced Space Propulsion for The Rocketdyne Division, Rockwell Int. Corp., *AN INTERIM RESEARCH PROGRESS REPORT: Concerning High-Pressure Droplet Gasification Submodel Refinement for ARICC-LJ Spray Combustion Code*, August 1989.
2. Litchford. R. J. and Jeng S-M, "LOX Vaporization in High-Pressure, Hydrogen-Rich Gas," AIAA Paper 90-2191, 26^{th} Joint Propulsion Conference, Orlando FL, July 1990.
3. Litchford. R. J. and Jeng S-M, "A Computationally Efficient Model for Turbulent Droplet Dispersion in Spray Combustion," AIAA Paper 90-2443, 26^{th} Joint Propulsion Conference, Orlando FL, July 1990.

SSME LEAK DETECTION FEASIBILITY INVESTIGATION BY UTILIZATION OF INFRARED SENSOR TECHNOLOGY

Ahmad A. Shohadaee and Roger A. Crawford
Center for Advanced Space Propulsion
The University of Tennessee Space Institute
Tullahoma, Tennessee

Abstract

This investigation examined the potential of using state-of-the-art infrared (IR) thermal imaging systems combined with computer, digital image processing and expert systems for Space Shuttle Main Engines (SSME) propellant path leak detection as an early warning system of imminent engine failure. A radiation transport model was used to evaluate and estimate the wavelength sensitivity of infrared sensors for detection of cryogenic and hot gas leaks. Laboratory experiments were devised and experimental approaches were established. The system was installed, checked out, and experiments were carried out using an IR thermal imaging systems combined with computer and digital image recording and processing system to validate the results of the model. The conclusion from this investigation is that both numerical and experimental results indicate that the leak detection by using infrared sensor technology proved to be feasible for a liquid rocket engine health monitoring system.

Nomenclature

a	absorption coefficient
I	source function, Eq. (2)
i	radiation intensity
K	extinction coefficient, $a + \sigma_s$
R	outer radius
r	radius
S	coordinate along path of radiation
T	temperature
Y	depth of plume at a cross section
k	optical thickness
λ	wavelength
σ_s	scattering coefficient
Φ	phase function for scattering
ω	solid angle

Subscripts

i	incident;incoming
p	plate
s	scattering
λ	spectrally dependent

Superscripts

$*$	dummy variable of integration
$'$	directional quantity

Introduction

The complicated high-pressure cycle of SSME propellant path provides many opportunities for external propellant path leaks while the engine is running. This mode of engine failure may be detected and analyzed with sufficient speed to save critical engine test hardware from destruction. The leaks indicate hardware failures which will damage or destroy an engine if undetected. Failure of high pressure lines, fittings, welds and internal component parts caused by thermal shock, thermal cyclic stress, stress corrosion, erosion, and high cycle fatigue can result in failure modes with sufficiently long time constants to allow detection and safe shut down. Mili-second detection rates (sampling rates) and 10 m-sec data processing can result in failure detection to shut down signal of less than 100 m-sec. Although the required time constants for detection and action have not been established from analysis of expected leaks, it appears that significant improvement in the engine protection system can be obtained with an advanced propellant path leak detection system. Optical sensors combined with computer, image analysis, and expert systems is one approach to address this problem.

Commercial infrared thermal imaging systems and associated digital image processing computers have been evaluated first because of their ability to sense both very cold cryogenic leaks as well as hot gas leaks.

Objectives and Approach

The primary objective of this investigation is the experimental validation of techniques for detecting and analyzing propellant path external leaks which have a high probability of occurring on the SSME. The selection of candidate detection methods requires a good analytic model for leak plumes which would develop from external leaks and an understanding of radiation transfer through the leak plume.

Gas species, temperature and plume shape have been used to characterize a leak plume. Due to the compact and complex propellant path configuration of the SSME, plume shape and resulting temperature distributions are difficult to predict. Significant plume impingement on other components results in solid sur-

face radiation as well as gas plume radiation. This investigation considers both cryogenic fuel and oxidizer leaks as well as hot gas leaks that may occur anywhere between the propellant inlet lines, turbomachinery, valves, preburner, etc., and the main thrust chamber.

It is recognized that detection of internal gas path anomalities is more difficult and most likely requires a different approach, such as the plume spectral analysis for metals. The combination of internal flow path performance measurements with external component temperature distributions may allow rapid diagnosis of internal flow problems. This investigation has not considered the main thrust chamber/nozzle plume spectral analysis. Observations have been restricted to the region above the nozzle throat to avoid the intense radiation field from the nozzle exit. Thus only those internal leaks which change the external temperature distribution on fluid lines or components upstream of the main thrust chamber would be detectable.

Selected leak detection sensors, primarily optical operating in the IR range have been evaluated on subscale laboratory plumes to determine sensitivity, signal to noise, and general suitability for the application. The actual detection system requires sensor integration into high data acquisition rate video to digital image conversion for processing. The current state-of-the-art infrared detectors and associated array processors are advancing at such a high rate that suitable sensor systems are commercially available. The application of expert system techniques will be required during image processing to evaluate and identify real leaks.

The objectives of this investigation have been categorized in three parts: first, developing a theoretical radiation transport model including emission, absorption and scattering. Second, developing a computational model to calculate radiation heat transfer in the participating media. Applying this numerical code to a simple leak model to evaluate infrared sensor sensitivity to engine and plume radiation patterns. Third, validate the model results with simple laboratory experiments.

I. Theoretical Radiation Transport Model

The first part of the investigation concentrated on developing theoretical models to predict radiation transport in absorbing, emitting and scattering media. These models have been used to evaluate the intensity of both background and plume radiation reaching a sensor location.

Theoretical study of the radiation heat transfer in an absorbing, emitting, and scattering medium is complicated due to the following characteristics of the medium. In a participating medium, absorption, emission and scattering of energy occur not only at the system boundaries, but also at all locations within the medium. A complete solution of the energy exchange problem requires the knowledge of temperature, radiation intensity, and physical properties at every point within the medium. The mathematics describing the radiation model is inherently complex, in particular, by the non homogeneity of temperature field. In addition, the highly non-gray spectral effects of gases are often much more pronounced than for solid surfaces, and a detailed spectrally dependent analysis is required. Most of the simplifications introduced in gas radiation problems are aimed at dealing with one or both of these complexities.

The transport equation describing the spectral intensity field is (from Ref. [1]):

$$\frac{di'}{dS} + i'_\lambda(k_\lambda) \equiv I'_\lambda(k_\lambda, w) \tag{1}$$

This equation together with boundary condition constitutes the basic mathematical problem to be solved for the description of intensity field. The source function $I'_\lambda(k_\lambda, w)$ is defined as:

$$I'_\lambda(k_\lambda, w) \equiv (1 - \Omega_{0\lambda}) i'_{\lambda b}(k_\lambda) + \frac{\Omega_{0\lambda}}{4\pi} \int_{wi=4\pi} i'_\lambda(k_\lambda, w_i)\Phi(\lambda, w, w_i)dw_i \tag{2}$$

This is the source of intensity along the optical path from both emission and incoming scattering. The albedo for scattering $\Omega_{o\lambda}$, defined as the ratio of the scattering coefficient, $\sigma_{s\lambda}$, to the extinction coefficient, $K_\lambda = \sigma_{s\lambda} + a_\lambda$,

$$\Omega_{0\lambda} \equiv \frac{\sigma_{s\lambda}}{K_\lambda} = \frac{\sigma_{s\lambda}}{\sigma_{s\lambda} + a_\lambda} \tag{3}$$

For scattering alone $\Omega_{0\lambda} \rightarrow 1$, while for absorption alone $\Omega_{0\lambda} \rightarrow 0$. The optical depth or opacity when both scattering and absorption are present is defined as:

$$k_\lambda(S) = \int_0^s k_\lambda(s^*)ds^* = \int_0^s (a_\lambda + \sigma_{s\lambda})ds^* \tag{4}$$

where s^* is a dummy variable of integration.
Equation (5) is the integrated form of the equation of transfer

$$i'_\lambda(k_\lambda) = i'_\lambda(0)\ exp\,(-k_\lambda) + \int_0^{k_\lambda} I'_\lambda(k^*_\lambda, w)\ exp\,[-(k_\lambda - k^*_\lambda]dk^*_\lambda \tag{5}$$

where k_λ^* is a dummy variable of integration.

Equation (5) is interpreted physically as the intensity being composed of two terms at optical depth k_λ. The first term is the attenuated incident radiation arriving at k_λ (including, however, the contribution of induced emission along the path). The second term is the intensity at k_λ resulting from spontaneous emission and incoming scattering in the S direction by all thickness elements along the path, reduced by exponential attenuation between each point of emission and incoming scattering k_λ^* and the location k_λ.

The equation of transfer in the form of (1) or (5) cannot be used by itself to obtain the local intensity because it contains the unknown temperature and scattering distributions that are in the source function [the temperature is in the $i_{\lambda b}(\lambda T)$ within I_λ']. The temperature distribution is also needed to determine the absorption and scattering coefficients so that the local optical depth $k_\lambda(s)$ can be computed from (4), and the physical coordinate S thereby related to the optical coordinate k_λ. The temperature distribution depends on conservation of energy within the medium, which in turn depends on the total absorbed radiation in each volume element along the path. This total energy will be obtained by utilizing the intensity passing through a location and integrating over all incident solid angles and all wavelength.

II. Numerical Radiation Model

Numerical results are very important since they are basis for confirming or disqualifying any attempt at an experimental model. The information from numerical results provides excellent guidance to select the best wavelength range for infrared detector, filters and equipment set up for generation of better experimental data. From observation of the numerical data and evaluation of the available information one can make a more realistic experimental model.

The numerical radiation transport model has been developed for a simple geometry containing a hot plate (background gray body), a simulated leak plume model in the form of cylindrical shape and a sensor location. Water vapor H_2O was selected as the molecule to concentrate detection on due to the water vapor present in the hot hydrogen propellant lines. The high absorption of water vapor in the infrared wavelength region may be exploited for plume detection. The models developed cover a temperature range of 300 to 2700° K at atmospheric pressure and wavelengths of 2 to 5 μm

The absorption and scattering coefficient is required for calculation of radiation heat transfer in participating medium. The absorption of radiation by the gas depends on the radiative properties of the gas medium. Analytical determination of the absorption coefficient requires quantum-mechanical calculations. Penner [2] presented a detail analytical study of quantum-mechanics to calculate the absorption coefficient.

In view of the progress that has been made in the field of radiation transfer there is still a crucial need for radiation properties such a spectral scattering coefficients. Most analyses are performed on a non-dimensional basis, this yields results which give important information with respect to general treads but does not reveal much information for numerical design values.

In this investigation absorption coefficients were taken from ref. [3] which is based on a combination of experimental data and analytical results. The radiation transport calculations have been done for a non-gray and non-isothermal gas (H_2O). A Gaussian temperature profile has been assumed for the plume temperature distribution. The radiation transport equation in the absorbing, emitting and scattering media with a non-gray and non-isothermal gas is much more complicated in cylindrical geometry than parallel plate case which can be found in literature. These models have been used to evaluate the intensity of background and leak plume radiation reaching the IR detector.

II.1 Discussion of Numerical Results

In this section consideration is given to numerical solution of the equation of transfer. Attention is limited to a plane source of radiation, incident on a cylindrical plume shape of water vapor as the absorbing, emitting, and scattering medium . The numerical models developed cover a temperature range of 300 to 2700°K at atmospheric pressure and wavelengths of 2 to 5 μm.

The schematic used in the radiation transport model for a simulated leak plume with a cylindrical geometry is shown in Figure 1. The plume cross section for numerical calculation of directional intensity is shown in Figure 2.

II.1.1. Background Radiation Source with Absorbing and Emitting Media

Figures 3 and 4 are typical numerical results calculated for variations of intensity at different Y locations. Figure 3 indicates that for a cold background plate with hot gas plumes, the intensity has its maximum value at $Y = 0$ and minimum value at $Y = R$. The intensity peak values are different for each wavelength due to the strong dependence of the absorption coefficients on wavelength. Figure 4 shows the

minimum intensity at $Y = 0$ and maximum value at $Y = R$, for the case of hot background plate and cold gas plume. Most of the incoming intensity will be absorbed by cold gas at the center line, where $Y = 0$ and S is maximum. At larger values of Y the travel path, S, becomes shorter resulting in a reduced absorption of the incoming energy.

The variation of intensity with wavelength is shown in Figures 5 and 6. For the case of hot plate with a cold plume, Figure 5 indicates that most of the incoming energy has been absorbed at wavelengths between 2.4 and 2.9 μ for different Y values, but for the case of cold plate with a hot gas plume, Figure 6 shows that most of the energy has been emitted by the hot gas. The results from Figures 5 and 6 suggest the choice of a band filter of 2.4 - 2.9μ for the IR camera. This indicates that at this wavelength range, most of the incoming energy has been absorbed or emitted by the gas plume.

Intensity variation as a function of wavelength for three different plume diameters is shown in Figure 7. This figure indicates that the absorption effect is significant even for very small plume sizes. The cold gas absorbs the incoming intensity from warm background, mostly in the wavelengths between 2.36 and 3.2μ and some in the wavelength range of 4.52 to 5.0 μ. The amount of absorption depends on the actual size of the plume which reflects the increase of the plume depth absorption.

The integrated directional intensity across the plume is defined by

$$i'(S) = \int_2^5 i'_\lambda(S) d\lambda \qquad (6)$$

which is equivalent to the area under the curves in Figure 7. The integrated results for $2.0 \leq \lambda(\mu) \leq 5.0$ is shown in Figure 8. The integrated intensities as a function of the actual distance across the plate geometry is plotted in this figure. The integrated intensities is related to the ambient, plate, and plume temperatures. The ambient integrated intensity is shown in the interval of $-15.0 \leq Y(cm) \leq -12.0$ and $15.0 \geq y(cm) \geq 12.0$. The observed value of plate integrated intensity is shown for $-12.0 \leq y(cm) \geq +12.0$; it decreases in the center, due to the presence of the absorbing plume. This decay of the integrated intensity is due to the combination of absorption and emission of the plume. The reduction of the integrated intensity for three different plume sizes is shown in Figures 8. The maximum absorption effect appears at the center of the plume (at $Y = 0$) which is directly related to the plume sizes. The value of the minimum of integrated intensity decreases with increased plume diameter. The leak size is directly proportional to the extend of reduction in the integrated intensity and this can be used as a calibration technique for the leak size.

Figure 9 show the reduction of the incoming integrated intensity from background radiation for a wavelength range of $2.5 \leq \lambda(\mu) \leq 2.86$. This wavelength range has been chosen with the aid of Figures 7 which indicates that most of the the background intensities will be absorbed by the plume in this range. The result from this figure prove that for this narrow wavelength band, the integrated intensity reduces much more than for the case of the wavelength band $2.0 \leq \lambda(\mu) \leq 5.0$ indicated by Figure 8. One should notice that even though the integrated intensity reduction is much more noticeable in the case of $2.5 \leq \lambda(\mu) \leq 2.86$ than for the case of $2.0 \leq \lambda(\mu) \leq 5.0$, the incoming intensity is also reduced dramatically. This can be seen by comparing the maximum values of the intensity on these figures.

These important results indicate that there is a distinct region of the spectra that can be utilized to monitor the major types of gas leaks. The region from 2.5 to 2.9 μ can be used to monitor leaks which are along the line-of-sight of the optical system and hence are blocking the radiation emitted by the hotter surfaces in the background. The significant reduction of the background radiation intensity in this region is a very sensitive warning signal for leaks of this type.

The effect of emission is more pronounced in hotter gases, since the emission of gases at low temperatures will be negligible compared to their absorption. To demonstrate this phenomenon, the following set of results are presented for the case of a hot gas plume against several different background radiation sources.

Figure 10 shows the effect of absorption and spontaneous emission for a fixed plume temperature and a fixed plume cross section. This figure indicate that the integrated incoming intensity increases mainly due to the emission from the hot gas. When the temperature of the background radiation source increases, the effect of the absorption will be more noticeable. The effect of the media absorption is shown in this figure which indicates that with the plate temperature above $700°K$, the absorption effect can be seen near the edges of the plume, since the plume has its lowest value ($300°K$) in that areas. Near the center of the plume, where the temperature has its maximum value, the emission of the gas becomes more significant than the absorption effect.

II.1.2 Background Radiation Source with Absorbing Emitting, and Scattering Media

Since all atmospheric cryogenic hydrogen leaks are surrounded by condensed water droplets on the outer edge of the plume which scatter the background

radiation, thus scattering effects has to be considered in this investigation.

Numerical results for radiation transport in absorbing, emitting, and isotropic scattering media will be presented in this section. The scattering coefficient is needed to calculate the radiation transport equation when scattering is included in the radiation calculation [1]. The scattering coefficients have been calculated with the help of the absorption coefficients and non-dimensional values of albedo, $\Omega_{0\lambda}$.

Equation (3) can be rearranged to give a equation for scattering coefficient which is

$$\sigma_{s\lambda} = \left[\frac{\Omega_{0\lambda}}{1-\Omega_{0\lambda}}\right] a_\lambda \qquad (7)$$

Since absorption coefficients are known at fixed temperatures and wavelength intervals, scattering coefficients can be evaluated at a fixed albedo for those temperatures and wavelengths. Figure 11 show the variation of absorption coefficient for albedos of 0.1 - 0.5 with plume temperature of 300^oK and wavelength interval of 2 to 5 μm

For isotropic scattering, the source function equation (2) is coupled with the local intensity in the media. This equation must be solved simultaneously with the intensity equation (5). Variation of intensity with wavelength is shown in Figures 12 which indicates that for low values of albedo and moderately warm background plates the scattering effect is not very noticeable. The effect of scattering for higher values of $\Omega_{0\lambda}$ is shown in this figure. Increasing the values of the $\Omega_{0\lambda}$ will expand the wavelength band which means incoming intensity will be reduced much more for the $\Omega_{0\lambda}$ =0.9 than for $\Omega_{0\lambda}$ =0.1. This effect can be understood because for higher values of albedo the extinction coefficients become greater and so does the optical thickness. The larger optical thickness means less energy can pass through the gas, thus the wavelength range will be larger for this case. In addition, it can be noticed that the emission at $\lambda = 2.7$ is much more pronounced for low values of albedo than high values of albedo which means that scattering effect is stronger than emission. It is important to notice that since the gas is fairly cold, most of the spectral intensities are absorbed and scattered in wavelength interval of 2.36 - 3.2 μ. In wavelength range of 2.0 to 2.36 μ and 3.56 to 4.4 μ, this figure shows that the intensity is totally independent of absorption, emission, and scattering of the media. This important result indicates that at these wavelength ranges, the presence of the plume will not affect the incoming intensity which means the integrated incoming intensity will pass through the plume without being affected. Hence the background radiation can be readily monitored independent of scattering or any other intervention by the plume media.

Figure 13 is the plot of the directional spectral intensity versus wavelengths at $Y = 0$. These intensity variations are for the case of hot gas at temperature of 1200^oK and are presented for three different fixed values of albedos of 0.0, 0.5, and 0.9. The albedo of 0.0 is for the case when no scattering is present and the participating gas media is effected by absorption and emission only. The extreme case of $\Omega_{0\lambda} = 0.9$ indicates that the media is predominantly a scattering media and albedo of 0.5 is the value between these two extreme cases. The effects of absorption, emission, and scattering are also shown in this figures. Figure 13 shows that for $\Omega_{0\lambda}$=0.0 absorption overcomes the emission in most of the wavelength band and for $\Omega_{0\lambda}$=0.5, the emission is being somewhat effected by combination of absorption and scattering of the media and for $\Omega_{0\lambda}$=0.9, the effect of scattering is strongly visible over the absorption and emission.

The effect of absorption, emission, and scattering of the gas media over the integrated intensity across the plume cross section at $(Y = 0)$ are shown in Figure 14. This figure has been plotted for a fixed radiation source temperature of 700^oK, gas temperature of 1200^oK and albedo of 0.2 and 0.9.

Figure 14 shows that for $\Omega_{0\lambda} = 0.2$ the emission from the gas dominates the scattering effect. Increasing $\Omega_{0\lambda}$ will decrease the peak value of the emission at the center of the plume. For ($\Omega_{0\lambda}$=0.9) shows that the peak value of the emission no longer exists. From a different point of view, it can be expressed that the effect of scattering for this case is similar to the effect of absorption with the difference that the incoming energy has been scattered rather than absorbed.

III. Experimental Analysis

The feasibility of the infrared detection scheme was evaluated on subscale simulated laboratory plumes to determine sensitivity, signal to noise ratio, and general suitability of IR imaging system for a liquid rocket engine propellant path leak detection. The experimental analysis was undertaken with the objective of testing simple models to differentiate the amount of radiation coming from a radiation source passing through an absorbing, emitting and scattering medium. The feasibility experiments were completed using the Thermovision® system to obtain images of both hot and cold plumes in front of a warm background plate.

The Thermovision 870 system operating in the 2-5 μm band was selected for the initial experiments. The

system can measure 10,000 point array at a rate of 25 times a second. The optical system with an extensive choice of lenses and filters enables measurement of a large range of temperatures -20°C to 1500°C over various distances.

III.1 Cold Plume in Front of Warm Background Plate

These experiments were accomplished with the cold plume in front of a warm background plate at several different plate temperature settings. Figure 15 shows a distinct cold plume against the warm background plate as obtained by the IR camera and enhanced by the VICOM digital image processing system.

The thermal images of the background radiation with and without the plume were stored on a VCR for later analyses. This image was enhanced using computerized image analyzer by subtracting the thermal images of the background radiation with the plume from the thermal images of the background radiation without the plume in order to enhance the difference.

The presence of cold plume reduced the incoming infrared radiation by absorbing the incoming thermal energy. This reduction of thermal energy by the plume is identified by the shaded area in Figure 15 and can be easily distinguished from the warm background radiation source.

Figure 16 shows the thermal image of warm background radiation source without the plume at temperature of 485 ^oK. The integrated intensity variation across the plate is also plotted in this figure. Figure 17 is the thermal image of warm background plate with the plume of cold water vapor in front of it. These images were stored on VCR for analyses. The integrated intensity across the image of the plate with the plume was accomplished using the computerized images by subtracting the image of the plate with the plume from the image of the plate without the plume, in order to enhance the differences.

These important results have confirmed the sensitivity of the IR system to low intensity signals.

III.2 Hot Plume in Front of Warm Background Plate

The hot plume for this experiment was created by burning pure hydrogen gas in air. The thermal image of the hydrogen flame in front of the warm background plate is shown in Figure 18. The emission energy from the flame is strongly noticeable in this figure. The emission level from the hydrogen flame was much more intense than the emission from warm background radiation source. The IR camera easily distinguished the different between the warm background and the hydrogen flame.

This result confirms the sensitivity of the IR system to high intensity signals.

III.3 Impingement Cooling

The complicated high pressure cycle of a rocket engine provides many opportunities for propellant path leaks while the engine is running. In such situations where a cryogenic fuel leak develops, the leak may impinge on local surfaces cooling parts of the engine which then result in sudden surface temperature changes in the areas around the leak. Therefore, a leak detection method can be proposed that is based on monitoring surface thermal radiation and its time-variations.

In order to verify the effects of impingement cooling on surface radiation for leak detection, a simple experiment was performed. This experiment was done by impinging cryogenic nitrogen on the hot plate. Figures 19 and 20 show the integrated intensity variation across the plate before and after cryogenic impingement on the surface of the hot plate. These figures show two important results caused by the impingement cooling. First, the temperature was reduced from 485 to 445 ^oK and second the integrated intensity dropped significantly in the area of impingement cooling. The impingement cooling effect can easily be noticed by the darker shaded area on the right hand side of the plate in Figure 20.

The result from this experiment confirms that a suitably selected infrared detector is capable of detecting the cryogenic leaks that cause impingement cooling which effects a sudden change of the surface temperature of areas around the leak.

III.4 Comparison of Numerical and Experimental Results

The numerical results were obtained with the assumption of the plate being a uniform source of radiation from side to side. Figures 21 and 22 are the comparison of the numerical and experimental results obtained at plate temperature of 485 ^oK for plume diameter of 5.0 *cm*.

The integrated intensities as a function of the the actual distance across the plate geometry are plotted in these figures. Figure 21 is for the case of no scattering and Figure 22 is for the case which scattering is present with albedo of 0.4. The plate integrated intensity is decreased in the center due to the presence of the plume. This decay of the integrated intensity is due to the absorption of the incoming intensity by the plume. The maximum absorption effect appears at

the center of the plume which is directly related to the plume size. The percentage of the integrated intensity reduction ($2 \leq \lambda(\mu) \leq 5$), from numerical results, is about 12% for Figure 21 and 14% in Figure 22.

The plate in the laboratory experiment was not a uniform source of radiation. Eventhough the radiation source is not uniform but the results can still be used to prove the sensitivity of the IR detector for general purpose of this investigation. The plate (see Figures 16 or 17) has two distinct uniform temperatures, a circular area in the center of the plate with a fairly uniform temperature and another uniform temperature for the rest of the plate.

The experimental results indicated that the integrated intensity variation has a profile similar to the numerical results where the integrated reduction appear at the center of the plate where the plume is present. The percentage of the integrated intensity reduction ($2 \leq \lambda(\mu) \leq 5$), is about 11 % compared to the 12 % and 14% of Figures 21 and 22. The differences in the percentage of the intensity reduction can be explained in the following manner. First, the plume diameter in experimental cases are not as uniform as the numerical model. Second, the concentration of the water vapor for experiment was a mixture of air and water vapor, but for numerical calculations were based on the assumption of pure water vapor. Third, the experimental background plate was not isothermal but the numerical calculations were based on uniform plate temperature. Finally, the numerical calculations were accomplished for the cases of no scattering and with scattering. The scattering effect in the experimental cases could not be measured directly, but based on these comparisons it shows that there is no scattering present in the laboratory simulations.

The overall comparison of the experimental and numerical show that the experimental results generally validated the numerical radiation transport model.

Conclusion and Recommendations

Optical sensors operating in the infrared wavelength range have been evaluated and determined to be strong candidates for use in a liquid rocket engine leak detection system. The proposed on-line advanced propellant path leak detection system would help to analyze intelligently the source and location of external leaks with sufficient speed to save critical engine test hardware from destruction by allowing safe shut down of the engine. The results of this research provide a basis for developing and employing an intelligent leak detection system in a real-life complex environment. An important step in developing such a system is the integration of IR sensor technology, digital imagine processing, and expert systems into a real time analysis system. It is recommended that additional laboratory simulations be conducted for analyzing and comparing with numerical results.

Acknowledgments

This work was performed with the Center for Advanced Space Propulsion (CASP) and was supported by NASA Grant NAG8-140. The CASP is part of the University of Tennessee Space Institute (UTSI).

References

1. R. Siegel and J.R. Howell, *Thermal Radiation Heat Transfer*, McGraw-Hill Book Company, 1981.
2. S.S. Panner, *Quantitative Molecular Spectroscopy and Gas Emissivities*, Addison-Wesley Publishing, Mass., 1959.
3. C.B. Ludwing, et. al, *Handbook of Infrared Radiation from Combustion Gases*. NASA SP-3080, 1973.

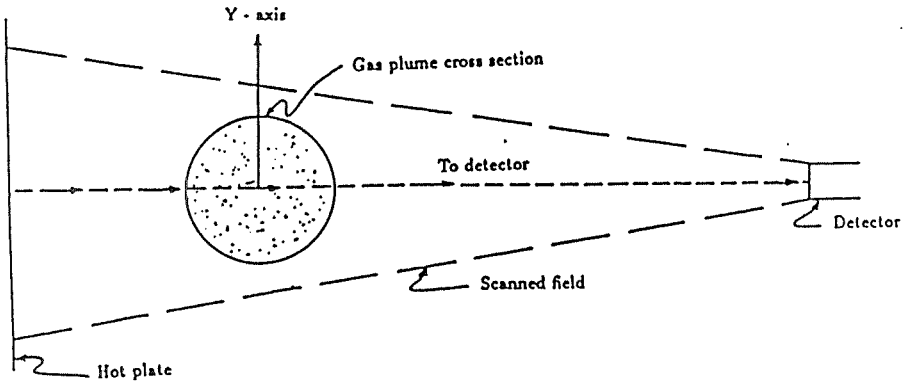


Figure 1 Geometrical arrangement.

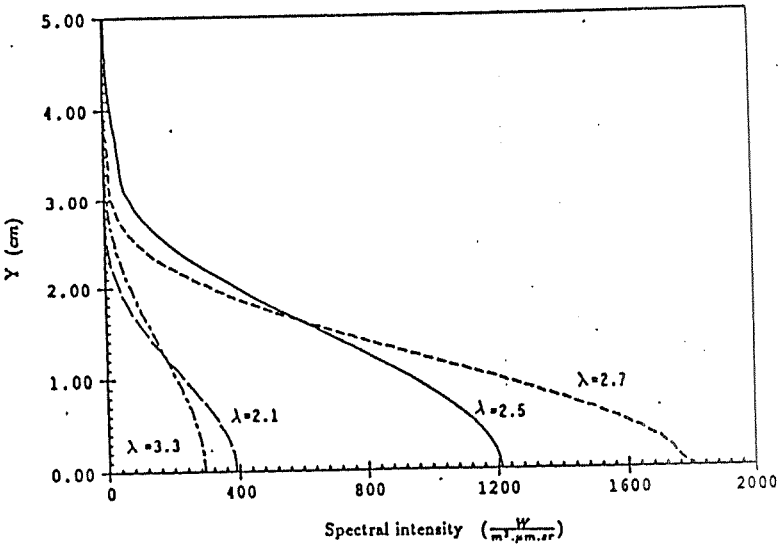


Figure 3 Spectral intensity vs. Y for cold background plate with hot gas plume.
At $T_p = 300°K$, $T_{max} = 1000°K$, $T_{min} = 300°K$.

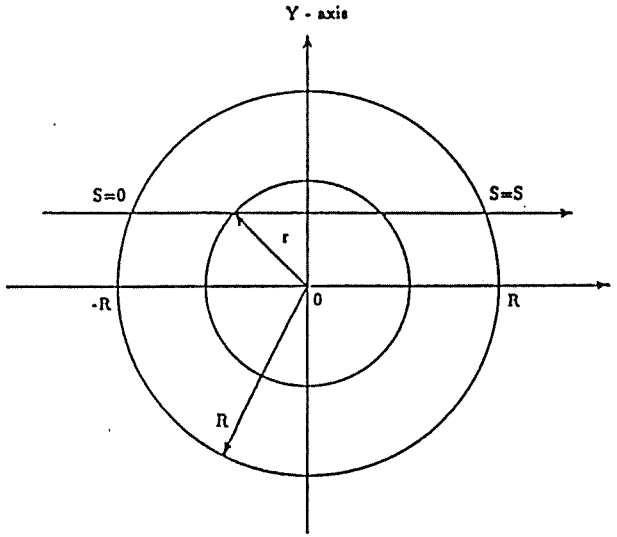


Figure 2 Plume cross section for numerical calculation

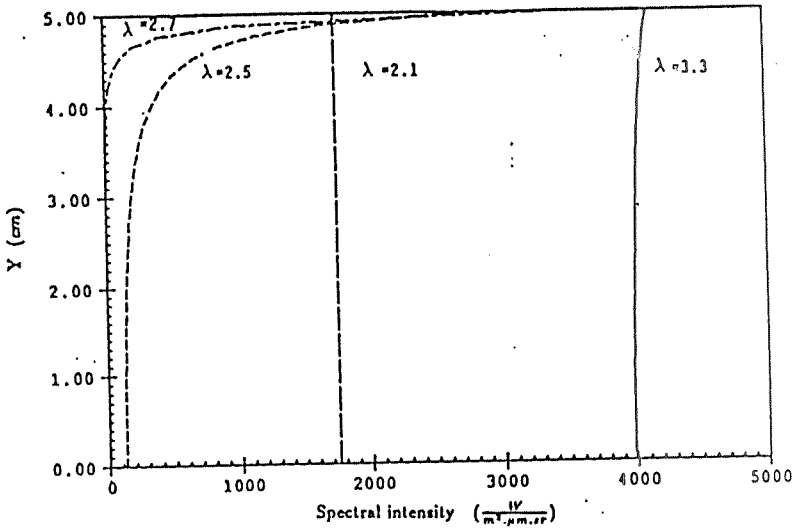


Figure 4 Spectral intensity vs. Y for hot background plate with cold gas plume.
At $T_p = 800°K$, $T_{max} = 400°K$, $T_{min} = 300°K$.

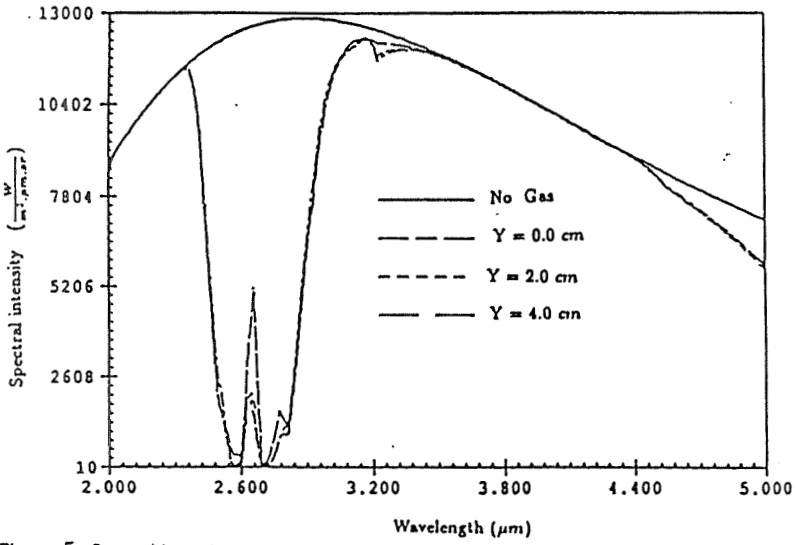


Figure 5 Spectral intensity vs. wavelength for hot background plate with cold gas plume at different Y
At $T_p = 1000°K$, $T_{max} = 600°K$, $T_{min} = 300°K$.

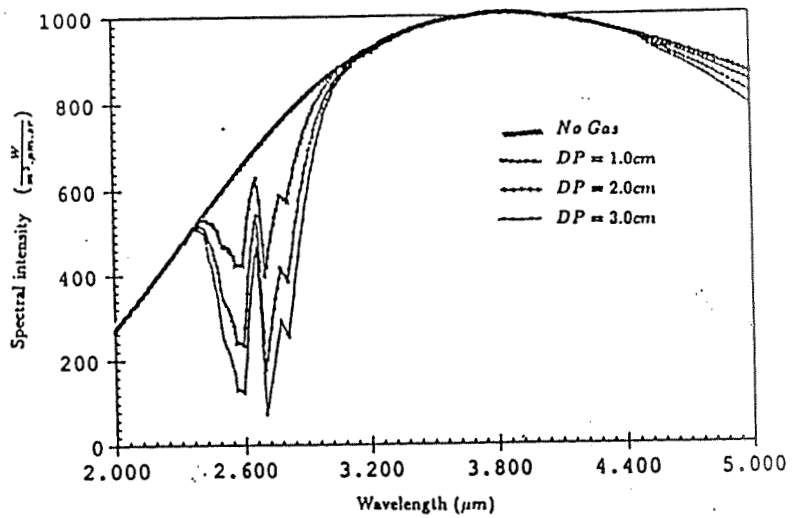


Figure 7 Spectral intensity vs. wavelength at different plume diameters.
At $T_p = 755°K$, $T_{max} = 373°K$, $T_{min} = 300°K$.

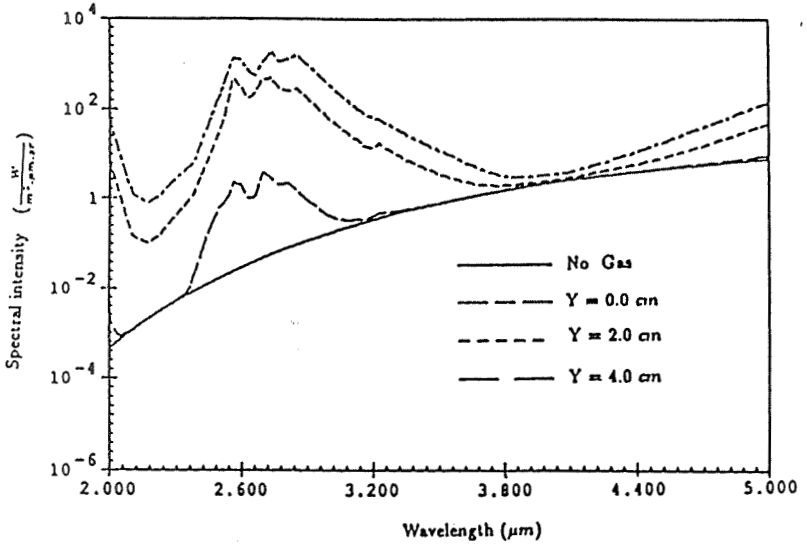


Figure 6 Spectral intensity vs. wavelength for cold background plate with hot gas plume at different Y location
At $T_p = 300°K$, $T_{max} = 1000°K$, $T_{min} = 300°K$.

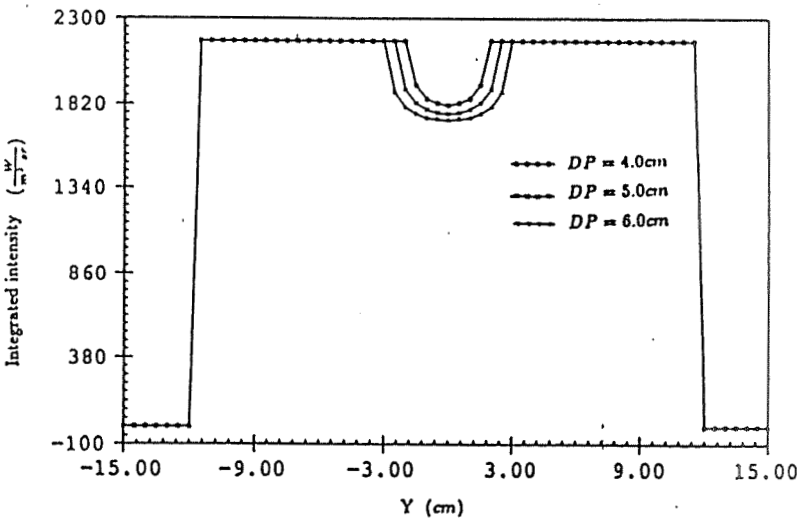


Figure 8 Integrated intensity vs. Y at different plume diameters.
At $T_p = 755°K$, $T_{max} = 373°K$, $T_{min} = 300°K$ and $2.0 \leq \lambda \leq 5.0$

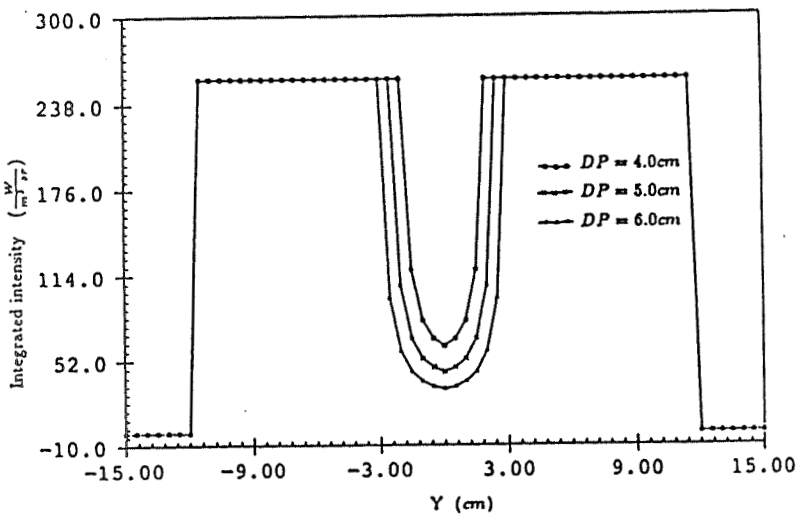


Figure 9 Integrated intensity vs. Y at different plume diameters.
At $T_p = 755°K$, $T_{max} = 373°K$, $T_{min} = 300°K$ and $2.5 \leq \lambda \leq 2.86$

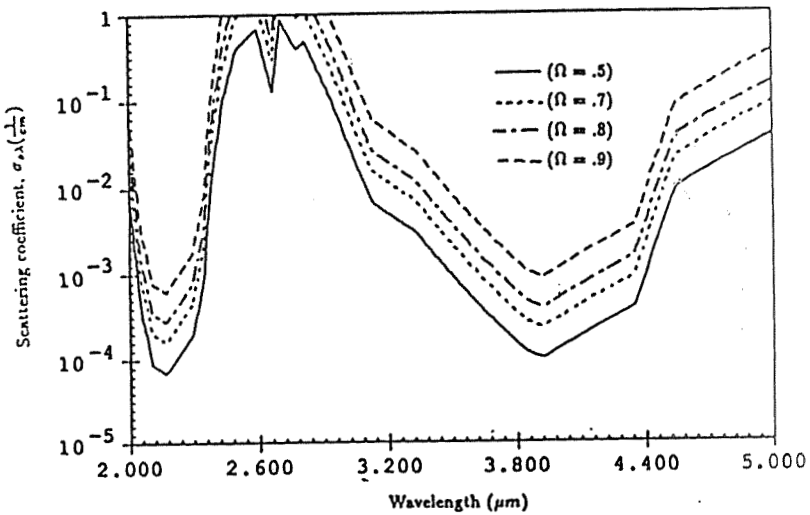


Figure 11 Scattering coefficients vs. wavelength at a fixed temperature of $300°K$.

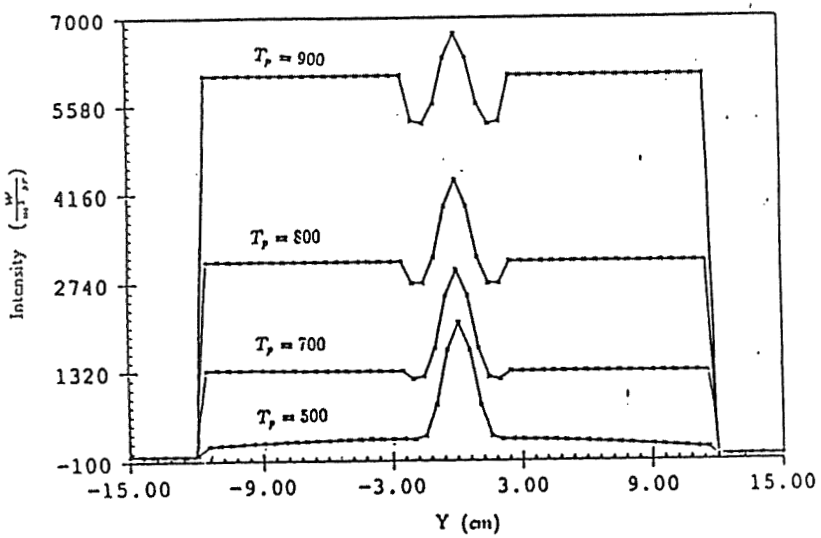


Figure 10 Total intensity vs. wavelength with a fixed plume diameter of 5.0 (cm)
At $T_{max} = 1200$, $T_{min} = 300°K$

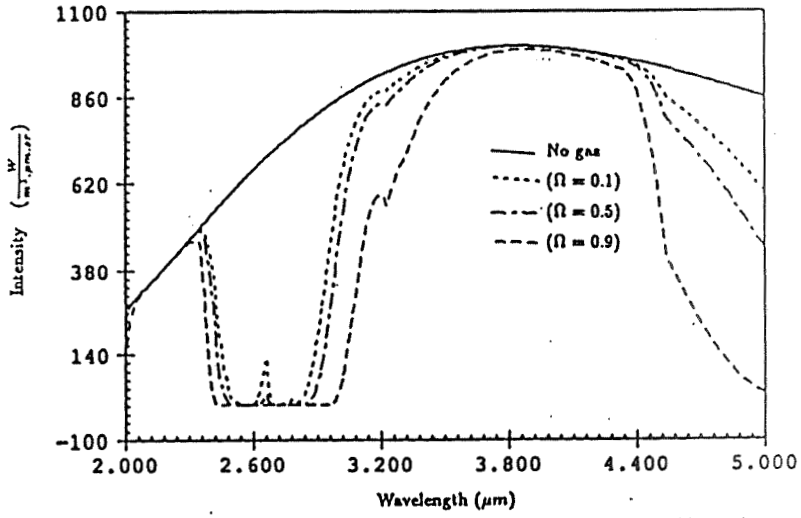


Figure 12 Intensity vs. wavelength for hot background plate with cold gas plume
At $T_p = 755°K$, $T_{max} = 373°K$, $T_{min} = 300°K$, and $DP = 10$(cm)

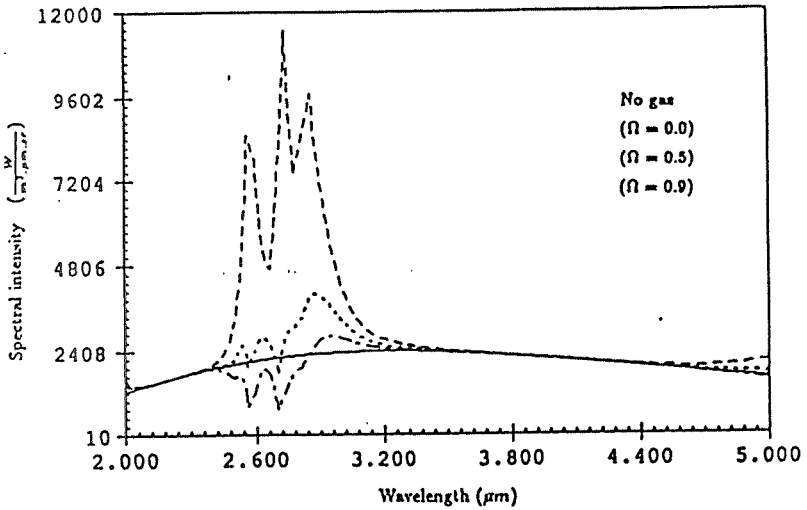


Figure 13 Spectral intensity vs. wavelength for hot background plate with hot gas plume
At $T_p = 900°K$, $T_{max} = 1200°K$, $T_{min} = 300°K$, and $DP = 5(cm)$

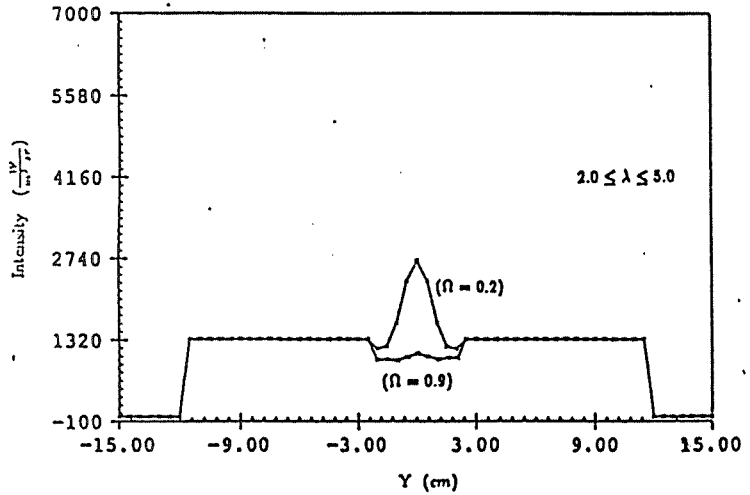


Figure 14 Total intensity vs. wavelength for a fixed albedo,
At $T_p = 700$, $T_{max} = 1200$, $T_{min} = 300°K$, and $DP = 5(cm)$

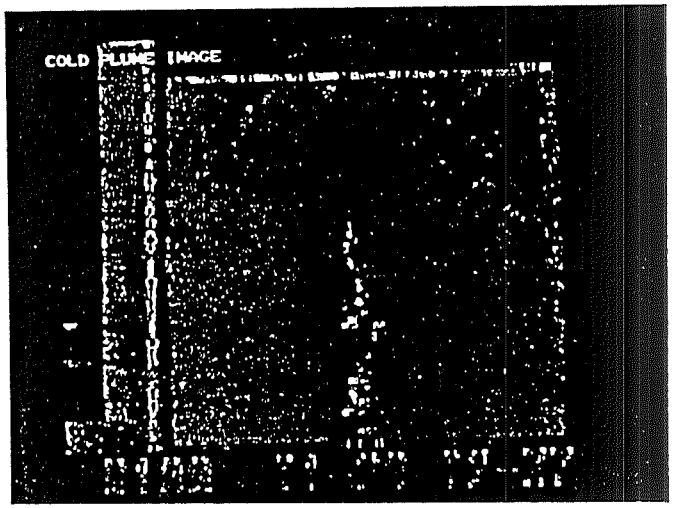


Figure 15 Thermal image of the cold plume in front of warm background radiation source

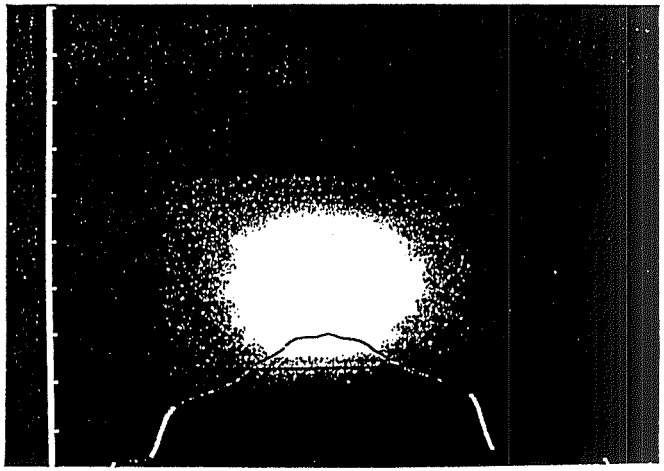


Figure 16 Thermal image of the warm background radiation source without the plume at $T_p = 485°K$

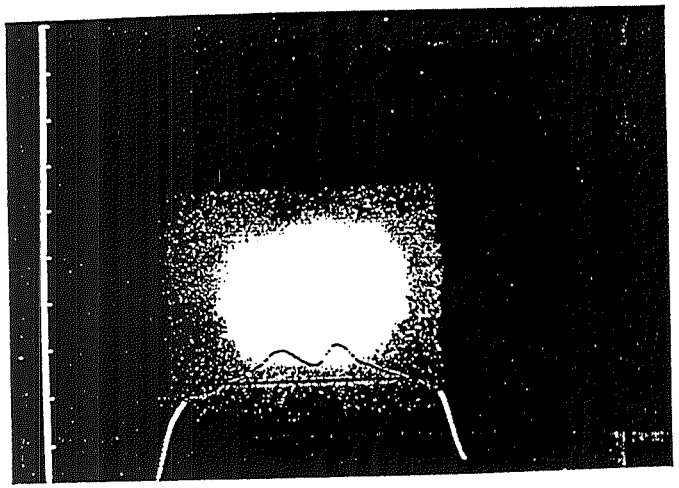


Figure 17 Thermal image of the warm background radiation source with the plume at $T_p = 485°K$

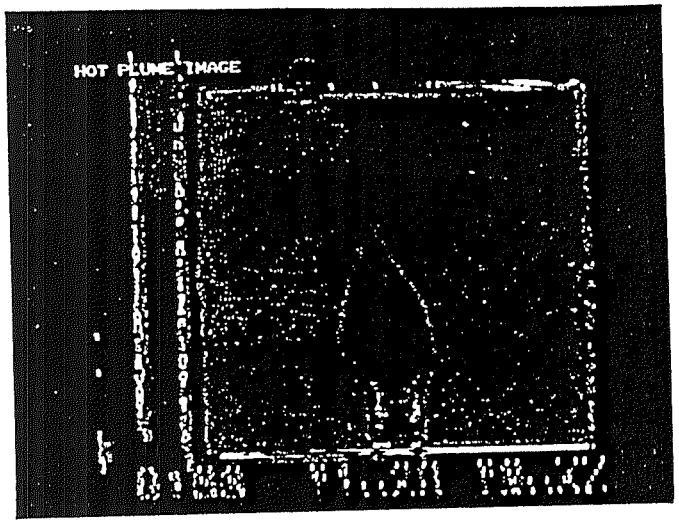


Figure 18 Thermal image of the hot plume in front of the warm background radiation source

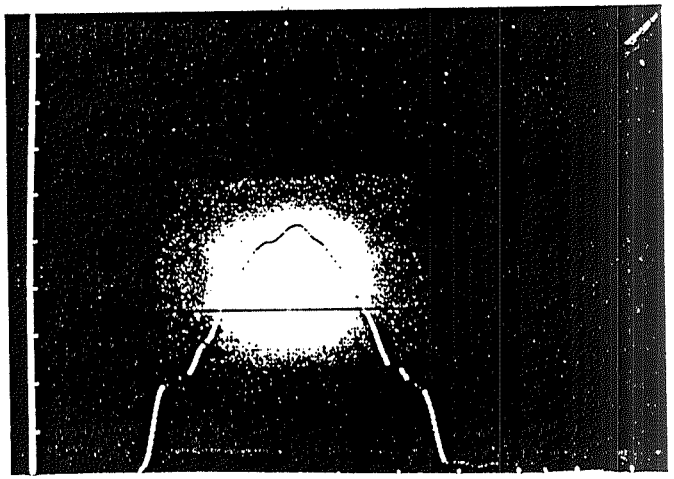


Figure 19 Thermal image of the warm background radiation source at $T_P = 485°K$

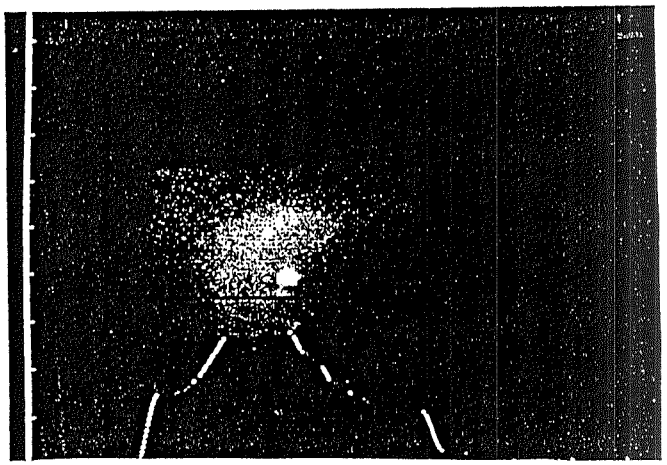


Figure 20 Thermal image of the impingement cooled plate at $T_P = 445°K$

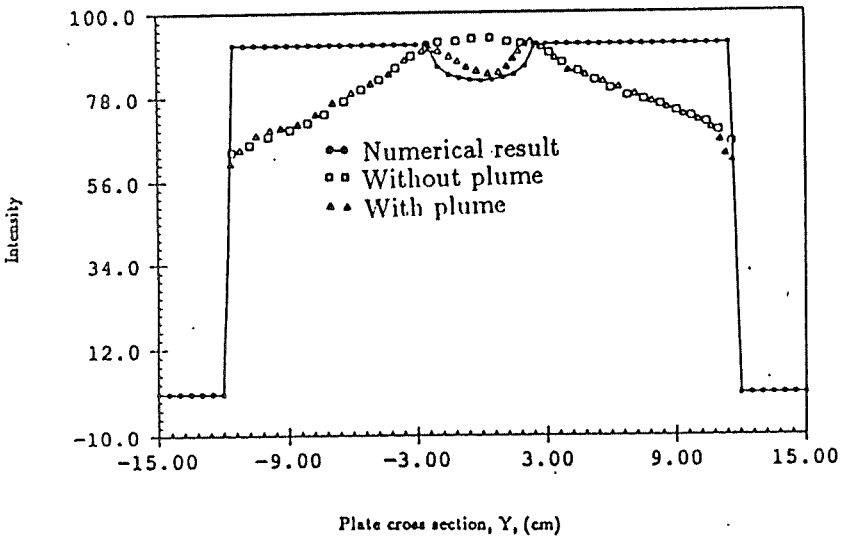


Figure 21 Numerical result of total intensity across the plate
($T_P = 485$, $T_{max} = 300$, $T_{min} = 300°K$, $DP = 5(cm)$, and $\Omega = 0.0$)

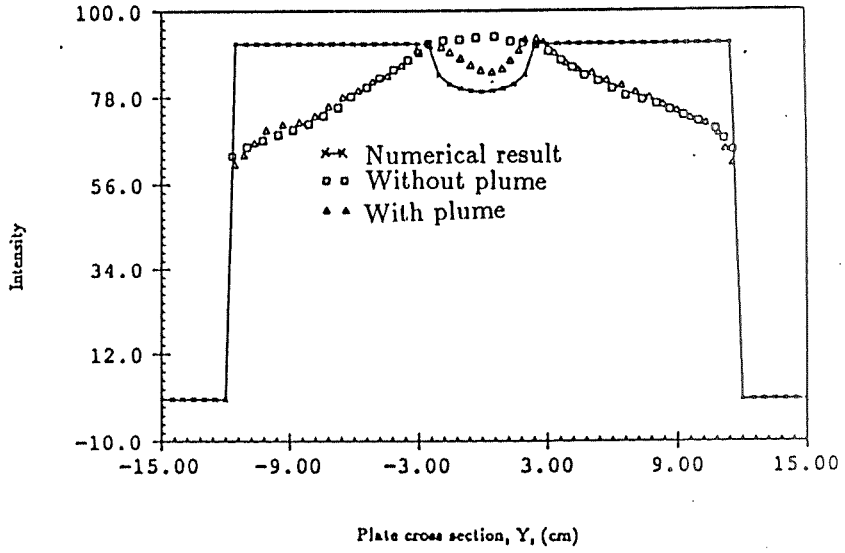


Figure 22 Numerical result of total intensity across the plate
($T_P = 485$, $T_{max} = 300$, $T_{min} = 300°K$, $DP = 5(cm)$, and $\Omega = 0.4$)

SPACE TRANSPORTATION

Development of the COMmercial Experiment Transporter (COMET)

Joseph F. Pawlick, Jr.

Development of the COMmercial Experiment Transporter (COMET)

By Joseph F. Pawlick, Jr.
Assistant Director for Commercial Space Transportation
Center for Advanced Space Propulsion
November 1, 1990

Overview

In order to commercialize space, this nation must develop a well defined path through which the Centers for the Commercial Development of Space (CCDS's) and their industrial partners and counterparts can exploit the advantages of space manufacturing and processing. Such a capability requires systems, a supporting infrastructure, and funding to become a viable component of this nation's economic strength.

This paper follows the development of the COMmercial Experiment Program (COMET) from inception to it's current position as the country's first space program dedicated to satisfying the needs of industry: an industry which must investigate the feasibility of space based processes, materials, and prototypes. With proposals now being evaluated, much of the COMET story is yet to be written, however concepts and events which led to it's current status and the plans for implementation may be presented.

Funding for this program was provided by NASA Headquarters Office of Commercial Programs through the Center for Advanced Space Propulsion Grant NAGW-1195.

This concept would not have come to fruition were it not for the support of Dr. George Garrison, Mr. E.G. Allee and those CCDS members who make up the COMET team. Invaluable assistance was provided by Messrs. N. Bowles and D. Lang of the Department of Transportation's Office of Commercial Space Transportation.

Program Initiation

At the May, 1989 Quarterly CCDS Directors Meeting, Mr. R. Ott, Director of Commercial Development Division of the NASA Office of Commercial Programs (Code C) presented the results of his office's investigation into available spaceflight opportunities for CCDS payloads. While the short range forecast showed a suitable number of flight opportunities, the quantity of commercial payloads for the outyears, were seen to outstrip Shuttle Get Away Special Canister (GAS Cans) and mid deck locker availability. Mr. James T. Rose, NASA Associate Director for Commercial Programs suggested that a "service" CCDS could be created which would investigate, optimize, and provide transportation opportunities to all CCDS's requiring spaceflight support.

While the Directors saw many benefits in such an approach, it was believed that the necessary skills existed within the present 16 CCDS's to accomplish this task. Mr. Rose asked the Directors to present their alternative approaches at the November, 1989 meeting at Houston, Texas.

Based on inputs from the Directors during the ensuing three months, Mr. Rose approved a structure for developing transportation and infrastructure in support of CCDS needs. Dr. Lundquist was nominated as the Chairman of the Commercial Space Services Procurements Subcommittee. Members added to the Subcommittee included Dr. G. Garrison, Director of the Center for Advanced Space Propulsion, Dr. R. Askew, Director of the Space Power Institute, Dr. A.D. Patton, Center for Space Power, Dr. M. Luttges, BioServe Technology, Dr. J. Wallace, Center for Materials for Space Structures, and Dr. A. Ignatiev, Space Vacuum Epitaxy Center.

During the next three months Drs. Garrison, Speer, Mr. E.G. Allee and the author, all of CASP, met with the ELV Subcommittee. These principles evolved:

1. Short programs are more successful than their long lived counterparts.

2. Procurement of systems isn't necessary if only services are desired.

3. The provider must be ultimately responsible for defining system quality and reliability.

4. Shuttle's man rating requirements place extreme demands on commercial experiments and prototypes.

5. The program must result in a U.S. based commercial infrastructure which offers total services to its customers.

6. Commercial approaches to procurement will result in lower costs.

7. CCDS's would be actively involved in the management of the program but not in it's technical decisions nor as contractor "monitors".

A survey of potential CCDS and commercial payloads was accomplished by CASP. Once the scope of physical attributes and experiment support needs were characterized, it became evident that an approach could be formulated which would satisfy the diverse needs of the CCDS's and their commercial partners. A concept of an ELV launched, commercially controlled vehicle which would be recovered on land within the continental United States evolved. If successful, a competitive procurement could yield, low cost service could be provided by U.S. industry. With a strawman approach identified, work areas, costing, and procurement strategies were developed.

At the February, 1990 CCDS Directors Quarterly Meeting, an overview of a three tier support approach was presented to CCDS and Code C personnel. Consisting of a progression of capabilities ranging from a "micro-bus", a system capable of supporting a 200# package of experiments to a 8000# "factory in space", the concept was based on currently designed systems which could be formed into a commercial venture to service the spaceflight needs of the CCDS experimenters. It was proposed that funding would be made available via augmentation to existing grants to fund service contracts. Each participating CCDS would be funded to accomplish it's technical monitoring and contract observation role.

An alternative was presented by Dr. M. Luttges of BioServe, University of Colorado, Boulder. The BioServe proposal consisted of a series of flights which would take advantage of the remaining Discoverer recovery capsules and be launched from a Pegasus ELV. BioServe's selected contractors would procure the hardware and accomplish the first mission in 1991.

Dr. Lundquist's committee was asked to review all aspects of the two plans and present a consolidated approach to Code C for their consideration. Since the BioServe variant provided the same capability as the micro-bus already considered by the Subcommittee, the microbus was eliminated and a plan was prepared which included BioServe's Commercial Interim Recovery Capsule (CIRC), and a mid-range system termed COMET. The high end, factory in space variant was eliminated since emphasis was placed on solving the near term problem of insufficient flight opportunities for CCDS payloads.

Since the two remaining systems would be competing for limited funding in the earlier years, it was decided to proceed with planning for a system which would place at least 6 cubic ft. of CCDS payloads into a 250 to 350 nm. equatorial orbit. Other non-recoverable experiments would be housed in the payload adapter and termed the "Service Module". All aspects of the procurement would be completed by the CCDS's and, if required, NASA would have the opportunity to review the final COMET selections during the subcontract review which would be completed by the Grants Officer prior to contract signature. This consolidated plan was presented to Mr. J. Rose on May 20, 1990. Four days later Mr. Rose notified the COMET Program Manager of his decision to pursue the COMET plan.

Less than two months later, a draft RFP was released, followed two weeks later by a discussion with industry. Less than Three months from it's activation, the COMET Team put out an RFP for all aspects of the program. On October 3, 1990, fourteen prime contractors placed proposals with CASP for COMET's various work areas. Proposal evaluation is being accomplished by the six CCDS's who make up the COMET Management

team. A goal of July, 1992 has been set for first flight.

Market Analysis

In a 1989 study commissioned by NASA's Office of Commercial Programs, SAIC Inc. surveyed a number of commercial experimenters to determine the size, type, and requirements of payloads they might wish to put into space. Figures 1 and 2 are extracted from that study and show the weight distribution and number of flights for payloads which could fly between 1989 and 2000. These two figures portray parameters which have great impact on system development: weight to orbit and total potential markets.

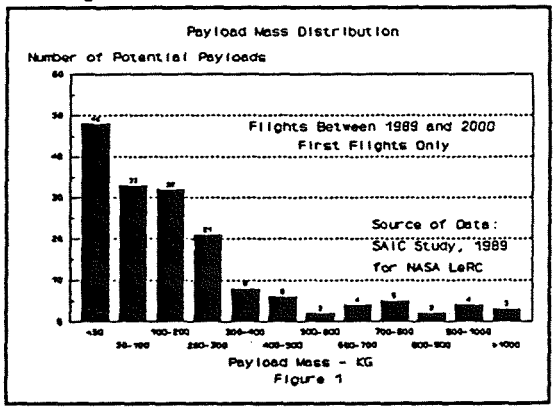


Figure 1

Figure 1 indicates that the largest number of commercial payloads will weigh less than 50 kilograms and that a system capable of orbiting 300 kg. of experiments will be able to service almost 85% of the identified payloads.

Figure 2 shows that a large number of payloads are not single flight experiments. Rather, many are to be refined through successive flights as the decade progresses. Note that the data exhibits a bow wave pattern which seems typical of a schedule for projected systems of any type but in particular, spacecraft and boosters. Of equal importance to the total numbers identified is the indication that some experiments are ready to fly now while many will become ready for flight when flight opportunities are available. SAIC study data further noted that one third of all experiments require recovery. These experiments cannot achieve meaningful results without the final product in hand, id. crystals from a macromolecular crystal growth experiment, layered substrate from epitaxial growth,

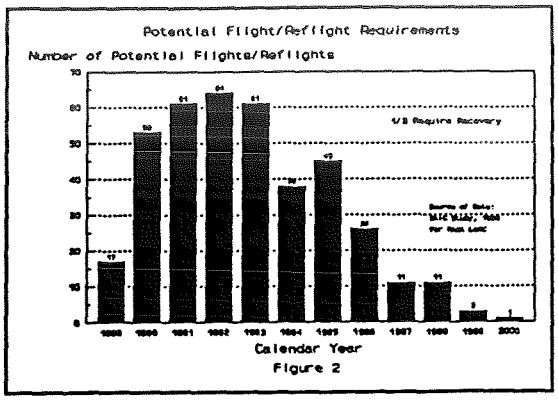


Figure 2

and finished product from materials processing work. Others, not requiring recovery may be on ELV's to satisfy their primary objectives.

While the SAIC statistical sample included CCDS payloads, it was difficult to determine the statistical presence of the CCDS's in the data. Additionally, it was believed that some CCDS's might respond more positively if they were given the opportunity to participate in structuring a commercial space program. CASP initiated a follow-up survey shortly before Christmas, 1989. Results of this survey identified needs for power, thermal, micro "g", and other salient experiment parameters and provided weight and physical dimensions.

Figure 3 superimposes the results of the CASP survey on the weight profile generated by SAIC. Each (1) represents a CCDS payload whose data was taken from this survey and clearly shows that the CCDS requirements are a subset of the larger data base. Given experiment maturation, all portions of the SAIC curve could reasonably be expected to be reproduced by CCDS sponsored payloads.

For example, payloads at the right of Figure 3 represent the University of Houston Epitaxial Growth and Clarkson Crystal Growth experiments. These are most like experiments which can evolve into a commercial "factory in space". In one sense, Figure 3 can be segregated into three major categories of payloads based on their weight and objectives: to the left are those

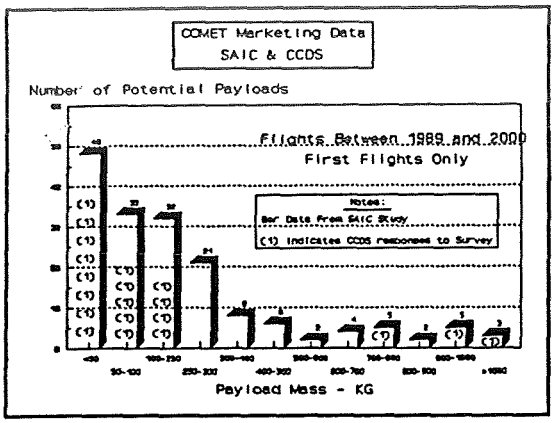


Figure 3

involving commercial experimentation, to the center are those who's goals are to refine the earlier experiments into commercial prototypes, those on the right are directed at commercial production.

It appears that the needs of those commercial payloads which reside on the left of the weight scale might be met with current, "off the shelf" technology. Those in the central area can be satisfied by systems already designed but not flown (within COMET capability) and those to the far right require systems which require Shuttle support or ELV systems which are beyond the weight capability envisioned for COMET. Commercialization will ultimately involve all three segments but funding availability will focus our attention on the first two.

While systems must be capable of supporting commercial spaceflight needs, an industrial capacity in space requires that an infrastructure be developed: one which facilitates the process of getting experiments and prototypes into space, gives investigators latitude to optimize the materials and processes of their experiments, allows them to influence mission planning, interact with their orbiting experiments, gather data when needed, and finally, have access to their experimental apparatus shortly after its return.

Infrastructure must be supported by systems which fulfill the needs of the commercially oriented experiments. Such systems must accommodate the spectrum of commercial payloads at a low cost and acceptable reliability. Reliability and quality need not be dictated since these are, rightfully, requirements of the supplier, not of the purchaser. We expect to bring forth new approaches to support of space systems. Developing an infrastructure along the lines of current aerospace models would force us to repeat many of the less efficient aspects of prior programs: those in which regulation, intricate cost accountability systems and numerous quality checkpoints drown out technical excellence and professionalism.

Since COMET will purchase the services of suppliers rather than their hardware or components, interactions with the supplier will be restrained. In contrast to an aerospace program under government contract, this more commercial approach dictates that providers will insure themselves against failure. In support of this tact, the COMET team's approach will be to review contractor progress to the extent necessary to assure the COMET team of the supplier's ability to meet major program milestones. In depth safety, quality, reliability, and maintainability reviews and procedures can give way to "best commercial practice".

Finally, if the program is to make an impact on the established way of doing business, there must be opportunities for new, emerging, "fledgling" companies to demonstrate their technical prowess.

In summary, a coherent approach to servicing commercial development requires a program, not a project. To be of continuing value, the program must be capable of growth: a growth dictated by products which are derived from experiments in space and lead to production in space. Our continuing objective must be to develop an infrastructure which outlasts current needs, even current vision, and allows this country to move steadily and surely toward a fully commercial approach to accessing space.

Mission Concept

Supporting the diverse elements of space commercialization brings with it a need to serve a wide spectrum of users. Low weight, low power experiments may be all that's required to serve a crystal growth investigation while commercial prototyping of a materials processing technique may greatly increase the levels of power and thermal systems need. To adequately support the broad

spectrum of requirements and build on successes, hardware capability must expand to satisfy increasingly larger and more complex payloads. However, regardless of the size of payload, the concept of a CCDS monitored, contractor operated, launch and recovery program provides the most viable method for creating an infrastructure.

To avoid unnecessary program complexity, COMET will make use of expendable launch vehicles (ELV's) rather than Shuttle. The talents of as many as five hardware and one systems engineering contractor will be combined to launch, command, control, re-enter and recover commercial payloads. CCDS members will structure statements of work and evaluation criteria, select the most responsive proposals, monitor contract performance and observe technical compliance.

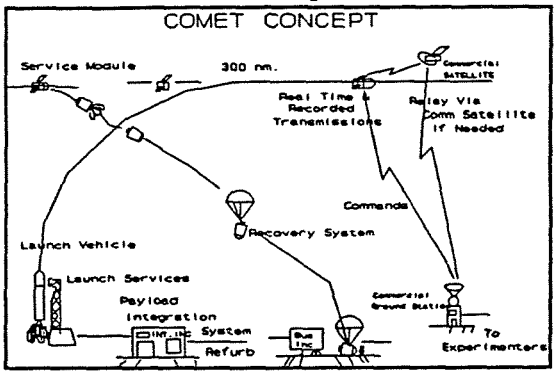


Figure 4

As shown in Figure 4, experiments will be moved directly from investigators to the Payload Integration contractor's facility. They will be matched with other compatible experiments, mated to a payload structure and integrated into a common command and data system. The experiments will be checked out and the payload integration contractor will insure that the built-up payload assembly meets the necessary interfaces with the FreeFlyer.

A checked out payload assembly will be transported to the Recovery System contractor location where the Recovery System (R/S) contractor will mate the payload to the R/S. After subsystem operability is confirmed, the Service Module will be mated to the R/S. The Systems Engineering contractor will assure proper performance of the integrated R/S, Payload, and Service Module checkout. Upon successful completion of these tests, the FreeFlyer will be moved to the launch site where the Launch Vehicle contractor will mate the FreeFlyer to the launch vehicle. Provisions will be made wherein experimenters will have access to their packages up to three hours before launch.

A successful launch will leave the FreeFlyer in an orbit of approximately 300 nm. and at a nominal inclination of 40°. Either of these orbital parameters will be modified to match experiment needs, however, land recovery at a suitable southeast U.S. locations dictates an inclination of at least 38°.

Once in orbit, the Orbital Operations contractor will activate FreeFlyer systems and command the FreeFlyer as required by the experimenters. Data recorders, real time telemetry, and experiment control will be commanded by the Orbital Operations team. Automated power, thermal, and attitude control systems will be provided by the Recovery System and Service Module and will be monitored by the orbital operations team along with experiment health and status.

Upon successful completion of mission objectives (a minimum of 30 days), the R/S contractor will calculate the necessary commands which must be sent to the FreeFlyer to allow the R/S and it's payloads to re-enter and land at the designated landing site. This information will be provided to the Orbital Operations team who, after assuring that payloads are configured for re-entry, will send the appropriate command sequence. After a successful re-entry, the R/S will be maneuvered for a soft landing at a location within the U.S.. Ram Air and other controllable parachute systems have been suggested as suitable devices for allowing steering corrections at the terminal portion of the flight and for controlling landing loads.

If the FreeFlyer has been equipped with experiments in the support module, the orbiting systems will continue to provide data until the systems are de-activated, the experiment goals have been accomplished, or the Service Module re-enters. Minimum life of the Service Module is to be 100 days.

COMET's payloads will be removed either at the recovery site or support area by the R/S contractor. Similarly, the Payload Integration contractor will remove the individual experiments at

either the recovery or support areas as requested by the investigator. Both the Recovery System and payload support structure will be returned to the appropriate location for refurbishment and return to the inventory for future flights.

A Systems Engineering contractor will address overall reliability, survivability, and compatibility questions and will assure the technical acceptability of experiments and system changes. This organization will also oversee or accomplish integrated testing of COMET components.

System Description

Since COMET's design will be determined by those selected to provide systems and infrastructure, Figure 5 is only a conceptual overview of major hardware components. COMET's FreeFlyer consists of two functional components: A Recovery System and a Service Module. In basic terms, the Service Module is a satellite, complete with power systems, thermal and attitude control, and experiment sensors. While attached to the Recovery System, the Service Module provides power, telemetry, command and telemetry recorders, and attitude control for its own systems and payloads and those located in the Recovery System.

While many payloads could be characterized in terms of their weight, CCDS Directors have found it more appropriate to express their payload needs in terms of payload volume.

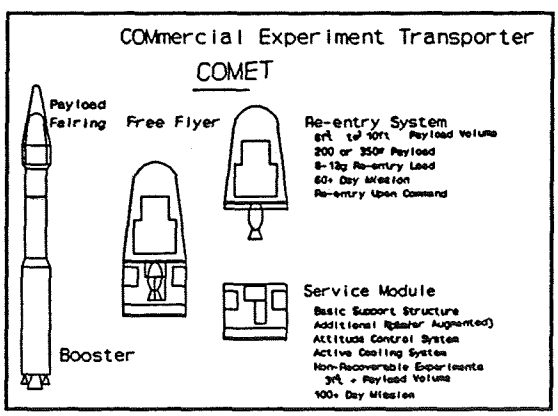


Figure 5

Based on the CASP survey and another, more in-depth view by the CCDS at Battelle, the minimum viable payload capacity is considered to be 6 ft^3 with a strong desire to increase to 10 ft^3 if possible. Note that the "advertised" R/S weight and payload capability are based on an estimate of payload density. While R/S manufacturers quote a density of 80 to 100#/ft^3, we believe a figure of 30 to 50#/ft^3 is appropriate based on experiences of the Consortium for Materials Development in Space (CMDS) in its Consort and Joust programs.

Candidate launch vehicles appear to be in the class of Orbital Sciences' "Taurus", Space Service's uprated "Conestoga" and American Rocket's "AMROC". Each have been designed but not yet flown. Other small launch vehicle manufacturers are expected to propose similarly capable systems.

Likewise, Recovery Systems for the COMET have been designed but not yet flown. Cheops, a recovery system marketed by COR Aerospace, an offspring of Lifesat, (currently entering phase B design), and an uprated Discoverer, are likely designs for COMET's Recovery System. While Lifesat is seen as a candidate, the commercial version must be much less sophisticated and considerably less costly.

Addition of a Service Module on COMET greatly enhances the capability of the FreeFlyer by eliminating R/S weight and housing systems capable of providing virtually an unlimited time in space prior to reentry and recovery. This Service Module design philosophy is similar to that of BIOSAT and Gemini. In these vehicles, the support, or service module, housed major systems: communications, power and conditioning, thermal control, and data/command storage devices. Like BIOSAT and Gemini, COMET will be designed to allow late access to the payload area through access techniques which will be proposed in the contract competition.

Schedule

Contacts with industry indicate that contracts for COMET class system could lead to launch within 18 months after contract go ahead. Potential Recovery System suppliers suggest that their vehicles could also launch within 18 months of Authorization to Proceed (ATP). There is a distinct possibility that systems already in production for micro-sats or other on-going satellite programs could offer a shorter lead time for orbiting systems. However, based on discussions with potential

suppliers, many are burdened with a 12 to 16 month cycle for long lead items.

While little testing will be required by our approach, adequate schedule time must be allocated by the hardware contractors to allow both the contractor and COMET Management Team to gain confidence in the system's capability to be successful.

Management Plan For COMET

Management of the program will take advantage of the depth of experience and talent within the CCDS team. A matrixed organization (Figure 6) will be headed by a Program Manager, assisted by a Systems Engineer, and will include a small permanent program management staff. Initial staffing within the Program Management Office includes permanently assigned personnel plus those with specialized talents to assist in the start-up of

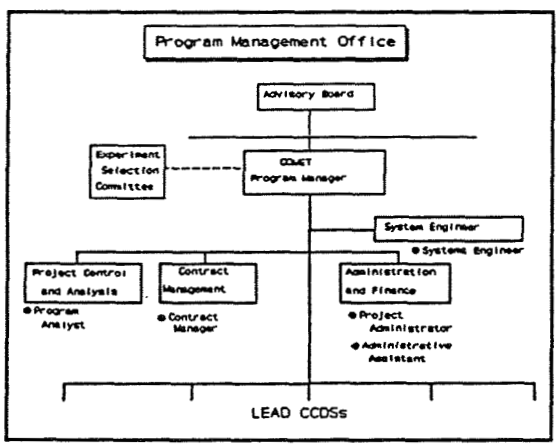


Figure 6

the program. In most cases these will be detailed from both the University of Tennessee, Arvin Industries and/or Calspan organizations. Legal assistance, for example, is not the product of one individual's time, but rather a composite of legally talented people who will assist in contract preparation and proposal evaluation. Experiences on other multi-contractor endeavors indicates that while lean, the staffing level is deemed adequate for efforts approaching $50M per year provided that government contract monitoring approaches are abandoned.

Through their intimate knowledge of the subject area, CCDS's who wish to make a significant contribution to the success of COMET were invited to involve their personnel in a contract observation and technical monitoring role.

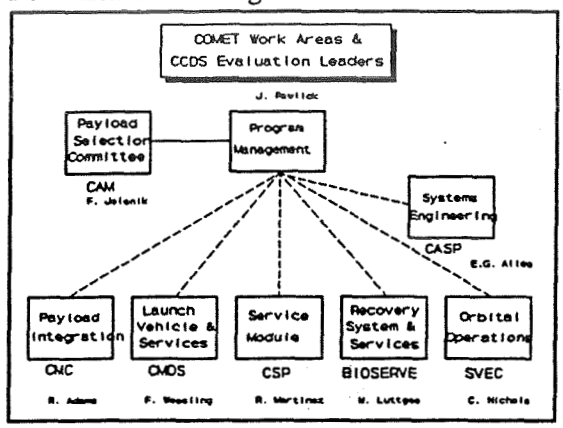


Figure 7

For each of the six major contracts let under the proposed COMET Program, a CCDS was chosen to prepare statements of work, evaluate proposals and provide a technical and contract assessments within the selected work area. These CCDS's, termed "lead CCDS's" were picked from the ranks of interested and experienced CCDS's and are the primary program interface with the contractor. Figure 7 shows CCDS's, which have been selected to make up the COMET Management team.

As a member of the Program Management Organization, the CCDS Monitor at the Lead CCDS's provide a voice within the technical area for formulating plans, defining changes, evaluating status, and determining the future of the COMET Program. As the Program member in closest communication with the contractor, the CCDS Monitor is in the best position to provide assessments of mission preparedness and contractor performance. With the addition of Lead CCDS's and their CCDS Monitors, the management structure becomes one in which CCDS members are directly responsible for the success of the program.

Lead CCDS's will be funded through an augmentation to their respective grants based on a budget provided to OCP by the Program Manager. OCP requires that each CCDS provide budgetary backup to their "bogey" similar to that currently used in requesting augmentation.

An Advisory Council will be formed by the CCDS Directors. Headed by a CCDS Director, the Council will be comprised of other CCDS Directors, and may include members from Code C, the Commercial Program Advisory Council, the Department of Transportation and the Department of Commerce. They will provide overall program assessment and guidance, assist in development of new concepts or areas of interest, and accomplish high level interfaces with Governmental and commercial organizations. Meeting semiannually, the Advisory Council will receive presentations from the Program Manager and Systems Engineer who will present program status and provide an insight into potential opportunities and problem areas.

An Experiment Selection Committee, headed by a CCDS Director, will consist of CCDS Directors representing Material Sciences, Life Sciences, Earth Sciences, and one or more from the "infrastructure" CCDS's (power, propulsion and robotics). They will be joined by members from NASA and industry and together will review proposed experiments to assure their compatibility with the goals and objectives of the commercial space initiative, assess the ability of the experiment to interface with other packages, and assure that the experiment has a reasonable probability of success when flown aboard the COMET. Based on the state of hardware readiness, they will recommend a window for flight and provide their recommendations to the Program Management Office for technical review and scheduling.

As currently formulated, there is no requirement for NASA to take an active role in management of the COMET program, instead they will review final subcontracts and control the flow of systems funding through a well structured and existing grant augmentation process.

Planning and Control

We recognize that planning is not a static process. In order to take advantage of unforeseen opportunities, a process of reviewing and evaluating the effect of proposed changes will be in place and is depicted in Figure 8. Since the need for change may take many paths, an element within the Program Management Office will be the focal point for planning and changes to existing plans. Certainly, the overall cost and schedule effect of changes due to events such as a member CCDS identifying a hardware anomaly, a new concept for

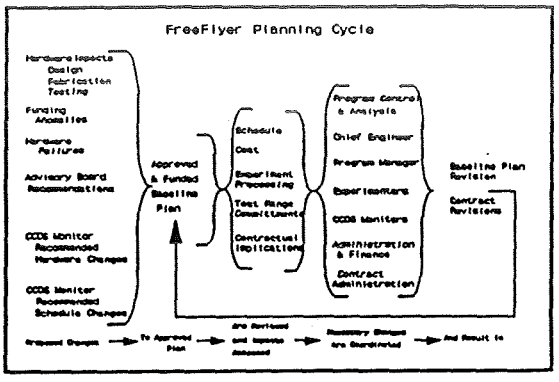


Figure 8

operations, or vendor suggested savings, must be ascertained before action will be taken. Similarly, suggested changes from the Advisory Council must be viewed in light of their effect on the entire program before the selected path is changed.

Proposed changes will flow to Project Control and Analysis where their effect on other work areas will be reviewed for both cost and schedule impact. Upon a successful review, the Baseline Plan will be modified and the program change instituted.

Annual Budget requests to Code C for COMET will be based on the latest revisions to the Baseline Plan. Upon funding of the Plan, it will then be designated the Baseline for the fiscal year.

Technical Management

The Program Office Systems Engineer will assume the responsibility for COMET's technical leadership. In that role the Systems Engineer will interface with the Payload Selection Committee, will provide technical guidance to member or Lead CCDS's, and when required, will assure that the Systems Engineering contractor is prepared to lead integrated systems tests.

Assisted by the Lead CCDS's and in close contact with the Systems Engineering contractor, the Systems Engineer will monitor technical aspects of the program and continuously evaluate the capability of program participants and their hardware to meet objectives. In this role the Systems Engineer will depend on the ability of the

Lead CCDS's to evaluate progress, identify problems, and highlight opportunities within their respective contractors.

Communication

While the Program Management personnel are located in the close proximity of each other, the Lead CCDS's are expected to be somewhat removed. Our approach recognizes this challenge and finds it imperative that modern communication aides be merged with a management philosophy and structure which stresses frequent, open and frank communications between the various team members. We have great respect for the contractors who will be supporting the FreeFlyer. In many instances, they will place the entire future of their company on the success of COMET. We share in their future and will strive to maintain a free-flow of communications consistent with good business practice.

Monthly summaries of activity will be distributed by the Program Management Office based on inputs from all members. Hardware status, launch readiness successes, opportunities, and problem areas will all be highlighted. Information gathering will lean toward informal means. Formal letters (numbered or otherwise) between Program members will be kept to a minimum consistent with good commercial business practice.

An annual summary of activity will be prepared by the Program Office to fulfill requirements of the Grant and Cooperative Agreements Handbook. Similar documents from Lead CCDS's will not be required.

Financial Information

Each of the work area Contractors will submit monthly summaries of project expenditures and financial forecasts. To this, Lead CCDS's will add milestone status, and technical achievements.

Contractor, Lead CCDS's, and program office expenditures will be gathered and maintained at the Program Management Office. Summaries of this information will be made available to all COMET Management Team members monthly along with a brief analysis of variances.

Financial data pertaining to the COMET program which is required by OCP will be generated by the Program Management Office and will include information pertinent to all participants in the program. Budget or expenditure information which concerns data deemed proprietary by either the contractors or Lead CCDS's will not be divulged outside the Program office. Such information will be requested by the Program Office only when needed to substantiate estimates or billings, or evaluate specific problem areas.

Procurement Plan

While it was possible to structure the procurement of the COMET launch vehicle, etc. into a single package, it was preferred to divide the overall project into six major elements. This led to increased probability of competition, and enhances the ability of the COMET team to select the best mix of contractors to accomplish the COMET program. Dividing the work into more compact packages also led to increased participation from smaller, specialized companies whose capital structure may not have allowed them to bid on the entire COMET Program.

Our approach of breaking the procurement into smaller pieces does not, by it's nature, rule out a single contractor or consortium from proposing on the entire effort. Such a proposal would be evaluated using the same criteria as required by each segment of work.

To assure a coherent approach during program startup, procurement goals were established:

a. One launch vehicle and a COMET FreeFlyer would be procured in FY91 for launch in mid FY92. After 1992, launches will be conducted annually until both budget and commercial need dictate a greater rate.

b. Systems Engineering will be initiated in FY91 and continue for the life of the program.

c. A low level of Orbital Operations and Payload Integration will be funded during 1991 for support to be provided to the launch in 1992.

d. Program Management and a number of CCDS's will prepare the RFPs, proposal evaluation, and program structuring in FY90.

While some may be qualified to bid on the contracts, CCDS's and other non-profit organizations or companies will not be considered as prime contractors for the work elements of the COMET. Similarly, we believe that participation of prime contractors should be limited to those in which both control and ownership rests within the U.S.

These guidelines recognize that an objective of the COMET program is to develop a lasting, U.S. based infrastructure for servicing commercial needs in space. It is the COMET Management Teams belief that this goal can be best achieved by qualified U.S., for profit firms.

To take advantage of the economies of an open marketplace, all COMET contracts will be competitively bid using a streamlined variant of the procedure generally followed by governmental agencies and large aerospace firms. We have followed the schedule outlined in Figure 9 in the preparation of statements of work and the RFP. All proposers have been advised that their proposal may be accepted without negotiation or further discussion. Results of the review process will determine whether oral presentations or best and final offers will be requested.

SIGNIFICANT COMET EVENTS
Program Initiation

Event	Date
Concept Finalized	May 20 '90
OCP Authorization	May 24
Draft RFP Issued	July 18
Industry Briefing	July 31
RFP Issued	August 22
Proposals Received	October 3
Evaluations Started	October 5

Figure 9

COMET's procurement plan was activated with the release Announcement of Opportunity was released and respondents provided qualifications and area(s) of interest. A draft RFP approach was selected so that comments could be solicited from industry concerning and evaluated for inclusion into the RFP. As part of the information flow to industry, prospective bidders were invited to a industry briefing conducted at the University of Tennessee Space Institute. Over one hundred representatives from sixty one industrial firms participated.

Questions posed during the review and during the ensuing two weeks were answered and, where appropriate, folded into the RFP. Six weeks were allocated for responses to the RFP and sixteen prime contractors have submitted proposals.

Based on these responses, Lead CCDS's have formed Selection Evaluation teams which include the appropriate legal, contracts, financial and technical personnel. Using evaluation criteria generated during Statement of Work preparation, each proposal will be graded on its managerial and technical approach, cost and other factors. Technical talent, risk, and cost realism will also be viewed by the evaluation teams. The CCDS based Evaluation Teams will weigh each proposal based on it's strengths, weaknesses, and the ability of the proposer to accomplish the COMET mission. These teams will provide their results to the Selection Advisory Council who will consolidate the various proposals onto approaches which would accomplish all COMET work areas. These optimum approaches will be presented to the Program Manager who will determine the Contractor(s) best suited to implement the COMET program.

If required, oral presentations , best and final offers, and negotiations will be then be conducted with contract signing planned before the end of CY90.

After Subcontract approval is received from NASA followed by Grant funding, the contractors will be given Authority To Proceed (ATP) by the Program Manager.

Contracts

To take advantage of the flexibility gained by defining six distinct areas of performance, contract types will vary based on the peculiarities of the area of performance. In general, fixed price contracts may be used where work statements and performance standards are more easily defined, i.e. launch vehicle and services and the FreeFlyer components. Exact levels of support required for

Orbital Operations, Systems Engineering, and Payload Integration are more difficult to define. Contracts written for these would most probably be cost plus fixed fee contracts. These levels of contract effort are envisioned:

a. Launch Vehicle

Contract for 3 launches with an option for 2 additional launches. If exercised, options for the fourth launch would be executed prior to the launch of the third vehicle. Insurance requirements of DOT/OCST must be met by contractor.

b.Recovery System

Contract for three flights with two options for an additional flights. With a projected lifetime of 10 or more refurbishment cycles, the Recovery System would be capable of supporting a continuing experiment load throughout the decade. Insurance requirements of DOT must be adhered to by contractor.

c. Service Module

Three modules will be required on the basic program. At a launch rate of one per year, the program will be at a point to exercise options in late 93 or early 1994 or re-bid the effort. An option for two modules would allow for an increased launch rate as the program matures or continues to sustain a once/year rate through 1996.

d. Systems Engineering

A contract of sufficient length to include start and three launches with two one year options. Initial activity will focus on system integration interface specifications and overall mission assessment. Re-flight insurance will be purchased by the contractor.

e. Orbital Operations and Payload Integration

A contract sufficient to cover startup and three flights with two one year options. Equipment purchases necessary for startup may cause the contracts to be slightly front end loaded however, once operational capability is reached, the dominant cost will be manpower.

Options

COMET will be structured to provide three years of system operation based on a once per year launch rate. Approximately eighteen months before integration of the third COMET begins, a bid or option decision will be made. If options on any of the six efforts are not exercised, the RFP cycle will be separated to acquire the needed capabilities. Options offer the COMET the capability to continue proven systems while offering the ultimate performance incentive, continued sales, to the COMET contractors.

Competition

An over-riding goal of competition is to get the optimum infrastructure and services at the most realistic cost. Our initial market survey indicated that there were at least three potential proposers for each of the work areas. More surfaced when the Announcement of Opportunity was released. We have observed that teaming opportunities abound and smaller "specialty" companies have joined with larger, better endowed firms.

The probability of achieving adequate competition was enhanced by issuing one RFP with six work areas. There were no restrictions on the number of efforts on which a proposer could bid however, should a proposer be awarded more than one effort, all elements of the allied statements of work will be included in resulting contract. Cost realism will be a closely monitored issue for contracts awarded which have multiple statements of work.

Only US or US owned (51% or greater) companies were eligible to compete as prime contractors. Companies who are members of a CCDS were not be ruled ineligible on the basis of their membership, however that CCDS was not be allowed to participate in the evaluation process of their proposal.

Universities and non-profit or not-for-profit companies or organizations were also not allowed to compete as prime contractors since investment of profit is seen as a primary mode of developing infrastructure.

Sub Contract Approval

Since all funding will be in the form of Grant augmentation, contracts which result from this program will be included in CASP's request for augmentation and the provisions for the Grant and Cooperative Agreements Handbook will be followed. Contracts will be issued by the Center for Aerospace Research, and, after CAR and contractor signature, will be forwarded to the NASA Grants Officer for review and approval. After approval of funding by the NASA Office Of Commercial Programs, the Program Manager will authorize the contractor(s) to proceed.

Role of CCDS's in the Contracting & Procurement Process

CCDS's have played a key role in the structuring the COMET program philosophy, distilling its concepts into practical work areas, and evaluating contractor responses in light of our generalized requirements.

Those who desired to take an active part in the program prepared the statements of work for each of the six contracts. As a part of the work statement, they have identified hardware and software concepts, mission requirements, and standards of quality and safety performance for the contractor. These major definitions provided the structure around which the six SOWs were built.

Based on their intimate knowledge of the tasks to be performed, the Lead CCDS's have structure the evaluation guide for the contract they have prepared, review the RFP in it's various stages of completion, and bring together a Selection Panel. This panel will review and evaluate the technical, management, and cost aspects of all proposals received for that effort. They will make recommendations to a Selection Advisory Council which consists of all CCDS's involved in the evaluation process. The Council will bring the various proposals into viable program options, evaluate risks and opportunities, and make recommendations to the Program Manager who will select the COMET Contractors.

Upon signature of the contract, the Lead CCDS will assign a Monitor to accomplish the contract oversight and technical monitoring of the contractors performance. They will evaluate contractor performance, and provide the Program Management Office their assessment of schedule and milestone achievements. If required, the Lead CCDS may be requested to investigate inconsistencies in performance or cost structure, and participate in buy off and critical test activity.

Conclusion

COMET offers a capable system for orbiting and retrieving commercially oriented prototypes and experiments. By integrating and focusing the efforts of as many as six contractors, we have the opportunity to invigorate the commercial space sector while we build expertise within the CCDS community.

Our approach is to free contractors from extensive specs and standards, place the burden of cost efficiency and quality with the supplier and challenge the contractor team to meet or exceed our performance requirements and stringent schedule.

COMET represents the opportunity for Commercialized Space to become a reality. CCDS's will provide crucial planning and contracting service to assure proper stewardship of the grant funding. For it's part, industry will be called upon to provide hardware and develop an infrastructure for the continuing commercialization of space. Structuring the RFP into six distinct work packages brings with it the challenge of integration but the potential of involving a large number of small firms: firms whose future can be assured by the COMET program.

IF we were to envision the COMET program at its most successful end point, hardware and infrastructure will have matured to the point where one of the participants, or a new start, will find it commercially attractive to market and operate COMET without the CCDS Management Team. At that time COMET will have reached its goal.

ELECTRIC PROPULSION

Preliminary Tests of the Electrostatic Plasma Accelerator

G. Aston
T.L. Acker

Experimental Facilities for Electric Propulsion Testing

Wilhelmus M. Ruyten
Verlin J. Friedly
Xiaohang Peng
Dennis Keefer

Plasma Particle Simulation of Electrostatic Ion Thrusters

Xiaohang Peng
Wilhelmus M. Ruyten
Verlin J. Friedly
Dennis Keefer

PRELIMINARY TESTS OF THE ELECTROSTATIC PLASMA ACCELERATOR

N96-70672

G. Aston* and T. Acker**
Electric Propulsion Laboratory, Inc.
Tehachapi, CA

Abstract

This report describes the results of a program to verify an electrostatic plasma acceleration concept and to identify those parameters most important in optimizing an Electrostatic Plasma Accelerator (EPA) thruster based upon this thrust mechanism. Preliminary performance measurements of thrust, specific impulse and efficiency were obtained using a unique plasma exhaust momentum probe. Reliable EPA thruster operation was achieved using one power supply.

Introduction

The lifetime of a typical commercial, communications satellite is limited by the mass of station keeping propellant which is carried on-board. For a typical communications satellite at geosynchronous altitude with a ten year useful life, it is generally true that each one second increase in specific impulse results in a station keeping propellant mass reduction of approximately one kilogram.

Arcjet and ion engine electric propulsion systems offer significantly increased specific impulse compared to chemical thrusters. The performance advantages offered by these relatively sophisticated propulsion systems, in general, outweigh the more complex spacecraft integration requirements their use imposes on the spacecraft. However, because these more advanced spacecraft propulsion systems are more complex, and costly, their use will probably be limited to the more expensive, large communication satellites.

Smaller, relatively inexpensive spacecraft are being developed for earth orbital mission applications. Although the mission lifetimes of these spacecraft are usually short (months, or one or two years), minimizing mass launched to orbit is still very important because of the low payload capability of the small launch vehicles. A need exists for an inexpensive electric propulsion thruster, with minimal complexity, but which still offers significantly improved specific impulse over chemical thruster systems.

The EPA thruster concept consists of a hollow cathode surrounded by a discharge chamber which is completely open at the end opposite the cathode. No accelerating electrodes are used and the specific impulse is a direct function of the applied discharge voltage and the propellant mass. Briefly, during thruster operation, a plasma is created by electron bombardment of the propellant gas in a plasma arc discharge which is sustained by the applied voltage between the cathode and the discharge chamber anodes. An initial start-up transient will occur in which more electrons will leave the thruster than ions. This will cause the discharge plasma to acquire a potential more positive than the ambient space plasma potential to prevent the further unbalanced loss of electrons. As the discharge plasma potential becomes more positive, ions will begin to be accelerated from the thruster. The potential difference between the discharge chamber plasma and the ambient space plasma will automatically adjust itself so that the loss rate of ions and electrons are equal; thus a neutral plasma will be exhausted from the thruster. The net ion velocity (and ultimately the specific impulse) is determined by the magnitude of this potential difference.

From the above discussion, it is apparent that the electrostatic acceleration process requires, in principle, only one power supply. Furthermore, this power supply could be the spacecraft bus since even voltages of 28 Vdc are sufficient to achieve significant plasma acceleration. The promise of thruster operation off the spacecraft bus, with minimal power interface requirements, makes the EPA thruster concept attractive for low cost small spacecraft.

The thrust producing mechanism for the EPA thruster was first discovered by chance in 1979 in the course of the extended SERT II (Space Electric Rocket Test II) test program.[1] In these tests, one of the SERT II mercury ion engines was operated in a discharge only mode without high voltage applied to the accelerator system. While operating in this mode, it was discovered that the thruster was producing a measured thrust level of 0.8 mN. Based on the measured thrust of 0.8 mN, Ref. 1 calculated a specific impulse of 300 sec. at an input power of 80 W, resulting in a thrust-to-

---

* Director, Member AIAA
** Research Engineer, Member AIAA

power ratio of 10 mN/kW for operation in this mode. The calculated 300 sec. specific impulse level was much higher than could be accounted for by simple gas dynamic expansion of heated mercury vapor.

This paper describes the results of a program to verify, through ground based tests, the electrostatic plasma acceleration concept and to identify those parameters most important in optimizing a thruster based upon this thrust mechanism.

Proof of Concept Thruster

A prototype EPA thruster was fabricated to perform verification testing of the electrostatic plasma acceleration concept. Primary thruster components included a magnetically protected plasma discharge chamber and a high current hollow cathode. Specific design details of these components are described below.

The EPA thruster discharge chamber required a magnetic field distribution which could prevent direct access to the anode electrodes by cathode electrons. In the interests of minimizing costs to this present program, a magnet retainer shell was loaned from the Jet Propulsion Laboratory (JPL). This structure was fabricated several years ago during a NASA/LeRC funded program concerned primarily with ion accelerator system development.[2] Figure 1 shows this magnet retainer shell. As can be seen from this figure, each side of the structure is lined by five rows of samarium cobalt magnets, forming a strongly diverging line cusp magnetic field distribution. This magnetic field distribution was not selected a priori as being optimum for the EPA thruster proof-of-concept tests. However, because the magnets in the retainer shell shown in Fig. 1 are water cooled, it was felt that a much wider range of operating conditions could be explored without any concern for hardware damage.

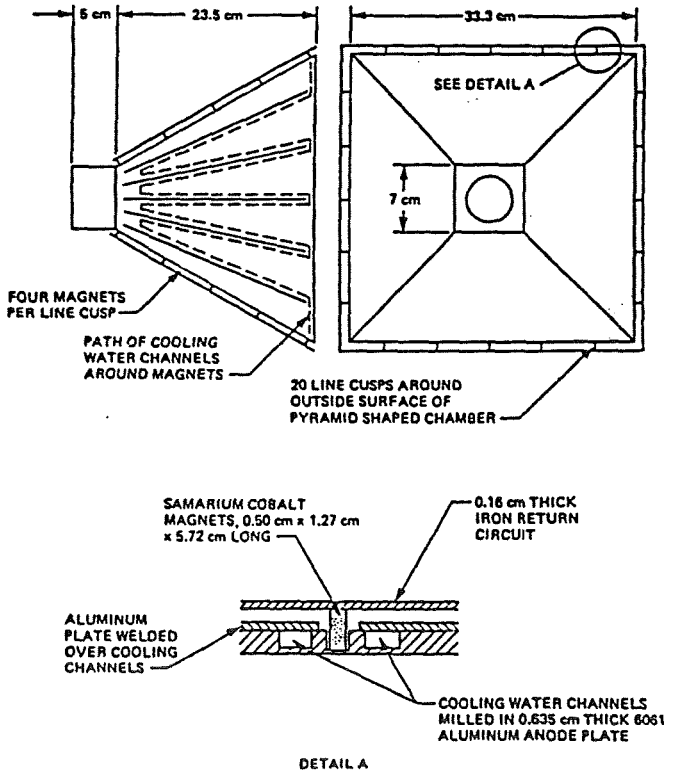


Fig. 1 Magnet retainer shell used during EPA thruster proof-of-concept tests.

A hollow cathode was designed to provide the discharge plasma electron current for the prototype EPA thruster and is shown in Fig. 2. This cathode was designed to provide maximum operating flexibility, providing emission currents of a few ampere up to over 100 ampere. This large current capability allowed a wide range of input powers to be investigated during proof-of-concept testing of the prototype EPA thruster.

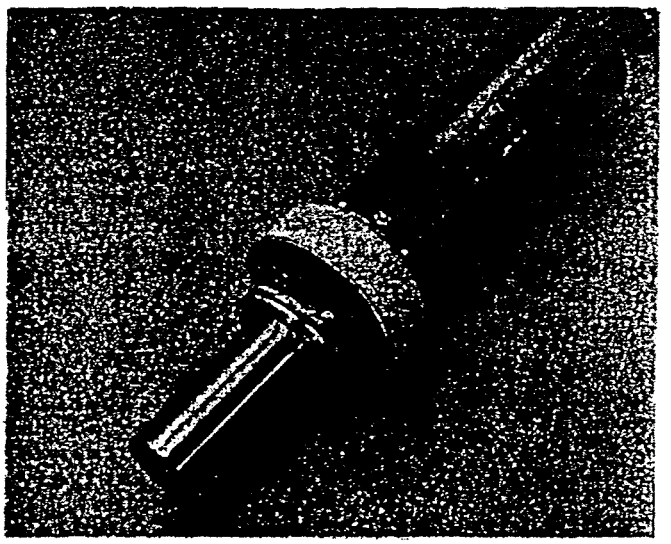


Fig. 2 Large, high current, hollow cathode used in prototype EPA thruster.

Vacuum Test Facility

Figure 3 shows the test facility used to perform the EPA thruster proof-of-concept tests. A chamber, 0.59 m in dia. and 1.83 m long, was pumped to high vacuum by a diffusion pump with a nominal pumping speed for air of 17,500 liters/sec. Backing this diffusion pump was a mechanical pumping system which included a blower and single stage roughing pump. With this vacuum tank and pumping system, background pressures were typically in the mid 10^{-5} Torr pressure range during EPA thruster operation with xenon propellant.

Power Supplies

Several power supply sets were used to operate the EPA thruster. Cathode

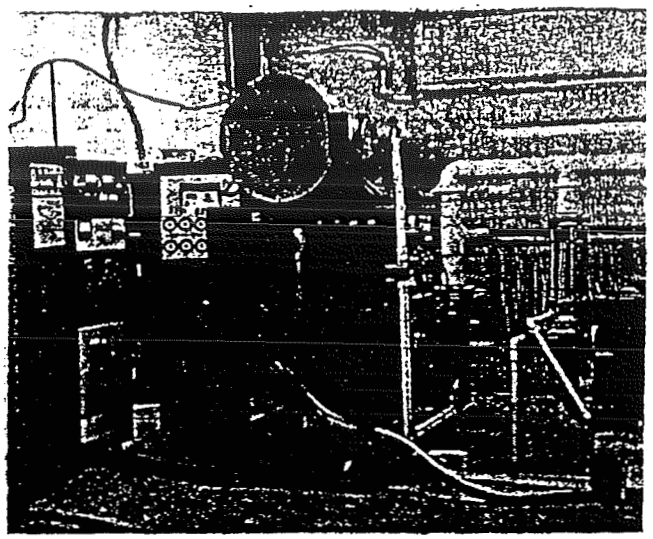


Fig. 3 Vacuum facility, power supplies and instrumentation.

heater power was provided by a ferroresonant supply which initially provided about 60 W to the heater immediately after turn-on with a gradual power increase to approximately 110 W after about 10 min. The relatively long cathode preheat time was a function of the large cathode size and mass. No attempts were made to reduce this preheat time since this was not an important consideration during this proof-of-concept test program. For cathode-to-keeper electrode breakdown and cathode start up, five series connected ferroresonant supplies provided an open circuit voltage of approximately 750 Vdc with a constant current capability of 3.5 A. As with the cathode preheat time, the cathode-to-keeper discharge starting voltage requirements were a function of the specific general purpose hollow cathode design chosen for these proof-of-concept EPA thruster tests and no attempt was made in this study to optimize this parameter.

Two sets of cathode-to-anode discharge power supplies were used during this EPA thruster test program. The first of these power supply sets comprised eleven ferroresonant power supplies each of which had a nominal output voltage of 28 Vdc and a constant current capability of 25 A. These supplies could be arranged in a variety of series and parallel combinations to simulate different spacecraft bus voltages. Figure 3 shows this power supply assembly which is incorporated into the rack at the left of this photo. The adjacent rack contains the keeper supply assembly discussed above. Next to this keeper power supply system is the second power supply set used to maintain the cathode-to-anode discharge. The principal power supply used in this rack to control the EPA thruster was a transistor regulated constant current/constant voltage unit which was always operated in the constant current mode.

Momentum Probe

To demonstrate proof-of-concept of the electrostatic acceleration process, it was necessary to quantify the thrust, specific impulse and efficiency of the prototype EPA thruster. A momentum probe was developed as a direct thrust measuring device in order to determine these parameters.

The main design features of the momentum probe are shown in Fig. 4. Plasma beam ions and neutrals from the EPA thruster impinge upon a very thin mica surface. The momentum exchanged between these particles and the mica surface exerts a small thrusting force which moves the mica probe to the left and the meter movement pointer to the right, away from its initial null position. To reestablish the meter pointer null, and therefore resolve the out of balance thrust force on the mica probe surface, a thin gold chain is unwound until the meter pointer is returned to its original null reading. The subsequent thrust force on the mica surface can then be calculated using the following trigonometric relationship:

$$T_d = (m\ f)/(2A \tan \theta)$$

In this equation, T_d is the thrust density; m is the mass increment of unwound chain; A is the area of the mica probe

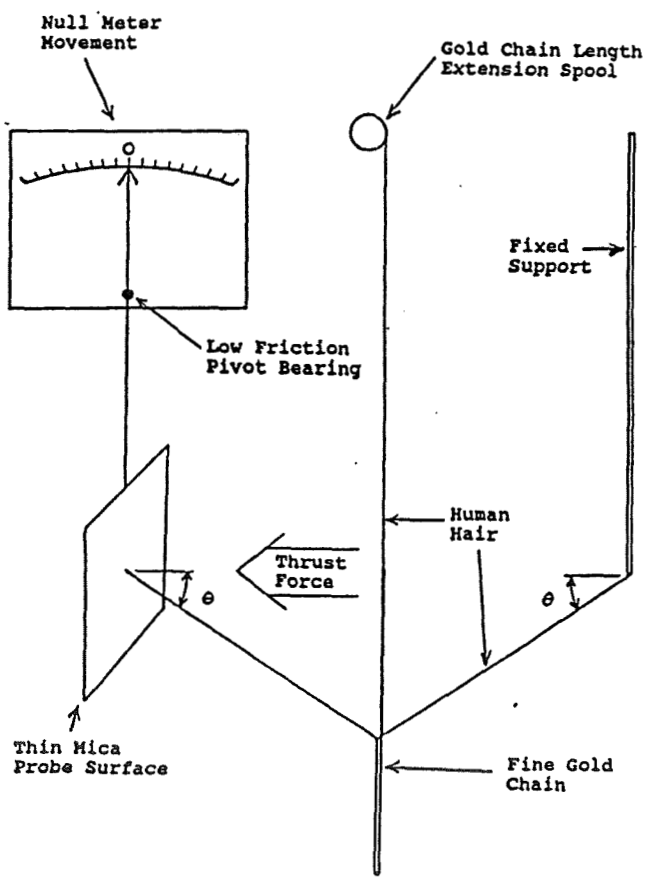


Fig. 4 Primary components of the momentum probe.

surface; θ represents the angle depicted in Fig. 4; and f is a multiplication factor that accounts for reflection of the impinging xenon particles off the mica probe surface. A value of f = 0.5 corresponds to an assumption of pure specular reflection, whereas f = 1.0 corresponds to the measured thrust density. The actual thrust density lies somewhere between these limits since the particles undergo diffuse reflection from the mica probe surface. While the probe system shown in Fig. 4 is believed to be a unique design, other workers have successfully used similar direct force measurement techniques to determine the thrust of other low thrust electric propulsion devices.[3,4]

Probe Fabrication:

Figure 5 shows the assembled momentum probe positioned in front of the EPA thruster. Construction of the probe was accomplished using a 4.1 cm x 8.2 cm thin mica plate connected directly to a microampere meter via a thin aluminum rod. A human hair was used to link the fine gold chain to the chain length extension spool as well as to the mica probe surface and the fixed support.

Fig. 5 Momentum probe mounted in front of EPA thruster.

Adjustment of the chain position was performed by rotating a flexible cable that was connected to the extension spool and a vacuum chamber feedthrough. During operation of the momentum probe, it was possible to resolve thrust differences equivalent to one link of the thin gold chain which corresponds to a mass, or thrusting force, of 2.05×10^{-3} g (2.01×10^{-5} N). The probe was mounted to a motorized carriage allowing thrust measurements to be taken at different locations along the horizontal axis of the thruster. Both the probe position in front of the thruster and the extension length of the gold chain were measured using precision potentiometers.

Cold Flow Gas Tests:

Calibration of the momentum probe was performed by measuring the cold flow specific impulse of xenon gas expanded through a conical nozzle. An illustration of the experimental set up is shown in Fig. 6. The nozzle was constructed of paper and attached to a stainless steel tube with an orifice diameter of 2.5 mm. The area ratio and half angle of the nozzle were approximately 300 and 20°, respectively.

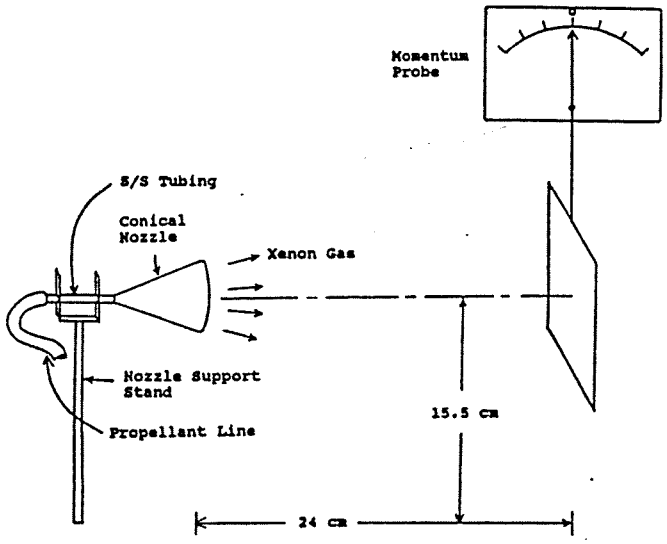


Fig. 6 Experimental setup of the cold flow specific impulse test.

The background pressure maintained in the vacuum chamber during the cold gas flow calibration experiment was 1.0×10^{-4} torr. Simple, one dimensional isentropic flow calculations indicated that a choked flow condition should exist at the throat of the nozzle at a xenon flow rate of 30 sccm. While operating the nozzle at 30 sccm, the thrust density data shown in Fig. 7 was obtained (half profiles are shown in this figure). This data corresponds to a cold flow specific impulse of 8.4 seconds (f=0.5) or 16.8 seconds (f=1.0). The actual value of the specific impulse should fall between these two values.

Expanding xenon in a well made expansion nozzle should deliver a cold flow specific impulse of approximately 25 seconds[5]. Comparing this value to the specific impulse measured with the momentum probe suggested that the probe was functioning reasonably accurately since at low mass flow rates through the nozzle, viscous effects impede the expansion of the xenon and so lower the measured thrust and specific impulse.

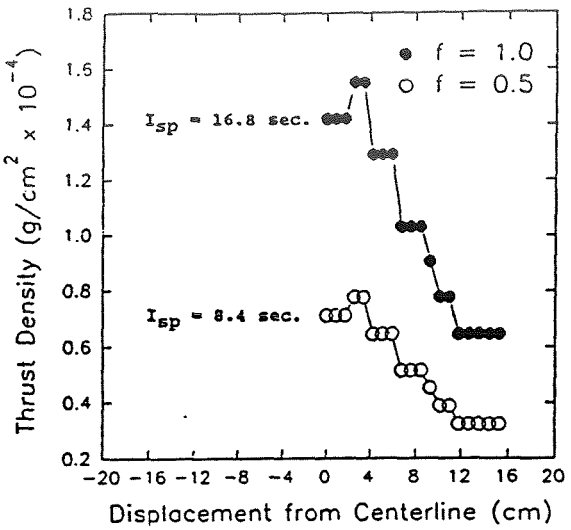


Fig. 7 Cold flow thrust density measurements.

Probe Error Analysis:

To complete the discussion describing the momentum probe, consideration must be given to the error inherent in measuring a given data point. The major factors influencing the accuracy of a thrust measurement were the measurement of the chain length, mass flow rate, exhaust area of the beam, and the angle with the horizontal subtended by the hair connecting the gold chain with the mica probe surface. Table 1 lists the uncertainty associated with each of these variables. Though the length of the gold chain could be resolved to within one link, the 3% error related to the high precision potentiometer reading corresponded to a chain length differential of about two links and was used in all error calculations. The absolute errors in thrust density, specific impulse and efficiency are shown in Table 2.

EPL Thruster Operation

The first set of runs with the EPA thruster were used to verify operation of the momentum probe and to investigate the effects of different parameters on thruster operation. Upon analyzing the data to determine thrust, specific impulse, and efficiency, subsequent modifications were made to the thruster configuration in order to further study thruster performance. During these tests, the thruster was started 55 times, accumulating a total run time of 9 hours and 40 minutes. Results of several of these tests are summarized below.

Preliminary Tests:

Initial tests of the EPA thruster included a total of 20 data runs in which several thruster operating parameters were varied during operation of the momentum probe. The momentum probe was extremely stable during transit across the thruster exit plane and the typical time required to obtain a probe thrust density measurement was approximately 30 seconds. Specific thruster operating conditions corresponding to these separate runs are presented in Table 3. Runs #1 - #5 showed thrust density profiles within the EPA thruster plasma exhaust. Due to a mechanical problem in the probe carriage, data runs #6 through #12 shown in Table 3 were taken at a fixed position in front of the thruster.

Table 1
Uncertainty in Momentum Probe Parameters.

Variable	Uncertainty
Chain length	± 1 link
Potentiometer reading	± 3 %
Hair angle, θ	± 1°
Beam radius	± 2 cm
Mass flow rate	± 1 sccm

Table 2
Percent Error in Performance Parameters.

Parameter	% error lower bound	% error upper bound
Thrust density	-26 %	+29 %
Specific impulse	-36 %	+50 %
Efficiency	-61 %	+136 %

Table 3
Preliminary EPA Thruster Performance Tests.

$\dot{m}_c$ sccm	$\dot{m}_m$ sccm	V_D V	I_D A	f = 0.5 T g(10^{-4})	f = 0.5 I_{sp} sec	f = 0.5 Eff %	
12	80	0	0	0.20	17.6	-	*
12	80	0	0	0.30	25.8	0.35	§
12	80	0	0	0.44	37.8	0.42	▼
6	40	0	0	0.24	39.0	0.20	▼
6	40	40	10	0.48	76.6	0.32	▲
6	40	43.5	15	0.54	87.6	0.28	▲
6	40	40	10	0.44	71.4	0.19	▲
6	40	47	20	0.64	104	0.30	▲
6	40	51	25	0.80	130	0.35	▲
6	30	45.5	15	0.48	96.4	0.27	▲
6	50	40	15	0.58	79.0	0.30	▲
3	50	41	15	0.62	87.8	0.34	▲
6	50	47	25	0.84	111	0.56	▲
6	50	46.5	30	1.00	133	0.44	▲

Comments: * cold gas flow
§ heater on
▼ heater/keeper on
▲ discharge on

The analysis used to produce the specific impulse and efficiency numbers listed in Table 3 was deliberately very conservative. For example, the thruster mass flow rate was assumed to include the mass flow rate of background gas from the vacuum tank which was assumed to be re-ingested by the thruster. This calculated background mass flow rate is a function of facility pressure and was assumed to be composed entirely of xenon. For these preliminary data, EPA thruster performance calculations were performed assuming purely specular reflection off the probe wherein the measured thrust density values were halved, that is f was assumed to be 0.5. Given these assumptions, Table 3 shows that maximum specific impulse values of 100 - 130 sec. were achieved. Note that the accuracy of these larger specific impulse values obtained during runs #6 - #12 was questionable since only one thrust density point was taken per run and the exhaust plume area was only approximated.

Propellant Manifold Redesign:

After the initial set of data runs, it was realized that one contributor to the low values of specific impulse measured in Table 3 was the thruster propellant injection configuration and the small plasma production volume caused by the far downstream cathode axial position. To overcome these problems, the cathode was relocated far back in the rear of the thruster and the downstream propellant injection manifolds were replaced by a single propellant injection manifold that surrounded the rear of the cathode. This revised EPA thruster configuration is pictured in Fig. 8.

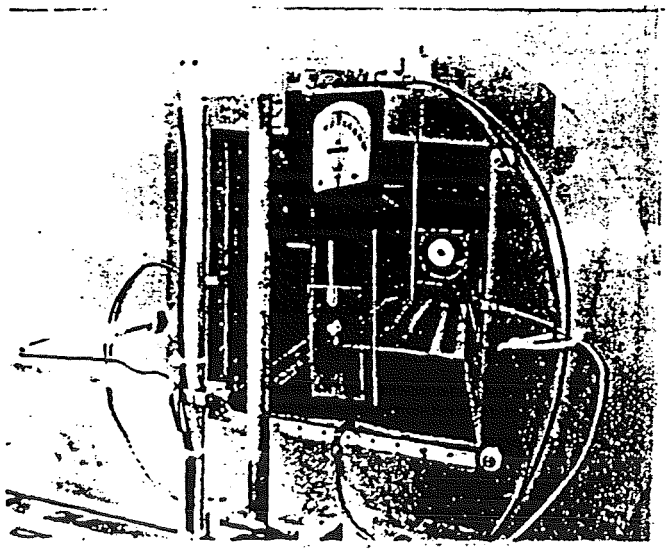


Fig. 8 EPA thruster with relocated cathode and propellant manifold.

Figure 9 documents the discharge current and voltage characteristics of the modified EPA thruster as a function of the propellant flow rate through the cathode and rear manifold. Comparing the results of Fig. 9 with those from Table 3 shows similar discharge current and voltage characteristics are attained with this reworked EPA thruster configuration but that the required propellant flow rates have been reduced by a factor of four. The results of Fig. 9 also show that for a fixed manifold flow rate and discharge current, the discharge voltage was reduced substantially as the cathode flow rate was increased (squares, upside down triangles and diamonds).

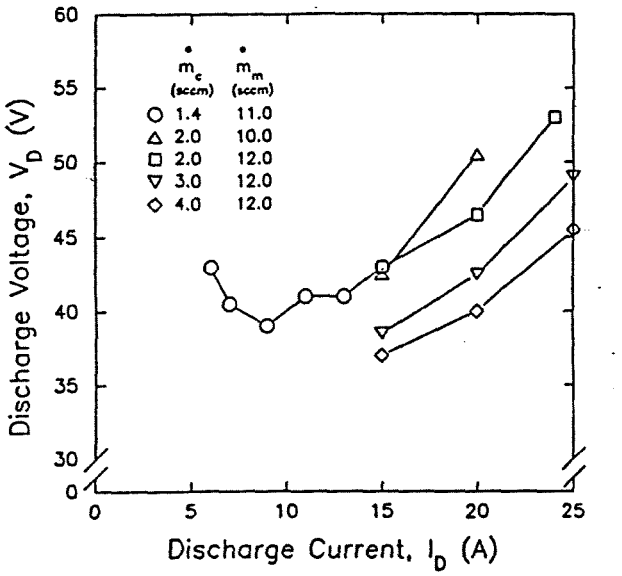


Fig. 9 EPA thruster discharge V-I characteristics after reducing the thruster propellant flow rate requirement by a factor of four.

In addition to modifying the propellant injection manifold, the thruster ground screen was also modified slightly in Fig. 8 compared to the preliminary thruster configuration shown in Fig. 5. Figure 10 shows thrust density profiles corresponding to various operating conditions of the thruster configuration shown in Fig. 8. Data at a discharge voltage of 41 V were taken within several minutes of each other with the thruster at the same operating conditions. These runs show that the reproducibility of the momentum probe measurement system was very good. Also in Fig. 10, the effect of turning off the cathode heater was seen to increase the discharge voltage and discharge power. This voltage and power increase reflects the plasma compensating for the lack of heater power by increasing the energy of the backstreaming ions to the cathode to increase cathode self-heating by ion bombardment.

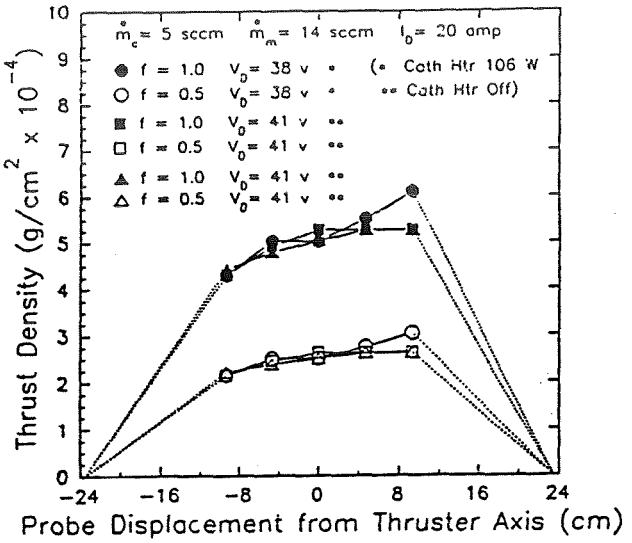


Fig. 10 Thrust density profiles after propellant manifold and ground screen modifications.

Figure 10 shows that the thrust density profiles are skewed. This behavior was directly related to the plasma distribution within the thruster which was non-uniform. Because of this plasma non-uniformity, a data analysis procedure was applied to the thrust density profiles shown in Fig. 10, and later results, wherein a straight line drop-off was assumed between the last probe data point and the beam edge (dotted lines in Fig. 10). The beam edge was determined by a separate ion flux probe in the same plane as the momentum probe but positioned far enough off the thruster axis to detect the accelerated plasma exhaust boundary. The thrust density data was analyzed by integrating under the profiles shown in Fig. 10 to obtain total thrust and then using this thrust value to calculate specific impulse and efficiency. Table 4 shows the results of these calculations for f = 0.5 and f = 1.0 (corresponding to the actual measured thrust density). Comparing these results with the cold flow specific impulse measurements of Fig. 7, shows more than an order of magnitude increase in specific impulse.

Magnet Removal:

Visual examination of the EPA thruster showed a severe non-uniform plasma discharge which was evidenced by the skewed thrust density profiles shown in Fig. 10. In particular, most of the discharge was seen to concentrate in opposite corners of the thruster. Figure 11 shows a picture of the thruster operating with this severely non-uniform plasma. Superimposed on this tendency to concentrate plasma production in opposite corners, was an additional effect wherein every second magnet line cusp showed no visible evidence of plasma luminosity.

Table 4
EPA thruster performance tests with partial optimization of propellant injection system and modified ground screen

	f = 0.5			f = 1.0		
V_D V	Thrust g(10^{-4})	I_{sp} sec	Eff %	Thrust g(10^{-4})	I_{sp} sec	Eff %
38	0.33	117	0.22	0.66	234	0.86
41	0.31	108	0.19	0.61	215	0.77

Comments: $\dot{m}_c$ = 5 sccm
$\dot{m}_m$ = 14 sccm
I_D = 20 A

Fig. 11 Thruster operation showing plasma concentration in large corner cusps.

Inspection of the magnetic field distribution in the corner of the EPA thruster suggested that the second row of magnets in from each corner may have been short circuiting the magnet cusp at the corners, effectively eliminating these cusps from the discharge. To overcome this problem, all corner magnets were removed from the magnet retainer shell. Figure 12 shows a photograph of thruster operation following this change and clearly shows much less concentration of the plasma at the corners. However, this figure still shows the alternating magnet cusp luminosity and subsequent plasma non-uniformity which appeared to be a permanent feature of the pyramid shaped magnet retainer design.

Table 5 shows the thruster performance parameters calculated after the magnets were removed. Of most interest is the large relative efficiency increase caused by small flow rate changes. In particular, stable thruster operation was achieved at a discharge voltage of 38 V. These data

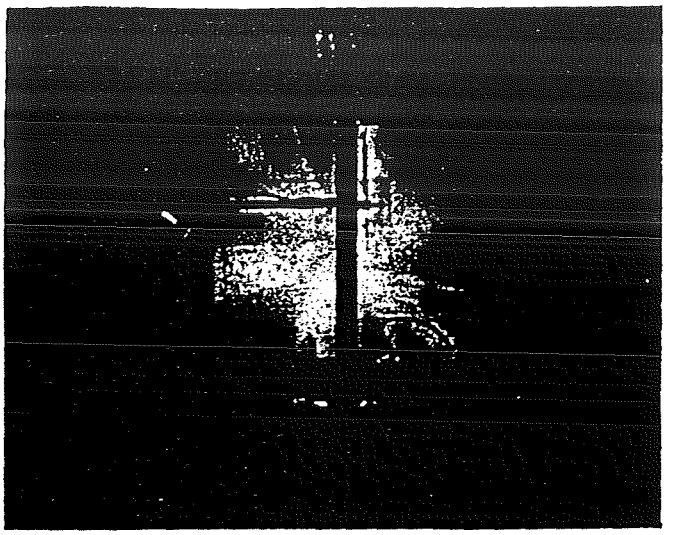


Fig. 12 Thruster operation after magnet removal showing absence of large plasma filled corner cusps but still showing alternating magnet cusps with and with out plasma present.

Table 5
Thruster performance parameters after magnet removal

f = 0.5			f = 1.0			
Thrust $g(10^{-4})$	I_{sp} sec	Eff %	Thrust $g(10^{-4})$	I_{sp} sec	Eff %	
0.25	89	0.14	0.49	178	0.55	*
0.28	94	0.22	0.55	189	0.88	§

Comments: $\dot{m}_c$ = 2 sccm
I_D = 15 A
* $\dot{m}_m$ = 17 sccm and V_D = 51 V
§ $\dot{m}_m$ = 21 sccm and V_D = 38 V

suggested that discharge voltages of order 28 V could be expected with further thruster design optimization.

Comparison with the results of Table 4, shows that specific impulse did not decrease significantly following removal of the corner magnets. Since one third of the magnets were removed from the engine in making this change, this behavior suggests that the discharge chamber magnetic field geometry was far from an optimum configuration. Poor plasma electron magnetic field containment was the principle reason for very low thruster efficiency values during this proof-of-concept program.

Discharge Chamber Masking:

After testing the thruster with the corner magnet cusps removed, a mica mask of 20 cm dia. was placed over the thruster exit plane. The purpose of this test was to pressurize the discharge chamber by preventing direct escape of a portion of the neutral propellant. A photograph of the masked down thruster during operation is shown in Fig. 13.

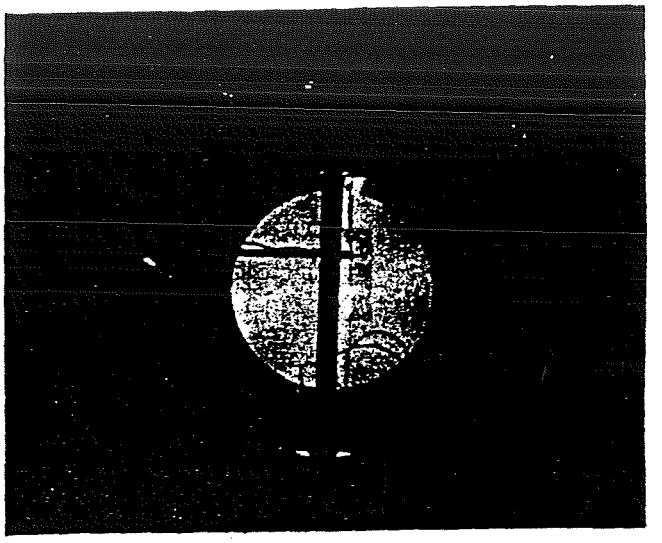


Fig. 13 Thruster operation with circular mica mask over the discharge chamber exit plane.

Figure 14 plots the measured thrust density characteristics of the masked down EPA thruster. The thruster performance parameters calculated for these profiles are presented in Table 6. It should be noted that only one data point was obtained at a discharge current of 30 A and then the momentum probe hair broke allowing the gold chain to be lost to the inside of the vacuum system diffusion pump. This single data point was analyzed by assuming a thrust density profile identical in form to that obtained with a 25 A discharge current.

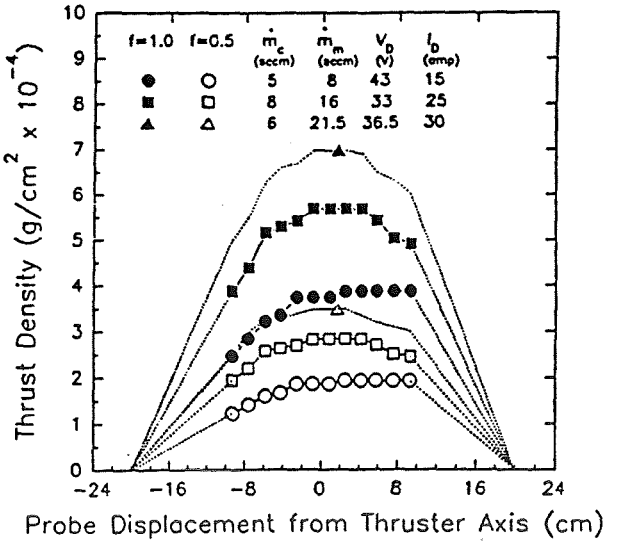


Fig. 14 Thrust density profiles with masked down EPA thruster.

Table 6
Thruster performance parameters after fitting mica mask over the discharge chamber exit plane

$\dot{m}_c$ sccm	$\dot{m}_m$ sccm	V_D V	I_D A	f = 0.5 Thrust g(10^{-4})	f = 0.5 I_{sp} sec	f = 0.5 Eff %	f = 1.0 Thrust g(10^{-4})	f = 1.0 I_{sp} sec	f = 1.0 Eff %
5	8	43	15	0.13	74	0.06	0.25	148	0.24
8	16	33	25	0.18	59	0.05	0.35	117	0.21
6	21.5	36.5	30	0.27	83	0.09	0.55	166	0.36

In general, thruster performance with the discharge chamber mask was worse than without it. This occurred since the mica mask created a very large surface for discharge plasma recombination with correspondingly large reductions in plasma production efficiency. Nevertheless, data shown with a 15 A discharge current is significant because stable thruster operation was achieved at relatively low flow rates. The cathode heater was turned off while operating at a 15 A discharge current. After several minutes, thruster operation stabilized at a discharge voltage of 51 V. These conditions represented a discharge power increase of 120 W which was similar to the normal cathode heater power requirement of 106 W. This large cathode power requirement was a further cause for very low thruster efficiency values during this program.

Summary

A systematic experimental program was undertaken to investigate thrusting effects caused from electrostatic plasma acceleration from a moderately dense plasma discharge. Preliminary results obtained with prototype hardware showed accelerated plasma specific impulse levels at least an order of magnitude larger than could be achieved by cold gas flow expansion. While thruster efficiency levels were very low this appeared to be a result of the inherent operating power requirements and geometrical deficiencies in the prototype hardware used for this investigation. Reliable EPA thruster operation was achieved using one power supply.

Acknowledgement

This work was supported by the Center for Advanced Space Propulsion (CASP) under subcontract CAR/EPL 89-01, and by in-kind support from the Electric Propulsion Laboratory, Inc.

References

1. Kerslake, W. R. and Ignaczak, L. R., "SERT II 1980 Extended Flight Test Thruster Experiments," AIAA Paper No. 81-0665, April 1981.

2. Aston, G. and Deininger, W. D., "Test Bed Ion Engine Development," NASA CR-174623, March 1984.

3. Snyder, A. and Banks, B.A., "Thrust measurements of a hollow cathode discharge", NASA TN D-6705, March 1972.

4. Nakanishi, S., Private communications, Mar.-April 1990.

5. Holcomb, L. B., "Satellite Auxiliary Propulsion Selection Techniques," JPL Tech. Rep. No. 32-1505, Nov. 1970.

EXPERIMENTAL FACILITIES FOR ELECTRIC PROPULSION TESTING*

Wilhelmus M. Ruyten**, Verlin J. Friedly†, Xiaohang Peng‡ and Dennis Keefer§

University of Tennessee-Calspan
Center for Advanced Space Propulsion
UTSI Research Park
Tullahoma, TN 37388

ABSTRACT

Experimental facilities for electric propulsion testing at the University of Tennessee Space Institute, jointly operated by the Center for Advanced Space Propulsion, include a cryogenically cooled, computer-controlled, 9 ft diameter by 20 ft long vacuum chamber, capable of maintaining a baseline pressure of 10^{-7} Torr, a 30 cm diameter ion thruster, and laser diagnostic and electrical diagnostic instrumentation. These facilities are described here in some detail, with emphasis on the measurement of ion thruster parameters such as neutral and ion densities, temperature, and velocity in the exhaust, as well as plasma potential and electron temperature.

I. INTRODUCTION

Although electric propulsion devices such as ion thrusters and arcjets have been around for many years, and much ground-based testing has been performed, critical issues remain unsettled, limiting the deployment of these devices in actual space missions. With regard to ion thrusters, one of these critical issues is the erosion of the extraction grids [1]. Ground based studies of ion thrusters indicate that this erosion is a serious issue indeed, but that the erosion in ground-based tests overestimates the erosion that would take place in an actual space flight due to the inability to maintain a space-like vacuum in a ground test.

Two parallel developments at the Center for Advanced Space Propulsion are underway that might resolve the issue of erosion rates in an ion thruster operated in a space-like vacuum. One of these is the completion of a large, cryogenically cooled vacuum facility. While this facility may not actually reach the ultrahigh vacuum of space for all operating conditions of the thruster, the background pressure will be sufficiently low to allow a reasonable extrapolation of the results to those for a space vacuum. The other development, described in detail in the accompanying paper [2], is that of a numerical code capable of predicting erosion rates for a wide range of operating parameters for the thruster. However, in order for this numerical code to provide not only qualitative but also quantitative results, accurate values for a number of experimental parameters must be obtained, for example, plasma potential, electron temperatures, and ion velocities in various regions of the thruster flow.

This paper describes the experimental facilities that will be used to obtain these thruster data. The vacuum facility is described in Section II. Section III gives a brief description of the ion thruster that is slated for testing in this facility, and Sections IV and V discuss the optical and electrical diagnostic methods, respectively, that are under development. Conclusions are presented in Section VI.

II. VACUUM FACILITY

A schematic of the vacuum chamber, jointly operated by The University of Tennessee and the Center for Advanced Space Propulsion, is shown in Fig. 1. The chamber has a steel shell that is 20 ft long and 9 ft in diameter. To establish the desired low vacuum

* This work was jointly supported by The University of Tennessee-Calspan Center for Advanced Space Propulsion under NASA grant NAGW 1195, Boeing Aerospace and Electronics, Seattle, WA and The University of Tennessee Space Institute, Tullahoma, TN.

** Calspan/CASP, Research Engineer
† UTSI/CASP, Research Engineer
‡ UTSI/CASP, Graduate Research Assistant
§ UTSI/CASP/CLA, Professor, ES&M

in the facility, a combination of mechanical pumps and cryopumps is used. Pumping the chamber down from atmospheric pressure to 10^{-3} Torr is achieved with a 150 ft³/min mechanical pump. Cryopumping of the system is achieved primarily by the helium cooled surface located inside the chamber. This cryosurface has an active area of about 36,000 in² (23 m²), and is cooled to about 20 K with a refrigeration unit capable of absorbing, at 20 K, a 900 W heat load. This cryopanel is surrounded with a liquid nitrogen cooled shroud, which is estimated to reach a maximum temperature of about 100 K. The main purpose of this shroud is to shield the heat radiated from the chamber walls.

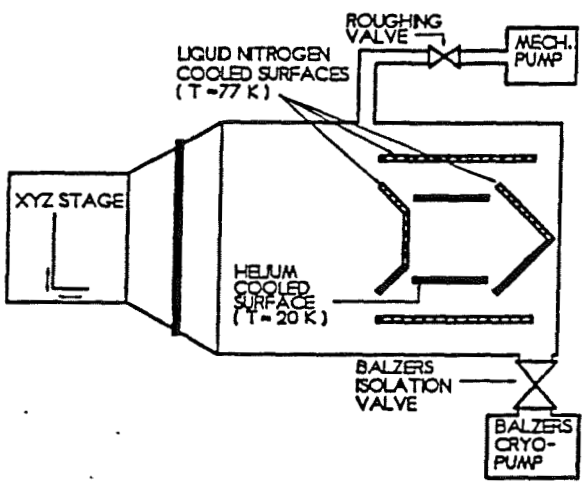


Fig. 1 *Schematic of vacuum chamber.*

To estimate the steady-state pumping capabilities of the cryopanel for a given gas flowing into the chamber, we have used the following equation [3]:

$$P_{ch} = \dot{m} \frac{\sqrt{2 \pi m k T_g}}{c A_c m}, \qquad (1)$$

where P_{ch} is the chamber pressure, $\dot{m}$ is the mass flow rate into the chamber, c is the fraction of gas particles that stick to the cryopanel, A_c is the surface area of the cryopanel, m is the atomic mass of the gas being pumped, k is Boltzmann's constant (1.38×10^{-23} J/K) and T_g is the gas temperature. Using Eq.(1), we calculate chamber pressures as a function of mass flow rate as shown in Fig. 2, where an argon propellant has been assumed. For this calculation, c was taken to be 0.9*0.75, where 0.9 is the sticking probability on the cryopanel [3], and 0.75 is the estimated fraction of cryopanel that will have gas condense on it; m = 6.68 x 10^{-26} kg; and T_g was either 300 K or 77 K as is illustrated in the plot. Although the exact gas temperature for particular operating conditions is yet to be determined, we estimate that it will be between that of the chamber walls (300 K) and that of the liquid nitrogen cooled shroud (77 K) for most applications.

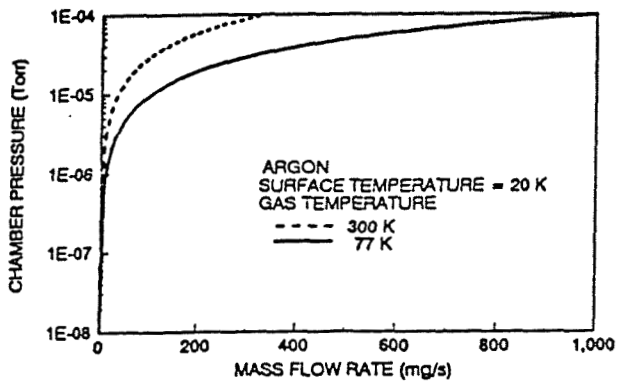


Fig. 2 *Estimated steady-state chamber pressure as a function of flow rate into the chamber.*

One limiting factor of the cryopanel is the heat load which the refrigeration unit can absorb while maintaining a temperature of 20 K. Equation 2 gives an estimate of this heat load Q_{cool} to the cryopanel as a function of the mass flow rate $\dot{m}$ into the chamber [3]:

$$Q_{cool} = \dot{m}[C_{pf}(T_v - T_c) + C_{pg}(T_g - T_v) + cE_s] + Q_{rad}. \qquad (2)$$

Here, C_{pf} is the specific heat for the cryodeposited frost (0.649 kJ/kg-K for argon [3]), C_{pg} is the specific heat for the pumped gas (0.520 kJ/kg-K), T_c is the cryopanel temperature (20 K), T_g is the gas temperature (assumed, here, to be 300 K), and T_v is the vapor sublimation temperature (50 K). Other parameters in Eq.(2) are c as defined above, E_s, the sum of the latent heats of fusion, vaporization and sublimation, for which a value of E_s = 16,300 kJ/kg was used [3], and Q_{rad}, the radiated heat load which was estimated to be 60 W by assuming a black body at 100 K. Figure 3 shows a plot of the required refrigeration power as a function of argon mass flow rate, obtained using Eq.(2). Noting that the heat load limit of the refrigeration unit is 900 W (while maintaining a helium temperature of 20 K), we thus

conclude that the maximum flow rate for these operating conditions is on the order of 60 mg/s for argon.

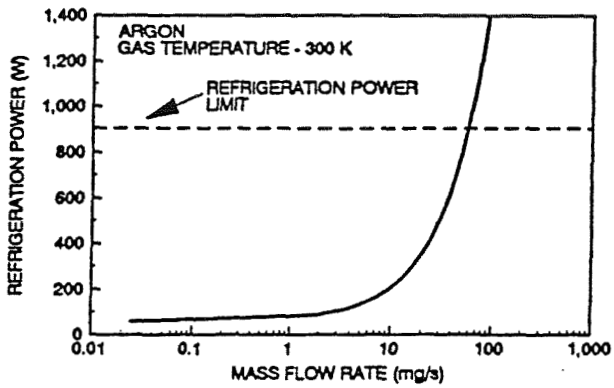


Fig. 3 *Required refrigeration input power versus mass flow rate into the chamber.*

In addition to the helium-cooled cryopump described above, the chamber is equipped with a 16 in. Balzers cryo-adsorption pump, also shown in Fig. 1. This pump is capable of pumping the lighter gases such as hydrogen that do not condense on the helium-cooled cryopanel. The pumping speed for the Balzers cryopump is given in table 1 for different gases. A 16 in. diameter air actuated vacuum valve is used to isolate this pump from the chamber.

Table 1 *Pumping speed of Balzers cryopump.*

	Pumping Speed (l/s)
Argon	5,500
Hydrogen	6,000
Nitrogen	6,500
Water	18,000

A computer program has been written to control and monitor the vacuum system. This program will be used to start the initial pump down of the chamber. When the chamber has reached a specified pressure, the program starts the cool-down of the liquid nitrogen shroud. This allows the user to let the computer start the chamber evacuation and cool-down cycles the night before a test, so that he/she can start the helium refrigeration system early in the morning upon arrival. The program then monitors a range of conditions in the chamber, and outputs these conditions to a computer screen and to a file for future reference.

A three dimensional translation stage (XYZ stage) has been installed inside the chamber as shown in Fig. 1. This stage will have the device of interest, for example, the ion thruster, mounted on it. The user can then position the device from outside the vacuum chamber using a computer control unit for the XYZ stage. This stage has a 12 inch reach in each of the three directions, and has a weight limit of 50 lbs.

III. ION THRUSTER

One of the proposed devices to be tested in the vacuum chamber is an ion thruster. Figure 4 gives a simple schematic of such a thruster, which, generally, is run either with xenon or argon as the propellant gas. The propellant enters the thruster from the left, and is ionized by electrons that are emitted from the (hollow) cathode, and subsequently drawn to the anode, located near the outside diameter of the thruster. These ions are then accelerated from inside the thruster through the grid system. The screen grid is maintained at either cathode or anode potentials. The potential of the accelerator grid is held several hundred volts negative relative to the screen grid. This acceleration process produces a thrust on the order of a fraction of a pound. However, specific impulses in excess of 2000 seconds can be obtained in this manner. The magnets on the inside of the thruster are used to confine the electrons and ions and help maintain the efficiency of the thruster. Since positive ions are extracted from the thruster, a neutralizer is located downstream of the thruster to introduce electrons into the ion beam so as to maintain charge neutrality.

Due to the small thrust of an ion thruster, these devices generally require operation times greater then 10,000 hours to accomplish a given mission, and the useful thruster lifetime becomes an issue of critical importance. For this reason, one important goal of our research is to develop sophisticated diagnostics that can help predict the lifetime of an ion thruster without running extended life tests. The proposed diagnostics are discussed in the following sections. A 30 cm diameter thruster shell, along with the required magnets, two sets of grids, and a cathode have been donated to CASP by NASA Lewis Research Center in Cleveland,

Ohio. The testing of the thruster will be conducted inside the vacuum chamber mentioned above. Necessary power supplies have been purchased, which will allow operation of the thruster with ion beam currents of up to 2 Amps and an electron current from the cathode to the anode of up to 15 Amps.

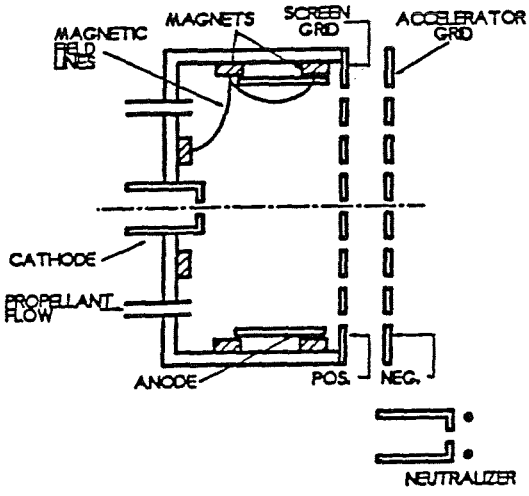


Fig. 4 *Schematic of an ion thruster.*

In order to predict the rise in background pressure in the vacuum chamber due to the operation of the ion thruster, the propellant flux for given operating conditions of the thruster must be known. In this regard, the thruster beam current is determined by the propellant flow through the thruster. This correlation is given by the following equation:

$$J_B \approx \frac{\dot{m}\, e\, \eta}{m}, \qquad (3)$$

where J_B is the beam current, $\dot{m}$ is the propellant flow rate, m is the mass of the propellant, e is the electronic charge, and η is the total propellant utilization efficiency. By rearranging Eq.(1), and incorporating it into Eq.(3), we then find the maximum allowable beam current for a given chamber pressure:

$$J_B \approx \frac{P_{ch}\, c\, A_c\, e\, \eta}{\sqrt{2\,\pi\, m\, k\, T_g}}. \qquad (4)$$

Some numerical results, obtained using Eq.(4), are shown in Fig. 5, which is a plot of the estimated beam current that can be obtained for a given chamber pressure. For this plot, η was taken to be 0.81, c = 0.68, T_g = 300 K and A_c = 0.75 x 23 m^2. Because, for optimal ion thruster operation, a facility pressure of less than 1 x 10^{-5} Torr is desired, Figure 5 implies that the thruster can be operated at beam currents up to 10 Amps.

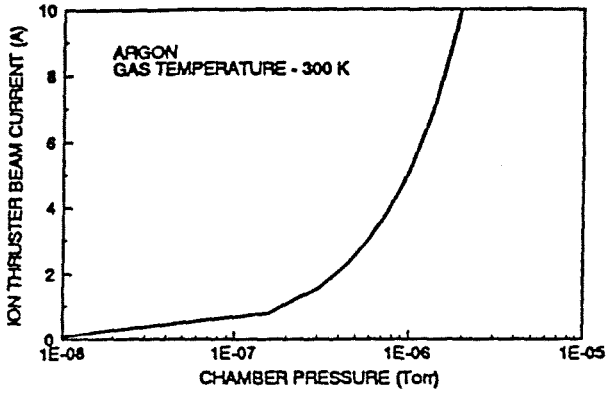


Fig. 5 *Ion thruster beam current that can be obtained as a function of chamber pressure.*

IV. OPTICAL DIAGNOSTICS

There are many candidate optical techniques that can be used to characterize flowfields in electrical propulsion systems. Of these, the primary one that we have singled out is that of laser induced fluorescence (LIF) [4]. In this technique, a laser beam is focused into the region of interest, exciting one of the species in the flow. Subsequently, fluorescence is detected, usually at right angles to the laser beam, by an optical detector, for example, a photo diode. The information that can be extracted from these LIF data includes temperature, density, and velocity components of the absorbing species along the laser beam. Furthermore, the technique is non-intrusive, and is capable of high spatial and temporal resolution. Some of the details of this method are discussed below.

First, Fig. 6 shows a schematic diagram of the experimental setup. A 10 Watt argon ion laser pumps a narrowband Coherent 699-21 ring dye laser, which is frequency stabilized to about 1 MegaHertz, or about 0.00002 Angstroms. This linewidth is much less than any other relevant broadening mechanism (Doppler,

Stark, pressure, power, and natural broadening), so, for all practical purposes, the laser can be regarded as a true single frequency source. The laser beam is then focused into the vacuum facility, and intersects the flow field at some predetermined position. Since the thruster is mounted on an XYZ positioning system, most optical components can be kept fixed, thereby simplifying optical alignment procedures, and the laser beam is scanned across the flowfield by translating the thruster instead. Because of the large size of the vacuum chamber, most critical optical components ate mounted inside the chamber, for example, to achieve optimum collection efficiency for the fluorescence radiation.

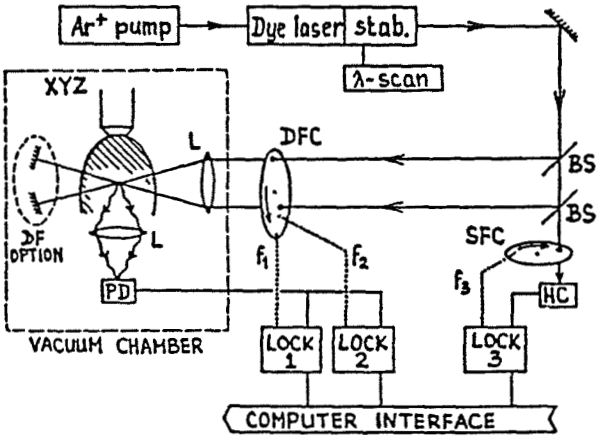


Fig. 6 *Schematic of optical diagnostics arrangement. BS = beam splitter; L = lens; PD = photo diode; HC = hollow cathode lamp; SFC/DFC = single/double frequency chopper, DF option = Doppler free option. LOCK = lockin detector. The thruster is mounted on an XYZ positioning system, while most optical components remain in fixed positions.*

To separate the LIF signal from background and plasma radiation in the flow, the laser beam is chopped, and all signals are measured with lockin amplifiers, which take the chopper frequency as reference (dotted lines in Fig. 6). Thus, great sensitivity is achieved. To measure more than one velocity component, the laser beam is split and recombined at a specified position in the flow.

The different laser paths are chopped at different frequencies (for example, with a dual frequency chopper as in Fig. 6), allowing, through determination of the associated Doppler shifts, the reconstruction of the velocity components of the absorbing state species at the probing point. Although only two velocity components can be extracted with the setup shown in Fig. 6, a three beam geometry could also be realized.

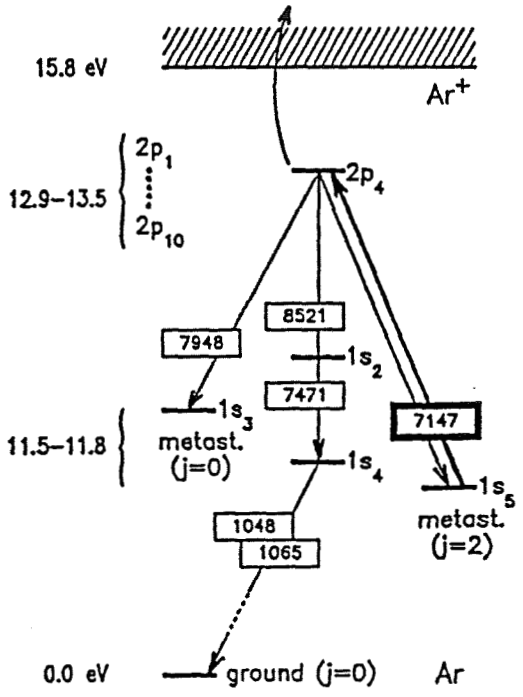


Fig. 7 *Partial energy level diagram of the argon atom, showing ground state, first and second excited states, and ionization continuum. Wavelengths in boxes are given in Angstroms. Excitation at 7147 Angstrom is shown as an example. Collisionally induced ionization from the 2p states plays an important role in the optogalvanic effect, while the fluorescence lines at 7147, 7471, 8521, and 7948 Angstroms form the basis for the LIF measurements.*

Unlike the well known laser Doppler velocimetry system, which relies on scattering of the laser light off of macroscopic particles in the flow, the LIF technique does not require seeding particulates into the flow, but relies instead on the resonant excitation of a specified ground state species naturally occurring in the flow. As an example, Fig. 7 shows part of the relevant energy level diagram of the neutral argon atom, which will be used as a propellant in our initial studies. Because argon, as do other noble gas atoms, has metastable states ($1s_3$ and $1s_5$ in Fig. 7 --- Paschen notation) that can be excited at optical wavelengths, the use of these metastable states as ground states for the LIF scheme

is particularly suited to LIF studies. Figure 7 shows how, for example, the $1s_5$ metastable state can be excited, at a wavelength of 7147 Angstrom, to one of the 2p states ($2p_4$ in Fig. 7), and how this state fluoresces back to the 1p states, and, by cascade emission in the far ultraviolet, to the neutral groundstate. In the LIF application, the fluorescence around 7000 to 8500 Angstroms, in particular, is detected and used to analyze the physical properties of the flowfield.

Because the fluorescence signal, at least for an optically thin plasma, is linearly proportional to the absorbing species density, relative density variations across a region of interest are easily monitored with LIF. More difficult is the extraction of absolute values for the density, which requires careful calibration of the entire optical system. Because the thruster flow is not expected to be in thermal equilibrium, the usual "Boltzmann plot" method is not applicable for the determination of plasma temperatures. Even for a nonequilibrium system, however, temperatures and velocities can, and will be measured, namely by measurements of the Doppler profiles of the absorbing species.

To measure such a Doppler profile, the wavelength of the laser is manually tuned to a particular absorption line of the gas. Subsequently, the laser is locked onto this line, and is frequency stabilized to within 1 MHz. Then, the laser is scanned electronically over a very narrow wavelength range around the linecenter. This range is typically several GigaHertz, sufficient to span the absorption line of interest. An example of such frequency (that is, wavelength) scans is shown in Fig. 8, which shows two simultaneous scans recorded in hollow cathode lamps (with argon buffer gas) at 7273 Angstroms ($1s_4$ to $2p_3$ transition). The two lines are 1.582 and 1.715 GHz wide, respectively, due to the combined effects of Doppler, Stark, natural, and power broadening. Although these measurements are preliminary, they indicate that, even for GigaHertz-wide lines, the linecenters can be determined to MegaHertz precision. For example, in Fig. 8, the two lines are Doppler shifted relative to each other by 10.5 MHz (that is, the shift is almost too small to be obvious by mere inspection of the figure), corresponding to a relative drift velocity of only 7.6 meters/sec. Thus, the capability for measuring velocity components in an electrical thruster to meters per second precision is established.

Actually, the data in Fig. 8 were obtained not with an LIF scheme, but by measuring the voltage change across the hollow cathode lamps as the laser is tuned through the absorption line. This method, known as optogalvanic spectroscopy [5], is very similar to LIF, except that, instead of measuring a fluorescence signal, one essentially measures an ionization signal. The mechanism behind this effect is illustrated schematically in Fig. 7: even though the final state of the laser absorption process is a neutral one ($2p_4$ in Fig. 7), collisional coupling with the plasma environment causes ionization of the gas (requiring only 2.5 to 3 eV), thereby changing the impedance of the discharge. Although optogalvanic spectroscopy is not, in itself, suitable for determination of the thruster flow parameters, it provides a very convenient reference for calibration of the velocity scale with regard to the Doppler profiles obtained with the LIF scheme. Indeed, although the linecenter of a GigaHertz wide line can be determined to MegaHertz precision, absolute determination of its center requires an absolute wavelength standard. Simultaneous measurement of the optogalvanic signal in a hollow cathode lamp provides just this standard.

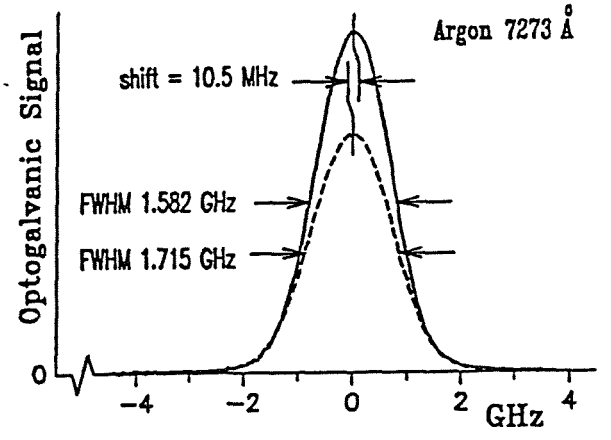


Fig. 8 *Simultaneous scans of the optogalvanic signal in copper (dashed curve) and uranium (solid curve) hollow cathode lamps with argon as buffer gas, revealing a relative Doppler shift of only 10.5 MHz in the presence of GigaHertz broadening.*

Figure 6 shows how this reference measurement is incorporated into the LIF scheme. Part of the laser beam is split off, chopped separately, and focused into a hollow cathode lamp, as the actual LIF measurement

on the thruster flow field takes place as described above. The optogalvanic signal from the hollow cathode lamp is then measured with a third lock-in detector, and is recorded simultaneously with the LIF signals from the other two lock-in detectors (one, two, or three detectors, depending on how many velocity components the LIF scheme is set up for). Thus, data are recorded similar to those in Fig. 8, where one of the profiles is known to be at exact linecenter, namely the optogalvanic one. From the shift of the LIF signal relative to the optogalvanic reference signal one can then determine the absolute velocity component of the absorbing species along the laser beam, and from the width of the Doppler profile one can extract the temperature. Again, it should be mentioned that this measurement can be performed at any position in the flow, thereby characterizing the flowfield with great accuracy.

Although we have described here the general principle of the LIF scheme, many additional details of the technique need to be addressed, for example, the determination of the various broadening mechanisms to the observed, composite, lineshapes. One powerful method, in this regard, is that of Doppler free LIF [6], or Doppler free optogalvanic spectroscopy [7]. Briefly, this technique consists of using overlapping, counterpropagating beams (for example, by reflecting the beams upon themselves, as indicated in Fig. 6). By carefully arranging the experiment, the LIF or optogalvanic signals then contain a component originating only from those atoms in the flowfield or in the lamp that experience equal Doppler shifts from the counterpropagating beams, that is, atoms with zero velocity. In this manner, inhomogeneous Doppler broadening can be eliminated, thereby rendering the purely homogeneously broadened lineshape. From this, we expect to be able to extract the Stark parameters of the line, yielding, among other parameters, values for the local electron density in the plasma [8]. Also, this method provides an independent check on the linecenter frequency calibration described above.

In conclusion, laser induced fluorescence will be used as the key optical diagnostic for spatially resolved measurements of density, temperature, and velocity of the thruster flowfield. Although we have used the example of an argon propellant here, the technique is readily extendable to other inert gases such as helium and xenon, or other species such as hydrogen and nitrogen.

V. ELECTRICAL DIAGNOSTICS

Although the nonintrusive nature of the optical diagnostics from Section IV are unparalleled by most other techniques, not all relevant plasma parameters are readily obtained from LIF data. In particular, the contribution from electrons is rather elusive to optical probes, since typical optical interactions in the visible region of the spectrum take place only in the much heavier neutral and ionized atomic species. Thus, to determine such parameters as plasma potential and electron density and temperature, one must generally rely on electric probes such as Langmuir probes and emissive probes. In addition, Faraday cages and retarding plate analyzers are electric probes that can be used for the determination of ion currents and ion energy distributions. Many excellent review articles on electric probes are available [9,10], and we will limit the discussion here to some of the salient features of these probes, and the extraction of plasma parameters from them.

In their simplest form, an electric probe (specifically, a Langmuir probe) consists of an insulated metal wire (perhaps with a small metallic sphere at the end) that is inserted into a plasma (that is, the thruster flowfield), and is connected to an electrical power supply. As the voltage on the probe is varied, the current to or from the probe is recorded as a function of the probe voltage. This measurement can either be made point by point, or the whole scan can be performed with a supply that is swept through the desired voltage range in a few microseconds. Furthermore, this measurement can be made at any position in the flowfield, by either maneuvering the probe, or, in our setup, by keeping the position of the probe fixed, and scanning the XYZ positioning system of the thruster. Unlike optical diagnostics, electrical probes can significantly modify the local plasma environment, and care must be taken in the interpretation of the data. If properly carried out, however, the details of the measured voltage-current characteristics can be related to the plasma parameters in the absence of the probe.

Figure 9 shows a typical voltage-current trace as obtained with a Langmuir probe. As the voltage V on the probe is varied, three distinct regions for the electron current I are manifest (labeled A, B, and C in Fig. 9). Region A, where the electron current saturates, marks the plasma potential V_p, where the probe voltage is equal to the plasma voltage. Similarly, region C,

where the ion current saturates, marks the floating potential V_f. At this floating potential, no current flows to or from the probe. The intermediate region, region B in Fig. 9, is the retarding field region, from which the electron temperature can be obtained from the slope of the signal in a singly logarithmic plot ("Boltzmann plot"). Finally, the electron density can be obtained from the magnitude of the probe current once the thermal electron velocity has been established.

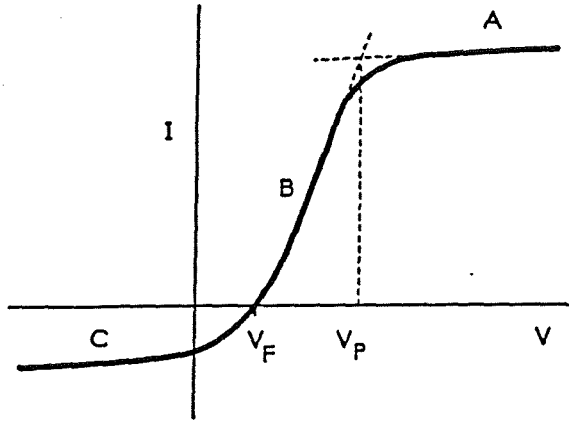


Fig. 9 *Typical voltage (V) - current (I) characteristic obtained with a Langmuir probe.*

An example of the extraction of these plasma parameters from Langmuir probe data is shown in Fig. 10. These data, which are discussed in more detail in Ref. [11], were taken in the downstream region (on axis) of a hollow cathode, electron source similar to those used in ion thrusters (the central leftmost part in Fig. 4 in Section III). In Fig. 10, the effects of different discharge currents on plasma potential, electron temperature, and electron density are all easily observed.

Of course, not all regions in a thruster are easily accessible to either electrical or optical probes. In particular, the sheath regions near the grids fall in this category, and we expect that the most fruitful approach in such regions is the input of experimentally determined upstream and downstream plasma parameters into numerical models, like the one described in the accompanying paper [2].

VI. CONCLUSIONS

A new 9 ft diameter by 20 ft long, cryogenically cooled vacuum facility at the University of Tennessee

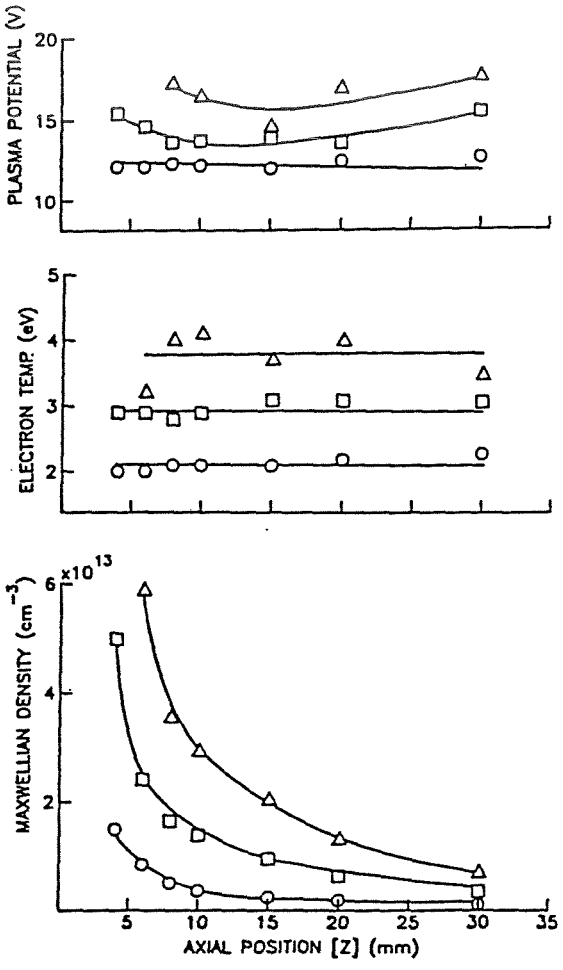


Fig. 10 *Data obtained with a Langmuir probe, on the centerline downstream of a hollow cathode. Circles, squares, and triangles are for discharge currents of 20, 40, and 60 Amps, respectively.*

Space Institute, jointly operated by the Center for Advanced Space Propulsion, has become operational. This facility will provide a baseline pressure of 10^{-7} Torr, and will allow for the testing of such electric propulsion devices as ion thrusters and arcjets at mass flow rates of up to 60 milligrams per second while maintaining a facility background pressure of 10^{-5} Torr. The first test, on a NASA supplied 30 cm diameter ion thruster, will commence shortly.

Apart from the evaluation of general thruster performance parameters such as gas flow and electrical characteristics, advanced optical and electrical diagnostics are under development. Specifically, laser induced fluorescence will be used to perform spatially resolved measurements of neutral gas density,

temperature, and velocity fields, and Langmuir probes will be used to obtain plasma potential contours and electron temperatures. The results of these tests should be instrumental to advance the understanding of the operation of these electric propulsion devices. For example, with precise experimental data on ion flowfield parameters as input for computational studies, a parallel effort on numerical modeling of the ion extraction dynamics in ion thrusters may, for the first time, yield accurate predictions for the grid erosion rates that are not easily obtained with ground testing alone.

ACKNOWLEDGMENTS

The contract monitor for this work was Dr. Jere Meserole of Boeing Aerospace and Electronics. We are grateful to NASA Lewis Research Center of Cleveland, Ohio for donating the ion thruster that will be used in these studies.

REFERENCES

[1] Rawlin, V. K., "Internal Erosion Rates of a 10-kW Xenon Ion Thruster," AIAA Paper 88-2912, AIAA/ASME/SAE/ASEE 24th Joint Propulsion Conference, Boston, MA, July 11-13, 1988.

[2] Peng, X., Ruyten, W. M., Friedly, V. J., and Keefer, D., "Plasma Particle Simulation of Electrostatic Ion Thrusters," this symposium, Tullahoma, Tennessee, November 1990.

[3] Tempelmeyer, K. E., *Heat Transfer at Low Temperatures*, ed. by Frost, W. (Plenum, New York, 1975), pp. 213-227.

[4] Ruyten, W. M., "Aspects of laser induced fluorescence employing pulsed and modulated excitation," Ph.D. thesis, the University of Tennessee, Knoxville, August 1989.

[5] Keller, R. A., and Zalewski, E. F., Appl. Opt. **19**(19), 3301 (1980), and references therein.

[6] Letokhov, V. S. and Chebotayev, V. P., *Nonlinear Laser Spectroscopy* (Springer, Berlin, 1977), Chap. II.

[7] Lawler, J. E., et al., Phys. Rev. Lett. **42**(16), 1046 (1079).

[8] Copley, G. H. and Cam, D. M., J. Quant. Spectr. Rad. Transf., **14**, 899 (1974); see also Griem, H. R., *Plasma Spectroscopy* (McGraw Hill, New York, 1964).

[9] Chen, F. F., in *Plasma Diagnostic Techniques*, ed. by Huddlestone, R. H., and Leonard, S. L. (Academic, New York, 1965), p.113.

[10] Hershkowitz, N., in *Plasma Diagnostics*, Vol. 1, ed. by Auciello, O., and Flamm, D. L. (Academic, New York, 1989), p.113.

[11] Friedly, V. J. and Wilbur, P. J., "High Current Hollow Cathode Phenomena," AIAA paper 90-2587, Orlando, Florida, July 1990; Friedly V. J., "Hollow Cathode Operation at High Discharge Currents," NASA report CR-185238, April 1990.

PLASMA PARTICLE SIMULATION OF ELECTROSTATIC ION THRUSTERS

Xiaohang Peng*, Wilhelmus M. Ruyten**, Verlin J. Friedly†, and Dennis Keefer‡
Center for Advanced Space Propulsion
Tullahoma, Tennessee 37388
(615) 454-9294

ABSTRACT

Charge exchange collisions between beam ions and neutral propellant gas can result in erosion of the accelerator grid surfaces of an ion thruster. Here we report on a combined particle-in-cell (PIC) - Monte Carlo method to simulate these erosion processes. The initial simulation is two-dimensional axisymmetric and uses three velocity components (2d3v) to investigate the influence of charge exchange collisions on the ion sputtering of the accelerator grid surfaces. An example calculation has been performed for operation on xenon. The simulation shows that the greatest sputtering occurs on the downstream surface of the grid, but some sputtering also occurs on the upstream surface as well as on the interior of the grid aperture. Extension to a fully three-dimensional model, taking into account the geometric structure of the grid surfaces, is also discussed.

I. INTRODUCTION

Accelerator grid and discharge baffle erosion may limit the useful lifetime of high specific impulse, high power ion thrusters using rare gas propellants [1]. Since damaging grid erosion occurs over hundreds of hours of operation, it is difficult to predict the useful lifetime of these critical components without expensive, long duration testing. It is thought that the accelerator grid erosion arises from sputtering caused when slow ions are produced by charge exchange. This charge exchange occurs between beam ions which have been accelerated and either neutral propellant atoms or, in the case of ground testing, with atoms of the residual gas in the test facility. The suspection that the observed sputtering rate depends also on the residual gas in the test facility, makes it difficult to interpret the results of ground tests and to predict the lifetime of an engine in space operation.

To better understand these erosion processes, and to help interpret experimental data that will be obtained in a parallel effort (see the accompanying paper [2]), we are currently developing a computational model, based on the PIC (particle in cell) method developed by Birdsall [3], and on Monte Carlo techniques to investigate the accelerator grid erosion produced by charge exchange in the ion beam. The initial, two-dimensional model simulates the charge exchange and sputtering which occur in the axisymmetric region around a single accelerator grid aperture. This provides insight into the details of the charge exchange process. For example, the code correctly predicts the experimentally observed maximum erosion on the downstream side of the accelerator grid. However, a three-dimensional model is required to take into consideration the geometric effects of the, generally hexagonal, array of apertures constituting the screen and accelerator grids. This three-dimensional model is discussed in Section IV. For earlier accounts of this work, see references [4] and [5].

II. SIMULATION MODEL

The PIC method utilizes a large number of test particles whose trajectories are followed in time to simulate a plasma. Each particle carries a representative charge which is distributed on a computational grid. The resulting charge distribution is used with Poisson's equation to calculate the electric field distribution over the computational domain. During the subsequent time step, this electric field is used to determine the acceleration of each of the test particles. The new positions of the particles in the computational grid are then calculated, the charge is redistributed over the grid and the process repeated.

Our two-dimensions-in-space and three-dimensions-in-velocity model simulates the plasma in an axisymmetric region near a single set of screen- and accelerator grid apertures. Upstream of these apertures the plasma is neutral at the discharge chamber conditions. There is an extraction sheath that forms upstream of the screen grid which repels electrons. The ions extracted from the plasma are accelerated and pass through the accelerator grid reaching zero potential somewhere downstream of the accelerator grid surface. A schematic representation of the computational domain is shown in Figure 1. The applied potentials are typical of those expected in a xenon thruster.

---

* Graduate Research Assistant
** Calspan/CASP, Research Engineer
† UTSI/CASP, Research Engineer
‡ UTSI/CLA/CASP, Professor of ES&M

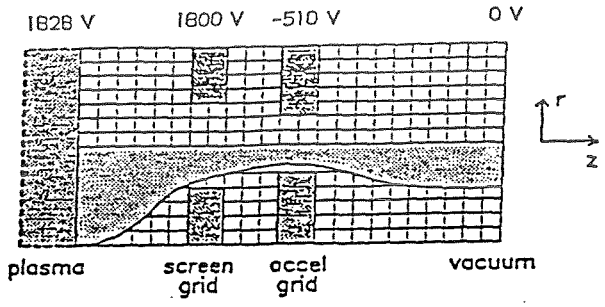


Figure 1: *Schematic of the computational domain used for the simulation of a single ion thruster aperture.*

Within the grid aperture region of an ion thruster few electrons are present since they are repelled by the negative potential applied to the grids. However, in the ion extraction sheath formed upstream of the screen grid, the sheath structure is determined by the electron density. In our model the electron density is not determined by tracking individual electron test particles but, rather, by assuming that the electron density is determined by the electric potential according to

$$n_e = n_0 \exp\left[-\frac{|e|}{kT_e}(\phi_0 - \phi)\right] \qquad (1)$$

where n_e is the electron density, n_0 is the ion density in the neutral plasma, e is the electron charge, k is Boltzmann's constant, T_e is the electron temperature, ϕ_0 is the plasma potential and ϕ is the local potential.

Once the ions reach the grid region, the electron density is small and the electric field in the acceleration region is determined by the ion dynamics. This assumption permits larger time steps to be utilized in the computation since the plasma frequency no longer sets the minimum time step as it would if the electrons were included as simulation particles. When collisions are neglected, the PIC method provides an efficient way to calculate the dynamics of the ion motion through the grids. Downstream of the accelerator grids, the ion beam is neutralized by injection of electrons into the beam. We have not yet attempted to include this process in our simulation.

Collisions between the accelerated ions and the neutral propellant gas, which also passes through the grid apertures, results in charge exchange. The result of a charge exchange collision is a neutral particle moving with the beam velocity and an ion having a thermal velocity. The slow ions created by this charge exchange are then accelerated toward the grids by the applied potentials and can strike the grid surfaces with sufficient energy to cause erosion by sputtering of the grid material.

We have used a Monte Carlo method to include the effects of charge exchange collisions in the PIC code. At the present time we have assumed a constant number density for the neutral particles in the computational domain. The Knudsen number for this flow is typically greater than 10 and a more exact simulation will require a computation of this rarefied flow. The Direct Simulation Monte Carlo (DSMC) method of Bird [6] has matured in the past few years [7], and might be used to more accurately model the neutral flow in the simulation.

The simulation procedure begins by first calculating the charge distribution on the computational domain using a linear distribution of the charge carried by each test particle onto its neighboring grid points. Next, the electric potential is determined by solving Poisson's equation using this charge distribution, and the electric field is determined from the calculated electric potential. The particles are then allowed to accelerate and move under the influence of the electric field for a prescribed time increment. During this time the test particles may experience charge exchange collisions which produce a fast neutral test particle and a new slow ion test particle. The new slow ion is also allowed to accelerate and move under the influence of the electric field for the duration of the time step, and the fast neutral atom is followed until it exits the computational domain.

Since the cross-section for the charge exchange process depends on the ion energy, the collision frequency is not constant, and we have employed the null collision concept [8] to simulate the collisions. For a constant collision frequency the probability of a free-flight time between t and $t + dt$ is,

$$P(t)dt = \nu e^{-\nu t} dt \qquad (2)$$

where ν is the collision frequency. Its cumulative distribution function is,

$$C(t) = \int_0^t P(t')dt' = 1 - e^{-\nu t} \qquad (3)$$

If r is a random number between 0 and 1 chosen from a distribution that is uniform, then the corresponding free time is given by

$$t = -\nu^{-1} \ln(1 - r) \qquad (4)$$

When the collision frequency is not constant, it would appear that the simple expression of Equation 4 could not be used. However, Skullerud [8] introduced the concept of a null collision process to permit the use of this simple expression. In a null collision no charge exchange or other change takes place. The null collision frequency at any energy is chosen so the sum of the real collision frequency and the null collision frequency adds to a constant collision frequency ν,

$$\nu_{\text{null}} = \nu - \nu_{\text{real}} \qquad (5)$$

This provides a distribution of collisions between real collisions and null collisions as shown in Figure 2.

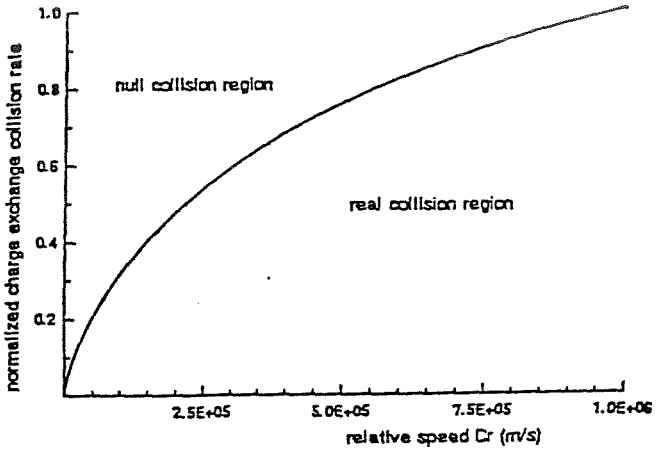


Figure 2: *The charge exchange collision rate as a function of the relative speed Cr.*

The procedure is to choose a random number which determines a free flight time according to Equation 4. The particle is then allowed to move during this time under the influence of the electric field to a new position, and then a second random number r_2 is chosen. If $r_2 > \nu_{real}/\nu$ then the event happens in the null collision regime (Figure 2) and no change occurs. Otherwise, the collision is counted as a real charge exchange and a new slow ion is added to the test particles at this location. This process is repeated until the computational time increment has been reached. Then the process is repeated using the current positions of the particles to determine the charge distribution.

In our simulation ions and neutral particles are continually being lost through the downstream end of the computational domain and must be replaced by ions entering through the upstream end. At each time step these new particles are distributed at random positions into the first grid cells to produce a uniform density in the cylindrical domain. The particle velocities are randomly chosen according to a Maxwellian distribution at the ion temperature with a mean extraction velocity in the axial direction. The mean extraction velocity and the initial ion density are chosen to provide the desired beam current.

III. EXAMPLE COMPUTATION

The simulation model has been used to calculate the ion flow, charge exchange collisions and sputtering in the region of a single aperture in the screen and accelerator grids. The computational domain is shown in Figure 3 and was chosen to correspond to the experimental configuration used by Rawlin [1] to study the erosion rates in a xenon thruster. It is assumed that a neutral plasma exists at the upstream boundary of the computational domain which has an ion density of $1 \times 10^{18} m^{-3}$, an ion temperature of 1000 K and an electron temperature of 1.5 eV. The plasma potential is assumed to be at 1828 V, the screen grid is at a potential of 1800 V, the accelerator grid is at a potential of -510 V and the downstream boundary is at a potential of 0 V. The neutral gas density is assumed uniform throughout the computational domain at a density of $1 \times 10^{19} m^{-3}$. Thus it exceeds the reported test facility pressure of 1.4×10^{-4} Torr in Rawlin's work, which corresponds to a number density of approximately $5 \times 10^{18} m^{-3}$.

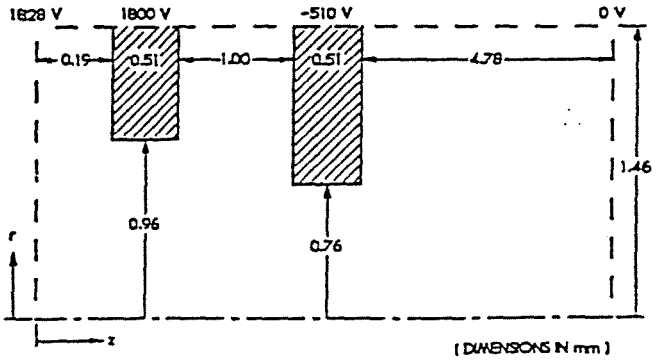


Figure 3: *Computational domain used for the example computation. Dimensions are in millimeters.*

Calculations were performed both with and without charge exchange collisions. Figure 4 shows a comparison of the simulated electric potentials. Charge exchange collisions cause little change in the calculated electric fields. This is because the number of slow ions created by charge exchange is small compared with the number of fast ions.

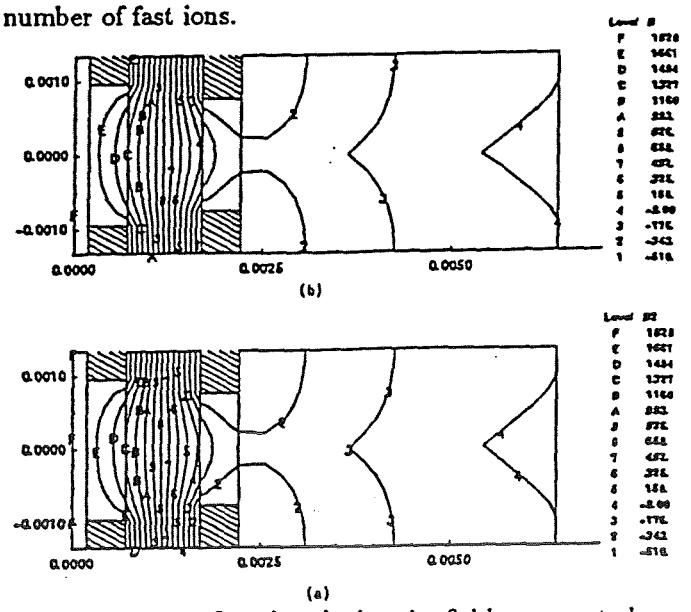


Figure 4: *Simulated electric fields computed without (a) and with (b) charge exchange collisions.*

The logarithm of the number density is shown in contour plots in Figure 5. The difference between the two cases is most pronounced near the accelerator grid because of the slow charge exchange ions which are being accelerated toward the grids. The difference in the average ion velocity between the two cases is significant, as shown in Figure 6. In this figure the values are given only at the computational grid points. Figure 7 shows the potential contours and velocities of the slow ions in the region just downstream of the accelerator grid. These are the instantaneous velocities after 43.5 microseconds of simulation. It can be seen from Figure 6b and Figure 7 that the slow ions produced downstream of the accelerator are accelerated back towards the accelerator grid, and may strike it with sufficient energy to produce significant sputtering and erosion.

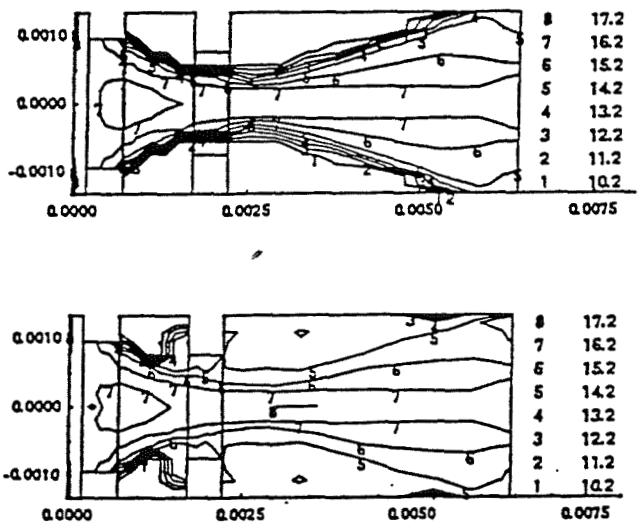


Figure 5: *Log_{10} ion density contours computed from the simulation without (a) and with (b) charge exchange collisions. Maximum density is along the axis.*

The calculated collision rate is shown in Figure 8. Most of the charge exchange collisions occur near the centerline and more collisions occur further downstream. This is because the collision rate depends upon both the collision cross-section and the relative velocity, and both the cross-section and the relative velocity increase as the ions are accelerated through the grids. The calculated sputtering rate for the accelerator grid surfaces is shown in Figure 9. The largest sputtering rate occurs on the downstream side of the grid as has been observed experimentally. The sputtering yield for xenon atoms on molybdenum was taken from reference [9]. The values shown are per square meter of surface and are for a time period of 43.5 microseconds. A smaller amount of sputtering occurs on the interior of the aperture and the upstream side of the grid. This sputtering is caused by the slow ions created near the center of the beam which are then accelerated back into the interior.

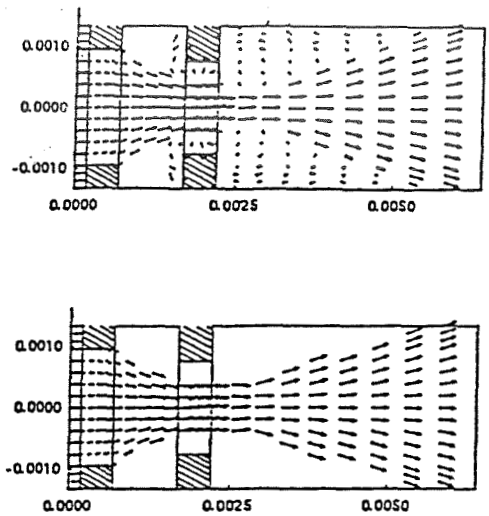


Figure 6: *Average ion velocity computed from the simulation without (a) and with (b) charge exchange collisions.*

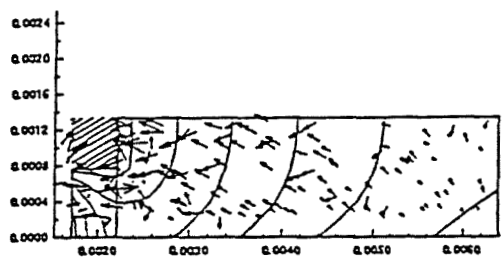


Figure 7: *Electric potential contours and velocities of the slow ions created by charge exchange at 43.5 μs in the simulation. The region shown is downstream of the accelerator grid.*

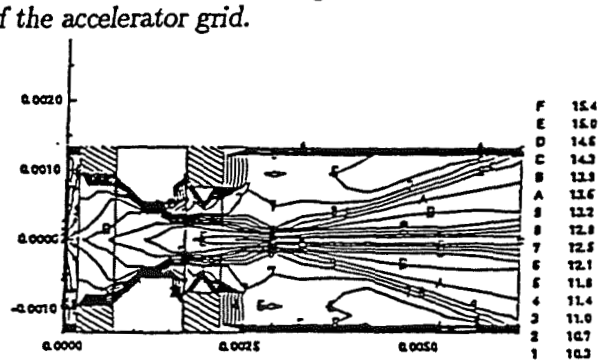


Figure 8: *Log_{10} of the charge exchange collision rate $[m^{-3}s^{-1}]$ calculated from the simulation.*

IV. THREE DIMENSIONAL MODEL

Since actual extraction grid systems consist of arrays with many adjacent apertures, the two-dimensional model with a single set of apertures is necessarily limited in predicting three-dimensional effects of the system. Such three-dimensional effects are, however, of significant importance for lifetime analyses of the grids. Specifically, experimental observations of

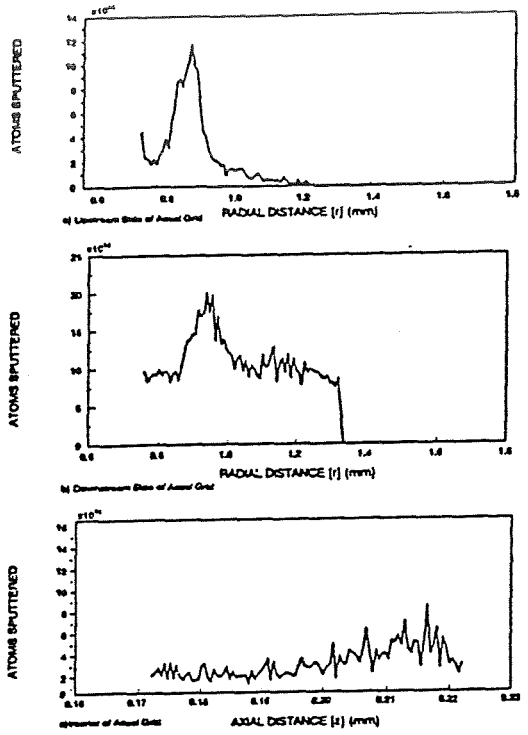
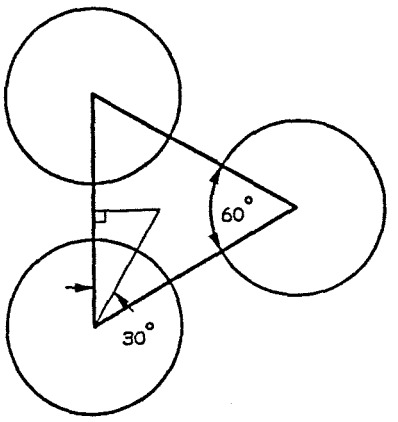


Figure 9: *Sputtering from accelerator grid surfaces. Sputtered atoms per square meter of surface during a simulation time of 43.5 μs.*

grid erosion show that the maximum erosion occurs at the geometric center of a group of adjacent apertures [1]. The electric fields produced by such an array will tend to focus slow ions created in the downstream beam toward the center of the aperture array, increasing the erosion rate at these geometric centers between apertures. Although the results of the two-dimensional model from Figure 9 already give a strong indication that this focusing of slow ions toward the grid material between apertures should occur, a three-dimensional simulation model is required to accurately simulate this effect. Such a three-dimensional model is presently under development, and is discussed briefly below.

Basically, the same steps are required in the 3D calculation as in the 2D model described in Section II. Specifically, the PIC method and the modeling of charge exchange collisions remains unchanged. The primary difference is that the computational grid is now 3D rather than 2D (note that, even in the 2D model, all three velocity components of the particles are taken into consideration). However, rather than modeling an entire grid system, the usual hexagonal symmetry of the aperture arrays is exploited to choose as small a 3D domain as is necessary. Figure 10a shows three apertures in such an array, and indicates the 30 degree by 60 degree right triangle that constitutes the minimum cross sectional cell of the 3D domain. Figure10b shows how this cell is subdivided into smaller triangular cells. Because this triangular mesh is taken to be constant along the direction of flow of the ions, the shapes of the triangles in this mesh are chosen such as to accurately model the contours of the screen and accelerator grid apertures, while roughly maintaining the uniformity of the mesh. Figure 11 shows the complete wedge-shaped computational domain for the 3D model, including the screen and accelerator grid apertures.



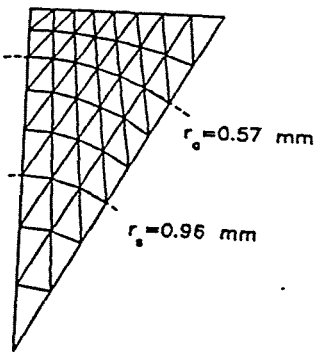


Figure 10: *Minimum cell computational cross-section for a hexagonal aperture grid (a), and nonuniform triangular mesh on this minimum cell (b), for a screen grid open fraction of 0.67.*

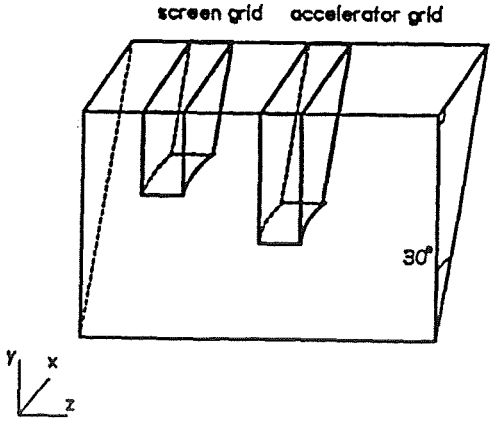


Figure 11: *View of the entire three-dimensional, wedge-shaped computational domain. Ion flow is from left to right.*

As with the 2D model, the 3D geometry is "closed" in the transverse direction. That is, the aperture array is assumed to extend to infinity along the xy plane in Figure 11, and no particles leave the computational region in this direction but, rather, are reflected back at the boundaries of the wedge-shaped domain shown in Figures 10 and 11. Of course, the model is still not yet fully 3D in that edge effects at the circumference of an actual grid system are not accounted for. However, these edge effects are insignificant compared to the fairly uniform erosion across the grids. Indeed, with this 3D model, we expect to be able to predict quantitatively erosion of the grids for various operating conditions of the thruster, and, in particular, the geometric distribution of this erosion around individual apertures.

V. CONCLUSIONS

A simulation of the ion flow through a single aperture in the screen and accelerator grids of a xenon ion engine has been carried out using a PIC and Monte Carlo method. The method provides details of the ion flow in the exhaust beam and the charge exchange collisions which are thought to be responsible for erosion of the accelerator grid. The two-dimensional, axisymmetric model cannot yet be expected to give quantitative agreement with experimental measurements. However, quite satisfactory qualitative features of the ion acceleration process can already be obtained, and an extension of the 2D model to a 3D geometry is under development. A further refinement would be the incorporation of accurate neutral gas density profiles, rather than the use of a constant background density for the entire computational domain.

ACKNOWLEDGMENT

This work was jointly supported by Boeing Aerospace and Electronics, Seattle, WA and the University of Tennessee-Calspan Center for Advanced Space Propulsion under NASA Grant NAGW-1195. The contract monitor was Dr. J. S Meserole.

REFERENCES

[1] Rawlin, V. K., "Internal Erosion Rates of a 10-kW Xenon Ion Thruster," AIAA Paper No. AIAA 88-2912, AIAA/ASME/SAE/ASEE 24th Joint Propulsion Conference, Boston, MA, July 11-13, 1988.

[2] Ruyten, W. M., Friedly, V. J., Peng, X., and Keefer, D., "Experimental Facilities for Electric Propulsion Testing," this symposium, previous paper, Tullahoma, Tennessee, November 1990.

[3] Birdsall, C. K. and Langdon, A. B., Plasma Physics Via Computer Simulation, McGraw-Hill, 1985.

[4] Meserole, J. S., Keefer, D., Ruyten, W. M., and Peng, X., "Ion Propulsion," First annual CASP technical symposium, Tullahoma, Tennessee, October, 1989.

[5] Peng, X., Keefer, D., and Ruyten, W. M.,"Plasma Particle Simulation of Electrostatic Ion Thrusters," AIAA paper 90-2647, Orlando, Florida, July 1990.

[6] Bird, G. A., Molecular Gas Dynamics, (Oxford University Press, London, 1976).

[7] Bird, G. A., "Nonequilibrium Radiation During Re-entry at 10 km/sec," AIAA Paper No. AIAA 87-1543, 1987.

[8] Skullerud, H. R., "The Stochastic Computer Simulation of Ion Motion in a Gas Subjected to a Constant Electric Field," Brit. J. Appl. Phys. (J. Phys. D.), Ser. 2, Vol. 1, p. 1567, 1968.

[9] *Sputtering Yield Data in the* $100 - 600$ *eV Energy Range*, General Mills Report 2309, July 15, 1962.

ARTIFICIAL INTELLIGENCE

Identification and Detection of Anomalies Through SSME Data Analysis

Lisa Pereira
Moonis Ali

Learning and Diagnosing Faults Using Neural Networks

Bruce A. Whitehead
Earl L. Kiech
Moonis Ali

Machine Learning of Fault Characteristics from Rocket Engine Simulation Data

Min Ke
Moonis Ali

Development of an Intelligent Hypertext Manual for the Space Shuttle Hazardous Gas Detection System

Ching F. Lo
George Z. Shi
Eric Cegielski

Engineering Monitoring Expert Systems Developer

Ching F. Lo
George Z. Shi

N96- 70674

IDENTIFICATION AND DETECTION OF ANOMALIES THROUGH SSME DATA ANALYSIS

Lisa Pereira
Moonis Ali

Center For Advanced Space Propulsion
The University of Tennessee Space Institute
Tullahoma, TN 37388

Abstract

The goal of the ongoing research described in this paper is to analyze real-time ground test data in order to identify patterns associated with the anomalous engine behavior, and on the basis of this analysis to develop an expert system which detects anomalous engine behavior in the early stages of fault development. A prototype of the expert system has been developed and tested on the high frequency data of two SSME tests, namely Test #901–0516 and Test #904–044. The comparison of our results with the post-test analyses indicates that the expert system detected the presence of the anomalies in a significantly early stage of fault development.

I. Introduction

A complex physical system, such as the Space Shuttle Main Engine (SSME), is subject to component failures at any time during its testing or operation. When a fault does occur, it is most likely that its effects will lead to damages in its components, possibly including partial or total destruction of the SSME. While testing the SSME, engineers are not able to see the onset of a fault, so when the fault fully develops, it could cause either early engine shutdown or serious damage to the SSME. Worse scenarios could happen if a slowly occurring fault goes undetected throughout all of the SSME's ground and flight tests, and once the engine is inserted into the shuttle for an actual flight, the fault fully develops, causing catastrophic damage to the shuttle and the crew.

When a fault does occur, extensive post-test data analysis is performed by examining a set of sensor plots from different time slices in order to determine a fault's behavioral characteristics, such as the time the sensor data started to indicate the faulty behavior and how the sensor data showed the fault's trend throughout its development. Many times some of the sensors start indicating abnormal SSME behavior considerably before the fault is noticeable by the engineers and before the parameters set for the redline criteria are met, especially if it is a slowly occurring fault. If the engineers or flight crew have an early indication of the developing anomalous engine behavior, preventative measures can be taken in order to avoid or minimize the consequences of the fault.

Automating the sensor data analysis process performed by engineers will result in an online detection system that can discover faults in a physical system during their early development stages by noticing the behavioral changes in all of the sensors' data. Such a system can be a useful tool in aiding engineers during a test, since it would warn them about the onset of anomalies occurring in the SSME during the test as opposed to afterward when a fault might cause early engine shutdown and damage to the SSME.

When analyzing the sensor data, engineers integrate the results from several sensors in order to come up with a more substantial hypothesis or conclusion about what has occurred with the SSME. By integrating the results from all of the generated sensor hypotheses, a detection expert system can provide a better and more precise indication of the health of a rocket engine, thereby diminishing the possibility of false alarms located generated by noise found in the data or by sensors located away from the monitored component.

Many diagnostic expert systems have been developed for rocket and jet engine domains [1-8]. The following is a review of the different approaches taken by researchers in solving fault detection and diagnostic problems.

The system described in [2] uses explicit knowledge representation and explicit reasoning to detect and diagnose faults in a jet engine. The detection of faults is based on stored explicit knowledge incorporated into each node in the domain dependent diagnostic tree. If a certain symptom matches a node in the tree, then reasoning rules, which are specific to a given node, are applied in order to guide the traversal through the fault diagnosis tree.

This research was supported by the System Dynamics Laboratory of NASA–MSFC Grant No. NAG8–121.

Rule-based diagnostic systems were used by the systems in [5,8]. In the system described in [5], the detection of faults is done by looking for abnormalities based on analytical redundancy contained in the Kalman Filter. The diagnostic process involves two approaches, one in which the origin of a failure is determined by the elements most likely to have caused the problem, and the other in which the mathematical modeling corresponds to aircraft configuration changes due to the origins of a fault. Rules are used for transforming the scheduled tasks and failure accommodation into a search problem, which schedules and selects the actions taken by the system. In the second system [8] rules are used to actually perform the diagnosis of a fault. It has two approaches in diagnosing faults. The first approach involves fault modeling based on the operation of the engine; it identifies the problems and determines the fault following trouble-shooting procedures. In the second approach, a qualitative model is used to generate fault hypothesis; the system then chooses one of the hypothesis and looks into the physical layout in order to infer which problems could occur. If the problems do not match the symptoms determined by the system, then that given hypothesis is eliminated.

The system described in [7] uses qualitative and temporal data to perform diagnostics in rocket engine data. Comparisons between the input and previously seen data are executed by the reasoning processor, and similarities found are used to diagnose abnormalities found in the sensor data. In this way, the system can detect faults that have not yet exceeded the safety parameters set for the rocket engine.

Another fault diagnostics approach implemented in [1] is a diagnostic system that combines two parallel approaches when detecting and diagnosing anomalous behavior in propulsion systems. In the first approach, the system learns and identifies sensor data behavior patterns, and generates hypothesis based on the behavior of the system. At the same time, the subsystem involved with the second approach processes the sensor data and reasons with the processed data, the design and functional knowledge of the propulsion system, and the knowledge of the principles of physics and mechanics of the propulsion system in order to generate a fault hypothesis. Results from both approaches are then integrated to form a final hypothesis about the propulsion system.

Neural networks have also been used as a method for detecting and diagnosing faults in rocket and jet engines [3,4,6]. They analyze temporal rocket or jet engine data represented in the form of sensor data curves. In [3] faults are detected and diagnosed by matching an incoming curve with known stored patterns with which the neural network has been trained, whereas in [4,6] detection and diagnosis is performed by looking at the activation values of the middle layer nodes.

The goal of the ongoing research described in this paper is to analyze real-time ground test data, to identify patterns associated with the anomalous engine behavior, and on the basis of this analysis to develop an "Identification and Detection Expert System (IDES) which detects anomalous engine behavior in the early stages of fault development significantly earlier than the indication provided by the redline detection mechanism. A prototype of IDES has been developed and tested on the high frequency data of the two tests where anomalous behavior of a High Pressure Oxidizer Turbo-Pump (HPOTP) was the cause of the fault. IDES's detection approach is based on the methodology applied by rotordynamics experts when they analyze the post-test high frequency sensor data. The system is designed to look for any kind of anomalies present in the monitored information found in the sensor data. The system also integrates each sensor's information into a single hypothesis about what the sensor sees as the behavior of the HPOTP, and then all of the generated sensor hypothesis are integrated in order to determine whether there is an actual fault occurring, or whether one of the sensors is just picking up feed through frequencies from other components.

In order to detect HPOTP faults more accurately and in their very early developing stages, the Isolator and Weld Strain Gage sensors were selected for monitoring. These sensors have been determined by the experts to be the best indicators of HPOTP faults because they are the closest sensors to the source of the problems we have analyzed in the two tests. The Isolator Strain Gages, when present in the HPOTP, are internal to the pump, while the Weld Strain Gages are located on the outside casing of the HPOTP.

It should be noted that the sensor data is not in a steady state from the beginning to the end because of scheduled events that have an impact on the data, causing sensor data to become transient for a while. In order to deal with transient state sensor data, the system utilizes knowledge about the scheduled thrust level changes in order to determine if the sensor data should be sampled and processed. If the system determines that the current time slice falls within the non-transient monitoring time period, it allows for the sensor data to be sampled and analyzed; otherwise, the system ignores the sensor data for that time slice and waits for the next possible processing time slice. The system also keeps information on the scheduled vehicle commands and informs the user whenever a scheduled event occurs during the safe monitoring period, so that if an anomaly is detected, the user is made aware that it could have been caused by the scheduled event.

The system also monitors and identifies intermittent frequencies found in a sensor's window of data. This allows for the system to warn engineers about frequencies that are significantly appearing in the sensor data, possibly indicating faults in other components of SSME. While some of the intermittent frequencies are known to engineers, others are of an unknown nature, giving engineers more information about HPOTP's behavior. The system also tracks and informs engineers about intermittent frequencies which it has detected in the past for each of the monitored sensors. In this way engineers can see which intermittent frequencies are always appearing in a sensor's data.

II. Expert System Architecture

The architecture of the Identification and Detection Expert System (IDES) is shown in Figure 1. It is comprised of four modules. These modules are: the Monitor, the Frequency Extractor, the Data Analyzer/Fault Detector, and the Sensor Integrator. The Profile of Scheduled Events provides information to the Monitor in order to allow IDES to differentiate anomalies from scheduled events. A user interface is also integrated with IDES in order to facilitate the interaction between the user and the system.

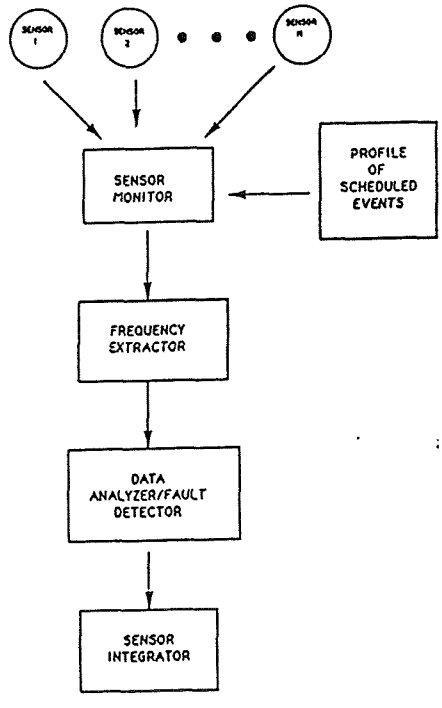


Figure 1. The System's Architecture.

The Monitor

The Monitor is responsible for sampling sensors from sequential windows of 0.4 seconds duration from the non-transient portion of data. The non-transiency of data is determined by utilizing the information provided by the Profile of Scheduled Events. A fast fourier transform of the data of each window is considered for further analysis. The Monitor is also responsible for checking with the Profile of Scheduled Events in order to avoid mistaking a scheduled event for an anomaly. The interface also informs users about the temporal events when a thrust level change command or any other scheduled vehicle command is executed.

The Frequency Detector

This module is responsible for analyzing each window of data and detecting the presence of frequencies of unusually high amplitude. First, attempts are made to identify these frequencies as some of the known frequencies. The frequencies which cannot be identified as known frequencies are treated as intermittent frequencies. Through the user interface, experts/operators are informed about the presence of these abnormal activities in the data.

The Data Analyzer/Fault Detector

The purpose of this module is to perform a comparative analysis of the newly extracted values of the monitored sensor information in order to determine if any of them are detecting anomalous HPOTP behavior. Given each of the extracted values and their respective expected normal values, the Data Analyzer/Fault Detector (DA/FD) calculates an abnormality score for each value, representing the degree by which it is detecting anomalous behavior. Based on the calculated abnormality scores, DA/FD decides whether an anomaly exists in any monitored information of a given sensor. When processing normal data, this module is also made responsible for learning what the expected normal values for the sensor information are.

The Sensor Data Integrator

This module has two components that are responsible for performing sensor data integration. The first component integrates all the available information on a sensor's monitored frequencies and generates a consistency score, which states how consistently any of the monitored sensor information has been detecting anomalies. After each sensor has been processed, the second component then integrates all of the generated sensor consistency hypotheses into a single overall hypothesis about whether the HPOTP's current status is normal or if any of the sensors are showing anomalous behavior.

The User Interface

In order to organize the display of IDES's output and to simplify a user's interaction with IDES, a user interface was integrated with the overall system. The layout for the interface consists of four sensor analysis output display windows and a set of system commands that allow the user to interact with IDES, as shown in Figure 2.

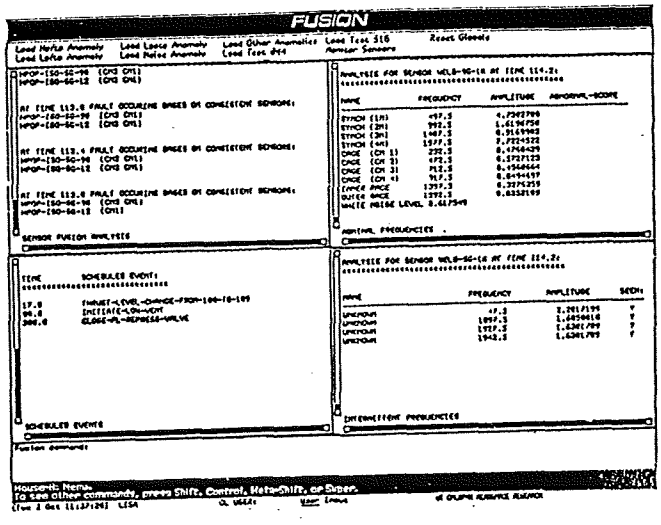


Figure 2. The User Interface.

Each of the windows shows a specific set of information that has either been extracted from the sensor data, supplied by the Profile of Scheduled Events, or deduced during IDES's reasoning process. The Nominal Frequency Window (NFW) and the Intermittent Frequency Window (IFW) display the frequency names, the frequency values, and the frequency amplitudes for all of the frequencies displayed in the respective windows. In the NFW, if a frequency is detecting an anomaly, its abnormality score is also output to the screen. This window also displays the white noise value for each of the sensors. In IFW, if any of the currently extracted frequencies have been seen before by the respective sensor, then it informs the user of that by printing a Y under the Seen column. Any time that a vehicle command is scheduled during IDES's current monitoring time slice, it is displayed in the Scheduled Events Window, along with the previously and next scheduled events. In the Sensor Fusion Analysis Window, the final hypothesis generated by IDES about the HPOTP is displayed for each monitored time slice; the sensors that contributed to the final hypothesis have their names indicated under the hypothesis.

The set of system commands displayed on the top of the interface's screen allows a user to interact with IDES without knowing any of the required parameters and function calls needed to run IDES. These commands provide the user with the flexibility of loading different test cases to be run through IDES. They are activated by either clicking the mouse on the desired option, or by typing a desired command in the command line.

III. Scheduled Events and Monitoring Strategy

IDES maintains a profile of scheduled events. The information contained in this profile is used by the Monitor in order to determine SSME's transient time period between scheduled thrust level changes, the new thrust level, and any scheduled vehicle commands that may affect the amplitude values of the monitored frequencies. The Profile of Scheduled Events is composed of two parts, one which contains the available information on the thrust level events, and the other which contains the available information on all the vehicle commands scheduled for a given SSME test.

A change in thrust level causes a transient state in the sensor data and brings instability to the frequency and amplitude relationship. These instabilities may mislead the system to erroneous frequency analysis. This is especially true if the thrust level change is drastic, as when it changes from 104 to 65 %. In order to avoid erroneous results, IDES employs a scheduled event profile to skip the transient and unstable period of data and analysis data of non-transient and stable durations only.

At each thrust change, the Profile of Scheduled Thrust Level Change Events (PSTLCE) provides the IDES with the SSME's new thrust level, and the stable data monitoring start and end times for the given thrust level. With this information IDES determines the time duration during which it can sample the sensor data.

Although anomalous behavior found in the sensor data usually indicates the development of an anomaly, it could be caused as an after-effect of a vehicle command scheduled to occur at a given time. In order to account for anomalous sensor data behavior, the Profile of Scheduled Vehicle Events provides the user with the current scheduled event, the previously

scheduled event, and the next scheduled event. If anomalous behavior is detected during a time scheduled for a vehicle command, the user is informed that the event may have had an effect on the sensor data, especially if some of the sensors detect anomalies in their monitored frequencies and no fault hypothesis is generated for the next few seconds.

As explained earlier, the monitor is responsible for sampling sensors from sequential windows of 0.4 seconds duration. The frequency spectrum of a window of strain gauge 12 data is given in Figure 3 which illustrates amplitudes of all frequencies between 0 and 2000 Hz. Each frequency in this spectrum represents the mid point of the 5 Hz bin which is used in sampling the data by the monitor.

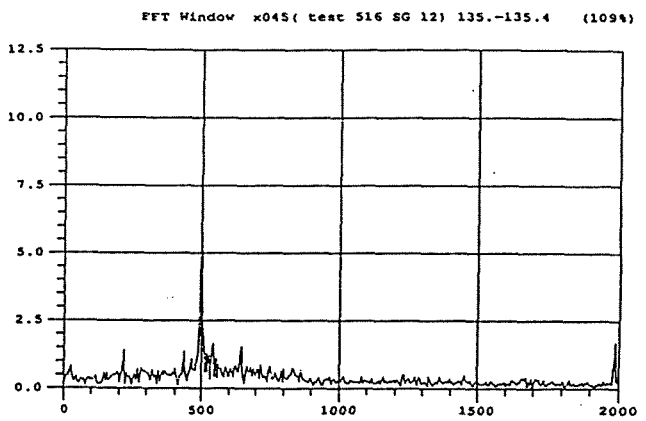


Figure 3. A Window of Sensor Data

IV. The Frequency Extractor (FE)

Human experts employ certain specific frequencies and their harmonics (termed as basic frequencies in this paper) in analyzing anomalous engine behavior. A list of such frequencies is given in Table 1. The presence of high amplitude at any arbitrary frequencies other than basic frequencies (termed as intermittent frequencies in this paper) may also indicate the presence of an anomaly. The Frequency Extractor, composed of two modules, scans each sensor data window and identifies and retrieves the needed frequency information for the analysis of HPOTP's health status. The two modules of FE are the Basic Frequency Extractor and the Intermittent Frequency Extractor, and each is described in the following subsections.

Design of the Basic Frequency Extractor

The Basic Frequency Extractor (BFE) design is

MONITORED SENSOR DATA INFORMATION	
1. FUNDAMENTAL SYNCHRONOUS FREQUENCY	(1N)
2. SECOND HARMONIC OF SYNCHRONOUS FREQUENCY	(2N)
3. THIRD HARMONIC OF SYNCHRONOUS FREQUENCY	(3N)
4. FOURTH HARMONIC OF SYNCHRONOUS FREQUENCY	(4N)
5. FUNDAMENTAL CAGFE FREQUENCY	(1X)
6. SECOND HARMONIC OF CAGE FREQUENCY	(2X)
7. THIRD HARMONIC OF CAGE FREQUENCY	(3X)
8. FOURTH HARMONIC OF CAGE FREQUENCY	(4X)
9. INNER RACE FREQUENCY	(1R)
10. OUTER RACE FREQUENCY	(OR)
11. WHITE NOISE LEVEL	

Table 1. Extracted Sensor Data Information

based on the heuristics applied by the experts when they identify the synchronous and cage fundamental frequencies, their respective harmonics, and the inner and outer race frequencies. Each frequency has a certain range in the data window where it can be found, and it is identified as the highest amplitude peak within that range. If there is not an apparent peak found within the expected frequency range, the given frequency is said to be absent at that current data window, or not significantly appearing in the sensor data.

When analyzing sensor data at different thrust levels, one can notice a shift in the frequency locations as shown in Figure 4. BFE needed a way to determine the necessary frequency ranges so that it can search for the desired frequencies, regardless of SSME's current thrust level. Sensor data was not available for all the possible thrust levels, thus eliminating the possibility of containing a table with frequency ranges for all the thrust levels. Another possibility could have been to keep ranges wide enough so that the desired frequencies could be found no matter what the thrust level is, but this method could produce wrong frequency identifications due to overlapping frequency ranges, such as the fundamental synchronous frequency and the first cage harmonic ranges, and to aliasing of frequencies around a given frequency's range.

This frequency range problem was solved by first manually analyzing several windows of normal data at different thrust levels, extracting the fundamental synchronous frequency from each of the windows, and then fitting the frequency points through a second degree polynomial curve that determines the approximate location of the first synchronous frequency

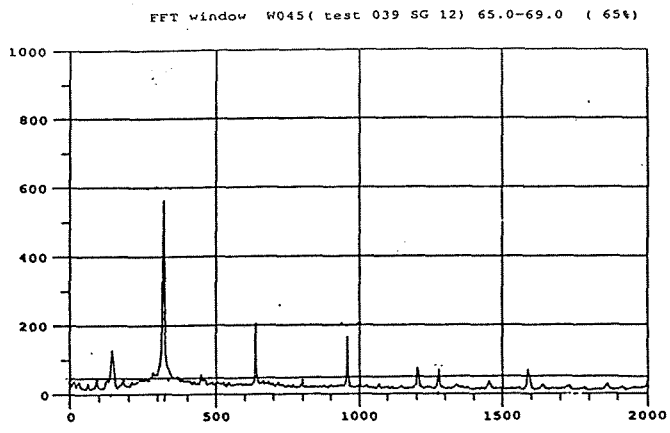
at any given thrust level. In this way, in analyzing a new test data, a shorter range of ± 30Hz is put around the approximate first synchronous frequency location (determined from the second degree polynomial) to insure that if the synchronous frequency has shifted, it can still be found within the expected range, especially since error is introduced by the sampling rate of the data and by the approximation of the interpolated function. Once the fundamental synchronous frequency has been identified for a given thrust level, a range of ± 10 Hz is put around it, creating a shorter range by which BFE must search for the fundamental synchronous frequency in subsequent windows. This process was not repeated for the other frequencies since their ranges are determined by the correct identification of the fundamental synchronous frequency.

In order to determine the frequency ranges for the synchronous harmonics, BFE generates a range of ± 10 Hz around the first synchronous frequency. It then multiplies this range by two, three, and four to get the first, second, and third harmonic ranges respectively.

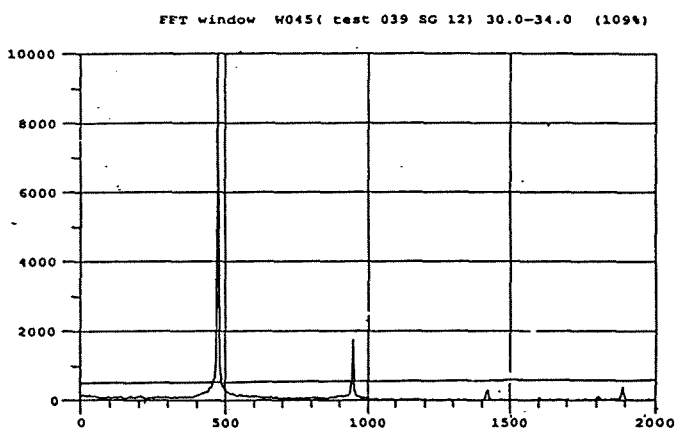
Experts in the field have determined that the fundamental cage frequency is always found within a range of 42-47 % of the fundamental synchronous frequency. The system computes this range and identifies the fundamental cage frequency as the highest peak in that range. Once the first cage frequency is identified, BFE generates a range of ± 10 Hz around it, and multiplies the new range by two, three, and four in order to generate the first, second, and third cage harmonic ranges.

Once all of the harmonic ranges are found, BFE matches the synchronous harmonics to the fundamental synchronous frequency, and the cage harmonics to the fundamental cage frequency. The top three peaks within each of the ranges are identified as possible matches to their respective fundamental frequencies. Out of the three selected peaks, BFE identifies the desired harmonic as the peak whose frequency is the closest multiple to the respective fundamental frequency, given a range of error of ± 15 Hz. In the case where there is not a match between the top three peaks of a harmonic's range with the respective first harmonic frequency, meaning that the given harmonic is not significantly showing in the sensor data, BFE uses the calculated harmonic frequency position and extracts the information for that given harmonic at the calculated frequency.

The reasoning behind looking for the closest multiple instead of an exact multiple is due to the sampling of the high frequency data input to the system, which is set at 5 Hz. A point is picked from each frequency bin as the representative amplitude for the given bin, and the frequency value selected is



A. Sensor Data at 65 Percent Thrust Level.



B. Sensor Data at 109 Percent Thrust Level

Figure 4. Frequency Shift at Different Thrust Level

the halfway point of the bin, making the exact location of a frequency unknown to BFE. The range of ± 15 Hz allows for a harmonic to be found within three bins to the right or left of the expected harmonic frequency value, due to the error introduced by the sampling rate of the sensor data.

Since the inner and outer race frequencies are exactly determined by a window's fundamental synchronous frequency, the cage to synchronous ratio, and the number of balls in a bearing, they do not go through the same identification process as the synchronous and cage frequencies. To find the inner race frequency information, BFE applies the following

formula:

INNER RACE FREQUENCY=1X * NB

The outer race frequency is computed by almost the same formula, except that instead of multiplying the fundamental cage frequency, BFE uses the complement of that value, or one minus the fundamental cage frequency:

OUTER RACE FREQUENCY=1N*(1-(1X/1N))*NB

where

1N = Fundamental Synchronous Frequency

1X = Fundamental Cage Frequency

NB = Number of Balls in a Bearing

Once the frequencies are identified, BFE uses the specific frequency point and extracts the corresponding amplitude value.

The Design of the Intermittent Frequency Extractor (IFE)

Intermittent frequencies are defined as the frequency peaks found in a sensor's data window that cannot be categorized as basic frequencies and are significantly appearing in the data. If any of the frequencies in the data window are selected as intermittent frequencies, they are sent to the Intermittent Frequency Classifier (IFC), either to be identified as one of the possible feed-through frequency signals from the other SSME components, shown in Table 2, or to be classified as unknowns.

KNOWN INTERMITTENT FREQUENCIES
1. 60 Hz LINE
2. HPOTP CASE MODE
3. HPOTP FIRST ROTOR MODE
4. LPOTP SYNCHRONOUS AND HARMONICS
5. LPFTP SYNCHRONOUS AND HARMONICS
6. HPFTP SYNCHRONOUS AND HARMONICS

Table 2. Possible SSME Component Feed-through Frequencies

In determining the intermittent frequencies, IFE employs a threshold significantly above the white noise level. The White Noise Level Extractor (WNLE) determines the white noise level by analyzing a sensor's data window. In order to avoid the monitored frequencies' peaks from influencing the white noise value, they are removed from the data window before the computation; in this way, only the noisy amplitudes contribute to the white noise value. WNLE then averages the remaining amplitude values to determine the white noise level value for a given sensor at each time slice. Because this value is also used as a threshold basis for the extraction of some frequencies, its standard deviation is also computed.

Any frequency whose amplitude is significantly above the white noise level of a sensor's data window is extracted by the Intermittent Frequency Extractor (IFE). In order to distinguish the intermittent frequencies from the noise found in a sensor's current data window, a threshold based on the sensor's white noise level is applied to the remaining frequencies. This threshold is set at three standard deviations above a sensor's current white noise level, and if any frequency's amplitude is above this threshold, it is extracted from the window and sent to the Intermittent Frequency Classifier for possible identification.

The choice for using three standard deviations above the white noise was determined after applying 2 and 3 standard deviations above the noise to the sensor data, and selecting the threshold which best eliminated the noisy frequencies from the ones that are clearly above the noise level, as shown in Figure 5, over a set of data windows.

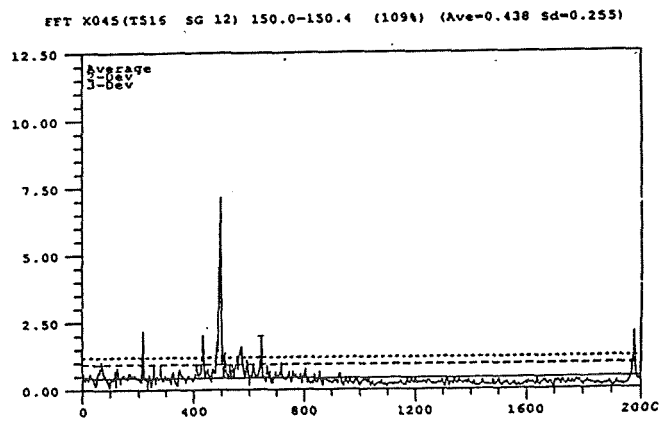


Figure 5. Intermittent Frequency Thresholding Criteria

The Intermittent Frequency Classifier, IFC, attempts to identify the selected frequencies as one of the possible feed-through frequencies that, at times, may appear in the sensor data. Since these frequencies also shift based on SSME's current thrust level, average ratios between the nominal frequencies and the first synchronous frequency were derived from the values used by experts and shown in Table 3, in order to determine the approximate location of the frequency in a data window. These approximate locations are then used by IFC to attempt to match an intermittent frequency as one of the other possible feed-through frequencies.

FREQUENCY NAME	100%	104%	109%
1. LPOTP	84	86	88
2. LPFTP	260	264	273
3. HPOTP	444	470	490
4. HPFTP	572	586	605

Table 3. Feed-Through Frequency Positions at Different Thrusts

The calculated ratios for identifying intermittent frequencies are encoded in a set of data-driven production rules. IFC applies each production rule, computing the approximate ranges for the feed-through frequencies, and attempts to match an incoming intermittent frequency as one of the possible other nominal frequencies. If the incoming frequency does not fit in any of the ranges, then it is classified as an unknown frequency; otherwise, it is identified with the name of the nominal frequency at that range.

Another function IFC performs is to learn intermittent frequencies that it has not seen, and to recognize the ones that it has seen in a previous data window for each of the monitored sensors. From the beginning of a run, IFC starts learning the intermittent frequencies for each of the sensors, and after the first time slice, it starts to recognize previously seen intermittent frequencies. For each intermittent frequency that IFC extracts and recognizes as a previously seen frequency, IFC flags the output information for that frequency as a recognized frequency; otherwise, IFC learns the new frequency for that given sensor so that it can remember it later if the frequency reappears in a different time slice.

V. Training Methodologies

During the development of IDES, several training methods were studied in order to analyze and detect anomalies through sensor data. This section presents and explains each of the approaches that have been tested.

Off-line Training Method

This method involved training the system to distinguish between normal and abnormal sensor data off-line. This was done by using test data which NASA engineers considered normal, and then using that normal data as a basis of comparison with the data from any of the other tests. When comparing the incoming data with the expected normal data in order to detect anomalous frequencies, IDES calculates the number of how many deviations the new frequency amplitude is above the expected normal amplitude.

In order to train the system, the amplitudes for each of the frequencies were extracted from several windows at the different thrust levels present in test 901-039. The mean and standard deviation values were calculated for each of the frequencies at each thrust level. Then, for each of the frequencies, an interpolating function was generated utilizing the different thrust levels and the mean amplitudes at each specific thrust level. In this way, an interpolated mean amplitude can be found for any thrust level, even the ones that were not represented in test 901-039.

Since the abnormality score is based on how many deviations an amplitude value is above the expected interpolated mean, the deviation selected for each frequency was the highest deviation value of that frequency over all of the represented thrust levels. Choosing the highest deviation value for a frequency eliminated the possibility of a normal frequency amplitude being classified as an anomaly in case the selected deviation had been lower than that frequency's deviation at the given thrust level.

In this method the abnormality score was computed by subtracting the interpolated normal amplitude of a frequency from the newly extracted amplitude, and then dividing the result by the highest deviation of that frequency. Anomalous frequencies were detected whenever their deviation number was above a threshold of 5 deviations. Since sensor data have different data scales in different tests (as shown in Figure 6), a scaling factor for each frequency was computed based on the first window of data. In this way, subsequent extracted frequencies' amplitudes were multiplied by their respective scaling factors in order for them to be scaled into representative amplitude values that were comparable to the data scaling of the trained normal amplitudes.

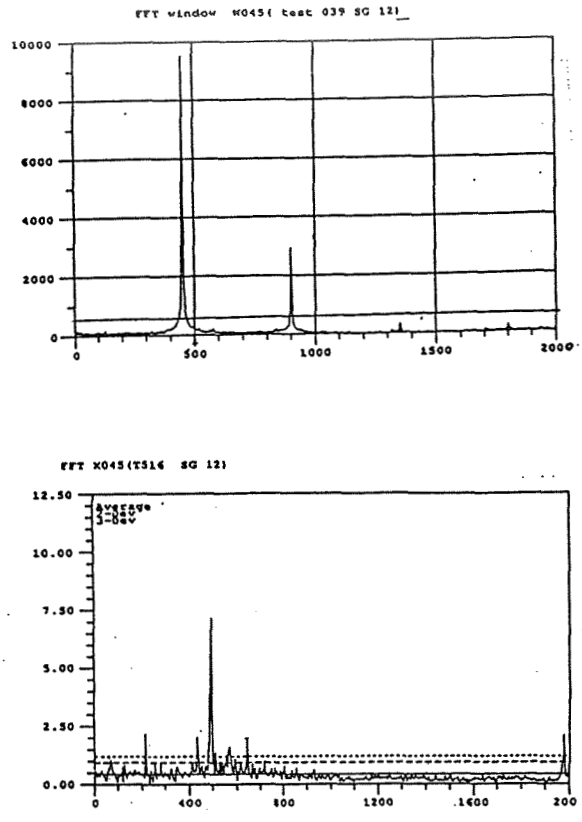


Figure 6. Same Sensor Data From Different Tests.

On-line Training Method

In this approach IDES learned online (from the same test) the expected normal value for each of the frequencies' amplitudes, thus eliminating the need for computing scaling factors between sensor data from different tests. Since the sensor data stabilized within a few seconds after thrust change and can be assumed to be representative of the normal frequency amplitudes, the expected normal values were based on the first window of data, after the sensor data stabilized from a thrust change. Subsequent frequency amplitudes then compared to the amplitudes found in the first data window, and the abnormality scores were determined based on the percentage by which the new amplitude increased from the learned expected normal amplitude. Anomalies were detected whenever a frequency's amplitude exceeded the specified percentage threshold for that specific frequency.

Two improvements were made to the online training algorithm. In the first improvement, the system was modified to compute a running average of the normal amplitude values over a series of data windows. Before adding an amplitude to the running average of a frequency, IDES checked for the presence of a violation. In the absence of a violation the amplitude was added to the running average; otherwise, the running average was not updated for that time slice. The training terminated when either twenty-five windows had been averaged, or when one of the sensors had detected anomalies in three consecutive windows. In the second modification, the algorithm was changed for the inclusion of non-violating amplitudes in the running average. All the frequencies which did not show anomalous behavior during the training time period, were considered normal and were allowed to update their respective running averages.

Since, in this approach, IDES learned online about the expected normal amplitudes for each of the frequencies, it saved the learned information along with the respective thrust level for possible future use during the test. After each thrust level change, IDES readjusted its expected normal amplitude values to reflect the new thrust level. For every previously unobserved thrust, IDES retrained itself for the data of this new thrust. But if the new thrust level had been observed earlier then the system retrieved the trained amplitude information for that thrust level, rather than learning it again.

Modified On-line Training Method

This approach borrowed concepts from the two previous training and detection methods. It utilized a modified detection criteria for the training algorithm to track all of the frequencies' amplitudes that were being included in each frequency's running average. During the training period, the detection of anomalies was still based on the percentage method, but after the system stops training, the detection of anomalous frequencies was switched to a different algorithm. The mean and standard deviation for each of the frequency's amplitudes were computed during the training period, and the running average values were replaced by the sum of the mean and a multiple of the standard deviation of each frequency after the training period. Detecting anomalies in the data then was based on the new frequencies' amplitudes exceeding the threshold set by the sum of each mean with its respective multiple of the standard deviation.

VI. Sensor Integration Module

When looking at sensor data in an attempt to fuse the information provided from several sensors, experts have always emphasized the consistency of a sensor in detecting an anomalous signal as being very important. The sensor integration algorithm developed for this system uses this consistency heuristic as the basis for generating and integrating the sensors's hypotheses.

The Sensor Integration Module is divided into two parts:

1. Single Sensor Integrator
2. Multiple Sensor Integrator

The first module generates a consistency hypothesis about HPOTP's behavior as seen by a single sensor and based on its monitored frequencies. The second module integrates all the hypotheses generated for all of the sensors into a single overall hypothesis about HPOTP's behavior.

Single Sensor Integrator

The Single Sensor Integrator (SSI) is designed to look at all of a sensor's currently extracted information to see if anomalies have been detected, so that it can combine the new results with the previous ones in order to generate a sensor's consistency hypothesis about possible anomalous behavior of HPOTP. The generated hypothesis is based on a sensor's consistency in detecting anomalous behavior within the same monitored sensor information over a period of three sequential windows.

Each sensor is associated with a symptom object that tracks the consistency of anomalous behavior found in the extracted sensor information along with the hypothesis generated for that given time slice. For each extracted sensor information, there is a sliding window that holds knowledge about whether a sensor has detected an anomaly within the last three time periods. At each new time slice, when a sensor data has been processed and the abnormality scores have been generated, each sliding window is updated to either contain a Y, indicating that an anomaly has been detected at that time for that given sensor information, or an N, indicating that no anomaly has been identified for that given piece of sensor information. In addition, the oldest element found inside each sliding window is removed from it.

Once all of the sliding windows of a sensor's symptom object have been updated, SSI inspects them, giving each sensor a consistency score of either three, two, or zero. As shown in Table 4, at 100.6 seconds a sensor gets a consistency score of three when one of its monitored frequencies has consistently shown an anomaly during three consecutive time slices; in this case, the fundamental synchronous frequency generated the score of three. If a sensor does not receive a consistency score of three, then a consistency score of two is given if one of the frequencies has detected an anomaly in two out of the three tracked time slices, as shown in Table 4 at 98.2 and 98.6 seconds. Otherwise, if a sensor has not received a score of either three or two, then it is assigned a score of zero, indicating that either no anomalies have been detected by the sliding windows at 97.0 seconds in Table 4, or that there is only one anomaly detected in any of the sliding windows at 97.4 seconds in Table 4.

TIME	1N	2N	3N	4N	1X	2X	3X	4X	1R	0R
96.6										
97.0										
97.4	X									
97.8										
98.2	X									
98.6	X									
99.0										
99.4										
99.8	X									
100.2	X									
100.6	X									

Table 4. A Sensor's Data Behavior.

The Multiple Sensor Integrator

Once all of the sensors have had their consistency hypothesis generated for a given time slice, the symptom objects are sent to the Multiple Sensor Integrator (MSI) to have the overall hypothesis about HPOTP's health status generated for the current time. MSI generates three possible hypothesis:

1. Possibility of a fault
2. Good possibility of a fault
3. Fault occurring based on N consistent sensors

In order for MSI to generate a hypothesis, it must first look into the consistency score found in the symptom objects of each sensor, and then determine how many sensors have received consistency scores of either three or two. For MSI to assign the first hypothesis (the possibility of a fault) as HPOTP's status, only one of the monitored sensors must have received a consistency score of three, indicating that it has consistently detected anomalous HPOTP behavior within three consecutive time slices. For the second hypothesis (a good possibility of a fault) to be assigned, one of the monitored sensors must have received a consistency score of three, while at least one

other sensor has received a consistency score of two, showing that one sensor has consistently detected anomalous behavior in three consecutive time slices, and at least one other sensor must have detected anomalous behavior in two out of three consecutive time periods. When two or more monitored sensors have received a consistency score of three, indicating that at least two sensors have detected anomalous behavior, the third hypothesis (fault occurring based on N consistent sensors) is assigned as HPOTP's status, where N is the number of consistent sensors showing the fault. No hypotheses are generated in the cases where all the sensors have consistency scores of either two or zero.

Examples of the sensor data conditions in which the three hypotheses are generated can be seen in Table 5. At 107.4 seconds, MSI generates the first hypothesis of a possibility of a fault. In this case, the Weld Strain Gauge 3A is the cause for the hypothesis based on its anomalous second harmonic of the synchronous frequency. An example of when the second hypothesis is generated is at 108.2 seconds, where now the Isolator Strain Gauge 12 has detected anomalous behavior in two out of the three time slices. At 108.6, MSI generates the third hypothesis which says that a fault is occurring based on two consistent sensors: Isolator Strain Gauge 12 and Weld Strain Gauge 3A.

	ISO SG 12	ISO SG 90	WELD SG 1A	WELD SG 3A
TIME	1n 2n 3n 4n 1x 2x 3x 4x	1n 2n 3n 4n 1x 2x 3x 4x	1n 2n 3n 4n 1x 2x 3x 4x	1n 2n 3n 4n 1x 2x 3x 4x
106.6	x	x x	x x	x
107.0			x	x
107.4				x
107.8	x	x		x
108.2	x		x	x
108.6	x x	x	x x	x x
109.0	x x	x x	x x	x
109.4	x x	x x		x x

Table 5. Sensors' Time Slice Outputs.

VII. Conclusion

We have developed a prototype of the Identification and Detection Expert System (IDES) which has been tested on the high frequency data of two SSME tests, namely Test \$901–0516 and Test # 904–044. The comparison of our results with the post-test analysis, performed by human experts, indicates the high potential capabilities of IDES in detecting anomalies significantly earlier than other methods currently being applied. However, the implementation of the prototype is the first step in achieving our goals. The success of IDES must be tested on a number of tests of different faults as well as on the same fault occurring with different severities and speeds. We expect that several modifications will be needed for the successful testing of IDES on the data of a large number of engine tests. Though we have performed preliminary analysis of intermittent frequencies, a large amount of domain knowledge will be needed in order to successfully interpret and employ the analysis of these frequencies to the data of a large number of tests. Furthermore, the identification of patterns and the detection of anomalies in early stage may not be enough. The diagnosis of faults, finding causes and sources of the problem, and determining of the possible corrective actions will be important extensions which we intend to perform in the future.

Acknowledgement

This research was supported by the System Dynamics Laboratory of NASA Marshall Space Flight Center, Grant No. NAG8–121. We thank Pat Vallely, Tom Fox, and Carlyle Smith for their assistance throughout this project and Kewei Wu for assisting in data storage and retrival process.

References

1. Abu-Hakima, S., Davidson, P., Halasz, M. and Phan, S.,"Jet Engine Technical Advisor (JETA)," Proceedings of the Second International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, Tullahoma, TN, pp. 154-159, June 6-9, 1989.

2. Dietz, W. E., Kiech, E. L. and Ali, M., "Classification of Data Patterns using Autoassociative Neural Network Topology," Proceedings of the Second International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, Tullahoma, TN, pp. 1028-1036, June 6-9, 1989.

3. Dietz, W. E., Kiech, E. L. and Ali, M., "Jet and Rocket Engine Fault Diagnosis in Real Time," Journal of Neural Network Computing, vol. 1, no. 1, pp. 5-18, Summer 1989.

4. Dietz, W. E., Kiech, E. L. and Ali, M., "Pattern-Based Fault Diagnosis using Neural Networks," Proceedings of the First International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, Tullahoma, TN, pp. 13-23, June 1-3, 1988.

5. Dietz, W. E. and Ali, M., "Qualitative and Temporal Reasoning in Engine Behavior Analysis and Fault Diagnosis," The Third Conference on Artificial Intelligence for Space Applications, Huntsville, AL, November 2-3, 1987.

6. Forsyth, G. F. and Larking, M. D., "Concept Demonstration of the Use of Interactive Fault Diagnosis and Isolation for TF30 Engines," Proceedings of the Second International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, Tullahoma, TN, pp. 15-20, June 6-9, 1989.

7. Gupta, U. K. and Ali, M., "LEADER-An Integrated Engine Behavior and Design Analysis Based Real-Time Fault Diagnostic Expert System for Space Shuttle Main Engine," Proceedings of the Second International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, Tullahoma, TN, pp. 135-145, June 6-9, 1989.

8. Handelman, D. A. and Stengil, R. F., "Combining Expert System and Analytical Redundancy Concepts for Fault-Tolerant Flight Control," Journal of Guidance, vol. 12, no. 1, pp. 39-45, January-February 1989.

N96-70675

LEARNING AND DIAGNOSING FAULTS USING NEURAL NETWORKS

Bruce A. Whitehead
Earl L. Kiech
Moonis Ali

Center For Advanced Space Propulsion
The University Of Tennessee Space Institute
Tullahoma, TN 37388

Abstract

Neural networks have been employed for learning fault behavior from rocket engine simulator parameters and for diagnosing faults on the basis of the learned behavior. Two problems in applying neural networks to learning and diagnosing faults are (1) the complexity of the sensor data to fault mapping to be modeled by the neural network, which implies difficult and lengthy training procedures; and (2) the lack of sufficient training data to adequately represent the very large number of different types of faults which might occur. Methods are derived and tested in an architecture which addresses these two problems. First, the sensor data to fault mapping is decomposed into three simpler mappings which perform sensor data compression, hypothesis generation, and sensor fusion. Efficient training is performed for each mapping separately. Secondly, the neural network which performs sensor fusion is structured to detect new unknown faults for which training examples were not presented during training. These methods were tested on a task of fault diagnosis by employing rocket engine simulator data. Results indicate that the decomposed neural network architecture can be trained efficiently, can identify faults for which it has been trained, and can detect the occurrence of faults for which it has not been trained.

Introduction

The objective of our research described in this paper is to employ neural networks for *(i)* learning fault behavior from rocket engine simulator parameters perturbed with noise (termed as sensor data in this paper), and *(ii)* diagnosing faults on the basis of the learned behavior. In a complex system (such as a liquid-fuel rocket engine), there are many possible ways in which components of the system may fail. Only a fraction of these possible failures have been observed and are known to human experts. Human experts have not seen all possible instances of all faults and hence cannot describe the features of the faults sufficiently well to make diagnostic decisions.

This state of affairs is problematic for training a neural network to recognize component failures based on sensor data. Neural networks learn to recognize faults by being trained with examples of these faults. They are capable of generalizing from some examples of a fault to other examples of the same fault, but they are not capable of recognizing a new fault for which no training examples have been given. If a neural network is trained to recognize a set of faults and then presented with an example of a completely new fault, it will typically either *(i)* find the closest match to the new example among the previously trained faults, or *(ii)* classify the new example as an interpolative "blend" of previously trained faults.

Neither of these classification strategies is appropriate for recognizing a failure which is different from the classes of failures for which the neural network has been trained. What is needed, rather, is an ability to recognize that the new failure is not an example of any previously trained fault. In other words, the neural network must be capable of recognizing the classes for which it has been trained as well as one additional "unknown" class, even though no training examples are available for this unknown class.

If the weights in a neural network are determined solely by the training examples, then the subsequent behavior of the network is also determined by these training examples. Such a network would only be able to classify new examples on the basis of the examples it has seen, and would not be expected to reliably recognize an "unknown" class. We therefore have developed a neural network architecture in which the weights are not determined solely by the training examples. Instead, the weights are determined partly by expert judgment about the type of classification to be performed, and partly by conventional backpropagation training from examples.

This architecture has been tested in the task of sensor fusion of data from the rocket engine simulator. The purpose of the sensor fusion architecture is

This work was supported in part by NASA Grant No. NAGW-1195 and Rocketdyne Contract No. R04QBZ90-032709.

to classify faults of sensor readings as either *(i)* examples of normal steady-state operation, *(ii)* examples of known classes of component failures, or *(iii)* examples of an unknown class of anomalous behavior.

Sensor data typical of a rocket engine fault (with and without the addition of simulated noise) are depicted in Fig. 1. Four sensors, high pressure fuel turbopump (HPFT) temperature, thrust, chamber coolant valve (CCV) pressure, and main fuel valve (MFV) pressure are monitored during engine operation. The data are normalized with steady-state values for convenience. In the present effort, it is assumed that the engine is operating at some steady-state condition when the fault condition occurs. The fault will be manifested as a deviation of sensor values from the steady-state condition. In the present effort, a time window containing 40 sensor readings spanning four seconds is used.

It is not sufficient for the purposes of diagnosis to simply detect when and whether a deviation from steady-state conditions has occurred; how the deviation is manifested over time is also important. For instance, an observation that a particular sensor parameter is decreasing linearly will likely result in a different diagnosis than that obtained from observing an asymptotic decrease[1]. Therefore, to be effective, a diagnostic system must be responsive to the qualitative (as well as the quantitative) behavior of the engine. The diagnostic process must also exhibit resilience to noise. A noise-corrupted version of the fault is therefore depicted in Fig. 1 for comparison. The 2% noise level means that the standard deviation for a Gaussian distribution of perturbations about the noise-free curves is 0.02. A 2% noise level would be considered excessive for most instrumentation; however, satisfactory operation at this noise level is used as a goal in the present effort.

Neural networks have been employed at UTSI in the past to diagnose the development of fault conditions in jet engines[2-4], using conventional feedforward networks trained with well-known back-propagation algorithms[5]. Although this method was effective in diagnosing faults when given samples of data corresponding to a fault for which the networks were trained, several deficiencies of the conventional feedforward model were noted:

1). Networks can be trained to associate a set of patterns with a set of fault conditions quite readily. However, when presented with an input pattern qualitatively different from those included in the

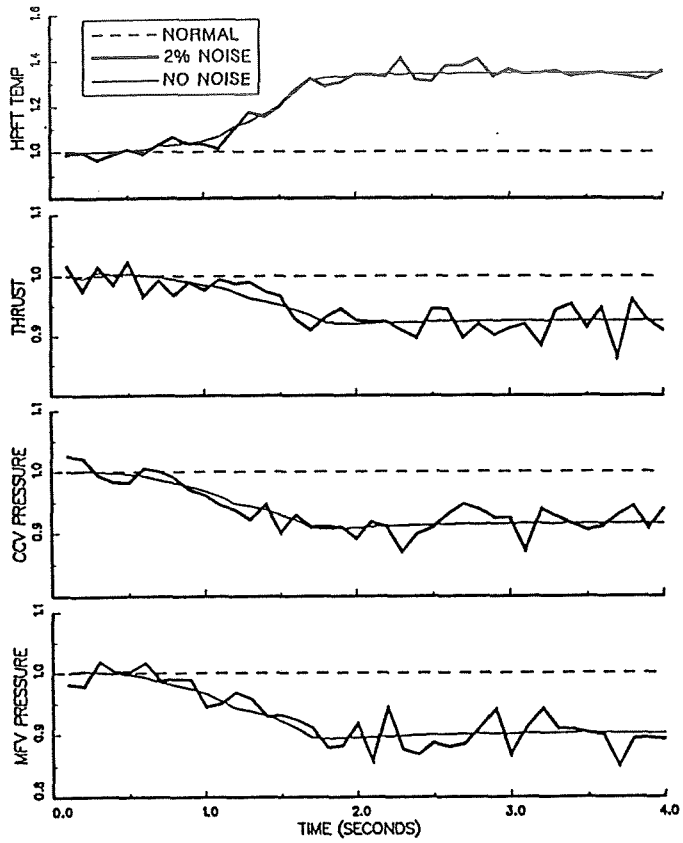


FIGURE 1. Sensor outputs for a 30% blockage of the main fuel valve with 1.5 second onset interval.

training examples (e.g., a pattern representing a previously unobserved fault condition), the networks can produce spurious categorizations and false positive identifications. There is no method to train a conventional feedforward network to, for instance, activate an output node if the input pattern is not similar to those patterns for which the network is trained.

2). Conventional feedforward networks are limited in the number and type of training examples which can be used. Some of the networks in the jet engine diagnostic system required long training times, which depended not only on the number of training examples required, but also on how similar training examples representing different faults were to each other.

These deficiencies can adversely effect the application of neural networks to the classification of fault patterns. The present work attempts to formulate a neural network architecture and training regimen which will successfully extend the capabilities of feedforward network models.

Architecture and Method

Consider the problem of inferring the operating conditions of a system from sensor data. Suppose there are M different possible system conditions $C_1 \ldots C_M$ and N different sensors $S_1 \ldots S_N$. In general, each possible condition C_i may be quantitatively characterized by a set of P parameters c_{ip}, $p = 1,2,\ldots,P$. (In our current research, each condition is characterized by two parameters, relating to the severity and onset interval of faults in jet and rocket engines.) The set of all possible system conditions can then be characterized by the set of parameters $\{c_{ip}\}$, where i enumerates the possible conditions and p enumerates the parameters associated with each condition. To infer these conditions, we may sample the value of each sensor S_j over a period of time. If each sensor is sampled at discrete times $t = 1 \ldots T$, then the complete set of sensor data is represented by the set of values $\{s_{jt}\}$, where j enumerates the different sensors and t enumerates the sampling times.

From the point of view of its operation, a system can be represented as a function

$$F : \{c_{ip}\} \rightarrow \{s_{jt}\}$$

which represents the cause-and-effect relationship from system conditions to sensor values. Ideally, if all conditions other than those enumerated could be held constant, F would be a deterministic function. In practice, however, F is usually characterized as a stochastic function in which values of s_{jt} are assumed to be influenced by noise as well as by the conditions $\{c_{ip}\}$. This noise results both from variability in testing conditions and from unreliability of the sensors. Sample data points for this function F can be derived by observing the physical system (under known test conditions) and/or by manipulating a simulation of the physical system.

In either case, the problem of inferring the condition of the system from observed sensor values is the problem of deriving the inverse function

$$F^{-1} : \{s_{jt}\} \rightarrow \{c_{ip}\}$$

which is the mapping from sensor values to hypothesized underlying conditions. For complex real-life systems such as jet and rocket engines, the nature of the function F^{-1} is unknown. However, observations of sensor values caused by known conditions (i.e. observations of F) can be used as example data points for F^{-1}. It might be thought that F^{-1} could readily be learned from these examples by a neural network trained with back-propagation. Each observed set of sensor values $\{s_{jt}\}$ would be input to the network, and the known conditions $\{c_{ip}\}$ would be the desired outputs for back-propagation training. In trials using jet and rocket engine data, however, we observed that two different fault conditions may cause very similar sensor behavior and also that low-severity, slowly developing faults are very difficult to distinguish from normal behavior. Such *differential diagnoses* have proven very difficult for an unstructured back-propagation algorithm to learn.

We propose that learning of the function F^{-1} from sample data points can be more efficient if F^{-1} is decomposed and structured into special-purpose constituent functions, each of which can be learned separately. The proposed decomposition is

$$F^{-1} = R * M * A$$

(where * denotes the composition of functions)

The purpose of the function R is data compression: to reduce the dimensionality of the sensor data $\{s_{jt}\}$ without losing the information necessary to make valid inferences. The purpose of the function M is to map this reduced sensor data $\{s'_{jh}\}$ into hypothesized conditions $\{c'_{jp}\}$. The functions R and M are applied separately to each sensor j. Finally, the purpose of the function A is to perform sensor fusion, i.e., to arbitrate among the hypotheses nominated by different sensors to determine which hypothesis is most likely. Each special-purpose component function R, M, and A is represented independently and trained separately. While the sensor fusion architecture A is the focus of the present study, the preprocessing functions R and M are discussed briefly in the following two subsections. The alternative sensor fusion architectures that were compared are then discussed in detail, and the results of the comparison are given.

Step 1: Data Compression

The data compression function R above is performed by an autoassociative network[6] with at least one layer of semi-linear hidden nodes. There are T input nodes and T output nodes, one for each discrete sampling time $t = 1,\ldots,T$. Data compression is accomplished by connecting the T input nodes to the T output nodes through a much smaller set of

$H << T$ hidden nodes. Node activations are continuous variables. The network is trained to associate input patterns with identical patterns clamped at the output nodes. As a result, training is unsupervised; it is left to the weights between the input, hidden, and output layers of the network to organize in a fashion whereby the mapping is performed correctly.

Once training has been completed, classification is accomplished by examining the activations of the hidden nodes. In a sense, the hidden node activations are employed as the "output" of the network. The mapping from input nodes to hidden nodes reduces the dimensionality of the input data from T (the number of time-series samples for a given sensor) to H (the number of hidden node activations). The mapping from hidden nodes to output nodes, on the other hand, has been used to reproduce the original input pattern from the hidden node activations. To the extent that training is successful, therefore, the H hidden node activations will contain sufficient information to reproduce the T time-series sensor values. The H hidden node activations for each sensor S_j thus compress that sensor's time-series values $\{s_{jt}\}$, $t = 1,...,T$ into its hidden node activations $(s'_{j1},...,s'_{jH}) = R(\{s_{jt}\})$, where R is the weighted-summation performed by the connections from the input layer to the hidden layer of the autoassociative network.

This special-purpose network for reducing the dimensionality of the sensor data can be trained very economically. Suppose that it is desired to train the overall architecture $R * M * A$ to classify a large number of temporal patterns (where each temporal pattern consists of a set of data curves, i.e. a set of time-series values for all sensors). While all of these curves will be classified by the mapping M of step 2 below, only a small subset of the training curves is needed to train the data compression network R. For training an autoassociative network to perform R, a small representative subset of the total set of curves is sufficient. Training is done on this subset of curves using an autoassociative back-propagation algorithm with output nodes clamped to the same values as the input nodes. After training is complete, data compression is accomplished by the resulting set of weights from the T input nodes to the H hidden nodes, as explained above.

Step 2: Hypothesis Generation

Once training has been completed for the autoassociative network above, the network is then run (with fixed weights) on the entire set of desired training data. Since no learning is involved at this stage, each sensor data curve $\{s_{jt}\}$, $t = 1,...,T$ can be collapsed into its hidden-node representation $\{s'_{jh}\}$, $h = 1,..,H$ in one iteration of the network. The H hidden node activations for each sensor constitute a compressed representation of the input data for that sensor. The entire set of training data can therefore be converted into a compressed representation with very little computation, i.e., only one iteration of the network per data curve per sensor.

Each data curve will evoke a specific hidden node response which can be represented as a single point $(s'_{j1},...,s'_{jH})$ in the H-dimensional parameter space of hidden node activations. Our studies have shown that data curves generated by smoothly varying the quantitative parameters $\{c_{ip}\}$ of a given fault condition C_i result in hidden node responses which map out a surface. If H hidden nodes are used, the hidden node activations will define a surface in an H-dimensional parameter space. For example, Figure 2 illustrates surfaces generated by training an autoassociative network, with three hidden nodes, on thrust data from the Space Shuttle Main Engine. The larger surface is generated from hidden node activations which correspond to blockage at the main oxidizer valve; the smaller surface corresponds to blockage of the main fuel valve. Both surfaces represent fault conditions over a range of severities and onset intervals. The hidden-node activation resulting from normal (steady-state) operation is represented as a single point in the parameter space.

In the present example, sensor curve characteristics are functions of two parameters: severity and onset interval for each condition C_i. Curves which vary by only one parameter c_{i1} (e.g., main oxidizer valve blockages of the same severity, but of different onset intervals) will evoke corresponding hidden node activations $R(F_j(c_{i1}))$ which define one coordinate direction of the main oxidizer valve surface. That is, $F_j(c_{i1})$ denotes the set of time-series data curves of sensor S_j when parameter c_{i1} is varied. R in turn compresses each curve into a point in the hidden-node space. Curves which vary by another parameter c_{i2} (e.g., main oxidizer valve blockages which vary in severity, but are constant in onset time) will evoke hidden node activations $R(F_j(c_{i2}))$ which define the other coordinate direction of the surface. The entire surface can therefore be mapped out by systematically varying both c_{i1} and c_{i2}. If $c_{i1} \times c_{i2}$ denotes the set of all such combinations of values of the fault parameters c_{i1} and c_{i2}, then $F_j(c_{i1} \times c_{i2})$ denotes the set of time-series curves for sensor j's responses over this range

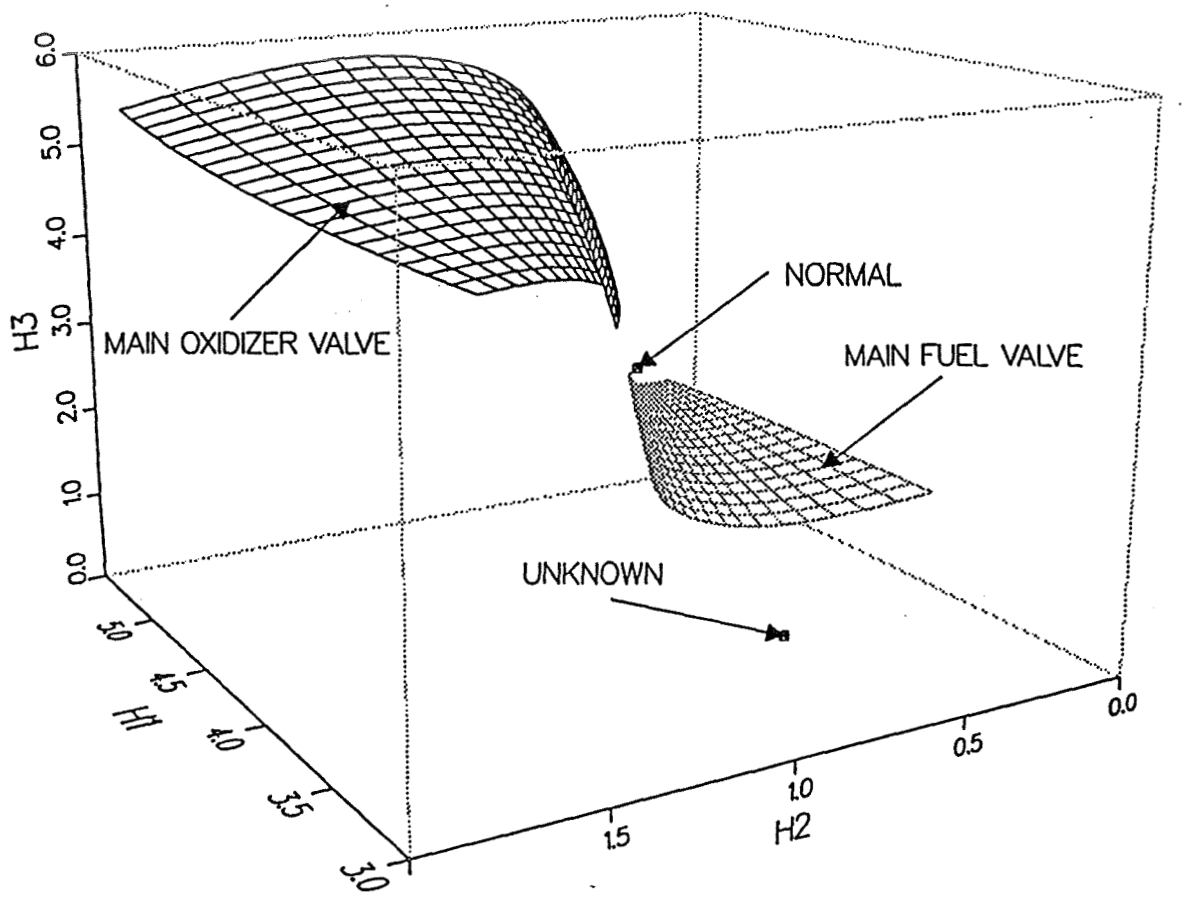


FIGURE 2. Surfaces representing Main Oxidizer Valve and Main Fuel Valve blockages for Thrust sensor. The autoassociative network maps each possible input pattern into one point H1, H2, H3, representing the hidden node activations which result from that input pattern. A class of related input patterns will generate a surface in this space.

of parameters. $R(F_j(c_{i1} \times c_{i2}))$ in turn denotes the surface obtained by compressing these sensor curves into points in the H-dimensional hidden-node space. Similarly, a set of main fuel valve blockages covering a range of severities and onset intervals can be used to define another surface.

As explained above, the entire set of training data can be converted into points in the hidden-node parameter space in one iteration of the autoassociative network per data curve per sensor. In the present implementation, these points are simply interpolated to generate a parameter surface for each fault condition. "Training" the map M from hidden-node activations to parameters consists merely of (i) generating a point in the hidden-node parameter space for each training example; (ii) labeling each such point with the known parameters (e.g. severity and onset interval) of the example which generated the point; and (iii) interpolating a coordinatized surface from the set of points obtained by varying the parameters of a given condition.

Note that each different sensor S_j will generate its own set of surfaces in H-dimensional data compression space, showing the responses of that sensor to the range of conditions in the training data. There will therefore be a separate map M_j for each sensor S_j.

This training process does not in any way require that only three hidden nodes be used, but is fully extensible to higher numbers of hidden nodes. If higher numbers of hidden nodes are used, higher-dimensional surfaces (hypersurfaces) will be generated.

Once a network is trained and the surfaces generated, it may be used for classification of new input patterns. In the example of Figure 2, a new input will be mapped into a new point in multi-dimensional space. If the point lies on or near the surface defined by the training examples, then the resulting hypothesis is that the fault condition represented by that surface is indeed occurring. The closer the new point is to the surface, the stronger is the evidence it provides for that hypothesis. If the new point is close to more than one surface, then more than one hypothesis will be generated, but with different levels of confidence if the distances to the surfaces are different.

After training is complete for steps 1 and 2 above, a new sensor curve $\{s_{jt}\}$, $t = 1,...,T$ can be converted into a new point $(s'_{j1},...,s'_{jH})$ in the hidden-node parameter space, where $(s'_{j1},...,s'_{jH}) = R(\{s_{jt}\})$, R being the mapping performed by the data compression network of step 1. This point in turn can be projected onto each surface $R(F_j(c_{i1}\times c_{i2}))$ yielding values for the parameters c_{i1} and c_{i2} of the hypothesized condition C_i. The degree of evidence for each hypothesized condition is a function of the distance from the new point to the surface representing that condition, as discussed below. If j sensors are operating simultaneously, then this process is carried out separately for each sensor and each surface, projecting to the surfaces that were generated for that particular sensor during training. This will yield a different opinion from each sensor as to the likelihood of various hypothesized conditions. The purpose of the third component of our decomposed architecture is to fuse the information from different sensors into a reliable inference.

Step 3: Sensor Fusion

The key to obtaining a reliable overall inference is the reliability of the differential diagnoses which can be contributed by each sensor. Figure 3 shows a typical problem of differential diagnosis. For this sensor, the main oxidizer valve surface and the main fuel valve surface intersect. The set of hidden-node-activations near this intersection are therefore consistent with blockage of either the main oxidizer valve or the main fuel valve. In this region of the hidden-node space, differential diagnosis by this sensor would be quite unreliable. On the other hand, for points which are near one surface but not near to the other surface, the data is consistent with only one interpretation and therefore the differential diagnosis is more reliable. Finally, points far from both surfaces may indicate either an unknown condition or a faulty sensor. These different possibilities may be represented by the set of distances from a given point $(s'_{j1},...,s'_{jH})$ in the hidden-node space to each of the M surfaces $\{R(F_j(c_{i1}\times c_{i2}))\}$, $i = 1,...,M$ in that space (as well as the distance to the "normal" point $R(F_j(c_0))$ derived from sensor values under normal steady-state conditions).

More specifically, let us define d_{ij} to be the distance from the point $(s'_{j1},...,s'_{jH})$ to the nearest point on the surface $R(F_j(c_{i1}\times c_{i2}))$, or to the normal point $R(F_j(c_0))$ when $i = 0$. Since the point $(s'_{j1},...,s'_{jH})$ was derived from sensor S'_j's data, and since the surface $R(F_j(c_{i1}\times c_{i2}))$ gives the set of such points predicted by hypothesis C_j, then d_{ij} indicates how far sensor S'_j's data is from the predictions of hypothesis C_i. This distance can be turned into a consistency measure by defining a tolerance D_i for each surface derived from the variance observed in the set of training data used to determine each surface $R(F_j(c_{i1}\times c_{i2}))$. (A tolerance D_0 can also be defined around the normal point $R(F_j(c_0))$. A new data point closer than this tolerance to the surface should be taken as evidence in favor of the hypothesis, while a data point farther away should be taken as evidence against. We therefore take $D_i - d_{ij}$ as a measure of the *consistency* of the data from sensor S_j with hypothesis C_i.

Suppose, then, that these consistency measures are represented as activations of input nodes in a layered neural network to perform sensor fusion, as in Figure 4. In general, such a network would require $(M+1)\times N$ input nodes X_{ij}, $i = 1,...,M$, $j = 1,...,N$, that is, one node X_{ij} for each pairing between $M+1$ hypothesized conditions and N sensors (counting the steady-state condition as one hypothesis). Each input node X_{ij} of the sensor fusion network receives a scalar input $D_i - d_{ij}$, where D_i is the tolerance associated with surface $R(F_j(c_{i1}\times c_{i2}))$ and where d_{ij} is the distance between the point $(s'_{j1},...,s'_{jH}) = R(\{s_{jt}\})$ and the surface $R(F_j(c_{i1}\times c_{i2}))$. (Recall that R is the data compression function performed by the autoassociative network trained in Step 1 above, and $F_j(c_{i2})$ gives the series of points generated by parameter c_{i2} of condition C_i during the training of Step 2.) Each such input node activation $D_i - d_{ij}$ therefore represents the *consistency* of data from sensor S_j with hypothesis C_i.

The desired output of the sensor fusion network is the most likely hypothesis based on all sensor data. The network therefore contains $M+1$ output

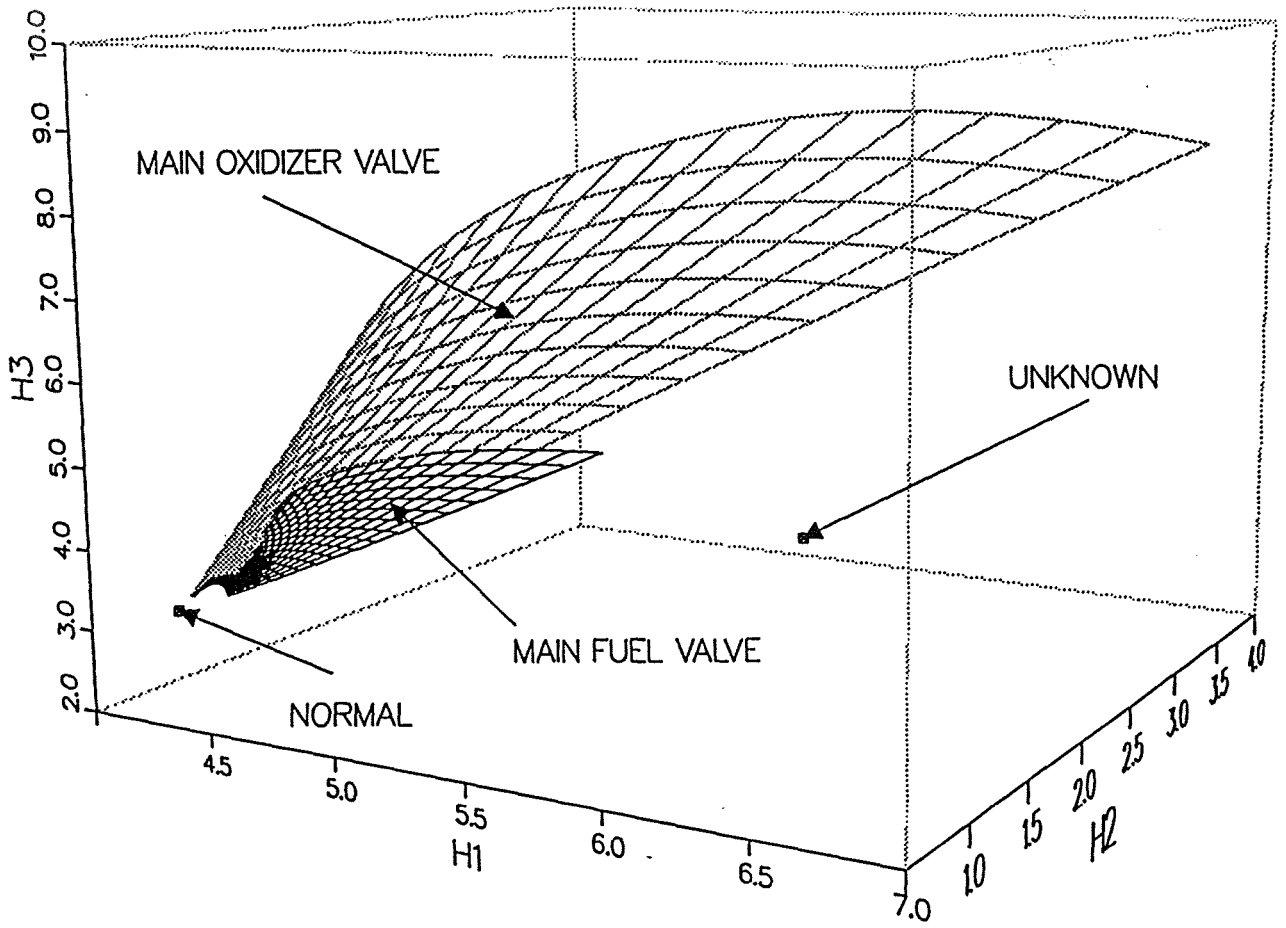


FIGURE 3. Surfaces representing Main Oxidizer Valve and Main Fuel Valve blockages for High Pressure Fuel Turbopump Inlet Temperature sensor.

nodes, one node Y_i for each possible hypothesis of a fault condition C_i, $i = 1,...,M$, and one node Y_0 for the hypothesis of normal steady-state operation. Given these sets of input nodes and output nodes, and restricting ourselves to one layer of $(M+1)^2$ hidden nodes, we can then define two possible architectures which differ only in the connectivity from the input nodes to the hidden nodes and from the hidden nodes to the output nodes. These two architectures are described below:

Fusion Architecture A is shown in Figure 4. A structured connectivity pattern is defined which contains $(M+1)^2$ hidden nodes H'_{ik}, $i = 0,...,M$, $k = 0,...,M$. First let us consider the case where $i \neq k$. Each hidden node H'_{ik}, $i \neq k$ has random excitatory connections from the set of input nodes $\{X_{ij}\}$, $j = 1,...,N$, and random inhibitory connections from the set of input nodes $\{X_{kj}\}$, $j = 1,...,N$. Hidden node H'_{ik} therefore calculates a weighted sum of evidence from all sensors. In this weighted sum, evidence in favor of hypothesis C_i counts positively, but evidence against hypothesis C_i counts negatively. Conversely, evidence in favor of condition C_k counts negatively in the weighted sum, but evidence against hypothesis C_k counts positively. Hidden node H'_{ik} is thus prewired to receive all data relevant to a differential diagnosis of hypothesis C_i over hypothesis C_k. (Conversely, hidden node H'_{ik} would receive data relevant to a differential diagnosis of hypothesis C_k over hypothesis C_i. For example, if M=2 classes of fault conditions, then there are $M(M+1)$=6 hidden nodes H'_{ik} with $i \neq k$ to perform all pairwise differential diagnoses between hypotheses

C_i and C_k. These are illustrated as the upper 6 hidden nodes in the middle layer of Figure 4.

The M+1 remaining hidden nodes (designated H'_{ii} for convenience, $i = 0,...,M$) are also prewired to perform differential diagnoses, but in this case each differential diagnosis H'_{ii} is between (i) the hypothesis that the current condition of the system is C_i, and (ii) the hypothesis that the current condition of the system belongs to some *unknown* class of anomalies which has not been seen during training. (In the case where M=2, there would be 3 such hidden nodes, illustrated as the lower 3 hidden nodes of Figure 4.) In essence, each of these hidden nodes H'_{ii} is attempting to detect unknown anomalies in general, and more specifically to differentiate the class of such unknown anomalies from a particular known condition C_i. Since by definition no training examples are available for unknown anomalies, the hidden nodes H'_{ii} are prewired with inhibitory connections suitable for detecting unknowns. Recall that the activity D_i–d_{ij} of each input node X_{ij} is negative if the distance d_{ij} from the nearest known condition C_i is greater than the preset tolerance D_i. Therefore, a negative activation in any input node should count as a contribution to the evidence for an unknown anomaly. Moreover, a negative activation in the particular input nodes X_{ij} representing distance from the known condition C_i should especially count as evidence to differentiate an unknown anomaly from the known condition C_i. Each hidden node H'_{ii} for detecting unknown anomalies should therefore have some inhibitory input from all input nodes (as evidence for an unknown anomaly), and stronger inhibitory weights from the particular input nodes X_{ij}, $j = 0,...,M$ (as evidence to differentially diagnose an unknown anomaly from the condition C_i). In figure 4, dotted lines from input nodes to the lower 3 hidden nodes indicate connections which are initially set to strongly inhibitory weights to perform each differential diagnosis. As we have explained, there are also weaker inhibitory connections (not shown in the figure) from all other input nodes to each of the lower 3 hidden nodes.

The weights leading to any given hidden node define a discriminant function which is customarily thought of as a hyperplane in the space of all possible input vectors. All points (input vectors) on one side of this hyperplane result in a positive activation of the given hidden node; all points on the other side result in a negative activation. This discrimination is sharpened by the nonlinear sigmoid function applied to the activation to yield the hidden node's output. The

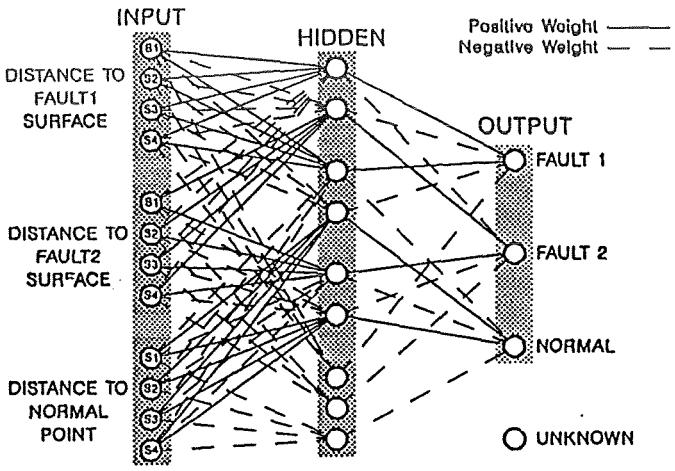


FIGURE 4. Architecture A: Prestructured to perform differential diagnosis (all connections not shown). Solid lines indicate connections initially set to positive wieghts (before training begins); dotted lines indicate connections initially set to negative weights.

weight changes of back propagation training in effect move each hyperplane in a direction which lessens the mean-squared error of the network's output over the training set.

From this standpoint, the differential diagnosis performed by each hidden node H'_{ik}, $k \neq i$, can be thought of as a hyperplane intended to separate the data points representing training examples of condition C_i from those representing training examples of condition C_j. While the initial setting of weights biases each hidden node H'_{ik} to perform this differential diagnosis, back propagation will move this hyperplane in whichever direction minimizes the mean-squared error of the network over the set of training examples.

Similarly, the role of each hidden node H'_{ii} can be viewed as a hyperplane intended to distinguish all known conditions C_i on one side, from unknown anomalies on the other side of the hyperplane. M+1 such hyperplanes are created by the hidden nodes H'_{ii}, $i = 0,...,M$. These M+1 hyperplanes are initially placed in different positions due to the stronger inhibitory weights assigned to the inputs X_{ij}, $j = 0,...,M$ than to the other inputs by each particular hidden node H'_{ii}, as explained above. Since no training examples are available for unknown anomalies, back propagation might conceivably reduce the number of known training examples erroneously classified as unknown, but would not be expected to improve the recognition of unknown anomalies as such. We hypothesized, however, that the effect of this inherent limitation of example-based training would be les-

sened by the prestructuring (initial setting of connections and weights) of architecture A described above.

Finally, each output node Y_i receives initially excitatory connections from the hidden nodes $\{H'_{ik}\}\ k = 1,...,M, (k \neq i)$, initially inhibitory connections from the hidden nodes $\{H'_{ki}\}\ k = 1,...,M, (k \neq i)$, and an initially inhibitory connection from the hidden node H'_{ii}. Each output node Y_i thus receives excitatory input from all differential diagnoses favoring hypothesis C_i, and inhibitory input from all differential diagnoses opposing hypothesis C_i.

Fusion Architecture B is shown in Figure 5. This architecture has the same number of hidden nodes as architecture A, and the same number of input and output connections per hidden node, but with random connectivity from input to hidden and hidden to output layers, and random assignment of initial weights. This architecture is intended to serve as a "control" case against which the effects of the initial structuring of architecture A can be evaluated.

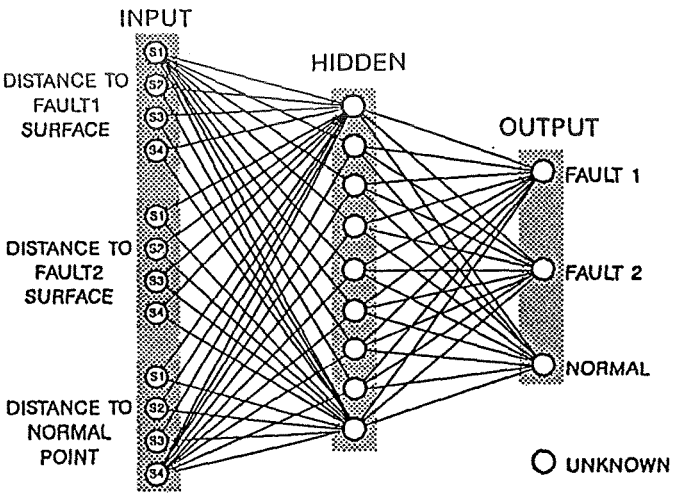


FIGURE 5. Architecture B: Fully connected with random initial weights (all connections not shown). All connections are initially set to random weights drawn from a uniform distribution between -0.1 and 0.1.

Both architectures for sensor fusion were trained with the same back-propagation algorithm, using the outputs of the training data that were used in step 2 to determine the surfaces $\{R(F_j(c_{i1} \times c_{i2}))\}, i = 1,...,M$ for each sensor j. Our hypothesis was that the back-propagation algorithm constrained to the connectivity of architecture A would (i) result in a set of weights from input nodes to hidden nodes which allow the hidden nodes to perform pairwise differential diagnoses, (ii) would reliably differentiate *unknown* engine conditions (not present during training) from the known classes of engine conditions present during training (a capability expected to be lacking in the unstructured back propagation architecture B), and (iii) that this would be accomplished without sacrificing diagnostic performance for known engine conditions, in comparison to the unstructured back propagation architecture.

Testing Procedure

The decomposed architecture (performing data compression, hypothesis generation, and sensor fusion) was trained as described in steps 1-3 above on simulated SSME data for the four sensors illustrated in Figure 1: high pressure fuel turbopump temperature, engine thrust, chamber coolant valve pressure, and main fuel valve pressure. The training set consisted of normal data and two fault conditions, main oxidizer valve (MOV) blockage and main fuel valve (MFV) blockage. Each fault condition was included in the training set at three different levels of severity, and three different onset intervals for each level of severity. In order to compare the performance of sensor fusion architectures A and B, they were each trained on the identical output of Steps 1 and 2 of the training procedure.

After training was completed, the performance of architectures A and B were compared on test data containing the 3 "known" conditions used in training (normal, MOV blockage, and MFV blockage) and 2 additional fault conditions that were not presented during training, Oxidizer Preburner Valve (OPV) blockage and Fuel Preburner Valve (FPV) blockage. These two additional "unknown" conditions were included in order to test each architecture's ability to detect fault conditions which had not been included in the training set. 500 instances of each condition were generated which differed only in the amount of noise added to the simulated data. For each noise level, 100 instances of each condition were generated with the inclusion of random noise at that noise level. During testing each example was classified as one of the three training conditions (normal, MOV blockage, or MFV blockage) based on the maximally active output node, or as "unknown" if none of the output nodes was activated above its threshold value. Each architecture (A and B) was tested using a range of different thresholds, and the threshold yielding the best performance for that network was used.

Table I

Percent correct classification
at various noise levels performed by prestructured architecture A.

Fault	Noise Level (%)				
	0.0	0.5	1.0	1.5	2.0
Normal (Known)	100	100	95	81	70
MOV (Known)	100	100	100	100	100
MFV (Known)	100	97	86	85	80
OPV (Unknown)	100	100	100	100	100
FPV (Unknown)	100	100	100	100	100

Table II

Percent correct classification
at various noise levels performed by fully connected architecture B.

Fault	Noise Level (%)				
	0.0	0.5	1.0	1.5	2.0
Normal (Known)	100	100	94	76	69
MOV (Known)	100	100	100	100	98
MFV (Known)	100	94	86	79	79
OPV (Unknown)	0	0	0	1	5
FPV (Unknown)	0	0	0	0	0

Results and Discussion

Table I shows the results of testing sensor fusion architecture A. As explained above, this neural network architecture was structured prior to training to perform differential diagnoses among the different conditions to be presented during training, and between each of these conditions and the class of unknown faults not presented during training. Table II shows the results of testing architecture B using the same training data and same testing data. Architecture B differed from architecture A in that the connection weights were not structured prior to training, but rather were set to random initial values as typically done in neural networks.

Each column in Tables I and II shows the results obtained for a given noise level in the test data. Within each column, classification performance (percent correct classifications) is shown separately for the normal operation condition and for each fault condition.

The first three rows of each table show the performance on test data representing the three known conditions (normal and two types of faults) that were presented during training. The performance of both architectures declined with the addition of greater levels of noise to the test data, with architecture A performing very slightly better at the higher noise levels, As expected, the two architectures did not differ greatly in their classification performance on these known conditions. This is consistent with our hypothesis that the initial structuring would not detract from the network's ability to correctly classify new examples of the same conditions presented during training.

The last two rows of each table show the results of testing with the two unknown fault conditions which had not been presented during training. Since a conventional back propagation network learns to classify based on its training examples only, we expected architecture B not to be able to recognize new fault conditions as unknown, but rather to classify them into one of the classes that had been presented during training. This is indeed what happened. As shown in the last two rows of Table II, its performance was five percent or less correct classifications at each noise level of these two unknown faults. Recall that results are shown for the best threshold setting for each network. In other words, there was no threshold setting which would allow the output nodes of architecture B to differentiate known examples from unknown examples on the basis of its output node activity.

Architecture A, by contrast, was able to correctly identify new faults as unknown, based on below-threshold activity in all output nodes. This is shown

in the last two rows of Table I, in which all examples of new faults at all noise levels are correctly classified on this basis. Since the only difference between architectures A and B is in the initial structuring of connections prior to training, it is reasonable to conclude that the initial structuring of architecture A to perform differential diagnoses allows this network to detect examples of unknown classes that were not presented during training.

The structured back-propagation network of architecture *A* could be viewed as a hybrid between a knowledge-based expert system and an example-trained neural network. In knowledge-based systems, both the general format of the rules and the exact instantiations of the rules are extracted from human experts. In conventional back-propagation networks, hidden nodes serve a function analogous to rules in an expert system. The general format of such a "hidden-node rule" is determined by which input nodes have significant connections to each hidden node, while the exact instantiation of such a rule is given by the exact weights which result from training. In conventional back-propagation, both the format of the hidden-node rules and the exact instantiation of these rules are implicitly determined from training examples by the back-propagation algorithm. The structured back-propagation architecture *A* above represents a hybrid between these two approaches. The general format of its "hidden-node rules" is determined in advance by the connectivity specified for architecture A, and is based on expert judgment about the general utility of rules based on differential diagnosis. Within this general format, however, the specific instantiation of each differential diagnosis rule is determined by the exact weights which are learned from training examples via back propagation. This initial structuring of the back-propagation architecture using expert human judgment allows the neural network to detect the occurrence of faults for which no training examples were presented. This dependence on expert human judgment is much less than in a rule-based expert system, however, since the exact instantiations of the rules are still learned from training examples. Training based on examples should make these "hidden-node" rules easier to maintain than in a conventional rule-based expert system. However, this reduced dependence and implicit learning of "hidden-node-rules" makes it more difficult to provide explanations to the user about the inference process. Nevertheless, forcing the "hidden-node rules" into a predefined format allows the behavior of the network to be more easily compared with the behavior of human experts than in the case of an unstructured back-propagation network. This potentially would allow the network to "explain" the reasons for its diagnosis in terms understandable by a human operator.

Acknowledgement

This work was performed within the Center for Advanced Space Propulsion (CASP) and was supported in part by NASA Grant NAGW-1195 and Rocketdyne Contract No. R04QBZ90-032709. The Center for Advanced Space Propulsion is part of The University of Tennessee-Calspan Center for Aerospace Research, a not-for-profit organization located at UTSI. The authors would like to thank A. M. Norman for his assistance during the course of this effort.

References

1. Cikanek, H.A., "Space Shuttle Main Engine Failure Detection", presented at the 1985 American Control Conference, Boston, MA, June 19-21, 1985.

2. Dietz, W.E., Kiech, E.L., and Ali, M., "Jet and Rocket Engine Fault Diagnosis in Real Time", *Journal of Neural Network Computing*, Vol. 1, No. 1, 1989.

3. Dietz, W.E., Kiech, E.L., and Ali, M., "Pattern-Based Fault Diagnosis Using Neural Networks", The First International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, Tullahoma, TN., pp 13-23, June 1-3, 1988.

4. Dietz, W.E., Kiech, E.L., and Ali, M. "Classification of Data Patterns Using an Autoassociative Neural Network Topology", The Second International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, The University of Tennessee Space Institute, Tullahoma, TN, pp. 1028-1036, June 6-9, 1989.

5. Rumelhart, D.E., Hinton, G.E., and Williams, R.J., "Learning Internal Representations by Error Propagation", pp. 318-364, in *Parallel Distributed Processing*, Vol. I, The MIT Press, Cambridge, MA, 1987.

6. Kuczkewski, R.M., Myers, M.H., and Crawford, W.T., "Exploration of Backward Error Propagations as a Self-Organizational Structure", TRW MEAD, One Rancho Carmel, San Diego, CA 92128.

N96- 70676

MACHINE LEARNING OF FAULT CHARACTERISTICS FROM ROCKET ENGINE SIMULATION DATA

Min Ke
Moonis Ali

Center for Advanced Space Propulsion
The University of Tennessee Space Institute
Tullahoma, TN 37388

Abstract

Transformation of data into knowledge through conceptual induction has been the focus of our research described in this paper. We have developed a Machine Learning System (MLS) to analyze the rocket engine simulation data. MLS can provide to its users fault analysis, characteristics, and conceptual descriptions of faults, and the relationships of attributes and sensors. All the results are critically important in identifying faults.

I. Introduction

An important component of intelligent diagnostic systems is the knowledge which human experts employ in analyzing and diagnosing faults. However, this knowledge employed is very limited in the sense that it is based on a very small number of observed situations. Since human experts have not seen all possible instances of all faults, they cannot describe fault characteristics sufficiently well to make diagnostic decisions. We have developed a Machine Learning System (MLS) for analyzing the SSME simulator data to generate characteristics about engine faults. In MLS inductive heuristics and domain knowledge are employed to guide the inductive process. With two phases of abstractions as well as a knowledge management system, MLS can be applied to a wide spectrum of domain tasks. MLS has been tested with SSME simulator data. MLS consists of two levels of abstractions. Section II presents the general algorithm, Section III describes the basic abstraction, Section IV describes the advanced abstraction, and Section V discusses the tests and results.

II. The General Algorithm

MLS consists of two levels of abstraction: the basic abstraction which generates the characteristic descriptions, discriminant descriptions and aggregational descriptions for a concept such as an engine fault; and the advanced abstraction which groups similar concepts into a concept hierarchy to form a higher level of concepts. MLS incorporates concept instance data incrementally.

In MLS the paradigm for inductive inference can be formulated as follows:

Before induction we have:

- A hierarchy of nodes with leave nodes representing basic concepts, internal nodes representing clusters, and descriptions of concepts at each node;
- Knowledge bases which include knowledge about attributes, components, relationships and basic concepts, and also include deductive rules, generalization rules, transformation rules and aggregation rules;
- A new instance description of a basic concept represented in MLS's representation language.

After induction we get:

- An extended hierarchy of concepts such that either a new basic concept node is added or an old basic concept's descriptions are modified to cover the new instance; the structure of the hierarchy and the cluster node's descriptions are modified to incorporate the new instance.

Let C be a basic concept, and let e be a new instance of C. A raw description is a description of raw data in our representation language. In the knowledge base, appropriate transformation rules, aggregation rules, deductive rules, generalization rules and inductive heuristics are provided. It is assumed that e is a set of simple expressions (i.e., atoms). The following is the outline of the general algorithm of MLS:

1) Read in e, raw description of an instance of concept C.
2) Apply simple transformation and attribute-level aggregation on e, generating result e1.
3) Apply deductive transformation and component-level aggregation on e1, generating result e2. New attributes and relationships are generated in this step.
4) Generalize concept C's characteristic descriptions to cover e2.

This research was supported by NASA Grant Nos. NAGW-1195, and NAG-1-513 and Rocketdyne Contract No. R04QBZ90-032709.

5) Generalize concept C's discriminant descriptions and specialize (or eliminate) discriminant descriptions of other concepts.

6) Modify the concept-level aggregation descriptions.

7) Modify the generalization hierarchy above the basic concept level. The most frequent operation is to modify (adding, generalizing, or changing the weight of) the descriptions of the concept class on the higher levels. Other operations include creating a new class and deleting an unqualified class.

The algorithm described here is only one process of a single instance. As an incremental method, the above algorithm can be repeatedly applied to many incoming instances.

III. Methods and Algorithms for the Basic Abstraction

MLS is an incremental learning system, so it is efficient to incorporate new instances. MLS has a capability of rich logic representation. Multiple-valued nominal attributes and inexact value matching are MLS's capabilities not shared by most other systems.

Each time an instance of a specified concept is incorporated into the concept hierarchy, the object of the concept and the object of the instance are sent to the procedure that incrementally modifies the characteristic description of the concept. The algorithm for construction of characteristic descriptions can be described as follows:

1. Check whether the basic concept is a new concept. If it is a new concept, the instance description is taken as the characteristic description of the concept and the concept hierarchy should be extended to incorporate the new basic concept. If it is not a new concept, then perform the following steps.
2. For each expression in the instance object, try to find the matching expression in the concept object.
   If not found, put the expression into the Alternative List AL.
   If found, generalize the two matching expressions, use the resultant expression as one expression of the characteristic descriptions, increase the count of instances that imply this expression, and compute the worth of the new expression.
3. Apply the Add-Alternative Rule to the unmatched expressions of the concept characteristic descriptions and expressions in AL. The resultant expressions are put into the concept characteristic description. Also some unqualified expressions are eliminated from the characteristic description of the basic concept.
4. Call the bottom-up modification procedure in the advanced abstraction. Use only confident features of the concept to modify the concept hierarchy. If the consistency factor of an expression is greater than the consistency factor threshold, then the expression is considered a confident feature. Since all characteristic features are complete (completeness factor is 1.0), only the consistency factor needs to be considered.

The Algorithm for Constructing Discriminant Descriptions is given below:

The inputs to this algorithm are an object of concept C1 and an object of C1's instance.

For each expression EXP in the instance description perform the following steps:

1. Obtain all matching expressions from the uniqueness table UT.
2. If there is no matching expression from UT, then EXP is a unique feature.

   Add EXP to the discriminant description of C1.

   Add an entry of EXP in UT. Exit.
3. If there are matching expressions, then try to find inconsistency. Let ML be the list of matching expression entries.
   For each entry in ML, check for inconsistency (Loop A)
   1) If the expression EXP is a discriminant feature of another concept C2, then an inconsistency situation is found.

   Delete EXP from the discriminant description of C2.

   Set the entry uniqueness flag off.

   Exit Loop A.
   2) If EXP is covered by a discriminant feature of another concept C3, then an inconsistency situation is found.

   Specialize the discriminant feature of C3.

   Modify the entry in UT.

   Exit Loop A.
   3) If EXP is covered by an inconsistent feature (an inconsistent feature is a feature that is already identified as nondiscriminant), then EXP is inconsistent. Exit Loop A.
4. If EXP is not found to be inconsistent in step 3, then check for partial inconsistency.
   For each entry in ML, if the expression in the entry is partially covered by EXP, then specialize both expressions;

modify the discriminant descriptions of the concept specified in the entry;

modify the entry.

5. If EXP is matching a discriminant feature EXP1 of C1 (in which case EXP is said to be compatible), then

perform the least generalization on EXP and EXP1;

modify C1's discriminant description

modify the entry in UT.

6. If EXP is neither inconsistent nor compatible, then

add EXP to C1's discriminant description; create an entry for EXP in UT.

IV. Methods and Algorithms of the Advanced Abstraction

The advanced abstraction is an integration of two incremental processes: 1) modification of the cluster hierarchy, and 2) modification of cluster descriptions. The results of the advanced abstraction are a clustering of basic concepts and conceptual descriptions of the clusters. The combination of incremental processes with an expressive logic representation language makes MLS a unique system in the conceptual cluster area. In this section, we illustrate how the matching factor is computed; describe how to measure the quality of clustering; discuss the operations on the concept hierarchy as well as the algorithms of hierarchy extension and modification; discuss the cluster parameters; and analyze the time cost of the advanced abstraction.

Features in a cluster description will be constantly modified during the incremental inductive process. A feature has a *confidence count* which may be increased or decreased. The confidence count of a feature determines whether the feature is confident or not. Only the set of confident features is used as the description of a cluster.

When the first instance of a new basic concept is incorporated into the concept hierarchy, a top-down extension process is performed. The procedure to perform top-down extension of the concept hierarchy is a recursive one. The first call uses the root node as one of the parameters. The root is a universal cluster node which incorporates all instances input to the inductive system. The description of the root is a generalization of all the incorporated instances.

A description of the procedure, nameed extend-hierarchy, is given below where *currentnode* is the current cluster, and *newconcept* is the new basic concept to be incorporated:

extend-hierarchy (currentnode newconcept)

(1) Modify the current cluster:

a. Increase the size (by 1) of the current cluster.

b. For each expression E in the cluster description perform the following steps:

i. If there is no matching expression of E in *newconcept*'s confident features, then see whether the expression confidence count of E is still greater than the feature retaining threshold FR-TH. Delete E if the expression confidence count of E is less than FR-TH. Retain E if the expression confidence count of E is not less than FR-TH.

ii. If there is a confident expression E1 in *newconcept* that matches E and E1 is covered by E, then increase the expression confidence count of E by 1, and compute the worth of the expression.

iii. If the confident feature E1 is not covered by E, but the value matching factor of E1 and E is greater than the feature matching threshold FM-TH, then generalize E to incorporate E1 and increase the expression confidence count of E by 1, and compute the worth of the expression.

iv. If E1 is not covered by E and value matching factor of E1 and E is less than FM-TH but greater than the feature conflict threshold FC-TH, then do nothing. If the value matching factor is less than FC-TH, then decrease the expression confidence count of E by 1, and see whether the expression confidence count of E is less than FR-TH. Delete E if the count is less than FR-TH. Retain E if the count is not less than FR-TH.

v. Add all unmatched expressions of *newconcept's* confident features to *currentnode*'s description.

(2) In the current cluster find each subcluster S that is close to *newconcept* and whose description does not violate constraints of *newconcept*. Closeness is measured by the matching factor between S and *newconcept*. A closeness threshold C-TH is used to determine whether a subcluster is close enough to *newconcept*.

(3) Find each basic concept directly under the current cluster that is close enough to *newconcpt*.

(4) If there are no close subclusters and basic concepts, then put *newconcept* directly under the current cluster.

(5) Find the best object from those close subclusters and basic concepts according to the quality measure of clustering. Here the quality measure is modified to include basic concepts. Since a basic concept does not have a subcluster or subconcept, the cluster matching factor is replaced by the matching factor of the basic concept and *newconcept*.

(6) If the best node is a basic concept, then combine the best node with *newconcept* to form a new subcluster of the current cluster: Create a new cluster. Generalize expressions of the descriptions of the two concepts. Two expressions are generalized only if their value matching factor is greater than FM-TH (0.7). The new expression's confidence count is the sum of the two counts of the generalized expressions. The new expression's worth is the average worth of the two generalized expressions.

(7) If the best node is a subcluster, then recursively call:
extend-hierarchy (bestnode newconcept)

When incorporating a new instance to an existing basic concept, the description of the basic concept may be generalized to cover the new instance. This modification may cause further modifications on the predecessors of the basic concept.

For each expression EXP in the instance's description, if it is not covered by a matching expression of the basic concept, the following procedure is performed:

Let P be the parent cluster of the basic concept C, OLDEXP be the expression in C that matches EXP, NEWEXP be EXP or the generalization of EXP and OLDEXP.

modify-hierarchy(C, OLDEXP, NEWEXP)

Get C's parent P.

Try to find EXP1, the expression in P that matches OLDEXP.

If not found (EXP1 is empty), then add NEWEXP to P's description.

If EXP1 is not empty, NEWEXP is EXP, OLDEXP covers NEWEXP, and the value matching factor of OLDEXP and EXP1 is greater than FM-TH, then increase the confidence count of EXP1 (Now NEWEXP is taken as the supporting feature of EXP1).

If EXP1 is not empty, NEWEXP covers OLDEXP, and the value matching factor of NEWEXP and EXP1 is greater than FM-TH, then generalize EXP1 to incorporate NEWEXP and increase the confidence count of EXP1.

If EXP1 is not empty, NEWEXP covers OLDEXP, and the value matching factor of EXP1 and NEWEXP is less than FC-TH, then NEWEXP is considered to be contradict to EXP1.

- The confidence count IC of EXP1 is decreased by 1.
- Test IC to see whether it is less than FR-TH.
- If IC is less than FR-TH, EXP1 is deleted from P's description; check the number of expressions in P's description; if the number is less than the cluster retaining threshold CR-Th, then delete P from the concept hierarchy and reassign all P's leaves and subclusters under P's parent (note: the root can never be deleted).

If P has a parent, then recursively call:
modify-hierarchy(P, OLDEXP, NEWEXP).

After modifying the concept hierarchy, try to combine C with one of its sibling concept nodes and create a new cluster, since after the modifying of the concept description, the concept may become close to another basic concept under the same cluster node.

V. Analysis and Results

Engine test analysis is one of several application areas of inductive learning, where conceptual descriptions about different faults can be automatically generated from a large number of fault instances, and similar faults can be classified into clusters. The inductive results, including high level characteristic descriptions and discriminant descriptions of faults and the concept hierarchy with descriptions of clusters, can be used to aid the fault test analysis and be used for fault diagnosis.

The Space Shuttle Main Engine (SSME) is one of the most complex reusable liquid-fuel (oxygen and hydrogen) rocket engines. Each time a test on SSME is performed, a huge amount of data is collected from many sensors. Many highly-trained engineers are required to perform a thorough investigation of the tests. Two difficulties are presented for the improvement of test analysis: (1) As more tests are performed and more thorough investigations are required, more experienced engineers are needed; (2) more senior staff with many year's experience are leaving. To overcome these difficulties, a computer conceptual induction technique is used to aid the engineers in analyzing the test data. In addition to its efficiency in forming concepts and generating concept descriptions, a computer inductive system can accumulate knowledge from both the data of many tests and the expertise of the engineering staff.

After SSME simulator data is input into MLS, a concept hierarchy is built by the inductive system. The concept nodes on the hierarchy represent various engine faults. The cluster nodes on the hierarchy represent higher level concepts, each of which describes a group of similar engine faults. Descriptions of engine faults or fault groups are also generated by the inductive system and stored in the nodes on the hierarchy. Features about any attributes, sensors or relationships can be easily accessed by a user. The expertise of the engineering staff can be incorporated into the system as concept constraints, deductive rules and various inductive biases.

SSME simulator generates the raw data about sensors for a fault. The raw data is simply a list of time-value pairs of each sensor. All values are in real

number form. Usually this kind of data is used to plot diagrams (Figure 5.1) for representing sensor behavior, and then human experts analyze the diagrams to find the characteristics of each fault. This human analysis process is usually time-consuming when the number of diagrams is large, and is complicated when the features of faults involve the relationships between sensors. Since a human expert describes the features of a fault by a set of attributes (which are shown in Table 5.1 and Table 5.3), MLS will automatically generate those human-oriented descriptions from the raw data.

Before sending the raw data to the induction system, preprocessing is performed which smooths the curves of sensors, divides the curves of sensors into segments and denotes each segment as an event. The collection of sensor descriptions constitutes the description of an instance of an engine fault which, in turn, is used for the induction process. MLS emphasizes a significant event for each sensor, since most characteristic features of a fault exist in the significant event. The preprocessor extracts basic attribute features from the raw data (as shown in Table 5.1) and deduction process generates more attribute features to describe a fault. The derived attributes are shown in Table 5.3.

Although MLS can perform induction on a domain without all the related domain knowledge, domain knowledge makes the induction more time-efficient and produces better results. A discussion of the domain knowledge needed in the SSME application area is given below.

A basic concept is an abstraction of a class of real world entities which share common features. A real world entity is a thing (such as an animal or a computer) or a situation (such as a disease, a machine fault, or a state of a process). In the SSME fault test analysis, each type of engine fault is a basic concept. MLS assumes that every real world entity belongs to only one basic concept. In the SSME domain we assume each fault instance represents only one fault. MLS takes in the instances of the basic concepts to incrementally generalize the descriptions about the basic concepts and to classify them into clusters. The basic concepts should be the main focus of an application domain if the purpose is to find the features of the basic concepts. In the SSME fault test analysis domain, the purpose is to find features for each fault and possible classification of faults. That is why engine faults are taken as basic concepts. In some application areas where the purpose is clustering, we can use basic concepts to represent every real world entity. In this case, MLS does not perform the generalization in the basic abstraction; the main task performed is in advanced abstraction.

The matching threshold for grouping basic concepts into classes is related to the number of levels in the hierarchy. In the SSME domain, faults can be grouped into several classes such as injector faults, control faults, duct faults, manifold faults, valve faults, high pressure oxidizer turbopump faults, and high pressure fuel turbopump faults. The possible number of levels of interesting high level concepts is one or two. By this kind of domain knowledge and purpose of clustering we choose the maximal level number to be three in MLS.

A structural concept has components whose features and relationships collectively constitute the description of a concept. For example, in a block world domain each block can be a component of a basic concept. In a cancerous cell analysis domain cell bodies are components of the basic concepts – cells. In a rocket engine fault analysis domain, an engine fault is described by features of temperature, pressure, flow, speed, etc. These parameters are measured by many sensors. Therefore, sensors are the components.

In a multi-concept inductive system components usually have different relevancies with different concepts. For example, a sensor may have distinct features for an engine fault and show nothing about other faults. A large number of sensors exist in the rocket engine, but only a few of them are related to a specific fault. To pay equal attention to all sensors for every fault is inefficient. Therefore, a component in MLS can be assigned different relevancies for different concepts. In building MLS the assignment of sensor-fault relevancies depends on domain knowledge such as functional relationships and structural relationships of the engine parts as well as locations of sensors and faults. Examples of component-concept relevancies in MLS are shown in Table 5.2. In this table, s10, s22, etc. are sensors; CCV, MOV, MFV, OPOV, and FPOV are five types of engine faults. In MLS sensors can be in one of three types with respect to each engine fault: critical sensors — which show strong evidence of and are closely relate to the fault; irrelevant sensors — which do not show any changes when the fault occurs; and unspecified sensors — which may show some change with the fault and whose relationship to the fault is unknown. As indicated in Table 5.2, critical sensors are assigned a high relevancy value (10.0); irrelevant sensors are given a low relevancy value (2.0); and unspecified sensors are given a value between those two values (7.0 is assigned to a temperature sensor and 8.0 is assigned to a pressure sensor). From the domain knowledge we know that pressure sensors are usually more important in identifying a fault than temperature sensors. Thus, pressure sensors are given higher relevancy values than temperature sensor.

Relationships of component features play an

important part in the description of concepts. Components may have positional relationships, temporal relationships, or some relationships governed by domain theory. MLS allow a system developer to indicate interesting component pairs. Then, deductive rules, which derive component-relationship descriptions from component features, are automatically generated. Examples of these kinds of rules in MLS are shown in Figure 5.2. Known relationships between component features can be used as concept constraints.

Attributes are the basic vocabulary to describe basic concepts. After the identification of basic concepts the system developer needs to find out what attributes should be used in modelling the domain problem. The choice of attributes is based on the *availability* and *utility*. Availability tells what attributes can be abstracted directly from the input raw data. Usually too many attributes can be abstracted from the raw data, but only a small portion is relevant and useful to the concept description. The utility of an attribute, which tells what attributes are usually used to describe a basic concept, is determined by experts with domain knowledge. Attributes determined by domain experts may not be available, so the constructive rules should be formulated so as to derive the unavailable attributes from the available attributes.

In MLS attributes are of different importance in describing a concept. Domain knowledge can be used to assign different worth values to different attributes. For example, in the SSME domain the attribute `direction` of change is more important than the attribute `rate` of change because a sensor's direction of change is usually the same for the same fault while the rate of change can be different for different severities and durations of the fault. In many cases, relationships between attributes are more important than individual attributes. MLS represents every attribute by an object and supports the knowledge acquisition facilities to help the system developer to define attributes. The "curve-pattern" attribute has a hierarchical domain which is represented by a list representation of tree. The slot "correspondence" is the transformation rule which transforms the input values into the symbolic values (for example, the value 'rf' stands for a two-event curve pattern with the first event as 'rising' and the second event as 'falling').

Inductive rules include generalization rules and transformation rules. Generalization rules are supplied by the inductive system. Transformation rules are domain related. Domain knowledge is needed to determine how to divide a real number value-domain into categories, and what symbol represents a value range. For example, the value of attribute `temperature` can be categorized into { `high, very-high, medium, low, very-low` } . For different domains the categories may cover different value ranges. There are no universal rules of transformation; the only criteria are that categories of values should correspond to the categories of concept instances, and that symbols need to reflect the value ranges in real world applications.

Since useful attributes may not be available directly from the raw data, deductive rules are used to derive them by applying various domain knowledge such as domain theory, physical laws, operational principles and domain experience. From domain knowledge in the flight engine test, we know that the attributes "starting time", "ending time", "changing rate", and "magnitude" have little value to characterize an engine fault, because they all change with the severity or duration of an engine fault. Different faults may have the same changing rate, and faults of the same type may have different changing rates. We found some relationships have more importance in characterizing engine faults. For example, a temperature sensor and a pressure sensor at the same location of the engine have certain relationships for a specific fault. New attributes START-TIME-DIFF (difference of the starting time) and END-TIME-DIFF (difference of the ending time) are used to represent the temporal relationships; RATE-RATIO (ratio of the changing rate) and MAGNITUDE-RATIO (ratio of the magnitude) are used to represent the quantitative relationships. The derived attributes in MLS are shown in Table 5.3. Logical relationships, like the concurrency of changing trend, is represented by the logical connective AND. In the domain of SSME test analysis, human experts recognize certain features and relationships for different faults. This kind of knowledge can be used as concept constraints. For example, (assuming an open loop situation) an increasing pressure of a valve inlet will cause an increasing pressure of the valve outlet. This rule is applicable to all valve blockage faults. In MLS this rule is expressed as:

IF (direction (s23) = ?x)
THEN (direction (s27) = ?x)

and

IF (direction (s23) = ?x)
THEN (direction (s28) = ?x)

where s23 is the pressure at the outlet of the high pressure oxidizer pump booster which is also the inlet to FPOV (fuel preburner oxidizer valve) and OPOV (oxidizer preburner oxidizer valve), s27 is the pressure of the outlet of OPOV, and s28 is the pressure of the outlet of FPOV. Examples of the deductive rules in MLS are shown in Figure 5.3.

The SSME simulation data of 61 instances about five valve-blockage faults is used to run the indutive system. Attributes have different worth values and sensors have different relevancy values based on whether a sensor is a critical sensor to a fault, an

irrelevant sensor or an unspecified sensor.

The concept hierarchy of the induction is shown in Figure 5.4. We can see that the induction gives quite good clustering. There are no faults grouped with different types of faults on the second level. On the first level of the hierarchy, the clusters show strong regularity. Cluster104 corresponds to the MFV blockage fault, cluster81 corresponds to the CCV blockage fault, cluster78 corresponds to the MOV blockage fault, and cluster73 corresponds to the OPV and FPOV blockage faults. In the hierarchy we can see that the OPV and FPOV faults have similar sensor behavior. Examples of the MSL output are shown in Table 5.4. and Table 5.5. In Table 5.4. a part of the characteristic description of Main Fuel Valve (MFV) blockage fault is given, where S122, S8, S9 etc, shown in the first column of the table, are sensor labels. The rest of the columns in the table illustrate the association of attributes with the corresponding values for various sensors. The association of an attribute with a value of a sensor is called a feature. The columns two to four illustrate atomic features which involve only one attribute. The other three columns illustrate the compound features which are conjunctions of two atomic features. There are different forms of attribute values in the Table. For example, sensor S122's attribute direction has a single-form of value POS; sensor pair S38/S41's attribute Rate-ratio has a range-form of value 4..5; sensor pair S9/S10's attribute Magnitude-ratio has a or-form of value 5V1. Similarly, in Table 5.5.a part of the discriminant description of MFV is given, which consists of features possessed by MFV's instances but not by any instances of other faults.

VI. Conclusion

We have developed an inductive machine learning system, MLS, for the acquisition of knowledge about SSME faults. Given fault data from an engine simulator as input, MLS will generate characteristic, discriminant and aggregational descriptions about each engine fault. MLS also generates a concept hierarehy which groups related faults into clusters. The descriptions about each cluster are higher level descriptions of fault groups. The output from MLS can be used for assisting engineers in analyzing engine tests and for engine fault diagnosis.

We have tested MLS with 61 fault instances from the SSME Simulator. Five valve blockage faults are included in those instances. MLS can correctly classify the faults with a high rate of success. Human oriented descriptions about faults and fault groups are generated.

Domain knowledge plays an important role in MLS. More knowledge enables the learning system to become more efficient and more accurate. For its future development, more domain knowledge about the engine needs to be added. Further research efforts are also needed to combine AI techniques with traditional statistical data analysis techniques.

Acknowledgement

This work was performed within the Center for Advanced Space Propulsion (CASP) and was supported in part by NASA Grant NAGW-1195, NAG-1-513 and Rocketdyne Contract No. R04QBZ90-032709. The Center for Advanced Space Propulsion is part of The University of Tennessee–Calspan Center for Aerospace Research, a not-for-profit organization located at UTSI. The authors would like to thank A.M. Norman for his assistance during the course of this effort.

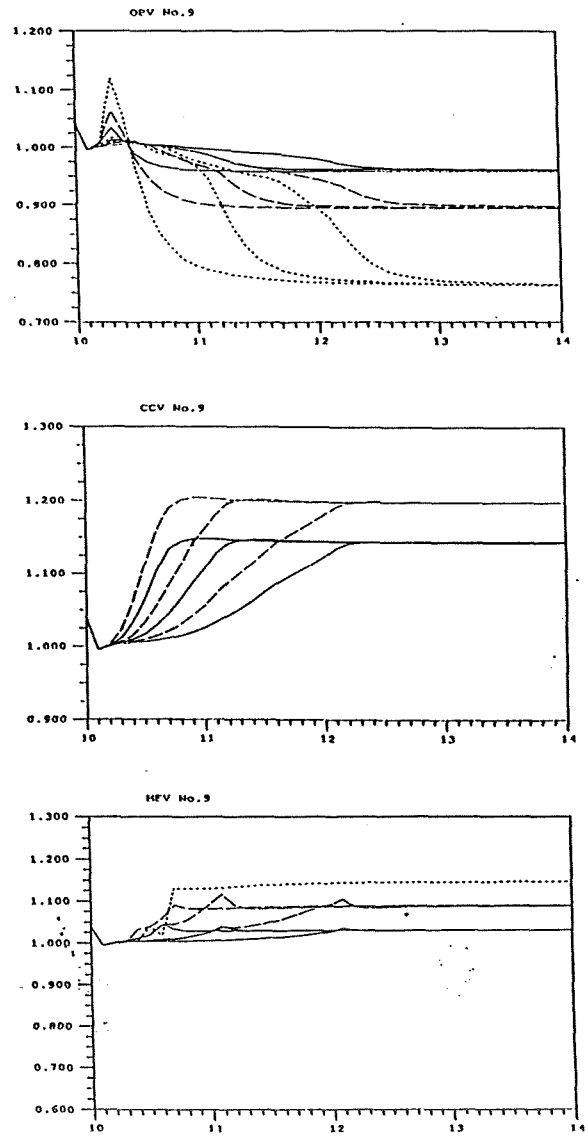


Figure 5.1 Sensor Behavior.

```
(d-rule4 if ((direction (?comp1) = pos)
             (direction (?comp2) = pos))
         then add ((increase-together (?comp1 ?comp2))))

(d-rule5 if ((direction (?comp1) = neg)
             (direction (?comp2) = neg))
         then add ((decrease-together (?comp1 ?comp2))))

(d-rule6 if ((direction (?comp1) = pos)
             (direction (?comp2) = neg))
         then add (( opposite-trend (?comp1 ?comp2))))

(d-rule7 if ((start-time (?comp1) = ?st1)
             (start-time (?comp2) = ?st2)
             (= ?st1 ?st2))
         then add ((change-concurrently (?comp1 ?comp2))))
```

Figure 5.2 Deductive Rules for Generating Component-Component Relationships.

```
(d-rule1 if ((start-time (?comp) = ?x)
             (end-time (?comp) = ?y))
         then add ((duration (?comp) = (- ?y ?x))))

(d-rule12 if ((start-time (?s1) = ?v1)
              (start-time (?s2) = ?v2)
              (<= ?v1 ?v2))
          then add ((activated-before (?s1 ?s2))))

(par4 if ((magnitude (s23) = ?m1)
          (magnitude (s27) = ?m2))
      then add
          ((magnitude-ratio (s23/s27) = (/ ?m1 ?m2))))

(d-rule11 if ((start-time (?comp1) = ?v))
          actions ((add-minmax 'first)))

(dd0 if ((direction (s9) = ?r1)
         (direction (s36) = ?r2))
     then add ((and (direction (s9) = ?r1)
                    (direction (s36) = ?r2))))

(rr0 if ((rate (s9) = ?r1)
         (rate (s36) = ?r2))
     then add ((rate-ratio (s9/s36) = (/ ?r1 ?r2))))
```

Figure 5.3 Deductive Rules in ETID.

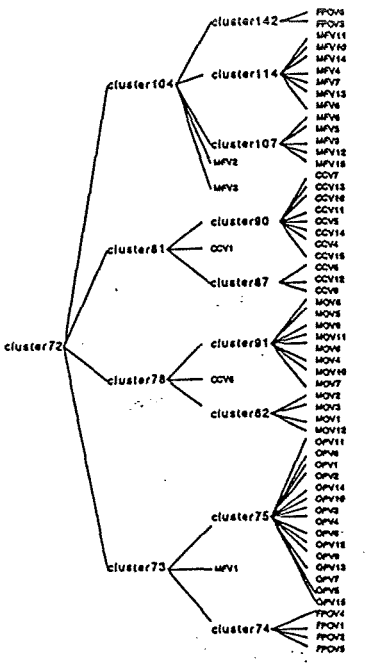


Figure 5.4 Hierarchy of Induction with Differentiating Attribute Worth and Component Relevancy.

ATTRIBUTES	EXPLANATIONS
START-TIME	starting time of an event
END-TIME	ending time of an event
RATE	average changing rate of an event
DIRECTION	indicate changing trend of an event
MAGNITUDE	the absolute amount of change of an event
AVG-VALUE	the average value of an event
VALUE-TO-NORMAL	indicate whether the average value of an event is above, same of or below the normal value
CURVE-PATTERN	indicate the general pattern of the curve
STABLE-LEVEL	the value of the sensor after it became stable
STABLE-LEVEL-TO-NORMAL	indicate whether the stable value of a sensor is above, same of or below the normal value
NUM-OF-EVENT	the number of events for a sensor
DURATION	the difference between the START-TIME and END-TIME

Table 5.1 Attributes Directly Extracted from Raw Data.

SENSORS	CCV	MOV	MFV	OPOV	FPOV
S10	10.0	8.0	10.0	8.0	8.0
S22	2.0	10.0	8.0	8.0	8.0
S30	8.0	8.0	8.0	10.0	8.0
S55	2.0	7.0	2.0	7.0	7.0
S121	7.0	7.0	2.0	7.0	7.0

Table 5.2 Sensor-Fault Relevancies.

ATTRIBUTES	EXPLANATIONS
INCREASE-TOGETHER	both events are rising together
DECREASE-TOGETHER	both events are falling together
OPPOSITE-TREND	two events have the opposite trend
CHANGE-CONCURRENTLY	two events have the same starting time
TEMPORAL-REL-TYPE1	two events have the same starting time and ending time
FIRST-START-TIME	the event has the earliest starting time
ACTIVATED-BEFORE	one event activated before the other
LARGEST-DURATION	the event has the largest duration
RATE-RATIO	the ratio of two events' rate
MAGNITUDE-RATIO	the ratio of two events' magnitude
AVG-VALUE-RATIO	the ratio of two events' AVG-VALUE
STABLE-VALUE-RATIO	the ratio of two events' stable level
START-TIME-DIFF	the difference of two events' starting time
END-TIME-DIFF	the difference of two events' ending time

Table 5.3 Derived Attributes.

sensor	direct-ion	curve-pattern	rate-ratio	stable-level-to-normal & value-to-normal	direction & direction	magnitude-ratio & stable-level-ratio
S122	POS	Rising		higher, greater		
S38/S41			4..5			
S8	POS	FF		higher, greater		
S9, S10					POS, POS	
S9/S10						5 V1, 5

Table 5.4 Characteristic Description of MFV.

sensor	direct-ion	curve-pattern	rate-ratio	stable-level-to-normal & value-to-normal	direction & direction	magnitude-ratio & stable-level-ratio
S12	NEG	FSFVFR		lower, less		
S33	NEG	Falling		lower, less		
S38/S41			4..5			
S9/S38						6, 1
S9/S10						1, 5

Table 5.5 Discriminant Description of MFV.

Development of an Intelligent Hypertext Manual for the Space Shuttle Hazardous Gas Detection System

Ching F. Lo*, George Z. Shi† and Eric Cegielski‡

Center for Advanced Space Propulsion
The University of Tennessee Space Institute
Tullahoma, TN 37388

Abstract:

This research is designed to utilize artificial intelligence (AI) technology to increase the efficiency of personnel involved with monitoring the Space Shuttle Hazardous Gas Detection System (HGDS) at the Huntsville Operations Support Center (HOSC). The project will result in a software system that has the potential to 1) enhance interpretation of real time data; 2) recognize and identify possible faults in the Space Shuttle subsystems related to hazardous gas detections; 3) locate applicable documentation related to procedures, constraints, and previous fault histories; and 4) assist in training of personnel.

The objective is to create a computerized service manual in the form of a hypertext manual and expert system which stores experts' knowledge and experience. The combination of hypertext manual and expert system for hazardous gas detection systems is termed a Hazardous Gas Intelligent Hypertext Manual (HIHM). A knowledge acquisition method is able to assist the user in modifying and updating the HIHM content and knowledge base. This will be investigated in the final phase of the project after the successful completion of a prototype. The resulting HIHM will assist the user in interpreting data timely, in identifying possible faults, in locating the applicable documentation efficiently, in training inexperienced personnel effectively and updating the HIHM frequently as required.

Introduction:

The complexity of space shuttle launch operations is well recognized. Monitoring the shuttle main propulsion system for abnormal gas leakage, for example Hydrogen gas leakage, during the pre-launch phase is one method used to detect subsystem faults. In general, Launch Commit Criteria (LCC) has been utilized as guideline to gauge the amount of leakage for various gases at different locations. Operation and Maintenance Requirements and Specification Document (OMRSD) requirements and Operation, Maintenance and Instruction (OMI) procedures have been developed to attempt, on a very limited basis, to isolate a leak prior to scrubbing a launch. Once a launch is scrubbed, more detailed isolation procedures are in place to isolate a leak prior to detanking. The Hazardous Gas Intelligent Hypertext Manual in the HOSC may display detailed mass spectrometer information and Main Propulsion System (MPS) and other information for the console engineer and management. This information includes the hardware configuration, the LCC/Haz different phase pre-launch operation time, the sensor port measurement ranges and locations, the gas species "real" concentration, and other related data.

The system will provide a significantly improved mechanism for the Hazardous Gas Detection System in data interpretation, faults recognition, and locating action documentation. The intelligent hypertext manual can also be utilized in training personnel and can be easily maintained and updated. The application scenarios of the hypertext system are:

First, the intelligent hypertext manual can furnish "quick and easy" access to appropriate information for fault detection and action options.

Second, assistance in training personnel can be provided by the large volume of accessible material for a specific area of space shuttle hardware.

Third, the management of documentation for the HGDS is greatly simplified.

Therefore, it seems apparent that an intelligent hypertext manual will provide accurate service information in a rapid and efficient manner. It eliminates most human errors and reduces the required resources including manpower and material cost to achieve the same specified goal. In short, the system will have the

This work was supported by NASA Grant NAG 8-835, NASA-MSFC, at the Center for Advanced Space Propulsion, UT Space Institute, Tullahoma, TN 37388.

* Professor of Aerospace Engineering, Principal Investigator.
† Senior Research Engineer.
‡ Graduate Research Assistant.

potential to increase productivity and the quality of service for fault detection from the HGDS.

Intelligent Hypertext Manual:

To couple an expert system with a computerized manual is an idea recently developed independently by Lo & Steinle (Ref 1) and Williamson (Ref 2). This system is contrasted with the traditional expert system which usually does not include a large volume of information. The system will place in computer memory a large volume of text/graphics documentation, such as schematics, optional actions, and service procedures in the form of hypertext database.

It should be pointed out that there are several reasons why the proposed form of an intelligent manual now becomes achievable.

1) The idea of an intelligent hypertext manual was conceived only recently (Ref 3);

2) The commercial hypertext software system became available only recently;

3) The mass storage technology, such as erasable optical disks, which is required for storing large volumes of data is also a recently available technology.

The effort for this project concentrates specifically on the engineering domain of the subsystems. However, the underlying software architecture and the other computer technologies will be applicable to other domains, such as engineering management or intelligent tutoring (Ref 4).

Implementation Consideration:

The selection of software and hardware has been completed. The decision was made to use Microsoft Windows-3 (Ref 5) to integrate the hypertext software Guide-3 (Ref. 6) and expert system shell VP-Expert (Ref. 7) or Level5 (Ref. 8). These software packages are available or already purchased. The PC-386 is available at CASP but required upgraded to have sufficient memory-RAM.

The topics selection criterion has been investigated and determined as follows. Three criteria based on to choose the topics are 1). Topics Readiness, such as documents in the computerized format and the completeness of documents. 2). Topics Priority, importance for the Hazardous Gas Detection System, 3). Support from NASA/contractor expert personnel. Based on these criteria, the Launch Commit Criteria and Background and Sensor's Location have been chosen for the initial module for the HIHM system.

Knowledge Organization:

The preliminary knowledge structure design including hypertext documents and expert systems is shown in Figure 1. The design encompasses the major information related to the Hazardous Gas Detection System from manuals,procedures, criteria, data etc. The knowledge structure of Launch Commit Criteria (LCC) and Sensors Location is part of the complete hazardous gas detection system. The LCC document is the most important and often used document during the launch monitoring and its section LCC/HAZ is specially important for the Hazardous Gas Analysis.

LCC Module:

The LCC Module has been selected as the initial effort. The construction of hypertext has been naturally started from LCC/HAZ documents. The division of LCC includes Hazardous gas sample and ground support equipment. The hypertext LCC/HAZ documents are arranged in such a way that the user can access any documents (open files) either by the document's name/I.D., or by the type of gas as shown in Figure 2. By document's name, the user opens documents by click on its name from a table of contents as shown in Figure 3. For example, Hydrogen Concentration limits chart is given in Figure 4. Some short notes (called Note Button), e.g. Note 6, are built in the document so the user can read them by pressing down the mouse button and holding it.

On the Other hand, the user select the type of gas in Figure 2, and then click on a followed-up chart which contains the measurement location and the launching operating time period for specific gas such as Hydrogen Concentration shown in Figure 5. Once a LCC/HAZ document is opened, the user can read it by scrolling it upward or downward as depicted in Figure 6. Any additional references in the current document can be opened by simply clicking on its name (built- in Reference Button) and the reference document will replace the current reading document. If the user wants to return to the original document after finishing with the reference. The user merely needs to click on a button "Go back", the reference will close and the original document will reappear.

Several figures are also included in the LCC/HAZ documents. The figures were first construct in a graphical software, and then converted and built into hypertext documents. Electronic links (buttons) are also built-in as shown in Figure 7–Hazardous Gas Detection System Schematic with a Note Button for sensor port location. The LCC/HAZ documents and LCC/GSE

(ground support Equipment) are completed and will be demonstrated at MSFC in the coming month.

Concluding Remarks:

All the documents related to the Hazardous Gas Analysis will be built to hypertext documents individually first but linked to each other as knowledge organization required. Currently, we are working on another document, "Hazardous Gas and Fire Detection System", which contains the measuring sensor's locations and the detailed descriptions of each sensor. This document will be built as the notes for LCC/HAZ documents as well as a set of separate documents. Other document, such as OMRSD and OMI, will be selectively built into hypertext form in the next year for the complete hypertext system.

An expert System will be constructed after certain amount of hypertext documents have been completed. Some database such as "Hazardous Gas Detection System Summary– All KSC Loading Operations" information will be built into the system to produce knowledge base to the expert system. The crucial point for constructing this expert system is the knowledge acquisition from NASA/Contractor experts. A close contact with NASA personnel should be on a regular bases to ensure the success of the system.

Acknowledgements:

This research project was awarded on June 18, 1990 as a NASA research grant NAG 8-835 for a period of two years. The software aspect developed at CASP in a period of 12 months was sponsored by CASP and was reported in the last CASP annual meeting.

References

1. Lo, C. F. and Steinle, F. W., "Application of Intelligent Systems to Wind Tunnel Test Facilities", AIAA Paper No. 88-0193, January, 1988.

2. Williamson, M., "On Frames, Semantic Nets, and Hypertext," Computers in Science, January/February 1988.

3. Lo, C. F. et al, "Intelligent Hypertext Systems for Aerospace Knowledge Representation," Final Report, NASA, SBIR-88-1, August 1989.

4. Lo, C. F. "An Overview of Intelligent Systems for Aerospace Ground Testing", AIAA 90-1422, June 1990.

5. Microsoft WINDOWS-3, Microsoft Corp., 1990.

6. GUIDE 3 - Professional Hypertext System, Owl International, Inc., 1990.

7. VP-Expert, Paperback Software International, 1989.

8. LEVEL5, Expert System Software, International Builders, Inc., 1990.

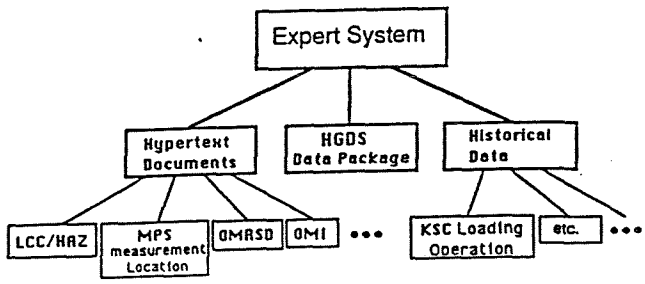


Figure 1. Knowledge Structure Design for the Hazardous Gas Detection System.

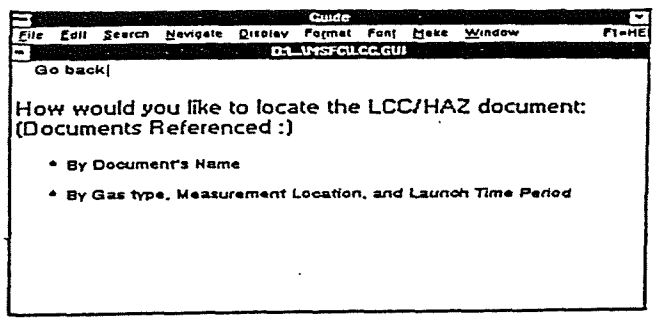


Figure 2. First Screen of LCC Module for a Typical Inquiry Session.

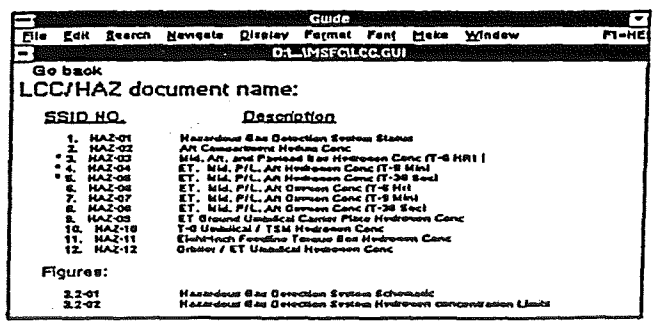


Figure 3. LCC/HAZ Document Content Screen.

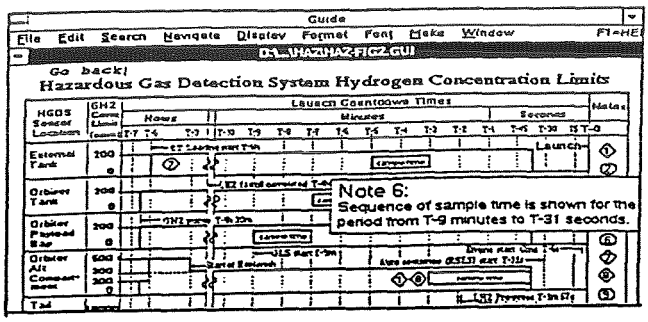


Figure 4. HGDS Hydrogen Concentration Limits Chart.

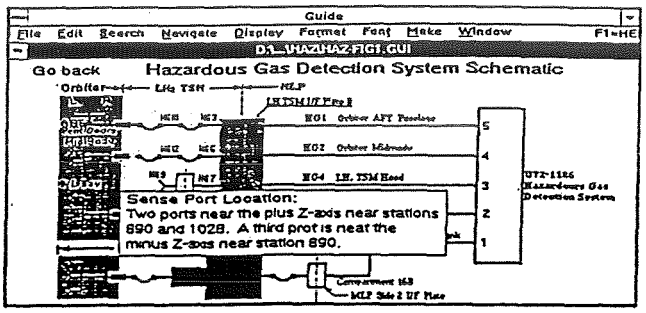


Figure 7. Schematic of the Hazardous Gas Detection System with a Note Button Display.

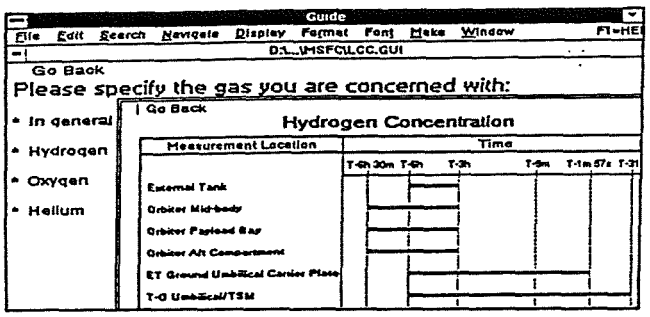


Figure 5. LCC/HAZ Selection Manual for Measurement Location and Launch Operation Time.

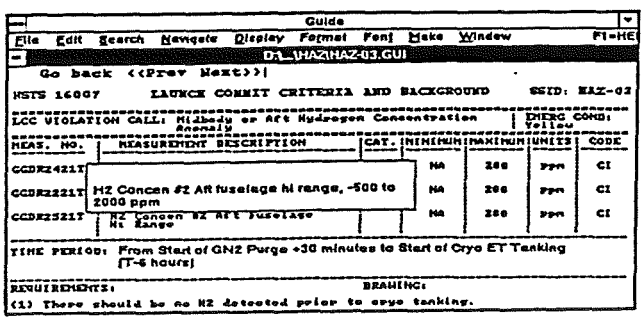


Figure 6. LCC/HAZ-03 Document with a Note Button Display.

Engineering Monitoring Expert Systems Developer

Ching F. Lo* and George Shi†

Center for Advanced Space Propulsion
The University of Tennessee Space Institute
Tullahoma, TN 37388

Abstract

This research project is designed to apply artificial intelligence technology including expert systems, dynamic interface of neural networks, and hypertext to construct an expert system developer. The developer environment is specifically suited to building expert systems which monitor the performance of ground support equipment for propulsion systems and testing facilities. Monitoring data will be acquired online through a set of measurement sensors. A dynamic database will be generated by conventional data processing techniques and/or a neural network interface module. The expert system developer, through the use of a graphics interface and a rule network, will be transparent to the user during rule constructing and data scanning of the knowledge base.

The project will result in a software system that allows its user to build specific "monitoring type" expert systems which monitor various equipments used for propulsion systems or ground testing facilities and accrues system performance information in a dynamic knowledge base.

The specific area which was chosen as an example to demonstrate and develop the general expert system developer is Compressor Stall Monitoring system. This resulted sample monitoring expert system can be implemented for its own usage. The resulting expert system developer at the end of the project has great potential for improving productivity in the construction of monitoring expert systems.

1. Objective

The objective of this research project is to design and construct an Expert System Developer (or shell) specifically for building equipment monitoring expert systems. Users can employ the Developer to build an expert system in which the knowledge base links directly to processed sensor processing through conventional techniques and neural networks. The objectives of Phase I are to design the architecture of the expert system, to identify the appropriate areas, and to implement acquired data processing methods including conventional and neural networks algorithms.

In Phase II (2nd year of effort), a prototype Expert System will be constructed which links the data processing facility constructed in Phase I to the knowledge base. The integration of results from the conventional pattern recognition algorithm and neural network output to the expert system forms the dynamic database. The prototype expert system will be demonstrated and evaluated. Updates database and refinements of the knowledge rule base will be accomplished in the process of evaluation.

In Phase III (3rd year of effort), an Expert system Developer and Tutoring Facility will be developed and will include hypertext and graphics interfaces. The effort of Phase III includes the graphic user interface for knowledge design, editing and overviewing the knowledge base. A tutorial for building and using the neural network and expert system builder will also be completed during this phase. Hypertext technology will be used for accessing and displaying information. At the end of Phase III, an expert system builder for engineering monitoring will be available.

2. Topic Identification:

The topic of stall monitoring of the 16T/S compressor has been identified as the specific area to be studied for the future construction of expert systems developer. The architecture of the expert system development environment has been designed. The neural networks model for classification and conventional pattern recognition paradigm have been investigated and selected. The prototype expert system for demonstration has been implemented on the PC MS/DOS platform successfully using the 16T/S compressor data.

This work was supported by NASA Grant NAG 2-596, NASA-ARC, at the Center for Advanced Space Propulsion, UT Space Institute, Tullahoma, TN 37388. This project is also contributed by Dr. H.H. Liu, Vanderbilt University, sponsored by CASP under Grant NAGW-1195.

* Professor of Aerospace Engineering, Principal Investigator.

† Senior Research Engineer.

In order to create an Expert System Developer, it has to start to construct specific expert systems of some selected typical topics, and later to be able to expand into an expert system developer. In the area of test facility operation, the continuous monitoring system is mostly concerned either the performance or the mechanical vibration of the rotating machinery. Thus, the following two specific areas which are the similar type problems existing at NASA/ARC and AEDC/AF have been considered to be investigated.

2.1 Compressor Stall Monitoring

The stall monitoring for the compressor in the AEDC 16-Foot Transonic/Supersonic (16T/S) wind tunnel is considered as the first candidate of the selections. The primary monitoring data are based on the time traces of rotor blade stresses during the operation of 16T/S compressors. The sensors data are recorded in the Compressor Monitoring System disk records and Compressor Monitoring Room oscillograph traces. Some rotor/stator stress data are recorded as oscillograph traces as shown in Figure 1. An early stall warning and detection expert system is intended to be constructed utilizing these time traces data and other auxiliary parameters.

2.2 Vibration Analysis for Rotating Equipment

The Vibration Analysis for Rotating Machinery in the Engine Test Facilities at AEDC is considered as the second candidate of the selections. The system should improve maintenance program based on machinery condition as diagnosed by vibration analysis. The ultimate goal is to reduce the dependence on specialists in vibration analysis and to transfer the first level of diagnosis of problems to plant personnel.

The topic of stall monitoring of the 16T/S compressor has been selected to study first, because the format of the data and the support from domain experts are readily available for immediate usage. These are the critical issues to justify the initiation of this study. The area of the vibration analysis for rotating machinery in the Engine Test Facilities at AEDC is deferred for the future application.

3. Monitoring System Architecture

A monitoring system architecture has been designed consisting of three major subsystems: knowledge-based diagnostic subsystem, neural network subsystem, and conventional algorithm data analysis and pattern recognition subsystem. Figure 2 shows the relations between these subsystems. Sensor signals usually require some type of preprocessing such as digital-to-analog conversion and filtering before they are submitted to data analysis and feature extraction.

The conventional data analysis subsystem utilizes conventional spectral, waveform and statistical techniques for data and pattern recognition analysis. This part is comprised largely of well established engineering analysis techniques. Results of data analysis are supplied to the knowledge-based subsystem which is responsible for symbolic reasoning of the diagnostic process. The knowledge-based subsystem employs heuristic knowledge such as rules acquired from domain experts for problem solving. The neural network subsystem extracts useful features and classifies data trends. Neural networks can be trained by examples and are more tolerant to noise. Such networks should be used in conjunction with conventional techniques to enhance the problem solving capability.

4. Testing Data Acquisition and Simulator

The primary monitoring data are based on the time traces of rotor blade stresses during the operation of 16T/S compressor. The data of Normal run for rotors and stators of Compressor C-1 of 16T at AEDC were recorded for various flow conditions. The flow Mach Number covered includes 0.6, 0.9, and 1.2. The original data from stress sensors were recorded on the magnetic tape in the analog format. The Rotor stress data include sensors A13 (for Rotor A and sensor #13), A17, B1, B2, C16, and C17. The Stator stress data include sensors S1, S2, and S4. The rotor and stator locations of Compressor C-1 are shown in Figure 3.

The analog data were converted to the digital format by a commercial program named STAR System and the spectrum analysis in frequency domain was also carried out by the STAR system. With the output binary data by STAR, a program for data reading and plotting was implemented on the PC. The resulted typical frequency spectrum data for A13 and C17 at Mach Number 0.6 are shown in Figure 4. More than 50 sets of digital data of amplitude vs frequency for each sensor were converted for utilizing in the techniques of Neural Network and Pattern Recognition later.

Simulator: For the rule-based expert system as well as neural network process, it is required to acquire sufficient data samples under normal and abnormal operating conditions, specially under the stall conditions. Since the most operation of 16T/S compressor is under normal condition, the stress data for the normal condition are relatively easy to acquired. On the other hand, the data under stall condition are very limited.

It is not advisable to force the compressor to enter stall situation to acquire such data. The alternative is to create a numerical Simulator which may provide those stall dynamic stress data. It is apparent that this simulator becomes necessary to generate such stall conditions data which otherwise would not be available. Therefore, the task to built a Simulator to produce the signals of dynamic stress data was completed.

The Simulator is basically the superposition of several sinusoidal, for example, sine waves form at various amplitudes, frequencies, and phase angles. Amplitudes (coefficients), frequencies and phase angles of sine waves are determined from the actual analog spectrum or, if necessary, from expert's drawing. A Fortran code of fast Fourier transform algorithm has been adopted from the existing subroutine implemented for the simulator running on an IBM-PC. The first test case is the Compressor C1 at AEDC for the rotor's "Normal" and "Rotating Stall" operation conditions. The results are very satisfactory. The rotating stall is based upon domain expert description to simulate the input data.

Another method to simulate rotational stall data has also been developed. The simulator starts from the Rotor Normal data and then incorporates inputs symptoms of rotational stall which are supplied by "domain expert." The typical Simulated results are shown in Figure 5. These simulated data will be used to develop the expert system until the real test data are obtained at later date.

5. Neural Network Models for Classification

5.1 Evaluation of Neural Network Software

The neural network software evaluated was the NeuralWorks from NeuralWare, Inc. (Ref 1). This software package seems to be quite appropriate for the project needs. First, it runs on several platforms ranging from PC, MacIntosh to Sun workstation to even N-Cube parallel computer with the same network specifications. Networks can be constructed using either the InstaNet facility or the Network Editor. Over one dozen well known networks such as Perceptron, Hopfield, Back-Propagation, and BSB are available as standard networks in InstaNet to facilitate quick prototyping. Users can define any customized networks using Network Editor. One can define the specific processing elements (PE) including transfer function and learning rule. Layer is made of processing elements, and network is made of layers and connections. Control strategy can be specified for the entire network for both learning and recall. Input/output can be done through either keyboard or file; file I/O will be more useful for the present application. The entire system of NeuralWorks is menu-driven, interactive and has plenty of graphic supports. It is rather versatile, powerful and easy-to-use. The overall evaluation about NeuralWorks is very good, and the decision has been made to use this package for the project.

5.2 Network Architecture

A three-layer back-propagation (BP) network (Ref. 2) has been selected for classification of air compressor operation conditions. The multi-layer BP networks have been studied extensively and widely used for hetero-association and pattern classification. Multi-layer networks are able to classify non-linearly separable classes. Back-propagation is the technique selected to solve the present problem. That is, the errors due to misclassification will be properly distributed to and rectified by all the connecting weights. In the present case, a three-layer—input layer, hidden layer and output layer—network is utilized as shown in Figure 6. The input layer takes the peak amplitude of stress frequency spectrum as feature inputs and sends them to the hidden layer. There are presently 11 nodes at the input layer representing 11 major peaks of frequency spectrum, 3 nodes at the output layer representing normal, near stall and deep stall condition of the air compressor, and 6 nodes at the hidden layer. Using NeuralWorks, the number of nodes, configuration and other parameters can be changed rather easily. The data from Row B of C3 Rotor Blade were supplied by Calspan/AEDC and chosen to investigate the BP since the data are available for three distinguished conditions–normal, near stall and deep stall.

5.3 Learning and Recall Procedure

Learning: Three frequency spectra for normal, near stall and deep stall conditions from Row B of C3 Rotor Blade were used for training the neural network. The network converged very quickly and the network classified the original training patterns correctly. The input/output file was set up as follows where i stands for input and d stands for desired output:

Training data for C3 Rotor Blade Row B. Data were taken from actual stress frequency spectra.

```
i    0.3800   0.4500 0.4800 0.5300 0.3800 0.3600
     0.2200   0.1600 0.1500 0.3500 0.0300
d 1.0 0.0 0.0
i    0.4700   0.4700 0.3700 0.4300 0.2800 0.4000
     0.3100   0.2400 0.2700 0.4600 0.0800
d 0.0 1.0 0.0
```

```
i    0.1800    0.1400 0.0300 0.1200 0.0050 0.1400
     0.0050    0.0050 0.0050 0.5500 0.0800
d  0.0 0.0 1.0
```

Recall: Since there was no additional data for testing the neural network, some test data based on the original training data are generated. A 5% random noise with Gaussian distribution was added to the original normal, near stall and deep stall data. Three test samples are generated for each group. All nine test data were classified correctly. If the original training data are good representative of the underlying distribution, then this test shows quite excellent performance. The graphical display for the results shown in Figure 7 represents three layers: input, hidden and output layers. The size of rectangular symbol represents the magnitude of the value of each node. The output layer Symbol 19 indicates that this result is Normal run condition given in Figure 7a. The results of near stall and deep stall conditions are shown in Figures 7b-7c.

6. Pattern Recognition Algorithm

The pattern recognition paradigm (Ref.3) divides the procedure into two phases: training and test. The basic concept of the present pattern recognition paradigm is to recognize the feature vector and apply an algorithm called Nearest Neighborhood. The feature vector is a set of measurements which are utilized to condense the description of a set of properties into a Euclidean feature space of multi-dimensions. Each point in feature space represents a value for the feature vector applied to a different category. Ordinarily, during a training phase, feature vectors from known sample data are used to partition feature space into regions representing the different classes. During a test phase the feature space partition are used to classify feature vectors from unknown data.

6.1 Minimum-Distance Pattern Classification Method

First, all feature vectors are normalized to eliminate the absolute magnitude of each signal. The class-average vector is calculated from all feature vectors of a single class. Figure 8 shows three classes A, C and R in a two-dimensional (two features) hyperplane domain. The classification of the signal is based on the minimum distance from the signal vector to each class-average vector center (one center for each class). The test of each data set will examine the distance from the data signal vector to the center of each pattern class. The data signal will be classified as that specific class which the minimum distance occurs between the data signal and one of the class centers.

6.2 Clustering Method

The basic idea about the clustering method is that the given signals can be partitioned into multiple cluster domains. For each cluster, there is a cluster center (average of all the data in that cluster). In each class, there are several clusters and each of them has a cluster center. The corresponding clusters centers will become representatives of them. The determination of a given signal belonging to certain class is based upon the signal's relations to the multiple cluster's centers.

The clustering method is preferred over the minimum distance (one center for each class) method of a single cluster in some cases. If the signals are evenly distributed over their regions, the two method are the same. But if the distributions are not perfectly even, one center cannot adequately represent its overall characteristics.

6.3 Cluster-seeking Algorithm

The Maximim (Maximum-Minimum)-distance algorithm is applied for cluster-seeking. This method is a heuristic procedure based upon the Euclidean distance concept. The algorithm consists of the following steps:

1. Arbitrarily choose a signal to be the center of the first cluster.
2. From the remaining signals, find that one whose distance to the first cluster is farthest and assign it to be the center of the second cluster.
3. For each remaining signals, calculate the distances to each center and store the minimum. Among all the minimum distances, choose the maximum. If the maximum (of the minimum group) is greater than one-half of the (largest) distances between centers, that signal becomes another center and then perform (3) again; otherwise the algorithm terminates.

After having the centers, each remaining signal is assigned to its nearest cluster center. To obtain a representative cluster center for each group, the mean of each cluster signals is designated as the new cluster center.

6.4 Cluster-Belonging Criteria

The Maximim (Maximum-Minimum)-distance algorithm for the cluster-seeking described above has been applied to the data sets of (A1306, R1306) and (C1706,

R1306). Data of A1306 and C1706 were obtained from Row-A blade-13 and Row-C blade-17 at Mach Number 0.6 Normal running condition, respectively. Data of R1306 are Simulated Rotational stall condition at Mach Number 0.6. Four clusters (therefore four cluster centers) for A1306 sets of signals were obtained and five clusters for R1306. For a given signal, two cluster-belonging criteria have established as follows:

1. Among all (total nine for A1306 and R1306) distances for a signal to cluster centers such as (A1306, R1306), the minimum one decides the class which it belongs to. (One- Minimum Nearest Neighbor)
2. The signal nearer to the distance average of cluster center belongs to the corresponding class. (Distance to Cluster Centers Average)

6.5 Results of (A1306 vs R1306) and (C1706 vs R1306):

The above cluster-belonging criteria are applied to the classification of (A1306 vs R1306) and (C1706 vs R1306). The following results obtained are listed in Table 1. Criterion-1 (One-Minimum Nearest Neighbor) can get better overall results. Criterion-2 (Distance to Cluster Centers Average) yields reasonable good results. But there is no perfect way among the above two criteria to classify the signals. For the purpose of comparison, the Minimum-distance method of non- clustering method (one circle) also is applied to the present case. The results is very close to One-Minimum criterion.

In general, the experience has shown that any reasonable methods will work for a good physical features. It is critical to select a good representative features of a physical problem. The feature vector selected for the current problem is the same as those in the neural network: the peaks of amplitude at various frequencies. Similarly the results can be obtained as expected.

7. Prototype Monitoring Expert systems

The basic parts of the prototype monitoring Expert Systems are shown in Figure 9. A spectrum analyzer for fast Fourier transform is used to process the sensor data from time domain to frequency domain by applying a commercial program STAR system. A feature extraction has developed in C code to extract the characteristics from the spectrum data obtained by the STAR system. The network weights and database classifier template were trained in neural networks and pattern recognition from the data feature characteristics. Then the new data sets are classified by neural networks in a commercial system NeuralWorks (Ref 1) and pattern recognition (Ref 3) based the pretrained weights and database template. The knowledge base, inference engine and user interface contain in a commercial expert system shell - - VP Expert (Ref 4). The user interface facility contains graphics capability to assist users examining the data characteristics. The graphics program is a compiled Microsoft Quick-Basic code. A demonstration expert system has been implemented running on any IBM PC. The functions for the expert system include the classification (identification) of monitoring test data as well as the acquisition of new data training for the storage in the weights or database template as a learning process at users (experts) requested.

The demonstration system which links output from conventional pattern recognition is constructed on the limited knowledge base in the present time to check the architecture of the proposed expert system. The task which links output from neural networks to the database of the monitoring expert system will be followed up as the next step.

8. Concluding Remarks

The prototype neural network based expert system has been implemented in the commercial tools. The combination of neural networks and conventional pattern recognition has resulted in optimal data processing procedures for the expert system.

In the future work, it appears feasible and beneficial to develop a generic monitoring expert system which would be extended to an expert system developing environment and which, in turn, would be used for constructing other specific expert systems.

Acknowledgements

This work was supported in part by NASA Grant No. NAG 2-596, funded jointly by AEDC and NASA/ARC, with Frank Steinle and G.G. Nordstrom as Technical Monitors. The authors would like to thank Dick Womack and Gary Jarrel of Calspan/AEDC and Bruce Bomar of UTSI for serving as domain expert and providing compressor's sensors data.

References

1. Neuralworks-Neural Network Software, Neural Ware. Inc.,1988

2. Rumelhart, D. E. and McClelland, J. L., Editors "Parallel Distributed Processing" Vol I, Foundations, MIT press, 1988

3. Tou, J. T. and Gonzalez, R. C. " Pattern Recognition Principles", Addison-Wesley, Reading, MA, 1974

4. VP-Expert Expert System Developer, Paperback Software, Berkeley, CA, 1989

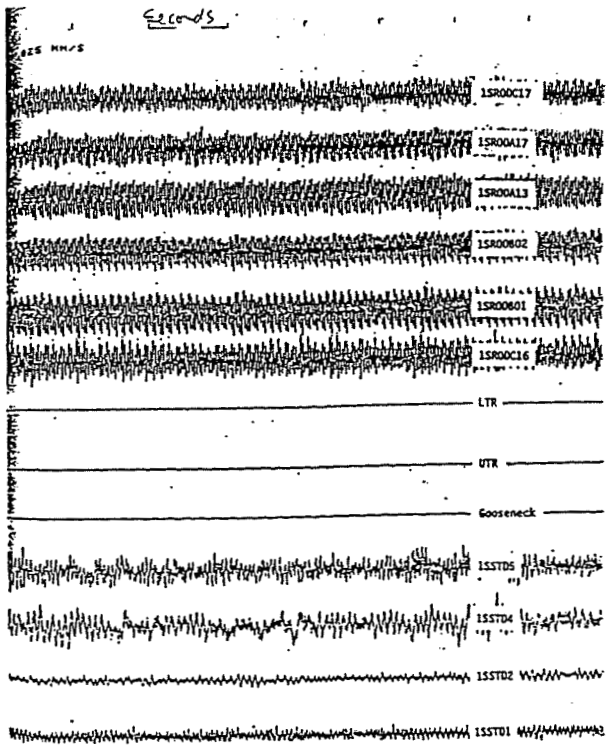


Figure 1a. Typical C1 Stress Traces (25 mm/sec).

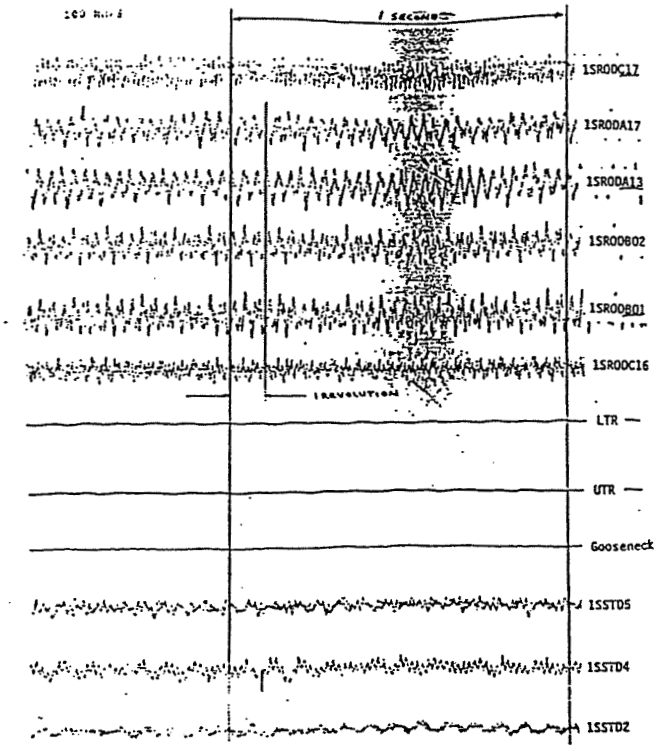


Figure 1b. Typical C1 Stress Traces (100 mm/sec).

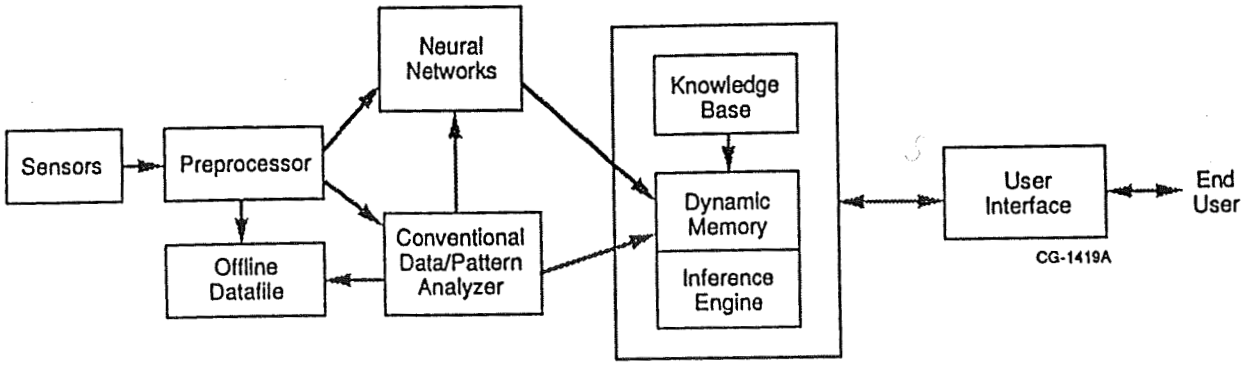


Figure 2. The Architecture of an Engineering Monitoring Expert System.

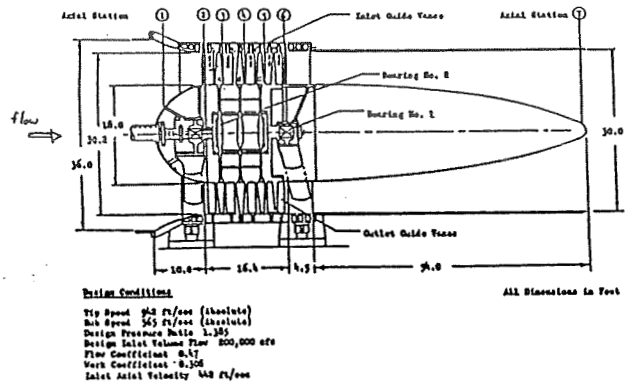


Figure 3. Dimensions for COMPRESSOR C1

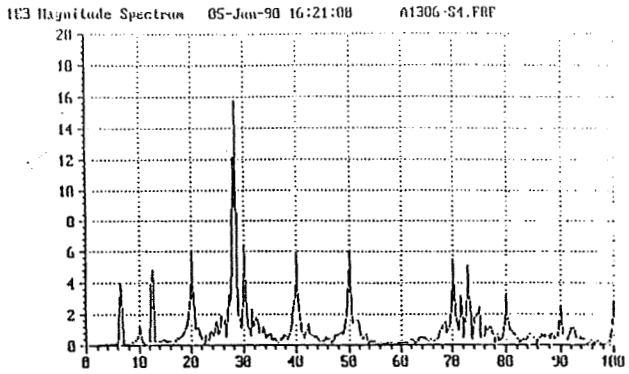


Figure 5. Similated Frequency Spectrum Data for Rotational Stall.

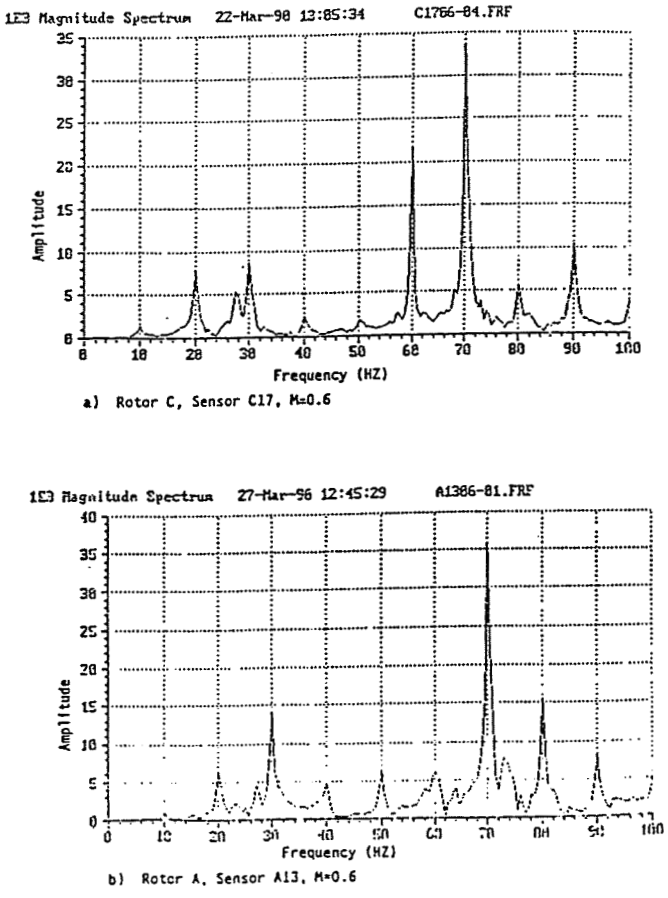


Figure 4. Typical Frequency Spectrum Data Plot.

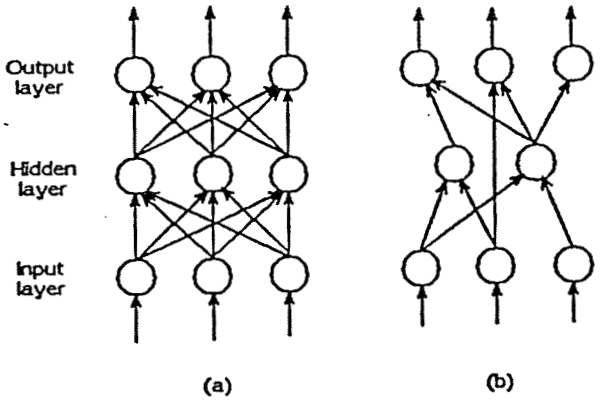


Figure 6. A Neural Network Model with (a) full connection and (b) user-defined connection.

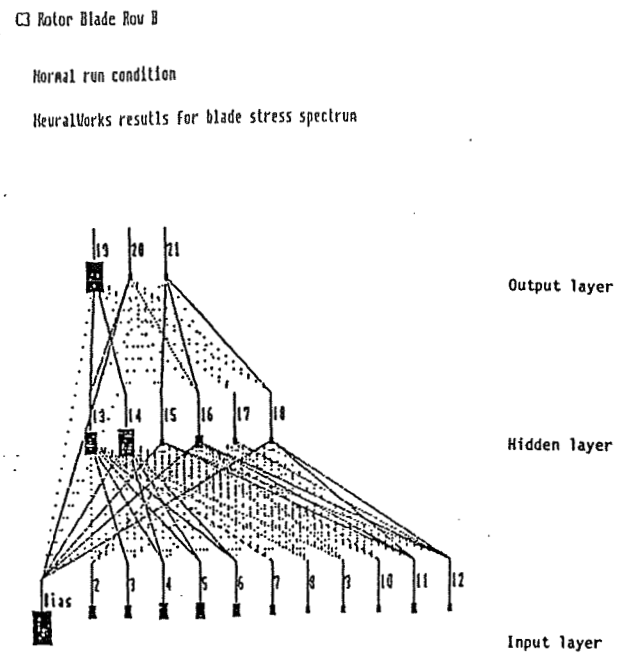


Figure 7a. A Three-layer BP Network for Rotor Blade Row B. (Normal Run Condition).

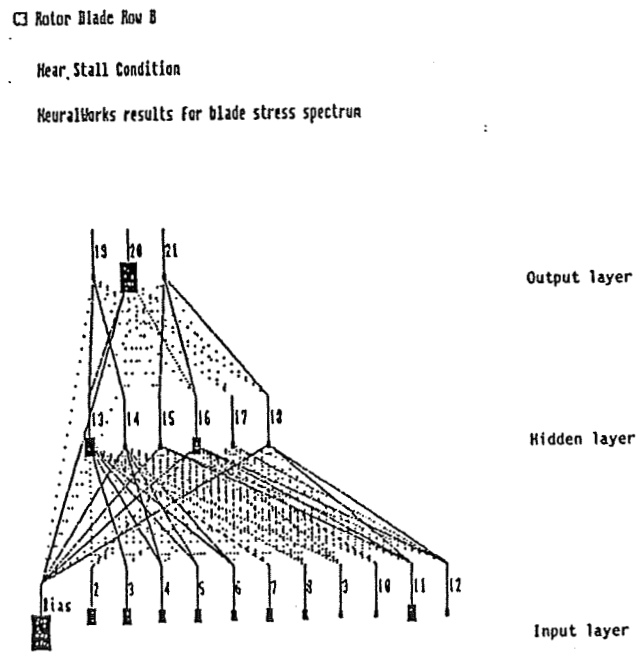


Figure 7b. A Three-layer BP Network for Rotor Blade Row B. (Near Stall Condition).

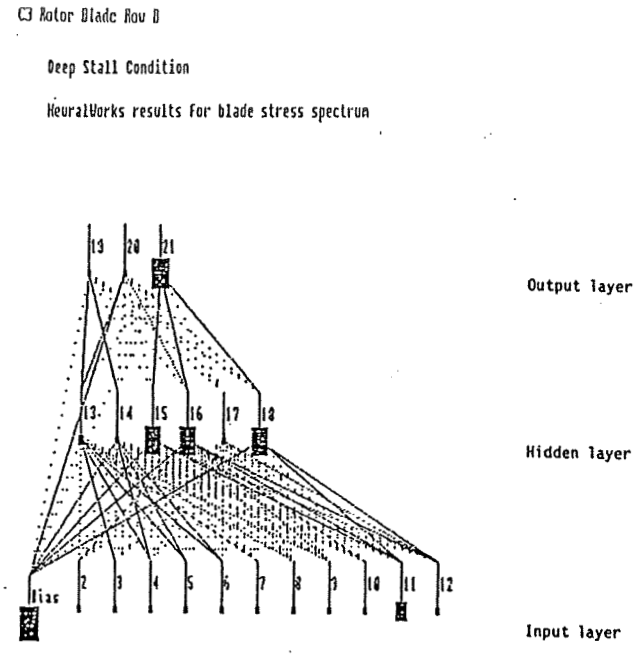


Figure 7c. A Three-layer BP Network for Rotor Blade Row B. (Deep Stall Condition).

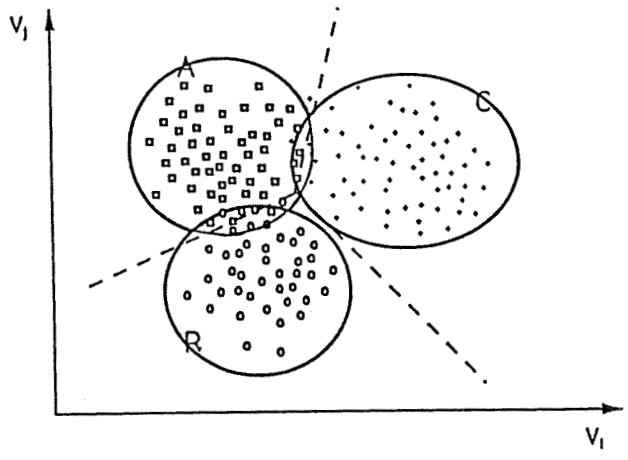


Figure 8. Pattern Classification Regions.

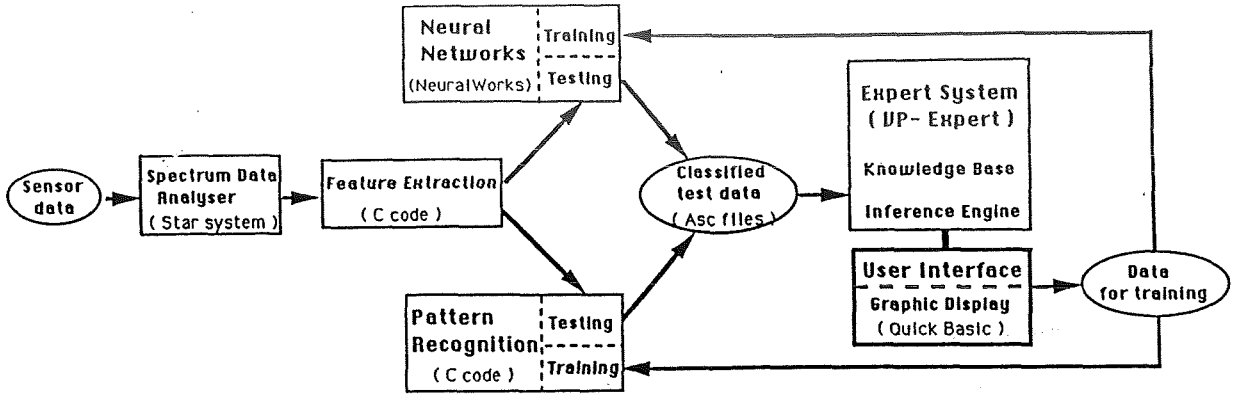


Figure 9. Prototype Expert System for Compressor Stall Monitoring.

Table I.

A1306 vs R1306

The initial point is the one of un-normalized longest from origin.

(The approach finds 4 groups in A1306, 5 groups in R1306.)

The radius of A1306 is 0.868789.
The radius of R1306 is 0.969805.

The distance between two circles is 0.322254.

	One minimum Nearest Neighbor	Distance to Cluster Centers Average
A1306	0 (100%)	0 (100%)
R1306	-1 (98%)	-5 (90%)

Remark: the minus means misclassification number.
the percentage in the bracket means correct rate.

C1706 vs R1306

The initial point is the one of un-normalized longest from origin.

The radius of C1706 is 0.867510.
The radius of R1306 is 1.036084.

The distance between two circles is 0.433841.

	One minimum Nearest Neighbor	Distance to Cluster Centers Average
C1706	0 (100%)	-4 (93%)
R1306	-2 (96%)	-4 (92%)

Remark: the minus means misclassification number.
the percentage in the bracket means correct rate.

LOW-G FLUID MANAGEMENT

Transfer Line Chilldown In Low Gravity Environment

B.N. Antar

The Subscale Orbital Fluid Transfer Experiment (SOFTE)

Frank G. Collins
Basil N. Antar

Study of Two-Phase Flow in Helical and Spiral Coils

Edward G. Keshock
An Yan
Adel Omrani

TRANSFER LINE CHILLDOWN IN LOW GRAVITY ENVIRONMENT

B. N. Antar*

Center for Advanced Space Propulsion

The University of Tennessee Space Institute

Tullahoma, TN 37388

1. Procedure:

It is well known[2] that vapor-liquid two-phase flow in a pipe may take any one of the following patterns: annular, inverted annular, slug, dispersed, bubbly and stratified. All of which, except the last, occur in vertical pipe flow, while the last is unique to horizontal pipe flows. The formation of any one of these two-phase flow patterns and the transition between any two depends very strongly on several parameters. In addition to the pipe orientation the flow thermal environment plays an important role in determining the flow patterns. Among the key parameters in vertical diabetic flow are the input liquid flow rate, the temperature difference between the saturation temperature and the pipe wall, and the direction of the input flow, whether up or down. Under terrestrial conditions, the type of vapor-liquid flow regime that occurs in any situation have been identified and classified through numerous experiments. Thus it is possible to predict the flow pattern that can occur from assessing the values of the controlling parameters. However, it is possible that some of the flow patterns enumerated above may not occur or may be modified under low-gravity conditions. There is just simply not enough data available on two-phase flows in low gravity to asses the picture.

*This work was jointly supported by The University of Tennessee Space Institute, CALSPAN, Center for Advanced Space Propulsion under NASA Grant NAGW1195, Boeing Aerospace Corporation, Seattle, WA, and The University of Tennessee Space Institute, Tullahoma, TN.

We undertake in here to formulate a rational procedure for developing a realistic predictive tool for the two-phase flow during the cooling of hot tube in low-gravity environment. The basic element for this capability will be an accurate numerical computer code with the minimum of empirical inputs. The code will then be validated using ground based experiments. Low-gravity model flight experiments will then be conducted to evaluate the dominant effects of gravity on the flow field. Finally, the computer code will be modified to reflect these effects. Each component of this program is described below.

1A. Two-Phase, Flow Numerical Model

The problem of two-phase flow modeling is quite complex even for simple flow configurations such as pipe flow. The problem is further complicated when there is exchange of heat between the the fluid and the pipe. This is the situation for the early stages of the transfer line chilldown. For low-gravity fluid transfer technology applications the essential data requirements include time and mass for full quench and the amplitude of the pressure fluctuations. The former requirement may be estimated through a lumped parameter thermal calculations while the latter requires detailed analysis of the governing conservation equations. Of course the second approach will also give a better estimate to the quench requirements. For this reason, and others that will be obvious later on, a mechanistic model requiring the least amount of empiricism was chosen for development here.

One of the earliest studies using a mechanistic model to predict transfer line chilldown

characteristics was that of Steward[3]. A more refined version of the model developed later and presented in Steward et al.[4]. In that model the flow field was described by the single phase conservation equations with continuously changing properties. The pipe flow was divided into the three sequential regions comprising of fully liquid, two-phase and fully vapor, respectively. In the two-phase segment of the model empirical transport coefficients were used for both the friction and the heat flux. Although this model did not explicitly account for the different two-phase flow patterns, the pressure peaks correlated well those measured in the laboratory. However, this model relied heavily on empirical data. Similar models were subsequently used by other investigators to predict pipe wall temperature histories as well pressure fluctuations[5,6].

The model that best describes non equilibrium two-phase flow is the mechanistic multifield model. A two-field version of such a model has been successfully used by Chan and Banerjee[7] for predicting the refilling and rewetting of a horizontal tube. Recently, Kawaji and Banerjee[8] successfully used a similar model for predicting the reflooding of a heated vertical tube. In a recent article Ng et al.[9] proposed using such a model to predict the transfer line cool-down characteristics for liquid helium applications. The code developed here for the transfer line chilldown in low-gravity fluid management applications follows the model employed by Kawaji and Banerjee[8] and is discussed in detail below.

In the mechanistic model the transient flow field in the pipe is divided into three consecutive distinct regions. It is assumed that the coolant entering the heated pipe is initially liquid which is followed by a two-phase flow region due to the heat exchange with the pipe wall and finally at a sufficient distance downstream the fluid is fully vapor. An essential requirement for the mechanistic two fluid model is that the vapor-liquid flow pattern need to be identified *apriori*. One method of identifying the flow regimes is to use local flow condition and wall temperature. Another method is to use flow regime maps such as that of Baker[10] and Scott[11]. However, for low-gravity, diabetic two-phase flow there does not exist any data from which to extrapolate a possible flow pattern. The only flow characteristic that is possible to predict under these conditions is the absence of stratification. This leaves only the flow in a vertical pipe as a possible model for the flow pattern in pipes under low-gravity conditions.

In a series of experiments to investigate the reflooding phenomena during the core cooling phase of the Loss-of-Coolant Accident in a pressurized water reactor, the following model was used. A liquid coolant, usually water, is injected in a vertical heated pipe whose temperature is much higher than saturation. In the constant injection, bottom reflood case, two distinct two-phase flow regimes were found to occur near the quench front[12]. If the coolant is injected rapidly and remains subcooled at the quench front, inverted annular flow occurs. On the other hand, if the coolant is injected slowly and becomes saturated below the quench front the flow changes to an annular droplet flow regime. These two flow regimes are sketched in Fig. 1.

In the fast reflood experiments the liquid column was found to extend above the quench front, and is separated from the wall surface by a thin vapor film. Heat transfer rates are very high in this region but would decrease significantly with height as the vapor film becomes thicker. Considerable precursory cooling takes place ahead of the quench front which is believed to account for the relatively high quench speeds observed in this regime. Further downstream the inverted annular flow regime breaks up and a dispersed flow regime is established where the vapor phase may be significantly superheated. Drops of various sizes are entrained in the vapor stream and evaporate as they move downstream.

Based on these observations, a one-dimensional, two fluid model of the two-phase region in tubular flow channel was constructed. The model is based on the solution to the volume, and time averaged conservation equations for one-dimensional, two-phase flow. Separate conservation equations are derived for each of the liquid and the vapor phases, for each regime of two-phase flow. These equations, subject to the assumption that the average of the products of the field variables is equal to the products of the averages, take the following form:

Liquid continuity:

$$\frac{\partial}{\partial t}(\alpha_l \rho_l) + \frac{\partial}{\partial z}(\alpha_l \rho_l u_l) = -\dot{m}_l''' \qquad (1)$$

Vapor continuity:

$$\frac{\partial}{\partial t}(\alpha_v \rho_v) + \frac{\partial}{\partial z}(\alpha_v \rho_v u_v) = \dot{m}_l''' \qquad (2)$$

Liquid momentum:

$$\frac{\partial}{\partial t}(\rho_l \alpha_l u_l) + \frac{\partial}{\partial z}(\rho_l \alpha_l u_l^2) + \alpha_l \frac{\partial P_l}{\partial z} - \Delta P_{li} \frac{\partial \alpha_l}{\partial z} = \tau_i''' - \dot{m}_l''' u_i + \tau_{wl}''' \qquad (3)$$

Vapor momentum:

$$\frac{\partial}{\partial t}(\rho_v \alpha_v u_v) + \frac{\partial}{\partial z}(\rho_v \alpha_v u_v^2) + \alpha_v \frac{\partial P_v}{\partial z} - \Delta P_{vi} \frac{\partial \alpha_v}{\partial z} = -\tau_i''' + \dot{m}_l''' u_i + \tau_{wv}''' \qquad (4)$$

Liquid energy:

$$\frac{\partial}{\partial t}(\rho_l \alpha_l h_l) + \frac{\partial}{\partial z}(\rho_l \alpha_l u_l h_l) = -\dot{m}_l''' h_{l,sat} + q_{s,l}''' \qquad (5)$$

Vapor energy:

$$\frac{\partial}{\partial t}(\rho_v \alpha_v h_v) + \frac{\partial}{\partial z}(\rho_v \alpha_v u_v h_v) = \dot{m}_l''' h_{v,sat} + q_{s,v}''' \qquad (6)$$

In the above equations, ρ, u, P, h, and α are the density, the averaged axial velocity, the pressure, the enthalpy, and the void fraction respectively. The independent variables for this one-dimensional model are the time, t, and the axial distance along the tube, z. The subscripts l and v denote the respective variables for the liquid and vapor phases. The term ΔP_{li} or ΔP_{vi} is the pressure difference between the liquid-vapor interface and the liquid phase or the vapor phase, respectively. The governing equations as written above contain the following unknowns:

$$u_l,\ u_v,\ \alpha_l,\ \alpha_v, \rho_l,\ \rho_v,\ P_l,\ P_v.$$

Equations similar to (1) - (6) apply to each individual two-phase flow regime with the understanding that the variables must be matched at the axial interface between any two consecutive regimes. These equations were derived using averaging procedure and the dependent variables (the unknowns shown above) are in terms of quantities averaged over the cross section of the tube. Also the averaging process gives rise to a number of source terms which need to be modeled before a solution to these equations can be found. These terms are the following:

$$\tau_{wl}''',\ \tau_{wv}''', \tau_i,\ q_{s,l}''',\ q_{s,v}''',\ \dot{m}_l''',$$

which appear on the right hand sides of equations (1) - (6). The source terms account for the radial fluxes of momentum and heat between the tube walls and the adjacent fluid and also at the liquid-vapor interfaces. These functions are normally determined empirically and depend primarily on the type of two-phase flow occurring at that location. These functions affect the solution to the averaged quantities and that they should be determined with great deal of care. Normally, in terrestrial applications these functions are modeled based on numerous, statistically meaningful measurements for each two-phase flow regimes. Alternatively, good working models for these functions may be obtained from conservation principles for each specific flow regime. In low-gravity environment, due to the lack of experimental data, only the latter method may be pursued. However, before these functions are modeled it is necessary to perform a minimum number of experiments on one-dimensional, two-phase flow in low-gravity in order to predict the most probable two-phase flow regime that occurs for a given set of operating conditions.

The specific model used in this effort is the one shown in Fig. 1 for the the fast bottom reflood case. In this model the transient two-phase flow will encompass four separate regions. Starting upstream of the quench front the flow is fully liquid with some nucleate boiling taking place just behind the front. This is superseded by an inverted annular flow region immediately downstream of the quench front. This regime is in turn superseded by a dispersed flow region which is situated behind a fully vapor region. In the fully liquid and fully vapor regions only the liquid and the vapor equations are used, respectively.

Once viable models for the source terms are formulated for each individual flow pattern it is then possible to integrate the governing equations (1)-(6) to obtain the time histories of the various averaged variables. Equations (1) - (6) are coupled, nonlinear partial differential equations whose solutions can only be obtained through numerical approximations. The numerical method chosen for solving these equa-

tions the finite difference technique. Depending on the equations themselves and whether a steady or a transient solution is required it is possible to use either an implicit or an explicit finite difference method. There are advantages as well as drawbacks to each method which will not be explored here. Now, since as outlined earlier in the transfer line chilldown process we are interested in both the initial pressure transients as well as final steady solution, a semi-implicit finite difference technique is best suited for the present case. Complete details of the numerical technique used with the results obtained from the code are given in Hedayatpour et al.[13].

2B. Ground Base Experiments:

The numerical model described above was tested as it applies to cryogenic transfer line chilldown characteristics in terrestrial environment with an experimental set up. The experiment also served the purpose of code validation. The working fluid for this experiment was purposely chosen to be a cryogen and specifically in this case it was liquid nitrogen. The test section was made up of 0.5 *in* O.D. copper tube which was nominally 12 *ft* long and kept at room temperature. This configuration allowed for the cooling liquid to be basically at saturation temperature. The test section tube was supported vertically and insulated in order to allow for better control of the experiment. The exit end of the test section was open to the atmosphere.

A test stand was constructed whose flow loop is schematically shown in Fig. 2. The cooling liquid was introduced into the test section by pressurizing the liquid nitrogen with dry nitrogen gas. The liquid mass flow rate into the test section was varied by varying the pressurant pressure. The test section was instrumented with 24 thermocouples placed in pairs at 12 *in* intervals in the axial direction along the outside wall of the tube. At each temperature monitoring location two thermocouples were placed at diametrically opposite positions along the circumference of the tube. One pressure transducer was placed at the inlet end of the test section. The flow was monitored with a flow meter placed upstream of the test section.

At the beginning of each test special care was exercised to insure that only liquid entered the test section. This was accomplished by first chilling down all the flow plumbing components up to the inlet of the test section to a temperature close to the saturation temperature. Each test commences with the opening of the inlet valve to the test section. The individual tests lasted until the temperature of last axial thermocouple steadied at the saturation temperature. Full details of the experiment, instrumentation and the results can be found in McGee[14].

2C. Low Gravity Experiments:

The primary objective of this effort is to model the transfer line chilldown transients in low-gravity environment. As discussed earlier, one of the most crucial data necessary for the type of modeling undertaken here is the precise knowledge of the two-phase flow patterns that may occur during the chilldown process. Unfortunately such data for low gravity environment is not available at the present time. To this end a sequence of three low-gravity, two-phase pipe flow experiments simulating the transfer line chilldown process have been developed. The main objective of these experiments is to provide the necessary data base on the flow patterns that occurs during the cooling of a hot tube with a liquid cryogen.

There are basically two methods of simulating low-gravity environment without accessing Earth's orbit. These are drop tubes and aircraft flight trajectories. The maximum zero-g period available at the present time in drop towers is of the order of 5 seconds while aircraft can achieve times of the order of 20 seconds. Ground based experiments on chilldown show that, depending on the flow rates and pipe lengths, a complete quench of a short length tube may be achieved in under 20 seconds. Aircraft parabolic trajectories appear to be best suited for studying the two-phase pipe flow patterns that are needed for the present effort.

There have been at least three aircraft, low-gravity experiments on two-phase flows in tubes, all under adiabatic conditions, however. The earliest was conducted by the Sundstrand Corporation under contract with NASA[15]. The test flow loop for this experiment comprised of the Sundstrand Two-Phase Thermal Management System which was instrumented for flow, temperature and pressure measurements. The two-phase flow was generated by pumping nearly saturated water into a swirl flow evaporator in which the vapor mass was varied by changing the evaporator heat load. A transparent test section was used for visual obser-

vation of the flow field. The experiments indicated the occurrence of two prevalent flow regimes, namely slug and annular flows. Upon comparing with similar ground based experiments, it was concluded that reduced gravity flow regimes and pressure drops can be satisfactorily predicted using existing two-phase flow predictive techniques.

Another low-gravity aircraft experiment was conducted by Battelle PNL with collaboration with Texas AMU[16]. The test section in this case was an instrumented 8 mm I.D. quartz tube placed in a horizontal configuration wrapped with a nicrome wire. Pressure, temperature, heat power and flow rate measurements were recorded at the test section. In addition an accelerometer and gamma ray densitometer were also used in the experiment. The working fluid was two-phase nitrogen-water at adiabatic conditions. The main data were collected through flow visualization means. It was found that for very low gas velocities, at both high and low liquid velocities, the flow regime was bubbly comprising of small bubbles dispersed nearly uniformly in a liquid field. At higher gas velocities and lower liquid velocities the flow consisted of well defined slugs separated by regions of distinct annular flow. At high flow rate tests annular flow was observed. These results should be contrasted with the stratified flow regime found to be dominant in terrestrial tests in horizontal pipes.

Dukler et al.[17] describe flow patterns observed during gas-liquid flow in conduits in low-gravity environment. They recorded visually the flow of air-water in round tubes at adiabatic conditions. Although three different patterns were observed, namely: slug, annular and bubbly flows, the flows observed may be characterized by two distinct regions, bubbly and annular.

It should be noted that all of the above experiments were conducted under adiabatic conditions where the temperature of the vapor-liquid fluid was the same as that of the tube wall. This is not the case in the chilldown process. Thus the reason for the present series of low-gravity experiments. The first of the three experiments conducted in the present study has freon 113 for the working fluid, while liquid nitrogen is used in the latter two. In all of the three experiments the main data acquisition is through flow visualization means. The test section in all three experiments is made of quartz to facilitate visual recording of the flow patterns. In the first two a 0.5 in nominal O.D. tube is used with different tube lengths. In the first experiment the test section is heated while in the latter two the test section is maintained at room temperature. All three experiments are or will be conducted aboard the NASA KC-135 aircraft in which the first experiment is scheduled for a June 1990 flight while the latter two are scheduled for October 1990 and January 1991.

Fig. 3 shows a schematic of the test flow loop for the first experiment. The test section is 12.0 mm I.D. and 1.2 m long quartz tube which is heated externally by a nichrome heating tape connected to an AC powerstat. the tube in this experiment will be inclined at 45 degrees to the floor of the aircraft. The tube wall temperature will be raised to 300 - 500 $^{\circ}C$ as operating temperature for each test. The working fluid is freon-113 (R-113), chosen for its low saturation temperature of 47.6 $^{\circ}C$ at one atmosphere. A positive displacement pump, consisting of a piston cylinder connected to a DC motor driven linear actuator, is used to inject the coolant into the test section. Varying the DC input voltage to the actuator controls the flow rate. Flow rates in the range of 2.6 - 20 mL/sec can be obtained. The pump is designed to be used in a "one-shot" manner, since the aircraft develops low-gravity conditions for only 20 to 30 seconds during each parabola. The freon vapor is condensed in an ice condenser after exiting the test section. The condenser is made of ten rectangular, hollow copper fins filled with ice contained in a lucite housing. The condensed freon is collected in a resevoir and is used for the next run.

In the second flight experiment liquid nitrogen is used as the working fluid. The test section for this experiment is made of a quartz tube 0.5 in O.D. and 3 ft long which is heated externally by a nichrome heating tape. The tube in this experiment will be positioned at 90 degrees with respect to the aircraft floor. The operating temperature in the tests for this experiment is room temperature, at approximately 70 $^{\circ}F$. A schematic diagram of the flow test loop is shown in Fig. 4. Before each parabola (test) the liquid nitrogen will be collected in an accumulator constructed of stainless steel bellows which expands to a maximum capacity of 50 mL. The LN2 flow into the test section will be controlled by positive displace-

ment of the upstream end of the bellows using dry helium for pressurant. The pressurant will be used to control the coolant flow rate using a pressure control valve. Also before each test the temperature of the test section is brought to room temperature using the heating tape. The exiting nitrogen vapor will be measured with a flow meter positioned at the exit end of the test section. The pipe leading to the flow meter will be heated to insure only vapor enters the meter. The nitrogen vapor will be dumped overboard by connecting the exit end of the flow loop to the aircraft purge line.

2. Results and Discussion:

In this section we present results for all three segments of the study undertaken here, namely, the code development, the ground based experiments and finally the flight experiment. It should be noted however, that all three segments of the study interact with each other dynamically and are not independent separate entities. In other words, results from the ground based experiments are needed to validate the code while at the same time the code requires specific data inputs from the experiments. The low-gravity flight experiments, on the other hand, will be compared with the ground based experiments to identify the effects of gravity on the flow field. Also the results of the low gravity experiments will be used to modify the code since a realistic predictive capability in low-gravity environment is the final outcome of this study.

A numerical code was developed to simulate the diabetic two phase flow in a tube as discussed in the previous section. The code was written to treat the following four specific flow regimes consecutively: fully liquid, two-phase inverted annular, two-phase dispersed and finally fully vapor. The code as developed requires the following data inputs to initiate the calculations: the initial conditions of the tube including tube wall temperature distribution, coolant mass flow rate and temperature. The type of fluid as well as the geometry and conditions of the tube are also needed. Another necessary input is the quench front history throughout the experiment. The heat transfer and friction models employed were those developed and documented in the literature appropriate for each of the four flow regimes modeled. The code solves for all the flow transients including tube wall temperature, void fraction and liquid and vapor velocities and temperatures, all functions of space and time. All of the details of the model as well as the computational requirements are fully discussed in Hedayatpour et al.[13].

Fig. 5 shows the outside wall temperature as measured and as calculated with the code at different axial locations for the experiments of Kawaji and Banerjee[18]. In that experiment the coolant was water at 80 $^{\circ}F$ and the tube wall temperature was initially set at 1000 $^{\circ}F$. The pipe in this case is 0.625 in O.D. and 12 ft long made of inconel 600. For the temperature histories shown in Fig. 5 a coolant flow rate of 3 in/sec is used.

Fig. 6 shows the outside wall temperature obtained from the calculations using the code together with the experimental temperature histories at several axial positions along the tube. The experiments in this case are the ground base tests discussed in section 2. above. The working fluid in these tests is LN2 and the tube geometry and the test flow loop are described in section 2. The results in this figure are for the case of 5 $psig$ pressurization test. The quench front history used for the numerical results of Fig. 6 is the one appropriate for this case and is shown in Fig. 8. Since no flow rate data was available for all of the tests shown for the ground base experiments a value was interpolated in order to produce the results shown in Fig. 6. It was found that the calculated temperature histories were not very sensitive to the flow rate as long as the latter was within a specific bound. The wall temperature history is shown at the three axial locations along the test section at 3, 6, and 8 ft, respectively from the entrance to the tube.

It can be clearly seen that the agreement between the calculated and the measured temperatures is not as good as that for the water case, shown in Fig. 5. There are several reasons for the apparent discrepancy. First, since no visual means were available to precisely identify the two-phase flow patterns occurring in the LN2 case it was assumed that the flow pattern was physically identical to the high flood rate pattern observed by Kawaji and Banerjee[8] for water. This is not as bad an assumption as might first appear. A precisely similar, vapor-liquid, flow pattern was also identified by Laverty and Rohsenow[19] in their experiments on liquid nitrogen flow in a vertical tube. However, it should be remembered that the inverted

annular region was found by Kawaji and Banerjee to be a function of the flow rate. It is not possible at this time to classify the flow for the case shown in Fig. 6 as fast or slow flow rate. Another problem can be identified in Fig. 6 is that the pattern of the predicted cooling rate changed from under predicting at the lower positions, the 3 *ft* position, to over predicting at the higher location, the 8 *ft* position. This inconsistency is disturbing and may indicate to errors in the assumed two-phase flow patterns used in the code. The cooling pattern calculated for the mid-tube region, however, appear to conform with the measured data reasonably well. Also, the time at which, and the magnitude of the sharp drop in temperature correlate well with the measured data.

Fig. 7 shows the tube temperature history as a function of two liquid inlet pressurization pressures. The tube wall temperature is shown as a function of time at several tube axial locations for the 5 *psig* and 2 *psig* pressurant pressure cases. It is seen that the temperature variation with time is similar at all locations and for the two different flow rates. Basically, the figure shows initial rapid cooling which then decreases followed by a sudden drop in temperature at approximately -160 ^{o}C to the quench temperature indicating that the tube has quenched at that point. All of the temperature curves shown in Fig. 7 are identical with the temperature variations obtained by other investigators for liquid nitrogen. The only characteristic difference between the two cases shown in Fig. 7 is the scale of the quench time.

Fig. 8 shows the quench front arrival time at the various axial locations along the tube for five different pressurant pressures. Fig. 9 shows the total time to quench the 12 *ft* test section again for the various pressurant pressures tried.

3. Acknowledgment:

The work described here constitutes a collaborative effort among the author and the following: Dr. Ali Hedayatpour, Prof. Frank Collins, Mr. Terry McGee, all of UTSI and Prof. M. Kawaji of the University of Toronto.

4. References:

1. Stark, J. A. Low-G Fluid Transfer Technology Study: Summary. NASA contractor report NASA CR-135020, 1976
2. Whalley, P. B. Boiling, *Condensation and Gas-Liquid Flow*. Oxford University Press, 1987.
3. Steward, W. G. Transfer Line Surge. *Advances in Cryogenics*, vol. 10,pp 313-322, 1964.
4. Steward, W. G., Smith, R. V. and Brennan, J. A. Cooldown Transients in Cryogenic Transfer Lines. *Advances in Cryogenics*, vol. 15, pp 354-363, 1970.
5. Srinivasan, K., Seshagiri Rao, V. and Krishna Murthy, M. V. Analytical and Experimental Investigation on Cool-Down of Short Cryogenic Transfer Lines. *Cryogenics*, vol. 14, pp. 489-494, 1974.
6. Bentz, M. D. Cryogen Flow Code. The Boeing Aerospace Company Document No. D180-29035-1, IRD Project 2-3633, pp 1-20, 1985.
7. Chan, A. M. C. and Banerjee, S., Refilling and Rewetting of a Hot Horizontal Tube, Part III: Application of Two-Fluid Model to Analyze Rewetting. *J. Heat Transfer*, vol. 103, pp. 653-659, 1981.
8. Kawaji, M. and Banerjee, S., Application of Multifield Model to Reflooding of a Hot Vertical Tube: Model Structure and Interfacial Phenomena. *J. Heat Transfer*, vol. 109, pp. 204-211, 1987.
9. Ng, Y. S., Lee, J. H. and Kittel, P. Proposed Mechanistic Model to Simulate Transfer Line Cool-Down Process Using Liquid Helium. *J. of Spacecraft and Rockets*, vol. 24, pp. 115-121, 1987.
10. Baker, O. Simultaneous Flow of Oil and Gas. *Oil Gas J.*, vol. 53, p 185, 1954.
11. Scott, D. S. Properties of Cocurrent Gass-Liquid Flow. *Advances in Chemical engineering*, vol. 4, pp. 199-277, 1963.
12. Kawaji, M., Ng, Y. S., Banerjee, S. and Yadigaroglu, G. Reflooding with Steady and Oscillatory Injection: Part I: Flow Regimes, Void Fraction, and Heat Transfer. *J. Heat Transfer*, vol. 107, pp. 670-678, 1985.
13. Hedayatpour, A., Antar, B. N. and Kawaji, M., Analytical and Numerical Investigation of Cryogenic Transfer Line Chilldown. AIAA paper No. AIAA 90-2373, 1990.
14. McGee, T. An Experimental Study of a Vertical Cooldown Line with LN2. M.Sc. Thesis, The University of Tennessee, 1990.
15. Hill, D. and Downing, R. S., A Study of Two-Phase Flow in Reduced Gravity Environment. Final Report DRL No. T-1884, Sundstrand Enrgy Sysytems, 1987.

16. Cuta, J. M. and Krotiuk, W. J., Reduced Gravity Boiling and Condensing Experiments simulated with the COBRA/TRAC Computer Code. Fist Nat. Fluid Dyn. Cong., AIAA paper No. AIAA 88-3634, 1988.
17. Dukler, A. E., Fabre, J. A., McQuilen, J. B. and Vernon, R., Gas-Liquid Flow at Microgravity Conditions: Flow Patterns and their Transitions. *Int. J. Multiphase Flow*, vol. 14, pp. 389-399, 1988.
18. Kawaji, M. and Banerjee, S., Two-Phase Flow Characteristics During Reflooding of a Hot Vertical Tube. EPRI report No. NP-2820, 1983.
19. Laverty, W. F. and Rohsenow, W. M., Film Boiling of Saturated Nitrogen Flowing in a vertical tube. *J. Heat Transfer*, vol. 88, pp. 90-98, 1967.

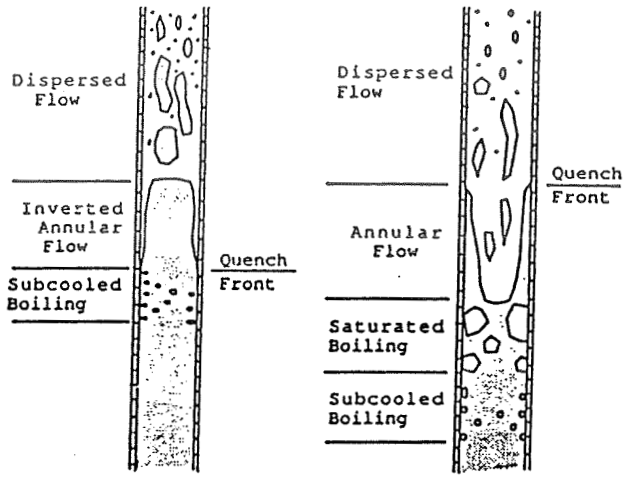


Figure 1 - Flow regimes and heat transfer in reflood: a) fast flooding; b) slow flooding.

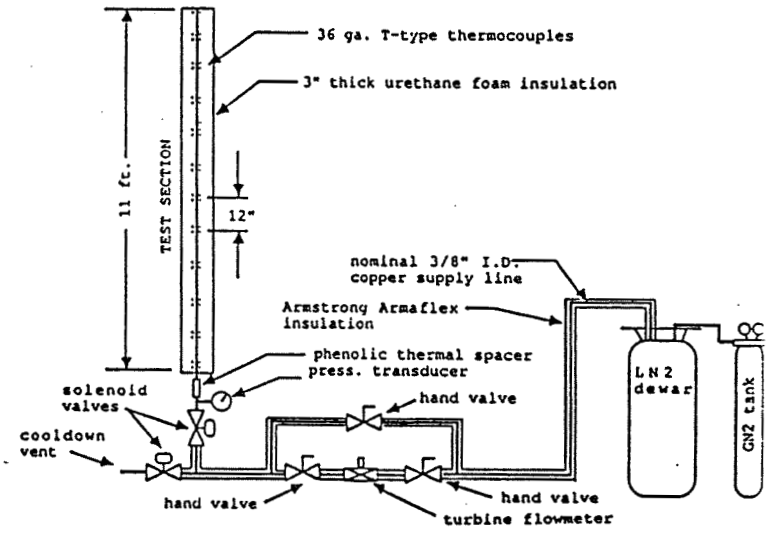


Figure 2 - Ground based liquid nitrogen experiment test flow loop

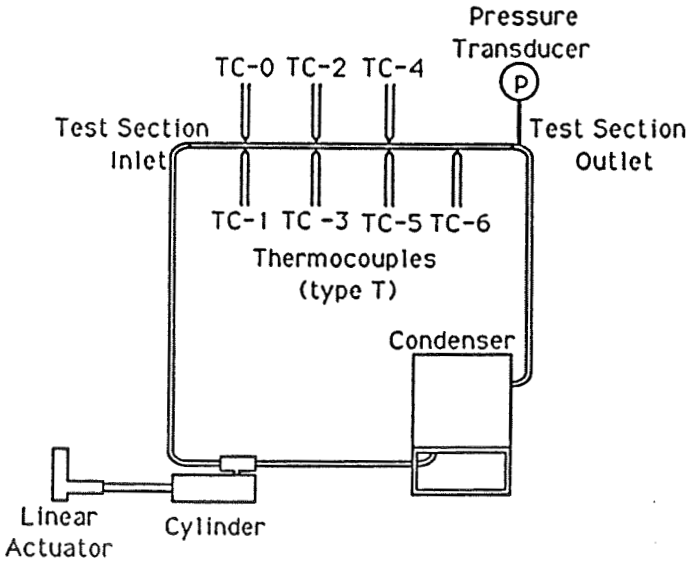


Figure 3 - KC-135 Freon 113 test flow loop

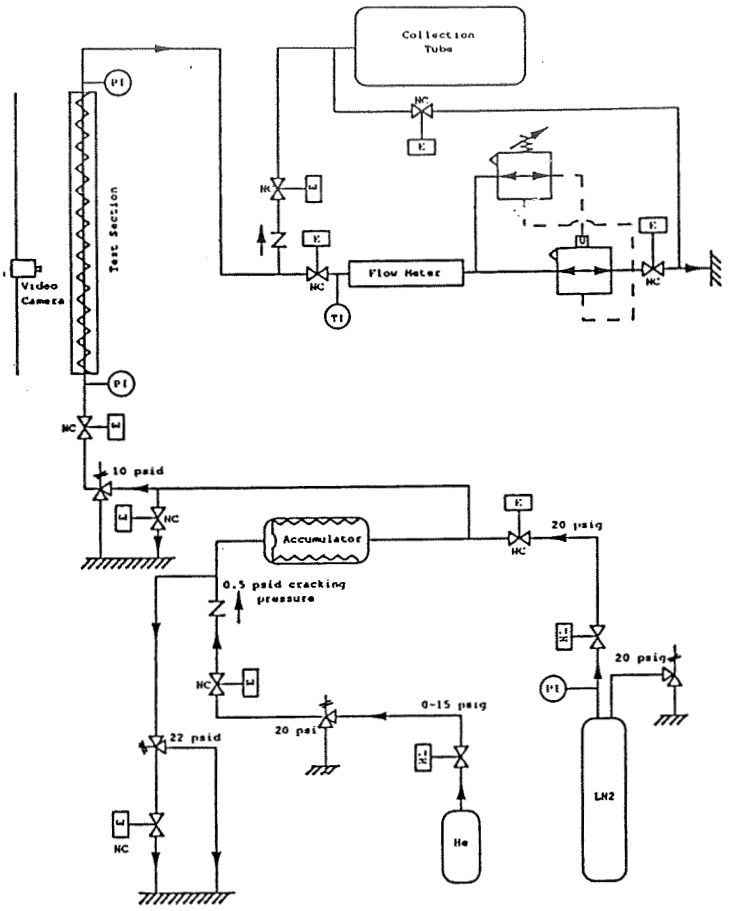


Figure 4 - KC-135 liquid nitrogen test flow loop

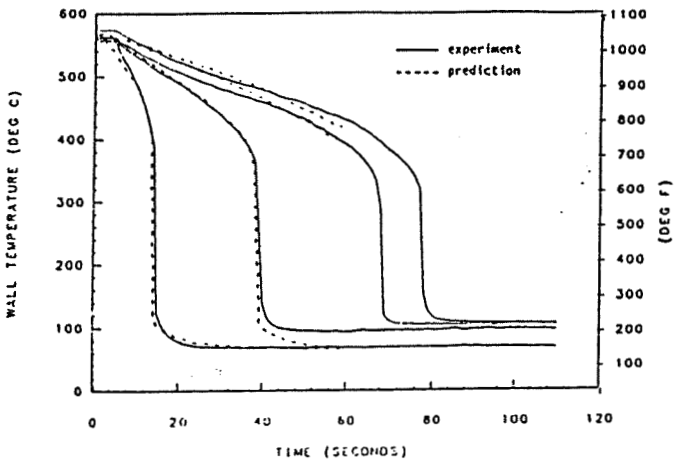


Figure 5 - Calculated and measured temperature profiles for water constant measured data from ref. 18.

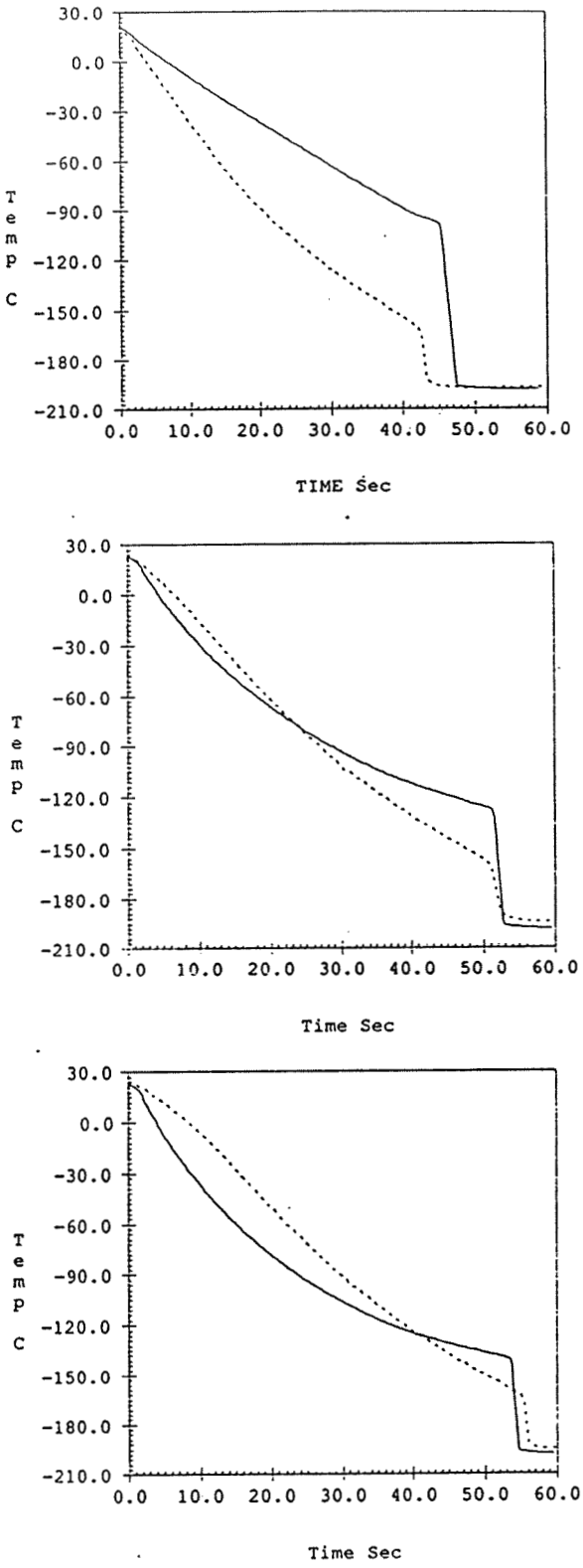
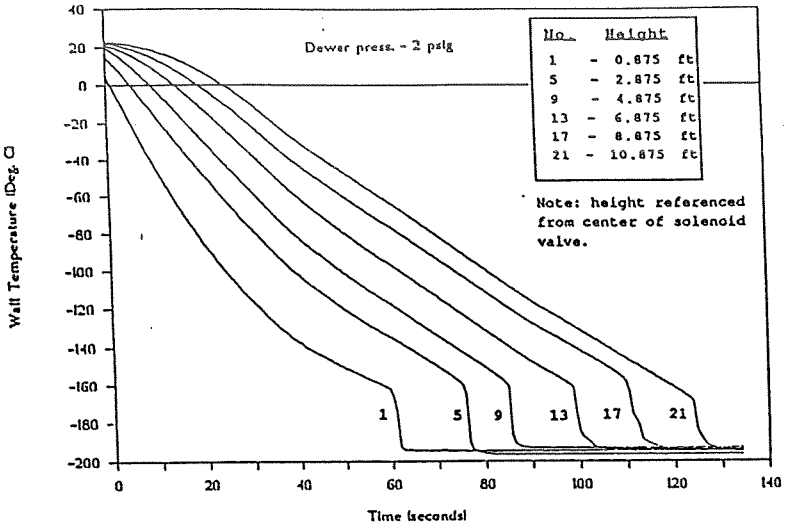


Figure 6 - Calculated and experimental temperature variations as function of time at axial three axial locations, a) 3 ft.; b) 6 ft.; c) 8 ft., from origin of test section.



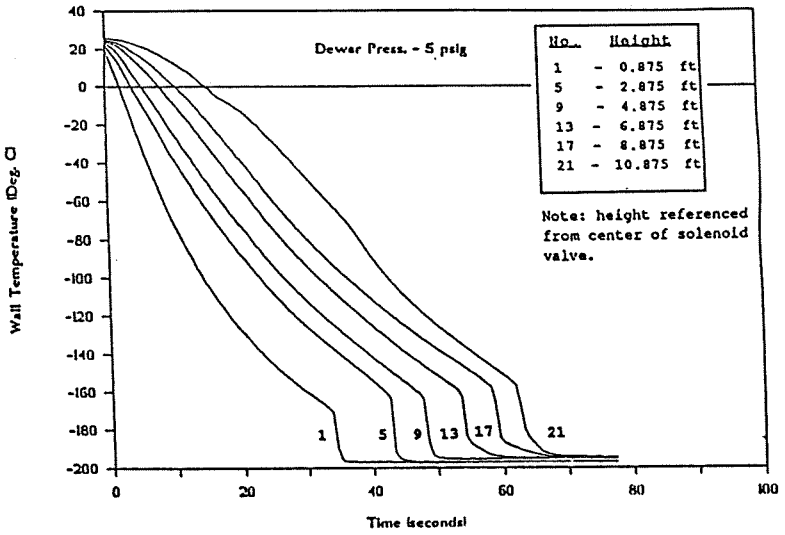


Figure 7a and 7b - Experimental temperature variations for 5 psig and 2 psig pressurant.

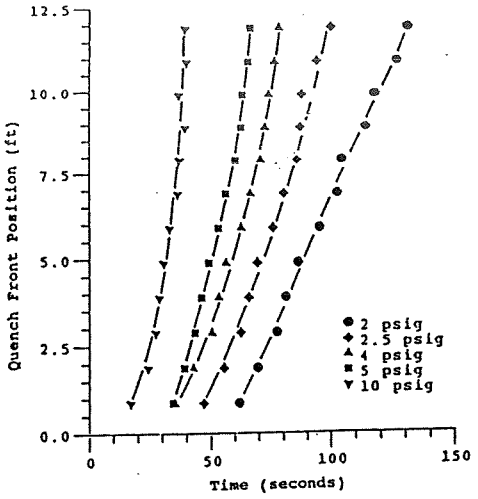


Figure 8. Experimental quench front history for liquid nitrogen tests.

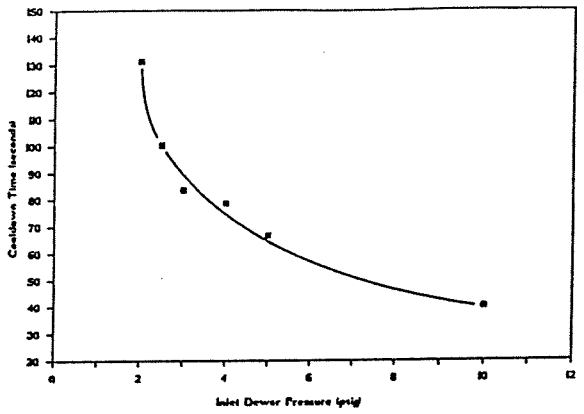


Figure 9. Total quench time for 12 ft test section as function of pressurant pressures.

THE SUBSCALE ORBITAL FLUID TRANSFER EXPERIMENT*

Frank G. Collins**
Basil N. Antar+

Center for Advanced Space Propulsion
The University of Tennessee Space Institute
Tullahoma, TN 37388

Reinhard W. Menzel++
Center for Science and Engineering
UT-Calspan Center for Aerospace Research
Tullahoma, TN 37388

Jere S. Meserole
Ogden Jones
Boeing Aerospace and Electronics
Seattle, WA 98124

Abstract

The Subscale Orbital Fluid Transfer Experiment (SOFTE) is a planned Shuttle Orbiter fluid transfer experiment. CASP performed certain aspects of the conceptual design of this experiment. The CASP work consisted of the conceptual design of the optical system, the search for alternative experimental fluids, the determination of the flow meter specifications and the examination of materials to use for a bladder that will empty one of the tanks in the experiment.

Introduction

The Subscale Orbital Fluid Transfer Experiment (SOFTE) is a planned Shuttle Orbiter experiment designed to obtain fluid-dynamic data pertinent to on-orbit liquid acquisition and tank filling operations. The experiment will mount in two Get Away Special canisters. It will involve transferring Freon 113 between two transparent plexiglas tanks. A screened-channel liquid acquisition device with plexiglas channels will be included in one tank. The experiments will examine axial and tangential no-vent fills between the two tanks plus the vapor breakdown and channel refill process of the liquid acquisition device. Video cameras will record the processes. The work at The Center for Advanced Space Propulsion consisted of certain aspects of the conceptual design of the SOFTE experiment. Only those aspects of the conceptual design performed by CASP will be described here.

Optical System Design

The primary data from the SOFTE experiment will consist of video recordings of the flow processes. The Get Away Special placed severe requirements on the lighting and recording of the fluid motion. In addition, distortion due to the curvature of the tanks posed problems. Various alternative optical systems were considered and tested on a model of the SOFTE experiment. These included miniature video cameras, either mounted external to the tank or on the tank wall, a borescope (endoscope) and a fiber optic bundle. The fiber optic bundle was rejected because of lack of resolution, large size and cost. It was shown that a video camera mounted next to the tank was almost as useful as a

*This work was jointly supported by The University of Tennessee-Calspan Center for Advanced Space Propulsion, UT Space Institute, Tullahoma, TN under NASA Grant NAGW-1195 and by Boeing Aerospace and Electronics, Seattle, WA.

**Professor of Aerospace Engineering, Principal Investigator

+ Professor of Engineering Science and Mechanics, Co-Investigator

++ System Technology Area Manager, Center for Aerospace Research

borescope. For that reason, miniature video cameras were recommended.

Alternative Fluids

A fluid that was saturated at 300K with a vapor pressure on the order of 0.5 atmosphere was desired. It was also desired to have a fluid which could simulate the fluid dynamic characteristics of cryogenic fluids during a no-vent fill in microgravity. Finally, the fluid must be safe for Shuttle Orbiter use. The baseline fluid was Freon 113. Thirty-four other fluids which satisfied the vapor pressure criteria were identified, including a DuPont alternative to Freon 113. All of these other liquids were rejected either because they were not safe for Shuttle Orbiter use or they were incompatible with the SOFTE construction materials. It was recommended that the experiment fluid remain Freon 113.

Flowmeter

The flowmeter will be the primary means of gauging the mass of fluid in each of the two tanks. Since one of the objectives of the SOFTE experiment will be to determine the maximum amount of fluid that can be transferred from one tank to the other under no-vent fill conditions, the accuracy of the flow meter was of primary concern. The flowmeter had to meet the other requirements of size, pressure drop, ruggedness and read-out requirements. All types of flowmeters were considered but only the turbine meter met most of the requirements. This meter will not operate accurately under two-phase flow conditions so means for preventing vapor from occurring within the meter were examined. To operate accurately, it was recommended that the meter be calibrated with Freon 113.

Bladder Materials

A bladder will be used to empty the receiver container. A bladder material that was compatible with the optical requirements, was not degraded by Freon 113 and was very flexible had to be identified. Various materials were tested in Freon 113 for an extended period of time. It was found that a particular polyurethane/binder material satisfied the experiment requirements.

N96- 70678

STUDY OF TWO-PHASE FLOW IN HELICAL AND SPIRAL COILS

by

Dr. Edward G. Keshock
Mechanical Engineering Department
Cleveland State University

Ms. An Yan Mr. Adel Omrani
Mechanical & Aerospace Engineering Department
University of Tennessee
Knoxville, TN

INTRODUCTION AND SCOPE OF STUDY

The principal purposes of the present study were to:

(1) Observe and develop a fundamental understanding of the flow regimes and their transitions occurring in helical and spiral coils; and

(2) Obtain pressure drop measurements of such flows, and, if possible, develop a method for predicting pressure drop in these flow geometries.

Elaborating upon the above, the general intent is to develop criteria (preferably generalized) for establishing the nature of the flow dynamics (e.g. flow patterns) and the magnitude of the pressure drop in such configurations over a range of flow rates and fluid properties.

Additionally, the visualization and identification of flow patterns were a fundamental objective of the study. From a practical standpoint, the conditions under which an annular flow pattern exists is of particular practical importance.

In the possible practical applications which would implement these geometries, the working fluids are likely to be refrigerant fluids. In the present study the working fluids were an air-water mixture, and refrigerant 113 (R-113).

In order to obtain records of flow patterns and their transitions, video photography was employed extensively. Pressure drop measurements were made using pressure differential transducers connected across pressure taps in lines immediately preceding and following the various test sections.

PRINCIPAL OBJECTIVES OF STUDY

Flow regime identification was made using video photography and direct visual observation for each test condition. These observations were presented in terms of plotted data points on various existing flow regime maps, as in [1]. Of particular interest were the conditions at which the transition to an annular flow pattern occurs for each given liquid flow rate. One of the ultimate desirable results of a study such as the present one is to characterize the flow pattern behavior in terms of a parameterized flow map accounting for the range of fluid dynamics and fluid properties. (Several maps of such data obtained in the present study are presented herein. Efforts are in progress in attempting to generate a generalized flow map applicable to flow in coiled systems.)

The pressure drop measurements obtained in the present study will be compared with predictions of several models appearing in the literature, as has been done in [2]. An attempt will be made to utilize the measured results to develop a model for accurately predicting pressure drop in coiled geometries. (This data presentation and comparison is currently in progress.)

This work was jointly supported by the University of Tennessee-Calspan Center for Advanced Space Propulsion, UT Space Institute, Tullahoma, TN under NASA Grant NAGW-1195, and Sundstrand Corporation, Rockford, IL.

EXPERIMENTAL PHASE

An existing two-phase flow loop within the University of Tennessee (Knoxville) Mechanical and Aerospace Engineering Department Laboratories was used to generate a two-phase, liquid/vapor R-113 flow over a range of thermodynamic qualities. This flow served as the inlet flow to individual tests of a variety of transparent coiled test sections. Similarly, using an apparatus (instrumented flow loop) provided by the Sundstrand Corporation, air/water flows were also studied using the same coiled test sections. Schematic diagrams of each of the test loops are presented in Figures 1 and 2.

Description of Test Sections

Four different helical coil test sections were utilized in the study. The first test section consisted of a 0.325" (8.255mm) I.D. glass tube, wound into a helical coil of pitch 2.4 cm, height 16 cm, and coil diameter of 7.5 cm. The total length of the glass tubing was 1.59 meters.

Three additional helical coil test sections were fabricated from plexiglas. Rectangular cross-section channels were machined into the outer surface of solid plexiglas cylinders, and then covered with a snug-fit outer plexiglas tube. One flow channel cross-section was very nearly square (8.89 mm height, H, by 8.255 mm width, W). The coil diameter was 7.6 cm, the pitch , P, was 2.54 cm, and the total coil length was 2.01 m. The second had dimensions of H = 6.88 mm, W = 10.3 mm, P = 2.54 cm, and a total coil length of 2.04 m.

Dimensions of the third coil were H = 10.3 mm, W = 6.88 mm, P = 2.54 cm, and total coil length of 2.28 m. Thus, for the three rectangular cross-section test sections, H/W ratios were approximately 1, 0.667, and 1.5. In all of the rectangular cross-section coils a short circular cross-section tube was inserted into the test piece body, so that a short transition section existed between the circular cross-section insert and the rectangular cross-section helical coil flow channel.

Finally, a spiral coil glass test section was also investigated. The inner diameter of the tube was 8.26 mm. The innermost spiral was of diameter (approximately) 5.5 cm, while the outermost spiral diameter was approximately 16 cm. The total length of the coil was 1.49 m.

In order to make measurements of pressure drop, pressure differential transducers were connected across pressure taps immediately preceding lengths of rubber tubing that were connected (clamped) to the test sections. The distance between the inlet tap and the inlet of the coil test sections was 19 cm. The distance between the outlet tap and the exit of the test sections differed for the different test sections. The complete listing of dimensions associated with each of the test sections tested in the study is not presented here, but will appear in the final project report.

Methodology and Procedures

General:

The test fluids and conditions anticipated to be tested in the original project specifications called for (1) air-water at a nominal temperature and pressure of 70 F and 30 psia, and (2) refrigerant-113 nominally at 130 F and 18 psia. The range of liquid flow rates to be investigated were to vary from 0.02 to 0.20 gpm. By adjusting the vapor phase flow rate while maintaining a constant liquid phase flow rate, a "quality" range from 0 to 50% was anticipated to be studied at five separate liquid flow rates, resulting in a total of 50 data points to be obtained for each test section. For the five test sections and two working fluids, the total number of data points to be obtained would be 250. No specification was made as to the orientation of the test pieces or the directions of flow.

The original intent was that each of the preceding tests would be conducted using a single orientation and a single flow direction (upward or downward, for example, with either a horizontally or vertically oriented test section).

After the test program was initiated, however, additional requests were made to expand the scope of the study. Consequently, for the air-water tests, all helical coil test sections were studied in both the vertical and horizontal orientations, with the inlet flow being in an upward direction Additionally, in one test of the circular cross-section helical coil oriented vertically, measurements were made for a downward flow. Also, for the R-113 tests, a very large number of test conditions was utilized to obtain a complete "picture" of the trend of pressure drop data over a range of liquid flow rates and vapor qualities. (In this test alone a total of 136 data points were obtained -- more than half of the total number

of data points originally specified for the entire test program.) Consequently, the total number of test conditions far exceeded the 250 values originally planned(actual number, 609).

Air-Water Test Procedures:

The apparatus depicted schematically in Figure 1 was first filled with water and circulated about the loop to flush out any contaminants before draining, refilling the loop and repeating the process again. After taking this means to insure the cleanliness of the system, liquid was circulated at a predetermined, constant mass flow rate. The air flow rate was adjusted so as to create a slug flow within the coiled test section that was slightly below a condition of annular flow. After making videotapes and pressure drop measurements of this flow condition, the vapor flow was increased carefully up to the point at which it was judged that a hydrodynamically defined annular flow just existed. This was designated as the transition condition. The air flow was increased again to a level somewhat above the transition condition. Additional operating conditions spanning the highest and lowest operating points (in terms of vapor quality) were attained, where additional measurements of pressure differential and flow observations were made.

The air flow was then stopped, the liquid flow rate adjusted to the next higher predesignated flow rate, and a range of air flow rates was spanned so as to define the transition condition at that flow rate, and obtain pressure drop data over a range of vapor phase qualities.

These procedures were repeated for each of five test sections. In the case of the helical coils, tests were conducted with the coil axis oriented vertically and horizontally. In the vertical orientation the flow was directed upward through the coil section.

R-113 Test Procedures:

For the refrigerant tests the general procedure was similar to that of the air/water tests. However, the vapor phase flow was generated by means of an electrically heated circular tube preceding the coiled test section(s), as depicted in the schematic diagram of Figure 2. At the beginning of a test run, the flow loop was completely filled with liquid, and circulated through it. Power was applied to the electrically heated tube, generating vapor. Vapor was vented through a release valve at the uppermost part of the loop to insure that any trapped air would be eventually discharged from the system. After so evacuating air from the system, the liquid flow was adjusted to a fixed rate, and power applied to generate a predetermined vapor quality, resulting in a slug flow condition, just below an annular flow condition.

The same procedure as in the air/water tests was followed to generate measurements and observations over a predetermined span of test conditions. It should be noted that at each test condition equilibrium conditions were attained before taking data, i.e. the liquid saturation temperature was required to be (at) the saturation temperature of the refrigerant corresponding to the measured system pressure.

DATA

PRESSURE DROP

The complete set of pressure drop measurements obtained cannot be presented here in report form due to the proprietary nature of the information, but will only be presented in visual display form. Additionally, a significant amount of the raw data is still being corrected for further analysis and comparison with predicted values. For example, because of the straight lengths of tubing preceding and following the coil test sections, a pressure drop associated with those lengths must be subtracted from the measured values across the entire flow length (inlet, test section, and exit). The average pressure drop along these straight lengths computed by the homogeneous and Lockhart-Martinelli methods are subtracted from the overall measured pressure drop. Also, since there is a difference in elevation between the inlet and exit of the helically coiled test sections, whether oriented horizontally or vertically, the measured pressure differentials must be corrected for the elevation head existing in each situation.

Once having obtained the true, corrected pressure drop measurements across the coils, these values are compared with several conventional methods of predicting two-phase pressure drop, such as Lockhart-Martinelli[3], Chisholm[4], homogeneous [see 2], and others. However, all such methods are intended to be applicable to straight channel flows, and so cannot take into account any effects that channel curvature might have upon pressure drop. Such methods of modifying straight channel pressure drops for applicability to curved geometries have been

proposed by Hart et al [5] and Lopina and Bergles [6]. Comparisons of the corrected measured pressure drop data are also made to these predictive methods. Also, a method has been devised and will be utilized for predicting two-phase pressure drops in helical coil geometries based upon the method developed by Hart et al [5] for single phase flow, for comparison with the measured values.

FLOW REGIME OBSERVATIONS

Flow pattern observations were made at each test condition (i.e. 609 data points). The determinations were generally made from direct visual observations of the flow. However, this became possible only after coordination of these visual observations with the video photographic results which were played back and scrutinized in detail, including freeze-frame analysis of the 1/1000 shutter speed video frames.

Each test sequence recorded on video tape contains a voice description of the test conditions, the test run, and a qualitative description of the flow details observed by the naked eye for that sequence.

Each of the flow patterns observed is identified in terms of the conventionally accepted descriptive nomenclature It should be pointed out that an annular flow was identified as that condition for which a continuous liquid film existed, or was sustained, about the entire interior periphery of the flow channel wall. Just prior to the transition from a slug flow to annular flow, a surge, or wave of liquid might wet the upper portion of the channel wall. However, the liquid film would drain downward from the upper wall, resulting in a dry wall condition until the next surge or wave would rewet the wall completely. Such an intermittent hydrodynamic condition cannot reasonably be interpreted as an annular flow, but the point at which the periphery is continuously covered with a liquid film can very logically be considered the transition condition to an annular flow pattern. The development of flow regime maps, in fact, are based entirely upon such logical hydrodynamic definitions or descriptions of flow patterns.

Whether or not such a transition represents an optimum condition for heat transfer may be another matter, however. Generally, a higher rate of heat transfer occurs when a liquid film is wavy, i.e. has an irregular interface, rather than a smooth one. Of course in an annular flow occurring under earth gravity conditions, the downward liquid drainage that occurs results in a larger thickness of liquid along the channel bottom. Generally, it is this portion that develops a wavy or irregular character first, while at higher and higher vapor flow rates (for the same liquid flow rate), the thinner regions of the annular liquid film then become wavy. Consequently, one might attempt to identify a type of annular flow condition that has special relevance with respect to heat transfer -- that condition for which the film is wavy or irregular about the entire channel wall periphery. However, flow pattern maps, which are intended to identify flow regime transitions are not based upon heat transfer criteria, but upon hydrodynamic criteria. The slug flow/annular flow transitions identified in the present study are based upon the conventional hydrodynamically-based flow pattern descriptions.

The flow regime observations obtained in the present study have been compared with several different flow pattern maps[7-10], but may only be presented in terms of visual projections due to the proprietary nature of the information. One of the flow maps [10] has been developed by Uddin specifically for helical coil geometries. Attempts were made to obtain flow visualization records of the details of secondary flows that have been postulated to occur in curved geometry single- and two-phase systems. Fluorescein dye injection at the inlet to helical coil test sections was used, in combination with high speed photography, video photography, and laser light sheet illumination. None of these techniques proved successful, primarily because of the rapid diffusion of the dye before entering the coiled section itself.

DISCUSSION OF RESULTS

PRESSURE DROP

The actual pressure drop across the coiled test sections are in the process of being extracted from the raw data.

Therefore it is not possible at this time to draw any conclusions regarding trends or generalizations relative to pressure drop. Additionally, however, the proprietary nature prevents any discussion other than that presented in connection with a visual display of the data.

FLOW REGIMES AND FLOW REGIME TRANSITIONS

For every operating condition the flow pattern within the coil test section was identified. Each operating condition thus constituted a data point on one of several flow pattern maps [7-10]. The maps of Baker, Choe and Taite-Dukler are all applicable to horizontally oriented flow channels, and so should not really be expected to predict flow patterns in coiled geometries, which furthermore are slightly inclined with the horizontal in the case of the vertically oriented helical coil systems. The map proposed by Uddin, however, was developed specifically for vertically oriented helical coils, for upward directed two-phase air-water flows. It is worth noting that even for the Uddin map, however, significant differences with the present studies exist. First, only air-water flows were studied by Uddin, with the basis for his map being the experimental observations of air-water flows. Additionally, the coil geometries studied therein had a coil diameter to tube radius ratio (R/r) of 20 to 25, whereas the ratio in the present study was between 9 and 10.

Preliminary examination of the flow regime data indicate that the actual transition (slug-to-annular flow) line for R-113 tests is significantly lower than that predicted by Uddin's transition line. Since the map is semiempirical, and based upon air-water data, it is reasonable to consider modifications to the map which account for variations in properties of fluids, as has been done for the Baker map [9], and the Mandhane map [11], for example.

Preliminary examination of the data seems to indicate little difference in the transition line for circular- or square-cross section channels. The implication is that the hydraulic diameter is the factor of significance in the transition criteria, rather than the channel shape.

Additional observations and conclusions must be deferred until all of the flow pattern maps are available for scrutiny.

REFERENCES

1. "A Reduced Gravity Flight Experiment: Observed Flow Regimes and Pressure Drops of Vapor-Liquid Flow in Adiabatic Piping," I.Y. Chen, R.S. Downing, R. Parish and E.G. Keshock, AIChE Symposium Series, V. 84, No. 263, 1988.

2. " An Experimental Study and Predictions of Two-Phase Pressure Drop in Microgravity," I.Y. Chen, R.S. Downing, E>G. Keshock and M.M. Alsharif, AIAA Aerospace Sciences Meeting, Reno, Nevada, Jan., 1989 (submitted for publication to AIAA Jnl. Thermophysics.)

3. "Proposed Correlations of Data for Isothermal Two-Phase Two-Component Flow in Pipes,' Chem. Eng. Prog., V. 45, pp. 39-48, 1949.

4. "Pressure Gradients Due to Friction During The Flow of Evaporating Two-Phase Mixtures in Smooth Tubes and Channels," Int. J. Heat Mass Transfer, V. 16, pp. 347-358, 1973.

5. "Single- and Two-Phase Flow Through Helically Coiled Tubes," J. Hart, J. Ellenberger and P. J. Hamersma, Chem. Eng. Sci., V. 43, No. 4, 1988, pp. 775-783.

6. "Heat Transfer and Pressure Drop in Tape Generated Swirl Flow," R. F. Lopina, A. E. Bergles, WRDC Technical Report No. 702891-47, 1967.

7. "Observation and Correlation of Flow Pattern Transitions in Horizontal Cocurrent Gas-Liquid Flow," W. G. Choe, L. Weinberg, and J. Weismann, Two-Phase Transport and Reactor Safety, Hemisphere Publishing Co., 1978.

8. "A Model for Predicting Flow Regime Transitions in Horizontal and Near Horizontal Gas-Liquid Flow," Y. Taitel and W. E. Dukler, AIChE Jnl., V. 22, No. 1, Jan. 1976, pp. 47-55.

9. "Simultaneous Flow of Oil and Gas," O. Baker, Oil & Gas Jnl., V. 53, No. 185, July, 1954.

10. "A Model to Predict Two-Phase Flow Patterns in Helically Coiled Tubes," A. K. Uddin, Particulate Phenomena and Multiphase Transport, V. I, Hemisphere Publishing Co., 1988.

11. "A Flow Pattern Map for Gas-Liquid Flow in Horizontal Pipes", M. M. Mandhane, et al, Int. Jnl. Multiphase Flow, V. 1, pp 537-553, 1974.

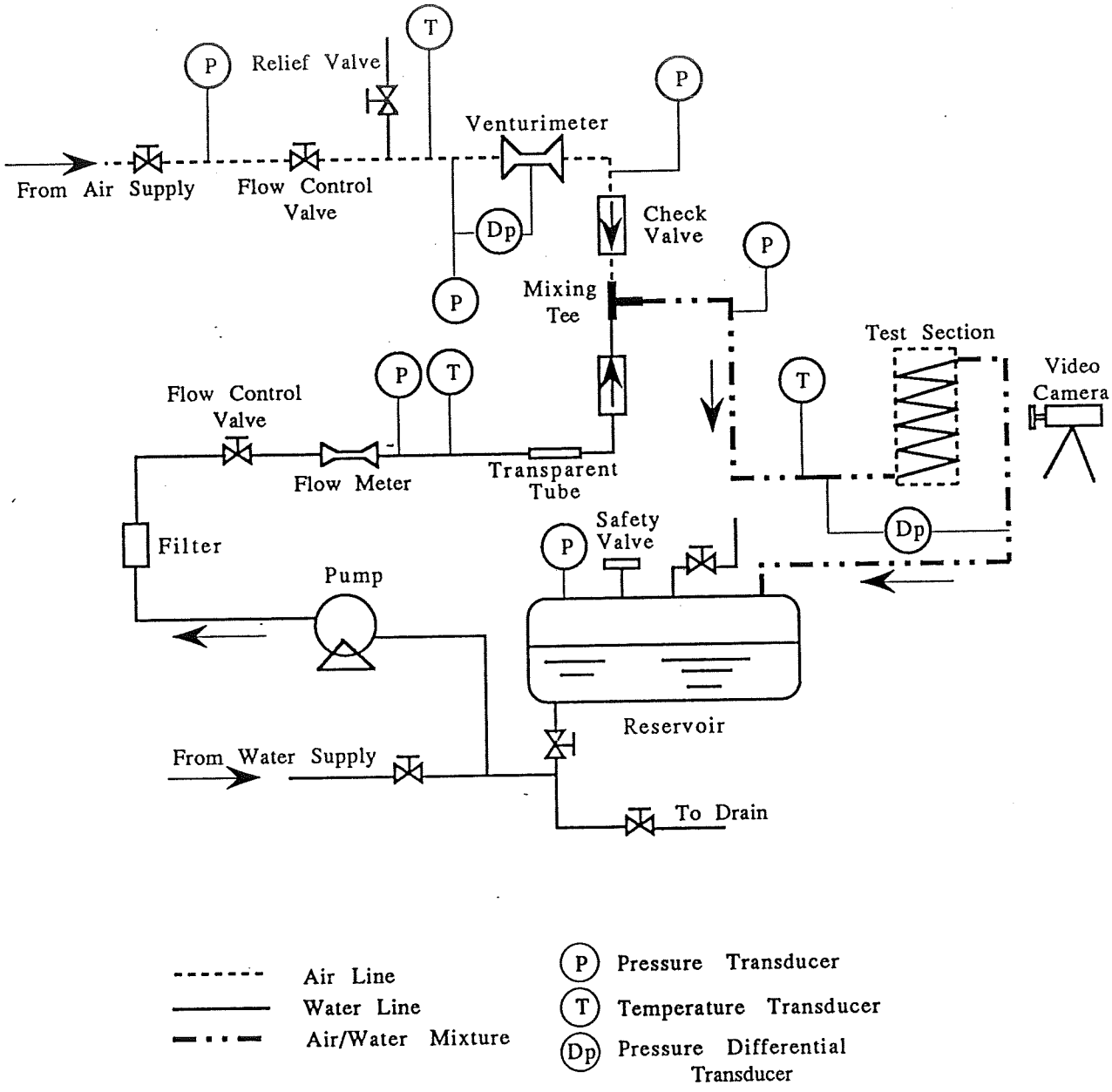


Figure 1 Schematic Diagram of Air-Water Two-Phase Flow Experimental Loop

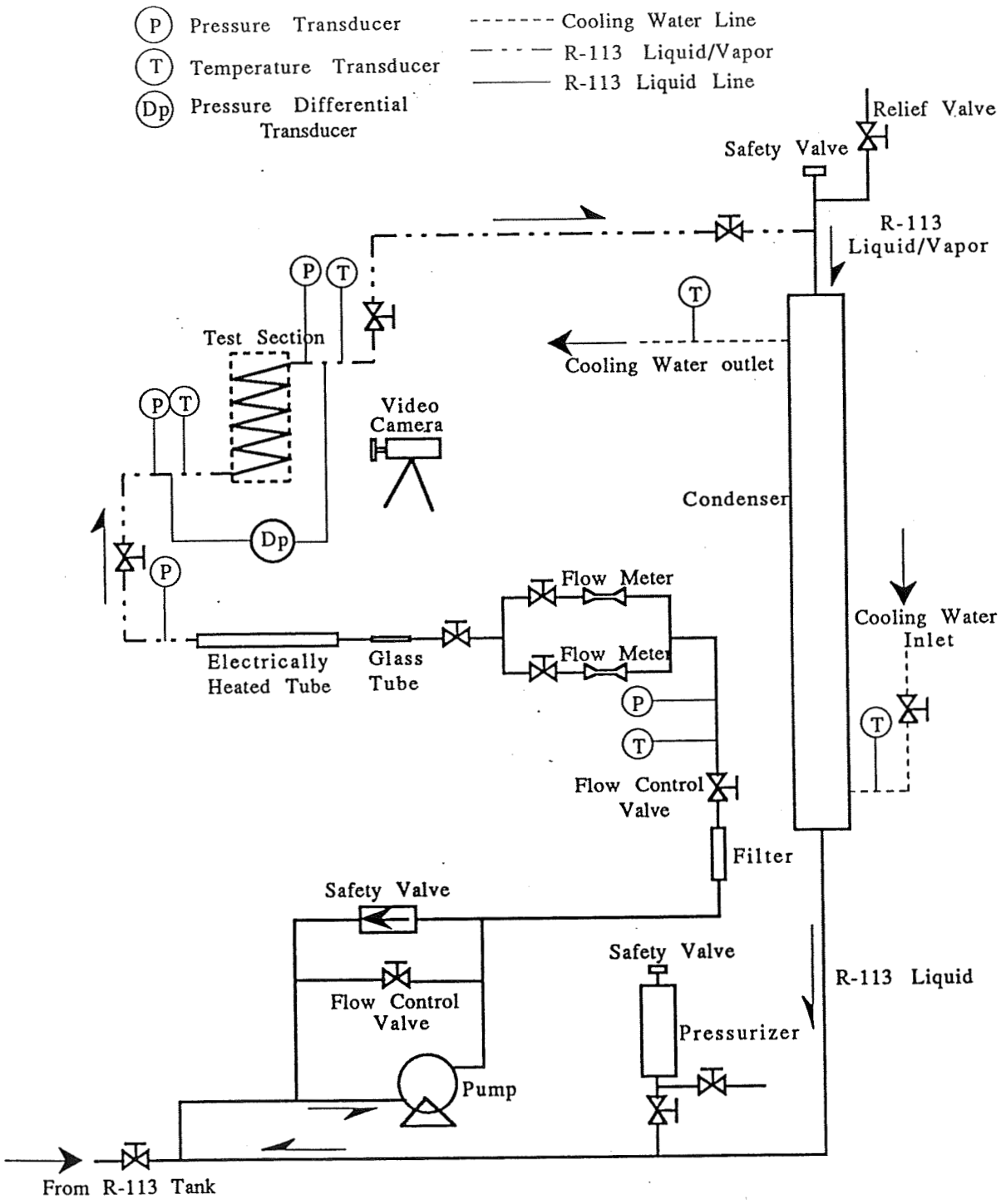


Figure 2 Schematic Diagram of R-113 Two-Phase Flow Experimental Loop

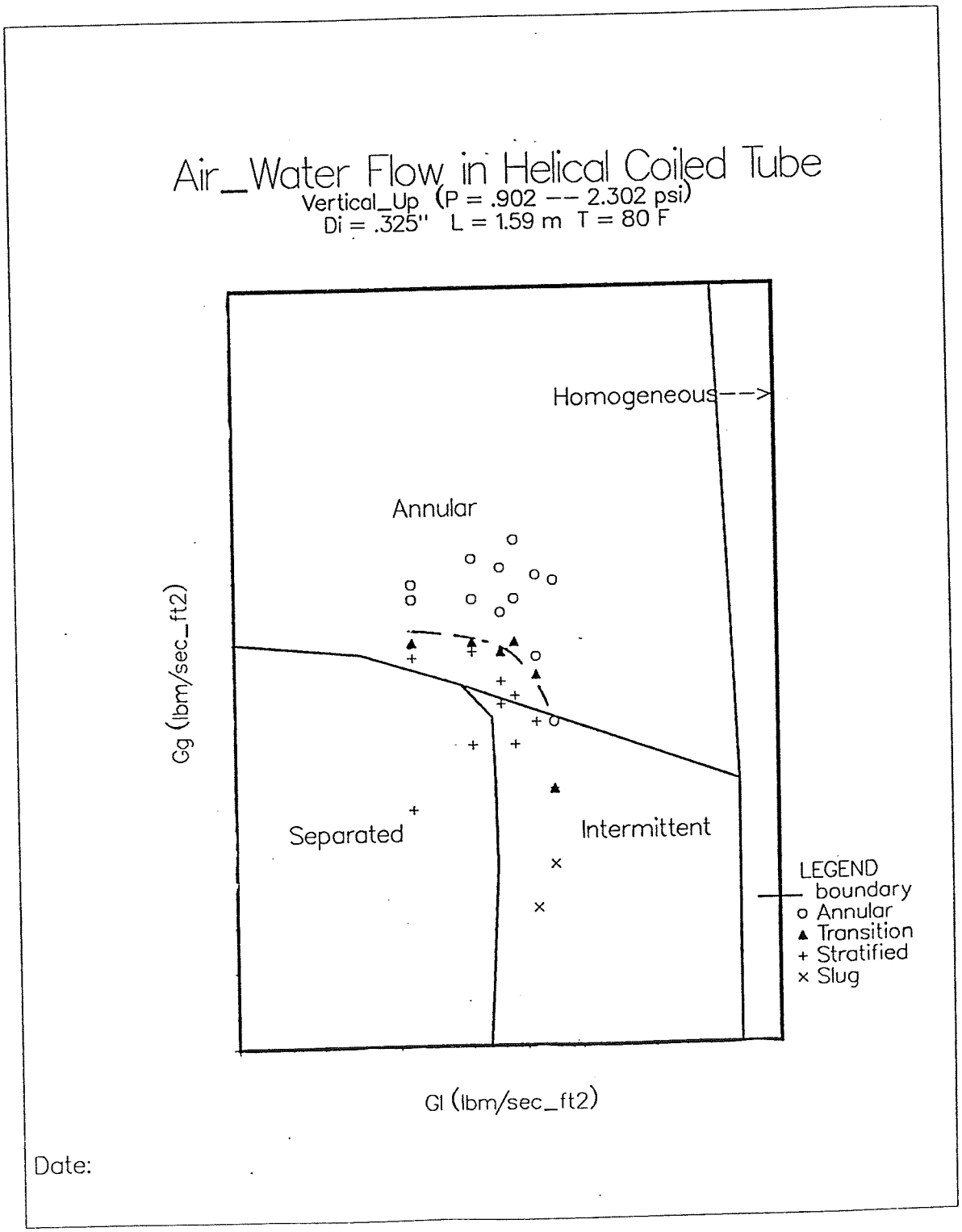


Figure 3. Comparison of flow regime observations with Choe et al map. (Circular cross-section vertically oriented coil axis, upward flow.)

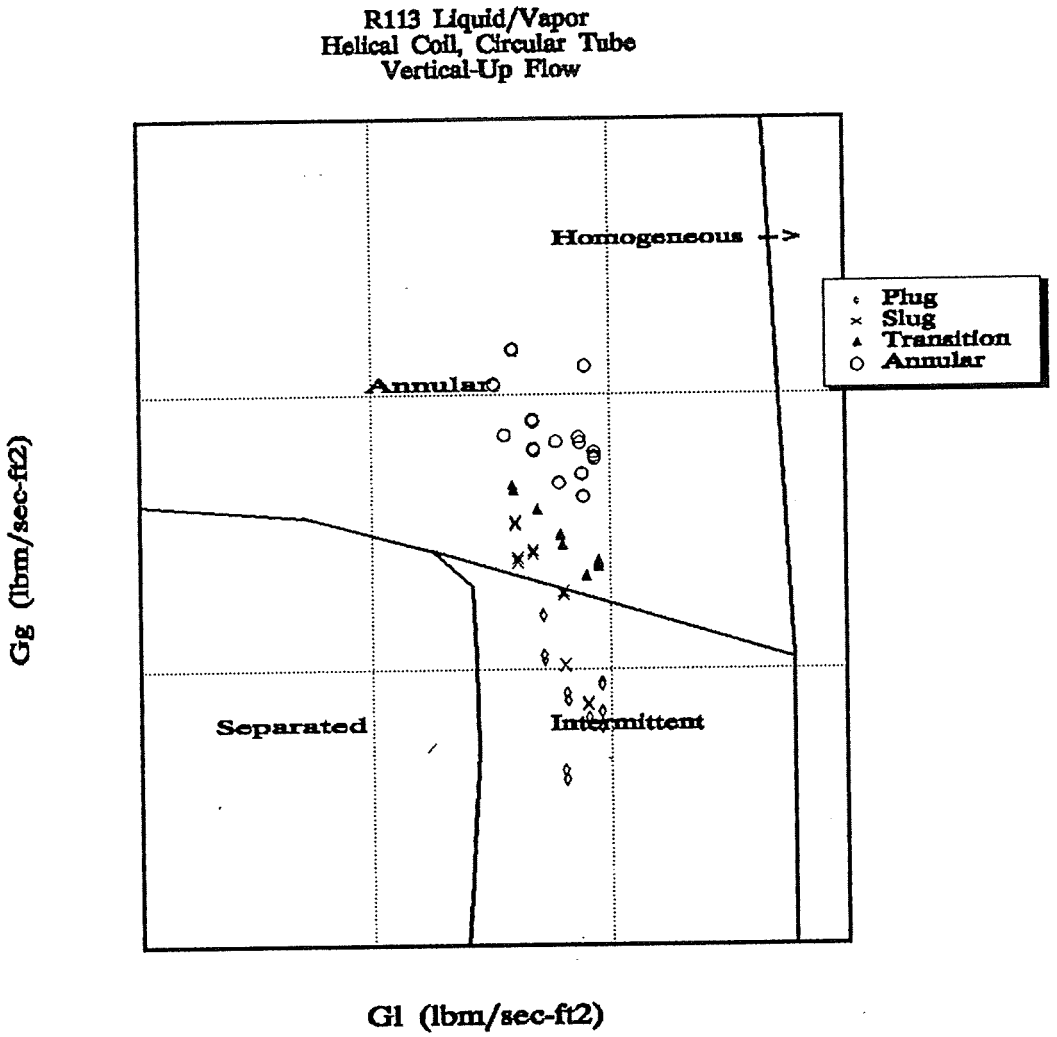


Figure 4. Choe, Weinberg, & Weisman Map

Air-Water, Vertical-Up Flow
Helical Coil, Circular Tube

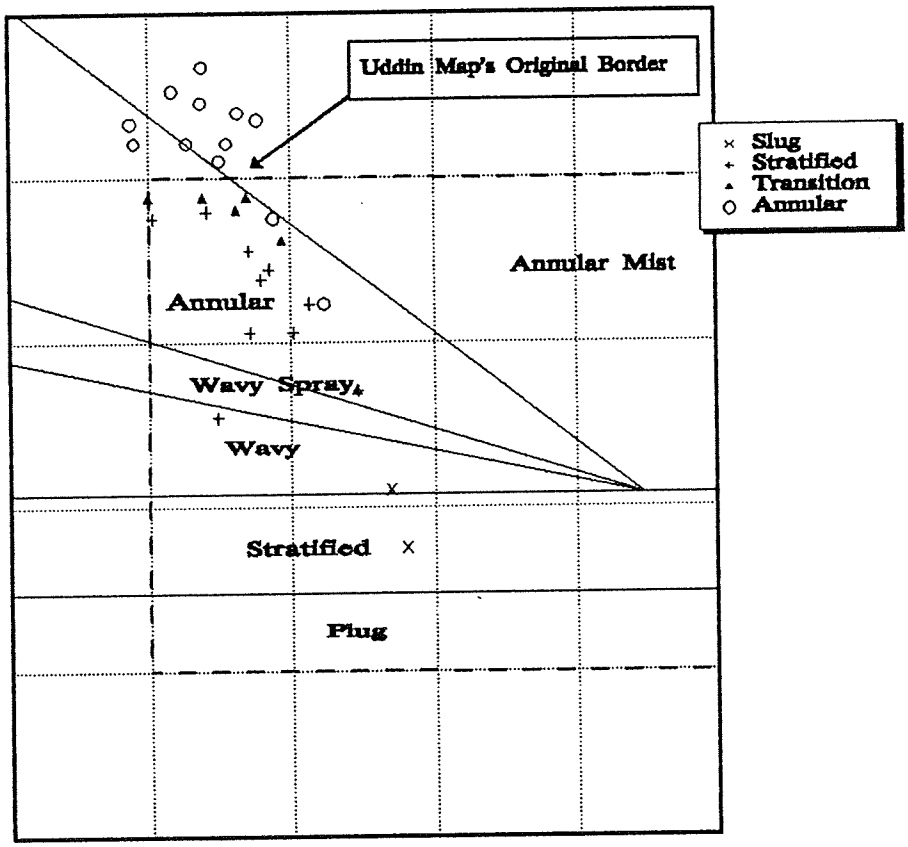


Figure 5. Uddin Flow Map

R113 Liquid/Vapor
Helical Coil, Circular Tube
Vertical-Up Flow

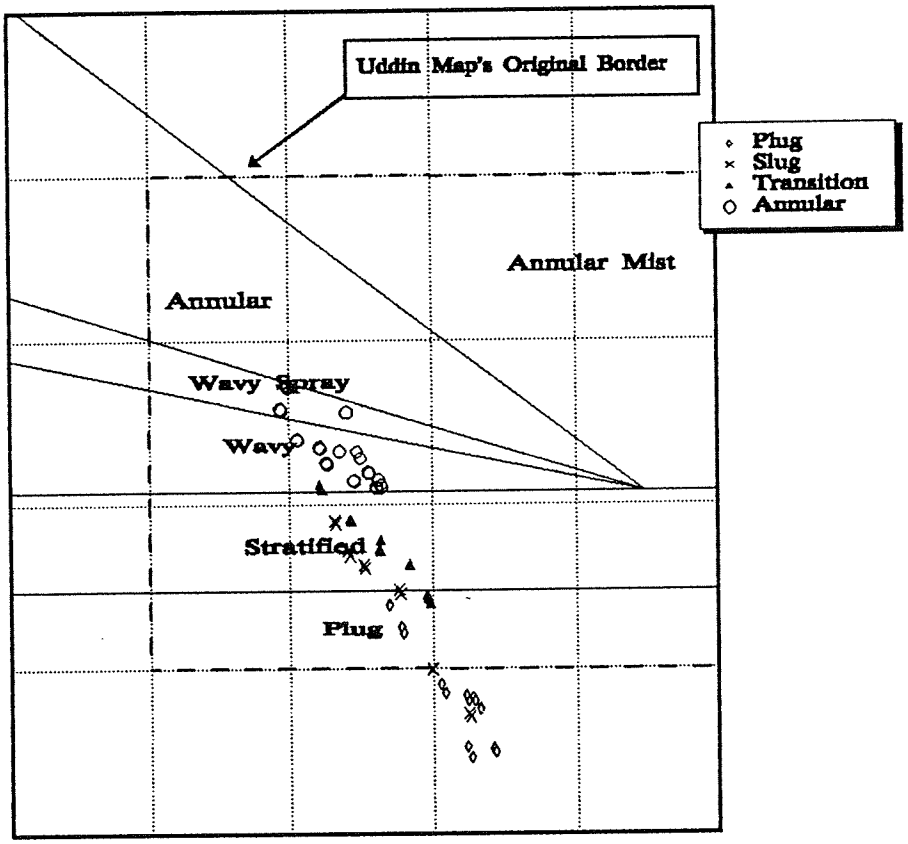


Figure 6. Uddin Flow Map

ROCKET ENGINE MATERIALS

The Effect of Laser Pulse Tailored Welding of Inconel 718

T. Dwayne McCay
Mary Helen McCay
C. Michael Sharp
Michael G.Womack

The Effect of Laser Pulse Tailored Welding of Inconel 718†

T. Dwayne McCay,‡ Mary Helen McCay,* C. Michael Sharp,** and Michael G. Womack***

Center for Advanced Space Propulsion
University of Tennessee Space Institute
Tullahoma, Tennessee 37388
USA

Abstract

Pulse tailored laser welding has been applied to wrought, wrought grain grown and cast Inconel 718 using a CO_2 laser. Prior to welding, the material was characterized metallographically and the solid state transformation regions were identified using Differential Scanning Calorimetry and high temperature x-ray diffraction. Bead on plate welds (restrained and unrestrained) were then produced using a matrix of pulse duty cycles and pulsed average power. Subsequent characterization included heat affected zone width, penetration and underbead width, the presence of cracks, microfissures and porosity, fusion zone curvature, and precipitation and liquated region width. Pedigree welding on three selected processing conditions was shown by microstructural and dye penetrant analysis to produce no microfissures, a result which strongly indicates the viability of pulse tailored welding for microfissure free IN 718.

I. Introduction

Inconel 718 was developed over thirty years ago by the International Nickel Company. Since that time its use has grown until it accounts for thirty-five percent of superalloy production.[1] Due to its exceptional properties from cryogenic temperatures up to 650° C, the material is routinely employed in gas turbine disks for resistance to creep and stress rupture at elevated temperature and in cryogenic piping for environmental cracking resistance. Additional uses are continually being developed as techniques for processing and fabricating the alloy advance.

Welding is one of the more desirable fabrication techniques for turbine and aircraft engine components and hence has been seriously investigated for IN 718. The main problem is microfissuring and cracking although the reduction of properties in the heat affected zone (HAZ) is also a concern. Common welding techniques are[2] shielded-metal arc, gas-tungsten arc, gas-metal arc, resistance, and electron beam. The common requirement for these processes is that they use minimal heat input in order to prevent cracking. Since laser welding uses a very localized heat source and produces a narrow heat affect zone (similar to electron beam welding), it has recently begun to be investigated as a technique for joining IN 718.

The purpose of this report is to present the results of recent studies and analyses of laser pulse tailored welding on three microstructures of Inconel; wrought, wrought-grain grown, and cast.

II. Background

Alloy 718 is known as a nickel-iron base alloy. Its nominal composition is 53% Ni, 18.6% Cr, 18.5% Fe, 3.1% Mo, 5% Nb, 0.4% Al, 0.9% Ti, 0.2% Mn, 0.3% Si and 0.04% C. The matrix is austenitic and the major strengthening components are gamma double prime (Ni_3Nb,Al), an ordered body centered tetragonal phase, and gamma prime (Ni_3Al,Nb) which is a face centered cubic phase. Although they are both primarily nickel, the gamma double prime is richer in niobium while the gamma prime is richer in aluminum. Other phases of significance which occur in present day alloys are delta (Ni_3Nb), Laves (A_2B) and MC.[1] Time-temperature-precipitation (TTP) diagrams have been developed by a number of investigators[3,4,5] to describe the formation of these phases in IN 718. A composite of those results is shown in Figure 1.[2] The state of the alloy can shift the curves to the right or the left, hence differences in grain size or thermal history can produce differences in phase formation.

Although Alloy 718 was originally developed as a wrought alloy, increasing need for its properties has resulted in its recent use as an investment casting alloy. In this form, the microstructure is significantly different from the wrought as can be seen from the

† This work was supported in part by the UT-Calspan Center for Advanced Space Propulsion under NASA Grant No. NAGW-1195.

‡ Professor, Engineering Science & Mechanics, Principal Investigator, UTSI
* Associate Professor, Engineering Science & Mechanics, UTSI
** Processing Engineer, UTSI
*** Graduate Research Assistant, UTSI

comparisons in Figure 2. There is significant concentration of niobium in the Laves phase (Ni_3Nb) eutectic which forms interdendritically and in the grain boundaries. The extensive niobium segregation is extremely difficult to eliminate by thermal treatment[6]. As would be expected, this creates difficulties in processes such as welding when going from wrought to cast material.

During the welding of Alloy 718, microfissures often occur in the heat affected zone creating problems in fabrication and repair. The microfissures are generally thought to be due to the combination of welding stresses and liquation, a phenomena that is influenced by microstructure and constituent inhomogeneities.[7] Laser welding with its inherently small heat affected zone has been considered as a potential solution to these difficulties. A study[8] on the welding and cutting of Inconel 718 with a 15kW CW laser indicated that it could be an effective joining method although there were some isolated instances of microfissuring. Additional studies[9,10] using YAG lasers produced cracks in the nailhead which were attributed to solidification cracking. More recently, pulsed CO_2 laser welding has been applied to wrought Inconel 718 with excellent success in eliminating the microfissures.[11]

III. Material characterization

To better understand the response of the IN 718 structures that were to be welded, Differential Scanning Calorimetry (DSC) was applied to the wrought, grain grown wrought (1 hr. at 1177° C, air cool), and cast. The DSC tests were conducted in argon on 150-160mg samples at two heating and cooling rates. Typical curves for the 40° C/min rate are shown in Figures 3, 4, and 5. The temperatures at which transformations occurred are identified in Table 1 for the three material states being studied. Figure 6 shows the relationship of the transformations (40° C/min runs) with the regions of phase formation for IN 718.[12]

After determining the temperatures for the phase transformations via DSC, high temperature x-ray analysis was conducted to aid in the eventual identification of the phase structures. The data was obtained first at room temperature by scanning through 20° to 120° for 20 angle at 1°/min. The data was collected for each sample state at discrete temperatures (wrought: 540, 570, 600, 730, and 760° C, wrought grain grown: 590, 630, 730, 770, 800° C, and cast: 530, 580, 620, 720, 770, and 820° C). A slow rate of scanning (0.5°/min) was selected for high temperature diffractometry to ensure the detection of slow as well as secondary transformations. In order to avoid the surface oxidation effect on the data, the experiments were conducted in an argon atmosphere. Following the runs, the room temperature x-ray scans were again run on the same samples. Figure 7 shows the x-ray data obtained over these temperatures. Similar analyses are presently being conducted on the fusion zones and heat affected zones of welded samples.

IV. Description of laser and worktable

Bead-on-plate welds were made using a Rofin-Sinar RS3000 fast axial flow CO_2 laser which produces a peak power of 3,700 watts. The laser is radio-frequency excited at 27.12 MHz and is configured to give a TEM_{10} beam mode which has a spatial power distribution described as a ring with a central peak. The wavelength of the beam is 10.6 microns (far-infrared).

Restrained welds were made using a fixture designed specifically for this project. The fixture, in the welding position, is shown in Figure 8. Sample restraint was provided by a pair of jaws that produced both horizontal and vertical pressure on the material (see Figure 9). A slot in the base accommodated the weld underbead. The fixture was mounted on a numerically controlled Aerotech work table which moved the material horizontally beneath the beam.

The beam was focused on the workpiece with an off-axis parabolic mirror having a focal length of 150 mm, equivalent to an f/6 system. The focused laser beam is directed perpendicular to the workpiece through a conical gas flow nozzle. The nozzle delivers a coaxial flow of inert gas which serves to shield the weld while it solidifies and to protect the laser optics from weld spatter. Helium was used as a shielding gas because of its high ionization potential and plasma suppression capabilities. The coaxial gas flow rate was 26 liters/min.

A nozzle positioned to the side of the beam delivery nozzle directed an additional flow of helium towards the workpiece at an angle of 45 degrees from the vertical and thus served as a separate means of plasma control. The plasma control gas flow rate was 20 liters/min.

A. Pulsing and delivered power

The laser output can be continuous-wave (CW) or pulsed. The pulsing format chosen for this experiment was the mixed frequency format (5 kHz pulse width modulation on/gated pulse on) shown in Figure 10.

When the laser is operated in the mixed frequency pulse format the output is gated into pulses with a frequency which is set by independent adjustments of the time on and time off for one pulse cycle. In addition, each gated pulse is composed of multiple high frequency (5 kHz) pulses. The 5 kHz duty cycle is ad-

justable (pulse width modulation) such that the power delivered during the time on of each gated pulse (pulse average power) is adjustable from 7 to 100 percent of peak power (260 to 3,700 watts).

The duty cycle of the low frequency gated pulses is defined as the time on divided by the total time for one pulse cycle. Therefore, the average power delivered during welding is equal to the duty cycle multiplied by the pulse average power.

Welds were attempted using the lowest specific energy input (energy per unit length of weld) possible in an effort to increase the cooling rate, reduce the heat affected zone width, and decrease the incidence of microfissures while maintaining full penetration. Pulse tailoring using the mixed frequency pulse format afforded the greatest flexibility in the selection of welding parameters with lower specific energies than those obtainable with continuous welding processes.

B. General procedure

Prior to the experiment the delivered power was measured using a Scientech water cooled calorimeter. The results are shown in Figure 11. The delivered power is slightly less than the laser controls indicate because of losses in the beam delivery and focussing system. Because the losses were minimal, the laser control indicator was used for subsequent power settings.

The focal position was determined by quickly passing a plexiglass specimen, held on a special fixture, through the beam at an angle of 6 degrees from vertical. Seven measurements were made at different powers. A hyperbolic impression was burned into the plexiglass, and the narrowest point in the impression was considered to be the focal point of the laser beam.

The distance from the bottom (datum) edge on the plexiglass to the narrowest point in the impression was measured using a 20X optical comparator. The distance from a known datum on the laser weld head to the bottom of the plexiglass was measured using the positional outputs from the CNC motion system. The focal position was then determined by subtracting the plexiglass datum to focal impression distance from the weld head datum to plexiglass datum distance. The focal position location did not appear to be functionally related to the laser power, and a mean value of the seven measurements was used as a determination of focal position.

Previous Inconel welds at UTSI were produced with the beam focused at 3 mm below the plane of the top surface. To determine the optimum focal position for the present research, unrestrained bead-on-plate welds were produced at focal positions from 1 to 4 mm below the sample surface. The best penetration for the 3.2 mm thick material was produced when the focus was from 2 to 3 mm below the surface. A focal position of 3 mm below the surface was maintained in order to be consistent with the previous work.[11]

The following parameters were kept fixed for all experiments:

1. Beam mode (TEM_{10});
2. Focussing system (f/6 off-axis parabola);
3. Focus position (3 mm below workpiece surface);
4. Pulse format (mixed frequency);
5. Nozzle standoff distance (6 mm);
6. Coaxial gas (helium, 26 liters/min);
7. Side gas jet (helium, 20 liters/min).

Four experiment matrices were selected for this study. They are: unrestrained weld matrix, restrained weld matrix, variable duty cycle weld matrix, and pedigree weld matrix. These are discussed in detail in the following sections.

C. Unrestrained weld matrix

During the first phase of the experiment, unrestrained bead-on-plate welds were produced on 1/8-inch (3.2 mm) wrought Inconel 718. Pulse average power, pulse duty cycle, and weld translation speed were the welding parameters that were varied in order to find a matrix of weld parameters which met the penetration criterion. This criterion was established as a continuous underbead with a minimum width between 0.8 and 0.9 mm. The function of the penetration criterion was to give the researchers a measurable standard by which to compare welds produced with different welding parameters.

After processing, the weld underbead was measured with a scaled loop (Magnification: 10 X, Scale Divisions: one-tenth of a millimeter). For welds which did not meet the penetration criterion, the pulse average power was either increased or decreased while the pulse duty cycle and translation velocity were held constant. In this manner the parameters which give acceptable penetration were determined. The final weld conditions are given in Table 2. A "U" precedes the sample number to designate the weld as unrestrained.

During the welding just described, it was observed that welds produced at the lowest duty cycles could be made to meet the penetration criterion over a wider range of pulse average powers than could the high duty cycle welds. Weld conditions U28, U29, U30, U31, and U32 were selected for further examination of this

behavior because they spanned a range of duty cycles from 37 to 71 percent at the same translation velocity.

These five conditions were rewelded, and this time the pulse power was increased until excessive penetration (minimum underbead width greater than 0.9 mm) was observed. The welds were again made and the pulse power lowered until under penetration was achieved (minimum underbead width less than 0.8 mm).

D. Restrained weld matrix

Restrained weld samples were cut to 3.5 cm by 0.95 cm from the as-received stock of thickness 3.2 mm (1/8-inch). Paint and oil were removed from the surface with an acetone flush.

The samples were clamped into the fixture (Figure 9) by first tightening the four vertical jaw bolts until they contacted their seating surfaces, preventing vertical jaw movements. Next, the two horizontal jaw bolts were hand-tightened, thus applying pressure horizontally. The vertical jaw bolts were further tightened, thus applying downward pressure. After the samples were welded, they were left in the fixture until the sample was cool to the touch and the tightening procedure was reversed.

Restrained welds were performed at five duty cycles (37, 52, 61, 71, and 80 percent), four pulse average power levels (2600, 2250, 2050, and 1900 watts), and a translation velocity of 12 mm/s; twenty conditions were used to produce the samples. Four fully penetrated conditions were repeated from the unrestrained matrix. These were U28, U29, U30, and U31.

The restrained weld conditions were selected so that the effect of duty cycle and/or the effect of pulse average power on the weld microstructure (microfissures and porosity) could be examined along lines of constant duty cycle (with changes in pulse average power) or along lines of constant pulse average power (with changes in duty cycle). In this manner a matrix of conditions (Table 3) was established. An "R" precedes the sample number to designate a restrained weld.

E. Variable duty cycle weld matrix

The purpose of this experiment was to investigate the microstructural relationship between pulsed laser welds in small grain wrought (ASTM #8 or smaller), grain grown wrought (ASTM #5), and cast (ASTM #2) Inconel 718. Restrained bead-on-plate welds were carried out using the same welding fixture and restraint as used in the restrained weld matrix.

The small grain wrought and grain grown wrought samples were cut from sheet stock while the cast samples were cut from a cast bar using an abrasive cut-off saw. The thickness for all samples was 1/8-inch (3.2 mm). All samples were cleaned with acetone immediately prior to welding.

The translation speed was kept constant at 12 mm/s and the pulse average power was also kept constant at 2250 W. Weld duty cycle was the experimental variable and the values used are shown in Table 4. The designation RW indicates that these samples were wrought Inconel 718 welded in the restrained condition. Likewise, RG indicates restrained welds on grain grown material, and RC indicates restrained welds on cast material.

F. Pedigree weld matrix

The purpose of the pedigree weld matrix was to demonstrate that three welds from the restrained matrix, which were free from microfissures and porosity, could be reproduced with a quantitative level of confidence. The conditions for the pedigree matrix are listed in Table 5. Along with being free from microfissures and porosity, these conditions are near the median of pulse average power, and near the median of the duty cycles used for the restrained matrix.

The samples were cut to the standard size (3.5 × 0.95 cm), cleaned, and placed in the welding fixture in the same manner as the restrained matrix samples. After welding, the samples were left in the fixture until they were cool to the touch. Nine samples (three for each condition) were welded.

G. Metallographic analysis of samples

The restrained weld matrix, pedigree weld matrix, and variable duty cycle weld matrix samples were transversely cut to begin the metallurgical analysis. Longitudinal sections of selected restrained weld samples were ground with a surface grinder in 0.2 mm increments parallel to the surface. Part of each variable duty cycle specimen was slant-ground using the technique shown in Figure 12. All sections were mounted in thermosetting mounting material and then ground with silicon-carbide abrasive paper. Paper of 80, 120, 180, 240, 320, 400, and 600 grit was used in succession for grinding. The samples were then polished with 6-micron diamond paste on a cloth wheel to remove visible scratches.

Immediately after the samples were polished they were etched with Waterless Kallings reagent to reveal the microstructure. The reagent was applied with a cotton swab and left on for about 15 seconds or until the weld was clearly visible. The samples were

rinsed in running water, then ultrasonically cleaned in distilled water and air dried to remove the remaining reagent. An Olympus Model PME inverted-stage metallurgical microscope was used at magnifications from 13X to 1000X to examine and photograph the transverse cross sections.

Images of the longitudinal restrained matrix weld sections were obtained with a video camera and macro-lens attached to a Vicom digital image processor. The outline of the longitudinal section was traced, and the area and perimeter enclosed by the trace were calculated by the Vicom. Longitudinal curvature was calculated by dividing the perimeter by the enclosed area. Since the welds have a three dimensional structure dependent on duty cycle, measurements of the longitudinal and transverse curvature were made as a basis for comparison between welds.

In order to calculate the transverse curvature, the outlines of the weld from the first, third, fifth, and seventh cross sections were stacked together in one image and the resulting contour map was photographed. The spacing between the lines on the contour map was measured at ten stations spaced every 3.5 mm. This spacing was selected because the measurements spanned the length welded by two pulses of the condition having the greatest pulse spacing (duty cycle of 37 percent, pulse average power of 2050 watts).

The method for establishing an approximate transverse cross sectional shape at an individual station is depicted for a typical weld in Figure 13. Transverse curvature is calculated by dividing the perimeter of the transverse cross section by the area (sum of areas A, B, and C).

In practice, most of the microfissures that are observed after welding this material are found in the heat affected zone (HAZ) beneath the weld nailhead. Microhardness measurements to determine HAZ width were made on samples U9, U28 and an unrestrained CW weld using a Buehler Micromet II hardness tester with 500 gram mass impact. The impact time was set at 15 seconds.

V. Results

A. Unrestrained welds

Unrestrained bead-on-plate welds were produced on 1/8-inch thick wrought Inconel 718. The process parameters (pulse average power, pulse duty cycle, and weld translation velocity) were varied in order to find a matrix of weld parameters which met penetration and continuous underbead criteria.

1. Relationships between weld parameters

Unrestrained, fully penetrated welds (0.8 to 0.9 mm minimum underbead width) were produced over a wide range of pulse duty cycles, pulse average powers, and weld translation velocities. As seen in Figure 14, a near-linear relationship exists between the pulse average power and the pulse duty cycle required to produce full penetration for a given weld translation velocity.

Penetration measurements also revealed a dependence of the acceptable pulse average power range on duty cycle (Figure 15). For example, sample U31 (37 percent duty cycle, 2600 W, 12 mm/s) had a range of 700 watts, while sample U28 (71 percent duty cycle, 1900 W, 12 mm/s) had only a 100 watt range between overpenetration and underpenetration. Therefore, the low duty cycle welds were more tolerant to pulse average power variations.

2. Centerline cracks

Large centerline cracks were observed in some unrestrained welds. These were thought to occur because the small samples had no surrounding unaffected material to provide restraint during weld pool melting and resolidification. Figure 16 shows that fully penetrated welds having a duty cycle less than 54 percent exhibited no centerline cracks regardless of changes in the other weld parameters (pulse average power, speed). This indicates that the long off-times that are characteristic of low duty cycle welds give sufficient time for the weld to solidify and strengthen before additional heat from a subsequent pulse is added. Low duty cycle welds are therefore less prone to centerline cracking when mechanical restraint is not provided.

3. Heat affected zone width

Microhardness measurements across the weld and into the matrix metal (Figures 17, 18 and 19) were used to determine the width of the heat affected zone (HAZ). Material which had no increase in hardness above that of the matrix was presumed to be outside of the HAZ. Weld condition U9 (61 percent duty cycle) had a heat affected zone width averaging 48 percent of the weld width, whereas weld condition U28 (71 percent duty cycle) had a heat affected zone width averaging 65 percent of weld width. In comparison, the HAZ of a similar (unrestrained) CW weld (2500 W, 1.5 M/min) on the same material was 61 percent of weld width. The results indicate that a weld having a duty cycle of 71 percent or more has a HAZ width that is approximately equal to the HAZ width of a CW weld.

B. Restrained welds

The unaffected material surrounding a weld acts as restraint during solidification. The small sample size used for this experiment necessitated the use of mechanical restraint since there was a limited amount of unaffected material surrounding the weld.

1. Nailhead microfissures

Figures 20, 21, and 22 show typical weld cross sections. Figures 23, 24, and 25 show typical weld microstructures at increasing magnifications. Although precipitates have accumulated in the grain boundaries within the heat affected zone beneath the nailhead, nailhead microfissures were not observed in any of the restrained weld cross sections.

The absence of cracking in the restrained welds produced by the laser for this study may be attributed to the rapid cooling rate and its associated small heat affected zone and limited extent of grain boundary liquation. In addition, the low total heat input to the welds, which is made possible by the use of pulse tailored laser welding, lessens the localized strains which occur in the HAZ.

2. Porosity

Restrained welds having large centerline pores are indicated on Figure 26 along with the other welds in that matrix. Pores were more likely to occur in welds which were processed with the lowest duty cycles and/or the lowest pulse average powers. During the "beam-off" portion of the pulsing cycle, the internal pressure in the keyhole decreases to the point that it will no longer support the walls and the keyhole partially collapses. Keyhole gases (metallic plasma), trapped by the collapsing of the keyhole, can nucleate into spherical pores which are engulfed by the rapid solidification front before they rise to the surface. These gases would then condense and solidify as the weld cools, leaving a void.

The high duty cycle and high pulse power welds did not exhibit porosity. The higher specific energy inputs significantly increase the time that the weld pool remains molten, thus allowing any pores which have nucleated the time to escape the molten material.

3. Nailhead curvature

Longitudinal curvature was calculated by dividing the perimeter of the longitudinal section by the longitudinal area. The results are found in Tables 6 and 7. Transverse curvature was calculated from cross sections obtained using the technique shown in Figure 13. Table 8 lists the mean value of ten measurements for each weld. The results must be regarded as approximate since the assumption of linearity between contours was made; the errors should be small since the difference in depth between contours is small.

It is desirable to reduce the two curvature values for each weld to a single curvature rating in order to compare the welds within the matrix. The individual values were first normalized by expressing them as a fraction of the largest value obtained by the same method.

It was assumed that the transverse and longitudinal curvature were equally important and therefore given equal weight. The curvature rating was calculated as the sum of the normalized longitudinal and transverse curvatures. As seen in Table 9, the curvature rating is inversely related to duty cycle and pulse average power.

C. Variable duty cycle weld matrix

Figure 2 shows the base material microstructures that were used in this study. The wrought grain grown (ASTM #5) has significantly larger grains than the wrought (ASTM #8). It also exhibits fewer transformation regions during DSC analysis, a result of the recovery and solutionizing which occurs during the grain growth treatment. The cast structure is large grained (ASTM #2) and obviously dendritic with extensive segregation. DSC analysis shows similar transformations to the wrought and wrought grain grown but with a significant transformation occurring around 1180° C which is probably incipient melting as a result of the segregation.

Subsequent to welding, the three Alloy 718 materials exhibited precipitation and evidence of grain boundary liquation and interdendritic remelting in the heat affected zones. Figure 27, 28 and 29 show microstructures in the region of the nailhead. The visually affected region extends 0.098±0.015 mm from the fusion zone in the wrought material, 0.139±0.001 mm from the fusion zone in the wrought grain grown material, and 0.309±0.147 mm from the fusion zone in the cast material.

D. Pedigree welding trials

Five metallurgical cross-sections were prepared from each of the nine samples welded using the conditions shown in Table 4. The cross-sections were examined at 100X magnification. Although some grain boundary precipitation was observed in the region adjacent to the nailhead, no microfissures and porosity were found. The cross-sections were then crack tested using the dye-penetrant technique. No cracks were detected.

Acknowledgments

The authors would like to acknowledge the contributions of the Marshall Space Flight Center, Materials and Processes Laboratory and in particular Mr. R. J. Schwinghamer and Mr. Ernie Bayless for providing the wrought Inconel. The High Temperature Materials Laboratory at Oak Ridge National Laboratory allowed access to their Differential Scanning Calorimeter and their high temperature x-ray diffractometer through their user program. Discussions and collaboration with ORNL personnel through this program have been invaluable. In particular the authors would like to thank the Center for Advanced Space Propulsion for their sponsorship of this work. This work was performed within the Center for Advanced Space Propulsion and was supported in part by NASA Grant No. NAGW-1195. The Center for Advanced Space Propulsion is part of the UT- Calspan Center for Aerospace Research, a not-for-profit organization located at UTSI.

References

1. John R. Radavich, "Metallography of Alloy 718," Journal of Metals, July 1988, pp. 42-43.
2. C. T. Sims, N. S. Stoloff and W. C. Hagel, Superalloys II, John Wiley and Sons, New York (1987).
   John R. Radavich, "The Physical Metallurgy of Cast and Wrought Alloy 718," **Superalloy 718 - Metallurgy and Applications**, ed. E.A. Loria, The Minerals, Metals and Materials Society, pp. 229-240, 1989.
3. R. Cozar and A. Pineau, Met. Trans. 4, 47 (1973).
4. H. L. Eiselstein, "Advances in the Technology of Stainless Steels and Related Alloys," STP 369, ASTM Philadelphia, PA, pp. 62-67, 1965.
5. W. J. Boesch and H. B. Canada, J. Met., 21 (10), 34 (1969).
6. T. J. Kelly, "Investigation of Elemental Effects on the Weldability of Cast Nickel-Based Superalloys," **Advances in Welding Science and Technology TWR '86**, ed. S. A. David (May 1986) pp. 623-627.
7. R. G. Thompson, "Microfissuring of Alloy 718 in the Weld Heat-Affected Zone," Journal of Metals, July 1988, pp. 44-48.
8. C. M. Banas, Final Report, Contract NAS8-36306.
9. L. A. Weeter, C. E. Albright and W. H. Jones, Welding Research Supplement, March 1986, 51s-62s.
10. J. P. Reynolds, H. W. Kerr, P. J. Fenrenbach, L. Bourque, and R. D. Davidson, **Advances in Welding Science and Technology, TWR '86**, Proceedings of an International Conference on Trends in Welding Research, May 1986, Ed. S. A. David, ASM International.
11. M. Sharp, M. H. McCay, T. D. McCay, N. B. Dahotre, A. Sedghinasab and S. Gopinathan, "Pulsed Laser Welding of Inco 718," **ICALEO '89 Proceedings**, Volume 69, Laser Material Processing, October 1989, pp. 229-238.
12. R. G. Carlson and J. F. Radavich, "Microstructural Characterization of CAST 718," Superalloy 718 - Metallurgy and Applications, ed. E. A. Loria, The Minerals, Metals & Materials Society, 1989, pp. 79-95.

Table 1. Transformation Temperatures in IN 718.

Sample	DSC Treatment	Transformation Temperatures (° C)
Wrought	Heating 10°C/min	568, 702, 1030, 1220, 1280
	Cooling 10°C/min	760, 1270
	Heating 40°C/min	615, 760, 912, 1025, 1220, 1275
	Cooling 40°C/min	730, 1100, 1150, 1265
Wrought Grain Grown	Heating 10°C/min	592, 775, 1230, 1280
	Cooling 10°C/min	765, 1272
	Heating 40°C/min	655, 850, 1220
	Cooling 40°C/min	1100, 1266
Cast	Heating 10°C/min	538, 725, 1175
	Cooling 10°C/min	1265
	Heating 40°C/min	585, 770, 1130, 1180, 1205, 1280
	Cooling 40°C/min	1160, 1267

Table 2. Unrestrained Weld Matrix.

Designation	Translation Velocity (mm/s)	Pulse Average Power (W)	On Time (ms)	Off Time (ms)	Duty Cycle (%)	Specific Energy (J/mm)	Underbead Width (mm)
U1	25	2500	17	3	85	85.0	1.0
U2	25	3000	17	7	71	85.2	1.0
U3	25	3000	17	21	45	54.0	0.6
U4	25	2700	17	7	71	76.7	0.8
U5	25	2900	17	11	61	70.7	0.8
U6	25	2475	17	4	81	80.2	0.8
U7	18	2000	23	4	85	94.4	0.6
U8	18	2333	22	9	71	92.0	0.8
U9	18	2567	23	15	61	87.0	0.8
U10	18	2857	24	21	53	84.1	0.9
U11	18	2125	22	6	79	93.3	0.9
U12	18	3000	22	33	40	66.7	0.8
U13	18	2775	22	22	50	77.1	0.8
U14	15	1900	28	5	85	107.7	1.0
U15	15	2150	29	12	71	101.8	1.0
U16	15	2150	28	18	61	87.4	0.8
U17	15	2600	28	25	53	91.9	1.0
U18	15	2500	28	24	54	90.0	1.2
U19	15	2650	28	33	46	81.3	0.9
U20	15	3000	28	76	27	54.0	0.0
U21	15	2530	28	28	50	84.3	0.8
U22	15	2200	28	42	40	58.7	0.0
U23	15	2000	28	7	80	106.7	0.8
U24	15	2725	28	42	40	72.7	0.9
U25	15	2250	28	19	60	90.0	0.8
U26	15	2125	28	12	70	99.2	1.0
U27	12	1850	34	6	85	131.0	1.0
U28	12	1900	36	15	71	112.4	0.9
U29	12	2050	36	23	61	104.2	0.8
U30	12	2250	35	32	52	97.5	0.9
U31	12	2600	35	60	37	80.2	0.8
U32	12	2400	35	43	45	90.0	0.8
U33	12	3000	35	198	15	37.5	0.0
U34	12	2400	35	42	45	90.0	0.8

Table 3. Restrained Weld Matrix, Translation Velocity = 12 min/s.

Designation	Pulse Average Power (W)	On Time (ms)	Off Time (ms)	Duty Cycle (%)	Specific Energy (J/mm)
R1	1900	36	9	80	127
R2	1900	36	15	71	112
R3	1900	36	23	61	97
R4	1900	35	32	52	83
R5	1900	35	60	37	58
R6	2050	36	9	80	137
R7	2050	36	15	71	121
R8	2050	36	23	61	104
R9	2050	35	32	52	89
R10	2050	35	60	37	63
R11	2250	36	9	80	150
R12	2250	36	15	71	132
R13	2250	36	23	61	114
R14	2250	35	32	52	98
R15	2250	35	60	37	69
R16	2600	36	9	80	173
R17	2600	36	15	71	153
R18	2600	36	23	61	132
R19	2600	35	32	52	113
R20	2600	35	60	37	80

Table 4. Variable Duty Cycle Weld Matrix, Translation Velocity = 12 min/s.

Designation	Pulse Average Power (W)	On Time (ms)	Off Time (ms)	Duty Cycle (%)	Specific Energy (J/mm)
RW1	2250	35	60	35	69.4
RW2	2250	35	32	52	97.5
RW3	2250	36	23	61	114.4
RW4	2250	36	15	71	133.1
RW5	2250	36	9	80	150.0
RW6	2250	36	4	90	168.8
RW7	2250	-	-	CW	187.5
RG1	2250	35	60	35	69.4
RG2	2250	35	32	52	97.5
RG3	2250	36	23	61	114.4
RG4	2250	36	15	71	133.1
RG5	2250	36	9	80	150.0
RG6	2250	36	4	90	168.8
RG7	2250	-	-	CW	187.5
RC1	2250	35	60	35	69.4
RC2	2250	35	32	52	97.5
RC3	2250	36	23	61	114.4
RC4	2250	36	15	71	133.1
RC5	2250	36	9	80	150.0
RC6	2250	36	4	90	168.8
RC7	2250	-	-	CW	187.5

Table 5. Pedigree Weld Matrix, Translation Velocity = 12 min/s.

Designation	Pulse Average Power (W)	On Time (ms)	Off Time (ms)	Duty Cycle (%)	Specific Energy (J/mm)
P1	2600	36	23	61	132
P2	2250	36	23	61	114
P3	1900	36	15	71	112

Table 6. Longitudinal Curvature Data for Restrained 61 Percent Duty Cycle Welds.

Sample	Depth (mm)	Perimeter (mm)	Area (sq. mm)	Curvature (1/mm)
R3	0.20	28.152	23.296	1.208
	0.38	26.916	20.759	1.297
	0.53	26.568	18.491	1.437
	0.66	26.452	16.137	1.639
	0.81	28.113	15.314	1.836
	1.04	27.186	14.877	1.827
	1.37	27.070	15.432	1.754
R8	0.20	28.036	22.310	1.257
	0.41	26.723	20.952	1.275
	0.53	27.070	19.221	1.408
	0.66	26.337	17.725	1.486
	0.84	28.152	16.240	1.733
	1.07	25.719	14.540	1.769
	1.37	26.259	15.528	1.691
R13	0.20	26.568	23.427	1.134
	0.41	28.229	23.860	1.183
	0.53	27.070	22.096	1.225
	0.69	27.109	20.696	1.310
	0.86	27.186	18.950	1.435
	1.07	25.873	16.113	1.606
	1.40	26.877	16.040	1.676
R18	0.20	28.113	27.164	1.035
	0.38	27.881	27.013	1.032
	0.51	27.611	25.611	1.078
	0.69	26.877	24.380	1.102
	0.86	27.070	23.548	1.150
	1.04	26.646	20.108	1.325
	1.35	28.152	18.896	1.490

Table 7. Longitudinal Curvature Data for Restrained 2050 Watt Pulse Average Power Welds.

Sample	Depth (mm)	Perimeter (mm)	Area (sq. mm)	Curvature (1/mm)
R6	0.00	26.852	26.573	1.010
	0.23	28.547	28.193	1.013
	0.51	29.047	27.657	1.050
	0.66	28.046	25.368	1.106
	0.94	27.853	22.651	1.230
	1.14	27.006	20.876	1.294
	1.30	26.813	20.030	1.339
R7	0.00	28.624	23.035	1.243
	0.33	29.394	24.765	1.187
	0.61	28.084	22.445	1.251
	0.79	27.352	20.966	1.305
	0.99	27.006	18.855	1.432
	1.19	26.543	16.354	1.623
	1.37	26.890	15.802	1.702
R8	0.20	28.036	22.310	1.257
	0.41	26.723	20.952	1.275
	0.53	27.070	19.221	1.408
	0.66	26.337	17.725	1.486
	0.84	28.152	16.240	1.733
	1.07	25.719	14.540	1.769
	1.37	26.259	15.528	1.691
R9	0.00	28.277	23.291	1.214
	0.28	27.468	22.642	1.213
	0.51	27.661	20.020	1.382
	0.79	26.813	16.544	1.621
	0.94	27.314	15.090	1.810
	1.14	25.773	12.978	1.986
	1.32	26.697	13.201	2.022
R10	0.00	29.510	24.160	1.221
	0.28	29.047	23.504	1.236
	0.48	29.202	18.983	1.538
	0.76	27.314	12.780	2.137
	0.94	27.584	12.617	2.186
	1.12	28.316	12.516	2.262
	1.32	28.277	12.647	2.236

Table 8. Transverse Curvature of Restrained Welds.

Sample +	Mean Curvature* (1/mm)	Standard Deviation
R3	0.231	0.005
R8	0.207	0.006
R13	0.215	0.006
R18	0.215	0.008
R6	0.186	0.005
R7	0.204	0.007
R9	0.211	0.006
R10	0.206	0.007

+ Samples R3, R8, R13, and R18 are 61 percent duty cycle welds. Samples R6, R7, R8, R9, and R10 are 2050 watt welds.
* Represents mean values of ten measurement stations per contour plot.

Table 9. Weld Curvature Summary.

Sample	Mean Longitudinal Curvature (Normalized Value)	Mean Transverse Curvature (Normalized Value)	Curvature Rating
R3	1.571 (0.858)	0.231 (1.000)	1.858
R8	1.517 (0.829)	0.207 (0.896)	1.725
R13	1.367 (0.747)	0.215 (0.931)	1.678
R18	1.173 (0.641)	0.215 (0.931)	1.572
R6	1.149 (0.628)	0.186 (0.805)	1.433
R7	1.392 (0.760)	0.204 (0.883)	1.643
R9	1.607 (0.878)	0.211 (0.913)	1.791
R10	1.831 (1.000)	0.206 (0.892)	1.892

Table 10. Weld Metallographic Observations.

Sample	Nailhead Microfissures	Large Pores
R1	N	Y
R2	N	N
R3	N	Y
R4	N	Y
R5	N	Y
R6	N	N
R7	N	N
R8	N	Y
R9	N	Y
R10	N	Y
R11	N	N
R12	N	N
R13	N	N
R14	N	N
R15	N	Y
R16	N	N
R17	N	N
R18	N	N
R19	N	N
R20	N	Y
P1	N	N
P2	N	N
P3	N	N

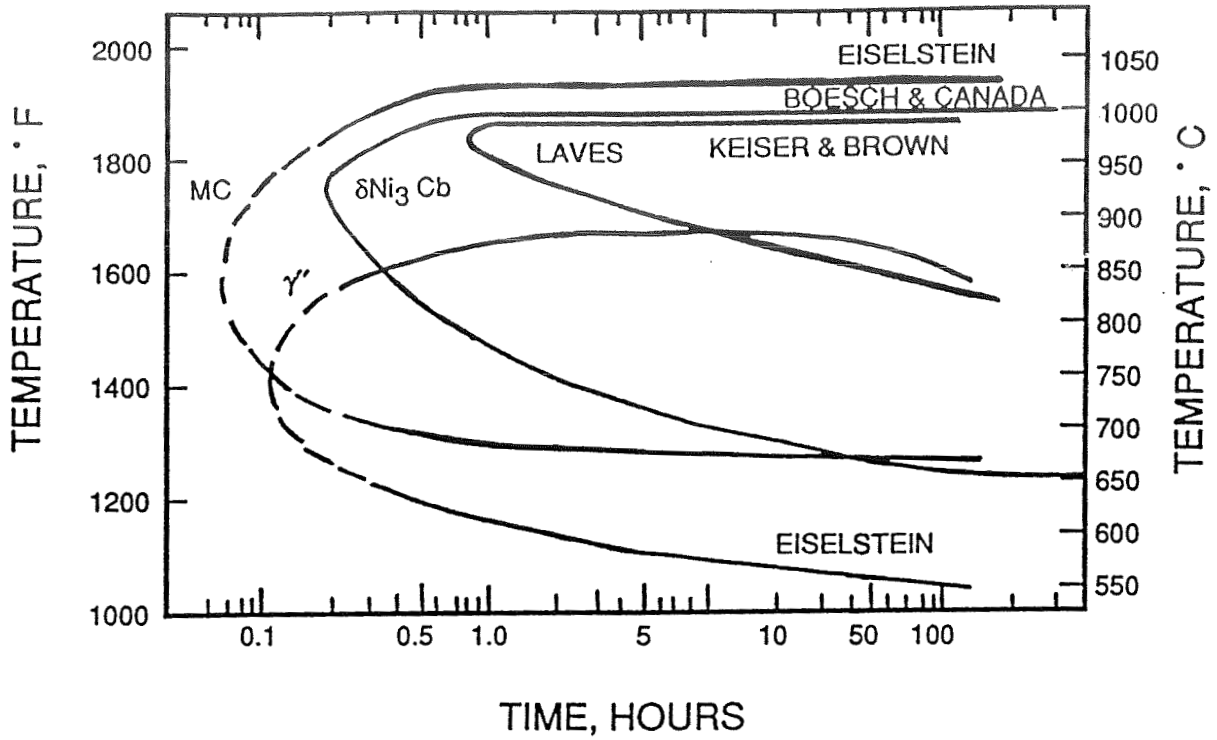
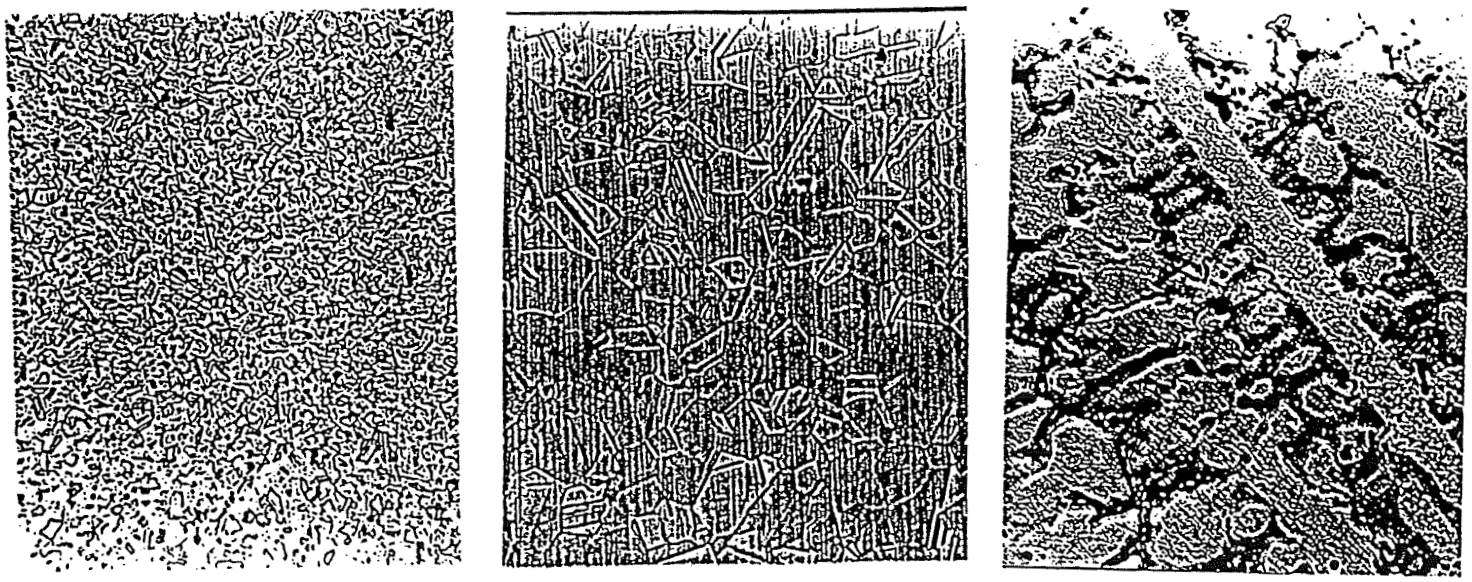
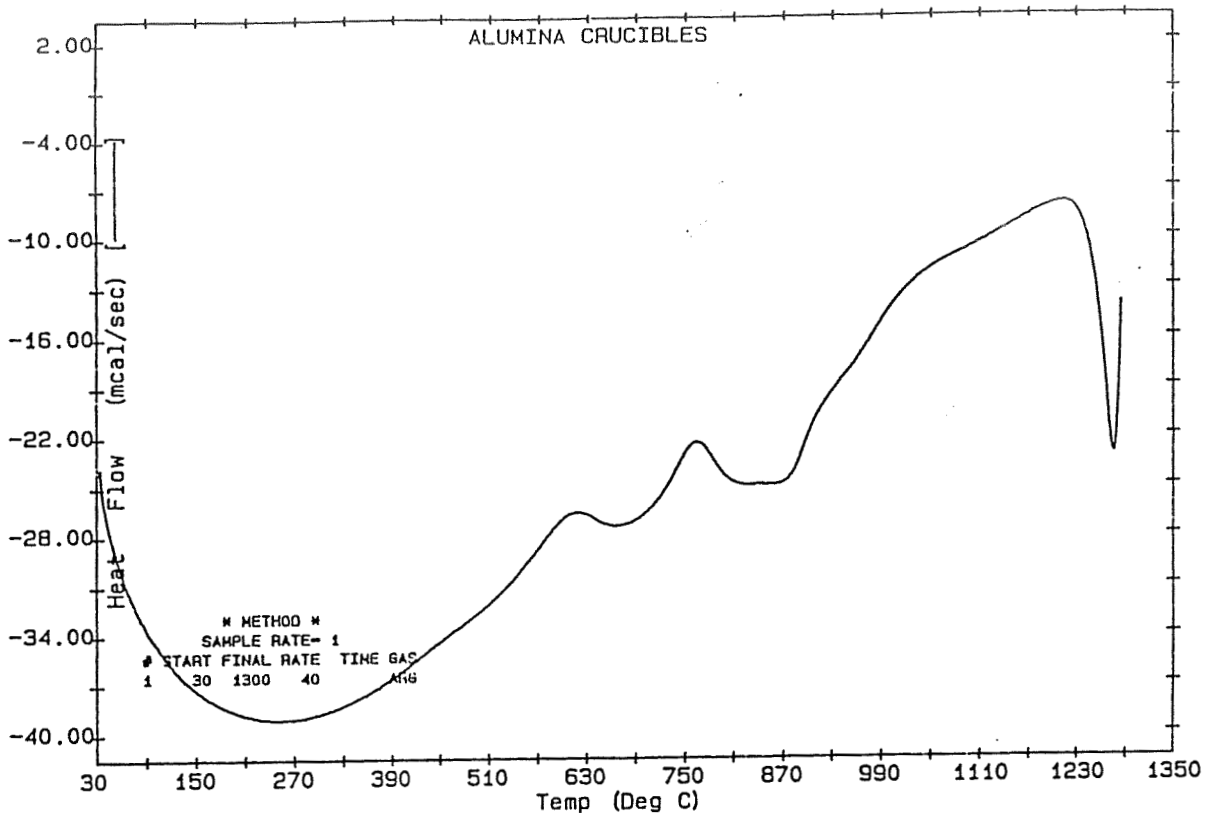


Figure 1. Time-temperature precipitation (TTP) diagrams for Inconel 718 illustrating the kinetics of formation for MC, M_6C, Laves, δ, and γ' phases.[2]

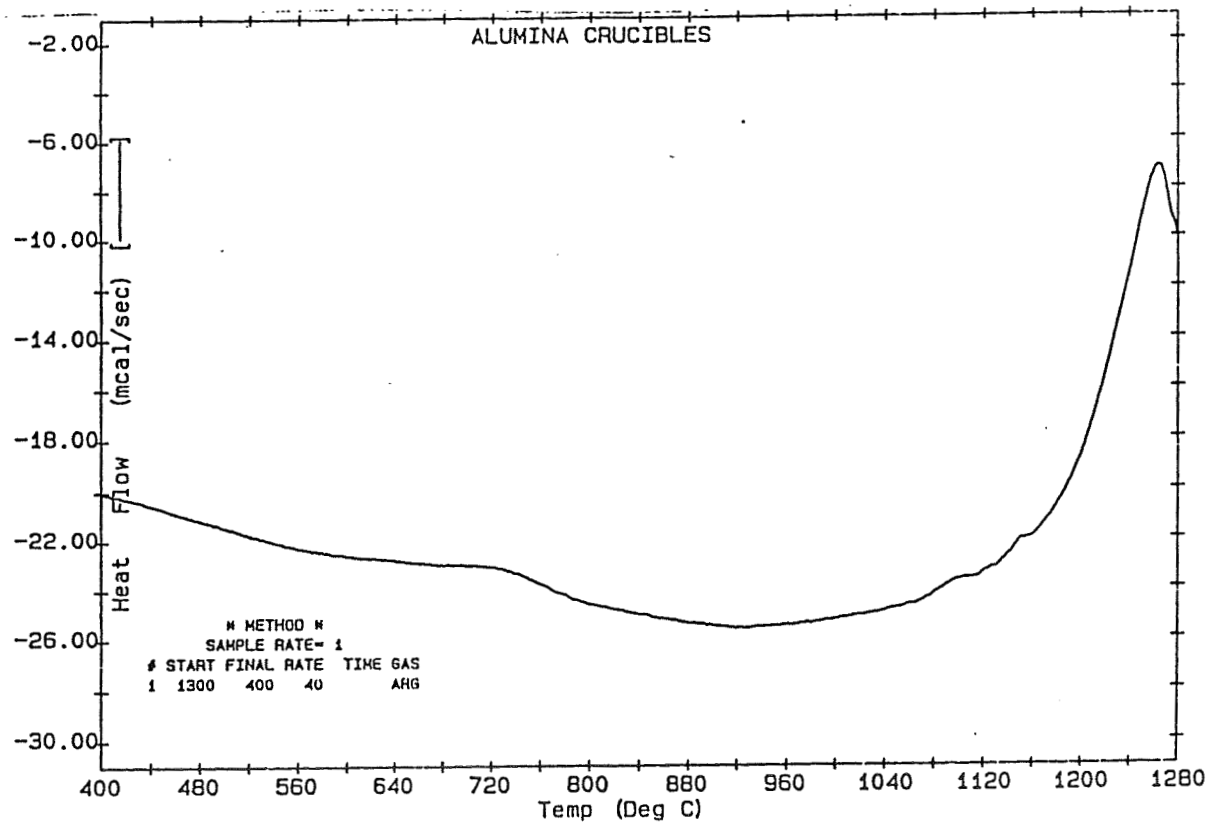


(a) (b) (c)

Figure 2. Typical microstructures, of IN 718 (a) wrought, (b) wrought grain grown, and (c) cast (100X).

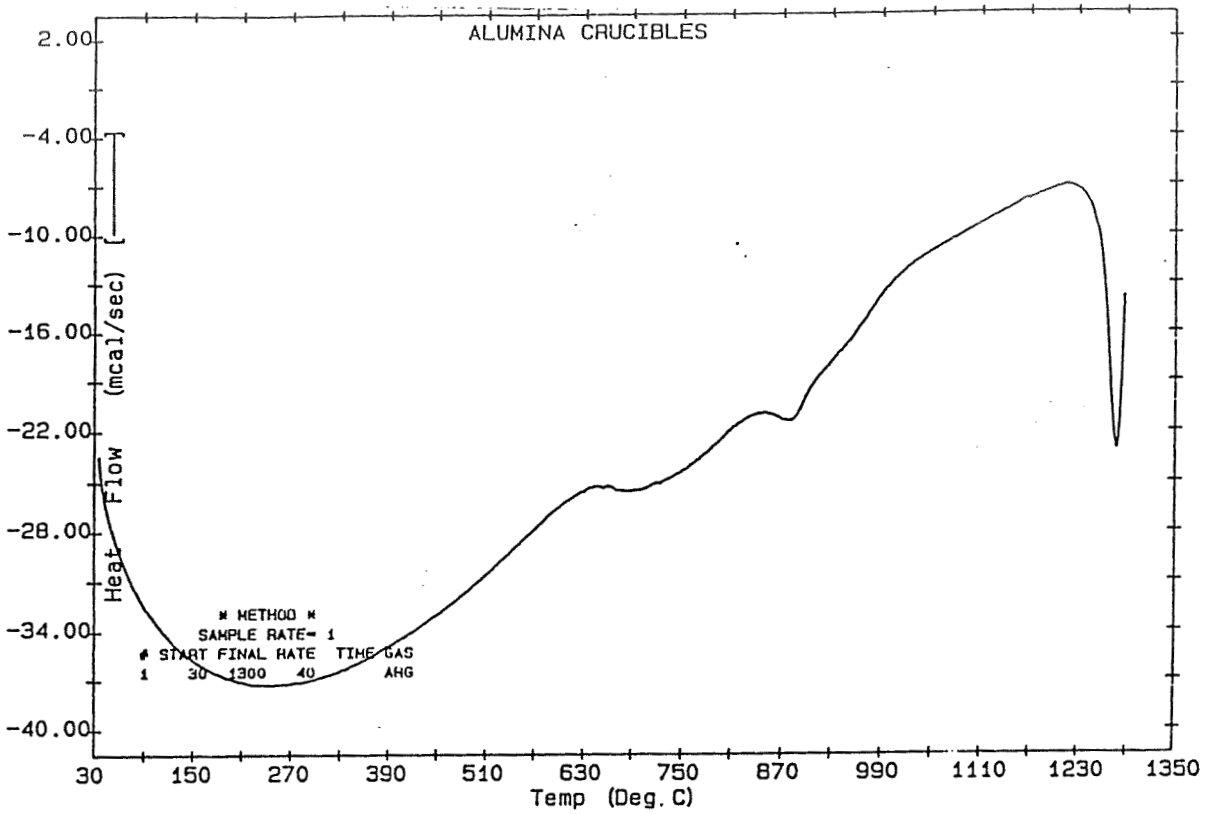


(a)

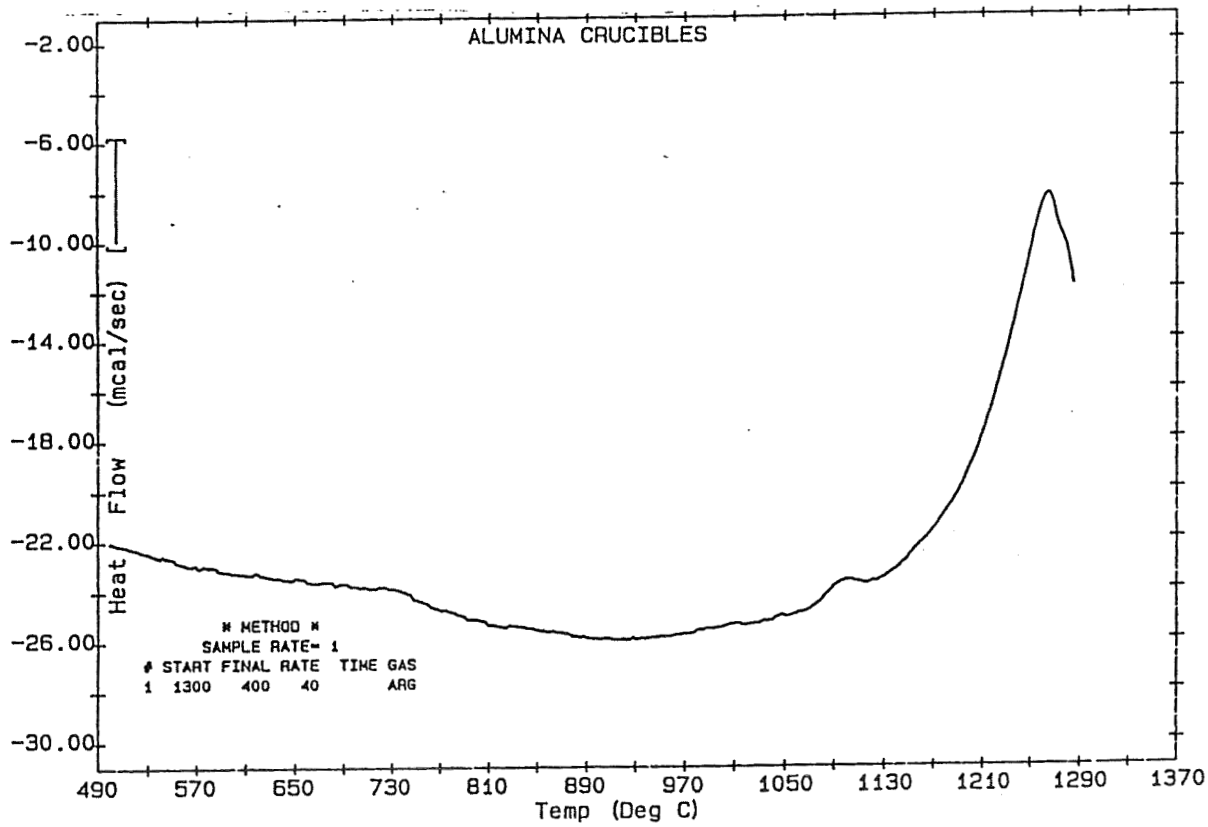


(b)

Figure 3. Differential Scanning Calorimetry curves of wrought IN 718 (a) Heating at 40°C/min, (b) cooling at 40°C/min.

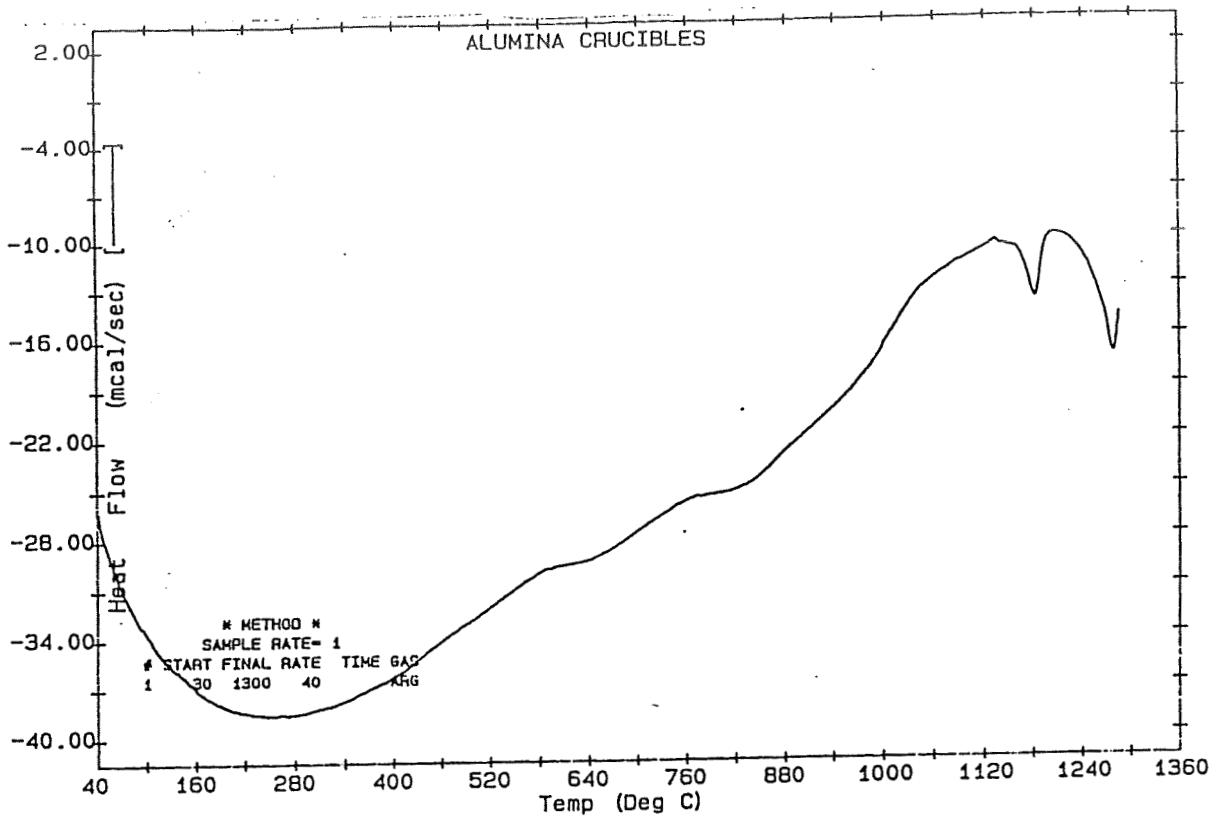


(a)

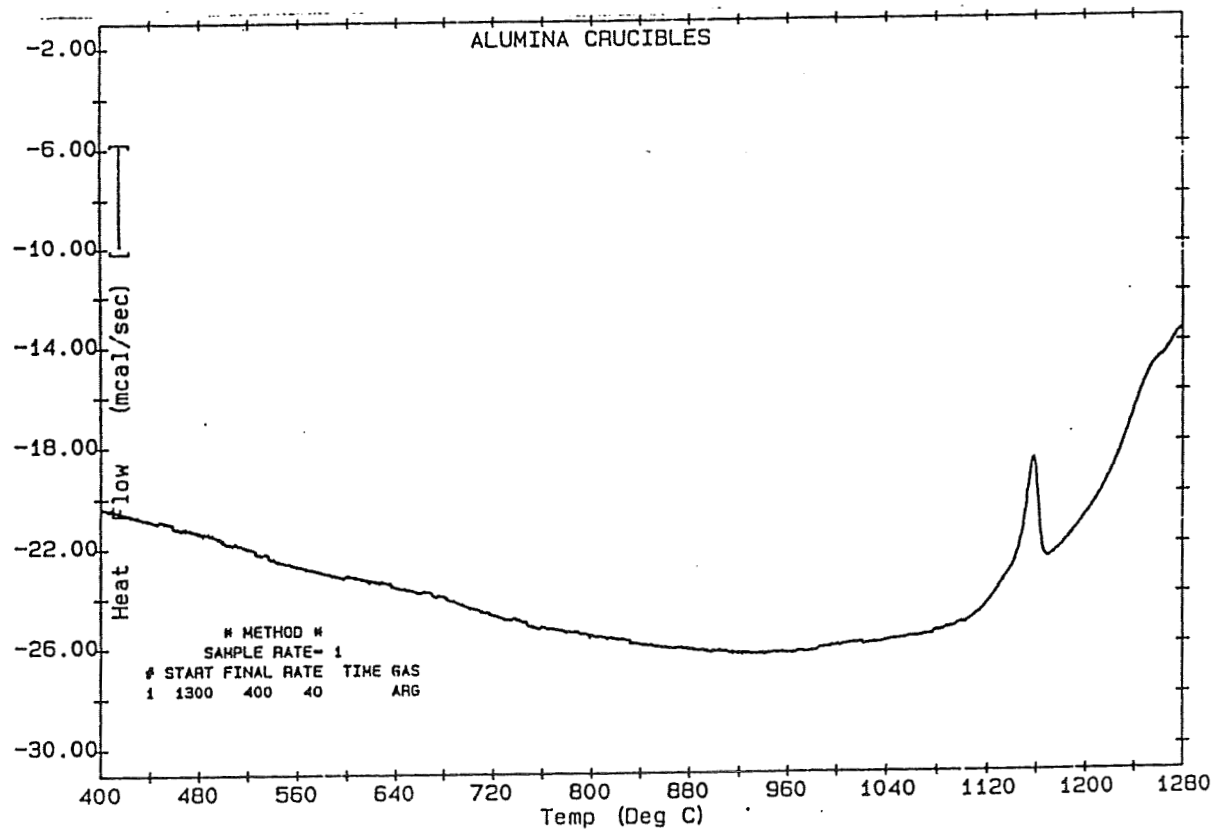


(b)

Figure 4. Differential Scanning Calorimetry curves of wrought grain grown IN 718 (a) Heating at 40°C/min, (b) cooling at 40°C/min.



(a)



(b)

Figure 5. Differential Scanning Calorimetry curves of cast IN 718 (a) Heating at 40°C/min, (b) cooling at 40°C/min.

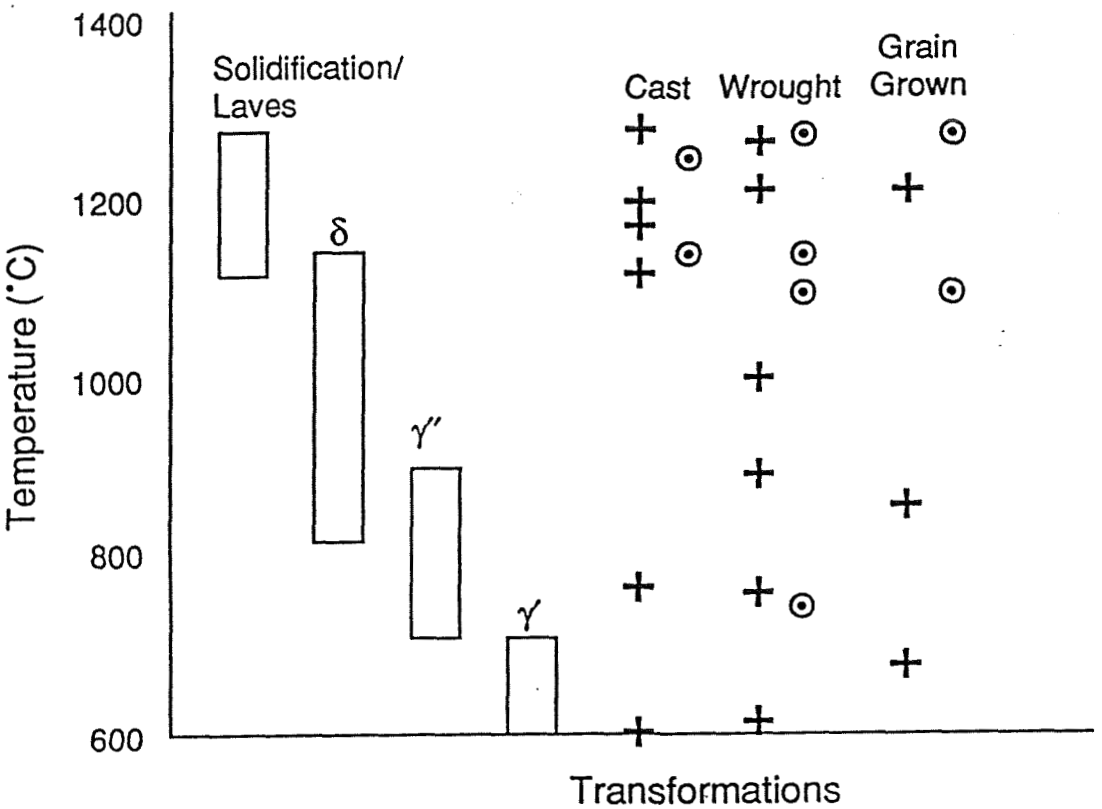
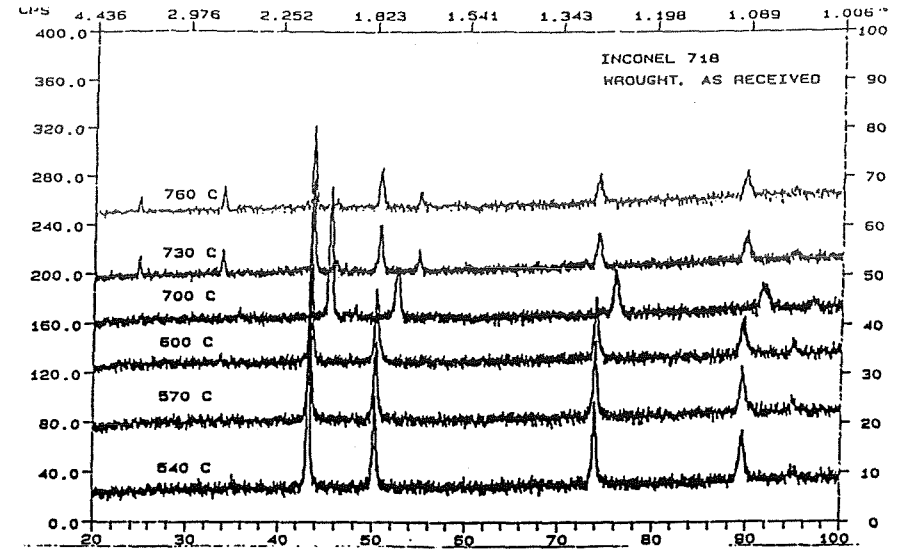
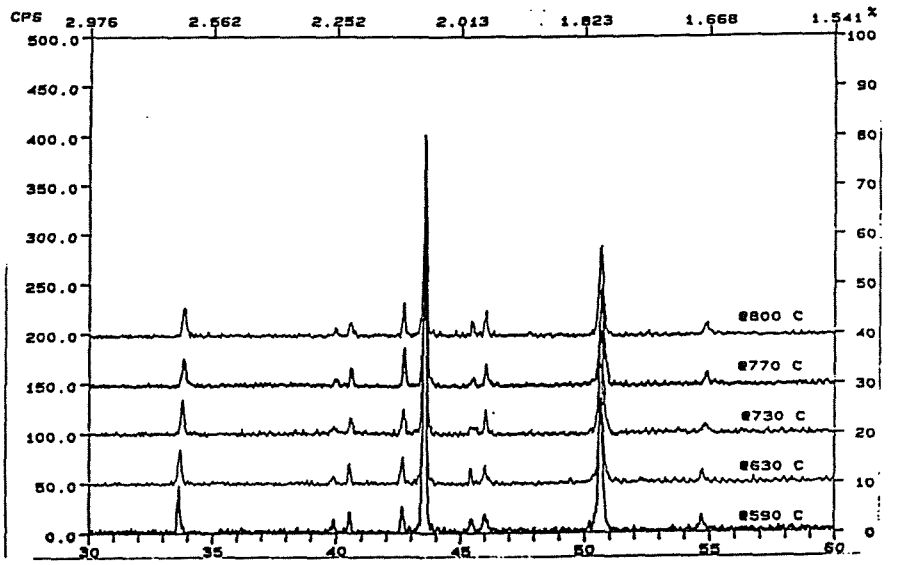


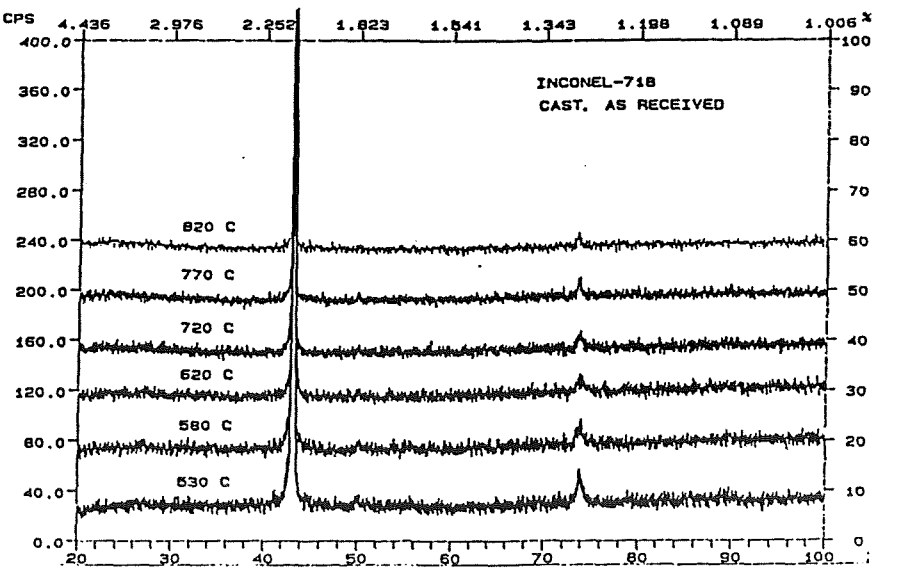
Figure 6. Transformation temperatures for phases in Inconel 718.[2]



(a)



(b)



(c)

Figure 7. High temperature x-ray diffraction of scans of wrought Inconel 718.

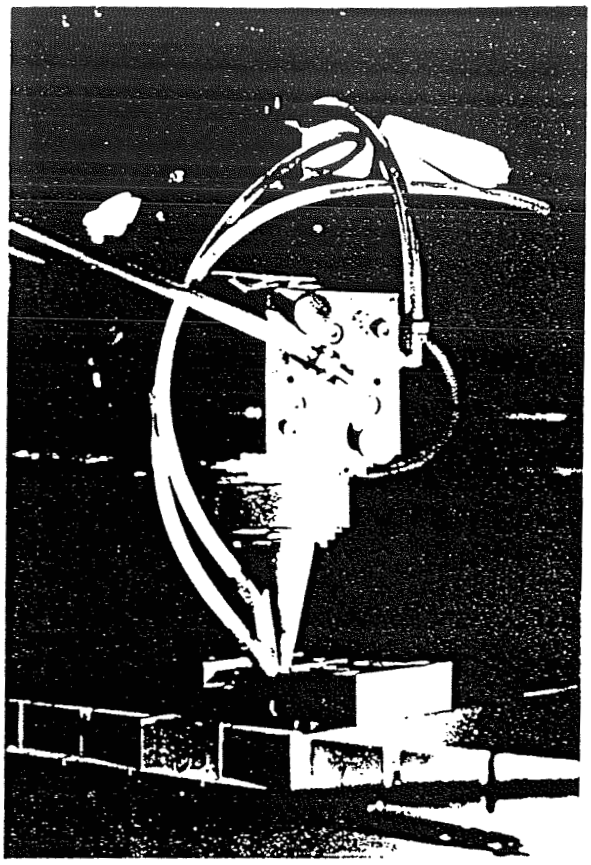


Figure 8. Welding fixture at nozzle.

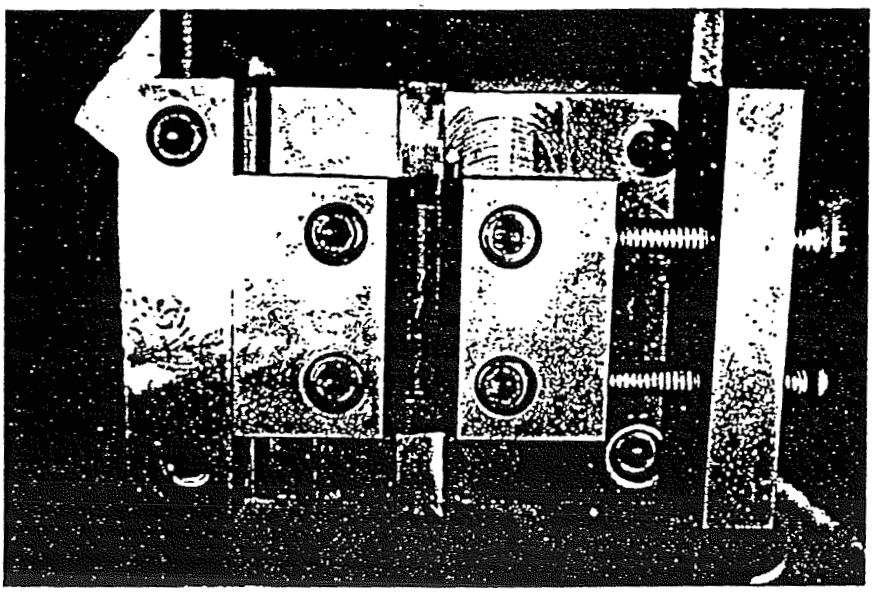


Figure 9. Welding fixture.

Mixed Frequency Pulsing (5kHz pulse width modulation on/Gated pulse on)
Pulse On Time = 35 ms
Pulse Off Time = 60 ms
Gate Duty Cycle = 37%
5kHz PWM Duty Cycle 50%
Peak Power 370 W
Pulse Average Power = 1850 W
Mixed Frequency Average Power 685 W

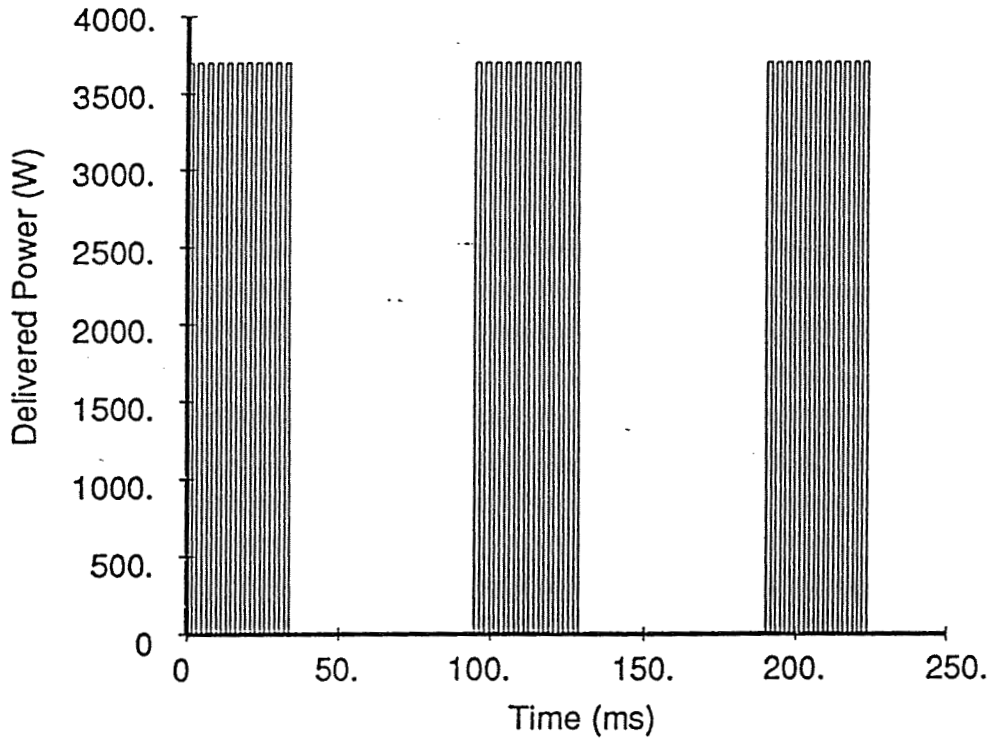


Figure 10. Mixed frequency pulsing format.

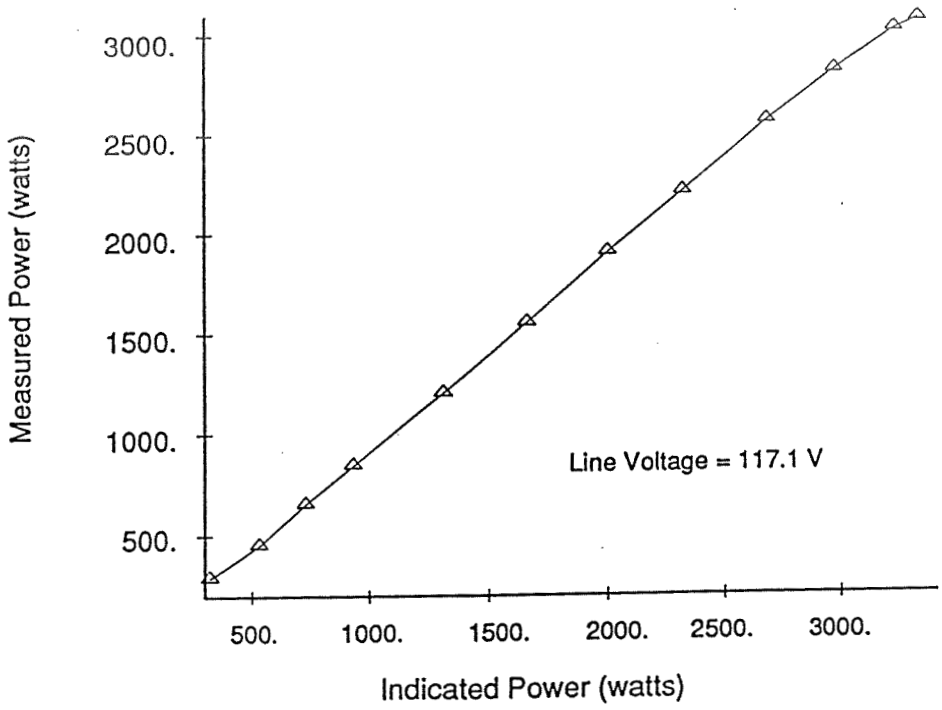


Figure 11. Results of laser power calibration.

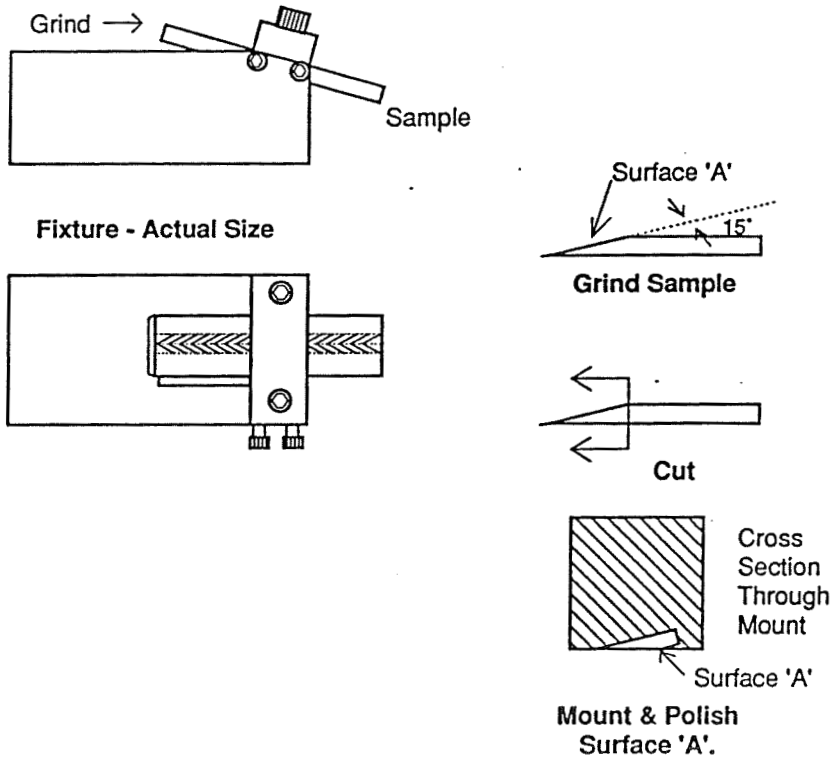


Figure 12. Preparation of slant-grind specimens.

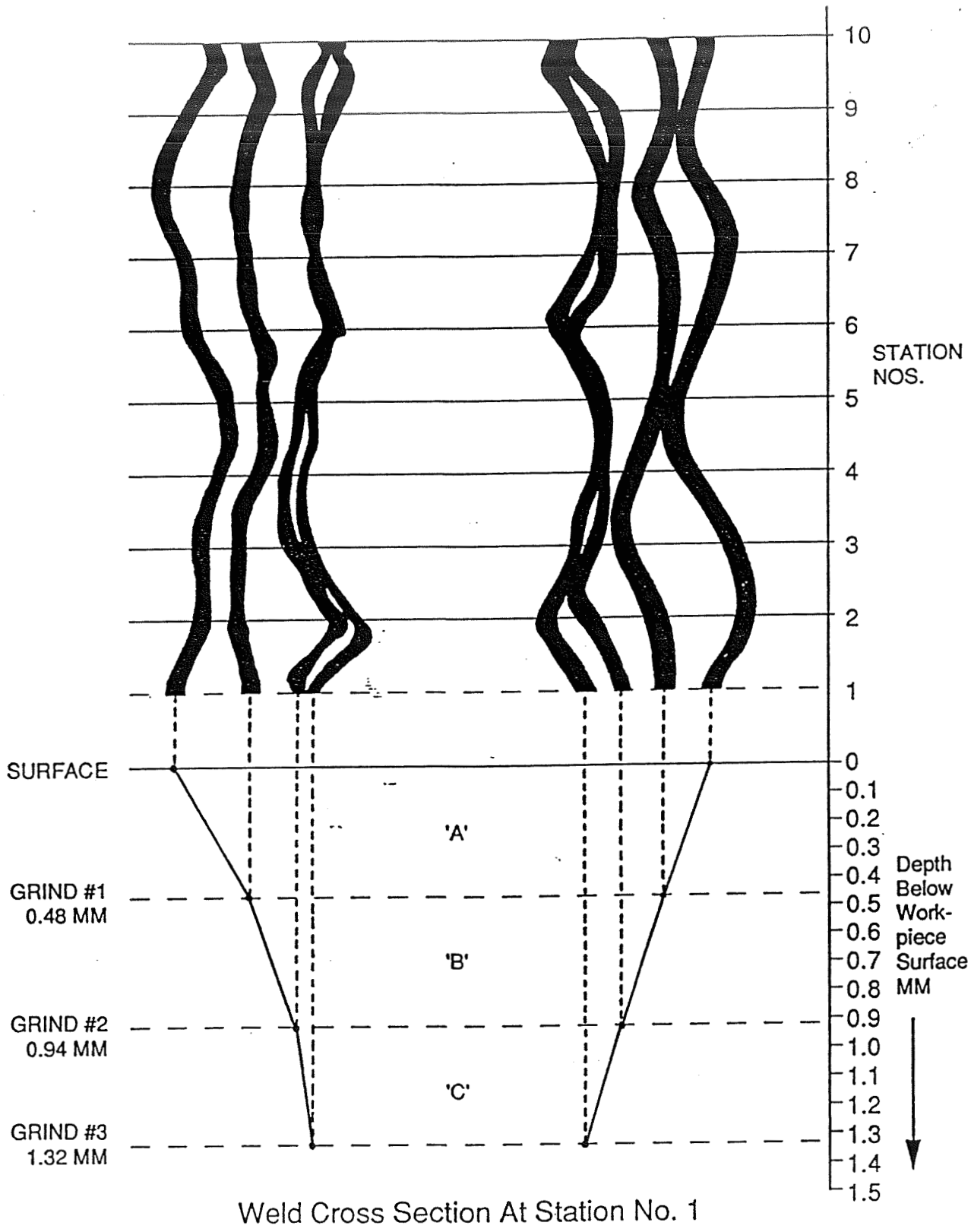


Figure 13. Transverse curvature measurement.

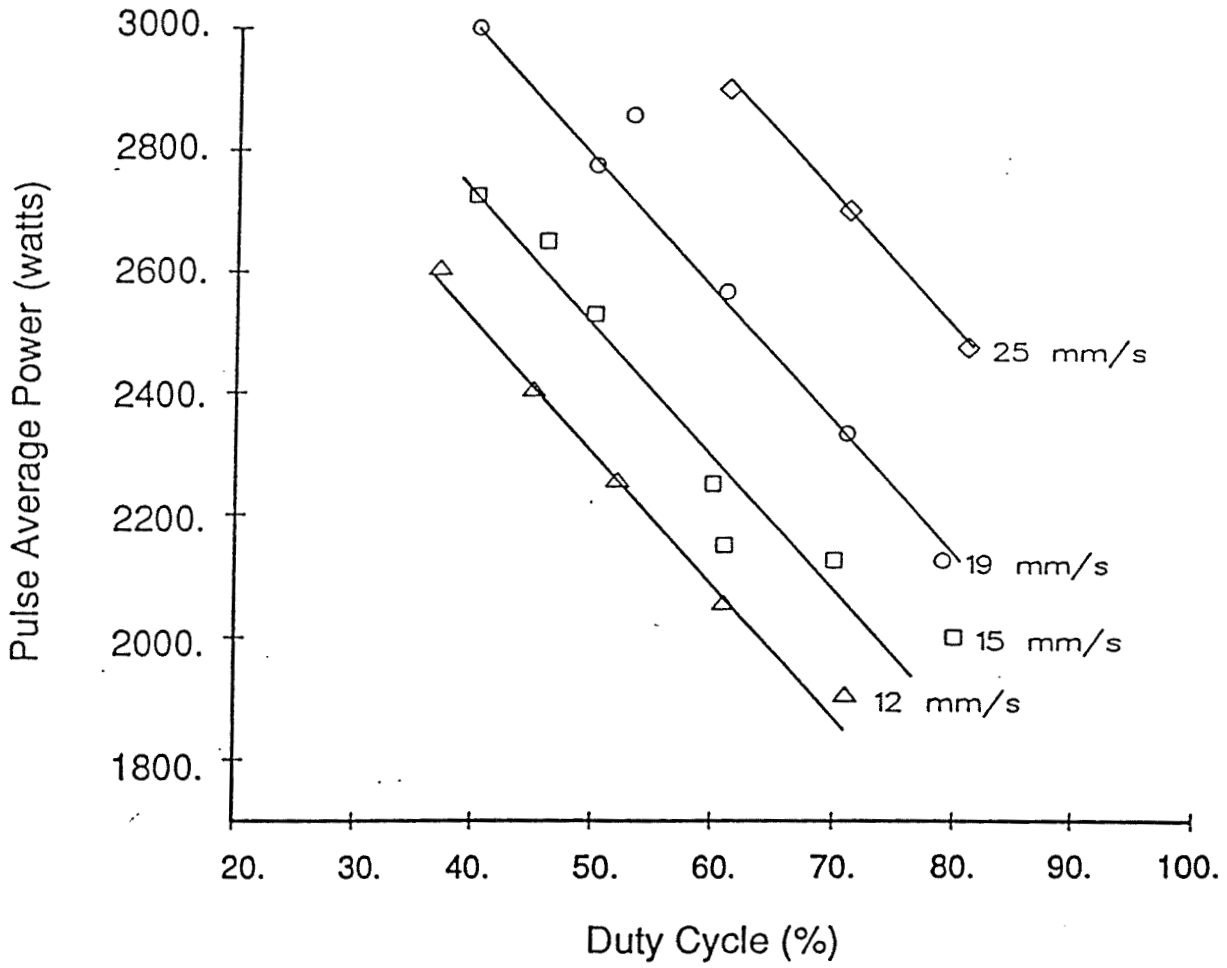


Figure 14. Unrestrained welds having full penetration - welded at four speeds.

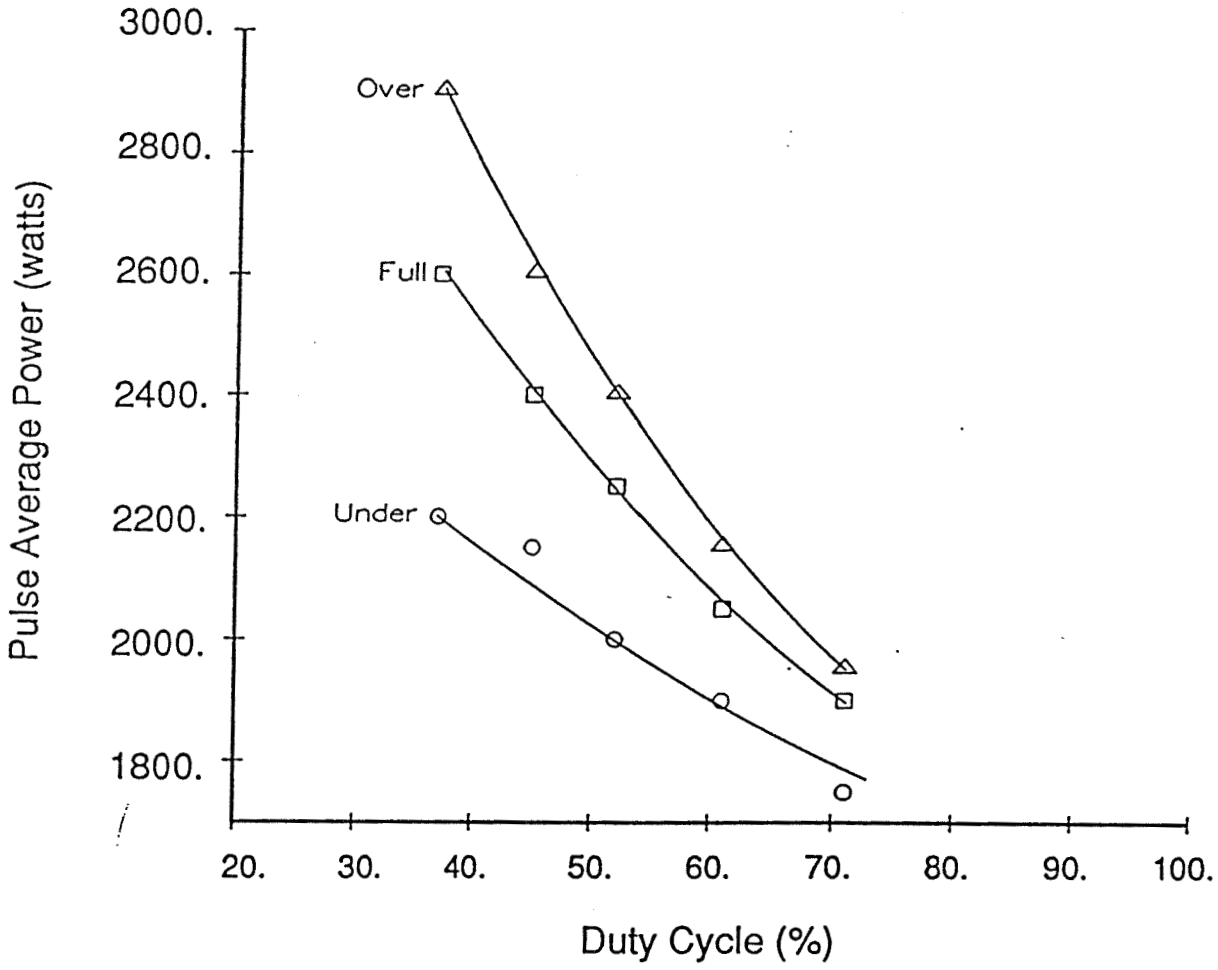


Figure 15. Penetration as it relates to pulse average power and duty cycle.

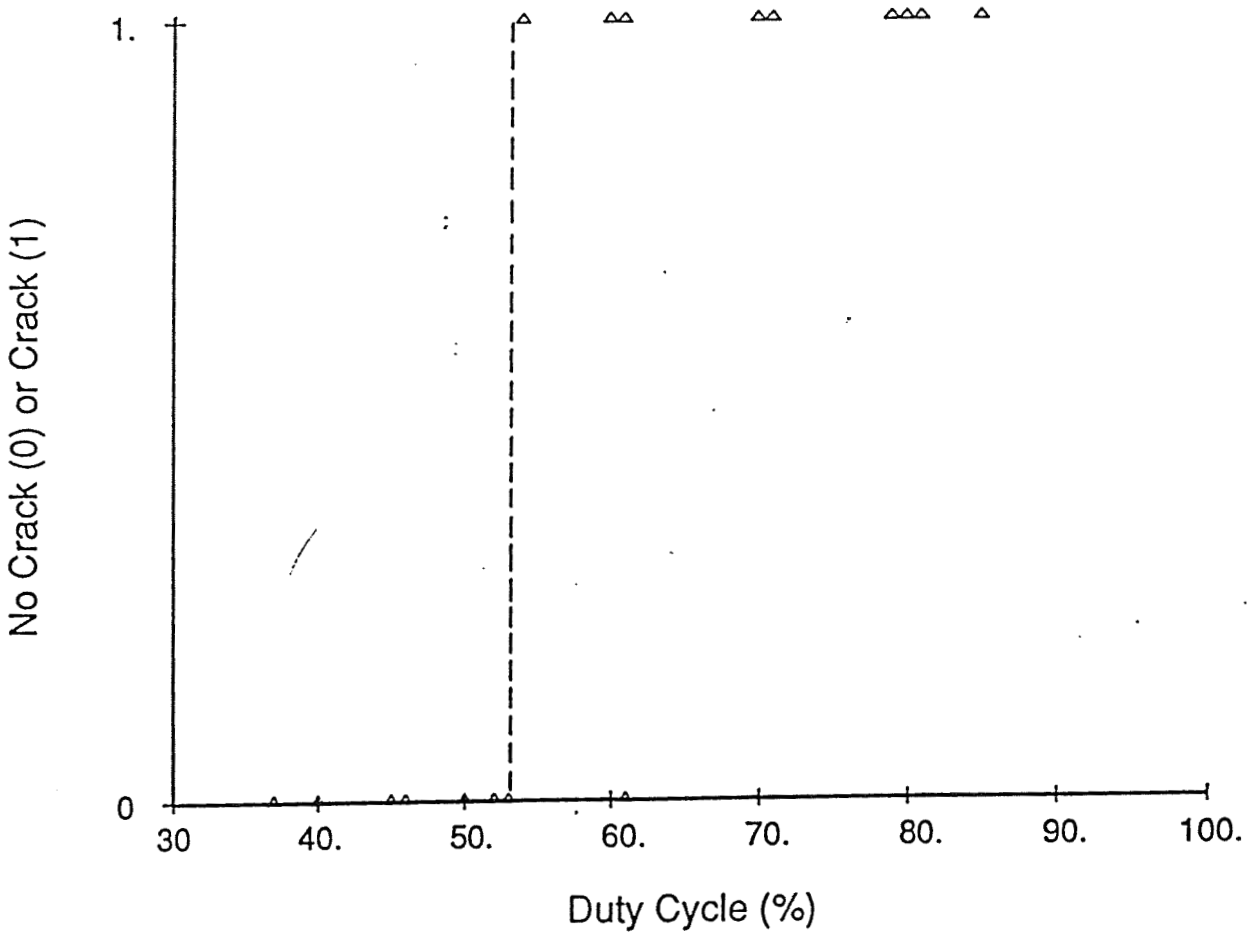


Figure 16. Centerline cracking as it relates to duty cycle in unrestrained welds.

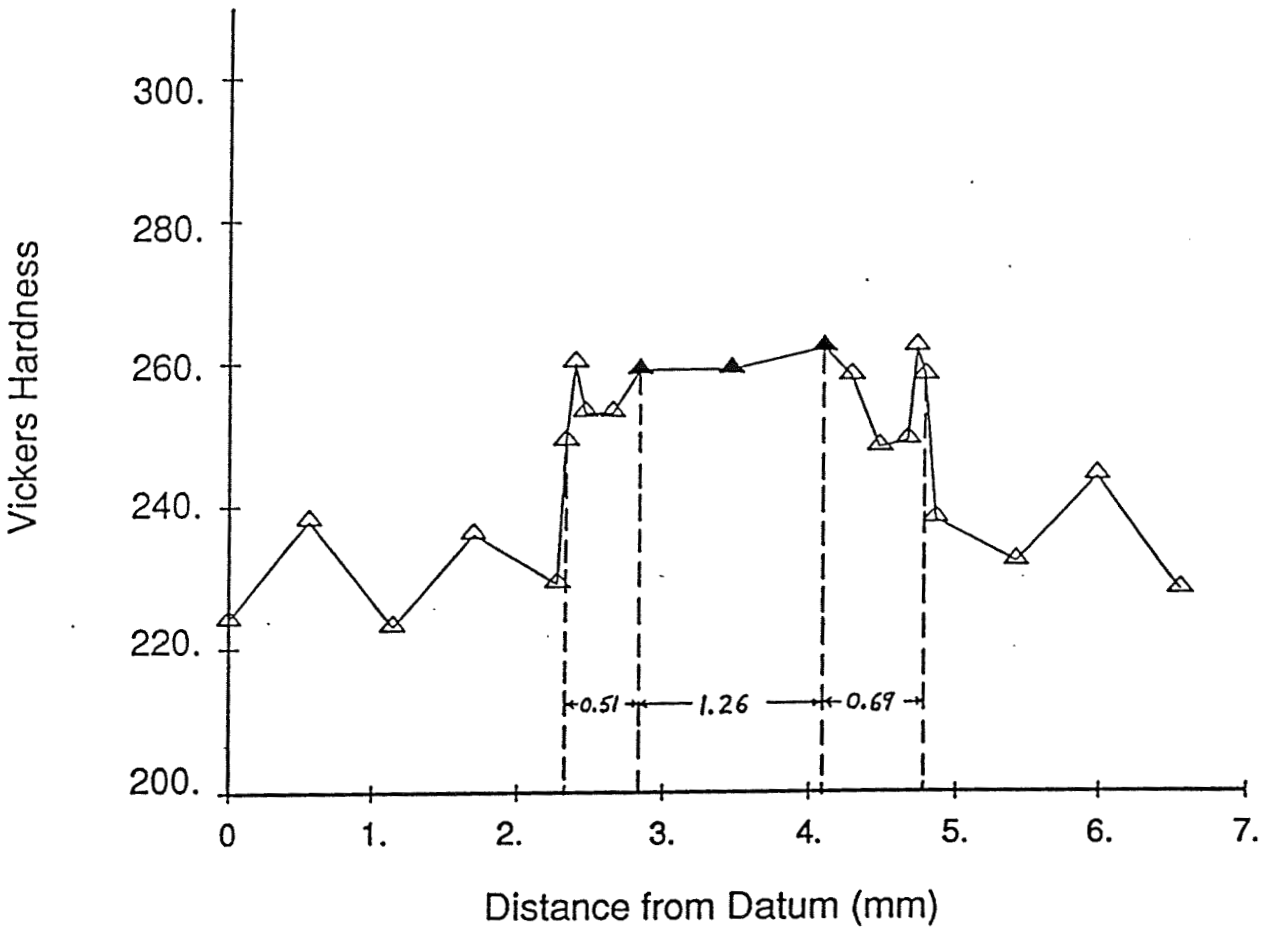


Figure 17. HAZ measurement for sample U9, 61% duty cycle, 2567 watts, speed 18 mm/s (Darkened points visually determined to be in fusion zone).

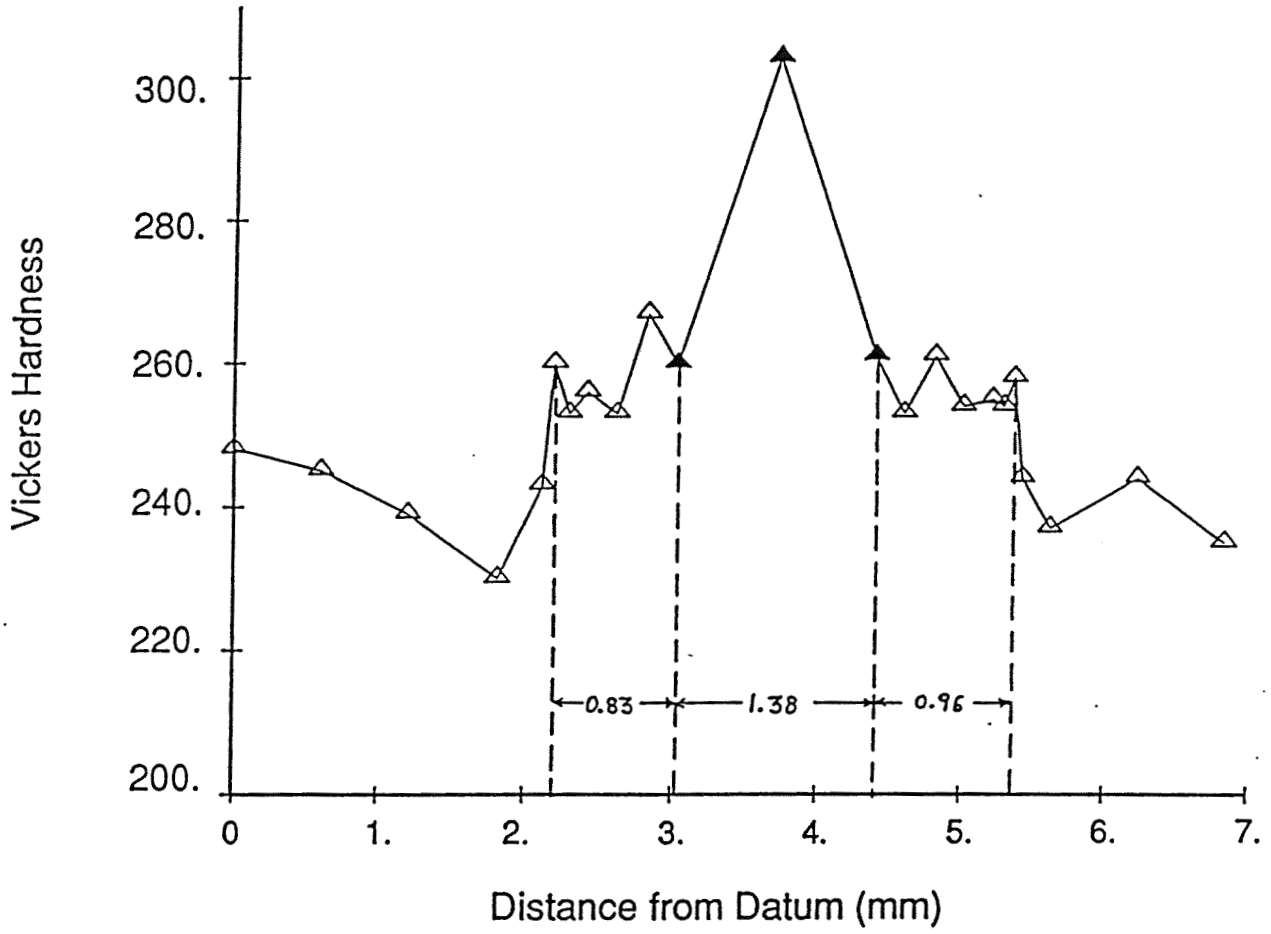


Figure 18. HAZ measurement for sample U28, 71% duty cycle, 1900 watts, speed 12 mm/s (Darkened points visually determined to be in fusion zone).

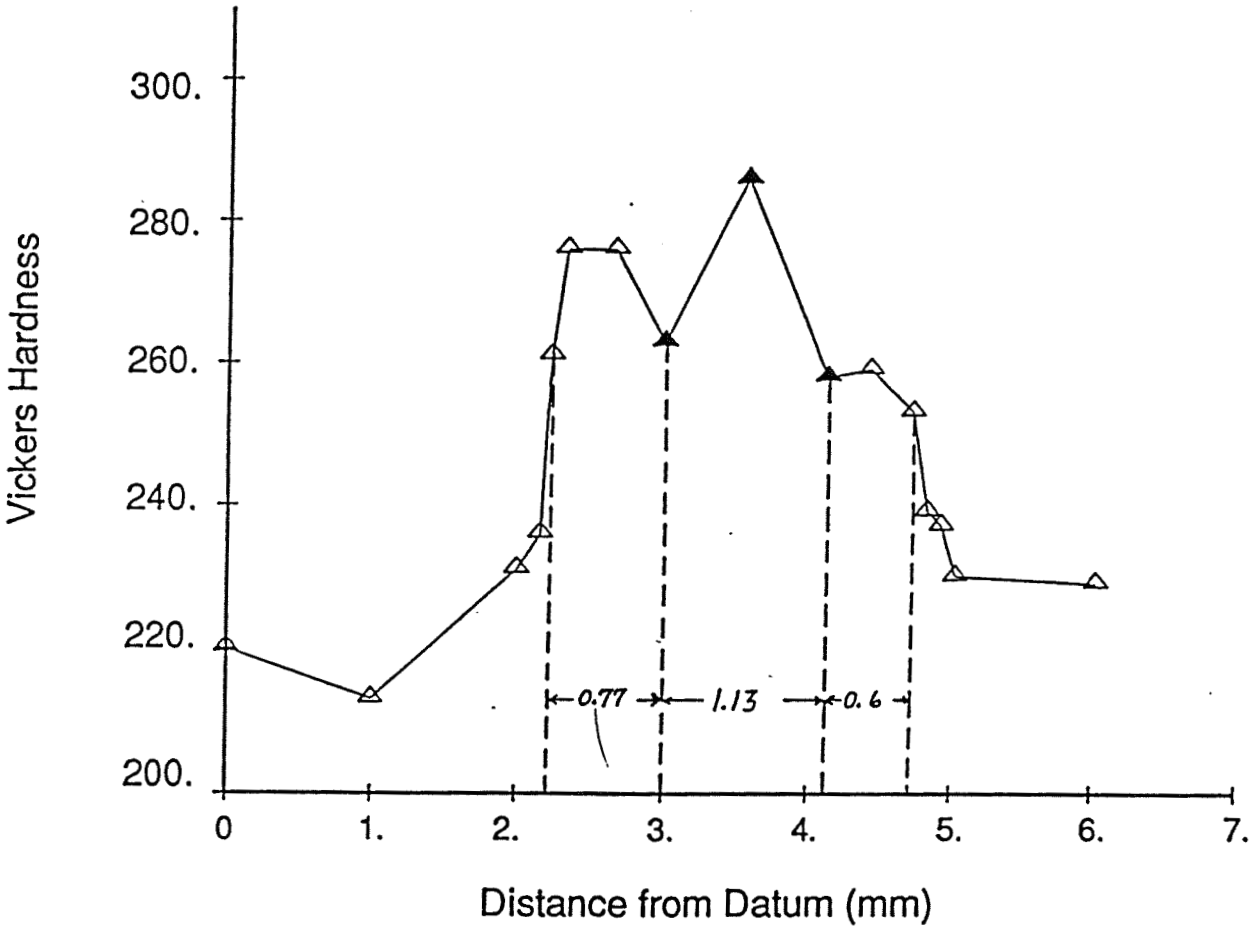


Figure 19. HAZ measurement for a CW weld, 2500 watts, speed 25 mm/s (Darkened points visually determined to be in fusion zone).

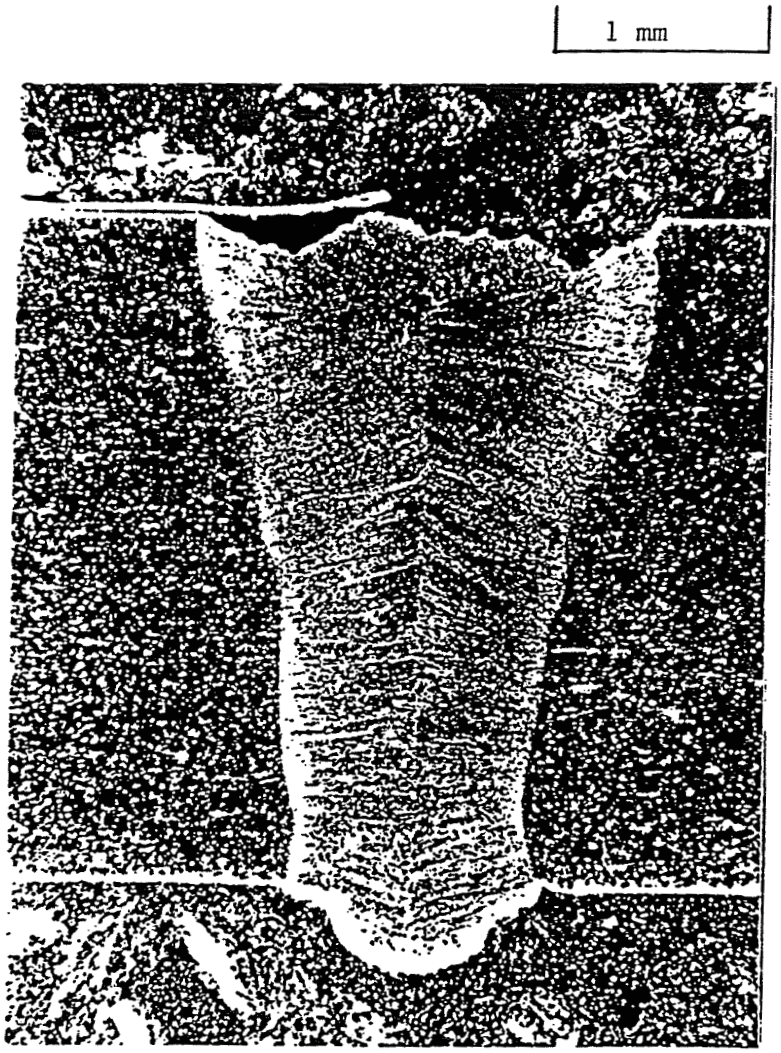


Figure 20. Fully penetrated restrained weld, 2250 watts, 61 percent duty cycle (Condition R13).

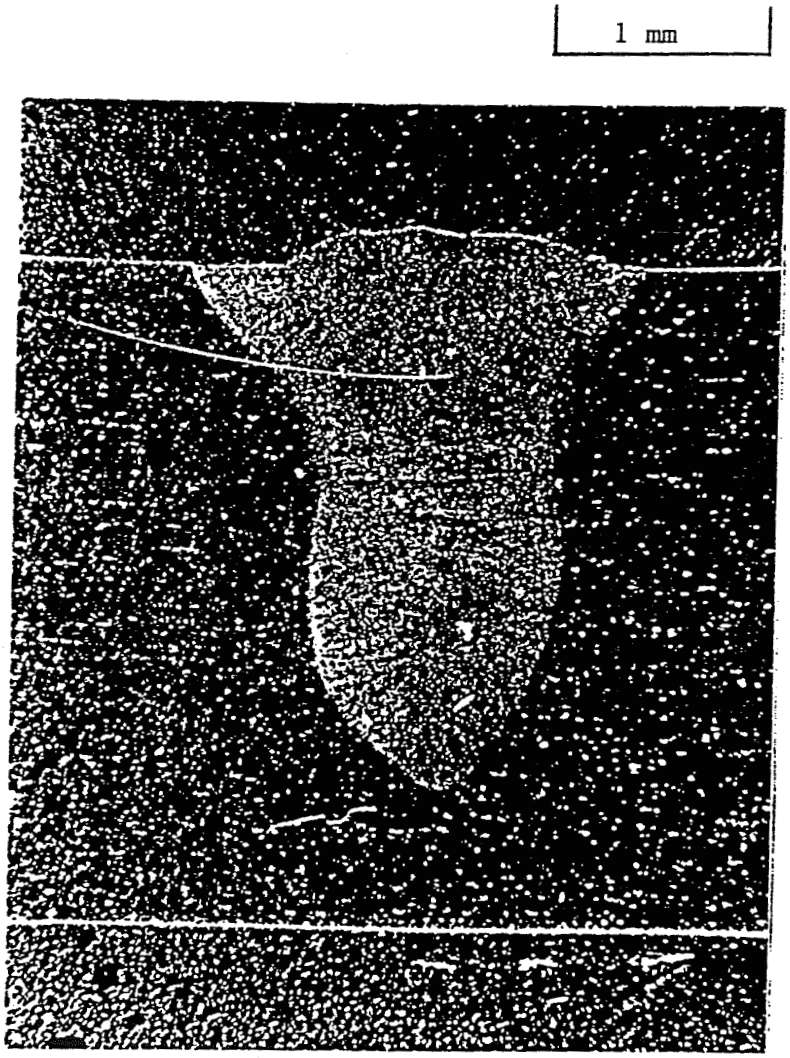


Figure 21. Underpenetrated restrained weld, 2050 watts, 37 percent duty cycle (Condition R10).

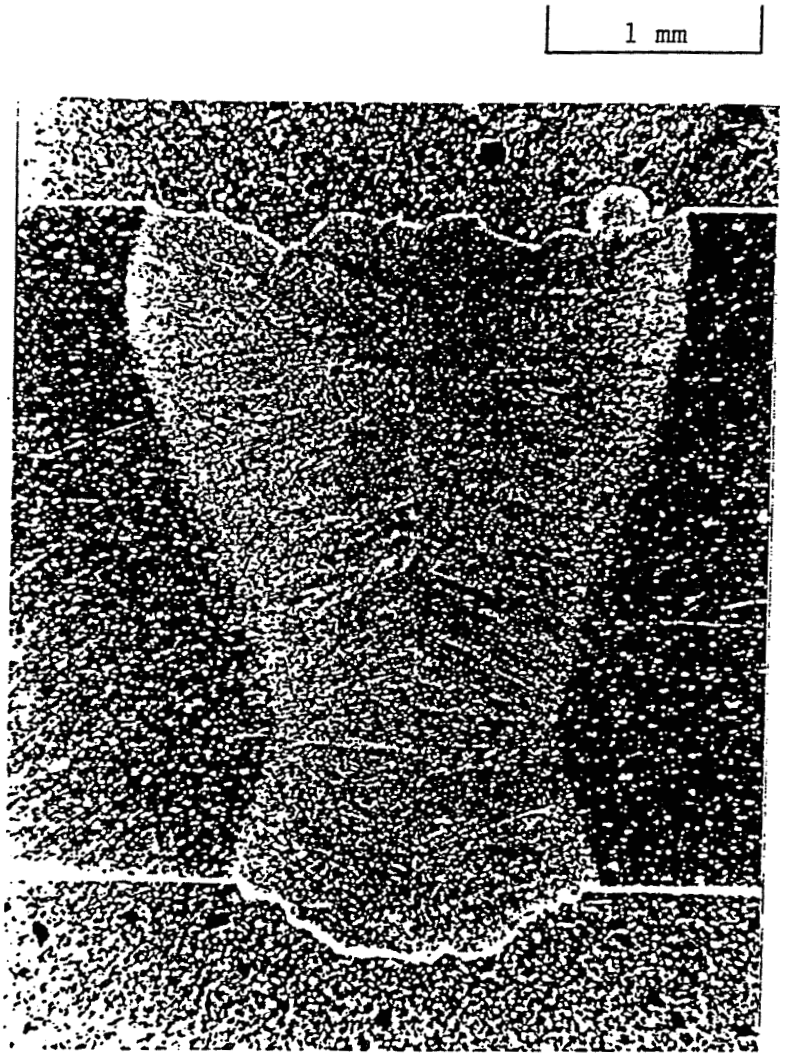


Figure 22. Overpenetrated restrained weld, 2600 watts, 71 percent duty cycle (Condition R17).

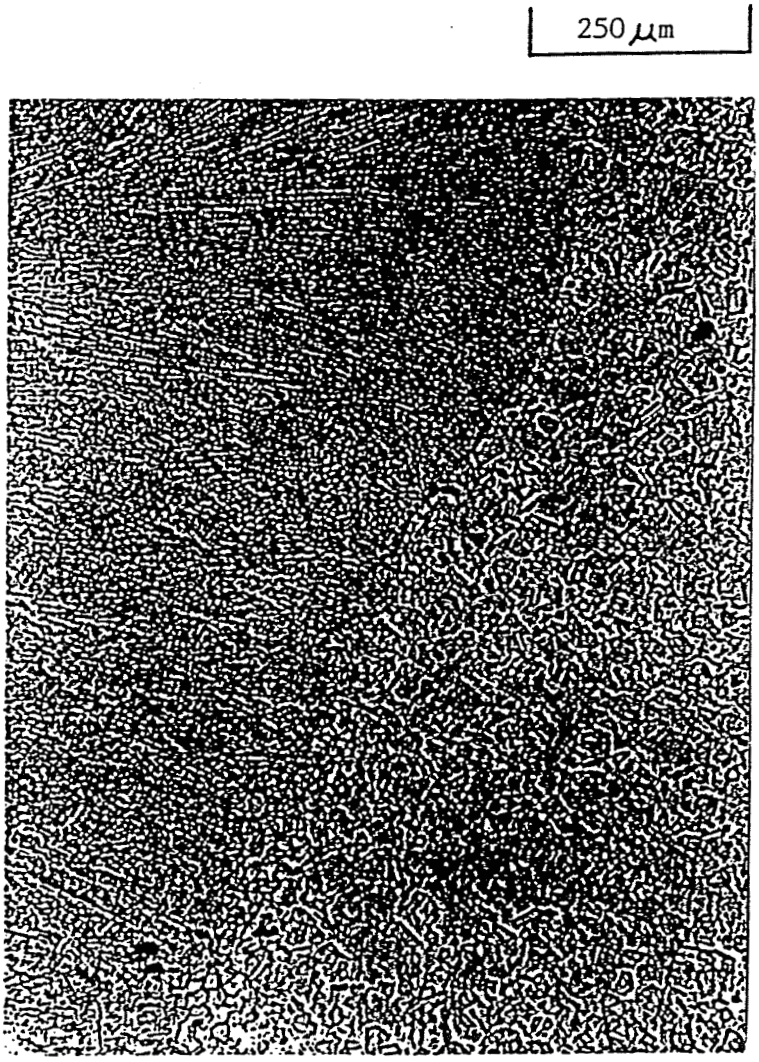


Figure 23. Typical view of weld nailhead/HAZ interface for Condition R6 (2050 watts, 80 percent duty cycle).

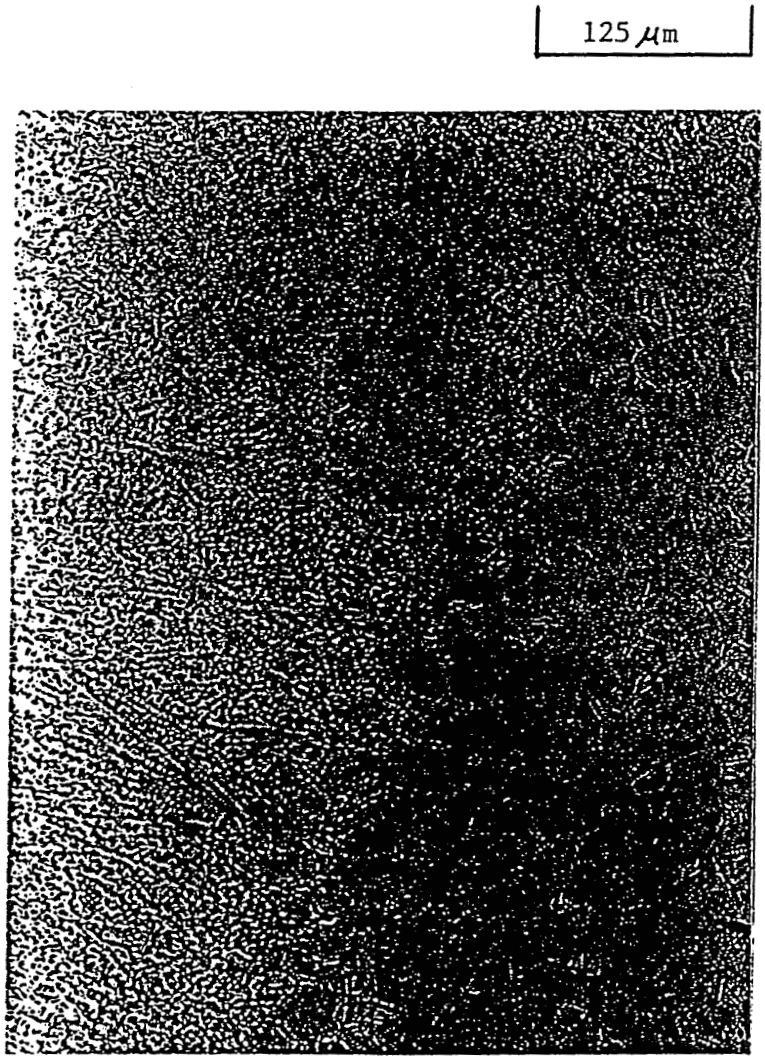


Figure 24. Weld nailhead showing weld structure and grain boundaries for Condition R6 (2050 watts, 80 percent duty cycle).

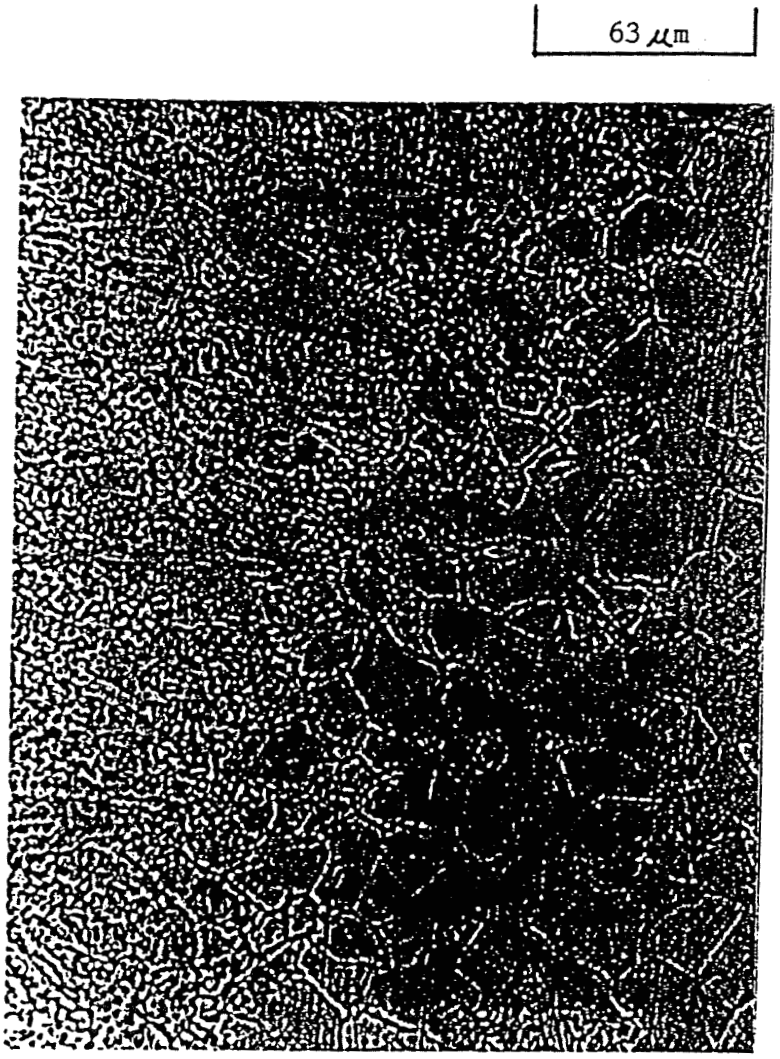


Figure 25. Typical weld HAZ - note the dark grain boundary precipitates. Condition R6 (2020 watts, 80 percent duty cycle).

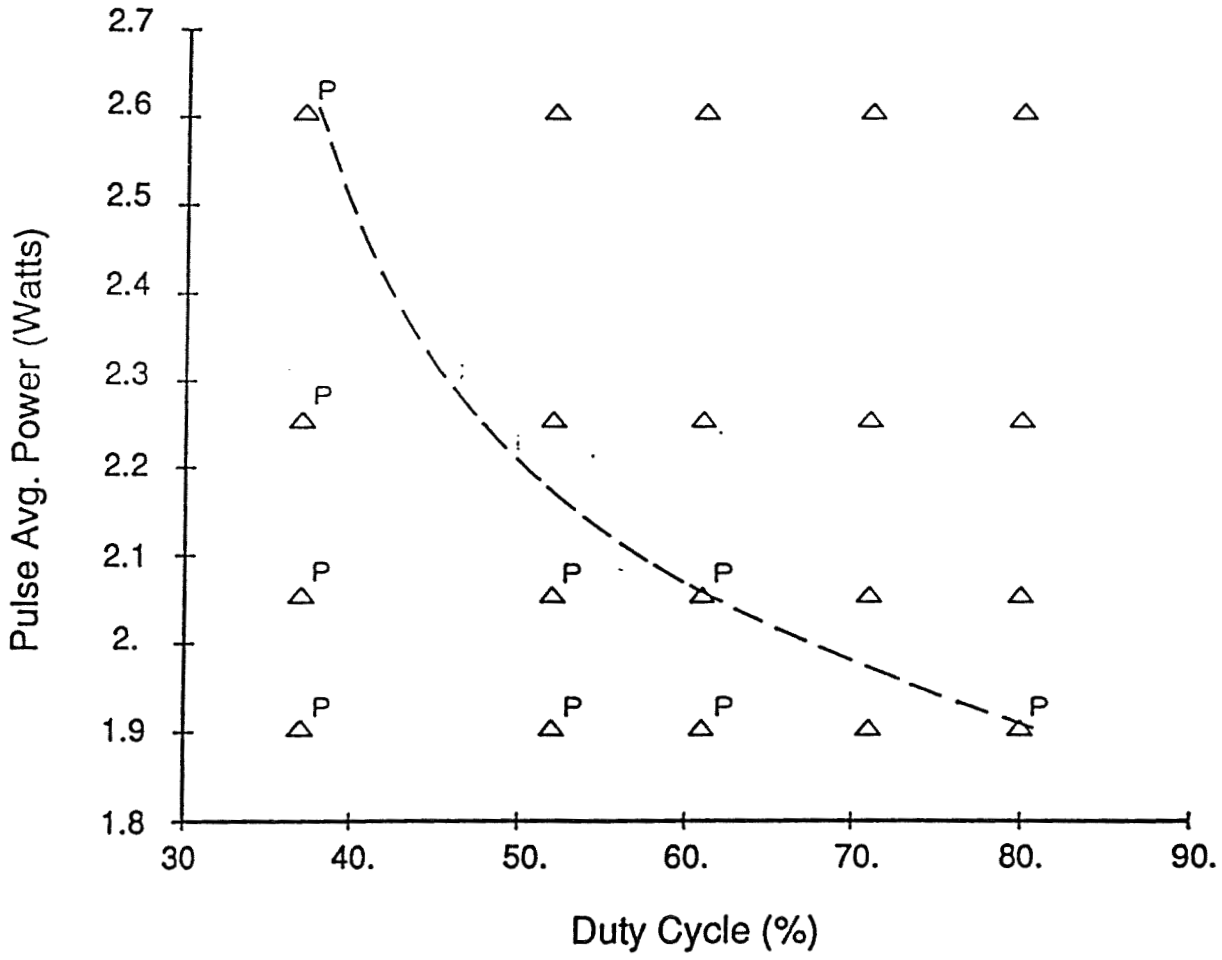
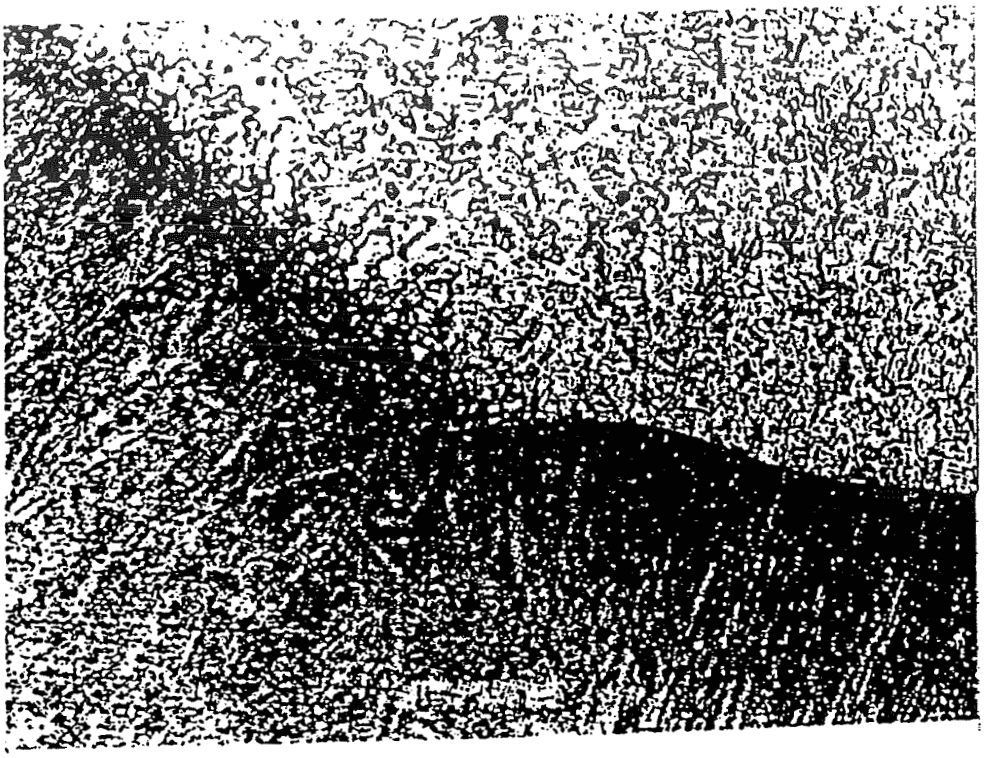
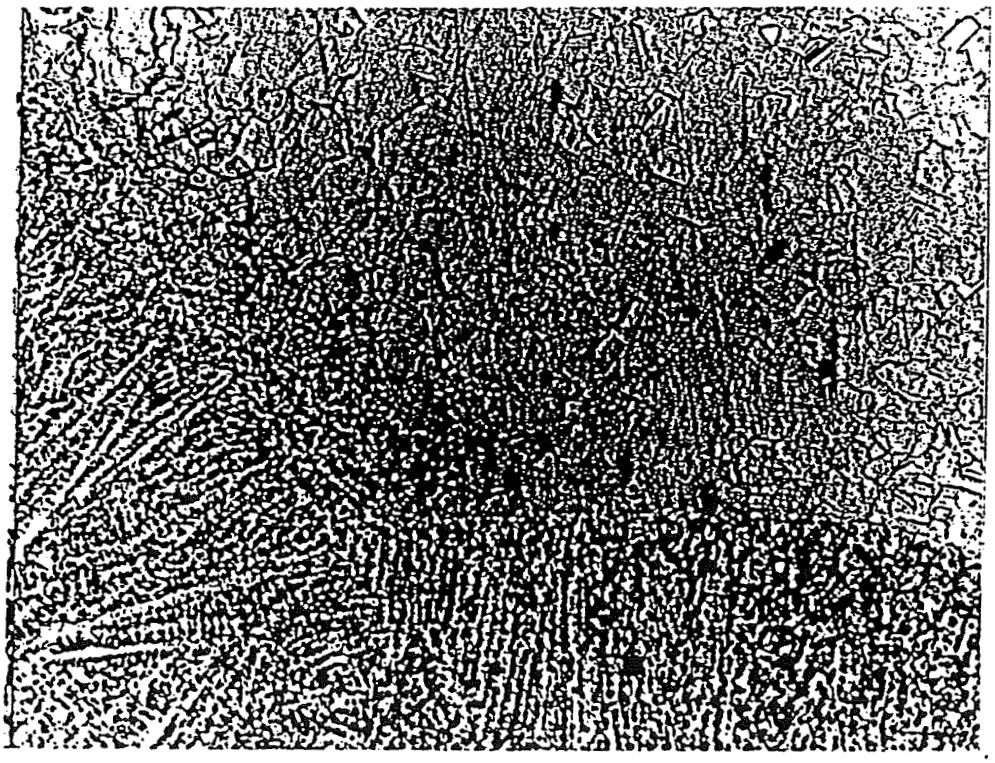


Figure 26. Restrained weld matrix (P indicates samples containing pores).

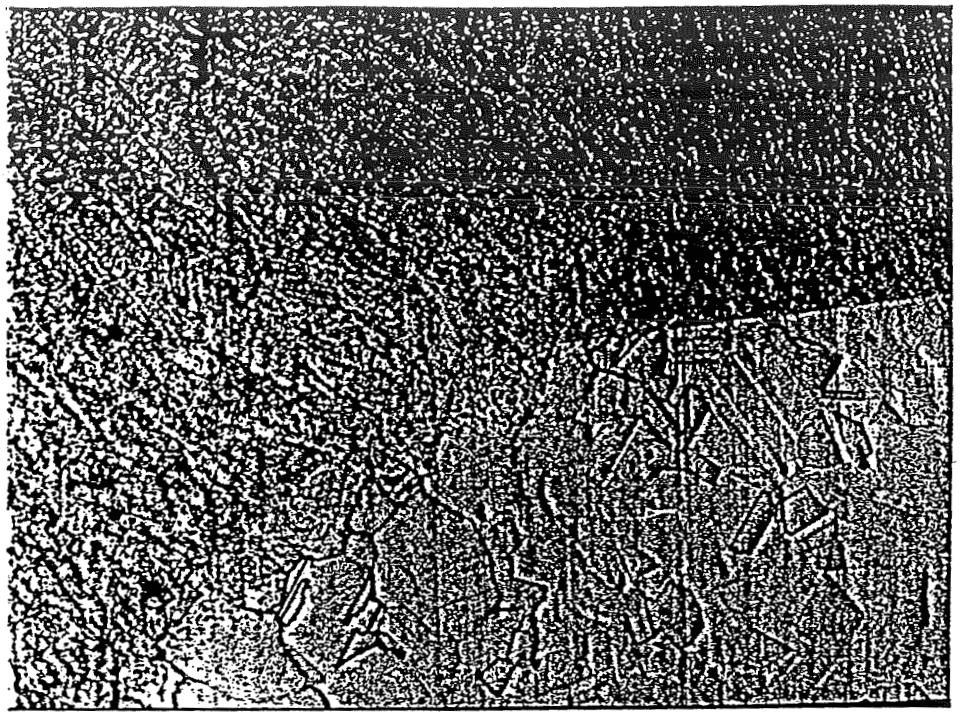


(a)

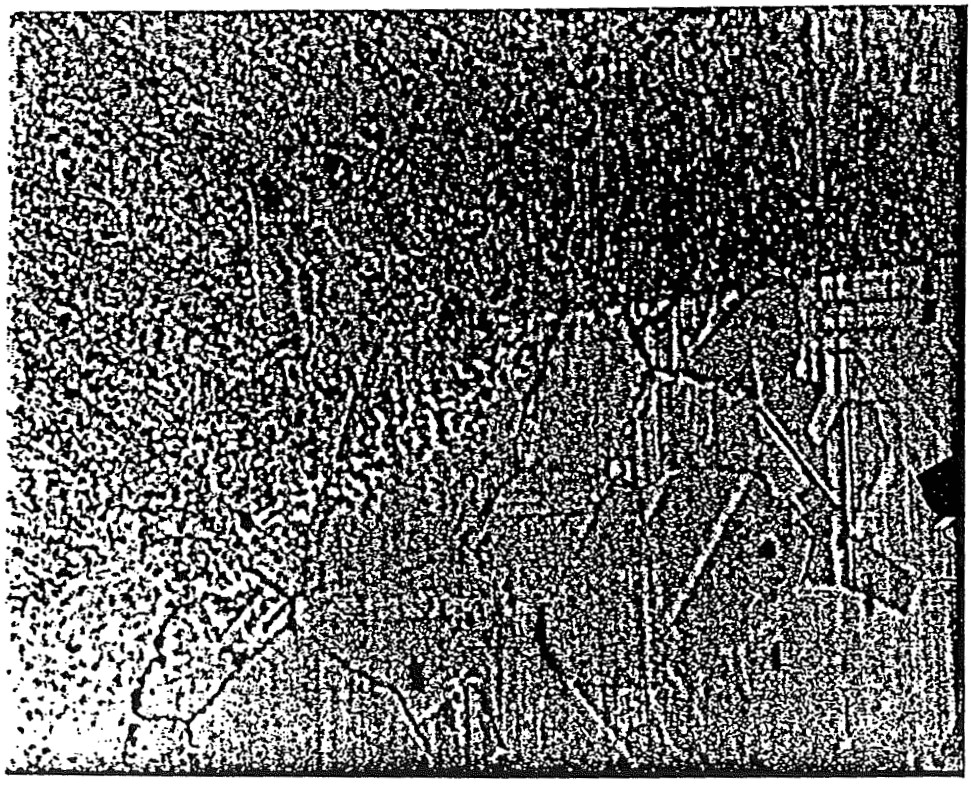


(b)

Figure 27. Microstructure of wrought IN 718 in the region of the nailhead (a) 100X and (b) 200X.

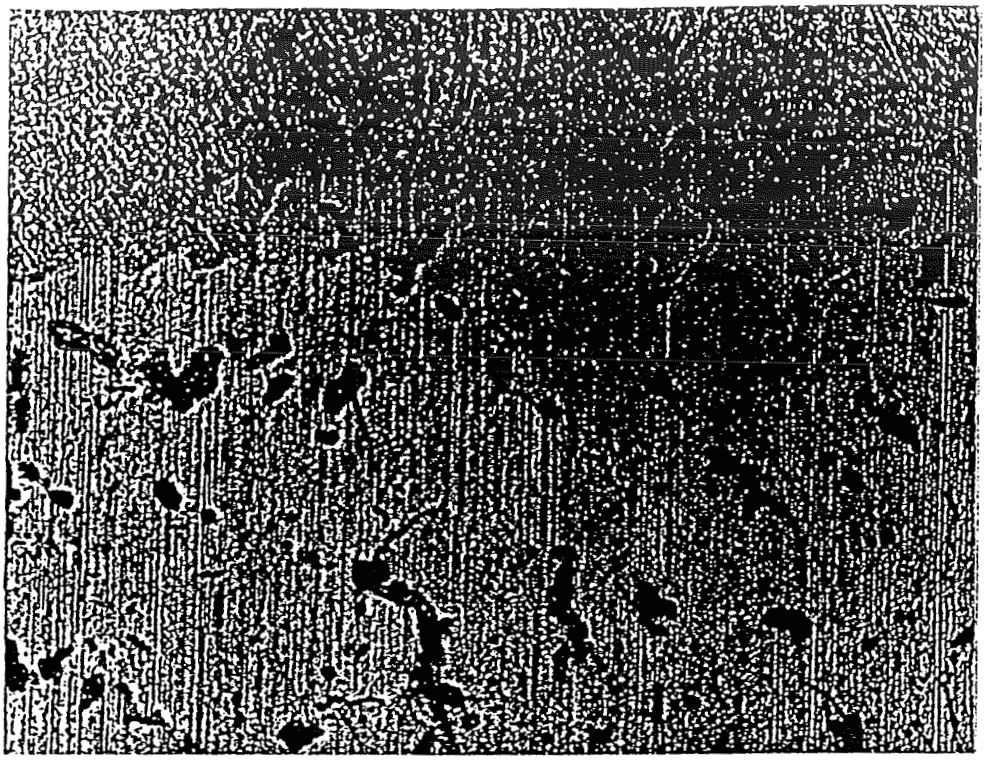


(a)

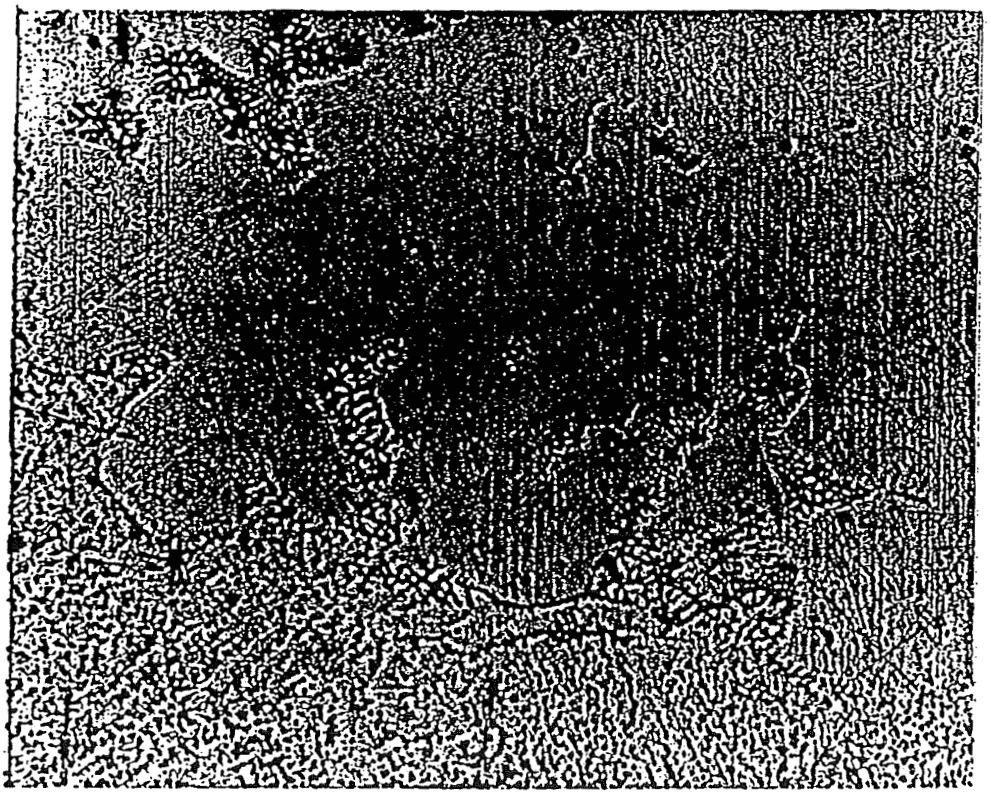


(b)

Figure 28. Microstructure of wrought grain grown IN 718 in the region of the nailhead (a) 100X and (b) 200X.



(a)



(b)

Figure 29. Microstructure of cast IN 718 in the region of the nailhead (a) 100X and (b) 200X.