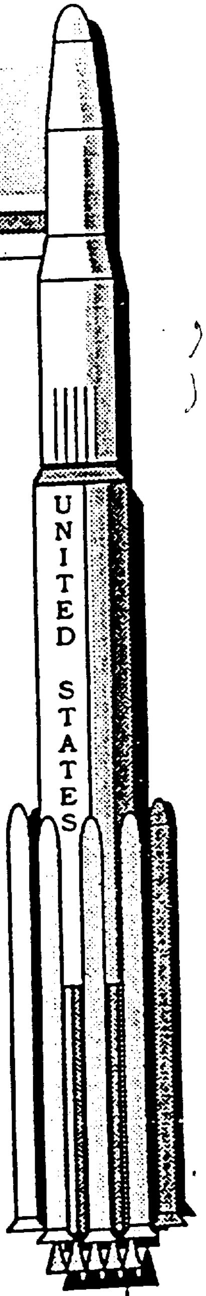
NASA-CR-199689

C A S P

Center for Advanced Space Propulsion



Annual Report

November 1, 1988
—
October 31, 1989
NAGW - 1195

ATTACHMENT D
NAGW-1195
CSTAR-FINAL REPORT

GL-0736AR

PAGE

ANNUAL FIGURE/TABLE LIST

EXECUTIVE SUMMARY

The Center for Advanced Space Propulsion (CASP) is one of sixteen NASA sponsored Centers established to facilitate the Commercial Development of Space. It is the only Commercialization Center with a technology concentration in propulsion. CASP has completed the second year of operation with a significant list of accomplishments:

The number of industrial partners has increased to nine (9), almost double the original number.

The number of active projects has increased to fourteen (14), almost three times the original number. Fifteen (15) other projects are in various stages of development.

Four projects leading to flight experiments are being actively pursued.

The selection of a contractor for the CASP office/laboratory building is underway. Occupancy is expected in the Fall of 1990.

Fifteen Technical papers were published or presented in national conferences by CASP principal investigators distributing results of their research effort.

The first CASP Annual Technical Symposium was held with over 70 people in attendance. Positive feedback on the one-day of presentations was received and the symposium will probably be expanded to two days next year.

During the past year we have continued to solidify and add projects in five focus areas:

- Advanced Chemical Propulsion
- Electric Propulsion
- AI/Expert Systems
- Microgravity Fluid Management
- Propulsion Materials Processing

It has become evident that CASP projects fall into two general categories; 1) technology development usually supported by the larger Aerospace companies and the NASA field centers, and 2) product development and flight experiments usually found in teaming arrangements with smaller companies. Figures 1 and 2 depict the relationships of our current projects with their anticipated role in CASP's development.

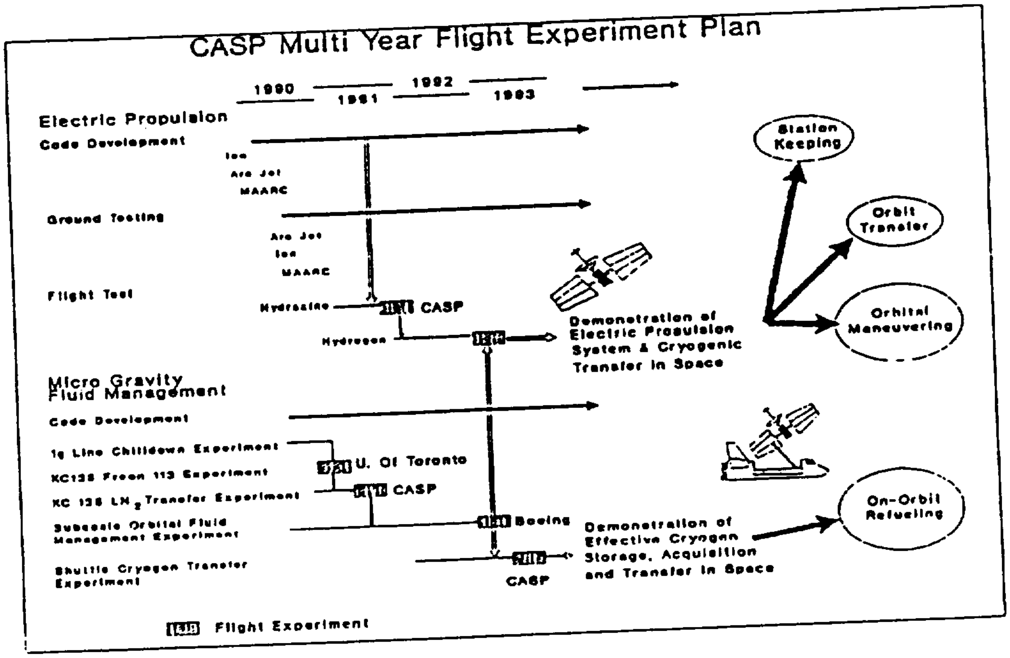


Figure 1.

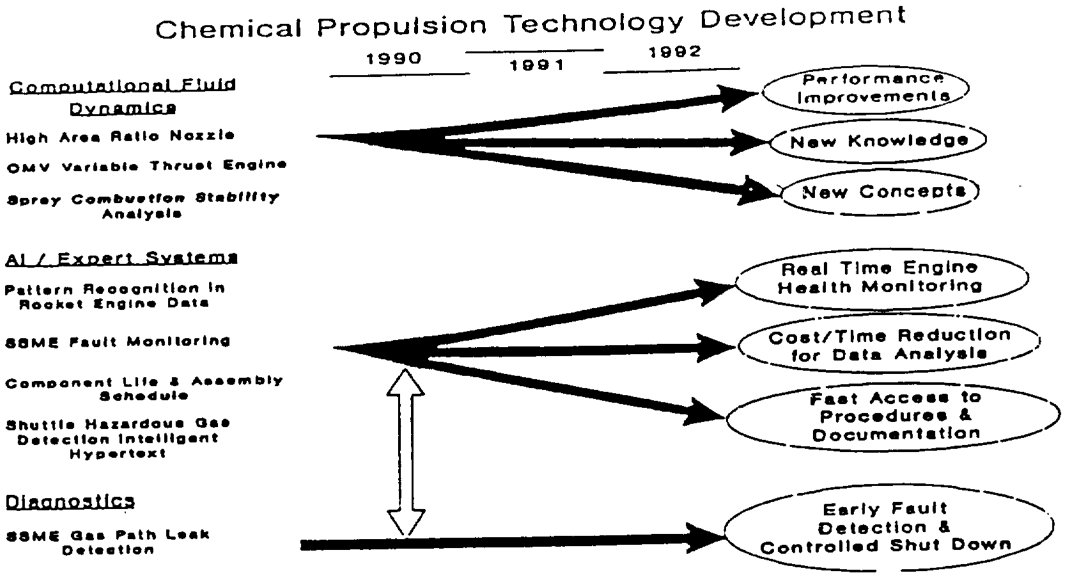


Figure 2.

During the first two years CASP has developed the capability to interface effectively with NASA and industrial partners. These communications have matured into a concentrated project workload in the microgravity and chemical propulsion areas.

In order to meet CASP long term objectives, emphasis will be placed on the following:

- Develop Electric Propulsion
- Develop and Process Flight Experiments
- Form Partnerships with Small Business

To adequately respond to the increasing interest by both industry and government in electric propulsion, CASP, CAR, and UTSI elected to jointly fund the refurbishment of a 10 x 28 ft. space chamber. Upon completion, the chamber will include provisions for state-of-the-art laser diagnostics to support electric thruster testing and subsequent space flight experiments.

During the year Dr. George Garrison assumed the position of Director and Dr. F. Speer became CASP's Associate Director. CASP continued to assemble a uniquely qualified team of investigators and two additional faculty were added as Principal Investigators. At the close of the period, requisitions for three propulsion specialists were being advertised by UTSI.

Personnel census at CASP's interim facility in the UTSI Research Park increased this year. Twenty seven investigators are now involved in fourteen active projects. CASP's microcomputer center was completed with the addition of one Compaq and four CompuAdd 386 microcomputers and a SUN 386i workstation.

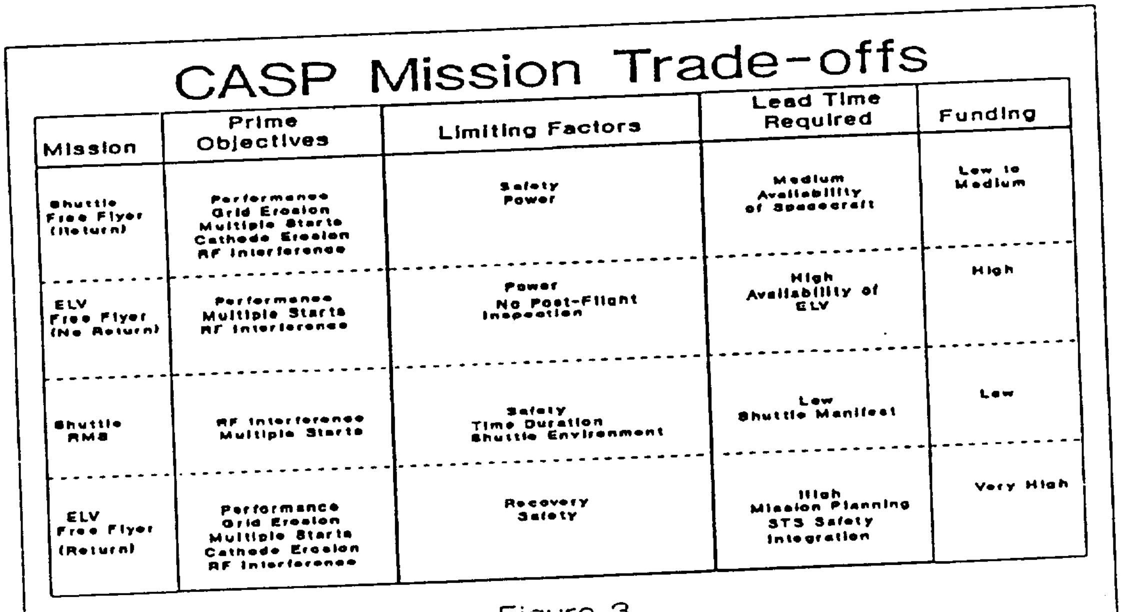
To bring together the growing number of investigators and assistants working on CASP projects, a 10,000 ft.^{2} office and laboratory facility is being funded by the State of Tennessee and the Appalachian Regional Commission. Site selection, design and floor plan layout of the new building have been completed, and bids for construction are being reviewed. Ground breaking is scheduled for early January, 1990.

NASA code C organized and hosted quarterly meetings of the CCDS directors, NASA field center representatives, and Office of Commercial Program (OCP) support contractors. There is strong support for the CCDS program at NASA Headquarters and, in spite of budget reductions across all of NASA, grants for the commercialization centers remained unchanged.

It has become very evident that the CCDS's are expected to fly experiments and that OCP will do everything possible to facilitate flights and shorten, where possible, the time from concept to flight. CASP continues to evaluate the skill and experience required to meet future flight experiment needs and identify the specific role the Center will take in the development of a follow-on to the current CCDS sounding rocket program. The matrix, show in figure 3, displays the tradeoffs and implied management challenges in developing future flight vehicles.

CASP ended the current year slightly under the financial goals set in the original 5-year plan, but finished the year with income from grants and cost sharing being slightly larger than expenses. Growth in project funding was slower than expected and in-kind contributions closely tracked the plan. Future CASP growth will demand greater income. The total was far below that necessary to support a flight program.

Along with the continuing challenge of daily operations, CASP's management emphasis now shifts to the next major set of design, fabrication, testing and flight of a commercially viable propulsion experiment. Of CASP's focus areas microgravity fluids management and electric propulsion related systems are materializing as prime candidates for space flight experiments.



CASP Mission Trade-offs

Mission	Prime Objectives	Limiting Factors	Lead Time Required	Funding
Shuttle Free Flyer (Return)	Performance Grid Erosion Multiple Starts Cathode Erosion RF Interference	Safety Power	Medium Availability of Spacecraft	Low to Medium
ELV Free Flyer (No Return)	Performance Multiple Starts RF Interference	Power No Post-Flight Inspection	High Availability of ELV	High
Shuttle RMS	RF Interference Multiple Starts	Safety Time Duration Shuttle Environment	Low Shuttle Manifest	Low
ELV Free Flyer (Return)	Performance Grid Erosion Multiple Starts Cathode Erosion RF Interference	Recovery Safety	High Mission Planning STS Safety Integration	Very High

Figure 3

INTRODUCTION & OVERVIEW

The Center for Advanced Space Propulsion is part of The University of Tennessee-Calspan Center for Aerospace Research (CAR). Formed in 1985 to take advantage of the extensive research faculty and staff of The University of Tennessee and Calspan Corporation, CAR is a not-for-profit organization established to conduct aeronautical and aerospace research and development. The broad technical knowledge and assistance provided by CAR make it a strong parent organization for the Center for Advanced Space Propulsion.

On October 31, 1989 CASP marked it's second successful year as a NASA Center for the Commercial Development of Space (CCDS). In carrying out it's mission to strengthen U.S. competitiveness in space propulsion technology, CASP added new industrial partners to the Management Advisory Council and new projects to it's list of active pursuits. Based on investigators' qualifications in propulsion system development, and matching industry strong intent the Center focused its efforts in the following technical areas:

- Advanced Chemical Propulsion
- Electric Propulsion
- AI/Expert Systems
- Fluids Management in Microgravity
- Propulsion Materials Processing

Throughout the year, the Center continued to integrate academic excellence with commercial space requirements and increased the number of active projects to 15. These are supported by 10 Principal Investigators and 19 Graduate Research Assistants and six staff. Each project has a defined commercial application and either enables or enhances the related technologies with respect to space commercialization.

CASP marked it's second anniversary in the Center's interim facility at the UTSI Research Park. Design and funding of CASP's new 10,000 ft.^{2} office building was completed during the year. Construction is scheduled to begin in late November, 1989.

Valuable inputs from both CASP's Management Advisory Council and NASA's Office of Commercial Programs formed a framework for CASP's performance goals. Figures 1 and 2 of the Executive Summary depict the relationship of the various projects accomplished by CASP during the year, and relates each to it's proposed objective. Critical for reaching these goals are the ground based facilities which are used by the investigative teams in performing the project work. Supercomputer support continued to be provided by MSFC and was the empowering tool for our advances in Computational Fluid Dynamics. Installation at UTSI of an Alliant computer to compliment the existing VAX allowed some programs which had been run on the NASA Marshall CRAY to be moved to UTSI. To supplement these computational systems, CASP completed outfitting it's interim computer laboratory with three additional 386 PC's and a Sun 386i. With these existing computers, CASP offers a full spectrum of computational hardware for it's researchers.

Earlier emphasis on marketing paid dividends this year with addition of two projects to the active list and three more industrial organizations which have expressed an interest in joining CASP in future commercial space ventures. Figure 3 tabulates the projects which have a potential for being funded during the next year. Those most closely matching the commercial objectives of CASP will be activated.

CASP's second year was marked by increased financial stability, but there is a continuing problem with the extended time to consummate project agreements. Project costs continued to vary somewhat from their budgets, however total CASP spending ended somewhat less than earlier projections. Cash flow problems caused by delayed project funding continued to restrict programs in some areas. To accommodate the shortfall, management action was taken to contain costs. CASP ended the year with expenditures within one percent of the funds available.

NASA gave impetus for increasing the CCDS presence in space by providing a number of GAS can and mid deck experiment opportunities. As briefed by NASA, flight openings are becoming available on Shuttle, Spacehab, and the Space Station. Such "reserved" space allows CASP's industrial partners to take immediate advantage of space experimentation without encountering a lengthy queue. In preparation for possible space flight, CASP selected experiments in Microgravity Fluids Management and Electric Propulsion for flight in the '92-'93 time frame.

A precursor micro-gravity fluids management experiment, planned to be flown on a KC-135 during the summer of 1990, was being formulated with MSFC as the year ended. Knowledge gained from this experiment will flow directly into Boeing's SOFTE space flight experiment design.

As the year ended, CASP's goals remained unchanged and designs of several space flight experiments were in progress. The industrial commitments and project prospects appear to be solid and should result in growth in each of the CASP focus areas.

In CASP's annual review with NASA Headquarters, Office of Commercial Programs, Jim Rose expressed approval of the path chosen and the progress made by the Center. He noted the need for elements of infrastructure which must be developed to service the continuing needs of all users of commercial space. CASP has the objective of being a significant player in development of one of the critical infrastructure elements - transportation.

ADMINISTRATIVE / BUSINESS STATUS

Funding / Business Goals

CASP is one of two operating elements of the Center for Aerospace Research, a not for profit organization founded by The University of Tennessee and Arvin/Calspan Corporation. CASP's primary source of funding is an Annual Grant from NASA code C with industrial contracts and agreements providing additional sources of revenue.

Figure II.1 describes the major elements of CASP funding and reflects the business goals set by CASP during the past two years. One significant philosophical change emerged during the year: NASA has agreed that a meaningful flight experiment cannot generally be formulated, tested, and flown within the five year limit imposed by the initial grant. As a result. NASA has accepted the need for Center funding beyond the five year limit. CASP's flight plans mirror the requirement for extended funding of core needs. It is our belief that a usable microgravity or electric propulsion experiment cannot be accomplished before FY 93. To reach the point of a flight experiment will require a great deal of nurturing (cost sharing) with both a longer period of baseline and augmented funding required to be successful.

In FY '89, the Center was not able to meet it's overall goal for funding. While in-kind contributions met the planned level, cash from contracts and grants was somewhat lower. A carry-over of industrial contributions from the prior year allowed project work to proceed at a level which was slightly above the year's income. In general, there is still a shortfall in funds caused by delays in the receipt of funds from NASA Centers and industrial partners. The effect of the delays can be seen in figure II.2 where project workload was restricted so it remained below the level of funds. Note that receipt of the CASP's NASA Headquarters Grant typically funds all work early in the year while the Center finalizes terms with our industrial partners.

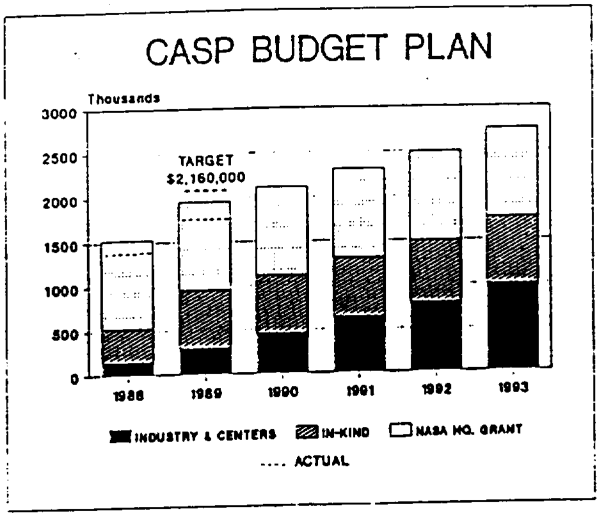


Figure II.1

Personnel

On February 1, 1989, Dr. George Garrison assumed the position of Director with Dr. F. Speer named CASP's Associate Director. CASP continued to assemble a uniquely qualified team of investigators and additional faculty were added as Principal Investigators for the Ion Propulsion and Inconel Welding projects. At the close of the period, CASP had filled one of the three open requisitions for propulsion specialists. Mr. Verlin Friedly accepted a position as a Research Engineer and will begin working primarily in electric propulsion. He will continue the work started by Dr. Y.K. Chen who left late in the year to take a position with Ames Research Center.

Initiation of the Helical Coil project with Sundstrand Inc. marked the first opportunity to tap expertise at the UT Knoxville campus. Dr. Edward Keshock, of the Engineering Department and two assistants are investigating two phase flow in helical coils for this project.

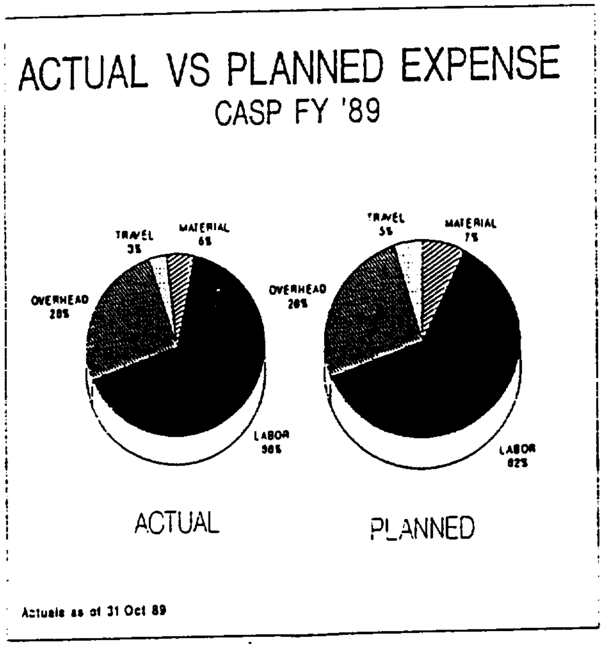


Figure II.2

Administrative support remained constant during the year and workload peaks were accomplished by using temporary hires.

Facilities

An expanded project workload and initiation of two projects requiring laboratory space to conduct ground testing pushed CASP's interim facility to capacity. Design of the new CASP Building was completed and offers the most immediate prospect for relief of the crowded conditions being experienced in the interim facility.

Figure II.3 shows the floor plan of the proposed building. A Fiber-optics link will connect the building with the UTSI Computer Center which is the port through which researchers access the CRAY at MSFC, and VAX and Alliant at UTSI. Included within the 10,000+ ft.^{2} building is a small experiment build-up area and laboratory. These areas have gained additional importance as CASP researchers have begun flight experiment fabrication.

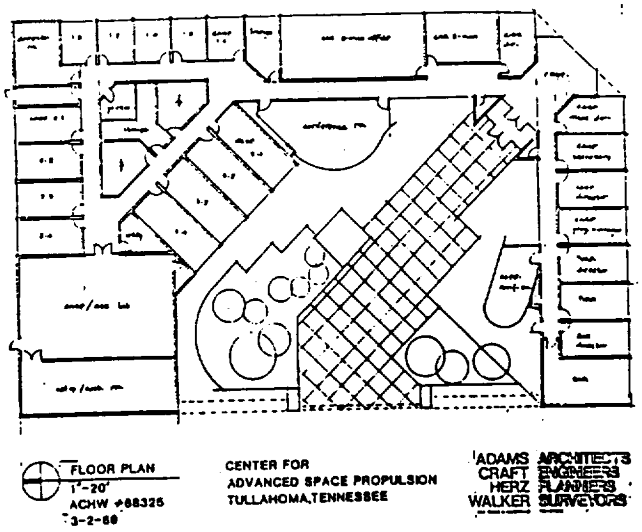


Figure II.3

Renovation of the 10' x 28' Space Chamber at UTSI was initiated during the year. This chamber will support electric propulsion system and component testing and will be initially configured to perform a detailed experimental investigation of ion thrust grid erosion. Testing of arc jets and resisto jets can be accommodated, but depending on the flow rate, additional pumping capacity may be required. The system is complete with a LH_2 pump and liner system, roughing pumps, and control system. Test hardware includes an X-Y positioning system and laser diagnostics equipment. Refurbishment of the chamber is scheduled for completion before June, 1990. Funding for activation is provided by CAR, UTSI, and CASP non-grant resources.

Addition of a Sun Microsystems workstation and three 386 microprocessors completed hardware acquisitions for the interim CASP computer center. Two PC's plus two Visual terminals are connected to the UTSI Computer Center switch through Codex 9600 baud modems. From the switch, investigators link to NASA & DOD Supercomputers and VAX and Alliant computers within the University system.

Seminar & Publications

In October 1989, CASP hosted it's first annual Technical Seminar. Lasting one day, the seminar allowed each investigator to present the salient aspects of his efforts to date and field questions from the industrial and Government attendees. Feedback from the attendees was favorable and their comments will be included in our planning for the next annual seminar.

Publications which have been either presented or published by CASP investigators are listed below.

FY 89 CASP PUBLICATIONS

AUTHOR	TITLE	FORMAT
LITCHFORD, RON	"FORMULATION AND EVALUATION OF A HIGH-PRESSURE, DISCRETE DROPLET GASIFICATION MODEL FOR BIPROPELLANT LIQUID ROCKET COMBUSTOR MODELING APPLICATIONS", Master's Thesis, The University of Tennessee, November 1989	MASTER'S THESIS
M. H. McCAY and T. D. McCAY, et. al.	"PULSE LASER WELDING OF INCO 718", Presented October 15-20, 1989, ICALEO 1989, Orlando, Florida	PRESENTATION
M. H. McCAY and T. D. McCAY, et. al.	"EFFECT OF PULSE DUTY CYCLE ON INCONEL 718 LASER WELDS", SAE TECHNICAL PAPER SERIES #892299, The Use of Lasers in Manufacturing, September 25-28, 1989	PRESENTATION PUBLICATION
G. GARRISON	"STATUS OF THE CENTER FOR ADVANCED SPACE PROPULSION" International Conference on Engineering Management, September 1989	PRESENTATION
W.E. DIETZ, E.L. KIECH and MOONIS ALI	"CLASSIFICATION OF DATA PATTERNS USING AN AUTOASSOCIATIVE NEURAL NETWORK TOPOLOGY" The Second International Conference on Industrial & Engineering Applications of Artificial Intelligence & Expert Systems, Tullahoma, TN, June 6-9, 1989	PRESENTATION PROCEEDING
W.E. DIETZ, H.J. FERBER and MOONIS ALI	"SPACE SHUTTLE MAIN-ENGINE COMPONENT ASSEMBLY, ASSIGNMENT AND SCHEDULING EXPERT SYSTEM", The Second International Conference on Industrial & Engineering Applications of Artificial Intelligence & Expert Systems, Tullahoma, TN, June 6-9, 1989	PRESENTATION PROCEEDING
U.K. GUPTA and MOONIS ALI	"LEADER - AN INTEGRATED AUTOMATIC LEARNING AND FAULT DIAGNOSTIC EXPERT SYSTEM FOR SPACE SHUTTLE MAIN ENGINE (SSME)", The Second International Conference on Industrial & Engineering Applications of Artificial Intelligence & Expert Systems, Tullahoma, TN, June 6-9, 1989	PRESENTATION PROCEEDING
W.E. DIETZ, E.L. KIECH and MOONIS ALI	"NEURAL NETWORK MODELS APPLIED TO REAL-TIME FAULT DIAGNOSIS" Journal of Neural Network Computing, Vol. 1, No. 1, 1989.	PUBLICATION
C. F. LO	"INTELLIGENT HYPERTEXT SYSTEMS FOR AEROSPACE ENGINEERING APPLICATIONS" TABES '89 Technical and Business Exhibition and Symposium, May 16-17, 1989	PRESENTATION
J. ORR	"PARC CODE APPLICATION TO HIGH AREA RATIO NOZZLES" AIAA Student Conference, April 20, 1989.	PRESENTATION MASTER'S THESIS

Professional Society:
AIAA - American Institute of Aeronautics & Astronautics
ASME - American Society of Mechanical Engineers
ASEM - American Society of Engineering Managers

STATUS REPORTS FOR ACTIVE PROJECTS

SECTIONS

Advanced Chemical Propulsion

A.I. Propulsion Applications

Microgravity Fluid Management

Electric Propulsion

Laser Material Processing

ADVANCED CHEMICAL PROPULSION

- OMV Variable Thrust Engine Analysis
- Spray Combustion Stability
- Advanced High Area Ratio Nozzles

Project Title: OMV/VTE Variable Thrust Engine Analysis*

Principal Investigators: L. Larosillere, R. Litchford and S. M. Jeng

Project Report

SUMMARY

The objective of the present work is to develop a predictive CFD based analytical tool for the variable thrust engine (VTE) in the orbital maneuvering vehicle (OMV). This objective is being accomplished within the framework of the Los Alamos KIVA computer code for chemically reactive flows with sprays. For the OMV application, the main structure of KIVA is to be retained while reformulating many of the phenomenological submodels, enhancing some of the numerics, and adding more features.

The analytical model consists of the general conservation equations for two-phase reactive flows and of submodels for turbulence, chemical reactions, and bipropellant sprays. Tailoring this model to the OMV engine brings about the added complexities of combustion and flow processes that occur in a liquid hypergolic propellant rocket chamber.

This report exposes the foundation upon which the analytical tool is being constructed and developed. Results from a cursory computational exercise involving the simulation of the flow and combustion processes in a hypothetical N_2H_4/N_2O_4 rocket engine thrust chamber is presented and discussed.

INTRODUCTION

TRW is currently developing a hypergolic bipropellant variable thrust engine to be qualified specifically for OMV. The available analytical prediction methodology is incapable of providing physical and phenomenological insight about the complex spray combustion and flow processes occurring in the engine thrust chamber. Thus, an adequate CFD model would significantly contribute towards the identification of existing design flows and the evaluation of alternative new designs.

A complete description of the combustion and flow processes occurring in a liquid hypergolic rocket chamber does not exist. It's known that a wide range of interrelated and coupled physical phenomena manifest themselves in the combustor. For example, both homogeneous and heterogeneous chemical reactions may occur; the spray droplet trajectory and life history is determined by the chamber gas environment: all gas fluid in the chamber is either from the liquid evaporation process or the reaction process in the mixed liquids. Many physical processes occurring in the VTE are similar to other complex flow phenomena and have been modelled and applied to the gas turbine combustor, diesel engine, cryogenic liquid propellant rocket; however, the advanced CFD model for the hypergolic propellant rocket has never been established.

This report is a current status update of an ongoing effort to develop a predictive tool for the combustion chamber of the VTE. A framework for constructing the model is presented in particular, results from a cursory computational exercise are presented in support of the proposed framework.

DEVELOPMENT OF SPRAY COMBUSTION AND FLOW MODEL

The case is considered in which liquid sprays of fuel and oxidizer are injected into the combustion chamber of a rocket engine. Upon entering the chamber volume, the two continuous liquid phases will encounter forces leading to the formation of fuel and oxidizer drops with a spectrum of properties. Commencing at the point of drop formation, the aim is to track the drop dynamics, the gas dynamics, and the ensuing interactions. Hence, the realm of two-phase chemically reacting flows is entered. A separate-phase flow model is actually used: gas phase is described using Eulerian coordinates and dispersed-phase liquid droplets are described via Lagrangian coordinates.

Government/Industry Partners - MSFC/TRW

MODEL DESCRIPTION

Gas-Phase

The unsteady conservation equations for a turbulent mixture of chemically reactive ideal gases are formulated. Specific equations will not be listed here because they are well known and are available elsewhere. Suffice to say that they are a system of coupled partial differential equations having the following structure:

$$\frac{\partial}{\partial t}(\rho\phi) + \text{ div } (\rho\vec{u}\phi) = \text{ div } (\Gamma_\phi \text{ grad } \phi) + S_\phi \quad (1)$$

where the dependent variable Φ may represent bulk velocity, internal energy, gas composition, turbulence kinetic energy, and rate of decay of turbulence kinetic energy. The variable Γ_ϕ is an effective diffusion coefficient for $_\phi$, and S_ϕ is an appropriate source term representing effects such as droplet/gas interactions, turbulence production, and chemical reactions. The primary limitations of this formulation are the modeling of the source terms and the numerical solution technique.

Two models are available to represent the effects of turbulence. A standard version of the k-E turbulence model, modified to include volumetric expansion effects, and spray/turbulence interactions has given results of sufficient accuracy for engineering purposes in a variety of applications. The subgrid scale turbulence model[1] is also available. This model reduces to the k-E model near walls where all turbulence length scales are too small to be resolved by the computational mesh.

The confinement of the gas is taken into account by the boundary conditions. Velocity boundary conditions on rigid walls can be free slip, no slip, or turbulent law-of-the-wall. For turbulent law-of-the-wall conditions, the tangential velocity components are determined by matching to appropriate wall function. Temperature boundary conditions on rigid walls are introduced by specifying either the wall temperature or the wall heat flux. The boundary conditions for the turbulence kinetic energy and its decay rate are zero gradient of the kinetic energy at the walls and equilibrium between generation and dissipation of turbulence. Boundary conditions for the individual species are zero gradients of the mass fractions at the walls, corresponding to chemically-neutral surface and no fluxes through the wall.

Dispersed-Liquid Phase

Each spray (fuel or oxidizer) is followed by employing a discrete droplet model which represents the spray by a statistically meaningful number of individual droplets. A computational droplet represents a parcel of droplets all having the same characteristics such as position and size. This formulation is taken from Sukorvicz[2] who stresses a statistical rather than a deterministic approach to the modeling of sprays.

A droplet's trajectory is tracked using the ballistic equations as derived via Newton's second law. Hence, a Lagrangian technique is implied. The turbulence effects on droplet motion are simulated statistically by superimposing upon the gas field turbulent eddies, each having a length, lifetime, and velocity fluctuations. This model is based on a Gaussian distribution of the local gas velocity perturbations.

Boundary conditions are needed in order to confine the sprays within the chamber walls. A question which must be answered is what happens when a droplet impinges on a hot rigid wall? Some sketchy results obtained by Watchers and Westerling[3] with water show that the observed phenomena are strongly linked to the Weber number. The technique employed in the current model is to set the droplet velocity equal to the wall velocity and not to allow any heat transfer between the droplet and wall. This is a provisional model for spray/wall interactions and needs to be critically reevaluated. Conditions at the location of drop formation need to be specified also. This is presently achieved by specifying a droplet mass flow rate and a distribution of droplet sizes, velocities, temperatures, and other relevant parameters.

PHYSICAL SUBMODELS AND MATERIAL PROPERTIES

Chemistry

Both equilibrium and kinetic chemical effects need to be accounted for. Unfortunately, chemical kinetic mechanisms and rates are not readily available for the proposed bipropellant combustion. The following nine step reaction is being enlisted for the N_2H_4/N_2O_4 system:

$$2N_2H_4 \longrightarrow 2NH_3 + H_2 + N_2 \quad (1)$$

$$N_2O_4 \longrightarrow 2NO_2 \quad (2)$$

$$2NO_2 \longrightarrow 2NO + O_2 \quad (3)$$

$$H_2 + NO_2 \longrightarrow H_2O + NO \quad (4)$$

$$2H_2 + O_2 \longrightarrow 2H_2O \qquad (5)$$

$$2H_2 + O_2 + 2NO \longrightarrow 2H_2O + 2NO \qquad (6)$$

$$2NH_3 + 3NO_2 \longrightarrow 3H_2O + 3NO + N_2 \qquad (7)$$

$$4NH_3 + 3O_2 \longrightarrow 6H_2O + 2N_2 \qquad (8)$$

$$2NH_3 + H_2 + 2O_2 \longrightarrow 4H_2O + N_2 \qquad (9)$$

The rate constants are assumed to be of a generalize Arrhenius form. Reaction rates are determined from the law of mass action or any other experimentally derived rule. Once the chemical scheme has been established, the chemical source term in the species continuity equation and the chemical heat release term in the energy equation are obtained.

Droplet Evaporation

The hypergolic propellants of primary interest for liquid rocket applications are hydrazine and its derivatives - MMH (monomethyl hydrazine) and UDMH (unsymmetrical dimethyl hydrazine). For high ambient temperatures, as occurs in rocket chambers, conditions are conducive for formation of an exothermic decomposition flame enveloping the fuel droplet near the surface as depicted in Figure 1. This process of pure monopropellant combustion can substantially enhance the gasification rate over the nonreactive case. If, in addition, an oxidation propellant is introduced, as is the case for hypergolic bipropellant systems, the decomposition envelope flame may be further surrounded by an oxidation flame where the decomposition products react with the oxidizing agent. This process of hybrid droplet combustion is depicted in Figure 2. The additional presence of the oxidation flame would further enhance the gasification rate over the pure decomposition case.

Since an MMH/N_2O_4 hypergolic bipropellant system has been designated for the OMV-VTE, the preceding phenomena will occur in the combustor and should be properly modeled as part of the analysis effort. Depending upon the local flow conditions the fuel will gasify in one of three possible modes: (1) pure evaporation, (2) pure monopropellant droplet combustion, or (3) hybrid monopropellant droplet combustion (dual flame model). Since the KIVA code is equipped with a pure evaporation model, no modifications are required in this case; however, theoretical models must be obtained and coded for the single and dual flame cases.

Much work has been done in the past on monopropellant combustion at steady-state condition, but little effort has been devoted to the development of transient engineering submodels which include the heat-up period and are applicable for time dependent CFD applications. An exception is some work by Faeth[4] on pure monopropellant droplet combustion at high ambient temperatures. He considered quasi-steady, spherically-symmetric, gas-phase transport coupled to a time dependent constant property liquid phase with one step global decomposition reaction rate kinetics of the Arrhenius form. The theory also allowed for corrections in transfer rates due to forced convection from empirical correlations for the film thickness. The gas-phase solution results in a system of nonlinear algebraic equations easily solved with an iterative algorithm, and the liquid-phase is governed by two ordinary differential equations.

Faeth performed calculations for EN (ethyl nitrate) and PGDN (propylene glycol dinitrate) and fitted the results to experimental data by selecting a reaction order and activation energy and adjusting the Arrhenius constant. The procedure correlated experimental data reasonably well and appears justified as a good, efficient engineering submodel for pure monopropellant combustion.

Because Faeth's theory is good and readily available it was selected as the basis submodel for implementation in the KIVA code. Before incorporation in the code, however, the model was studied for applicability to hydrazine type fuels. Calculations were made for hydrazine, MMH, and UDMH monopropellant combustion in an atmosphere of inerts and compared with the experimental results of Allison and Faeth[5]. The computations were made such that the process reached steady-state and the droplet diameter attained the size corresponding to experiment. The kinetic parameters were adjusted to fit the data as prescribed by Faeth with good correlation obtained. The comparison of theory with experiment is depicted in Figures 3, 4, and 5 for hydrazine, and MMH, and UDMH, respectively. With the activation energies chosen for hydrazine and MMH, the burning rate constant falls off sharply at decreasing temperatures. If a smaller activation energy is chosen, the theory may be extended to lower temperatures. Also note that a reaction order of unity was specified in the current calculations while a second order reaction may be necessary to correlate data at higher pressures as stated by Faeth. These issues will be examined and addressed in future study. Calculations for pure evaporation are also shown as dashed lines on the figures to indicate enhancement of the gasification rate.

A transient model for hybrid droplet combustion is not available at this time although Allison and Faeth[5] have formulated and correlated a steady-state hybrid combustion model for hydrazine type fuels. At this time, two alternatives are being considered. They are to modify Allison and Faeth's theory to a quasi-steady approach or make ad hoc modifications to the pure monopropellant combustion model. Future study will determine the approached deemed appropriate.

Because it is unwise to expend effort modifying the KIVA code until complete droplet gasification models have been established and experimentally validated, the original KIVA pure evaporation model will be retained during modifications to other parts of the code structure. Artificial amplification of the gasification rate may be used to reflect actual combustion processes for code development purposes until model development is completed.

Drop/Drop Interactions

Dense sprays exist in a region close to the inspector. Hence, considerations of drop collisions are warranted. There are two types of drop/drop interactions: collisions between drops of the same kind and collisions between two different kinds of drops (i.e. fuel and oxidizer). A literature search and review is being conducted in order to formulate an appropriate model.

Thermochemical Data Base

Thermophysical properties for hydrazine and its derivatives have been collected and placed in a block data subprogram. Properties for N_2O_4 are also available.

CALCULATION TECHNIQUE

Grid Generation

A rapid grid operation technique has been constructed specifically for the VTE geometry. Further developments are proceeding as the program evolves. The grid consists of arbitrary hexahedrons and is able to be arbitrarily packed.

Numerics

Presently, the numerics are taken exactly from KIVA[1]. KIVA employs a fully 3-D finite volume approach over a staggered mesh. Time marching is performed by use of operator splitting and the solution is advanced in a series of phases. Variable implicitness is employed to circumvent the current stability requirements. Enhancements to the inflow/outflow boundary condition treatment are contemplated.

RESULTS AND DISCUSSION

A cursory calculation is presented for an assessment of the proposed modeling framework. It is performed using a 15 x 40 grid (Fig. 6), covering a chamber radius of 1.83 cm and a length of 12 cm. A small amount of mesh packing is done near the walls. At time t=0, pressure, temperature, and density are set uniform in the chamber. The gas composition at t=0 is virtually CO_2 with initial temperature of 1600 °k and pressure of 4.5 atm. A constant back pressure equal to the initial pressure and adiabatic wall conditions are applied for all times.

Fuel (N_2H_4) and oxidizer (N_2O_4) droplets are injected approximately 2 cm downstream of the back wall and 0.3 cm from the center line. The injection parameters are:

Sauter mean diameters - SMD = 50 μ m

Injection velocity - $V_{\text{inj}} = 15.5$ m/s for N_2H_4

$V_{\text{inj}} = 13.3$ m/s for N_2O_4

Injector orientation - tilt = 30°, cone angle = 25°

Total injected propellant flow rate - 197. g/s

Fuel oxidizer ratio - F/O = 0.83

A λ-squared distribution about the SMD is used for the sizes of injected droplets. Hence, the drop sizes are statistically sampled at the beginning of injection.

Figure 7 presents the velocity field inside the chamber at two different times. It can be seen that the velocity field is mostly one directional away from the injection point. Figure 8 is a temporal evolution of the N_2H_4 droplet parcel ballistics within the chamber. A high percentage of N_2H_4 droplets can be seen impinging on the chamber wall and adhesing to it. Figure 9 shows similar trends for the N_2O_4 droplets. An immediate deduction from these results is that the droplet vaporization rates are grossly under-determined using the pure droplet evaporation (i.e. non-combusting) model. Figures 10 and 11 show the temporal evolution of the droplet parcel velocities for N_2H_4 and N_2O_4, respectively.

The temporal evolution of the temperature field is presented in Figure 12. A uniform region of high temperature (H) can be seen propogating downstream from the injection point where the temperature is relatively low (L). The maximum temperature within the chamber is well below the adrabatic flame temperature due to incomplete droplet gasification.

Figure 13 shows the fuel oxidizer ratio distribution within the chamber for different times. Again, this points to the low droplet gasification rate. Mass fraction contours of nitrogen, which is the primary combustion product in this system, are presented in Figure 14 for different times. The N_2 concentration is seen to be maximum near the chamber exit tending towards an equilibrium value as time evolves.

FUTURE DEVELOPMENTS

The plan is to fully develop all of the aforementioned submodels and efficiently incorporate them into the overall computational scheme. Some basic research will be undertaken specifically in the areas of drop/drop interactions and hypergolic bipropellant gasification modes. It is believed that the performance of this hypergolic system is strongly controlled more so by the gasification process than the gas phase chemical kinetic rates.

REFERENCES

1. A. A. Amsden, P. J. O'Rouke, T. D. Butller, "KIVA-O: A Computer Program for Chemically Reactive Flows with Sprays," LA-11560-MS.
2. J. K. Dukowicz, J. Comput. Phys. 35, 229 (1980).
3. L. H. J. Watchers, N. A. J. Westerling, Chemical Engineering Science, Vol. 21, pp. 1047.
4. Faeth,G. M., "Prediction of Pure Monopropellant Droplet Life Histories," AIAA J., Vol. 8, No. 7, pp. 1308-1314, 1970.
5. Allison, C. B. and Faeth, G. M., "Hybrid and Decomposition Combustion of the Hydrazine Fuels," NASA CR-72977, 1972.

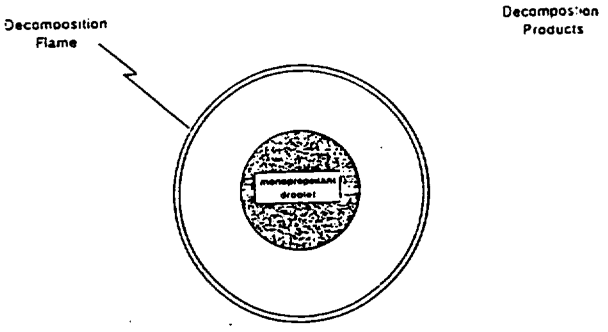


Figure 1. Schematic of pure monopropellant droplet combustion

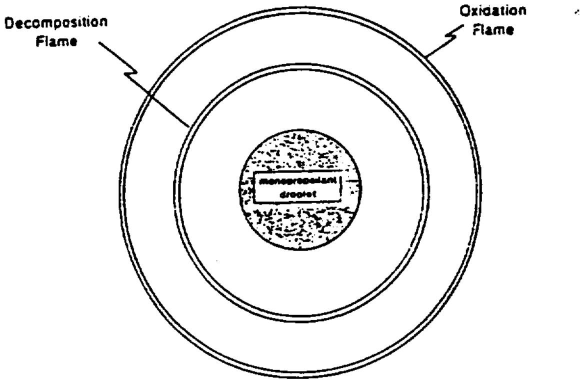


Figure 2. Schematic of hybrid monopropellant droplet combustion

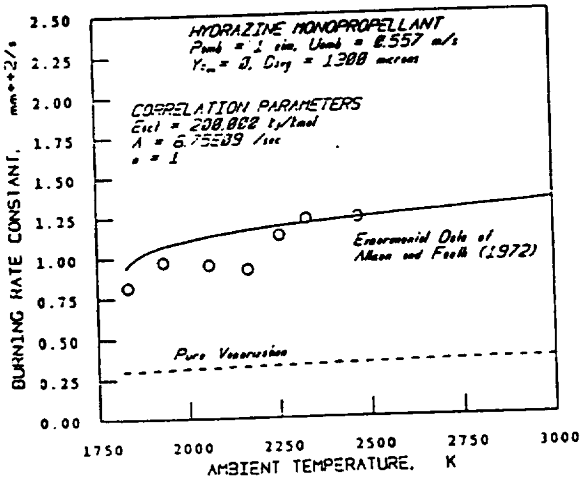


Figure 3. Theoretical and experimental burning rate constants with various ambient temperatures for hydrazine monopropellant droplets

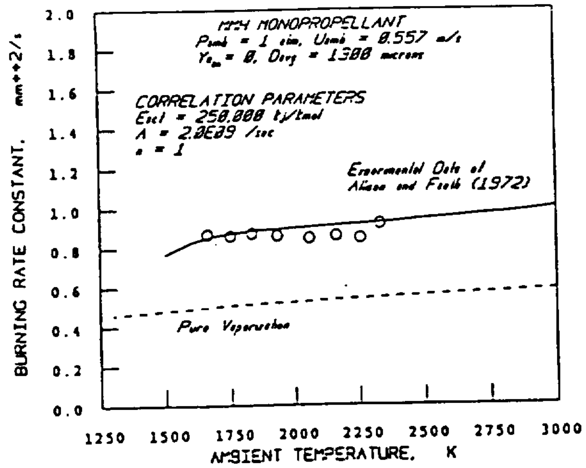


Figure 4. Theoretical and experimental burning rate constants with various ambient temperatures for MMH monopropellant droplets

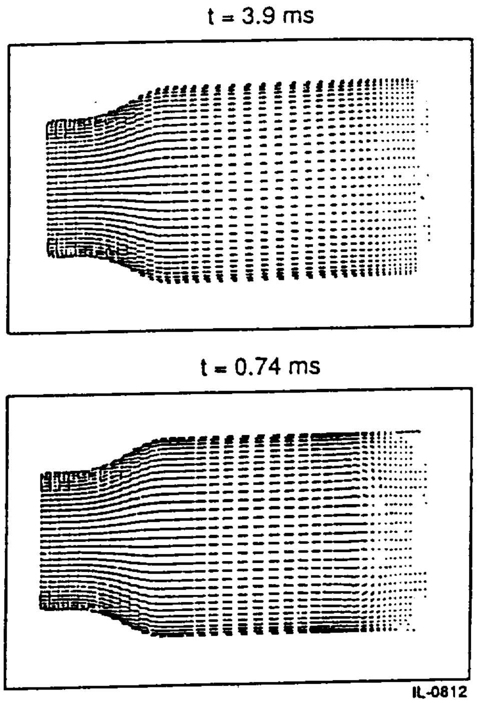


Figure 7. Gas velocity fields

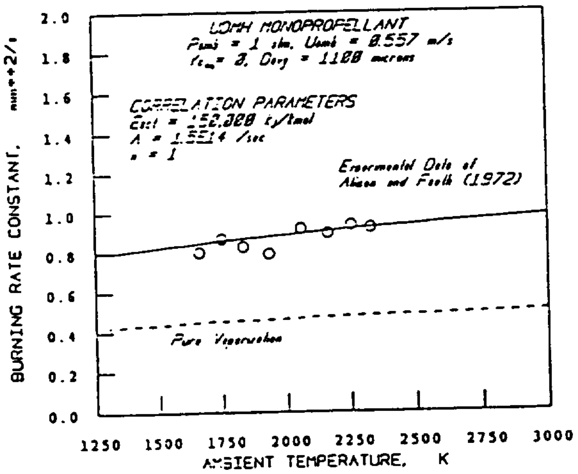


Figure 5. Theoretical and experimental burning rate constants with various ambient temperatures for UDMH monopropellant droplets

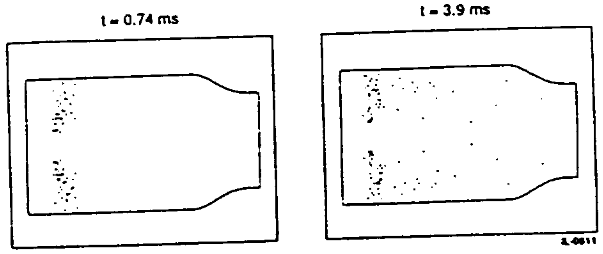


Figure 8. N_2H_4 droplet trajectories

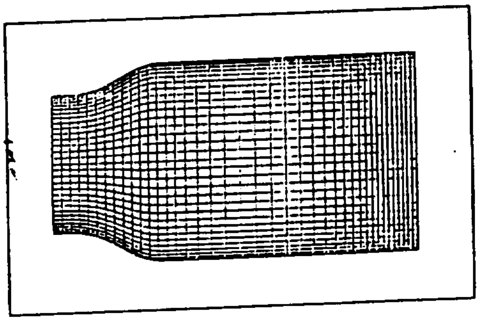


Figure 6. The computing mesh (flow is from right to left)

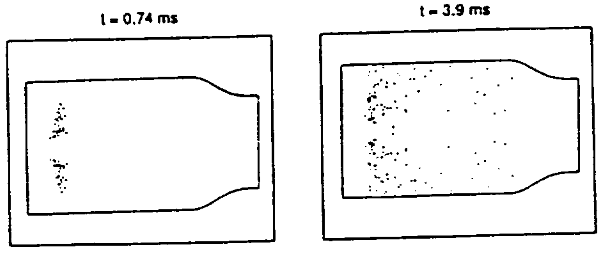


Figure 9. N_2O_4 droplet trajectories

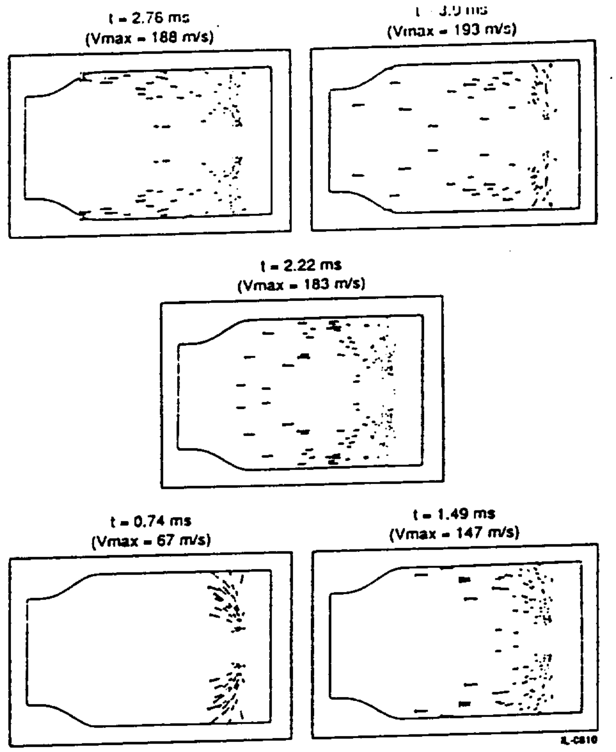


Figure 10. Temporal evolution of N_2H_4 droplet parcel velocity

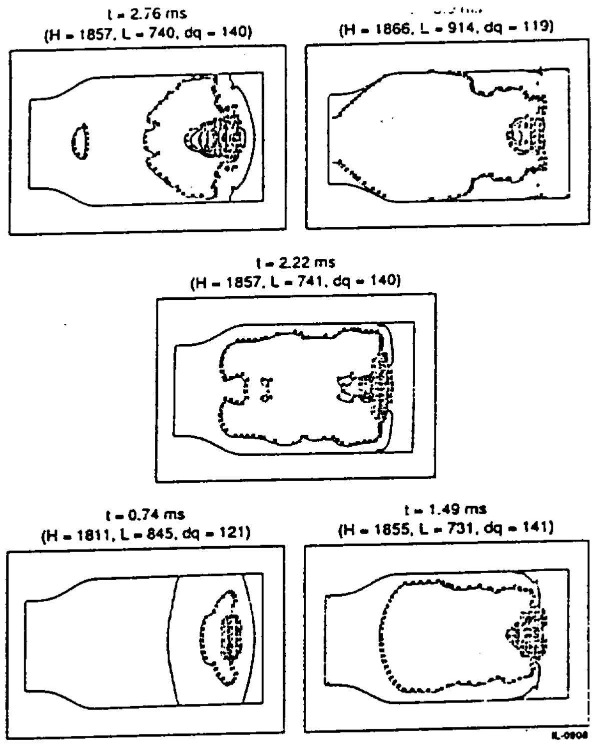


Figure 12. Evolution of the temperature field

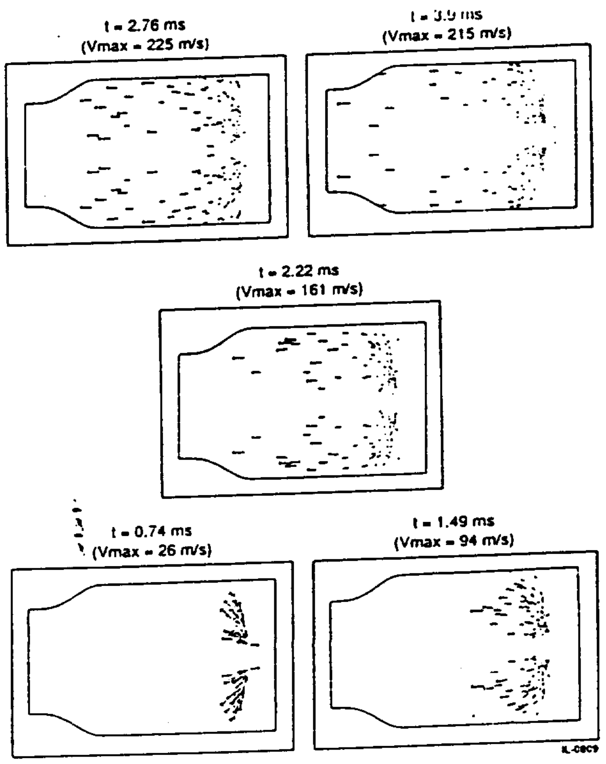


Figure 11. Temporal evolution of N_2O_4 droplet parcel velocity

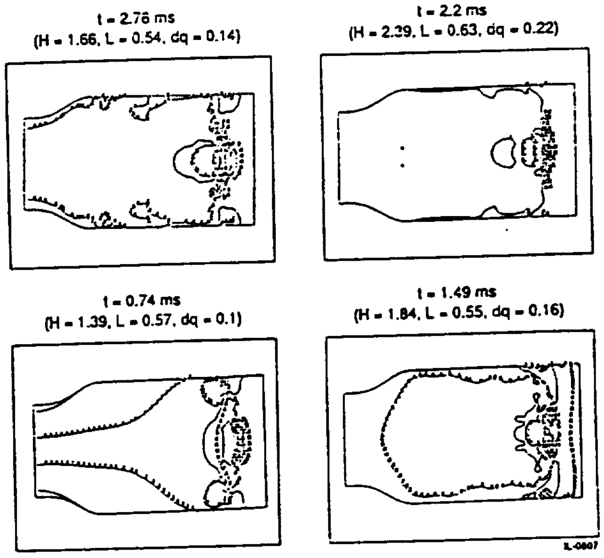


Figure 13. Temporal evolution of the fuel/oxidizer ratio (F/0)

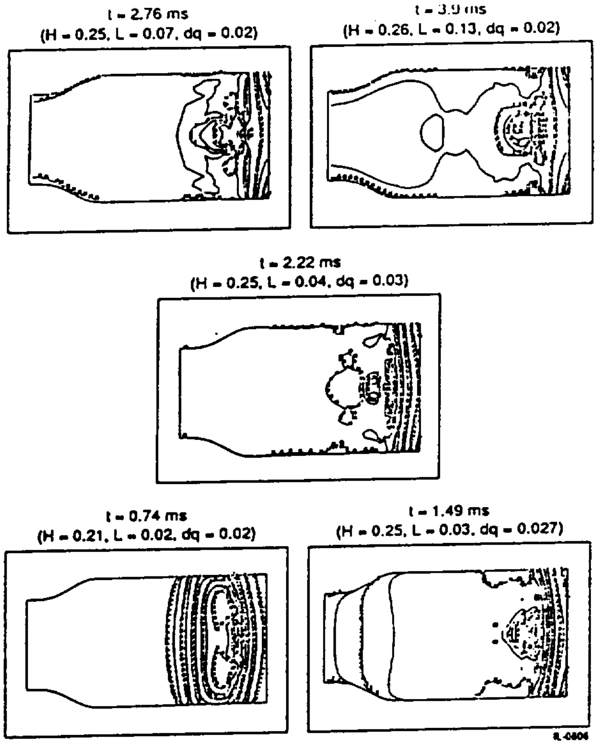


Figure 14. Temporal evolution of N_2 mass fraction

Project Title: Spray Combustion Stability

Principal Investigators: Dr. Pak-Yan Liang and Dr. S-M Jeng

Other Investigators: Mr. Ronald Litchford

Project Report

PROJECT OBJECTIVES

The central purpose of this project is the improvement of liquid-fueled rocket motor design technology in order to assist the establishment of economical commercial access to space through the development of engines with enhanced performance and reliability. Specific research effort in the project is focused on spray physics and associated combustion instability phenomena. Results garnered from this work will contribute to the development of new computational tools for design of stable liquid propellant rocket engines.

The specific objectives of the research effort include identifying and evaluating physical submodels which pertain to spray combustion stability with the idea of enhancing or refining existing submodels with a more comprehensive approach. In particular, any refinements to the spray combustion physical submodels which are achieved during the project will be channeled back to Rocketdyne for incorporation in their ARICC liquid rocket combustor code as second-generation improvements. Also, as the ARICC code forms the basis of future CFD development, some effort is devoted to an evaluation of the code's capability for modeling oscillating pressure waves within the combustor.

Over the past year, personnel associated with this project have changed significantly. Dr. Jeng assumed the role of principal investigator in May, and Mr. Litchford joined the project in June. The major research activities concerned with this project during 1989 may be summarized in the following three areas: (1) evaluation of state-of-the-art submodels for liquid rocket combustor instability study, (2) improvement of droplet combustion model at elevated pressure, (3) development of new, computationally efficient droplet tracking models.

ACCOMPLISHMENTS

Existing physical submodels have been evaluated with special emphasis on their capability for spray combustion instability study. The ARICC code has also been tested for transient spray and instability calculations. From these evaluations, submodel development work crucial to the establishment of a predictive liquid rocket combustion instability code have been identified. It was realized that submodel improvements must be made before any CFD code can accurately calculate combustion instability, and a decision was made that initial work be devoted to physical submodel development. Three principal tasks were derived as the focus of a three year research effort.

Task I - High Pressure Droplet Gasification Model

The droplet vaporization rate is believed to be one of the most important parameters in determining liquid rocket combustion instability. Furthermore, currently available models for rocket combustor calculations are based on derivations for low ambient pressure, whereas practical combustors operate at supercritical chamber pressures. Because high-pressure thermodynamics equilibria calculations have shown a significant deviation in equilibria concentrations and heat of vaporization from low pressure assumptions, a high-pressure vaporization model is required to accurately predict the evaporation rate. Work toward development of an engineering model with the desired high-pressure characteristics has comprised much of the research effort during 1989 and has been concentrated on LOX droplet vaporization within the SSME environment.

Task II - Droplet Turbulent Dispersion Model

Lagrangian tracking for droplet positions and associated properties has been recommended by rocket researchers as the best method for spray combustion modeling. However, due to limited computer speed and memory, a "computational droplet" in a CFD code generally represents 10^4 - 10^6 physical droplets. In addition, it is uneconomical and impractical in the engineering design practice to have a massive, statistically significant number of computational droplets in a CFD code. The reverse problem is that an insufficient number of computational droplets will generate computational noise and forbid spray combustion instability calculations. During the past year, initiation of a new model requiring a small number of computational droplets and capable of averting numerical difficulties has been made. The basic concept of the model has been established, and calculations have been compared against simple spray flows.

Task III - Spray Atomization Model

Spray angle and droplet size distributions calculated from the existing spray atomization models are based upon semiempirical correlations. The pressure and flow fluctuations upon the atomization process have not been taken into consideration, and this limits code capability for instability study. It has been recognized that a fully predictive method for spray atomization processes will not be mature within the next decade, and the intention in this study is to develop a transient spray atomization process based on a semi-empirical method. This task will commence during 1990.

High Pressure Droplet Gasification

Research effort during 1989 has been concentrated on evaluation of existing and newly formulated high-pressure droplet gasification models. A full report of this study can be found in Reference [1]. Such attention is warranted by the role of propellant vaporization as a rate limiting participant in instability mechanisms and by the fact that practical liquid rocket combustors operate at supercritical mean chamber pressures. Furthermore, pressure sensitivity of the propellant vaporization rate must be correctly reflected by the mathematical submodel in response to acoustic pressure perturbations. Because the current droplet evaporation submodel in the ARICC code is a heuristic extension of conventional low-pressure theory, reformulation of the model has been undertaken from first principles, following an extensive review of the literature to assimilate a basis for submodel development.

A thorough study of previous work in this area has indicated that in approaching the modeling of droplet gasification at elevated pressures, the fundamental physical mechanisms which distinguish the high-pressure case from the low-pressure case must be recognized and taken into consideration. The peculiarities encountered at elevated pressures derive primarily from thermodynamic considerations. Most notably, with increasing pressure absorption of ambient gas into the liquid occurs and real gas effects become important in determining thermophysical properties and thermodynamic vapor-liquid equilibria. Therefore, multi-component thermodynamics based on a realistic equation of state must be considered in the model formulation. Inclusion of realistic thermodynamics further implies the attainment of higher droplet temperatures which in turn demand that the surface regression effect and bulk thermal expansion be included in the transport model. The preceding considerations formed the basis for a high-pressure droplet vaporization model formulation.

The fundamental approach used in the formulation is based on a quasi-steady gas phase coupled to a constant property liquid phase in time. Transport is modeled as a spherically symmetric, diffusion controlled process which can be corrected for convective effects using steady-state heat and mass transfer correlations. To account for the high-pressure effects, surface regression and bulk thermal expansion of the drop are considered, and high-pressure multi-component thermodynamics is employed to couple the liquid-phase and gas-phase solutions at the interface.

The transport processes during gasification of a propellant droplet are depicted in Figure 1. The derived governing system of equations reduce to a set of ordering differential equations in time where the heat and mass transport coefficients are obtained from the simultaneous solution of two analytic equations [Reference 1].

Because the droplet surface, under supercritical ambient pressures, can be heated to the critical state where the surface tension and heat of vaporization tend to zero, finite rate supercritical gasification can occur. Furthermore, since the preceding derived transport model is conceptually valid only to the critical state and due to the absence of definitive theory, the stripping vaporization criteria as proposed by Dr. Liang is tentatively retained. In this way, finite rate vaporization may be heuristically considered.

Because the conventional low-pressure thermodynamic model is invalid at moderate and increasingly higher pressures, proper evaluation of vapor-liquid equilibria and enthalpy of vaporization demands rather rigorous formulation from multi-component thermodynamics. From fundamental theory, the vapor-liquid equilibria constraints may be enforced such that the temperature, pressure, and component fugacities are identical across the phase boundary. If an equation of state is specified, the general thermodynamic relations for the fugacity coefficients may be integrated to provide analytic equations for the equilibria concentrations in each phase. An application of the Redlich-Kwong equation of state with the mixing rules of Prausnitz and Chueh, [References 2 and 3] known to yield good, verifiable results, is the conventional approach for droplet analyses at high pressure and has been effected in this study. In addition, the heat of vaporization may be determined from the rigorous thermodynamics relation between the partial molar enthalpies and the fugacity coefficients for each phase. The resulting thermodynamic model may be partially simplified in special cases to neglect ambient gas absorption while retaining the real gas effects.

The equilibria model was numerically solved for several hydrocarbon/inert gas mixtures for evaluation with respect to experimental data. Figure 2 shows the calculated results for a n-hexane/nitrogen binary system in comparison with experiment. Excellent Agreement with theory is obtained. Prediction of the low-pressure ideal-gas model is depicted by the dashed line to illustrate the associated error with increasing pressure. The calculated enthalpy of vaporization for the n-pentane/nitrogen binary system is shown in Figure 3 indicating substantial deviation from the latent heat of vaporization with increasing pressure.

Upon validation of the equilibria model, the formulated transport/thermodynamic governing equations were combined and numerically solved in a single droplet vaporization code. The general numerical procedure adopted for the study was to evaluate thermodynamic and thermophysical properties, heat and mass transfer coefficients, and vapor-liquid equilibria explicitly and assume these values hold over the numerical time step used in the solution of the governing ODE's. Calculations were conducted for comparison with experiment and some results are shown in Figures 4 and 5. Reasonable correlation with the data was obtained without adjusting any empirical constants.

With the basic validity of the theory established, attention was turned toward modification of the model to include the cryogenic propellants (O_2/H_2) of primary interest. Since hydrogen is a quantum gas, the mixing rules were modified in the manner proposed by Chueh and Prausnitz to accommodate the desired propellant system. Equilibria calculations for the H_2/O_2 system using the quantum gas mixing rules are given in Figures 6 and 7. Expected qualitative trends are obtained. As an investigative precursor to evaluation in a CFD code, the droplet vaporization code was then modified to a simplified 1-D rocket combustor code for the H_2/O_2 propellant system. LOX droplets were assumed to be axially injected as discrete, identically sized particles in a cylindrical chamber. The axial gas velocity was assumed to increase in proportion to the amount of mass evaporated from the droplets, and the temperature and pressure were taken to be constant throughout the combustor. Calculations with the highpressure equilibria model are compared with calculations using the current thermodynamic model in the ARICC code in Figures 8 and 9, with the principal result being that the high-pressure model predicts a much faster rate of development and a droplet lifetime roughly half that predicted by the currently used model.

The high-pressure gasification model formulated for implementation into already computationally intensive CFD codes is obviously a rather sophisticated submodel with computational demands exceeding those encountered for the low-pressure models currently used. However, the required additional computing time is demonstrably not excessive and the ability to encompass the physics of high-pressure gasification is crucial in instability analyses where the vaporization response to pressure perturbations acts as the rate-controlling Rayleigh mechanism for acoustic wave reinforcement.

Direction concerning future refinement of the currently proposed high-pressure gasification model is focused on extension for supercritical gasification and will depend heavily upon experimental results available for validation purposes. Currently, our thinking is to incorporate an integral equation formulation for convective liquid-phase energy transport to obtain the internal surface temperature gradient so that finite-rate vaporization may be considered once the surface reaches criticality. Other possible improvements include the addition of an envelope flame for conjugate supercritical combustion of LOX droplets and inclusion of H_2O in the equilibria calculation.

Droplet Turbulent Dispersion Model

The spatial distribution of propellant chemical energy release within a liquid rocket engine is primarily controlled by the droplet life history - droplet trajectory and vaporization rate. For spray combustion modeling, the Lagrangian droplet tracking method has been most recommended. The stochastic separated flow (SSF) models which utilize the Lagrangian tracking method consider both slip and fluctuations relative to the gas flow and have been well calibrated for simple flows.

In the SSF model, droplets are assumed to interact with a succession of turbulent eddies using random-walk calculations. For example, the droplet acceleration is calculated by

$$\frac{du_p}{dt} = \frac{3}{4}\frac{\rho C_D}{\rho_p d_p}(u - u_p)|u - u_p|$$

where $u_p, u, d_p, \rho_p, \rho, C_D$ are the droplet velocity, gas velocity, droplet diameter, droplet density, gas density, and drag coefficient. The gas velocity is calculated from its mean velocity plus random sampling from the fluctuating velocity.

In extension of the SSF model for rocket combustion instability study, however, two major difficulties are encountered. First, if an SSF model is used for time-dependent, three dimensional spray calculations, a rather sufficient number of computational droplets are required to cover the droplet probability distribution over nine independent variables - three spatial coordinates, three velocity components, droplet temperature, droplet size, and time. The random walk sampling over nine independent variables for each droplet property is still computationally too expensive or impossible even with the aid of supercomputers. Second, since each droplet in the SSF model represents a Delta function over nine independent variables, "computational shot noise" will be generated if an insufficient number of computational droplets are used in the code.

Schematic representation of a typical spray into stagnant air is shown in Figure 10. If a large number of computational droplets are used (number of droplets/number of computational cells > 100), the distribution of droplets per computational cell will be very smooth. However, if a small number of droplets are used (number of droplets/number of computational cells < 10) the simulated droplet distribution will be significantly altered from real flows. This zigzag distribution represents spatially non-uniform energy release and hinders successful combustion instability calculations.

Because of this inadequacy, it is desirable to modify the SSF model for time dependent spray combustion within rocket chambers. The modified model should require a small number of computational droplets and should generate less computational noise. Effort on this task began in August, and, thus far, the preliminary model has been established and calibrated against nonevaporating spray flows.

The proposed model considers that a computational droplet has a probability density function of temperature, size, number of physical droplets, and locations at different times. The shape of these distributions over each variable will not be changed, although the variance will be altered. In the original SSF model, the random sampling over whole PDF distributions are needed. For the proposed method only the mean and variance of the PDF are calculated. This results in a significant savings in cpu time and memory requirements. The model has been tested against simple non-evaporating spray flows as shown in Figure 10.

The calculations for the gas phase were performed using a modified version of GENMIX. The droplet trajectory was computed using a second order finite difference algorithm. Details of the SSF model may be found in Reference 4.

Several test cases using from 100 to 8000 computational droplets have been performed with 50 gas-phase computational cells radically. Figures 11-13 show the radial distribution of the number of droplets per gas-phase cell calculated from the original SSF model and the modified SSF model. Results for the original SSF model using 200 and 600 droplets show significantly irregular distributions, whereas the modified SSF model greatly reduced the noise. Both models predict similar distributions when a large number of computational droplets ($N_p = 8000$) were used. Clearly, the modified SSF model can predict a better spatial distribution when a small number of computational droplets are used.

Figure 14 shows the predictions of three different dispersion models. The original deterministic separated flow (DSF) model assumes that the gas-phase fluctuation velocity has no effect on the droplet trajectory and that each computational droplet represents a point source for a group of physical droplets. In the modified DSF model, droplet dispersion is considered through the turbulence effect on the position variance. The results from both SSF models using 8000 computational droplets to simulate the spray are also plotted in the figure. Both DSF models simulated the spray using 40 computational droplets which represents a significant reduction in the number of droplets considered compared to the SSF model. Clearly, the DSF model without dispersion considerations underpredicted the distribution width. The DSF model with dispersion rate calculations can predict the global radial distribution of droplets fairly well, although it did a poor job in calculating the local structure.

REFERENCES

1. An Interim Research Progress Report: Concerning High-Pressure Droplet Gasification Submodel Refinement for ARICC-LJ Spray Combustion Code, A Report prepared by The Center for Advanced Space Propulsion for the Rocketdyne Division, Rockwell International Corporation, August 1989.

2. Chueh, P.L. and Prausnitz, J.M., "Calculation of High-Pressure Vapor-Liquid Equilibria," I&EC, Vol. 60, No. 3, pp. 34-52, 1968.

3. Prausnitz, J.M. and Chueh, P.L., Computer Calculations for High-Pressure Vapor-Liquid Equilibria, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1968.

4. Shuen, J-S, Solomon, A.S.P., Zhang, Q-F, and Faeth, G.M., "Structure of Particle-Laden Jets: Measurements and Predictions", AIAA J., Vol. 23, No. 3, pp. 396-404, 1985.

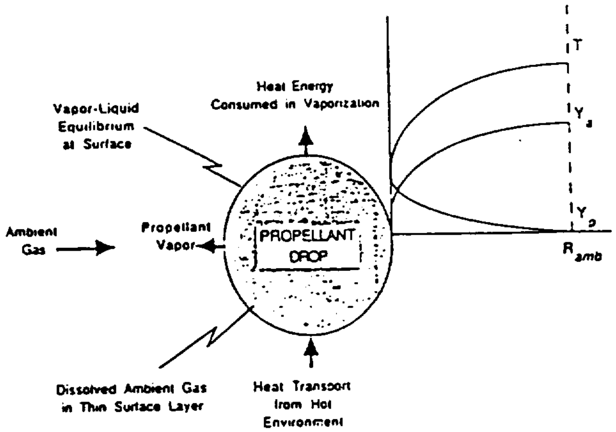


Figure 1 Schematic of transport processes during gasification of a propellant droplet.

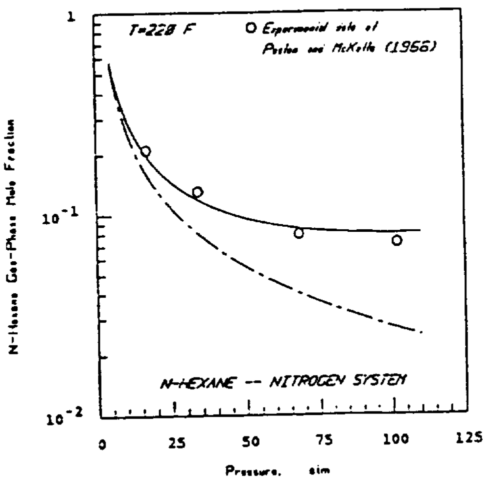


Figure 2 Comparison of theoretical equilibria from no-absorption model and ideal gas model (dashed line) with experimental data for the n-hexane—nitrogen binary system at T_∞ = 220 F.

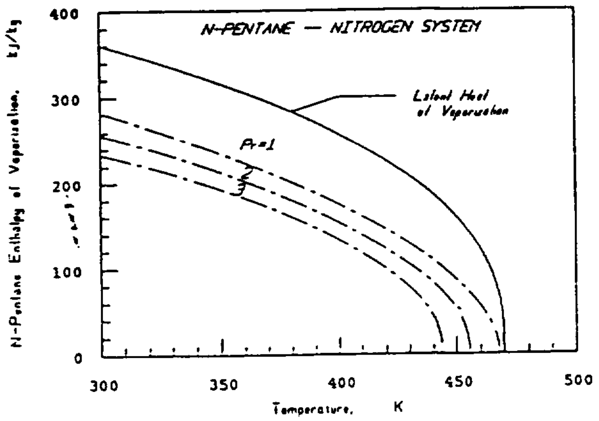


Figure 3 Comparison of theoretical enthalpy of vaporization from general model including ambient gas absorption at supercritical pressures with the latent heat of vaporization for the n-pentane—nitrogen binary system.

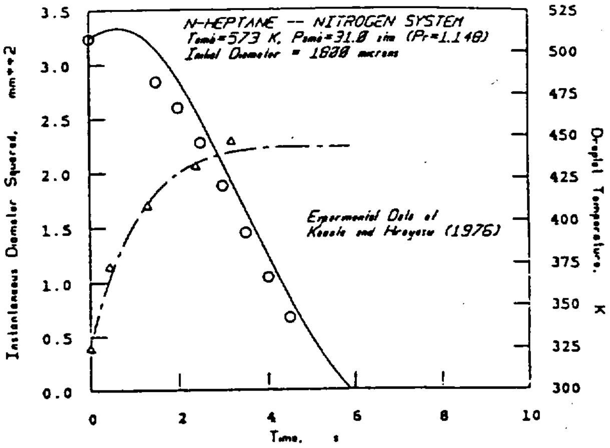


Figure 4 Comparison of theoretical and experimental vaporization histories of n-heptane droplet in nitrogen gas at 573 K and 31 atm.

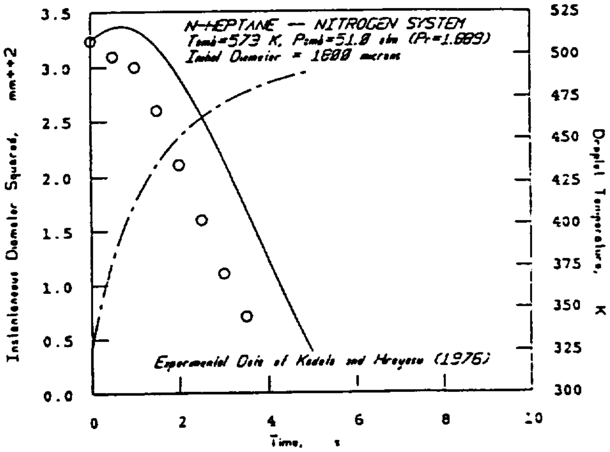


Figure 5 Comparison of theoretical and experimental vaporization histories of n-heptane droplet in nitrogen gas at 573 K and 51 atm.

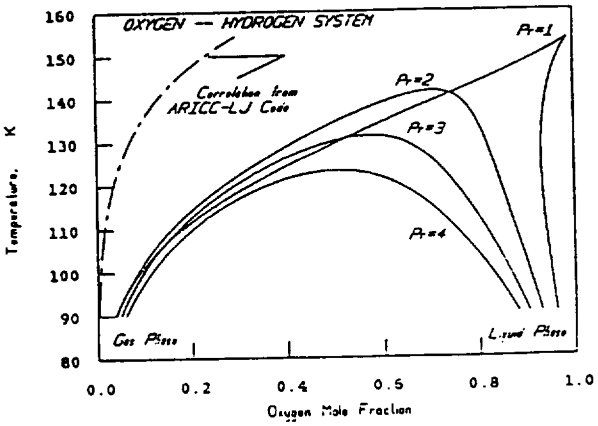


Figure 6 Comparison of theoretical equilibria from general model including ambient gas absorption with current correlation in ARICC-LJ code for oxygen—hydrogen binary system.

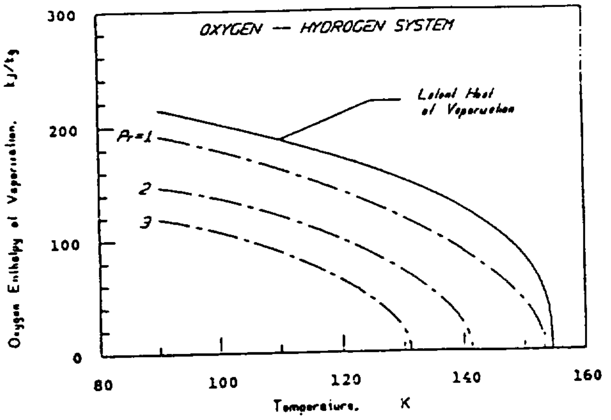


Figure 7 Comparison of theoretical enthalpy of vaporization from general model including ambient gas absorption at supercritical pressures with the latent heat of vaporization for the oxygen—hydrogen binary system.

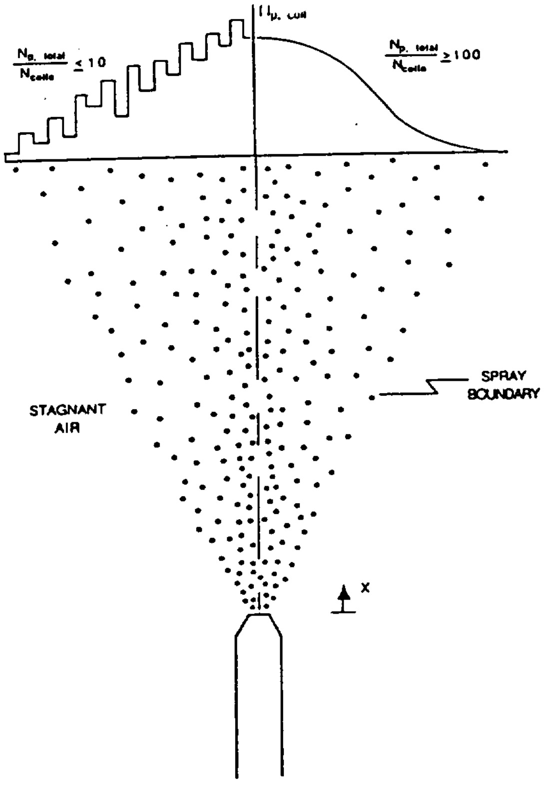


Figure 10. Sketch of droplet positions in spray flows[25]

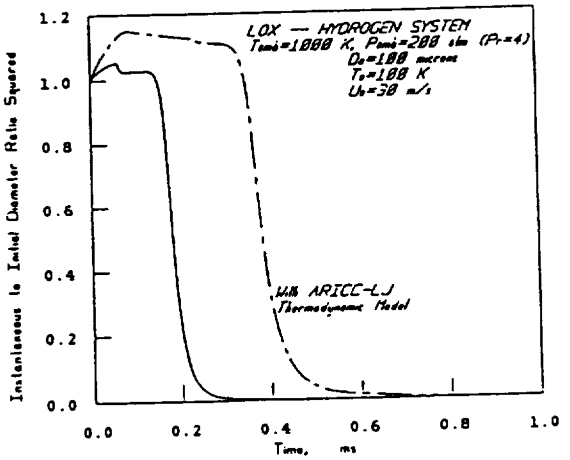


Figure 8 Calculated $d^2_{s}/d^2_{s_0}$ ratio histories for LOX/H_2 liquid rocket combustor.

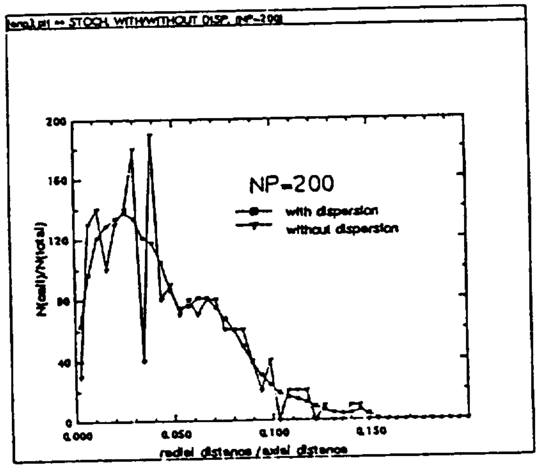


Figure 11. Comparison of calculated results from three different models.

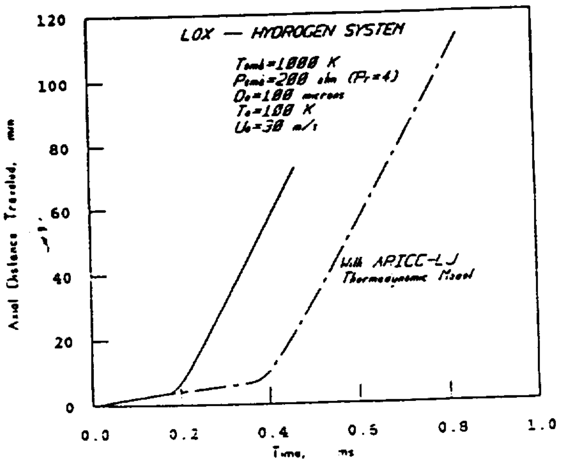


Figure 9 Calculated axial distance traveled by LOX droplet in LOX/H_2 liquid rocket combustor.

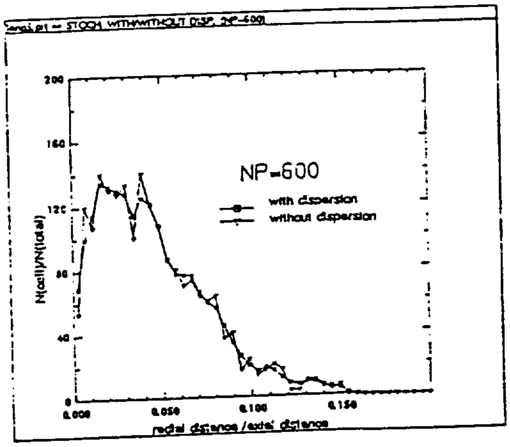


Figure 12. Radial variation of number of droplets per gas phase computational cell, total computational droplets = 600

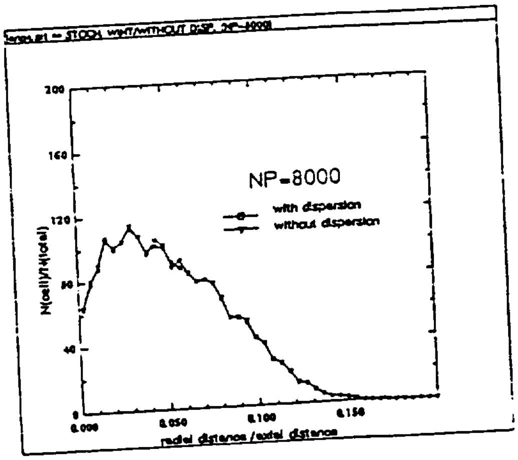


Figure 13. Radial variation of number of droplets per gas phase computational cell, total computational droplets = 8000

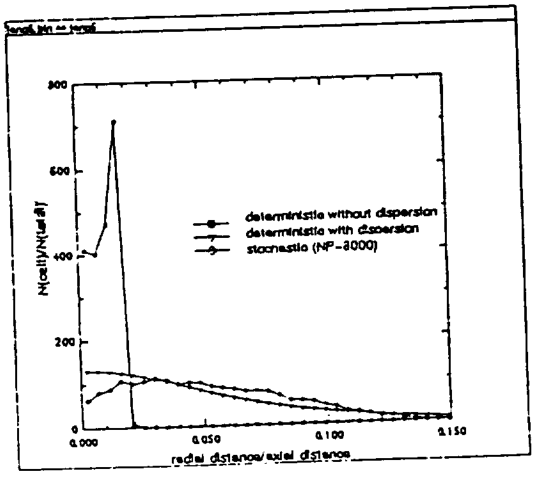


Figure 14. Radial variation of number of droplets per gas phase computational cell, total computational droplets = 200

Project Title: Advanced High Area Ratio Nozzles

Principal Investigators: Dr. Farhad Raiszadeh and Dr. Frank G. Collins

Other Investigators: Mr. Joseph L. Orr, Jr. and Mr. Brian Myruski

Project Report

PROJECT OBJECTIVES

The objective is to develop computational techniques for the design of high-area-ratio nozzles and to validate these models by comparison with experiments and computations using other codes.

FY89 ACCOMPLISHMENTS

Progress was made in two areas during the past year. First, performance computations were added to the PARC2D code and the performance of the SSME nozzle was computed for inviscid, laminar and turbulent flow assuming a perfect gas with $\gamma = 1.2$. Second, the PARC2D code was modified in a non-CASP project to compute equilibrium flow about hypersonic blunt bodies. Progress has been made toward modifying this code to compute equilibrium H_2/O_2 flow through the SSME and related nozzles.

First consider the perfect gas computations in the SSME nozzle. Last year it was reported that the PARC2D code in the Euler mode gave excellent agreement with inviscid solutions to the same flow computed by the Rockwell USA code and by the JANNAF method of characteristics code. This year laminar and turbulent solutions to the flow field were obtained and are reported in the thesis of Joseph L. Orr, Jr. Several computational grids were used, including one that clustered grids in the location of the embedded shock wave plus a tight clustering along the wall. The other grids had various wall packings. The results of the computation were found to be somewhat dependent upon the grid, especially for the boundary layer temperature distribution and the predicted adiabatic wall temperature. After this work had been completed it was discovered that other investigators found that the first grid point must have $y+ < 0.2$ for the solution to be independent of the grid [J.S. Ryan, J. Flores and C.Y. Chow, "Development and Validation of CNS (Compressible Navier-Stokes) for Hypersonic Applications," AIAA-89-1839, presented at AIAA 20th Fluid Dynamics, Plasma Dynamics and Laser Conference, June 12-14, 1989]. All future computations will use this criteria to ensure grid-independent solutions. Typical velocity and temperature boundary layer profiles are shown in Figures 1 and 2.

The solutions used a new artificial viscosity model for the PARC2D code, developed by personnel at Sverdrup Technology, Inc., AEDC Group. This model decouples the artificial viscosity in the two coordinate directions and reduces the fourth-order dissipation (based on the local cell Reynolds number) in the boundary layer to zero at the wall. This new artificial viscosity model is not nearly as robust as the original model and great care had to be taken to get a converged solution. The turbulent solution was even more difficult to obtain and L2 decreases of only 10^4 to 10^5 could be achieved.

Performance was added to the code by integrating across the nozzle exit plane to get the mass flux and the thrust. Integration across the throat gave exactly the same value of mass flux. Because of the thin boundary layer in the SSME nozzle the predicted mass flux and specific impulse were very independent of the computational grid. The various grids are described in Table 1 and the performance results are given in Table 2. The specific impulse decrement for the laminar boundary layer was -1.4 sec and -5.5 sec for the turbulent boundary layer, compared to the inviscid performance.

The specific impulse for a perfect gas depends upon γ and M, the molecular weight. Examples of the sensitivity to these parameters are given in Table 3. The predictions given in Table 2 assumed O/F = 8 so that only H_2O would appear after combustion. Thus, M was assumed to be equal to 18. However, the actual SSME operates with O/F = 6, giving an average molecular weight of 14 after complete combustion. In addition, γ will vary through the nozzle. The first part of Table 2 compares the changes expected in specific impulse if $\gamma = 1.2$, $M = 16$ while the second assumes $\gamma = 1.18$, $M = 16$. In both cases the specific impulse increased compared to the computed results for $M = 18$, $\gamma = 1.2$ given in Table 2. Only an equilibrium code can predict the actual specific impulse.

During the past two years the PARC2D code was modified to compute the flow field about blunt bodies moving at hypersonic speeds through the upper altitudes of the earth's atmosphere. The thermodynamic properties are computed from statistical mechanics, incorporating the latest spectroscopic values to compute the partition function and including a significant number of excited electronic states. The mixture is assumed to be that of perfect gases because of the relatively low density, even after the bow shock, at the high altitude of interest. The composition at a given temperature and pressure is computed from the law of mass action, using statistical mechanics to compute the equilibrium constants. Diffusion is ignored except in the computation of the coefficient of thermal conductivity.

The transport properties are obtained using the latest approximation to the exact kinetic theory relations and the latest computations of the collision integrals. The latter were obtained using accurate representations of the interaction potentials. There are three contributions to the coefficient of thermal conductivity of the mixture, from the transport of translational energy, the transport of internal energy and a reactive contribution caused by the diffusion that occurs in regions of dissociation, for example. This last contribution can be an order of magnitude greater than the other contributions to the thermal conductivity and can have a significant influence on the flow field.

The equilibrium code was applied to the flow field about the Aeroassist Flight Experiment (AFE) vehicle which skips along the outer edge of the atmosphere at a Mach number of 33. The perfect gas temperature was over 40,000 K after the shock wave but the predicted equilibrium temperature was around 8,000 K. The stagnation point heat transfer was found to be significantly influenced by the real transport properties, compared to the usual Sutherland approximation.

This equilibrium version of PARC2D is presently being modified for H_2/O_2 . At the present time the thermodynamic properties of the molecules present in H_2/O_2 combustion, namely, H_2, O_2, H, O, OH, H_2O and HO_2, can be accurately computed up to 7,000 R, as compared to the JANNAF tables. Computed values of the specific heat are shown in Figure 3. The composition for various O/F ratios, pressures and temperatures as well as the mixture properties can be accurately computed, when compared to the results tabulated in NASA SP-3011. The adiabatic flame temperature can also be accurately computed. As examples, the temperature and pressure computed along the centerline of the SSME nozzle by the perfect gas code were used to compute the composition and ratio of specific heats for various locations in the nozzle. These results are shown in Figures 4 and 5. Although these results will not be repeated by the equilibrium code, they do indicate how the composition and gamma will later be found to vary through the nozzle flowfield.

Final modifications to the equilibrium version of PARC2D are presently underway and this new code will be used for various projects during the next year. These will include the SSME nozzle as well as the enlarged-throat SSME nozzle that was tested at NASA MSFC. The computed results for the latter nozzle will be compared with the measured exit flowfield.

PUBLICATION

Joseph L. Orr, "SSME Nozzle Flow Field Calculations Using PARC CFD Code", M.S. Thesis, The University of Tennessee, Knoxville, December 1989.

TABLE 1

Grid	Description
Case 1	Inviscid, non-shock-fitted grid
Case 2	Shock-fitted grid with wall packing
Case 3	Non-shock-fitted, slight wall packing
Case 4	Non-shock-fitted, moderate wall packing

TABLE 2

Assumption/Case	$\dot{m}$ (lbm/sec)	Thrust (x10⁶ lbf)	I_{sp} (sec)
Ideal, 1-D	1223.5	5.377	424.4
Inviscid, Case 1	1186.0	4.561	393.3
Inviscid, Case 2	1210.0	4.763	392.7
Laminar, Case 2	1209.0	4.732	391.4
Laminar, Case 3	1218.0	4.770	391.5
Laminar, Case 4	1216.0	4.764	391.3
Laminar, Case 2*	1215.0	4.754	391.3
Turbulent	1220.0	4.724	387.3

*With an existing enhanced artificial dissipation model

TABLE 3

	I_{sp} (sec)	
Flow Assumption	$\gamma = 1.2$, $M = 15$	$\gamma = 1.18$, $M = 15$
Ideal	450.1	470.7
Inviscid	416.7	435.7
Laminar	415.2	434.2
Turbulent	410.8	429.5

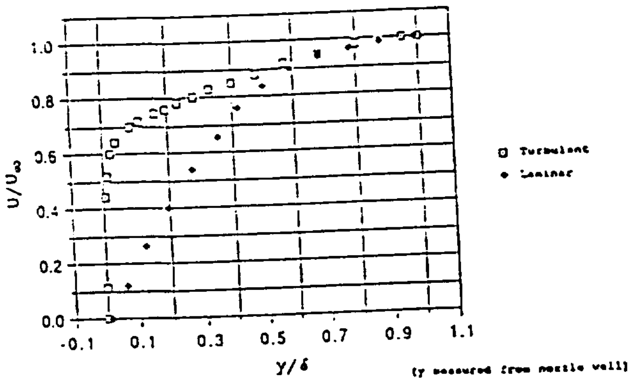


Figure 1. Boundary Layer Velocity Profile Comparison.

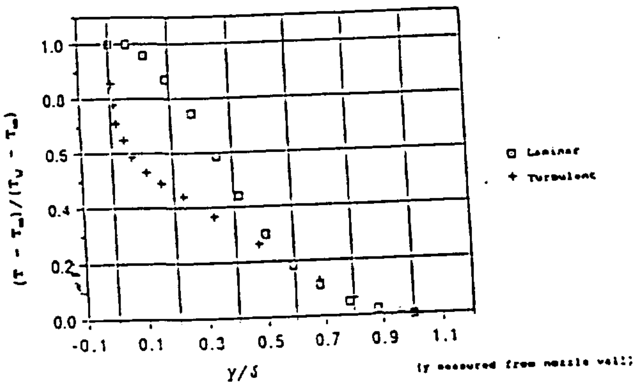
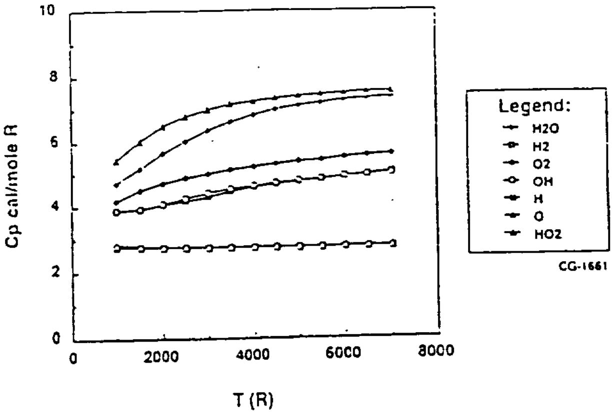
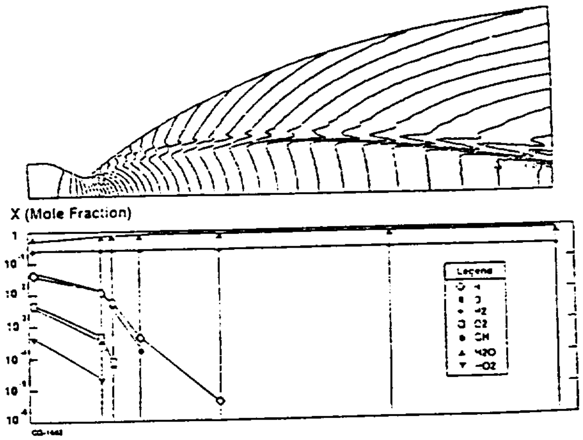


Figure 2. Boundary Layer Temperature Profile Comparison.



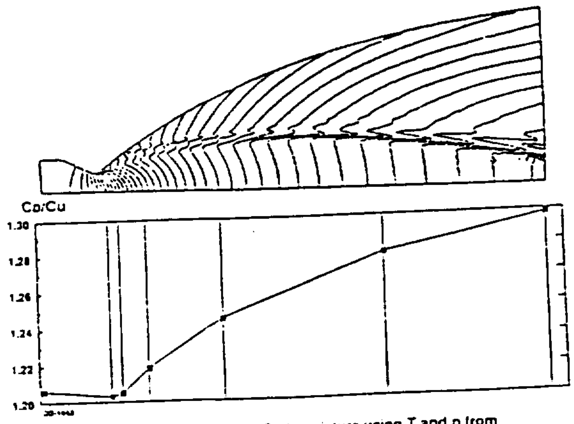
Specific heat for constituents in SSME Nozzle flow

Figure 3



Computed composition at various locations in SSME nozzle using locations from perfect gas solution O/F = 6.0

Figure 4



Computed ratio of Cp/Cu for mixture using T and p from perfect gas solution in SSME nozzle O/F = 6.0

Figure 5

A.I. PROPULSION APPLICATIONS

- Identification and Interpretation of Patterns in Rocket Engine Data: Artificial Intelligence and Neural Network Approaches
- SSME Propellant Path Leak Detection
- Intelligent Hypertext Manual Development for the Space Shuttle Hazardous Gas Detection System
- Classification of Data Patterns Using an Autoassociative Neural Network Topology
- SSME Component Assembly and Life Management Expert System
- Leader - An Integrated Engine Behavior and Design Analyses Based on Real-Time Fault Diagnostic Expert System for Space Shuttle Main Engine

Project Title: Identification and Interpretation of Patterns in Rocket Engine Data: Artificial Intelligence and Neural Network Approaches

Principal Investigator: Moonis Ali

Other Investigators: Bruce Whitehead, Uday K. Gupta and Harry Ferber

Sponsors: CASP/NASA

Project Report

INTRODUCTION

The Space Shuttle Main Engine (SSME) is a complex system whose health is monitored by a set of sensors placed at various points within the engine. While serious anomalies in engine behavior will ultimately cause large fluctuations in the sensor data, these large fluctuations may not occur until it is too late to shut down the engine or to take other corrective actions. The purpose of the work reported here is to automatically detect anomalies in engine behavior as early as possible, when their indications in the sensor data are much subtler. Early detection of these anomalies is challenging because of the large amount of noise present in the data. In the presence of this noise, early indications of anomalies become very difficult to distinguish from fluctuations in normal steady state operation due to the noise.

Detection of anomalies is also challenging because there is an extremely large set of possible fault conditions. Characteristics of only a fraction of these possible fault conditions are known, and only a very small fraction have ever been observed.

This paper describes an expert system which is designed to perform automatic data analysis, identify anomalous events and determine the characteristic features of these events. We have employed both artificial intelligence and neural net approaches in the design of this expert system. The artificial intelligence approach is useful because it provides:

1. The use of human experts' knowledge of sensor behavior and faulty engine conditions in interpreting data.
2. The use of engine design knowledge and physical sensor locations in establishing relationships among the events of multiple sensors.
3. The use of stored analysis of past data of faulty engine conditions.
4. The use of knowledge-based reasoning in distinguishing sensor failure from actual faults.

The neural network approach appears promising for four reasons:

1. Neural networks can be trained on extremely noisy data, and produce classifications which are more robust under noisy conditions than other classification techniques.
2. The neural network technique thus avoids the necessity of noise removal by digital filtering, and therefore avoids the need to make assumptions about frequency bands or other signal characteristics of anomalous behavior.
3. Neural networks can, in effect, generate their own feature detectors based on the characteristics of the sensor data used in training. This is in contrast to other techniques which are based on the analysis of predetermined features of the sensor data. If anomalous engine behavior is indicated by some unanticipated feature present in the sensor data, a neural network is more likely than other techniques to detect this unanticipated feature.
4. Neural network algorithms are inherently parallel, and therefore are potentially implementable in special-purpose parallel hardware. Neural network algorithms therefore have the potential of being useful for the detection of anomalous engine behavior quickly enough for suitable corrective action to be taken.

Architecture A

The architecture of the Data Analysis Expert System (DAES) is shown in Fig. 1. The high frequencies of the noisy data are first filtered through a two-band digital filter so that the significant features of data corresponding to anomalous engine conditions are not lost. The upper and lower thresholds of the filter were determined by minimizing the number of segments in the filtered data. A segment is defined as the time duration between two consequent zero crossings in the first derivative of the filtered data. The Segmentation Module, shown in Fig. 2, identifies the segments. These segments are input to the Significant Event Selection Module (SESM) shown in Fig. 3. SESM compares the features of each input segment with the features of the most deviant segment belonging to normal steady state sensor data of the corresponding sensor stored in the data base. The input segments which are more deviant than the segments of normal data are given as the significant events at the output of SESM. The event identifier, shown in Fig. 1, separates abnormal significant events from scheduled events by employing the knowledge about the scheduled events. This knowledge includes a planned event schedule and computer commands profile as well as specific features of scheduled events.

Results of Prototype A

A prototype called 'PROTOTYPE A' has been developed on the basis of the above described architecture A, shown in Fig. 1. In order to illustrate the results, we selected the analysis of HPFTP (High Pressure Fuel Turbo Pump) failure in SSME test #901-364, specifically, hot gas intrusion to rotor cooling. The probable cause of the failure was a new HPFTP thermal shield retainer nut assembly used for the first time on this test. The geometry of the nut allowed a direct leak path through the heat shield for the high temperature gas which produced two jets impinging directly on the turbine end cap (Kaiser helmet) and reducing material properties in the impingement zone. Main Combustion Chamber Hot-Gas Injection Pressure (MCCHGP) sensor data from 320 to 400 seconds after the start of the test is shown in Fig. 4. The occurrence of the fault is noticeable from the global pattern which is superimposed by the high frequencies. The filtered data is shown in Fig. 5 in which the global pattern is more visible. SESM identifies three significant events which are then further analyzed and their features are determined. These features are shown in Fig. 6. The identification of the significant events by DAES is compared with the NASA and Rocketdyne human experts' analysis. DAES findings were identical to human experts'

Architecture B

In Architecture A, digital filtering, segmentation, and significant event selection are performed by three separate modules, as shown in Figure 1. This method of identifying significant events is expensive computationally, and requires filtering out high frequencies in order to extract a reasonable number of segments for further analysis. The alternative is to replace these three modules with a neural network. Because of the noise resistance of neural networks, it is unnecessary to remove high frequencies by digital filtering. Instead, the neural network accepts data in the form of a complete power spectrum for each successive window of data. The purpose of the neural network is to identify significant events on the basis of these spectra, without further filtering or segmentation. Figure 7 depicts the neural network architecture we are currently investigating as an alternative to the digital filter, segmentation, and significant event selection modules of Figure 1. Fast Fourier Transform (FFT) preprocessing is employed to produce data which is time-shift invariant within each FFT window, but without filtering out the high frequencies. The resulting spectral data is compared to the average power spectrum of normal steady-state data, and the difference in power at each frequency is fed into the neural network. The neural network is trained with this spectral data using Kohonen's [1988] self-organizing feature map.

Current tapes of SSME tests typically include not only sensor data, but also control information specifying the overall commands (such as thrust changes) given to the engine and the resultant commanded value settings. The overall sequence of steps in training the network and classifying the data proceeds as follows:

1. If control information for the test is available, remove scheduled transients from the new data. If control information is not available, analyze the entire data stream.
2. Perform Fast Fourier Transform of sequential windows of the time series test data.
3. If normal steady-state data for the same sensor is in the database, subtract the normal power spectrum from the power spectrum of the new data, yielding a delta (difference from normal)

power spectral density (P. S. D.)

4. Set the weights in the neural network to random initial values.
5. Using Kohonen's unsupervised learning algorithm to adjust weights within the network, sequentially feed in the delta P. S. D.'s of successive windows of test data.
6. Turn off the learning algorithm. The weights in the network now represent the result of the learning algorithm.
7. Again feed in the delta P. S. D.'s of successive windows of test data. For each window of data, the output of the network reflects the neural network's classification of that data.
8. Apply a threshold to the output activation level based on the observed outputs of normal steady-state data. The amount by which activity exceeds this threshold represents the level of confidence that an event has occurred.

Preliminary Results of Prototype B

Our initial implementation of the architecture above has been operational for a few weeks, during which time we have evaluated its performance on SSME tests, for which both raw sensor data and final NASA Investigative Board Reports were available. Our initial analysis was confined to each sensor meeting both of the following conditions: (i) The NASA final report indicated an anomaly in the data for that sensor which occurred significantly before engine cutoff; and (ii) availability of data for that sensor. Results for the three sensors meeting these conditions are shown in Table 1. The original time-series sensor data (Figure 8), and plots of neural network activation obtained from FFT window sizes of 64 data points (Figure 9) and 32 data points (Figure 10) are shown for one of the sensors, Main Combustion Chamber (MCC) Hot-Gas Injection Pressure. The neural network activation plotted is the result of Step 8 above, and represents the level of confidence that an event has occurred, as computed by the network based on the data with which it has been trained in each case.

As shown in Table 1, and illustrated in Figures 9-10, the neural network detects events in agreement with the NASA Investigative Board analysis. Obviously, a smaller FFT window size yields more precise timing of the events detected. On the other hand, on the basis of this pilot data, it appears that decreasing the FFT window size makes the anomaly detection process less resistant to steady-state noise. (Compare Figures 9 and 10). There would thus appear to be a trade-off between precision and reliability in the anomaly detection process. Our next step will be to investigate the effect on this trade-off of various alternatives in preprocessing steps 1-3 above, such as window shape, window overlap, and the method of comparing new data with known normal data in Step 3. While the initial results of this technique appear promising, much more work is needed before this neural network approach can be adequately compared with other approaches to identifying and classifying anomalies in SSME sensor data.

CONCLUSION

The preliminary results obtained from both prototypes are in agreement with analyses done by NASA human experts. While much more work is needed for a complete evaluation, it appears that these architectures are a promising first step in automating the extraction of features and subsequent diagnosis of rocket engine data.

REFERENCE

Kohonen, T. Self-Organization and Associative Memory. Springer-Verlag, 2nd ed., 1988.

SENSOR	TIME OF OCCURRENCE OF FIRST ANOMALY		
	As analyzed by NASA Board	As indicated by neural network with FFT window size (in data points	
		32	64
Fuel Preburner Ch Pr	385.88	385.7 - 386.9	385.7 - 386.3
HP Fuel Pump Speed	385.92	385.7 - 386.9	(not detected)
MCC Hot-Gas Inj Pr	386.00	385.7 - 386.9	(not detected)

TABLE 1: Anomalies in sensor data detected by neural network activations. The time interval given for each anomaly is that of the FFT window in which the anomaly was detected.

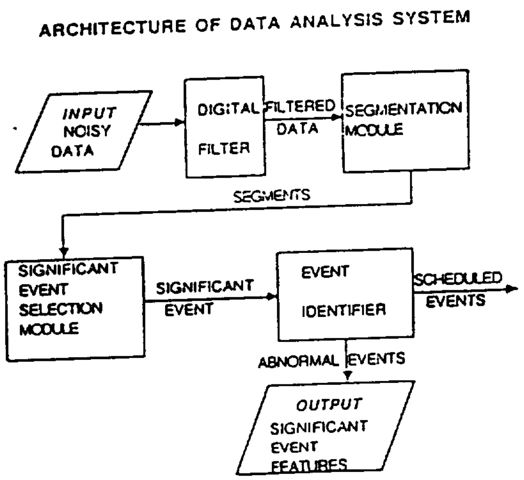


FIGURE 1

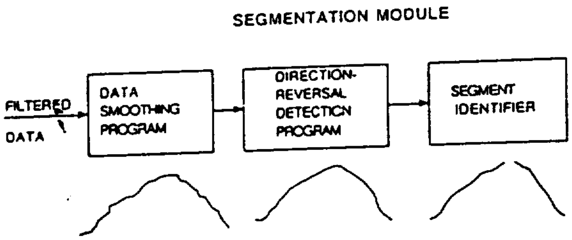


FIGURE 2

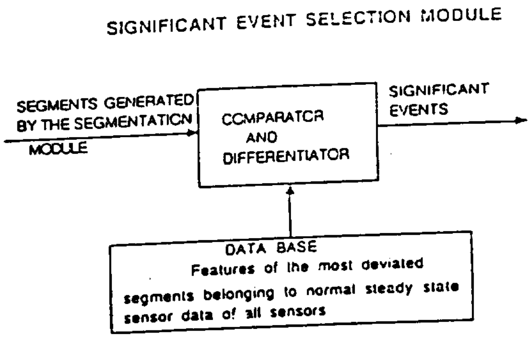
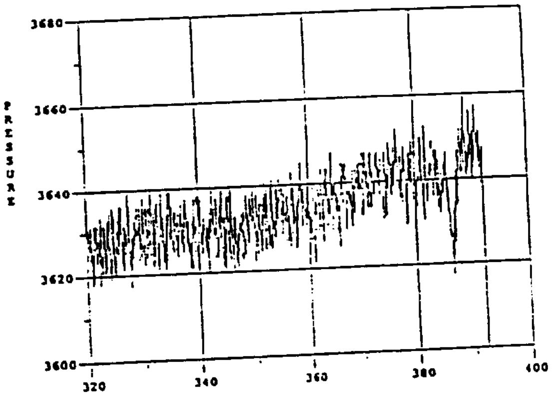


FIGURE 3



MCC HGP
(RAW DATA, TEST#901-364, TAPE#4, 50HZ)

FIGURE 4

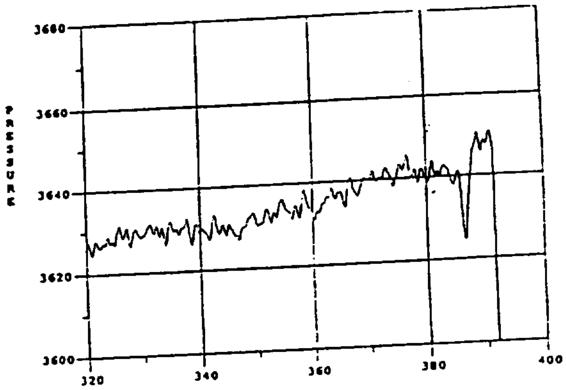
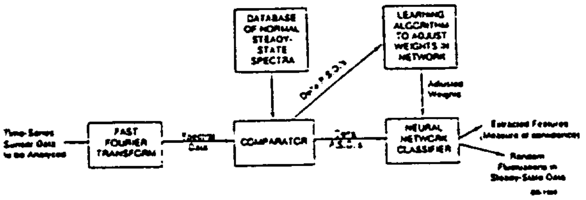


FIGURE 5: MCC HGP
(FILTERED DATA, TEST#901-364, TAPE#4, 50HZ)

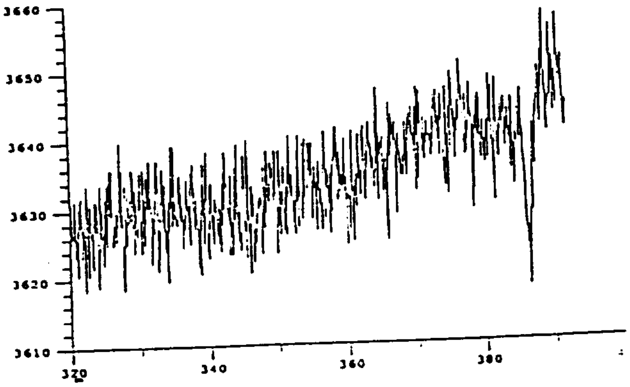
```
                 EVENT 1                                   EVENT 2

START-TIME     :  385.429              START-TIME     : 386.789
END-TIME       :  386.349              END-TIME       : 387.939
RATE-OF-CHG    :  LOW (-16 DEG)        RATE-OF-CHG    : LOW (13.5 DEG)
TOTAL-CHG      :  10.5791              TOTAL-CHG      : 16.21328
SHAPE          : "CONCAVE-CONCAVE"     SHAPE          : STRAIGHT


                                                EVENT 3
                                       START-TIME     : 391.320
                                       END-TIME       : 392.100
                                       RATE-OF-CHG    : HIGH (-83.2 DEG)
                                       TOTAL-CHG      :  301.5295
                                       SHAPE          : STRAIGHT
```

FIGURE 6: FEATURES OF THE THREE SIGNIFICANT EVENTS



AUTOMATIC FEATURE EXTRACTION
FROM SSME SENSOR DATA:
CURRENT ARCHITECTURE

FIGURE 7



Tape #4, Test 901-364, MCC Hot-Gas Inj Pressure

FIGURE 8

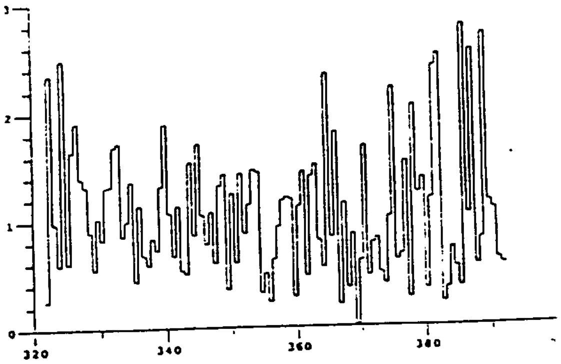


FIGURE 10: 901-364 MCC HG Inj Pr: Neural Net Activ

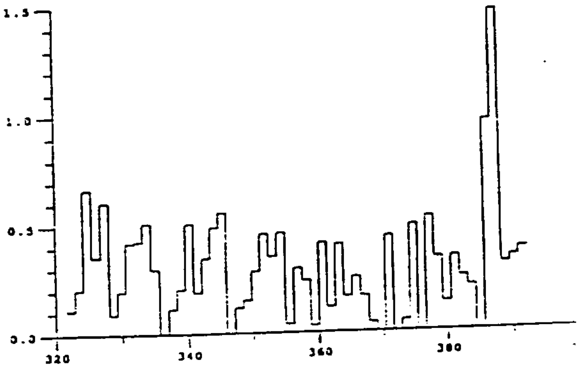


FIGURE 9: 901-364 MCC HG Inj Pr: Neural Net Activ

Project Title: SSME Propellant Path Leak Detection

Principal Investigator: Roger Crawford

Other Investigators: Ahmad Ali Shohadaee and W.T. Powers

Project Report

PROJECT OBJECTIVES

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, therefore, detection of both cryogenic and hot gas leaks is the objective of this investigation.

The primary objective of this phase of the 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.

One advanced propellant path leak detection technique is obtained by using state-of-art technology of infrared (IR) thermal imaging systems combined with computer, digital image processing and expert systems for the engine protection. The feasibility of the IR leak plume detection will be evaluated on subscale simulated laboratory plumes to determine sensitivity, signal to noise, and general suitability for the application. The theoretical analysis was undertaken with the objective of developing and testing simple, easy-to-use models to predict the amount of radiation coming from a radiation source, background plate (BP), which can be absorbed, emitted and scattered by the gas leaks.

THEORECTICAL ACCOMPLISHMENT

During this period the investigation concentrated on developing theoretical and numerical models to predict radiation transport in absorbing, emitting and scattering media.

The transport equation describing the spectral intensity field is:

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

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:

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

and the albedo for scattering $\Omega_{0\lambda}$, defined as the ratio of the scattering coefficient, $\sigma_{s\lambda}$, to the extinction coefficient, $K_\lambda = \sigma_{s\lambda} + a_\lambda$,

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

where a_λ is absorption coefficient.

Equation (3) is the integrated form of the equation of transfer

$$\begin{aligned} i'_\lambda(k_\lambda) &= i'_\lambda(0)\ exp\,(-k_\lambda) \\ &+ \int_0^{k_\lambda} I'_\lambda(k^*_\lambda, w)\ exp\,[-(k_\lambda - k^*_\lambda]dk^*_\lambda \qquad (3) \end{aligned}$$

where k^*_λ is a dummy variable of integration.

Equation (3) 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 radiation transport calculations have been done for a non-gray and non-isothermal gas (H_2O). Water vapor was selected as the molecule to concentrate detection on due to the water vapor present in the hot hydrogen propellant lines. The radiation transport equation in the absorbing and emitting 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 will be used to evaluate the intensity of background and leak plume radiation reaching the IR detector. The current radiation transport model for a simulated leak plume in form of cylindrical geometry has been developed (Figure 1). Numerical calculation of directional intensity passing through this cylindrical plume of water vapor at each y location (see Figure 2) has been completed. Figure 3 and 4 are the typical numerical variations of intensity at different y locations. Figure 3 indicates that the intensity has maximum values at $y = 0$ and minimum values at $y = R$, which has the same pattern for each value of wave length but the peak values appear at different intensity due to the absorption coefficient at that particular 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. When y values increases, the S value decreases, therefore, there is not enough distance for the incoming energy to be absorbed. Also, closer to the outer edge of the plume the temperature decreases with increasing r.

The variation of intensity with wavelength is shown in Figure 5 and 6. For case of hot plate with cold plume Figure 5 indicates that most of the incoming energy has been absorbed between 2.4 and 2.9 μ for different y values, but for case of cold plate with hot gas, Figure 6, shows that most of energy has been emitted by the hot gas.

EXPERIMENTAL RESULTS

The Thermovision 870 system operating in the 2-5 micron 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. The availability and successful utilization of several Thermovision 870 systems at the U.S. Air Force Arnold Engineering Development Center, aided the selection process. A formal loan request from CASP to AEDC has been approved, and the IR instrumentation was made available in July 1989.

The feasibility experiments have been completed using the Thermovision system to obtain images of both hot and cold plumes in front of a warm background plate (350°F). The cold plume was provided by a room temperature mixture of air, water vapor and 2μ water droplets. Figure 7 shows a distinct plume against the warm grey background plate as obtained by the IR camera and enhanced by the VICOM digital image processing system. These results have confirmed the sensitivity of the Thermovision system to low intensity signals.

Planning is being conducted for more quantitative experiments with leak plumes which closely simulate SSME propellant path properties. These experimental results will be used to validate the radiation transport model.

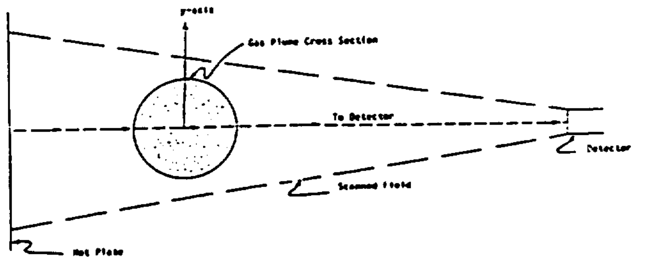


Figure 1. Schematic Illustration of Experimental Set-up

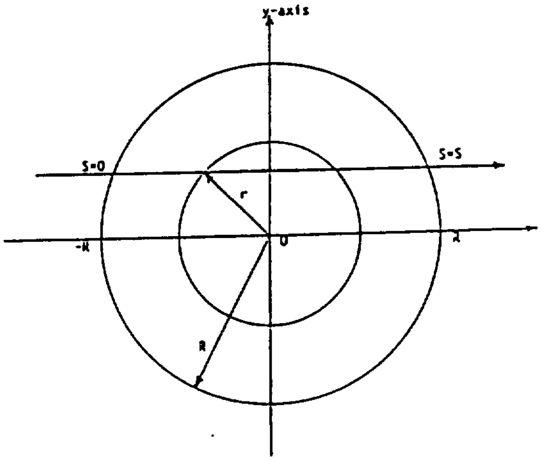


Figure 2. Plume Cross Section

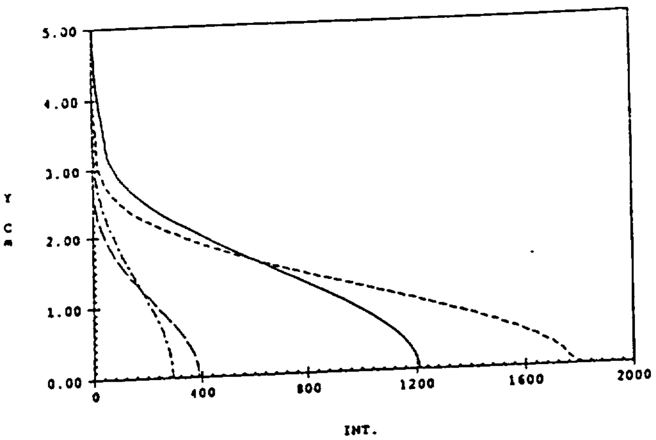


Figure 3. INTENSITY VS. Y for Cold Background Plate With Hot Gas Plume

(Tp=300.0, Tmax=1000.0, Tmin=300.0 K)

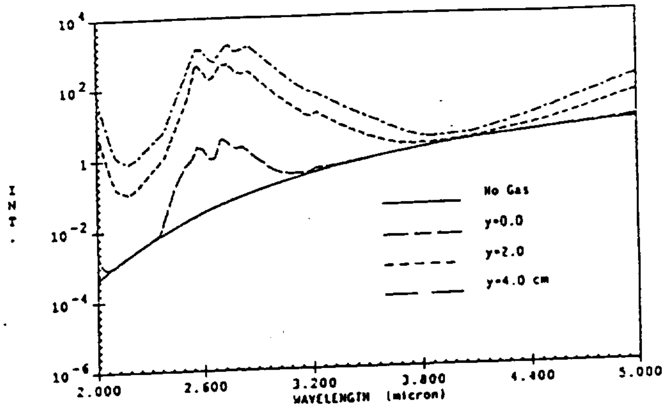


Figure 6. INTENSITY VS. LAMBDA for Cold Background Plate With Hot Gas Plume

(Tp=300.0, Tmax=1000.0, Tmin=300.0 K)

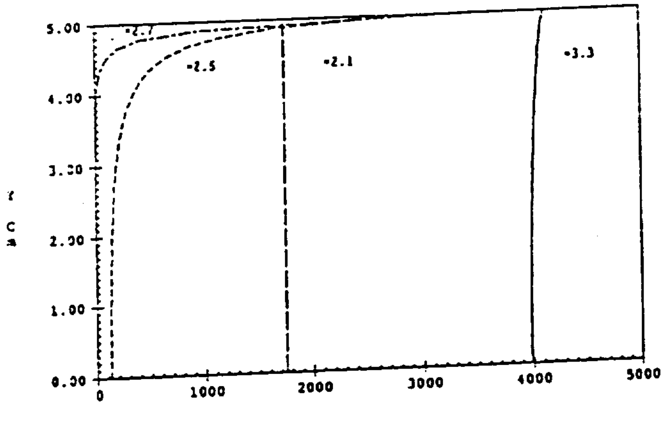


Figure 4. INTENSITY VS. Y for Hot Background Plate With Cold Gas Plume

(Tp=800.0, Tmax=400.0, Tmin=300.0 K)

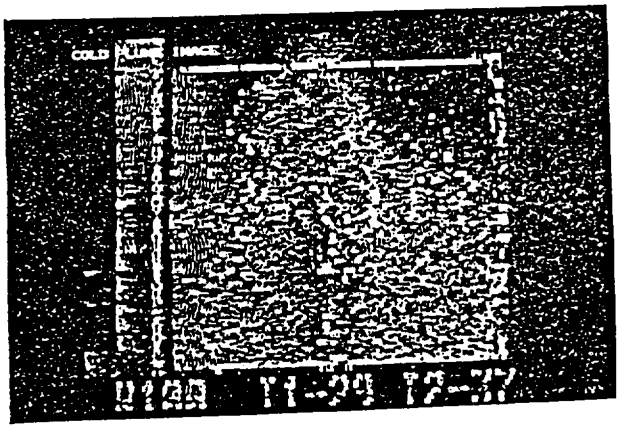


Figure 7. Cold Plume From IR Sensor, Digital Image Enhanced

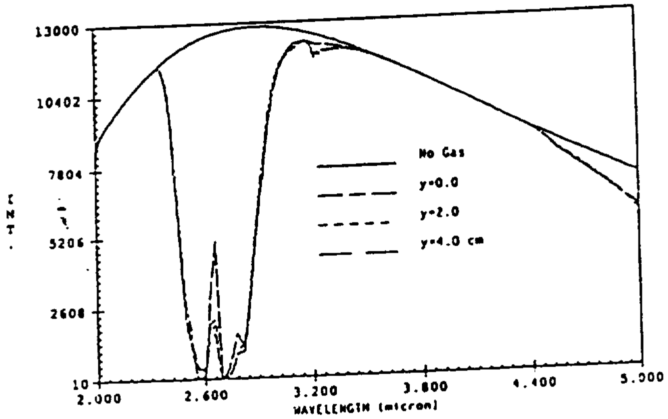


Figure 5. INTENSITY VS. WAVELENGTH for Hot Background Plate With Cold Gas Plume

(Tp=1000.0, Tmax=600.0, Tmin=300.0 K)

Project Title: Intelligent Hypertext Manual Development
for the Space Shuttle Hazardous Gas
Detection System

Principal Investigator: Ching F. Lo

Other Investigator: W. Andes Hoyt

Project Report

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 systems at the Marshall Space Flight Center. The objective is to create a computerized service manual in the form of a hypertext and expert system which stores experts' knowledge and experience. The resulting Intelligent Manual 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 manual frequently as required.

INTRODUCTION

The complexity of space shuttle launch operations is well recognized. Monitoring the shuttle main propulsion system for abnormal gas leakage during the launch phase is one method used to detect subsystem faults. In general, Operation and Maintenance Requirements and Specification Document (OMRSD) information 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 proposed Hazardous Gas Intelligent Hypertext Manual (HIHM) will provide a significantly improved mechanism for the Hazardous Gas Detection System (HGDS). The proposed system will furnish "quick and easy" access to appropriate information for fault detection, data interpretation and action options for the Huntsville Operations Support Center (HOSC). It may display detailed Mass Spectrometer information and Main Propulsion System (MPS) information for the Console Engineer and Manager. This information might include the hardware configuration, the sensor port locations, the gas species "real" concentration, etc.

Additional Application Scenarios. The Intelligent Hypertext Manual (IHM) offers a system for use in training personnel through the large volume of material available for a specific area of space shuttle. The management of documentation for the HGDS will also be greatly simplified through easy maintenance and update capabilities.

Benefits. The Intelligent Hypertext Manual will provide accurate service information in a rapid and efficient manner. It will eliminate most human errors and reduce the required resources including manpower and material cost to achieve the same specified goal. In short, the system will assist to improve productivity and quality of service for fault detection from the HGDS.

INTELLIGENT HYPERTEXT MANUAL TECHNICAL DESCRIPTION

A. Hypertext Manual

The traditional printed service manual restricts material to a linearl arrangement. In searching for a simple procedure, one must constantly make references to other chapters or sections, index entries, footnotes, figures, tables, etc. Often, a text invites the user to skip a section if he is at a different technical competence level. This reference trail among paper documents requires users to exert substantial physical effort and time to gather the complete procedure information. The present proposal offers a solution by converting the linear printed text into a "hypertext/hypermedia" system supported by current computer technology. Hypertext (Ref. 1) is an advanced document management system in which a user can manipulate a database via graphical objects on a computer screen that represent objects inside the computer memory.

The basic feature of the computer-based hypertext is that windows on the screen are associated with objects in a database, and links are provided between these objects graphically as well as in the database. Usually, screen objects are manipulated with a mouse, and several blocks of documentation can be displayed simultaneously in different windows as shown in Fig. 1. For example, specific documents of the Hazardous Gas Detection System for the Space Shuttle could be displayed in a window and at the same time the detailed schematic for the MPS flight measurement locations could be displayed in another window simultaneously. The on-line customized document can be organized to meet users' current requirements. In short, the hypertext manual has the advantages of tracing references easily, of creating new references, of creating customized information display, and modularity of information.

Fortunately, it is not required to design and program a hypertext system from the basic AI programming environment, since several commercial systems are readily available (Ref.1), e.g. KMS (Knowledge Management system Ref. 2) which runs on the SUN workstation and enables users to create a hypertext of a knowledge base. The software package Guide 2 (Ref.3) runs on the IBM Personal Computer or compatible or the Macintosh computer and has a complete development and management system for large volumes of documentation. The Hypertext technology has been given much attention recently. In a special issue of the ACM Journal(Ref. 4), authors presented six papers at HYPERTEXT '87 (University of North Carolina). Also, an in-depth review of Hypertext appeared in BYTE (Ref. 5). In addition, it should be possible to extend hypertext to hypermedia which will enable the user to capture knowledge in a specific domain with graphics and other media. Both Guide 2 and KMS offer graphic display capability.

B. Rule-Based Expert Systems

In order to more effectively utilize the computerized manual a rule-based expert system will be created. For selected systems of the shuttle, experts' experience and knowledge will be stored in the form of rules. Expert systems can typically explain their reasoning and offer a user-friendly interface. A typical consultation session consists of a series of dialogues between the expert system and the user. The user answers questions asked by the system. The expert system will provide the appropriate recommendation, comment, and advice in response to the user's answers via a search algorithm in the system. For example, the user inputs to the computer the gas sample information from the mass spectrometer. The expert system will recommend a corrective action and supporting information if the data indicate a problem, otherwise the expert system will indicate no action is required but may display a comparison with similar data observed in the past. Recommendations for corrective actions will be based on experts' stored opinions and historical data.

The organization and representation of the knowledge base for the expert system will be another key effort in the project. A model will be selected from the available paradigms which include frame-rule representations and objective oriented techniques. Interleaving of the expert advisory knowledge base and the computerized manual will create an Intelligent Hypertext Manual.

C. Intelligent Hypertext Manual

To couple an expert system with a computerized manual is an idea recently proposed independently by Lo & Steinle (Ref 6) and Williamson (Ref 7). This system is contrasted with the traditional expert system which usually does not include a large volume of information. The proposed system will place in memory a large volume of text/graphics documentation, such as, schematics, optional actions, and service procedures from the hypertext database. the architecture for Intelligent hypertext manual is shown in Fig. 2 with the selected softwares.

The effort under 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.

SOFTWARE AND HARDWARE CONSIDERATIONS

A. Hypertext Software Evaluation

The examination of commercial Hypertext programs resulted in the purchase of two programs for evaluation purposes. KMS operates on the Sun 386i workstation. The hypertext system, Guide 2 (Ref. 3), runs on an IBM personal computer.

A demonstration of hypertext systems using KMS and Guide 2 was prepared using the current HGDS IHM project plan as its data base. The demonstration package takes advantage of all the hypertext database features incorporated in KMS and Guide 2, including text and graphics.

KMS offers a programming language for automation of repetitive actions, which is equivalent to macros.

Three features of Guide 2 are judged superior to that of KMS. Guide 2 runs on IBM PC compatible computers which makes hardware acquisition, software installation and portability better than a workstation-based system. Guide 2 is currently installed on a COMPAQ portable computer which is for demonstrations away from the CASP office. Also, Guide 2 allows more than two windows, albeit smaller than a single KMS window, of information to exist on the computer screen simultaneously. Guide 2 provides three types of links; notes, expansion buttons and references expansion to the database as opposed to the single type of KMS link frames. The three links in Guide 2 offer the user greater flexibility than the KMS frames.

KMS offers more powerful features than Guide 2 for formatting and displaying information on the monitor. KMS displays two full pages of information simultaneously as opposed to Guide's several partial pages of information. The KMS screen is a WYSIWYG (what you see is what you get) display that offers better resolution.

Both KMS and Guide 2 offer limited graphic image editing capabilities. With Guide 2 the user can resize images in two dimensions. Guide 2 also allows moving images around the document. The user cuts images to the Microsoft Windows Clipboard for pasting into new locations. KMS gives the hypertext user the same graphics editing features as Guide 2. Additionally, KMS contains drawing tools to include lines, rectangles, polygons, ellipses and arcs in a document. Completely closed figures such as a circle can also be filled with varying levels of shading.

KMS and Guide 2 each provide only one method for importing graphics into a hypertext document. Guide 2 uses the Windows Clipboard program for pasting images from external programs into a document. Therefore the computer system used to build a hypertext database must have programs capable of cutting or pasting to the Clipboard. The Clipboard limits the amount of an image that can be pasted into Guide. A sample A4-page-size image had to be cut to the Clipboard in five individual pieces for pasting into Guide 2. KMS imports images using a single command that is attached to a point in the document. The command inserts the entire image file into the KMS document. The size of the file is not limited to the display. One image tested occupied two full A4 pages.

Although, KMS offers only one type of link, its links can be shared over directories and even networks. The access performance of KMS on the Sun 386i, about one second to fill a screen with new data, exceeds that of Guide 2 running on a 386 PC.

KMS offers the user powerful indexing capability for rapid navigation through the hypertext database. The index feature gives the hypertext manual a definite advantage over the IBM-PC based Guide 2 hypertext. However KMS does not offer pop-up note capability like Guide 2. Advantages of KMS is the ability of the SUN to run several programs simultaneously.

Guide 2 only runs under the Microsoft Windows Environment using DOS 2.0 or higher as an operating system which is subjected to the RAM 640k limit. On the other hand, KMS runs under standard UNIX operation system directly and has no limitations as restricted to DOS operation system.

B. Expert System Software Evaluation

VP-Expert and EXSYS are DOS compatible programs that offer rule based artificial intelligence capabilities. The inference engine in VP-Expert and EXSYS consists of both backward and forward chaining with confidence factors assigned to each knowledge based production rule. The shell is also compatible with spreadsheet, database and ASCII text external programs and is able to exchange data with these files. The executable programs such as filename.exe, filename.com, or filename.bat files can be accessed from the shell. The intelligent knowledge bases are constructed from these external programs in addition to knowledge-based rules acquired from domain experts.

C. Integration of Hypertext and Expert Systems

Calls to/from the hypertext document must be made to/from an expert system for full implementation of the intelligent hypertext manual. Guide 2 offers a "launch" capability in conjunction with command buttons to interface external programs. However, when Guide 2 launches an external program, that program must also be compatible with the Microsoft Windows environment. In one application a reverse process worked successfully using VP Expert and Guide 2. Instead of Guide 2 making an external reference, VP Expert was used to call Guide 2 when the expert system user wanted to search the hypertext database.

KMS, when operated under the SUN View environment can easily interface with a DOS window that contains an expert system. To switch between DOS and KMS the user establishes two windows on the screen and simply points the mouse cursor to the window of choice. Unix, which is the operating system on the SUN, also provides a powerful "piping" capability for passing information from one application to another. The evaluation of the piping feature is now beginning.

D. Database Mass Storage

In anticipation of the large volumes of documentation that will have to be assimilated, the mass-storage technology for text as well as for graphic images is being investigated. Text requires approximately 1 Kbyte of storage per page. For graphics at a low pixel density, the typical file of compressed images requires 15 to 30 Kbytes. Magnetic disk storage may satisfy the requirement for the prototype intelligent Manual. However, it is a key objective of the proposed research to evaluate mass-storage technology, such as Erasable Optical,CD-ROM and WORM, for the future extension of the project. This technology is also required to implement the system on an information server computer. The CD-ROM is a technology similar to digital audio compact disk (CD) used for the storage and retrieval of digitally encoded (machine-readable) information. For mass information storage, CD-ROM is a reliable and permanent medium. The memory capacity of one CD-ROM is 600 Mbytes which is equivalent to 1500 (5.25-inch) floppy disks. One disadvantage is that CD-ROM has longer access time and a slower data transfer rate compared to a magnetic disk. WORM is another technology which is just like CD-ROM but allows the user to "Write Once and Read Many" times. The CD-ROM/WORM technology will be evaluated for storing the large database required for an intelligent hypertext manual.

The new erasable optical disks offer another alternative to storage of large databases. Current first generation systems store 300 megabytes of data per side per disk. Costs range from $4700 to $6500 per unit. This technology is new and each vendor offers unique formats so data is not interchangeable between machines. (Ref 13). Due to its rewrite capability, the erasable optical drive offers significant advantages over both the WORM and the CD-ROM disks.

Disk storage requirements for text and graphical files were established. Text data in ASCII formats occupies 1 bit per character. The different hypertext formats were also evaluated and it was determined that text storage requirements will not be a problem for the Intelligent Hypertext Manual. Three types of graphical files were evaluated for size requirements and included the SUN graphics image format, PC-Paintbrush format and Microsoft Windows format. These files can occupy up to 1 Megabyte of storage for an eight by ten inch figure. The UNIX operating system on the SUN offers a compress program which gave significant reductions in graphic file storage needs. A compress program was found for the IBM-PC files, from Parson's technology for $49.95.

E. Optical Scanners

Requirements. An Intelligent Hypertext Manual will consist of large volumes of textual and graphical information which must be available, on-line, to the user at the HOSC. To meet the goal of demonstrating a hypertext manual of the HIHM, the assimilation of the many volumes of information linked to the HGDS must be accomplished rapidly.

The use of electronic, desktop scanners for input of information to the Space Shuttle HIHM will expedite the task of making required systems information available to the expert system. Scanners alleviate the need for information to be retyped, redrawn or rephotographed. They also assure greater accuracy than manual input through a keyboard and/or mouse.

Selection Criteria. Scanners perform two functions, graphical input and optical character recognition (OCR). The scanner for use with the IIIM must perform both functions with a high level of reliability and accuracy. The manuals for the HGDS contain text, figures and photographs which the user will need in the IHM.

Although the hardware for scanners is similar the quality of scans and the capability of OCR determine the usefulness of particular systems. Scanners accomplish both tasks with software, and usually it is the software that determines the capabilities of a given piece of hardware. OCR uses software that reads type and stores it in ASCII, or other, character formats. Graphics are stored using bit-mapped formats that can be manipulated using paint or similar programs. Most scanners require separate scans for each type of input if a document contains both text and graphics. The scanner must possess the capability to read multiple styles of type for input into the hypertext database.

The scanner may need to handle the bulky and bound books of the Operation and Maintenance Requirements and Specification Document (OMRSD) and Operation, Maintenance and Instruction (OMI) procedures volumes. The sheet-fed scanners require that bound documents be taken apart before input. The compromise between sheet-fed scanners with automatic document feed and scanners that handle bulky documents will become one of time and availability. If extra copies of the documents are available for breaking down into single sheets, a high capacity ADF machine may be most efficient. On the other hand, if the documents must be rebound a flatbed or overhead scanner may be the most efficient, depending upon the speed of the scanner.

SUMMARY OF ACCOMPLISHMENT

The objective of the present Phase I is to establish the feasibility and merit of the proposed system. An intelligent hypertext manual will be created to demonstrate the feasibility of a selected subsystem with a limited knowledge base. This hypertext manual will be demonstrated to NASA personnel at the completion of Phase I. NASA's comments and recommendations will be incorporated into the final design of the hypertext manual. The capability of the expert system's inference mechanism will be illustrated by a few production rules and facts.

As the first step, it was initiated to evaluate hypertext and expert systems software, mass-storage technologies, scanners and host computers. Results of these evaluations are included in this report. A demonstration System was constructed using both KMS and Guide-2 software. The integration of the expert system shell EXSYS into the hypertext systems is in progress.

FUTURE WORK

In the future work, the development of the expert systems for a selected knowledge base will be completed aided by NASA's domain experts. The expert system alone will be demonstrated separately before integrating with the hypertext manual. The major task that will be accomplished during this phase will be the integration of the hypertext manual and expert system into the prototype Intelligent Hypertext Manual. A demonstration of the system to NASA personnel at this stage will be crucial for the implementation of all improvements recommended by NASA. The acquisition of the knowledge base for the HIHM will continue in the selected subsystems. A critical evaluation of HIHM will be performed by NASA in the later part of the work, after assembling and incorporating the improvements recommended by NASA personnel.

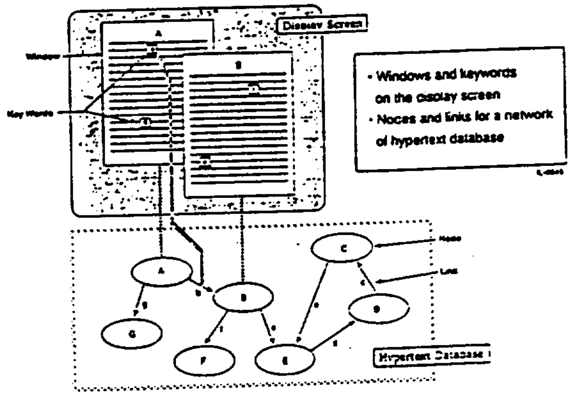
REFERENCES

1. Conklin, J., "Hypertext: An Introduction and Survey," IEEE Computer, September, 1987.
2. KMS Reference Manual, Knowledge Systems, Murrysville, PA. 1988.
3. Guide 2, Professional Hypertext Software, OWL International, Inc., Bellevue, WA, 1988.
4. Special Issue - "Hypertext" - communications of the ACM, July 1988.
5. Special Issue - "Hypertext - In-Depth Review", BYTE, October 1988.
6. Lo, C. F. and Steinle, F. W., "Application of Intelligent Systems to Wind Tunnel Test Facilities", AIAA Paper No. 88-0193, January, 1988.
7. Williamson, M., "On Frames, Semantic Nets, and Hypertext," Computers in Science, January/February 1988.
8. Sickel, V. P., et al., "Documentation Management for Large Systems of Equipment", Alcoa Lab, Alcoa Center, PA 1988.
9. Handy Scanner HS-3000, Diamond Flower Electric Instrument Company, Inc., West Sacramento, CA, 1988.
10. PC Magazine, October, 1987, "Scanners Take Off." Desk top Publisher.

11. Buyer's Handbook, Spring, 1989.

12. PC Magazine, March 28, 1989. "Alternate Input from Fax to Finish Scanners Build a Better Image."

13. "Optical Storage Technology," BYTE, Oct., 1989.



Text Segment Linkage

Figure 1. Hypertext System

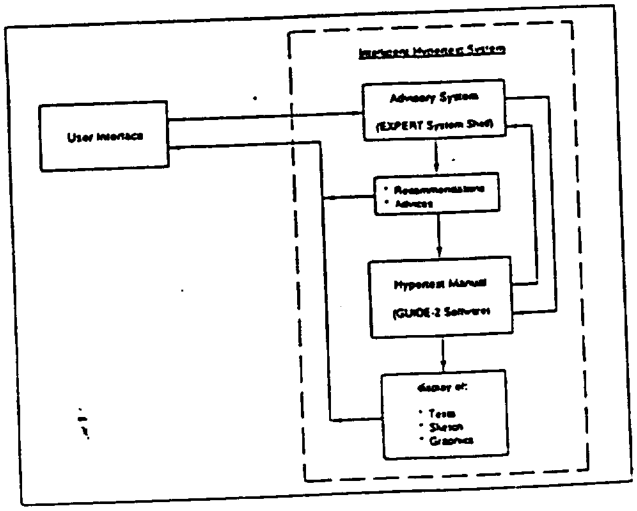


Figure 2. The Conceptual Model of the Implemented Intelligent Hypertext Manual System

Project Title: Classification Of Data Patterns Using An Autoassociative Neural Network Topology

Principal Investigator: M. Ali

Other Investigators: W. E. Dietz, E. L. Kiech

Sponsors: CASP/NASA

Project Report

Abstract

An effort is underway at the University of Tennessee Space Institute to develop diagnostic expert system methodologies based on the analysis of patterns of behavior of physical mechanisms. In this approach, fault diagnosis is conceptualized as the mapping or association of patterns of sensor data to patterns representing fault conditions. Neural networks are being investigated as a means of storing and retrieving fault scenarios. Neural networks offer several powerful features useful in fault diagnosis, including 1) general pattern mapping capabilities, 2) resistance to noisy input data, 3) the ability to be trained by example, and 4) the potential for implementation on parallel computer architectures.

Results of a prototype neural-network-based diagnostic system applied to jet and rocket engines are documented in Refs. 1 and 2. Although the performance of the earlier diagnostic system was quite good, some desirable improvements were identified. The most important improvement deals with how the networks respond to fault classes for which the networks have not been specifically trained. (These fault classes are referred to as "unknown" classes in this paper). The networks used in Refs. 1 and 2 were trained to categorize inputs representing known fault conditions. However, there was no method whereby a network could be trained to recognize when an input pattern corresponded to a previously unobserved, i.e. unknown, fault condition. When presented with sensor patterns representing an unknown behavioral scenario, the networks would sometimes incorrectly categorize the patterns as a known fault condition. As a result, the diagnostic system rarely produced false negative results (i.e., reporting no fault condition when given a set of patterns indicative of a known fault condition), but would sometimes produce false positive results when given input patterns representing a behavioral scenario for which the networks were not specifically trained.

A neural network topology is presented which effectively addresses these problems. The network topology is autoassociative, i.e. the network input and output is identical when properly trained, and hence learning is unsupervised. Classification of input patterns is accomplished by examining the activations of the nodes in the hidden layer. A range of training curves indicative of a particular fault class will result in hidden node activations which map out a surface in N-dimensional space, where N is the number of hidden nodes. An input curve presented to the network for classification will invoke hidden node responses which can be conceptualized as a point in N-space. If the hidden node activation lies near a surface corresponding to a particular fault, the input pattern will be associated with that fault. If, on the other hand, the resultant hidden node activation lies away from any surfaces defined by the training set, the input will be classified as an unknown pattern. Therefore, an input pattern produces a hidden node response which is sufficient to either 1) associate the input with a known fault scenario, or 2) reject the input altogether as not representative of any known fault condition for which the network has been trained. If an input is associated with a fault condition, the hidden node activations will also serve to identify the severity and duration (i.e., rapidity of fault onset) of the fault.

Another improvement to the system reported in Refs. 1 and 2 concerns the number of networks required for diagnosis. The earlier system required five networks for each sensor; one to diagnose the fault type, and two additional networks to identify fault severity and duration for each fault. Two main faults were considered; as a result, the entire system required 20 networks. Adding another fault type to the system's diagnostic capability required the training of twelve additional networks.

With the new approach, a single network can be trained to respond accurately to input curves from all sensors; as a result, a single network performs

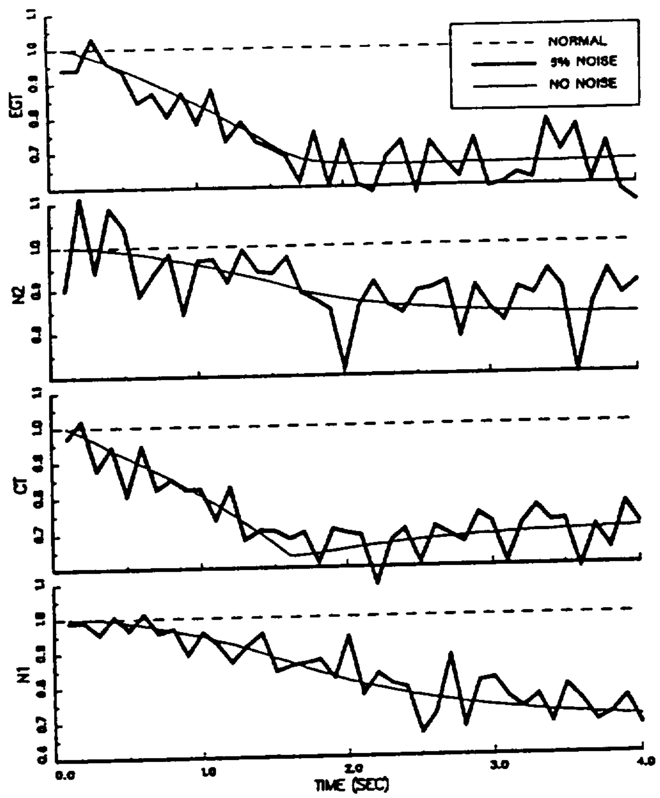


Figure 1. Sensor Outputs For A 75%, 1.5 Second Fuel Interruption

the functions of all 20 networks required in the earlier system. In the present approach, a single network is trained to identify and categorize all fault conditions. As a result, adding the capability of identifying additional faults requires the training of only a single network.

This paper describes 1) the network topology used in the present implementation of the diagnostic system, 2) the training regimen used, and 3) the response of the system to inputs representing both previously observed and unknown fault scenarios. The effects of noise on the integrity of the diagnosis are also evaluated.

Introduction

The University of Tennessee Space Institute (UTSI) is currently involved in the development of advanced methodologies for the detection and diagnosis of faults in aircraft and space systems. Faults in these domains often develop rapidly, requiring diagnostic systems to respond in near real-time. Also, these problem domains require diagnostic systems to exhibit resiliency in the event of noisy data and sensor degradation or failure.

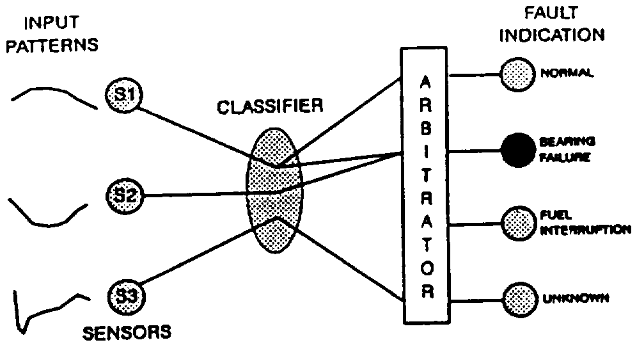


Figure 2. Mapping Of Sensor Patterns To Fault Conditions

Sensor data typical of a jet engine fault are depicted in Fig. 1. Four sensors, exhaust gas temperature (EGT), combustion temperature (CT), and the rotational speeds of the low and high pressure compressors (N1 and N2, respectively) 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 (Ref. 3). 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 depicted in Fig. 1 for comparison. The 5% noise level means that the standard deviation for a Gaussian distribution of perturbations about the noise-free curves is 0.05. A 5% 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.

Many fault diagnosis problems may be conceptualized as the mapping or association of temporal patterns representing sensor data to patterns representing fault conditions. This process is illustrated in Fig. 2, which depicts a mapping process from patterns of incoming sensor data to a pattern of real numbers which can be associated with a particular fault condition. Each sensor inputs a pattern into a classifying system, which attempts to determine which fault condition is indicated by that sensor. In the example of Fig. 2, sensor S1 has been classified as indicative of either normal operation or a bearing failure. This ambiguity may occur if the bearing failure is of a low severity, or has just started to manifest itself in the sensor patterns. Also, sensor S1 may not be highly responsive to bearing failures. Sensor S2 has more definitely indicated a bearing failure, while sensor S3 is unable to make a positive diagnosis of any known failure. The ambiguities are resolved by an arbitrator, which contains domain knowledge about the diagnostic process, such as the sensitivity of certain sensors to different faults. Using this knowledge, the final diagnosis of bearing failure is performed correctly.

The conceptualization of the diagnostic process as a pattern association or mapping process leads naturally to the investigation of pattern mapping and classifying methodologies which are applicable to real-time fault diagnosis. Neural networks exhibit several features useful in diagnostic systems, such as 1) general mapping capabilities, 2) resistance to noisy input data, 3) the ability to be trained by example, and 4) the potential for implementation on parallel computer architectures.

A system based on neural networks to diagnose the development of fault conditions in jet engines was designed and developed at UTSI (Refs. 1 and 2). The system utilized 20 neural networks arranged in a hierarchical fashion which 1) input unprocessed data from four sensors, 2) identified the fault condition, and 3) estimated the severity and duration (i.e., whether the fault was developing slowly or rapidly) of the identified fault. The networks were trained using data from a jet engine simulation program (Refs. 4 and 5). The system exhibited the ability to render accurate and reliable diagnoses at high noise levels, and was able to generalize to correctly identify fault patterns which were not included in the training sets.

Neural network models were also employed in the development of diagnostic techniques applied to the rocket engine domain. The requirement for real-time response in the rocket engine domain is even greater than that of the jet engine domain. The rocket engine domain is characterized by the possibility of very rapid development of fault conditions, with potentially catastrophic results if remedial action is not taken quickly.

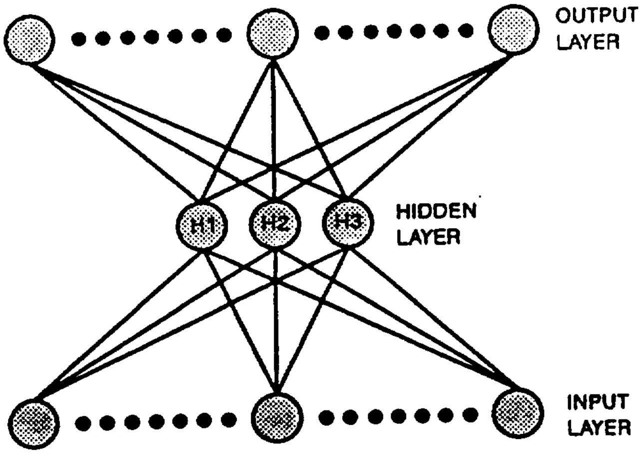


Figure 3. Autoassociative Neural Network Topology

The neural network models used in the prototype engine diagnostic systems were conventional feedforward networks trained using well-known back-propagation algorithms (Ref. 6). Training was accomplished by clamping the input and output nodes to patterns representing sensor data and fault conditions, respectively, and then applying the back-propagation algorithm until the network weights effected a proper mapping between input and output patterns. Although these methodologies were 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 training examples (e.g., a pattern representing a previously unobserved fault condition), the networks can produce spurious categorizations in an unpredictable manner and produce 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. Instead, the diagnostic systems had a tendency to produce a positive categorization corresponding to the fault in the training set which, in some sense, was "closest" to the input. Referring back to Fig. 2, the output node marked "Unknown" cannot be implemented with conventional neural network topologies.

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 were a function not only of the number of training examples required, but also 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 topology and training regimen which will successfully extend the capabilities of feedforward network models.

The Autoassociative Back-Propagation Model

The network topology employed in the present work is autoassociative[1], with at least one layer of semi-linear hidden nodes, as depicted in Fig. 3. Node activations are continuous variables. The autoassociative network contains the same number of input and output nodes and is trained to map or 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 such that the mapping is performed correctly. In the present effort, three layers of nodes are used. The input and output layers contain 40 nodes; the hidden layer contains three nodes. Only the hidden layer nodes are semi-linear; the input and output nodes are linear. Therefore, the activations of nodes in the input and output layer are not restricted to specified ranges.

Training The Network

In the present effort, training the autoassociative model is accomplished in two distinct steps. Suppose, for instance, it is desired to train a network to classify a large number of temporal patterns from engine sensors. In the first step a small representative subset of the total set of curves is selected. This subset of curves is used to obtain a system of weights which will perform the autoassociative mapping, using a back-propagation algorithm.

In the second step, the entire set of available data is input sequentially to the network. Each pattern will invoke a specific hidden node response. When patterns are related, i.e. exhibit similar qualitative features, the hidden node responses will map out a surface. If N hidden nodes are used, the hidden node activations will define a surface in N-dimensional space. For example, Fig. 4 illustrates surfaces generated by training an autoassociative network with three hidden nodes on exhaust gas temperature data from a jet engine. The larger surface is generated from hidden node activations which correspond to a fuel interruption; the smaller surface corresponds to a bearing failure. Both surfaces represent fault conditions over a range of severities and durations. A single point in three-dimensional space represents the hidden node activation resulting from normal (non-transient) operation.

The first step of training, using the back-propagation algorithm, requires most of the total computational effort in the two-step training process. However, the amount of computational effort required using the subset of examples is much less than would be required if all available patterns were used in the training set. For instance, the first step of training for the case in Fig. 4 used a representative set of 24 patterns out of an available 450. In the present effort, failure scenarios of moderate severity were used. It is known that some subsets of data produce better classification performance than others; methods of selecting an optimal subset are presently under investigation.

In the second step of training, the only computational burden is to obtain hidden node activations for every available curve. The second step of training for the case in Fig. 4 utilized all 450 examples, but required a relatively insignificant amount of computational effort compared to the first step. The end result is a network that, in effect, is trained on all 450 patterns, with a total training time far less than would be required to conventionally train a network on all 450 examples using back-propagation alone.

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

Classification Of Input Patterns

Once a network is trained and the hypersurfaces generated, it may be used for classification of input patterns. An input pattern will be mapped into a point in multi-dimensional space. If the point lies on or near a surface associated with a specific fault, the pattern is classified as an example of an input corresponding to that fault. Fig. 4 depicts the point in

---

[1] A variation of the autoassociative network topology has been previously applied to 2-D image analysis and image compression, as reported in Ref. 7.

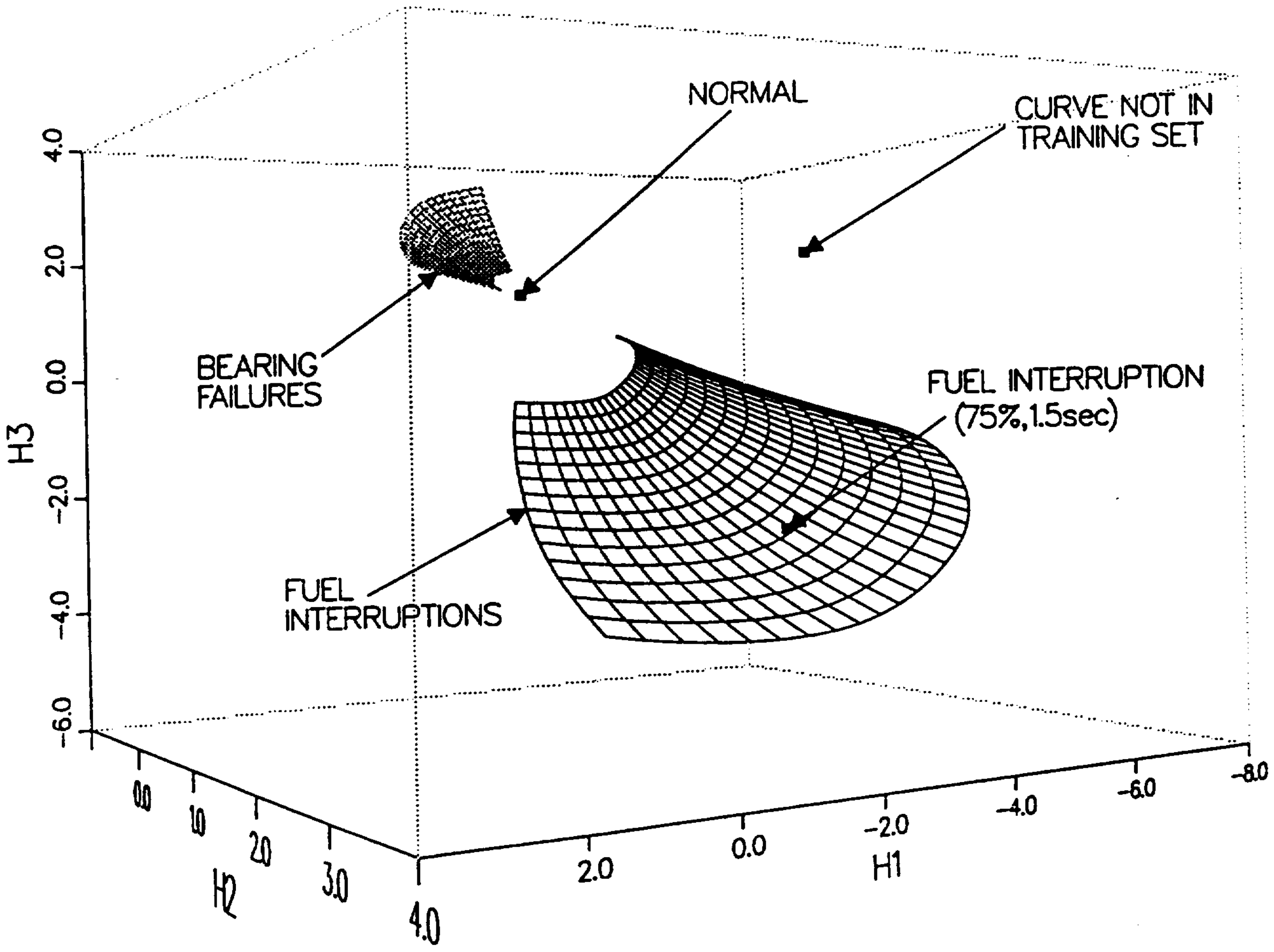


Figure 4. EGT Surfaces

three-dimensional space representing a fuel interruption scenario. If the point lies in a region off the surface (beyond a preset tolerance), the input is classified as corresponding to a type for which the network was not trained, as illustrated in Fig. 4. In this sense a network can determine if it has been properly trained to recognize an input; if the input is not recognized as belonging to a class corresponding to the training set, it can be identified as an instance of another class and processed further, if desired, by other (e.g., knowledge-based) methods.

The position of a point on the surface also may be used to further quantify aspects of an identified input pattern. In the case of the jet engine diagnostic system, each fault pattern is associated with a severity and duration. As a result, each point on a hypersurface is associated not only with a particular fault, but is also associated with a unique severity and duration. Therefore, an input pattern which results in hidden node activations which fall on a surface can be associated with a fault severity and duration. In the example in Fig. 4, the hidden node activation corresponding to a fuel interruption scenario falls on a point on the fuel interruption hypersurface corresponding to a severity and duration of 75% and 1.5 seconds, respectively.

Diagnostic Process

Diagnosis proceeds in two steps. First, it is determined whether a fault condition is occurring, and, if so, which fault scenario is most likely. This step is fault identification. Secondly, the severity and duration of the fault is determined.

Identification of Fault Scenarios

From the standpoint of fault identification, it is desirable if hypersurfaces corresponding to different faults are widely separated. The hypersurfaces depicted in Fig. 4 are widely separated, and therefore can be used reliably to distinguish between fuel interruptions and bearing failures. However, this fortunate circumstance will not always occur. The position of the hypersurfaces is related to

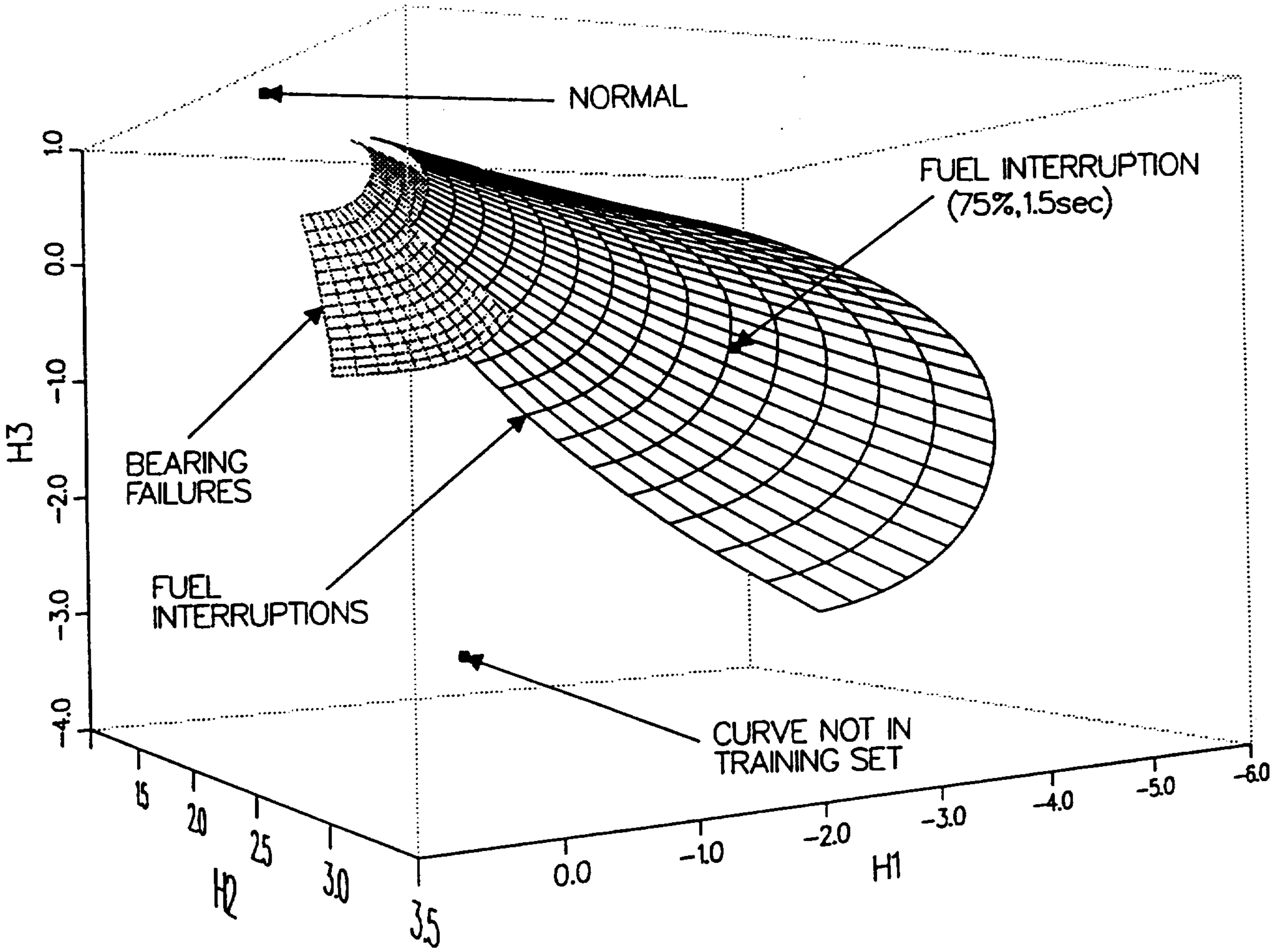


Figure 5. N1 Surfaces

the qualitative and quantitative features of a fault scenario. Different fault scenarios may produce similar responses in a given sensor. For instance, Fig. 5 depicts N1 hypersurfaces for fuel interruption and bearing failure scenarios. The proximity of the two surfaces indicates that, in some cases, the sensor responses to the two faults will be similar. In addition, since the two surfaces intersect, some combinations of fault severity and duration will invoke identical sensor responses. Hence, the usefulness of the N1 sensor in discriminating between bearing failures and low severity fuel interruptions is limited. However, the use of other sensors will almost always remove any ambiguities introduced by the N1 sensor (see Fault Arbitration, below). Moreover, the response of N1 to higher severity fuel interruptions may be used successfully, since the corresponding region of the hypersurface is isolated enough to avoid ambiguities.

Although failure scenarios are represented as hypersurfaces, normal (noise-free) engine behavior is represented as a point in hyperspace. The desirable spatial relationships between the normal point and the hypersurfaces are identical to those between hypersurfaces, i.e., it is desirable for the point representing normal behavior to be separated from the hypersurfaces. In Fig. 4, it is clear that this desirable condition is not met to a sufficient degree. The point representing normal steady-state behavior is quite close to the hypersurface representing bearing failures. This is because bearing failures do not cause severe deviations from the steady-state values for EGT. As a result, EGT is not a reliable sensor to use for discriminating between normal and bearing failure scenarios, especially in high noise environments.

The situation is much better in the case of N1, however, as can be seen in Fig. 5. In this case the point representing normal behavior is almost equidistant from both the fuel interruption and bearing failure hypersurfaces, indicating that N1 is a reliable means of discriminating between normal behavior and fault conditions. Therefore, EGT is best used to discriminate between bearing failure and fuel interruption, and N1 is best used to discriminate between normal behavior and either fault condition. Knowledge of the diagnostic strengths and limitations of each sensor can be used very effectively to resolve situations where, for instance, a particular sensor detects an incoming pattern that has characteristics of more than one fault scenario. This process is performed by the arbitration system (see Fig. 2).

Fault Arbitration

Fault arbitration is essentially a process whereby final fault identifications are made by weighting some sensors more than others, based on the ambiguous fault identifications initially obtained. For example, consider a low-severity bearing failure. The hidden node activations for EGT and N1 will be near the point representing normal engine behavior. In the case of EGT, the hidden node activation will also be close to the hypersurface corresponding to a bearing failure. As a result, EGT will identify the incoming signal as being characteristic of either a bearing failure scenario or normal behavior. The N1 sensor will unambiguously identify the fault as a bearing failure. Overall, the diagnostic system at this point has identified either a bearing failure or normal behavior; a "tie" exists. However, it is known that N1 is quite good at breaking ties between bearing failure and normal behavior; therefore, the results of N1 are weighted more heavily than those of EGT in this case. The final result is an unambiguous identification of a bearing failure.

The weighting would be different in the case of a tie between a fuel interruption and bearing failure scenario. In this instance, it is known that EGT is superior in discriminating between fuel interruptions and bearing failures; as a result the weight associated with the results from EGT would be high and the weight associated with N1 would be low. The final result would be an unambiguous identification based on the results of EGT.

In the present effort, four sensors in the jet engine are monitored. The weights assigned to each sensor in the event of ambiguous initial diagnoses are presented in Table I[2]. As stated earlier, normal/bearing-failure ties are broken by N1. No sensors are suppressed in the event of a tie between fuel interruption and normal behavior scenarios, since all sensors can adequately discriminate between those scenarios. N1, as indicated previously, cannot reliably distinguish between fuel interruptions and bearing failures; in the event of a tie of that nature, all sensors but N1 are used for a final diagnosis.

Table I

Sensor				Dual Fault Indication
N1	CT	N2	EGT	
1.0	1.0	1.0	1.0	Normal & Fuel
1.0	0.0	0.0	0.0	Normal & Bearing
0.0	1.0	1.0	1.0	Fuel & Bearing

Determination of Optimal Radius of Influence

As stated previously, fault identification is accomplished by determining the proximity of a hidden node activation produced by an input pattern to a hypersurface. In the present system, a preset tolerance is used to determine how close a hidden node activation should be to a hypersurface to be considered "on" the hypersurface. In effect, a hidden node activation has an associated "radius of influence"; if the distance to a hypersurface is less than this radius, the hidden node activation is considered to be close enough to be identified with the attributes of the hypersurface.

The ideal radius of influence for the present system was determined by experiment. A large number of fault scenarios were presented to the system, with an associated noise level of 5%. The radius of influence was varied from 0.0 to 2.0, and the resultant percentages of correct fault identifications were obtained, as depicted in Fig. 6. For small radii (e.g., less than 0.1), all input fault patterns were classified as unknown fault classes. This is expected, as it is unlikely that any noisy input pattern will invoke a hidden node activation close enough to a hypersurface to be classified properly. As the radius of influence increases, the percentage of correct fault identifications increases. Correct identifications of fuel interruptions and bearing failures increase to 100% and remain at that level for radius values of greater than ~ 0.6. However, the percent of correct classifications for noisy normal input patterns peaks at a radius of about 0.9. This is because, as the radius increases beyond 0.9, some of the normal input patterns are incorrectly classified as faults. This peak indicates clearly that an optimal radius of influence exists, and that the

---

[2] In the present effort, binary weightings are used; however, fractional weightings are entirely plausible and may be necessary for optimal performance. Optimal weighting assignments are currently being investigated.

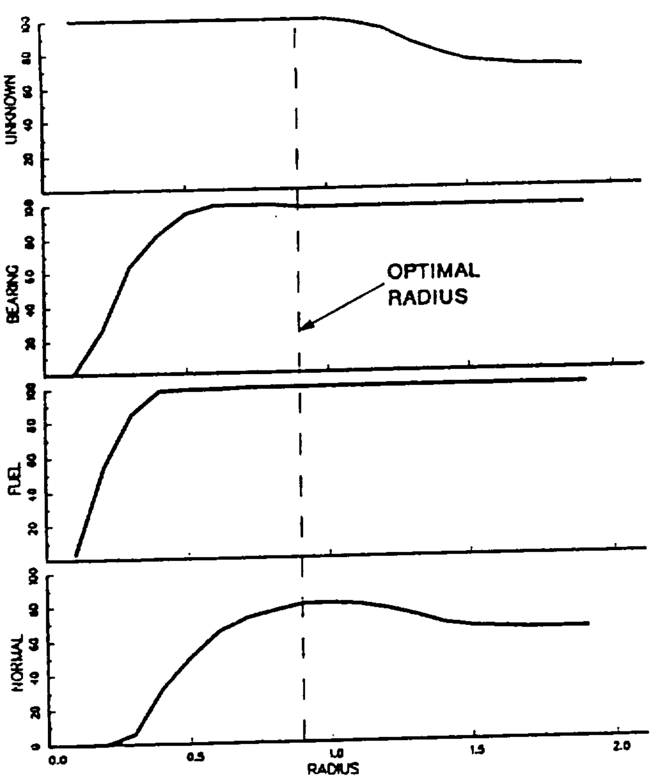


Figure 6. Diagnostic Accuracy vs Radius Of Influence

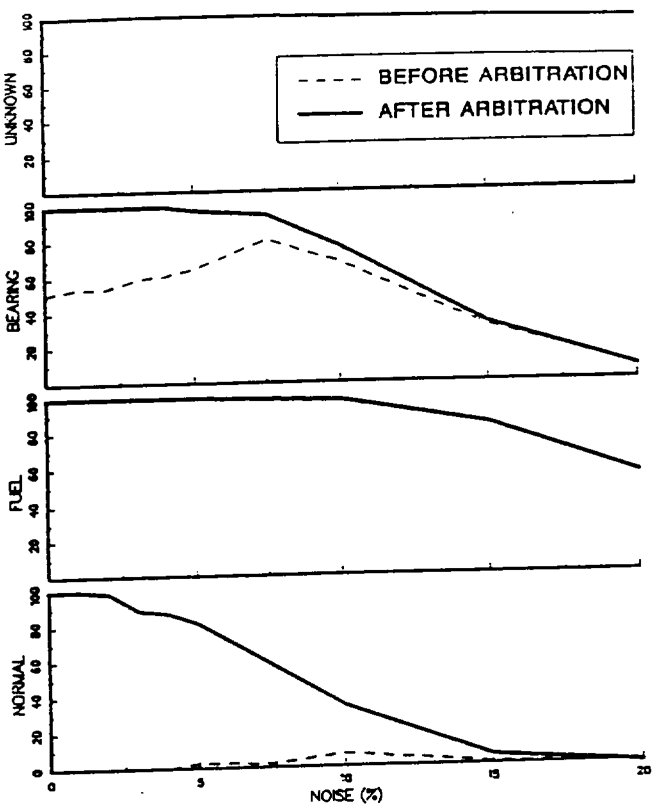


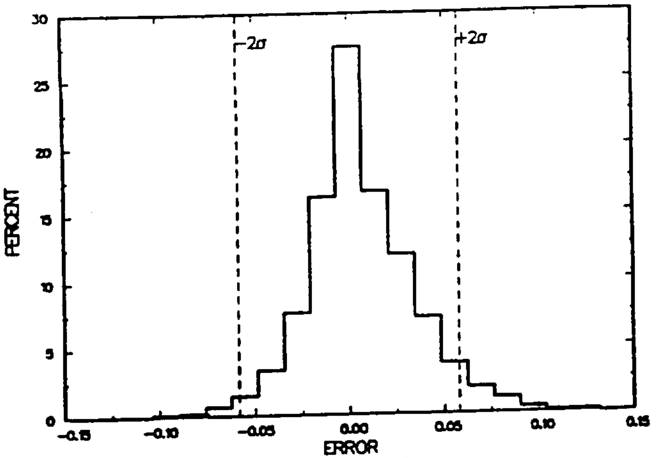
Figure 7. Diagnostic Accuracy vs Noise Level

diagnostic system can in effect be "tuned" to maximize performance for a given noise level.

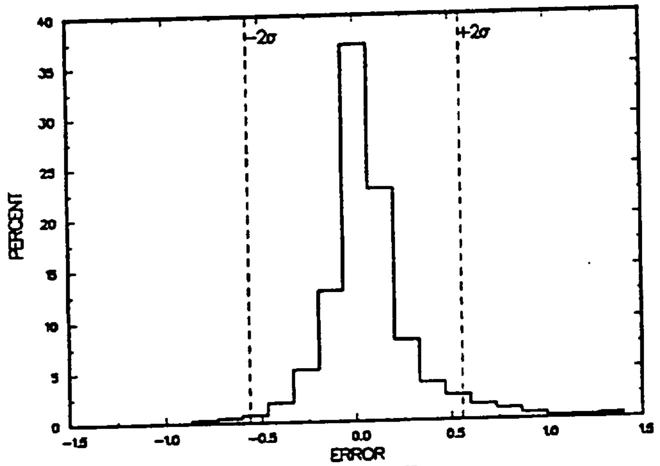
Results

Results of diagnoses for a radius of influence of 0.9 are presented in Fig. 7 for noise levels from 0% to 20%. A large number of normal and failure (fuel interruption, bearing failure, and the unknown bleed pipe rupture) scenarios were processed by the diagnostic system. The illustration depicts the percentage of correctly identified faults before and after arbitration, as a function of noise level. The effects of the arbitration process on the overall fault identification process is pronounced, particularly in the identification of normal input patterns. Initially, all normal scenarios are identified as either normal or bearing failure scenarios. After the arbitration process, normal scenarios are correctly identified up to fairly high noise levels. Bearing failure scenarios initially confused with normal scenarios are, after arbitration, correctly identified at a level of 100% up to nearly 5% noise levels. As expected, fuel interruption identifications are only very slightly affected by noise levels up to 10%. At extremely high noise levels, the hidden node activations of normal and fault scenarios are usually far from any hypersurface, and are therefore classified as unknown patterns. Identification of a fault condition for which the network was not trained (bleed pipe rupture) is constant at nearly 100% for all noise levels. This is also expected, since the addition of noise will usually cause the associated hidden node activations to fall even farther away from the hypersurfaces when compared to the no-noise case.

It is interesting to note that, before arbitration, the bearing failure and normal scenarios are actually correctly identified more often under noisy conditions. This is because random noise fluctuations may produce hidden node activations which are less ambiguous than those produced by the noise-free conditions. For instance, a noisy normal behavioral pattern may produce a hidden node activation which is farther away from the bearing failure hypersurface than that produced in a noise-free environment. In this case, there will be no confusion between normal behavior and a bearing failure scenario. A similar process occurs with true bearing failure scenarios; noise may move a hidden node activation away from the point designating normal behavior, thereby removing any ambiguity. However, arbitration significantly improves results both in noisy and noise-free environments.

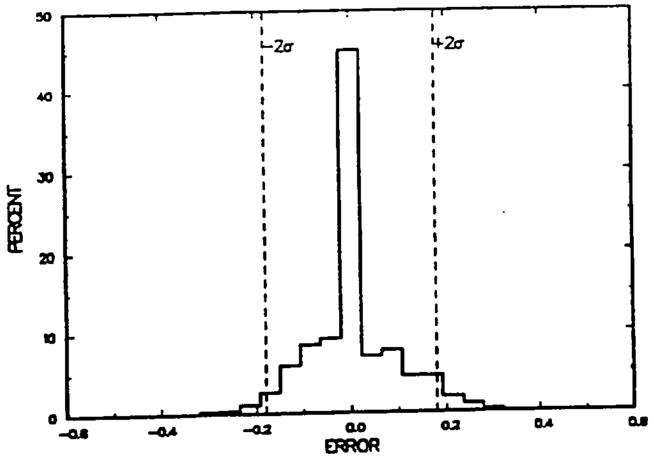


(a) Severity Error

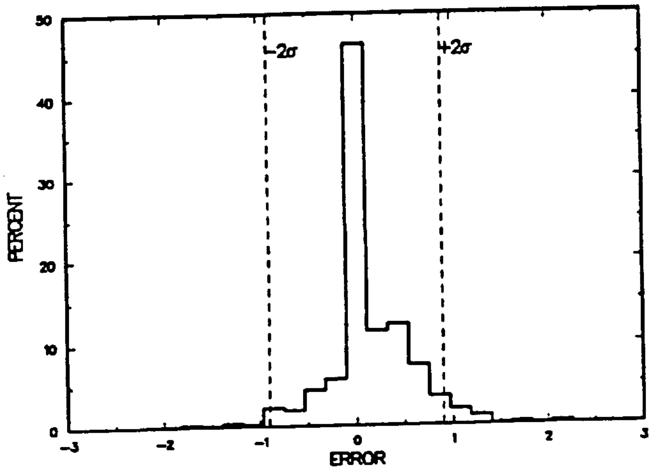


(b) Duration Error

Figure 8. Error Distribution For Fuel Interruption Severity (a) and Duration (b)



(a) Severity Error



(b) Duration Error

Figure 9. Error Distribution For Bearing Failure Severity (a) and Duration (b)

Determination of Fault Severity and Duration

Once a fault has been identified, the hypersurfaces may be used to estimate the severity and duration of the fault. Each point on a hypersurface is associated with a unique fault severity and duration. During the fault identification process, a hidden node activation corresponding to the input pattern will be generated. To estimate fault severity and duration, the closest point on the hypersurface to the point representing the hidden node activation may be found; the closest point will correspond to a unique fault severity and duration.

Fault identification is mostly affected by the separation of the hypersurfaces from each other and from the point representing normal steady-state behavior. Estimates of fault severity and duration are affected primarily by the size of the hypersurface representing the identified fault. This is because, in noisy environments, the perturbations of the hidden node activations will be relatively more severe for small hypersurfaces than large ones. Figs. 8 a) and b) depict histograms of fuel interruption severities (expressed as decimals) and durations, respectively, for a 5% noise level. The standard deviation (σ) for severities is about 0.03, which means that approximately 99% of estimated severities will be within 0.06 (2σ) of the true values. The corresponding standard deviation for fuel interruption durations is 0.3 seconds. As can be seen in Figs. 4 and 5, the hypersurfaces for fuel interruption scenarios are quite large. As a result, good estimates of fuel interruption severity and duration may be obtained. However, the hypersurfaces associated with bearing failures are relatively small. This has a detrimental impact on the ability to accurately estimate the severity and duration of bearing failures. As can be seen in Figs. 9 a) and b), the 2σ limits for bearing failure severity and duration are 0.18 and 0.9 seconds, respectively.

Conclusions

A diagnostic system based on neural networks has been developed and applied to the real-time diagnosis of jet and rocket engines. The approach used in the present work addresses deficiencies inherent in many feedforward neural network models. The present system greatly reduces the number of false positive diagnoses that can be produced when applying standard feedforward networks to pattern classification. In addition, the present approach greatly reduces the number of networks necessary to 1) identify the existence of a fault condition, and 2) estimate the duration and severity of the identified fault. The present method also incorporates a two-stage training approach which greatly reduces the overall training time.

Acknowledgement

This work was performed within the Center for Advanced Space Propulsion (CASP) and was supported in part by NASA Grant NAGW-1195. 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. Dietz, W.E., Kiech, E.L., and Ali, M., "Neural Network Models Applied to Real-Time Fault Diagnosis", Journal of Neural Network Computing, Vol. 1, No. 1, 1989.

2. 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, The University of Tennessee Space Institute, Tullahoma, TN., pp 13-23, June 1-3, 1988.

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

4. Fishbach, L.H., "Computer Simulation of Engine Systems", AIAA-80-0051, 1980.

5. Chappell, M.A., "Advanced Turbine Engine Simulation Technique", AEDC-TR-86-19, 1986.

6. 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.

7. 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.

Project Title: SSME Component Assembly And Life Management Expert System
Principal Investigator: M. Ali
Other Invesitgators: W. E. Dietz, H. J. Ferber
Sponsors: CASP/Rocketdyne

Project Report

Abstract

The Space Shuttle utilizes several rocket engine systems, all of which must function with a high degree of reliability for successful mission completion. The Space Shuttle Main Engine (SSME) is by far the most complex of the rocket engine systems. In earlier spacecraft, rocket systems (and, in fact, the entire spacecraft) were designed for use on only a single mission and were discarded after use. In a major departure from earlier practices, almost all systems on the Space Shuttle, including the rocket systems, are designed to be reusable; only the external tank is discarded during each mission.

The reusability of spacecraft systems introduces many problems related to testing, reliability, and logistics. Components must be assembled from parts inventories in a manner which will most effectively utilize the available parts. Assembly must be scheduled to efficiently utilize available assembly benches while still maintaining flight schedules. Assembled components must be assigned to as many contiguous flights as possible, to minimize component changes. Each component must undergo a rigorous testing program prior to flight. In addition, testing and assembly of flight engines and components must be done in conjunction with the assembly and testing of developmental engines and components.

The development, testing, manufacture, and flight assignments of the engine fleet involves the satisfaction of many logistical and operational requirements, subject to many constraints. The purpose of the SSME Component Assembly and Life Management Expert System (CALMES) is to assist the engine assembly and scheduling process, and to insure that these activities utilize available resources as efficiently as possible.

PROBLEM DESCRIPTION

Fig. 1 illustrates the general problem CALMES is designed to address. Typically, a priori knowledge consists only of parts inventories and a proposed flight schedule. Using this input, CALMES performs four basic functions: 1) determining how components may be optimally assembled from the parts inventory, 2) assigning components to flights, 3) scheduling component testing, 4) scheduling component assembly. The first function deals exclusively with component assembly, while the remaining functions are considered to be scheduling operations. The order in which these functions are performed is indicated by the direction of the arrows in Fig. 1. The four major functions are explained in more detail below:

Part-Component Assignment

For the purposes of the present discussion, the SSME is considered to be composed of 1) major components, and 2) minor components (or parts). Major components include the four turbopumps, i.e., the high pressure fuel and oxidizer turbopumps (HPFTP and HPOTP, respectively) and the low pressure fuel and oxidizer turbopumps (LPFTP and LPOTP, respectively). Minor components or parts comprise the constituent parts of the major components, and cannot be subdivided further. Minor components or parts include turbine blades, impellers, impeller shafts, housings, and turbine discs. Both major and minor components are rated according to remaining allowable life, and remaining al-

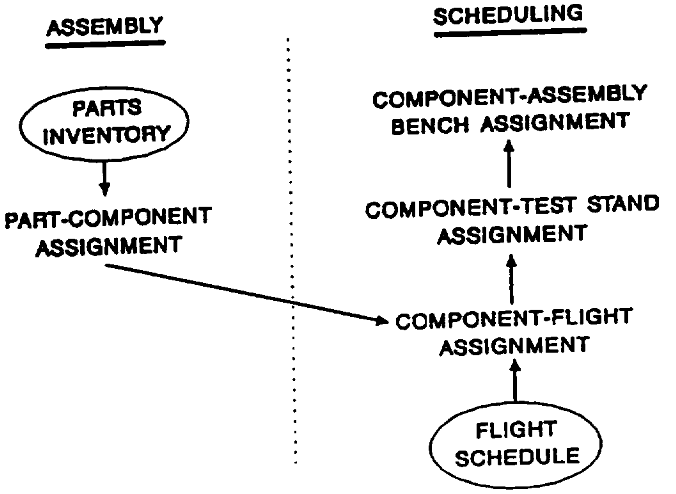


Figure 1. CALMES Input And Main Functions

lowable starts. If, for example, the HPOTP has 800 seconds of remaining allowable life, and two remaining allowable starts, it may be used once for 800 seconds, or twice for a total of 800 seconds.

It is relatively easy to calculate allowable lifetimes and number of starts for minor components. Each minor component is assigned an allowable life and number of starts at the time of manufacture; these estimates are based on engineering design considerations and extensive testing. During the life of a part, the length of time the part is used and the associated number of starts are simply subtracted from the initial values each time the part is used.

The allowable life and number of starts of a major component, however, are determined by the smallest allowable lifetime or number of starts of its parts. Therefore, it is critical that major components be assembled from parts with sufficient remaining life and number of starts to accommodate anticipated test and flight requirements. Generally, there will be many ways in which available parts may be combined into major components. A single turbopump consisting of five parts, where each part comes from an inventory of 60 parts, may be assembled in almost a billion different ways. However, many of the possible component configurations will not satisfy the required number of starts or allowable lifetimes, based on anticipated flight schedules and durations. It is the task of the present system to assemble a set of major components which efficiently utilizes the available resources from a large parts inventory.

Component-Flight Assignment

Each flight of the shuttle requires three engines which have remaining allowable lifetimes greater than the longest anticipated shuttle flight (i.e., 770 seconds, which would be required in an abort situation). Of course, each engine, and therefore each engine component, must have at least one allowable start. As in the case of major components, the remaining life and number of starts of engines are determined by the smallest remaining life or number of starts of its components. Generally, it is desirable to assign an engine or component to several contiguous flights to minimize engine changes. An engine spanning more than one flight requires longer allowable lifetimes and multiple allowable starts. The task of CALMES is to assign engines and major components to flights in an efficient manner. Generally, this means that 1) engines and components will be assigned to as many contiguous flights as possible, and 2) the remaining life of an engine or component after its use on flights will be smaller than that of any flight remaining on the schedule.

Component-Test Stand Scheduling

There are presently several types of engine ground tests, including 1) development, 2) certification, 3) flight certification extension, 4) acceptance, and 5) component "green runs". Development tests are performed on hardware which has not been certified for flight. Certification tests are performed on hardware which has matured to a point where use on a flight is desired. Flight certification extension tests are performed on flight certified hardware to extend the certified hardware lifetime or number of allowable starts. Acceptance tests are performed on entire engines prior to installation on a shuttle orbiter. Green run tests are designed to identify any problems in a component in a controlled environment before it is installed in an engine which will be used on a manned flight. Usually, green runs consist of three engine firings of 1.5, 250, and 520 seconds over the course of one week. Other types of tests may range from one to several weeks, and may involve a wide range of engine firing times and starts.

CALMES is designed to assign test stands to components or engines to fulfill test requirements. The assignment process attempts to 1) schedule an available test stand within a specified time range, and 2) to preferentially assign a particular test stand usually used for particular types of tests. The former requirement is more important than the latter, and therefore takes precedence.

Component-Assembly Bench Scheduling

The final function of CALMES is to assign assembly benches to components which are required for ground tests. Presently, there are six available assembly benches. Components may take several weeks to assemble, depending on the type of component and the condition of its parts. CALMES is designed to assign assembly benches to either 1) utilize as many assembly benches as possible, and hence minimize the overall time required for the assembly of all components, or 2) use as few assembly benches as possible, but assembling the components as late as possible.

Simplifying Assumptions

The approach used in the prototype has been designed to be able to accommodate the general problem outlined above. However, the following simplifying assumptions have been made for the purpose of verifying the approach used in CALMES.

1) A single type of major component (high-pressure oxidizer turbopumps, or HPOTPs) is considered. More major components will be considered in the operational program.

2) HPOTPs are assumed to consist of five life-limiting parts, i.e. blades, impeller, housing, impeller shaft, and turbine disc. Additional parts may be considered in the operational program.

3) A single launch vehicle is assumed. The operational program will consider three launch vehicles.

4) Only green run tests are scheduled. Engine certification and development tests, etc. are not presently incorporated into test schedules.

Problem Formulation for the Present System

The problem subset being addressed by the expert system prototype may be summarized by the following processes:

1) Given an inventory of parts (e.g., turbine blades, etc.), assemble a virtual fleet of major components (e.g., turbopumps). A virtual fleet is a set of pumps that could be built if desired; however, only pumps actually required by the flight schedule will be assembled.

2) Assign components to shuttle flights. This step determines which components from the virtual fleet are actually required.

3) Schedule green run ground tests on components required for flight.

4) Schedule assembly benches on components required for ground tests.

Hard and Soft Constraints

All of the above steps must be performed subject to many constraints. For example, the pumps must be assembled from parts which will result in allowable lifetimes sufficient to accommodate the desired flight schedule. Resources such as test stands and assembly benches must be scheduled to meet flight deadlines, testing requirements, and time restrictions. Also, engines must be assigned to flight vehicles in a manner which minimizes engine changes over the proposed flight schedule.

Constraints on the assembly and scheduling processes may be categorized as being either hard or soft. Hard constraints are inviolable; an example is a requirement of three months between the completion of a test on a pump and the use of the pump in a flight engine. Soft constraints are preferences which may be relaxed, if necessary, to accommodate hard constraints. An example of a soft constraint is the preferred use of a particular test stand for certain types of tests, or the use of as few assembly benches as possible during assembly of a major component. The approach used in the present work allows the incorporation of both hard and soft constraints into the assembly and scheduling process. In addition, soft constraints may be given relative degrees of importance.

General Approach

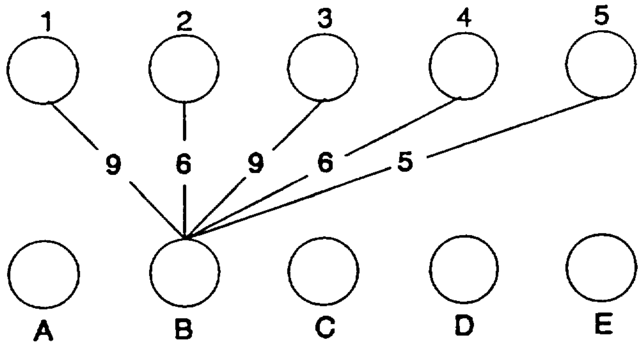
All of the major functions of CALMES have one aspect in common: they all require assignments of one object or entity to another. For instance, the part-component assignment problem assigns parts from an inventory to instances of components. Component-test stand assignment requires the assignment of components to available test slots on particular test stands. The four major functions of CALMES are each cast as instances of the formalized assignment problem, a familiar problem in operations research.

The assignment problem is illustrated in Figs. 2 and 3. Fig. 2 a) depicts a bipartite graph, i.e., a graph which consists of two distinct sets of nodes. The two sets of nodes are connected by weighted links. The links between node B and the nodes of the other side of the graphs are illustrated in Fig. 2 a). Each link is associated with a weight, which in the present implementation is a positive real number. Links between the other nodes are omitted for clarity.

The system of weights can be conveniently represented in matrix form, as depicted in Fig. 2 b). An optimal match is obtained when 1) each node on one side of the graph is matched with a single node on the other side of the graph, and 2) the sum of the weights (the cost) is a minimum. The circled weights in Fig. 2 b) illustrate the optimal matching for this graph. Note that each link assigned to the optimal match occupies a unique row and column of the weight matrix.

Optimal Matching

The process of finding an optimal match is illustrated in Fig. 3. The first step of the matching process is to select, as an initial matching, all pairings represented by the smallest weights in the columns of the matrix. This process results in the pairings A2, C1, and D5, as depicted in Fig. 3 a). Nodes 3 and 4 cannot be matched to nodes on the other side of the graph, because their minimum weighted links conflict with matchings A2 and C1,



a) BIPARTITE GRAPH

$$\begin{array}{c|ccccc}
 & 1 & 2 & 3 & 4 & 5 \\
\hline
A & 7 & 2 & \textcircled{1} & 9 & 4 \\
B & 9 & 6 & 9 & 6 & \textcircled{5} \\
C & \textcircled{3} & 8 & 3 & 1 & 8 \\
D & 7 & 9 & 4 & \textcircled{2} & 2 \\
E & 8 & \textcircled{4} & 7 & 4 & 8
\end{array}$$

b) MATRIX REPRESENTATION

Figure 2. Optimal Matching Of Bipartite Graph

respectively. If nodes 3 and 4 are matched arbitrarily with nodes B or E, the lowest sum of weights possible is 20.

At this point, three nodes on both sides of the graph have been temporarily committed (i.e., they have been paired with nodes on the other side of the graph). The matching is now improved by introducing the concept of augmenting paths. An augmenting path through the graph consists of alternating uncommitted and committed segments, where the first and last segments are uncommitted. Fig. 4 shows an augmenting path in Fig. 4 a). At first, there is only one committed segment (i.e., A2). If the committed segments are made uncommitted, and vice versa, the path will have two committed segments and only one uncommitted segment, as in Fig. 4 b). An augmenting path will always increase the number of pairings by one. In effect, augmenting paths "undo" initial, inferior pairings and replace them with pairings which are closer to the optimal match. After a single augmentation, depicted in Fig. 3 b), the lowest possible cost has been reduced from 20 to 16. After two augmentations, illustrated in Fig. 3 c), the cost has been reduced further to 15. At this point, all possible pairings have been made, and the optimal match has been attained.

The choice of augmenting paths is obviously of primary importance. Many methods have been studied; CALMES uses a modified simplex method to search for augmenting paths and to obtain globally optimal matchings. Discussion of the matching algorithm, sometimes referred to as the Hungarian method, is beyond the scope of this paper; however, the basic approach is fully documented in Refs. 1 and 2.

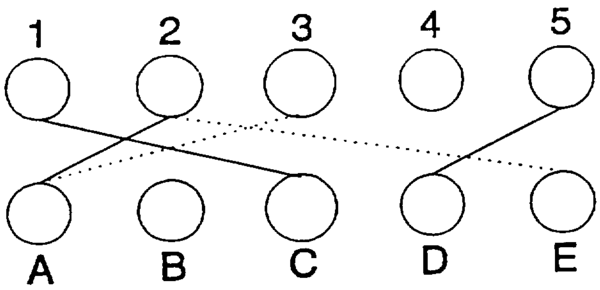
Results

Assignment of Parts to Pumps

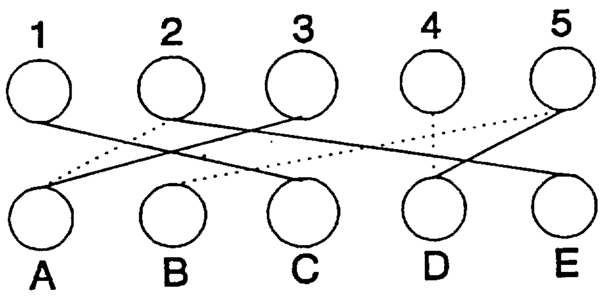
The first function CALMES performs is the assembly of a fleet of pumps from an inventory of parts. In the present simplified implementation, only high pressure oxidizer turbopumps are considered; in addition, only five different parts are assumed: turbine blades, impeller shaft, housing, turbine disc, and impeller. Any number of these parts can be in the inventory, each with associated allowable lifetimes and starts. The task of CALMES is to "build" as many pumps from the available inventory, maximizing either the allowable lifetime or the number of starts of the fully assembled fleet. The resultant set of pumps is "virtual", in that any of the pumps may possibly be actually built; however, CALMES does not at this stage determine whether or not a particular pump should actually be assembled.

The virtual assembly is initiated by optimally matching, for instance, the turbine blade inventory with the inventory of impellers. This is accomplished by associating nodes on one side of the bipartite graph with instances of turbine blades in the inventory, and associating nodes on the other of the graph with instances of impellers. A matching between the two sides of the bipartite graph will in effect generate a set of partially assembled virtual pumps.

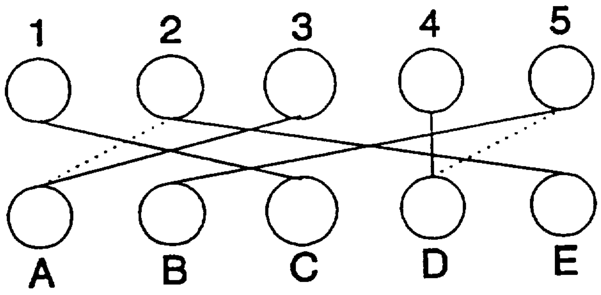
The determination of the weights between the nodes is accomplished by applying appropriate heuristics. The resultant lifetime of any blade-impeller matching will be the lower of the blade lifetime and the impeller lifetime. Therefore, it



a) INITIAL MATCHING (COST = 20)



b) ONE AUGMENTATION (COST = 16)

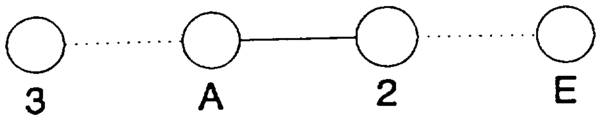


c) TWO AUGMENTATIONS (COST = 15)

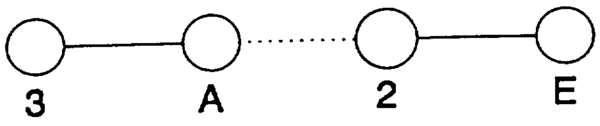
Figure 3. Development Of Optimal Matching

would not be productive to match a blade with a high lifetime with an impeller of a low lifetime, providing a blade with a lifetime close to that of the impeller is available in the inventory. Therefore, the weights between blades and impellers are made proportional to the absolute value of the difference of the blade and impeller lifetimes. As a result, a blade and impeller with equal lifetimes will most likely be matched; a match between blades and impellers with greatly differing lifetimes would be unlikely, but of course would not be impossible if such a matching resulted in a globally optimal matching.

Once a virtual fleet of partially assembled pumps is obtained, the matching process is repeated between the partially assembled pumps and another parts inventory. In the present system, the fleet of partially assembled pumps will be matched



a) BEFORE PATH AUGMENTATION



b) AFTER PATH AUGMENTATION

Figure 4. Path Augmentation

with turbine housings. Again, the weights between the partially assembled pumps and housings are made proportional to the difference in the respective allowable lifetimes. The second matching process results in another virtual fleet of partially assembled pumps; however, the pumps now have three parts: blades, impellers, and housings.

Assignment of the partially assembled pumps with the two remaining parts is accomplished with two more matching steps. Weights are determined on the basis of allowable lifetimes of the partially assembled pump inventory and the inventories of turbine discs and impellers. The final result is a virtual, completely assembled pump fleet.

Abbreviated output from CALMES illustrating the part-component assignment process is reproduced in Figs. 5-7. Fig. 5 shows the assignment of blade sets to individual pumps. The pumps are designated by alphabetic characters. Since both upper and lower cases are used, 52 pump instances can be designated. The first entry states that the second blade set (serial number 1000002, allowable life of 1723 seconds, 9 allowable starts) is assigned to pump "T". Assignments for the remaining four parts are listed by CALMES in a similar manner, but are not reproduced here.

At the bottom of Fig. 5 are blade sets which are not assigned to any pump. Parts with the highest lifetimes will generally not be assigned, because the only possible assignments would group them with other parts with lower lifetimes. Such an assignment would not increase the overall lifetime of the

```
              PART-PUMP ASSIGNMENTS
                     BLADES
PART  NAME      S/N     LIFE   STARTS  PUMP
----  ------  -------   ----   ------  ----
  2   BLADES  1000002   1723      9     T
  3   BLADES  1000003   1020      3     c
  7   BLADES  1000007   1951     13     M
  9   BLADES  1000009   2003     12     I
 10   BLADES  1000010    773      7     h
  .     .        .        .       .     .
  .     .        .        .       .     .
  .     .        .        .       .     .
 81   BLADES  1000081    509     15     p
 82   BLADES  1000082   1975     15     K
 87   BLADES  1000087   1791      1     S
 88   BLADES  1000088   1549      1     U
 89   BLADES  1000089   1929      2     N

              UNASSIGNED PARTS
PART  NAME      S/N     LIFE   STARTS
----  ------  -------   ----   ------
  1   BLADES  1000001   1317     10
  4   BLADES  1000004    988     13
  5   BLADES  1000005   2187      3
  6   BLADES  1000006   2298      5
  8   BLADES  1000008   2082     11
  .     .        .        .       .
  .     .        .        .       .
  .     .        .        .       .
 92   BLADES  1000092   1494      1
 93   BLADES  1000093    858     11
 94   BLADES  1000094   2409     10
 95   BLADES  1000095    350      8
```

Figure 5. Assignment Of Parts (Blade Sets) To Pumps

virtual fleet, and would thereby be a suboptimal assignment. Parts with low remaining lifetimes will not be assigned, because their assignment would lower the overall fleet lifetime.

Parts are assigned to pumps in a manner which ensures an optimal pump fleet, in the sense that no rearrangement of parts will result in a mix of pumps with greater allowable lifetimes. The assignment algorithm allows pump fleets to be optimized on the basis of life, starts, or an arbitrary combination of desirable factors.

Fig. 6 is a table summarizing the virtual pump fleet resulting from the parts assignments. Forty-seven pumps can be assembled from the parts inventory, with allowable lifetimes from 2324 seconds to 79 seconds. As stated above, this is the best possible fleet of pumps which can be obtained from the given parts inventory. Fig. 7 is a table which details which parts (by serial number) are assigned to each pump.

As stated earlier, the results of Fig. 5-7 indicate a virtual fleet of pumps which can, if required, be assembled from the available parts. Which pumps should be actually assembled will depend on the requirements of the flight schedule.

```
                     VIRTUAL PUMP FLEET
INSTANCE  PUMP     U/N     LIFE  STARTS  DESIGNATION
--------  -----  -------   ----  ------  -----------
    1     HPOTP  2000059   2324     1         A
    2     HPOTP  2000027   2311     4         B
    3     HPOTP  2000021   2259     3         C
    4     HPOTP  2000064   2229     2         D
    5     HPOTP  2000091   2203     1         E
    .       .       .        .      .         .
    .       .       .        .      .         .
    .       .       .        .      .         .
   43     HPOTP  2000044    401     3         q
   44     HPOTP  2000009    398     2         r
   45     HPOTP  2000046    389     3         s
   46     HPOTP  2000031    345     3         t
   47     HPOTP  2000094     79     1         u
```

Figure 6. Optimal Pump Inventory

```
                   PUMP-PART ASSIGNMENTS
                                         TURBINE    MAIN
PUMP  BLADES   HOUSING   SHAFT    DISC     IMPELLER
----  -------  -------  -------  -------  --------
 A    1000080  2000059  3000018  4000042  5000014
 B    1000046  2000027  3000025  4000021  5000037
 C    1000074  2000021  3000032  4000054  5000012
 D    1000058  2000064  3000013  4000058  5000030
 E    1000061  2000091  3000038  4000078  5000009
 .       .        .        .        .        .
 .       .        .        .        .        .
 .       .        .        .        .        .
 q    1000076  2000044  3000005  4000015  5000043
 r    1000055  2000009  3000026  4000037  5000015
 s    1000031  2000046  3000017  4000048  5000028
 t    1000036  2000031  3000053  4000064  5000038
```

Figure 7. Parts-To-Pump Assignment Table

Assignment of Pumps to Flights

Once a virtual pump fleet has been assembled, the pumps must be assigned to engines which are scheduled to fly on the shuttle. At this point, knowledge of the proposed flight schedule is necessary. Under normal conditions, each flight requires the engines to fire approximately 550 seconds. To illustrate the flexibility of the present approach, a fictitious flight schedule with a range of engine firing requirements from 500 to 770 seconds is depicted in Fig. 8 a). (In real life, each flight would be assumed to require 770 seconds of engine firing time, to account for the possibility of worst-case flight abort scenarios).

Pumps may, if remaining lifetime and starts permit, be used on multiple flights. Obviously, it is advantageous for pumps to be used on multiple flights which are contiguous, to minimize pump changes. Therefore, use of a pump on multiple flights is a soft constraint, i.e., it is a preferred feature in the assignments of pumps to flights, but it is not absolutely necessary for an assignment scenario to be acceptable.

FLIGHT	WEEK	DURATION
1	55	500
2	56	550
3	57	650
4	58	555
5	59	600
6	60	500
7	61	665
8	62	550
9	63	722
10	64	770
11	65	519
12	66	700

a) Input Flight Schedule

```
E-1  ......................................................TTGGGeQQKKRR....
E-2  ......................................................fFFFHHHNNCCC....
E-3  ......................................................XXdIIIcBBBPP....
     0000000001111111111222222222233333333334444444444555555555566666666667
     1234567890123456789012345678901234567890123456789012345678901234567890
                                    WEEK
```

b) Flight Assignment Timeline

Figure 8. Flight Schedule and Assignments

The assignment of pumps to flights is initiated by grouping contiguous flights. Although the groups may be arbitrarily large, in the present implementation the groups are restricted in size to five contiguous flights. Each group is assigned a node on one side of a bipartite graph. Nodes on the other side of the bipartite graph are associated with the virtual pump fleet generated in the previous step. Weights are assigned to links between the pump fleet and flight groups in the following manner:

1. If a pump has a greater allowable lifetime and allowable starts than that represented within a flight group, the weight is equal to the difference between 1) the pump lifetime and 2) the total length of the flights in the flight group.

2. If a pump has a smaller allowable lifetime than a flight group, or fewer number of allowable starts than are represented in the flight group, the weight is set to a large number. Since a pump must have a greater lifetime and number of starts than that represented by a flight group, an optimal assignment which includes a link with a large-numbered weight would represent the violation of a hard constraint.

Using these heuristics, a pump will tend to be matched to a flight group with a lifetime requirement nearly equal to (but less than) the lifetime of the pump. If a matching occurs which includes a link with a large-numbered weight, the hard constraint (2) has been violated. In an operational version of CALMES, some soft constraint could be relaxed until the assignment no longer contained the large-numbered matching.

Generally, a flight schedule will allow several different arrangements of multiple flight groups. For instance, a 10-flight schedule will allow seven different arrangements of five flights. One arrangement is two groups of five flights (flights 1 through 5 and 6 through 10). The other arrangements involve only a single group of five contiguous flights. The resulting groups will be flights 1 through 5, 2 through 6, 3 through 7, etc. Each of these seven arrangements requires a separate matching process with the pump fleet. CALMES will retain the matching with the lowest global cost.

Once all possible five-flight groups are matched, the process is repeated with four-flight groups. If no five-flight groups were matched, the 10-flight schedule in the example above would allow 13 arrangements of four-flight groups. Each arrangement of four-flight groups is optimally matched with the pump inventory; the best optimal match is retained. The process is repeated with flight groups of three, two, and finally single flights. At this point all possible matchings will have been made. The original pump inventory will have been assigned to single or multiple flights in an optimal manner, i.e., in a manner which minimizes the overall differences

```
                               TEST SCHEDULE
                               =============
A-1   ......................................................................
A-2   ...........................................d...cB..P..................
B-1   .......................................XfTFGIHeQNKCR..................
      0000000001111111111222222222233333333334444444444555555555566666666667
      1234567890123456789012345678901234567890123456789012345678901234567890
                                   WEEK

                              FLIGHT SCHEDULE
                              ===============
E-1   ......................................................TTGGGeQQKKRR....
E-2   ......................................................fFFFHHHNNCCC....
E-3   ......................................................XXdIIIcBBBPP....
      0000000001111111111222222222233333333334444444444555555555566666666667
      1234567890123456789012345678901234567890123456789012345678901234567890
```

Figure 9. Test Schedule Timeline

```
                             ASSEMBLY SCHEDULE
                             =================
AL-6  .GGGGGGGGGGffffffffffTTTTTTTTTTFFFFFFFFFFRRRRRRRRRR...................
AL-5  .QQQQQQQQQQIIIIIIIIIIHHHHHHHHHHeeeeeeeeeePPPPPPPPPP...................
AL-4  ccccccccccNNNNNNNNNNKKKKKKKKKKddddddddddCCCCCCCCCC....................
AL-3  ............................XXXXXXXXXXBBBBBBBBBB......................
AL-2  ......................................................................
AL-1  ......................................................................
      0000000001111111111222222222233333333334444444444555555555566666666667
      1234567890123456789012345678901234567890123456789012345678901234567890
                                   WEEK

                               TEST SCHEDULE
                               =============
A-1   ......................................................................
A-2   ...........................................d...cB..P..................
B-1   .......................................XfTFGIHeQNKCR..................
      0000000001111111111222222222233333333334444444444555555555566666666667
```

Figure 10. Assembly Schedule Timeline

between the lifetimes of the assigned pumps and the lifetime requirements of the flight groups.

Fig. 8 depicts a) the input flight schedule (provided by the user), and b) the assignment of pumps to each scheduled flight. The input flight schedule is not realistic, as it schedules a shuttle flight each week from week 55 to week 66. However, it is included as an example of a "worst case" scenario, since it is most demanding of ground test and component assembly resources.

The assignment of pumps to flights is depicted graphically in a timeline format spanning 70 weeks (Fig. 8 b). The assignment graph indicates that pump "F" is assigned to engine E-2 for flights at weeks 56, 57 and 58, while pump "T" is assigned to engine E-1 for flights at weeks 55 and 56. The pump-flight assignment algorithm allows pumps to be assigned to flights in a preferential manner. In the example, single engines are to be assigned to as many flights as possible, with early flights taking precedence over later ones. Other preferences may be included conveniently if they are expressible in terms of weight assignments.

Pump Test Scheduling

Once pump-flight assignments are made, pump test scheduling is performed. This function is cast as an optimal match between nodes representing 1) dates on which test stands are available, and 2) dates on which pumps are required for flight. A hard constraint requiring three months between a pump test and its use on a flight must be accommodated by the test schedule. In addition, two soft constraints are imposed, i.e., it is preferred that tests occur on test stand B-1, no earlier than four months before the flight.

As in the pump-flight assignments, node matchings which would represent a violation of a hard constraint are weighted with a large number. As a result, links between 1) a given flight and 2) test stand dates later than three months before the flight will be given large weights. The soft constraints are imposed by assigning small weights to links between nodes representing test stand B-1 and those representing pumps which have been previously assigned to flights. Slightly larger weights are assigned to links between the pump nodes and nodes representing test stand A-2. In this manner the preference of using test stand B-1 over test stand A-2 is imposed. An optimal match of the resulting bipartite graph will result in 1) a (hopefully) total avoidance of any match between nodes which would violate a hard constraint, and 2) a preferential match between nodes representing time slots on test stand B-1 and the pump nodes. If an optimal match cannot be attained by assigning pump tests to test stand B-1, then test stand A-2 will be assigned some pumps for testing. If an optimal match can be obtained only by violating a hard constraint, a soft constraint must be relaxed until the hard constraint is satisfied.

The results of scheduling pump tests for the given flight schedule are illustrated by the timelines depicted in Fig. 9. The flight assignment timeline is repeated for comparison. Note that some tests are scheduled for test stand A-2 so that no test is performed earlier than four months before the flight. In this example, all hard constraints are satisfied.

Pump Assembly Scheduling

The final step, scheduling of pump assemblies, is depicted in Fig. 10. As in the previous step, the scheduling algorithm allows preferences to be accommodated if possible. In this example, it is desired to schedule assemblies as late as possible, using as few assembly benches as possible. (Alternatively, a user could prefer to fill up all benches if possible, to only use certain benches during certain periods, or to specify variable assembly times, etc.) An obvious hard constraint is that pumps must be assembled before they are tested.

The two sides of the associated bipartite graph will represent 1) available assembly benches and dates, and 2) pump testing dates determined in the previous step. As in the previous step, links between node pairs which would violate a hard constraint if the nodes were matched are given large weights. Node pairings which represent preferences, or soft constraints, are weighted with small weights; the magnitudes of these weights depend on the relative desirabilities of the associated node pairings. In the example, all required pumps can be assembled in time for their green run tests. Four assembly benches are required for the pump assembly.

Summary

A prototype of the SSME Component Assembly and Life Management Expert System (CALMES) has been developed and implemented at CASP/UTSI. The four major functions of CALMES, i.e., 1) the assignment of parts from an inventory to instances of high pressure oxidizer turbopumps, 2) the assignment of pumps to flights, 3) the scheduling of pump tests, and 4) scheduling of pump assembly benches are conceptualized as general optimal assignment problems. A graph-theoretical optimal matching algorithm, based on a modified simplex method, is applied to the major functions required by the SSME component assembly and scheduling processes. Heuristics pertaining to ideal and preferred scheduling and assembly constraints are incorporated into weighted links in the graphical representations. The prototype successfully demonstrates the accommodation of hard and soft constraints in assembly and scheduling processes.

Acknowledgement

This work was performed within the Center for Advanced Space Propulsion (CASP) and was supported in part by NASA Grant NAGW-1195. 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 the Rocketdyne Division of Rockwell International for their cooperation in the use of rocket engine ground test data. The authors would also like to thank A. Daumann and E. L. Kiech for their assistance during the course of this effort.

References

1. Christofides, N., "Graph Theory: An Algorithmic Approach", pp. 373-380, Academic Press, New York, 1975.

2. Papadimitriou, C.H. and Steiglitz K., "Combinatorial Optimization: Algorithms and Complexity", Prentice-Hall, Englewood Cliffs, NJ, 1982.

Project Title: Leader - An Integrated Engine Behavior and Design Analyses Based Real-Time Fault Diagnostic Expert System for Space Shuttle Main Engine (SSME)

Principal Investigators: U. K. Gupta and M. Ali

Project Report

ABSTRACT

An expert system, called LEADER, has been designed and implemented for automatic learning, detection, identification, verification and correction of anomalous propulsion system operations in real time. LEADER employs a set of sensors to monitor engine component performance, and to detect, identify and validate abnormalities with respect to varying engine dynamics and behavior. Two diagnostic approaches are adopted in the architecture of LEADER. In the first approach fault diagnosis is performed through learning and identifying engine behavior patterns. LEADER, utilizing this approach, generates few hypotheses about the possible abnormalities. These hypotheses are then validated based on the SSME design and functional knowledge. The second approach directs the processing of engine sensory data and performs reasoning based on the SSME design, functional knowledge, and the deep-level knowledge, i.e., the first principles (physics and mechanics) of SSME subsystems and components. This paper describes LEADER's architecture which integrates a design-based reasoning approach with neural network-based fault pattern matching techniques. The fault diagnosis results obtained through the analyses of SSME ground test data are presented and discussed.

INTRODUCTION

Most of the present day knowledge based fault diagnostic systems are not capable of performing fault diagnosis in real time. Many of them use a large number of production rules, that make the system response slow. Also, the interactions among the rules make addition and deletion of rules difficult. Furthermore, a number of these diagnostic systems perform shallow reasoning that fails to diagnose faults which require deep level knowledge. A number of these systems also suffer from a major drawback, namely, the user is expected to understand the fault symptoms because of the lack of facilities to acquire knowledge automatically from the sensor data. Since these diagnostic systems are generally unsuitable for real-time diagnostic applications, alternative approaches for developing real-time diagnostic systems are needed.

In recent years, a number of researchers have proposed diagnostic approaches for complex mechanical systems, some of them are based on causal reasoning [1], qualitative reasoning with mechanism models [2, 3], sensor-based reasoning [4], and pattern matching using neural networks [5]. These approaches have been applied to various domains such as digital-circuits, jet engine and other engineering domains. The expert system LEADER discussed in this paper is based on two different but complementing approaches that show a potential for real-time diagnosis of faults occuring in a complex mechanical system such as a rocket engine. The two approaches are based on engine bahavior and engine design analyses, respectively. LEADER has been applied for jet engine and rocket engine fault diagnosis. The approaches and the techniques used in developing LEADER, along with the results of SSME fault diagnosis obtained through the analysis of ground test data are discussed in this paper.

LEADER ARCHITECTURE & APPROACH

The architecture of LEADER is divided into two diagnostic subsystems, and an integration subsystem. One of the diagnostic subsystem performs fault diagnosis based on the analysis of SSME behavior, whereas the other diagnostic subsystem analyzes SSME design and functions to diagnose SSME faults. Each of the diagnostic subsystems performs diagnosis independently of the other by sharing the processor time, or whenever possible by running in parallel. The two diagnostic subsystems allow LEADER to perform SSME fault diagnosis using two different but complementing approaches - one based on utilizing fault patterns obtained from SSME behavior corresponding to the faults, and the other based on applying deep-level reasoning to SSME design. Whenever the results from the two approaches concur. no further analysis is needed. Otherwise, suitable analysis followed by integration of the results from the two approaches is performed by the integration subsystem of LEADER. This is expected to result in an overall improvement in LEADER's performance in terms of its reliability, speed, effeciency, and accuracy of results. As shown in figure 1, LEADER's architecture is divided into the following three subsystems:

A. Engine Behavior Analysis Based Diagnostic Subsytem (BABS).

B. Engine Design Analysis Based Diagnostic Subsystem (DABS).

C. Integration Subsystem (IS).

The three subsystems. their modules, and the overall control strategy and approaches used to design and develop LEADER are described below.

A. Engine Behavior Analysis Based Diagnostic Subsystem

This subsystem performs diagnosis by analyzing the SSME behavior data representing various abnormal conditions. SSME behavior is analyzed by learning the discriminant behavior-patterns associated with various SSME faults, and using them to diagnose unknown faults by matching the learned fault-patterns with unknown fault patterns. The results of this subsystem indicate that this subsystem performs fault diagnosis fast and successfully, if the unknown fault indeed represents one of the fault categories present in the training data. This subsystem is comprised of the following modules:

(1) The Machine Learning Module.

(2) The Diagnostic Module.

(3) Hypothesis Validation Module.

(1) The Machine Learning Module

This module uses the available SSME test sensor data and identifies several parameter characteristics which are important in diagnosing SSME faults represented by the test data. These characteristics include sensor value slope variations and inter and intra sensor slope-trend relationships, slope-magnitude relationships, temporal relationships, magnitude relationships and duration relationships. Figure 2 shows the slope and temporal relationships. For example, if a sensor shows a positive trend whereas another shows a negative trend during the occurence of a fault, the two sensors are considered to have a slope-relationship Typ-1 as shown in figure 2. Also, if one of the sensors starts showing the trend earlier than the other sensor and both continue to show the same trend upto end time, they are related by temporal relationship Typ-8 as shown in figure 2.

The machine learning program associates each fault with a set of relevant sensor characteristics. In our implementation thirteen types of SSME faults and their associated discriminant descriptors were identified by the machine learning module, using SSME sensors data obtained during the twenty-seven ground tests. The approach and the learning techniques used in developing this module are similar to those discussed and described in detail in our previous paper [6], and will not be discussed in detail here.

(2) The Diagnostic Module

After the diagnostic patterns are learned by the MLM in the form of characteristic and discriminant descriptors, the diagnostic module is responsible for using those patterns in diagnosing a fault represented by the given sensors' data. In our present implementation, the diagnostic module uses a neural network based pattern matcher that has been trained with the behavior patterns obtained from the machine learning module. The network is trained to associate a particular set of relational sensor behaviors with a particular fault. During the diagnosis phase, the neural network performs matching of patterns associated with the faults it was trained with, to the patterns associated with the unknown fault. The patterns used in training the neural networks are binary representations of discriminant sensor-pair relationships found to be relevant to a particular fault. The binary representation used here should not be confused with the standard binary value representation of a number, and is used only to indicate the presence or absence of a relationship. Further explanation on these patterns is given later in this section. As a result of this pattern analysis and matching, the system derives several possibilities of faults in terms of hypotheses and their estimated likelihood of occurences which are available for further analysis and validation by the hypothesis validation module.

A neural network consists of a system of interconnected processing elements. The networks used in the present effort consist of three layers. The first layer consists of input nodes, the outputs of which are connected to a second layer of nodes. The second layer is called the "hidden" layer, because it is only indirectly affected by the application of external stimuli to the input nodes. The outputs of the hidden nodes are connected to the third layer, which consists of the output nodes of the neural network.

Each connection is associated with a weight. The input nodes are activated by an external stimulus derived from the behavior relations between various sensor pairs. The activations of the input nodes are multiplied by the weights of the connections between the input and hidden layer; the resulting values become the inputs to the hidden layer. The activations of the hidden layer are then multiplied by the weights of the connections between the hidden and output layer; the resulting values become the inputs to the output layer. The activations of the output layer represent the degree to which the relational features between two sensors match the previously learned features corresponding to different faults. In cases where more than one fault category matches the data, additional reasoning must be applied in order to determine the most likely fault.

Table 1 shows a few examples of the learned descriptors provided by the machine learning module. These descriptors were used to train the neural networks. The first column of table 1 shows pairs of related sensors, the second column shows the faults associated with the sensor pairs, and the third column shows the binary representations of the learned descriptors based on the five types of sensor relationships, namely, slope-type, slope-magnitude, temporal, magnitude and duration relationships respectively; the presence of binary 0 or 1 indicates the presence or absence of a corresponding relationship. For example, the first descriptor for sensor pair HPFTDT1-MCCCDT has a pattern 0001; since the fourth bit from the left is on, this indicates the presence of a slope relationship of Typ-4 between the two sensors.

The diagnostic module has been tested by first training it with the patterns provided by the learning program (using a small set of engine behavior examples repesenting about thirteen fault groups). When presented with these examples and about five unknown examples representing some of the thirteen types of faults, it identified them with reasonable accuracy. Table 2 shows the results of the diagnostic module for the five test cases excluded from the training set, and presented to the diagnostic module to identify the corresponding fault classes. The first column of table 2 shows the fault instances used to test the diagnostic module performance. The test cases were obtained from actual SSME ground tests and represented SSME faults. The second and third columns of table 2 show the fault hypotheses and the corresponding relative likelihood of occurences, respectively. These results form the basis for a preliminary diagnosis by the BABS subsystem. The fault-hypothesis with highest vote-count in each case shown in table 2 is considered to be the most likely fault corresponding to the sensor data of the test-case. Further validation of the preliminary results is performed as described in the next section.

(3) Hypothesis Validation Module (HVM)

This module is responsible for validating the fault hypotheses resulting from the diagnostic module in the light of SSME design. It determines if the events associated with a hypothesis and described in terms of SSME design indeed occured, and are supported by the sensor data of the fault currently being diagnosed by the diagnostic module. If the events associated with a hypothesis are all found to have occured as deduced by direct or inferred evidence, the corresponding hypothesis is considered to be a valid diagnosis, representing the fault currently being diagnosed.

The strategy used by the hypothesis validation module involves knowing the events associated with the hypothesis being validated and using the processes of event-justification and event-contradiction to determine whether the events associated with a hypothesis have occured or not. The processes of event-justification and event-contradiction use the unknown fault sensor data and SSME design to infer the knowledge or evidence that would confirm or deny the occurence of events associated with the hypothesis being validated.

The process of event-justification proceeds as follows:

An event-list associated with the hypothesis to be validated is accessed. Events in the event-list give a general description of the way the faulty component itself and other interacting components in the upstream, downstream, and proximity of a faulty component get affected. Since the description is general, any missing information in unknown fault data or evidence during the process should be deduced/inferred from the SSME design knowledge and parameter-translation knowledge. Two approaches for event-justification are used by the event-justification process, namely:a direct-justification approach, and an indirect-justification approach. The direct-justification approach involves using available unknown-fault sensor data directly to support the occurence of events associated with the hypothesis being validated. All those events for which contradicting evidence is obtained directly are placed under a contradicted-events list, and are processed later by the event-contradiction process. On the other hand, all those events for which no supporting or contradicting evidence is found are processed using the indirect-justification approach. The indirect-justification approach uses SSME design knowledge and parameter-translation rules to find other ways of finding supporting or contradicting evidence. For example if an event suggests that temperatures in the proximity of a component should go up, but does not explicitly mention the sensors, then the components in the proximity of the component are located from the SSME design and their temperature sensors are obtained and checked for a positive trend in the unknown fault sensor-data. An example illustrates parameter translation: when an event expects the temperature of hot-gas in a component to increase, the pressure of hot-gas should be checked if the temperature sensor does not indicate a rise in temperature. If the pressure of the hot-gas seems to be rising, it should be considered as a supporting evidence since a temperature rise can be translated to a pressure rise.

The Event-contradiction processing involves finding out whether the events contradicted belong to the upstream, downstream, proximity, or within a component, and determining the nature of contradiction. Contradictions of events expected to occur within a component (for a given fault-hypothesis to be true) are not tolerated, although the contradicted-events belonging to upstream or proximity may be considered more tolerable than the contradictions of events belonging to downstream or within a component indicated faulty by the hypothesis being validated. As a result of determination and short analysis of the locations of contradicted events, the event-contradiction process gives a list of intolerable contradictions and a separate list of contradictions that can be overlooked.

If none of the hypotheses is found to be valid based on the validation process mentioned above, the occurence of a new type of fault may be indicated. In cases of new types of faults the deep-level reasoning based subsystem is most likely to succeed in performing fault diagnosis using any available results from The Hypothesis Validation Module.

Figure 3 illustrates some of the results obtained from the HVM module for a test case in which the module was given the sensor data of the fault currently being diagnosed, and three fault-hypotheses to be validated. The three fault hypotheses represented Main Combustion Chamber Manifold (MCC-MANIFOLD-FAULT) fault, Liquid Oxygen Injector (MCC-LOX-INJ-FAULT) fault, and High Pressure Fuel Pump (HPFP-FAULT) fault. As shown in figure 3, the occurences of various expected events within, upstream, downstream, and/or proximity of faulty components represented by hypotheses MCC-LOX-INJ-FAULT and MCC-MANIFOLD-FAULT are contradicted, thus invalidating the occurences of these two hypotheses for the sensor data of the fault currently being diagnosed. The HPFP-FAULT hypothesis does not have any contradicted events in any of the four categories shown in figure 3. These results from the HVM module are correct because the unknown fault sensor data was obtained from ground test data representing the HPFP fault.

B. Engine Design Analysis Based Diagnostic Subsystem (DABS)

This subsystem performs diagnosis using design based reasoning. In order to cope with the temporal engine dymanics, the concept of Time Frame (TF) is used. The TF concept is particularly important because abnormality and trend variations are indicated by different sensors in a unique temporal sequence for each fault. A TF is that particular instance of time when one or more sensors start indicating an abnormal parameter condition. Each TF is given a time frame number which is incremented by one for the successor time frame and so on. Also associated with each TF is the group of sensors that start showing abnormal engine parameters and trends at that TF. Every time one or more sensors start indicating an abnormal condition of the engine, a new TF is created and given a time frame number higher than the previous TF number by one. Figure 4 shows the time frames created for simulating real time sensor indications towards abnormal engine parameters and their trends for a High Pressure Fuel Pump Inlet blockage fault. The time frame numbers are indicated by arrows, and the sensors indicating abnormal parameters and their trends are mentioned in a list along the trend portion of the corresponding sensor data curve. In order to perform diagnosis in a real-time environment, information obtained from the sensors indicating abnormal engine parameters and their trends is mapped into various TFs before presenting it to the deep-level reasoning module.

The DABS subsystem consists of the following modules:

(1) SSME Design and Functional Module (DFM).

(2) Deep Level Reasoning Module (DLRM).

(3) Integration Module (IM).

(1) SSME Design & Functional Module (DFM)

This module is responsible for providing the design and functional knowledge pertaining to SSME subsystems, components, or subcomponents whenever requested by the other modules of LEADER. The SSME design and functional module incorporates an SSME design and functional knowledge base (DFKB). The major SSME subsystems such as the oxidizer-subsystem, the fuel subsystem, the main combustion subsystem, their major components and subcomponents along with their functional and physical relationships, interactions, and associated sensors are represented in DFKB. Object representation is used to represent each subsystem and its major components and subcomponents. For example, the oxidizer-subsystem is defined as an object which, in addition to the subsystem-level knowledge, points to component objects such as the High Pressure Oxidizer Turbine (HPOT). Component-level objects, besides storing component-level and some of the parent level knowledge, also point to their subcomponents such as the main oxidizer pump. Thus, a structural hierarchy of knowledge is established. Figure 5 shows a schematic of a Space Shuttle Main Engine design. The major components have been indicated. Figure 6 shows the component-object representation of High Pressure Oxidizer Pump (HPOP).

The various object attributes include pointers to the parent subsystem, children subcomponents, upstream, downstream, proximity, and high interaction components. Some of the object attributes shown in figure 6 indicate functions of HPOP in terms of the change in temperature and pressure of liquid oxygen at the input and output of HPOP. Other attributes describe the types of fault-primitives that can occur in HPOP, the associated sensors and other relevant information.

DFKB also includes a number of causal-symptom rules on the mechanism involved with SSME component design, performance, and interaction. The causal-symptom rules have been divided into five groups depending on the nature of the physical relationships described by them, as discussed below.

The physical relationships between a basic component such as a turbine and a substance such as hot-gas which it carries have been represented in the form of causal-symptom relationship rules involving the heuristics and the physics that describes the cause(s) for a certain sensor indication towards an abnormal engine parameter and its trend. The causes may be described in terms of other engine parameters and their trends, or in terms of fault-primitives such as a leak, blockage, burn, structural-failure occuring in one or more components or in upstream, downstream or within the reference component(s). An example of such a rule is:

SYMPTOM (Rising Turbine Speed)
CAUSE (Rising Turbine Hot-gas Temperature)

This rule suggests that the cause of an increase in turbine speed may be attributed to a rise in the temperature of hot-gas flowing into the turbine.

The physics behind thermal/hydraulic parameter trends shown by a substance such as fuel or hot-gas has also been described in the form of causal-symptom laws/rules. For example, whenever the hot-gas temperature increases beyond the one desired in SSME, a basic cause is the fuel and liquid oxygen ratio imbalance in the hot-gas, with more liquid oxygen and comparatively less fuel in the hot-gas mixture. An example of such a rule is:

SYMPTOM (Rising Hot-gas Temperature)
CAUSE (OR (Increasing Liquid-Oxygen Flow))
(Dropping Fuel Flow)

This rule suggests that a trend towards a rise in hot-gas temperature in SSME may be attributed to either an increase in the flow of liquid-oxygen or a drop in fuel supply/flow in the engine.

Another set of causal-symptom rules/laws based on the mechanics and design of SSME and its components, describes the interaction between any two components used in SSME design. For example, High Pressure Fuel Turbine and High Pressure Fuel Pump are two components of types Turbine and Pump, respectively, used in SSME design. This allows deducing the influences that components with high interaction in a design like SSME would have on each other. These laws also allow inferring the possible causes for a component's parameter trend in terms of other components interacting highly with the symptom parameter. For example, a decrease in load on a pump such as the High Pressure Fuel Pump can be one of the possible causes for the increase of speed of its driving turbine, the High Pressure Fuel Turbine. An example of a causal-symptom rule representing this group is as follows:

SYMPTOM (Dropping Turbine Speed)
CAUSE (Increasing Pump Load)

This rule suggests that a possible cause for a trend towards a drop in turbine speed may be an increasing load on the driven pump.

In order to be able to infer the possible causes or predicted-effects of a parameter trend in terms of its upstream or downstream component, heuristic rules describing the causes and predictions in terms of upstream and/or downstream component connections in SSME design are used. These heuristics are based on the type of a sensor such as a temperature, pressure, speed/rpm, or a flow measurement sensor, as well its trend. For example, an increase in pressure of a liquid such as fuel would indicate that there is a blockage in the downstream component(s). These heuristics are more of a general nature and depend on the type of the given parameter. An example of such a causal-symptom rule is as follows:

```
SYMPTOM   (OR (Dropping Liquid Flow))
              (Dropping Liquid Pressure)
CAUSE     (OR (Leaking Liquid Upstream))
              (Blocked Liquid Upstream)
```

This rule suggests that the possible causes for a symptom like a drop in flow or pressure of a liquid in SSME may be the occurence of a liquid leak or blockage upstream of the location where the symptoms have been indicated.

Finally, another set of domain-specific heuristics has been used, which indicates the cause(s) for a general scenario presented by the parameters showing certain trends. An example of such a heuristic is shown by a causal-symptom rule below:

```
SYMPTOM   (AND (Rising Temperature Sensors))
               (Dropping Pressure Sensors)
CAUSE     (OR (Leaking Fuel))
              (Blocked Fuel)
```

This rule suggests that if most of the temperature sensors in SSME show rising trends, whereas the majority of pressure sensors indicate a droping trend, the likely cause of the indicated abnormality is fuel leak or blockage in SSME.

(2) Deep-Level Reasoning Module (DLRM)

This module is responsible for performing the design based forward reasoning to infer the causes, location and associated events of a given unknown fault by utilizing the unknown fault sensor data, SSME design and functional knowledge, and domain heuristics. The deep level reasoning module is comprised of three problem-solvers, namely the subsystem level problem solver (SLPS), the component level problem solver (CLPS), and the subcomponent level problem solver (SCLPS). It uses a hypotheses blackboard to write the intermediate results, such as most likely faulty subsystem(s), faulty components, associated events, and supporting sensors. A hierarchy of known fault hypotheses and associated subsystems or components is also maintained and used by the various problem solvers of this module.

Our approach in developing the three problem solvers is focussed more towards identifying the root causes as well as the affected subsystems, components, and subcomponents, that are responsible for the abnormal behavior of SSME, and predicting future faults. This approach is adopted because after a fault occurs, its symptoms and affects propagate throughout the system as time passes. Thus, if the cause(s) of a fault is known, the fault propagation may be contained or stopped by taking corrective measures to eliminate as much as possible the causes of the fault. The three problem solvers mentioned above work interactively to identify and diagnose new faults. Each of these problem-solvers is discussed in detail below.

The purpose of the subsystem-level problem-solver is to identify the subsystems appearing to be at fault, and prioritize them in the order in which they should be further investigated to identify the specific components in which the fault has occurred. As soon as the information associated with a TF is presented to the DLRM, the SLPS is triggered. It uses certain heuristics on sensor data. For example, in general the speed, vibrational, and flow measuring sensors are more reliable in pointing towards faults than pressure and temperature measuring sensors. Heuristics like these allow the SLPS to attach weights or varying degrees of importance to the various sensors belonging to a TF. From the SSME design and functional module, the location of each sensor associated with the given TF is obtained in terms of the subsystem they belong to. The total weight of all the sensors of a SSME subsystem indicating an abnormal engine condition gives that subsystem's fault-likelihood number. The various subsystems, corresponding fault-likelihood numbers, and associated sensors and TFs, are written on the subsystem-level of the hypothesis-blackboard. The subsystem with the highest fault-likelihood number is then investigated further by the component-level problem-solver.

The major objective of the component-level problem-solver is to identify the components at fault and the root-causes that led to the SSME fault(s). It generates a fault-mechanism tree to achieve this objective, and also attempts to predict the components affected and conditions that will result due to the present fault(s). A fault-meachanism tree (FMT) is employed. The FMT is a tree-structure that describes the mechanism of fault(s) in terms of the basic underlying causes, effects, and predictions. The FMT consists of causal nodes at top, pointing downward to the corresponding symptom nodes, and the symptom nodes pointing towards the predictions or future symptom nodes. A symptom-node can be considered as a causal-node relative to the node pointed to at the lower-level. Apart from the SSME design and functional knowledge, the causal-symptom rules of physics, mechanics, and heuristics behind the various SSME components and parameters are utilized to generate the FMT. Figures 7a, 7b, 7c illustrate various stages of development of a FMT by the Deep-Level Reasoning Module when presented with the sensor data representing "Blockage of High Pressure Fuel Pump Inlet" fault. Each node represents a mechanism-event, the TF when the mechanism-event most likely occured, and the likelihood referred to as Confidence-Level (CL) that the node lies on the specific path of the FMT which explains the causes and consequences of the fault currently being diagnosed with highest confidence-level values at each node of the path. Such a path is referred to as the best path of the FMT. Various nodes of the FMT are created during the process of reasoning. Figure 7a shows the FMT developed during the TF 1. Each node of the FMT is shown by a block, with the causal/symptom/prediction event represented in the form of a list. The CL associated with each node, and the TF during which the event(s) associated with the node may have occurred is represented by the CL value and the TF is indicated with the node. In figure 7a, the events associated with the top nodes may be responsible for triggering the events associated with the lower level nodes. Using pointers from parent(s) to children, a hierachy is maintained, with events associated with the parent nodes giving birth to the events associated with corresponding children nodes. The top right node in figure 7a suggests that a drop in fuel flow to HPFP may be a possible cause for a decrease in HPFP pump load as indicated by the corresponding child node on the right. A fault-event prediction made during the TF 1 is shown by the node at the bottom of figure 7a. The predicted fault-event indicates an increase in speed of the HPFT turbine. The confidence-level of each node is calculated by adding the CLs of its children nodes and incrementing the resulting CL value by one, the initial CL value of a node. The FMTs in figures 7b, and 7c can be interpreted similarly. The node with no pointers coming into or out of it in figure 7b does not support any of the previously occured events, or no cause has been obtained corresponding to the event associated with this node. Also, this node does not have any child, which indicates that it has not been found to cause any other fault-events. Figure 7c shows (in thick lines) the best path of the FMT. The best path in figure 7c may be formalized into a hypothesis indicating that the HPFP fault is due to a drop in the flow of fuel to HPFP.

Triggering of the CLPS performs the following steps:

- It reads the subsystem-level part of the hypothesis-blackboard to obtain the subsystem with the highest likelihood of fault, and the sensors associated with the latest TF. If no such sensor is available, then it either goes for the next likely faulty subsystem or simply returns, depending on the relative likelihood of the various subsystems to be at fault.

- Once the sensors obtained during the latest TF and supporting the highest priority subsystem are read, the location of the sensors in terms of the components of this subsystem are obtained from the design knowledge. Also, the substance(s) associated with each of those component-names is obtained from the SSME design and functional module. The component-type such as turbine, pump, injector is obtained for each component-name.

- Knowing the components, the corresponding component-types, the associated active sensors and trends, and wherever applicable the associated substance with each component, one or more lists following certain formats called templates are created. Theses templates mostly represent symptoms of the fault currently being diagnosed, and are matched with the symptom side of appropriate causal-symptom rules of the DFKB to heuristically generate the possible causes and future predictions from the sensors and trends of the fault currently being diagnosed. Templates representing the symptoms of the fault currently being diagnosed are called symptom-templates. The five types of symptom templates defined by us and understood by the DLRM are described later in this section. During this step, the CLPS creates all possible symptom-templates for each set comprising of a related sensor, its trend, indicated component-type, and substance, as applicable following the format of the five symptom-template.

- Templates created from related information comprising an abnormal parameter and its trend, component-type, and substance (as applicable), are matched with the symptoms encoded in the causal-symptom rules, preferably of a high success rate category known for that template-type. In this way a number of possible causes for each symptom-template is generated, and pointers set from the causes towards the matching symptoms-templates. Whenever a symptom-template fails to match any causal-symptom rule, an attempt is made to match it with the causal side of the causal-symptom rules. Whenever a match is found with the causal side of a rule, the corresponding symptom-part is considered as a prediction, whose other (new) possible causes are then searched by matching this prediction template with the symptom-sides of the causal-symptom rules.

- The above step is carried out for each set of related information comprising an abnormal parameter and its sensor, trend, component, component-type, substance, for the sensors indicating abnormal behavior of SSME during the latest TF.

- As a result of the above steps a fault-mechanism tree/hierarchy is created, which is used to find the most likely search path leading from the cause(s) of fault(s) to the symptoms and onto the predictions wherever the chain points to the symptoms expected to occur after the latest TF.

- The above steps are repeated for relevant components every time another TF information is presented to the CLPS, and any additional causes and symptoms generated are added appropriately to the fault-mechanism tree created so far.

- The next step is to find the best path in the fault-mechanism tree. Search for the best path is performed by choosing the path with highest confidence-level value at each node of the path. The best path of the fault-mechanism is selected and formalized into a component-level fault hypothesis. Component-level results from the CLPS are written on the component-level part of the hypothesis-blackboard.

In our implementation five types of symptom-templates have been defined, each of them may have a higher rate of success in matching with a certain category/set of causal-symptom rules and less with other categories of causal-symptom rules. The five types of symptom-templates are as follows:

Template-typ-1: (Trend Component-typ Substance Parameter-typ)

Template-typ-2: (Trend Component-typ Parameter-typ)

Template-typ-3: (Trend Substance Parameter-typ)

Template-typ-4: (Majority-trend Parameter-typ)

Template-typ-5: (Trend Parameter-typ)

Template types 1, 2 have a higher rate of matching with causal-symptom rules on the interaction between two components, or between a component and the substance contained in it. Template type 3 has more successful matching with causal-symptom rules on substances (liquid/mixture). Template type 4 has a higher rate of successful matching with the general domain heuristics, and Template type 5 has more success with the rules on sensor-behavior heuristics.

After diagnosing the component-level faults, the DLRM triggers the SCLPS in order to investigate the subscomponent-level faults corresponding to the already identified component-level faults. Knowing the components at fault, and the description of the component-level fault(s) such as causes, symptoms, and associated plausible fault-primitives, the SCLPS looks into the design of each of the components. From the design of each faulty component, information on corresponding subcomponents and fault-primitives possible with those sub-components is obtained. Knowing the nature of a fault that occurred inside a component, its subcomponent with the best match between the nature of component fault, and the subcomponent function, location, design, and plausible fault-primitives associated with its subcomponents is chosen to be the faulty subcomponent. The described approach for the SCLPS is being implemented in LEADER. Results from the SCLPS are written on the subcomponent-level part of the hypothesis-blackboard.

The overall fault-hypotheses are formalized and a brief explaination is then generated from the best path(s) of resulting FMT and the results on the three levels of the hypothesis blackboard.

C. Integration Subsystem (IS)

The diagnostic results based on the SSME behavior and design analyses are integrated by the Integration Subsystem and the resulting hypothesis is considered final. IS checks if the results obtained from both the design and behavior analyses based approaches concur, indicating the same hypothesis. If they concur, then the hypothesis is considered final; otherwise, a brief analysis of the certainty of results from each approach is carried out, so that the subsystem either combines both the hypotheses or validates one of them to give the final results.

RESULTS

The results of fault diagnosis obtained from LEADER illustrate the diagnostic power which results from integrating two complementary approaches, one based on the analysis of engine design and the other on the analysis of the engine behavior. We have tested LEADER with a limited amount of rocket engine data and expect to test it further for numerous types of SSME faults as more data on SSME becomes available. The approach discussed for the DLRM has been tested for some of the fuel subsystem components' faults and the main-combustion subsystem components' faults. The FMTs developed for each fault case has been correctly analyzing the sensor data of the test case, but in most cases the speed and reliability of the component-level fault diagnosis depends on the accuracy and reliability of the causal-symptom rules in the DFKB. We expect to implement the SCLPS and the integration module using the approach discussed here.

Although we are still far from complete and exhaustive testing of LEADER, the results obtained so far with the available data show that the approaches discussed in this paper have a potential for performing real-time fault diagnosis in a complex mechanical system such as SSME.

ACKNOWLEDGEMENTS

This work was performed within the Center for Advanced Space Propulsion (CASP) and was supported in part by NASA Grants NAGW-1195 and NAG 1-513. The Center for Advanced Space Propulsion is part of The University of Tennessee-Calspan Center for Aerospace Research, a non-profit organization located at UTSI. The authors would like to thank the Rocketdyne Division of Rockwell International for their cooperation in the use of rocket engine ground test data. The authors would also like to thank Dr. Roger Crawford for his assistance during the course of this effort.

REFERENCES

[1.] Randall Davis,"Diagnosis via causal reasoning: Paths of interaction ad the locality principle," Artificial Intelligence Laboratory, MIT.

[2.] Raman Rajagopalan, "Qualitative Modeling in the Turbojet Engine Domain," M.S. Thesis, CSL Tech. Rept. T-139, Dept of Electrical Engineering, Univ. of Illinois, Urbana, IL, March 1984.

[3.] Jeff Yung-Choa Pan, "Qualitative Reasoning with deep-level mechanism models for fault diagnosis of mechanism failures," IEEE proceedings, pp. 295-301, 1984.

[4.] M. Ali, D. A. Scharnhost, "Sensor-based fault diagnosis in a flight expert system," IEEE Proceedings of the Second Conference on Artificial Intelligence Applications, Miami, FL., pp. 49-54, Dec. 11-13, 1985.

[5.] W.E. Dietz, Earl L. Kiech, M. Ali, "Pattern based fault diagnosis using neural networks," First International conference on industrial and engineering applications of atificial intelligence and expert system, Tullahoma, TN., June 1-3, 1988.

[6.] U.K. Gupta, M. Ali. "Hierarchical Representation and Machine Learning from Faulty Jet Engine Behavioral Examples to Detect Real Time Abnormal Conditions," Proceedings of the First International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert System, Tullahoma, TN., pp. 710-720, June 1-3, 1988.

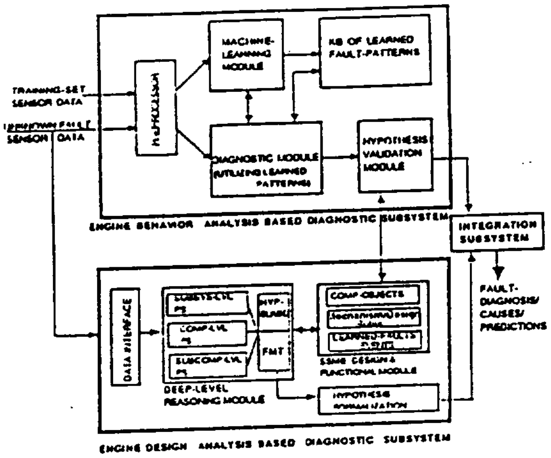


Fig. 1 - Architecture of LEADER

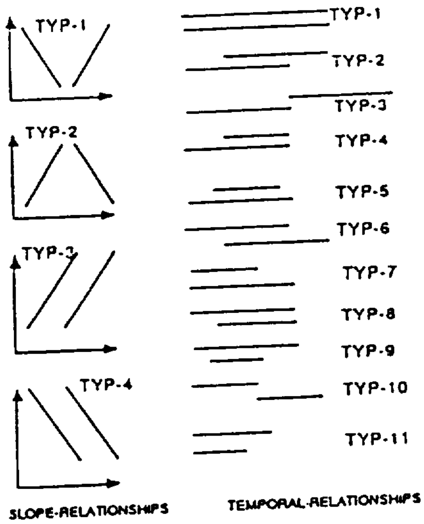


Fig. 2 - Slope & Temporal Relationships

SENSOR PAIRS	FAULT-HYPOTHESES	LEARNED FAULT-PATTERNS
HPFTDT1-MCCCDT	MCC-MANIFOLD	0001 100 10000000000 010 100
HPFPN-PFPDELP	MOXV	0100 001 00100000000 001 001
HPFPDP-OPOVPOS	CONTROLLER	0001 001 00000000001 001 001
HPOTDT1-HPOTXPDRT	HPOT-BEARING	0010 001 00000000010 001 001
	HPOT-TYP1	0100 100 00100000000 100 100
HPFPN-HPFPCLP	HPFP	0010 001 00100000000 100 001
MCCHGP-SFPDELP	MCC-LOX-INJ	0100 100 00000010000 100 100

Table 1. Some of the Learned Fault Patterns (Binary representations of learned descriptors)

TEST-CASES/INSTANCES (SSME fault instances excluded from learning program training set)	RESULTING FAULT-HYPOTHESES	VOTE-COUNT (Measure of match between test-case data & known fault patterns on resulting hypotheses)
MCC-LOX-INJ-1	MCC-LOX-INJ	81
	MPV-FUEL-LEAK	4
	MCC-MANIFOLD	1
MCC-LOX-INJ-2	MCC-LOX-INJ	136
	MCC-MANIFOLD	4
	CONTROLLER	4
	MOXV	1
HPFP-FAILURE-1	HPFP	42
	MCC-MANIFOLD	3
HPFP-FAILURE-2	HPFP	50
	CONTROLLER	2
MCC-MANIFOLD-1	MCC-LOX-INJ	12
	MCC-MANIFOLD	6
	HPFP	2
	CONTROLLER	2

Table 2. Diagnostic Module Results on Five Test-cases

```
PRIORITIZED-HYPOTHESES=
(MCC-MANIFOLD-FAULT MCC-LOX-INJ-FAULT HPFP-FAULT)

HYPOTHESIS= MCC-MANIFOLD-FLT
CONTRADICTIONS=
((WITHIN ((FUEL TEMP NEG (MCCCDT))))
 (DNSTREAM ((TURBINE SPEED POS (HPFPN))
            (FUEL PRESSURE NEG (LPFPDP))))
 (UPSTREAM NIL)
 (PROXIMITY NIL))

HYPOTHESIS= MCC-LOX-INJ-FLT
CONTRADICTIONS=
((WITHIN ((LOX DELP NEG (PFPDELP SFPDELP))))
 (DNSTREAM ((HOT-GAS DELP POS (MCCHGP))
            (EXHAUST-GAS PRESSURE NEG (MCCCHP))))
 (UPSTREAM ((LOX PRESSURE NEG (LPOPDP))))
 (PROXIMITY ((HOT-GAS TEMP POS (HPOTDT1 HPFTDT1))
             (SUBSTANCE TEMP POS *))))

HYPOTHESIS= HPFP-FAULT
CONTRADICTIONS= ((WITHIN NIL)
                 (DNSTREAM NIL)
                 (UPSTREAM NIL)
                 (PROXIMITY NIL))
```

Fig. 3 - Hypothesis Validation Module results Test case represted HPFP-Inlet Blockage fault)

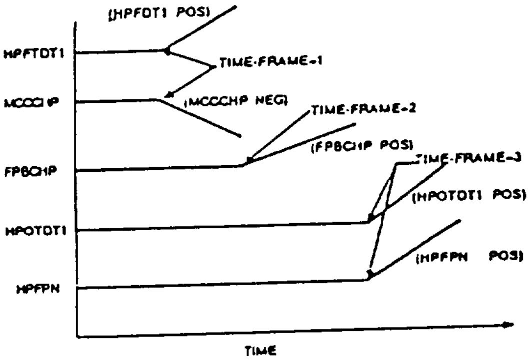


Fig. 4 - Time-Frames corresponding to sensor data on HPFP-Inlet-Blockage Fault

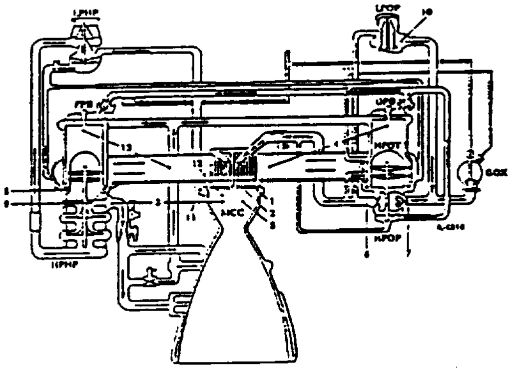


Fig. 5 - Schematic of the Rocket Engine design

```
NAME:            "HIGH PRESS OXIDIZER PUMP"
TYPE:            PUMP
PARENT:          OXIDIZER-SUBSYSTEM
FUNCTION:       "SUPPLIES LOX AT HIGH PRESS TO SSME"
ASSOC-SENSORS:  (HPOPN HPOPDP)
UPSTREAM-COMP:  (LPOP HPOT)
DNSTREAM-COMP:  (MOV OPB FPOV OPOV)
PROXY-COMP:     (OPB HGMC HPOT)
SUB-COMP:       (MOP OPOP OXIDUCT)
INTERACTION:    ((HIGH HPOT HGMC)
                 (MED OXIDIZER-SUBSYSTEM))
FN-INPUT:       ((LOX (PRESS MED LPOPDP)))
FN-OUTPUT:      ((LOX (PRESS HIGH HPOPDP)))
IN-POWER:       HPOT
OUT-POWER:      NIL
ASSOC-EVENTS:   ((WITHIN (LOX PRESS * (HPOPDP))
                         (SELF SPEED * (HPFPN)))
                 (UPSTREAM (LOX PRESS * *)
                           (LOX TEMP * *))
                 (DNSTREAM (LOX FLOW DROP * *)
                           (HOT-GAS TEMP DROP *))
                 (PROXIMITY TEMP RISE *))
PLAUS-DEFECTS:   (LEAKS BURNS)
HISTORY:         ((FREQ 1/9))
```

Fig. 6 - Object-Representation of HPOP (High Pressure Oxidizer Pump)

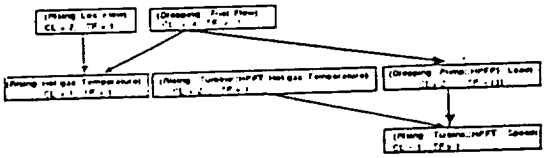


Fig. 7a - Fault Mechanism Tree developed during Time-Frame 1 shown in figure 4

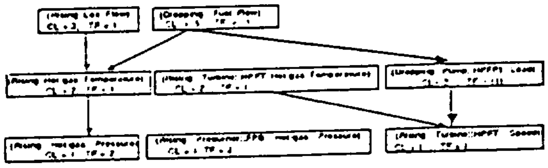


Fig. 7b - Fault Mechanism Tree resulting after Time-Frame 2 shown in figure 4

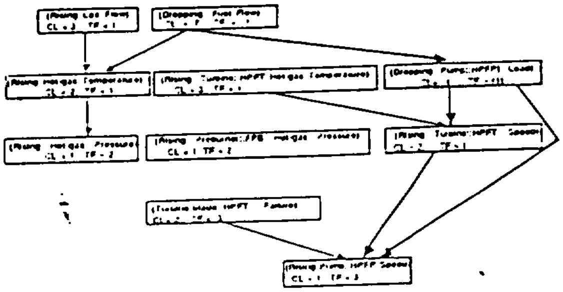


Fig. 7c - Fault Mechanism Tree resulting after Time-Frame 3 shown in figure 4

MICROGRAVITY FLUID MANAGEMENT

- Two-Phase Flow in Helical and Spiral Coils
- The Subscale Orbital Fluid Transfer Experiment
- Low Gravity Transfer Line Chilldown

Project Title: Two-Phase Flow in Helical and Spiral Coils

Principal Investigator: Edward G. Keshock

Other Investigators: Mia L. Bush, Adel Omrani and An Yan

Project Report

Coiled tube heat exchangers involving two-phase flows are used in a variety of application areas, extending from the aerospace industry to petrochemical, refrigeration land power generation industries. The optimal design in each situation requires a fundamental understanding of the heat, mass and momentum transfer characteristic of the flowing two-phase mixture. However, two-phase flows in lengths of horizontal or vertical straight channels with heat transfer are often quite difficult in themselves to understand sufficiently well to permit accurate system designs.

In the case of possible applications to systems operating in a space environment, coiled flow channels may be advantageously employed, since the centrifugal forces experienced by the two-phase fluid acts to promote contact with the channel wall, thereby tending to insure good contact with the wall, and an enhanced heat transfer. Because of such possible improved heat transfer performances it becomes important to identify the conditions under which heat transfer rates are optimized. Correspondingly, it is of importance to know the "price" that must be paid for such optimal performance. From a fundamental, physical viewpoint, these questions are resolved to a very large extent by (1) prediction/identification of the two-phase liquid-vapor flow patter, and (2) the accompanying pressure gradient to achieve or maintain specific flow patterns.

In the case of horizontal or vertical straight flow channels a variety of flow pattern maps have been generated and proposed for application in a 1-g environment [1-4]. Similarly a variety of methods for predicting two-phase pressure drops have been developed (see [5], for example). Relatively few studies have been made of helical coil flows, however [6-9]. The most recent significant study of helical two-phase flows is that of Uddin [10] who observed a variety of flow patterns under various flow conditions, and generated a map identifying the various flow regime boundaries (Figure 1). The transition are considered to be dependent upon the interactive effects of (1) the gravitational force of the test environment, (2) the induced centrifugal force field, and (3) shear stresses induced by secondary flow within the vapor phase portion of the flow.

In general, two-phase flow in coiled geometries can be expected to be much more complex than in straight channels because of the additional factors of an induced centrifugal force field and the resulting induced secondary flows. One's intuition would lead one to expect that the liquid phase would produce a phase separation, with the liquid phase being centrifuged to the outer wall. However, the increased complexity of helical flows becomes apparent from the studies of Bannerjee et al. [6], who observed that under some conditions the liquid phase flowed along the tube wall nearest the coil axis – a phenomenon labeled "film inversion."

A second factor tending to complicate coil flows is the inducement of secondary flows, as depicted schematically in Figure 2. The secondary flow depicted in Figure 2(a) would be expected even in a single-phase flow. The pattern depicted in Figure 2(b), proposed by Uddin, indicates an annular liquid film surrounding a vapor-core flow in which a rotational secondary flow pattern is set up. It should be noted that this latter flow pattern has not been experimentally verified, but has only been hypothesized.

Sundstrand Corporation developed the basic concepts for two-phase helical flows and the present project is an extension of experimental work conducted at Sundstrand.

The present study has the following general objectives:

1. Observe two-phase flow patterns of air-water and R-113 working fluids over a range of flow conditions, for helical and spiral coil geometries, of circular and rectangular cross-section.
2. Compare observed flow patterns with predictions

of existing flow maps.

3. Study criteria for flow regime transitions for possible modifications of existing flow pattern maps.
4. Measure associated pressure drops across the coiled test sections over the range of flow conditions specified.

The experimental measurements and observations will be conducted using an apparatus donated by the Sundstrand Corporation, and now located and installed in the Two-Phase Flow Laboratory of the Mechanical and Aerospace Engineering Department, UTK. A schematic diagram of the modified apparatus is shown in Figure 3.

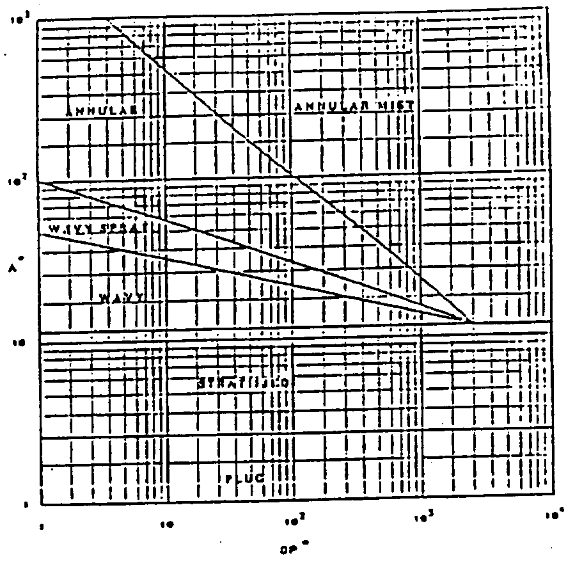
Flow regime observations will be made using a high shutter speed video camcorder. Additionally, several selected operating points will be observed using a high-speed motion picture camera.

Pressure drop measurements will be made for each operating condition, and compared with values calculated using several methods from the literature intended for use in straight channel flows.

The system is now operational and the first set of data has been only recently obtained for a helical coils glass test section of 0.325 inch I.D., and 3 inch diameter helix axis, with a pitch of 1." Future tests are planned for a 0.325" I.D. spiral coils section, and three rectangular cross-section of three different aspect ratios.

REFERENCES

1. Baker, O., Simultaneous Flow of Oil and Gas, Oil and Gas Journal, vol. 53, no. 12, pp. 185 (1954).
2. Choe, W.G., Weinberg, L., and Weisman, J., "Observation and Correlation of Flow Pattern Transitions in Horizontal Concurrent Gas-Liquid Flow," Two-Phase Transport and Reactor Safety, Hemisphere Publ. Co., Washington, 1978.
3. J.M Mandhane, et al., "A Flow Pattern Map for Gas-Liquid Flow in Horizontal Pipes," Int. J. Of Multiphase, vol. 1, no. 4, pp. 537-554, 1974.
4. Taitel, Y., and Dukler, A.E., "A Method for Predicting Flow Regime Transitions in Horizontal and Near-Horizontal Flow," AIChE Journal, 1976, 22, 47-55.
5. Idsinga, W., Todreas, N., and Bowing, R., "An Assessment of Two-Phase Pressure Drop Correlations," Int. Journal Multiphase Flow, vol. 3, pp. 401-413, 1977.
6. Bannerjee, S., E. Rhodes, and D.S. Scott. Film Inversion of CoCurrent Two-Phase Flow in Helical Coils, AIChE Journal, vol. 13, pp. 189 (1967).
7. Bell, K.J., J. Taborak and J.T. Genoglio, Interpretation of Horizontal In-Tube Condensation Heat Transfer Correlations with a Two-Phase Flow Regime Map, Chemical Engineering Progress Symposium Series, vol. 66, no. 102, pp. 150 (1970).
8. Boyce, B.E., and Collier, J.G., Hold-Up and Pressure Drop in the Two-Phase Flow of Air-Water Mixture in Helical Coiled Tube. Proceedings of the International Symposium on Research in Concurrent Gas-Liquid Flow, September 18-19, pp. 203 (1963).
9. Uddin, A.K.M., A Study of the Two-Phase Mechanism of Air-Water Mixture and Development of a Flow Regime Map for Prediction of Flow Regimes in Helically Coiled Tubes. Project Report. submitted to the Department of Engineering Science, Trinity University, September 1986.
10. A.K.M. Uddin, A Model to Predict Two-Phase Flow Patterns in Helically Coiled Tubes, V.I., Particulate Phenomena and Multiphase Transport, T.N. Veziroglu, Hemisphere Publ. Co., 1988.



$A^* = \frac{V_G^2}{g_c R}$

$DP^* = D^* X_{tt}$

where

A^* = Dimensionless radial acceleration of the gas phase

DP^* = Dimensionless characteristic parameter

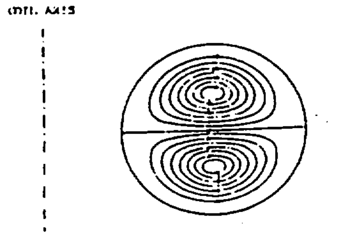
D^* = Dimensionless velocity of the secondary flow

X_{tt} = Lockhart-Martinelli parameter

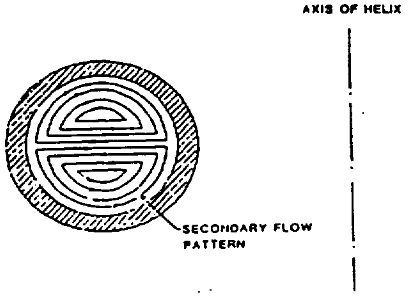
V_G = Gas velocity

g_c = Gravitational acceleration

Figure 1. Flow regime map proposed by Uddin (10) for prediction of flow regimes in vertically upflow helically coiled tubes.



(a) Single Phase Flow



(b) Two-Phase Flow

Figure 2. Illustration of Secondary Flow in the Cross-Section of a Helically Coiled Tube.

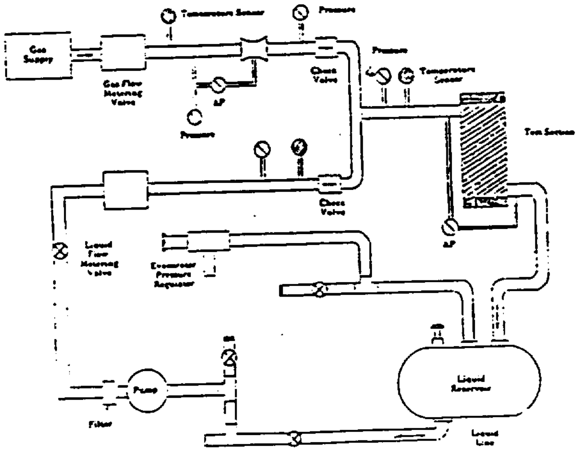


Figure 3. Schematic of Two-Phase Flow Test Apparatus

Project Title: The Subscale Orbital Fluid Transfer Experiment

Principal Investigators: Dr. J. S. Meserole and Dr. Frank G. Collins

Other Investigators: Mr. Ogden Jones, Dr. Basil Antar, Dr. Reinhard Menzel
and Mr. Perry Gray

Project Report

PROJECT OBJECTIVES

CASP is a subcontractor to Boeing Aerospace Corporation to provide support for the concept definition and design of a subscale orbital fluid transfer experiment.

FY89 ACCOMPLISHMENTS

The Subscale Orbital Fluid Transfer Experiment (SOFTE) contract has three phases: concept definition, preliminary design and final design. Most of the work during this year concentrated upon the first phase.

SOFTE is an experiment that will look at the fluid mechanics of the process of transfer of a saturated fluid between two containers. The experiment will be placed in two GAS can containers; the tanks will be in one container and the power and electronics will be in a second container. Since GAS cans are being used, the experiment will be autonomous.

The fluid supply tank will contain a liquid acquisition device (LAD) to assist in fluid removal. The LAD will be in four bands, each containing wire screens. The receiver tank will contain a elastomeric diaphragm that can be inflated to push the fluid back to the supply tank. Vaporized fluid will be used to inflate the elastomeric diaphragm. The experiment will be designed to investigate the no-vent fill process of a cryogenic fuel tank, LAD breakdown, LAD refill and vented fill.

The work during the present year consisted of examining concepts for visual observation of the fluid transfer process, examination of methods for accurately metering the amount of fluid transferred between the two tanks, examination of possible test fluids, and consideration of the materials to use for the elastomeric diaphragm.

The objective of the visual observation is to locate the fluid-vapor interfaces and, if possible, quantify the amount of vapor and the area of the interface. It is proposed to use video cameras to view the overall process in each tank and to place borescopes or other devices through the tank walls to obtain detailed, undistorted views inside the tanks of critical portions of the transfer process. Further work will continue to find an economical means for providing this detailed view, which clearly would increase the data obtained from the experiment.

Fluid gauging of cryogenic fluids at 0-g remains one of the greatest challenges for orbital resupply of cryogenic liquids. In the present experiment the flow meter is the primary means of gauging the fill level of the receiver tank and should be as accurate as possible. In addition, the meter must be rugged enough to survive lift-off, have a small pressure drop and be of small size. It is desirable to measure mass flow directly because of the expected density change of the fluid caused by the temperature change during the course of the experiment. No commercial meter met these requirements and it was determined that the best compromise would be a turbine meter. The turbine meter can accurately measure volume flow if the fluid passing through the meter is known to always be in a liquid state. To measure mass flow, additional pressure and temperature measurements must be made. The overall expected accuracy and reliability of this system will be examined during the next phase.

Numerous fluids having a vapor pressure in the range 0.5 to 0.75 atmosphere were examined. The base-line fluid was Freon 113, which is compatible with the tank materials although it has a vapor pressure that is lower than desired. No safe alternative could be found; however, the duPont Freon alternatives are still under consideration.

During the next phase the preliminary design of the optical system and the elastomeric diaphragm will be completed. Also, the overall accuracy and reliability that can be achieved with the turbine flow meter will be determined. This information will be incorporated into the overall experiment design, which is being conducted by Boeing Aerospace Corporation.

Project Title: Low Gravity Transfer Line Chilldown

Principal Investigators: B. N. Antar and Jere S. Meserole

Other Investigators: F. G. Collins, A. Hedayatpour and T. McGee

Project Report

Significant Accomplishments in the Past Year

A code has been developed that solves for the transfer line chilldown time and flow and heat transfer characteristics in one-g environment. The code solves the transient, one dimensional, space averaged mass, momentum and energy conservation equations for liquid-vapor two-phase flow in tubes. The physical configuration solved is that appropriate for bottom coolant injection in a vertically supported heated tube. Four distinct regions are considered consecutively: fully liquid, inverted annular, dispersed, and fully vapor flow. The conservation equations for both the liquid and the vapor are solved in each region separately. Also, in each region the mass and energy transport between each phase as well as the energy and momentum transport between the tube wall and the fluid are accounted for. A finite wall thickness is also considered.

The model described above was solved numerically through a mixed finite difference scheme with forward time marching. The inverted annular regime was resolved using a semi-implicit finite differencing while the dispersed regime was solved explicitly. Also, a staggered mesh was used in which the velocity was resolved at mesh boundaries while all other field variables were resolved at the mesh centroids. Different mesh sizes were used depending on the region of solution. A coarse mesh was used in the dispersed flow region while a much finer mesh was used in both the inverted annular flow region and the tube walls.

The code was written and verified against the results of Kawaji and Banergee (J. Heat Transfer, vol. 110, 1988). In that work a model was developed for a vertical tube chilldown with bottom flooding with water. They validated their model with an experiment specifically constructed for that purpose. The results from our code compared favorably with the results of Kawaji and Banergee for the case of a 12 ft long and 0.5 in I.D. heated tube at 1000 °F with water injected at 80 °F. Two inlet water speeds were used 5 in/sec and 3 in/sec. The results of both cases compared well with both the experiments of Kawaji and Banergee and their calculations. Figures 1 and 2 show wall temperature histories at four locations along the tube axis for both flow rates. Figures 3 and 4 show the void fraction for the two flow rate cases at three axial locations.

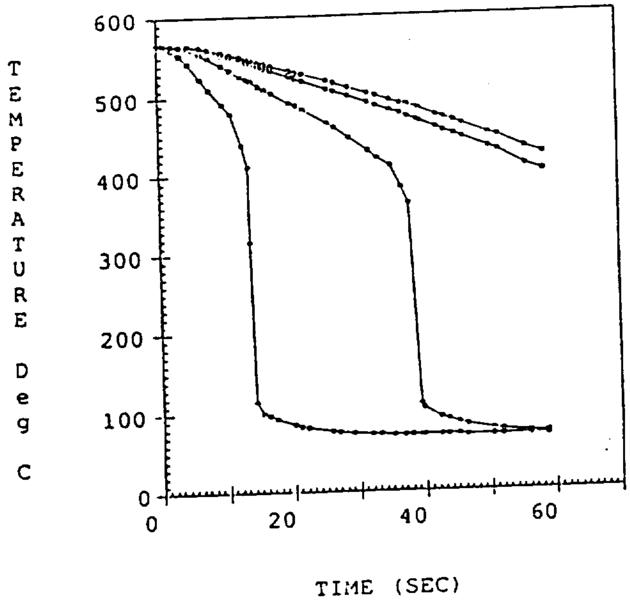
Focus of Current Research Activities

The code validations will be completed by comparing other field variables (e.g. vapor velocity, heat fluxes, etc.) with the experiments and numerical analyses of Kawaji and Banergee. Next, the code will be tested for cryogen chilldown characteristics by simulating an a liquid nitrogen chilldown experiment. In that experiment the cooling fluid will be liquid nitrogen while the test section will be a 0.5 in O.D. by 12 ft long copper tube at ambient temperature. The tube will be supported vertically with liquid nitrogen injected at the bottom of the tube while the top is open to the atmosphere. A test stand for this model is presently under construction. This experiment will be run and numerically simulated for different liquid nitrogen mass injection rates.

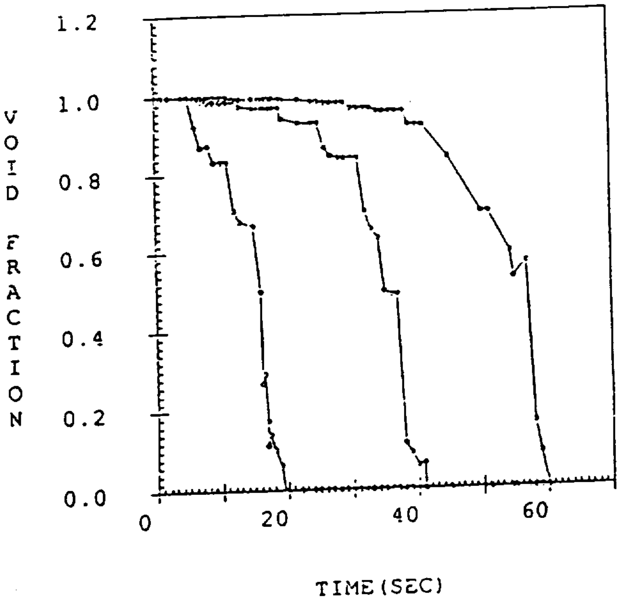
Plans for the Next Year

For the next year the plans are to validate the code with liquid nitrogen, bottom injection experiments. Also, the code will be modified to handle the case for which the tube will exit into a receiver tank at any prescribed temperature and pressure. The code will also be modified for low gravity conditions. For this end a series of low gravity aircraft experiments will be performed to gather all the necessary data needed for upgrading the code. The first low gravity experiment is scheduled for flight in June, 1990 and it is hoped it will soon be followed by the next two flights. In the first flight experiment the working fluid will be freon 113 and the test section will be 0.5

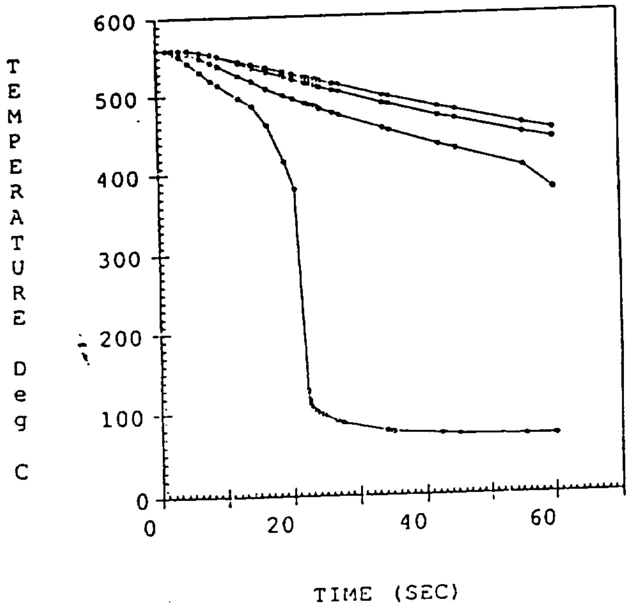
in I.D. and 4 ft long quartz tube heated to approximately 300 °C. In the following two flight experiments liquid nitrogen will be used for the working fluid.



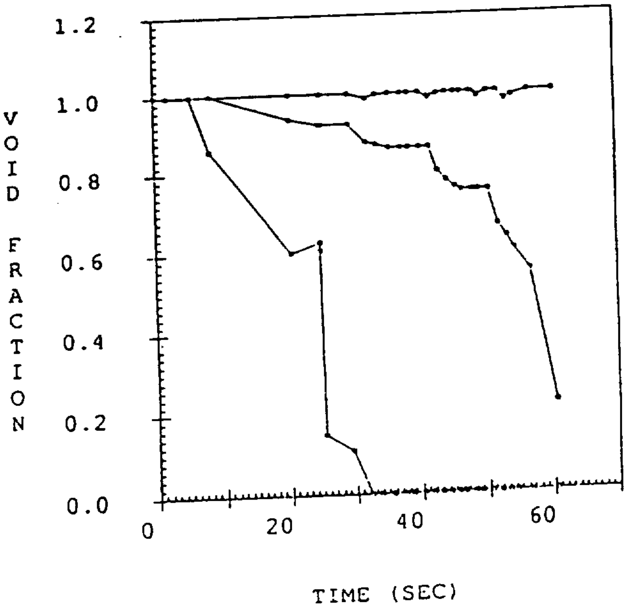
WALL TEMPERATURE HISTORY (10-5-80)



VOID FRACTION HISTORY (10-5-80)



WALL TEMPERATURTE HISTORY (10-3-80) LOCATION



VOID FRACTION HISTORY (10-3-80)

ELECTRIC PROPULSION

- Electrostatic Plasma Accelerator (EPA)
- Ion Propulsion
- Magnetic Annular Arc (MAARC)

Project Title: Electrostatic Plasma Accelerator (EPA)

Principal Investigators: John R. Brophy and Graeme Aston

Project Report

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. A variety of electric propulsion thrusters are being developed which promise specific impulse levels hundreds of seconds greater than those possible with chemical thrusters. For a fixed launch vehicle delivery mass to geosynchronous orbit, the propellant mass savings enabled by high specific impulse thrusters can be used to either significantly increase the satellite station keeping lifetime or to increase the number of on-board transponders. As a result, the economic benefits to commercial users of higher specific impulse electric propulsion thrusters can be very large.

The application of electric propulsion to communications satellites, however, has been limited to the use of hydrazine thrusters with electric heaters for thrust and specific impulse augmentation. These electrothermal thrusters operate at specific impulse levels of approximately 300 s with heater powers of about 500 W. Low power arcjets (1-3 kW) are currently being investigated as a way to increase specific impulse levels to approximately 500 s. Ion propulsion systems can easily produce specific impulses of 3000 s or greater, but have yet to be applied to communications satellites. The reasons most often given for not using ion propulsion systems are their high level of overall complexity, low thrust with long burn times, and the difficulty of integrating the propulsion system into existing commercial spacecraft busses.

The Electrostatic Plasma Accelerator (EPA) is a thruster concept which promises specific impulse levels between low power arcjets and those of the ion engine while retaining the relative simplicity of the arcjet. The EPA thruster produces thrust through the electrostatic acceleration of a moderately dense plasma. No accelerating electrodes are used and the specific impulse is a direct function of the applied discharge voltage and the propellant atomic mass.

PHENOMENOLOGICAL MODEL

The EPA thruster consists of a hollow cathode surrounded by a cylindrical discharge chamber shell which is completely open at the end opposite the cathode. The discharge chamber walls are maintained at anode potential, which may be between 28 to 60 volts positive of the cathode. A magnetic field is applied to the discharge chamber through the use of external permanent magnets to improve the plasma generation efficiency as well as the efficiency with which the plasma is exhausted from the thruster.

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 anode and the cathode. The plasma created in the discharge chamber will characteristically assume a potential slightly higher (a few volts) than the applied voltage. Electrons, by virtue of their much smaller mass, are far more mobile than the discharge chamber ions and will attempt to leave the discharge chamber plasma. An initial start-up transient will occur in which more electrons will leave the thruster than ions. This will cause the discharge chamber 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 ambient space plasma will automatically adjust itself so that the loss rate of ions and electrons are equal; thus a neutral plasma is exhausted from the thruster. The net ion velocity (and ultimately the specific impulse) is determined by the magnitude of this potential difference.

The magnitude of the potential difference between the discharge chamber plasma and the ambient space plasma will be approximately equal to the applied discharge voltage provided the potential of the cathode can be maintained within a few volts of the ambient space plasma. This will be the case provided sufficient electrons can escape from the discharge chamber to neutralize the electrostatically accelerated ions. In order for an electron to escape from the discharge chamber it must have sufficient energy to overcome a potential barrier approximately equal to the applied voltage. Only electrons in the high energy tail of the Maxwellian energy distribution have sufficient energy to overcome this potential difference. Furthermore, there will be sufficient numbers of such electrons only if the electron temperature is above some threshold level (which is a function of the potential barrier to be overcome).

The number of electrons in a Maxwellian energy distribution, with energies above a certain level, can be readily calculated from statistical thermodynamics. For an applied voltage of 32 volts, the electron temperature must be between 4 and 5 eV for there to be sufficient electrons to neutralize the ions leaving the thruster. At 60 volts, the required electron temperature is between 7 and 8 eV. Electron temperatures of these magnitudes can easily be produced in an electron bombardment type of discharge chamber.

SERT II Thrust Mechanism Verification

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 the 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. Other tests indicated that the thrust level decreased as the discharge voltage was reduced (which is consistent with the phenomological model discussed above). Based on the measured thrust of 0.8 mN, Ref. 1 calculated a specific impulse of 300 s at an input power of 80 W, resulting in a thrust-to-power ratio of 10 mN/kW for operation in this mode. The 300 s specific impulse level (for mercury propellant) is much higher than can be accounted for by simple gas dynamic expansion. These tests demonstrate the validity of the EPA thruster thrust producing mechanism, even though the SERT II ion thruster configuration was far from optimum for this mode of operation.

TASK OBJECTIVES

For the EPA thruster to be a viable propulsion concept, overall thruster efficiencies approaching 30approaching 1000 s must be achieved. To achieve these efficiency levels, a discharge chamber which can produce ions for an energy cost on the order of 120 eV/ion while ionizing at least 90mechanisms have been identified including: the power required to ionize the propellant (equivalent to the frozen flow loss in an electrothermal or chemical engine), the loss of unionized and unaccelerated propellant from the thruster, cosine losses resulting from beam divergence, and the production of a non-uniform exhaust velocity profile due to multiply charged particles. The major efficiency loss mechanisms are believed to be the power required to ionize the propellant, and the loss of unionized propellant.

Consideration of the characteristics of the collisional phenomena which accompany the electron bombardment ionization process, together with the electron emission processes required to maintain the arc discharge, result in a minimum energy expenditure of approximately 60 eV to produce each ion in the discharge chamber plasma. It is expected that not all of the ions created in the discharge chamber will be accelerated into the exhaust. Some fraction of the ions will recombine with electrons at the walls of the discharge chamber. This ion recombination process will increase the overall average energy expenditure per exhaust ion above 60 eV/ion. Advances in plasma discharge chamber design and understanding have progressed to the point where discharge chambers for plasma contactors can now ionize 90133 eV/ion.[2] Thus, the performance objectives for the EPA thruster appear to be within reach.

The specific objectives of the EPA thruster proof of concept program are:

1. Operate the Test Bed Ion (TBI) Engine* as an EPA thruster and parametrically vary the thruster geometry and operating conditions to quantify the electrostatic plasma acceleration effect.
2. Based on the experimental results of (1), develop an analytical model for EPA thruster design and performance prediction.

---

* This device was developed under a prior NASA/LeRC funded program at the Jet Propulsion Laboratory and is currently on loan to CASP and EPL.

3. Design and build a radiation cooled EPA thruster based upon the output from (1) and (2) and performance optimize this thruster.

PRESENT PROGRAM STATUS

A brief description of the present program status is given below. The Test Bed Ion Engine discharge chamber has been received from JPL. A support stand has been fabricated and the engine has been mounted to the support stand and installed in an oil diffusion pumped vacuum facility. A novel hollow cathode configuration was designed for use with this discharge chamber and has been fabricated. This cathode uses a shaped porous tungsten insert which functions both to provide a reservoir of low work function material and also as a choked flow orifice to separate the relatively high pressure hollow cathode plasma from the much more dilute discharge chamber plasma.

Eleven ferroresonant DC power supplies have been installed in an instrument rack to serve as the discharge power supply for the EPA thruster. These power supplies can be combined to give a maximum current of 275 A at 28 V. A variety of plasma probes are being assembled for probing the EPA thruster exhaust, including: a Faraday/Langmuir probe rake combination and a retarding potential analyzer. The Faraday/Langmuir probe rake will be used axial and radial position downstream of the thruster. The retarding potential analyzer will be used to verify the beam ion energy and any possible distribution in ion energies.

SUMMARY

The goal of the present program is to demonstrate feasibility of the EPA thruster concept. The program includes both experimental and theoretical investigations of the EPA acceleration mechanism and discharge chamber performance. The theoretical investigations will include the development of a discharge chamber model which describes the relationships between the engine size, plasma properties and overall performance. One use of the model will be to determine the discharge chamber operating conditions which produce sufficiently high electron temperatures for ion beam neutralization. The experimental investigations will be used to demonstrate ion beam acceleration and neutralization, and discharge chamber performance.

REFERENCES

1. Kerslake, W. R., and Ignaczak, L. R., "SERT II 1980 Extended Flight Thruster Experiments," AIAA 81-0665, April 1981.

2. Beattie, J. R., Williamson, W. S., and Matossian, J. N., "Plasma Contactor Neutralizer System for ATAS Missions," IEPC-88-030, October 1989.

Project Title: Ion Propulsion

Principal Investigators: Dr. J. S. Meserole and Dr. Dennis Keefer

Other Investigators: Dr. Wilhelmus M. Ruyten and Mr. Xiaohang Peng

Project Report

INTRODUCTION

An ion engine is a plasma thruster which produces thrust by extracting ions from the plasma and accelerating them to high velocity with an electrostatic field. The ions are then neutralized and leave the engine as high velocity neutral particles. The advantages of ion engines are high specific impulse and efficiency and their ability to operate over a wide range of input powers. In comparison with other electric thrusters, the ion engine has higher efficiency and specific impulse than thermal electric devices such as the arcjet, microwave, radiofrequency and laser heated thrusters and can operate at much lower current levels than the MPD thruster. However, the thrust level for an ion engine may be lower than a thermal electric thruster of the same operating power, consistent with its higher specific impulse, and therefore ion engines are best suited for missions which can tolerate longer duration propulsive phases.

The critical issue for the ion engine is lifetime, since the prospective missions may require operation for several thousands of hours. The critical components of the ion engine, with respect to engine lifetime, are the screen and accelerating grid structures. Typically, these are large metal screens that must support a large voltage difference and maintain a small gap between them. Metallic whisker growth, distortion and vibration can lead to arcing, and over a long period of time ion sputtering will erode the grid structures and change their geometry. In order to study the effects of long time operation of the grid structure, we are developing computer codes based on the particle-in-cell (PIC) technique and laser induced fluorescence (LIF) diagnostic techniques to study the physical processes which control the performance and lifetime of the grid structures.

RESEARCH OBJECTIVES

In order to verify the results of the code and to aid in the development of physical models used in the code, we are developing an experimental capability for the operation and diagnostics of ion engines. The test chamber is approximately 3 m in diameter by 9 m in length and will be pumped cryogenically using approximately 1 kW of 20 K gaseous helium. A three axis computer controlled translation system will permit movement of the engine to facilitate precise spatial diagnostics of the exhaust flow. Mr. Jim Sovey of NASA Lewis Research Center has agreed to provide us with an ion engine for diagnostic testing.

In additional to conventional electric probe diagnostics, we are developing a LIF technique to measure the velocity, density and temperature of the neutral and, possibly, ion species in the exhaust flow. The diagnostic technique uses the single frequency line from a ring-dye laser to excite fluorescence from the species of interest. As the laser frequency is scanned across the doppler broadened absorption line, the emitted radiation is measured using an optical detector. The strength of this optical signal is proportional to the velocity distribution function of the absorbing species, resolved along the laser beam direction. Calculation of the zeroth, first and second moments of this measured distribution provides a measurement of the density, mean velocity and temperature of the absorbing species.

The basic function of the grid structure is to extract ions from the plasma maintained in the interior of the engine and to accelerate them through the grids and out of the engine. In this extraction and acceleration process some of the ions will strike the surfaces of the grids and remove metal atoms from the structure by sputtering. In addition, some collisions between fast ions and neutral particles will occur which result in charge exchange that produces slow ions near the grid surfaces. These slow ions will be accelerated to the grid surfaces by the strong electric field and will cause additional sputtering. Collisions between ions or electrons and neutral particles, particularly in the rare gas propellants, can also produce metastable neutral particles which may also produce sputtering when they impinge on the grid surfaces. We are developing a detailed computer code in order to study the effect of grid geometry and the imposed potential field on these life-limiting processes and the efficiency of the acceleration process.

Particle-in-cell code

To model ion-thruster performance, we are developing a particle-in-cell code [1], in which a single pair of screen-grid and accelerator-grid holes spans an axially symmetric computational domain, as illustrated in figure 1. Thus, the (generally nonuniform) computational grid extends from the plasma inside the ion engine out toward the vacuum of space. Under typical operating conditions, the electrical potential depends not only on the fixed voltages on the metal grids, but also on the extracted ion flux. The particle-in-cell (PIC) method is particularly suited to model such complex dynamical behavior. Its main features are illustrated in figure 2:

During each iteration in time, one first solves Poisson's equation, whose inhomogeneous part is given by the instantaneous charge density. Then, one obtains the electric field for each grid point. The next two steps are at the core of the PIC-concept: Each particle, representing a "cloud" of charge density, and generally located within some cell of the computational grid, is accelerated and displaced according to the interpolated values of the electric field at the grid points. Also, new particles are added into the computational domain by random sampling from the plasma phase-space, and particles that impinge on the electrodes or exit through any of the other boundaries are removed from the computation. To close the cycle, one calculates the new charge density on the grid points from the local particle positions. Interpolations from grid coordinates to particle coordinates and vice versa are performed using linear volume-weighting, as illustrated to the right of the flow diagram in figure 2. When the system reaches steady-state, or at any other point during the computation, such quantities as thrust, electrical efficiency, and electrode impingement currents can be obtained.

Many sophistications can be made to this model. Already, electrons are taken into account, assuming instantaneous thermalization with the ion flow. In the near future, neutral atoms will be added to the code, whose charge exchange interactions with the ions and electrons will be modeled using Monte-Carlo techniques. Also, metastable neutrals may be taken into account, since these are particularly accessible by laser diagnostic techniques. Ultimately, we expect the code to predict accurate values of the specific impulse of the device, and reasonable estimates of grid lifetimes, so that optimization studies of grid geometry and different grid structures (for example, floating metallic or dielectric screen grids) can be performed.

PRELIMINARY RESULTS

To illustrate some of the capabilities of the PIC code, a representative calculation for a single-hole ion engine operated with neon gas is shown in figure 3. This calculation corresponds roughly to the geometry shown in figure 1, with a screen grid aperture of 2 mm, and voltages as indicated in figure 1. Initial conditions are for a plasma temperature 2500 K, ion density 10^{16} m^{-3}, and a perveance of 10^{-9} $A/V^{3/2}$. Some 15,000 particles were included in the calculation. In figure 3(a), contours of the electrical potential are shown. Note that, even downstream from the accelerator grid, these contours are bent around the centerline, due to the ion density. Contours of the ion density are shown in figure 3(b). The impingement on the electrodes, on the order of twenty percent in figure 3, and the divergence of the ion plume are highly sensitive to the geometry and operating conditions of the ion-thruster. Finally, figure 3(c) shows a plot of average velocities of the ions, as obtained from the PIC code.

Not only will the PIC code be run for different geometries and operating conditions of the thruster, but also comparisons with experimental results are planned, notably with work performed at Colorado State University by Paul Wilbur and coworkers (see, for example, ref. [2]), and, once our own facilities become operational, with detailed measurements performed here at the Space Institute.

[1] C. K. Birdsall and A. B. Langdon, *Plasma Physics via Computer Simulation* (McGraw-Hill, New York, 1985).

[2] G. Aston, *Ion Extraction from a Plasma*, Ph.D. thesis (Colorado State University, Fort Collins, 1980).

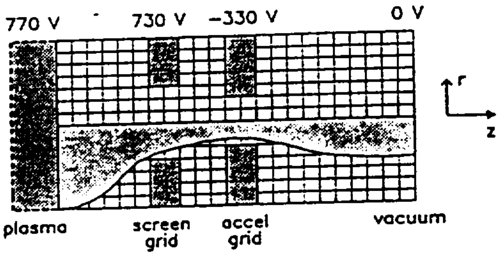


Fig. 1: Schematic of ion-thruster computational grid.

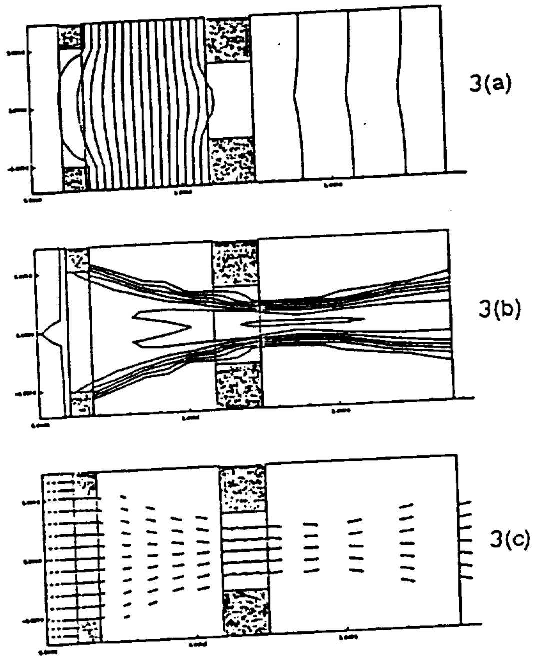


Fig. 3: Representative results of PIC code: (a) electrical potential, (b) ion density, (c) average ion velocities; experimental conditions are given in the text.

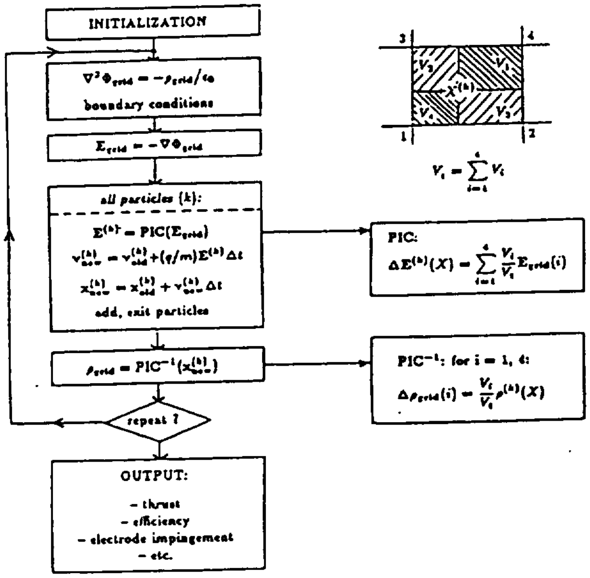


Fig. 2: Schematic of particle-in-cell (PIC) code.

Project Title: Magnetic Annual Arc (MAARC)

Principal Investigator: Herbert Thomas

Project Report

Electromagnetic propulsion systems offer high specific impulses (> 1500) and high thrust-density as compared to conventional chemical rocket propulsion systems. Understanding and simulating the physical processes in electromagnetic propulsion devices (as well as testing with refined diagnostic equipment) hold the key to improved performance and longevity of these devices. This in turn is an essential requirement for a practical application in future space endeavors.

The physical processes involved in electromagnetic accelerators, especially in Arc-Jets and MPD-thrusters, are quite complicated and deny an accurate description by analytical means. Therefore it is necessary to develop computer-codes to simulate the main effects occurring in electromagnetic accelerators. Depending on the density levels at which a particular device operates, tow different approaches are necessary. For Low density devices a statistical description of the flow is necessary, while for higher density where collisional processes dominate the continuum equations are required. The Magnetic Annular Arc (MAARC) model being developed assumes a collision dominated plasma and the conservation equations for mass, momentum and energy including terms for the electromagnetic effects are as follows.

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho u}{\partial x} + \frac{\partial \rho v}{\partial y} = 0$$

$$\frac{\partial(\rho u)}{\partial t} + \frac{\partial}{\partial x}(\rho u^2 + p) + \frac{\partial}{\partial y}(\rho u v) = (\bar{J} \times \bar{B})_x + \frac{\partial}{\partial x}\tau_{xx} + \frac{\partial}{\partial y}\tau_{yx}$$

$$\frac{\partial(\rho v)}{\partial t} + \frac{\partial}{\partial x}(\rho u v) + \frac{\partial}{\partial y}(\rho v^2 + p) = (\bar{J} \times \bar{B})_y + \frac{\partial}{\partial x}\tau_{xy} + \frac{\partial}{\partial y}\tau_{yy}$$

$$\frac{\partial(\rho w)}{\partial \zeta} + \frac{\partial}{\partial x}(\rho u w) + \frac{\partial}{\partial y}(\rho v w) = (\bar{J} \times \bar{B})_\theta + \frac{\partial}{\partial x}\tau_{xz} + \frac{\partial}{\partial y}\tau_{yz}$$

$$\frac{\partial(e)}{\partial t} + \frac{\partial}{\partial x}[u(e + p)] - \frac{\partial}{\partial y}[v(e + p)] = \bar{E} \cdot \bar{J} + \nabla \cdot (k \nabla T) + \Phi$$

Maxwell's equations in combination with Ohm's law are used to determine the electric and magnetic fields and the current distribution:

$$\frac{\partial \bar{B}}{\partial t} + \nabla \times \bar{E} = 0$$

$$\frac{1}{c^2}\frac{\partial \bar{E}}{\partial t} - \nabla \times \bar{B} = -\mu_o \bar{J}$$

$$\bar{J} = \sigma(\bar{E} + \bar{u} \times \bar{B}) - \frac{\omega_e \tau_w}{|B|}(\bar{J} \times \bar{B})$$

While Maxwell's equations can be used to solve electromagnetic problems having a relatively low electrical conductivity, numerical instability problems arise if the electrical conductivity is high. The numerical instability prohibits the use of Maxwell's equations in their original form for electric thrusters. However, it is possible to combine Maxwell's equations and Ohm's law to obtain the general induction equation.

$$\frac{\partial \bar{B}}{\partial t} + \nabla \times \frac{\nabla \times \bar{B}}{\mu_o \sigma} - \nabla \times (\bar{u} \times \bar{B}) + \nabla \times \left[\frac{\beta}{\mu_o}(\nabla \times \bar{B}) \times \bar{B}\right] = 0$$

The resulting induction equation is an elliptic, 2nd order non-linear differential equation. It is solved by a successive over-relaxation method (SOR), which enable the simultaneous solution of a system of tri-diagonal equations, sweeping along lines of constant ξ and η (the coordinates in computational domain). The effect of an applied magnetic field are calculated once and are part of the right-hand-side. The MAARC geometry is depicted in Figure 1.

The boundary condition for the insulator results from the parallel-current condition, fixing the value of the magnetic field times the radius along the insulator. The conditions imposed on the electrodes consist of:

a) the total current passing through the electrodes (input variable)

b) The magnetic field value is zero for downstream and

c) The electric field is perpendicular to the electrodes.

Calculations of a test-case showed convergent behavior. Because of the differencing used in those calculation, the accuracy of the magnetic field strength is only approximately one per mille. This inaccuracy leads to errors of the order of 1 to 10 for the current density and the electric field, which are derived from the magnetic field strength. It is therefore necessary to reformulate the Induction equation in a strong conservative form to reduce the inaccuracy of the magnetic field strength, such that the error of the current density and the electric field becomes tolerable. This is especially important, since the boundary conditions are derived from the current densities and the electric field values.

Reformulation of the Induction equation into a strong conservative form is now in process.

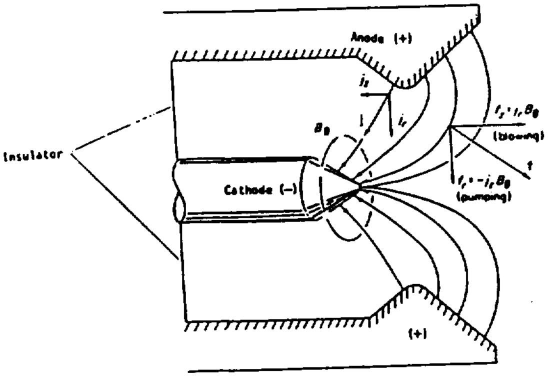


Figure 1. MAARC Geometry

LASER MATERIALS PROCESSING

- Effect on Pulse Duty Cycle on Inconel 718 Laser Welds

Project Title: Effect of Pulse Duty Cycle on Inconel 718 Laser Welds

Principal Investigators: M. H. McCay, T. D. McCay and N. B. Dahotre

Other Investigators: C. M. Sharp, A. Sedghinasab and S. Gopinathan

Project Report

ABSTRACT

Crack sensitive Inconel 718 was laser pulse welded using a 3.0 kW CO_2 laser. Weld shape, structure and porosity were recorded as a function of the pulse duty cycle. Within the matrix studied, the welds were found to be optimized at a high (17 ms on, 7 ms off) duty cycle. These welds were superior in appearance and lack of porosity to both low duty cycle and CW welds.

BACKGROUND

Since its development thirty years ago (1), Inconel 718 has grown in use until it accounts for 35% of all superalloy production (2). The material is routinely employed in gas turbine disks because of its resistance to creep and stress rupture at elevated temperatures. Additional uses which take advantage of its environmental cracking resistance have developed in nuclear power plant and cryogenic facility piping.

Inconel 718 is known as a nickel-iron base alloy. It contains chromium for oxidation resistance, molybdenum and tungsten as solid-solution strengtheners, titanium, aluminum and columbium to combine with nickel as precipitation strengtheners, and carbon and other elements for various complex effects (3). In contrast to alloys that are strengthened by the ordered face centered cubic (fcc) gamma prime phase, Inconel 718 is strengthened by the ordered body centered tetragonal (bct) gamma double prime. The nominal composition is given in Table I.

When Inconel 718 is welded, metallurgical evaluation often reveals microfissures in the heat affected zone. In general, these are thought to be due to liquation, a phenomena that is influenced by microstructure and constituent inhomogeneities (4). The microfissures create problems both in hardware fabrication and repair and, therefore, the welding of Inconel 718 has continued to be a challenge for both researchers and manufacturers.

Laser welding with its inherently small heat affected zone has been considered for Inconel 718. A study (5) on the welding and cutting of Inconel 718 with a 15kW CW (continuous wave) CO_2 laser indicated that it could be an effective joining method for up to one half inch thicknesses. The CW laser welds produced isolated instances of microfissuring and some porosity. Several additional studies have been made of pulse laser welding on Inconel 718 using YAG lasers (6,7). In these instances the material cracked in the nailhead in both individual and overlapping welds. The cracks were intergranular and attributed to solidification cracking.

The purpose of this investigation is to study the solidification morphology and weld structure, including porosity, of Inconel 718 as the laser welding process proceeds from CW to individual non-overlapping pulses.

EXPERIMENTAL PROCEDURE

WELDING - Bead-on-plate laser welding of the Inconel 718 was accomplished with a Rofin Sinar RS3000 laser. The RS3000 is a commercially available fast axial flow, RF (27.12 MHz) excited, CO_2 (10.6 μm wavelength). For welding the laser may be operated in two different configurations: 1) Quasi CW - output is pulse width modulated at 5 KHz to give control of average power from 0.2 to 3.7 kW; 2) Gated Pulse - output can be gated from a CW condition up to 25 kHz; peak power remaining fixed at 3.7 kW. The cavity can be configured to produce beams with several different spatial modes including TEM_{00}, TEM_{10}, and TEM_{20}.

One-eighth inch (1/8") thick wrought alloy (ASTM grain size #6) was welded using the standard RS3000 welding head (150 mm f/6 optics, circular polarization). The welding head nozzle permits shielding gases, both coaxial flow and plasma control, to be used to moderate the welding process and protect the molten metal during solidification. Helium, at a flow rate of approximately 60 liters per minute, was selected as the shielding gas. The laser beam was directed perpendicular to the workpiece. Figure 1 is a schematic of the welding nozzle and the sample orientation. The beam was operated in the TEM_{10} mode (ring with a central peak). Gated pulsing was employed using the parameters shown in Table II. The measured peak power, i.e. delivered at the work piece, was 3.2 kW. This is below the 3.7 kW output due to the optical system losses in the beam delivery system. The energy per pulse was kept constant by maintaining the same pulse on-time and only varying the pulse off-time. A pulse on-time of 17 ms was necessary to provide complete penetration of the sample. Pulse off- time was varied from 1 ms to 21 ms in increments of 2 ms. The samples are designated STA 1, STA 3, ... STA 21, indicating the pulse off-time used for the particular weld. Laser power output calibration was done prior to and after welding with a Scientech Model 390 water cooled calorimeter. Figure 2 shows the delivered average power as a function of the selected pulse off-times. For comparison, a CW bead on plate weld was made under similar gas flow and focus conditions, using 2.5 kW at a weld speed of 1.5 m/min. Several individual (isolated) pulses were also made for characterization. a function of pulse off-time.

DIAGNOSTICS - Plasma Spectroscopy - The radiation emitted by the metallic plasma formed above the weld pool was focussed onto the entrance slit of a 1/4 meter spectrometer. The transfer optics consisted of a 6:1 telescope which reduced the size of the plasma image to about 1 mm at the plane of the entrance slit. A grating with 1200 lines/mm was chosen for obtaining high resolution spectra, providing a spectral window of approximately 60 nm. Five separate spectral windows of constant extent (60 nm) were used to examine the many spectral lines of interest, most of which are in the 250 - 500 nm range. An average of 100 temporal (20 msec each) scans over the length of the bead was obtained by using an optical multichannel analyzer (OMA). This enhanced the signal to noise ratio and negated the transients within the plasma which result from flow field, melting and solidification process effects. Each spectral window was calibrated using suitable emission line calibration lamps. The spectral data was later transferred to a personal computer for post-facto reduction and analysis.

Since the nature of the decaying (during the beam off-time) plasma may effect the coupling of the next laser pulse to the workpiece, the plasma rate of decay is of significance in pulsed laser welding. It is currently impossible to model such a process due to unknowns such as the appropriate recombination coefficients, collision cross sections and other atomic properties. This is chiefly due to the plasma containing most of the constituents of the substance being welded as well as any shield gas. In order to estimate the rate of decay of the plasmas an experiment was set up wherein the radiation was monitored using a nanosecond response silicon p-i-n photodiode. The photodiode output was stored on an oscilloscope. This data allowed the time constant of the plasma decay to be estimated.

Video - A videocamera with a 4.2 neutral density filter recorded the plasma image on videotape during the weld passes. The video frames were subsequently digitized and analyzed using a VICOM image processor.

Sample Preheat - The effect of preheating/cooling on the microstructure of the Inconel 718 weldment was also studied. One of the samples was heated to 780°C in an electric resistance furnace. After soaking for one hour, it was immediately subjected to laser pulse welding using a pulse off-time of 21ms. A second sample was immersed in liquid Nitrogen for one hour and then subjected to laser pulse welding using the above conditions. A third specimen of Inconel was pulse welded at room temperature, also using the same conditions.

Metallography - The welds were metallographically examined in the XY, YZ, and XZ planes (See Figure 1) using an optical microscope and a video camera connected to a VICOM digital image processing system. Top and bottom weld bead width and the variation in bead width (maximum width - minimum width) were measured in the YZ plane. Gas pore diameter, gas pore spacing and gas pore area fraction were measured in the XZ plane. Gas pore area fraction is the total centerline pore area divided by the total centerline weld area. Since the weld shape and solidification structure were expected to vary in three dimensions, successive sections were examined by polishing in 0.1 mm increments in the YZ and XY planes for samples with pulse off-times of 7 ms, 11 ms, 15 ms, and 21 ms.

RESULTS

Figures 3 and 4 show typical top and bottom views, (XY planes) respectively, of the pulsed welds. The periodic changes in width correlate with the laser pulse rate. Since the workpiece travel speed was maintained constant for all of the runs, the physical distance between pulses increases with increasing off-time. This effect can also be seen in the longitudinal views along the weld centerline (XZ plane) in Figure 5. As the pulse off-time increases, the melt zones from the individual pulses begin to separate.

Figures 6 and 7 show the minimum and maximum top and bottom (respectively) bead widths as the pulse off-time is increased. For both the top and bottom beads there is an overall general increase in variation of the bead width (Figure 8) which reflects the periodic correlation of width with laser pulse rate. In addition, pores begin to appear more frequently within the weld center as the pulse off-times increase. Figure 9 demonstrates the increase in poring, showing the pore area fraction as a function of pulse off-time.

Using an average radius for the large pores of 0.025 cm and a surface energy value (8) of 1770 dynes/cm, the pressure across the wall of the gas bubble forming the pore can be calculated from the following equation:

$$P = 2\sigma/r$$

where σ is the surface energy and r is the radius of the gas bubble. These values yield a pressure of 106 mm Hg. The temperature of nickel at that vapor pressure (8) is 2370°C. For the other Inconel 718 constituents one obtains temperatures of 2140°C for chromium, 4200°C for molybdenum, and 2370°C for iron. Except for molybdenum, the anticipated weld pool temperatures are thus well above those required to vaporize the Inconel constituents and produce these gas bubbles. The dendrite arm spacings adjacent to the pores are significantly smaller than those further distant, indicating that the cooling rates are greater in the vicinity of the pores. This is consistant with vaporization of the metal to form the pores.

These same gases would be expected to make up the weld keyhole and the plasma above the surface. Figure 10 presents the radiating constituents within the plasma which is comprised primarily of neutral atoms, singly ionized metal atoms and electrons. Nickel, iron, chromium and niobium are definitely identified as being present in the plasma.

Figures 11 and 12 are mid-section views in the XY plane for STA 7 and STA 21, respectively. The weld pool in STA 21 appears to have had sufficient time to almost completely solidify between pulses. The 100X views show the growth structure and the appearance and location of the pores.

Figure 13 is a series of photographs taken from the video image of the plasma obtained during welding. The plasma ignites, grows, encounters the destruct jet, and dissipates with each laser pulse. This can also be seen in the photodiode data (Figure 14). The central (plateau-shaped) trace in the figure is the excitation pulse for the single laser pulse indicated by the arrow at the bottom of the scope display. The plasma ignites within 2 milliseconds behind the beam "on" trigger and dissipates within 2 milliseconds after the trigger is "off".

Successive cross-section (YZ plane) polishes can be used to reveal the change of weld pool geometry throughout a laser pulse. Figure 15 shows schematic cross-sections throughout one pulse for two different duty cycles. Figure 16 is an optical micrograph showing a typical cross-section for CW and a single (17 msec) laser pulse. The geometry for a series of pulses varied with pulse off-time. The short off-time sample had the most uniform depth weld pool geometry with significantly less variation. The weld middle width of STA 7 was 0.8740 ± 0.0320 mm and that of STA 21 was 0.7312 ± 0.1295 mm over the duration of a pulse.

The width to depth ratios for the top nailhead are presented in Figure 17. As the pulse off-times increase, so do both the ratio and its maximum and minimum values. The nailhead top width (W_T) to neck width (W_N) is 1.853 ± 0.077 mm for STA 7 and decreases to 1.73 ± 0.244 for STA 21 where 0.077 mm and 0.244 mm represents the standard deviation in the measured widths. The W_T/W_N ratio was largest for the CW sample, being 2.269 ± 0.088. This ratio for a single pulse (17 msec) was 1.995. Since convection exaggerates the nailhead, its effect must be lessening as pulse off-time increases. This is confirmed by a similar reduction in the amount of banding/perturbations in the solidifation process. The number of visible bands was highest for STA 7 (six) and lowest for STA 21 (one).

The short off-time weld (STA 7) is fairly uniform with columnar grains growing toward the weld centerline. There is a distinguishable nailhead the long off-time weld (STA 21) shows considerable variation in size and shape. Columnar solidification proceeds perpendicular to the laser beam. These grains appear equiaxed on cross-sections removed from the pulse centerline. The nailhead is small and in some sections almost non-existant.

CONCLUDING REMARKS

Figure 18 is a schematic of the welding process as it occurs for pulses with two different off-times. The laser beam ignites a metallic plasma which influences the thermal coupling to the alloy. Although the metallic plasma extinguishes between the STA 7 pulses, the molten zone is continuous, producing a large nailhead with a small width to depth ratio. Solidification occurs at a consistent angle with respect to the weld direction. There is significant convective motion as evidenced by the size of the nailhead, the solidification banding, and the minimal porosity.

As the pulse off-time increases (e.g. STA 21), the size, shape and continuity of the molten zone begins to vary. The nailhead area decreases, but the width to depth ratio increases. The plasma extinguishes between pulses and the entire molten region nearly solidifies. As a result, the solidification direction becomes perpendicular to the beam and convection is reduced. Pores caused by vaporization of the alloy constituents within the keyhole are trapped by the more rapid solidification and the lessened convection.

The pulsed welds at short off-times were significantly superior to both the long off-time pulsed welds and CW welds. A more definitive study to optimize the laser weld parameters is now underway. The cracking behavior of the pulsed welds will be reported at the completion of that work.

ACKNOWLEDGMENTS

The authors would like to thank Mr. Robert J. Schwinghamer former Director of the Materials & Processes Laboratory at Marshall Space Flight Center, Alabama for providing the material, and Mr. Ernie Bayless from the same organization for discussions on welding Inconel 718. The authors would also like to thank Mr. Keven Zysk who investigated laser welding of Inconel 718 for his Master's Degree.

REFERENCES

1. H. L. Eiselstein, ASTM STP369 (1965) pp. 65-79.
2. E. A. Loria, Journal of Metals July 1988, pp. 36-41.
3. Superalloys II, ed. C. T. Sims, N. S. Stoloff, & W. C. Hagel, John Wiley & Sons, 1978.
4. R. G. Thompson, Journal of Metals July 1988, pp. 44-48.
5. C.M. Banas, Final Report, Contract NAS8-36306, 1988.
6. L.A. Weeter, C.E. Albright and W.H. Jones, Welding Research Supplement, March 1986, 51s-62s.
7. J.P. Reynolds, H.W. Kerr, P.J. Fehrenbach, 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.
8. CRC Handbook of Chemistry and Physics, 62nd Edition, R.C. Weast ed., M.J. Astle, Assoc. Ed, CRC Press, Inc., Boca Raton Florida.

Table I. Composition of Inconel 718

ELEMENT	WT%
Ni	53.0
Cr	18.6
Mo	3.10
Nb	5.00
Al	0.40
Ti	0.90
Fe	18.5
Mn	0.20
Si	0.30
C	0.04

CG-1498

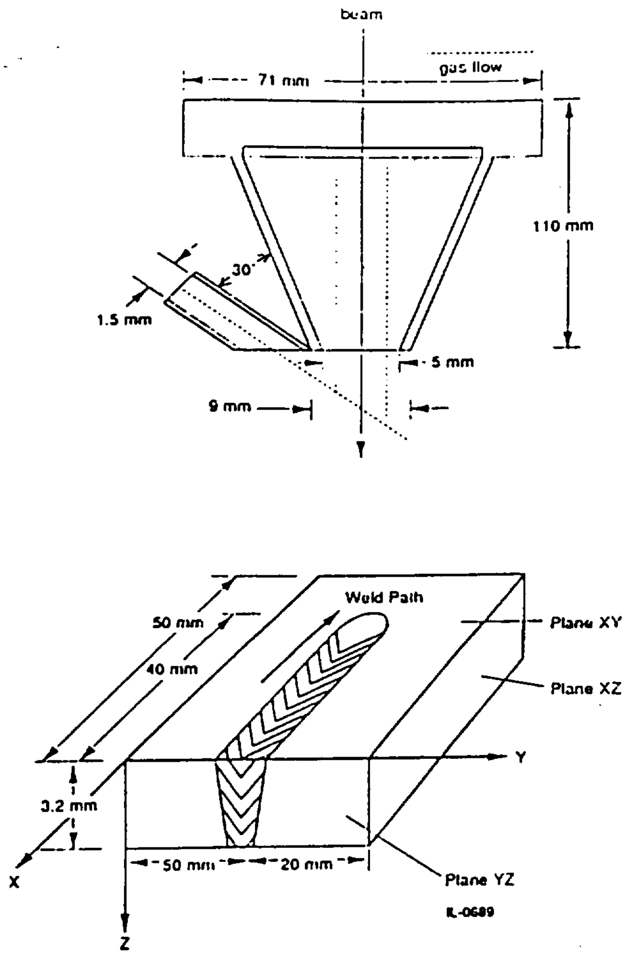


Fig. 1 - Schematic of Welding Nozzle and Sample

Table II. Inconel 718 Pulse Weld Parameters

Variable	UNITS
Weld Speed	1.5 *m/min*
Focus	1 *mm* Below XY Plane
Pulse On Time	17 *ms*
Pulse Off Time	1-21 *ms* (by 2 *ms*)
Peak Power	3.2 *kW*
Gas	Helium
Flow Rate	60 ℓ/*min*

CG-1473

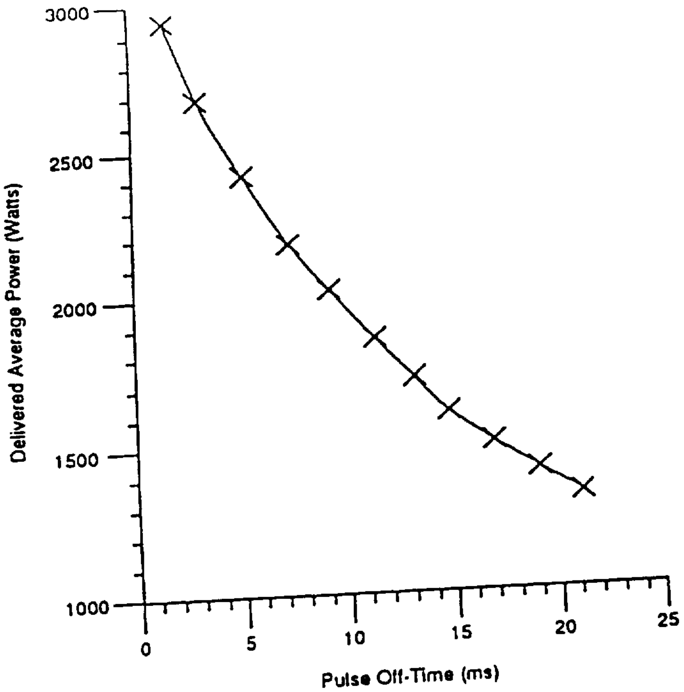


Fig. 2 - Delivered verage power output of the Rofin-Sinar laser as a function of pulse off-time.

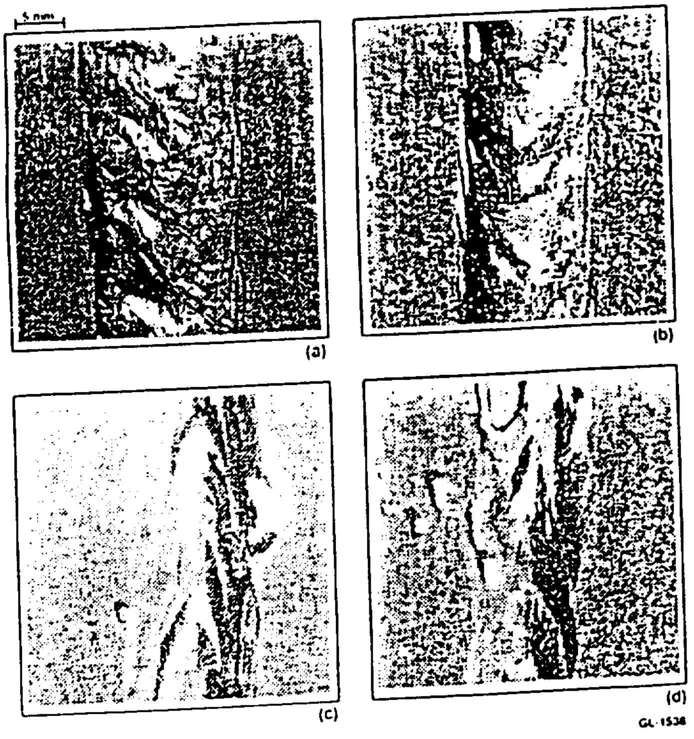


Fig. 3 - Top View of Pulsed Welds at:
a) 7 ms off-time; b) 11 ms off-time
c) 15 ms off-time; d) 21 ms off-time

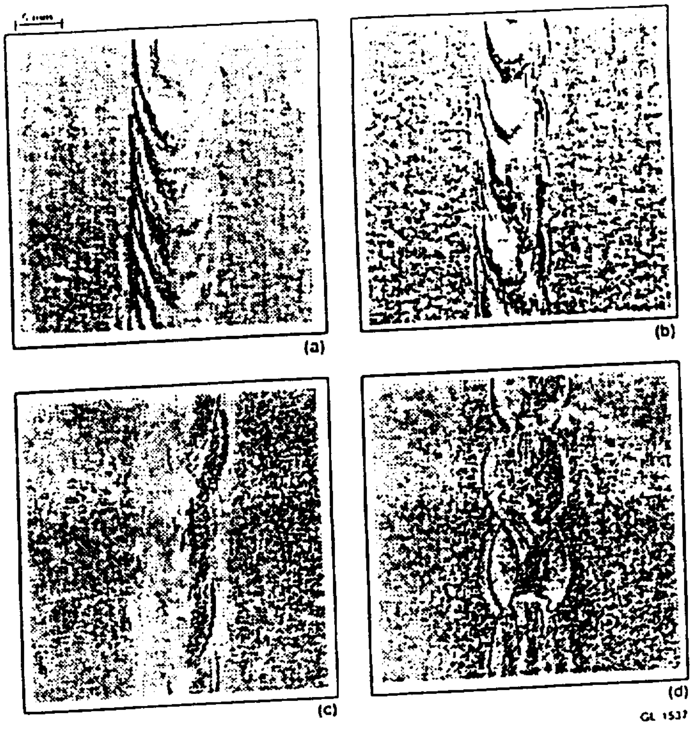


Fig. 4 - Bottom View of Pulsed Welds at:
a) 7 ms off-time; b) 11 ms off-time
c) 15 ms off-time; d) 21 ms off-time

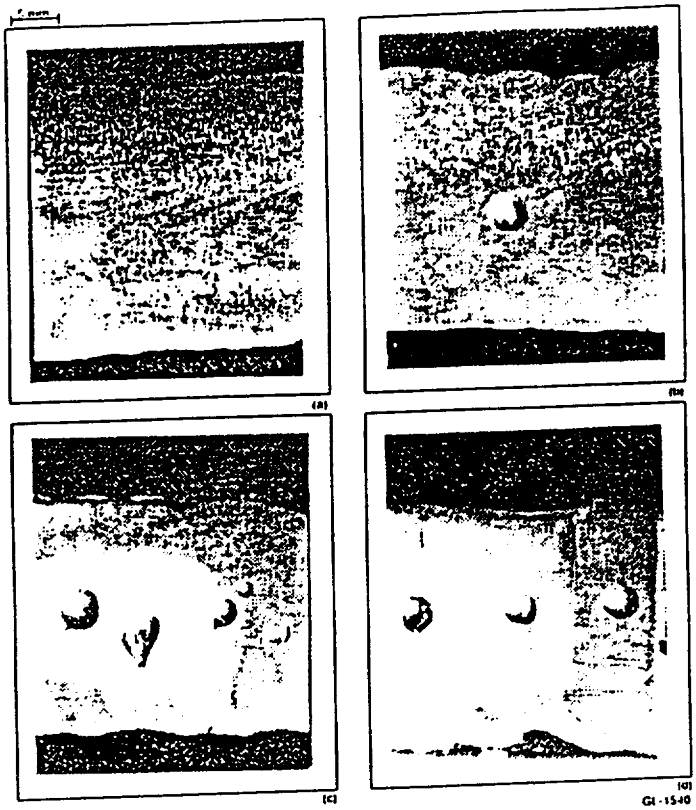


Fig. 5 - Longitudinal Centerline View of Pulsed Welds at:
a) 7 ms off-time; b) 11 ms off-time
c) 21 ms off-time; d) 28 ms off-time

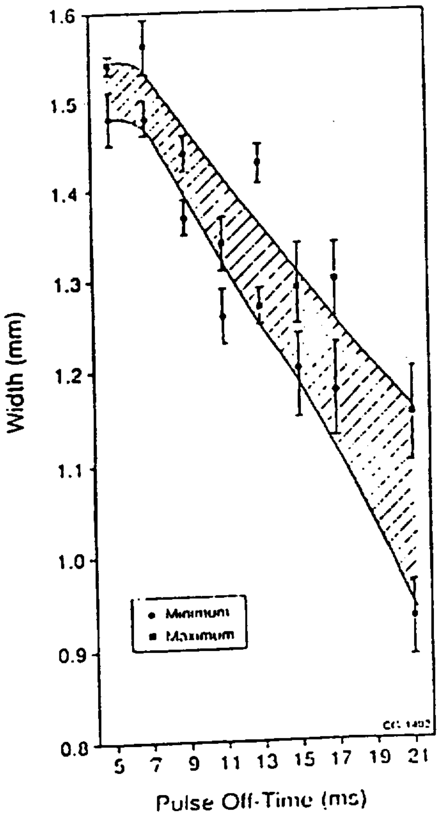


Fig. 6 - Maximum and minimum top bead widths as a function of pulse time-off.

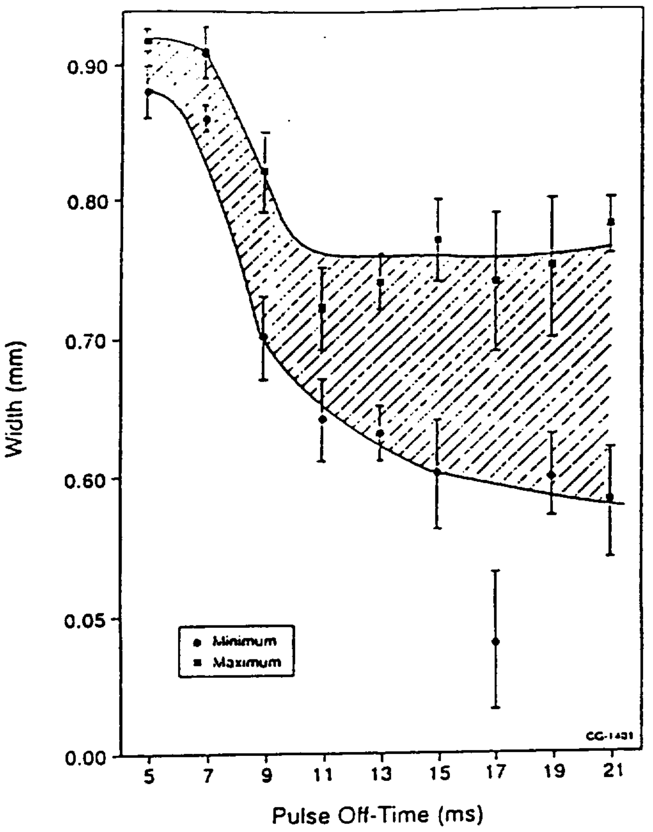


Fig 7 - Maximum and minimum bottom bead widths as a function of pulse off-time.

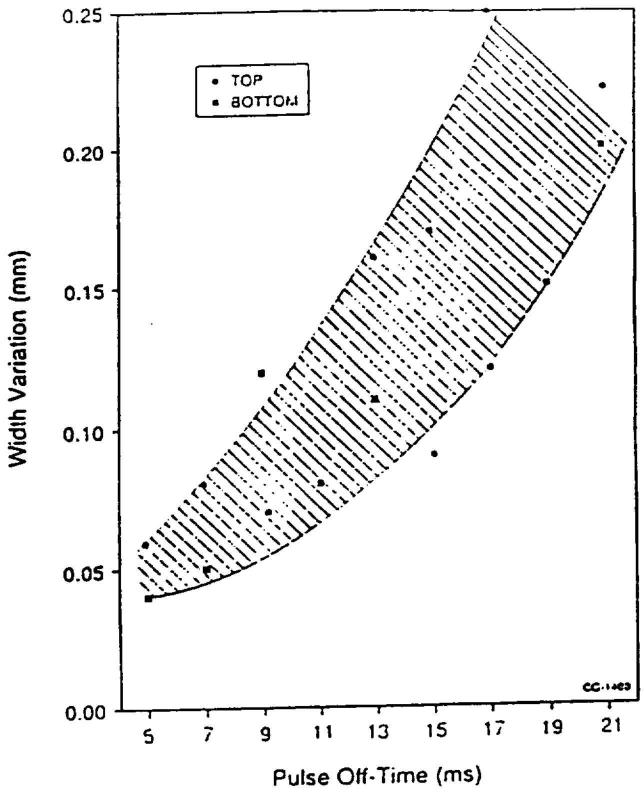


Fig. 8 - Variation in top and bottom bead widths as a function of pulseoff-time.

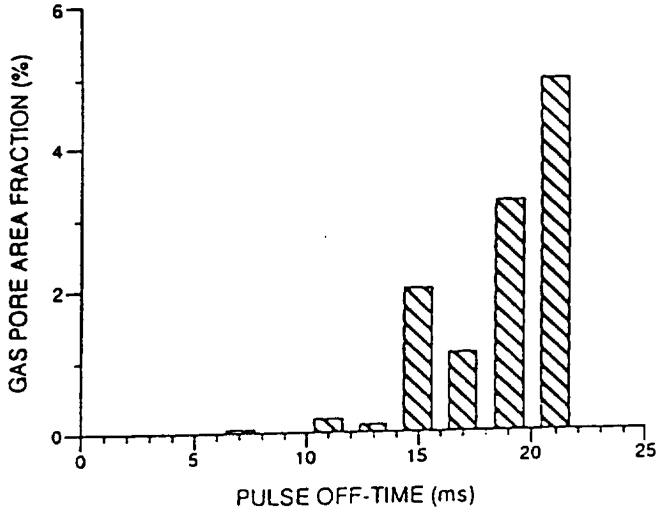


Fig. 9 - Gas pore area fraction as a function of laser pulse off-time.

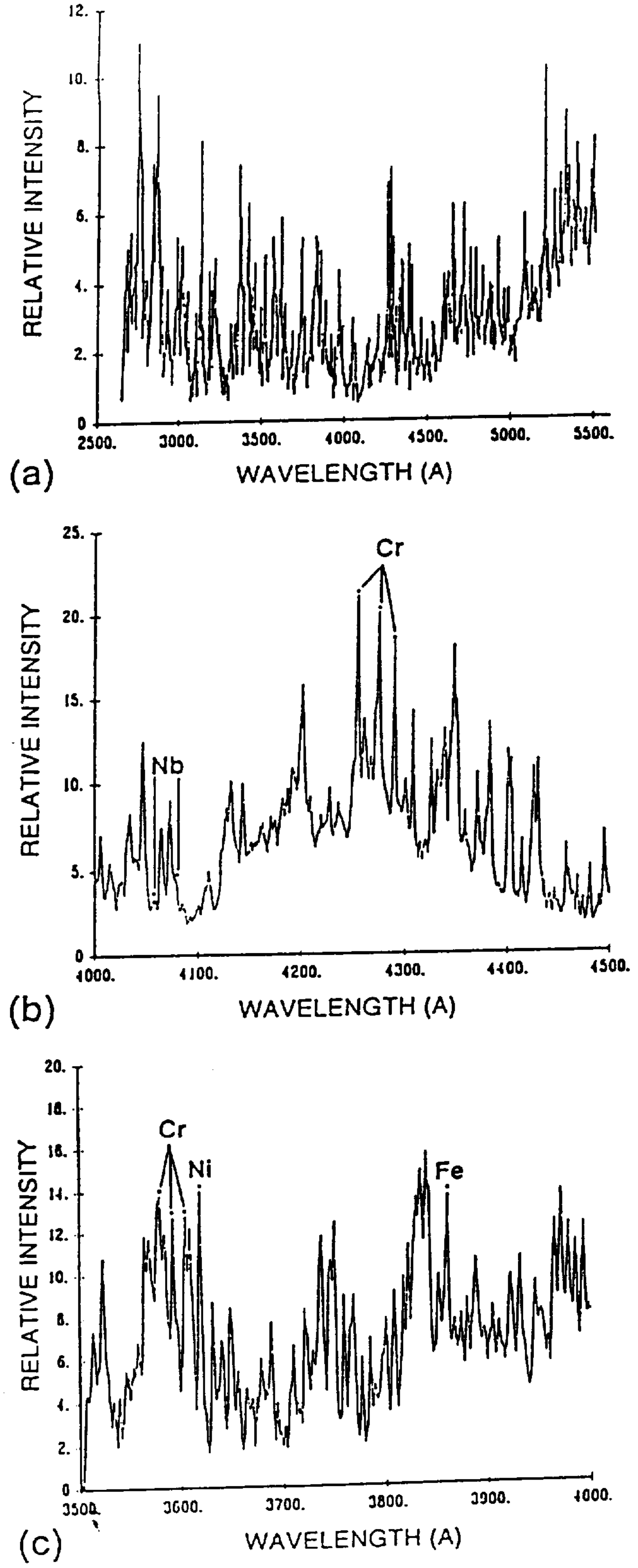


Fig. 10 - The emission spectrum of Inconel 718 in helium shield gas for (a) the spectral range of 2500 to 5000 A°, (b) the spectral range of 3500 to 4000 A°, and (c) the spectral range of 4000 to 4500 A°.

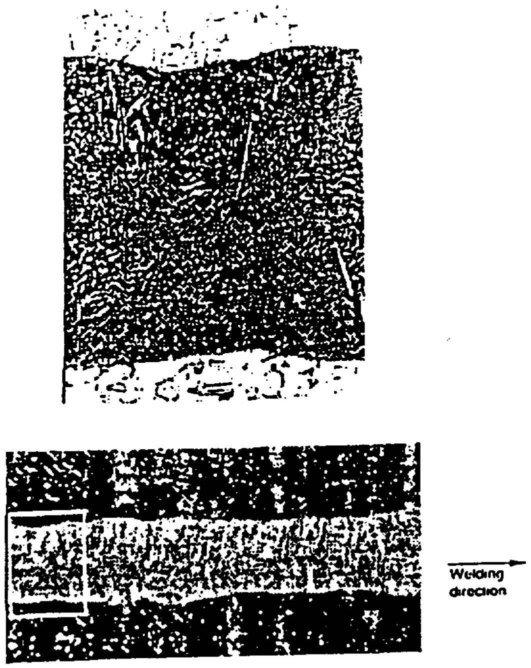


Fig. 11 - A midsection (XY plane) overview of the STA 7 weld (28X) and higher magnification (100X) view of the solidification structure.

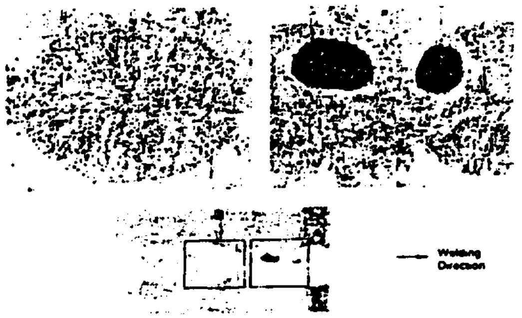


Fig. 12 - A midsection (XY plane) overview of the STA 21 weld (28X) and higher magnification (100X) views of the solidification and pore structure.

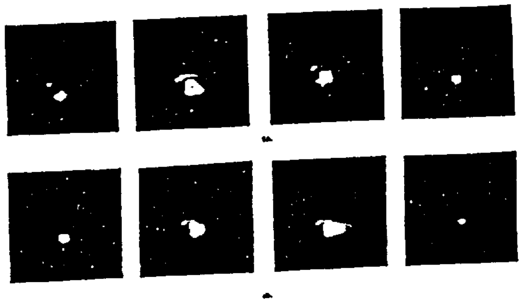


Fig. 13 - Sequential photographs of plasma shapeduring a single pulse for (a) 17 ms on / 7 ms off and (b) 17 ms on / 28 ms off.

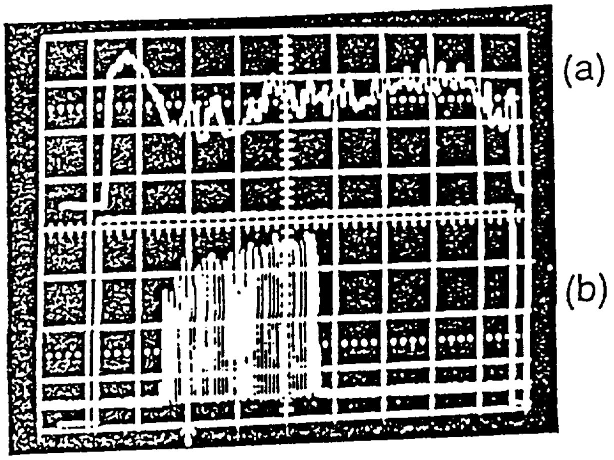


Fig. 14 - Oscilloscope trace of photodiode output during (a) single pulse and (b) multiple pulses (21 ms off-time).

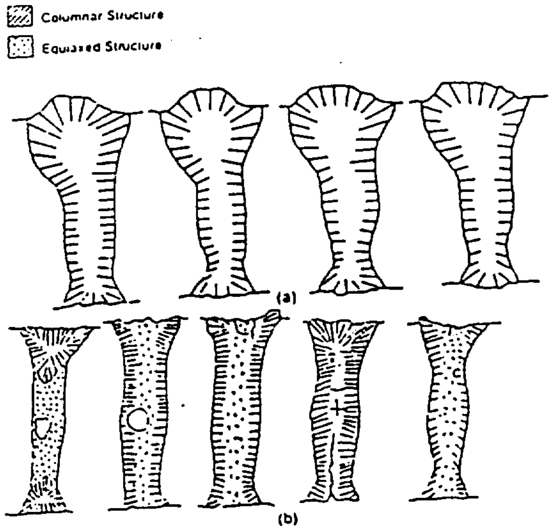


Fig. 15 - Typical cross section grain structures throughout one pulse for (a) STA 7 and (b) STA 21.

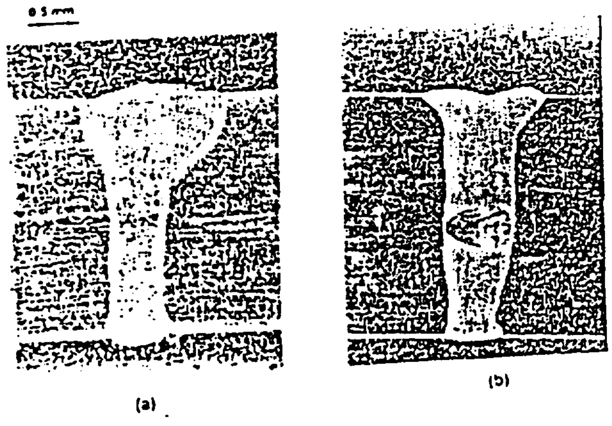


Fig. 16 - Optical micrographs of (a) CW and (b) single (17 ms) pulse laser welds.

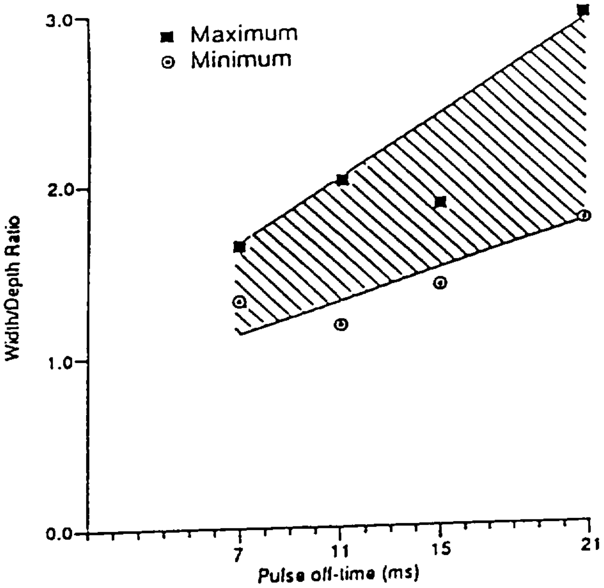


Fig. 17 - The top nailhead width to depth ratio as a function of pulse off-time.

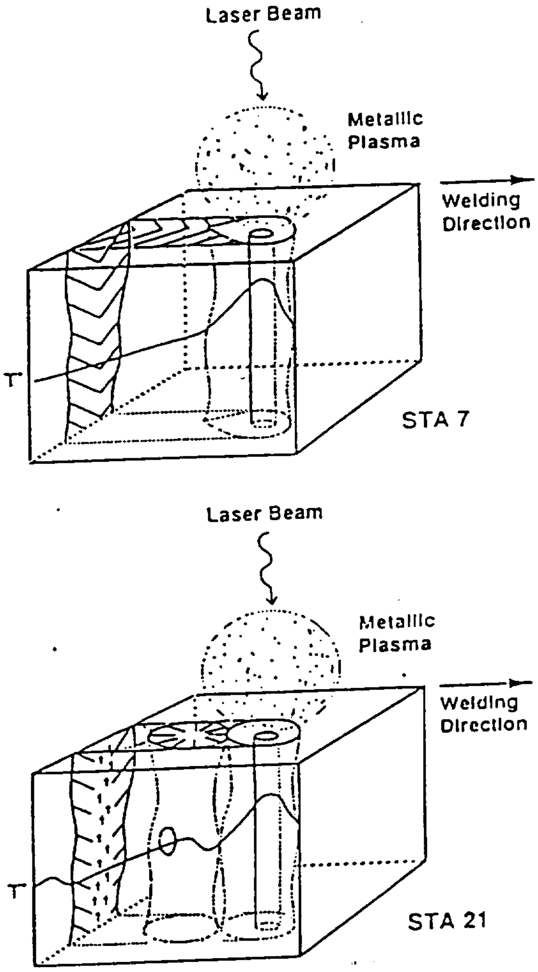


Fig. 18 - Schematic of pulsed welding process with a representative temperature profile for (a) short (7 ms) and (b) long (21 ms) duration pulse off-times.