NASA Conference Publication 3263

Dual-Use Space Technology Transfer Conference and Exhibition

Volume II

Proceedings of a conference and exhibition held at
Lyndon B. Johnson Space Center
Houston, Texas
February 1-3, 1994

NASA Conference Publication 3263

Dual-Use Space Technology Transfer Conference and Exhibition

Volume II

Compiled by
Kumar Krishen
Lyndon B. Johnson Space Center
Houston, Texas

Proceedings of a conference and exhibition held at
Lyndon B. Johnson Space Center
Houston, Texas
February 1-3, 1994

National Aeronautics and
Space Administration

1994

CONTENTS

Session P2: POWER I

Session Chair: Dave Belanger

ELECTRIC AUXILIARY POWER UNIT (EAPU) FOR THE SPACE SHUTTLE

omit

by
Michael Le
David Belanger
NASA Johnson Space Center

Presented at
Dual-Use Space Technology Transfer Conference
February 1-3, 1994

Johnson Space Center - Houston, Texas

Electric Auxiliary Power Unit (EAPU) for Space Shuttle	PROPULSION AND POWER DIVISION	
	Michael Le	2/2/94

BACKGROUND

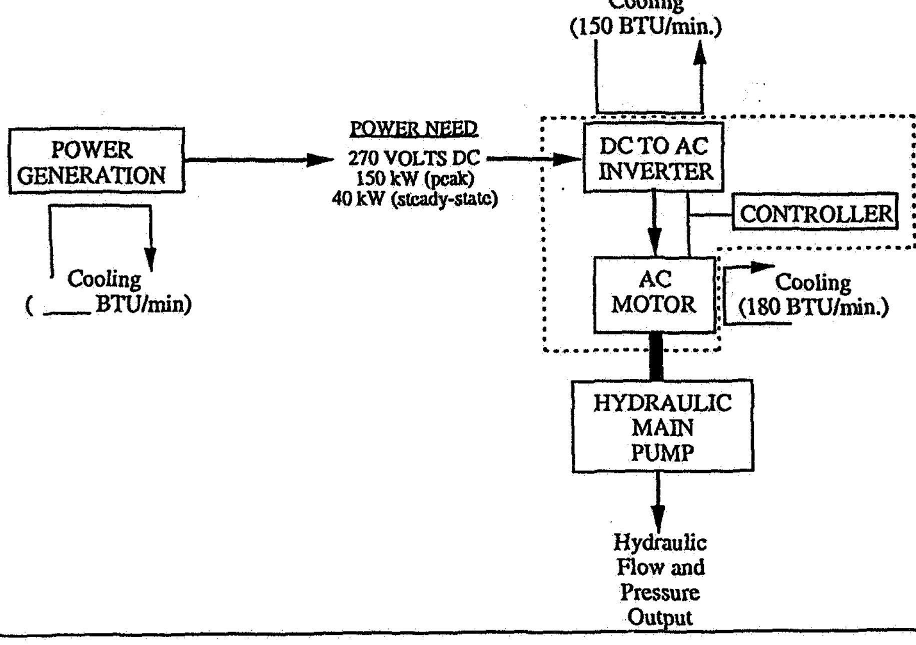
O While the hydrazine fueled Auxiliary Power Unit (APU) continues to support successful Space Shuttle flights, there have been many concepts proposed to replace the APU's for the following reasons:

- o Hazards associated with servicing and operating hydrazine fuel system
- o Extensive ground servicing and refurbishment
- o Reliability issue

O In 1992, Lamar University conducted a study to determine if an alternate power system could replace the APU favorably from a weight and performance standpoint.

- o CONCLUSION: A high power density fuel cells would trade evenly in weight with increased operational life and on-orbit power availability

ELECTRIC APU SYSTEM CONCEPT



Johnson Space Center - Houston, Texas

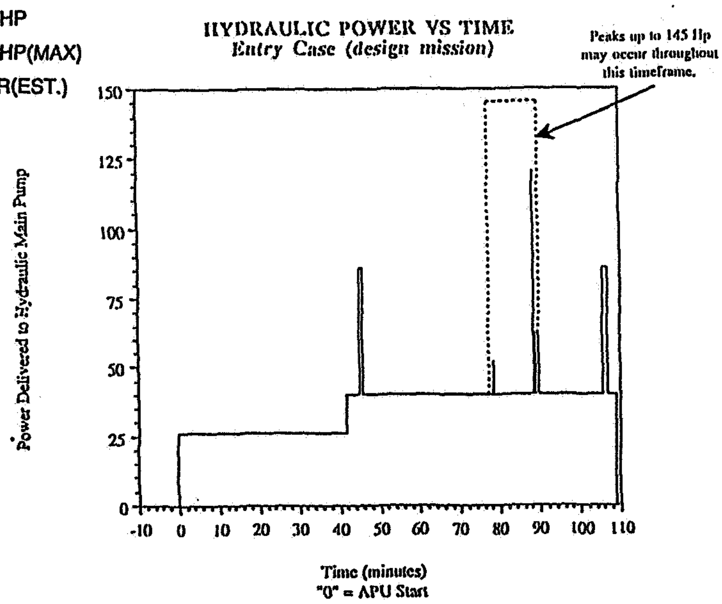
Electric Auxiliary Power Unit (EAPU) for Space Shuttle | PROPULSION AND POWER DIVISION | Michael Le | 2/2/94

EAPU Power And Energy Requirements

BASE POWER: 25-40HP

PULSE POWER: 105 HP(MAX)

ENERGY: 64.2 KW-HR(EST.)



Johnson Space Center - Houston, Texas

Electric Auxiliary Power Unit (EAPU) for Space Shuttle	PROPULSION AND POWER DIVISION	
	Michael Le	2/2/94

EAPU- TEST OBJECTIVES

O TO DEMONSTRATE FEASIBILITY OF REPLACING THE HYDRAZINE FUELED APU WITH AN ELECTRIC DRIVEN SYSTEM.

- o DEMONSTRATE FEASIBILITY OF A HIGH CURRENT DENSITY FUEL CELL AS POWER SOURCE
- o OBTAIN DESIGN AND OPERATIONAL PARAMETERS FOR THE EAPU SYSTEM COMPONENTS
- o IDENTIFY ADDITIONAL DEVELOPMENT NEEDS

O TO PROVIDE TRAINING OPPORTUNITIES FOR ENGINEERS

O TO DEVELOP TESTING CAPABILITIES NOT PRESENTLY IN EXISTENCE

O TO PROVIDE POTENTIAL TECHNOLOGY TRANSFER OPPORTUNITIES

Johnson Space Center - Houston, Texas

Electric Auxiliary Power Unit (EAPU) for Space Shuttle	PROPULSION AND POWER DIVISION	
	Michael Le	2/2/94

EAPU - TEST PROGRAM DESCRIPTION

* 4-CELL SUBSTACK FUEL CELL OFF LIMIT TEST

- INVESTIGATE OPERATIONAL LIMITS OF THE EXISTING ORBITER FUEL CELL AT THE ENERGY AND POWER DEMANDS REQUIRED TO OPERATE THE HYDRAULIC SYSTEM

* 1.3 KW-HR PROTOTYPE FLYWHEEL SYSTEM TEST

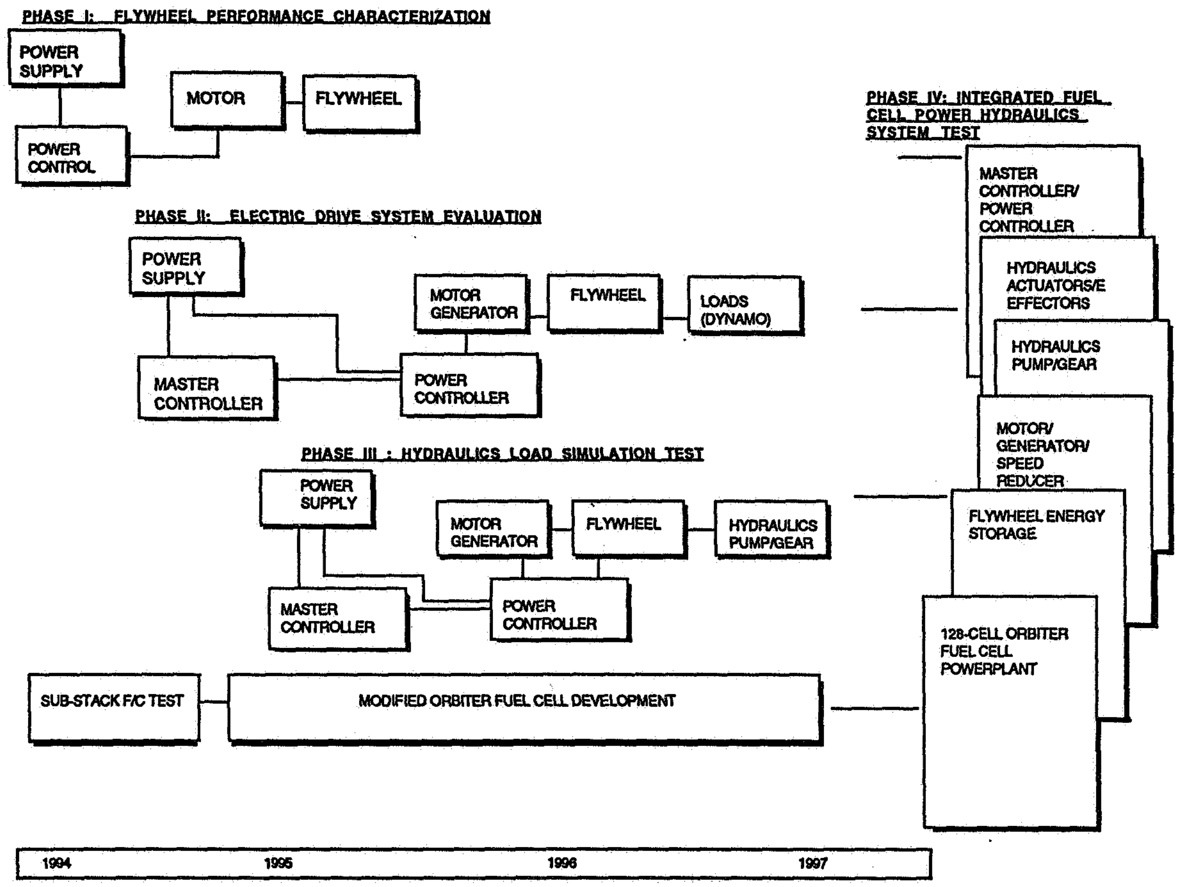
- OBTAIN DESIGN AND OPERATIONAL PARAMETERS
- DETERMINE THE ABILITY OF THE FLYWHEEL TO PROVIDE PEAK POWER AS DEMANDED BY AN ELECTRIC MOTOR

* HYDRAULIC SYSTEM LOADS TEST

- DEFINE HYDRAULICS LOADS AND RATES PROFILE
- DEVELOP POWER AND ENERGY REQUIREMENTS
- DETERMINE COOLING REQUIREMENTS
  - MOTOR/CONTROLLER/SPEED REDUCER

* INTEGRATED EAPU SYSTEM TEST

- DEMONSTRATE FUNCTIONAL AND PERFORMANCE CAPABILITY OF AN EAPU



Johnson Space Center - Houston, Texas

Electric Auxiliary Power Unit (EAPU) for Space Shuttle	PROPULSION AND POWER DIVISION	
	Michael Le	2/2/94

EAPU STATUS

o FUEL CELL TEST

- Initiated check-out testing of the test support equipment
- Anticipate power-on test in February '94

o MOTOR/GENERATOR

- Commercial 25HP motor is being procured; delivery in Mar. '94

o 1.3 Kw-Hr PROTOTYPE FLYWHEEL TEST

- Flywheel assembly and test being completed
- Plan to deliver to JSC in Feb-Mar. '94

DUAL-USE TECHNOLOGY OPPORTUNITY

* Lessons learned from this test could be used to aid in a design and development of an electric car or bus system.

  o Performance and operational characteristics of the fuel cells, flywheel system, power management system

  o Effective use of government facilities and experiences in the technology transfer process

Session P3: POWER II

Session Chair: Dave Belanger

S1-35
411760

A PC BASED TIME DOMAIN REFLECTOMETER FOR SPACE STATION CABLE FAULT ISOLATION

Michael Pham and Marty McClean
Propulsion and Power Division
NASA/JSC Houston, Texas

Sabbir Hossain, Peter Vo and Ken Kouns
Propulsion and Power Department
Lockheed Engineering and Sciences Co.
Houston, Texas

KEYWORDS

Time Domain Reflectometer, Fault Detection and Isolation, Space Station Cable

ABSTRACT

Significant problems are faced by astronauts on orbit in the Space Station when trying to locate electrical faults in multi-segment avionics and communication cables. These problems necessitate the development of an automated portable device that will detect and locate cable faults using the pulse-echo technique known as Time Domain Reflectometry. A breadboard time domain reflectometer (TDR) circuit board was designed and developed at the NASA-JSC. The TDR board works in conjunction with a GRiD lap-top computer to automate the fault detection and isolation process. A software program was written to automatically display the nature and location of any possible faults. The breadboard system can isolate open circuit and short circuit faults within two feet in a typical space station cable configuration. Follow-on efforts planned for 1994 will produce a compact, portable prototype Space Station TDR capable of automated switching in multi-conductor cables for high fidelity evaluation. This device has many possible commercial applications, including commercial and military aircraft avionics, cable TV, telephone, communication, information and computer network systems.

This paper describes the principle of time domain reflectometry and the methodology for on-orbit avionics utility distribution system repair, utilizing the newly developed device called the Space Station Time Domain Reflectometer (SSTDR).

INTRODUCTION

The utility distribution system for the Space Station delivers essential fluid and avionics utilities to all system elements such as the nodes, habitation and laboratory module, and many of Freedom's external orbital replacement units (ORU's). The utilities are distributed with the aid of two integrated fluid and electrical utility trays per segment of the Space Station connected together at termination points of trays. These trays provide environmental protection to the lines from contamination and damage from micrometeroid/orbital debris (MMOD), atomic oxygen (AO), and ultraviolet (UV) rays. By design, the utility avionics line connections are to be performed by EVA using EVA-compatible Zero-G electrical connectors. Maintainability and repairability of these utility and avionics lines are major concerns.

Design requirements were identified early in the development of the Space Station. For utility distribution, the design must meet the following requirements:

- minimize on-orbit installation and maintenance time
- allow for EVA assembly while meeting EVA time constraints
- provide accessibility for on-orbit repair.

In general, the electrical cable on-orbit may fail in two different ways; (1) a short circuit caused by damaged wire insulation, bent connector pins and metallic particles in connectors, or (2) an open circuit caused by a broken wire due to EVA damage or meteoroid impact or an open connector.

In the event of such a cable failure, the space station contractor proposes to use the following EVA maintenance scenarios:

Avionics systems that have suffered damage will be removed end-to-end from the utility tray and a replacement line will be installed end-to-end. The sections of "the remove and replace units" of one truss segment vary in length from 23 to 45 ft. Avionics line maintenance is considered as a one crew member task. The EVA crew-member would perform the steps shown in the timeline of Table 1. The timeline shows that the remove and replace scenario of one avionics line segment is about 41 minutes. And this will be repeated until the exact failed segment is found and replaced. Some systems have as many as 6 segments to be replaced. This remove and replace approach would be EVA intensive and has extremely high cost for logistics of all avionics lines segments.

Table 1
Timeline for avionics line removal and replacement task

No.	Crew	Task description	Assumptions	Time (Min-sec)
1	EV1	Open tray cover of segment end exposing Segment -to-segment connection of avionics lines	In position on PWP Secure tray cover in open position	1:00
2	EV1	Disconnect 0-g connector of one end of failed line		0:30
3	EV1	Translate along length of segment tray, opening tray covers and releasing clamps on failed line.	Translate via SSRMS; Loosen lines from clamps	5:00
4	EV1	Disconnect 0-g connector of second end of failed line		0:30
5	EV1	Remove failed line		10:00
6	EV1	Connect 0-g connector of replacement line		0:30
7	EV1	Translate along length of segment tray, installing line in tray, closing line clamps, and closing tray covers.		15:00
8	EV1	Connect 0-g connector of second end of replacement line		0:30
9	EV1	Close tray covert		
		Estimated Task Time with 20% overhead		41:24

NASA/JSC has developed an alternate technique to pin-point the location of the fault. It is based on the time domain reflectometer (TDR) concept. The TDR technology has been around for years. Several companies such as Tektronix, Biddle Instruments, Hewlett Packard, Cabletron, Anristu offer such devices. But they are industrial or commercial units, and none would meet the requirements of the space station environment. Hence, a prototype TDR system was developed at JSC, referred in this paper as the Space Station Time Domain Reflectometer (SSTDR), that is compact, portable for astronaut usage and has the ability to pin-point faults in a multi-conductors cable.

Unlike any TDR device on the market, the SSTDR is integrated with a lap-top computer. The computer in the SSTDR system is used to store cable and connector data, and to interpret the TDR waveform. Because, the software which is used for controlling the operation of the TDR and manipulation of data, is stored on the computer hard disk, the future program changes can be made with little or no hardware change. Thus program improvements and program modifications to fit specific type cable or cable configuration needs may be implemented easily. The biggest advantage of a PC based TDR is that the base line waveform and configuration of a cable can be stored on the computer hard disk or optical disk which can be retrieved for comparison with the real time data of the cable under test.

The SSTDR system also contains a switch matrix so that any set of two conductor cables can be selected from a bundle of cable under computer command. The ease of use, the flexibility, the custom programmability, and the ability to locate electrical faults multi-conductor cable in a faster and more efficient manner, make the SSTDR a unique instrument for cable fault detection and location.

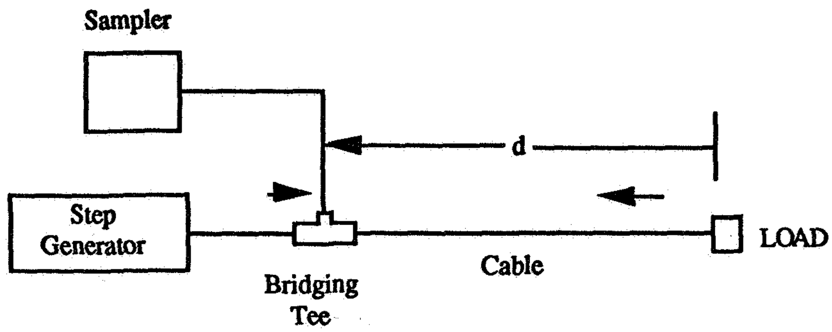
This paper describes in detail the Space Station Time Domain Reflectometer. It also describes the implementation and capabilities of the system and gives insight into potential growth features.

SSTDR PRINCIPLES OF OPERATION

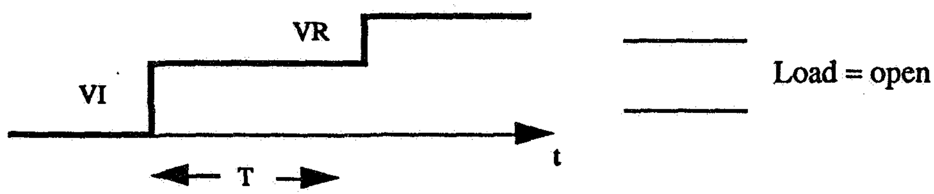
A time domain reflector works on the same principles as radar. Ultra-fast rise time voltage steps (Vi) are sent down the cable under test. The step signal travels along the cable until it finds an impedance mismatch causes a reflection. The reflection returns to the cable end and is received by a sampler circuit (Figure 1).

If the reflection is caused by a short or an open in the cable, a voltage step is reflected back. The reflected voltage (Vr) is superimposed on the advancing initial step and will appear as a step-up or step-down transition on the display, depending on whether it is reflected in-phase or out-of-phase with respect to the initial step. The reflected voltage is step-up for an open circuit and step-down for a short circuit fault.

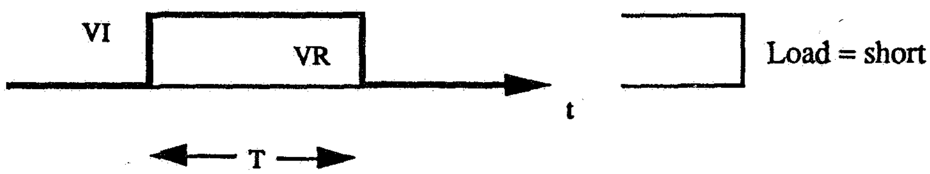
The position of the mismatch will appear at a point on the reflection-time data directly proportional to the linear position of the cause. The distance of the mismatch, the fault, is derived from measuring the time lapse, in nanoseconds, between the sending of the step and the reception of its reflection. Because the approximate velocities of propagation of such step functions in cables of different construction are known, the distance to the fault in feet is the product of the measured time lapse and the known velocity of propagation in feet per nanosecond. Because the pulse travels "there and back", it is convenient to use half the value of the velocity of propagation.



(a) A basic TDR Set-up



(b) Open Circuit: Total reflection is in-phase at the load



(c) Short Circuit: Total reflection is out-of-phase at the load

Figure 1 (a)-(c): Principle of Fault Location using TDR.

SSTDR HARDWARE

The SSTDR design is the only time domain reflectometer system that uses a portable GRiD lap-top computer integrated with the TDR circuit. The GRiD computer contains the high level software which controls the operation of the TDR, the data base for the cable under test, and the base line cable signature. The TDR circuit board consists of a microcontroller, a pulse generator, a timebase circuit, a sampling circuit and a switching circuit. A block diagram of the hardware used in the pre-prototype SSTDR is shown in Figure 2.

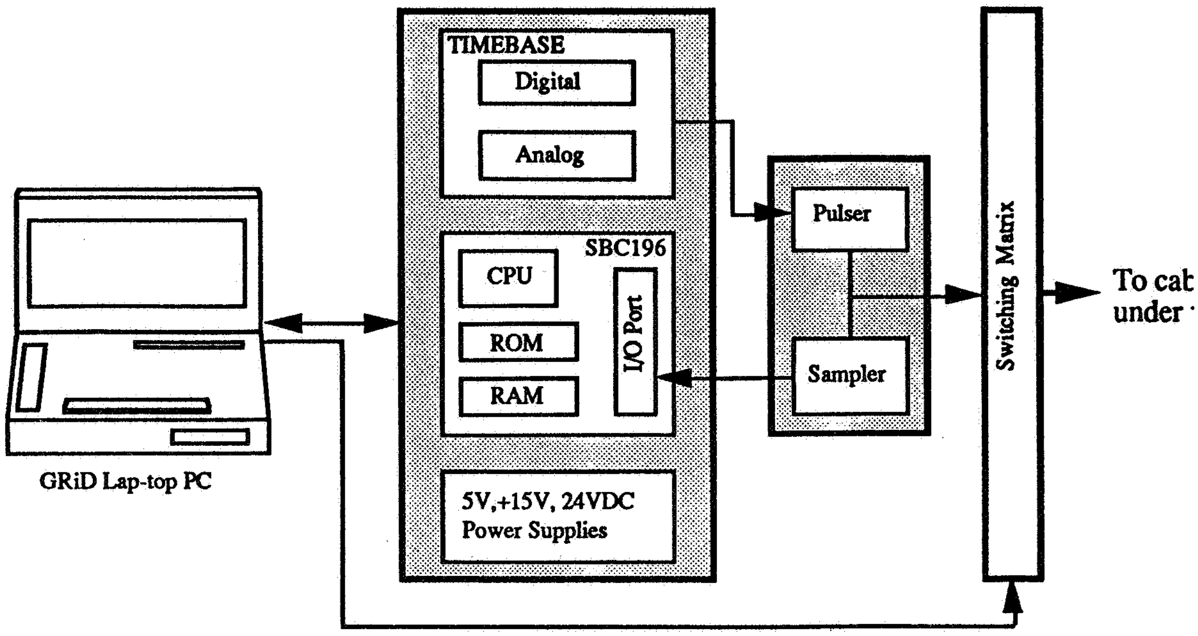


Figure 2: A Block diagram of SSTDR system.

The pulse generator produces a 25 microsecond, 600 milivolt pulse every 250 microseconds. The pulse rise time of 1 nanosecond gives a fault location resolution of approximately 1/3 of a foot.

The timebase circuit generates precisely timed strobes to the pulse driver and sampler circuit. The digital portion of the timebase contains a 20 MHz precision clock and a programmable digital counter to produce a pulse every 250 microseconds. The output of the digital counter is used to trigger a delay counter which provides 50 nanosecond resolution to the sampler time delay. The end of the delay counter signal is used to generate a timebase interrupt request to the processor to inform that a sample is being taken. The output of the delay counter is also provided to an analog ramp circuit for further control delay. The voltage of the ramp circuit is compared to the output of a digital to analog converter such that, every time the comparator produces a sampling pulse, it is delayed by 1 nanosecond from the previous pulse. Thus, the TDR waveform can be broken down into discrete samples, and each advanced in time by 1 nanosecond.

The sampling pulse activates the sample/hold circuit. The signal from the sample hold circuit is fed to the 10 bit A/D converter of the Intel 80C196KC microcontroller. The A/D conversion is activated by an interrupt request signal from the digital timebase delay circuit. The single board microcontroller operates at 16 MHz. A 128K EPROM is used to store the system and application software and a 64K RAM to store temporary data and to perform all necessary functions of the TDR board. The communication to the host computer (GRiD) is done via a RS-232 port.

The host computer is a 386SL, 25MHz, lap-top system manufactured by GRiD Inc. The software for data manipulation, TDR control, and graphic display is resident on the hard disk.

A computer controlled switching circuit is used to select a particular set of wires from a cable bundle. The operator can select any pair of cables by entering appropriate connector pin numbers.

SSTDR SOFTWARE

The SSTDR software consists of two separate components - the software resident in the TDR board microcontroller and the other in the GRiD computer. The software in the TDR board is written in FORTH language and is primarily used to control the timing function of the pulse generation and sampling circuit, data acquisition and for temporary storage of the waveform data (Figure 3).

The host software resident in the GRiD computer is written in ADA language. It stores the data, processes it and displays the wave form. Besides data manipulation, a graphical user interface was developed to interface with the operator. The operator can select the mode of operation, cable configuration, type of connector and particular set of cables to be tested. The flow chart for operation sequence of the SSTDR is given in the appendix.

The SSTDR has two operation modes: manual mode and automated mode. The manual mode enables the user to select any particular pair of wires and analyze the cable. In the manual mode, the SSTDR can also be used as a general purpose cable analyzer. In the automated mode, however, the operator can test a specific space station cable configuration. In this mode, the SSTDR , under software control, tests all individual cables in sequence. By using a switching matrix, the SSTDR sequentially switches through all conductors in the cable system and tests them.

In both of these modes, the SSTDR can automatically detect and locate any fault if the data base is resident in memory for a particular cable under test. The data bases consists of nominal information about the cable under test and acceptable tolerance windows for impedances of these cable paths. The data base may also incorporate a base line waveform of a fault-free cable which can be used for comparison with the waveform from a cable under test to perform fault diagnosis.

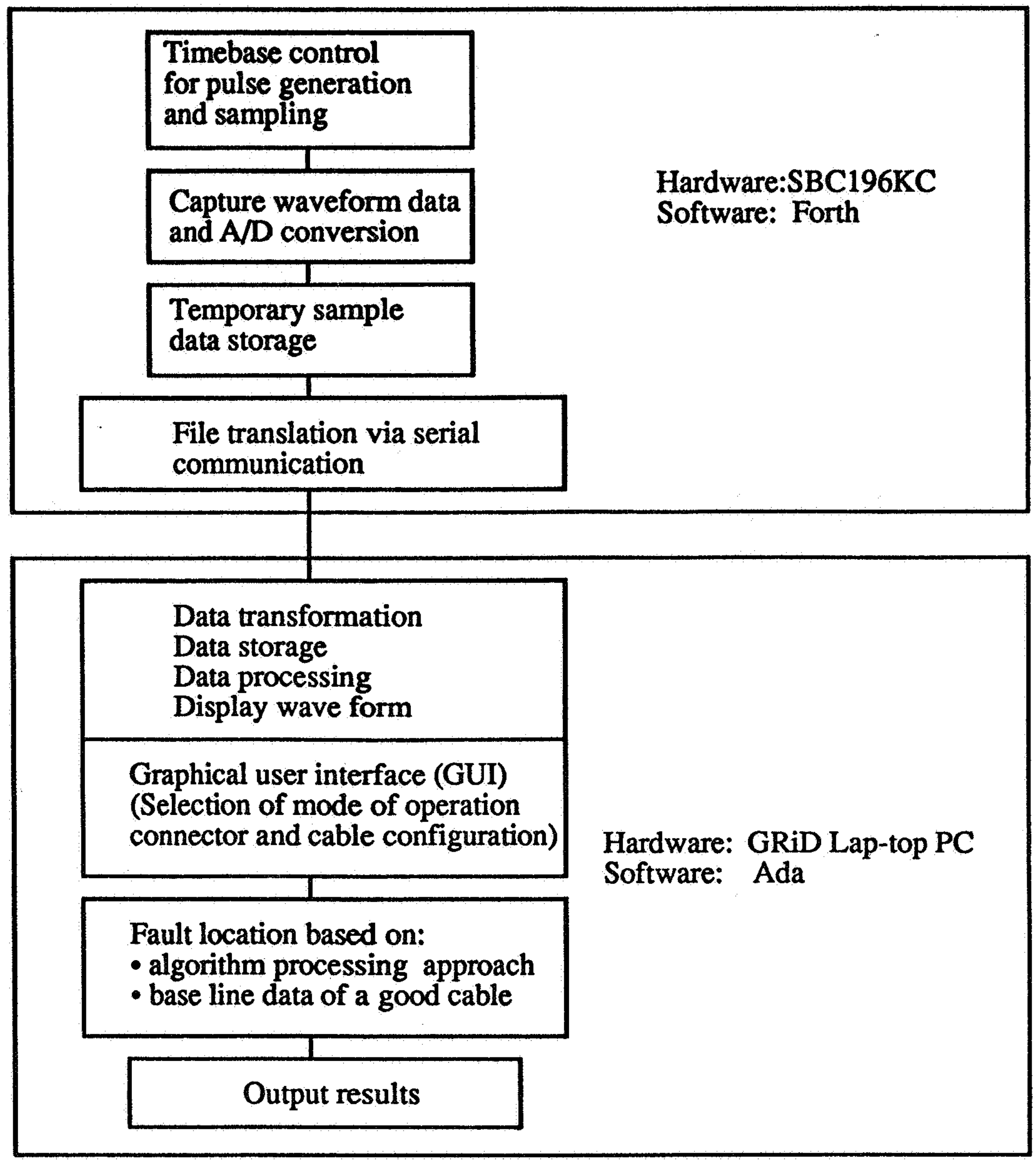
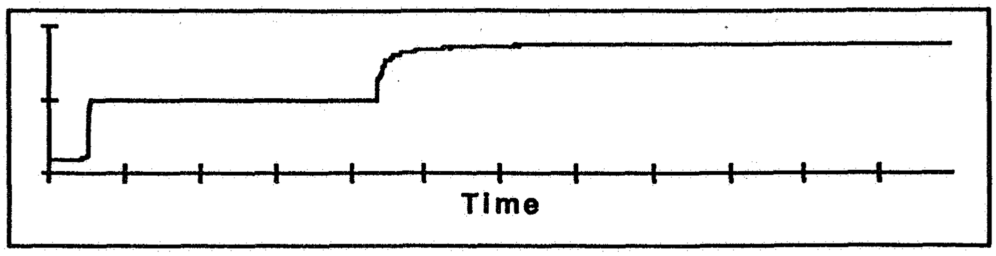


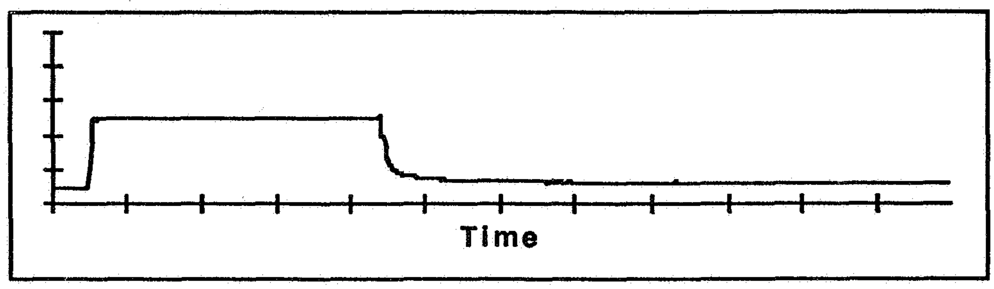
Figure 3: General system software structure.

LABORATORY TEST RESULTS

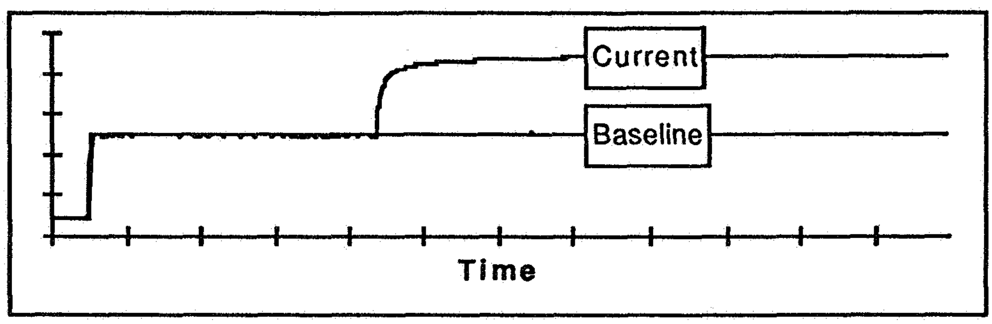
During the course the SSTDR development, cables with known fault locations were tested with a prototype unit and both hardware and software were verified for their accuracy in estimating the location of the fault. Figure 4 (a)-(3) shows some of the results obtained during our test run. The test results show that a typical open or short faults can be located within a couple of feet.



(a) Case 1: Open circuit fault.



(b) Case 2: Short circuit fault.



(c) Current waveform superimposed on baseline showing an open circuit fault.

Figure 4(a)-(c): Examples of waveforms obtained in the laboratory tests.

CONCLUSION

Time Domain Reflectometry can be successfully applied to detection and location of faults in a multi-conductor cable system. The PC based SSTDR unit can successfully locate faults to within a couple of feet. The advantage of a PC based system that any future enhancement of the software can be easily installed in the system. The custom feature of this system lends itself to many commercial and diversified applications. including commercial and military aircraft avionics, cable TV, telephone, communication, information and computer network systems.

REFERENCES

Clegg, B. and Lord, N.G.: "Modern cable-fault-location methods," Proc. IEE, Vol. 122, No. 4, April, 1975.

Scherer, R., Ray, R. and Trush V.: "A Portable Scalar Network Analyzer for the Flight Environment," Application Notes, Lockheed Sanders, Inc.

Kuan, K.K. and Warwick, K.: "Real-time expert system for fault location on high voltage underground distribution cables," Proc. IEE, Vol. 139, No. 3, May 1992.

P 409

SPACE BATTERIES AND THEIR ROLE IN DUAL USE TECHNOLOGY

Bobby J. Bragg and Betsy Kluksdahl
NASA Johnson Space Center
Battery Office
Mail Code EP5
Houston, TX 77058

Dual Use Space Technology Transfer Conference
NASA Johnson Space Center
Houston, TX

February 1 -3, 1994

Space Batteries Have a Multi-directional Technology Transfer

SOURCES OF SPACE BATTERIES

- Commercial Cells - Technology transfer into NASA
- Specially Contracted Cells - Technology transfer out of NASA

COMMERCIAL CELLS

- Technology from commercial sector is transferred into NASA
- Alkaline, Zinc-Air, and Ni-Cd cells used in instruments, cameras, radios, etc. in crew compartment
- Ni-MH cells being investigated for hazard issues prior to allowing use in crew compartment
- Other new chemistries evaluated as they arise for applicability to NASA

SPECIALLY CONTRACTED CELLS

- NASA contracts with industry for cells specifically designed for NASA use
- Commercial sizes or level of hazard protection require special NASA design modification to provide required size and/or sufficient hazard protection
- Research contracts to develop new cell sizes and/or chemistries

---

NASA-JSC Specific Cell Designs

SOURCES

- Direct contract with JSC battery office
- Small Business Innovative Research Program

DIRECT CONTRACT WITH BATTERY OFFICE

- Mutually beneficial contract between NASA and vendor
  - Li-BCX cells - needed specific cell design for desired rates and required hazard protection
  - Investigation into feasibility of thermally sensitive thin film that could be incorporated into jellyroll cell designs to prevent internal cell shorts
  - Research into innovative solid-state cell
  - Development of large capacity, high rate zinc-air cells

NASA-JSC Specific Cell Designs (continued)

SMALL BUSINESS INNOVATIVE RESEARCH (SBIR) PROGRAM

- Contract with the government through the SBIR program but monitored by the battery office
  - Development of lithium polymer cells
  - Development of rechargeable zinc-air cells
  - Development of rechargeable nickel-zinc cells
  - Development of improved metal hydrides for better nickel-metal hydride (Ni-MH) cells
  - Development of rechargeable silver-metal hydride (Ag-MH) cells

---

Features of NASA-JSC Specific Cells

Cell Chemistry	Voltage [V]	Capacity [Ah]	Energy Density [Wh/kg]
Li-BCX C, D, and DD cells Available and in use.	3.2	6, 13, & 25	350
High Rate Zinc-Air In development and under test.	1.1	30	350
Large Capacity Zinc-Air In development and under test.	1.2	200	440
Rechargeable Zinc-Air In development - values are estimates.	1.2	15	150 - 200
Lithium polymer In development - values are estimates.	3 - 4	2 - 10	75
Silver-Metal Hydride In development - values are estimates.	1.1	50	80

Battery Test Capabilities at JSC

FACILITY AVAILABLE FOR CELL AND BATTERY TESTING

- Acceptance
  Open Circuit Voltage, Load Voltage, Weight, Dimensional Check

- Electrical Characterization and Performance
  Capacity Discharge (constant current or constant resistance), Calorimetry, Shelf-life

- Abuse
  Vibration, Mechanical Shock, Short Circuit, "Smart Short," High Temperature Exposure

- Certification for Flight
  Combination of tests mentioned above

<u>Battery Technology Transfer is Mutually Beneficial</u>

- NASA benefits from the research and development activities of industry when commercial battery technology is transferred into NASA.

- Industry benefits from NASA and Government research funds that allow the development of new and/or improved battery technology that can be used by NASA and then commercialized by industry.

Session P4: POWER III

Session Chair: Bill Boyd

52-44
44763

PROTON EXCHANGE MEMBRANE FUEL CELLS FOR SPACE AND ELECTRIC VEHICLE APPLICATIONS - FROM BASIC RESEARCH TO TECHNOLOGY DEVELOPMENT

Supramaniam Srinivasan, Sanjeev Mukerjee, A. Parthasarathy, A. Cesar Ferreira, Masanobu Wakizoe, Yong Woo Rho, Junbom Kim, Renaut A. Mosdale, Ronald F. Paetzold, James Lee, Shinichi Hirano and A. John Appleby
Center for Electrochemical Systems and Hydrogen Research
Texas Engineering Experiment Station
Texas A&M University System
College Station, TX 77843-3402

ABSTRACT

The proton exchange membrane fuel cell (PEMFC) is one of the most promising electrochemical power sources for space and electric vehicle applications. The wide spectrum of R&D activities on PEMFCs, carried out in our Center from 1988 to date, is as follows (1) Electrode Kinetic and Electrocatalysis of Oxygen Reduction; (2) Optimization of Structures of Electrodes and of Membrane and Electrode Assemblies; (3) Selection and Evaluation of Advanced Proton Conducting Membranes and of Operating Conditions to Attain High Energy Efficiency; (4) Modeling Analysis of Fuel Cell Performance and of Thermal and Water Management; and (5) Engineering Design and Development of Multicell Stacks. The accomplishments on these tasks may be summarized as follows:

1. A microelectrode technique was developed to determine the electrode kinetic parameters for the fuel cell reactions and mass transport parameters for the H_2 and O_2 reactants in the proton conducting membrane.

2. High energy efficiencies and high power densities were demonstrated in PEMFCs with low platinum loading electrodes (0.4 mg/cm^2 or less), advanced membranes and optimized structures of membrane and electrode assemblies, as well as operating conditions.

3. The modeling analyses revealed methods (i) to minimize mass transport limitations, particularly with air as the cathodic reactant; and (ii) for efficient thermal and water management.

4. Work is in progress to develop multi-kilowatt stacks with the electrodes containing low platinum loadings.

INTRODUCTION

The proton exchange membrane fuel cell (PEMFC) is one of the most promising electrochemical power sources for space and electric vehicle applications. As illustrated in Figure 1, this type of fuel cell has the best prospects for attaining high energy efficiencies and high power densities. The other advantages are:

1. It's a low temperature fuel cell system, and, thus, the start-up time could be quite fast.

2. It uses a polymeric electrolyte (a perfluorinated sulfonic acid membrane) which is made conducting, by the absorption of only water.

3. This type of fuel cell has very good prospects for long lifetimes, as has been demonstrated by a 100,000 hour lifetime in single cells.

The wide spectrum of R&D activities on PEMFCs, carried out in our Center from 1988 to date, is represented in the flow chart. In the following sections, presented are the highlights of the work which has been carried out to date and also in progress.

ELECTRODE KINETICS AND ELECTROCATALYSIS OF OXYGEN REDUCTION

The slow kinetics of the oxygen reduction reaction is the main cause of efficiency losses in fuel cells. This loss in efficiency also has an effect of reducing the attainable power density. However, in order to have a viable power system, a minimum efficiency (in terms of power density) of 40% to 50% is desirable. To elucidate the electrode kinetics of oxygen reduction at the platinum Nafion interface, a microelectrode technique (1) was developed in our Center. The schematic of the cell is shown in Figure 2. This technique was used to determine the electrode kinetic parameters for oxygen reduction on platinum as well as of the diffusion coefficients and of the solubilities of oxygen in the electrolyte. The results showed that the Tafel behavior for oxygen reduction is very similar to that in aqueous acid electrolytes but that the exchange current densities were higher at the platinum proton exchange membrane interface. As in the case of aqueous electrolytes, reaction order for the reaction with respect to oxygen is unity. The studies at different temperatures yielded the activation energy for the reaction which was again of the same order as the aqueous electrolytes. This study also was shown to yield the diffusion coefficient and solubilities of oxygen in the membrane. It was found that the diffusion coefficients about an order of magnitude lower in the polymer electrolytes, but the solubilities are, however, considerably higher. However, the product of the diffusion coefficient times solubility is higher in the polymer electrolytes than in the aqueous electrolytes, thus accounting for lesser mass transport problems in the proton exchange membrane fuel cells. An interesting aspect of the microelectrode study is that the electrode kinetic parameters which were obtained on the smooth platinum microelectrode were within an order of magnitude of those obtained on the high surface area electrocatalyst in the fuel cell electrodes in PEMFCs.

In a separate study, platinum alloy electrocatalysts were evaluated for the oxygen reduction reaction in proton exchange membrane fuel cells. This involved comparison of the electrocatalytic activity for oxygen reduction reaction (ORR) on carbon supported binary platinum alloys (Pt/Cr, Pt/Mn, Pt/Fe, Pt/Co, Pt/Ni) with that of a conventional Pt/C electrocatalyst (all electrodes containing the same Pt loading, 0.3 mg/cm^2). Figure 3 shows the *i*R corrected Tafel plots indicating the enhancement in electrocatalytic activities towards ORR exhibited by all the binary platinum alloys (2). In addition to the electrode kinetic studies in proton exchange membrane fuel cells, *in situ* X-ray absorption spectroscopic studies were conducted to determine the electronic and geometric properties of the alloys at potentials in the fuel cell operating range. Of all the alloys, the platinum chromium alloy has shown the highest enhancement in electrocatalytic activity. The results of the electrochemical and X-ray absorption spectroscopic studies showed correlations between the electrocatalytic activities and the geometric as well as electronic properties of the alloys, i.e. Pt-Pt bond distances and Pt d-band vacancies (Pt 5 *d* orbitals) in the form of volcano type relationships.

OPTIMIZATION OF STRUCTURES OF ELECTRODES AND OF MEMBRANE AND ELECTRODE ASSEMBLIES

Following the work of Srinivasan and co-workers at Los Alamos National Laboratory from 1985 to 1988, a wide variety of studies were carried out in our Center to enhance the performance of proton exchange membrane fuel cells for space and electric vehicle applications. These studies consisted of:

1. A determination of the effect of the platinum content on the fuel cell performance.

2. Preparation of electrodes with different Teflon contents in the active and diffusion layer and their evaluation in fuel cells.

3. Changing the conditions of hot-pressing during the preparation of the membrane and electrode assemblies.

The effect of increase of platinum content on the performance is akin to that of increase of pressure and can be attributed to the increase in electrocatalytic activities. However, the performances obtained even with a very ultra low platinum loading of 0.1 mg/cm^2 was quite good. The effective utilization of the platinum with the ultra low platinum loading was quite high (3). Figure 4 illustrates the effect of platinum loading on the power density vs. current density plot. This figure shows that at current density of about 500 mA/cm^2 the amount of platinum required per kilowatt is only 1.2 grams per kilowatt while with the state-of-the-art electrodes it is about 10 times this value. Optimization of the Teflon content in the diffusion layer and the catalytic layer is therefore of vital importance. The Teflon content of about 50% in the active layer and 40% in the diffusion layer showed the best performance in our studies (4). In respect to the optimization of the conditions for the preparation of the membrane and electrode assemblies, it was found that a high temperature is more favorable for attaining a better performance in fuel cells. The temperature also depends on the glass transition temperature for the membrane. With some experimental membranes from the Asahi Chemical Company, it was found that even though the glass transition temperature is around 140° hot-pressing at about 155 - 160° produces a better behavior.

SELECTION AND EVALUATION OF ADVANCED PROTON CONDUCTING MEMBRANES AND OF OPERATING CONDITIONS TO ATTAIN HIGH ENERGY EFFICIENCIES AND HIGH POWER DENSITIES

A detailed study of the effect of the membranes and of the operating conditions was carried out using DuPont's Nafion®, Dow experimental and Asahi Chemical's Aciplex®-S membranes (5). A comparison of the performance of fuel cells with the three types of membranes is shown in Figure 5. The results of the studies such as the one as shown in this figure as well as those at different temperatures and pressures have been analyzed. Correlations have been drawn with the physicochemical properties of the membranes. A typical example is shown in Table I. One of the most important physicochemical parameters is the water content of the membrane. (The membranes with higher water contents have higher sulfonic acid contents and thus show better performances in PEMFCs.) The conductivity is directly related to the water content of the membrane and attaining the maximum conductivity is essential from the point of view of reducing the slope of the linear region in the current potential plot.

MODELING ANALYSIS OF FUEL CELL PERFORMANCE AND OF THERMAL AND WATER MANAGEMENT

One of the toughest challenges in developing proton exchange membrane fuel cells is to reduce the mass transport overpotentials particularly when air is used as the cathodic reactant. For space applications there is hardly any choice other than using pure hydrogen and pure oxygen as reactants but for terrestrial applications air is the unique choice for the cathodic reactants. A comparison of the performances of PEMFCs with H_2/O_2 and H_2/air as reactants illustrates the significant problems when air is used. Further, for terrestrial applications such as electric vehicles, it is essential to work at very low pressures (ideally at 1 atm pressure). Mass transport overpotentials have an effect on the slope of the linear region of the cell potential vs. current density plot and also can cause the departure of this plot from linearity at low current densities. A systematic study was conducted in our Center to elucidate other problems causing these effects (6). In this study, where O_2/He, O_2/Ar and O_2/N_2 gas mixtures were used, the cell potential current density plot with these gas mixtures is shown in Figure 6. This figure shows that with a lighter inert gas, He, the mass transport effects are somewhat less than with Ar or N_2. Ar, N_2 and air exhibit the same type of current potential behavior. The mathematical analysis of this behavior showed that the increase of the slope in the linear region is due to the mass transport problem in the catalyst layer while the departure from linearity is caused by problems by films of water or droplets of water in the diffusion layer.

A problem which is related to the mass transport problem is also thermal and water management. At the present time, the fuel cell performance is quite satisfactory at low to intermediate current densities, but with the need for operating the fuel cells at relatively high power densities, it is necessary to operate at higher current densities. As stated in the previous section, the proton conducting membrane operates most satisfactorily when the water content has its maximum value. In the fuel cells which have been developed to date, the humidification of the membrane has been shown to be important. At the higher current density, it is also necessary to remove a considerable amount of heat. For instance, if the fuel cell were to operate at about 2 A/cm^2, the electricity generation corresponds to 1.2 watts/cm^2 while the heat generation rate is about 1.8 watts/cm^2. A modeling analysis was carried out of the thermal and water management problems (7). Different methods of cooling (air cooling, liquid cooling and evaporative cooling) were examined. The results showed that when air is used as the cathodic reactant, the evaporative cooling method appears to be most beneficial (Table II).

ENGINEERING DESIGN AND DEVELOPMENT OF MULTI-CELL STACKS

During the period from about 1984 to date, considerable progress has been made in attaining high power densities and high energy densities in single cells. The work at Ballard Power Systems, Inc. has also demonstrated that high levels of performance could be obtained in multi-cell stacks. At Ballard, the fuel cells have been developed using electrodes with high platinum loading. At the present time, the toughest challenge is of an engineering nature, i.e. developing multi-cell stacks with low platinum loading electrodes and also finding solutions to the mass transport problems and thermal and water management. Our Center is currently engaged in the development of multi-cell stacks (1 and 10 kW) using electrodes made in-house. The electrodes which would be used will be fabricated either using the rolling or the spraying method. Due consideration of the flow patterns in the bipolar plate, humidification of the reactant gases and of the operating conditions (temperature, pressure, flow rate) will be taken into account. Modeling analysis of the multi-cell stack in respect to the electrochemical performance, thermal and water management are also being carried out. For the 1 kW stack, the electrode area will be 150 cm^2, whereas for the two 10 kW stacks, the electrode areas would be either 300 or 600 cm^2.

CONCLUSIONS

The accomplishments on these tasks may be summarized as follows:

1. A microelectrode technique was developed to determine the electrode kinetic parameters for the fuel cell reactions and mass transport parameters for the H_2 and O_2 reactants in the proton conducting membrane.

2. Three fold activity enhancements for ORR have been demonstrated with Pt/cr alloys electrocatalysts. Studies using *in-situ* XAS spectroscopy have revealed important foundamental insights into ORR electrocatalysis.

3. High energy efficiencies and high power densities were demonstrated in PEMFCs with low platinum loading electrodes (0.4 mg/cm^2 or less), advanced membranes and optimized structures of membrane and electrode assemblies, as well as operating conditions.

4. The modeling analyses revealed methods to (i) minimize mass transport limitations, particularly with air as the cathodic reactant; and (ii) for efficient thermal and water management.

5. Work is in progress to develop multi-kilowatt stacks with the electrodes containing low platinum loadings.

ACKNOWLEDGEMENTS

The work summarized in this paper was carried out with financial support from the U.S. Defense Advance Research Project Agency, NASA Johnson Space Center, NASA Lewis Research Center, NASA Sponsored Center for Space Center, Texas A&M University, Energy Research Applications Program - Texas Higher Education Coordinating Board, Energy Research Corporation, Asahi Chemical Industry Company, Ltd., Mazda R&D of North America, Mechanical Technologies, Inc., and Korea Gas Corporation.

REFERENCES

1. Parthasarathy, A., Martin, C.R. and Srinivasan, S., J. ELECTROCHEM. SOC., Pennington, NJ, Vol. 138, 1991, pp. 916.

2. Mukerjeev, S. and Srinivasan, S., J. ELECTROANAL. CHEM., Vol. 357, 1993, pp.201.

3. Ferreira, A.C., Srinivasan, S. and Appleby, A.J., J. ELECTROCHEM. SOC., Pennington, NJ, Extended Abstracts, 181st Meeting of the Electrochemical Society, St. Louis, MO, May 1992, Vol. 92-1, 1992, pp. 11.

4. Ferreira, A.C., Srinivasan, S. and Appleby, A.J., J. ELECTROCHEM. SOC., Pennington, NJ, Extended Abstracts, 183rd Meeting of the Electrochemical Society, Honolulu, Hawaii, May 1993, Vol. 93-1, 1993, pp. 46.

5. Wakizoe, M., Velev, O.A., Srinvasan, S. and Appleby, A.J., J. ELECTROCHEM. SOC., Pennington, NJ, Extended Abstracts, 182nd Meeting of the Electrochemical Society, Toronto, Ontario, Canada, October, 1992, Vol. 92-2, 1992, pp. 156.

6. Rho, Y.W., Velev, O.A., Srinivasan, S. and Appleby, A.J., J. ELECTROCHEM. SOC., Pennington, NJ, Extended Abstracts, 182nd Meeting of the Electrochemical Society, Toronto, Ontario, Canada, October 1992, Vol. 92-2, 1992, pp. 732.

7. Koch, H., Nandi, A., Anand, N.K., Velev, O.A., Swan, D.H., Srinivasan, S. and Appleby, A.J., J. ELECTROCHEM. SOC., Pennington, NJ, Extended Abstracts, 178th Meeting of the Electrochemical Society, Seattle, WA, October 1990, Vol. 90-2, 1992, pp. 176.

Table I. Physico-chemical Characteristics of Perfluorinated Sulfonic Acid Polymer Membranes for PEMFCs

Membrane	Equiv. Weight, g/mols SO^{-}_{3}	Thickness in dry state, μm	Water Content, %	Conductivity, Ω^{-1} cm^{-1}
Dow	800	125	69	0.118
Aciplex®-S	1000	120	43	0.108
Nafion®	1100	100	34	0.059

Membrane	Eo, V	Tafel slope, mV/dec	Resistance, Ω cm^2	i_{900} mA/cm^2
Dow	0.983	62	0.106	25
Aciplex®-S	1.017	63	0.111	50
Nafion®	0.990	52	0.168	40

Table II. Mass Flow of Ambient Air Required for Cooling PEMFC

Current Density	Cell Potential (V)	Heat to be Removed	# of Stoichiometric Mass of Air
200	0.702	3.488	1.42
400	0.604	8.384	2.34
600	0.513	14.527	2.71
800	0.426	21.881	3.00
1000	0.340	30.430	3.40

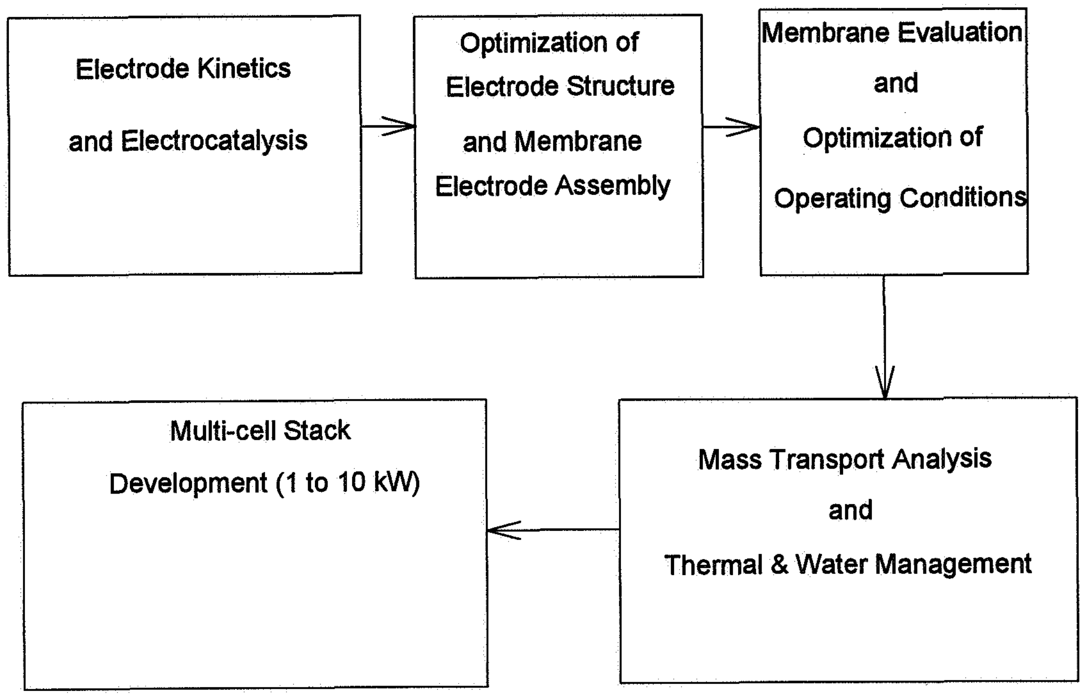


Figure 1 Flowchart of Activities on Proton Exchange Membrane Fuel Cell Project at Texas A&M University.

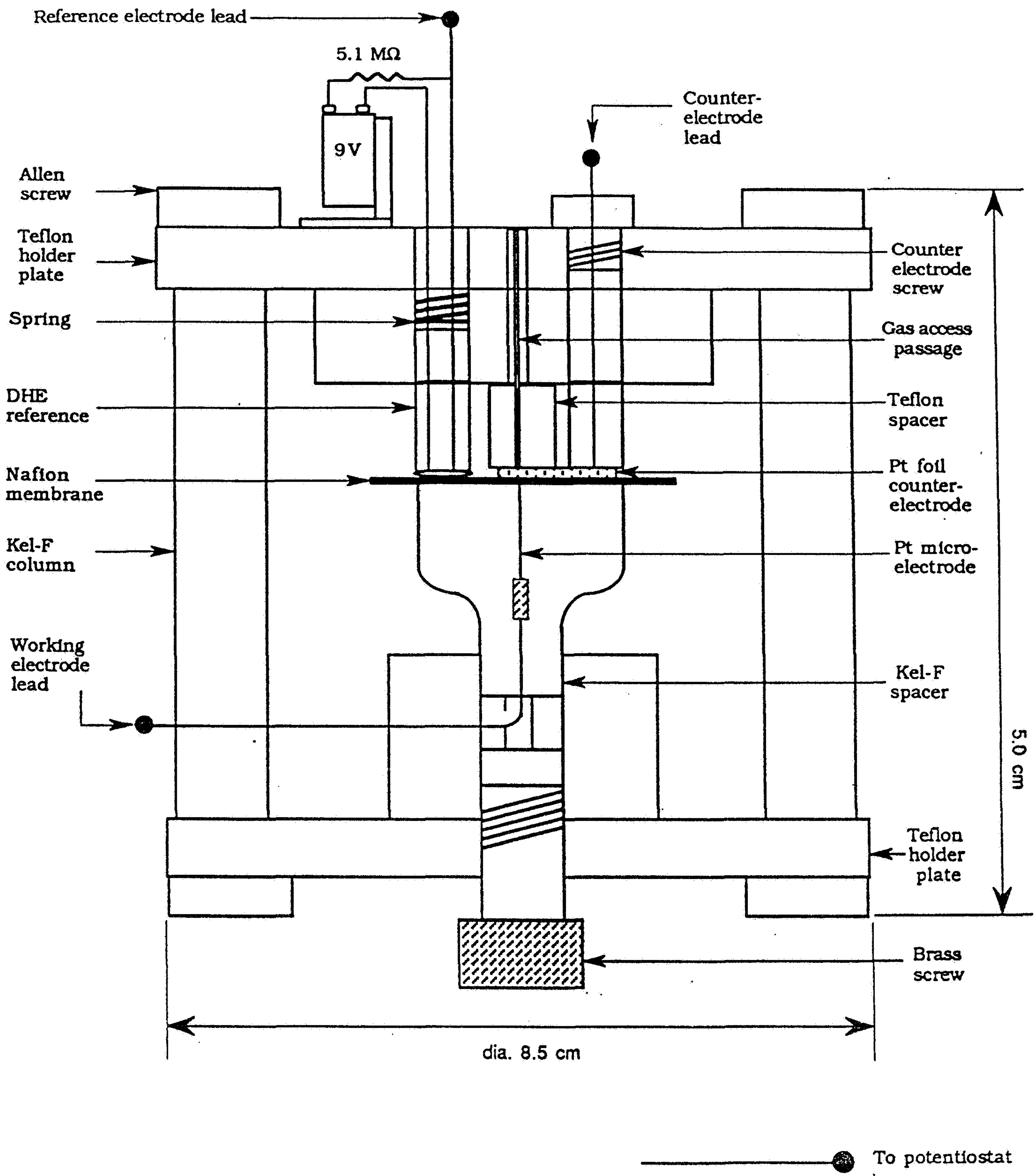


Fig. 2. Schematic of a micro-electrode set up for evaluation of mass transport and electrode kinetic parameters in a PEM environment

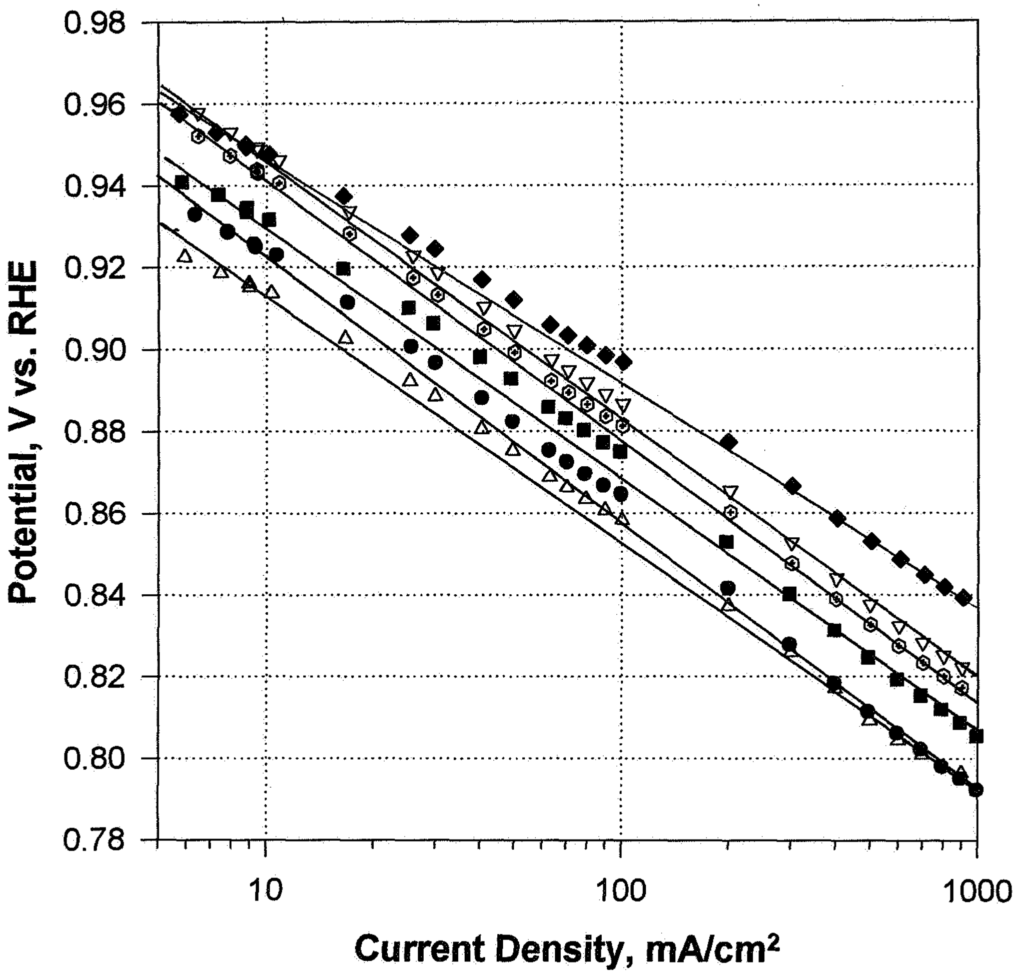


Figure 3 iR-corrected Tafel plots for oxygen reduction in proton exchange membrane fuel cells at 95°C and 5 atm pressure for Pt (ETEK) and Pt alloy electrocatalyts (JMRC). Pt loading on electrodes, 0.3 mg/cm^2. Pt/C (Δ), Pt/Cr (◆), Pt/Mn (⊕), Pt/Fe (∇), Pt/Co (■) and Pt/Ni (●).

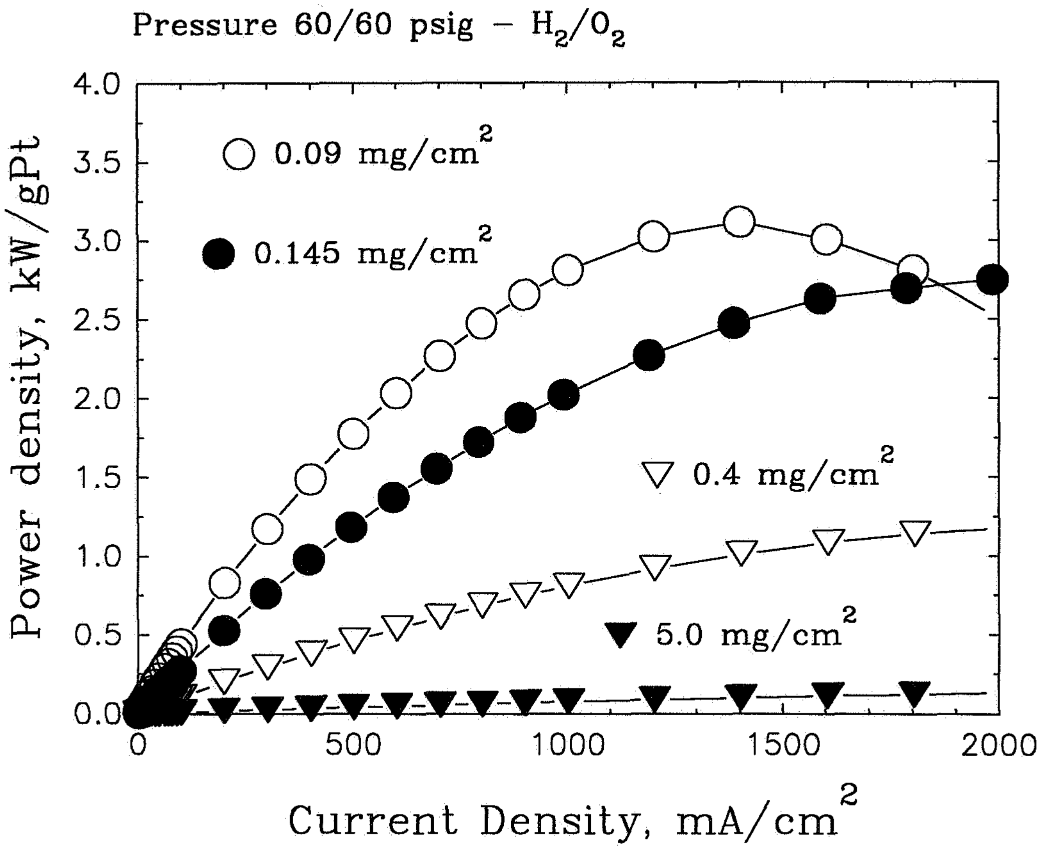


Figure 4. Effect of Pt loading in the electrode on power vs current density in PEMFC (○) 0.090 and (●) 0.145 mg Pt/cm^2-CESHR; (▽) 0.4 mg Pt/cm^2 ETEK, Inc.; (▼) 5 mg Pt/cm^2-Johnson & Mathey at 95°C and 5 atm.

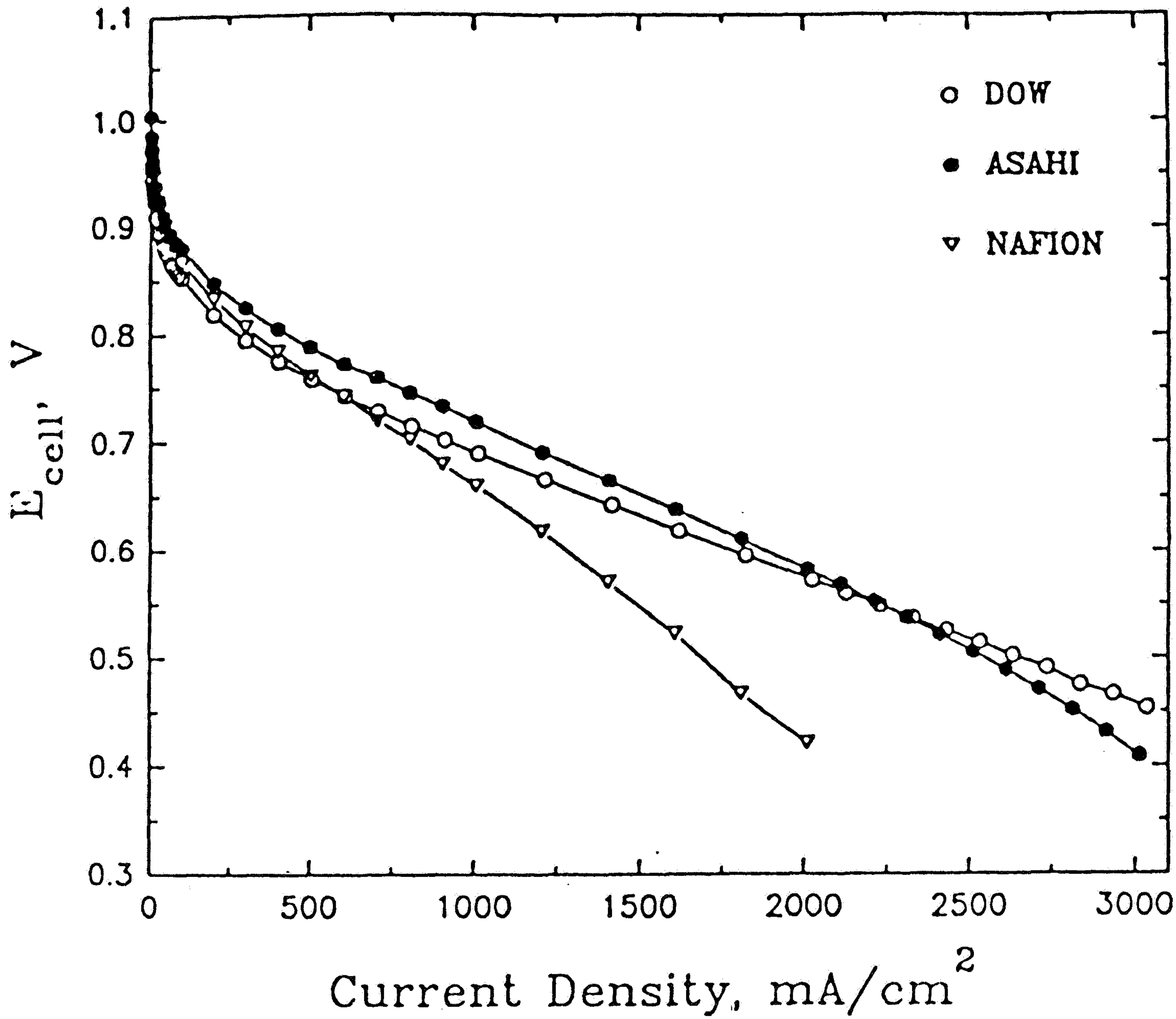


Fig. 5. Comparison of Dow Nafion and Aciplex-S membrane in a PEM fuel cell environment

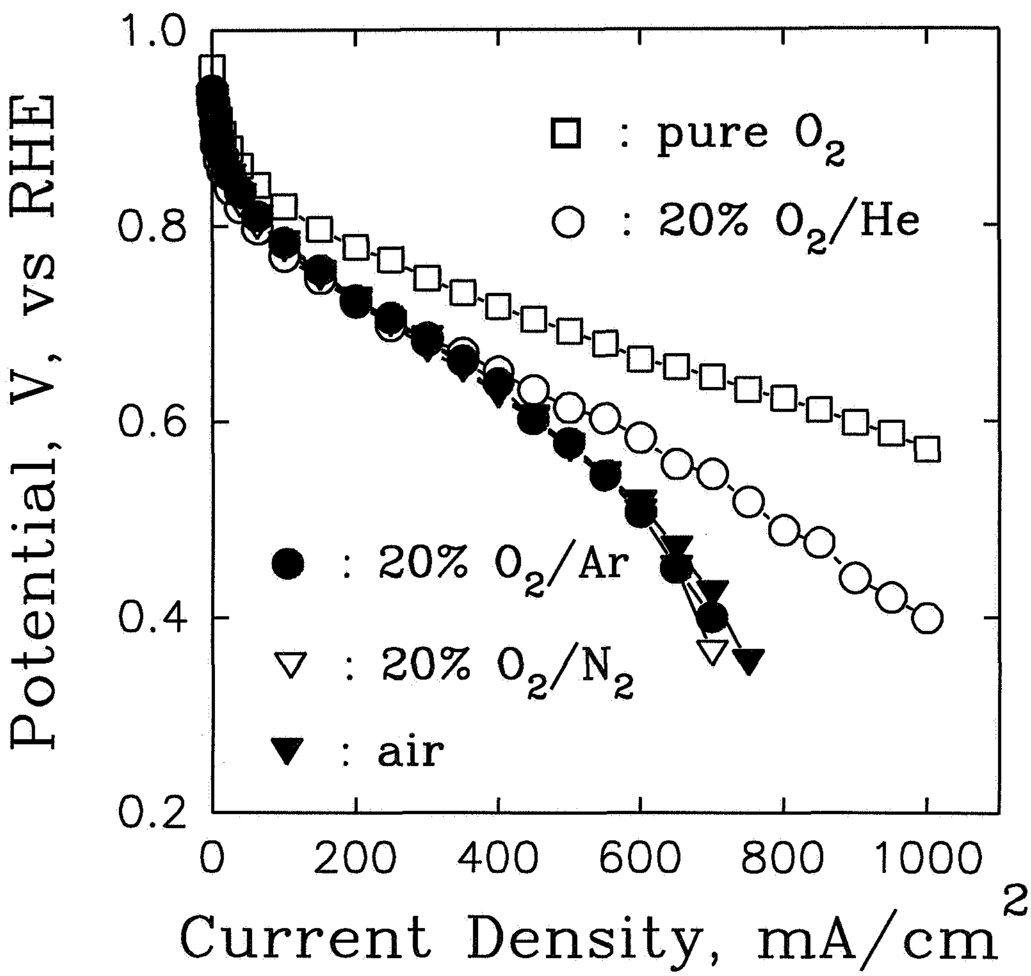


Figure 6 Effect of different gas compositions: pure oxygen, air, 20% oxygen in He, Ar and N_2, on the performance of a proton exchange membrane fuel cell at 70°C and 1 atm pressure.

53-25
44767

OZONE AS A LAUNDRY AGENT ON ORBIT AND ON THE GROUND

William N. Agosto, Lunar Industries, Houston, Texas

ABSTRACT

Ozone (O_3), is the strongest commercial oxidizing agent for aqueous systems and may be ideal for space station laundering operations. It can be generated electronically from air in situ. It kills virtually all microorganisms, attacks many organics and inorganics, and breaks down stable ring structures of benzene and related oils when coupled with ultra violet radiation. It cleans and disinfects in cold water without the need for detergent. It leaves no residues. Ozone permits up to 90% wash water recycling and it eliminates wash time, water volume, and recycling problems of a detergent rinse. Ozone is self purging and converts spontaneously to oxygen. It can be rapidly purged by well established catalytic and thermal processes. Scaling of an ozone laundering system for space station may have commercial applications in a consumer model for home use.

INTRODUCTION

Ozone, (O_3) is the triatomic form of oxygen and the strongest commercial oxidizing agent next to fluorine, which is the most chemically active element. Ozone oxidizes many organic and inorganic compounds and kills a broad spectrum of bacteria, algae, and viruses thousands of times faster than chlorine. It converts to oxygen spontaneously without leaving a residue. Accordingly, it cannot be stored practicably but is easily generated just before use by uv irradiation or corona discharge in air. It has a lifetime of several hours in air and about a half hour in water at room temperature which is ample for chemical action and sterilization. It decomposes rapidly over 100^{o}C and by catalytic action at room temperature.

COMMERCIAL USES OF OZONE

Ozone is 12 to 13 times more soluble in water than oxygen, accordingly its most widespread use is in water purification and waste water treatment (1).

Potable Water Treatment

Ozone destroys virtually all forms of bacteria and viruses in potable waters thousands of times faster than chlorine. It increases settling of suspended colloids, removes tastes, odors, and colors. It oxidizes cyanides, phenols, algae, and sulfides like H_2S to sulfates. Since it destroys most dissolved organics, it prevents formation of carcinogenic trihalomethanes that result when chlorinated waters react with residual organics. It also destroys detergents, pesticides, humic, fulvic, and tannic (organic) acids. And it precipitates the hydroxides of iron, manganese, lead, mercury, nickel, cerium, and silver by raising the metals to their highest oxidation state (1). The Los Angeles municipal water system uses 10,000 pounds of ozone per day and has been operating for 8 years. Many European cities ozonate their water supply.

Wastewater Disinfection

One of the largest applications of ozone is the disinfection of secondary or biologically treated wastewater. Ozone is used when a high quality chemical-free effluent is required for flow into reservoirs or fishing areas. Chlorine kills trout and other biota, and unlike chlorine, ozone does not affect the pH (acidity) of the water body (1).

Industrial Applications

Multiton quantities of ozone are used to destroy cyanides in the paint, plating, photographic and mining industries; as well as phenols in paper mills, coke mills, oil refineries, and thermoplastic resin manufacture. Currently, these are the largest industrial waste applications. The use of ozone obviates the need to transport hazardous chemicals or wastes (1).

Ozone is also used to clean water for HVAC chillers and industrial cooling towers where it is reported to reduce scale build up and inhibit corrosion associated with chemical additives that would otherwise be used to prevent scaling. In addition, ozone is used in aqua culture for fish farms hydroponics and in agriculture to remove pesticides from produce.

High Purity Water and Water Polishing

Breweries and bottled water plants ozonate their water to remove residual taste, odor, and for sterilization. Sterility of pharmaceutical deionized and distilled waters is maintained with ozone. Organic contaminants on the surface of electronic components are removed with ozonated deionized water (1).

Odor Control

Probably the largest number of ozone generators in the U.S. is used for sewage odor control. More than 1000 ozone generators are used in wastewater treatment plants and wastewater pumping stations. Other industrial odors controlled by ozone include those from dairy processing plants, fish processing, rubber compounding, commercial kitchens, rendering plants, food processing, and pharmaceutical fermentation (1).

Commercial Laundries

In a new application, ozone has been introduced as the prime cleaning agent some in large institutional laundry systems. It reduces washing time by 20%, chemicals by 45%, and water consumption up to 90% by recycling the wash water with concomitant cost reduction. Ozone laundering is a cold water process and energy consumption is also substantially reduced according to reports from a Westin Hotel in Rancho Mirage, California where a pilot system is now in operation. Similar systems are in operation at nursing homes and prisons in Florida. Since ozone destroys fecal matter as well as body fluids and odors, little or no detergent is required for these applications. Bleach is compatible with ozone and is used for whitening (2).

Aqueous Ozone Dosages

Ozone concentrations vary anywhere from 0.5 mg/liter for mild potable water disinfection to 15 mg/liter for organics destruction in municipal wastewaters. Accordingly, low concentrations are effective for aqueous purification and disinfection (1).

OZONE GENERATION PROCESSES

There are many ways to generate ozone in the laboratory but only two processes are commercial, 1) ultraviolet irradiation of air or oxygen, and 2) silent corona discharge through air or oxygen with high voltage ceramic or glass electrodes (1).

Ultraviolet (uv) Generators

UV models produce only grams/hour (fractions to a few ounces per day) of ozone and are used primarily for odor control in food processing plants, kitchens, air ducts, and dumpsters. The energy efficiency of uv generators are low at about 20 kwh per pound of ozone generated but they are typically lighter, draw less power, and require less air preparation than the corona discharge systems.

Corona Discharge Generators

Energy consumption for corona discharge generators is typically 7-10 kwh per pound of ozone generated from air. In addition, air preparation draws 2-3.5 kwh/lb. Accordingly, power consumption is the main operating cost.

Typically air feed corona discharge systems require several stages of air filtering down to submicron size, a compressor as well as an air cooler and dryer before ozonation. A contact sparger system mixes ozone with influent water and residual ozone is neutralized in a destruct unit. In addition, cooling water or air is normally required for the electrodes of corona discharge systems to prevent degradation in the efficiency of ozone generation.

OZONE AS A LAUNDERING AGENT ON SPACE STATION

Ozone may be the ideal cleaning agent for laundering operations on space station. It is generated in situ out of ambient air, it requires no additives, cleans and disinfects in cold water, destroys body fluids, wastes, and odors, and, unlike soap, it leaves no residue. Soap removal is a major problem in wash water recycling of closed system laundries. The use of ozone as a cleaning agent would obviate the problem since up to 90% of the ozonated water can be recycled. Wash time and water volume are also reduced because detergent rinsing is unnecessary.

Ordinarily, ozone does not attack stable ring structures like oils and benzene. However, it does react with those compounds in the presence of ultra violet radiation. Accordingly, it may be possible to clean up oils and related soils in a space station laundry by irradiating the washload with solar uv during the ozone wash cycle.

One commercial application of an ozone laundry system on space station may be a scaled down washer suitable for the consumer market. Penetration of the home laundry market could be highly profitable. Reduction in detergent effluents and water recycling are attractive features, especially where water is scarce or expensive. New designs for highly efficient and simplified ozonators that are under development enhance the prospects.

Ozone Hazard

Ozone is not a poison in the sense of entering into body chemistry. For example, it does not combine with the blood the way oxygen does. However, as a strong oxidizing agent, it is highly irritating to the respiratory tract even at low concentrations. At least one research group has reported that ozone preferentially attacks cancer cells over normal tissues (3).

Ozone has the advantage that it is self purging. It destructs rapidly between 350 and 400°C and at approximately 38°C in commercially available catalytic reactors. All moderate to large scale commercial ozonation systems include catalytic destruct units to prevent ozone contamination of the ambient. The end product of ozone destruction is oxygen.

CONCLUSIONS

Ozone laundering systems appear ideal for space station applications for the following reasons:

1. Ozone is a cleaning and disinfecting agent that can be generated in situ.

2. It is a cold water wash process.

3. Water volume is reduced because detergent rinsing is eliminated.

4. Ozone and leaves no residue. Accordingly, up to 90% wash water recycling can be achieved.

5. Wash time is reduced 20%.

6. Residual ozone can be readily converted to oxygen in commercial catalytic and/or thermal units.

Due to scaling considerations, the development of an ozone laundry system for space station may lead to consumer applications.

REFERENCES

1. Kirk-Othmer Encyclopedia of Chemical Technology, 3rd. Ed., Vol. 16, pp. 689-713, 1978-1984.

2. Merli, Richard, Westin Hills Achieves Huge Savings, LAUNDRY NEWS, pp. 4, October 1993.

3. Sweet, Frederick et al., Ozone Selectivity Inhibits Growth of Cancer Cells, SCIENCE, Vol 209, pp. 931-932, 1980.

omit

Session R1: PERCEPTION AND VISION TECHNOLOGIES

Session Chair: Thomas W. Pendleton

54-54
44776

OBJECT RECOGNITION AND POSE ESTIMATION OF PLANAR OBJECTS FROM RANGE DATA

Michael Magee

Computer Science Department
University of Wyoming
Laramie, Wyoming 82071-3682

Chiun Hong Chien
Mark L. Littlefield

Lockheed Engineering and Sciences Company
Mail Code C-19
2400 NASA Road 1
Houston, Texas 77058

Thomas W. Pendleton

NASA Johnson Space Center
Engineering Directorate
Automation and Robotics Division
Robotics Intelligence Section
Mail Code ER-4
Houston, Texas 77058

Abstract

The Extravehicular Activity Helper/Retriever (EVAHR) is a robotic device currently under development at the NASA Johnson Space Center that is designed to fetch objects or to assist in retrieving an astronaut who may have become inadvertently de-tethered. The EVAHR will be required to exhibit a high degree of intelligent autonomous operation and will base much of its reasoning upon information obtained from one or more three-dimensional sensors that it will carry and control. At the highest level of visual cognition and reasoning, the EVAHR will be required to detect objects, recognize them, and estimate their spatial orientation and location. The recognition phase and estimation of spatial pose will depend on the ability of the vision system to reliably extract geometric features of the objects such as whether the surface topologies observed are planar or curved and the spatial relationships between the component surfaces. In order to achieve these tasks, three-dimensional sensing of the operational environment and objects in the environment will therefore be essential.

One of the sensors being considered to provide image data for object recognition and pose estimation is a phase-shift laser scanner. The characteristics of the data provided by this scanner have been studied and algorithms have been developed for segmenting range images into planar surfaces, extracting basic features such as surface area, and recognizing the object based on the characteristics of extracted features. Also, an approach has been developed for estimating the spatial orientation and location of the recognized object based on orientations of extracted planes and their intersection points. This paper presents some of the algorithms that have been developed for the purpose of recognizing and estimating the pose of objects as viewed by the laser scanner, and characterizes the desirability and utility of these algorithms within the context of the scanner itself, considering data quality and noise.

1. Introduction

There has been considerable recent research devoted to the development of intelligent free-flying robots that can assist in space operations.[1,2] One such robotic device, the Extra Vehicular Activity Helper/Retriever (EVAHR), is intended to operate in relatively close proximity to a human operator, assisting with tasks such as fetching a tool, retrieving objects that may have drifted away from the primary work area, or even retrieving an astronaut who may have inadvertently become de-tethered. Early results from tests using a Manned Maneuvering Unit (MMU) propelled EVAHR on a Precision Air Bearing Floor (PABF) to simulate the frictionless environment of space demonstrated that it was possible to retrieve both large and small objects using computer vision to sense the operational environment and to employ a speech recognition system for understanding human voice commands to direct the robot's actions.[3,4] Studies are currently underway to assess the operational characteristics of the sensors and robot control mechanisms in microgravity with experiments on NASA's KC-135 aircraft.

The ability of the EVAHR to sense its operational environment is central to its functionality as an autonomous or semi-autonomous device since it must be able to recognize objects, track them, estimate their spatial poses, and estimate their motion parameters over time.[5,6,7] Because of the heterogeneous nature of these tasks, it is ultimately likely that several sensors with complementary capabilities will be employed to achieve different goals depending upon the current state of the world (the world model), the task to be achieved, and the characteristics of the sensors themselves.[8] For example, images from a color camera are useful for identifying objects based on their visible spectral characteristics but it difficult to estimate pose from two-dimensional images. Conversely, a laser scanner can provide three-dimensional coordinates for points on a scanned object, but no color information is available. The remainder of this paper focuses on processing actual image data from a laser scanner, and documents a method for segmenting objects into their primary planar regions, recognizing them, and estimating their spatial poses.

2. Laser Scanner Characteristics

The sensor employed for the studies whose descriptions follow is a laser range scanner that measures distances based on the phase shift of a modulated signal carried on an infrared laser beam. The range values returned by the scanner are represented by 12 bit integers that span a single ambiguity interval of approximately 15.2 meters. This means that a difference of one range unit (out of 4096) represents a distance change of about 4 mm. The scanner is able to produce a dense range image by employing a rotating mirror whose rotation axis can be tilted. The scanner simultaneously provides two separate range and reflectance (intensity) images that are fully registered.

The quality of the range data provided by the scanner is affected by several factors which generally relate to the composition of the surface material, its reflectivity characteristics, its geometry, and the orientation of surface normals relative to the scanner itself. The most influential among these factors is the reflectivity of the surface material. For extreme cases in which a scanned region is composed of a highly specularly reflective material, reliable range estimates are not expected since the laser beam will be reflected away from the sensor.

For less extreme cases involving diffuse reflective surfaces, however, the quality of the data is highly dependent on the albedo of the surface. These dependencies can best be illustrated by examining the quality of the range images acquired by scanning black and white planar surfaces (sheets of paper) that were oriented perpendicular to the optical axis of the scanner. As a measure of data stability, the local standard deviation (sigma) for range values was computed within a row. This local standard deviation was based on the center range value and the nearest 8 neighbors within the row. It was observed that the local sigma varied by as much as 3 range units. For such cases, in excess of 99% of the range samples could be expected to fall within 3 sigma (± 9 range units) of the mean value. For the test case under discussion, this translates into a local variation of approximately ±33 mm over a distance of 8 mm. For the black surface, the quality of the data was significantly worse. Local standard deviations as high as 9 range values were observed meaning that a 3 sigma test would include range values as far as ± 100 mm over this limited region of a scan line. The local standard deviations for reflectances varied up to 30 units for the white surface and up to 8 units for the black surface.

The implications of these observed local variations are very important when designing algorithms that attempt to segment the image into component regions such as planes and curved surfaces. For example, the magnitude of the local variations in range values makes it extremely difficult to segment planar surfaces based on a local geometric constraint such as surface normal consistency. Furthermore, even on white objects, it is difficult to recognize the curvature of objects smaller than 100 mm since the magnitude of local range variation is large relative to surface size. If the data is smoothed by a classical filtering mechanism, finer details that are necessary to recognize an object and/or estimate its pose may be lost. Hence, algorithms that depend on local geometry are less likely to succeed than those that take a more global approach to object analysis. The results of both local and global algorithms that were developed are presented in the next section.

3. Finding Planes, Recognizing Objects and Estimating their Spatial Poses

The local instability of range values observed for the laser scanner makes scene segmentation using locally computed surface normals exceptionally difficult unless the range values are smoothed using a reasonably large filter. Applying such a filter, of course, results in a loss of scene detail but does make it possible to find planes that are large relative to the size of the filter.

An approach that was found to be both more computationally efficient and robust was to grow surfaces based on local range and reflectance difference constraints. It was determined that after applying a 7X7 mean filter, planes that were not highly oblique to the sensor axis could be successfully grown by adding to regions neighboring image elements whose smoothed reflectance and range values did not differ by more than 40 and 1.5, respectively. This provided the basis by which planar regions could be segmented and the segmented planes used for object recognition and pose estimation. Figure 1 shows one object, a simulated Orbital Replacement Unit (ORU), to which the plane segmentation algorithm was applied. This ORU consists of a rectangular solid to which an H-shaped handle is attached by an intermediate short cylindrical section. When viewed by the laser scanner and rendered as a solid model, the ORU appears as in Figure 2. With respect to the observed noise characteristics that had to be dealt with algorithmically, Figures 3 and 4 are more revealing, however.

Figure 3 shows a wireframe rendering of the scanned ORU with the bright line profile across the main body and H-shaped handle being isolated in Figure 4. It should be noted that the raw data across the major left surface should be linear, but is extremely "busy". It is this effect that makes the segmentation of planes using surface normals difficult since inconsistent directions based on local patches are computed unless large smoothing filters are applied. On the other hand, using a region growing approach based on propagating the local constraints of reflectance and range similarities, it is possible to successfully segment the scene into its planar regions as shown in Figure 5. It should be noted, however, that these successfully segmented planar regions are somewhat deceptive since when viewed from the perspective of the laser scanner the true variation of the original data is not evident. Figure 6 shows the same segmentated image data but from a different viewpoint. It should be noted that there are several areas of high variation. In particular, the H-shaped handle has range values that vary by as much as the width of the handle's vertical substructures. Hence, the level of noise is relatively large compared to the feature itself.

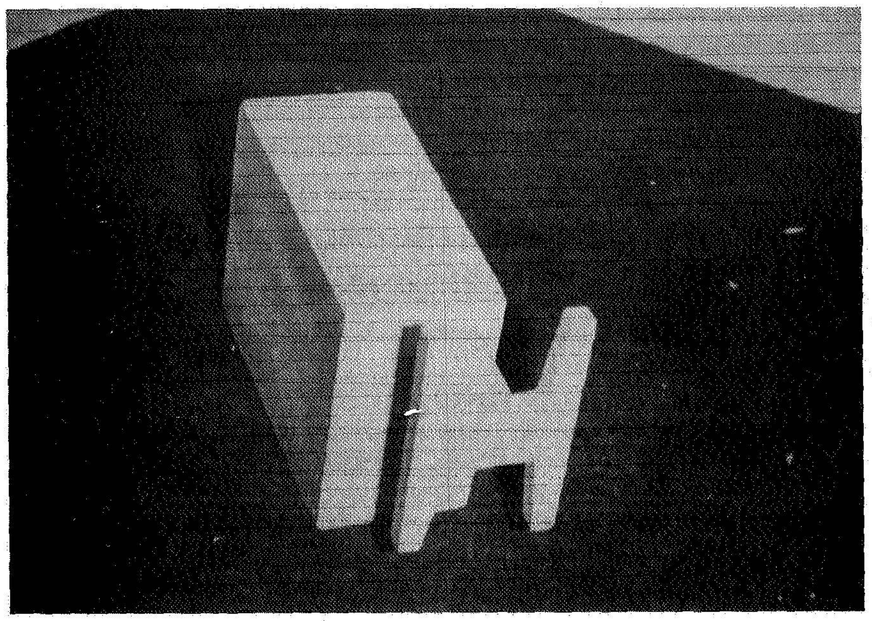


Figure 1: simulated orbital replacement unit (ORU)

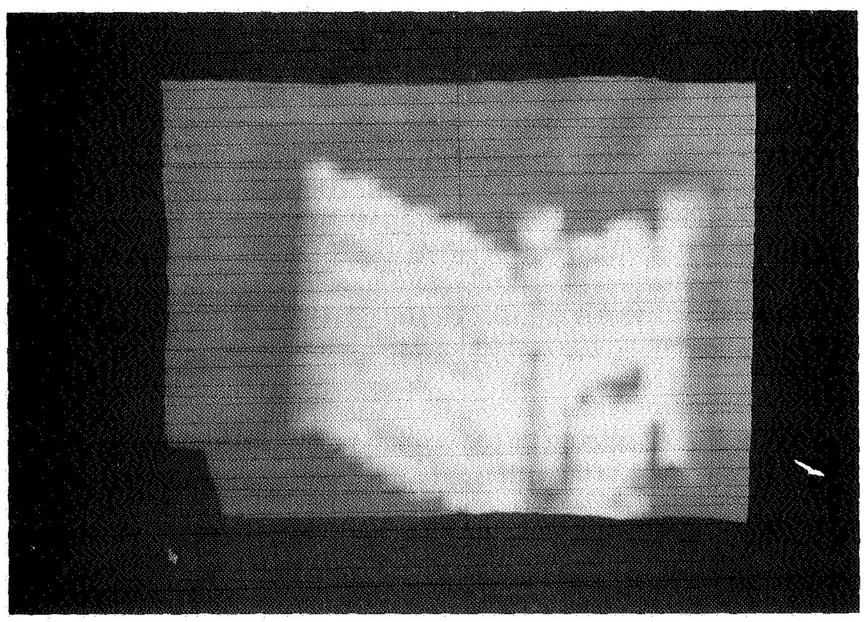


Figure 2: ORU as a shaded model graphically reconstructed from range data

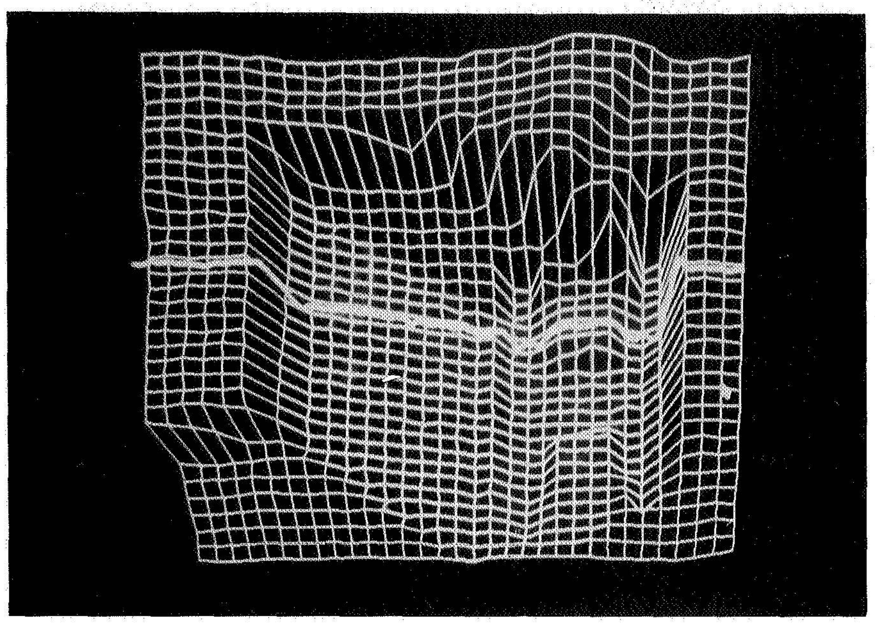


Figure 3: ORU as a wireframe model graphically reconstructed from range data

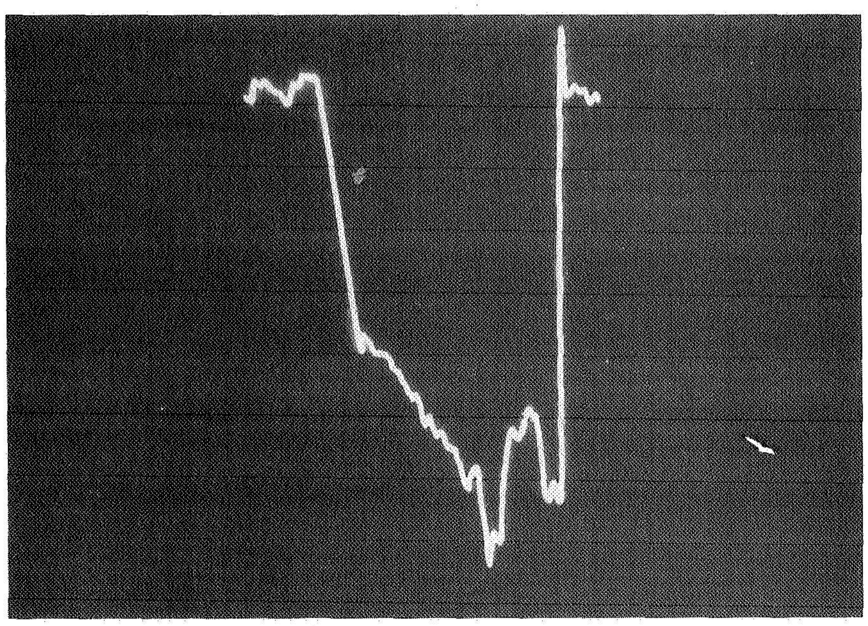


Figure 4: a single line (profile) of laser range data across the ORU

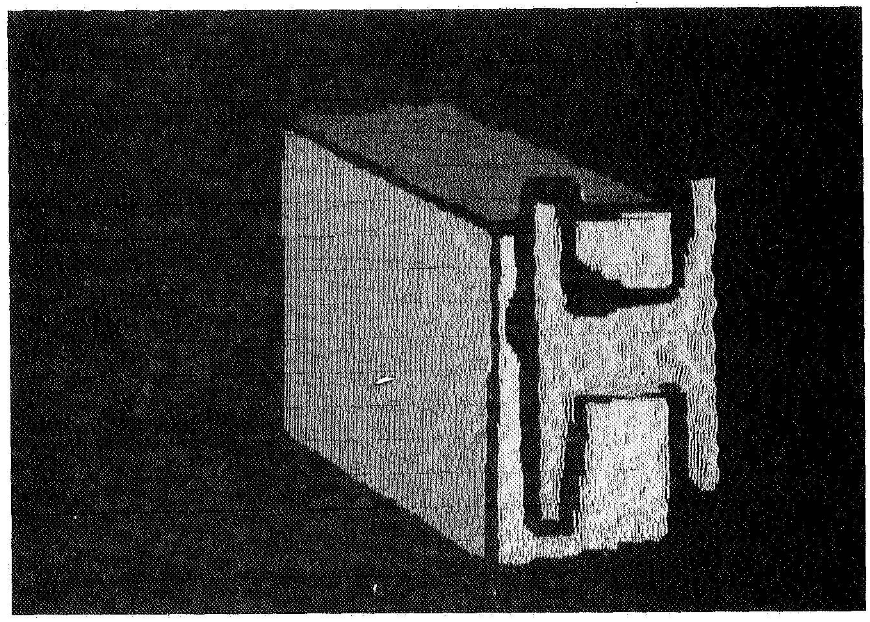


Figure 5: laser range data segmented into planar regions

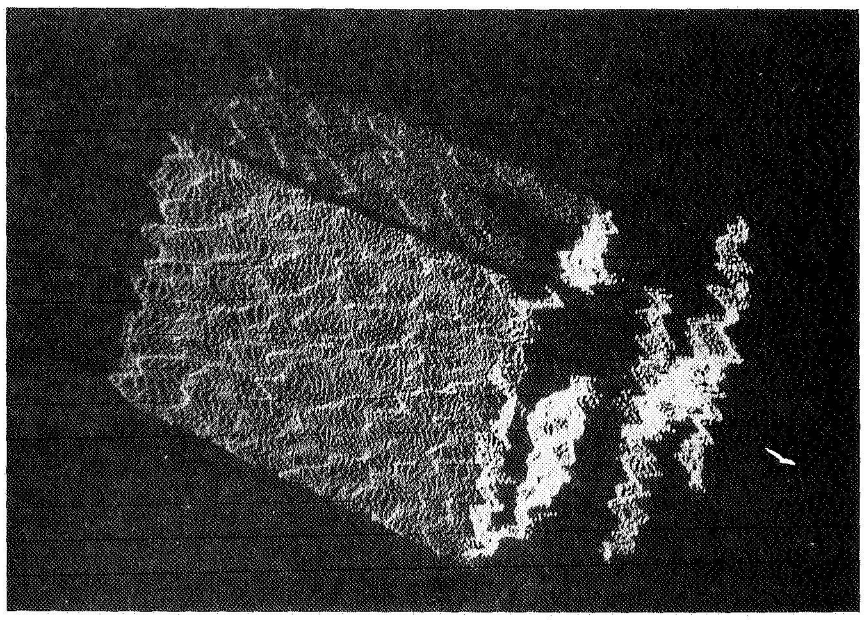


Figure 6: variations in laser range data from planar regions

The method by which the primary planar features of the ORU were recognized was based on the areas of the observed planes. Since the original sensor data as shown in Figure 6 has extreme variations in the form of hills and valleys, however, incorrect areas for the planar features would be computed unless the data were forced to conform to the best plane equation that fits all of the range points belonging to a segmented feature. This was achieved by computing the plane equation using a least squares fit of all the points in each segmented planar feature and backprojecting each point in the segmented planar feature onto the computed plane. Figure 7 shows the points in the adjusted three-dimensional range image that results when this process is applied to the data in Figure 6.

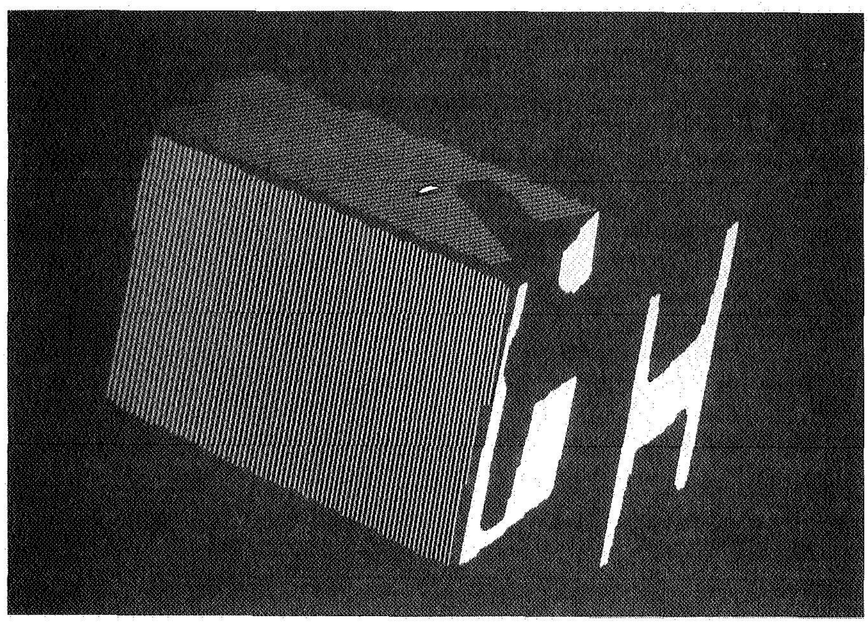


Figure 7: laser range data after conformal mapping to extracted planes

After the planar conformal mapping of the original data has been achieved, the area of each feature is computed and compared against the areas of planar features in the model base, and correspondences between observed and model features are established. Since, this feature matching method is based on computed surface areas, it is necessarily sensitive to occlusion. However, once surfaces have been grown, it is possible to compute other features that would be useful for recognition such as the vertices and line segments that result from the intersections of planes. Four or more features are sufficient to provide the basis for feature matching and pose estimation.

For the current study, pose is estimated by orienting the model such that three of its surface normals match the orientations of the corresponding planes in the observed data and such that the intersection point of these three planes is translated to be consistent with the analogous observed intersection point. The wireframe overlay in Figure 8 demonstrates that the proper spatial pose for the ORU model is computed such that its features correspond to those in the original range image data.

4. Conclusions

A method has been presented for segmenting planar regions from laser range and reflectance data which is useful for recognizing objects and estimating their spatial poses. The method, which is based on local constraint propagation, permits successful planar segmentation even in the presence of significant noise, but postprocessing of the three-dimensional data in the segmented regions is required to accurately characterize and use the planar regions.

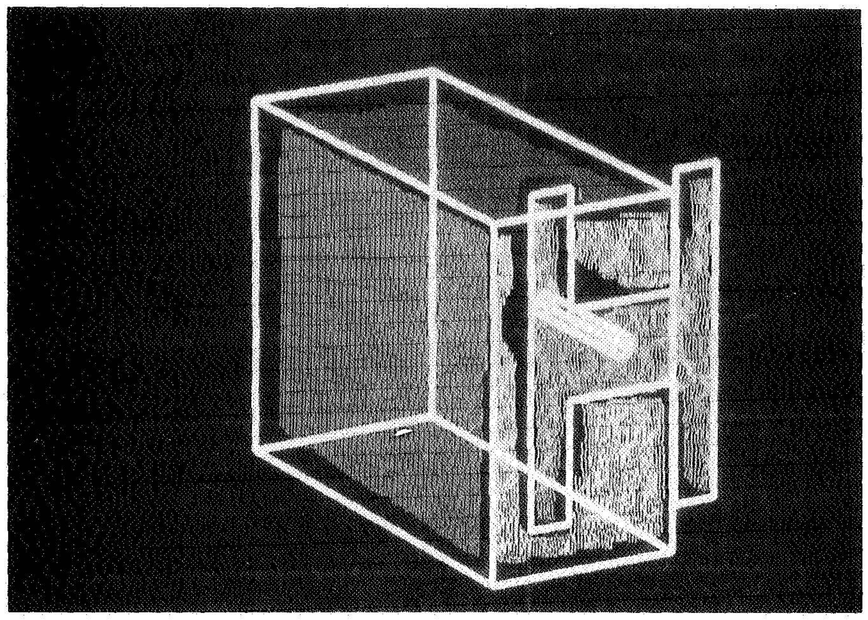


Figure 8: overlay showing correctly estimated pose for ORU model

5. Acknowledgements

The support of the NASA Johnson Space Center Summer Faculty Fellowship Program and the University of Wyoming Planetary and Space Science Center is gratefully acknowledged.

6. References

1. G.J. Reuter, C.W. Hess, D.E. Rhoades, L.W. McFadin, K.J. Healey, and J.D. Erickson, "An Intelligent Free-Flying Robot", *SPIE Symposium on Advances in Intelligent Robotic Systems, Space Station Automation IV*, SPIE 1006-03, Cambridge, Massachusetts, Nov. 6-11, 1988.
2. J.D. Erickson, C.R. Price, D. Cooke, "Future Needs for Space Robots for SEI", *Proceedings of the SPIE Symposium on Cooperative Intelligent Robotics Systems in Space II*, Boston, Massachusetts, November 10-14, 1991.
3. J. Erickson, R. Goode, K. Grimm, C. Hess, R. Norsworthy, G. Anderson, L. Merkel, and D. Phinney, "Technology Test Results from an Intelligent Free-Flying Robot for Crew and Equipment Retrieval in Space", *SPIE Symposium on Cooperative Intelligent Robotic Systems in Space II*, Boston, MA, November 10-14, 1991.
4. K.A. Grimm and J.D. Erickson, "An Experiment in Vision Based Autonomous Grasping within a Reduced Gravity Environment", *Spie Cooperative Intelligent Robotics in Space III*, Cambridge, Massachussetts, November, 1992.
5. C.H. Chien, "Multi-view Based Pose Estimation from Range Images", *Spie Cooperative Intelligent Robotics in Space III*, Cambridge, Massachussetts, November, 1992.
6. M.L. Littlefield, "Adaptive Tracking of Objects for a Mobile Robot Using Range Images", *Spie Cooperative Intelligent Robotics in Space III*, Cambridge, Massachussetts, November, 1992.
7. L. Hewgill, "Motion Estimation of a Freely Rotating Body in Earth Orbit", *Spie Cooperative Intelligent Robotics in Space III*, Cambridge, Massachussetts, November, 1992.
8. M. Magee, C.H. Chien, and T.W. Pendleton, "A Vision System Planner for the Extravehicular Activity Retriever", *Proceedings of the International Conference on Intelligent Autonomous Systems*, Pittsburgh, Pennsylvania, February, 1993, pp. 696-705.

DUAL-USE SPACE TECHNOLOGY TRANSFER CONFERENCE
AND EXHIBITION

MICROWAVE IMAGING OF METAL OBJECTS

S5-32
44772

Jiang Li
C. Tolliver
Hsiang H. Yeh
College of Engineering and Architecture
Prairie View A&M University

Kumar Krishen
NASA
Johnson Space Center

ABSTRACT

The procedure of microwave imaging by maximum entropy method is discussed. First, the relationship between the induced current on the metal object surface and the scattered field is introduced. Our imaging concept is to reconstruct the induced current on the object surface from the measured scattered field. The object configuration will be provided by the induced current which is zero everywhere except on the object surface. Future work is also included with focus on the application of microwave imaging to both NASA and industry.

I INTRODUCTION

To achieve the highest resolution in microwave imaging has been a goal for a long time for scientist and engineers attributed largely to its academic significant and the understandable commercial and military values. The quality of microwave imaging is judged by how faithfully the microwave image represents the spatial distribution of the object of interest.

In the most direct form, imaging a three-dimensional object can be accomplished by using a range-gated, short-pulse radar with a pencil-beam antenna. The antenna beam and the range gate are systematically scanned throughout the three-dimensional volume, and the intensity of the signals received is displayed as a function of the spatial coordinates being interrogated. Because the spatial resolution is established by the angular and range resolution of the sensor, the image is obtained directly, without subsequent processing.[1]

The relative simplicity of forming the image, which requires minimal data processing, and the fact that the image can be obtained while the object is stationary constitute the advantages of the direct imaging method. The principal disadvantages of the direct method are:

• A high degree of spatial resolution requires subnanosecond pulses and large apertures, posing practical limitations.

• The cross-range resolution, obtained from the antenna beamwidth, degrades as the range increases.

• The spatially limited irradiation of the object omits interactions and coupling that may occur when the entire object is simultaneously irradiated.

Synthetic imaging means the imaging is obtained by synthetic means when results of many observations of the object at different frequencies and angles are coherently combined, and the short-pulse, and large aperture antenna are still very important to produce a faithful imaging of the objects.

In this paper, we will investigate the imaging of a metal object by single frequency microwaves. First we will introduce our imaging concept. The relationship of the induced current on a metal object surface and scattered field is introduced, and then maximum entropy method will be used to reconstruct the induced current on the object surface, thus reveal the object configuration.

II. THE MAXIMUM ENTROPY METHOD FOR MICROWAVE IMAGING

From electromagnetic theory we know that once the radiated field is known, the induced current (tangential electric field) on the surface of a metal object can be found by the boundary conditions. The induced current is therefore known at every point. According to the Huygens' principle, each point can then be treated as a radiating element, and the total scattered field can be obtained by integrating over the entire object surface.

For microwave imaging, this means that if we are able to determine the induced current which is

$$\mathbf{J} = 0 \quad \text{inside the object}$$
$$\mathbf{J} \neq 0 \quad \text{outside the object}$$

we will be able to reconstruct the image of the object. The general relationship between the induced current on the metal object surface and the scattered field can be expressed as surface integration of the Green's function. For example, if the induced current on the object surface is in the x-direction, we will have[2]

$$E_x(\mathbf{r},\mathbf{r}') = \iint_{s'} U(\mathbf{r},\mathbf{r}')J_x(\mathbf{r}')ds'$$

where

$$U(\mathbf{r},\mathbf{r}') = \frac{\exp(-ik|\mathbf{r}-\mathbf{r}'|)}{4\pi|\mathbf{r}-\mathbf{r}'|} - \frac{1}{k^2}\frac{\partial^2}{\partial r \partial r'}\left(\frac{\exp(-ik|\mathbf{r}-\mathbf{r}'|)}{4\pi|\mathbf{r}-\mathbf{r}'|}\right)$$

the numerical model of the above imaging problem with the random error added is

$$\mathbf{E} = \mathbf{U}\mathbf{J} + \varepsilon$$

where $\mathbf{J} = [J_1, J_2, \ldots\ldots, J_N]$ is the vector which contains the data of the image to be reconstructed, N is the number of current elements on the metal object surface, $\mathbf{E} = [E_1, E_2, \ldots\ldots, E_M]$ is the vector which contains the measured data for the scattered field from which the induced current is to be reconstructed, and $\varepsilon = [\varepsilon_1, \varepsilon_2, \ldots\ldots, \varepsilon_M]$ is the vector representing the error, M is the number of measurement points and $\mathbf{U}$ is a matrix of size $M \times N$.

The concept of maximum entropy is related to that of probability density and it determines an image $\mathbf{J} = [J_1, J_2, \ldots\ldots, J_N]$ which maximizes the function

$$S(\mathbf{J}) = \sum_{k=1,}^{N} \ln J_k - \lambda \sum_{i=1}^{M} \frac{1}{\sigma_i^2}\left(\sum_{k=1,}^{N} U_{ik}J_k - E_i\right)$$

where the first term is the entropy of the image, and the second term is a quadratic term which represents noise. This maximum problem can be solved iteratively.

The maximum entropy criterion is as follows. First, the difference between the computed fields and the measured fields is to be minimized at the same time the entropy of the scattering current is also to be minimized. The entropy is an indication of the fluctuation of the scattering current from cell to cell. The global maximum entropy is the case where the scattering current is a constant, but constant scattering current may produce a large discrepancy between the computed fields and the measured fields. By minimizing the difference, together with maximizing the entropy, a compromise is reached where the variation of scattering current from cell to cell is smooth but at the same time the resulting computed field is different from the measured field by a tolerable amount.

III. THE SCATTERED FIELD BY A FLAT METAL OBJECT

Consider a planar radiating aperture A, defined as the area on the aperture plane $z=0$ over which the tangential field is non-zero. The electric field in the aperture plane is entirely x-directed

$$E_{ax}(x,y)=E_x(x,y,0)$$

the radiated field exits in the half-space $z\geq 0$. In the far field region, the electric field at point $P(x,y,z)$ or $P(r,\theta,\phi)$ with direction cosines (α,β,γ) can be expressed as[3]

$$\mathbf{E}(r,\theta,\phi)=j\frac{2\pi}{kr}\exp(-jkr)F_x(\alpha,\beta)[\mathbf{u}_x\gamma-\mathbf{u}_z\alpha]$$

and the corresponding magnetic field is

$$\mathbf{H}(r,\theta,\phi)=j\frac{2\pi}{Zkr}\exp(-jkr)F_x(\alpha,\beta)\left[-\mathbf{u}_x\alpha\beta+\mathbf{u}_y\left(1-\beta^2\right)-\mathbf{u}_z\beta\gamma\right]$$

where Z is the plane-wave impedance of the medium and

$$F_x(\alpha,\beta)=\frac{1}{\lambda^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}E_{ax}(x,y)\exp\left[jk(\alpha x+\beta y)\right]dxdy$$

is the angular spectrum which is Fourier transformation of the radiation aperture field.

Suppose that we have a flat metal object which is parallel to the incident aperture of which the radiating field is expressed by their angular spectra and the distance between the object and the incident aperture is z.

Over the surface of the conducting plate, the total tangential electric field must be zero to satisfy the boundary condition. When dimension of the object is large in comparison to the wavelength, the local diffraction effects are negligible. By the tangent-plane approximation of physical optics, the aperture field assumes that the field on reflection is the same over the conducting plate as if it were part of an infinite plane. This is the simplest the aperture field corresponding to the scattered field can be taken to be the negative of the tangential component, therefore the induced current, of the incident field over the conducting plate, and zero elsewhere in the x-y aperture plane. From the above discussion we can see that when the incident aperture field is entirely x-directed, the induced current on the object surface will be also x-directed. Therefore the same angular spectrum method can be used to find the scattered field.

One example of the plane radiating aperture is horn antenna. The aperture field distribution of a horn antenna can be expressed as

$$E_{ax}(x,y)=E_o\cos\left(\frac{\pi y}{b}\right)\exp\left[-\frac{jk}{2}\left(\frac{x^2}{l_E}+\frac{y^2}{l_H}\right)\right]$$

The angular spectrum describing the radiating field is

$$F_x(\alpha,\beta) = \frac{E_0}{2\lambda^2} X_f \left(Y_f^+ + Y_f^- \right)$$

where

$$X_f = \sqrt{\frac{\lambda l_E}{2}} \exp\left(\frac{j\alpha^2 l_E \pi}{\lambda}\right) \left[F\left(\sqrt{\frac{2}{\lambda l_E}} \frac{a}{2} - \sqrt{\frac{\lambda l_E}{2}} \frac{2\alpha}{\lambda} \right) + F\left(\sqrt{\frac{2}{\lambda l_E}} \frac{a}{2} + \sqrt{\frac{\lambda l_E}{2}} \frac{2\alpha}{\lambda} \right) \right]$$

$$Y_f^+ = \sqrt{\frac{\lambda l_H}{2}} \exp\left(\frac{j\beta_+^2 l_H \pi}{\lambda}\right) \left[F\left(\sqrt{\frac{2}{\lambda l_H}} \frac{b}{2} - \sqrt{\frac{\lambda l_H}{2}} \frac{2\beta_+}{\lambda} \right) + F\left(\sqrt{\frac{2}{\lambda l_H}} \frac{b}{2} + \sqrt{\frac{\lambda l_E}{2}} \frac{2\beta_+}{\lambda} \right) \right]$$

$$Y_f^- = \sqrt{\frac{\lambda l_H}{2}} \exp\left(\frac{j\beta_-^2 l_H \pi}{\lambda}\right) \left[F\left(\sqrt{\frac{2}{\lambda l_H}} \frac{b}{2} - \sqrt{\frac{\lambda l_H}{2}} \frac{2\beta_-}{\lambda} \right) + F\left(\sqrt{\frac{2}{\lambda l_H}} \frac{b}{2} + \sqrt{\frac{\lambda l_E}{2}} \frac{2\beta_-}{\lambda} \right) \right]$$

and

$$\beta_+ = \beta + \frac{\pi}{bk}$$

$$\beta_- = \beta - \frac{\pi}{bk}$$

and F(x) is the Fresnel Integral

$$F(x) = \int_0^x \exp\left(-j \frac{\pi u^2}{2} \right) du$$

The advantage to use the angular spectrum method to find the radiating field is that we can assume the aperture field incidence instead of assuming the plane wave incident. From the above example we can see that the incident wave can be expressed analytically.

In the process of imaging reconstruction, to find the scattered field from the known source is called the forward modeling and to find the imaging from the measured (or simulated) scattered field is call the inversion. Because the inversion process by maximum entropy method is to be solved iteratively, the simplicity of the forward model is a must for a successful imaging method.

IV. FUTURE WORK

As stated in the previous section, the imaging of a flat metal object with known planar incident aperture field consists of the following steps

• The radiated field distribution in the z≥o space is calculated by the radiating aperture field.

• The radiating aperture for the scattered field is obtained, which is the negative tangential part of the radiated field obtained in the previous step; the diffraction effect is neglected.

• The scattered field is calculated.

• The measured scattering field is simulated by adding a random error to the scattering field obtained from the calculation.

• Maximum entropy is used to reconstruct the tangential component on the object surface, thus the shape of the object.

Our future work on the microwave imaging will be set up the experiment system and display the reconstructed metal object image on the computer screen. In the mean time, we will study the effects of a) frequency of the incident wave, b) viewing angle and area, c) size, shape, material and orientation of the object on the quality of the image of the object. The object image obtained from the magnitude and phase of the scattered field and the object image obtained from scattered field magnitude only will also be studied.

We will also work on the image of the metal object covered by dielectric materials. The near field image will be emphasized because of the practical importance and the potential high quality of the image because of the high signal/noise ratio.

ACKNOWLEDGMENT

Founding of this project is provide by NASA/JSC.

REFERENCES:

[1]Mensa, D.L., ***High Resolution Radar Cross-Section Imaging***, Artech House, 1991

[2 Baribaud, M., "Microwave imagery: analytical method and maximum entropy method", J. Phys. D: Appl. Phys., vol.23, no 3, p.269-289

[3] Clark, R.H., and Brown, J., ***Diffraction Theory and Antennas,*** Ellis Horwood Ltd., Chichester, England,1980.

56-63
44775

Robotic Vision Techniques for Space Operations

Kumar Krishen, Ph.D.
NASA Johnson Space Center
Houston, Texas 77058

ABSTRACT

Automation and robotics for space applications are being pursued for increased productivity, enhanced reliability, increased flexibility, higher safety, and for the automation of time-consuming tasks and those activities which are beyond the capacity of the crew. One of the key functional elements of an automated robotic system is sensing and perception. As the robotics era dawns in space, vision systems will be required to provide the key sensory data needed for multifaceted intelligent operations. In general, the three-dimensional scene/object description, along with location, orientation, and motion parameters will be needed. In space, the absence of diffused lighting due to a lack of atmosphere gives rise to: (a) high dynamic range (10^8) of scattered sunlight intensities, resulting in very high contrast between shadowed and specular portions of the scene; (b) intense specular reflections causing target/scene bloom; and (c) loss of portions of the image due to shadowing and presence of stars, Earth, Moon, and other space objects in the scene. In this work, developments for combating the adverse effects described earlier and for enhancing scene definition are discussed. Both active and passive sensors are used. The algorithm for selecting appropriate wavelength, polarization, look angle of vision sensors is based on environmental factors as well as the properties of the target/scene which are to be perceived. The environment is characterized on the basis of sunlight and other illumination incident on the target/scene and the temperature profiles estimated on the basis of the incident illumination. The unknown geometrical and physical parameters are then derived from the fusion of the active and passive microwave, infrared, laser, and optical data.

I. Vision for Vision and Remote Sensing

NASA has identified five strategic enterprises: Mission to Planet Earth (MTPE), Aeronautics, Human Exploration and Development of Space (HEDS), scientific research, and space technology. In each of these areas there is a need for remote sensing and vision. For the MTPE Program, one major task is the observation of the Earth and its atmosphere in order to provide estimation of resources, and sense, monitor, and model the environment. The HEDS program deals with space infrastructure including robotic missions and human expeditions and settlements. The need for autonomous systems for servicing, maintenance, repairs, docking, assembly, planning, monitoring, diagnosis, control, and fault recovery is of paramount importance for this enterprise. The NASA scientific missions are aimed at acquiring knowledge about the universe, matter, and the process for the evolution of life. Much of the research in this area involves remote observations of celestial bodies, phenomena and processes. These observations are carried out in various parts of the electromagnetic spectrum. In all the NASA strategic enterprises, the need for automation and robotics has been established for orbiter/station/satellite servicing, astronaut assistance, equipment transfer, docking and berthing, inspection, remote monitoring, rocket staging, telescience, and assembly of structures and systems. In support of all the tasks involving vision and sensing, technologies will be required to assure the perception of objects, scenes, and phenomena in space. In general, key sensing data will be needed for the three-dimensional scene/object description, including location, orientation, motion parameters, and surface/subsurface properties. The need and status for the vision technology applied to robotics and automation have been research by NASA over the past decade and have been summarized by Krishen [1,2]. For the remote sensing applications, the processed sensor data should provide surface properties such as roughness, dielectric constant, emissivity, reflectivity, temperature, orientation, and slope.

The ultimate goal of vision and sensing systems is to be able to acquire required scene/object parameters at any time, location, or illumination condition. This assured vision has been the subject of fusion by Collin, Krishen, and Pampagnin[3]. The possibility of this happening rests on understanding the reflectance, emission/absorption, and scattering properties of objects/scenes based on physical interaction of the illuminations and the material composition of the object/scene. Once this is approximately realized thorough mathematical and empirical methods, it is then applied to identifying sensing strategies. For example, the environment is monitored in terms of the radiation received form the sun or any other source. Based on this known radiation environment, the scattered and emitted radiation in various parts of the electromagnetic (EM) spectrum can be estimated. Then a selection of the sensor parameters can be accomplished and vision data acquired. The data can be used to provide the needed parameters using physical and phenomenological models. One key element of this approach is to use active and passive sensors located in various parts of the EM spectrum. The combination of needed sensors and modes for these sensors will depend on the object/scene parameters and the accuracy to which they need to be determined. In this paper, a scheme for new multisensor approach to sensing and vision will be described.

II. Object/Scene Parameter Estimation

The natural space environment consists of intense light and dark period. At a nominal altitude of 270 nmi, the sunlight intensity will fluctuate between about 60 minutes of extreme brightness (13,000 ft-c) and 30 minutes of darkness. Furthermore, due to the absence of atmosphere, light is not diffused/scattered. Consequently, the unenhanced images have large contrast with intensity changes of the order of 10. The intense specular reflections combined with camera performance can cause bloom and Faunhofer/Airy rings resulting in scene obscurity. Further complexity results from other objects, such as stars, moon, sun, Earth, and other satellites in the field of view (FOV). Object reflectivity can also pose problems for the vision systems. Most space systems are painted white or finished with smooth, specular materials to provide highly reflective surfaces. The ubiquity of white surfaces intensifies the problem of relying on photometric data for object identification/discrimination. A secondary source of concern affecting vision is the absence of gravity. For free-flying and tethered objects there would be an increased number of positions and orientations in which the objects may be found due to the lack of disturbances caused by aerodynamic and gravitational forces.

There is also a wide range of temperatures from a few degrees Kelvin to hundreds of degrees exhibited by objects and scenes. The surface reflectance can range from 0.07 for lunar soil and thermal protection to 1 for polished metallic surfaces.

A crucial step in the estimation of object/scene parameters is the monitoring and intelligent use of environmental information in space. The spacecraft and other space objects to be recognized by a robot have a definite geometry and consist of materials manufactured here on the Earth. Furthermore, their orbits and general orientations can also be initially available or estimated. For the lunar or Mars surface, there is similar data available in terms of surface orientation as a function of sun and Earth location or time. With this type of information available from objects and scenes in space, one can develop the surface illumination intensity pattern using computerized model of the sun radiation intensity. This can also be extended to provide initial temperature and emissivity estimates using Planck's law and the Rayleigh-Jeans approximations. Figure 1 illustrates the initially estimated data from orbital and illumination considerations.

The second step in the approach is to calculate scattering and emission patterns based on known geometry and orbital parameters. For the lunar surface these calculations can be done on the basis of surface height and slope distributions. Once the scattering and emissions are estimated, the selection of active and passive sensors can be effectively done. The sensors include microwave, millimeter wave, laser, infrared, and optical types.

The operating frequencies, polarization and look angles for these sensors can also be selected for the object/scene viewing. This selection provides the initial estimation of the required parameters. The laser scanner can be utilized as an initial estimate for the velocity and orientation of the object. This can be followed by another sensor such as an active microwave radar to estimate the roughness of the scene/object. Once some roughness estimate is made, a radiometer can be used to estimate the dielectric constant. The dielectric constant, along with the initial value of the incident radiation can then be utilized to select appropriate infrared sensor to map the scene/object. In the last step, if the optical image is available, the data can be fused to provide the synergism needed to refine the estimates of roughness, dielectric properties, and temperatures. The overall flow is briefly shown in Figure 2.

The fusion of data including sensor information takes into account physical scattering and emissivity models and estimates more accurately the object/scene parameters and/or provides new parameters that could not be estimated by the sensors on an individual basis. For example, the fusion of active and passive microwave data could lead to the estimation of roughness, dielectric properties, and the root mean square height distribution.

III. Status of the Sensor Fusion

The initial work on sensor fusion as described in this paper was carried out by identifying objects, parts of which are occluded, shadowed, or wiped by intense specular reflection. This work resulted in a patent[4]. The algorithm developed calculates radar scattering cross sections (RCS) for the visible portion of the object. Then a minimization technique is used to compare measured RCS with the calculated ones. The object geometry is then perturbed to minimize the difference between measured and calculated values. This procedure leads to a

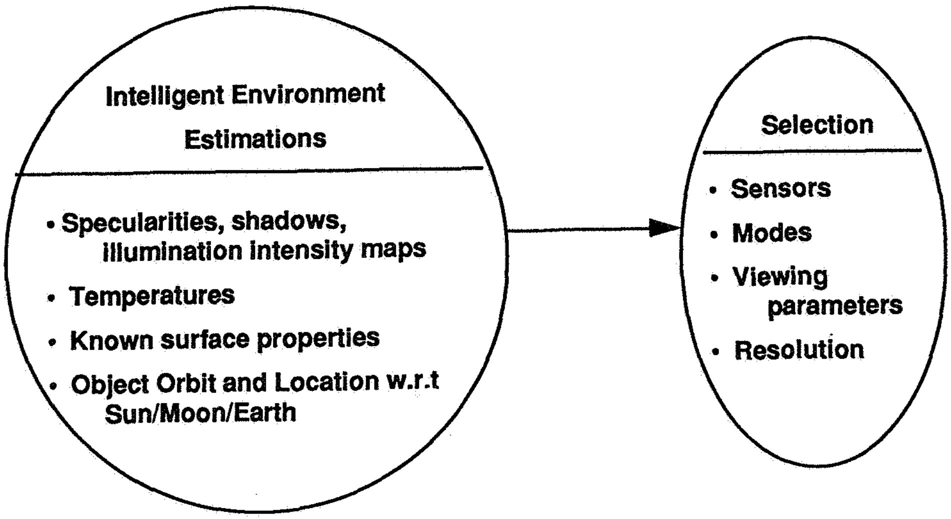


Figure 1. Estimated Environment

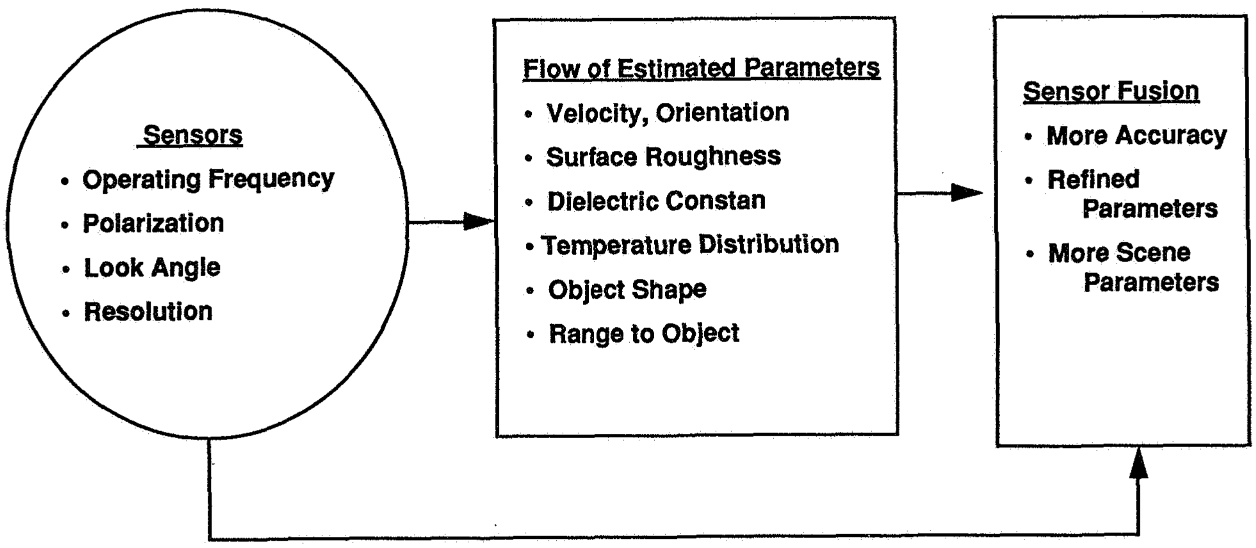


Figure 2. Sensor Fusion Results

better geometrical description of the object. This work was then generalized for the lunar outpost geometry using multisensors[5,6]. An algorithm was developed based on the general discussion presented in Section II of this paper. The description of the object/scene is provided by fusing the appropriate data using physical models. Fuzzy description models are used to alleviate the difficulty of collecting highly-calibrated data. So that the responses of the sensors are identified as very low, low, medium, high, very high, etc., as opposed to numbers. This approach has yielded results using simulated lunar scenes. A spatial map of the needed parameters can be generated through this scheme[6].

IV. Concluding Remarks

The sensor fusion described in this paper advances the state of the art by using unique algorithms based on physical models of scattering and emission from space objects and scenes. The ultimate objective is to be able to switch sensors automatically according to the parameters that need to be estimated through observation. Currently, these models are being refined at JSC and experimental verification plans being developed for the multisensor fusion using laser, optical, visible, infrared, microwave, and radiometer data.

References

[1] Robotic Vision/Sensing for Space Applications, Krishen, K., Rui J. J. deFigueiredo, and Olin Graham, Proceedings of the 1987 IEEE International Conference on Robotics and Automation, 1987, Raleigh, North Carolina, Vol.1, pp. 138-150.

[2] Robotic Vision Technology and Algorithms for Space applications, K. Krishen, Acta Astronautica, Vol. 19, No. 10, 1989, pp. 813-826.

[3] Adaptive Multisensor Fusion for Planetary Exploration Rovers, M. Collin, K. Krishen, and L. Pampagnin, Proceedings of the International Symposium on Artificial Intelligence, Robotics and Automation in Space, organized by CNES, Cepadues Editions, Toulouse Labege, France, September 30-October 2, 1992, pp. 113-116.

[4] Method and Apparatus for Sensor Fusion, K. Krishen, S. Shaw, and R. J. P. deFigueiredo, Patent No. 5,005,147 issued on April 2, 1991.

[5] Sensor Fusion for Assured Vision in Space Applications, Marie-France Collin and Kumar Krishen, Proceedings of Space Operations, Applications and Research 1992 (SOAR '92), NASA Conference Publication 3187, Vol. I, August 4-6, 1992, ppgs. 72-85.

[6] Multiple Sensor Data Integration for Space Operations, M. F. Collin, Kumar Krishen, and Roland Pesty, Proceedings of the SPIE's 1993 Symposium on Aerospace and Remote Sensing, April 12-16. 1993, Marriott Orlando World Center, Orlando, Florida.

57-63
44779

PERCEPTION FOR MOBILE ROBOT NAVIGATION: A SURVEY OF THE STATE OF THE ART

David Kortenkamp
The MITRE Corporation
1120 NASA Road 1
Houston, TX 77058
korten@aio.jsc.nasa.gov

ABSTRACT

In order for mobile robots to navigate safely in unmapped and dynamic environments they must perceive their environment and decide on actions based on those perceptions. There are many different sensing modalities that can be used for mobile robot perception; the two most popular are ultrasonic sonar sensors and vision sensors. This paper examines the state-of-the-art in sensory-based mobile robot navigation. The first issue in mobile robot navigation is safety. This paper summarizes several competing sonar-based obstacle avoidance techniques and compares them. Another issue in mobile robot navigation is determining the robot's position and orientation (sometimes called the robot's *pose*) in the environment. This paper examines several different classes of vision-based approaches to pose determination. One class of approaches uses detailed, *a priori* models of the robot's environment. Another class of approaches triangulates using fixed, artificial landmarks. A third class of approaches builds maps using natural landmarks. Example implementations from each of these three classes are described and compared. Finally, the paper presents a completely implemented mobile robot system that integrates sonar-based obstacle avoidance with vision-based pose determination to perform a simple task.

1 INTRODUCTION

In order for mobile robots to navigate safely in unmapped and dynamic environments they must perceive their environment and decide on actions based on those perceptions. Although there are many different ways to sense the world, most robots use ultrasonic sonar sensors and visions sensors. Typically, mobile robots use sonar sensors to avoid collisions with objects in their environment and vision sensors to localize themselves within their environment. The former is necessary in unmapped or dynamic environments to ensure the safety of the robot and the safety of its surroundings. The latter is necessary because, over time, errors accumulate in the robot's internal location sensors. This paper first looks at several different approaches to sonar-based obstacle avoidance and then examines several different approaches to using vision sensing to localize.

2 OBSTACLE AVOIDANCE

Obstacle avoidance for mobile robots is a topic of much current research. In recent years, two sophisticated obstacle avoidance methods have been implemented: Vector Field Histogram (VFH) [3] and Navigational Templates (NaTS) [20]. These two methods take very different approaches to representing the obstacles in the mobile robot's environment. VFH is the more traditional approach and will be examined first.

2.1 Vector Field Histogram

The VFH obstacle avoidance algorithm uses a two-dimensional Cartesian grid, called the Histogram Grid (illustrated in Figure 1), to represent data from ultrasonic range sensors. Each cell in the Histogram Grid holds a certainty value that represents the confidence of the algorithm in the existence of an obstacle at that location. This representation was derived from the certainty grid concept originally presented in [18].

The central idea behind a certainty grid is to fuse sonar readings over time to eliminate errors in individual sonar readings. In a typical implementation, each cell of the grid array would represent say a 10cm by 10cm square of the

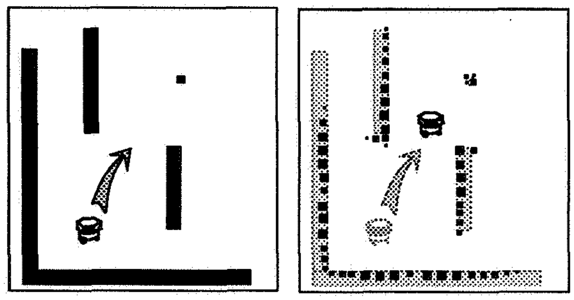


Figure 1: Left figure a robot and an environment. Right figure shows the Histogram Grid of obstacle locations that VFH has created after the robot has moved.

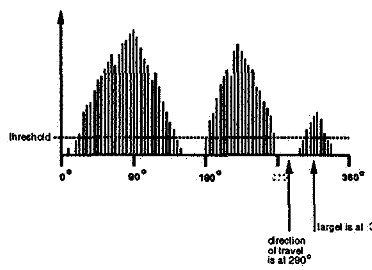


Figure 2: The Polar Histogram has "mountains" in the direction of obstacles. VFH steers the robot toward a "valley" in the direction of the target.

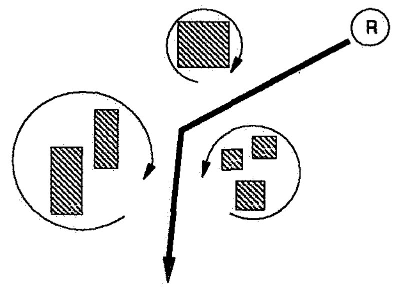


Figure 3: In NaTs each obstacle or group of obstacles is spun. The spin tells the robot to which side of the obstacle it should pass.

environment and the array covers the entire environment space. As the robot travels through the environment, its sonar sensors are continuously being fired and returning range readings to objects. Since the approximate location of the robot at any time is known through odometry and the direction of each sonar sensor is known, the location of objects in the grid array can be estimated. Each time an object is detected at a particular cell location, the value of the cell is increased and the values of all the cells between the robot and this cell are decremented (since they must be empty). The cells have minimum and maximum values, which have been chosen arbitrarily for computational convenience.

To perform the actual obstacle avoidance using the Histogram Grid, VFH creates an intermediate data representation called the Polar Histogram. The purpose of the Polar Histogram is to reduce the amount of data that needs to be handled for real–time analysis while at the same time retaining the statistical information of the Histogram Grid, which compensates for the inaccuracies of the ultrasonic sensors. In this way, the VFH algorithm produces a sufficiently detailed spatial representation of the robot's environment for travel among densely cluttered obstacles, without compromising the system's real–time performance. The spatial representation in the Polar Histogram can be visualized as a mountainous panorama around the robot, where the height and size of the peaks represent the proximity of obstacles, and the valleys represent possible travel directions (see Figure 2). The VFH algorithm steers the robot in the direction of one of the valleys, based on the direction of the target location.

When VFH is combined with an ultrasonic filtering routine called EERUF (Error Eliminating Rapid Ultrasonic Firing) [2] it is uniquely suited to high speed obstacle avoidance. Experiments on a robot at the University of Michigan have demonstrated obstacle avoidance in the most difficult obstacle courses at speeds of up to 1.0 m/sec

2.2 Navigational Templates

One problem with VFH and other obstacle avoidance methods is that it can be difficult for high-level processes to influence low-level behaviors. For example, there is no way to tell VFH to which side of an obstacle it should pass. Navigational Templates (NaTs) are a method for combining high-level, qualitative guidance with low-level, quantitative control. NaTs come in two forms: substrate NaTs (s-NaTs) and modifier NaTs (m-NaTs). An s-NaT defines a gradient field indicating for each position in space the best direction of travel to accomplish the navigation task. This substrate is independent of any obstacles in the environment. m-NaTs are used to model obstacles in the environment. An obstacle or group of obstacles is represented by an m-NaT that has a spin, which constrains the robot's motion. The spin can be clockwise or counter-clockwise. The robot uses the s-NaTs to determine the general direction of travel and uses the m-NaTs to determine how to avoid obstacles. Figure 3 shows an environment with several m-NaTs and the robots path through the environment.

NaTs has been implemented on several mobile robots. Experiments demonstrate effective obstacle avoidance at top speeds of about 0.2 m/sec. Most importantly, NaTs allow higher-level processes to influence the spin of obstacles and determine the robot's response when it encounters those obstacles. For example, when passing people in a hallway, the robot could always pass to the right of person, even if this wasn't the optimal strategy. This is not possible with traditional obstacle avoidance techniques, such as VFH.

3 LOCALIZATION

All of the obstacle avoidance algorithms discussed in the previous section rely on the robot knowing where it is in the environment, so that it can calculate a travel path to its goal location. Robots keep track of their position and orientation (or their *pose*) in their environment by using wheel encoders. However, due to wheel slippage, etc. the robot's internal calculations of its location will accumulate errors. Over time these errors can become significant. Therefore, the robot needs a way of determining its pose in the environment using exterior references, which is called *localization*. This section will look at three different approaches to localization. The first approach uses CAD-like models of the environment, the second uses triangulation to fixed landmarks and the third uses naturally occurring visual cues.

3.1 Using CAD-like models

A popular and successful way of determining position and orientation is to match visual features to a 3-D, CAD-like world model that has been given to the robot. There are several different, working instantiations of this approach including Harvey [8], Mobi [13], FINALE [12] and COSIM [19]. The basic approach of each system is the same: 1) The robot has a detailed, 3-D model of its environment; 2) A robot action is performed and the model is used to predict the location of visual features that the robot should see after performing the action; 3) The locations of these visual features is compared with corresponding features in an actual image taken by the robot and the robot position is then updated. The major difference between systems is in the visual features that they use. Harvey uses vertices, Mobi uses vertical edges that fall within a small band across the center of the image, FINALE uses horizontal and vertical edges and CoSiM uses horizontal edges. Because many of these systems are similar and in the interest of brevity, only one will be examined in detail–COSIM.

3.1.1 COSIM

The COSIM system consists of a mobile robot with a single black and white camera and a 3D model of the robot's environment. COSIM determines its position and orientation by comparing a simulated image generated from the 3D model with an actual camera image. This process consists of five components:

1. Camera calibration routines that use the model and images to determine effective focal length, and the x and y pixel dimensions.
2. A fast and crude orientation corrector that uses vanishing point analysis.
3. An algorithm to match 2D image points to 3D model features.
4. A reverse projection algorithm to produce two sets of 3D match points from a 3D model and 2D image points.
5. An algorithm for determining a registration vector that gives the rotation and translation necessary to align the two sets of 3D points. This registration vector can be used to determine the camera's position and orientation relative to some fixed coordinate frame that is inherent to the 3D model.

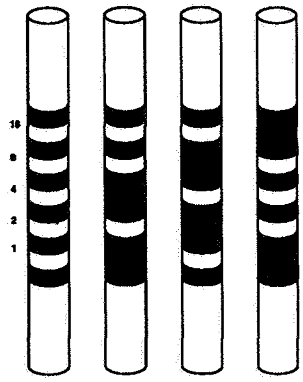


Figure 4: Example object tags showing the basic pattern and objects with bit pattern of 0, 5, 10 and 17.

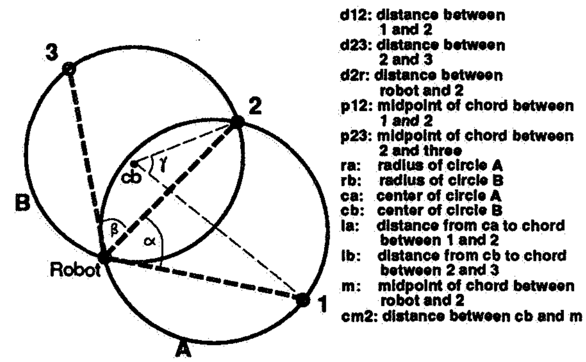


Figure 5: CARMEL's three object triangulation using the circles method.

These five components work together as follows: First, camera calibration (1) is performed to obtain the camera's intrinsic values, which are crucial in all of the other calculations. Calibration will have to be repeated anytime the camera is replaced or when the optical system is adjusted. Next the robot starts moving through the modeled environment and takes images. Each image is analyzed to determine the deviation between the position of the vanishing point in the image from its expected position based on the model and dead reckoning information. This step (2) assumes that there exists a pair or more of dominant parallel lines in the environment (the baseboard lines between the walls and the floor are used in the actual implementation). This technique corrects gross orientation errors very quickly, which allows for expectation matching of features to be performed robustly. COSIM matches image features to model features using a technique based on searching small regions in the image space for particular oriented features (lines) and their intersections. Then a global consistency check is performed in 2D space that eliminates gross matching errors. This produces a set of image points and a matching set of model points. These points are then reverse projected to 3D space. Finally, using a technique based on [1] the two sets of matched 3D points are used to find a correction vector $\vec{q}$. The correction vector is made up of two vectors, a unit rotational quaternion and a translation vector. The correction vector can be used to align the robot with the model, both in position and orientation.

3.2 Using fixed landmarks

Model-based system require a complete model of the environment. This can expensive to produce and maintain. Triangulating to fixed landmarks in the environment can provide position and orientation information without the overhead of a 3-D model. Landmark-based systems rely on a small set of distinctive landmarks that can be seen over a large portion of the environment. Triangulation requires distinct, point-like landmarks that can be recognized from any direction and from long distances. These landmarks can be natural (see [16, 22] for a system that use natural landmarks) or artificial. There are numerous references for triangulation, including [21, 14]. The triangulation approach will be demonstrated in this subsection by examining a robot system, CARMEL, which uses artificial landmarks whose locations have been acquired previously by the robot. First, the landmark recognition system will be described and then the landmark triangulation algorithm will be given.

3.2.1 Landmarks

CARMEL's vision system consists of an object tag design and an algorithm for detecting and distinguishing those tags. The object tag consists of a black and white stripe pattern. An example object tag is shown in Figure 4. The basic stripe pattern is six evenly spaced horizontal black bands of 50 mm width, with the top of the top band and the bottom of the bottom band spaced 1000 mm apart. The white gaps between the black bands correspond to the bit positions in a five-bit word. A white space between two bands corresponds to an "off" bit, while filling the space with black corresponds to an "on" bit. The five bits between the six bands can then represent 32 unique objects.
The actual algorithm for extracting objects from an image requires no preprocessing of the image. The algorithm

makes a single pass over the image, going down each column of the image looking for a white-to-black transition that would mark the start of a potential object. A finite state machine keeps track of the number and spacing of the bands. After finding enough bands to comprise a tag the algorithm stores the tag id and pixel length. Once a column is complete, the eligible objects are heuristically merged with objects found in previous columns. The distance between the top of the top band and the bottom of the bottom band, in terms of the number of pixels in the image, is then used to estimate the actual distance from the camera to the object. The algorithm has an effective range of about 12 meters. See [9] for a detailed description and analysis of the vision algorithm.

3.2.2 Triangulation

The landmark recognition algorithm returns a heading and a distance to any landmark. The triangulation algorithm requires only the relative heading between any three known landmarks to determine the robot's location and orientation. Referring to Figure 5, landmark 1, landmark 2, and the robot form one circle (circle A). Even though the robot's location is not known, there are only two possible circles because the angle between landmarks as viewed by the robot is known. Landmark 2, landmark 3, and the robot also form another unique circle (B). From the information available (landmark locations and the angles α and β) the equations of the circles can be determined. The two circles intersect at landmark 2 and the robot's location. Landmark 2's location is already known and the robot's location is the other intersection point. Below is the algorithm for circle intersection. Refer to Figure 5 for variable definitions used. A landmark's position is denoted by (Lxi, Lyi) where i represents the number of the landmark. A landmark's orientation from the robot is denoted by Loi.

1. Properly order landmarks, see below.
2. $\alpha = \mathrm{Lo2} - \mathrm{Lo1}$. if α is too small or equals 90° or 270°, return with error because division by 0 will occur.
3. $\beta = \mathrm{Lo3} - \mathrm{Lo2}$, if β is too small or equals 90° or 270°, return with error because division by 0 will occur.
4. $ra = \frac{d12}{2\sin(\alpha)}$
5. $rb = \frac{d23}{2\sin(\beta)}$
6. $la = \frac{d12}{2\tan(\alpha)}$
7. $lb = \frac{d12}{2\tan(\beta)}$
8. Let $v12x$ and $v12y$ be the unit vector from landmark 1 to landmark 2.
9. Let $v23x$ and $v23y$ be the unit vector from landmark 2 to landmark 3.
10. $cax = p12x - la(v12y)$
11. $cay = p12y + la(v12x)$
12. $cbx = p23x - lb(v23y)$
13. $cby = p23y + lb(v23x)$
14. Return an error if the centers of the two circles are too close (we used 10 units).
15. If γ is very large, then return error.
16. Let $cbax$ and $cbay$ be the unit vector from the center of circle B to the center of circle A.
17. $d2r = 2rb\sin(\gamma)$
18. $c2m = rb\cos(\gamma)$
19. Robot x position $= Rx = 2mx - Lx2 + 0.5$

20. Robot y position = $Ry = 2my - Ly2 + 0.5$

21. $\phi = \arctan(\frac{Ly1-Ry}{Lx1-Rx})$, the heading of landmark 1 from the true robot position.

22. If $\phi > 0.0$ then robot orientation error $= -(Lo1 - \phi)$ else robot orientation error $360° + \phi - Lo1$

23. Return with solution

For this algorithm to work properly, both α and β must be less than 180°. When this condition holds we say that the landmarks are "ordered" properly. Properly ordered landmarks assure that the desired solution (out of the two solutions possible) is found. Properly ordered landmarks have the following two features:

1. They are labeled consecutively (1,2,3) in a counterclockwise fashion,
2. The angle between landmarks 1 and 2 (β) and the angle between landmarks 1 and 3 (α) must be less than 180°.

This triangulation algorithm is analyzed and compared with other triangulation algorithms in [5].

3.3 Using visual cues

A drawback to landmark triangulation is that it relies on knowing the exact location of fixed landmarks and on being able to see them from anywhere in the robot's environment. Such assumptions make these algorithms difficult to use in environments such as office buildings, where the field of view is very restricted. In these environments it may not be possible to localize at any place, instead, the robot may have to do *place recognition.* Place recognition involves using naturally occurring visual cues to determine whether or not the robot is at one of a number of predefined locations in the environment. For example, in an office environment, the robot could use visual cues to determine in which room it is located. However, when the robot is traveling between rooms it does not know its precise location in the environment. Representations that store only interesting places and information about how to travel between interesting places are called *topological maps* and were pioneered by Brooks [4] and Kuipers [15]. Topological maps have become quite popular in mobile robotics and there are many implementations, including [6] and [17]. This section will focus on one particular system, RPLAN [10].

RPLAN consists of two modules, a sonar sensing module that detects interesting places in an indoor, office-like environment and a vision module that identifies those interesting places. Interesting places are openings that the robot finds as it follows a hallway or follows a wall. These openings presumably lead to new spaces that the robot can explore. At each of these openings, the robot takes an image. Vertical edges are extracted from a black-and-white image by a modified Sobel edge detector (see Figure 6). A second image is analyzed in the same way as the first, but it is offset by 18cm from the first image. Once there is a list of edges from this pair of images the edges are matched across the image using the direction of transition of the edge (i.e., was the edge from light to dark or dark to light?), length, and location. The pixel shift in the matched edges from the first image to the second image (called the disparity) is calculated and used to determine a rough distance to the edge. Each visual cue, thus, has three scene-independent features: direction, length, and distance. These cues are stored and then, when the robot comes back to the same place, it can match the cues that it sees with those that it has stored and determine its position. Experiments in natural, unaltered environments, show that by using a combination of sonar and vision information, place recognition can be extremely robust.

4 AN INTEGRATED SYSTEM

Obstacle avoidance and vision sensing are not issues that should be studied separately, as they are both necessary for an intelligent mobile robot. In this section, I look at a completely implemented system that performs a task using both sonar-based obstacle avoidance and vision-based sensing. The robot is CARMEL, a Cybermotion K2A base with a ring of 24 sonar sensors and a rotating camera (see Figure 7). CARMEL has all of its processing entirely on-board in the form of three PCs. CARMEL's task is to explore a 20m by 20m arena, locate ten objects and approach each object. This task was part of the AAAI'92 robot competition, which CARMEL won [7].

CARMEL's strategy was to mark the 10 objects with the barcode tag described in subsection 3.2.1 of this paper. CARMEL then moved to predetermined points in the arena and performed a visual sweep for objects, noting

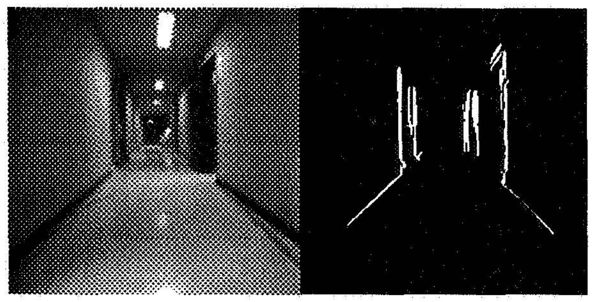


Figure 6: An actual image and its visual cues.

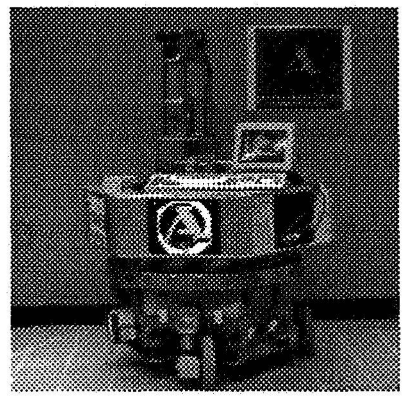


Figure 7: The mobile robot CARMEL.

their locations. When all of the objects were mapped, CARMEL then proceeded to approach each one, choosing the object nearest its current location as the goal. While moving between objects, CARMEL avoided randomly placed obstacles using the VFH obstacle avoidance algorithm described in subsection 2.1. For the competition run, CARMEL's maximum speed was set at 0.5 m/sec, approximately twice the speed of the other robots. CARMEL also had various high-level error recovery routines that allowed it to cope with unexpected situations, such as blocked paths or objects that had been incorrectly located. CARMEL could also triangulate its position using the algorithm described in subsection 3.2 of this paper. However, due to some last minute code changes, the triangulation routines were never invoked during the competition run.During the competition, CARMEL found and visited all 10 objects in just under 10 minutes, about twice as fast as any other robot. Details of CARMEL's software design can be found in [11].

5 CONCLUSION

Perception is an important issue in mobile robot research. If a mobile robot cannot perceive its environment correctly and efficiently, then it will not be able to perform even simple tasks. Sonar sensors and vision sensors are often used for mobile robot perception and they are each suitable for different things. Sonar sensors are typically used to avoid collisions, whereas vision sensors are typically used to perform localization. Robot designers should carefully chose the correct sensor for their task. A successful robot system, such CARMEL, can be developed by carefully integrating sonar and vision sensing.

6 ACKNOWLEDGMENTS

Ulrich Rashke and Dr. Johann Borenstein reviewed the VFH sections of this paper. Charles Cohen and Frank Koss reviewed the triangulation section of this paper. Clare Congdon drew several of the figures. The CARMEL robot is part of the University of Michigan Artificial Intelligence Laboratory. Funding for its purchase and for research using it is provided by Department of Energy grant number DE-FG02-86NE37969.

REFERENCES

[1] Paul J. Besl and Neil D. Mckay. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2), 1992.

[2] Johann Borenstein and Yoram Koren. Noise rejection for ultrasonic sensors in mobile robot applications. In *The Proceedings of the IEEE Conference on Robotics and Automation*, 1992.

[3] Johann Borenstein and Yorem Koren. Real-time obstacle avoidance for fast mobile robots. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(5), 1989.

[4] Rodney A. Brooks. Visual map making for a mobile robot. In *Proceedings IEEE Conference on Robotics and Automation*, 1985.

[5] Charles Cohen and Frank Koss. A comprehensive study of three-object triangulation. In William J. Wolfe and Wendell H. Chun, editors, *Mobile Robots VII*. SPIE, Bellingham, Washington, 1993.

[6] Thomas Dean, Kenneth Basye, Robert Chekaluk, Seungseok Hyun, Moises Lejter, and Margaret Randazza. Coping with uncertainty in a control system for navigation and exploration. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, 1990.

[7] Thomas Dean and R. Peter Bonasso. 1992 AAAI robot exhibition and competition. *AI Magazine*, 14(1), 1993.

[8] Claude Fennema and Allen R. Hanson. Experiments in autonomous navigation. In *Proceedings Image Understanding Workshop*, 1990.

[9] Marcus J. Huber, Clint Bidlack, Kevin Mangis, David Kortenkamp, L. Douglas Baker, Annie Wu, and Terry Weymouth. Computer vision for CARMEL. In William J. Wolfe and Wendell H. Chun, editors, *Mobile Robots VII*. SPIE, Bellingham, Washington, 1993.

[10] David Kortenkamp. *Cognitive maps for mobile robots: A representation for mapping and navigation.* PhD thesis, The University of Michigan, 1993.

[11] David Kortenkamp, Marcus Huber, Charles Cohen, Ulrich Raschke, Clint Bidlack, Clare Bates Congdon, Frank Koss, and Terry Weymouth. Integrated mobile robot design: Winning the AAAI-92 robot competition. *IEEE Expert*, 8(4), August 1993.

[12] A. Kosaka and Avi C. Kak. Fast vision-guided mobile robot navigation using model-based reasoning and prediction of uncertainties. *Computer Vision, Graphics, and Image Processing*, 56(2), 1992.

[13] David J. Kriegman, Ernst Triendl, and Thomas O. Binford. Stereo vision and navigation in buildings for mobile robots. *IEEE Transactions on Robotics and Automation*, 5(6), 1989.

[14] E. Krotkov. Mobile robot localization using a single image. In *Proceedings IEEE Conference on Robotics and Automation*, 1989.

[15] Benjamin J. Kuipers and Yung-Tai Byun. A robot exploration and mapping strategy based on a semantic hierarchy of spatial representations. *Robotics and Autonomous Systems*, 8, 1991.

[16] Tod S. Levitt and Daryl T. Lawton. Qualitative navigation for mobile robots. *Artificial Intelligence*, 44(3), 1990.

[17] Maja K. Mataric. Integration of representation into goal-driven behavior-based robots. *IEEE Transactions on Robotics and Automation*, 8(3), 1992.

[18] Hans P. Moravec and Alberto Elfes. High resolution maps from wide angle sonar. In *Proceedings IEEE Conference on Robotics and Automation*, 1985.

[19] Yuval Roth, Annie S. Wu, Remzi H. Arpaci, Terry Weymouth, and Ramesh Jain. Model-driven pose correction. In *Proceedings IEEE International Conference on Robotics and Automation*, 1992.

[20] Marc G. Slack. Navigation templates: Mediating qualitative guidance and quantitative control in mobile robots. *IEEE Transactions on Systems, Man, and Cybernetics*, 23(2), 1993.

[21] K. Sugihara. Some location problems for robot navigation using a single camera. *Computer Vision, Graphics, and Image Processing*, 42(3), 1988.

[22] Saburo Tsuji and Shigang Li. Memorizing and representing route scenes. In Jean-Arcady Meyer, Herbert L. Roitblat, and Stewart W. Wilson, editors, *From Animals to Animats 2: Proceedings of the Second International Conference on Simulation of Adaptive Behavior*. MIT Press, Cambridge, MA, 1993.

omit

Session R2: AUTOMATION TECHNOLOGIES

Session Chair: Kathleen Jurica

COMPUTER-AIDED OPERATIONS ENGINEERING WITH INTEGRATED MODELS OF SYSTEMS AND OPERATIONS

Jane T. Malin
NASA
Automation and Robotics Division
Johnson Space Center, ER2
Houston, TX 77058

Dan Ryan
MITRE
Johnson Space Center, ER2
Houston, TX 77058

Land Fleming
Lockheed Engineering & Sciences Co.
Houston, TX 77058

ABSTRACT

CONFIG 3 is a prototype software tool that supports integrated conceptual design evaluation from early in the product life cycle, by supporting isolated or integrated modeling, simulation, and analysis of the function, structure, behavior, failures and operation of system designs. Integration and reuse of models is supported in an object-oriented environment providing capabilities for graph analysis and discrete event simulation. Integration is supported among diverse modeling approaches (component view, configuration or flow path view, and procedure view) and diverse simulation and analysis approaches. Support is provided for integrated engineering in diverse design domains, including mechanical and electro-mechanical systems, distributed computer systems, and chemical processing and transport systems. CONFIG supports abstracted qualitative and symbolic modeling, for early conceptual design. System models are component structure models with operating modes, with embedded time-related behavior models. CONFIG supports failure modeling and modeling of state or configuration changes that result in dynamic changes in dependencies among components. Operations and procedure models are activity structure models that interact with system models. CONFIG is designed to support evaluation of system operability, diagnosability and fault tolerance, and analysis of the development of system effects of problems over time, including faults, failures, and procedural or environmental difficulties.

INTRODUCTION

The core of engineering design and evaluation focuses on analysis of physical design. Thus, today's computer-aided engineering software packages often do not provide enough support for conceptual design early in the life cycle or for engineering for operation, fault management, or supportability (reliability and maintainability). Benefits of engineering for operations and supportability include more robust systems that meet customer needs better and that are easier to operate, maintain and repair. Benefits of concurrent engineering include reduced costs and shortened time for system development. Integrated modeling and analysis of system function, structure, behavior, failures and operation is needed, early in the life cycle.

Conventional system modeling approaches are not well suited for evaluating conceptual designs early in the system life cycle. These modeling approaches require more knowledge of geometric or performance parameters than is usually available early in design. More abstracted models can support early conceptual design definition and evaluation, and also remain useful for some later analyses. Component-connection models provide one such useful abstraction, and discrete events are another. Discrete event simulation technology combines both abstractions, for evaluation of conceptual designs of equipment configurations in operations research (3). CONFIG uses these abstractions, with some enhancements, to define and evaluate conceptual designs for several types of systems.

The initial CONFIG project goal was to support simulation studies for design of automated diagnostic software for new life-support systems (9). The problem was to design an "expert system" on-line troubleshooter before there was an expert. The design engineer could use a model of the system to support what-if analyses of failure propagation, interaction, observability and testability. This activity is similar to Failure Modes and Effects Analysis (5), but uses comparative simulations of failure effects to develop diagnostic software. Conventional simulation software was not up to this challenge, but discrete event simulation software has been. CONFIG supports the use of qualitative models for applying discrete event simulation to continuous systems.

A major design goal for CONFIG is to support conceptual design for operations and safety engineering. Major tasks in conceptual design are design definition, evaluation (by simulation and analysis) and documentation. Operations engineering focuses on the design of systems and procedures for operating, controlling and managing the system in normal or faulty conditions. Safety engineering focuses on prevention of hazardous effects and conditions in the physical system or its operation. In these types of engineering, complex interactions and interfaces among system components and operations must be a focus.

Another design goal of CONFIG is to bridge the gaps between physical design engineering and other types of engineering. Component-connection representations are well suited for modeling and defining physical system designs (as structures of interacting components) and operations designs (as structures of interacting actions), as well as interactions between system components and operational actions. Discrete event models have been used for this type of modeling for queueing and scheduling problems, but can be extended to support conceptual modeling in operations and safety engineering. This type of modeling is also compatible with systems engineering function diagrams (1).

CONFIG 3

The project approach has been to incrementally integrate advanced modeling and analysis technology with more conventional technology. The prototype integrates qualitative modeling, discrete event simulation and directed graph analysis technologies for use in analyzing normal and faulty behaviors of dynamic systems and their operations. The prototype has been designed for modularity, portability and extensibility. A generic directed graph element design has been used to standardize model element designs and to promote extensibility. This directed graph framework supports integration of levels of modeling abstraction and integration of alternative types of model elements.

Enhanced Discrete Event Simulation Capabilities

In traditional discrete event modeling and simulation, state changes in a system's entities, "events", occur discretely rather than continuously, at nonuniform intervals of time. Throughout simulation, new events are added to an event list that contains records of events and the times they are scheduled to occur. Simulation processing jumps from one event to the next, rather than occurring at a regular time interval. Computation that results in creation of new events is localized in components, which are connected in a network. Any simulation run produces a particular history of the states of the system. Statistical simulation experiments, using random variables in repeated simulation runs, are used to compare design alternatives.

To enhance this discrete event simulation approach to accommodate continuous systems, a number of new concepts and methods were developed. These include component models with operating modes, types of links connecting components ("relations" and "variable clusters"), state transition structures ("processes"), methods for representing qualitative and quantitative functions ("process language"), and a new simulation control approach.

Digraph Analysis Capabilities

The CONFIG Digraph Analyzer (DGA) makes graph analysis techniques available for evaluating conceptual designs of systems and their operations. The DGA is based on reachability search, and is implemented generically for application to the many types of graph data structures in CONFIG. DGA can support analyses of completeness, consistency and modularity. Analysis of failure sources and impacts can be done by tracing the paths from a given failure.

System Modeling

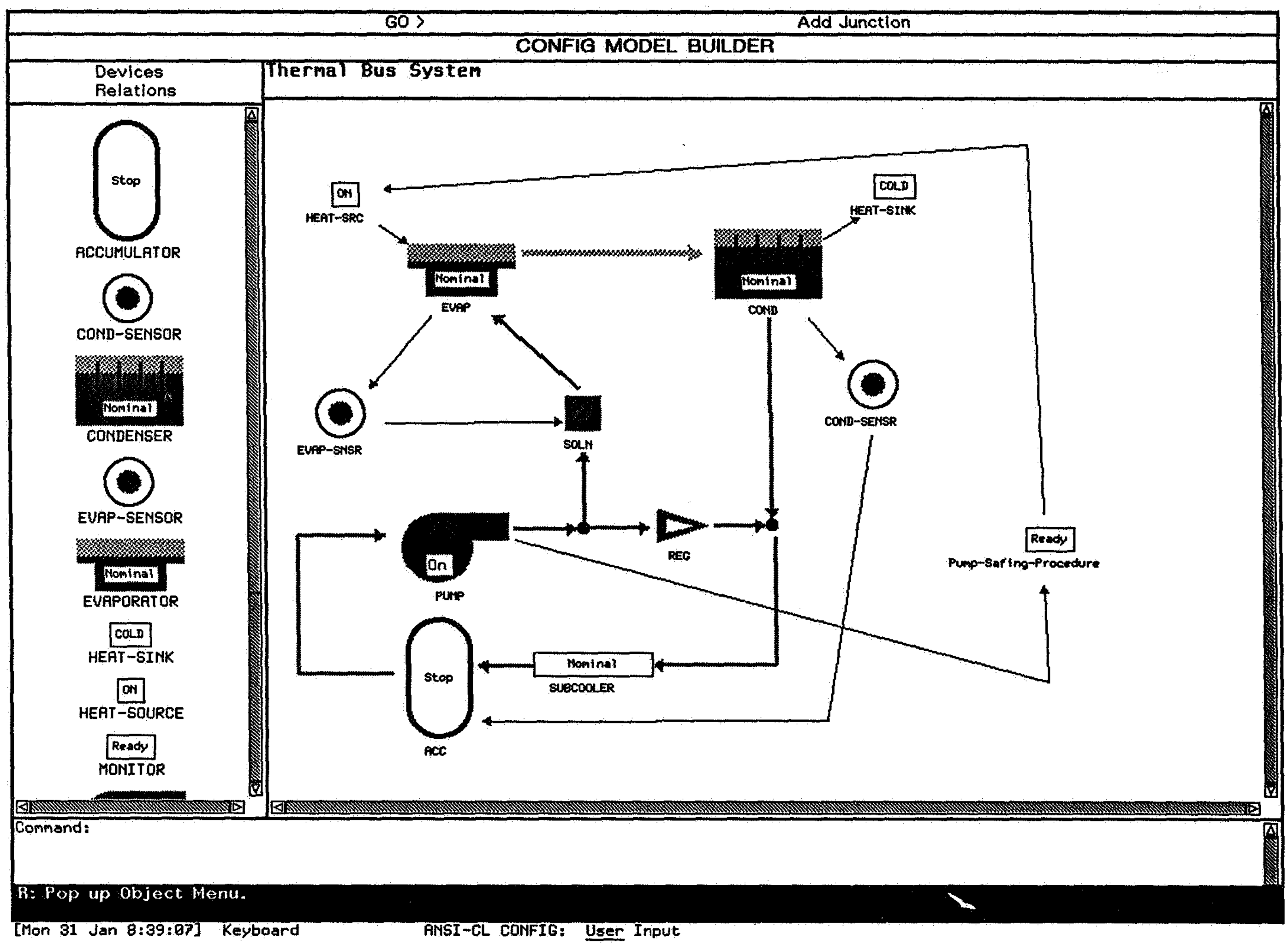
Devices are the basic components of a CONFIG system model. Relations are basic connectors for building topological model structures with devices. Device behavior is defined within operating and failure Modes, which contain mode dependent and mode transition Processes. Modes are connected together in a mode transition digraph which delineates the transition dependencies among the individual modes. A Thermal Bus System model, shown in the Figure, illustrates a CONFIG model of a system of connected devices (heat exchangers, valves, sensors, accumulator, etc.) for removing heat. The behavior of each device depends on its current mode and on changes in its input received from other devices via relations or via the global flow path manager.

Device Processes define changes in device variables as events with time delays, which are conditionally invoked and executed during a simulation. Processes define time-related behavioral effects of changes in device input variables, both direction of change and the new discrete value that will be reached, possibly after a delay. Faults and failures can be modeled in two distinctly different ways. Failure modes can be used to model device faults. Mode-transition processes can be used to model latent failures or triggering failures that prevent or cause mode changes. Relations connect devices via their variables, so that state changes can propagate along these relations during simulations. Related variables are organized into variable clusters, to separate types of relations by domain (e.g., electrical vs. fluid connections). Relations can also be used to make connections between system Devices and Activities in operations models.

Flow Path Modeling

Flow is a property of many systems, whether the substance flowing is a liquid or information. There are two difficulties in modeling flows with local device processes. First, flow is a global property of the topology of the modeled system and the substances flowing within it. Second, while dynamic changes in system structure and flow can occur during operations, process descriptions involving flow must often rely on assumptions of static system topology. These factors would limit the reusability of device descriptions to a limited set of system structures.

A flow-path management module (FPMM) has been implemented to address these problems. The FPMM is separate from the module implementing local device behavior, but the two modules are interfaced via flow-related state variables in the devices. When FPMM is notified during simulation of a local change in device state, it recomputes the global effects on flows produced by the local state change. The FPMM then updates the state of flow in all affected


GO >
Add Junction
CONFIG MODEL BUILDER
Devices
Relations
Thermal Bus System
Stop
ACCUMULATOR
COND-SENSOR
Nominal
CONDENSER
EVAP-SENSOR
Nominal
EVAPORATOR
COLD
HEAT-SINK
ON
HEAT-SOURCE
Ready
MONITOR
ON
HEAT-SRC
COLD
HEAT-SINK
Nominal
EVAP
Nominal
COND
EVAP-SNSR
SOLN
COND-SENSR
On
PUMP
REG
Ready
Pump-Safing-Procedure
Stop
ACC
Nominal
SUBCOOLER
Command:
R: Pop up Object Menu.
[Mon 31 Jan 8:39:07] Keyboard
ANSI-CL CONFIG: User Input

devices. This design permits the user to write reusable local device process descriptions that do not depend on any assumptions concerning the system topology.

FPMM uses a simplified representation of the system, as a collection of aggregate objects, or "circuits." Further abstraction is achieved by identifying serial and parallel clusters in the circuit (6). In many cases, configuration determination alone is sufficient to verify flow/effort path designs, or to establish flow paths for a continuous simulation, for reconfiguration planning, and for troubleshooting analysis (see Ref. [2] on cluster-based design of procedures for diagnosis, test, repair and work-around in a system).

Operations modeling

Activities are the basic components of a CONFIG operations model, and are connected together in action structures with Relations. They represent procedures or protocols that interact with the system, to control and use it to achieve goals or functions. Each activity model can include specifications for what it is intended to achieve or maintain. Activity behavior is controlled in a sequence of phases, ending in an evaluation of results. Activity behavior is defined by processes that model direct effects of actions, or that control device operation and mode transitions to achieve activity goals. Relations define sequencing and control between activities and connect Devices with device-controlling Activities.

Operations models are designed to support operation analysis with procedure models. These models are designed to support analysis of plans and procedures for nominal or off-nominal operation. The procedure modeling elements are designed for reuse by intelligent replanning software, and for compatibility with functional modeling in systems engineering.

Model Development and Integration Capabilities and Approach

CONFIG provides intelligent automation to support nonprogrammer and nonspecialist use and understanding. CONFIG embeds object-oriented model libraries in an easy-to-use toolkit with interactive graphics and automatic programming.

CONFIG provides extensive support for three separable yet tightly integrated phases of user operation during a modeling session: Library Design, Model Building, and Simulation and Analysis. This includes a graphical user interface for automated support of modeling during each of the phases including the development of object-oriented library element classes or templates, the construction of models from these library items, model inspection and verification, and running simulations and analyses.

The integration between the phases enables an incremental approach to the modeling process. Lessons learned from analyses can be repeatedly and rapidly incorporated by the user into an initially simple model. Support for these phases as separate user activities fosters the achievement of concurrent engineering goals. Different users can define library elements, build models, and analyze models at different times depending on area of expertise and availability of resources. Support for the model building phases spans all types of modeling that can be performed in CONFIG including component structure, behavior and flow, and activity goals and structure.

Hosting

CONFIG is implemented in software that is portable to most Unix work stations. The Common LISP Object System (CLOS) is a highly standardized language, with compilers for most of the commonly available work stations. The user interface was implemented using the Common LISP Interface Manager (CLIM), another standardized tool built on CLOS.

CONCLUSIONS

CONFIG is designed to model many types of systems in which discrete and continuous processes occur. The CONFIG 2 prototype was used to model and analyze: 1) a simple two-phase thermal control system based on a Space Station prototype thermal bus, 2) a reconfigurable computer network with alternate communications protocols, and 3) Space Shuttle Remote Manipulator System latching and deployment subsystems (7). The core ideas of CONFIG have been patented (8). CONFIG 3 has added capabilities for graph analysis and for modeling operations and procedures.

The CONFIG prototype demonstrates advanced integrated modeling, simulation and analysis to support integrated and coordinated engineering. CONFIG supports qualitative and symbolic modeling, for early conceptual design. System models are component structure models with operating modes, with embedded time-related behavior models. CONFIG supports failure modeling and modeling of state or configuration changes that result in dynamic changes in dependencies among components. Operations and procedure models are activity structure models that interact with system models. The models support simulation and analysis both of monitoring and diagnosis systems and of operation itself. CONFIG is designed to support evaluation of system operability, diagnosability and fault tolerance, and analysis of the development of system effects of problems over time, including faults, failures, and procedural or environmental difficulties.

ACKNOWLEDGEMENTS

The authors wish to thank Bryan Basham, Leslie Ambrose, Ralph Krog, Debra Schreckenghost, Brian Cox, Daniel Leifker and Sherry Land for their contributions to CONFIG design and implementation.

REFERENCES

1. Alford, M., "Strengthening the System Engineering Process," ENGINEERING MANAGEMENT JOURNAL, Vol. 4, No. 1, March, 1992, pp. 7-14.

2. Farley, A. M., "Cluster-based Representation of Hydraulic Systems," PROC. 4TH CONFERENCE ON AI APPLICATIONS, March, 1988, pp. 358-364.

3. Fishman, G. S., PRINCIPLES OF DISCRETE EVENT SIMULATION, Wiley, New York, NY, 1978.

4. Forbus, K., "Qualitative Physics: Past, Present, and Future," EXPLORING ARTIFICIAL INTELLIGENCE, Eds. H. Shrobe and AAAI, Morgan Kaufmann, San Mateo, CA, 1988.

5. Fullwood, R. R. and Hall, R. E., PROBABILISTIC RISK ASSESSMENT IN THE NUCLEAR POWER INDUSTRY: FUNDAMENTALS AND APPLICATIONS, Pergamon Press, 1988.

6. Liu, Z. and Farley, A. M., "Structural Aggregation in Common-Sense Reasoning," PROC. 9TH NATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE (AAAI-91), AAAI Press, Menlo Park, CA, July, 1991, pp. 868-873.

7. Malin, J. T., Basham, B. D., and Harris, R. A., "Use of Qualitative Models in Discrete Event Simulation for Analysis of Malfunctions in Continuous Processing Systems," ARTIFICIAL INTELLIGENCE IN PROCESS ENGINEERING, Ed. M. Mavrovouniotis, Academic Press, 1990, pp. 37-79.

8. Malin et al., U. S. PATENT 4,965,743, "Discrete Event Simulation Tool for Analysis of Qualitative Models of Continuous Processing Systems," October, 1990.

9. Malin, J. T., and Lance, N., "Processes in Construction of Failure Management Expert Systems from Device Design Information," IEEE TRANS. ON SYSTEMS, MAN, AND CYBERNETICS, Vol. SMC-17, 1987, pp. 956-967.

10. Malin, J. T. and Leifker, D. B., "Functional Modeling with Goal-Oriented Activities for Analysis of Effects of Failures on Functions and Operations," INFORMATICS AND TELEMATICS, Vol 8, No. 4, 1991, pp. 353-364.

59-63

44784

A PROTOTYPE SUPERVISED INTELLIGENT ROBOT FOR HELPING ASTRONAUTS

J. D. Erickson, K. A. Grimm, and T. W. Pendleton
Automation and Robotics Division/ER
NASA-Lyndon B. Johnson Space Center

ABSTRACT

The development status is described of a prototype supervised intelligent robot for space application for purposes of (1) helping the crew of a spacecraft such as the Space Station with various tasks such as holding objects and retrieving/replacing tools and other objects from/into storage, and for purposes of (2) retrieving detached objects, such as equipment or crew, that have become separated from their spacecraft. In addition to this set of tasks in this low Earth orbiting spacecraft environment, it is argued that certain aspects of the technology can be viewed as generic in approach, thereby offering insight into intelligent robots for other tasks and environments.

Also described are characterization results on the usable reduced gravity environment in an aircraft flying parabolas (to simulate weightlessness) and results on hardware performance there. These results show it is feasible to use that environment for evaluative testing of dexterous grasping based on real-time visual sensing of freely rotating and translating objects.

INTRODUCTION

Numerous facets contribute to achieving robotic intelligence. This paper, based on a more complete presentation in [1], describes many of these facets and attempts to relate them to the central theme of a software architecture that enables a sufficient level of robotic intelligence, and thus, real work in real environments under supervision by exception.

The essence of intelligent systems is that they are capable of collecting and applying knowledge of the situation gained at execution time and correlating it with other knowledge to take effective actions in achieving goals. Intelligent systems are composed of sensors for perceiving both the external and internal environments, effectors for acting on the world, and computer hardware and software systems providing intelligent connection between the sensors and effectors.

Part of the processing by these computer systems is symbolic in a nonnumeric sense and thus enables practical reasoning, or the behavior which in humans we call intelligent. The intelligent system we will be addressing, the Extravehicular Activity Helper/Retriever (EVAHR), is a supervised, intelligent, mobile robot with arms and end effectors Intelligent robots of this nature are required for long-term operations in space and are mandatory for space exploration to improve safety, reliability, and productivity, while enabling large cost savings through minimizing logistics [2].

PROBLEM STATEMENT

Long-term space operations such as the Space Station have requirements for capabilities for rescue of EVA crew and retrieval of equipment. A space station cannot chase separated crew or equipment, and other vehicles such as the Space Shuttle will not usually be available. In addition to the retrieval of drifting objects, another need is for robotic help to EVA crewmembers in various tasks such as holding objects; retrieving and replacing tools and other items from and into storage; performing inspections; setting up and dismantling work sites; performing servicing, maintenance, and repairs; and deploying and retrieving payloads. Modeling, simulation, and analysis studies of space exploration missions have shown that supervised intelligent robots are enabling for human exploration missions [3, 4].

The free-flying, supervised intelligent robot called EVA Helper/Retriever (EVAHR) is being prototyped as a potential solution to the crew helper and detached crew and equipment retrieval need. EVAHR is a technology test-bed providing evaluation and demonstration of the technology included for the following three purposes: 1) Robotic retrieval of objects which become detached from their spacecraft; e.g., astronauts adrift from the Space Station. 2) A robotic crew helper around a spacecraft; e.g., inspector, "go-fer," holder, maintainer, servicer, tester, etc. 3) A "generic" prototype supervised, intelligent autonomous robot (for planetary surfaces with different mobility such as wheels or tracks and for terrestrial applications with appropriate adaptations).

Early supervised intelligent robotic systems with initial capabilities to meet real needs are beginning to emerge from laboratories and manufacturers. It is now possible, in our opinion, to construct robots capable of accomplishing several specific high level tasks in unstructured real world environments.

The ability to acquire and apply knowledge and skills to achieve stated goals in the face of variations, difficulties, and complexities imposed by a dynamic environment with significant unpredictability is our working definition of "robotic intelligence." This does not require a broad-based general intelligence or common sense by the robot. However, doing the work needed to accomplish goals does require, in general, both mobility and manipulation in addition to reacting, or deciding "intelligently" at each step what to do. Further, supervised intelligent robots are required for human-robot teams where supervision is most naturally provided by voice.

Certain aspects of the EVAHR technology, which provide the capability for performing specified tasks in a low-Earth-orbiting spacecraft environment, can be viewed as generic in approach, thereby offering insight into intelligent robots for other tasks and environments. This is because the design of the software architecture, which is the framework (functional decomposition) that integrates the separate functional modules into a coherent system, is dictated in large measure by the tasks and nature of the environment. And because both the goal-achieving tasks and the partially unpredictable nature of the environments are similar on Earth and in space, the software architecture can be viewed as generic - as can many of the software modules, such as the AI planner, world model, and natural language interface. Other software is bundled with certain hardware. This leads to the concept of a modular, end-user customized robot, put together from modules with standard interfaces [5-7] such as users do with a personal computer, yet maintaining real-time response.

APPROACH

The end goal for intelligent space robot development is one or more operational robots as part of human/robot teams in space. Prior to that, an evaluation of performance in space will be required.

Our approach to development of operational robots as part of human-robot teams in space is a systems engineering approach with an iterative, three-ground-phase requirements prototype development, tested in both ground and aircraft simulations of space, followed by evaluation testing of a flight test article in space. We adapt and integrate existing technology solutions.

The EVA Helper/Retriever ground-based technology demonstration was established to design, develop, and evaluate an integrated robotic hardware/software system which supports design studies of a space borne crew rescue/equipment retrieval and crew helper capability. Goals for three phases were established. The Phase I goals were to design, build, and test a retriever system test-bed by demonstrating supervised retrieval of a fixed target. Phase II goals were to enhance the test-bed subsystems with significant intelligent capability by demonstrating arbitrarily oriented target retrieval while avoiding fixed obstacles. The objectives for Phase III, which is currently in progress, are to more fully achieve supervised, intelligent, autonomous behavior by demonstrating grasp of a moving target while avoiding moving obstacles and demonstrating crew helper tasks. Phase III is divided into two parts. Phase IIIA goals are to achieve real-time complex perception and manipulator/hand control sufficient to grasp moving objects, which is a basic skill both in space retrieval and in accomplishing the transition from flying to attaching to a spacecraft. Phase IIIB goals are to achieve a software architecture for manipulation and mobility, with integrated sensing, perception, planning, and reacting, which guarantees safe, robust conduct of multiple tasks in an integrated package while successfully dealing with a dynamic environment.

Our overall testing approach is short cycle run-break-fix with increasing integration and more relevant environments; such an approach finds design and implementation problems early when they are lowest cost to fix.

The performance characteristics of the EVAHR hardware enable (or defeat) the "intelligent" behavior of the robot as "animated" by the software. We are testing only a subset of the Phase IIIB hardware in Phase IIIA.

The hardware subset includes a seven degree of freedom arm (Robotics Research K807i), a five degree of freedom, compliant, force-limited dexterous hand, a laser range imager (Perceptron), a stereo video camera system (Teleos Prism 3), a pan/tilt unit, a 700 Megaflop computational engine employing i860s and transputers, and an Inertial Measurement Unit (IMU) of accelerometers and gyros.

During Phase IIIA we are using a subset of our reaction plan architecture while we are exploring two new approaches to the software architecture for Phase IIIB. The first is a version of the 3-tiered, asynchronous, heterogeneous architecture for mobile robots [8-10] adapted to include manipulation. The second is a version of the SOAR architecture [11] applied to robots [12]. SOAR is of interest because of its capabilities in learning, including recent work in situated, interactive natural language instruction [13].

For each approach we are conducting evaluation testing of minimal prototype architecture implementations to obtain some evidence of their strengths and weaknesses for our tasks before selecting one for larger scale implementation in Phase IIIB.

Safety is a major issue in human-robot teams, especially in space. Since robotic motion control programs cannot be considered safe unless they run in hard real time, an approach which addresses this issue in a different manner from that of the 3-tiered architecture is needed for comparative evaluation. We are pursuing the development of one such approach [14].

The pivotal problem in successfully coupling symbolic reasoning with the ability to guarantee production of a timely response is that the timing of actions taken by a real-time system must have low variances, so that the effects of those actions on unfolding processes can be predicted with sufficient accuracy. But intelligent software reserves the option of extended searching, which has very high variance. Therefore, when building a system that must act in real time as well as reasoning, one can choose to either 1) Subject the AI component of the system to hard deadlines. This effectively embeds the AI reasoner within the real-time system, and under time pressure, results in loss of intelligent function. 2) Refuse to subject the AI component of the system to hard deadlines, and have the real-time subsystem "do its best" with whatever commands the AI subsystem can generate in time. This effectively embeds the real-time subsystem within the AI system, and under time pressure, results in loss of timely control. 3) Refuse to subject the AI component of the system to hard deadlines, but let the AI components "negotiate" with the real-time subsystem to obtain a feasible schedule for task execution. This does not embed either subsystem within the other, and with proper selection of the real-time executive's task schedule, has the promise of remaining functional under time pressure. The 3-tiered approach is a category three approach, whereas we interpret SOAR to be a category two approach.

We can now summarize the state of the art. Simple control systems can get away with seeming to be "fast enough," but that approach becomes potentially very dangerous in more complex systems, particularly in intelligent systems where the set of tasks being executed changes over time. In a system that may perform any subset of N possible tasks, there are 2^N possible combinations of tasks, and it becomes impossible to test the performance of each combination by hand when N is large. Therefore, it becomes imperative to have automated support for obtaining a guarantee that the system can always perform in hard real time.

THREE-TIERED SOFTWARE ARCHITECTURE

Combining all prior knowledge and knowledge sensed during a task requires that planning in advance can only be guidance, with control decisions as to what to do postponed until such time as the situation is being sensed and the task is being executed. This is the essence of Agre and Chapman's theory of plans-as-advice [15], and is a design principle underlying the 3-tiered approach. The three tiers are the planner, the sequencer, and the reactive controller. The responsibility of the planning layer is to determine which tasks would accomplish the goal, and in what approximate order. Thus, the planning layer forms a partially ordered set of tasks for the robot to perform, with temporal constraints. The AI planner which we are evaluating for this application is the AP Planner [16]. It may be possible to use SOAR for this application.

The sequencing "middle" layer is responsible for controlling sequences of primitive physical activities and deliberative computations. Operating asynchronously from the planner, yet receiving inputs from that layer, the sequencer takes the sketchy plan and expands it, based on the current situation. Thus, the hierarchical plan expansion happens at execution time rather than at the deliberative stage. To implement the sequencer, data structures called Reactive Action Packages (RAP's) are used to represent tasks and their methods for executing [9].

At the lowest level, the reactive controller accepts sensing data and action commands, sensorimotor actions that cannot be decomposed any further, from the sequencer. For example, "move," "turn," or "grasp" are all examples of action commands that are passed on to the hardware. The reactive controller also monitors for success or failure of these commanded activities.

PHASE IIIA RESULTS TO DATE

Results from Phase II have been reported previously [17]. Some preliminary results from Phase IIIA have also been reported [18-23].

SOAR Evaluation for Phase IIIB

SOAR was selected for study as a promising candidate system for the EVAHR planning system. SOAR is a symbolic AI architecture which emphasizes problem-solving, planning, and learning.

One major advantage of SOAR is its ability to learn by taking a new experience, and saving the sequence of steps to the goal as a "chunk." This chunk is in the form of a set of production rules, and if the same scenario is encountered in the future, the associated chunk will execute without having to search for the correct sequence as it did initially.

From our experience with Hero-SOAR, a subset of SOAR for a Hero robot, we know that the reactivity of SOAR is an important capability needed to respond to the environment quickly. SOAR may be seen as a system with a planner, which plans in the traditional sense yet with no actual data structure produced; a mechanism to execute the plan; and a fast replanning ability.

Phase IIIA Computer Simulation Results

Software modules for grasping of free-floating objects in a zero-g, 6-DOF environment have been described in previous sections. Results of performance testing of these modules as subsystems are described in this section. The modules have also been integrated and tested in our orbital and KC-135 simulations[24], and these results are also described below.

Search is the first visual function to be performed when there is no knowledge about the location of an object of interest. It is carried out as follows[25,26]: EVAHR's front hemisphere is divided into concentric "rings," and each ring is further divided into sectors, each of which is enclosed by the FOV of the sensor. Each search starts from the center ring and spirals outward until an object is found.

Algorithms for image-based pose estimation have been implemented. Several objects were chosen for testing. These objects include some orbital replaceable units (ORU), a star tracker, a jettison handle, and some wrenches.

To test the robustness of the software, 500 tests were run on each test object with actual poses of the object randomly oriented using a random number generator in (simulated) images. Noise was added to the "range" component of the image to test the sensitivity of the algorithms to noise. There were two indications from the test results: 1) Most estimation errors are less than 5 degrees (with up to 3 percent noise in range); 2) The performance of the pose estimation software gradually degraded with increasing noise in range measurements.

The rotational state estimator uses intermittent delayed poses from the pose estimator software to provide the arm trajectory planner with current estimates of the target's rotational state at the rate of 100 Hz. The estimator utilizes an extended Kalman filter because of the inherent nonlinear nature of rotational dynamics. The effects of varying various parameters on the performance of the standalone rotational state estimator have been reported [19]. Testing on the integrated rotational state estimator shows it converges within 4 pose estimates (about 4 sec) and maintains error estimates of less than 3 degrees, which meets requirements.

The relative translational state estimator used for the KC-135 experiment does not use an inertial coordinate system. The equations describing the dynamics are nonlinear. Therefore, the estimator design is based on an extended Kalman filter. The results of its performance in the KC-135 simulator show an accuracy similar to that for the orbital case[27].

Integrated software testing in the orbital simulation has concentrated on and produced results in two areas: (1) determining the overall system performance against grasping different moving targets with random initial states and (2) determining the computational requirements for the pose estimation software, using rate and delay as parameters. Grasp impact dynamics calculations are made to verify that the target is not knocked away during the grasp or by a prior collision with the arm. Under these conditions, the system has achieved a >70-percent successful grasp rate for both objects tested. The state estimates have less than 1 inch and five degrees of error.

Results from the second suite of tests show that pose estimation rate and delay also have a direct effect on the time-to-grasp in successful tests. Assuming pose estimation rate and delay of 0.1 sec, we were able to estimate that six i860 processors would be sufficient to achieve these rates and delays.

AIRCRAFT REDUCED GRAVITY ENVIRONMENT

Some microgravity research can be conducted inside an aircraft simulating space by flying vertical parabolic flight paths, but only for very limited amounts of time. During Phase IIIA we are flying a subset of the EVAHR Phase IIIB hardware and software aboard the NASA Reduced Gravity Program's KC-135 aircraft. This aircraft flies a series of parabolic trajectories resulting in approximately 15 sec of near microgravity (<.01-g) in the cabin during each parabola. The robotic arm, hand, vision sensor with pan/tilt system, and IMU (Inertial Measurement Unit of accelerometers and gyroscopes) is attached to the floor of the aircraft. During microgravity, an object is released, tracked by the vision system, and grasped by the hand. All of these objects have a complex construction with multiple graspable points.

On several KC-135 preliminary flights, data characterizing the reduced gravity was collected from an IMU placed on the cabin floor. Video recordings also were made of objects floating during the reduced gravity interval. The vertical acceleration fluctuated significantly about zero-g. Fluctuations between 75 mg and -75 mg were commonplace. These fluctuations caused the released object to accelerate toward either the ceiling or floor of the airplane. Lateral accelerations were also observed and were due to air turbulence, flight path corrections, or other effects.

An evaluation of 38 parabolas was performed, and the trajectory duration determined. This interval started when the target was released and continued until the target hit the inside of the airplane fuselage, or was touched by personnel, or left the FOV of both video cameras. The results indicate 2/3 of the parabolas have 4 seconds or less of usable microgravity.

These results, especially the trajectory durations, do not match well with the extrapolation to the KC-135 of time-to-grasp results from the orbital simulation presented above.

In a separate flight of the KC-135, we exercised the unintegrated hardware subsystems (except the stereo cameras) independently. All of the hardware is designed to operate in a 1-g environment and might behave differently in the KC-135 in microgravity or after the 1.8-g pullout at the bottom of the parabolas. Motions and operations representative of those that will be used in later object tracking and grasping evaluations were used in these tests. All equipment was determined to operate without measurable changes in behavior from that expected.

CONCLUSIONS

The need for crew help and retrieval of detached crew and equipment in space has been identified. Evaluation of the practical realization of a potential solution has passed several successful milestones, but is still ongoing, with many of the critical developments yet to come. The potential solution described here is an initial attempt to build and understand a prototype of a supervised intelligent robot for use in space. It is also potentially useful in terms of the software architecture for many U.S. economy-related robot applications on Earth.

Both our Phase II and Phase IIIA results demonstrate that manipulation requires greater accuracy of sensing and perception than does mobility. Integrated testing with our Phase IIIA computer simulation has not only shown that we have a workable software design, but has also afforded us systems engineering analyses supporting computer hardware design for achieving real-time complex perception processing (sensor to percept) and grasp control (percept to action) for freely moving objects.

Our future plans are first to complete the metrology of the manipulator and joint calibration of both vision-system-manipulator pairs. We are recoding the laser scanner pose estimation software to run in real time on the i860 network. The tracker and translational state estimator are currently running in real time on i860's. The manipulator trajectory controller and grasp planner are running in real time on the transputer network. Grasp testing using targets mounted on the object-motion unit are being conducted in preparation for the KC-135 vision-guided grasping flights. Then we have several moving object grasp evaluation flights to conduct. Phase IIIB developments are dependent on the selection of a final software architecture from the preliminary prototyping efforts which are underway using a set of crew helper tasks, scenarios, and computer simulation environments with human-injected unpredictable events to assess the value of the many goal-planning and real-time reaction aspects of the supervised intelligent robot design.

ACKNOWLEDGMENTS

This effort is internally supported by the Center Director's Discretionary Fund and by the Automation and Robotics Division of the NASA Lyndon B. Johnson Space Center.

REFERENCES

1. J. D. Erickson et al., "An Intelligent Space Robot for Crew Help and Crew and Equipment Retrieval," International Journal of Applied Intelligence, accepted for publication in 1994.
2. J. D. Erickson et al., "SEI Planet Surface Systems Humans/Automation/ Robotics/Telerobotics Integrated Summary," document JSC-45100, NASA Johnson Space Center, Houston, TX, Mar. 1991.
3. J. D. Erickson, "Needs for Supervised Space Robots in Space Exploration," 1992 World Space Congress, IAF-92-0800, International Astronautical Federation, Paris, France, Aug. 1992.
4. J. D. Erickson et al., "Some Results of System Effectiveness Simulations for the First Lunar Outpost," 1993 AIAA Space Programs and Technology Conf., Huntsville, AL, Sept. 1993.
5. R. P. Bonasso and M. G. Slack, "Ideas on a System Design for End-User Robots," in Proc. SPIE 1829 Cooperative Intelligent Robotics in Space III, Nov. 1992, p. 352.
6. R. O. Ambrose, M. P. Aalund, and D. Tesar, "Designing Modular Robots for a Spectrum of Space Operations," in Proc. SPIE 1829 Cooperative Intelligent Robotics in Space III, Nov. 1992, p. 371.
7. F. daCosta et al., "An Integrated Prototyping Environment for Programmable Automation," in Proc. SPIE 1829 Cooperative Intelligent Robotics in Space III, Nov. 1992, p. 382.
8. E. Gat, "Integrating Planning and Reacting in a Heterogeneous Asynchronous Architecture for Controlling Real-World Mobile Robots," in AAAI-92, Proc. 10th Natl. Conf. on AI, San Jose, CA, July 1992.
9. R. J. Firby, "Adaptive Execution in Dynamic Domains," Ph.D. Thesis, Yale University, 1989.
10. R. P. Bonasso, "Using Parallel Program Specifications for Reactive Control of Underwater Vehicles," Jour. of Applied Intelligence, Kluwer Academic Publishers, Norwell, MA, June 1992.
11. J. E. Laird, A. Newell, and P. S. Rosenbloom, "SOAR: An Architecture for General Intelligence," Artificial Intelligence, Vol. 33 , No. 1, pp. 1-64, 1987.
12. J. E. Laird et al. "RoboSOAR: An Integration of External Interaction, Planning, and Learning Using SOAR," Robotics and Autonomous Systems, Vol. 8, 1991.
13. S. B. Huffman and J. E. Laird, "Learning Procedures from Interactive Natural Language instructions," in Proc. 10th Intl. Conf. on Machine Learning, June 1993.
14. M. Schoppers, "Ensuring Correct Reactive Behavior in Robotic Systems," AIAA/NASA JSC Workshop on Automation and Robotics '93, NASA Johnson Space Center, Houston, TX, Mar. 24, 1993.

15. P. Agre and D. Chapman, "Pengi: An Implementation of a Theory of Activity," in Proc. AAAI, 1987, pp. 268-272.
16. C. Elsaesser and T. R. MacMillan, "Representation and Algorithms for Multiagent Adversarial Planning," MTR-91W000207, MITRE Corp, McLean, VA, Dec. 1991.
17. J. D. Erickson et al., Technology Test Results from an Intelligent, Free-Flying Robot for Crew and Equipment Retrieval in Space," Proc. SPIE 1612, Cooperative Intelligent Robotics in Space II, Boston, MA, November 1991.
18. M. L. Littlefield, "Adaptive Tracking of Objects for a Mobile Robot Using Range Images," in Proc. SPIE 1829, Cooperative Intelligent Robotics in Space III, Boston, MA, Nov. 1992, pp. 433-443.
19. L. Hewgill, "Motion Estimation of a Freely Rotating Body in Earth Orbit," in Proc. SPIE 1829, Cooperative Intelligent Robotics in Space III, Boston, MA, Nov. 1992, pp. 444-457.
20. K. A. Grimm et al., "Experiment in Vision-Based Autonomous Grasping Within a Reduced Gravity Environment," in Proc. SPIE 1829, Cooperative Intelligent Robotics in Space III, Boston, MA, Nov. 1992, pp. 410-420.
21. C. H. Chien, "A Computer Vision System for Extravehicular Activity Helper/Retriever, in Proc. SPIE Conference on Sensor Fusion and Aerospace Applications, Orlando, FL, Apr. 1993.
22. R. S. Norsworthy, "Performance Measurement of Autonomous Grasping Software in a Simulated Orbital Environment," in Proc. SPIE 2057, Telemanipulator Technology and Space Robotics, Boston, MA, Sept. 1993.
23. G. Anderson, "Grasping a Rigid Object on Zero-G," in Proc. SPIE 2057, Telemanipulator Technology and Space Robotics, Boston, MA, Sept. 1993.
24. R. S. Norsworthy, "Simulation of the EVAHR/KC-135 for Software Integration/Testing," Proc. AIAA/NASA Conf. on Intelligent Robotics in Field, Factory, Service, and Space, Houston, TX, March 1994.
25. M. L. Littlefield, "Real-Time Tracking of Objects in a KC-135 Microgravity Experiment," Proc. AIAA/NASA Conf. on Intelligent Robotics in Field, Factory, Service, and Space, Houston, TX, March 1994.
26. E. L. Huber and K. Baker, "Object Tracking with Stereo Vision," Proc. AIAA/NASA Conf. on Intelligent Robotics in Field, Factory, Service, and Space, Houston, TX, March 1994.
27. L. Hewgill, "Motion Estimation of Objects in KC-135 Microgravity," Proc. AIAA/NASA Conf. on Intelligent Robotics in Field, Factory, Service, and Space, Houston, TX, March 1994.

S10-38

44789

FAILURE ENVIRONMENT ANALYSIS TOOL APPLICATIONS

Ginger L. Pack
NASA Johnson Space Center
Houston, Texas 77058

David B. Wadsworth
Lockheed Engineering and Sciences Company
Houston, Texas 77058

ABSTRACT

Understanding risks and avoiding failure are daily concerns for the women and men of NASA. Although NASA's mission propels us to push the limits of technology, and though the risks are considerable, the NASA community has instilled within it, the determination to preserve the integrity of the systems upon which our mission and, our employees lives and well-being depend. One of the ways this is being done is by expanding and improving the tools used to perform risk assessment. The Failure Environment Analysis Tool (FEAT) was developed to help engineers and analysts more thoroughly and reliably conduct risk assessment and failure analysis. FEAT accomplishes this by providing answers to questions regarding what might have caused a particular failure; or, conversely, what effect the occurrence of a failure might have on an entire system. Additionally, FEAT can determine what common causes could have resulted in other combinations of failures. FEAT will even help determine the vulnerability of a system to failures, in light of reduced capability. FEAT also is useful in training personnel who must develop an understanding of particular systems. FEAT facilitates training on system behavior, by providing an automated environment in which to conduct "what-if" evaluation. These types of analyses make FEAT a valuable tool for engineers and operations personnel in the design, analysis, and operation of NASA space systems.

INTRODUCTION

FEAT was developed as part of an effort to find ways to better identify and understand potential failures that threaten the integrity of NASA systems. Past and current methods of failure assessment consists of developing often enormous amounts of documentation in the form of Failure Mode Effect Analysis (FMEA) worksheets. Engineers create these worksheets by attempting to exhaustively enumerate potential system failures and consequences. Hazards analysis is performed in a similar manner; experts are gathered together and are asked to brainstorm about the hazardous manifestations of various failures. System knowledge and experience are necessary for ensuring the comprehensiveness of this approach. However there are troubling drawbacks to this technique. First, there exists the difficulty of anticipating every scenario. Analysis is also inherently constrained by the limits of actual experience. Further, such methods lack consistency and do not enforce a standard level of coverage. Although there is certainly much to be credited to knowledge acquired through experience, it is not sufficient to avoid unanticipated interactions which may lead unexpectedly to undesirable consequences. As many industries have learned, sometimes experience comes at too high a cost. Those at NASA have been looking for better ways to anticipate failure and for tools to assist in "designing out" potential problems. FEAT was developed to address this problem.

TECHNICAL APPROACH

FEAT is a software application that uses directed graphs or, digraphs, to analyze failure paths and failure event propagation. The behavior of the systems to be analyzed is represented as a digraph. Then, the digraph model of the system, is used by FEAT to answer questions concerning the cause and effects of events which are captured in the model. Therefore, the first step in using FEAT is to create the digraph model of the system in which one is interested. Once FEAT has analyzed the digraph, it has the information it needs to perform cause and effect analysis.

What are digraphs? Directed graphs are graphs that consists of a set of vertices and a set of edges, where there is an edge from one vertex a to another vertex b. The vertices are drawn as circles and the edges are drawn as arrows. The direction of the arrows indicates a causal relationship between the vertices (see figure 1). The vertex from which the edge begins, is called its source; and the vertex at which the edge terminates, is called its target. Direct graph theory is an accepted and established area of mathematical study. Therefore we will only introduce it in this paper, to the extent necessary for an understanding of how it is used in FEAT. The interested reader may find further information by consulting the literature.

Figure 1

The structure of the digraph can be represented by a matrix, and consequently can be easily implemented in a computer. The conversion from digraph to matrix is straightforward and is illustrated below in figure 2. This matrix is called the *adjacency* matrix (reference 1), and is the basis from which other information about the graph can be derived. The matrix of the graph is obtained by entering either zero or one, depending on whether or not an edge connects two vertices. The presence of an edge from a to b in figure 1, indicates an entry of one (1) into the corresponding matrix entry. However, since there is no edge from a to c, a zero (0) would be entered in the corresponding matrix entry.

	a	b	c
a	0	1	0
b	0	0	1
c	0	0	0

Figure 2

Additional information can be added to the digraph, by applying logical operators to express conditional statements. FEAT uses AND and OR operators to accomplish its analysis. The AND operator is represented on the graph as a vertical bar with a horizontally placed arrow at its center. An OR operator is simply two or more edges whose target is the same vertex. Theses operators [figure 3], and their use in FEAT [figures 4 & 5], are described below.

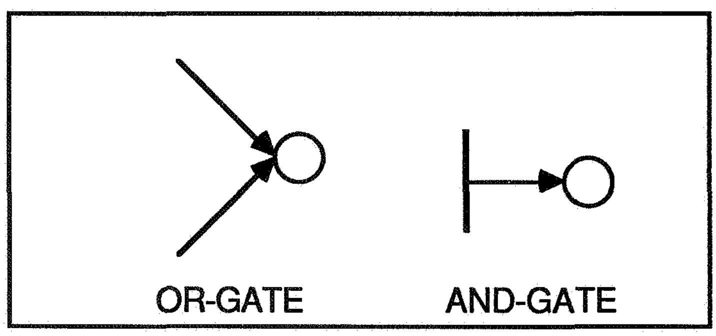


Figure 3

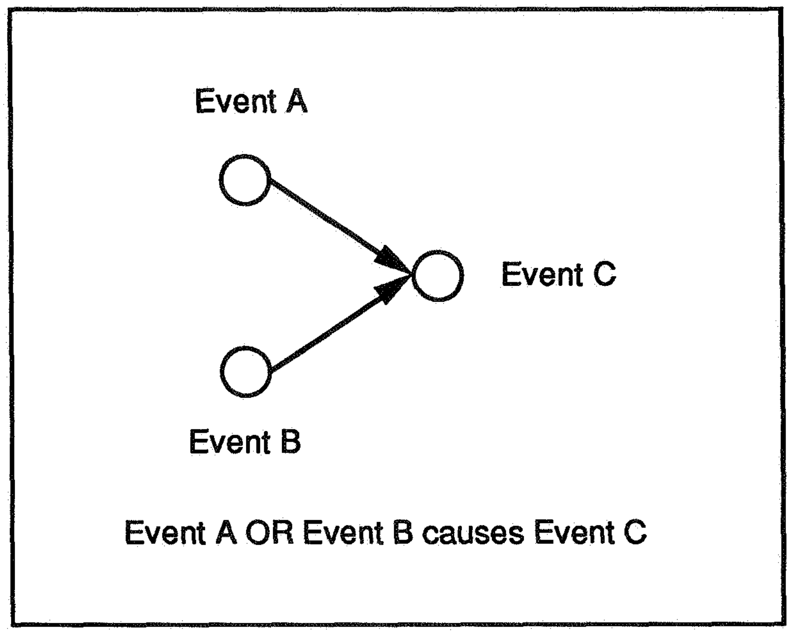


Figure 4

The "AND" gate is shown in Figure 5. The AND gate is used when both event A and Event B must occur in order for Event C to occur. Conversely, if only Event A occurs or, if only Event B occurs, then Event C does not occur.

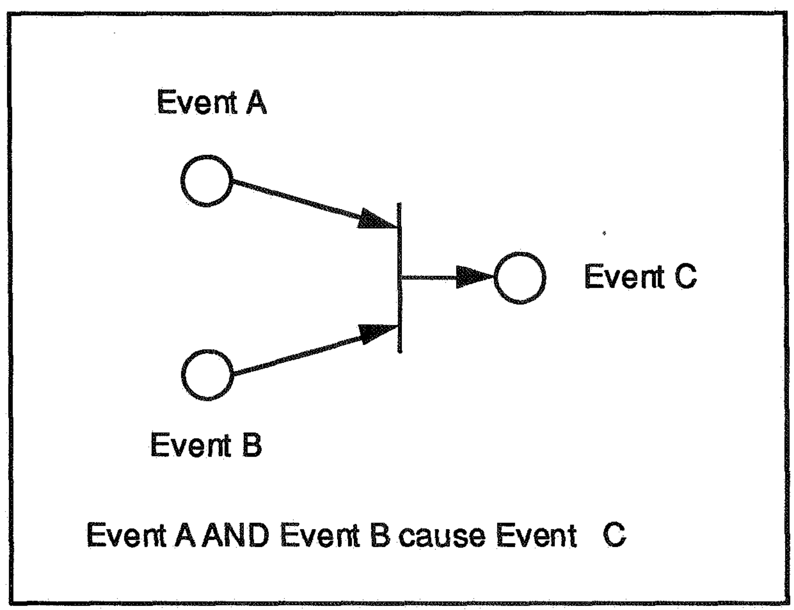


Figure 5

Analytical Capabilities The *reachability* of an event refers to whether there is a path by which other events in the digraph can be reached. A given event is said to reach another, if the first event can cause the second through some path of the graph. Using the adjacency information derived from the digraph, reachability can be computed for every event and pair of events in the digraph. Analysis can be conducted upstream or downstream from an event node. (References 2, 3 and 4 provide a much more detailed discussion of digraphs and reachability.)

Reachability information allows FEAT to answer the following questions about a modeled system:

A. What happens to the system if "Event A (and Event B and Event C and ...)" occurs?
B. What are the possible causes of "Event A"?
C. What common cause could account for the simultaneous indication of numerous events?
D. What is the susceptibility of the system to new events given that one or more events has already occurred, or the system has been reconfigured due, for example, to maintenance?

Digraph Example The following example demonstrates how a digraph might be implemented for a light and switch. The digraph provides a methodical way in which to express the topology and behavior of a system. It is worth noting that the digraph itself may have various constructions for the same information contained in it, depending on who created it. Different modelers may lay out the digraph differently. However, for a properly constructed digraph, the same information will be captured. In the following example [figures 6 & 7], power source A provides current to switch A which connects to the bulb. Similarly, power source B can energize the bulb.

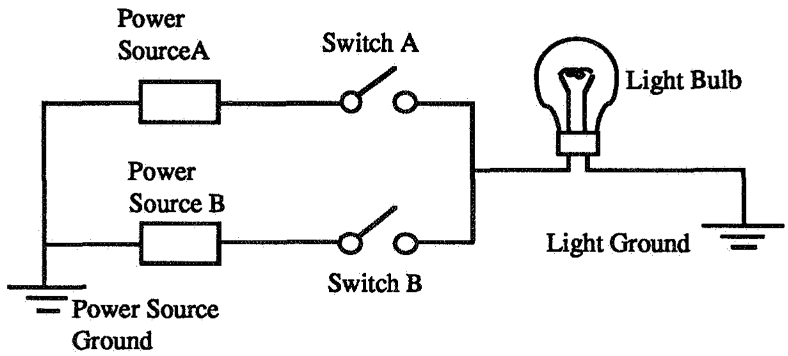


Figure 6
Light bulb and Power Source Schematic

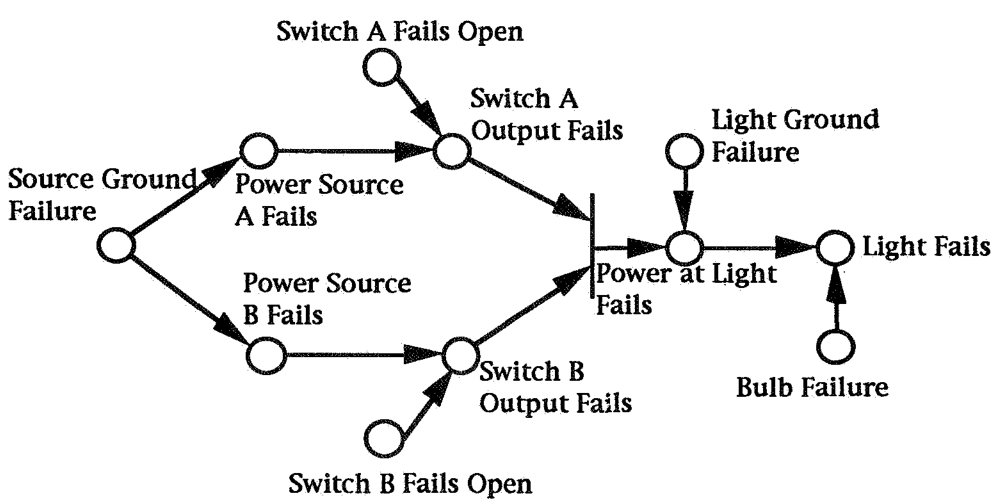


Figure 7
Digraph of Light bulb Schematic

- If "Power Source A Fails" or "Switch A Fails Open" then "Switch A Output Fails". This is an example of OR logic and is shown in the digraph by the arrows leading into "Switch A Output Fails".

- If output from both switches A and B fail, then they will cause the "Power at Light to Fail". This logic appears as an **AND** gate on the digraph (the vertical line). In this case, the **AND** gate reflects redundancy designed into a system.

Why digraphs?

Directed graphs are useful because they visually depict the logical topology and dependency relationships of physical and conceptual systems and processes. Because they capture causal effects between events, they can be used to describe system behavior. Directed graphs are also easily converted into a matrix and, because of this, can be readily analyzed in a computer. Creating and laying out the digraph of a system, also formalizes the method of evaluation during the analytical process, and provides a standard representation convention. Finally, digraph analysis is mathematically sound, since methods for determining connectivity paths of the digraph vertices can be mathematically proved.

DIRECTED GRAPHS AND FEAT

Digraph construction is facilitated by use of an editor specifically designed for the task. Such an editor is included in the FEAT package which consists of two programs: Digraph Editor and FEAT.

Digraph Editor

The Digraph Editor facilitates construction of the digraph model by allowing the user to create event nodes, edges, and the logic operators, and to connect and arrange them into a digraph. Event nodes and edges are laid out and connected using the logic operators. The pieces that make up a digraph are supplied in a digraph toolbox from which items may be selected. These items are placed on the screen and arranged to produce the system digraph.

Other information is needed to complete the digraph and to make it usable by FEAT. Event nodes have an associated text block, which includes information that will identify the event node to FEAT, describe the event for the user, and relate the event to a drawing which contains the component to which the event pertains. This information is extracted from tables that the user creates. Digraph Editor uses the tables to automatically generate a mnemonic reference that FEAT will use to identify the event.

Digraph Editor also provides a number of tools for validating and verifying the model as it is being developed. Digraph Editor will check tables for duplicate entries, check nodes for incorrect form, and determine whether a selected node has a duplicate in the digraph. Digraph Editor also contains an algorithm that allows the user to analyze small or incomplete digraphs, while still in the editor. Once the digraph is completed and the paths in it are analyzed, FEAT can return answers to questions regarding the behavior of the modeled system.

Currently, digraph models are created manually by selecting and arranging digraph components; the modeler must interpret drawings and other sources of information to generate the digraphs. This is a laborious task. Consequently, efforts are underway to develop methods to automatically translate schematics and drawings into corresponding digraph models.

Digraph Editor is currently only available for the Macintosh II class of computer.

FEAT

FEAT is the portion of the package that analyzes single or multiple digraphs, and graphically displays causes and effects of events. Propagation results are shown both on the digraphs and on an another associated graphical representation, such as a schematic or block diagram. FEAT uses a multi-step algorithm, described in Reference 2, to compute reachability for each event and pair of events in the digraphs. Events are identified to FEAT through the mnemonic that is generated by Digraph Editor. Queries about the behavior of the system are made by selecting events and telling FEAT to return all of the causes of that event (targeting), or by telling FEAT to return all of the effects of that event (sourcing). FEAT displays all of the single events, and all pairs of events that may cause a selected event. Multiple events may also be selected and analyzed. FEAT allows some events to be temporarily removed from the analysis so that answers can be obtained about a reconfigured system.

FEAT also contains a feature which allows users to attach to a schematic, formatted database information and graphics. In this way, component descriptions, parts lists, drawings, etc, may be displayed in conjunction with a schematic.

One of the major advantages of FEAT, as discussed in Reference 2, is that it allows the analysis of very large systems. Large systems can be digraphed by creating and connecting a series of smaller digraphs. FEAT understands when propagation occurs across the digraphs.

Planned enhancements to FEAT include the following: increasing the speed with which reachability is computed by improving FEAT's computational algorithm; provision of a method for computing and displaying probabilities of events occurring; and computation and display of the time it takes for an event to propagate through the graph.

FEAT is currently available for the Macintosh II class of computer and for UNIX/X-Windows/OSF-Motif systems. No programming skill is required to use FEAT. However, a course in digraph modeling is quite helpful in learning how to construct system models.

DIGRAPHS AT NASA

Why NASA chose digraphs

NASA's interest in digraphs began as part of the Shuttle Integrated Risk Assessment Project (SIRA). SIRA was initiated in the wake of the Challenger accident, in an effort to find better ways of assessing risk and preventing failure. Digraphs support such analysis by providing end to end cause and effect analysis of modeled systems. Digraphs also provide a standard and methodical approach for conducting safety analysis and risk assessment. Digraphs capture information in an easily retrievable format, and facilitate the transfer of design information. FEAT takes advantage of these characteristics in a way that aids engineers and analysts with design, assists safety engineers with risk assessment, and promotes understanding of system behavior, thereby making FEAT a good tool for training inexperienced persons.

What has been done at NASA?

The first system to which digraph analysis was applied was the Space Shuttle Main Engine System (SSME). Since then, acceptance of digraphs and the use of FEAT has extended in several directions. Most recently, FEAT has been formally released to the Space Station Freedom Program (SSFP) Technical Management Information System (TMIS), as Digraph Data System (DDS) Release 1.0. DDS will, through TMIS, be available to SSF Engineering and Integration, SSF Combined Control Center, and the various Work Packages and their contractors. A Macintosh Powerbook version of FEAT will be deployed as a Development Test Objective (DTO) on the STS-52 flight scheduled for October 1992. Reliability and Maintainability personnel at NASA-JSC, are using FEAT to construct a model of the Simplified Aid for Extra-Vehicular Activity (EVA) Rescue (SAFER). FEAT is also being used to model the redesigned Servo Power Amplifier (SPA) for the Remote Manipulator System (RMS).

Proponents have used FEAT for a variety of analytical tasks, such as Fault Tolerance Analysis and Redundancy Management (FT/RM), Fault Detection, Isolation, and Recovery (FDIR), and "What-If" analysis. Within the Space Station Freedom Program, FEAT is being used in the performance of Integrated Risk Assessment for the station, which includes Failure Mode and Effects Analysis (FMEA), Hazards Analysis (HA), and FT/RM. FEAT has also been established as a baselined tool in the Mission Operations Combined Control Center, where flight controllers will use FEAT models to assist with real-time monitoring tasks. FEAT's role is expanding in both Space Station and in Space Shuttle.

Space Station The Space Station Engineering Integration Contractor (SSEIC), is using FEAT to perform integrated risk assessment. This tasks consists of performing the analysis to assure the station design is safe, reliable, and has an acceptable level of risk (reference 5). The space station design consists of modules designed and built by the United States, and of modules which will be designed and built by NASA's international partners. The work to be performed by NASA is divided into four Work Packages distributed among different centers. Additionally, a variety of contractors are working in support of the Work Packages. Consequently, system integration is a paramount concern of the program. SSEIC is tasked with ensuring the integration of these various factions and is using digraph-based FEAT, to work the integration problem. Specifically, FEAT supports the following areas of the Integrated Risk Assessment process:

1. Reliability Analysis
2. Safety Analysis
3. Integrated Risk Analysis
4. Integrated Risk Assessment

The models being developed for the station Integrated Risk Assessment will eventually be provided to Mission Operations personnel for use in FDIR of the on-orbit station.

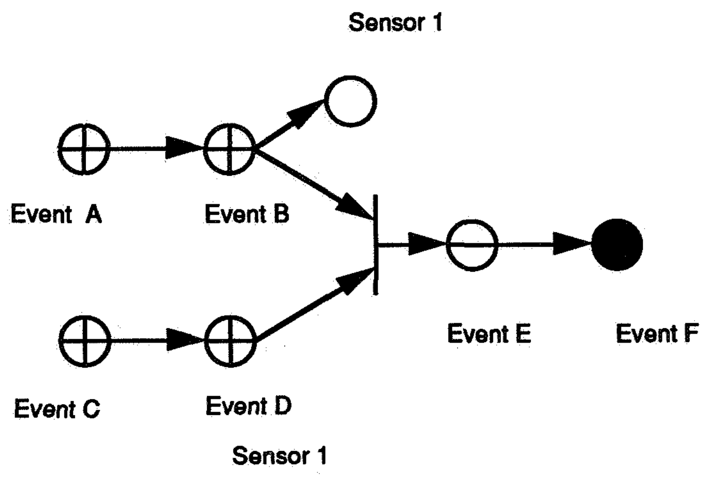
Space Shuttle FEAT is scheduled to fly on STS-52 as a Detailed Test Objective (DTO). A FEAT model of the S-band Communications System has been installed on an Apple™ Powerbook™, which will be flown aboard the shuttle. Astronauts will use the model to perform on-board fault isolation for the S-band Communication System. They will be able to configure the model to match the actual S-band system configuration, and then will use FEAT to identify possible causes of failures of the S-band system.

FUTURE APPLICATIONS OF DIGRAPHS

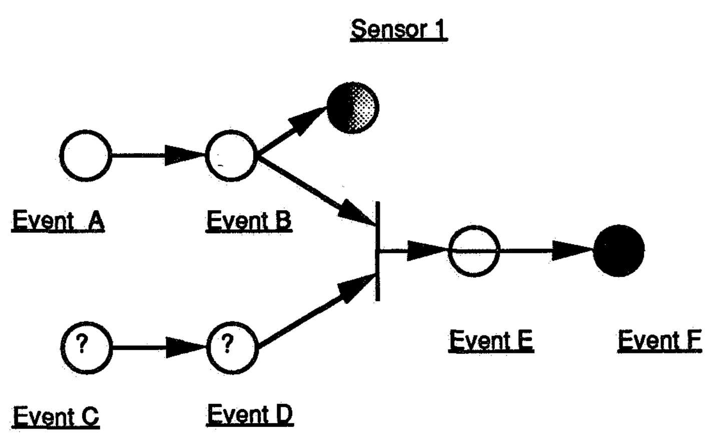
Digraphs are gaining acceptance, within the NASA community, as a viable method for conducting many kinds of analysis. Space Station Freedom Program and Operations, has mandated the use of digraph analysis for the Space Station Level II Integration effort; and many others are beginning to take up the banner. Some of the potential areas of application include the following:

Fault Isolation/Testability

FEAT's ability to model and analyze system failures make it a natural candidate for fault isolation efforts. If a failure event occurs, FEAT can display all of the possible single and paired causes for that event. However, in a large system, potential causes can be enormous in number. A method of pruning the list of possible causes is then necessary. Sensor information associated with the system can be used to remove candidate causes which occur downstream of a known nominal condition. Incorporation of sensor data, into the analysis, can help to reduce the number of candidate failures to a manageable sum. Then using traditional techniques, further isolation can be accomplished. Figure 8 shows an example of such a case.



Event F may be caused be Events E or F directly or by Events A or B paired with Events C or D through the AND gate.
Sensor 1's state is unknown.



If Sensor 1 reports nominal conditions, this implies that neither Event A or Event B occurred. Since neither Event A or B occurred, then the AND gate condition cannot be met, and the only way Event F can occur is if Events E or F occur. The number of events necessary to be checked to evaluate the cause of Event F has been reduced from 6 to 2.

Figure 8

Sensor data may also be combined with FEAT to identify the potential for cascading alarms. For instance, if a fault occurs downstream from a sensor, the sensors upstream will eventually alarm as a result of the fault. FEAT can show the effects of a fault on the downstream sensors.

This solution is being implemented by NASA, in an extension of FEAT, called Extended Real-time FEAT (ERF). ERF automatically prunes the list of possible faults, according to sensor information. ERF is being developed as a part of the FDIR system for the On-orbit Control Center Complex. Mission Controllers will use ERF to resolve off-nominal system behavior, by reducing the potential number of failure causes.

FEAT developers are pursing the possibility of incorporating, or interfacing with, a testability analysis tool which will help to evaluate sensor coverage in systems, and make recommendations regarding appropriate sensor locations. ERF is dependent upon adequate sensor information and proper placement of the sensors. Properly placed sensors provide information to quickly and accurately locate faults. The combination of FEAT, ERF and testability tools will make a very powerful fault isolation system.

Temporal Analysis

Not every event immediately affects the next downstream event. There may be appreciable delays within an event and between events. For example, an inappropriately shut valve, may not for some time, cause the pressure in the system to rise to an unacceptable level. In such a situation, time delay is an important aspect of calculating the potential failure space.

This issue will be addressed in FEAT when a modification is made to Digraph Editor to allow modelers to include time delays within events, and delays between events. FEAT will then compute the maximum and minimum time delay between selected events. This capability will be supplied in a future version of FEAT.

Software Modeling

Physical systems are not the only candidates for digraph analysis. Software functions and data flow can be modeled as well. Particularly, the flow and effect of invalid/improper data can be modeled. This can provide insight to the designer in determining mission critical software functions. Additionally, the effect of invalid data on other system functions (both software and hardware) may be shown. For instance, a software functional component that generates invalid data as an event; may then provide that data to other software and hardware as an invalid data input event. FEAT can be used to model these behaviors too.

Design Evaluation and Redundancy Management

Digraph models can be used to determine whether or not a system design provides sufficient redundancy. Maintenance and configuration effects on the system, can be evaluated by selectively removing (setting) components from the system. The reconfigured system can then be evaluated for induced single and paired events. This can be particularly useful in determining new vulnerabilities after a system has encountered failures and/or has portions of the system secured for maintenance.

FEAT contributes to design evaluation by rapidly displaying all single events caused by the event of interest, and all pairs of events that will result in that event. Unexpected single point common cause events are also quickly identified. As the design is modified to provide additional redundancy, the digraph model can be updated to reflect the changes, and the new set of single events and pairs of events can be evaluated.

Logistics Analysis

Logistics analysis addresses corrective and preventive maintenance tasks, and determines the kinds and numbers of repair parts needed for a system. This type of analysis is associated with the reliability and availability (reference 6), of systems. Reliability is defined as the measure of the mean time between failure (MTBF) and, concerns the probability that a system will operate over a specified period of time. No provision is made for repair when calculating reliability. Availability varies from reliability, in that it is a measure of the mean time to repair (MTTR), or, the probability that the system will operate over a period of time considering that something can be done to restore functionality lost as a result of a failure. How system repairs can be supported, or supportability, is important to determining availability. If repairs can be made instantaneously, availability is increased. However, long delays between failure and repairs makes the system proportionally less available.

FEAT models can help to identify critical components and the effect of their failure upon the system. Digraph models of the system can, along with specific part reliability, help to determine priorities for inventory stocks, and schedules for maintenance. Spare parts inventories are a major factor in determining supportability. For example, spares for parts that cause single point common cause events should have higher priority for stocking than parts that contribute to pairs of events.

Maintainability concerns the time it takes to remove and replace a component. Digraph models can identify components prone to low reliability, and single common cause failure. Designers can then either improve the reliability of the component or ensure that such items are accessible and easily replaced.

SUMMARY

As NASA continues to search for better and innovative approaches to new and old problems, directed graph analysis has emerged as a viable addition to the methods applied to Risk Assessment. Directed graphs are a well established area of mathematical study and analysis, and provide an easily comprehendible visual representation of cause and effect relationships. Conversion of the digraph to an equivalent matrix is straightforward, and allows analysis of digraphs to be mathematically calculated and verified. The nature of matrices also makes them ideally suited for computerized calculations, which in turn provides a vehicle for automating the task of risk assessment and failure analysis.

FEAT uses directed graph theory to provide engineers and analysts with a powerful and flexible automated analytic helper. FEAT can provide end to end analysis of cause and effect events. Very large systems can be modeled in modules, then connected to form the entire system. This feature also allows digraphs to be arranged in mix and match fashion. FEAT can detect and return information about single point failure vulnerability, failure event pairs, common cause events, and reduced capability analysis. FEAT shows the results of event propagation on system schematics and on the associated digraph. Digraph Editor provides a helpful way for the analyst to create digraphs.

The FEAT Project is funded by the NASA Space Station Freedom (SSF) Advanced Programs Development Office (Code MT) and the SSF Program Office (Code MS).

REFERENCES

1. L. Levy, *Discrete Structures of Computer Science*. John Wiley & Sons, 1980.
2. R. Stevenson, J. Miller and M. Austin. Failure Environment Analysis Tool (FEAT) Development Status. *AIAA Computing in Aerospace VIII, AIAA-91-3803*. October 1991.
3. I. Sacks. Digraph Matrix Analysis. *IEEE Transactions on Reliability, Vol. R-34, No. 5*. December 1985.
4. I. Sacks, G. Keller, and R. Rauch. Application of Digraph Matrix Analysis to the Space Station. *RDA , Logicon, R & D Associates, RDS-TR-148400-001*. September 1987.
5. J. Schier. Integrated Risk Assessment (IRA): Defining the Level II Safety & Reliability Job and Implementation Plan using Digraphs. *Unpublished Grumman presentation*. September, 1992.
6. B. Blanchard. *Logistics Engineering and Management*. 4th Edition. Prentis-Hall, Inc. Englewood Cliff, NJ. 1992.

BIBLIOGRAPHY

D. Haasl, N. Roberts, W. Vesely, F. Goldberg. *Fault Tree Handbook*. GPO Sales Program, U.S. Nuclear Regulatory Commission, Washington, DC. 1981.

J.Pearl, *Probabilistic Reasoning in Intelligent Systems*. Morgan and Kauffman, 1988.

Dual Use Space Technology Conference

NASA Scheduling Technologies

A compilation of information concerning
planning and scheduling tools designed by NASA

Jerry R. Adair
Intelligent Systems Branch
NASA/JSC

February 1994

Table of Contents

Abstract

This paper is a consolidated report on ten major planning and scheduling systems that have been developed by the National Aeronautics and Space Administration (NASA). A description of each system, its components, and how it could be potentially used in private industry is provided in this paper. The planning and scheduling technology represented by the systems ranges from activity based scheduling employing artificial intelligence (AI) techniques to constraint based, iterative repair scheduling. The space related application domains in which the systems have been deployed vary from Space Shuttle monitoring during launch countdown to long term Hubble Space Telescope (HST) scheduling. This paper also describes any correlation that may exist between the work done on different planning and scheduling systems. Finally, this paper documents the lessons learned from the work and research performed in planning and scheduling technology and describes the areas where future work will be conducted.

Introduction

NASA has performed an extensive amount of work in the area of planning and scheduling technology both in terms of research and real-world applications. This paper will present this work and discuss the goals, strategies, methods of implementation and results of each project on a case by case basis. Each project will be described with introductory remarks, planning and/or scheduling technology details of implementation, and closing remarks. In addition, the relation of a project to the private sector and how it could be used therein (i.e., its dual-use potential) will be outlined.

NASA Planning and Scheduling Projects

SPIKE

We begin with an activity-based scheduler called SPIKE that was designed to perform scheduling tasks for the Hubble Space Telescope (HST). It was developed at the Space Telescope Science Institute (STScI) in Baltimore, Maryland. SPIKE exploits AI techniques for constraint representation and for scheduling search. The aim of SPIKE is to allocate observations to timescales of days to a week, observing all scheduling constraints, and maximizing preferences that help ensure that observations are made at optimal times. SPIKE has been in use operationally for HST since shortly after the observatory was launched in April 1990. Although developed specifically for HST scheduling, SPIKE was carefully designed to provide a general framework for similar (activity-based) scheduling problems.

HST scheduling is a large scale problem: some 10,000 to 30,000 observations per year must be scheduled, each subject to a large number of operational and scientific constraints. The scheduling of the HST has been divided into two processes: the first is long-term scheduling, which allocates observations to week-long segments over a scheduling period of one year or more in duration.

This task is the responsibility of the SPIKE system. Individual weeks are then scheduled in more detail by the Science Planning and Scheduling System (SPSS), which orders observations within the week and generates a detailed command sequence for the HST control center at NASA Goddard Space Flight Center.

Scientists who wish to use the HST write observing programs to achieve their goals and these are sent to STScI in machine-readable form over national and international computer networks. They are then translated by an expert system called Transformation into a form suitable for scheduling within SPIKE. SPIKE treats the construction of a schedule as a constrained optimization problem and uses a heuristic repair-based scheduling search technique. The SPIKE algorithm has desirable "anytime" characteristics: at any point in the processing after the initial guess has been constructed, a feasible schedule can be produced simply by removing any remaining activities with constraint violations. The repair heuristics used by SPIKE are based on a very successful neural network architecture developed for SPIKE and later refined into a simple symbolic form which has made the original neural network obsolete. The adopted initial guess heuristic selects the most-constrained activities to assign first, where the number of minimum conflicts times is used as the measure of degree of constraint. If there remain gaps when all conflicting activities have been deleted (removed), then a simplistic best-first pass through the unscheduled activities is used to fill them.

SPIKE provides support for rescheduling operations in several ways. Two worth mentioning in particular are the task locking and conflict-cause analysis features. Tasks or sets of tasks can be locked into place on the schedule, and will thereafter not be considered during search or repair. These tasks represent the fixed points on the schedule. Conflict-cause analysis permits the user to force a task onto the schedule, and then display both what constraints have been violated and by which other tasks. The conflicting tasks can be unassigned if desired either individually or as a group and returned to the unscheduled tasks pool. The primary lesson learned from SPIKE development (which applies to similar systems) is to build in the expectation of change from the outset.

The implementation of SPIKE started in early 1987 and was initially based upon Texas Instruments Explorers as the hardware and software environment. The SPIKE graphical user interface was implemented in KEE Common Windows, but the remainder of the system, used only Common Lisp and the Flavors object system. At HST launch, STScI had a complement of 8 TI Explorers and microExplorers user for SPIKE operation, development and testing. In late 1991 SPIKE was moved from Explorers to Sun SparcStation IIs as the primary operations and development workstation. All of the Flavors code was automatically converted to the Common Lisp Object System. The Lisp used on the SparcStation is Allegro Common Lisp from Franz Inc. Allegro CL supports a version of CommonWindows based on X-windows, and so the user interface continues to operate on UNIX platforms as it did on the Explorers.

The dual use potential for SPIKE lies in large-scale scheduling problems that need to employ heuristics in the search algorithm. SPIKE can be modified to accommodate a new problem domain that might be found in private industry. For more information on SPIKE, contact the Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, Maryland, 21218. The contact is Mark D. Johnston.

Additional HST Scheduling Research

Next, we take a look at research work performed for HST scheduling by the Robotics Institute at Carnegie Mellon University which is related to the SPIKE system described in our previous project description. This work focuses on the short term scheduling of the HST and can best be described through an examination of its attributes: decomposability and scalability.

Existing and planned space-based observatories vary in structure and nature, from very complex and general purpose, like the HST, to small and targeted to a specific scientific program, like the Submillimeter Wave Astronomy Satellite (SWAS). However the fact that they share several classes of operating constraints (i.e., periodic loss of target visibility, limited onboard resources, etc.) does suggest the possibility of a common approach. The complexity of this problem stems from two sources. First, both types of observatories have the difficulty of the classical scheduling problems: optimization of objectives relating to overall system performance (e.g., maximizing return of science data), while satisfying all constraints imposed by the observation programs (e.g., precedence and temporal separation among observations) and by the limitations on the availability of capacity (e.g., observations requiring different targets cannot be executed simultaneously). Secondly, a safe mission operation requires the detailed description of all the transitions and intermediate states that support the achievement of observing goals and are consistent with an accurate description of the dynamics of the observatory; it is this aspect that constitutes a classical planning problem. Yet another characteristic of the problem is its large scale. To effectively deal with problems of this size, it is essential to employ problem and model decomposition techniques, which is where the research in this project was focused.

The problem in the HST short term scheduling domain is the efficient generation of schedules that account for the major telescope's operational constraints and domain optimization objectives. The basic assumption is to treat resource allocation, or scheduling, and auxiliary task expansion, or planning, as complementary aspects of a more general process of constructing behaviors of a dynamic system. The natural approach to solving the problem is then an iterative posting of constraints extracted either from the external goals or from the description of the system dynamics; consistency is tested through constraint propagation. In addition, the use of abstraction, model decomposability and incremental scaling, and exploiting opportunism to generate good solutions are techniques researched in this effort. These three are described in more detail in the following paragraph.

Briefly, where models are expressed in terms of the interacting state variables of different components of the physical system and its operating environment, an abstract model is one which summarizes system dynamics in terms of more aggregate structural components or selectively simplifies the represented system dynamics through omission of one or more component state variables. In HST, problems are naturally approached by decomposing them into smaller sub-problems separately and then assembling the sub-solutions. We can judge how the problem solving framework supports modularity and scalability by two criteria: 1) the degree by which heuristics dealing with each sub-problem need to be modified when adding sub-problem assembly heuristics to the problem solver, 2) the degree of increase of the computational effort needed to solve the problem versus the one needed to solve the component sub-problems. Finally, for overall sequence development, a simple dispatch-based strategy was used: simulating forward in time at the abstract level, the candidate observation estimated to incur the minimum amount of wait time (due to HST reconfiguration and target visibility constraints) was repeatedly selected and added to the current sequence. This heuristic strategy, termed "nearest neighbor with look-ahead" (NNLA), attends directly to the global objective of maximizing the time spent collecting science data. However, one critical tradeoff that must be made in space-based observatory scheduling is between maximizing the time spent collecting science data and satisfying absolute temporal constraints associated with specific user requests.

A second sequencing strategy (to NNLA) of comparable computational complexity that directly attends to the objective of minimizing rejection of absolutely constrained goals is "most temporally constrained first" (MCF). Under this scheme, the sequence is built by repeatedly selecting and inserting the candidate goal that currently has the tightest execution bounds.

Both NNLA and MCF manage combinatorics by making specific problem decomposition assumptions and localizing search according to these decomposition perspectives. The difference between these two comes in that NNLA assumes an event based decomposition while MCF assumes that the problem is decomposable by degree of temporal contrainedness.

The research in this effort is for work done in the HST scheduling project, but like the SPIKE effort, could be applied to similar systems (i.e., large scale) in the private sector. More information on this research work can be found through The Robotics Institute at Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, Pennsylvania, 15213. The contacts are Nicola Muscettola and Stephen F. Smith.

ERE

The next project is the work performed at NASA Ames Research Center in Moffett Field, California. It is related to previous work mentioned in that it is an integrated system and involves the planning, scheduling, and control for automatic telescopes. The project is called the Entropy Reduction Engine (ERE) project and is focusing on the construction of integrated planning and scheduling systems. This project serves as the focus for extending classical AI techniques involving automatic planners (i.e., systems that synthesize a plan to solve a problem) in a variety of ways. In the context of closed-loop plan execution, the ERE project is a focus for research as it relates to planning and scheduling. The eventual goal of the ERE project is a set of software tools for designing and deploying integrated planning and scheduling systems that are able to effectively control their environments. ERE has two important sub-goals: first the integration of planning and scheduling and second the study of plan execution as a problem of discrete event control.

The ERE project defines an integrated planning and scheduling system as a system that would able to consider alternative sets of actions, unlike the stand-alone scheduler, which is unable to deviate from its given action set. Moreover, the ERE project views plan execution as a problem in discrete event control; specifically, it formalizes a plan as a simple type of feedback controller, and this yields a new view on plan execution.

ERE itself is an architecture for producing systems that look ahead into the future, and by so doing, choose actions to perform. The essential idea is as follows: if a system has a limited amount of time to plan, and having planned, is allowed to plan no further, then it makes sense for the system to make the best use of the available time by incrementally improving its current plan until time runs out. The algorithm ERE employs is called traverse and robustify and follows this idea. ERE can provide a solid base for the development of integrated telescope planning, scheduling, and control systems that help to make this simplified management structure a reality.

ERE could be applied to private industry where a need for scheduling of automatic telescopes in an integrated planning and scheduling environment is required. Additional information on ERE can be obtained at NASA Ames Research Center, Mail Stop: 269-2, Moffett Field, California, 94035. The contact is Keith Swanson.

OMP

Next is the application and its underlying research which were performed by the Jet Propulsion Laboratory in Pasadena, California. It deals with a scheduling approach called Iterative Refinement and the application is called the Operations Mission Planner (OMP).

Iterative refinement is a heuristics-based approach to the scheduling of deep space missions which are modeled on the approach used by expert human schedulers in producing schedules for planetary encounters. Whenever the Voyager spacecraft encounters a planet, the science experiments to be conducted must be preplanned and ready to execute. This is a difficult scheduling problem due to the number and complexity of the experiments and the extremely limited resources of such a spacecraft. In general, not only are the schedules oversubscribed, they are also dynamic. As the scientists learn more about their objectives, the experiment requests themselves are updated.

Thus, the mission schedule is a dynamic entity. To solve the problem, OMP is centered on minimally disruptive replanning and the use of heuristics to limit the scheduler's search space.

In OMP, a resource tracks how a variable describing a state of the system changes through time and the steps which presently reserve this resource. There are five fundamental types of resources to OMP: capacity, consumable, renewable, continuous-state, and discrete-state. A brief description of each follows: a capacity resource is basically a pooled resource but can have non-integer value and may have a time varying initial capacity. A consumable resource is one for which there is a limited supply, and once it is used, it is no longer available (e.g., spacecraft fuel). A renewable resource is a generalization of a consumable resource, where the resource can be replenished (e.g., storage tape; it is used up during recording, and "replenished" during playback). A state resource represents a resource whose state (i.e., configuration, position, etc.) must be a certain value in order to support some activity. A continuous-state resource is a resource in which the state of the resource can best be described by a continuous variable (e.g., the direction that an antenna is pointing). A discrete-state resource however is represented by discrete values (e.g., on/off, low-gain/medium-gain/high-gain).

Associated with each of these types of resources is its definition of conflict. A conflict for a capacity resource occurs if the system reserves more than the limit of the pool at any moment in time. The resource is in conflict at the temporal interval for which a oversubscription occurs.

For additional definitions, OMP defines a step as a temporal interval which contains resource reservations. To OMP, an activity is a set of steps and a set of constraints that link the steps together. The temporal constraints are the "glue" that bind the steps into a logical unit. The user views an activity as the "primitive" action that must be scheduled to satisfy a user scheduling request.

OMP's iterative planning consists of a series of techniques. Each technique is responsible for a different aspect of the overall planning process. The first of these techniques roughs out the plan and identifies areas of high resource-conflict. The later techniques use the knowledge of the resource conflicts to refine the plan and solve many of the schedule problems. The final techniques try to solve the last of the conflicts and "optimize" the plan.

The OMP load phase is responsible for drafting an initial schedule. During the resource centered phase, OMP becomes resource oriented. The scheduler focuses upon a resource region which contains conflicts and uses quick and simple techniques to fix these regions before processing another resource. By focusing on just one resource region at a time the scheduler may fix one portion of the schedule but create additional conflicts in other regions. The scheduler discovers the bottlenecks by tracking these interactions between the separate regions. Once a bottleneck has been identified, it is classified and OMP attempts to resolve that bottleneck using techniques specialized for the type of bottleneck.

Once the conflict regions of the schedule have been resolved, OMP takes another look at the high priority activities which have been deleted from the schedule and tries to fit them in. At this point, OMP will perform its deepest search in an effort to schedule just one more activity. This phase is called the Optimization, although it doesn't produce a truly optimal schedule as would be defined in an operations research sense. Rather, it refers to fitting in additional activities after a conflict-free schedule has been produced. By specializing the planning techniques, each technique can be made more efficient.

The basic concept of self-reflective search in OMP is focusing the search by using knowledge gained from monitoring the search process. The OMP architecture, operating as outlined in the previous discussion, provides the mechanisms for supporting self-reflective search: the chronologies gather the raw information, the assessment heuristics analyze the information and feed the results to the control heuristics which focus the dispatch heuristics.

During the scheduling process, OMP keeps a chronology of the effort expended to resolve resource conflicts. During the resource centered phases, OMP focuses on a temporal interval within a given resource that is in conflict. Simple heuristics are used to reduce the level of conflict in the focus region. The chronologies keep track of the effect of these actions within the region and on other regions which are changed as a result of the scheduling actions. OMP first attempts to find a set of resource assignments which reduces the total amount of conflict in the entire schedule. If the system cannot lower the total conflict then it will increase the effort level for the focus region. This process will eventually cause OMP to cycle through the same regions.

The iterative planning approach to scheduling employed by OMP arose from attempts to heuristically control the search space of mission scheduling. The source of the heuristics were the human schedulers of Voyager, Viking, and SpaceLab who provided information on the stages of those scheduling processes.

Most of these heuristics in OMP assume that the scheduler knows which resources are the bottlenecks and which tasks are causing the most difficulty for the scheduler. The iterative planning approach assumes that the information gained by earlier techniques can be used by the later techniques to constrain the search space. Iterative planning also assumes that the schedule will not be changed dramatically by the later techniques.

OMP could be applied to the private sector in cases where an iterative approach utilizing heuristics for scheduling search is required to schedule in most any domain. The platform for OMP is a Mac II fx or Quadra. Additional information can be obtained by contacting Eric Biefeld at the Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, California, 91109-8099.

<u>APS</u>

Next is the Autonomous Power System (APS) project developed at the NASA Lewis Research Center. APS is designed to demonstrate the applications of integrated intelligent diagnosis, control and scheduling techniques to space power distribution systems. The APS project consists of three elements: the Autonomous Power Expert System (APEX) for Fault Diagnosis, Isolation, and Recovery (FDIR); the Autonomous Intelligent Power Scheduler (AIPS) to efficiently assign activities, start times and resources; and power hardware (Brassboard) to emulate a space-based power system. The AIPS portion of APS served as a learning tool and an initial scheduling testbed for the integration of FDIR and automated scheduling systems. Many lessons were learned from the AIPS scheduler and were integrated into a new scheduler called SCRAP (Scheduler for Continuous Resource Allocation and Planning). The APS project domain is the intelligent hardware and software of an electrical power system.

Similar to the scheduling of a telescope, the resources onboard a complex spacecraft will be vastly oversubscribed, having many times more resource requests than available resources. This makes it a paramount objective to efficiently utilize the available resources in order to complete as many activities as possible. The goal of the APS project is automated scheduling for space systems with proof-of-concept demonstrations on a power system testbed. In this scenario, the scheduler must not only know how to generate the schedule, but must also know how to implement the schedule, and how to recover from system or load induced deviations in the schedule.

The APS Brassboard is a power system testbed that contains a set of power supplies, switchgear, and loads that simulate a space-based power system. In order to more closely emulate this system, each load is given a set of attributes resembling those of the space-based system. Each activity (i.e., load) has a time varying profile of power demand, earliest start time and latest completion time constraints, a priority, and a temporal placement preference. In general, two modes of schedule generation are needed for any integrated scheduling system. The ability to generate an initial schedule and the ability to modify (i.e., reschedule) an already executing schedule in the case of an anomaly. The AIPS scheduler has two modes of schedule generation used for scheduling and rescheduling. The scheduling engine itself is an incremental scheduler that uses a set of activity selection and placement heuristics.

Finally, the SCRAP scheduling tool employs the theory of delineating two general categories in scheduling: predictive and reactive systems. Specifically, predictive scheduling allows the efficient allocation of available resources to activities by generating schedules based on predicted knowledge of the activity and resource states. On the contrary, reaction provides easier implementation in dynamic domains, but sacrifices resource usage efficiency caused by the lack of knowledge used to generate schedules. SCRAP also employs the theory that it may not be necessary to construct the initial schedule with a great level of detail. Therefore SCRAP schedules far term activities with less effort or detail than near term activities. It accomplishes this by using multiple levels of abstraction when scheduling activities. Further into the future the schedule is constructed abstractly, while nearer to the execution time more precision is used.

The APS project is an ongoing effort to demonstrate the use of knowledge-based diagnosis and scheduling software in advanced space-based electrical power systems. The SCRAP paradigm will allow for more efficient use of the available resources in such a system.

The APS project has dual use potential in electrical power system domains that can be modeled and which would allow for less detail in long term scheduling and more detail in the short term. Additional information can be obtained through Mark J. Ringer, Sverdrup Technology Inc., NASA Lewis Research Center Group, Cleveland, Ohio, 44135.

MAESTRO

Next is a scheduling and resource management system named MAESTRO developed by the Martin Marietta Astronautics Group in Denver Colorado. As its problem domain, MAESTRO was interfaced with a Space Station Module Power Management and Distribution (SSMPMAD) breadboard at the Marshall Space Flight Center (MSFC). The resulting combined system serves to illustrate the integration of planning, scheduling, and control in a realistic, complex domain.

Briefly, the functionality of MAESTRO is as follows: the Activity Editor is used to create definitions for activities which accomplish goals desired by the user. MAESTRO is used to select and schedule a subset of these activities, and to save the resultant schedule(s) out to files. The Transaction Manager (TM) serves as a communications port, facilitating specific types of communications between MAESTRO and the rest of the system during breadboard operation. The Front End Load Enable Scheduler (FELES) creates schedulers of power system events (such as closing switches) from saved schedule files. The Communications and Algorithmic Controller (CAC) distributes schedules among Load Centers (LCs), into which are incorporated Lowest Level Processors (LLPs). The Fault Recovery And Management Expert System (FRAMES) performs fault isolation, diagnosis and recovery for the power scheduler during real-time contingencies.

During normal operations, a user will interact with the activity editor to create a set of activities to be scheduled, saving these activities' definitions in an activity library. In that or another session, the user will run the scheduler to create one or more initial schedules of these activities. These schedules will be saved into a schedule library.

When a power system anomaly occurs, MAESTRO will get a set of information from FRAMES through the TM. MAESTRO follows a three-step process to handle these messages and revise the schedule. It a) modifies the schedule to reflect changes made to it by the power system and to remove resource and temporal constraint violations for activities not yet begun, b) tries to find ways to create and schedule continuations for interrupted activities, and c) tries to schedule any activities that can take advantage of the resources released by the interruption of others.

Insofar as planning is concerned, activity continuation is the single automated planning function within MAESTRO. However, in terms of scheduling, MAESTRO can represent temporal constraints between activities, sometimes necessitating the consideration of more than one continuation model at once.

MAESTRO is a tool that could be applied to a need for a system executive that could monitor and control individual subsystems of a commercial rocket launch. Additional information on MAESTRO can be obtained from Daniel L. Britt and Amy L. Geoffroy at the Martin Marietta Astronautics Group, P. O. Box 179, Mail Stop: XL4370, Denver, Colorado, 80201.

GERRY

Next is a well-known scheduling tool called GERRY that was also developed at NASA Ames Research Center. GERRY is a general scheduling system applied to the Space Shuttle ground processing problem at Kennedy Space Center (KSC). It employs the approach called constraint-based iterative repair that provides the ability to satisfy domain constraints, address optimization concerns, minimize perturbation to the original schedule, and produce modified schedules quickly; all of which rescheduling systems should be capable of in general. GERRY utilizes a novel approach to rescheduling that addresses the concerns and gives the user the ability to individually modify each criteria's relative importance.

In terms of problems specific to GERRY's domain, there is an additional complicating factor introduced by the Space Shuttle problem and that is preemption. In preemptive scheduling, each task is associated with a calendar of legal work periods that determine when the task must be performed. Preemption effectively splits a task into a set of subtasks. It requires additional computational overhead since for each task the preemption times must be computed and appropriate constraint manipulation for each time assignment must be performed.

In the Space Shuttle processing domain, rescheduling is necessitated by changes that occur in the environment. In response, GERRY adopts the approach of repairing the constraints that are violated in the schedule. Constraint-based iterative repair itself begins with a complete schedule of unacceptable quality and iteratively modifies it until its quality is found satisfactory. In GERRY repairs are associated with constraints. Repairing any violation typically involves moving a set of tasks to different times: at least one task participating in the constraint violation is moved, along with any other tasks whose temporal constraints would be violated by the move. Simply put, all temporal constraints are preserved after the repair. Similar to many scheduling evaluation tools, at the end of each iteration, the system re-evaluates a cost function to determine whether the new schedule resulting from the repairs is better than the current solution. The system sometimes accepts a new solution that is worse than the current solution in order to escape local minima and cycles. In summary, the algorithm is interruptable, restartable, and outputs a solution when terminated.

To examine the effects of this scheduling strategy, the STS-43 Space Shuttle processing flow was used as an experiment. The results were very positive and the experiments suggest that the constraint framework and the knowledge encoded in this framework is an effective search tool that allows one to adjust the importance of schedule perturbation and other objective criteria.

GERRY could be adapted to an environment in the private sector which employed assembly line processing for construction of a product. For more information concerning GERRY, contact Eugene Davis, Brian Daun, or Michael Deale at the NASA Ames Research Center, Mail Stop: 269-2, Moffett Field, California, 94035.

TMSA

Next is a concept prototype known as Time Management Situation Assessment (TMSA) developed by the Advanced Computing Technologies Group at Boeing at the Kennedy Space Center in Florida. TMSA was designed to support NASA Test Directors (NTDs) in schedule execution monitoring during the later stages of a Shuttle countdown. The system detects qualitative and quantitative constraint violations in near real-time.

The problem domain involves the NTDs primary concern with the orderly and timely execution of the countdown process. The cognitive model they reason with is relatively high-level and includes a nominal (i.e., planned) model of the countdown and a set of qualitative and quantitative constraints that define such a countdown by specifying temporal duration and ordinal relationships between countdown events. The NTDs monitor the current countdown and assess its compliance with their nominal countdown model. The countdown schedule may be revised by reordering events and/or adjusting the durations of intervals between events. The characteristics of the domain are as follows:

A) The situation is highly structured with a well formulated, proven set of constraints on the schedule, and

B) Although this is an advisory system used by experts, the criticality of the situation places a premium on timeliness and correctness beyond that of many applications.

The verification and validation issues in TMSA's software environment in conjunction with the above mentioned characteristics led to an approach of the problem algorithmically and avoided using heuristics.

Furthermore, while the countdown is formulated in terms of both events and intervals, the constraints between intervals are such that intervals were represented as start and end pairs of events. From the NTD's perspective countdown time is discrete. Pseudo events are used and are defined as events that are not members of the universe of countdown events employed by the NTDs. Uncertainty arises in the countdown schedule situation in that many of the qualitative constraints between countdown events are ambiguous. In addition, ambiguity also occurs in some quantitative duration constraints on the length of intervals.

To solve this problem, two algorithms were developed for TMSA. These form the reasoning kernel of the system and are designed to monitor and interpret the legality of the temporal duration and sequential unfolding of a countdown. The first algorithm, called the ConstraintChecker, is used to maintain a qualitative representation of the current status of a countdown and to check the consistency of that status with the qualitative constraints that define the legality of a countdown. The second algorithm, known as the ScheduleMaintainer, is used to maintain both a qualitative and quantitative representation of a countdown. This representation includes both the current status of the countdown and the quantitative constraints that define the legality of a countdown. The representation is also used to generate relational assertion vectors as input to the consistency checking algorithm. Together, the ConstraintChecker and the ScheduleMaintainer form the core of TMSA and exchange relations between the ordering of events.

The TMSA prototype is implemented in Smalltalk and runs on a 25 mhz 486, under MS DOS. TMSA could be applied to a private industry domain with similar constraints and events found in that of a Space Shuttle countdown, like a launch of a commercial rocket perhaps. It could also be utilized in an environment where testing of a device of some sort underwent a well-defined procedure that employed a set of rules for its own execution. An example of this might be the process testing of an airplane in a wind tunnel. For further information concerning TMSA, contact Michael B. Richardson of the Advanced Computing Technologies Group, Boeing Aerospace Operations, FA-71, Kennedy Space Center, Florida, 32899.

COMPASS / ISAID

Finally, there is an interactive and highly flexible scheduling system called COMPASS developed by the McDonnell Douglas Space Systems Corporation in Houston, Texas. COMPASS is a constraint-based scheduler and with its interactive trait, it provides an environment where a mixed initiative is possible; that is, it lets the computer do what it does best (i.e., check constraints and calculate feasible intervals) and lets the human do what he/she does best (i.e., provide heuristic and subjective inputs into the schedule). This results in a cooperative production of a schedule which reflects both the hard constraints and subjective preferences. Written in ADA with the user interface in C and X-Windows, COMPASS is a well-designed, highly flexible scheduler that could be modified to fit the needs of a wide variety of scheduling problems. It operates on a number of platforms, including Sun3/4, Sun Sparc, Rational, RS6000, and VAX/VMS. COMPASS provides both forward and backward scheduling modes and displays a schedule in the Gantt chart format. In addition, a histogram of resource usage can be displayed. An example of COMPASS's modification ease follows.

COMPASS was used as the scheduling engine in a highly successfully project known as the Interactive Scheduling AID (ISAID). The target environment for this tool was the Systems Engineering Simulator (SES) located at NASA Johnson Space Center (JSC). The SES is a facility which houses the software and hardware for a variety of simulation systems such as the Manned Maneuvering Unit and the Remote Manipulator System. Due to the highly flexible and well thought out design of COMPASS, it was modified to meet the requirements of the SES. This entailed work in three discrete areas:

A) A new user interface. The SES scheduling experts had grown accustomed to performing scheduling duties with paper "worksheets" which represented a schedule in development in a calendar-type format. The default interface for COMPASS represented activities in a Gantt chart format (H bars) which was unacceptable to the SES users. Therefore, a completely new interface which appears like a calendar (very much like the previous method) was developed.

B) Implementation of domain-specific SES constraints. These constraints (or "rules") were incorporated into the scheduling engine of COMPASS to emulate the procedure the SES scheduling experts used when developing a schedule.

C) Optimization. This feature had not been implemented into COMPASS prior to the ISAID project. The addition of this feature gave COMPASS the ability to allow the user to interleave schedule creation, revision, and optimization. The optimization procedure itself was based on the simulated annealing of metal. It begins with high temperatures and allows many changes to the schedule to occur, evaluates the result according to an objective function, compares that to the best schedule generated so far, and decrements the temperature as time passes. At low temperatures, very little change is allowed to occur to the schedule.

With these three enhancements to COMPASS, the ISAID tool found much success and is still used on a weekly basis to generate SES schedules.

COMPASS could be applied to many scheduling domains in the private sector due to its highly flexible design and ease of modification. For more information about COMPASS, contact Barry Fox of the Planning and Scheduling Technology Group at the McDonnell Douglas Space Systems Company, 16055 Space Center Boulevard, Houston, Texas, 77062.

Conclusion

This paper has presented a number of scheduling applications and research areas, all of which are capable of being applied to the private sector. They were designed with the capability of being modified to meet the requirements of an application domain other than the one for which they were developed.

Furthermore, the projects described in this paper use various planning and scheduling strategies to meet particular goals and therefore one of them will fulfill a number of scheduling needs in the private community. A contact person or persons was (were) listed at the end of each project description for your convenience.

References

All technical information contained in this paper was referenced through:

NASA Ames Research Center, Artificial Intelligence Research Branch. *Technical Report FIA-92-17: Working Notes form the 1992 AAAI Spring Symposium on Practical Approaches to Scheduling and Planning.* NASA Ames Research Center, 1992.

CONFIG - INTEGRATED ENGINEERING OF SYSTEMS AND OPERATIONS

ABSTRACT. CONFIG 3 is a prototype software tool that supports integrated conceptual design evaluation from early in the product life cycle, by supporting isolated or integrated modeling, simulation, and analysis of the function, structure, behavior, failures and operation of system designs. Integration and reuse of models is supported in an object-oriented environment providing capabilities for graph analysis and discrete event simulation. Integration is supported among diverse modeling approaches (component view, configuration or flow path view, and procedure view) and diverse simulation and analysis approaches. Support is provided for integrated engineering in diverse design domains, including mechanical and electro-mechanical systems, distributed computer systems, and chemical processing systems.

INTRODUCTION. The core of engineering design and evaluation focuses on analysis of physical design. Thus, today's computer-aided engineering software packages often do not provide enough support for conceptual design early in the life cycle or for engineering for operation, fault management, or supportability (reliability and maintainability). Benefits of engineering for operations and supportability include more robust systems that meet customer needs better and that are easier to operate, maintain and repair. Benefits of concurrent engineering include reduced costs and shortened time for system development. Integrated modeling and analysis of system function, structure, behavior, failures and operation is needed, early in the life cycle.

Conventional system modeling approaches are not well suited for evaluating conceptual designs early in the system life cycle. These modeling approaches require more knowledge of geometric or performance parameters than is usually available early in design. More abstracted models can support early conceptual design definition and evaluation, and also remain useful for some later analyses. Component-connection models provide one such useful abstraction, and discrete events are another. Discrete event simulation technology combines both abstractions, for evaluation of conceptual designs of equipment configurations in operations research (3). CONFIG uses these abstractions, with some enhancements, to define and evaluate conceptual designs for several types of systems.

The initial CONFIG project goal was to support simulation studies for design of automated diagnostic software for new life-support systems (9). The problem was to design an "expert system" on-line troubleshooter before there was an expert. The design engineer could use a model of the system to support what-if analyses of failure propagation, interaction, observability and testability. This activity is similar to Failure Modes and Effects Analysis (5), but uses comparative simulations of failure effects to develop diagnostic software. Conventional simulation software was not up to this challenge, but discrete event simulation software has been. CONFIG supports the use of qualitative models for applying discrete event simulation to continuous systems.

A major design goal for CONFIG is to support conceptual design for operations and safety engineering. Major tasks in conceptual design are design definition, evaluation (by simulation and analysis) and documentation. Operations engineering focuses on the design of systems and procedures for operating, controlling and managing the system in normal or faulty conditions. Safety engineering focuses on prevention of hazardous effects and conditions in the physical system or its operation. In these types of engineering, complex interactions and interfaces among system components and operations must be a focus.

Another design goal of CONFIG is to bridge the gaps between physical design engineering and other types of engineering. Component-connection representations are well suited for modeling and defining physical system designs (as structures of interacting components) and operations designs (as structures of interacting actions), as well as interactions between system components and operational actions. Discrete event models have been used for this type of modeling for queueing and scheduling problems, but can be extended to support conceptual modeling in operations and safety engineering. This type of modeling is also compatible with systems engineering function diagrams (1).

CONFIG 3. The project approach has been to incrementally integrate advanced modeling and analysis technology with more conventional technology. The prototype integrates qualitative modeling, discrete event simulation and directed graph analysis technologies for use in analyzing normal and faulty behaviors of dynamic systems and their operations. The prototype has been designed for modularity, portability and extensibility. A generic directed graph element design has been used to standardize model element designs and to promote extensibility. This directed graph framework supports integration of levels of modeling abstraction and integration of alternative types of model elements.

Enhanced Discrete Event Simulation Capabilities. In traditional discrete event modeling and simulation, state changes in a system's entities, "events", occur discretely rather than continuously, at nonuniform intervals of time. Throughout simulation, new events are added to an event list that contains records of events and the times they are scheduled to occur. Simulation processing jumps from one event to the next, rather than occurring at a regular time interval. Computation that results in creation of new events is localized in components, which are connected in a network. Any simulation run produces a particular history of the states of the system. Statistical simulation experiments, using random variables in repeated simulation runs, are used to compare design alternatives.

To enhance this discrete event simulation approach to accommodate continuous systems, a number of new concepts and methods were developed. These include component models with operating modes, types of links connecting components ("relations" and "variable clusters"), state transition structures ("processes"), methods for representing qualitative and quantitative functions ("process language"), and a new simulation control approach.

Digraph Analysis Capabilities. The CONFIG Digraph Analyzer (DGA) makes graph analysis techniques available for evaluating conceptual designs of systems and their operations. The DGA is based on reachability search, and is implemented generically for application to the many types of graph data structures in CONFIG. DGA can support analyses of completeness, consistency and modularity. Analysis of failure sources and impacts can be done by tracing the paths from a given failure.

System Modeling. Devices are the basic components of a CONFIG system model, which are connected together in topological model structures with Relations. Device behavior is defined in operating and failure Modes, which contain mode dependent and mode transition Processes. Modes are connected together in a mode transition digraph which delineates the transition dependencies among the individual modes.

Device Processes define change events in device variables, and are conditionally invoked and executed with appropriate delays during a simulation. Processes define time-related behavioral effects of changes to device input variables, both direction of change and the new discrete value that will be reached, possibly after a delay. Faults and failures can be modeled in two distinctly different ways. Failure modes can be used to model device faults. Mode-transition processes can be used to model latent device failures that cause unintended mode changes. Relations connect devices via their variables, so that state changes can propagate along these relations during simulations. Related variables are organized into variable clusters, to separate types of relations by domain (e.g., electrical vs. fluid connections). Relations can also connect Devices with device-controlling Activities in operations models.

Flow Path Modeling. Flow is a property of many systems, whether the substance flowing is a liquid or information. There are two difficulties in modeling flows with local device processes. First, flow is a global property of the topology of the modeled system and the substances flowing within it. Second, while dynamic changes in system structure and flow can occur during operations, process descriptions involving flow must often rely on assumptions of static system topology. These factors would limit the reusability of device descriptions to a limited set of system structures.

A flow-path management module (FPMM) has been implemented to address these problems. The FPMM is separate from the module implementing local device behavior, but the two modules are interfaced via flow-related state variables in the devices. When FPMM is notified during simulation of a local change in device state, it recomputes the global effects on flows produced by the local state change. The FPMM then updates the state of flow in all affected devices. This design permits the user to write reusable local device process descriptions that do not depend on any assumptions concerning the system topology.

FPMM uses a simplified representation of the system, as a collection of aggregate objects, or "circuits." Further abstraction is achieved by identifying serial and parallel clusters in the circuit (6). In many cases, configuration determination alone is sufficient to verify flow/effort path designs, to establish flow paths for a continuous simulation, for reconfiguration planning, and for troubleshooting analysis (see Ref. [2] on cluster-based design of procedures for diagnosis, test, repair and work-around in a system).

Operations modeling. Activities are the basic components of a CONFIG operations model, and are connected together in action structures with Relations. They represent procedures or protocols that interact with the system, to control and use it to achieve goals or functions. Each activity model can include specifications for what it is intended to achieve or maintain. Activity behavior is controlled in a sequence of phases, ending in an evaluation of results. Activity behavior is defined by processes that model direct effects of actions, or that control device operation and mode transitions to achieve activity goals. Relations define sequencing and control between activities and connect Devices with device-controlling Activities.

Operations models are designed to support operation analysis with procedure models. These models are designed to support analysis of plans and procedures for nominal or off-nominal operation. The procedure modeling elements are designed for reuse by intelligent replanning software, and for compatibility with functional modeling in systems engineering.

Model Development & Integration Capabilities and Approach. CONFIG provides intelligent automation to support nonprogrammer and nonspecialist use and understanding. CONFIG embeds object-oriented model libraries in an easy-to-use toolkit with interactive graphics and automatic programming.

CONFIG provides extensive support for three separable yet tightly integrated phases of user operation during a modeling session: Library Design, Model Building, and Simulation and Analysis. This includes a graphical user interface for automated support of modeling during each of the phases including the development of object-oriented library element classes or templates, the construction of models from these library items, model inspection and verification, and running simulations and analyses.

The integration between the phases enables an incremental approach to the modeling process. Lessons learned from analyses can be repeatedly and rapidly incorporated by the user into an initially simple model. Support for these phases as separate user activities fosters the achievement of concurrent engineering goals. Different users can define library elements, build models, and analyze models at different times depending on area of expertise and availability of resources. Support for the model building phases spans all types of modeling that can be performed in CONFIG including component structure, behavior and flow, and activity goals and structure.

Hosting. CONFIG is implemented in software that is portable to most Unix work stations. The Common LISP Object System (CLOS) is a highly standardized language, with compilers for most of the commonly available work stations. The user interface was implemented using the Common LISP Interface Manager (CLIM), another standardized tool built on CLOS.

CONCLUSIONS. CONFIG is designed to model many types of systems in which discrete and continuous processes occur. The CONFIG 2 prototype was used to model and analyze: 1) a simple two-phase thermal control system based on a Space Station prototype thermal bus, 2) a reconfigurable computer network with alternate communications protocols, and 3) Space Shuttle Remote Manipulator System latching and deployment subsystems (7). The core ideas of CONFIG have been patented (8). CONFIG 3 has added capabilities for graph analysis and for modeling operations and procedures.

The CONFIG prototype demonstrates advanced integrated modeling, simulation and analysis to support integrated and coordinated engineering. CONFIG supports qualitative and symbolic modeling, for early conceptual design. System models are component structure models with operating modes, with embedded time-related behavior models. CONFIG supports failure modeling and modeling of state or configuration changes that result in dynamic changes in dependencies among components. Operations and procedure models are activity structure models that interact with system models. The models support simulation and analysis both of monitoring and diagnosis systems and of operation itself. CONFIG is designed to support evaluation of system operability, diagnosability and fault tolerance, and analysis of the development of system effects of problems over time, including faults, failures, and procedural or environmental difficulties.

REFERENCES

1. Alford, M. Strengthening the System Engineering Process, Engineering Management Journal, Vol. 4, No. 1, March, 1992 , pp 7-14.

2. Farley, A. M. Cluster-based Representation of Hydraulic Systems, Proc. 4th Conference on AI Applications, March, 1988, pp. 358-364.

3. Fishman, G. S. Principles of Discrete Event Simulation. New York, NY: Wiley, 1978.

4. Forbus, K. Qualitative Physics: Past, Present, and Future. In Exploring Artificial Intelligence (H. Shrobe and AAAI, eds.). San Mateo, CA: Morgan Kaufmann, 1988.

5. Fullwood, R. R. and Hall, R. E. Probabilistic Risk Assessment in the Nuclear Power Industry: Fundamentals and Applications. Pergamon Press, 1988.

6. Liu, Z. and Farley, A. M. "Structural Aggregation in Common-Sense Reasoning". Proc. 9th National Conference on Artificial Intelligence (AAAI-91), 1991, pp. 868-873.

7. Malin, J. T., B. D. Basham and R. A. Harris, "Use of Qualitative Models in Discrete Event Simulation for Analysis of Malfunctions in Continuous Processing Systems." Artificial Intelligence in Process Engineering (M. Mavrovouniotis, ed.), Academic Press, pp. 37-79, 1990.

8. Malin et al., U. S. Patent 4,965,743, "Discrete Event Simulation Tool for Analysis of Qualitative Models of Continuous Processing Systems" October, 1990.

9. Malin, J. T., and Lance, N. "Processes in construction of failure management expert systems from device design information". IEEE Trans. on Systems, Man, and Cybernetics, 1987, SMC-17, 956-967.

10. Malin, J. T. and Leifker, D. B. "Functional Modeling with Goal-Oriented Activities for Analysis of Effects of Failures on Functions and Operations". Informatics and Telematics, 1991, 8(4), pp 353-364.

ACKNOWLEDGEMENTS. The authors wish to thank Bryan Basham, Leslie Ambrose, Ralph Krog, Debra Schreckenghost, Brian Cox, Daniel Leifker and Sherry Land for their contributions to CONFIG design and implementation.

Title: CONFIG - INTEGRATED ENGINEERING OF SYSTEMS AND OPERATIONS
Funded By: ETB
Principal Investigator: Jane T. Malin, ER2
Task Performed By: JSC, MITRE, LESC
Contact: Jane T. Malin
713 483-2046, FAX (713) 483-3204, Mail Code ER2
Alternate Contact: Dan Ryan, MITRE

Om/T

Session R3: ROBOTICS TECHNOLOGIES

Session Chair: Reginald Berka

LYNDON B. JOHNSON SPACE CENTER (JSC) PROPOSED DUAL-USE TECHNOLOGY INVESTMENT PROGRAM IN INTELLIGENT ROBOTICS

S12-63
44796

Jon D. Erickson
National Aeronautics and Space Administration Lyndon B. Johnson Space Center
Houston, Texas 77058

Abstract

This paper presents an overview of the proposed Lyndon B. Johnson Space Center (JSC) precompetitive, dual-use technology investment project in robotics. New robotic technology in advanced robots, which can recognize and respond to their environments and to spoken human supervision so as to perform a variety of combined mobility and manipulation tasks in various sectors, is an objective of this work. In the U.S. economy, such robots offer the benefits of improved global competitiveness in a critical industrial sector; improved productivity by the end users of these robots; a growing robotics industry that produces jobs and profits; lower cost health care delivery with quality improvements; and, as these "intelligent" robots become acceptable throughout society, an increase in the standard of living for everyone. In space, such robots will provide improved safety, reliability, and productivity as Space Station evolves, and will enable human space exploration (by human/robot teams).

The proposed effort consists of partnerships between manufacturers, universities, and JSC to develop working production prototypes of these robots by leveraging current development by both sides. Currently targeted applications are in the manufacturing, health care, services, and construction sectors of the U.S. economy and in the inspection, servicing, maintenance, and repair aspects of space exploration. But the focus is on the generic software architecture and standardized interfaces for custom modules tailored for the various applications allowing end users to customize a robot as PC users customize PC's. Production prototypes would be completed in 5 years under this proposal.

1. Introduction

This paper suggests a large number of opportunities for robotic manufacturers, integrators, potential buyers/users of robots, commercial technology developers, and universities to work with the NASA JSC Automation and Robotics Division, with NASA funding a major portion of the development. The focus is intelligent robotics as partial solutions to productivity problems in several sectors of application. The stage of development addressed is precommercial. In each case dual use is a prerequisite: there must be a space use as well as a nonspace, commercial use. Generally, this is easily the case.

The specific motivation and rationale for this NASA JSC proposed technology investment program is detailed in Erickson [1]. The general policy that sets the context for the NASA technology investment program, which will begin in 1994, is given in Clinton and Gore [2].

It is important to understand that although a set of objectives, an approach, and a number of tasks are suggested here, these are meant to stimulate the creative thought process of those in nonaerospace and aerospace industry to propose objectives, approaches, and tasks that they believe, due to their involvement with their commercial buyers/users, would be economic and profitable as a result of jointly funded developmental efforts with NASA.

Intelligent robotics is the use of robotic systems in solving problems in tasks and environments where the robot's ability to acquire and apply knowledge and skills to achieve stated goals in the face of variations, difficulties, and complexities imposed by a dynamic environment having significant unpredictability is crucial to success. This means the robots can recognize and respond to their environments at the pace of their environments and to spoken human supervision so as to perform a variety of mobility and manipulation tasks. This does not require broad-based general intelligence or common sense by the robot.

These robots are capable of significant, autonomous reaction to unpredictable events, yet are subject to optional human supervision during operation in a natural way, such as by voice. We refer to this capability in the supervised robot as "adjustable autonomy." Also, a key essence is that previously acquired knowledge is combined with knowledge acquired at the instant of task performance.

The overall approach can be summarized as capitalizing on a software architecture that can be viewed as generic and modular, and hardware approaches that are modular, reconfigurable, and extendible. Many of the software modules, such as a deliberative planner, world model, and natural language interface, can also be viewed as generic. Other software is bundled with certain hardware; e.g., sensing software is bundled with specific sensor hardware. This leads to the concept of a modular, end-user customized robot, put together from modules with standard interfaces [3, 4, 5] such as users do with a personal computer. An integrated computer aided concurrent engineering environment that we are working on [6] is a way to achieve close teamwork by geographically distributed "virtual" teams to develop the production prototypes.

JSC can be a key partner in this dual-use technology investment program in intelligent robotics for two reasons: (1) human space exploration missions require supervised intelligent robotics as enabling tools [7, 8] and, hence, must develop or have developed supervised intelligent robotic systems and (2) intelligent robotic technology is being developed for space applications at JSC (but has a strong crosscutting or generic flavor) that is advancing the state of the art [9, 10] and is producing both skilled personnel and adaptable developmental infrastructure, such as low cost simulation environments for software testing and integrated testbeds for complete prototype testing. JSC also has a Small Business Innovative Research (SBIR) program [11] for intelligent robotics, which is underutilized and has no commercial cost sharing requirement. It is limited in scope to about $0.6 million and 2 years in Phase II efforts.

A key element in the cutting edge intelligent robotics technology work at JSC is an understanding of and solution approach to the key issue of melding artificial intelligence planners with reactive capabilities. Artificial intelligence planners offer goal-achieving planning, but also high-time variance due to searching. Reactive capabilities are needed to deal safely in real time with dynamic, unpredictable environments at the pace of the dynamics [9]. A second key element that JSC brings is an approach to improved robotic reliability as required for space, but also useful in industry. A third key element that JSC brings to cutting edge technology is an understanding of and solution approach to the key issue of robotic safety while maintaining productivity.

Of all these elements, the personnel skilled in the state of the art and knowledgeable about the technology are the most important.

2. Overview of Proposed Activities

New robotic technology in advanced robots that can recognize and respond to their environments and to spoken human supervision so as to perform a variety of mobility and manipulation tasks in various sectors is an objective of this proposed effort. In the U.S. economy, such robots offer the benefits of improved global competitiveness in a critical industrial sector; improved productivity by the end users of these robots; a growing robotics industry that produces jobs and profits; lower cost health care delivery with quality improvements; and, as these "intelligent" robots become acceptable throughout society, an increase in the standard of living for everyone [12]. In space, such robots will provide improved safety, reliability, and productivity as Space Station evolves, and will enable human space exploration (by human/robot teams).

The proposed effort consists of partnerships between manufacturers, users, universities, and JSC to develop working production prototypes of these robots by leveraging current development by manufacturers and JSC. Currently targeted applications are in the manufacturing, health care, services, and construction sectors of the U.S. economy and in the inspection, servicing, maintenance, and repair aspects of space exploration. But the focus is on the generic software architecture and standardized interfaces for custom modules tailored for the various applications, allowing end users to customize a robot as personal computer users customize PC's. Production prototypes would be completed in 5 years under this proposal, as would automated developmental environments and integrated testbeds.

JSC possesses the required core skills in its civil service and contractors to form the nucleus of the multiple partnerships. Current technology integration efforts at JSC include the EVA helper/retriever supervised intelligent robot [10], the mobile robotics testbed project, and the Soda-Pup entry in the AAAI national robotics competition (1992 award winner). In addition, JSC is responsible for engineering upgrades to the Shuttle Remote Manipulator Systems, integration of the Mobile Servicing Systems and Special Purpose Dexterous Manipulator into Space Station, and numerous robotics technology efforts.

User coordination involves interested manufacturers with deployed robots. Joint facility sharing and temporary personnel exchange are possible.

The overall set of activities has been grouped into the following seven related categories of tasks, each with its own objectives and approach.

1. Problem-Solving Insertion of Robot Intelligence Technology
2. Generic Intelligent Robotics Software Architecture
3. Modular Manipulation and Mobility for Robotics
4. Integrated Sensing and Perception Capabilities for Robotics
5. Robotic Surrogates for Human Grasping and Manipulation
6. Integrated Prototyping Environment for Robotics
7. Robotic Applications in Advanced Manufacturing, Health Care, Service Industries, and Construction

The following sections present the objectives, approach, and benefits for each of these categories of tasks and give the titles of the set of tasks grouped into that category. One-page task descriptions are available [13] for all tasks, giving task objectives, proposed effort, major milestones, benefits, and other information.

Problem-Solving Insertion of Robot Intelligence Technology

The objectives are (1) to work as a team with end user industries whose productivity problems can be solved by integrating adaptive robots into the advanced manufacturing or service process, and in so doing to develop a new paradigm of product line development for robot manufacturers and (2) to provide the robotics industry sensor/software control techniques that will make the robots more flexible and attractive for use by end user industries. This will impact the end users of these robots by improving the end users' efficiency and productivity and thus improved global competitiveness. This will also stimulate robot demand and provide a new way of doing business for robot manufacturers. The benefits for space will be a healthier robotics industry capable of supplying quality robotics at lower costs.

The proposed effort consists of partnerships between manufacturers, integrators, nonprofits, and JSC to solve end user problems by integrating adaptive robots into end user operations. As part of that effort the robot manufacturers' products must first be upgraded with sensing and intelligent reaction capabilities from new sensor/software control technology. A key product is the development, documentation, and refinement of the problem-solving insertion process for intelligent robotics technology, including end user problem identification techniques; problem selection criteria; requirements definition; development of a solution; integration with the end user people, processes, and equipment; user training; and continuing user support.

In related work, JSC is responsible for integration of the Mobile Servicing System and Special Purpose Dexterous Manipulator into Space Station, which gives us the necessary experience and insight to help users.

The eight tasks in this category are the following:

- End User Target Problems Identification
- End User Problem Selection
- Selected Problem Requirements Definition
- Design, Development, Test, and Evaluation of Solution
- Integration with User Equipment, Processes, and People
- User Training Definition
- Continuing User Support Definition
- Problem Solving Insertion Process Development, Documentation, and Refinement

The benefits of this problem-solving insertion are that the end user businesses obtain a useful solution to their problems The robot manufacturers and integrators obtain a better understanding of the integration process, not only as part of the problem-solving insertion of their products but also as part of the requirements for capabilities in their products. JSC gets a benefit for space applications due to understanding of capabilities of intelligent robots required to solve certain types of problems.

Generic Intelligent Robotics Software Architecture

The objective is a generic, supervised intelligent robotics software architecture that provides a portable software approach that integrates intelligent planning and reactive control with sensing and internal representation of environment to enable advanced robots that can recognize and respond to their surroundings and to spoken human supervision in order to perform a variety of manipulation and mobility tasks. The benefits of such an architecture are the faster development time, lower cost, and increased adaptable and flexible performance. In turn, these provide improved productivity by the end users of these robots, whether terrestrial or space.

The proposed effort consists of partnerships between manufacturers, nonprofits, and JSC to develop the software and evaluate its characteristics and robustness in several tasks and environments. The design of the software architecture, which is the framework (functional decomposition) that integrates the separate functional modules into a coherent system, is dictated in large measure by the tasks and nature of the environment. Because both the goal-achieving tasks and the partially unpredictable nature of the environments are similar on Earth and in space, the software architecture can be viewed as generic. Many of the software modules, such as a deliberative planner, world model capability, and natural language interface, can also be viewed as generic. Other software is bundled with certain hardware; e.g., sensing software is bundled with specific sensor hardware.

Current work on the EVA helper/retriever supervised intelligent robot, the mobile robotics testbed project, and Soda-Pup project at JSC have provided us the necessary insight and experience. Not just any architecture will do here. It must solve the key issue of combining deliberative goal-achieving planning with reactive capabilities in such a way as not to limit the intelligence of the planner or the safety of the reactive execution. The JSC work is believed to offer such a solution. It is a practical implementation of the mathematical theory of intelligent robots [14].

The ten tasks in this category are the following:

- Artificial Intelligence Planning Software
- Sequencing and Scheduling Software
- Reactive Controller Software
- Integration of Natural Language Understanding Into Architecture
- Real-Time Speech Planning Software
- World Modeling Software
- Software Development Environment
- Integrated Software Architecture
- Integrated Testing Against Simulated Environments
- Skill Acquisition

A generic software architecture for supervised intelligent robots will enable portability and reuse, major time and cost savings in development and testing, more robust and higher quality software, and maintenance and training cost reductions. People will have a natural means of supervision by including task limited natural language understanding and speech generation software in the robotics software architecture. Improved safety of operations is also a benefit. These benefits apply in space and in the U.S. economy.

Modular Manipulation and Mobility for Robotics

The objective here is to develop a set of standardized modular components that can be reconfigured as required into modular robots offering a broad spectrum of tasks, reduced system costs, reduced weight, reduced mean time to repair, changeout of broken components, and reduced operator training. As components for an integrated prototyping environment for evaluating alternate approaches to design of robotic systems, these contribute to making adaptive robots "faster, better, and cheaper."

The proposed effort consists of partnerships with robotic manufacturers, nonprofits, and universities to develop working production prototypes of a set of standardized modular components The development of standards for mechanical and electrical connections and similar modular interfaces will be a product as well. Both manipulator and mobility systems with robot body structures would be developed Arm sockets, links, joints, actuators, and sensors would be designed and developed to standards for manipulators. Wheels, tracks, suspension, drive train motors, gears, brakes, drive control electronics, structure, pan/tilt units, power and communications subsystems, and sensors would be designed and developed to standards for mobility and body systems.

We have a current effort in designing modular components for space manipulation [4].

The eight tasks in this category are the following:

- Manipulator Socket, Link, Joint, Actuator, and Sensor Modular Component Standards Development
- Manipulator Modular Component Designs
- Manipulator Modular Component Development
- Mobility Modular Component Standards Development
- Mobility Modular Component Designs
- Mobility Modular Component Development
- Modular Robotics Testbed Development
- Modular Prototype Testing on the Testbed

Modular, reconfigurable manipulator arm and mobility subsystems as part of modular, reconfigurable intelligent robots will reduce cost, reduce development time and cost, enable more uses through reuse and reconfiguration, reduce maintenance and repair time and costs, and increase availability (uptime). This approach also enables low cost, rapid prototyping, and rapid development of intelligent robots with testing against intended tasks and environments to improve quality. All applications are beneficial, especially those for space.

Integrated Sensing and Perception Capabilities for Robotics

Development of the capability to select and tailor sensors and real-time perception processing to the task- and environment-driven requirements of adaptive robots is the objective of this portion of the effort. Perception is the extraction of useful information about the environment needed to understand the situation to complete the task successfully.

These capabilities must be integrated with the interface standards of a generic, supervised intelligent robotics software architecture. These are the most important capabilities enabling reactive behavior and deliberative, goal-achieving planning and actions. By enabling advanced robots to recognize their dynamic environments so as to respond appropriately, this effort leads to improved productivity by the end users of these robots, a growing robotics industry that produces jobs and profits, and improved global competitiveness. In space, these capabilities enable robots to provide the flexible support that enables space exploration (by human/robot teams).

Integration of sensing and perception into planning and control in a robust way is a challenge for at least two fundamental reasons. First, the time available to sense and perceive the many dynamic and unpredictable elements of the situation is limited. Second, perception attaches meaning to the link between a conception of the environment and the objective environment. Perception is the process of inference that recognizes regularities in sensor data that are known on the basis of a model of the world to be reliably related to causal structure of objects and their relations in the environment and then conveys this to cognition. Sensory data underdetermines world structure; therefore, a model of world structure is required.

Perception involves understanding generic, generally applicable models of world structure (not merely specific object models) and how that causal structure evidences itself in sensor data. Causal structure is of interest so as to be able to predict consequences, anticipate events, and plan actions so as to achieve goals. Perception is generally focussed by needs for information that supports planning and reasoning for goals achievement. Designing perception involves converting the understanding and inference processes into calculational steps (algorithms and inferences) and designing computation hardware systems to meet the requirements of information at rates and latencies required to deal with a dynamic environment.

The proposed effort consists of partnerships between manufacturers, nonprofits, universities, and JSC to develop a set of sensors and perception processing appropriate to numerous task- and environment-driven requirements for adaptive robot applications, both in the U.S. economy and in space. Included here are vision sensing and visual perception, along with speech recognition and task limited natural language understanding (speech perception). The unification of visual and speech perception is also included here. Proximity sensing, tactile/slip sensing, and force/torque sensing, which are critical aspects of many manipulation tasks, are addressed in the next category of tasks.

Current sensing and perception efforts at JSC include focused developments for EVA helper/retriever (laser scanner, stereo video, torque and proximity sensors, speech recognition and task limited, natural language understanding, etc.) and the mobile robotics testbed project (real-time stereo vision).

The six tasks in this category are the following:

- Vision Sensors and Sensing Software Development
- Finding, Recognizing, Locating, and Tracking Objects and Humans
- Visual Perception of Objects' Spatial Relations
- Visual Perception of Objects' Condition and Process Participation
- Speech Recognition and Natural Language Understanding
- Unification of Visual and Speech Perception

The benefit of sensing and perception capabilities is to enable the supervised intelligent robot both to extract needed information about the changing task environment, including humans, on a real-time basis so as to react safely and appropriately, and to build and continuously update internal representations of the changing environments so as to plan safe goal-achieving actions. People will have a natural means of supervision through task limited, natural language understanding software. The unification of visual and speech perception adds power to the human/robot team. These benefits apply in space and in the U.S. terrestrial economy.

Robotic Surrogates for Human Grasping and Manipulation

Robots and humans must be capable of interacting with the same environment in terms of grasping and manipulation for certain tasks. Dexterous robotic grasping and manipulation capability must be developed to achieve this capability. The robot may operate in conjunction with a human as an apprentice or may be substituted for a human (e.g., in hazardous operations). The benefits to the U.S. economy from robots with such capability would be very large: improved global competitiveness; improved productivity by the end users; a growing robotics industry meaning more jobs and profits; and an increased standard of living in the United States. In space, robots with these capabilities are required to interface with space hardware on astronaut/robot teams. This would reduce the cost of designing the robotic environment and allow more tasks to be done robotically.

The proposed effort consists of partnerships between manufacturers and JSC to develop working production prototypes of human-scale versions of robot hands by leveraging current development by both sides. Integration of tactile, slip, force, and torque sensing; adaptive grasping; stable grasp recognition; and manipulation strategy approaches will be accomplished. However, it should be recognized that the resulting robot hands are not expected to be equivalent to human hands. Limited multitask capability is all that is expected in the 5-year term of this effort.

EVA helper/retriever and the dexterous, anthropomorphic robotic testbed (DART) are two of the current related efforts at JSC, as well as some SBIR developments.

The nine tasks in this category are the following:

- Hand designs
- Integrated hand, wrist, and arm designs
- Tactile/slip sensors, sensing software, and perception software
- Proximity sensors, sensing software, and perception software
- Force/torque sensors, sensing software, and perception software
- Integrated sensing with hand, wrist, arm to provide stable grasp recognition and other intelligent functions
- Grasping and manipulation strategies
- Collision avoidance strategies
- Compliance strategies

Supervised intelligent robots and human ability to interact with the same environment in terms of dexterous grasping and manipulation will provide major benefits in U.S. industry, service applications, and in space. Costly special designs and structuring of the robot environment will be minimized or eliminated, thus reducing costs. Robots will be able to operate in conjunction with humans as robot apprentices to humans on human/robot teams.

An Integrated Prototyping Environment for Robotics

The objective is to develop an integrated rapid prototyping and rapid development environment for building robotic systems "faster, better, and cheaper" based on modularity, reconfigurability, and extendibility, including a library of hardware modules (such as manipulators, tools, and sensors), complementary software modules (such as sensing and perception strategies and manipulator control), and software advisors designed to reduce the cost of programming robots. This effort would also leverage development of a generic intelligent robotics software architecture in a related subcategory.

The proposed effort consists of partnerships between manufacturers and JSC to develop an integrated prototyping environment that will allow users to generate and evaluate alternate approaches to the design of a robotic system quickly. This effort would also leverage development of modular manipulation and mobility for robots and other related subcategories.

Task-directed process design with a systems engineering focus and reconfigurable modular designs are strong points of our experience at JSC that are critical to success here.

The five tasks in this category are the following:

- Requirements for Prototyping Environment
- Design of Prototyping Environment
- Development of Prototyping Environment
- Testing of Prototyping Environment
- Knowledge Support System

Automation of the process of designing and developing intelligent robots reduces costs and development time. Automation of the testing of intelligent robots also reduces costs and development time while providing the user early feedback that the robots will solve the problems. All markets benefit: advanced manufacturing, health care, service industries, construction, mining, space, etc.

Robotic Applications in Advanced Manufacturing, Health Care, Service Industries, and Construction

This effort's objectives are to enable the manufacture and marketing of supervised intelligent robotic systems for applications in advanced manufacturing, health care, service industries, and construction by developing working production prototypes. Production prototypes will also be developed for inspection, servicing, maintenance, and repair tasks for space exploration. Such advanced robotic systems offer the benefits of improved productivity by the end users and improved global competitiveness to the U.S. economy. In space, such robots provide improved safety, reliability, and productivity as Space Station evolves and enables human space exploration (by human/robot teams).

The proposed effort consists of partnerships between manufacturers, nonprofits, doctors and hospitals, universities, and JSC.

The required core skills are available at JSC in its civil service and contractors to form the nucleus of the multiple partnerships. Current technology integration efforts include the EVA helper/retriever supervised intelligent robot, the mobile robotics testbed project, and the Soda-Pup entry in the AAAI national robotics competition. In addition, JSC is responsible for numerous applications of robotics.

In our ongoing relationship with the Texas Medical Center, recent interest by Drs. Steve Kroll and Chuck Van Duren in robotic microsurgery and arterial catheterization has been shown.

The four tasks in this category are the following:

- Robotic Applications in Advanced Manufacturing
- Robotic Applications in Health Care
- Robotic Applications in Service Industries
- Robotic Applications in Construction

The benefits of intelligent robots to advanced manufacturing are spelled out in detail in Erickson[1]. The benefit to health care is lower cost health care delivery with quality improvements due to improvements in productivity. The benefits to other service industry applications are improvements in productivity. The benefits of intelligent robots to construction include improved construction time and productivity.

3. Concluding Remarks

We have presented a "straw man" pre-competitive, dual use technology program in intelligent robotics intended to stimulate the creative aspects of nonaerospace and aerospace industry to propose their own objectives, approaches, and tasks for new jointly funded partnerships with NASA JSC. It is evidence of our "earnest" and that we are ready to proceed with our end of the partnerships.

4. References

1. Erickson, J.D. "Intelligent Robotics Can Boost America's Economic Growth." Proceedings of AIAA/NASA Conference on Intelligent Robotics in Field, Factory, Service, and Space, Houston, TX, Mar. 1994.

2. Clinton, William J.; and Gore, Albert, Jr. "Technology for America's Economic Growth, A New Direction to Build Economic Strength." White House, Washington, DC, Feb. 22, 1993.

3. Bonasso, R.P.; and Slack, M.G. "Ideas on a System Design for End-User Robots." Proceedings of SPIE 1829, Cooperative Intelligent Robotics in Space III, Boston, MA, Nov. 1992.

4. Ambrose, R.O.; Aalund, M.P.; and Tesar, D. "Designing Modular Robots for a Spectrum of Space Operations." Proceedings of SPIE 1829, Cooperative Intelligent Robotics in Space III, Boston, MA, Nov.1992.

5. daCosta, F.; et al. "An Integrated Prototyping Environment for Programmable Automation." Proceedings of SPIE 1829, Cooperative Intelligent Robotics in Space III, Boston, MA, Nov. 1992.

6. Erickson, J.D.; and Lawler, D. "Integrated Computer Aided Concurrent Engineering." Dual-Use Space Technology Conference and Exhibition, NASA JSC, Feb. 1994.

7. Erickson, J.D. "Supervised Space Robots Are Needed in Space Exploration." Proceedings AIAA/NASA Conference on Intelligent Robotics in Field, Factory, Service, and Space, Houston, TX, Mar. 1994.

8. Miles, G.E.; and Erickson, J.D. "Robotics in Controlled Environment Agriculture." Proceedings of AIAA/NASA Conference on Intelligent Robotics in Field, Factory, Service, and Space, Houston, TX, Mar. 1994.

9. Erickson, J.D.; et al. "An Intelligent Space Robot for Crew Help and Crew and Equipment Retrieval." International Journal of Applied Intelligence, accepted for publication in 1994.

10. Erickson, J. D.; Grimm, K.A.; and Pendleton, T.W. "An Intelligent Robot for Helping Astronauts." Proceedings of AIAA/NASA Conference on Intelligent Robotics in Field, Factory, Service, and Space, Houston, TX, Mar. 1994.

11. NASA Small Business Innovation Research Program Solicitation, SBIR 93-1. NASA Headquarters, Washington, DC, May 1993.

12. Office of Technology Assessment. "Exploring the Moon and Mars: Choices for the Nation." Congress of the U.S., Washington, DC, Aug. 1991.

13. Erickson, J.D. "NASA JSC Proposed Dual Use Technology Investment Program in Robotics." JSC-37922, Automation and Robotics Division, NASA JSC, Nov. 1993.

14. Saridis, G.N. "Analytic Formulation of the Principle of Increasing Precision with Decreasing Intelligence for Intelligent Machines." Automatica, Vol.25, Nov. 3, 1989.

DUAL-USE TECHNOLOGIES FOR THE MINING, PROCESSING, AND ENERGY INDUSTRIES

S13-31
44799

by

Barney B. Roberts
Futron Corporation
1120 NASA Road 1
Houston,TX 77058

Kris Kimel
Kentucky Science and
Technology Council
P.O. Box 1049
Lexington, KY 40588-1049

ABSTRACT

Over the years, NASA has utilized several approaches towards transferring space technologies into the private sectors. Some of these approaches have been successful, others have had mixed results. The conventional approach usually involves identifying advanced NASA technologies and then searching for applications. Some approaches involve joint sponsorship, but mostly focus on technologies for space. The greatest success has occurred when market forces are used to determine technology initiatives. (See Reference 1).

This paper describes an unconventional approach that was structured to drive out customer requirements for advanced technologies where NASA is also a customer on par with others. Using the models defined in reference 1., the approach used herein is best described as entrepreneurial deal-making. This approach is new and is working very well so far, but it is still too early, and the process is too immature, for quantitative evaluation of success. However, it is appropriate to share these experiences at this time in order to obtain feedback and improve our chances for success.

One of the distinguishing factors of this approach is that NASA is not the "sole customer" nor the "sugar daddy". In the needs identification stage, NASA is one of many users (customers), and in the subsequent development stage, NASA is one of many suppliers along with industry, academia, and other government organizations. This specific characteristic of the approach was a primary goal that was incorporated from inception. It was the viewpoint of the instigators (the authors) that if the activity was customer focused, it would:

1. Have a higher probability for success since it will be driven by those who will reap the benefits.
2. Be able to advocate and promote action if necessary, since it would be founded outside the federal government.
3. Not be self-perpetuating; that is, if no common need could be found that had a reasonable return on investment, it would self-destruct.
4. Have increased stability from a broader base of support and not be dependent on NASA being the principal funding source.

To date, the workshop activities have identified a collection of potential customers in NASA, other federal government, private industry, and academia who have common needs for advanced technologies. These potential customers are beginning to collect around the following application categories:

1. Mining technologies
2. Materials processing technologies
3. Energy technologies

It must be noted here that not all benefits are derivable from the utilization of a "new technology" per se'. For example, Universities need to *educate* and the *focus* of the research is a lower priority; the mining industry is required to invest in returning mined areas into useful areas, such as a test bed for surface or subsurface robotic vehicles; and etc. As these "strange bedfellows" have shared ideas in the workshops, some dual use technologies have been identified and some lessons have been learned on how to encourage and nurture this process.

INTRODUCTION

The NASA Space Exploration Initiative began in the early 1980s albeit under different names. That was not the first time that the Space Exploration Initiative was proposed. The first time was in the 50s with studies by Werner

VonBraun and revisited again in the late 60s by the Space Task Group in a study for Vice president Agnew. Each time the human exploration of space was revived, new technologies changed the configuration or the performance of the systems. In the 1969 Agnew report, nuclear thermal propulsion was the latest technological innovation, approximately doubling the trans-Mars space ship performance. In the revival of the 1980s, the literature contained substantial data and information about the Apollo lunar samples. This database of information brought a new perspective on exploration. A lunar base study by the University of Houston suggested the use of indigenous resources for propellants, life support and construction.

The Planet Surface System Office at the Johnson Space Center sponsored the development of a surface system and operations simulation model. This model, developed by the Large Scale Programs Institute in Austin, Texas, demonstrated that the use of lunar produced propellants reduced the costs of space transportation to the lunar surface by 1/3.

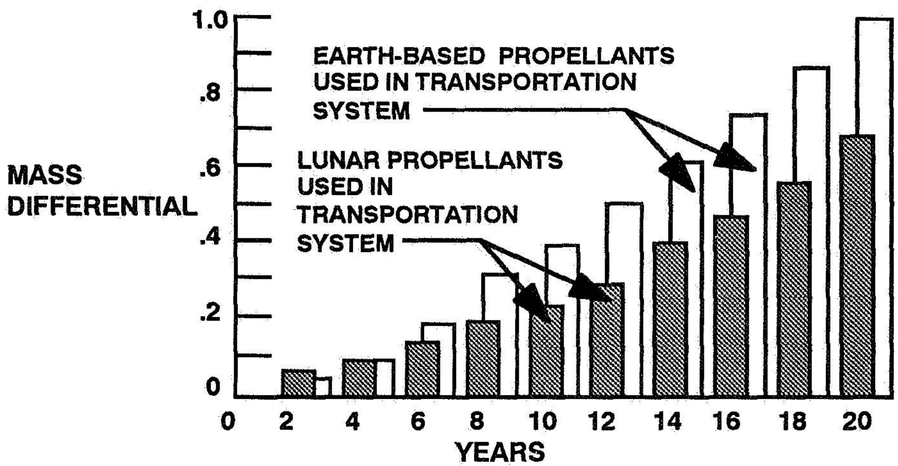


Figure 1. *For a fixed mission set delivered to the lunar surface, the use of lunar propellants in the lunar transportation system can reduce the transportation costs (~mass delivery requirements) to the lunar surface by 1/3 and shows a return on investment for the mining and processing equipment in 3 years.*

With this new perspective, NASA needed access to new skills and began to cultivate new government, industrial and academic partners. These partners included Bureau of Mines, Department of Energy, US Army Corps of Engineers, architect and engineering firms, construction industry, mining and processing industries, and the energy industry.

One of these interested parties was the Kentucky Science and Technology Council (KSTC). The KSTC works in partnership with the State Government of Kentucky, the colleges and universities of Kentucky, the Kentucky Public Education System, and the mining and processing industry of Kentucky. KSTC is chartered to advance science and technology in the state of Kentucky. In 1991, KSTC entered into discussions with NASA to seek common interests in technology developments that were aligned with the Space Exploration Initiative.

THE DUAL-USE PROCESS AND RESULTS

Strategy and Planning.

In the initial strategic planning discussions, as much effort was given to the methodology of cooperation as to the content of the cooperative technology developments themselves. In the beginning, we wanted to be successful and we saw many problems with the typical approach of NASA as technology customer and the KSTC associates as technology supplier:

- NASA's commitment to follow-through is unilateral and may be terminated without regard to the interests of the KSTC partnership
- The resulting technologies will be focused for application in space, and upon completion of the space objective, there is no basis for continuing commitment to the KSTC partnership
- For long term commitment within the state of Kentucky and the mining and processing industry, direct benefits must be designed into the program from the beginning

Thus, we made an early decision that we would not proceed with the concept of NASA as technology customer and the KSTC associates as technology supplier.

We agreed to go forward with a joint technology development approach ("Dual-use" was not yet coined). Unfortunately, there were no cookbooks on how to proceed.

Process

The process was not initially formalized, but was allowed to evolve as we learned how to proceed. We kept our focus on the main goal that this project would develop joint technologies equally benefiting all parties involved. We then employed short-term procedures as needed. We were very conscious of the competing needs of imposing structure to assure convergence, versus the need to remove boundaries to assure the emergence of creative concepts. The lessons learned from this experiment could be helpful in meeting the balance in future initiatives of this type.

The Getting to Know you Stage. Our first meeting was informal. It was for the purpose of introductions and learning who each potential partner was and discovering each other's skills and needs in very general detail. The meetings were kept small and limited to only a few representatives of each interest, however, *every* interest was represented. The purpose was to gain enough information to plan the next meeting. It was determined that the next meeting was to be introductory, addressing skills, capabilities, and needs, but open to a much larger participation and in a more formal presentation format.

This initial meeting was very beneficial in that a small investment transferred a significant amount of understanding of each other's needs. Most importantly it identified the diversity of the needs. All needs were "technology related" but not necessarily "technology-focused". For example, the Kentucky University System "needs" to retain its doctoral graduates in the state and a high-technology initiative in the State is a contributing objective. The mining industry is required to return "value" to the regions in which it is removing resources; supporting a high-technology initiative is one possible means towards this end.

Formal Presentations of Skills, Capabilities, and Needs. The format of the second meeting was formal presentation with ample time allowed for discussion. The attendance (about 40) was limited by invitation-only but broadly covered the KSTC/NASA interest group. This meeting was very effective in providing detail of each participant's skills and capabilities, but fell short of identifying *mutual* needs. The predominate paradigm of the meeting was one of "understanding NASA's needs and how to meet them" as opposed to the objective of joint needs. Several times the question was asked "What [technologies] does NASA want us to do?" which clearly indicated the difficulty in effecting the desired paradigm shift *from:* customer-supplier *to:* customer-customer, then supplier-supplier. Figure 2. illustrates the desired paradigm shift.

This was our first discovery of how to improve the process. That is, the greatest difficulty in identifying dual-use technologies is creating within the participants the concept that in the initial phases of definition, NASA is a "customer" on par with every other participant.

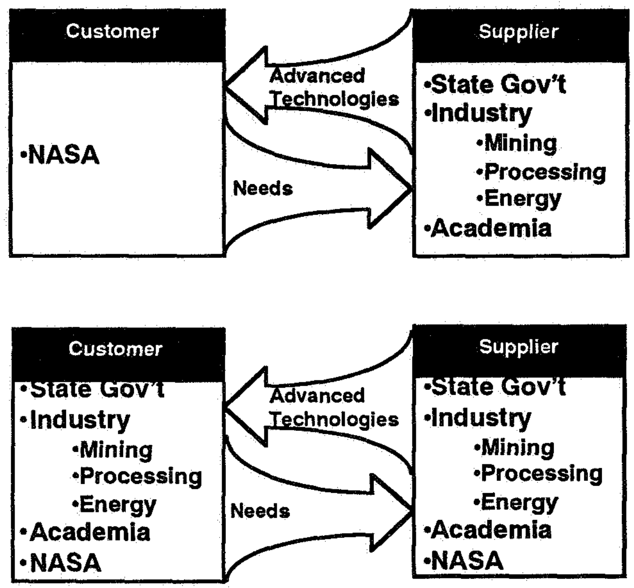


Figure 2. *The greatest problem was getting the participants to internalize the desired paradigm shift that NASA was on par with all other participants as a customer.*

Figure 2 is intended to graphically depict this problem. Had we known a' priori the extent of this difficulty, the use of a graphic similar to Figure 2 may have helped.

The next stage was to be a needs workshop with open attendance with announcements in technical publications and by mailings. In order to better set the stage, the needs workshop was organized into (1) Identifying the Customers and the Needs, and (2) Meeting the Needs.

The Needs Workshop. The needs workshop was well attended, exceeding expectations by 50%. This was a clear indicator of the interest in dual-use technology initiatives in the mining, processing, and energy industries. The workshop was organized into alternating sessions of working groups and plenaries with a kickoff of keynotes to set the stage. On the first day, the working groups were focused on identifying the customer and the needs. On the second day the working groups were focused on meeting the needs. Although the paradigm shift problems continued to predominate, it appears as though consensus was achieved on a vision and an initial objective.

Results

It is important to caution the reader that the results reported herein are preliminary in that they were assembled post-workshop and have not had the review and concurrence of the attendees.

Vision. It was clear that the vision of this emerging consortium was to be a part of the solution to the world energy problem. There is overwhelming evidence that the energy requirements for the world's populace will outstrip terrestrial supply early in the next century (see Figure 3). It was also very clear that the solution was going to require a broad range of new technologies involving space and terrestrial applications.

potential dual-use technologies that meet most of the customers' needs.

Near-Term Objective. There is growing consensus that to meet these needs, a robotic system test bed is required. Some initial discussions have begun to organize a consortia to develop the robotic test bed in Kentucky.

The Next Steps. The next steps are to:

(a) Develop a strategic plan:
  (1) Carefully craft a vision statement
  (2) Develop the goals and objectives that define ends towards achieving this vision
  (3) Prepare a SWOT analysis (Strengths, Weaknesses, Opportunities, Threats)

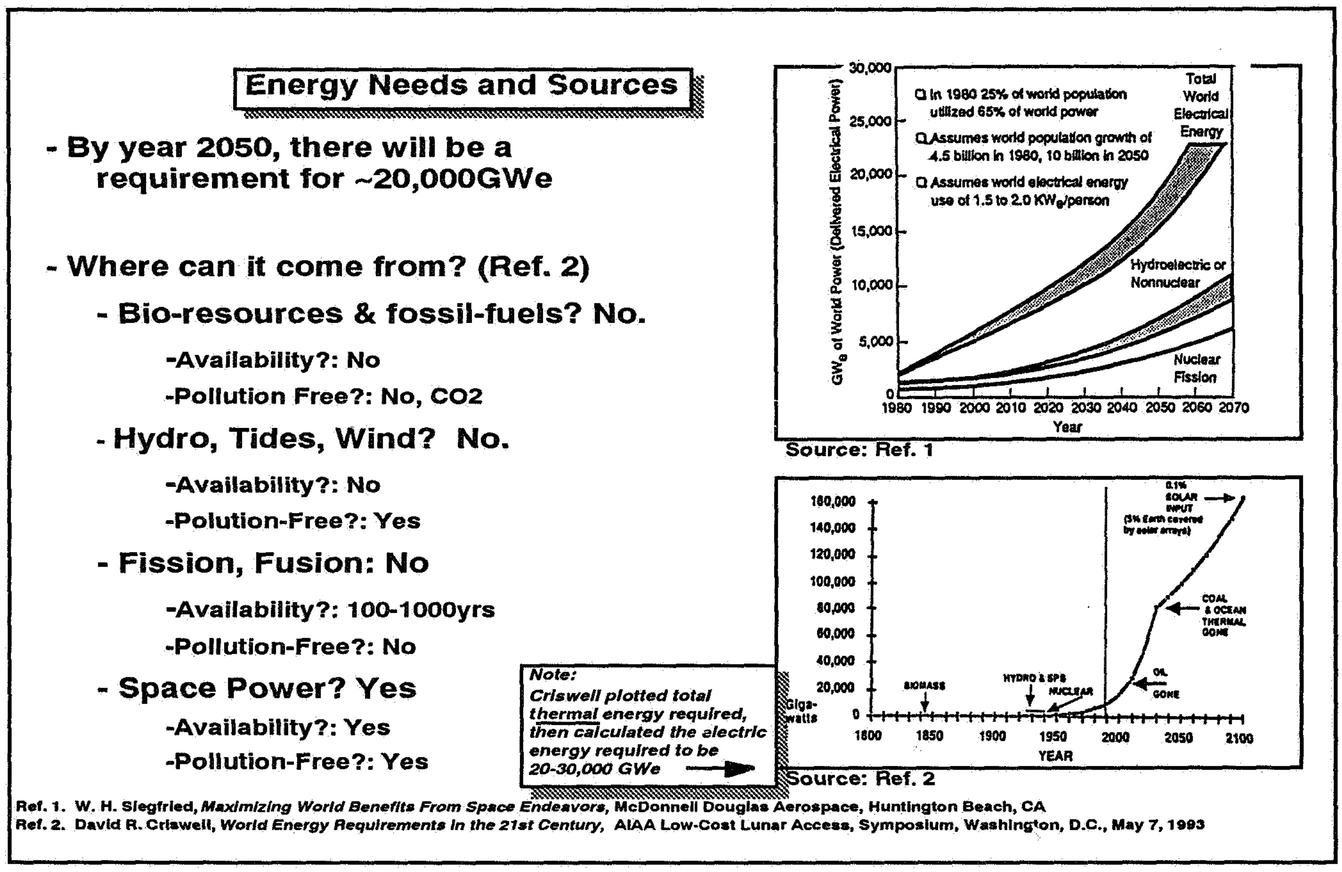


Figure 3. *Energy demand will surpass the total terrestrial energy supply early in the 21 st century.*

A possible vision and strategy is shown in Figure 4. The vision is to move the energy industry into space and the near term objective is to begin the development of the robotic mining and processing technologies that are needed for both (1) energy from space and (2) increased mining efficiencies to bridge the gap from terrestrial energy supply to space supplied. Figure 4. fades to gray in the out-years to indicate the increasing uncertainty in the future. Figure 5 and 6 are possible mappings of the

(b) Define the near-term program:
  (1) Define the needs and achieve consensus among the "customers"
  (2) Analyze the needs, execute trades and develop requirements for the robotic test bed
  (3) Prepare concepts for the test bed, develop proposals
  (4) Develop business plans and consolidate the consortium

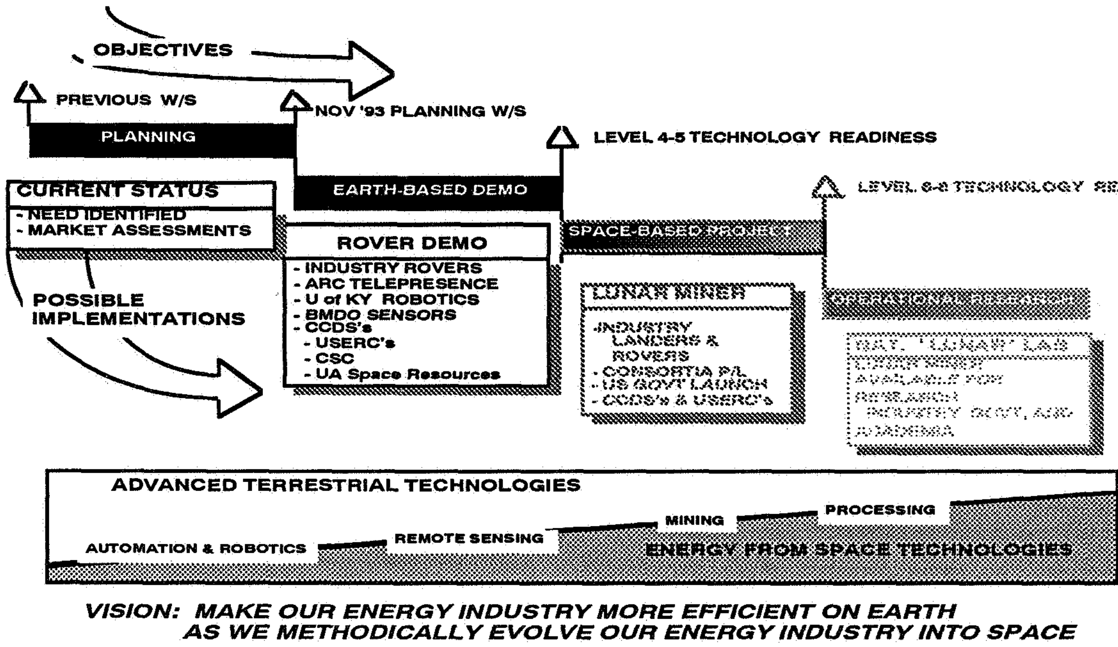


Figure 4. *The workshop participants were gravitating towards a vision that they wanted to be a part of the solution to the energy problem by contributing technologies that solve near-term energy problems as well as long-range energy problems*

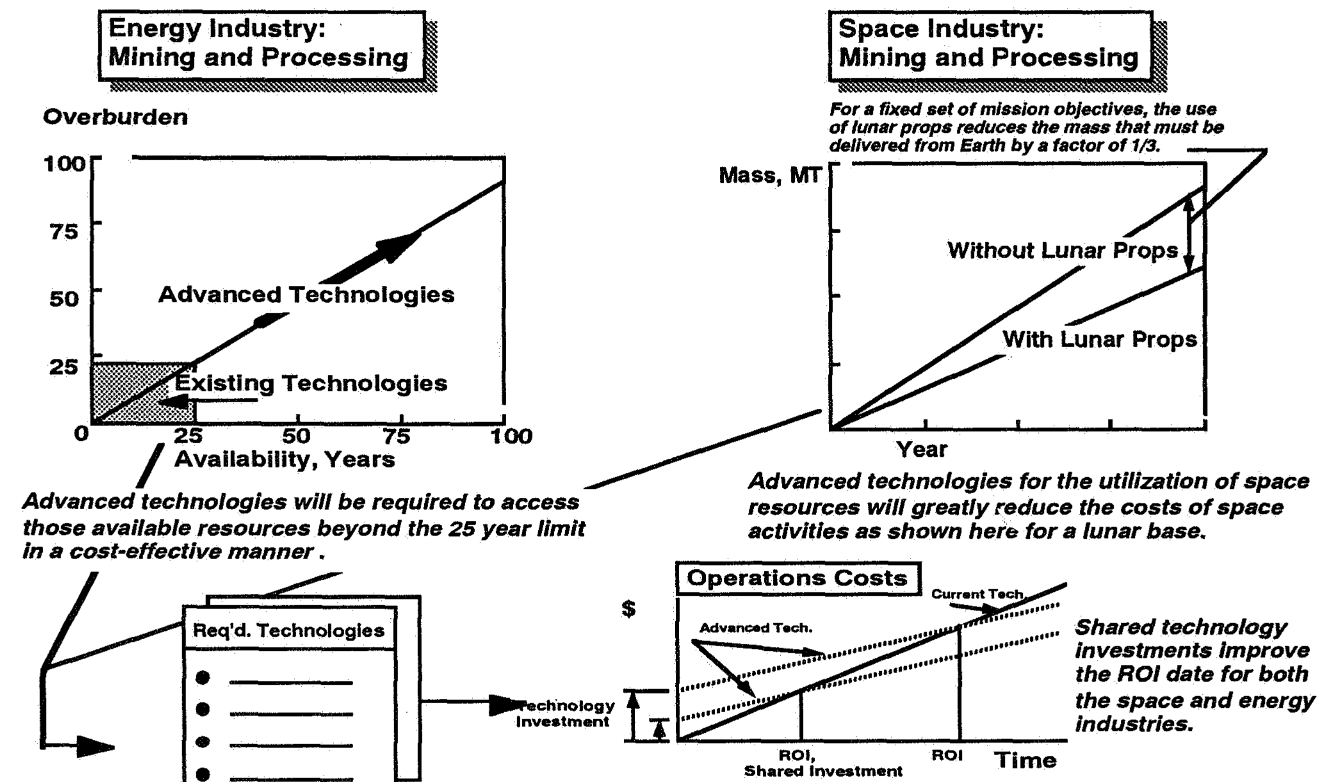


Figure 5. *The workshop participants were able to identify common technology needs. Mining and processing technologies that are needed now can directly contribute to future space missions*

Terrestrial Application	Technology	Space Application
•Mineral Analysis, Yield Estimation - Deep Mine Vein Location and Tracking •Wall and Ceiling Integrity	•Advanced Sensors	•Mineral Analysis, Yield Estimation Surface Mineral Analysis and Resource Location
•Deep Mine Robotic Operations • Mining • Beneficiating • Removal	•Advanced Robotic Mining	•Surface Mining Operations • Mining • Beneficiating • Transportation
•Improved Automated Processing; Increased efficiency	•Automated Processing: Advanced FDIR	•Remote, Low-Maintenance, Processing
•Reliable, Low-Polution Personal Trans. •Regenerable Energy Economies •Small, Decentralized Power Systems for Remote or 3rd World Applications	•Alternative, Regenerable Energy Economies • Methane/O2 • H2/O2	•ISRU Based Engines •Regenerable Energies •Hi-Density Energy Storage
•Environmentally-Safe Energy Production	•Space-Based Energy Generation & Transmission	• Surface Power Gen & Beaming

Note: Technology "Spin-In" = →
"Spin-Off" = ←
"Both" = ↔

Figure 6. *This is a tabulation of the common technologies identifies by the workshop participants*

CONCLUSION/LESSONS LEARNED

Regarding the Program

- There is a need to begin developing technologies that jointly contribute to future energy supply solutions (some involving energy from space) and bridging the gap by improving the efficiency of terrestrial mining capability.
- The near term needs are highly automated mining and processing technologies.
- There exists potential support within the attendees to begin a discussion of a potential robotic test bed.

Regarding the Process

- The *approach* appears to be viable and properly focused on the needs of the "customer".
- The *execution* was flawed, but not stymied, by a paradigm shift problem.
- The basic lesson-learned was that it is very difficult to communicate the concept of all participants being on par as customers as well as suppliers --dual roles.
- Many "needs" are abstract, relating loosely to the technologies, such as the need to retain Phds in the state, or the need for the mining industry to return "value" to the mined regions. These needs are sometimes the strongest ones and must be openly considered.

REFERENCES

1. Sacknoff, Scott, Jennings, John, Lacobie, Kevin, "Technology Transfer: Models, Methods, and Summaries, Futron Corporation, Bethesda, MD, December 15, 1993.

514-63
44801

VIRTUAL REALITY APPLICATIONS IN ROBOTIC SIMULATIONS

David J. Homan, Charles J. Gott, S. Michael Goza
Automation and Robotics Division
A&R Lab Management Branch
NASA Johnson Space Center
Houston, TX 77058

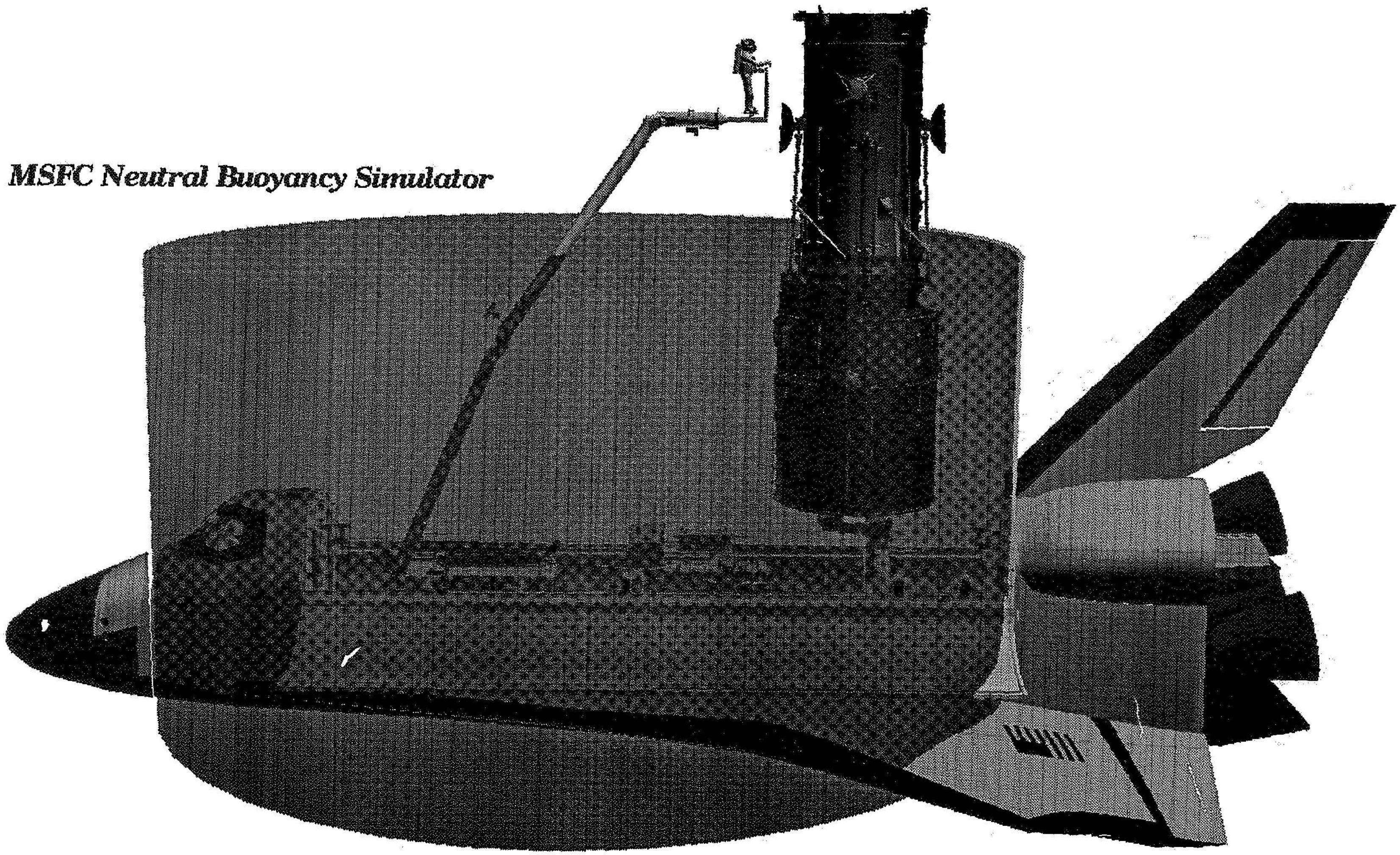
The preparation for the repair of the Hubble Space Telescope (HST) by the seven member crew of the Space Shuttle Endeavour on mission STS-61 in December, 1993, provided a perfect opportunity for developing and evaluating the use of Virtual Reality technology as applied to the solution of a real-world problem. The problem, which was not unique to that flight, arises during ground-based training for space shuttle missions that involve extravehicular activities (EVA) in conjunction with Remote Manipulator System (RMS) operations, in general; and more specifically, the integrated operations that occur when an EVA person is being maneuvered around the payload bay while standing on the end of the manipulator. Current ground-based simulators do not provide complete integrated training for these scenarios due to physical limitations or safety concerns. VR provided a means to practice integrated EVA/RMS operations in the on-orbit configuration with no discomfort or risk to the crewmembers involved.

The neutral buoyancy facilities at the Johnson Space Center (JSC) and the Marshall Space Flight Center (MSFC) provide excellent training for EVA work, but because of limited volume (Figure 1) cannot encompass the entire RMS reach envelop to provide adequate RMS training. The Manipulator Development Facility (MDF) and the Shuttle Mission Simulator (SMS) at the JSC provide excellent RMS training, but do not support EVA training. Virtual Reality afforded the STS-61 crew the luxury of practicing the integrated EVA/RMS operations in an on-orbit configuration prior to the actual flight. The VR simulation (Figure 2) was developed by the Automation and Robotics Division's Telepresence/Virtual Reality Lab and Integrated Graphics, Operations, and Analysis Lab (IGOAL). The RMS Part Task Trainer (PTT) was developed by the IGOAL for RMS training in 1988 as a fully functional, kinematic simulation of the shuttle RMS and served as the RMS portion of the integrated VR simulation. The "EVA person" was tied into the system via a head mounted display system, two (2) data gloves (both right and left hands), four (4) electro-magnetic tracking sensors (1 for head tracking, 2 for hand tracking, and 1 for tracking the object being handled by the EVA person), and the software required to generate 3-D graphics, do collision detection between the subjects hands and other objects in the virtual environment, and control the flow of data from sensors-to-simulation and simulation-to-simulation. The entire simulation resides on a Silicon Graphics Onyx Reality Engine 2 System with three (3) Graphics pipelines, and six (6) 150 MHz R4400 CPU's. In this configuration, one graphics display is used by the RMS simulation, while the other two displays are used to drive the head mounted display system.

Members of the STS-61 crew used the system on eight separate occasions for a total of 16 hours of preflight work. During those sessions, the crew was able to rehearse complete EVA/RMS tasks by taking advantage of the systems capability to present the on-orbit configuration which allowed the full range of the RMS to be simulated. This allowed the EVA person on the end of the RMS to be placed in a position that would be required during the flight, but not attainable in the ground-based facilities (see Figure 1). Not only could the RMS operator see the correct RMS configuration, but the EVA person could see the configuration from the correct vantage point. By integrating the two simulation capabilities, the RMS operator and the EVA person were also able to develop the command protocol between them and have confidence that each new what the other meant when the maneuvers were performed during the actual EVA's. Because the EVA crewmember could get a realistic view of the shuttle and payload bay in the VR simulation, he/she could explore different positions and views to determine the best method for performing a specific task, thus greatly increasing the efficiency of use of the neutral buoyancy facilities. A number of task procedures and RMS positions derived in the neutral buoyancy facilities were changed when the integrated VR simulation showed them to be unsuitable for achieving the task. One other added benefit of the VR system noted by the crew was that when using VR, the EVA crewmember relies only on visual cues to determine his/her orientation (as when in space) instead of the gravity cues received in the neutral buoyancy facilities.

Similar EVA/RMS training challenges face future astronauts in preparing for the assembly of the Space Station. The Virtual Reality Training Simulation, with its present capabilities, will not replace any of the current ground-based training facilities in the foreseeable future, but will provide additional dimensions to those facilities and fill in the gaps in EVA/RMS training that are inherent in today's training scenarios.

HST Repair Mission



IGOAL Figure 1. Neutral Buoyancy Facility volume limitations

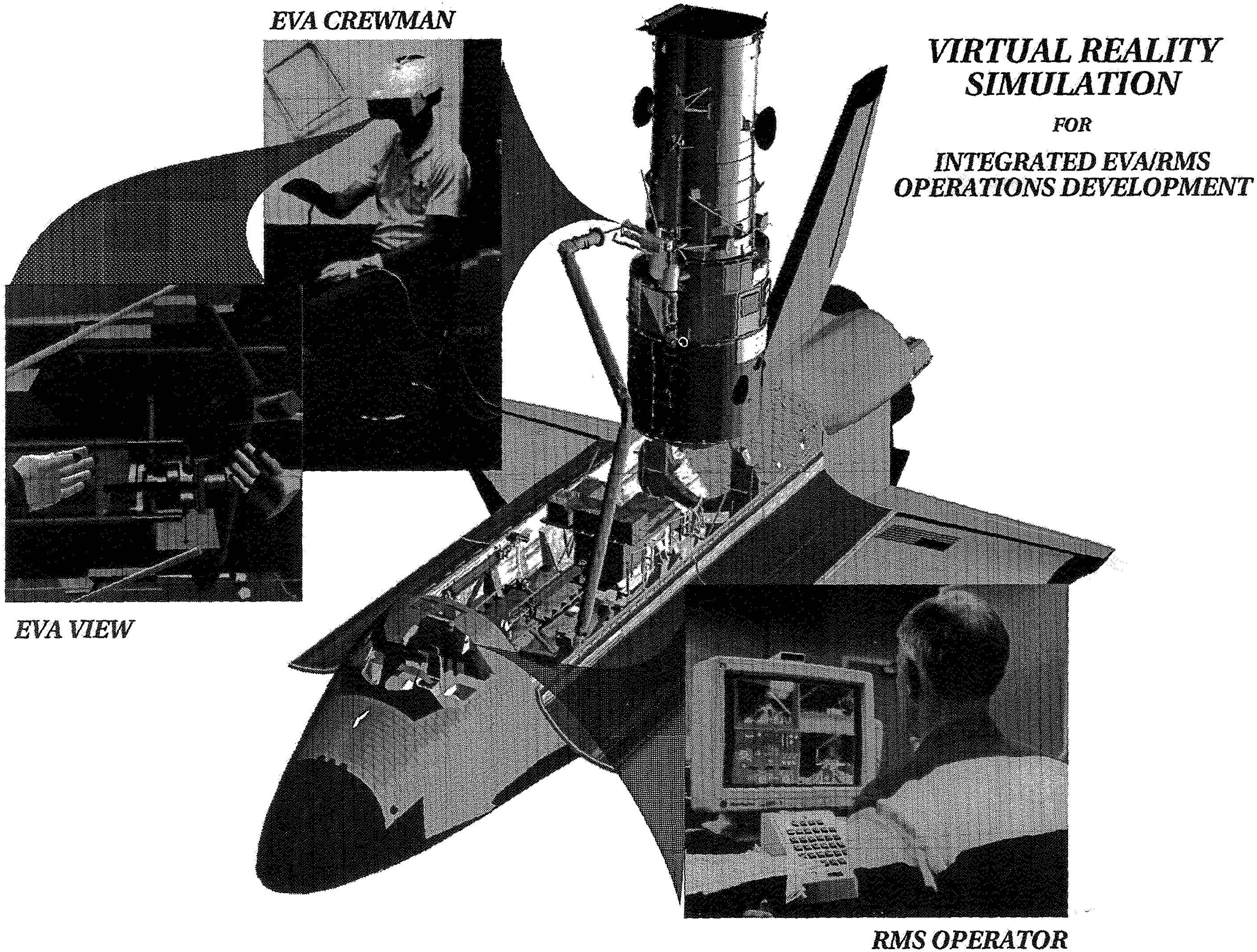


Figure 2. Integrated EVA/RMS Virtual Reality Simulation.

S/5-37
44805

A Distributed Telerobotics Construction Set

James D. Wise
Department of Electrical and Computer Engineering
Rice University, Houston, Texas 77251

February 1, 1994

ABSTRACT

During the course of our research on distributed telerobotic systems, we have assembled a collection of generic, reusable software modules and an infrastructure for connecting them to form a variety of telerobotic configurations. This paper describes the structure of this "Telerobotics Construction Set" and lists some of the components which comprise it.

1. Introduction

The Universities Space Automation and Robotics Consortium (USARC) was formed in 1989 to promote research into robotics and telerobotics for space-based applications. It consists of five universities (Rice University, Texas A&M University, the University of Texas at Arlington, the University of Texas at Austin, and the University of Texas at El Paso) in conjunction with NASA's Johnson Space Center. One of our major areas of research has been teleoperation over long distances, typically involving long delays, low bandwidth, and possible loss of data. In the course of this research we have assembled an experimental telerobotics system which connects robots and control sites at the universities and JSC.

In order to conduct research efficiently in this distributed environment we have built our system using a construction set approach: it consists of a number of independent modules which may be connected together in any (reasonable) configuration. This independence allows each group of researchers to concentrate on their own area of expertise and have their work immediatly integratable into the overall system.

2. The Construction Set Approach

Figure 1 is a block diagram of one of our basic configurations for teleoperation: a robot worksite with two remote operator control sites. It consists of a number of clearly defined functional blocks communicating via streams of data packets using a few well defined formats.

Our experience with these diagrams has been that simply drawing the modules is enough to define the system: the interconnections largely draw themselves. Each module accepts a particular type of input data and produces another type as output. By matching the data types, we form the required connections.

This idea of letting the data define the connections, rather than imposing them externally gives rise to what we call a *data centered* approach to modularity using *undirected messages.* In such a system, consumer processes declare the types of data they are interested in and producer processes declare the type of data they distribute. Based on this data centered approach, we have developed a data exchange mechanism that we call the Telerobotics Interconnection Protocol, or TelRIP.

TelRIP is described in detail elsewhere [1]. Here ia a brief summary of the key points:

- It supports a uniform interface to a variety of underlying communications media.
- Processes communicate by exchanging Data Objects.
- Byte order and format translation are handled automatically.
- Undirected messaging provides data distribution based on properties of the Data Objects.

The last item is the key to the successful modularity of our system. Rather than data being sent to a specific

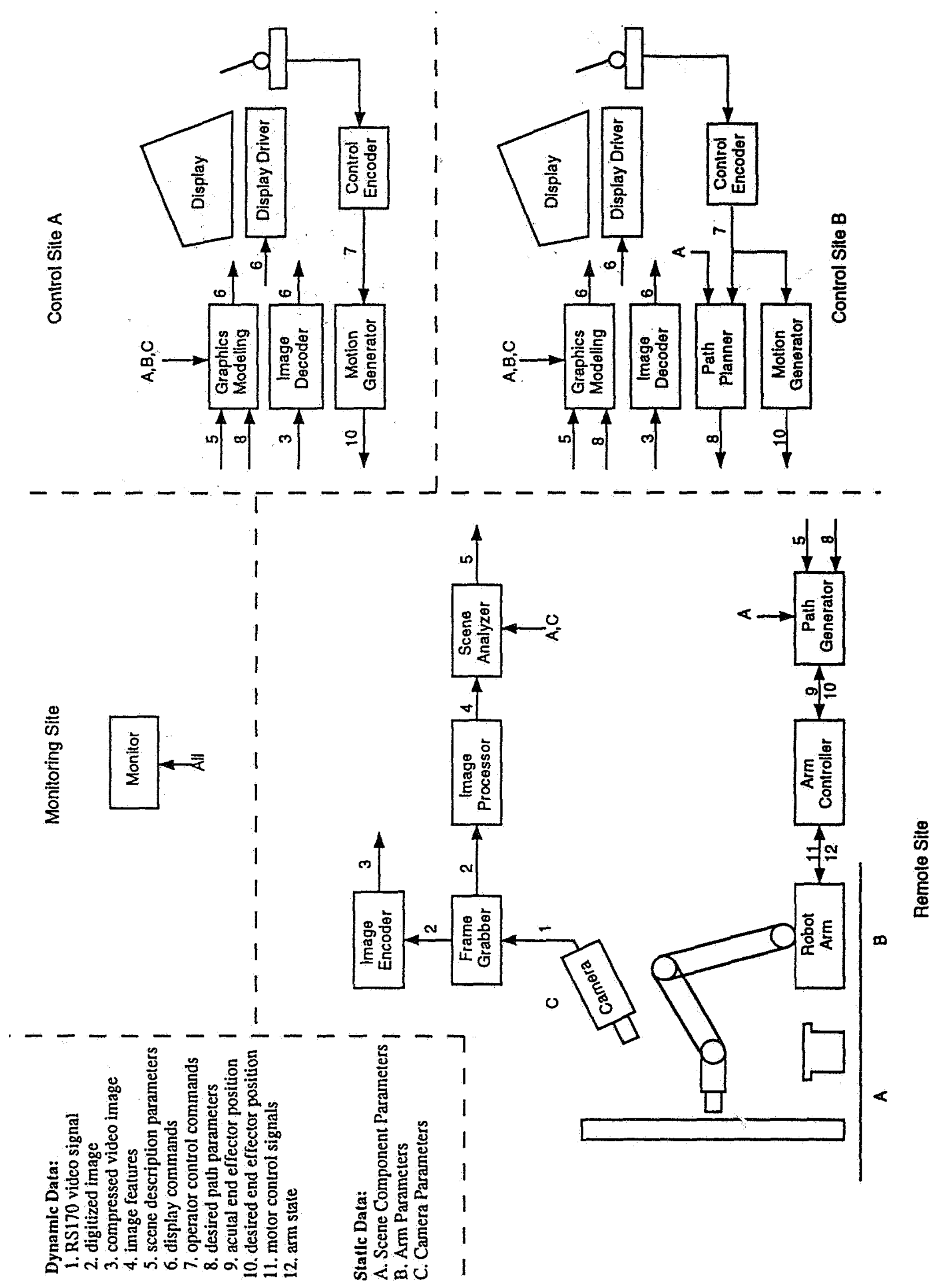


Figure 1: A distributed Telerobotics System

destination, resulting in a "hard wired" connection, modules specify the types and other characteristics of the data they wish to receive. An "automatic connection" will be made to whichever other modules provide the required data. Since this connection is dynamic, modules may be added to or removed from the configuration at any time with minimal impact on the rest of the system. This gives the system a high degree of both flexibility, and robustness.

Conceptually, our construction set consists of three components:

- A set of Functional Modules which perform fairly narrowly defined, hopefully generic, functions.
- A standard, minimal vocabulary of data objects which encapsulate the information communicated among the modules.
- A data exchange mechanism which delivers the data objects from the producers to the consumers, that is, it forms the connections between modules.

Each of these conceptual components has a physical realization:

- The functional modules are implemented as programs, or "Processing Modules"
- A common vocabulary is achieved by defining standards for data objects which are used by all programs. These standards define both the format of the data, and the semantics.
- Data exchange among modules is handled by TelRIP.

3. System Components

TelRIP provides the framework, or building board, for our construction set. The remaining components are a set of programs to implement the required functional modules, and a common data object vocabulary.

3.1. Data Object Vocabulary

The atomic unit of communication between modules is the TelRIP Data Object. A data object is similar to an object oriented programming object in that it contains a self identifying data record. TelRIP data objects lack an explicit class hierarchy and imbedded method code. However, they incorporate additional ancillary information, or *Properties* which are used in classifying and directing objects.

These properties include:

- A timestamp, which gives the time at which an object was distributed. TelRIP maintains synchronization of timestamps among systems.
- A source identifier or *address.*
- An intended destination address.
- Qualifying properties (e.g. compressed, operator generated, etc).

Although there is no explicit class hierarchy, it is convenient to group the object types into classes. Table 1 is a brief description of the major object types used in our system.

3.2. Functional Modules

Like the data object types, the functional modules may be grouped into classes:

Class	Functionality
Operator Interface	
	Hand Controller Interface
	Controller Semantics
	Simulation Display
	Video Image Display
	Control Panels
	Interoperator Communication
	Audio Response
Robot Interface	
	one for each type of robot
Motion Control	
	Motion Semantics
	Transform Generator
	Force Control
Video	
	Frame Grabber
	Image Compressor
Monitoring	
	Human Performance
	System Analysis/Debugging
Utilities	
	Video Recorder
	Audio Recorder

Table 2: Functional Modules

Class	Type	Function	Contents
sensor	IMAGE	transmit a video image	image size pixels
	FORCE_TORQUE	transmit force and torque data	f_x, f_y, f_z n_x, n_y, n_z
control	COMMAND	send a general command	command code optional arguments
	HAND_CTLR_STATE	transmit raw hand controller data	axis values button states
	ERROR	return an error condition	error code
motion	JOINT_STATE	specify joint angles	$\theta_1, ..., \theta_n$
	JOINT_TRAJECTORY	specify a path in joint space	array of joint states
	CARTESIAN_STATE	specify position and orientation of end effector	x,y,z roll,pitch,yaw
	CARTESIAN_TRAJECTORY	specify a path in cartesian space	array of cartesian states
	HOMOG_TRANSFORM	specify viewpoint to effector transformation	matrix coefficients
	MOTION_SEMANTICS_CONFIG	specify mapping from hand controller to end effector	axis modes reference frame
parameters	HAND_CTLR_DESCRIPTION	characterize a specific hand controller	gains axis assignments
	CAMERA_DESCRIPTION	describe camera view	camera location field of view
communication	TEXT_MESSAGE	printed interoperator communication	text characters
	AUDIO_RECORD	spoken interoperator communication, audio command response	sample rate audio samples

Table 1: Data Object Types

3.3. Processing Modules

The functional modules are implemented by a set of programs or *Processing Modules.* Table 3 lists the major programs in our construction set. The origins of some of the program names are, as they say, lost in antiquity.

A few of these will be described in more detail:

TSM Although the automatic interconnections provided by data type matching are adequate for most single robot, single controller scenarios, when multiple instances of a particular functional module are present in a configuration, it is necessary to be able to distinguish them. This is done using the address property mechanism: each instance of a duplicated module has a unique value of the address property. This property may be specified either as a physical address or a functional address.

To simplify the dynamic management of these properties in situations where control is passed from one robot or controller to another, we have developed a program called the Telerobotics Session Manager (TSM). TSM allows the connections among modules to be displayed and modified while the configuration is running.

VCP Currently the largest class of modules is what we call the Operator Interface Components. By implementing these as a set of narrowly scoped components rather than a single monolithic entity, we can construct a Virtual Workstation (VWS), where the operator interface for a variety of configurations can be assembled on a single physical workstation.

We have extended the system independence concepts of TelRIP to the construction of operator interfaces with a system called the Virtual Control Program (VCP). VCP allows a user interface to be defined in a very straightforward way, independent of the system on which it will be displayed. This allows a programmer to write a single program with a graphical user interface which will run on Open Look, Motif, Windows, or any other window system without having to know how to program any of them.

Interfaces are defined in terms of commands, parameters, adjustments, etc. and are implemented using buttons,

Program	Function
spaceball	convert device specific codes into generic hand controller objects
hc_robot	generate robot control objects from hand controller objects
TDM	graphical simulation display
tdm_sim	interface TDM to TCS
displayx	display a video image on a workstation screen
cmdks	generic command line interface
vcp	generic graphical user interface
chat	printed interoperator communication
phone	audio interoperator communication
audioresp	voice response to commands
remote_*	translate robot control objects into device specific commands
force_dsp	display forces and moments
motion	motion semantics controller
tfb	tool frame builder
pcgrab	digitize one frame of video
vcompress	compress a frame of video for long haul transmission
sggrab	grab and compress video with remote pan and zoom
tsm	Telerobotic Session Manager
rmonitor	gather human performance data
xmon	network monitor
vcr	video recorder
dat	audio recorder
siggen	audio signal generator

Table 3: Processing Modules

text items, sliders, or whatever appropriate constructs are available from the target windowing environment. A system dependent program (VCP main) runs on the target display and communicates with the client Processing Module to generate the GUI display and updates and commands between the PM and the display.

4. Demonstration Configurations

We have built or are currently building a number of demonstration configurations using these components. These include:

4.1. Manual Controller Performance

A standardized task was performed by a number of operators using several different manual controllers. The data streams were analyzed to determine the human performance parameters of each session.

4.2. Multiple Robots, Multiple Controllers

This configuration demonstrates the interoperability and dynamic reconfigurability of our system modules. Two robot worksites (A and B) and two robot control sites (Y and Z) performed the following scenarios:

- A task at site A was begun by controller Y. Midway through the task control was switched to site Z, who completed the task.
- Controller Y performed a task at site A. Using the same control components, he performed a second task at site B.
- Controller Y performed a task at site A simultaneous with controller Z performing a task at site B.

4.3. Extended Teleautonomous Control

Modules intended to increase the performance of an operator controlling a remote manipulator under adverse circumstances (delays, restricted visibility, etc) will be added to the basic configuration. These include: Time and position clutches, time brake, voice control, and force control.

4.4. Network Stress Testing

The reliability and robustness of the system are examined by increasing the stress (typically an increased data rate) on various components.

4.5. Workload Analysis in Shared Human-Autonomous Tasks

Real-time measurement of human workload and overall system performance will be made on a set of tasks having varying degrees of operator involvement and autonomous control.

5. Acknowledgments

This work has been supported by NASA under grant NAG9-461 and by the State of Texas under TATP grant 999903-267.

References

[1] J.D. Wise and L.A. Ciscon, "TeleRobotics Interconnection Protocol Operating Manual," Tech. Rep. 9103, Universities Space & Automation Research Consortium, 1991.

S16-37 44807

"DUAL USE DISPLAY SYSTEMS FOR TELEROBOTICS"

Michael J. Massimino
Michael F. Meschler
Alberto A. Rodriguez
McDonnell Douglas Aerospace
16055 Space Center Blvd.
Houston, Texas 77062

Abstract

This paper describes a telerobotics display system, the Multi-mode Manipulator Display System (MMDS), that has applications for a variety of remotely controlled tasks. Designed primarily to assist astronauts with the control of space robotics systems, the MMDS has applications for ground control of space robotics as well as for toxic waste cleanup, undersea, remotely operated vehicles, and other environments which require remote operations. The MMDS has three modes: 1) Manipulator Position Display (MPD) mode, 2) Joint Angle Display (JAD) mode, and 3) Sensory Substitution (SS) mode. These three modes are discussed in the paper.

1. Introduction

Manual control of a remote manipulator can be a difficult task due, in part, to a lack of useful feedback to the operator on the position of the manipulator with respect to its desired position, destination, or target object to be manipulated. For example, to control many remote manipulator systems, including the space shuttle remote manipulator system (SRMS), the operator relies largely on visual feedback from direct views through windows and indirect views from cameras. However, the visual information can be insufficient in providing the operator with adequate cues, due to obstructions, poor viewing angles, camera failures, or problems with resolution or camera control. Our first mode, the Manipulator Position Display (MPD) mode addresses these problems.

Another area of which poses concerns for the operator is avoiding undesired positions which cause joint limits or singularities. The operator may not get an indication that such a problem is developing until the problem has already occurred. For example, a warning light may indicate that a reach limit has occurred. Such indications do not always warn the operator ahead of time so that he/she can avoid the undesired position, or provide useful cues to tell the operator how to get out of the situation once it has occurred. Our second mode, the Joint Angle Display (JAD) mode was designed to help alleviate some of these control difficulties.

A third area of interest is force feedback through sensory substitution. Force feedback has been shown to be preferable to non-force feedback in many teleoperation studies.[1] However, providing force reflection in the form of a force to the operator's arm and hand muscles can have its disadvantages. Systems that provide force feedback are often bulky master/slave manipulators that are impractical in many environments. Further, presenting force feedback to the operator's hand or arm in the presence of even small time delays has been shown to create operator induced instabilities. The third mode of the MMDS is the Sensory Substitution (SS) mode and addresses these issues.

At the time of the writing of this paper, the MPD mode has undergone testing and is further along in the development cycle than both the JAD mode which is in its initial development and the SS mode which is still in its design phase.

2. Manipulator Position Display (MPD) Mode

The MPD mode provides six degree of freedom hand controller positioning cues to the operator in a graphical format. This mode was designed to help alleviate the problems associated with poor visual feedback caused by obstructions, poor viewing angles, poor resolution, camera control, or camera failure. The MPD mode relies on six degree of freedom information obtained from manipulator sensors, such as joint position encoders, or, if available, a computer based vision system which can calculate current position relative to a target or desired position. The MPD's algorithms perform the necessary calculations and provide the operator with "fly-from" or "fly-to" cues that alleviate from the operator the burden of calculating the appropriate system inputs.[2]

In order to operate effectively, the MPD mode requires knowledge of the current and desired (or target) positions. The current position of the manipulator arm can be obtained through real time position data from the system sensors (encoders or vision system) in six degrees of freedom. The desired position of the arm in six degrees of freedom can be entered into the MPD program if they are known aprioi, or can be obtained from vision system or telemetry data if such data are available. With this knowledge, the MPD mode can present the deviation or error that exists in each degree of freedom to the operator in an easy to use format. The MPD mode not only has applications for the space manipulators, but also for other human-machine applications (aircraft, deep sea manipulators, toxic waste cleanup, etc.) which require the operator to control multi-degree of freedom systems under limited viewing conditions when the desired target points can be identified.

2.1 MPD Mode Display for the SRMS - The Rotational/Translational Display (RTD)

The display to be used by the operator in MPD mode can be tailored to the application area if necessary. As an example, we will examine the Rotational/Translational Display (RTD) which is the MPD mode display designed for astronauts who control the Space Shuttle Remote Manipulator System (SRMS).

Figure 1 shows the format of the RTD.[3] The RTD separates the rotational and translational cues by depicting those cues through the motion of two separate objects. Two separate hand controllers are used by the astronauts to control the SRMS: 1) a Translational Hand Controller (THC) to control all translational motions, and 2) a Rotational Hand Controller (RHC) which is used to control all rotational motions. The RTD was designed so that one object on the display would correlate exclusively to the translational inputs on the THC, while the second object would correlate exclusively to rotational inputs on the RHC. This concept and the format of the display was developed by working directly with several astronauts. An explanation of how the RTD works and a summary of its features follows.

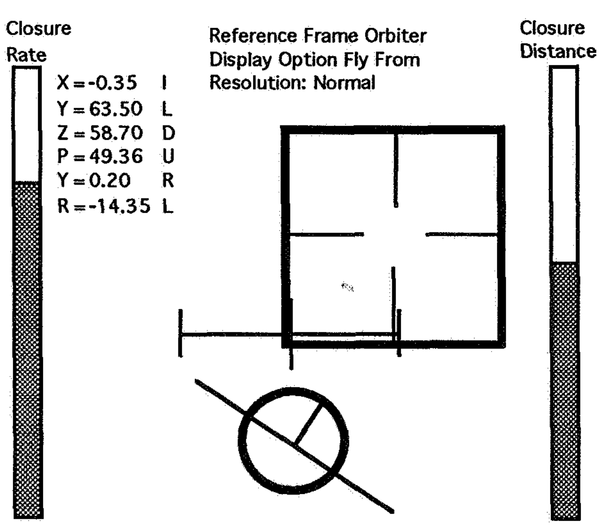


Fig. 1. Rotational/Translational Display (RTD) Format

The line in the center with the three tick marks in Fig. 1 is stationary and acts as the reference line. The operator drives the translational cues using the square with the tick marks shown. Deviation in Z-translation (up an down motion) is depicted by the square being above or below the reference line, while Y-translation deviation (side to side motion) is shown by the square being to the left or right of the center of the reference line. For X-translation (in and out motion), the operator relies on the size of the square relative to the length of the reference line.

For rotational cues the operator looks to the circular object shown in Fig. 1. The position of the circle with respect to the reference line provides the rotational deviation information. If the circle is above or below the reference line, a deviation in pitch exists. A deviation in yaw is depicted by the circle being to the left or right of the center of the reference line. Roll cues are provided by the orientation of the extended line running through the center of the circle and the shorter line in the center of the circle. If those lines are tilted to the left or to the right, then a deviation in roll exists.

On either side of the display are bar graphs which represent the closure rate, on the left, and the closure distance, on the right. These cues become useful when the tip of the manipulator is approaching its final destination. In addition, the operator is provided with a digital readout of the deviations in each of the six degrees of freedom. This digital readout can be seen in the upper left hand corner of Figure 1, and would be helpful in the final stages of a task to ensure that the deviations are within the desired limits (i.e. close to zero).

The RTD also includes a number of other features to provide the operator better assessment of the manipulator's position. One of these features is the highlighting of cues. This feature becomes most useful when the manipulator is reaching its target position and the RTD cues are converging on the stationary reference line. A task will usually have defined tolerance limits for each degree of freedom within which the manipulator is considered to be at its desired final position. Based on this information the highlighting feature indicates to the operator when the manipulator is within the defined limit for each degree of freedom. This indication is achieved by increasing the width of specific lines on the rotational and translational cues. For example, when the position of the manipulator is within the specified range in the X-axis the square becomes bolder than the other lines. When all of the lines which comprise the translational cue are bold, the operator will know that the manipulator tip is within tolerance in the X, Y, and Z axes. The rotational cues work similarly. For example, the circle becomes bold when the manipulator's attitude is within the yaw limit. As with the translational cue, when the manipulator attitude is within limit in yaw, pitch, and roll the entire rotational cue will be bold. Fig. 1 shows an example of the bold feature indicating that the X-axis and the yaw axes are within range. The tolerances can be set to different values for each degree of freedom and for each task.

In addition to the highlighting feature, the RTD display also provides color cues to help distinguish between the translational and rotational cues, and the stationary reference line. The use of color is useful when the manipulator position is close to its final destination and it can be difficult to differentiate between the translational cue, rotational cue, and the reference line. In the current MPD implementation the translational cue is drawn in red, the rotational in green and the reference line in white.

The RTD also displays hand controller Direction Cues which provide the operator with cues for the necessary hand controller deflections. The Direction Cues can be seen in Figure 1 as letters following the deltas in the upper left-hand corner of the display. The letters I or O are used to indicate in or out deflection of the translational hand controller, L or R for left or right deflection of the translational hand controller, and U or D for up or down deflection of the translational hand controller. For the rotational Direction Cues the letters U, D, L , and R are used in the same way as with the translational Direction Cues. With Direction Cues the operator is presented with clear indications of the necessary hand controller deflections eliminating the possibility of unnecessary and potentially dangerous movement of the manipulator.

The RTD also provides the operator with a choice for displaying the cues in fly-from (outside-in) or fly-to (inside-out) formats. [4] At the beginning of each task the operator is given the choice of which convention to use. Once the selection is made, the RTD lists the selection being used in the top center part of the screen as shown in Figure 1.

The RTD also has the capability to select between the different coordinate frames in which the manipulator position and attitude can be commanded. The RTD can be operated in three different coordinate frames: orbiter, end effector, and payload. These choices correspond to the reference frame options for commanding the SRMS on the space shuttle. Once selected the choice is displayed in the top center part of the main display screen above the fly-from fly-to selection (see Figure 1). The choice of coordinate frames can be modified to include any number of frames.

To quantify the effectiveness of the RTD, experiments with human operators were conducted. The RTD was presented to four trained and experienced test subjects on a GRID 1660 laptop computer. A space shuttle SRMS task was simulated using the Manipulator Analysis - Graphic, Interactive, Kinematic (MAGIK)[5] simulation system which runs on Silicon Graphics computers. The task was a space station assembly task, which focused on the installation of a Pressurized Mating Adapter (PMA) to a space station module. The experimental results concluded that using the RTD significantly improved operator performance by 33% over performing the same task without the use of the RTD.[3,6]

The RTD is a useful tool for SRMS operations. Changes can be made to the RTD to tailor an MPD mode display for applications other than SRMS such as undersea, rovers, or toxic waste cleanup.

3. Joint Angle Display Mode

Another area of telerobotics operation where the operator can use assistance is the avoidance of unwanted joint positions such as joint limits or singularities. Reaching such limitations could shut down the system with a software stop (soft stop) or a mechanical hardware stop (hard stop). This type of situation could force the operator to control the manipulator in a single joint mode where each joint must be driven individually to alleviate the problem. The goal of the Joint Angle Display (JAD) is to present the operator with graphical cues which provide information on the current position of each joint relative to software stops and hardware stops.

The JAD is comprised of a set of bar graphs which represent the position of each joint of a manipulator. The bar graphs are updated on a real-time basis using data from the position sensors at each joint. The JAD mode has three submodes: 1) nominal operations, 2) joint limits, and 3) single joint operations.

3.1 Nominal Operations Display Submode

The nominal operations submode display provides the current joint positions to the operator. As can be seen in Figure 2, each joint is identified at the top of each bar graph: SY = shoulder yaw, SP = shoulder pitch, EP = elbow pitch, WY = wrist yaw, WP = wrist pitch, WR = wrist roll. As the position of the joints change the bar graphs are updated keeping the operator informed of the position of each joint. Used in conjunction with the MPD mode display, the JAD provides the operator with sufficient information to keep the manipulator from reaching unwanted joint positions while being driven to its final POR.

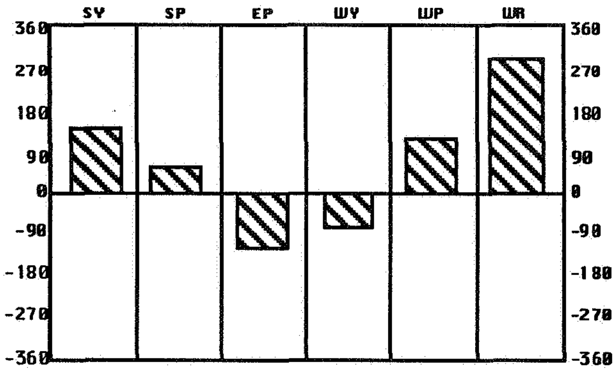


Figure 2. Nominal Operations Joint Angle Display

3.2 Joint Limits Display Submode

The second submode of the JAD includes all the features of the first submode plus cues to indicate the location of the task specific joint limitations. These limits are specific to different tasks being performed by the manipulator and can be used to keep the operator from positioning the manipulator in undesired areas. As can be seen in Figure 3 the joint limits are indicated by the small triangles to the right of each bar graph. Further, when a joint limit is reached the pattern or color of the associated bar graph can also change as an added cue for the operator. This feature eliminates the burden on the operator to recall the limit of each individual joint when trying to identify which joint has reached its limit. In addition, this display can also emit an audible tone when any joint reaches a limit. By including the audible tone the operator will be notified of a joint limit error without having to constantly monitor each joint.

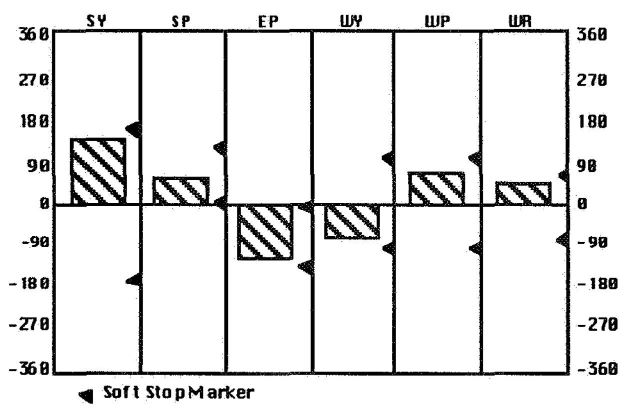


Figure 3. Joint Limits in the Joint Angle Display

3.3 Single Joint Operations Display Submode

Another application for the JAD mode will be single joint operations when the operator needs to drive the arm through a sequence of single joint movements. This operational scenario could occur during failure modes which make controlling all joints concurrently impossible. During these operations, the Single Joint Operations submode will not only provide the operator with information on the current joint positions and joint limits, but will also provide the operator with operational cues. These cues will include the amount of deflection needed for each joint, and the joint sequence. The sequence is shown by highlighting the bar graph associated with the joint to be commanded while the desired position is indicated by a triangle on the left-hand side to the bar graph. Once the joint reaches the desired position its bar graph is displayed normally and the bar graph associated with the next joint in the sequence is highlighted. Figure 4 provides an example of the Single Joint Operations Submode display. In this example the display indicates that the Wrist Pitch joint should be moved to -86 degrees.

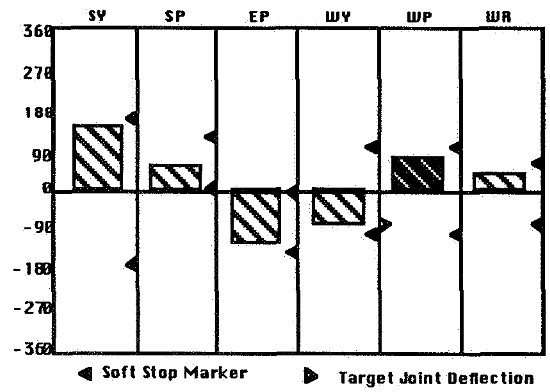


Figure 4. Wrist pitch joint indication.

4. Sensory Substitution Mode

To provide force feedback information to the operator, force reflection is the method used for most master-slave systems. Force reflection has had a long history of success. For example, Hill and Salisbury[7] found in their experiments that with force feedback task completion times were significantly shorter than without force feedback for peg-in-hole tasks. However, providing force reflection in the form of a force to the operator's arm and hand muscles can have its disadvantages. Systems that provide force feedback are often bulky master/slave manipulators that are impractical in many environments. Further, presenting force feedback to the operator's hand or arm in the presence of even small time delays has been shown to create operator induced instabilities. Ferrell[8] suggested that the advantages of force sensitivity could be maintained in the presence of a time delay if the force feedback were substituted through the auditory or tactile modalities, and that a tactile display to the active hand might be especially compatible. The third mode of the MMDS is the Sensory Substitution (SS) mode and addresses these issues.

Bach-y-Rita, Webster, Tompkins, and Crabb[9] define sensory substitution as "the provision to the brain of information that is usually in one sensory domain (for example visual information via the eyes and visual system) by means of the receptors, pathways and brain projection, integrative and interpretative areas of another sensory system, (for example visual information through the skin and somatosensory system). Some examples include sign language for the deaf, and Braille for the blind." Sensory substitution has been successfully used for many years in helping people who are fully or partially deficient in one or more of their sensory systems, for example, sensory aids for the blind or deaf.[10]

4.1 Tactile and Auditory Displays

To provide the sensory substitution information, we are concentrating on tactile and auditory feedback devices for the operator. The tactile, in particular vibrotactile, and auditory modalities are of interest for several reasons. Such displays might be particularly useful for presenting force information because they provided non-reactive representations of force feedback. Non-reactive means sense modalities that do not induce operator movements like force reflection does when providing force information.[11] Such movements may be undesirable in certain situations, and can cause instabilities in the presence of a time delay. They are desirable for generic task information as well, because the auditory and vibrotactile modalities can present information while not placing any extra burden on the operator's visual system which is normally intently viewing the remote task environment via television monitor.

Vibrotactile and auditory displays may also provide cost benefits by reducing the need for expensive bilateral force reflecting manipulators. Further, auditory and vibrotactile displays may also reduce the need for expensive or complicated visual systems. Massimino and Sheridan[1] showed that force feedback could decrease the need for visual feedback, since force feedback combined with low frame rate conditions (3 frames per second) provided performance that was comparable to performance under high frame rate conditions (30 frames per second) without force feedback. In addition, Bliss, Hill, and Wilber[12] concluded that the utility of tactile feedback increased under poor visual conditions, and provided highly useful information that required a relatively low bandwidth channel. Thus a potential benefit of vibrotactile or auditory feedback is a possible reduced need for high quality visual feedback which could lead to decreased cost of teleoperation.

We are currently developing auditory and tactile displays to present manipulator force and position information to the operator. These displays will be incorporated into our multi-mode system to provide the operator with an integrated visual, auditory, and tactile feedback display system.

5. Conclusions

The MMDS can be expected to provide significant operational benefits that include providing the operator with useful manipulator position information when viewing conditions are constrained, assisting with recognizing and avoiding unwanted manipulator position, and providing force information under conditions which would normally make the presentation of such information impractical. The MMDS can also reduce operator workload, reduce training time, and assist the operator with performing unscheduled or unpracticed procedures. The MMDS has space based application for the space shuttle and the space station as well as for ground control of space based manipulators. It is a generic system which can be utilized for dual use application areas such environmental, hazardous waste, nuclear, and undersea remote manipulation environments.

6. References

1. Massimino, M.J., and Sheridan, T.B. Variable Force and Visual Feedback Effects on Teleoperator Man/Machine Performance. Proceedings of the NASA Conference on Space Telerobotics. Pasadena, CA, January 31-February 2, 1989.

2. D.W. Collins, "Payload Deployment and Retrieval System Overview Workbook," NASA NASA TD383, NAS9-18000, NASA Johnson Space Center, Mission Operations Directorate, Space Flight Training Division, Houston, TX, Feb. 1988.

3. M.J. Massimino, M.F. Meschler, and A.A. Rodriguez, "Human-Machine Interface Aids for Space Telerobotics," 44th Congress of the International Astronautical Federation, Graz, Austria, Oct. 1993, paper # IAF/IAA-93-G.3.154.

4. L. R. Young, "Human Control Capabilities," Bioastronautics Data Book, Second Edition, pp 751-806, NASA Scientific and Technical Informarion Office, Washington, D. C., 1973.

5. R.G. Boettger, K.E. Harvey, A.S. Mediavilla, W.C. O'Donnell, "Manipulator Analysis - Graphic, Interactive, Kinematic (MAGIK) Version 5.4 User's Guide", NASA TM-5.24.11-25, NASA Johnson Space Center, Automation and Robotics Division, Houston, TX, Sept. 1992.

6. M. Massimino and M. Meschler, "Experimental Results Using a Manipulator Position Display as a Human-Machine Interface Aid for Controlling Space Robotic Tasks," AIAA Space Programs and Technologies Conference and Exhibit, Huntsville, AL, Sept. 1993, paper # AIAA 93-4114.

7. Hill, J.W. and Salisbury, J.K. Study to Design and Develop Remote Manipulator Systems. Annual Report, NASA Contract NAS2-8652, Nov., 1977.

8. Ferrell, W.R. Delayed Force Feedback. Human Factors. Vol.8, No.5, October, 1966. pp.449-455.

9. Bach-y-Rita, P., Webster, J.G., Tompkins, W.J., and Crabb, T.. Sensory Substitution for Space Gloves and for Space Robots. In Proceedings of Workshop on Space Telerobotics, Vol.2.,pp.51-57.

10. Mann, R.W. Technology and Human Rehabilitation: Prostheses for Sensory Rehabilitation and/or Sensory Substitution. Advances in Biomedical Engineering, Vol. 4, 1974, pp. 209-353.

11. Massimino, M.J. and Sheridan, T.B. Sensory Substitution for Force Feedback in Teleoperation. Analysis, Design and Evaluation of Man-Machine Systems 1992, Selected Papers from the 5th IFAC/IFIP/IFORS/IEA Symposium, The Hague, The Netherlands, June, 1992, H.G. Stassen, Editor, Oxford: Pergamon Press, 1993.

12. Bliss, J.C., Hill, J.W., and Wilber, B.M. Tactile Perception Studies Related to Teleoperator Systems. NASA-CR-1775, Stanford Research Institute, Menlo Park, CA.

omit

Session S1: INFORMATION ACCESS

Session Chair: Chris Culbert

Information Processing & Management Vol. 29, No. 6, pp. 765-774, 1993
Printed in Great Britain.

0306-4573/93 $6.00 + .00

517-54
44813

A NEW MACHINE CLASSIFICATION METHOD APPLIED TO HUMAN PERIPHERAL BLOOD LEUKOCYTES

MARK E. RORVIG
Software Technology Branch, PT4, Lyndon B. Johnson Space Center, National Aeronautics and Space Administration, Houston, TX 77058-3696

STEVEN J. FITZPATRICK
Measurement and Evaluation Center, University of Texas at Austin, Austin, TX 78712-2219

CHARLES T. LADOULIS
Department of Pathology, Mercy General Catholic Hospital, Darby, PA 19023-0000

SANJAY VITTHAL
Reservoir Research and Engineering, Halliburton Services Company, Duncan, OK 73536-0416

Abstract – Human beings judge images by complex mental processes, whereas computing machines extract features. By reducing scaled human judgments and machine extracted features to a common metric space and fitting them by regression, the judgments of human experts rendered on a sample of images may be imposed on an image population to provide automatic classification.

1. INTRODUCTION

Pattern classification of imagery by computational devices is usually approached in two phases. The first phase is the specification of image exemplars representing the classes by an expert as a training set, with a subsequent classification phase occurring as the joining of image features extracted from the target image population with the features similarly extracted from the specified exemplars (Duda & Hart, 1973). Various difficulties arise with these techniques in both phases. For example, in the training phase, the expert's knowledge must be properly decoded to record accurately the salient features used for exemplar classification, a process of recognized difficulty with many pitfalls (Hayes-Roth *et al.*, 1983). Additionally, in the classification phase, information from the expert must often be encoded as specific programs for identification and matching, thus restricting the applicable domain of the algorithm (Young & Fu, 1986). Even the most robust of these methods, the Fisher linear discriminant (where neither the features of the exemplar nor the domain features of the target population of images need be exactly specified) suffers from the noise introduced in exemplars when the expert makes judgments on only a few features of a multi-featured image.

The method described in this paper, however, requires neither explicit decoding of expert judgments nor domain-specific feature matching. Further, it removes from consideration the noise introduced in the Fisher method. This method, called the Two-Domain Method, introduces two unique processes in both the training and classification phases. First, expert knowledge is acquired through multidimensional scaling (Young & Hamer, 1987) of judgments of dissimilarities rendered by an expert on a sample of images from the target population. Second, general pattern features extracted from images of the target population are transformed to points in a Euclidean space. With this method, the problem of image classification is reduced from the complex one of creating machine-based validity rules to the simple matter of creating a linear mapping between two datasets derived from the human domain and the machine domain, respectively.

This paper describes a NASA owned invention (MSC-21737). Inquiries for use may be made to Mr. Hardie Barr, Patent Counsel, Lyndon B. Johnson Space Center, NASA, Houston, TX 77058-3696, telephone 713-483-1003.

2. THE TWO-DOMAIN-METHOD

Consider a collection of images denoted *C*. Let the goal of the expert be to define pairwise dissimilarities among a sample of these images chosen by a random process. These dissimilarities judgments may be collected by presenting all possible pairs of the images in the sample, and asking the expert to place a mark on a line labeled dissimilar at one end and similar at the other. (A ruler applied to these lines establishes a matrix of dissimilarity values among the sampled images.) By processing these judgments in an n-dimensional space using conventional multidimensional scaling (MDS) techniques, a unique, real-valued ordering of these images by their dissimilarity may be produced. Let this ordering be denoted Φ. With this procedure it becomes unnecessary to know explicitly the portions, features, or aspects of the image, or even the deductive rules used by the expert, in rendering the judgments. Whatever features, aspects, or rules the expert may have attended to or employed are already implicit in the ordering, Φ.

Consider again the collection *C*. Let it be assumed that each image in this collection has been digitized and processed so as to extract a number of general, primitive features rendered as histograms. (In the application of this paper, six features are extracted: grey levels; edge intensity; edge slope; line length; line distance from the origin; and angle distance from the origin. No claim is made that these features are the only possible features that might be used, or even that these features are optimal. These features are used only because they are very general, convenient ones.) By converting the histograms for each image into Lorenz information measures (Chang & Yang, 1973), and calculating the Euclidean distance among all pairs of images over all feature measures, a matrix, denoted M, of primitive machine image interpretations may be produced. In this manner, the complex problem of image classification is reduced to the far simpler one of creating a linear mapping of Φ on M.

In this method, the mapping is performed by extracting from *C* the original machine measures matching the subset of *C* judged by the human expert, calculating Euclidean distances for both machine measurements and human coordinates, deriving weights, β, by multiple regression (where the Euclidean distances from the MDS solution for the human judgments are the dependent variable and the Euclidean distances among images based on machine measurements are the independent variable), and multiplying M by β. By resubmitting the predicted values to the multidimensional scaling process, the final ordering is produced, segregated into classes in an n-dimensional space. Let this last result be denoted Φ'. The complete procedure is displayed as a diagram in Fig. 1, with an example of the complete calculations used in the application below available in Appendix A.

3. AN APPLICATION OF THE TWO-DOMAIN-METHOD TO THE CLASSIFICATION OF TWO POPULATIONS OF HUMAN PERIPHERAL BLOOD LEUKOCYTES

In this article, we have chosen to apply the Two-Domain Method to a problem of discriminating two populations of microscopic images of circulating human white blood cells (leukocytes).

Specifically, the Two-Domain Method was tested for its power to discriminate two distinct patterns of human blood leukocyte distribution: an abnormal pattern associated with acute liver failure exhibiting abnormal circulating white blood cell frequency and distribution (Subject 1), and a normal pattern from a normal, healthy subject (Subject 2).

Circulating human leukocytes were separated by flotation from red blood cells by a standard flotation method, and uniform monolayer films prepared and cytochemically stained by a routine clinical laboratory automated instrument using hematoxylin and eosin dyes. The resulting slides therefore include all nucleated circulating white blood cells, predominantly neutrophils, eosinophils, lymphocytes, and monocytes, as well as platelets.

Eight representative sample fields were selected for each subject. The photographic recording was standardized using one continuous film strip of Ektachrome color reversal film rated at ASA 200. All slides were photographed at the same magnification. Effects of exposure variations and background density were tested in the Two-Domain Method

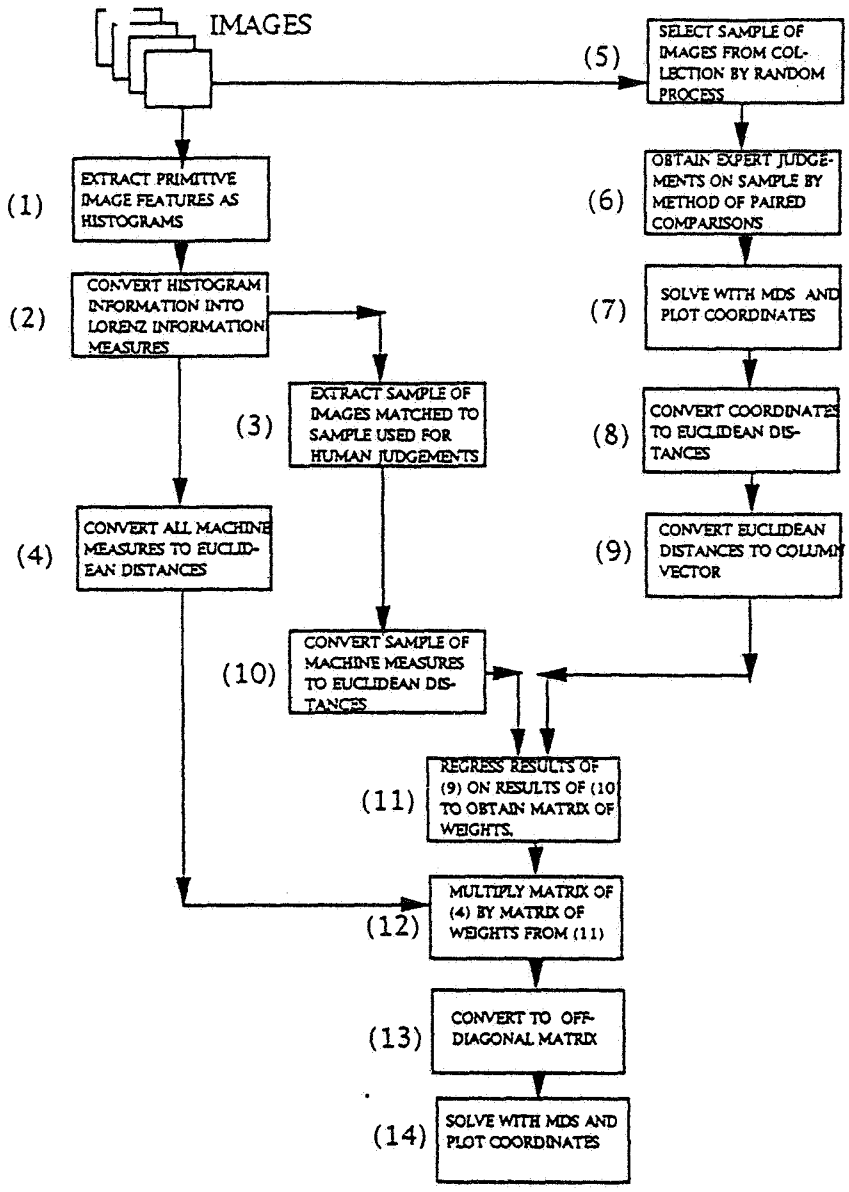


Fig. 1. The procedural steps necessary to execute the Two-Domain Method for any collection of images. Major formulas (exclusive of those used in general MDS (multidimensional scaling) procedures as applied in the application on human peripheral blood may be found in Appendix A.

by recording each image at two different exposures. Set A images (numbered 1–16) were exposed at ASA 200, and Set B images (numbered 17–32) were exposed at ASA 400. Samples used in the test thus consisted of 16 images from each subject, at two levels of exposure, on the same photographic film strip.

The difference in exposure levels substantially alters the machine measurements of these images, and is typical of problems that confound image pattern classification generally, in that "noise" artificially introduced by one element or another distort the machine classification algorithms. Reproductions of both Set A and Set B are presented following the Appendix. The purpose of this application is thus to demonstrate that the Two-Domain Method is sufficiently robust not only to properly classify Set A (by segregation in an n-dimensional space), but also to reduce or eliminate the noise introduced by the difference in Set B film exposure levels.

Expert judgments of dissimilarities were made by an experienced pathologist (C.T.L.), primarily on the basis of the segmentation of leukocyte nuclei, and lymphocyte and monocyte shape and size. Other cell types present in the images were ignored for judgment purposes. Judgments were provided in a single session on slides 1–8 of Set A according to the procedure described in Section 2, and submitted (as are all datasets discussed in this section) to the ALSCAL procedure in SAS, a common multidimensional scaling package.

In Fig. 2, Plots 1 and 2 exhibit a strong separation between the cell populations of the two subjects. The primitive machine interpretations derived from both Set A and Set B, scaled by ALSCAL, appear in Fig. 3 as Plots 3 and 4, respectively.

The images represented by datapoints in Plot 3 appear to have some natural clustering tendency along the same lines as those provided directly by human judgments, proba-

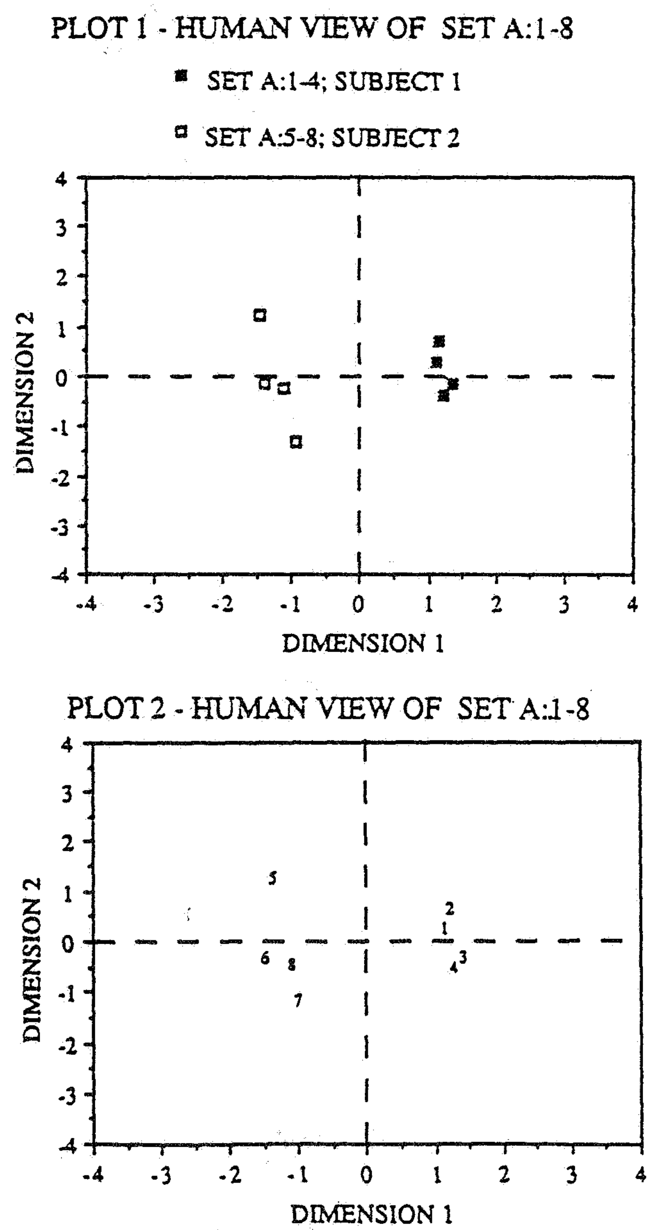


Fig. 2. MDS ALSCAL plots of the original human view of a sample of eight images of peripheral white blood cells. The human judgments were collected through the method of paired comparisons, and show a clear separation between the slides from Subject 1 and Subject 2.

2. THE TWO-DOMAIN-METHOD

Consider a collection of images denoted C. Let the goal of the expert be to define pairwise dissimilarities among a sample of these images chosen by a random process. These dissimilarities judgments may be collected by presenting all possible pairs of the images in the sample, and asking the expert to place a mark on a line labeled dissimilar at one end and similar at the other. (A ruler applied to these lines establishes a matrix of dissimilarity values among the sampled images.) By processing these judgments in an n-dimensional space using conventional multidimensional scaling (MDS) techniques, a unique, real-valued ordering of these images by their dissimilarity may be produced. Let this ordering be denoted Φ. With this procedure it becomes unnecessary to know explicitly the portions, features, or aspects of the image, or even the deductive rules used by the expert, in rendering the judgments. Whatever features, aspects, or rules the expert may have attended to or employed are already implicit in the ordering, Φ.

Consider again the collection C. Let it be assumed that each image in this collection has been digitized and processed so as to extract a number of general, primitive features rendered as histograms. (In the application of this paper, six features are extracted: grey levels; edge intensity; edge slope; line length; line distance from the origin; and angle distance from the origin. No claim is made that these features are the only possible features that might be used, or even that these features are optimal. These features are used only because they are very general, convenient ones.) By converting the histograms for each image into Lorenz information measures (Chang & Yang, 1973), and calculating the Euclidean distance among all pairs of images over all feature measures, a matrix, denoted M, of primitive machine image interpretations may be produced. In this manner, the complex problem of image classification is reduced to the far simpler one of creating a linear mapping of Φ on M.

In this method, the mapping is performed by extracting from C the original machine measures matching the subset of C judged by the human expert, calculating Euclidean distances for both machine measurements and human coordinates, deriving weights, β, by multiple regression (where the Euclidean distances from the MDS solution for the human judgments are the dependent variable and the Euclidean distances among images based on machine measurements are the independent variable), and multiplying M by β. By resubmitting the predicted values to the multidimensional scaling process, the final ordering is produced, segregated into classes in an n-dimensional space. Let this last result be denoted Φ'. The complete procedure is displayed as a diagram in Fig. 1, with an example of the complete calculations used in the application below available in Appendix A.

3. AN APPLICATION OF THE TWO-DOMAIN-METHOD TO THE CLASSIFICATION OF TWO POPULATIONS OF HUMAN PERIPHERAL BLOOD LEUKOCYTES

In this article, we have chosen to apply the Two-Domain Method to a problem of discriminating two populations of microscopic images of circulating human white blood cells (leukocytes).

Specifically, the Two-Domain Method was tested for its power to discriminate two distinct patterns of human blood leukocyte distribution: an abnormal pattern associated with acute liver failure exhibiting abnormal circulating white blood cell frequency and distribution (Subject 1), and a normal pattern from a normal, healthy subject (Subject 2).

Circulating human leukocytes were separated by flotation from red blood cells by a standard flotation method, and uniform monolayer films prepared and cytochemically stained by a routine clinical laboratory automated instrument using hematoxylin and eosin dyes. The resulting slides therefore include all nucleated circulating white blood cells, predominantly neutrophils, eosinophils, lymphocytes, and monocytes, as well as platelets.

Eight representative sample fields were selected for each subject. The photographic recording was standardized using one continuous film strip of Ektachrome color reversal film rated at ASA 200. All slides were photographed at the same magnification. Effects of exposure variations and background density were tested in the Two-Domain Method

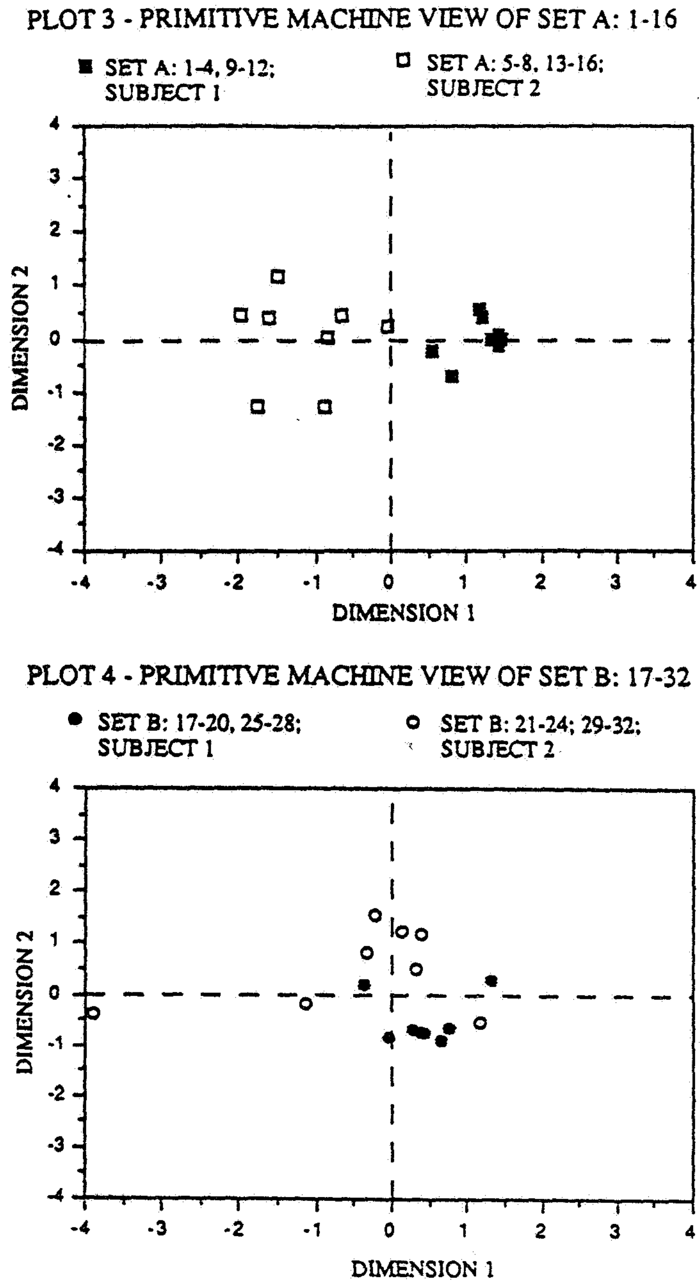
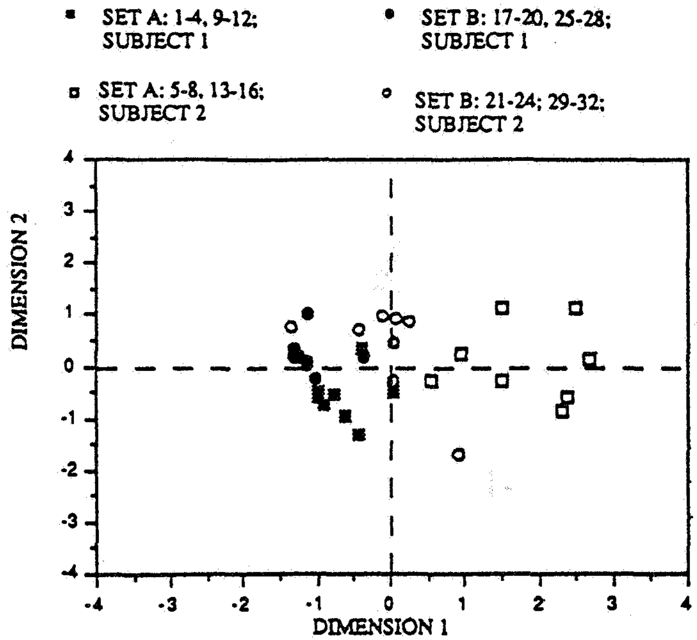


Fig. 3. MDS ALSCAL plots of the primitive machine views of Set A and Set B, including Subject 1 and Subject 2. Plot 3 of Set A (from film rated at ASA 200 and exposed at ASA 200) exhibits some natural clustering by machine features alone, whereas Plot 4 of Set B (from film rated at ASA 200 but exposed at ASA 400) exhibits little machine differentiation between the two subjects.

bly due to the increased light levels in the images produced from Subject 1 and caused by the generally lower levels of white blood cells in the sample drawn from that subject. Plot 4, on the other hand, derived from the deliberately overexposed images, reveals very little meaningful segregation.

In Fig. 4, Plot 5 reveals the strong confounding effect of the Set B data when combined with Set A and scaled together. When the sets are combined, each item acts to influence the scale value of every other item, so that the pure machine view, or interpretation, of these images becomes extremely confused. There is, for example, some segregation of

PLOT 5 - PRIMITIVE MACHINE VIEW: SETS A AND B



PLOT 6 - HUMAN WEIGHTED MACHINE VIEW: SETS A AND B

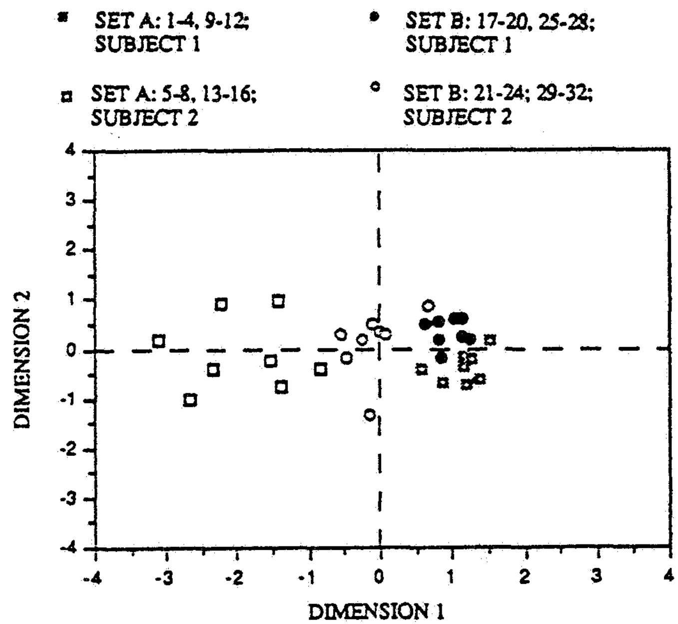


Fig. 4. MDS ALSCAL plots of Sets A and B, Subjects 1 and 2. Plot 5 exhibits distortion of the natural clustering effect displayed in Set A of Plot 3 when Set A and B are combined. Plot 6 exhibits the reordering of Subject 1 and Subject 2 classes when weighted by the human view and displayed in Plot 1. Numbered displays of these datapoints are available in Fig. 7, Plots 7 and 8, in Appendix A.

Subject 1 and Subject 2, but still much less than that appearing in the human classification of these images provided in Plot 1.

Plot 6 exhibits the effect of the Two-Domain Method on the disordered data of Plot 5. Plot 6 was produced according to the procedures of Fig. 1 with the detailed calculations provided in Appendix A. In Plot 6, Subject 1 and Subject 2 data are perfectly segregated

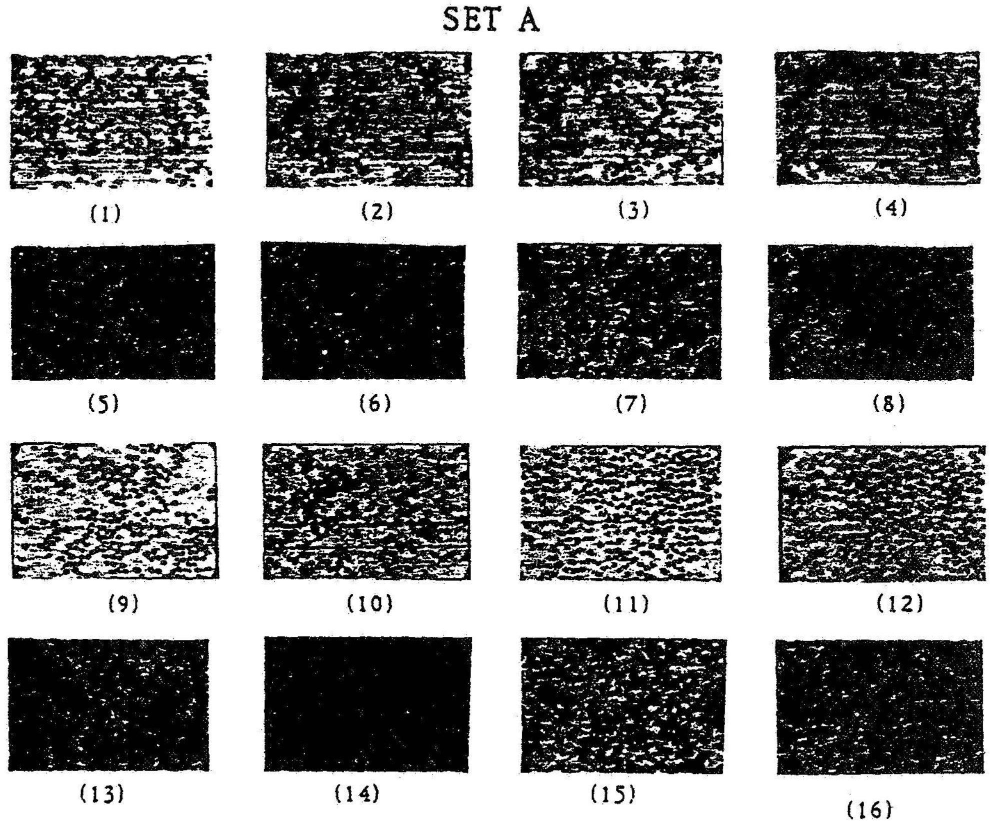


Fig. 5. Reproductions of Set A of 16 slides of peripheral blood cells used in this experiment. All slides were prepared by the same machine-assisted process. Slides 1–4 and 9–12 were extracted from a traumatized subject. Slides 5–8 and 13–16 were extracted from a normal subject. The film used to create the slides was rated ASA 200 and exposed at ASA 200.

for Set A (Fig. 5), and, with the exception of one image, also perfectly segregated for Set B (Fig. 6). Clearly, the strong, confounding effect introduced by combining Set B with Set A images is eliminated.

4. DISCUSSION

The Two-Domain Method, considered very generally, is effective simply because it reduces the intense machine activity associated with image pattern matching to the simple operations of interval scale value relations. Moreover, the scaling theory underlying the method is easily transferable to operations involving classifications among higher dimensions. Indeed, multidimensional scaling has, for some time, been more often used to record human judgments in higher dimensions for a variety of marketing applications (Green & Carmone, 1969). Finally, by using replicated multidimensional scaling methods, the opinions of multiple experts (as opposed to the single expert used in this application) may be combined in the creation of Φ.

The Two-Domain Method is also applicable to image classification systems that routinely use Bayesian methods. In this case, the operations of the Bayesian classifiers would use, as their inputs, the dissimilarity values output from multidimensional scaling matrix transforms, ignoring the plotted values (which are derived from the dissimilarity values anyway.) Along these same lines, the Two-Domain Method may facilitate neural net image classification, both by making the net more efficient due to the reduction of information that must be submitted (dissimilarities or Euclidean distances rather than vectors of pixel values) and by the increased rigor of the training set expression, which reduces noise when aspects of images are judged rather than images as wholes.

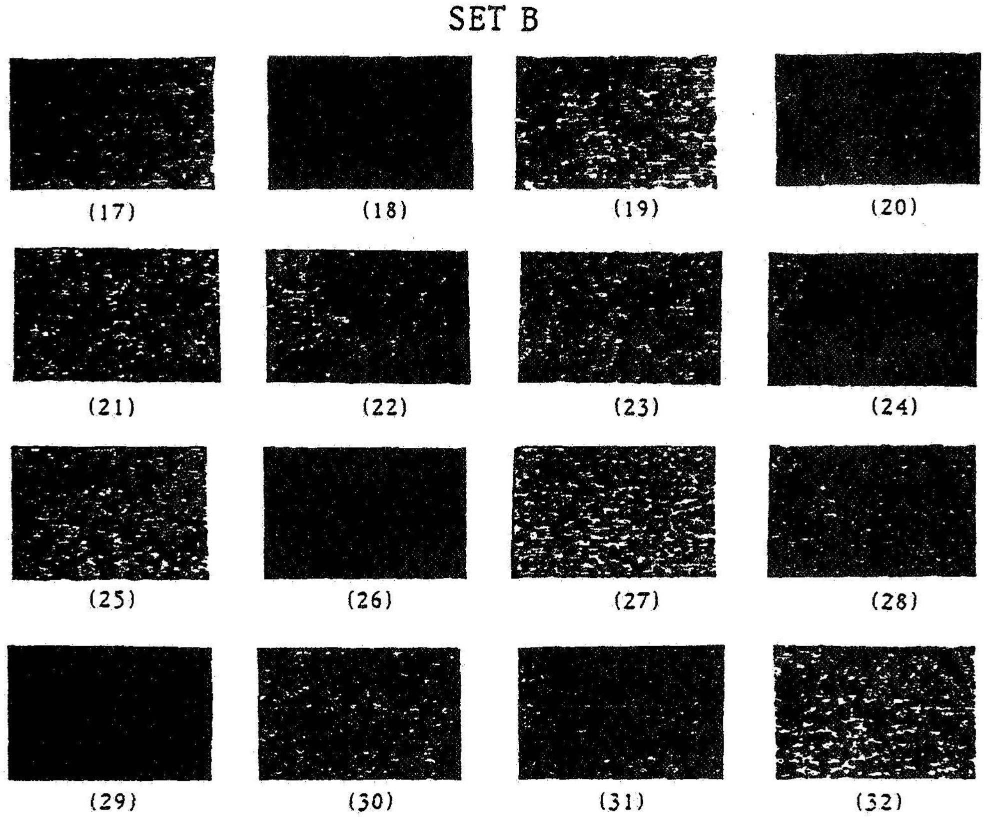


Fig. 6. Reproductions of Set B of 16 slides of peripheral blood cells used in this experiment. All slides were prepared by the same machine-assisted process. Slides 17–20 and 25–28 were extracted from a traumatized subject. Slides 21–24 and 29–32 were extracted from a normal subject. The film used to create the slides was rated ASA 200 and exposed at ASA 400.

Finally, as expressed in some earlier work (Rorvig, 1988), the Two-Domain Method may be used in the searching of large databases of images, where image representations are stored as feature components (Chang & Yang, 1983). In this application, the method would be applied to image classes iteratively, by segregating and mapping successively smaller classes of imagery. This application may be critical in locating desired sets of images that cannot be described linguistically because of either intellectual or economic constraints.

REFERENCES

Chang, S.K., & Yang, C.C. (1983). Picture information measures for similarity retrieval. *Computer Vision, Graphics, and Image Processing, 23*, 366–375.

Duda, R.O., & Hart, P.E. (1973). *Pattern classification and scene analysis.* New York: John Wiley.

Green, P.E., & Carmone, F.J. (1969). Multidimensional scaling: An introduction and comparison of nonmetric unfolding techniques. *Journal of Marketing Research, 6*, 330–341.

Hayes-Roth, F., Waterman, D.A., & Lenat, D.B. (1983). *Building expert systems.* Reading, MA: Addison-Wesley.

Rorvig, M.E. (1988). Psychometric measurement and information retrieval. In M.E. Williams (Ed.), *Annual Review of Information Science and Technology, 23*, 157–189.

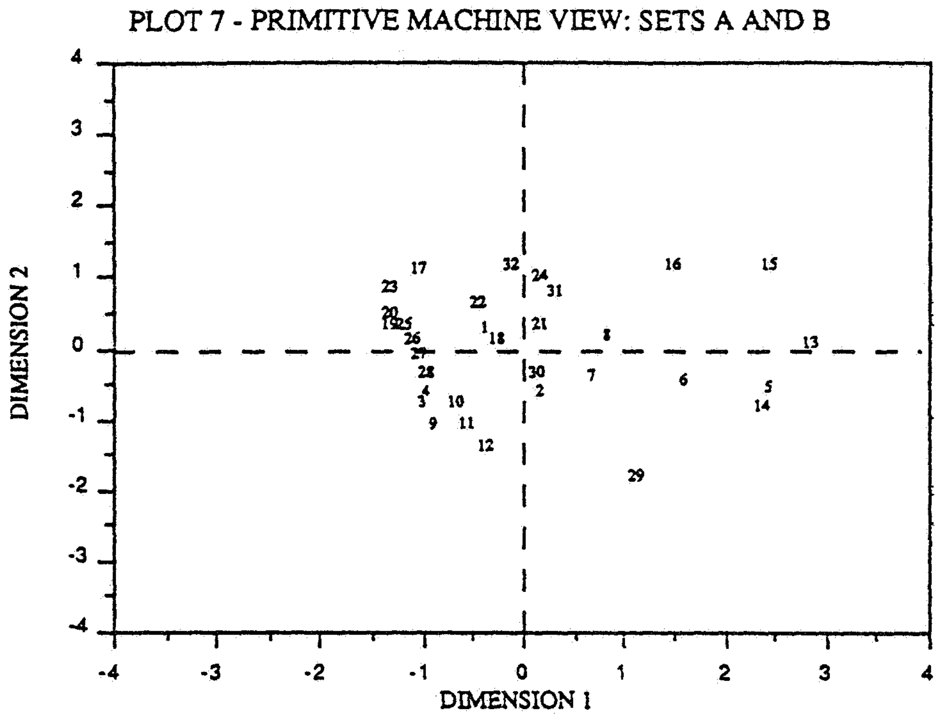
Young, F.W., & Hamer, R.M. (1987). *Multidimensional scaling: History, theory, and applications.* Hillsdale, NJ: Lawrence Erlbaum.

Young, T.Y., & Fu, K.-S. (1986). Part I. Pattern recognition. In *Handbook of pattern recognition and image processing* (pp. 3-167). Orlando, FL: Academic Press.

APPENDIX A

The details of Plot 6 production follow. First, the primitive machine measurements (Lorenz information measures (Chang & Yang, pp. 369–370)) for images 17–24 corresponding to the human judgments rendered on Set A for images 1–8 were converted to six sets of squared Euclidean distances (one for each machine measurement) according to the following equation:

$$Q_k = (p_{ik} - p_{jk})^2; \quad i < j, \; k = 1,6 \qquad (1)$$



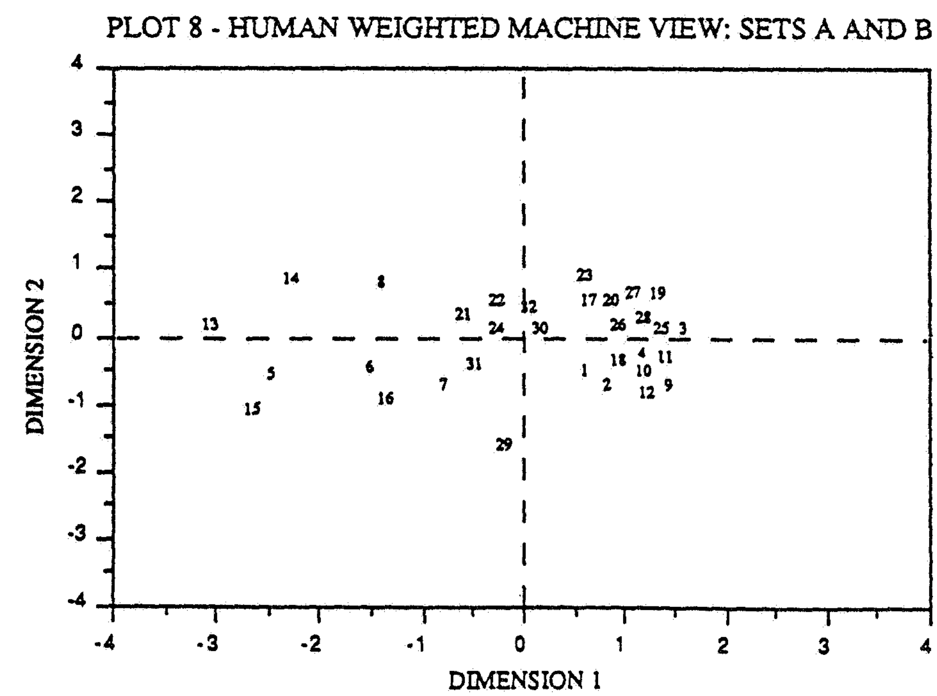


Fig. 7. MDS ALSCAL plots (in numbered display) of both primitive and human weighted views of 32 peripheral blood cell slides. The lower plot exhibits the substantial "learning" effect created by imposition of human judgments on machine interpretations.

where

Q = a matrix of 28×6,
Q_k = a column of matrix Q,
p = a matrix of 8×6,
p_{ik} = the machine measurement k for image i, and
p_{jk} = the machine measurement k for image j.

Since a column of Q contains the squared difference between all pairs of images on the corresponding machine measurements, there are $[n(n-1)]/2$ elements in each column, where n is the number of images.

Second, the squared Euclidean distances between all pairs of slides 1–8 of Set A, that is, Φ, were computed from the spatial coordinates of the MDS solution for the human judgments of Plot 1 according to eqn. 2:

$$D = \sum_{k}^{r} (x_{ik} - x_{jk})^2, \quad i < j,\ k = 1, r \tag{2}$$

where

D = the square symmetric matrix,
x_{ik} = the coordinate of image i on dimension k,
x_{jk} = the coordinate of image j on dimension k, and
r = the number of dimensions in the solution.

Third, the square symmetric matrix was converted to a column vector containing the top off-diagonal elements (for convenience also denoted D) and regressed on the matrix Q of eqn. 1 to produce the vector of weights, β. Equation 3 is the multiple regression equation in standard form and eqn. 4 is the standard least squares solution.

$$D = Q\beta' \tag{3}$$

$$\beta = (Q'Q)^{-1Q'D} \tag{4}$$

Fourth, the procedure of eqn. 1 was applied to all machine data, images 1-32, denoted M, and multiplied by the vector of weights, β, or

$$\mathbf{V} = M\beta' \tag{5}$$

where

$\mathbf{V}$ = the final vector converted to an off-diagonal matrix for submission to MDS, and
M = the 496×6 matrix from the procedure of eqn. 1.

$\mathbf{V}$, submitted to MDS and scaled, thus results in Φ' as displayed in Plot 6.

5/8-82
421816

SPATIAL DATA MANAGEMENT SYSTEM (SDMS)

Mark W. Hutchison
hutch@algol.jsc.nasa.gov

Science Applications International Corporation, Comsystems
17049 El Camino Real, Suite 202, Houston, Texas 77058
NASA/Johnson Space Center
Software Technology Branch
Information Science and Technology Group

Abstract:
The Spatial Data Management System (SDMS) is a testbed for retrieval and display of spatially related material. SDMS permits the linkage of large graphical display objects with detail displays and explanations of its smaller components.

SDMS combines UNIX workstations, MIT's X Window system, TCP/IP and WAIS information retrieval technology to prototype a means of associating aggregate data linked via spatial orientation. SDMS capitalizes upon and extends previous accomplishments of the Software Technology Branch in the area of Virtual Reality and Automated Library Systems.

1. INTRODUCTION

In early 1993, training systems supporting the Hubble Space Telescope rescue mission were being engineered and deployed throughout the Johnson Space Center and divisions within NASA began augmention of training for flight personnel and the Hubble flight crew. To accomplish the training, detailed familiarization with the vast array of pieces ranging from large assemblies to tiny subcomponents, their location on the satellite and their orientation was required. Spatial Data Management System (SDMS) is a prototype conceived to fulfill the need for spatial linkage of large graphical images with component images and text. The ISAT team's recent experience in the production of an automated electronic library system (NELS [1]) suggested a novel approach to the solution of this problem.

The approach was to couple previous work combining large images of the satellite from a modest distance with text, and highly detailed photographs of the bays and equipment with the recently completed alpha version of an automated library system -- thereby providing a combination macro/micro view with automated information retrieval of text.

Previous coding furnished by Mr. Lac Nguyen [2] in conjunction with his virtual reality project provided the visual "macro" views of the satellite, while high-detail, high-quality photos of components were obtained from Van Steinburg [3] for "micro" views of the satellite and subcomponents. The text portion of the system was composed from detailed debriefings of the crew who originally placed the satellite in orbit.

The virtual reality mockup images of the Hubble Space Telescope were joined with high quality photographs of the telescope and the descriptive text recorded from deployment crew debriefings and features of the X Window System. This union was accomplished using Wide Area Information Service (WAIS) as the information retrieval engine. Detail photographs consisted of complete spans of all bays, compartments and components of the telescope.

2. METHODOLOGY

The initial framework of "macro" views of the satellite provide the user with a profile view of the satellite from a generous positive Y-distance, and a small control panel. The control panel consists of the following buttons and associated actions:

- Prev - Move to previous macro image
- Next - Move to next macro image
- Quit - Exit program
- Help - Provide context sensitive help

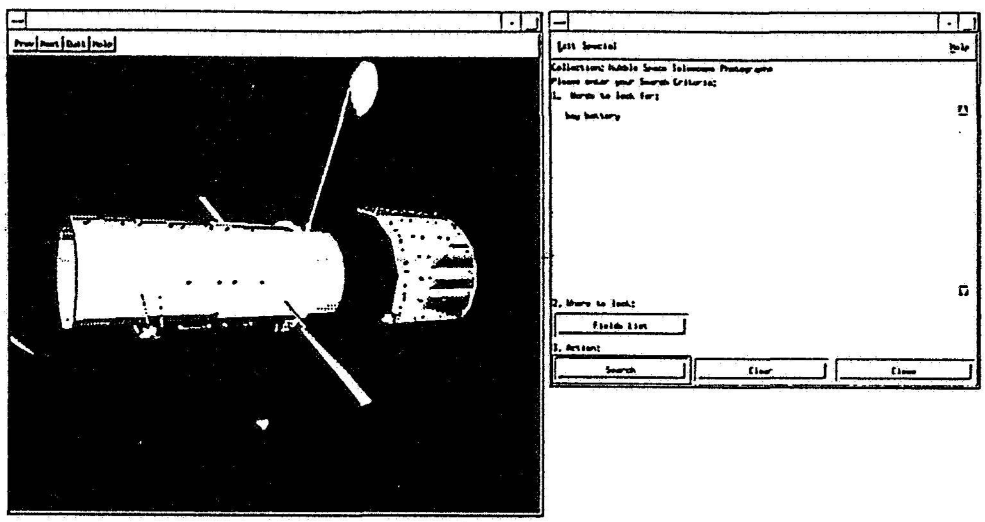


Figure 1.
A profile of the Hubble. The spacecraft may be rotated about its longitudinal axis, permitting views of various bays and subcomponents.

From this vantage point, the user may rotate the satellite about its longitudinal axis a full 360 degrees by manipulation of forward/backward buttons available on the control panel. The motion is accomplished by a rapid projection sequence of a previously processed set of X Window images, providing the user with a smooth transition between views. The number of views has alternately been coded at 4 views (90 degrees) and 6 views (60 degrees). These parameters are purely platform-dependent, as the processes involved are very memory-intensive.

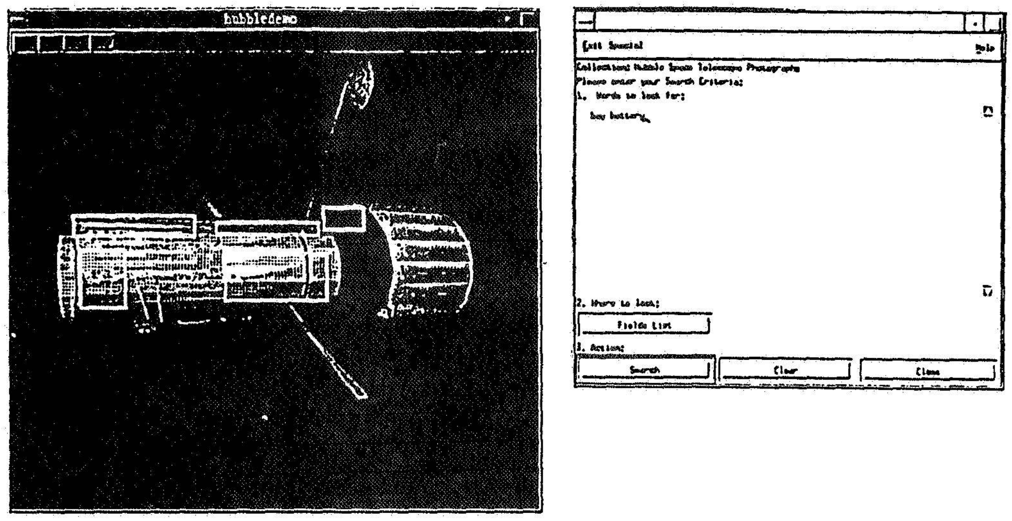


Figure 2.
Bays and subcomponents are highlighted by the right mouse button. Clicking upon the regions activates an independent viewer which brings up a a highly detailed view of selected component.

The next feature is activated by pressing the right mouse button. (See figure 2.) When this button is pressed, the image displays "outlines" of regions on the satellite which are mouse-sensitive. By clicking upon one of the sensitive regions with the left mouse button, the user activates an independent pop-up viewer, providing a close-up "micro" view of the desired component. Text associated with the specific region is available through a NELS/WAIS query presently. This must be operated by hand in the prototype, but would be automated in a full implementation of the project. Access to spatially related items, their metadata, or further queries are available via the familiar NELS/WAIS interface.

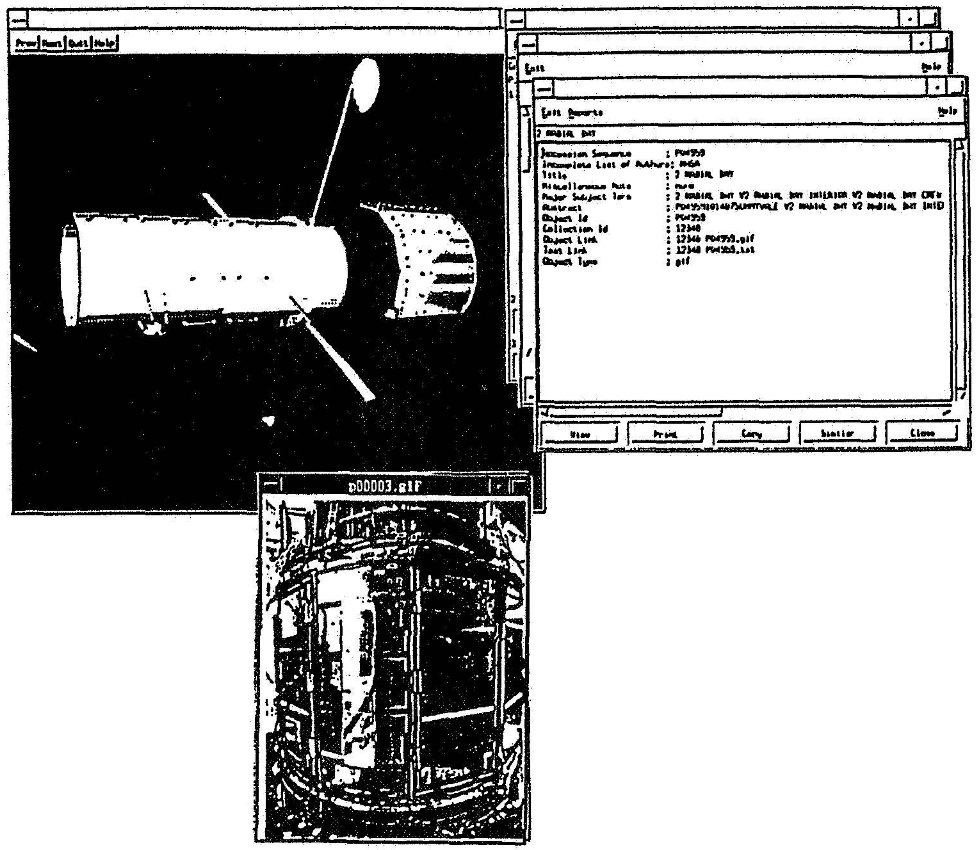


Figure 3.
Both a high quality, detailed photograph and textual information corresponding to the selected component have been triggered by the mouse and displayed.

(See figure 3.) Selecting a desired component activates an independent viewer and retrieves text providing details of the component. Services provided in WAIS and NELS such as printing, copying and reporting features are also available.

3. FEATURES OF THIS PROTOTYPE

- Hotspots - Regions of the display are sensitive to activation via mouse buttons.
- Plus/minus rotation about longitudinal axis - The spacecraft may be spun about its long axis through by selecting the "Prev" and "Next" buttons.
- Basic aural - Basic sound capabilities have been added in the form of sound files for contextual help.
- Basic spatial - Rudimentary spatial orientation is accomplished thru WAIS.

- Distributed information via WAIS - Distributed text, image, sound and graphical user interface are provided through client/server model.

4. AREAS FOR FUTURE ENHANCEMENT

- Full integration of sound - Training may be enhanced with aural components added to the present textual presentation of information.
- Integration of motion/video - Full motion video could be applied to the detail image presentation.
- Mosaic/Wide World-Web interface - Integrating the current spread of Mosaic/HTML technology into the interface may produce an interactive teaching tool capable of accepting text, multi-answer questions, or other forms of computer based training.
- Generalized 3 dimensional view - Permit the user to translate, or navigate around the satellite rather than providing a fixed view from a specific orientation.

5. CONCLUSIONS

Demonstrations of SDMS astonish and fascinate its viewers. The incorporation of sound, or other multimedia and the completion of several other features would make it a most appropriate and adaptive training tool suited for linkage of graphical objects with component objects.

6. ACKNOWLEDGMENTS

The work described was supported by the Software Technology Branch of the Johnson Space Center. Dr. Mark E. Rorvig served as scientific monitor of the project and provided the motivation and stimulus for the broad combination of emerging technologies. Mr. Lac Nguyen of I-NET provided invaluable help and experience with the adaptation of his original model for the purposes of this prototype, design issues and the "everpresent" coding bug. Mr. Terry McGregor of I-NET participated in early design and algorithm issues, and many thanks to Ms. Stephanie Smith of Hernandez Engineering who championed user interface design issues and who diligently demonstrates this prototype.

7. REFERENCES

[1] NASA Electronic Library System (NELS) 2.0 Product Specification, Software Technology Branch, NASA/Johnson Space Center, September 1993.

[2] Virtual reality programming in progress by Mr. Lac Nguyen of Inet, Inc. in support of the NASA/JSC Software Technology Branch. Fall 1992.

[3] High quality targa photographic images provided by Michael Van Steinburg, in support of Goddard Space Flight Center, Maryland, the Hubble Space Telescope project, and the Center for AeroSpace Information. Spring 1991.

[4] The Wide Area Information Servers (WAIS) doucmentation. Public domain distribution of WAIS, version 8-b5. April, 1991.

8. GLOSSARY

COSMIC: NASA's COmputer Software Management and Information Center (COSMIC). COSMIC controls the dissemination of NASA software available to the public.

HST: Hubble Space Telescope.

ISAT: Information Science and Technology team/group. The team of engineers employed by the Software Technology Branch of the NASA/Johnson Space center who developed NELS and SDMS.

MIT: Massachusetts Institute of Technology.

Mosaic: An Internet information browser and World Wide Web client developed at NCSA.

NASA: National Aeronautics and Space Administration.

NELS: NELS is the NASA Electronic Library System, is an information management tool for creating distributed repositories of documents, drawings, and code for use and reuse by the aerospace and other communities. The NELS retrieval engine loads metadata and source files of full text objects, perform natural language queries to retrieve ranked objects, and create links to connect user interfaces.

NELS is designed for use on heterogeneous hardware and software systems, which may locally or remotely. Some of the key features of NELS are network operation, natural language support, object retrieval, interface.

NCSA: National Center for Supercomputing Applications. NCSA provides services and facilities to a variety of users and institutions. These users are involved in computational science research that complements the traditional methods of laboratory experimentation and theoretical investigation. Computational science allows researchers to recreate numerically, or simulate, natural phenomena on a high-performance computer. In many cases, such phenomena cannot be investigated in the laboratory or fully evaluated theoretically due to such constraints as safety, cost, speed, or time. In other cases, the supercomputer is being coupled directly to laboratory instruments or observational facilities for extensive data computing. The federal High Performance Computing and Communications Program, as well as the communications technology initiatives envisioned by the current administration, support the creation of a permanent national infrastructure for high-performance computing. NCSA's direction reflects these national initiatives.

SDMS: Spatial Data Management System The subject of this technical paper.

STB: Software Technology Branch of the NASA/Johnson Space Center (PT4)

The Software Technology Branch is the home of the Software Technology Laboratory which demonstrates, evaluates, prototypes and develops new software technologies. Branch software is available to the NASA/Contractor community from PT4 and to the general public through the Computer Software Management and Information Center (COSMIC). Source code and documentation for programs may be purchased at a nominal fee for unlimited copies with no royalties.

TCP/IP: Transmission Control Protocol/Internet Protocol

Targa: One of the many electronic picture formats. Noted for high definition and color.

UNIX: Strange arcane operating system originally developed by Bell Laboratories.

Virtual Reality: An alternate reality provided by electromechanical means.

WAIS: Wide Area Information Service. WAIS is a product to help end-users find and retrieve information over networks.

WWW: Wide World Web.

A Method for Automatically Abstracting Visual Documents*

Mark E. Rorvig
Software Technology Branch, Lyndon B. Johnson Space Center, National Aeronautics and Space Administration, Houston, TX 77058

Visual documents—motion sequences on film, videotape, and digital recordings—constitute a major source of information for the Space Agency, as well as all other government and private sector entities. This article describes a method for automatically selecting key frames from visual documents. These frames may in turn be used to represent the total image sequence of visual documents in visual libraries, hypermedia systems, and training guides. The performance of the abstracting algorithm reduces 51 minutes of video sequences to 134 frames; a reduction of information in the range of 700:1.

Introduction

Although the application of visual documentation techniques has been expanded manyfold in the last decade due to steady reductions in cost, methods for summarizing these documents have remained bound by human editing procedures. Such procedures are typically subject to high costs as well as variations and biases introduced by individual editors possessing different training backgrounds and aesthetic temperaments (Pryluck, Teddlie, & Sands, 1982). While significant work has been done in identifying sources of descriptive information for visual documents (O'Connor, 1985, 1986), it is curious that no work has been done to abstract or index visual documents with visual exemplars directly. Indeed, the most closely related work has been conducted around the problem of data compression algorithms (Yeh et al., 1991).

In the method introduced in this article, however, frames of the visual document are digitized and subjected to a structural decomposition process that reduces all information in the image to sets of values. These values are in turn normalized, further combined to produce only one value per frame, and fitted to a normal distribution of all values in a defined training set of frames. By selecting only those values at specified areas at the tails of the distribution, key frame images may be abstracted from their surrounding frames.

Methodology

Consider a visual document composed of 30 frames of interleaved video or film per second as a sampling universe. For each second, little change in any frame occurs such that, in the method of this article, a sampling rate of one frame of video imagery per every 5 seconds constitutes the sampling frame. The problem of visual abstraction thus devolves into the determination of a method for selecting significant frames from among the reduced set extracted from the original run of frames. (It should be noted that, in some cases, a higher density of sampling may be preferred. The interval of 5 seconds was chosen arbitrarily for the demonstration of this method. No claim is made for any optimal sampling rate.) The collection of images thus sampled from the visual document shall be denoted "C."

Consider the collection of images C. Assume that each image in this collection has been digitized and processed so as to extract a number of general, primitive features rendered as histograms. Specifically, in the demonstration of this article, NTSC standard VHS video frames were sampled at intervals of 5 seconds each, digitized by a commercially available analog frame digitizer and stored as PICT format files with a common *XY* dimension. Further, although Figure 1 suggests the use of hue, chroma, and

*This work was performed under the terms and conditions of the *Memorandum of Understanding Between the National Aeronautics and Space Administration Lyndon B. Johnson Space Center (JSC) and the University of Texas at Austin* as signed and dated by the authorities of the respective institutions on March 26, 1991 and March 13, 1991 and transmitted by NASA JSC Reply Reference AL4-91-105. The method described in this article is a NASA-owned invention (MSC-22093-1). Inquiries for use may be made to Mr. Hardie Barr, Patent Counsel, NASA Johnson Space Center, AL3, Houston, TX 77058; tel.: (713) 483-1003.

Received May 6, 1992; revised September 4, 1992; accepted September 4, 1992.

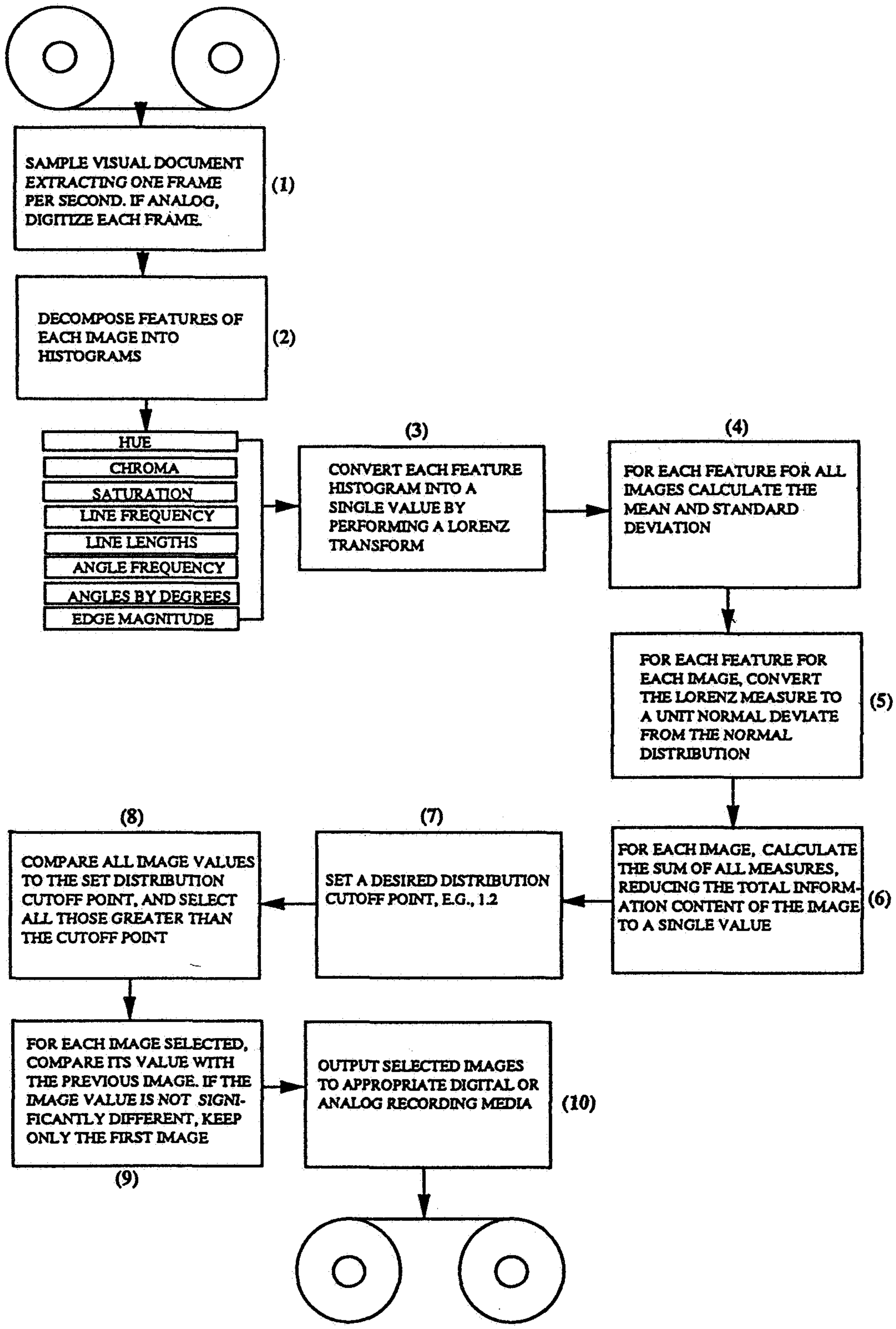


FIG. 1. The complete method for automatic abstraction of visual documents, consisting of 10 steps.

saturation, for this demonstration only grey levels were extracted. Other extracted features were edge intensity, edge slope, line length, line distance from the origin, and angles. The feature extraction programs were written in C language. The pixel intensity histogram accumulated grey-scale values in 64 intervals. All other values were calculated after performing a Hough transform. Edge intensity was defined as constant grey scale values of greater than five pixels in width and accumulated in 64 histogram intervals. Edge slopes were accumulated in 45 histogram intervals of 2 degrees each. Lines were defined as constant grey scale values of less than 5 pixels and accumulated in 64 histogram intervals of 4–256 pixels. Line distance from the origin was defined as the number of pixels between the center of the line and the largest values of *XY* coordinates of the image and accumulated in 64 histogram intervals of 45–256 pixels. Angles of lines were accumulated in 45 histogram intervals of 2 degrees each. The algorithms for these processes are quite standard (Ballard & Brown, 1982). No claim is made that these features are the only possible features that might be used, nor even that these features are optimal. Other work suggests that fuzzy measure approaches may be more effective (Leigh, 1992).

By converting the histograms for each image into Lorenz information measures (Chang & Yang, 1983), each image of collection C may be reduced to a small set of real numbers, where each number constitutes a structural attribute of the entire image. However, since the abstraction method selects images based on the relative position of each image in the normal distribution of all images in collection C, each individual structural feature of the images must be assumed to be normally distributed. This assumption may be incorrect however, depending upon the composition of any set of particular images decomposed by any particular feature. To overcome the weaknesses of such an assumption, the structural features of each image are themselves averaged by calculating the mean and standard deviation for each measure across the collection C and converting the individual measures to unit normal deviates of the normal curve. Thus, all values are rendered to a common unit of measurement. By simply summing all the measures of an image, each image may be represented by a single value encompassing its entire structure (Shelton, 1991). These image values, now implicitly constituting a training set, are also processed to derive their mean and standard deviation, with the selection rule set to retain all values representing images at the tails of the normal distribution.

To demonstrate the method, approximately 51 minutes of video including one shuttle launch sequence, one shuttle landing sequence, one satellite deployment, and numerous onboard experiments were processed by the method of Figure 1, steps 1–8. Sampling at 5-second intervals from the 51 minutes yielded 604 frames. With the selection threshold set at ± 1.2 *SD*, 134 (approximately 22%) of the images were retained (Project ICON Laboratory, 1991). It should be noted that both the training set and the threshold parameters may be altered for the convenience of the user. Table 1 lists all scenes, the number of frames sampled from each scene at 5-second intervals, and the number of frames selected by the algorithm. Scenes discussed in detail in the results section are included in the Appendix and marked in Table 1 with an asterisk. Frames selected by the algorithm in the Appendix are those without cross marks. The number of frames in a scene may not be equal to the number of frames noted below due to space constraints for publication.

Results

Shuttle launch and landings (Figs. 2–3 and 15–17 in the Appendix) are significant events. Thus, the composition choice of the training set for these scenes at 11% of the total frames resulted in higher sampling rates by the algorithm than for the collection as a whole. Specifically, the average sampling rate of 22% for the collection is increased to 50% for the launch and to 40% for the landing. Essentially, the algorithm samples frames in inverse proportion to the appearance of scene types in the document, where scene "type" refers to the average composition of the image set in terms of its component light levels, edges, and other correlates of the image decomposition method described in the methodology. More significantly, however, is that the frames extracted from the total also represent significant launch and landing events. For example, Figures 2–3 contain frames from shuttle roll, full thrust, and Solid Rocket Booster separation; three of the more volatile events in any launch. Similarly, Figures 15–17 include significant frames from the first appearance, shift to descent attitude, first appearance on the horizon, landing gear deployment and roll-by. The algorithm failed, however, to select two frames in both these sequences that are also important, specifically ignition in the launch sequence and touchdown in landing.

Figures 6–7 and 11–14 represent the inverse frequency effect also. Onboard activity sequences formed 64% of the training set. Therefore, they were sampled at a lower rate than the entire collection of frames. Specifically, for these two scenes, the sampling percentages were approximately 10%. Figures 6–7 contain a long sequence of frames. The sampled images collapse this sequence into a storyboard of (1) an astronaut removing a panel from the shuttle aft bay; (2) a close-up of the removal; (3) display of an experimental panel with its documentation highlighted in the background; and (4) display of another experimental panel with its documentation also highlighted. Figures 11–14 consist of a food preparation sequence in microgravity. Although such scenes are generally of little scientific interest, they capture an important dimension of space flight in human terms. In this case, the sampled images collapse the sequence into a storyboard of (1) three frames showing successively more items prepared for cooking; (2) setting of a timing device; (3) two additional frames displaying the loading of the items into the microwave oven; and (4) one final frame of the astronaut apparently checking cooking instructions.

Figure 4 displays a flaw in the algorithm, corrected in Figure 1 by steps 9 and 10. Specifically, Figure 4 exhibits frames that fell at the tails of the normal distribution of images. Due to their similarity they were all selected. The 9 frames of the Remote Manipulator System (RMS) shown in Figure 4 represents this aspect of the algorithm in its worst case. First, the RMS is a member of the 25% of scenes comprising vehicle deployment and related activity, so that, due to the inverse frequency effect noted earlier, the sampling rate would tend to be higher than the onboard scenes in any case. Second, since the RMS moves very slowly during the satellite deployment process,

TABLE 1. Scene sequence of video used in the demonstration of the automatic abstracting method.

Scene Description	No. of frames	No. Selected
1. Shuttle launch*	32	17
2. Shuttle bay doors opening prior to deployment	23	2
3. Flight deck (forward) scene shift before deployment	8	0
4. Remote Manipulator System lift of vehicle from bay	31	6
5. Flight deck (aft) scene shift	6	1
6. Astronaut exercising in weightless environment	26	22
7. Detail of the Remote Manipulator System effector	18	4
8. Remote Manipulator System arm and effector*	9	9
9. Interior room of the mission control building during flight	6	0
10. Scene shift to space vehicle emerging from shuttle bay	1	1
11. Unidentified out of focus scene (appears as flat grey panel)	7	0
12. Split screen for astronauts and vehicle emerging	5	0
13. Astronauts standing on flight deck (aft)	10	1
14. Alternating frames of mission control and shuttle in space on Earth limb*	5	3
15. Window reflection of interior light	4	0
16. Spacecraft view against Earth at oblique angle	8	2
17. Astronaut emerging from rigid sleep sation and mid-deck (starboard aft)	13	2
18. Astronaut being thrown towel in weightless environment	24	4
19. Pan scene of mid-deck	9	1
20. Pan scene of mid-deck	4	0
22. Astronaut (standing) describing equipment readout facility for camera	24	4
23. Alternating views of flight deck (aft) and mission control room	5	5
24. Astronaut pulling experiment racks for camera on flight deck (aft)*	44	3
25. Astronaut emerging from forward to aft flight decks	3	2
26. Pan shot of empty flight deck (aft)	3	0
27. Astronaut examining flight hardware*	24	0
28. Astronaut examining manuals and experiment locker on flight deck	47	6
29. Equipment floating in flight deck (forward)	10	3
30. Water globule experiment	4	2
31. Acoustic levitation experiment*	15	0
32. Meso-scale lightning experiment*	21	4
33. Astronaut emerging from rigid sleep station	9	2
34. Miscellaneous onboard activities including view of oscilloscope	36	2
35. Alternating view of mission control and ocean with clouds	5	0
36. Shuttle onboard food preparation*	66	8
37. Sunrise over the Earth limb	5	1
38. Shuttle landing*	34	17

all the scenes decompose into very close values. Steps 9 and 10 of Figure 1 could be performed in many ways. However, the most direct one would simply be to observe the items selected and set a cut-off value relevant to the training set and the characteristics of the population of visual documents to be filtered by the algorithm.

Finally, of the 38 scenes in the visual document, 10 were unsampled. Of these 10 scenes, only 2 unsampled scenes (Figs. 8 and 9) pose a problem. In all other cases, the scenes skipped include material of little interest. Figures 8 and 9 are events of interest. However, the algorithm fails to include even one frame because in both cases, there was simply insufficient contrast in the images. Unfortunately, the cylinder examined by the astronaut in Figure 8 is almost the same light level as the background cabinet. Similarly, the glass beads suspended by acoustic pressure in Figure 9 do not differ significantly from the plate background because they are clear. This is similar to the case in Figure 6 when the astronaut is pulling experiment racks; only the frames showing racks with the documentation in white held behind them are selected by the algorithm. Since the technique used in this demonstration examines only grey levels, appropriate frames from Figure 9 remained unsampled. In Figure 1, however, it is suggested that a more sophisticated measure of light should be used, specifically a three-set histogram for hue, saturation, and chroma. Whether or not this procedure would correct this deficiency of the algorithm's performance remains unknown at this time. Another potential solution also exists. Specifically, the normal curve area values could be expanded to select a greater number of frames. This solution, coupled with a parameter to reject frames with closely similar values (steps 9 and 10 in Fig. 1), might produce the best results.

Discussion

Given the light level analysis technique and normality assumptions of the algorithm used for this demonstration, the performance of the abstracting method appears to be acceptable for a large variety of visual documentation representation tasks. The errors of the algorithm are, in contrast to human editing procedures, at least due to traceable causes and exhibited as systematic error. Knowing that the filtering algorithm will focus on scenes of high contrast is quite different from attempting to determine the aesthetic sense of a human editor.

Although the most important near term requirement for the algorithm is the conservation of storage space in machine-readable libraries, the broader application is the general conservation of human attentiveness. It has been observed that the fundamental problem of librarianship is that data are infinite and attention is scarce. Thus, the broadest application of the abstracting method is in the presentation of the central contents of visual documents in a manner such that minutes of visual imagery may be reduced to merely a few frames. Another class of applications is the general problem of machine vigilance. This problem arises in security systems, where cameras might scan constantly, while human attention is limited to a few intermittent pans. Machine surveillance of many industrial processes are also candidates for this algorithm. In all cases, a computer could be programmed to sound an alarm whenever a scene fell outside the prescribed normal distribution limits.

Overall, the general performance of the algorithm may also indicate that the assumption of Marr regarding the paucity of information necessary for the human neocortex to assess imagery is correct. Consider the model of human attentiveness on which the abstracting method is based: at all times, we create a running probability distribution of scenes in our minds; the distribution changes with the addition and deletion of sensation; we attend to outliers in the distribution. Although sparse, the model has the explanatory power to account for a phenomena which we have all experienced: in a crowded room we are much less likely to notice the entry of one more person; in a nearly vacant room, we are much more likely to notice it. Moreover, when it is considered that the model of attentiveness used for this algorithm is operationalized by processing frames to a single representative value, the temptation to construe human vision as primarily an internal mental function rises significantly.

Appendix

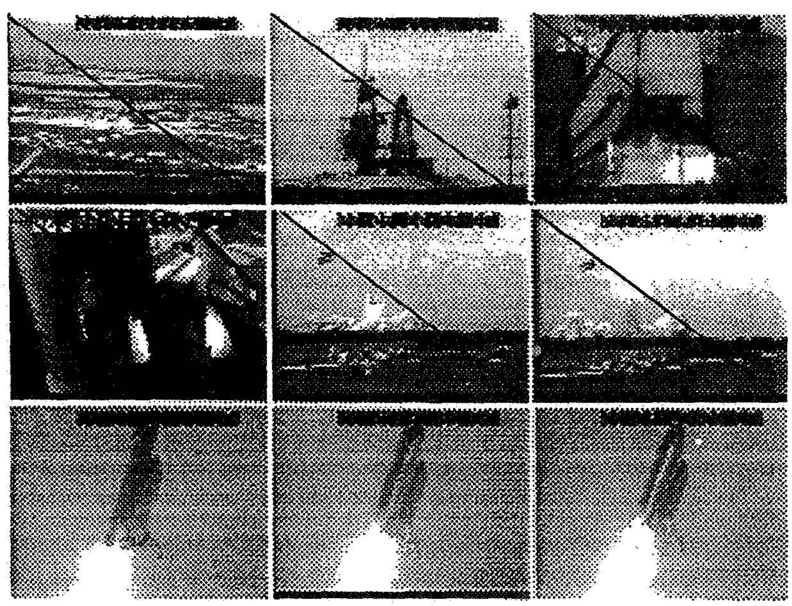


FIG. 2. Scene 1: Shuttle launch sequence, part 1.

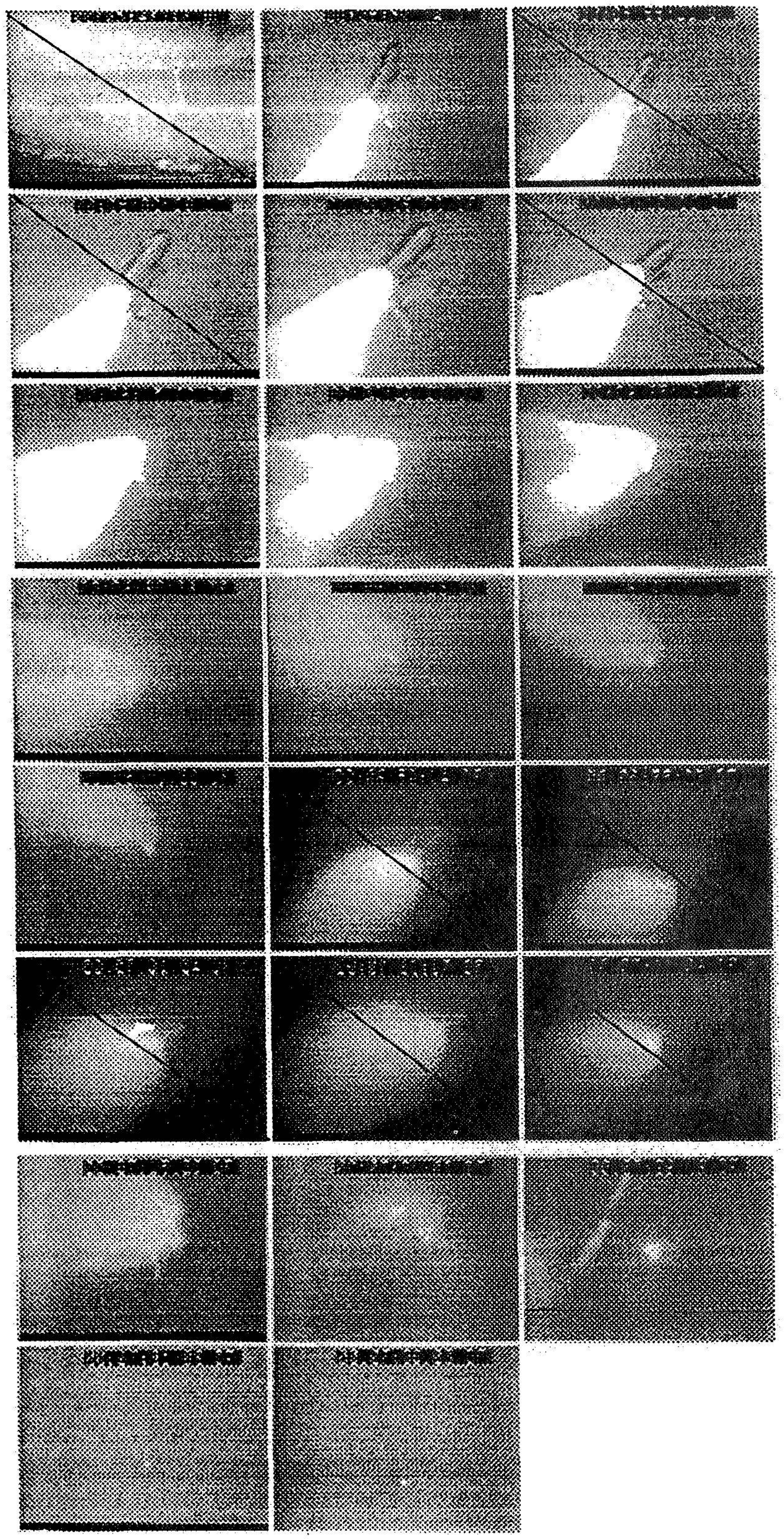


FIG. 3. Scene 1: Shuttle launch sequence, part 2.

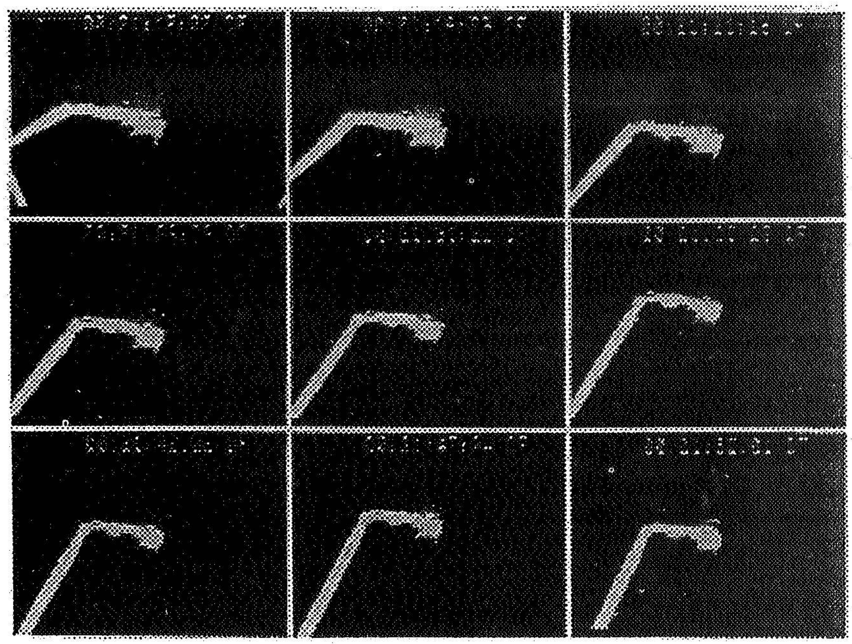


FIG. 4. Scene 8: Remote Manipulator System arm and effector.

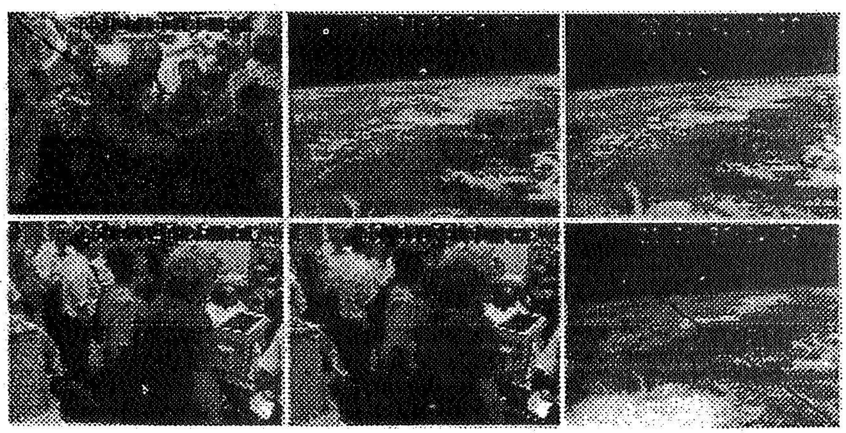


FIG. 5. Scene 14: Alternating frames of mission control and shuttle in space on Earth limb.

FIG. 6. Scene 24: Astronaut pulling experiment racks, part 1.

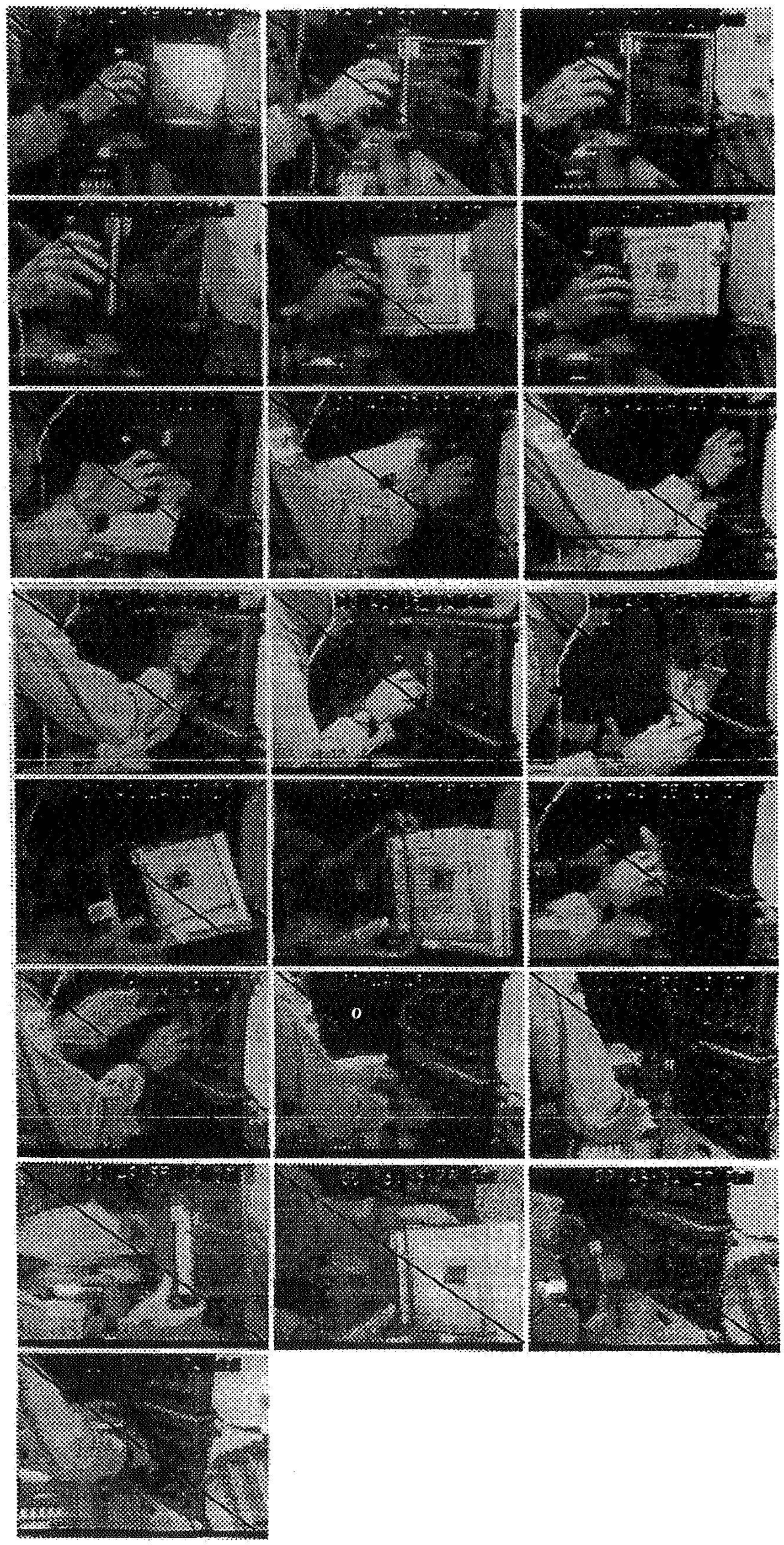


FIG. 7. Scene 24: Astronaut pulling experiment racks, part 2.

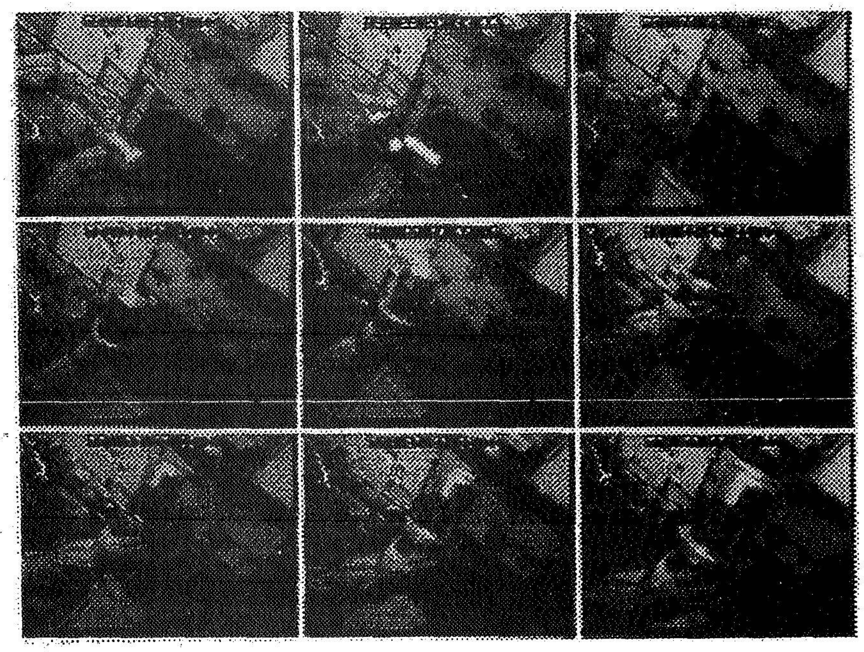


FIG. 8. Scene 27: Astronaut examing flight hardware, frames 4-12.

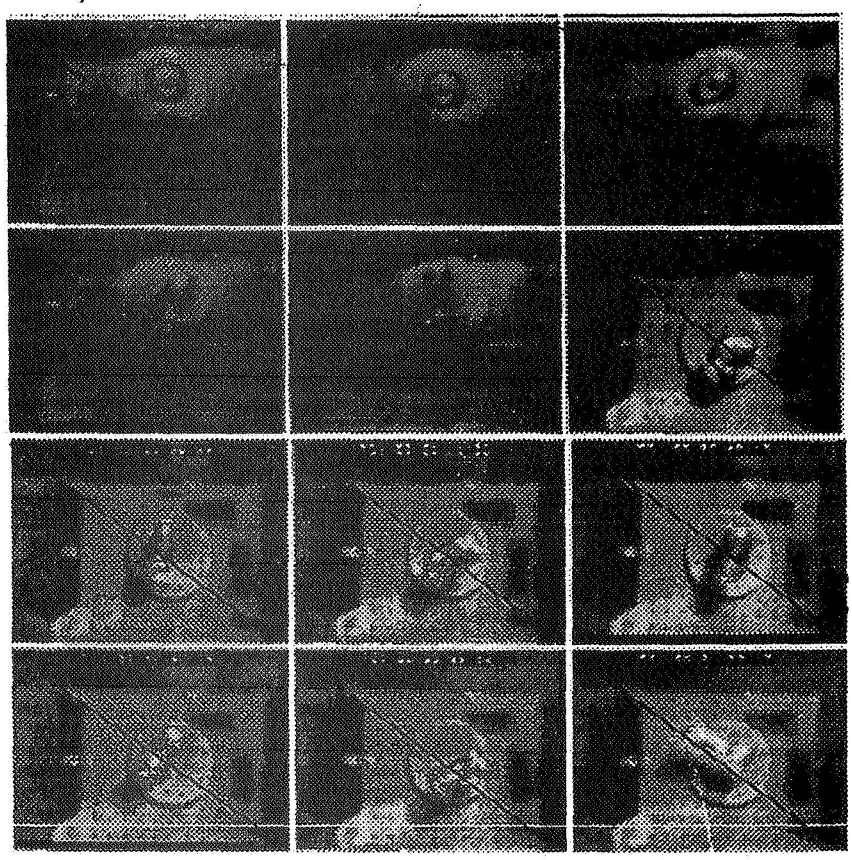


FIG. 9. Scene 31: Acoustic levitation experiment, frames 2-13.

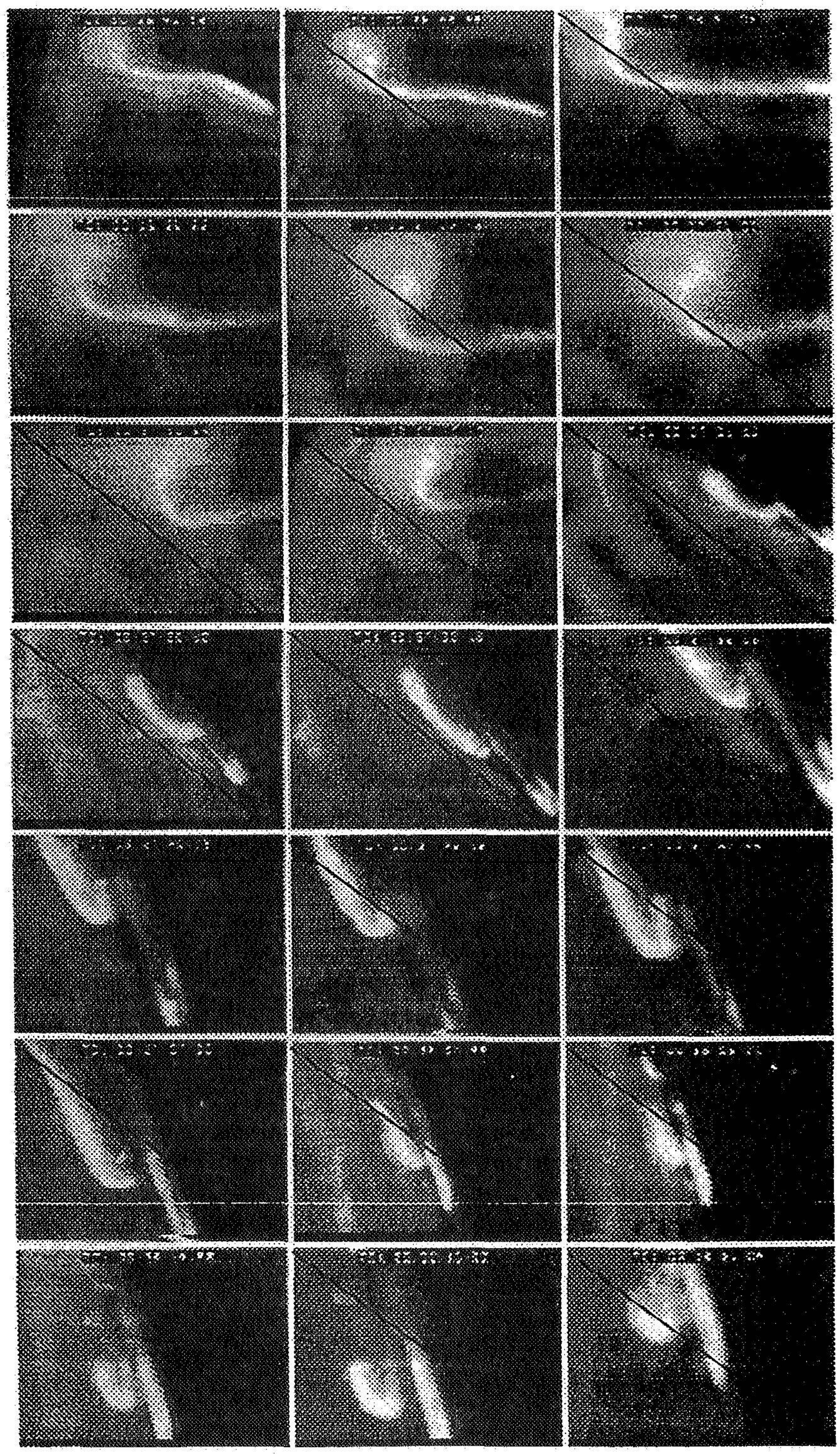


FIG. 10. Scene 32: Meso-scale lightning experiment.

FIG. 11. Scene 36. Shuttle onboard food preparation, part 1.

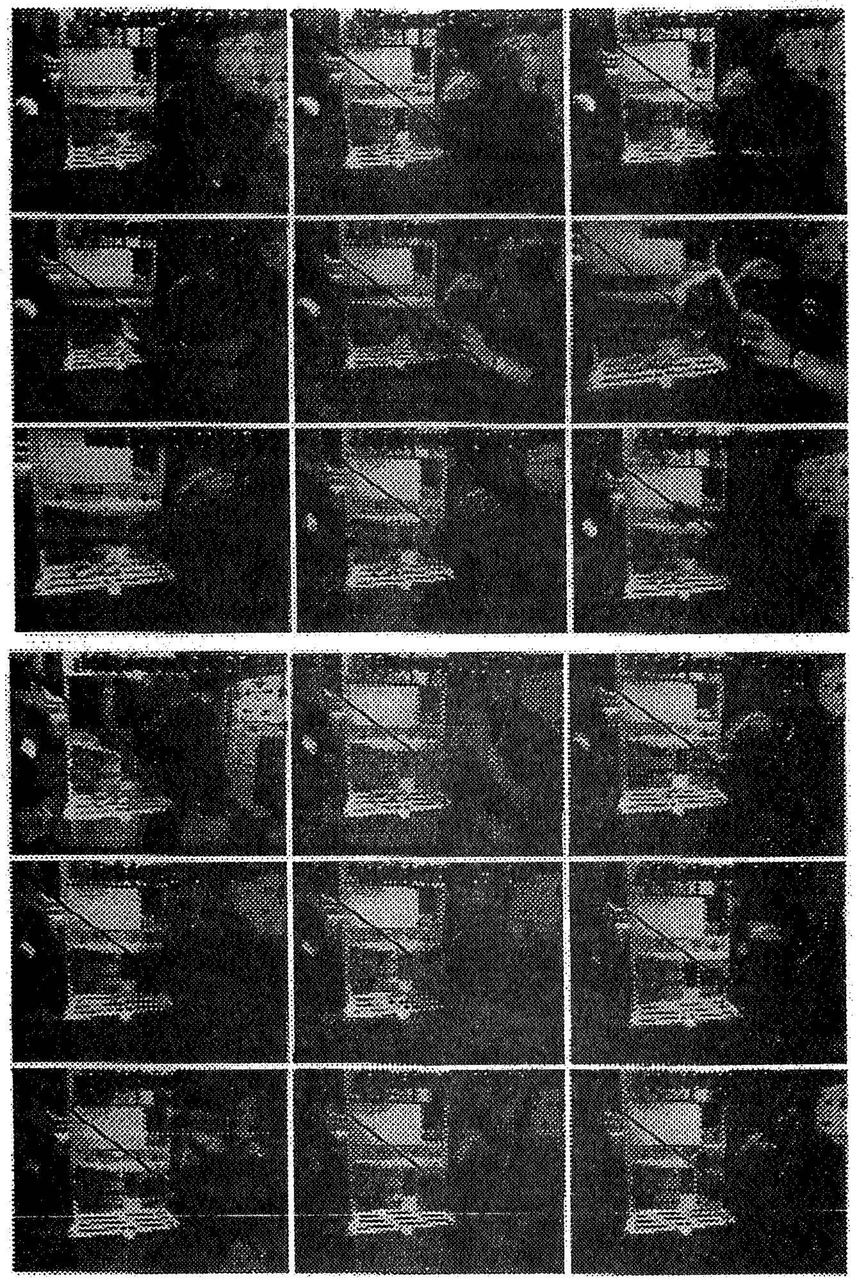


FIG. 12. Scene 36: Shuttle onboard food preparation, part 2.

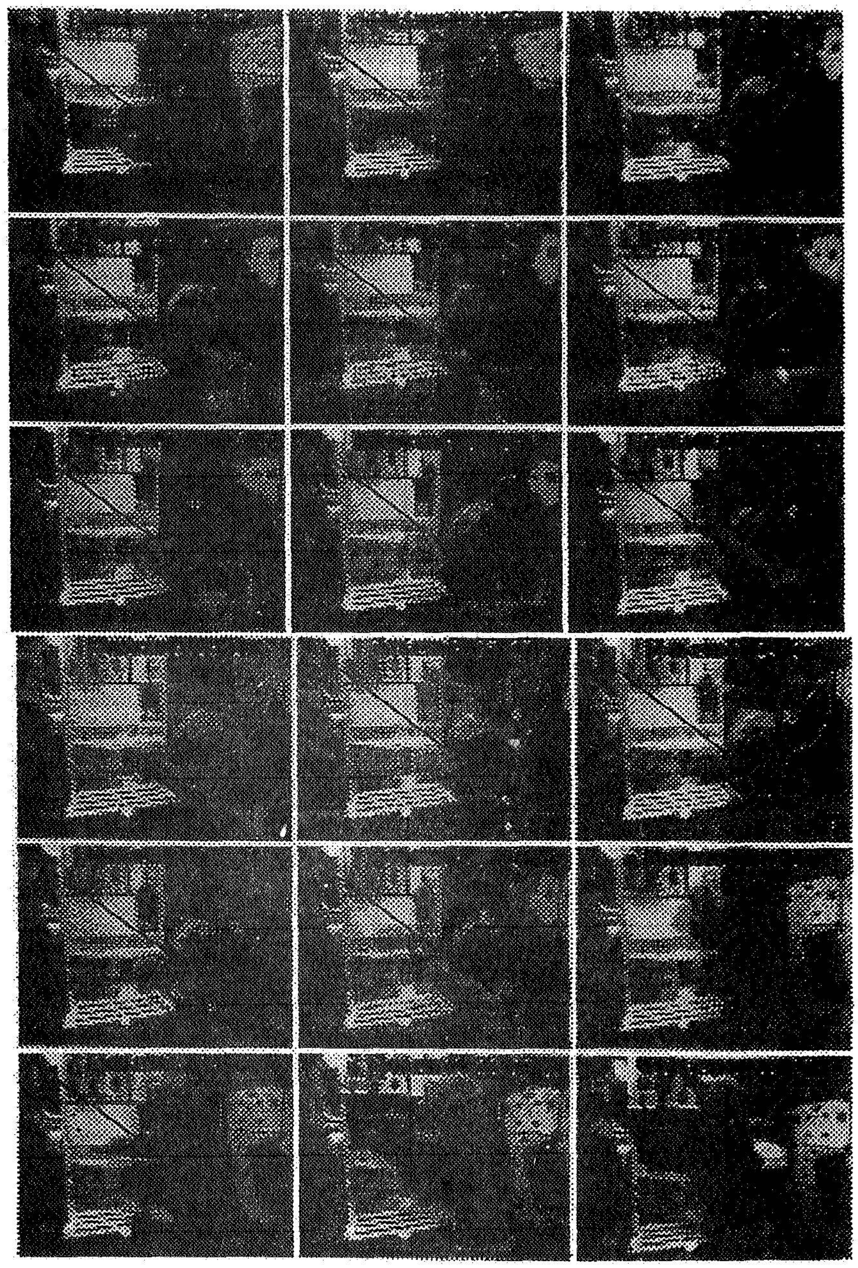


FIG. 13. Scene 36: Shuttle onboard food preparation, part 3.

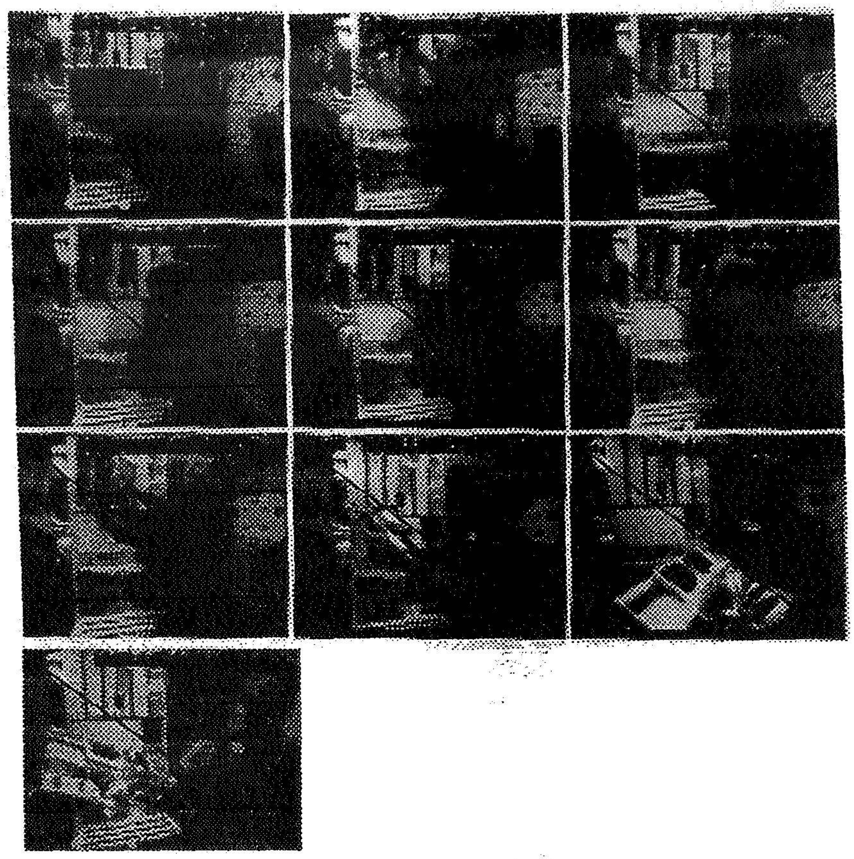


FIG. 14. Scene 36: Shuttle onboard food preparation, part 4.

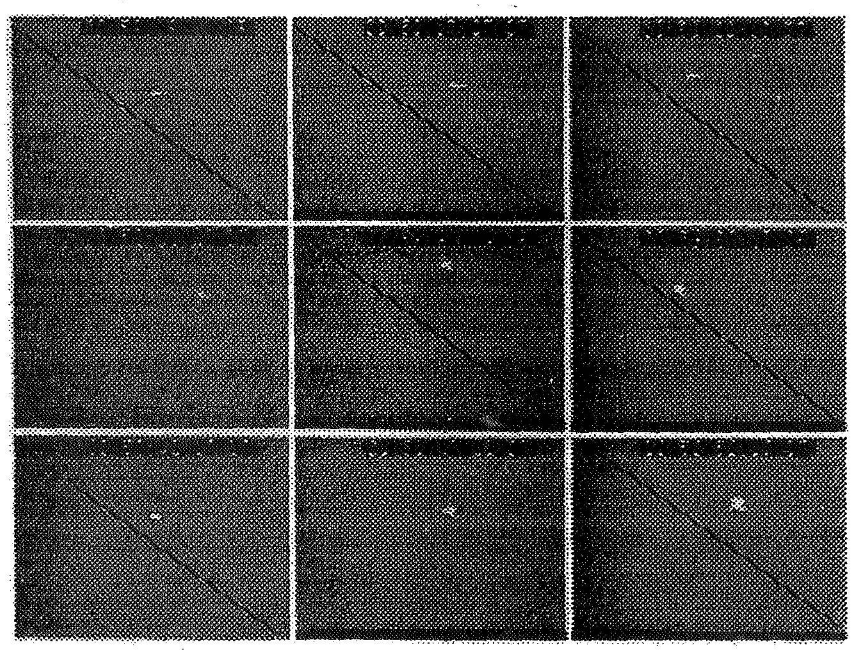


FIG. 15. Scene 38: Shuttle landing sequence, part 1.

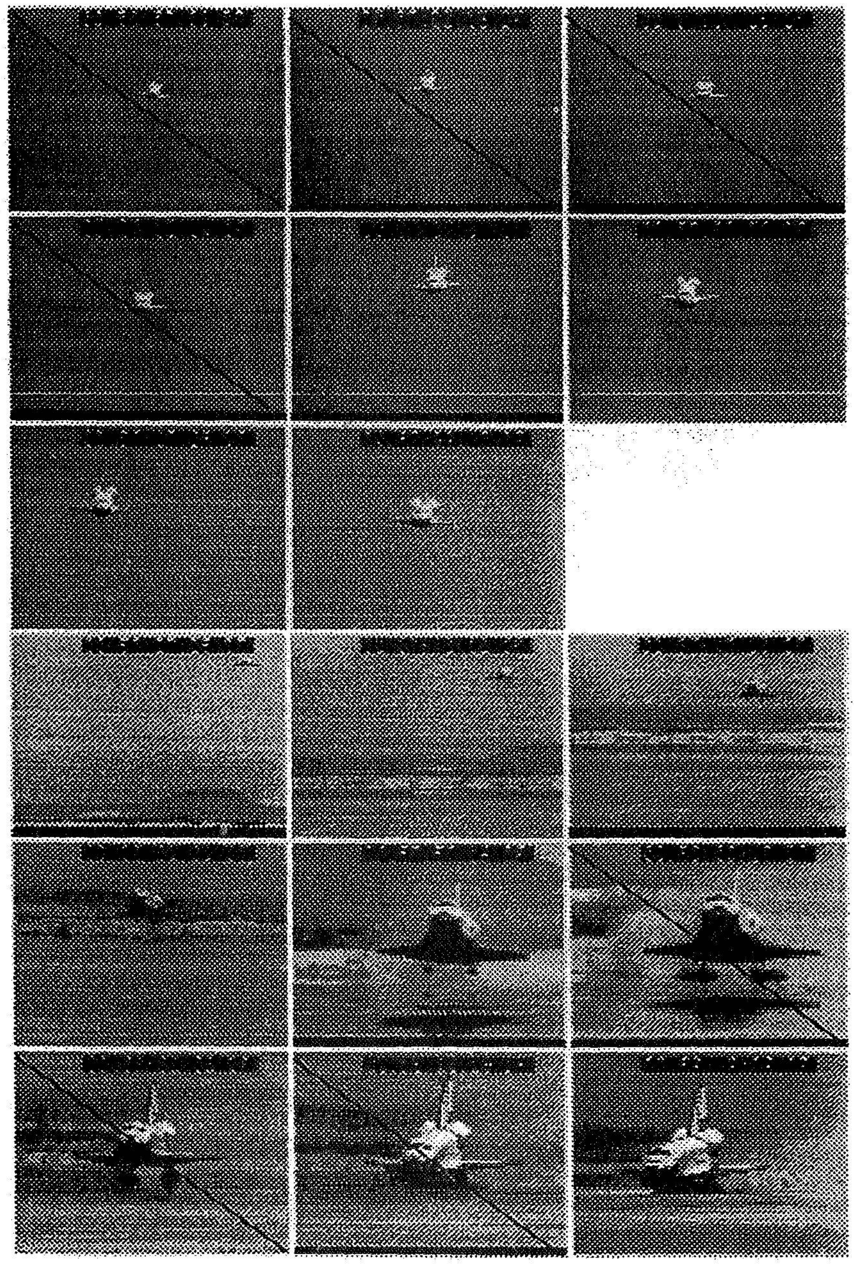


FIG. 16. Scene 38: Shuttle landing sequence, part 2.

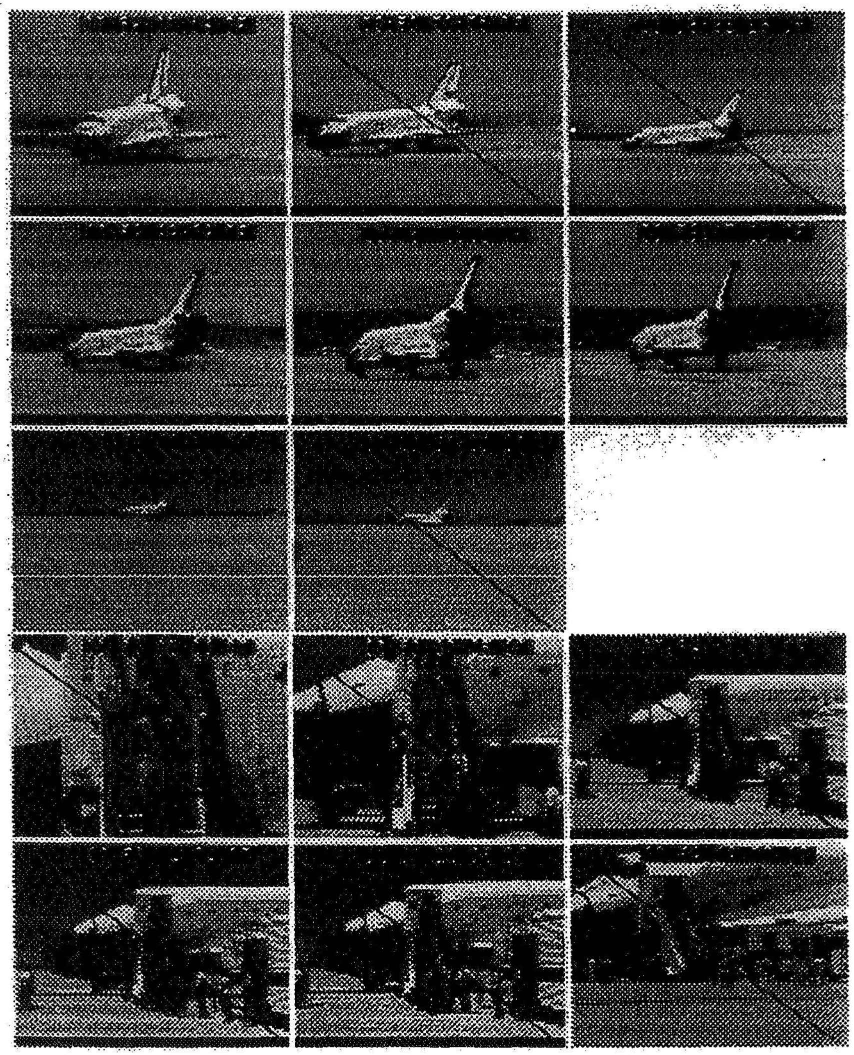


FIG. 17. Scene 38: Shuttle landing sequence, part 3.

References

Ballard, D. H., & Brown, C. M. (1982). *Computer vision,* Englewood Cliffs, NJ: Prentice-Hall.

Chang, S. K., & Yang, C. C. (1983). Picture information measures for similarity retrieval. *Computer Vision, Graphics, and Image Processing, 23,* 366–375.

Leigh, A. B. (1992). *A fuzzy measure approach to motion frame analysis for video data abstraction,* Unpublished M.S. Thesis, Department of Computer Science, University of Houston–Clear Lake, Houston, TX.

O'Conner, B. C. (1985). Access to moving image documents: Background concepts and proposals for surrogates for film and video works. *Journal of Documentation, 41,* 209–220.

Project ICON Laboratory, University of Texas, Austin (1991). Project staff: Mary Ann Albin, Dr. Steven Fitzpatrick, Hee-Sook Choi, Abby Goodrum, Diane Luccy, Stephanie Smith, John Stansbury, Nackil Sung, Margaret Whitehead, Dr. Ronald Wyllys.

Pryluck, C., Teddlie, C., & Sands, R. (1982). Meaning in film/video: Order, time and ambiguity. *Journal of Broadcasting, 26,* 685–695.

Shelton, R. O. (1991, April). Personal communication with Robert Shelton, Computer Engineer, Software Technology Branch, Johnson Space Center.

Yeh, P. -S., et al. (1991, February). On the optimality of Code Options for a Universal Noiseless Coder. *JPL Publication 91-2,* Pasadena, CA: Jet Propulsion Laboratory.

omit T

Session S2: INTELLIGENT SYSTEMS

Session Chair: Wade Webster

S20-61
44822

CLIPS: THE C LANGUAGE INTEGRATED PRODUCTION SYSTEM

Gary Riley
Software Technology Branch
NASA Johnson Space Center
Mail Stop PT4
Houston, TX 77058

ABSTRACT

Expert systems are computer programs which emulate human expertise in well defined problem domains. The potential payoff from expert systems is high: valuable expertise can be captured and preserved, repetitive and/or mundane tasks requiring human expertise can be automated, and uniformity can be applied in decision making processes. The C Language Integrated Production System (CLIPS) is an expert system building tool, developed at the Johnson Space Center, which provides a complete environment for the development and delivery of rule and/or object based expert systems. CLIPS was specifically designed to provide a low cost option for developing and deploying expert system applications across a wide range of hardware platforms. The commercial potential of CLIPS is vast. Currently, CLIPS is being used by over 5,000 individuals throughout the public and private sector. Because the CLIPS source code is readily available, numerous groups have used CLIPS as the basis for their own expert system tools. To date, three commercially available tools have been derived from CLIPS. In general, the development of CLIPS has helped to improve the ability to deliver expert system technology throughout the public and private sectors for a wide range of applications and diverse computing environments.

INTRODUCTION

Conventional programming languages, such as FORTRAN and C, are designed and optimized for the procedural manipulation of data (such as numbers and arrays). Humans, however, often solve complex problems using very abstract, symbolic approaches which are not well suited for implementation in conventional languages. Although abstract information can be modeled in these languages, considerable programming effort is required to transform the information to a format usable with procedural programming paradigms.

One of the results of research in the area of artificial intelligence has been the development of techniques which allow the modeling of information at higher levels of abstraction. These techniques are embodied in languages or tools which allow programs to be built that closely resemble human logic in their implementation and are therefore easier to develop and maintain. These programs, which emulate human expertise in well defined problem domains, are called expert systems. The availability of expert system tools has greatly reduced the effort and cost involved in developing an expert system.

The C Language Integrated Production System (CLIPS) [1, 2, 3, 4] is an expert system tool developed by the Software Technology Branch at NASA's Johnson Space Center. The prototype of CLIPS, version 1.0, was developed in the spring of 1985 in a UNIX environment. Subsequent development of CLIPS greatly improved its portability, performance, and functionality. The first release of CLIPS, version 3.0, was in July of 1986. The latest version of CLIPS, version 6.0, was released in August of 1993. A version of CLIPS written entirely in Ada, CLIPS/Ada, has also been developed. CLIPS is currently available to the general public through the Computer Software Management and Information Center (see appendix).

KEY FEATURES OF CLIPS

CLIPS was designed to address several issues key to NASA. Among these were the ability to run on a wide variety of conventional hardware platforms, the ability to be integrated with and embedded within conventional software systems, and the ability to provide low cost options for the development and delivery of expert systems.

CLIPS is written in C for portability and speed and has been installed on many different computers without changes to the source code. At the time of its original development, CLIPS was one of the few tools that was written in C and capable of running on a wide variety of conventional platforms. CLIPS can be ported to any system which has an ANSI compliant C compiler including personal computers (IBM PC compatibles, Macintosh, Amiga), workstations (Sun, Apollo, NeXT), minicomputers (VAX 11/780, HP9000-500), Mainframes (IBM/370), and supercomputers (CRAY).

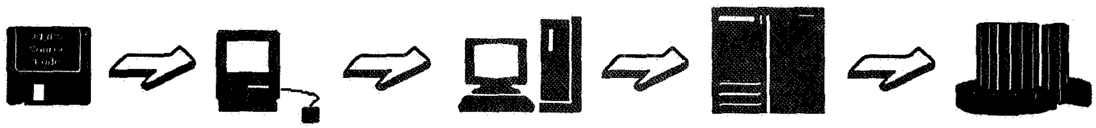


Figure 1. CLIPS is Easily Ported From One Environment to Another

To maintain portability, CLIPS utilizes the concept of a portable kernel. The kernel represents a section of code which utilizes no machine dependent features. The inference engine contains the key functionality of CLIPS and is used to execute an expert system. Access functions allow CLIPS to be embedded within other systems. This allows an expert system to be called as a subroutine (representing perhaps only one small part of a much larger program). In addition, information stored in CLIPS can be accessed and used by other programs. Integration protocols allow CLIPS to utilize programs written in other languages such as C, FORTRAN, and Ada. Integration guarantees that an expert system does not have to be relegated to performing tasks better left to conventional procedural languages. It also allows existing conventional code to be utilized. The CLIPS language can also be easily extended by a user through the use of the integration protocols.

To provide machine dependent features, such as windowed interfaces or graphics editors, CLIPS provides fully documented software hooks which allow machine dependent features to be integrated with the kernel. The I/O router system allows interfaces to be layered on top of CLIPS without making changes to the CLIPS kernel. The standard interface for CLIPS is a simple, text-oriented, command prompt. However, three interfaces are also provided with CLIPS that make use of the I/O router system and integration protocols to provide machine specific interfaces. These interfaces are provided for Apple Macintosh systems, IBM PC MS-DOS compatible systems, and X Window systems. Figure 2 shows the CLIPS interface for the Macintosh computer.

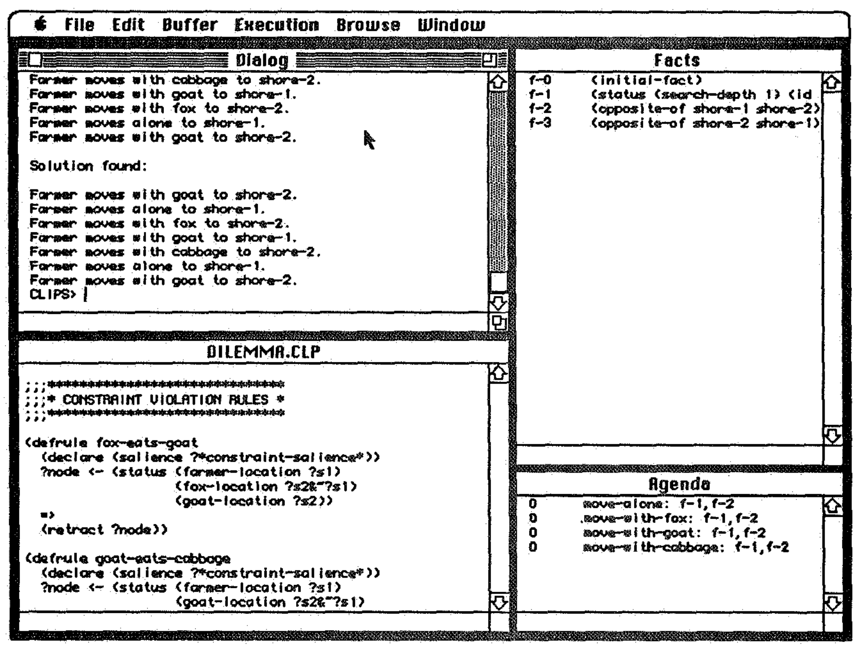


Figure 2. CLIPS Macintosh Interface

One of the key appeals of the CLIPS language results from the availability of the approximately 40,000 lines of CLIPS source code. Because the development of an expert system tool can require many man-years, the benefits of using CLIPS as a starting point for research and the creation of special purpose expert system tools cannot be understated. CLIPS users have enjoyed a great deal of success in adding their own language extensions to CLIPS due to the source code availability and its open architecture. Many users have also developed their own interfaces and interface extensions [5, 6].

KNOWLEDGE REPRESENTATION

Expert system tools are designed to provide highly productive environments by allowing knowledge to be represented flexibly. A flexible representation scheme allows the application developers to try several different approaches or to use an approach best suited to their problem. CLIPS provides a cohesive tool for handling a wide variety of knowledge with support for three different programming paradigms: rule-based, object-oriented, and procedural. In addition, CLIPS also supports the concepts of iterative refinement (refining an expert system with small iterative changes) and rapid prototyping (demonstrating proof of concept) which are found in many expert system tools.

Rule-Based Programming

The first (and originally the only) programming paradigm provided by CLIPS is rule-based programming. In this programming paradigm, rules are used to represent heuristics, or "rules of thumb", which specify a set of actions to be performed for a given situation. A rule is composed of an *if* portion and a *then* portion. The *if* portion of a rule is a series of patterns which specify the facts (or data) which cause the rule to be applicable. The process of matching facts to patterns is called pattern matching. CLIPS provides a mechanism, called the inference engine, which automatically matches facts against patterns and determines which rules are applicable. The *if* portion of a rule can actually be thought of as the *whenever* portion of a rule since pattern matching always occurs whenever changes are made to facts. The *then* portion of a rule is the set of actions to be executed when the rule is applicable. The actions of applicable rules are executed when the CLIPS inference engine is instructed to begin execution. The inference engine selects a rule and then the actions of the selected rule are executed (which may affect the list of applicable rules by adding or removing facts). The inference engine then selects another rule and executes its actions. This process, illustrated by Figure 3, continues until no applicable rules remain.

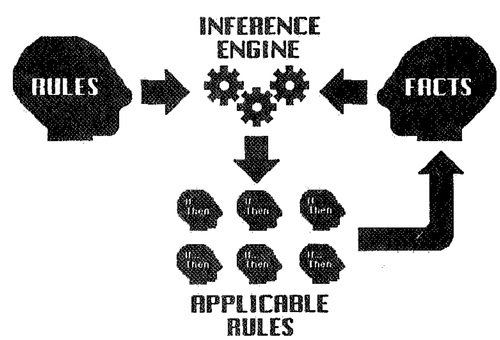


Figure 3. Execution of a Rule-Based Program

To illustrate the advantages of rule-based programming, consider the problem of monitoring a series of sensors. The following example program written in the C programming language illustrates how these sensors could be monitored using a procedural programming paradigm to determine if any two of the sensors have *bad* values (which a hypothetical expert indicates represents an overheated device).

```
#define BAD 0
#define GOOD 1
#define DEVICE_OVERHEATED 0
#define DEVICE_NORMAL 1

int CheckSensors(sensorValues,numberOfSensors)
  int sensorValues[];
  int numberOfSensors;
  {
   int firstSensor, secondSensor;

   for (firstSensor = 1;
        firstSensor <= numberOfSensors;
        firstSensor++)
      {
       for (secondSensor = 1;
            secondSensor <= numberOfSensors;
            secondSensor++)
          {
           if ((firstSensor != secondSensor) &&
               (sensorValues[firstSensor] == BAD) &&
               (sensorValues[secondSensor] == BAD))
             { return(DEVICE_OVERHEATED); }
          }
      }

    return(DEVICE_NORMAL);
   }
```

The *CheckSensors* function is implemented by storing the values of the sensors as integers in an array and then using two *for* loops to compare all combinations to determine if any two sensors have *bad* values. This function is relatively efficient if the sensors only need to be checked once. However, if this check is performed each time a sensor's value is changed, then all possible combinations are rechecked which is inefficient. In addition, the programmer has the responsibility for calling this function whenever an update is made to a sensor's value. An additional function could be written to check only one sensor against all other sensors, however, this increases the burden on the programmer. For contrast, the equivalent CLIPS code for a rule which performs the same task is shown following.

```
(defrule Two-Sensors-are-Bad
   (Sensor (ID-number ?id)  (status Bad))
   (Sensor (ID-number ~?id) (status Bad))
   =>
   (assert (Device (status Overheated))))
```

The first line of the rule contains the keyword *defrule* which indicates that a rule is being defined. The symbol *Two-Sensors-are-Bad* is the name of the rule. The next two lines beginning with the symbol *Sensor* are the patterns that form the *if* portion of the rule. Essentially, the first pattern searches for any *Sensor* fact that contains a *status* value of *Bad* and the second pattern searches for another *Sensor* fact with a *status* value of *Bad* that does not have the same *ID-number* as the *Sensor* fact matching the first pattern. The => symbol serves to separate the *if* portion of the rule from the *then* portion of the rule. Finally, the *assert* command in the *then* portion of the rule creates a new fact which indicates that the device has overheated.

Because of the overhead associated with the inference engine and the generality provided through pattern matching, a rule-based program generally does not execute as quickly as a procedural program. However, significantly less code is required and the programmer does not have to explicitly check for applicable rules when sensor values are changed. Rules are always looking for new facts which satisfy their conditions. Indeed, careless implementation of pattern matching capabilities in a procedural language may result in a program which runs much less efficiently than its rule-based counterpart. CLIPS's inference engine is based on the Rete algorithm [7] which is an extremely efficient algorithm for pattern matching.

Object-Oriented Programming

The second programming paradigm provided by CLIPS is object-oriented programming. This programming paradigm allows complex systems to be modelled as modular components (which can be easily reused to model other systems or to create new components). Object-oriented programming encompasses a number of concepts including data abstraction (the ability to define complex objects using high level representations), encapsulation (the ability to hide the implementation details of an object, thereby increasing its modularity and potential for reuse), inheritance (the ability to define new classes of objects by reusing existing classes), and polymorphism (the ability of different objects to respond to the same "command" in specialized ways).

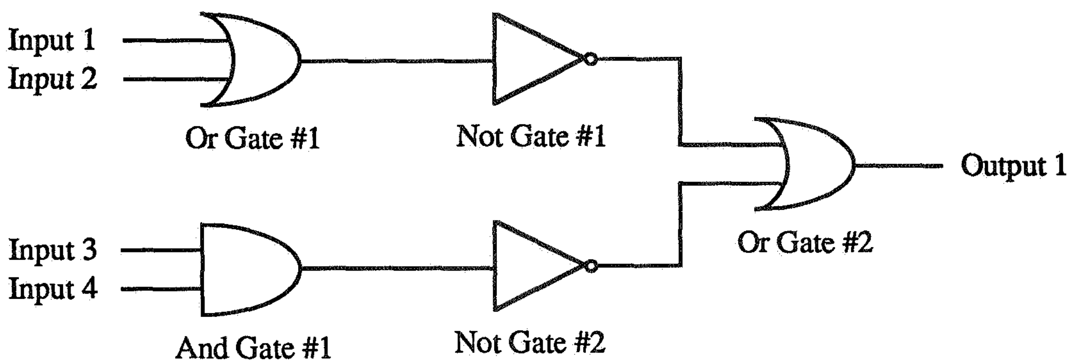


Figure 4. A Simple Electronic Circuit

Figure 4 shows a diagram of an electronic circuit consisting of *and*, *or*, and *not* gates. In electronics, a gate is a circuit that has an output dependent on some function of its input. The gates shown in Figure 4 all have boolean inputs and boolean output values. Physically, these boolean values would correspond to high and low voltages. Conceptually, these boolean values could be considered as *On* and *Off* or *True* and *False*. An *and* gate has an output value of *True* if all of its inputs are *True*, otherwise its output value is *False*. An *or* gate has an output value of *True* if any of its inputs are *True*, otherwise its output value is *False*. A *not* gate has an output value of *True* if its input value is *False* and an output value of *False* if its input value is *True*. In Figure 4, if *Input 1* and *Input 2* are both *False* and *Input 3* and *Input 4* are both *True*, then the output of *Or Gate #1* would be *False* and the output of *And Gate #1* would be *True*. The output of *Not Gate #1* would be *True* since its input (the output of *Or Gate #1*) is *False*. The output of *Not Gate #2* would be *False* since its input (the output of *And Gate #1*) is *True*. Finally, *Output 1* from *Or Gate #2* would be *True* since at least one of its inputs (the output from *Not Gate #1*) is *True*.

Using object-oriented programming methodologies, it is relatively easy to model the behavior of the electronic circuit shown in Figure 4. The first step in modelling the circuit is to define classes which can be used to describe the gates used in the circuit. Since all of the gates might have some attributes in common (such as a part number), it would be useful to first define a *Gate* class. Another class, *One Input*, could be used to describe the attributes associated with a single input gate (such as a *not* gate). Since a two input gate is essentially a one input gate with an additional input, the *Two Input* class could inherit the attributes of the *One Input* class and then define additional attributes for the second input. Similarly, a *One Output* class and *Two Output* class could also be defined. Figure 5 illustrates the basic classes used to describe the gate circuits in Figure 4. The classes described illustrate the basic concepts of data abstraction and inheritance. Note that even though the circuit gates shown in Figure 4 would not need to utilize the *Two Output* class, other types of gates could utilize this class. For example, a *splitter* gate (which splits its one input into two identical outputs) could make use of this class.

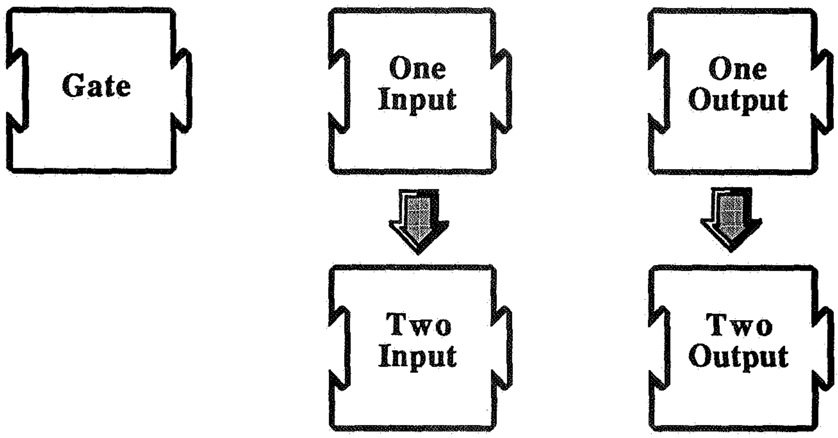


Figure 5. Classes Used to Describe Electronic Circuit Gates

Once the base classes for the gates are defined, it is possible to describe the gates in terms of these classes. Figure 6 conceptually illustrates how this could be done for the *not* gate and the *and* gate. The type of inheritance shown in Figure 6 is called multiple inheritance since a single class is inheriting attributes from more than one class. For example, the *And Gate* class inherits attributes from the *Two Input*, *One Output*, and *Gate* classes. In contrast, the inheritance shown in Figure 5 is called single inheritance since a single class inherits attributes from at most one other class (such as the *Two Input* class inheriting attributes from the *One Input* class). Some object-oriented programming languages support only single inheritance. CLIPS provides support for full multiple inheritance.

Once the gate classes have been defined, it is possible to define instances (or objects) of these classes. For example, *Or Gate #1* would be a specific instance of the *Or-Gate* class as would *Or Gate #2*. It would have its own data areas for storing its input and output values. Thus a class serves as the prototypical definition which is used for creating objects belonging to that class.

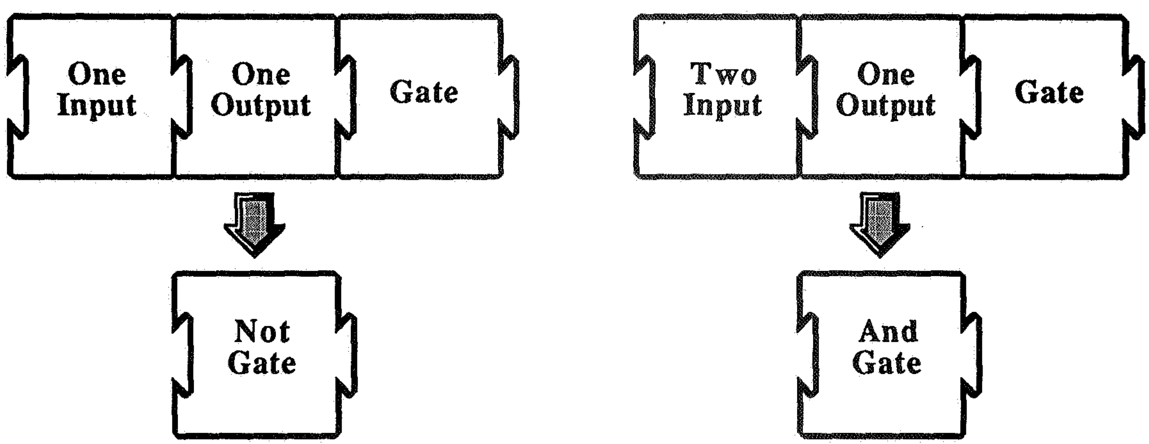


Figure 6. Building New Classes from Existing Classes

In CLIPS, objects are manipulated by sending them messages which specify an action to perform. For the circuit example, an appropriate action might be to recompute the output of a gate based upon its inputs. Notice that even though the *or* gates and *and* gates are both *Two-Input One-Output Gates*, their outputs are computed differently. In object-oriented programming, procedures as well as data can be associated with objects. Rather than writing one routine to compute the output values for all gate types given their inputs, the routines for computing outputs for objects can be encapsulated inside the classes themselves. When an *or* gate is sent a *Compute Output* message, its output is computed to be *True* if either of its inputs are *True*, otherwise its output is *False*. When an *and* gate is sent a *Compute Output* message, its output is computed to be *True* if both of its inputs are *True*, otherwise its output is *False*. Thus, both objects respond differently, yet appropriately, to the same message. This behavior is the essence of polymorphism and is illustrated by Figure 7. The procedures attached to classes are referred to as message-handlers.

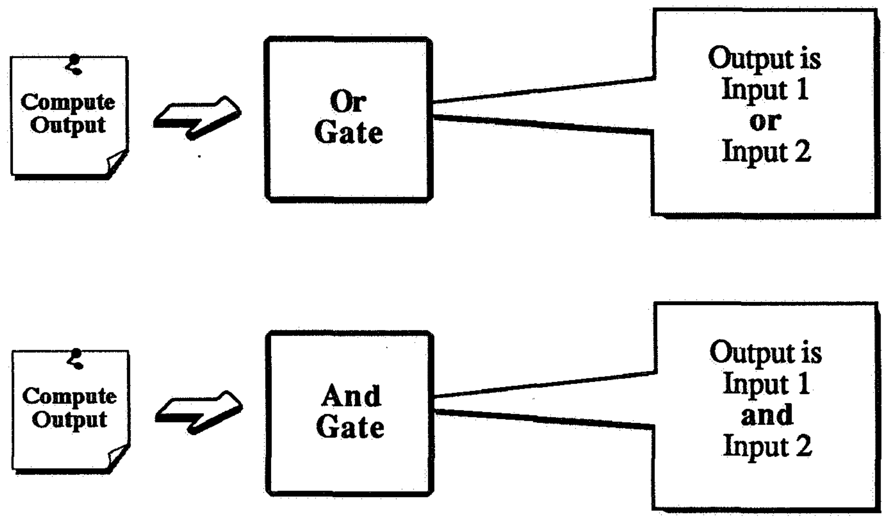


Figure 7. Two Different Objects Responding Differently to the Same Message

Procedural Programming

The third programming paradigm provided by CLIPS is procedural programming. This programming paradigm provides capabilities similar to those found in languages such as C, Pascal, Ada, and LISP. With respect to building expert systems, these are the least interesting capabilities provided by CLIPS. However, the ability to define procedural code directly within CLIPS allows new procedural capabilities to be added to CLIPS without the need of a compiler or linker. To add new capabilities to CLIPS which have been written in languages such as C, FORTRAN, or Ada, a compiler and linker are required to recompile and relink the new source code with the CLIPS source code. CLIPS allows the definition of global variables, functions, and generic functions. Generic functions are the most interesting feature of the CLIPS procedural programming language in that they allow different pieces of procedural code to be executed depending upon the arguments used when calling a function. This capability is called function overloading. As an example, the addition function could be overloaded so that numeric data types are numerically added and string data types are concatenated.

CURRENT USES

Although CLIPS was originally developed to aid in the construction of aerospace related expert systems, it has been put to widespread usage in a number of fields. CLIPS is being used by over 5,000 users throughout the public and private community including: all NASA sites and branches of the military, numerous federal bureaus, government contractors, universities, and many companies At the First and Second CLIPS Conferences held in August 1990 and September 1991 respectively, over 120 papers were presented on a diverse range of topics. In addition to aerospace and engineering applications, some other examples of CLIPS applications include: software engineering, network security, genetics, medicine, botany, and agriculture [5, 6]. Several CLIPS based programs have been recognized at the Innovative Applications of Artificial Intelligence Conferences [8, 9, 10, 11, 12]. CLIPS has been used as the core of at least three commercial products and two college level textbooks are available which come bundled with CLIPS.

CONCLUSION

Because of its portability, extensibility, capabilities, and low-cost, CLIPS has received widespread acceptance throughout the government, industry, and academia. The development of CLIPS has helped to improve the ability to deliver expert system technology throughout the public and private sectors for a wide range of applications and diverse computing environments.

REFERENCES

1. *CLIPS Reference Manual*, Version 6.0, NASA document JSC–25012, Houston, Texas, June 1993.

2. Giarratano, J., and Riley, G. *Expert Systems: Principles and Programming*, 2nd Edition, Boston, PWS Publishing Company, 1994.

3. Brooke, T. "The Art of Production Systems," *AI Expert*, January 1992.

4. Mettrey, W. "A Comparative Evaluation of Expert System Tools," *Computer*, February 1991.

5. *Proceedings of the First CLIPS Conference*, Houston, Texas, August 1990.

6. *Proceedings of the Second CLIPS Conference*, Houston, Texas, September 1991.

7. Forgy, C. "Rete: A Fast Algorithm for the Many Pattern/Many Object Pattern Match Problem," Pages 17-37, Artificial Intelligence 19, (1982).

8. Franier, R., et al. "PI-in-a-Box: A Knowledge-based System for Space Science Experimentation," *Proceedings of the Fifth Innovative Applications of Artificial Intelligence Conference*, July 11–15, 1993, Washington, D.C.

9. Robey, B., et al. "The DRAIR Advisor: A Knowledge-Based System for Materiel Deficiency Analysis," *Proceedings of the Fifth Innovative Applications of Artificial Intelligence Conference*, July 11–15, 1993, Washington, D.C.

10. Dutton, T. "HUB SIAASHING: A Knowledge-Based System for Severe, Temporary Airline Schedule Reduction," *Innovative Applications of Artificial Intelligence 4*, Klahr, Philip, and Scott, A. Carlisle ed., 1992.

11. Muratore, J., et al. "Space Shuttle Telemetry Monitoring," *Innovative Applications of Artificial Intelligence*, 1989, AAAI Press/The MIT Press, Menlo Press, Schoor, Herbert, and Rappaport, Alain ed.

12. Loftin, B., et al. "An Intelligent Training System for Space Shuttle Flight Controllers," *Innovative Applications of Artificial Intelligence*, 1989, AAAI Press/The MIT Press, Menlo Press, Schoor, Herbert, and Rappaport, Alain ed.

APPENDIX

CLIPS version 6.0 and CLIPS/Ada version 4.4 are currently available. CLIPS is free to NASA, USAF, and their contractors for use on NASA and USAF projects by calling the Software Technology Branch Help Desk between the hours of 9:00 AM to 4:00 PM (CST) Monday through Friday at (713) 286-8919. The STB Help Desk can also be reached via electronic mail at <stbprod@fdr.jsc.nasa.gov> or by FAX at (713) 286-4479. Government contractors should have their contract monitor call the Software Technology Branch Help desk to obtain CLIPS. Others may purchase CLIPS from COSMIC at a nominal fee ($350 for the source code and $115 for printed documentation as of August 1993). Price discounts are available to U.S. academic institutions. Further information is available by calling COSMIC at (706) 542-3265.

A Feasibility Study for Long-Path Multiple Detection Using a Neural Network

G.A. Feuerbacher, Texaco Inc. and T.A. Moebes, SAIC

S21-63
44824

Section 0.0 Abstract

Least-squares inverse filters have found widespread use in the deconvolution of seismograms and the removal of multiples. The use of least-squares prediction filters with prediction distances greater than unity leads to the method of predictive deconvolution which can be used for the removal of long path multiples.

The predictive technique allows one to control the length of the desired output wavelet by control of the predictive distance, and hence to specify the desired degree of resolution. Events which are periodic within given repetition ranges can be attenuated selectively. The method is thus effective in the suppression of rather complex reverberation patterns.

A back propagation(BP) neural network is constructed to perform the detection of first arrivals of the multiples and therefore aid in the more accurate determination of the predictive distance of the multiples. The neural detector is applied to synthetic reflection coefficients and synthetic seismic traces. The processing results show that the neural detector is accurate and should lead to an automated fast method for determining predictive distances across vast amounts of data such as seismic field records. The neural network system used in this study was the NASA Software Technology Branch's NETS system.

Section 1.0 Introduction

The Wiener filter (least-squares inverse filter) is one of the most effective tools for the digital reduction of seismic traces. It is the most important element of many deconvolution methods. In one application this filter is used to deconvolve a reverberating pulse train into an approximation of a zero-delay unit impulse. More generally it is possible to arrive at Wiener filters which remove repetitive events having specified periodicity. Multiples are such events and the periodicity are the arrival times or "predictive distances" of the multiples.

In this paper we develop a method using the BP neural network to detect multiples and their first arrivals. This would enable us to automatically determine predictive distances for each seismic trace and thus remove multiples more accurately and with a minimum effect on good data. The neural detector is applied to synthetic reflection coefficients and synthetic seismic traces. The processing results show that the neural detector is accurate and should lead to an automated fast method for determining predictive distances across vast amounts of data such as seismic field records. The neural network system used in this study was the NASA Software Technology Branch's NETS system.

Section 2.0 Synthetic Data: Reflection Coefficients

Figure 1. represents an idealized noise-free model of an offshore seismic situation(Peacock and Treitel, 1968). Reflector 1 is the water surface, reflector 2 is the water bottom, reflector 3 is a strong interface beneath the water bottom, and S is the surface location just beneath the water surface. The associated normal incidence reflection coefficients are 1, c1, and c2, respectively, while the transmission coefficient across reflector 2 is t1. If c is the downward reflection coefficient, the corresponding upward reflection coefficient is -c. From physical considerations, we know that the magnitude of all reflection coefficients are less than unity. Figure 2. shows the deconvolution of a first-order ringing system

$$(1,0,0,\ldots n..0,-c1,0,0,\ldots n..0,c1^2,0,0,\ldots n..0,\ -c1^3,0,0,\ldots n..0,\ c1^4,\ 0,0,\ldots n..0\ \ldots\ldots\ldots)\ \ldots\ldots\ldots\ldots(I)$$

where n is the predictive distance. The deconvolution operator that removes the multiples $-c_1, c_1^2 -c_1^3, c_1^4)$ is $(c_1,0,0,...n..0,1)$. The results after deconvolution is the multiple free signal (1,0,0,...n..0).

Figure 3. shows the deconvolution of a second-order ringing system

$(1,0,0,...n..0,-2c_1,0,0,...n..0,3c_1^2,0,0,...n..0, -4c_1^3,0,0,...n..0, 5c_1^4, 0,0,...n..0 ....)$ ....(II)

where n is the predictive distance. The deconvolution operator that removes the multiples $-2c_1, 3c_1^2 -4c_1^3, 5c_1^4)$ is $(c_1^2,0,0,...n..0,2c_1,0,0,...n..0,1)$. The results after deconvolution is the multiple free signal (1,0,0,...n..0).

We simulated the neural detector for the above synthetic seismic trace (I) above by training an NN with data of the form $(-c_1, c_1^2 -c_1^3, c_1^4)$, $0 < c_1 < 1$, as input where +0.5 indicated a multiple, and with data not of the form $(-c_1, c_1^2 -c_1^3, c_1^4)$, $0 < c1 < 1$, as input where -0.5 indicated an event that is not a multiple. The topology of the NN is shown in Figure 4. The network has four input nodes, two hidden nodes, and one output node. An example of training input indicating a multiple is (-0.5, 0.25, -0.125, 0.0625) for input and +0.5 for output. An example of training input indicating a non-multiple is (-0.5, 1.25, -0.145, 0.1117) for input and -0.5 for output. Another example of training input indicating a non-multiple is (2.0,-0.5, 0.25, -0.125) for input and +0.5 for output. This last example would simulate a window moving over a multiple but not quite covering the multiple. We trained the network with 5 multiples and 5 non-multiples and achieved 100% accuracy for 30 test cases.

We performed a similar experiment for a second-order ringing system and also achieved 100% accuracy.

Section 3.0 Synthetic Data: Seismic Data

Marine seismic data are frequently plagued by the presence of multiple reflections from the water bottom and by water-layer peg-leg multiples. This problem is especially severe in "hard bottom" areas where the reflection coefficient at the water-sediment is large, resulting in high-amplitude multiple reflections. Essentially, the water-bottom multiples arise because the water layer acts as a wave guide, resulting in repetitions of the water-bottom bounce. The peg legs arise from primary reflections that take an extra bounce or two in the water layer. There are two approaches to suppression of multiples, each depending upon one of the two distinguishing characteristics of multiples, namely periodicity and velocity.

At shallow water depths (say, less than 250 ft.) the multiples are periodic, after normal moveout(NMO) correction, with a repetition period equal to the two-way travel time through the water layer. Predictive deconvolution is very effective in suppression of such multiples.

When the water is deep, successive multiples are no longer periodic at far offsets, nor do they have the proper amplitude relations for predictive deconvolution to be successful. However, multiples typically spend more of their travel time in the lower velocity water layer than the primaries do, and as such have lower velocities. The differential moveout between multiples and primaries caused by the difference in their velocities has been successfully exploited in the common depth point (CDP) stacking scheme, as well as in exponential stacking routines.

At intermediate water depths (250 -1250 ft.) the multiples are not periodic, nor is the velocity difference between the primary and its associated (peg legs) multiples sufficient to allow for the application velocity-separation schemes mentioned above. The non periodicity of the multiples is shown on a synthetic field record in Figure 5, where the spread is over

10,000 feet and the two-way water-bottom time is 200 msec. The two-way travel time to the primary is 1 second, and the velocity of the primary event is 6000 feet/second. Notice that on the near trace the period of the (peg leg) multiples is 300 msec, whereas on the far trace the multiples are no longer periodic. This synthetic field record confirms the lack of periodicity of the multiples at a large offset, which precludes the use of conventional deconvolution schemes. Deconvolution schemes operate with a specified distance dependent upon the time span between a primary and its first multiple, and this time span is assumed to apply between successive multiples.

Observing that (peg-leg) multiples on intermediate water depths are no longer periodic with increasing offset Hilderbrand(1978) proposed a non periodic form of prestack gapped deconvolution to attenuate (peg leg) multiples. This report uses synthetic data to demonstrate the feasibility of such a deconvolution procedure being improved upon and extended to deep water depths using a neural network to determine the predictive distance for multiples. Extensive recording of a program originally coded by Hildebrand would be required.

Section 3.1 Description of The Synthetic Seismic Traces.

1) The "water layer" model(field record) was created by generating spikes with a normal move-out equation and a model velocity of 5000 feet per second. The reflection coefficients were determined by the equation

$$r = (V1 - V2)/(V2 + V1), \text{ neglecting density.}$$

The half spaces were assumed to have velocities of 1100 ft./sec.(air) and 6500 ft./sec/ (rock). The data is sampled every two milliseconds. Convolution of the spikes(reflection coefficients) with a minimum phase wavelet produced the final seismic trace(see Figure 6.). The first trace of the primary starts at 200 msecs and extends to 300 msecs. The first trace of the first multiple starts at 400 msecs and extends to 500 msecs. The first trace of the second multiple starts at 600 msecs and extends to 700 msecs(the second multiple is faint). We trained a BP network with 70 input nodes, 20 hidden nodes, and 2 output nodes to detect the multiples. Traces 1, 3, 5, 10, and 15 were input as part of the training data. The first 70 samples from each trace multiple were used. For the output nodes, true was indicated as +0.5 and false was indicated as -0.5. Five traces that were not multiples were generated and entered as training data in a similar manner. This gave the network a grater training scope. For test data we entered the first 70 samples of the remaining multiples and five new non multiple examples. We created one network for the first-order multiples and a second network for the second-order multiples. We achieved 100% accuracy in both tests.

2) The Three-Layers Model (20 trace field record) data were generated by a wave-equation program (PARX) at Texaco, Incorporated. The assumed reflection times (at zero-offset) were 350 msec, 500 msec, and 750 msec. The velocities of the three rock formations were 5000 ft./sec., 6500 ft/sec., and 9500 ft./sec., respectively(see Figure 7).

The three primaries start at 400, 500, and 750 milliseconds, respectively. The three corresponding first-order multiples start at 1200, 1450, and 1650 milliseconds, respectively. The topology and setup of the network was the same as in the "water layer" model. We worked only with first-order multiples. Three networks were created for each of the three different primaries' respective multiples. We achieved 100% accuracy in all three tests.

The topology for the above networks is shown in Figure 8.

Section 4.0 Mathematics Of Gapped, Predictive Deconvolution.

For zero-offset traces, and for normal incidence ray paths, peg legs from a primary reflector are periodic, the period being equal to the two-way travel time in the water-bottom(Backus, 1959). Mathematically, the (peg-leg) multiple generation process can be described by the difference equation:

$$m(n) - -2Rm(n-T) - (R**2)mn(n-2T) + p(n)$$

where

n = sample number for some fixed sampling interval
m = composite(primary + pegleg) signal received
R = water-bottom reflection coefficient
T = two-way travel time in water layer in samples
p = primary signal giving rise to peg leg) multiples

and,

$$p(n) = m(n) + 2Rm(n - T) + (R**2)(n - 2T).$$

The primary p(n) can be recovered, theoretically, from a weighted sum of the trace and its delayed samples. This is an finite impulse response operator with two gaps of T each-hence the terminology "double gapped " operator. Non periodic models of p(n) and m(n) have been postulated where both m(n) and p(n) are functions of T1 and T2 where T1 is the separation between primary and first (peg-leg)multiple in sample intervals, and T2 is the separation between (peg-leg)first multiple and second (peg-leg) multiple in sample intervals. If T1 = T2, this later model collapses to the periodic (peg-leg) multiple model.

The arrival time of the (peg-leg) multiples can be computed using Dix's formula in intermediate water depths. Applications have shown many inaccuracies in intermediate depth water. It is even recommended not to use this approach in deep water due to inaccurate calculations of T1 and T2, the predictive arrival times for the (peg-leg) multiples. Our feasibility study suggests that the predictive arrival times can be found automatically by identifying the multiples with a neural network. We could approximate the location of the multiples with Dix's equation, and fine tune their location with the neural networks.

Section 5.0 Summary

Neural Networks are now part of the leading edge in Geophysical data processing and interpretation. They have been recently successful in locating subsurface targets (Poulton, et al, 1992) and in obtaining seismic reflectivity sequences from seismic data(Wang, 1992). In this study we have shown the feasibility of developing a BP neural network for estimating the predictive distances of multiples where other traditional methods are not as adequate as desired. This seems to be particularly true in the case of deep water bottom multiples. The simulation data processing results showed that 1) the accuracy of the predictive distance of multiples can be enhanced with use of a neural detector over existing methods, and 2) the software implementation could be much faster since NN applications are potentially faster than traditional numerical and statistical methods. We look forward to the opportunity to implement an NN application and apply it to real data.

FIGURES

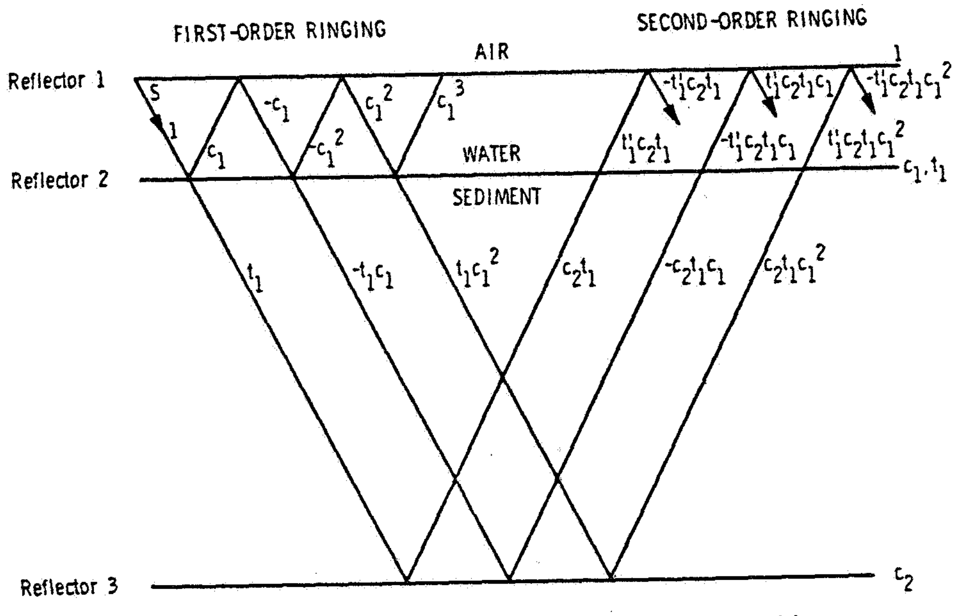


FIG. 1. First- and second-order ringing in a 2-layer marine model.

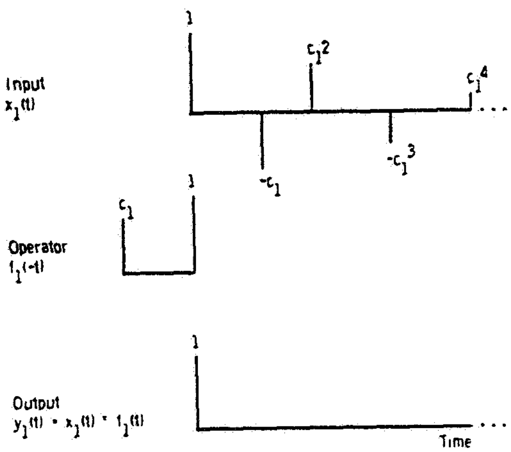


FIG. -2. Deconvolution of a first-order ringing system. The operator is shown in time-reversed form.

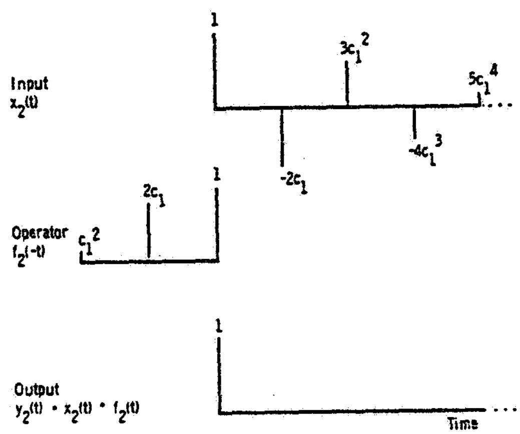


FIG. 3. Deconvolution of a second-order ringing system. The operator is shown in time-reversed form.

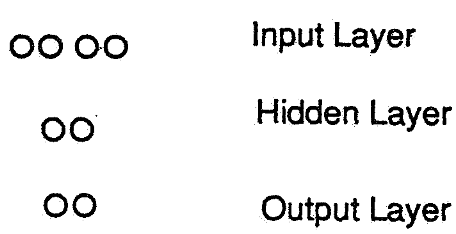


Figure 4. NN Topology For The Reflection Coefficients Data

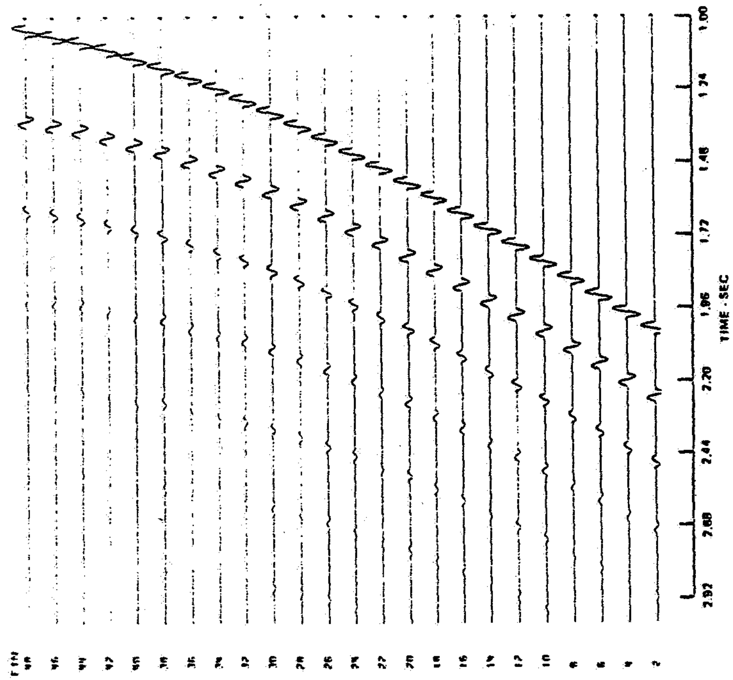


Figure 5. Synthetic gather with primary and multiple reflections

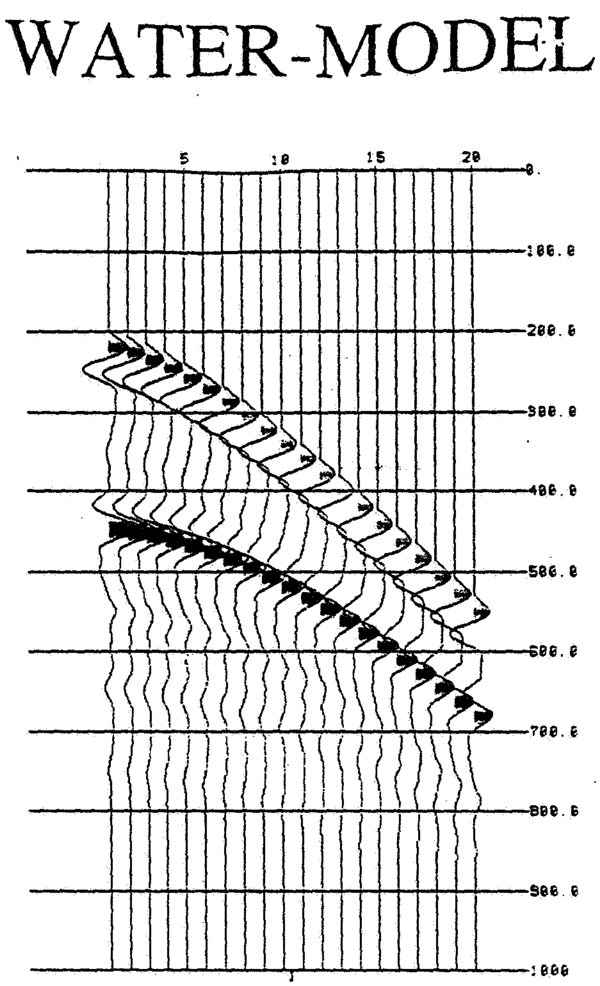


Figure 6. Water-Layer Model of a 20 Trace Field Record

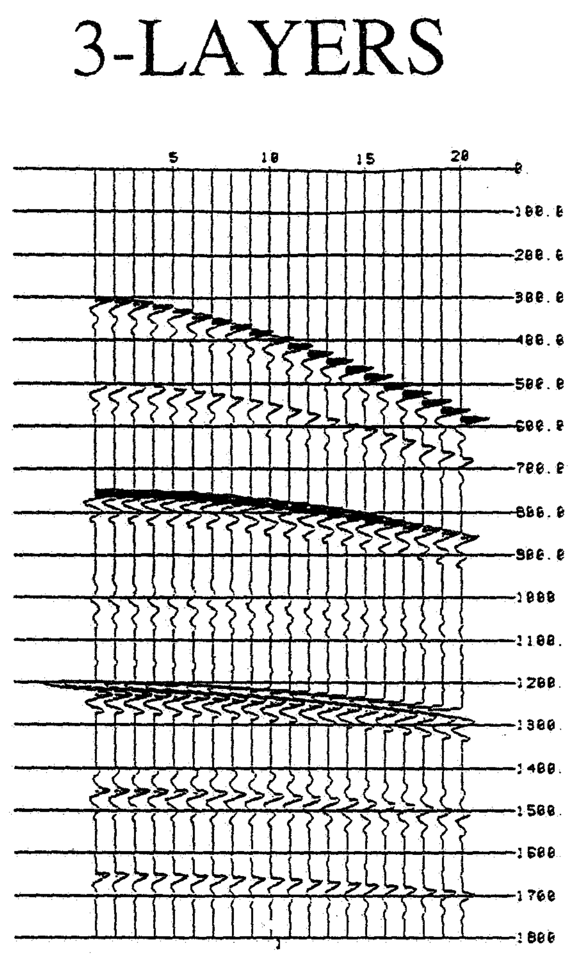


Figure 7. Three Layer Model of a 20 Trace Field Record

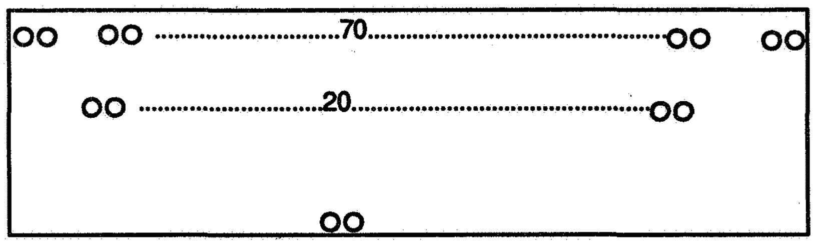


Figure 8. Network Topology For Multiple Detection BP Network on Synthetic Seismic data

REFERENCES

1. Bachus, M.M. (1959), Water reverberations-Their nature and elimination. Geophysics, vol.24, p. 233-261.

2. Hildebrand, H.A. (May 1980), The estimation of primary waveforms in the presence of peg-leg multiples. Geophysics, vol.12, p. 2353-265.

3.) Poulton M., Sternberg, and GLass, C.E.(December 1992) Location of subsurface targets in geophysical data using neural networks. Geophysics, vol.57, p. 1534-1544.

4) Wang, Li-Xin, A neural detector for seismic reflectivity sequences. IEEE Transactions On Neural Networks, Vol 3, No. 2, March 1992, p. 338-340.

S22-61
44827

Intelligent Computer-Aided Training Authoring Environment

Robert D. Way
way@gothamcity.jsc.nasa.gov
(713) 483-1899

LinCom Corporation
1020 Bay Area Blvd. Suite 200
Houston, TX 77058

ABSTRACT

Although there has been much research into intelligent tutoring systems (ITS), there are few authoring systems available that support ITS metaphors. Instructional developers are generally obliged to use tools designed for creating on-line books.

We are currently developing an authoring environment derived from NASA's research on intelligent computer-aided training (ICAT). The ICAT metaphor, currently in use at NASA, has proven effective in disciplines from satellite deployment to high school physics. This technique provides a personal trainer (PT) who instructs the student using a simulated work environment (SWE). The PT acts as a tutor, providing individualized instruction and assistance to each student. Teaching in an SWE allows the student to learn tasks by *doing* them, rather than by *reading about* them.

This authoring environment will expedite ICAT development by providing a tool set that guides the trainer modeling process. Additionally, this environment provides a vehicle for distributing NASA's ICAT technology to the private sector.

INTRODUCTION

"Industrial, business and commercial training accounts for about half of the total educational expenditure in the United States."

(Richardson 1988)

Education, training and re-training are frequently documented as expensive and inefficient by the media. Recent research in intelligent tutoring systems (ITS) and intelligent computer-aided instruction (ICAI) are often promoted as the remedy to educational problems.

Anecdotally, these claims seem well founded. Schank's "Case-Based Teaching" (Riesbeck 1991) and Woolf's "Discourse Management" (Woolf 1991) demonstrations both show amazing ability to intelligently interact with students. Unfortunately, the theories behind these systems are still the subject of intensive research. It may take several years before training based on these metaphors is commercially available. At any rate, many ITSs have been around long enough to show an influence on the commercial market. Equally as innovative, and more often quoted, Anderson's Geometry and LISP Tutors, (Anderson 1985) Johnson & Soloway's PROUST system for Pascal programming (Johnson 1984), and Hollan & Hutchins' industrial trainer Steamer (Hollan 1984), were all well published successes before 1988.

Yet despite being updated with the latest multi-media effects and hyper-text links, most available educational products do not employ any of the instructional techniques pioneered by ITSs during the past decade.

One reason for this lack of transfer is that while there has been much research into intelligent tutoring systems, there are few authoring systems available that support intelligent tutoring concepts. For example, no commercially available authoring tool supports all five common components (Burns 1991) of an ITS: domain expert, instructional expert, student model, intelligent interface, and simulation.

Instructional designers are commonly obliged to use general purpose authoring tools. These tools, like Authorware and ToolBook, shield designers from the complexities of color graphics, digitized sound and video, but do nothing to assist in the intricacies of ITS design. Deprived of basic student modeling capabilities, designers are discouraged from creating systems which vary remediation based on the ability of each student. Worse yet, lacking new instructional metaphors, general purpose tools tend to promote systems which are either high-tech slide shows or on-line books.*

Intelligent Computer-Aided Training

We are developing an authoring environment derived from NASA's research on intelligent computer-aided training (ICAT). The ICAT instructional metaphor, currently in use at NASA, has proven effective in disciplines from satellite deployment to high school physics (see related work section). This technique provides a personal trainer (PT) who instructs the student within a simulated work environment (SWE). The PT acts as a personal tutor, providing individualized instruction and assistance to each student. Teaching in an SWE allows the student to learn tasks by *doing* them, rather than by *reading about* them.

Figure 1 shows a simulated environment developed to train astronauts in operating the Spacehab module. Concepts are presented in the personal trainer window on the right hand side of the screen. Students respond to the trainer by identifying and manipulating objects in the environment. Navigation around the shuttle is performed using the displays in the top right corner. Switches and dials are directly manipulated using the mouse.

Different fidelities of simulated environment and personal trainer are used depending on the needs of the training. In this case the personal trainer is implemented using only text. Other systems have required more personality. In these cases the PT is presented using video of a real person. This helps capture the feeling of interacting with a human trainer.

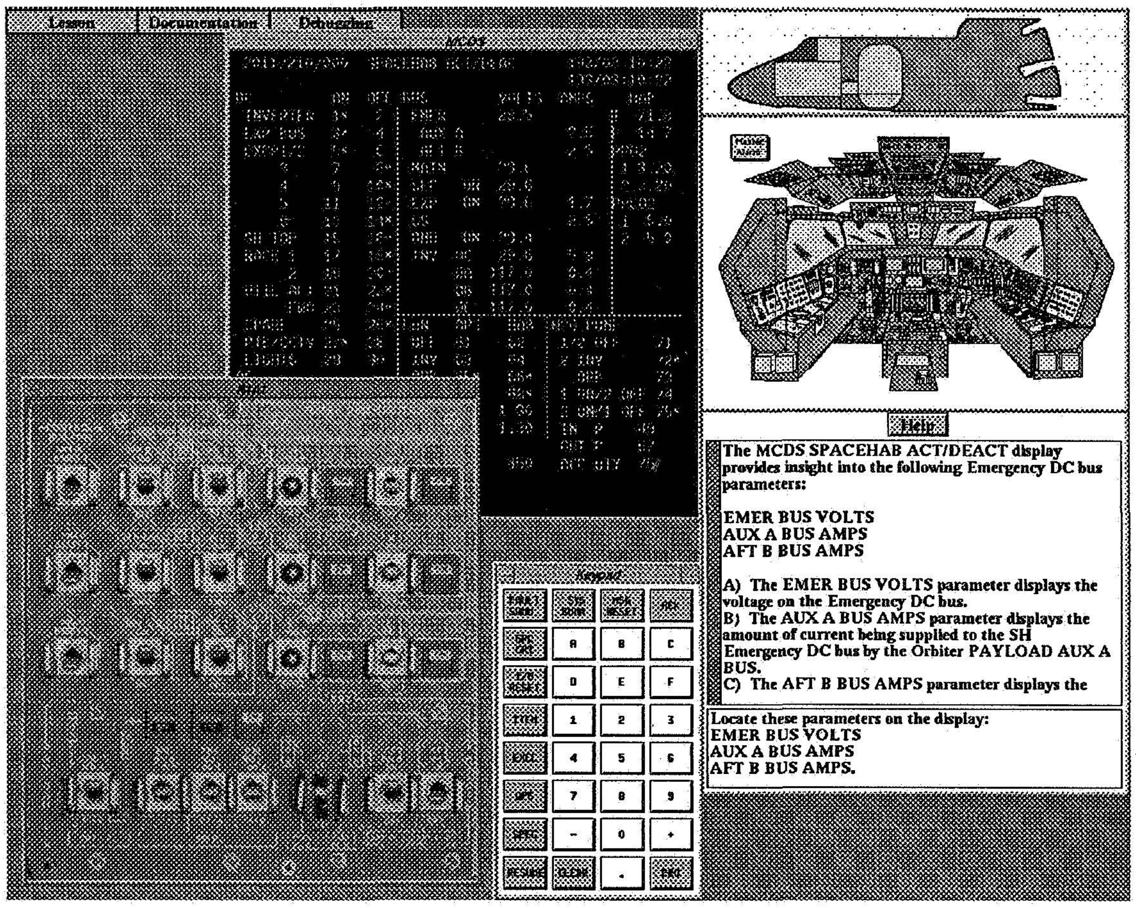


Figure 1
Simulated Environment for Space-Hab Training

* Capable instructional systems *have* been developed using current tools, but these systems owe more to the experience of the author than to the support of the tool.

ARCHITECTURE

The STB's authoring environment plans elaborate on NASA's second generation ICAT architecture (see Figure 2). NASA's architecture, shown in the lower level, defines which modules are necessary to build an ICAT, what each should do, and how they communicate. This architecture does not, however, specify *how* to implement each module. Previous ICAT system have used different methods of implementation. This project will provide a common library for all new ICAT systems and tools to guide the instructional designer through the development process.

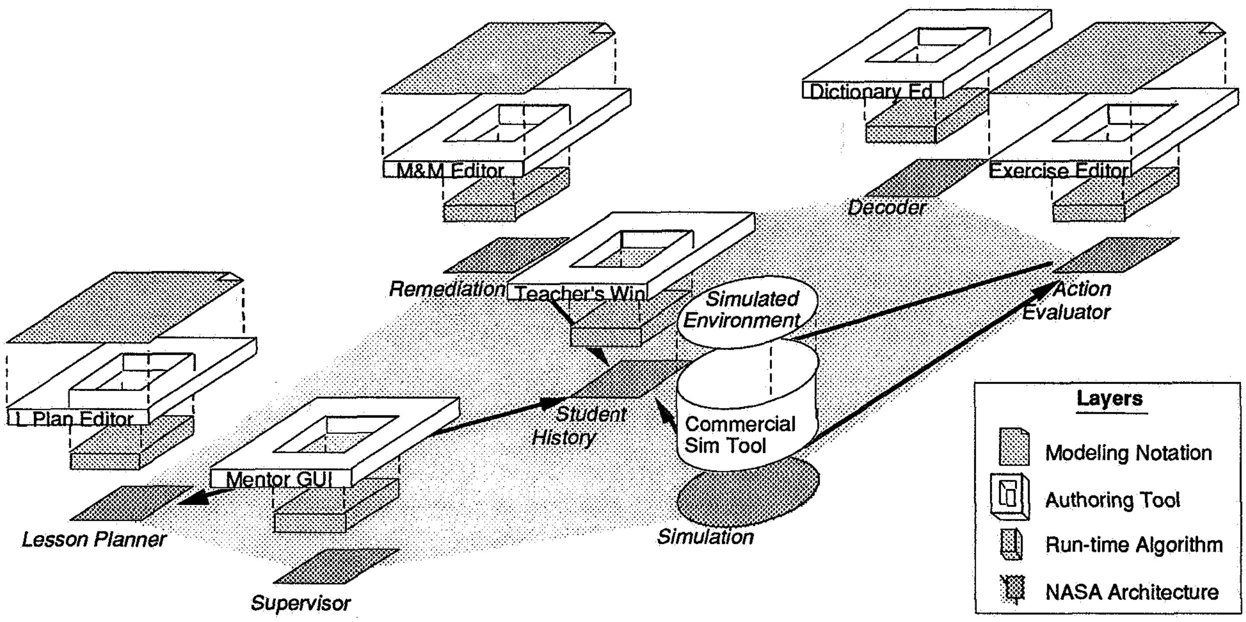


Figure 2
ICAT Architecture

Building an ICAT system based on this architecture requires both simulating the work environment and modeling a personal trainer. Unlike previous systems, Space station ICATs will use a pre-existing simulated work environment built for Part-Task Trainers. To simplify this process we have minimized the interaction between the SWE and the personal trainer modules of the ICAT system. The SWE only interacts with the PT through the student history and the action evaluator. The SWE is shown as the lone round module in Figure 2.

Rectangular modules in Figure 2 represent sections of the personal trainer. A personal trainer has three main duties, each assigned to a separate module:

Lesson Planner: Assign appropriate exercises to each student.

Action Evaluator: Watch what the student does and compare it objectively to the exercise's expected behavior.

Remediation: Point out the student's mistakes and give assistance tailored the student's past performance.

The three other modules shown in Figure 2 provide basic support to these core routines. The **Student History** remembers what the student did, what the student knows, and what we told the student. The **Supervisor** provides the personal trainer's graphical interface with the student and coordinates execution of the ICAT. The **Decoder** acts as a mini database for the ICAT. It maintains a list of all actions, concepts, exercise names, messages, and misconceptions used in a particular ICAT. These support modules are well understood (for the purposes of this proposal) and are not discussed further.

After reviewing previous ICAT implementations and relevant literature, The STB has selected synergistic implementations for each ICAT module (Lesson Planner, Remediation, etc.). On top of each module, The STB has layered a graphical modeling notation. These notations enable an instructional designer to easily diagram a new ICAT design.

But, while these notations greatly aid in specifying ICAT knowledge, someone with programming knowledge of these particular algorithms must still hand translate the diagrams into computer data files. The main goal of this project is to construct an authoring environment which helps an instructional designer create a new ICAT system. This environment will provide graphical editors for each modeling notation, and automatically translate diagrams into the required data files. In addition, these editors will be integrated with the run-time algorithms. This will allow the editors to also act as debugging tools . While a student is working an exercise, the editors will automatically highlight the ICAT diagrams to show the student's progress.

The remainder of this document is broken into four sections, one explaining the SWE and three explaining to the core modules above.

Simulated Work Environment

Figure 3 shows one of the approximately twenty-five panels that make up the Spacehab Intelligent Facilities Trainer (SHIFT) Simulated Work Environment. Each of the switches, buttons and lights on the panel is active. An astronaut throws switches using the mouse. Indicator lights are controlled by an underlying engineering model of the Spacehab module.

Astronauts learn procedures which involve throwing a series of switches and checking for the appropriate indicators to light up. They also monitor the system, listening for alarms and looking for problem "signatures".

The GUI and engineering models of previous SWEs have required up to two-thirds of the total effort allotted to creating an ICAT system. By using pre-existing part task trainers we expect to dramatically reduce the development time of space station ICAT systems.

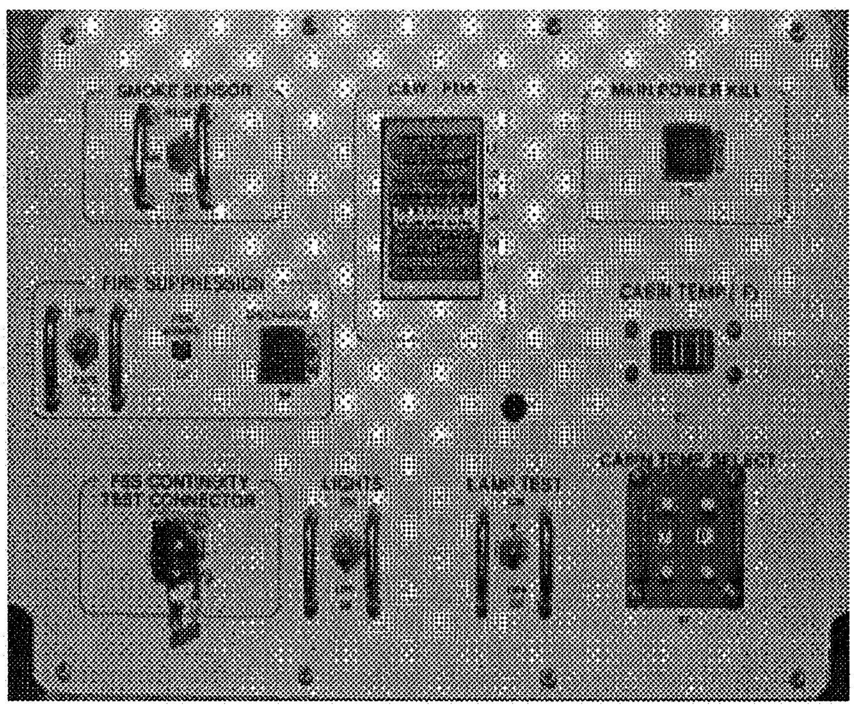


Figure 3
Panel from the SHIFT simulated work environment

Action Evaluator

Once an SWE is defined, the instructional designer must develop exercises to be performed in the environment. ICAT theory defines two different types of exercises: those which present new material to the student and those which apply the material. These concept application lessons also serve as the way the ICAT system evaluates the student's progress.

The STB has developed a graphical notation called operating procedure language (OPL) which enables an instructional designer to specify what things a student should do during a lesson. Figures 4 and 5 show examples of this notation. The STB has also implemented algorithms which allow the action evaluator to compare the student's behavior to the specified notation.

Figure 4 shows the OPL notation for an exercise which presents a new procedure to the student. Rectangles represent actions the student is expected to carry out. Rounded rectangles represent the sequence interactions given to the student. Circles represent commands to the simulated work environment. Notice that a presentation exercise normally has a linear structure.

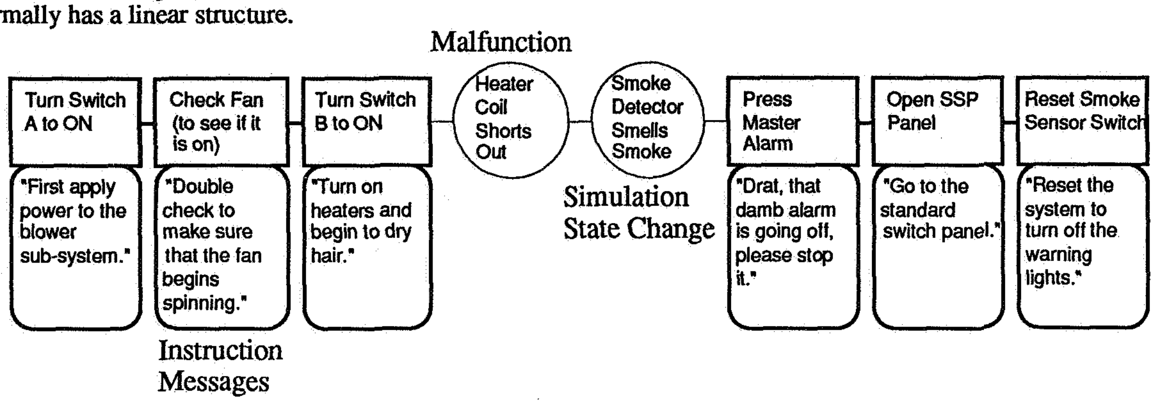


Figure 4
Material Presentation Exercise

Figure 5 shows an exercise which allows the student to practice a procedure. Practice exercises are not necessarily linear, they allow students to work a procedure using any equivalent series of steps. They also support branching of procedures based on the state of the SWE. Practice exercises may also be annotated to show common mistakes.

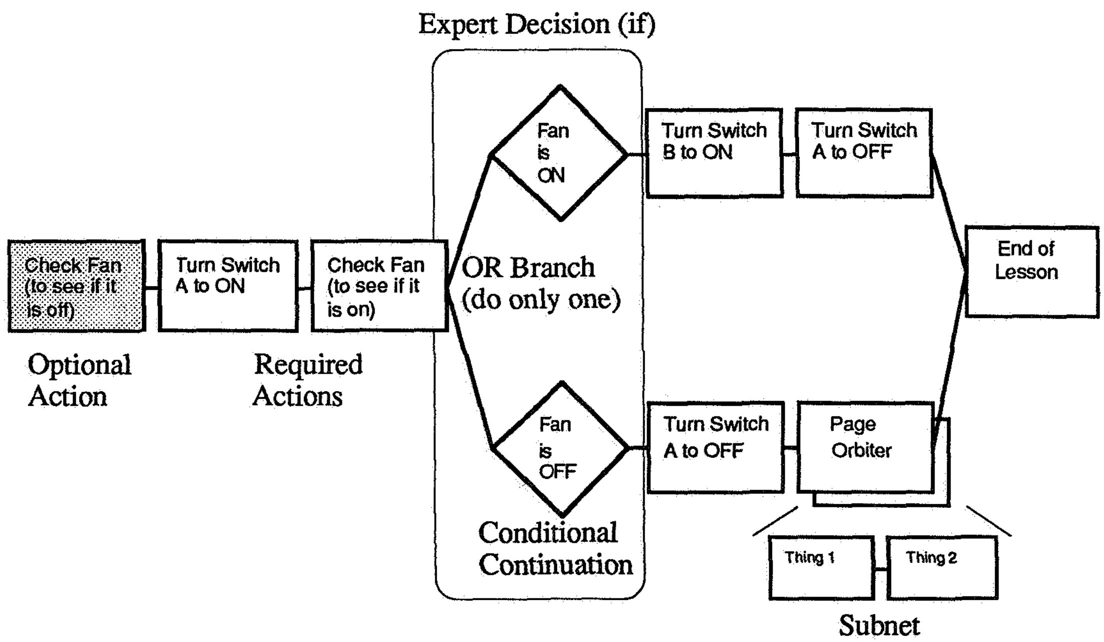


Figure 5
Practice/Evaluation Exercise

During the SHIFT project, OPL greatly simplified the creation of exercises. Although OPL provides an efficient way of specifying exercises, each exercise must be translated by hand into the data files required for implementation. The new exercise editor will facilitate both drawing OPL diagrams and generating the data files.

Remediation

In the course of working an exercise, the student will complete a number of actions. Each action, either correct or incorrect, gives information about the student's understanding of the material. The major part of trainer modeling is watching the series of actions that the student performs, and identifying patterns of behavior which signify a misunderstanding of the material. Once a misunderstanding is identified, the trainer gives the student an appropriate message.

Trainer modeling is based on the assumption that we have available a human "trainer" who understands the material and could teach it well in a one-on-one setting. If no one exists who could teach the proposed material we have found it extremely difficult to create an effective ICAT system. If the assumption is true, however, trainer modeling simply reduces to "doing what the human trainer would do."

In creating various ICAT systems, we have talked to many teachers and trainers. These experts have told us that each student comes into the training program with a personal conceptual model of the world. This conceptual model drives the student's assimilation of new material. If the student's conceptual model is correct relative to the material being taught, the assimilation is usually easy. If the student's conceptual model is wrong, however, it can often lead to misinterpretation of the material. These misinterpretations often cause the student to make mistakes during an exercise.

Fortunately, many students have similar backgrounds and, therefore, have similar conceptual models of the world. Because these similar models cause students to make similar mistakes, good teachers learn to quickly spot behavior patterns which indicate misinterpretations of the material. This is a principle called "Misconception Theory." (Way 1991) Misconception theory holds that teachers can easily describe what common mistakes students make, what misconceptions causes these mistakes, and how to remediate them.

The remediation module of the personal trainer implements this process. The STB has developed an efficient matching algorithm which stores patterns of behavior (student actions) and maps them to misconceptions. This algorithm is facilitated by imbedding the pattern matching in a binary search tree (see Figure 6). In addition, the algorithm provides for structured relationships between misconceptions. This allows the system to "fallback" and remediate a more general concept if several related misconceptions have been diagnosed (see Figure 7). Thirdly, it stores one or more types of remediation for each misconception. This allows the system to give appropriate levels of explanation based on the student's background.

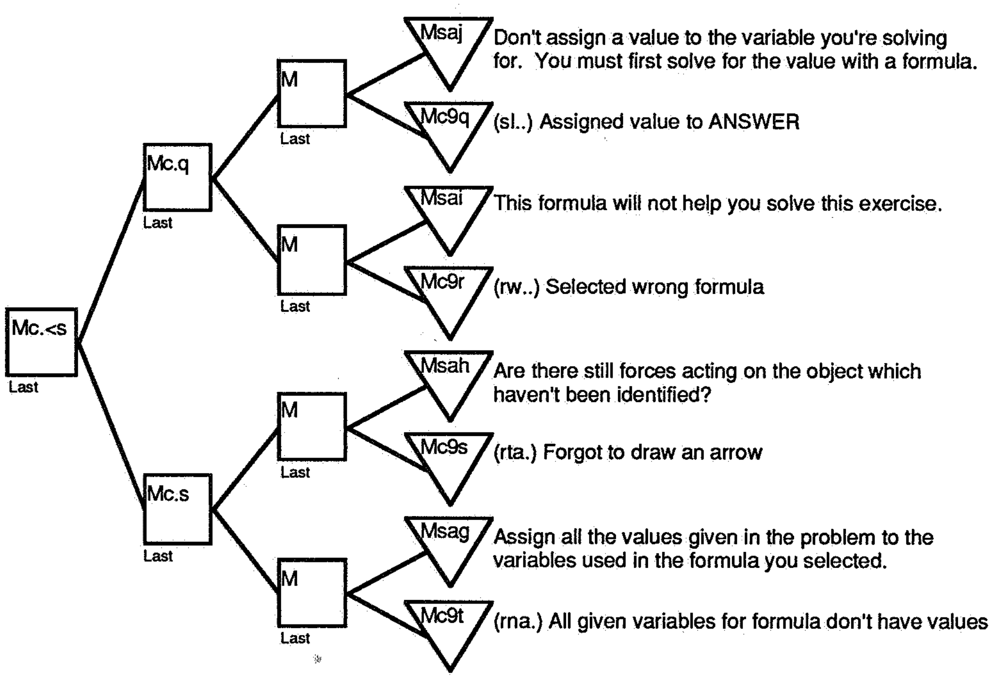


Figure 6
Remediation Pattern Matching Tree

This remediation mechanism was used with great success in the Intelligent Physics Tutor. However like OPL notation, the misconceptions must be translated by hand into the tree notation shown in Figure 6. This project will implement an editor to automate this process.

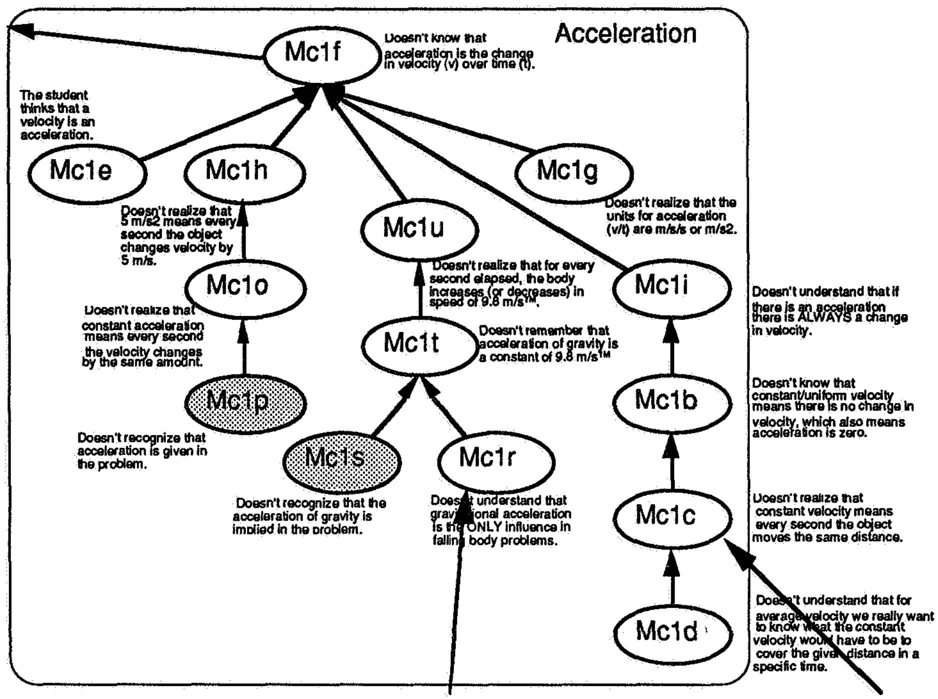


Figure 7
Related Misconceptions

Lesson Planner

The STB has developed a second algorithm, as a part of misconception theory, which allows the personal trainer to select appropriate exercises for each student. Exercise selection is based on knowledge which the personal trainer gained from watching the student complete previous exercises. This process is implemented as the Lesson Planner module.

Selecting an appropriate exercise requires indexing meta-knowledge about each exercise. Meta-knowledge includes concepts that must be understood to successfully complete the exercise and misconceptions commonly revealed during this exercise.

As the student works each exercise, the lesson planner uses this meta-knowledge to compile a list of the concepts the student has applied. The lesson planner also builds a list of misconceptions diagnosed by the remediation module. Using these two lists, the lesson planer creates a third list recommending concepts for future study. Selecting exercises simply becomes a matter of matching the concepts recommended to the exercise's meta-knowledge (see Figure 7).

This technique was used with great success in the Intelligent Physics Tutor. Like OPL and misconceptions, these meta-data indexes must be translated by hand into executable data files. This project will develop a tool for graphically editing these data files.

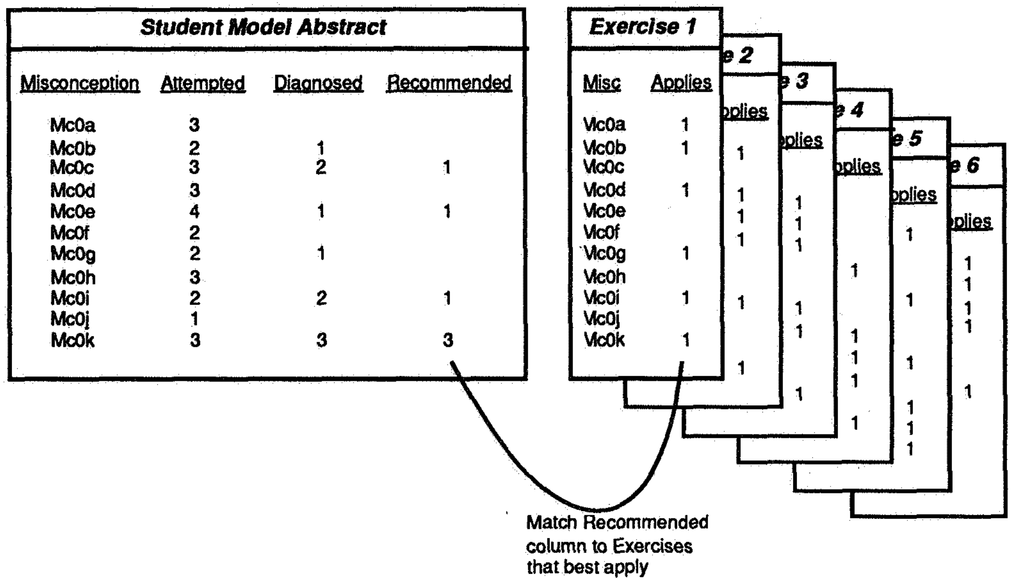


Figure 7
Meta-Knowledge Index

CONCLUSIONS

The ICAT authoring environment will allow space station trainers to develop ICAT training without outside programming help. By eliminating dedicated knowledge engineers and reusing existing simulations, we should see an inherent 50% reduction in development cost. This coupled with ICATs ability to provide more people, more training in an equivalent time, will provide a substantial cost savings in required personnel. Additionally, when completed, this environment will represent a product which could distribute NASA's ICAT technology to the private sector.

REFERENCES

Anderson, J.R., Boyle, C.F., & Yost, G. (1985). The geometry tutor. In A. Joshi (Ed.), Proceedings of the Ninth International Joint Conference on Artificial Intelligence (pp. 1-7). Los Altos, CA: Morgan Kaufmann.

Burns, H., & Parlett, J.W. (1991). The Evolution of Intelligent Tutoring Systems. *Intelligent Tutoring Systems Evolutions in Design.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Hollan, J.H., Hutchins, E.L., & Weitzman, L. (1984). Steamer: An interactive inspectable simulation-based training system. *AI Magazine*, 5, 15-27.

Johnson, L., & Soloway, E. (1984). Intention-based diagnosis of programming errors. *Proceedings of American Association of Artificial Intelligence Conference* (pp. 162-168). Los Altos, CA: Morgan Kaufmann.

Richardson, J.J. (1988) Directions for Research and Applications (page 251). *Foundations of Intelligent Tutoring Systems* . Hillsdale, NJ: Lawrence Erlbaum Associates.

Riesbeck, K.C. & Schank, R.C. (1991) From Training to Teaching: Techniques for Cased-Based ITS. *Intelligent Tutoring Systems Evolutions in Design.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Way, R.D. (1991) An Intelligent Tutoring System for Physics Problem Solving. *NASA's 1991 Conference on Intelligent Computer Aided Training,* NASA Conference Publication 10100, Vol.I

Woolf, B. (1991) Representing, Acquiring, and Reasoning About Tutoring Knowledge. *Intelligent Tutoring Systems Evolutions in Design.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Intelligent Computer-Aided Training Authoring Environment

by

Robert D. Way
LinCom Corporation
1020 Bay Area Blvd. Suite 200
Houston, TX 77058

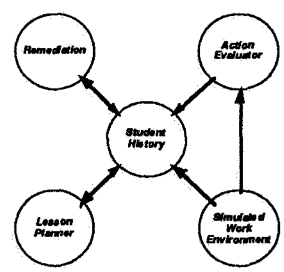
Road Map

What is ICAT?
How are ICATs built?
Why an authoring environment?
Authoring environment vision?

What is Intelligent Computer-Aided Training?

Simulated Work Environment
- Students learn by doing real tasks
- Presents new concepts as on-the-job training

Automated Personal Trainer
- Customizes lessons for each student
- Provides immediate feedback and help
- Summarizes student progress for the instructor



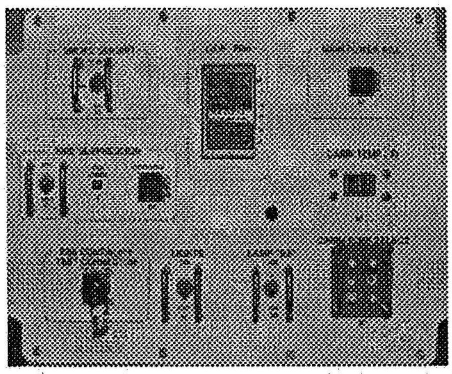
SHIFT

ICAT Tenets

Present the concepts in the real environment with the real objects. (Simulate the Work Environment)
Have the student do what you want him to learn.
Watch as the student does the job, take note of everything.
Watch quietly as the student works.
Give appropriate help when asked.
Point out errors while they're still in context.
Let the student recover from errors if possible.
Never give a formal "Test".

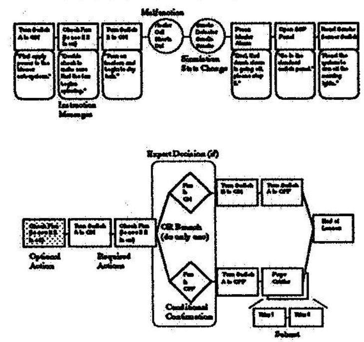
NASA Johnson Space Center
Information Systems Directorate
Information Technology Division

Simulated Work Environment



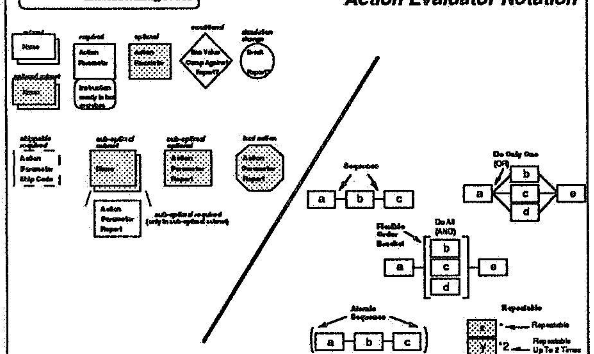
NASA Johnson Space Center
Information Systems Directorate
Information Technology Division

Action Evaluator Procedures



NASA Johnson Space Center
Information Systems Directorate
Information Technology Division

Action Evaluator Notation



NASA Johnson Space Center
Information Systems Directorate
Information Technology Division

Teaching Principles

Teacher's recognize a set of common misunderstandings.

They normally gear their responses more to these misunderstandings than to the students queries.

If their initial instinct appears incorrect they fall back to other commonly related misconceptions.

They double-check themselves by watching to see if the student exhibits other signs of the misconceptions.

NASA Johnson Space Center
Information Systems Directorate
Information Technology Division

Modelling the Teacher

Most of these misconceptions are easily recognizable from patterns of student actions.

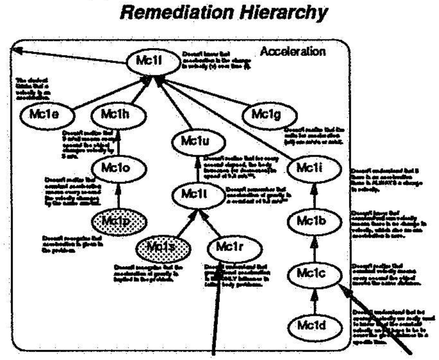
We then respond by mimicking a teacher who suspects a certain misunderstanding.

Once the misconceptions are known we are able to relate common ones as a teacher would.

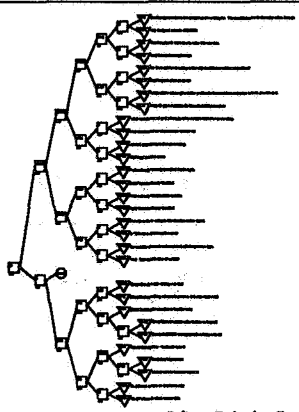
We then adjust the curriculum to double-check the student's understanding.

NASA Johnson Space Center
Information Systems Directorate
Information Technology Division

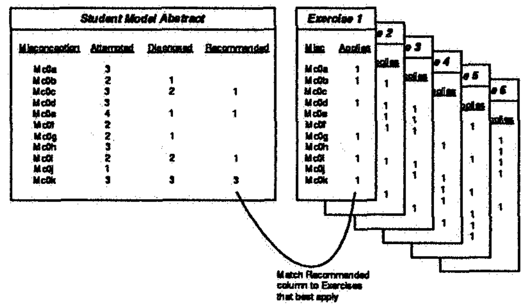
Remediation Hierarchy



Remediation Tree



Lesson Planner

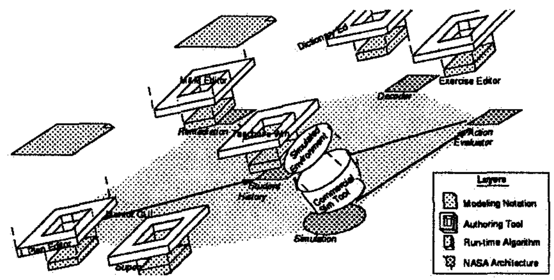


Why Authoring Tools?

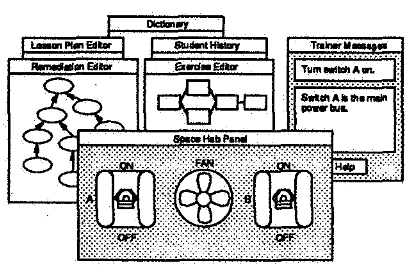
Current Systems
- Labor intensive to build
- Require programmers trained in ICAT to maintain

Space Station Training
- Multiple ICATs training five separate subsystems
- Must use pre-existing simulation models
- ICATs must be maintainable by without programming

Tool Architecture



ICAT Authoring Tool



Where's the Intelligence?

Immediate "teacher like" feedback to student actions.

Context sensitive help at all times.

Adjustment of material based on the demonstrated understanding of the student.

Teacher level summaries of both student and class progress.

Session S3: SOFTWARE ENGINEERING I

Session Chair: Susan Gerhart

REENGINEERING LEGACY SOFTWARE TO OBJECT-ORIENTED SYSTEMS

C. Pitman, D. Braley, E. Fridge
NASA/JSC/PT4
Houston, TX 77058

A. Plumb
LinCom, Inc.
Houston, TX 77058

M. Izygon, B. Mears
I-NET, Inc.
Houston, TX 77058

ABSTRACT

NASA has a legacy of complex software systems that are becoming increasingly expensive to maintain. Reengineering is one approach to modernizing these systems. Object-oriented technology, other modern software engineering principles, and automated tools can be used to reengineer the systems and will help to keep maintenance costs of the modernized systems down. The Software Technology Branch at the NASA/Johnson Space Center has been developing and testing reengineering methods and tools for several years. The Software Technology Branch is currently providing training and consulting support to several large reengineering projects at JSC, including the Reusable Objects Software Environment (ROSE) project, which is reengineering the flight analysis and design system (over 2 million lines of FORTRAN code) into object-oriented C++. Many important lessons have been learned during the past years; one of these is that the design must never be allowed to diverge from the code during maintenance and enhancement. Future work on open, integrated environments to support reengineering is being actively planned.

1. INTRODUCTION

Like many other organizations, NASA has a legacy of complex software systems that were developed during the past three decades and are still being used today. These systems are critical to NASA's mission; they represent an enormous investment; and they are very expensive to use, maintain, and enhance. The aerospace community can continue to maintain the systems, using improved tools and techniques, but at some point that will become ineffective: the skills for old languages will be harder to find; the few people who understand how to maintain the systems will retire; or the number of patches on top of patches will drive maintenance costs beyond all reason.

The systems could be scrapped and redeveloped, but it would take enormous amounts of time and money to redevelop them from first principles, because of the vast quantity of software involved, because the necessary engineering skills would have to be found and reassigned to the tasks, etc. More importantly, the systems themselves may be the only repository for detailed engineering knowledge of the application at the algorithm level, because system requirements documents are not kept up-to-date by most organizations as new requirements are continually added. If the current systems were scrapped, all of this engineering knowledge of the current algorithms and interfaces would be lost, and for mission-critical systems this knowledge was gained at great expense.

Another option is software reengineering, i.e., modernize the systems using up-to-date software technologies, in order to improve maintainability, without losing the embedded engineering knowledge or current capabilities. However, software reengineering itself is not a trivial task. It is a young field, and organizations need support (training, tools, etc.) in order to realize benefits. In particular, methods and tools to help reengineer software systems are becoming increasingly important.

2. REENGINEERING TERMS

Reengineering terminology is not always used in the same way by different authors, and so it is best to begin with brief definitions of some terms (Figure 1). Forward engineering refers to the usual direction of software development (whether a waterfall, spiral, incremental, or rapid prototyping process): begin with requirements, then design , code, test, and deliver. In contrast, reverse engineering refers to starting with the old code and recovering from it the essential semantics, the design, and/or the requirements. Reengineering is NOT synonymous with reverse engineering. Reengineering is the combination of reverse engineering followed by forward engineering into the new, modernized software system. (Actually, this definition is itself oversimplified: in many reengineering processes, the reverse engineering is not done completely in a single phase up front, but is done as needed and fed back into the forward engineering effort at various times.) Note that reengineering that goes all the way back to recover requirements is very similar to rebuilding the software system from scratch, but the difference is that the reverse engineering portion of the reengineering process helps to assure that all of the critical requirements which are captured only in the old code will appear in the new system.

3. OBJECT-ORIENTED TECHNOLOGY

Object-oriented methods provide a modern approach to building software. There are many methods available on the market today, such as those by Rumbaugh, Booch, Shlaer-Mellor, or Wirfs-Brock. Some-second generation methods, such as Fusion, are also beginning to emerge.

Probably, the most widely used method today is the Object Modeling Technique (OMT) by James Rumbaugh, et al. [1] OMT provides a method for incorporating good software engineering principles during the requirements analysis and design phases. Many of these principles, such as data abstraction, encapsulation and modularization, commonality, and reuse, have been discussed for many years now, but are still all too often ignored when the pressures of project deadlines loom close at hand.

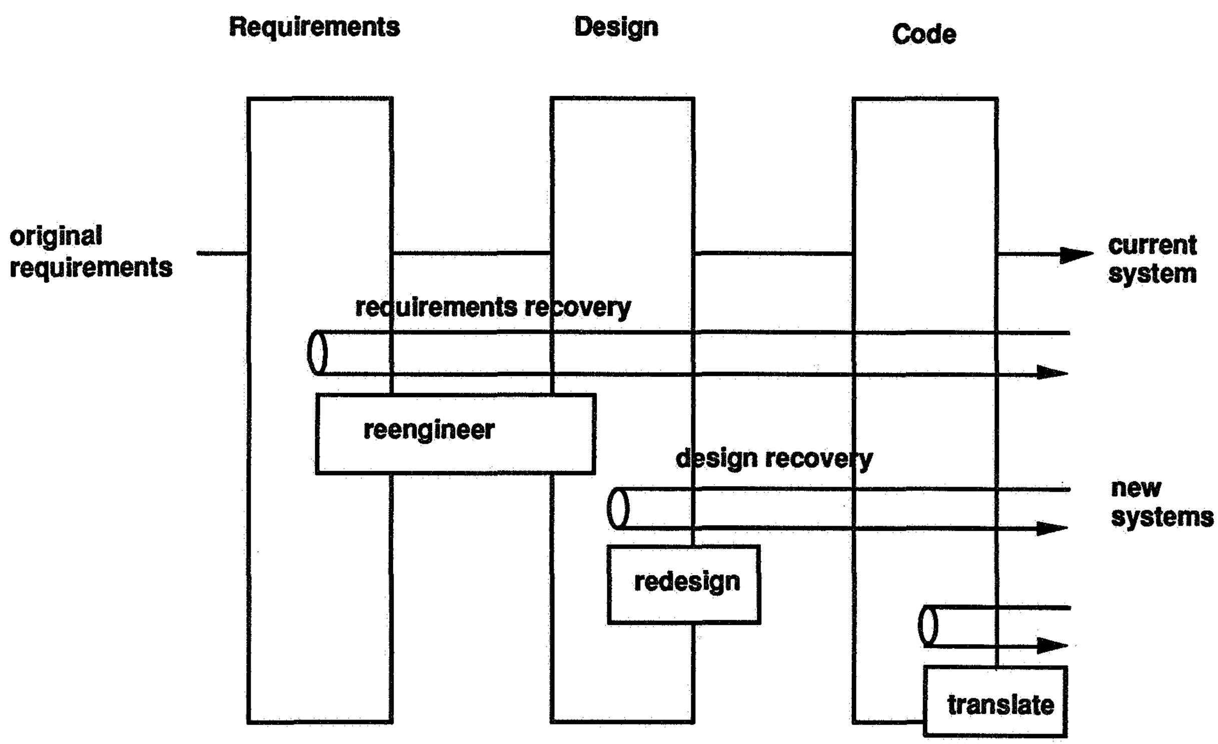


Figure 1. Reengineering is the combination of reverse engineering (analyzing the current code to recover its meaning, design and/or requirements) followed by forward engineering into the new, modernized software system. Three levels of reengineering are shown: translation, redesign, and complete reengineering.

OMT also helps to smooth the transition between life cycle phases. For example, OMT models provide a standard language that can be used for defining guidelines for transforming designs into code; indeed, some tools exist today that can automatically generate C++ "header" files directly from the OMT models.

Perhaps most importantly, large software engineering projects that are done by teams are inherently a social enterprise, and OMT models are an extremely effective communications tool for software development or reengineering teams.

4. TECHNOLOGY TRANSFER SERVICES

The Software Technology Branch (STB) helps teams to learn, adopt, and use reengineering and object technology. The STB has led seminars on reengineering and on OMT, and is providing mentoring and consulting support to some JSC reengineering projects. The STB also performs evaluations of Computer-Aided Software Engineering (CASE) tools for these projects, and has developed some evaluation criteria as a part of these efforts. The STB in turn gains from these experiences because they also offer opportunities for testing and refinement of reengineering methods on real applications.

5. REENGINEERING EXPERIENCES

The Software Technology Branch has been developing and testing reengineering methods and tools for several years. [2] The first project involved a large orbital mechanics program named the Orbital Maneuver Processor (OMP). The project recovered the design of the OMP FORTRAN code, modified the design into an object-based one, and then implemented the new design in Ada. [3, 4] Several tools that aid in the recovery of the design from FORTRAN code (e.g., COMMON block structures, calling trees, ...) were used and enhanced for this and following projects; these tools were originally developed by JSC during the period of development of the Space Shuttle flight planning software.

Work on support environments for reengineering has resulted in the Reengineering Applications (REAP) environment, which provides a uniform presentation and invocation of reengineering tools and a suggested sequence for their use. As mentioned above, the STB is providing reengineering support to other projects: a command system (Mission Operations Computer), a database reconfiguration system (Reconfiguration Tools), and a solar thermal analysis and optimization system (in conjunction with the University of Houston, for Sandia National Laboratories).

6. REENGINEERING SUPPORT FOR THE ROSE PROJECT

Another of the more significant reengineering projects with which the STB is associated is the Reusable Objects Software Environment (ROSE). This project is seeking to recover the requirements from the flight analysis and design system (over 2 million lines of FORTRAN code), enhance the requirements for reuse if necessary, and then redesign the system into an object-oriented one and implement it in C++. The first-year phase of this four-year project will be concluded in February, 1994; so far, very promising results have been obtained.

On the ROSE project, the reverse engineering method used for FORTRAN design recovery was developed over the course of a few years and then tailored to the ROSE project. Several Computer-Aided Software Engineering (CASE) tools are used to facilitate this process, such as data and control structure analysis tools, complexity metrics tools, and restructuring/refining tools (to a limited extent). The forward engineering part of the ROSE reengineering project is using the Object Modeling Technique (OMT), an Object-Oriented Analysis and Design (OOAD) method developed by James Rumbaugh, et al. OMT is excellent for Analysis, providing the basis for communication and clarification of the problem to be solved and of the requirements. (Section 3)

7. FILLING THE GAPS IN COMMERCIAL TOOLS

Sometimes commercially available tools do not support all of the steps in an organization's software development or reengineering process; this leaves holes or "gaps" in the process where custom-built tools might be needed. Usually, these gaps occur because many organizations' processes are more complex than those found in text books, and it is the text book processes that are most often supported by commercial tools.

Custom-built "gap" tools make sense when all of the following conditions occur:
(i) available tools do not support a specific part of the organization's process;
(ii) the gap tools will be used to automate process steps that would otherwise be performed manually;
(iii) the gap tools must save more time than they take to build and maintain; and
(iv) there should be easy, stable methods for interfacing to the commercial tools used by the organization.

For example (Figure 2), the ROSE project is reengineering a large FORTRAN system. The project has selected Reasoning Systems' Refine/FORTRAN as a FORTRAN front end, working with Software through Pictures (STP). However, Refine/FORTRAN does not completely handle the COMMON formats that ROSE has encountered, and one of the JSC Maintenance Aids, CREATE, was designed precisely for that type of COMMON format. The STB built a gap tool to integrate this information into the Refine/FORTRAN - STP communications, and thus to correctly show the internals of the COMMON blocks.

8. INTEGRATION TECHNOLOGY

As one of its more forward looking research projects, the STB has also begun investigating open, integrated reengineering environments. (Figure 3) Today, many companies are working on integrated support environments for forward engineering, but not too much work has been done on reengineering environments, even though the requirements are very similar. An integrated environment should provide presentation, data, control, and process integration. The National

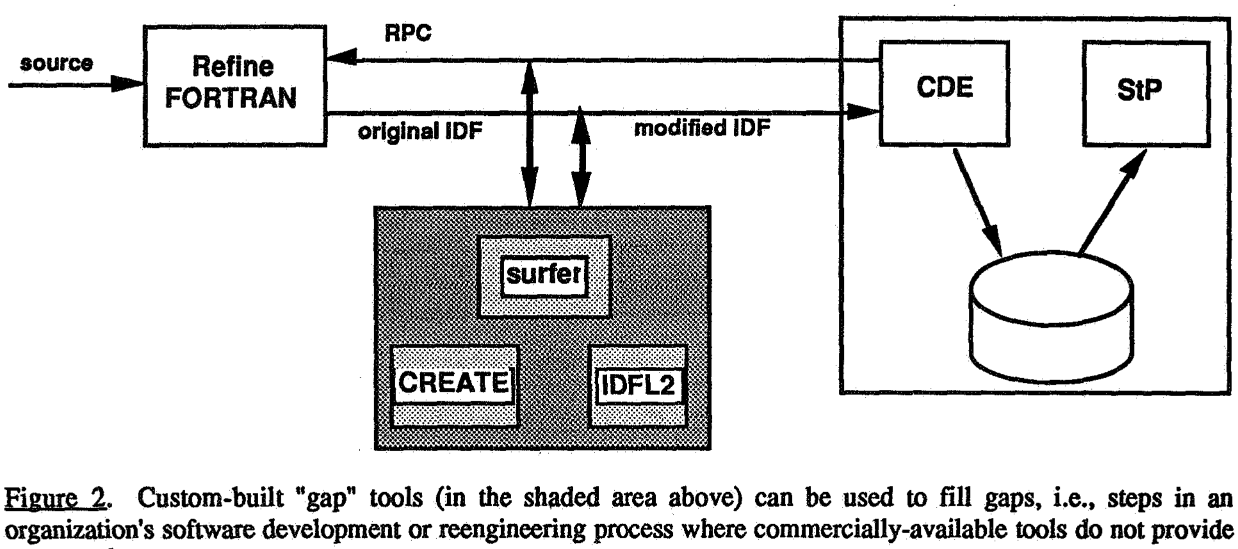


Figure 2. Custom-built "gap" tools (in the shaded area above) can be used to fill gaps, i.e., steps in an organization's software development or reengineering process where commercially-available tools do not provide automated support.

Institute of Standards and Technology (NIST) and the European Computer Manufacturers Association (ECMA) have proposed a standard model for discussing such an integrated environment. The STB, in conjunction with IBM, has worked to develop a reengineering data model for a repository, which would provide data integration in a reengineering environment based on the NIST/ECMA model. Much good initial work has been done to date, but this is a challenging problem and much work still remains. Nevertheless, the objective is to produce the REAP II environment, based on this integration work, within one to three years.

9. LESSONS LEARNED

One of the most important lessons learned over the past two decades of software development and maintenance at JSC is that the design must never be allowed to diverge from the code during maintenance and enhancement; i.e., the design must be kept closely tied to the code. For the ROSE project, it was discovered that, during design, OMT must be supplemented and extended to permit a closer tie between the design and the C++ code that will eventually implement it.

Another very important lesson learned is that it is critical to choose tools that fit the processes and data types of an organization and project, and not vice versa. All too often an organization will buy a tool and try to force-fit it into its processes, without considering the way it does business, much less if the processes themselves might need reengineering.

Finally, an important lesson learned is that it is critical to provide appropriate training in the methods, processes, tools, and language used for a project, BEFORE the project team members are expected to meet project deadlines and deliverables. The ROSE project adopted this approach and it is paying off. The STB was asked to help coordinate the ROSE training and to provide initial training in the OMT method.

10. CONCLUSIONS

The STB has been researching and developing reengineering of legacy software systems for many years, and it has pulled together some methods, training, and tool products that can greatly facilitate the tasks associated with reengineering legacy software. These products and services are proving useful for JSC software systems, and can be useful for reengineering software in other scientific and technical domains as well. The STB is committed to applying and transferring this

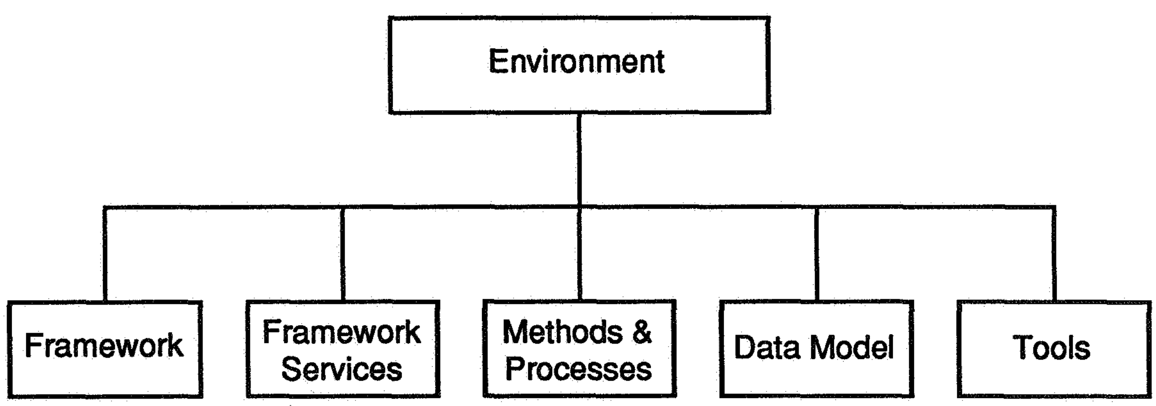


Figure 3. An environment to support software reengineering is composed of many parts. A framework is the backbone upon which an environment is built, and it supplies services like data repository services. Methods and processes for doing reengineering should drive the environment's configuration, and the data model, stored in the framework's repository, contains required information about the processes and tools. Tools automate some of the reengineering work.

reengineering technology to other projects, including those in industry. Please contact us if you have projects where this technology might help.

11. REFERENCES

1. Rumbaugh, J., Blaha, M., Premerlani, W., Eddy, F., and Lorensen, W., *Object-Oriented Modeling and Design*, Prentice-Hall, Englewood Cliffs, New Jersey, 1991.

2. Pitman, C. L., Erb, D. M., Izygon, M. E., Fridge III, E. M., Roush, G. B., Braley, D. M., and Savely, R. T., "The Development and Technology Transfer of Software Engineering Technology at Johnson Space Center", *Fifth Annual Workshop on Space Operations Automation and Robotics (SOAR ' 91)*, Houston, Texas, July 9-11,1991.

3. Fridge III, E., Braley, D., & Plumb, A., "Maintenance Strategies for Design Recovery and Reengineering," Vols. 1-4, NASA Johnson Space Center, Houston, TX, June, 1990.

4. Plumb, A., & George, V., "A Method for Conversion of FORTRAN Programs," Barrios Technology, Inc., Houston, TX, March, 1990.

A MEASUREMENT SYSTEM FOR LARGE, COMPLEX SOFTWARE PROGRAMS

S24-61
44834

Kyle Y. Rone, Kitty M. Olson, Nathan E. Davis
IBM Corporation
3700 Bay Area Blvd.
Houston, TX., 77058-1199

ABSTRACT

This paper describes measurement systems required to forecast, measure, and control activities for large, complex software development and support programs. Initial software cost and quality analysis provides the foundation for meaningful management decisions as a project evolves. In modeling the cost and quality of software systems, the relationship between the functionality, quality, cost, and schedule of the product must be considered. This explicit relationship is dictated by the criticality of the software being developed. This balance between cost and quality is a viable software engineering trade-off throughout the life cycle. Therefore, the ability to accurately estimate the cost and quality of software systems is essential to providing reliable software on time and within budget.

Software cost models relate the product error rate to the percent of the project labor that is required for independent verification and validation. The criticality of the software determines which cost model is used to estimate the labor required to develop the software. Software quality models yield an expected error discovery rate based on the software size, criticality, software development environment, and the level of competence of the project and developers with respect to the processes being employed.

A MEASUREMENT APPROACH

Thirty years of experience with programs for the National Aeronautics and Space Administration (NASA) has shown that the primary key to customer satisfaction is the capability to concurrently and consistently deliver compliant products, on time, within budget, and with an appropriate quality level. Figure 1 illustrates a measurement approach that leads to stabilization and control of the project and, ultimately, to improving the processes involved.

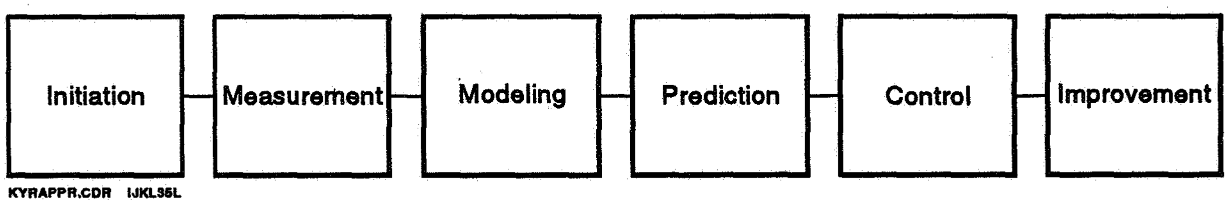


Figure 1. A measurement approach to stabilize and control a project.

Initiation consists of defining, tailoring, and stabilizing the required processes. Processes are institutionalized through standards, education, and procedures. Procedures based on process models which include precedence order and identified interim products permit visibility and control of the process.

Measurement includes defining a consistent set of measures which relate to the key goals of the project and the processes being utilized. It is important to measure with integrity. The purpose is not only to show that work is progressing, but to provide an understanding of how the processes are working. Measurements must be collected and stored in a manner to facilitate historical analysis.

Modeling involves presenting the data in a form that can be related to the process models and important, identifiable project parameters. These parameters include size, complexity, criticality, and process proficiency. The models must also be related to project schedules so that staffing needs can be analyzed.

Prediction begins by calibrating the new problem using historical data, tailoring the process to the new problem, and selecting an appropriate model. The model and the current product definition are used to forecast the key project parameters of size, complexity, and criticality. A plan is developed based on the process model and the predicted parameters.

Control points are established as key schedule milestones to permit assessments. After the previous steps are complete, the stage is set to activate the control loop: reaching a defined control point, collecting measurements, evaluating the data, acting on the data, and assuring that changes to the system are reflected in the plan.

Improvement can be initiated only after completion of these steps--initiation, measurement, modeling, prediction, and control. To improve the process, a change must be proposed, the impact on the key measurements predicted, and the first five steps repeated to re-establish and stabilize the process. Improvements can be assessed only when taken individually so that the impact of a change can be isolated. This approach applies to all the "keys to customer satisfaction" as illustrated in Figure 2. The consistent application of this approach is required to assure that all customer satisfiers are met.

Keys	Steps					
	Initiation	**Measurement**	**Modeling**	**Prediction**	**Control**	**Improvement**
Product	• Process • Interim product • Precedence order • Tailoring mechanism	• Size • Process proficiency	• Process models	• Process tailoring	• Control points	• Modify process • Modify ordering • Automation
Cost		• Function driven cost • Schedule driven cost • Complexity • Criticality	• Factor models • % Models • Phasing	• Calibrate to process • Function driven	• Cost management	• Modify cost models
Schedule		• Process elapsed time • Process order	• Schedule rules of thumb	• Phased cost and errors	• Schedule management	• Modify schedule rules of thumb
Quality		• Inspection errors • Process errors • Product errors • Total errors	• Life cycle errors • % Models • Phasing	• Calibrate to process • Function driven • Cost driven	• Quality management	• Modify quality models

Figure 2. Measurement approach applied to the keys to customer satisfaction.

COST AND QUALITY RELATIONSHIP

In developing software systems the cost and quality of a product can be traded against one another. By attempting to minimize development costs many projects simply defer error correction into the product time frame where the cost of error correction is more expensive. To prevent this from occurring a careful balance of product cost versus product quality must be established. The trade-off between cost and quality is dictated by the criticality of the function being developed. As criticality increases, it becomes imperative that the software be error free.

Reducing errors can be accomplished by thoroughly testing the software throughout the development life cycle. As shown in Figure 3 the cost models are imposed on a curve which relates the product error rate to the percent of the project labor that is required for independent verification and validation. Independent verification and validation involves monitoring the test strategy, plans, and procedures for a project and may also require an organization independent of the development organization to conduct system tests. The criticality level determines which cost model is used to estimate the labor months required to develop a software system. For example, a software component classified as low criticality will incur verification costs and indirect costs which are a relatively low percentage of the overall total development cost. In contrast, a software component which is classified as high criticality will incur verification costs and indirect costs which are a relatively high percentage of the overall total development cost.

Each cost model is associated with a specific product error rate. For example, the low criticality cost model is related to a product error rate of one error per one thousand source lines of code (KSLOC). In contrast, the high criticality cost model is related to a product error rate of one tenth (0.1) error per one KSLOC.

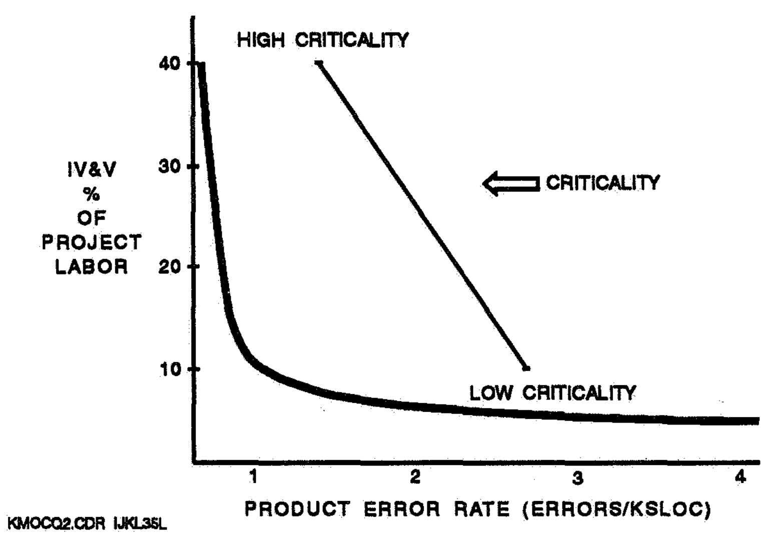


Figure 3. Product Error Rate versus Independent Verification & Validation Percentage (IV&V) of Project Labor

MODELING SOFTWARE COST

Modeling software cost is based on the technique of stepwise refinement. This technique involves decomposing the software system requirements into software functional components. These components are further decomposed into as many independent elements as possible. Decomposition terminates whenever a reused software element or a Commercial-Off-The-Shelf (COTS) software product is identified or whenever the component is decomposed to the lowest level. These software elements are sized and classified according to release, language, complexity, and criticality.

Software size, usually measured in source lines of code, is an important factor that ultimately affects the accuracy of the labor estimate. For example, as the size of the software system increases, a parallel increase in the interdependency among the various software components also occurs.

Release represents either an incremental product release, a release of the software development environment, or the learning curve associated with the software development process. Language is the programming language in which each software component will be implemented.

Complexity, the relative difficulty of developing each software component, is an important factor affecting development costs. Some types of software systems are more difficult to develop than others, e.g., developing an operating system versus developing utility software.

Criticality is the level of effect of a failure of a software component. Software for certain medical diagnostic or treatment systems and air traffic control systems must not fail or human lives could be lost. In contrast, an inventory control system should not fail, but the impact of the failure would not result in the loss of human life. As illustrated in Figure 4 these inputs--size, release, language, complexity, and criticality-- are used by the cost model to generate an estimate of the overall effort required to develop the software components and to determine how the effort will be distributed.

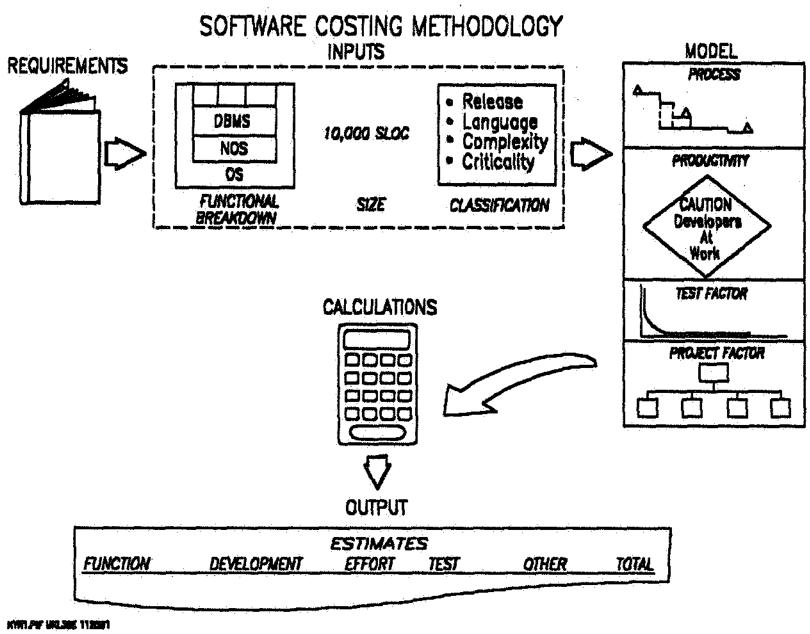


Figure 4. Software Cost Estimation Methodology

MODELING SOFTWARE QUALITY

The software development process is inherently complex; therefore, many opportunities exist to make errors. It is essential that an organization project an expected error distribution for each software development project. By plotting the actual error distribution against the expected error distribution curve, management can judge whether or not work is proceeding within expected bounds. Using this information, management can determine how well their error prediction and error prevention practices are working.

As in the cost estimation process, the requirements are first decomposed into functional components. These components are sized and classified according to release, criticality, project proficiency and development proficiency. Project proficiency represents the level of competence of a project with respect to its software engineering process. This is, how well a project as a whole is able to implement a software engineering process. Project proficiency determines how many errors will be inserted in the product per one thousand source lines of code. Development proficiency represents the level of competence of the developers with respect to their process. Development proficiency determines the total number of errors that will be discovered early in the development cycle. Development proficiency is dependent on situational factors such as experience level, availability of mentors, the ability to integrate new technology and especially on how well software product inspections are conducted. Criticality determines the project's product error rate, i.e., the number of product errors per one thousand source lines of code.

As illustrated in Figure 5, these inputs--size, release, criticality, project proficiency, and development proficiency--are used by the quality model to generate an expected error distribution pattern of early, process, and product errors.

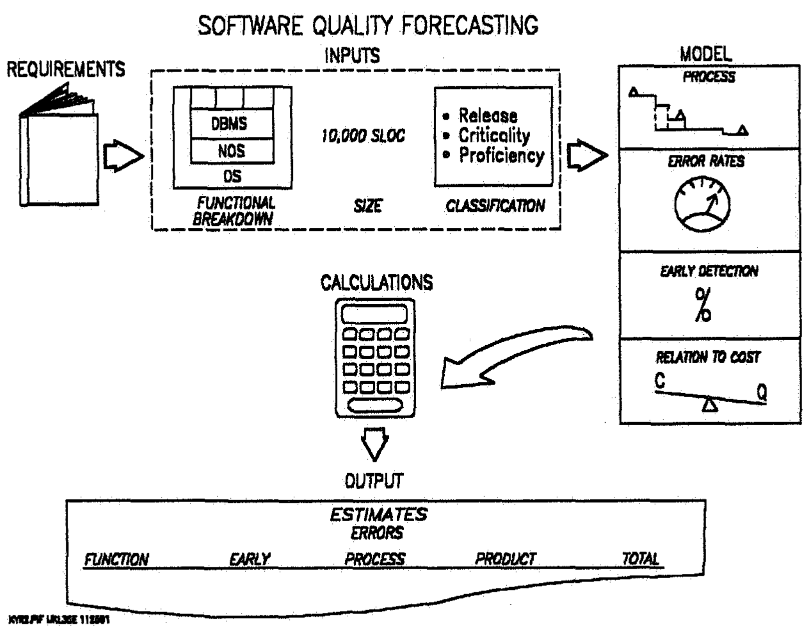


Figure 5. Software Quality Forecasting Methodology

MODELING A STAFFING PROFILE

Once the cost and quality estimates are determined a Rayleigh curve can be used to phase either estimate over time. A Rayleigh curve is a plot of a mathematical function which describes a life cycle phenomena. The Rayleigh curve illustrates whether the slope of a staffing curve is too steep or whether the error density is too great at certain points in the process.

Figure 6 illustrates the staffing profile for an ideal project. A minimum level of critical skills is required during the maintenance phase. This steady-state staffing level forms the support line. It includes critical skills for requirements, design, implementation, testing, and management. The support line is a function of system size and productivity as well as unique skill requirements specific to the software being maintained. In Figure 6 the area below the support line and above the maintenance tail of the Rayleigh curve represents the capability for new development work.

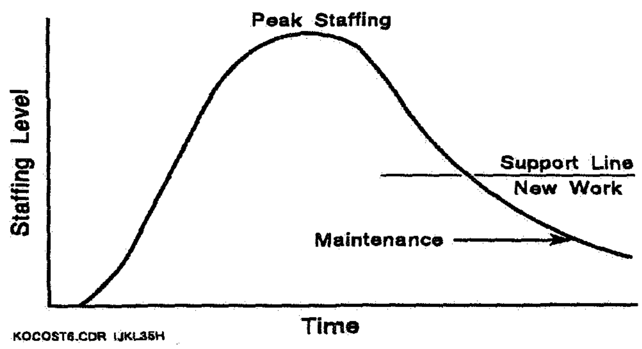


Figure 6. Staffing level is modeled using a Rayleigh curve.

As shown in Figure 7 the total maintenance effort can be modeled as the sum of a sequence of Rayleigh curves. The sizing and scheduling of new development activities should be planned to provide a stable level of effort as illustrated by the total development line. Software maintenance which handles Problem Reports can continue at a lower support level as illustrated by the total maintenance line. The total development line should not fall below the critical skills required by the project as determined by the initial staffing model.

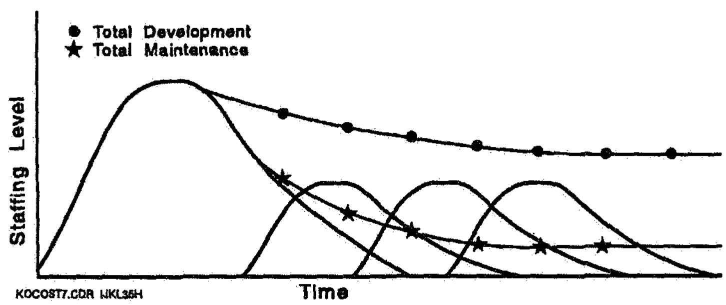


Figure 7. Maintenance effort modeled as a sequence of Rayleigh curves.

SUMMARY

Software now controls everything from the nation's telephone communication system to the world's financial systems. Delivering reliable software on time and within budget depends on an accurate measurement approach. An integrated measurement approach provides the information needed to effectively plan, manage, and control the software development process.

The cost and quality models described in this paper provide the capability to quickly generate estimates for diverse types of projects. Changes in assumptions such as size, complexity, or criticality are easily factored into the cost and quality estimates. The estimates can be phased across time to determine if the staffing profile or the error density is too steep at certain points in the process.

REFERENCES

Boehm, B. W., "Improving Software Productivity," *COMPUTER*, Vol. 20, No. 9, September, 1987, 43-57.

Kan, S. H., "Modeling and Software Development Quality," *IBM Systems Journal*, Vol. 30, No. 3, 1991, 351-362.

Kemerer, C. F., "An Empirical Validation of Software Cost Estimation Models," *COMMUNICATIONS of the ACM*, Vol. 30, No. 5, May, 1987, 416-429.

Putman, L. H. and W. Myers, *Measures For Excellence*, Yourdon Press, Englewood Cliffs, New Jersey, 1992.

Rone, K. Y., "Cost and Quality Planning for Large NASA Programs," *Proceedings of the Fifteenth Annual Software Engineering Workshop*, NASA Fifteenth Annual Software Engineering Workshop, Greenbelt, Maryland, Nov. 28-29, 1990.

omit

Session S4: SOFTWARE ENGINEERING II

Session Chair: Ernest Fridge

UNDERSTANDING SOFTWARE FAULTS AND THEIR ROLE IN SOFTWARE RELIABILITY MODELING

John C. Munson
Department of Computer Science
University of West Florida

This study is a direct result of an on-going project to model the reliability of a large real-time control avionics system. In previous modeling efforts with this system, hardware reliability models were applied in modeling the reliability behavior of this system. In an attempt to enhance the performance of the adapted reliability models, certain software attributes were introduced in these models to control for differences between programs and also sequential executions of the same program [8]. As the basic nature of the software attributes that affect software reliability become better understood in the modeling process, this information begins to have important implications on the software development process.

A significant problem arises when raw attribute measures are to be used in statistical models as predictors, for example, of measures of software quality. This is because many of the metrics are highly correlated [1, 4]. Consider the two attributes: lines of code, *LOC*, and number of program statements, *Stmts*. In this case, it is quite obvious that a program with a high value of *LOC* probably will also have a relatively high value of *Stmts*. In the case of low level languages, such as assembly language programs, there might be a one-to-one relationship between the statement count and the lines of code. When there is a complete absence of linear relationship among the metrics, they are said to be orthogonal or uncorrelated. Usually the lack of orthogonality is not serious enough to affect a statistical analysis. However, for the purposes of some statistical analysises such as multiple regression, the software metrics are so strongly interrelated that the regression results may be ambiguous and possibly even misleading. Typically, it is difficult to estimate the unique effects of individual software metrics in the regression equation. The estimated values of the coefficients are very sensitive to slight changes in the data and to the addition or deletion of variables in the regression equation.

Since most of the existing metrics have common elements and are linear combinations of these common elements, it seems reasonable to investigate the structure of the underlying common factors or components that make up the raw metrics. The technique we have chosen to use to explore this structure is a procedure called principal components analysis. Principal components analysis is a decomposition technique that may be used to detect and analyze collinearity in software metrics. When confronted with a large number of metrics measuring a single construct, it may be desirable to represent the set by some smaller number of variables that convey all, or most, of the information in the original set. Principal components are linear transformations of a set of random variables that summarize the information contained in the variables. The transformations are chosen so that the first component accounts for the maximal amount of variation of the measures of any possible linear transform; the second component accounts for the maximal amount of residual variation; and so on. The principal components are constructed so that they represent transformed scores on dimensions that are orthogonal [4].

Through the use of principal components analysis, it is possible to have a set of highly related software attributes mapped into a small number of uncorrelated attribute domains. This definitively solves the problem of multicollinearity in subsequent regression analysis [3]. There are many software metrics in the literature, but principal component analysis reveals that there are few distinct sources of variation, i.e. dimensions, in this set of metrics. It would appear perfectly reasonable to characterize the measurable attributes of a program with a simple function of a small number of orthogonal metrics each of which represents a distinct software attribute domain.

SOFTWARE ATTRIBUTES AND SOFTWARE FAULTS

There is now sufficient evidence to support the conclusion that there is a distinct relationship between software faults and measurable program attributes and that this information will yield specific guidelines for the design of reliable software. In particular, software complexity measures are distinct program attributes that have this property [3]. If a program module is measured and found to be complex, then it will have a large number of faults. These faults may be detected by analytical methods, e.g., program inspections. The faults may also be identified based on the *failures* that they induce when the program is executing. A program may preserve a number of latent faults over its lifetime in that the particular manner that it is used may never cause the complex code sequences to execute and thus never expose the faults. Alternatively, a program may be forced to execute its complex code segments early in its life cycle and thus fail frequently early on followed by reliable service.

Code faults are not inserted by some random process analogous to the assignment of chocolate chips to cookies in the cookie manufacturing business. Faults occur in direct relationship to the complexity of the programming task. It is said about one of our recent presidents that he could not walk and chew gum at the same time. A programmer faced with the task of converting a complex requirement into a complex algorithm in a rich programming language has a much more difficult task than just walking and chewing. It is quite reasonable to expect that the programmer will make errors. These errors will express themselves as faults in the program. From a maintenance perspective, it will be very expensive and time consuming to find and to fix these faults. The real problem is to identify design rules that will restrict code faults from being introduced in the first place.

The concept of module granularity is relevant in terms of our description of a software system. At the lowest level of granularity, the micro level, we might define a module to be the smallest compilable unit of code. The next level of module granularity might be called the top level module granularity. In this case, all `include` compiler directives have been preprocessed by the compiler system. Typically, these top level modules will be the smallest units that may be measured by an execution profiler, a concept that will be necessary later in this discussion. Yet another level of granularity would be the functional subsystem level, i.e. the collection of modules grouped together in such a subsystem that would work in concert to solve a particular problem.

Whole classes of faults can be characterized as within-module faults. An example of this might be the case where a local identifier within a module is inadvertently assigned the wrong value by a programmer. One stage of fault analysis will examine the relationship of micro complexity measures such as the Halstead software science measures to faults occurring within module boundaries.

At the macro complexity level, we are interested in software measures that describe the interaction of program modules. This would include the macro measures of software coupling, call graph features, or more global software system metrics such as the aggregate complexity metric [2]. At this level we are interested in the relationship between macro complexity measures such a coupling complexity metrics and faults that are found on the inter-module interface level.

Yet another class of software faults do not relate to physical problems in the code but to temporal sequencing anomalies in a real-time control environment. This fault class will represent the highest level granularity. This is the operating system task management level. In this case the operating system or operating environment is responsible for the sequencing of the operation of program modules and the successful rendezvous of these modules.

A DOMAIN MODEL OF SOFTWARE ATTRIBUTES

In most linear modeling applications with software metrics, such as regression analysis and discriminant analysis, the independent variables, or metrics, are assumed to represent some distinct aspect of variability not clearly present in other measures. In software development applications, the independent variables, in this case, the complexity metrics, are strongly interrelated or demonstrate a high degree of multicollinearity. In cases like this, it will be almost impossible to establish the unique contribution of each metric to the model. One distinct result of multicollinearity in the independent measures is that the regression models developed using independent variables with a high degree of multicollinearity have highly unstable regression coefficients. Such models may be subject to dramatic changes due to additions or deletions of variables or even discrete changes in metric values. To circumvent this problem, principal components analysis has been used, quite successfully, to map the metrics into orthogonal attribute domains [4]. Each principal component extracted by this procedure may be seen to represent an underlying common attribute domain.

To aid in the visualization of this concept, the principal components analysis of a large avionics software system will be examined. In this specific example, a set of 13 metrics were identified. These metrics were chosen because they were thought to represent a number of distinct software complexity attributes. Each of the 13 measurements were obtained from a set of 1300 program modules that comprise the avionics system.

Many of the 13 metrics are known to be highly correlated. The first step in the investigative process is to determine how many distinct orthogonal attributes are really being measured by these 13 raw complexity metrics. Principal components analysis will reveal this orthogonal attribute structure. From Table 1, which contains the results of the principal components analysis of the 1300 software modules, we can see that most of the information represented by the original 13 metrics can be reduced to three orthogonal attribute domains. This table contains the values that assess the strength of the relationship between each raw metric and the three resulting orthogonal domains. There are a group of six metrics, for example, that are closely related to the first domain in the new orthogonal attribute domain structure. These are highlighted with boldface type. Similarly there are four raw metrics associated with the second attribute and three with the third. In some cases it is useful to associate names with the domains. The metrics most closely associated with **Domain1** in Table 2 are those whose table values are the largest in column one of the table. For clarity, they have been typed in bold. All of these metrics seem to share a common property of **Size** measurement. Similarly, the metrics associated with **Domain2** seem to share a **Structure** measurement and those in **Domain3** a **Control** measure.

If we carefully examine the set of 13 measures used above we can observe that there are some conceptual areas of software attributes that are not represented in this set of metrics. For example, there is the matter of data structure complexity. Clearly, there are no measures present in the set for this software attribute. The problem is that we do not know just how many distinct, measurable attributes a software system might have. But we do know that these 13 metrics are only measuring three distinct, uncorrelated attributes.

Attribute Domain Model

The objective now is to begin to build and extend a model for software attributes. This model will contain a set of orthogonal attribute domains. Once we have such a model in place we would then like to identify and select from the attribute domain model those attributes that are correlated with a software quality measure, such as number of faults. Each of the orthogonal attributes will have an associated metric value that is uncorrelated with any other attribute metrics. Each of these attributes may potentially serve to describe some aspect of variability in the behavior of the software faults in a program module.

One the primary purposes in the use of principal components analysis is to reduce the dimensionality of the attribute problem. In the specific case of the avionics system under current scrutiny, there were initially thirteen apparent software attributes that were being measured. Through the use of principal components analysis, we find that there are three distinct attribute domains that will serve to explain most of the variation observed in the original set of thirteen metrics. Further, we can transform the thirteen raw attribute measures for each of the 1300 program modules into three new uncorrelated metrics, one for each of the new orthogonal attribute domains. While this reduction in the number of metrics has simplified the problem somewhat, we really would like to represent each program with a single metric that would allow us to compare (order) each of the program modules in comparison to the rest of the program modules.

There have been some tragically ill-considered attempts to design software systems reflecting the complexity of the object being designed. The most notable of these attempts relates to the use of McCabe's measure of cyclomatic complexity V(g). For some unknown reason, magic values of cyclomatic complexity are now being incorporated into the requirements specifications of some software systems. For example, we might choose to specify that no program module in the software system should have a cyclomatic complexity greater than an arbitrary value of, say 15. This is a very good example of how software measures might well be used in the design process.

There are potentially catastrophic consequences associated with this univariate design criterion. First, there is little or no empirical evidence to suggest that a module whose cyclomatic complexity is greater than 15 is materially worse than one whose cyclomatic complexity is 14. Secondly, and most important is the fact that if, in the process of designing a software module, we find that the module has a cyclomatic complexity greater than 15, the most obvious and common solution to the problem is to divide the software module into two distinct modules. Now we will certainly have two modules whose cyclomatic complexity is less than 15. The difficulty here, is that instead of one program module, we have created two, or possibly three, in its place. This will increase the macro complexity of measures related to complexity. In other words, we have decreased cyclomatic complexity, but we have increased coupling complexity. The result of the ignorant decision may well be that the total *system* complexity will increase. This, in turn will lead to a concomitant increase in total faults.

From the attribute domain model, we can see that there are many distinct software complexity domains. If we make design decisions on the basis of incomplete measurements, we run the risk of doing ourselves real damage. As above, we may make a design decision to reduce a measurement in one domain but this may in fact cause a concomitant increase in measures in other complexity domains. The net result of this univariate decision is that the net complexity of a system may rise. While it is true that we may have fewer faults associated with aspects of control complexity, we run the risk of increasing the count of faults associated with coupling considerations.

In order to simplify the structure of software complexity even further than the orthogonal domains produced by the principal components analysis it would be useful if each of the program modules in a software system could be characterized by a single value representing some cumulative measure of complexity. The objective in the selection of such a function, g, is that it be related in some manner to software faults either directly or inversely such that $g(x) = ax + b$ where x is some unitary measure of program complexity. The more closely related x is to software faults, the more valuable the function, g, will be in the anticipation of software faults. Previous research has established that the relative complexity metric, ρ, has properties that might be useful in this regard. The relative complexity metric, ρ, is a weighted sum of a set of uncorrelated attribute domain metrics [6, 7]. This relative complexity metric represents each raw metric in proportion to the amount of unique variation contributed by that metric.

A sample of the relative complexity values for the avionics software system is shown in Table 3, together with the domain metrics. This table contains three compartments. The first compartment shows the program modules of the least complexity; the second compartment shows those of average relative complexity; and the third compartment contains the complexity values from the most complex modules (by relative complexity). The last column in the table lists the total faults found in each of the associated modules. We can see an obvious relationship between the relative complexity of a program module and the number of faults found in it. The bivariate correlation between these two variables is approximately 0.80 for all of the data from the 1300 program modules.

Maintaining a Software Measurement Data Base

From the standpoint of design methodology, we would like to be able to use the measurements from past software development efforts for the purposes of developing a preliminary assessment of current or active software design project. In other words, we would like to have the ability to use a past system to serve as a baseline for a current development project. In essence the objective is to use an existing data base to baseline subsequent measures from a software system currently in development. We might choose, for example, to take the first build of a real-time control software system that was developed in the past and use this for a real-time control software system currently being developed. In this sense, all subsequent software measures on new systems will be transformed relative to the baseline system.

The ability to use information from past development projects in current design work is most important. This is due to the fact that many of the software quality and reliability attributes of a system can only be measured after the system has been in service for some time. If a software system under current development is directly comparable to one that had demonstrated quality and/or reliability problems in the past, there is evidence to suggest that the design had better be modified, and soon.

The attribute measures presented so far are static measures of the program. They measure such features of the program as its size and the complexity of its control structure. If the functionality of a program was extremely restricted, these static measure might well be sufficient to describe the program entirely. Most modern software systems, however, have a broad range of functionality. Consider, for example, the software system for a typical spreadsheet program. The number of distinct functions in such a system and the number of ways that these function might be exercised are both very large numbers. In addition to static measures of program attributes, we must also be concerned with dynamic measures of programs as well.

Execution Profile

A software system is written to fulfill a set of functional requirements. It is designed in such a manner that each of these functionalities is expressed in some code component. In some cases there is a direct correspondence between a particular program module and a particular functionality. That is, if the program is expressing that functionality, it will execute exclusively in the module in question. In most cases, however, there will not be this distinct traceablility of functionality to modules. The functionality will be expressed in many different code modules. In addition to the user profile derived from the customer profile, one can derive a functional profile for each system mode of operation. Of course, each mode of operation has its own occurrence probability. For a given system mode, it is possible to refine that mode into the functions it requires and each function's occurrence probability providing a quantitative distribution of the relative use of various functions.

As a program is exercising any one of its many functionalities, it will apportion its time across one to many program modules. This temporal distribution of processing time is represented by the concept of the execution profile. In other words, if we have a program structured into n distinct modules, the execution profile for a given functionality will be the proportion of time spent in each program module during the time that the function was being expressed.

Another way to look at the execution profile is that it represents the probability, p_i, of execution occurring in module m_i at any point in time. When a software system is running a fixed function there is an execution profile for the system represented by the probabilities, $p_1, p_2, p_3, \cdots, p_n$. For our purposes, p_i represents the probability that the i^{th} module, in a set of n modules, is in execution at any arbitrary time.

Each functionality will have is own, possibly unique, execution profile. The specific implications for software design in the concept of the execution profile may be drawn from information theory. In this case, each scenario under which the program may execute has its own set of execution profiles. Each of the execution profiles may be assessed in terms of its entropy defined as follows:

$$h = -\sum_{i=1}^{n} p_i \log_2 p_i.$$

The point of maximum entropy will occur when all modules will execute with an equal probability. In which case,

$$h_{\max} = \log_2 n - 1.$$

This point of maximum entropy would occur in a circumstance where all modules of a program were executed for precisely the same amount of time under a given input scenario. It is a point of maximum surprise. There is no information as to where the program is likely to be executing at any point in time. At the other extreme, the point of minimum entropy, 0, is the point at which the program will execute in only one of its modules. The probability of executing occurring in this module is 1.0. The probability of it executing any other module is 0.

The concept of entropy may also be used to evaluate a design in terms of its modularity. We can see that the entropy for a set of n program modules is bounded above by $\log_2 n - 1$. From the standpoint of entropy, the maximum effect of a decision to increase the number of modules in a design from n to $n+1$ is basically the difference between $\log_2 n$ and $\log_2 (n+1)$. For small n, this difference might be substantial. For large n, the incremental effect of the additional module is not that great.

The entropy of a design is not a sufficient condition in its own right for the evaluation of a design. There is clearly an interaction between entropy and complexity. A design of low entropy implies that a great deal of time is spent in relatively few modules. If the associated modules have small values of relative complexity then its reliability should be high. If, on the other hand, the most frequently executed modules are also very complex, we would anticipate a low reliability for the system.

Functional Complexity

As per the earlier discussion of relative complexity, it can be seen that it is possible to characterize a program module by a single value such as relative complexity. This relative complexity measure, ρ, of a program is a measure of the program at rest. When a program is executing, not every one of its modules will receive the same level of exposure. This is evident in the concept of execution profile. Some executions might result in very complex program modules being executed. Other program input stimuli may cause the program to execute only its least complex modules.

There is, then, a dynamic aspect to program complexity that is related to its entropy under a particular test scenario. The functional complexity, ϕ, of the system running an application with an execution profile defined above is then:

$$\phi = \sum_{j=1}^{n} p_j \rho_j$$

where ρ_i is the relative complexity of the j^{th} program module and p_j is the execution probability of this module. This is simply the expected value of relative complexity under a particular execution profile. The execution profile for a program can be expected to change across the set of program functionalities. In other words, for each functionality, f_i, there is an execution profile represented by the probabilities $p_1^i, p_2^i, \ldots, p_n^i$. As a consequence, there will be a functional complexity ϕ_i for each function, f_i execution, where

$$\phi_i = \sum_{j=1}^{n} p_j^i \rho_j$$

This is distinctly the case during the test phase when the program is subjected to numerous test suites to exercise differing aspects of its functionality. The functional complexity of a system will vary greatly as a result of the execution of these different test suites.

Given the relationship between complexity and embedded faults, we would expect the failure intensity to rise as the functional complexity increases. If an application is chosen in such a manner that there are high execution probabilities for the complex modules, then the functional complexity will be large and the likelihood of a failure event during this interval would be relatively high. Given the relationship between software complexity and software faults, some applications will lead to consistent failures while others will not. The most important message, here, is that a software system will fail in direct relationship its functionality. This is a very predictable and understandable relationship.

SOFTWARE RELIABILITY MODELING

Almost all of the current software reliability modeling approaches are simply derivatives of hardware reliability models. Our view of complex software systems is colored by our experience with complex mechanical or electronic systems. In the case of software systems it is evident that these systems are clearly a different animal. Consider the case of a jet engine as a mechanical system. When this system is operating, essentially all of its components are operating together. An operational test of this system may be achieved by simply breathing life into it. This is not the case with software systems. These systems contain many independent components, or modules. Taken in their totality there are potentially a vast number of distinct paths that may be taken through these systems as a function of the different inputs to the system. Hence, the dynamic complexity of the software system will depend on the inputs to the system. The net effect of differing inputs to the system is that the operational or functional complexity of the system will change in response to the varying inputs. Given the association between module complexity and errors, it follows that as applications change over time intervals, so too, will the likelihood of faults change with respect to time.

If we closely examine the attempts to fit simple Non-homogeous Poisson Process models or simple two parameter Weibull type models to empirical data, a consistent pattern emerges. There are systematic departures in the observed values from the predicted values. These departures represent significant trends in the residuals. There is something systematically happening in the execution of the software. This clearly does not represent noise in the system. It is the thesis of this paper that the departure of the residuals from the random pattern that we would expect due to measurement errors and other sources of noise is directly attributable to the functional complexity of the software in execution.

In order to demonstrate the integration of functional complexity into the modeling process, the Weibull distribution will be used. This is basically a three parameter distribution whose cumulative distribution for a random variable, x, is given as follows:

$$F(x;\alpha,\beta,\gamma) = 1 - e^{\left[\frac{x-\gamma}{\alpha-\gamma}\right]^{\beta}}$$

where $\beta > 0$, $\alpha > 0$ and $\gamma \geq 0$. The location parameter, γ, will represent the minimum life of the system. The parameter, β, is the shape parameter or Weibull slope. The parameter, α, is the scale parameter.

In the particular case where we wish to represent the cumulative failure probability of a system with a minimum life of zero, the γ term vanishes yielding a two parameter Weibull distribution,

$$F(x;\alpha,\beta) = 1 - e^{\left[\frac{x}{\alpha}\right]^{\beta}} .$$

Let us now surmise that the scale of the distribution represents the functional complexity of the software varying by functionality over time. The former parameter, α, will now be derived from empirical data. This form of the Weibull distribution is a one parameter distribution in β. The bottom line of this new modeling process is the observation that the probability of the failure of a software system is not just strictly a function of time. It also appears to be a function of the actual code that is executing at any point in time. This new process may be summarized quite succinctly as follows: Tell me what the software will be executing and I will tell you what its reliability will be for that application.

SUMMARY

It is increasingly evident that the reliability of a software system is largely determined during program design. One distinct aspect of the software design process that lends itself to measurement is the decomposition of the functionality of a system into modules and the subsequent interaction of the modules under a given set of inputs to the software. The actual distribution of execution time to each software module for a given input scenario is recorded in the execution profile for the system. For a given input scenario to a program, the execution profile is a function of how the design of the software has assigned functionality to program modules.

The reliability of a software system may be characterized in terms of the individual software modules that make up the system. From the standpoint of the logical description of the system, these functional components of the larger system are, in fact, operating in series. If any one of these components fails, the entire system will fail. The likelihood that a component will fail is directly related to the complexity of that module. If it is very complex the fault probability is also large. The system will be as reliable as its weakest component.

It is now quite clear that the architecture of a program will be a large determinant of its reliability. Some activities that a program will be asked to perform are quite simple. They will be easily understood by designers and programmers alike. The resulting code will not likely contain faults. If, on the other hand, the specified functionality of a program is very complex and, as a result ambiguous, then there is a good likelihood that this functionality will be quite fragile due to the faults introduced through the very human processes of its creation. A more realistic approach to software reliability modeling will reflect the software reliability in terms of the functionality of the software.

REFERENCES

[1] V. Cote et al., "Software Metrics: An Overview of Recent Results," *Journal of Systems and Software*, **8**, 1988, pp. 121-131.

[2] T. M. Khoshgoftaar and J. C. Munson, "A Measure of Software System Complexity and Its Relationship to Faults," *Proceedings of the 1992 International Simulation Technology Conference* Houston TX November, 1992.

[3] T. M. Khoshgoftaar and J. C. Munson, "Predicting Software Development Errors Using Complexity Metrics," *IEEE Journal on Selected Areas in Communications* Vol. 8, No. 2, February 1990, pp. 253-261.

[4] J. C. Munson and T. M. Khoshgoftaar, "The Dimensionality of Program Complexity," *Proceedings of the 11th Annual International Conference on Software Engineering*, Pittsburgh, May 1989, pp. 245-253.

[5] J. C. Munson and T. M. Khoshgoftaar, "Regression Modeling of Software Quality: An Empirical Investigation," *Journal of Information and Software Technology*, Vol. 32 No. 2 March 1990, pp. 105-114.
[6] J. C. Munson and T. M. Khoshgoftaar, "Applications of a Relative Complexity Metric for Software Project Management," *Journal of Systems and Software*, Vol. 12, No. 3, 1990, pp. 283-291.
[7] J. C. Munson and T. M. Khoshgoftaar, "The Relative Software Complexity Metric: A Validation Study," *Proceedings of the Software Engineering 1990 Conference*, London, 1990, pp. 89-102.
[8] J. C. Munson and T. M. Khoshgoftaar, "The Use of Software Complexity Metrics in Software Reliability Modeling," *Proceedings of the IEEE International Symposium on Software Reliability Engineering*, May, 1991, pp. 2-11.

	Domain1 Size	**Domain2 Structure**	**Domain3 Control**
N_2	**0.964**	0.106	0.096
LOC	**0.950**	0.119	0.166
N_1	**0.946**	0.117	0.064
η_2	**0.921**	0.029	0.057
Stmts	**0.740**	0.225	0.472
Comm	**0.662**	0.056	0.395
Max–Path	0.109	**0.914**	0.303
Path	0.115	**0.909**	0.296
Paths	0.128	**0.853**	0.189
Cycles	0.045	**0.761**	0.238
Edges	0.170	0.376	**0.866**
Nodes	0.169	0.384	**0.864**
η_1	0.222	0.341	**0.647**
Eigenvalues	4.715	3.470	2.615

Table 1. Complexity Domain Structure for Avionics Software

Module	Domain1	Domain2	Domain3	ρ	DR Count
1	-0.783	-0.016	0.369	43.36	0
2	-0.782	-0.017	0.368	43.36	0
3	-0.782	-0.017	0.368	43.36	0
4	-0.778	-0.021	0.361	43.37	0
5	-0.777	-0.022	0.359	43.37	0
6	-0.772	-0.027	0.349	43.39	0
7	-0.769	-0.031	0.343	43.40	0
8	-0.763	-0.001	0.314	43.53	0
9	0.069	-0.119	-0.183	49.8	0
10	-0.230	0.456	0.049	49.89	0
11	0.220	-0.078	-1.329	49.92	2
12	0.486	-0.888	-0.526	49.98	9
13	-0.258	0.436	0.412	50.03	1
14	-0.092	0.173	0.125	50.05	0
15	0.259	-0.054	-1.568	50.05	0
16	0.230	0.027	-1.566	50.13	1
17	3.161	3.277	2.557	95.44	9
18	7.575	-5.399	1.667	97.8	25
19	3.752	3.198	1.314	98.80	4
20	3.452	4.460	3.065	103.6	6
21	4.828	2.456	0.269	104.02	4
22	5.989	3.087	6.093	124.72	10
23	7.956	4.452	-4.040	134.83	3
24	8.242	5.139	-0.865	144.42	15

Table 2. Sample Program Modules with Domain Metrics, Relative Complexity and Faults

EXPERIENCES IN IMPROVING THE STATE OF THE PRACTICE IN VERIFICATION AND VALIDATION OF KNOWLEDGE-BASED SYSTEMS

S26-61
414842

Scott W. French
IBM FSC
Rockville, MD

Chris Culbert
NASA JSC
Software Technology Branch
Houston, TX

David Hamilton
IBM FSC
Houston, TX

ABSTRACT

Knowledge-based systems (KBSs) are in general use in a wide variety of domains, both commercial and government. As reliance on these types of systems grows, the need to assess their quality and validity reaches critical importance. As with any software, the reliability of a KBS can be directly attributed to the application of disciplined programming and testing practices throughout the development life-cycle. However, there are some essential differences between conventional software and KBSs, both in construction and use. The identification of these differences affect the verification and validation (V&V) process and the development of techniques to handle them. The recognition of these differences is the basis of considerable on-going research in this field. For the past three years IBM (Federal Systems Company - Houston) and the Software Technology Branch (STB) of NASA/Johnson Space Center have been working to improve the "state of the practice" in V&V of Knowledge-based systems. This work was motivated by the need to maintain NASA's ability to produce high quality software while taking advantage of new KBS technology. To date, the primary accomplishment has been the development and teaching of a four-day workshop on KBS V&V. With the hope of improving the impact of these workshops, we also worked directly with NASA KBS projects to employ concepts taught in the workshop. This paper describes two projects that were part of this effort. In addition to describing each project, this paper describes problems encountered and solutions proposed in each case, with particular emphasis on implications for transferring KBS V&V technology beyond the NASA domain.

BACKGROUND

Before this project began, KBS V&V was the subject of several ongoing research projects and articles in popular trade journals. Much of this discussion and research was based on several conjectures about differences between KBS V&V and V&V of conventional software. These conjectures were based on theoretical arguments and limited personal experiences. However plausible these conjectures were, no systematic effort had been made to determine the extent to which they impacted industry's ability to deploy reliable KBSs. That is, there was no evidence that the state of the practice in KBS V&V needed any improvement. After all, many successful KBS systems had been developed and, presumably, verified and validated in some fashion.

To understand the state of the practice in KBS V&V, we performed an extensive survey of several KBS V&V projects within NASA, other government agencies, and commercial companies ([5]). This survey validated the conjectures to some extent. It showed that deployed KBSs were generally less accurate or reliable than users and developers expected. We could not determine whether this was because the KBSs were of low quality or because expectations were unrealistically high. In either case, it pointed to a need to better define and meet accuracy and reliability requirements. More importantly, the survey indicated that the development of KBSs differed significantly from generally advocated practices. For example, less than half of the projects had any form of documented requirements. There were also some indications that KBS V&V was of some difficulty. For example, over sixty percent of the projects indicated that test coverage determination was a particular problem, but through follow-up interviews we learned that many were unfamiliar with existing test coverage techniques. The interviews also showed that many developers were domain experts with little programming experience; this may account for the unfamiliarity with traditional V&V approaches.

Results from the survey were instrumental in providing direction for both long-term and near-term work in the V&V of KBS. Though the survey appeared to justify the need to research new methods for KBS V&V, it also pointed out the need to inform KBS developers about V&V methods that already existed. The near-term direction we chose was to train KBS developers both in known KBS V&V techniques and in conventional V&V techniques (many of which had been shown to be useful in KBS V&V, despite the differences between KBS and conventional software). We sought to educate KBS developers in a way that both convinced them of the importance of V&V and gave them some confidence through hands-on experience with techniques so they could effectively use them. This appeared to be the most immediate way to make a significant impact in the state of the practice in KBS V&V.

We developed a four day workshop, teaching the underlying theory supporting V&V (i.e., why V&V is important), a wide-range of testing techniques, and a set of guidelines for applying these techniques. The material was based on a broad survey of V&V methods and was reviewed by several leading KBS V&V researchers. This material included:*

- a review of basic V&V concepts
- an explanation of the key differences between KBSs and conventional software
- a summary of over fifty V&V techniques
- examples, worksheets and guidelines for the techniques that were considered most useful
- an extensive set of references, cross referenced to each technique and concept

The workshop was taught several times to many NASA KBS developers. Although the results of the workshop have been very favorable (see [8]), the responsibility for applying the material taught lay entirely in the hands of the students. We contacted a small sampling of students several weeks after each course in an attempt to find out how well they were able to apply the material. By far, the most frequent answer was that they had not yet had an opportunity to apply them. (Perhaps the reason that they were able to attend the class was that they were "in between" projects.) To improve the impact of the workshop, we looked for (and found) ongoing KBS projects within NASA that would be willing to apply concepts taught in the workshop. The remainder of this paper describes this work.

PROJECT DESCRIPTIONS

This section describes two projects within NASA that we worked with to develop a KBS V&V approach. Each project fit within different development organizations within mission operations (Space Shuttle and Space Station Freedom). The first group, called users, was composed primarily of flight controllers who develop applications to automate and assist flight control activities within the mission control center. The second group, facility developers, developed the mission control complex itself. This development included both development of the key parts of the mission control center and incorporation of users group applications as part of the control center baseline. We worked with one project from each of these groups. We worked with a users group that developed a monitoring application called the Bus Loss Smart System or BLSS. We also worked briefly with facility developers for the space station mission

---

* See [6] and [7] for a discussion of the workshop contents.

control complex to develop criteria for assessing model-based user applications for inclusion into the control center baseline. This section gives insight into these projects by describing their environment, procedures and problems.

OVERVIEW OF THE USERS GROUP

In preparing to work with these groups, we taught a condensed (one day) version of the workshop to both flight controllers and application developers. There was a two-fold objective in teaching this workshop: (1) understand the kinds of problems they are working on and (2) teach them techniques that address those problems. Most of the problems they faced related directly to how these user applications are built. By this we mean that the basic development practices that support V&V were not practiced. We found that in most cases, the flight controller is the expert, developer and user. Inspections are viewed as being too expensive (both in dollars and time) and are, therefore, not done. Requirements were considered more of an enemy than a friend. For this reason, they rarely document the requirements that did exist. Their systems are viewed as prototypes, not as flight certified mission control applications. Becoming certified means that the application is added to the baselined set of mission control center applications. Few user applications had become certified. Testing emphasizes the functionality and user-interface aspects of their systems and not other important kinds of correctness such as safety and resource consumption.

Based on this insight into their development environment, several techniques were presented to the group. Most of these focused on helping them specify what the system should do and how it should do it. The following list of techniques was presented:

- Inspections ([18] and [2])
- Cause-Effect Graphing ([18], [19], [20] and [21])
- State Diagrams ([23])
- Decision Tables ([17])
- Assertion Analysis ([10])
- Object-oriented Analysis ([23], [11] and [26])
- Connectivity Graphs ([12] and [22])
- Petri Nets ([22], [15] and [1])
- Minimum Competency ([24] and [25])
- Pre/Post Conditions ([4], [3], [14], [9] and [13])

Bus Loss Smart System (BLSS)

BLSS is a flight control application designed to monitor electrical equipment anomalies, loss or malfunction within key electrical systems onboard the orbiter. Since it was a prototype it only acted as an assistant to the flight controller in analyzing telemetry data sent from the orbiter to the ground.

Like most of the other flight control applications it was developed using G2. Schematics of the electrical systems were created using G2 graphics capabilities. When anomalies were discovered in the electrical system, the flight controller was notified by BLSS via messages and highlighted items on the schematic. The flight controller then interacted with BLSS by indicating whether the anomaly should be ignored or further analysis was needed. BLSS then performed some deeper investigation into the anomaly.

Two primary methods were being used for testing BLSS. Both were system or "black-box" methods. With the first method, the flight controller supplied the programmer with simulation "scripts" (very much like operational scenarios). A simulation was then run based on this script to see that required outputs (as stated on the script) were generated. These simulations used actual telemetry data as supplied by the mission control complex.

The second method was also a simulation, but not a simulation that uses telemetry data from the mission control complex. Instead, special rules were inserted into the knowledge-base that caused certain events to happen at specific times while running in G2 simulation mode. A series of ten or so of these special cases had been developed to test the system. If the system passed all of these special cases, then testing was considered to be done.

FACILITY DEVELOPERS GROUP OVERVIEW

The purpose of the "models assessment" effort was to capitalize on existing Space Station Freedom (SSF) advanced automation projects. In these advanced automation projects, prototype systems were built in order to prove or demonstrate the ability to automate SSF operations using advanced technology. These prototype systems were not intended to be used operationally (i.e., the were not to be used directly by an SSF flight controller during actual flight operations). However, rather than building operational tools by completely re-implementing these systems, it was hoped that the prototypes could be turned into operational tools through additional development and/or additional V&V.

Models Assessment

The models were evaluated according to their usefulness and correctness. The usefulness of a prototype was judged by how well it met the needs of its target flight controllers. This involved more than just the functionality of the prototype. Issues such as usability were also considered. Judging the correctness of a prototype depended on its current level of correctness and the additional effort required to make the prototype sufficiently correct. Factors that impacted the assessment of correctness for a prototype were their lack of good requirements, their need to be stand-alone applications (i.e., the failure of one application should not affect another), their required role and function (e.g., advisor fault detection, diagnosis, etc.) and the role of their experts (users/experts may or may not be the developer).

APPROACH

Both projects had been studied in sufficient detail to define a V&V approach. In this section we describe our approach for each of these projects and the specific activities implementing that approach.

Bus Loss Smart System.

The most urgent need for the BLSS seemed to be to develop a good set of requirements that supported testing. The requirements that did exist lacked sufficient detail (i.e., they were very ambiguous) to support testing and maintenance. They also failed to address other important aspects of requirements such as safety, resource consumption, user profiles, etc.. Fortunately, most of the information needed for their requirements did exist. Our approach was to collect these requirements into a complete document based on DOD Std 2167A that would support testing. Our objective was to demonstrate the value of following standards and teach them how to write good requirements.

To complement the DOD 2167A format we provided the flight control group with a requirements handbook that describes the format of the document, the characteristics of good requirements, a step-by-step requirements definition method and a verification method for requirements. The approach we advocated was to define the overall goal of the system, the high-level tasks the system must perform (separated into competency and service tasks as discussed in [24] and [25]), user profiles, operational scenarios, and a state model for each task (see [4]). The tasks were then integrated through the definition of pre/post conditions and task invariants. Another urgent need for the BLSS was to have a good design specification that supported verification. The BLSS developers began defining this specification using an outline based on the DOD Std 2167A. We helped them incorporate a data dictionary based on the state models described in both the requirements and the design along with pre/post conditions for each procedure in the implementation. We had also planned to help them use inspections as a way to increase the quality of these specifications.

The last area where we are helped with BLSS was during user acceptance testing. This was different from the certification testing we described previously. This is primarily a "black-box" test activity performed by BLSS users to convince themselves that the system works. We had convinced them to use a statistical testing approach based on their simulation scripts. Simulation scripts were to be created that include at least one failure for each bus being monitored. The tester would keep track of the number of BLSS errors (based on severity - failing to identify a bus failure would be the most severe error) versus how long BLSS is in operation. Using these statistics we would apply a reliability model to quantify the quality of BLSS, in order to "certify" it.

Results

The BLSS development project had not yet been completed at the end of our consulting engagement. A draft requirements document and a draft design document had been developed but formal testing had not yet begun. Though they expressed great interest in the V&V approach that we had defined for them, it does not appear that they have continued following it as well as we had hoped.

Models Assessment

The general V&V approach defined for the Models Assessment was as follows. The first step was to develop requirements for each prototype. The requirements format developed for the BLSS project was to used as a base for a models assessment requirements format. Requirements were to be divided into requirements for evaluating prototype usefulness and requirements for evaluating prototype correctness. Initially, only the requirements supporting usefulness evaluation needed to be written. Then, if they were deemed useful, additional requirements would need to be documented to support evaluation of the correctness of the prototype. The initial requirements should include a description of the current operation of the system with emphasis on the problem(s) that the prototype was intended to address, the goals of the prototype (e.g., rapid diagnosis of faults or comprehensive identification of every possible failure condition) and a high-level description of the user interface to the system. This needed only be high level at this point, since the user would have the opportunity to interact with the tool and judge, firsthand, the usefulness of the interface.

Once the prototype had been deemed useful, the more difficult task of assessing correctness would begin. At this point, the tool should no longer be considered a prototype because it is being "certified" for operational use. There are two major types of correctness to be considered: safety and minimal functionality. With regard to safety, we wanted to show that the failure of any application would not interfere with other control center applications. For minimal functionality we wanted to demonstrate that both minimal service and minimal competency requirements are satisfied. Competency requirements (see [24] and [25]) define the "knowledge" or "intelligent ability" of the system. Service requirements would be all requirements that were not competency requirements. These would include, but are not be limited to, input and output formats, response time, processor the tool should run on, etc..

The general approach for this phase of V&V of the tool would be to validate the requirements by inspection, require the developer to verify the tool against the requirements, and then perform final validation via statistical testing. Statistical testing would involve running the tool in an operational environment for some period of time, recording any failures that might occur. This failure information will be used to predict an expected mean time in between failures (MTBF) of the system in operational use. We considered measuring MTBFs for safety, minimal service and minimal competency requirements.

Results

Unfortunately, we did not have an opportunity to apply the approach because no prototype made it to the point of being assessed for correctness. The primary reason for this was that much of the SSF architecture had changed by the time the prototype was ready for evaluation. So the only assessment that could be made was whether the prototype was a useful automation demonstration.

SUMMARY AND IMPLICATIONS FOR FURTHER TECHNOLOGY TRANSFER

Although it is felt that the material and approaches developed in this project have great potential to improve the state of the practice in KBS V&V (and in all software V&V), the results to date have only been moderately successful at best. Our initial concern and risk was that projects would be unwilling to try a sophisticated V&V approach because of the perceived cost. This is because we were targeting KBS projects which tend to be small and follow a rapid development (i.e., prototyping) process. To mitigate this risk, we strove to develop a streamlined V&V approach that involved a small number of techniques that had the best cost/benefit ratio (i.e., requirements and inspections) and/or directly addressed the problem of certification (i.e., statistical testing). The initial interest we received from the projects led us to believe that we were successful in this aspect.

Another problem that we did not fully appreciate was the length of time that would be required for a project to become self sufficient in following an established V&V approach. Our consulting engagement needed to be much longer than the six to eight months that we had, so that we could have followed each project to successful conclusion of at least a first release of the system.

A final problem in transferring technology to the target projects was the lack of a defined and enforced process. KBS projects have historically been small and involve rapid, highly iterative, development. This is true of KBS projects inside and outside of NASA ([5]). (And it may be true for most software development projects outside of NASA.) Because of this, there was no way for our suggested V&V approach to be officially adopted and enforced beyond our consulting engagement.

This project has important implications for the transfer of software engineering technology outside of NASA. NASA's software engineering methods and technology are among the best and NASA has a good reputation for building high quality software; therefore, NASA has much that could benefit others who do software development. However, many commercial projects are unlike the typical large, well-defined and safety critical NASA projects. KBS projects have many similarities with the typical commercial projects in that they are usually small, ill-defined applications that must be developed quickly. This does not necessarily imply that NASA's software technology is unsuitable for commercial projects. Because there is also a growing realization that more discipline and rigor is needed in many industries where software and KBSs are becoming key parts of safety critical systems, such as in medical devices.

Just as traditional and well-accepted V&V methods had to be adapted to fit the KBS projects discussed in this paper, NASA's software engineering methods will need to be adapted to fit the commercial software development environment. But, as evidenced by the survey discussed in this paper, such methods do appear to be needed. Also, based on the experiences discussed in this paper, transferring these adapted methods will require a systematic concerted effort. Simply making the techniques available to interested commercial software developers, as we tried to do with our KBS V&V workshop, will likely have minimal impact. These conclusions are consistent with the experiences of other attempts to transfer software engineering technology, such as those reported in [9].

REFERENCES

1. Becker, S.A. and Medsker, L., "The Application of Cleanroom Software Engineering to the Development of Expert Systems." Heuristics: The Journal of Knowledge Engineering. Quarterly Journal of the International Association of Knowledge Engineers (IAKE) Volume 4 Number 3, pp. 31-40. Fall 1991.

2. Fagan, M.E., "Design and Code Inspections to Reduce Errors in Program Development." IBM Systems Journal Volume 15 No. 3 pp. 182-211, 1976

3. Gries, D., The Science of Programming. Springer- Verlag New York, Inc. 1981.

4. Hamilton, D. and French, S., "A Design Language for Testable Expert Systems." Workshop Notes from the Ninth National Conference on Artificial Intelligence - Knowledge Based Systems Verification, Validation and Testing. July 17, 1991.

5. Hamilton, D., Kelley, K. and Culbert, C., "KBE V&V - State-of-the-Practice and Implications for V&V Standards." Workshop Notes from the Ninth National Conference on Artificial Intelligence - Knowledge Based Systems Verification, Validation and Testing. July 17, 1991.

6. Hamilton, D. and French, S.W., Workshop on Verification and Validation of Expert Systems. University of Houston/Clear Lake RICIS Contract #69 Deliverable #2, February 1992.

7. Hamilton, D., French, S.W. and Culbert, C., "An Approach to Improving the State-of-the-Practice in Verification and Validation of Expert Systems." Workshop Notes for the AAAI-92 Workshop on Verification and Validation of Expert Systems. July 16, 1992.

8. Hamilton, D. and French, S.W., Workshop on Verification and Validation of Expert Systems - Final Report. University of Houston/Clear Lake RICIS Contract #69 Deliverable #5, August, 1992.

9. Hamilton, D. and French, S.W., "Advancing the State of the Practice in Formal Verification", proceedings of AIAA Computing in Aerospace 9, October 19-21, 1993

10. Hoare, C.A.R., "Introduction to Proving the Correctness of Programs." ACM Computing Surveys. pp. 331-353, September 1976.

11. Howden, W.E., "Comments Analysis and Programming Errors." IEEE Transactions on Software Engineering. Volume 16 Number 1 pp. 72-81, January 1990.

12. Korson, T. and McGregor, J.D., "Understanding Object-oriented: A Unifying Paradigm." Communications of the ACM. Volume 33 No. 9 pp. 40-60 September 1990.

13. Landauer, C.A., "Correctness Principles for Rule- Based Expert Systems." Expert Systems with Applications. Pergamon Press. Volume 1 Number 3 pp. 291-316, 1990.

14. Linger, R.C., Mills H.D. and Witt, E.I., Structured Programming: Theory and Practice. Addison-Wesley Publishing Company 1979.

15. Liskov, B. and Guttag, J., Abstraction and Specification in Program Development. McGraw-Hill Book Company 1986.

16. Liu, N.K. and Dillon, T.. "An Approach Toward the Verification of Expert Systems Using Numerical Petri Nets." International Journal of Intelligent Systems. Volume 6 Number 3, pp. 255-276. June 1991.

17. Montalbano, Decision Tables. Science Research Associates, 1974 18. Myers, G.J.. The Art of Software Testing. John Wiley & Sons, Publishing 1979.

19. Myers, G.J.. Software Reliability Principles and Practices. John Wiley & Sons, Publishing 1976.

20. Myers, G.J.. Reliable Software Through Composite Design. Mason/Charter Publishers 1975.

21. Myers, G.J.. Composite/Structured Design. Litton Educational Publishing 1978.

22. Nazareth, D.L.. An Analysis of Techniques for Verification of Logical Correctness in Rule-Based Systems. pp. 80-136, Catalog #8811167-05150. UMI Dissertation Service, Ann Arbor, MI 48106, 1988.

23. Rumbaugh, J.. Object-Oriented Modeling and Design. Prentice-Hall, Inc. 1991.

24. Rushby, J. and Crow, J.. Evaluation of an Expert System for Fault Detection, Isolation and Recovery in the Manned Manuevering Unit. Final Report for NASA contract NAS1-182226 (NASA/Langley)

25. Rushby, J.. Quality Measures and Assurance for AI Software. NASA contractor report #4187, NASA Langley.

26. Yourdon, E. and Coad, P.. Object-Oriented Analysis . Prentice Hall, Inc. Englewood Cliffs, NJ 1990.

Session S5: MATHEMATICS, MODELING AND SIMULATION

Session Chair: Sam Veerasamy

RECENT ADVANCES IN WAVELET TECHNOLOGY

R. O. Wells, Jr.
Computational Mathematics Laboratory
Department of Mathematics
Rice University
Houston, Texas 77251
wells@rice.edu

March 24, 1994

1 INTRODUCTION AND OVERVIEW

In this paper I want to report on some recent developments in wavelet technology and, in particular, how it relates to some of the research activities at NASA. First, I want to indicate the nature of our research effort at Rice University in this direction. We have developed over the last four years a **Computational Mathematics Laboratory** (CML) housed in the **Computer and Information Technology Institute** (CITI) at Rice. This laboratory has as its primary focus research in the theory and applications of wavelets and more generally multiscale phenomena in mathematics, science and engineering. The researchers in the CML are:

- R. O. Wells, Jr., Professor of Mathematics (Rice), Director of CML
- C. S. Burrus, Professor of Electrical and Computer Engineering (Rice) and Director of CITI
- W. W. Symes, Professor and Chairman of Computational and Applied Mathematics (Rice)
- Roland Glowinski, Cullen Professor of Mathematics, University of Houston
- 4 Post Doctoral Fellows and 5 Graduate Students
- *Principal Support:* ARPA, NASA, Aware, Inc., Texas Instruments, Texas Higher Education Coordinating Board

Wavelet research has been developing rapidly over the past five years, and in particular in the academic world there has been significant activity at Rice, Yale, MIT, Delaware, Brown, S. Carolina, Washington Univ., Minnesota, Dartmouth, and numerous other universities. In the industrial world, there has been developments at Aware, Inc., Lockheed, Martin-Marietta, TRW, Kodak, Exxon, and many others. The government agencies supporting wavelet research and development include ARPA, ONR, AFOSR, NASA, and many other agencies. The recent literature in the past five years includes a recent book [6] which is an index of citations in the past decade on this subject, and it contains over 1,000 references and abstracts.

2 WAVELET MATHEMATICS

Fundamentally, wavelets are a new type of function which provide an excellent orthonormal basis for functions of one or more variables. They provide a localized basis, and can represent square-integrable functions, but also constant and, more generally, polynomial functions in a locally finite manner.

In 1988 Daubechies' fundamental paper on wavelets [1] appeared. In this paper we find for the first time a parametrized family of orthonormal systems of functions in $L^2(\mathbf{R})$ with certain important complementary properties. Each system of functions has the following properties:

- each system is generated from a scaling function $\varphi(x)$ and a wavelet function $\psi(x)$ by rescalings (by powers of an integer) and translations (e.g., $\varphi_{j,k}(x) := 2^{j/2}\varphi(2^j x - k)$ and $\psi_{j,k}(x) := 2^{j/2}\psi(2^j x - k)$. The wavelet system
$$\{\psi_{0,k}(x), \psi_{j,k}(x), \quad j,k \in \mathbf{Z}, \quad j \geq 0\} \tag{1}$$
is an orthonormal basis for $L^2(\mathbf{R})$ and more general functions as well (including constants and higher order polynomials, depending on the wavelet system chosen).

- each element in a given system has compact support and is continuous or can be chosen to be smooth up to a given finite order and by the rescaling above, the supports of the basis functions becomes very small for large scaling index j.

- There are fast algorithms for computing the coefficients of the expansion of a given digitized (sampled function). This is the *discrete wavelet transform* (from the sampled function to the wavelet expansion coefficients), and it is an $O(N)$ algorithm.

- The classical discrete Fourier and cosine transforms appear as a special case of the general discrete wavelet transform (DWT)

- The discrete wavelet transform is parallelizable and can ben implemented on massively parallel machines as well as can be designed into specializedd VLSI chips (e.g., for digital video editing).

In general a scaling function and corresponding wavelet function satisfy the *scaling equation*

$$\varphi(x) = \sum_{k=0}^{2g-1} a_k \varphi(2x-k) \tag{2}$$

and the corresponding *wavelet defining equation*

$$\psi(x) := \sum_{k=0}^{2g-1} b_k \varphi(2x-k), \tag{3}$$

where the coefficients of the scaling equation a_k must satisfy linear and quadratic constraints of the form:

$$\sum a_k = 2, \tag{4}$$

$$\sum a_k a_{k+2l} = 2\delta_{l,0}, \tag{5}$$

and where $b_k := (-1)^{k+1} b_{2g-1-k}$.

One of the powers of wavelet technology is the ability to choose the defining coefficients for a given wavelet system to be best adapted to the given problem. Daubechies developed in her original paper [1] specific families of wavelet systems which had maximal vanishing moments of the ψ function and which were very good for representing polynomial behavior. In Figure 1 we see the corresponding Daubechies scaling and wavelet function for the case of 4 coefficient ($g = 2$) where

$$\{a_0, a_1, a_2, a_3\} = \{\frac{1+\sqrt{3}}{4}, \frac{3+\sqrt{3}}{4}, \frac{3-\sqrt{3}}{4}, \frac{1-\sqrt{3}}{4}\} \tag{6}$$

and

In Figure 2 we see the contrast between the Fourier representation and wavelet representation for a given example of a transitory signal, and that the wavelet representation does provide a superior representation for this particular example.

3 WAVELET MULTISCALE REPRESENTATION OF DATA

If we consider such a wavelet system, and assume that there is a certain amount of smoothness (C^2, for instance), then we can try to use these functions as basis elements for representing discrete data at different scales.

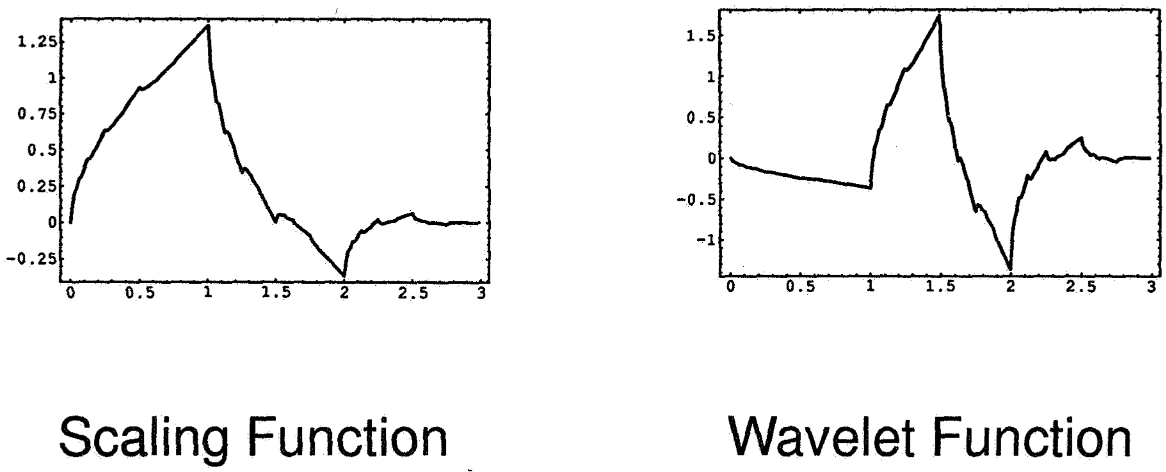


Figure 1: On the left is the 4-coefficient Daubechies scaling function and on the right is the corresponding wavelet function

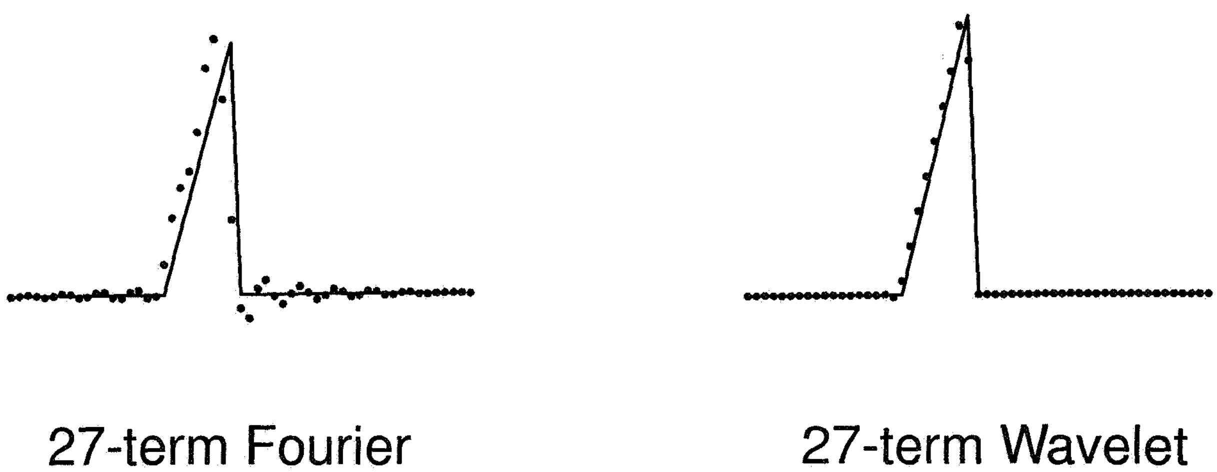


Figure 2: Comparison of a wavelet and Fourier representation of a transient signal

Namely, if we let for fixed $J \in \mathbf{N}$,

$$\tilde{f}(x) = \sum_k c_{Jk}\varphi_{Jk}(x), \qquad (7)$$

where c_{Jk} represents a sampling of a given function $f(x)$ at the points $x = k/2^J$, then $\tilde{f}$ is a smooth wavelet interpolation of our original sampled $f(x)$ *at the scale* J (or, what is the same thing, on a mesh with mesh size $h = 1/2^J$). Mallat [5] showed that from the scaling equations defining φ and ψ one can reexpress $\tilde{f}$ in terms of scaling and wavelet functions at coarser scales, namely:

$$\tilde{f}(x) = \sum_k c_{Jk}\varphi_{Jk}(x) = \sum_k c_{0k}\varphi_{0k}(x) + \sum_k \sum_{j=0}^{J-1} d_{Jk}\psi_{jk}(x). \qquad (8)$$

In (8) we see that the left hand side (LHS) represents the data at a single "fine" scale J, while the right hand side (RHS) gives a multiscale representation of the data at the coarser scales $\{0, 1, \ldots, J-1\}$. The Mallat transform consists of mapping the coefficients at the single scale on the LHS of (8) to the multiscale coefficients on the RHS of (8), and conversely (inverse Mallat transform). This transform consists of convolution with the filters which define the scaling and wavelet functions along with downsampling (and upsampling for the inverse transform).

4 IMAGE COMPRESSION AND TELECOMMUNICATIONS TECHNOLOGY

A major application of wavelets to technology has been in the area of data compression. The following list indicates the breadth of this application area. In each case the compression ratios indicated are what is roughly currently available, and are all products of Aware, Inc., of Cambridge, Mass., which is the leading commercial supplier of wavelet-based compression algorithms, in the form of software, chips, and plug-in boards for various application areas. Moreover, the compression ratio indicates compression to a version of the original signal which is indistinguishable from the original signal for the purposes at hand, and has been verified and tested by the industry experts in that given area. As one example, audio compression, listed at 8-1 compression ratio, has the property that the human ear cannot normally distinguish the compressed signal from the original, and the compression algorithm uses information about how the ear perceives sound and at what frequency scales.

- Audio compression - high fidelity at 8:1
- Still-image compression 20:1 (BW), 100:1 (Color)

- Seismic compression 20:1
- Radiology images 20:1
- Fingerprint images 25:1
- Video compression (color) 140:1

The basic idea in a compression algorithm in all of the above examples is to represent the digitized signal in terms of a wavelet expansion (the coefficients of this expansion will be the Discrete Wavelet Transform). Using a statistical analysis of the data type involved one carries out a systematic dropping of bits of these wavelet expansion coefficients at specific scales (this is the quantization process) to represent the same signal effectively with less bits, and an additional lossless compression is then applied to the result, which can then be either transmitted or archived. To recover the signal, one reverses the process with the exception of the quantization step, as those bits cannot be recovered. For further details about this compression process in the context of images, see, e.g., [12], and more information about specific technologies in all of the areas above is available, in particular, from Aware, Inc. in Cambridge, Mass.

One important feature of all of these algorithms is that one can download a compressed signal (or even an uncompressed signal represented in terms of its DWT), at any desired scale to obtain "snapshots" of the data, and download additional information later (or in the case of audio, to increase the fidelity at a later time). This technology is undergoing rapid development at the present time, and there is still much to be learned and understood in terms of modeling these compression ideas.

A second important area in which the DWT has played an important role is that of Assymetric Data Subscriber List (ADSL) technology. This is the basic copper wire twisted-pair communications link between American homes and their telephone companies. The spectral bands of this communication link are divided into three regions, the lowest being POTS ("Plain Old Telephone Service"), the second being a band for sending conventional digital data (linking computers for instance), and the high end of the band is reserved for digital video commmunication. The problem was that this was such a noisy channel that it was difficult to send video signals over this band in a meaningful manner. Very recently, Aware, Inc. announced a partnership with Analog Devices (a second Boston area company) to build transceivers which will implement such video communication in an effective manner, and this will be marketed to the telephone industry by an Alliance involving this partnership plus Westel, Newbridge, and AT&T, all of whom are involved in various aspects of the telecommunication industry. The technical report [8] which will appear soon in the proceedings of the International Communications Conference to be held in New Orleans in 1994 gives further information about this new advance in wavelet communications technology.

5 WAVELET-BASED NUMERICAL SOLUTIONS OF DIFFERENTIAL EQUATIONS

The wavelet represention of a sampled function of the form (7) allows one to use the scaling functions at a given scale (in this case at the arbitrary scale J corresponding to a mesh size of $h = 1/2^J$) as finite-element or Galerkin-type basis elements in a discrete approximation to some continuous problem (e.g., solving a partial differential equation numerically). In a number of recent papers these ideas have been carried out for various types of elliptic boundary value problems [11,10,9,2,4]. In addition one can use the multiscale representation of data as given in (8) to implement multigrid iterative schemes for solving such elliptic boundary problems where the solution by direct methods or by iterative methods at a single scale is prohibitive. In particular, one obtains an efficient multiscale algorithm for solving the model problem involving Laplace's equation for a domain with a very general boundary [3]. Moreover, a second model problem involving anisotropic coefficients in two dimensions with periodic boundary values admits a robust multigrid algorithm whose condition number is independent of the mesh size and of the anisotropy parameter [7]. In these multigrid applications of wavelets to numerical analysis the linear Mallat algorithms (transform from single scale to multiscale and conversely) play a major role. They allow one to map simply from one adjacent scale to another in a very effective manner, and that, along with the implicit orthogonality (and hence lack of redundancy), is one of the keys to their success in this applications (which is also true in their application to digital signal processing).

References

[1] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Comm. Pure Appl. Math.*, 41:906–966, 1988.

[2] R. Glowinski, J. Periaux, M. Ravachol, T. W. Pan, R. O. Wells, Jr, and X. Zhou. Wavelet methods in computational fluid dynamics. In et al. M. Y. Hussaini, editor, *Algorithmic Trends in Computational Fluid Dynamics*, New York, pp. 259-276, 1993. Springer-Verlag.

[3] R. Glowinski, A. Rieder, R. O. Wells, and X. Zhou. A wavelet multigrid preconditioner for Dirichlet boundary value problems in general domains. Computational Mathematics Laboratory TR93-06, Rice University, 1993.

[4] Roland Glowinski, T. W. Pan, R. O. Wells, and Xiaodong Zhou. Wavelet and finite element solutions for the neumann problem using fictitious domains. Technical Report 92-01, Rice University, 1992. Computational Mathematics Laboratory.

[5] S. Mallat. Multiresolution approximation and wavelet orthonormal bases of $L^2(\mathbf{R})$. *Trans. Amer. Math. Soc.*, 315:69–87, 1989.

[6] S Pittner, J. Schneid, and C. W. Ueberhuber. *Wavelet Literature Survey.* Institute for Applied and Numerical Mathematics, Technical University Vienna, Wien, Austria, 1993.

[7] A. Rieder, R. O. Wells Jr., and X. Zhou. A wavelet approach to robust multilevel solvers for anisotropic elliptic problems. *Applied and Computational Harmonic Analysis*, 1994. To appear.

[8] Michael A. Tzannes, Marcos C. Tzannes, John Proakis, and Peter N. Heller. DMT systems, DWMT systems and digital filter banks. Conference Proceedins of ICC '94, New Orleans. IEEE, 1994. (Aware, Inc. Technical Report AD940308).

[9] R. O. Wells and Xiaodong Zhou. Representing the geometry of domains by wavelets with applications to partial differential equations. In J. Warren, editor, *Curves and Surfaces in Computer Graphics III*, volume 1834, pages 23–33. SPIE, 1992.

[10] R. O. Wells and Xiaodong Zhou. Wavelet solutions for the Dirichlet problem. Technical Report 92-02, Rice University, 1992. Computational Mathematics Laboratory.

[11] R. O. Wells and Xiaodong Zhou. Wavelet interpolation and approximate solutions of elliptic partial differential equations. In R. Wilson and E. A. Tanner, editors, *Noncompact Lie Groups.* Kluwer, 1994. To appear. Proceedings of NATO Advanced Research Workshop.

[12] W. Zettler, J. Huffman, and D. Linden. Application of compactly supported wavelets to image compression. In *Proc. SPIE: Image Processing Algorithms and Techniques*, volume 1244, pages 150–160. Intl. Soc. for Optical Engineering, 1990.

A METHOD TO COMPUTE SEU FAULT PROBABILITIES IN MEMORY ARRAYS WITH ERROR CORRECTION

Gokhan Gercek
Lockheed Engineering and Sciences Company

S28-61
44849

Abstract:

With the increasing packing densities in VLSI technology, Single Event Upsets (SEU) due to cosmic radiations are becoming more of a critical issue in the design of space avionics systems. In this paper, a method is introduced to compute the fault (mishap) probability for a computer memory of size M words. It is assumed that a Hamming code is used for each word to provide single error correction. It is also assumed that every time a memory location is read, single errors are corrected. Memory is read randomly whose distribution is assumed to be known. In such a scenario, a mishap is defined as two SEUs corrupting the same memory location prior to a read. The paper introduces a method to compute the overall mishap probability for the entire memory for a mission duration of T hours.

I. INTRODUCTION

The radiation effects in spacecraft electronics evolving into a more significant problem with advances in semiconductor technology. The minituraztiation trends in microelectronics technology have created a new set of radiation problems for the designers of space avionics. As explained in [1], space radiation is a significant cause of errors in space borne memory devices. There are various ways to deal with radiation related problems. These can be avoidance, hardening, fault tolerance and SEU tolerance [2]. Avoidance is about, given a choice, operating in a less severe radiation environment. Another way to reduce the effects of the radiation is a technique called hardening. Hardening involves both processing changes which affect material and junction properties and circuit changes which reduce or eliminate degradation and failure mechanisms. Fault tolerance is associated with redundancy and voting mechanisms to reduce or eliminate radiation caused (and sometimes due to other reasons) errors and failures. The final technique, the SEU tolerance can be also considered in fault tolerance category. SEU tolerance is about those methods, tools and designs that would reduce SEUs or their effects [3], [4], [5].

There are several aspects of SEU related problems. First, SEUs create no significant damage to the circuit but only transient error conditions. This is mostly due to the fact that effects of SEUs are confined to (albeit not exclusively) bistable flip-flop storage elements. Secondly, SEUs mostly affect single bit storages and therefore single error correction techniques are accepted as a sufficient method of dealing with SEUs.

Despite the fact that error detection and correction mechanisms are quite effective in dealing with SEUs, one must remember the cumulative effects of SEUs in such designs. The cumulative effect of the SEUs refer to the situations where number of SEUs can cause multiple error conditions in a given word in a memory array every time. Obviously, this becomes a more pronounced problem when SEU rates are higher. It must be considered in all designs when the risk (i.e. the probability of occurrence times the cost incurred from the occurrence) is fairly high due to a SEU failure. The errors induced by space radiation are known as Single Event Upsets (SEUs).

Accumulation of errors in memory arrays with error detection and correction circuits can be reduced by deploying periodic "refresh" cycles (**scrubs**) where each memory cell is read and if it is in error corrected. By selecting sufficiently small refresh cycle durations, the probability of SEU error accumulation can be minimized. Another way of improving the SEU immunity in memory arrays is refreshing memory locations during the accesses. Everytime the program executing in the CPU accesses the memory, an error checking is performed on the contents of the memory word and when an error is found the contents of the memory location is refreshed.

The refresh approach can be costlier in CPU performance since periodic refreshes steals cycles from useful CPU operations. Memory accesses for refreshes introduce additional wait states resulting in slower CPU operations.

In this paper a memory array M words is considered. It is assumed that memory contents are refreshed overtime the memory is accessed. Furthermore, a simplifying assumption is that the memory locations are accessed for reading only. This is due to the fact that when a memory location is written into, the errors in pre-write state are irrelevant since they can not cause any failures.

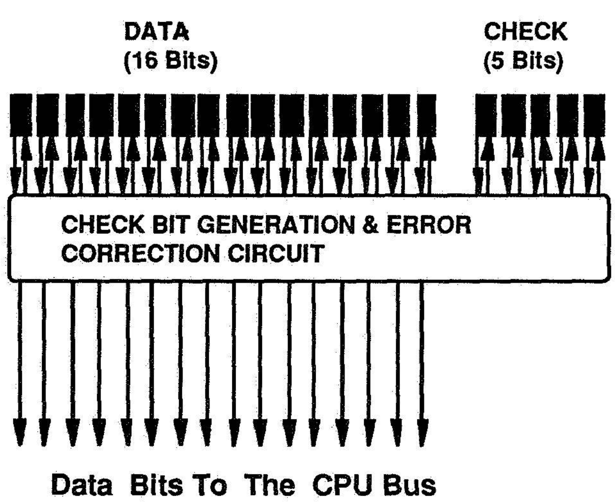


Figure 1. Word organization with check and data bits.

The access pattern to memory locations in a memory array is random in general. Therefore a memory access probability distribution is introduced to model the randomness. A bi-rectangular distribution is assumed for the derivations. However, the analysis can be carried out for any type of distribution without loss of generality.

As a memory array of M words with D data bits and C check bits is considered (i.e. total word length L=D+C). Figure 1 shows a word organization example with D=16 and C=5. The check bits are assumed to be capable of correcting single bit errors (such as Hamming code). It is also assumed that an SEU can not upset more than one bit of storage at a given time [1]. We define a "Mishap" as an error condition with more than one error accumulating in a memory location prior to a refresh. The reason for using the word "mishap" instead of "failure" is that, not every mishap can result in a failure necessarily. For example, if a memory array has some words which may never be accessed during the scrub period, then the Mishap can not result in a failure. We also assume a memory access rate of k times (randomly) per second. k can be taken roughly as the MIPs rating of the processor. We denote Λ as the SEU upset rate (upsets per unit time) for the entire memory. Thus the SEU arrival rate per word becomes $\lambda = \Lambda / M$ which is assumed to be Poisson distributed. We define the time unit, t_u, as a **quantum** which is the access time to the memory. Thus $t_u = 1/k$.

II. MEMORY PROFILE MODEL

Since the CPU accesses memory locations in a random manner, we define a memory access distribution profile or simply memory profile as the probability distribution of accessing any one of the M memory locations at a given time. Figure 2 shows the bi-rectangular distribution adopted for the subsequent analyses. Note that this is a discrete distribution with the independent variable being the address of a memory location. Although we use the bi-rectangular distribution for analytical simplicity, it can be shown that the analysis can be extended to any other type of distribution.

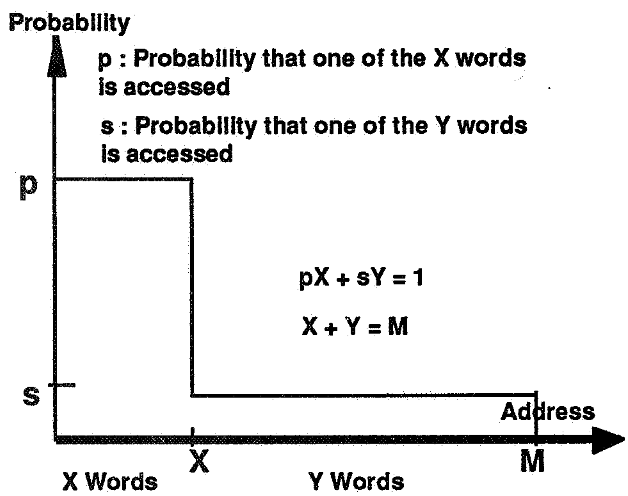


Figure 2. Bi-rectangular memory access probability distribution profile.

The memory profile in Figure 2 can be interpreted as follows: During a given access to the memory, there is "p' probability that we will read from a particular memory location between addresses 1 and X and there is q=1-p probability that we will not read from that particular location. We define Y = M - X. The asymmetric profile reflects the fact that certain parts of the memory (i.e. first X words or "X" type) are more frequently accessed than the next Y words or "Y" type words. Similarly during a given access, we have probability s that a "Y" type memory will be read and a probability r=1-s that particular location will not be read. Note also that due to conservation of probability, pX+sY=1.

III. ANALYSIS

Consider a X type memory location "A" in the memory profile. Assume that "A" is just accessed. For our subsequent analyses we will call the interaccess time for a given memory location the location inter read time (LIRT). The probability that "A" will be accessed again after N quanta is:

$$P\{ \text{LIRT for A} = N \} = P_N = p\, q^{N-1} \qquad (1)$$

As Equation (1) suggests, the LIRT of a given memory location is geometrically distributed.

Now let's consider the probability of two or more SEUs striking this memory location during the LIRT of N quanta. Note that if two or more SEUs corrupt the memory location, this would result in a mishap.

$$P\{ \text{two or more SEUs in N quanta} \} = 1 - P\{ 0 \text{ SEU in N quanta} \} - P\{ 1 \text{ SEU in N quanta} \}$$

Since SEU arrivals are assumed to be Poisson distributed with parameter λ,

$$P\{ 0 \text{ SEU in N quanta} \} = e^{-\lambda N}$$

$$P\{ 1 \text{ SEU in N quanta} \} = \lambda N e^{-\lambda N}$$

Thus

$$P\{ \text{two or more SEUs in N quanta} \} = 1 - e^{-\lambda N} - \lambda N e^{-\lambda N}$$

or

$$P\{ \text{Mishap in N} \} = 1 - e^{-\lambda N} - \lambda N e^{-\lambda N} \qquad (2)$$

And probability of success in N quanta will then be:

$$P\{ \text{Success in N} \} = 1 - P\{ \text{Mishap in N} \}$$

$$P\{ \text{Success in N} \} = 1 - e^{-\lambda N} - \lambda N e^{-\lambda N} \qquad (3)$$

Since LIRT is geometrically distributed as given by Equation 1, the expected value of LIRT is 1/p. This means that an X type of location is read on the average once every 1/p quanta. Thus

$$E\{ LIRT_X \} = 1/p \qquad (4)$$

In Equation (3), the probability of success is given under the assumption that the $LIRT_X$ is given as N. Since $LIRT_X$ geometrically distributed, the average probability of success during $LIRT_X$ can be found by:

$$P_s = P\{ \text{average success in } LIRT_X \}$$

$$P_s = P\{ \text{average success in } LIRT_X \} = \sum_{N=1}^{\infty} (e^{-\lambda N} + \lambda N e^{-\lambda N})\, p q^N \qquad (5)$$

Equation 5 can be separated into two infinite series and each can be individually computed to yield:

$$P_s = \frac{p e^{-\lambda}}{1 - q e^{-\lambda}} + \frac{p \lambda e^{-\lambda}}{(1 - q e^{-\lambda})^2} \qquad (6)$$

It should be noted that Equation 6 is a computation sensitive equation since the numbers involved are very small (e.g. $p \approx 10^{-6}$, $q \approx 1 - 10^{-6}$ and $\lambda \approx 10^{-15}$). If Equation 6 is computed using a typical set of numbers with a calculator, the resulting value for P_S would likely to be 1.0 due to the computation sensitivity of Equation 6.

In order to facilitate the computational problem, we can introduce the following form for P_S:

$$P_s = 1 - \varepsilon_x \qquad (7)$$

In Equation 7, ε_X is a very small number which represents the probability of mishap during an average $LIRT_X$. By using first order Taylor series approximation it can be shown that:

$$P_s = 1 - \frac{\lambda^2}{p} \qquad (8)$$

or equivalently

$$\varepsilon_x = \frac{\lambda^2}{p} \qquad (9)$$

Now let's assume that all memory locations are scrubbed every T many quanta. For a given memory location of type X, the probability of success in T quanta is:

$$P_{sT} = (1 - \varepsilon_x)^m \qquad (10)$$

Where m=T/(1/p) or the number of average size LIRTs in T. The probability that all the locations of type X survives during T quanta is:

$$P_{sT} = (1 - \varepsilon_x)^{mX} \qquad (11)$$

Since ε_x is a very small number and mX is a very large number Equation 11 can be approximated as:

$$P_{sTX} = 1 - mX\varepsilon_x \qquad (12)$$

Equation 12 is the survival probability for the first X words of the memory for a duration of T. By using similar arguments, for the Y type locations, the survival probability can be found as:

$$P_{sTY} = 1 - nY\varepsilon_Y \qquad (13)$$

In Equation 13, n=Ts , s being the probability of accessing a Y type location at a given read. ε_Y is defined in a similar way as ε_x in the following way:

$$\varepsilon_Y = \frac{\lambda^2}{s} \qquad (14)$$

The survival probability for the entire memory can then be computed as:

$$P_{sT} = (1 - nY\varepsilon_Y)\ \ (1 - mX\varepsilon_x)$$

or

$$P_{sT} = (1 - nY\varepsilon_Y - mX\varepsilon_x + mnXY\varepsilon_x\varepsilon_Y) \tag{15}$$

and the probability of a mishap in the entire memory for a duration of T can then be found as:

$$P_{mishapT} = nY\varepsilon_Y + mX\varepsilon_x - mnXY\varepsilon_x\varepsilon_Y \tag{16}$$

Example: Let's assume a memory of 250 KWords with a word size of 32 bits (without the checkbits), a memory profile as shown in Figure 2, an access rate of 5 million reads per second (i.e. quanta = 0.2 μsec.) and an SEU arrival rate of 10^{-15} upsets/word/quantum. Let's also assume that the memory is never scrubbed during the entire mission which lasts 30 days (i.e. T=720 hours). Using the analysis given in the paper, the probability of a mishap during the entire mission can be computed as $P_{mishapT} = 5\ 10^{-12}$.

IV. CONCLUSION

It is shown that SEU reliability of memory arrays with single error correction feature is predictable when a memory profile can be associated with the memory access patterns. Although the derivation is performed for a bi-rectangular profile, it is possible to extend the approach to general profile models. In case periodic scrubs are used, the analyses yield the result for one scrub cycle. The mishap probability for the entire mission can then be found by multiplying the number of scrubs in a mission with the mishap probability in one scrub cycle.

REFERENCES

[1] J. Bradford and P. Bail, "Cosmic ray effects in spaceborene electronics" Proc. Third Digital Avionics Systems Conf., Forth Worth, TX, Nov. 1979.

[2] R. D. Rasmussen, "Spacecraft Electronics Design for Radiation Tolerance", Proceedings of the IEEE, Vol. 76, No. 11, November 1988.

[3] G. Burdick, E. Kopf and D. Meyer, " The Galileo Single Event Upset Solution and Risk Assessment", Proc. Ninth Ann. Rocky Mountain Guidance and Control Conf. February 1986.

[4] P. Lewkowicz and L. Richter, "Single event upsets in spacecraft digital systems," Proc. Int. Telemetering Conf., October 1984.

[5] M. Sievers and G. Gilley, " An architectural approach to tolerating radiation effects," IEEE Ann. conf. on Nuclear and Space Radiation Effects Short Course, Monterey, CA, July 1985.

S29-61
44852

COMPUTER SIMULATION: A MODERN DAY CRYSTAL BALL?

Michael Sham
Lockheed Space Operations Company
1100 Lockheed Way
M/S LSO-003
Titusville, FL 32780

Andrew Siprelle
Siprelle Associates
402 North Maple Street
Maryville, TN 37801

ABSTRACT

It has long been the desire of managers to be able to look into the future and predict the outcome of decisions. With the advent of computer simulation and the tremendous capability provided by personal computers, that desire can now be realized. This paper presents an overview of computer simulation and modeling, and discusses the capabilities of *Extend*. *Extend* is an iconic-driven Macintosh-based software tool that brings the power of simulation to the average computer user.

An example of an *Extend* based model is presented in the form of the Space Transportation System (STS) Processing Model. The STS Processing Model produces eight shuttle launches per year, yet it takes only about ten minutes to run. In addition, statistical data such as facility utilization, wait times, and processing bottlenecks are produced. The addition or deletion of resources, such as orbiters or facilities, can be easily modeled and their impact analyzed. Through the use of computer simulation, it is possible to look into the future to see the impact of today's decisions.

OVERVIEW OF COMPUTER SIMULATION AND MODELING

Computer simulation and modeling is one of many types of decision support tools. A decision support tool, defined as something that aids in making a decision, can take on many different forms. The simplest of the tools is a coin toss, hopefully employed when the stakes are low, and they increase in sophistication from experience/intuition, spread sheets, math models, and computer simulation. Each decision support tool has its strengths and weaknesses, and therefore must be applied to the situation for which it is best suited. The advantage of computer simulation is that it provides the ability to model processes to predict outcomes in an interactive fashion. Although computer simulation and modeling has been used for many years, it has traditionally been used solely by large companies, universities, and government agencies due to the large investment in hardware, software, and specialized personnel required. Recent changes in the availability of software and hardware have brought the power of computer simulation and modeling to everyone. Figure 1 provides an overview of the three

Format	Strength	Weakness
Language Based	Can be Tailored Across a Range of Applications	Specialized Software Language General Simulations Requires Experts in Coding and Simulation to Operate
Data Driven	Uses Pre-Built Graphical Blocks	Limited to the domain (i.e., manufacturing) for which the blocks were built
Hybrid	Easy to use Customizable Blocks	

Figure 1. Simulation Software Formats and Their Characteristics

types of simulation formats that are currently available.

STS PROCESSING MODEL

An example of the use of simulation is provided by the STS Processing Model. The purpose of developing the Model was to determine the impact of changes such as the number of orbiters required to be processed, facility shutdown for modifications (to either the facility or to an orbiter inside the facility), major flight part unavailability, or GSE disruptions. Other changes, such as the processing impact of a new launch vehicle upon the facilities and the ability of the launch site to effectively process both vehicles can also be modeled. These changes can seriously impact facility utilization and annual launch rate. The facilities at Kennedy Space Center, as in all types of manufacturing or processing operation, are limited. The processing of the space shuttle is performed mainly in three types of facilities. The three Orbiter Processing Facility (OPF) bays are used to de-service and remove payload unique equipment from the orbiter after a mission, perform repairs and modifications, and install equipment and supplies in preparation for the next mission. The two Vehicle Assembly Building (VAB) bays are where the external tank (ET) is attached to the solid rocket boosters after they are stacked, and is also where the orbiter is attached to the ET. The two launch pads are used to prepare the vehicle for launch, including payload installation, fueling, and final checkout. Figure 2 presents an overview of the current STS processing flow.

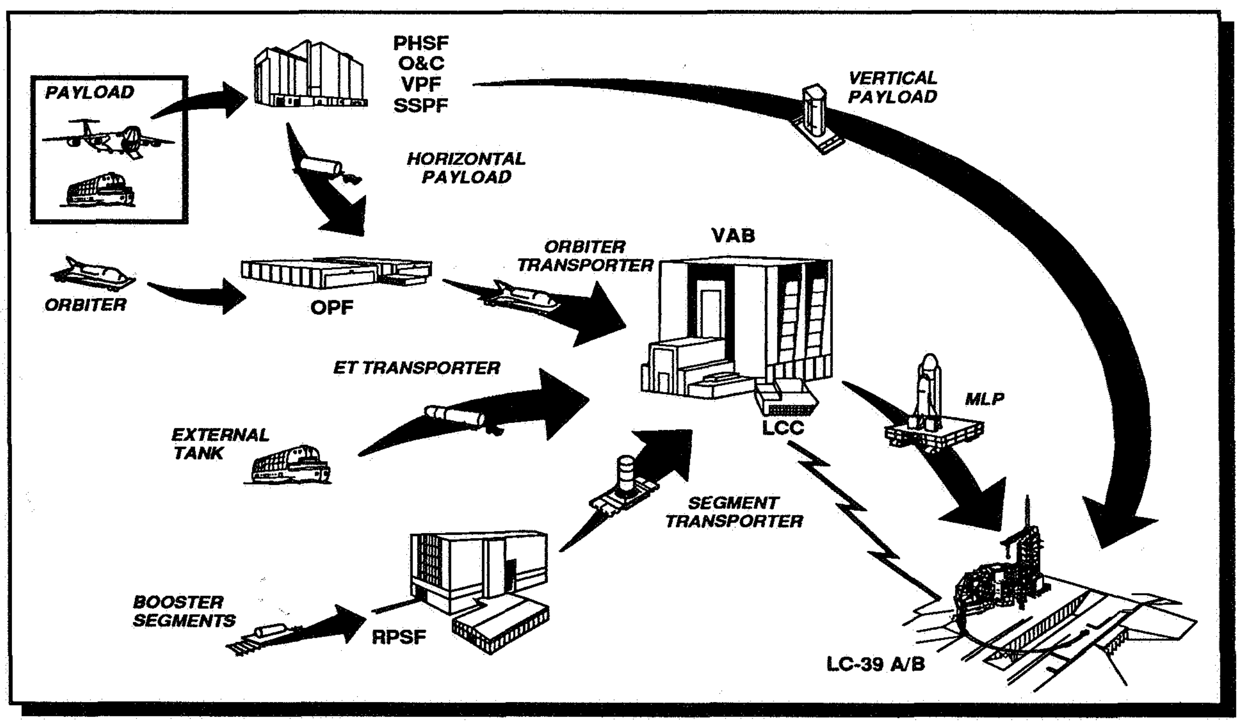


Figure 2. STS Processing Flow

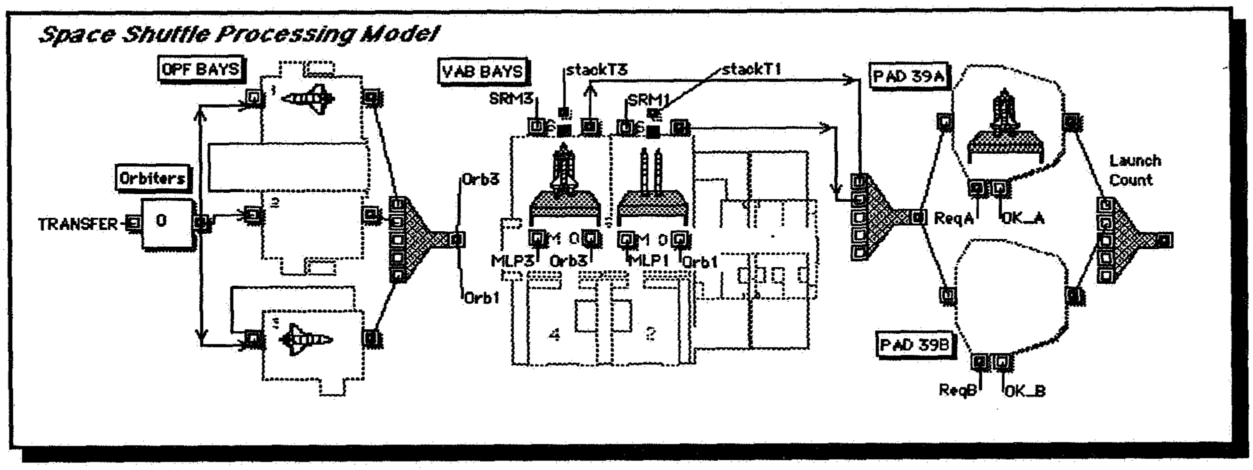


Figure 3. STS Processing Model-Overview

The STS Processing Model was developed through the use of the simulation software *Extend + Manufacturing*™ available from Imagine That, Inc., San Jose, Ca. *Extend* is hybrid, library based, iconic-block (graphical element) commercial-off-the-shelf (COTS) simulation software package. The Model operates in a discrete event mode, where events are orbiter movements or other status changes. Event times are driven by orbiter process durations and the resolution of resource conflicts. An orbiter's process duration is selected from a statistical distribution of achieved processing durations or a default constant. As can be seen in Figure 3, the simulation model of the launch site is very intuitive as the OPF, VAB, and launch pad footprints are used as part of the Model. The process flow on the screen is the same as the orbiter movement toward the launch pad. Additionally, icons of the shuttle, solid rocket boosters, and external tank provide visual clues as to the status of the integrated flow.

Through the use of an input/output screen, called the Notebook, assets such as orbiters or mobile launcher platforms (MLPs) can be added or deleted in order to perform "what-if" analyses. These type of changes take about 10 seconds to make, and it takes about 5 minutes to model a years worth of processing to determine the effects on the launch site. Processing times that the Model pulls from the statistical database, such as for the OPF, VAB, or pads, are also shown in the Notebook input screen (Figure 4) as the Model is running.

The output of the Model is shown in Figure 5. The output shows the achieved launch rate, yearly launch rate, facility utilization for each of the facility processing bays and launch pads, and MLP and orbiter availability. A spreadsheet within the Notebook also captures the as-run data for each processing flow so that comparisons and statistical analyses can be made to determine the results of each "what-if" run. Additional data elements can be added to the Notebook as desired.

Each of the facilities represented consist of a hierarchical block. A hierarchical block is composed of a series of logic blocks that represent the logic and events that occur within the facility. Through the use of hierarchical blocks, it is very easy to add or delete facilities to determine the effect on the processing flow, launch rate, or facility utilization. After a hierarchical block is created, it can be added to a library, such as the STS Processing Library, and used to add the facility in the processing flow as desired. It is also easy to delete a facility, simply by selecting and deleting it, to determine the subsequent effect on the processing flow. Either change,

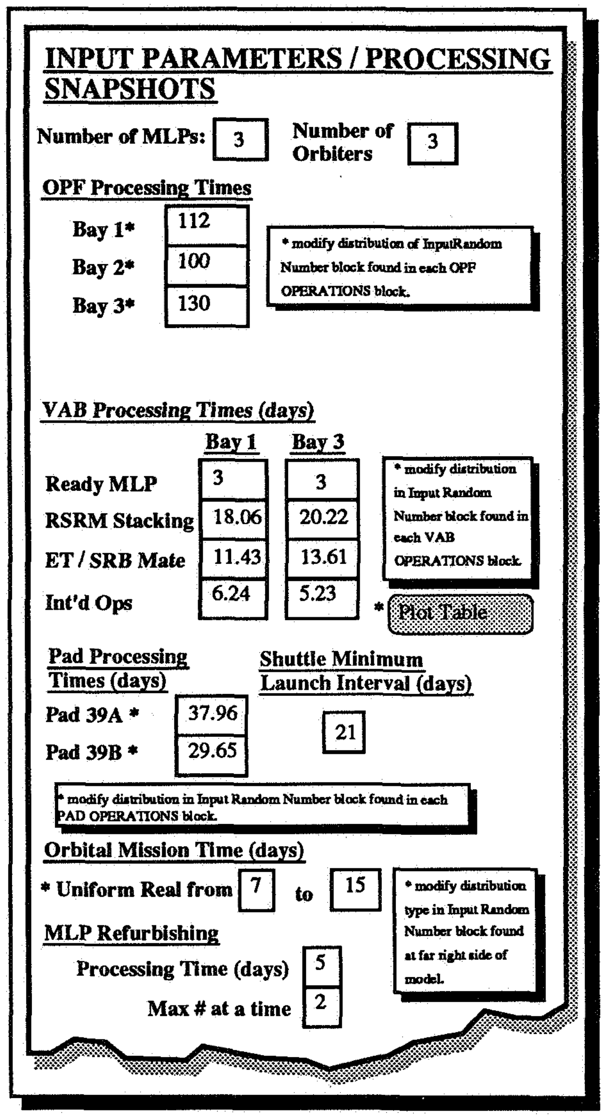


Figure 4. STS Processing Model NotebookInput Screen

whether adding or deleting a facility, takes less than 15 seconds to implement.

Due to its nature as a hybrid simulation package/language, *Extend* enables people with a wide range of ability to change the Model at many levels of detail. The user can double-click on block icons and change dialog parameters. From libraries supplied, the user can get new blocks, connect them, and enter parameters. The simpler groundrules are represented in Equation blocks, so the user can change these or add new blocks to modify groundrules. For the most flexibility, the user can create new primitive blocks by using *Extend*'s built-in C compiler and dialog/icon editors to either modify a pre-existing block or build one from scratch. Most blocks needed are already pre-built. In fact, all but one of the blocks used to build the Model are pre-built.

OTHER USES OF SIMULATION

In addition to the STS Processing Model, *Extend* and simulation have been used in a wide variety of instances where it was important to

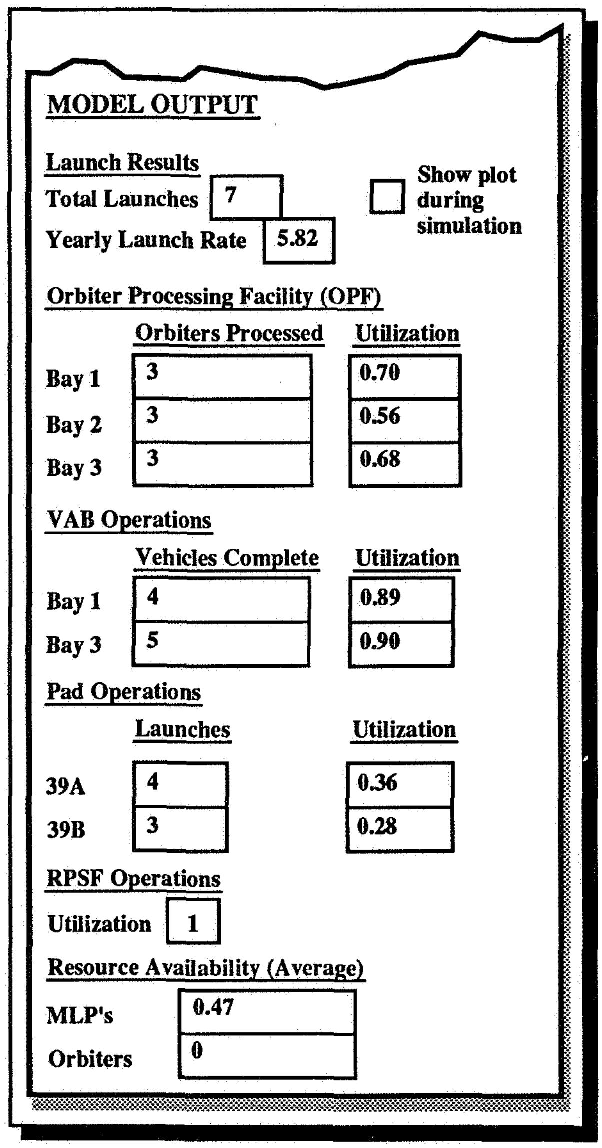


Figure 5. STS Processing Model Notebook Output

see the impact of changes before they were implemented. Lennon Associates, an architectural firm, uses simulation as part of the design and specification process. They modeled the emergency room requirements for Grossmont Hospital in preparation for construction of a new emergency room. Using data compiled from the hospital's records, Lennon found that there were nearly 400 different outcomes that could happen to a patient upon entering the emergency room door. By simulating the actual emergency room requirements, bed requirements were determined to be lower than what the hospital thought they needed, lowering construction and operating costs. Although the emergency room has not yet been built, it is estimated that the simulation will save the hospital $1 million in construction costs and $300,000 per year in operating costs.

BDM International used *Extend* to verify the engineering design of equipment for a military contract, saving $10-15 million. The *Extend* simulation reduced the need for testing and simplified the analysis of the project. It also reduced the size and number of reports required. The main savings occurred because the project team used *Extend* to develop a hardware design that was less expensive than what was currently available, and used simulation as a proof of concept.

Gentek, Inc, a company that specializes in engineering and financial modeling, used *Extend* for an expansion project undertaken by Ben and Jerry's Ice Cream Co. The simulation modeled their 10-year requirements for electricity and the effects and trade-offs (including operation costs) of using alternative sources of energy such as thermal storage. The co-generation plant they selected allowed Ben & Jerry's to expand without forcing the utility companies to increase their capacity.

CONCLUSION

Until recently, the use of computer simulation was limited to those entities with the large amount of resources required to take advantage of the technology. However, due to the recent advances made in hardware and software technology, computer simulation is a tool that is available to anyone desiring to take a peek into the future.

BIOGRAPHIES

Michael Sham (407-383-2200 ext. 2427) is a member of the Advanced Programs Study Team for Lockheed Space Operations Company at Kennedy Space Center. Mr. Sham's current responsibilities include the analysis and evaluation of new space launch systems in terms of their impact to the existing launch infrastructure. He received his BS in Marketing and Management from Oklahoma State University and a MBA with an emphasis in statistics from the Florida Institute of Technology.

Andrew Siprelle (615-982-7046) is President of Siprelle Associates, a firm which provides simulation consulting, training, and custom models. Mr. Siprelle's industry experience includes creation and analysis of deterministic and stochastic models for strategic planning, application of Experimental Design for process improvement, and training employees in industrial statistics. He has used *Extend* to help numerous companies analyze operations and increase productivity. Mr. Siprelle specializes in statistical analysis of simulation output and simulation experiments. He received his BS in Industrial Engineering and Operations Research from Virginia Polytechnic Institute and State University.

amt

Session S6: NETWORKS, CONTROL CENTERS, AND DISTRIBUTED SYSTEMS

Session Chair: Zafar Tacqvi

530-62
44855

AVERAGE WAITING TIME IN FDDI NETWORKS WITH LOCAL- PRIORITIES

Gokhan Gercek
Lockheed Engineering and Sciences Company

Keywords: FDDI networks, priority queues, computer networking, average waiting times, local priority access, network performance.

Abstract: A method is introduced to compute the average queuing delay experienced by different priority group messages in an FDDI node. It is assumed that no FDDI MAC layer priorities are used. Instead, a priority structure is introduced to the messages at a higher protocol layer (e.g. network layer) locally. Such a method was planned to be used in Space Station Freedom FDDI network. Conservation of the average waiting time is used as the key concept in computing average queuing delays. It is shown that local priority assignments are feasable specially when the traffic distribution is asymmetric in the FDDI network.

I. INTRODUCTION

Analysis of priority based queuing systems have attracted first interests from mathematicians and operation research analysts. With the explosive growth of computing systems and computer networks it has found new application grounds and thus has become a research area in these fields as well. Priority queuing disciplines are more complex in nature than their non-priority counterparts due to multidimensional state variables involved in the description of the system.

Among the queuing disciplines that impose a static priority structure (exogenous), head-of-the-line (HOL) discipline is perhaps the most common and most intuitive one. HOL discipline was first studied by Cobham [1] who derived an expression for the average waiting times associated with different priority groups in the queue. Later, Kesten [2] derived a formula for calculating the Laplace transform of the waiting time in an HOL queue, which is usually very difficult to invert. Davis [3] introduced a method to evaluate the waiting time distributions associated with a two level priority queue. A way of calculating the probability density functions for a multilevel HOL queue under identical service time assumption is given in [4]. An important study concerning the moments of the waiting times in an HOL queue is performed by Miller [5].

In computer network performance analysis, the derivation of an average waiting time expression associated with a computer network is an elaborated task. Under priority access, the derivation becomes even more complex and in most cases intractable. As will be shown in the subsequent section, it is possible to find the waiting time for an FDDI network with local priorities since an analysis for the no-priority counterpart is readily available.

The underlying assumptions in the average message time derivation are:

1. Exhaustive service discipline: The FDDI station which is using the communication medium transmits all its messages starting with the highest priority ones. It should be noted that in practice, the token holding time used by MAC layer does not permit exhaustive transmission of messages. Therefore, exhaustive service assumption is an approximation. In implementation,

however, this issue can be circumvented by setting the token holding time as large as possible.

2. Local Priority Assignment: Message priority assignments have only local meanings. That is other FDDI stations are not aware of the message priorities in the buffers of a particular station.

Organization of the paper:

The subsequent sections are organized in the following way. In section II a method is introduced to transform the FDDI network into a classical priority queuing system. Then the solution for the resulting priority queuing system is given. In section III the method is applied to FDDI networks with symmetric and asymmetric traffic patterns. Section IV is the conclusion.

II. WAITING TIME IN FDDI NETWORKS WITH LOCAL PRIORITIES

A. Modeling of FDDI networks with Local Priorities.

As will be explained in the following paragraphs, the FDDI network considered in this paper can be analyzed in terms of standard priority queues. The results from HOL priority systems will constitute the framework of the subsequent discussions.

Average Waiting Time in HOL queues:
For an N level HOL queue, with level N as the highest and level 1 as the lowest priority level (Fig. 1), the average waiting time, W_p, associated with each level can be found according to the formula [1]:

$$W_p = \frac{W_R}{(1-\sigma_p)(1-\sigma_{p+1})} \quad , \qquad p = 1,2,......N \qquad (1)$$

where

$$\sigma_p = \sum_{k=p}^{N} \rho_k \qquad and \qquad \sigma_{N+1} = 0$$

W_R is the remaining service time of the entity found it the service and ρ_k is the traffic intensity into level k. Using renewal theory arguments [6], and assuming Poisson arrivals, W_R can be calculated as:

$$W_R = \sum_{j=1}^{N} \rho_j \frac{b_j^{(2)}}{b_j} \qquad (2)$$

where $b_j^{(2)}$ and b_j are the second and first moments of the message length distributions respectively. As can be noticed from Eq. 1, the waiting time of a level k entity is not affected by the entities in levels k-1, k-2,...1, except for the contribution of these levels to W_R. Furthermore assuming Poisson arrivals to all levels, the average waiting times satisfy the conservation law of the waiting times [7].

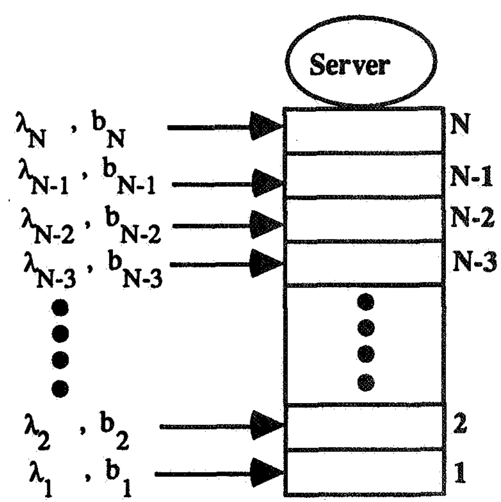


Figure 1. Head Of the Line (HOL) queue.

B. Vacation Time of the Server

Here we consider an FDDI network with M stations sharing the common fiber optic communication medium (the server). When a station "i" gains the access right to the transmission medium, it transmits all of its messages exhaustively and then passes the token to the next station according to FDDI protocol. At this point, from the viewpoint of the station "i" , the server is considered to be in vacation. The vacation time of the server is a random variable whose distribution is generally unknown. The station "i" can reaccess to the medium at the completion of the vacation time (Fig. 2).

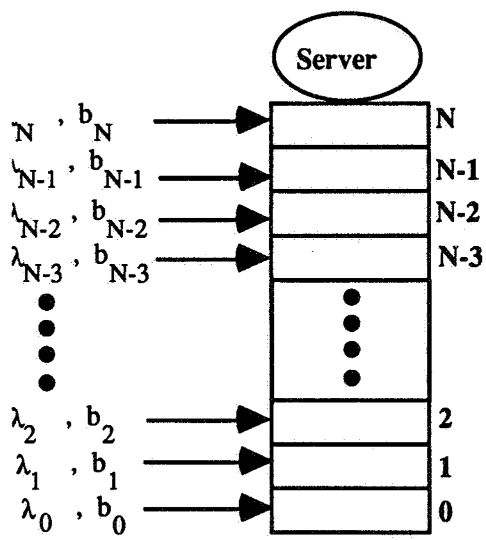


Figure 2. Representation of an M station local priority LAN in terms of an HOL queue.

C. Incorporating Vacation Time into Priority Queue Model

From the viewpoint of a station i, the vacation time of the server can be treated as just another cause (or a hypothetical message) which keeps the server busy. Moreover this message is processed always after the completion of the services of the other messages. Therefore if the service time of the hypothetical message representing the vacation time is a random variable b_0, then the vacation time can be incorporated into the priority queue structure associated with the station of interest as an additional priority level (level 0). It should be noted that, in order to properly mimic vacation time, there must always be a message available to service in this priority level. When the server completes the service of all messages in N levels, if it can not find a message in the level 0, then the modeling will not be valid.

Fig. 2 shows modeling of a local area network in terms of an N+1 level HOL queue. The requirement that there must be a pool of entities in the priority level 0 can be satisfied by adjusting the Poisson arrival rate into this level so that the overall traffic rate approaches unity. The waiting times associated with the new N+1 level priority queue are given by:

$$W_p = \frac{\overline{W}_R}{(1-\sigma_p)(1-\sigma_{p+1})} \quad , \qquad p = 0,1,2,\ldots\ldots N \tag{3}$$

where

$$\overline{W}_R = \rho_0 \frac{b_0^{(2)}}{b_0} + \sum_{j=1}^{N} \rho_j \frac{b_j^{(2)}}{b_j} \tag{4}$$

It should be noted that the terms $b_j^{(2)}$ and b_0, in Eq-4 are not known since the distribution of the vacation time is not available. Nevertheless W_R can be determined by using the results for a queue with N=l (i.e. non-priority) which has a traffic intensity same as that of the N level queue. Now, let's assume an FDDI network which has the same access protocol as the assumed N priority level FDDI network. Also let us assume that the overall traffic intensity at a station 'k' are same for all stations in both cases. Remembering that the vacation time can be modeled as an additional level, we can write the following relation using the conservation law of the waiting time.

$$\rho_0 W_0 + \sum_{j=1}^{N} \rho_j W_j = \hat{\rho}_0 \hat{W}_0 + \hat{\rho}_1 \hat{W}_1 \qquad (5)$$

The variables marked with caret belong to N=l case. Obviously W_0 is same as W_0, and ρ_0 and ρ_0 are equal. Therefore;

$$\sum_{j=1}^{N} \rho_j W_j = \hat{\rho}_1 \hat{W}_1 \qquad (6)$$

The term W_1 in Eq. 6 is the waiting time of the messages in a non-priority FDDI network and it should satisfy Eq. 3 for N=1. Assuming that ρ_1 and σ_1 are the same, we have;

$$\sum_{j=1}^{N} \lambda_j b_j = \hat{\lambda}_1 \hat{b}_1 \qquad (7)$$

and if further assume that W_R's are the same in both networks then:

$$\rho_0 \frac{b_0^{(2)}}{b_0} + \sum_{j=1}^{N} \rho_j \frac{b_j^{(2)}}{b_j} = \hat{\rho}_0 \frac{\hat{b}_0^{(2)}}{\hat{b}_0} + \rho_1 \frac{\hat{b}_1^{(2)}}{\hat{b}_1} \qquad (8)$$

but since the priority level-0 arrival and service length distributions are the same;

$$\sum_{j=1}^{N} \rho_j \frac{b_j^{(2)}}{b_j} = + \rho_1 \frac{\hat{b}_1^{(2)}}{\hat{b}_1} \qquad (9a)$$

or

$$\sum_{j=1}^{N} \lambda_j \frac{b_j^{(2)}}{2} = \hat{\lambda}_1 \frac{\hat{b}_1^{(2)}}{2} \qquad (9b)$$

If $(\rho_1 + \rho_0)$ is allowed to approach to unity, then the queue lengths in level-0 become instable and thus in these levels constant presence of entities will be assured.

Now we are in a position to summarize the method for finding waiting times in an FDDI station with N priority levels. It is assumed that the arrival and message length distributions are known for all M stations in the network and average waiting time expression is available for the non-priority version of the same network. The following three steps outline the approach.

1. Treating the station as a non-priority station and using the average waiting time is calculated.

$$\hat{b}_1 = \sum (\lambda_j\, b_j\,) / \hat{\lambda}_1$$

$$b_1^{(2)} = \sum (\lambda_j\, b_j^{(2)}\,) / \hat{\lambda}_1$$

2. Using Eq. 3 for N=l, W_R is calculated

$$\hat{W}_R = W_1\,(1 - \hat{\rho}_1\,)$$

3. W_R substituted in Eq. 3 to determine W_p's for p=1,2,....N.

$$W_p = \frac{W_R}{(1-\sigma_p)(1-\sigma_{p+1})} \quad , \qquad p = 1,2,......N \qquad (10)$$

III. COMPUTATION OF LOCAL PRIORITY FDDI NETWORK AVERAGE WAITING TIMES

In this section we apply the algorithm developed above to two types of traffic patterns: an FDDI network with a symmetric traffic pattern and an FDDI network with an asymmetric pattern.

A. FDDI network with symmetric traffic pattern:

The network is assumed to have a symmetric overall traffic pattern for all M stations. Stations may have different number of priority levels with the provision that the overall traffic intensity to all N levels is constant and same for all the stations. The average waiting time for the non-priority version of the polling protocols is given by [8]:

$$\hat{W}_1 = \frac{1}{2}\left(\frac{M\hat{\lambda}_1 \hat{b}_1^{(2)}}{(1-M\hat{\rho}_1)} + rC_r^2 + \frac{(1-\hat{\rho}_1)Mr}{(1-M\hat{\rho}_1)}\right) \quad (11)$$

where M is the number of stations, $\hat{\lambda}_1$ is the arrival rate into a station, $\hat{\rho}_1$ is the traffic intensity into a station, $\bar{r}$ is the average walking time of the token (time to transfer the access right from one station to the next one) and C is the coefficient of variation of the walking time.

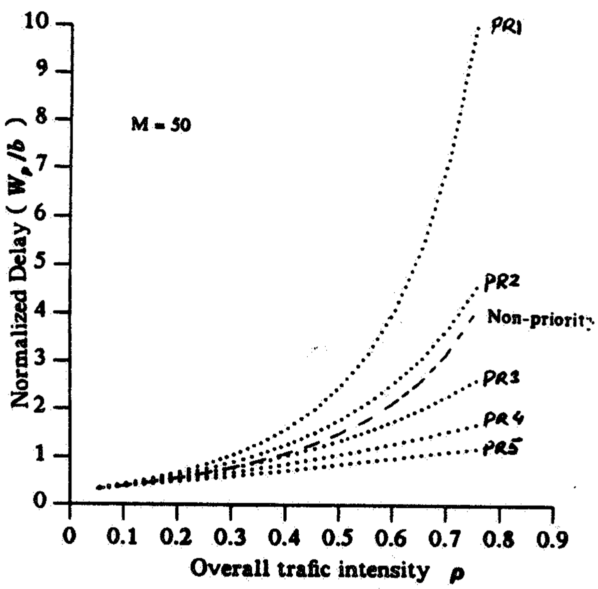


Figure 3. Local Priority FDDI protocol with symmetric arrivals ($\bar{r}$=0.01).

Substituting $\hat{W}_1$ from Eq. 11 in Eq. 10 yields the average waiting times associated with a FDDI protocol which has N distinct priority levels. In Fig. 3 and Fig. 4 the results for a 5 priority level FDDI protocol are shown. The message lengths are assumed to be exponentially distributed and arrivals are assumed to be Poisson distributed for all levels with identical parameters.

The figures also give the average waiting times associated with a non-priority FDDI network. In both cases, r is assumed to be constant. As can be observed it is possible to achieve significant improvements in the average waiting time of the high priority entities at the expense of the low priority ones.

B. FDDI network Asymmetric Traffic Patterns:

In this case the FDDI network is assumed to have a heavily unbalanced arrival pattern. While a particular station generates all the traffic in the network, other stations idle and pass the token to the subsequent station. Other assumptions are same as the symmetric case. The average delay expression for this case is given by [8]:

$$\hat{W}_1 = \frac{1}{2}\left(\frac{\hat{\lambda}_1 \hat{b}_1^{(2)}}{(1-M\hat{\rho}_1)} + \bar{r}C_r^2 + M\bar{r}\right) \quad (12)$$

The definitions for $\bar{r}$, C and M are same as before. $\hat{\lambda}_1$ is the arrival rate into the station, $\hat{\rho}_1$ is the traffic intensity into the station, $\hat{b}_1^{(2)}$ is the second moment of the message length distribution. Proceeding in similar way as done for symmetric arrival pattern, the average waiting time associated with different priority levels can be determined. Fig. 5 shows the results for an asymmetric FDDI protocol under similar assumptions as the previous case.

In Fig. 4 an interesting (and counter intuitive) trend is observed for high priority messages (priority 5 and 4).

The delays experienced for those classes of users decrease with the traffic intensity. A qualitative explanation for this situation can be given as follows. At low traffic intensities, waiting time of a message is due to token circulation time. When ρ approaches to 0 all the entities experience an average of $M\bar{r}/2$ delay. At high traffic intensities token spends less time circulating freely and more time serving the station since with a high probability there will be some entities waiting in the buffer. And high priority entities enjoy the increased availability of the token by not waiting for the token to arrive. This effect becomes less visible as Mr decreases.

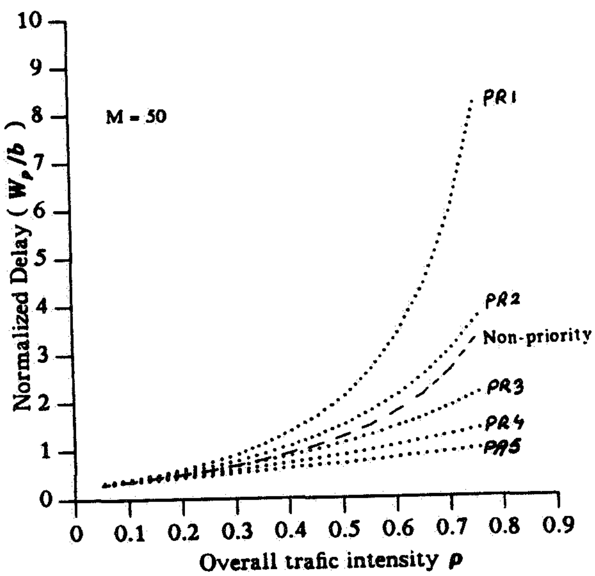


Figure 4. Local priority FDDI protocol with assymetric arrivals ($\bar{r}$=0.01)

IV. CONCLUSIONS

Average waiting time analysis of FDDI networks under local priority assignments become possible by modeling the vacation time of the server as an additional priority level. Then the problem can be treated as a standard HOL priority queuing problem. Lack of information about the distribution of the vacation time of the server can be overcome since analytical results due to the non-priority version of the same FDDI network is readily available. Two important assumptions which affect the validity of the results are the locality of the priority assignments and the exhaustiveness of the service discipline.

REFERENCES

[1] A. Cobham, "Priority Assignment in Waiting Line Problems," *Operations Research-ch,* 2, pp. 70-76, 1954.

[2] H. Kesten and J. T. Runnenberg, "Priority in Waiting-Line Problems I and II," *Nederlandse Akademie van Wetenschappen,* Amsterdam, Proceedings, Series A, 60, pp. 312-324, 325-336, 1957.

[3] R. Davis, "Waiting time distribution of a multiserver queuing system," *Operations Research,* vol. 14, pp. 133-136, 1966.

[41 G. Gercek and G. Davis, "Waiting time pdf's in multilevel static priority queues," *IEEE Phoenix Conference on Computers and Communications,* 1988 .

[5] R. G. Miller, "Priority queues," Ann-Math.Stat., vol. 31, pp. 86-103.

[6] D. R. Cox,*Renewal Theory.* London: Methuen, 1962.

[7] L. Kleinrock, "A conservation law for a wide class of queuing disciplines *Naval Research Logistics Quarterly,* vol. 12, pp. 181-192, 1965

[8] Rubin and L. F. M. De Moraes, "Message delay analysis for polling and token multiple-access schemes for local communication networks," IEEE *Journal on Selected Areas in Communicatons,* vol. SAC-1, No. 5, November, 1983.

S31-82
44857

The Electronic Documentation Project In The NASA Mission Control Center Environment

Lui Wang
Albert Leigh[1]
Software Technology Branch (PT4)

Information Systems Directorate
National Aeronautics And Space Administration Agency
Lyndon B. Johnson Space Center
NASA Road 1
Houston, TX 77058

NASA's space programs like many other technical programs of its magnitude is supported by a large volume of technical documents. These documents are not only diverse but also abundant. Management, maintenance, and retrieval of these documents is a challenging problem by itself; but, relating and cross-referencing this wealth of information when it is all on a medium of paper is an even greater challenge. The Electronic Documentation Project (EDP) is to provide an electronic system capable of developing, distributing and controlling changes for crew/ground controller procedures and related documents. There are two primary motives for the solution. The first motive is to reduce the cost of maintaining the current paper based method of operations by replacing paper documents with electronic information storage and retrieval. And, the other is to improve the efficiency and provide enhanced flexibility in document usage.

Initially, the current paper based system will be faithfully reproduced in an electronic format to be used in the document viewing system. In addition, this metaphor will have hypertext extensions. Hypertext features support basic functions such as full text searches, key word searches, data retrieval, and traversal between nodes of information as well as speeding up the data access rate. They enable related but separate documents to have relationships, and allow the user to explore information naturally through non-linear link traversals. The basic operational requirements of the document viewing system are to: provide an electronic corollary to the current method of paper based document usage; supplement and ultimately replace paper-based documents; maintain focused toward control center operations such as Flight Data File, Flight Rules and Console Handbook viewing; and be available NASA wide.

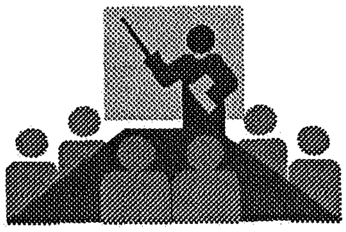
[1] Albert Leigh is currently with LinCom Corporation on the Technology Development Contract to support the Software Technology Branch, NASA Johnson Space Center, Houston, TX.

NASA Johnson Space Center
Information Systems Directorate

EDP

The Electronic Documentation Project In the NASA Mission Control Environment

Dual-Use Space Technology Transfer Conference and Exhibition



Presented by

Lui Wang/PT4

Albert Leigh/LinCom

Software Technology Branch

ABL- 3/30/94 1

NASA Johnson Space Center
Information Systems Directorate

Overview

- The EDP Team
- Background
- Documents
- Objectives
- Viewer System
- Workflow System
- System Infrastructure
- Conclusion

Software Technology Branch

ABL- 3/30/94 2

EDP Team

NASA Johnson Space Center

- Mission Operation Directorate (EDP project management, hardware/software evaluation and selection)
- Information Systems Directorate (software development, hardware/software consultation)

Other NASA Centers

- Ames Research Center (Flight Planning System to be compatible with EDP viewer)
- Jet Propulsion Lab (hardware consulting for storage, software consulting for library management)

Background

Space Program Support

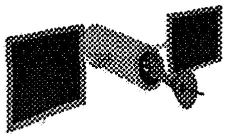
- Large volume of technical documents
- Challenging problem associated with paper-based system:
  - management and maintenance
  - storage and retrieval
  - cross-reference information
  - distribution and exchange

New Direction

- Emergence of new industrial/international standards and hardware/software systems

NASA Johnson Space Center
Information Systems Directorate

Documents



- Flight Data File (FDF)
  - procedural checklists, timelines, schematics, charts, cue cards, photos, uplinked text messages, etc
- Flight Rules
- Operator Console Handbook
- Space Shuttle System Handbook Drawings

Software Technology Branch

ABL- 3/30/94 5

NASA Johnson Space Center
Information Systems Directorate

Objectives

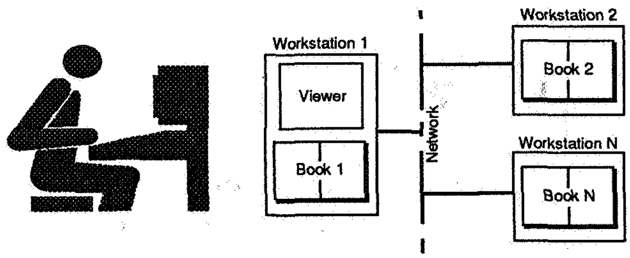
- Develop and electronic documentation system for flight controllers
- Two major goals to provide:
  1. An electronic document <u>viewing system</u> in the office environment and in the Mission Control Center (MCC)/Control Center Complex (CCC)
  2. An integrated electronic <u>work flow system</u> which includes FDF distribution, 482 tracking, e-mail, and electronic signature, etc.
- Expand EDP from MOD to JSC, NASA centers, commercial/educational sectors

Software Technology Branch

ABL- 3/30/94 6

NASA Johnson Space Center
Information Systems Directorate

Viewer



EDP

- Developed Level-A Requirements
- Evaluated NASA, COTS and Internet products
- Selected NASA/JSC (PT4)'s HyperMan 2.0 viewing software that provides full-range of hypertext capabilities

Software Technology Branch

ABL- 3/30/94 7

NASA Johnson Space Center
Information Systems Directorate

Viewer

Requirements

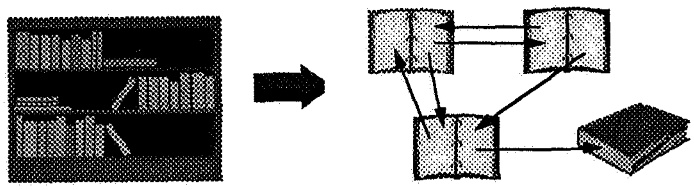
- Page-based presentation
- Topic-based browsing
- Multiple user access to a document
- Multiple document access by a user
- Configuration controlled document viewing
- Hypertext features

Software Technology Branch

ABL- 3/30/94 8

Hypertext

- On-line presentation of large amounts of loosely structured information
- Non-linear traversal
- Electronic cross-referenced information
- Search capability
- Annotations



Evaluation

- NASA Products: HyperMan, CID, HyLite
- Internet Products: Mosaic, etc.
- COTS: Interleaf, Frame, DynaText, etc.

NASA Johnson Space Center
Information Systems Directorate

Workflow

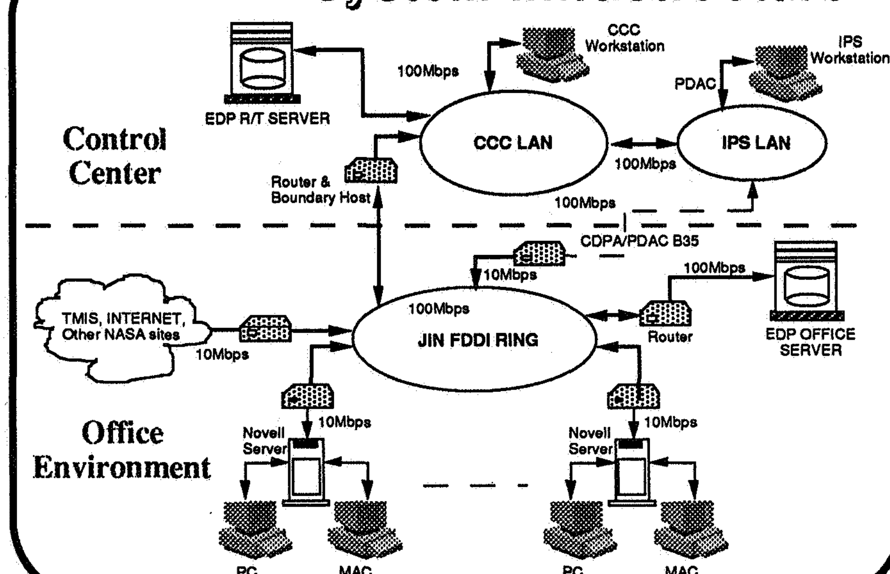
- Building preliminary, basic, and final editions of FDF
- Making change requests
- Crew procedures control board review
- Implementing approved changes
- Procedures Validation and Data Source Information

Software Technology Branch

ABL- 3/30/94 11

NASA Johnson Space Center
Information Systems Directorate

System Infrastructure



Software Technology Branch

ABL- 3/30/94 12

Conclusion

- Reduce costs (printing, distribution, etc.)
- Reduce storage requirements
- Increase efficiency in information access, retrieval and exchange
- Improve reliability
- Spinoffs to external organizations

omit

Session T1: MARKETING AND BARRIERS

Session Chair: Robert Brown

Invention-Driven Marketing

William E. Carlson
Intermetrics, Inc.
Cambridge, MA 02138

Suppose you have just created a revolutionary bicycle suspension which allows a bike to be ridden over rough terrain at 60 miles per hour. In addition, suppose that you are deeply concerned about the plight of hungry children. Which should you do:

a. Be sure all hungry children have bicycles?
b. Transfer the technology for your new suspension to bicycle manufacturers worldwide?
c. Start a company to supply premium sports bicycle based on your patented technology, and donate the profits to a charity which feeds hungry children?

Woven through this somewhat trivial example is the paradox of technology transfer - the supplier (owner) may want to transfer technology; but to succeed, he or she must reformulate the problem as a user need for which there is a new and better solution.

Successful technology transfer is little more than good marketing applied to an existing invention, process, or capability. You must identify who needs the technology, why they need it, why the new technology is better than alternatives, how much the customers are willing and able to pay for these benefits, and how to distribute products based on the technology to the target customers.

In market-driven development, the term "technology transfer" is rarely used. The developers focus on studying user needs and designing solutions. They may have technology needs, but they don't have technology in search of a use.

Technology transfer is really a process for exploiting serendipity. Any time you do an experiment, there is the potential for a surprising and valuable result which holds widespread applicability outside the domain that originally motivated the experiment. Any time you build a large system, there will be hundreds of small inventions; and each of them may have broad applicability.

Some managers dislike the unexpected and unplanned. Nonetheless, certain inventions in search of a market have proven too valuable to ignore: velcro, xerography, electronic mail, personal computers. Many of the inventions that have revolutionized our lives were created by technologists, driven to push the limits of the possible. Why does someone climb a mountain? . . . . Because it's there. Great inventions are often created for the same reason; but achieving global impact requires effective marketing of the new invention. Hence, I prefer the phase "Invention Driven Marketing" to "Technology Transfer".

This panel deals with technology for technology transfer. As is apparent from the above discussion, I view marketing of a new invention as a social process. The challenges are human and institutional, rather than technological. Nonetheless, innovative information technology can facilitate the human interactions and institutional changes which must occur. Several examples will be presented to stimulate questions and comments from the audience.

Keywords: Language, Design, Programmer Productivity Tools, Digital Signal Processing, System Simulation, Network Security, Guidance, Navigation, and Control, Information Systems Integration.

Invention-Driven Marketing:

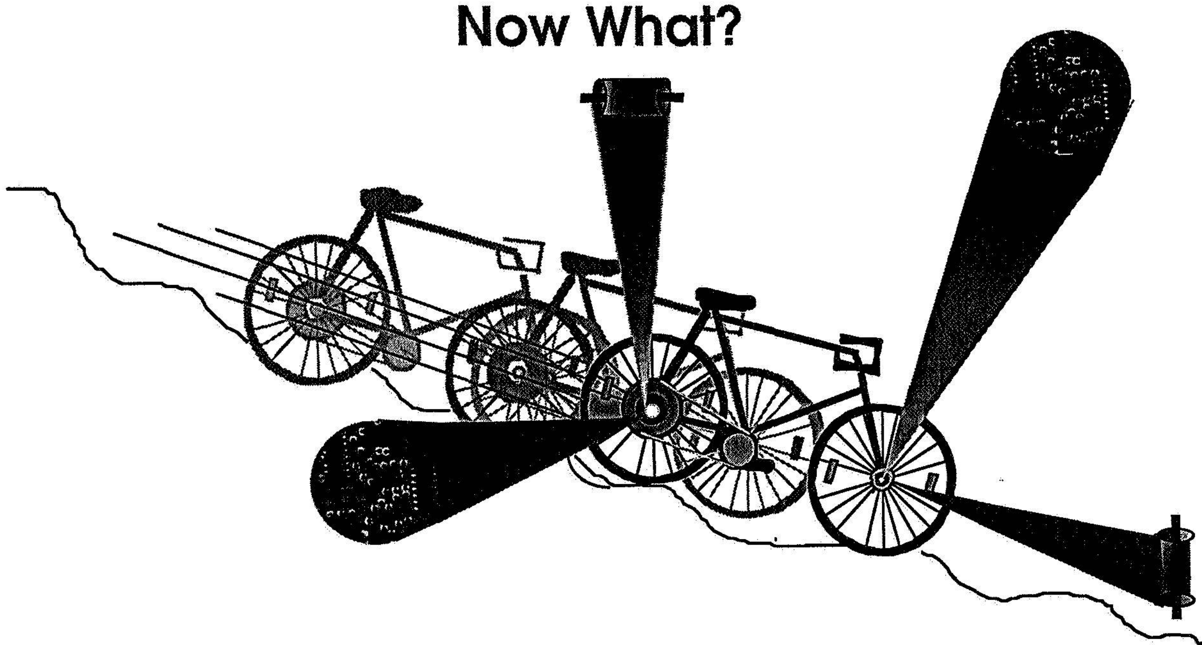
An Approach to Technology for Technology Transfer

William E. Carlson
Chief Technology Officer
Intermetrics, Inc.

February 1, 1994

Intermetrics

HERE'S THE TECHNOLOGY; Now What?



Intermetrics

SOME OPTIONS
If Driven to Help Hungry Children

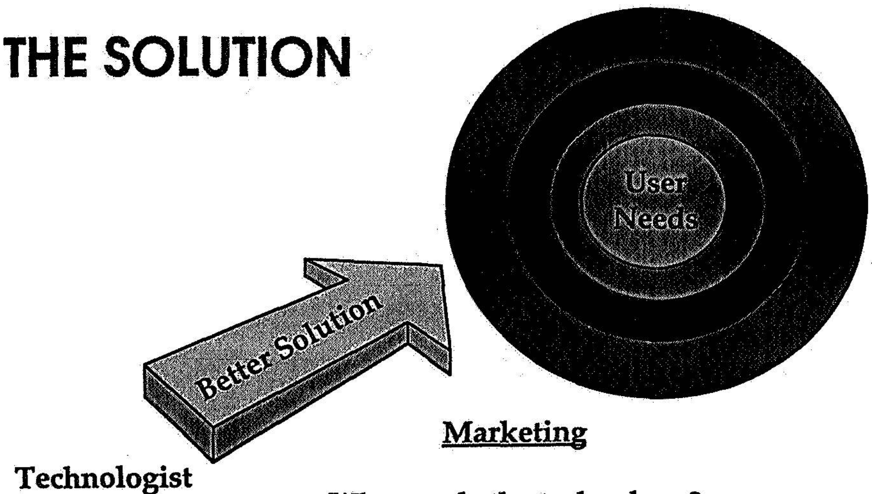
- ❑ Be Sure All Hungry Children Have Bicycles
- ❑ Transfer the Technology for the New Suspension to Bicycle Manufacturers Worldwide
- ❑ Start a Company
  - Premium sports bicycles
  - Patented technology
  - Donate profits to charity which feeds hungry children

Intermetrics

THE TECHNOLOGY TRANSFER GAP

intermetrics

THE SOLUTION



Marketing

Who needs the technology?
Why they need it?
Alternatives?
How much can they/will they pay for the benefits?
Cost effective distribution?

Intermetrics

WHY TECHNOLOGY TRANSFER

- Serendipity
  - Surprising and valuable results of an experiment
  - Applicability outside planned domain
  - Small inventions on path to big system
- Marketing to Exploit an Existing Invention Vs. Inventing to Satisfy an Identified Market Need

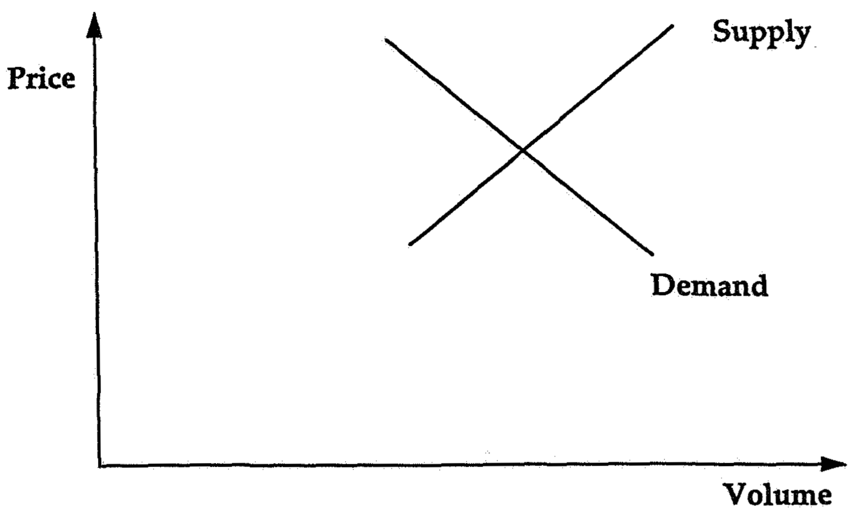
Intermetrics

ECONOMIES OF SCALE AND THE LEARNING CURVE

- ❑ Conventional Supply-Demand Curve
- ❑ Demand for High Tech Goods
- ❑ Supply of High Tech Goods

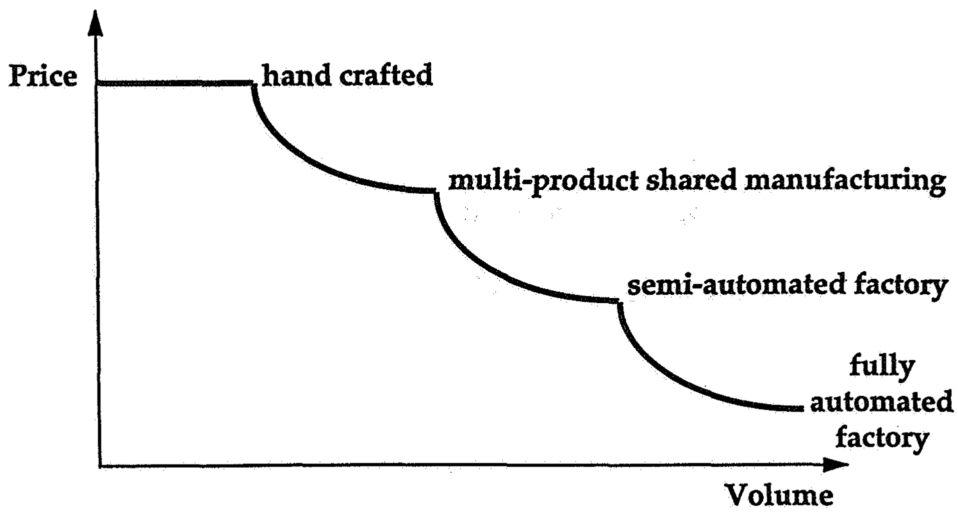
Intermetrics

CONVENTIONAL SUPPLY-DEMAND CURVE



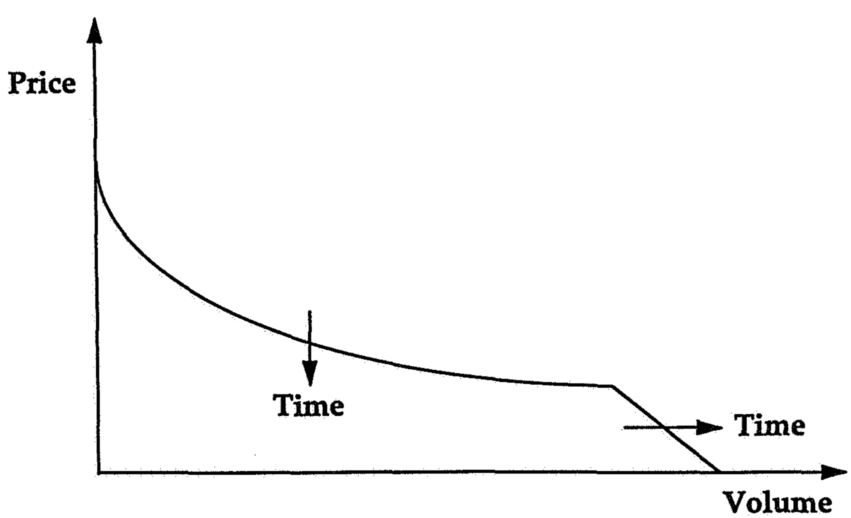
Intermetrics

SUPPLY OF HIGH TECH GOODS



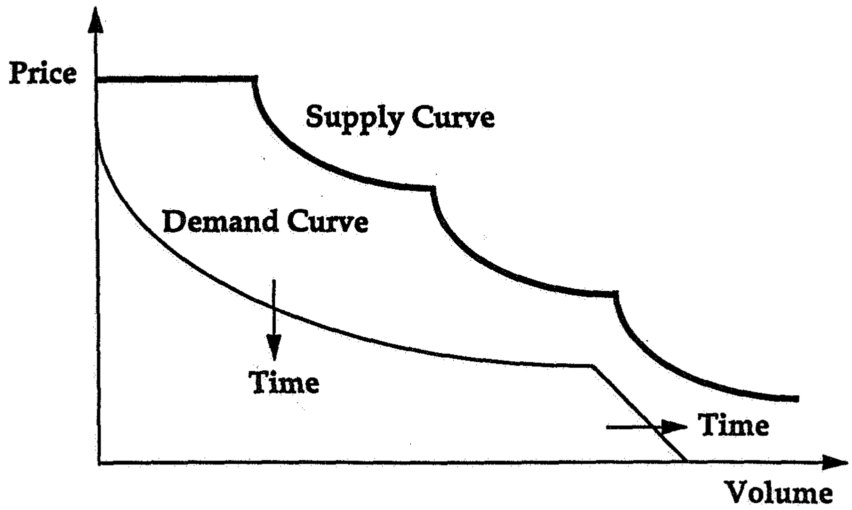
Intermetrics

DEMAND FOR HIGH TECH GOODS



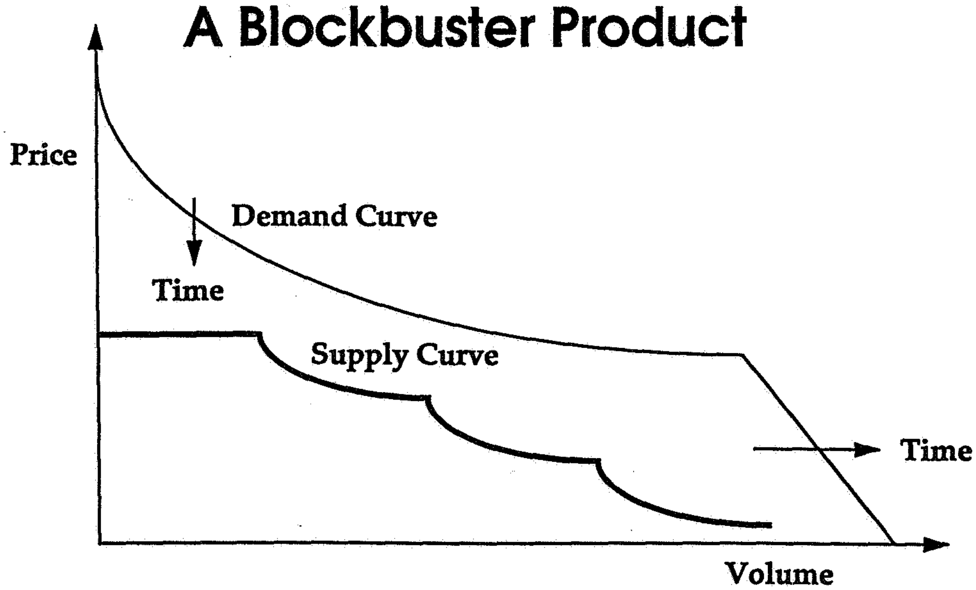
Intermetrics

MARKET FOR HIGH TECH GOODS
An Uncompetitive Product



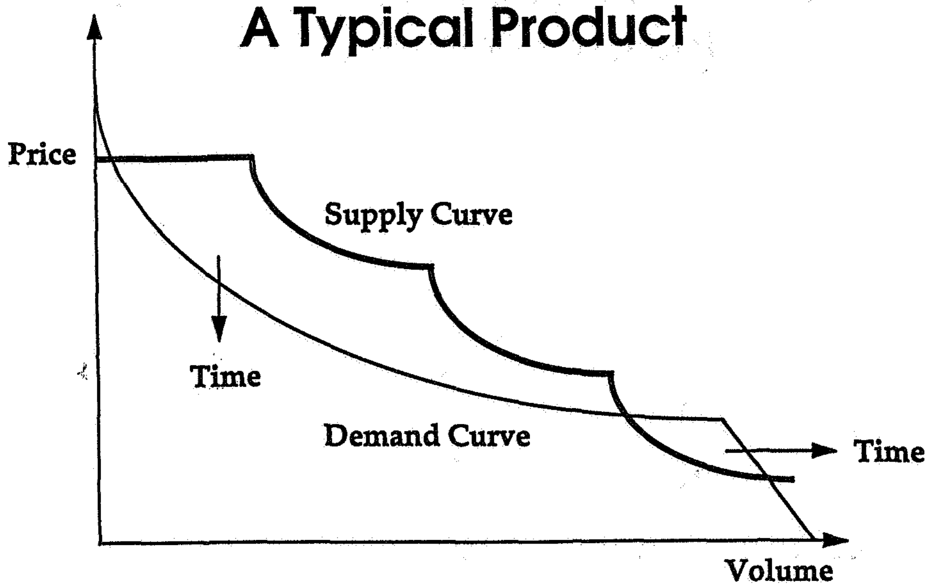
Intermetrics

MARKET FOR HIGH TECH GOODS
A Blockbuster Product



Intermetrics

MARKET FOR HIGH TECH GOODS
A Typical Product



Intermetrics

SOME TECHNOLOGIES FOR TECHNOLOGY TRANSFER

- ❑ Flexible Factories
  - Lower manufacturing costs for small volumes
- ❑ Information Retrieval
  - Lower cost of market research
- ❑ Groupware
  - Better market research
- ❑ Multi-Media Catalogs
  - Lower cost of educating the market

Intermetrics

SUMMARY

- ❑ Invention Driven Marketing
- ❑ Human and Institutional Barriers
- ❑ Marketing is a Social Process
- ❑ Benefits From Improved Information Interchange

Intermetrics

omit

Session T2: BUSINESS PROCESS AND TECHNOLOGY TRANSFER

Session Chair: Russ Cargo

WEAPONS TO WIDGETS: ORGANIC SYSTEMS AND PUBLIC POLICY FOR TECH TRANSFER

Russell A. Cargo, Ph.D.
Adjunct Assistant Professor
Department of Public and International Affairs
George Mason University
10836 Burr Oak Way
Burke, VA 22015-2419

ABSTRACT

Large cuts in defense spending cause serious repercussions throughout the American economy. One means to counter the negative effects of defense reductions is to redirect federal dollars to temporarily prop up defense industries and, over the longer-term, stimulate growth of new non-defense industries. The creation of non-defense products and industries by channeling ideas from public laboratories into the private sector manufacturing facilities, known as technology transfer, is being undertaken in a massive program that has high visibility, large amounts of money, and broad federal agency involvement. How effectively federal money can be directed toward stimulating the creation of non-defense products will define the strength of the economy, (i.e., tax base, employment level, trade balance, capital investments, etc.), over the next decade. Key functions of the tech transfer process are technology and market assessment, capital formation, manufacturing feasibility, sales and distribution, and business organization creation. Those, however, are not functions typically associated with the federal government. Is the government prepared to provide leadership in those areas? This paper suggests organic systems theory as a means to structure the public sector's actions to provide leadership in functional areas normally outside their scope of expertise. By applying new ideas in organization theory, can we design government action to efficiently and effectively transfer technologies?

INTRODUCTION

Technology transfer is expected to revitalize the economy. At the federal level we have dedicated sizable programs and 10-20% of the federal research and development (R&D) budget to that effort. The question I wish to raise is whether we have structured our organizations to accomplish that goal? Or, will we diminish our outcomes because we have not given close attention to the organizational structures to facilitate technology transfer? This paper will make observations about technology transfer processes and suggest that organic systems theory will help us organize ourselves to succeed.

PERSPECTIVES ON THE ISSUE OF SYSTEMS THEORY AND TECH TRANSFER

This segment of the Dual Use Technology Conference is designed to investigate, "Technologies for Technology Transfer," and this particular session is "Business Processes and Technology Transfer." My comments are based on the premise that our most valuable technology is our pool of human resources. How those resources, the *people* involved, are directed is unquestionably a matter of management, and management is changing. Effective management is no longer considered to be quantification-centered. Nor is it based on endless planning and forecasting. After decades of measuring *everything*, we have learned that our world is complex, forecasts are usually wrong, and detailed plans are nearly worthless nine months after they are completed. Far too many managers find it safer to demand "quantified certainty" than to make operational decisions. After all, if the numbers add up, the manager can't be faulted even if there is a bad outcome. And, if the bad outcome is proved to be unforeseen, then the manager is innocent -- or so goes traditional thinking. In fact, managers are guilty of wasting time, money, and perhaps worst of all, they work in organizations where it takes great courage to make decisions. The results of the traditional management approach are well known: disappointing productivity, difficulty in meeting foreign competition, and a shrinking manufacturing base.

However, progressive organizations have begun to retreat from the traditional MBA -- quantitative -- methods. What must evolve in order to fill that vacuum is a revised paradigm based on experience and research. Organic systems theory may be the answer. It builds on a substantial body of knowledge about systems, yet is updated to correct the faults of the traditional machine systems theory. We now understand that we are not simple cogs in a machine system, rather we are like complex cells or organs in a large dynamic organism, subject to all types of external and internal stimuli. Although there is a level of predictability to our functions, it is by no means certain.

THE ESSENCE OF THE ALTERNATIVE

Peter Senge has written an important book on this transformation in management thought, *The Fifth Discipline: The Art and Practice of the Learning Organization.* Senge proposes that organizations have the capability to unlock enormous potential in creativity if they can observe the whole system in which the organization operates. He writes:

> The essence of the discipline of systems thinking lies in a shift of mind: seeing interrelationships rather than linear cause-effect chains, and seeing processes of change rather than snapshots (Senge, p.73)

He goes on to explain,

> Systems thinking finds its greatest benefits in helping us distinguish high- from low-leverage changes in highly complex situations. In effect, the art of systems thinking lies in seeing trough complexity to the underlying structures generating change. ....What we most need are ways to know what is important and what is not important, what variables to focus on and which to pay less attention to -- and we need ways to do this which can help groups or teams develop shared understanding. (Senge, p. 128)

Shared understanding is precisely what is troubling about the sudden emphasis on technology transfer for many government agencies and workers. Do they have a shared understanding? Buckminister Fuller spoke of our ideas being two hundred years ahead of our social development. What he meant by that was that we have the capability to do more than we do. What impedes our progress is that humans avoid change and this is clearly built into our species. It is as much genetic as it is cultural. Unlike other animals, we spend years with our dependent offspring before they go out into the world on their own. That in itself is a major contributor to a pattern that slows social and cultural change in organizations.

When we institute a major change in the way a large organization like a federal agency operates we must account for these patterns of human behavior that affect the process of change. Organic systems theory can help us create mechanisms within the organization that ameliorate the negative aspects of resistance to change and speed the assimilation of new ideas. This technology for technology transfer is a methodology for thinking about the task in terms of the organization and its members. The important search is for appropriate structures and processes within the organization that will lead to fulfillment of our goals.

ISSUES INTERNAL TO THE TECHNOLOGY TRANSFER PROCESS

There are four major internal problem areas facing the technology transfer process that I believe can be successfully addressed through organizational and process change:

1. **PERSONNEL**: Federal lab staff and scientists, to date, are not thoroughly committed to tech transfer tasks of commercialization. There is not wide-spread experience with or a commitment to technology identification and assessment for commercialization, CRADA development, and solicitation of industry support for transfer of technologies.

2. **CONTACTS AND ACCESS**: Relatively few commercial enterprises are "plugged-in" to federal labs to the degree that they can efficiently gain access to available ideas, or work through the process of development.

3. **CAPITAL**: Capital is not available in the quantity necessary for most small firms to sustain the technology identification, assessment, and R&D process to bring an idea out of a federal lab.

4. **BUSINESS EXPERTISE IN LABS**: Federal agencies do not have sufficient expertise in manufacturing and market assessment, finance, or sales and distribution to assist commercial firms with those functions. Consequently, it is hard to advise or know what is commercializable.

Leadership in the technology transfer process whether it be agency-level or from Congress can define the organization and process of tech transfer to eliminate these areas of weakness.

SYSTEMS THEORY AND PUBLIC POLICY

What can be done? First, it is clear that public policies will determine the nature of the technology transfer initiative. The task we face is to enact policies that will maximize the effectiveness of the transfers. It is imperative that we look at the whole environment if we are to successfully employ systems theory to this issue. For example, the technology transfer program within the Department of Defense (DoD) and NASA is part of the larger policy outline framed by the broad national objectives of:

1) protecting our national research capabilities,

2) preserving the industrial capabilities of our defense contractors,

3) stimulating growth in the manufacturing sector to create jobs to off-set lost defense/federal jobs,

4) regaining national competitiveness in the global marketplace to help correct a serious trade imbalance, and

4) stimulating the economy to create additional revenues from tax receipts to reduce the federal budget deficit.

Second, those objectives cannot be met through a single program within one agency. On the other hand, they can be met by a systemic approach that includes coordinated programs within a variety of agencies and with modifications to existing legislation that runs counter to national objectives. For example, the Technology Reinvestment Project is an undertaking coming from the Defense Conversion, Reinvestment, and Transition Assistance Act of 1992. The solicitation for proposals came from the Advanced Research Projects Agency (ARPA) representing five different agencies (DoD, Commerce, DOE, NSF, and NASA). This is a good example of a coordinated effort among five agencies, leading directly to the transfer of technologies. However, are policy-makers developing other coordinated programs that address the other national objectives and tie together the related activities across the government? In the macro-view, have Congress and the Administration clarified their policies so that other critical agencies are operating in supportive roles? Is the Internal Revenue Service involved? Have tax laws been changed to place emphasis on capital investment in plant and equipment? Do capital gains taxes encourage investment? What will the Department of Labor's role be in direct support of the tech transfer initiative. Have the Treasury Department and the Federal Reserve made policies that will free capital through instruments of debt?

Similarly, in the micro-view, we can address the four internal problem areas mentioned above; 1) personnel, 2) contacts and access, 3) capital, and 4) business expertise in the labs. Beginning with personnel, have position descriptions and mission statements been rewritten to reflect the changed objectives of the organization? The incentives built into the Technology Transfer Act of 1986 provide for personal remuneration from lab activities but do not effect everyone involved in the process. Although such policies have generated change, it is a lengthy process and requires commitment from leadership and ongoing reinforcement in personnel policies; promotion decisions, award programs, and, if necessary, penalties for those who do not accept the changing mission of their agency.

Are contacts fostered by public programs? Is access truly encouraged? Not long ago I attended a multi-day event put on by one of the military services to help introduce business to the availability of technology and to encourage cooperation between the federally sponsored research and the manufacturing community. Immediately following a warm welcoming appeal to the audience by the General Officer in command of the laboratory he was whisked away to other "important" business. He was followed by a civilian presenter from the contracting office who made it very clear that no matter who you were or what technology you wanted, if you didn't have "his" forms filled out properly, there was no way you would be awarded a contract.

The message we in the audience understood was that in the next two days we would hear only more baloney, and only those insiders with experience in the system would be able to weed through the bureaucracy. Can such a break-down be fixed? Certainly, but it takes time and considerable effort. Obviously the contracting officer hadn't gotten the message and, one can only wonder whether the commander was sincere in his invitation to work with them, since he only attended the program long enough to hear himself.

Although capital availability is primarily driven by tax law and Federal Reserve Policy, in this context, through programs such as the Technology Reinvestment Project and the Small Business Innovation and Research (SBIR) Program, small amounts of capital are controlled at the agency-level. Within those Departments of government, do we have the best possible structure and system in place to direct tech transfer money? Are the evaluation and award processes most effective at transferring technologies or are SBIR funds, for instance, used as an "off budget" means of pursuing insiders' pet projects? Is it necessary for this process to be administered in house? Are the awards objectively made when lab personnel provide the evaluation and make the awards? Or, would the process be better if an outside, rotating, panel of scientists and business persons was formed to review and recommend awards for approval by a top agency administrator who would be publicly accountable? Is it possible to build a checks and balances protection into the system that awards millions of dollars rather than keeping the process in the hands of the same people who have interests in certain technologies, personal ties to "insiders" and little accountability for long-term outcomes once the grants are awarded?

CONCLUSION

Public policy that directs technology transfer efforts throughout the federal laboratory system could benefit from the application of systems theory. The weaknesses in the present structure described above are not serious flaws as organizations go. But without correction, the potential exists for technology transfer efforts to be undermined from within and to remain of limited benefit to potential inventors, manufacturers, and investors who are positioned to make a difference in the American economy in the coming decade. I propose that leaders in this Administration and in Congress evaluate the current technology transfer process and question its validity from a systems theory perspective. Further, those at every level in the system can do likewise. I am confident that improvements can be made and that the pay-off will be significant.

Research for this presentation uncovered titles that may provide a warning. (For example, "Let's Improve Technology Transfer," "Perry to Ruffle Feathers to Preserve Defense Industrial Base," "NASA's Controversial Quest for a Broader Mission," and, "Is Technology Transfer a Flop?") However, what we do with the enormous potential we have available from the outstanding work of the scientists in the federal laboratories will make a significant difference in the way we live and in the way our children fit into the shrinking world of the twenty-first century. We should take advantage of the technologies we have at our disposal -- *all* the "technologies," including organization theory. Let us use organic systems theory as we formulate our public policies to plan organizations that can unlock the creativity and potential of our human resources. We have much to gain.

REFERENCES

Benson, Tracy E., "The Learning Organization; Heading Toward Places Unimaginable," INDUSTRY WEEK, Vol. 242, n. 1, Jan 4, 1993, pl 35.

Chaffee, C. David, "NASA's Controversial Quest for a Broader Mission," PHOTONICS SPECTRA, Vol. 27, n. 1, Jan. 1993, p. 126.

Christie, John, "Let's improve Technology Transfer," MACHINE DESIGN, Nov. 7, 1991, vol. 63, n. 22, p. 191.

Heath, Les J., "Creating Learning Organizations: Growth Through Quality," An interactive Video Conference, The AED Foundation, Oak Brook, IL, February 26, 1993.

Kleiner, Kurt, "Courting High Technology: Programs, Projects Encourage Growth of Firms." BALTIMORE BUSINESS JOURNAL, Vol. 11, n. 18, Oct. 1, 1993, p. 19.

Lytle, David, "Is Technology a Flop?" PHOTONICS SPECTRA, Vol. 24, n. 2, Feb 1990, p. 60.

"Perry to Ruffle Feathers to Preserve Defense Industrial Base," DEFENSE DAILY, vol. 181, n. 6, Oct. 8, 1993, p. 46.

Senge, Peter M., THE FIFTH DISCIPLINE: THE ART AND PRACTICE OF THE LEARNING ORGANIZATION, Doubleday Currency, New York, 1990.

omit to p 679

Business Process Re-engineering -- A Precursor to Technology Transfer

Presented by

Guy Thomas
Insource Management Group, Inc.

IMG™
INSOURCE MANAGEMENT GROUP, INC.

"The Business Integration Company"

Technology Creates Wealth

- "Technology is the major determinant of wealth because it determines the nature and supply of natural resources"
- "The advance of technology is determined mainly by our *ability to process information*"
- "The backlog of unimplemented technological advances (that is, the technology gap) is the true predictor of economic growth for both the individual and society"

Source: Pilzer, Paul Z., Unlimited Wealth, 1990

IMG™
INSOURCE MANAGEMENT GROUP, INC.

Technology is a Change Driver

- The technology exists to address the issues and solve the problems facing companies today
- Today's competitors will use 90's technology to create competitive advantage
- No longer good enough to just have a high-quality product, or the lowest price, or the best customer service
- If you don't take the steps to offer it all in terms of quality, cost and service, someone else will

IMG™
INSOURCE MANAGEMENT GROUP, INC.

"The Business Integration Company"

Necessary, but not sufficient...

- Total Quality Management
- Self-directed Teams/Case Teams
- Benchmarking
- Updating information technology
- Rightsizing
- Downsizing
- Outsourcing
- Downcosting
- Re-engineering

IMG™
INSOURCE MANAGEMENT GROUP, INC.

Dangers of isolated technical solutions

- Can alienate people
- Can create chaos in the organization
- Do not recognize essential non-technical solutions
- Can automate mistakes

IMG™
INSOURCE MANAGEMENT GROUP, INC.

"The Business Integration Company"

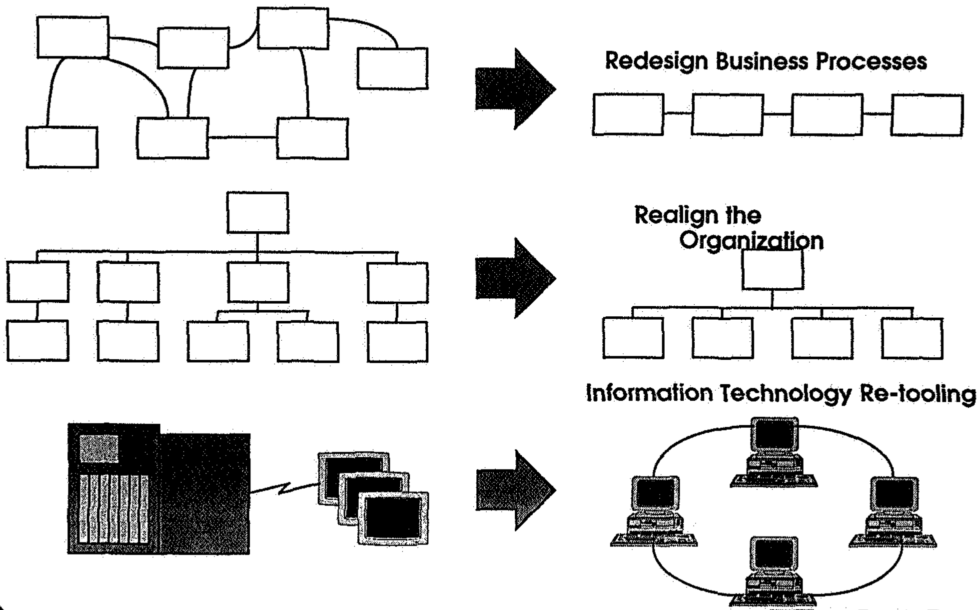
What is required?

- Commitment to change
- Common vision
- Employee involvement and commitment
- Management commitment and follow-through
- Balanced approach dealing with People, Process and Technology
- Readiness for change
- Predictability of the outcome

IMG™
INSOURCE MANAGEMENT GROUP, INC.

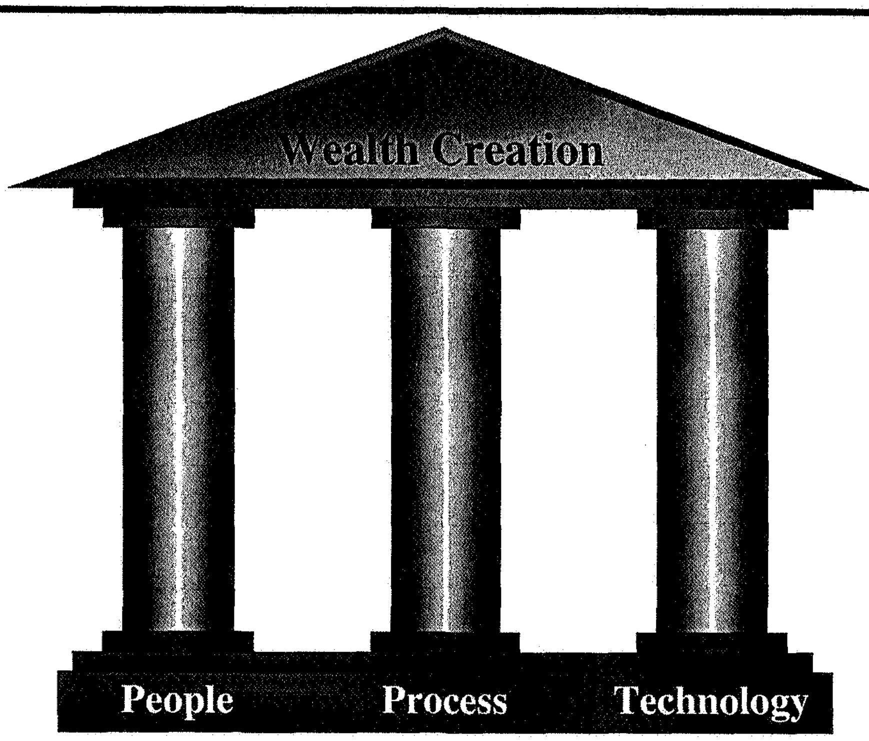
"The Business Integration Company"

Business Driven Solutions



IMG™
INSOURCE MANAGEMENT GROUP, INC.

"The Business Integration Company"



IMG™
INSOURCE MANAGEMENT GROUP, INC.

534-85
44868

FuzzyCLIPS From Research to Product

Dan Bochsler
Edgar Dohmann

Togai InfraLogic, Inc.
Technology Systems Division
17000 El Camino Real #208
Houston, Texas 77058

***Abstract* - This paper describes the commercial productization of FuzzyCLIPS which was developed under a NASA Phase II SBIR contract. The intent of this paper is to provide a general roadmap of the processes that are required to make a viable, marketable product once its concept and development are complete.**

I. INTRODUCTION

FuzzyCLIPS was developed under a NASA Phase II SBIR contract. FuzzyCLIPS combines fuzzy reasoning capability with conventional rule based technology. It retains the flexibility and portability of CLIPS, targets development of both stand-alone and embedded systems, and extends rule syntax and user definitions of membership function types.

The rule based portion of FuzzyCLIPS was derived from NASA's version 5.1 of CLIPS (C Language Integrated Production System) which was developed by NASA's Johnson Space Center Software Technology Branch. The fuzzy portion of FuzzyCLIPS is based upon Togai InfraLogic's fuzzy logic products.

A number of demonstration examples were developed for FuzzyCLIPS during the Phase II SBIR implementation. Demo and alpha versions of FuzzyCLIPS were shown at several conferences during the Phase II SBIR implementation. Based upon reactions to these demonstrations and initial product inquiries, FuzzyCLIPS appears to have a bright future as a tool for developing intelligent decision making and control systems.

The technical development of FuzzyCLIPS is only part of the overall story of creating a commercial software product. The final productization process which is described in this paper required an additional 4 months of work following the final SBIR delivery to NASA.

II PRODUCT DEVELOPMENT

Market planning is a crucial part of the product development cycle. Since a primary objective of the Phase II SBIR program was to eventually commercialize the FuzzyCLIPS product, several key decisions related to its marketing were made early in the design phase of the project. These decisions then became important objectives of the overall development process.

CLIPS is available for IBM PC, Macintosh, and UNIX platforms. We decided that FuzzyCLIPS would also support these same platforms so that all current users of CLIPS would be able to take advantage of the fuzzy logic enhancements.

Once a decision was made on the platforms to be supported, the user interface was considered. We decided to retain the limited graphics approach that is used in CLIPS 5.1. This decision made portability of the final product across the 3 platforms fairly straightforward. A more extensive graphic interface would have been nice but would have required significantly more development effort, made the platform porting task more difficult, and increased our maintenance and support costs.

The FuzzyCLIPS extensions were developed in C to maintain maximum compatibility with CLIPS 5.1 and to simplify porting across the 3 supported platforms.

Testing new products is a major concern for developers. We decided to implement a Verification and Validation (V&V) tool utility to serve as an application example of FuzzyCLIPS and to provide a tool for evaluating and testing our own extensions. This V&V tool is included with the final FuzzyCLIPS product.

To further assist users with testing new applications, we developed several example applications which can be used as templates or benchmarks for user's own applications.

V PRICING

Pricing is the final factor that must be considered in productization. Competitive product pricing, general price range of other products the customers are used to purchasing, business revenue goals, and production costs are some of the elements that may influence a product's sale price.

Discounts must also be taken into account when establishing a product's price to ensure that there is adequate revenue margin. Discounts may be offered for quantity purchases, educational institutions, special time offers, rebate coupons, etc. Discounts may also be required when selling through resellers or distributors.

Terms of sale must also be determined. Sales can be offered via credit card, check, cash only, purchase orders, or credit lines. The specific methods that will be accepted must be determined before the sales effort begins.

There is no direct competitive product for FuzzyCLIPS at this time so competitive product pricing information was not available to us. We used the COSMIC pricing policy for CLIPS as a general guideline to determine a fair price range for products that our target customers are accustomed to paying.

To address the needs of a variety of potential customers, we decided to offer FuzzyCLIPS in three product versions: User, Developer, and Professional at prices of $199, $499, and $899 respectively.

FuzzyCLIPS Professional includes all CLIPS 5.1 source code, source code for the TIL fuzzy logic extensions, and the interactive executable files. FuzzyCLIPS Developer includes all CLIPS 5.1 source code, a run-time library of the TIL fuzzy logic extensions, and the interactive extensions. FuzzyCLIPS User contains the interactive executables for CLIPS 5.1 and the TIL fuzzy logic extensions. All versions include the TIL FuzzyCLIPS User's Manual, verification and validation files, and sample FuzzyCLIPS examples. An economical migration path is offered for users who start with the User of Developer version and wish to upgrade as their needs evolve.

At the present time we are not offering any discounts for FuzzyCLIPS. We are handling all sales direct so reseller discounts are not an issue. The User and Developer versions offer two low-cost options to organizations that are price sensitive such as educational institutions. These two versions are priced to be attractive to such organizations and the margins do not lend themselves to any further discounts.

We are presently accepting prepaid orders for FuzzyCLIPS via check, money order, or credit cards. We will accept purchase orders from major companies or educational institutions. Purchase orders may be accepted from other companies if accompanied by acceptable credit information.

Sales taxes and shipping costs are other factors that must be taken into consideration. We are currently applying a fixed shipping and handling fee to all orders and shipping orders via regular UPS ground service. We will ship via overnight Federal Express if the customer provides an account number for billing the overnight fee. All FuzzyCLIPS orders are processed at our Irvine, California office so all California residents must add California sales tax to their order or provide us with their tax-exempt ID number.

VI CONCLUSION

As discussed in this paper, product development is only 1 of 4 factors in the commercialization of a new product. Promotion, Distribution, and Pricing must also be considered.

Once a product is actually being marketed and sold, there are still other considerations such as: maintenance and support, upgrades, and enhancements.

We have established a 1-800 number for customers to call for questions or problems with FuzzyCLIPS or any other TIL product. Each software manual includes a Software Trouble Report that customers may mail or fax to us to describe any problems or offer suggestions.

We will maintain a database of all FuzzyCLIPS customers who return their postage-paid product registration form. When updates to FuzzyCLIPS are announced, we will mail all registered users information and instructions on how they may obtain an upgrade. We also offer an economical migration path for users who start with the User or Developer version and wish to upgrade to another version as their needs evolve.

A number of enhancements and add-on products are currently under development for extending the capabilities of FuzzyCLIPS. TIL also offers consulting and implementation services to customers who need assistance in developing their applications.

REFERENCES

Carreno, L., and Jani, Y.: "*FuzzyCLIPS User's Manual*", Version 1.5, Togai InfraLogic, Inc., 1993.

omit to
p. 756

Session T3: NEW WAYS OF DOING BUSINESS

Session Chair: John P. Van Blois

IMPROVED WAYS OF DOING BUSINESS BETWEEN INDUSTRY AND GOVERNMENT

EVOLVED FROM STRATEGIC AVIONICS TECHNOLOGY WORKING GROUP (SATWG) ACTIVITIES

Dr. Kenneth J. Cox
February 1994

INITIAL SATWG ORGANIZATION

- NASA initiated a Space Avionics Technology Symposium in 1989 to promote improved cooperation between technologists, engineers, operators and program managers, and to focus on customer driven products

- SATWG established in 1990, and presently involves
  - All NASA Centers
  - Major space integration contractors
  - Space avionics contractors

- Activity based upon mutual respect, trust, and open dialogue

- Scope includes development and maintenance of space flight elements and ground/flight infrastructure

SATWG MISSION FOCUS

- Provide a forum for development of avionics strategies and associated roadmaps
- Facilitate teaming and partnership relationships between NASA, government agencies, industry, academia, and professional societies
- Connect and network national pockets of expertise and knowledge
- Search for common ground in aerospace exploration and commercialization
- Construct creative paradigms for quality, environmental, and economic improvements in governance
- Enhance the ability of the aerospace community to adapt and evolve with cultural change

DANCE OF PERMANENT WHITE WATER MANAGEMENT

THE CHALLENGE

- Manage within a continuous, turbulent, change upon change environment with high stakes

THE FEELINGS

- Similar to being on a raft in white water with only a pole
  - Limited control over where going
  - Raft is basically unmanageable without rudder
  - Pole is reactive only/fend against the rocks
  - If successful, must do the same thing tomorrow
  - Do not know what will be ahead

THE MYTH

- Someone in upper management is in calm water and knows the big picture

THE RESPONSE

- Organizations must plan for and adapt to significant levels of continuous change to remain viable and to properly evolve

THE FOCUS

- Encourage individual awareness and empowerment
- Develop continuous individual adaptation and growth
- Honor and utilize differences and diversity in people
- Develop a significant commitment to service
- Encourage development of creative intuitive skills

- Develop group behavior to encourage system esteem and sharing
- Encourage continuous improvement of relationships, processes, and products
- Develop practical open dialogue skills

- Facilitate effective communicating and partnering within, between, and among organizations
- Foster a commitment to the long view
- Develop transformational skills for building global communities

TOP INDUSTRY/GOVERNMENT CHALLENGES

- An ordered strategic improvement strategy has evolved since 1992
  - Understand current processes and eliminate inefficiencies
  - Improve processes
  - Apply new technology
- The goals and issues have been sorted into
  - Multiple use technology
  - Programmatic drivers
  - Systems challenges and community building

MULTIPLE USE TECHNOLOGY

- Avionics Architectures
  - Open Architectures
  - Living Standards
  - Fault Tolerant Systems
  - Modular/Portable Software
  - Design for Operation
- Guidance, Navigation, & Control
  - Rapid Systems Development
  - Rendezvous & Capture
  - Smart Sensors/Effectors
  - Advanced Algorithms
  - Autonomous Systems
- Vehicle Health Management
  - Autonomous Check-out
  - Fault Prediction Built-in Test
  - System Model Generation
  - Health Monitoring Systems
  - Data Base Management
- Communications & Tracking
  - Hi-Rate/Density Transmission
  - Expert Space Systems
  - Advanced Antenna System
  - Radio Frequency/Optic Sensors
- Power Management & Control
  - Advanced Energy Storage
  - High Frequency Distribution
  - Electric Control/Actuation
  - Fault Tolerant (Modular) Systems
  - Power Electronics

PROGRAMMATIC DRIVERS

A mature base of programmatic requirements for technology has been developed by structured quality function deployment methods, and involve results from commercial ELV, NLS, ACRV, and SATWG

- Operability - All aspects of the system operates in a simple and effective fashion (both ground and flight)
- Affordability - Improve program life-cycle costs
- Dependability - System always operates as required, in a predictable fashion
- Mission Suitability - Total system infrastructure meets mission requirements

SYSTEMS CHALLENGES AND COMMUNITY BUILDING

- Cultural Change
- Dual Use Technology Initiatives
- Concurrent Engineering/Integrated Product Design Processes
- Incentives
- Innovation
- Connectivity/Networking/Sharing
- Teaming
- Creative knowledge generation and capture
- Risk/Cost assessment
- Education initiatives

A PARADIGM FOR IMPROVED WAYS OF DOING BUSINESS

- Horizontal domain of product
  - Parts model
  - Content based
  - Tends toward specialization
  - Discipline oriented
- Vertical domain of process
  - Wholeness model
  - Systemic and cultural based
  - Tends toward generalization
  - Strategy oriented
- Out of plane domain of community
  - Global Model
  - Society based
  - Tends toward interconnectiveness
  - Sovereign state oriented - industry, government, academia, professional societies, and consortiums
- Domain of environment changing with time
  - Universe model
  - Consciousness based
  - Reflects transformational dynamics and processes

A PARADIGM FOR IMPROVED WAY OF DOING BUSINESS (CONT.)

- A dimensional domain representation contains interesting properties
  - Concept of a sphere of information, communication, and intelligent systems changing with time
  - Basic properties include quantum physics, chaos theory, crises management, and complexity theory - concerned with random events, changing energy levels, probabilities, discontinuities, resonances, and non-local connections
- Industry has recommended that NASA decrease emphasis on creation of avionics technology (domain of product) and increase emphasis on integration of avionics technology (domain of process)
- As a consensus, both government and industry recommend an increased partnership emphasis on domain of community in order to compete and cooperate at the global level
- A holistic definition of technology has been suggested
  - Hard Technology - Associated with things and objects
  - Soft Technology - Associated with people, culture, and community

IMPROVED WAYS OF DOING BUSINESS

- Transition from autonomous, discipline-oriented processes to product-oriented, multi-discipline teams with virtual systems capabilities
  - Orient toward information, intelligent systems, and operations
  - Motivation, mutual respect, and trust required for teaming
  - Need small groups of generalists to integrate specialists' knowledge across organizational and geographical boundaries

- Proposed life cycle teams

• Rapid Development	(Requirements)
• Integrated Product	(DDT&E)
• Integrated Operations	(Operations)
• Continuous Improvement	(Evolution)
• Strategic Management	(Phaseout)

- Establish limits on infrastructure

IMPROVED WAYS OF DOING BUSINESS IN SPACE

- Innovative aerospace industry/government relationships are needed to effectively compete and cooperate at the international level

- Traditionally, industry/government roles and responsibilities have moved in a dynamic pendulum between excessive bureaucracy and excessive monopoly

- The contemporary culture seeks an improved dynamic balance away from excessive bureaucracy
  - Reduce regulatory burdens
  - Change government oversight
  - Lower reporting demands
  - Eliminate unique accounting practices

IMPROVED WAYS OF DOING BUSINESS IN SPACE (CONTINUED)

- A new model is proposed involving a third sovereign party acting as a matchmaker
  - This entity is referred to as a project broker or trading company
  - Its functions are to propose and establish multiple company interactions and deals within the industry environment with the government acting as a facilitator, catalyst and partner
  - Trading company characteristics include capability to utilize distributed expertise to respond rapidly to market driven changes, and satisfy streamlined government laws, regulations, and policies
  - Strong networking and communication capabilities are required
  - A characteristic trading company generally has low investment in aerospace infrastructure, including large workforce groups, laboratories, equipment, and launch/mission control sites
- Third party trading company concepts need to be examined for
  - Market driven commercial space applications
  - NASA based exploration and science missions
  - Civil space technology, manufacturing, and infrastructure development to support defense readiness

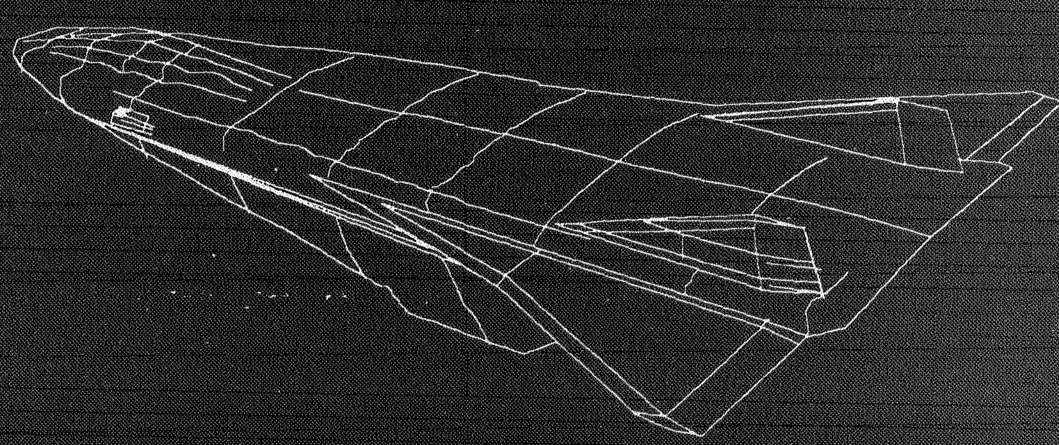
FUTURE CHALLENGES

- Initiatives should be developed for a quality, environmental, and economic revolution in governance
- Specific initiatives to transition government operations from a rule/regulation based system to a market driven system are desired where the government role is a facilitator/catalyst and a partner with industry
- Motivate individuals to work for the common good and to search for common ground in space exploration and commercialization
- Future technological, political, ecological, and economical paradigms will be driven by Earth resource limitations and by active exploration of space
- Place greater emphasis on soft technology associated with people and community

NASP National Aero-Space Plane

Value Of NASP To This Country

The Return On Investment From A National Perspective



12 November 1992

National Aero-Space Plane
Joint Program Office
ASC/NAR
Bldg. 52, M Street
Wright Patterson AFB, OH 45433-6503

Value of NASP to This Country
The Return on Investment from a National Perspective

Table of Contents

Value of NASP to This Country
The Return on Investment From a National Perspective

The magnitude and diversity of technological challenges associated with the National Aero-Space Plane (NASP) program are resulting in advancements across a broad range of technologies. This work is being accomplished by five major aerospace companies, over 500 subcontractors (including universities) and 15 major U.S. government research laboratories. This range of affected technologies and the scope of interaction with national institutions and firms are significantly enhancing the basis for U.S. technological preeminence in the international aerospace marketplace.

The more obvious, longer-term benefits consist of operational vehicles for flexible, efficient access to space (i.e., Earth orbit) and options for military and civil hypersonic aircraft. Both long and shorter-term benefits will accrue from the applications of NASP developed technologies. In addition to their impact on the aerospace industry, short-term benefits are beginning to be identified in a variety of industries, including those for chemicals, energy, medicine, and automobiles.

Historical Perspective

The history of technology development and application can be considered in terms of a tangled set of technological paths: surges of progress result when several crucial technologies converge in the presence of economic or military opportunity. The dynamic development of western civilization in the past 600 years has been highly influenced by such conjunctions of technologies and needs.

The emergence of NASP technological applications in the last few years has been affected by several factors. First, some information has been moved only recently beyond the barriers of security classification or proprietary constraints. Second, and more conventionally, any commercially astute firm waits to announce a promising application until after it achieves a secure market position. The integral effect of these factors ensures that both the range and significance of NASP benefits will be much better understood after a few more years.

Current expectations for applications can be based on several factors. The scope of technological advances to date gives rise to strong expectations. Furthermore, historical studies indicate that striving to meet strong challenges, such as those of NASP, lead to valuable applications. The ability to predict accurately the resulting course of technological history; however, remains as difficult as ever.

Economic Studies

A series of studies of the economic impact of NASP technologies have predicted an increase in the competitiveness of U.S. industries from productivity gains brought about by NASP technologies. One study projected that the Gross National Product may expand by a total approaching $26 Billion over the next 10 to 15 years, yielding an average annual increase in employment of nearly 40,000. While these predictive macroeconomic studies should be considered with the full list of assumptions and caveats, they do serve to highlight the potential for beneficial trends, particularly in key segments of the economy. For example, these economic studies show that NASP research should significantly benefit several critical U.S. industries: (1) aerospace, where the U.S. enjoys a favorable trade balance, (2) motor vehicles, where our balance of trade has continued to erode, and (3) emerging industries, such as ceramics. Moreover, the advanced technologies of NASP will help shift the country's employment distribution towards higher productivity industries. This potential is further articulated in a list of over 4000 intersections of NASP technology with standard industrial categories established by the Department of Commerce.

Near Term Benefits

The most tangible evidence of the national benefits being realized from NASP are the technology spin-offs currently being identified. Many of the specific examples will take time to implement because of further technology development, re-tooling, capital investment, and the regulatory approval processes. In spite of this, some very promising candidates for applications have been identified in a variety of industries. The majority of the near-term benefits involve materials science and advanced computational techniques.

<u>Materials Science:</u>

A. **Titanium and Titanium Aluminide (TiAl) Metal Matrix Composites (MMC)** Silicon carbide fibers immersed in a matrix of titanium or titanium aluminide alloys.

- Technology developed in part by NASP is being utilized for MMC compressors for demonstration gas-turbine engines. This yields higher performance than with conventional titanium components (such as Ti-6-4).

  -- Benefits: Include lighter weight (40%), higher temperature tolerance (1500°F), higher speed (15%), less fuel consumption (1.5%) and significant dollar savings IR&D for materials. Both mechanical properties and processing information generated by NASP have provided the benefits. Transition to Garrett, General Electric, and Pratt & Whitney engine companies is already in process for DOD and commercial applications.

B. **(TIMETAL™ 21S) Titanium.** Developed as matrix material for NASP MMCs, this material exhibits 100 times the corrosion resistance of "standard" aircraft titanium. It has the same high strength and temperature capability of current alloys.

- Benefits:

  -- Strip producible

    --- Economically rolled to foil
    --- Mitigates high ingot cost

  -- Very high strength-to-weight ratio

  -- Very high corrosion resistance

  -- Better cold and warm formability than conventional Ti Alloys

- TIMETAL™ 21S use is anticipated in piping for sour oil and gas wells. Sour wells refer to sulfuric underground conditions (i.e. those producing sulfuric acid) where TIMETAL™ 21S piping can withstand the corrosive gases. Test pipes have been produced.

  -- Benefits: At approximately half the weight, a deeper well can be drilled with a subsequent increase in the number of oil and natural gas sources that can be tapped. Well depths of 30,000 ft are now achievable (versus 20,000 ft for steel-clad pipe). Also, the lighter weight allows larger diameter pipe with up to a 50% increase in flow rate. Up to 150,000 lb of TIMETAL™ 21S would be required per well. Significant long-term increases in energy production is foreseen.

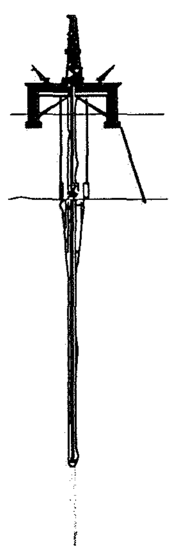


Figure 1. Depth Extension Using TIMETAL™ 21S

- TIMETAL™ 21S has great potential for use in offshore drilling structures. Current oil platforms require biannual replacement of 2 million pounds of nickel-based superalloy bolts because of corrosion damage. A significant expense is the replacement labor cost. An international industry consortium is currently evaluating TIMETAL™ 21S (operational platforms) bolts that will be permanent. Substantial savings are expected:

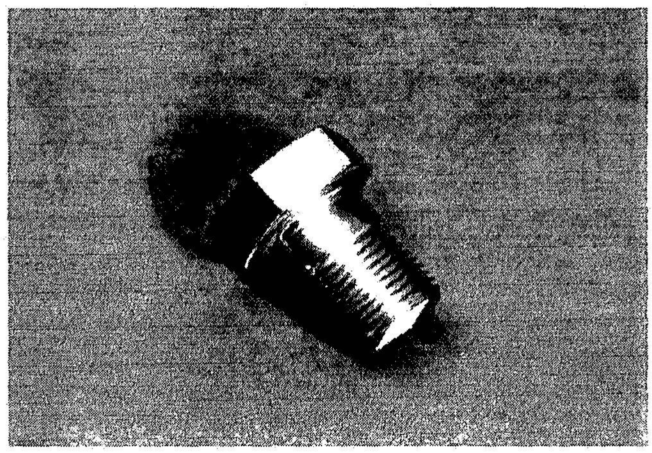


Figure 2. Anti-Corrosive Bolt For Chemical Plant

-- Long life through corrosion resistance

-- Substantially reduced labor cost over platform lifetime

-- Safety

-- 100 million pound market potential for just the North Sea Platforms

- A Navy program is evaluating TIMETAL™ 21S bolts for extensive surface/subsurface application.

- TIMETAL™ 21S used in a jet-engine tail cone where atmosphere is hot and corrosive. The part is in production for F404 engine.

  -- Benefits: Reduced weight component, higher aircraft fuel efficiency and reduced maintenance costs.

- TIMETAL™ 21S is baselined into nacelles for B-777 replacing superalloys, to eliminate corrosion hazards from leaks of hot "Skydrol". Significant weight savings:

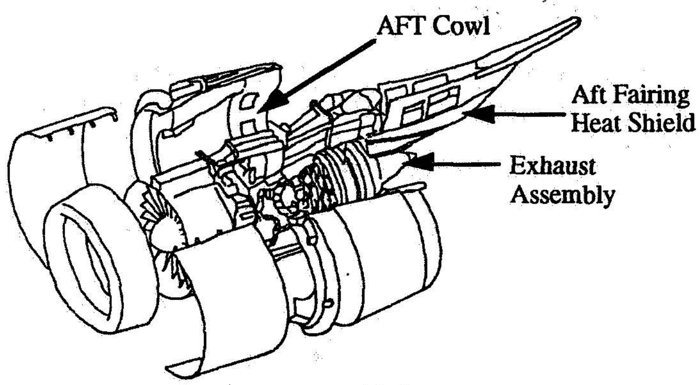


Figure 3. TIMETAL™ 21S Applications On Boeing 777

--	Plug & Nozzle	550lb/nacelle
--	Aft cowl	200lb/nacelle
--	<u>Heat Shield</u>	<u>100lb/engine</u>
--	Total	850lb/engine

-- Life Cycle Cost savings have been estimated in excess of $3-5 million per aircraft. Fleet savings approaching one half billion dollars are anticipated.

- TIMETAL™ 21S used in manufacturing.

  -- Benefits: Corrosion resistance to acidic chemicals results in longer life chemical tanks and lower process costs due to tank longevity. Also, this increases safety over previous tanks. Tests on a chemical tank liner article are in progress.

  -- Other parts produced include anti corrosive bolts, tubing for nuclear plants, heat-exchanger parts for oil-refinery plants and high-speed conveyor bearings.

- Potential exists to use TIMETAL™ 21S as human implant material for prosthetic joints, pacemaker components, orthopedic wire, and several other applications.

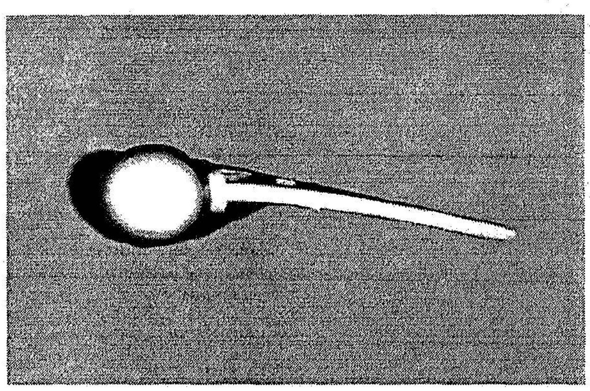


Figure 4. TIMETAL™ Titanium Hip Implant

-- Benefits: Potentially 30% to 50% fewer surgical procedures. Elasticity closer to human bones results in a more natural joint for such applications.

- The higher oxidation resistance and low density appear excellent for armor plating for military ground vehicles.

- The material and market did not exist prior to NASP. The point of contact (POC) is Dr Paul Bania, V.P. Research, Timet at (702) 564-2544. Marketing estimates (Timet) include:

-- Government:

Aircraft	200,000-lb/yr
Spacecraft	200,000-lb/yr
Ground Vehicles	500,000-lb/yr

-- Commercial:

Aircraft	400,000-lb/yr
Chemical Processing	300,000-lb/yr
Oil Drilling	3-5,000,000-lb/yr
Biomedical	100,000-lb/yr
Total	5-7,000,000-lb/yr

- Titanium is a highly competitive international industry. The major players are the U.S., Japan, and now Russia (CIS), which has the production capability of the U.S. and Japan combined. TIMETAL^{tm} 21S, patented by Timet, will maintain U.S. competitiveness in this strategic market. Timet expects dramatic growth in the titanium industry primarily because of TIMETAL^{tm} 21S.

C. **Gamma Titanium Aluminide Development.** NASP research enabled the scale-up and productivity of Martin Marietta XD^{tm} gamma titanium aluminide. (Several major programs have enabled the general development of XD^{tm}, an advanced process for synthesis of composites with in situ precipitation of reinforcements. The XD^{tm} process was applied to the production of gamma titanium for NASP.) Lower cost and much higher strength uniformity versus other composites is available using this process. Only button size elements existed before NASP. Because of NASP, 10,000 lb ingots are now made. NASP support also helped to enable work to apply the process to other titanium aluminides.

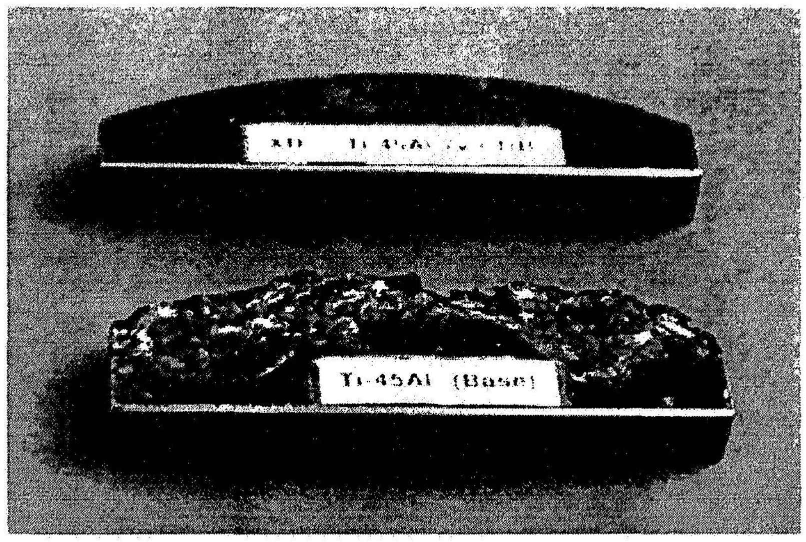


Figure 5. Composition Comparison Between XD™ And Base Metal

- Gamma titanium small-engine impeller test articles have been produced for gas-turbine engine and automotive turbochargers. This is currently under development for single piece compressor stage in a cruise-missile engine.

  -- Benefits: Light weight, high-strength, and high-temperature capability allows for faster spin-up to operating speed for launch at lower altitude. Not only is survivability increased, but the material also allows for low-cost manufacturing.

- A compressor blade was investment cast and machined.

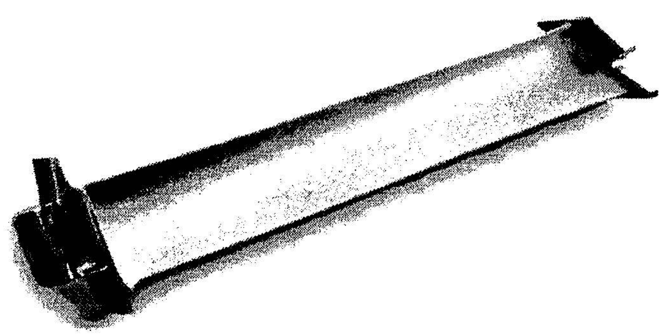


Figure 6. Low Pressure Turbine Blade For GE Engine (XD™ Titanium)

- A gun blast diffuser for the F/A-18 aircraft has been built as a replacement component.

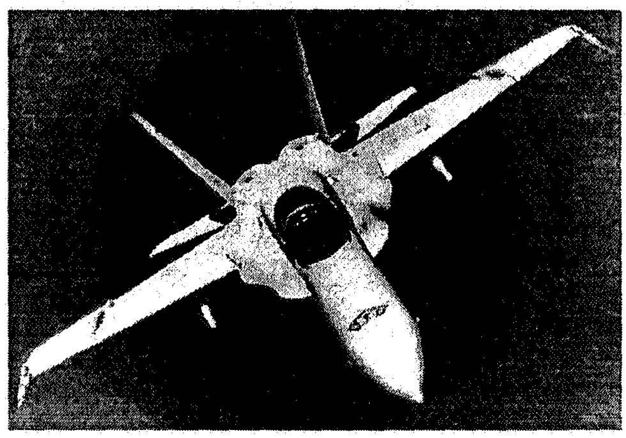


Figure 7. F-18 Gun Blast Diffuser Cast XD™ Component

-- Increased erosion resistance.

-- 40 lb weight savings attributed to this component.

- Cast XD^tm titanium aluminide airfoil (for jet-engine applications) have been produced under DARPA contract.

  -- Benefit: less than half the weight versus a conventional superalloy.

- TiAl missile fins and other aerospace structures. Prototypes have been produced for Navy weapons. (AQM-127A)

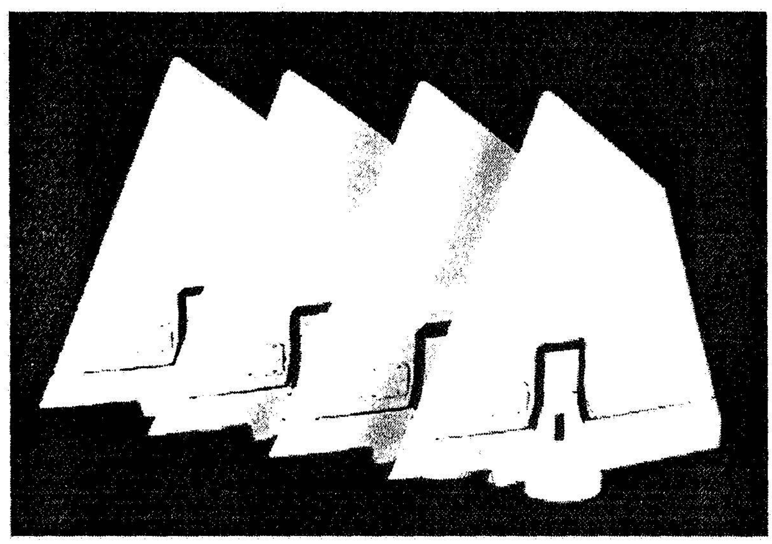


Figure 8. Cast XD™ Missile Fins

- -- Benefits: Lower cost, easily producible, and lighter weight weapon components with a 50% increase in tensile strength. Potential for a weight saving of up to 150 pounds in supersonic Navy missile resulting in a 40% range (95% area) increase.
- -- Coated XD™ fins have been built for thrust vectoring missiles. Dramatic increases in maneuverability become available.

- Test articles for TiAl auto exhaust valves are undergoing tests by major automotive manufacturers.

  - -- Benefits: TiAl valves are 50% lighter than current steel valves; they allow for a 500 RPM increase in engine speed resulting in cost effective ways of increasing engine performance. Lighter engines with increased efficiency are anticipated.

- NASP support has helped to expand the technology base for Martin Marietta to apply the XD™ process to other materials. Materials manufacturing licenses are in place with the following companies.

  - -- Cast Ti-based XD with Howmet Corp.
  - -- Cast Al-based XD with Cercast Corp.
  - -- Semi-Solid Al-based XD with AMAX.
  - -- XD Copper production under discussion.

- Martin Marietta POC: Dr Steven Kampe, (301) 247-0700.

D. Beryllium Alloy Development

- NASP heat exchanger requirements have initiated renewed interest in a beryllium alloy (Lockalloy) not used since the 1960s when the original manufacturer ceased production. NASP requirements improved the formula (characteristics), resulting in significantly better manufacturing processes. NASP also substantially improved the characterization data base for this material. This new data base has allowed consideration of this material for commercial applications. The new NASP-developed material is trademarked AlBeMet™.

  - -- Benefits: Low density and high strength have resulted in reduced electrical subsystem weight. Excellent thermal conductivity will result in increased component life.
  - -- Applications:
    - --- Spacecraft structure for GE Astro Spacecraft
    - --- Electrical subsystems on the F-22

  - -- A major commercial spin-off of this product is underway. An AlBeMet rotary actuator for high-performance computer disk drives is being produced. The new disk drive actuator contains arms which are lighter, stiffer and thinner than the equivalent component made out of traditional materials. This allows both a more rapid access time to data stored on the hard disk without exciting the resonance frequency of the arms as well as the ability to stack about twice as many disks (i.e., provide greater memory capacity) within a given volume. Firm orders are in place ($12M FY92). The market for this product is conservatively estimated at $137M over the next three years. With an initial investment of $657,000 made by the NASP program, the return on investment is over 208 to 1.

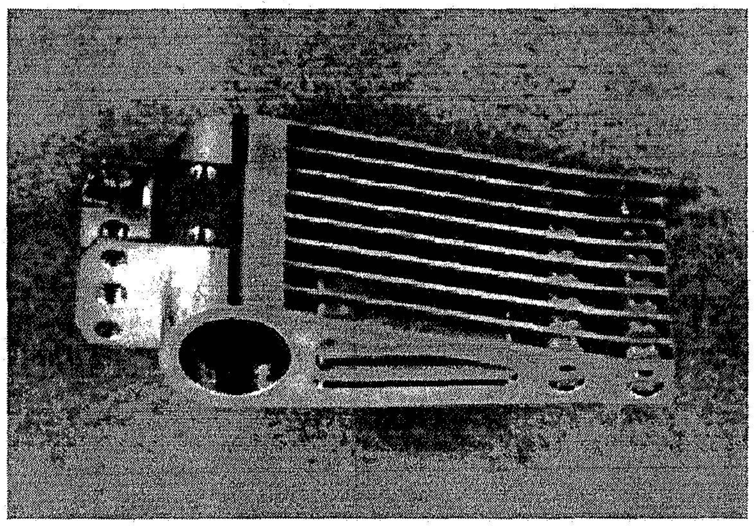


Figure 9. AlBeMet™ 160 Disc Drive Actuator Arm

--- Beryllium has traditionally been used exclusively for high-performance defense applications. Recently, the sole U.S. manufacturer, Brush-Wellman, was close to ceasing production of beryllium. (DOD/DOE Beryllium Coordination Committee Report, Feb 92.) Now with an established commercial vendor this material will be available for DOD applications as well as other commercial options.

-- Brush-Wellman, Inc: Mr Jim Marder (216) 383-4068

E. **Thermo Set/Thermo Plastic Material Development.** Advanced graphite epoxies and polyimides were applied for NASP requirements on large cryogenic fuel tanks. The enhanced data base has opened options for such large cryogenic structures.

F. **Metallic Coatings Development.**

- Initial research is underway to study NASP developed coatings for engine fuel injector nozzles which would allow increased use of methanol as an automotive fuel.

-- Benefits: This should result in less reliance on non-renewable energy sources, elimination of new injector material development costs, and a potential reduction in automotive emissions.

- Research is underway on NASP coatings used in industrial plant machinery operating in corrosive environments, e.g., fertilizer production.

  -- Benefits: Potential for an order-of-magnitude increase in serviceable life and maintenance costs. Current operations often dictate extensive sandblasting and repainting every few months.

Manufacturing Process Technologies:

A. **Thin-Sheet Rolling.** The NASP program has led to enhanced capabilities for NASP, which required thin-rolled sheets (3-6 mil) of titanium aluminides in order to make metal-matrix composites (MMC). These sheets did NOT exist prior to NASP. Under contract to NASP, Texas Instruments (TI) invented, developed and patented a highly complex (yet efficient) process for rolling thin-gage sheet.

- Solves ductility problem of titanium variants
- Allows manufacture of titanium MMC
  -- Large scale NASP Task D tank demonstrated hot structure concept
- TI POC: Mr Richard Delagi, (508) 699-5218

The NASP TI rolling process has allowed the development of new classes of materials:

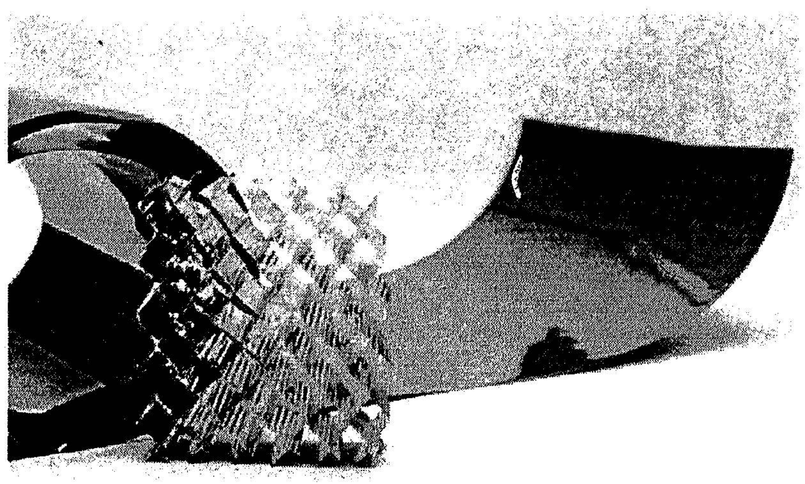


Figure 10. High Temperature Aluminum Steel And Honeycomb

- **High-Temperature Aluminum (HTA).** Allied Signal has developed a new aluminum alloy with the same temperature capability as titanium (Ti-6-4). However, the HTA alloy could only be cast into one-inch slabs because of ductility limitations. TI felt the same rolling process that solved NASP material problems could be applied here; TI has successfully rolled large quantities of HTA. Allied Signal has successfully produced honeycomb aircraft structure for Rohr Industries. Rohr is baselining these components for the Boeing 777. (Note that while NASP did not fund this effort, the NASP-developed process was critical to the final HTA advancement.)

  -- Benefits:

    --- Twice the thermal capability of aircraft-grade aluminum (2024 and 2219)
    --- Superior corrosion resistance in comparison with Ti-6-4
    --- 2/3 the weight of titanium
    --- Substantial cost saving (by a factor of 10)

  -- This material development had reached a dead end until the NASP-derived rolling process was applied.

- **Other Applications**: Other foils have been rolled by TI using the NASP-developed rolling process.

  -- Artificial valves for hearts require light weight, high strength and high durability to last over millions of beats. A Texas company is developing artificial valves using NASP titanium foil.
  -- The biomedical industry is developing a case for heart pacemakers requiring increased corrosion resistance.
  -- The thin foils have been bonded to steel in the manufacture of piping for water desalination plants. The pipe is cheaper (than clad steel) and has potentially increased operating life.
  -- Titanium foil is being applied to camera shutters for high-performance photography (large lens), which require light weight and high strength.

Figure 11. High Performance Camera Shutter From TI Titanium Foil

-- A major speaker manufacturer is developing new high-performance speakers using TI (NASP-enabled) foils.

- NASP-enabled foils are under evaluation in automobile catalytic converters to replace ceramic substrate.

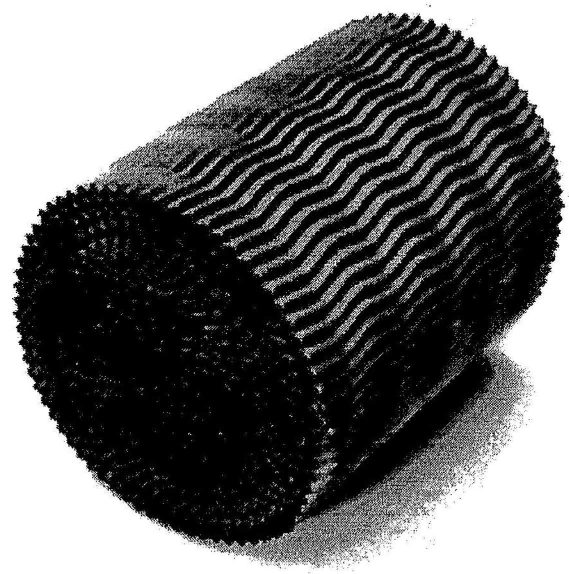


Figure 12. NASP Enabled Metallic Catalytic Converter Foil

-- Benefits: This reduces requirement for strategic materials (platinum and palladium catalyst) and significantly improves converter performance during cold operation.

B. **Megalaminates.** After TI found they could roll thin sheets of material using the NASP developed rolling process, they were asked to develop a new NASP engine material: copper niobium. Copper has excellent thermal characteristics but poor strength. Niobium could increase the strength but does not mix with copper; the niobium separates onto the surface. TI found by rolling the Cu-Nb into thin sheets, stacking the sheets (many hundreds of layers have been demonstrated) and then rolling the combined stack they could uniquely produce the required material with all the desired characteristics. This is a NASP-derived material developed using a NASP-derived process.

- Benefits:

-- Increased wear resistance
-- Excellent thermal characteristics

- Potential Applications:

  -- Throughout the NASP engine/combustion system
  -- Rocket combustors
  -- Copper-niobium tooling produced via the megalaminate process is in use throughout industry for commercial wire manufacturing and annealing to increase wire ductibility. Wire manufacturing tools made of copper-niobium last over four times longer than conventional tools. There is the potential for a 20% increase in production speed and a 10% reduction in power consumption.

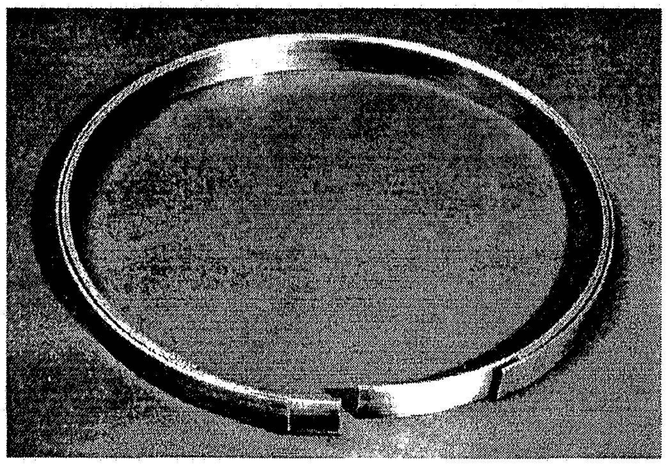


Figure 13. Copper-niobium Tooling For Wire Drawing

  -- A Cu-Nb megalaminate is under development for engine piston rings to replace chrome-clad steel rings. Engine companies have long desired to replace steel rings with copper; however, it was impractical before this NASP development.

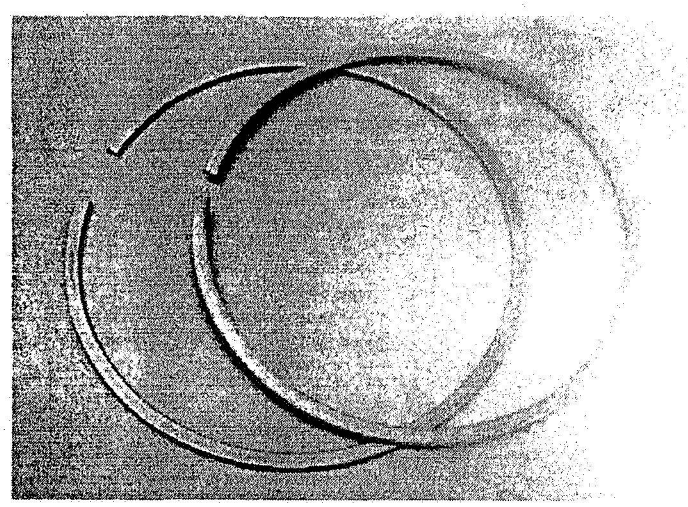


Figure 14. Auto Piston Rings Made Of Cu-Nb Megalaminate

-- Heat exchanger components for space systems and the chemical-processing industry have been built and are in test.

- **SILVAR™ Megalaminate**: TI found that the stacked Cu-Nb had superior performance (hardness) than predicted. TI began experiments using other materials. These materials exhibited similar superior characteristics. TI has patented a Silver-Invar (SILVAR™) material for electronics circuit boards.

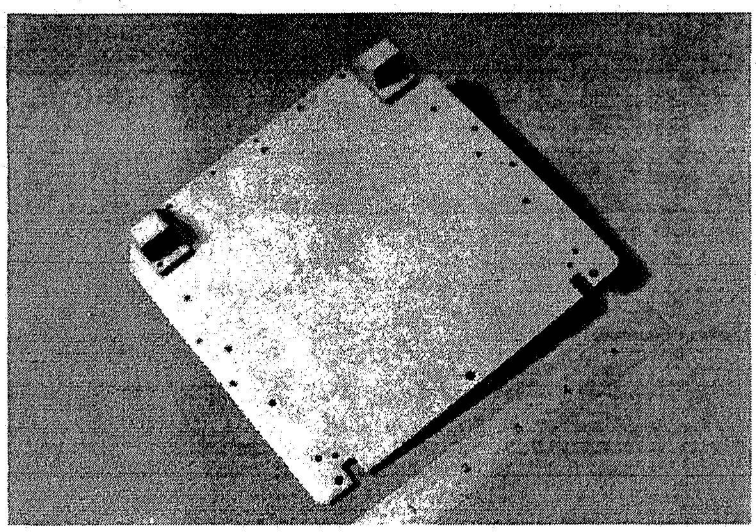


Figure 15. Standard Electronic Module From Isobaric™ Rolling-Cladding

-- Benefit:

--- Excellent conductivity
--- Good machineability
--- Durability substantially increased by better matching of thermal expansion characteristics
--- Higher temperatu:e electronic circuits

-- This is patented and in production for industry applications. (The material was unavailable prior to the NASP rolling process.)

- Other Megalaminate: TI has developed other materials concepts using their thin foil concepts. The process involves multiple stacking and rolling of several hundreds layer laminates. Ten titanium varieties and two aluminum varieties are currently in production.

-- Benefits:

--- Over twice the material hardness
--- Superior corrosion resistance/durability
--- Ease of manufacturing

C. **Other Manufacturing Processes:** In addition to improved material characteristics (foils) and new materials (megalaminates), this Isobaric^{tm} rolling technology is showing promise as a manufacturing technology for other classes of materials.

- Stainless-steel heat exchangers typically require a brazing mixture of TiCuNi powder. When combined, these materials are extremely brittle and therefore require manual manufacture (brazing) of heat exchangers. TI now rolls thin sheets of Ti, Cu, and Ni and then clads these materials to stainless steel. Next, sheets of steel are brazed together into complex heat exchangers using automated technique. Benefits:

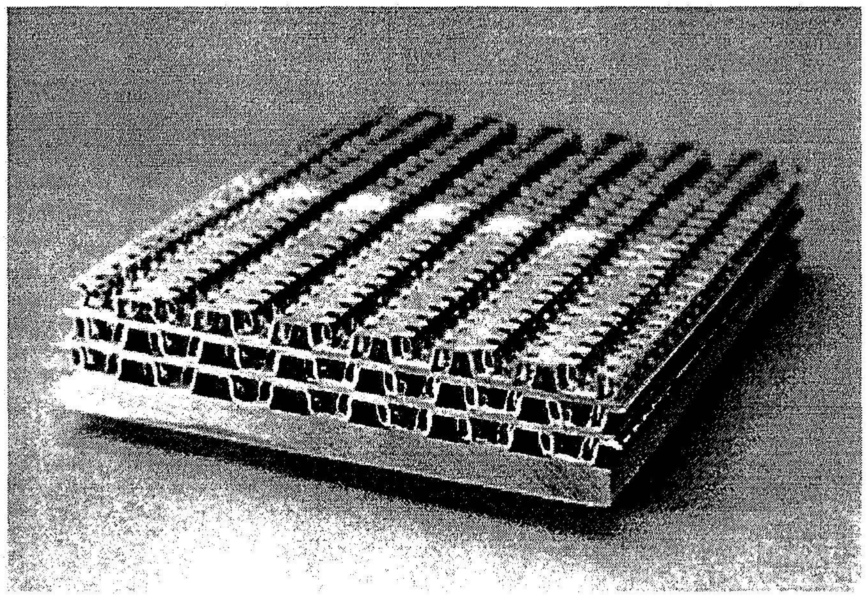


Figure 16. Industrial Heat Exchanger Produced From Megalaminates

-- Substantially lower cost
-- Higher quality
-- Higher performance
-- Parts are in production by ITT

- TI is evaluating use of megalaminates in the production of rapid solidification rate (RSR) powder metals. These metals have substantially better performance than natural metals. A critical part of their manufacture is exact proportion control as material is melted. TI is using their megalaminate (multi-ply) technology to precisely control RSR material feed stock.

Other Process Benefits:

- Material process technology developed for NASP will potentially impact other material process developments. For example, this should enhance General

Motors' work to explore aluminum metal matrix composites for wheels and suspensions; many of the NASP manufacturing processes should be of interest.

- NASP vehicle production will develop the required large-scale manufacturing infrastructure. Textron is planning a production plant for NASP titanium metal-matrix composites with initial operational capability in the second half of 1993. It will produce a variety of shapes, from panels to stringers.

- NASP research has enhanced the development of diffusion bonding and super-plastic-forming technology for advanced titaniums. Higher strength and 50% cost reductions in manufacturing are expected.

- NASP has contributed to the development of Rapid Solidification Rate (RSR) technology for existing metals. NASP has demonstrated fabrication of specific high-temperature and high-strength metal components. A prototype RSR production facility was demonstrated for NASP purposes.

  -- Benefits of (RSR) technology allow the use of lower-weight alloys in high-temperature environments. Specific enhancements include 400°F, 800°F, and 1600°F maximum temperature increases for aluminum, titanium, and columbium respectively.

- NASP is developing new production techniques for advanced ceramics.

  -- Benefits: Oxidation resistant coatings and their application have demonstrated over 200-percent increase in durability.
  -- Benefits: A 300-percent improvement in high-temperature capability (versus existing aircraft materials) has been demonstrated. The world's largest structural elements of carbon-carbon potentially counter foreign (French) world leadership.
  -- Benefits: Carbon-silicon carbide combustor for cruise-missile engines were tested at temperatures over 3000°F. Lighter weight, higher performance, and greater efficiency are obtained.

<u>Computational Fluid Dynamics (CFD)/Computational Techniques:</u>

A. **CFD.** The NASP program has been a major factor in the dramatic growth and improvements in computational fluid dynamics (CFD) within the U.S. over the last five years. The primary impetus from NASP came in the domains of high-speed (supersonic/hypersonic) flows around complex contours with extremes of aerothermochemistry, flow injection, and/or thermochemical boundary interactions. More generally applicable advances were made in grid generation and techniques for comparing computational results with sophisticated experiments for code validation.

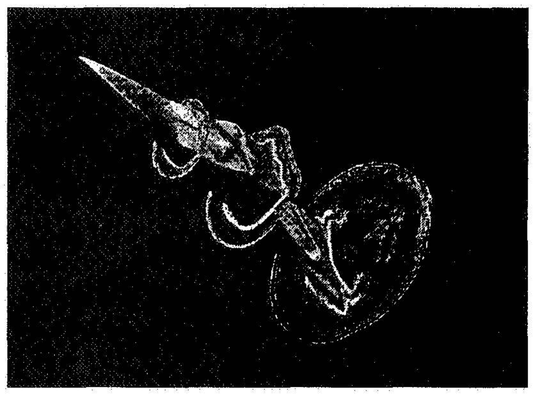


Figure 17. NASP Vehicle CFD

Looking simply for direct applications of specific segments of software from NASP to other tasks misses a major NASP contribution. The extent of the challenges, the demanding schedules, and the many thousands of hours of supercomputing vastly accelerated the overall growth of CFD. A significant proportion of the U.S. expertise in CFD was utilized and challenged by NASP. Also, NASP participants have been forced by the limits of ground-based facilities to learn how to rely heavily on CFD. The benefits of this approach are beginning to be felt for less severe fluid-dynamic environments as validation work continues and overall experience grows. NASP has significantly advanced certain types of validation work too. The combined impact on CFD of NASP and other major program has been enormous.

- NASP combustion research: The NASP program has developed higher-order predictive tools to model complex fuel injection, mixing and combustion. Efforts are currently underway to apply these techniques to the design of improved combustion chambers for automotive, aerospace, and electric power generation, yielding higher efficiencies and lower pollution rates.

- System/subsystem analysis, integration, and evaluation: The unprecedented degree of vehicle integration required by the X-30 has led to the development of detailed analysis tools for evaluating complex aeromechanical/aerothermal interactions. Several of these tools have found applications in the aerospace field yielding cost and time savings due to a reduction in the required amount of testing.

- As part of the development of these computational technologies, the NASP program has been responsible both for new codes and for enhancing and upgrading of existing codes.

  -- During the early phases of the program, NASA Langley Research Center took the existing NASA-developed CFL3D and SPARK codes and upgraded them to accommodate analysis for NASP configurations. CFL3D was to be used to analyze NASP configurations for global aerodynamic/propulsion integrated performance, while SPARK was intended to be used for the analysis of a 3-D supersonic flow.

-- The AFWAL PNS code (PNS being an acronym for parabolized Navier-Stokes) was modified by Science Applications International, Inc. (SAIC), adding high temperature (equilibrium air) capability and an advanced turbulence model.

-- The NASP contractor team developed the PENGUIN and FALCON 3-D codes, which include equilibrium chemistry capability, high-resolution shock wave capture, and mass conservation for accurate propulsion performance prediction. The contractor team also developed the Integrated Structural/Thermal Analysis Program to allow modeling of complex thermal/mechanical interactions.

- Each of these analysis tools has found application in non-NASP analyses, as documented by the following examples:

  -- NASA Ames Research Center modified SPARK to study hydrogen/air detonation. They are also using it to establish the credibility of a more efficient Ames combustion code and to improve combustion rate models with the goal of improving the SPARK code itself.

    --- Benefit: This enhanced basic research and stimulated development of additional improved codes.

  -- Douglas Aircraft Company applied CFL3D to compute transonic flow solutions for such configurations as the C-17 wing/body and the MD-11/12. The solutions obtained by the NASP-enhanced CFL3D code were found to agree with experimental and flight data better than those obtained from in-house Euler/boundary-layer and full-potential/boundary-layer codes.

    --- Benefit: The code supports more accurate, timely aircraft design capability; cost savings result from reduced dependence on wind-tunnel testing.

  -- United Technology Research Center is using SPARK for turbulence and chemical model development, analysis of combustor subscale configurations, and assessment of mixing schemes. A modified version of SPARK is being applied to analysis of the High Speed Civil Transport (HSCT).

- General Dynamics applied the PENGUIN and FALCON-3D codes to the design and analysis of hypersonic engine inlets and nozzles for the INCAAPS program. The strong shock wave capture and mass conservation capabilities of these codes enabled configuration analysis which had not been possible with previous tools.

  -- Benefit: These codes extend analytic capability to flight regimes previously unreachable by numerical simulation.

- Douglas Aircraft used the Integrated Structural/Thermal Analysis Program to perform evaluation and selection of an acceptable design for an F-15 heat exchanger upgrade. Repeated failures in designs under study were threatening aircraft delivery schedule and cost. The NASP-developed code enabled duplication of the failure modes and identification of flaws in the designs which led to the procurement of an alternate design.

  -- Benefit: This saved significant labor, test time, and procurement cost.

B. <u>**Neural Network Fault Diagnostics:**</u> NASP small business initiative for diagnostics of the vehicle management system. A spin-off to a major computer manufacturer (Cray) is in progress.

- Benefits: This form of artificial intelligence provides more efficient, less labor intensive, and more accurate fault diagnosis for complex electronics applications such as supercomputers.

Long-Term Benefits

<u>Technical Infrastructure</u>

NASP continues to have a significant impact on the test and analysis capability in the U.S. The technical infrastructure now ranges from hardware for sophisticated wind-tunnel tests for hypersonic propulsion to equipment for producing large quantities of slush hydrogen to unique facilities for imposing severe thermo-mechanical loads on large structures. Of even greater importance is the cadre of scientists and engineers who know how to maximize results from these facilities.

At the beginning of the program, some participants were active in resurrecting hypersonic-related facilities that had been in the process of being dismantled or destroyed. A surplus wind tunnel from NASA Langley (the Langley Expansion Tube) was recovered from storage and rebuilt at the General Applied Science Laboratory on Long Island. The NASA Lewis Plumbrook research facility was being excessed by the GAO for sale on the open market. Today the complex (near Cleveland, OH) is vibrant and conducting research on advanced cryogenics, potentially benefiting all future space systems.

To meet urgent needs across the entire flight regime, several small scale research and development facilities have been built by NASP that create a foundation for growth and understanding of how to use such capability in the future. For example, the U.S. now has a Mach 11-12 propulsion capability (100MW arcjet facility) at NASA Ames Research Center. Also, free-piston-driven shock tunnels for Mach 20+ test conditions now exist at the California Institute of Technology. WYLE Laboratory, Norco, CA, has created a new Hydrogen Technology Center capable of testing hydrogen-carrying components and subscale systems to high-temperature conditions; the ability to generate high heat loads on moderate size samples has been increased by an order of magnitude. NASP will grow this capability over time to system scale components at the Phillips Laboratory, Albuquerque, NM; at Edwards Air Force Base, CA; and at the Wright Laboratory, Dayton, OH. NASP has pioneered the use of lasers for a high-temperature nonintrusive test capability; this capability will be expanded for use in facilities throughout the country.

NASP is developing facilities that will evolve into a powerful capability for future operational systems. The Arnold Engineering Development Center (AEDC), Tullahoma, TN, has two national test facilities for structures and propulsion. The NASA Stennis Static Test Stand and Component Test Facilities will also be expanded to meet NASP needs. Similarly, static-structural test capability at Edwards AFB will be expanded to test the vehicle structure.

Additionally, NASP is supporting new test techniques and technical expertise across the entire range of facilities being developed. These new facilities will represent substantial new capability for all future aerospace developments in aerospace and other high-technology programs.

<u>Civil (NASA) and Commercial Space Launch</u>:

If the NASP program is completely successful, NDVs could provide reliable, frequent, space launches with low operating cost; this is pivotal to the competitive commercial space-launch market where cost of the satellite and the cost of launch are about equal. Because of high cost and limited infrastructure, 50% of commercial launches are now accomplished on foreign systems; this is a capability formerly dominated by U.S. launch systems. The President's National Space Launch Policy, 24 Jul 91, addressed this issue: "...actively consider commercial space launch needs and factor them into decisions on improvements in launch facilities and launch vehicles."

Currently, the U.S. is not competitive in commercial space launch. For example, as of January 1992, Hughes Commercial Spacecraft has 23 customer orders for services. Of those, only two are scheduled to be launched on U.S. boosters (one Atlas and one Delta). In a global market, commercial payloads have moved to the best available service.

The multi-mission capability of NDVs makes commercial needs synergistic with military requirements. The characteristics of high reliability, flexibility, and responsiveness that make NDVs attractive for military missions also make them a viable low-cost commercial launch vehicle. The payload community can take advantage of cost savings and design to lower cost, more reliable, new systems. The NDVs attribute of assured space access should equate to preferred customer service and enable U.S. leadership in an increasingly competitive world.

<u>Military Utility</u>:

As a science and technology program, the NASP is intended to demonstrate and validate the technological basis for single-stage-to-orbit space launch vehicles.

Operational vehicles derived from a successful NASP program could provide very significant increases in our strategic, tactical and deterrent capabilities. Requirements for future NDVs are articulated in a joint Mission Need Statement (MNS 202-92, Military Aerospace Vehicle) signed by Lt Gen Moorman and Gen Butler of Air Force Space Command and Strategic Air Command, respectively. The Mission Need Statement is currently in the final review process. These two commands have also executed a Memorandum of Agreement (MOA, dated 27 Nov 90) to support and advocate the NASP development effort. More importantly, their involvement ensures the operational relevance of NASP technology demonstrations.

The fundamental capability to be obtained through NDVs is the ability to deliver large payloads (about ten tons) to low Earth orbit in an assured aircraft like operation. This capability is critical for the military to exploit the inherent advantages of operating systems in space. Additionally, tremendous long-term cost advantages are expected for single-stage-to-orbit transportation systems.

Even more significant is the multi-role capability NDVs provide for surveillance, force enhancement, and other military missions. The attributes inherent to future operational hypersonic vehicles embody some of the fundamental tenets of warfighting capability in a dramatically changing world: specifically, speed, range and survivability. The ability for global overflight from the continental United States, within an hour of launch, can be coupled with atmospheric, airbreathing maneuvering that enhances their unpredictability and survivability.

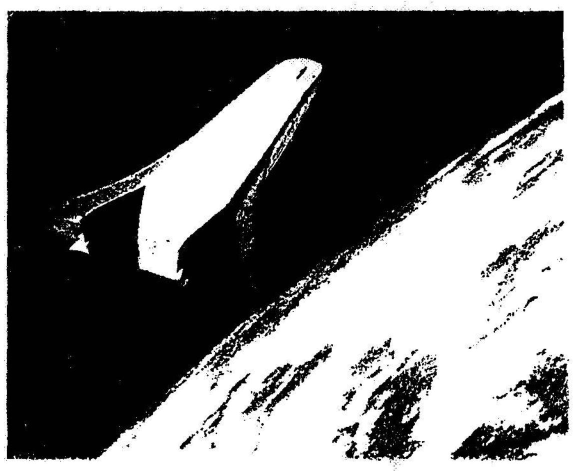
Summary

The objective of the NASP program is to provide the key technologies to permit the Nation to develop both civil and military vehicles for space launch and for hypersonic cruise in the atmosphere. The X-30, the NASP flight-research vehicle, is envisioned to use airbreathing primary propulsion and hydrogen fuel to achieve single-stage-to-orbit flight with horizontal takeoff and landing.

The total value of NASP to the United States must be seen from many perspectives to be fully appreciated. The long-term benefits center around flexible, efficient access to space using advance technologies that enhance the international technological preeminence of the Nation. Operational vehicles for civil and military missions will provide the U.S. a significant, clearly visible measure of strength. Perhaps of even greater importance is the invigoration and growth of the national technology base. Near-term applications in materials, materials processing and computer-related sciences have already been identified. This trend gives evidence that the NASP investment will continue to pay significant dividends as the U.S. develops the technology for a revolutionary class of vehicles, the world's first aerospace planes.

Mr Terry Kasten/ASC/NAR, WPAFB OH 45433-6503/(513) 255-3165/kbc/3 Aug 92 - updates by Code RN/CM

NATIONAL AERO-SPACE PLANE PROGRAM



OVERVIEW

- GENERAL PROGRAM SUMMARY
- GENERAL TECHNOLOGY TRANSFER SUMMARY
- SPECIFIC TECHNOLOGY TRANSFER EXAMPLES
- PROJECTED 20-YR ECONOMIC IMPACT
- VALUE SUMMARY

INDUSTRIES THAT WILL BENEFIT FROM NASP TECHNOLOGIES

AEROSPACE

AUTOMOBILE

COMPUTER

DEPARTMENT OF DEFENSE

DEPARTMENT OF ENERGY

MANUFACTURING

MEDICAL

OIL DRILLING

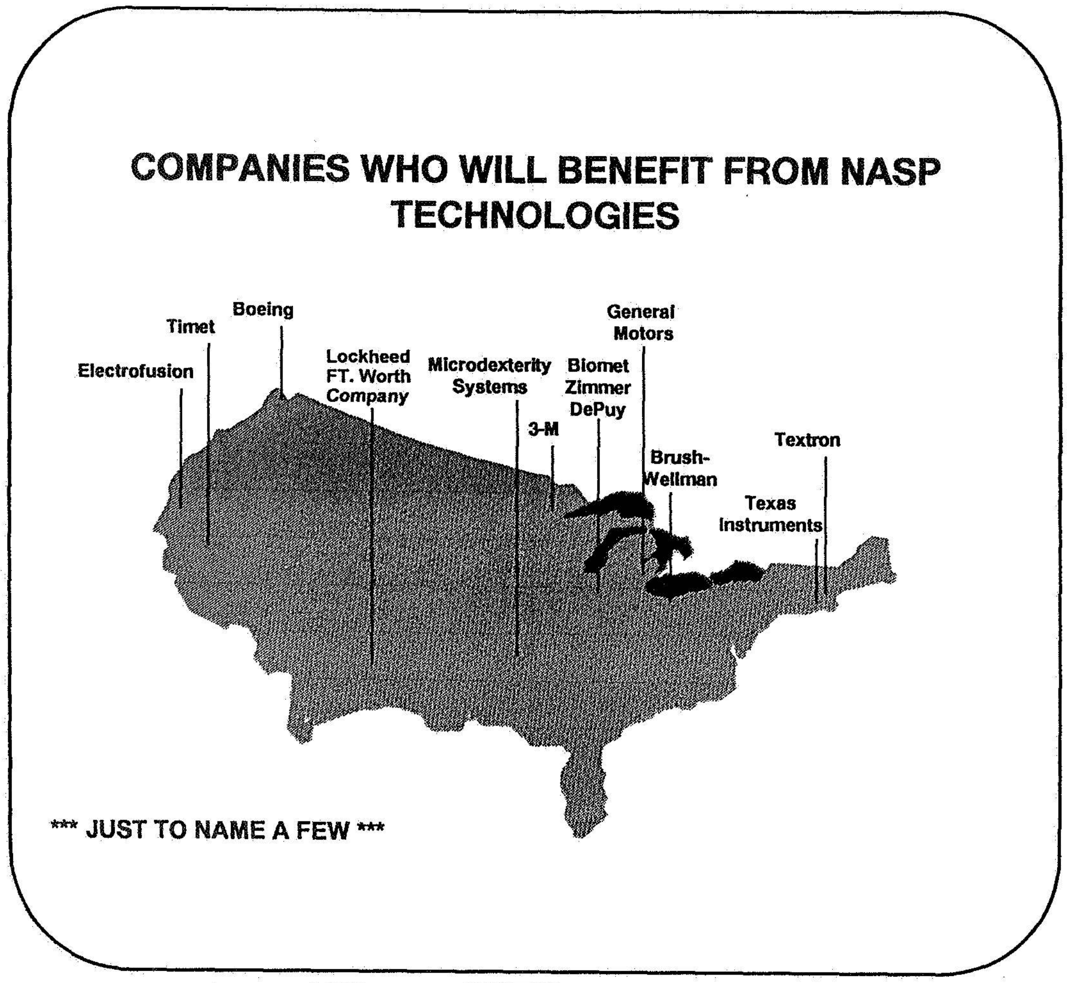
SPORTING GOODS

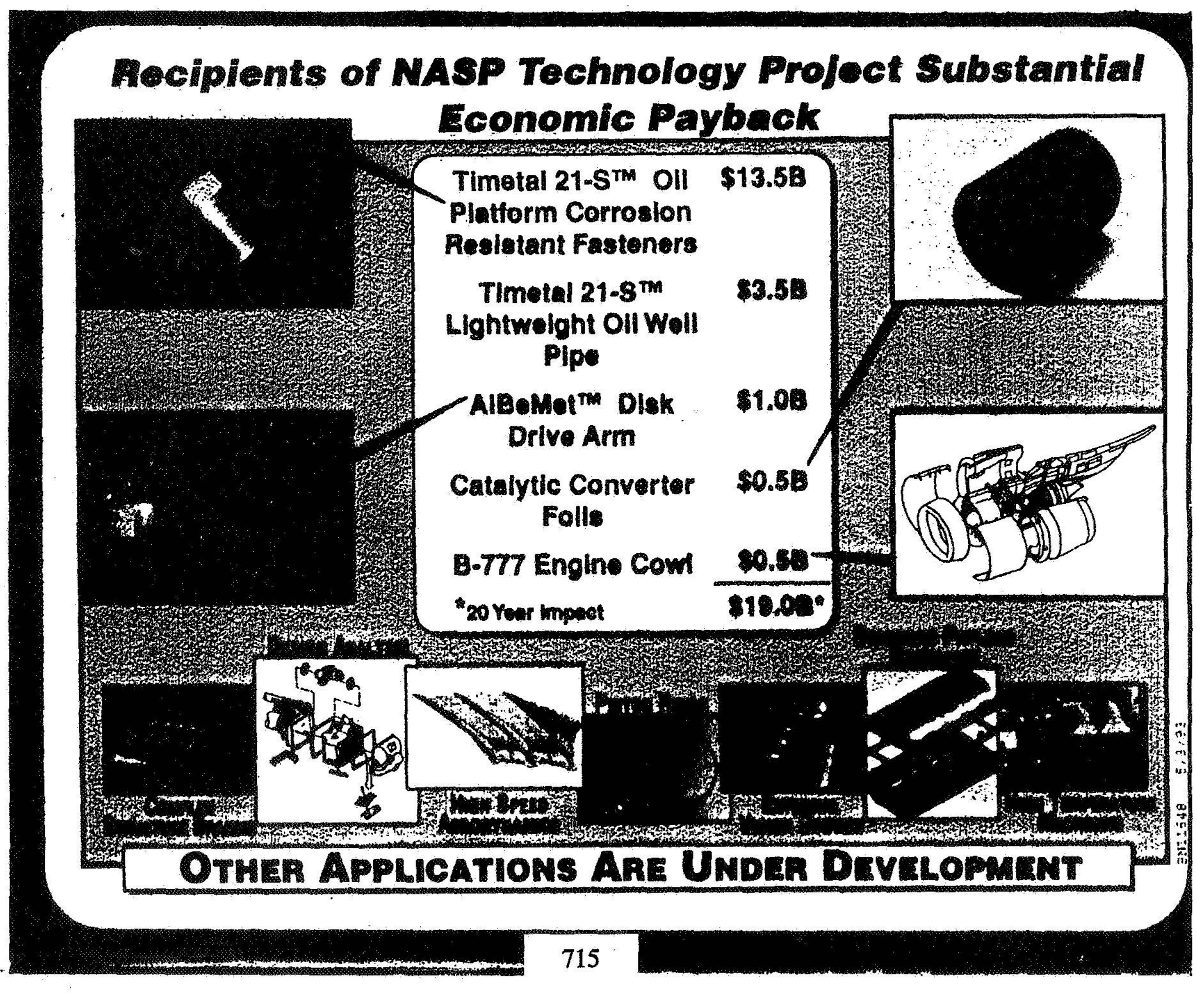
WATER PURIFICATION

Technology Transfer Examples

- Oil Well Pipe
- B-777 Naccelles
- XD™ Compressor Blades
- XD™ Missile Components
- AlBeMet™ Disk Drive Actuator Arm
- Automobile Catalytic Converter Substrate
- Off Shore Oil Platforms
- High Performance Camera Shutters
- Electronic Circuit Boards
- Machine Tool Manufacturing
- F-18 Gun Blast Deflector
- Artificial Hips
- Electronic Cooling Components
- Speaker Foils
- High Temperature Aluminum
- Heart Artificial Valves
- Engine Rings, Valves, Pistons
- Advanced Heat Exchangers

NAR/Kasten
92-0127


COMPANIES WHO WILL BENEFIT FROM NASP TECHNOLOGIES
Boeing
Timet
Electrofusion
Lockheed FT. Worth Company
Microdexterity Systems
Biomet Zimmer DePuy
General Motors
3-M
Brush-Wellman
Textron
Texas Instruments
*** JUST TO NAME A FEW ***

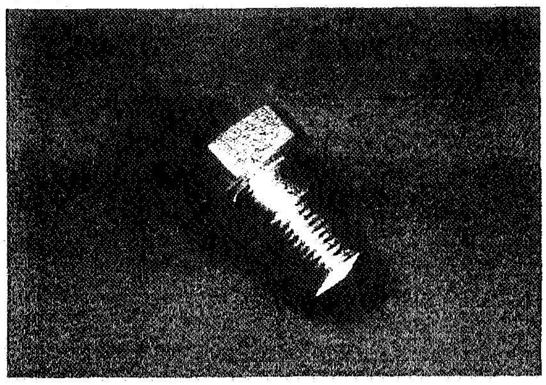

Recipients of NASP Technology Project Substantial Economic Payback
Timetal 21-S™ Oil Platform Corrosion Resistant Fasteners $13.5B
Timetal 21-S™ Lightweight Oil Well Pipe $3.5B
AlBeMet™ Disk Drive Arm $1.0B
Catalytic Converter Foils $0.5B
B-777 Engine Cowl $0.5B
20 Year Impact $19.0B
OTHER APPLICATIONS ARE UNDER DEVELOPMENT

Materials & Structures

MATERIAL DEVELOPMENT

- Beta Titanium Alloy for Chemical Plant Use

  ✓ TIMET



NASP CONTRIBUTION

- Development and Characterization of TIMETAL® 21S Alloy
- Production of Commercial Scale Billets

BENEFITS

- 60 Times the Corrosion Resistance of Prior Alloys

BN11302 4/23/92

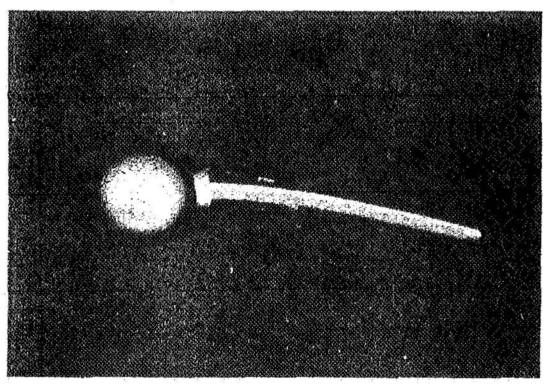
NASP

Materials & Structures

MATERIAL DEVELOPMENT

- TIMETAL® 21S Prosthetic Implants

  ✓ TIMET



NASP CONTRIBUTION

- Development and Characterization of TIMETAL® 21S

BENEFITS

- Longer Implant Lifetime, 30% to 50% Reduction in Surgical Procedures
- Enhanced Bone/Prosthesis Compatability
- Reduced Toxicity

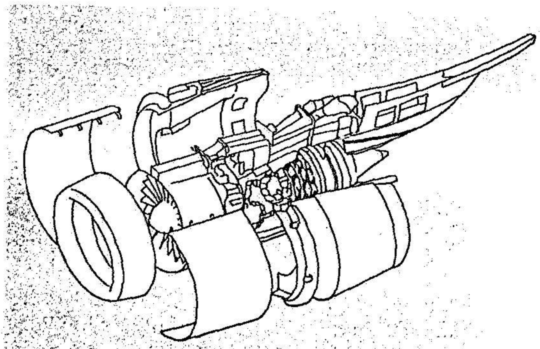
BN11475 7/20/92

NASP

Materials & Structures

MATERIAL DEVELOPMENT

- Timetal 215™ Application to Boeing 777
  - ✓ Timet Corporation



NASP CONTRIBUTION

- Development and Characterization of Timetal 215™

BENEFITS

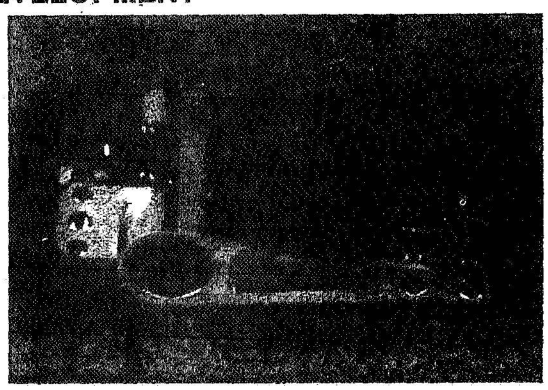
- Resistant to Corrosion by Aircraft Hydraulic Fluid ("Skydrol")
- 1700lb Weight Savings Per Aircraft
- 20 Year Life Cycle Cost Savings of $2M to $5M Per Aircraft

BN11505 8/12/92

Materials & Structures

MATERIAL DEVELOPMENT

- AlBeMet™ Disk Drive Actuator Arm
  - ✓ Brush Wellman, Inc.
  - ✓ Electrofusion



NASP CONTRIBUTION

- Improved Fabricability
- Improved Material Characterization Database

BENEFITS

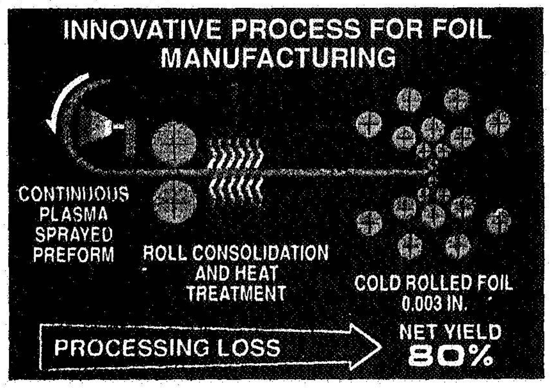
- 8.5 ms Access Time (15% Improvement)
  - ✓ Low Weight Allows Rapid Movement
  - ✓ High Stiffness Reduces Damping Time
- Improved Thermal Properties Result in Simpler Hard Drive Design

BN11517 8/17/92

Materials & Structures

PROCESS IMPROVEMENT

- Cold Rolled Foil Processing of Gamma Titanium Aluminide
  - ✓ Texas Instruments
  - ✓ General Electric



NASP CONTRIBUTION

- Initial Development and Characterization of Ti-Al
- Development of Processing Techniques for Hard-to-Work Materials

BENEFITS

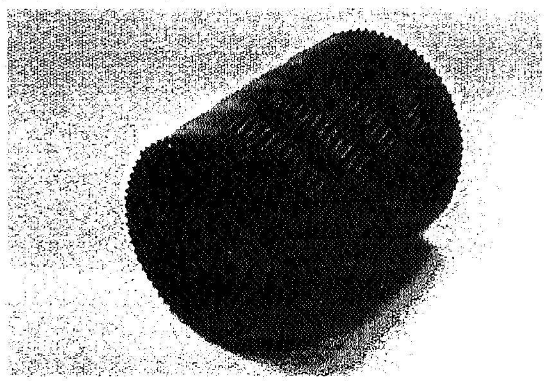
- Existing Process (Chemical Milling) Has 5% Yield, Cost = $3000/lb
- New Isobaric Process Yields 50% of Original Ingot With Potential for 80% Yield and a Cost of About $300/lb

BN11405 7/13/92

Materials & Structures

PROCESS IMPROVEMENT

- Nickel-Aluminum Foil Automobile Catalytic Converter Element
  - ✓ Texas Instruments



NASP CONTRIBUTION

- Development of Isobaric Processing Technique

BENEFITS

- No Cracking Failure Mode (Foil Not Brittle as are Ceramics)
- Reduced Amounts of Platinum Catalyst Required
- Significant Reduction in Exhaust Emissions

BN11511 8/12/92

HYPERSONIC ECONOMIC IMPACT MACROECONOMIC ASSESSMENT

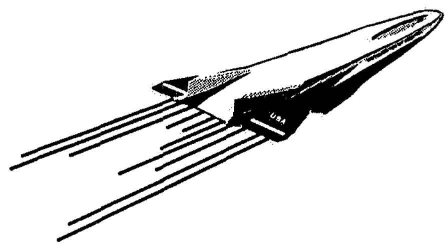
- BACKGROUND
  - ~ PRINCETON ECONOMIC RESEARCH, INC
  - ~ UNIVERSITY OF MARYLAND INFORUM INPUT/OUTPUT MODEL
  - ~ COMPLETED IN 1990
- STUDY CONTENT
  - ~ MACROECONOMIC ASSESSMENT
  - ~ HYPERSONIC TECHNOLOGIES
  - ~ IMPACT ON WORLD ECONOMY
- STUDY RESULTS
  - ~ U.S. GNP INCREASE - $126B (CUM - 20 YRS)
  - ~ U.S. EXPORTS INCREASE - $44B (CUM - 20 YRS)
  - ~ U. S. EMPLOYMENT INCREASE - 38,000 JOBS

VALUE OF NASP TO THE NATION

- DUAL-USE TECHNOLOGY INVESTMENT
  - ~ Military Requirement for Ready Space Access
  - ~ Civilian Space Launch
- TECHNOLOGY TRANSFER INVESTMENT
  - ~ Proactive Program
  - ~ $19B To Date/ $26B Projected (over 20-yr period)
  - ~ Oil, Medical, Aerospace, Automotive, and Military
- INFRASTRUCTURE INVESTMENT
  - ~ Test Facilities
  - ~ Revolutionary Design Tools
  - ~ Innovative Management Tools

NASP PROGRAM SUMMARY

- NASP Is Developing Technology For The Next Generation
- NASP Technology Satisfies Important U.S. Civil And Military Needs
- NASP Spinoffs Have Already Paid Significant Dividends
- Strong Support Is Vital To Ensure Continued Progress



Session T4: TECHNOLOGY TRANSFER METHODS

Session Chair: Ken Cox

MCC

Speeding the Payoff: New Paradigms in Technology Transfer

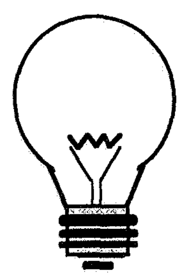
Joseph Sims
Vice President, Business Development
Microelectronics and Computer
Technology Corporation (MCC)

MCC

Illuminating Opinions...

- "When the Paris Exhibition closes, electric light will close with it and no more will ever be heard of it." — *Erasmus Wilson, professor at Oxford University, 1878*

- "[Edison's ideas are] good enough for our transatlantic friends...but unworthy of the attention of practical or scientific men." — *Report of a committtee set up by the British Parliment to look into Edison's work on the incandescent lamp, c. 1878*



From Cerf, Christopher and Navasky, Victor, *The Expert's Speak: The Definitive Compendium of Authoritative Misinformation* [New York: Pantheon] 1984.

MCC

...That Didn't Ring True

- "Well informed people know it is impossible to transmit the voice over wires and that, were it possible to do so, the thing would be of no practical value." — *Editorial in the Boston Post, 1865*

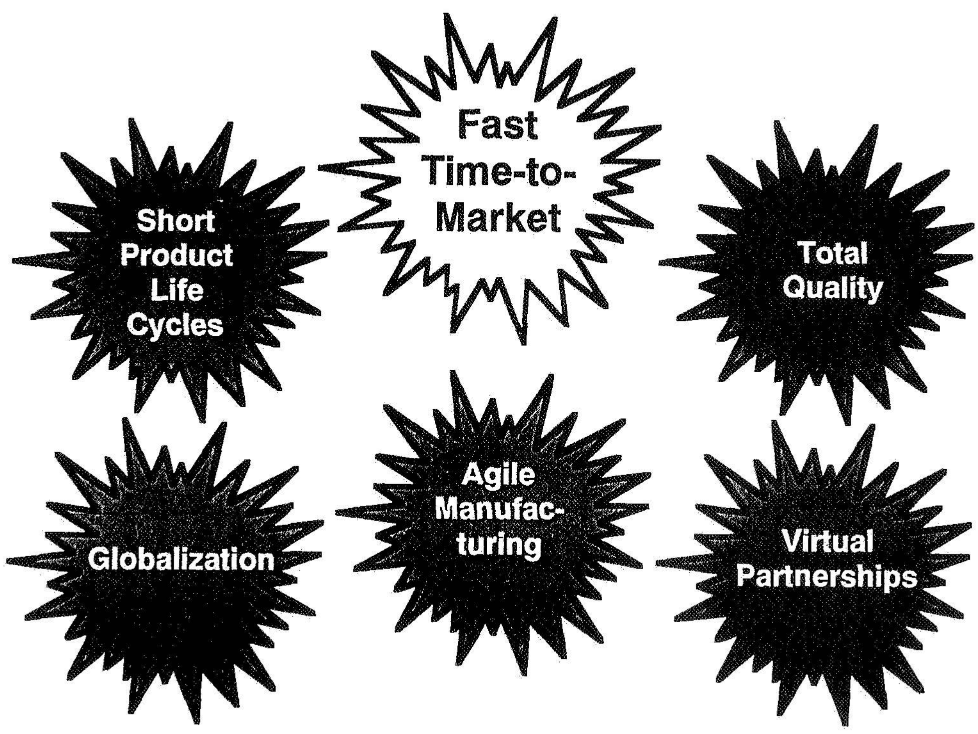
- "That's an amazing invention, but who would ever want to use one of them?" — *President Rutherford B. Hayes, after participating in a demonstation call.*

- Historical note: Bell patented the telephone in 1876, and offered to sell it to the Western Union Telegraph Company for $100,000. Western Union declined.

From Cerf, Christopher and Navasky, Victor, *The Expert's Speak: The Definitive Compendium of Authoritative Misinformation* [New York: Pantheon) 1984.

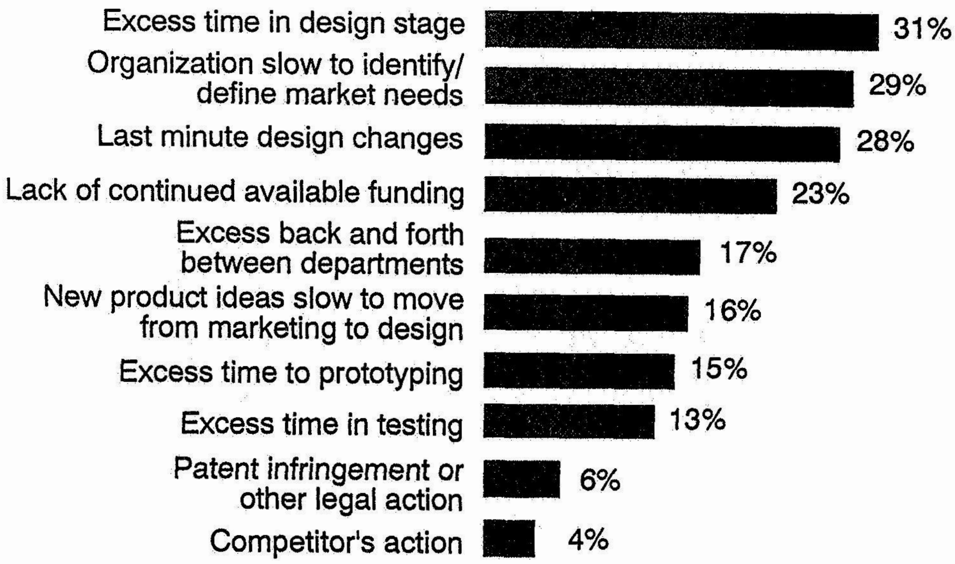
MCC

The New Environment



MCC

The Key: Time-to-Market



Data from a study by Ernst & Young

*percent of respondants rating each significant

MCC

Time-to-Market: The Bottom Line

If your company is late to market by:					
6 mo.	5mo	4mo.	3mo.	2mo.	1mo.
Your gross profit is reduced by:					
-33%	-25%	-18%	-12%	-7%	-3%
Improve time-to-market by only 1 month, and profits improve:					
+11.9%	+9.3%	+7.3%	+5.7	+4.3%	+3.1
For revenues of $25 million, annual gross profit increases:*					
+$400k	+$350k	+$300K	+$250k	+$200k	+$150k
For revenues of $100 million, annual gross profit increases:*					
+1600k	+$1400k	+$1200k	+$1000k	+$800k	+$600k

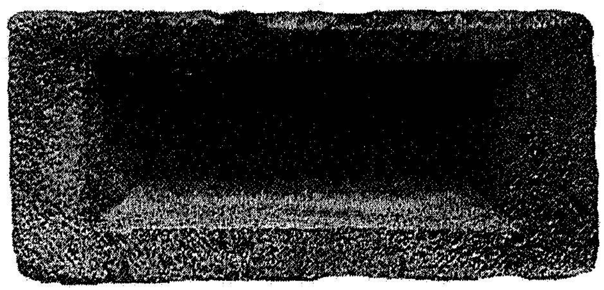
6 months late to market = one-third less profit!!!

Source: McKinsey & Co., in Vesey, Joseph, "Speed to Market Distinguishes the New Competitors," *Research and Technology Management*, Nov-Dec 1991, p. 34

*Over course of 5 years

MCC

THIS IS THE LATEST THING IN MICROPROCESSORS. IN 12 MONTHS IT WILL BE OBSOLETE.



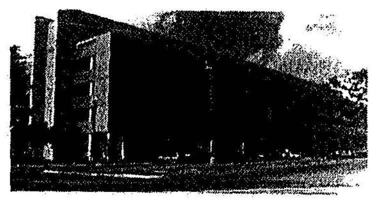
THIS IS A BRICK. IN 12 MONTHS IT WILL STILL BE THE LATEST THING IN BRICKS.

from an advertisement in the Wall Street Journal

MCC/VENTURES

MCC is a National Leader in Advanced R&D

- For profit, Delaware corporation
- $45M in annual research
- $500M in total research since inception
- 130 patents
- Current portfolio includes 30+ projects
- $50M in project backlog



MCC
Headquarters
Austin, Texas

MCC High Value Electronics
Austin, Texas

BKVentures

MCC MCC Members by Industrial Concentration

AEROSPACE
Aerojet Electronics
E-Systems
Hughes Aircraft
Lockheed
Martin Marietta
Rockwell International
TRW, Inc.
United Technologies

COMPUTER-AIDED DESIGN
Cadence Design
Mentor

COMPUTER SYSTEMS
Apple Computer
Austin Business Computers*
Ceridian Corporation
Digital Equipment Corp.
Hewlett Packard
Honeywell
IBM
AT&T Global Info. Solu. Co.
Tandem Computers
Unisys

CONSUMER ELECTRONICS
Eastman Kodak

DISPLAYS
Crystaloid Electronics Co.*
ProjectaVision

ELECTRONICS
General Electric
ITAC Systems, Inc.*
Jamar Technology
KDT Industries, Inc.*
Marlow Industries, Inc.*
MIC Technology

ELECTRONICS (con't.)
RTB Technology
Shipley Company
Teledyne
Westinghouse

ENERGY
Amoco Production

INVESTMENTS
Warburg, Pincus Ventures, Inc.

MATERIALS
SI Diamond Technology, Inc.*

PUBLIC AGENCIES/ NON-PROFIT
The Markel Foundation
National Security Agency

RESEARCH
EMPF
Embedded Computing Institute, NAWC
Environ. Res. Inst. of Michigan
Lawrence Livermore Lab.
Mitre
Sandia National Labs

RETAIL
The Limited
Lindsay Office Products, Inc.

SEMICON. EQUIP. & MATERIALS
3M
AMP, Inc.
E.I. Dupont Denemours
Elmo Semiconductor Corp.*
Olin Corp.
Optical Associates*
Temptronic Corp.*

SEMICONDUCTORS
Advanced Micro Devices
Elmo Semiconductor Corp.*
Harris
Motorola
Multichip Technology, Inc.
National Semiconductor
Space Systems Loral
Texas Instruments
VLSI Technology

SERVICES
American Express
Andersen Consulting
SAIC

SOFTWARE
Corporate Memory Systems
Knowledge Based Systems
Microsoft
Interval Research Corporation

TELECOMMUNICATIONS
AT&T
Bellcore
Broadband Technologies
DSC Communications
Northern Telecom
Sprint Corporation

UNIVERSITIES
Boston University
Drexel University
Inst. for Micromanufacturing, La Tech Univ.
New Jersey Institute of Technology
Rice University
Tx. Hospital Educa. & Research Fdn.
University at Albany, SUNY
University of Southern California

* Small Business Associates

MCC MCC Projects Address a Diverse Range of Cutting Edge Technologies

High Value Electronics

Semiconductor Packaging/Interconnect
Low Cost Interconnect
Multichip Modules
Advanced Tape Automated Bonding
Flip Chip Technology Development
Laser Bonding
Multichip Systems Design Advisor
Reliability of Electronic Systems without Hermeticity Requirements
Laser Direct Write for Interconnect
Technology for Single Chip/Few Chip Systems
Packaging for Portables
Optoelectronic Interconnect

Flat Panel Displays

Electronic Applications of Superconductivity

Information Systems

Intelligent Systems
Neural Nets
Common Sense Knowledge Base for Enhanced Decision Support(Cyc)
Knowledge Based Natural Language

Agile Manufacturing Technology
Group Coordination Technology
Network Services and Technology for Electronic Commerce (Enterprise Integration Net)
Distributed Heterogeneous Information Management and Integration

Distributed and Parallel Processing
Object Management Facilities for Scalable, Distributed/Parallel Applications
Rapid Prototyping for Scalable, Parallel Computing Systems (High Performance Modules)
Embedded Fault Recovery Facilities for Real-time Ada Applications (Ada Fault Tolerance)

MCC

How Did We Get Here...

MCC Technology Transfer Success

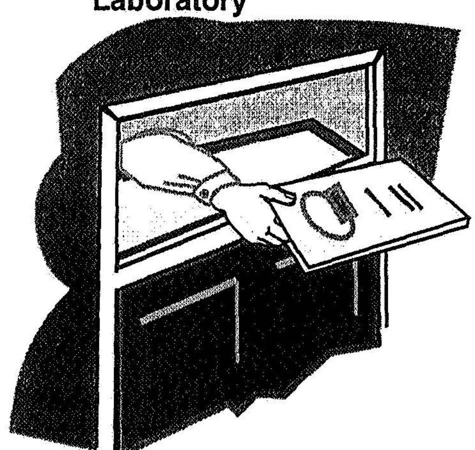
Transferred >200 technologies
Produced >400 videotapes
Published >2700 technical reports
130 Patents assured

MCC

The Traditional View of Technology

Corporate Research Laboratory → **Technology Transfer** → Product Design and Manufacturing

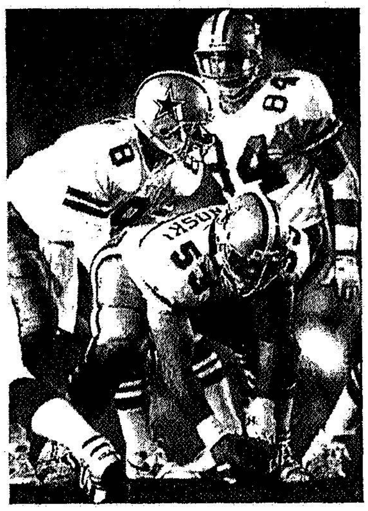
External Research Laboratory/ Consortium → Corporate Research Laboratory → Product Design and Manufacturing



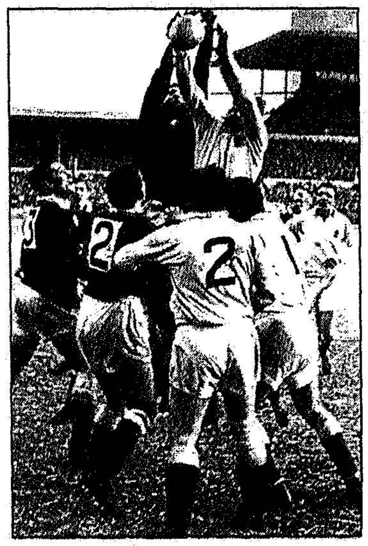
MCC

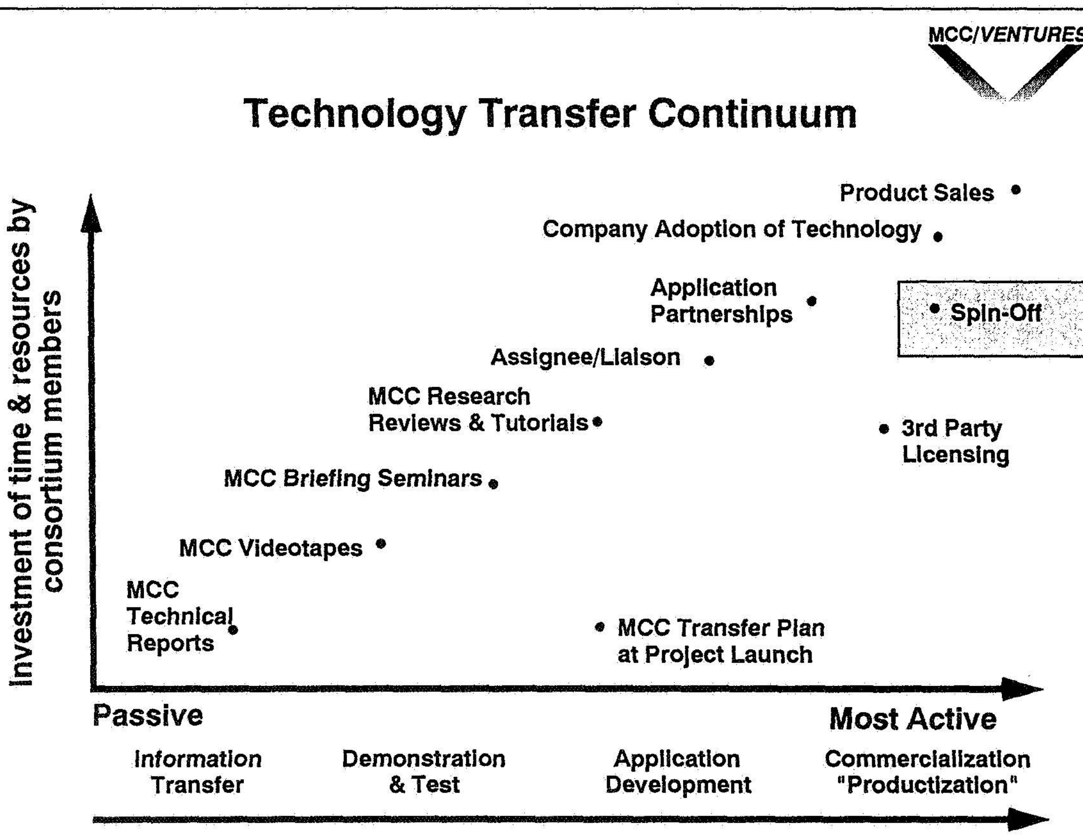
But the Reality is...

...Seldom a Smooth Handoff



...More Often a Frantic Scramble

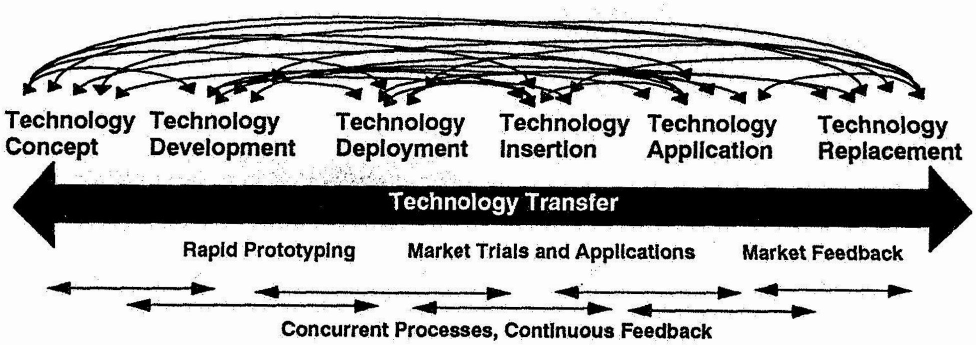




BKVentures

MCC

Technology Transfer



Research Input

Business Input

MCC/VENTURES

Enterprise Creation:

- Technology Development as the Basis for Business Start-ups
- The Result: Technology Transfer to the Purchasing Office Rather than the Research Lab
- Examples:
  - ThermoElectron
  - Xerox Technology Ventures
  - MCC Ventures
- Challenges
  - Financing
  - Entrepreneurial Leadership
  - Product Development
  - Market Development
  - Administrative Support

MCC/*VENTURES*

MCC FAMILY OF COMPANIES

Member Companies	MCC	*VENTURES*	Spin-Offs
	– Consortium Research	– Product Maturation	– Products

BKVentures

MCC/*VENTURES*

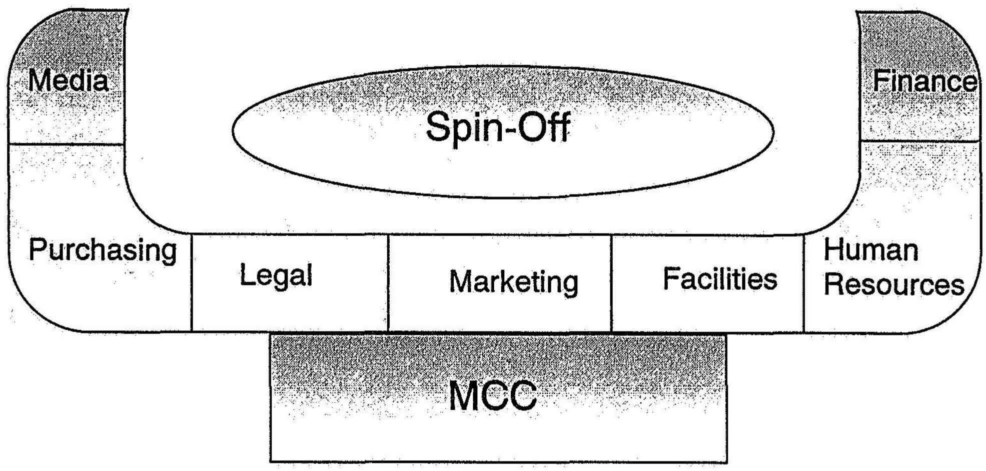
What Does *VENTURES* Do?

- Market Risk Reduction
- Infrastructure Support
- Access to Technology
- Seed Round Funding

BKVentures

MCC/*VENTURES*

Infrastructure Support



BKVentures

MCC/*VENTURES*

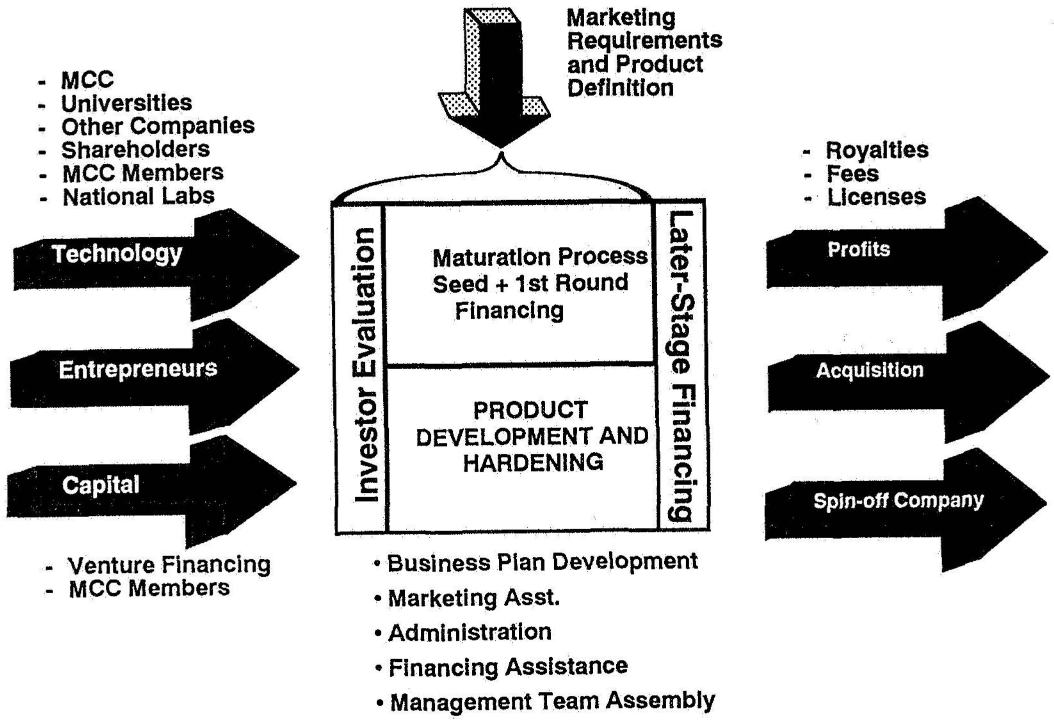
Why Do Small High Tech Businesses Fail?

- Investable CEO
- Customers
- Market Access
- New Products
  - Technology
  - Capital
- Management

BKVentures

MCC

The MCC Ventures Process



MCC/VENTURES

Evolutionary Technologies, Inc.

- Product: EXTRACT Tool Suite which automates the migration of data between dissimilar data storage environments
- Regular royalty stream now flowing to MCC:>$200,000 in payments to date
- Completed second round of funding. Source: Menlo Ventures, Arete Ventures, private sources
- Sales and marketing
  - — About 17 customers/20 installations
  - — Making marketing expansion into international markets
- ~60 employees (started with 3)

ET Evolutionary Technologies, Inc.

MCC/VENTURES

Pavilion Technologies

- Product: Process Insights, a neural net-based software tool for modeling, optimizing,and controlling complex processes
- First profitable quarter achieved in 3rd quarter of operations
- Version 2.0 of Process Insights software shipped
  - Full graphic user interface
- New product, Software CEM (continuous emission monitoring) gaining widespread acceptance as cost effective tool for complying with Air Quality Act requirements
- Sales and marketing:
  - 23 paid customer installations
  - 20 installations for evaluation purposes
- Staffing currently @30 full and part time (started with 6)

Pavilion
Turning Your Data Into Gold!

MCC/VENTURES

Corporate Memory Systems, Inc.

- Product: CM1, a computerized groupware tool to support group decisionmaking and documentation of decision processes
- Four installations now operating
  - Nabisco and Pacific Gas and Electric well established
  - Launching new sites at NYNEX, Chemical Bank
  - In discussions with MCI
- Port of CM1 to Unix complete
- Beginning next round of financing
- *"Three of the most promising areas in the booming groupware field are: capturing organizational memory, codifying organizational knowledge, and supporting and/or eliminating meetings... Corporate Memory Systems has a commanding lead in this important segment of the workgroup computing market... the software is easy to learn, seductive and offers immediate benefits to the groups who use it." -- Patricia Seybold Group*

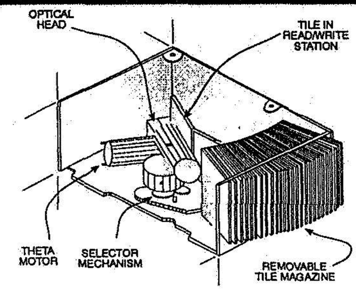
CM

MCC/ VENTURES

Tamarack Storage Devices, Inc.

- Product: Commercializing holographic storage technology for use in the personal computer and workstation markets
- Joint development agreement signed with Projectavision
- Prototype of initial product operational
- Holostore technology recognized in Business Week "Best of 92" issue and as "Technology of theYear" for 1993 by Industry Week

TAMARACK
STORAGE DEVICES

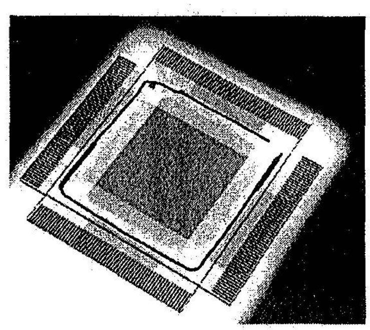


MultiStore (left front view)

MCC/ VENTURES

SI Diamond Technology (SIDT)

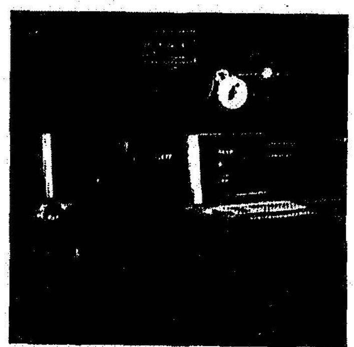
- Joint Venture Agreement with MCC completed
  - — Goal: To produce a new generation of flat panel displays based on diamond field emission display (DFED) technology
  - — MCC and SIDT pool patents and know-how
  - — SIDT receives license to manufacture DFEDs
  - — Manufacturing to begin in 1994
  - — MCC retains exclusive licensing rights after 2 years


Photograph of SIDT Diamond Field Emission Display prototype

MCC

Development Initiatives: Connectivity and Distributed Systems

- Carnot
  - Suite of technologies to unify physically distributed information, enterprise-wide
  - Information integration and management across heterogeneous systems
  - Complete communication framework, from interface to network communications
  - Knowledge-based natural language interface



Carnot technology addresses the need for integration of distributed information management, including flexible, adaptable interfaces

BKWorkshop 9/93

Carnot Application Partnerships

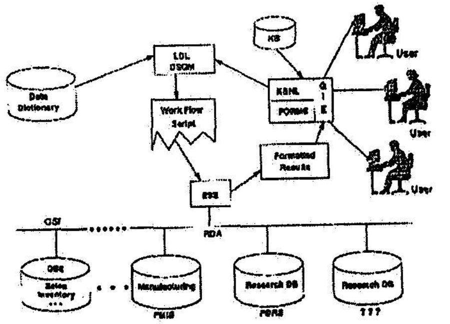
Partner	Purpose	Status
Dept. of Defense	Text retrieval from distributed databases	Beginning coding for implementation
Pacific Bell	Enterprise/data modeling	Project complete
Bellcore/Ameritech	Service provisioning	Detailed project plan completed and approval pending
Eastman Chemical	Accessing multiple distributed databases Custom interface design/natl. language	Implemented and operational Continued enhancement by adding data bases

Carnot Application Partnerships: Status

MCC

Significant Technical Accomplishments: Carnot

Eastman Chemical Application Partnership Prototype Operational



Technical Significance

- Application partnership objective: Allow direct retrieval of information by chemists
  - — Easy to use form-based interface
  - — Transparent access to multiple, heterogeneous databases
  - — Allow concurrent access
- Knowledge discovery system now being implemented
  - — Knowledge-based induction and multiple strategy explanation
  - — For error detecting and cleaning, pattern retrieval, schema generation

Business Value

- Faster, less costly responses to queries
- Reduce tendency to repeat costly experiments by recapturing relevant data
- Knowledge discovery benefits potentially significant:
- Error detection and prevention
- Analytical assistance and discover of new relationships in data

—

—

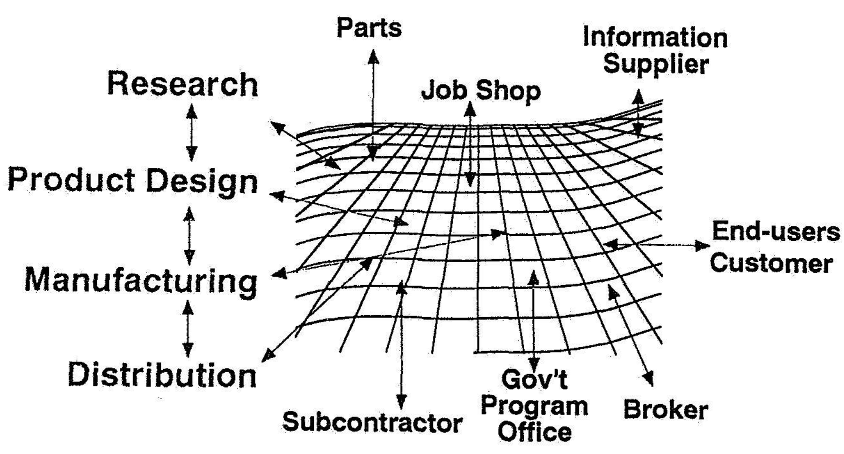
Project Status

- Users now regularly accessing system using forms-based interface built on MCC's LDL technology (2 databases) and natural language interface built using MCC's KBNL (1 database)
- NL Interface being extended to additional databases during 1Q94
- Other future work — generating initial templates automatically; improving templates interactively

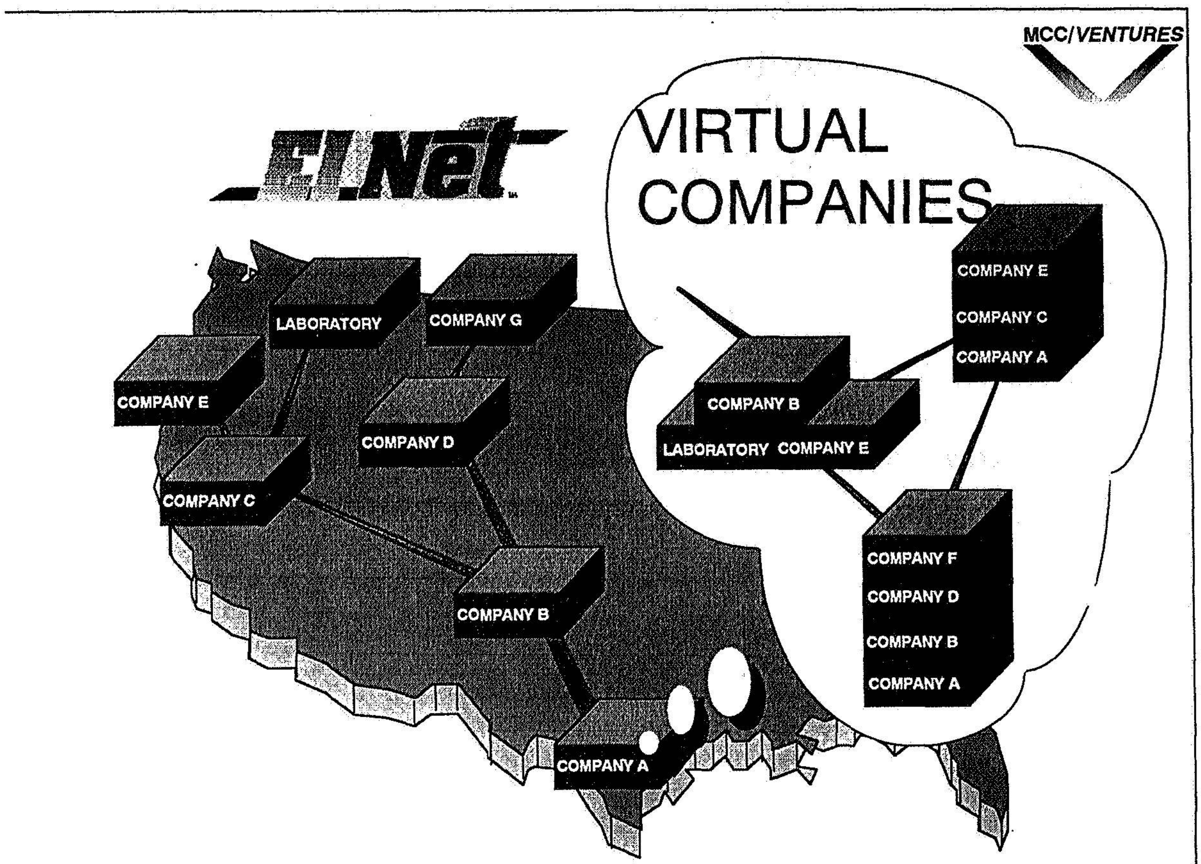
MCC

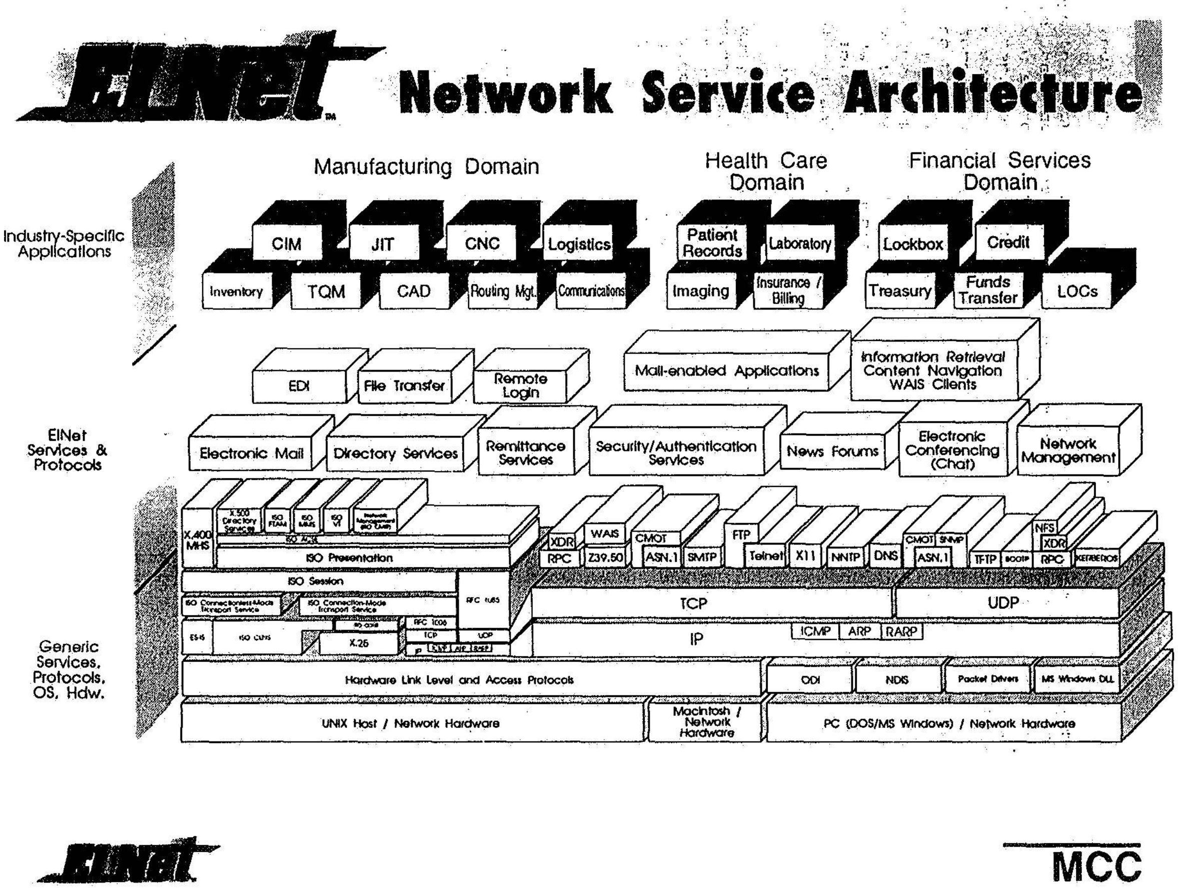
MCC/*VENTURES*

Building an Electronic Commerce Infrastructure



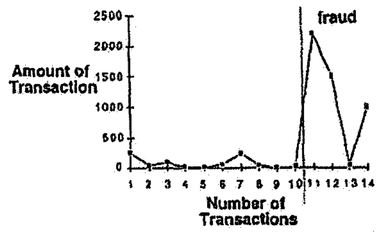
Industrial Internet:
- Electronic Commerce
- Multi-Company Product Development
- Accelerated Technology Transfer
- Manufacturing Supply Chains


MCC/VENTURES
VIRTUAL
COMPANIES
COMPANY E
COMPANY C
COMPANY A
LABORATORY
COMPANY G
COMPANY E
COMPANY D
COMPANY B
LABORATORY COMPANY E
COMPANY C
COMPANY F
COMPANY D
COMPANY B
COMPANY A
COMPANY B
COMPANY A


Network Service Architecture
Manufacturing Domain
Health Care Domain
Financial Services Domain
Industry-Specific Applications
CIM
JIT
CNC
Logistics
Inventory
TQM
CAD
Routing Mgt.
Communications
Patient Records
Laboratory
Imaging
Insurance / Billing
Lockbox
Credit
Treasury
Funds Transfer
LOCs
EDI
File Transfer
Remote Login
Mail-enabled Applications
Information Retrieval Content Navigation WAIS Clients
ElNet Services & Protocols
Electronic Mail
Directory Services
Remittance Services
Security/Authentication Services
News Forums
Electronic Conferencing (Chat)
Network Management
Generic Services, Protocols, OS, Hdw.
X.400 MHS
ISO Presentation
ISO Session
X.25
XDR
RPC
WAIS
Z39.50
CMOT
ASN.1
SMTP
FTP
Telnet
X11
NNTP
DNS
ASN.1
TFTP
NFS
XDR
RPC
TCP
UDP
IP
ICMP
ARP
RARP
Hardware Link Level and Access Protocols
ODI
NDIS
Packet Drivers
MS Windows DLL
UNIX Host / Network Hardware
Macintosh / Network Hardware
PC (DOS/MS Windows) / Network Hardware
MCC

Significant Technology Accomplishments: Neural Nets

MCC/VENTURES



Visa/BankAmerica Fraud Detection Application Successful

- Fraud detection system on-line at BankAmerica (San Francisco)
- Providing real-time transaction support

- Reduced fraud costs (which were $1B in 1992) — marginal improvements can yield substantial returns
- System has already more than recouped investment:
  - First week, identified $100K in fraudulent transactions
  - Second week, identified $120K
  - Third week, identified $150K
- Success of "live" system indicates potential for new business opportunity

- Application provides foundation for new initiative in Intelligent Systems for Financial Services (ISFS)

Accelerating the Improvement of Software Practice

Squeezing 20 Years into an 8-Year Time Capsule

Bill Hefley

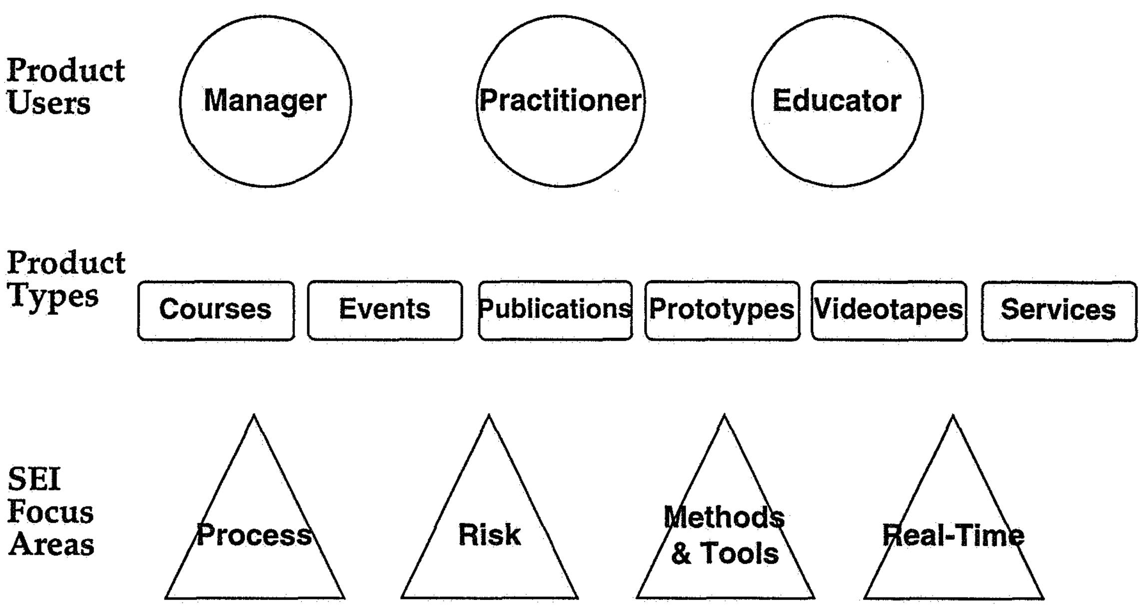
Dual-Use Space Technology Transfer Conference

February 2, 1994

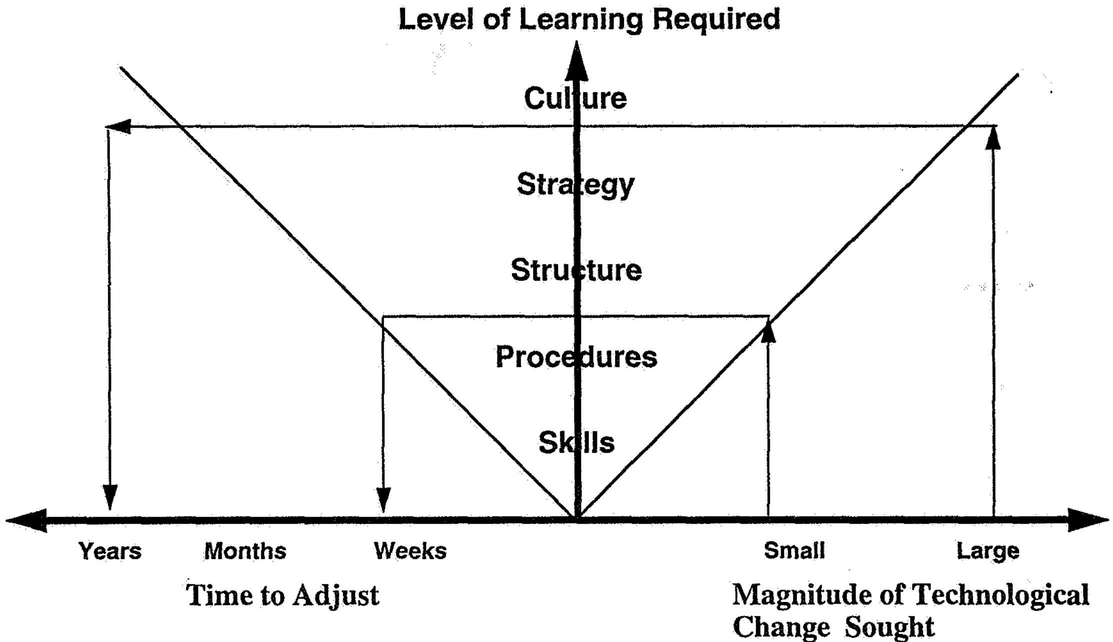
Software Engineering Institute
Carnegie Mellon University
Pittsburgh, PA 15213

Sponsored by the U.S. Department of Defense

Products & Services



Software Engineering: A Big Change



Critical Success Factors

A number of critical factors have been shown to play a major role in successful transitions of software technology into practice, including:

- implementation history
- sponsorship
- involvement
- change agents
- resistance
- culture
- technology fit

Barriers

Barriers to successful technology transition include:

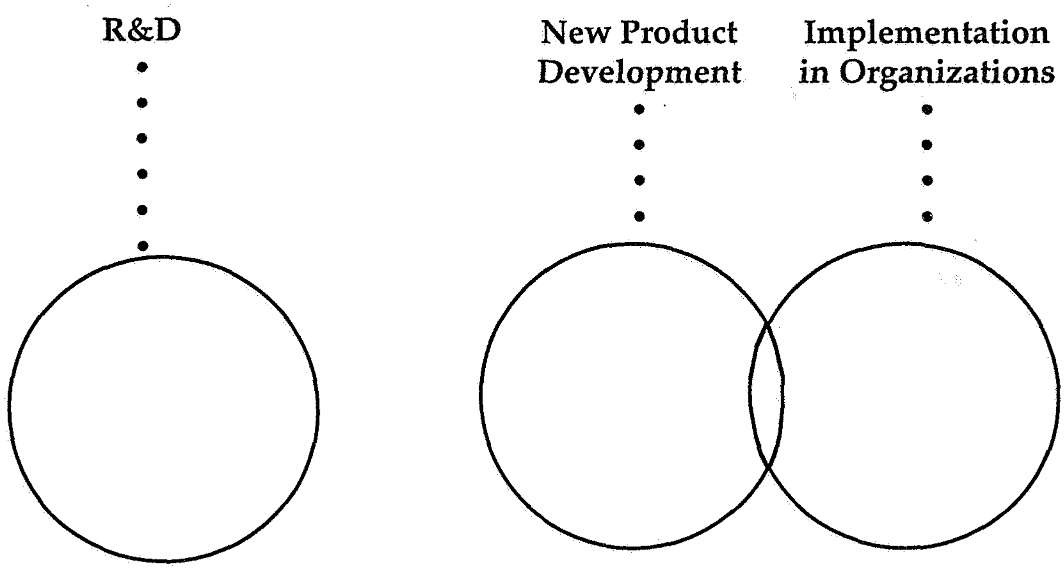
- psychological hurdles
  - not invented here (NIH)
  - not in my back yard (NIMBY)
- ability to adopt
  - readiness
  - receptivity
  - recognition of need
- high costs
- contracting disincentives
- packaging (training, documentation, consulting)
- appropriateness (technology fit)

Hurdling the Barriers

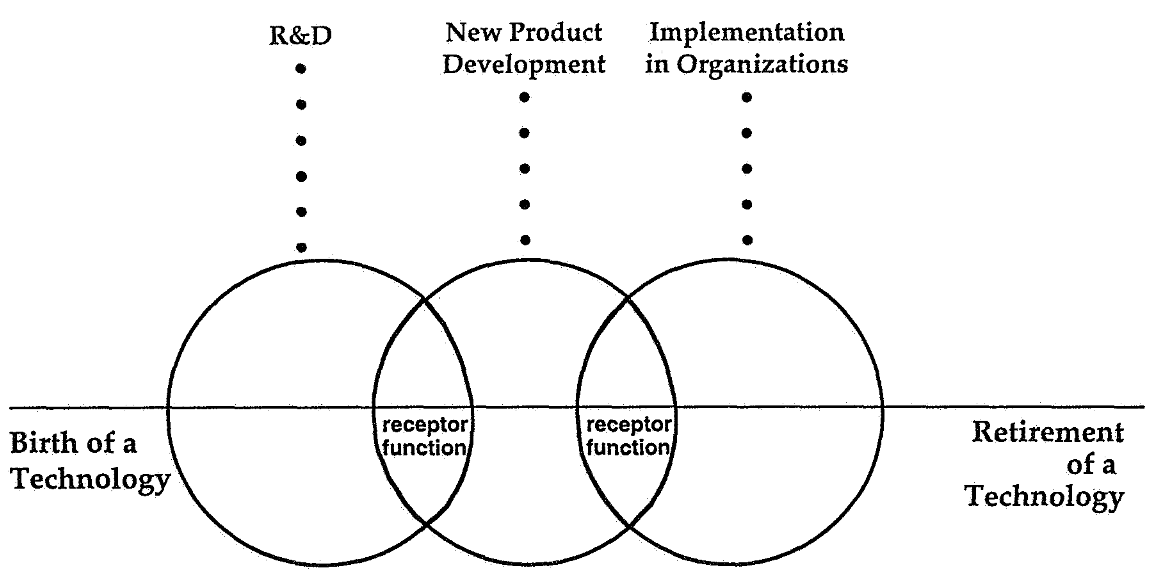
The SEI has confronted these barriers in maturing:

Area of Focus	Implemented Improvements
software technology	rate monotonic scheduling theory
software process	software process assessment, definition, modeling, and measurement
software profession	professional-level software engineering education

Solution to the Problem



Technology Transition Framework



Improving the Context

Successful technology transition has been accomplished, in part, by improving the context in which transition occurs.

- providing a focusing goal
- improving the technical capabilities of the community
- assuring that the technology is supported
  - training materials
  - personnel
- assuring that the technology is packaged
  - conceptually coherent
  - implementable improvements

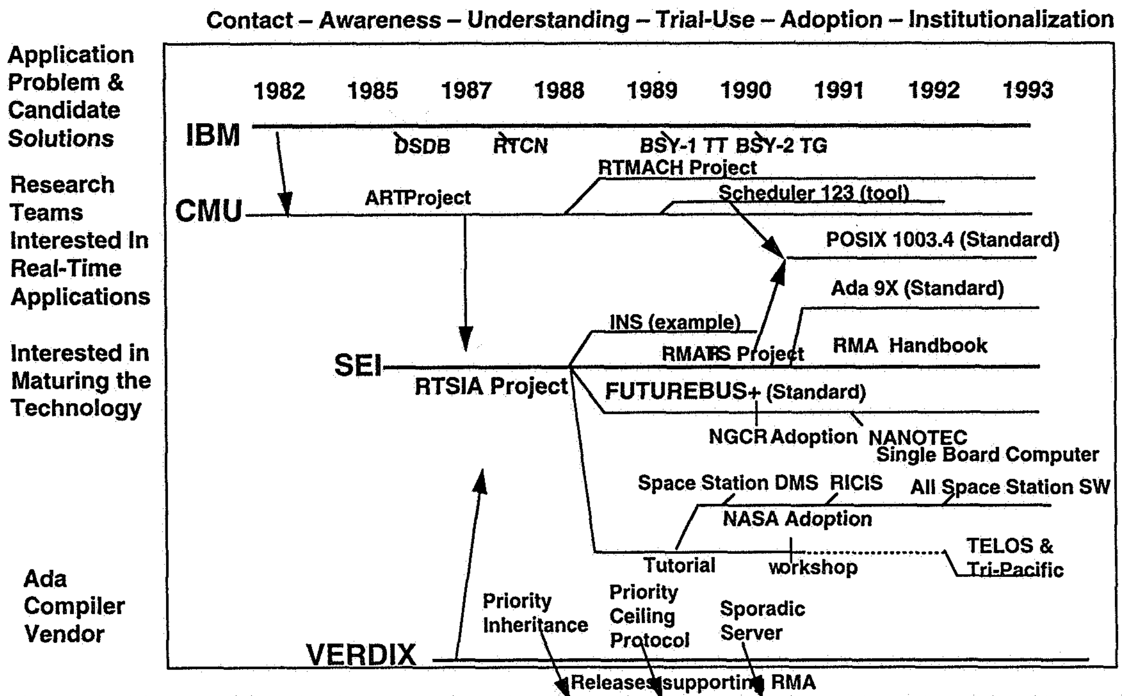
Maturing a Technology

Developed Rate Monotonic Analysis (RMA)

- provides a sound engineering basis for ensuring predictable performance of real-time systems
- stimulated industrial and academic research
- adopted in published and proposed standards: FutureBus+ (P896.3), Ada9x, and real-time POSIX (P1003.4A, P1003.4B, and P1003.21)
- adopted by NASA (Space Station Freedom), European Space Agency (ESA), and by a variety of defense contractors
- licensed training services to commercial vendors
- specified usage in handbook published commercially

Maturing RMA

Rate Monotonic Analysis - a technique for assuring that timing constraints are met



Maturing the Process - 1

Process is a commonly accepted component of software engineering practice today as evidenced by:

Software Capability Evaluations (SCE)

- 690 evaluators trained to date
- over 50 major acquisition programs have used SCE

Software Process Assessments (SPA)

- the SEI has data on 173 assessments across 750 projects since 1987, but many more have been held to date
- 10 licensed vendors

Maturing the Process - 2

Quality and productivity
- growing evidence of improved quality and productivity after software process improvement activities

Schedule improvement
- evidence coming in regarding schedule improvements after software process improvement activities

SEPG National Meeting
- in 5 years has grown from 46 to 520 in attendance

SPINS
- 7 exist today with over 20 projected in 1995

Maturing the Process - 3

Documents distributed
- over 10,000 copies of "A Method for Assessing the Software Engineering Capability of Contractors"
- over 3,000 copies of CMM 1.0 (8/91)
- over 2,000 copies of CMM 1.1 (3/93)
- copyright release letter for non-profit use, anonymous FTP

Advisory boards
- 5, with 83 members from industry, government, and academia

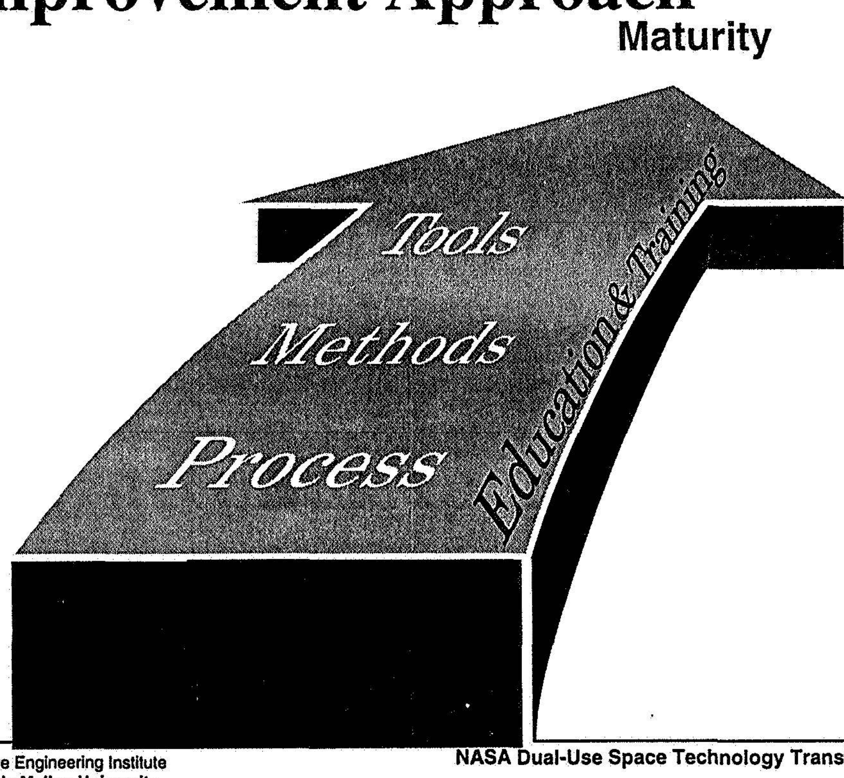
Industry
- increasing funding for software process improvement
- contribution to SEI via technical collaboration agreements, and resident affiliates are increasing

Improving the State of the Practice

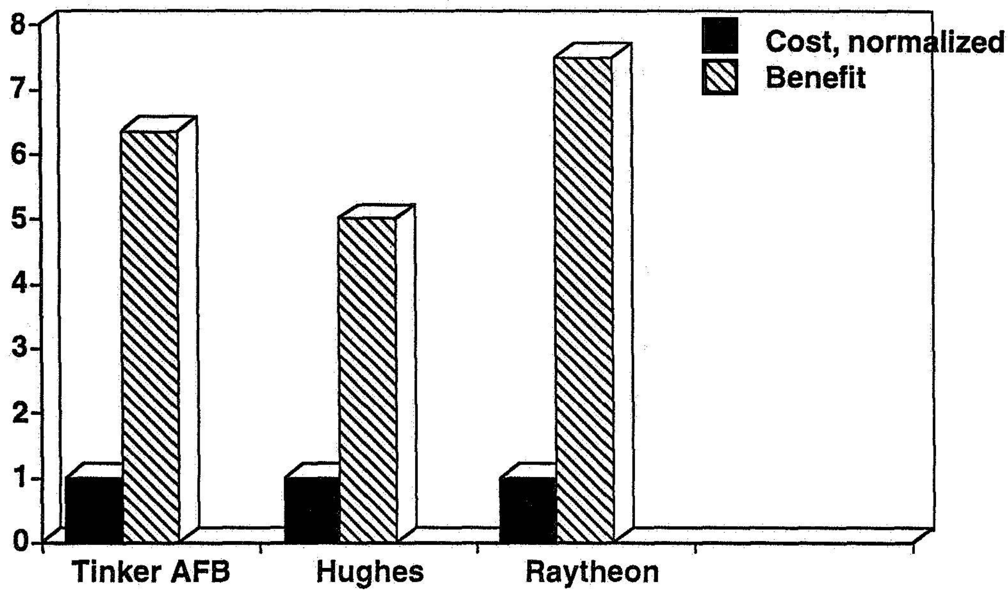
The SEI improves software practice by:

- **being actively involved with customers**
- **helping organizations build their own capacity for continuous improvement**
- **developing a structured approach to improvement**
- **transferring our expertise to the community**

Improvement Approach

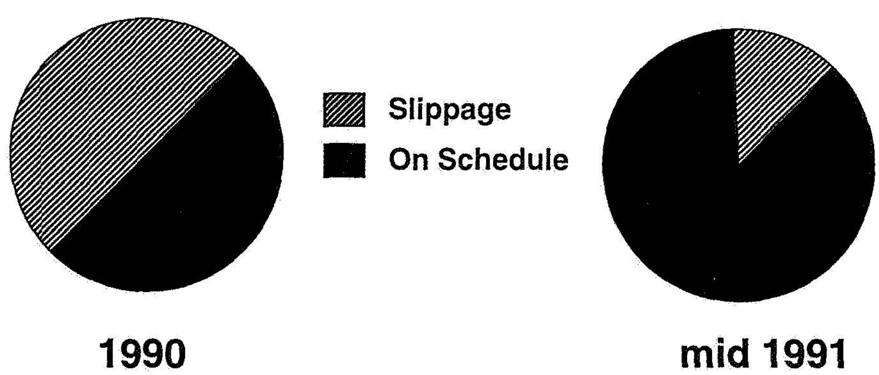


Benefits of Process Improvement



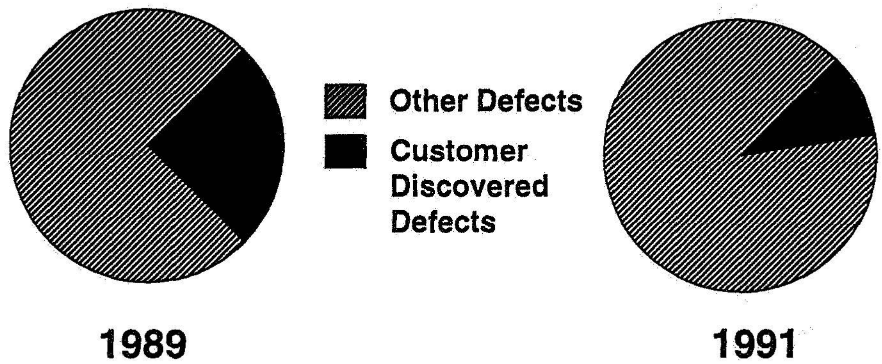
Schlumberger Benefits -1

Project Tracking and Oversight



Schlumberger Benefits -2

Software QA Improvements



Source: ICSE Proceedings, IEEE 1993, pp. 212-219

Maturing the Profession

More than 100 instructors have been trained to teach SEI-developed courses.

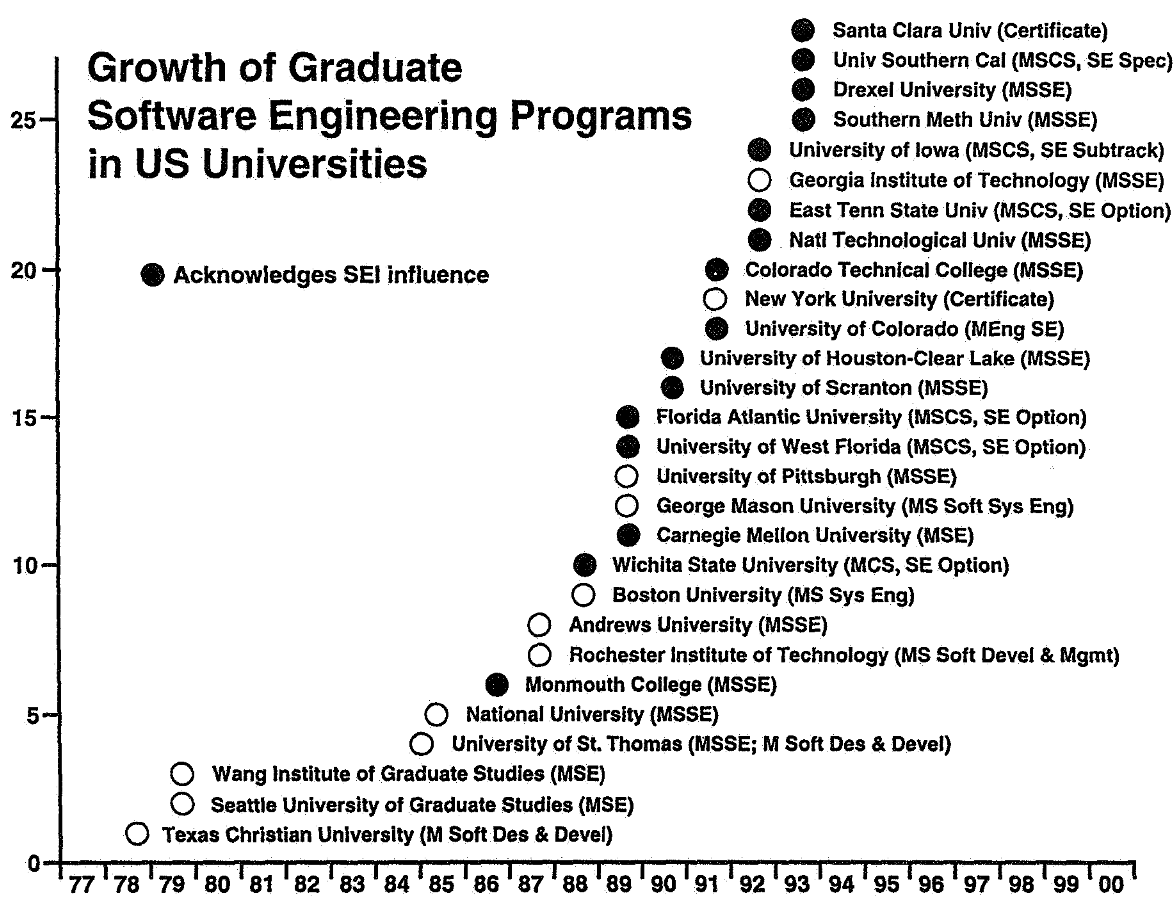
SEI educational materials have been used by more than 300 colleges to deliver over 500 software engineering courses.

SEI courses are broadcast via satellite by the National Technological University (NTU).

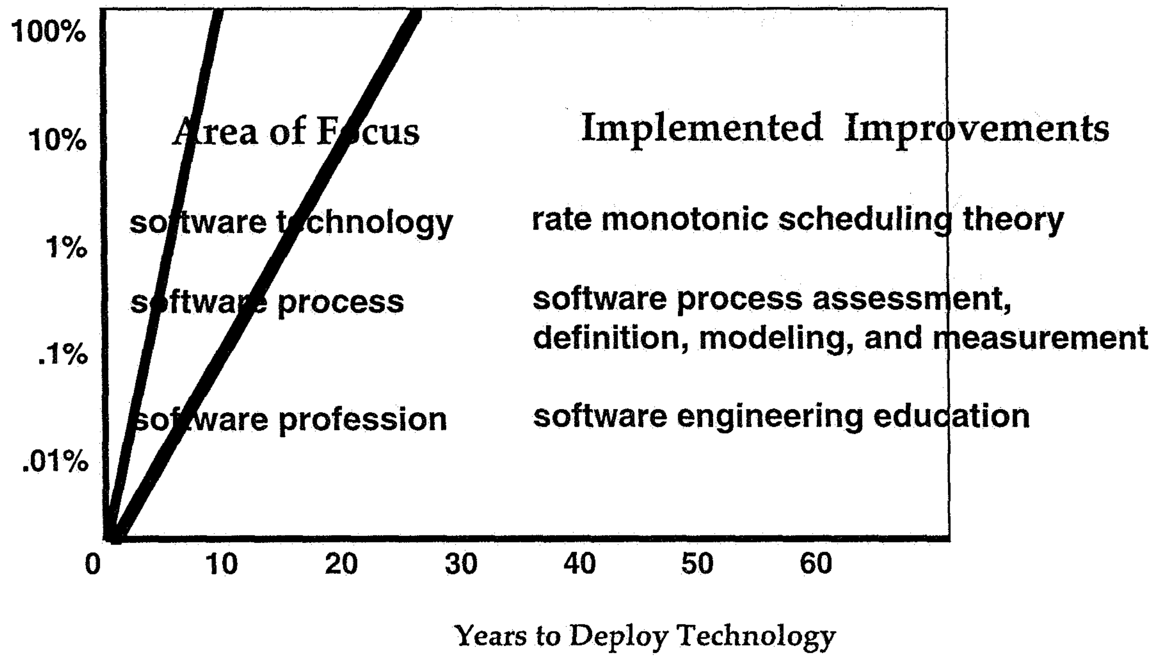
Software Engineering Education

SEI software engineering education activities reach:

Directly	Indirectly
executives	their organizations
instructors/trainers	practitioners
faculty	students and practitioners



Squeezing 20 Years into an 8-Year Time Capsule



Contact Information

Bill Hefley
Technology Transition Initiatives
Software Engineering Institute
Carnegie Mellon University
5000 Forbes Ave.
Pittsburgh, PA 15213-3890

Phone: (412) 268-7793
Fax: (412) 268-5758

E-mail: weh@sei.cmu.edu

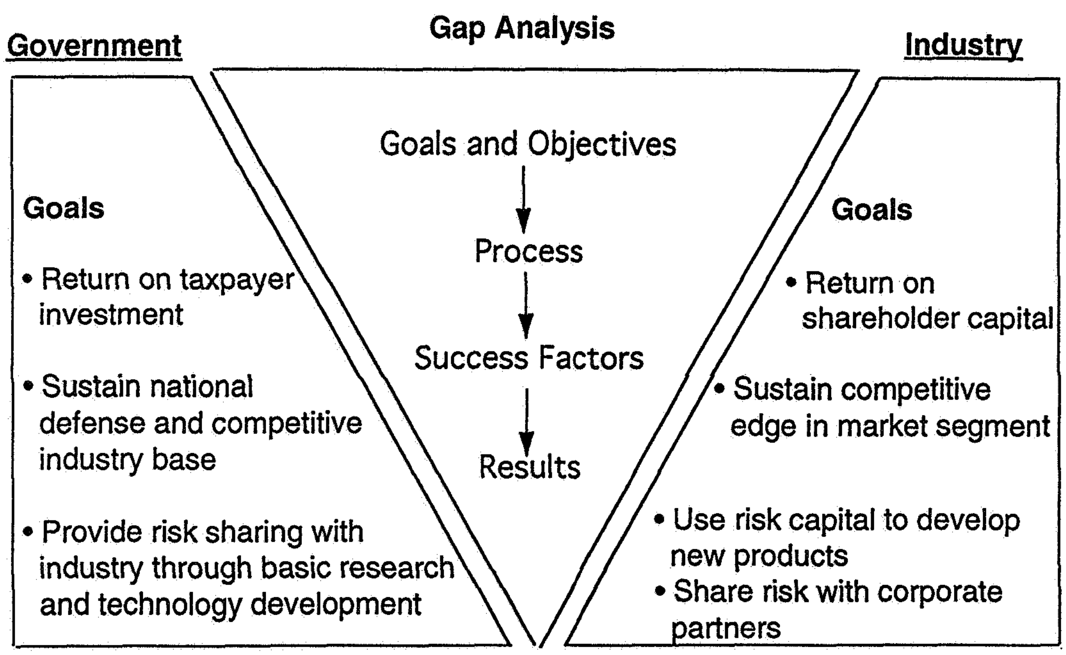
The Meeting of Needs:

Success Factors in Government/Industry Technology Transfer

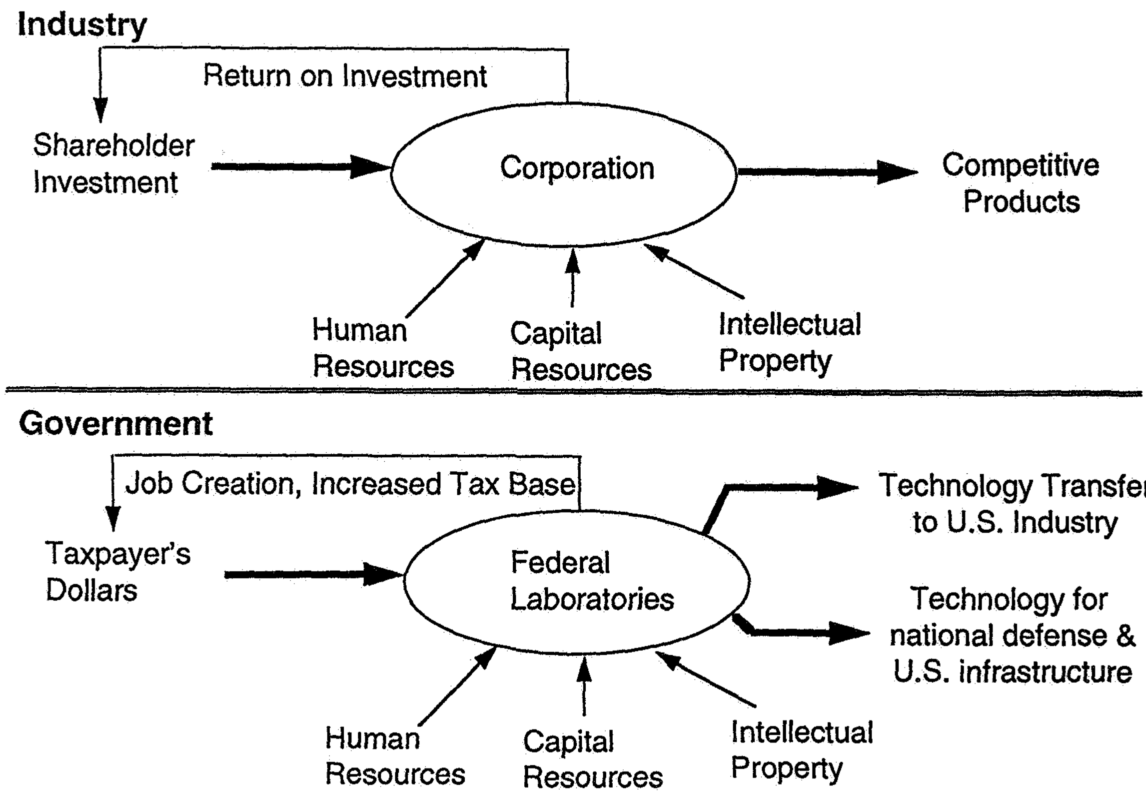
Jim Havelka
KPMG Peat Marwick
Space and High Technology Group
(713) 280-0460

February 2, 1994

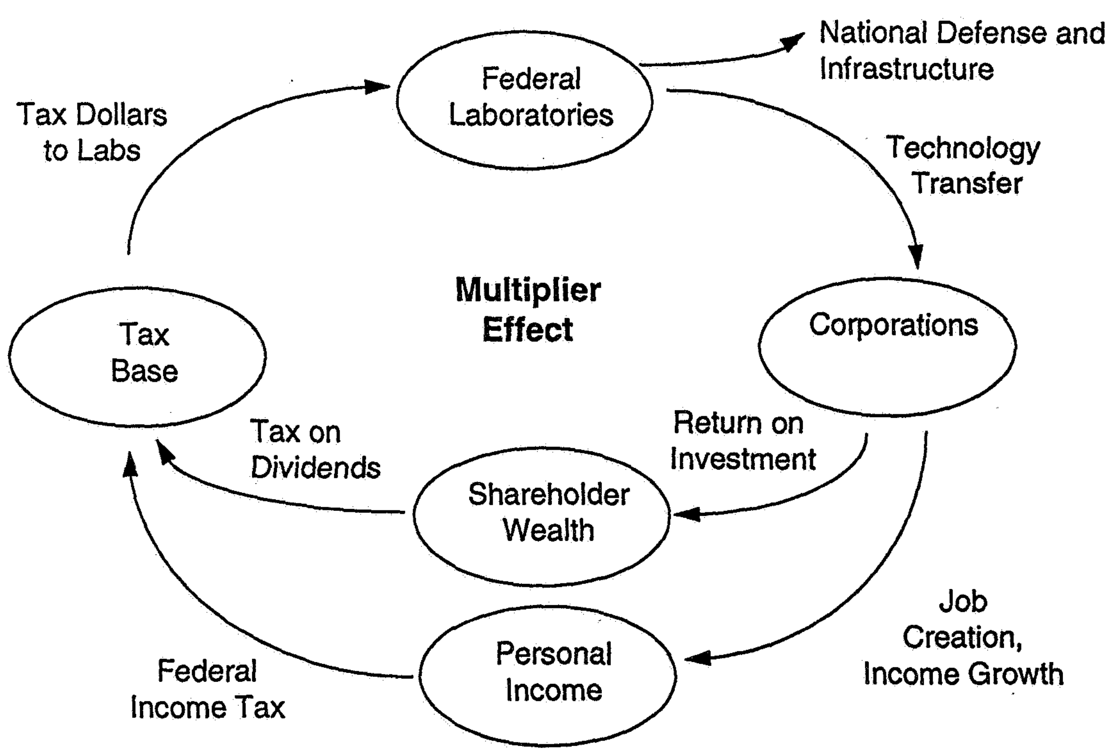
Technology Transfer Needs of Government/Industry



Processes for Government/Industry

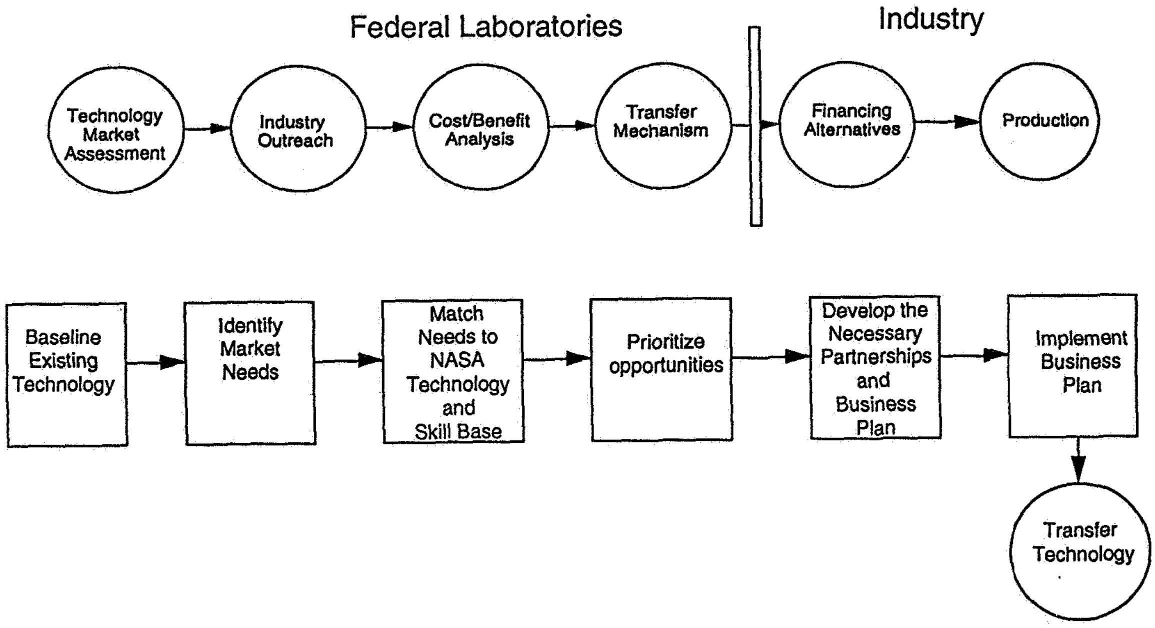


Integrated Government/Industry Model



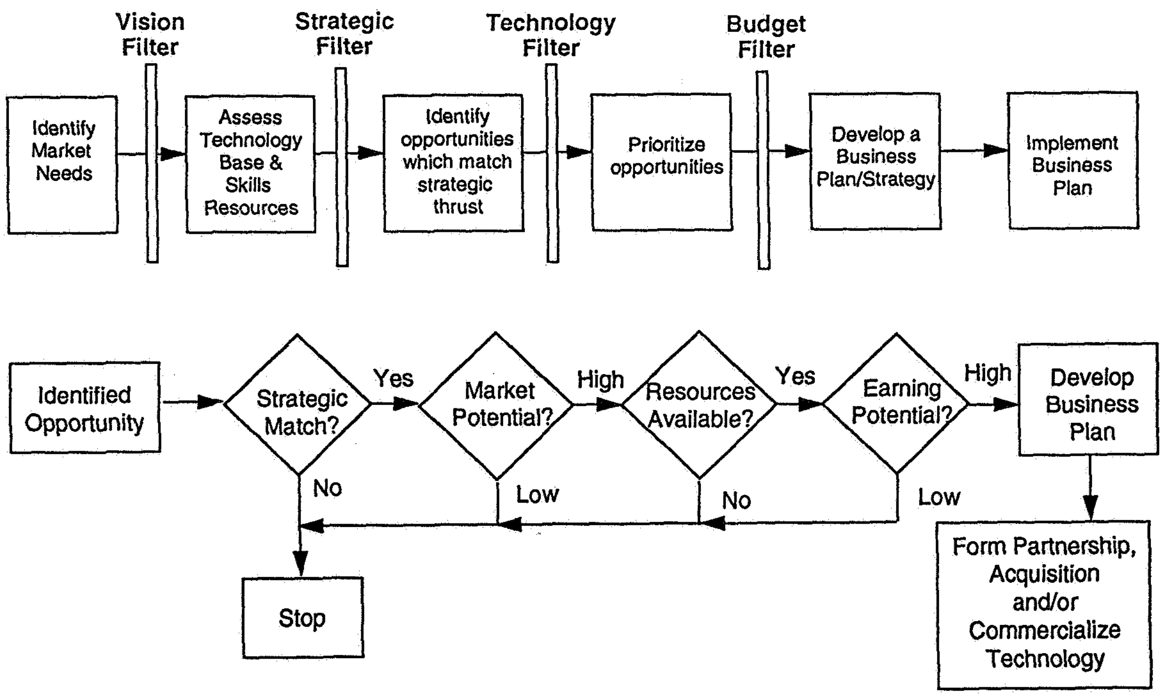
Federal Laboratory Technology Transfer Process

Decision Models

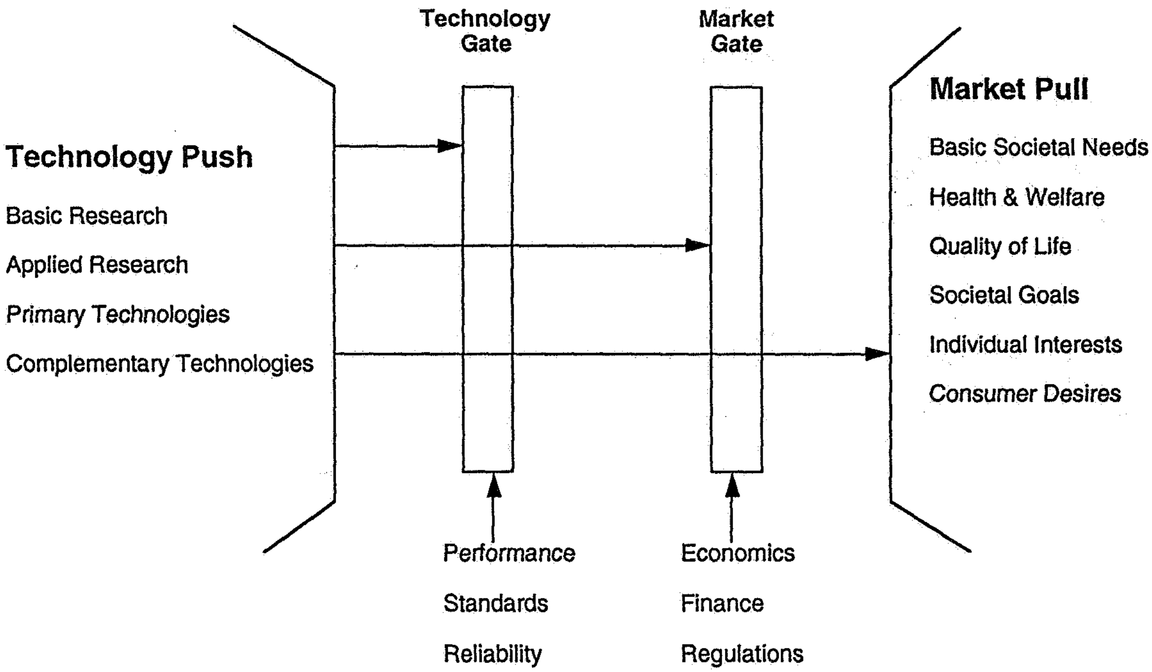


Industry Assessment of a Technology Transfer Opportunity

Decision Models



Mapping Technology to Commercial Products



Key Similarities between Government and Industry

- Leverage capital to benefit the capital provider and achieve organizational goals
  - taxpayers - shareholders
- Technology, human resources, and physical resources are key organizational assets
- Strategic use of limited resources to achieve organization objectives
- Each must respond to the needs of the external environment
  - Congress - shareholders
  - U.S. Industry - market forces

Key Differences between Government and Industry

Federal Laboratories	Industry
• Federal laboratories maximizes technology development and job creation	• Industry maximizes shareholder wealth
• Federal laboratories seek to create broad transfer of technology	• Industry seeks to create monopolistic situations to maximize profits
• Federal laboratories primarily develop basic research and technology prototypes	• Industry utilizes technology to develop commercial applications

Key Success Factors for Technology Transfer

Federal Laboratories	Industry
• Results-oriented research and technology development	• Integrate technology with human and physical resources to support a successful commercial product
• Alliance with industry organizations	• Effective product positioning and development of intellectual property
• Entrepreneurial culture	• Entrepreneurial culture
• Establish a business/marketing resource which can link technology and business together	• Focusing on market requirements, product life cycle and core strengths
• Utilize Technology Push and Market Pull Processes	• Developing an infrastructure and cost structure to support a commercial products

Summary

- Technology transfer can be an efficient generator of national industry competitiveness and wealth creation, if the needs of Government and Industry are both understood and meet.

- A successful technology transfer takes into account the Goals, Processes, and Success Factors of both transfer parties: the Federal Lab and licensing corporation.

UNIQUE STRATEGIES FOR TECHNICAL INFORMATION MANAGEMENT AT JOHNSON SPACE CENTER

Vijay Krishen
Documentation Data Management Systems, Inc.
4127 Long Grove Drive, Seakbrook, TX. 77586

Abstract

In addition to the current NASA manned programs, the maturation of Space Station and the introduction of the Space Exploration programs are anticipated to add substantially to the number and variety of data and documentation at NASA Johnson Space Center (JSC). This growth in the next decade has been estimated at five to ten fold compared to the current numbers. There will be an increased requirement for the tracking and currency of space program data and documents with National pressures to realize economic benefits from the research and technological developments of space programs. From a global perspective the demand for NASA's technical data and documentation is anticipated to increase at local, national, and international levels. The primary users will be government, industry, and academia. In our present national strategy, NASA's research and technology will assume a great role in the revitalization of the economy and gaining international competitiveness. Thus, greater demand will be placed on NASA's data and documentation resources. In this paper the strategies and procedures developed by DDMS, Inc., to accommodate the present and future information utilization needs are presented. The DDMS, Inc., strategies and procedures rely on understanding user requirements, library management issues, and technological applications for acquiring, searching, storing, and retrieving specific information accurately and quickly. The proposed approach responds to changing customer requirements and product deliveries. The unique features of the proposed strategy include: (1) To establish customer driven data and documentation management through an innovative and unique methods to identify needs and requirements. (2) To implement a structured process which responds to user needs, aimed at minimizing costs and maximizing services, resulting in increased productivity. (3) To provide a process of standardization of services and procedures. This standardization is the central theme of the strategic approach. It will allow Division level Data and Documentation Libraries (DDLs) to function independently and optimize efficiency at the Directorate level. This process also facilitates interconnectivity between Division level DDLs and makes them transparent to the users, (4) To implement the process of "cost savings", and at the same time the objective is to gain substantial improvement in the organization, categorization, and preservation of JSC-generated data and documentation, and (5) To find, locate, retrace, restore, and preserve the Center-generated crucial scientific and technical information that has been and is being provided by the engineers and scientists of JSC. This is important to the preservation of "lessons learned". Preliminary estimates of the possible cost savings which will result from the implementation of this process will also be discussed in this paper.

---

1 This strategy was developed by the author while working for Lockheed Engineering and Sciences Company and Futron Corporation in support of NASA JSC.

I. Introduction

Access to data and documentation generated by NASA Johnson Space Center (JSC) programs is increasing rapidly; however, efficient access to the information is not increasing proportionally. This effect gives rise to the personal office library, which is common throughout (JSC) and contractor community. Such a library is highly focused, not updated properly, haphazard, and consumes crucial office space and professional time. These limitations are commonly accepted as a matter of convenience and necessity. To improve this condition each Division/Office of JSC has embarked on a program to develop a Division/Department level Data and Documentation Library (DDL). The goals are to reduce dependence on office libraries and to deliver an easily accessible resource whose total value offsets the office style convenience.

At the present time, there are more than 45 DDLs, in addition to the JSC Scientific and Technical Information Center (STIC) located in building 45. Out of these, more than 17 are outside JSC, establihsed by support contractors and more than 28 are in throughout JSC Divisions/Offices. These DDLs contain information related to programs/projects for which a Department/Division/Office is responsible. The types of information residing in these DDL's includes: paper documents, video, audio, photos, magnetic tapes, CDs/CD-ROM Systems, optical disks, microfilm, blueprints, drawings, slides, viewgraphs. Books, journals, periodicals, and international publications are contained in the STIC and not generally in the DDLs.

Tables I and II provide approximate numbers of of video recordings, photographic images, and printed materials at only one of the distribution sites at JSC. The retaped and reprinted data is estimated for the 45 DDL's. In addition to this estimate there are closet and office libraries containing valuable information. On the basis of an average of 25K documents per DDL, the overall estimate of documents in and around JSC is 112.5 M, assuming that each document has only 100 copies distributed throughout these libraries. The personal distribution copies are not included in this estimation.

Table I. Current scope of data and documentation

APPROXIMATE NUMBERS	
CURRENT SCOPE	**YEARLY INCREASES**
• **VIDEO**	
- 4,750 Hours Recorded	- 2,500 Hours Mission Recorded
- 4,000 Hours of Retaped Data	- 700 Hours of Retaped Data
• **PHOTOGRAPHIC IMAGES**	
- 1 Million Images	- 30,000 Images
- 2 Million Reprints	- 20,000 Images
• **PRINTED PUBLICATIONS/MATERIALS**	
	- 22,800 Documents
	- 112 Million Impressions
	- More than 200 Million Xerox copies

* Audio Data Estimates Not Available

Table II. Estimated contents of DDLs

APPROXIMATE NUMBERS

- **EXAMPLES OF CURRENT DDL CONTENT**
  - **JSC Mission Operations Directorate - 83 K**
  - **JSC Engineering Directorate - 79 K**
  - **JSC History Library - 696 K**
  - **Loral's 5 area Documentation Centers - 210 K**
- **ESTIMATED CURRENT DATA AND DOCUMENT CONTENT OF 45 DDL's**
  - **Average Content 25 K**
  - **Total Content 1.125 M**
- **ESTIMATED DUPLICATION/MULTIPLICATION RATE**
  - **Average 100**
- **TOTAL CONTENT IN OFFICES & DDL's 112.5 M**

Several support contractors operate the DDLs. As a result of organizational and contractor differences, there are diverse management and operating procedures among the DDL functions. Such diversity has introduced non-standard storage and retrieval features, confusing data and document structuring, cataloging, distribution, record keeping, and limited on-line viewing of the DDL contents. As a result, the quality of services has suffered, and costs are unreasonably high. Figure 1 presents an analysis of the current situation in regard to information status and considerations. The main feature of this diagram is the CONSEQUENCE of the situation.

In addition to the current NASA manned programs of Space Station, the introduction of the space exploration programs are anticipated to add substantially to the number and variety of data and documentation at JSC. This growth is being estimated at seven to ten fold compared to current situation. With the national presures there is an increased requirement to realize economic benefit from the research and technological developments of space programs. Moreover, the accountability for data and documents will continue to grow as the space technologies find their uses in our commercial sector. We believe that the investing in information infrastructure or highways, as referred to in the Presidents technology policy, will provide greater demand for space information.

Unique strategies and procedures relying on understanding of user requirements, library management issues, and technological applications for acquiring, searching, storing, and retrieving specific information accurately and quickly will be needed to provide services and achieve significant cost savings. Our vision is to have all this information collected, organized, categorized according programs and projects, and networked within Divisions/Departments/Offices, Directorates, and JSC, and contractor community. To achieve this, we need to establish a strategy that defines an organization's directions and requirements.

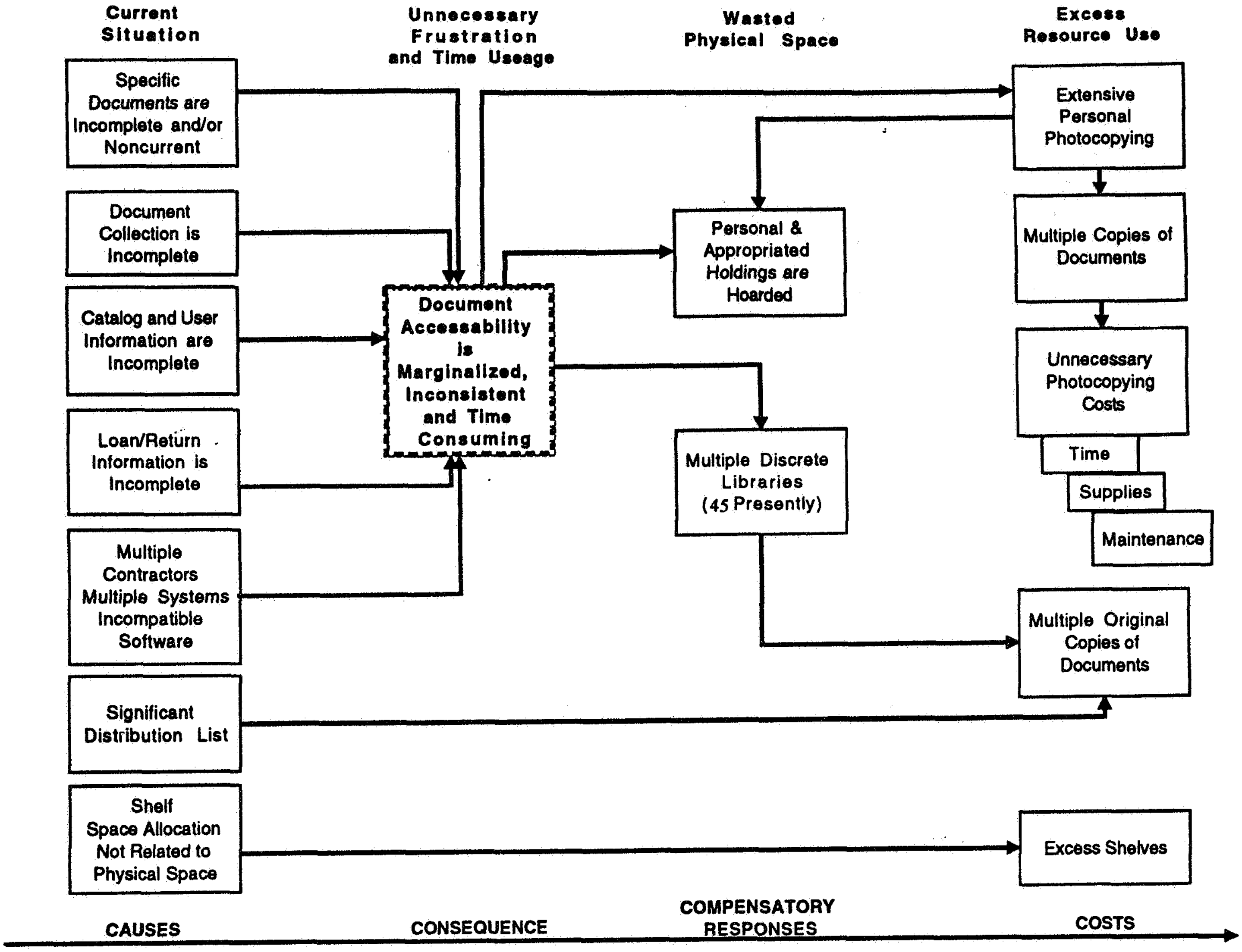


Figure 1. Data and Document status, consideration and consequence

II. Strategy For Information Management

The strategy for managing information through the DDLs will consist of the following key components:

(1) Development of DDL for each Division/Department which is responsive to the needs and requirements of users. Each DDL should contain data and documents for the projects and programs for which the Division/Department is responsible. Each DDL should be physically located such that it is easily accessible to the Division/Department users. This will help to reduce the need for "personal" office libraries, cut down duplication/multiplication of information, and it will also reduce the access distribution.

(2) Implementation of flexible management and utilization approaches which uses existing infrastructure and assets. The flexibility is needed to reflect the changing needs and requirements for information. In particular, changes in projects, programs, and technology should be accommodated in the process. Use of existing facilities and assets is a part of cost savings goal.

(3) Development and implementation of a standardized process in the management and operation of the DDLs. The elements of the process include acquisition, organization, distribution, categorization, retrieval, database design, interconnectivity, storage, and physical structure of the DDLs.

(4) Development and implementation of performance measurement procedures. This includes customer feedback, frequent surveys, and analysis of the data to provide a "customer driven" process (Figure 2).

(5) Incorporating a deliberate procedure to ensure improved services and/or reduced costs for each activity related to the development and operation of DDLs. This process is captured in Figure 3.

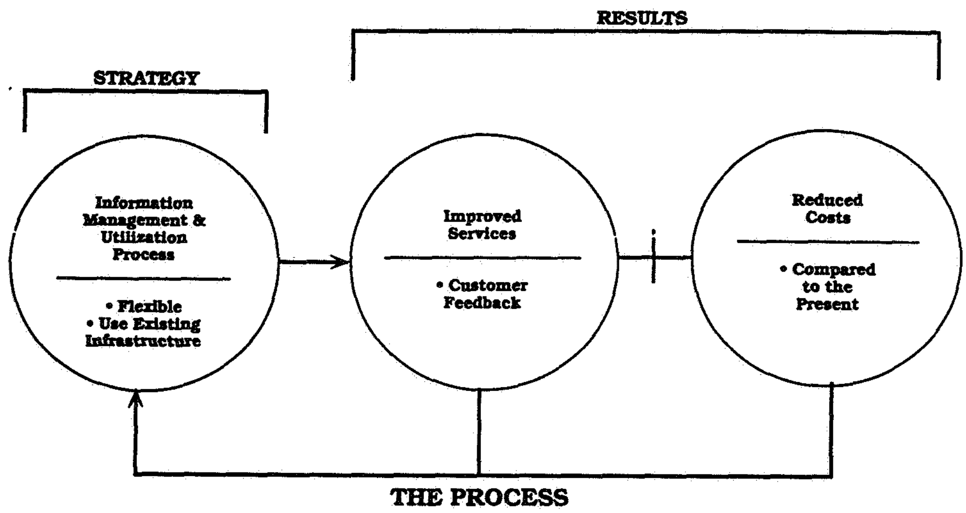


Figure 2. Customer driven strategic process

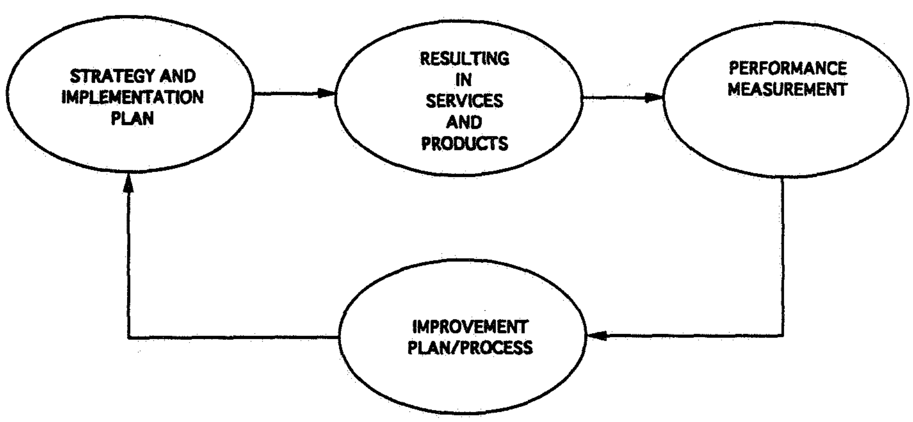


Figure 3. Process for continuous improvement

The key to a successful strategy is the formulation of present and future data and documentation needs and requirements for each Division/Department at JSC and within the contractor community. A structured approach to evolve a DDL management system which will satisfy these requirements then forms the core of the strategy. A crucial part of the strategy is to introduce appropriate procedures and technology which will enhance the services and reduce the costs. The result of improved services will be increased productivity from engineers, scientists, and managers. The process adopted should use existing infrastructure and be flexible. The result of the strategy should be reduced costs and/or improved services.

Information management and utilization process includes knowledge capture, transfer, and change/update of data and documentation. Additional aspects include, standardization, search/research, acquisition, updating, database design, networking, structured physical layout of the DDL, and streamlining of reception/distribution of data and documentation

The current assessment of the DDLs leads to the areas of improvement which include:

(1) Identification of needs and requirements at Division level
(2) Overall configuration control
(3) Intercontractor communication for standardization and configuration control
(4) Interface and integration of multiple contractor software design implementation at Division level
(5) Efficient physical storage and retrieval
(6) Standardized computer logging and retrieval
(7) Streamlined approach to distribution and duplication/multiplication of data and documentation
(8) Accurate records of loan and hold
(9) Archival and preservation of critical information
(10) Division, Directorate, and contractor level interconnectivity
(11) Provide process to incorporate new information such as Space Station, new and advanced programs
(12) Provide periodic performance assessment/measurement

A part of the strategy is to gain control over the current situation by implementing enhancement procedures. Several such steps are shown in Figure 4.
The progress through the improvement process adds successive levels of value through: 1) the increased utilization of the libarial system; 2) more efficient use of space; 3) reducing storage requirements for data and documents; 4) using relevant procedures to expedite the time required to log, store, and retrieve data and documents; 5) reducing photocopying requirements; 6) capitalizing on configuration control to standardize the operating features among DDLs; 7) establishing a systematic process for managing revisions of information for each DDL; 8) removing obsolete documents and maintaining only required number of copies. In parallel with the near-term improvements, an activity leading toward refining the advantages and disadvantages of keeping the existing information system posture; pursuing a standard operating system for all 45 libraries which relies on unifying current technologies within JSC; or modernizing with outside data management processing tools. A considerable number of these modern tools have proven features for reliably, accurately, and rapidly retrieving and integrating a great volume of topical information represented by a wide variety of media. The gateway to the formulation of options is a cost/benefit analysis and assessment.

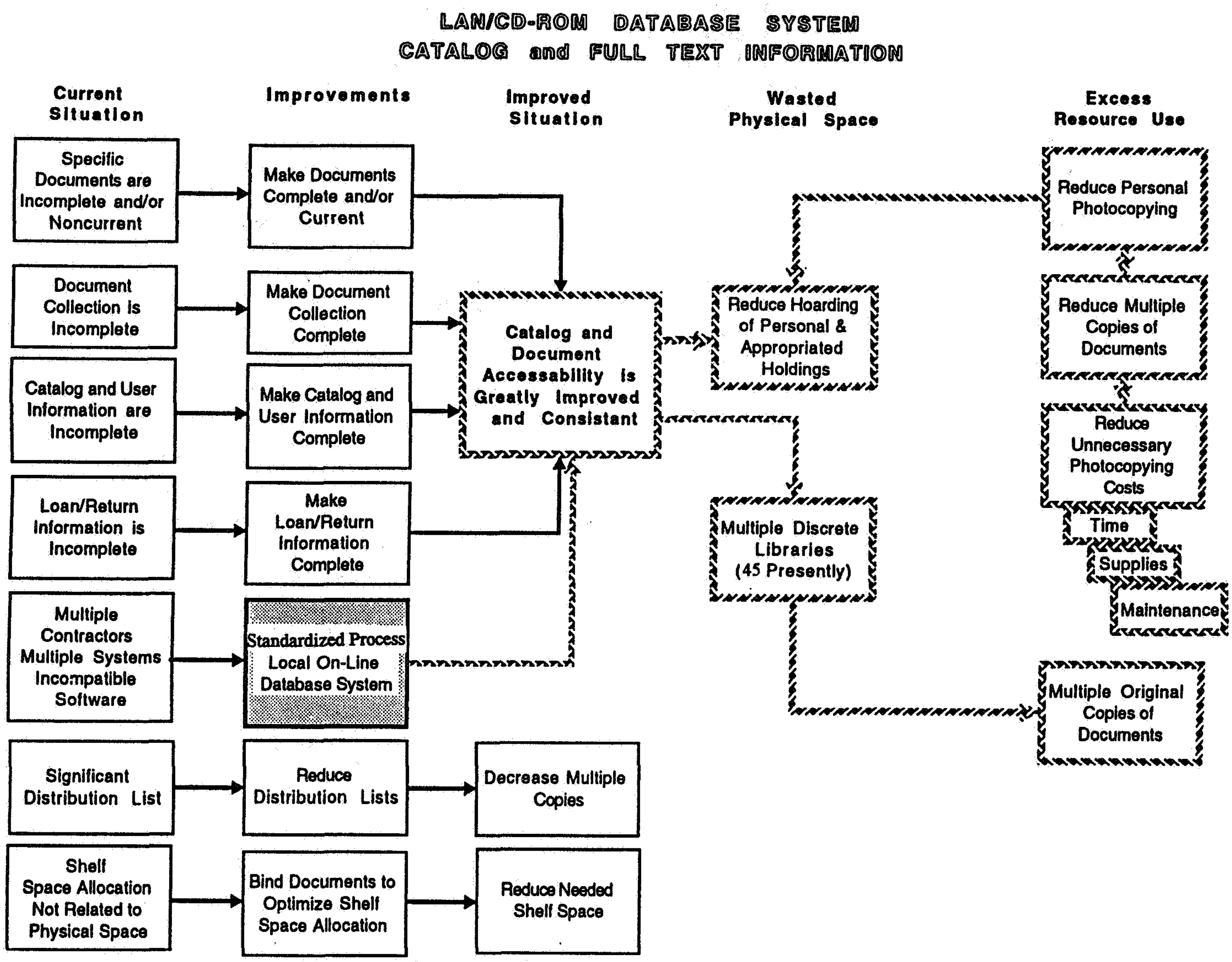


Figure 4. Information accessibility

III. Implementation Strategy

The plan is to implement the project in the following manner: Data and documents for manned flights are to be categorized under specific projects such as National Space Transportation System (NSTS), Space Transportation System (STS), Assured Crew Return Vehicle (ACRV), Space Station (SS), Lunar Outpost/Mars Missions (LO/MM), and Research and Technology Development (R&TD). Subdivisions of these categories will reflect breakdown in terms of the disciplines such as power, propulsion, tracking, communications, data systems, crew systems, thermal systems, automation, robotics, aeronautics, navigation, control, software, systems engineering and others. We believe that the overall flow of the relational and hierarchial organizational process will be cost effective. The benefits will be obtained by the following procedures; 1) Make documents complete or current by incorporating changes, and revising and updating the information. 2) Make information collection complete by finding /locating program/project related documentation required by users, and order it from various distributors. 3) Make categorization complete by analysis of programs and projects and needed information, categorize it by the subjects and Division functions, then sort out the information for inputting process.

Phased approach has been used to implement this strategy. It has been devided in four phases, I through phase IV. This approach can be applied to a single Division/ Department, however, the phases and tasks can be extended to several Divisions/ Departments, and then to the Directorates. The ultimate goal is to extend this strategy to entire JSC and to support contractors to realize center-wide benefits. This phased approach is employed to lead to standardized procedures for the development and operation of DDLs. These phases are described briefly in the following:

Phase I

In this phase, a procedure is developed to conduct a study/survey of needs/ requirements of each Division/Department. The results of the survey (s) will be prepared for each Department/Division. A report identifying steps, procedures, and processess to alleviate problems and respond to customer needs will be generated. Specifically, the following areas need to be covered: 1) survey current capabilities (strengths and weaknesses); 2) survey user needs; 3) formulate requirements; 4) develop trade reports assessing cost/benefit/analysis and options; 5) identify impacts and benefits of savings that can be obtained from specific approaches. The forecast of information needs/requirements allows; a) to scope information infrastructure b) to formulate Division/Department level strategy, and c) to have realistic estimate of resources needed to develop and operate the DDL. Figure 5 shows the strategy for the identification of the requirements.

Phase II

In this phase, the option of a standardized database design program and its inter-connectivity within the Division/Department is researched. This option includes modification of existing database or incorporating new or advanced software system, such as CD-ROM or scanning-based data input. The pros and cons of each option are described and documented. Figure 6 depicts the features of database design which will enable interconnectivity and standardization.

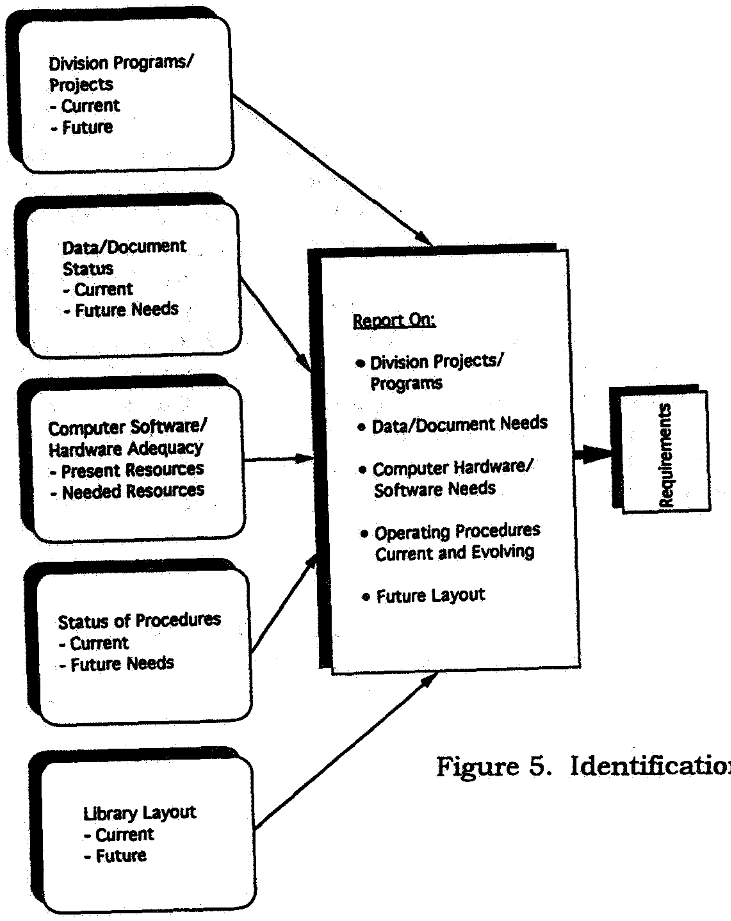


Figure 5. Identification of needs/requirements at Division level

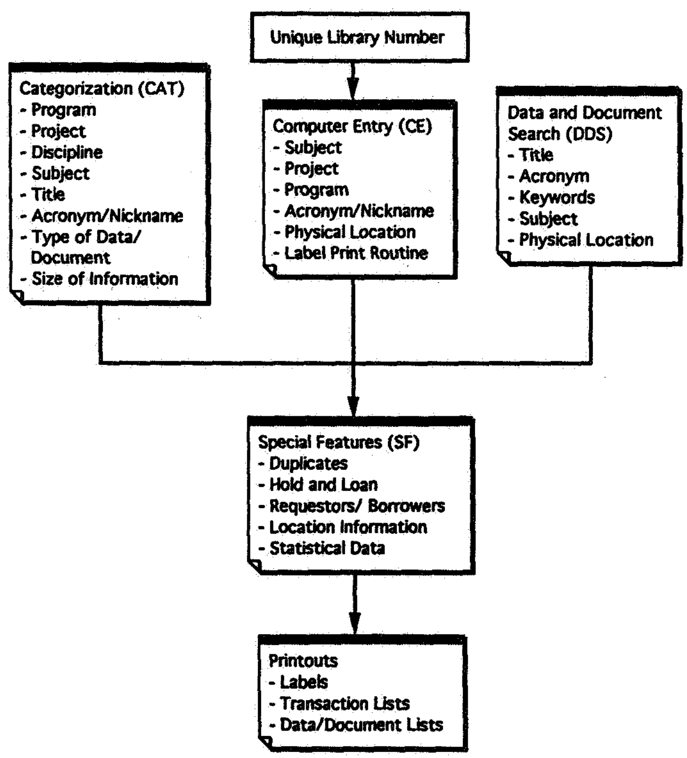


Figure 6. Database program-menu design (DBP-MD) and fields

Phase III

This phase contains a series of tasks which end with an operational DDL system. The phase is analogous to a conventional Design, Development, Test, and Evaluation (DDT&E) activity. It follows a set of milestones from a Preliminary Design Review (PDR) through a customer acceptance review. Verification/acceptance tests will demonstrate the success of implementing the approved system recommendations. The DDL will be put under formal configuration control at the completing of the Critical Design Review (CDR). The configuration management file will contain hierarchical information extending from Phase I and Phase II requirements (approved recommendations) to the system design standards and operating policies. Specifically the following tasks are to be accomplished: 1) Implement the database standardized menu design and software for interconnectivity; 2) verify and check data/documents to capture/retrieve information; 3) plan, develop, design, structure, maintain, and operate Division/Department data and documentation library (DDL) management system; 4) Provide ongoing support to maintain currency, and refine end-to-end procedures to follow.

Phase IV

A standardized process has already been developed, implemented, and established at the Division/Department level during previous phases. In this phase Program Office/ Directorate level implementation of DDL interconnectivity will then be extended to entire JSC and contractor community (Figure 7).

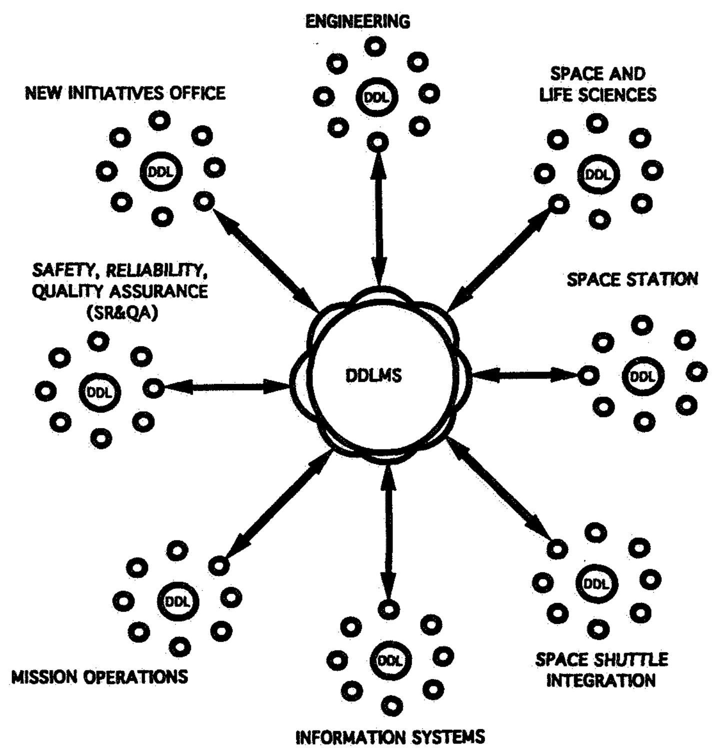


Figure 7. Inter-Directorate DDL implementation strategy

A key task in this phase is to implement a standardized document request, receive and distribution method. Figure 8 and 9 shows a coordinated efficient process to handle incoming and outgoing data and document distribution diagram. This process will include providing a list of possible sources for documents, change requirements, and new revisions to each DDL. These DDLs will then process their requests electronically or manually. A coordinated information requesting, receiving, and distributing system will be designed according to programs/projects of each Division/Department and implemented. The result will be to provide improved customer service while at the same time reduce multiplication of documents and associated handling efforts (Figure 8 and 9).

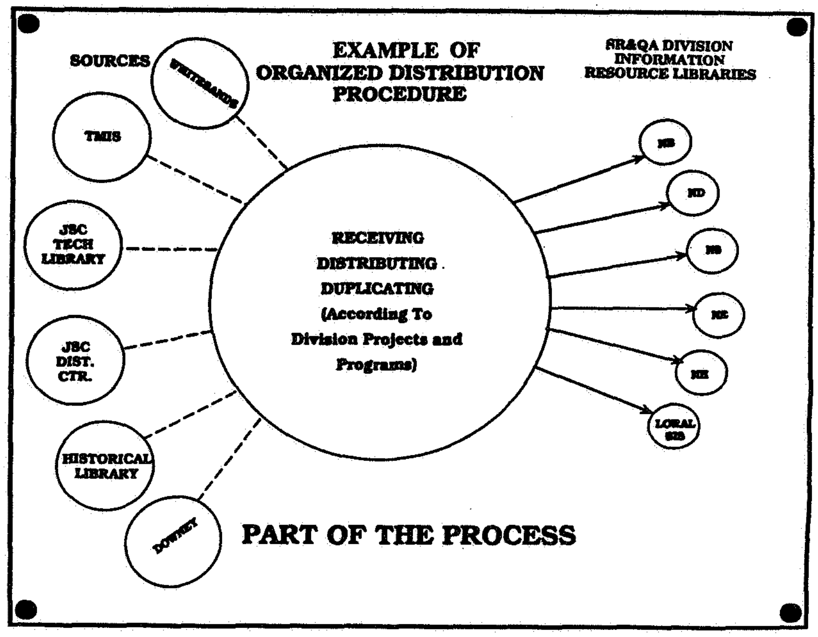


Figure 8. Organized/categorized distribution procedure

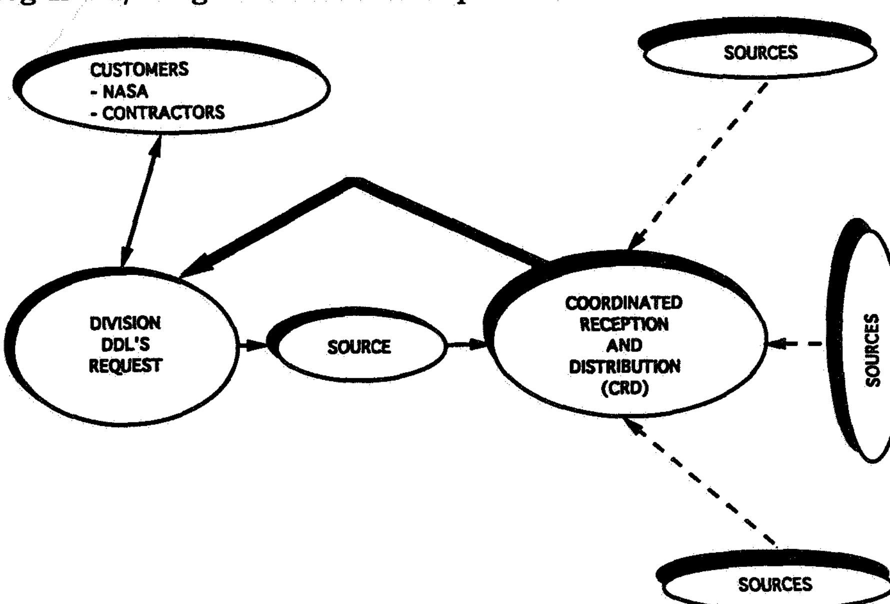


Figure 9. Coordinated process of reception and distribution

Specifically, this phase also includes the following areas: 1) design efficient networking approach at Directorate level; 2) link Division/Department level DDLs with rest of the Divisions and contractor DDLs; 3) provide NASA JSC and contractor personnel of each Division an easy access to all information, and 4) to find/locate, retrace, restore, and preserve information that has been and is being provided by discipline specialists, software specialists, engineers and scientists of Johnson Space Center (JSC) in the past and at present.

The benefits of the implementation approach described in this section of the paper can be summarized as follows: 1) reduction in resources to upgrade and maintain systems software and documentation; 2) reduction in duplication of documentation; 3) reduction in space to house libraries; 4) improved availability to user; 5) decreased user search and acquisition time; 6) all information in most updated form; 7) traceable history of cost-savings and productivity enhancement; 8) and lessons learned preserved with knowledge capture.

IV. Impacts of Technology Advancements

NASA and many other organizations are developing new technologies to manage information. Electronic capture, transfer, and retrieval of all kinds of information is progressing at a fast pace. In particular, NASA JSC is involved in the following research areas: 1) NASA JSC Electronic Library System (NELS); 2) Cooperating Network Systems (CNS); 3) Hyperman - a Hypermedia Viewing System (HVS); 4) Multimedia Presentation of Text, Audio, and Vedio from large Multimedia Libraries; 5) The Wide Area Information Server (WAIS); 6) Digital Document Storage and Retrieval Wavelet Technology; 7) Artificial Intelligence and Expert Systems, Knowledge Capture, Transfer, Display and Processing; 8) Highspeed and Parallel Information Processing. The early phasing in of this technology is one of the important areas of implemention flexibility for the DDLs. Utilizing new technology options for accessibility, database, cataloging, indexing, display, and retrieval will be constantly evaluated. Cost-effective approaches will then be implemented through trade off studies. The options available are given in Figure 10. The trend will be to develop electronic information sharing architectures for DDLs.

V. Conclusion

The Clinton - Gore technology policy identified technology transfer from federal laboratories as a key factor for the economic revitalization of this Nation. To enable this transfer, NASA JSC has to share information with industries and academic institutions. It is vital that DDLs provide the necessary information to enable technology transfer. The approach to enhancing DDLs is to use disciplined methods to preserve and disseminate data and documentation. Furthermore, the objective is to provide cost-effective and quality controlled process, so that Divisions/Departments have control over their information needs and uses. NASA JSC is an ideal organization to develop and test an organized approach to information management. The overall network will consist of more than 45 DDLs serving more than ten thousand users. Once this interconnectivity and accessibility is established and optimized, the process can then be transferred to the industry to realize commercial benefits.

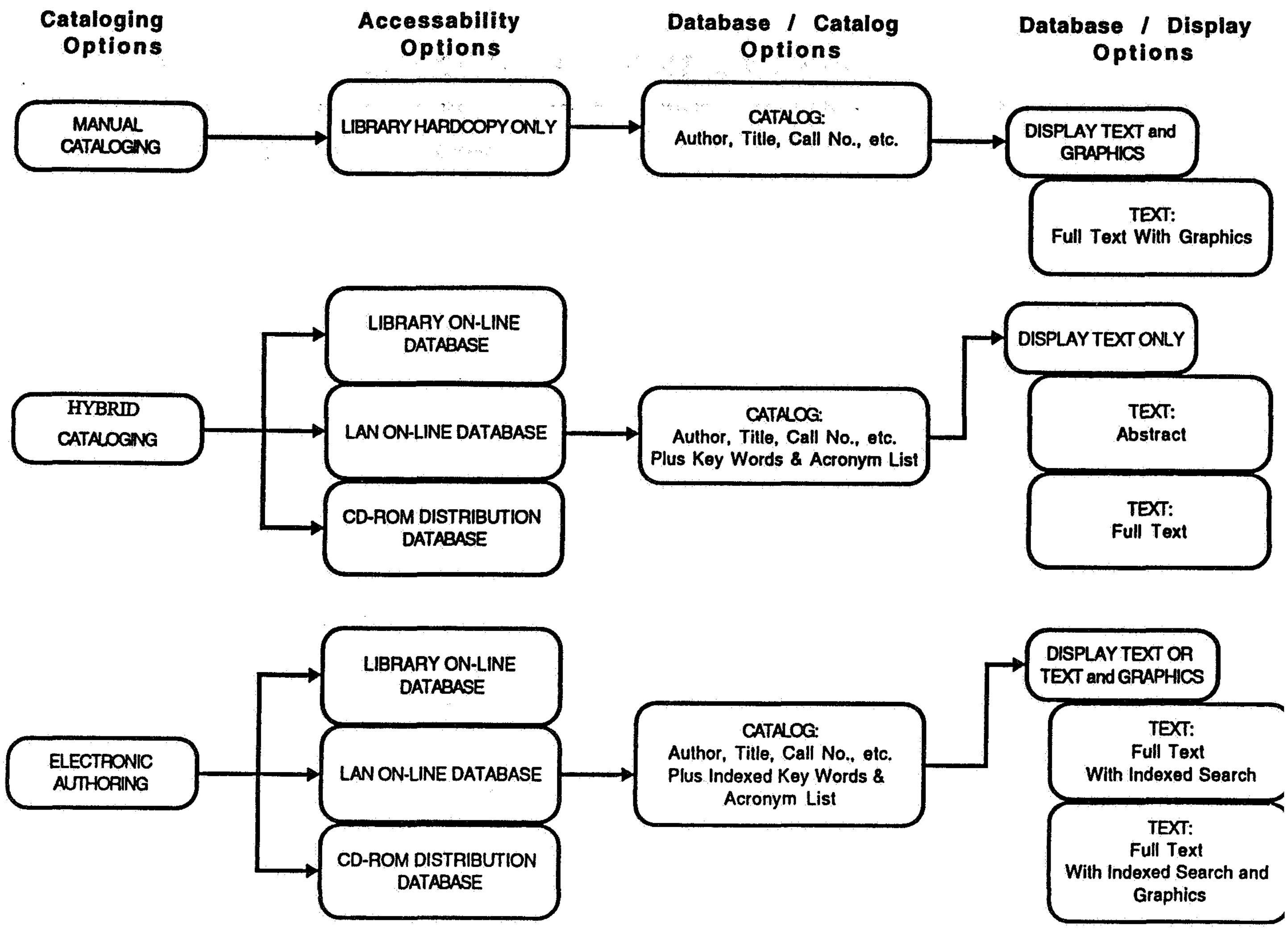


Figure 10. Technology implementation options

VI. Acknowledgments

Many valuable suggestions and comments regarding the process in this paper were made by Mr. Michael Ruiz and Mr. Don Brown. The author is grateful to Dr. Kenneth J. Cox and Mr. Aldo Bordano for their encouragement and support

Dual-Use Space Technology Transfer Conference
NASA/Johnson Space Center, February 2, 1994

THE RETINAL FUNDUSCOPE DEMONSTRATION PROJECT

by

Jim Wilhelm
Assistant Director
Mid-Continent Technology Transfer Center
The Texas A&M University System

Mid-Continent
Technology Transfer Center
A NASA Regional Technology Transfer Center

JAMES P. WILHELM
Assistant Director

8920 Business Park Dr.
Austin, TX 78759-7405

The Texas A&M University System

TEL: 512-346-1444 • FAX: 512-338-1464 • EMAIL: Wilhelm@technology.com

TECHNOLOGY TRANSFER ?????s

1. WHAT'S THE TECHNOLOGY & COMMERCIAL VALUE?

2. WHY IS A DEMONSTRATION NEEDED?

3. HOW DID WE GET SEVEN PARTIES TO WORK TOGETHER?

RETINAL FUNDUSCOPE DEMONSTRATION PROJECT

1. WHAT'S THE TECHNOLOGY?

Life Sciences

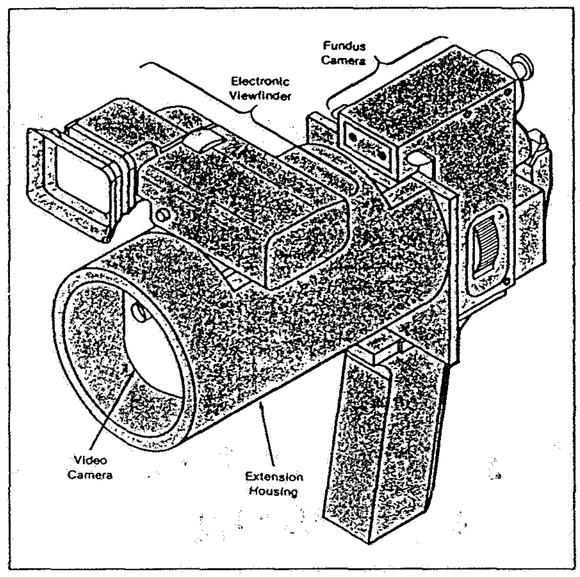
Hardware, Techniques, and Processes

Portable Video/Digital Retinal Funduscope

An inexpensive instrument can be operated with minimal training, under hospital or field conditions.

Lyndon B. Johnson Space Center, Houston, Texas

A lightweight, relatively inexpensive electronic and photographic instrument has been developed for the detection, monitoring, and objective quantification of ocular/systemic disease or physiological alterations of the retina, blood vessels, or other structures in the anterior and posterior chambers of the eye. The instrument can be operated with little training. It can function with a human or animal subject seated, recumbent, inverted, or in almost any other orientation; and in a hospital, laboratory, field, or other environment. The instrument produces video images that can be viewed directly and/or digitized for simultaneous or subsequent analysis. It can also be equipped to produce photographs and/or fitted with adaptors to produce stereoscopic or magnified images of the skin, nose, ear, throat, or mouth to detect lesions or diseases.



The **Portable Funduscope** has a modular design. The extension housing, video camera, and electronic viewfinder can be removed and replaced with a 35-mm film camera. The fundus camera can be equipped with a variety of lenses, prisms, and the like.

The instrument is an assembly of commercially available equipment, some of which has been modified slightly to make it compatible with the other equipment and the overall design. One major component is a portable fundus camera with lenses, filters, and prisms as required for various viewing configurations. The image produced by the fundus camera is either recorded on 35-mm film or sensed by a low-light-level charge-coupled-device (CCD) video camera (see figure). The other major components are an electronic viewfinder, a video monitor for viewing in real time, and a computer that digitizes the video image.

Equipment to stabilize the subject's head is included, but the instrument can be operated without such stabilization. The operator can adjust the focus on the fundus camera while viewing through an eyepiece, using either the 35-mm-film-camera attachment or looking directly at the electronic viewfinder mounted on the CCD camera. When the desired image is obtained, the camera is secured in place. Excitation and barrier filters can be inserted into the fundus camera for fluorescein angiography. The images from the CCD camera can be directly digitized by the computer for storage or transferred via telephone lines, computer networks, or satellite to remote locations.

By providing for the digital analysis of images, the instrument helps physicians to compare sequential images from a given patient to detect subtle disease progressions earlier. In addition, the acquisition of images as digital information facilitates storage, transfer, and manipulation to enhance features of interest. It also enables extensive analyses of images, including quantitative analyses of the diameters of blood vessels and the detection and monitoring of changes in retinas caused by hypertension, diabetes, atherosclerosis, vasculitis, uveitis, macular degeneration, glaucoma, and infections. The instrument can be upgraded easily as advanced sources of light, optical equipment, CCD cameras, computers, and image-analyzing computer programs become available.

This work was done by Gerald R. Taylor of **Johnson Space Center;** *Richard Meehan of the University of Colorado; and Norwood Hunter, Michael Caputo, and C. Robert Gibson of Krug International. For further information, Circle 37 on the TSP Request Card.*

This invention is owned by NASA, and a patent application has been filed. Inquiries concerning nonexclusive or exclusive license for its commercial development should be addressed to the Patent Counsel, Johnson Space Center [see page 16]. Refer to MSC-21675.

1. WHAT'S THE TECHNOLOGY & COMMERCIAL VALUE?

- **Funduscope Aids Detection of Disease** or Physiological Alteration of the Retina, Blood Vessels or Other Structures of the Eye as well as the Skin, Nose, Ear, Throat, and Mouth

- Remote Diagnosis **Allows More Specialists to Participate** Resulting in:
  -Faster and More Complete Diagnosis
  -Fewer Rescheduled Visits
  -Access to Specialists Otherwise Not Available
  -Improved Health Care Delivery

- Remote Consultations will **Reduce Physician Overhead Costs and Travel Time by $132,000,000 Per Year** (A.D. Little, 1992)

- **Continuing Medical Education Faster,** More Effectively and **More Efficiently Saving $100,000,000 Annually** in Conference and Travel Expenses

RETINAL FUNDUSCOPE DEMONSTRATION PROJECT

2. WHY IS A DEMONSTRATION NEEDED?

1. **Field Test** Portable Funduscope by Health Care Providers

2. **Assess** SatelliteTelemedicine **Cost/Benefit** of Using the Portable Funduscope

3. **Assess** Commercial **Market Potential** of Portable Funduscope

4. **Assess Cost/Benefit** of Health Care Staff **Continueing Education** via Satellite Videoconferencing

RETINAL FUNDUSCOPE DEMONSTRATION PROJECT

3. HOW DID WE GET SEVEN PARTIES TO WORK TOGETHER?

- ***NASA Headquarters*** ($350,000): Demonstration Project Funds to Reduce Commercial Risk

- ***Johnson Space Center***: Protocols and Results with NASA Telemedicine Projects

- ***KRUG Life Sciences***: Funduscope Availability & Training ($4,500 in-kind)

- ***Scott & White Memorial Hospital*** ($95,000 & $400,000 in-kind): Funduscope Evaluation for Telemedicine & Health Care Telelearning

- ***James Rumsey Technical Institute*** ($361,800 plus $50,000 in-kind) sponsored by the West Virginia Dept. of Education and the Appalachian Regional Commission: VoTech Skills Training

- ***SpaceTech*** ($85,000 in-kind): Network Integration, Operation & Expansion

- ***MCTTC*** ($40,000 in-kind) : Project Management & Funduscope Market Assessment for Commercialization

3. HOW DID WE GET SEVEN PARTIES TO WORK TOGETHER?

JOINT-SPONSORED DEMONSTRATION AGREEMENT

1. PREAMBLE
2. DEFINITIONS
3. PARTIES & KEY PERSONNEL
4. ATTACHMENTS
5. TERM OF AGREEMENT
6. PURPOSE OF AGREEMENT
7. STATEMENT OF WORK
8 TASKS & DELIVERABLES
9. FUNDING & IN-KIND CONTRIBUTIONS
10. PROJECT MANAGEMENT
11. RIGHTS TO INTELLECTUAL PROPERTY
12. RIGHTS TO PHYSICAL PROPERTY
13. RIGHTS TO PUBLICATION
14. TRANSPONDER TIME ALLOCATION
15. INSURANCE & LIABILITY
16. DISPUTES
17. TERMINATION
18. NASA PROVISIONS
19. EXPANSION OF DEMONSTRATION PROJECT
20. GENERAL PROVISIONS

RETINAL FUNDUSCOPE DEMONSTRATION PROJECT

EXPECTED RESULTS

1994-95: Demo Expansion to Include Additional Medical Centers & Technologies

1995-97: Commercialization of Portable Retinal Funduscope by KRUG & Distribution Partner

1996-98: Commercialization of Portable Funduscope with Attachments for Ear, Nose, Throat, Mouth and Skin

1995-99: Telemedicine Becomes Established Medical Practice

RETINAL FUNDUSCOPE DEMONSTRATION PROJECT

OMIT.

Session T5: TECHNOLOGY TRANSFER—EXAMPLE I

Session Chair: Jack Aldridge

DEPTH IMAGING

FROM

EASTMAN KODAK COMPANY

TOPICS TO BE COVERED

1. Brief look at the technology
2. Current Kodak structure to bring new technology into the market
3. Problems with current structure
4. Approach with Depth Imaging
5. Problems?
6. Ramifications for Technology Transfer
7. Questions

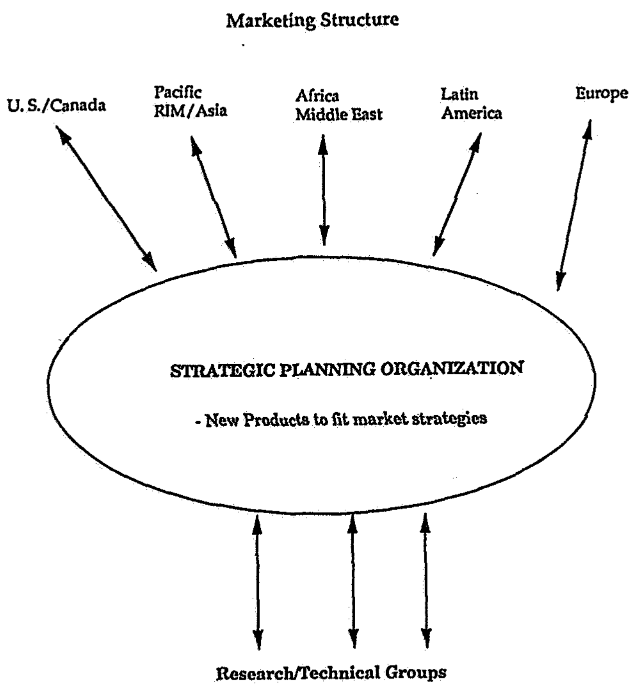
DEPTH IMAGING TECHNOLOGY

What makes it unique?

Breaks through the clutter!!

1. Look around capability - Holography
2. Color found in photography
3. Digital - can:
   - Merge photo and computer generated
   - Totally computer generated
4. Change images

KODAK INNOVATION STRUCTURE



APPROACH WITH DEPTH IMAGING

- Marketing and technology are combined into one organization.
  - Sit together, eat together, work together, and argue directly with one another

- Market place reaction is fed back directly to technical people

- Technical people keep marketing updated on developments.
  Driving key development focus:
  - Larger size
  - Lower cost

- Quicker reaction to market place
  - Images are driven by customers
  - Partnerships that appear advantageous are explored immediately
  - Future technology is tried now

PROBLEMS?

- See me in one year

- Possible burn out

- Maintaining focus

- Keeping team energized and excited

- The critical human element

RAMIFICATIONS FOR TECHNOLOGY TRANSFER

- Is there a champion who can make decisions?
- Can it be focused - technology and human resources?
- Who is accountable?
- Are the key contributors with necessary expertise sitting and working together?
- Is there immediate marketplace feedback?

Session T6: TECHNOLOGY TRANSFER—EXAMPLE II

Session Chair: Kyle Fairchild

Dual Use Space Technology Transfer Conference

Artificial Bone From Space Shuttle Tile

Casey Fox, Ph.D.

BioMedical Enterprises, Inc.

Motivations for Technology Transfer

Government:

- Protect taxpayer's investment in technology
- Augment funding for projects with dual use
- Generate laboratory support through royalties
- Motivate laboratory staff through new technical challenges and royalties

Industry:

- Obtain technology to gain a business edge
- Recruit expertise from Federal Laboratories
- Gain support for risky high return technology

Mechanisms of Interaction Between Industry and Government

License: (14 CFR Part 1245)

- Straight forward (follow the CFR)
- Well defined (terms of license)

Grant or Contract:

- Small Business Innovative Research (SBIR)
- Small Business Technology Transfer (STTR)
- Technology Reinvestment Project (TRP)
- Advanced Technology Program (ATP)

Research Agreement:

- Cooperative Research and Development (CRADA)
- Technology Exchange Agreement
- Unreimbursed Space Act Agreement

Technology Transfer Capitalization

Grants or Contracts:

- Limited funds
- Restricted use of funds
- Significant Industry investment to apply
- Low probability of award
- An award can make a company or project

Research Agreements:

- Not a source of industry capital
- Capital for colleagues in Federal Laboratories
- Excellent source of intellectual capital

Private/Public Sector:

- Private capital (Angels & Equity)
- Venture capital
- Public Offerings
- Corporate capital

Interaction Mechanisms

and

Capitalization

Licensing and Capitalization

- Partial exclusivity (field of use)
- Attractive to all sources of capital
- Gives licensee a real opportunity
- Well defined relationship with the Government

Grants, Contracts and Capitalization

- Often sufficient for proof-of-concept
- Proof may lead to capitalization
- Builds investor confidence
- Provides cash flow
- Can defocus effort
- Shared intellectual property with Government
- Some Government rights may deter capitalization

Research Agreements and Capitalization

- Provides expertise and laboratory resources
- Provides motivation for Government employees
- May provide funding for Federal Laboratories
- May establish shared intellectual property
- May be unattractive to capital sources

Intellectual Property and Licensing

Government Relationship:

- Type of License: exclusivity
- Field of Use
- Practical Application Period
- Term:
- Fee:
- Royalty: minimum and running

Business Impact:

- Attract capital
- Little effect on business operations

Intellectual Property, Grants and Contracts

Government Relationship:

- Government data rights
- Government purpose license rights
- March-in rights

Business Impact:

- Decreases motive to apply
- May cause decreased interest in the project
- May result in loss of technology
- May result in loss of competitive edge

Intellectual Property and Cooperative Research Agreements

Government Relationship:

- Rights to disclose foreground and background data
- exclusive, partially exclusive, revocable, royality-bearing license

Business Impact:

- Must limit openness with government
- Results in shared property
- Results in uncertainty concerning future rights

DEVELOPING
TECHNOLOGY COMMERCIALIZATION INFRASTRUCTURE

Presented at the

Dual-Use Technology Conference

Gilruth Center
NASA Johnson Space Center
Houston, TX

FEBRUARY 2, 1994

GEORGE ULRICH
1702 SILVERPINE
HOUSTON, TX 77062

DEFINITIONS
(non-attribution food for thought)

Research: The systematic inquiry into why the world works the way it does.
Basic Research - "*What I am doing when I don't know what I am doing.*" - Wernher Von Braun.
Applied Research - "*If we can put a man on the moon, then why can't we...*"

Development: The link between a scientific idea and a prototype of that idea. Because development is more expensive that ideas it uses up to 2/3 of the funding in R&D.

Science: "*There is something fascinating about science. One gets such wholesale return of conjecture out of such a trifling investment of fact.*" - Mark Twain

Engineering: The practical application of scientific principals.

Technology: The practical knowledge, know-how skills and artifacts that can be used to develop a new product/service and/or new production/delivery service. It can be embodied in people, materials, cognitive and physical processes, plant, equipment and tools.

Commercialization Issue: Not, can it do the job?? but can it do the job profitably???!!!

AGENDA

1. What's Happening at JSC?
2. What Support is Available?
3. What is Industry Doing?
4. How Can Academia Support Technology Transfer
5. What are some solutions to increase and expedite commercialization?
6. How has JSC-Clear Lake Area served as a model for technology commercialization?

1. What's happening at JSC?

Small Business Innovative Research (SBIR)
Research & Technology Objectives and Plans (RTOP)
New Technology Report (NTR) - over 160 reports in 1993
AE Projects
Innovation Magazine
JSC Directorate's Facility and lab DocumentSpin-Offs
Tech Briefs
Spin-Offs
List of Technologies Historically and Current

NETWORKING! The road to proactive participation.

2. What Support Is Available?

FEDERAL	STATE	LOCAL	OTHER/ASSOCIATION
Advance Research Program Agency (ARPA)	Texas Innovative Network System (TINS)		
EASTIC	University Grants	Greater Houston Partnership (GHP)	
JSC Technology Commercialization Center (JTCC)	Texas Capital Network	Texas Space Business Roundtable	AAAS
Mid-Continent Technology Transfer Center (MCTTC)	Texas Product Development Center	MIT Enterprises	AIAA
Centers for the Commercial Development of Space (CCDS)	Office of Advanced Technology	CLAEDF	Austin Technology Incubator
Center for Aerospace Information	Dept. of Commerce	Houston Advanced Research Center (HARC)	IC2/ Austin Technology Incubator (ATI)
Mid-Continental Technology Transfer Center (MCTTC)	Texas Information Procurement Service (TIPS)		IEEE
Federal Coordinating Council for Science, Engineering and Technology	Small Business Development Center (SBDC)	Institute for Biomedicine and Biotechnology	
COSMIC	Texas Space Commission	Tech Tuesday	Southwest Research Institute (SRI)
NASA - JSC TU Office	Texas Marketplace		
National Technology Transfer Center (NTTC)			
NIST			
Research Triangle Institute (RTI)			
SBIR			
University Space Research Associates (USRA)			

THE MAIZE

Who can help and who is responsible to help - knowing the difference

Many organizations available
- competition is keen to show successful transfers
- they really want to help

BUT

You have to help them help you.

THE Process
(fact or fiction)

Over 67 processes/models on technology transfer in library search.

"Each technology transfer case is different."

Have you adequately defined the technology and its capability?

Do you have a business plan?

Have you done marketing research and analysis?

Do you have the rights to the technology?

Where do you start!!!!

I've heard what the government wants to do? What should I do?

1. Define your technology field:
   Library sources - Listing of all NASA Patents by topic.
   - Spin-off Magazines (published annually)
   - Innovation magazine
   - Tech Briefs (they're free) a response provides packages of related data.
   - subject search

   MCTTC - They have existing database.
   JTCC - Full range of support.
   Patent Office - for patent and license information.
   TU Office - For information about current activities at the Center.
   SBDC - Full range of support, not necessarily focused at NASA Tech.
   Small Business Development Class - Sam Bruno, UHCL.
2. Market Analysis, iterative
3. Business Plan, iterative
4. Financial Plan, iterative

Sources of Supportive Material

Plans (Business, Market, Financial)
- Library, Book Store,
- Texas Dept. of Commerce " Road Map to Starting Your Own Business"
- Weekend course at UHCL with Dr. Sam Bruno
- MCTTC
- JTCC
- SBDC
- Texas Product Development Center

Patent/License - Patent Office, Ed Fein

3. What is industry doing?

Perception of Industries in R&D vs. Operations

Results of local survey - the interest is high for technology commercialization.

VIsion	Majority believed technology commercialization would be beneficial to the company.
Current Status	Majority did not have program for commercialization of federal technology nor did they have knowledge of it but express great interest.
Pathway to the future	Majority were interested in future involvement.

Aerospace - accounts for 28% of the Clear Lake Area economy.

Medical and Health

Petrochemical

Environmental

Tourism - accounts for approx. 30% of the CL economy

Is your company involved in technology transfer?

o Who is the representative (usually in business development or marketing).

o If not, talk with marketing group, express interest, try to develop ties with government and academia for support.

o If no company support look to other options. Self, joint venture, etc.

OR

Maybe you want to go it alone. Will your company support your efforts?

4. How can academia support technology transfer

How can we induct technical commercialization into engineering and should we. (Who is responsible for commercialization at federal lab?)

<u>The Unwritten Laws of Engineering,</u> a reprint

University of Houston - Clear Lake, MCTTC, Small Business Seminar

University of Houston - Small Business Development Center
Texas Product Development Center

University of Texas - IC2, ATI

Texas A&M - MCTTC, Co-op program

Numerous other University and Education Organizations

"The facts, just the facts" Sgt. Friday

Technology at JSC - the numbers story

o Since 1976 - 107 JSC specific spin-off articles. Averaging about 5 per year.
o New Technology Reports (NTR) - 130 annually.
o JSC patent office successfully processes 30 patents per year.
  Multiple license requests per patent.
  Many patents go untapped.
o Tech Briefs - 210,000 subscriptions - just domestic. Requires update annually.

Year	# of Requests (JSC)	
1985	5,941	
1986	10,441	
1987	12,520	
1988	8,073	
1989	8,324	
1990	12,415	
1991	12,293	
1992	14,641	136,000 from all centers

o COSMIC - Thousands of users from JSC developed software. Help desk.
o List of 148 technologies for flavor attached.

Technology Transfer - Support your local economy!

Where do the technologies at JSC end up?

Spin-offs Statistics Since 1976.

These numbers apply only to documented SPIN-OFF records and do not account for other types of transfer such as from Tech Briefs, other Patents, MCTTC, TU or other transfer agents.

Of 86 total JSC recorded Spin-offs:

State	Number of Spin-offs	Patents	Remarks
Unk	11	1	76-78
CA	10	1	
NY	10	2	
MA	9	1	2X
PA	6	0	
TX	5	0	
CO	4	0	
WA	4	0	2X
RI	2	0	2X

Twenty five states were accounted for atleast one spin-off by 1991.
Texas is rated 3 in venture capital behind California and New Hampshire.
Texas is rated 3 in manufacturing.

What About Patents?

Of the 86 JSC spin-offs recorded since 1976 only 8 involved patents. The last one was in 1985. Are patents a necessary part of the process? Are they necessary for new business?

Patents:
1991: CT, OH, PA(2), WA
1990: FL, WA, WI, CA(2)
1989: DE, MA, TX

Why - Do we have the infrastructure we need?

Government
- NASA Technology Utilization Office
- Mid-Continental Technology Transfer Center
- JSC Technology Commercialization Center

Texas Department of Commerce
- Advanced Projects
- Small Business Development Center
  University of Houston

What are the local business doing? (See Questionnaire)
What is academia doing? (See MOU)
What are you doing?

Is your company involved in technology transfer?

o Who is the representative (usually in business development or marketing).

o Do you have an R&D unit? Advanced Projects Office?

o Is there funding set up for R&D or S&T?

o If not, talk with marketing group, express interest, try to develop ties with government and academia for support.

o If no company support look to other options. Self, joint venture, etc.

OR

Maybe you want to go it alone. Will your company support your efforts?

6. How has JSC-Clear Lake Area served as a 'model for technology commercialization

Technology development
Technology transferred out
Limit industrial base

MEMORANDUM OF UNDERSTANDING

Clear Lake Area Economic Development Foundation
and
University of Houston - Clear Lake

Purpose

The purpose of the agreement is to promote and support cooperation between the Clear Lake Area Economic Development Foundation (CLAEDF) and the University of Houston - Clear Lake (UHCL) for the purpose of promoting economic and social interests of the area, especially as related to technology transfer.

Goals

To accomplish the purpose, CLAEDF and UHCL agree to:

a. cooperate in the sponsorship, planning, and implementation of programs and projects as mutually identified and agreed upon that support the creation and retention of jobs throughout the area;

b. cooperate in the sponsorship, planning, and implementation of programs and projects as mutually identified and agreed upon that support the creation and retention of education and training opportunities throughout the area; and

c. cooperate to facilitate the technology transfer process across the area to increase the number of jobs in the area.

Strategies

The strategies for the CLAEDF and UHCL to help create and retain jobs through technology transfer processes include:

- Planning jointly activities related to the goals;
- Assisting one another in implementation and assessment of plans;
- Coordinating an area-wide strategic approach to technology transfer and jobs creation;
- Assessing the nature and levels of current and planned activities related to technology transfer;
- Encouraging active participation with business, academia, and government in pursuit funds that will assist the economic development of the area, as appropriate to CLAEDF and UHCL; and
- Assisting one another, to the extent possible, in all other ways that will achieve the goals of the MOU.

Agreement

Agreed to this day, August 16, 1993:

James Reinhartsen, President
Clear Lake Area Economic Development Foundation

Dr. Glenn A. Goerke, President
University of Houston - Clear Lake

UH-CL/CLAEDF TECHNOLOGY TRANSFER SURVEY RESULTS

survey distribution: app. 70

responses:
9 aerospace
4 non-aerospace
3 technology transfer
16 TOTAL

PART ONE: VISION - Texas Gulf Coast Area Technology Transfer

1.) Are you (company) interested in building a stronger technology transfer infrastructure within the Texas coast area?

Yes (14 responses)
Yes, especially in the area of biotechnology as it relates to Space Station.
Yes, if business opportunities are associated with such an effort.

2.) Would you (company) like to increase the percentage of federal funds used in technology transfer within the local area?

Yes (15 responses)
Yes. However, the plan must be well thought out and have a high value added prospectus.

3.) Would you (company) be willing to support technology transfer/commercialization if it reduced your risk and helped stabilize the local economy? (Such as occurred in the Silicon Valley from 1950s to the 70s.)

Yes (14 responses)
Yes, economic stabilization would allow the value of our homes in the Clear Lake area to remain stable.
Yes, if risks and payback are clearly defined.

4.) With the decrease in NASA's budget, there is a need to create meaningful jobs that will keep technical talent in the local area. Do you believe that increased emphasis on commercialization of technology can open new industries and strengthen local companies?

Yes (14 responses)
Yes, but only if U.S. and State governments are serious about supporting the programs.
Yes, but it will probably take several (5-10) years for us to see meaningful results.

PART TWO: CURRENT STATUS - Clear Lake Area Technology transfer

5.) What is your (company) present involvement and or position on technology transfer?

Present involvement limited or none but supportive and interested in opportunities especially if there are financial benefits. (5 responses)

We facilitate technology transfer and commercialization. (3 responses, TT Offices)

Attempting to commercialize a product based on NASA technology.

We are pursuing some new initiatives that involve technology transfer and support any opportunity for transfer of technology involvement.

Looking for specific technologies that may be beneficial to a specific project start-up I am pursuing.

Involved in television systems design, multi-media.

Several attempts, few results. Very expensive to market.

Our corporation has one formal TRP. We have attempted several locally, but have yet to be successful.

6.) How would you rate your (company) awareness on current technology transfer efforts?

Low (7 responses)

Fair/moderate (4 responses)

Average

We believe we have very good awareness, but we welcome any opportunity to expand our awareness.

Moderately high. We respond to Technology Transfer questionnaires with regard to our government contracts, but we seldom receive any feedback on the technology transfer results.

High (2 responses, TT Offices)

7.) To support the economy, people need to learn how to utilize the numerous federal and state programs available which provide various degrees of funding. Do you understand all the programs?

No (14 responses)

Have you tried or are you interested in tapping into these programs?

Interested (11 responses)

Some experience, interested in more (3 responses)

Interested especially as these programs relate to small businesses. The CLAEDF should publish a brochure on "How to Start a Small Business" and ask the programs that are available for assistance.

We are heavily involved in the programs.

PART THREE: PATHWAY TO THE FUTURE?

8.) Do you perceive ways to increase the existing mutual support between industry, government and academia? How?

Yes, through increased communication and cooperation. This proposed meeting is an excellent start.

Yes, should start by sitting down and talking between all parties concerned.

Yes, better define needs and quantity through local councils or task forces.

Yes, by finding common objectives which support the goals of both organizations.

Yes, concentrate on real world application. Stop micro-managing and measure results instead.

Yes, more focused effort by local NASA technical people to spin off technology.

Yes, collaborative technology seminars/groups, infrastructural interfacing, mutual trust and support paradigms - all demand cultural revolutions which are NOT easy to make.

Yes, use service providers at Tech Transfer centers or university economic centers to coordinate.

Yes, middle management exchange program (details TBD).

Yes, develop an on-line interactive computer network.

Yes, seminars and the "How to Start a Small Business " brochure.

Yes, we need to add capital resources.

Yes, Technology Transfer.

None other than those we are already aware of.

Open to any good ideas.

It is a natural fit for government, academia, and industry to come together to facilitate the progress of technology transfer and can be a benefit to all parties.

9.) Are there other areas such as consortiums or partnerships with academia, industry or government which you would like to see that can benefit your company?

Yes (6 responses)

Have an interest in exploring this further.

Consortiums and partnerships could be very beneficial under the right circumstances.

A "forum" may be more appropriate than a consortium.

UH-CL could offer some special courses on starting small businesses. UH-CL and CLAEDF could sponsor these courses together.

More government seed money for small and economically disadvantaged businesses.

Education is primary, definition of competencies, quotas of students (can't prepare everybody for few jobs, target people & jobs & companies)

I have not been impressed with consortiums. Academia lacks the business motive to make a product economically successful.

None at the present time.

I'll think about it.

10.) The federal and state governments are increasing their support of technology transfer through various programs. Is your company interested in developing plans to tap into these programs with matching funds? Consortiums, joint projects?

Yes (9 responses)

Yes, depending on required funding and delivered benefits (3 responses)

Matching funds - maybe, consortiums - no

NASA is interested but the laws and regulations are somewhat restrictive. How will the government distinguish "winners" form "losers" in today's economy" NAFTA should help many small business in the Houston area after it is in place.

We have no funds to invest in matching programs, joint projects or consortiums.

11.) Does your company have future plans to become more involved in technology transfer?

Yes (4 responses)

Yes, with NASA technology (2 responses)

Yes, increased focus on SBIR programs with planned increase in sources of cash.We are a small business and must look at each opportunity carefully. We would like to select one project to pursue.

Will continue to look for appropriate technology - with reasonable "strings attached"

We are currently looking for new opportunities. Technology transfer is in an area of great potential.

We are now investigating reasons for meeting with others in the area. There is going to be a need for a group to do some nurturing and channeling of products and interests. This is not a mainstream activity for a chemical company.

Currently looking at several ideas.

Within the scope that you have defined here, not much.

Not at this time.

I do not know.

12.) The Governor's Office is supporting the local economy by funding transition programs. What do you see as a way to provide long term support to this effort?

Would like to know more (3 responses)

Set up a UH-CL program to teach new business techniques and ways to find money

The Governor's Office should support businesses trying to transition into new markets.

The Governor's office should set up a special commission to review and reduce government red-tape and bureaucratic requirements which limit and restrict small business start-ups. They could also work with local banks to make start-up capital readily available in the Clear Lake area.

The key to long term support is to eliminate the need for the support - i.e., to develop new markets for our companies which decrease our dependence upon the government.

The goal is to create jobs. First the "gap" of technology need and development of suitable service or product has to occur. I doubt very seriously if most companies have any person(s) assigned to developing suitable opportunities.

Subsidies or tax incentives might help at the State level.

Technology Transfer efforts should pull from across petrochemical as well as aerospace community in some sort of joint effort.

How do you feel about a two pronged approach - one at educating and the other at developing new business through technology transfer?

Believe this to be a good approach (5 responses)

Good approach if properly weighted (25% education, 75% development)

I believe education must support the R&D phase, but follow the other phases.

We are heavily involved in both education and technology transfer and sponsor quarterly "Building Your Company VIA Technology Transfer" seminars.

What is the goal of the "educating" approach? It's time to get on with doing the new business development.

Not enough info at this point to evaluate.

Session T7: TECHNOLOGY TRANSFER "HOW TO"

Session Chair: George Ulrich

	JOHNSON SPACE CENTER	Office of Patent Counsel Edward K. Fein

LICENSING NASA TECHNOLOGY

	JOHNSON SPACE CENTER	Office of Patent Counsel Edward K. Fein

NASA POLICY

It is NASA policy to protect the Government's interest in, and to provide for the widest practicable and appropriate dissemination, early utilization, expeditious commercial development, and continued availability of, inventions made by Government employees and by employees of NASA contractors in the performance of work under NASA contracts.

JOHNSON SPACE CENTER

Office of Patent Counsel

Edward K. Fein

WHAT TECHNOLOGY CAN NASA LICENSE?

Intellectual Property

- Patents
- Copyrights
- Trademarks
- Trade Secrets

JOHNSON SPACE CENTER

Office of Patent Counsel

Edward K. Fein

WHERE DO NASA INVENTIONS COME FROM?

- NASA Employees
- NASA Contractors
  - Small Business Firms and Nonprofit Institutions
  - Other than Small Business Firms and Nonprofit Institutions, "Big Business"

	JOHNSON SPACE CENTER	Office of Patent Counsel
		Edward K. Fein

WHY NOT PLACE TECHNOLOGY IN THE PUBLIC DOMAIN?

- Technology in the Public Domain is Freely Available to All
- No Incentive to Invest Risk Capital Required to Commercialize Technology

  (Competitors may Freely Copy and Market)
- Exclusivity keeps Competition from Imminent Access to the Technology

	JOHNSON SPACE CENTER	Office of Patent Counsel
		Edward K. Fein

WHO CAN OBTAIN COMMERCIAL RIGHTS?

- Contractors Who Make and Report "SUBJECT INVENTIONS"

  (Inventions Conceived or First Reduced to Practice in the Performance of Work under a NASA Contract)
  - Waiver of Title (Big Business)
    - Advance Waiver
    - Waiver after Reporting an Invention
  - Election to Retain Title

    (Small Business/Nonprofits)
- Inventors
  - Contractor Employees
  - NASA Employees
- Licensees

JOHNSON SPACE CENTER	Office of Patent Counsel Edward K. Fein

HOW TO FIND INFORMATION ON NASA'S PATENTS

- Patent and Technical Literature Searches
- NASA Tech Briefs
- NASA Patent Abstract Bibliography

  Indexes and Abstracts all of NASA's Patent and Patent Applications Arranged by Technical Subject Matter, Inventor, Source, and Number. Available for Purchase from the National Technical Information Service (NTIS), Springfield, VA 22161 (703)487-4600
- NASA Regional Technology Transfer Centers (RTTCs); e.g., Mid-Continent Technology Transfer Center (MCTTC)
- NASA Field Center Technology Utilization Offices
- NASA Field Center Patent Counsel
- NASA Headquarters Office of General Counsel

JOHNSON SPACE CENTER	Office of Patent Counsel Edward K. Fein

HOW TO OBTAIN NASA COMPUTER PROGRAMS

The Computer Software Management and Information Center (COSMIC) provides public access (at a reasonable cost) to computer programs developed by or for NASA, as well as selected programs of the Department of Defense and other Government agencies.

COSMIC
382 East Broad Street
Atlanta, GA 30602

Telephone: (706) 542-3265
FAX: (706) 542-4807

JOHNSON SPACE CENTER

Office of Patent Counsel

Edward K. Fein

HOW TO APPLY FOR A NASA LICENSE

(14 CFR §1245.207)

(a) Identify the invention for which license is sought (NASA Case Number; Title; Patent/Patent Application Number; Date)

(b) Type license requested (exclusive, non exclusive)

(c) Name and address of applicant (person or organization)

Citizenship or place of incorporation

(d) Name, address, telephone number of applicant's representative, if different from (c)

(e) Nature and type of applicant's business

Products or services which applicant has successfully commercialized

Number of employees (approximate)

(f) Statement regarding how applicant discovered invention was available for licensing

JOHNSON SPACE CENTER

Office of Patent Counsel

Edward K. Fein

HOW TO APPLY FOR A NASA LICENSE (continued)

(14 CFR §1245.207) (continued)

(g) Statement regarding small business firm status as defined in §1245.202(d)

(h) Detailed description of applicant's plan for development and/or marketing of the invention, including:

(1) Statement of anticipated requirements to achieve practical application as defined in §1245.202(e) in terms of time and investment of capital and other resources

(2) Statement of applicant's capability (resources) for fulfilling the plan, including manufacturing, marketing, financial, and technical

(3) Fields of use in which applicant plans to practice the invention

(4) Geographic area of intended manufacture and of intended use or sale

(i) Identification of any previously held licenses of Federal inventions

(j) Statement of applicant's knowledge (if any) regarding: (1) extent of use of the invention (by Government or others); and (2) its commercial availability.

PROCESSING OF LICENSE APPLICATIONS § 1245.208

- Applications should be sent to: the Patent Counsel of the NASA field center responsible for the invention
- Patent Counsel reviews application, preliminarily negotiates terms and conditions, and makes recommendation to NASA Headquarters
  - Grant as requested
  - Grant with modification
  - Deny
- Exclusive and Partially Exclusive Licenses require a Federal Register notice providing opportunity for filing written objections within a 60-day period
- Headquarters makes final determination
- Patent Counsel and prospective licensee enter into final negotiations
- Right of Appeal
  - Applicants who have been denied license
  - Objectors to the Federal Register notice

JOHNSON SPACE CENTER

Office of Patent Counsel

Edward K. Fein

TECHNICAL ASSISTANCE

Technical assistance may be available to licensee from NASA scientists and engineers to aid in the development and commercialization of selected inventions and technologies. Technical assistance is generally negotiated directly with the NASA Field Installation where the licensed invention was made, and it may be provided on a cooperative or a cost-reimbursable basis.

TECHNOLOGY AS A COMMERCIAL RESOURCE

Jill D. Fabricant, Ph.D.
Director
Johnson Space Center
NASA Technology Commercialization Center

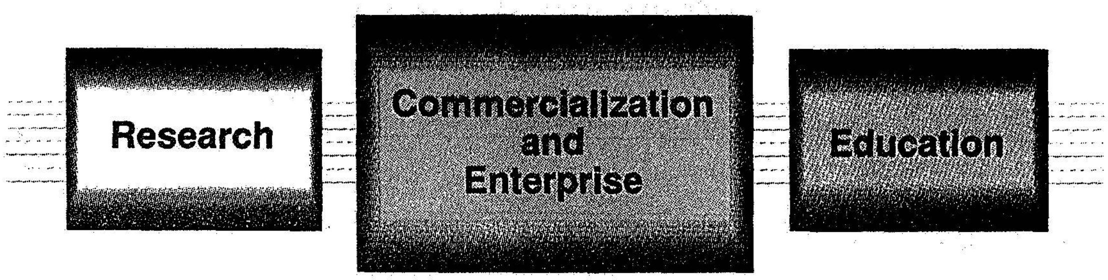
Presentation for the 1994
Dual-Use Space Technology
Transfer Conference
Houston, Texas
February 3, 1994

Technology can be regarded as a powerful added value to any corporate institution. Technology leads to new products and services which creates value in companies. The not-invented-here syndrome serves to stifle growth in many companies, but it can be overcome and now there is a growing move toward outsourcing of new technology in many institutions. This has lead to the epoch of technology transfer first created at the University of Wisconsin over 2 decades ago. Today, technology transfer is commonplace in many institutions, and our center, the Johnson NASA Technology Commercialization Center, a part of the IC2 Institute, at the University of Texas is involved in the transfer of NASA technology to the private sector. Dr. George Kozmetsky, the founder of the IC2 Institute, was recently awarded the Science and Technology Award from President Clinton for his decades of work in technology transfer. Dr. Kozmetsky is the Principal Investigator on this project and was one of our keynote speakers at the beginning of this conference. Today, I would like to discuss our program at the NASA Technology Commercialization Center, the background behind the experiment, and how we are presently involved in the placement and licensing of NASA technology.

IC 26

IC2 Institute

Innovation, Creativity, and Capital



IC^2 NASA Technology Commercialization Centers

IC^2 NASA Technology Commercialization Centers
8920 Business Park Drive
Austin, TX., 78759
Phone: (512) 794-9994
Fax: (512) 794-9997

NT911C

Ames Research Center
Mountain View, California

Ames Technology Commercialization Center
ATCC, 155A Moffet Park Drive
Sunnyvale, CA., 94089
Phone: (408) 734-4700
Fax: (408) 734-4946

NT911C

Johnson Space Center
Houston, Texas

Johnson Technology Commercialization Center
JTCC, 2200 Space Park Boulevard
Houston, TX., 77258
Phone: (713) 335-1250
Fax: (713) 333-9285

IC² Institute

Successful Technology Commercialization

- Market need / benefit
- Value-added chain
- Networks
- Quality accessibility to
  - Markets
  - Technology
  - Management
  - Capital

IC² Institute

Major Trends

- Market driven economies
- Worldwide democracy
- New infrastructures
- New technologies
- Accountability

 IC 28

Know How Networks

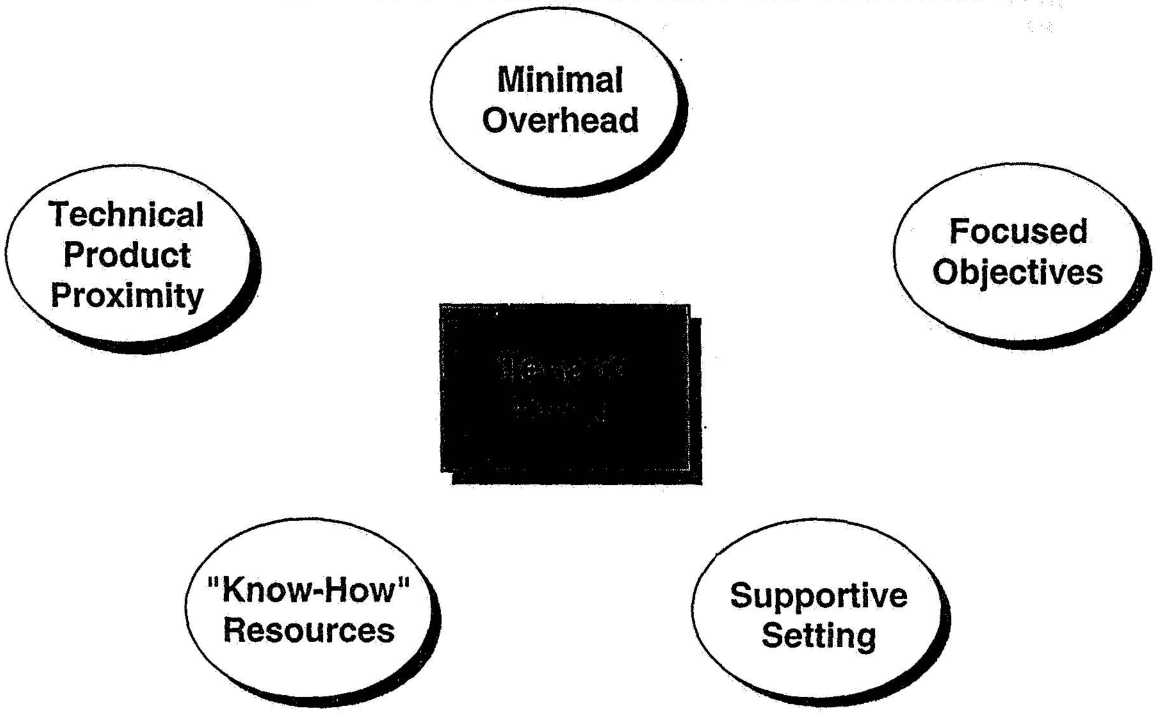
- Business know-how
- Market research know-how
- Financial know-how
- Distribution, sales, service know-how
- Entrepreneurial know-how
- Managerial know-how
- Other technological know-how

vg 2640

Culture

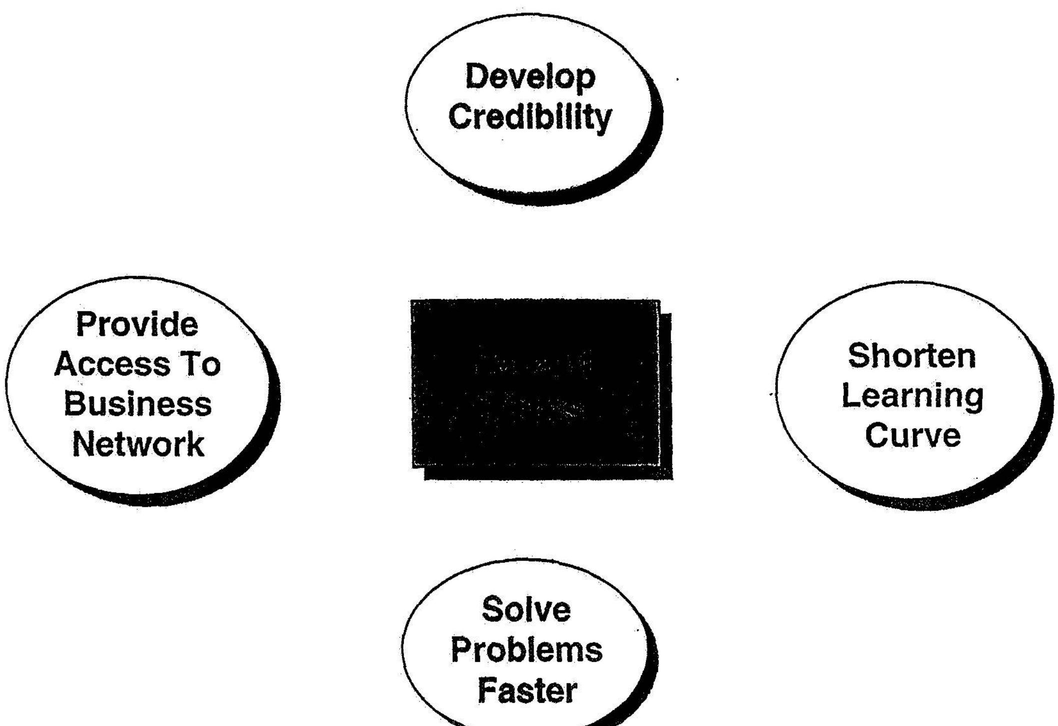
- Must be physically separate from NASA R&D
- Must an entrepreneurial environment
- Must provide entrepreneurial training
- Must be internally cooperative and externally competitive
- Must be an "open" organization
- Must provide continuous learning for commercialization

Benefits of Technology Commercialization Center Sysytem



IC.11

Benefits of Technology Commercialization Center System



IC.12

NT911C

Purpose

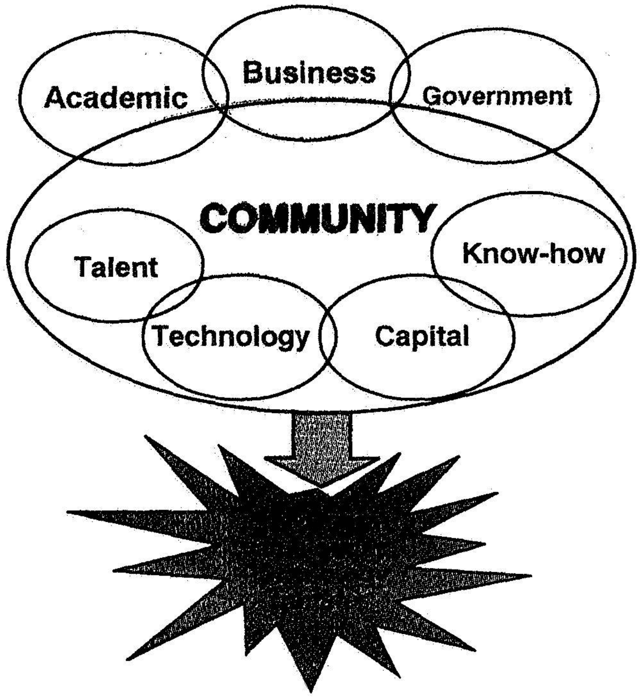
- Promote and support vision and leadership
- Utilize free market mechanisms
- Leverage research and development

NT911C

Value Added Intermediaries

- Meet goals of NASA, community, tenants
- Improve U.S. competitiveness
- Foster dual use of technologies
- Coordinate NASA missions with emerging industries

Technology Incubation



vg2855c IC.32

☆U.S. GOVERNMENT PRINTING OFFICE: 1994-561-486

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	May/94	Conference Publication - February 1-3, 1994

4. TITLE AND SUBTITLE

Dual-Use Space Technology Transfer Conference and Exhibition - Volumes I and II

5. FUNDING NUMBERS

6. AUTHOR(S)

Kumar Krishen, Compiler

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

Lyndon B. Johnson Space Center
Houston, Texas 77058

8. PERFORMING ORGANIZATION REPORT NUMBERS
S-765

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

National Aeronautics and Space Administration
Washington, DC 20546

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
NASA CP-3263

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Unclassified/Unlimited
Available from the NASA Center for AeroSpace Information
800 Elkridge Landing Road
Linthicum Heights, MD 21090
(301) 621-0390 Subject Category: 99

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This document contains papers presented at the Dual-Use Space Technology Transfer Conference and Exhibition held at the Johnson Space Center February 1-3, 1994. Possible technology transfers covered during the conference were in the areas of information access; innovative microwave and optical applications; materials and structures; marketing and barriers; intelligent systems; human factors and habitation; communications and data systems; business process and technology transfer; software engineering; biotechnology and advanced bioinstrumentation; communications signal processing and analysis; new ways of doing business; medical care; applications derived from control center data systems; human performance evaluation; technology transfer methods; mathematics, modeling, and simulation; propulsion; software analysis and decision tools; systems/processes in human support technology; networks, control centers, and distributed systems; power; rapid development; perception and vision technologies; integrated vehicle health management; automation technologies; advanced avionics; and robotics technologies. More than 77 papers, 20 presentations, and 20 exhibits covering various disciplines were presented by experts from NASA, universities, and industry.

14. SUBJECT TERMS
microwave, biotechnology, intelligent systems, software analysis, automation, bioinstrumentation, software engineering, advanced avionics and robotics

15. NUMBER OF PAGES
921

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited