NASA CR-202756

1422

1996 Research Reports

NASA/ASEE Summer Faculty Fellowship Program

John F. Kennedy Space Center
and
University of Central Florida

1996 RESEARCH REPORTS

NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

JOHN F. KENNEDY SPACE CENTER

UNIVERSITY OF CENTRAL FLORIDA

EDITORS:

Dr. Roger Johnson, University Program Director
Department of Mechanical, Materials and Aerospace Engineering
College of Engineering
University of Central Florida

Mr. Gregg Buckingham, NASA/KSC Program Director
Administration Office
John F. Kennedy Space Center

PREPARED FOR:

John F. Kennedy Space Center
Merritt Island, Florida

NASA Grant NGT10-52605

Contractor Report No. CR-202756

October 1996

PREFACE

This document is a collection of technical reports on research conducted by the participants in the 1996 NASA/ASEE Summer Faculty Fellowship Program at Kennedy Space Center (KSC). This was the twelfth year that a NASA/ASEE program has been conducted at KSC. The 1996 program was administered by the University of Central Florida (UCF) in cooperation with KSC. The program was operated under the auspices of the American Society for Engineering Education (ASEE) and the Office of Educational Affairs, NASA Headquarters, Washington, D.C. The KSC program was one of nine such Aeronautics and Space Research Programs funded by NASA Headquarters in 1996. The basic common objectives of the NASA/ASEE Summer Faculty Fellowship Program are:

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

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

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

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

The KSC Faculty Fellows spent ten weeks (June 3 through August 9, 1996) working with NASA scientists and engineers on research of mutual interest to the University faculty member and the NASA colleague. The editors of this document were responsible for selecting appropriately qualified faculty to address some of the many problems of current interest to NASA/KSC. A separate document reports on the administrative aspects of the 1996 program. The NASA/ASEE program is intended to be a two-year program to allow in-depth research by the University faculty member. In many cases a faculty member has developed a close working relationship with a particular NASA group that has provided funding beyond the two-year limit.

TABLE OF CONTENTS

1996 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

PHASE IX FIBER OPTIC CABLE MICROBENDING AND TEMPERATURE CYCLING TESTS

Dr. Mustafa A.G. Abushagur, Professor
Electrical and Computer Engineering Department
The University of Alabama in Huntsville
Huntsville, Alabama

KSC Colleagues - Po T. Huang and Larry Hand
Communications/Fiber Optics

Contract Number NASA-NGT10-52605

July 26, 1996

ABSTRACT

Optical fibers represent the back bone of the current communication networks. There performance in the field lacks long term testing data because of the continuous evolution of the manufacturing of fibers and cables. An optical fiber cable that is installed in NASA's KSC have experienced a dramatic increase in attenuation after three years of use. The attenuation has increased from 0.7 dB/km to 7 dB/km in some fibers. A thorough study is presented to assess the causes of such attenuation increase. Material and chemical decomposition testing showed that there are no changes in the composition of the fiber which might have caused the increase in attenuation. Microbending and heat cycling tests were performed on the cable and individual fibers. It is found that the increase in attenuation is due to microbending which is caused by excess stress exerted on the fibers. This was the result of manufacturing and installation irregularities.

Phase IX Fiber Optic Cable Microbending and Temperature Cycling Tests

Mustafa A.G. Abushagur

1. INTRODUCTION

Optical fiber networks are considered as an ideal medium for communications because of their extremely high data rates, light power throughput, security and immunity to interference. Fiber networks have made the current revolution in information possible. Optical fibers are very small in size vulnerable to environmental influences and need to be cabled to allow ease of handling and protection. The cabling and installation cause an increase in fiber attenuation. After installation and during the life of the cable, estimated to be well beyond 20 years, the fiber attenuation should not increase more than a fraction of a dB/km. Increase in attenuation after installation causes serious problems in system performance such as decrease in signal-to-noise ratio (SNR) and increase in bit-error-rates (BER). Cables after being installed are susceptible to the environmental changes. Penetration of hydrogen in the fiber either due to water infusion or a chemical reaction between the cable components is of a great concern [1-3]. Macrobending and microbending are also major contributors to attenuation increase [4-5].

In this report we present a case study on an optical fiber cable that was installed in 1993 at NASA's Kennedy Space Center. This particular cable demonstrated dramatic increase in attenuation over a very short period of use. The cause of such increase is investigated and reported in this paper. In Section 2 we present the background of the problem. The methodology of testing is introduced in Section 3. Results of material testing is included in Section 4. Microbending test results are given in Section 5. Summary and conclusions are given in Section 6.

2. BACKGROUND

Kennedy Space Center is one of the pioneer users of fiber optic technology because of its need for high data rate networks. It has installed since the early 1970's more than 13,000 miles of fibers between its Space Shuttle facilities. In September 1993, PHASE IX network was installed between five nodes as shown in Figure 1.

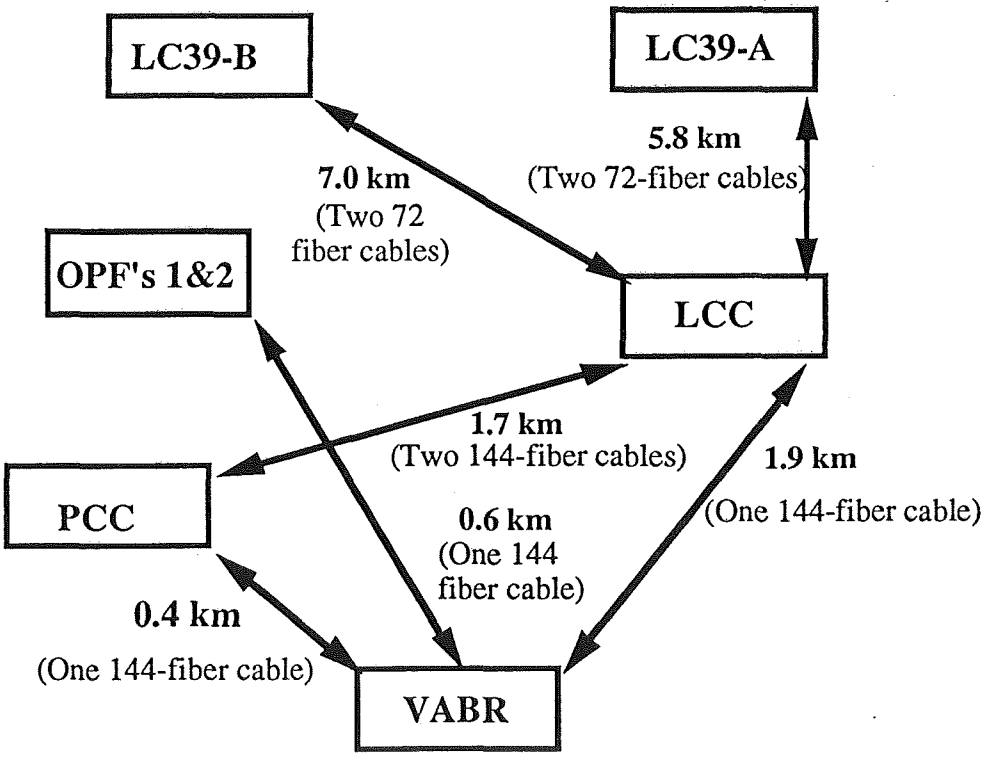


Figure 1. Fiber optic network between the Space Shuttle facilities.

This network is made out of 9 cables totaling 31.9 km. Most of the fibers are graded index multi-mode fibers and three of cables have few single-mode fibers. All the fibers were tested for their attenuation after installation and met

the specification of less than 1 dB/km. The fibers were tested again in 1994 and were within specification. In early 1996, a deterioration of performance and an increase in the BER was noticed. Measurement of the attenuation of the fibers on April 26, 1996 showed that most of the fibers attenuation has increased drastically. Attenuations were measured and found to be between 0.8 and 7.8 dB/km with an average of 2.61 for a particular cable (LCC-PCC). These measurements were made at wavelength of 1300 nm. Similar attenuation increase took place at 1550 nm. The single-mode fibers in cable (VABR-LCC) did not show any increase in attenuation. The LCC-PCC cable was then removed from underground for further testing that can not be done while installed. The attenuation of the fibers were immediately measured on April 29, 1996. The attenuation of the fibers were dropped after removal to the range from 0.7 to 3.1 dB/km with an average of 1.23 dB/km , see Figure 2.

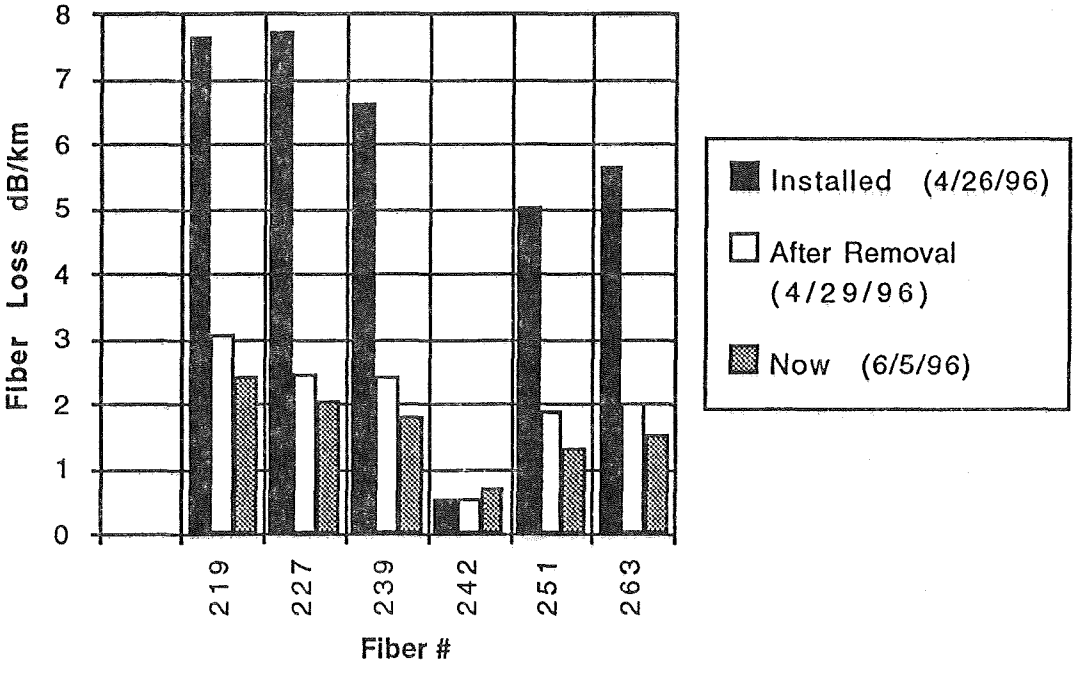


Figure 2 Test fiber attenuation while installed and after removal.

The increase in attenuation can be due to several factors [4]. These are in general can be classified in two categories: material composition change and microbending. The material change is due to infusion or generation of mainly hydrogen especially in an O-H chain. This is the result of water infusion in the core of the fiber or might be generated by a chemical reaction between the different components of the cable, especially the gel compound and fiber coating. Microbending can be caused by cable aging, which results in cable shrinkage, manufacturing and installation problems, or deterioration of the fiber protective coating or the gel.

3. TESTING METHODOLOGY

The removed cable was tested to determine the cause of attenuation increase. The tests were designed to investigate both attenuation sources and were conducted on a set of test articles listed in Table 1. PHASE V cable is a cable that was installed prior to PHASE IX cable in similar environment but still within attenuation specification and is used for comparison purposes. Both cables were manufactured by the same company (Chromatic) using fibers from two different vendors Spectran and Corning). The Spectran and Corning fiber test spools are similar to those fibers in PHASE IX and PHASE V cables, respectively, but never been cabled.

Two separate sets of tests were conducted. The first set focused on the possibilities of a chemical change in the fiber and cable components. The second set focused on microbending and what might have caused it from both fiber and cable elements. The following two section present the tests and results.

TABLE 1 1 km PHASE IX test fibers

Length	Item
1 km	PHASE IX FM66 cable
50 meters	PHASE IX FM66 cable
50 meters	PHASE V cable
50 meters	Spectran fiber test spool
50 meters	Corning fiber test spool
50 meters	Fibers, from PHASE IX cable, in buffer tube
50 meters	Fibers, from PHASE IX cable, without buffer tube

4. MATERIAL AND CHEMICAL TESTS

Fibers used for communication are made from fused silica, SiO_2, and GeO_2 is used to allow the index of refraction variations required. The existence of other impurities in the fiber are the major cause for absorption of the light energy. The lead factor in absorption in the wavelengths of interest is the OH. The tests carried in this investigation were based on determining the contents of the fiber and cable elements. A comparison between a number of test items listed in Table 1 were carried. The samples were analyzed by optical microscopy, inductively coupled plasma (ICP) spectroscopy, anion-ion chromatography, and scanning electron microscopy (SEM) with energy dispersive spectrometry (EDS) and wavelength dispersive spectrometry (WDS). The test data showed that all fiber cores have Ge, Si and O. Spectran fibers contain some P in the core. From all the tests there was no evidence of the presence of any hydrogen increase which might have caused the attenuation problem. The material content of the fiber coatings, the gel, and the cable sheeting were found to be the same for all test samples. There was a visible difference between the primary coating of the Sepctran and Corning fibers. The first seems to have a coarse texture when examined under the microscope. The secondary coating of the Spectran fiber is found to be much softer than that of Corning fibers. In order for the fiber to be protected against outside stresses the primary coating need to be soft and the secondary coating to be hard. If the secondary is soft it fails to protect the fiber and causes an increase in microbending.

White light test was also performed to investigate the material composition change in the fibers. This test provides the means to measure the attenuation of the fiber across the spectrum. The results of this test showed that the fibers in the cable and those on the test reels have similar attenuation spectrum. This concludes that there was no change in the fiber composition and there was no hydrogen infustion into the fiber. However the test showd that Spectran fibers have same attenuation at both 1300 and 1550 nm wavelengths, while Corning fibers have lower attenuation at 1550 nm.

In summary all the chemical and material tests performed leads to the conclusion that the attenuation increase in PHASE IX cable was not caused by material composition change in the fibers but must be some other such microbending.

5. MICROBENDING TESTS

Microbending loss occurs when small, periodic perturbations are introduced in fiber [6-8]. These perturbations in the optical fiber cause it to deviate from being a perfectly circular cylindrical waveguide. So any external stresses on the fiber will cause such loss. This can be the result of the cabling process of the fiber or after installation as a result of the cable shrinkage due to aging or environbmental factors. The fibers are made in such a way to resist microbending by coating them with a soft then a hard plastic coating [9], also by being placed in buffer tubes. These measures do improve the microbending resistance of fibers. Aging problems of the coating itself is

need to be studied. In this section we report the investigation that has been carried to determine the microbending loss of the fiber cable understudy. By isolating the different factors that cause the microbending we should be able to determine the major source of this loss. We have performed two separate tests one on the 1 km PHASE IX cable and the second test on the 50 m samples listed in Table 1. The first of these tests is a temperature cycling which simulates the effects of cable shrinkage. By decreasing the temperature of the cable it shrinks and induces stresses on the fibers. Increasing temperature higher than room temperature causes the cable to expand and releases the fibers from the stress. The temperature cycling test started at 20 ^{o}C then lowered to -20 ^{o}C and then raised gradually to 60 ^{o}C and then back to 20 ^{o}C. The second test is a microbending resistance test performed on short lengths of fibers.

5.1 1 km Cable Test

The 1 km cable was placed in a thermal chamber where both temperature and humidity were controlled and monitored, see Figure 3. The cable was on a spool.

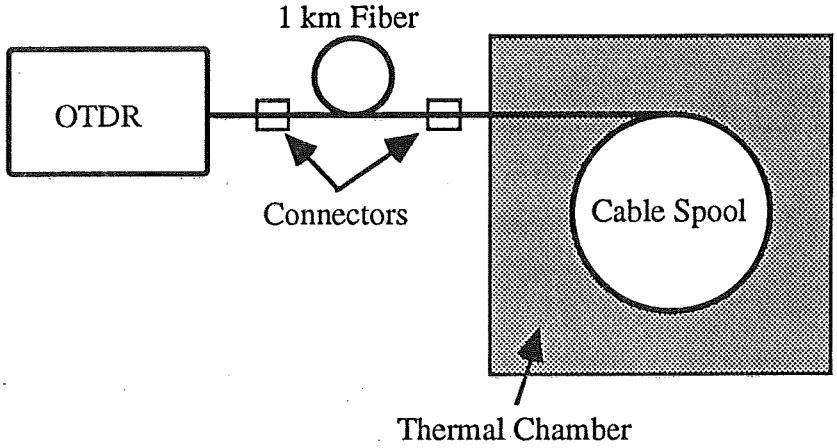


Figure 3. Experimental setup for temperature-attenuation measurement.

The temperature was measured with two thermo-couples one on the surface of the spool and the other was buried will within the cable. The test started at room temperature then reduced to - 20 ^{o} C and then raised to 60 ^{o}C. The temperature in the chamber was left at each step overnight to allow the temperature to reach the inside of the cable. The attenuation of six fibers was measured at each of these temperatures. The attenuation measurements are plotted in Figure 4.

The fiber attenuation is very high at -20 ^{o}C since the cable sheeting contracted and induced stress on the fibers. As the temperature increased the attenuation dropped very rapidly till about 0 ^{o}C and decreased with much slower rate onward. The attenuation of the tested fibers is different. The attenuation range of the fibers is about 3 dB/km at -20 C and all converge to within 0.2 dB/km range at 60 ^{o}C. Fiber 242 tested within specification throughout the tests even when the cable was installed except for the range of temperatures lower than 0^{o}C. This fiber shows the minimum attenuation and does not change much between 0 and 60 ^{o}C. While fiber 219 which demonstrated the highest attenuation while installed displays the highest attenuation throughout the temperature test shown in Figure 4. Fibers 219 and 227 are in the same buffer tube as well as 242 and 251 are in another buffer tube.

The temperature test simulates a number of factors in the field. First, the temperature change throughout the year. Second, the aging effect which results in shrinkage of the cable. These results show the extent of the ability of the fiber to cope with such effects. The fibers are designed to resist such stress. The major function of the primary and secondary coatings of the fiber is to protect the fiber from external stresses [10]. The fibers in the cable show a continuous increase of attenuation with cable shrinkage (temperature decrease). The attenuation dropped as the stress on the fibers was released. The attenuation variations that are shown in Figure 4 might not be due to the cable shrinkage only. The fiber itself might experience attenuation change due to its coating. The second set of experiments attempts to isolate these different factors.

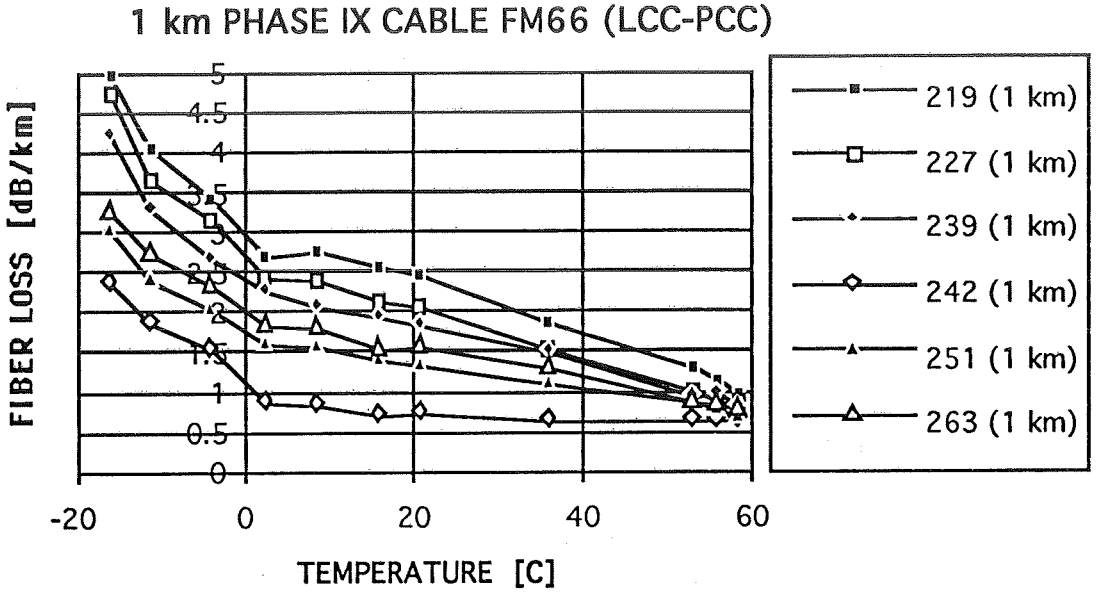


Figure 4. The attenuation of 6 fibers from PHASE IX cable is plotted as the temperature was increased from -20 °C to 60 °C.

5.2 50 m Cable and Fiber Tests

The second test conducted is on the 50 m fibers listed in Table 1. The attenuation measurements were made using the power meter instead of an OTDR since the fiber lengths are too short. The fibers and cables are placed in the same thermal chamber. The experimental setup is shown in Figure 5.

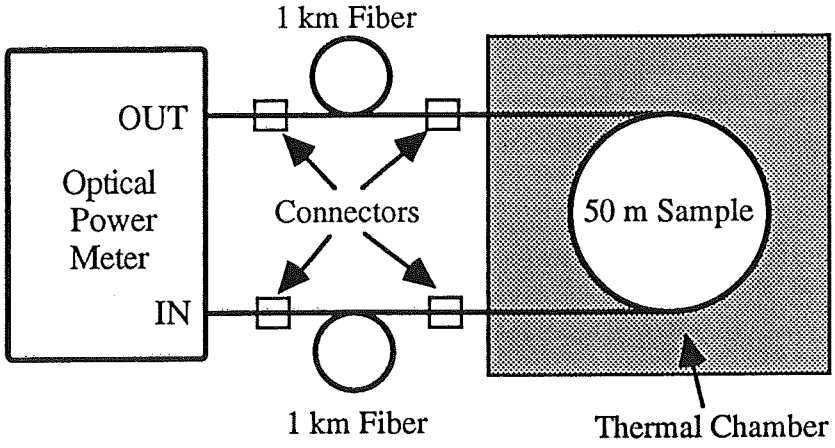


Figure 5 Experimental setup for the 50 m fibers and cables testing.

The temperature was changed from +20 to -20 and then the attenuation was measured after at least four hours from the time the chamber's temperature was changed. Two 1 km fibers were placed one between the light source and the test fiber and the second between the test fiber and the photodetector. The power meter uses as a reference the power output when the network in Figure 5 is connected excluding the test fibers. The attenuation reading of the test fibers includes the loss of one set of connectors, and one fusion splice. The readings for each fiber were taken several times to assure repeatability and accuracy.

The attenuation of the single fibers, two of the test spools and two are taken from PHASE IX cable were measured over the entire temperature range and the results are shown in Figure 6. The attenuation of the test spool fibers stayed almost constant over the temperature range from -20 to 60 ^{o}C. The fibers taken from the cable demonstrate the same behavior as those in the cable shown in Figure 5. Their attenuation increases rapidly after the temperature drops below freezing. The only difference between the two sets of fibers is that the environment they have been in over the last four years and the color coating, since the test spool fibers do not have any color coating. The fibers taken from the cable were cabled and installed in the field. Attenuation increase at low temperatures is either due to the color coating or a damage in the primary coating. The fiber glass itself do not exhibit such attenuation change with temperature. As temperature decreases the coatings of the fibers will shrink much more than the fiber itself. This intern causes an induced stress on the fibers leads to microbending losses. The coefficients of expansion of the fused silica is about $5x10^{-7}$ /^{o}C, while it is the range of $10\text{-}22x10^{-5}$ /^{o}C for Polyethylene (PE) which is typically used for the primary and secondary coatings of the fiber.

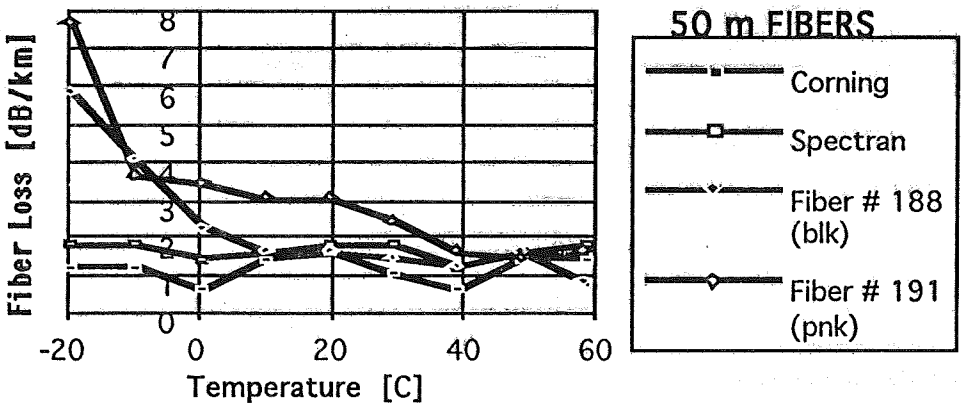


Figure 6

The next set of 50 m fibers tested are those removed from PHASE IX cable but left inside their loose buffer tubes which are filled with gel. These are four fibers with ID# 233, 239, 253 and 263. The attenuation curves as a function of temperature are shown in Figure 7 All fibers show slight attenuation increase for temperatures lower than 0 ^{o}C. Fiber 253 demonstrates the best performance. The reason for this increase is similar to that mentioned for the previous set of fibers in Figure 6. Also it can be seen from the comparison between the losses at room temperature and higher for these fibers and those in Figure 6 that the buffer loose tube has no effect on the fiber losses.

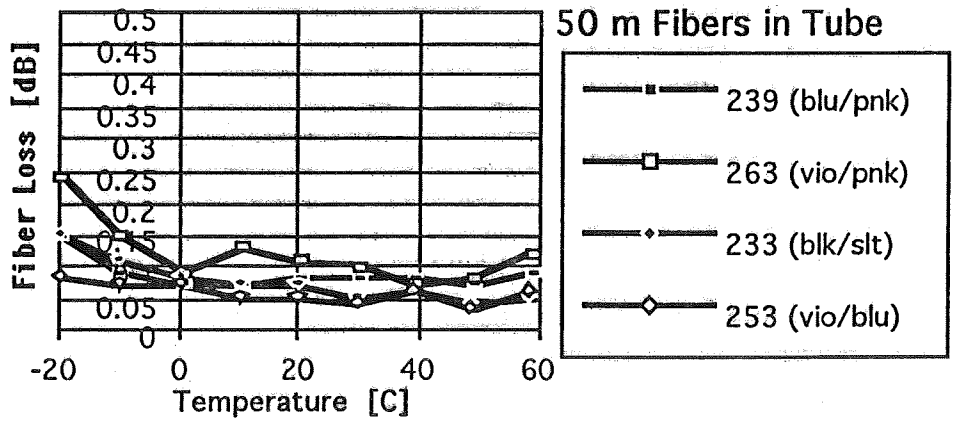


Figure 7.

The third set of 50 m fibers tested are those in PHASE IX cable. The fibers test are those with ID# 219, 227, 239, 242, 251, and 263. These are the same test fibers in the 1 km PHASE IX cable. The attenuation shows

similar behavior as those of the 1 km cable. A large increase in attenuation as the temperature drops below freezing. The attenuation is shown in Figure 8. Again here also fiber 242 shows the best performance. The high attenuation of fiber 239 may be due to a bad connector or this particular segment of the fiber has such an attenuation.

The last group of 50 m fibers tested was those in PHASE V cable. The attenuation of these fibers are given in Figure 9. The four fibers test are identified by their colors and those of the loose tube. Two of the fibers did not show any change in attenuation as temperature changes while the other two showed slight increase in attenuation for temperatures below freezing. This leads to fact that either this cable does not shrink as that in PHASE IX or the fibers and in particular their primary coating are able to alleviate the effect of any induced stress. This cable continuously functions within attenuation specifications after five years of operation.

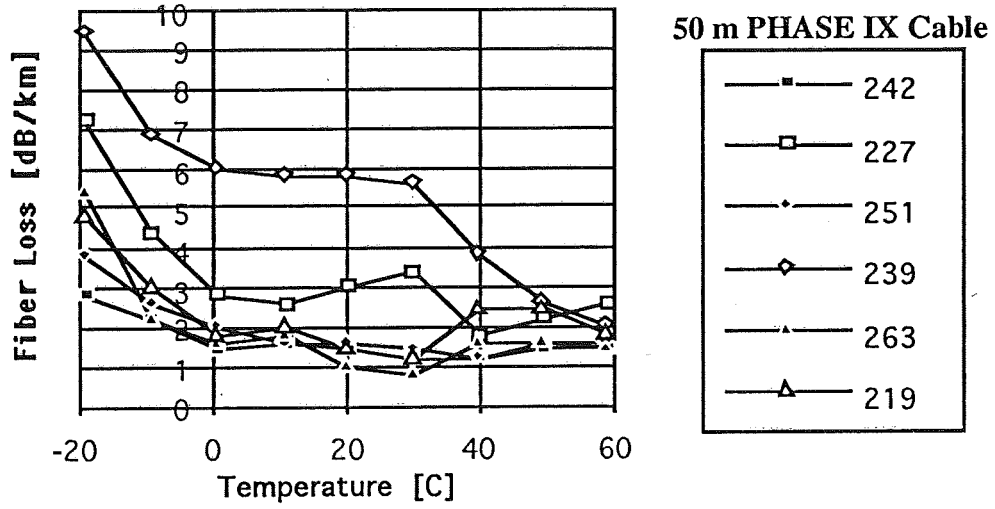


Figure 8

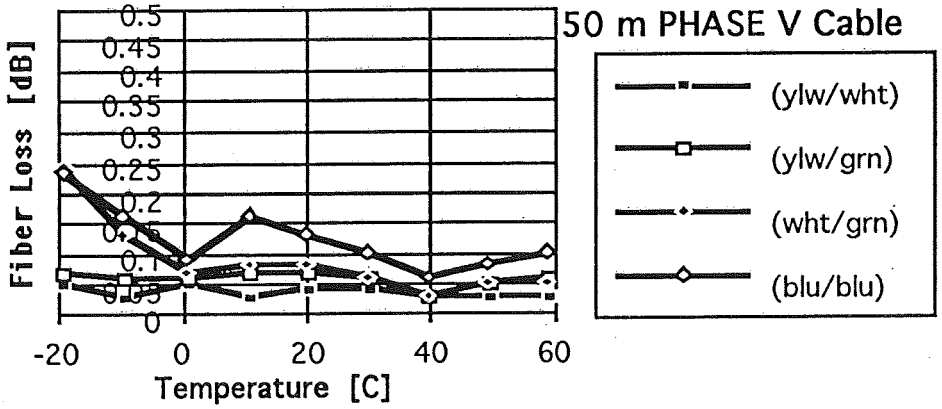


Figure 9

6. SUMMARY AND CONCLUSIONS

The increase in the fiber attenuation in PHASE IX cable installed at NASA's KSC was investigated and analyzed. Material and microbending tests were performed on a set of fibers from the defective cable and others. The chemical composition of the fibers showed that there is no difference between the fibers in the defective cable and others which will cause such an increase in attenuation. There is clearly a difference in attenuation performance between the fibers in PHASE IX cable and others which can be attributed to microbending. The losses of the fibers in the cable decreased when they were taken outside of the cable and left in their loose buffer tube. The increasing rate in attenuation for temperatures below freezing is much higher for cabled fibers. The primary coating of the

fibers in PHASE IX cable have more granularity than the test spool fibers which have a more homogeneous primary coating when it is examined under the scanning electron microscope. The reason behind this increase in microbending loss is the result of a number of factors, namely:

1. The shrinkage of the outer jackets of the cable due to aging and temperature change.
2. The degradation of the primary coating of the fiber by either aging or chemical reaction which might have taken place between the gel and the coating.
3. The shrinkage in the outer cable jackets after being stretched during installation. This might cause the cable core central member to be compressed and causes an outward pressure on the fibers.
4. An excess of fiber in the cable during cable manufacturing. This causes the fibers to be forced against each others in the loose buffer tubes.

The question which is still need to be answered is why the fiber attenuation was higher when it was installed? and what causes this increase to appear almost three years after installation? From the preceding tests and the different attenuation contributing factors this increase might be explained as follows. The cable when it was manufactured it has an inherent stress on the fibers caused by excess length in fiber and core strength member. This stress was within the tolerance range of the fiber coating so it did not show in the attenuation measurements. As the cable settled after installation and shrunk due to aging it induced an intolerable stress on the fiber which showed a sudden increase in attenuation. The increase of attenuation might be also a result of the deterioration of the fiber primary coating. This reasoning might justify the increase in attenuation after three years of use but it does not justify the decrease in attenuation resulted when the cable was removed from the duct. The reason which might justify this decrease is that when the fiber was installed it was stretched and by aging the shrinkage triggered the increase in attenuation. As soon as the fiber was removed from the duct the attenuation decreased because it was released from the tension it was under. In summary a number of the factors outlined earlier in this section contributed collectively to the attenuation increase.

REFERENCES

[1] D.L. Philen, "Measurements of OH Diffusion in Optical-Fiber Cores," Bell System Technical Journal, Vol. 61, No. 3, pp. 283-293 (1982).

[2] R.J. Araujo, "Model for Hydrogen Aging in Multimode Fibers," Journal of Light Wave Technology, Vol. 6, No. 2, pp. 197-202 (1988).

[3] S.R. Barnes and M.J. Pitt, "Prediction of Optical Cable Losses Due to Hydrogen," International Wire and Cable Symposium Proceedings, pp. 102-106 (1985).

[4] B. Wilshire and M.H. Reeve, "A Review of the Environmental Factors Affecting Optical Cable Design," Journal of Light Wave Technology, Vol. 6, No. 2, pp. 179-184 (1988).

[5] S. Tanaka and M. Honjo, "Long-Term Reliability of Transmission Loss in Optical Fiber Cables," Journal of Light Wave Technology, Vol. 6, No. 2, pp. 210-217 (1988).

[6] W.B. Gardner, "Microbending Loss in Optical Fibers," Bell Syst. Tech. J., Vol. 54, No. 2, pp. 457-465 (1975).

[7] D. Gloge, "Optical Fibre Packaging and its Influence on Fiber Straightness and Loss," Bell Syst. Tech. J., Vol. 54, No. 2, pp. 245-262 (1975).

[8] T. Kokubun et.al., "Microbending Loss Characteristics of Small Diameter Dual Coated Optical Fiber," Trans. Inst. Elec. Commun. Eng. Japan, Vol. J67-B, No. 6, p. 688 (1984).

[9] E. Suhir, "Stresses in Dual-Coated Optical Fibers," Journal of Applied Mechanics, Vol. 55, pp. 822-830 (1988).

[10] Hiroshi Murata, Handbook of Optical Fibers and Cables, Mercel Dekker, Inc., New York (1988).

1996 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

INTERFACING LABVIEW WITH INSTRUMENTATION FOR ELECTRONIC FAILURE ANALYSIS AND BEYOND

Mr. Randy K. Buchanan, Lecturer
Engineering Technology Department
Pittsburg State University
Pittsburg, Kansas

KSC Colleagues - Coleman Bryan and Larry Ludwig
Material Science

Contract Number NASA-NGT10-52605

August 8, 1996

Acknowledgments

I would like to thank Coleman Bryan for providing me the opportunity to engage in the research activities involved during my summer at Kennedy Space Center. I look forward to the opportunity to work with you more in the future.

Thanks also to Larry Ludwig for sharing his laboratory and office with me, and providing direction for this research. Without Larry, the final result would have been much less prolific. I hope my anal retentiveness was not too much of a nuisance.

Several people were very helpful throughout the summer in various aspect of providing support. Thanks goes to Kurt Leucht for all his computer and network support, and to Jeff Rauwerdink for helping to establish a "place" for me, as well as daily diversion.

Andy Finchum and Jon Bayless were instrumental in acquainting me with various areas of the laboratories and providing me with my first few days of entertainment. Thanks to Andy for your many personal tours of places where we probably should not have been -- they were the best. Thanks to Jon for supporting my efforts and being my first guinea pig (that's a compliment).

Steven Huff was also a help in day-to-day tasks, but was especially helpful in aiding me with visuals for my mid-term presentation. I had no idea graphics files could be so large.

Thanks to Bill Dearing for computer support and for servicing my E-mail account, sometimes on a daily basis. He was gracious in accepting that I wouldn't be working in his lab after my arrival.

Thanks to Farshid Sepehri, graduate student, for assistance during the first half of the summer.

Thanks to Gregg Buckingham for all your special effort in making this summer such a dynamic experience for all involved. Your work and efforts are noticed, and appreciated. Also thanks for taking the time to meet with me and my student.

Thanks to Kari Stiles and Richard Johnson for their summer-long efforts and behind the scenes coordination. The resultant program you helped produce was both informative and enjoyable.

Author Notes

Preparation of this manuscript was supported in part by a NASA Faculty Summer Fellowship sponsored by the University of Central Florida and the American Society for Engineering Education. The tutorial guide referenced throughout this paper is available directly from the author or from Larry Ludwig, NASA/LO-MSD-2E, Electronic Failure Analysis Laboratory, NASA Kennedy Space Center, FL 32899. Correspondence should be addressed to Randy Buchanan, Department of Engineering Technology, Pittsburg State University, Pittsburg, KS 66762, or call (316) 235-4370.

Abstract

The software, **Lab**oratory Virtual **I**nstrumentation **E**ngineering **W**orkstation (LabVIEW), is designed such that equipment and processes related to control systems can be operationally linked and controlled by the use of a computer. Various processes within the Failure Analysis Laboratories of NASA Kennedy Space Center demonstrate the need for modernization, and in some cases automation, using LabVIEW. An examination of procedures and practices within the Failure Analysis Lab resulted in the conclusion that some device was necessary to elevate the potential users of LabVIEW to an operational level in minimal time. This paper outlines the process involved in creating a tutorial application to enable personnel to apply LabVIEW to their specific projects. Suggestions for furthering the extent to which LabVIEW is used are provided in the areas of data acquisition and process control.

INTERFACING LABVIEW WITH INSTRUMENTATION FOR ELECTRONIC FAILURE ANALYSIS AND BEYOND

Randy K. Buchanan

1. INTRODUCTION

Various processes within the Failure Analysis Laboratories of NASA Kennedy Space Center demonstrate the need for modernization, and in some cases automation. The commonalty which exists with all processes is the need for computer control, either in the form of improving existing hardware and software, or the creation of computer control where none has previously been used. Laboratory Virtual Instrumentation Engineering Workbench (LabVIEW) software has been indicated by the Failure Analysis Laboratory to be a desirable solution to accomplish most of these tasks. In addition, LabVIEW is widely accepted in industry as a premier solution to facilitate or enhance computer control processes. Standardization utilizing this software would place the lab in a state-of-the-art continuum with leading technology in the field, allowing for quick/flexible control with the ability to easily adapt to changes and advances.

Processes targeted for modernization included the Slack Vacuum Chamber, Electrostatic Discharge Chamber, High-Amperage DC Test Fixture, High-Voltage AC Test Fixture, Electronic Failure Analysis Test Fixture, and any custom configurations for units under test. During initial investigations into these processes, it became apparent that the gap between LabVIEW programming and implementation must be bridged before modernization initiatives could be seriously undertaken.

2. BACKGROUND AND SIGNIFICANCE

LabVIEW is a powerful and flexible instrumentation and analysis software based on the programming language "G". The programming language "G" is in some ways similar to common languages, like C or BASIC, in that it is general purpose with extensive libraries of functions for any programming task including conventional program development tools. It is different however, in that it is graphical in nature. In addition, it includes extensive libraries of special functions, data acquisition, General Purpose Interface Bus (GPIB) and serial instrument control, data analysis, data presentation, and data storage. (ref. 1) It is one of the few languages that is platform independent, meaning that the same program can function in Windows on a PC, and the operating systems of UNIX, Macintosh, and Sun. Modular programming is also a feature which makes LabVIEW so attractive. Programs may be created with segments, much like sub-routines in a traditional language, in any order, and compiled into a singular high level program which functions as one. A unique characteristic of LabVIEW is its data-flow principle of processing inputs and outputs. Whereas typical languages use control-flow commands and a program counter to sequence through programming steps one at a time, LabVIEW processes input data as soon as all data is received at the proper input node. More than one data line may be processed at any given instant in time, as shown in Figure 1.

One should note that LabVIEW is not limited to General Purpose Interface Bus (GPIB) type communication. It is designed to operate over VME Extensions for Instrumentation (VXI) and serial mediums, as well as digital and analog data acquisition and process control systems.

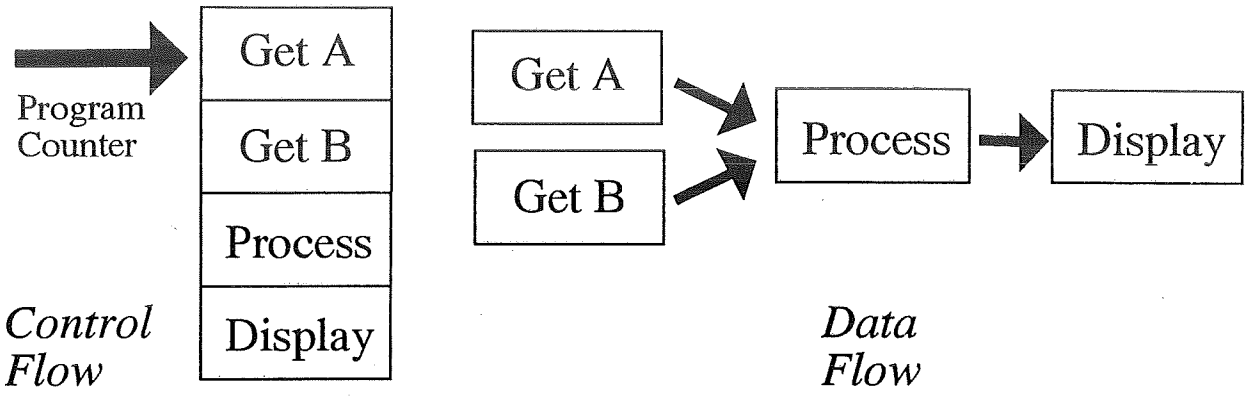


Figure 1. LabVIEW Data-Flow Principle (ref. 2, p. 78)

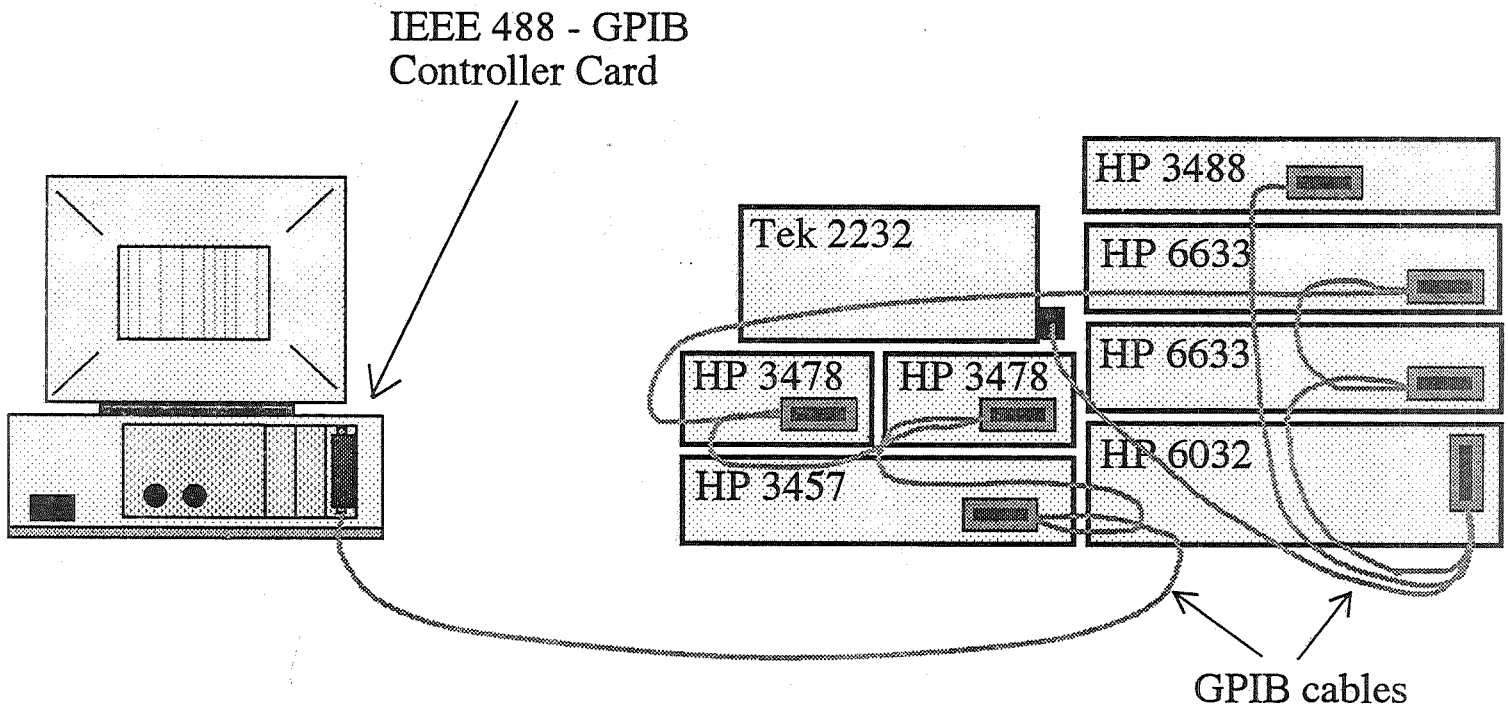
Many problems surround the implementation of a software such as LabVIEW. Learning curves for new software can be fairly flat, especially in the beginning. The time required to learn new software removes the engineer from the daily tasks at hand. Often, it is the case that the engineer will delay committing the time to learn the new software in lieu of work responsibilities. An examination of procedures and practices within the Failure Analysis Lab resulted in the conclusion that some device was necessary to elevate the potential users of LabVIEW to an operational level in minimal time. Thus became the scope of my fellowship.

Search of available literature and documentation revealed that there exists a lack of available training in the areas of LabVIEW implementation. In only one instance was it found that an attempt had been made to initiate software manipulation for simplification of LabVIEW software/hardware utilization. However the scope of this project, which was performed by a private contractor, is limited as its results have not been made available for public consumption. (ref. 3) Opportunities exist for training in terms of LabVIEW programming directly from National Instruments, the maker of LabVIEW. Training for implementation of the software with physical instruments, however appears to be non-existent. Therefore, it was deemed necessary to devise a method by which all lab personnel would be able to acquire the tools necessary to create applications using LabVIEW in each of their own areas. Ultimately, a step-by-step instruction reference procedure, designated "Tutorial Guide for Interfacing LabVIEW with GPIB Addressable Instrumentation" (unpublished guide, see Author Notes) was created to achieve this result. In addition, a computer with GPIB and LabVIEW capability was established with 8 GPIB addressable instruments, in an arrangement conducive to training and instruction.

3. APPARATUS

For the training apparatus, instrumentation operation and control was facilitated via the GPIB standard, while coalesced with a computer station possessing an IEEE 488.2 interface board, and operated under a LabVIEW environment. The Electronic Failure Analysis (EFA) Test Fixture (see Figure 2) was chosen as the system for the training apparatus. This test fixture was deemed to possess the most instrumentation which was common to all areas of the lab, especially in the Electrical/Electronic Failure Analysis area. More importantly, however, it was deemed that the EFA would be the most user-friendly system from which to learn.

The EFA Test Fixture was altered to support LabVIEW and GPIB communication. Equipment was coupled via GPIB parallel cables, from the IEEE-488 card in the computer, and sequentially to each piece of equipment in the fixture, as indicated in Figure 2.



Computer with LabVIEW

Gateway 2000 P5-60 Computer

National Instruments AT-GPIB/TNT
IEEE-488.2 Controller Card

LabVIEW for Windows Version 3.1

GPIB Instruments

Tektronix 2232 100 MHz Digital Storage Oscilloscope
Hewlett Packard 3478A Digital Multimeter (2)
Hewlett Packard 3457A Digital Multimeter
Hewlett Packard 3488 Switch/Control Unit
Hewlett Packard 6633A System DC Power Supply (2)
Hewlett Packard 6032A System Power Supply

Figure 2. LabVIEW and GPIB Instrumentation Interfacing Utilizing the Electronic Failure Analysis Test Fixture

Specific ordering of connections to the GPIB cables is not important, as long as all equipment is connected at some point to the GPIB "bus".

LabVIEW Virtual Instrument (VI) programs which are designed to operate specific equipment were obtained from National Instruments and modified or re-configured for the test fixture instruments. (for instructions refer to unpublished guide, see Author Notes) Most VIs were to designed to consume most, if not all, of the visible area of the computer monitor display screen. Since the training apparatus included 8 different instruments, a method of incorporating all instrument VIs into a single screen VI was necessary. Therefore, several VIs incorporating all 8 instruments were developed for the EFA fixture. Multiple VIs were made to suit users' various preferences.

Finally, GPIB communication was established, and all instruments verified for GPIB operation via LabVIEW. Hardware and software issues were resolved as necessary, and procedures for solving issues not specific to this particular fixture, were included in the tutorial guide (unpublished guide, see Author Notes).

4. TESTING

Once each instrument VI was configured to operate within the existing system, they were individually executed to assure proper operation. Errors which occurred, if any, were responded to and appropriate actions taken to resolve the problem. All individual VIs within the EFA user VI were initially tested for proper operation by first physically connecting the output of each of the three power supplies to the input of one of the three multimeters. The oscilloscope was also connected to one of the power supplies to provide an input for measurement. All VIs were concurrently ran after adjusting the power supplies for the desired voltage output. Upon program completion, the EFA user panel multimeter outputs were read to verify that their voltage measurements concurred with original power supply settings.

It was then important to provide a set-up which lab personnel could relate to, as well as train with. To provide for training on the test fixture, EFA employees were consulted for suggested training environments. Per recommendations, a unit-under-test arrangement was constructed involving a Marotta Solenoid Valve. The test Fixture was designed to provide a means to analyze the functioning of the valve, as well as to study GPIB control. The solenoid was powered from a 0 V resting position to a 28V dc activated state. The response voltage waveform was observed by connecting the oscilloscope to a resistor network as shown in Figure 3.

Specifications for the unit-under-test Marotta solenoid valve with a magnetic armature reference are as follows. The valve position is read by an onboard electronics module which provides a 28 volt output for an activated valve. The electronics module uses a Hall Effects Device (HED) sensor to activate a transistor driven output stage, providing either a 0.6 volt (or less) output or a 28 volt dc output according to the valve position. The two most important electronic valve parameters are the current waveform through the solenoid valve coil and the electronics module output waveform. From the current waveform the actual activation time of the valve can be determined, including the duration of armature movement. The output waveform from the electronics module, when captured at the same time as the current waveform, shows the delay time between armature movement and the electronics package response. It also shows whether the electronics package is providing a correct output per the valve armature's actual position. Operations and measurements were verified to be correct between LabVIEW and the actual instrument readings.

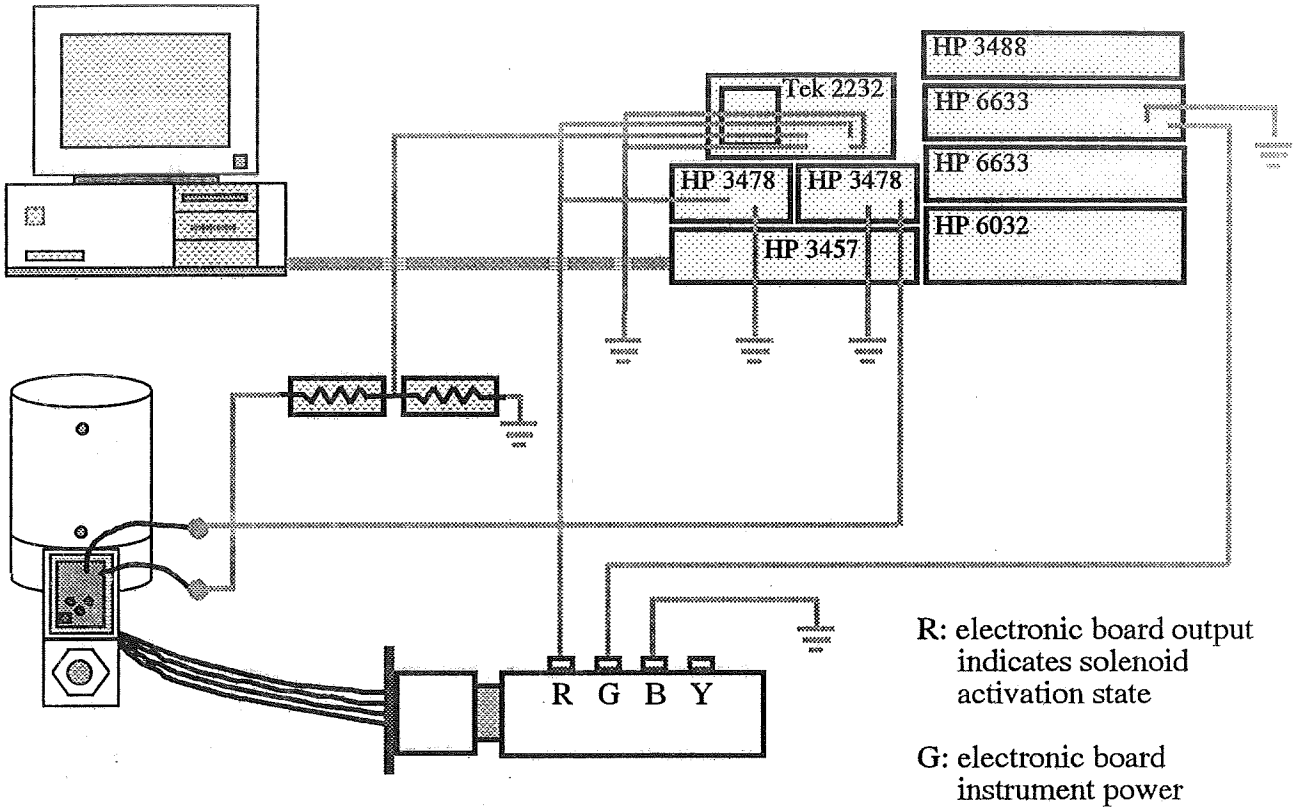


Figure 3. EFA Marotto Solenoid Valve Test Fixture

5. TRAINING

With the test configuration in place, real training opportunities for EFA employees to utilize LabVIEW in needed control systems is available. To this end, the EFA LabVIEW and GPIB Instrumentation Training Configuration has been placed into operation, and the tutorial guide (unpublished guide, see Author Notes) has been submitted for use by the Failure Analysis and Physical Testing Branch.

The tutorial guide provides background information into GPIB, including controller card requirements and system cable connection. Sources for VI acquisition are included, and most importantly, the LabVIEW File Transfer Protocol (FTP) Internet address for VI downloading is provided. Most of the step-by-step procedures outlined in the guide relate to creating customized application-oriented VIs. These procedures include instructions for creating both the user panel, which is visible to the operator via the computer screen, and the user diagram, or program, which actually enables the panel and the VI as a whole to function as desired. Illustrations of the relevant panels and diagrams are included to increase understanding and to provide a visual guide through the process.

Many opportunities exist for personnel to use the tutorial guide in their specific areas, whether or not related to failure analysis. The guide may be copied and distributed to the appropriate parties outside the EFA. The steps outlined in the tutorial deal with creating a method by which to facilitate the interfacing of LabVIEW with GPIB instruments, not with failure analysis itself. Essentially, the programming and communication methods are the same, regardless of the type of instrument which is being implemented.

The tutorial guide has already been utilized by laboratory personnel, with positive results. A member of the Electronic Failure Analysis Lab was successful in following the tutorial instructions and creating an application in a matter of hours. The process included downloading VIs from the Internet, re-configuring to fit his application, and programming a user interface panel which incorporated all the desired instruments into one panel. This particular engineer had no formal training in the use of LabVIEW previous to using the tutorial.

Much can be said as to the advantages of having hardware to "see and touch", when attempting to enlighten oneself on methods of programming and implementation. A system such as the Electronic Failure Analysis LabVIEW and GPIB Instrumentation Training Configuration provides personnel with an opportunity to train with actual instrumentation before them. The Marotto Solenoid Valve Test Apparatus was devised to provide a set-up which Electronic Failure Analysis Laboratory personnel could relate to, as well as train with. Similar relative apparatus could be established in other areas to provide motivation and a method to conduct training specific to an area. This not only holds greater interest, but also provides instant visual feedback as a method of indicating whether certain operations would or would not be successful.

6. DISCUSSION

Establishing a dependable reference test station requires alleviation of all real and potential problems. Implementation of the Electronic Failure Analysis LabVIEW and GPIB Instrumentation Training Configuration required a test of all associated wiring. Power supply panels, which provide voltage to the test fixture via voltage buses, required re-wiring to attain a standardized polarity and ordering in the form of a predictable formation. All GPIB cables were installed or re-routed to achieve reasonable use from the various cable lengths involved. Eventually one GPIB cable was found to be defective and was consequentially replaced. As such, physical communication interfaces should be established before investing time in the programming side of implementation.

Like any other operation which involves a computer programming, if all parameters are not met or are declared incorrectly, the system will not operate properly. Therefore, it is important to take time to learn of the details necessary to make the system work right the first time. It is fortunate, however, that LabVIEW tends to be much more forgiving than traditional programming languages, in the sense that many times it will help you locate the problem. Often it appears that it has foreseen the mistakes you may make and ultimately contains a built-in remedy. LabVIEW has a unique method of troubleshooting by which it steps though your program (even if it is not written sequentially) and visually indicates the numeric or string values of each node. This, as well as other features makes programming in LabVIEW enjoyable in comparison.

Although GPIB is the target medium of communication for implementing LabVIEW in this case, it is not the only medium available. Specific procedures outlined in the tutorial guide correlate with GPIB instrumentation. Similar methods of implementation could, however, be utilized for other environments. In fact, this GPIB configuration could be used as a model for development of other systems with minimal difficulty.

Limitations to the use of LabVIEW within a control system, such as for process control, would include the application to high-speed operations. Typically, internal timing functions should be limited to 0.1 second. (ref. 2, p. 31) This timing constraint varies between applications and depends upon the characteristics and type of platform (PC, Macintosh, etc.) which is being used. External "smart" controllers may be used to all but alleviate the timing issue. In such a system the controller assumes control responsibilities and LabVIEW provides the Man-Machine Interface (MMI).

6. CONCLUSIONS AND RECOMMENDATIONS

LabVIEW provides a solid yet flexible foundation for configuring computer controlled instrumentation. Instrument control using LabVIEW as covered is the fundamental tool necessary for various levels of testing and analysis. The tutorial guide developed for this purpose will enable laboratory personnel to create LabVIEW applications in less time than previously possible, by providing step-by-step instructions for acquisition, augmentation, and implementation of virtual instrument programs. The EFA fixture will provide the hands-on experience necessary for staff training and development, by serving as a model for others to emulate. Although the tutorial and EFA test fixture were designed for the Electronic Failure Analysis area, their application and use should not be limited to such. Procedures dealt with relate to the implementation of LabVIEW with instrumentation, not with failure analysis itself.

Beyond the realm of instrument control, many situations involve controlled and manipulated variables. LabVIEW-linked instrumentation, to the extent explored in this investigation, may be used to aid in the measurement and manipulation of some of these variables, but would fall short of providing full control. The inclusion of additional plug-in boards and/or the cooperative use of external controllers, would provide the ability to use LabVIEW to effect control over a particular process. Therefore, it is recommended that similar training in the areas of data acquisition and process control be developed.

REFERENCES

[1] National Instruments. LabVIEW Graphical Programming for Instrumentation: User Manual for Windows. National Instruments Corporation; Austin, TX; 1992.

[2] Johnson, Gary W. LabVIEW Graphical Programming: Practical Applications in Instrumentation and Control. McGraw-Hill, Inc.; New York; 1994.

[3] I-NET. Generic LabVIEW. KSC-11703 Kennedy Space Center Report; year unknown.

1996 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

ELECTROCHEMICAL IMPEDANCE SPECTROSCOPY OF CONDUCTIVE POLYMER COATINGS

Dr. Luz Marina Calle, Professor
Chemistry Department
Randolph-Macon Woman's College
Lynchburg, Virginia

KSC Colleague - Louis G. MacDowell
Material Science

Contract Number NASA-NGT10-52605

August 9, 1996

AKNOWLEDGEMENT

I would like to express my deepest appreciation to NASA/ASEE for providing me with the opportunity to participate in their summer faculty fellowship program. My professional development has been enhanced considerably due to my participation in the program. Gregg Buckingham, Dr. Roger W. Johnson, and Kari L. Stiles provided the leadership that makes this program such a wonderful experience for the faculty fellows. Gregg, with his enthusiasm and interest in the program, provided us with unique opportunities to learn about NASA and the Kennedy Space Center. Roger and Kari were always there for us when we needed them.

The people in the Materials Section of the Materials Science Division provided me with a warm and friendly environment during long hours of research. I am specially grateful to my NASA colleague, Louis MacDowell, for introducing me to the field of corrosion research. I treasure my collaboration with him in a project that is at the cutting edge of research in conductive polymer coatings.

The library staff at the Kennedy Space Center allowed me to find the background information I needed for my project in a very efficient and cordial manner. I'm also grateful to Lenee B. Washington, a student at Clemson University, for her invaluable help in conducting some of the experiments.

Finally, I would like to thank my husband, Carlos, for coping so patiently with the inconvenience of moving with me to Florida this summer.

ABSTRACT

Electrochemical Impedance Spectroscopy (EIS) was used to investigate the corrosion protection performance of twenty-nine proprietary conductive polymer coatings for cold-rolled steel under immersion in 3.55% NaCl. Corrosion potential as well as Bode plots of the data were obtained for each coating after one hour immersion in 3.55% NaCl. All coatings, with the exception of one, have a corrosion potential that is higher in the positive direction than the corrosion potential of bare steel under the same conditions. The EIS spectra of the twenty-nine coatings were characterized by an impedance that is higher than the impedance of bare steel. The twenty-nine coatings were classified into two groups based on their EIS spectra. Group A consisted of twenty-one coatings with Bode plots indicative of the capacitive behavior characteristic of barrier coatings. An equivalent circuit consisting of a capacitor in series with a resistor simulated the experimental EIS data for these coatings very well. Group B consisted of eight coatings that exhibited EIS spectra showing an inflection point which indicates that two time constants are present. This may be caused by an electrochemical process taking place which could be indicative of coating failure. These coatings have a lower impedance than those in group A.

ELECTROCHEMICAL IMPEDANCE SPECTROSCOPY OF CONDUCTIVE POLYMER COATINGS

Luz Marina Calle

1. INTRODUCTION

Electrically conductive polymers have attracted a great deal of interest since their discovery about two decades ago. Soon after their discovery, it became clear that their unique properties could be used in several technological applications, such as the development of a new class of superconductors, light emitting plastics, polymer-based switching devices, and sensing devices. The sheer volume of fundamental and applied research in this field makes it inevitable that conductive polymers will find an increasing range of applications.[1]

Polyaniline (PAN) has attracted much attention as a unique electrically conductive polymer.[2] Many publications revealed that this material exhibits unusual electrical and optical properties. The reversibility of these properties combined with its good environmental stability and its low cost of production makes this polymer suitable for the development of the aforementioned technological applications.[3] Another possible application of PAN involves its use in protecting metals and semiconductors from corrosion.[4,5] Investigations aimed at following the improvement of this application are justified.

In the mid 1980's, researchers at the Kennedy Space Center (KSC) in Florida became interested in PANs as protective coatings for metallic surfaces. Research has been ongoing for over 20 years at KSC to find coating materials to protect launch site structures and equipment from the extremely corrosive conditions present at the launch complexes. The combination of proximity to the Atlantic ocean and acidic combustion products from solid rocket boosters results in corrosive stresses unique to KSC.

Extensive coating testing at KSC lead to the conclusion that inorganic zinc-rich primers (ZRPs) significantly outperformed organic zinc-rich type primers in the marine atmosphere of Florida. This was partially attributed to the increased conductivity of the inorganic ZRP coating film. The materials typically used to produce the organic zinc-rich films (e.g., epoxies, vinyls, etc.) produced an undesirable insulating effect on the zinc particles. This effect resulted in decreased galvanic activity of the zinc for protection of the carbon steel substrate. On the other hand, the organic zinc-rich primers had the advantage of allowing for less than perfect surface preparation on steel to achieve performance. The organic polymers provided better adhesion to marginally prepared substrates than the inorganic materials. This result led researchers at KSC to the idea of using conductive organic materials to formulate these zinc coatings in order to develop a conductive organic zinc-rich primer. The idea being that the conductive organic vehicle would provide both the increased conductivity needed for superior galvanic protection of the steel substrate and the better adhesion to less than perfectly prepared surfaces. Hence the work on

conductive organic polymers and the search for materials that would allow the production of a new generation of protective coatings based on this technology began.

The main objective of this work was to use Electrochemical Impedance Spectroscopy (EIS) to study recently developed proprietary conductive polymer coatings. These materials are either doped forms of PAN or short chain compounds based on PAN. The compounds were applied on cold-rolled steel panels as corrosion protective coatings. Corrosion potential versus time measurements were collected for an hour followed by EIS measurements. Both sets of measurements were obtained using 3.55% (w/w) sodium chloride (NaCl) solution as the electrolyte.

2. EXPERIMENTAL PROCEDURE

Test Samples

Specimens were flat cold-rolled steel type S panels (Q Panel Corporation, Cleveland) 20.2 cm (7.8 in) x 10.1 cm (4.0 in) coated on one side with a conductive polymer formulation supplied by Dr. Ed Weil from Polytechnic University in New York, NY. The panels were photographed for documentation purposes prior to any testing or exposure. Coating thickness measurements were performed using a magnetic pull-off thickness gauge and found to be in the range between 1.2 and 2.0 mils (.003 cm - .005 cm).

Corrosion Potential Measurements

Corrosion potential measurements were performed using a system manufactured by EG&G Princeton Applied Research Corporation. The system used includes: (1) the Model 273 Computer-Controlled Potentiostat/Galvanostat, (2) the Model 5210 Computer-Controlled Lock-In Amplifier, and (3) the Model 352/342 SoftCorr™ II Corrosion Measurement Software. The electrochemical cell included a Ag/AgCl (silver/silver chloride) electrode, a platinum counter electrode, the sample working electrode, and a bubbler/vent tube. The flat specimen holder in the electrochemical cell is designed such that the exposed surface area is 1 cm^2. Corrosion potential values were gathered for one hour in aerated 3.55% NaCl. All solutions were prepared using deionized water. Aeration with dry air was maintained throughout the test.

Electrochemical Impedance Measurements

A Model 378 Electrochemical Impedance system manufactured by EG&G Princeton Applied Research Corporation was used for all EIS measurements. The system includes: (1) the Model 273 Computer-Controlled Potentiostat/Galvanostat, (2) the Model 5210 Computer-Controlled Lock-In Amplifier, and (3) the Model M388 Electrochemical Impedance Software including circuit modeling routines. Data were gathered in the frequency range from 100 kHz to 0.01 Hz. Three experiments were performed in a sequence covering the specified frequency range, and the data were automatically merged and saved. The frequency ranges for the three experiments were

100 kHz to 5 Hz, 10 Hz to 0.1 Hz, and 0.1 Hz to 0.01 Hz. The AC amplitude used for the experiments was 10 mV. Each sample was studied after one hour immersion in aerated 3.55% NaCl.

Bode magnitude plots of the data (showing the logarithm of the modulus of the impedance, $\log|Z|$, as a function of the logarithm of frequency and phase angle, alpha in degrees, as a function of the logarithm of frequency) were obtained for each coating after one hour of immersion in 3.55% NaCl. The impedance data were analyzed using the Equivalent Circuit computer simulation program by B. A. Boukamp.[6]

3. RESULTS AND DISCUSSION

Table 1 shows the thickness and color of the coatings as well as the corrosion potential and the type of EIS spectra after one hour of immersion in 3.55% NaCl. All coatings, with the exception of coating 16, exhibit a corrosion potential after one hour of immersion in 3.55% NaCl that is higher in the positive direction than the corrosion potential of steel under the same conditions. Similar results have been reported for PAN coated metals like iron, steel, and stainless steel in salt water[7].

The EIS data for all 29 coated and 1 uncoated panels were plotted in the Bode format. A qualitative comparison of all plots revealed that the coated panels have a higher impedance than the uncoated panel. Figure 1 shows the Bode plot of one of the coatings in comparison with the Bode plot obtained for an uncoated panel. Furthermore, the impedance spectra of all panels can be classified into two groups (labeled as A and B in Table 1): 21 coatings exhibit the characteristic Bode plot of a barrier coating (group A). Figure 2 shows Bode plots for several coatings in group A. These spectra were analyzed to find the corresponding equivalent circuit. The best fit was obtained for a circuit composed of a resistor in series with a capacitor (Figure 3). Figure 4 shows the experimental as well as the fitted data using the aforementioned equivalent circuit. The resistance has a value of 413.67 ohms and the capacitance was determined to be 0.103 nF. After one hour of immersion, the coatings in group A behaved as purely capacitive dielectrics. This behavior is similar to what has been found for epoxy or polyester coated steel panels in 3.5% NaCl for up to 1000 hours of immersion.[8] Bode plots for the remaining 8 coatings show an inflection point which indicates that two time constants are present (group B). This may be caused by an electrochemical process taking place which could be indicative of coating failure. These coatings have a lower impedance than those in group A. Figure 5 shows Bode plots for selected coatings in group B (a group A coating was included in the figure for comparison). Due to complexity of the spectra and time constraints, the analysis of these plots will be performed at a later time.

Table 1. Thickness, color, corrosion potential, and EIS group of conductive polymer coatings

Panel No.	Thickness mils	Color	E_{corr} Volts	EIS Group
1	1.0	Clear	-0.4570	A
2	3.0	Clear	0.0090	B
3	3.0	Clear	-0.2210	A
4	1.0	Amber	-0.3860	A
5	2.0	Gray	0.0030	A
6	3.0	Clear	0.0580	B
7	1.0	Green	-0.0600	A
8	1.0	Gray	-0.0010	A
9	1.0	Clear	-0.0990	A
10	2.0	Green	-0.0080	B
11	1.0	Clear	-0.2090	A
12	1.0	Clear	-0.0900	A
13	1.0	Gray	-0.0570	A
14	2.0	Blue	-0.4560	A
15	3.0	Blue	-0.3100	B
16	1.0	Amber	-0.6380	A
18	3.0	Amber	-0.1110	A
20	3.0	Amber	-0.1600	A
19	1.0	Amber	-0.1260	A
20	3.0	Black	0.0020	B
21	3.0	Clear	-0.3040	B
22	1.0	Gray	-0.1320	B
23	1.0	Clear	-0.4450	B
24	1.0	Black	0.0230	A
25	1.0	Gray	-0.1160	A
26	1.0	Black	-0.4790	A
27	1.2	Clear	-0.1730	A
28	3.0	Amber	-0.1440	A
29	1.0	Green	-0.1400	A
Steel			-0.5400	

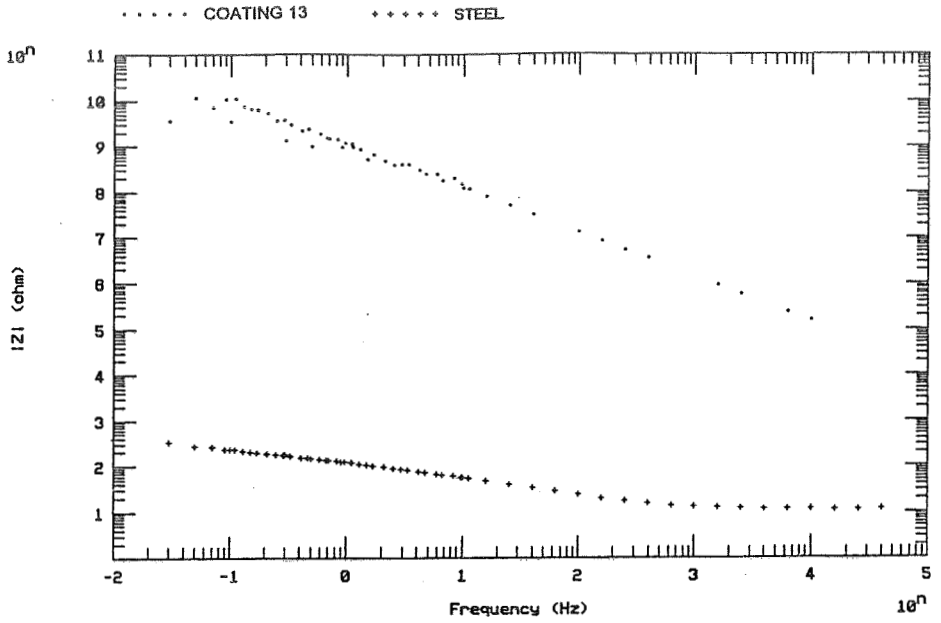


Figure 1. Bode plots for coating 13 and bare steel

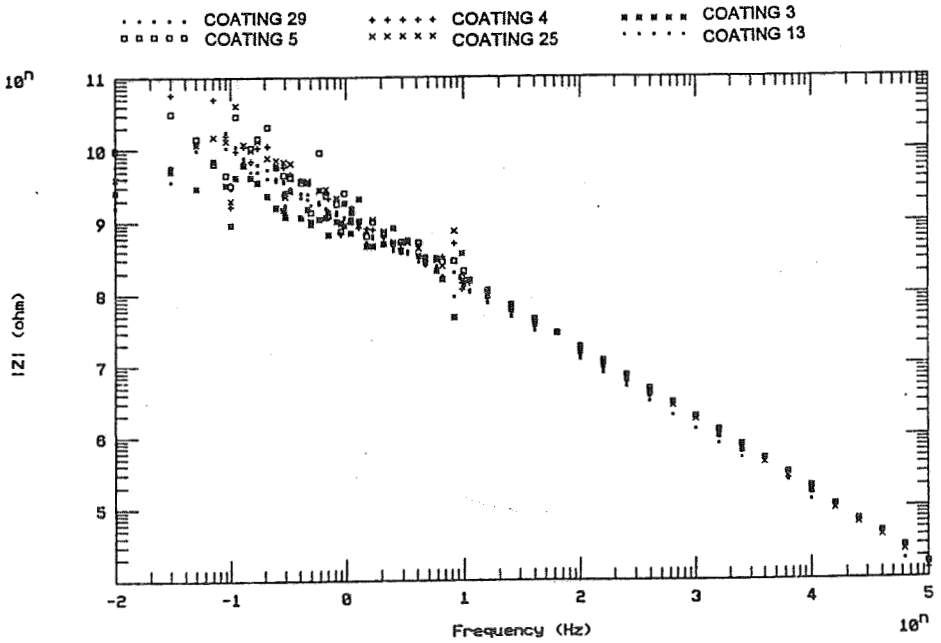


Figure 2. Bode plots for several coatings in group A

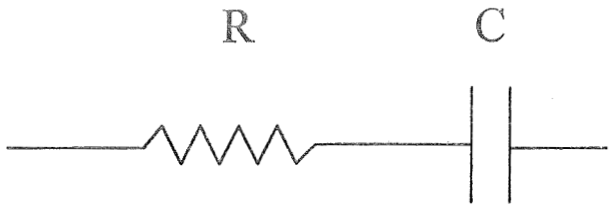


Figure 3. Equivalent circuit for Group A conductive polymer coatings

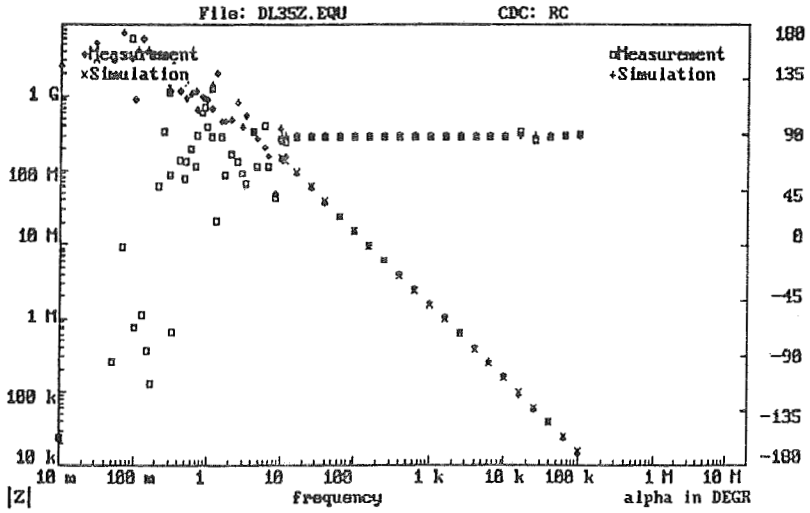


Figure 4. Experimental and simulated EIS data for conductive polymer coating 3

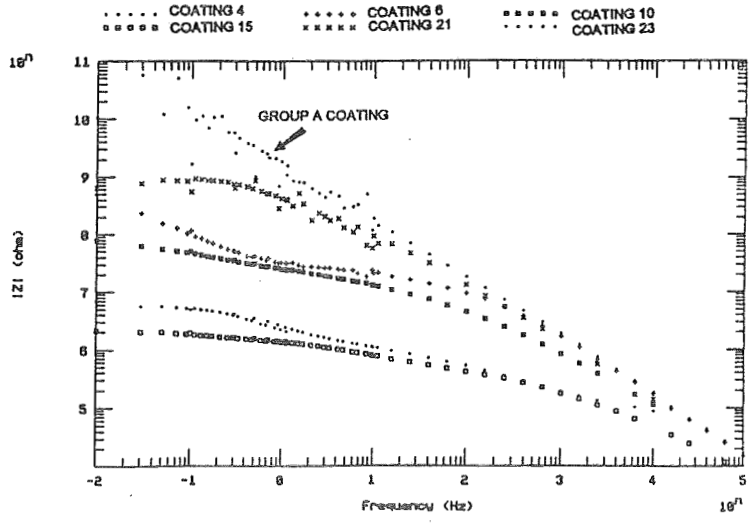


Figure 5. Bode plots for one coating in group A and several coatings in group B

4. CONCLUSION

Corrosion potential measurements of the twenty-nine coatings after one hour of immersion in 3.55% NaCl indicate that all but one of the coatings exhibit a corrosion potential that is shifted in the direction of noble metals.

EIS measurements showed that the coated panels have a higher impedance than bare steel. Qualitative observation of the Bode plots, allowed the classification of the coatings into two groups: coatings in group A exhibit an EIS spectra indicative of capacitive behavior; coatings in group B exhibit an EIS spectra with an inflection indicative of an electrochemical process. The experimental EIS data for coatings in group A can be simulated by using an equivalent circuit including a resistor in series with a capacitor. The EIS spectra of coatings in group B will be analyzed to find an equivalent circuit at a later date. Experiments using longer immersion times are currently underway.

REFERENCES

[1] S. Haddlington, Chemistry and Industry 1 (1995): p. 11.

[2] M. Morita, Journal of Applied Polymer Science 52 (1994): p. 711.

[3] G. D'Aprano, M. Lecrerc, G. Zotti, G. Schiavon, Chem. Mater. 7 (1995): p. 33.

[4] G. Mengoli, M.M. Musiani, B. Pelli, E. Vecchi, J. Appl. Polym. Sci. 28 (1983): p. 1125.

[5] R. Noufi, A. J. Nozik, J. White, L.F. Warren, J. Electrochem. Soc. 129 (1982): p. 2261.

[6] B. Boukamp, "Equivalent Circuit Programme" v. 3.99, University of Twente, The Netherlands (1992).

[7] B. Wessling, Adv. Mat. 6 (1994): p. 226.

[8] T.A. Strivens and C.C. Taylor, Mater. Chem., 7 (1982): p. 199.

1996 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

ELECTROLYTIC REMOVAL OF NITRATE FROM CELSS CROP RESIDUES

Dr. Guillermo Colon, Associate Professor
Chemical Engineering Department
University of Puerto Rico
Mayaguez, Puerto Rico

KSC Colleague - John C. Sager
Life Sciences

Contract Number NASA-NGT10-52605

August 9, 1996

ACKNOWLEDGMENTS

I would like to express my appreciation for being selected to participate in the 1996 NASA/ASEE Summer Fellowship Program. To Dr. Roger W. Johnson (University of Central Florida) and to Mr. Gregg Buckingham (NASA University Program Director) my congratulations, for their capable administration of the program.

I also want to thank my NASA colleague Dr. John C. Sager for his invaluable support and friendship during this summer. Special thanks are extended to Mr. Barry Finger (excellent Engineering Lead), Mr. Michael Alazraki, Mr. Joe Martinez, Jay L. Garland, and Dr. Richard F. Strayer for their continuous support and arranging of the laboratory facilities to carry out the experimental work reported in this document. Thanks to Teresa Englert for their support with the analyses of inorganics and TOC, and Mrs. Callen R. Coiner for her attention expressed to me.

My gratitude goes also to Ms. Kari Stiles for her dedication and invaluable assistance during my stay in KSC, to Dr. Choul-Gyun Lee for his friendship and numerous rides from Titusville to NASA-KSC facilities, and to Miss. Ivonne Díaz-Brufau (Environmental Engineering undergraduate student, University of Central Florida) for her assistance to carry out the experimental work.

In general, I made lots of friends in NASA Kennedy Space Center that I will never forget.

ABSTRACT

The CELSS resource recovery system, which is a waste processing system, uses aerobic and anaerobic bioreactors to recover plants nutrients and secondary foods from the inedible biomass. Studies have been carried out using aerobic and anaerobic bioreactors to recover plants nutrients and secondary foods from inedible crop residues. Crop residues contain significant amount of nitrate. There are actually two major problems concerning nitrate: 1) both CELSS biomass production and resource recovery consume large quantities of nitric acid, 2) nitrate causes a variety of problems in both aerobic and anaerobic bioreactors. Previous studies with aerobic biodegradation have experienced nitrogen losses through denitrification when low level of dissolved oxygen are reached due the rapid organic matter degradation. During the anaerobic biodegradation of inedible biomass, nitrate is converted to ammonia which reduces the system capacity to convert the organic matter to volatile fatty acids and needs further downstream processing steps to convert the ammonia back to nitrate. The nitrate anion causes several problems in the resource recovery system in such a way that, removal prior to the process is highly desirable.

The technique proposed to remove the nitrate from potato inedible biomass leachate and to satisfy the nitric acid demand was a four compartment electrolytic cell. In order to establish the electrolytic cell performance variables, a series of experiments were carried out using potato crop residues aqueous leachate as the diluate solution. The variables studied were the potato biomass leachate composition and electrical properties, preparation of compartment solutions to be compatible with the electrolytic system, limiting current density, nutrients removal rates as a function of current density, fluid hydrodynamic conditions, applied voltage, and process operating time during batch-recirculation operation.

Results indicated that the limiting current density (maximum operating current density) was directly proportional the solution electrical conductivity an a power function of the linear fluid velocity in the range between 0.083 to o.403 m/s. During the electrolytic cell once-through operation, the nitrate, potassium, and other nutrients removal rates were proportional to the current density and inversely proportional to fluid velocity. The removal of monovalent ions was found to be higher than divalent ones. Under batch-recirculation operation at constant applied voltage of 4.5 and 8.5 volts, it was found that the nutrient removal rates were independent of applied voltage, but proportional to the ions concentration and operating time. From material balances it was found that 2.2 moles of oxygen gas and 4.4 moles of hydrogen gas were produced at the electrodes surfaces per mole of nitrate transferred.

SUMMARY

The National Aeronautical and Space Administration (NASA) has been studying the viability of Controlled Ecological Life Support System (CELSS) for long term space missions for several years. The purpose of this system is to provide all the basic human needs required for life support. The CELSS breadboard project is a large-scale integration system with two main components, a full-scale Biomass Production Chamber (BPC) and the Resource Recovery system. The BPC is a closed system where plants are grown under controlled hydroponic conditions. The plants in CELSS generate oxygen, purified water, and produce edible foods.

The CELSS resource recovery system, which is a waste processing system, uses aerobic and anaerobic bioreactors to recover plants nutrients and secondary foods from the inedible biomass. Studies have been carried out using aerobic and anaerobic bioreactors to recover plants nutrients and secondary foods from inedible crop residues. Crop residues contain significant amount of nitrate. There are actually two major problems concerning nitrate: 1) both CELSS biomass production and resource recovery consume large quantities of nitric acid, 2) nitrate causes a variety of problems in both aerobic and anaerobic bioreactors. Previous studies with aerobic biodegradation have experienced nitrogen losses through denitrification when low level of dissolved oxygen are reached due the rapid organic matter degradation. During the anaerobic biodegradation of inedible biomass, nitrate is converted to ammonia which reduces the system capacity to convert the organic matter to volatile fatty acids and needs further downstream processing steps to convert the ammonia back to nitrate. The nitrate anion causes several problems in the resource recovery system in such a way that, removal prior to the process is highly desirable.

The technique proposed to remove the nitrate from potato inedible biomass leachate and to satisfy the nitric acid demand was a four compartment electrolytic cell. In order to establish the electrolytic cell performance variables, a series of experiments were carried out using potato crop residues aqueous leachate as the diluate solution. The variables studied were the potato biomass leachate composition and electrical properties, preparation of compartment solutions to be compatible with the electrolytic system, limiting current density, nutrients removal rates as a function of current density, fluid hydrodynamic conditions, applied voltage, and process operating time during batch-recirculation operation.

Results indicated that the limiting current density (maximum operating current density) was directly proportional the solution electrical conductivity an a power function of the linear fluid velocity in the range between 0.083 to o.403 m/s. During the electrolytic cell once-through operation, the nitrate, potassium, and other nutrients removal rates were proportional to the current density and inversely proportional to fluid velocity. The removal of monovalent ions was found to be higher than divalent ones. Under batch-recirculation operation at constant applied voltage of 4.5 and 8.5 volts, it was found that the nutrient removal rates were independent of applied voltage, but proportional to the ions concentration and operating time. From material balances it was found that 2.2 moles of oxygen gas and 4.4 moles of hydrogen gas were produced at the electrodes surfaces per mole of nitrate transferred.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

ABBREVIATIONS, ACRONYMS AND NOMENCLATURE

BPC = Biomass Production Chamber
CELSS = Controlled Ecological Life Support System
ED = Electrodialysis

A = empirical constant
A_m = effective membrane area, m^2
b = empirical number, dimensionless
C = bulk solution ion concentration, Kg-equiv./m^3
C_0 = boundary layer solution ion concentration, Kg-equiv./m^3
e = current efficiency, dimensionless
F = Faraday's constant, 96,500 Coul./g-equiv.
i = current density, A/m^2
i_{lim} = limiting current density, A/m^2
I = current, A
J_e = ion flux by electrotransport, Kg-equiv./s-m^2
J_D= ion flux by diffusion, Kg-equiv./s-m^2
k = specific electrical conductivity, Mho/m
N_f = diluate stream feed concentration, kg-equiv./m^3
N_p = diluate stream product concentration, kg-equiv./m^3
Q = volumetric flowrate, m^3/s
t_m = ion transport number in solution, dimensionless
t_s = ion transport number in membrane, dimensionless
U = linear fluid velocity, m/s

δ = boundary layer thickness, m

I. INTRODUCTION

The National Aeronautical and Space Administration (NASA) has been studying the viability of Controlled Ecological Life Support System (CELSS) for long term space missions for several years. The purpose of this system is to provide all the basic human needs required for life support. The CELSS breadboard project is a large-scale integration system with two main components, a full-scale Biomass Production Chamber (BPC) and the Resource Recovery system. The BPC is a closed system where plants are grown under controlled hydroponic conditions. The plants in CELSS generate oxygen, purified water, and produce edible foods.

The CELSS resource recovery system, which is a waste processing system, uses aerobic and anaerobic bioreactors to recover plants nutrients and secondary foods from the inedible biomass. Studies have been carried out using aerobic and anaerobic bioreactors to recover plants nutrients and secondary foods from inedible crop residues. Crop residues contain significant amount of nitrate. There are actually two major problems concerning nitrate: 1) both CELSS biomass production and resource recovery consume large quantities of nitric acid, 2) nitrate causes a variety of problems in both aerobic and anaerobic bioreactors. Previous studies with aerobic biodegradation have experienced nitrogen losses through denitrification when low level of dissolved oxygen are reached due the rapid organic matter degradation. During the anaerobic biodegradation of inedible biomass, nitrate is converted to ammonia which reduces the system capacity to convert the organic matter to volatile fatty acids and needs further downstream processing steps to convert the ammonia back to nitrate. The nitrate anion causes several problems in the resource recovery system in such a way that, removal prior to the process is highly desirable.

1.1 BACKGROUND

Previous studies associated with the CELSS Breadboard project indicated that the water soluble extract of inedible biomass have been suitable for use as plant nutrient source in hydroponic systems (1,2). Water extraction or leaching, can be viewed as one step in an overall biological oxidation scheme (3), or as a pretreatment of biomass to remove nutrient prior to further treatment. Leaching was effective at removing the majority of nutrients from the inedible biomass (2). The aqueous leachate contains both inorganic and organic molecules. Biological pretreatment of leachate prior to its use in hydroponic systems have been recommended to both reduce phytotoxic effects and to convert soluble organic to edible products in regenerative systems (4). Based on the inorganic analyses of crop residues, leachate contains higher levels of nitrate and potassium when compared to others nutrients (5). Three Intermediate-Scale Aerobic Bioreactors have been designed, fabricated and operated (6). They utilized mixed microbial communities to bio-degrade plant residues. It was determined that the soluble organic compounds in the leachate were metabolized by microorganisms resident in the nutrient delivery system. It was found that this lead to increase total root respiration, the development of microbial biofilms on surfaces of roots and nutrient delivery system components, and increased denitrification (conversion of nitrate to nitrogen) by the rhyzosphere microbial community. It

was also observed that the aerobic decomposition of soluble organic compounds, the pH increased from 6.5 to 8.5. To control the pH during crop residues decomposition, a 1 N nitric acid solution was injected into the bioreactor.

Anaerobic digestion of inedible biomass is a natural occurring biological process in which organic material is converted to reduce end products. The anaerobic decomposition of solid organic wastes proceeds in two consecutive steps, liquification and methanogenesis (7,8). The first phase of the anaerobic digestion is known as the acidigenic phase (volatile fatty acids formation). In the second phase a second group of organisms, the methanogens, complete the digestion process converting the volatile fatty acids to highly reduced CH_4 and CO_2. The presence of highly levels of nitrate in solution during the anaerobic decomposition of organic matter conducts to the undesired formation of ammonia. The aerobic dinitrification process consumes part of the medium dissolved oxygen, which is highly required to solubilize the organic matter present. The undesired anaerobic decomposition of nitrate to ammonia needs further nitrification process to convert back the ammonia to nitrate to be use as plant's nutrient.

Many researchers (9, 10, 11, 12) have described processes for producing weak acids from its salt via electrodialysis (ED), but do not address the process variables, e.g., the limiting current density, the membrane fouling effect, and hydrodynamic conditions on the solution-membrane mass transfer. Theoretical concepts and experimental research on the area of producing acids from its salts and desalination of saline water by electrodialysis gives better understanding of the process variables (9, 13, 14, 15, 16, 17). In a typical desalination system, the channels through which saline water passes are made of alternating cation and anion permselective membranes. The passage of a direct current through the stack of channels results in the depletion and enrichment of the salt in the alternating compartments. The rate of ion transfer is proportional to the current density. However, the current density at which the system can be operated is to a large extent limited by the polarization which take place in the lialysate channels. Polarization is a current saturation condition associated with a high ohmic resistance of the fluid because of the decreased ionic concentration. Theoretically, the limiting current is reached when the ionic concentration, which is lowest at the membrane surface, approaches zero. In practice, this is not attained because water ions themselves enter the transfer process. It has been long understood (13) that increasing the turbulence level in the fluid streams generally increases the limiting current by means of decreasing the fluid boundary layer thickness. This results from the increased rate of diffusion of ions towards the membrane surface to replenish those which have been depleted. It is normal to operate practical systems under turbulent conditions, through not excessively high Reynolds numbers, because of the associated higher pumping costs and inter-spacers leaking. In, practice, the Reynolds number may be low enough to indicate that the flow would be laminar, but turbulent promoters used result in a strong eddying motion.

Mass transfer through permselectives membranes consists of two steps:
1) the reduction of ionic concentration in the solution by electrotransport of ions from the boundary layer near the membrane and 2) the diffusion of ions to the partially de-ionized boundary layer. The kinetics of the first step is given by the Nernst equation

$$J_e = (t_m - t_s)/(i\ F) \quad (1)$$

where J_e is the flux of ions by electrotransport, I is the current density, F is the Faraday's number, t_s is the ion transport number in solution, and t_m is the ion transport number on membrane. The second step is given by the Fick's first law:

$$J_D = D(C-C_0)/\delta \quad (2)$$

where J_D is the flux of ions by diffusion, D the diffusion coefficient, C the ion concentration of bulk solution, C_0 the ion concentration at the boundary layer, and δ the thickness of the boundary layer. The thickness of the boundary layer is a function of the fluid velocity and geometry of the spacers (18). From the equations (1) and (2), the following can be derived:

$$i = DF(C-C_0)/\delta(t_m-t_s) \quad (3)$$

Increasing the voltage of the stack raises the current density. The flux of ions by electrotransport is also increased until the concentration of the solution boundary layer approaches zero ($C_0 = 0$). Under these conditions the flux of ions by diffusion is maximal; therefore, the current density is the limiting one ($i = i_{lim}$)

$$i_{lim} = DFC/\delta(t_m-t_s) \quad (4)$$

A further increase in J_D can be achieve only by raising the fluid velocity of the solution in the cell to a level at which the pressure drop across the cell will not cause internal leakage. A further increase in the stack voltage will raise the current density. Most of this additional current will cause dissociation of water rather than mass transfer from the diluate cell compartment. When the concentration of the solution in the boundary layer decreases, the electrical resistance of cell dilate solution increases. Therefore, it is important that the current density be prevented from approaching the limiting current-density value. A practical equation for limiting current density, derived from empirical results and suggested by several authors (19,20) is

$$i_{lim}/k = AU^b \quad (5)$$

where U is the fluid velocity, A is an empirical constant number, b in an empirical number, generally between 0.5 to 0.9, and k is diluate solution average specific electrical conductivity.

Material balances over the membrane area across the diluate solution is flowing through,

and in which a direct current is being passed, the following equation can be obtained for a electrolytic cell continuous mode of operation (9):

$$\ln\frac{N_f}{N_p} = \left[(\frac{e}{F})(\frac{i}{N_d})(\frac{A_m}{Q})\right] \quad (6)$$

From an empirically determined relationship between (I/N_d) and the fluid velocity, values of (I/N_d) and A_m may be chosen. For batch recirculation electrolytic cell mode of operation, it is apparent that multiple passes through a single stack are equivalent as to passing the solution through several stacks in series. The performance equation for a batch recirculation process is as follow:

$$\frac{N_f}{N_{nn}} = \exp\left[\frac{A_m}{FQ}\sum_{k=1}^{n}(\frac{i}{N_d})_k e_k\right] \quad (7)$$

If the process is carried out at constant average value of $(I/N_d)_{AV}$ and assuming an average current efficiency, $e_{av,}$ Eq. 7 can be simplified to:

$$\ln\frac{N_f}{N_{pn}} = (\frac{ne_{av}}{F})(\frac{i}{N_d})_{av}(\frac{A_m}{Q}) \quad (8)$$

The value of $(I/N_d)_{av}$ may be simulated partially by operating at a constant voltage determined empirically from polarization studies.

Permselectives membranes for electrodialysis (ED) contain either group of positive ions (anion-exchange membrane) or negative ions (cation-exchange membrane). In an applied electric field and in aqueous solution of ions, an anion-exchange membrane permits the passage of anions; a cation-exchange membrane permits the passage of cations only. The most important characteristics of permselectives membranes used for ED are: low electrical resistance, good permselective qualities for cations and anions, good mechanical stability, and high chemical stability. A nitrate-specific anion-exchange membrane was prepared from chloromethylated polysulfone by partially aminating it with a secondary amine for completion (21). When the nitrate-chloride ratio in solution was equal, nitrate flux was larger than chloride by a factor of 1.6 in alkaline medium and a factor of 2 in acid medium.

1.2 PROPOSED TECHNOLOGY

The technique proposed to remove the nitrate from potato inedible biomass leachate and

to satisfy the nitric acid demand is a four compartment electrolytic cell (see Fig. 1). This cell contains two cation and one anion permselectives membranes for defining the cell compartments. The compartments are separate by means of polypropylene tortous spacers which contains turbulent promoters to enhance the mass transfer from the bulk solution to the membrane surface, where the separation is carried out, thus permitting the cell to operate at a high current density. When an electrical potential is applied to the electrolytic cell by mean of a power supply, potassium ions and other cations move to the cathode while nitrate ions and other anions move to the anode. At the cathode compartment potassium ions are combined with hydroxide ions product of water electrolysis to produce potassium hydroxide. The electrolytic reduction of water also produces hydrogen gas which may be used as a potential fuel. The nitrate ions removed from the leachate compartment pass through the anion-exchange membrane to the nitric acid compartment and combine with the hydrogen ions that come from the anode compartment to enrich the nitric acid solution. At the anode compartment water oxidizes under acid conditions and produces oxygen gas and hydrogen ions. Oxygen gas may be use to satisfy any human being needs, as well as, to be combined with hydrogen gas for energy purposes.

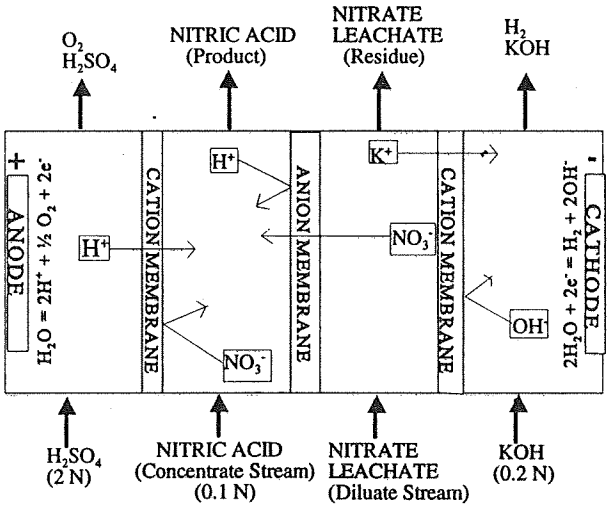


Fig. 1. Configuration of Electrolytic Cell

1.3 OBJECTIVES

The main objective of this summer project was to design, and test an electrolytic system to remove nitrate from crop residues, as well as, to satisfy biomass production chamber (BPC) and anaerobic and aerobic nitric acid demand. The potato crop residues was leachated and the resultant leachate electrodialyzed to remove nitrate. In the design of this process, variables such as current density, applied voltage, fluid hydrodynamic conditions and leachate ionic concentration were correlated to optimize nitrate removal.

II. EXPERIMENTAL SET-UP

2.1 ELECTROLYTIC CELL

The experimental system consists of a 0.23 x 0.25 m four compartments electrolytic cell (Fig. 2). The electrodes are a stainless steel cathode and a titanium substrate plated anode. The membranes selected were two heavy duty cation permselective membranes (Type 61-AZL-389, from IONICS, Inc., Watertown, MA) and one anion permselective membrane (Type 103-QZL-386, from IONICS, Inc., Watertown, MA) . The separation between electrodes and cation membranes and between membranes were done with polypropylene tortous path spacers that contain the channeling and turbulent promoters. The thickness of electrodes spacers is 2×10^{-3} m and the intermembrane spacer is 1×10^{-3} m. The open flow area is 6×10^{-6} m^2 for intermembrane compartments and 1.2×10^{-5} for electrodes compartments. The effective membrane area is 0.023 m^2.

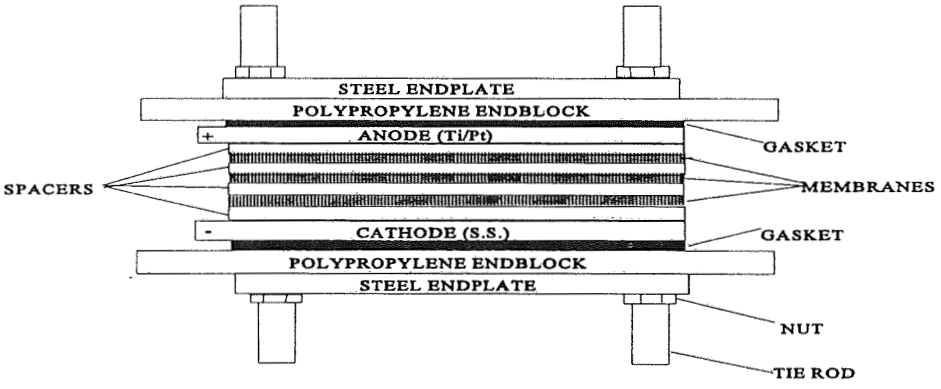


Fig. 2. Four Compartments Electrolytic Cell Construction

2.2 COMPARTMENTS SOLUTIONS AND HYDRODYNAMIC CONDITIONS

The cathode compartment feed solution was a 0.2 N KOH. The solution feed to the compartment next to the cathode was a filtered potato leachate. The specific protocol for leaching potato inedible ground biomass is elsewhere (1). An overview of the relevant conditions were as follows: 1) loading rate of 50 g of biomass per liter of deionized water, 2) duration time of 2 hr, 3) aeration rate of 20 LPM, and 4) rising with an equal volume of deionized water in five separate washings. Then the leachate was filtered with a 200 μm stainless-steel wire mesh to remove the suspended matter that may clog the electrolytic cell flow channel. The composition of the leachate obtained by the previously mentioned method is shown

in Table 1. The nitric acid feed solution was a 0.1 N HNO_3 and the anode feed solution was a 2 N H_2SO_4. The nitric acid and potassium hydroxide solutions were prepared in such a way, to keep the ratio of the specific electrical conductivities of these solutions to the leachate solution below 10 to prevent back diffusion and thus to avoid short circuit through the manifold ($k_{HNO3}/k_{Leachate}$ =3.4, $k_{KOH}/k_{Leachate}$ = 4.2). The fluids velocities flowing through the tortous spacers were varied from 0.083 to 0.403 m/s given Reynold's numbers (based on the equivalent hydraulic diameter of 1.714 x 10^{-3} m) of 142 to 691. The flowrates through the electrodes compartments were doubled as compared to the leachate and nitric acid solutions which kept the same fluid velocity.

Table 1. Potato Leachate Average Composition

Analysis	Composition
Spec. Cond., µMho/cm	8650
pH	5.81
NO_3, mg/L	**1474**
PO_4, mg/L	154
K, mg/L	**2877**
Ca, mg/L	7.5
Mn, mg/L	157.1
Fe, mg/L	4.26
Cu, mg/L	0.16
Mg, mg/L	0.95
Zn, mg/L	0.75
NH_4	56.9
Cl, mg/L	70.4
TOC, mg/L	3039.8

2.3 LIMITING CURRENT DETERMINATION

The determination of the limiting current density (maximum operating current density) was carried out by operating the system with the respective solutions under continuous operation mode. A pH electrode, a specific electrical conductivity flow cell, and thermocouples were

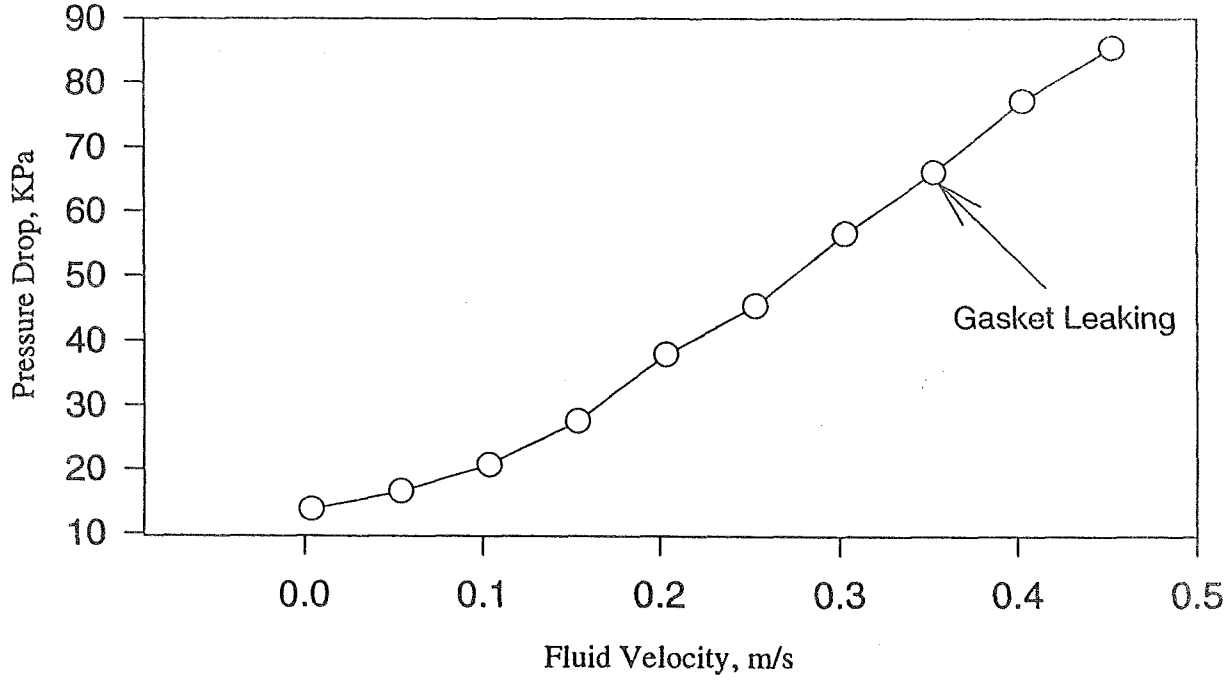
installed at the outlet of the potato leachate stream to monitor pH, specific electrical conductivity, and temperature difference, respectively, as a function the applied current density and hydrodynamic conditions. Also the applied voltage vs current density was monitored to determine the cell electrical resistance.

III. RESULTS AND DISCUSSION

3.1 ELECTROLYTIC CELL PRESSURE DROP

The electrolytic cell was tested for external leakage by pumping de-ionized water at room temperature through each compartment at the same fluid velocity. As can be observed in Fig. 3., the pressure drop through the electrolytic cell increases rapidly with fluid linear velocity (flowrate). At fluid velocities higher than 0.35 m/s external leakage was observed between spacer-membrane junction. Thus, the maximum permissible flowrate is limited by high pressure drop and external leakage. At a fluid velocity of 0.35 m/s, in which the external leakage began the pressure drop, was approximately 66.2 kPa (9.6 psi). Therefore, it is also more economical to operate below the maximum allowable pressure of 66.2 kPa.

Fig. 3. Electrolytic Cell Pressure Drop with Fluid Velocity



3.2 POLARIZATION (LIMITING CURRENT DENSITY)

"Current density" in general is defined as the current per unit area of available membrane through which the current passes. The effective membrane area of the electrolytic cell is of 0.023 m^2 (230 cm^2). In solution electrical current is carried by ions, it is follow that more dilute the solution the lower will be the current transmittal capacity of the solution. Application of too

much voltage will result in ion starvation at the membrane-solution interface with the resultant decomposition of solvent and/or other materials that may be present. This phenomenon is referred to as polarization and manifest itself through pH disturbances in solution, increase in electrical resistance, and loss of current efficiency. Consequently, the limiting current density is the first operating variable that should be established experimentally.

Fig. 4. shows the applied voltage vs current relationship for the operation of the electrolytic cell under continuous mode of operation at different fluid velocities ranging from 0.083 to 0.403 m/s. From the Ohm's law (E = IR), the slope of the voltage-current relationship at a given current is the cell electrical resistance. As can be observed in this figure the slope (cell electrical resistance) is small at low current, but as current is increased there is current where it suddenly changes from a finite value to almost infinity. The current at which this drastical change occur is known as the maximum permissible current or limiting current. The limiting current increases with increasing fluid velocity. As the fluid velocity is increased, the boundary layer thickness decreases, thus higher permissible current can be used. The effect of increased flow velocity included the enhancement of mass transfer, reduction of membrane fouling, and therefore minimization of polarization. The average cell electrical resistance is approximately 2 ohms when the operation is carried out below limiting current conditions. When the cell electrical resistance increases drastically some of the energy required to transport ions, dissipates as heat due to the Joule heating effect (see Fig 5). Thus, this effect increases the fluid temperature as it passes through the cell. It is suggested to operate at low temperature as possible to avoid membrane overheating and thus, to maintain the chemical stability of the permselective membranes for long period of operation. Most of the commercial polymeric-basis membranes can be used for an extended period at a temperature up to 40°C, warmer operation entails a great reduction of useful life. As can be observed in Fig. 6, the operation at current density higher than the limiting one produce pH disturbances in the leachate stream, as well as, loss of current density (leachate specific electrical conductivity is independent of current density). Thus, the operation at current density higher than the limiting one will may result in decomposition of solvent and other materials presented in the leachate. As the fluid velocity increases the pH disturbances and loss of current efficiency are minimized by increasing the mass transfer of ions from the solution bulk to the membrane interface. Fig. 7, shows the limiting current density as a function of leachate fluid velocity. These data were correlated very well by the power function model: $i_{lim} = 576.41\ U^{0.667}$. The resulting correlation coefficient obtained is 0.995. This correlation agrees with the one suggested by several authors (19,20). Once the limiting current density is known at a given fluid velocity, selection of operating current density below the limiting one will warranty an optimum electrolytic cell performance. Due to the fluid velocities range used (laminar flow conditions), the tortous path flow channels through which the solution is passing appear to function as causing a degree of localized eddying rather than inducing turbulence flow.

Fig. 4. Electrolytic Cell Applied Voltage vs Current at Different Fluid Velocities

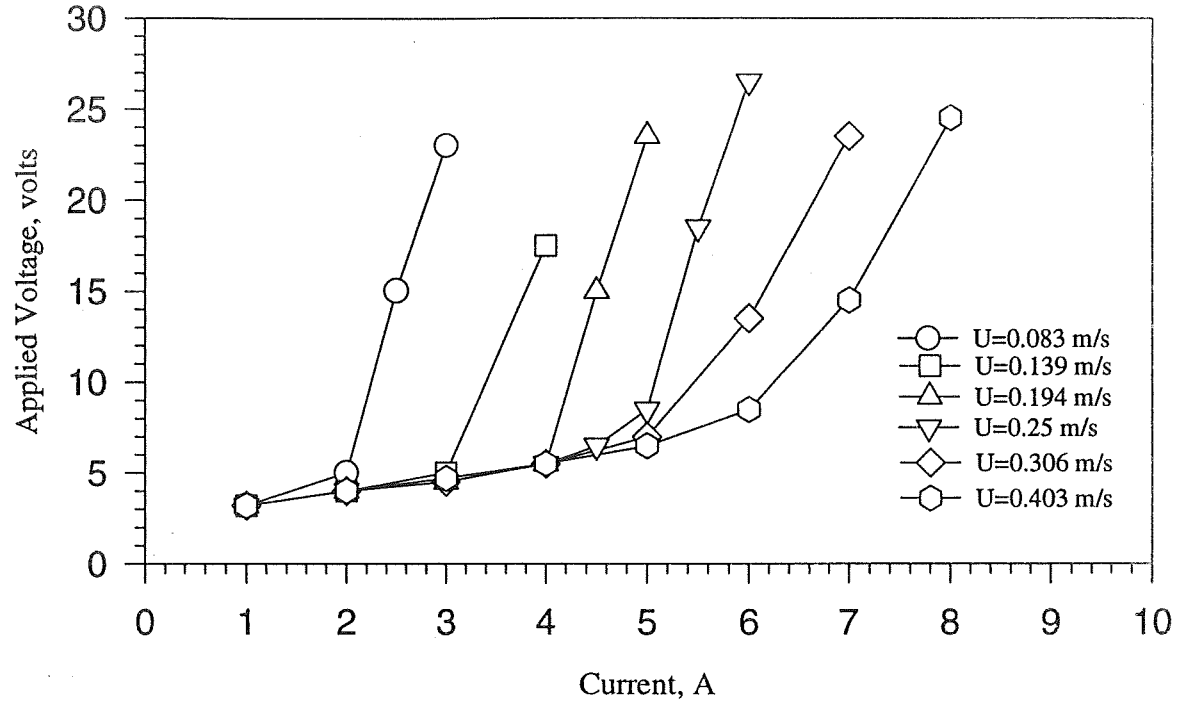


Fig. 5. Leachate Temperature Change Through the Electrolytic Cell as Function of Fluid Velocity and Applied Current

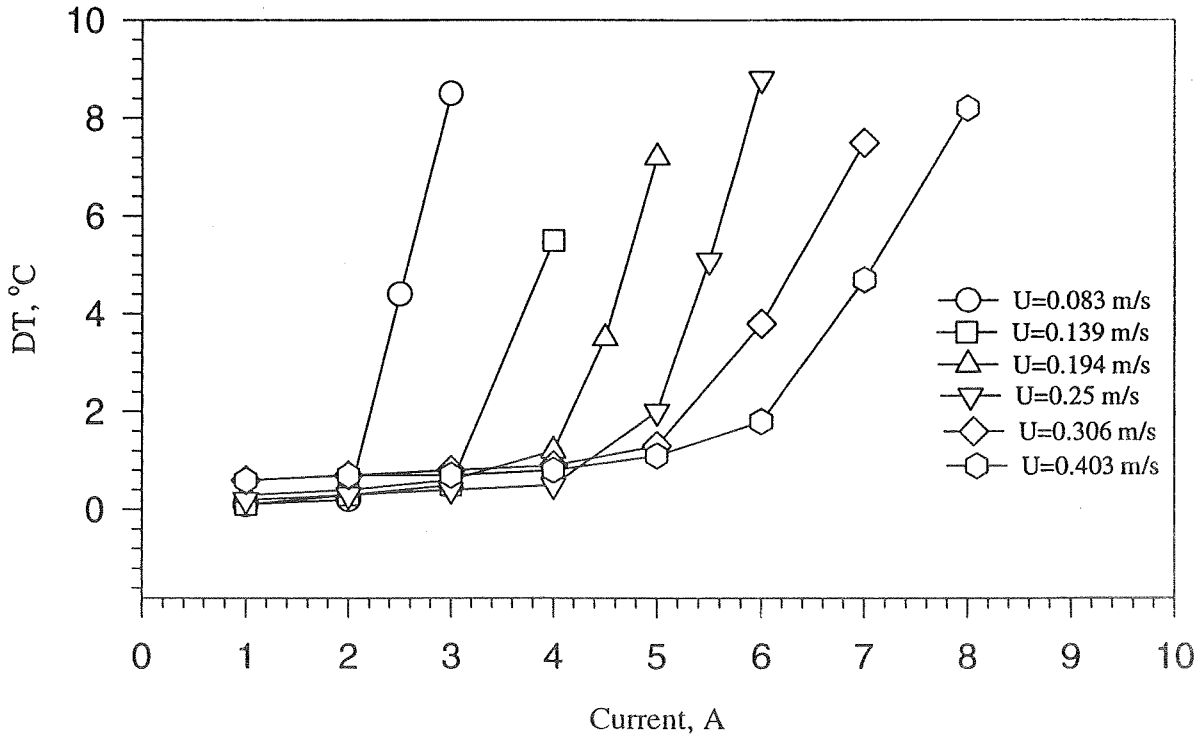
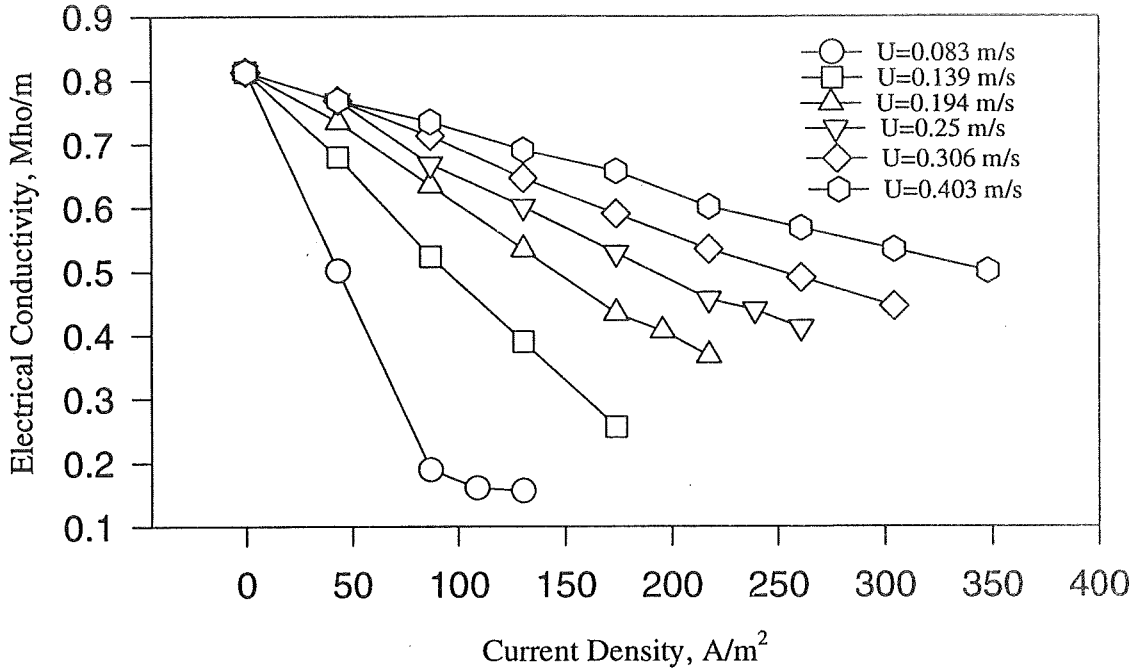


Fig. 6. Leachate Effluent Specific Electrical Conductivity and pH vs Current Density during Once-Through Continuous Operation



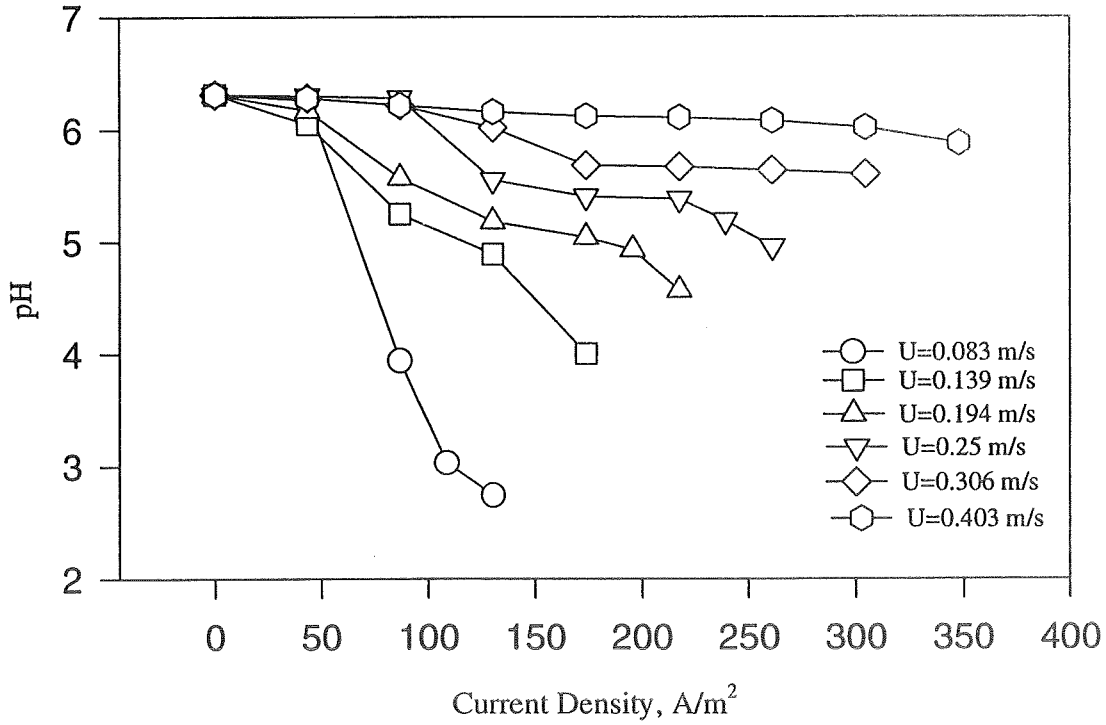
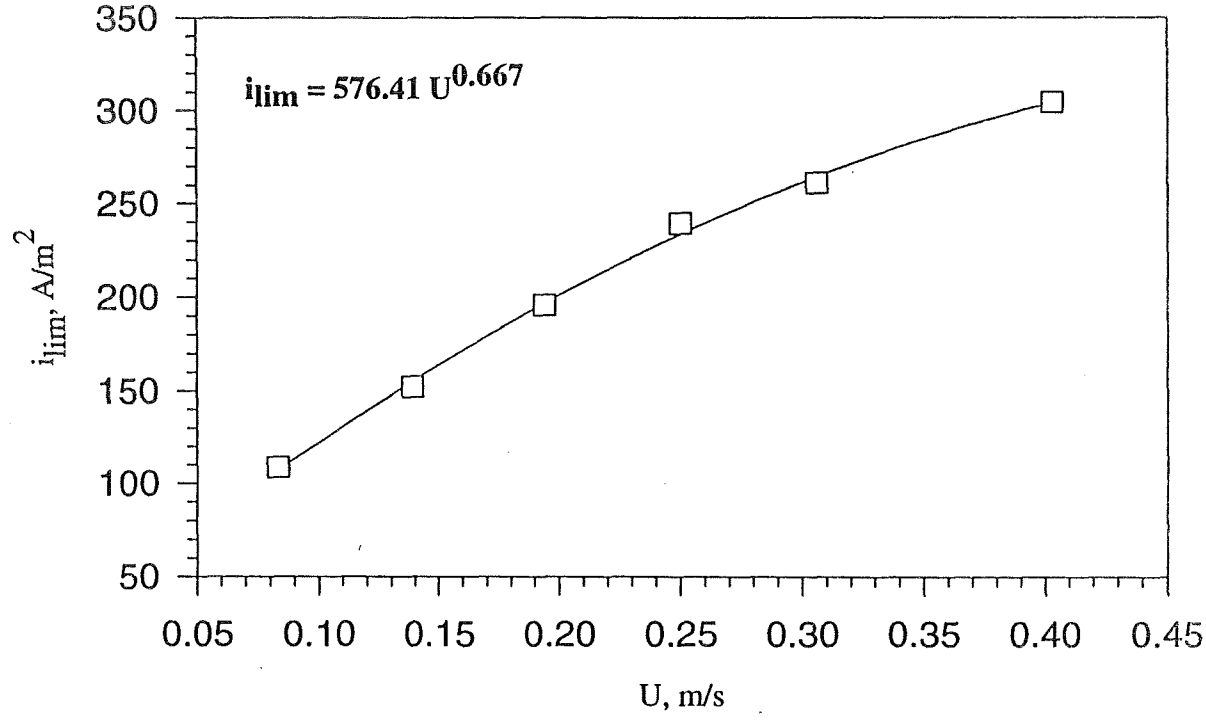


Fig. 7. Effect of Fluid Velocity on Electrolytic Cell Limiting Current Density



3.3 ONCE-THROUGH CONTINUOUS OPERATION

This operation was carried by pumping the solutions once through the electrolytic cell from feed tanks at a constant flowrate. As the diluting stream (leachate solution) passes through, the electrolytic cell becomes more demineralized. The concentrating stream (nitric acid solution) becomes more concentrated. In a continuous (once-through) operation the solution to be processed was passed through the electrolytic cell only once. This is convenient when one "pass" suffices to achieve the desired degree of demineralization.

Fig. 8 shows the leachate demineralization as a function of current density at a fluid velocity of 0.083 m/s. As can be seen, the removal per pass of each nutrient contained in the potato leachate increased with current density. The removal of nitrate, potassium, and chloride ions are much higher than the other ions present. At low current density of 44 A/m^2 is possible to remove up to 62.6% nitrate, 40.3% potassium, 26.5% chloride, 33.7% magnesium, 30.6% manganese, 38.2% zinc, 9.41% iron, and 19.8% phosphate, respectively. Although the removal of nutrients from potato leachate increase with current density, operation at current density higher

than 108.7 A/m^2 (limiting current density) is not recommended due to excessive solution and membrane overheating. Operation at the limiting current density condition is possible to remove up to 92% nitrate, 92.5% potassium, 22.1% phosphate, 39.8% calcium, 71.4%magnesium, 39.9% iron, 42.3% zinc, 56.7% manganese and 92.4% chloride. The removal of monovalent ions was found to be higher than divalent ions because monovalent ions have higher mobility (current transmittal capacity) than divalent ones. At current density higher than the limiting one the removal of ions is almost independent of current density. At these conditions a big portion of electrical energy is being spent to heat the solutions (joule heating effect). As the fluid velocity is increased, the fluid residence time in the electrolytic cell decreases; therefore, lower demineralization per pass is obtained (see Figure 9 and 10). Increasing the fluid velocity will permit the operation at higher current density with minimum polarization, fouling, and solution and membrane overheating. At a fluid a velocity of 0.25 m/s and operating at the lower current density of 43.8 A/m^2 is possible to remove 12.4% nitrate, 6.2% potassium, 9.8% phosphate, 23.5% calcium, 9% magnesium, 6.4% iron, 10.4% zinc, 6.9% manganese, and 15.7% chloride, respectively. While the operation at the limiting current density (239.13 A/m^2) makes possible to remove up to 74.7% nitrate and 46% potassium; the operation at fluid velocity of 0.306 m/s and limiting current density (260.9 A/m^2) makes possible to remove up to 43.6% nitrate and 40.1% potassium (see Fig 10). It is recommended to operate the electrolytic cell process at fluid velocities between 0.20 to 0.25 m/s and current densities from 200 to 225 A/m^2. Operation at these conditions minimize solution and membrane overheating, as well as, polarization and membrane fouling effects.

3.4. BATCH-RECIRCULATION OPERATION

In batch-recirculation operation of the electrolytic cell system, the solutions to be processed were placed in their respective reservoir tanks. They were passed through the stack and then recycled back to the reservoirs. Thus, the leachate solution became increasingly depleted while the nitric acid solution became more concentrated. The electrolytic process was considered complete when the nitrate content drop to desired degree.

Batch operation has the advantage that the process cycle can be geared toward the desired end product and is relatively independent of the feed, the concentration of which may vary from batch to batch. Even if the feed varies, batch recirculation is equivalent to operate a number of electrolytic cell stack in series, with each stack receiving an increasingly depleted nitrate solution and concentrated nitric acid solution feed.

There are two ways to operated the electrolytic cell under batch-recirculation mode of operation, at constant-current and at constant-applied-voltage. The operation at constant-current would seem attractive for assuring a constant rate of nitrate removal from potato leachate, but this will be true only if the feed composition remain constant. One alternative is to operate so that the cell current is constant and below the value of the initial limiting current. For part of the process duration, the operating current lies below the limiting one, and this represents an under-utilization of the system capacity. When the operating current exceeds the limiting one,

Fig. 8. Removal of Selected Nutrients from Potato Leachate vs Current Density at U = 0.083 m/s

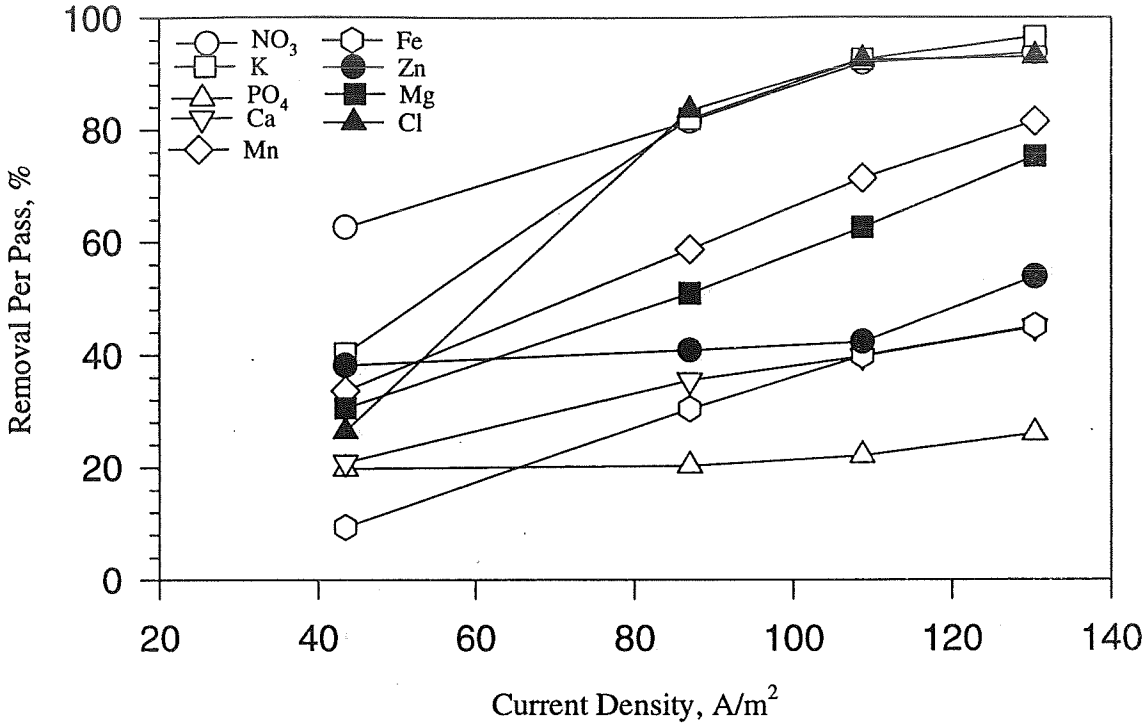
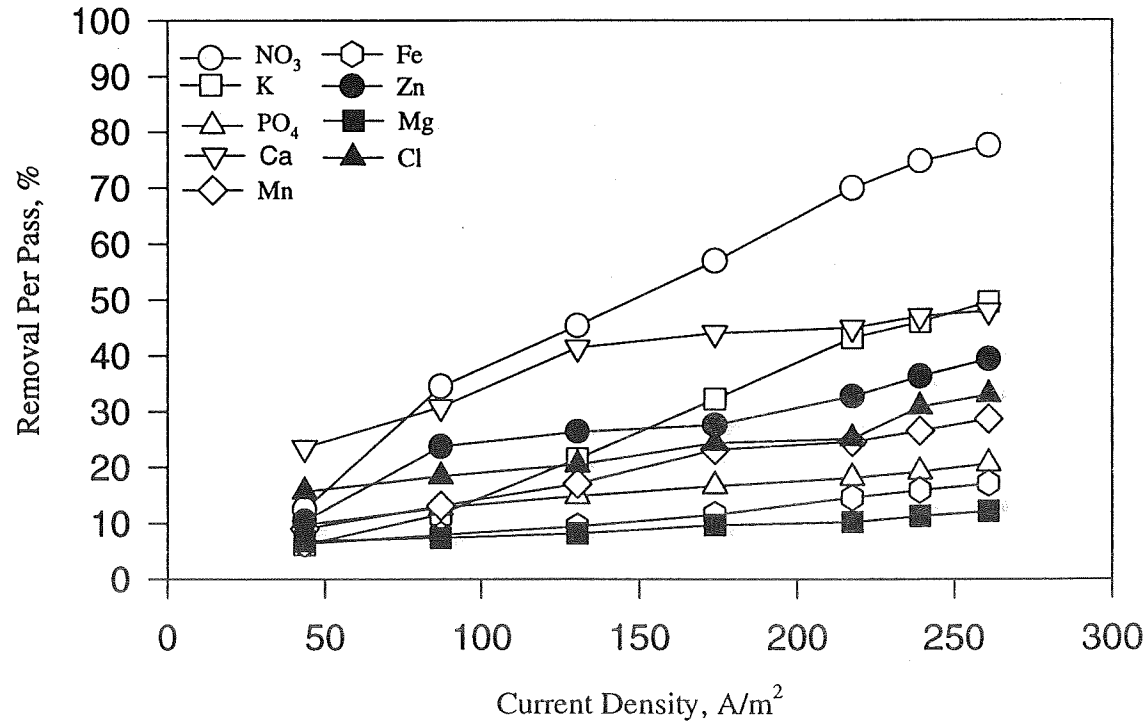


Fig. 9. Removal of Selected Nutrients from Potato Leachate vs Current Density at U = 0.25 m/s



secondary reactions begin, resulting in power wastage. Therefore, constant current operation is impractical for use with batch-recirculation systems, which would be operated over a wide ionic concentration range. This realization leads to the concept of using high current levels during the initial part of the process and reduced levels when the ionic concentration of the leachate falls. One way of accomplishing this, is to switch to a constant applied voltage operation. The operation was carried out in two trials at constant applied voltages of 4.5 and 8.5 volts, respectively. The reservoir tanks were charged with 2 liters of each solution. The flowrate was adjusted to give a fluid velocity of 0.25 m/s for all the trials.

Fig. 10. Removal of Selected Nutrients from Potato Leachate vs Current Density at U = 0.306 m/s

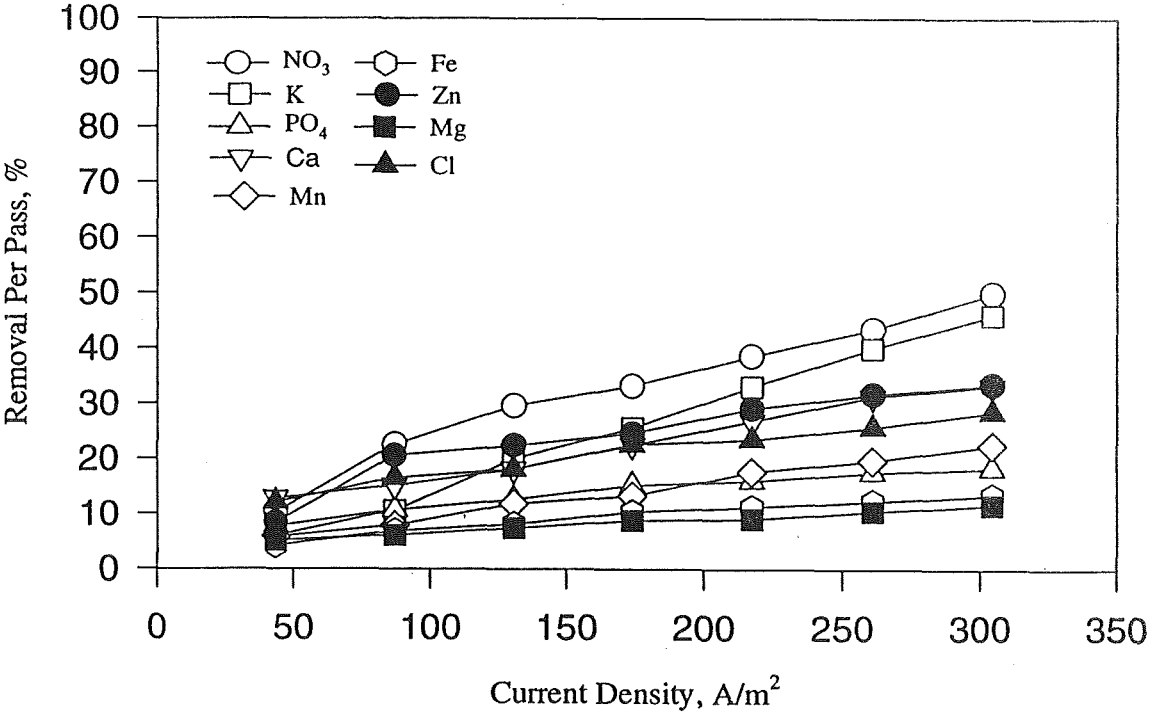


Fig. 11 shows the current density versus time relationship for the batch-recirculation operation at constant applied voltage conditions. At a constant applied voltage of 8.5 volts, the current density declines rapidly during first the 20 minutes, and then levels off with a diminishing rate for the remaining period of time. This occurred as consequence of a rapid demineralization of the leachate at high current density. As the ionic concentration of the leachate diminishes, the solution electrical resistance increases rapidly, thus to keep the same applied voltage, the current density is itself adjusted to lower values. The operation at constant applied voltage of 4.5 volts, required lower initial current density, thus the initial rate of demineralization was low enough to

keep small variation of solution electrical resistance. After the period of 20 minutes of operation the patterns of current density variation vs time at both applied voltages were very similar. This means that the rate of demineralization is slow at low current density. As can be observed in Fig. 12, the specific electrical conductivity versus time at both applied voltages have a similar pattern. The leachate solution electrical conductivity decreases with operating time as the leachate solution is depleted. Also, as being seen in Fig. 13, the leachate solution pH decreases slowly from 5.8 to 4.2 at E = 8.5 volts and from 5.7 to 4.53 at E = 4.5 volts. At a low applied voltage (E = 4.5 volts), pH is more stable than at high applied voltage (E = 8.5 volts), which means a low degree of water de-ionization or low polarization effect.

Figures 14 and 15 show the demineralization of potato leachate as function of time. The initial rate of demineralization is high and then levels off. The initial rate of demineralization is higher for two reasons; initially the leachate solution contains a high concentration of ionic species and the initial current density is also high. As the leachate become more depleted, both current density and the rate of nutrient removal decrease. The average rate of nutrient removal was almost independent of applied voltages, but depends on operating time. During the operation of the electrolytic cell at constant applied voltage of 4.5 volts for 158 minutes, made possible to remove up to 92.1% nitrate, 26.5% phosphate, 79.2% potassium, 39.2% calcium, 68.2% magnesium, 38.5% iron, 38.8% copper, 48.2% manganese, 44% zinc, 88.7% chloride, and 22.7% TOC, respectively. Approximately the same overall removal was obtained when the system was operated at constant applied voltage of 8.5 volts, but for a period of 148 min. As the cations are transferred to the cathode compartment, the alkali concentration increases with time (see Fig. 16). The concentration of alkali solution was found to change from 0.209 to 0.266 N for the operation at 4.5 volts (27.3% conversion) and from 0.201 to 0.257 N at 8.5 volts (27.9% conversion). The alkali formation is the product of the electrochemical reaction at the cathode surface that produces hydroxide ions which are combined with the cations to form alkali compounds. Also, as the anions are transferred to the nitric acid compartment, the acid concentration was found to increase with time. The concentration of the acid solution in the nitric acid compartment was found to increase from 0.098 to 0.151 N for the operation at 4.5 volts (54.1% conversion) and from 0.098 to 0.152 at 8.5 volts (55.1% conversion). From material balances was found that the average oxygen gas (O_2) produced at the anode surface under acid condition was approximately 2.2 moles O_2/mole NO_3 transferred. The average production of hydrogen gas at the cathode surface was found to be approximately 4.4 moles H_2/mole NO_3 transferred.

Fig. 11. Current density vs Time during Batch-Recirculation Operation at Constant Applied Voltages and U = 0.253 m/s

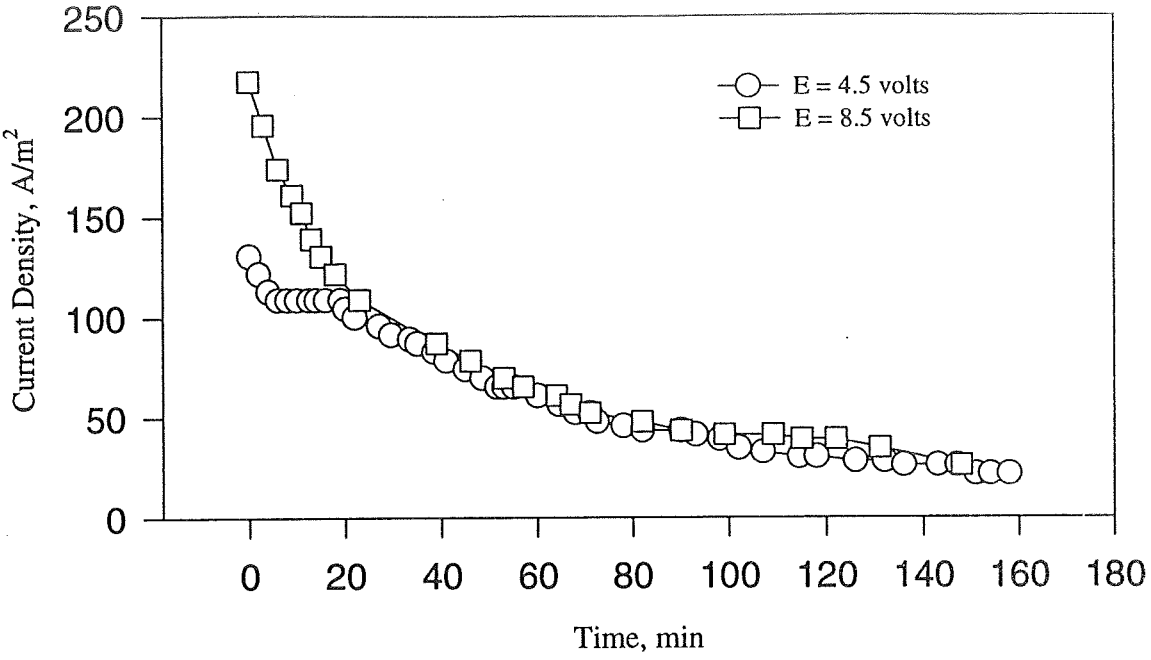


Fig. 12. Leachate Specific Electrical Conductivity vs Time during Batch-Recirculation Operation at Constant Applied Voltages and U = 0.253 m/s

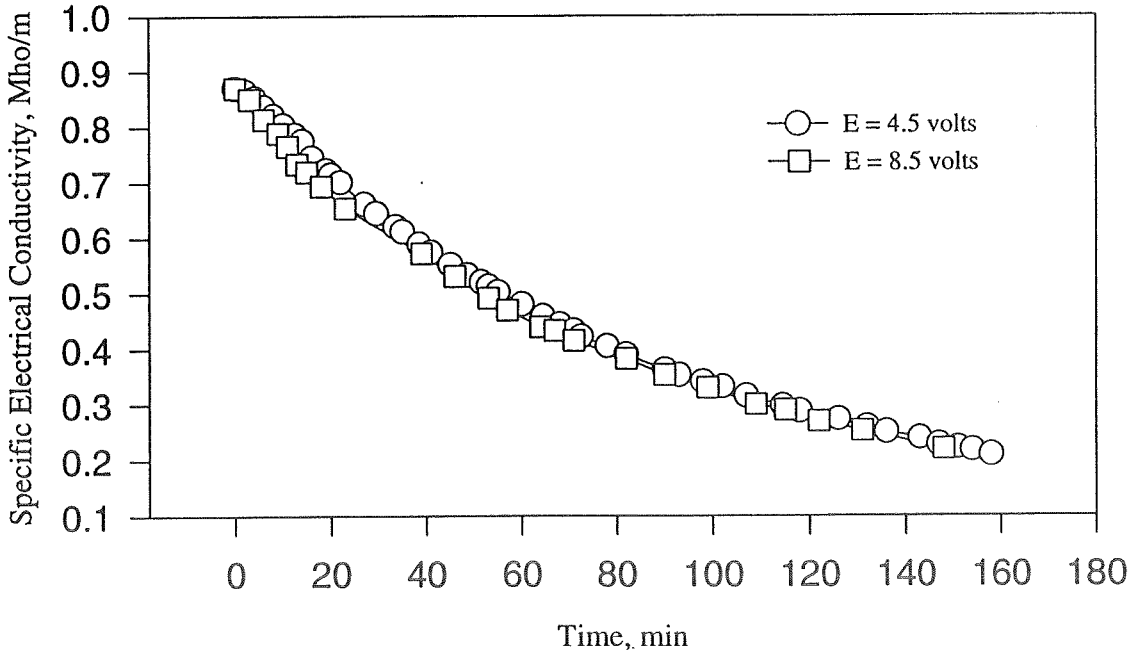


Fig. 13. Leachate Solution pH during Batch-Recirculation Operation at Constant Applied Voltages and U = 0.253 m/s

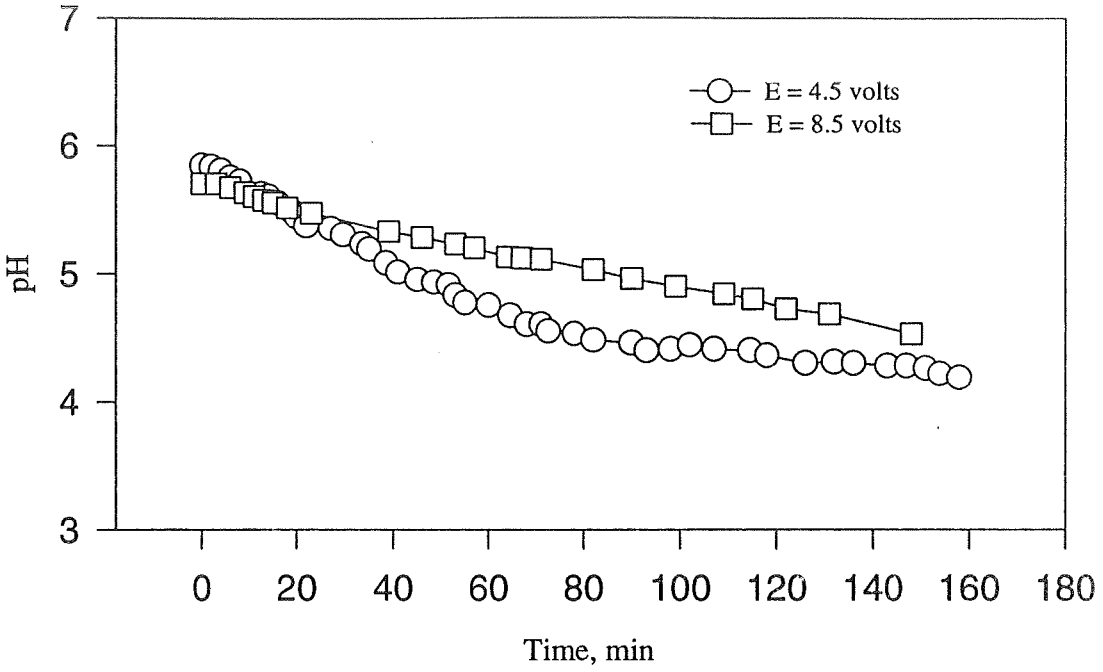


Fig. 14. Nutrient Removal during Batch-Recirculation Operation at E = 4.5 volts and U = 0.253 m/s

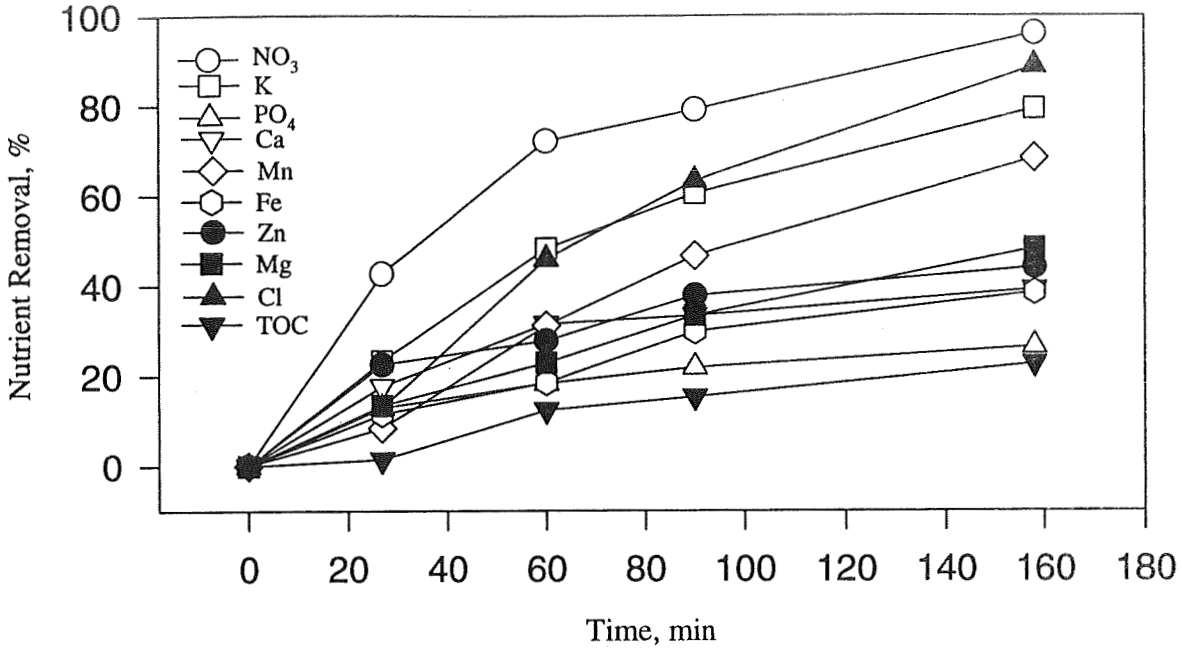


Fig. 15. Nutrient Removal during Batch-Recirculation Operation at E = 8.5 volts and U = 0.253 m/s

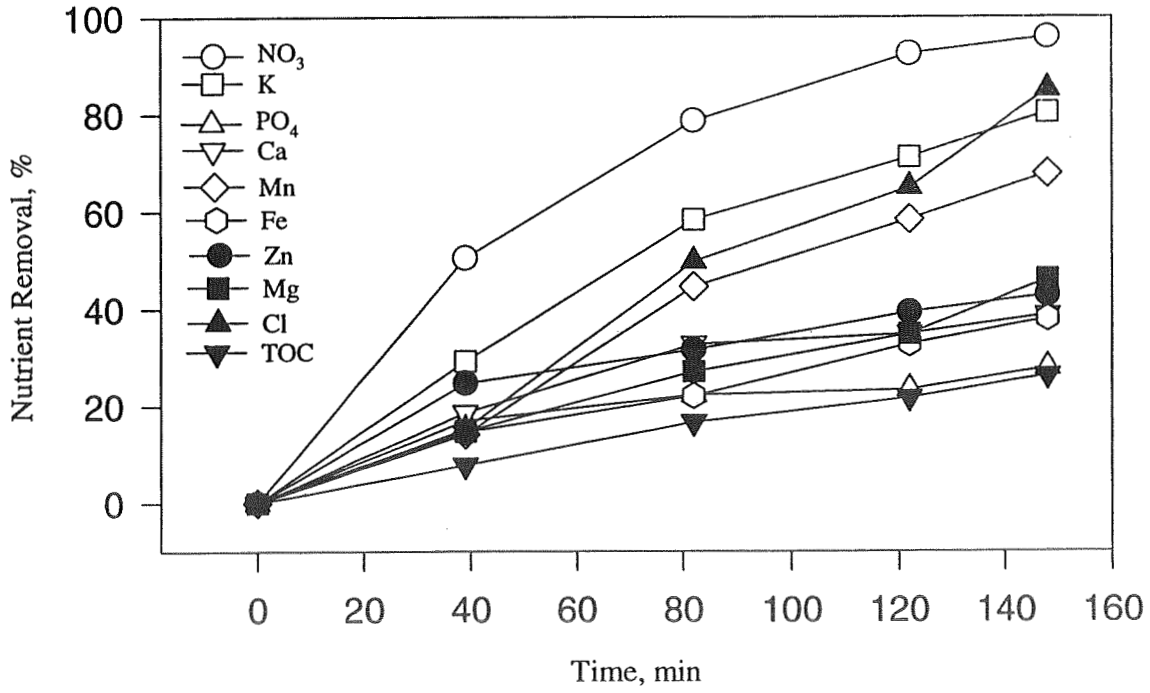
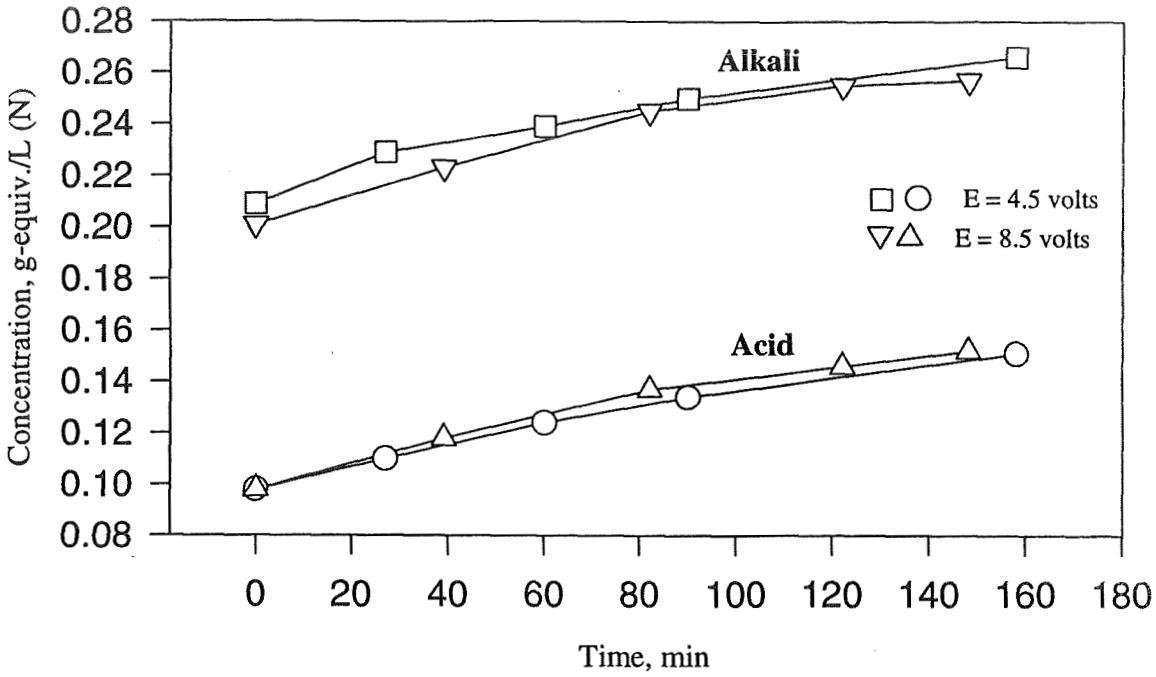


Fig. 16. Nitric Acid and Potassium Hydroxide Concentration vs Time during Batch-Recirculation Operation at Constant Applied Voltages and U = 0.253 m/s



IV. CONCLUSIONS

The limiting current density (maximum operating current density) was found to be directly proportional the solution electrical conductivity an a power function of the linear fluid velocity in the range between 0.083 to o.403 m/s. The operation at current density higher than the limiting one, results in pH disturbances which may decomposes pH sensitive leachate components, excessive solution and membrane overheating, which may reduce membrane useful life , and loss of process current efficiency.

During the electrolytic cell once-through operation, the nitrate, potassium, and other nutrients removal rates were proportional to current density and inversely proportional to fluid velocity. Although the removal of nutrients from potato leachate increase with current density, operation at current density higher than the limiting one is not recommended due to excessive solution and membrane overheating. The removal of monovalent ions was found to be higher than divalent ones, because monovalent ions have higher mobility in solution, as well as, are more membrane selective than divalent ones. The operation at high fluid velocities reduced the ion removal capacity per pass, but increases bulk solution-membrane interface mass transfer, which reduces membrane fouling to extent membrane useful life

Under batch-recirculation operation at constant applied voltage of 4.5 and 8.5 volts, it was found that the nutrient removal rates were independent of applied voltage, but proportional to the ions concentration and operating time. The initial nutrients removal rates were found to be high because of the high initial solution ionic concentration and current density. This operation appeared to work as a current self-regulating process, because as the leachate becomes depleted the solution electrical resistance increases, therefore to maintain the constant applied voltage condition the current is self adjusted to lower values. From material balances it was found that 2.2 moles of oxygen gas and 4.4 moles of hydrogen gas were produced at the electrodes surfaces per mole of nitrate transferred.

V. REFERENCES

1. Garland, J.L. and C.M. Mackowiack. Utilization of water soluble fraction of wheat straw as a plant nutrient source. NASA Tech. Mem. 103497, (1990).

2. Garland, J.L., C.M. Mackowiack and J.C. Sager. Hydroponic crop production using recycled nutrients from inedible crop residues. SAE Tech. Pap. 932173, (1993).

3. Modell, M. Supercritical waste oxidation of aqueous wastes. NASA Tech. Mem. 88215, (1986).

4. Garland, J.L. Coupling plant growth and waste recycling systems in a controlled ecological life support system (CELSS). NASA Tech. Mem. 107544, (1992).

5. Garland, J.L. Characterization of the water soluble components of inedible residues from candidate CELSS crop. NASA Tech. Mem. 107577, (1992).

6. Finger, B.W. and R.F. Strayer. Development of an Intermediate-Scale Anaerobic Bioreactor to regenerate nutrients from inedible crop residues. SAE Tech. Pap. 941501, (1994).

7. Chynoweth, D.P. Overview. In D.P. Chynoweth and R. Isaacson (Ed.) Anaerobic Digestion of biomass. Elsevier Applied Science, London, (1987), p. 171.

8 Rejken, B.A. Anaerobic digestion of solid organic matter. In O.Kitanic and C.W. Hall (Ed.) Biomass Handbook, Gordon and Beach Science Publishers, New York, NY, (1989), p. 319.

9. Colon, G. And J.R. Rosenau. Lactic acid recovery from fermented acid-cheese whey via electrodialysis. AIChE Tech. Pap. , 1987 AIChE Annual Meeting, New York, NY, (1987).

10. Bodamer, G.W. Electrolytic process for producing amines. U.S. Pat. No. 2,737,486, (1956).

11. Bodamer, G.W. Electrolytic conversions with permselectives membranes. U.S. Pat. No. 2,921,005, (1960).

12. Giuffrida, A.J. Electrodialysis apparatus and process ion modification. U.S. Pat. No. 3,964,985 (1976).

13. Hittman Associated, Inc. Electrodialysis desalting state of the art. Office of Saline Water Res. Develop. Progr. Rep. No. 610, (1970).

14. Sonin, A.A. and R.F. Probstein. A hydrodynamic theory of desalination by electrodialysis. Desalination, 5, 1908, (1968)

15. Sonin, A.A. and R.F. Probstein. Turbulent flow theory of electrodialysis. U.S. Department of the Interior, Research and Development Progress Report No. 771, (1972)

16. Resenberg, N.W. and C.E. Tirrell. Limiting current in membrane cells. Ind. Chem. 49, 780 (1957).

17. Cowan, D.A. and J.H. Brown. Effect of turbulence on limiting current in electrodialysis cells. Ind. Eng. Chem., 51, 1445, (1959).

18. Korngold, E. Electrodialysis - Membranes and Mass Transport. In Synthetic Membrane Processes Fund. And Water Applications. Bedford George (Ed.). Academic Press, (1984), p. 191.

19. Davis, A. And E. Lacey. Forced-flow electrodesalination. O.S.W. Research and Development Report No. 557, Webster, South Dakota, (1970).

20. Mason, R. Wester test facilities and electrodialysis test bed plant. O.S.W. Research and Development, Webster, South Dakota, Report No. 568, (1970).

21. Indusekhar, V.K., G.S. Trivedi, and B.G. Shah. Removal of nitrate by electrodialysis. Desalination, 84, 213-221, (1991).

1996 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

STATISTICAL PROCESS CONTROL FOR KSC PROCESSING

Dr. Roger G. Ford, Associate Professor
Engineering Department
St. Mary's University
San Antonio, Texas

KSC Colleagues - Hector Delgado and Randy Tilley
Systems Safety

Contract Number NASA-NGT10-52605

August 9, 1996

ABSTRACT

The 1996 Summer Faculty Fellowship Program at Kennedy Space Center (KSC) served as the basis for the research effort into Statistical Process Control for KSC Processing. The effort entailed several tasks and goals. The first was to develop a customized Statistical Process Control (SPC) course to be taught to the Safety and Mission Assurance Trends Analysis group. The actual teaching of this course took place over several weeks. In addition, an Internet version of the same course complete with animation and video excerpts from the course when it was taught at KSC was developed. The Application of SPC to Shuttle Processing took up the rest of the summer research project. This effort entailed the evaluation of SPC use at KSC, both present and potential, due to the change in roles for NASA and the Single Flight Operations Contract (SFOC). Individual consulting on SPC use was accomplished as well as evaluation of SPC software for KSC use in the future. A final accomplishment of orientation of the author to NASA changes, terminology, data format, and new NSA task definitions will allow future consultation when the needs arise.

STATISTICAL PROCESS CONTROL FOR KSC PROCESSING

ROGER G. FORD, PH.D., P.E.

1. INTRODUCTION

The use of Statistical Process Control (SPC) at Kennedy Space Center (KSC) was the focus of this research effort for the Summer of 1996. The research effort is broken down into several tasks that were accomplished over the course of the summer. Those tasks were:

- Development of a customized KSC Statistical Process Control course to be taught to Safety and Mission Assurance Trend Analysis Group and any others interested
- Teaching the developed SPC course
- Developing an Internet version of the SPC course for the benefit of Safety and Mission Assurance (S&MA) management
- The evaluation of the application of SPC to Shuttle processing both present and future
- Evaluation of SPC software for use by S&MA personnel
- Establish a working relationship with the people involved with Shuttle processing analysis, the terminology and procedures involved in Shuttle processing, and become familiar with the type of data gathered in the course of Shuttle processing in order to be able to act as consultant to the analysis effort in the future

NASA has been required, due to budget considerations, to change its role in Shuttle Operations. From an historic partnership and combined effort with contractors in the pursuit of space flight, now NASA is faced with assuming a caretaker or managerial "hands off" role for the future of the Shuttle program. This is a significant cultural change requiring the need to utilize all of the techniques from industry available to increase efficiencies, reduce rework, gain control and identify causes for excessive costs, reduction in inspections and surveillance routines, reductions in manpower, and reduce the throughput time necessary for each cycle of the orbiters from blastoff to blastoff. Statistical Process Control is a significant tool to assist in the success of NASA's new managerial role because it can assist in determining the reasons for process variation leading to decreased costs.

In order to use SPC within the KSC environment, pertinent personnel need training in the proper use of the technique. The training must extend from the top down, meaning upper management must be knowledgeable in SPC techniques as well as the analyst working with the data. Training has been and will be accomplished in two ways: first, a customized SPC course was developed and taught to the analysts who will be performing the majority of the analyses; and second, an Internet version of the same course will be finished soon for the use of anyone, especially top management, who desires or is required to utilize SPC techniques. The web-based course is full of animation, video from the course when it was taught at KSC, interactive tools for statistical calculation, and the charts, search capabilities, and "Frequently Asked Questions" (FAQ) needed to make the product useful at any time.

The need to assess current use of SPC techniques as to appropriateness, correctness, completeness, and existence where needed was foremost in the task of application of SPC to Shuttle processing. Then came the consultation with various individuals as to the approach recommended to gain control of the various processes. The three approaches to SPC analysis discussed, in order of recommended application, are metrics (descriptive statistics), attribute charting technique for nonconformance data, and variable charting techniques for measurement data. This effort, in particular, is an on-going one covering, perhaps, the next three to five years. Management at the highest levels within NASA must understand what to expect in order to get processes under statistical control as well as understand what decisions to make and when to make them based on the SPC techniques.

SPC software for KSC use was evaluated. The SAS Institute's JMP statistical software package was in residence and deemed to be the most powerful and speedy even though explanations were scant.

2. Statistical Process Control Course Development, Teaching, and Internet Version

SPC can be a very easy to use technique and a very difficult one as well. There is a definite sequence that is recommended to be followed when implementing SPC into any situation, especially one in which the knowledge level of SPC is relatively low or there are misconceptions. Unfortunately, these conditions exist virtually everywhere and certainly existed at NASA at the beginning of the Summer of 1996.

Because of this need to know the proper sequence in which to implement SPC, and the possible difficulty that some have with understanding of the statistical theory involved with SPC, a course customized to KSC was developed this summer. The course was entitled "Quality Control NASA Style", and it was designed for six sessions of two hours each for a total of 12 hours. The last session was to be a computer lab class where SPC software could be demonstrated and practiced upon.

The SPC course was divided into six days of PowerPoint presentations covering the following topics:

I. Day One
- A. Concepts of Quality - Definitions of Quality
- B. Inspection versus Prevention
- C. Quality by Design
- D. Accuracy and Precision
- E. History of Quality Control
- F. Seven Basic Tools
- G. Statistical Inference
- H. Probability
- I. Descriptive Statistics
- J. Constructing a Frequency Distribution

II. Day Two
- A. Advanced Statistics and Probability
- B. Measures of Central Tendency
- C. Measures of Dispersion
- D. Definitions of Probability
- E. Permutations
- F. Combinations
- G. Important Discrete Distributions

III. Day Three
- A. Important Continuous Distributions
- B. The Central Limit Theorem
- C. Statistical Process Control
- D. SPC is an Attitude
- E. Variation
- F. Control Chart Basics
- G. Kinds of Control Charts
- H. Operating Characteristic Curves
- I. Implementing SPC
- J. Traditional Applications of SPC
- K. Non-Traditional Applications of SPC

L. Control Charts for Attributes
M. Kinds of Attribute Charts
N. Construction of a p Chart
O. p Chart Interpretation

IV. Day Four
A. Control Charts for Variables
B. Construction of a Variables Control Chart
C. Interpret Variables Control Chart

V. Day Five
A. Basic Questions on Variable Charts
B. Process In Control
C. Process Out of Control
D. Process Capability
E. Calculate Capability Index
F. Interpretation of Capability Index
G. Other SPC Techniques
H. Other Control Charts
I. Reliability
J. Reliability Measurement
K. Types of Failure
L. Historic Bathtub Curve
M. Cost of Reliability
N. Taguchi Loss Function

VI. Day Six - Computer Lab Instruction on SPC Software

The course was presented to personnel twice per week for three weeks from the Safety and Mission Assurance Directorate and specifically the Safety group, Trend Analysis group, some summer interns, and members of the I-NET multimedia group who would be working on the Internet version of the course.

The Internet version of the course is entitled "Introduction to Statistical Process Control". The opening page of the Internet course is depicted in Figure 1.The course, at the time of this final research report, is approximately 75% complete and should be ready to go by September 1, 1996. The NASA internal Internet address is http://kosh/spc/. The Internet version contains

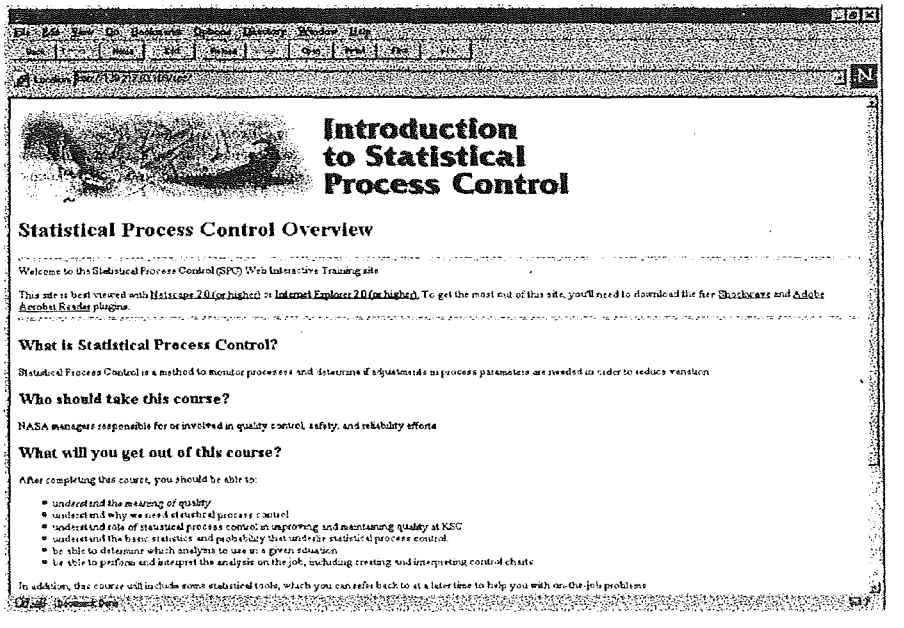


Figure 1. Sample SPC Course Web Page

essentially everything that the original course contained with the addition of animated graphics, interactive statistical calculation tools, exercises that need to be completed before progression into the course, a search engine fully referenced to the entire course, "Frequently Asked Questions" (FAQ) section, and links to the original PowerPoint presentations. If utilized by management, this on-line course will make the successful difference for SPC as a cost reduction too at NASA KSC.

3. **Application of SPC to Shuttle Processing**

NASA in Transition

There is a cultural change and a transitional process currently underway at NASA KSC. The best manner in which to describe what is going on is to examine a time line. There ***was*** the way NASA conducted Shuttle Operations in the past, there is the ***transition*** or how things are happening at the present time, and what ***will be*** when the transition is over and NASA's role has changed.

Was

The cultural change now underway relates back to the way things used to be done at KSC. Ever since the beginnings of NASA itself and manned space flight in the late 1950's and early 1960's, NASA has worked alongside contractors who provided the hardware for the missions. NASA did "hands on" work on a day-to-day basis, was involved in the on-site inspection and testing of the components, and had a partner relationship with the contractors over the years. This compatible relationship was successful in a mission sense. Goals like putting man into space, orbiting the earth, going to the moon, putting a man on the moon, the Space Shuttle program, etc., were all accomplished with only a few problems along the way.

The culture that saw this sympathetic and compatible relationship between NASA and its contractors also had few concerns concerning money. The programs were goal oriented, not cost conscious. If something was a problem, they just kept working on it until it was fixed and operational.

Transition

The current situation at NASA KSC can be identified as one in a state of transition. The United States government over the last ten to fifteen years has amassed an enormous national debt for a number of reasons. The annual deficit has continually added to the national debt year after year. Therefore, there is tremendous pressure on the politicians to decrease the annual deficit and thereby not adding to the national debt. Budget cutting of governmental agencies is one manner in which reductions in the U.S. government budget decreases deficit spending. NASA's budget has fallen victim to these budget cuts necessitating the need for NASA to change its role in the manner in which it conducts Shuttle Operations.

NASA is changing into a caretaker or oversight role from the contractor/partner relationship it enjoyed for so many years and over so many programs. There is a transition period when NASA must first realize that there is a change needed due to the different budgetary constraints, and secondarily figure out just how that change effects daily operations as well as the realization that the contractors will be doing all of the "hands on" work while NASA only verifies the results.

During the current transition period, which will probably last only a short time, the NASA efforts should be concentrated in the collection of pertinent data to be analyzed as comparisons to "before the transition" and "after the transition". The pertinent data to be collected is one of the decisions to be made. What metric technique for the chosen processes is another decision to be made. It should be pointed out that the recommended sequence for metric choice is simple before complex. Most processes can be evaluated with simple metrics and simple analysis

in order for management decisions to be made and verification of data to be made. It is easier to increase the complexity of analysis from an elementary level than it is to start with complexity and lose credibility.

Another important outcome of the transition period is a defined role for the analysis group after the transition. It is a reality that no one really knows what the post-transition environment will be like. Therefore, the data that is collected and analyzed, by what analytical methods, and how often will define the role of the analysis function to a great degree during the transition period. Goal determination for the analysis function should also be defined during the transition because of the timing of the contracts that take control of Shuttle Operations by contractors. If goals for analysis by NASA are not part of the contracts, it will be very hard to implement analysis after contract award and contract start.

Management is the key to success of any statistical analysis effort. If management either does not understand the statistics involved or how to interpret the results of the analysis in order to improve the conditions of the processes, then, essentially, the analysis effort is wasted. The true payback of the SPC technique actually ***begins*** with the charting of the process. Management at the highest levels (Center Director, Director, on down) must be able to assess the statistical control status of the selected processes in order to make decisions about reducing variation leading to reduced costs and reduced pressures. The transition period is a prime time for management to become trained (through the web-based course) in SPC and the associated interpretation of the analyses. Continuous improvement is the goal of SPC. Continuous improvement means lower costs. Proper use of SPC means processes under control. Processes under control mean higher quality at reduced costs.

Will Be

The new oversight role that NASA KSC is transitioning into is analogous to management in private industry. Industry management must always seek to produce at higher quality, reduced costs, and faster delivery because of competition. NASA KSC has all three of these components influencing their very survival in the next twenty years unlike what they experienced in prior years. Verification of contractor data now becomes the primary function of Mission Assurance. But just checking for validity in contractor reporting of data is not enough to determine if contractors are performing up to contract expectations. NASA must maintain the ability to perform selective data evaluation from contractor provided data independent of the contractor's reporting. In this way, the contractor can be held accountable for virtually anything that they do and not just on what they report.

A most important point can be made here also. That is that NASA KSC now can benchmark their operation to those of industry. Because of the competition factor, the cost sensitivity factor, and the continuing need for high quality, concepts such as Total Quality Management (TQM), Statistical Process Control (SPC), Just-In-Time (JIT) techniques, ISO 9000 requirements and certification standards, the criteria of the Malcolm Baldrige National Quality Award, and individual benchmarking measures such as wage levels, overhead rates, quality levels, safety performance, maintenance rates, etc., all now apply to Shuttle Operations at NASA KSC. In fact, these concepts and comparisons should contribute greatly to the overall goal of NASA KSC of a successful Shuttle program run efficiently and on budget.

Statistical Process Control

The use of Statistical Process Control is varied, versatile, and structured. Many users assume that the more complexity of the technique, the more help is received. This may be true in some cases. However, more than 75% of all data analysis can be done successfully and completely with just simple metrics such as bar charts, Pareto analysis, flow charts, etc. Why jump into a variable control chart analysis with all of the associated knowledge needed about statistics and how to interpret the chart correctly when Pareto analysis might give the order of the most significant problems to address.

The recommended sequence for the implementation of SPC techniques is: first, metrics; second, attribute control charting; and, third, variable control charting.

Metrics

SPC techniques such as bar charts, Pareto analysis, check sheets, flow charting, cause/effect diagrams, frequency distributions, and scatter diagrams can be classified as easy to learn but very revealing for data analysis. More than 75% of all data analysis can be completely accomplished through the use of these "seven tools". Therefore, they should always be the starting point of any problem areas in need of analysis. They are easy to use, easy to teach, easy to understand, easy to interpret, easy to implement, and highly effective. Management can make decisions concerning cost reductions, efficiencies, and scheduling of manpower based on them. Risk analysis can be based on them. The use of simple metrics to define problems in processes by simple analysis should always be the first tool pulled from the SPC toolbox.

Statistical Analysis - Attribute Charting

When the data that needs to be analyzed falls under the category of attribute data, or go-no go, conforming/ nonconforming, good/bad, etc., and statistical analysis is needed to determine if the data is operating in statistical control or not, then attribute charting is required. The data is target oriented meaning that there should be some chosen desired value for a process target. The data has a mean, but the mean may or may not be at the desired target. When the process control limits are calculated, the attribute data target is used to see if the process is even close to the desired value of the target. Therefore, the first management decision may well be to reset the desired target value to the process mean or to find out why the process mean is so far from the target value. An example of NASA KSC lost time cases rate for every 200,000 hours worked is shown in Figure 2.

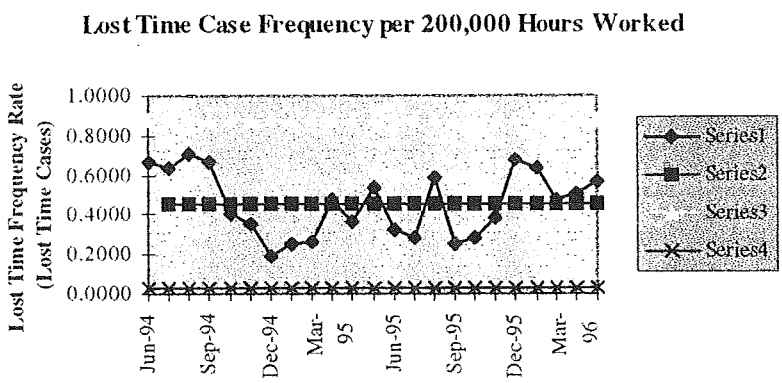


Figure 2. Sample Attribute Chart

When the process mean is used as the target and the control limits are calculated, outliers can be found which are points outside of the control limits. These points warrant investigation as to the reasons for the assignable causes relating to the points out of control. Then limits are recalculated without the outlier data, and the new chart with its recalculated control limits is evaluated again. This is done repeatedly until there are no points out of control. The process is then said to be "in control" which simply means that only random error is causing variation in the process. In this manner, 99.73% of all data should fall within the control limits. If any point falls outside of the limits, then an assignable cause is associated with the occurrence, the cause can be fixed, and management can continue to be confident in the output of the process without adjustment. There are also many other interpretation skills necessary to utilize control charts efficiently. These are discussed in the Internet version of the SPC course.

Statistical Analysis - Variable Charting

Whenever descriptive statistics, metrics, or attribute charting has highlighted an area that is in need of further or a

greater degree of scrutiny, then, perhaps, variable SPC charting is warranted (see Figure 3). The data must be in the form of continuous measurement data unlike attribute data which is, essentially, binary in nature. The difficulty with the understanding of the level of statistics involved with variable charting makes it the last resort which gives the greatest amount of information of all of the SPC techniques. For these reasons, variable charting should be used cautiously and only when necessary. In the author's experience, this technique is the most frequently misused, misapplied, and misinterpreted.

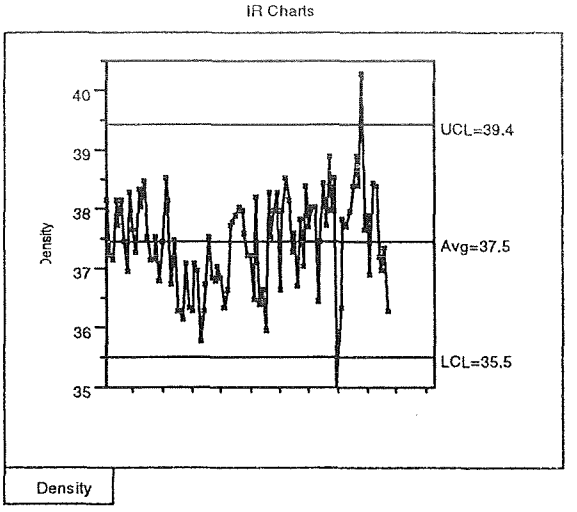


Figure 3. Sample Variable Chart - BTA Material

As with attribute charting, variable charting gives the state of control of the process, assesses the variability of the process, and provides management basis for decisions concerning the quality of the processes. The actual manipulation of the data is different from attribute charting. However, the interpretation of the chart is essentially the same. It is very important that the pertinent people involved in the process be the ones to investigate the outliers because of the need to involve everyone in the gaining of control of the process in question. The process drives continuous improvement and continuous improvement is, or should be, the goal of any process.

4. Expertise Shared as Internal Summer Consultation

There were several opportunities to do Statistical Process Control and statistics consulting with the people from RM. Some of the significant ones were:

- Jerrace Mack - RM → Inspection data correlation with structured surveillance data; first time quality data analysis
- Lisa Pantano - Safety → Monthly Mishap Severity Rate chart consultation from p chart to u chart because of Poisson distribution of rates
- Jim Lichtenthal - QE group → candidates for MIP reduction SPC analysis

5. SPC Software Evaluation

If SPC is to be utilized to its maximum efficiency at NASA KSC, software has to be utilized that is capable, accurate, easy to use, and fast. After looking at several packages provided by KSC and one brought by the author, the SAS Institute's JMP Statistical Discovery Software was found to be the most powerful. It also had the advantage of already being on-site and in limited use.

When data analysis is being performed on large data bases, this software almost immediately can formulate data into any number of needed methods of analysis from histograms to Pareto analysis to frequency distributions to control charts of both attribute and variable data. In short, this software is impressive. The only drawback is the lack of explanation for the novice. This again accentuates the worth of the on-line SPC course to provide the tutorials necessary for efficient use of the software.

6. Orientation for Future Consultation

One of the original goals of this summer's activities was to provide the author with sufficient orientation to NASA KSC's terminology, data gathering, organizational structure, and facilities so that an on-going relationship could be maintained for SPC consultation. In the immediate and distant future when questions arise concerning the

statistical analysis of contractor data by NASA, the need for consultation may become necessary. Now that orientation provided by this summer's research fellowship is completed, phone communication, fax communication, etc., can be utilized to provide direct and timely consultation.

7. Conclusions and Recommendations

As a result of the Summer Faculty Fellowship time spent and the analyses done, the following conclusions and recommendations are offered:

- Assess current attribute charts concerning visibility of ±3 sigma standard deviations from the process mean
- Assess current attribute charts for showing process average as the process target
- Recalculate control limits quarterly to determine if change has occurred - narrowing of limits indicates decreased variation and a tighter degree of control and a widening set of limits indicates increased variation and less control
- List all processes in the VAB, OPFs, launch pads, etc., where attribute data or measured data is collected as candidates for analysis by SPC techniques
- Do SPC analysis using JMP software on all candidate processes for determination of status of statistical control
- Based on SPC analysis, if in control, then basis for continued control is established
- Based on SPC analysis, if out of control, then begin determination of causes for out of control points, elimination of assignable cause data from database, recalculation of control limits, determination of causes of out of control points, etc., until stabile in control limits are found
- Management decisions can be made during this process of gaining control and after control is established to accomplish the following:
    - ⇒ optimize the use of resources
    - ⇒ reduce amount of inspections
    - ⇒ eliminate, correct, or reduce inspections (material qualification, test specimen preparation, correlation with weather data, consistent employee procedures, etc.)
    - ⇒ Reduction in manpower requirements
    - ⇒ Risk assessment
    - ⇒ Correlation of charting with Problem Resolution (PR) rates and PR severity data (number of pages per PR)

1996 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

DEVELOPMENT OF A COTS-BASED DISTRIBUTED COMPUTING ENVIRONMENT BLUEPRINT APPLICATION AT KSC

Dr. Isaac Ghansah, Professor
Department of Computer Science
California State University Sacramento
Sacramento, California

KSC Colleague - Bryan Boatright
Communications/Networks

Contract Number NASA-NGT10-52605

August 9, 1996

ACKNOWLEDGMENTS

I would like to thank my NASA colleague, Bryan Boatright for the opportunity he gave me to work on an interesting problem this summer. He was available when I needed clarification for what was required as well as getting in touch with appropriate personnel. He also gave me a private tour of network facilities at KSC as well as provide an environment which made me productive.

Mark Page and Henry Yu, both employees of INET were available to answer questions, professional or otherwise. I appreciate their patience with me.

Mr. Greg Buckingham (NASA program director), Dr. Roger Johnson (UCF Program Director), and Ms. Kari Stiles (Administrative Secretary) did an amazing job of organizing and coordinating many social and professional activities that made it an enjoyable summer.

John Schnitzius (NASA Advanced Network Systems) freely gave me ride to and from work. John gave my family and I so much help in many other ways that I do not have enough words to say thanks to him and his wife Loretta. Anna Maria Ruby of NASA and Carlos of INET were always ready to talk about all kinds of things. Thank you.

Finally, I want to thank my wife Becky and my children (Fred, Joy, and Kofi) for being patient with me while I was busy spending many hours on this project.

ABSTRACT

This paper describes a blueprint that can be used for developing a distributed computing environment (DCE) for NASA in general, and KSC in particular. A comprehensive, open, secure, integrated, and multi-vendor DCE such as OSF DCE has been suggested. Design issues, as well as recommendations for each component has been given. Where necessary, modifications were suggested to fit the needs of KSC. This was done in the area of security and directory services. Readers requiring a more comprehensive coverage are encouraged to refer to the eight-chapter document prepared for this work.

1.0 INTRODUCTION

There is currently a need to develop a distributed computing infrastructure at NASA Kennedy Space Center (KSC) to support current and future activities involving both national and international partners and customers. The deployment of such an infrastructure represents a foundation for a secure distributed computing base. Future services will be integrated into the environment efficiently and effectively. This infrastructure also provides the framework for developing distributed applications in a coherent manner. The objective of this research is to develop a commercial off-the-shelf (COTS) based open distributed computing environment (DCE) for application at KSC. The main benefit of an open system is availability of low-cost application software.

1.1 Current Distributed Computing Environment at KSC

In the case of network protocols, KSC currently has a myriad suite including AppleTalk, SNA, IPX(Novell), DECnet (Phase IV), and TCP/IP. Current applications include e-mail (SMTP MIME, Microsoft, ccmail, POP, X.400 and Quickmail), real time packetized telemetry (via PC GOAL), and data warehousing.

1.2 Future Distributed Computing Needs at KSC

From the management and policy direction point of view NASA is moving in two main directions. First, there is the performance based contracting shift which will allow NASA to divorce itself from the day-to-day oversight of contractors. Second, future operations of NASA will involve the international space station. All the major industrialized nations are involved with building parts of the international space station although the United States to a large extent and Russia to a lesser extent will share most of the burden.

Technically, the future of distributed computing within NASA is as follows. From the network protocol point of view, NASA is moving towards TCP/IP suite of protocols. From the point of view of e-mail/messaging systems, NASA is embracing two key technologies, including the Microsoft Exchange Mail and standard POP mail clients. As far as user interface is concerned, NASA is looking to using World Wide Web technology extensively. In the future, general applications will involve client-server computing/databases. In the light of this a standard naming convention is desired such that one could look at the database and retrieve any information automatically. Other future applications involve transaction processing for on-line procurement such as handling purchase orders, travel vouchers, time cards, etc. on line.

As pointed out earlier, the future operations of NASA involve the International Space Station. For that reason there will be need for communication with international partners, contractors, and Space itself (space station and other space components). The international partners also use different protocols. The foregoing calls for open standards, ones supported on multiple vendor platforms. In addition, the current view is to move

away from the mainframe environment (a central point of failure) towards distributed client/server environment (a decentralized system with no single point of failure).

Therefore, a comprehensive, integrated, open, secure, and multivendor distributed computing environment (DCE) is needed that is transparent to applications. The only DCE we know of that satisfies these requirements is OSF (Open Software Foundation) DCE. OSF is a consortium of all the major computing firms including IBM, HP, DEC, etc. Visibly absent from OSF are Microsoft, AT&T and Sun Microsystems. Although Microsoft makes DCE-like products, it does not currently have a comprehensive integrated DCE. AT&T and Sun have a different DCE called ONC (Open Network Computing). Major components of ONC DCE are the popular Sun RPC and NFS. Both Sun RPC and NFS are published standards. Unfortunately, ONC DCE is not comprehensive.

The rest of this report is organized according to the major components of a generic DCE. Section 2 discusses Remote Procedure Call (RPC). Section 3 elaborates on Distributed Time Service (DTS). In section 4, Distributed Security Service is extensively discussed using the SAAINT acronym. Section 5 discussed Distributed Directory Service (DDS), section 6 discusses Distributed File Service (DFS) , and section 7 covers Distributed Transaction Service (DTRS). Finally, section 8 concludes this paper. In each section we will look at each of these components from the point of view of why OSF DCE satisfies its requirements at NASA KSC. In addition, where necessary, we will state modifications that should be made to satisfy KSC requirements.

2.0 REMOTE PROCEDURE CALLS (RPC)

RPC is the foundation of any complex distributed computing environment. All communication that takes place in a DCE uses the RPC. RPC is based on the client/server model. Its objective is to simplify distributed programming so that the distributed programmer is productive.

2.1 RPC-based Design Issues

The first step in designing an RPC-based distributed application is to identify all the operations and specify them formally using an interface description language (IDL). Subsequently, an IDL compiler can be used to produce the necessary client stub code and server stub code automatically. A client program (written by the programmer) and the client stubs are linked into one object code. The server stub and server program (written by programmer) also are linked into one object code. Note however, that the client code runs on the client and the server code runs on the server. The interface specification is the key to the RPC. To check a design of a distributed application one has to look at the interface description. If the interface description is clear, concise, and complete, the analyst does not have to look too much into the other parts of the code (i.e. stubs, client program, server program etc.) to figure out what is going on.

2.2 Recommendation

OSF DCE RPC is based on the HP/Apollo RPC. Its interface definition language is clear, concise, and complete. For example, each interface is uniquely identified globally with a UUID (universal unique identifier). In addition although at-most-once is the default semantics, the interface specification allows the designer to specify other appropriate semantics (e.g. maybe, idempotent). Finally, if an operation (which is essentially a procedure) has parameters they are qualified with as input ([in],), output ([out]), or both as the case might be. On the other hand Sun RPC is at-least-once by default but has no way of specifying alternate semantics at the interface.

3 DISTRIBUTED TIME SERVICE (DTS)

A time service is needed for coordinating activities in a DCE. The time service keeps a single, global, monotonically increasing virtual clock which is used to synchronize machine (both client and server) clocks across the network. There are many situations in a DCE where synchronized physical clocks are needed. They include although not exhaustively: security, distributed file service, replicated systems, and concurrency control.

3.1 DTS Design Issues

Time service is obtained by arranging for some time servers to be connected to some authoritative time sources such as Universal Coordinated Time (UTC). Accurate UTC sources exist around the world. Machines acting as clients obtain synchronized time from these time servers. The challenge is to filter the variable network delays experienced by the time packets, and to make sure that time always goes forward and never backwards (this can be caused by inevitable clock drifts). The time service must be distributed so that there is no single point of failure. This is done by using multiple time servers. If the time service is to scale to a large number of machines, a hierarchical (multi-level) scheme is necessary for organizing the time servers. The most accurate servers are connected to the most accurate UTC receivers. These servers will be the highest level servers. Lower level servers receive their time from these higher level servers. Fault-tolerance is achieved by obtaining time information from at least three servers at the same time and rejecting ones which are totally out of line. The best estimate is selected from the midpoint of the accepted time intervals.

3.2 Recommendation

A DTS which satisfies all of these requirements is the OSF DCE DTS. The way the time service is organized is follows. Assuming that KSC has one DCE cell, at least three DTS servers will be required. Each NASA center will similarly have as least three DTS servers. At least one of the DTS servers will be designated as global DTS server. It will synchronize itself to other NASA centers' global DTS servers and to other KSC local servers.If there is the need to synchronize to machines in other countries, a similar

arrangement can be made for a higher level (say country) DTS servers to synchronize to other country DTS servers, and so on. A flexibility of the OSF DTS also allows coexistence with other time services such as the Internet NTP (Network Time Protocol).

4.0 DISTRIBUTED SECURITY SERVICE

Security should be one of the primary goals of a distributed system for an organization such as NASA KSC which is involved in a mission-critical enterprise. We believe a good, comprehensive, distributed security service should have the characteristics of a SAAINT (will not allow anything to go wrong). We use SAAINT as an acronym for Secrecy (which means the same as Privacy or Confidentiality in security circles), Authentication, Authorization (Access Control), Integrity, Non-repudiation, and Trust. Let us look at the technical aspects of the components of a distributed security service and show how they can be applied to a DCE for NASA in general, and KSC in particular.

4.1 Distributed Security Design Issues

There are two major ways of ensuring secrecy: secret key and public key. In the secret key case the same key, $K_{A\text{-}B}$ used to encrypt and decypt. $K_{A\text{-}B}$ is a secret known only by A and B. In the public key case, A has two keys: $KPub_A$ and $KPriv_A$. $KPub_A$ is a public key used to encrypt and whereas $KPriv_A$ (a private key that A keeps secret) is used to decrypt. Examples of secret key system are DES, IDEA, RC2, RC4, and RC5. Examples of public key systems include Diffie-Hellman key exchange, RSA, and DSS.

Authentication is the most difficult problem to solve in a DCE because of the openness of the network. In such an open network, someone (an intruder) can impersonate, replay , or modify a message. The best known authentication protocol is Kerberos. In Kerberos, a principal has to prove itself to another principal by obtaining a ticket from a key distribution server (KDS). The KDS has secret key it shares with each principal based on the principal's password. Kerberos allows mutual authentication and also allows A and B to perform secret communication after the authentication, using a random secret session key generated by the KDS.

Authorization is implemented by using an access control list (ACL). It is useful to think of ACL as a matrix. An element of the matrix, ACL_{ij} , is the set of operations that an authenticated principal (person, application program), i, can perform on an object (file, printer), j. Because keeping a long list for all principals might not scale well in a large organization such as NASA it useful to use a hierarchical organization based on groups. A possible organization for human principals is as follows. Every employee at KSC belongs to the group called KSC-Empl. KSC-Empl group in turn belongs to another group called NASA-Empl. With this scheme the management of the ACL is distributed because someone at each site keeps track of employees only at that site. However, the authorization decision will take a longer time. Note that the hierarchy can be extended to cover the entire world, such as International space agency.

Integrity protection is useful in documents that have been digitally signed. It is also useful in situations where we have an unsigned message such as e-mail which we want the computer to automatically detect if it has been changed by someone. Integrity is typically implemented using a **message digest (MD).** A message digest is a one-way function which takes a variable length message, *m*, and produces a fixed length bit string *MD(m),* usually 128 bits.*)*. Suppose Ann wants to send a message, *m*, to Bryan. Ann computes $MD(m, K_{A-B})$ and sends it along with the message, *m*, to Bryan. $MD(m, K_{A-B})$ is called **message integrity check (MIC)** or **message authentication code (MAC).** Another way of producing the MIC is to compute $K_{A-B}\{MD(m)\}$. Examples of message digests include MD2, MD4, MD5, and SHA. SHA seems to be the most secure because it produces a 160-bit hash.

Non-repudiation is used in the context of **digital signatures**. This is essential for unclassified memos, some e-mail, and electronic commerce (in the Internet for example).The digital signature scheme should be such that if A delivers a message to B, 1) B can verify, the identity of A; 2) A cannot later repudiate the contents of the message; 3) it is not possible for B to concoct the message A sent; and 4) the signature cannot be used for another document. Secret key signatures do not provide strong non-repudiation because the KDS (who knows everybody's secret) is involved in both signing and verification. Public key non-repudiation is strong. All real-world digital signature standards use the public key scheme. RSA is versatile, and has been around for a relatively long time. It is by far the defacto world standard. DSS, developed by NIST, is newer but not popular. In particular, RSA can be used to provide both secrecy and signatures, whereas DSS is designed only for signatures.

For the security service to be trusted, **key management** is essential. As usual let us look at how this is handled in both secret key and public key systems. In the case of secret key systems we have seen that the KDS keeps a shared secret for each principal. In a large system there could be multiple KDS's each responsible for users in its own **domain** (or **realm** as used in Kerberos). For example, in the case of NASA there will be a domain KDS serving the domain for each center. There will be one for KSC. Then for cross-domain authentication, every domain has to establish trust relationships with every other domain. In reality this means that each KDS has to be registered as a principal in the KDS of every domain it trusts. In a large system (such as international/intercorporate/interagency internetwork) this chain of trust relationships is not scaleable. A more scaleable arrangement is to allow a principal in domain A, to communicate with a principal in domain C, through domain B, without being registered in domain C. What we mean by registered is that they share a secret key. In this case there are two questions to answer. First, how does a principal in A find the domain path to principal in C? Second, how do we decide whether the domain path was acceptable? The first question is answered by making sure that the path followed is embedded in the naming convention. Since X.500 or DNS both use a hierarchical naming scheme this problem is solved. For example in the case of NASA the domain relationships will be in the from of the tree shown in figure 1. In figure1, KSC is a master domain at Kennedy Space Center.

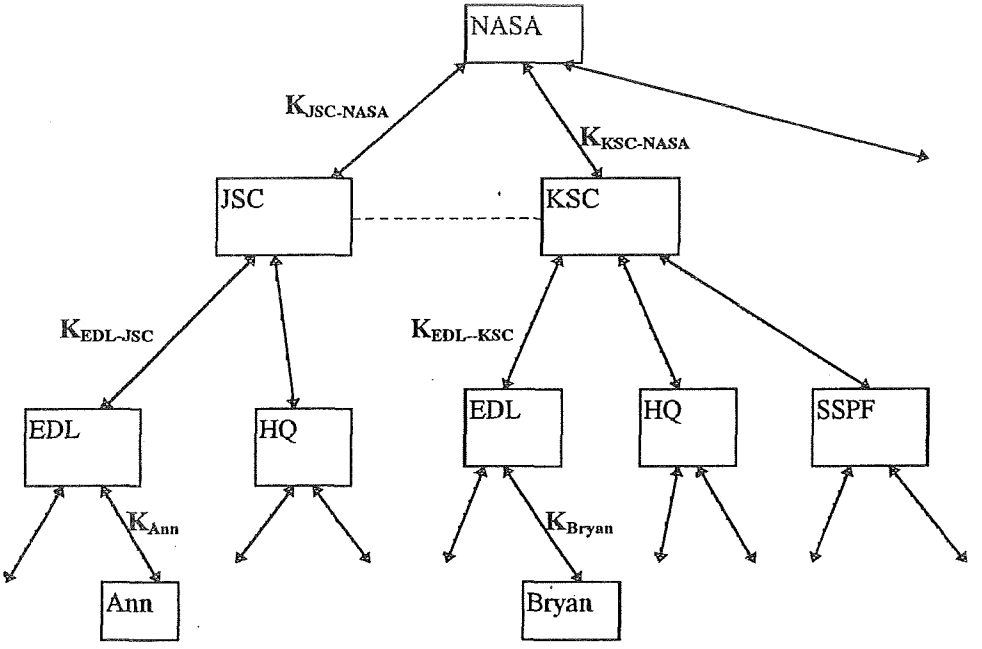


Figure 1 Hierarchical Trust Relationships in Multi-Domain Secret Key Systems

Suppose Bryan (or workstation on his behalf) wants to perform secure communication with Ann, Bryan will have Ann's X.500 name which is of the form /c=usa/o=nasa/ou=jsc/ou=edl/cn=Ann. Because the name follows the path in the hierarchy, Bryan can ask for a ticket to talk to Ann from EDL, who will ask for a ticket from KSC, and so on, until finally, Bryan can receive a ticket to talk to Ann. This is exactly how Kerberos V5 works. Kerberos V5 is used in OSF DCE. The answer to the second question (how do we know if the path transited is acceptable?) is left to the application by Kerberos V5. There is a TRANSIT field which lists all the domains transited in order to obtain the ticket. Ann will look at this list to determine if it should authenticate Bryan. Note that we do not have to follow the default hierarchy. We can use shortcuts. For example, in figure 1, if KSC's KDS trusts JSC's KDS (have registered shared keys) then we do not have to go through NASA.

Key management in public key systems is necessitated by the fact that we want to make sure that somebody's public key is authentic, otherwise, we could have impersonation. Authentic public key is ensured by using a KDS called **Certification Authority (CA).** The CA keeps the public keys of all principals. Anyone needing someone's public key asks the CA who returns a signed public key. A scalable CA is organized similar to the hierarchy in figure 1. Thus each KDS in a domain is a CA. Instead of shared keys a

domain CA issues certificates for its children and it obtains certificates from its parent. This hierarchy is sometimes called **certification hierarchy.**

4.2 Recommendation

The author recommends using Pretty Good Privacy (PGP) for secure e-mail and OSF DCE Security Service for everything else. The reasons are as follows. PGP is in the public domain, has received the test of time, and is more efficient than PEM. For example, PGP does compression. PEM does not. The only advantage PEM seemingly has over PGP is its specification of how an authentic public key is discovered. When an X.500-like certification hierarchy becomes available this argument will be moot. OSF DCE uses Kerberos V5 security system which has a better trust (key management) system than anything else available in its class such as Microsoft's Windows NT security.

5.0 DISTRIBUTED DIRECTORY SERVICE (DDS)

A distributed directory service (DDS) is a general purpose service which provides information to clients about services available in a distributed system. The distributed directory service is a fundamental building block of any distributed system because it is essential for network transparency and in many cases provides convenience for running applications automatically. For example a DDS can store a mapping of host names to IP addresses of servers so that clients can find them. Similarly, it can provide information about location, type, status of printers, e-mail address, phone number, and addresses of employees, to name only a few.

5.1 DDS Design Issues

Because a DDS deals with a large number of objects, the naming mechanism will affect its scalability. As noted earlier in this document, a hierarchical arrangement is the best way to design any scaleable system. Both DNS and X.500 provide hierarchical name space. Their naming formats reflect the hierarchical arrangement. Details can be found in chapter 5. X.500 is more flexible. On the other hand DNS has the larger installed base. Because a DDS is queried frequently for information (almost every operation requires access to it), its performance (availability and response time) is enhanced with replication and caching. Because the information in a DDS does not change often primary/secondary update protocol is used.

5.2 Recommendation

We strongly recommend X.500 for NASA in general, and KSC in particular. Because many current and almost all future applications will involve client-server computing a *standard naming convention* is desired such that one could look up and *retrieve any information automatically*. For example a print queue can be placed in a directory service such as X.500 so that different printers can be automatically selected for printing. In general, the future calls for a X.500-based directory service database that is a central

repository of all information. Note that the database will contain information about resources and not the resources themselves. Such a directory service will provide the following services and more for KSC.

- It will reduce network traffic as opposed to some current protocols that use broadcasting to find information about who has the service.
- It can answer a query such as "Where is the closest fax machine, printer, etc.?"
- It's database could be updated as needed. For e.g. printer status updated once a day.
- The X.500 directory service can be used as the basis for certificate management and distributed security service. This infrastructure is needed for verifying signatures, doing private/secret communication (e.g. e-mail) via encryption, etc. in the public key environment.
- The X.500 DDS can also be used as the basis for hierarchical trust relationships in secret key system such as Kerberos V5 used in OSF DCE. The current Microsoft Windows NT domain trust architecture being promulgated for NASA [ ] has the drawback of not being scaleable.

Currently, no technology exists that provides X.500-based DDS that is a central repository of most information. We do not know the exact reason for this. However, we suspect that it is because of its complexity. All is not lost because it is possible to custom-design such a system to fit the needs of an organization such as KSC.

We recommend using OSF DCE's DDS as a starting point for KSC. There are a number of reasons for this. First, its naming mechanism closely resembles X.500 although not exactly the same. Later, we will suggest recommendations to fit KSC needs. Second, it can be integrated with DCE security service. For example, a CDS server will not allow a user to access the database unless that user has been authenticated. Similarly, its authorization mechnism can be used to ensure that only the appropriate operation is performed. Third, because it is X.500-like the key management within secret key or public key case can be integrated relatively easily into its hierarchy.

How is DCE DDS organized? Basically, the system is organized into *cells*. Thus the service is called cell directory service (CDS). The client interface to CDS is called *clerk* and the directory database is called a *clearinghouse*. The clearing house is replicated and the clerk caches a successful information lookup to improve performance. Strictly speaking, CDS is not a central repository of all information in OSF DCE. OSF DCE actually has about four directory services. The first one is CDS which itself is basically a *name service*. It provides a mapping of host or RPC interface names to IP addresses. The second directory service is the *endpoint map* used for RPCs. The endpoint map is a mapping of RPC interface names and port numbers. When an RPC call is made, the client stub uses the interface name to fetch IP address from CDS. Then, the IP address plus interface name is used to fetch the port number from the endpoint map. The endpoint map is accessed from a well-known port number. The third directory service is the *file system's directory service*. This service is used to create, delete, etc. file directories as well as name, remove, etc., of files. Note that one of the major functions of a file directory service is to provide a mapping of file names to file IDs. The designers of DCE made the decision

to place the file directory service at the same location as the file service. To show that files are handled differently, every file name in CDS has fs (file system) as its prefix. The fourth directory service in OSF DCE is the *security directory service*. This directory which is next to the security server is called Registry database. It provides a mapping of principal names to principal attributes and privileges. Another directory which is closely related to it, *access control list* (used for authorization) is placed at the application server and accessed there. Once again, to show that principals are handled differently, the designers of DCE decided to prefix principal names with sec (security).

In order to design a directory service that is a central repository of all information such as in OSF DCE just described, we can pose the question. Is it possible to integrate these four (or five) directory services into a single one? To answer the question, let us look at how we could organize such a system. Basically we want a directory service that provides a mapping of names (file, RPC interface, persons, e-mail address, etc.) into IP adresses, port number, privilege attributes, etc. The key to doing this is to place in the DDS, for each unique name entry, not only the corresponding IP address but also all other relevant information such as port#, fileID, privilege attribute, ACL, etc. For example, in the case of RPC interface names, both IP address and port number will be in the DDS database. This way, there will be no need for an endpoint map at the server. If security is important for this interface, the privilege attributes will be entries as well. We could make similar suggestions for files or any other objects. We have answered the question.

However, we see a two disadvantages to such an integrated directory service. The first is security. It may not be a good idea to place security information in a system which is accessed by everyone unless we address all security issues associated with it and be sure it will work. The second issue is performance. We could potentially have more data crossing the network. To guard against that, it is better to place the data where it is needed. For example, in the case of RPC, having port information in the directory service means that we will transport it twice across the network, once during its retrieval from the DDS server database and then during the time the application server is accessed. On the other hand this scheme requires two accesses. So the tradeoff is between network transit time and database access time. With high speed network such as fiber the access time could be the more dominant. A research study has to be done here.

6 DISTRIBUTED FILE SERVICE (DFS)

The distributed file service (DFS) should be a major component in any useful distributed system. In particular, all application programs will be developed using the distributed file system. In addition, a distributed file service simplifies the way remote files are handled.

6.1 DFS Design Issues

Issues involved with the design of a DFS include: file directory service, file service interface, replication, and caching. The file directory service provides a mapping of symbolic names to universal file identifier (UFID). The file directory service is the user's

interface to the file system. Historically, files have always been named with a hierarchical organization. It is convenient if all machines have a uniform view of the DFS. Having the file directory service on a separate machine from the file service provides flexibility but performance is enhanced if they coexist on the same server. The file service interface issues involve whether the entire file is cached from the server or whether most accesses go to the server. The former scheme provides simplicity, scalability, and performance. Caching can be used to enhance both response time and availability and can be applied to the client or the server. Most DFS's apply it to both. Client caching poses the bigger problem - cache coherence. Replication can be used to enhance response time, availability, throughput via parallelism, and reliability. Note that replication is not used only for files. It can be used for directory services, key management services, time services, etc. to provide all the above requirements. Because replication produces consistency problems we must find ways of dealing with it. Methods range from primary/slave (the simplest) to quorum consensus (the most complex).

6.2 Recommendation

We highly recommend a DFS for NASA in general, and KSC in particular. A DFS simplifies the way remote files are handled. With NASA and its partners spread across five to seven continents in 15 time zones, with potential in the future of large volumes of data flowing between countries involved with the international space station, a DFS will make file transfer more efficient. For example, most data transfer today is done by e-mail or by FTP. Consider using FTP for a large file which you might decide to make changes to before sending it back to the remote location. A convenient way of handling such large files is to use a DFS. Instead of FTPing you can use a copy command. You can create a file remotely without too many problems. You can also update remote files easily. In three to five years ease of availability of data might be important because the international space station will be in place. A distributed file service will certainly help. The biggest problem NASA might have to face with distributed file service is security.

We recommend using OSF DCE's DFS. This DFS is similar to Andrew File System (AFS), developed at CMU and currently marketed by Transarc. The nice thing about OSF DFS is that it can be integrated within DCE. That is, DCE security can be used to provide authentication and authorization. This DFS also is scaleable (designed to support thousands of workstation) in an international scale. It uses X.500-like naming scheme, provides a uniform view of all files for all clients, provides replication, caching, and can be fully integrated into other components of OSF DCE such as the distributed time service.

7 DISTRIBUTED TRANSACTION SERVICE (DTRS)

A distributed transaction service (DTRS) is a service that supports atomic transactions involving multiple servers. In this service a client specifies the sequence of operations that are to comprise the transaction. The service guarantees to preserve the atomicity of the whole sequence. Typical examples of this include electronic funds transfer, on-line databases, distributed diaries, distributed editors, etc. Note that transactions are not only

needed for conventional bank-oriented databases but also for designing certain reliable distributed systems.

7.1 DTRS Design Issues

A transaction has the so called ACID characteristics of *Atomicity* (all or nothing), *Consistency* (stable states), *Isolation* (no interference), and *Durability* (permanence of commitment). The main issue that needs to be resolved with transactions is transaction management. There are two parts to it. The first is *recovery* which means that in the face of server or client crashes, the atomicity commitment will still be met. The second is *concurrency control* which deals with making sure that interleaved transactions produce the correct result even though they may be concurrently accessing common data. Recovery is handled by using the distributed two phase commit protocol. Concurrency control is handled by using locking, optimistic schemes, and timestamps. Details of each of these issues are covered in chapter 7.

7.2 Recommendation

Although transaction processing can be used for on-line applications such as handling POs, travel vouchers, time cards, etc., they are not required for these applications unless these applications involve multiple concurrent read/update. We do not see this in these applications. Where they will most likely be used is client server database type applications. We have not done enough studies to figure out where these applications are but we suspect that there are numerous applications where they could be used. In general, any application where multiple servers are solving a problem which is of atomic nature (reliable distributed system) will involve use of a transaction manager. If such a need arises we recommend using Transarc's Encina. Encina is an On Line Transaction Processing (OLTP) middleware designed to run top of OSF DCE. That is its major advantage. For example, it uses DCE directory and security services. It provides locking and two phase commit protocols for transaction management. Encina is marketed by most of the major OSF companies such as IBM, HP, etc.

8 CONCLUSIONS

This report developed a COTS-based DCE for NASA in general, and KSC in particular. We found out that based on current and future needs of KSC, OSF DCE holds the best promise of satisfying NASA application requirements. Given that NASA is interested in a commercial-off-the-shelf, comprehensive, open, secure, integrated, and multi-vendor distributed computing environment, OSF DCE is the only product we know of that comes close to meeting those requirements. Where necessary, we suggested modifications to DCE to satisfy NASA needs. This was done in the area of directory and security services.

Officially, NASA has not committed itself to any specific vendor's products. However, unofficially, there is a defacto commitment to Microsoft due to the fact that Microsoft

systems are pervasive. This was not necessarily done intentionally. For example, around 1994 or so a NASA-wide poll showed that about 80% of systems were Microsoft. It is true that Microsoft products are cheaper than UNIX-based systems. For example it costs less to manage an NT server that a UNIX server. In addition, it is understood that Microsoft is one of the few vendors which has committed itself to providing cross-platform systems for applications such as spreadsheets. On the other hand, Microsoft products are not necessarily open, scaleable, or secure. We have provided OSF DCE as an alternative. Indeed, in reading Microsoft white pages about future trend of their products, we see Microsoft providing OSF DCE-like solutions. For example, although Windows NT is supposed to have a time service, there is no mention of the time service protocol although we suspect that something similar to Berkeley algorithm (implemented as timed in 4.3BSD) is being used. The security system is different from Kerberos and it uses an unscalable key management scheme. The directory service is similar to DNS. Yet, the claim in their white pages seems to indicate that future systems will have X.500 and Kerberos. NASA does not have to wait for the future Microsoft release. OSF DCE is already here.

References

[CHA96] Chadwick D., *Important Lessons Derived from X.500 Case Studies*, IEEE Network, Mar/April 1996 pp. 22-34.

[COU94] Couloris G., Dollimore J., and Kingberg T., *Distributed Systems: Concepts and Design 2ed*, 1994, Addison Wesley.

[DIL95] Dilley, J., *Experiences With the OSF Distributed Computing Environment*, Proceedings of the 3rd IFIP TC 6/WG 6.1 International Conference on Open Distributed Processing, Brisbane, Australia, February 1995, pp. 465-475.

[KAU95] Kaufman C., Perlman R., and Speciner M., Network Security, 1995 Prentice Hall

[LIN95] Linington, P., *Reference Model for Open Distributed Processing: The Architecture*, Proceedings of the 3rd IFIP TC 6/WG 6.1 International Conference on Open Distributed Processing, Brisbane, Australia, February 1995, pp. 15-33.

[MIC1] Microsoft White Paper, *The Microsoft Strategy for Distributed Computing and DCE Services*, 1995, Microsoft Corporation.

[MIC2] Microsoft White Paper, *The Microsoft Directory Services Strategy*,, 1995, Microsoft Corporation.

[NAS96] NASA MSFC, *NASA Windows New Technology (NT) Enterprise Domain Strategy*, Revision 2, June 1996

[ROS90] Rose, M., *The Open Book*, 1990, Prentice Hall.

[ROS92] Rosenberry W., Kenney D., and Fisher G., *Understanding DCE*, 1992, Oreilly and Associates

[ROS93] Rosenberry W., Teague J., *Distributing Applications Across DCE and Windows NT*, 1993, Oreilly and Assoc.

[SCH96] Schneier B., *Applied Cryptography: Protocols, Algorithms, and Source Code in C, 2 Ed.*, 1996 Wiley and Sons.

1996 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

MICROINVERTEBRATES IN CELSS HYDROPONIC RHIZOSPHERE: EXPERIMENTAL INVASION AS A TEST OF COMMUNITY STABILITY AND A TEST OF A METHOD TO MEASURE BACTERIVORY

Dr. David G. Jenkins, Assistant Professor
Biology Department
University of Illinois, Springfield
Springfield, Illinois

KSC Colleagues - John Sager and Jay Garland
Life Sciences

Contract Number NASA-NGT10-52605

August 9, 1996

ACKNOWLEDGEMENTS

I wish to thank the following persons for so ably administering the NASA/ASEE Summer faculty fellowship Program: Ms. Kari Stiles, Mr. Gregg Buckingham, and Dr. Roger Johnson. I learned a lot about KSC and had a very productive summer thanks in large part to their efforts. I am grateful to have been here.

Many people helped me start, conduct and finish my work, most notably: Cheryl Atkinson, Kim Cook, Colleen Loader, Doug Schiedt, and Liz Stryjewski. I would especially like to thank my NASA colleagues Dr. John Sager and Dr. Jay Garland for their support and for the opportunity to work in the CELSS program, and the Garland family for their hospitality. I also thank Dr. Guillermo Colon for his friendship.

TABLE OF CONTENTS

ABSTRACTS

The following report consists of two separate draft manuscripts, each prepared for submittal to a peer-reviewed journal after KSC colleague editorial review and final revisions. Manuscripts were prepared in journal-specific format. References for the 2 papers have been combined for this report. The abstracts of the two manuscripts are as follows:

1. Experimental invasion of aquatic rhizosphere habitat and invertebrate communities. Jenkins, D.G. and J.L. Garland. To be submitted to Oecologia.

Invasion of communities or ecosystems may be considered a disturbance, by which the stability of a community may be judged and the mechanisms regulating community composition inferred. We experimentally invaded aquatic rhizospheres of hydroponically-grown plants with three sets of epiphytic communities (invaders). Invaded rhizospheres were either sterile or pre-inoculated with organisms that normally inhabit the rhizospheres (residents). About 1/2 of invaders were successful with or without residents present. Few invader species were clearly excluded by the presence of residents, and invasion did not displace residents. Environmental conditions partially excluded invaders and biotic interactions were far less important in determining community composition after invasion.

2. Lysozyme analysis (Gonzalez et al. 1993 and Vrba et al. 1993) is neither protistan- or bacterivore-specific. Jenkins, D.G., C. Atkinson and J.L. Garland. To be submitted to Microbial Ecology.

Enzymatic assays of protistan bacterivory have been proposed recently (Gonzalez et al 1993 and Vrba et al. 1993). Both methods rely on lysozyme hydrolysis of the β(1-4) glycosidic bond between N-acetylglucosamine and the fluorogenic compound 4-methylumbelliferyl. We evaluated the specificity of the Gonzalez et al. method (acid lysozyme) to protistan bacterivory, and found that 5 of 6 protists and 20 of 21 (95%) metazoan genera tested had acid lysozyme. Therefore, the methods of Gonzalez et al. (1993), and by inference Vrba et al. (1993), are not protistan-specific. Lysozymes (e.g., acid lysozyme, β-N-acetylglucosaminidase) hydrolyze both peptidoglycan and chitin, meaning that bacterivory cannot be distinguished from chitin digestion by lysozyme analysis. Therefore, the methods of Gonzalez et al. (1993), and by inference Vrba et al. (1993), are not bacterivore-specific. Consequently, lysozyme-based analyses of bacterivory can not be applied reliably to samples more complex than isolated protistans that have been demonstrated to not also digest chitin, as some predatory protistans may.

EXPERIMENTAL INVASION OF AQUATIC RHIZOSPHERE HABITAT AND INVERTEBRATE COMMUNITIES

1.1 INTRODUCTION

"Invasion" and "colonization" may be used synonymously, but studies of invasion usually focus on the success of introduced, sometimes exotic species, the effects of those species on resident communities, or the traits of a community that make it invasible (Lodge 1993, Robinson and Dickerson 1984, Crawley 1987). Colonization studies tend to focus on timing of arrival, distances from source populations, species richness, and extinction, usually related to MacArthur and Wilson's island biogeography theory (MacArthur and Wilson 1967). Robinson and Dickerson (1984) characterized invasion theories in two ways: (1) invasibility is independent of community composition and implies nothing about community stability (e.g., MacArthur and Wilson 1967); or (2) invasibility depends on community composition, because composition affects stability, niche availability and niche overlap (e.g., Elton 1958, May 1974). The link between complexity and stability is largely intuitive and has long been the subject of an unresolved debate (May 1973), but invasibility has been used as a test of community niche availability, complexity, and stability (Crawley 1987, Robinson and Dickerson 1984, Lodge 1983).

In the long run, community composition is the net result of regional-scale processes (dispersal) and local-scale processes, such as competition and predation (Ricklefs 1987), and communities will always change as a result of the interplay between these ongoing processes. In the short run, we desire certain communities to be stable and predictable: organisms inhabiting the rhizosphere of hydroponically-grown plants at Kennedy Space Center, FL (KSC) is one of these communities. Various crops (e.g., wheat, potato) are grown hydroponically under carefully-regulated conditions during development and testing of a Controlled Ecological Life Support System (CELSS) for long-term space missions (Wheeler et al. 1996). Long-term dependence on life support functions (gas and water processing, food) in space requires that systems be stable. Microbial communities in hydroponic rhizosphere are considered stable (Strayer 1994), and invertebrate community composition has been limited and repeatable (Jenkins 1995, unpublished results). However, stability is best measured by a system's response to change.

We experimentally introduced three sets of epiphytic communities from natural submerged aquatic vegetation into the rhizosphere of hydroponic plants. If simultaneous invasion by multiple species can be considered a disturbance to the resident community, then this experiment was a test of that resident community's stability in response to that disturbance. We inferred the effects of residents on the invasion process by comparing invasion in habitat with and without residents.

Three possible mechanisms may limit resident community composition: 1) Environmental conditions. Hydroponic plants are grown in a nutrient medium intended to promote plant growth, but that may inhibit growth of some aquatic invertebrates. In addition, rhizospheres are kept dark to inhibit algal growth: should aquatic invertebrates require algae for food and not be be able to subsist on bacteria, they may be excluded. 2) Biotic interactions. Invertebrates that exist in the rhizospheres may exclude subsequent species via competition and predation. 3) Dispersal. Hydroponic rhizospheres are in environmental chambers or the Biomass Production Chamber (BPC; Wheeler et al. 1996), all of which are housed indoors. Little opportunity likely exists for aquatic invertebrates to arrive.

Specific questions addressed in this study were: (1) Can "invaders" successfully colonize hydroponic systems that lack resident organisms? If so, then the environment is not exclusive, and dispersal would be indicated as limiting community composition. If not, environmental conditions would be indicated as limiting community composition. (2) Can invaders successfully colonize hydroponic systems that contain residents? If so, biotic interactions would not be indicated as excluding invaders. If not, and if invaders can succeed in sterile hydroponic systems, then biotic interactions would be indicated as limiting community composition.

1.2 MATERIALS AND METHODS

The experiment consisted of two treatments (resident organism presence/absence and invader inoculum source), organized in randomized blocks. Each block contained one replicate of each treatment combination. Blocks were used to account for a potential effect of position relative to the mist source, which was anticipated to affect plant growth and could serve as a micro-invertebrate contamination source. Therefore, experimental units were arranged in four blocks, from right to left, with one replicate of each treatment combination in each block.

Wheat plants (*Triticum aestivum* L., cv. Yecoro rojo) were grown hydroponically in jugs within an environmental control chamber, which permitted regulation of light and humidity. Wheat is commonly used in CELSS experiments; it grows quickly in hydroponic medium and provides abundant rhizosphere for aquatic organisms within several weeks. Humidity was provided as deionized water mist, blown from the right end of the chamber (Figure 1). Relative humidity cycled between 56% (mist off) and 90% (mist on), and light intensity ranged from 125 µE at 25 cm elevation (jug top) to 363 µE at 80 cm elevation (approximate plant height at experiment end).

All seeds and materials were sterilized and handled aseptically to minimize contamination by non-experimental organisms. Each of the 24 2.2-L Nalgene wide-mouth jugs contained one wheat plant, growing hydroponically in filter-sterilized (0.2 µm) ½-strength Hoagland's solution (Morales 1995). Seedlings were 5 days old at experiment start-up, and were inserted in a split polyurethane foam plug between two strips of 83-um mesh Nitex mesh (to act as a wick for hydroponic medium until roots grew longer). The plant and foam plug were then inserted in a hole at the center of the jug lid, and the lid was placed on the jug. Each jug was wrapped in black polypropylene to inhibit algal growth.

Each lid had two other holes, stoppered with rubber septa: one served as an aeration port, the other as a sampling port. Facility air was passed through a liquid trap and then a 0.2-µm filter to a syringe needle, which was inserted through an aeration septum. A Pasteur pipette fitted into the bottom of the septum passed air to the bottom of the container. Hydroponic medium samples were collected through the sampling port. The septum was removed and a sterile 10 mL serological pipette was inserted to the bottom of the container. As suction was applied, the pipette was lifted up in a spiral motion to sample the medium throughout the container.

Resident organisms came from hydroponic wheat rhizosphere of an ongoing EGC experiment at Advanced Life Sciences Support, Hangar L, Kennedy Space Center. A section of root mat was cut with scissors and interstitial medium was squeezed by hand into a beaker: this extract was the resident inoculum. Ten mL of mixed resident inoculum was inoculated into appropriate experimental containers immediately before seedling placement, on 4 June 1996 (Day -7 of the experiment). Organisms in remaining resident inoculum were identified and scored on a 1-5 relative abundance scale (1=rare, 5= very abundant). Plants and resident organisms were then left undisturbed for one week before invader inoculation. Hydroponic medium samples were collected on 11 June 1996 (Day 0), prior to invader inoculation, and examined as above.

Invader inoculum sources were 3 ponds in the Merritt Island National Wildlife Refuge, adjacent to KSC: ponds were selected to encompass a salinity range, and were designated as fresh (<1 ppt), brackish (7 ppt), and salt (16ppt). We expected that fresh, brackish and saltwater inocula would provide a range of invasion intensity, given that organisms adapted to saline conditions would not readily adapt to the hydroponic medium (< 1 ppt).

Invader inocula were collected on 10 June 1996 by simply filling a wide-mouth 1 L jar with submerged aquatic vegetation and pond water. Samples were placed on ice until return to the laboratory, where they were held at room temperature overnight. The next morning (Day 0), samples were shaken vigorously and sieved on 0.5 mm mesh to separate plants and macroorganisms from inocula. This procedure was repeated with approximately 50 mL of the sieved water to further extract organisms from plants. Extract was then mixed and 10 mL were pipetted into appropriate jugs. Organisms in invader inoculum were identified and scored for relative abundance as above.

Combination of the above treatments resulted in the following treatment combinations: freshwater invaders, non-resident (FNR); freshwater invaders, resident (FR); brackish invaders, non-resident (BNR); brackish invaders, resident (BR); saltwater invaders, non-resident (SNR); saltwater invaders, resident (SR).

Further samples were collected from hydroponic medium and examined microscopically (as above), using 10 mL pipettes (as described above), on the following dates: 14 June (Day 3), 25 June (Day 14), and 8 July (Day 27). The experiment ended 9 July (Day 28), when plants were removed after 28 days of hydroponic growth. Plant roots were cut with scissors and placed in a sterile, 50-mL centrifuge tube containing 30 mL sterile saline solution (0.85% NaCl). Sterile glass beads were added and samples were inverted 10 times to extract invertebrates from rhizosphere. A 10-mL subsample was taken for microscopic examination, as above. Plant roots and shoots were dried at 70° C for 48 hrs and weighed.

Data were analyzed by both univariate and multivariate techniques. Invertebrate species number and plant dry weights were analyzed by ANOVA and Tukey's HSD (SAS 1990). Invertebrate relative densities (Day 27 media and rhizosphere samples) were analyzed by detrended correspondence analysis (DCA) and cluster analysis (using centroid distances) was run for rhizosphere results.

1.3 RESULTS

Wheat plants grew to approximately 2 g total dry weight. Some plants that received resident inocula had brown roots and lesser root mass, probably related fungal infection, but root, shoot, and total dry weights were not significantly different among treatments or blocks.

Organisms present in resident, freshwater, brackish and saltwater inocula are listed in Table 1. Note that all resident organisms had been observed in previous samples from CELSS rhizospheres. In addition, two other rotifers, *Lecane inermis* and *Lecane hamata*, have been observed in previous CELSS rhizosphere samples and were observed in subsequent samples from jugs inoculated with residents: it is likely that they were present as resting eggs or were very rare in initial inoculum.

Composition in hydroponic media (and presumably rhizosphere) underwent successional change. For example, *Actinophrys sol* was numerous in initial resident inoculum, but faded away during the experiment, likely related to static conditions in the experimental jugs compared to flow-through conditions in EGC and BPC experiments.

Species number in hydroponic media decreased after inoculation in all experimental treatments (Figure 1). Resident inoculum contained 8 species at Day -7, but media inoculated with residents had an average of 3 species at Day 0, immediately prior to addition of invaders. Fresh and saltwater species numbers dropped most sharply after inoculation (Figure 1). Jugs that were not inoculated with resident organisms at Day -7 were uncontaminated at Day 0.

Species numbers generally increased slightly following Day 0: all jugs had received invaders, and ½ of the jugs had also received residents one week earlier. As expected, saltwater and brackish invaders contributed little to species number (Figure 1). The greatest increase occurred in FNR media, reaching an average of 15 species at Day 14. However, FR and FNR treatments had similar species numbers by Day 27 (Figure 1).

Invader source had a significant effect on species number in hydroponic medium (ANOVA; $p=0.0001$), and jugs receiving freshwater invaders had significantly more species than jugs receiving brackish or saltwater invaders ($p=0.05$, Tukey's HSD). The presence/absence of resident organisms prior to invasion had no significant effect on species number. Blocks were also not significantly different, and no significant interaction existed between the two main treatments.

Rhizosphere samples were analyzed separately from media samples. Invader source, resident presence/absence, and source-resident interaction significantly affected species number (ANOVA; $p=0.0001$, 0.005, 0.014, respectively). Communities invaded by freshwater organisms had more species than those invaded by saltwater and brackish organisms, and communities with resident organisms had more species than those without residents (Tukey's HSD, $p=0.05$). Also, rhizosphere communities were significantly more diverse in FR jugs than other jugs (Figure 1).

Multivariate analyses also indicated differences among treatments. Detrended correspondence analysis (DCA) of rhizosphere data yielded three groupings along the first axis: BNR and SNR communities, BR and SR communities, and FR and FNR communities (Figure 2). The DCA results were consistent with cluster analysis of rhizosphere data (Figure 3). Finally, the same general pattern among treatments was shown by DCA of Day 27 media + rhizosphere data, although media and rhizosphere patterns were separated, especially for freshwater treatments (Figure 4). This analysis indicated that treatment effects were consistent in different micro-habitats, and that media samples yielded different composition than rhizosphere samples, as was expected.

Inspection of FR and FNR rhizosphere community composition revealed that FR and FNR treatments shared 12 of the 24 species observed in both treatments (Table 1). Of those 12 species, 6 were in resident inoculum or have been observed in resident communities previously (Table 1). This overlap in invader and resident composition is consistent with previous chance colonization of BPC and EGC rhizospheres from regional freshwater sources. Nine other species occurred in FR rhizospheres only (although infrequently), and 3 of those 9 species (*Cephalodella gracilis, Cyclidium, and Stylonichia*) were in the initial resident inoculum. The other six species have never been observed in samples of CELSS rhizosphere and probably entered as freshwater invaders. Three genera were exclusively observed in FNR rhizospheres: *Litonotus, Hartmanella, and Onychodromus*, indicating that they originated from invader inocula.

Fresh, brackish, and saltwater invaders were used because we anticipated a gradient in invasion strength due to varying tolerance of low-salinity culture medium. Invader sources did vary in invasion strength, but not due to osmotic stress. The fractions of fresh, brackish, and saltwater species that survived to Days 27 or 28 were comparable (52%, 58%, and 56%, respectively). Freshwater invader inoculum was more diverse than brackish or saltwater inocula, thus providing greater invasion pressure, regardless of osmotic tolerance among species from different habitats.

Species were individually judged for their ability to colonize the experimental systems. The compatibility of the environment and the effect of biotic interactions were inferred from data for rhizosphere samples (Day 28) or media samples at Day 27 (Table 2). We considered presence/absence at 27 or 28 days post-invasion to be a reasonable test of success for communities composed of protists and microscopic metazoans. We judged taxon presence in 3 or 4 replicate jugs of a treatment to indicate success: presence in 1 or 2 replicates was considered unclear evidence of success or exclusion.

One-half (51%) of the invader species clearly survived 28 days: more invader species were present at Days 27 and 28 than residents (Table 2). We considered this evidence of successful colonization for those species. In addition, consider the number of species present in FNR treatments (Figure 1): clearly, invaded jugs accrued invading species. However, about one-third (39%) of all invader species did not successfully invade the hydroponic habitat (by our criterion of presence at Days 27 or 28). Environmental conditions apparently excluded those species, assuming that biotic interactions among members of the inoculated community did not alter their survivorship.

About two-thirds of the invaders (68%) successfully colonized with no apparent effects by residents, and 3 species (*Blepharisma, Colurella*, and *Encentrum*: 8%) were more successful in the presence of residents. Only 5 (13%) were clearly excluded from habitats containing residents, and four other species (11%) may have been excluded: mechanisms of exclusion are not known. Most invaders (76%) could successfully colonize systems with residents, and the number of species excluded by resident presence were outnumbered by those excluded by environmental conditions.

1.4 DISCUSSION

Can invaders successfully colonize hydroponic systems that lack resident organisms? The majority of species successfully colonized non-resident rhizospheres, although about 1/3 of the species did not. Environmental conditions apparently excluded some species, assuming that biotic interactions among members of the inoculated community did not alter their survivorship. Therefore, environmental conditions partially limited invertebrate community composition, and this effect was not limited to species

from brackish or saltwater habitats: 9 of 24 excluded species were in initial freshwater inoculum.

Given the partial overlap in invader and resident taxonomic composition, it is likely that regional habitats served as source pools for some species when they first colonized the facility and became current rhizosphere residents. It is also likely that similar colonization events will continue, albeit at a low rate due to isolation of hydroponic habitats within the facility. This process is suggested by the observation of *Floscularia* and *Stentor* in initial resident inoculum: neither genus had been observed previously in CELSS hydroponic systems and are probably recent colonists. Other organisms were successful upon experimental invasion but have not been observed in resident communities, probably because they disperse less readily.

Can invaders successfully colonize hydroponic systems that contain residents? About 3/4 of the invaders did, and environmental conditions excluded more species than biotic interactions. Lodge (1993) pointed out that reliable statistics on invasion success rates are rare for a variety of reasons, including difficulties in recording unsuccessful invaders, biases to easily-observed species, paucity of data on many taxa, and a focus on potentially-successful species in planned invasions for biological control. Nevertheless, Lodge (1993) cited values of 10% (Williamson 1989) and 1-24%, ranging up to 46% for one taxon (Vermeij 1991). The much higher invasion rate (76%) attained in our experiment supports our conclusion that rhizosphere communities were invasible.

Resident communities were relatively simple, potentially containing up to 12 species. Simple communities should be more invasible than speciose communities, related to cumulative effects of resident predators and competitors on invaders (Lodge 1993). This logic most clearly applies to single invaders entering speciose resident communities. In the case of a speciose invasion of a simple resident community, cumulative effects of invader predators and competitors might be expected to exert strong pressure on residents. However, resident species were not excluded and did not clearly have reduced densities upon invasion. We find it difficult to attribute the invasibility of the communities to biotic interactions by diverse invaders.

The greater diversity of FR treatments (versus FNR), and the similarity of FR and FNR in multivariate analyses and species lists all indicate that invasion did not simply displace resident communities. Likewise, resident communities did not seem to exclude more than a few invader species (e.g., *Litonotus* and *Hartmanella*). Therefore, it seems unlikely that biotic interactions were strong enough to yield detectable changes in community composition. If anything, invaders supplemented the resident community, and may have reinforced resident populations in cases of taxonomic overlap (e.g., *L. hamata, L. inermis, R. rotaria*). Lodge (1993) stated that "many, if not most, colonists have little impact on invaded communities," and Vermeij (1991) showed from paleobiological evidence that invasion often simply increases species richness. Of course, strong direct and indirect interactions have been documented, with predation and habitat change most often invoked as mechanisms of interaction between invaders and residents; competition is more difficult to demonstrate and rarely invoked (Lodge 1993). Predators and potential competitors existed in both invader and resident communities, but neither interaction strongly affected community composition. It is unlikely that invertebrates would have significantly changed rhizosphere habitat.

Crawley (1987) stated that "a community is invasible when an introduced species is able to increase when rare," and that "it is impossible to to recognize the existence of a vacant niche without the empirical attempt to establish a given species in a given environment." By these criteria, our results show that rhizosphere communities were invasible many times over, because multiple open niches exist. Hydroponic rhizospheres appear to contain vacant niches, in that invader organisms successfully colonized without displacing resident organisms. An invasible community is often considered to have an open niche for the invader, although it is usually difficult to establish that open niches exist. Crawley (1987) listed three problems with the concept of open niche: (1) investigators measure niche differently (i.e., niche is in the eye of the beholder); (2) invasion alone does not indicate an open niche, because invaders may displace residents, rather than fill an open niche; and (3) apparently unexploited resources (perceived open niches) may not be exploitable for hidden reasons (e.g., low resource quality, etc.). We avoided these three pitfalls by making no assumptions about niche dimensions, tracking both resident and invader

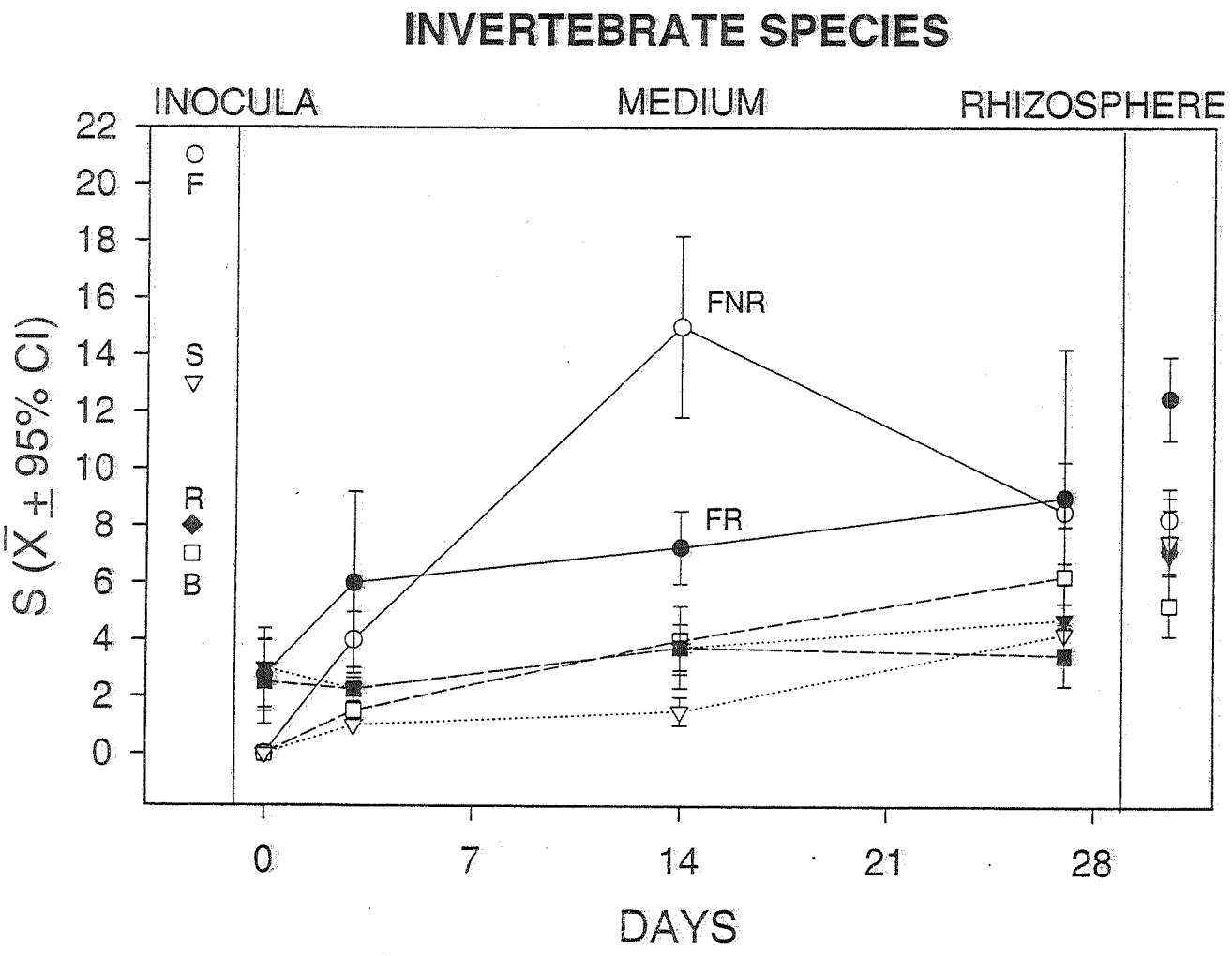
populations, and by evaluating invader success rather than perceived resources.

An alternative to multiple open niches may explain the invasibility of the rhizosphere communities: abundant resources. Bacterial densities in the CELSS rhizospheres are commonly about 10^{10} cells / cm^2 root surface area (Strayer 1994). Some residents may occupy niches that largely overlap or coincide with some invader niches, but competitive exclusion would not take effect until resources became limiting; an unlikely event at such high bacterial density. Limiting resources would not occur until either bacterivory became intense (very high invertebrate density), or microbial production was reduced (low nutrient levels, etc.). Experiments with more diverse and dense resident communities and varying resource availability may be needed to detect competitive exclusion.

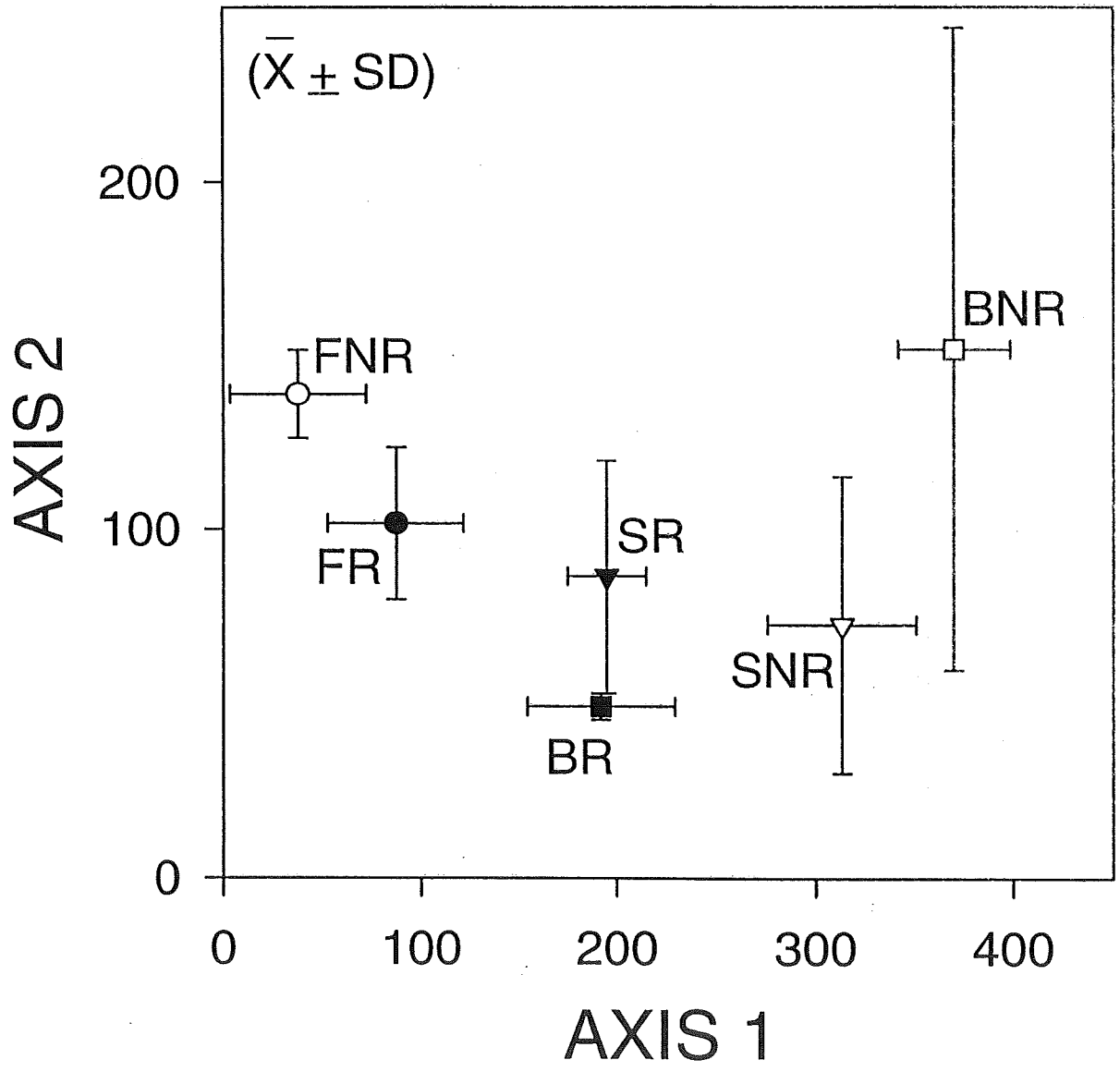
Invaded communities attained different end states. Our experimental design focused on the invasibility of rhizospheres with and without the presence of resident organisms: we can not compare invaded end states to uninvaded end states to statistically evaluate differences. Likewise, we cannot evaluate temporal patterns as a measure of stability, because media and rhizosphere communities differed and rhizospheres were sampled only at the end. However, the rhizosphere environment (without residents) was invasible by about one-half of the tested organisms, and invasion of those environments by organisms from different sources led to different outcomes. This result is not unexpected, given different assemblages in original inoculum.

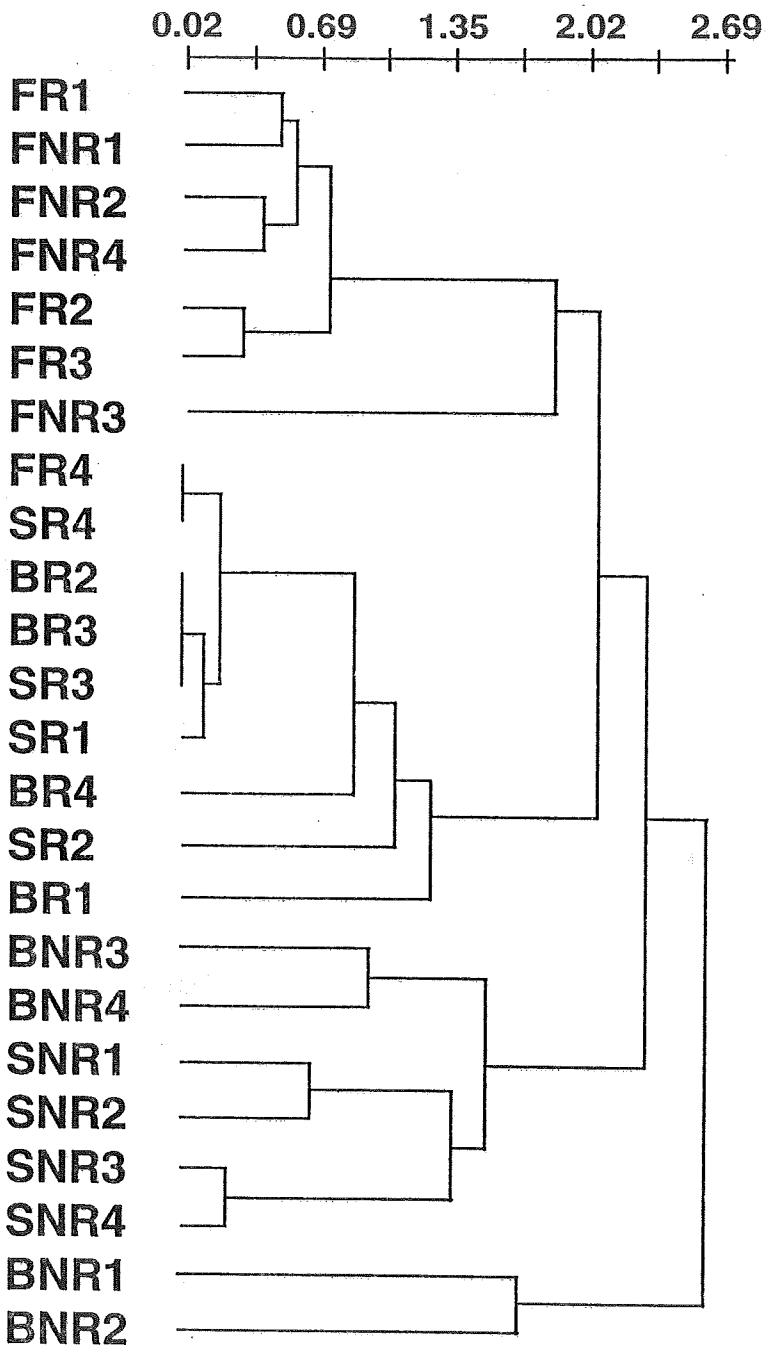
More importantly, invasion by different communities led to different outcomes despite the presence of residents. Differences among invader source treatments were lessened by the common denominator of residents, but remained. The similarity of freshwater-invader communities in multivariate analyses indicates that invaders were important in characterizing both treatments: this pattern would not exist if residents had resisted invasion. Brackish and saltwater treatments with residents were more clearly separated from those without residents: residents played a greater role in defining those communities, due to the lesser numbers of invader species in brackish and saltwater inocula.

1.5 LIST OF FIGURES


INVERTEBRATE SPECIES
INOCULA
MEDIUM
RHIZOSPHERE
S ($\bar{X}$ ± 95% CI)
22
20
18
16
14
12
10
8
6
4
2
0
F
S
R
B
FNR
FR
0
7
14
21
28
DAYS

RHIZOSPHERE INVERTEBRATES DCA




0.02
0.69
1.35
2.02
2.69
FR1
FNR1
FNR2
FNR4
FR2
FR3
FNR3
FR4
SR4
BR2
BR3
SR3
SR1
BR4
SR2
BR1
BNR3
BNR4
SNR1
SNR2
SNR3
SNR4
BNR1
BNR2

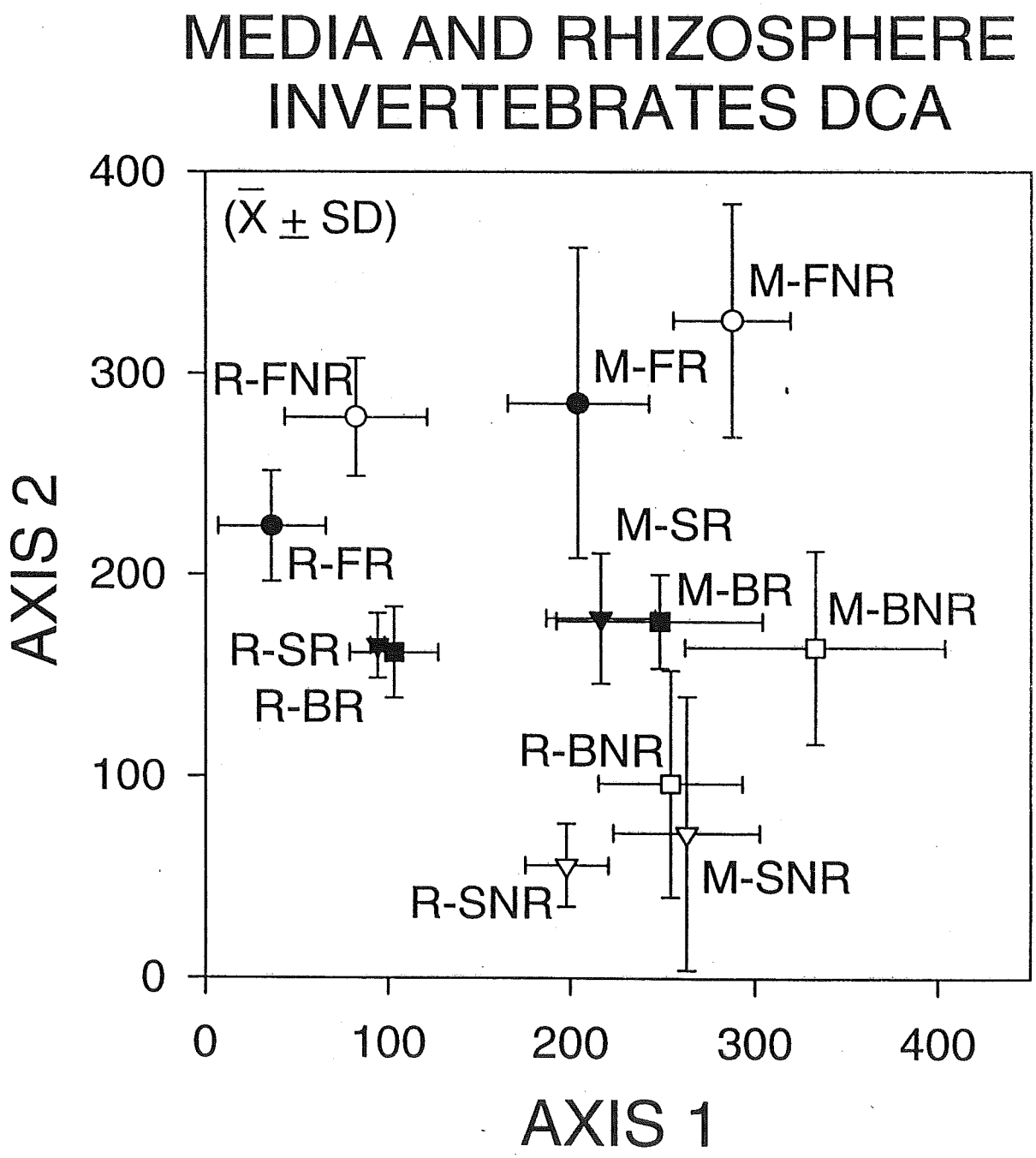

MEDIA AND RHIZOSPHERE
INVERTEBRATES DCA
(X̄ ± SD)
M-FNR
R-FNR
M-FR
M-SR
R-FR
M-BR
M-BNR
R-SR
R-BR
R-BNR
M-SNR
R-SNR
AXIS 2
AXIS 1
400
300
200
100
0
0
100
200
300
400

Table 1. Invertebrate taxa observed in initial resident inoculum, freshwater, brackish, and saltwater invader inocula, and rhizosphere samples after 28 days. An X indicates presence; multiple X's for rhizospheres indicate the number replicate jugs in which taxa was observed (maximum = 4).

					RHIZOSPHERE (DAY 28)					
TAXA	RESIDENTS	FRESH	BRACKISH	SALT	FR	FNR	BR	BNR	SR	SNR
Actinophrys	X							X		
Arcella					XXXX	XXX				
Aspidisca					X					
Blepharisma					X		XXX		XX	XX
Bodo saltans	X				XX	X	XX	XX	XX	XXXX
Bodo sp.	X									
Bryophyra		X								
Cephalodella	X	X			XX		XX		XXX	X
Chaetonotus		X			X					
Cinetochilum								XX		X
Coleps		X		X						
Colurella		X	X		XXXX	XXXX			XX	
Cyclidium	X				XX		XXXX	X	XXX	XX
Dileptus anser								X		
Encentrum				X			XXX		XX	X
Eosphora		X								
Euchlanis dilatata		X								
Euglena		X								
Euplotes					X	XX		X		
Floscularia	X			X						
Gonostomum					X			X		XX
Hartmanella						XXX				
large ciliate			X							
Lecane asthena								XX	XX	
Lecane cornuta					X					
Lecane hamata					XXXX	XXX	X		X	
Lecane inermis		X	X		XXXX	X	XXX		XXXX	X
Lecane obtusa							X			XXX
Lecane pyriformis					XXXX	XXX				
Lecane tethis										XXX
Lepadella					XX	X				XX
Limnias		X								
Litonotus				X		XXXX		X		
Mayorella		X			XXX	XXX	XX		X	
Microdalyiella			X	X						
nematode		X		X	XXXX	X	X			
Notomatta		X			XX					
Onychodromus						X				
ostracod		X		X						
Peranema								XXX		XXX
Pleuronema	X									
Rotaria rotaria		X			XXXX	XX	XXXX		XXXX	
small ciliate		X	X	X	XX	X		XX		XXX
snail			X							
Squalophyra		X								
Stenostomum				X						
Stentor	X	X								
Stylonichia					X					
Trachelophyllum		X								
Trichocerca		X								
Vaginocola								X		X
Vorticella		X	X				XXX	XXX	XX	X

1: FR = freshwater invaders, residents present; FNR = freshwater invaders, no residents present; BR = brackish wa Invaders, residents present; BNR = brackish water invaders, no residents present; SR = saltwater invaders, resid present; SNR = saltwater invaders, no residents present.

2: In addition to taxa listed above, others were observed in media samples at Days 3, 14, or 27, including: *Actinosphaerium, Chilodonella, Colpidium, Cyphoderia, Difflugia, Dissotrocha, Gastrostyla, Gromia fluvialis, Halteria, Heteromonas, Mesodinium, Metopus, naidid oligochaete, Nuclearia, Oxnerella maritima, Pamphagus, Salpingoeca, Tetrahymena, Thecamoeba, Urocentrum turbo, Uroleptis.*

Table 2. Summary of invasion patterns for the rhizosphere environment and biotic interactions between invaders and resident communities. The rhizosphere environment was judged to be compatible for invaders if they survived in non-resident treatments after 27 or 28 days. Biotic interactions were judged for those invaders that survived in non-resident treatments.

Environmental Compatibility for Invaders

	Number	Percent of Invaders
Survived:	31	51%
Not Survived:	22	36%
Unclear (rare):	8	13%
Total Invaders:	61	100%
Residents and Inferred Residents:	12	

Biotic Interactions for Invaders that Survived Environmental Conditions

	Number	Percent of Invaders
Excluded by Residents:	5	16%
Possibly Excluded by Residents:	4	13%
Promoted by Residents:	3	10%
Unclear (mixed or rare):	19	61%
Total:	31	100%

LYSOZYME ANALYSES (GONZALEZ et al. 1993, VRBA et al. 1993) ARE NEITHER PROTISTAN- OR BACTERIVORE-SPECIFIC.

2.1 INTRODUCTION

Microbial food webs have been acknowledged for some time as being important in aquatic ecosystem energetics and materials cycling (Azam et al. 1983, Cole et al. 1988, Pomeroy and Wiebe 1988). Protozoans are most often considered the major grazers of bacteria in aquatic systems, but metazoans also feed on bacteria and can indirectly regulate bacterial composition and productivity via predation on protozoans (Porter et al. 1985, Berninger et al. 1991, Jurgens et al. 1994). Bacterivory by protozoans and metazoans has most often been estimated by quantifying fluorescently- or radio-labelled bacteria, or by inference from changes in bacterial density (e.g., Sherr et al. 1987, Nygaard and Hessen 1990, Jurgens et al. 1994). Various problems exist with these techniques, and alternative methods that are based on analysis of lysozymes have been described (Gonzalez et al. 1993, Vrba et al. 1993). We use the term "lysozymes" here to include all forms of the enzyme that hydrolyzes peptidoglycan, regardless of specific nomenclature (Stryer 1981, Cabezas 1989).

Gonzalez et al. (1993) developed the acid lysozyme (L_{ACID}) assay to measure lysozyme activity in protistan cell lysates at acid pH, which has been demonstrated to occur in protistan food vacuoles (Sleigh 1989). Gonzalez et al. (1993) demonstrated maximal L_{ACID} activity at pH 4.5, and a strong relationship between bacterivory rates (fluorescently-labelled bacteria intake) and acid lysozyme activity in a variety of marine water samples and protistan cultures.

While clearly a valuable tool for measuring bacterivory by some protistan isolates, the L_{ACID} assay must pass another test for it to be considered protistan-specific: metazoan bacterivores must not produce L_{ACID} as well. Gonzalez et al. (1993) argued that the L_{ACID} assay is protistan specific because marine bacterial exoenzymes are inactive at pH < 5 and function maximally at seawater pH. Water samples often contain a variety of metazoan taxa, especially samples from inland, estuarine, or benthic marine waters. Some metazoan invertebrates (e.g., rotifers) are comparable in size to some protists, and may not be easily separated from protists in samples prior to L_{ACID} analyses. If metazoan bacterivores contribute to measured "protistan" L_{ACID} activity, protistan bacterivory would be overestimated and metazoan importance underestimated.

In addition, protistan L_{ACID} must be shown to differ from protistan and metazoan lysozymes that can act at acid pH but that may be present for reasons other than bacterivory. Lysozymes hydrolyze the β(1-4) glycosidic bond between N-acetylmuramic acid (NAM) and N-acetylglucosamine (NAG) in peptidoglycan, but also hydrolyze chitin, which is composed of only NAG molecules joined by β(1-4) glycosidic bonds (Stryer 1981). Therefore, protistan and metazoan predators of chitin-bearing prey (e.g., rotifers, nematodes) and chitin-bearing metazoans that molt should have lysozymes that hydrolyze the analog substrate, 4-methylumbelliferyl β-D-N,N',N''-triacetylchitotriose (MUF-CHT), used by Gonzalez et al. (1993). Chitotriose is a trimer of NAG molecules and an intermediate breakdown product of chitin digestion (Vrba et al. 1993, Gooday 1990). If organisms have lysozymes that are not involved in bacterivory but that react with MUF-CHT, the method of Gonzalez et al. (1993) would overestimate protistan bacterivory.

Overlap in metazoan and protistan sizes and the presence of chitinolytic lysozymes should also affect the method of Vrba et al. (1993), which uses 4-methylumbelliferyl β-N-acetylglucosaminide (MUF-NAG) hydrolysis as a measure of protozoan bacterivory. The substrates used by Gonzalez et al. (1993) and Vrba et al. (1993) differ only in the presence of two additional NAG monomers: otherwise the MUF-CHT and MUF-NAG substrates are identical (Sigma Chemical Co. Product Structure Data Sheets M5639 and M2133). The enzyme assayed by Vrba et al. (1993), β-N-acetylglucosaminidase, is a lysozyme by definition (Stryer 1981). We did not specifically test the method of Vrba et al. (1993), but the fact that both methods assay lysozymal cleavage of the same bond suggests that our findings for the method of Gonzalez et al. (1993) apply to the method of Vrba et al. (1993) as well.

We began this study with the intent of testing the protistan-specificity of L_{ACID} analysis. We wished to know if the method would discriminate protistan from metazoan bacterivory in complex communities. Alternatively, we thought that a general measure of community bacterivory might still prove useful. In the process of explaining our results, we more closely examined the chemistry underlying the Gonzalez et al. (1993) and Vrba et al. (1993) methods.

2.2 MATERIALS AND METHODS

Organisms were obtained from three small ponds on the Merritt Island National Wildlife Refuge, Florida and from Carolina Biological Supply Co. The ponds varied in salinity (<1, 7, and 16 ppt). A pond sample was collected by scooping or pushing submerged aquatic vegetation into a 1 L plastic bottle. The bottle was filled with pond water and placed on ice until return to the laboratory for processing. A second bottle was also filled with pond water and chilled to ensure sufficient water for maintaining animals in the lab.

We expected *a priori* that several genera would not have L_{ACID}, based on basic information about feeding: *Stentor* is predacious on other protists (Pratt and Cairns 1985); *Chilomonas* (a cryptomonad) is not phagotrophic (Lee et al. 1985); tardigrades have stylet mouthparts and feed by piercing plants or small metazoans (Pennak 1989); and cyclopoid copepods are raptorial feeders (Pennak 1989).

Samples were held at room temperature during processing. Each field sample jar was vigorously shaken to dislodge organisms attached to plants. Aliquots of field-collected and cultured samples were then examined in a Petri dish with a dissecting scope. Animals were individually isolated by micropipetting them in a series of filter-sterilized (0.2 μm) pond or culture water. Three to four sequential transfers were used to obtain an isolate for acid-lysozyme analysis.

Isolated organisms were sonicated and analyzed according to Gonzalez et al. (1993). We measured fluorescence at 360 nm with a Perkin-Elmer LS 50 B luminescence spectrometer (excitation at 330-400 nm, emission filter = 470 nm. Readings from boiled controls (1 per extract) were subtracted from results to correct for background fluorescence. Results were expressed as µM MUF / mL extract. Expression as bacteria mL^{-1} h^{-1} would require calibration with ingested fluorescently-labelled bacteria (Gonzalez et al. 1993); an unnecessary step in a simple test of the presence of acid lysozyme among taxa. In addition, we did not standardize extracted biomass among taxa, which would be expected to affect magnitude of results. We focused at this point on the presence or absence of the enzyme in diverse taxa, not the magnitude of results.

2.3 RESULTS

Results are listed in Table 1: taxonomy is per Pennak (1989). Values with confidence intervals overlapping a value of 0 (no activity) were considered to indicate a negative result. Genera listed twice were analyzed on separate occasions.

Of the 6 protistan genera tested, only *Stentor* unequivocally lacked L_{ACID} (Table 1): a response we predicted because *Stentor* is a predator of other protists. Results for *Actinosphaerium* were low, and in one case variable enough to be considered a negative result by our criterion. The latter is probably a result of low density of isolated organisms (26 individuals isolated in 6 mL) in the analyzed extract. Two values are listed for *Paramecium*: the first for organisms isolated in original culture solution, and the second from organisms isolated in an optically-clear culture solution. Because we expected *Paramecium* to have L_{ACID}, we were surprised to have a negative result for organisms in the original culture solution. Analysis of organisms in clear solution indicated that *Paramecium* do in fact have L_{ACID}, and that dissolved organic compounds (e.g., humic and tannic acids) apparently interfered with fluorescence detection at 360 nm.

We analyzed 10 micrometazoan taxa (Platyhelminthes, Rotifera, Nematoda, and Tardigrada). Nine had L_{ACID}, including organisms we thought would not have L_{ACID} (tardigrades and cyclopoid copepods). *Brachionus* was marginally positive, probably due to low density of extracted individuals. Another rotifer,

Lepadella, had a greater mean value than *Brachionus*, but subsamples varied in activity enough to be considered negative. *Stenostomum* was also low in density, which may explain the mixed results among extracts. One unidentified nematode extract was marginal, while the other sample was clearly positive, again probably related to density effects on detection limit.

All 12 of the macrometazoan genera (Annelida and Crustacea) were clearly positive (Table 1). Greater biomass for some of these genera was probably reponsible for greater mean values. Surprisingly, cyclopoid copepods clearly had L_{ACID}.

In summary, 5 of 6 protistan genera tested had L_{ACID}: the one lacking L_{ACID} (*Stentor*) was predicted. However, one other protistan genus, *Chilomonas*, was predicted to lack L_{ACID} but had it. Of the 21 metazoan taxa tested, 20 (95%) had L_{ACID}. The one genus (*Lepadella*) with activities not significantly different from 0 may actually have L_{ACID}, but biomass may have been too low to get a clear reading.

2.4 DISCUSSION

Acid lysozyme analysis is not protistan-specific. This conclusion stands even if we discount the genera that had marginal or low activities. All five of the metazoan phyla tested were positive for L_{ACID}. Clearly, L_{ACID} analysis of field samples can overestimate protistan bacterivory if metazoans are in the extracted sample, especially if whole water samples (e.g., Gonzalez et al. 1993) or large-mesh plankton netting are used. Likens and Gilbert (1970) found that 35 µm mesh is needed to quantitatively sample rotifers (i.e., remove rotifers from filtrate). Use of mesh > 35 µm to exclude metazoans (e.g., Vrba et al. 1996) will permit small metazoans to pass into samples, potentially leading to overestimates of protistan bacterivory. Although it may exclude some large protists, use of 20-µm (e.g., Sherr et al. 1992) or 35-µm mesh will more clearly isolate bacterivorous protists from metazoans.

Acid lysozyme analysis is not bacterivore-specific. Gonzalez et al. (1993) correctly stated that peptidoglycan only occurs in eubacteria cell walls, but lysozymes hydrolyze β(1-4) glycosidic bonds of both peptidoglycan and chitin (Stryer 1981). Any samples that contain organisms with chitinolytic activity will overestimate "protistan" bacterivory when analyzed for L_{ACID} by the method of Gonzalez et al. (1993). In addition, it is possible that some protistan predators use lysozymes to degrade chitin of captured prey (e.g., rotifers, nematodes, gastrotrichs, etc.) and peptidoglycan of bacterial cell walls. Thus, L_{ACID} cannot be expected to discern between protistan and metazoan bacterivory, let alone bacterivory and chitinolysis in general.

Although we did not test the method of Vrba et al. (1993), we must infer the same conclusion for their method. The methods of Gonzalez et al. (1993) and Vrba et al. (1993) both analyze enzymatic hydrolysis of the β(1-4) glycosidic bond attaching MUF to NAG: Gonzalez et al. (1993) uses NAG trimers (chitotriose), while Vrba et al. (1993) uses NAG monomers. The methods differ (e.g., sonication of cells versus whole-cell analysis for extracellular enzyme activity), but the basic lysozyme-glycosidic bond reaction is central to both methods. Vrba et al. (1993) considered interference by bacterial chitinolytic enzymes (lysozyme), but did not rule out invertebrates as sources of chitinolytic lysozyme. Given that chitin-bearing invertebrates that molt may release lysozyme into the water, the method of Vrba et al. (1993) may be subject to the dual problems of non-specificity for Protista and bacterivory. In addition, it is possible that some protistan predators use lysozymes to digest chitin-bearing prey, which would further overestimate protistan bacterivory by lysozyme analysis.

The fact that "acid lysozyme" has maximal activity at pH 4.5 is not by itself indicative of a protistan-specific assay. Multiple metazoan phyla exhibited lysozyme activity at that pH in our analysis. Gonzalez et al. (1993) did not compare protistan and invertebrate metazoan lysozyme pH optima, but only compared protistan and bacterial lysozyme pH ranges.

Our results indicate two additional problems with the Gonzalez et al. (1993) method. First, the color interference with *Paramecium* culture solution indicates that natural waters colored by dissolved organic matter may be difficult to analyze by this method. Secondly, *Chilomonas* had L_{ACID}, although *Chilomonas* is a cryptomonad that does not feed phagotrophically (Lee et al. 1985). This result is consistent with

extracellular release of lysozyme, as measured by Vrba et al. (1993). Is it possible that intracellular L_{ACID} and extracellular β-N-acetylglucosaminidase are two versions of the same lysozyme, operating at different pH optima? Vrba et al. (1996) observed both high- and activity enzyme kinetics, and offered extracellular release of vacuolar enzymes as a possible explanation for low-activity enzymes.

Enzymatic assays of bacterivory offer the potential to mitigate errors introduced by other methods (Gonzalez et al. 1993), but only if enzymatic assays themselves are less likely to introduce greater error. Our results indicate that protistan bacterivory could be grossly overestimated by lysozyme analyses (e.g., Gonzalez et al. 1993, Vrba et al. 1993) of samples more complex than protistan isolates. In addition, potential chitinolysis by predatory protistans may introduce error in protistan isolates. At this point, the tradeoffs involved in deciding on a method to analyze bacterivory in aquatic habitats do not point to abandonment of other approaches in favor of lysozyme analyses.

Table 1. Acid lysozyme analysis results among invertebrate taxa. Results include organisms isolated from both field and cultured samples. Each mean represents two analyzed extracts from one sample, with 95% confidence intervals (CI). See text for further descriptions.

	uM MUF / mL	
	Mean	95% CI
Phylum Sarcomastigophora		
Subphylum Sarcodina, Class Actinopoda		
Actinosphaerium	0.063	0.080
Actinosphaerium	0.051	0.005
Subphylum Sarcodina, Class Rhizopoda		
Amoeba proteus	0.313	0.007
Amoeba proteus	0.262	0.148
Subphylum Mastigophora, Class Phytomastigophorea		
Peranema	0.493	0.002
Chilomonas	1.974	0.227
Phylum Ciliophora		
Paramecium (brown soln)	-0.037	0.049
Paramecium (clear soln)	8.509	0.355
Stentor	-0.005	0.002
Stentor	0.001	0.001
Phylum Platyhelminthes, Class Turbellaria		
Microdalyiella	0.299	0.031
Stenostomum	0.010	0.003
Stenostomum	-0.018	0.007
Phylum Rotifera		
Brachionus	0.007	0.005
Lecane	0.020	0.002
Lepadella	0.024	0.030
Philodina	0.783	0.082
Rotaria	0.018	0.007
Phylum Nematoda		
Cephalobus	2.489	1.495
Unidentified nematode	0.008	0.005
Unidentified nematode	0.027	0.002
Phylum Tardigrada		
Unidentified tardigrade	0.472	0.252
Phylum Annelida, Class Oligochaeta		
Stylaria	17.588	0.074
Tubifex	15.583	1.197
Unidentified Naidid	12.136	1.369
Phylum Arthropoda, Class Crustacea		
Subclass Branchiopoda		
Artemia	15.576	0.782
Daphnia	15.770	0.111
Scapholeberis	1.110	0.126
Subclass Ostracoda		
Unidentified Ostracod	4.721	0.082
Unidentified Ostracod	2.025	0.140
Subclass Copepoda		
Cyclopoid Copepod	5.757	0.313
Harpacticoid Copepod	0.013	0.004
Subclass Malacostraca		
Asellus	16.911	1.512
Gammarus	18.506	3.023

3. REFERENCES

Azam F, Fenchel T, Field JG, Gray JS, Meyer-Reil LA, Thingstad F (1983) The ecological role of water-column microbes in the sea. Mar Ecol Prog Ser 10:257-263

Berninger UG, Findlay BJ, Kuupo-Leinikki P (1991) Protozoan control of bacterial abundances in fresh water. Limnol Oceanogr 36:139-147

Cabezas JA (1989) Some comments on the type references of the official nomenclature (IUB) for β-N-acetylglucosaminidase, β-N-acetylhexosaminidase, and β-N-acetylgalactosaminidase. Biochem J 261:1059-1060.

Cole JJ, Findlay S, Pace ML (1988) Bacterial production in fresh and saltwater ecosystems: a cross-system overview. Mar Ecol Prog Ser 43:1-10

Crawley MJ (1987) What makes a community invasible? In: Gray AJ, Crawley MJ, and Edwards PJ (eds). Colonization, succession and stability. 26th Symposium of the British Ecological Society. Blackwell Sci Publ, Oxford pp 429-453

Elton CS (1958) The ecology of invasions by animal and plants. Chapman and Hall, London

Gonzalez JM, Sherr BF, Sherr EB (1993) Digestive enzyme activity as a quantitative measure of protistan grazing: the acid lysozyme assay for bacterivory. Mar Ecol Prog Ser 100:197-206.

Gooday GW (1990) The ecology of chitin degradation. In: Marshall KC (ed) Advances in microbial ecology, vol. 11. Plenum Press, New York, pp 387-430.

Jenkins DG (1995) Invertebrate diverstiy and density in hydroponic systems of the closed Biomass Production Chamber. Report to Dynamac Corp, CELSS Project, Kennedy Space Center, FL 13 pgs

Jurgens K, Arndt H, and Rothhaupt KO (1994) Zooplankton-mediated changes in bacterial community structure. Microb Ecol 27:27-42.

Lee JJ, Hutner SH, Bovee EC (1985) Illustrated guied to the protozoa. Soc. of Protozoologists, Lawrence KS.

Likens GE and Gilbert JJ (1970) Notes on the quantitative sampling of natural populations of planktonic rotifers. Limnol Oceanogr 15:816-820.

Lodge DM (1993) *Biological invasions: lessons for ecology.* Trends in Ecol Evol 8:133-137

MacArthur RH, Wilson EO (1967) The theory of island biogeography. Princeton University Press, Princeton, NJ

May RM (1973) Stability and complexity in model ecosystems. Princeton University Press, Princeton, NJ

May RM (1974) On the theory of niche overlap. Theor Pop Biol 5:297-332

Morales A (1995) Survival of potentially pathogenic human-associated bacteria in the rhizosphere of hydroponically-grown wheat. MS Thesis, University of South Florida, 64 pgs

Nygaard K and Hessen DO (1990) Use of ^{14}C-protien-labelled bacteria fro estimating clearance rates by heterotrophic and mixotrophic flagellates. Mar Ecol Prog Ser 68:7-14.

Pennak RW (1989) Freshwater invertebrates of the United States, 3rd ed. John Wiley & Sons, New York.

Pomeroy LR, Wiebe J (1988) Energetics of microbial food webs. Hydrobiologia 159:7-18.

Porter KG, Sherr EB, Sherr BF, Pace ML, Sanders RW (1985) Protozoa in planktonic food webs. J Protozool 32:409-415.

Pratt JR, Cairns JJ, Jr. (1985) Functional feeding groups of Protista. Trans. Am. Microsc. Soc.

Ricklefs RE (1987) Community diversity: relative roles of local and regional processes. Science 235:167-171

Robinson JV, Dickerson JE, Jr. (1984) Testing the invulnerability of laboratory island communities to invasion Oecologia 61:169-174

SAS/STAT User's Guide, version 6, 4th edition, volume 2. SAS Institute, Cary, NC.

Sherr BF, Sherr EB, Fallon RD (1987) Use of a monodispersed, fluorescently labelled bacteria to estimate in situ protozoan bacterivory. Appl Environ Microbiol 53:958-965.

Sherr BF, Sherr EB, McDaniel J (1992) Effect of protistan grazing on the frequency of dividing cells in bacterioplankton assemblages. Appl Environ Micro 58:2381-2385.

Sleigh M (1989) Protozoa and other protists. Edward Arnold, New York.

Strayer RF (1994) Dynamics of microorganism populations in recirculating nutrient solutions. Adv Space Res 14:(11)357-(11)366

Stryer L (1981) Biochemistry, 2nd ed. WH Freeman and Co., San Francisco.

Vermeij GJ (1991) Science 253:1099-1104. (As cited in Lodge 1993).

Vrba J, Simek K, Nedoma J, Hartman P (1993) 4-methylumberiferyl-β-N-acetylglucosaminidase hydrolysis by a high-affinity enzyme, a putative marker of protozoan bacterivory. Appl Environ Micro 59:3091-3101.

Vrba J, Simek K, Pernthaler J, Psenner R (1996) Evaluation of extracellular, high-affinity β-N-acetylglucosaminidase measurements from freshwater lakes: an enzyme assay to estimate protistan grazing on bacteria and picocynobacteria. Microb Ecol 32:81-99.

Wheeler RM, Mackowiak CL, Stutte GW, Sager JC, Yorio NC, Ruffe LM, Fortson RE, Dreschel TW, Knott WM, Corey KA (1996) NASA's Biomass Production Chamber: a testbed for bioregenerative life support studies. Adv Space Res 18:(4/5215-(4/5)224

Williamson M (1989) In: Drake JA et al. (eds) Biological invasions: a global perspective (SCOPE 37). Wiley, New York pgs 329-350 (As cited in Lodge 1993)

1996 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

SOFTWARE RELIABILITY ISSUES CONCERNING LARGE AND SAFETY CRITICAL SOFTWARE SYSTEMS

A Literature Survey of Current Techniques

Dr. Khaled Kamel, Professor and Chairman
Engineering Computer Science Department
University of Louisville
Louisville, Kentucky

KSC Colleague - Barbara Brown
Computer Science

Contract Number NASA-NGT10-52605

July 19, 1996

Software Reliability Issues Concerning Large and Safety Critical Software Systems

Khaled Kamel and Lawrence Lowe, University of Louisville

ABSTRACT

This research was undertaken to provide the National Aeronautics and Space Administration (NASA) with a survey of state-of-the-art techniques being used in industry and academia to provide safe, reliable, and maintainable software to drive large systems. Such systems need match the complexity and strict safety requirements of NASA's shuttle system. In particular, the Launch Processing System (LPS) is being considered for replacement. The LPS is responsible for monitoring and commanding the shuttle during test, repair, and launch phases. NASA built this system in the 1970's using mostly hardware techniques to provide for increased reliability, but it did so often using custom-built equipment, which has not been able to keep up with current day technologies. This report surveys the major techniques used in industry and academia to ensure reliability in large and critical computer systems.

1. Introduction

After many years of expensive and highly customized maintenance of the Shuttle's Launch Processing System, NASA has decided that the time to redevelop the LPS has come. A few factions exist within the NASA hierarchy, each with slightly different but complementary ideas on how to approach such a giant step. All the involved persons, however, share a mutual goal: to implement the new Launch Processing System taking advantage of today's well-established standards and commercial-off-the-shelf (COTS) equipment. Standards and commercial equipment are necessary to keep operation and development costs minimized. Without them, the customized path becomes increasingly more expensive with time, and obsolescence is quick to come. An added advantage is that most commercial vendors implement the latest technology in their components to provide high speed and high reliability during operation with backwards-compatible upgrades. Since commercial equipment has become widely used in a number of environments, our confidence in its ability to perform correctly is strengthened.

Regardless of how safe and reliable hardware is, it is usually software that drives it, and software is expected to play an even larger role at NASA with upcoming efforts such as Space Station and future vehicles. It is well known that the potential for software to fail is significantly higher than that of hardware. Many will attribute this software

characteristic to the high number of states and transitions that a software system can reach. Thoroughly testing all such states and paths is often not a feasible task (Kim 95). The task is complicated even further with the possibility of unforeseen states. As a result, software practices have focused on detecting, tolerating, and recovering from failures related to software rather than preventing them. The key mechanism for tolerating faults is *redundancy,* and it is the primary focus of those techniques presented within this paper. Since the fault handling of both hardware and software within a system are closely related, the architectures and methods presented here often address these reliability concerns from both sides.

2. Terminology of Reliability

Software controls an increasingly large number of systems in operation throughout the world today. Many of these systems either directly or indirectly affect people, and that is why there can be no practical allowance of error. Ideally, software would be designed and implemented without any inherent flaws, but Pham suggests that "we cannot reasonably expect a large system to be error free" (Pham 92). Any person having experience in the development of a large software system knows that exhaustive testing is a difficult undertaking. Despite any amount of effort, there is a point at which further testing produces minimal improvement with exponential increases in time and cost. Thus, the "final release" of any software system is expected to require future maintenance and the uncovering of flaws missed in the development stage.

Since software cannot practically be completely error-free, there is a factor of *reliability* associated with a system. Gersting defines reliability as "the capability of a system to perform over a period of time in adherence to its specifications" This definition places a strong connection between reliability and requirements. Later sections of this paper show that imperfect requirements, and not software faults, are the cause of many system failures.

Any conceptual error in the way that a unit of software handles a state or set of inputs is referred to as a *fault.* If the state or set of inputs occurs for which the software unit is faulty, the fault will result in an *error.* This error is an internal state that does not relate to the set of allowable states. An error, then, has a potential to cause a system *failure.* The failure is an observable deviation from the specifications. This terminology is important in that software can possess faults that never manifest themselves as errors, and errors that do not trigger failures. Furthermore, we can have software that possesses a number of faults, but is still considered reliable since the faults may not result in failures.

Gersting suggests that knowledge of the *operational profile* of the software unit in question can be used to help estimate the unit's reliability (Gersting 92). The operational profile gives an indication of how often different sub-components within a unit are executed. If a sub-component sees a large amount of traffic, then any faults in that sub-component are more likely to generate failures. It follows then that a fault in a low-traffic sub-component has a lesser impact on the reliability of the unit.

Software faults can be further classified according to *Independence* and *Persistence* (Laprie 92). Under the Independence class, a fault is either *related* or *independent*. Related faults often result from a general misunderstanding of the specifications or from a specification fault. When one fault is encountered, its related faults have a high chance of creating similar errors. This set of similar errors can lead to *common-mode failures*, the set of failures attributable to the same fault. Faults that are independent, however, have no relation and usually do not produce failures under common conditions.

The Persistence class separates faults as either being *solid* or *soft*. A soft fault is a fault synonymous with an intermittent error when discussing hardware faults. That is, once having occurred, it is not likely that the fault conditions will be reproduced. Furthermore, a fault is soft only if it is recoverable. A solid fault is just the complement of the soft fault. It either continues to occur under normal operations or is unrecoverable. This classification of independence and persistence will play a large role when discussing fault-tolerant architectures later in this paper.

3. Redundancy Techniques

Assuming that all software will contain faults, the developer is left with the task of finding a way for software to operate reliably in the presence of such faults. Pham suggests that, "to achieve ultrareliability in computing, it is necessary to adopt the strategy of defensive programming based on redundancy" (Pham 92). Redundancy suggests that a collection of programs, referred to as versions, performing the same function be used in place of a single program. In theory, any faults will be masked by the collection of programs (i.e., the collection will often succeed as a whole where a single program might fail). This technique, allowing faults to exist while still operating reliably, is referred to as *fault-tolerance*. Indeed, fault-tolerance has become the key focus in industry and academia in providing for reliable software systems. The two techniques discussed below are the foremost in implementation and are the main focus of this paper. There are issues, however, that foil the theory. The theory of fault-tolerance

depends on faults being independent so that the majority of the collection's versions agree on the correct output. If faults are related, a common-mode failure might cause the majority of the collection to rule in favor of an incorrect output or produce no majority at all. Experiments have shown that, even in the presence of related faults, these redundant systems still give some improvement over their nonredundant counterparts. The results that they give make it clear that redundancy is worth further exploration.

3.1. Recovery Block Method

Using the recovery block approach, as well as other methods, the designer wishes for a system to perform a task or calculation with a high degree of confidence in its results. Under the recovery block method, a system is comprised of three components: (1) a primary module responsible for executing critical software functions and the desired task or calculation, (2) an acceptance test by which the primary module's output can be verified, and (3) a set of alternate modules, each capable of performing the same function as the primary module (Pham 92). According to Stark and Dugan, when the recovery block is entered, a checkpoint is created so that the system's state can be restored. The primary module is then executed to perform the desired task, and the results are verified by the acceptance test. If the acceptance test detects erroneous output, the system is rolled back to the checkpoint, and an alternate module is used to perform the operation. Alternates will continue to be tried by the recovery block until either a module passes the acceptance test or all alternates have been tried. A system failure occurs only in the latter case.

The recovery block concept is very similar to the hardware redundancy technique known as *standby sparing*. The

Recovery Block Method

```
Ensure T
      By M1
      Else by M2
        Else by M3
            •
            •
            •
          Else by Mn
Else Error
```

Where T denotes the acceptance test, M_1 denotes the primary module, and M_k, $2 \leq k \leq n$, denotes the alternates.

gain in reliability with software redundancy, however, is not as large as with redundant hardware (Stark 94). In addition to the problems of correlated faults already discussed, the redundancy management itself introduces opportunities for further modes of failure.

Gersting notes that a key advantage to the recovery block strategy is that, if the primary module is determined to be successful, no additional resources or time are required to process the alternate modules (Gersting 92). In a sense, the system has performed its function with a near-minimal amount of work. A well designed and implemented system should produce correct results far more often than incorrect ones, so it follows that a recovery block should not seriously degrade average system performance.

A disadvantage to the recovery block method is also observed by Gersting. He states that the overhead in execution time required to perform a system rollback after a failed version can adversely affect system performance in the event of a failure . This is especially true in real-time environments.

The acceptance test is of special concern if the recovery block method is to provide reliable output. As stated earlier, a module will be considered successful and be used to provide the system's output if the module's output passes the acceptance test. This places a heavy responsibility on the acceptance test to accurately decide when an output is or is not a reasonable output. Because knowledge of the exact result expected does not exist, the acceptance test is restricted to items such as limit-checking and past knowledge when making its decisions. If the acceptance test is not well designed, it may pass results that are erroneous. Although this is not flagged as a failure by the system, it is clearly a deviation from the desired mode of operation. At the other extreme, a faulty acceptance test might reject a valid output.

It can be concluded then that a few things are necessary for the recovery block method to be a viable improvement on a system. First, it is desirable that the primary and alternate modules contain as few faults as possible. At a minimum, we would like the existing faults to be independent among versions. Second, the acceptance test must accurately decide the correctness of a module's output. If we are to tolerate faults, we must first detect them correctly. Regardless of the difficulty in meeting these requirements in practice, this technique has been demonstrated to give some improvement in reliability.

3.2. N-Version Programming Method

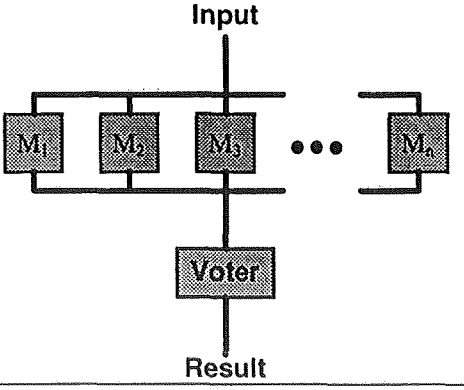
Under N-version programming, N independent programs are executed in parallel on identical input. The system determines results by voting on the outputs provided by the set of programs (Pham 92). With basic N-version programming, this vote is often just the majority agreement.

It is necessary that the N versions involved be as independent as possible. By independent it is meant that the faults contained within one version be independent of those contained within another. One cannot purposely select the faults that will exist within a version, so it is suggested that independence might be attained through diverse development environments. Each version should be implemented using different algorithms, programming languages, staff, and tools . The less in common between versions, the higher the chance of independence. It is important, however, that all versions use the exact same set of requirements and that the requirements be as clear and complete as possible.

The N-version programming method is very similar to the hardware redundancy technique known as N-modular redundancy. As with recovery blocks, the redundant software technique does not provide as significant of a gain in reliability as achieved with hardware. This is true mostly because the N versions often cannot be as thoroughly tested as hardware units. Hardware provides a fixed mapping from a set of inputs to a set of outputs, while software tends to take unanticipated turns during execution. This is not a flaw in software, but in the human's inability to foresee all possibilities.

A strong advantage of N-version programming is that when a version fails, no additional time is required to rollback

N-Version Programming Method



Where M_k, $1 \leq k \leq n$, denotes the set of versions to be executed in parallel. The voter may use any voting logic.

the system and reperform the computation as with recovery blocks . The results will be made available as soon as the outputs from all of the versions have been voted on. This means that the same amount of time is required to "reliably" perform a calculation whether a version fails or not.

A disadvantage is that "resources and overhead are required to execute, synchronize, and vote upon multiple versions for every computation" (Gersting 92). This means that even when all version successfully perform the desired computation, the overhead required to execute and decide each version is required. This makes basic N-version programming a poor choice for systems that are either resource-limited or time-critical.

Another disadvantage is that straight-forward comparison of version outputs is often not possible when voting . This is true in cases when outputs are floating-point types, where correct answers are expected to fall within a tolerance range. This complicates the voting logic. In fact, some studies have shown that more reliable results may be achieved by voting strategies other than a direct majority rule.

3.2.1. Voting Techniques

An important situation arises when basic N-version programming provides a set of outputs on which no correct majority exists (e.g., either N different outputs occur or the majority of the versions output an identical incorrect result). Under basic N-version programming, the system has no choice but to produce a system failure. A study done by Gersting compares different voting strategies by which the probability of success in such a situation is improved.

The computation performed in Gersting's study outputs a one-dimensional vector. Three redundant versions of the software were used, each being numbered according to decreasing confidence in reliability (i.e., the most reliable version is labeled 1, the next most reliable is labeled 2, and so on). Each of these versions produces an output vector for each iteration. The elements in a vector are fields of data that have some meaning to the recipient of the vector. The need for a composite result such as this may not always be the case. Some of the voting algorithms presented here may not apply in those cases.

3.2.2. Composite/Version

With this algorithm, a counter is maintained for each of the versions. Each field in the vector is compared among the three versions. The majority answer for a field is taken as the correct answer for that field. Those versions whose vectors agree with the chosen field result have their counters incremented. In the case where no majority exists for a

field, no version is allowed to increment its counter. The overall result vector is the vector belonging to the version with the highest counter value after all fields have been voted on. In other words, a "correct" value for each field is decided using the majority for that field. The winner is the vector that most often agrees with the correct field values. Note that the result vector is always taken exactly from one of the version outputs. Had the result vector been built from each field majority, inconsistencies between fields may have been introduced.

3.2.3. Weighted Composite/Vector

This algorithm is identical to the Composite/Version algorithm except that a version's counter is not always incremented by one when a field within its output vector agrees with the majority. Instead, weights are assigned to different fields based on importance. The weight is the value added to a version's counter on successful field comparisons. Gersting investigated assigning weights based on how often a field changes. If a field's value does not change often, it received a lower weight since a version probably had a higher chance of getting it right. The end result was that a version was chosen that most often got the critical fields correct. Of course, this requires that knowledge of the application be known in order to assign meaningful weights.

3.2.4. Composite

The Composite algorithm does not require a counter for each version. Instead, the majority of each field is taken as the correct result for that field in the result vector. When a field has no majority, that field in the result vector will be filled by the field value from Version 1. This means that the result vector may very likely not match any of the version outputs. This method increases the chance of internal inconsistencies that the first two algorithms were trying to avoid.

3.2.5. History

For this algorithm, the voter maintains a count for each version that represents the number of times that a version agreed with the majority during past computations. For a computation, the algorithm selects the result vector from the version that both agrees with the majority and has the highest count. If no majority exists, the agreeing version with the highest confidence is chosen.

3.2.6. Acceptance Test

This algorithm first attempts to produce a result vector by taking a majority vote. If such a vote fails to produce a majority, the algorithm then submits each version's output vector to an acceptance test in order of decreasing

reliability. The result in such a case is the first vector to pass the acceptance test. As with the Weighted Composite/Vector algorithm, this algorithm requires knowledge of the application in order to define the acceptance test criterion.

Gersting made a few noteworthy conclusions about his experiment. First, he found each algorithm to reduce the failure rate by roughly 50%. There was little difference in the actual improvement between versions. He also found that success was a bit higher when only allowing the result vectors to be one of the versions' outputs (i.e., do not create a composite vector in a piecewise manner from the field majorities). However, there were times when such a technique provided correct results in the presence of three failed versions. The History algorithm proved to be least successful of all. The most successful algorithm was the Acceptance Test. Gersting ends with a useful rule of thumb: "if an output must be forced in the no-majority case, do almost anything reasonable - preferably simple - and you'll often be right" .

Conclusion

Large and safety critical computer systems similar to NASA LPS are required to have reliable hardware. Software that derive the hardware is continuing to grow and play major role in the system operation. This report detailed the most commonly used techniques in implementing reliability and safety requirements in software systems. The replacement of the LPS should carefully consider the overall system reliability including hardware and software during all phases of system operation, system analysis, system evaluation, final implementation, and throughout the system life cycle.

Works Cited

Babcock IV, Philip S. "The Application of Analytic Tools to the Design and Verification of Robust Systems." 1994 Proceedings Annual Reliability and Maintainability Symposium 1994.

- Dugan, Joanne Bechta, and Stacey A. Doyle. "Simple Models of Hardware and Software Fault Tolerance." 1994 Proceedings Annual Reliability and Maintainability Symposium 1994: 124-29.

Dunham, Janet R., and John L. Pierce. An Experiment In Software Reliability. NASA Contractor Report 172553. Research Triangle Park: Software Research and Development Center for Digital Systems Research, 1986.

Eagle, Kenneth H., and Ajay S. Agarwala. "Redundancy Design Philosophy for Catastrophic Loss Protection." 1992 Proceedings Annual Reliability and Maintainability Symposium 1992: 1-4.

Eckhardt Jr., Dave E., et al. An Experimental Evaluation of Software Redundancy As a Strategy for Improving Reliability. NASA Technical Memorandum 102613. Hampton: Langley Research Center, 1990.

France. North Atlantic Treaty Organization. Advisory Group for Aerospace Research and Development. Fault Tolerance Design and Redundancy Management Techniques. Lecture Series No. 109. Neuilly Sur Seine: 1980.

Gersting, Judith L., et al. "A Comparison of Voting Algorithms for N-Version Programming." In Fault-Tolerant Software Systems: Techniques and Applications. Proceedings 24th Annual Hawaii International Conference on Systems Sciences. January 1991. Los Almitos: IEEE Computer Society Press, 1992.

Kim, K. H., and Howard O. Welch. "Distributed Execution of Recovery Blocks: An Approach for Uniform Treatment of Hardware and Software Faults in Real-Time Applications." In Fault-Tolerant Software Systems: Techniques and Applications. IEEE Transactions on Computers. Vol. 38 Number 5. Los Almitos: IEEE Computer Society Press, 1992.

Lala, Jaynarayan, et al. Study of a Unified Hardware and Software Fault Tolerant Architecture. NASA Contractor Report 181759. Cambridge: The Charles Stark Draper Laboratory, Inc., 1989.

Laprie, Jean-Claude., et al. "Definition and Analysis of Hardware- and Software-Fault-Tolerant Architectures." In Fault-Tolerant Software Systems: Techniques and Applications. IEEE Computer. Vol. 23 Number 7. Los Almitos: IEEE Computer Society Press, 1992.

National Aeronautics and Space Administration. Launch Processing System Description. Cape Canaveral: John F. Kennedy Space Center, 1984.

Pham, Hoang. Fault-Tolerant Software Systems: Techniques and Applications. Los Almitos: IEEE Computer Society Press, 1992.

Stark, George E. "Technologies For Improving the Dependability of Software-Intensive Systems: Review of NASA Experience." 1994 Proceedings Annual Reliability and Maintainability Symposium 1994.

Vouk, Mladen A., Amitkumar M. Paradkar, and David F. McAllister. "Modeling Execution Time of Multi-Stage N-Version Fault-Tolerant Software." In Fault-Tolerant Software Systems: Techniques and Applications. Proceedings IEEE Computer Software and Applications Conference. 1990. Los Almitos: IEEE Computer Society Press, 1992.

1996 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

PROJECT MANAGEMENT FRAMEWORK TO ORGANIZATIONAL TRANSITIONS

Dr. Tim Kotnour, Assistant Professor
Industrial Engineering and Management Systems Department
University of Central Florida
Orlando, Florida

KSC Colleague - Saul Barton
Human Resources

Contract Number NASA-NGT10-52605

August 2, 1996

ABSTRACT

The contribution of this paper is a description of a project management framework and associated models for organizational transitions. The framework contains an integrated set of steps an organization can take to lead an organizational transition such as downsizing and change in mission or role. The framework is designed to help an organization do the right work the right way with the right people at the right time. The underlying rationale for the steps in the framework is based a set of findings which include: defining a transition as containing both near-term and long-term actions, designing actions which respond to drivers and achieve desired results, aligning the organization with the external environment, and aligning the internal components of the organization. The framework was developed based on best practices found in the literature, lessons learned from heads of organizations who have completed large-scale organizational changes, and concerns from employees at the Kennedy Space Center (KSC). The framework is described using KSC.

Managers can use this framework to help design their transitional activities. Change leaders can use the methodology and framework to develop a framework specific to their organization. Researchers can use this framework to further research the effect of different change activities on transition and organizational performance.

PROJECT MANAGEMENT FRAMEWORK TO ORGANIZATIONAL TRANSITIONS

Tim Kotnour, Ph.D.

1. INTRODUCTION—THE NATURE OF KSC'S TRANSITION

The Kennedy Space Center (KSC) as part of the National Aeronautics and Space Association's (NASA) efforts to perform "better, faster, cheaper" have begun a large-scale organizational change effort. Their effort is similar to the other private, public, and government organizations who have attempted efforts such as downsizing to meet performance requirements. The drivers for KSC's transition include: the reduction in NASA's budget, the development a single flight operations contractor (SFOC) for shuttle processing, and a definition of KSC's roles based on NASA's plans (e.g., strategic, HEDS, and enterprise). These drivers have led to the change in role of and possible reduction of the civil servants at KSC. The change in the civil servants' roles has been described as moving from oversight to insight. These circumstances lead KSC to undergo a large-scale organizational change.

Best practices can help guide an organization to complete successful large-scale change. Other studies have been completed which summarize the set of "best practices" for organizational change and downsizing [5, 6, 11, 23, 28, 34, 35]. However, the translation of the findings into an all encompassing working plan needs to be completed. Some authors [1, 20] do offer a set of explicit steps. A project management approach is adopted to integrate the disparate set of research findings and case studies into a useful tool for managers to consider when leading a large-scale organizational change. The project management approach emphasizes the need to explicitly define a set of steps and the linkage among those steps with goals (e.g., cost, schedule, performance). The purpose of this research is to help mangers understand, discuss, and take action on their organizational change and transition. To accomplish this purpose we seek to assimilate the best practices and the theory behind the practices into a framework. We use four related models to describe a set of implications to consider in conducting an organizational transition. The four models are:

1. Doing the right work the right way with the right people at the right time.
2. An organizational transition is an "All-term" transition.
3. Transition drivers lead to actions which lead to results.
4. An organization must be aligned both externally and internally.

From these four models and associated implications, a set of steps are defined to lead an organization through its transition and are represented as a work breakdown structure. In this paper, I explain the models, provide the rationale for them, and highlight a set of steps to implement the implications of the models. The objective of the research was to develop a "Transition Framework" including a plan which KSC management can use to answer: What is the integrated and sequenced set of activities we need to perform over the next four-plus years to ensure a "successful" transition to a new KSC state?

We propose a model for organizational change, shown in Figure 1, which describes the fundamental problem that KSC is experiencing when managing a large-scale transition or change:

Doing the right work the right way with the right people at the right time.

We use this model to connect the other three models together. Addressing the problem of the right work is a process of flowing requirements down from the mission/vision set forth by NASA HQ and senior management of KSC. The requirements are based on the current, transition, and future states of the organization. The right work is a function of items such as mission, agency requirements, meaningful work, and products and services the customer desires. The right way includes the processes, structures,

and tools by which the work is completed and managed. The right people is derived from understanding how people are organized to complete the work and the tools they use. The right people encompasses issues such as skill needs from a process and structure perspective, number aligned with budget, and training. The right time is the proper timing of doing the work to meet near-term work requirements, near-term transition actions to ensure long-term success, and long-term work requirements. The box in the middle of the "triangle" shows a basic approach to addressing the "right" set of issues and is the foundation for the framework. An organization must decide on the approach to align these components (i.e., work, processes, structures, tools and people) [13]. In summary the implications of Figure 1 include:

- KSC is transitioning from it's current to future state to ensure the right work is completed;
- the requirements for KSC flow down from NASA's definition of KSC's mission;
- the right work (i.e., products and services) are defined from the mission;
- the right way includes the processes, structures, and tools to produce the products and services;
- the right people (i.e., skill mix) is derived from the above three implications; and
- the right work includes the transition actions to move the organization from it's current to future state.

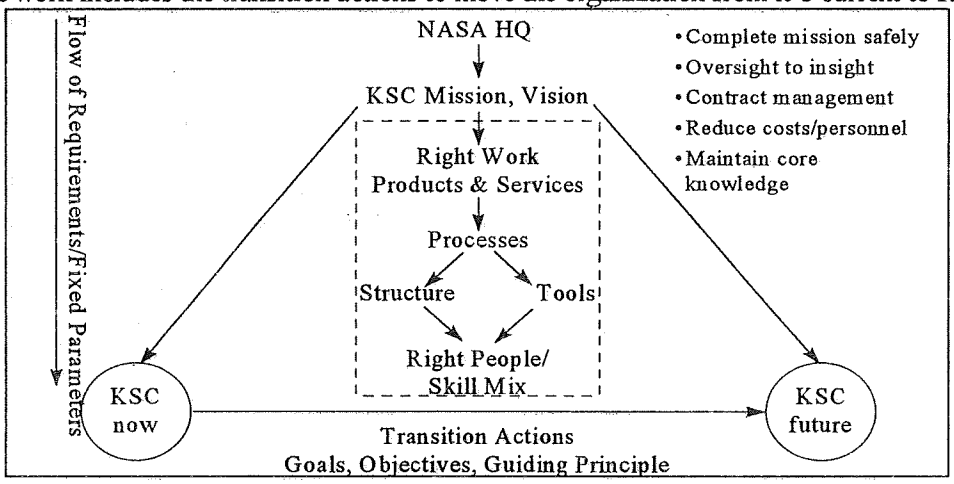


Figure 1. Doing the right work the right way with the right people at the right time.

2. METHODOLOGY

The methodology was designed to ensure a theoretically sound, valid framework was developed which reflected KSC's unique environment, goals, objectives, and concerns because there isn't a universal change model [12, 21]. The methodology was also designed to address validity issues associated with conducting change studies in the field aimed at meeting a specific organizational need [7, 9]. A goal of the project was to understand the unique KSC environment and to ensure the models reflected the concerns of KSC. The boxes represent the steps and the ellipses represent the major outputs from the steps. Three types of studies were executed: literature reviews, internal studies, and external field studies. A theoretical, literature basis was used to develop conceptual models, interpret data from interviews, and elicit best practices. KSC internal interviews and focus groups provided insight into the concerns of the KSC civil servants. The "future-state" focus groups reviewed and discussed the results of the senior management interviews. The "transition" focus groups reviewed and discussed an initial draft of the transition framework. The focus groups consisted of selected members from KSC's middle management. Past studies of KSC's organization were also reviewed [14, 27]. External interviews with heads of organizations who have completed large-scale transitions provided lessons learned.

3. RESULTS: A FRAMEWORK FOR ORGANIZATIONAL TRANSITIONS

The results of the study are best displayed in a framework consisting of four models and a project management work breakdown structure. The first model was given as the problem statement in Figure 1. This model was developed after the contractor interviews and before the "transition" focus groups were completed. The implications of the problem statement model were discussed in the introduction.

3.1. "All-Term" Transition

The second model, given in Figure 2, represents KSC's transition as an "all-term" transition. This model was developed based on Beckhard and Harris [2], personal communication, and the "future state" transition focus groups. Beckhard and Harris [2] portray an organizational transition as containing a present and future state. When we started, we were talking about two circles (i.e., year 1996 for the present state and year 2000 for the future state). However, after the senior management interviews and the "future-state" focus groups we added the third circle to reflect the later time period. The focus groups emphasized the need for KSC to define their future state beyond the near-term time frame. They felt the year 2000 mid-term state was transitory and wanted to define and take action for the long-term state. The word "all-term" was developed by a senior executive at Westinghouse who described his management style as "all-term" to reflect the need to balance the organizations for both the near and long term. The general implications of this model include:

- requirements to be met by the transition flow from the definition of the future state;
- the organization must understand the current, near term, and long-term future states and the relationships among them;
- the transition actions move the organization from the current to the desired future state;
- actions must balance survival in the near term with long-term development;
- long-term actions can be defined in more general terms than the near-term actions [12];
- earlier actions should be chosen with care because they impact choices for longer-term actions [12];
- the strategic direction defines requirements and time frames from which actions are taken; and
- all employees including those that survive the transition must be planned for [3, 25].

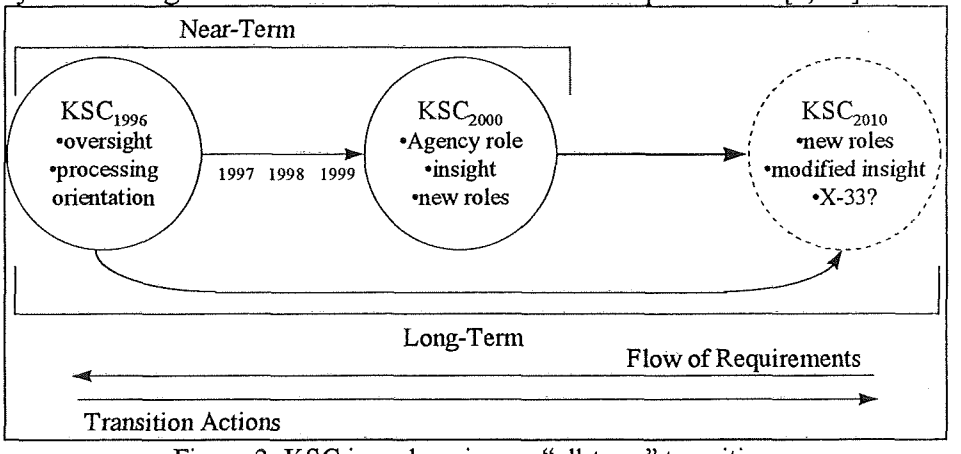


Figure 2. KSC is undergoing an "all-term" transition.

The senior management interviews, focus groups, and contractor interviews provided the data to support this notion of an all-term transition. As shown in Figure 2, the model directly applies to KSC. The roles of KSC are changing from oversight (i.e., current) to insight (i.e., near), and modified insight (i.e., long). KSC must define the actions they need to take today in the current and near term to ensure the long-term

state is reached. What actions should KSC perform to ensure they have the capability and are chosen to be a launch and landing site for X-33? KSC can use this model to balance their actions to ensure current, near-term, and long-term needs are met.

3.2. Best Practices and Lessons Learned for Organizational Transitions

The third model, given in Figure 3, represents the best practices gleaned from the literature. The representation is based on the antecedents-behavior-consequences model applied to organizational actions[17]. This model was develop by integrating the comprehensive studies of organizational change/downsizing [5, 6, 11, 23, 28, 34, 35] and works describing organizational change in general [15, 16, 18, 22, 26, 30].

The model shows that organizations take transition actions in response to drivers with the goal of producing positive and minimizing the negative results. The "drivers" represent the forces acting on the organization. These drivers provide the basis for defining specific goals and objectives of the transition. The literature has provided a description of some of the actions associated with positive results. The actions include a strategic, systematic orientation to the change. Typical negative results resulted from a loss of institutional memory, knowledge, and skill to perform the work. These results occurred because the transition was not planned systematically. The use of across the board cuts or attrition to reduce the workforce left the organization void of critical skills. The people remaining in the organization did not have the skills to complete the job. Long-term innovation and improvement suffers from the lack on in-depth knowledge of the organizational systems. This set of best practices is supported by both the contractors and the concerns of KSC's focus groups. The implications of this model include:

- an organization must design actions which produce positive and minimize the negative results;
- an organization must clearly understand the forces or drivers of the change;
- strategic, systematic actions lead to more positive results;
- a change without planning leads to negative results for the organization; and
- the potential positive and negative results from an action must be understood before action is taken.

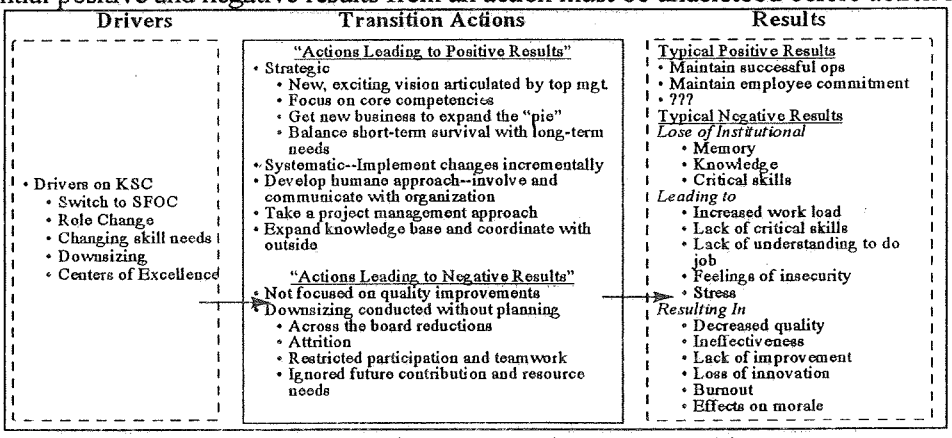


Figure 3. Best practices for organizational transitions.

Given in Figure 4 is a description of the lessons learned contractors described and the concerns of the "future state" focus groups communicated. Contractors described a process of defining the future state for the organization based on a shared understanding of all senior managers. Contractors strongly suggested that senior managers conduct an off-site session in which objective, open dialogue can be developed to

address the fundamental issues facing the organization. During organizational transitions the need to overcome existing "fiefdoms" and recognize the unity of the organization is important to success. Contractors, focus groups, and senior managers were consistent in this message. Once the future state is defined and a time frame for which the future state must be achieved, a set of objectives can be developed to accomplish the future state. A project management approach was also emphasized.

By providing the definition of the future state, the existing organization can be systematically analyzed to determine how the work being completed matches with requirements of the desired future state. Priorities are set to determine which work to continue doing. For example, one organization prioritized their work into three groups based on their level of importance to meeting their mission. Only the most essential work was still completed. Another organization defined a goal to reduce flexibility while maintaining capability (e.g., reduce the number of shifts from three to one). Once the work is determined and the process by which the work is completed, the needed skills can be defined. This systematic analysis leads to process improvements focused on the mission critical areas. The focus group's supported this systematic analysis focused on mission critical work when they wanted KSC to define "meaningful work" for the civil servants. Two benefits can be received by KSC if they follow this approach. First, employees see that KSC senior management is working a process to attack the transition problem. Employees may not personally like the results of the process to themselves but they are more comfortable knowing that KSC is being proactive. Second, the Agency's requirements will be met because both the future state definition and analysis of the organization is based on the center's mission as defined by the Agency. KSC can use Figure 4 as a concise description of a transition process.

- Senior Management establish a
  - **UNIFIED FUTURE STATE/VISION** with a specific **TIMEFRAME** based on shared understanding
  - set of objectives and ground rules
  - project management approach (leadership, task team, outside help)
- which allows the **SYSTEMATIC ANALYSIS** of
  - requirements--fixed/known and unknown
  - priorities
  - tasks--processes
  - personnel skills
- to support **PROCESS IMPROVEMENT** by
  - surgically eliminating unneeded work processes and positions
  - focusing on **MISSION CRITICAL** and **CORE COMPETENCY** areas
- with benefits of
  - regularly communicating a systematic process or rationale for decisions and actions to stakeholders
  - providing value added service to the agency. (Contractors & Focus Groups)

Figure 4. Lessons learned and concerns describe an approach for organizational transitions.

3.3. Aligning the Organization

Figure 5 portrays the fourth model which shows the need to align KSC both externally and internally with the needs of the customers. The external alignment is in matching the organization's products and services with the market and customer needs. This alignment also includes the high-level definition of the organization's roles and core processes. Based on this high-level definition, the organization can align the internal components of the organization. The model is adapted from Kurstedt's [17] management system model by emphasizing the macro representation of the organization. We use the model to show how an organization needs to align its components: the process by which work is completed, the tools people use to complete the work, and the people completing the work [10]. The alignment comes through the interfaces

of the components: the structure by which people are organized to do the work (i.e., processes and people), the information available to the worker (i.e., people and tools), and the metrics used (i.e., processes and tools). When changing the organization, the organization must align these components and interfaces. For example, an organization cannot do the same work (i.e., products and services) in the same way (i.e., process, tools, and structure) by reducing the people. The changes in components must be balanced and aligned with each other. The dotted arrows inside the circle show the systematic, logical way to use the model. The analysis and implementation starts from the products and moves to the people. An across the board reduction and then adjustment in process may not produce the customer's desired results. The implications of this model include:

- the external world defines the market's needs;
- the organization must align its products and services with customer needs;
- the products and services are developed from core processes;
- the organization must align processes, structure, tools, and people to deliver products;
- an organization cannot reduce people without affecting the process and process performance; and
- changes in any of the one components will affect the other components and organizational performance.

This model can be used for both macro level (i.e., KSC as a whole) and at a micro level (i.e., an individual's work process). KSC can use this model to address the issue of what the product and processes of KSC for a given range of civil servants are. KSC can also use this basic model to address issues about the "insight role": What is the insight process? How will civil servants and contractors be organized to complete insight? What measurements and data will be used for insight? What tools will be used to convert the data to information for insight? What is the skill mix needed for insight?

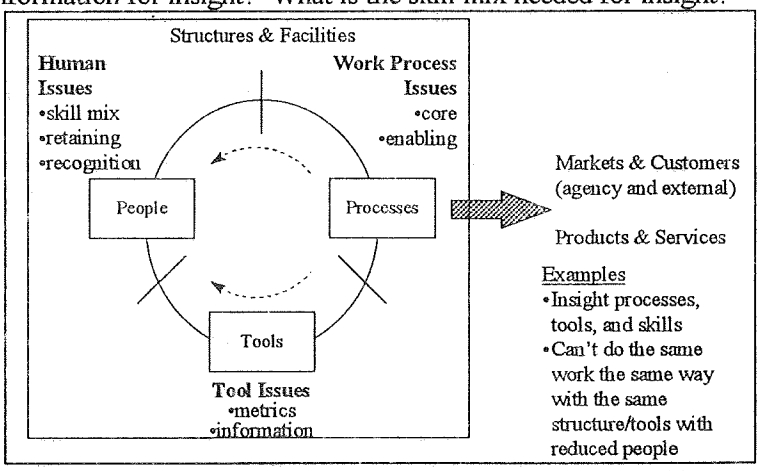


Figure 5. An organization must be aligned both externally and internally.

3.4. Level-1 Work Breakdown Structure

Based on the models, best practices, lessons learned, and concerns developed from the literature, contractor interviews, senior management interviews, and focus groups, we have developed a work breakdown structure (WBS) to implement the concepts in the models. Figure 6 is a flow representation of the six steps in the level-1 WBS. In the paper, we concentrate on level-1; a full WBS has been defined.

Strategic direction provides both the future state and overall philosophy (e.g., goals, ground rules, focus) for the transition [4, 24, 31]. In this step senior management articulates and communicates a unified

message to lead the transition. Defining roles provides an explicit description of what KSC will do and how it relates to its external world. For example, KSC would define it's relationship to other Agency centers and the contractors [29]. Aligning processes aims to improve KSC's process to meet the defined roles and to ensure KSC's processes contain only value-added activities [19]. Aligning structures aims to ensure KSC's people are organized to complete the work processes. Aligning the workforce ensures the right people (skills and numbers) are available to meet the requirements as defined by the previous four high-level steps of the transition and addresses options to align the workforce [8, 33]. Managing the process ensures a smooth proactive transition. This steps includes the crucial steps of leadership, communication [32] and a project management approach. The implications of this model include:

- the strategic direction drives the actions;
- aligning the workforce is the last step;
- the "right work" is define by aligning the strategic direction and roles;
- the "right way" is defined by aligning the processes and structures;
- the right people is defined by aligning the workforce.

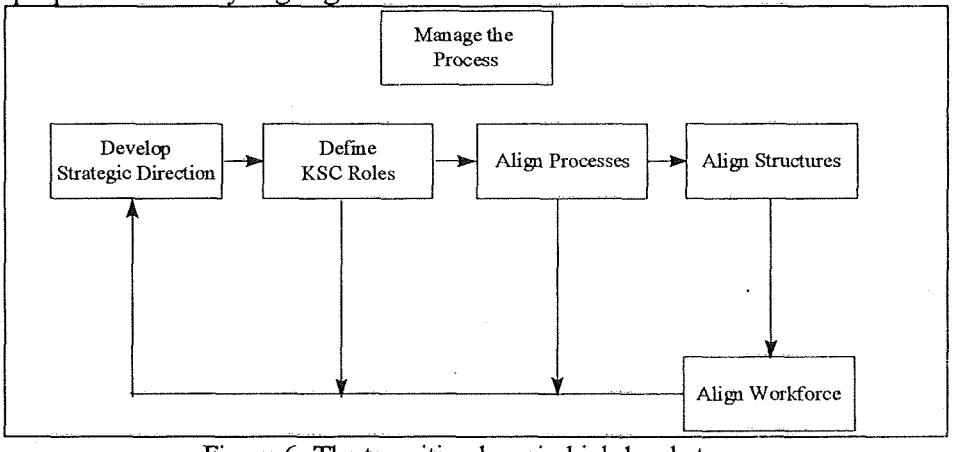


Figure 6. The transition has six high-level steps.

5. CONCLUSION

In this paper, we presented a transition framework based on a project management approach and a set of models. The models and associated implications provide an integration of the key issues to consider and the work breakdown structure provides a set of steps to use when undergoing a large-scale change. The framework is not an answer, but an evolving set of actions to validate, add, delete, and refine to reflect further research and the unique characteristics of a given organization. As with all projects, a process must be used to assign responsibilities and monitor progress. The initial framework and WBS is provided in this paper for an organization to begin its change efforts. The framework is designed to help an organization do the right work the right way with the right people at the right time.

6. ACKNOWLEDGMENTS

The worked complete in this report would not have been possible without the combined efforts of many people at the Kennedy Space Center. I thank Saul Barton for sharing his experience in organizational change activities and insights on successful implementation of efforts in organizations. This work has truly been a collaborative effort with Saul. The individuals on KSC's strategic planning and business integration teams provided an additional source of ideas and concerns. Jim Jennings and Warren Camp provided a sounding board and provided their insight to make the product "fit" the KSC environment. I thank them for

their guidance. I thank Tammy Belk and Betty Cromie for helping ensure the summer went smoothly. I thank Gregg Buckingham, Kari Stiles, and Roger Johnson for their management of the NASA/ASEE summer faculty program. Finally, I thank the people of NASA/ASEE who make this program possible.

REFERENCES

[1] Appelbaum, S. H., Simpson, R. and Shapiro, B. T., "The Tough Test of Downsizing", *Organizational Dynamics*, 1987, pp. 68-79.
[2] Beckhard, R. and Harris, R. T., *Organizational Transitions*, Reading, MA: Addison-Wesley, 1987.
[3] Brockner, J., Grover, S., O'Malley, M. N., Reed, T. F., and Glynn, M. A. "Threat of Future Layoffs, Self-Esteem, and Survivors' Reactions: Evidence from the Laboratory and the Field", Strategic Management Journal, 14, 1993, pp. 153-166.
[4] Bruton, G. D., Keels, J. K., and Shook, C. L., "Downsizing the Firm: Answering the Strategic Questions", *The Academy of Management Executive*, 10(2), May 1996, pp. 38-45.
[5] Cameron, K., Freeman, S. J. and Mishra, A. K., "Best Practices in White-Collar Downsizing: Managing Contradictions", *The Academy of Management Executive*, V(3), August 1991, pp. 57-73.
[6] Cameron, K., Freeman, S. J. and Mishra, A. K., "Downsizing and Redesigning Organizations", In *Organizational Change and Redesign*, (Eds: Huber, G. P. and Glick, W. H.), New York, NY: Oxford University Press, 1993, pp. 19-65.
[7] Cunningham, J. B., *Action Research and Organizational Development*, Westport, Connecticut: Praeger, 1993.
[8] Feldman, D. C. and Leana, C. R., "Managing Layoffs: Experiences at the Challenger Disaster Site and the Pittsburgh Steel Mills", Organizational Dynamics, 1989, pp. 52-64.
[9] Glick, W. H., Huber, G. P., Miller, C. C., Doty, D. H., and Sutcliffe, K. M., "Appendix: Studying Changes in Organizational Design and Effectiveness: Retrospective Event Histories and Periodic Assessments", In *Organizational Change and Redesign*, (Eds: Huber, G. P. and Glick, W. H.), New York, NY: Oxford University Press, 1993, pp. 411-433.
[10] Hanna, D.P., *Designing Organizations for High Performance*, Reading, MA: Addison-Wesley, 1988.
[11] Huber, G. P. and Glick, W. H., "What Was Learned About Organization Change and Redesign", In *Organizational Change and Redesign*, (Eds: Huber, G. P. and Glick, W. H.), New York, NY: Oxford University Press, 1993, pp. 383-392.
[12] Kanter, R. M., Stein, B. A., and Jick, T. D., *The Challenge of Organizational Change: How Companies Experience It and Leaders Guide It*, New York, NY: The Free Press, 1992.
[13] Keidel, R. W., "Rethinking Organizational Design", *The Academy of Management Executive*, 8(4), 1994, pp. 12-30.
[14] Kennedy Space Center Loyalty Study Team, "Loyalty and Quality of Relationships", Kennedy Space Center, FL: October 1995.
[15] Kilmann, R. H., Covin, T. J., and Associates, *Corporate Transformation*, San Francisco, CA: Jossey-Bass Publishers, 1988.
[16] Kirkpatrick, D. L., *How to Manage Change Effectively*, San Francisco, CA: Jossey-Bass Publishers, 1985.
[17] Kurstedt, H. A. *Management Systems Engineering*, Blacksburg, Va.: Author, 1994.
[18] LaMarsh, J., *Changing the Way We Change: Gaining Control of Major Operational Change*, Reading, MA: Addison-Wesley, 1995.
[19] Leifer, R. and Burke, W. J., "Organizational Activity Analysis: A Methodology for Analyzing and Improving Technical Organizations", *IEEE Transactions*, 41 (3), August 1994, pp. 234-244.
[20] Marshall, R. and Yorks, L. "Planning for a Restructures, Revitalized Organization", Sloan Management Review, Summer, 1994, pp. 81-91.
[21] McKinley, W., Sanchez, C. M., and Schick, A. G., "Organizational Downsizing: Constraining, Cloning, Learning", *The Academy of Management Executive*, IX(3), August 1995, pp. 32-44.
[22] Mohrman, A. L., Mohrman, S. A., Ledford, G. E., Cummings, T. G., Lawler, E. E., and Associates, *Large Scale Organizational Change*, San Francisco, CA: Jossey-Bass Publishers, 1989.
[23] National Academy of Public Administration, "Effective Downsizing: A Compendium of Lessons Learned for Government Organizations", Washington, DC, 1996.
[24] Neal, J. A. and Tromley, C. L., "From Incremental Change to Retrofit: Creating High-Performance Work Systems", *The Academy of Management Executive*, 9(1), February 1995, pp. 42-54.
[25] O'Neill, H. M. and Lenn, J., "Voices of Survivors: Words that Downsizing CEOs Should Hear", *The Academy of Management Executive*, IX(4), November 1995, pp. 23-34.
[26] Price Waterhouse Change Integration Team, *Better Change: Best Practices for Transforming Your Organization*, Burr Ridge, IL: Irwin Professional Publishing, 1995.
[27] Right Associates, "Organization Measurement Consulting Services offered to NASA Kennedy Space Center", Prepared by Right Associates, February 1996.
[28] Robertson, P. J., Roberts, D. R., and Porras, J. I., "Dynamics of Planned Organizational Change: Assessing Empirical Support for a Theoretical Model", *Academy of Management Journal*, 36(3), June 1993, pp. 619-634.
[29] Serlin, M. D., "The Competitors", *Government Executive*, June 1996, pp. 29-33.
[30] Sink, D. S. and Morris W. T., *By What Method?*, Industrial Engineering and Management Press, Institute of Industrial Engineering, 1995.
[31] Sink, D. S. and Tuttle, T. C., *Planning and Measurement in Your Organization of the Future*, Industrial Engineering and Management Press, Institute of Industrial Engineering, 1989.
[32] Smeltzer, L. R. and Zener, M F. "Development of a Model for Announcing Major Layoffs", Group & Organizational Management, 17(4), 1992, pp. 446-472.
[33] Tomasko, R. M., *Downsizing: Reshaping the Corporation for the Future*, American Management Association, 1987.
[34] United States General Accounting Office, "Workforce Reductions: Downsizing Strategies Used in Selected Organizations", Washington, D.C.: US GAO, March 1995.
[35] Wyatt Company, The, "Best Practices in Corporate Restructuring", Chicago, IL: The Wyatt Company, 1993.

1996 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

DISPERSION COMPENSATION OF FIBER OPTIC SYSTEMS FOR KSC APPLICATIONS

Dr. Samuel P. Kozaitis, Associate Professor
Division of Electrical, Computer Science and Engineering
Florida Institute of Technology
Melbourne, Florida

KSC Colleague - Larry Hand
Communications/Fiber Optics

Contract Number NASA-NGT10-52605

July 12, 1996

ABSTRACT

Installed fibers such as those at Kennedy Space Center are optimized for use at 1310 nm because they have zero dispersion at that wavelength. An installed fiber system designed to operate at 1310 nm will operate at a much lower data rate when operated at 1550 nm because the dispersion is not zero at 1550 nm.

Using dispersion measurements of both installed and dispersion compensating fibers, we compensated a 21.04 km length of installed fiber with 4.25 km of dispersion compensating fiber. Using the compensated fiber-optic link, we reduced the dispersion to 0.494 ps/nm·km, from an uncompensated dispersion of 16.8 ps/nm·km.

The main disadvantage of the compensated link using DC fiber was an increase in attenuation. Although the increase was not necessarily severe, it could be significant when insertion losses, connector losses and fiber attenuation are taken into account.

1.0 Introduction

We wanted to increase the information-carrying capacity of an existing fiber-optic communication system. Our system is optimized to operate at a nominal wavelength of 1310 nm and is near its maximum data capacity. Replacing the existing system with a new system is not a practical option, nor is the use of multiple wavelengths. The use of multiple wavelength requires expensive equipment that is not always easily obtained, and the modification to the existing system could be complicated.

Recent developments in fiber optic technology are not always optimized for the 1310 nm wavelength. The invention of the erbium-doped fiber amplifier[1] (EDFA) has changed fiber-optic system design because its has reduced the importance of loss in fiber systems. Most EDFAs operate at a wavelength around 1550 nm; therefore, many recent developments in fiber optic technology have been for the 1550 nm wavelength.

An installed fiber system designed to operate at 1310 nm will operate at a much lower data rate when operated at 1550 nm. Installed fibers such as those at Kennedy Space Center (KSC) are optimized for use at 1310 nm because they have zero material dispersion at that wavelength. However, if components operating at 1550 nm are used with this fiber, material dispersion will dominate and the speed (bit-rate) of the system will decrease when compared to the bit-rate at 1310 nm.

The most straightforward way to reduce the effect of dispersion is to use dispersion compensating (DC) fiber. DC fiber has the characteristic that the dispersion at 1550 nm is generally opposite to that of the fiber it is to compensate. DC fibers can be designed to compensate the dispersion of standard fibers (optimized for 1310 nm) through their index of refraction profile.[2-3] Using the proper length of DC fiber in series with a standard fiber, the total dispersion may be zero. Most references in the literature on this topic deal with the use of DC fibers in series with standard fibers and EDFAs to increase the distance between repeaters of a fiber system.[4]

2.0 Chromatic dispersion

Chromatic dispersion is caused by a variation in the index of refraction of a material with wavelength. Since it can cause pulse spreading in an optical communication system, it can degrade system performance. This is shown schematically in Fig. 1. As a pulse of light travels down an optical fiber, different wavelengths travel at different speeds. By the time the pulses arrive at the receiver, they may have spread over several bit periods and cause errors. To reduce

the effects of dispersion, we must somehow compensate for its efffect, or transmit data at a lower rate. In this section, we discussed some of the theory of chromatic dispersion followed by the relationship between dispersion and bandwidth.

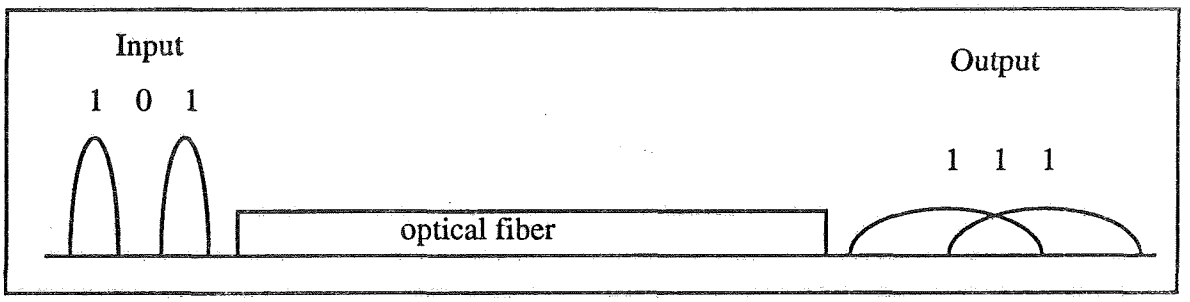


Figure 1 Error caused by chromatic dispersion in an optical fiber system

The dispersion of a fiber is defined as the broadening per unit spectral width for a fiber of unit length.[5] It is based on the second derivative of the index of refraction as a function of wavelength as is usually written as

$$D = - (\lambda/c)(d^2n/d\lambda^2), \quad (1)$$

where c is the speed of light. The units are ps/km·nm, which is the amount of broadening in picoseconds that would occur in a pulse with a bandwidth of one nanometer while propagating through one kilometer of fiber. The maximum bandwidth of a system can be written as

$$\text{maximum bandwidth} = 1/|LD\Delta\lambda|, \quad (2)$$

where L is the length of the fiber, and $\Delta\lambda$ is the spectral width of the source. For example, given a length of fiber of 10 km, D = 100 ps/km·nm, and a source with a spectral width of 1 nm, the maximum bandwidth is 1 GB/sec. The bandwidth will increase linearly with a decrease in dispersion or length of fiber.

3.0 Compensation of chromatic dispersion using dispersion compensating fibers

To compensate for the dispersion of a standard fiber at a particular wavelength, a DC fiber can be used in series with a standard fiber as shown in Fig. 2.

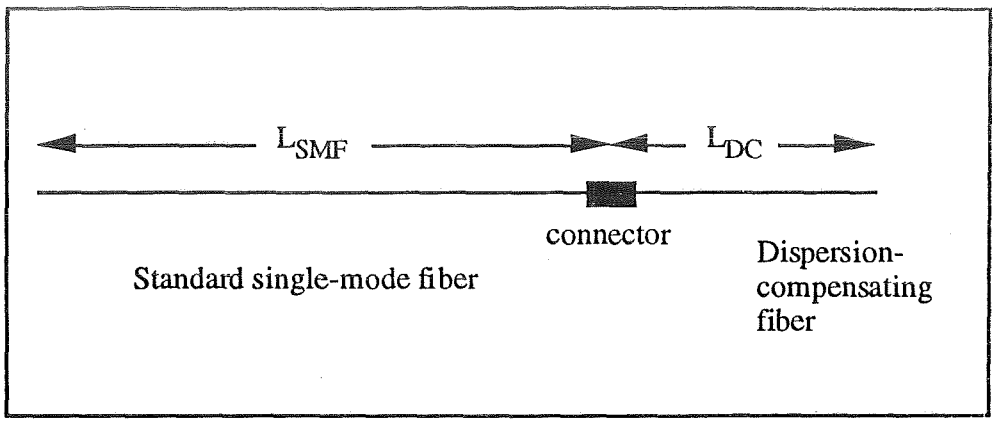


Figure 2 Schematic diagram of compensation of standard fiber

In this case, the total dispersion is

$$\Delta\tau = L_{SMF}D_{SMF} + L_{DC}D_{DC}, \tag{3}$$

where L_{SMF} and L_{DC} are the lengths of the standard and DC fibers, and D_{SMF} and D_{DC} are the dispersion of the standard and DC fibers respectively. Because length is always positive, the total dispersion will go to zero at the wavelength of interest if the DC fiber has a negative dispersion of the proper magnitude. To compensate a fiber of length L_{SMF}, we set $\Delta\tau$=0 in Eq. (3) and solved for L_{DC}. We found that the length of DC fiber needed was

$$L_{DC} = - L_{SMF}D_{SMF}/D_{DC}, \tag{4}$$

where the dispersion of the DC fiber is negative at the wavelength where compensation is to take place.

For a length of standard fiber, the length of DC fiber needed for compensation at a particular wavelength is the negative of the ratio of the dispersions at that wavelength. This was described as a compensating factor and written as

$$\text{compensating factor (CF)} = - D_{SMF}/D_{DC}, \tag{5}$$

where small numbers lead to short lengths of DC fiber.

The figure of merit (FOM) measure in a sense measures the efficiency of a fiber.[6] It is defined as, the ratio of the dispersion of the DC fiber and its attenuation, at a particular wavelength. The FOM can be written as

$$\text{FOM} = D_{DC} / \alpha_{DC}, \tag{6}$$

where α_{DC} is the attenuation of the DC fiber. A more negative number indicates better efficiency.

4.0 Dispersion and attenuation measurements of fibers

To measure dispersion we followed the corresponding fiber-optic test procedure (FOTP) standard[7] and used a York S18 chromatic dispersion system. The fiber we were attempting to compensate was Corning SMF-28CPC3 single-mode fiber. We measured its characteristics with a fiber of length 8.87 km. The zero-dispersion wavelength was found to be 1323.02 nm with a slope of 0.0926 ps/nm^2·km at that wavelength. The dispersion at 1550 nm was found to be 16.83 ps/nm·km and the attenuation 0.37 dB/km at 1550 nm.

We performed dispersion measurements on two types of DC fiber. They were both labeled as Corning FiberGain Module fiber. Therefore, we referred to the fibers here as FGM1 and FGM2. We performed five experiments and averaged their results. We measured a length of 1.01 km of FGM1 fiber and found its dispersion at 1550 nm to be -115.0 ps/nm·km. We measured a length of 4.25 km of FGM2 fiber and found its dispersion at 1550 nm to be -79.13 ps/nm·km. Note that both types of fibers each had SMF fiber pigtails with two permanent connectors measuring a total of approximately 1.23 m. A graph showing the dispersion data in Fig. 3.

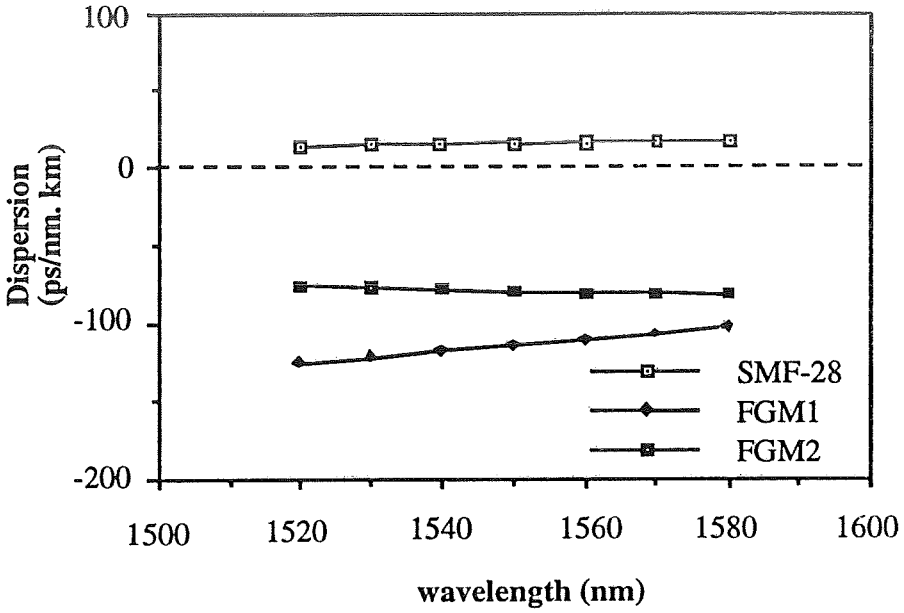


Figure 3 Dispersion data in vicinity of 1550 nm

A summary of the measurements is given in Table 1.

Table 1 Characteristics of dispersion-compensating fibers

Fiber	Compensating factor	Dispersion at 1550 nm (ps/nm·km)	Dispersion slope at 1550 nm (ps/nm^2·km)	Attenuation at 1550 nm (dB/km)	FOM at 1550 nm (ps/nm·dB)
SMF-28CPC3	---	16.83	0.060	0.37	---
FGM1	0.147	- 115.0	0.367	1.8	-63.9
FGM2	0.213	- 79.13	-0.106	0.75	-106

5.0 Compensated fiber-optic links

We calculated the dispersion of compensated links in the vicinity of 1550 nm. The results are shown in Fig. 4.

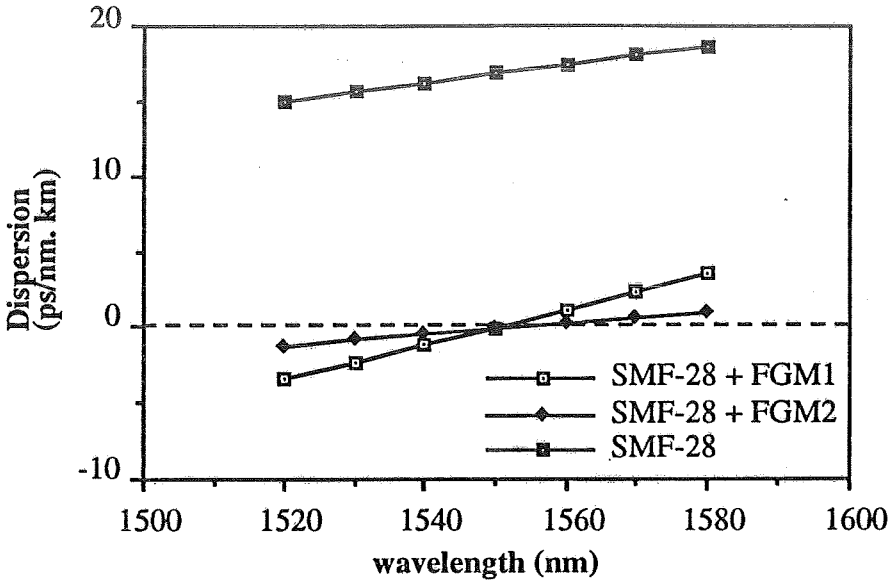


Figure 4 Calculated dispersion of compensated links for two types of DC fiber

A summary of the calculations on compensated links is given in Table 2.

Table 2 Characteristics of dispersion-compensated links

Fiber link	Dispersion slope at 1550 nm ($ps/nm^2 \cdot km$)	Attenuation at 1550 nm of total fiber (dB/km)
uncompensated	0.0601	0.37
SMF-28 + FGM1	0.114	0.55
SMF-28 + FGM2	0.0373	0.44

5.2 Selected link

To test a dispersion compensated system in a realistic setting, we connected a fiber-optic link using existing installed fibers at KSC. We used two links in series that both originated from the engineering development lab (EDL) that measured approximately 5.2 km, and 15.8 km. The links were identified by numbers on the fiber optic panel, 39-40, and 43-44, and their circuit numbers, 6TEM9948, and 6TEM9945 respectively. The links were connected with a 1 m length of SMF fiber, and connected to the dispersion measuring instrument through a total length of

49.2 m of SMF fiber. The total fiber loop went through the following KSC buildings: EDL-O&C-CDSC-O&C-EDL-CDSC-VABR-CDSC-O&C-EDL. The total length of fiber was measured to be 21.04 km.

5.3 Measurements of compensated link

We used the total length of the FGM2 fiber to compensate the selected link; ideally, the fiber would compensate 20 km so there should be a small dispersion in our link. We added the FGM2 fiber to the selected link and measured the total loop to be 25.29 km.

Using Eq. (3) we calculated the dispersion of the compensated link to be 0.704 ps/nm·km. We performed 25 dispersion measurements with the FGM2 fiber at the source side of the dispersion measuring instrument and found the dispersion to be 0.582 ps/nm·km, with a standard deviation of 0.304. With the FGM2 fiber on the opposite end, we found the dispersion to be 0.407 ps/nm·km, with a standard deviation of 0.408. Therefore, it appears that the difference in measurements is within statistical error and we used the average of the 50 experiments for the value of the measured dispersion as 0.494 ps/nm·km. The dispersion results of the compensated link are shown in Table 3.

Table 3 Measurements of dispersion-compensated links

Fiber link	length (km)	calculated dispersion (ps/nm·km)	measured dispersion (ps/nm·km)
uncompensated	21.04	-	16.83
compensated SMF-28 + FGM2	25.29	0.704	0.494 σ=0.371

6.0 Summary and conclusion

To compensate for the dispersion of an installed standard fiber, a DC fiber can be used in series with the standard fiber. Because the dispersion of SMF fiber is positive, we need a fiber with a dispersion that is negative for compensation. Large negative dispersions of the compensating fiber will lead to short lengths of DC fiber and reduce space requirements.

Using measurements of dispersion for both SMF and DC fibers we compensated a 21.04 km length of installed SMF fiber with 4.25 km of DC fiber. Using the compensated fiber-optic

link, we reduced the dispersion to 0.494 ps/nm·km, from an uncompensated dispersion of 16.8 ps/nm·km. We did not find a difference in dispersion measurements when the DC fiber was placed at the beginning or end of the link.

The main disadvantage of the compensated link using DC fiber was an increase in attenuation. Although the increase was not necessarily severe, it could significant when insertion losses, connector losses and fiber attenuation is accounted for.

7.0 References

[1] A. F. Elrafaie, R. E. Wagner, D. A. Atlas, and D. G. Daut, *J. Lightwave Technol.* 6, 704-709 (1988)

[2] R. Lundin, "Minimization of the chromatic dispersion over a broad wavelength range in a single-mode optical fiber," *Applied Optics* 32, 3241-3245 (1993)

[3] A. J. Antos, D. W. Hall, and D. K. Smith, "Dispersion-compensating fiber for upgrading existing 1310-nm-optimized systems to 1550-nm operation," *OFC/IOOC '93 Technical Digest*, 204-205 (1993)

[4] H. Taga, et. al., "Performance evaluation of the different types of fiber-chromatic-dispersion equalization for IM-DD ultralong-distance optical communication systems with Er-doped fiber amplifiers," *J. Lightwave Technol.* 12, 1616-1621 (1994)

[5] C-L. Chen, "Elements of Optoelectronics and Fiber Optics," Irwin:Chicago, p. 464 (1996)

[6] J. M. Dugan, et. al., *OFC/IOOC '92 Technical Digest*, paper PD-14, 367-368 (1992)

[7] EIA Standard "FOTP-169, Chromatic dispersion measurement of optical fibers by the phase-shift method," EIA-455-169, Global Engineering Documents, Irvine:CA (1987)

1996 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

ADAPTIVE NOISE SUPPRESSION USING DIGITAL SIGNAL PROCESSING

Dr. David Kozel, Associate Professor
Engineering Department
Purdue University Calumet
Hammond, Indiana

KSC Colleague - Richard Nelson
Communications/Rf and Audio

Contract Number NASA-NGT10-52605

August 9, 1996

Abstract

A signal to noise ratio dependent adaptive spectral subtraction algorithm is developed to eliminate noise from noise corrupted speech signals. The algorithm determines the signal to noise ratio and adjust the spectral subtraction proportion appropriately. After spectral subtraction low amplitude signals are squelched. A single microphone is used to obtain both the noise corrupted speech and the average noise estimate. This is done by determining if the frame of data being sampled is a voiced or unvoiced frame. During unvoiced frames an estimate of the noise is obtained. A running average of the noise is used to approximate the expected value of the noise. Applications include the Emergency Egress Vehicle and the Crawler-Transporter.

1. Introduction

It is desired to incorporate adaptive noise suppression into the communications equipment on the Emergency Egress Vehicle and the Crawler-Transporter. In the case of the Emergency Egress Vehicle, people are fixed relative to the noise source. The spectral content of the noise source changes as a function of the speed of the vehicle and its engine. In the case of the Crawler-Transporter the people can move relative to the Crawler-Transporter. Thus, the noise a person hears will vary with their location relative to the Crawler-Transporter and if the hydrolic leveling device on the Crawler-Transporter is being used. Due to the varying nature of the noise, an adaptive algorithm is necessary for both applications. Furthermore, the noise frequencies produced by both applications are in the voice band range. Thus, standard filtering techniques will not work. A signal to noise ratio dependent adaptive spectral subtraction algorithm is developed to eliminate the noise. OIS-D microphones are used. OIS-D microphones have noise suppression of a mechanical nature, which provides approximately 15dB of noise suppression. This is sufficient to provide a signal to noise ratio favorable enough for spectral subtraction to perform very well.

2. Spectral Subtraction

The additive noise model used for spectral subtraction assumes that noise corrupted speech is composed of speech plus additive noise.

$$x(t)=s(t) + n(t) \tag{1}$$

where:

- x(t) noise corrupted speech
- s(t) speech
- n(t) noise

Taking the Fourier Transform of equation (1)

$$X(f) = S(f) + N(f) \tag{2}$$

$X(f)$, $S(f)$, and $N(f)$ are complex so they can be represented in polar form

$$|X(f)|e^{j\theta x} = |S(f)|e^{j\theta s} + |N(f)|e^{j\theta n} \tag{3}$$

Solving for the speech

$$|S(f)|e^{j\theta s} = |X(f)|e^{j\theta x} - |N(f)|e^{j\theta n} \tag{4}$$

Since the phase of the noise is in general unavailable, the phase of the noise corrupted speech is commonly used to approximate the phase of the speech. This is equivalent to assuming the noise corrupted speech and the noise are in phase. As a result the speech magnitude is approximated from the difference of the noise corrupted speech and noise magnitudes.

$$\hat{S}(f) = \left|\hat{S}(f)\right| e^{j\theta x} = \left(|X(f)| - |N(f)|\right) e^{j\theta x} \tag{5}$$

The inverse Fourier Transform yields the estimate of the speech.

$$\hat{s}(t) = \mathcal{F}^{-1}\left\{\hat{S}(f)\right\} \tag{6}$$

There are different types of spectral subtraction. The type described above is termed magnitude spectral subtraction, because the magnitude of the noise spectrum at each frequency is subtracted. A derivation for power spectral subtraction is given in [1]. In its most general form spectral subtraction is written as [2]

$$\hat{S}(f) = \left\{|X(f)|^b - \alpha(SNR(f))E\left[|N(f)|^b\right]\right\}^{\frac{1}{b}} e^{j\theta x} \tag{7}$$

The exponent, b, equals 1 for magnitude spectral subtraction and 2 for power spectral subtraction. The proportion of noise subtracted, α, can be variable and signal to noise ratio dependent. In general α is greater than 1 which is termed over subtraction. Over subtraction is used to reduce the distortion caused from approximating the phase. In equation (7), E[*] represents the expected value of [*].

2.1 Limitations of Spectral Subtraction

When using any algorithm, it is important to understand its limitations and restrictions. The phase approximation used in the speech estimate produces both magnitude and phase distortion in each frequency component of the speech estimate. This can be seen in Figure 1 by the vector representation of equation (4) and equation (5) respectively for any one frequency. If the magnitude of the noise, $|N|$, is "small" relative to the magnitude of the corrupted speech, $|X|$, the distortion caused by using the noise corrupted speech phase, θ_x, in place of the noise phase is minimal and unnoticeable to the human ear. Likewise, if the phase of the noise, θ_n, is "close" to the phase of the corrupted speech, θ_x, the resulting error produced by the approximation is

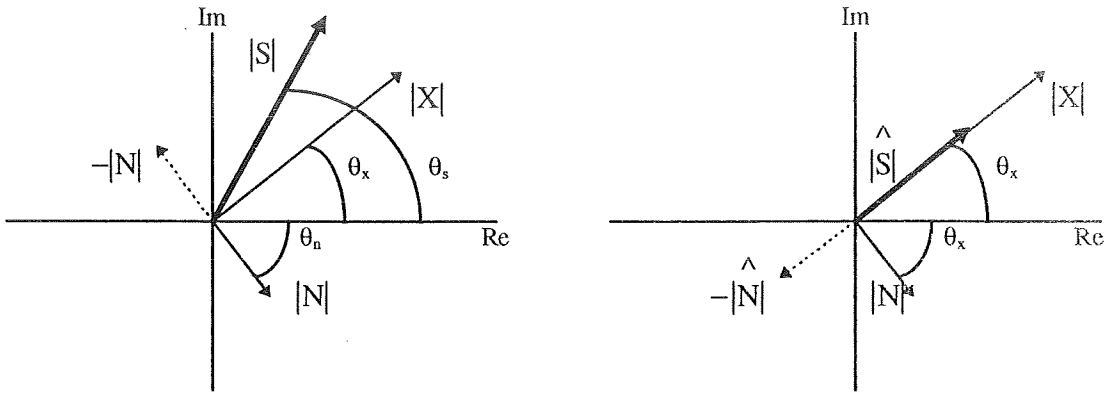


Figure 1. Vector Representation of Equations (4) and (5) Respectively

minimal and unnoticeable to the human ear. Since the relative phase between θ_x and θ_n is unknown and varies with time and frequency, the ratio between the magnitude of the noise corrupted speech and the noise is used as an indication of accuracy.

3. Signal to Noise Ratio Dependent Adaptive Spectral Subtraction Algorithm

A Diagram of the Signal to Noise Ratio Dependent Adaptive Spectral Subtraction algorithm (SNRDASS) is shown in Figure 2. Details of the algorithm are described in the following subsections.

3.1 Framing, Windowing, Zero Padding and Recombining the Signal

The process of windowing, zero padding, and recombination of the signal is shown in Figure 3. The sampled signal is segmented into frames each containing m points. Each frame is multiplied by a triangular window containing m points. This is required since the algorithm uses a Fast Fourier Transform (FFT) which assumes that the signal is periodic relative to the frames. If a window is not used spurious frequencies are produced due to signal levels at the ends of each frame not being equal. As a result of windowing each frame is required to overlap the previous frame in time by 50 percent. This allows the two triangular windowed components to add to the original signal when recombined. If a window type other than a triangular window is used the addition of frames can produce oscillation errors of up to approximately 9 percent of the original amplitude in the recombined signal. Spectral subtraction can be considered as a time varying filter[3] which can vary from frame to frame.

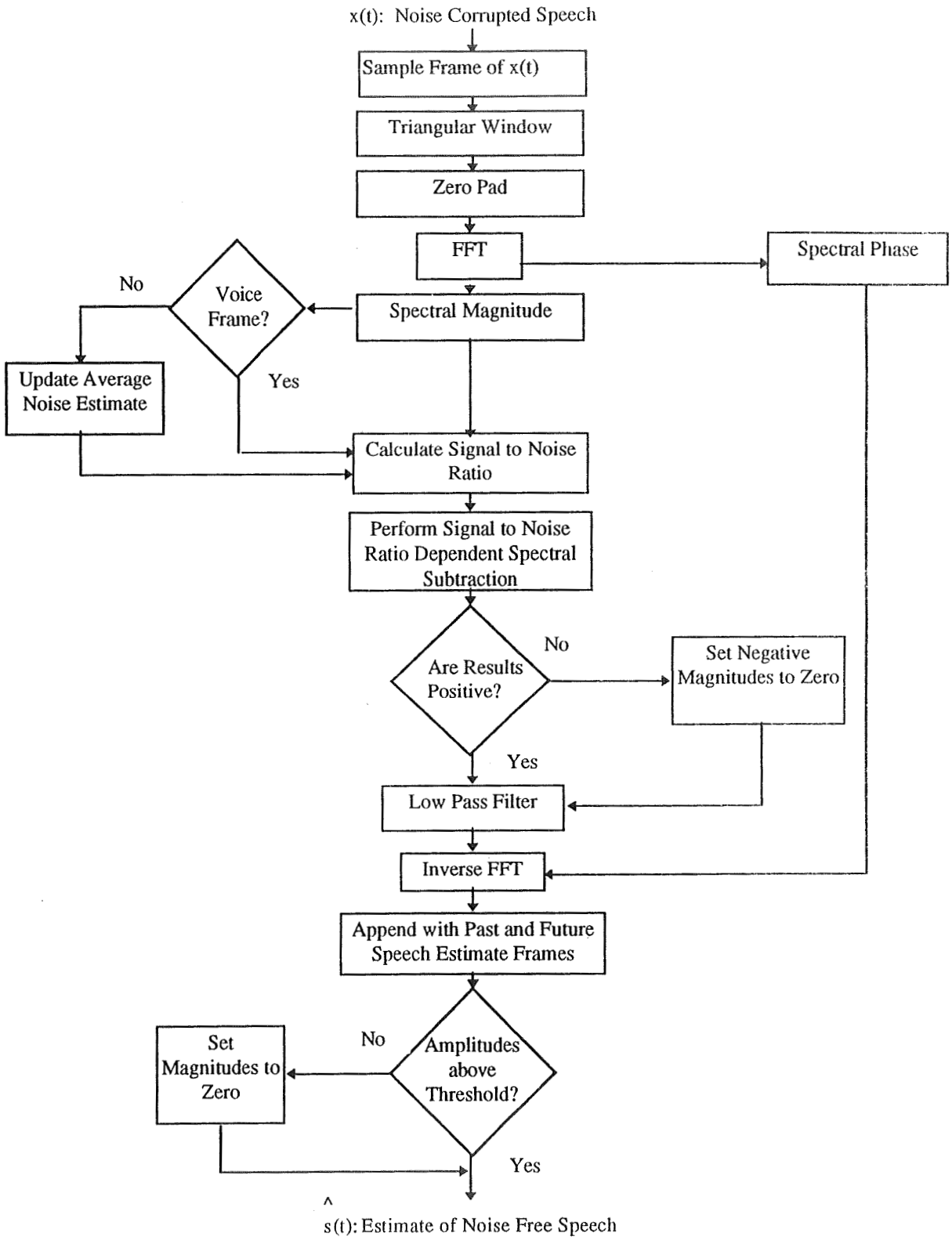


Figure 2 Signal to Noise Ratio Dependent Adaptive Spectral Subtraction Algorithm

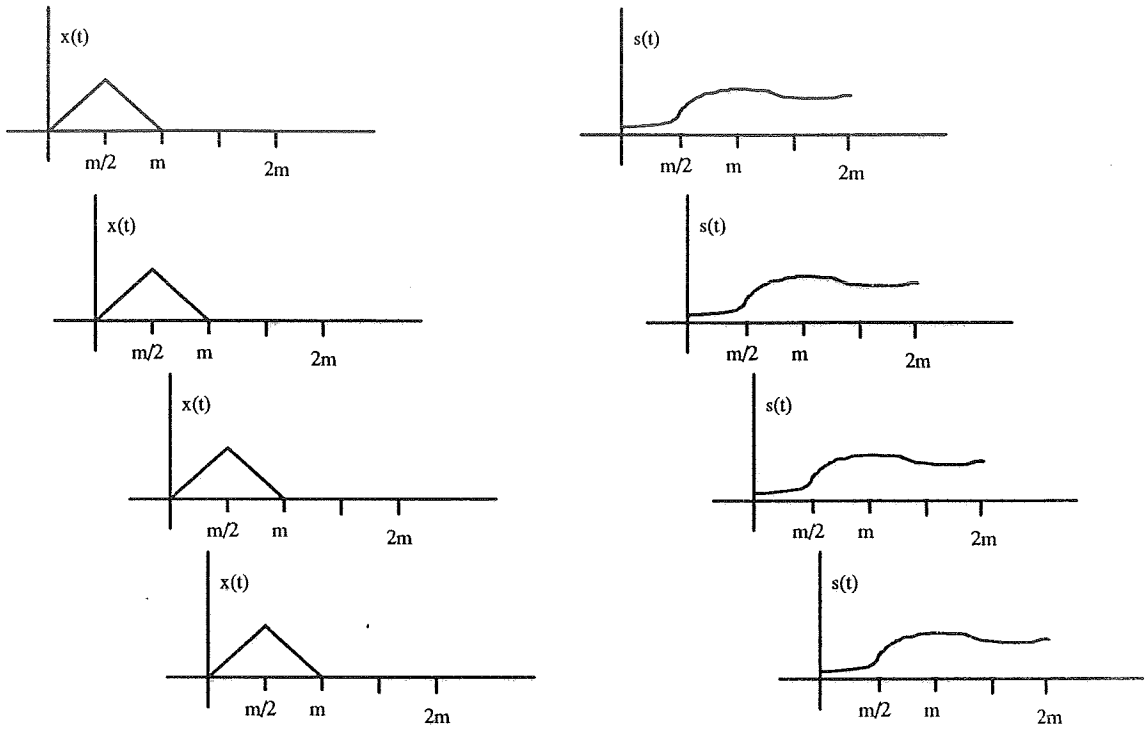


Figure 3. Windowing, Zero Padding, and Signal Recombination

$$\hat{S}(f) = \left| \hat{S}(f) \right| e^{j\theta x} = |H(f)|\,|X(f)|\,e^{j\theta x} = \left|1 - \frac{|N(f)|}{|X(f)|}\right| |X(f)|\,e^{j\theta x} = \left(|X(f)| - |N(f)|\right) e^{j\theta x} \tag{8}$$

Since the filter is obtained from the corrupted speech and noise it has a length of m points. The length of the time domain response of such a filter will be 2m-1. To eliminate the effects of circular convolution the windowed signal must be zero padded by m points to a total length of 2m points [4].

Since there is a 50 percent overlap in each frame only m/2 points of new information is obtained. But since the response last for 2m points four output frames overlap in time and must be combined to produce the correct output for each frame. This is a variation of the overlap add method described in [5].

3.2 Voice Frame Recognition

Once the signal has been windowed and zero padded the FFT is taken. The resulting magnitude and phase of the signal spectrum are determined. The phase is set aside for recombination with the spectral subtracted magnitude. The magnitude of the signal spectrum is used to determine if

the frame contains voice or is voice free. This is done by comparing the maximum value of the signal magnitude spectrum with a proportion, γ, of the maximum value of the average noise magnitude spectrum.

$$\text{if } \max(|X(kf)|) > \gamma \max(\overline{|N(kf)|}) \text{ for } k = 1,\ldots,m \text{ then consider frame to be voiced} \tag{9}$$

The proportion, γ, can be initialized by comparing the maximum magnitude of a known voice frame to the maximum magnitude of the average noise.

The average magnitude spectrum for the noise is obtained by the following procedure: When the algorithm is first being initialized an initial noise only sequence of frames must be obtained to get a baseline on the average magnitude spectrum of the noise.
For frame 1 of the initial noise only sequence:

$$\overline{|N(kf)|} = |X(kf)| \quad \text{for } k = 1,\ldots, m \tag{10}$$

for other frames of the initial noise only sequence:

$$\overline{|N(kf)|} = \delta \overline{|N(kf)|} + (1 - \delta)|X(kf)| \quad \text{for } k = 1,\ldots, m \tag{11}$$

where $0.70 \le \delta \le 0.95$.
Once the initial average noise estimate is obtained, from a known noise only test sequence, each frame of signal is checked for voice using equation (9). If equation (9) is not satisfied the frame is considered unvoiced and equation (11) is used with a predetermined value for δ that is in the specified range. In general δ determines how quickly the noise estimate can vary. The technique is simple, but works well, since voice frames are generally strong in specific frequencies due to excitation of the vocal cords. A more computationally intensive algorithm is used in [6].

3.3 Adaptive Signal to Noise Ratio Dependent Spectral Subtraction Algorithm

After the average noise magnitude spectrum is updated according to the rules outlined in Section 3.2, the magnitude spectrum of the signal and the average noise magnitude spectrum are used to perform spectral subtraction. The signal to noise ratio dependent proportion, α, is determined using the following equation:

$$\alpha = \frac{\eta \sum_{k=1}^{m} \overline{|N(kf)|}}{\sum_{k=1}^{m} |X(kf)|} \tag{12}$$

When the algorithm is first initialized η is determined by testing a signal frame that is known to contain voice. η is chosen such that α is approximately 1.78 in the voiced frames. Once α is determined spectral subtraction is performed using

$$\left| \hat{S}(kf) \right| = |X(kf)| - \alpha \overline{|N(kf)|} \qquad \text{for } k = 1,\ldots,2m \tag{13}$$

If any of the estimates for $\left| \hat{S}(kf) \right|$ are negative, they are set to zero. $\left| \hat{S}(kf) \right|$ is then lowpass filtered eliminate musical noise which is generally high frequency. The lower the 3dB frequency of the filter the more noise and speech eliminated. Depending on the results achieved from spectral subtraction this step may not provide appreciable improvement. Furthermore, the smoothing filter discussed in Section 4 will achieve some of the same results. If calculation time becomes an issue for real time implementation the lowpass filter can be omitted from the algorithm.. After lowpass filtering, the phase of the noise corrupted speech, θ_x, is combined with the magnitude of the estimate of the speech and the inverse FFT is taken. This provides one of the four offset output frames that must be combined using the overlap add method described in Section 3.1. The summing provides an averaging effect for phase errors which reduces the error. But, since each frame is filtered by a different transfer function, see equation (8), summing frames also produces discontinuities in the response causing musical noise.

3.4 Low Level Signal Squelching

The low level signal squelching algorithm looks at three frames of estimated speech: the past, present, and future frames. Future frame estimates of speech are obtained by delaying the speech estimate for one frame before being output. Thus, the signal to noise ratio dependent spectral subtraction algorithm is actually calculating the future output, while the present output is being held in a buffer to determine if low level squelching is required, and the past frame is being output through the D/A. The algorithm is described by the following equation:

$$\begin{aligned} &\text{if} \quad \left|\hat{S}(kf,i)\right| < \mu \ \max(\overline{|N(kf,L)|}) \qquad \text{for } k = 1,\ldots,m/2, \quad \text{and } i = L-1,\ L,\ L+1 \\ &\text{then} \ \left|\hat{S}(kf,L)\right| = 0 \quad \text{for } k = 1,\ldots,m/2 \end{aligned} \tag{14}$$

where μ is a user discretion proportion. A frequency domain squelching technique is given in [7].

4. Overview

A block diagram of the adaptive noise suppression system is shown in Figure 4. Noise or noise corrupted speech enters the microphone. A high gain amplifier is used to bring the voiced signal up to the ±2.5 Volt range used by the Analog to Digital (A/D) converter. Before entering the A/D converter the signal passes through an anti aliasing lowpass filter with 3dB attenuation at 3KHz and 30dB attenuation at 5.9KHz. It is then sampled by the A/D converter using 12 bit

resolution and a 14.925KHz sampling rate. At this point the Digital Signal Processor (DSP) performs noise suppression using signal to noise ratio dependent adaptive spectral subtraction. Next, the digital signal is converted back to an analog signal at a rate of 14.925KHz using the Digital to Analog converter. It is then sent through a smoothing filter, which for the data obtained in the testing was a lowpass Bessel filter with a 3dB frequency of 3KHz. This can be replaced with a voice band filter, which is a bandpass filter with low and high 3dB passband frequencies of 300 and 3KHz respectively. If the voice band filter does not have good damping characteristics the smoothing filter is necessary or transients will be produced from the step discontinuities resulting from the D/A conversion. After the voice band filter the signal is modulated and transmitted by the communication device.

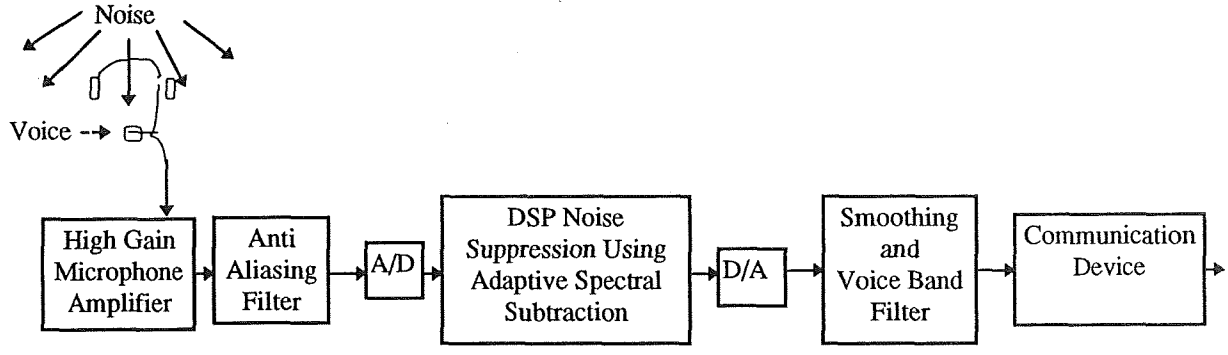


Figure 4. Adaptive Noise Suppression Block Diagram

5. Application

The emergency egress vehicle is shown in Figure 5. Basically it is an M113 tank, which is used to

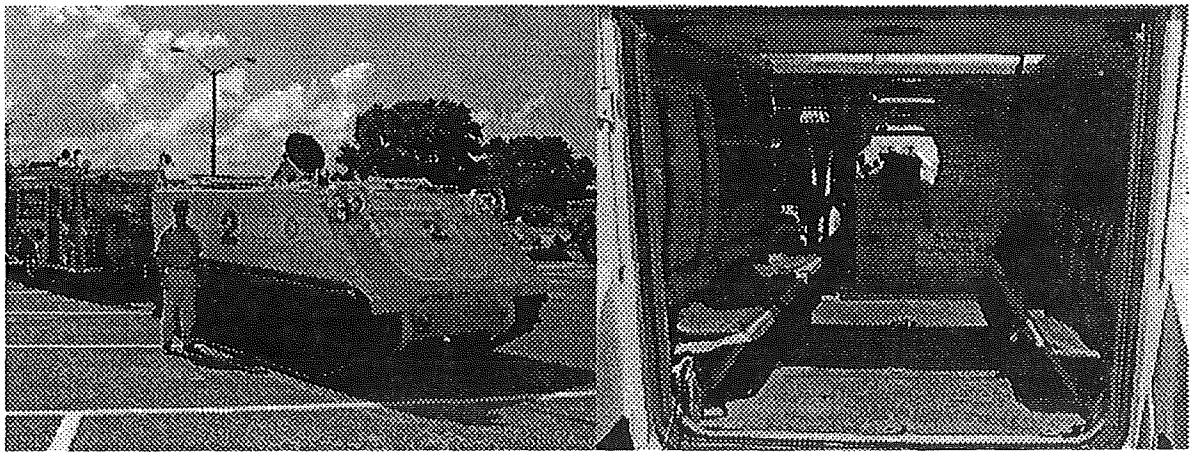


Figure 5. Exterior and Interior of Emergency Egress Vehicle

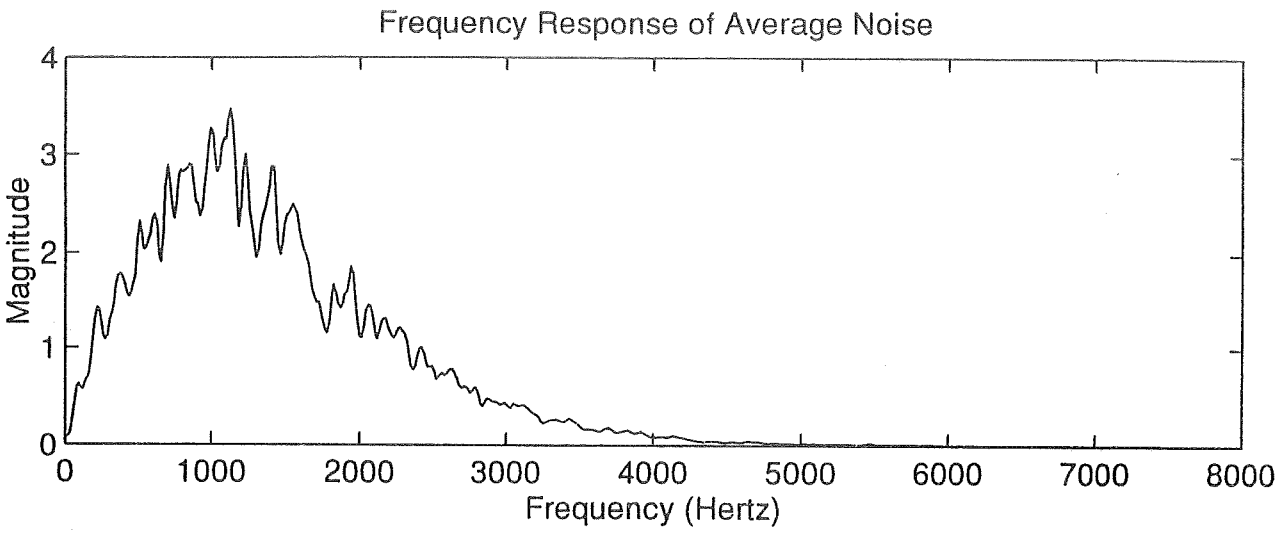


Figure 6. Frequency Response of the Average Noise

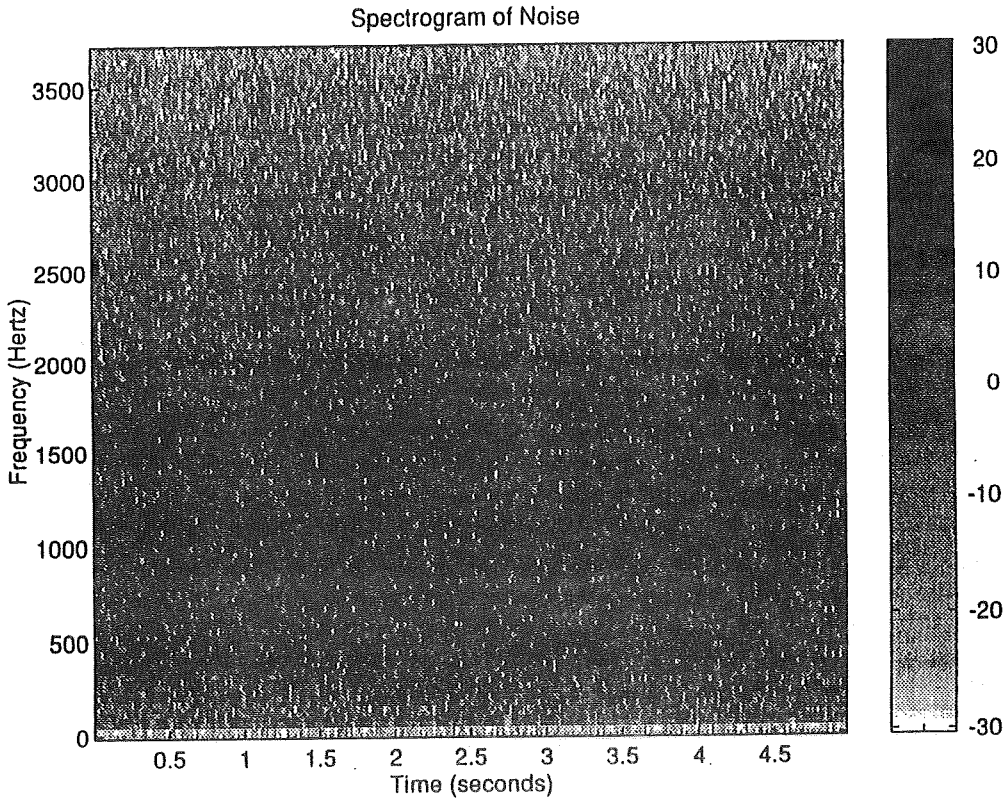


Figure 7. Spectrogram of the Noise

evacuate the astronauts if an emergency situation arises during launch. The noise level inside the vehicle is 90 decibels with the engine running and 120-125 decibels once the vehicle starts moving. As a result, it is impossible to hear what the M113 crew is saying during a rescue operation. The headsets used by the rescue crew have OIS-D microphones which have noise suppression of a mechanical nature, which provides 15 decibels of noise suppression. Furthermore, the frequency response of the microphone attenuates frequencies outside of the voice band range of 300Hz to 3kHz. The frequency response and spectrogram of the noise as obtained through the microphone are shown in Figure 6 and Figure 7 respectively. The decibel scale to the right in Figure 7 is determined using

$$|N_{dB}(f)| = 20\log_{10}|N(f)| \tag{15}$$

Where $|N(f)|$ is the FFT of the noise n(t) obtained from the A/D converter. Thus the spectrogram levels are relative to the voltage of the A/D, not absolute voice levels.

From Figure 6 it is apparent that the noise which is input by the microphone is directly in the range of voice band frequencies. Thus using standard filtering techniques to attenuate it will also attenuate speech by the same factor. The spectrogram of the noise shown in Figure 7 shows that the noise is not constant. As each track of the M113 tank hits the ground, the reaction force causes an impulse on the tank which excites its resonant frequencies.

6 Test Results

The signal to noise ratio dependent adaptive spectral subtraction algorithm was tested on the emergency egress vehicle shown in Figure 5 using the following parameter settings:

$m = 512$

$\gamma = 2.0$

$\delta = 0.90$

$\eta = 4.0$

$\mu = 0.025$

The words "test, one, two, three, four, five" were spoken into the microphone. The original sampled signal, signal after spectral subtraction, and signal after squelching are displayed in Figure 8. Spectrograms for the same three conditions are shown in Figure 9. It can be seen from Figures 8 and 9 that a signal to noise ratio of approximately 15dB exist for the original sampled signal. As mentioned in Section 5 the microphones provided approximately 15dB of noise attenuation. This provided a favorable signal to noise ratio, which is required for spectral subtraction to work well. Lowering the gain and talking louder also improved the signal to noise ratio without saturating the voltage limits of the A/D converter. It can be seen from Figures 8 and 9 that spectral subtraction provided approximately 20dB of improvement in the signal to noise ratio. Listening test verified that the noise was virtually eliminated, with little or no distortion due to musical noise.

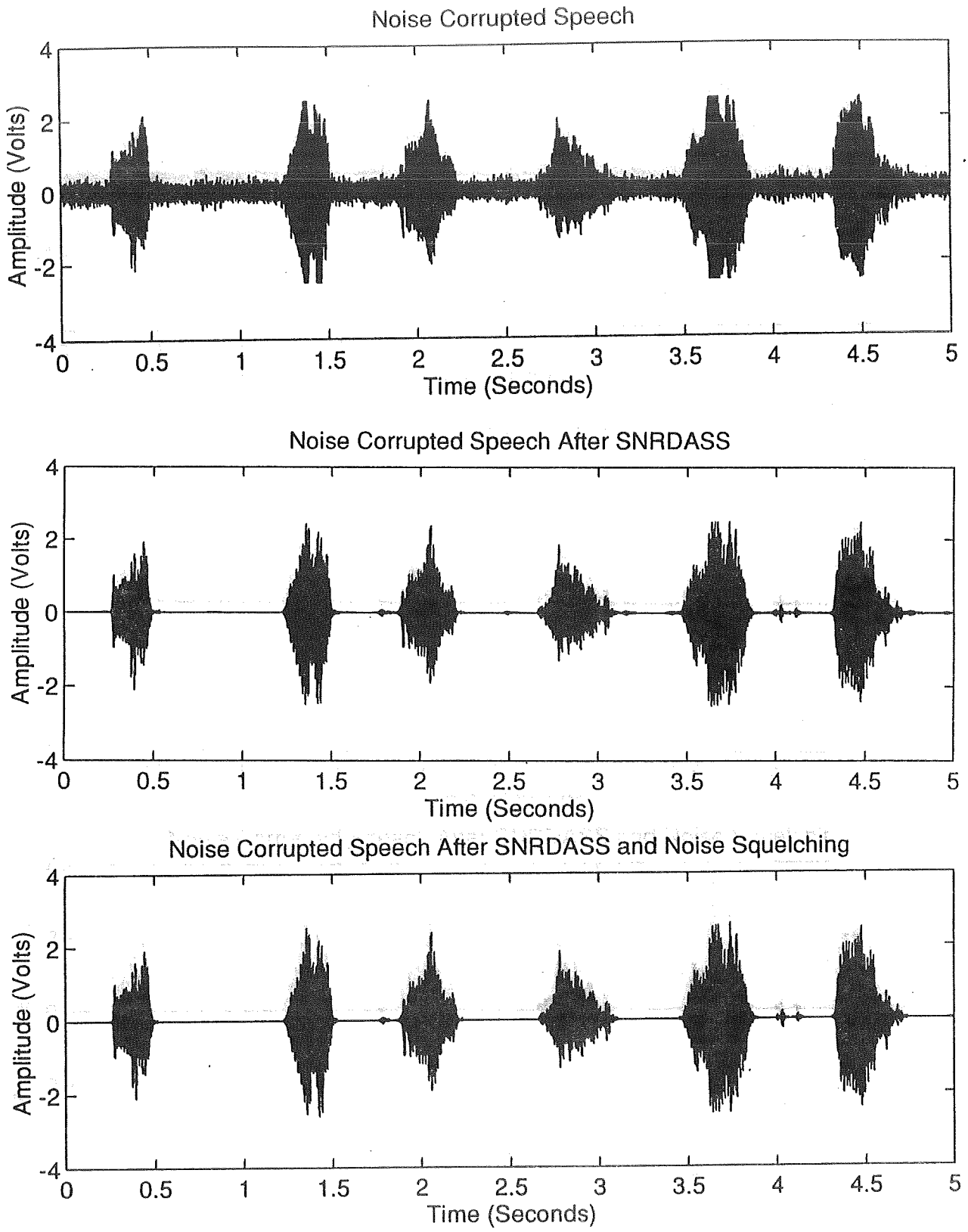


Figure 8. Time Domain Results

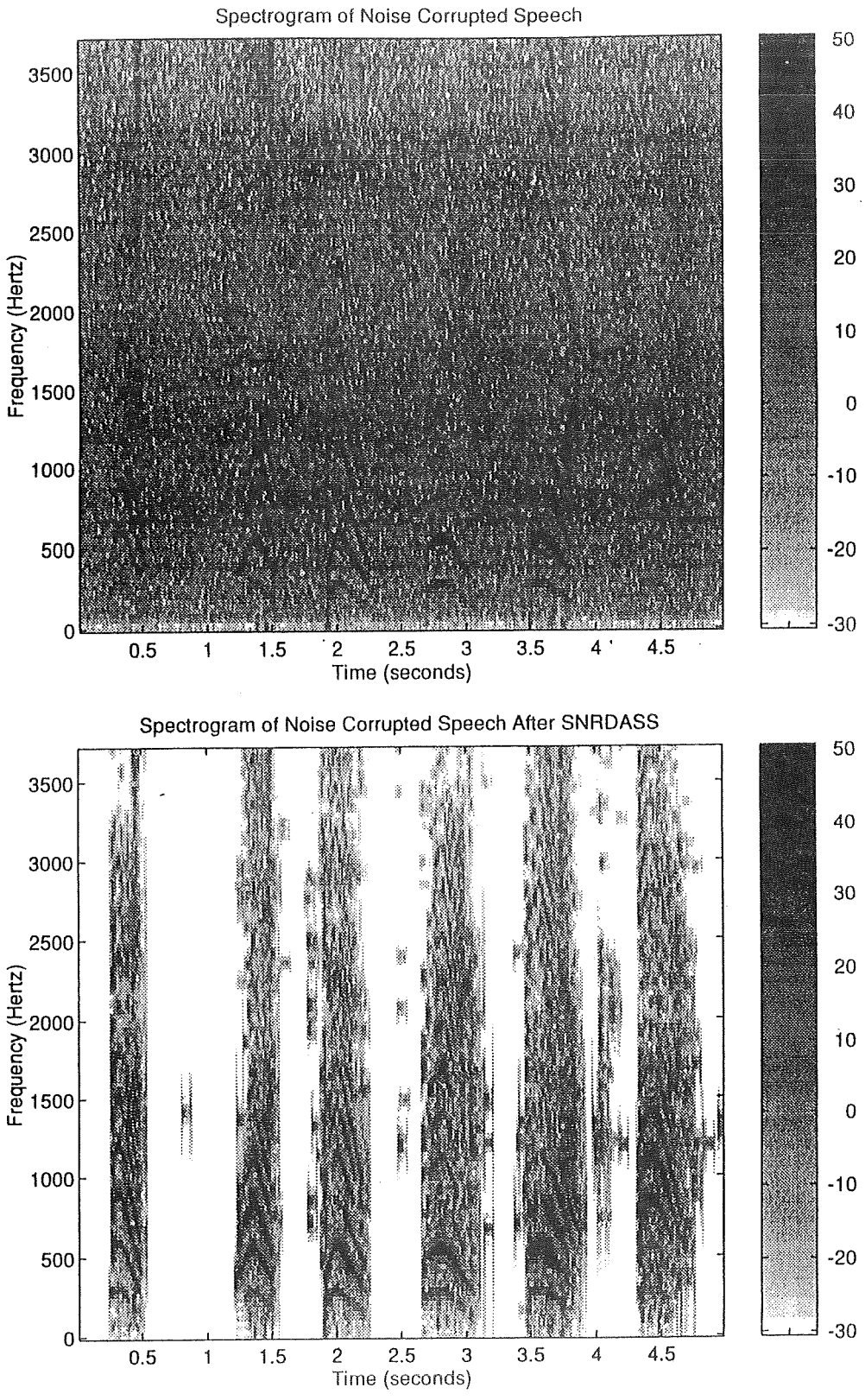


Figure 9. Spectrograms of the Results

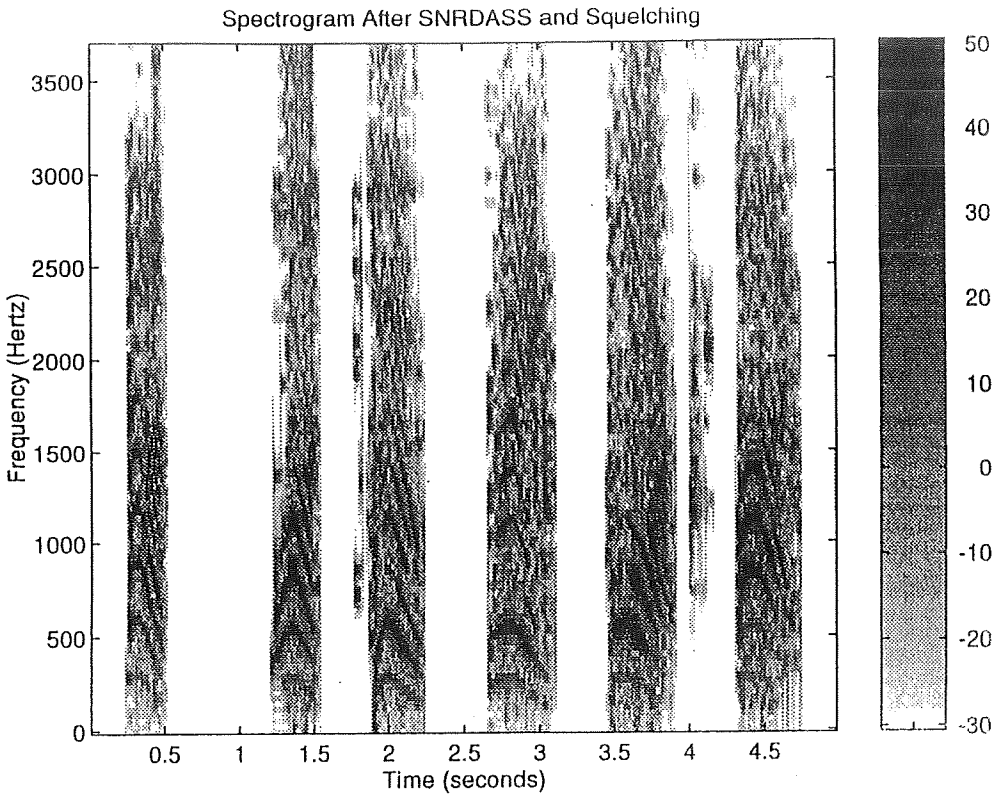


Figure 9 Continued. Spectrograms of the Results

References

[1] McOlash, Scott M., Niederjohn, Russell J., and Heinen, James A., A Spectral Subtraction Method for the Enhancement of Speech Corrupted by Non-White, Non-Stationary Noise, Proc. of the 1995 IEEE 21st International Conference on Industrial Electronics, Control, and Instrumentation, v2, pt2, Nov 6-10, pg 872-877.

[2] Milner, B. P., and Vaseghi, S. V., Comparison of Some Noise-Compensation Methods for Speech Recognition in Adverse Environments, IEE Proceedings: Vision, Image, and Signal Processing, v141, Oct. 5,1994, p280-288.

[3] Arslan, Levent, McCree, Alan, and Viswanathan, Vishu, New Methods for Adaptive Noise Suppression, Proc. of the 1995 20th International Conference on Acoustics, Speech, and Signal Processing, v1, pt1, May 12, 1995, pg 812-815.

[4] Cunningham, Edward P., Digital Filtering: An Introduction,Houghton Mifflin Co., Boston, MA, 1992.

[5] Proakis, John G., and Manolakis, Dimitris G., Digital Signal Processing Principles, Algorithms, and Applications, Second Edition, Macmillan Publishing Co., New York, NY, 1992.

[6] Hirsch, H. G., and Ehrlicher, C., Noise Estimation for Robust Speech Recognition, Proc. of the 1995 20th International Conference on Acoustics, Speech, and Signal Processing, v1, pt1, May 12, 1995, pg 153-156.

[7] Ching, Wee-Soon, and Toh, Peng-Seng, Enhancement of Speech Signal Corrupted By High Acoustic Noise, Proc. of the 1993 IEEE Region 10 Conf. on Computers Communications, Control, & Power Engineering, pt2, Oct 19-21, 1993, pg 1114-1117.

1996 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

A MODEL FOR NASA-KSC's PRIVATIZATION TRANSITION

Dr. Jerome P. Lavelle, Assistant Professor
Mr. Dennis W. Krumwiede, Graduate Student
Industrial and Manufacturing Systems Engineering Department
Kansas State University
Manhattan, Kansas

KSC Colleague - Jean Flowers
Industrial Engineering

Contract Number NASA-NGT10-52605

July 26, 1996

ACKNOWLEDGMENTS

The authors wish to express their deepest thanks and appreciation to all those who have made this summer fellowship a sublime and vastly rewarding experience. We earnestly have been invigorated by the wonderful people at KSC and feel confident that they are motivated and committed to ensuring that at NASA-KSC the space program will remain a world-class operation.

Thanks go to many people; we would like to name them explicitly. First, to the folks in the Industrial Engineering group at KSC, our sponsors, we wish to say thank you for the sedulous professionalism and positive attitude that you bring to your work: to Jean Flowers, Amanda Mitskevich, Colleen Orr, Tim Barth, and Tim O'Brien, thanks guys and best of luck this next year! To their bosses, Bob Sieck, Larry Ellis and Eric Redding, a big thanks for your time, energy, and encouragement. Also to the NASA/ASEE folks, Roger Johnson, Gregg Buckingham and Kari Stiles, we wish to express our deepest appreciation for the effort you put into developing and sustaining this excellent opportunity for ordinary people like us. Thanks too to other folks that we've enjoyed meeting and working with this summer: Pepper Philips, Nicole Passonno, Bill Carew, Scott Cilento, Jenny Lyons, Louie Garcia, Carlos Estrada, Mary Scholtz, Elaine Liston, Beverly Bush and Terrie Bunch

Many thanks to all of you again for helping to make this summer both enjoyable and professionally interesting and expansive. Best of everything to all!

ABSTRACT

This paper reports on the results of the authors' 1996 NASA/ASEE Summer Faculty Fellowship research. The authors have developed a model for government agencies that are considering privatization of all or part of their functions. Privatization encompasses the transitioning of government functions from "government run" to "contractor run". The model developed in this paper is used to analyze the National Aeronautics and Space Administration's (NASA's) decision to privatize space shuttle operations at the Kennedy Space Center (KSC). Several specific recommendations are given to KSC as they attempt to operationalize this privatization decision at the Center and to transition to a new relationship with their contractors.

A MODEL FOR NASA-KSC's PRIVATIZATION TRANSITION

Dr. Jerome P. Lavelle and Mr. Dennis W. Krumwiede

1. INTRODUCTION

The National Aeronautics and Space Administration (NASA), like many other United States government agencies, is undergoing a transition toward privatization of all or some agency functions. The process of transitioning toward privatization requires a systematic methodological approach. Given is a methodology for use at NASA-KSC as it transitions from its current and traditional role, with respect to shuttle processing and operations, to the use of contractor performance-based metrics.

2. BACKGROUND

Voters today are demanding a more lean and responsive government, and politicians and policy makers have taken notice. Concepts like "re-inventing government" which stress efficiency and decision making accountability in the public arena have set the stage for what people expect of their government. Voters are looking for reduced government size, fiscal responsibility through balanced budgets and privatization where practical and possible. The National Performance Review (NPR) is a reaction to that demand; it brings federal government employees together to work toward: eliminating red tape, putting the customer first, empowering employees, and cutting back to return to the basics (Government Reform, 1995). Governments, as with corporate America, are being asked to do more with less and do it better each year. Government branches and agencies are challenged to establish tangible strategic plans and justify all activities (and expenses) in light of stated goals. The Government Performance and Results Act of 1993 mandates that all U.S. federal agencies use a strategic planning process as a core tool in establishing work plans and budgets and in scoping resource needs and deliverables to the customer (the voter).

With this in mind, NASA has been very active in establishing a vibrant strategic planning process by promoting planning at the Agency, Center and sub-levels. KSC's first formal Strategic Plan was developed in the fall of 1987 and has continued to successfully evolve since that time. In the short term (the 1996-2002 time frame), NASA's strategic goal to *Revolutionize NASA* challenges NASA to consider the way it looks at itself and its customers (NASA Strategic Plan, 1996). This plan calls for continued excellence in research and development in aerospace technologies and to begin privatizing and de-emphasizing operations in areas where technologies are now matured or maturing.

Privatization is a method used by governments to obtain high quality goods and services at lower costs and on a more timely basis through the transfer of programs and functions to the private sector. Privatizers may include private citizens, businesses, and organizations (Kent, 1987; Brinkley, 1987). According to Dimeo (1991) there are three basic types of privatization:

- Contract services to private companies, paying the firms a fee for the work.
- The formation of a partnership with private companies, which will build facilities and operate the services for the government .
- The sale of public assets to the private sector.

NASA-KSC's past and present space shuttle operations transition falls into the first category of privatization given above.

A few of the Agency's strategic objectives that highlight the changing world for KSC are:

- Return NASA to a premier R&D agency.
- Do things that no one else can.
- Change the way that we work with contractors.
- Measure performance and communicate results to demonstrate relevance and contribution to national needs. (NASA Strategic Plan, 1996)
- Decrease Space Shuttle costs and improve the management and operations of the integrated government/contractor team. (HEDS Strategic Plan ,1996)

As of the launch of STS-78 on June 10, 1996 there have been over seventy successful launch and recoveries of NASA's space shuttle orbiter and payload. Clearly the processing of the orbiter and its payload is a mature technology, and, per the strategic goals, a valid candidate for a transition toward privatization.

3. PRIVATIZATION MODEL

Privatization requires a systematic methodology. A general model is suggested by the authors that addresses all phases of the privatization process. This model is given in Table 1 below. It combines the author's personal observations of the process at KSC as well as the United Kingdom Methodology from Miller (1993), and the Air Force Methodology from McSwain and Smith (1989).

Per the model, and with respect to privatization of space shuttle operations, the Agency has been successful in defining privatization goals and in identifying the Centers affected by those goals. The Kennedy Space Center is the NASA entity primarily affected by the privatization of space shuttle operations. The model emphasizes the importance of including Center-level input in the decision making process. NASA-KSC has been empowered to plan for and operationalize the privatization plan. The Center has been given latitude in terms of the time and process to transition from its current role to the privatized role. The model also points to the importance of a formal structure to evaluate privatization and to incorporate lessons and learning into the decision making process at the Agency and Center levels.

4. OPERATIONALIZATION OF PRIVATIZATION AT KSC

Step five of the General Agency-Level Model in Table 1 affects KSC specifically, with respect to the privatization of space shuttle operations. Given the decision to proceed with privatizing this function, the task falls on NASA-KSC to carry out that effort. In Table 2 a specific methodology to operationalize privatizing shuttle operations at KSC is given.

The method in Table 2 requires, as a first step, a clear definition of the goals and objectives of the privatization from the perspective of the Center. It is important that these are aligned with, or result from, privatization goals of the Agency. As with concurrent engineering in product design, it is important to have all relevant impacts planned for at the beginning stages of the process because such efforts affect 95% of what follows in the product life-cycle. This is why it is important to: have goals and objectives established from the beginning of the privatization initiative, have these goals established by top management, and have them communicated downward to those involved in actualizing the goals and objectives of the privatization.

Table 1: General Agency-Level Model for Privatization

1. Define Agency-Level Goals and Objectives With Respect to Privatization • Define through strategic planning process. • Communicate throughout the organization.
2. Identify Affected Centers of Potential Privatized Functions • Communicate Agency-level privatization goals and objectives.
3. Evaluate Privatization Alternatives at Agency and/or Center Level • Include Centers in the evaluation. • Identify potential functions to privatize. • Specify the type of privatization. • Perform cost and risk analyses on the Agency and Center(s). • Evaluate alignment with, or affect on, the strategic plan. • Understand the political implications.
4. If Privatization Chosen, Communicate • Insure communication methods saturate the organization. • Relate current and potential plans through communication channels. • Use complete and honest tones in communication. • Relate updates and future communications.
5. Empower Centers to Plan and Operationalize Privatization • Ensure Center-level privatization strategic plans align with Agency strategic plans.
6. Evaluate Effectiveness of all Privatization Efforts at Agency and Center Levels • Measure effectiveness against expected goals and objectives over time. • Decide to continue present effort, add to, or eliminate privatization portions.
7. Document Lessons and Learning at Agency and Center Levels • Document inputs, processes and outputs throughout. • Incorporate lessons and learning into the process of evaluation and execution.

Table 2: Methodology to Operationalize Privatization at KSC

Overall Methodology
1. Center top management should ensure that specific goals and objectives of privatization at the Center are aligned with Agency privatization strategic plans.
2. Develop a set of contract measures that incentivize the contractor to perform in alignment with the goals and objectives of privatization. These are the measures that are used for making contract evaluations and are explicitly incentivized.
3. Develop an exhaustive list of all processes involved in meeting the goals and objectives of the privatization.
4. In a matrix format, map the key processes from the exhaustive list to the contract measures that are incentivized explicitly through the contract. These key processes are insight measures that can be aggregated (either 1-to-1 or many-to-1) to the contract measurement level. These measures provide insight and confidence; they are not incentivized directly.
5. Do not measure or track other processes that are not primary in supporting the goals — these processes provide no insight or confidence, and are not incentivized through the contract.
6. Establish a system, for both the contract measures and the insight measures, to track, report, and manage those measures.

Once privatization goals have been defined a set of contract measures that explicitly support these goals should be developed. It is important that contract measures are incentivized in the contract. These contract measures should lead the contractor to perform with respect to them, thus performing with respect to the overall goals. A lack of goals makes it impossible to develop contract measures that promote contract performance in alignment with those goals.

The next step in the Methodology to Operationalize Privatization at KSC is to exhaustively identify all processes associated with the privatization, on both the NASA-KSC and contractor side, and list them explicitly. These processes then should be mapped, in a matrix format, to the contract measures developed previously. The identification of the processes, as well as the mapping of those processes to contract measures involves both NASA-KSC and contractor process owners. Mapping the processes to contract measures involves identifying those processes that associate to specific contract measures. These may associate in a one-to-one or a many-to-one fashion. However, these processes are not incentivized directly; their role is to both provide insight into the contractor's processes and to aggregate forward to form the contract measures. In mapping processes to contract measures not all processes may be accounted for. From a measurement perspective, it is important to track only those processes that either provide insight or are at the contract measurement level. Processes that are not insight or contract measures do not explicitly relate to the goals and thus do not require measurement.

The last task in the Methodology to Operationalize Privatization at KSC is to establish a standardized system of obtaining measurement data and in analyzing and reporting that data in a bi-directional fashion throughout the organization. As one defines the contract and insight measures described in the previous paragraphs (and tracks those over time to form a system of metrics), it is important to use a formal operational definition of each measure. This operational definition is the who, what, when, where, why, and how of each measurement.

5. CONCLUSIONS

The following are a series of conclusions and recommendations developed by the authors. These recommendations are based on the General Agency-Level Model for Privatization and Methodology to Operationalize Privatization at KSC described in the preceding sections and impressions of activities at NASA-KSC reached over the tenure of the 1996 NASA/ASEE Summer Faculty Fellowship assignment.

- At KSC it is important for the active measures-and-metrics groups to continue to discover, from the ground up, the goals and objectives that the SFOC promotes. Having understood these fundamentals, a system of measurements can be developed based on the methodology described above. If there is an apparent inconsistency between what is being understood as the goals and what the SFOC incentivizes the contractor to do an adjustment to one of these factors is necessary. As a first step the authors suggest formal training on the SFOC to at least the measures-and-metrics groups.
- The measures-and-metrics groups should utilize a standard set of resources, definitions and analysis methods as part of their system development work. Resources of interest may include: NASA Contractor Metrics Handbook, Measurement Planning Handbook and Measurement Workbook, Air Force Metrics Handbook, and Air War College for Developing a Successful Privatization Project.
- The measures-and-metrics groups should use a local benchmark to learn lessons and gain insight into the metrics development process. The McDonnell Douglas company at KSC

has extensive experience in developing metrics systems. Ms. Sherry Smith is the lead engineer in this area for the company.

- When goals are communicated there needs to be a repository for accessing those goals by all levels of the organization. Difficulty in obtaining information pertaining to those goals leads to misguided and non-uniform approaches to problem solving.
- Continuous improvement (CI) is a "way of life" that proactive organizations pursue, and the goal of promoting CI as a principle of operation is appropriate for such organizations. Given the organizational challenges that face NASA in the future, the goal of promoting CI seems appropriate. If this is a goal (to promote this in the contractor), it must be incentivized in the SFOC explicitly.
- The Goals-Contract-Measures-Metrics integrated system should be documented, communicated, and driven at all levels affected. It should serve as a common "language".
- In developing a system to track, report and manage contractor metrics, if the goal of promoting the CI philosophy in the contractor is chosen, then the use of statistical process control methods seem appropriate. See Kinlaw (1993) and the Air Force Metrics Handbook (1991). In this context, training session on SPC and other CI tools seems appropriate.
- It is important to track the progress of the privatization effort from cost and customer satisfaction perspectives. If either attribute fails to achieve levels consistent with Agency, or Center goals, privatization should be re-evaluated.
- The technical discipline of industrial engineering provides expertise in methods and process analysis for organizational efficiency and effectiveness. In the privatization effort at KSC, with its resulting need to operationalize changes in the organization, industrial engineering expertise should be emphasized. The industrial engineering group should be empowered to lead the engineering of the required changes. That role should be understood and communicated throughout the Center. To be effective in this role it is important to place such expertise at appropriate organizational levels.

6. REFERENCES

1996 Application for the President's Quality Award Program, National Aeronautics and Space Administration, John F. Kennedy Space Center, September 1995.

Air Force Systems Command, 1991. Metrics Handbook. U.S. Air Force Systems Command. Reproduced by U.S. Department of Commerce National Technical Information Services. 1st Edition. No. PB92-16643.

Bancroft, T., 1991. Pravatization? Forbes. June 24, pp 44.

Binkley, J., 1987. Reagan Appoints Privatization Unit. New York Times. September 4, p. 28.

Carter, F. G., 1991. Partial Privatization in a small city fleet. Public Works. pp 38-39.

Carter, J., Are You Using the Right Metrics for Continuous Improvement?, Production, September 1994, p. 16-17.

Conner, J., 1995. Privatization seen successful for firm set up by Congress. Wall Street Journal, Public Utilities Fortnightly. Eastern Edition. June 1, pp 38-42.

Deming, W.E., 1986. Out of Crisis, Cambridge, MA. p. 276.

Dimeo, J., 1991. Can privatization help stretch the local dollar? American City & County. September, pp 26-29.

Lessons in Privatization. Euromoney. February 1996, pp. 93-102.

ENR, 1992. Sydney privatization targets water supply. ENR. February 24, 1992.

Government Reform: GAO's Comments on the National Performance Review, US General Accounting Office, Washington, DC. June 27, 1995.

HEDS Strategic Plan, NASA's Enterprise for the Human Exploration and Development of Space, January 1996.

Hewlett, J. G. , 1990. Lessons from the attempted privatization of nuclear power in the United Kingdom. Energy Sources. 16, pp 17-37.

Kent, C. A., 1987. Entrepreneurship and the Privatizing of Government. Greenwood Press, Inc., Westport, Conn.

Kiernan, V., 1995. Hello, this is mission control, how can I help you? New Scientist. 145 (Mar. 25), p 6.

Kinlaw, D. C., 1993. Measurement Planning Handbook and Measurement Workbook. Kinlaw Associates. Norfolk, VA.

McSwain, T C. and Smith, W. E.,1989. Criteria for Developing a Successful Privatization Project, Air War College Research Report, AD-A217 523, Air University United States Air Force, Maxwell Air force Base, Alabama.

Miller, A. N., 1994. Privatization: Lessons from the British experience. Long Range Planning. 27 (6), pp 125-136.

Morris, J., 1987. Privatization and the Unions. American City and County. July, pp 68-72.

Morrocco, J. D., 1993. Navy shifting depot work to private industry. Aviation Week & Space Technology. June 28, pp 28-30.

New Zealand, 1992. New Zealand expands service while adjusting to privatization. Aviation Week & Space Technology. February, p 55.

NASA Strategic Plan, February 1996, National Aeronautics and Space Administration, Washington, DC, 20546

Norman, J. R., 1991. Unforeseen consequences. Forbes. September 16, pp 150-152.

O'Conner, R., 1991. Going private. International View. November/December, pp 30-66.

Peratta, E., 1995. Despite bumps in the road, privatization races on. American City and County. October, pp 50-61.

Prager, J., 1994. Contracting out government services: Lessons from the private sector. Public Administration Review. March/April Vol. 54 (2) pp 176-184.

Proctor, P., 1992. Privatization, New Sydney Runway boost Australian airport system. Aviation Week & Space Technology. January, pp 40-41.

Reinhardt, W. G., 1988. Privatization focus shifts to feds as practitioners bare shortcomings. ENR. January 28, pp 8-9.

Reinhardt, W. G., 1988. Defense Department warming up to public/private ventures. Public Works Financing. April, pp 6-12.

Rubin, D. K., and Karmer, D., 1995. DOE sees life with less. ENR. May 8, p 16.

SFOC, NonCompetitive, Draft Copy, 12-07-95, RFP 9-BV2-32-6-1P.

Sakita, M., 1989. Restructuring of the Japanese National Railways: review and analysis. Transportation Quarterly. 43, pp 29-45.

Schmeidler, N. F., 1996. Organization Performance Measurement. Lea Edwards Associates.

Tokyo, 1989. Deregulation, privatization spur to Diversify operations. Aviation Week & Space Technology. May, pp 42-43.

Wessel, R. H., 1995. Privatization in the United States. Business Economics. October, pp 45-50.

Zolkus, R. 1995. Huge privatization yields risk management challenge. Business Insurance. October 23.

1996 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

SCHEDULING SYSTEM ASSESSMENT, AND DEVELOPMENT AND ENHANCEMENT OF RE-ENGINEERED VERSION OF GPSS

Dr. Rasiah Loganantharaj, Associate Professor
Mr. Bushrod Thomas, Graduate Student
Center for Advanced Computer Studies
University of Southwestern Louisiana
Lafayette, Louisiana

KSC Colleague - Nicole Passonno
Ground Systems

Contract Number NASA-NGT10-52605

August 8, 1996

ABSTRACT

Scheduling of activities that maintain, repair and configure a space shuttle orbiter must satisfy temporal, resource and configuration constraints. The GPSS, which stands for ground processing scheduling system, has been successfully used to schedule tasks that prepare an orbiter for its next launch. The features of the GPSS that support specification of a task with configuration requirements and effects is what makes the GPSS unique when compared to other scheduling software.

The objective of this project is two-fold. First to provide an evaluation of a commercially developed version of GPSS for its applicability to the KSC ground processing problem. Second, to work with KSC GPSS development team and provide enhancement to the existing software.

The Red Pepper Software (RPS) company has developed a commercial scheduling product based on the original algorithm of GPSS. The RPS team consists of many of the original members of the NASA Ames, KSC and Lockheed development. The RPS system was installed and an initial evaluation regarding its suitability for solving the ground processing scheduling problem was performed.

The GPSS in use at KSC is written in Lisp and C and has very little or no documentation available regarding the development process, or software design and algorithms. System re-engineering is required to provide a sustainable system for the users and software maintenance group. As such, the USA development team is developing a re-engineered version of GPSS in C++. The GPSS uses an iterative repair method for scheduling. The algorithm first satisfy temporal constraints among the tasks, and then it iteratively repair the conflicts associated with resource and configuration constraints. Resolving the conflicts is a difficult and time consuming process for the algorithm. The deconfliction process of GPSS is a hill climbing method coupled with a weak simulated annealing implementation. The temperature for the annealing process is fixed at 75, and it is brought down to 25 after achieving some conditions. We propose a formula that determines the initial temperature based on: (1) the initial cost, and (2) the tolerance limit with which one is willing to accept a higher cost in order to avoid local minima. We have also laid down procedures to gradually cool down the temperature as the algorithm improves the solution.

We have implemented the resource deconfliction portion of the GPSS in common LISP using its object oriented features. We have used the LISP profile prototype code, which was developed by the GPSS reverse engineering group, as a building block to our implementation. Our prototype will help the GPSS developers re-engineer the deconfliction portion in C++. It corrects and extends some of the deficiencies of the current production version, plus it uses and builds on the classes from the development team's profile prototype, which they are now porting to C++.

SCHEDULING SYSTEM ASSESSMENT, AND DEVELOPMENT AND ENHANCEMENT OF RE-ENGINEERED VERSION OF GPSS

Rasiah Loganantharaj and Thomas Bushrod

1. Introduction

The objective of this project is two-fold: first, to provide an evaluation of Red Pepper Software's Production Response Agent (PRA), a commercially developed version of the ground processing scheduling system (GPSS), for its applicability to the KSC ground processing problem; second, to work with the KSC GPSS development team and provide enhancement to the existing software. We start with an introduction of scheduling, GPSS and its limitations

Scheduling is a process of assigning time slots for activities while satisfying their resource and configuration requirements, and temporal ordering among themselves. A typical temporal constraint specifies a successor, a predecessor, and possibly a minimum delay between them. For example, a temporal constraint may have a form such as: task T_2 is a successor of a task T_1, and T_2 can start only after 10 units of time from the completion of T_1. Satisfying all the temporal constraints among the tasks provides a schedule with early and late start time of each task, critical path, etc. There are plenty of sophisticated software algorithms to schedule tasks that are constrained only by temporal relations. Satisfying temporal constraints of this sort will take $O(N^2)$ time where N is the number of tasks. When we introduce resource requirements for tasks, the problem can no longer be solved optimally in polynomial time. The requirement that a task have two overhead cranes is an example of a resource constraint. A task can be scheduled only after it is assigned all of its requested resources. If we have an unlimited quantity of resources to the extent of satisfying all the resource requirements, the problem reduces to solving only the temporal constraints. However, in many real world problems, resources are limited. Thus, decisions must be made about the order in which requests are satisfied. An optimal schedule maximizes the resource utilization and minimizes the schedule length.

The complexity of a scheduling problem is further compounded when the problem is extended to include configuration requirements and configuration effects. By *configuration requirement*, we mean a requirement of a task requesting an attribute of an object to be in a specified state. For instance, the precondition that payload-bay doors of a space shuttle orbiter be opened 75 degrees to allow loading of a satellite is an example of a configuration requirement. Here, the attribute is the state of the payload-bay doors (which may have a range of possible values from closed to fully open), and the activity is "load satellite." Similarly, an activity may affect the state of some attribute of an object. According to the law of causality, the effect cannot precede the causing activity, and it must persist until some other subsequent activity changes the attribute. Suppose the bay area is loaded with hazardous material. The area remains hazardous until the material is removed from the bay. This is an example of an effect that persists to infinity. In the presence of tasks that cause changes to attribute values, finding a proper placement for a task with configuration requirements increases the search efforts of the scheduler.

An objective function of a scheduler is to obtain a placement for each activity such that the overall schedule maximizes resource utilization and minimizes schedule length. Except for some trivial cases, obtaining an

optimal schedule that satisfies temporal, configuration, and resource constraints is NP-hard. That is, no polynomial-time algorithm is available to solve these optimization problems. The algorithm that solves the problem optimally will take time on the order $O(2^N)$, where N is the number of tasks to be scheduled. When N is in the order of hundreds, it will take thousands of years to solve the problem, even if we run it on the fastest computer available today. Typical flows at KSC involve a thousand or more tasks. Therefore, a sub-optimal solution is being sought to solve scheduling problems of this sort.

There are two major approaches to obtaining the sub-optimal solution: the constructive method, and the iterative repair method. In the constructive method, one tries to build a feasible solution satisfying temporal, resource, and configuration constraints incrementally. When there is a failure, the algorithm backtracks and a new choice is made. This approach can further be categorized based on how the backtracking is done. The backtracking can vary from simple methods, such as chronological backtracking, to very sophisticated methods, such as intelligent backtracking. Heuristic methods are used to decide how to allocate resources among the competing tasks.

On the other hand, the iterative repair method starts with a rough schedule that satisfies only the temporal constraints. Then it repairs the resource and configuration constraint violations one by one until no more violations exist or no further improvements can be made. The iterative repair method is very suitable to tackling over-constrained problem instances because it will iteratively reduce the number of violations even though there may not be any feasible solutions (solutions satisfying all constraints, though not necessarily optimally) for the problem. If the constructive method is used to solve an over-constrained problem, it will consume a substantial amount of time exploring the entire search space before it eventually fails. Many of the ground processing scheduling problems at KSC are over-constrained, and hence, GPSS uses the iterative repair method to solve the scheduling problem.

1.1 GPSS

GPSS has been used successfully at KSC for scheduling orbiter processing facility (OPF) operations. The OPF processing has three phases: (1) making the orbiter safe for processing and gaining access to the orbiter through the installation of access platforms, (2) testing, maintenance and repair operations on the orbiter, (3) close-out and checkout of orbiter. Approximately 40% of OPF processing is routine and very predictable in advance. The remaining 60% of OPF activities are very dynamic and are driven by factors such as payload of the previous and the following missions, test requirements specific to the age of the orbiter, diagnostic in-flight anomalies, and unexpected damage caused to the insulators.

GPSS provides a graphical interface for specifying tasks. Each task has a duration, work calendar, resource requirements, state requirements, and state effects. Temporal constraints between predecessor-successor pairs are specified. The number of resources available for each resource type is specified. A task may require many resource types, and for each resource type the task can specify the number of resources it needs. GPSS assumes that resources are reusable, that is, once the task is completed, the number of resources used by the task is released and can be used by other tasks. Each attribute's initial state is specified. A task may request an attribute to be in a specified state. When a task is scheduled in a time period during which the attribute has a different value than the one it requested, we say *configuration violation* has occurred.

Given an initial specification of tasks that may include resource and configuration requirements along with temporal constraints, GPSS satisfies their temporal constraints first. Resources are allocated, and a *profile*

(a.k.a. *history*) is maintained for each resource. A resource profile consists of a sequence of intervals along the time line indicating the *users* and *changers* of each interval. In an interval where the total requests of the users exceed the number of available resources, the resource type is said to be *over-allocated* and it is an indication of a *resource constraint violation*. In a system modeling consumable resources, a changer could allow a task to produce or consume some quantity of the resource, but GPSS is not used in that way at present.

When there is a resource violation, a user can either resolve the conflict manually or use the system to deconflict the violation. A user can select a task from among the tasks in violation for the same resource, and move it to another time interval where the resource request of the task can be satisfied. After each move, the system runs the temporal constraint satisfying algorithm and accepts the move only if there exists a way to satisfy all temporal constraints. Otherwise the move is rejected and the previous system status is restored.

Similar to the resource profile, a profile is maintained for each configuration attribute. Each interval of an attribute profile maintains its users and changers. Unlike the reusable resource constraints, users and changers need not be the same. The changers affect the attributes value. When there are many chanagers in an interval, the value of the attribute is determined by the effect of the latest changer. A task is said to be in configuration conflict if the requested attribute does not match with the attribute value in the interval. As with resolving resource conflicts, a user can either resolve the configuration conflict manually or use the system to deconflict the violation.

1.1.1 Limitations of GPSS

While GPSS is successfully used for scheduling activities in each flow, it has some weaknesses in regard to deconfliction. The auto-deconfliction in GPSS is very slow and it lacks some user controllable features such as *fencing* (providing time boundaries on the working schedule within which to select violated constraints to resolve) and *user-fixing* activities (holding selected tasks in their scheduled intervals) during deconfliction.

1.2 Organization of the report

Following the introduction, we briefly summarize the evaluation of Red Pepper Software. In the next section we describe about repairing conflicts and discuss about the enhancement we propose to the GPSS. This is followed by summary and discussion.

2. Evaluation of Red Pepper Software (RPS)

2.1 Installation

We received Production Response Agent software, PRA (v1.5.2), from RPS on a medium density tape. We installed the tape on a Sparc 5 workstation under the operating system Solaris 2.5. The initial installation was not successful because a motif library file (libXm.so.2) was missing in the Solaris 2.5 distribution. We called the RPS hotline (1-800-578-3345) and explained the installation problem. The next day, RPS had set up an ftp access for us to download the missing file. With the missing file in the correct directory we were able to complete the installation of the software and to run the Production Response Agent (PRA) successfully.

2.2 Evaluation of PRA for scheduling purposes

For the purpose of our initial PRA evaluation, we constructed a simple scheduling problem with temporal and resource constraints, ignoring status constraints. We studied the PRA user manual to build a scheduling model to solve the example problem. We were unable to find any materials in the manual that talk about building a model to solve scheduling problems involving temporal and resource constraints. We then examined the PRA startup menu to see if there is a way either to build a model or to set the problem instance. We failed to find a way to enter our simple example into the system.

PRA has to be initiated with a command file which comes with the software distribution. The command file loads different built-in models. It is our understanding that PRA will only work to solve a problem that fits the defined models. The models that came with the distribution do not support the scheduling model that is used by GPSS at NASA KSC.

To confirm our understanding of PRA, we contacted the technical support at RPS. The technical personnel at RPS acknowledged the following concerning the capability of PRA (V1.5.2) towards building a new model:

1) The user manual of PRA does not provide any information on building a new model to solve problems.
2) To learn about building models using PRA, one has to attend a class at RPS on SPL, their proprietary Scheduling Programming Language.
3) During the tutorial classes at RPS, students copy the existing models and modify the parameters and instructions to solve their problem.

PRA provides a graphical user interface to access and to modify the instances defined by the model. The product seems to be very capable and useful for solving problems in high level production planning and scheduling. However, for PRA to be utilized for ground processing scheduling purposes it would have to support temporal, resource, and configuration constraints at the task and sub-task levels.

Since PRA does not support scheduling problems similar to the ones solved by GPSS we did not continue with a through evaluation. Instead, we worked with the GPSS re-engineering group, focusing on enhancing GPSS and building a prototype for constraint violation deconfliction.

3. Development and Enhancement of Re-engineered version of GPSS

3.1 Repairing Conflicts

GPSS is an iterative repair scheduler which first satisfies all the temporal constraints among the tasks and then repairs all the resource and the configuration violations iteratively. The Waltz algorithm, which was originally developed for the purpose of scene analysis, was modified to propagate temporal constraints and to achieve temporal constraint satisfaction. In the process of achieving temporal constraint satisfaction, the Waltz algorithm may move tasks that cause resource or configuration constraint violations. The Waltz algorithm cannot move the tasks that are user-fixed. In such cases when temporal satisfaction cannot be achieved, the algorithm reports failure.

When there is either a resource or a configuration violation, GPSS provides two options to the user: the user can manually resolve the conflicts by moving tasks, or use the auto deconfliction option to resolve the conflicts. When a user resolves conflicts by moving tasks, GPSS allows the user to move one task at a time. Temporal consistency is maintained by running the Waltz algorithm after each movement. When temporal constraints cannot be satisfied, the move made by the user is rejected. In this section we look into how the system deconflicts the violations.

When GPSS deconflicts the violations a user has limited options: (1) select the relevant resources to focus, and (2) control the time taken for deconfliction by setting the number of iterations and the window size. The window size determines how many violated resource constraints and configuration constraints are selected for deconfliction in each iteration. For example, if a user selects window size 5 and iteration 40, 5 resource constraints and 5 configuration constraints are resolved in each iteration, or until all the constraint violations are resolved. Let us describe the algorithm believed to be used in GPSS (the algorithm is written based on the notes provided by Mr. Jim Tulley).

```
for  i = 1 to iterations do
{       Construct the list of violated constraints and categorize them into
        (1) resource constraint violations
        (2) configuration constraint violations
        Construct a list of tasks from the resource constraint violation list.
        From this list randomly select k tasks (k is the window size) and call them tasks-to-focus
        for j = 1 to k do
        {       pick up a task from tasks-to-focus and update the list
                call Repair on the earliest violated constraint of this task
        }
        for j = 1 to k do
        {       repair configuration constraint violation
        }
}
```

A conflict, either resource or configuration, can be resolved by moving a task involved in the conflict to a new time period where the conflict does not arise. Let us consider how GPSS repairs the constraint violations. During each iteration, a list of tasks called tasks-to-focus is created randomly by selecting tasks of window size from the resource conflicting tasks, or the total of the resource conflicting tasks whichever

is the smaller. For each task in the tasks-to-focus list, find the earliest constraint violation and call a routine called Repair to repair the violation. The Repair routine, knowing the resource and the time period of the violation, determines all the tasks that request the resource during the time period. A heuristics table is constructed to determine which task to move to reduce the conflicts. The heuristics table consists of ten columns to represent (1) fitness, (2) nearest-fix, (3) task-active, (4) temporal dependents, (5) proximity to now, (6) cost, (7) task, (8) start time, (9) weighted sum of scores, and (10) next time to try. Each task is given two rows: one for forward direction and the other for backward direction.

The first column calculate the measure of fitness and it is given by the formula:

$$\text{fitness} = \frac{1}{(\text{request} - \text{over-allocation})^2 + 1}$$

The fitness of a task has a value of one if its request matches the over-allocation that indicates the conflict on the resource during the interval can be resolved by moving the task.

The nearest fix value of a task is the time span of the current start and the next-time-to-try column of the table. The next column is a flag indicating whether the task is active or not. Forward and backward temporal dependencies for each task are computed and stored in the temporal dependency column.

The cost in column 6 is computed by moving the task in both the forward and the backward directions within a temporary context. This cost is the approximate cost since temporal consistency is not checked after the move.

When all the computations are done for each row, the table is normalized. The task with the best weighted summation is selected for the actual move. The selected task is moved and the Waltz algorithm is run to satisfy the temporal constraints. If the move reduces the previous cost, the move is accepted and the iteration continues. When the move increases, the cost is accepted with some probability computed using simulated annealing techniques. We discuss the techniques and how they are being used to solve the problem.

3.1.1 Simulated Annealing

Simulated annealing is a technique used to avoid local minima when using a hill climbing technique to solve problems. This technique simulates some aspects of the natural annealing process, and hence got the name simulated annealing. In an annealing process, a metal is heated to a higher temperature and then cooled down gradually to obtain fine grain. When the temperature cools down rapidly, one obtains course grain in the annealing process.

In a typical hill climbing method, heuristics are applied to find the next, most promising move. The algorithm commits itself to that move and continues forward from the new state. Move with higher cost are rejected as the search progresses through moves that always decrease the cost. Very often the algorithm attains a local minima or plateau. To avoid local minima, moves with higher cost must also be accepted with some controlled probability. This is exactly what is done in simulated annealing. Suppose delta, Δ, is the increase in cost. A move which normally would be rejected for having higher cost will now be accepted with probability $e^{-\Delta/T}$, where T is the temperature. Initially T is set to a higher temperature and it is

decreased gradually. At the higher temperature, or in the initial phase of the solution, the probability of accepting moves with higher cost is high. This probability decreases as the temperature decreases.

3.1.2 How Simulated Annealing is used in GPSS

The deconfliction algorithm used in GPSS is a hill climbing algorithm coupled with a weaker version of the simulated annealing algorithm. The temperature is set at 75 and it remains at that temperature until the cost is less than 10 and the number of iterations completed is greater than 5. When these conditions are met, the temperature is reduced to 25.

Setting of initial temperature should be dependent on the problem size, the domain of the problem, and the heuristic strategy being used to solve the problem. In the GPSS implementation, temperature remains fixed for a wide range of solutions. Hence, there is a danger of accepting moves that increase the cost by more than 50% of the best found solution. Suppose the best cost is 25 and the new cost is 50. The increment of the cost is 25, 100% higher than the best known so far. In GPSS, the probability of accepting such a move is 0.71, which is very high. The annealing portion is so weak that it allows all sorts of moves with much higher cost. This makes the algorithm wander around the search space without any focus in reducing the cost.

3.2 Enhancement of the GPSS Deconfliction Process

We can improve the resolution of conflicts in GPSS by improving the quality of the solution while decreasing the computational cost of obtaining such solutions. In a typical flow, the total number of constraint violations will be in the order of several hundreds, and the span of the schedule will be around 85 days. Very often, a user is interested in getting a conflict free schedule for a selected period of time, say for 11 days. The current GPSS does not allow such option, instead the deconfliction process selects tasks randomly and resolves the conflicts in each iteration. The algorithm may be spending time to resolve conflicts at the tail end of the schedule which is currently of much less interest to the user. A *fencing capability* will provide the user with the option to select the time period on which to focus, and enable the removal of all conflicts occurring during that period.

With the fencing capability, a user can interactively select the time period over which deconfliction takes place. The deconfliction algorithm is changed to select only the violated constraints occurring during the fencing period. During deconfliction, the violated tasks are moved forward or backward to remove the conflicts. The algorithm should make sure that a task cannot be moved before NOW. Further, the algorithm tries to keep the tasks inside the fencing windows by storing the tasks moved out of the fencing window during deconfliction into a list called try-later. After all the violations are resolved, the algorithm tries to schedule the tasks from try-later into the fencing windows without violating the existing schedule in the fencing window.

In addition to the fencing capability, a user would like to have the option to select a certain set of tasks and make them static during the deconfliction. This option can be easily implemented. First, the algorithm ensures there is no constraint violation among the selected tasks. If constraint violation exists, the algorithm reports appropriate error message and rejects the set of tasks. Once the request is accepted, the algorithm excludes these fixed set of tasks from the candidates to be moved to get a conflict free schedule. It is also possible that there may not be a conflict free schedule once a set of tasks are fixed.

The current implementation of GPSS does not enforce space capacity constraint. Without such enforcement it schedules tasks in the same time slot even though space capacity constraint is violated. Consider three tasks T_1, T_2, and T_3 each requiring access to the crew module. Assume that T_1, T_2, and T_3 respectively require 3,2, and 2 technicians and can operate concurrently. If the capacity of the crew module permits at most 4 people at a time, T_1, T_2, and T_3 can not be scheduled during the same time period without violating the space capacity constraint. GPSS does not enforce such constraint, therefore, it will not report this as a violation. This problem can be corrected by adding space capacity constraint to GPSS. The supporting implementation will be very much similar to that of resource constraint.

In addition to providing these capabilities to the user, we can improve the deconfliction process by improving the heuristics and decreasing the computational cost. A close examination into the heuristics table reveals that GPSS captured many interesting aspects of schedules such as fitness, temporal dependency, nearest fix, and cost. The computation of all the parameters except cost will take a constant amount of time. Computing the cost of a move is computationally very expensive since it involves de-allocation of all the resources for the candidate task, re-allocation of the resource to the task at the new time period, and computation of all the violations with the new placement. This is clearly the most expensive item in the table. We would like to remove the cost column of the heuristics table and add two other heuristic parameters: (1) number of constraints, and (2) ratio of violated constraints to total constraints. When a task with only a few constraints is moved, the chance of perturbation to the schedule is smaller than the chance incurred when moving a task with many constraints. This column can be normalized as:

$$1 - \frac{\text{number of constraints}}{\text{maximum number of constraints among the conflicting tasks}}$$

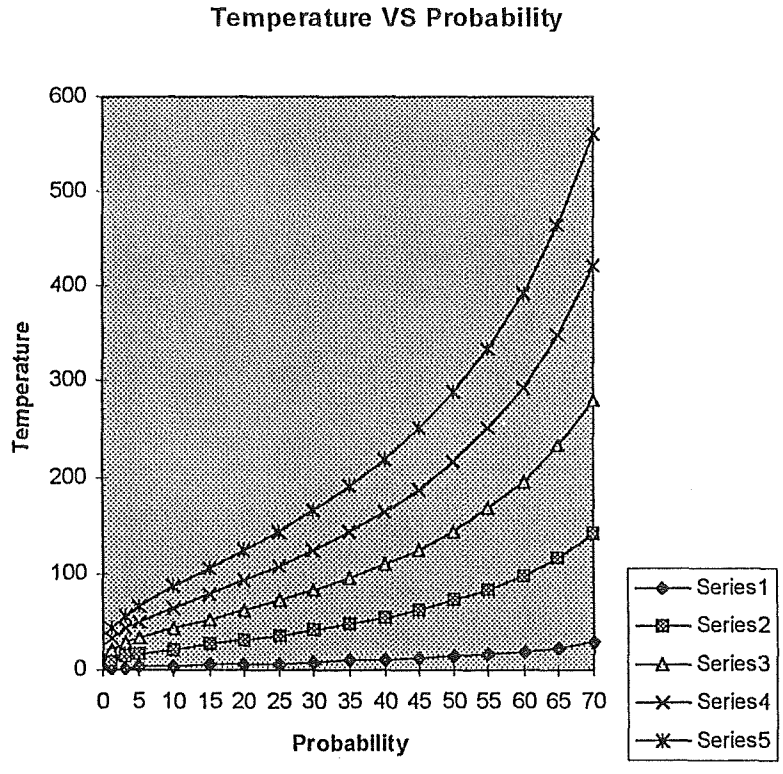
When comparing candidates for movement, we must take the percentage of violated requests into consideration. A task with a higher percentage of its requests violated should be favored as a potentially better candidate to move than the others. This is captured by the other, similarly normalized parameter.

The iterative repair method used in GPSS is basically a hill climbing algorithm coupled with simulated annealing to avoid local minima. Unfortunately, the simulated annealing as implemented in GPSS is very weak. It fixes the temperature at 75 until the cost is less than 10 and there have been at least 5 iterations. Suppose the current cost of a schedule is 15 and the new cost as a result of a new move is 30, then this move of 100% increase in cost will be accepted with 0.818 probability. Worst yet, it will accept 200% increase in cost with 0.67 probability.

The success of solving a problem using simulated annealing depends on setting the temperature at the correct value and gradually cooling it down. When the temperature is set at a higher value, as has been done in GPSS, the algorithm accepts moves with very high cost. Therefore, it may jump around the search space without focusing on the most fruitful direction of cost reduction. On the other hand, when the temperature is set too low it will not allow reasonable moves that may very well be leading to a better solution.

We reject fixing the temperature statically. Instead, we favor setting the temperature for each set of iterations. The initial temperature is based on the initial cost, say init-cost, percentage of allowable increases in cost, say per-incr-cost, with the probability, say p. Then the initial temperature is equal to:

$$\frac{\text{per-incr-cost x init-cost}}{100 \text{ x } \ln\left(\frac{1}{p}\right)}$$



This graph shows the variation of the initial temperature with the accepting probability for a series of increased cost. The series 1, 2 3, 4 and 5, respectively, represent the percentage of increased cost of 1, 5, 10, 15 and 20. The initial cost is assumed to be 1000.

4. Summary and Discussion

We have successfully completed the following overall objectives of the summer research sponsored by NASA/ASEE: (1) to provide an evaluation of a commercially developed version of GPSS for its applicability to the KSC ground processing problem; (2) to work with the KSC GPSS development team and provide enhancements to the existing software.

We have successfully installed and made an initial evaluation of PRA from Red Pepper Software for its suitability for scheduling ground processing activities. The models supplied with PRA do not support ground processing scheduling. It is a non-trivial task to build a model that supports ground processing scheduling activities using the PRA shell and SPL, their proprietary Scheduling Programming Language.

We then focused on improving and enhancing the ground processing scheduling system (GPSS). GPSS consists of a user interface and an engine. Major functional units of the engine are temporal propagation and deconfliction of violations. A modified version of the Waltz algorithm is used for propagating temporal constraints. The conflict resolution of the engine forms the bulk of the cognitive and computational complexity of the engine. The performance of deconfliction can be improved by improving the quality of the solution without sacrificing the computational cost. We have investigated into the fencing capability, and space capacity constraint and have provide some insight of how this capability can be implemented.

During deconfliction, GPSS selects a task and moves it to an interval where it can be scheduled without any violation. GPSS uses a set of heuristics to determine which task to move among the tasks in conflicts. The current GPSS system captures important parameters. We have added two more parameters that help to make a better decision.

The deconfliction process of GPSS is a hill climbing method coupled with a weak simulated annealing implementation. The temperature for the annealing process is fixed at 75, and it is brought down to 25 after achieving some conditions. We propose a formula that determines the initial temperature based on: (1) the initial cost, and (2) the tolerance limit with which one is willing to accept a higher cost in order to avoid local minima. We have also laid down procedures to gradually cool down the temperature as the algorithm improves the solution.

We have implemented the resource deconfliction portion of the GPSS in common LISP using its object oriented features. We have used the LISP profile prototype code, which was developed by the GPSS reverse engineering group, as a building block to our implementation. Our prototype will help the GPSS developers re-engineer the deconfliction portion in C++. It corrects and extends some of the deficiencies of the current production version, plus it uses and builds on the classes from the development team's profile prototype, which they are now porting to C++.

We feel that the resource deconfliction can be further improved by moving only a portion of a conflicted task if the resource is not dedicated to it, instead of moving the complete task as has been presently implemented in GPSS. Further investigation is required to substantiate the claim and fine tune the implementation.

The heuristics table is very static, and the normalization process is somewhat arbitrary. To fine-tune the weight of the table, an extensive experiment must be conducted. Use of an artificial neural network seems to be a promising approach to set the weights of the heuristics table so that the system can make better decisions about which tasks to move.

We have looked into some of the modeling capabilities provided by GPSS. We are intrigued, for instance, by the configuration requirements and effects. Suppose there are three tasks T_1 , T_2 , and T_3 which all require the bay area to be open. Assume that T_1 and T_3 are mutually exclusive, that is, they cannot be performed concurrently. A user models T_1 and T_3 with the requirement 'bay area open', and the effect 'cleared' that persists from beginning to end. The attribute here is 'bay area', which can be in an 'open' or a 'cleared' state. The effect here is not at all causal as one would suspect. T_1 and T_3 are modeled as having an effect of 'cleared' to ensure the mutual exclusion. This modeling will not permit T_2 to be carried

out concurrently with either T_1 or T_3. An in-depth study is required to provide a flexible way to model tasks and their behaviors naturally.

REFERENCES

[1] M. Deale, M. Yvanovich, D. Schnitzius, D. Kautz, M. Carpenter, M. Zweben, G. Davis, B. Daun. 1994, "The Space Shuttle Ground Processing Scheduling System," in *Intelligent Scheduling*, edited by M. Zweben, and M. Fox, *Morgan Kaufmann Publishers*, San Francisco, Calif., pp.423-449.

[2] S. Kirkpatrick and C. D. Gelatt, and M. P. Vecchi, 1983. "Optimization by Simulated Annealing," *Science* Vol. 220 #4598.

[3] D. Waltz 1985. "Understanding Line Drawing Scenes with Shadows," *The Psychology of Computer Vision*, edited by P. Winston, McGraw-Hill.

[4] M. Zweben, E. Davis and M. Deale. 1993. "Iterative Repair for Scheduling and Rescheduling," *IEEE Systems, man, and Cybernetics*, Special Issue on Planning, Scheduling, and Control.

[5] M. Zweben, B. Daun, M. Deale. 1994. "Scheduling and Rescheduling with Iterative Repair," in *Intelligent Scheduling*, edited by M. Zweben, and M. Fox, *Morgan Kaufmann Publishers*, San Francisco, Calif., pp.241-255.

1996 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

LIGHTNING STUDIES USING VHF WAVEFORM DATA

Dr. Mark B. Moldwin, Assistant Professor
Physics and Space Sciences Department
Florida Institute of Technology
Melbourne, Florida

KSC Colleague - Carl Lennon
Atmospheric Science

Contract Number NASA-NGT10-52605

June 28, 1996

ABSTRACT

Several atmospheric electricity studies were begun utilizing VHF lightning data obtained with the Lightning Detection and Ranging System (LDAR) at KSC. The LDAR system uses differences in the time of arrival of electromagnetic noise generated by the lightning process to seven antennas to calculate very accurate three dimensional locations of lightning. New software was developed to obtain the source location of multiple, simultaneous, and spatially separated lightning signatures. Three studies were begun this summer utilizing this data and are: (1) VHF observations of simultaneous lightning, (2) ground-based VHF observations of TIPPs, and (3) properties of intracloud recoil streamers. The principle result of each of these studies are: (1) lightning commonly occurs in well separated (2-50 km) regions simultaneously, (2) large amplitude pairs of VHF pulses are commonly observed on the ground but had not been previously identified due to the large number of signals usually observed in the VHF noise of close lightning, and (3) that VHF Q-noise and pulse signatures associated with K-changes within intracloud lightning propagate at velocities of $>10^8$ m/s. The interim results of these three studies are reviewed in this brief report. The results of these studies will be submitted to *Geophysical Research Letters* and the *Journal of Geophysical Research.*

LIGHTNING STUDIES USING VHF WAVEFORM DATA

Mark B. Moldwin, Brent M. Goode, and Carl L. Lennon

1. INTRODUCTION

Kennedy Space Center (KSC), the 45th Space Command at Cape Canaveral Air Station (CCAS), and the National Weather Service (NWS) in Melbourne operate extensive and sophisticated meteorological and atmospheric physics instrumentation and systems to help support the launch activities at KSC (the shuttle) and CCAS (unmanned rockets). One of these systems is the Lightning Detection and Ranging (LDAR) system developed at KSC [1]. This system is used operationally to assist in making lightning warnings and advisories at KSC and CCAS areas. The system uses differences in the time of arrivals (DTOA) at seven stations of VHF (66 MHz) electromagnetic radiation generated in the lightning process. The DTOAs are converted to the three-dimensional location of the source of the radio noise [2, 3]. This method is demonstrated in Figure 1. The radio source is assumed to be a point source that radiates isotropically. The radio signal then propagates out in a spherical pattern intercepting the different stations at different times. By knowing the relative locations of the 6 remote sites compared to the central site, six time differences can be determined. Any combination of three stations ideally would give you the (X, Y, Z) location of the source, however due to uncertainties in the timing, some combinations of sites give better results due to an effect known as the geometric dilution of precision (GDOPs) [4]. We have tested several algorithms to determine the best estimate of the source location using the 20 combinations of four stations possible with a seven antenna array. We found that a method similar to the one utilized by the LDAR system [5] works the best. The new system has been termed "TIPPs" since it was designed to study trans-ionspheric pulse pairs [6]. However, in our effort to study TIPPs we found that the system is also able to study several other phenomena as well. Therefore this report is divided into four sections. The first three sections describe the initial results from three studies and the fourth section summarizes some of the calibration results. The final results of these studies will be submitted to the *Journal of Geophysical Research* and reprints will be given to the NASA/ASEE office when available. In this report only very brief summaries of the later two science projects are presented.

2. VHF OBSERVATIONS OF SIMULTANEOUS (SYMPATHETIC?) LIGHTNING

2.1 Introduction

Visual observations from space of thunderstorms have shown that lightning flashes often occur simultaneously (to the eye) from well-separated regions and that the onset of one flash often appears to precipitate or trigger others over a wide area [7]. The latter phenomena has been termed "sympathetic" or "associated" lightning [8]. Mazur, using VHF radar, defined "associated" lightning as a sequence of echoes from lightning separated by at least 1 km and separated by less than 200 ms in time. He demonstrated for events that occur within 200 ms of one another that the probability that the events are independent and randomly occurring is very small.

The question that arises from such observations is what is the cause of the association between sympathetic lightning flashes. We utilized the Lightning Detection and Ranging (LDAR) system at KSC to determine the locations of VHF lightning signals during one severe winter storm that passed over the LDAR system on 30 April, 1996. We found that within our 200 μs window 42% (41 of 98 records) of our triggered events contained a lightning signal located at least 2 km from the trigger event. Of course the second VHF signal maybe due to a continuation of the same flash since it is common to observe multiple VHF signals within a 200 μs window [e.g., 3].

However, since we know the location of the source of the VHF noise and the time-delay between the arrival of the signals to the LDAR system, we can estimate a velocity of propagation if the two sources are related. Typical VHF propagation velocities observed for intra-cloud step leaders are several times 10^5 m/s. Therefore by examining the apparent propagation velocity we can attempt to discriminate between observing two parts of the same flash and two separated flashes. Though we cannot uniquely determine the source mechanism of sympathetic lightning, we can restrain the possible causes. Mazur [8] hypothesized that associated lightning maybe caused by the interdependence of electric fields between neighboring electrically active cells (EACs) in a multicell thunderstorm. Specifically it was thought that the collapse of an electric dipole in one of the neighboring EACs due to a lightning flash would cause an electric field pulse which could trigger a neighboring cell. This study can test this hypothesis by examining whether an electromagnetic pulse has the time to propagate between spatially separated EACs.

2.2 Observations and Methodology

LDAR is a passive array of seven antennas that detect VHF pulses at 66 MHz with a bandwidth of 6 MHz. The antennas are sensitive to both horizontally and vertically polarized signals and are separated from 7 to 10 km from a central site (See Figure 1). The system and its performance are described in [4], [5] and [1] and therefore details of its operation are not given here. Though the hardware used in this study is the LDAR system, a new revised data management and new software routines are used. These software changes allow us to analyze the VHF waveform data in 200 μs windows with a sampling rate of 50 ns. The new software also aides in the selection of the "correct" pulses observed at the different stations hence alleviating one of the major disadvantages of long-baseline VHF time-of-arrival systems [9]. This disadvantage, namely the difficulty in identifying VHF pulses from two or more simultaneous separated sources because the pulses can arrive in a different order at each receiver, is alleviated by having the computer help pick the correct peaks depending on an initial "best guess." This semi-automated routine allows one to calculate the 3-Dimensional locations of virtually all the identifiable pulses within one 200 μs window. This is the first system with this capability and allows us to determine the locations of multiple VHF signals separated by only a few microseconds.

Though our sampling rate is 20 MHz (50 ns time resolution), we use a "zero-stuffing" technique [e.g., 10] to increase our effective time resolution to approximately 6 ns. This is accomplished by interpolation using packing in the frequency domain. The total uncertainty in our timing is however approximately 30 ns due to errors in such factors as the speed of the VHF propagation in air, transmission line delays, noise, bandwidth, the geometry of the antenna array with respect to the position of the source (Geometric Dilution of Precision or GDOPs) and quantizing. The main contributor to the error is the uncertainty in calculating the transmission line delays (or K-factors). We use a lightning simulator located at a known source near the central site to determine the K-factors for each of the six remote sites. Due to noise in the signals and quantizing the exact timing of the calibration signal has an uncertainty which is typically 10-15 ns. Due to having an array of 7 antennas we effectively have 20 configurations of 4 antennas to do ranging. This almost virtually assures us of having a combination with good GDOPs to complete our calculations. The lightning simulator allows us to remove any bias or systematic errors from our analysis so we are left with only the random errors outlined above. The system is triggered when a signal is received at the central site that has an amplitude above a preset level. This level was set fairly high, so the pulses triggered were probably large intracloud pulses emitted during the initial or active phase [11, and 12] and have been associated with first streamers within the cloud [3] (though some of the signals move at very high speeds which have been associated with return stroke recoil streamers). Data are saved to disk 5 μs prior to the trigger event and 195 μs after. Due to the spacing of the array, many more than one signal can fall within this 200 μs window. Since we save the waveform data we can inspect each record and determine the individual peaks which

correspond to a single event as observed at each station. This gives us 6 time delays from the 6 pairs of stations (each remote site compared to the arrival at the central site). From this we have 6 equations and three unknowns. Therefore any 3 equations (or combination of three sites) gives us a solution of X,Y, Z for the source. However, due to the errors in timing and to the GDOPs, not all combinations necessarily give a good solution. We calculate the lightning location for each of the 20 combinations of three equations and filter out the sites with bad GDOPs. The algorithm used is similar to the LDAR algorithm [4] except that we initially utilize all 20 combinations whereas LDAR uses the 2 combinations that have the best overall GDOPs if their solutions agree within 5%.

A typical set of waveform data from this storm is shown in Figure 2. The seven panels show from bottom to top the signal from the central site (site 0) to site 6. Notice the sequence of distinct pulses that arrive at the different stations at different times. We use this time delay information and the known locations of the 7 sites to determine the 3-dimensional (X,Y,Z) location of the VHF noise source. For sources within approximately 15 km of the central site the accuracy is within 1% in X, and Y and 2% in Z. For events that occur within the baseline of the array, the uncertainty is on the order of 10s of meters.

For the interval shown in Figure 2 twelve pulses have been identified and their locations and absolute time differences determined. The locations are giving in Table 1. Note that pulses K and L are located 11.47 km apart while pulse pair H and I were located coincidentally in space (within the errors). The two pairs (KL) and (HI) are examples of two types of multiple signals we typically see in our 200 μs window. Pair KL is an example of apparently separate lightning flashes while HI are examples of multiple signals from the same flash as reported by [3, 5, 13]. We also occasionally trigger on Q-noise which [3, 14] attributes to recoil streamers preceding K-changes in intracloud lightning flashes. The properties of Q-noise is examined separately and is described in section 4.

2.3 Main Result

We examined the major amplitude pulses observed in ninety-eight 200 μs records obtained during the 30 April, 1996, winter thunderstorm. For each set of pulses observed at all 7 sites we calculated its 3-dimensional location relative to the central observing site, the relative time delay between it and any subsequent pulses, the amplitude, the absolute time of the pulse, and the apparent propagation velocity between each possible combination of pulse pairs. On average we identified 2.31 VHF lightning discharges per record. As seen in Figure 2 however, it should be noted that there is often many other pulses observed that were not identified due to changing amplitude of the signal between stations, proximity to other pulses, presence of Q-noise, or having the signal arrive prior to opening or after the closing of the 200 μs window.

Figure 3 shows a histogram of the number of 200 μs records that contained 1, 2, 3, or more separate lightning signals within the record. Most records contained multiple signals but from the same source, though 42% of the records contained signals coming from well separated sources.

We have found that it is common to observe multiple lightning discharge VHF signals from well separated (2-50 km) sources within several 10s of μs of each other. We present these results as evidence of simultaneous lightning in a winter multicell thunderstorm. The natural question that arises from these observations is if these lightning discharges are dependent or independent of one another. Mazur [8] demonstrated that the interdependence hypothesis fails for lightning flashes that begin within about 200 ms of each other and are separated by at least 1 km to the 5% confidence level. Because of our non-continuous records we cannot uniquely determine the start of the lightning flash (i.e., we don't know if we triggered at the first pulse of the flash or somewhere in the middle). Therefore we cannot directly extend Mazur's results to our data. However we do demonstrate that multiple flashes do often occur simultaneously in time over extended distances.

The next question that arises is if they are dependent, what is the mechanism of their relationship? We suggest two possibilities; (1) as suggested by [8], an EMP is launched from one EAC to another changing the electric field configuration in such a way as to trigger the second lightning discharge, and (2) a extra-terrestrial source such as a cosmic ray shower passing through two widely separated EACs generating new ionization trails which leads to lightning discharges.

3. GROUND OBSERVATIONS OF TIPPs

TIPPs were originally observed from space using a broadband VHF radio receiver on the Alexis satellite [6]. They are characterized by a pair of dispersed (in frequency) signals separated from one another by 4 to 100 μs (with a median of 50 μs). Their VHF source power is estimated to be considerably stronger than typical lightning discharges (about 100 kW compared to the 1-1000 W for typical lightning). Though the source mechanism of the signals could not be determined it was suggested they came from thunderstorms and perhaps were related to sprites, jets, elves, or terrestrial gamma-ray bursts observed in the upper atmosphere. In observations of close lightning (within 15 km) there are multitude of pulses and Q noise trains within a 200 μs window (see Figure 2). However, as the thunderstorm moves away from the LDAR system the number of pulses (particularly Q noise pulses) rapidly diminish. This suggests that Q noise maybe an electrostatic process that falls off as the inverse cube of the distance. This is in contrast to EM radiation signals whose fields fall off as 1/R. (However, this fall off with distance may be due to the inherent relative weakness of Q-noise compared to pulses). For lightning far away from the system several examples of large amplitude, pulse pair signals have been observed. The characteristics of the pulse pairs such as pulse duration, and inter-pulse timing are identical for those found for TIPPs. However, a distribution of the source power of these signals show that they are considerably weaker than those estimated for TIPPs. Despite this difference, we suggest that TIPPs are just radio observations of very energetic pulse paired lightning. The open question that remains is what lightning process occurs in pairs?

4. 3-D MAPPING OF VERY FAST INTRACLOUD RECOIL STREAMERS

Recoil streamers associated with ground return strokes have been found to have the highest propagation velocity of any lightning phenomena typically exceeding one-third the speed of light [9]. The TIPPs system offers the first 3 dimensional velocity information of these events since earlier estimates were made photographically or with interferometers which could only make 2D measurements. We found that recoil streamers travel at approximately 2×10^8 m/s and travel several kilometers at a time with no marked stepping. This phenomena has been observed by [3] using a VHF DTOA system and [15] using VHF interferometric techniques. However, the former identified this phenomena with Q-noise trains. We found that this phenomena has both Q-noise and pulse waveforms. In addition, we observed Q-noise pulses of <10 μs duration whereas in a histogram of 310 Q-noise durations [3] demonstrated that he found none of that short a duration (though in one of his examples he presents a 7 μs long Q-noise burst). We were able to map the propagation of several of these recoil streamers over many kilometers and found that the apparent velocity between pulses remained fairly constant along a single path.

5. CALIBRATION AND PROGRAMMING

This section describes the miscellaneous tasks that were completed. The WVWSSTRXYZL program was commented and debugged. The jitter (a measure of the uncertainty) was changed to the standard deviation of the X, Y, and Z values for the different antenna combinations used in the solution. A Grid Search/Linear Least Squares algorithm was developed and compared to the "LDAR" method. It was found that the LDAR solutions were equal to this more computationally

intensive method using calibration data. New K-factors were determined and are listed in the WVWSSTRXYZL program. A calibration power curve for Site 0 was made to determine the absolute electric field amplitudes for the TIPPs study. It was found that the relationship between counts and power followed the following linear relation

$$dB(\text{above a microvolt}) = (\text{count} + 426.71)/72.165 \quad (1)$$

The antenna response pattern changes with elevation angle and was estimated to follow the following linear relationship from 30 to 0 degrees

$$dB = [30\text{-elevation angle}] \times (0.40) \quad (2)$$

or 11.87 dB fall off from 30 to 0 degrees. The antenna response is from 0-1 dBi from 90 to 30 degrees.

6. ACKNOWLEDGMENTS

I would like to thank Carl Lennon for being an excellent mentor, colleague, and host. The LDAR system he has developed has made and will continue to make important contributions to the field of lightning studies. Brent Goode has been an excellent summer research assistant and I look forward to working with him on completing the studies next fall. The KSC NASA/ASEE staff have made the summer faculty fellowship program a intellectually rewarding experience and I wish them continued success.

7. REFERENCES

[1] Lennon, C. L., LDAR-A new lightning detection and ranging system, *EOS,* AGU, 56(12), 991, 1975.
[2] Oetzel and Pierce, VHF technique for locating lightning, *Radio Sci., 4,* 199, 1969.
[3] Proctor, D. E., VHF radio pictures of cloud flashes, *J. Geophys. Res., 86,* 4041-4171, 1981.
[4] Poehler, H.A., An accuracy analysis of the LDAR system, NASA Technical Report FEC-7146, 1977.
[5] Maier, L., C. Lennon, T. Britt, and S. Schaefer, Lightning detection and ranging (LDAR) system performance analysis, Proc. Intn'l. Conf. Cloud Phys., Dallas, TX, paper 8.9, Amer. Meteorol. Soc., Boston, MA., 1995.
[6] Massey, R.S., and D.N. Holden, Phenomenology of Trans-Ionospheric Pulse Pairs, *J. Geophys. Res.,* 1995.
[7] Vonnegut, B., O.H. Vaughan, Jr., M. Brook and P. Krehbiel, Mesoscale observations of lightning from Space Shuttle, NASA Technical Memorandum 86451, 1984.
[8] Mazur, V., Associated lightning discharges, *Geophys. Res. Lett., 9,* 1227-1230, 1982.
[9] Uman, M. A., *The Lightning Discharge,* Academic Press, New York, 1987.
[10] Bracewell, R.N., The Fourier Transform and its Application, McGraw-Hill, New York, 1986.
[11] Kitagawa, N., and M. Brook, A comparison of intracloud and cloud-to-ground lightning discharges, *J. Geophys. Res., 65,* 1189-1201, 1960.
[12] Weidman, C.D., and E.P. Krider, The radiation field wave forms produced by intracloud lightning discharge processes, *J. Geophys. Res., 84,* 3159-3164, 1979.

[13] Rustan, P.L., M.A. Uman, D.G. Childers, and W.H. Beasley, Lightning source locations from VHF radiation data for a flash at Kennedy Space Center, *J. Geophys. Res., 85*, 4893-4903, 1980.

[14] Proctor, D. E., R. Uytenbogaardt, and B.M Meredith, VHF radio pictures of lightning flashes to ground, *J. Geophys. Res., 93*, 12683-12727, 1988.

[15] Shao, X.M., P.R. Krehbiel, R.J. Thomas, and W. Rison, Radio interferometric observations of cloud-to-ground lightning phenomena in Florida, *J. Geophys. Res., 100*, 2749, 1995.

Table 1. The locations of the pulses identified in Figure 2.

Event	x (km)	y (km)	z (km)
A	-6.464	8.845	6.905
B	-7.563	9.945	6.034
C	-8.545	8.356	6.746
D	-7.868	7.452	7.666
E	-7.081	5.935	4.460
F	-4.369	6.268	7.483
G	-4.401	6.310	7.650
H	-6.403	8.906	7.935
I	-6.656	9.140	7.132
J	-6.447	8.891	7.033
K	-11.620	15.084	15.188

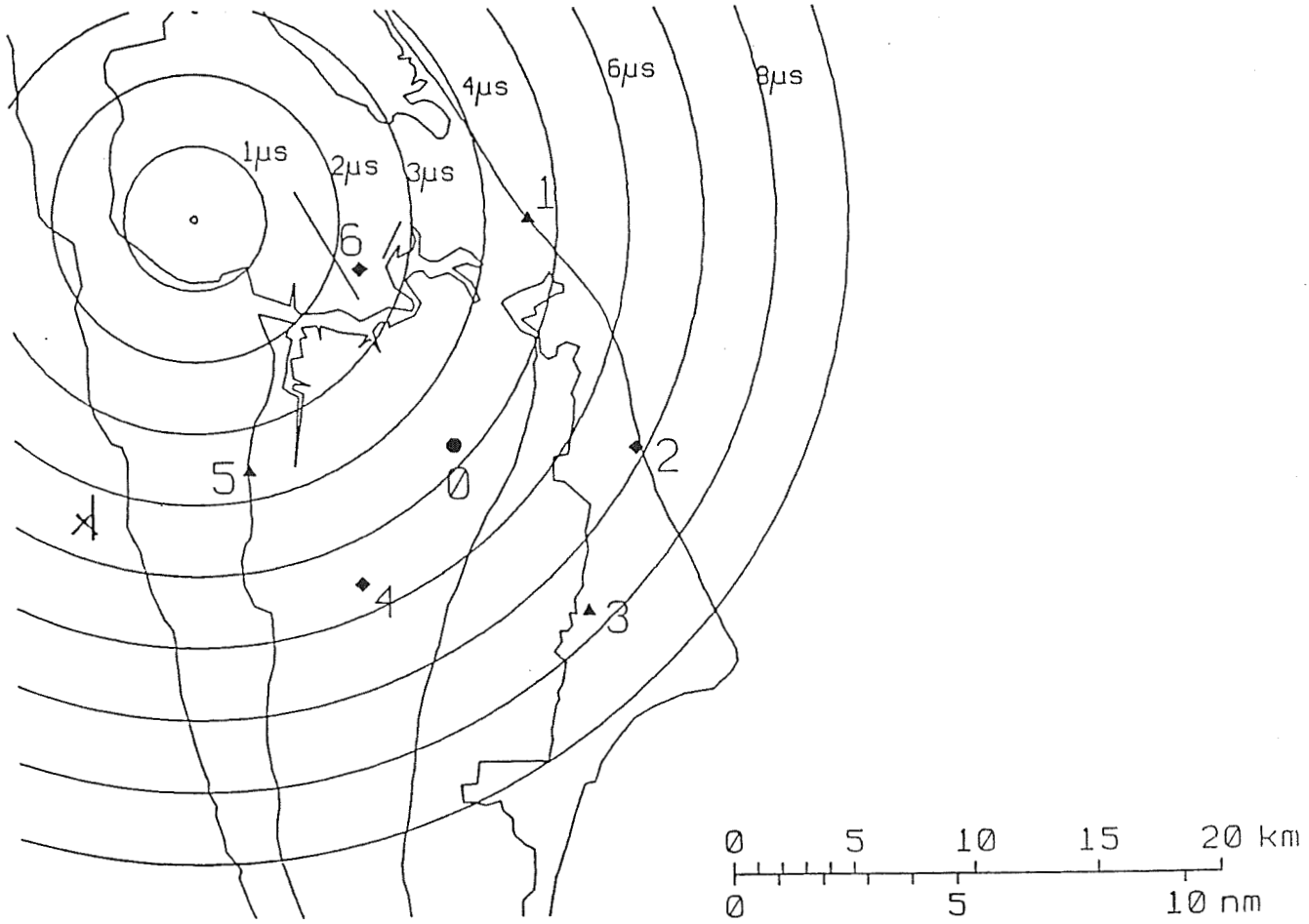


Figure 1. The locations of the seven LDAR sites on KSC and the location of the EMP as it propagates out from a example lightning pulse.

Figure 2. An example LDAR "TIPPs" waveform plot. The plot contains 200 μs of data and shows the data from the central site (bottom) through the 6th remote site.

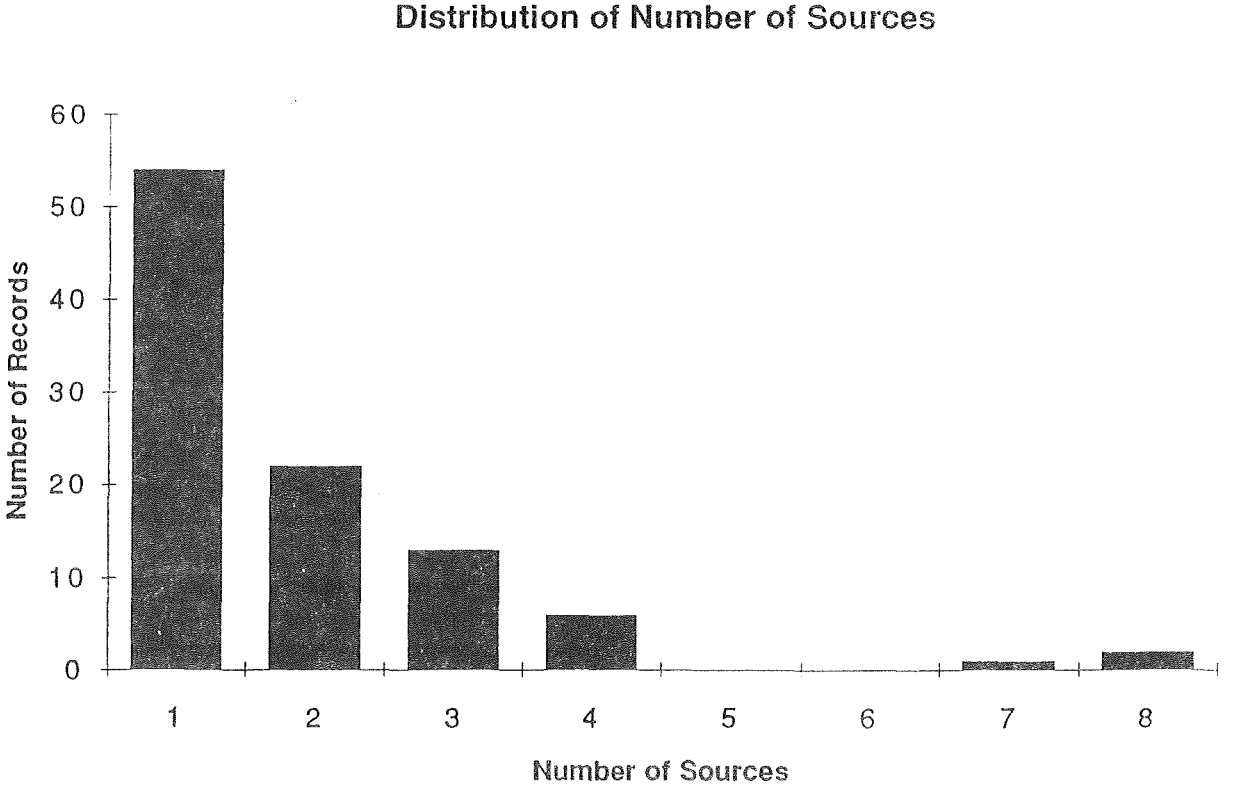


Figure 3. A histogram showing the distribution of records that contained multiple pulses.

1996 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

WHITE LIGHT SCHLIEREN OPTICS USING BACTERIORHODOPSIN AS AN ADAPTIVE IMAGE GRID

Dr. Robert E. Peale, Assistant Professor
Mr. Boh Ruffin, Graduate Student
Mr. Jeff Donahue, Undergraduate Student
Physics Department
University of Central Florida (Peale and Donahue)
University of Michigan (Ruffin)

KSC Colleague - Carolyn Barrett
Instrumentation and Hazardous Gas Monitoring

Contract Number NASA-NGT10-52605

August 5, 1996

ABSTRACT

A schlieren apparatus using a bacteriorhodopsin film as an adaptive image grid with white light illumination is demonstrated for the first time. The time dependent spectral properties of the film are characterized. Potential applications include a single-ended schlieren system for leak detection.

1. INTRODUCTION

Remote imaging of leaks is needed at large industrial complexes including, for example, the space shuttle launch facilities at Kennedy Space Center. Backscatter/Absorption Gas Imaging (BAGI) [1] is useless for gases which lack strong IR absorption, such as O_2, H_2, He, and N_2. This and other laser techniques, e.g. Raman imaging, may be dangerous for personnel and equipment. A system using white light, preferably ambient light, is desirable.

The schlieren method is well established for viewing plumes of foreign gases in air[2]. The test region is sandwiched between a source grid and its negative, onto which the source grid is imaged using appropriate optics. Only rays deflected by a plume in the test region may pass the image grid. These rays are used to image the plume. By thus eliminating the uninteresting background, the signal-to-noise is greatly enhanced, and otherwise invisible plumes appear in high-contrast detail.

All usual schlieren set-ups sandwich the test region between optics, which limit the field of view to industrially uninteresting scales. Peale and Summers[3] showed that the optics on one side can be replaced by a high contrast pattern on flexible reflecting cloth. A zoom lens images the pattern onto its negative. This scheme can be scaled to large fields of view with only modest cost increases, in principle. An obvious extension of this idea is to substitute a naturally occurring high contrast scene for the pattern-cloth combination, thereby creating a single-ended schlieren system. The negative can be photographic, but exposure, development, reinsertion, and realignment are time consuming operations, during which the scene and its illumination may change.

Erasable photochromic films offer an attractive alternative. Films of bacteriorhodopsin (BR) from the purple membrane of Halobacterium Halobium are promising because larger absorbance changes and more cycles can be achieved than with man-made photochromic chemicals[4]. In the simplest model, the BR ground state (bR state) has a strong absorption in the yellow-green. Absorption by this band pumps BR into its long-lived excited M-state, which absorbs in the blue. Thus, a negative of a scene in blue light can be created in a BR film using yellow-green light.

Downie has demonstrated a BR schlieren apparatus using blue and green laser light [5]. Use of white light and color filters was unsuccessful, however. We demonstrate successful white-light BR schlieren with performance only 2 times worse than a traditional schlieren apparatus. Our system is sufficiently sensitive to observe the heat waves generated by the human body. This result is a first step in realizing a truly single-ended remote imaging system for leaks using white

light. A variety of laboratory applications can also be envisioned.

2. EXPERIMENT

Transmission spectra were collected using a diverging incandescent source, a 400 to 700 nm variable interference filter having 20 nm bandwidth, and a large area photovoltaic detector. Low intensity measurements were performed with this set up by placing the BR film just in front of the detector where the beam spot size is about 2 cm. Here, no time dependence is observed in the transmission. Higher intensity measurements were performed by placing the film just after the interference filter where the spot size is still just a few mm, and strong time dependence is observed. Data were recorded on a strip chart recorder.

Time dependent transmission data were also recorded using a variety of blue-violet band pass filters, long pass filters, and a common flood lamp. This source and two of the filters were used subsequently for schlieren experiments. A large-area Si detector was used in photovoltaic mode with a variety of low impedance load resistors to maintain linear response. The output was recorded on a strip chart. Neutral density filters were moved from front to back of the BR film to provide a range of incident intensities at the BR film while keeping the average intensity at the detector constant.

Standard schlieren optics in the configuration shown in Fig. 2.1 collected images of a variety of phase objects for comparison with BR schlieren. The BR schlieren set up is shown schematically in Fig. 2.2. A ground glass plate diffused the flood lamp illumination and reduced UV emissions. The heat absorbing filter was Schott KG2. The long pass filter was Schott OG550. The blue-violet band pass was Schott BG12, which was combined with a neutral density filter having an optical density of 1. The 19-inch, six-foot focal length spherical mirror imaged the source grid either onto the image grid (Fig. 2.1) or onto the bacteriorhodopsin film (Fig. 2.2). The grid(s) were Ronchi ruling(s) having 50 lines per inch from Edmund Scientific. A Javelin JE2062IR black-and-white ccd camera monitored the test region, which was located just in front of the spherical mirror. A zoom lens optimally filled the camera image plane with the test region.

The bacteriorhodopsin film was obtained from Bend Research and had a nominal optical density of 2.8 at 570 nm (absorbance of 6.5). The wild-type BR was incased in polyvinyl alcohol to form a ~100 μm film. The nominal M-state lifetime was 1 to 5 s.

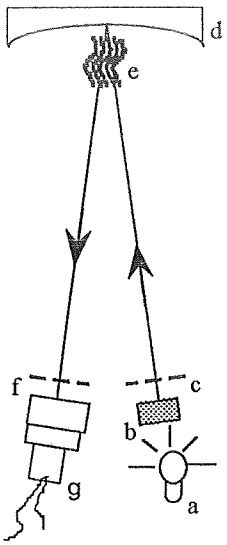


Fig. 2.1. Schematic of standard schlieren optics. a) Source; b) diffuser; c) source grid; d) spherical mirror; e) phase object; f) image grid; g) camera with zoom lens.

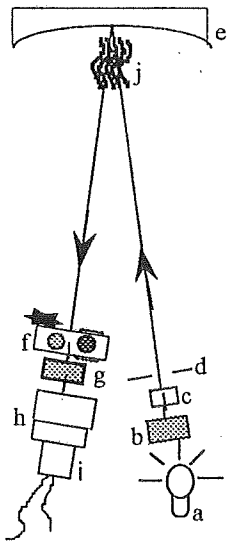


Fig. 2.2. Schematic of a bacteriorhodopsin schlieren set-up. a) Source; b) diffuser; c) heat-absorbing filter; d) source grid; e) spherical mirror; f) filter holder with two interchangeable filters; g) bacteriorhodopsin film; h) zoom lens; i) camera; j) phase object.

3. RESULTS

The relevant optical properties of our BR film were characterized first. Fig. 3.3 presents absorbance spectra. The solid circles were taken at an intensity sufficiently low that no time dependence of the transmission was observed. The cross symbols in Fig. 3.3 give the absorbance spectrum immediately after completion of a higher-intensity bleach transient using 525 nm light. Since BR has no absorption at 1 μm, the experimental absorbance at 1 μm was subtracted to eliminate contributions of reflection and scattering. After the bleach, the absorbance is everywhere lower except near 400 nm. The solid diamond symbols in Fig. 3.3 represent the difference in the two curves, which reveal the decreased absorbance in the yellow-green region and the increased absorption in the blue-violet. These changes result from the light induced population of the BR metastable M-state.

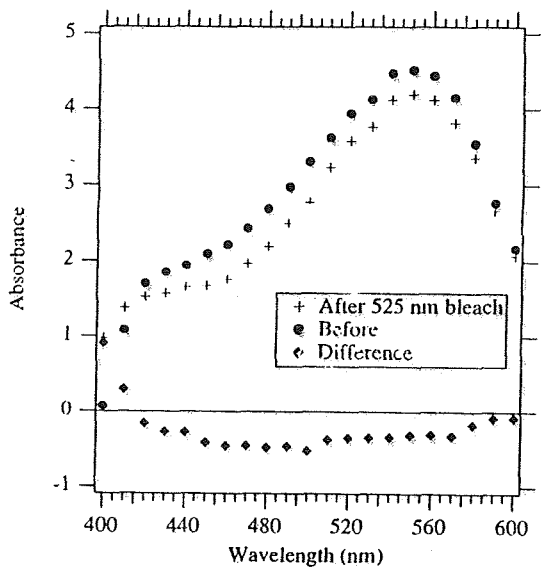


Fig. 3.3. Absorbance spectra before and after creation of M-state population in the bacteriorhodopsin film and their difference.

Fig. 3.4 shows the recovery dynamics of the bR-state absorption near 550 nm. The solid symbols show the absorbance normalized to its maximum value as a function of time spent in the dark after completion of a bleaching transient. A half life of about 10 s is observed for this recovery. The open symbols show the absorbance as a function of time spent under 450 nm illumination. These

data reveal the well known effect that wavelengths coincident with M-state absorption hasten M-state depopulation [4]. Since schlieren results are taken with blue light, Fig. 3.4 shows that the read light should be attenuated if long read times are required.

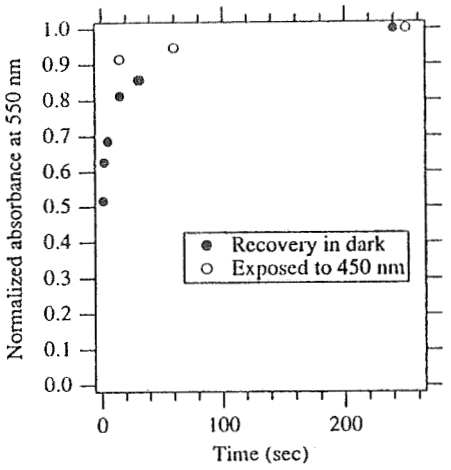


Fig. 3.4. Recovery dynamics of bacteriorhodopsin ground-state absorption in the dark and with blue illumination.

Fig. 3.5 compares the efficiency of different bleaching-wavelengths on the 400 nm absorbance change vs. intensity. Long-pass filters for writing and a 400 nm band-pass filter for reading were used. The shorter wavelengths included with the 475 nm long-pass filter result in a smaller change at 400 nm. These shorter wavelengths may be absorbed by the tail of the M-state absorption, thus helping to reduce its population.

Fig. 3.3 suggests that the largest absorbance increases occur at 400 nm or below. However, wavelengths below 400 nm should give only smaller changes[4]. This was tested using 550 nm long-pass filter and a number of blue-violet band pass filters. Fig. 3.6 presents the results, where the read wavelength is given for each solid symbol. The largest absorbance increases are confirmed to occur near 400 nm.

Figs. 3.5 and 3.6 reveal that the best filter combination for the schlieren experiments is the 550 nm long-pass for writing the image grid and the 400 nm band pass for reading it. The largest expected absorbance change at the read wavelength is about 1.3. An additional observation is the appearance of an optimum intensity well below the film's damage threshold. This phenomenon is unexplained by the simple two state model but is consistently observed.

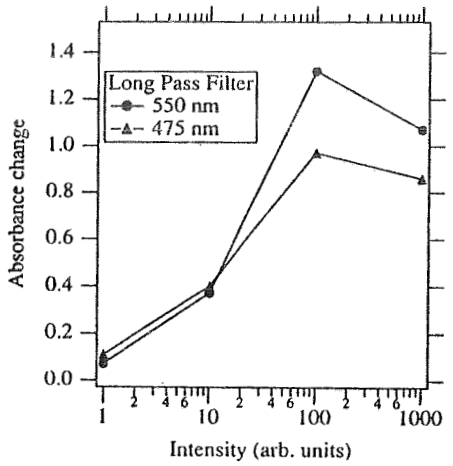


Fig. 3.5. Absorbance change at 400 nm as a function of intensity for different write filters.

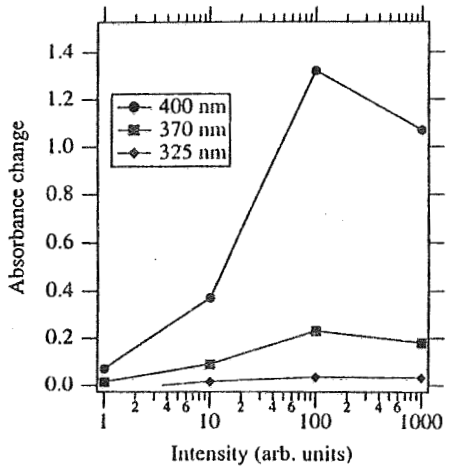


Fig. 3.6. Absorbance change at various read wavelengths as a function of intensity using a 550 nm long pass as the write filter.

Fig. 3.7 presents BR schlieren results for a low velocity flow of He gas. Turbulence is observed. Identifiable regions are darker on the left side and lighter on the right since schlieren is sensitive to index gradients which change sign when going across symmetric phase objects. For reprographic purposes, this image was processed with a ramp filter to enhance the contrast. Original digital data can be obtained by contacting the authors. In this experiment, a 1:1 image of the source grid was bleached into the BR film at an intensity sufficient to cause an absorbance change of 0.8 at 400 nm. The bleaching time was 10 s. The phase object was absent from the test region while writing the image grid. Then read filter and the phase object were simultaneously inserted, and an image of the test region was immediately recorded. Real time images of the test region were observed on a video monitor. While observing in the blue, the picture gradually brightened, and the contrast of the phase objects gradually faded to invisibility after several minutes. Then a reference image was collected.

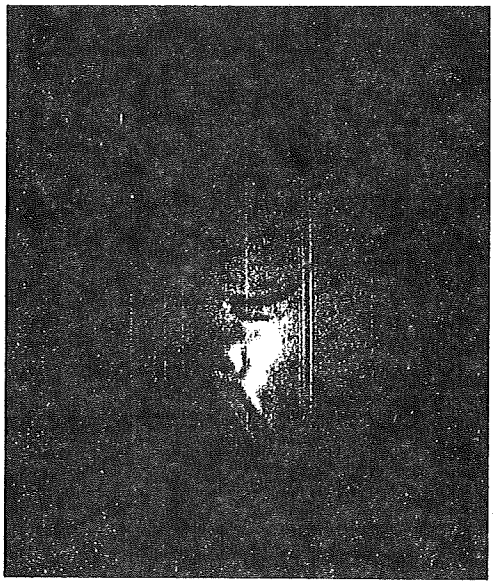


Fig. 3.7. Bacteriorhodopsin schlieren image of a low velocity He gas plume.

Fig. 3.8 presents an intensity profile across a portion of the unprocessed image taken immediately after the write procedure (lower heavy trace). It shows intensity variations of ~10% with spatial frequencies of a few tens of profile points. This is compared with a profile taken several minutes later (upper heavy trace). The long time profile reveals an overall increase in the transmission of the bacteriorhodopsin film and the nearly complete fading of the intensity variations that characterize the phase object. Also included is a profile taken with our standard schlieren set-up (faint trace). A strong slope to the baseline is observed because the phase object was located near

the edge of the field of view. The short period intensity variations are only about 2 times larger than the results obtained using BR. This suggests that BR schlieren can be as sensitive as standard schlieren but with the advantages of adaptability and automatic image-grid alignment.

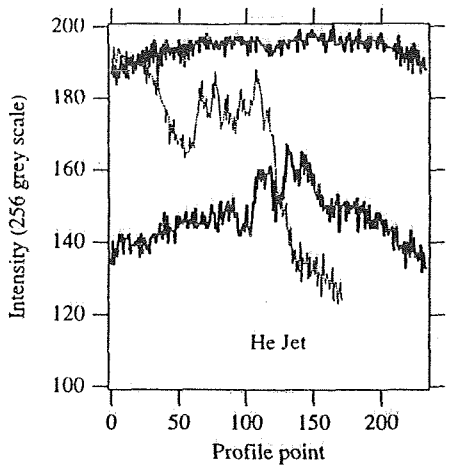


Fig. 3.8. Image profiles for the low velocity He plume. The thin curve is from standard schlieren results. Thick curves are from bacteriorhodopsin immediately after the write procedure (lower) and several minutes later (upper).

4. SUMMARY

An adaptive schlieren apparatus using bacteriorhodopsin films as the medium for writing image grids has been demonstrated using white light as the illumination source for the first time. An image a He gas plume revealed only 2 times less sensitivity than a standard schlieren apparatus.

5. REFERENCES

[1] Herbert Kaplan, "A Big Advance in Gas Imaging," Photonics Spectra, Feb. 1996, p 44-46.

[2] Selected Papers on Schlieren Optics, SPIE Milestone Series, vol. MS61, J. R. Meyer-Arendt, ed., (The International Society for Optical Engineering, Belingham, Wash., 1992).

[3] R. E. Peale and P. L. Summers, "Zebra schlieren optics for leak detection," Appl. Optics 35, x-y (1996).

[4] Robert R. Birge, "Protein-based optical computing and memories," Computer, November pp. 56-67 (1992).

[5] John D. Downie, "Application of bacteriorhodopsin films in an adaptive-focusing schlieren system," Applied Optics 34,6021-6028 (1995).

1996 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

MODELING ADVANCE LIFE SUPPORT SYSTEMS

Dr. Marvin Pitts, Associate Professor
Biological Systems Engineering Department
Washington State University
Pullman, Washington

KSC Colleagues - John Sager, Colleen Loader and Alan Drysdale
Life Sciences

Contract Number NASA-NGT10-52605

August 1, 1996

ABSTRACT

Activities this summer consisted of two projects that involved computer simulation of bioregenerative life support systems for space habitats. Students in the *Space Life Science Training Program* (SLSTP) used the simulation, SpaceStation, to learn about relationships between humans, fish, plants and microorganisms in a closed environment. One student, Ed Eaton, completed a six week project to modify the simulation by converting the microbes from anaerobic to aerobic, and then balancing the simulation's life support system. A detailed computer simulation of a closed lunar station using bioregenerative life support was attempted, but there was not enough known about system restraints and constants in plant growth, bioreactor design for space habitats and food preparation to develop an integrated model with any confidence. Instead of a completed detailed model with broad assumptions concerning the unknown system parameters, a frame work for an integrated model was outlined, and work begun on plant and bioreactor simulations. The NASA sponsors and the summer Fellow were satisfied with progress made during the 10 weeks, and we have planned future cooperative work.

Modeling Advance Life Support Systems

Marvin Pitts

1. Introduction

Scientists and engineers within the National Aeronautics and Space Administration (NASA) are conducting research which will lead to the development of advanced life-support systems that use plants, and microbes to solve long term life-support problems in space (1). The Controlled Ecological Life-Support System (CELSS) is a complex, extensively-controlled, bioengineered system for human life support. It relies on plants, and microbes to perform gas exchange and food and potable water reclamation to supply principal elements needed for human existence.

Information on gas exchange of plants in controlled environments is vital for assessing the use of crops for human life-support in closed space habitats. Fortson et al. (2) studied gas exchange and biomass production for wheat stands (*Triticum aestivum* L. c.v. Yecora Rojo) grown from planting to maturity in a 20 m^2 canopy area closed growth chamber. They determined that more biomass is needed to produce oxygen than to produce food grain.

A single-person demonstration unit was developed by Owens and Hall (3) to evaluate ecological processes and hardware requirements necessary to assess feasibility and to define design criteria. The system consisted of a 1 m^2 plant growth area, a 500 l fish culture tank, and computerized monitoring and control hardware. Nutrients in the hydroponic solution were derived from fish metabolites and fish food leachate.

Strayer (4) characterized the microbial constituents of the CELSS Biomass Production Chamber during production tests of hydroponically-grown crops of wheat and soybeans. Bacterial and fungal viable counts were determined for the hydroponic solution, dehumidifier condensate water, and atmosphere.

Corey and Wheeler (5) measured gas exchange in the Biomass Production Chamber for various combinations of plants and fish. They described the life-support needs for humans in space in their manuscript.

The computer simulation, SpaceStation©, uses simplified relationships and typical steady-state values defined by the CELSS researchers cited above. SpaceStation© biology consists of the organisms— plants, fish, people and microbes— that share four quantities: oxygen, carbon dioxide, water and organic waste. The SpaceStation© biology modeled in the program is highly interrelated.

In the simulation, the people consume oxygen, fish protein and cereal grain, and they produce carbon dioxide and organic waste (in water passed to the fish). The fish consume oxygen and cereal grain, and they produce carbon dioxide, organic waste (in water passed to the microbes) and protein. The microbes consume carbon dioxide and organic waste and produce oxygen and plant nutrients (in water passed to the plants). The plants consume carbon dioxide and plant nutrients and produce oxygen and cereal grain. The growth and/or death of any organism depends on the state of the other organisms in the system.

People and fish share the oxygen and cereal grain in the station, and both produce carbon dioxide and organic waste. Both people and fish will suffer (slower growth or death) if the quantities of oxygen or grain decrease too far, or if carbon dioxide or organic waste accumulate excessively. The interaction between people and fish is complicated by the predator / prey relationship.

The interaction between plants and microbes is less complicated. The plants and microbes share the carbon dioxide in the station. The plants are dependent on the microbes to convert organic waste into plant nutrients, and the animals that produce the organic waste depend on the plants for food.

The stated goal of the simulation is for the station's life-support manager to change food and water allocations between people, fish and plants to keep the system in balance when disruptions occur due to human arrivals and departures from the station. The simulation objectives are to:

1. maintain a suitable environment for people on the station
2. maintain suitable environments for fish, microbes and plants, and
3. minimize changes in fish, microbe and plant populations.

2. PROJECT DESCRIPTIONS AND RESULTS

Project 1: Using SpaceStation with SLSTP students (with Colleen Loader)

The Advance Life Support section of Biomedical Operations hosted eleven college students in the Space Life Science Training Program (SLSTP). Our goal was to introduce concepts related to system dynamics and engineering design in biological systems to the students.

Early in the six week program, I presented a two hour workshop on the relationship between people, fish, plants and microbes in a closed environment, in the use of SpaceStation, and gave a design exercise to determine how much crew fluctuation the station could accommodate and remain in ecological balance. Teams of three students were tasked to determine the maximum number of people who could enter or leave the station, and needed changes to the environmental control system. We did not get a chance to meet again as a group to hear team reports.

I also mentored Edward Eaton, a ceramic engineering junior at Rutgers University, in his student project. Ed changed the SpaceStation simulation by converting the microbes from anaerobic to aerobic. This change meant that the microbes began to consume oxygen rather than produce the gas. In his presentation to the SLSTP students, Ed listed key concepts he gained during the project:

- Biological systems are complex, time variant and nonlinear
- Different time constants in various organisms made balancing life support difficult
- Negative feedback control stabilizes populations
- Need to infer the unobservable from what is observable

Project 2: Simulation of a Lunar Station (with Alan Drysdale)

The second project I was involved with was to write a computer simulation of the biomass and related energy flows through a bioregenerative life support system for a lunar station. A lunar station was selected over the International Space Station (ISS), a Mars transit vehicle or a Mars station because the lunar mission duration and distance from Earth would preclude frequent resupply as is possible with the ISS, and yet is not as far in the future as a Mars mission. The goal of the project was to understand the dynamics of biomass and related energy transport in a closed system.

To meet this goal, we planned to develop a model based on the science developed in the CELSS Breadboard project. We would consult with the engineers and scientists working on CELLS projects to determine system parameters and constants. Once the model was completed, its predictions would be compared to other models, and the differences in model predictions would then give direction to the assumptions made in developing the different models.

Once we began to implement this plan, I soon realized that the integrated model envisioned was not possible at this time. Critical relationships in Plant and Bioreactor systems intended for space are not defined, or system parameters are not known. The engineers and scientists involved with CELSS (as well as myself) were skeptical that the integrated model would accurately estimate system dynamics. Conversation with Russ Fortson, Lockheed Corp. based at JSC, indicated that a model that was used by design engineers at KSC and JSC would be most useful. Consultation with engineers and scientists at JSC was needed to gain their support of the simulation. In addition, expertise in crew requirements and physical-chemical systems was at Johnson Space Center. Every indication we had was that an integrated simulation developed using only CELSS expertise and current knowledge would not be accepted or used.

Clearly, I needed to redefine my goals for the summer. My long term goal is to develop an integrated computer simulation of a lunar station that uses a combination of bioregenerative and physical-chemical life support systems. My revised goal for this summer was to lay the groundwork needed to build the integrated simulation. To meet this goal, I completed the following activities:

- Sketched a framework for an integrated model using modules to simulation major life support systems
- Began to define Resource Recovery and Plants modules using CELSS expertise
- Began to define Crew module and physical-chemical devices using JSC expertise (met with JSC engineers and scientists August 8th and 9th)

The framework I outlined consists of the following major groupings: crew, plants, resource recovery, food preparation, energy (electrical, mechanical and heat), water and other liquids, air and associated gasses, physical - chemical units and environmental control. Additional parts of the simulation used for design are: flow of carbon, nitrogen, other elements and organic compounds, and data sets of system constants for various station operation assumptions.

Plant Module: Plant growth and grain production are influenced by the following factors: light source and intensity, CO_2 concentration in the air, NO_3 and NH_4 concentrations in the plant nutrient solution, other major nutrients, water and evapotranspiration. Based on these relationships, I built three models: one which modeled the growth of a single set of plants of the same age and under the same environmental conditions, a second model which simulated a series of eleven sets of plants grown at weekly intervals, and a model of a proportional control system used to meter NH_4 into the plant nutrient solution to regulate solution pH.

The single set plant model is based on the following equation defined by Volk et. al, (6):

$$CGR = K[H*P - (24-H)R] \qquad 2.1$$

where CGR = crop growth rate
K = unit conversion constant
H = hours per day that growth lights are on
P = biomass accumulated through photosynthesis
R = biomass consumed by respiration when the growth lights are not on

The photosynthesis term is a function of the light energy absorbed by the plants, the conversion of light energy into plant mass, and the age of the plant. The respiration term is a function of the same factors. Curve 1 in Figure 2.1 is a plot of the wheat growth under the environmental conditions used in the CELSS Biomass Production Chamber (BPC) (7). The second and third curve in Figure 2.1 is a modification of equation 2.1 which partitions the plant biomass into vegetation and grain. After antithesis, most of the additional plant mass is channeled into grain fill. Two terms were defined in the simulation which control the mass partition: day of antithesis and potential grain fill. Day of antithesis is self explanatory. Potential grain fill is a term used to model the significant effects of environmental conditions during antithesis on the ability

of the plant to produce grain. There is a qualitative understanding of the effect of some environmental conditions on grain fill, but additional study is needed to quantify these effects. Figure 2.2 demonstrates the significant reduction in grain yield from delaying antithesis. Figure 2.3 demonstrates significant reduction in grain yield due to lower grain potential constants.

I expanded the single plant model into eleven sets of plants, each containing producing a weekly ration of grain for a crew of 4. The oldest plants set was harvested each week, and a new set planted. This model was used to demonstrate the effect of various environmental factors in common with all eleven sets on grain production. For example, Figure 2.4 shows the reduction in harvest index following a light system failure of 5 days. Even though the light failed on days 10 through 15, the most severe grain shortage was five weeks after the light disruption, and 16 weeks until grain production reached normal levels.

The third plant–related model was a proportional control of plant nutrient solution pH via the addition of NH_4. Plant nutrient uptake uses an ion exchange in the absorption of nitrogen based molecules. The primary nitrogen source for plants grown in hydroponic systems is NO_3. For each mole of NO_3 the plant uptakes, it releases 2 moles of OH. Over time, the pH of the solution will increase to toxic levels due to the increasing concentration of OH ions. Traditionally, pH is corrected via the addition of nitric acid (7). Nitric acid is not produced in a closed system such as a lunar station, and must be supplied from external sources. In a closed system that includes animals, there is a constant source of ammonia from animal waste. Plants readily absorb ammonia, in fact at a rate twice that for nitrate (8). Ammonia uptake must be limited because plants release 4 moles of H ions per mole of ammonia (which would drive the nutrient solution pH down), and because excess absorption of ammonia is toxic to plants (maximum $NH_4 : NO_3$ ratio is 0.5). Figure 2.5 demonstrates that a proportional control based on pH levels can maintain the correct chemistry for plant growth without the addition of nitric acid to the plant nutrient solution.

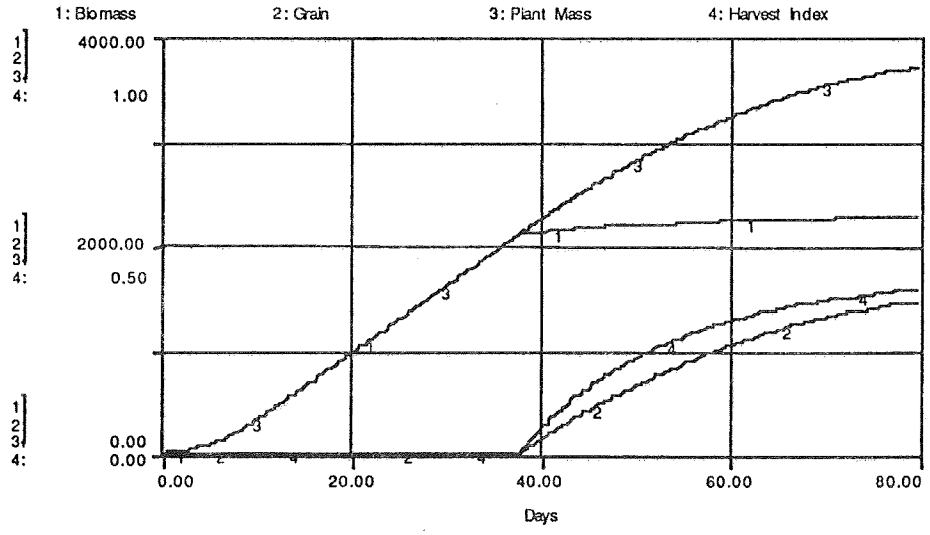


Figure 2.1 Graph of plant growth using equation developed by Volk et. al. for wheat grown under environmental conditions similar to the biomass production chamber used in the NASA CELSS project. Curve 1 is plant vegetative mass; Curve 2 is grain mass, Curve 3 is the total plant mass (vegetative and grain), and Curve 4 is the Harvest Index (grain mass / total plant mass).

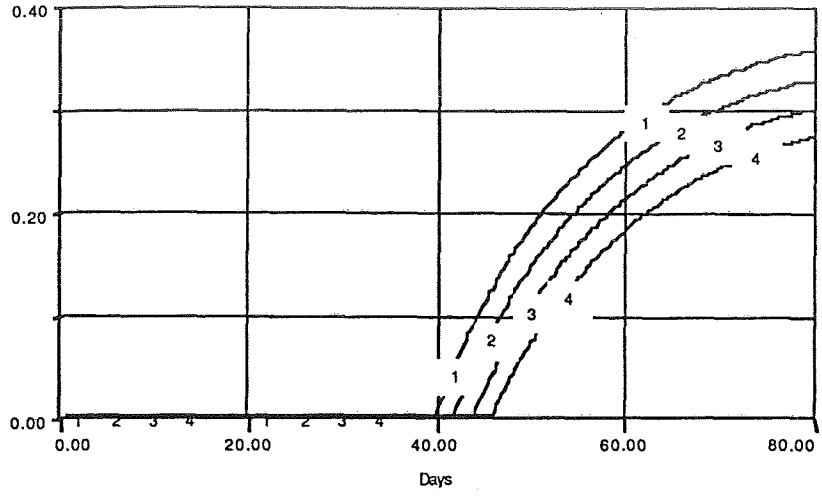


Figure 2.2. Effect of day of antithesis on Harvest Index using the equation developed by Volk et. al. for wheat grown under environmental conditions similar to the biomass production chamber used in the NASA CELSS project. Day of antithesis were 40 (Curve 1), 42 (Curve 2), 44 (Curve 3) and 46 (Curve 4).

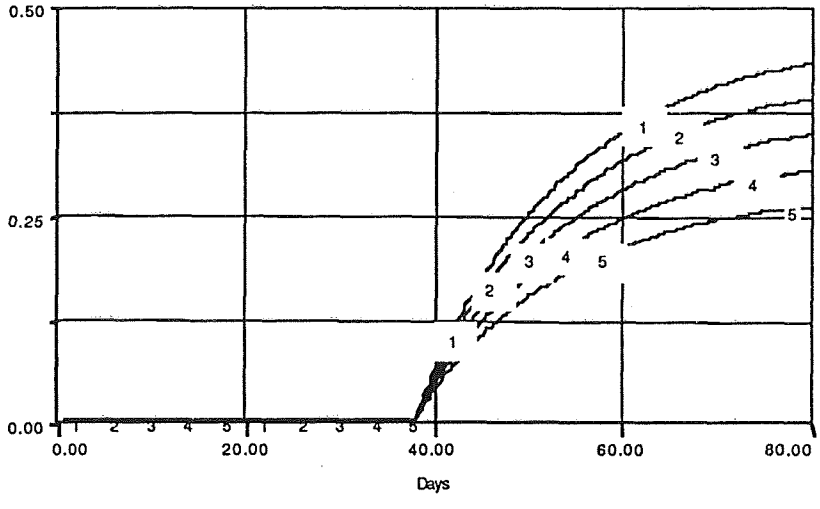


Figure 2.3. Effect of "potential grain fill" on Harvest Index using the equation developed by Volk et. al. for wheat grown under environmental conditions similar to the biomass production chamber used in the NASA CELSS project. Potential grain fill values were 1 (Curve 1), 0.9 (Curve 2), 0.8 (Curve 3), 0.7 (Curve 4) and 0.6 (Curve 5).

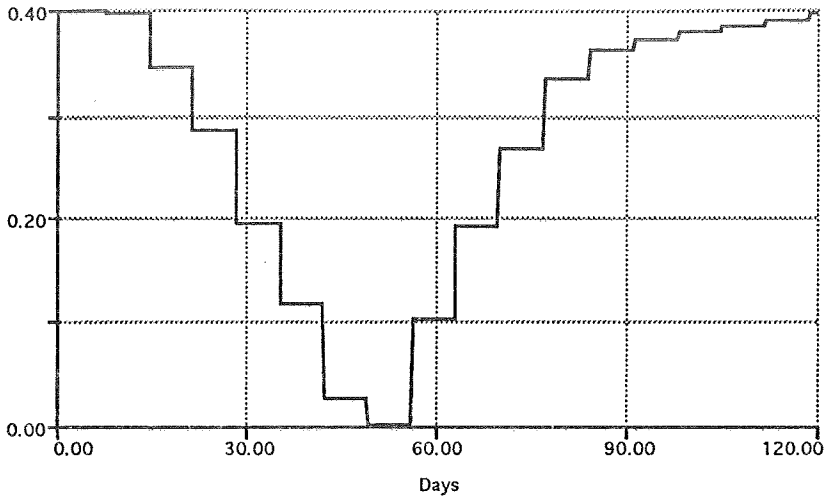


Figure 2.4. Effect of light failure in a staggered growing system on Harvest Index using the equation developed by Volk et. al. for wheat grown under environmental conditions similar to the biomass production chamber used in the NASA CELSS project. A weekly harvest based on eleven plant sets were modeled. Lights were off from Day 10 to Day 15.

Curve 1 shows a characteristic finite error in pH due to the proportional control. The rise in pH would not affect plant growth. The ratio of NH_4 : NO_3 remains below 0.50. Rising pH from OH produced by nitrate uptake is first controlled by the buffer in the solution, and as ammonia is released into the solution in response to the rise in pH, the H ions produced by the ammonia uptake combine with OH ions to produce water molecules.

Bioreactor modeling: Waste Recovery in a lunar station has different goals than existing wastewater treatment plants in regards to nitrogen. Terrestrial wastewater treatment encourages microbe growth because the microbes bind organic carbon and nitrogen, reducing BOD and nitrate levels in the effluent. The microbes are removed from the treatment plant and disposed via burial, incineration or spread onto agricultural fields (9). In a lunar station, these elements must be recycled. While incineration will release the carbon in the microbes in a useful form as carbon dioxide, the nitrogen in the microbes will be released as nitric oxides, requiring further treatment to convert the nitrogen into compounds suitable as plant nutrient. Thus, a design goal for waste recovery in a lunar station is to reduce organic carbon and nitrogen in waste streams without adding nitrogen to untreated biomass that will be incinerated.

Many of the constants needed to model a bioreactor are system specific, and there is not a design procedure which will estimate these parameters for a lunar–based bioreactor. However, I developed a fully–mixed, continuous flow bioreactor model using typical values for demonstration purposes to indicate how a lunar–bioreactor might be operated. Figure 2.6 indicates the effect of microbe mass on the amount of nitrogen that is in the untreated biomass (organic waste that was not treated plus microbes that were removed from the bioreactor). For this system, an optimum microbe mass is indicated by curves 2 and 3 which have the minimum amount of untreated nitrogen. Low microbe levels (curve 1) cannot consume all the organic matter in the waste, allowing untreated organic matter out of the bioreactor. High microbe levels (curves 4 and 5) result in a large amount of microbes removed from the bioreactor.

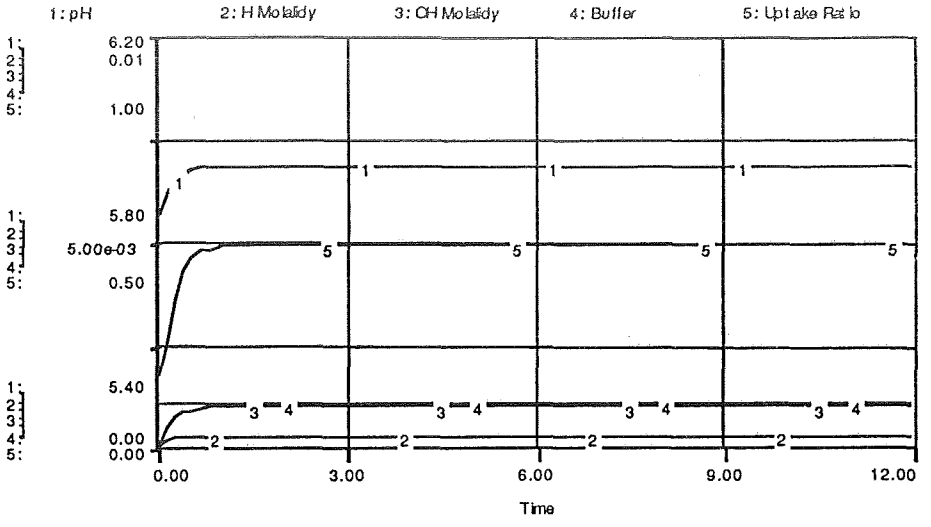


Figure 2.5 Modeled nutrient chemical response to proportional control of pH via the addition of ammonia.

Few Bioreactor design constants or processes exist for lunar station based waste recovery systems. Some of the unknown values I needed for modeling were:

- Microbe growth rate
- Substrate consumption
- How to maximize consumption while minimizing microbe growth?
- How to scale bioreactor volume and flow rate?
- How these values change in alternate bioreactor designs?

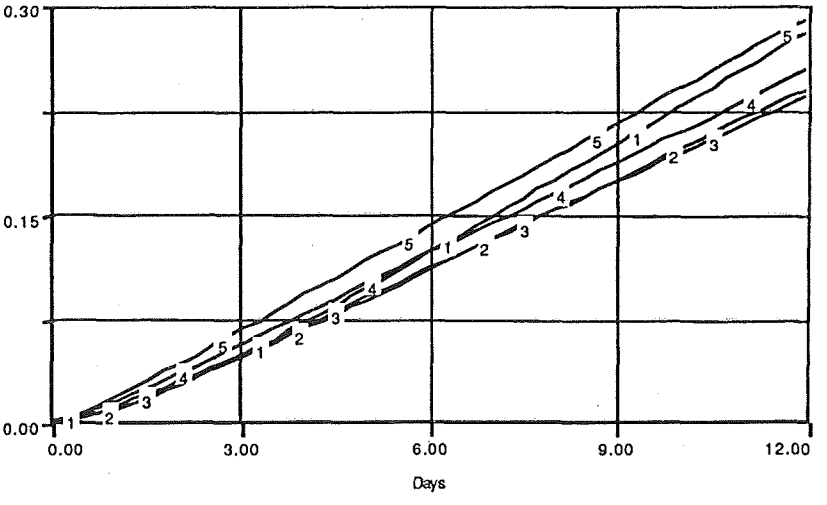


Figure 2.6 Effect of microbe mass in a continuous flow, fully mixed bioreactor on nitrogen level in effluent biomass (untreated organic carbon and microbes removed from the reactor). Microbe mass increases by a 0.1 kg from Curve 1 (0.1 kg) to Curve 5 (0.5 kg).

3. CONCLUSIONS AND RECOMMENDATIONS

During the ten week summer program I developed the following products:

- An improved SpaceStation model
- Workshop task sheets for use with SpaceStation with emphasis on engineering design in biological systems
- Framework for a model of integrated life support systems, and
- Prototype Bioreactor and Plant modules

During the summer, a number of needs surfaced which require further cooperative work. The most urgent areas are to:

- Determine relationships that are oversimplified
- Verify trends pointed out in the prototype models, and
- Determine constants needed in the models

4. REFERENCES

1. Chamberland, D., W. M. Knott, J. C. Sager, and R. M. Wheeler. 1992. Controlled Ecological Life-Support System. *J. Florida M.A.* 79(8): 537-544.

2. Fortson, R. E., G. W. Stutte, J. C. Sager, and R. M. Wheeler. 1993. CELSS Reliability and Plant Response to Environment Stress. ASAE Paper #93-4006.

3. Owens, L. P. and C. R. Hall. 1989. Biomass Production and Nitrogen Dynamics in an Integrated Aquaculture/Agricultural System. Editor: Robert D. MacElroy. Controlled Ecological Life Support Systems (CELLS) '89 Workshop.

4. Strayer, R. F. 1991. Microbiological Characterization of the Biomass Production Chamber During Hydroponic Growth of Crops at the Controlled Ecological Life Support System (CELSS) Breadboard Facility. 21st International Conference on Environmental Systems, July 15, pp. 35-48.

5. Corey, K. A. and R. M. Wheeler. 1992. Gas Exchange in NASA's Biomass Production Chamber. *BioScience* 42(7): 503-509.

6. Volk, T., B. Bugbee and R. Wheeler. 1995. An Approach to Crop Modeling with Energy Cascade. Life Support and Biosphere Sci. (1):119-127.

7. Wheeler, R., W. Berry, C. Mackowiak, K. Corey, M. Heab and W. Knott. 1993. A database of Crop Nutrient Use, Water Use and Carbon Dioxide Exchange in a 20 m^2 growth Chamber. I. Wheat as a Case Study. Journal of Plant Nutrition 16(10): 1881-1915.

8. Raven, J. 1986. Biochemical disposal of excess H+ in growing Plants. New Phytol. 104: 175-206.

9. Tchobanoglous, G. and F. Burton. 1989. Wastewater Engineering Treatment, Disposal and Reuse. ISBN 0-07-04169-7. McGraw-Hill. pp 74-75

1996 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

A FEASIBILITY STUDY ON A PARALLEL MECHANISM FOR EXAMINING THE SPACE SHUTTLE ORBITER PAYLOAD BAY RADIATORS

Dr. Rodney G. Roberts, Assistant Professor
Electrical Engineering Department
Florida State University
Tallahassee, Florida

KSC Colleague - Eduardo Lopez del Castillo
Robotics

Contract Number NASA-NGT10-52605

August 9, 1996

ACKNOWLEDGMENTS

The author would like to express his appreciation to Todd Graham and Eduardo Lopez del Castillo for an interesting problem to work on. Thanks also goes to Jefferson M. Trumpower of Florida State University for the development of the MATLAB routines, and to Carey Cooper of Intergraph Corporation for his help with the IGRIP simulations. Special thanks goes to Roger Johnson, Gregg Buckingham, and Kari Stiles.

ABSTRACT

This report summarizes the author's work as a 1996 NASA/ASEE Summer Faculty Fellow at Kennedy Space Center. The goal of the project was the development of the necessary analysis tools for a feasibility study of a cable suspended robot system for examining the Space Shuttle orbiter payload bay radiators. These tools were developed to address design issues such as workspace size, tension requirements on the cables, the necessary accuracy and resolution requirements, and the stiffness and moment requirements of the system. The report describes the mathematical models for studying the inverse kinematics, statics, and stiffness of the robot. Each model is described by a matrix. The manipulator Jacobian matrix characterized both the inverse kinematics and the statics of the robot. The manipulator Jacobian was also related to the stiffness matrix, which characterized the stiffness of the system. Analysis tools were then developed based on the singular value decomposition (SVD) of the corresponding matrices. It was demonstrated how the SVD can be used to quantify the robot's performance and to provide insight into different design issues.

1. INTRODUCTION

The Space Shuttle orbiter payload bay radiators are inspected before and after each mission in the Orbiter Processing Facility (OPF). These inspections are labor intensive and require access to overhead bridge crane buckets. The buckets are platforms that are attached to a telescoping tube which is attached to the overhead bridge crane trolley. Technicians ride in the buckets and visually inspect the radiator panels for defects. When defects are found, they need to be quantified and their location logged.

The inspection of the orbiter's radiators is an ideal candidate for automation. The use of robotics for the inspection of payload bay radiators will eliminate the use of the buckets, reduce coordination effort and paper work required, reduce inspection time, and increase personnel and equipment safety. Furthermore, a robotic system can prepare electronic maps of the radiator's surface and automate the generation of problem reports. Robotic concepts have been developed; however, these concepts need to be rigorously evaluated for technical feasibility.

McDonnell Douglas Aerospace has developed a series of unique tendon suspended robots under the trademark CHARLOTTE™ [1,2]. The robot translates and rotates on and about three axes by synchronously modulating the length of the suspending tendons. Charlotte™ was originally designed for attending Space Shuttle experiments and has flown on two Space Shuttle missions. Another version of Charlotte™ supports virtual reality simulations of extra-vehicular activities at the robotics laboratories of NASA's Johnson Space Center.

This report will consider the feasibility of using a cable suspended robot such as Charlotte™ for the radiator inspections. First, mathematical models for a cable suspended robot were developed. These models were used to study the inverse kinematics, the required cable tensions, and the stiffness of the robot. Analysis tools were then developed to address design issues such as workspace size, tension requirements on the cables, the necessary accuracy and resolution requirements, and the stiffness and moment requirements of the system. The report describes the mathematical models for studying the inverse kinematics, statics, and stiffness of the robot. Each model is described by a matrix. The manipulator Jacobian matrix characterized both the inverse kinematics and the stiffness of the robot. The manipulator Jacobian was also related to the stiffness matrix, which characterized the stiffness of the system. Analysis tools were then developed based on the singular value decomposition (SVD) of the corresponding matrices. It was demonstrated how the SVD can be used to quantify the robot's performance and to provide insight into different design issues.

2. MODELING THE ROBOT

2.1 INVERSE KINEMATICS

The robot is basically a box with cables attached to its eight vertices. Let the location of the i-th external cable connection be denoted by $\mathbf{p}_i$ and the i-th onboard cable connection be denoted $\mathbf{q}_i$. The vector representing the i-th cable is

$$\mathbf{l}_i = \mathbf{p}_i - \mathbf{q}_i = \mathbf{p}_i - \mathbf{x} - Q\mathbf{v}_i$$

where $\mathbf{x}$ denotes the position of robot and the rotation matrix

$$Q = R_z(\phi)R_x(\theta)R_y(\psi) = \begin{bmatrix} c\phi c\psi - s\phi s\theta s\psi & -s\phi c\theta & c\phi s\psi + s\phi s\theta c\psi \\ s\phi c\psi + c\phi s\theta s\psi & c\phi c\theta & s\phi s\psi - c\phi s\theta c\psi \\ -c\theta s\psi & s\theta & c\theta c\psi \end{bmatrix}$$

represents the robot's orientation. A similar analysis can be found in [3].

2.2 INVERSE VELOCITY KINEMATICS

The inverse velocity kinematics relates the linear and angular velocities of the robot to the required rate of change in the cable lengths. This relationship is given by a manipulator Jacobian matrix L which is a function of the position and orientation of the robot. This matrix can be used to study the kinematics of the robot, e.g., identify kinematically singular configurations, determine optimal configurations, etc.

Let $\mathbf{l}$ denote the vector of cable lengths. The inverse velocity kinematics for Charlotte™ was derived and is given by

$$\dot{\mathbf{l}} = L\begin{bmatrix} \dot{\mathbf{x}} \\ w \end{bmatrix}$$

where

$$L = \begin{bmatrix} -\hat{\mathbf{l}}_1^T & (\hat{\mathbf{l}}_1 \times Q\mathbf{v}_1)^T \\ -\hat{\mathbf{l}}_2^T & (\hat{\mathbf{l}}_2 \times Q\mathbf{v}_2)^T \\ \vdots & \vdots \\ -\hat{\mathbf{l}}_8^T & (\hat{\mathbf{l}}_8 \times Q\mathbf{v}_8)^T \end{bmatrix}.$$

2.3. STATICS MODEL

For the robot to be in equilibrium, the sum of the forces and the sum of the moments on the robot must be zero. Mathematically, this can be written as

$$\sum_a \mathbf{F}_a = \mathbf{0}$$

$$\sum_a \mathbf{M}_a = \mathbf{0}.$$

From these relations, it can be shown that the forces on the cables satisfy

$$\mathbf{AF} = \mathbf{W}$$

where $\mathbf{F}$ is the vector of cable tensions, $\mathbf{W}$ is the vector consisting of the gravity term and moment of the robot, and $\mathbf{A} = -\mathbf{L}^T$.

The general solution of the force/moment equations is

$$\mathbf{F} = A^+\mathbf{W} + (I - A^+A)\mathbf{z}.$$

where A^+ is the pseudoinverse of A and $\mathbf{z}$ is an arbitrary vector. This equation gives a two dimensional family of solutions. However, since there are cables involved, each component of the force vector must be nonnegative. A configuration will be called statically stable if such a

force vector exits. Intuitively, it would seem that the pseudoinverse solution would result in an unfeasible solution so the null space term $(I - A^{+}A)\mathbf{z}$ is necessary.

One may want to choose a solution that solves the following optimization problem:

$$\text{minimize} \quad \|\mathbf{F}\|_{\infty}$$
$$\text{subject to} \quad F_i \geq B_i$$

where B_i is some positive bound to insure that the cable does not become slack.

2.4 MODELING STIFFNESS

For the stiffness study, the cables were assumed to be under sufficient tension to be modeled as a spring. The stiffness for the i-th cable is given by

$$k_i = \frac{EA}{l_i}$$

where E is Young's modulus of elasticity, A is the cross sectional area of the cable, and l_i is the cable length. The stiffness matrix of the robot is given by

$$K = L^T K_D L$$

where $K_D = \text{diag}(k_1, k_2, \ldots, k_8)$. The stiffness matrix is symmetric and positive semi-definite. A similar analysis can be found in [4].

3. SOME IMPORTANT KINEMATIC AND STATICS ISSUES

3.1 KINEMATICALLY STABLE CONFIGURATIONS

It is desirable that each motion of the robot result in at least one cable length getting longer and at least one getting shorter. Otherwise there is a motion that results in the cable lengths getting shorter. This could result in cables becoming slack. Kinematic instability is also related to the stiffness of the robot. We will say that a configuration is kinematically unstable if there is a motion that results in some of the cables becoming shorter with no cable becoming longer.

The following theorem, which characterizes kinematic stability by the left null space of L, was proven.

THEOREM A nonsingular configuration is kinematically stable if and only if there is a left null vector of L with the property that each of its components is positive.

This theorem is very useful for identifying kinematically unstable configurations and is easily implemented numerically. An example of a kinematically unstable configuration for the planar version of Charlotte™ is shown in Figure 1. Rotating the robot counter-clockwise with its center fixed results in each cable becoming shorter. One can sense from the figure that this configuration is not an equilibrium. The question of whether such configurations can be equilibrium points will be answered in the next subsection.

3.2 STATICALLY STABLE CONFIGURATIONS

An important question is the existence of a feasible force vector. We say that a configuration is statically stable if there is a force vector whose components are positive. Clearly, if there is a null vector whose components are positive, we can add enough of this vector to the pseudoinverse solution to obtain a force vector with positive components. Hence, we have the following.

THEOREM If there is a null vector of L with strictly positive components then the robot is in a statically stable configuration.

We thus have that a kinematically stable configuration is statically stable. However, a kinematically unstable configuration may possibly be statically stable depending on effects of gravity. For example, it is possible to hold the robot in static equilibrium while the robot is under all of the exterior cable connections. Such a configuration then is statically stable but not kinematically stable due to a pendulum effect. An example of a statically stable configuration which is kinematically unstable is the situation when the robot is lower than all eight cable connections. Below is an example of a statically unstable configuration for a planar version of Charlotte™.

EXAMPLE Consider the robot shown in Figure 1. Assume that the center of gravity is at the center of the robot and that the weight of the robot is 1 N. The outer cable connections are at (2,2), (2,-2), (-2,-2) , and (-2,2) and the onboard cable connections are at (1,0) , (0,-1) , (-1,0) , and (0,1) . The $\hat{\mathbf{l}}_i$ vectors are

$$\hat{\mathbf{l}}_1 = \frac{1}{\sqrt{5}}\begin{bmatrix}1\\2\end{bmatrix},\ \hat{\mathbf{l}}_2 = \frac{1}{\sqrt{5}}\begin{bmatrix}2\\-1\end{bmatrix},\ \hat{\mathbf{l}}_3 = \frac{1}{\sqrt{5}}\begin{bmatrix}-1\\-2\end{bmatrix},\ \text{and}\ \hat{\mathbf{l}}_4 = \frac{1}{\sqrt{5}}\begin{bmatrix}-2\\1\end{bmatrix}.$$

It then follows that the forces and moments satisfy

$$\frac{1}{\sqrt{5}}\begin{bmatrix}1 & 2 & -1 & -2\\ 2 & -1 & -2 & 1\\ 1 & 1 & 1 & 1\end{bmatrix}\mathbf{F} = \begin{bmatrix}0\\1\\0\end{bmatrix}.$$

The general solution is

$$\mathbf{F} = \begin{bmatrix}0.2\\-0.1\\-0.2\\0.1\end{bmatrix} + \alpha\begin{bmatrix}1\\-1\\1\\-1\end{bmatrix}$$

where the first vector on the right is the pseudoinverse, the second vector is a null vector, and α is an arbitrary constant. It is necessary to choose α so that each component of $\mathbf{F}$ is positive. This is clearly impossible as choosing α to make the third component positive makes the second component more negative. One can see from the figure that this configuration is not kinematically stable as rotating the robot counterclockwise with its center fixed would correspond to having each of the cables getting shorter.

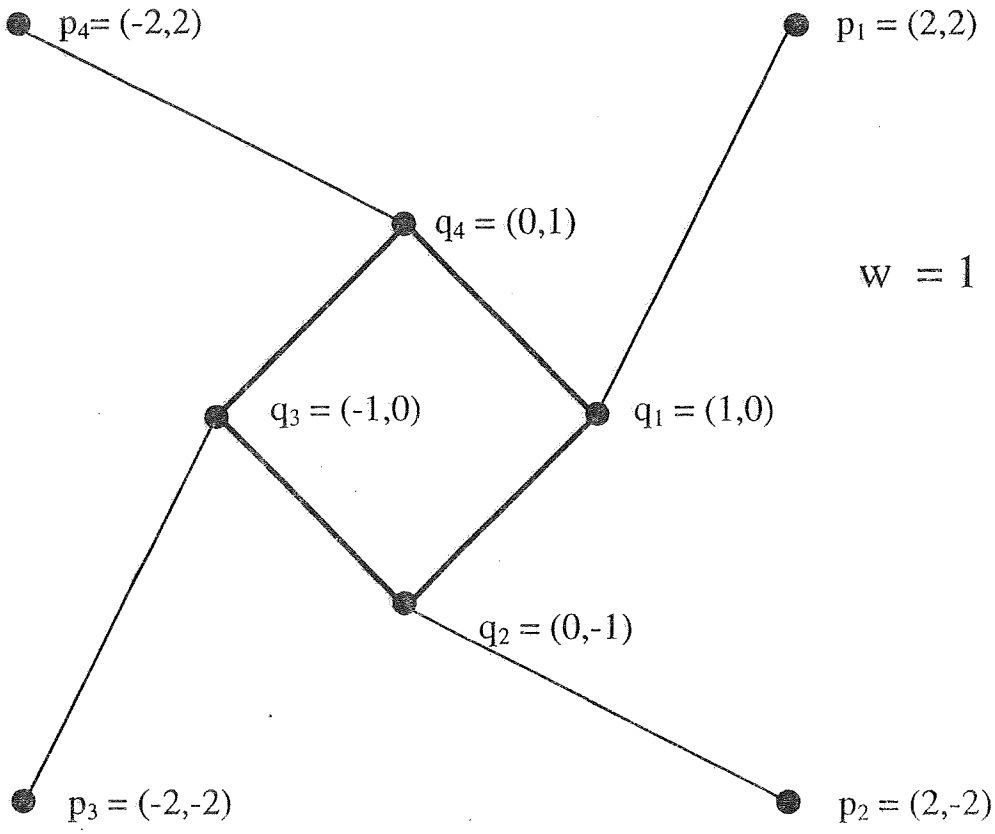


Figure 1 A configuration that is kinematically and statically unstable.

4. ANALYSIS USING THE SINGULAR VALUE DECOMPOSITION

The singular value decomposition (SVD) is a matrix decomposition that provides insight into the input/output relationship of a matrix [5]. For example, it identifies which input vectors result in large output vectors and which ones result in small output vectors. Applied to the manipulator Jacobian, the SVD tells us the amount of change in cable lengths necessary to produce a given robot motion . This is important for analyzing the amount of resolution/accuracy required for the cable lengths. From the SVD we can determine the robot motion that results in the least change in the cable lengths. This tells us the direction in which the robot is most flimsy.

The SVD determines the null space of a matrix. This then tells us whether or not a configuration is statically stable. This is useful information for determining workspace limitations. Stiffness is also an important design issue. The SVD of the stiffness matrix tells us the directions in which

the robot is most stiff and least stiff and gives us the values of stiffness in these directions. This information can help in the choice of cable material.

Several programs using MATLAB were developed using the SVD to evaluate performance capabilities of the robot. The programs and results were provided to the NASA colleague.

5. RESULTS AND DISCUSSIONS

The goal of this the project was the derivation of the mathematical models for a cable suspended robot and the development of the analysis tools to evaluate the performance of the robot. The analysis tools were based on the singular value decomposition and are applicable to a large class of cable suspended robot systems. Computer software in the form of .m files was written to use a widely available commercial program called MATLAB. These programs will provide a basis for the evaluation and the design of a cable suspended robot system for the Space Shuttle orbiter payload bay radiators.

REFERENCES

[1] P. D. Campbell, P. L. Swaim, and C. J. Thompson, "Charlotte Robot Technology for Space and Terrestrial Applications," SAE Technical Series paper #951520.

[2] F. Eichstadt, P. Campbell, T. Haskins, "Tendon Suspended Robots: Virtual Reality and Terrestrial Applications," SAE Technical Series paper #951571.

[3] J. Albus, R. Bostelman, and N. Dagalakis, "The NIST ROBOCRANE," Journal of Robotic Systems, 10(5), 709-724 (1993).

[4] N. G. Dagalakis, J. S. Albus, B. L. Wang, J. Unger, and J. D. Lee, "Stiffness study of a parallel link robot crane for shipbuilding applications," Trans. of the ASME J. of Offshore Mechanics and Artic Engineering, 111, 183-193 (1989).

[5] R. A. Horn and C.R. Johnson, "Matrix Analysis," Cambridge University Press, 1985.

1996 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

LEAKAGE THROUGH A CHANNEL FORMED BY A GASKET, A SEALING SURFACE, AND A FILAMENT TRAPPED BETWEEN THEM

Dr. John M. Russell, Associate Professor
Aerospace Engineering Program
Florida Institute of Technology
Melbourne, Florida

KSC Colleague - Frederick Adams
Instrumentation and Hazardous Gas Monitoring

Contract Number NASA-NGT10-52605

August 9, 1996

ABSTRACT

Much critical equipment at Kennedy Space Center is plumbing for the transport of liquid Hydrogen or liquid Oxygen. Every piece of hardware for handling such a *hazardous cryogen* is subject to testing prior to installation and use.

Safe, realistic testing of all such hardware is prohibitively expensive, which leads, perforce, to expedients, such as: (1) Leak testing with non-flammable tracer fluids (*e.g.* liquid Nitrogen); and (2) Leak testing with room temperature tracer fluids (*e.g* gaseous Helium).

Such expedients undermine the realism of the tests. If, however, one could apply rational fluid dynamics methods to derive an general analytical expression with which one could relate the throughput of gaseous Helium through a given leak channel to the throughput of liquid Hydrogen through the same channel then one could recover much of the information that one would otherwise forfeit through these expedients. These facts lead to following questions: (1) What would be an example of a generic flaw in a gasket?; and (2) How can one calculate the flow of fluid it?.

The following report addresses these questions. It considers a particular leak geometry, namely one formed by a gasket, a sealing surface, and a filament trapped between them (so that the cross section of the leak channel is a *flat bottomed curvilinear triangle*, two sides of which are circular arcs and which has cusps on all three corners).

LEAKAGE THROUGH A CHANNEL FORMED BY A GASKET, A SEALING SURFACE, AND A FILAMENT TRAPPED BETWEEN THEM

JOHN M. RUSSELL

1. INTRODUCTION

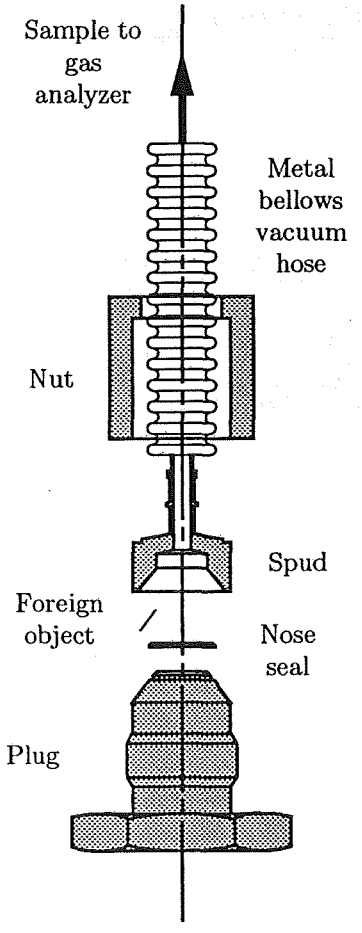


Figure 1

Consider an apparatus of the form illustrated in Figure 1, at left. One end of the metal bellows hose is plugged and the other end is attached to a residual gas analyzer capable of measuring small partial pressures of tracer fluid. The hose is evacuated and the plugged end is subject to any of a variety of test conditions, such as: (1) Immersion in a bath of liquid Nitrogen; (2) Exposure to cold gaseous Nitrogen by suspension above the liquid level in a liquid Nitrogen Dewar; (3) Exposure to gaseous Nitrogen at room temperature; (4) Exposure to gaseous Helium at room temperature. In this way, one may measure the leakage of tracer fluid that leaks from the outside of the plugged end of the hose to the inside.

If the seal on the plug were flawless no tracer fluid would leak. One may, however, simulate the kind of flaw that may arise in practice by draping a fine wire or other hard filament across the nose seal prior to assembly (see Figure 1, where the filament is labeled "foreign object").

One would like to be able to model the flow through a leak channel of the above sort, but to do so, one must describe the shape of the resulting leak channel in detail. To this end, consider what a hypothetical observer might see if he or she could observe the filament along a line of sight parallel to the filament itself during assembly of the apparatus (see Figure 2 on the next page).

I have illustrated the spud and the plug as jaws of a vise. I have illustrated the nose seal in light shading. As the soft jaws close over the assembly, the soft seal material gives way to the hard filament as illustrated in Panel (b). One may surmise that the wire crevice will be effectively flat bottomed and that the arc of the crevice corresponding to the wire will be circular. There is no physical reason to assume that the arc corresponding to the gasket is also circular. The analysis simplifies considerably under this assumption, however, so I will make it in the remainder of this work. I will also use the term *wire crevice* to denote the kind of curvilinear triangle (with a flat bottom, two circular arcs, and cusps in the corners) illustrated in Panel (c) at the bottom of Figure 2.

The remainder of this report is an analysis of the flow through a single wire crevice. Section 2 presents the necessary scientific background on the flow of liquids and gases through capillaries. Section 3 formulates the theory appropriate to the wire crevice. Section 4 presents the main result, namely a set of contours of

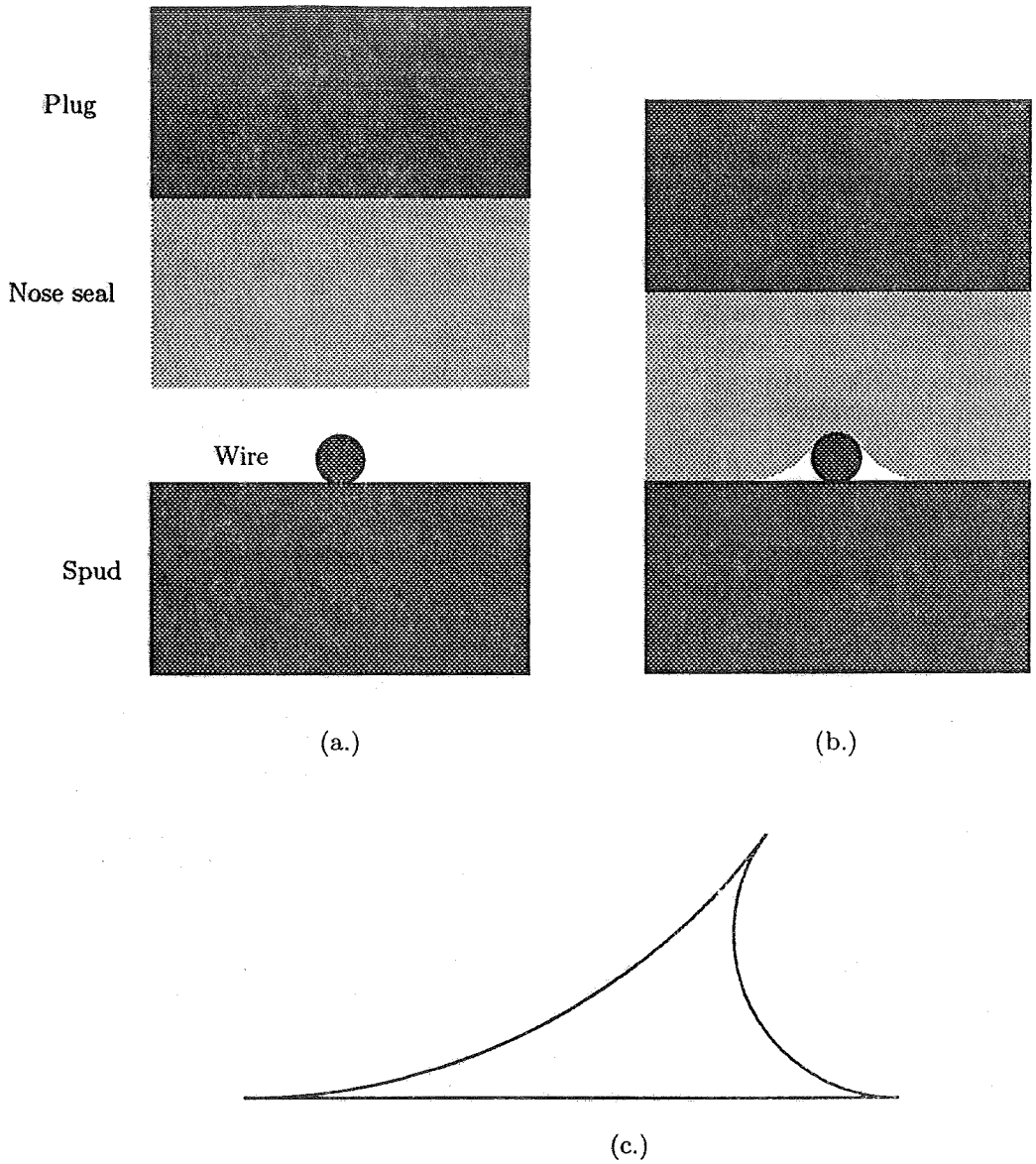


Figure 2. Generation of a wire crevice. Panel (a) represents the placement of the plug, nose seal, wire, and spud prior to assembly. Panel (b) represents these components after assembly. Panel (c) is a detail of the left member of the pair of leak channels illustrated in white in Panel (b).

constant streamwise velocity (in nondimensional variables) for the flow in question, and Section 5 presents the conclusions and recommendations.

2. SCIENTIFIC BACKGROUND

Authors of textbooks on fluid mechanics assign pride of place to the classical problem of *fully-developed laminar flow* through a pipe or channel and this problem applies to the present case when the fluid is a liquid. Many of the assumptions and conventions one introduces in the liquid-flow problem apply equally to the flow of a gas, so I will state them here with all due ceremony.

Assumption of a prismatic boundary, orientation of coordinates and notation for velocity components and fluid parameters. One assumes that the boundary of the region occupied by the fluid has straight-line generators. Let (X, Y, Z) be cartesian coordinates (each of which has the dimensions of length) with the positive Z-axis parallel to the generators of the boundary surface and in the direction of the flow of fluid. Let (U, V, W) be the components of the fluid velocity vector (each of which has the dimensions of velocity) belonging to the coordinates (X, Y, Z). Let P be the fluid pressure (a positive valued parameter with the dimensions of force-per-unit-area). Let ρ be the fluid density (a positive valued parameter with the dimensions of mass-per-unit-volume). Let μ denote the dynamic viscosity of the fluid (a positive valued parameter with the dimensions of stress divided by strain rate, or, equivalently, force times time divided by area).

Notation for typical lengths. Let L denote the diameter of the smallest circle into which the cross section of the channel can be fit. One may align the X-axis so that the two most widely separated points on a typical cross section are points on the X-axis and I will assume that the coordinate axes have this feature. Let δ denote a typical dimension of the cross section in the direction of the Y-axis. The assumptions I have introduced thus far ensure that $\delta \leq L$. Finally, let ℓ denote the *streamwise length* of the capillary.

Slenderness assumption. In capillary flow theory, one assumes that $L/\ell \equiv \epsilon \ll 1$. I will refer to the parameter ϵ as the *slenderness parameter* in what follows.

On the neglect of gravitational terms. One may interpret the differential equation for the rate of change of translational momentum of a fluid in the form of the statement "The acceleration of a fluid particle equals the resultant force-per-unit-mass (from all causes) to which that particle is subject." One of the force-per-unit mass terms in the streamwise component of the momentum equation is the expression $-(1/\rho)(dp/dZ)$, which represents direct driving by the unequal pressures at the inlet and the outlet. The presence of the minus sign reflects the fact that the fluid in a capillary flows in the direction of *decreasing* pressure. Another of the force-per-unit-mass term is $-g\sin\gamma$, in which $g > 0$ is the acceleration due to gravity and γ is the angle of inclination of the long axis of the capillary above the horizontal. Here γ is positive or negative according to whether the fluid flows uphill or downhill, respectively.

The neglect of the gravitational term, $-g\sin\gamma$, in comparison with the term due to direct driving by the pressure gradient, $-(1/\rho)(dp/dZ)$, thus amounts to the assumption that the ratio of the orders of magnitude of these terms is very much less than one, *viz.*

$$\frac{g}{(1/\rho)|dp/dZ|} \ll 1\,. \tag{1}$$

In the case of gas flow, we have the equation of state of an ideal gas, namely $p = \rho RT$, in which R is the gas constant particular to the gas in question (a parameter with the dimensions of energy per unit mass per unit temperature) and T is the absolute temperature. If one eliminates ρ from (1) by means of $p = \rho RT$, one obtains a restatement of (1) in the form

$$\frac{p/|dp/dZ|}{RT/g} \ll 1\,. \tag{2}$$

One may estimate the numerator of the fraction in the left member of (2) by the capillary length ℓ. The denominator RT/g has the dimensions of length and represents the change in elevation over which the

density of a stationary isothermal atmosphere decreases by a power of e. This so called *scale height of the atmosphere* is on the order of kilometers (about ten kilometers for air at 15° C). For a capillary whose length is on the order of millimeters, therefore, one may apply the assumption (2) with complete confidence.

For a liquid, the assumption (1) amounts to the statement that the pressure difference due to the hydrostatic head of the working fluid between the inlet and the outlet is small compared to the applied pressure difference. If the applied pressure is on the order of atmospheres and the capillary length is on the order of millimeters, then this one may again apply (2) with confidence.

Steadiness. I will restrict attention to the case when the velocity and pressure fields are independent of time.

On the neglect of inertial reaction to streamwise acceleration and the exclusion of turbulent flow. A standard exercise in classical fluid mechanics is to estimate the ratio so called *inertial* to *viscous* terms in the momentum equation. If there were only one representative length scale λ in the problem, then this ratio would be the REYNOLDS number $Re = \rho W\lambda/\mu$, where W is a typical streamwise velocity. In capillary flow, however, there are two relevant length scales, namely the capillary length, ℓ, and the smallest cross-stream length, δ, of the channel. A measure of the inertial force-per-unit-mass is thus W^2/ℓ while a measure of the viscous force per unit mass is $\mu W/(\rho\delta^2)$. One thus arrives at the following estimate of the ratio of the inertial to the viscous terms:

$$\left(\frac{\delta}{\ell}\right)\left(\frac{\rho W\delta}{\mu}\right) \ll 1 \,. \tag{3}$$

The first factor in the right member is the slenderness factor, ϵ, and the second factor is the REYNOLDS number, Re, based on the smallest cross-stream length of the channel. One may therefore restate the condition for the neglect of inertial terms in the momentum equation as follows:

$$\epsilon\, Re \ll 1 \,. \tag{4}$$

If ℓ is on the order of millimeters and δ is on the order of microns then ϵ will be on the order of 10^{-3}. In this case, Re need not be small and may even be in the hundreds without violating assumption (4). I will, however, confine attention to the case of non-tubulent flow, which generally requires that Re be less than 2,000.

Assumption of constant temperature. I will assume that the temperature of the fluid (whether gaseous or liquid) is of uniform temperature. Since the fluid viscosity, μ, is primarily a fuction of temperature, the constant temperature assumptions effectively guarantees that the viscosity is uniform throughout the fluid.

The Poisson equation for the streamwise velocity. Every viscous term in the equation of motion of a fluid involves a second derivative of a velocity component with respect to some combination of the spatial coordinates. In view of the slenderness assumption $\epsilon \ll 1$, one may separate the various viscous terms into classes distinguished by their respective orders of magnitude. The largest such terms are those in which both spatial derivatives are with respect to the cross-stream coordinates (X, Y). Keeping these terms and making a few additional uses of the assumption $\epsilon \ll 1$, one arrives at the following dominant balance of terms in the streamwise momentum equation:

$$\frac{\partial^2 \mathrm{W}}{\partial \mathrm{X}^2} + \frac{\partial^2 \mathrm{W}}{\partial \mathrm{Y}^2} \approx \frac{1}{\mu}\frac{dp}{d\mathrm{Z}} \,, \tag{5}$$

in which P is a function of Z only.

For a particular cross section Z = constant, equation (5) takes the form of a POISSON equation for the distibution of streamwise velocity W across that station. The boundary condition for (5) is the condition of no-slip between a viscous fluid and an adjacent solid wall:

$$\mathrm{W}\big|_{\mathrm{wall}} = 0 \,. \tag{6}$$

Introduction of nondimensional variables. One may introduce nondimensional cross-stream position coordinates (x, y) by the definitions

$$x \equiv \mathrm{X}/L \quad , \quad y \equiv \mathrm{Y}/L \; . \tag{7 a, b}$$

In the foregoing discussion, I have used the symbol W for a measure of the typical streamwise velocity across the cross section of the channel. One may make this definition more specific by taking

$$W \equiv -\frac{L^2}{\mu}\frac{dp}{d\mathrm{Z}} \; . \tag{8}$$

Note that fluid flows from greater to lesser pressure and the positive Z-axis in the direction of the flow. It follows that $dp/d\mathrm{Z} < 0$, so the right member of (8) is positive as expected. Having a scale for the streamwise velocity, one may define the nondimensional velocity component w by

$$w \equiv \mathrm{W}/W \; . \tag{10}$$

If one multiplies the POISSON equation (5) by L^2/W, one obtains, after some reductions, the following normalized form:

$$\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} = -1 \; . \tag{11}$$

whose solution is subject to a boundary condition equivalent to (6), *viz.*

$$w\big|_{\text{wall}} = 0 \; . \tag{12}$$

The solution of the boundary-value defined by (11) and (12) will be a function $(x, y) \mapsto w$. Note, in particular, that w *does not depend upon anything other than* x *and* y.

Conservation of mass along the channel and its consequences. Let $\mathcal{S}'$ denote a surface of the X-Y plane corresponding to the channel cross section. By the same token, let the unpunctuated symbol S denote the corresponding surface with respect to the nondimensional variables (x, y). One may write an expression for the rate of transport of mass, $\dot{m}$, across a typical cross section of the channel by

$$\dot{m} = \iint_{\mathcal{S}'} \rho \mathrm{W} \, d\mathrm{X} d\mathrm{Y} \; . \tag{13}$$

Under the assumptions stated thus far, the mass density, ρ, is, at worst, a function of the streamwise coordinate Z. One may therefore move the coefficient ρ outside of the integral sign in (13). If one carries out this step and also introduces the nondimensional coordinates (x, y) and the nondimensional streamwise velocity w by means of the definitions (7 a, b) and (10), one obtains

$$\dot{m} = \rho W L^2 \iint_{\mathcal{S}} w \, dx dy \; . \tag{14}$$

If one eliminates W by means of (8) and rearranges slightly, one obtains

$$\frac{\dot{m}\mu}{L^4 \iint_{\mathcal{S}} w \, dx dy} = -\rho \frac{dp}{d\mathrm{Z}} \; . \tag{15}$$

The assumption of steadiness and the law of conservation of mass imply that $\dot{m}$ is independent of Z. All of the other terms in the left member of (15) are also independent of Z under assumptions already stated. One concludes that for either liquids or gases

$$\rho\frac{dp}{dZ} = \text{independent of Z}\,. \tag{16}$$

In the particular case of liquid flow, ρ is constant, so constancy of $\rho\,dp/dZ$ with respect to Z implies constancy of dp/dZ. It follows that P varies linearly with respect to Z and we have

$$\rho\frac{dp}{dZ} = \rho\frac{p_2 - p_1}{z_2 - z_1} = \rho\frac{p_2 - p_1}{\ell} \tag{17}$$

for liquid flow. Here, the subscripts "1" and "2" denotes conditions at the inlet to and the outlet from the capillary, respectively.

In the particular case of gas flow, ρ varies according to the equation of state $P = \rho RT$, so

$$\rho\frac{dp}{dZ} = \frac{p}{RT}\frac{dp}{dZ} = \frac{1}{RT}\frac{d(p^2/2)}{dZ}\,. \tag{18}$$

Since the leftmost member of (18) is independent of Z (by (16)), one concludes from (18) and the assumption of constant temperature that p^2 varies linearly with respect to Z. It follows that

$$\rho\frac{dp}{dZ} = \frac{1}{2RT}\frac{p_2^2 - p_1^2}{z_2 - z_1} = \frac{1}{2RT}\frac{p_2^2 - p_1^2}{\ell} \tag{19}$$

for gas flow.

One may combine the results expressed by (15), (17), and (19) in the following form

$$\dot{m} = \begin{cases} \dfrac{L^4}{\mu}\displaystyle\iint_S w\,dx dy\,\rho\frac{p_1 - p_2}{\ell}, & \text{for liquid flow;} \\ \dfrac{L^4}{\mu}\displaystyle\iint_S w\,dx dy\,\dfrac{1}{2RT}\dfrac{p_1^2 - p_2^2}{\ell}, & \text{for gas flow:} \end{cases} \tag{20}$$

The two cases of equation (20) furnish an analytical correlation between the throughputs of liquid and gases through the same capillary.

Authors of works on leak detection sometimes employ the symbol S to denote the rate of transport of fluid volume across a cross section of a fluid conduit. These same authors also sometimes write Q for the product pS and refer to the resulting quantity as the *gas throughput*. In view of the equation of state of an ideal gas, one may relate Q to $\dot{m}$ by the following identities

$$\dot{m} = \rho S = \frac{p}{RT}S = \frac{Q}{RT}\,. \tag{21}$$

If one multiplies (20) (in the case of gas flow) by RT, one obtains, in view of the outermost equality in (21), the identity

$$Q = \frac{L^4}{\mu}\iint_S w\,dx dy\,\frac{p_1^2 - p_2^2}{2\ell}\,. \tag{22}$$

The quantity Q so defined has the dimensions of pressure times volume transport rate. Thus, the expression atm·cm^3/s (also known as *standard cubic centimeters per second*) are suitable units for the measurement of Q. Of course, the outermost identity in (21) allows one to relate Q to the mass transport rate $\dot{m}$.

The case of a circular cross section; Classical formulas. The practical implementation of the foregoing formulas (which apply to all capillaries regardless of the cross sectional shape) depends upon the evaluation of the integral

$$\iint_S w\, dxdy\,, \tag{23}$$

which, in turn, depends upon the solution of the normalized boundary value problem defined by (11) and (12). The simplest example of a solution of this system is that of the capillary whose cross section is a circle of diameter L. In nondimensional coordinates, the diameter becomes one and, thus the radius becomes 1/2. The equation of the boundary shape (*i.e.* a circle of radius 1/2 centered on the origin) thus becomes $(1/2)^2 - x^2 - y^2 = 0$. If one notes that the the operator

$$\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \tag{24}$$

takes the expression $(1/2)^2 - x^2 - y^2$ to a constant (in this case, -4) in the interior of the circle and that the expression $(1/2)^2 - x^2 - y^2$ vanishes on the boundary, it becomes plain that multiplication of $(1/2)^2 - x^2 - y^2$ by a suitable constant (in this case, 1/4) yields a solution of the boundary value problem (11), (12) for the circular geometry. One concludes that

$$w = \frac{(1/2)^2 - x^2 - y^2}{4}\,. \tag{25}$$

in circular geometry.

If one introduces (r, θ) for polar coordinates in the (x, y) plane, then (25) becomes

$$w = \frac{(1/2)^2 - r^2}{4} \tag{26}$$

and the integral (23) is equivalent to

$$\iint_S w\, dxdy = \int_0^{1/2} w(r)\, 2\pi r\, dr = \frac{\pi}{8}\frac{1}{2^4}\,. \tag{27}$$

If one substitutes this result into (20) and notes that $p_1^2 - p_2^2 = (p_1 + p_2)(p_1 - p_2)$, one obtains

$$\dot{m} = \begin{cases} \dfrac{\pi}{8\mu}\left(\dfrac{L}{2}\right)^4 \rho \dfrac{p_1 - p_2}{\ell}, & \text{for liquid flow;} \\ \dfrac{\pi}{8\mu}\left(\dfrac{L}{2}\right)^4 \left(\dfrac{(p_1+p_2)/2}{RT}\right) \dfrac{p_1 - p_2}{\ell}, & \text{for gas flow.} \end{cases} \tag{28}$$

The liquid-flow form of (28) is associated with the names of HAGEN and POISEUILLE, both of whom studied the flow experimentally in the nineteenth century. The result was derived analytically by GEORGE GABRIEL STOKES. The earliest analytical derivation of the gas-flow form of (28) appears to be that of OSKAR EMIL MEYER in 1866 (Ref. 1, page 269).

In view of the ideal gas law, the constant valued expression

$$\frac{(p_1+p_2)/2}{RT}$$

has the dimensions of density, and one may regard it as an *effective constant density* for the gas in a capillary (based on the ideal gas law and the arithemetic mean of the inlet and outlet pressures). While this interpretation is convenient as a mnemonic device, it is important to realize that the derivation of the MEYER formula does not assume constancy of the density even as an approximation. Indeed, the MEYER formula holds for *any* expansion ratio. The foregoing derivation supports the last assertion nothwithstanding contrary statements by modern authors in the literature on leak detection.

3. FLOW THROUGH A WIRE CREVICE

Transformation to a Dirichlet problem. Let $(x, y) \mapsto w_p$ be a function defined by the equation

$$w_p = -y^2/2. \tag{29}$$

It follows that

$$\frac{\partial^2 w_p}{\partial x^2} + \frac{\partial^2 w_p}{\partial y^2} = -1\,, \tag{30}$$

One concludes that w_p is a particular solution of the differential equation (11) satisfied by w. Let $(x, y) \mapsto \phi$ be a function defined by the equation

$$w = w_p + \phi\,. \tag{31}$$

If one subtracts (30) from (11), and uses the identity $w - w_p = \phi$ (which follows from (31)), one obtains

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} = 0\,. \tag{32}$$

If, now one substitites (31) into the boundary condition (12), one obtains

$$\phi\big|_{\text{wall}} = -w_p\big|_{\text{wall}} = -(-y^2/2)\big|_{\text{wall}}\,,$$

or

$$\phi\big|_{\text{wall}} = \frac{(y_{\text{w}})^2}{2}\,, \tag{33}$$

in which the subscript $(\,)_{\text{w}}$ denotes evaluation on the wall.

The partial differential equation (32) that ϕ satisfies is LAPLACE's equation. A function that satisfies LAPLACE's equation is called a *harmonic* function. The boundary-value problem consisting of LAPLACE's equation and a statement of the boundary values of the unknown (as distinguished from the normal derivative of the unknown or some combination of the unknown and its normal derivative) is called the DIRICHLET *problem.* The pair of equations (32) and (33) thus constitutes a DIRICHLET problem for ϕ.

Construction of harmonic functions from differentiable functions of a complex variable. Experience has shown that one may expedite the solution of the DIRICHLET problem in two space dimensions considerably through the use of complex variables. There are two reasons why this is so. To illustrate the first reason, let $\zeta = \xi + i\eta$ be a complex number. Here $i = \sqrt{-1}$ and the pair of real numbers (ξ, η) represent a point in a plane (the complex ζ-plane). If $\zeta \mapsto F$ is an *analytic* (*i.e.* differentiable) function that maps ζ to the complex number F (with the expansion, say, $F = \phi + i\psi$), then the real and imaginary parts of F

(namely ϕ and ψ, respectively) are related to the real and imaginary parts of ζ (namely ξ and η, respectively) by the CAUCHY-RIEMANN equations

$$\frac{\partial\phi}{\partial\xi} = \frac{\partial\psi}{\partial\eta} \quad , \quad \frac{\partial\phi}{\partial\eta} = -\frac{\partial\psi}{\partial\xi} \, . \tag{34 a,b}$$

If one eliminates ψ from the (34 a, b) by cross differentiation, one obtains, immediately,

$$\frac{\partial^2\phi}{\partial\xi^2} + \frac{\partial^2\phi}{\partial\eta^2} = 0 \, . \tag{35}$$

In this way, any analytic function of a complex variable (*e.g.* $\zeta \mapsto F$) is related in a systematic way to a harmonic function (*e.g.* $(\xi, \eta) \mapsto \phi$).

A class of changes of independent variable that leaves Laplace's equation invariant. A second reason why complex varibles are useful in solving the DIRICHLET problem in two dimension concerns the invariance of LAPLACE's equation under a class of transformations of the independent variables. Suppose, for example, that $z \mapsto m$ is an analytic function that takes the complex number z to the complex number m and that the expansions of z and m into real and imaginary parts are

$$z = x + iy \quad , \quad m = m_r + im_i \, . \tag{36}$$

Then the function $z \mapsto m$ induces a transformation $(x, y) \mapsto (m_r, m_i)$ (called a *conformal transformation*) between points in the complex z-plane and points in the complex m-plane. A basic exercise in the theory of functions of a complex variable is the demonstration that

$$\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} = \left|\frac{dm}{dz}\right|^2 \left(\frac{\partial^2}{\partial m_r^2} + \frac{\partial^2}{\partial m_i^2}\right) \tag{37}$$

(*cf.* JEFFREYS & JEFFREYS, p534) (Ref. 2). If, for example, $(x, y) \mapsto \phi$ is harmonic, then

$$0 = \frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} = \left|\frac{dm}{dz}\right|^2 \left(\frac{\partial^2\phi}{\partial m_r^2} + \frac{\partial^2\phi}{\partial m_i^2}\right) \, . \tag{38}$$

Division of the outermost equality by $|dm/dz|^2$ yields

$$\frac{\partial^2\phi}{\partial m_r^2} + \frac{\partial^2\phi}{\partial m_i^2} = 0 \, . \tag{39}$$

Thus, if ϕ is a harmonic function of one set of independent variables, and those variables are related to a new set of variables by a conformal transformation then ϕ will also be harmonic with respect to the new variables.

Transformation of the wire crevice to the interior of the unit circle. The book *Analytic function theory* by EINAR HILLE (Ref. 3) gives several examples of conformal transformations, of which two are particularly germane to the present problem. Figure 3 illustrates regions in the planes of four distinct complex variables. The transformation between what I have called the m-plane and the τ-plane appears on

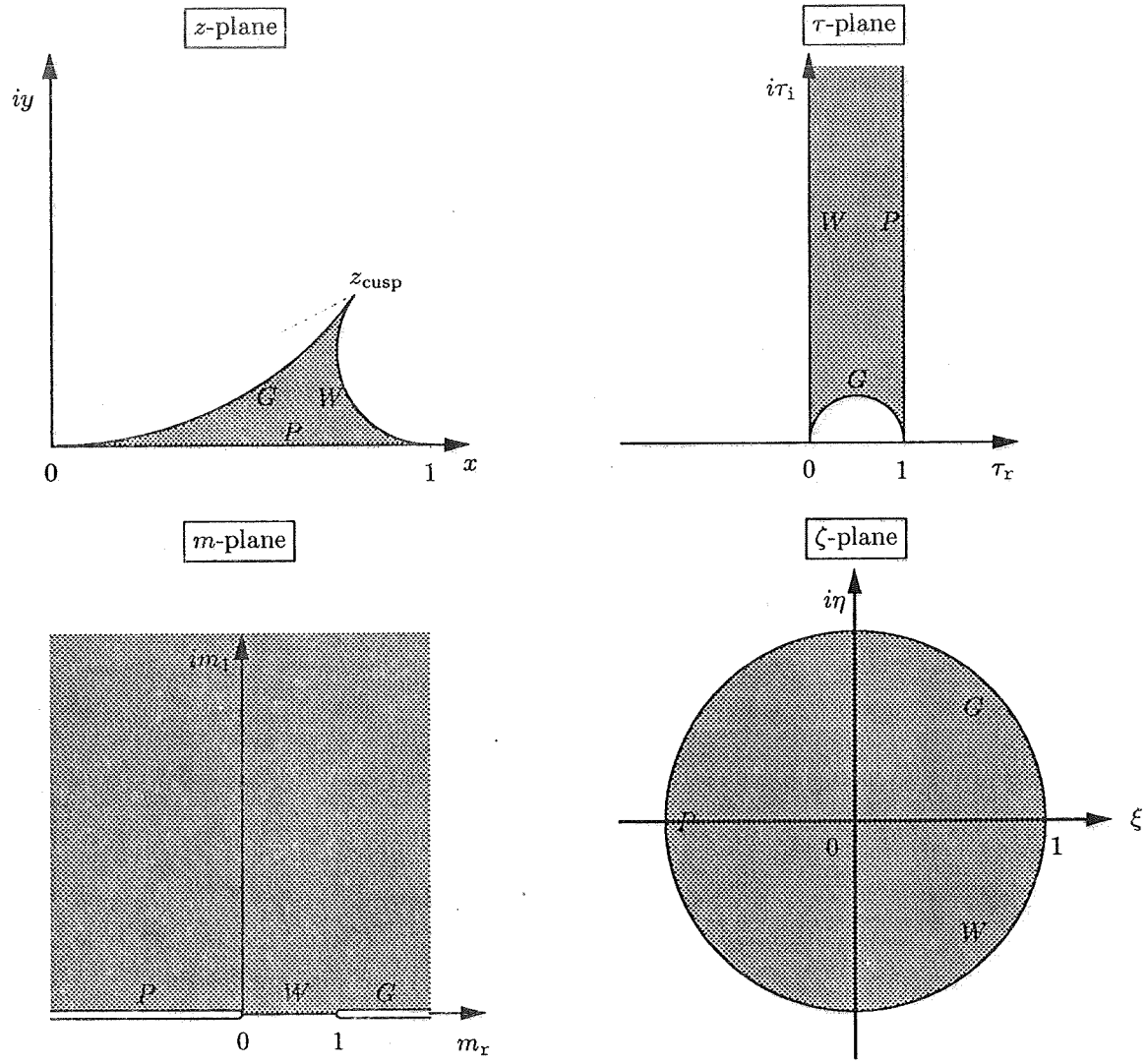


Figure 3. Three step conformal transformation of a wire crevice (in the z-plane) to the interior of a unit circle (in the ζ-plane).

(and near) page 177 of HILLE's book and the transformation between what I have called the τ-plane and the ζ plane appear on (and near) page 370 of that source.

HILLE's formula for the transformation $m \mapsto \tau$ is equivalent to

$$\tau = \frac{iK(1-m)}{K(m)} , \tag{40}$$

in which $m \mapsto K(m)$ is a so-called *complete elliptic integral of the first kind* (*ibid.*, p150). The inverse of (40) is

$$m = \left(\frac{\vartheta_2(0|\tau)}{\vartheta_3(0|\tau)} \right) , \tag{41}$$

in which $(z,\tau) \mapsto \vartheta_i(z|\tau)$ $(i \in \{1,2,3,4\})$ are the so-called *theta functions*, (*ibid.*, p156). Some modern software packages, such as *Mathematica*, enable evaluation of $m \mapsto K(m)$ and $(z,\tau) \mapsto \vartheta_i(z|\tau)$ $(i \in \{1,2,3,4\})$ for complex arguments as easily as the evaluation of sines and cosines. The use of such higher functions is therefore not a practical limitation on a the development of a computational algorithm for the present problem.

HILLE's formula for the transformation $\zeta \mapsto m$ (or a trivial modification therof) is equivalent to

$$m = i\left(\frac{1-i\zeta}{1+i\zeta}\right) , \tag{42}$$

whose inverse is

$$\zeta = i\left(\frac{m-i}{m+i}\right) . \tag{43}$$

HILLE (*ibid.*, p 46, *et seq.*) discusses a class of conformal transformations which he associates with the name MÖBIUS. In these transformations, two complex variables are related by a formula which equates the first variable to a ratio of linear functions of the second (The inverse of such a transformation equates the second variable to a ratio of linear functions of the first). HILLE calls attention (*ibid.*, page 51) to the class of geometric figures consisting of circles and straight lines. A basic feature of the MÖBIUS transformation is that it takes a circle or line in the plane of the first complex variable to a circle or line in the plane of the second. In this sense, the class of circles and lines is invariant under MÖBIUS transformations. With this theorem in mind, it is not difficult to find (by trial and error) a MÖBIUS transformation that takes the wire crevice in the z-plane to the interior of the shaded figure in the τ-plane. Such a MÖBIUS transformation is given by

$$\tau = \frac{1 - z/z_{\text{cusp}}}{1-z} , \tag{44}$$

whose inverse is

$$z = \frac{\tau - 1}{\tau - (1/z_{\text{cusp}})} . \tag{45}$$

Here z_{cusp} is the intersection of the gasket and the wire in the wire crevice problem.

The condition that the gasket and the wire be tangent to each other restricts values of z_{cusp} to points on a semicircle of diameter one centered at the point 1/2 on the real axis. One may write a formula which encompasses all possible values of z_{cusp} that fulfill this condition in the form

$$z_{\text{cusp}} = \frac{1}{1-iA} , \tag{46}$$

in which $A > 0$ is a real parameter. In fact $A^2 = R_w/R_g$, where R_w and R_g are the radii of curvature of the wire arc and the gasket arc, respectively.

Boundary value of ϕ in the ζ-plane. The foregoing conformal transformation formulas enable one to associate any point (x,y) in the in or on the wire crevice in the z-plane (where $z = x + iy$) to one and only one point (ξ,η) in or on the unit disk in the ζ-plane (where $\zeta = \xi + i\eta$). Thus, the boundary values (33) for the unknown harmonic function $(x,y) \mapsto \phi$ in (32) carry over to corresponding boundary values of ϕ in the ζ-plane. Figure 4 illustrates the corresponding distribution of these boundary values expressed in terms of the polar coordinate angle θ in the ζ-plane (The reader should note that my use of (r,θ) as polar coordinates in the ζ-plane here is a new use of these symbols, not to be confused with my earlier use of them in the text between (25) and (26) above).

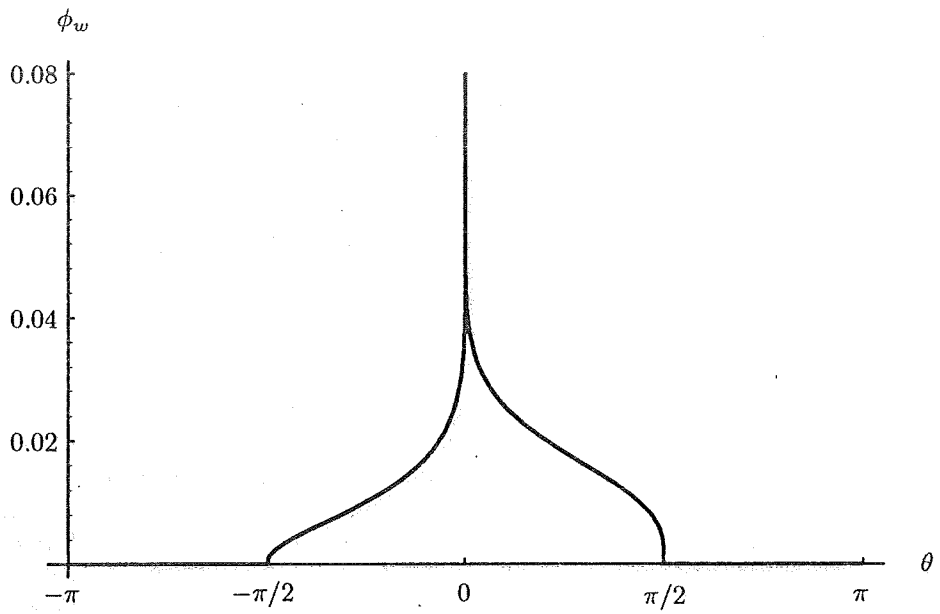


Figure 4. Distribution of the boundary value of the harmonic function ϕ as a function of polar angle in the ζ-plane

Solution for ϕ in the ζ-plane as an infinite series. One may relate the LAPLACE operator in cartesian coordinates (ξ, η) in the ζ-plane to polar coordinates (r, θ) in that plane by the identity

$$\frac{\partial^2(\,)}{\partial\xi^2} + \frac{\partial^2(\,)}{\partial\eta^2} = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial(\,)}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2(\,)}{\partial\theta^2} \,. \tag{47}$$

One may verify that for any integer n, the operator in the right member of (47) takes expression $r^n\cos(n\theta)$ to zero. The same operator takes $r^n\sin(n\theta)$ to zero and one arrives at the conclusion that any function of the form

$$\phi = \frac{a_0}{2} + \sum_{n=1}^{\infty} r^n[a_n\cos(n\theta) + b_n\sin(n\theta)] \,, \tag{48}$$

whose constant, real-valued coefficients (a_n, b_n) are subject to the condition that the series (48) converge, is a solution of LAPLACE's equation in the ζ-plane. In view of the foregoing discussion under the heading "A class of changes of independent variable that leaves LAPLACE's equation invariant" such a representation for ϕ must also satisfy LAPLACE's equation in the z-plane.

In view of the complex variable identities $\cos\theta = (e^{i\theta} + e^{-i\theta})/2$ and $\sin\theta = (e^{i\theta} - e^{-i\theta})/(2i)$, one may arrange (48) in the following equivalent form

$$\phi = \frac{a_0}{2} + \sum_{n=1}^{\infty} r^n \left[\left(\frac{a_n - ib_n}{2}\right)e^{in\theta} + \left(\frac{a_n + ib_n}{2}\right)e^{-in\theta}\right] \,. \tag{49}$$

One may shorten the notation by defining values of a_n and b_n for zero and negative subscript as follows (*e.g.* LIGHTHILL (1959, p3), Ref 4):

$$a_{-n} \equiv a_n \quad , \quad b_{-n} \equiv -b_n \tag{50}$$

(so $b_0 = 0$, by construction). Moreover, for any integer n, we write

$$\frac{a_n - ib_n}{2} \equiv C_n \,. \tag{51}$$

With these conventions, equation (49) takes the form

$$\phi = \sum_{n=-\infty}^{\infty} C_n r^{|n|} e^{in\theta} \,. \tag{52}$$

If one substitutes $r = 1$, one obtains the boundary value of ϕ, namley ϕ_w, so

$$\phi_w = \sum_{n=-\infty}^{\infty} C_n e^{in\theta} \,. \tag{53}$$

I have already remarked that the boundary-value function $\theta \mapsto \phi_w$ is known (*cf.* Figure 4 above). Let k be an arbitrary integer. If one multiplies (53) by $e^{-ik\theta}$ and integrates with respect to θ from zero to 2π, one obtains

$$\int_0^{2\pi} \phi_w e^{-ik\theta} d\theta = \sum_{n=-\infty}^{\infty} C_n \int_0^{2\pi} e^{i(n-k)\theta} d\theta \,. \tag{54}$$

But

$$e^{i(n-k)\theta} = \begin{cases} \frac{d}{d\theta}(\theta) & \text{if } k = n; \\ \frac{d}{d\theta}\left(\frac{e^{i(n-k)\theta}}{i(k-n)}\right) & \text{if } k \neq n. \end{cases} \tag{55}$$

It follows that

$$\int_0^{2\pi} e^{i(n-k)\theta} d\theta = \begin{cases} 2\pi & \text{if } k = n; \\ 0 & \text{if } k \neq n. \end{cases} \tag{56}$$

If one substitutes this last result into (54), one obtains

$$\int_0^{2\pi} \phi_w e^{-ik\theta} d\theta = 2\pi C_k \,, \tag{57}$$

or, equivalently,

$$C_n = \frac{1}{2\pi} \int_0^{2\pi} \phi_w e^{-in\theta} d\theta \,, \tag{58}$$

(which holds for all integer n. Equation (58) thus determines the coefficients in the series representation (52).

Solution for w in the z-plane. Knowing ϕ at any point in the ζ-plane, one may associate the same value of ϕ with the corresponding point in the z-plane (*via* the conformal transformations given in equations (40)–(45)). One concludes that ϕ is now known through the whole wire crevice. Since w_p is specified by (29), one concludes that the whole right member of (31) for w is known throughout the wire crevice.

4. RESULTS

Figure 5 gives a plot of the contours of constant w obtained from a numerical evaluation. The slight raggedness of the outermost contour is probably due to an overly course discretization of θ in the calculation.

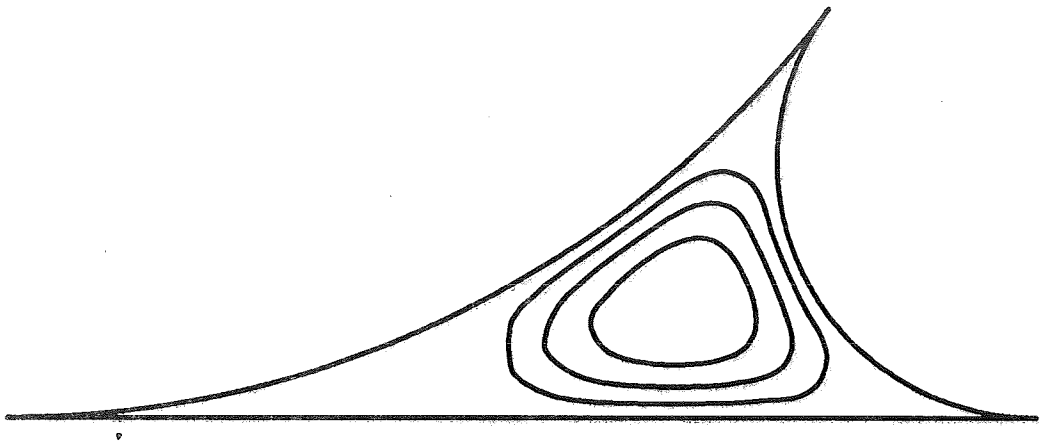


Figure 5. Contours of constant streamwise velocity in the flow through a wire crevice. The innermost contour represents 3/4 of the maximum value of w, the next innermost contour represent 1/2 the maximum, and so on.

5. CONCLUSIONS AND RECOMMENDATION

The foregoing work supports the following conclusions:

1. The liquid-gas flowrate correlation embodied by (20) (which holds for arbitrary shaped cross sections) and (28) (which holds for circular cross sections) are on a firm scientific basis and are suitable for immediate application.
2. One may solve the boundary-value problem for the flow through a wire crevice analytically and the solution is expressible in terms of a single infinite series.
3. The solution for wire-crevice geometry is computationally feasable on a small computer.

I had hoped to refine the calculations so as to eliminate the raggedness of the w-contours before completion of the summer term, but ran into some snags during the last two weeks. The recommendation, of course, is to continue the effort including the calculation of the integral (23), and thence, the throughput.

ACKNOWLEDGEMENTS

I am indebted to my NASA colleagues RICK ADAMS and CAROLYN BARRET for continued moral support and lobbying on my behalf during this effort (and previous ones). I am also indebted to the I-NET personnel in the Hazardous Gas Detection Lab, especially LARRY LINGVAY, GUY NAYLOR, RICH HRITZ, CURT LAMPKIN and DAVID FLOYD for providing continued encouragement, material support, space, and practial knowhow. Finally, I am indebted to the people who ran the NASA/ASEE Summer Faculty Program this year, namely, KARI STILES and Dr. ROGER JOHNSON of the University of Central Florida and GREGG BUCKINGHAM of NASA.

REFERENCES

1. MEYER, OSKAR EMIL Über die Reibung der Gase/ über die Strömung der Gase durch Capillarröhren. *Annalen der Physik*, **117**, pp 253–281, 1866.
2. JEFFREYS, HAROLD & JEFFREYS, BERTHA *Methods of mathematical physics.* Third edition, Cambridge University Press, 1956.
3. HILLE, EINAR *Analytic function theory.* Second edition, Chelsea Publishing Company, Volume II, 1962.
4. LIGHTHILL, M.J. *Introduction to Fourier analysis and generalized functions.* Cambridge University Press, 1958.

1996 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

VISUALIZATION OF REAL-TIME DATA

Dr. Ryan Stansifer, Associate Professor
Computer Science Department
Florida Institute of Technology
Melbourne, Florida

KSC Colleague - Peter Engrand
Artificial Intelligence/Expert Systems

Contract Number NASA-NGT10-52605

August 9, 1996

ABSTRACT

In this project we explored various approaches to presenting real-time data from the numerous systems monitored on the shuttle to computer users. We examined the approach that several projects at the Kennedy Space Center (KSC) used to accomplish this.

We undertook to build a prototype system to demonstrate that the Internet and the Java programming language could be used to present the real-time data conveniently. Several Java programs were developed that presented real-time data in different forms including one form that emulated the display screens of the PC GOAL system, this system is familiar to many at KSC. Also, we developed several communicating programs to supply the data continuously. Furthermore, a framework was created using the World Wide Web (WWW) to organize the collection and presentation of the real-time data.

We believe our demonstration project shows the great flexibility of the approach. We had no particular use for the data in mind, instead we wanted the most general and the least complex framework possible. People who wish to view data need only know how to use a WWW browser and the address (the URL). People wanting to build WWW documents containing real-time data need only know the values of a few parameters—they do not need to program in Java or any other language. These are stunning advantages over more monolithic systems.

1 Introduction

1.1 Real-Time Data

During the operation of the shuttle, sensors are monitoring many of the shuttle systems. This data goes to special hardware at the Launch Control Center (LCC) known as the Common Data Buffer (CDBF). Lots of data is collected, approximately 30,000 measurements. These measurements are continually changing—some of them can change rapidly at certain times. The data is used in monitoring the operation of the shuttle and in analyzing substems for saftey, performance, technolgical improvements, etc.

The goal of this project to design a new information system structure to meet all these needs.

2 Related projects

Several systems use the real-time data from the CDBF. These systems vary vastly in purpose and sophistication.

2.1 PC GOAL

The goals of the PC GOAL system presents shuttle data on schematic like screens described by character-oriented DSP files. Figure 1 shows one of the PC GOAL display screens. The system PC GOAL system requires Intel computer and special networking hardware and software to work. The data collection engine is fast and reliable.

2.2 PAT

Rockwell's propulsion advisor tool (PAT) is designed to do intelligence analysis. It is not this aspect that is relevant here at the moment, but rather its the overall architecture. PAT is an X Window System application. This provides excellent and efficient graphics.

The X Window System requires that the software be ported and distributed to all the platforms. Also the computer users would have to be running the X Window System. Actually more different platforms run the X Window System than support Java, but currently more users are familiar with WWW browsers than with the X Window System.

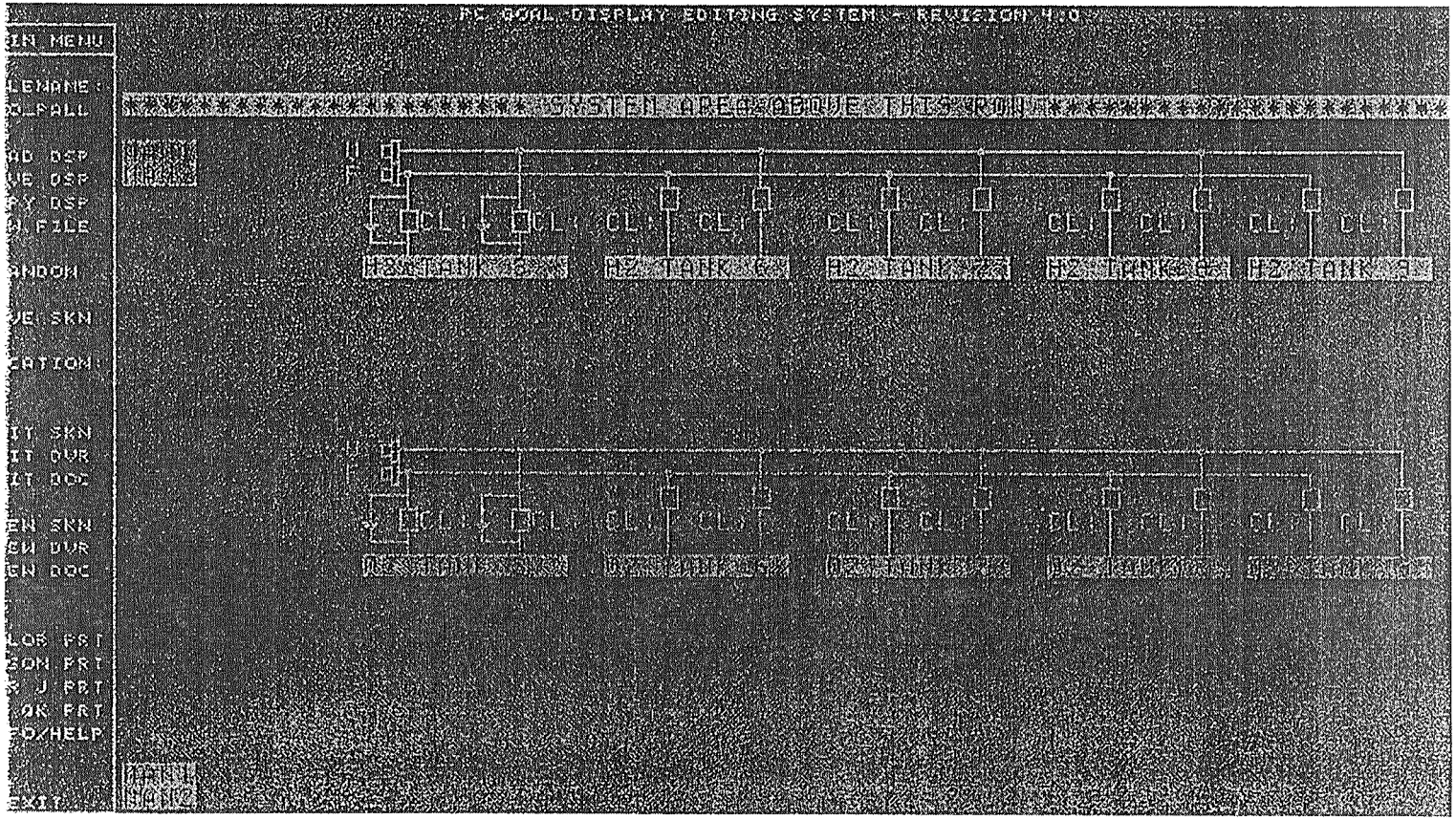


Figure 1: One of PC GOAL display screen

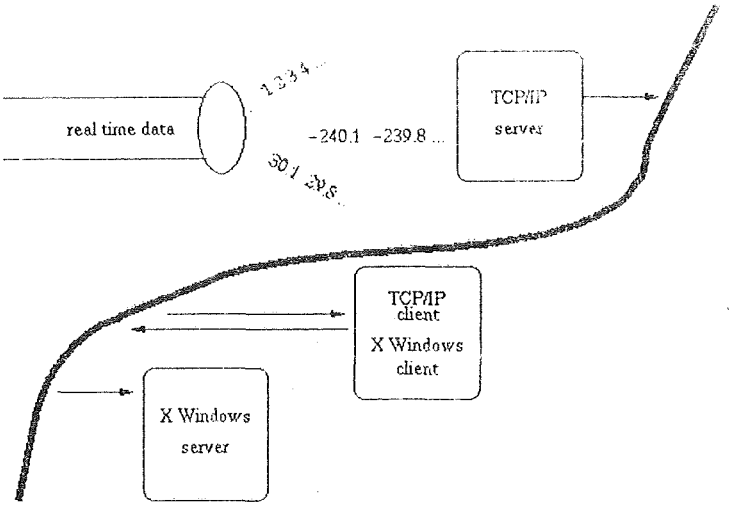


Figure 2: Real-time data reaching an X Window System application

2.3 Exodus

The Exodus project displays its data on WWW documents using CGI programs. Figure 3 shows one the documents produced by the CGI program with some real-time data. (For some reason, unreasonable numbers and strange characters appear in the place of the good data, but that is not the point here.) Periodically another document is produced with the data on it and sent to the WWW browser. This requires a lot of redundant network traffic when on a few numbers have changed. Also graphics are hard to produce on an HTML document without resorting to graphic images that would have to be transmitted each time to the browser.

3 Java applets

The key difference with using Java is that the applet makes a connection to a data server and only the data is transmited across the network after the initial class is transmited. Figures 4 and 5 depict the situation.

The advantages are in using Java are even greater than narrow technical merits suggest. Administratively, the operation of real-time data service using Java is much better. The Java applet is written once and executed remotely; no porting has to be done. Also, the

Figure 3: Four FD's monitored by Exodus DMON

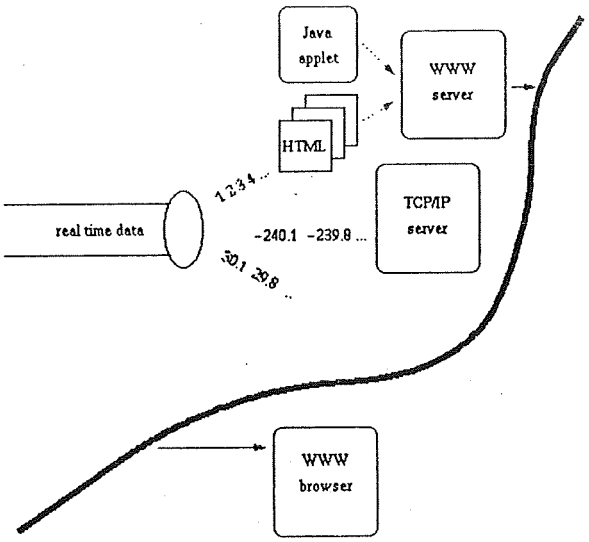


Figure 4: A WWW server distributes a Java applet

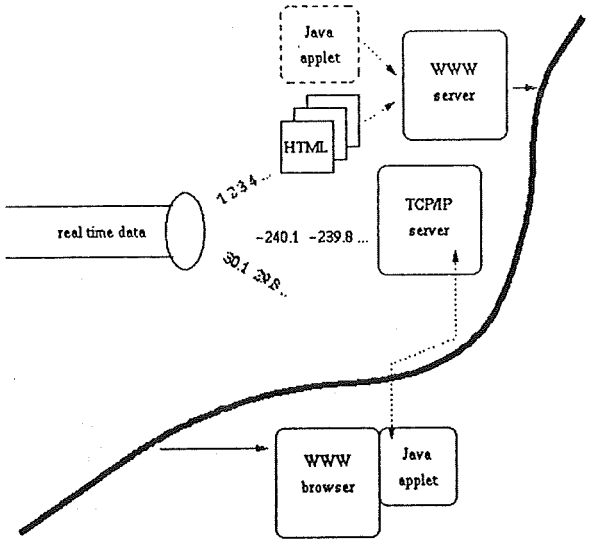


Figure 5: Then, the applet connects with the TCP/IP data server

latest version of the applet is always distributed to the user; there is no version control problem. Finally, the operation of the service is easy as browsers are ubiquitous; no training is required.

4 Using the Display Applet

Adding real-time data to an HTML document is easy in the framework we have developed. Here is an an example of using the display applet. The indicated FD's are displayed on the PC GOAL display background `LOXPRIM`. The numbers are used into place the data in a rectangle region of the image. The format governs how the data is displayed.

```
<applet code="DisplayApplet06" width=576 height=465>
   <param name="server"      value="ecto1">
   <param name="orbiter"     value="104">
   <param name="background" value="LOXPRIM.DSP">

   <param name="0FD"      value="GL0P0006A">
   <param name="0class"   value="TextDisplayClass">
   <param name="0format" value="%5.1f">
   <param name="0pos"     value="80,28,40,14">

   <param name="1FD"      value="GL0Q0229A">
   <param name="1class"   value="TextDisplayClass">
   <param name="1format" value="%4.0f">
   <param name="1pos"     value="184,28,32,14">
</applet>
```

5 Conclusions and Future Work

We have completed a prototype implementation in the Java programming language in which real-time data can be display on top of images, in PC GOAL DSP files, or in individual components. Figure 6 shows a Java applet emulating the PC GOAL screen `LOXPRIM`.

The graphical user interface is primitive, but until demand requires a particular function, making the interface more elaborate would be less productive. The work now needs to turn to the client/server architecture and design of appropriate network protocols and communication programs.

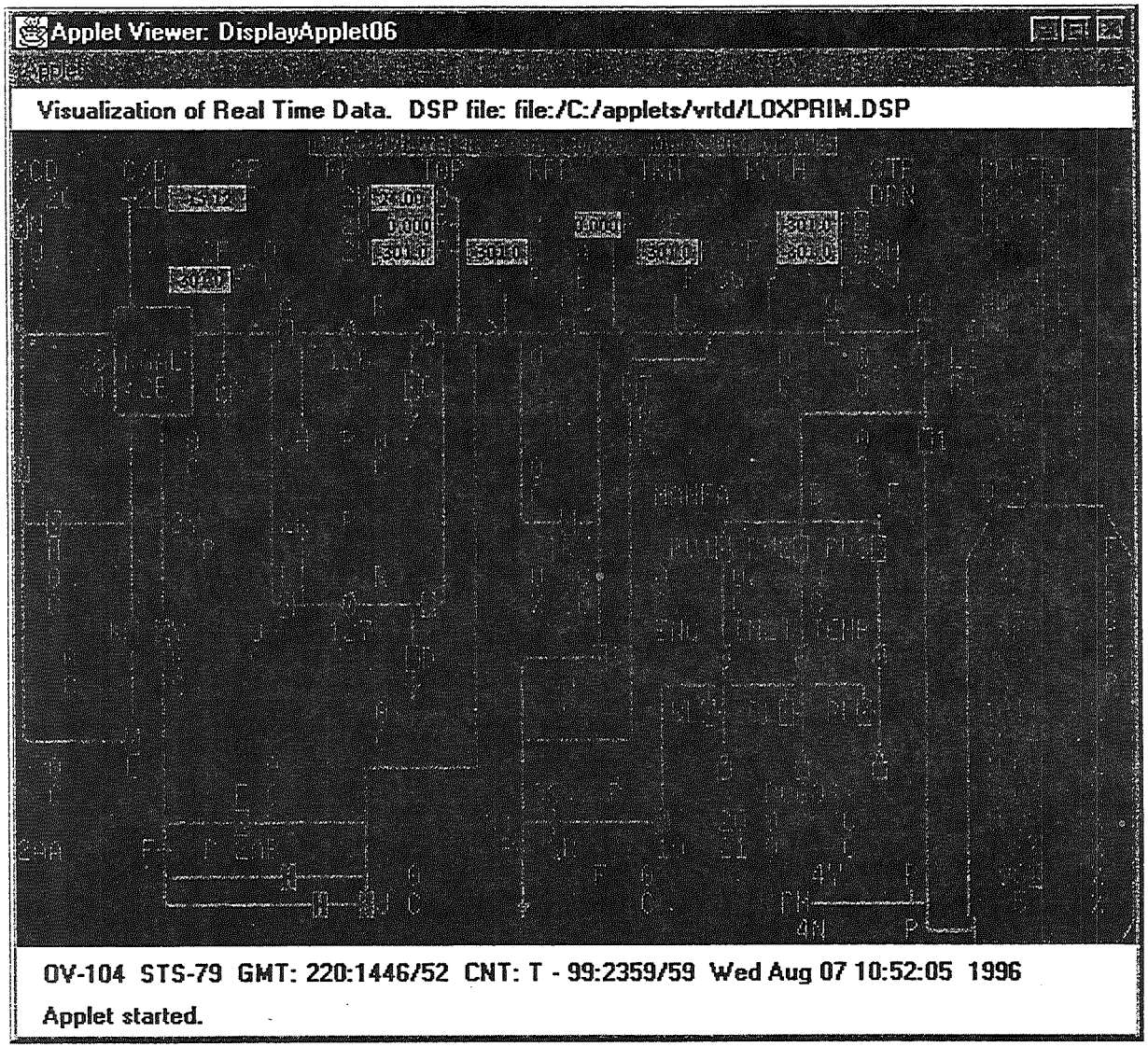


Figure 6: Real-Time Data in an Emulated PC GOAL screen

6 Acknowledgments

The impetus of this work is due to Peter Engrand (NASA) and Charlie Goodrich (I-Net). The rest of the Advanced Software group, in particular, Steve Beltz and John Dockendorf, contributed greatly in hosting me and providing facilities, The group at Rockwell led by James Engle, in particular Randy Lane and Van Bullington, made the project possible by providing a server for the real-time data. A special thanks is due to Roger Lee (USA) for answering questions about the "scan server." I gratefully acknowledge the contribution of Roger Johnson (UCF) and Greg Buckingham (NASA) who ran the NASA/ASEE Summer Faculty Fellowship Program with efficiency and enthusiasm.

Finally, I wish to thank the summer interns Paul Bryer (Florida Tech) and Kevin Gillette (Princeton). Paul wrote the Java component classes and Kevin wrote the Java code for interpreting DSP files and the code communicating with the "scan server."

A Links

The scientific literature is no particular use in understanding the context or methods of this work, so in lieu of references I have listed a number of links to material on the WWW that provides additional information pertaining to the project.

Books – a list of books on Java

`http://www.cs.fit.edu/~ryan/ogi/java-books.html`

Demonstration using DSP files – a browser that understands Java can view real-time data by staring at this URL and picking one of the PC GOAL display screens.

`http://ecto1.ksc.nasa.gov:8600/ryan/dsp/`

Exodus – The Exodus project

LPS `http://lpsweb.ksc.nasa.gov/`

PC GOAL `http://lpswebksc.nasa.gov/SDC/PCGOAL/homepage.html`

`Personel of the Advanced Systems and Analysis Division`

`http://www-de.ksc.nasa.gov/de/dm/dm-asd/personnel.html`

Ryan Stansifer – author of this report

`http://www.cs.fit.edu/~ryan/`

1996 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

CROSS: A GDSS FOR THE EVALUATION AND PRIORITIZATION OF ENGINEERING SUPPORT REQUESTS AND ADVANCED TECHNOLOGY PROJECTS AT NASA

Dr. Madjid Tavana, Associate Professor and Chairman
Management Department
La Salle University
Philadelphia, Pennsylvania

KSC Colleague - Seunghee Lee
Industrial/Business Management

Contract Number NASA-NGT10-52605

July 1996

ABSTRACT[1]

Objective evaluation and prioritization of Engineering Support Requests (ESRs) is a difficult task at the Kennedy Space Center (KSC) -- Shuttle Project Engineering Office. The difficulty arises from the complexities inherent in the evaluation process and the lack of structured information. The purpose of this project is to implement Consensus Ranking Organizational Support System (CROSS), a multiple criteria decision support system (DSS) developed at KSC, that captures the decision makers' beliefs through a series of sequential, rational, and analytical processes. CROSS utilizes Analytic Hierarchy Process (AHP), subjective probabilities, entropy concept, and Maximize Agreement Heuristic (MAH) to enhance the decision makers' intuition in evaluating ESRs. Some of the preliminary goals of the project are to:

- Revisit the Structure of the Ground Systems Working Team (GSWT) steering committee.
- Develop a template for ESR originators to provide more complete and consistent information about ESRs to the GSWT steering committee and stakeholders.
- Develop an objective and structured process for the initial screening of ESRs.
- Extensive training of the stakeholders and the GSWT steering committee members to eliminate the need for a facilitator.
- Automate the process as much as possible.
- Create an environment to compile Project Success Factor data on ESRs and move towards a disciplined system that could be used to address supportability threshold issues at the KSC.
- Investigate the possibility of an organization-wide implementation of CROSS.

[1] I would like to express my gratitude to NASA/ASEE for providing me with this wonderful research opportunity. I am greatly indebted to my NASA colleague, *Seunghee Lee* for her patience and expert guidance throughout this project. I am also grateful to *Jeff Wheeler* and *Tom Mullin* for their support and encouragement. In addition, I would like to thank the GSWT Steering Committee. And last but not least, I wish to express my appreciation to *Dr. Roger Johnson* of UCF and *Greg Buckingham* of NASA for their expert leadership and *Kari Stiles* of UCF, for her professionalism and enthusiasm. Each of them made participation in the program a pleasurable and rewarding experience.

CROSS: A GDSS for the Evaluation and Prioritization of Engineering Support Requests and Advanced Technology Projects at NASA

Madjid Tavana

Introduction

Decreased availablity of funding and an increasing number of Engineering Support Requests (ESRs) has created more competition among the stakeholders at NASA - Kennedy Space Center. There is clearly a need to replace the current unstructured ESR evaluation and selection process. The current process lacks the accountability, ignores the participation, and limits the objectivity that can be achieved through Consensus Ranking Organizational Support System (CROSS). The more comprehensive and structured framework provided by CROSS promotes the participation and harmony among Management, The Ground System Working Team Steering Committee (GSWT-SC), ESR Originators, and Stakeholders.

A total of 30 ESRs as shown in Table 1 are being considered by the GSWT-SC for 1997 fiscal year.

Insert Table 1 Here

CROSS is a three-phase, eleven-step procedure which systematically assesses ESRs and provides a final ranking of these ESRs by calculating their Project Success Factor (PSF). The three phases of assessment as represented in Figure 1, include the *Interaction Phase*, *Integration Phase*, and *Interpretation Phase*. These phases along with their respective steps are described below:

Insert Figure 1 Here

I. Interaction Phase:

During this phase, Decision Makers (DMs) and stakeholders interact through a series of automated systems for the purpose of data gathering and processing. This phase includes the following steps:

1. DMs Identify stakeholder groups: In this step, DMs identify the stakeholders to participate in the evaluation process and obtain management approval. This identification is in line with the organizational mission, objectives, and management's fiscal year goals. Three groups of stakeholders are identified to evaluate fiscal year 1997 ESRs include: *Safety and Reliability, Supportability and Obsolescence,* and *Cost Benefit and Process Enhancement.*

2. DMs utilize AHP and EC to determine the importance weight of each stakeholder group (first round): AHP was introduced by Saaty (1972 and 1977a) to assist a DM in evaluating

complex judgmental problems. AHP helps the DM assign numerical values to qualitative attributes by making trade-offs among them. The process which is described in Step 5, is confined to a series of pairwise comparisons. Saaty (1972) argues that a DM naturally finds it easier to compare two things than to compare all the items in a list. AHP also evaluates the consistency of the DM and allows for the revision of the responses. Because of the intuitive nature of the process and its power in resolving the complexity in a judgmental problem, AHP has been applied to many diverse decisions. A comprehensive list of the major applications of AHP, along with a description of the method and its axioms, can be found in Saaty (1972, 1977a, 1977b, 1980, and 1990), Weiss and Rao (1987) and Zahedi (1986). At the beginning of each evaluation cycle, DMs individually use EC software which is based on AHP to determine the importance weights of each stakeholder group. The results from the first round are presented in Table 2.

Insert Table 2 Here

3. DMs utilize AHP and EC to determine the importance weight of each stakeholder group (second round): DMs meet and review the first round anonymous feedback concerning individual and group weights. They are encouraged to share their viewpoints and perceptions during this feedback session. At the end of the meeting, DMs are given the opportunity to revise their weights with EC, given their new insight and understanding from other individuals. The second round results are presented in Table 2.

Insert Table 2 Here

4. Stakeholder groups identify their subcriteria: Each stakeholder group holds a separate meeting and develops their own set of subcriteria to be used in the evaluation of ESRs. A listing of all subcriteria fro all stakeholders along with their importance weights (which are described next) are presented in Figure 2 and Table 3.

Insert Figure 2 and Table 3 Here

5. Stakeholder groups utilize AHP and EC to determine the importance weight of their subcriteria: Members of different stakeholder groups use EC in brainstorming sessions and determine their group weight for each subcriterion identified earlier in step 4. Assuming that in the stakeholder i's mind, $c_1, c_2, \ldots, c_{N_i}$ are the N_i subcriteria that contribute to an ESR success. The stakeholder's goal is to assess the relative importance of these subcriteria. Saaty's Analytic Hierarchy Process (AHP) is a method of deriving a set of weights to be associated with each of the N_i subcriteria. Throughout the AHP, Stakeholder i is asked to compare each possible pair c_j, c_k of subcriteria and provide quantified judgments on which one of the subcriteria is more important and by how much. These judgments are represented by an N_i x N_i matrix:

$$A = (a_{jk}) \qquad (j,k=1, 2, \ldots, N_i)$$

If c_j is judged to be of equal importance as c_k, then $a_{jk}=1$
If c_j is judged to be more important than c_k, then $a_{jk}>1$
If c_j is judged to be less important than c_k, then $a_{jk}<1$

$$a_{jk} = 1/a_{kj} \qquad a_{jk} \neq 0$$

Thus, the matrix A is a reciprocal matrix (i.e., the entry a_{jk} is the inverse of the entry a_{kj}). a_{jk} reflects the relative importance of c_j compared with subcriteria c_k. For example, $a_{12}=1.25$ indicates that c_1 is 1.25 times as important as c_2.

Then, the vector w representing the relative weights of each of the N_i subcriteria can be found by computing the normalized eigenvector corresponding to the maximum eigenvalue of matrix A. An eigenvalue of A is defined as λ which satisfies the following matrix equation:

$$A\,w = \lambda\,w$$

where λ is a constant, called the eigenvalue, associated with the given eigenvector w. Saaty has shown that the best estimate of w is the one associated with the maximum eigenvalue (λ_{max}) of the matrix A. Since the sum of the weights should be equal to 1.00, the normalized eigenvector is used. Saaty's algorithm for obtaining this w is incorporated in the software Expert Choice.

One of the advantages of AHP is that it ensures that stakeholders are consistent in their pairwise comparisons. Saaty suggests a measure of consistency for the pairwise comparisons. When the judgments are perfectly consistent, the maximum eigenvalue, λ_{max}, should equal N_i, the number of subcriteria that are compared. In general, the responses are not perfectly consistent, and λ_{max} is greater than N_i. The larger the λ_{max}, the greater is the degree of inconsistency. Saaty defines the consistency index (CI) as $(\lambda_{max} - N_i) / (N_i - 1)$, and provides the following random index (RI) table for matrices of order 3 to 10. This RI is based on a simulation of a large number of randomly generated weights. Saaty recommends the calculation of a consistency ratio (CR), which is the ratio of CI to the RI for the same order matrix. A CR of 0.10 or less is considered acceptable. When the CR is unacceptable, the stakeholder is made aware that his or her pairwise comparisons are logically inconsistent, and he or she is encouraged to revise their judgment. These importance weights are presented in Figure 2 and Table 2.

n	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
RI	*0.58*	*0.90*	*1.12*	*1.32*	*1.41*	*1.45*	*1.49*	*1.51*

Insert Figure 2 and Table 3 Here

6. Stakeholder groups identify probabilities of occurrence of their subcriteria for the ESRs: Each stakeholder group receives a listing of all ESRs under consideration from the GSWT-SC. The stakeholder group will assign a probability to each subcriterion under each ESR. The assignment of probabilities is done by the group in a brainstorming session. This result is presented in Table 4.

II. Integration Phase:

In this phase all the data collected during the Interaction Phase are integrated and processed using a series of software programs including EXCEL, EC, ENTROSYS, and MAHS.

7. ENTROSYS is utilized to revise the importance weight of each stakeholder group determined in the second round: Entropy concepts will be used to revise the second round weights of the subcriteria based on the information provided by the stakeholders concerning the probabilities. Entropy Measurement Sub-System (ENTROSYS), an automated system will be used to perform all necessary calculations. Given that each subcriterion is an information source, the more information is revealed by a subcriterion, the more relevant it is. This intrinsic information will be used in parallel with the stakeholder group weights. The probabilities of occurrence are used to measure this average intrinsic information. The more different the probabilities of a subcriteria are for a set of ESRs, the larger is the contrast intensity of the subcriterion and the greater is the amount of information transmitted by that subcriterion. The model views decision making as an information processing task and a large amount of information about the ESRs is processed through their subcriteria. Given the fact that subcriteria are information sources, the more information is revealed by the *j-th* subcriteria and the *i-th* stakeholder, the more relevant is the subcriteria in the decision analysis. Zeleny (1982) argues that this intrinsic information must be used in parallel with the initial weight assigned to various subcriteria by the DM. In other words, the overall importance weight of a subcriteria, F_{ij}, is directly related to the intrinsic weight, f_{ij}, reflecting average intrinsic information developed by a set of ESRs, and the subjective weight, W_i, reflecting the subjective assessment of its importance rendered by the DM. The probabilities of occurrence are used to measure this average intrinsic information. The more different the probabilities of a subcriteria are for a set of ESRs, the larger is the contrast intensity of the subcriteria, and the greater is the amount of information transmitted by that subcriteria. In this section, all formulas necessary for calculating the overall importance weight of opportunities are presented. Assume that vector $p_{ij} = (p_{ij}^1, ..., p_{ij}^q)$ characterizes the set P in terms of the *j-th* subcriteria for the *i-th* stakeholder and define:

$$P_{ij} = \sum_{m=1}^{q} p_{ij}^m \qquad (i = 1, 2, \ldots, N_i \text{ and } j = 1, 2, \ldots, N_j)$$

Then, the entropy measure of the *j-th* subcriteria for the *i-th* stakeholder is:

$$e(p_{ij}) = -K\sum_{m=1}^{q} \frac{p_{ij}^m}{P_{ij}} \ln \frac{p_{ij}^m}{P_{ij}}$$

Where $K>0$, $\ln$ is the natural logarithm, $0 \le p_{ij}^m \le 1$, and $e(p_{ij}) \ge 0$. When all p_{ij}^m are equal for a given i and j, then $p_{ij}^m / P_{ij} = 1/q$, and $e(p_{ij})$ assumes its maximum value, which is $e_{\max} = \ln q$. By setting $K = 1/e_{\max}$, we achieve $0 \le e(p_{ij}) \le 1$. This normalization is necessary for meaningful comparisons. In addition, the total entropy is defined as:

$$E = \sum_{j=1}^{N_j} e(p_{ij})$$

The smaller $e(p_{ij})$ is, the more information is transmitted by the *j-th* subcriteria for the *i-th* stakeholder and the larger $e(p_{ij})$, the less information is transmitted. When $e(p_{ij}) = e_{\max} = \ln q$, the *j-th* subcriteria for the *i-th* stakeholder is not transmitting any useful information. Next, the intrinsic weight is calculated as:

$$f_{ij} = \frac{1}{N_i - E}\left[1 - e(p_{ij})\right]$$

Since f_{ij} is inversely related to $e(p_{ij})$, $1 - e(p_{ij})$ is used instead of $e(p_{ij})$ and normalized to make sure $0 \le f_{ij} \le 1$ and

$$\sum_{j=1}^{N_j} f_{ij} = 1$$

The more different the subjective probabilities, p_{ij}^m, are, the larger f_{ij}, and the more important the *j-th* subcriteria for the *i-th* stakeholder is. When all the subjective probabilities, p_{ij}^m, are equal, then $f_{ij} = 0$. In order to calculate the overall importance weight of the *j-th* subcriteria for the *i-th* stakeholder, F_{ij}, the intrinsic weight, f_{ij}, is multiplied by the subjective weight, w_{ij}, and then the product is normalized:

$$F_{ij} = \frac{f_{ij}.w_{ij}}{\sum_{j=1}^{N_j} f_{ij}.w_{ij}}$$

The revised importance weights along with the initial and intrinsic weights are presented in Table 5.

8. EXCEL is utilized to calculate PSFs and the committee ranking of ESRs: The model described next will integrate *importance weights of stakeholders* with the *weights for subcriteria* and the *probabilities of occurrence* to arrive at a set of PSFs. The higher the PSF, the more desirable an ESR is. These calculations are done using a simple model developed with Microsoft EXCEL.

9. EXCEL and MAHS are utilized to provide committee and consensus rankings of the ESRs enhanced with sensitivity analysis: Microsoft EXCEL and Maximize Agreement Heuristic System (MAHS) are used to provide a consensus ranking of the ESRs. Assume that each one of our d DMs has ranked q ESRs. Assuming further that the opinions of the d DMs are to be valued equally, the Maximize Agreement Heuristic (MAH) seeks to arrive at the consensus ranking of the ESRs for the group as a whole. According to Beck and Lin (1983), MAH defines an agreement matrix, A, where each element a_{mn} represents the number of DMs who have preferred ESR m to ESR n. Strict preference is important. If a DM is indifferent between m and n, he or she is not counted in a_{mn}. The sum of a_{mn} for each ESR m across all columns represents the positive preference vector, C, where

$$C_m = \sum_{n=1}^{q} a_{mn}, \qquad (m=1,2,...,q)$$

Similarly, the sum of a_{mn} for each ESR across all rows represents the negative preference vector, R, where

$$R_m = \sum_{n=1}^{q} a_{mn}, \qquad (m=1,2,...,q)$$

If for ESR m, $C_m=0$, implying that no DM prefers ESR m to any other ESR, ESR m is placed at the bottom [in subsequent iterations, at the next available position at the bottom] of the final consensus ranking. However, if for ESR m, $R_m=0$, implying that no DM prefers any other ESR over ESR m, ESR m is placed at the top [in subsequent iterations, at the next available position at the top] of the ranking.

When there are no zero values in either C or R, the difference in total decision maker agreement and disagreement ($C_m - R_m$) is calculated for each ESR, and ESR m with the largest absolute difference $|C_m - R_m|$ is considered. If ($C_m - R_m$) is positive, ESR m is placed in the next available position at top of the final consensus ranking, and if the difference is negative, ESR m is placed in the next available position at the bottom of the consensus ranking. Any ties are broken arbitrarily. Once an ESR is assigned a position in the final consensus ranking, that ESR is eliminated from

further consideration. The remaining ESRs form a new matrix and the process is repeated until all ESRs are ranked. ESR rankings of the voting members of the GSWT-SC are presented in Tables 5 and 6 and Figure 3.

Insert Tables 5 and 6 and Figure 3 Here

III. Interpretation Phase:

During this phase all the synthesized data are presented to the GSWT-SC for the purpose of decision making.

10. DMs discuss the consensus and committee rankings and recommend a final ranking of ESRs to management: DMs meet and discuss the results of committee and consensus rankings. A final recommendation that includes a ranking of all ESRs will be forwarded to management for approval.

11. Management reviews the DMs' ranking of ESRs and makes the final decision: Management reviews the recommendation of the DMs and after considering various organizational implications, makes the final Selection.

The Model

To formulate an algebraic model , let us assume:

S^m = Project Success factor of the *m-th* ESR; ($m = 1, 2, ..., q$)

W_i = The importance weight of the *i-th* Stakeholder; ($i = 1, 2, ..., N_i$)

F_{ij} = The Overall Importance Weight of the *j-th* Subcriterion and the *i-th* Stakeholder; ($i = 1, 2, ..., N_i$ and $j = 1, 2, ..., N_j$)

P_{ij}^m = The *m-th* Probability of Occurrence of the *j-th* Subcriteria for the *i-th* Stakeholder;($m = 1, 2, ..., q$; $i = 1, 2, ..., N_i$; and $j = 1, 2, ..., N_j$)

N_i = Number of Stakeholders

N_j = Number of Subcriteria for the *i-th* Stakeholder

Given the above notations, the Project Success Factor for the *m-th* ESR is:

$$S^m = \sum_{i=1}^{N_i} W_i \left(\sum_{j=1}^{Nj} F_{ij} \left(P_{ij}^m \right) \right)$$

Where:

$1 \geq S^m \geq$

$$\sum_{i=1}^{N_i} W_i = 1$$

$$\sum_{j=1}^{Nj} F_{ij} = 1$$

and

$0 \leq P_{ij}^m \leq 1$

References

Beck, M. P. and Lin, B. W. "Some heuristics for the consensus ranking problem", *Computers and Operations Research*. 10 (1983), 1-7.

Saaty, T. L. *An Eigenvalue Allocation Model for Prioritization and Planning*. Energy Management and Policy Center, University of Pennsylvania, 1972.

Saaty, T. L. "A Scaling Method for Priorities in Hierarchical Structures", *Journal of Mathematical Psychology*, 15 (1977a), 234-281.

Saaty, T. L. "Modeling Unstructured Decision Problems: A Theory of Analytical Hierarchies", *Proceedings of the First International Conference on Mathematical Modeling*, (Conference) (1977b), 69-77.

Saaty, T. L. *The Analytic Hierarchy Process*. New York, NY: McGraw-Hill, 1980.

Saaty, T. L. *Multicriteria Decision Making: The Analytic Hierarchy Process*. Pittsburgh: RWS Publications, 1990.

Weiss, E. N. and Rao, V. R. "AHP Design Issues for Large-Scale Systems," *Decision Sciences*, 18 (1987), 43-61.

Zahedi, F. "The Analytical Hierarchy Process--A Survey of the Method and Its Applications", *Interfaces*, 16 (1986), 96-108.

Zeleny, M. *Multiple criteria decision making*. New York, NY: McGraw-Hill, 1982.

Figure 1: Consensus Ranking Organizational Support System (CROSS)

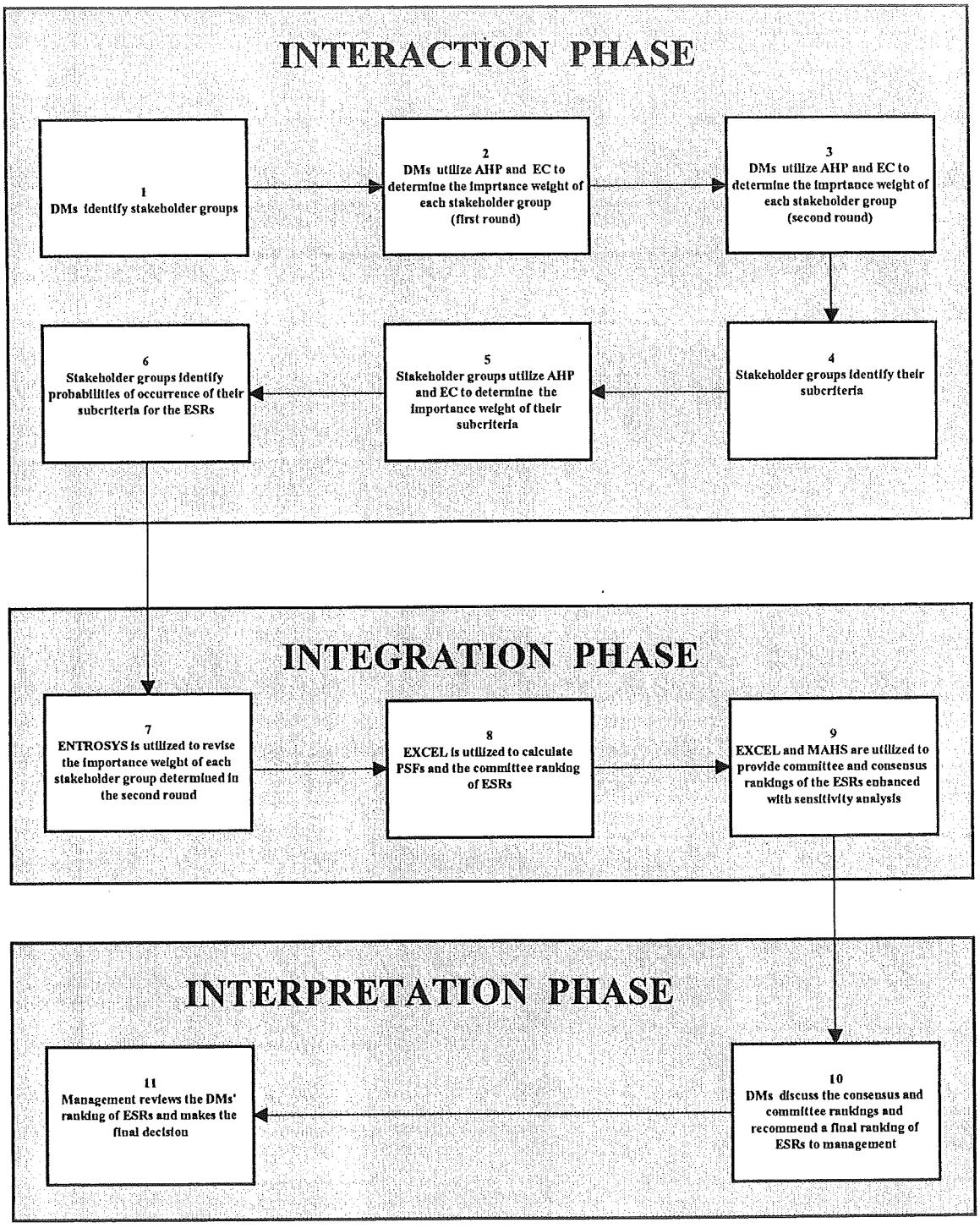


Figure 2: ESR Evaluation Subcriteria (Hierarchy and Relative Weights)

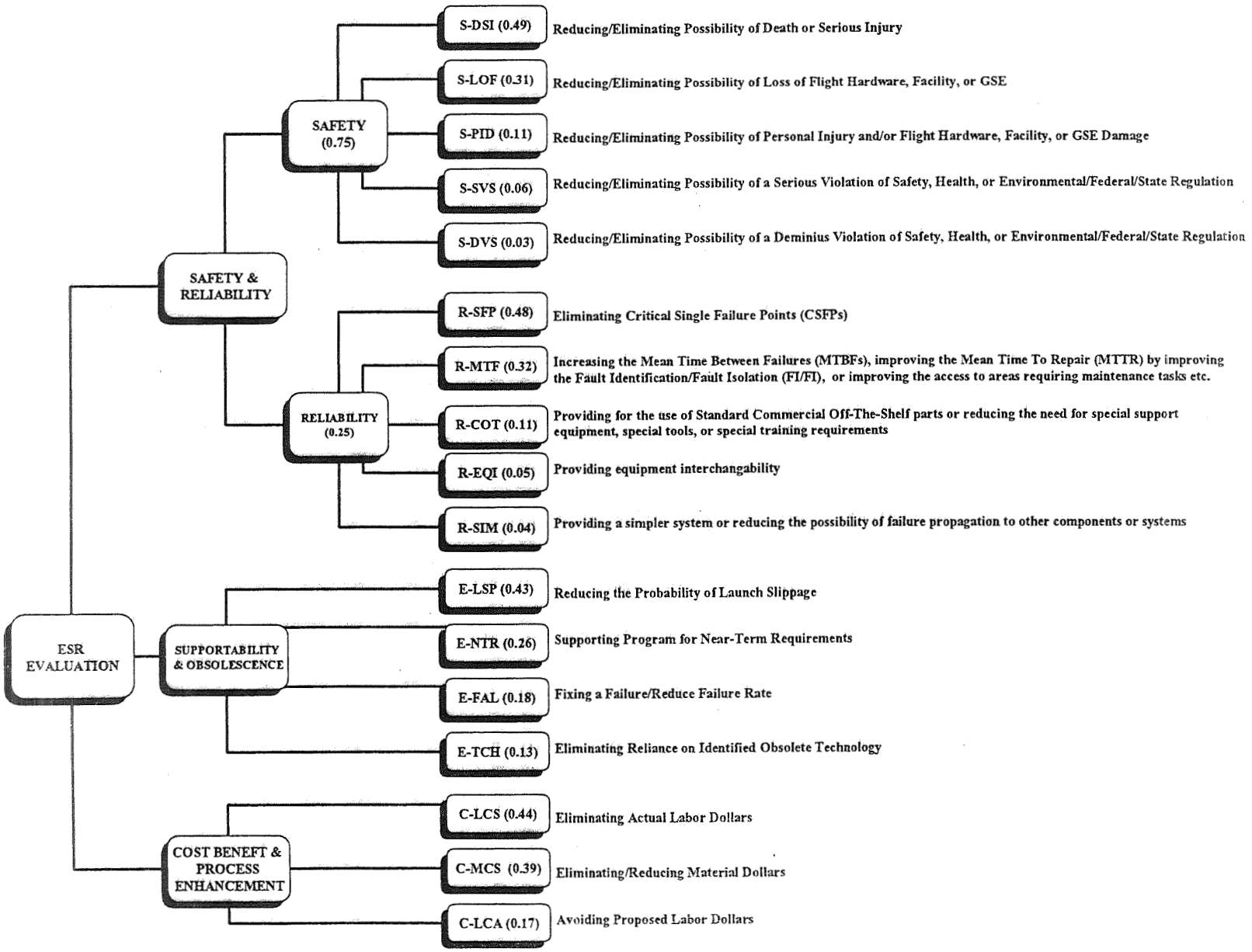


Figure 3: A Graphical View of the ESRs

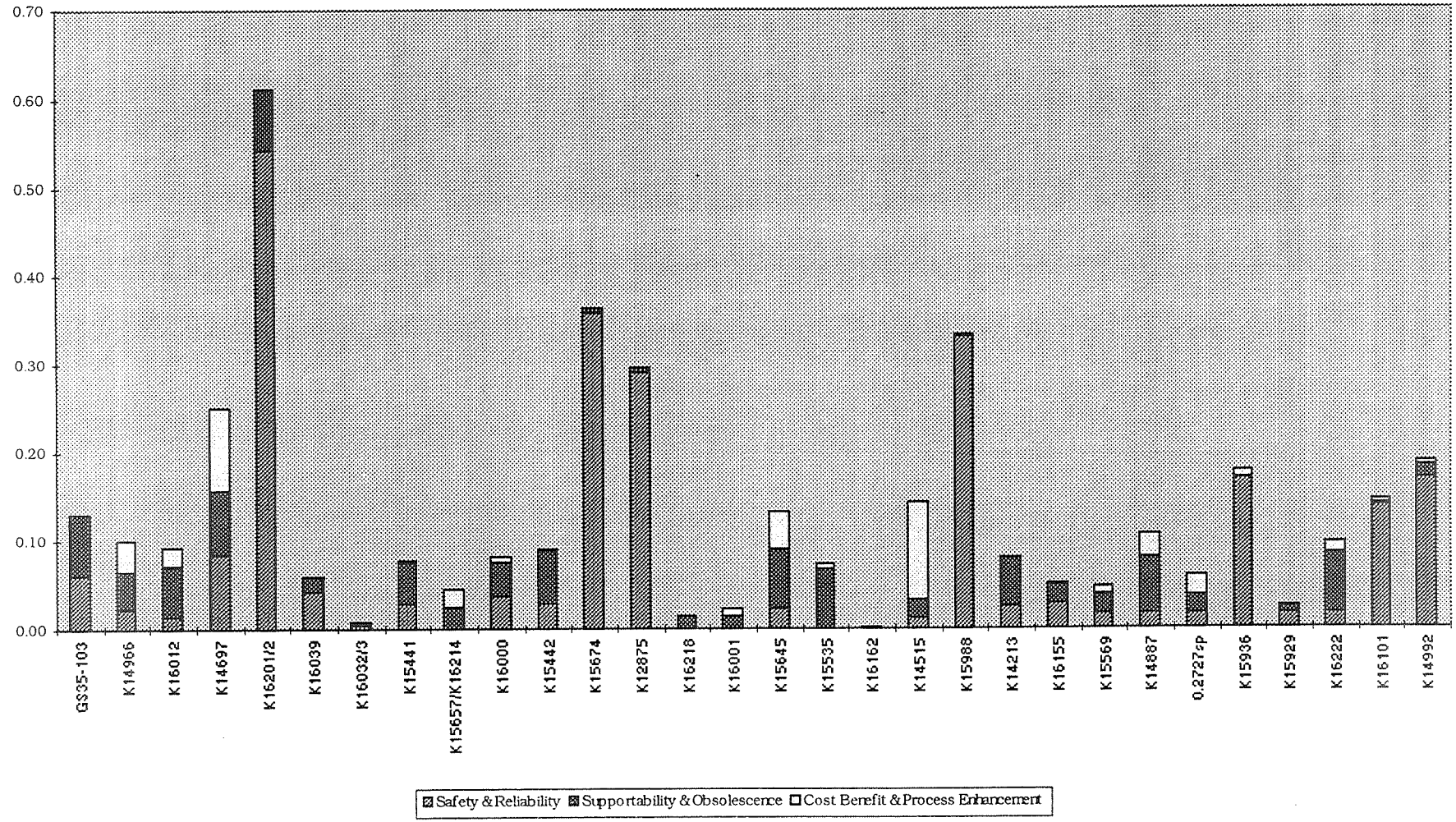


Table 1: Engineering Support Requests (Fiscal Year 1997)

Project No.	Organization	ESR No.	MOD
1	TE	GS35-103	$83,800
2	TE	K14966	$25,000
3	TE	K16012	$40,000
4	TE	K14697	$172,000
5	TE	K16201/2	$350,000
6	TE	K16039	$6,250
7	TE	K16032/3	$36,000
8	TE	K15441	$14,130
9	TE	K15657/K16214	$131,835
10	TE	K16000	$36,000
11	TE	K15442	$16,420
12	TE	K15674	$18,000
13	TE	K12875	$23,000
14	TE	K16218	$89,100
15	TE	K16001	$28,100
16	TV	K15645	$47,476
17	TV	K15535	$110,662
18	TV	K16162	$8,400
19	TV	K14515	$117,647
20	TV	K15988	$5,980
21	TV	K14213	$84,728
22	TV	K16155	$31,140
23	TV	K15569	$26,988
24	TV	K14887	$25,800
25	TV	0.2727sp	$37,000
26	TV	K15936	$29,000
27	TV	K15929	$13,938
28	TV	K16222	$20,000
29	TV	K16101	$4,200
30	TV	K14992	$2,856
		Total	$1,635,450

Table 2: Stakeholders' Relative Weights for the Voting Members of GSWT Steering Committee

Round 1

	Alexander	Allison	Kelley	Lee	Tootill	Normalized Mean
Safety & Reliability	0.71	0.65	0.76	0.76	0.63	0.71
Supportability & Obsolescence	0.16	0.28	0.15	0.15	0.28	0.20
Cost Benefit & Process Enhancement	0.13	0.07	0.09	0.09	0.09	0.09
Inconsistency Ratio	0.02	0.06	0.03	0.03	0.08	

Round 2

	Alexander	Allison	Kelley	Lee	Tootill	Normalized Mean
Safety & Reliability	0.60	0.57	0.76	0.63	0.53	0.62
Supportability & Obsolescence	0.25	0.33	0.15	0.26	0.33	0.26
Cost Benefit & Process Enhancement	0.15	0.10	0.09	0.11	0.14	0.12
Inconsistency Ratio	0.05	0.02	0.03	0.04	0.05	

Table 3: ESR Evaluation Subcriteria (Operational Definitions)

Code	Weight	Safety Subcriteria	Operational Definition
S-DSI	0.49	Reducing/Eliminating Possibility of Death or Serious Injury	Consequences of an action could be personal death or serious injury from potential safety and/or health hazard.
S-LOF	0.31	Reducing/Eliminating Possibility of Loss of Flight Hardware, Facility, or GSE	Consequences of an action could be loss of flight hardware, facility, or GSE from potential safety hazard.
S-PID	0.11	Reducing/Eliminating Possibility of Personal Injury and/or Flight Hardware, Facility, or GSE Damage	Consequences of an action could be personal injury and/or flight hardware, facility, or GSE damage from potential safety, health, and/or environmental hazard.
S-SVS	0.06	Reducing/Eliminating Possibility of a Serious Violation of Safety, Health, or Environmental/Federal/State Regulation	Consequences of an action could be a safety/federal citation and/or monthly fine arising from serious safety, health, and/or environmental standard violation.
S-DVS	0.03	Reducing/Eliminating Possibility of a Deminius Violation of Safety, Health, or Environmental/Federal/State Regulation	Consequences of an action could be a safety/federal citation arising from a deminius safety, health, and/or environmental standard violation.

Code	Weight	Reliability Subcriteria	Operational Definition
R-SFP	0.48	Eliminating Critical Single Failure Points (CSFPs)	Eliminates a component whose failure to perform its intended function could result in loss of life/vehicle, loss (damage) of a vehicle system, or loss of life/vehicle during the existence of a hazardous condition.
R-MTF	0.32	Increasing the Mean Time Between Failures (MTBFs), improving the Mean Time To Repair (MTTR) by improving the Fault Identification/Fault Isolation (FI/FI), or improving the access to areas requiring maintenance tasks etc.	Increases the average system operating time between failures of components of the system or of the entire system, reduces the average time it takes to repair an improperly functioning system by reducing the amount of time required to trouble-shoot and isolate a system problem, or improves the accessibility of maintenance personnel when their support is required.
R-COT	0.11	Providing for the use of Standard Commercial Off-The-Shelf parts or reducing the need for special support equipment, special tools, or special training requirements	Utilizes commercial, off-the-shelf components that have historical data available instead of unique, one-of-a-kind components or reduces the necessary support equipment, special tools, or unique training skills required to operate and maintain the system properly.
R-EQI	0.05	Providing equipment interchangability	Ensures that maintainability is not inhibited for all field units because of logistic problems associated with the special selection and storage of replacement components.
R-SIM	0.04	Providing a simpler system or reducing the possibility of failure propagation to other components or systems	Makes system success dependent on fewer items and thereby decreases the potential for failure of the system or reduces the chance that a component failure will propagate to another component within the system and/or to another system.

Code	Weight	Supportability & Obsolescence Subcriteria	Operational Definition
E-LSP	0.43	Reducing the Probability of Launch Slippage	Ability to reduce the probability of launch slippage.
E-NTR	0.26	Supporting Program for Near-Term Requirements	Ability to provide continuous support to launch and landing.
E-FAL	0.18	Fixing a Failure/Reduce Failure Rate	Ability of correcting an immediate (flow sensitive) failure.
E-TCH	0.13	Eliminating Reliance on Identified Obsolete Technology	Ability to eliminate identified and known obsolescence from occurring in systems and hardware.

Code	Weight	Cost Benefit & Process Enhancement Subcriteria	Operational Definition
C-LCS	0.44	Eliminating Actual Labor Dollars	Labor cost savings reflects actual labor dollars that are eliminated and are not available for other activities.
C-MCS	0.39	Eliminating/Reducing Material Dollars	Material cost savings reflect actual material dollars that are eliminated or reduced.
C-LCA	0.17	Avoiding Proposed Labor Dollars	Labor cost avoidance reflects proposed labor dollars that are avoided and are available for other activities.

Table 4: Stakeholders' Probability Judgments

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Safety and Reliability	S-DSI	0.10	0.00	0.00	0.00	1.00	0.10	0.00	0.00	0.00	0.00	0.00	1.00	0.60	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	S-LOF	0.10	0.00	0.00	0.20	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.20	0.80	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.90	0.00	0.00	0.80	0.00
	S-PID	0.10	0.00	0.00	0.40	0.50	0.10	0.10	0.00	0.00	0.00	0.00	1.00	0.60	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.90	0.00	0.00	0.00	0.90	0.00	0.00	0.90	1.00
	S-SVS	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	S-DVS	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	R-SFP	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
	R-MTF	0.00	1.00	0.50	1.00	1.00	1.00	0.00	1.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	1.00
	R-COT	0.50	0.50	0.50	1.00	1.00	0.00	0.00	1.00	0.00	0.50	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	R-EQI	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	R-SIM	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
Supportability and Obsolescence	E-LSP	0.10	0.00	0.10	0.10	0.10	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.00	0.20	0.00	0.00
	E-NTR	0.50	0.70	0.50	0.20	0.60	0.50	0.20	0.50	0.40	0.00	0.60	0.40	0.30	0.30	0.10	0.70	0.50	0.10	0.50	0.10	0.80	0.20	0.10	0.80	0.30	0.00	0.50	0.50	0.10	0.70
	E-FAL	0.40	1.00	0.50	0.80	0.40	0.20	0.00	0.50	0.00	0.40	0.60	0.00	0.00	0.30	0.40	0.00	0.90	0.00	0.40	0.00	0.10	0.50	0.00	1.00	0.50	0.00	0.00	0.80	0.00	0.10
	E-TCH	0.80	0.00	0.40	0.70	0.80	0.00	0.00	0.60	0.00	0.70	0.80	0.00	0.00	0.00	0.00	1.00	0.80	0.00	0.00	0.00	1.00	0.00	0.10	0.50	0.00	0.00	0.00	0.00	0.00	0.00
Cost Benefit and Process Enhancement	C-LCS	0.00	0.00	0.00	0.90	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.90	0.00	0.00	0.00	0.10	0.00	0.00	0.10	0.00	0.00	0.00	0.00
	C-MCS	0.00	0.80	0.50	0.90	0.00	0.00	0.00	0.00	0.40	0.10	0.00	0.00	0.00	0.00	0.00	1.00	0.20	0.00	1.00	0.00	0.00	0.00	0.00	0.50	0.40	0.00	0.00	0.30	0.10	0.00
	C-LCA	0.00	0.70	0.50	0.00	0.00	0.00	0.10	0.10	0.70	0.10	0.10	0.00	0.00	0.00	0.00	0.90	0.00	0.00	0.90	0.00	0.00	0.30	0.10	0.90	0.80	0.20	0.00	0.20	0.10	0.30

Table 5: ESR Rankings of the Voting Members of the GSWT Steering Committee

	Project Number																													
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**
Alexander	5	12	20	13	4	30	26	19	16	29	1	24	2	28	3	11	10	21	8	17	25	6	22	9	23	27	15	14	7	18
Allison	5	12	20	13	4	30	26	1	16	29	19	24	28	2	3	11	21	17	10	8	25	6	22	23	9	27	15	14	7	18
Kelley	5	12	20	13	4	30	26	29	1	19	16	2	24	10	11	28	3	8	21	6	22	25	17	23	9	27	15	14	7	18
Lee	5	12	20	13	4	30	26	29	19	1	16	24	2	28	3	11	10	21	8	17	6	25	22	23	9	27	15	14	7	18
Tootill	5	12	20	4	13	19	30	16	26	1	24	29	28	2	3	11	17	21	10	8	25	6	9	22	23	15	27	14	7	18
Consensus Ranking	**5**	**12**	**20**	**13**	**4**	**30**	**26**	**19**	**1**	**16**	**29**	**24**	**2**	**28**	**3**	**11**	**10**	**21**	**8**	**17**	**6**	**25**	**22**	**23**	**9**	**27**	**15**	**14**	**7**	**18**

Table 6: Consensus Ranking of the ESRs Using MAH

Rank	Project No.	Organization	ESR No.	MOD	CUM. MOD
1	5	TE	K16201/2	$350,000	$350,000
2	12	TE	K15674	$18,000	$368,000
3	20	TV	K15988	$5,980	$373,980
4	13	TE	K12875	$23,000	$396,980
5	4	TE	K14697	$172,000	$568,980
6	30	TV	K14992	$2,856	$571,836
7	26	TV	K15936	$29,000	$600,836
8	19	TV	K14515	$117,647	$718,483
9	1	TE	GS35-103	$83,800	$802,283
10	16	TV	K15645	$47,476	$849,759
11	29	TV	K16101	$4,200	$853,959
12	24	TV	K14887	$25,800	$879,759
13	2	TE	K14966	$25,000	$904,759
14	28	TV	K16222	$20,000	$924,759
15	3	TE	K16012	$40,000	$964,759
16	11	TE	K15442	$16,420	$981,179
17	10	TE	K16000	$36,000	$1,017,179
18	21	TV	K14213	$84,728	$1,101,907
19	8	TE	K15441	$14,130	$1,116,037
20	17	TV	K15535	$110,662	$1,226,699
21	6	TE	K16039	$6,250	$1,232,949
22	25	TV	0.2727sp	$37,000	$1,269,949
23	22	TV	K16155	$31,140	$1,301,089
24	23	TV	K15569	$26,988	$1,328,077
25	9	TE	K15657/K16214	$131,835	$1,459,912
26	27	TV	K15929	$13,938	$1,473,850
27	15	TE	K16001	$28,100	$1,501,950
28	14	TE	K16218	$89,100	$1,591,050
29	7	TE	K16032/3	$36,000	$1,627,050
30	18	TV	K16162	$8,400	$1,635,450

1996 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

THE USE OF A BLOCK DIAGRAM SIMULATION LANGUAGE FOR RAPID MODEL PROTOTYPING

Dr. Jonathan E. Whitlow, Associate Professor
Chemical Engineering Department
Florida Institute of Technology
Melbourne, Florida

KSC Colleague - Peter Engrand
Computer Science

Contract Number NASA-NGT10-52605

August 16, 1996

ACKNOWLEDGMENTS

I would like to express gratitude to my NASA colleague Peter Engrand for the opportunity to participate in the Summer Faculty Research Program for a second year. In addition, thanks are owed to Charlie Goodrich and John Dockendorf of INET and Larry Feinberg and Randy Lane of Rockwell for their help and technical support during my stay. I would also like to extend compliments to Dr. Roger Johnson and Kari Stiles of UCF as well as Greg Buckingham of NASA for providing a professional and stimulating summer program. Thanks also go to Joe Coffman for his valuable work and dedication to the project.

ABSTRACT

The research performed this summer was a continuation of work performed during the 1995 NASA/ASEE Summer Fellowship. The focus of the work was to expand previously generated predictive models for liquid oxygen (LOX) loading into the external tank of the shuttle. The models which were developed using a block diagram simulation language known as VisSim, were evaluated on numerous shuttle flights and found to work well in most cases. Once the models were refined and validated, the integration of the predictive methods were integrated into existing software utilized by Rockwell known as Propulsion Advisory Tool (PAT). Although time was not sufficient to completely integrate the models developed into PAT, the ability to predict flows and pressures in the orbiter section and graphically display the results was accomplished.

INTRODUCTION

Development of advanced software systems for launch support has been an ongoing task at KSC for a number of years. One launch support system which has received much attention is the loading of liquid oxygen LO_2. Two examples of software which have focused on this particular system for the purpose of fault detection and diagnosis are the Knowledge-based Autonomous Test Engineer (KATE) and the Propulsion Advisory Tool (PAT). The ability to develop predictive models for use in these or other software applications, can be a difficult task for a complex process such as LO_2 loading. The objectives of this work were two fold: 1)To develop predictive models for the LO_2 loading process using the block diagram simulation language VisSim and 2) Integrate the developed models into existing software for launch support.

MODEL DEVELOPMENT

The LO_2 predictive models were developed using a PC-based version of VisSim. VisSim is a block diagram simulation language which can be used to solve both differential and algebraic equations. Modeling equations which describe a physical system are transformed into a block diagram which is numerically solved as a function of time. A graphical interface is provided with pull down menus which allow blocks to be selected and wired together with the use of a mouse. In addition complex blocks can be formed in a hierarchial fashion to allow for natural organization

of the blocks which comprise the model. Figure 1 which gives a process flow and instrumentation diagram for the hardware in the vicinity of the orbiter associated with LO_2 loading is also a complex block using a bitmap image for representation. On top of the image are other complex blocks (i.e. pressure gauges, flow meters, valves, etc.), which if examined would reveal additional blocks underneath.

Before any modeling could be done on the LO_2 system, data from previous loading operations, pertaining to the instrumentation associated with the system had to be obtained. This was achieved by first broadcasting data recorded on a CD ROM using PC Goal over the network. The data was then recorded using KATE. Data was recorded from the T-6 hr mark which is prior to the LO_2 being loaded, until approximately two hours after the replenish phase had started. Once this data was recorded, it had to be put into a form which could be used by the VisSim program. This was done by running several PERL scripts which sorted through the recorded data and stored the data in a proper form in a series of ASCII files. These files have been compressed and stored on several floppy disks for future use if desired.

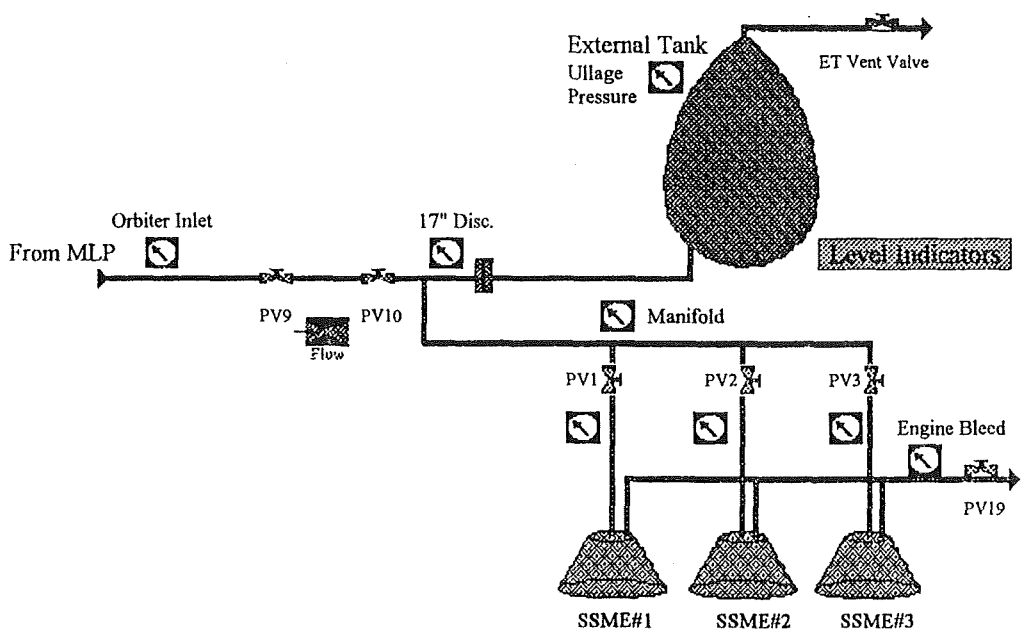


Figure 1 LO_2 Loading Process Flow & Instrumentation Diagram for the Orbiter Section

To completely model the loading of liquid oxygen into the external tank of the space shuttle, six separate phases as outlined in Table 1 must be accounted for. This work has primarily focused on the Fast Fill, Topping and Replenish phases and hence the description of the modeling effort outlined here pertains primarily to these regimes. Although the slow fill process is not explicitly modeled, the predictions should be applicable to this loading phase, since presumably only liquid oxygen is present. Chill down on the other hand can not be readily modeled and the models should not be used during this period, due to the presence of two phase flow, sensor saturation and rapidly changing dynamic conditions.

Table 1 Phases of Liquid Oxygen Loading

Chill down (T-8:00 to T-5:30)
Slow Fill ET to 2% (T-5:30 to T-5:15)
Fast Fill ET to 98% (T-5:15 to T-3:25)
Topping of ET to 100% (T-3:25 to T-3:20)
Auto Replenish (T-3:20 to Pre-launch)
Stop/Revert Flow (Abnormal Case)
Phases of Liquid Oxygen Loading

As with all physical systems which involve the transport of mass or energy, the principals of conservation apply. An equation for LO_2 conservation of mass for the external tank (ET) can be written as:

$$\rho A \frac{dh}{dt} = \rho F_{17'' \, inlet} - \rho F_{ET \, vent} \tag{1}$$

where:
ρ = *the density of* LO_2
A = *the cross sectional area of the ET*
h = *the level of* LO_2
$F_{17'' \, inlet}$ = *the flow of* LO_2 *through the* $17''$ *inlet*
$F_{ET \, vent}$ = *the flow of gaseous* O_2 *through the* $17''$ *inlet*

Due to the geometry of the ET, the cross sectional area changes as a function of liquid height and thus must be accounted for. This is done with VisSim using lookup tables contained in files in which values of area are obtained as a function of height. From the conditions modeled, the liquid density of the LO_2 is approximately constant although also included in the VisSim program are lookup tables for density as a function of temperature and pressure.

The flow of a fluid in a pipe between two points is proportional to the square root of the pressure difference between the points. Thus flow terms in equation 1 are defined by the equation:

$$F = \gamma(\Delta P)^{0.5} \tag{2}$$

where:
γ = *a flow coefficient* (*i.e. admittance*)
ΔP = *the pressure differential for the section of pipe considered*

Conservation of energy equations can also be written for key sections of the LO_2 loading process, however the observable variable temperature, has spatial variations as well as time dependency. This leads to complex solutions for partial differential equations as at each time step a temperature grid has to be solved by a numerical procedure such as finite differences. A simplification can be made if each section of pipe which is to be modeled is assumed to be a series of well stirred

compartments. The partial differential equations can thus be transformed into a series of ordinary differential equations. For a given section of pipe where only one compartment is assumed with one stream entering and one stream exiting the energy balance can be written :

$$\rho V C_p \frac{dT_o}{dt} = \rho C_p F T_i - \rho C_p F T_o + Q_{gain}$$

where:
T_o = *the temperature exiting the pipe*
T_i = *the temperature entering the pipe*
C_p = *the specific heat of the* LO_2
Q_{gain} = *the heat gain the surroundings*
F = *the flow of* LO_2 *through the pipe*
V = *the volume of the pipe*
ρ = *the density of the* LO_2 *inside the pipe* (3)

In equation 3 the heat gained from the surroundings can be written as:

$$Q_{gain} = UA(T_a - T_o)$$

where:
T_a = *the ambient temperature*
T_o = *the* LO_2 *temperature exiting the pipe*
U = *an overall heat transfer coefficent*
A = *the surface area of the pipe section* (4)

The LO_2 loading data for STS55 was used as the base case for model development. In developing the model it was assumed that all flow measurements were accurate and hence could be used to aid in the conservation of mass equations. Since there are two flow meters associated with the LO_2 pumps, the two values were averaged and then passed through a simple first order filter to eliminate some of the noise. From a mass balance, the flow of LO_2 going to the orbiter is equal to this filtered value for the pump flow, minus the flow recycled to the storage tank (bypass line). The first step in the model development was to assume a flow coefficient for the flow being returned to the storage tank as given by equation 2. The pressure drop used in the equation was from the pump discharge to the pressure in the bypass line. There are two separate return lines to the storage tank in the recycle stream. The smaller line is always open, while the larger line has a valve which is only opened after the fast fill operation has been completed. This leads to a need for different admittance values for the flow through this line. The flow coefficient for the bypass line can be readily determined during the replenish loading phase, since the total flow to the orbiter is equal to the flow through the replenish flow meter, since the valve on the transfer line fill valve is closed at that time.

Once the flow to the orbiter is set, the next step is to assume flow coefficients for the LO_2 flow through the main engines of the space shuttle. It was assumed that since these lines are the same dimensions with similar flow paths, that the flow and hence the flow coefficients would be equal

through each line. Subtracting this flow through the engines from that which is going to the orbiter, leaves the difference going up to fill the ET. Another flow coefficient which needs to be assumed is that for the ET vent valve during the times in which it is open.

All of the assumptions are checked through integration of the differential equation given in equation 1 to get the level as a function of time. Since level sensors are placed in the external tank at the 5%, 98%, 100-% and 100% marks, the values obtained from the numerical integration can be checked against the actual data. Once the replenish phase is entered the level remains approximately constant under closed loop control. The flow coefficients can then be iterated until reasonable values are achieved.

SIMULATION RESULTS

As previously discussed the LO_2 loading predictive models were developed using STS50 as the base case. Figure 2 shows how the measured values compare to predictive values for the orbiter inlet pressure. The prediction is generated from equation 2, by using the flow rate through the section of pipe between adjacent measurement points, the flow coefficient for the section and the measured pressure of the adjacent point. The other pressure measurements on the orbiter give accuracy similar to that shown in Figure 2.

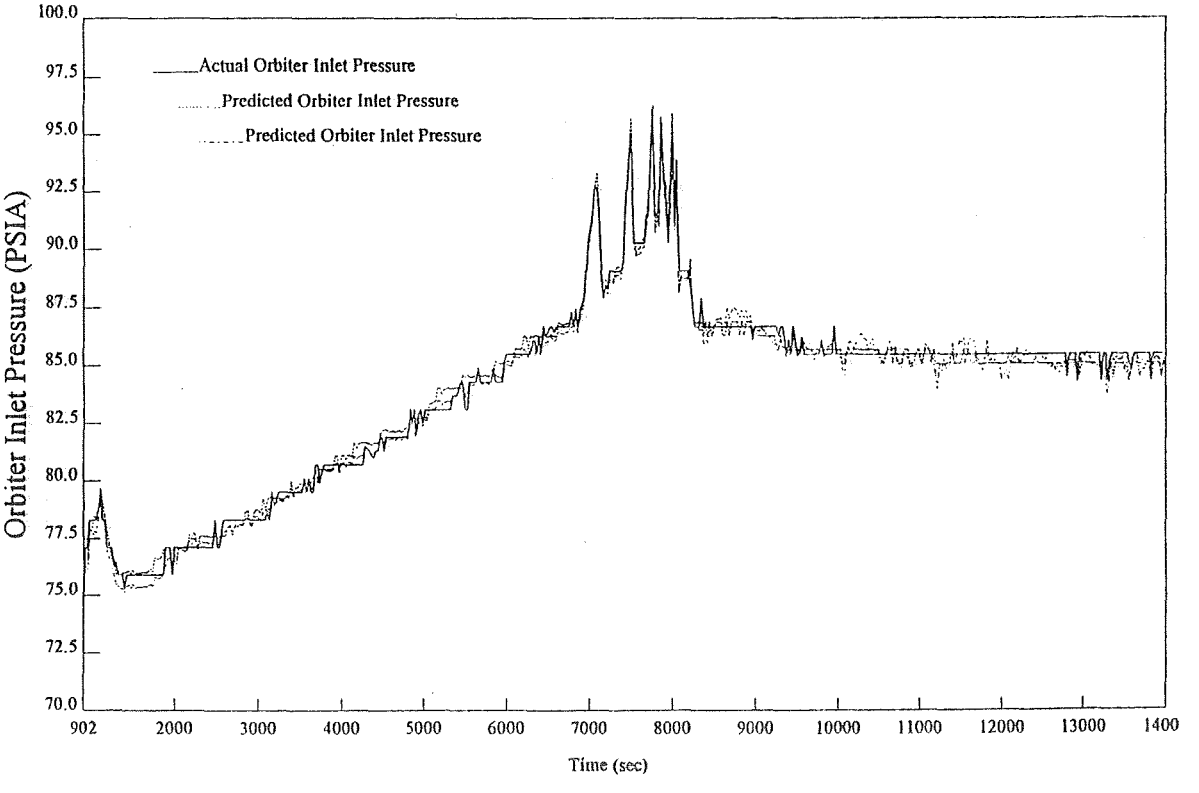


Figure 2 Predicted versus Actual Orbiter Inlet Pressure for STS50

Two predictive methods were used for each of these measurements as one section of pipe used was from the orbiter inlet to the 17 inch disconnect and the other section was from the manifold to each point. Two predictive methods were also used to predict the pressures in the engines, as the primary prediction used an average predicted flow through the engines, a flow coefficient and atmospheric pressure while the second prediction used the manifold pressure as the adjacent point. Since the manifold does not have an actual pressure measurement, one was generated in VisSim by taking a value between the orbiter inlet, 17 inch disconnect and engine inlet pressures. It was found that due to noise in the data and the relatively small differential pressure between measurements during the replenish phase, that an offset needed to be added to the orbiter inlet pressure, in order to insure that the manifold pressure was always between the measured values. While this worked well for STS50, using the generated manifold pressure for predicting adjacent pressures could not be repeated in other flights without adjusting this offset.

To evaluate the robustness of these models, other LO_2 data from different missions was also examined. Table 2 gives a comparison of the pressure predictions using the primary predictive method (i.e. the manifold pressure was not used) relative to the base case. It was found that the pressure predictions worked well for all flights, with the only difference coming from offset differences between orbiters. Figure 3 shows how the prediction for the orbiter inlet pressure has a relatively constant error or offset using the same offsets as those used for STS50. This evaluation of LO_2 on different flights shows that the models developed are robust and should be applicable to all loadings once an offset is determined for the particular orbiter.

Table 2 Comparison of Predicted Pressures Relative to Measured Pressures for Various Shuttle Missions

Orbiter	Mission	Orbiter Inlet Pressure	17" Disc. Pressure	SSME#1 Pressure	SSME#2 Pressure	SSME#3 Pressure
Columbia	STS40	OK	OK	OK	OK	OK
Columbia	STS52	OK	OK	OK	OK	OK
Columbia	STS55	OK	OK	OK	OK	OK
Atlantis	STS37	0-1 psi low	0-1 psi high	6-7 psi low	4-5 psi high	1-2 psi high
Atlantis	STS44	0-1 psi low	0-1 psi high	6-7 psi low	4-5 psi high	1-2 psi high
Atlantis	STS46	0-1 psi low	0-1 psi high	6-7 psi low	4-5 psi high	1-2 psi high
Endeavor	STS47	7-9 psi low	7-9 psi high	3-4 psi low	1-2 psi low	3-4 psi high
Endeavor	STS49	7-9 psi low	7-9 psi high	3-4 psi low	1-2 psi low	3-4 psi high
Endeavor	STS57	7-9 psi low	7-9 psi high	3-4 psi low	1-2 psi low	3-4 psi high

It was found that in general the predicted level in the ET as a function of time was close to the actual level, however there were errors as would be expected. Although in some cases the predicted level would be several feet off from the actual level, the error itself was small since the geometry of the ET has such a rapidly decreasing cross sectional area as it approaches being full.

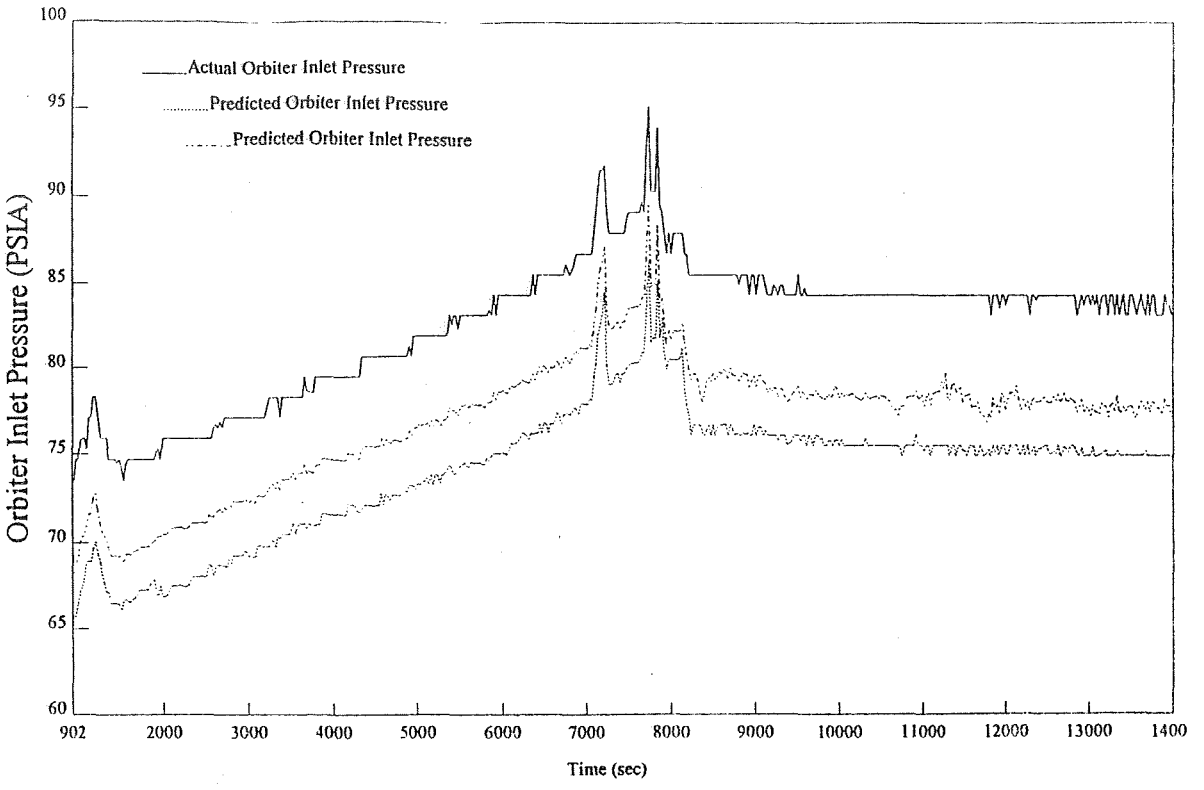


Figure 3 Predicted versus Actual Orbiter Inlet Pressure for STS57

To illustrate this, the difference in height between the 98% level sensor and 100% level sensor is greater than 3.5 feet, or roughly 7% of the height, while the volume of this section is only approximately 2% of the total volume.

Predictions were also generated for the temperatures in the orbiter section. For most of the fast fill phase the temperatures stay relatively constant and change only small amounts. Once the flow of LO_2 is decreased during the topping and replenish phases, all temperatures begin to rise and then ultimately come to a new steady state. This was modeled by performing a linear expansion of equation 3 and assuming that the dynamic response of the temperature can be modeled as a first order with time delay transfer function where the flow rate is assumed to be the only input variable effecting the output variable temperature. The response to the flow rate disturbance input can be seen in Figure 4. Although the model prediction has some error from the simplifications, it does seem to reasonably capture the dynamics.

MODELING FOR THE CASE OF LO_2 REVERT FLOW (DRAIN BACK)

As a separate case, an attempt was made to model the revert flow condition which occurs during a launch scrub. Data from STS71 was obtained for the entire drain period. No attempt was made to model temperatures for this case as only pressure predictions in the orbiter were evaluated. The cause for the scrub on this flight, was the failure of a fuse which had some of the key ground

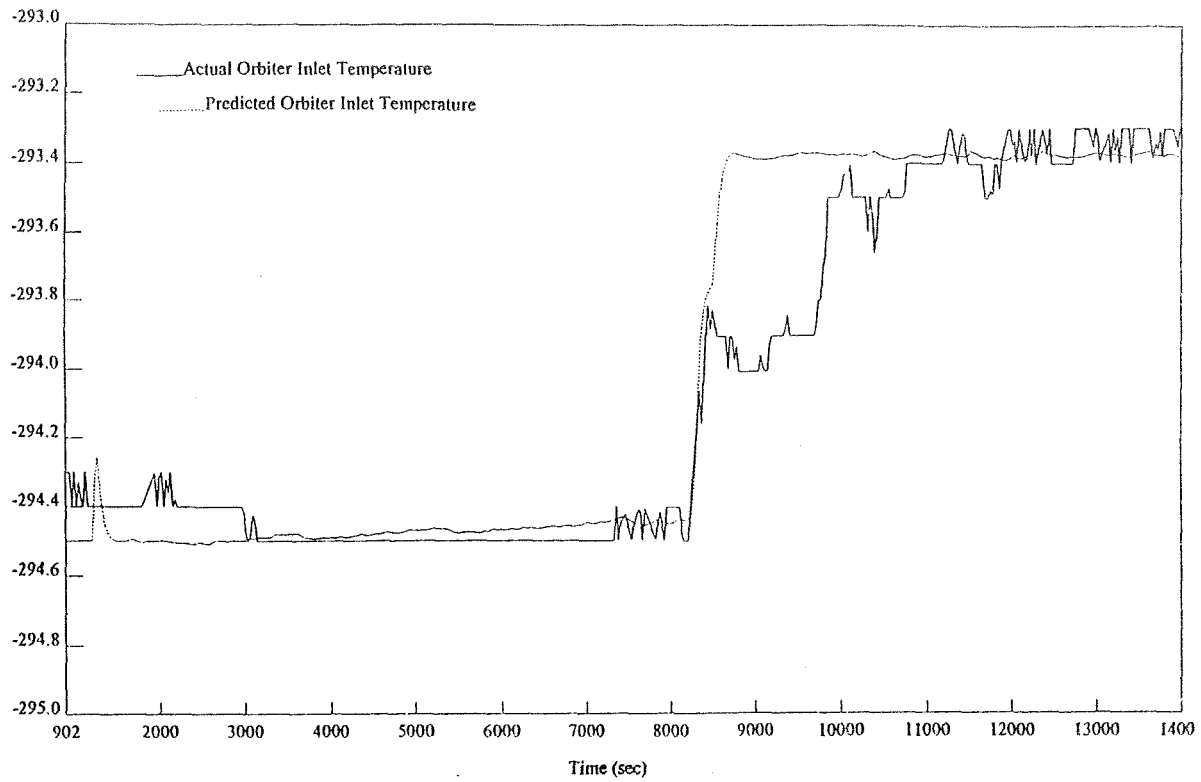


Figure 4 Predicted versus Actual Orbiter Inlet Temperature for STS50

measurements and valve state indicators. As a consequence the flow of LO_2 which was drained back to the storage tank was based on the measurements between the orbiter inlet pressure and the bypass pressure. A flow coefficient for this section of pipe was determined by iteration in a fashion similar to that outlined above. By having the integrated level and the actual level match at the times given by the data at which the 5% and 98% levels were reached.

In addition to the above analysis on the ET level, it was found that the pressures on the orbiter could also be predicted accurately once the appropriate offsets were included. Figure 5 shows the shuttle main engine #2 predicted and measured pressures as illustration of this fact. Although not shown, the orbiter inlet prediction had considerable error initially, due to the fact that the PV10 valve was actually closed during this period. This caused the actual pressure to drop, however once the valve was reopened the pressure prediction became valid.

INTEGRATION OF THE PREDICTIVE MODELS INTO PAT

The ultimate goal of this work was to use the predictive models which were developed using VisSim, in Rockwell's Propulsion Advisory Tool (PAT). PAT is a program written in G2 which is used by Rockwell for launch support. G2 is a graphical oriented programming tool which gives the user the ability to use rules, objects or procedures for software development. It is relatively user friendly, made evident by the fact that of the ten week period in the summer fellowship, only the last three weeks of the summer program were spent working on PAT.

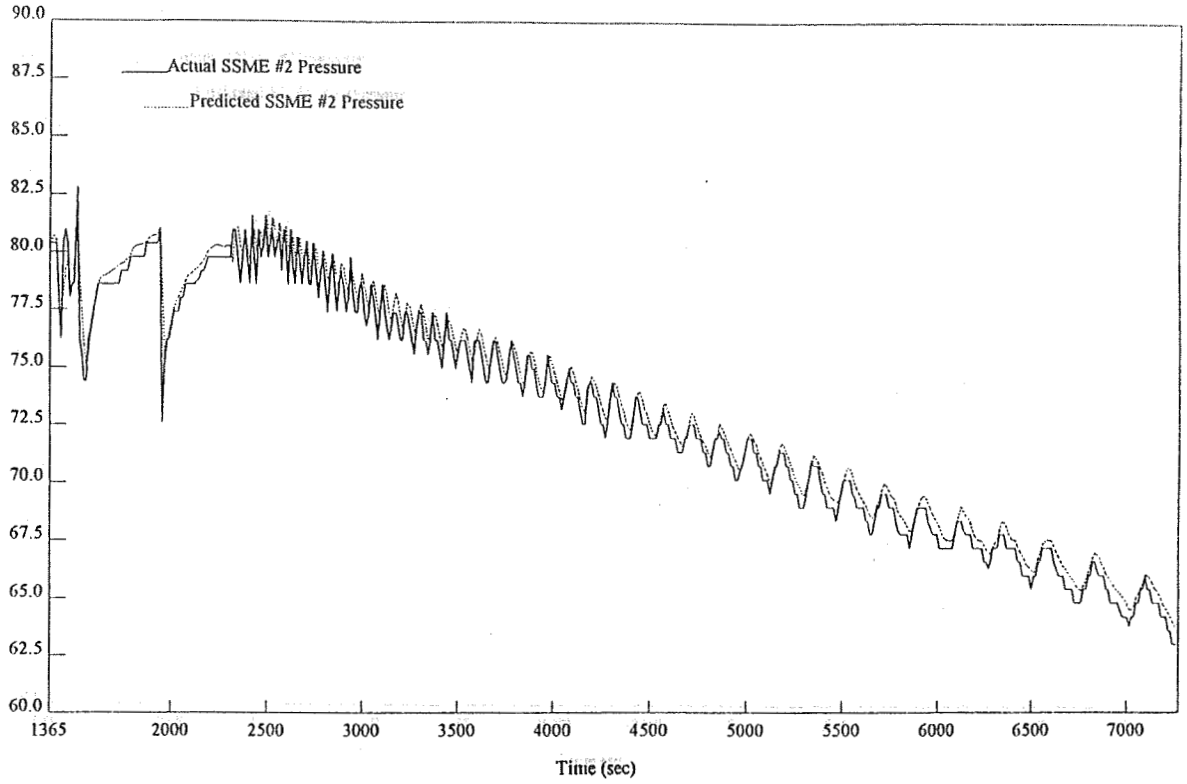


Figure 5 Predicted versus Actual SSME#2 Inlet Pressure for STS71

Due to the lack of time, all models were not added to PAT, several aspects of the model development were included, in particular pressure and flow predictions in key sections of the orbiter. The model components which were included in PAT were found to work well as they were evaluated using data from a recent LO_2 loading.

Generic functions were written in G2 for both flow prediction and pressure prediction and placed into a workspace. (i.e. G2 window) The workspace containing these functions in their native G2 language, along with some rules for the different loading phases are shown in Figure 6. Since many of the predictions in VisSim use simple first order filters, this capability was also included in PAT. Figure 7 illustrates the effect of the digital filter applied to the flow rate measurement from the LO_2 pumps.

CONCLUSIONS

The work which has been performed the past two years under the NASA Summer Faculty Fellowship has shown that a block diagram based simulation language such as VisSim can be used to rapidly develop robust predictive models. While additional work could be done to improve the robustness and predictive abilities of the models, the development effort has been largely successful. Furthermore, it has also been demonstrated that the results of the model development can also be incorporated into existing software with relative ease and similar success.

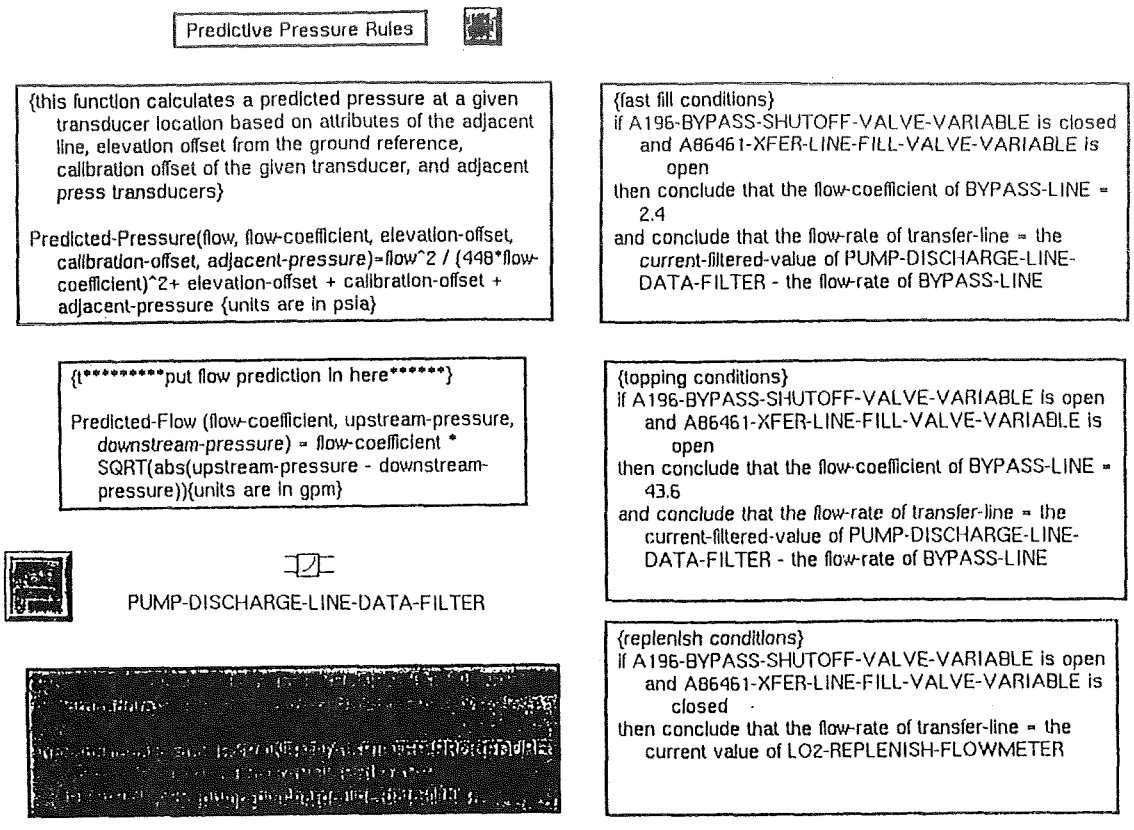


Figure 6 G2 Workspace Containing Predictive Rules and Procedures

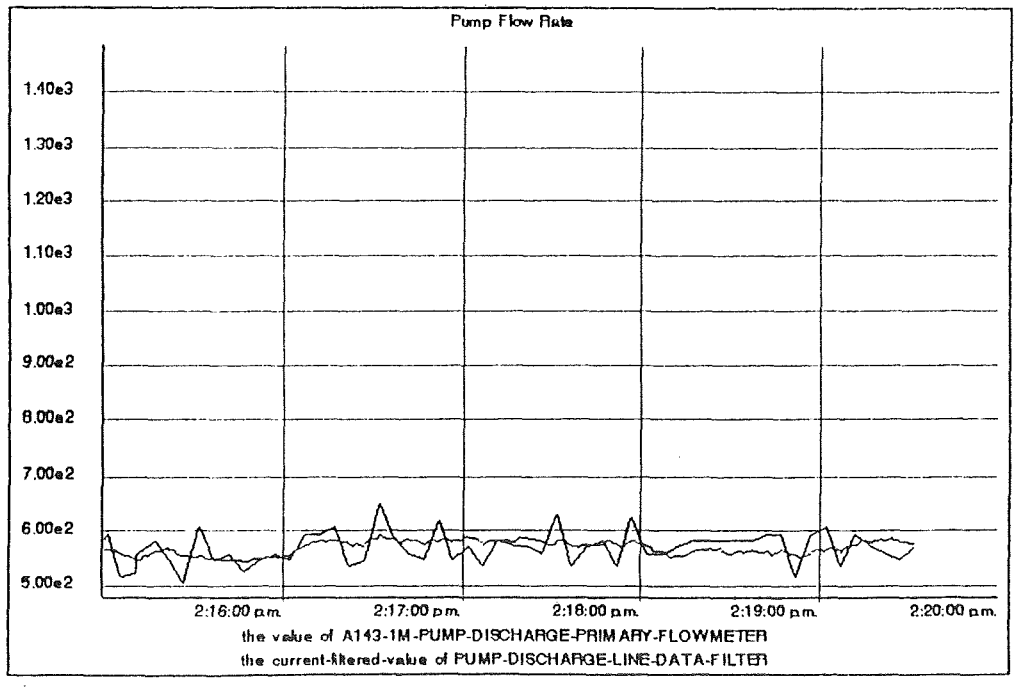


Figure 7 Comparison of Filtered and Actual LO_2 Flow Rates

1996 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

FOLLOW-THE-LEADER CONTROL FOR THE PIPS PROTOTYPE HARDWARE

Dr. Robert L. Williams II, Assistant Professor
Mechanical Engineering Department
Ohio University
Athens, Ohio

KSC Colleague - Thomas Lippitt
Robotics

Contract Number NASA-NGT10-52605

August 9, 1996

ABSTRACT

This report summarizes the author's summer 1996 work at NASA Kennedy Space Center in the Advanced Systems Division. The subject was the Payload Inspection and Processing System (PIPS). PIPS is an automated system, programmed off-line for inspection of Space Shuttle payloads after integration and prior to launch. PIPS features a hyper-redundant 18-dof serpentine truss manipulator capable of snake-like motions to avoid obstacles. During the summer of 1995, the author worked on the same project, developing a follow-the-leader (FTL) algorithm in graphical simulation which ensures whole-arm collision avoidance by forcing ensuing links to follow the same tip trajectory. The summer 1996 work was to control the prototype PIPS hardware in follow-the-leader mode.

This report summarizes improvements in the algorithm accomplished this summer. Angle-to-length mappings and length-to-LVDT voltage calibrations are presented; these were required for FTL hardware implementation. The algorithm was improved with a general feed-line for FTL (rather than straight out from the zero angles as last summer), reduced iterations for solution convergence, and the inclusion of joint limit checks for trajectories. Teleoperation was developed and implemented as the primary path planning mode for the prototype hardware. In this mode, the operator defines obstacle-free trajectories for the manipulator tip using a hand controller, either off- or on-line. Improvements in the existing low-level *C* code were made to enable FTL motions. *C++* code was developed to run the FTL algorithm on-line; this code was interfaced to the low-level control *C* code. A videotape was produced to document proof-of-concept FTL control of the prototype hardware.

The project was successful in providing FTL control in hardware. The STS-82 payload mockup was used in the lab to demonstrate serpentine motions to avoid obstacles in a realistic environment. Four trajectories are delivered for this payload in this report. A general FTL prototype hardware demonstration capability including teleoperation is the primary accomplishment of the summer.

This ten-page report presents highlights of the thirty-seven-page report delivered to Tom Lippitt of NASA KSC at the end of the project [6]. Please request the full version from the author or Tom Lippitt if desired.

1. INTRODUCTION

Inspection of Space Shuttle payloads after integration and prior to launch is essential for launch and mission safety. Currently, this inspection is completed by humans, which is dangerous, costly, labor-intensive, and not versatile in the cluttered and sensitive Shuttle bay environment. With shrinking budgets, development of efficient, labor saving methods are warranted. Therefore, the Advanced Systems Division at NASA Kennedy Space Center (KSC) is developing an automated tool, the Payload Inspection & Processing System (PIPS), for prelaunch inspection and light tasks in the Space Shuttle bay [1],[2], [3]. Figure 1 shows the design concept for PIPS. This unique device features a hyper-redundant serpentine truss manipulator (STM) for carrying a camera along obstacle-free trajectories to required goal points for inspection. The prototype PIPS hardware (built by Foster-Miller, Inc. [4] and modified by NASA) has eighteen degrees-of-freedom.

The author worked on the same project during the summer of 1995, where a follow-the-leader algorithm was successfully developed and implemented to the KSC serpentine truss manipulator in graphical simulation. Given a obstacle-free trajectory for the manipulator tip , the follow-the-leader algorithm ensures whole arm collision avoidance by forcing each ensuing link to follow the tip. The primary goal of the summer 1996 work was implementation of the follow-the-leader algorithm to the prototype hardware.

This report is organized as follows. Section 2 summarizes improvements in the general FTL algorithm including a general feed-line, teleoperation (human-based path planning), and joint limit checks. Section 3 discusses hardware implementation of the FTL algorithm and STS-82 hardware trajectory simulations. The conclusion follows, including a summary of accomplishments and design lessons learned from simulation.

This report presents project highlights because it is limited to ten pages in length. For detailed reports on two summers' work on this project, please see [5, 6].

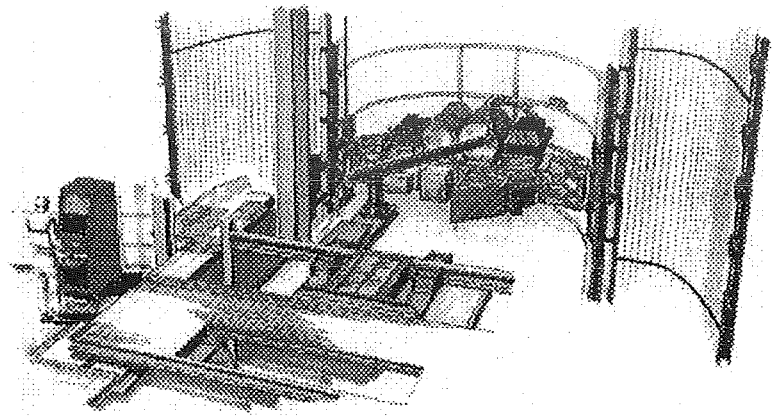


Figure 1.1 PIPS Conceptual Design

2. ALGORITHM IMPROVEMENTS

2.1 General Feed-Line

In last summer's algorithm, the STM was constrained to start follow-the-leader trajectories from the nominal reset position, where all joint angles are zero and the STM configuration is straight out along the prismatic joint. With this limitation, the crane joint θ_2 and the third base joint θ_3 are never used during trajectories. A significantly greater follow-the-leader workspace is enabled if the STM is fed onto the trajectory along general lines in space, rather than just along the prismatic track.

Figure 2.1 shows the general feed-line trajectory geometry. Length L_1 is the X_0 distance from the origin to point {4} in the reset position and length L_2 is the straight STM distance from spine points {4} to {18}. As discussed in Section 2, joints d_1, θ_2, and θ_3 are used to push spine point {4} onto trajectories. Since spine point {4} is the first to be pushed onto the trajectory, the feed-line starts at the nominal reset position for this spine point, as before. However, the initial position for the STM tip, spine point {18}, is determined by a sequence of two rotations: 1) α about Y_0; and 2) β about X_0. This sequence is a Y-X (α,β) fixed rotation, described by the rotation matrix:

$$ {}^0_F R = R_X(\beta)R_Y(\alpha) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & c\beta & -s\beta \\ 0 & s\beta & c\beta \end{bmatrix} \begin{bmatrix} c\alpha & 0 & s\alpha \\ 0 & 1 & 0 \\ -s\alpha & 0 & c\alpha \end{bmatrix} = \begin{bmatrix} c\alpha & 0 & s\alpha \\ s\alpha s\beta & c\beta & -c\alpha s\beta \\ -s\alpha c\beta & s\beta & c\alpha c\beta \end{bmatrix} \tag{2.1} $$

For all trajectories, the first two points are: ${}^F P_1 = \{0 \;\; 0 \;\; 0\}^T$ and ${}^F P_2 = \{0 \;\; 0 \;\; L_2\}^T$. (Note: ${}^B P_C$ is the vector to the origin of frame $\{C\}$ from the origin of frame $\{B\}$, expressed in the coordinates of frame $\{B\}$). The first point cannot be reached by the STM tip but must be defined in order to intersect the STM back onto the feed-line. If remaining path is determined in the $\{F\}$ frame, we must first transform all trajectory points to {0}:

$$ {}^0 P_i = {}^0_F T \, {}^F P_i \tag{2.2} $$

where (see Fig. 2.1):

$$ {}^{0}_{F}T = \begin{bmatrix} c\alpha & 0 & s\alpha & X_S \\ s\alpha s\beta & c\beta & -c\alpha s\beta & 0 \\ -s\alpha c\beta & s\beta & c\alpha c\beta & L_1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.3} $$

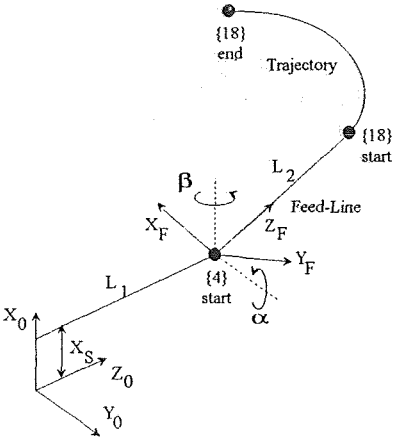


Figure 2.1 General Feed-Line Geometry

2.2 Teleoperation

Teleoperation is the most effective, reliable, and safe method for path planning of the PIPS system. Under teleoperation, a human operator enters commands to a robot system via a hand controller and the system responds. In the context of hyper-redundant serpentine manipulators, the operator enters obstacle-free trajectories for the manipulator tip, step-by-step. The follow-the-leader algorithm ensures that whole arm collision avoidance is maintained, also step-by-step. If any step results in a violation of joint limits, the operator is given the option to try again, but the bad command set is not sent to the manipulator. To extract the manipulator from a teleoperated trajectory, the command history is reversed.

For spatial teleoperation, three-dof input is sufficient, which controls relative XYZ positions. Orientation at the manipulator tip (camera pointing vector) is fixed by the relative locations of the last two trajectory points. The three-dof input could be chosen to be $\Delta X, \Delta Y, \Delta Z$. However, for follow-the-leader control, it is more convenient to use spherical coordinates to define next trajectory point relative to current trajectory point. Starting from the current trajectory point, the next point is defined using a hand controller to input a radius P and two spherical angles, ϕ, θ. As shown in Fig. 2.2, the hand controller can be aligned with the manipulator tip video monitor so teleoperation is natural. Input motion is relative to the manipulator tip coordinate frame $\{i\}$. Teleoperation is greatly enhanced by placing two or three cameras in the workspace to provide orthogonal views. In this report, the computer keyboard was used to simulate a virtual hand controller (typing numerical commands). Teleoperation would be much easier using an actual hand controller with proportional readings from each axis.

Figure 2.3 shows the i^{th} teleoperation step where the next trajectory point ${}^{0}\left\{{}^{0}P_{i+1}\right\}$ is determined based on the current trajectory point ${}^{0}\left\{{}^{0}P_{i}\right\}$ using the following vector-loop-closure equation:

$$ {}^{0}\left\{{}^{0}P_{i+1}\right\} = {}^{0}\left\{{}^{0}P_{i}\right\} + {}^{0}\left\{{}^{i}P_{i+1}\right\} \tag{2.4} $$

(Note: ${}^{A}\left\{{}^{B}P_{C}\right\}$ is the vector to the origin of frame $\{C\}$ from the origin of frame $\{B\}$, expressed in the coordinates of frame $\{A\}$.) Coordinate frame $\{i+1\}$ is obtained by two rotations relative to $\{i\}$: 1) ϕ about X_i; and 2) θ about Y_{i+1} (the Y-axis resulting from the first rotation). This sequence is an X-Y (ϕ,θ) Euler rotation (Note: the order of matrix multiplication is opposite from the fixed rotation sequence presented in Section 2.1).

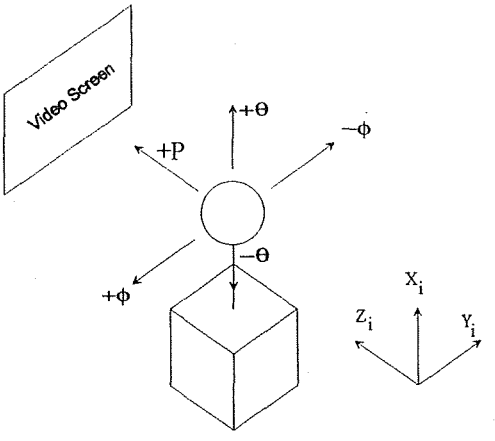


Figure 2.2 Teleoperation Hand Controller

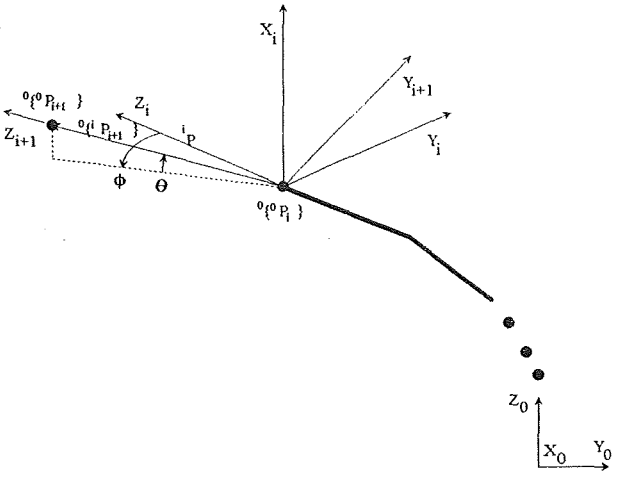


Figure 2.3 i^{th} Teleoperation Step

The relative vector ${}^{0}\left\{{}^{i}P_{i+1}\right\}$ for Eq. 3.6 is found from:

$$ {}^{0}\left\{{}^{i}P_{i+1}\right\}={}^{0}_{i}R\,{}^{i}\left\{{}^{i}P_{i+1}\right\} \tag{2.5} $$

The relative vector ${}^{i}\left\{{}^{i}P_{i+1}\right\}$ in frame $\{i\}$ is produced by rotating vector ${}^{i}P=\left\{0 \quad 0 \quad P\right\}^{T}$ through the X-Y (ϕ,θ) Euler rotation sequence described above:

$$ {}^{i}\left\{{}^{i}P_{i+1}\right\}=R_X(\phi)R_Y(\theta)\,{}^{i}P=\begin{Bmatrix} Ps\theta \\ -Pc\theta s\phi \\ Pc\theta c\phi \end{Bmatrix} \tag{2.6} $$

The rotation matrix forms are given in Eq. 2.1. In Eq. 2.5, the rotation matrix ${}^{0}_{i}R$ must be initialized to the starting orientation, at the second trajectory point (which is the first point the tip can reach). If the starting point is the nominal straight reset configuration, ${}^{0}_{1}R=I_3$. If a general feed-line is used, the initial orientation matrix is ${}^{0}_{1}R={}^{0}_{F}R$, given in Eq. 2.1. The rotation matrix ${}^{0}_{i}R$ must be updated after each successful teleoperation input as follows, to prepare for the next input step:

$$ {}^{0}_{i}R\rightarrow{}^{0}_{i+1}R={}^{0}_{i}R\,{}^{i}_{i+1}R \tag{2.7} $$

Where the rotation matrix ${}^{i}_{i+1}R$ comes from the current X-Y (ϕ,θ) Euler rotation sequence,

$$ {}^{i}_{i+1}R=R_X(\phi)R_Y(\theta) \tag{2.8} $$

2.3 Initial Q_i

As discussed in Section 2.2, the first step in calculating a follow-the-leader step is to shape the manipulator to the trajectory from tip to base. This is accomplished by intersecting manipulator segment link spheres of radii Q_i with the trajectory straight-line segments. Due to the offset universal joint structure of the prototype hardware, the intersect/inverse position kinematics computation is iterative. In last summer's work, the initial Q_i were taken as average values for generality [5]. This summer it was realized that if the initial manipulator configuration is straight (straight out or with a general feed-line), the initial Q_i values should be the maximum possible, $Q_i = \Delta_i + S_{i+1}$ because this leads to the exact solution the first iteration. Therefore, this improvement was made in the algorithm; the data is summarized in Table A.2.

In last summer's work, to achieve a spine point error tolerance of 0.1", two iterations were required at the initial step and only one thereafter since the Q_i values are continuously updated in the inverse kinematics solution. With the improvement in initial Q_i, only one iteration is required to achieve the error tolerance 0.1".

2.4 Joint Limit Checks

Last summer joint limit checks were based on plotting angular data resulting from the follow-the-leader data and inspecting to see if any limits are violated. This process was facilitated by bounding the graphs with the appropriate joint limit bounds, but it was not effective. This summer joint limit checks have been implemented in MATLAB and *C++* code in terms of logic statements. This is used in both automated (input *XYZ* data) and teleoperated modes. The variables **lolim** and **hilim** contain hardware joint angular (and d_1) limits which were derived from the actual measured limits on LVDT voltage and ball-screw actuator length. If joint limits are violated during simulation, the array **lims** is filled: with 0 if the corresponding joint does not meet a limit, and with 1 if the corresponding joint exceeds its limit. Then the array **jons** is displayed which contains the numerical values of the bad step, which of course cannot be sent to the hardware.

3. HARDWARE IMPLEMENTATION

3.1 Calibration

The follow-the-leader algorithm results in STM joint angle histories, eighteen joint commands for each input *XYZ* trajectory point. The STM hardware accepts eighteen LVDT voltage commands to drive each stepper motor / gear box / ball screw actuator combination. The LVDT voltage is the feedback sensor to ensure each joint moves to the commanded location. Therefore, a sequence of two transformations is required: 1) Joint angle to joint length mapping; and 2) Joint length to LVDT voltage calibration. The mathematics for these transformations is presented in the next two sub-sections. Both require extensive physical and electronic measurements, which are presented in appendices.

3.1.1 Angle-to-Length Mapping

This section presents the angle-to-length mapping required for each of the eighteen STM joints.

3.1.1.1 Prismatic Joint 1. The first joint is already a sliding joint, hence no mapping is required.

3.1.1.2 Crane Joint 2. Given a general θ_2, the corresponding L_2 must be calculated (see Figure 3.1):

$$L_2 = \sqrt{P_X^2 + P_Y^2} \tag{3.1}$$

The parameters for Eq. 3.1 are given below:

$$P_X = C + E\cos\beta$$
$$P_Y = -D + E\sin\beta$$

$$E = \sqrt{A^2 + B^2} \qquad \beta = \theta_2 + \phi_2 + 90^\square \qquad \phi_2 = \tan^{-1}\left(\frac{A}{B}\right)$$

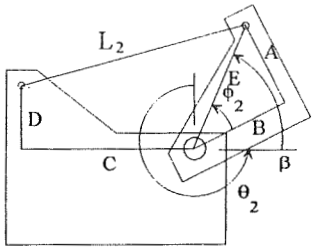


Figure 3.1 Crane Joint Model

3.1.1.3 Odd serpentine joints 3 through 17. The odd serpentine joints 3, 5, 7, 9, 11, 13, 15, and 17 have the same structure, called a type "A" box in [4]. In a type "A" Box, positive change in L_i corresponds to a positive change in θ_i (see Figure 3.2). At the nominal $\theta_i = 0$ position, the following relationships hold:

$$L_i = L - \delta \qquad \theta_0 = \tan^{-1}\left(\frac{L}{H}\right) - \tan^{-1}\left(\frac{\delta}{H}\right) \tag{3.2}$$

Given a general θ_i, the corresponding L_i can be calculated using Law of Cosines:

$$L_i = \sqrt{L_V^2 + H_V^2 - 2L_V H_V \cos(\theta_0 + \theta_i)} \tag{3.3}$$

where: $L_V = \sqrt{L^2 + H^2}$ $H_V = \sqrt{H^2 + \delta^2}$

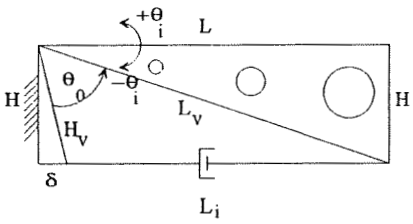


Figure 3.2 "A" Box Joint Model

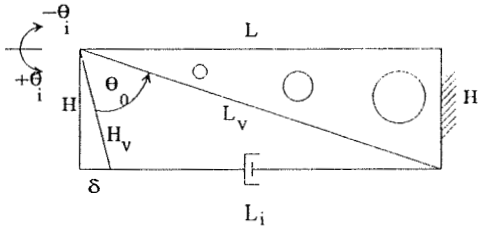


Figure 3.3 "B" Box Joint Model

3.1.1.4 Even serpentine joints 4 through 18. The even serpentine joints 4, 6, 8, 10, 12, 14, 16, and 18 have the same structure, called a type "B" box in [4]. In a type "B" Box, positive change in L_i corresponds to a negative change in θ_i (see Figure 3.3). At the nominal $\theta_i = 0$ position, Eq. 3.2 still holds. Given a general θ_i, the corresponding L_i can be calculated using Law of Cosines:

$$L_i = \sqrt{L_V^2 + H_V^2 - 2L_V H_V \cos(\theta_0 - \theta_i)} \tag{3.4}$$

3.1.2 Length-to-Voltage Calibration

This section presents the length-to-voltage calibration required for each of the eighteen STM joints. There are two different joint categories: 1) Motion base joints 1 and 2; and 2) Serpentine joints 3 through 18.

3.1.2.1 Motion base joints 1 and 2. Since the LVDT devices are linear, a linear calibration was performed between the minimum and maximum conditions. Given a general L_i, the LVDT voltage V_i is calculated using a linear equation for hardware feedback control.

$$V_i = V_{MIN} + \left(\frac{V_{MAX} - V_{MIN}}{L_{MAX} - L_{MIN}} \right) \left(L_i - L_{MIN} \right) \tag{3.5}$$

3.1.2.2 Serpentine joints 3 through 18. The calibrations for these joints are similar to Eq. 3.5. However, in order to ensure that the nominal STM zero position gives the measured nominal voltage V_0, a half-side calibration is used.

$$V_i = V_{MIN} + \left(\frac{V_0 - V_{MIN}}{L_0 - L_{MIN}} \right) \left(L_i - L_{MIN} \right) \tag{3.6}$$

3.2 Operational Scenario

There are several scenarios under which the hardware may be operated in follow-the-leader (FTL) mode. This section briefly discusses the options; please see [6] for more information.

The low-level control *C* code was developed by NASA and modified in this project. This code enables feedback control on each joint to achieve commanded LVDT voltages singly or in combinations.

Off-line MATLAB code was developed to program follow-the-leader motions, complete with graphics and animation. This code allows teleoperation, computation of *XYZ* trajectories, and input of externally-generated *XYZ* trajectories. The fidelity of the graphics is sufficient to avoid obstacles in the real hardware. The follow-the-leader algorithm generates histories of d_l and joint angles, which are mapped to lengths and the lengths are calibrated to equivalent LVDT voltages for each stepper motor / gearbox / ball screw actuator combination. The MATLAB code writes ASCII files to disk; the **VLT.DAT** or **TRAJ.DAT** files can be downloaded to the operational hardware *C* code to execute MATLAB-generated trajectories.

High-level follow-the-leader *C++* code was developed by graduate assistant James Mayhew to run the MATLAB code on-line. This code interfaces to the modified NASA low-level *C* code.

3.3 STS-82 Hardware Simulation

This section discusses hardware demonstration of follow-the-leader trajectories developed for the STS-82 payload, which is scheduled to fly in February, 1997. The hardware setup includes the eighteen-dof STM, Shuttle pallet, and mock-up STS-82 payload. The Shuttle pallet and STS-82 payload were modeled in MATLAB, along with the spine of the STM.

Four follow-the-leader trajectories were developed off-line using the MATLAB code and implemented in hardware control to demonstrate representative inspection locations and tasks for the STS-82 payload. All trajectories were developed free of hardware joint limits and proved to be free of collisions in hardware. The four trajectories are named below, along with output trajectory and voltage command data files.

1) *Remove WFPC SIPE Box Plastic Sheeting Task*	**TRAJ1_96.DAT**	**VLT1_96.DAT**
2) *Inspect Load Isolation System Strut Task*	**TRAJ2_96.DAT**	**VLT2_96.DAT**
3) *Inspect Keel Fitting Task*	**TRAJ3_96.DAT**	**VLT3_96.DAT**
4) *Retrieve Witness Plates Task*	**TRAJ4_96.DAT**	**VLT4_96.DAT**

A three minute videotape was produced by the PI to highlight the summer's accomplishments and demonstrate proof-of-concept follow-the-leader hardware control of the prototype STM hardware. Trajectory 3 is the featured trajectory and the motion is time-lapsed. Three final trajectories were developed and appear after the

narrated portion of the video. Figures 3.4 and 3.5 are photographs of the initial and final STM configurations for **TRAJ1_96.DAT**, respectively.

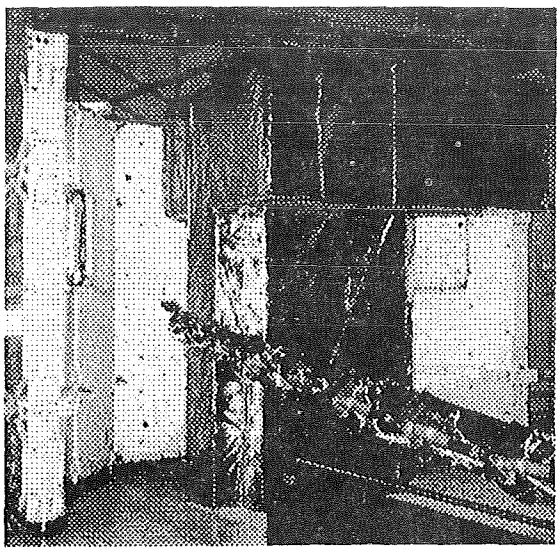


Figure 3.4 Initial STM Configuration

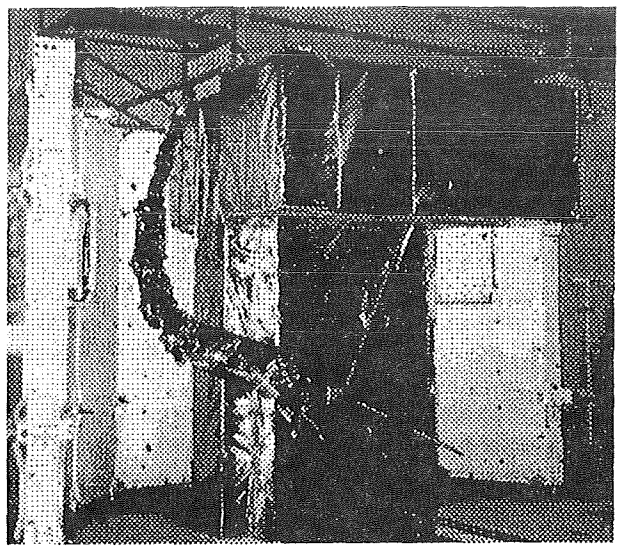


Figure 3.5 Final STM Configuration

4. CONCLUSION

This report highlights the PI's summer 1996 project in FTL hardware control of the prototype STM hardware. During last summer's project the PI developed the FTL algorithm and demonstrated it in graphical simulation. This year's project successfully accomplished FTL control in hardware, and several improvements were made to the algorithm in addition. Given an obstacle-free trajectory for the hyper-redundant manipulator tip, the FTL algorithm assures whole-arm collision avoidance for the entire trajectory by forcing all ensuing links to follow the tip.

Accomplishments for the summer 1996 project are as follows.

- Follow-the-leader (FTL) proof-of-concept demonstration in hardware
- Mapping and calibration from joint angles to LVDT voltages
- Determination and avoidance of hardware joint limits
- Upgrade FTL algorithm to include general feed-line and better convergence
- Implementation of teleoperation as primary path planning mode
- Adaptation of off-line MATLAB code to on-line C++ code
- Hardware simulation of STS-82 payload follow-the-leader trajectories
- Demonstration capability for future FTL tasks
- Videotape to document results and demonstrate hardware trajectories

The summer 1996 project was successful in demonstrating FTL trajectories in hardware. One product from this project over the past two summers is a list of design lessons learned from the prototype STM hardware and control system. These lessons should form part of the specifications for the final PCR PIPS hardware.

- STM joint offsets should be zero.
- The motion base translational travel should be equal to the STM length.
- The simplest FTL algorithm results from equal STM link lengths.

- All joint limits should be increased.
- The motion base must have more range in three dimensions.
- Servo controlled actuators should be used so all motors can reach their goals simultaneously.
- The hardware must be lighter yet also stiffer.
- Actuation redundancy should be provided in the event of joint failures.
- LVDT voltage noise must be reduced.

5. REFERENCES

[1] Richardson, B., Sklar, M., and Fresa, M., "PCR Inspection and Processing Robot Study, Final Report", McDonnell Douglas Space Systems - Kennedy Space Division, November, 1993.

[2] Pasch, K., "Self-Contained Deployable Serpentine Truss for Prelaunch Access of the Space Shuttle Orbiter Payloads", NAS-1659-FM-9106-387, Final Report, Contract NAS 10-11659, NASA Kennedy Space Center, FL, August, 1990.

[3] Herman, H., and Schempf, H., "Serpentine Manipulator Planning and Control for NASA Space-Shuttle Payload Servicing", CMU-RJ-TR-92-10, Carnegie-Mellon University, October, 1992.

[4] Snyder, M., "Self-Contained Deployable Serpentine Truss (SCDST) for Prelaunch Access of Space Shuttle Orbiter Payloads", NAS-1794-FM-9323-651, Final Report, Contract NAS 10-11794, NASA Kennedy Space Center, FL, October, 1993.

[5] Williams, R.L., II, "Follow-the-Leader Algorithm for the Payload Inspection and Processing System", Final Report, 1995 NASA/ASEE Summer Faculty Fellowship Program, NASA Kennedy Space Center, August, 1995.

[6] Williams, R.L., II, "Follow-the-Leader Control for the Payload Inspection and Processing System Prototype Hardware", Final Report delivered to Tom Lippitt, 1996 NASA/ASEE Summer Faculty Fellowship Program, NASA Kennedy Space Center, August, 1996.

[7] Craig, J.J., *Introduction to Robotics: Mechanics and Control*, Addison-Wesley Publishing Co., Inc., Reading, MA, 1988.

ATTACHMENT - REPORT DOCUMENTATION PAGE - ITEM #6 (Authors)

1996 Research Reports

NASA/ASEE Summer Faculty Fellowship Program

REPORT AUTHORS:

Dr. Mustafa A.G. Abushagur - University of Alabama-Huntsville

Mr. Randy K. Buchanan - Pittsburg State University (Kansas)

Dr. Luz M. Calle - Randolph-Macon Woman's College (Virginia)

Dr. Guillermo Colon - University of Puerto Rico-Mayaguez

Dr. Roger G. Ford - St. Mary's University (Texas)

Dr. Isaac Ghansah - California State University-Sacramento

Dr. David G. Jenkins - University of Illinois-Springfield

Dr. Khaled A. Kamel - University of Louisville (Kentucky)

Dr. Timothy G. Kotnour - University of Central Florida

Dr. Samuel P. Kozaitis - Florida Institute of Technology

Dr. David Kozel - Purdue University-Calumet (Indiana)

Dr. Jerome P. Lavelle - Kansas State University

Dr. Rasiah Loganantharaj - University of Southwestern Louisiana

Dr. Mark B. Moldwin - Florida Institute of Technology

Dr. Robert E. Peale - University of Central Florida

Dr. Marvin J. Pitts - Washington State University

Dr. Rodney G. Roberts - Florida A&M University-Florida State University College of Engineering

Dr. John M. Russell - Florida Institute of Technology

Dr. Ryan D. Stansifer - Florida Institute of Technology

Dr. Madjid Tavana - La Salle University (Pennsylvania)

Dr. Jonathan E. Whitlow - Florida Institute of Technology

Dr. Robert L. Williams - Ohio University

EDITORS:

Dr. Roger Johnson - University of Central Florida

Mr. Gregg Buckingham - John F. Kennedy Space Center

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	October 1996	Contractor Report - Summer 1996

4. TITLE AND SUBTITLE

1996 Research Reports
NASA/ASEE Summer Faculty Fellowship Program

5. FUNDING NUMBERS

NASA Grant
NGT10-52605

6. AUTHOR(S)

See attached list

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Central Florida
Orlando, Florida 32816-2450
John F. Kennedy Space Center
Kennedy Space Center, Florida 32899

8. PERFORMING ORGANIZATION REPORT NUMBER

NASA CR-202756

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

National Aeronautics and Space Administration
Washington, D.C. 20546

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Unclassified - Unlimited
Subject Category 99

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This document is a collection of technical reports on research conducted by the participants in the 1996 NASA/ASEE Summer Faculty Fellowship Program at the Kennedy Space Center (KSC). This was the twelfth year that a NASA/ASEE program has been conducted at KSC. The 1996 program was administered by the University of Central Florida in cooperation with KSC. The program was operated under the auspices of the American Society for Engineering Education (ASEE) with sponsorship and funding from the Office of Educational Affairs, NASA Headquarters, Washington, D.C. and KSC. The KSC Program was one of nine such Aeronautics and Space Research Programs funded by NASA in 1996. The NASA/ASEE Program is intended to be a two-year program to allow in-depth research by the University faculty member. The editors of this document were responsible for selecting appropriately qualified faculty to address some of the many problems of current interest to NASA/KSC.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Research and Technology	290
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL