NASA CR-166837

1989 RESEARCH REPORTS

NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

JOHN F. KENNEDY SPACE CENTER

UNIVERSITY OF CENTRAL FLORIDA

(NASA-CR-166837) NASA/ASEE SUMMER FACULTY
FELLOWSHIP PROGRAM (University of Central
Florida) 420 p CSCL 05I

N90-16686
--THRU--
N90-16702
Unclas
G3/99 0234818

1989 RESEARCH REPORTS

NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

JOHN F. KENNEDY SPACE CENTER

UNIVERSITY OF CENTRAL FLORIDA

EDITORS:

Dr. E. Ramon Hosler
Professor of Engineering
University of Central Florida

Mr. Dennis W. Armstrong
Systems Training and Employee Development Branch
Kennedy Space Center

PREPARED FOR:

John F. Kennedy Space Center
Merritt Island, Florida

NASA Grant NGT-60002 Supplement: 2

Contractor Report No. CR- 166837

October 1989

PREFACE

This document is a collection of technical reports on research conducted by the participants in the 1989 NASA/ASEE Summer Faculty Fellowship Program at Kennedy Space Center (KSC). This was the fifth year that a NASA/ASEE program has been conducted at KSC. The 1989 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. The KSC program was one of eight such Aeronautics and Space Research Programs funded by NASA Headquarters in 1989. 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 5 through August 11, 1989) 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 administration aspects of the 1989 program. The NASA/ASEE program is basically a two-year program to allow in-depth research by the University faculty member. In most 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

N90-16687

1989 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

FORMALISMS FOR USER INTERFACE SPECIFICATION AND DESIGN

PREPARED BY: Dr. Brent J. Auernheimer

ACADEMIC RANK: Associate Professor

UNIVERSITY AND DEPARTMENT: California State University - Fresno
Department of Computer Science

NASA/KSC

DIVISION: Data Systems

BRANCH: Real Time Systems Branch

NASA COLLEAGUE: Mr. Les Rostosky

DATE: August 28, 1989

CONTRACT NUMBER: University of Central Florida
NASA-NGT-60002 Supplement: 2

Acknowledgements

This work would not have been possible without the help of the following people — their assistance is sincerely appreciated. The NASA colleague for this project was Les Rostosky. Dave Roberts of NASA provided resources for word processing and access to the Internet.

Dr. Ramon Hosler guided me through NASA and UCF paperwork. His organization of the summer faculty program was conscientious and thorough. Kari Baird's enthusiasm was motivating at just the right times.

Finally, two academic colleagues were particularly helpful providing insights and references via electronic mail. I am grateful to Gary Perlman of the Ohio State University and Stephanie Doane of the Institute for Cognitive Science at the University of Colorado, Boulder for their help.

Abstract

This report describes the application of formal methods to the specification and design of human-computer interfaces. A broad outline of human-computer interface problems, a description of the field of cognitive engineering and two relevant research results, the appropriateness of formal specification techniques, and potential NASA application areas are described.

Table of Contents

I. Introduction

I.1 Organization of the Paper

This report describes the application of formal methods to the specification and development of human-computer interface software.

This paper is organized as follows: problems of interface specification and design and the inadequacies of traditional software engineering techniques are discussed. The emerging field of cognitive engineering as a new approach to solving user interface problems is described. Second, direct manipulation user interfaces are defined and contrasted with traditional command language interfaces. Third, a NASA context of shuttle and space station software is given. Fourth, two relevant research results are discussed. Both results are influenced by the emerging field of cognitive engineering. Techniques of reducing the confusing jumble of windows [1], and of formally specifying direct manipulation user interfaces [2] are presented. The feasibility and appropriateness of formal specification techniques for user interfaces is discussed, and user interface correctness criteria from a NASA user interface document are described. Finally, recommendations are given for using formal methods of user interface development and specification on NASA projects.

I.2 Background

In the beginning, the primary users of software were software developers themselves, or highly motivated users willing to endure bizarre idioms to take advantage of a computer's power. A typical user today is usually not a software developer and has little tolerance for productivity impediments imposed by software that is hard to use, learn, or is dangerous.

User interface software takes a significant amount of the total effort involved in the development of a software system. Estimates of of "40 to 50 percent of the code and runtime memory" [3, p. 15] being devoted to the user interface, and "50% of the coding effort in a typical data base application is usually spent on the implementation of the user interface" [4, p. 481], are common.

I.3 The Emerging Field of Cognitive Engineering

Possible solutions to the user interface crisis come from the emerging field of cognitive engineering.

Cognitive engineering is the offspring of cognitive science and software engineering. This multidisciplinary approach to human-computer interfaces is well-founded in the fields of human factors and human-computer interaction (HCI). [5] is a description of the interaction between cognitive psychology and computer science that results in the field of HCI.

A new, interdisciplinary academic discipline has formed to address general questions of human and computer cognition. Norman defines this new field of cognitive science [6, p. 326] as "a fusion of many disciplines, all concerned with different aspects of cognition.

It combines psychology, artificial intelligence, linguistics, sociology, anthropology, and philosophy."

Norman argues that the application of the disciplines of cognitive science to the "cognitive aspects of human-computer interaction" describes a new multidisciplinary approach of *cognitive engineering* [6, p. 326]. Throughout this paper, the term cognitive engineering will be used to include the field of HCI.

The important contributions of cognitive engineering are insights into what makes computers hard to use and learn. Norman argues [6, p. 331] that

> Computers can add to task difficulty when they stand between the person and the task, adding to the issues that must be dealt with to complete the task. When computer stand between, they act as an intermediary, requiring that the user be proficient in both the task domain and with the intermediary (the computer).

I.4 A Fallacy of Interface Design

A goal of cognitive engineers is to make the intermediary nature of the computer disappear allowing the user to exert their entire effort on the task.

Making the computer transparent was until recently not a concern of software developers. In the beginning, software developers themselves were the primary users of software. That is, the users of software had both computer and task knowledge. This naturally lead to what Landauer calls the "egocentric intuition fallacy" [7, p. 906]:

> One part of the egocentric intuition fallacy is the compelling illusion that one knows the determinants of one's own behavior and satisfaction. ... Even when a system is obviously hard [to use], intuition may not reveal the true reasons.

In addition to having flawed intuition about what makes a good interface, developers have mental models of software systems that are often very different from the operational mental model that users have of the same system [8]. A designer's natural tendency is to design to their mental model and not to the users' cognitive models.

I.5 The Contribution of Cognitive Engineering

The contribution that cognitive engineering makes to solving human-computer interface problems can be summarized as follows: cognitive engineering provides the sound experimental basis of human cognition necessary to build human-computer interfaces that enhance rather than hinder problem solving.

I.6 The Inadequacies of Present Software Engineering Techniques

Traditional software engineering techniques for interface design supported the development of command language interfaces. Command language interfaces consist of a typed dialogue between the system and the user. These software engineering tools were based on state machine and context free grammar models of command languages. Traditional tools, however, are inadequate for developing mouse-window interfaces. Fischer is convinced that "current life-cycle models of software engineering are inadequate for most of today's computing problems," [9, p. 44] particularly for user interfaces.

Recent developments in rapid prototyping, knowledge-based support tools, and specification systems are being applied to the development of user interfaces. Many of these second generation interface development tools are coming from cognitive science and engineering laboratories.

I.7 User Interface Management Systems

The class of interface development tools called *user interface management systems* (UIMSs) is rapidly maturing. Perlman defines a UIMS as "a set of software tools for developing and managing user interfaces" and elaborates [10, p. 823]:

> The fundamental role of user interface tools is to act as mediator: (1) to simplify the interface between users and the applications with which they interact; and (2) to simplify the task of the user interface developer to create the user-application interface.

Some UIMSs allow designers to compose interfaces with respect to a body of constraints or guidelines and have their designs critiqued [9], or design interfaces by demonstration, rather than coding [11]. In general, UIMSs support the development of good user interfaces by [12, p. 33]:

- Providing a consistent user interface between related applications
- Making it easier to change the user interface design when needed
- Encouraging development and use of reusable software components
- Supporting ease of learning and use of applications

Meyers further categorizes user-interface tools into user-interface *toolkits* and user-interface *development systems* (UIDS) [3, p. 16]. A user interface toolkit is a collection of templates for common interaction idioms such as windows, mice, or sliders. Application source code calls high level routines in toolkit libraries to use common services. The most well know toolkit is the X Windows system [13]. Meyers further defines a UIDS as "an integrated set of tools that help programmers create and manage many aspects of interfaces" [3, p. 16]. A UIDS provides both development and run-time support for user interfaces.

I.8 Independence

An important idea underlying user interface toolkits is that of independence. The X window system, for example, provides device independence while supporting sophisticated windowing and mouse input environments. Perlman explains [10, p. 823]:

> The fundamental concept behind the creation of useful tools is *independence*; the ability to independently insert new users, devices, and applications into information processing tasks without requiring major efforts by the application developer.

Hartson and Hix also point out that a primary motivation for UIMSs is independence. Hartson and Hix stress the importance of dialogue independence — using the same language to communicate with diverse applications or objects. They explain: "in UIMS, the concept of dialogue independence is explicitly recognized and supported. Most UIMS are based, at least to some extent, on dialogue independence" [14, p. 17].

I.9 Disentangling Application and Interface

The notion of independence has been taken further as the basis of disentangling applications and user interfaces [15]. Figure 1 shows an application in which the user interface is entangled in the application. Figure 2 shows application and user interface functionality clearly separated. This separation has some cost, however. Common, independent interfaces sometimes are the lowest common denominator [10] and may have performance problems [12, p. 36].

II. Direct Manipulation Interfaces, Specification, and Verification

Mouse-window user interfaces are examples of direct manipulation user interfaces. Jacob explains differences between direct manipulation and command interfaces [2, p. 283]:

> With a direct manipulation interface, the user seems to operate directly *on* the objects in the computer instead of carrying on a dialogue *about* them. Instead of using a command language to describe operations on objects that are frequently invisible, the user "manipulates" objects visible on a graphic display

II.1 The Challenge of Direct Manipulation Interfaces

Direct manipulation interfaces have been a challenge for traditional software engineering methods of specification and design. Fischer asserts that "static specification languages have little use in HCI software design," [9, p. 50] and DeMillo et al. also doubt that formal software engineering techniques can be applied to user interfaces [16, p. 277]:

> It has been estimated that more than half the code in any real production system consists of user interfaces and error messages - ad hoc, informal structures that are by definition unverifiable.

II.2 Overviews of Interface Development Techniques

[10] and [3] overview software tools for user interfaces, [4] describes formal methods for interface development, and [14] reviews the management of user interface development.

III. NASA Applications

The motivation of this research is the development of a new *Core* of test, control, and monitor software for the space station and shuttle. This system consists of three parts: a generic core, new checkout, control, and monitor software (CCMS-II) for shuttle, and test, control, and monitor software (TCMS) for the space station.

The hardware architecture of the project is highly distributed. Multiple processors are dedicated to data acquisition, application processing, data bases, data archival and retrieval, and display. These processors communicate through high speed networks.

Two things are significant in this effort. First, each user will have a powerful display processor (DP) as their interface to the system. A DP is capable of multiple sessions with remote applications and supports a mouse-and-window (direct manipulation) environment. Second, the system is required to use commercial off-the-shelf (COTS) hardware and software where feasible. Independence from particular hardware and software is emphasized throughout the Core system.

Casual use of COTS can result in a jumble of interfaces. To prevent this problem, the software must meet the Space Station Information Systems (SSIS) User Support Environment Interface Definition (USEID). USEID defines a common user interface (CUI), user interface development tools, and CUI standards.

The CUI is the most relevant to this research. CUI consists of direct manipulation interface (DMI) services, user interface language (UIL) services, and an intermediate language (IL) to describe interapplication communication.

Schoaff describes four scenarios for use of TCMS: monitoring applications (such as programs to sample pressure), controlling applications (such as opening and closing valves), testing applications (creating combinations of monitoring and controlling applications), and operating the interface (allowing users to customize their interface) [17].

III.1 Some NASA Related Goals

The preceding section produces motivates several goals. The remainder of this report should be read with the following in mind.

- Can present UIMS's support a direct manipulation, highly distributed system such as Core?
- What formalisms are available to support the specification of Core user interfaces? In particular, can a formalism for the use of COTS software be developed as a means of maintaining a common user interface?
- Given a formal specification of portions of a user interface, can the specification and implementation by verified with respect to some correctness criteria?

IV. A Discussion of Two Relevant Research Results

Two systems in particular will be examined: the Rooms UIMS [1], and a UIMS built around a formal specification system for direct manipulation user interfaces [2]. Rooms is an example of a system with roots in cognitive engineering, while the foundation of Jacob's work is traditional software engineering. Both systems are of the new generation of interface tools and are heavily influenced by cognitive engineering.

IV.1 The Rooms Window Management System

Computer scientists know that running programs access memory nonuniformly. That is, a plot of memory accesses shows distinct clumping of references. This characteristic is exploited in virtual memory operating systems which allow programs to run when physical memory is much smaller than the program's range of memory accesses. Virtual memory operating systems shuffle programs' code and data between a small physical memory and a larger secondary store giving each program the illusion of a large physical memory. Operating systems take advantage of programs' locality of memory references to perform this shuffling between physical and secondary memory in ways that tend to maximize performance.

Optimal memory shuffling is in general not attainable, and in the worst case a running program will experience thrashing - repeatedly referring to memory locations that are not in physical memory and must be shuffled in from secondary storage.

Henderson and Card applied the concepts of locality of reference and thrashing to direct manipulation interfaces. Their Rooms UIMS was designed as a solution to the problem of *window thrashing*. Window thrashing occurs because the small size of computer displays. Although mouse-and-window interfaces are often described as a metaphor for a physical desktop, "a standard office desk has the area of 22 IBM PC screens, 46 Macintosh screens, or even 10 of the "large" 19-inch Xerox 1186 or Sun-3 screens" [1, p. 212].

Four techniques of dealing with the small size of computer screens are: 1) alternating screen usage, 2) distorted views, 3) large virtual workspaces, and 4) multiple virtual workspaces. The present CCMS system use the first technique, allowing users to flip among screens. The second technique is commonly used - shrinking an entire window to an icon is an example of a distorted view. The second technique can be taken to an extreme by using fish-eye techniques in which objects are distorted based on some priority (e.g. most recently used or most important). The third technique uses windows as viewports into a large virtual world. Henderson and Card cite the NASA Ames virtual reality helmet as an example of the third technique taken to an extreme. Rooms is based on the fourth technique - a hypertext-like organization of users' worlds. A problem with hypertext applications is navigation. The authors state that [1, p. 238]

> Our experience suggests that navigation tends to be easier in a multiple-

> virtual-workspace system than in either a large single workspace or a hypertext system.

Henderson and Card use locality of window reference as the basis of Rooms. The authors postulate that just as programs in execution exhibit locality of reference, use of windows also occurs in related chunks. In the same way that locality of reference makes virtual memory feasible in that only a small amount of physical memory is necessary to run a program in what it thinks is a large virtual address space, locality of window reference should make multiple virtual workspaces feasible.

Henderson and Card elaborate [1, p. 221]:

> The design of the Rooms system is based on the notion that, by giving the user the mechanism for letting the system know he or she is switching tasks, it can anticipate the set of tools/windows the user will reference and thus preload them together in a tiny fraction of the time the user would have required to open, close, and move windows or expand and shrink icons. A further benefit is that the set of windows preloaded on the screen will cue the user and help reestablish the mental context for the task.

Thus, task-specific sets of windows aid in navigation.

IV.2 Specification of Direct Manipulation Interfaces

A formal specification of a system is a state of the functionality that a system is to perform. Formal specifications have been used to specify the behavior of complex system such as secure operating systems and terminals.

A formal specification can be used in several ways. First, if the specification is highly abstract, it is possible to reason about essential qualities of the system before it is implemented. The essence of functionality can be explored without superfluous details. Second, abstract specifications can be refined to more detailed specifications. The transformations, or mappings, between an abstract specification and a refinement are also formally defined and can be reasoned about. Third, if correctness constraints can be formally expressed, it is possible to verify that the specification is correct with respect to the correctness criteria.

It is possible to formally specify, possibly through multiple levels of abstraction, a system down to pre and post assertions corresponding to entry and exit assertions for code. If the specification is proved correct with respect to the correctness criteria, we are assured that the a program implementing the specifications will meet the correctness criteria.

Figure 3 shows the specification and verification process.

Although formal specification and verification is theoretically possible, its use has been limited to critical applications such as secure message systems. A typical correctness criteria for such systems is that a user cannot read an object having security level higher than theirs, and cannot write to an object that has security level lower than theirs. These properties are called the simple security property, and the star-property. In the ASLAN formal specification language [18, 19, 20] they would be expressed as

```
INVARIANT

FORALL u: user, o: object (
      (reading(u, o) -> level(u) >= level(o))
    & (writing(u, o) -> level(u) <= level(o)))
```

where -> is logical implication.

The disadvantages of formal specification and verification are many and well known. DeMillo et al. argue that a social process much like that for mathematical theorems must be applicable [16]. That is, much of our faith in mathematical theorems comes from the wide dispersion and perusal of conjectures and proofs. The authors argue that verifications are tedious and dull, lacking insight and that no one would read through an entire specification or verification proof. More recently, Fetzer has reexamined the DeMillo arguments and refined them further [21], concluding that DeMillo et al. "arrived at the right general conclusion for the wrong specific reasons" [21, p. 1062].

IV.2.1 Objections to Formal Specification and Verification

The major objections to formal specification and verification are summarized as follows:

1. Interesting constructs of modern programming languages are not formally defined, and are therefore not verifiable.
2. The volume of information that formal specification and verification develop make it infeasible to prove all but the smallest programs.
3. Specifications are not interesting reading and do not embody insights that would draw human reviews to peruse them.
4. To be useful, a specification must correspond to implementation code: high level specifications are worthless.

These arguments are too pessimistic and will be examined in turn. Our responses will be with respect to the user interface software. Our general assertion is that formal specification, when applied to isolated pieces of a system, with specific correctness constraints in mind, can be productively applied.

1. Hoare [22] has specified the semantics of Pascal, and Mallgren [23] has developed techniques for specifying graphics programming languages.
2. Although to specify and verify even a small program takes volumes of information, care must be taken to specify only well defined pieces. Also, the specification process does not have to be taken to code to be worthwhile.
3. High level specifications embody the essential functionality and criticality of a software system. The abstract specifications are often quite insightful, and not difficult to read. Also, Just as programmers usually do not read the assembly languages produced by compilers of higher level languages, specifiers don't have to reading all the output of specification and verification systems.
4. High level specifications are useful in isolating critical aspects, something physical that can be discussed in groups (the social aspect that DeMillo was looking for), and can serve as the foundation for user documentation and for traceablity of requirements in design.

IV.2.2 Reasonable Expectations for Formal Specification

Our view is that formal specifications can be productively applied to software development processes. In addition, formal specification is worthwhile even if formal verification is not the ultimate goal. For a formal specification effort to be worthwhile there must be 1) well defined areas of functionality to be specified, 2) well defined constraints on system behavior that can be used as high level correctness criteria, 3) abstraction techniques that can refined through multiple levels of abstraction, and 4) personnel able to read and evaluate specifications.

The general goal is to lower expectations for formal specification - the goal isn't necessarily provably correct software - but to specify important functionality and correctness criteria in a way that is reviewable by software engineers and integrates usefully in other software development efforts.

IV.2.3 Jacob's Formal Specification System for Direct Manipulation Interfaces

The creed of formal specification is "describe what the system does, not how". When specifying user interfaces it is easy to be trapped into specifying the entire system behavior - after all, what a user sees is the essential functionality of the system. As noted, for a specification to be successful it is essential to break up the system into interacting parts.

Thus, user interface specifiers should limit their work to defining only users methods of interaction. An application's functionality can be specified, and verified if necessary, separately.

That is, given separately specified user interface and applications, and a definition of their interaction, one can inductively argue that the system as a whole is specified.

Until recently, user interface specification languages, and their associated UIMSs have addressed the problem of dialogue specification, design, and support. These techniques have commonly been based on state machine models or formal grammars.

Although easier to use, direct manipulation user interfaces have proven to be much harder to specify than keyboard-and-screen based dialogues.

Jacob [2] has developed a collection of specification methods and a language for direct manipulation user interfaces. Jacob's specification language is object oriented and executable, allowing prototype interfaces to be rapidly developed.

Jacob uses a three level specification technique: the lexical level, syntax level, and semantic level. Tokens, such as button clicks, highlighting, and dragging, are defined at the lexical level. Sequences of tokens representing plausible user input make up the syntactic level. Finally, the functional requirements of the system are embodied in the semantic level. The semantics of an application's functionality can be specified separately using conventional specification techniques. Jacob's techniques, therefore, emphasize the lexical and syntactic levels and the way functions are invoked.

Jacob defines an interaction object as "the smallest unit in the user interface that has a state" [2, p. 290]. He then gives three general steps to specifying a direct manipulation user interface [2, p, 290]:

1. define a collection of interaction objects,
2. specify their internal behaviors, and
3. provide a mechanism for combining them into a coordinated user interface

A key point of Jacob's work is that user interfaces can be described as interacting state machines. However, direct manipulation interfaces have often been described as stateless, or modeless. Jacob argues that in direct manipulation user interfaces are "highly moded, but they are much easier to use than traditional moded interfaces because of the direct way in which the modes are displayed and manipulated" [2, p. 288]. By separating the specification into three levels and allowing objects to essentially inherit state machines from other objects, Jacob has brought the power and simplicity of state machines to the specification of direct manipulation user interfaces.

In particular, Jacob proposes that an interface be modeled as a high level executive process that invokes coroutines corresponding to individual dialogues. Such a user interface is highly moded (input has different meaning depending on its context) and can be modeled by state machines that interact in well-defined ways. The user interface as a whole is viewed as a simple executive and multiple independent dialogues.

V. Correctness Constraints

Formal specifications are particularly valuable when they are analyzed with respect to formally state correctness criteria. Correctness constraints can be categorized as being properties that must hold in every state (safety properties), or properties that must eventually be true (liveness properties).

Formal specification and verification is particularly well suited to ensuring that high-level safety properties are maintained. Liveness techniques are much harder to specify and prove.

A space station user interface requirements document [24] contains many safety properties. These properties could be used as correctness constraints for formal specifications. The following are example correctness constraints:

- 3.3.1.4 Graphics Panels shall be no larger than the maximum size of the Workspace Panel and no smaller than TBD. [24, p. 9]
- 3.3.2.1 Selection of a Radio Button shall invoke the highlight of the selected button and the removal of the Radio Button highlight from the previously selected Radio Button. [24, p. 10]
- 3.3.2.2.1 If the selected Momentary Button is not released and the cursor is moved outside the boundaries of the button, the highlight shall be removed from the button and the switch action canceled. [24, p. 10]
- 3.4.2.1 The Alarm Classification area shall contain an alphabetic identifier that signifies to the operator the classification of the alarm ... when no alarm is present, this area shall be left blank. [24, p. 16]

VI. Concluding Remarks

This report argues that the development of NASA human-computer interfaces can benefit from recent cognitive engineering research. It has also shown that formal specification techniques can be used to specify direct manipulation user interfaces.

In particular, a study on window use and management that combined results from cognitive psychology and operating systems theory [1] is applicable to designing interfaces for ground and flight software. These interfaces can automatically present users with familiar environments for specific tasks.

Second, a formal specification system for direct manipulation user interfaces [2] was described. The system is object oriented and breaks down the task of specification into three levels: lexical, syntactic, and semantic.

Finally, appropriate use and reasonable expectations of formal specification of user interfaces is discussed. The use of formally expressed correctness constraints is examined, and examples of such constraints from a NASA user interface document are given.

VI.1 Recommendations

The following are recommendations for development of NASA user interface software. These recommendations are realistic and can be used in addition to other user interface development techniques.

- Build data collection capabilities into user interface software. These capabilities should not be used to track specific individuals' work, but should be used to identify areas in which mistakes are frequently made, or to find task specific localities of interface use. The Rooms user interface system was a direct result of being able to monitor users' needs for interface resources.

- Use rapid prototypes of interfaces for early usability tests. Early useability tests frequently bring out interaction areas which are difficult to users but simple to the interface developers. Jacob's system for specification of direct manipulation user interfaces is not only a formal specification system, but also a rapid prototyping tool. Using executable formal specification languages has the advantage of both formal specification and rapid prototyping.

- Identify high level correctness constraints for interfaces. Identification of correctness criteria is useful even without formal verification. Although NASA user interface documents contain many examples of low level correctness criteria, quantitative high-level correctness criteria are rare.

- Consider a pilot study using formal techniques on a small, well-defined piece of a user interface. A good example to formally develop would be an alarm area of a screen. The function of alarm areas is typically straightforward, and high level correctness criteria are apparent.

VI.2 Goals Revisited

Finally, the three goals identified in section III.1 can be addressed. The first goal asked whether present UIMSs can support direct manipulation interfaces that NASA requires. At this point the answer is no, although there are excellent interface development tools that can be applied to pieces of interface development efforts.

The second goal was to seek formalisms to support the specification of user interfaces. Again, the tools are not mature, but are potentially useful for well defined pieces of interface development. The second part of the second goal concerned the use for formalisms for use of COTS software. This remains an open research question.

The third goal concerned the formal specification of interfaces and correctness criteria. We believe that with reasonable expectations, formal specification can be productively applied to the development of user interfaces.

References

[1] D. A. Henderson, Jr. and S. K. Card. Rooms: the use of multiple virtual workspaces to reduce space contention in a window-based graphical user interface. *ACM Transactions on Graphics* 5, 3 (July 1986).

[2] R. J. K. Jacob. A specification language for direct-manipulation user interfaces. *ACM Transactions on Graphics* 5, 4 (October 1986).

[3] B. A. Meyers. User-interface tools: introduction and survey. *IEEE Software* 6, 1 (January 1989).

[4] M. U. Farooq and W. D. Dominick. A survey of formal tools and models for developing user interfaces. *International Journal of Man-Machine Studies* 29 (1988).

[5] B. Auernheimer and J. L. Dyck. Human-computer interaction: where computer science and psychology meet. *Proceedings of the Twenty-second Annual Hawaii International Conference on System Sciences,* IEEE Computer Society Press, 1989.

[6] D. A. Norman. Cognitive engineering - cognitive science. In *Interfacing Thought* J. M. Carroll, ed. The MIT Press, 1987.

[7] T. K. Landauer. Research methods in human-computer interaction. In *Handbook of Human-Computer Interaction, M. Helander, ed. North-Holland, 1988.*

[8] S. Doane, W. Kintsch, and P. Polson. Action planning: producing UNIX commands. *Proceedings of the Eleventh Annual Meeting of the Cognitive Science Society.* Ann Arbor, 1989.

[9] G. Fischer. Human-computer interaction software: lessons learned, challenges ahead. *IEEE Software* 6, 1 (January 1989).

[10] G. Perlman. Software tools for user interface development. In *Handbook of Human-Computer Interaction, M. Helander, ed. North-Holland, 1988.*

[11] B. A. Meyers. *Creating User Interfaces by Demonstration.* Academic Press, 1988.

[12] J. Löwgren. History, state and future of user interface management systems. *SIGCHI Bulletin* 20, 1 (July 1988).

[13] R. W. Scheifler and J. Gettys. The X window system. *ACM Transactions on Graphics* 5, 2, (April 1986).

[14] H. R. Hartson and D. Hix. Human-computer interface development: concepts and systems for its management. *ACM Computing Surveys* 21, 1 (March 1989).

[15] M. J. Muller. Disentangling application and presentation in distributed computing: architecture and protocol to enable a flexible, common user interface. *IEEE Transactions on Systems, Man, and Cybernetics.* 18, 4 (July/August 1988).

[16] R. A. DeMillo, R. J. Lipton, and A. J. Perlis. Social processes and proofs of theorems and programs. *Communications of the ACM* 22, 5 (May 1979).

[17] W. D. Shoaff. The design of an intelligent human-computer interface for the test, control and monitor system. Technical report, grant NASA-NGT-60002, August 1988.

[18] B. Auernheimer and R. A. Kemmerer. ASLAN users manual. Technical report TRCS84-10. Department of Computer Science, University of California, Santa Barbara (March 1985).

[19] B. Auernheimer and R. A. Kemmerer. RT-ASLAN: a specification language for real-time systems. *IEEE Transactions on Software Engineering,* SE-12, 9 (September 1986).

[20] B. Auernheimer and R. A. Kemmerer. Procedural and nonprocedural semantics of the ASLAN formal specification language. *Proceedings of the nineteenth annual Hawaii international conference on system sciences* (January 1986).

[21] J. H. Fetzer. Program verification: the very idea. *Communications of the ACM* 31, 9 (September 1988).

[22] C. A. R. Hoare and N. Wirth. An axiomatic definition of the programming language Pascal. *Acta Informatica* 2 (1973), p. 335-355.

[23] W. R. Mallgren. *Formal Specification of Interactive Graphics Programming Languages.* The MIT Press, 1983.

[24] A. D. Cohen. User interface requirements document (DR SY 45.1). Work Package No. 2. McDonnell Douglas Space Systems Company (July 1989).

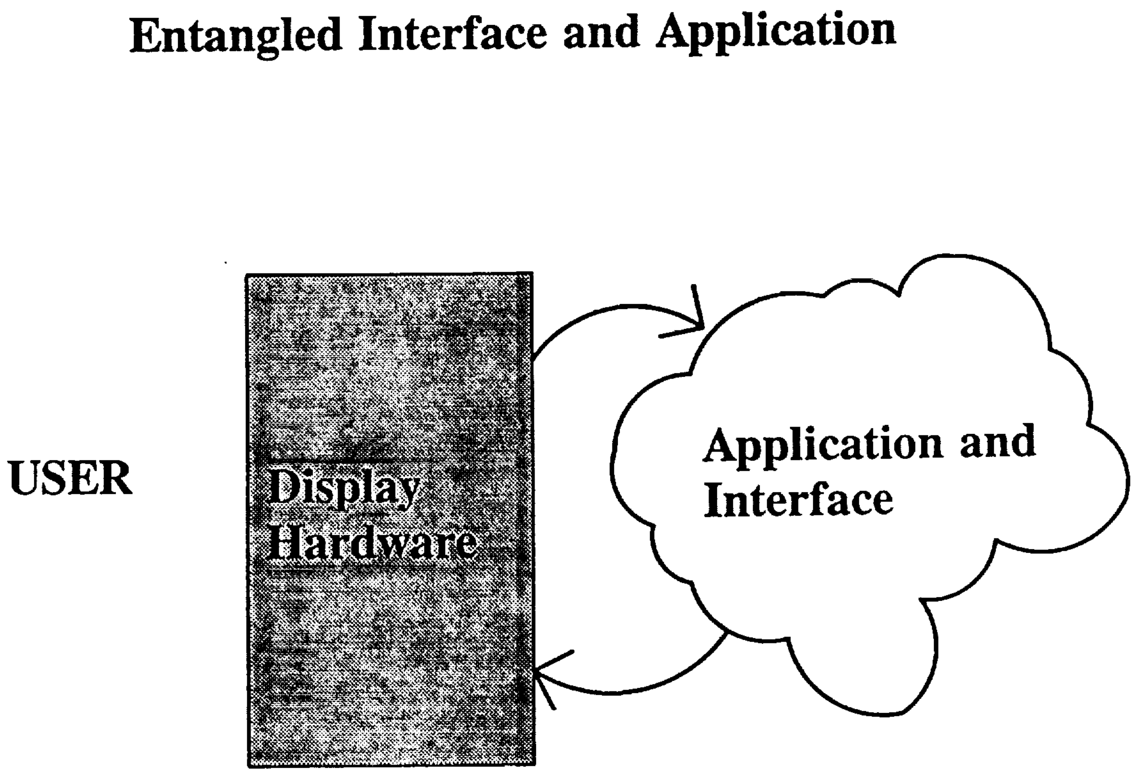


Figure 1. Entangled User Interface and Application

Disentangled Interface and Application

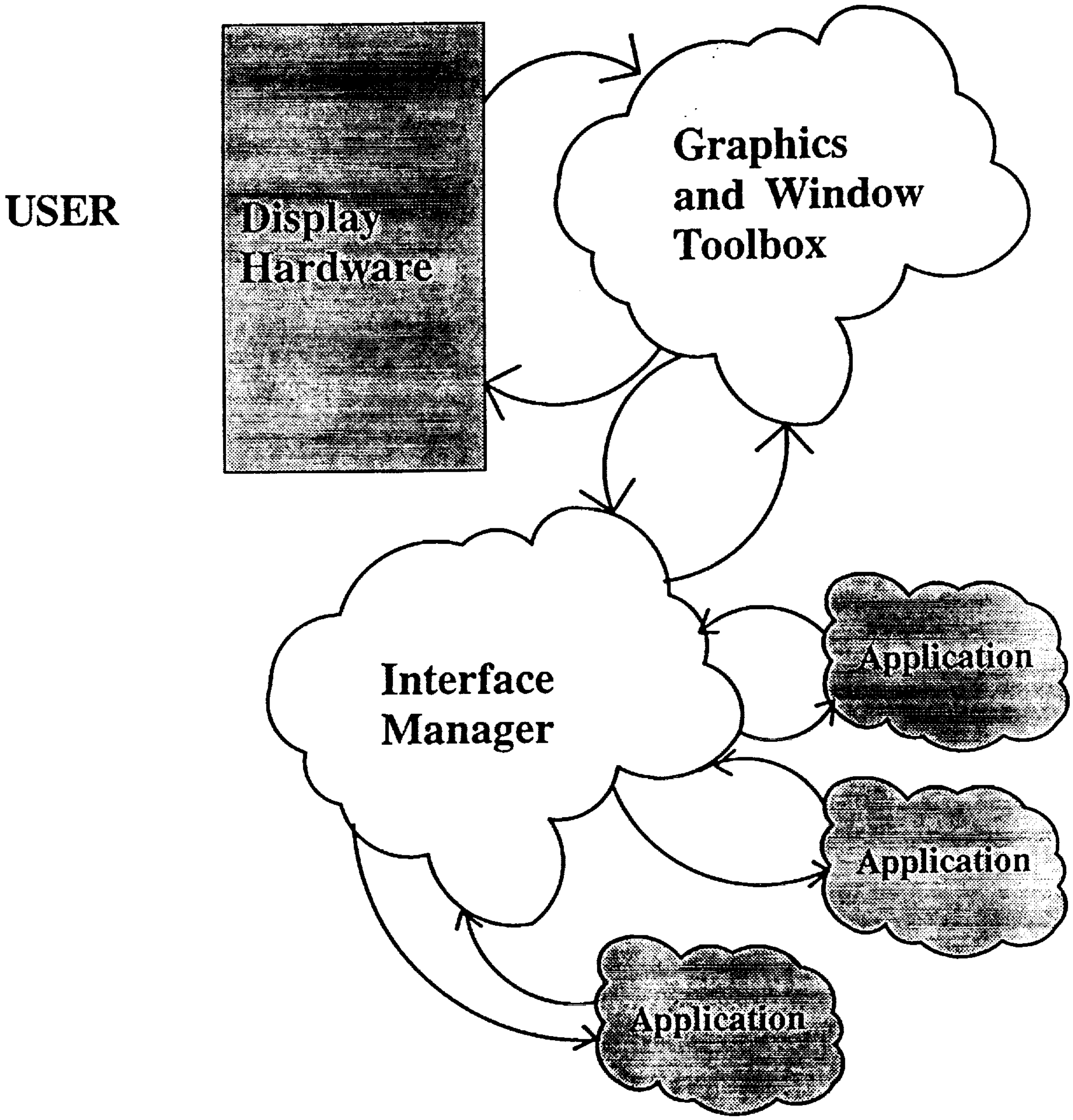


Figure 2. Disentangled User Interface and Applications

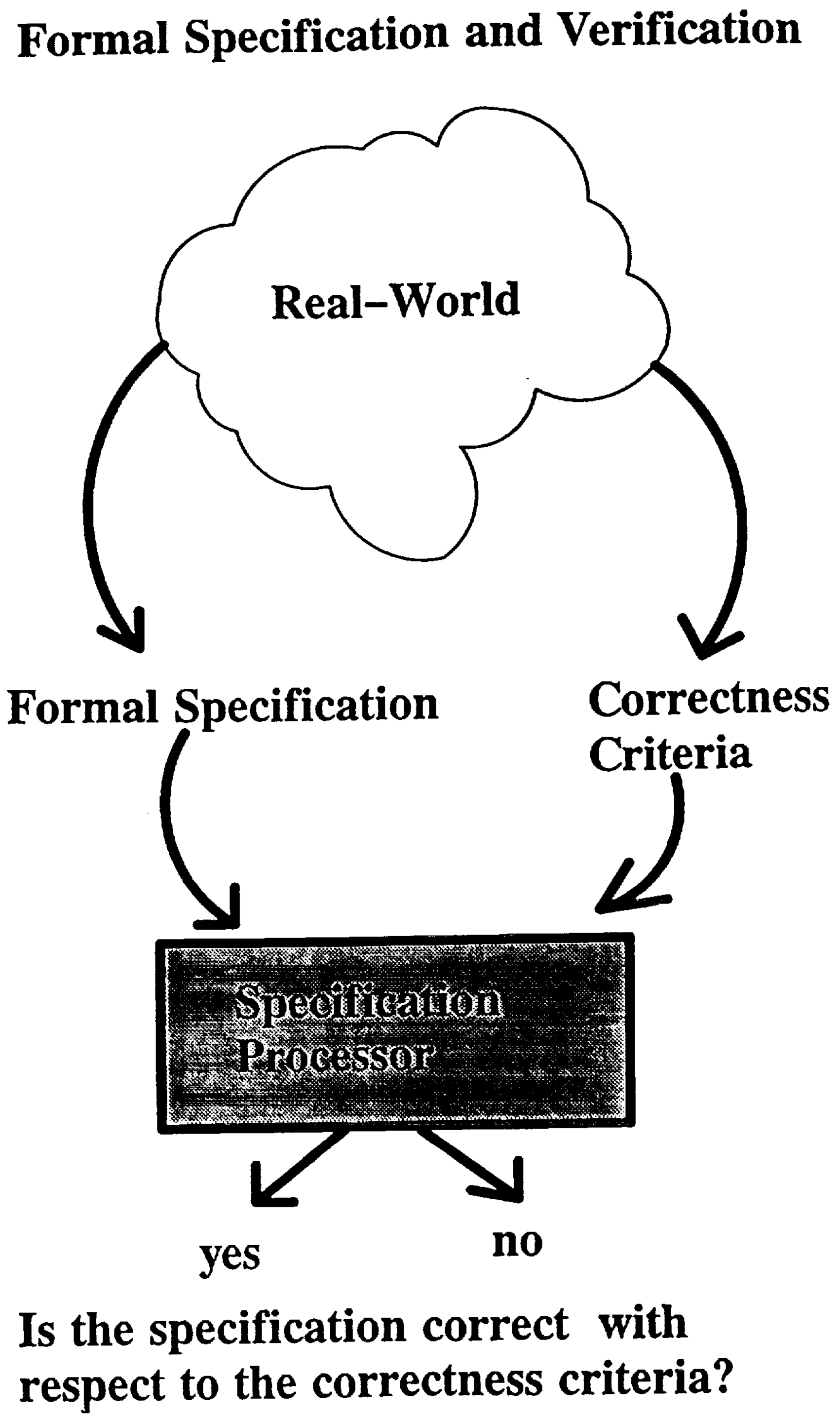


Figure 3. The Formal Specification and Verification Process

N90-16688

1989 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

STUDY OF METAL CORROSION USING AC IMPEDANCE TECHNIQUES IN THE STS LAUNCH ENVIRONMENT

PREPARED BY:	Dr. Luz M. Calle
ACADEMIC RANK:	Associate Professor
UNIVERSITY AND DEPARTMENT:	Randolph-Macon Woman's College Chemistry Department
NASA/KSC	
DIVISION:	Materials Science Laboratory
BRANCH:	Materials Testing Branch
NASA COLLEAGUE:	Mr. Louis G. MacDowell III
DATE:	August 11, 1989
CONTRACT NUMBER:	University of Central Florida NASA-NGT-60002 Supplement: 2

ACKNOWLEDGEMENT

I would like to thank my NASA colleague, Louis G. MacDowell, III, for providing me with the opportunity and orientation to pursue this research. I would also like to thank the people in the Materials Science Laboratory who made me feel welcome at the Kennedy Space Center. I am grateful to NASA/ASEE for my selection as a faculty fellow in the 1989 NASA/ASEE program at the Kennedy Space Center. The expertise and kindness of the program director, Dr. E. Ramon Hosler, is also gladly acknowledged.

ABSTRACT

AC impedance measurements were performed to investigate the corrosion resistance of 19 alloys under conditions similar to the STS launch environment. The alloys were: Zirconium 702, Hastelloy C-22, Inconel 625, Hastelloy C-276, Hastelloy C-4, Inconel 600, 7Mo + N, Ferralium 255, Inco Alloy G-3, 20Cb-3, SS 904L, Inconel 825, SS 304LN, SS 316L, SS 317L, ES 2205, SS 304L, Hastelloy B-2, and Monel 400. AC impedance data were gathered for each alloy after one hour immersion time in each of the following three electrolyte solutions: 3.55% NaCl, 3.55% NaCl-0.1N HCl, and 3.55% NaCl-1.0N HCl. The data were analyzed qualitatively using the Nyquist plot and quantitatively using the Bode plot. Polarization resistance, Rp, values were obtained using the Bode plot. Zirconium 702 was the most corrosion resistant alloy in the three electrolytes. The ordering of the other alloys according to their resistance to corrosion varied as the concentration of hydrochloric acid in the electrolyte increased. The corrosion resistance of Zirconium 702 and Ferralium 255 increased as the concentration of hydrochloric acid in the electrolyte increased. The corrosion resistance of the other 17 alloys decreased as the concentration of the hydrochloric acid in the electrolyte increased.

SUMMARY

AC impedance techniques were used to study the corrosion of 19 alloys under conditions similar to the STS launch environment which is highly corrosive. The 19 alloys were: Zirconium 702, Hastelloy C-22, Inconel 625, Hastelloy C-276, Hastelloy C-4, Inconel 600, 7Mo + N, Ferralium 255, Inco Alloy G-3, 20Cb-3, SS 904L, Inconel 825, SS 304LN, SS 316L, SS 317L, ES 2205, SS 304L, Hastelloy B-2, and Monel 400. AC impedance data were acquired for each of the alloys after one hour immersion in each of the following three electrolytes: 3.55% NaCl, 3.55% NaCl-0.1N HCl, and 3.55% NaCl-1.0N HCl.

The data were analyzed qualitatively by using the Nyquist plot and quantitatively by using the Bode plot. Rp values were obtained from the Bode plot. The corrosion resistance for 17 of the 19 alloys decreased as the concentration of hydrochloric acid in the electrolyte increased. The corrosion resistance of Zirconium 702 and Ferralium 255 increased as the concentration of hydrochloric acid in the electrolyte increased. The most corrosion resistant alloy in the three electrolytes was Zirconium 702. The ordering of the alloys in terms of their resistance to corrosion was different in the three electrolytes.

A good correlation was found between the ac impedance data and the dc polarization data even though the Rp values were different. It is postulated that the Rp values obtained by using ac impedance techniques are more accurate. A comparison with the beach corrosion data led to the conclusion that there is, in general, a good correlation between the materials that performed well in both tests as well as between those that performed poorly. However, the ordering of the materials according to their resistance to corrosion was different in both tests. It can be concluded that ac impedance techniques can be used to choose what materials should be subjected to long-term corrosion testing.

TABLE OF CONTENTS

LIST OF FIGURES

STUDY OF METAL CORROSION USING AC IMPEDANCE TECHNIQUES IN THE STS LAUNCH ENVIRONMENT

I. INTRODUCTION

Flexible metal hoses are used in various supply lines that service the Orbiter at the launch pad. These convoluted flexible hoses were originally constructed of 304L stainless steel. The severely corrosive environment at the launch site caused pitting corrosion in many of these flex hose lines. In the case of vacuum jacketed cryogenic lines, failure of the flex hose by pitting causes a loss of vacuum and subsequent loss of insulation.

The corrosive environment at the launch site is due to the very high chloride content caused by the proximity of the ocean and to the generation of seventeen tons of concentrated hydrochloric acid as a fuel combustion product of the Solid Rocket Boosters during a launch. These corrosive conditions cause severe pitting on some of the commonly used steel alloys.

A previous investigation was undertaken in order to evaluate 19 metal alloys with the purpose of finding a more corrosion resistant replacement material for 304L stainless steel. The tests performed in that investigation were: electrochemical corrosion testing, accelerated corrosion testing in a salt fog chamber, long term exposure at the beach corrosion testing site, and pitting corrosion tests in ferric chloride solution. These tests led to the conclusion that the most corrosion resistant alloys were, in descending order, Hastelloy C-22, Inconel 625, Hastelloy C-276, Hastelloy C-4, and Inco Alloy G-3. Of these top five alloys, the Hastelloy C-22 stood out as being the best of the alloys tested. The details of this investigation are found in report MTB-325-87A (1). Furthermore, on the basis of corrosion resistance combined with weld and mechanical properties, Hastelloy C-22 was determined to be the best material for the construction of flex hoses to be used in fuel lines servicing the Orbiter at the launch site.

The electrochemical corrosion testing done previously was based on the use of dc polarization techniques. In the present investigation, ac impedance techniques will be used in order to study the corrosion of the 19 alloys under three different electrolyte conditions: neutral 3.55% NaCl, 3.55% NaCl-0.1N HCl, and 3.55% NaCl-1.0N HCl. The 3.55% NaCl-0.1N HCl electrolyte provides an environment for the corrosion of the alloys similar to the conditions at the launch pad.

II. MATERIALS AND EQUIPMENT

2.1 CANDIDATE ALLOYS

The nineteen alloys tested and their nominal compositions in weight percent are shown in Table 1. The choice of these alloys for the previous investigation was based on their reported resistance to corrosion.

2.2 AC IMPEDANCE MEASUREMENTS

A model 378 Electrochemical Impedance system manufactured by EG&G Princeton Applied Research Corporation was used for all electrochemical impedance measurements. The system includes: (1) the Model 273 Computer-Controlled Potentiostat/ Galvanostat, (2) the Model 5301A Computer-Controlled Lock-In Amplifier, (3) the IBM XT Microcomputer with peripherals, and the Model 378 Electrochemical Impedance Software.

Specimens were flat coupons 1.59 cm (5/8") in diameter. The specimen holder in the electrochemical cell is designed such that the exposed metal surface area is 1 cm^2.

The electrochemical cell included a saturated calomel reference electrode (SCE), 2 graphite rod counter electrodes, the metal working electrode, and a bubbler/vent tube. Each alloy was studied under three different electrolyte conditions: aerated 3.55% neutral NaCl, aerated 3.55% NaCl-0.1N HCl (similar to the conditions at the launch site), and aerated 3.55% NaCl-1.0N HCl (more aggressive than the conditions at the launch site). All solutions were prepared using deionized water.

III. PROCEDURE FOR AC IMPEDANCE MEASUREMENTS

The test specimens were polished with 600-grit paper, wiped with methyl-ethyl ketone, ultrasonically degreased for five minutes in a detergent solution, rinsed with deionized water, and dried. Each specimen was observed under the microscope and weighed before and after each experiment to monitor changes caused by corrosion on its appearance and weight.

The electrolyte solution was aerated for at least 15 minutes before immersion of the test specimen. Aeration continued throughout the test.

AC impedance measurements were performed under each of the three electrolyte conditions chosen. After immersion in the electrolyte, the sample was allowed to equilibrate for 3600 seconds before the instrument started acquiring data. It was determined previously that after 3600 seconds, the corrosion potential had usually stabilized (2).

AC impedance measurements were gathered in the frequency range from 100 kHz to 0.1001 Hz. A combination of two methods was employed to obtain the data over this wide range of frequencies: (1) phase-sensitive lock-in detection for measurements from 5 Hz to 100 kHz, and (2) the FFT (fast Fourier transform) technique for measurements from 0.1001 Hz to 11 Hz. The data from lock-in (single-sine) and FFT (multi-sine) were automatically merged by the IBM XT microcomputer dedicated software.

The conditions for the lock-in experiments were: initial frequency, 100 kHz; final frequency, 5 Hz; points/decade, 5; AC amplitude, 5 mV; DC potential, 0 vs OC (open circuit); condition time, 0 seconds; condition potential, 0 V; open circuit delay, 3600 seconds. The open circuit potential was monitored with a voltmeter.

The conditions for the FFT experiments were: base frequency, 0.1001 Hz; data cycles, 5; AC amplitude, 10 mV; DC potential, 0 vs OC; open circuit delay, 0 seconds. The open circuit potential was monitored with a voltmeter.

The data for each experiment were plotted in the Nyquist and Bode plot format.

IV. RESULTS AND DISCUSSION

4.1 THEORETICAL BACKGROUND

AC impedance techniques offer some distinct advantages over dc techniques (3). First, the small excitation amplitudes that are used, generally in the ranges of 5 to 10 mV peak-to-peak, cause only minimal perturbations of the electrochemical system, thus reducing errors caused by the measuring technique itself. Second, the technique offers valuable information about the mechanisms and kinetics of electrochemical processes such as corrosion. Third, measurements can be made in low conductivity solutions where dc techniques are subject to serious potential-control errors.

Despite the advantages of the ac impedance techniques mentioned above, their application requires sophisticated techniques in order to interpret the data and extract meaningful results. The application of ac impedance measurements to study corrosion has so far resulted in the publication of a large amount of experimental data without much interpretation. The technique is at the present time in a transition from the data collection stage to the data analysis stage (4).

AC impedance measurements are based on the fact that an electrochemical system, such as those studied in this investigation, can be represented by an equivalent electrical circuit. The equivalent circuit for a simple electrochemical cell is shown in Figure 1 (5). The circuit elements R_Ω, Rp, and Cdl represent the uncompensated resistance (resistance from the reference to the working electrode), the polarization resistance (resistance to electrochemical oxidation), and the capacitance very close to the metal surface (at the double layer). There are several formats that can be used for the graphical representation of the ac impedance data (3,6,7). Each format offers specific advantages for revealing certain characteristics of a given test system. It was determined at the beginning of this research, that the most suitable formats for plotting the ac impedance data were the Nyquist and the Bode plots.

The Nyquist plot is also known as a Cole-Cole plot or a complex impedance plane diagram. Figure 2 (5) shows the Nyquist plot for the equivalent circuit shown in Figure 1. The imaginary component of the impedance (Z") is plotted versus the real component of the impedance (Z') for each excitation frequency. As indicated in Figure 2, this plot can be used to calculate the values of R_Ω, Rp, and Cdl.

The Bode plot for the equivalent circuit in Figure 1 is shown in Figure 3 (5). This graphical representation of the ac impedance data involves plotting both the phase angle (Θ) and the log of absolute impedance (log|Z|) versus the log of the frequency ($w = 2\pi f$). As indicated on the figure, values for R_{Ω}, Rp, and Cdl can also be obtained from the Bode plot. Of special interest for this research is the determination of the Rp values which can be used to calculate the corrosion rate of an electrode material in a given electrolyte (3,8).

4.2 RESULTS AND DISCUSSION

The Bode plots included in this report appear in the form of two separate graphs: log|Z| versus log Frequency (Hz) and θ versus log Frequency (Hz). Nyquist (at the top) and Bode (at the bottom) plots for the 19 alloys used in this investigation are shown in Figures 4-13. None of the Nyquist plots obtained in this investigation exhibited the ideal semicircle shown in Figure 2. Experimentally, it has been observed that deviations from the results expected for simple equivalent circuits occur for real, corroding systems (6,9). Some of the deviations that have been observed for real systems are: a semicircle with its center depressed below the real axis, a partial semicircle, and a partial semicircle that changes shape at the low frequency end. Impedance data that result in a Nyquist plot in the form of a depressed or partial semicircle can still be used to calculate Rp values. Several authors have described computer modeling of electrochemical impedance (10,11). The usual approach is to curve-fit the semicircle that results from a single time constant capacitive response. This approach allows an estimate to be made of the low frequency intersection of the semicircle response with the real axis. This procedure is especially important when the response still has a large imaginary contribution at low frequency resulting in a partial semicircle. Deviations that result in a Nyquist plot with the shape of a partial semicircle that changes at the low frequency end require a more complex computer program which contains more circuit elements. The time limitations of this research prevented the use of the methods just mentioned to analyze the Nyquist plots for the 19 alloys.

Valuable qualitative information can be extracted by comparing the Nyquist plots shown in Figures 4-13. Each Figure shows the change in the Nyquist plot for a one hour immersion time of the alloy in the three different electrolytes: (X) 3.55% NaCl, (□) 3.55% NaCl-0.1N HCl and (o) 3.55% NaCl-1.0N HCl. The change in the corrosion rate, which is inversely proportional to Rp, can be estimated qualitatively by looking at the change in the Nyquist plot. Zirconium 702 (Figure 4a) stands out as being the most corrosion resistant alloy under the conditions used in this study. Its Rp was not only the highest but it also showed the least change upon increasing the concentration of the hydrochloric acid from 0.0N to 0.1N to 1.0N; that is, Zirconium 702 became more corrosion resistant as the concentration of hydrochloric acid increased. This finding agrees with the known fact that Zirconium is resistant to hydrochloric acid at all concentrations up to boiling temperatures. However, there are indications that the metal is vulnerable to pitting in seawater (12). Ferralium 255 (Figure 4b) also became more corrosion resistant upon increasing the concentration of the acid. Its Rp values were similar in the three electrolytes but lower than those for Zirconium 702. The change in Rp for the other 17 alloys upon increasing the concentration of the acid in the electrolyte was in the opposite direction to that observed for Zirconium 702 and Ferralium 255; they became less resistant to corrosion as the concentration of the acid increased. Hastelloy C-22 (Figure 5a), Inco Alloy G-3 (Figure 5b), Hastelloy C-4 (Figure 6a), Inconel 625 (Figure 6b), Hastelloy C-276 (Figure 7a), and Hastelloy B-2 (Figure 7b) have similar Nyquist plots showing the decrease in Rp as the concentration of the acid increases. The decrease in Rp appears as an increase in the curvature of the partial semicircle. Monel 400 (Figure 8a) shows partial semicircles with a slight decrease in Rp caused by increasing the acid concentration. The semicircle obtained in 3.55% NaCl-0.1N HCl has a feature at the end (a straight line with a positive slope) that has been associated with a Warburg impedance (13). This behavior has been explained by postulating an extra impedance term in the equivalent circuit that is associated with diffusion controlled processes. 20Cb-3, ES 2205, and 7Mo + N (Figures 8b, 9a, 9b) show similar changes in the Nyquist plot. No drastic change in the Rp values is observed when the concentration of hydrochloric acid increases from 0.0N to 0.1N (as indicated by the two parallel lines in the Nyquist plot). However, there is a significant change in the Nyquist plot when the concentration of the acid is increased to 1.0N that results in a considerable decrease in Rp. The turn at the low frequency end of the curve is probably an indication of a diffusion process taking place. SS 304L (Figure 10a) shows a Nyquist plot that is different from all the others. It should be pointed out that one of the experiments involving SS 304L resulted in the partial breakdown of the surface of the metal sample. Data from that experiment were discarded. The complex Nyquist plot obtained for SS 304L in 3.55% NaCl-1.0N HCl is similar to the Nyquist plot obtained for a pin-holed coal tar epoxy coating on mild steel (13). SS 304LN (Figure 10b) showed good resistance to corrosion in neutral 3.55% NaCl but similar low resistance to corrosion (low Rp values) in 3.55% NaCl-0.1N HCl and 3.55% NaCl-1.0N HCl. Inconel 600 (Figure 11a) showed a similar behavior. Inconel 825, SS 317L, SS 904L, and SS 316L (Figures 11b, 12a, 12b, 13a) show a similar behavior indicating comparable resistance to corrosion in 3.55% NaCl and 3.55% NaCl-0.1N HCl and a considerably lower resistance to corrosion in 3.55% NaCl-1.0N HCl.

Since the Nyquist plots obtained in this investigation did not resemble the ideal Nyquist plot shown in Figure 2, calculation of Rp values from those plots was not pursued. It was decided that the Bode plot is a more straightforward means of presenting the data in order to calculate Rp. In a Bode plot, the impedance of a "perfect capacitance" can be represented as a straight line with a slope of -1 and a phase angle of -90°. A "resistor" will plot as a horizontal line for the log¦Z¦ with a phase angle of 0°. A Warburg impedance is a straight line with a slope of -1/2 and a phase angle of -45° (14). The data gathered for the 19 alloys, when plotted in the Bode format (lower plot in Figures 4-13) were interpreted as shown in Figure 3. The value of Z at the lowest frequency (0.1001 Hz) is the sum of Rp and R while the value of Z at the highest frequency (100.020 kHz) is R. The values for Rp obtained from the Bode plot data after one hour immersion in the three different electrolytes are given in Table 2. These values indicate that Zirconium 702 is the most corrosion resistant alloy under the conditions used in this study. The ranking of the other 18 alloys differs for the three electrolytes. In general, it can be concluded that for all the alloys, with the exception of Zirconium 702 and Ferralium 255, the Rp values decrease as the concentration of hydrochloric acid increases in the electrolyte. The changes in Rp can thus be followed qualitatively by examining the data in the Nyquist plot format and quantitatively by using the Bode plot format.

4.3 COMPARISON WITH DC POLARIZATION RESULTS

A comparison of the Rp values obtained in this investigation with the Rp values obtained by dc polarization techniques for the same alloys in 3.55% NaCl-0.1N HCl (ref. 2, Table 3) and for 10 of the alloys in 3.55% NaCl-1.0N HCl (ref. 2, Table 9), indicates that there is no correlation between the actual values of Rp. However, there is a good correlation between the ordering of the alloys according to their resistance to corrosion by both methods even though the ordering does not match exactly. The fact that ac impedance techniques use only very small signals which do not disturb the electrode properties to be measured can be used to support the validity of the Rp values obtained by this technique.

4.4 COMPARISON WITH BEACH CORROSION DATA

A comparison of the Rp values obtained from ac impedance measurements with the beach corrosion data for all the alloys in 3.55% NaCl-1.0N HCl (ref. 2, Figure 9) indicates that there is, in general, a good correlation even though the ranking of the metals does not match exactly. For example, Hastelloy C-22 was the most resistant alloy when tested at the beach corrosion site while it ranked seventh according to the Rp value in this study. It should be pointed out that, as a group, there is a good correlation between the alloys that performed at the top (Zirconium 702, Ferralium 255, Inconel 625, Inco Alloy G-3, Hastelloy C-4, Hastelloy, C-276, and

Hastelloy C-22) in both investigations. A tentative conclusion that can be drawn from this comparison is that the ac impedance technique can be used to choose what materials should be subjected to long-term corrosion testing at the beach testing site. The lack of a close correlation between the ac impedance data and the beach corrosion data may result from the fact that the ac impedance measurements were obtained for the alloys after one hour immersion in the aerated electrolytes at room temperature. The conditions at the beach testing site are obviously different (ref. 1, p. 8) and more similar to the conditions in the STS launch environment. The results from the present investigation may be more appropriate for testing the corrosion resistance of alloys that are going to be in contact with liquid electrolytes such as the ones used here.

V. CONCLUSIONS

1. AC impedance techniques, when used for corrosion testing, provide useful qualitative (Nyquist plot) and quantitative information (Rp values) that can be used to screen alloys to be subjected to long-term corrosion testing.

2. The Rp values obtained for the 19 alloys under three different electrolyte conditions can be used to rank the alloys according to their resistance to corrosion since Rp is inversely proportional to the rate of corrosion.

3. Zirconium 702 was found to be the most corrosion resistant alloy under the conditions used in this investigation.

4. There is a good general agreement between the results obtained using dc and ac techniques even though the actual Rp values were found to be different. It is postulated that the Rp values obtained by the ac technique are more accurate.

VI. FUTURE WORK

1. AC impedance measurements involving longer immersion times to investigate how the Rp value -and therefore the rate of corrosion- changes with time.
2. Implementation of the use of software to perform the analysis of the data in the Nyquist plot format in order to calculate Rp values.
3. Include testing of alloys after exposure to conditions as similar to the STS launch environment as possible.
4. Study the effect of protective coatings on the rate of corrosion of the 19 alloys.
5. Modify the electrolyte conditions to include other chemicals normally found at the STS launch environment.
6. Study the effect that a change in temperature, similar to the seasonal changes that occur at the STS launch environment, would have on the rate of corrosion.

REFERENCES

1. MacDowell, L.G. and Ontiveros, C., Evaluation of Candidate Alloys for the Construction of Metal Flex Hoses in the STS Launch Environment, Test Report, Document No. MTB-325-87A, National Aeronautics and Space Administration, Kennedy Space Center, Materials Testing Branch, August 23, 1988.

2. Ontiveros, C., Localized Corrosion of Candidate Alloys for Construction of Flex Hoses, 1987 NASA/ASEE Summer Faculty Fellowship Program Research Reports, Kennedy Space Center, 1987.

3. Application Note AC-1, Basics of AC Impedance Measurements, EG&G PARC, Princeton, NJ., 1984.

4. Mansfeld, F., Don't Be Afraid of Electrochemical Techniques -But Use Them with Care!, Corrosion, Vol. 44, pp. 856-868, 1988.

5. Rothstein, M.L., Electrochemical Corrosion Measurements for the Metal Finishing Industry, Application Note Corr-5, EG&G PARC, Princeton, NJ., 1986.

6. Mansfeld, F., Recording and Analysis of AC Impedance Data for Corrosion Studies. I. Background and Methods of Analysis, Corrosion, Vol. 37, pp. 301-307, 1981.

7. Mansfeld, F., Kendig, M.W., and Tsai, S., Recording and Analysis of AC Impedance Data for Corrosion Studies. II. Experimental Approach and Results, Vol. 38, pp. 570-580, 1982.

8. Lorenz, W.J. and Mansfeld, F., Determination of Corrosion Rates by Electrochemical DC and AC Methods, Corrosion Science, Vol. 21, pp. 647-672, 1981.

9. Williams, D.E. and Naish, C.C., An Introduction to the AC Impedance Technique, and its Application to Corrosion Problems, U.K. Atomic Energy Authority, Harwell Report AERE-M3461, pp. 1-10, 1985.

10. Moody, J.R., Quin, X.P., and Strutt, J.E., The Application of a Compurerized impedance Monitoring System to a Study of the Behavior of 347 ss in Nitric Acid, presented at the 166th Meeting of the Electrochemical Society, New Orleans, Louisiana, 1984.

11. Kendig, M.W., Meyer, E.M., Lindberg, G. and Mansfeld, F., A Computer Analysis of Electrochemical Impedance Data, Corrosion Science, Vol. 23, pp. 1007-1015, 1983.

12. Uhlig, H.H., Corrosion and Corrosion Control. An

Introduction to corrosion science and engineering, Second Edition, John Wiley & Sons Inc., p. 368, 1971.

13. Scantlebury, J.D., Ho, K.N. and Eden, D.A., Impedance Measurements on Organic Coatings on Mild Steel in Sodium Chloride Solutions, Electrochemical Corrosion Testing, ASTM STP 727, Mansfeld, F. and Bertocci, U., Eds., American Society for Testing and Materials, pp. 187-197, 1981.

14. Cahan, B.D. and Chien, C., The Nature of the Passive Film of Iron. II. A-C Impedance Studies, J. Electrochem. Soc., Vol. 129, pp. 474-480, 1982.

TABLE 1 CANDIDATE ALLOYS AND THEIR NOMINAL COMPOSITIONS (WT%)

ALLOY	Ni	Fe	Cr	Mo	Mn*	Co*	Cu	C*	Si*	P*	S*	Other
HASTELLOY C-4	Bal.	3.0	18	17	1.0	2.0		0.01	0.08	0.02	0.01	Ti 0.7
HASTELLOY C-22	Bal.	3.0	22	13	0.5	2.5		0.01	0.08	0.02	0.01	V 0.3, W 3
HASTELLOY C-276	Bal.	7.0	17	17	1.0	2.5		0.01	0.08	0.02	0.01	V 0.3, W 4.5
HASTELLOY B-2	Bal.	2.0	1	28	1.0	1.0		0.01	0.1	0.02	0.01	
INCONEL 600	Bal.	8.0	16		1.0		0.5	0.15	0.5		0.01	
INCONEL 625	Bal.	5.0	23	10	0.5	1.0		0.10	0.5	0.01	0.01	Cb 4.1
INCONEL 825	Bal.	22.0	21	3	1.0		2.5	0.05	0.5		0.03	
INCO G-3	Bal.	20.0	22	7	1.0	5.0	2.0	0.02	1.0	0.04	0.03	Cb 0.5, W 1.5
MONEL 400	Bal.	2.5			2.0		31	0.30	0.5		0.02	
ZIRCONIUM 702												Zr 99.2, Hf 4.5
SS 304L	10	Bal.	19		2.0			0.03	1.0			
SS 304LN	10	Bal.	19		2.0			0.03	1.0	0.04	0.03	N 0.13
SS 316L	12	Bal.	17	2.5	2.0			0.03	1.0	0.04	0.03	
SS 317L	13	Bal.	19	3.5	2.0			0.03	1.0			
SS 904L	25	Bal.	21	4.5	2.0		1.5	0.02	1.0	0.04	0.03	
20 Cb-3	35	Bal.	20	2.5	2.0		3.5	0.07	1.0			
7Mo + N	4	Bal.	28	2	2.0			0.03	0.6	0.03	0.01	N 0.25
ES 2205	5	Bal.	22	3	2.0			0.03	1.0	0.03	0.02	N 0.14
FERRALIUM 255	5	Bal.	26	3	1.5		2.0	0.04	1.0	0.04	0.03	N 0.17

* Values are max.

TABLE 2

POLARIZATION RESISTANCE IN 3.55% NACL

MATERIAL NAME	Rp (ohms)
ZIRCONIUM 702	60884
HASTELLOY C-22	23146
INCONEL 625	18338
HASTELLOY C-276	17121
HASTELLOY C-4	15797
INCONEL 600	15426
7MO + N	15394
FERRALIUM 255	15057
INCO ALLOY G-3	14787
20CB-3	14660
SS 904L	14449
INCONEL 825	14237
SS 304N	13103
SS 316L	12598
SS 317L	12290
ES-2205	11913
SS 304L	11383
HASTELLOY B-2	3779
MONEL 400	652

TABLE 3

POLARIZATION RESISTANCE IN 3.55% NACL-0.1N HCL

MATERIAL NAME	Rp (ohms)
ZIRCONIUM 702	59937
FERRALIUM 255	15675
20CB-3	15369
SS 317L	14745
SS 316L	14290
INCONEL 625	13798
INCONEL 825	13281
ES-2205	13227
SS 904L	12786
INCO ALLOY G-3	12716
7MO + N	12502
HASTELLOY C-4	11723
HASTELLOY C-276	10947
HASTELLOY C-22	9758
HASTELLOY B-2	1806
SS 304L	778
MONEL 400	586
SS 304LN	531
INCONEL 600	498

TABLE 4

POLARIZATION RESISTANCE IN 3.55% NACL-1.0N HCL

MATERIAL NAME	Rp (ohms)
ZIRCONIUM 702	57081
FERRALIUM 255	15133
INCONEL 625	12213
INCO ALLOY G-3	11665
HASTELLOY C-4	8698
HASTELLOY C-276	7201
HASTELLOY C-22	6800
HASTELLOY B-2	1487
MONEL 400	538
INCONEL 825	490
INCONEL 600	481
SS 304LN	431
20CB-3	422
SS 304L	325
SS 904L	300
SS 317L	203
SS 316L	158
ES-2205	101
7MO + N	49

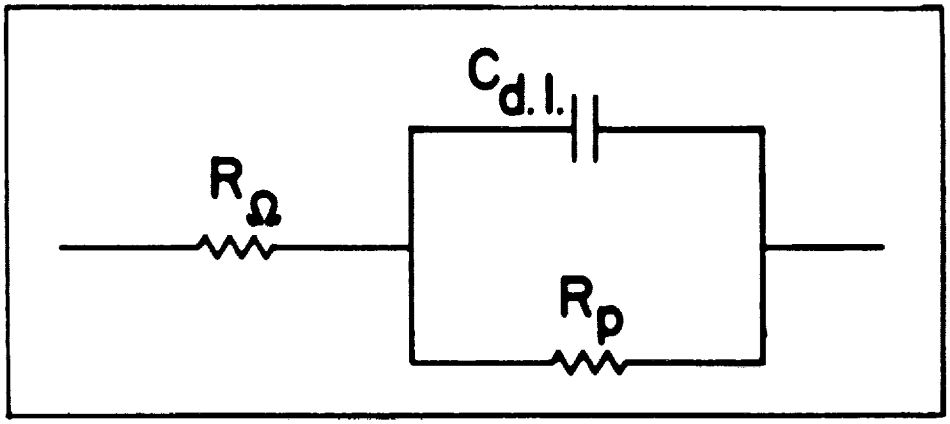


Figure 1. Equivalent circuit for a simple electrochemical cell.

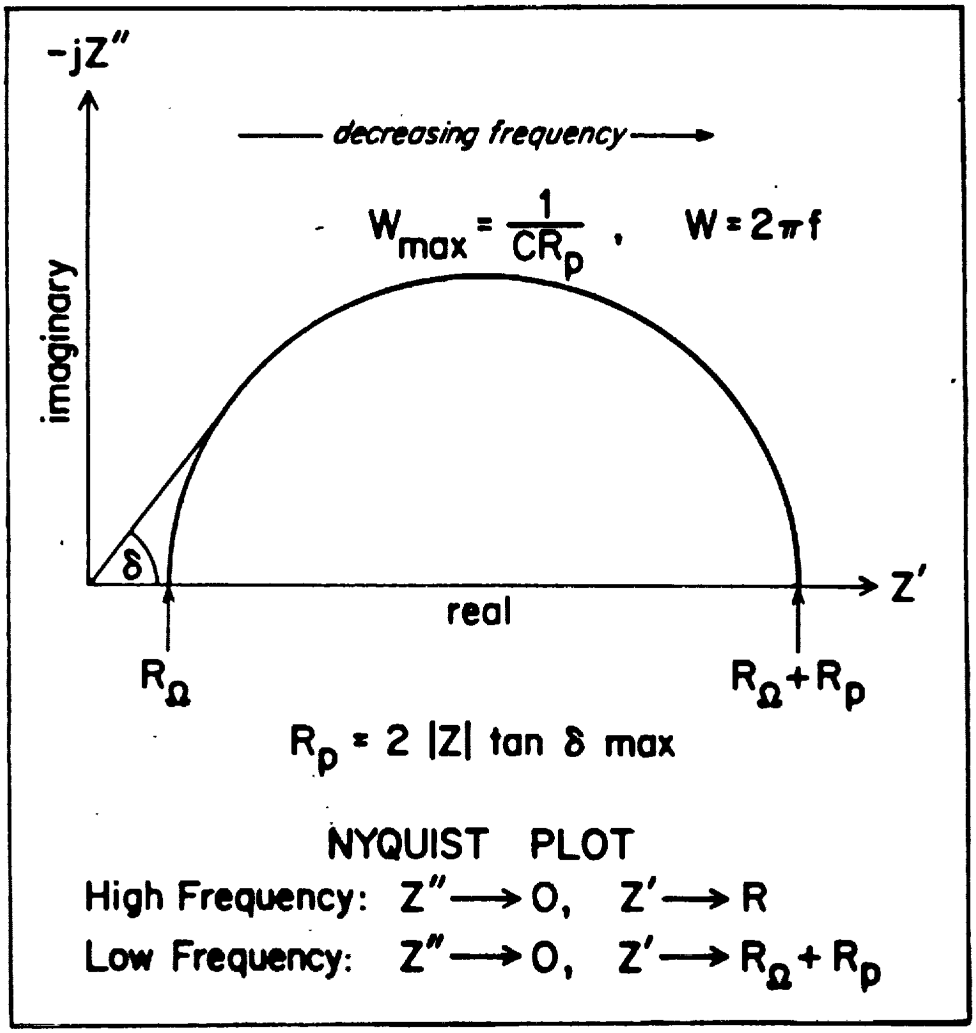


Figure 2. Nyquist plot for equivalent circuit in Figure 1.

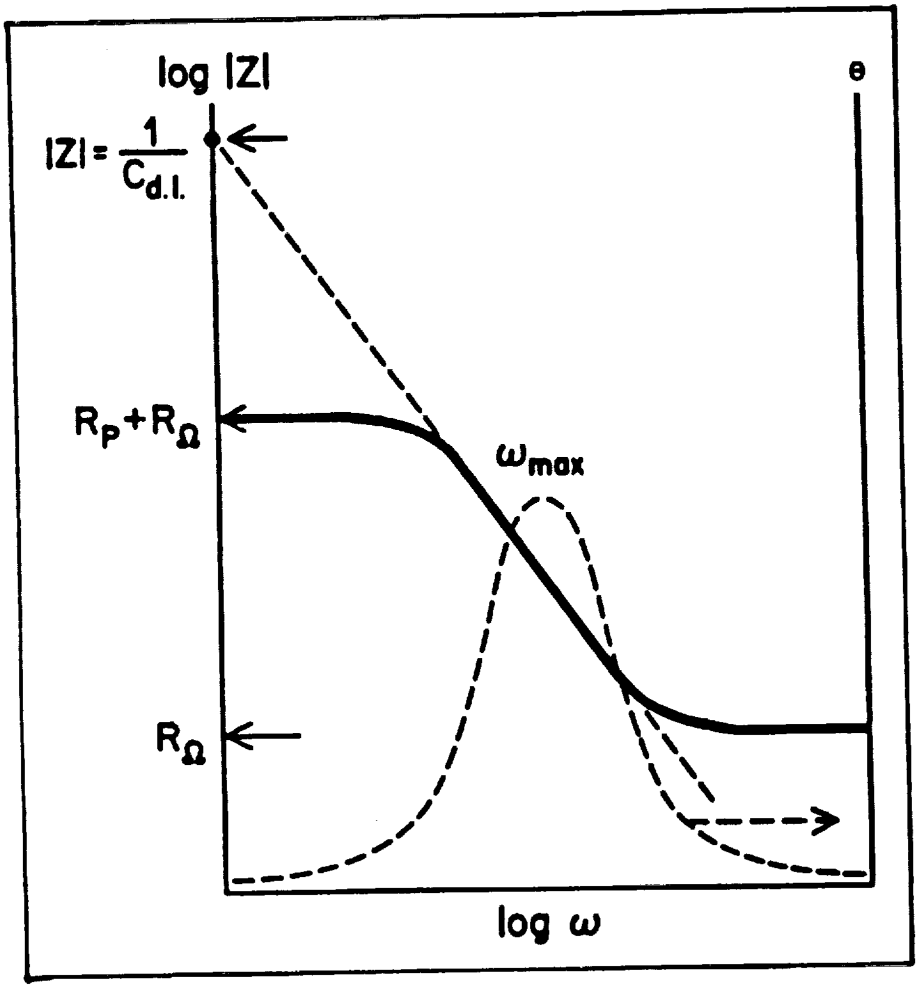


Figure 3. Bode plot for equivalent circuit in Figure 1.

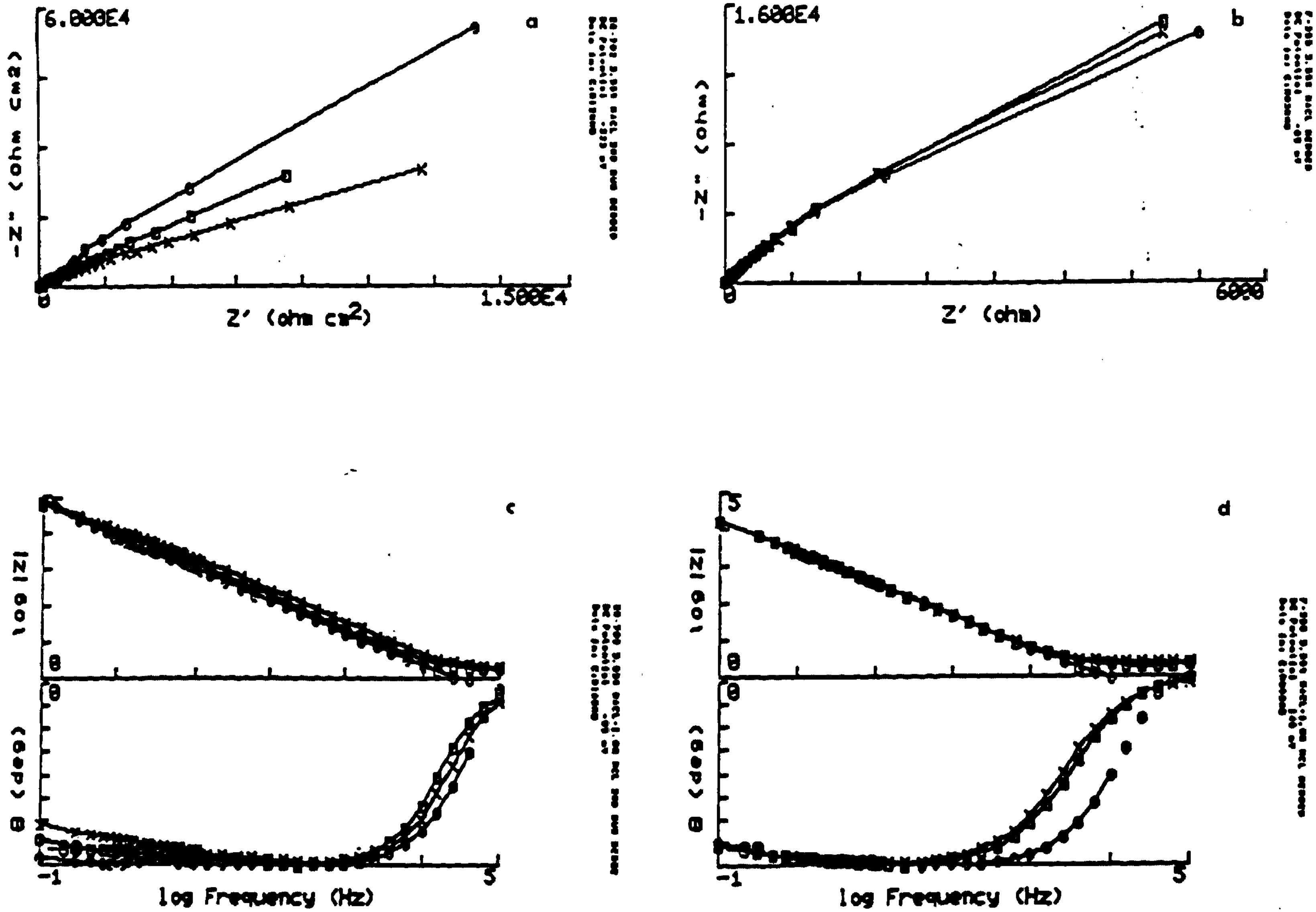


Figure 4. Nyquist and Bode plots for Zirconium 702 (a,c) and Ferralium 255 (b,d).

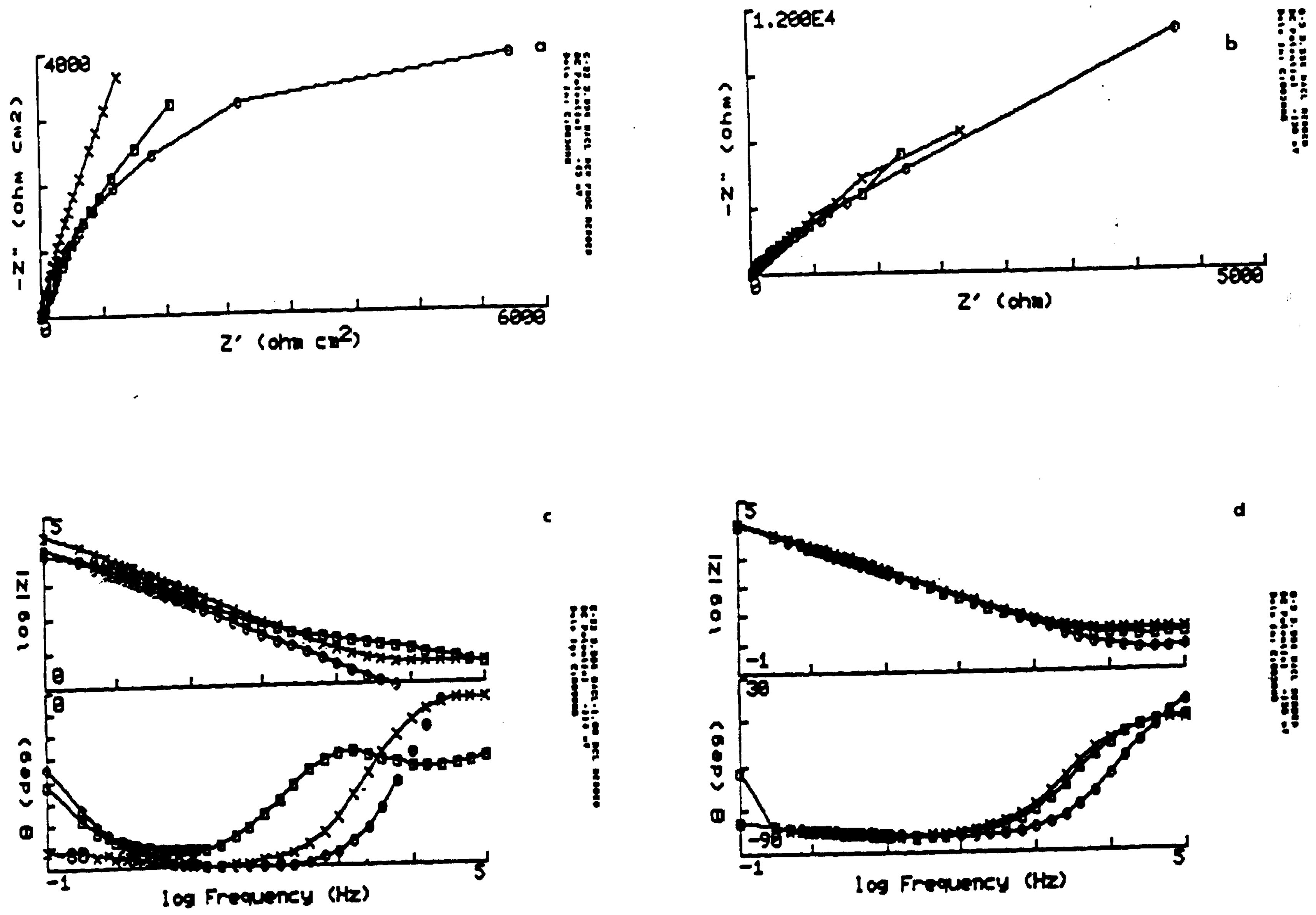


Figure 5. Nyquist and Bode plots for Hastelloy C-22 (a,c) and Inco Alloy G-3 (b,d).

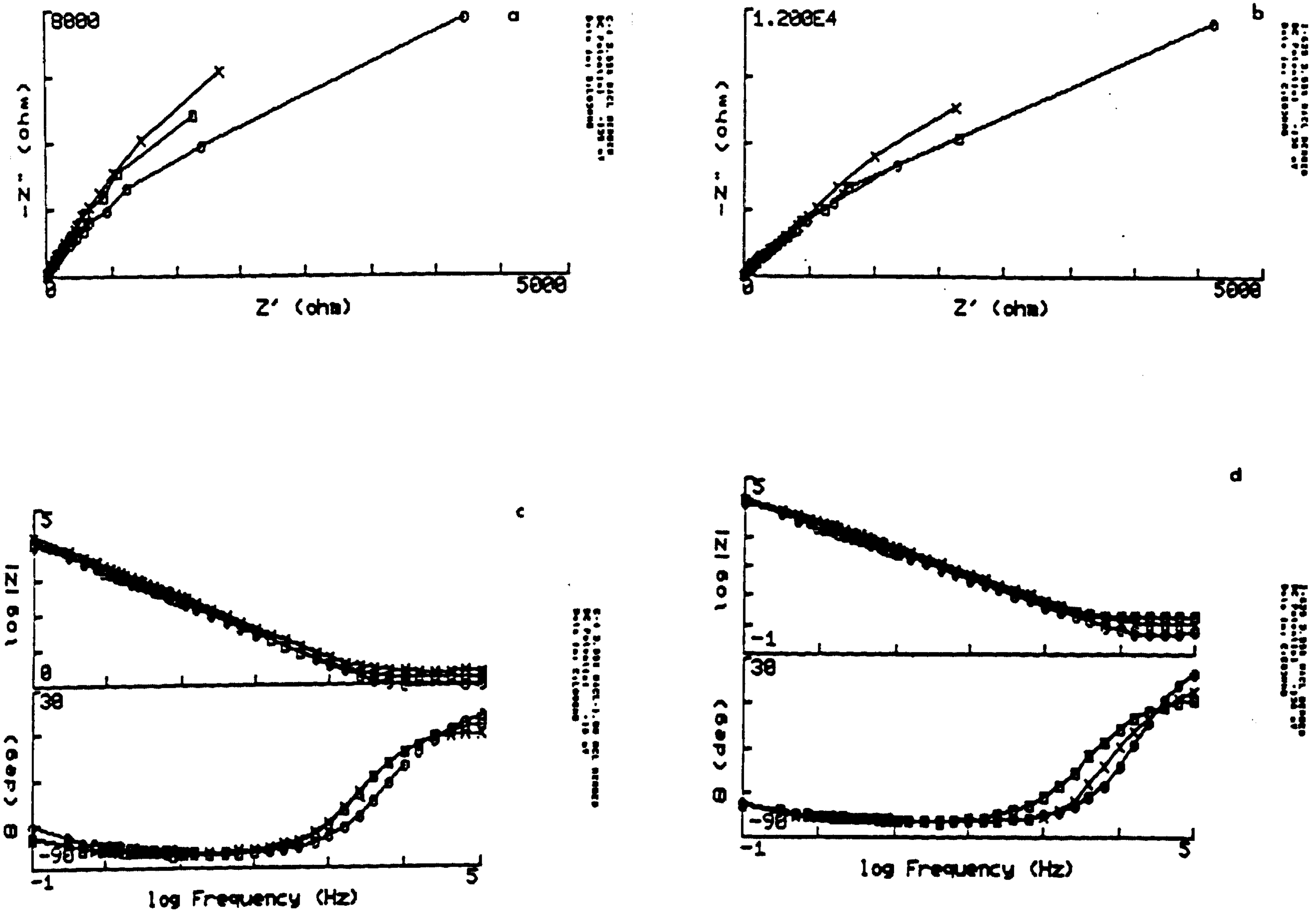


Figure 6. Nyquist and Bode plots for Hastelloy C-4 (a,c) and Inconel 625 (b,d).

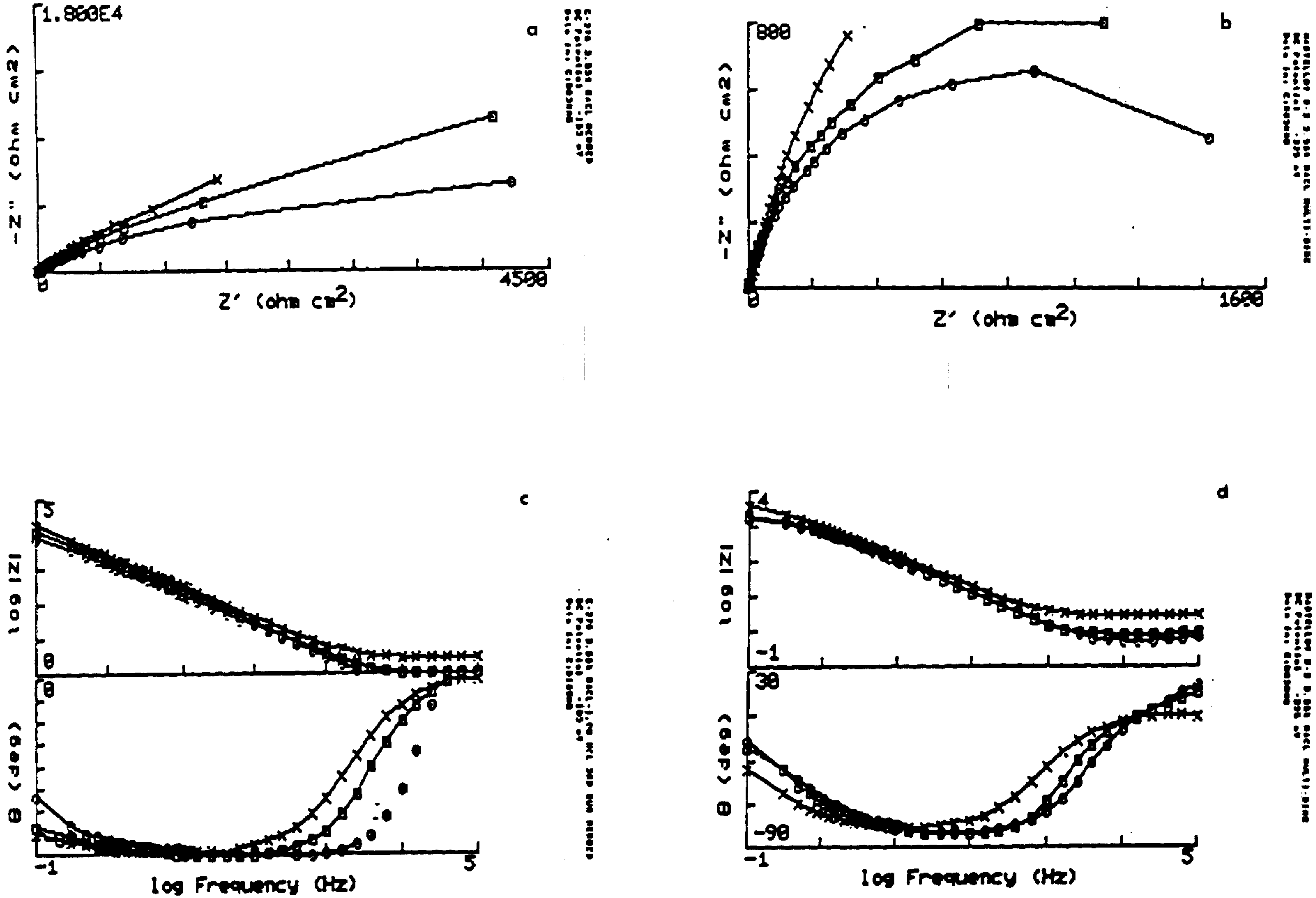


Figure 7. Niquist and Bode plots for Hastelloy C-276 (a,c) and Hastelloy B-2 (b,d).

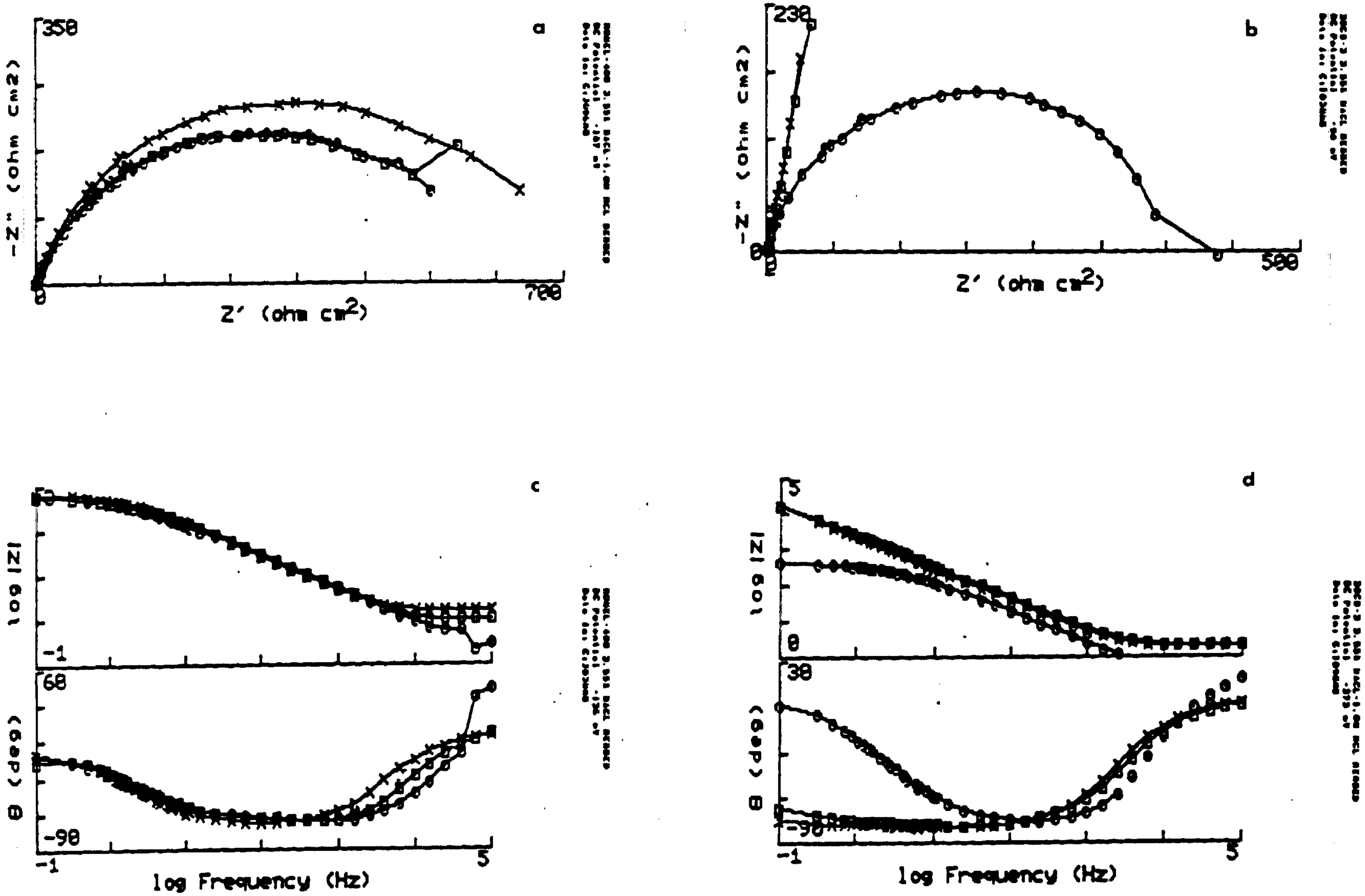


Figure 8. Nyquist and Bode plots for Monel 400 (a,c) and 20Cb-3 (b,d).

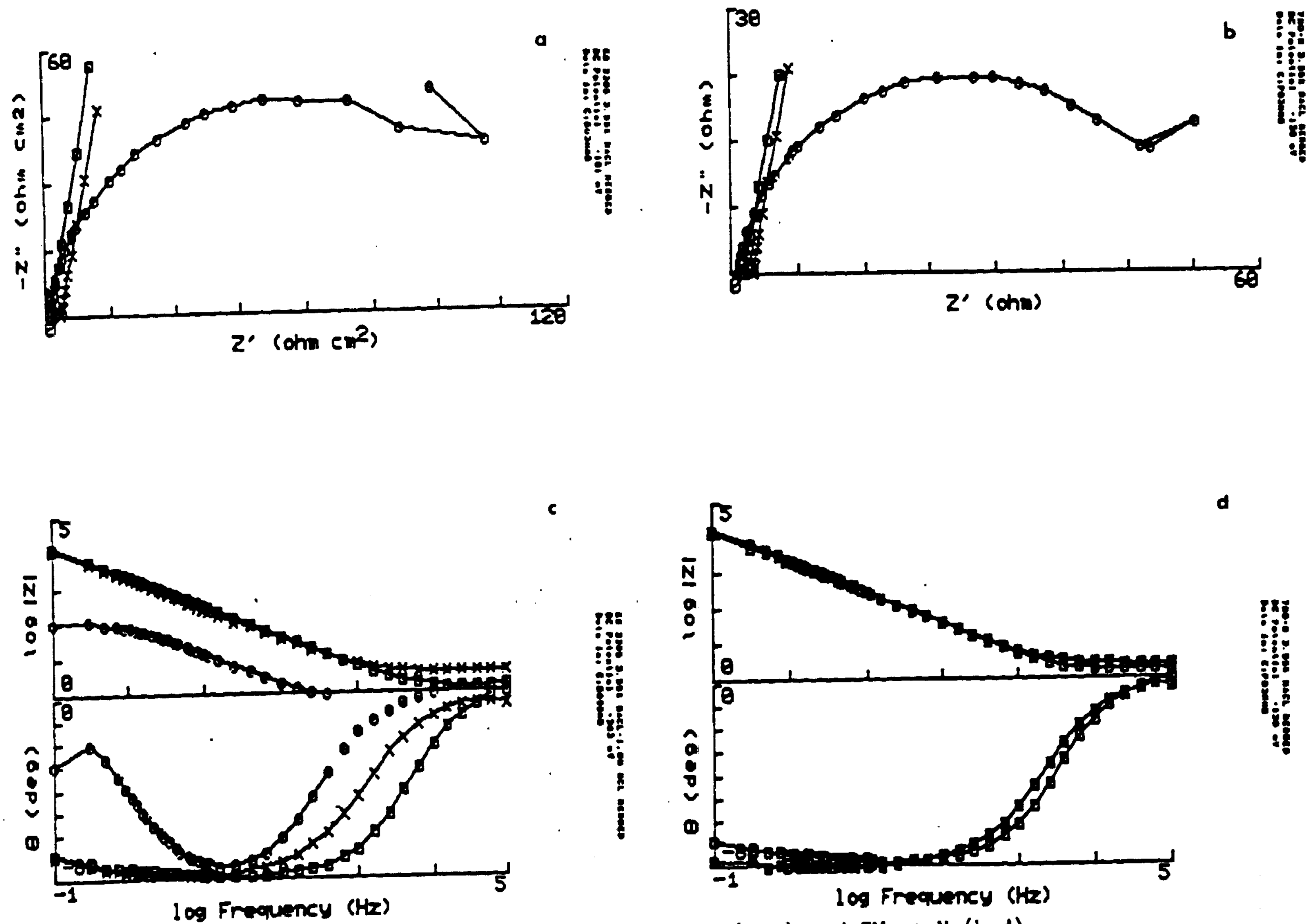


Figure 9. Nyquist and Bode plots for ES 2205 (a,c) and 7Mo + N (b,d).

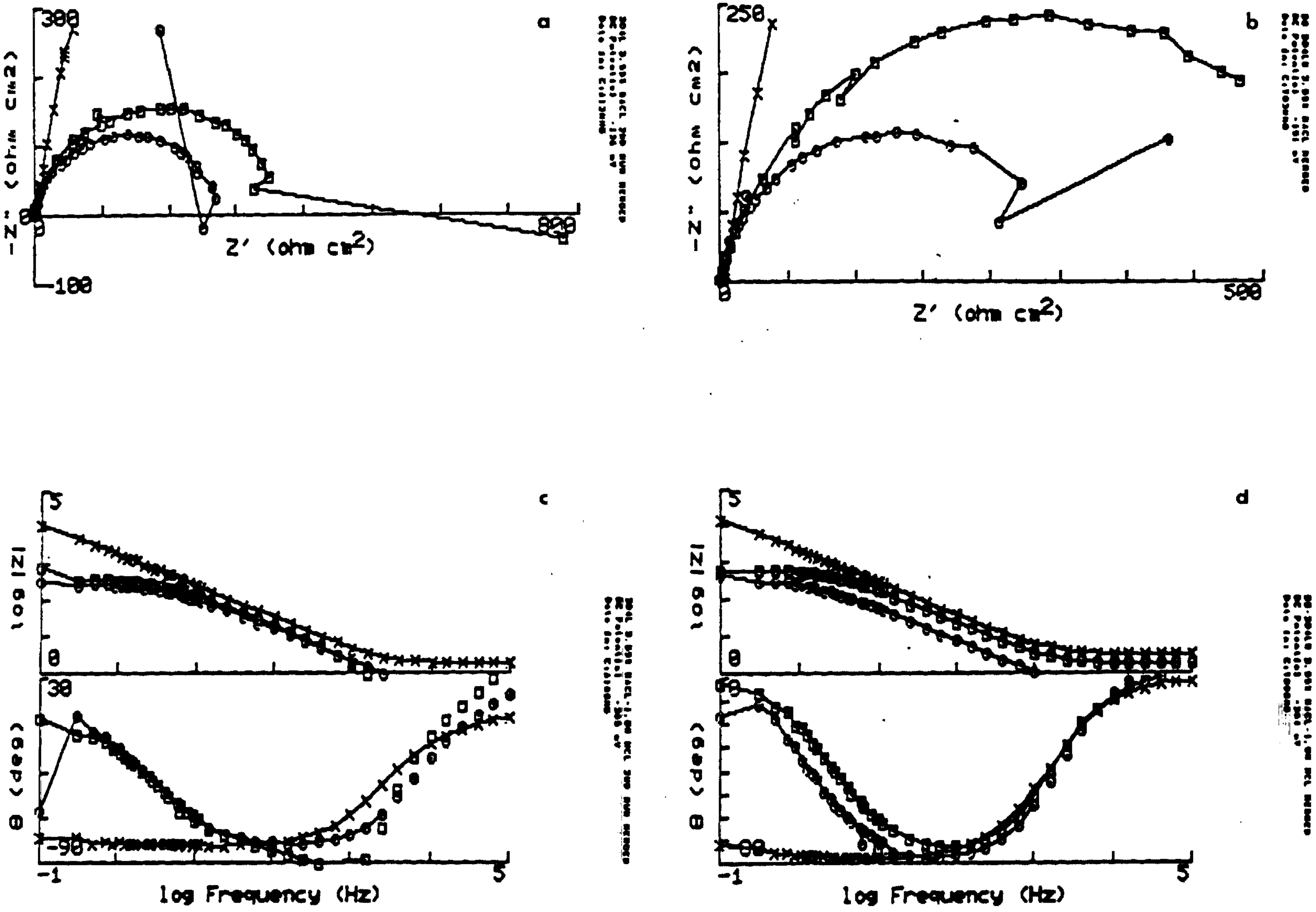


Figure 10. Nyquist and Bode plots for SS 304L (a,c) and SS 304LN (b,d).

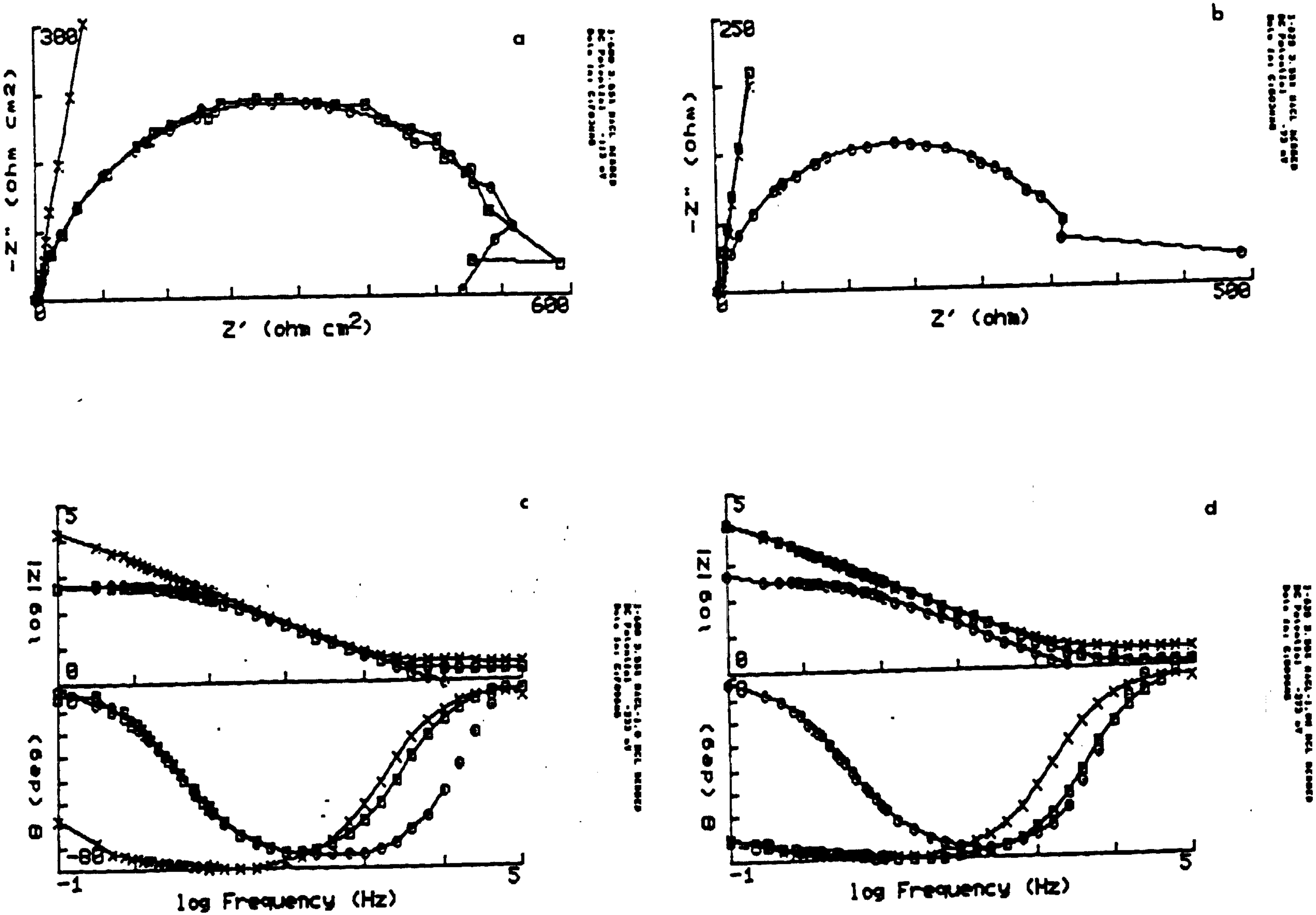


Figure 11. Nyquist and Bode plots for Inconel 600 (a,c) and Inconel 825 (b,d).

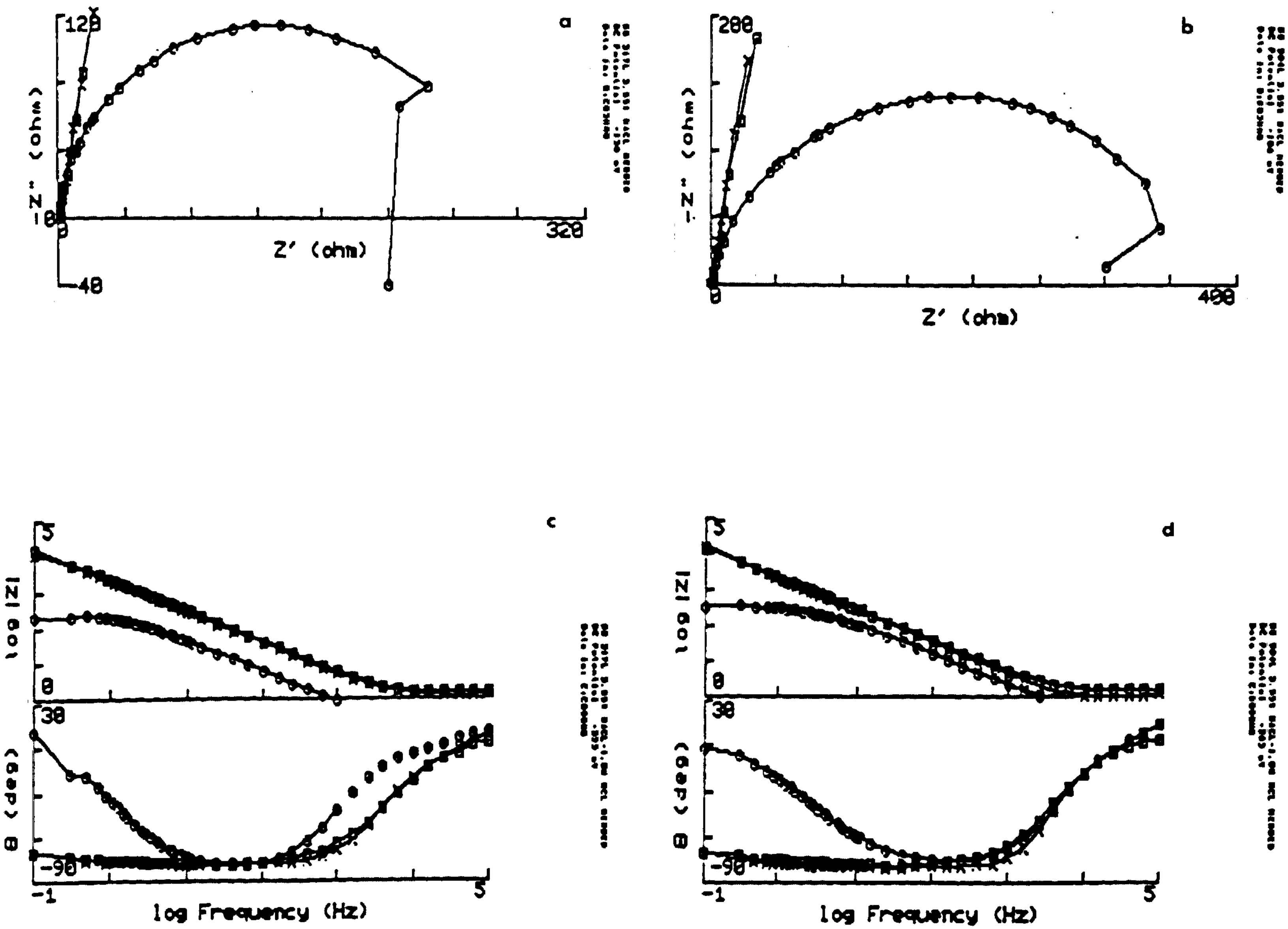


Figure 12. Nyquist and Bode plots for SS 317L (a,c) and SS 904L (b,d).

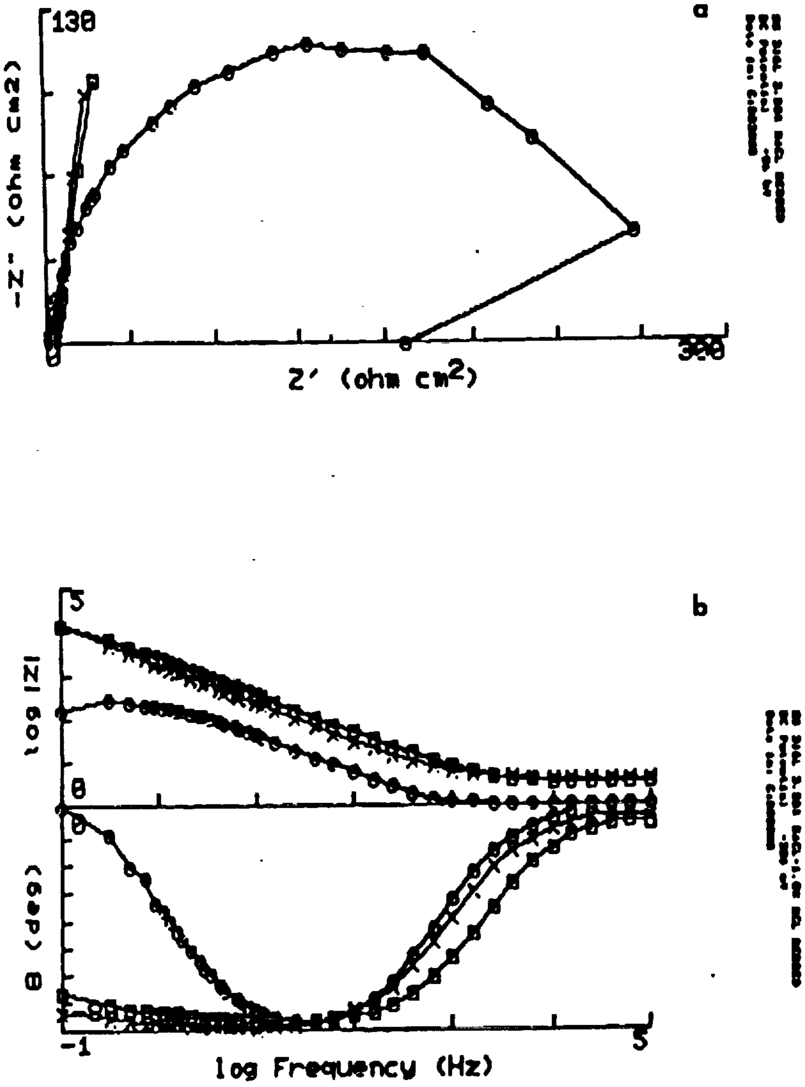


Figure 13. Nyquist and Bode plots for SS 316L (a,b).

N90-16689

1989 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

DYNAMICS OF CARBON DIOXIDE EXCHANGE OF A WHEAT COMMUNITY GROWN IN A SEMI-CLOSED ENVIRONMENT

PREPARED BY:	Dr. Kenneth A. Corey
ACADEMIC RANK:	Associate Professor
UNIVERSITY AND DEPARTMENT:	University of Massachusetts Department of Plant & Soil Sciences
NASA/KSC	
DIVISION:	Biomedical Operations and Research
BRANCH:	Life Sciences Research Office
NASA COLLEAGUE:	Mr. Ralph P. Prince
DATE:	August 4, 1989
CONTRACT NUMBER:	University of Central Florida NASA-NGT-60002 Supplement: 2

ACKNOWLEDGEMENTS

I am extremely grateful for the opportunity to be a part of the 1989 NASA/ASEE Summer Faculty Fellowship Program. Many individuals are responsible for helping to make this an enjoyable, productive, and rewarding experience.

Special thanks is due to program director E. Ramon Hosler and his administrative assistant Kari Baird for their smooth handling and coordination of tours, activities, special events, and general administrative details. To my KSC colleague, Ralph P. Prince, I extend my appreciation for making this wonderful opportunity possible and also for many stimulating discussions related to the CELSS program.

A most special thanks is extended to Raymond M. Wheeler, the project leader and my collaborator. It was only through Ray's leadership and experience that this project could be conducted. For his patience and time spent discussing experimental protocol and interpretations of data, I am most grateful.

I am grateful to John Sager with whom I've had many stimulating discussions and who introduced me to a valuable mathematics software package. I also acknowledge Glenn A. Markwell, for his assistance in data retrieval and for providing PLC input statements for experiments, Cheryl L. Mackowiak for planting and general crop maintenance, Katrina Chetirkin for her assistance in data acquisition, and Peter B. Heifetz for introducing me to a helpful graphics software package.

This project represents a small piece of work associated with the larger scale efforts undertaken by many dedicated individuals. To all those people who have contributed to the design, development, and testing of the BPC and who I have not mentioned, I extend my gratitude.

Finally, I would like to acknowledge the support of colleagues from my home institution, the University of Massachusetts; particularly Allen V. Barker, John H. Baker, and Lester F. Whitney for providing letters of reference which helped lead to this opportunity.

ABSTRACT

A wheat (Triticum aestivum 'Yecora Rojo') community was grown in the semi-closed conditions of the NASA/KSC Biomass Production Chamber (BPC). Experiments were conducted to determine whole community carbon dioxide exchange rates as influenced by growth and development, carbon dioxide concentration, time within the photoperiod, irradiance, and temperature. Plants were grown at a population of about $1500/m^2$ using a 20 hour light/4 hour dark daily regime. Light was supplied by HPS vapor lamps and irradiance was maintained in the range of 590 to 675 $\mu mol/m^2/s$. The temperature regime was 20 C light/16 C dark and nutrients were supplied hydroponically as a thin film.

Fractional interception of PPF by the community increased rapidly during growth reaching a maximum of 0.96, 24 days after planting. This time corresponded to canopy closure and maximum rates of net photosynthesis (NP). Net daily CO_2 utilization rates were calculated to day 48 and a 4th order regression equation integrated to obtain total moles of CO_2 fixed by the community. This procedure may be useful for monitoring and prediction of biomass yields in a CELSS.

Results of 5, 1-hour photosynthetic drawdowns of CO_2 during the photoperiod and from the data log of mass flow injections of CO_2 into the chamber indicated a constant rate of NP during the photoperiod. Net photosynthesis appeared to be relatively linear with CO_2 concentration down to about 400 ppm, suggesting minimal enhancement of NP by supra-ambient CO_2 concentrations with the light regime used. The CO_2 compensation point (CCP) was in the range of 45 to 55 ppm and was independent of irradiance. Carbon dioxide exchange rates were related linearly to irradiance up to 750 $\mu mol/m_2/s$, indicating that growth was not light saturated. Light compensation points (LCP) of the whole community were in the range of 193 to 223 $\mu mol/m^2/s$ and did not appear to change during growth between days 13 and 41. Net photosynthetic rates for a fully developed canopy were maximized between 16 and 20 C and dark respiration was minimized at 16 C.

This study resulted in the development of a broad database for the CO_2 exchange dynamics of a wheat community and will be valuable in the selection of environmental conditions in future biomass production efforts for CELSS. The work also demonstrated the unique research capabilities of the BPC and suggested an approach to the monitoring and prediction of crop biomass.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

BPC	Biomass Production Chamber
CER	Carbon dioxide Exchange Rate
CCP	Carbon dioxide Compensation Point
CELSS	Controlled Ecological Life Support System
conc	concentration
DR	Dark Respiration
f	fractional interception
HPS	High Pressure Sodium
I	Incident PPF
IRGA	Infra-red Gas Analyzer
LCP	Light Compensation Point
NCU	Net Carbon dioxide Utilization
NP	Net Photosynthesis
PAR	Photosynthetically Active Radiation
PHICS	Plant-Human Integration in Closed Systems
PLC	Programmable Logic Controller
PPF	Photosynthetic Photon Flux
ppm	parts per million
R	Reflected PPF
RH	Relative Humidity
T	Transmitted PPF

I. INTRODUCTION

1.1 BACKGROUND

Humans living for long durations on lunar or planetary bases, on planetary missions, and in space colonies will require the use of controlled ecological life support systems (CELSS) to produce, process, and recycle biomass (1,2,3). An idealized CELSS would sustain human life indefinitely without the necessity for materials resupply. Human oxygen and nutritional requirements would be met by the production and processing of plant biomass. Carbon dioxide from the respiratory activity of humans and other heterotrophic organisms would provide the necessary carbon for photosynthesis while nutritional requirements of plants would be supplied by recycling of human and plant wastes.

An important step by NASA in the development of a bioregenerative life support system was the initiation of the CELSS Breadboard Project (4,5) which has as its primary goals the design, construction, and evaluation of systems and techniques for the production, processing, and recycling of biomass. Implementation of these goals began with the construction and testing of a pre-prototype biomass production chamber (6). A hypobaric test vessel used in the Mercury and Gemini programs was modified and systems constructed for heating, ventilation, and air conditioning (HVAC), lighting, and nutrient delivery (6). In addition, the BPC was provided with capabilities for the monitoring and control of atmospheric variables (temperature, relative humidity, and carbon dioxide and oxygen concentrations), nutrient solution variables (quantity, flow rate, temperature, pH, and conductivity), and irradiance (PPF).

Currently, the BPC is being tested under semi-closed conditions for the production of wheat (Triticum aestivum) and it is planned that future tests will include other staple crops such as soybean (Glycine max), Irish potato (Solanum tuberosum), and sweet potato (Ipomeae batatas). A major constraint imposed on these crop evaluations is that they are conducted in just one experimental unit. However, relative to studies in conventional growth chambers, the high degree of environmental control and the large size of the plant community combine to provide the BPC with the advantage of a well-integrated sample. Moreover, the BPC has been designed and constructed to be operated as a closed, nearly gas-tight system. This characteristic is particularly valuable for the conduct of short term experiments involving the monitoring of gas exchange processes as influenced by environmental variables.

An important aspect of environmental control in a CELSS will be the monitoring and regulation of atmospheric gases (7) at concentrations necessary for the maintenance of all life forms. For example, it is important to know the rate of carbon dioxide utilization by a crop community under various environmental conditions so that appropriate designs and control systems for maintaining oxygen and carbon dioxide mass balances can be achieved. Major variables that influence the carbon dioxide exchange rates of plants are temperature, irradiance, and carbon dioxide concentration. Results of small-scale growth chamber studies (8,9) can provide the appropriate ranges of these variables to select for use in community-scale studies with the BPC. The unique capabilities of the BPC can then be used for the conduct of non-steady state gas exchange experiments with relatively large crop samples.

1.2 OBJECTIVES

The overall objective of this study was to begin the contruction of a gas exchange database for wheat grown in a semi-closed environment. Specific objectives were to determine: a) rates of net photosynthesis (NP) and dark respiration (DR) during growth and development, b) influence of irradiance on carbon dioxide exchange rates, c) carbon dioxide and light compensation points of the whole community, and d) effects of temperature on net photosynthesis and dark respiration.

II. MATERIALS AND METHODS

2.1 PRODUCTION REGIME

The hard red spring wheat cultivar Yecora Rojo was planted in the Biomass Production Chamber (BPC) in the Life Sciences Support Facility, Hangar L, Kennedy Space Center (KSC) on May 30, 1989. Seed presoaked for 30 min and held for 4 days at about 3 C, were planted at a rate of 1900 to 2000/m^2 into strips of polyethylene (white top/black bottom) attached to "T" shaped PVC support pieces crossing the tray tops. Seed were suspended above the tray bottom by opposing strips. The trapezoidally shaped trays were 5 cm in depth and each provided approximately 0.25 m^2 growing area. Each of the 4 levels in the BPC contained 16 trays, providing a total growing area of 16 m^2 excluding areas between trays and outside trays where foliage extended during growth.

Germinating seedlings were covered and maintained in the dark for 2 days following planting at a RH of 90 % and a temperature of 24 C in the BPC. Following an additional 2 days at 24 C, covers were removed and a daily temperature regime of 20 C light/16 C dark was used throughout growth and development except for 3, 3-day spans when temperature was varied. Relative humidity in the BPC was reduced to approximately 70 % on day 7 and maintained as such thereafter. An average stand of 1488 plants/m^2 was measured from a sample of 8 trays following emergence.

The nutrient delivery system consisted of supplying a thin film of a modified half-strength Hoagland's (10) solution at a flow rate of 1 liter/min/tray. The pH of the nutrient solution was maintained in the range of 5.8 to 6.2 through automatic injections of HNO_3 as the pH increased upon crop uptake of NO_3^--N.

The wheat community was provided with a daily regime of 20 hour light/4 hour dark. Lighting during the 20 hour photoperiod was supplied by high pressure sodium (HPS) vapor lamps attenuated when necessary to provide an average canopy top irradiance of about 600 $\mu mol/m^2/s$.

Carbon dioxide concentration was maintained at 1000 ± 50 ppm throughout growth and development except during daily dark respiration and photosynthetic drawdown episodes and when experimental drawdowns were conducted. Establishment of set-points, ranges, and timing of events for all BPC subsystem variables was directed by programmable logic controllers (PLC).

2.2 EXPERIMENTAL

2.2.1 IRRADIANCE MEASUREMENTS. Instantaneous photosynthetic photon flux (PPF) measurements were made at least once weekly using a sunfleck ceptometer (model SF-40, Decagon Devices Inc., Pullman, WA 99163). The ceptometer measures irradiance in the wavelength region of 400 to 700 nm or photosynthetically active radiation (PAR). An average of 5 readings from each of the 64 trays was taken in each of 3 positions; 1) upfacing ceptometer at the canopy top, termed the incident PPF (I), 2) upfacing ceptometer at the canopy bottom or transmitted PPF (T), and 3) inverted ceptometer at the canopy top, or reflected PPF (R). Calculations of the fractional interception, f, were made from the following equation using averages of all readings.

$$f = 1 - T/I - R/I$$

After 21 days from planting, irradiance levels at the canopy bottom approached zero and measurements of T were taken from only 4 trays on each level.

2.2.2 RATES OF NET PHOTOSYNTHESIS (NP) AND DARK RESPIRATION (DR). Daily determinations of the rates of NP and DR were made from changes in CO_2 conc in the upper and lower levels of the BPC atmosphere. Carbon dioxide conc was monitored with infra-red gas analyzers (IRGA) and computer-logged at 4-min intervals. At 0200 hr each day, the lights were set to turn off and the CO_2 conc in the BPC increased linearly due to respiratory evolution. At 0600 hr, the CO_2 conc was above the 1000 ppm setpoint, the lights turned on, and a linear decrease (photosynthetic drawdown) in CO_2 conc occurred. Carbon dioxide conc at 20-min intervals between 0300 and 0500 hr and at 12-min intervals between 0628 and 0728 were used to obtain slopes of equations fit by linear regression. Whole community rates of NP and DR were calculated from the slopes assuming a BPC volume of 112,060 liters (7). Leak rates of CO_2 over the short time span from which calculations were made were assumed to be negligible. A BPC leak rate of 0.417 %/hr was measured in April, 1989 following the retrofitting of inflatable door seals (John Sager, personal communication).

On days 44-48, 4 trays representing approximately 6 % of the total chamber biomass were removed for study of downward rolling of leaves. Data for NP and DR were corrected for the biomass removed during that period by using the factor 64/60 (1.0667).

Calculations of rates of NP for the 20 hr photoperiod and of DR for the 4 hr period on a whole chamber basis were made to determine a net daily rate of CO_2 utilization (NP - DR). A polynomial was fit to the net daily rates of CO_2 utilization (NCU) and the resulting function integrated over time to obtain a total quantity of CO_2 fixed. A solution to the definite integral was obtained using the Math CAD software program (Mathsoft, Inc., Cambridge, MA 02139).

2.2.3 DIURNAL CO_2 EXCHANGE RATES (CER). On day 36, photosynthetic drawdowns of CO_2 conc were measured during 5 time intervals of the photoperiod; 0628 - 0728, 1012 - 1112, 1508 - 1608, 2018 - 2118, and 0006 - 0106 (day 37). Linear regressions were fit to 6 equally spaced points over each 60-min interval. Prior to each drawdown, CO_2 conc in the BPC was increased to about 1200 ppm and the flow valves closed to avoid PLC-directed injection of CO_2 upon depletion of CO_2 below the 1000 ppm setpoint. Rates of DR were determined before and after the photoperiod.

2.2.4 DETERMINATION OF CO_2 COMPENSATION POINT (CCP). On days 18 and 19, the CCP of the wheat community was approximated at 2 irradiance levels (I = 675 µmol/m^2/s on day 18 and I = 497 µmol/m^2/s on day 19). This was achieved by closing the CO_2 injection flow valves, and allowing the BPC CO_2 conc to decline. The barrier separating the upper and lower levels of the BPC was removed to facilitate mixing of gases in the entire chamber. Carbon dioxide drawdowns were conducted until the rate of CO_2 decrease approached zero. On day 19, one of the flowmeters for CO_2 injection was slightly open and a period of injection occurred when the CO_2 conc was below 500 ppm. In handling this data, a 90 min segment of CO_2 conc data was spliced out in order to approximate the time-dependent drawdown pattern.

2.2.5 EFFECTS OF IRRADIANCE ON NP. On days 13, 20, 27, and 41, irradiance was varied by dimming the HPS lamps to provide 5 irradiance levels. At the start of the experiment, BPC CO_2 conc was increased to approximately 2000 ppm and then allowed to change for 1 hr at each irradiance. Carbon dioxide conc at 8 min intervals for a 32 min time span near the midddle of each PPF drawdown event was used in calculations of CER. Measurements of incident PPF for each irradiance level were made for all 64 trays the day after the experiment. Light compensation points for the community were calculated (x-intercept) from the linear regressions of CER vs. PPF.

2.2.6 EFFECTS OF TEMPERATURE ON NP AND DR. On days 21 - 23 and 41 - 43, temperature of the BPC was varied. For each 3-day experiment, temperature was set at 24 C for day 1, 20 C for day 2, and 16 C for day 3. Rates of NP and DR were calculated as described previously.

2.2.7 MASS FLOW MEASUREMENTS. Data logged for a cumulative record of CO_2 injected into the BPC was retrieved and used to calculate NP during uninterrupted times when the chamber was sealed. Calculations of NP from mass flow data were compared with those made from drawdown measurements.

III. RESULTS AND DISCUSSION

3.1 IRRADIANCE

Irradiance incident at the canopy top was maintained between 590 and 675 μmol/m^2/s throughout growth and development (Table 3-1). As the wheat plants increased in height, it was necessary to alter the dimming levels of the light banks in order to remain within this range. Sharp decreases in the irradiance transmitted to the bottom of the canopy (T) and reflected (R) at the top of the canopy occurred after day 9. During this rapid growth phase of the wheat community there was a near doubling in the fractional interception (f) of PPF by the canopy. Canopy closure occurred after about 3 weeks as indicated by f values of 0.94 to 0.96. After 24 days, transmitted PPF declined to negligible levels and remained negligible until day 42. It is anticipated that both transmitted and reflected PPF will increase during senescence of the canopy.

Table 3-1. Incident (I), transmitted (T), and reflected (R) PPF and fractional interception (f) in the BPC during growth and development of a wheat community.

DAYS AFTER PLANTING	PPF (μmol/m^2/s)[a,b]			
	I	T	R	f
9	628 ± 44	317 ± 28.0	104 ± 11.2	0.34
11	665 ± 42	173 ± 22.0	46 ± 5.5	0.67
14	675 ± 47	66 ± 9.0	26 ± 2.1	0.86
17	675 ± 30	37 ± 2.6	13 ± 2.9	0.93
21	620 ± 27	5 ± 0.6	30 ± 2.0	0.94
24	659 ± 33	2 ± 0.7	27 ± 1.3	0.96
28	610 ± 17	ND	ND	ND
31	601 ± 28	1 ± 0.5	22 ± 0.5	0.96
36	590 ± 30	1 ± 0.5	26 ± 3.3	0.96
42	616 ± 42	1 ± 0.5	36 ± 9.8	0.94

[a] Values represent the means of 4 levels ± 1 standard deviation.
[b] ND=no data.

3.2 GROWTH AND DEVELOPMENT - NP AND DR

Wheat seedlings became photosynthetically competent 6 days after planting. At this stage it was possible to obtain photosynthetic drawdown rates following the dark respiration period. The rate of NP increased sharply up to about 15 days and then remained in the range of 30 to 35 µmol/m^2/s for about 3 weeks, followed by a gradual downward trend up to day 48 (Figure 3-2). Rates of NP reported herein from whole community CO_2 drawdowns are comparable to those reported for single leaf blades (11). The trend for rates of dark respiration early in development was similar to that observed for rates of NP. After 2 weeks the rate of DR remained relatively constant. During the interval of days 6 to 14, day-to-day variability in rates of NP and DR were low. Subsequent fluctuations in the daily rates are attributable to the conduct of temperature experiments (days 22-24 and days 42-44), changes in canopy geometry (growth) with respect to lighting, and periodic adjustments in lamp output to maintain a relatively uniform level of canopy top irradiance throughout growth.

Maximum daily rates of CO_2 utilization (35 mol/chamber/day) occurred between 3 and 4 weeks after planting (Figure 3-3). The equation fit to the data in figure 3-3 is:

$$Y = -39.912 + 10.129\ X - 0.496\ X^2 + 0.010331\ X^3 - 0.000079541\ X^4,$$
$$R^2 = 0.881.$$

This equation can be integrated to obtain the total moles of CO_2 fixed by the entire wheat community. Solving the definite integral of the above equation between days 6 and 48 yields a total CO_2 fixed of 1,202 moles or 52.9 kg. Assuming that 1 mole of CO_2 yields 1 mole of CH_2O, then 52.9 kg of CO_2 should yield 36.1 kg of carbohydrate (12). Upon obtaining a complete growth and development data set, this procedure may be used in conjunction with data on ash and total N as percentages of dry weight to predict total community biomass.

A gradual downward trend in NP (Figure 3-2) is expected to occur. However, a gradual decrease in NP may be accompanied by a sharp increase in DR as the plants senesce leading to a large decrease in the daily net fixation of carbon dioxide.

3-3 DIURNAL PATTERN OF CARBON DIOXIDE EXCHANGE RATES

A question that arose from using morning photosynthetic drawdown rates of CO_2 as a measure of the daily rate of NP was the following. How representative of the daily rate was a 1 hr segment of CO_2 conc changes obtained at the beginning of a 20 hr photoperiod? Measurements of NP obtained from the beginning and end of the photoperiod and 3 intermediate times revealed some variation in the rate of NP computed in this manner, but not enough to discern any diurnal pattern in the rate (Figure 3-3). The regions of the dotted lines represent presumed extrapolations of the data. Actual plant reponses could differ only slightly from these extrapolations. The absence of a distinct diurnal pattern in NP for plants grown at constant irradiance and angle of incidence on the canopy contrasts with photosynthetic responses of plant canopies receiving only solar radiation, since the angle of incidence and the intensity of solar radiation varies throughout a day.

Further evidence to claim a relatively constant rate of NP was obtained from mass flow data logged during a complete photoperiod when the BPC doors remained sealed (Figure 3-4). The rate of NP calculated from the slope of the line presented in Figure 3-4 was 420 µmol/chamber/s as compared to a rate of 467 µmol/chamber/s obtained from the morning photosynthetic drawdown data. The disparity between the two methods of calculating rates of NP may derive from several sources. First, the leak rate of carbon dioxide from the chamber over a period of 18 hours used in the mass flow method of calculation could cause a significant error in the measurement. This should lead to an overestimate of the true rate. However, the drawdown computation method yielded a higher value for NP and therefore suggests that other factors contribute to the disparate measurements. Another possibility is the variability in IRGA measurements used to obtain the drawdown data. The IRGA is subject to drift and needs to be calibrated on a frequent basis to ensure consistency in the daily absolute values of CO_2 conc. Such instrument variability should be random and result in times when the mass flow method yields higher or lower measurements of the rate than the drawdown method. This contention was confirmed by comparison of data sets from other days and it was concluded that factors contributing to the disparity between methods were of a random nature.

3-4 CARBON DIOXIDE COMPENSATION POINT

Photosynthetic drawdowns of CO_2 at PPF values of 497 and 675 µmol/m^2/s were linear to CO_2 concentrations below 400 ppm, suggesting that CO_2 enrichment above ambient levels has a minor effect on the rate of NP of a wheat community at these irradiances and stage of development (Figure 3-6). However, it will be necessary to repeat this experiment, perhaps at a higher irradiance, to test this possibility.

The discontinuity in the presentation of the data for the 497 PPF treatment represents the time splice explained in section 2.2.4. The rate of CO_2 consumption was irradiance-dependent, but the CCP was irradiance-independent as indicated by the convergence of the two data sets after about 6 hours. The plateau portion of the plot approximates the CCP to be between 45 and 55 ppm which agrees with values reported previously for wheat and other C_3 plants (13). This range may be a slight overestimate of the CCP due to the long time required to ensure a complete drawdown. Such extreme drawdowns represent possible scenarios in a CELSS where atmospheric gas control systems may, on occasion, not be altered in response to significant changes in the proportions of

CO_2 evolved and utilized, or when resources (e.g. CO_2 supply, energy) become limiting and appropriate modifications must be made.

Additional CCP data obtained from complete drawdowns will need to be conducted under different temperature regimes and perhaps lower oxygen concentrations to broaden this database since CCP has been reported to be temperature- and oxygen- dependent (13,14,15,16). While decreased oxygen conc for C_3 plants such as wheat should improve photosynthetic efficiency by decreasing photorespiratory carbon loss, it would probably not be feasible to lower the partial pressure of oxygen in a CELSS with humans below about 0.15 atm, unless there was sufficient gas tight partitioning of human and plant atmospheres.

3-5 EFFECT OF IRRADIANCE ON NP

Carbon dioxide exchange rate was linearly related to PPF up to the approximately 750 $\mu mol/m^2/s$ limit used during the 4 separate irradiance experiments. An example of the results of one (day 41) of these experiments is presented in Figure 3-7. Irradiance levels used in this study were well below the light saturation levels suggested by growth responses measured at higher irradiances by Bugbee and Salisbury (9) and as summarized by Larcher (17).

Following canopy closure (about 20 days), there was an increase in the light compensation points (LCP) of the whole wheat community (Table 3-2).

Table 3-2. Light compensation points (LCP) of a wheat community during growth and development. See 2.2.5 for experimental details and calculations.

DAYS AFTER PLANTING	LCP (µmol/chamber/s)
13	209
20	193
27	229
41	223

Since fractional light interception by the canopy did not increase after about 20 days (Table 3-1), it would be expected that light would eventually become limiting considering that irradiance was nonsaturating and that continued increases in biomass were taking place. In general, the whole community LCP values reported here are substantially higher (about 5-fold) than those reported for single flag leaves (11) or those assumed for wheat in growth experiments (9). It is expected that a community LCP would be higher than that measured for single leaves or plants exposed to a higher irradiance than the average irradiance recieved by all leaf surfaces in a dense canopy.

Additional determinations of LCP later in development will need to be made to determine if changes occur during grain fill and senescence.

3-6 EFFECTS OF TEMPERATURE ON NP AND DR

Net photosynthetic rates for the 6-week old canopy was lowest at the highest temperature (24 C), but decreased when the temperature was lowered from 20 C to 16 C (Figure 3-8), suggesting an optimum temperature between 16 and 20 C for NP. The rate of DR increased with increasing temperature. The net daily rates of carbon dioxide utilization were 28.6, 29.6, and 22.6 mol CO_2/chamber/day for 16, 20, and 24 C, respectively. At 16 C, the rate of DR is decreased more than the rate of NP relative to 20 C, which explains the small difference between the 2 temperatures in the daily rate of CO_2 utilization. Considering this minor difference in carbon-fixation, it may be more economical from the perspective of energy efficiency to use 16 C as a growth temperature for wheat in a CELSS. It will also be important to conduct similar temperature experiments at different stages of development to determine if the same trend is observed.

IV. CONCLUDING REMARKS

The biomass production chamber served as an excellent pre-prototype CELSS test chamber for making determinations of the rates of carbon dioxide exchange of a wheat community as influenced by environmental variables. When sealed, carbon dioxide exchange studies in the BPC may be conducted in a non-steady state mode whereby CO_2 injection is prevented, or in a quasi-steady state mode whereby CO_2 injections are made to maintain near constant concentrations. In the quasi-steady state mode, absolute quantities of CO_2 injected can be used to calculate assimilation rates. The non-steady state mode allows rapid determinations of CO_2 utilization and evolution rates following alterations in environmental variables.

The whole community gas exchange responses measured in the BPC are judged to be more representative of crop responses than studies involving single leaf or single plant measurements. Daily measurements of NP and DR enable constant monitoring of crop 'activity' and may furthermore be used to estimate crop biomass at any stage of development using straightforward analytical techniques. Similar gas exchange studies conducted in the BPC with other crops selected for CELSS should provide valuable databases. However, the constraints of one experimental unit and the time required for conducting these studies necessitates judicious selection of variables and treatments.

An additional avenue of research that should be pursued during the BPC crop evaluations is the acquisition of gas exchange data for an integrated system of plants and humans. Plant-human integration in closed systems (PHICS) will enable experimental human ratings of the test chamber and also will serve to cross verify measurements of gas exchange rates of each biotic component measured in isolation.

V. REFERENCES

1. Salisbury, F.B., B.G. Bugbee, and D.L. Bubenheim. 1987. Wheat production in controlled environments. Adv. Space Res. 7: 123-132.

2. Salisbury, F.B. and B.G. Bugbee. 1988. Space farming in the 21st century. 21st Century Science & Tech. 1 (1): 32-41.

3. Olson, R.L., M.W. Oleson, and T.J. Slavin. 1988. CELSS for advanced manned mission. HortScience 23: 275-286.

4. MacElroy, R.D. and J.Bredt. 1984. Current concepts and future direction in CELSS. In NASA Conf. Pub. 2378. Life Support Systems in Space Travel.

5. Prince, R.P., W.M. Knott, S.E. Hilding, and T.L. Mack. 1986. A 'breadboard' biomass production chamber for CELSS. Amer. Astronautical Society, 33rd Annual Meeting, Boulder, CO. AAS-86-338.

6. Prince, R.P., W.M. Knott, J.C. Sager, and S.E. Hilding. 1987. Design and performance of the KSC Biomass Production Chamber. Soc. Auto. Engin. Tech. Paper Ser. No. 871437.

7. Sager, J.C., C.R. Hargrove, R.P. Prince, and W.M. Knott. 1988. CELSS atmospheric control system. Amer. Soc. Agric. Eng. Paper No. 88-4018.

8. Tibbitts, T.W., D.C. Morgan, and I.J. Warringon. 1983. Growth of lettuce, spinach, mustard, and wheat plants under four combinations of high-pressure sodium, metal halide, and tungsten halogen lamps at equal PPFD. J. Amer. Soc. Hort. Sci. 108: 622-630.

9. Bugbee, Bruce G. and Frank B. Salisbury. 1988. Exploring the limits of crop productivity. Plant Physiol. 88: 869-878.

10. Hoagland, D.R. and D.I. Arnon. 1950. The water culture method for growing plants without soil. Calif. Agr. Expt. Sta. Circ. 347.

11. Araus, J.L., L. Tapia. 1987. Photosynthetic gas exchange characteristics of flag-leaf blades during grain filling. Plant Physiol. 85: 667-673.

12. Holt, D.A., R.J. Bula, G.E. Miles, M.M. Schreiber, and R.M. Peart. 1975. Environmental physiology, modeling and simulation of alfalfa growth. I. Conceptual development of SIMED. Purdue University, West Lafayette, IN, Agricultural Experiment Station Research Bulletin 907.

13. Bykov, O.D., V.A. Koshkin, and J. Catsky. 1981. Carbon dioxide compensation concentration of C_3 and C_4 plants: dependence on temperature. Photosynthetica 15: 114-121.

14. Smith, E.W., N.E. Tolbert, and H.S. Ku. 1976. Variables affecting the CO_2 compensation point. Plant Physiol. 58: 143-146.

15. Dvorak, J. and L. Natr. 1971. Carbon dioxide compensation points of Triticum and Aegilops species. Photosynthetica 5: 1-5.

16. Gerbaud, A. and M. Andre. 1980. Effect of CO_2, O_2, and light on photosynthesis and photorespiration in wheat. Plant Physiol. 66: 1032-1036.

17. Larcher, W. 1980. In: Physiological Plant Ecology. 2nd. ed., Ch. 3: Carbon utilization and dry matter production. 303 pp.

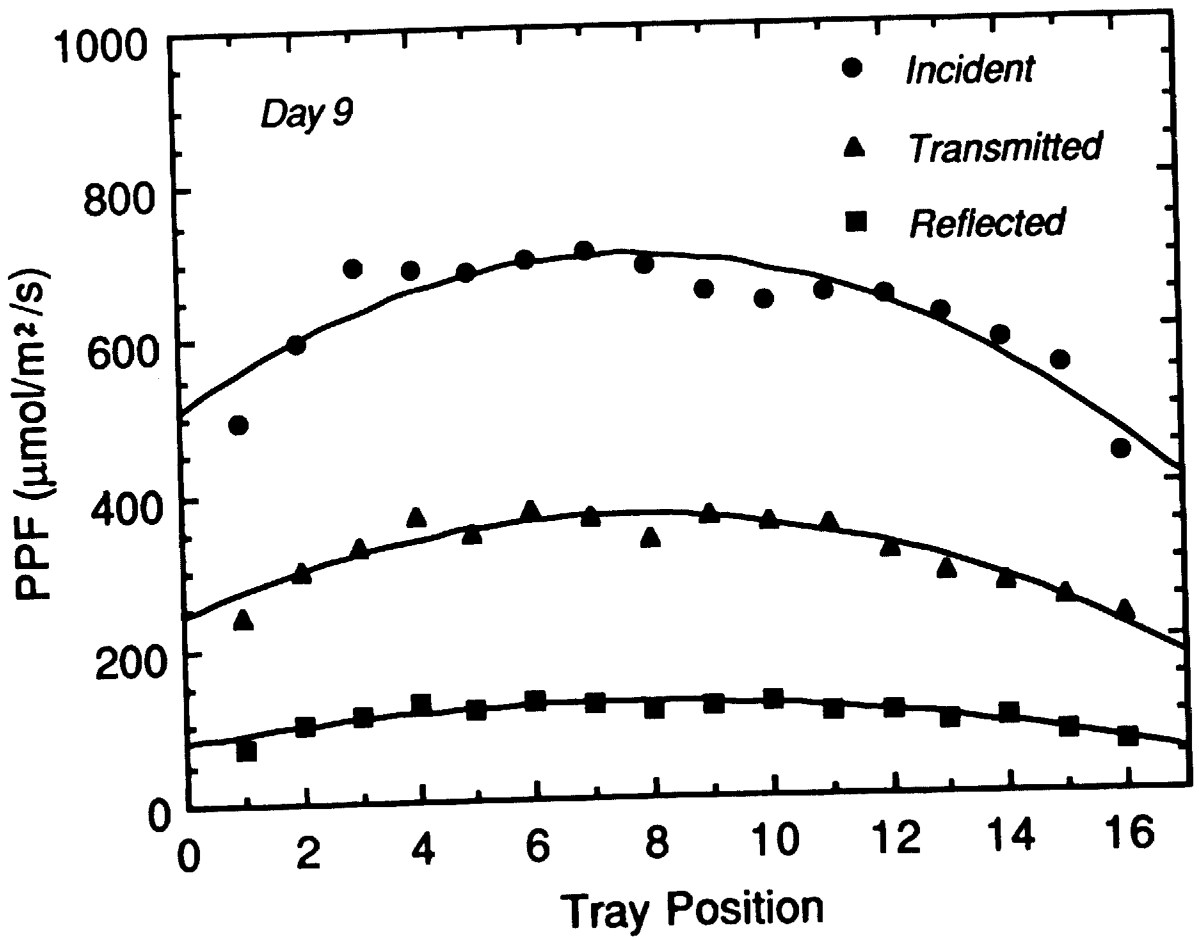


Figure 3-1. Irradiance levels under HPS lamps at various locations in the BPC.

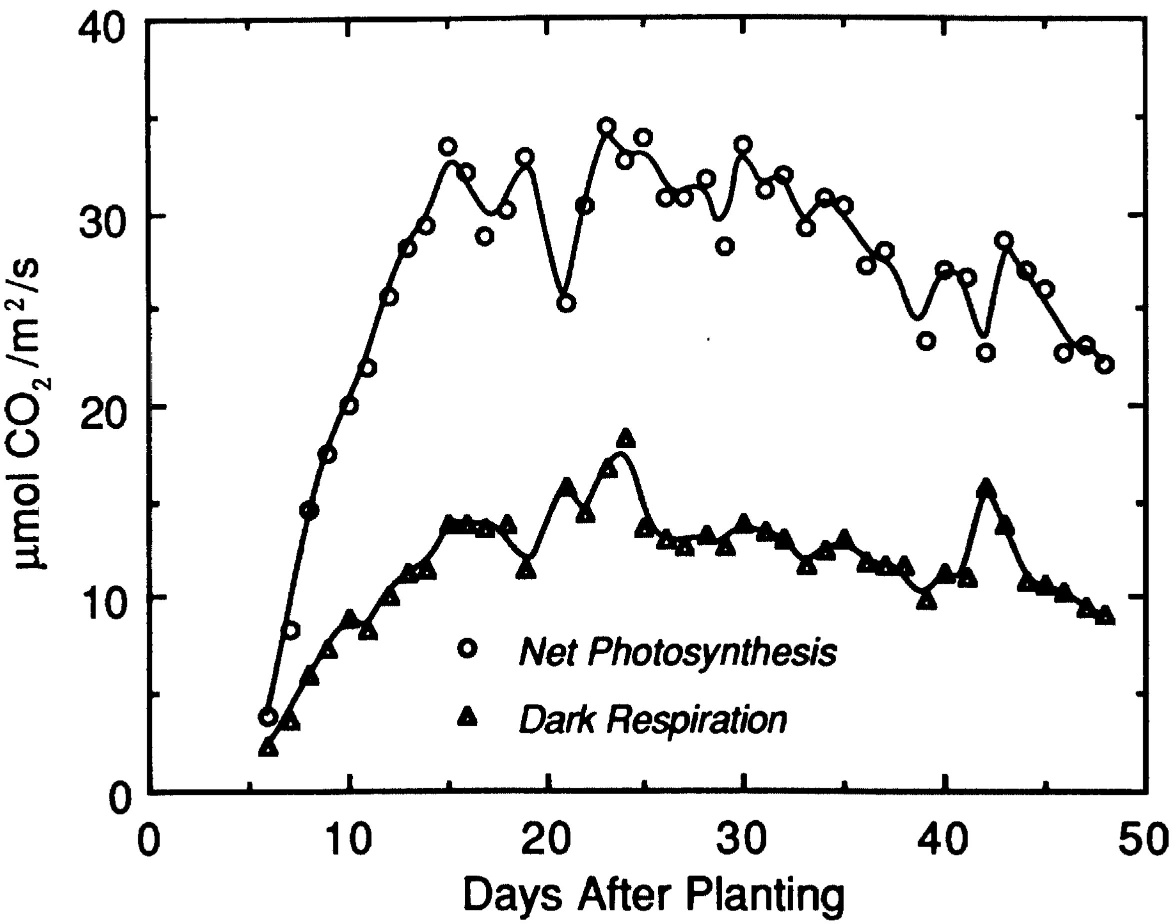


Figure 3-2. Rates of NP and DR during growth of a wheat community.

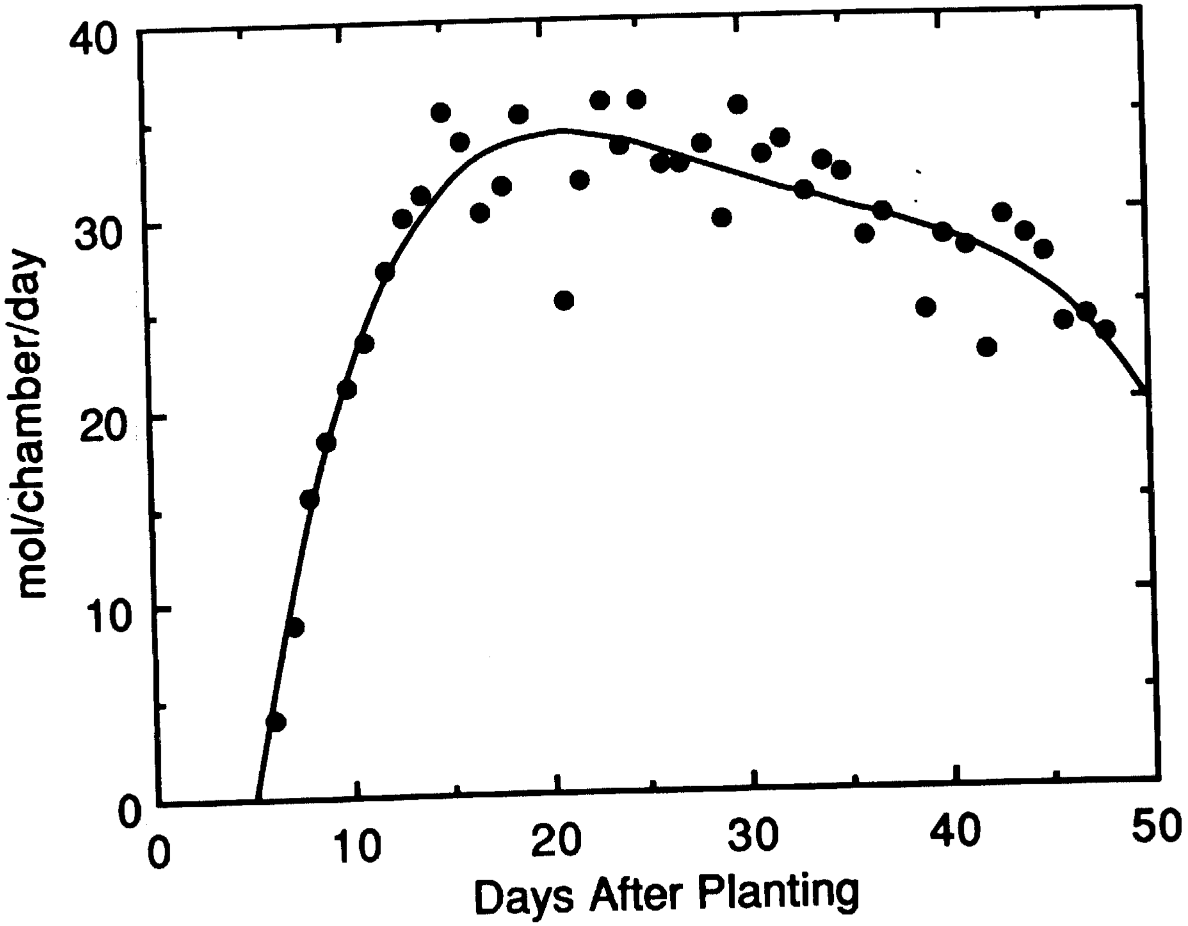


Figure 3-3. Net daily carbon dioxide utilization during growth of a wheat community.

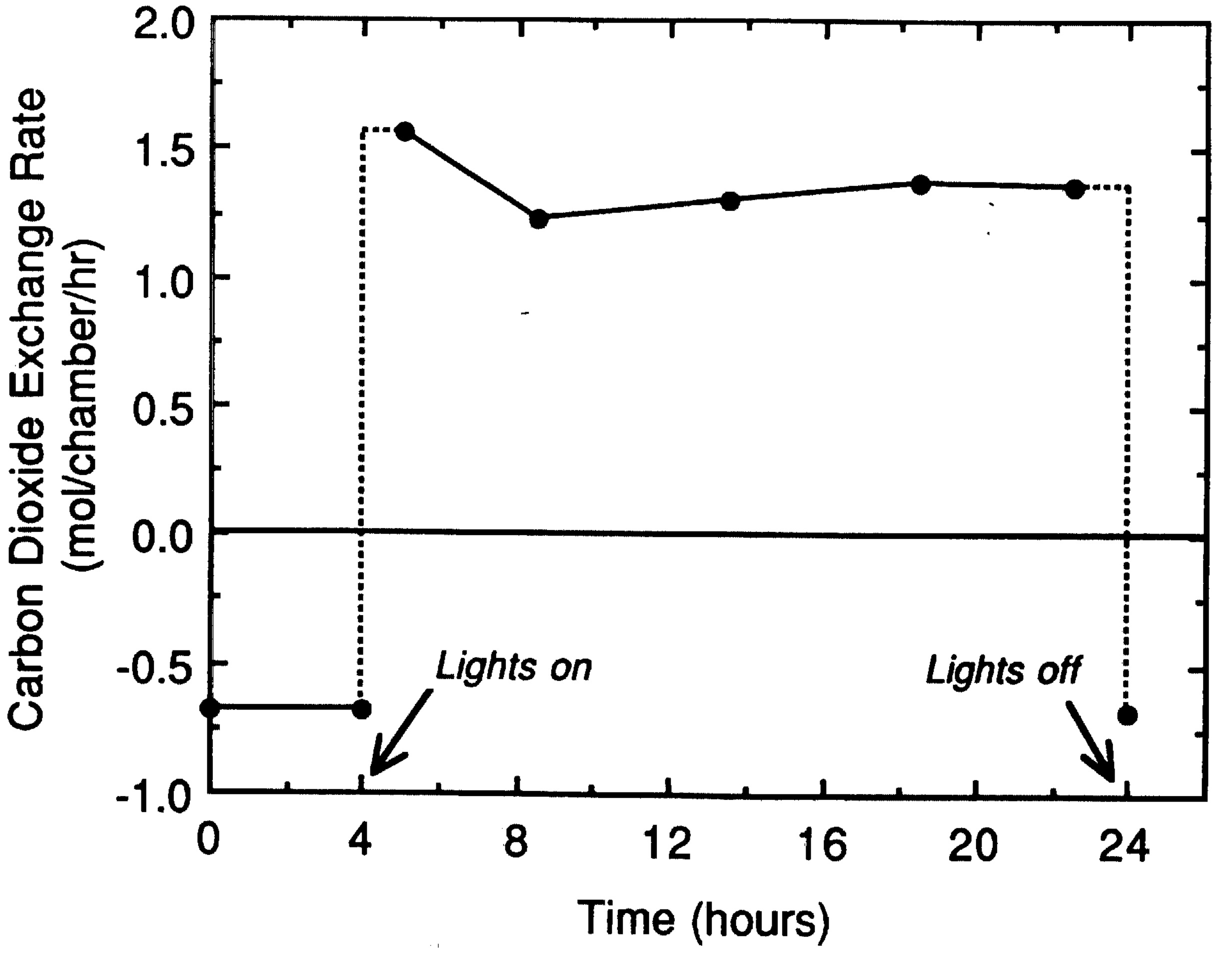


Figure 3-4. Daily pattern of carbon dioxide exchange rates.

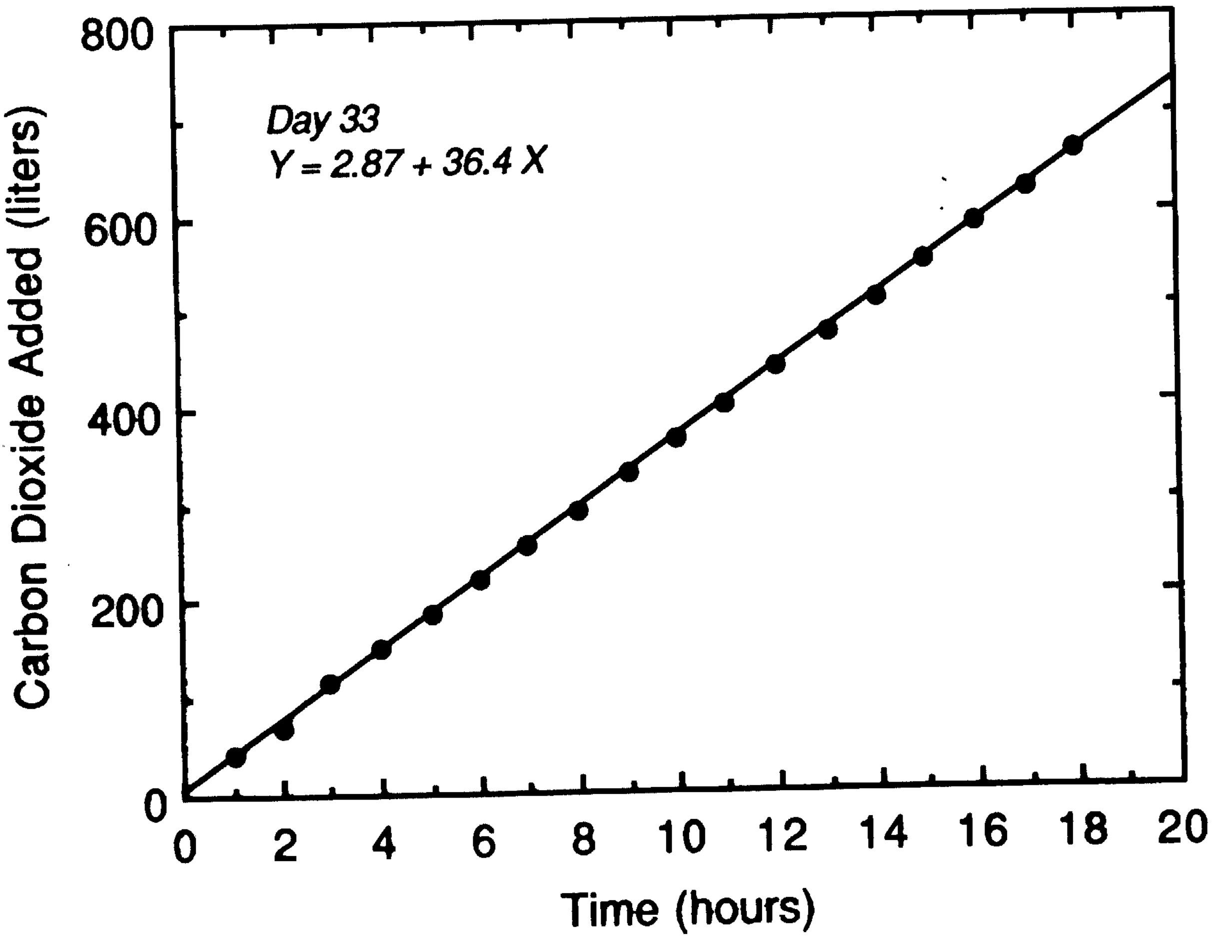


Figure 3-5. Cumulative additions of carbon dioxide to the BPC atmosphere during the photoperiod.

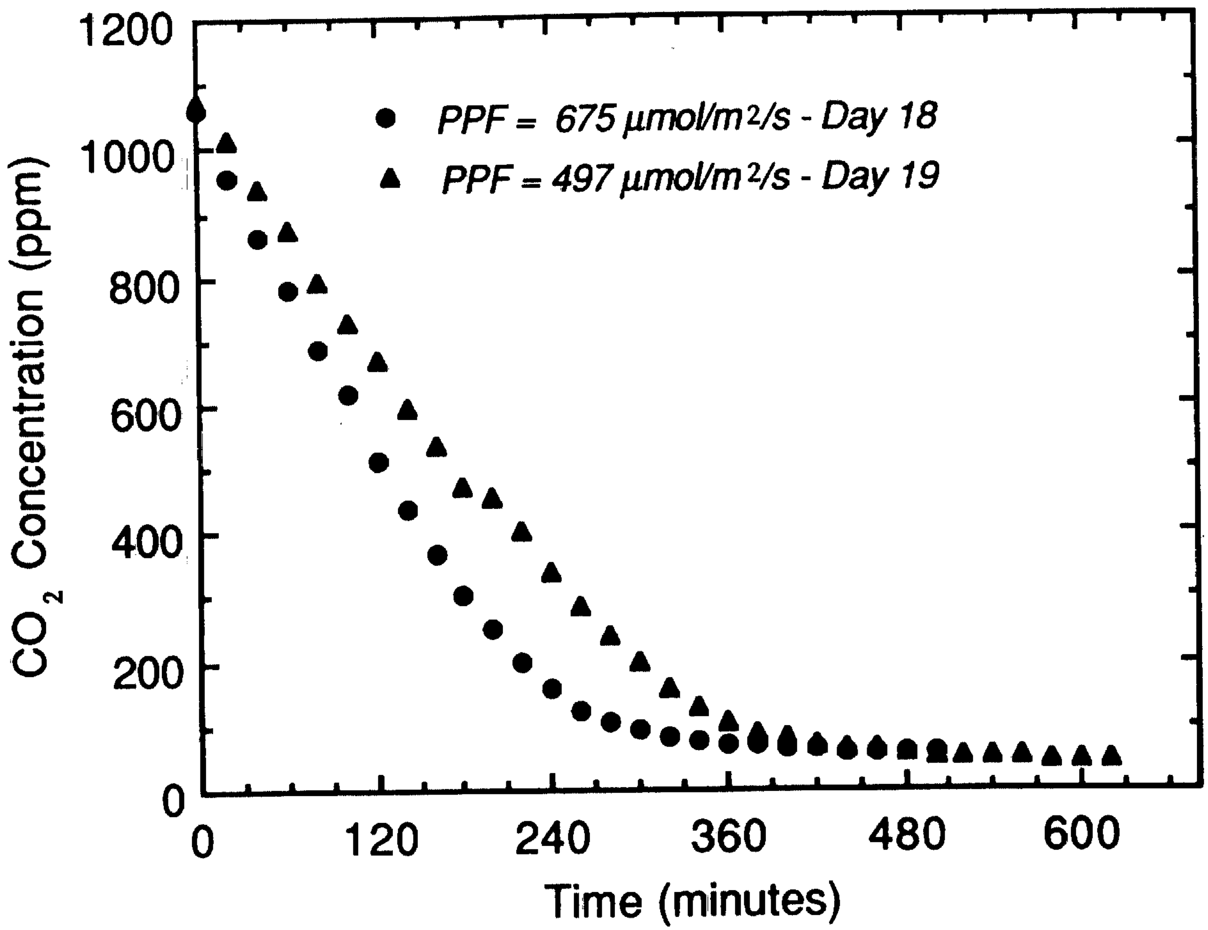


Figure 3-6. Photosynthetic drawdowns of CO_2 by a sealed wheat community at 2 irradiance levels.

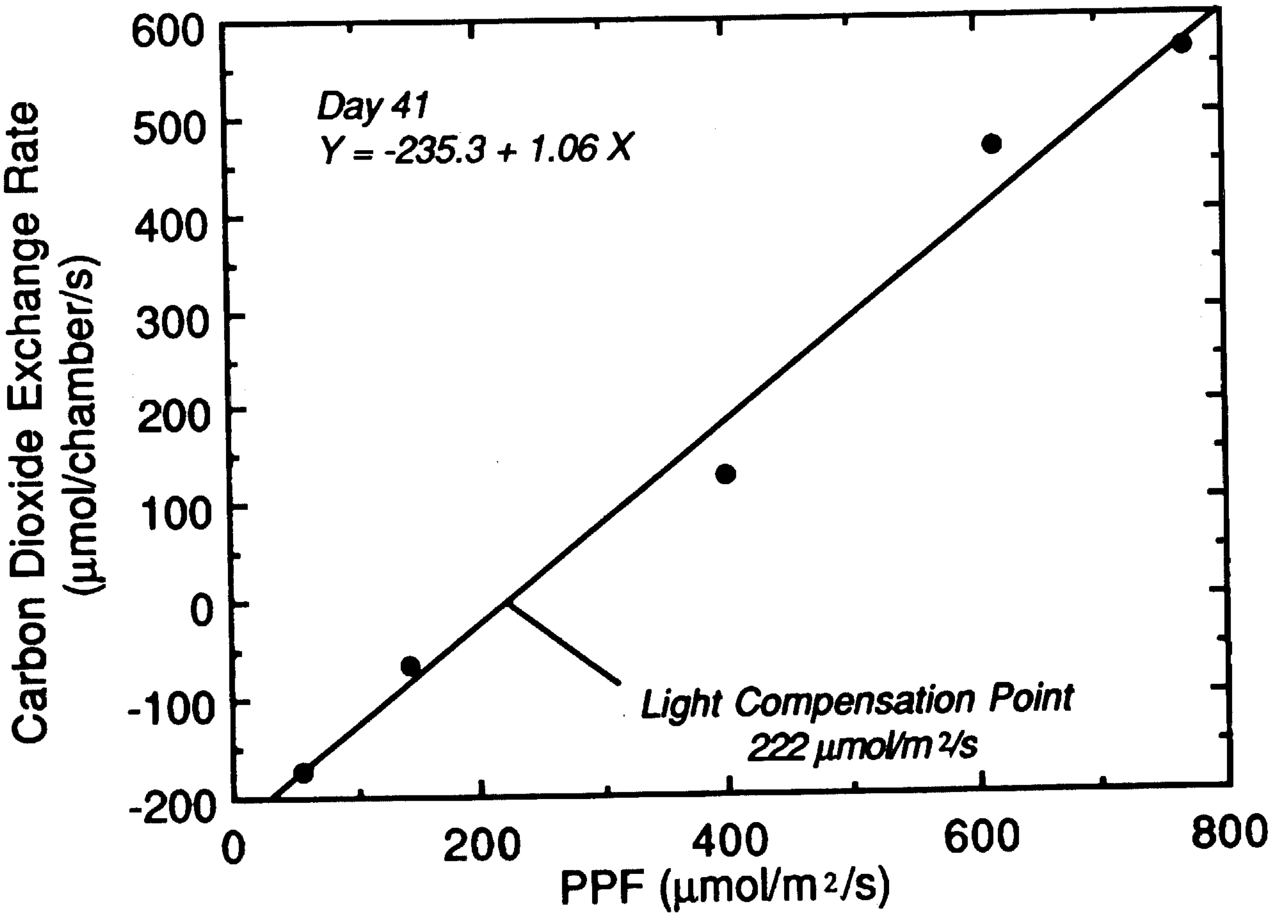


Figure 3-7. Relationship of carbon dioxide exchange rate with irradiance.

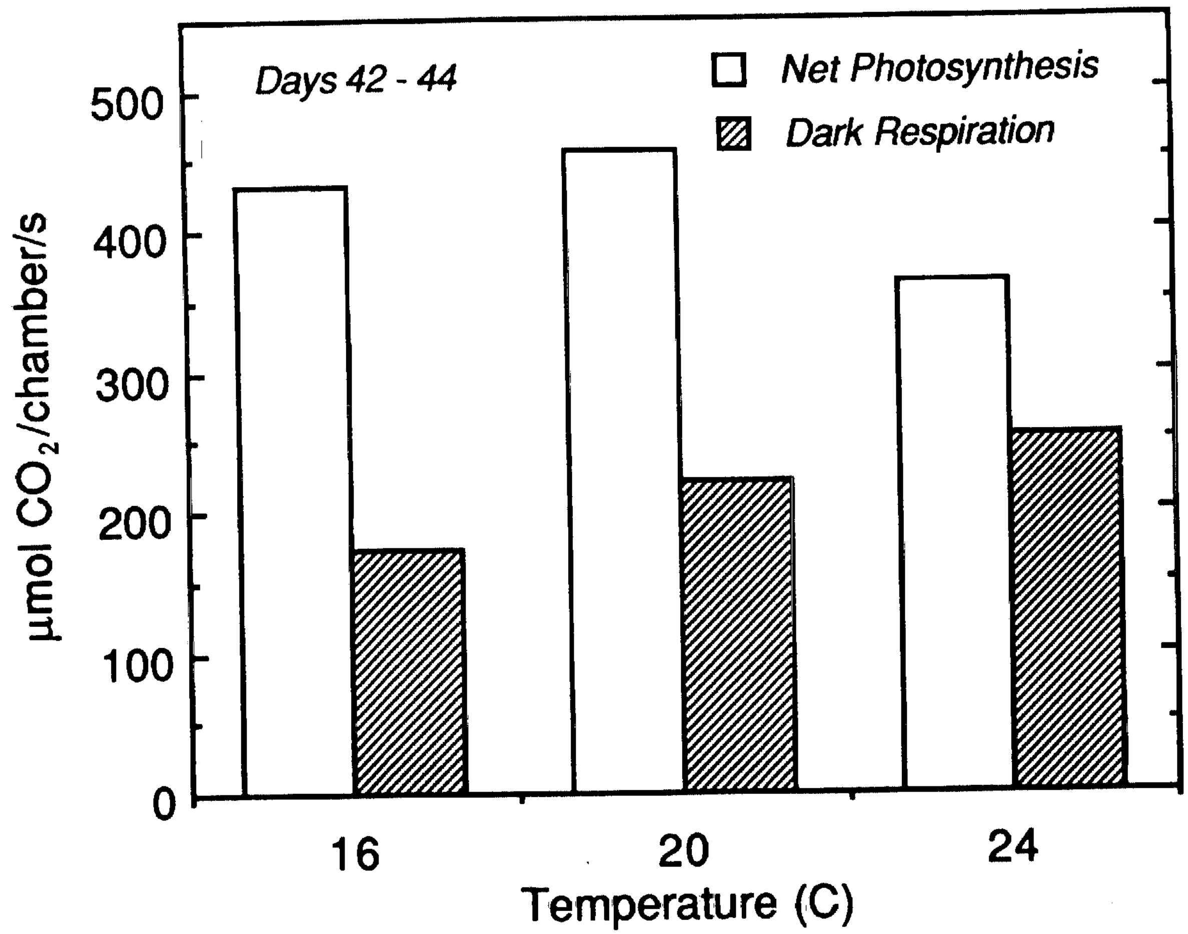


Figure 3-8. Effects of temperature on rates of net photosynthesis and dark respiration.

N90-16690

1989 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

ROCKET-TRIGGERED LIGHTNING STRIKES AND FOREST FIRE IGNITION

PREPARED BY:	Dr. James H. Fenner
ACADEMIC RANK:	Assistant Professor
UNIVERSITY AND DEPARTMENT:	Auburn University Aerospace Engineering Department
NASA/KSC	
DIVISION:	Electronic Systems
BRANCH:	Ground Support Equipment
NASA COLLEAGUE:	Mr. Bill Jafferis
DATE:	August 4, 1989
CONTRACT NUMBER:	University of Central Florida NASA-NGT-60002 Supplement: 2

Abstract

The report presents background information on the rocket-triggered lightning project at Kennedy Space Center (KSC), a summary of the forecasting problem there, the facilities and equipment available for undertaking field experiments at KSC, previous research activity performed, a description of the atmospheric science field laboratory near Mosquito Lagoon on the KSC complex, methods of data acquisition, and present results. New sources of data for the 1989 field experiment include measuring the electric field in the lower few thousand feet of the atmosphere by suspending field measuring devices below a tethered balloon. The report also details problems encountered during the 1989 field experiment, and lists future prospects for both triggered lightning and lightning-kindled forest fire research at KSC.

Summary

Kennedy Space Center (KSC) is the center for and its operations the focus of the world's most exacting single-point, short-range weather forecasting problems. Thunderstorms, with lightning, hail, strong winds, and possibly tornadoes, represent the greatest hazard at KSC.

The present Atmospheric Science Research Laboratory program at KSC includes ground and airborne electric field measuring instruments (field mills); a ground-based radar; numerical models; rocket triggered lightning experiments; and conventional, fairly dense network of reporting stations and rain gages. When available, KSC will add a high-resolution wind profiler now being developed at Marshall Space Flight Center.

KSC recognizes the critical nature of smaller scale weather phenomenon in the forecasting problem addressed, i.e. short-period, precise, local weather forecasts. No other group has ever attempted to forecast on a routine basis the weather events KSC desires to predict. KSC will first attempt to improve the general understanding of smaller scale weather phenomena. The research project coordinates actions of disparate groups in collecting and analyzing heterogeneous data, and in integrating results into a real-time data display system.

Some problem areas: Most of the individual research efforts by the various participating groups take place without coordination with either KSC or the other cooperating groups. KSC is trying to integrate the research program into a unified effort. Devising a reliable operational forecasting method may take many years and considerable effort from KSC, other government weather-forecasting units, and academia.

Work on lightning-kindled forest fires has just begun at KSC, and will continue.

Table of Contents

I

INTRODUCTION

1.1 BACKGROUND INFORMATION

The lightning program at Kennedy Space Center (KSC) began in the 1960's, when the National Aeronautics and Space Administration (NASA) began building taller structures on the Center. Lightning strikes to Apollo 12 and, more recently, to an Atlas-Centaur rocket, resulting in both its and its payload's destruction, gave rise to added research to understand lightning better. KSC used their own employees, as well as cooperated with academic institutions and private companies in developing its lightning program. The present program includes ground and airborne electric field measuring instruments (field mills), radar; numerical models, rocket triggered lightning experiments, and conventional mesometeorological network of reporting stations and raingages.

Most of the individual research efforts by the various participating groups take place without coordination with either KSC or the other cooperating groups. KSC needs to integrate the entire research program into a unified program. Moreover, KSC needs to use the results and techniques developed for its day-to-day operations. Additionally, the limited meteorological expertise at KSC has hampered the research effort, requiring KSC to rely heavily on outside personnel and equipment for this research.

Numerous disparate groups and organizations have some expertise in various aspects of thunderstorm and lightning phenomena. Railroads know about lightning's ability to travel long distances along rail tracks, and to cause damage far from the original strike. Electric power companies also know how lightning travels through its conductors to damage equipment far from the thunderstorm producing the lightning. They also know lightning can couple into lines not originally struck by lightning. Airlines and the military know lightning strikes aircraft both in the air and on the ground, and that aircraft can trigger lightning flashes even far from a thunderstorm cloud. Radio and television stations, as well as telephone companies know lightning strikes their towers and disrupts their transmissions and communications. It also couples into their equipment. Boaters, anglers, and golfers, among others, know their recreational equipment (rods, masts, golf clubs) may serve as conductors for lightning strikes--particularly newer graphite materials in rods, masts, and club shafts.

1.2 THE LIGHTNING FORECASTING PROBLEM AT KENNEDY SPACE CENTER

KSC is the center for and its operations the focus of the world's most critical single-point, short-range weather forecasting problems. Many operations at KSC are extremely vulnerable to weather, usually in such novel ways that the forecasting problem has no counterpart in any other realm. The forecaster must develop their own experience at KSC, they cannot rely on experience gained elsewhere to help them with unique KSC forecasting problems.

The special nature of weather at KSC, as well as extremely high economic and human costs if KSC launches (and therefore missions) fail leads to very precise forecasting criteria with extremely little margin for error. KSC success or failure also impacts directly and significantly on national and international opinion of United States' space effort and expertise. KSC failures draw considerable national and international attention! Thunderstorms, with lightning, hail, strong winds, and possibly tornadoes, represent the greatest hazard at KSC.

In its approach to forecasting extreme weather, KSC recognizes the critical nature of smaller scale weather features and phenomenon (mesoscale components) on the problem addressed: short-period, precise, local weather forecasts. Even the excellent world-wide weather data available through MIDDS cannot by itself make the local weather forecasting problem easier. KSC plans to integrate weather data from satellites, radar, its own local mesonetwork of weather stations, regional weather stations, and data on local lightning strike into the forecasting technique. When available, KSC will add a high-resolution wind profiler (now being developed at NASA Marshall Space Flight Center [MSFC]).

Other data for the weather forecasting scheme envisioned include dual-doppler radar, NEXTRAD at Melbourne, Florida; ground- and airborne electric field measurements from KSC-operated sites; and local lightning-locating data. If at all possible, KSC envisions using its rocket-triggered lightning data into an operational forecasting technique. Since no one, to our knowledge, has ever attempted to forecast on a routine basis the weather events KSC desires to predict, we can only describe the forecasting as experimental. Devising a reliable operational forecasting method may take many years and considerable effort from KSC, other government weather-forecasting units, and academia.

The approach KSC will take will first attempt to improve our understanding of smaller scale (mesoscale) weather phenomena. Only when we obtain an adequate knowledge of the systems we wish to forecast can we confidently try to predict that phenomenon. This approach requires, however, the close coordination and cooperation of disparate, heterogeneous data, and its integration into a real-time (preferably interactive) data display system. The forecasting problem will also almost require such a technique, because KSC must forecast weather events lasting less than one minute, thus requiring almost instantaneous data collection and display. This requirement may not be unique (airports would also like to have this capability), but the economic and political costs of delays and wrong decisions at KSC are much, much higher than anywhere else.

1.3 FACILITIES AND EQUIPMENT

1.3.1 ATMOSPHERIC SCIENCE FIELD LABORATORY (ASFL). Kennedy Space Center (KSC) lies in a region of the United States with one of the highest

C-2

frequencies of thunderstorms and lightning activity. **Figure 1** shows the location of the ASFL and other sites used for the RTLP. KSC operations involve some stupendously expensive equipment (the Shuttle, satellites, and launch vehicles) subject to critical and exacting time schedules. At the same time, launch support equipment such as towers, antennae, above ground and buried cable, are subject to damage or interruption of their function and use from lightning strikes. This combination of conditions make lightning both a hazard to and a significant factor in success or failure of KSC operations. As a result, KSC has been involved in and conducted extensive lightning studies for more than two decades. These studies involved characterizing lightning flashes, devising methods of protecting equipment from lightning strikes, and ways to locate and predict lightning and thunderstorms. Since the early 1980's, KSC, in conjunction with other Government organizations, private companies, and universities, has intensified its studies of thunderstorm and lightning phenomena.

KSC and the Eastern Space Missile Center (ESMC) weather group delve into thunderstorm and lightning forecasting, as well as devising methods of predicting other significant, adverse, or severe weather events (e.g. freezing precipitation, for, icing, or strong or gusty winds). The combination of KSC and ESMC have developed one of the finest facilities for forecasting short-range weather events. The KSC/-ESMC facilities include weather satellite and radar data, a mesoscale weather observation network (more than fifty stations), and the Meteorological Interactive Data Display System (MIDDS) which supplies world-wide meteorological data and soundings. The KSC also uses a tethered balloon for research on thunderstorms and lightning, and may be able to include this in future forecasting techniques.

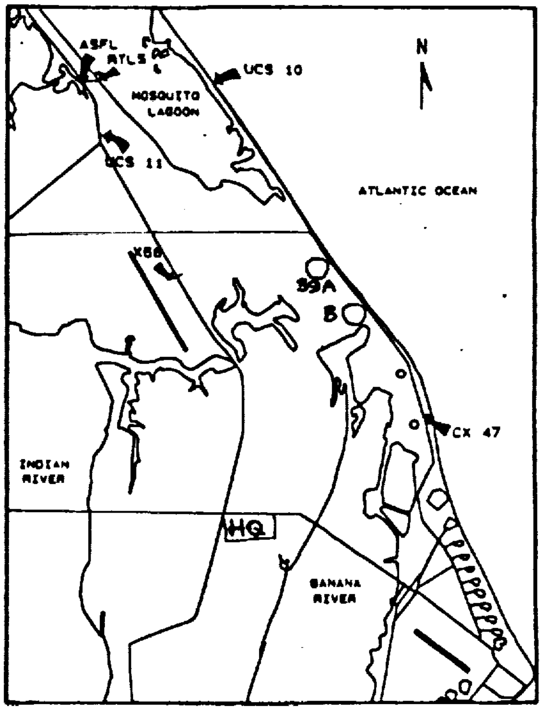


FIGURE 1-1 SITE LOCATIONS

KSC and ESMC wish to create and operate an advanced weather support and forecasting system in order to reduce weather-related hindrances to KSC operations. The KSC program also plans to transfer to other weather forecasting units (such as the US Air Force or the National Weather Service) the technology and knowledge gained through this research.

The Federal Aviation Authority (FAA), Air Force Wright Aeronautical Laboratories (AFWAL), and the US Naval Research Laboratory (NRL) are among the Government groups interested in lightning studies at KSC. KSC and other groups are interested in (1) characterization of lightning hazards to KSC operations, to communications, to power distribution, and to command and control systems; (2) remote lightning detection; and (3) understanding the "advent and demise" of thunderstorms. In addition, certain groups within the Government are interested in using lightning strikes to simulate the electromagnetic pulse (EMP) nuclear weapon bursts might send out.

After learning more about lighting and its effects on air- and spacecraft, KSC would like to transfer the techniques and knowledge gained from its studies to operational forecasting and to academic institutions training weather forecasters. This should ensure qualified forecasters for future operations.

1.4 DATA ACQUISITION

Items investigated in KSC lightning and thunderstorm studies include static and field charges using electric field measurements in and around KSC; locating and counting lightning discharges (cloud-to-ground strikes, mainly) within 200 miles of KSC; radar data from the KSC region; surface wind data using a mesoscale network of measuring stations within about 50 miles of KSC; electric and magnetic fields and lightning current measurements from the KSC area; and other meteorological data obtained from local, regional, and national sources. These data will, hopefully, be integrated into a forecasting method and applied to improving short-term weather forecasting and verification of numerical weather forecasting, and to evaluating lightning warning procedures.

The Maxwell current and its changes with time may help researchers understand when thunderstorms begin ("turn on") and when they quit ("turn off"). Maxwell current may thus ultimately lead to an approximate threshold for impending lightning strikes. (Lightning and its accompanying thunder define a thunderstorm; without these two phenomena, the event is merely a rain- or hailshower.)

Photo analysis of lightning by the State University of New York at Albany (SUNYA) may be used to quantify several parameters, such as size and shape of strokes. Streak images yield stroke propagation speed.

The program at KSC is the first program to measure all parameters (electric and magnetic field, current, electron temperature in the lightning plasma, luminosity, spatial orientation, and stroke propagation speed) at the same time, thus allowing case studies to test theoretical and numerical models of lightning behavior.

Photography on calibrated film can determine flash luminosity. If luminosity is a function of current, then we can measure lightning current directly. Further, time resolved lightning spectra would then yield electron temperature in the lightning channel. Photographic images can be analyzed by video densitometers, if digitized, or conventional densitometers if not.

Kennedy Space Center (KSC) receives wind observations from a network of over 50 instrumented towers covering an area about 53 by 57 km (about 1600 square kilometers) as shown in figure1-2. Most wind instruments have been mounted on top of standard 54-foot tall telephone poles, and set to record wind data at five-minute intervals.

Three direction-finding stations locate negative lightning flashes (where earth is positive relative to cloud). Lopez and Holle (1986) describe the lightning direction-finding method.

A United States Air Force WSR-74C radar located at Patrick AFB, approximately 30 km SSE of KSC, supplies data at five-minute intervals. A Weather Bureau radar at Daytona Beach, about 100 km NNW of KSC, also supplies radar imagery at irregular intervals. A Lightning Location and Protection, Inc, (LLP) Integrated Storm Information System (ISIS) records negative lightning flash information as well as Daytona Beach radar information. This ISIS equipment is currently located at the US Fish and Wildlife headquarters on KSC property, but KSC plans to move it to their Range Control building on Cape Canaveral Air Force Station during July.

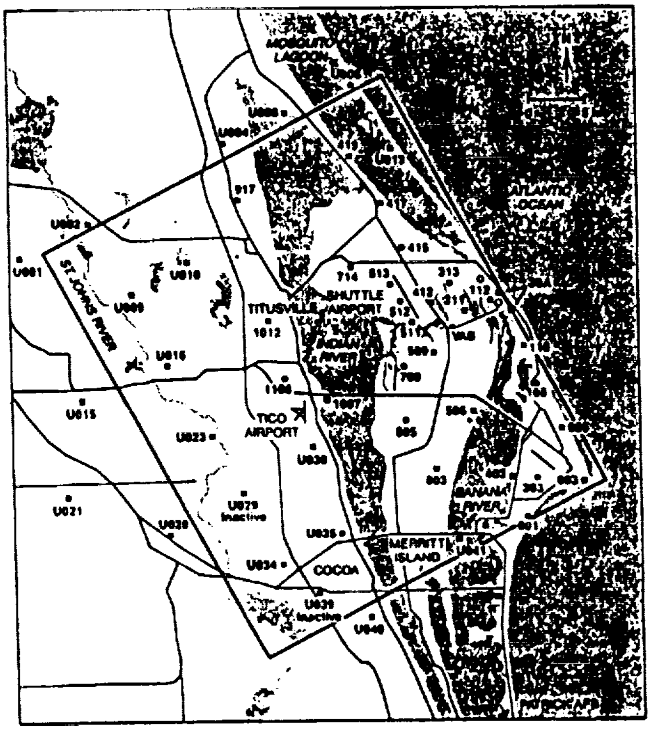


Fig.1-2 Kennedy Space Center and Cape Canaveral Air Force Station area. The rectangle marks the 1989 mesoscale network. Solid squares indicate meteorological stations. Patrick AFB (lower right) is the site of the USAF WSR-74C weather radar. (After Watson, et al.)

Lightning triggered when small rockets trailing a conductive wire behind them (rocket-triggered lightning, RTL) are launched near active thunderstorms provide several advantages for scientific study of lightning. First, lightning occurs at a pre-defined place and at a pre-determined time. This allows researchers to measure parameters seldom--and then only with extreme difficulty--measured in the natural atmosphere. Secondly, it allows a detailed look at the very long "leader" strikes propagating into un-ionized air, close to the conditions prevailing in an unmodified environment. Both of these advantages help researchers understanding lightning leaders, thus understand lightning itself better--and, more importantly, that triggered by aerospace vehicles traversing that region of the atmosphere.

Suspending an isolated metallic object (a cylinder about eight feet long and two feet in diameter) below a tethered balloon as a lightning strike object (LSO) may also simulate an aerospace vehicle-triggered lightning (ATL) object. Leaders observed and measured during such strikes will provide data for comparison with prior observations, hopefully to verify or refute the bi-directional ATL model commonly proposed. The series of field mills suspended along the tether cable provides electric field measurements around the LSO. This experimental set-up also allows negative leader current to be measured at the LSO site, possibly permitting return-stroke current measurements at ground and higher levels at the same time. Streak cameras and conventional photography record visual imagery for later quantitative study.

Data taken both over land and over water allows similarities and differences to be observed and measured. Rocket launches over water represent a "purer" electric lightning signature, since there is no distortion of the signal from the ground or support equipment around the launch pad. The 1989 RTLP includes launches from land and water RTLS; my proposal is to launch from each site alternately, or from each at short intervals, i.e. quasi-simultaneously.

Other sensors include microphones to record the sound of thunder; and current sensors in the ground (correlated with negative cloud-ground potential.)

II

PRESENT RESULTS

2.1 NEW METHODS OF DATA ACQUISITION

The 1989 rocket-triggered lightning strike research season envisions recording positive lightning strike data as well as negative. The State University of New York at Albany (SUNYA) plan to operate a system using satellite data to provide actual lightning strike location data.

KSC will also attach electric field measuring instruments ("Field mills") at intervals along the cable attached to a tethered balloon located near the rocket launch site. This will supply a vertical sounding of electric field strength near the triggering site, important data presently missing. Six field mills located along the tethering cable will supply field strength at heights of approximately 300, 400, 500, 600, 700, and 800 m above the ground.

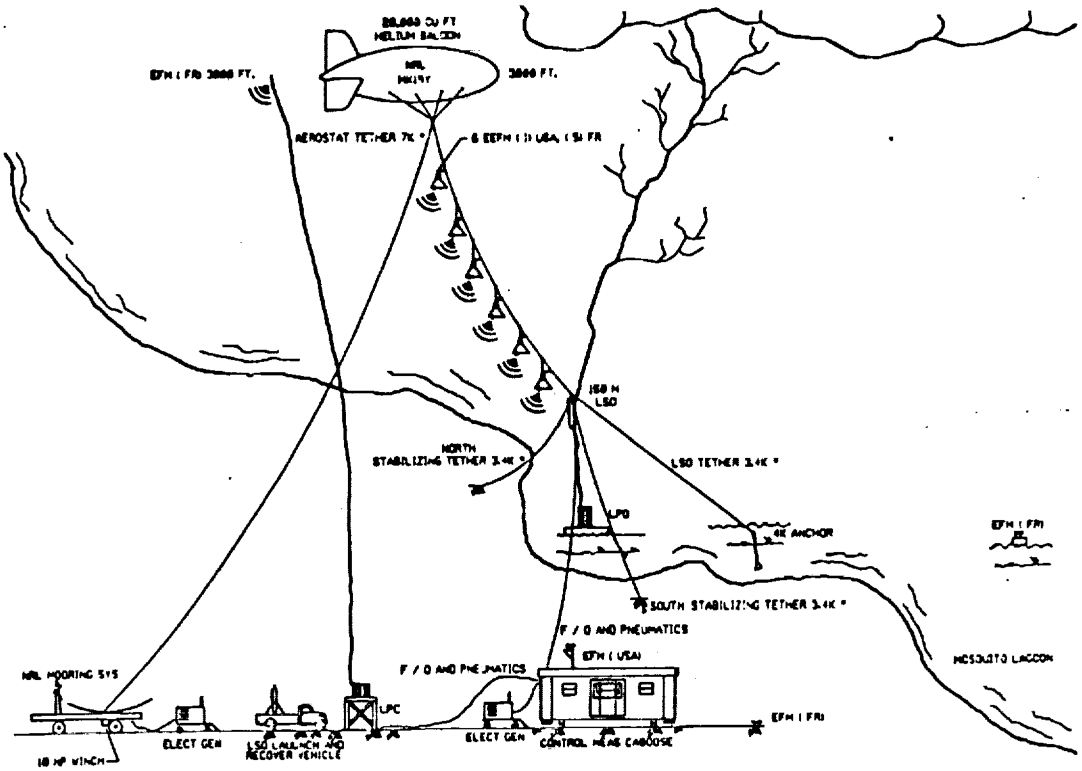


Fig. 2-1 1989 RTLP

2.2 RESULTS FROM 1989 FIELD EXPERIMENT

2.2.1 PROBLEMS WITH EQUIPMENT. Prior to beginning the 1989 field experiment at the RTLS and ASFL, KSC encountered several unexpected problems. These problems delayed the field experiments; consequently, the RTLP gathered little data prior to my departure. This report thus presents few results, and those only preliminary.

The 1988 field experiment ended August 31, 1988. From that date until just before my arrival June 12, 1989, the ASFL and RTLS buildings, as well as all equipment, sat idle. Preliminary inspection by my KSC colleague, Mr. Bill Jafferis, indicated some equipment needed service and calibration. Unfortunately, before KSC could accomplish this calibration and service, the fire marshal and building inspectors noticed a number of safety discrepancies. These discrepancies required immediate repair; they also precluded service to and calibration of equipment housed in those buildings until the buildings themselves had been fixed.

Adding to the delayed caused by building troubles, the tethered balloon leaked. It, too, required repair. However, since the much larger balloon, "Fat Albert," came down for patching the week I arrived, its repair delayed all repairs to the RTLP balloon. Further, one tether cable apparently needed to be replaced; a new cable required about five weeks for delivery. As of August 1st, the tethered balloon has not been flown. As a result, no airborne field mills supply data, because they have not yet been launched.

Another problem surfaced on June 25th, when the thunderstorm and lightning display system, ISIS, ceased receiving radar data from the Daytona Beach radar. Replacing the modem and other attempts to find and fix the problem failed. As of July 26th, ISIS is still not receiving a radar signal, even though the signal appears to be arriving at the site. The ISIS does record lightning strike location, which is useful, but lack of associated radar returns used to deduce the physical relationship between the lightning and the parent thunderstorm limit ISIS's value as a forecasting tool.

Since all attempts to find and correct the problems with ISIS failed, and we knew the telephone lines on the northern parts of KSC are not as good as those elsewhere, we decided to move the system. KSC personnel and I moved the ISIS from its previous location at the headquarters building of US Fish and Wildlife service, about ten miles north of the main KSC complex of buildings, to the Range Control building on Cape Canaveral Air Force Station. Unfortunately, trouble with the central computer in that building have thus far (Aug 1) prevented use of the ISIS at that location, too.

If these problems were insufficient to delay the start of the 1989 field experiment, others followed: The French researchers from Grenoble delayed their arrival by about ten days. Since they actually run the RTLS, launch the triggering rockets, and gather data, their delay pushed the start of operational rocket launches back still more. Moreover, even after they arrived in the middle of July, their equipment failed to arrive for another several days.

All RTLP equipment and personnel appeared to be ready for the first rocket launch only on July 21st. Unfortunately, for a number of days thereafter, weather conditions did not produce lightning at the launch site on Mosquito Lagoon. The first operational launch occurred on July 29th, but only one rocket was launched. The remainder of the day, we waited in vain for fields to build up to launch criteria (about 4500 V/m.)

2.2.2 SOME PRELIMINARY RESULTS FROM THE 1989 FIELD EXPERIMENT. Due to delays outlined above, coupled with my departure August 4th, this report gives only a few preliminary results from the 1989 field experiment.

Table 2-1, next page, gives a small sample of the lighting-strike data recorded and archived for the project. Table 2.1 gives azimuth (direction in degrees from true north) and range (in nautical miles) relative to the recording site. On this day, July 29, 1989 (Julian date 89210), about 1000 cloud-to-ground lightning strikes occurred within 25 nautical miles of the recording location on Cape Canaveral AFS. Table 2.1 lists only the first few strikes recorded. Time is given in hours, minutes, and seconds Universal Time (or Greenwich Mean Time).

Table 2-2, next page, provides a small sample of the wind data. Wind measuring equipment is generally mounted on top of telephone poles of varying lengths; thus, the heights range from about 45 to 65 ft, with a mean of 54. Some locations also record temperature and dew point in Kelvin, from which we derive relative humidity. Average wind speed and maximum wind gust recorded during the previous five minutes are given meters per second. Figure 1-2 shows the location of each station.

2.3 LIGHTNING-KINDLED FIRES IN FOREST PRODUCTS

The plan was to place piles of toothpicks, kindling sawn from 1/4-inch plywood, 1-by-1's, and 2-by-2's at the launch site to see which ones ignite. KSC also intends to measure the current in each strike. The idea of using well defined sizes of wood is to achieve reproducible results. Again, no results have been obtained due to paucity of triggered lightning.

Table 2-1. Sample Lightning Data for July 29, 1989

THE FOLLOWING INFORMATION DEPICTS LIGHTNING OCCURRING WITHIN
25NM OF THE RCC AND COMPLEX 1/2

DAY	TIME	AZIMUTH	RANGE (n.mi.)
89210	20405	235°	20.9005
89210	20539	233	22.6031
89210	20539	224	18.7438
89210	20753	235	21.5042
89210	20902	231	23.5604
89210	21057	237	20.9962
89210	43814	333	24.9781
89210	51134	330	22.1674
89210	60032	332	21.9072
89210	65848	339	15.3245
89210	71953	326	14.6215
89210	75912	329	20.9053
89210	183951	204	14.64
89210	184107	191	17.6217
89210	184159	206	18.0055
89210	184237	192	18.4356
89210	184301	191	16.9742
89210	184334	196	18.7138
89210	184458	190	17.8074
89210	185058	310	15.7288

Table 2-2. Sample Wind Data for July 29, 1989

				W	I	N	D		
Day	Time	Site	Ht	Dir	Speed	Gust	Temp	Dew Pt	RH
89210	235500	1	54	167	3.0	3.6	300.38		
89210	235500	3	54	213	4.1	4.6	300.38		
89210	235500	5	54	183	3.6	4.6	301.49	300.77	96.0
89210	235500	108	54	178	2.5	3.6	299.82		
89210	235500	110	54	188	3.0	4.6	300.93	296.54	77.0
89210	235500	112	54	190	2.5	3.6	299.27		
89210	235500	303	54	189	2.5	3.0	298.71		
89210	235500	311	54	203	3.6	4.1	298.71		
89210	235500	313	54	198	2.0	2.5	299.82	295.93	79.3
89210	235500	403	54	182	2.5	3.0	299.27		
89210	235500	412	54	199	2.0	2.5	298.71		
89210	235500	415	54	186	2.5	3.0	298.71		
89210	235500	506	54	195	3.0	3.6	299.82		
89210	235500	509	54	206	2.0	2.5	298.16		
89210	235500	714	54	165	3.0	3.0	299.27		
89210	235500	803	54	166	1.5	1.5	297.04		
89210	235500	419	65	7	1.0	4.1			
89210	235500	417	60	175	1.5	2.0			
89210	235500	1007	42	185	2.5	3.0			
89210	235500	709	44	185	1.5	2.0			
89210	235500	805	54	62	1.5	2.0			
89210	235500	917	45	0	0.0	0.0			
89210	235500	1012	48	173	0.5	1.0			
89210	235500	1108	47	32	1.0	10.8			
89210	235500	39	60	219	4.6	5.6	299.27		
89210	235500	39	60	202	4.1	4.6	299.27	294.93	77.0

III

CONCLUDING REMARKS

3.1 NEW DEVELOPMENTS AND PROPOSED METHODS FOR FORECASTING LIGHTNING

The current system for forecasting thunderstorm location and lightning strike location uses a composite technique including maximum radar reflectivity of thunderstorm cells; location, number, and frequency of negative lightning strikes; surface wind convergence; and surface wind pattern. (Surface wind pattern typically varies with stage of thunderstorm development. See Byers and Braham [1949] or Watson et al. [1989].)

As Shuttle, and other, launches become scheduled more frequently, KSC operations less tolerant of delays. Thus, KSC forecasters need to identify more low-risk launch window, requiring improved forecasting of weather events such as triggered lightning, wind shear, and turbulence with accuracy and timeliness unique to space programs. Measurements of electric fields, for example, have not yet been included in the forecasting process for triggered lightning. Moreover, many critical weather factors cannot even be measured directly; the forecasters infer their value from their relationship to other parameters they can measure.

Launch safety needs both accurate current weather data and forecasts for two hours or less. Observations limit the accuracy and quality of forecasts, particularly on this short-term forecasting, or nowcasting, time scale. KSC must improve its observations, including new instrumentation and measuring systems, to improve operational forecasts.

New instrumentation is no panacea, however. New instruments improve detection, not necessarily forecasting. Forecasting methods use the data available when those methods were developed. KSC needs to modify forecasting methods and techniques to include new data sources. Displays for lightning detection networks and new instruments to detect in-cloud and cloud-to-cloud lightning, for example, should be incorporated into KSC weather forecasting. Likewise, local weather analysis and forecasting techniques specific to KSC need to be developed. KSC should also develop an interactive, computer-aided weather decision-making system, and possibly numerical weather prediction models specific to KSC operations.

Local convergence of surface winds still induce thunderstorm formation at KSC. Byers and Rodebush (1948) and Byers and Braham (1949) suggested this cause, and many later experiments and studies supported them. The comparatively dense network of surface wind measurements at KSC allow use of local convergence for short-period forecasting. In particular, the forecaster must locate and follow the movements of the sea breeze, as it dominates all other convergence forces in and around KSC. A proposed, new prediction method (not yet completed) is to write computer programs to calculate and plot convergence over several sub-areas within the KSC research area (fig.I-2), and to locate lines and regions of convergence within the same area. Breaking the

KSC research area down into four or nine smaller regions, or sub-areas, should be adequate. Watson and Blanchard (1984) noted that smaller areas provide reasonably good predictions of thunderstorm development using average convergence data, but Watson et al. (1989) noted that larger ones do not, since the averaging process dilutes the convergence (large on a small scale) when averaged over an area as large as the KSC research area. My proposal to break the KSC area down into smaller units for automatic convergence computation would solve the apparent dilution problem.

Surface convergence does not, of course, take into account any dynamic processes occurring higher in the atmosphere. The MIDDS provides upper-level information. By writing programs to analyze and plot various combinations of data (the best combinations have yet to be determined), the forecaster should be able to predict at least the potential for triggered lightning. One thing the previous research at KSC has shown: Lightning appears to begin just after maximum convergence (averaged over a fairly small area), to peak before average divergence over the same area reaches a maximum, and to decrease as divergence decreases.

3.2 FUTURE PROSPECTS

From its start seven years ago on the shores of Merritt Island's Mosquito Lagoon, about eight miles north of KSC's Vehicle Assembly Building, NASA's Rocket-Triggered Lightning Program (RTLP) has developed progressively into a formidable research effort. NASA's desire to improve KSC lightning protection and lightning forecasting gave the RTLP emphasis. Each year adds new features to improve scientific knowledge. 1988 added a tethered balloon.

New elements added in 1989 included field mills suspended below the tethered balloon (at various heights above the ground), and attempts at quantifying lightning-initiated kindling of forest materials. Placing field mills at intervals between the ground and the height of the balloon (about 500 m) provides data on change of electric field strength with altitude, the better to help characterize lightning strike potential over land and water. Field mills detect and help locate lightning, as well as allow study of the electric field environment prior to lightning strikes. The series of field mills suspended below the tethered balloon provide a more complete view of weather conditions conducive to rocket- or aircraft-triggered lightning.

The future thrust should be in combining and assimilating the many diverse data sources into an integrated short-term predictive technique. One main thrust should lie in setting up an expert system or knowledge bank, a "forecaster's helper" along the lines of the artificial-intelligence based "doctor's associate" used by some physicians and in some hospitals. A second main effort, writing programs to analyze the myriad data sources (KSC local wind fields, electric fields, radar, and other data from MIDDS) automatically, should support development of an expert system. KSC apparently recognizes the fact that too little work has been done in integrating the excellent data,

IV

LITERATURE AND BIBLIOGRAPHY

Byers, H.R. and R.R. Braham, 1949: The Thunderstorm. U.S. Government Printing Office, Washington, D.C. 287pp.

Byers, H.R. and H.R. Rodebush, 1948: "Causes of thunderstorms of the Florida peninsula" J Meteor., Vol.5, pp.275-280.

Krider, E.P., R.C. Noggle, and M.A. Uman, 1976: "A gated, wideband magnetic direction finder for lightning return strokes" J. Appl. Meteor., Vol.15, pp.301-306.

Lopez, R.E. and R.L. Holle, 1986: "Diurnal and spatial variability of lightning in northeastern Colorado and central Florida during summer" Mon. Wea. Rev., Vol.114, pp.1288-1312.

Maier, M.W. and W. Jafferis, 1985: "Locating rocket-triggered lightning using the LLP lightning locating system at the NASA Kennedy Space Center" Paper presented at the 10th International Conference on Lightning and Static Electricity, Paris, France.

Pielka, R.A., 1974: "A three-dimensional numerical model of the sea breeze over south Florida" Mon. Wea. Rev., Vol.102, pp.115-139.

Watson, A.I. and D.O. Blanchard, 1984: "The relationship between total area divergence and convective precipitation in south Florida" Mon. Wea. Rev., Vol.112, pp.673-685.

Watson, A.I., R.L. Holle, R.E. Lopez, R. Ortiz, and J.R. Daugherty, 1989: "Use of the surface wind field as a predictor of thunderstorms and cloud-to-ground lightning at Kennedy Space Center" Paper to be presented at the 1989 International Conference on Lightning and Static Electricity, 26-28 Sep, Bath, England. 7pp.

Watson, A.I., R.E. Lopez, J.R. Daugherty, R. Ortiz, and R.L. Holle, 1989: "A composite study of Florida thunderstorms using radar, cloud-to-ground lightning, and surface winds" Paper presented at the 24th Conference on Radar Meteorology, Talahassee, FL, 27-31 Mar 1989. 4pp.

Watson, A.I., R.E. Lopez, R.L. Holle, and J.R. Daugherty, 1987: "The relationship of lightning to surface convergence at Kennedy Space Center: A preliminary study" Wea. Forecasting, Vol.2, pp.140-157.

Williams, E.R., 1985: "Large-scale charge separation in thunderclouds" J. Geophys. Res., Vol.90, pp.6013-6025.

Ulanski, S.L. and M. Garstang, 1978: "The role of surface convergence and vorticity in the life cycle of convective rainfall. Part I: Observation and analysis" J. Atmos. Sci., Vol.35, pp.1047-1062.

V

Acronyms and Abbreviations

AFWAL - Air Force Wright Aeronautical Laboratories

ASFL - Atmospheric Science Field Laboratory
Building near Mosquito Lagoon, about 15 miles north of the main KSC building complex, housing equipment, research space, and offices for conducting field experiments in lightning and other aspects of atmospheric science.

EMP - Electromagnetic Pulse
Pulse of electromagnetic radiation emitted by nuclear explosions.

ESMC - Eastern Space Missile Center

FAA - Federal Aviation Administration

ISIS - Integrated Storm Information System
System for storage and display of digital radar data and/or cloud-to-ground lightning strike location. Displays either radar or lightning data separately on the video terminal, or both together.

KSC - Kennedy Space Center

LLP - Lightning Location and Protection, Inc
Manufactures of ISIS.

MSFC - Marshall Space Flight Center

MIDDS - Meteorological Interactive Data Display System
World-wide weather data dissemination and display system.

NASA - National Aeronautics and Space Administration

NRL - Naval Research Laboratories

RTLP - Rocket Triggered Lightning Program
Program at KSC to launch small rockets into thunderstorm clouds, triggering lightning at the launch site.

RTLS - Rocket Triggered Launch Site
Site on Mosquito Lagoon, near the ASFL, where RTLP personnel launch small rockets into active thunderstorm clouds. Contains launch sites over land and water.

SUNYA - State University of New York at Albany

N90-16691

1989 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

THE DAB MODEL OF DRAWING PROCESSES

PREPARED BY: Dr. Larry W. Hochhaus

ACADEMIC RANK: Associate Professor

UNIVERSITY AND DEPARTMENT: Oklahoma State University
Psychology Department

NASA/KSC

DIVISION: Data Systems Division

BRANCH: Technical & Information Systems

NASA COLLEAGUE: Ms. Carrie Belton

DATE: August 11, 1989

CONTRACT NUMBER: University of Central Florida
NASA-NGT-60002 Supplement: 2

TABLE OF CONTENTS

I. ABSTRACT

The problem of automatic drawing was investigated in two ways. First, a DAB model of drawing processes was introduced. DAB stands for three types of knowledge hypothesized to support drawing abilities, namely, Drawing Knowledge, Assimilated Knowledge, and Base Knowledge. Speculation concerning the content and character of each of these subsystems of the drawing process is introduced and the overall adequacy of the model is evaluated. Second, eight experts were each asked to understand six engineering drawings and to think aloud while doing so. It is anticipated that a "concurrent protocol analysis" of these interviews can be carried out in the future. Meanwhile, a general description of the videotape database is provided. In conclusion, the DAB model was praised as a worthwhile first step toward solution of a difficult problem, but was considered by and large inadequate to the challenge of automatic drawing. Suggestions for improvements on the model were made.

II. INTRODUCTION

What does a person have when she has the ability to draw? At the broadest level, the purpose of this paper is to describe drawing processes. More narrowly, however, focus is on how engineering drawings are created from a knowledge base which abstractly represents the object to be portrayed. Additional concern is with development of a conceptual model of the drawing process.

The model is designed to assist thinking about the cognitive and information processing operations involved in translating from a knowledge base to a drawing of the system represented. The model is also designed to help understand the requirements of creating an automatic drawing mechanism to accomplish that task.

The model is called DAB because its activities are supported by three general knowledge systems, Drawing Knowledge, Assimilated Knowledge, and Base Knowledge. The parts of the DAB model and their relationships are shown in Figure 1.

The DAB model was inspired by recent efforts to expand the capabilities of KATE, an artificial intelligence project developed in the laboratories of NASA, the National Aeronautics and Space Administration. KATE (Knowledge-based Autonomous Test Engineer) is a reasoning system which uses stored knowledge about the structure and function of a variety of systems. Its purpose is to apply captured abstract thinking processes of engineers in the form of algorithms to the tasks of monitoring, diagnosis, and control of launch systems.

KATE represents in its memory both the components of the modeled system (e.g., electronic relays, valves, pumps) and the connections of those components. Ideally, with added dynamic drawing capabilities, KATE could produce visual representations of the target system. The drawings should portray both functional and structural characteristics. Such visual displays would speed any human's understanding of the modeled system and its components.

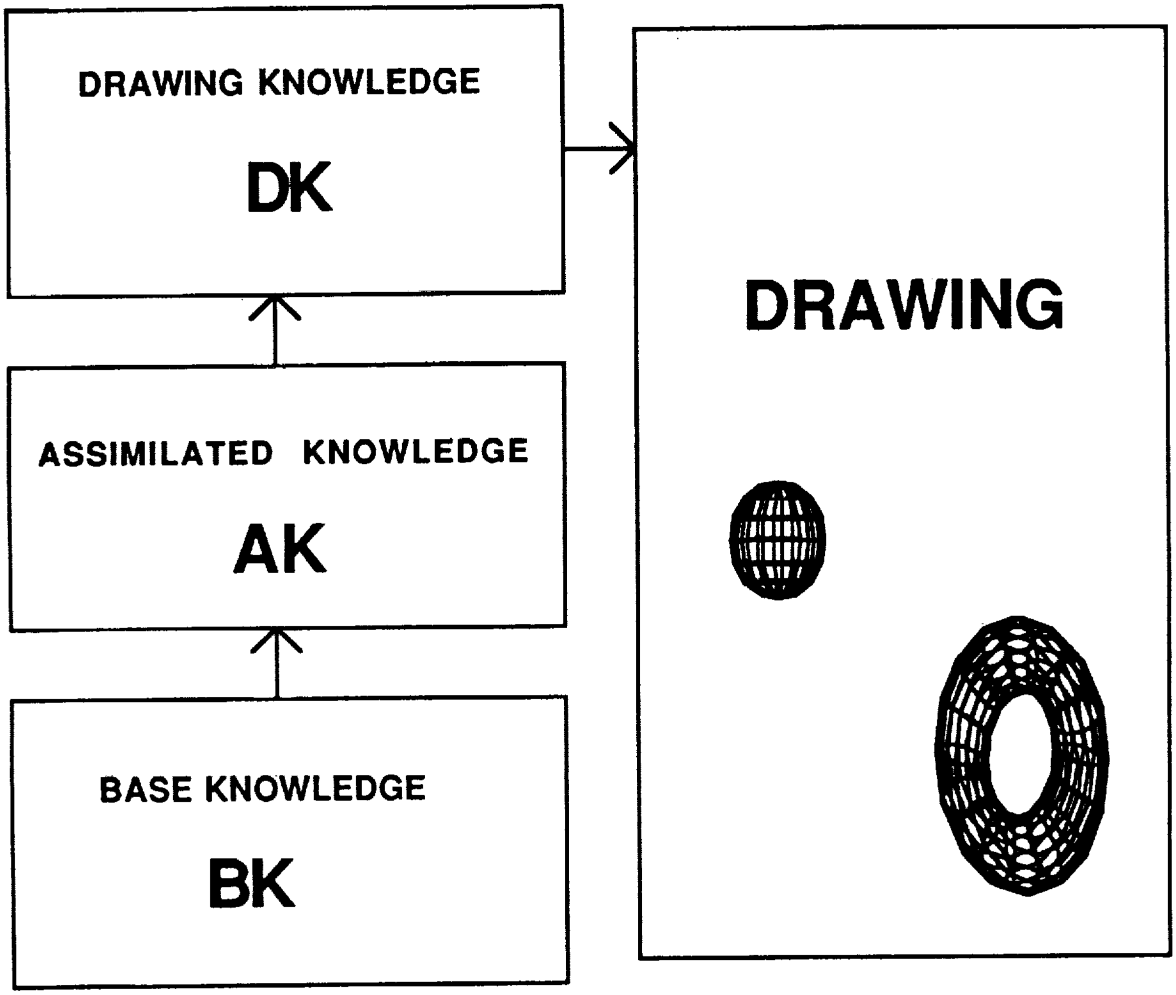


FIGURE 1 THE DAB MODEL

III. OVERVIEW OF KATE

As indicated, KATE is an artificial intelligence system designed to mimic the reasoning processes of an experienced engineer. At NASA, the initial applications of KATE were narrowly specific. Nevertheless, by continually readapting KATE to many systems and problems, the character of KATE has become continually more generic and pertinent to a wider variety of engineering systems.

The review by Scarl, Jamieson, and Delaune [1] provides a fuller description of KATE than will be offered here. Instead of another thoroughgoing review, the present summary will focus on KATE's drawing capabilities and on the descriptions of KATE which pertain to the creation of visual representations of the systems KATE is modeling. In doing so, it will help to keep in mind the distinction made by Simon [2] and others between declarative and procedural knowledge. Declarative knowledge pertains to facts and relationships between facts while procedural knowledge refers to knowledge about how to do things. As one might guess, KATE contains both types of knowledge.

KATE's declarative knowledge is primarily in the form of a knowledge base. This base knowledge represents the components of the system being modeled, the connections of these components to one another, mathematical values (such as pressure or temperature readings on a sensor in the system), and functional relationships between component values. The base knowledge of KATE is one of the three fundamental knowledge systems which support the DAB drawing model. More will be said about base knowledge characteristics in the next section of this report.

Procedural knowledge in KATE consists of how KATE uses its knowledge base. One example of this is KATE's diagnostic capacity in which an evaluation is made of possible causes of faulty sensor readings in the system being modeled. KATE attacks such problems by creating a suspect list. Next, through inference processes KATE attempts to logically rule out or determine the innocence of the various suspects. Under ideal circumstances, only the component actually at fault will remain on the suspect list. (KATE assumes only a single point of failure). When more than one suspect remains, further tests (e.g., application of commands to the system) may be needed to appropriately narrow the explanations of the erroneous reading or readings.

As mentioned in the introduction, it would be very helpful if KATE could automatically create visual representations of the system being modeled. Such visual representations would not augment KATE's declarative or procedural knowledge with respect to solving engineering problems, but the drawings could be very useful to any human user of KATE. Some of the uses which could be made of graphical representations of KATE's declarative knowledge are:

1. Creating alarms which alert humans to emergency situations.

2. Permitting humans to understand how KATE operates.

3. Teaching a novice about the modeled system.

4. Creating a user interface.

5. Aiding diagnosis of system faults.

These five functions, alarms, understanding KATE, teaching about a new system, creation of a human-machine interface, and fault diagnosis are the motivation for the present efforts to increase KATE's drawing ability.

At the time of this writing, KATE can do some drawings, but its visual representation ability is limited in two primary ways. These limitations correspond to KATE's two ways of visually portraying the modeled system, tree drawings and iconic drawings. Concerning the first of these, KATE by lines can create a visual tree-structure to show how the parts of the modeled system are connected to one another. To avoid line crossings, KATE, whenever a series of parts form a loop, will create a doubled representation of a component in the system and mark the duplicated node in the drawing with an asterisk to note its replication. Size limits to the drawings are accomplished by selecting a focal component and then moving upstream and downstream a criterion number of components.

In two ways, the tree diagram is not a faithful drawing of the modeled system. First, whenever they occur, duplicated representations of components in the system will guarantee the drawing will lack accuracy as a representation of a real world system. Second, the components of the tree drawings are not realistic representations of what they stand for. For example, pumps are not drawn to look like pumps, but are protrayed only as a labeled node in the network.

KATE additionally creates iconic drawings of the system being modeled but, in its current form, KATE sometimes generates these, but most often relies on human artists. The KATE programs (presently in LISP) contain "icons" or computer graphics images of the modeled parts. The layout, organization, and placement of these icons, however, is typically not automatically generated, but is given by the user in a template-type drawing system which simply places the selected icons at their predetermined locations on the screen.

At present, only a small portion of KATE's drawings are fully automatic. In most cases, KATE draws in iconic form only what has been predefined as a subsystem of the system being modeled. Hence, the visual drawings currently produced by KATE are more like memorized drawings than like sketches generated to suit the situation. Stated another way, the layout of iconic drawings is not currently created by procedural knowledge within KATE, but is a form of declarative knowledge based on what has been drawn by a human artist. Because

KATE's drawings are "canned" rather than synthesized, the creation of iconic drawings is an extremely time-consuming step in using KATE and is necessary for each new application. Were this drawing process more generic, considerable savings and generalizability could be realized for the KATE system.

IV. THE DAB MODEL

4.1 The Nature of Automatic Drawing

The scope of the problem of creating automatic drawings from a highly abstract engineering knowledge base is enormous. In magnitude, the task is akin to a request to simulate the creative talents of a design engineer and an engineering draftsman. One can reasonably ask, where does one begin? Where indeed? It is one purpose of the present paper to make such a beginning.

The present DAB model is partly the result of ideas stimulated by interviews of eight engineers and designers which are described in Section V of this paper. Primarily, however, DAB is simply a rationalistic effort to indicate what an automatic drawing system would have to know in order to produce reasonable drawings and schematics of systems such as those modeled within KATE. Put another way, what would an expert designer have to have to be able to create a good schematic representation of an electrical or mechanical system? DAB is thus a process model of the skill that results from such an ability.

It is an assumption of the present approach that knowledge of human drawing ability will ultimately aid development of automatic drawing abilities. Unfortunately, the lines of distinction between DAB as a process model of human abilities and DAB as a model of automatic drawing have not been kept clear throughout this paper. Primarily, DAB is a model of human skills and processes, yet it is hoped that by starting here the drawing process can be brought under fully automatic control.

The current DAB model of drawing is additionally a cataloging system by which to organize the varieties of declarative and procedural knowledge required to create effective functional representations of systems. Once again, DAB stands for Drawing Knowledge, Assimilated Knowledge, and Base Knowledge (see Figure 1). For the purpose of cataloging artistic abilities and processes, these three components are seen as mutually exclusive and exhaustive categories. Dynamics of information exchange between these subsystems, however, may at times blur their boundaries.

The Base Knowledge component has been reviewed somewhat already in describing KATE in the preceding section of this report. The Base Knowledge is the declarative knowledge of the KATE system and is a list of components, physical connections between components, and mathematical functions relating commands, components, and outputs. The mathematical expressions represent transfer functions. Such transfer functions are equations which relate changes in command values to changes in outputs.

Drawing Knowledge represents what is required to directly create the visual representation to be presented to humans. Such

Drawing Knowledge will include procedures for creating drawings which satisfy aesthetic constraints and will include other complex mechanical routines needed in the actual drawing process.

The Drawing Knowledge aspect of the DAB model will include additional knowledge in the form of a subsystem which knows how to create a general layout and knows where to begin in doing so. The source of this knowledge is considered one of the central keys in the drawing process. The possibility of including a Layout Knowledge processing component between Assimilated Knowledge and Drawing Knowledge was considered, but rejected due to a simplicity criterion. For now, knowledge of layout procedures are included as a part of the Drawing Knowledge component of DAB.

Expanded consideration of the Drawing Knowledge component is given in the next subsection, 4.2. It may help to keep in mind that all of the Drawing Knowledge is procedural knowledge. As such, the Drawing Knowledge stage merely tells the DAB model how to put pen to paper.

The third component of the DAB model, Assimilated Knowledge, represents the planning, organization, and preparation a designer must go through before actually beginning to draw. Interviews described in Section V of this report indicate such prior planning is a large measure of what is done when a designer puts a schematic together. The actual time spent in drawing (application of Drawing Knowledge) appears to be minor compared to the work that must be undertaken in getting ready to draw. Assimilated Knowledge is the link between the declarative knowledge of the Base Knowledge component and the procedural knowledge of the Drawing Knowledge stage of the DAB model.

At the start of this writing, it is not clear whether Assimilated Knowledge is declarative or procedural in character. Later, we will return to consideration of this issue. For now, I suspect it is a creative mixture of both. Perhaps the Assimilated Knowledge stage, via as yet unspecified procedures, creates a set of declarative knowledge structures out of the vast encyclopedic knowledge which it must possess and turns these synthesized chunks of knowledge over to the Drawing Knowledge component just prior to drawing.

The following quotation vividly conveys the mental processes involved between the time a designing problem is presented to introspective awareness and the moment a creative solution to that problem is discovered. It is not hard to imagine that similar mental steps would take place as an artist or a design engineer prepares to start the overt drawing process.

"I have this amplifier to design. It is supposed to operate at a center frequency of 100 megahertz. The output must be one volt and the input is one-tenth volt, which is a gain of 10. I better make the gain variable from 10 to 20 in case the other devices are off normal. The amplifier will have to be shielded. I can use a metal can for the shield with coaxial lines for all signals and high frequency reject for the power and control lines. Say, this amplifier sounds a lot like the amplifier I designed for the XYZ project - where did I put the design drawings for the XYZ project?" [3].

The designer in the quotation above has drawn on what numerous cognitive psychologists have called schemata (singular, schema). Schemata are broad generalizations based on past experience. Schemata, therefore, are distillations of the enormous quantities of knowledge accumulated through humans' memories of their encyclopedic repertoire of interactions with the world.

The concept of schemata is most often credited to Bartlett [4] who himself acknowledged inspiration in the writings of the neurophysiologist, Head. Head's thinking, however, was limited to the notion of procedural knowledge schemata, such as those an experienced player would employ in a game of tennis. Bartlett simply extended the concept of schemata to include declarative as well as procedural knowledge. For example, in recalling complex narratives, such as Bartlett's famous "War of the Ghosts" story, intrusions of words such as "canoe" for the word "boat" in the original story, led Bartlett to infer the operation of perceptual schemata which must have distorted the participant's original perception of the story.

Since Bartlett, numerous cognitive psychologists have extended the concept of schemata both theoretically and empirically. For example, at a theoretical level, Minsky [5] introduced the idea of "frames" and Schank [6] developed the notion of "scripts" which, among other things, made the concept of schemata both more precise and more functional. Empirical demonstrations of the operation of schemata have been frequent also. A common example along these lines was provided by Chase and Simon [7] who showed that an expert chess player surpassed a novice in the recall of sensible chess board configurations, but showed no differences in memory for senseless chess board patterns. The implication is that the expert's schemata for previous chess positions were responsible for the result.

To return to the main point, Assimilated Knowledge within the DAB model may represent a process of activation of the proper schemata from the individual's encyclopedic knowledge so that these schemata may be passed along to the Drawing Knowledge component in a usable form.

The expanded consideration of the nature of the Assimilated Knowledge component of the DAB (subsection 4.3) consists of further speculation about the varieties of schemata which must operate at various times in the drawing process as well as additional consideration of what those knowledge assimilation processes may be. Section 4.4 is an evaluation of the DAB model.

4.2 Drawing Knowledge

To appreciate the complexity of the drawing process, try this simple exercise: Write down a list of four components and name them for the letters A through D. Add to this list six connections between pairs of components (such as A-B, B-C, or B-D). Keep in mind that each letter should have at least one connection but that any letter could have multiple connections to other letters. From this list of parts and pathways between parts, create a two-dimensional drawing depicting the symbolic connections. The demonstration should convince anyone that with just a few more components and a few more pathways the drawing task could quickly turn into spaghetti chaos if it has not done so already.

In spite of this complexity, a number of papers have attempted to produce objective descriptions of such a drawing process [e.g., 8, 9, 10]. The brief but thoughtful paper by Batini et al. [8] covers several key contributions these efforts have produced and illustrates a number of essential features of the Drawing Knowledge process. Batini et al. describe four aesthetic criteria useful as guides to the drawing process: 1) minimization of the number of crossed lines, 2) minimization of bends along connections, 3) minimization of the global length of connections, and 4) minimization of diagram area.

By establishing a rigid priority between aesthetic criteria, Batini et al. were able to create a layout algorithm (called GIOTTO) consisting of a five-stage process. The stages are modeling, planarization, orthogonalization, compaction, and drawing. The last four stages of the GIOTTO model represent what may be called Drawing Knowledge in the terms employed in the present paper. The first stage, modeling, is the transition process between the conceptual schema of the depicted system and the earliest conceptual graph of the system. As such, modeling represents part of what is meant in the present report by Assimilated Knowledge.

Clearly, I would not want to suggest that all the problems of screen layout have been solved. Related papers [e.g., 11] on the topic of automatic circuit design further confirm the scope of such a challenge. Nevertheless, to allow more space to focus on the equally tough problems in specifying the nature of Assimilated Knowledge in the DAB model, further discussion of Drawing Knowledge will be herewith suspended. Because of its relevance to the problem of drawing by the KATE model, however, applause is issued for what researchers have done with Drawing Knowledge so far and encouragement is offered to keep up efforts along these lines.

4.3 Assimilated Knowledge

In one of the interviews described in Part V of this report, an experienced designer hinted at what has to be done before the drawing process is overtly initiated. When asked to elaborate on this planning process, he described taking field trips to carefully inspect and get to know the system to be drawn. If an electrical system was under consideration, lists of pin connections would be made, checked, and rechecked to ensure high accuracy. Overall, this preparatory stage was characterized as typically containing over 80% of the work involved in drawing.

Like a number of designers these days, the individual made extensive use of computer-aided design (CAD) programs to accelerate the actual drawing process. The almost trivial act of completing the remaining lines once the CAD templates had been put in place once again emphasized that establishment of the Base Knowledge and Assimilated Knowledge procedures constitute a major portion of the experienced designer's repertoire. Drawing Knowledge is important, but prior planning is essential.

Where then does the designer begin in creating or activating the Assimilated Knowledge needed prior to the overt expression of her drawing? The interview report of the designer in the paragraphs above suggests that, in humans, creation of a solid Base Knowledge is part of the planning process. If a thoroughgoing understanding of the Knowledge Base cannot be assumed, it is the designer's responsibility to create such a mental structure.

One way to create such a Base Knowledge is to physically interact with the system to be depicted. This might involve field visits to look at, possibly touch, and possibly dismantle the system to be understood. Conversations with other engineers or experts at the site of the system might also be exploited in order to rapidly establish the requisite understanding. In short, part of the contents of the Assimilated Knowledge component of the DAB model is an executive which can make high-level decisions about what to do to be sure all is known that needs to be known before starting the overt drawing process.

How can a system know it is missing key information without knowing exactly what it is that is missing? That is, how can a system know it should search for something if it doesn't know just what it is searching for? Such a "metasystem" process sounds altogether mysterious, yet need not be so. For example, an executive process such as the one described could merely ask if it was in the posession of a complete Base Knowledge for the system to be depicted. If not, the executive could initiate any or all of a list of activities (such as field trips or questions directed toward experts) which would heuristically lead toward completion of the missing Base Knowledge.

Creation of a Base Knowledge is not critical within KATE; presence of an intact Base Knowledge can be assumed. It should be noted that KATE does do some consistency checking. For example, KATE can determine if inputs and outputs are connected. This evaluation is in no way semantic, however. For example, KATE never asks if any configuration of parts makes reasonable sense. The problem of organizing the Base Knowledge component for drawing purposes, however, is an important area where the DAB model needs expansion. The question at issue is how can KATE achieve a meaningful parse of the system it is modeling in a way that it can capture functional as well as structural roles for the system's components. A related question is how KATE can subdivide the drawing problem into perceptual organization units or "chunks."

A good designer, like the one who reminded himself of the prior design of the XYZ project, should have an extensive repertoire of prior experience with drawings and a good system of retrieving more detailed memory of those drawings. The manner in which this access takes place is not immediately apparent, however.

This is not the place to launch a description of a novel analog memory retrieval scheme, but clearly something akin to continuous information access or image retrieval must be operating within Assimilated Knowledge processing. Consequently, a brief model will be sketched; elaboration and full development of this model of image recall will be deferred to others.

The problem is that when provided with discrete retrieval cues, how are continuous memory images accessed? Take Donald Norman's example... "As you approach from the outside the house you lived in three houses ago, did the door open on the left or the right?" How do discrete signals call forth continuous memory images?

Let's call the required memory process a Discrete to Analog Memory (DAM) retrieval system. DAM processing takes place by using the richly creative and generative imagery system [see 12] to build an analog image from the discrete question cues. Once established, the generated image could be matched to stored images hence triggering retrieval of the best matching continuous image in memory.

It is surprising no one has suggested anything like the presently proposed DAM system for recall of continuous experience, but I am not presently aware of such a system. Future research could be directed toward elaboration of tests of the implications of such a memory model.

In a nutshell then, the Assimilated Knowledge component of the DAB model requires a memory system which is capable of retrieval of continuous images stored in the designer's experience. For expository convenience, I have called this a DAM retrieval process.

A third possible aspect of the DAB model's Assimilated Knowledge processing is constraint of the proposed drawing on the basis of drawing standards in force within the professional community of the designer. Many engineers in the interviews summarized in Part V complained that the drawings shown to them during the interviews depicted the flow of fluids from right to left rather than the opposite way which they were used to seeing within the NASA community.

Because of my limited knowledge of engineering drawing and due to my short stay here at NASA, I have only a sketchy knowledge of such drawing standards. Nevertheless, some example constraints are as follows:

1. Put power supplies on the left.
2. Put electrical ground connections on the right.
3. Put sensor measurements and input commands on the top.

4. Put electro-mechanical components at the bottom.

5. Group similar devices on a horizontal line.

This third aspect of Assimilated Knowledge functioning will be called constraint processing. It could easily be argued that constraint processing should also be incorporated within the Drawing Knowledge stage of the DAB model, but it is realistic to believe these adjustments exert their influence as early as the Assimilated Knowledge stage. Constraint number five above, for example, suggests acknowledgement of repetition within the drawing is part of the planning that must take place before overt drawing is initiated. Therefore, it appears that both Drawing Knowledge and Assimilated Knowledge are subject to constraint processing.

The present elaboration of the Assimilated Knowledge component of the DAB model provides for three subprocesses or subsystems within it; namely, executive processes, DAM retrieval, and constraint processing. In future elaborations of the DAB model, additional processes surely must be incorporated into the Assimilated Knowledge component in order for it to function adequately. For now, however, it will consist of the present short list.

1. Executive Processes 2. DAM Retrieval 3. Constraint Processing

4.4 Evaluation of the DAB model

The greatest problem with the DAB model is its sketchiness and its incompleteness. Given little prior work to go on in this area, however, the present modest start should perhaps not be discounted. Much more time needs to be devoted toward outlining the character of the critical stages of drawing, assimilated, and base knowledge. The Assimilated Knowledge component needs elaboration in particular. The three subprocesses suggested are at best a very rough first guess at what might belong in the Assimilated Knowledge stage. Executive processing seems essential, though in need of further development. Specification of behaviors and activities triggered by the executive processor in reaction to missing Base Knowledge information needs to be made. The operation of organizing components of the Base Knowledge on the basis of functional knowledge of those parts is a key shortcoming in the present early sketch of the Assimilated Knowledge component.

DAM retrieval, the process by which pertinent schemata are activated is clearly underdeveloped in its present form. Also, the locus of the constraint processing mechanism must be resolved. As noted earlier, the nature of the knowledge within the Assimilated Knowledge stage (declarative vs. procedural) should be specified in future extensions of the DAB model. Perhaps the lack of specificity with respect to the character of the Assimilated Knowledge component's information is a result of the preliminary character of the DAB model. Finally, the Base Knowledge and Drawing Knowledge processes are generally well understood. While knotty problems in the drawing processes supporting Drawing Knowledge must still be worked out, it is felt that the spotlight of attention should remain on the Assimilated Knowledge component for quickest progress.

At present, plans for a major overhaul of the DAB model are already underway. It appears that the general character of the Assimilated Knowledge component of the model is procedural. In the revision, Assimilated Knowledge will likely be replaced with a general functioning set of control processes which reorganize the declarative knowledge structures of the Base Knowledge to create and deliver layout procedures to the Drawing Knowledge component. In doing so, use would be made of semantic knowledge concerning real world functional properties of components.

Regardless of these shortcomings, the DAB model has provided the springboard to thinking about visual thinking that it was supposed to. Its lack of elaboration seems to point to a gap in our understanding about the productive processes which drive artistic expression. In many areas of psychology we know far more about comprehension processes than about the converse operations of production, such as in the area of language models [13]. The reasons for our difficulties in generating good models of production processes seems to stem from our inability to conduct rigorously controlled experiments to test hypotheses about production processes. The targets of comprehension, however, are far easier to manipulate and control, hence comprehension processes are better understood. Such a state of affairs, however, should not daunt our efforts to make reasoned guesses about production processes such as those involved in drawing operations.

The gap between our understanding of production and comprehension processes is even more exaggerated in the areas of perception and drawing than it is in theories of language. A visit to any reasonably stocked library will document this. Shelves and shelves of books can be found on the topic of perceptual processes, but while examples of drawings may exist, written works describing the information processing operations in drawing are scarce indeed. Two exceptions are the excellent texts by McKim [14, 15]. While these are not fully developed cognitive models of processing operations in drawing, they go beyond what is presented here to a fair degree. McKim's works on the topic of drawing models are highly recommended.

V. OTHER SUMMER RESEARCH ACTIVITIES

5.1 Background

Work on the current project began with orientation to the NASA community, preliminary efforts to define a summer project, and reading background literature. After two weeks, a research plan was established, a 10 page proposal was drafted and typed, and the project described in the remainder of this section was undertaken.

Another week was spent in locating and selecting schematics to be used and in securing tentative agreements with potential participants. Once the data were collected, the rough plan for this report (brief theory paper and summary of data collection procedures) was adopted. The remainder of Section V is a description of the data collection activities undertaken and is organized somewhat along the lines of a typical scientific report.

5.2 Introduction

To advance understanding of the manner in which electronic and mechanical drawings are produced, steps were taken to collect basic data concerning the processes by which experienced engineers comprehend such drawings. This approach tacitly asserts that knowledge of drawing comprehension processes is propaedeutic to theories of drawing production. As indicated in Section 4.4 of this report, scientific rigor in the investigation of comprehension can be expected to excede the control available for research on production processes. Later, as the efforts begun here are carried foreward, the assumption that production and comprehension processes are converse can be more carefully evaluated.

A common research method for discovering the mental processes behind intelligent behavior has been "concurrent protocol analysis" or the "think aloud method." Newell and Simon used this technique to evaluate a number of skills from problem solving and cryptarithmetic to chess ability. Their summary of these efforts [16] is considered classic reading in the area of artificial intelligence. See also [17].

Letovsky [18] provides a recent example of the use of concurrent protocol analysis in the study of comprehension processes. Letovsky gathered verbal protocols from professional computer programmers as they attempted to understand and modify a computer program. Efforts were made to catalog cognitive events as the programmers were engaged in the comprehension portion of their task. These event types were used to derive a computational model of the programmers' mental processes.

With the think aloud method, data collection is relatively rapid while the protocol analysis itself is painstakingly slow. Given the brief time available for the present fellowship activities, it was decided that a data base would be created which would consist of videotape recordings of eight experienced engineers each evaluating six schematic drawings.

Detailed analysis of the contents of those recordings will await further investigation resources of time and money.

The purposes of Section V of this report then are 1) to describe the visual materials and participants that were employed in creation of the videotape recordings and 2) to provide a record concerning details of method behind establishment of the tapes so as to help one understand their contents at a later time.

5.2 Method

5.2.1 Materials

The six drawings were selected from various sources suggested by engineers and staff in the Artificial Intelligence laboratory here at the Kennedy Space Center. In the order presented to each participant, the schematics employed were 1) Apollo - Skylab-I Launch Complex 39 Environmental Control System Mechanical System (79K00076; sheet 29), 2) ibid. (sheet 12), 3) Hypergol Fuel Deservicing System (79K09247; sheet 82), 4) Apollo - Skylab-I Condenser Water (79K00076; sheet 19), 5) Hypergol Fuel Deservicing System (79K09247; sheet 4), and 6) Red Wagon - Simulation of Liquid Hydrogen Loading System (unclassified).

Drawings 3 and 5 were electrical and the remainder were mechanical schematics. Most of the drawings were cropped to retain a portion of the drawing which was judged to be both somewhat coherent and at an intermediate level of complexity. If still in view, bottom titles to drawings were removed. In several mechanical drawings, a few labels (such as "water glycol tank") were obliterated to make the comprehension process more challenging. For managability, Drawing 4 was made smaller (.67 original size) with a reduction copy machine.

5.2.2 Participants

Eight experienced NASA and Boeing employees were used as participants. In the order they appear on the two videotapes, highest academic degrees and disciplines represented were as follows: 1) PhD-Mechanical Engineering, 2) BS-Chemical Engineering, 3) BS-Electrical Engineering, 4) BS-Electrical Engineering, 5) BS-Computer Science, 6) MS-Computer Science and Electrical Engineering, 7) PhD-Mechanical Engineering, and 8) BS-Mechanical Engineering. Mean age was 38.1 years and mean years of work experience was 10.6. There were seven males and one female represented.

5.2.3 Procedure

Participants were individually interviewed by the present author. When possible, attempts to build rapport were made with participants by informal chatting prior to videotaping.

Taping occurred in a large conference room (104 Engineering Design Laboratories). The tripod-mounted videotape camera (Panasonic model WV-CC60) was placed on a table adjacent to the six foot diameter round work table. The field of view captured about one-half of the table and included the drawings and the left arm and upper trunk of the interviewee.

Participants were read a set of general instructions describing the think-aloud method. Primarily, the instructions described the goals of the research project and encouraged participants to keep talking throughout the interview.

Following general instructions, a page of specific instructions used three sets of questions to orient participants to the task of understanding the schematics. Verbatum, these questions were:

1. **What does it do?** What does the item or parts of items in the schematic do? (What is the function of the components you see?)

2. **Talk about the drawing.** Why was the schematic drawn in this way? (Are there ways you could improve the drawing?)

3. **What goes with what?** Tell us specifically about any structural or functional relationships between components you see. (How do the parts of the drawings fit together?)

Participants were allowed to refer to the specific instructions sheet whenever they wished during the task.

At this point, the schematics were introduced one at a time. The interviewer prompted participants in a variety of ways as a means to keep them talking about the drawings. Interviews ranged from 20 to 45 minutes in duration.

5.3 Results

Without typed protocols or other means to assist evaluation of the interview results, analysis of the videotapes in presently quite limited. The ideas expressed were influential concerning creation of the DAB model described in Section IV of this report. In any event, the tapes provide a data base, which in addition to protocol analysis, could be used to evaluate specific hypotheses concerning processes during visual perception.

5.4 Discussion

Generally speaking, the interviews accomplished their purposes. Ideas concerning the drawing process were gathered and a database which could support a detailed protocol analysis was created.

Originally, a goal of the present procedures was to learn what "perceptual chunks" or visual organizing units were employed during the comprehension process. It is for this reason the third specific orienting question was framed to detect how parts of a drawing were grouped. Other types of task orientation were condidered and rejected. For example, Letovsky [18] asked his programmers to make a meaningful modification to the programs they were studying thereby indirectly forcing participants to comprehend the computer programs.

Limits of time, however, precluded this approach. Another orienting task, memorization in preparation for identification of the function of parts of the drawing in a test to follow, was dismissed because it was felt the direct methods used here would provide better access to perceptual organization units.

VI. CONCLUDING REMARKS

A problem clearly stated is a problem half solved. The present DAB model begins to show some of the problems that exist in the areas of automatic drawing and models of human drawing processes, yet it has really only scratched the surface of the problem in doing so. In spite of this, the DAB model is a beginning. If it serves as a challenge to others to go on to elaborate a more workable model, it will have served its purpose well.

It is unfortunate such a large-scale project was undertaken with severe time limitations. In the empirical phase of the present research also, greater planning could have structured the data collection procedures to good advantage. The analytical scheme of Letovsky [18], for example, is particularly admirable. Questions, conjectures, and searches were grouped together where possible into higher order structures called inquires. Further subdivisions within questions, conjectures, and searches created a highly useful taxonomy. Once classified, these cognitive events were used to make inferences about the types of knowledge structures that make up programming expertise. The types of knowledge include: programming language semantics, goals, plans, efficiency knowledge, domain knowledge, and discourse rules.

It was hoped that a similar classification scheme for verbal/cognitive events and hypothetical knowledge structures in engineers in the present task could be established but, because the protocol analysis of computer programmers and schematics readers differ widely, it was not possible to adapt Letovsky's scheme for the present purposes. It appears a system for the specific purpose of classifying statements about engineering drawings will need to be created.

In spite of these shortcomings, the positive contributions of the present work seem to be: 1) creation of a database consisting of verbal protocols of experienced engineers attempting to comprehend technical drawings, 2) stimulation of thinking about possible processes underlying the drawing process (i.e., the DAB model), and 3) definition of the problems of learning more about drawing processes. With respect the these criteria, the project has been a great success.

VII. REFERENCES

[1] Scarl, E. A., Jamieson, J. R., & Delaune, C. I., *Diagnosis and Sensor Validation Through Knowledge of Structure and Function,* IEEE Transactions on Systems, Man, and Cybernetics, SMC-17 (1987), pp. 360-368.

[2] Simon, H. A., *The Sciences of the Artificial,* Cambridge, MA: MIT Press, 1969.

[3] Lunsford, G. H., & Efurd, R. B., *Design Automation: an Information Transformation Process,* Southcon/87, Atlanta, GA, 1987.

[4] Bartlett, F. C., *Remembering,* Cambridge University Press: New York, 1932.

[5] Minsky, M., *A Framework for Representing Knowledge.* In P. H. Winston (Ed.), The psychology of computer vision (pp. 211-277). New York: McGraw-Hill, 1975.

[6] Schank, R. C., *The Role of Memory in Language Processing.* In C. N. Cofer (Ed.), The structure of human memory (pp. 162-189). Freeman: San Francisco, 1976.

[7] Chase, W. G., & Simon, H. A., *Perception in Chess,* Cognitive Psychology, 10 (1973), pp. 55-81.

[8] Batini, C., Furlani, L., & Nardelli, E., *What is a Good Diagram? A Programatic Approach.* IEEE Procedings 4th International Conference Entity Relationship Approach, IEEE: Chicago, IL, 1985.

[9] Carpano, M., *Automatic Display of Hierarchized Graphs for Computer-aided Decision Analysis,* IEEE Transactions on Systems, Man, and Cybernetics, 10 (1980), pp. 705-715.

[10] Tamassia, R., di Battista, G., & Batini, C., *Automatic Graph Drawing and Readability of Diagrams,* IEEE Transactions on Systems, Man, and Cybernetics, 18 (1988), pp. 61-79.

[11] Soukup, J., *Circuit Layout,* Procedings of the IEEE, 69 (1981), pp. 1281-1304.

[12] Kosslyn, S. M., *Image and Mind,* Harvard University Press: Cambridge, MA, 1980.

[13] Carroll, D. W., *Psychology of Language,* Brooks/Cole: Belmont, CA, 1986.

[14] McKim, R. H., *Experiences in Visual Thinking,* Brooks/Cole: Belmont, CA, 1972.

[15] McKim, R. H., *Experiences in Visual Thinking, 2nd Ed.,* Brooks/Cole: Belmont, CA, 1980.

[16] Newell, A., & Simon, H. A., *Human Problem Solving*, Prentice-Hall: Englewood Cliffs, NJ, 1972.

[17] Ericsson, K. A., & Simon, H. A., *Protocol Analysis: Verbal Reports as Data*, MIT Press: Cambridge, MA, 1984.

[18] Letovsky, S., *Cognitive Processes in Program Comprehension*, In E. Soloway & S. Iyengar (Eds.), Empirical studies of programmers. Ablex Publishing: Norwood, NJ, 1986.

N90-16692

1989 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

ANALYSIS OF THE 60-Hz POWER SYSTEM AT KSC - THE ORSINO SUBSTATION

PREPARED BY:	Dr. Alex O. Kalu
ACADEMIC RANK:	Assistant Professor
UNIVERSITY AND DEPARTMENT:	Savannah State College Engineering Technology Department
NASA/KSC	
DIVISION:	Facilities Engineering
BRANCH:	Electrical Branch
NASA COLLEAGUE:	Mr. Julian V. King
DATE:	August 18, 1989
CONTRACT NUMBER:	University of Central Florida NASA-NGT-60002 Supplement: 2

ACKNOWLEDGEMENT

I am very grateful to the U.C.F. Coordinators of ASEE/NASA Fellowship Program and to the NASA KSC authorities for the opportunity to participate in what has been a rewarding experience this summer. Specifically, I wish to express my appreciation to Dr. Ray Hosler for making the summer program as much fun as it was a challenge. Thanks, Ray, for all the refreshing extra-morale activities. Ms Kari Baird also deserves many thanks for her willingness to help every time.

I would also like to express my indebtedness to my NASA colleague, Mr. Julian King, for his friendliness and assistance during this Study. Julian played a good host and answered all my questions - giving direct contacts to answer the very few questions he could not. He was very resourceful and a patient teacher. Mr. Steven C. Milton, a U.C.F. electrical engineering student who worked with me, deserves a mention. Steve produced all the graphics and did an 'A' job at it.

Many thanks also to the EG&G engineers who I interacted with during the Study. Finally, thanks to Mrs. Imogene Smith for typing the report.

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

An analysis of the Orsino Substation, a component (50%) of the 60-Hertz electric power system at the Kennedy Space Center, is presented. This report contains separate single-line diagrams of the sixteen feeder circuits to permit easy access to information on the individual feeders for future planning. The load condition of each feeder and load break switch are presented and a heuristic reliability analysis of the system is performed. This report contains information about the system organized in a useful fashion for decision making purposes. The beauty of it is in the simplified manner by which information about the system can be obtained.

SUMMARY

The various feeder networks comprising the Orsino Substation are decoupled and presented in easy-to-read diagrams without any loss of information. The connected loads to each load break switch (LBS) are computed to determine the exact load conditions of each LBS and feeder. These data are compared to the demands, a useful information for power coordination and growth planning. The load connected to each primary transformer is also compared to the demand and to the size of the transformer. The Study shows that, though the system is still very healthy, a systematic planning should be instituted. Hitherto, load has been added arbitrarily or perhaps based on geographic convenience. This cannot continue indefinitely if an optimal lifetime and performance of the system is to be expected.

I. INTRODUCTION

1.1 Purpose

The 60-Hertz electric power system at the Kennedy Space Center (KSC) is described in [1] the GP-900 document. This document, the authors point out, is for the benefit of present and potential users of the power system to help them determine whether their requirements can be met adequately or not. It, therefore, identifies the facilities and the types of power available to them in pursuant to the objectives of the system.

The focus of this paper is different. It is for the attention of the NASA/KSC staff whose functions might include a managerial role in the system expansion planning and reliability evaluation. This paper is also directed toward the system planners and operators and those responsible for system reliability. Much of the information contained in this paper is derived from [2] - the system one-line diagrams document 79K17429. The purpose of this paper is to provide an easy reference and guideline to the NASA staff overseeing system planning when making a decision on adding (or deleting) components to the system. This document is also invaluable for a quick appraisal of the system's load conditions and reliability status. There is a tremendous reduction in effort and time in obtaining information from this paper as compared to doing the same from the one-line diagram document 79K17429.

1.2 Scope and Methodology

This report describes and evaluates only one of the two substations comprising the 60-Hertz electric power distribution system at KSC - the Orsino Substation. This Substation is decoupled, and the Study is done on a feeder-by-feeder basis. A one-line diagram of each feeder in normal configuration is drawn. It is, hence, easy to describe and evaluate the electrical status of the isolated feeder. Both the diagrams and the tables showing the load conditions and a somewhat heuristic assessment of the reliability at various load points are useful tools for decision making.

1.3 Assumptions

The protective devices such as fuses, reclosers, and circuit breakers are assumed to be perfect and known. They are thus omitted in the diagrams (for the most part) so as to enhance clarity and for easy perusal which is a chief objective of this paper.

The feeder diagrams depict the normal configuration status of the individual feeders, i.e., in drawing the one-line diagram of a given feeder, the entire system is assumed healthy.

The load connected to a substation (transformer) cannot exceed the KVA rating of the substation.

The primary voltage of all the substations is understood to be 13.2 KV, and the secondary voltages of the substations are irrelevant to the assessment of the reliability conditions of a load break switch (LBS). Thus, these pieces of information are omitted in both diagrams and tables. Furthermore, they are easily assessible and readily available from other existing documents.

II ORSINO SUBSTATION BRIEF DESCRIPTION

2.1 General

The Orsino substation generally referred to as the industrial area substation is described in the GP-900 document. As is the general intent of that document the description is with regards to the facilities served by the substations rather than its configuration. To determine the margin between available power and demand, it is important to know what feeders are connected to a given transformer in the distribution station. The description that follows is given in the order depicted in the systems one-line diagram, document 79K17429. This description is also an attempt to simplify the diagram.

2.2 Substation Transformers and Assigned Feeders

The six individual transformers in the Orsino distribution station supply power to the following feeders under normal configuration.

T6: 10000/12 500 KVA, 3-0, 115 KV/13-2 KV

Supplies feeder 206

T5: 1000/12 500 KVA, 3-0, 115 KV/13-2 KV

Supplies feeders 211, 207, and 201

T4: 1000/12 500 KVA, 3-0, 115 KV/13-2 KV

Supplies feeders 210, and 202//203

T3: 10000/12 500 KVA, 3-0, 115 KV/13-2 KV

Supplies feeders 208, 209, 212 and 204//205

T2+T1: 5000 KVA, 3-0, 115 KV/13-2 KV

Supply feeders 101, 102, and 103.

III METHOD OF INVESTIGATION AND DESCRIPTION OF DATA

3.1 Feeder Single Line Diagrams

The preliminary drafts of the separate one-line diagrams of the individual feeders were drawn from the system one-line diagram - 79K17429 document. This was done by tracing the loads of each feeder from the distribution station (M6-996) through the load break switches connected to it to the various loads and interconnection points with other feeders. To determine the load supplied by a feeder at normal configuration coordination board diagram was visited. Here the normally open (N.O) and normally close (N.C.) switches are shown. Using this information and from discussions with the power coordinators the final draft of feeder one-line diagrams were drawn.

This report arranges the feeder diagrams in the same order (from left to right) they are found in order documents and computer data base for the system. However, the present and normal configuration shows some redundant feeders (as two feeders in parallel) such as feeders 202//203, 103//212, and 204//205. In these cases only one diagram is shown for both feeders and which ever feeder between the two that is at a higher position in the arrangement (to the left of the other in the system one-line diagram) determines its position in this document, e.g. feeder 103//212 is figure number 9 because FDR 212 is the nineth item in the feeder data base though FDR 103 is the thirteenth item.

The following symbols and notations are represented in the diagrams.

-----------o------------	Normally open (N.O)
#	Load Break switch member (LBS#)
(#)	Connector line number
#	Rated Load in KVA (Size of Substation)

The numbers shown in this report correspond to those assigned to the various components in other existing documents on the system.

3.2 The Connected Load Data

Table 1 is a tabular description of the load conditions of the various load break switches connected to each feeder. These are presented on a feeder by feeder bases in the same order as the single-line diagrams. For a given feeder, the load break switches (LBS) are listed from bottom of feeder up to the distribution station transformers supplying power to the feeder.

The second column of table 1 shows the feeders connected to the LBS. If this column shows more than one entry for a LBS, the first entry is the main supply feeder (the feeder under discussion) while the others are redundancies. The redundancies for a LBS are shown together with the LBS from which they are connected.

The third column of table 1 lists the LBS numbers to which the LBS under consideration is connected. The first of those LBS numbers is usually the bus from which the LBS under consideration draws power.

Column 4 of table 1 shows the load connected to the LBS. This is the sum total of the rated KVA of the transformers connected to the LBS either directly or through other LBS's. It will be observed that this figure increases as one goes up the ladder (down the column). Since some LBS's constitute loads for the preceding LBS. There may be more than one entry in this column. This happens when an LBS shares load with another LBS. The shared loads are denoted by superscript numbers which are explained below.

Explanation of superscripts:

1. The 1900 KVA Load is shared with LBS # 51 on feeder 201

2. The 2500 KVA Load is shared with LBS # 34 on feeder 201

3. The 325 KVA Load is shared with LBS # 23 on feeder 208

4. The 2500 KVA Load is shared with LBS # 59

5. The 1900 KVA Load is shared with LBS # 36 on feeder 211

6. The 150 KVA Load at LBS # 59 is shared with feeder 208 Via LBS #23

7. The 325 KVA Load is shared with LBS # 59

Column 5 of table 1 shows the redundancy ties of the LBS. The term "Failure" is used to indicate disconnection. It might be intentional - during planned switching - for instance, or unintentional as in the case of a fault. In any event the entries on this column show the additional load that will be connected to the LBS if a failure occurred above a designated LBS on designated feeder but no higher than the next LBS. Only single failures are considered, i.e. if more than one redundant line is connected to the would-be failed LBS, that case will not be considered. Simply, only inevitably added loads in the event of a failure are listed.

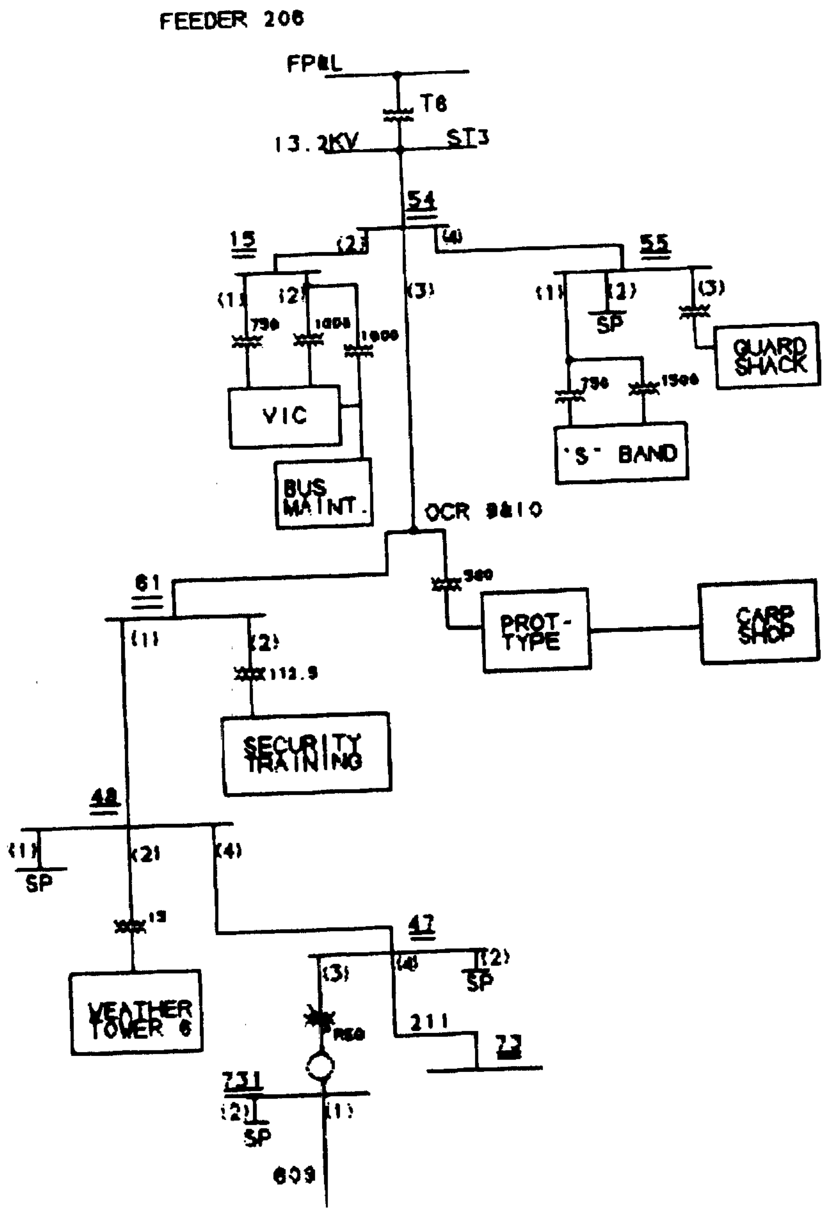

FEEDER 206
FP&L
T6
13.2KV
ST3
54
(2)
(4)
(3)
15
(1)
(2)
750
1000
1000
VIC
BUS MAINT.
55
(1)
(2)
SP
(3)
750
1500
'S' BAND
GUARD SHACK
OCR 9&10
500
PROT-TYPE
CARP SHOP
61
(1)
(2)
112.5
SECURITY TRAINING
48
(1)
SP
(2)
(4)
15
WEATHER TOWER 6
42
(3)
(4)
(2)
SP
REG
211
73
731
(2)
SP
(1)
609

FIGURE 1

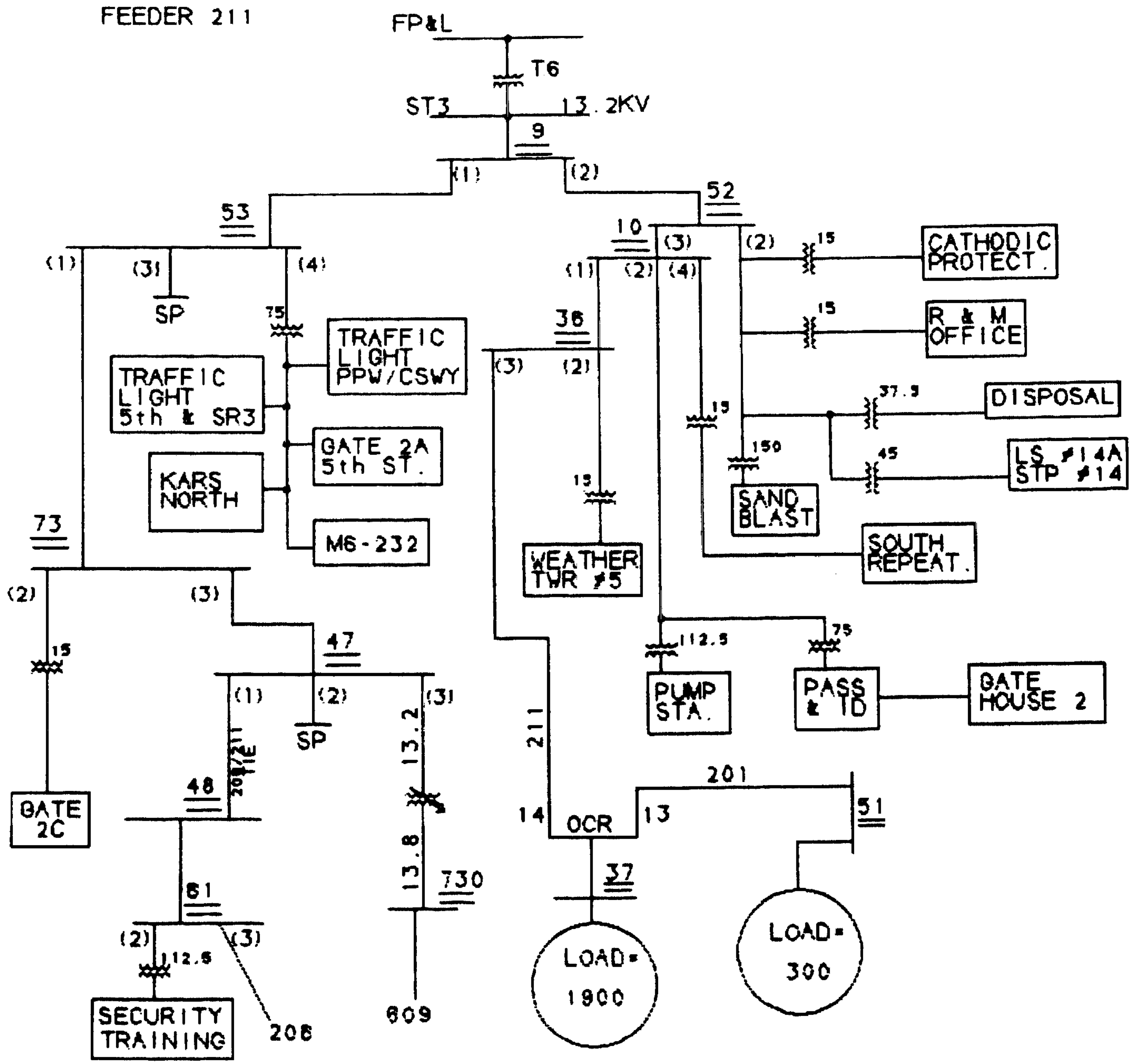

FEEDER 211
FP&L
T6
ST3
13.2KV
9
(1)
(2)
53
52
(1)
(3)
(4)
SP
75
TRAFFIC LIGHT PPW/CSWY
TRAFFIC LIGHT 5th & SR3
GATE 2A 5th ST.
KARS NORTH
M6-232
73
(2)
(3)
15
GATE 2C
47
(1)
(2)
(3)
SP
208/211
48
61
(2)
(3)
112.5
SECURITY TRAINING
208
13.2
13.8
730
609
10
(1)
(2)
(3)
(4)
(2)
15
CATHODIC PROTECT.
15
R & M OFFICE
37.5
DISPOSAL
45
LS #14A STP #14
150
SAND BLAST
15
SOUTH REPEAT.
36
(3)
(2)
15
WEATHER TWR #5
112.5
PUMP STA.
75
PASS & ID
GATE HOUSE 2
211
14
OCR
13
201
51
37
LOAD= 1900
LOAD= 300

FIGURE 2

FEEDER 207

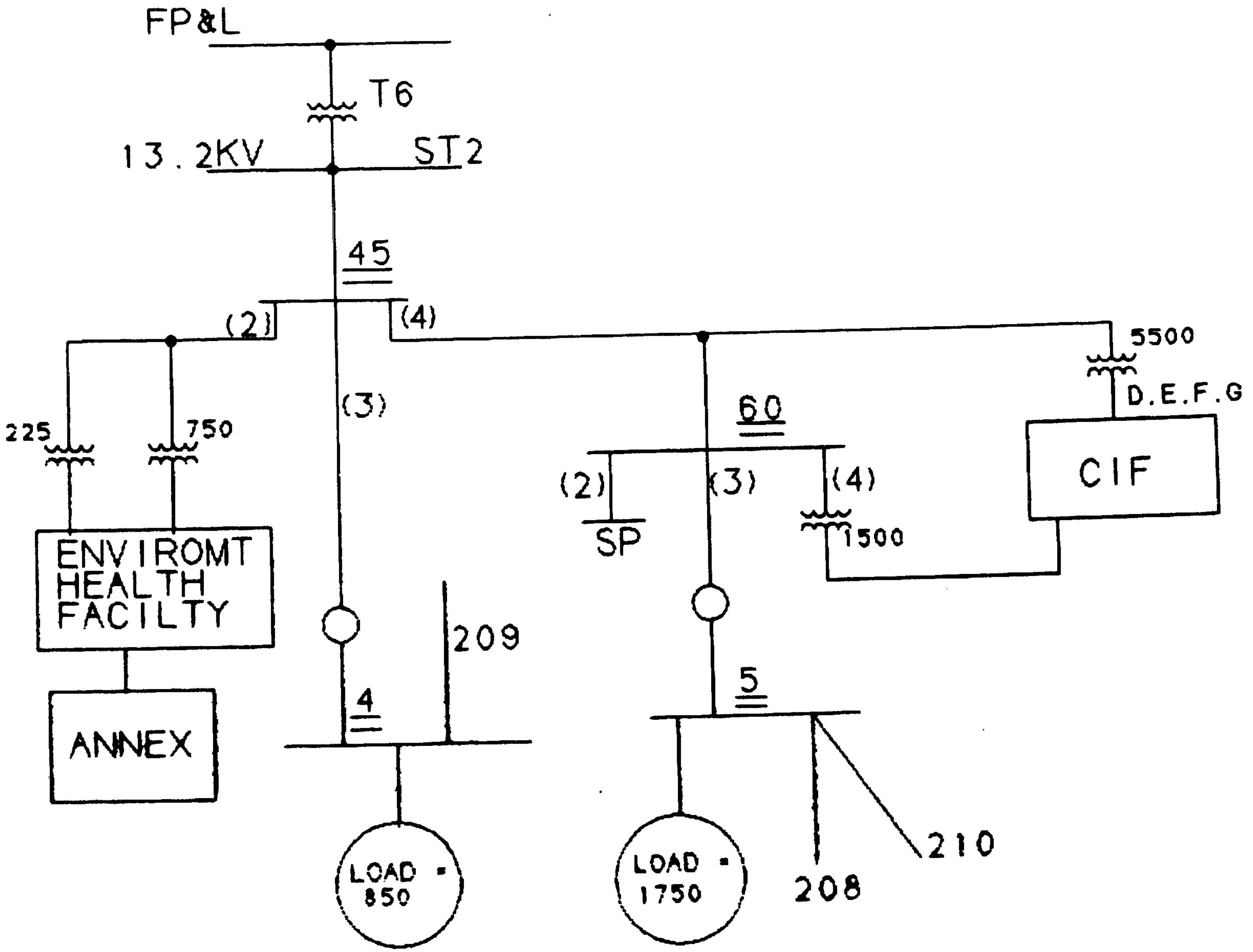


FIGURE 3

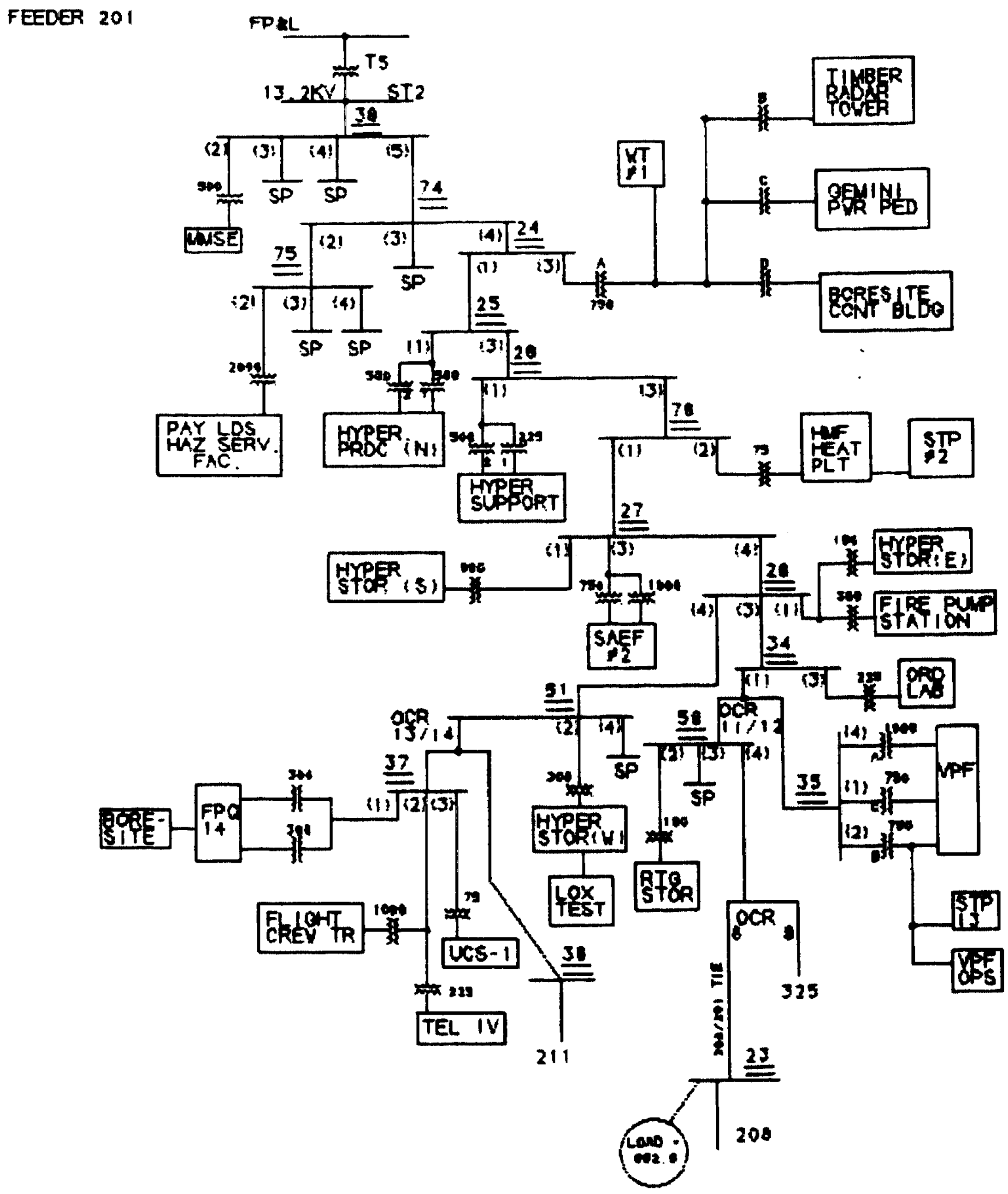

FEEDER 201
FP&L
T5
13.2KV
ST2
38
SP
MMSE
74
75
24
25
26
78
27
28
34
51
58
37
35
38
23
A
B
C
D
790
TIMBER RADAR TOWER
GEMINI PWR PED
BORESITE CCNT BLDG
VT #1
PAY LDS HAZ SERV. FAC.
HYPER PRDC (N)
HYPER SUPPORT
HMF HEAT PLT
STP #2
HYPER STOR (S)
SAEF #2
HYPER STOR(E)
FIRE PUMP STATION
ORD LAB
OCR 13/14
OCR 11/12
HYPER STOR(W)
LOX TEST
RTG STOR
VPF
STP 13
VPF OPS
BORE-SITE
FPQ 14
FLIGHT CREW TR
UCS-1
TEL IV
OCR
211
325
208/201 TIE
208
LOAD - 002.0

FIGURE 4

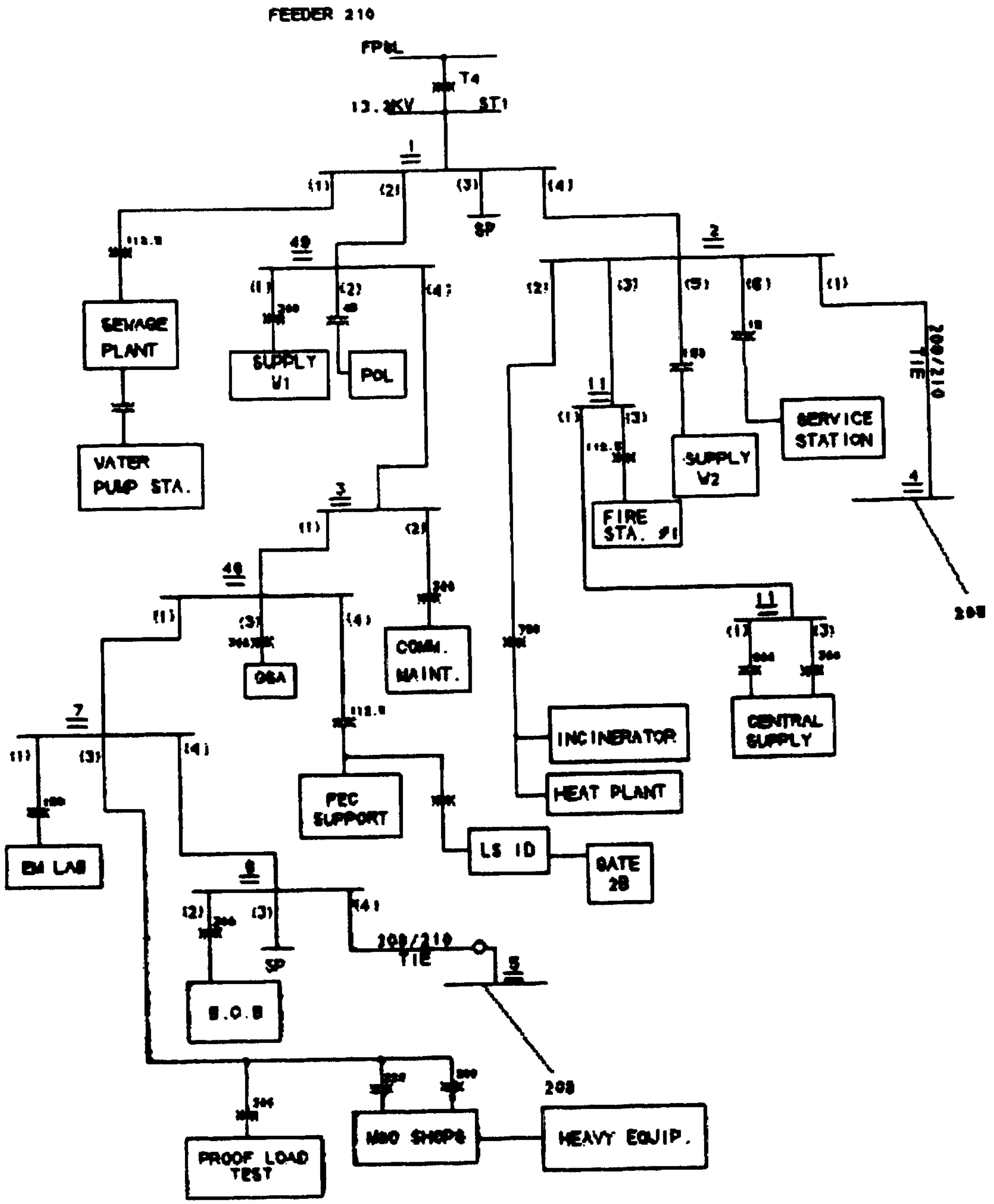

FEEDER 210
FP&L
T4
13.2KV
ST1
1
(1)
(2)
(3)
(4)
SP
112.5
SEWAGE PLANT
VATER PUMP STA.
49
SUPPLY V1
POL
2
(5)
(6)
SERVICE STATION
SUPPLY V2
11
FIRE STA. #1
CENTRAL SUPPLY
200/210 TIE
4
208
750
INCINERATOR
HEAT PLANT
3
COMM. MAINT.
48
GSA
112.5
PEC SUPPORT
LS ID
GATE 2B
7
EM LAB
8
S.O.B
208/210 TIE
5
208
PROOF LOAD TEST
M&O SHOPS
HEAVY EQUIP.

FIGURE 5

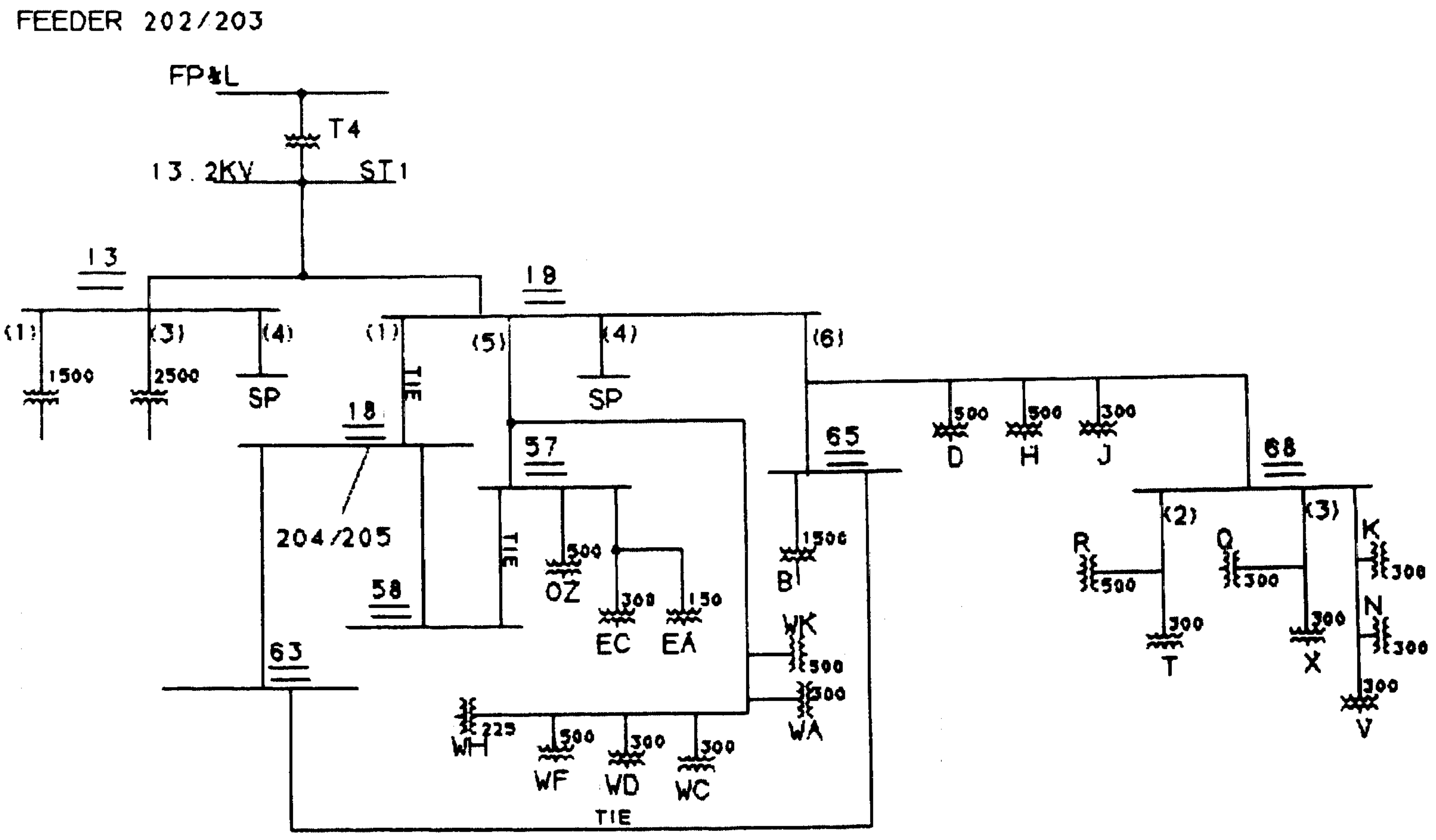

FEEDER 202/203
FP&L
T4
13.2KV
ST1
13
(1)
1500
(3)
2500
(4)
SP
18
(1)
TIE
(5)
(4)
SP
(6)
18
204/205
58
63
57
TIE
500
OZ
300
EC
150
EA
65
1500
B
WK
500
300
WA
225
WH
500
WF
300
WD
300
WC
TIE
500
D
500
H
300
J
68
(2)
R
500
300
T
Q
300
(3)
300
X
K
300
N
300
300
V

FIGURE 6

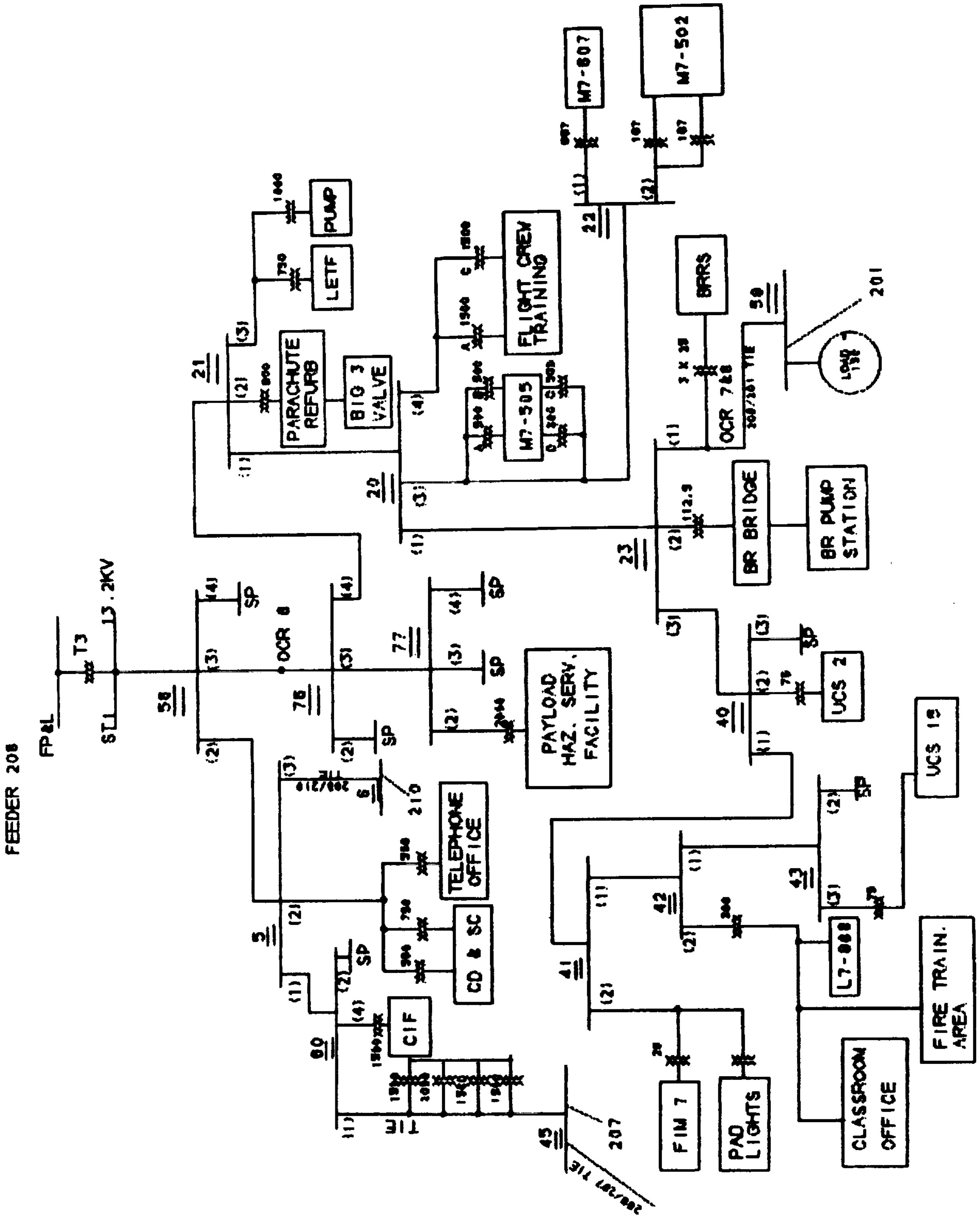

FEEDER 208
FP&L
ST1
T3
13.2KV
58
OCR 8
76
77
SP
PAYLOAD HAZ. SERV. FACILITY
5
6
210
208/210
TELEPHONE OFFICE
CD & SC
CIF
60
45
207
208/207 TIE
21
PARACHUTE REFURB
BIG 3 VALVE
LETF
PUMP
20
M7-505
FLIGHT CREW TRAINING
22
M7-607
M7-502
23
BR BRIDGE
BR PUMP STATION
BRRS
OCR 7&8
208/201 TIE
59
201
40
UCS 2
UCS 19
41
42
43
L7-068
FIRE TRAIN. AREA
CLASSROOM OFFICE
FIM 7
PAD LIGHTS

FIGURE 7

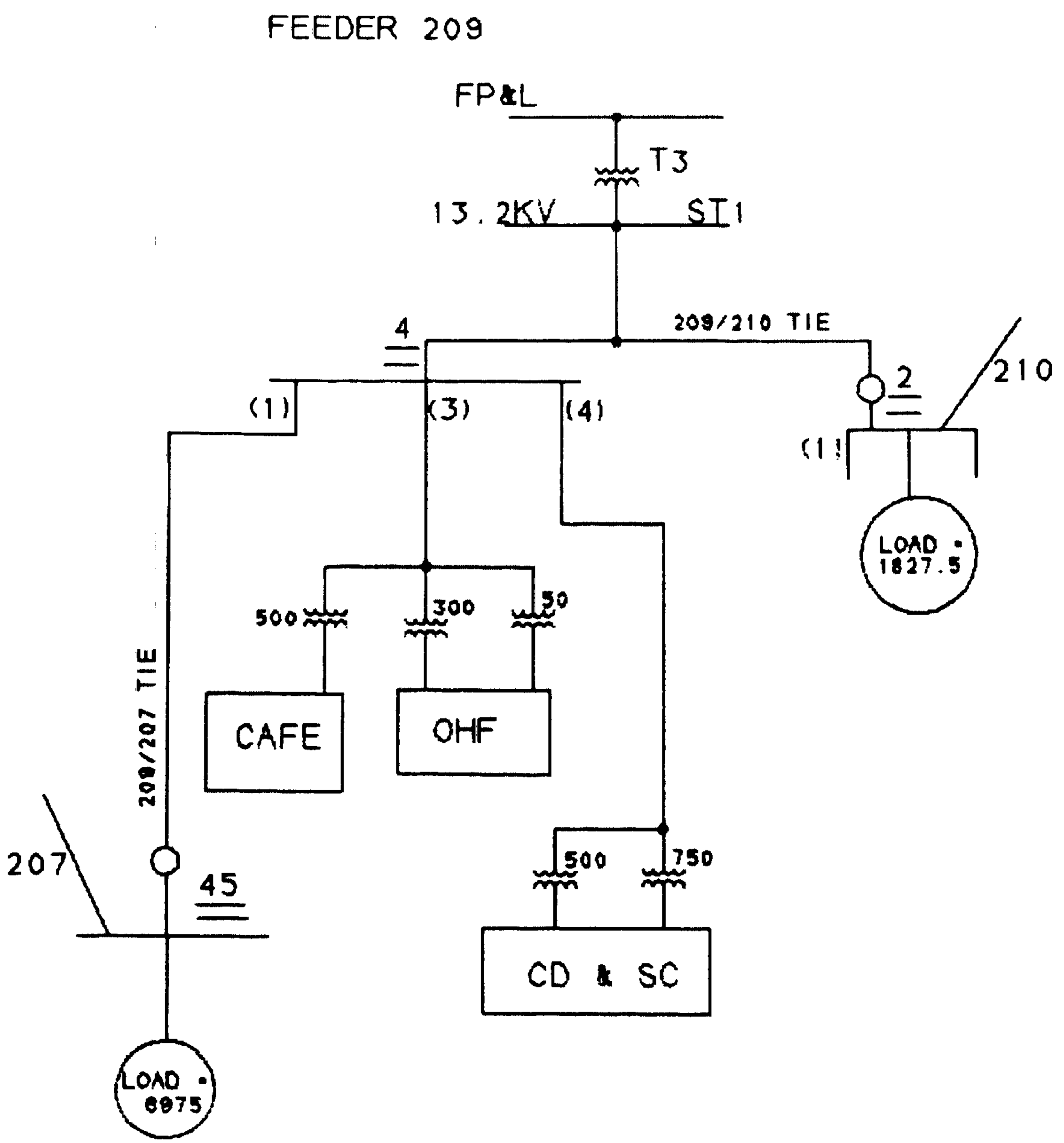

FEEDER 209
FP&L
T3
13.2KV
ST1
209/210 TIE
4
2
210
(1)
(3)
(4)
(1)
LOAD = 1827.5
500
300
50
CAFE
OHF
209/207 TIE
207
45
500
750
CD & SC
LOAD = 6975

FIGURE 8

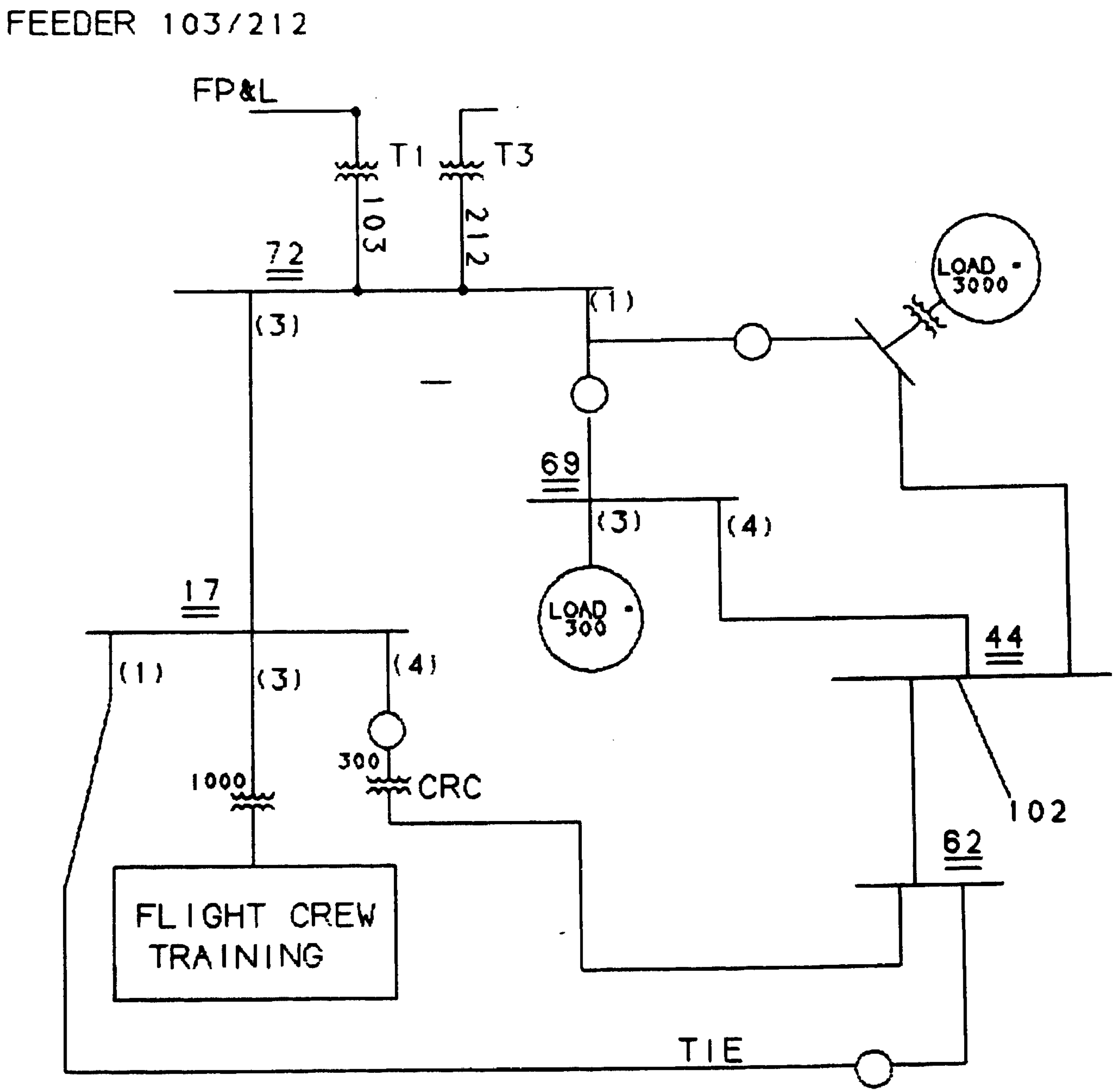

FEEDER 103/212
FP&L
T1
T3
103
212
72
(3)
(1)
69
(3)
(4)
LOAD • 300
LOAD • 3000
17
(1)
(3)
(4)
1000
300
CRC
44
102
62
FLIGHT CREW TRAINING
TIE

FIGURE 9

FEEDER 204/205

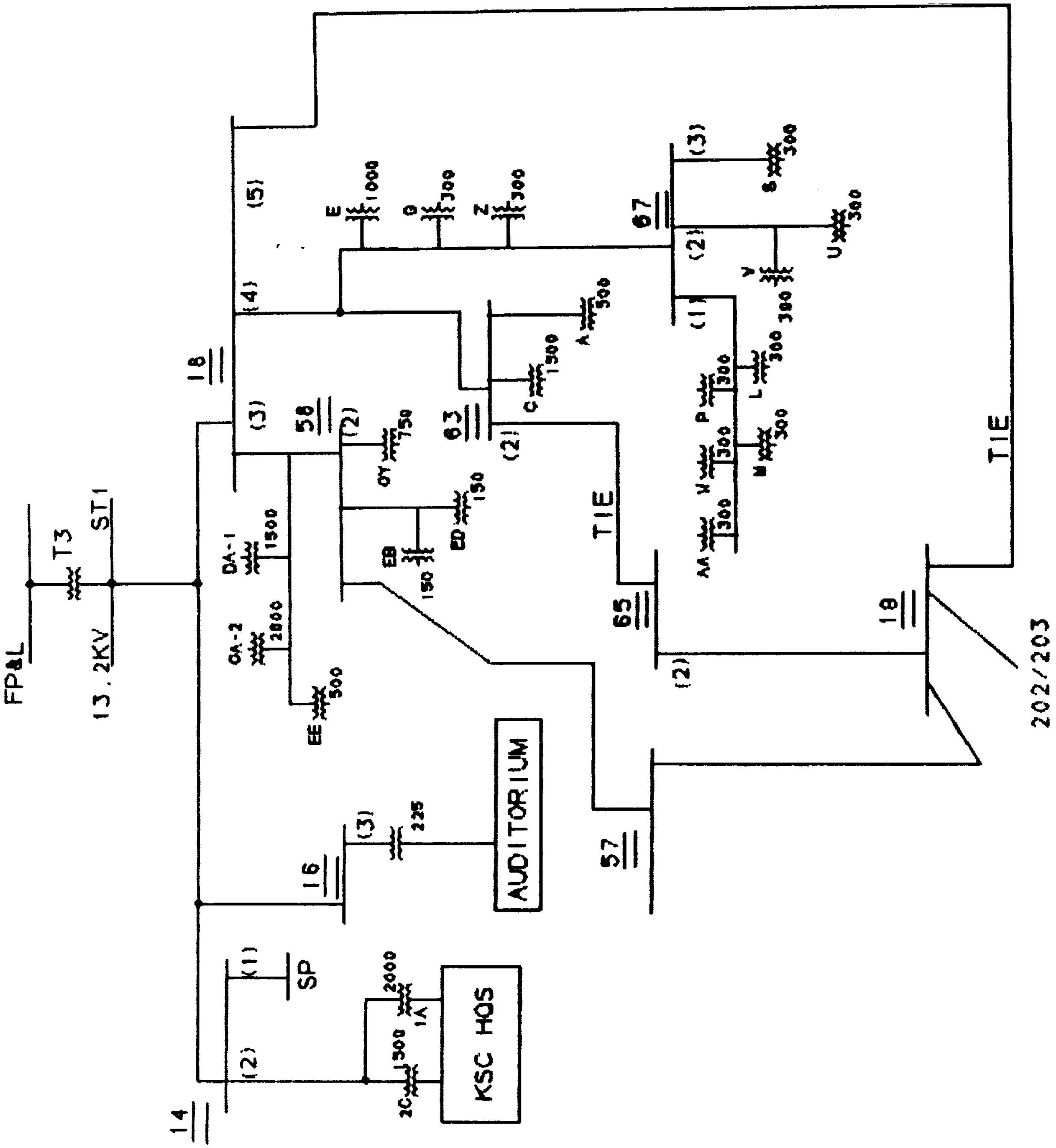


FIGURE 10

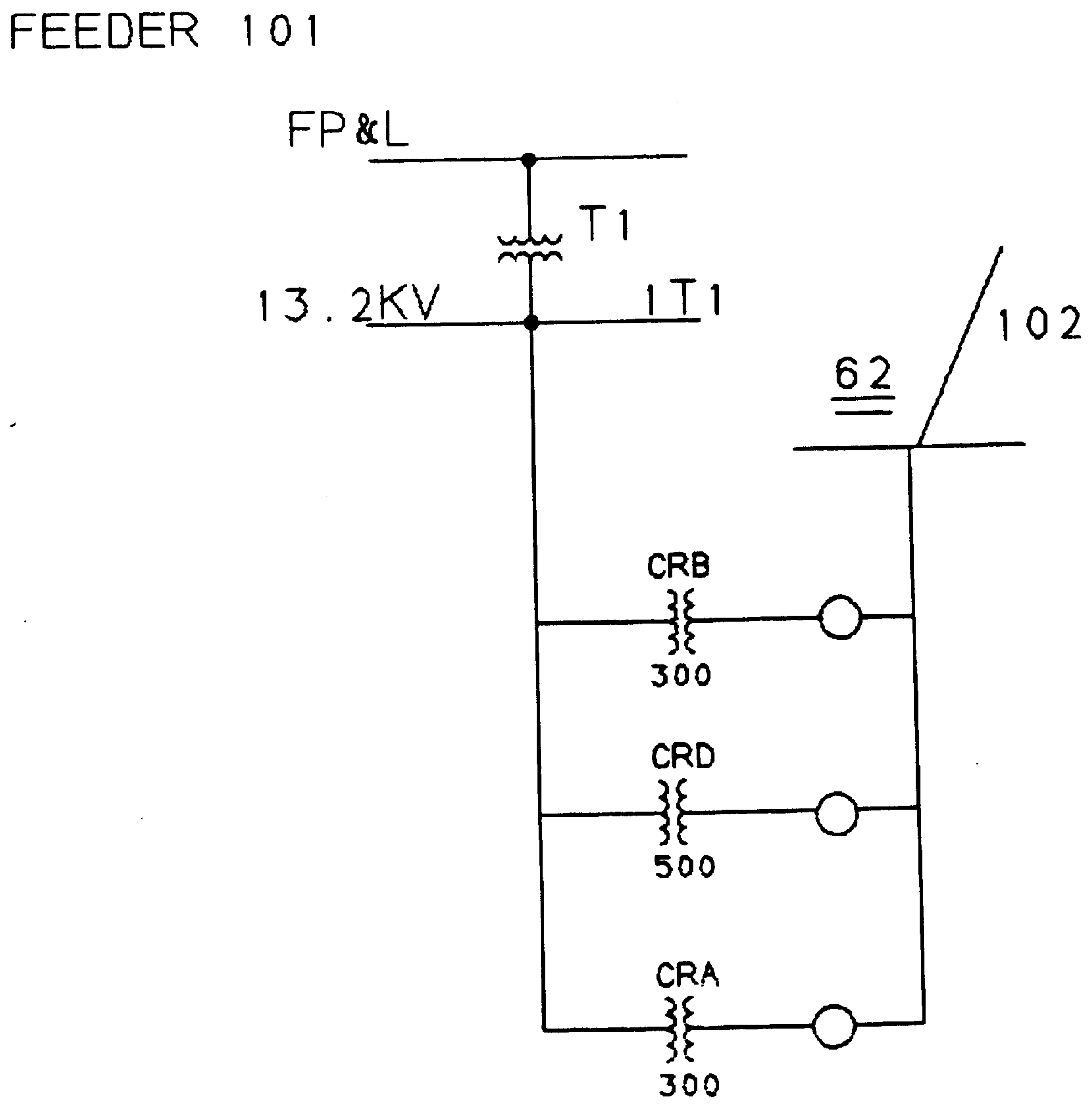


FIGURE 11

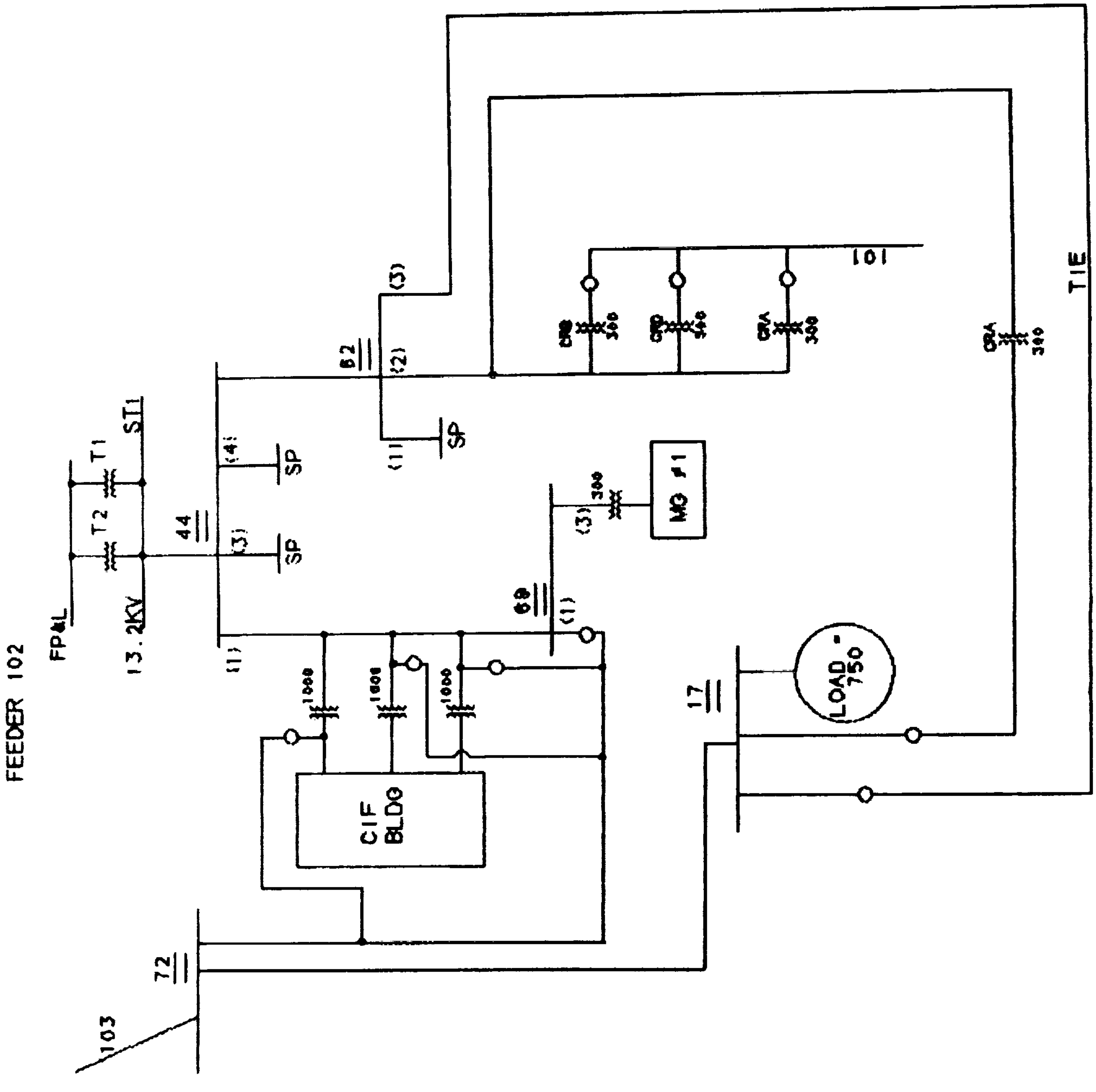

FEEDER 102
FP&L
13.2KV
T2
T1
ST1
44
(1)
(3)
(4)
SP
SP
82
(2)
(3)
(1)
SP
101
CRB 300
CRD 300
CRA 300
CRA 300
TIE
68
(1)
(3)
300
MG #1
1000
1000
1000
CIF BLDG
LOAD - 750
17
72
103

FIGURE 12

TABLE I

NORMAL CONFIGURATION LOADS ON THE LOAD BREAK SWITCHES AT THE ORSINO SUBSTATION

Lbs #	Supply Feeders	Adjacent LBS #s	Total Connected Load Under Normal Configuration	Added Load-if Failure Occurs Above LBS # - on Feeder	Special Features, Comments
	Feeder 206				
48	206 211 VIA LBS #47 609 VIA LBS #47	61, 47	15	Only single failures are considered. 112.5-61-206 612.5-OCR10-206	Tie point of 206 to 211 and 609
61	206	54, 58	127.5	500-OCR10-206	
55	206	54	2265		
15	206	54	2750		
54	206	T6,15, 55,61	5642.5		This is feeder 206.
	Feeder 211				
37	211 VIA OCR 14 201 VIA OCR 13	36,51	1900		This is a load bus. Connected load is shared by feeders 211 and 201.
36	211	10,37,51	15 1900[1]	300-51-201	Tie point of 211 to 201.
10	211	52,36	217.5	300-51-201	
52	211	9,10	480 1900[1]	300-51-201	
47	211,206,609 VIA REG	73,48,730		15-73-211 90-53-211 15-48-206 127.5-61-206 627.5-OCRIO-206 (C-5 not studied)	This is tie bus. 206/211 tie; and ORSINO/ C-5 tie.
73	211	53,47	15		

53	211	9.73	90		
9	211	T5,53, 52	570 1900[1]	300-51-201	This is feeder 211.

Feeder 207

60	207, 208 VIA LBS#5	45,5	1500	Only single failures are considered.	Tie bus for 207 to 208.
45	207, 209 VIA LBS#4	T5,60,4	7975	Only single failures are considered.	Tie point of 207 to 209. This is feeder 207.

Feeder 201

35	201	59,34	2500		Load bus end of 201
59	210 VIA OCR 12 208 VIA OCR 7	34,23,35	150, 2500[2] 325[3]	587.5-23-208 6388.5-20-208 8638.5-21-208 10638.5-76-208 ~~225-34-201~~	Tie bus., 201/208 tie
34	201, ~~208 VIA OCR 11~~	28,59,35	225 ~~2500~~[4]	150-59-208 ~~587.5-23-208~~ 6388.5-20-208 8638.5-21-208 10638.5-76-208	
37	201 VIA OCR 13 211 VIA OCR 14	51,36	1900		Load bus shared by feeders 201 and 211.
51	201, 211 VIA OCR 13	28,37,36	300 1900[5]	15-36-211 217.5-10-211 262.5-52-211	Tie point of 201 to 211
28	201	27,51,34	3475 1900[5] 150[6]	SEE LBS#59 SEE LBS#51	Failure above LBS#28 maybe assisted by 211 VIA LBS#36 or land by 208 VIA LBS#23.
27	201	78,28	5725, 1900[5] 150[6]	Same as for LBS#28.	Same as for LBS #28
78	201	26,27	5800 1900[5] 150[6]	Same as for LBS#27.	Same as for LBS #27

26	201	25,78	6525,78 1900[5] 150[6]	Same as for LBS#78.	Same as for LBS #78
25	201	24,26	7525 1990[5] 150[6]	Same as for LBS#26.	Same as for LBS #26
*24	201	74,25	8275 1900[5] 150[6]	Same as for LBS#25.	Same as for LBS#25
75	201	74	2000		
74	201	39,75,24	10275 1900[5] 150[6]	Same as for LBS#24.	Same as for LBS$24
39	201	T5,74	10775 1900[5] 150[6]	Same as for LBS#74.	This is feeder 201.
	Feeder 210				
6	210, 208 VIA LBS#5	7,5	300		Tie point of 210 to 208
7	210	46,6	1475		
46	210	3,7	1887.5		
3	210	49,7	2187.5		
49	210	1,3	2532.5		
12	210	11	800		
11	210	2,12	912.5		
2	210, 209 VIA LSB#4	1,11	1827.5	Only singlue failures are considered.	Tie point of 210 to 204
1	210	T4,2,49	4360		This is feeder 210
	Feeder 202/203				
68	202/203	19	2500		
65	202/203, 204/205 VIA LBS#63	19,63	1500	2000-63-204/205	

57	202/203, 204/205 VIA LBS#58	19,58	950	1050-58-204/205	
19	202/203, 204/205 VIA LBS#18	T4,57, 65.68,18	8175	7850-18-204/205	
13	202/203	T4	4000		
	Feeder 208				
43	208	42	75		Load Bus, end of 208
42	208	41,43	375		
41	208	40,42	400		
40	208	23,41	475		
23	208, 201 VIA OCR 8	20,40,59	587.5 325[7]	150-59-201 2650-OCR11-201 2875-34-201	Tie point of 208 to 201
22	208	20	1001		
20	208	21,22,23	6388.5 325[7]	Same as for LBS#23	
21	208	76,20	8638.5 325[7]	Same as for LBS#21	
77	208	76	2000		
76	208	56,77,21	10638.5 325[7]	Same as for LBS#21	
60	208, 207,209 VIA LBS#45	5,45	1500	Only single failures are consicered	Tie point 208 to 207 VIA LBS#45. Failure abover LBS#60 on208 can be assisted by 207 and/or 209.
5	208, 210 VIA LBS #6	56,60, 6	1750	300-6-210 1475-7-210 1887.5-46-210 2187.5-3-210 2532.5-49-210	Failure above LBS#5 on 208 can be assisted by 210 VIA LBS#6 or by 207; 209 VIA LBS#45.

56	208	T3,5,76	11638.5 3257	SEE LBS#23, LBS#5	This is feeder 208.

Feeder 209

4	209, 210 VIA LBS#2 201 VIA LBS#45	T3,2,45	2100	6975-45-207 1827-5-2-210	This feeder is mostly a redundancy for feeders 210 and 207.

Feeder 212/103

17	103,212, 102 VIA LBS#62	72,62	1000	300-62-102	
72	103,212, 102 VIA LBS#69	T1,T3, 17,69	1000	300-62-102 300-69-102 3300-44-102	This is feeder 212 in parpallel with feeder 103.

Feeder 204/205

67	204/205	18	2400		
63	204/205, 202/203 VIA LBS#65	18,65	2000	1500-65-202/203	
58	204/205, 202/203 VIA LBS#57	18,57	1050	950-57-202/203	
18	204/205, 202/203 VIA LBS#19	T3,58 63,67,19	8750	8075-19-202/203	
16	204/205,	T3	225		
14	204/205	T3	3500		
	Feeder Total		12475		Feeder 204/205

Feeder 101

T1	101, 102 VIA LBS#62	62	1100	Only single failures are considered.	

Feeder 102

62	102, 103 VIA LBS#17	44,17	300	1100,101 750-17-103	

69	102, 103 VIA LBS #72	44,72	300	Only single failures are considered.
44	102	T1,T2, 69,62	3600	

IV RESULTS AND DISCUSSION

The information obtained from this investigation is presented in table 2 through 5. This can form the bases for a decision on where any additional load to the system can be made. Much of the information such as comparing demand to connected load are relevant only to the extent that it can help the system planning engineers to shape his/her engineering judgement. It is also important that the power coordinator pays attention to this data not necessarily to achieve higher system reliability, but to assure system optimal use and longivity.

To err on the side if safety, conservative assumptions are used to estimate quantities when there is no data to enforce precision. For instance, if a load is shared by two feeders, the shared load is added in full to compute the total load connected to each of the feeders. Also, since there is no demand data for the primary transformers the demands for the feeders connected to each primary transformer is added to estimate its demand as if those feeders peak simultaneously. Once more, it should be noted that all data are based on the system normal configuration and thus results obtained depict the worst case situations. The actual system reliability status is improved by the presence of redundancies and assist tie lines.

TABLE 2

CONNECTED LOAD -- DEMAND TABLE FOR FEEDERS

FDR	CONNECTED LOAD	DEMAND AUGUST 1988	MARGIN	DEMAND AUGUST 1989	MARGIN
206	5642.5	1750.60	3891.90		
211	570 1900[1]	976.97	1493.03		
207	7975	1188.55	6786.95		
201	10775 1900[5] 150[6]	2523.96	10301.04		
210	4360	1498.74	2861.26		
202/203	12175	4006.33	8168.67		
208	11638.5 325[7]	2840.08	9123.42		
209	2100	853.54	1246.46		
212/103	1000	151.67	848.33		
204/205	12475	4417.27	8057.73		
101	1100	8.67	1091.33		

TABLE 3

CONNECTED LOAD -- DEMAND TABLE FOR PRIMARY TRANSFORMERS

T#	CONNECTED LOAD	DEMAND AUGUST 1988	MARGIN	DEMAND AUGUST 1989	MARGIN
T6	5642.5	1750.60	3891.90		
T5	21370	4689.48	16680.52		
T4	16535	5505.07	11029.93		
T3	27538.5	8117.02	19421.48		
T2+T1	5700	617.77	5082.23		

TABLE 4

PRIMARY TRANSFORMERS: SIZE VS CONNECTED LOAD

T#	SIZE	CONNECTED LOAD	MARGIN
T6	10000/12500	5642.5	4357.5
T5	10000/12500	21370	-11370
T4	10000/12500	16535	-6535
T3	10000/12500	27538.5	-17538.5
T1+T2	5000	5700	-700

TABLE 5

PRIMARY TRANSFORMERS: SIZE VS DEMAND TABLE

T#	SIZE	DEMAND AUGUST 1988	MARGIN	DEMAND AUGUST 1989	MARGIN
T6	10000/12500	1750.60	8249.40		
T5	10000/12500	4689.48	5310.52		
T4	10000/12500	5505.07	4494.93		
T3	10000/12500	8117.02	1882.98		
T1+T2	2500+2500	617.77	4382.23		

Orsino (Industrial Area) Substation Total Data:

Capacity = 45000 KVA

Connected Load = 76786 KVA

Capacity-Load Margin = - 31786 KVA

Demand: August 1988 = 20679.94 KVA

Capacity-Demand Margin = 24320.06 KVA

The above information showing demands (also in tables 3 and 5) is based on the assumption that peaks on all the feeders occur at the same time which is a very unlikely event. It is only used to estimate the worst case scenario. In fact, the system history reveal that no two feeder demands (Peaks) have ever occurred at the same time.

V CONCLUDING REMARKS

The data obtained during this study show that the industrial area (Orsino) Substation is healthy and reliable. There is a substantial margin between capacity and demand. History also does not show any need for additional protective devices or further redundancies.

However, the data reveals that loads have been added arbitrarily or perhaps based on geographic convenience (or associated costs) only, during growth. This philosophy does not lead to optimal system performance. The reduction or even elimination of outages should not be the only concern in power system operation. Well coordinated power and proper load distribution enhance several good results. The information such as contained in this report should be useful for making such decisions.

It will be desirable if a similar study is performed on the C-5 substation and thus on the entire 60-HZ power system. This could form the basis for a formal power systems planning here at K.S.C. This is needed in a rapidly growing facility such as K.S.C. With anticipated ever growing load, arbitrariness may lead to very poor results.

The initial intentions of this researcher was to do a classical load flow analysis and reliability analysis on the 60-HZ power system at K.S.C. The system which was constructed over twenty five years ago has undergone a lot of growth. Numerous load break switches have been added resulting in a new configuration. The diagrams shown in this report are based on the present system configuration.

However, data on the present form, such as the impedances of the lines from bus to bus are still being constructed. This made it impossible to perform a classical power flow analysis on the system. Further, the contractor, EG&G, that operates the system recently acquired a new package EDSA which is capable of performing such an analysis. Also, lacking at present is a complete inventory of component reliability parameters such as failure rates and repair rates. These are essential inputs for a classical system reliability analysis. These classical studies could not be performed at this time for the above reasons.

With the system being as healthy and reliable as it is; the capacity - demand margin being very substantial and history showing no reason for immediate concern. Attention can be sshifted to the effects of future growth. None the less, the author would still suggest that meter readings be taken at the Primary transformer levels and isolation devices be placed above several load break switches.

REFERENCES

1. GP-900 Revision A - 60 Hz Power Systems Integrated Plan - Prepared by Planning Research Corporation Systems Services Company, May 1983.

2. 79K17429 - KSC Power Distribution One-line Diagram - Prepared by EG&G Florida, September 1987.

N90-16693

1989 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

FACTOR ANALYTIC REDUCTION OF THE CAROTID-CARDIAC BAROREFLEX PARAMETERS

PREPARED BY:	Dr. David A. Ludwig
ACADEMIC RANK:	Associate Professor
UNIVERSITY AND DEPARTMENT:	University of North Carolina - Greensboro Mathematics Department
NASA/KSC	
DIVISION:	Biomedical Research
BRANCH:	Life Sciences Research
NASA COLLEAGUE:	Dr. Victor A. Convertino
DATE:	July 28, 1989
CONTRACT NUMBER:	University of Central Florida NASA-NGT-60002 Supplement: 2

ABSTRACT

Nine carotid-cardiac baroreflex parameters were measured on 30 middle aged human males and subsequently factored in an effort to determine the underlying dimensionality of the parameters. The results indicated that the variation in the nine variables could be explained in four dimensions with only a seven percent loss of information.

SUMMARY

An accepted method for measuring the responsiveness of the carotid-cardiac baroreflex to arterial pressure changes is to artificially stimulate the baroreceptors in the neck. This is accomplished by using a pressurized neck cuff which constricts and distends the carotid artery and subsequently stimulates the baroreceptors. Nine physiological responses to this type of stimulation are quantified and used as indicators of the baroreflex.

Thirty male humans between the ages 27 and 46 underwent the carotid-cardiac baroreflex test. The data for the nine response parameters were analyzed by principle component factor analysis. The results of this analysis indicated that 93 percent of the total variance across all nine parameters could be explained in four dimensions. Examination of the factor loadings following an orthogonal rotation of the principle components indicated four well defined dimensions. The first two dimensions reflected location points for R-R interval and carotid distending pressure respectively. The third dimension was composed of measures reflecting the gain of the reflex. The fourth dimension was the ratio of the resting R-R interval to R-R interval during simulated hypertension.

The data suggests that the analysis of all nine baroreflex parameters is redundant and future analyses should be performed on an unweighted linear composite of the variables that make up each of the four underlying dimensions. An alternative to an unweighted composite would be the selection of one parameter from each of the four principle components.

TABLE OF CONTENTS

I. INTRODUCTION

The carotid baroreceptors are afferent arterial pressure receptors located in the walls of the carotid artery. These receptors respond to changes in arterial blood pressure and initiate various autonomic responses in order to maintain blood pressure and profusion to the brain (1). The carotid baroreflex responds to theses baroreceptors by altering heart rate to compensate for increases or decreases in arterial pressure. It is believed that orthostatic hypotension results when this reflex mechanism is attenuated (2). The symptoms of this condition include light headiness and in the extreme case, fainting (intolerance).

Astronauts who have been exposed to microgravity usually experience some degree of orthostatic hypotension after returning to a one G environment (3). It has been hypothesized that one of the major reason this occurs is due to a loss of the responsiveness of the baroreflex (2). The reflex is believed to be impaired during space flight since cranial blood flow is not challenged by the force of gravity.

An accepted method for evaluating the responsiveness of the carotid-cardiac baroreflex involves the use of a pressure cuff which is placed around the neck providing an artificial stimulus to the carotid artery (4). When negative pressure is applied, the carotid artery distends providing a high pressure stimulus (hypertension) to the baroreceptors. A low pressure stimulus (hypotension) is provided when a positive pressure is applied and the carotid artery is constricted. Sequential changes in pressure are compared to changes in beat-to-beat (R-R) interval providing an index of baroreceptor responsiveness. The responsiveness is quantified by nine variables which characterize the relationship between neck cuff pressure and R-R interval.

The nine measures of baroreceptor responsiveness were chosen based on physiological considerations. There psychometric properties have only been evaluated on a limited basis (5). One such psychometric property is redundancy. It's possible that these nine measures can be expressed in fewer than nine dimensions with only minimal loss of information. If this were the case, greater efficiency in data collection and experimental analysis would be possible. It is therefore the purpose of this study to evaluate the intercorrelation structure of these nine variables and if possible, explain this structure in fewer than nine dimensions with a minimal loss of information.

II. METHODS

2.1 Subjects

Thirty healthy nonsmoking normotensive men between the ages of 27 and 46 gave written informed consent for the baroreflex test. Selection of subjects was based on the results of a clinical screening comprised of a detailed medical history, physical examination, urinalysis, complete blood count and chemistry, glucose tolerance test, chest X-ray, resting and treadmill electrocardiograms, and psychological evaluation. None of the subjects were taking prescription medication at the time of the study.

2.2 Experimental

Prior to testing and data collection, each subject was given an explanation of the baroreflex testing procedure and familiarized with the protocol. Carotid-cardiac baroreflex stimulus was delivered via a computer controlled motor driven bellows which provided pressure steps to a Silastic neck chamber. During held expiration, a pressure of about 40 mmHg was delivered to the chamber and held for about 5 seconds. Then, with the next R wave (heartbeat), the pressure sequentially stepped to about 25, 10, -5, -20, -35, -50, and -65 mmHg followed by a return to ambient pressure. Pressure steps were triggered by R-waves so that neck chamber pressure changes were superimposed upon naturally occurring carotid pulses. During each testing session the stimulus sequence was applied seven times and the data averaged for each subject. Subjects mean R-R interval for each pressure step was plotted against carotid distending pressure (resting systolic blood pressure minus neck chamber pressure) to produce a carotid-cardiac baroreceptor response function (Figure 1). Nine characteristics (parameters) of this response function were calculated and used for statistical analysis.

2.2.1 BAROREFLEX PARAMETERS. The nine baroreflex parameters are listed below and illustrated in Figure 1.

(1) Minimum R-R Interval
(2) Maximum R-R Interval
(3) R-R Interval at Baseline
(4) Range of the R-R Interval
(5) Maximum Slope of the Response Function
(6) Minimum Carotid Distending Pressure
(7) Maximum Carotid Distending Pressure
(8) Carotid Distending Pressure at Maximum Slope
(9) Ratio of Resting R-R Interval Minus Minimum R-R Interval to R-R Range

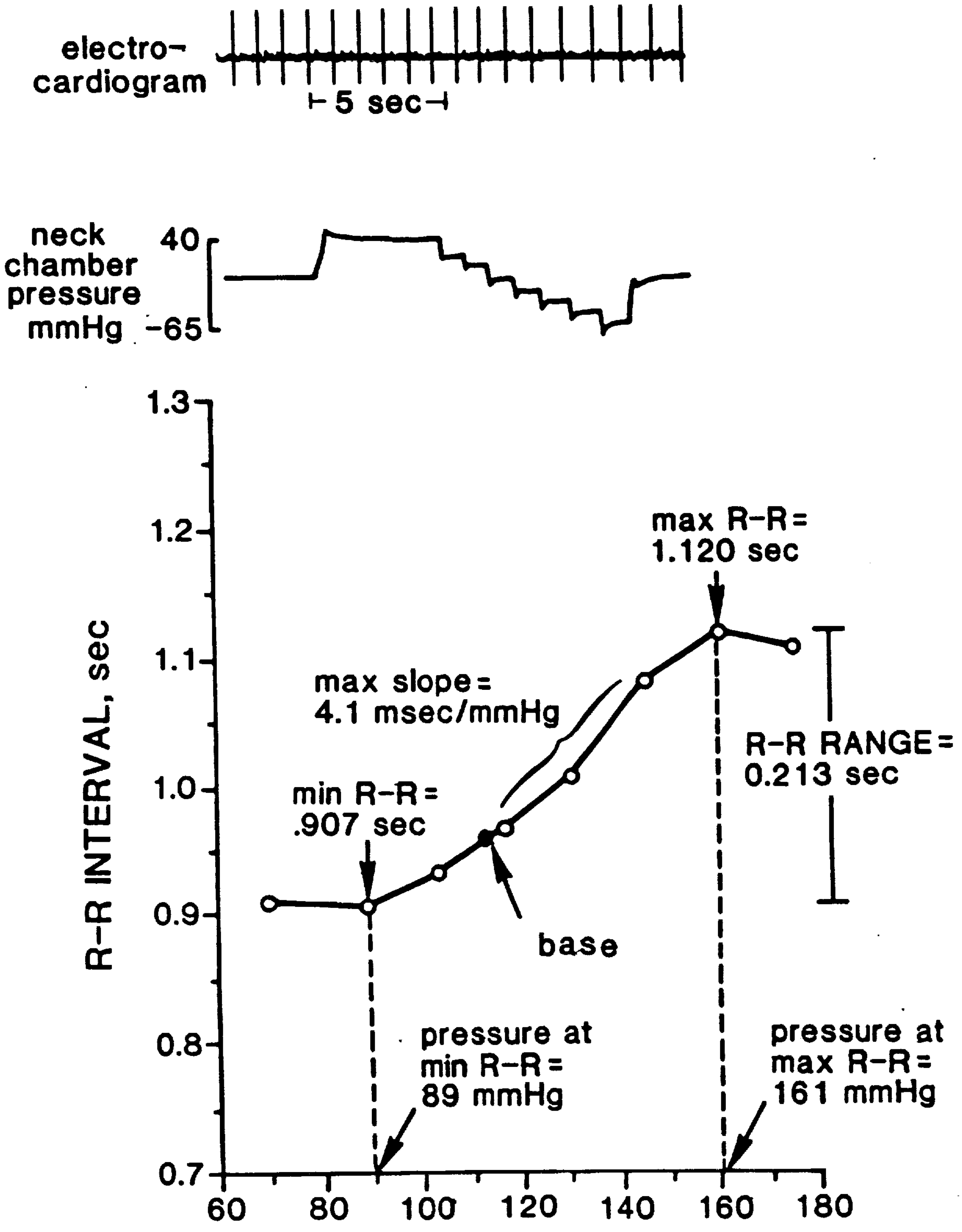


Figure 1. Baroreflex Response Function

Maximum slope was determined by least squares estimation of every set of three consecutive points on the response function. The slope estimate taken from the segment with the steepest slope was retained for statistical analysis. The carotid distending pressure at maximum slope was the point halfway between the pressures which defined the maximum slope.

2.3 Statistical

The 30 subject by nine variable data matrix was subjected to principle component analysis followed by an orthogonal rotation (varimax) of the initially extracted components. The number of components was determined by the solution which produced a simple structure of the factor loadings with a minimum number of factors needed to account for a majority of the variance in the nine parameters.

III. RESULTS AND DISCUSSION

Evaluation of the eigenvalues from the unrotated principle components, percent of explained variance, and rotated factor loadings indicated that a four factor solution was the most parsimonious. The four factor solution explained 92.5 of the of the variance in the original nine variables, and had a well defined (simple) structure. Table I presents the rotated factor loadings, post rotation eigenvalues and percent of explained variance for the four factor solution.

Table I. Rotated Factor Pattern

Parameter	Factor I	Factor II	Factor III	Factor IV
(1) RRMIN	.970*	-.137	-.071	-.169
(2) RRMAX	.921*	-.093	.340	-.151
(3) RRBASE	.968*	-.210	.015	.078
(4) RRRANG	.088	.075	.978*	.008
(5) MAXSLP	.058	-.187	.955*	.006
(6) CDPMIN	-.288	.791*	-.190	-.027
(7) CDPMAX	-.150	.887*	-.036	-.029
(8) CDPSLP	-.004	.896*	.075	-.265
(9) RRRATO	-.139	-.206	.015	.965*
Eigenvalue	2.86	2.37	2.03	1.06
% Variance	31.78	26.33	22.56	11.78
Cum % Var.	31.78	58.11	80.67	92.45

Interpretation of the factors seems relatively straight forward and makes good theoretical sense. Factor I is a location factor for R-R interval. It reflects the position of the baroreflex function on the y-axis. Factor II is the location factor for carotid distending pressure. It reflects the position of the baroreflex function on the x-axis. Factor III is a gain factor reflecting the change or responsiveness of the reflex. And factor IV is a single variable reflecting the percentage of the R-R range falling below the baseline point.

Since the factor loadings for this four factor solution are either very high or very low, a single variable from each factor can be used to account for the factor variance in lieu of an unweighted linear composite such as a sum. Selection of this one variable from each factor group is probably best left to the researcher. Previous research has indicated that the test-retest reliabilities for baseline R-R interval and maximum slope are slightly higher than the other possible factor representatives for factors I and III respectively (5). If carotid distending pressure were measured at baseline, this parameter would reduce to resting systolic blood pressure and provide a simpler measure of location on the x-axis. R-R ratio is a unique variable measuring the percent of the reflex function below the baseline point. Thus the only "true" baroreflex parameters are maximum slope and R-R ratio since factor I reflects heart rate and factor II reflects systolic pressure.

Certainly the slope of the reflex function is the best indicator of baroreflex responsiveness. Although the reliability of this parameter is already quit high (.92) (5), it could probably be improved by using all of the data points between minimum R-R interval and maximum R-R interval with very little change in the value of the parameter. This is true given the near linear response function for the baroreflex. When all the data points are used to calculating slope, the slope will be more stable even if one or two points tend to be extreme or out of range. In fact, the current calculation of maximum slope tends to seek out these extreme values. The practice of using the average of seven trials does smooth out the effects of possible outlying values.

IV. CONCLUSIONS

Researchers measuring the carotid-cardiac baroreflex need not measure or analyze all nine of the parameters calculated from the baroreflex response function. Nearly all of the variance in the original nine parameters can be accounted for by baseline R-R interval, a measure of carotid distending pressure, maximum or full slope of the response function, and the operational point (R-R ratio).

V. REFERENCES

1. Watson, R.D.S., Stallard, T.J., Flinn, R.J., & Littler, W.A. Factors determining direct arterial pressure and its variability in hypertensive men. Hypertension, 2: 331-341, 1980.

2. Convertino, V.A., Doerr, D.F., Eckberg, D.L., Fritch, J.M., & Vernikos-Danellis, J. Carotid baroreflex response following 30 day exposure to simulated microgravity. The Physiologist, 32: Suppl., 67-68, 1989.

3. Bungo, M.W., Charles, J.B. & Johnson, P.C. Cardiovascular deconditioning during space flight and the use of saline as a countermeasure to orthostatic intolerance. Aviation Space and Environmental Medicine, 56: 985-990, 1985.

4. Sprenkle, J.M., Eckberg, D.L., Goble, R.L., Scelhorn, J.J., & Halliday, H.C. Device for rapid quantification of human carotid baroreceptor-cardiac reflex responses. Journal of Applied Physiology, 60: 727-732, 1986.

5. Doerr, D.F., & Convertino, V.A. A technique for reproducible measurement of the carotid-cardiac baroreflex in man (Abstract). Aviation, Space and Environmental Medicine, 60: 508, 1989.

N90-16694

1989 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

MODELING THE NEAR ACOUSTIC FIELD OF A ROCKET DURING LAUNCH

PREPARED BY:	Dr. David W. Mauritzen
ACADEMIC RANK:	Assistant Professor
UNIVERSITY AND DEPARTMENT:	Indiana-Purdue University - Fort Wayne Department of Engineering
NASA/KSC	
DIVISION:	Mechanical Engineering
BRANCH:	Special Projects
NASA COLLEAGUE:	Dr. Gary Lin
DATE:	July 28, 1989
CONTRACT NUMBER:	University of Central Florida NASA-NGT-60002 Supplement: 2

ACKNOWLEDGEMENTS

I would like to thank NASA for sponsoring this program which allows faculty to participate in the space program. This experience is of both professional and personal interest to me, and will be of great interest to our students.

More specifically, I would like to thank my NASA colleague Gary Lin and branch manager Willis Crumpler for this opportunity.

The program was professionally managed by Professor E. R. Hosler of UCF, who has worked very hard to make the program successful and enjoyable. He has succeeded admirably.

I also appreciate the support given me by Josué Njock-Libii, Acting Engineering Department Chairman, and our Dean, Aly Mahmoud.

Additionally, I have enjoyed the tours and meetings which have been provided to afford us information on the nature and direction of the NASA program as well as interaction with other faculty members. In particular, I would like to thank Professor John M. Russell of the Florida Institute of Technology for his comments on this project.

ABSTRACT

The design of launch pad structures is critically dependent upon the stresses imposed by the acoustical pressure field generated by the rocket engines during launch. The purpose of this effort is to better describe the acoustical field in the immediate launch area. Since the problem is not analytically tractable, empirical modeling will be employed so that useful results may be obtained for structural design purposes. The plume of the rocket is considered to be a volumetric acoustic source, and is broken down into incremental contributing volumes. A computer program has been written to sum all the contributions to find the total sound pressure level at an arbitrary point. A constant density source is initially assumed and the acoustic field evaluated for several cases to verify the correct operation of the program.

SUMMARY

This report documents an effort to model the acoustic near field of a rocket. The generating mechanisms are sufficiently complex that an empirical approach was found to be potentially more useful.

It was assumed that the acoustic source could be modeled as a collection of volumetrically distributed sources and that the acoustic field at any point could then be evaluated by incoherently summing contributions from elemental volumes.

A computer program was written to perform the numerical integration. The program is divided into logical sections and therefore is readily modifiable and extensible.

Test runs were made to verify the correct operation of the program. The results indicate linearity with respect to the source intensity of an arbitrary single cell and additivity with respect to a randomly selected pair of cells. For large ranges, the field decays with inverse square law behavior as does a point source.

SECTION I

INTRODUCTION

1.1 PROJECT DEFINITION

The intent of this effort is to develop a model for the general properties of the near acoustic field of rocket powered vehicles during launch. The lower portion of the acoustic spectrum, nominally less than 100 Hz, is of particular interest since mechanical structures respond primarily to these frequencies. In addition, the maximum spectral density typically occurs within this band.

1.2 PURPOSE

Acoustic power levels in the vicinity of a rocket engine exhaust can approach levels which are physically damaging to support structures, facilities, and even to the payload of the vehicle. Consequently the estimation of the sound pressure level (SPL) is important for future designs.

1.3 BACKGROUND

The acoustic energy generated by a rocket is principally due to the turbulent mixing of its high velocity exhaust gasses with the atmosphere. Although turbulence has been of interest since the late nineteenth century, as indicated by the work of Raleigh (1, 2) and Reynolds (3,4), little interest was shown in the acoustic field generated by jet flow until the work of Lighthill (5) in 1952. The far field analysis was extended by Williams (6) to include the Doppler shift induced by the advection of sound. Moffatt (7) and Goldstein (8) have pointed out the potential importance of the interaction of large scale fluctuations and small scale turbulence with respect to sound generation. Experimental work has been performed by Mollo-Christensen and Narasimha (9), Mollo-Christensen, Kolpin, and Martuccelli (10), Liu (11), Michalke and Fuchs (12), and Browand and Weidman (13). Progress in the field has been reviewed by Ffwcs Williams (14) and Goldstein. (op cit)

SECTION II

ANALYTICAL EFFORTS

2.1 PROBLEM DEFINITION

The generation of acoustic noise by a jet was considered in a landmark paper by Lighthill in 1952. Under the assumptions that the perturbed pressure and density variations are small compared to their mean value, the variation of entropy is negligible, and that the turbulent Mach number is small, his theoretical work led to a second order equation for density which has the form of a non-homogenous wave equation involving the instantaneous Reynolds stress tensor. He finds the acoustic source to have the form of acoustic quadrupoles, and finds approximate expressions for the spatial and temporal variations of the density and the time autocovariance of the density fluctuation.

Unfortunately, the statistics of the instantaneous Reynolds stress are not known. Therefore the author is ultimately forced to resort to similarity arguments to extend the analysis to arbitrary jet diameters and velocities; he finds that the mean squared density fluctuation varies with the square of the jet diameter and the eighth power of the Mach number.

2.2 LIMITATIONS

In addition to the assumptions already stated in the development of the theory, it is also assumed that the nature of the acoustic source does not change with Mach number and that the problem is indeed separable into a turbulence problem and an acoustic generation/propagation problem. It is not obvious that these assumptions are valid.

Even if one accepts all the assumptions involved, the analysis has been made only for the far acoustic field, while the current engineering interest is in the near field.

SECTION III

PROGRAM

3.1 APPROACH

In order to obtain an engineering approximation for the near acoustic field strength within a reasonable time, it will be necessary to resort to empirical techniques. Since the principal acoustic noise source of a rocket is associated with the plume of the rocket where its exhaust gasses mix with the atmosphere, this turbulent mixing region will be modeled as a volumetric noise source.

In the program developed herein, the source region has been broken into elemental volumes, source strengths associated with them, and the total sound field then estimated by evaluating the propagation loss from the elemental volume to the observer and forming an incoherent sum of these contributions. Since the source region has the approximate shape of a truncated cone, it has been most convenient to utilize a cylindrical coordinate system.

The program has been divided into functional modules so that it may be refined by upgrading the functional algorithms employed. As an example, the source generation function is, for initial purposes, considered uniform; this is rather primitive, and will be changed. Further refinements, such as modifications to the propagation loss model for paths which traverse the plume could also be incorporated.

As presently written, no disk files are utilized; all arrays reside in RAM. Although this limits the number of rockets, number (and hence size) of the incremental volume elements, number of observers and observation times, and so forth, the computational speed was considered sufficiently important that these limitations were considered acceptable at this time. Some of these limitations may be alleviated through the use of dynamically allocated arrays, which can free memory space. Larger memories are becoming common place in personal computers, and an alternative approach would be to utilize a RAM disk, which would allow operation faster than magnetic disks. Ultimately, it may well be that the computer resource requirements will force implementation on a larger, faster computer.

3.1.1 Program Variables and Arrays.

Variables and arrays have been assigned mnemonically. Therefore the program is easily read, and, more importantly, easily modified. The following is an alphabetical listing and description of the variables and arrays used in the program.

- ACCEL - Acceleration of vehicle, assumed constant, feet/sec^2
- ALTITUDE(INDXTIME) - Altitude of vehicle at time TIME(INDXTIME), feet
- BAND - Number of the frequency band currently of interest
- CONEAN - Semi-vertex angle of truncated cone approximating rocket plume, radians
- CONELEN - Length of cone which contributes to acoustic noise, feet
- CS(INDXDTHT) - Cosine of INDXDTHT * DTHTC
- CURDRC - Current size of DRC at this zc, feet
- CURRC - Current size of cone radius at this zc, feet
- CURROC - Number of rocket currently under consideration
- CURZC - Axial position with respect to the cone apex currently considered, INDXDZC * DZC
- DEG2RAD - Conversion constant, degrees to radians

- DV2OBSR2(CUROBS, INDXTIME, CURROC, INDXDZC, INDXDRC, INDXDTHT)
  - Square of the range between elemental source dV and observer for the current observer CUROBS at time TIME(INDXTIME) for the current rocket under consideration CURROC for a particular dV at INDXDZC * DZC, INDXDRC * CURDRC, and angle INDXDTHT * DTHTC, feet squared

- DVSIZE(INDXDZC,INDXDRC) - Size of incremental volume for a particular distance from apex and angle, INDXDZC * DZC and INDXDRC * CURDRC, cubic feet

- DVX, DVY, DVZ - Scalar position coordinates equivalent to DVXPOS(INDXTIME, CURROC, INDXDZC, INDXDRC, INDXDTHT)
  DVYPOS(INDXTIME, CURROC, INDXDZC, INDXDRC, INDXDTHT)
  DVZPOS(INDXTIME, CURROC, INDXDZC, INDXDRC, INDXDTHT)

- DVXPOS(INDXTIME, CURROC, INDXDZC, INDXDRC, INDXDTHT)
  DVYPOS(INDXTIME, CURROC, INDXDZC, INDXDRC, INDXDTHT)
  DVZPOS(INDXTIME, CURROC, INDXDZC, INDXDRC, INDXDTHT)
  - X, Y, and Z coordinates of a particular elemental dV for a particular time, rocket, and cone related coordinates, feet

- DZC - Thickness of incremental axial "slice", feet

- EXDIA(CURROC) - Exit diameter of current rocket under consideration

- EXVEL(CURROC) - Exit velocity of exhaust of current rocket under consideration

- HELP$ - This parameter sets the help level. If set to Y or y program provides more verbose discussion of inputs, if N or n, omits discussion.

- INDXDRC - Index used to select a particular value of radius, INDXDRC * CURDRC

- INDXDTHT - Index used to select a particular value of angle theta, INDXDTHT * DTHTC

- INDXDZC - Index used to select particular value of zc coordinate, INDXDZC * DZC

- INDXTIME - Index used to select the time of observation, TIME(INDEXTIME)
- MAXRC - Maximum cone radius at a particular zc, CURZC * TGCONEAN, feet
- MININDDZ - Minimum index for zc such that zc is aft of the tail of the rocket
- NOFBANDS - Number of frequency bands of interest
- NOFDRC - Number of radial slices to use
- NOFDZC - Number of axial slices to use
- NOFOBS - Number of observers or locations to evaluate
- NOFOBSTM - Number of observation times to evaluate
- NOFROCS - Number of rockets in vehicle
- NOFTHTC - Number of angular slices to use
- OBSXPOS(CUROBS)
  OBSYPOS(CUROBS)
  OBSZPOS(CUROBS)
  \- X, Y, and Z coordinates of current observer, feet.
- PI - Constant, 3.14159
- RELDVXPOS(INDXDZC, INDXDRC, INDXDTHT)
  RELDVYPOS(INDXDZC, INDXDRC, INDXDTHT)
  RELDVZPOS(INDXDZC, INDXDRC, INDXDTHT)
  \- Coordinates of dV element relative to generic cone, x, y, and z
- ROCXPOS(CURROC)
  ROCYPOS(CURROC)
  ROCZPOS(CURROC)
  \- Initial position of current rocket relative to earth coordinates, X, Y, Z
- SN(INDXDTHT) - Sine of INDXDTHT * DTHTC

- SPL(CUROBS, BAND, INDXTIME)
  - Sound pressure level for current observer, band number, and time

- SPLDB(CUROBS, BAND, INDXTIME) - Sound pressure level in dB

- SRCSTRTH(BAND, CURROC, INDXDZC, INDXDRC, INDXDTHT)
  - Source strength for current band number, rocket, and cone relative coordinates of interest

- SUMSPL - Used to accumulate the total contributions of all the elemental volumes

- TGCONEAN - Tangent of the semi-apex cone angle

- THRST(CURROC) - Thrust of current rocket, pounds

- TOTTHRST - Total thrust of vehicle, pounds

- VEHCLWT - Vehicle weight, pounds

- XDRIFT(INDXTIME)
  YDRIFT(INDXTIME)
  - Drift of rocket in X - Y (earth relative) plane at time TIME(IINDXTIME), feet

3.1.2 PROGRAM FUNCTIONAL MODULES.

The elements of the program are divided into seven modules so that the program may be readily modified or extended. The following sections are a discussion of their names and functions.

3.1.2.1 VEHICLE GEOMETRY AND PARAMETER DEFINITION SECTION.

All relevant numbers describing the system geometry, weight, and performance are defined in this section. Whenever the same system is to be used for many runs, it will be desirable to fix the data rather than reenter it, as was done for the program verification tests. At present, the exhaust velocity of each rocket is included in the data, but is not used in the program. The length of the laminar core in the plume is a function of the exhaust exit velocity, and it is anticipated that future programs

C-3

will utilize this data.

The position of the rocket is described with respect to the earth. As viewed from above, the Y axis is defined to pass through the centers of the two SRBs and the external tank, and the X axis goes through the center of the external tank. Altitude, Z, is measured with respect to mean sea level.

For test purposes, the parameters for the program were selected to be approximately those of the shuttle system.

3.1.2.2 OBSERVER DEFINITION SECTION.

The sound pressure level (SPL) is evaluated at a number of points of observation, called observer locations. This section is used to define the earth relative positions (OBSXPOS, OBSYPOS, and OBSZPOS) for each point of interest.

3.1.2.3 GENERIC CONE GENERATION SECTION.

In order to save computation time, a "generic" cone is defined and divided into a large number of elemental volumes, dV. The semi-apex angle of the cone employed is 7.5 degrees, and the length of the cone which contributes to the generation of noise is considered to be 1000 feet. For test purposes, the cone is divided into 200 foot axial slices (DZC), 8 angular sectors (NOFTHTC), and 3 radial slices (NOFDRC). Since the maximum radius (MAXRC) changes with the current distance from the apex (CURZC), the value of the incremental change in radius changes (CURDRC). (It would also be possible to keep the incremental radial increment constant and change the summation appropriately.)

The size of the increments cited above are much too large to be considered incremental volumes, and will be reduced when the program is actually utilized. Smaller increments would have had severe impact on computing time and memory requirements; since inaccuracy in the integration is not a problem for program testing, the reduced time and memory were welcome.

The location and size of each dV are then calculated in a rectangular cone relative coordinate system, x,y,z. The cone relative system is defined to be parallel to the earth relative system.

3.1.2.4 SOURCE GENERATION SECTION.

The source generation section ascribes an acoustic intensity to each elemental volume for each rocket and each frequency band. The generic cone is effectively truncated at the top to the exit diameter of the rocket by summing only over an appropriate subset of the DZC index, from MININDDZ to NOFDZC.

3.1.2.5 RANGE AND RANGE SQUARED EVALUATION SECTION.

The range evaluation section calculates the position of each rocket relative to the earth at each time of interest. The initial acceleration is estimated using the vehicle weight (VEHCLHWT) and the total thrust (TOTTHRST). It is assumed that the change in the mass of the vehicle is negligible during the time of interest so that the altitude of a rocket is just its initial value plus one half the acceleration times the square of the elapsed time.

The possibility of horizontal drift is allowed through the variables XDRIFT AND YDRIFT. Such drifts could have drastic effects on the sound pressure level since the distance scales of interest are relatively small.

This section then calculates the square of the range (DV2OBSR2) from each elemental volume of each rocket to each observer for each time.

3.1.2.6 SOUND PRESSURE LEVEL EVALUATION SECTION.

This section sums the contributions from each dV to each observer for each time and band assuming inverse square law propagation. Intensities are added directly, which is tantamount to assuming that there is negligible coherence between sources.

3.1.2.7 TEST AND VERIFICATION SECTION.

This section contains subroutines which were generated to check the performance of components of the program and were called at the end of each section above, but have been deleted to save space in the report.

3.2 PROGRAM LISTING

The following is a listing of the computer program which performs the numerical integration to evaluate the total SPL at a point. Some cosmetic changes have been made in this listing.

```
   '  NEAR ACOUSTIC FIELD ROCKET NOISE PROGRAM (NAFRNOP?)
   '  RNFRPT17.BAS, 7/17/89; FOR SPL PROGRAM DOCUMENTATION
   '
   '  VEHICLE GEOMETRY AND PARAMETER DEFINITION SECTION
   '
   '  NORMALLY INPUT ALL DATA; FOR TEST PURPOSES DEACTIVATE
HELP,
   '  FIX SYSTEM; DE - REM FOR OPERATIONAL PROGRAM.  STATE-
MENTS
   '  REQUIRING ATTENTION ARE FULLY LEFT JUSTIFIED.
   '
REM:INPUT "DO YOU WANT HELP? (Y/N)", HELP$: PRINT " "
HELP$ = "N":           REM: FOR TEST PURPOSES.
REM:      PRINT "DEFINE VEHICLE GEOMETRY AND PARAMETERS:"
          PRINT " "
      IF HELP$ = "Y" OR HELP$ = "y" THEN PRINT "PLEASE INPUT
      VEHICLE WEIGHT IN POUNDS", VEHCLWT
      PRINT " "
VEHCLWT = 5000000: REM:  INPUT"VEHCLWT = ? ",VEHCLWT
      IF HELP$ = "Y" OR HELP$ = "y" THEN PRINT "HOW MANY
ROCKET ENGINES WILL BE ACTIVE?"
      PRINT " "
NOFROCS = 3:   REM:  INPUT"NOFROCS = ?",NOFROCS
      DIM ROCXPOS(NOFROCS), ROCYPOS(NOFROCS), ROCZPOS(NO-
FROCS),
          EXDIA(NOFROCS), EXVEL(NOFROCS), THRST(NOFROCS)
      IF HELP$ = "Y" OR HELP$ = "y" THEN PRINT "COORDINATE
      DISCUSSION":
      PRINT "
      THE xyz COORDINATE SYSTEM IS ASSOCIATED WITH THE
ROCKET.  IT REMAINS PARALLEL TO THE XYZ EARTH REFERENCE
SYSTEM."
      PRINT " "
      "PLEASE DEFINE THE ROCKET PARAMETERS, x, y, AND z
POSITIONS, EXIT DIAMETER, EXIT VELOCITY, AND THRUST FOR
EACH ENGINE."
'
'     (YOU MAY WISH TO FIX THE PARAMETERS WHICH ARE CONSTANT
```

```
'       FOR  A  SERIES OF RUNS TO  AVOID  EXCESSIVE  DATA
'       REENTRY.)
'      NOTE:  DATA APPROXIMATELY CONSISTENT WITH  SHUTTLE
'      SYSTEM
'
REM:       FOR CURROC = 1 TO NOFROCS
REM:       PRINT "FOR ROCKET ENGINE NUMBER "; CURROC
REM:INPUT"X =, Y =, Z =, EXDIA =, EXVEL =, THRST =";
ROCXPOS(CURROC),ROCYPOS(CURROC),ROCZPOS(CURROC),
EXDIA(CURROC),EXVEL(CURROC),THRST(CURROC)
REM:
REM: TEST DATA
ROCXPOS(1) = 28.08333: ROCYPOS(1) = 0: ROCZPOS(1) = 40:
EXDIA(1) = 13.567: EXVEL(1) = 10663: THRST(1) = 1198483:
ROCXPOS(2) = 0: ROCYPOS(2) = 20.8333:  ROCZPOS(2) = 40:
EXDIA(2) = 12.375: EXVEL(2) = 8202: THRST(2) = 264833
ROCXPOS(3) = 0: ROCYPOS(3) = -20.8333: ROCZPOS(3) = 40:
EXDIA(3) = 12.375: EXVEL(3) = 8202:  THRST(3) = 2648334:
REM:       NEXT CURROC
'
'
'     OBSERVER DEFINITION SECTION;
'     DEFINE NUMBER AND LOCATIONS OF OBSERVERS.
'
     IF HELP$ = "Y" OR HELP$ = "y" THEN PRINT "PLEASE
DEFINE NUMBER AND POSITIONS OF OBSERVERS RELATIVE TO XYZ
(EARTH) COORDINATES.":
' PRINT " "
REM: INPUT"HOW MANY OBSERVATION POINTS DO YOU WANT?",NOFOBS
REM: FOR TEST PURPOSES ONLY ONE OBSERVER
NOFOBS = 1:  REM: FOR TEST PURPOSES
     DIM OBSXPOS(NOFOBS), OBSYPOS(NOFOBS), OBSZPOS(NOFOBS)
REM:FOR CUROBS = 1 TO NOFOBS
REM:PRINT"X, Y, AND Z FOR OBSERVER NUMBER ";CUROBS;":"
REM:INPUT OBSXPOS(CUROBS),OBSYPOS(CUROBS),OBSZPOS(CUROBS)
REM:NEXT CUROBS     : REM: TEST DATA
     OBSXPOS(1) = 0: OBSYPOS(1) = 30000: OBSZPOS(1) = 0
'
'     GENERIC CONE GENERATION SECTION
'
     PI = 3.14159265#
     DEG2RAD = PI / 180
     CONEAN = 7.5 * DEG2RAD:  ' CENTRAL CONE ANGLE ASSUMED
     TGCONEAN = TAN(CONEAN):  ' TO BE 7.5 DEGREES.
     CONELEN  = 1000:
'  GENERATING  PART,  ASSUMED  =  1000'
```

```
DZC = 200:  ' 200 FOOT Z AXIS "SLICES" USED FOR TESTING.
     NOFDZC  =  INT(CONELEN / DZC + .5):
' CALC NO. OF Z SLICES.
NOFDRC = 3:  ' ASSUMES 3 SLICES ALONG RADIUS FOR TESTING.
NOFTHTC = 8:  ' ASSUMES 8 SEGMENTS IN CIRCLE FOR TESTING.
' CALCULATE SIZE OF DTHETA.
     DTHTC = 2 * PI / NOFTHTC
     DIM SN(NOFTHTC), CS(NOFTHTC),
     RELDVXC(NOFDZC, NOFDRC, NOFTHTC),
     RELDVYC(NOFDZC, NOFDRC, NOFTHTC),
     RELDVZC(NOFDZC), DVSIZE(NOFDZC, NOFDRC)
        FOR INDXDTHT = 1 TO NOFTHTC
        SN(INDXDTHT) = SIN(INDXDTHT * DTHTC)
        CS(INDXDTHT) = COS(INDXDTHT * DTHTC)
        NEXT INDXDTHT
        FOR INDXDZC = 1 TO NOFDZC
        CURZC = (INDXDZC - .5) * DZC
        MAXRC = CURZC * TGCONEAN
        CURDRC = MAXRC / NOFDRC
          FOR INDXDRC = 1 TO NOFDRC
          CURRC = (INDXDRC - .5) * CURDRC
            FOR INDXDTHT = 1 TO NOFTHTC
       RELDVXC(INDXDZC, INDXDRC, INDXDTHT) =
       CURRC * CS(INDXDTHT)
       RELDVYC(INDXDZC, INDXDRC, INDXDTHT) =
       CURRC * SN(INDXDTHT)
       RELDVZC(INDXDZC) = CURZC
       DVSIZE(INDXDZC, INDXDRC) = CURRC * CURDRC * DTHTC *
                                  DZC
            NEXT INDXDTHT
          NEXT INDXDRC
        NEXT INDXDZC
        PRINT "CONE RELATIVE DV POSITIONS CALCULATED."
 '
 REM: PRINT "ENTERING RELDVPOS/DVSIZE CHECK": GOSUB 2490:
 REM: STOP: REM: TEST.  OK 7/11
 '
 '    SOURCE GENERATION SECTION
 '
 NOFBANDS = 1:     REM: FOR TEST PURPOSES
      DIM SRCSTRTH(NOFBANDS, NOFROCS, NOFDZC, NOFDRC,
 NOFTHTC),
      MININDDZ(NOFROCS)
       FOR CURROC = 1 TO NOFROCS
       MININDDZ(CURROC) = INT(EXDIA(CURROC)
       / (2 * TGCONEAN) / DZC + .5)
```

```
      IF MININDDZ(CURROC) < 1 THEN MININDDZ(CURROC) = 1
      NEXT CURROC
      FOR BAND = 1 TO NOFBANDS
        FOR CURROC = 1 TO NOFROCS
          FOR INDXDZC = MININDDZ(CURROC) TO NOFDZC
            FOR INDXDRC = 1 TO NOFDRC
              FOR INDXDTHT = 1 TO NOFTHTC
              SRCSTRTH(BAND, CURROC, INDXDZC, INDXDRC,
INDXDTHT)
              = 2E+14 * DVSIZE(INDXDZC, INDXDRC)
              NEXT INDXDTHT
            NEXT INDXDRC
          NEXT INDXDZC
        NEXT CURROC
      NEXT BAND
'
REM:  FOR TEST, SET SELECTED dV SOURCES TO NON - ZERO
'
      PRINT "SOURCE STRENGTH CALCULATED."
'
REM:  GOSUB 2730:   STOP: REM:(TEST)  OK 7/11/89
'
'
'     RANGE AND RANGE SQUARED EVALUATION SECTION.
'     THIS SUBSECTION EVALUATES THE LOCATION OF EACH DV
'     ELEMENT FOR EACH ROCKET AT EACH TIME.
'
      IF HELP$ = "Y" OR HELP$ = "y" THEN PRINT
      "THE POSITION OF THE VEHICLE IS ESTIMATED AT A NUMBER
OF TIMES AFTER LIFT OFF.  PLEASE INDICATE HOW MANY TIME
VALUES YOU INTEND, THEN THE ACTUAL OBSERVATION TIMES"
PRINT " "
NOFOBSTM = 1:  REM:  INPUT"NOFOBSTM = ",NOFOBSTM
      DIM TIME(NOFOBSTM), ALTITUDE(NOFOBSTM), XDRIFT(NO-
FOBSTM),
      YDRIFT(NOFOBSTM)
      DIM DVXPOS(NOFOBSTM, NOFROCS, NOFDZC, NOFDRC,
NOFTHTC),
      DVYPOS(NOFOBSTM, NOFROCS, NOFDZC, NOFDRC, NOFTHTC),
      DVZPOS(NOFOBSTM, NOFROCS, NOFDZC)
REM:FOR INDXTIME = 1 TO NOFOBSTM
REM:INPUT"EVALUATE AT TIME AFTER LIFT OFF (SECS) = ",
          TIME(INDXTIME)
REM:NEXT INDXTIME
TIME(1) = 0
REM: FOR TEST PURPOSES, LET TIME = 0 IDENTICALLY.
```

```
      TOTTHRST = 0  REM: SET TOTAL THRUST TO 0
        FOR CURROC = 1 TO NOFROCS
        TOTTHRST = TOTTHRST + THRST(CURROC)
        NEXT CURROC
      PRINT "TOTAL THRUST = "; TOTTHRST
      ACCEL = (TOTTHRST - VECHLWT) * 32.16 / VEHCLWT
        FOR INDXTIME = 1 TO NOFOBSTM
        ALTITUDE(INDXTIME) = ACCEL / 2 * TIME(INDXTIME) ^ 2
        XDRIFT(INDXTIME)  =  0:
REM: CAN INTRODUCE DRIFT HERE
        YDRIFT(INDXTIME) = 0:  REM: IF DESIRED.
          FOR CURROC = 1 TO NOFROCS
            FOR INDXDZC = MININDDZ(CURROC) TO NOFDZC
              FOR INDXDRC = 1 TO NOFDRC
                FOR INDXDTHT = 1 TO NOFTHTC
DVXPOS(INDXTIME, CURROC, INDXDZC, INDXDRC, INDXDTHT) =
                ROCXPOS(CURROC) + XDRIFT(INDXTIME)
                + RELDVXC(INDXDZC, INDXDRC, INDXDTHT)
DVYPOS(INDXTIME, CURROC, INDXDZC, INDXDRC, INDXDTHT) =
                ROCYPOS(CURROC) + YDRIFT(INDXTIME)
                + RELDVYC(INDXDZC, INDXDRC, INDXDTHT)
DVZPOS(INDXTIME, CURROC, INDXDZC) =
               ROCZPOS(CURROC) + ALTITUDE(INDXTIME)
                - (INDXDZC + .5 - MININDDZ(CURROC)) * DZC
                NEXT INDXDTHT
              NEXT INDXDRC
            NEXT INDXDZC
          NEXT CURROC
        NEXT INDXTIME

        PRINT "DV POSITIONS CALCULATED FOR ALL ROCKETS."

REM:
REM:    PRINT "ENTERING DV POSITION CHECK.": GOSUB 3050:
REM:     (TEST) :  STOP : REM:    OK, 7/12/89
'
'
'    RANGE SQUARED SUBSECTION
'
DIM DV2OBSR2(NOFOBS, NOFOBSTM, NOFROCS, NOFDZC, NOFDRC,
NOFTHTC)
  FOR INDXTIME = 1 TO NOFOBSTM
    FOR CURROC = 1 TO NOFROCS
      FOR INDXDZC = MININDDZ(CURROC) TO NOFDZC
        FOR INDXDRC = 1 TO NOFDRC
          FOR INDXDTHT = 1 TO NOFTHTC
```

```
DVX = DVXPOS(INDXTIME, CURROC, INDXDZC, INDXDRC, INDXDTHT)
DVY = DVYPOS(INDXTIME, CURROC, INDXDZC, INDXDRC, INDXDTHT)
DVZ = DVZPOS(INDXTIME, CURROC, INDXDZC)
            FOR CUROBS = 1 TO NOFOBS
DV2OBSR2(CUROBS, INDXTIME, CURROC, INDXDZC, INDXDRC,
INDXDTHT)             = (DVX - OBSXPOS(CUROBS)) ^ 2 + (DVY -
OBSYPOS(CUROBS)) ^ 2
+ (DVZ - OBSZPOS(CUROBS)) ^ 2
            NEXT CUROBS
          NEXT INDXDTHT
        NEXT INDXDRC
      NEXT INDXDZC
    NEXT CURROC
  NEXT INDXTIME

PRINT "RANGE SQUARED VALUES CALCULATED FOR ALL dV AND
ROCKETS."
'
REM:
REM: PRINT "ENTERING RANGE SQUARED CHECK": GOSUB 4270:
STOP:
REM: (TEST)
REM: CHECKS ON TWO POINTS SELECTED AT RANDOM, 7/12/89
'
'
'    SOUND PRESSURE LEVEL EVALUATION SECTION
'
     DIM SPL(NOFOBS, NOFBANDS, NOFOBSTM),
     SPLDB(NOFOBS, NOFBANDS, NOFOBSTM)
       FOR CUROBS = 1 TO NOFOBS
         FOR INDXTIME = 1 TO NOFOBSTM
           FOR BAND = 1 TO NOFBANDS
           SUMSPL = 0
             FOR CURROC = 1 TO NOFROCS
               FOR INDXDZC = MININDDZ(CURROC) TO NOFDZC
                 FOR INDXDRC = 1 TO NOFDRC
                   FOR INDXDTHT = 1 TO NOFTHTC
                   SUMSPL = SUMSPL
+ SRCSTRTH(BAND, CURROC, INDXDZC, INDXDRC, INDXDTHT)
/ DV2OBSR2(CUROBS, INDXTIME, CURROC, INDXDZC, INDXDRC,
INDXDTHT)
REM:  PRINT "CURROC, INDXDZC, INDXDRC, INDXDTHT, AND SUMSPL
= ";
REM:  CURROC; INDXDZC; INDXDRC; INDXDTHT; SUMSPL
                   NEXT INDXDTHT
                 NEXT INDXDRC
```

```
                 NEXT INDXDZC
               NEXT CURROC
SPL(CUROBS, BAND, INDXTIME) = SUMSPL
SPLDB(CUROBS, BAND, INDXTIME) =
10 * LOG(SPL(CUROBS, BAND, INDXTIME)) / LOG(10)
             PRINT "FOR OBSERVER ";
CUROBS; ", BAND "; BAND; " AND TIME "; TIME(INDXTIME); "THE
SPL IS "; SPLDB(CUROBS, BAND, INDXTIME); " DB."
             NEXT BAND
           NEXT INDXTIME
         NEXT CUROBS
         PRINT " SPL AND SPLdB CALCULATED"
REM:
REM: PRINT "ENTERING SPL CHECK.": GOSUB 5490
REM:
     STOP
REM:
REM: END OF PROGRAM
```

SECTION IV

RESULTS AND DISCUSSION

4.1 PROGRAM TEST RESULTS AND DISCUSSION

The sections of the program were checked individually by hand calculations at a few randomly selected points. In this section the program was employed under varying circumstances to verify overall performance. Shuttle vehicle data was used in the program. Source strength was set to a constant density and the level set arbitrarily for test purposes. Neither the constant density nor the arbitrary level have been correlated to actual sound levels; these runs are for test purposes only.

As a first case, data was generated for a polar plot in the X - Y plane. This data revealed little variation with angle; the pattern is omnidirectional within a fraction of a dB. There was a slight rise on the positive X axis which is probably attributable to the shuttle main engines.

The solid line in figure 4-1 shows the variation of the SPL generated by the rocket, as modeled by the cited distributed source, along the X (or Y) axis as a function of the logarithm of the range. The ranges considered were from 100 to 30000 feet. For comparison purposes, the + symbols show the performance of an equivalent point source obeying an ideal inverse square law. For large ranges, approximately 2000 feet or more, the distributed model and equivalent point model are indistinguishable. For small ranges, less than 2000 feet, the distributed model indicates lower sound levels than the point source. This behavior is logical.

Next, the SPL was evaluated along the Z axis and compared to the SPL evaluated along the X or Y axis. Figure 4-2 shows the Z axis variations, solid line, and the X axis variations, + symbol, as a function of the logarithm of range. For a given range along the Z axis the sound level is less than or equal to the level at the same range on the X axis. Moving along the Z axis takes the observer directly away from the source while moving along the X axis increases the range to the distributed source more slowly, once again a logical behavior. For ranges approaching zero, the observation points approach the origin and the

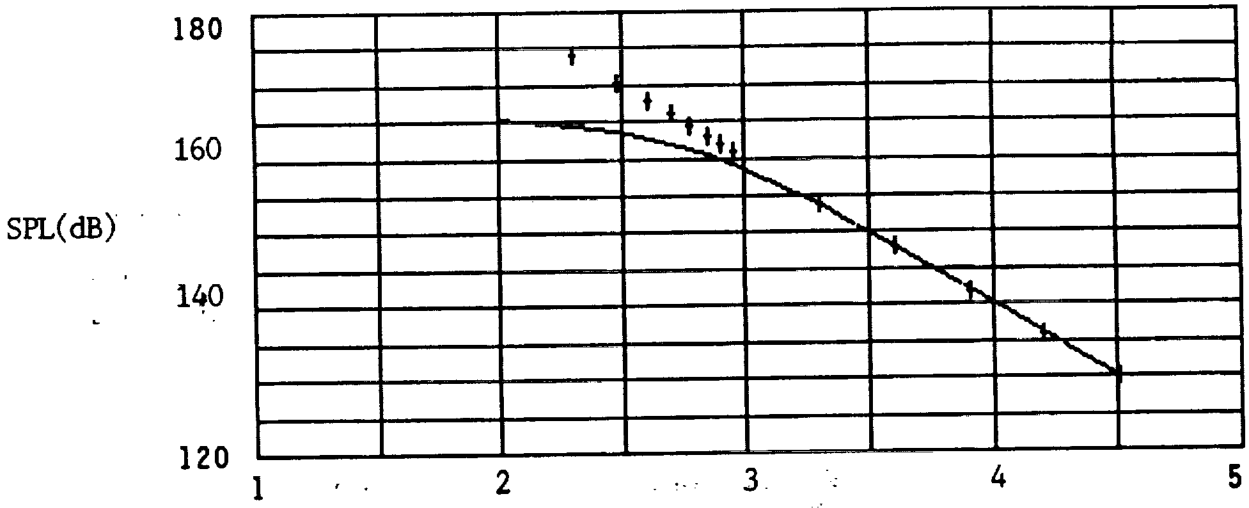


Figure 4-1. SPL IN dB AS A FUNCTION OF THE LOGARITHM OF X OR Y RANGE (point source and ideal inverse square law, + symbol)

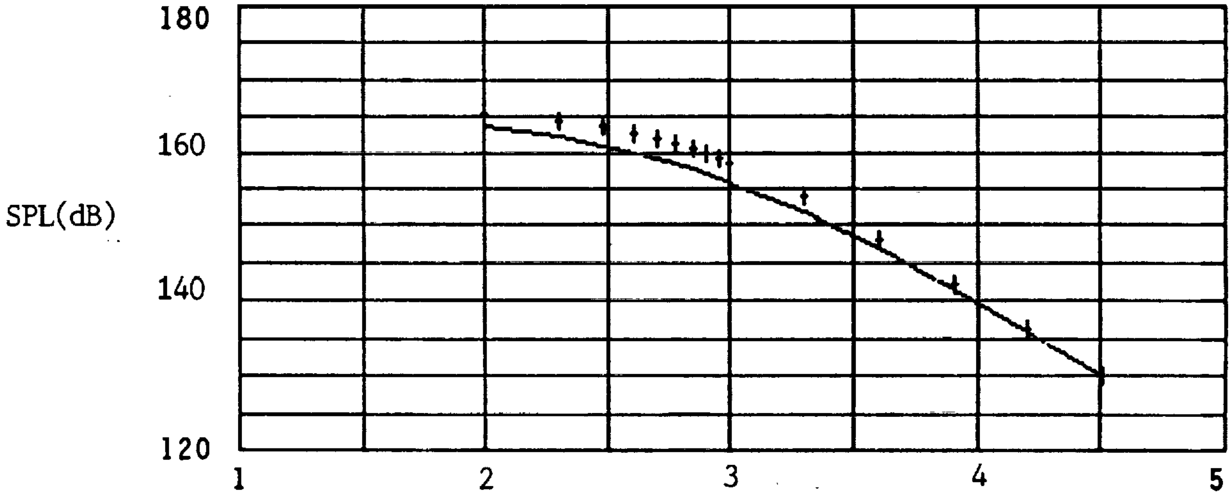


Figure 4-2. Z AXIS VARIATION IN SPL COMPARED TO X OR Y AXIS VARIATION (Z range, solid line, X or Y range, + symbol)

SPL must be the same; at large ranges, the distributed source behaves as a point source and both curves again coincide. The maximum difference appears to be at a range which is approximately the same as the length of the cone, about 1000 feet.

The SPL was then evaluated for an observer at a horizontal range of 100 feet, both as a function of vehicle altitude, figure 4-3, and time after lift off, figure 4-4. As the altitude of the vehicle increases, the distributed source is actually coming closer to the observer, so the SPL increases until its effective center passes the observer at a vehicle altitude of about 800 feet. The sound level then decreases as altitude continues to increase. In figure 4-4 the sound pressure level variation with time is graphed. The SPL increases very slowly at first since the vehicle position initially changes quite slowly, then rises to a peak at about 14 seconds, then falls as the vehicle recedes.

The SPL pattern was then explored around circular contours in the X - Z plane at ranges of 100, 500, 1000, and 3000 feet. Figures 4-5, 4-6, 4-7, and 4-8 are polar plots of the SPL calculated at these ranges respectively. Circles corresponding to 150 dB, 160 dB and 170 dB have been graphed on these plots for reference, while the SPL has been graphed with + symbols. For the 100 foot contour the variation in range to the source is small compared to the dimensions of the source and the variation of the SPL is only approximately 6 dB. Straight down, toward the bottom of the page, the observation point is near the top of the generating cones and the SPL exhibits a maximum. The maximum is skewed a bit toward the right (along the X axis), again presumably associated with the shuttle engines. The pattern exhibits much greater directionality for the 500 foot and 1000 foot contours, please see figures 4-6 and 4-7. These contours cut through the approximate middle and far end of the source cones, so the sound levels are most significantly different for these cases. The graph for the 3000 foot contour, figure 4-8, shows reduced directionality and is beginning to become omnidirectional as anticipated; the source is again approaching the behavior of a point source, as anticipated.

Finally, to compare these patterns more directly, the contours in the X - Z plane for ranges of 100, 500, 1000,

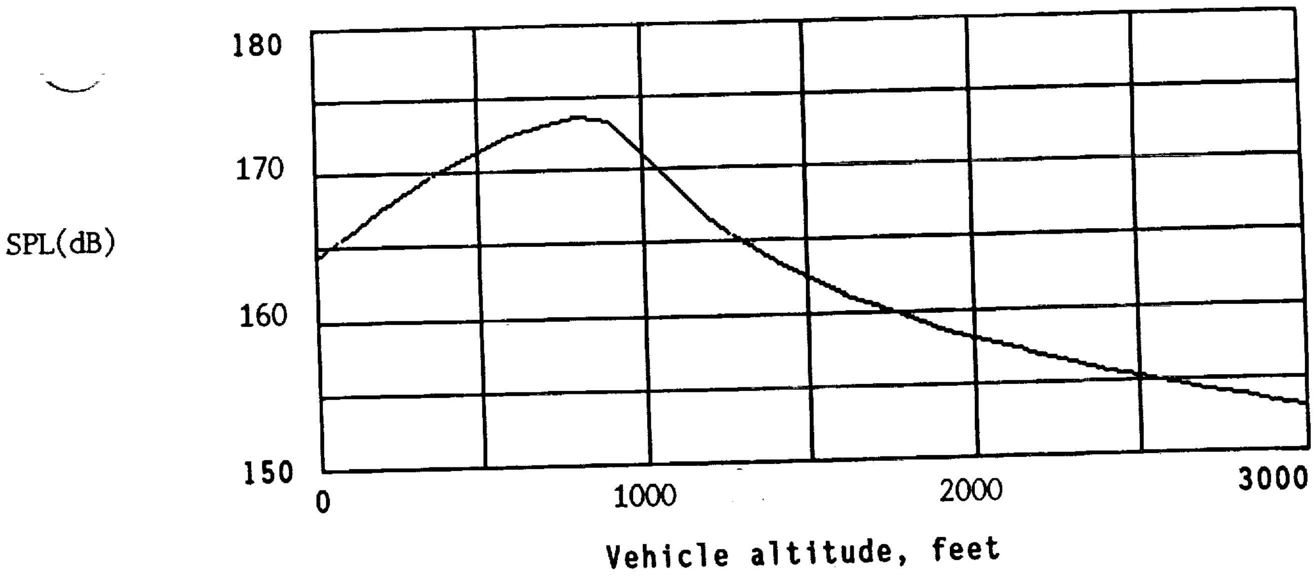


Figure 4-3. SPL AS A FUNCTION OF VEHICLE ALTITUDE at a horizontal range of 100 feet

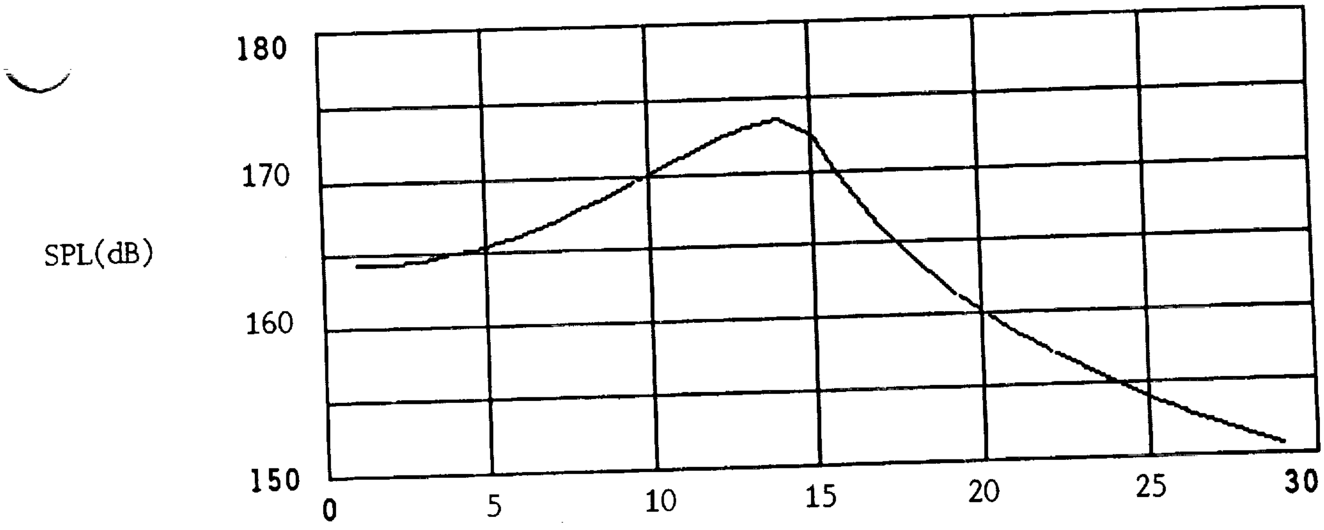


Figure 4-4. SPL AS A FUNCTION OF TIME AFTER LIFT OFF for a horizontal range of 100 feet

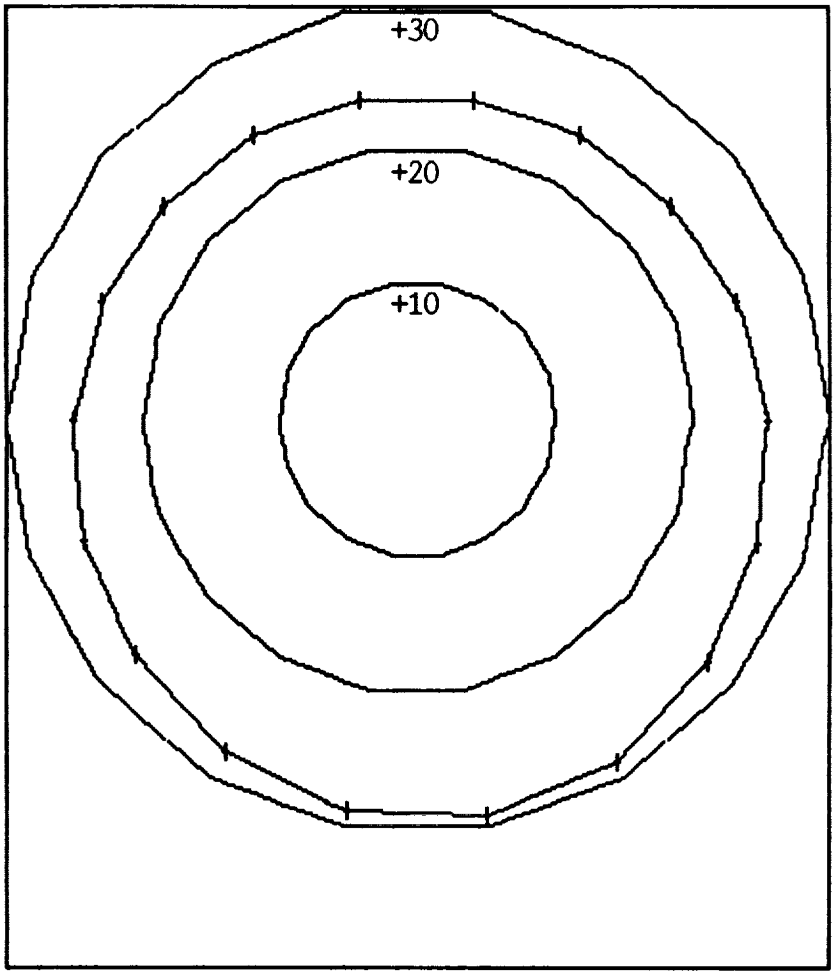


Figure 4-5. POLAR PLOT OF SPL re 140 dB FOR A RANGE OF 100 FEET (in X - Z plane)

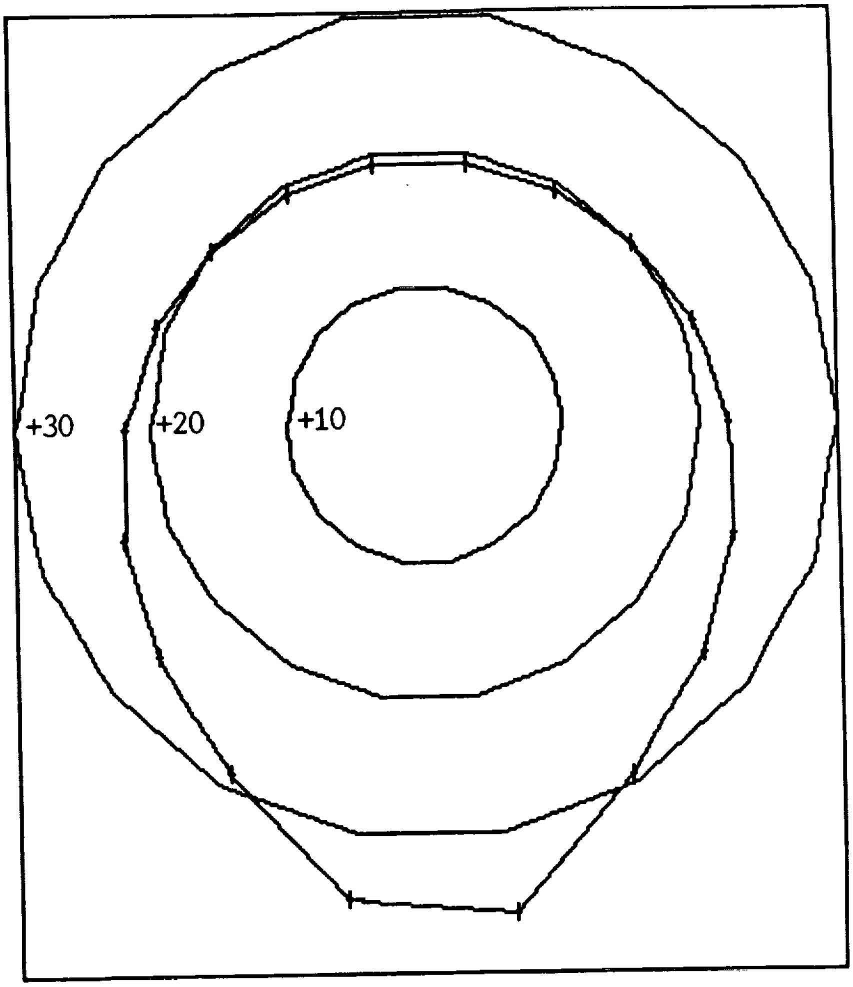


Figure 4-6. POLAR PLOT OF SPL re 140 dB FOR A RANGE OF 500 FEET (in X - Z plane)

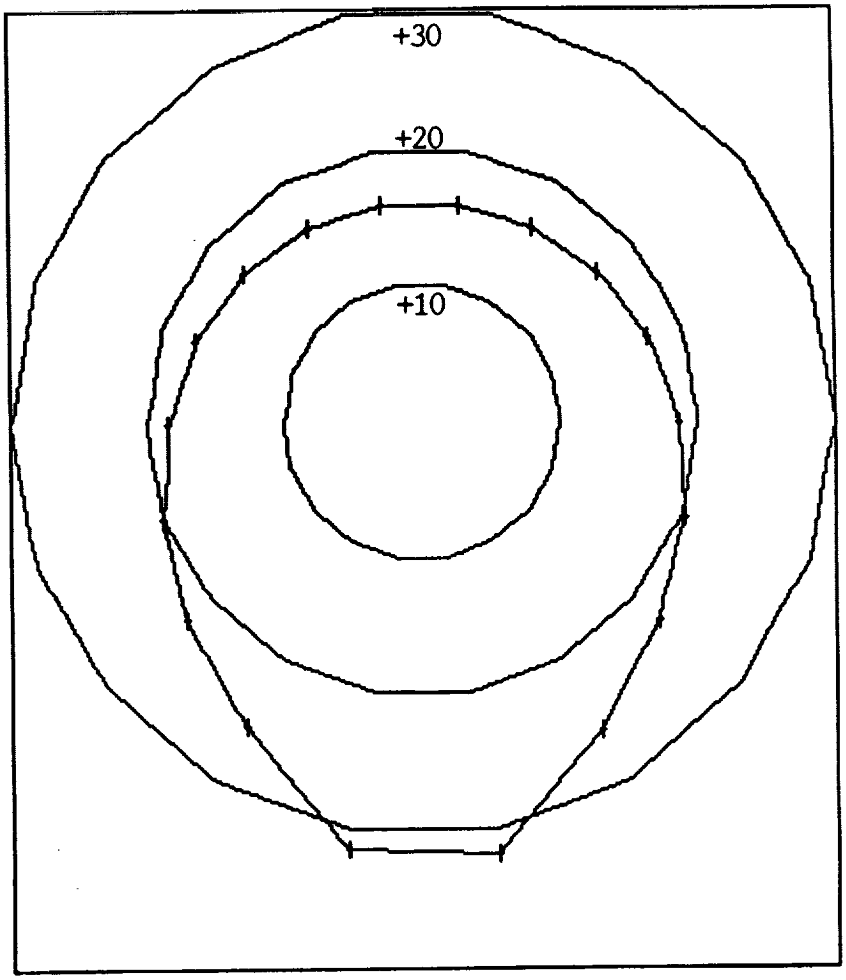


Figure 4-7. POLAR PLOT OF SPL re 140 dB FOR A RANGE OF 1000 FEET (in X - Z plane)

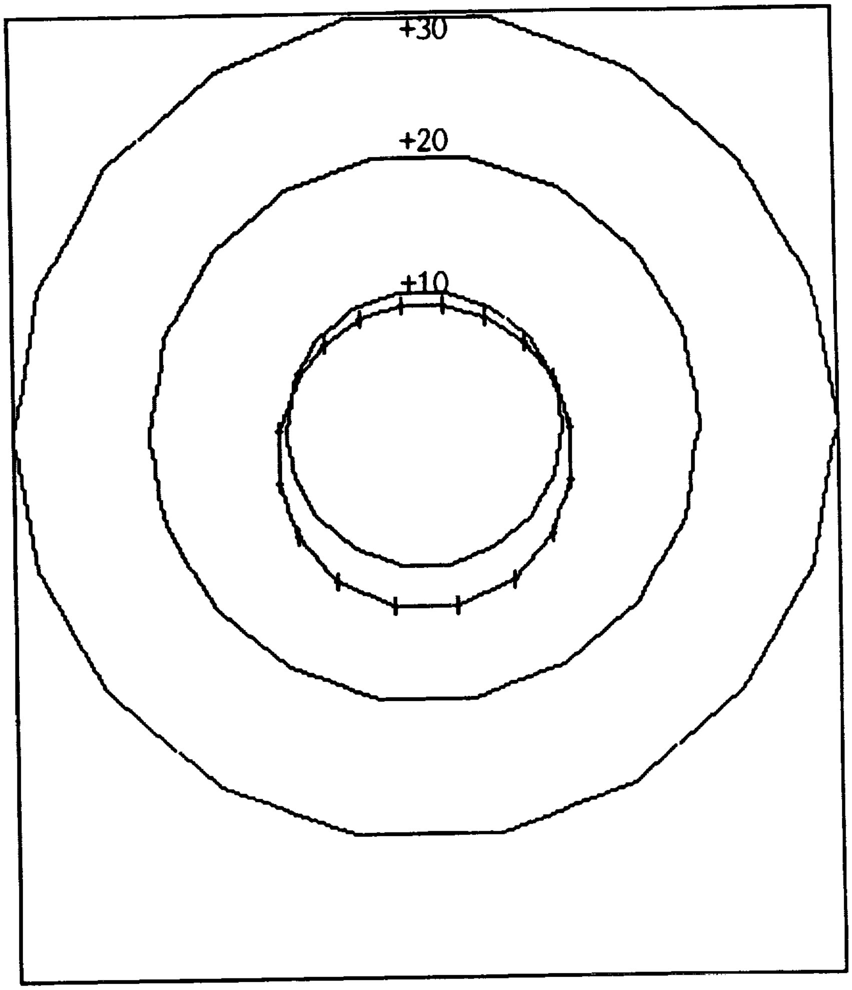


Figure 4-8. POLAR PLOT OF SPL re 140 dB FOR A RANGE OF 3000 FEET (in X - Z plane)

and 3000 feet were plotted in figure 4-9 using +, x, rectangle, and diamond shaped symbols. The reference circles at 150, 160, and 170 dB have been deleted to avoid clutter. Except for straight down, the region of the rocket plume, the levels decrease monotonically as the range increases. The directionality and slight asymmetry are consistent and apparent. (Note: the software used to graph these curves does not maintain scale between X and Y axes, nor does the screen presentation accurately reflect the appearance of the graph.)

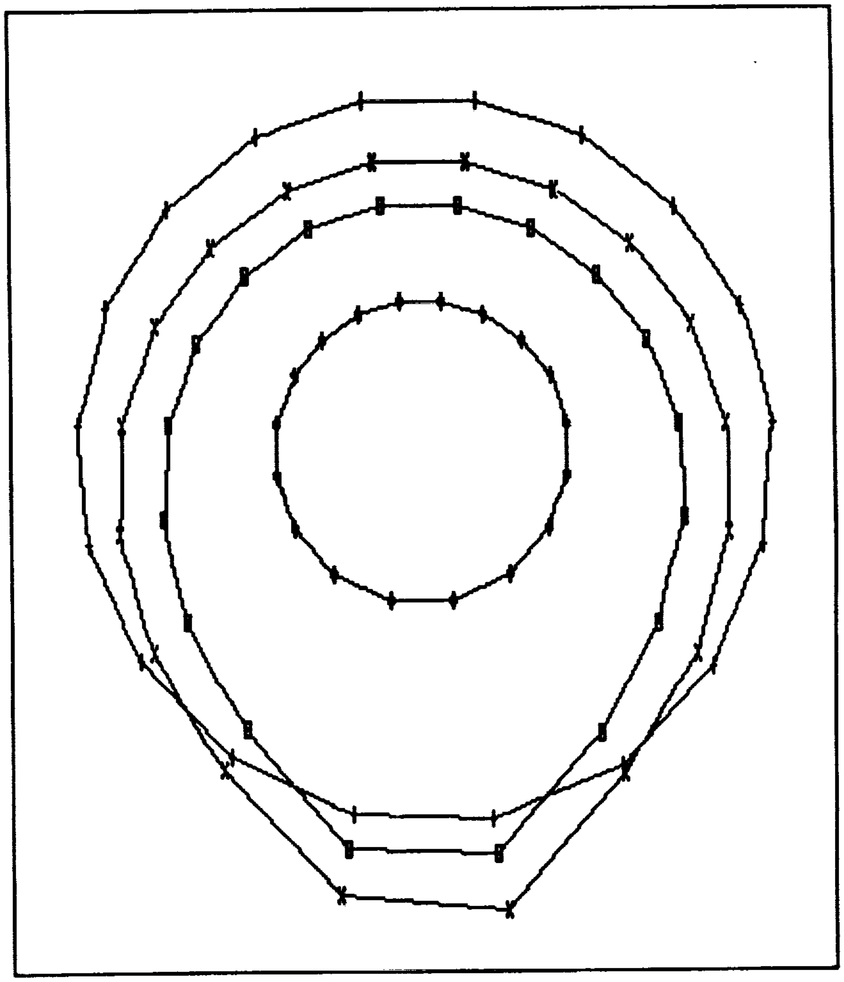


Figure 4-9. COMPARISON OF SPL PATTERNS IN THE X - Z PLANE FOR RANGES OF 100 (+), 500 (X), 1000 (RECTANGLE), AND 3000 (DIAMOND) FEET

SECTION V

CONCLUSIONS

5.1 CONCLUSIONS AND COMMENTS

The program designed and implemented to evaluate the near acoustic field of a rocket appears to function as intended. It employs superposition and assumes that the source is incoherent. For test purposes, a constant source density was used. During actual use experimental data need be employed to calibrate the model.

The use of BASIC and a personal computer were adequate for program definition and development, and were convenient. It may be desirable to utilize a more powerful computer in actual use, particularly if small integration steps are taken, many runs are required and/or the program is made more complex through the use of more elaborate mathematical models.

REFERENCES

1. Rayleigh, Lord. On the Instability of Jets. Proc. London Math. Soc. 10, 4-13; Scientific Papers Vol. 1, 1878, pp. 361-371, Cambridge University Press, Cambridge.

2. Rayleigh, Lord. On the Stability, or Instability of Certain Fluid Motions. Proc. London Math. Soc. 11, 57-70; Scientific Papers Vol. 1, 1880, pp. 474-487, Cambridge University Press, Cambridge.

3. Reynolds, O. An Experimental Investigation of the Circumstances Which Determine Whether the Motion of Water Shall be Direct or Sinuous, and the Law of Resistance in Parallel Channels. Philos. Trans. R. Soc. London 174, 1883, pp. 935-982.

4. Reynolds, O. On the Dynamical Theory of Incompressible Viscous Fluids and the Determination of the Criterion. Philos. Trans. R. Soc. London 186, 1894, pp. 123-161.

5. Lighthill, M.J. On Sound Generated Aerodynamically, 1. General Theory. Proc. R. Soc. London, Ser. A211 (1107), 1952, pp. 564-587.

6. Efwcs Williams, J.E. The Noise from Turbulence Convected at High Speed. Philos. Trans. R. Soc. London, Ser. A255, 1963, pp. 469-503.

7. Moffatt, H.K. Some Developments in the Theory of Turbulence. J. Fluid Mech. 106, 1981, pp. 27-47.

8. Goldstein, M.E. Aeroacoustics of Turbulent Shear Flows. Ann. Rev. Fluid Mech. 16, 1984, pp. 263-285.

9. Mollo-Christensen, E.L. and áNarasimha, R. Sound Emission from Jets at High Subsonic Velocities. J. Fluid Mech. 8, 1960, pp. 49-60.

10. Mollo-Christensen, E.L., Kolpin, M.A. and áMartucelli, J.R. Experiments on Jet Noise, Far Field Spectra and Directivity Patterns. J. Fluid Mech. 18, 1964, pp. 285-301.

11. Liu, J.T.C. Developing Large-Scale Wavelike Eddies and the Near Jet Noise Field. J. Fluid Mech. 62, 1974, pp. 437-464.

12. Michalke, A. and Fuchs, H.V. On Turbulence and Noise of an Axi-Symmetric Shear Flow. J. Fluid Mech. 70, 1975, pp. 179-205.

13. áBrowand, F.K. and Weidman, P.D. Large Scales in the Developing Mixing Layer. J. Fluid Mech. 76, 1976, pp. 127-144.

N90-16695

1989 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

PLANT FEATURES MEASUREMENTS FOR ROBOTICS

PREPARED BY: Dr. Gaines E. Miles

ACADEMIC RANK: Professor

UNIVERSITY AND DEPARTMENT: Purdue University
Department of Agricultural Engineering

NASA/KSC

DIVISION: Biomedical Operations and Research

BRANCH: Life Sciences Research Office

NASA COLLEAGUE: Mr. Ralph P. Prince

DATE: August 11, 1989

CONTRACT NUMBER: University of Central Florida
NASA-NGT-60002 Supplement: 2

Acknowledgements

This project would not have been possible without the assistance of many people and I am deeply grateful to them. Thank you, Ralph Prince and John Sager for providing the opportunity to participate in this most enjoyable experience and for finding all the apparatus necessary to conduct the experiments. For their advice on plant physiology and visual symptoms of deficiencies, I am deeply grateful to Ken Corey and Ray Wheeler. To Wyckliffe Hoffler, who provided the bellows which made the closeup images possible, I say thanks. Without Cheryl Mackowiak's help in setting up the plant growth experiments, no results would have been possible, and thus I am deeply indebted to her. Finally, my most deeply felt thanks go to Drew Amery, who worked beside me all summer; collecting data, writing and debugging programs, and plotting the data. Many thanks, Drew, and good luck in graduate school. I'm sure you'll do well in all you choose to do.

Abstract

Initial studies of the technical feasibility of using machine vision and color image processing to measure plant health were performed. Wheat plants were grown in nutrient solutions deficient in nitrogen, potassium, and iron. An additional treatment imposed water stress on wheat plants which received a full complement of nutrients. The results for juvenile (less than 2 weeks old) wheat plants show that imaging technology can be used to detect nutrient deficiencies. The relative amount of green color in a leaf declined with increased water stress. The absolute amount of green was higher for nitrogen deficient leaves compared to the control plants. Relative greenness was lower for iron deficient leaves, but the absolute green values were higher. The data showed patterns across the leaf consistent with visual symptoms. The development of additional color image processing routines to recognize these patterns would improve the performance of this sensor of plant health.

Summary

Wheat plants were grown in solutions deficient in nitrogen, potassium, and iron. Another treatment was to impose a water stress on a plant. Color images of individual leaves were acquired with a machine vision system. Color image processing routines were written to compare the quantified green values and the green trichromatic coefficients for the treatment and control leaves.

Green trichromatic values tended to decline with increased water stress. Green values were higher for nitrogen deficient leaves. Green trichromatic values were lower for potassium deficient leaves, and patterns in the curve matched the location of symptoms. In extreme conditions, iron deficiency resulted in a higher green value and considerably lower green trichromatic values. Patterns in the green curve seemed to match the expected striping pattern. Iron deficiency symptoms developed very quickly and images of classical symptoms were not captured.

This evidence gives rise to expectations that multispectral imaging, combined with additional image processing will be able to clearly detect nutrient deficiencies. A nondestructive means of measuring plant health will lead to the development of a sensor for automatic control of nutrient delivery systems.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

INTRODUCTION

1.1 DESCRIPTION OF THE PROBLEM

Plants grown for food are necessary components of life support systems for long-term space voyages. Current research on a controlled ecological life support system (CELSS) has shown that plants grown in liquid cultures may rapidly develop nutrient deficiencies which affect the performance of a CELSS. If this situation were to occur on a long-term space voyage, it would pose a serious threat to the crew's food supply and life support system.

In terrestrial production systems, low-cost labor is used to visually inspect plants and diagnose disorders. Often the investigation requires tissue analysis which is an invasive, destructive method. In astrocultural production (growth of plants in space) the procedure must be nondestructive and automated to reduce labor and provide the rapid sensing necessary for feedback control of nutrient delivery.

1.2 NUTRIENT DEFICIENCY SYMPTOMS

Much has been written describing the visual symptoms of nutrient deficiencies in plants(1,2,3,4). The following discussion provides an overview and is by no means exhaustive. The lack of sufficient **nitrogen** stunts plant growth, leaves are small and thin. Leaf color is pale-green, yellow-green, or uniformly yellow. Chlorosis (yellowing) is usually more pronounced in older tissue since nitrogen is mobile within plants and tends to move from older to younger tissue when nitrogen is limited. Older leaves first yellow at the tips. In severe cases, tips and margins of older, mature leaves may turn brown (*firing*).

Phosphorus deficiency also stunts plant growth, but leaves turn dark-green, bronzy, reddish-purple, or purplish. Root and grain development is poor and maturity is delayed. Tips of leaves turn brown and die.

Potassium deficiency slows the growth of plants, and results in small fruit or shriveled seeds. Yellowing starts at the tip of older leaves, proceeds along the leaf edge to the base. Finally, tips and margins turn brown (*burn*), starting on mature leaves. Interveinal areas are brown, yellow, or *scorched* in appearance. Stalks are weak and easily lodged. Internodes are shortened.

Iron deficiency results in an interveinal chlorosis of young leaves. Veins remain green, except in severe cases, creating a visibly striped effect along the full length of the leaf. In very severe cases leaves become almost white and growth is stunted.

Leaves of **calcium** deficient plants become hard and stiff, and the margins roll upwards. The growing points and root tips may die. Foliage appearance is abnormally dark-green. Stems become weak, and blossoms and buds shed.

Lack of sufficient **magnesium** stunts plant growth and causes interveinal areas of older leaves to turn yellow. Leaves curl upward along margins. Older leaves show a *chainlike* yellow streaking. Chlorotic areas turn brown and die, starting at the leaf tip.

Sulfur deficiency results in small, spindly plants having a light-green to yellowish-green color in the younger leaves. The top of the plant shows the yellowing first. Affected leaves become thick and firm. Stems are hard, woody, and abnormally elongated and spindly. Symptoms resemble nitrogen deficiency, except there is a more general loss of color.

Manganese deficiency symptoms appear as an interveinal chlorosis of younger leaves. There is a gradual shift from pale green to a darker green next to veins, but there is no sharp distinction between veins and interveinal areas as with iron deficiency. Leaves may become all yellow in severe cases with formation of necrotic spots. Irregular gray specks may develop in oats, interveinal white streaks in wheat, and brown spots and steaks in barley. Plants are stunted with narrow, erect leaves.

Water is perhaps the most important nutrient and symptoms of moisture stress are expressed in many ways. Leaves loose turgidity, wilt, and stomates close. Because transpiration ceases, leaf temperature increases. Stems and leaves usually follow a diurnal pattern of contractions following turgor pressure changes due to transpiration and root uptake.

1.3 PROPOSED SOLUTION

One possible solution to the problem of monitoring plant nutrient status is to use machine vision and image processing to capture images of leaves and extract features which are characteristic of deficiencies. Al-Abbas, Barr, Hall, Crane and Baumgardner (5) showed that the significant differences which exist in the reflectance and transmission properties of maize leaves in the visible spectrum, 400 to 750 nanometers (nm) were due to pigmentation differences (mostly chlorophyll) achieved by nitrogen, potassium, phosphorous, sulfur, calcium, and magnesium deficiencies. Since blue light has a wavelength of approximately 435 nm, green about 546 nm, and red about 700 nm, a solid-state, color video camera should be able to capture these spectral differences as well as the spatial patterns which are characteristic of the various nutrient deficiencies. By interfacing the camera electronically to a computer through a frame-grabber, the array of data from an image can be transferred and stored on disk or memory. Programs can be written to extract and quantisize the salient features and classify the leaf image into one or more deficiency categories.

1.4 OBJECTIVE

The objective of this study is to prove that machine vision and image processing can be used to detect nutrient deficiencies.

PROCEDURES

2.1 PLANT GROWTH

In order to obtain leaves of known, characteristic symptoms, wheat seeds (cv, *Yecora Rojo*) were presoaked, then germinated in a deionized water film. Five days later, a half-strength Hoagland's solution was added. Seven days after initiation, plants were transplanted to aerated, 2 liter jugs filled with a nutrient solution as shown in Table 1. There were 3 replicates of 4 treatments: deficiencies of nitrogen, potassium, and iron; and a normal, or control treatment. Three wheat plants were grown in each jug for a total of 36 plants. Two days after transplanting, all treatments showed visible symptoms characteristic of a deficiency. Each day thereafter, the symptoms became more striking, especially on the new leaf tissues. In the iron and potassium treatments, leaves which formed during germination or during the period when the half-strength Hoagland's solution was available, did not show symptoms to the extent that the newly emerged tissue did. On the other hand, nitrogen deficiency symptoms were observed on new and old tissues.

Lighting was provided by two 400 watt high pressure sodium lamps suspended approximately 75 centimeters (cm) above the plants. The jugs were placed in two rows and covered an area approximately 30 cm x 100 cm.

Water stress was imposed on normal plants by removing them from the nutrient solution. At first the plants were placed in an empty jar, but when excess moisture was observed collecting on the inside walls, the plants were moved into room air at approximately 22 degrees C and 70% relative humidity. Images were taken every 10 minutes during the stress period and the entire process videotaped at 30 frames per second for the 7.5 hour test.

2.2 MACHINE VISION

The apparatus to capture and process the images consisted of a personal computer with a 200 Megabyte disk and tape cartridge for image storage and a color image frame-grabber to simultaneously digitize the red, green and blue signals from a Panasonic CCD color camera (model WV-D5000). A bellows was added to the camera to provide close-up images. Fields of view were on the order of 1 inch or less.

2.3 IMAGE PROCESSING

From each leaf image regions were selected for processing. The first step of processing was to compute the green trichromatic coefficient for each picture element (pixel) according to:

$$G_N = \frac{3G}{R+G+B} * 255 , \qquad [1]$$

where R, G, B refer to the quantified intensity of red, green and blue colors in the image, which are integers varying from 0 to 255. Because leaf width was variable, the pixel data were normalized so that the first element in a column was the bottom edge of the leaf and the 100th element was the top margin. Edges were located with a Sobel operator. This resulted in an image that was independent of field of view (size of leaf) and light intensity. By summing the G_N values across a row and dividing by the number of elements in the row (which varied from one image to the next but was constant within a region), a single curve could be obtained. This gives a visible, quantifiable method of comparing images from each of the treatments.

Because the trichromatic measures the relative amount of a color, the green component was analyzed to provide a quantitative measure of greenness.

From these data two additional features were computed: the average and standard deviation of the green trichromatic coefficients and the green values. The trichromatic average may be considered a single value for *relative greenness* while the standard deviation is a measure of *texture* (6,7,8). These three features: texture, greenness, and curve shape are the basis for detecting nutrient deficiences in plants via machine vision and image processing.

RESULTS AND DISCUSSION

3.1 WATER STRESS

Figure 1 shows the third leaf (newest) of a wheat plant supplied with a control nutrient solution on the 8th day after plants were placed in the jugs. Figure 2 shows the same leaf 7 hours

---

* Mention of a manufacturer is for informational purposes only and does not constitute an endorsement of the product to the exclusion of others.

and 40 minutes later in a wilted state. The results of processing the images are shown in Figures 3 and 4. Inconsistency in the Sobel operator's ability to locate the leaf edge may be responsible for the erratic data near the margins. On average the green trichromatic values are higher for the nonstressed or turgid leaf (Table 2), but are not statistically different, possibly due to the difficulty in locating the leaf margin with software. This decline in green with time was consistent for all images taken during the water stress test. One possible explanation is that the loss of turgor also changes the leaf reflectance.

3.2 NITROGEN DEFICIENCY

Figure 5 is an image of the third leaf of a wheat plant after 7 days in a nitrogen deficient solution. A section of this leaf was processed and the results are plotted in Figures 6 and 7. Except at the leaf margins, the green trichromatic values for the control (section A of leaf A1WS1) are higher than the nitrogen-stressed plant. However, the reverse is true for the absolute green values: the nitrogen-stressed values are lower (Table 2). Since a light, or pale-green has a higher absolute value than a dark green, this result was as expected. The control green trichromatic was higher than the nitrogen-stressed plant because the red and blue values changed between the two leaves. On a normal healthy leaf green contributes a higher percentage to the overall reflectance than it does on a nitrogen-stressed leaf.

3.3 POTASSIUM DEFICIENCY

Figure 9 shows a portion of the first leaf of a wheat plant grown in a potassium deficient solution for 4 days. A section was taken and the results are shown in Figures 9 and 10. Although the human eye detects a significant difference in the greenness across the leaf, Figure 9 shows that the actual green values differ very little. On the other hand the green trichromatic values show significant differences. Notice the drop of approximately 100 points in the trichromatic value, beginning with leaf width of 70%. This corresponds to the brown area in Figure 9 which is an obvious symptom of potassium deficiency.

3.4 IRON DEFICIENCY

Figure 11 illustrates the second leaf of a wheat plant grown in iron deficient solution for 9 days. Figures 12 and 13 and Table 2 show the green and green trichromatic values for this leaf. The green values for the iron-stressed plant are higher than the control, but the reverse is true for green trichromatic.

From the earlier description of iron deficiency one would expect the green values to undulate across a leaf, but they do not in Figures 12 and 13. However, on second look at Figure 11, the leaf does not show the usual pattern of iron deficiency. It appears that the second leaf, which had partially formed before the plants were placed in the deficient solution, fails to show a dramatic change in green pattern. Such patterns were observed as early as 2 days, but digitized images were not obtained.

CONCLUSIONS AND RECOMMENDATIONS

This study proves that visual symptoms of some nutrient deficiencies on individual leaves are detectable with machine vision and image processing. Green trichromatic values tend to decline with increased water stress. Green values are higher for nitrogen deficient leaves. Green trichromatic values are lower for potassium deficient leaves, and patterns match the location of symptoms. In extreme conditions iron deficiency results in a higher green value and considerably lower green trichromatic value. Patterns in the green curve seem to match the expected striping pattern. This evidence suggests that other symptoms may be detectable.

Much work remains to develop the technology of machine vision and image processing into a practical sensor for real-time monitoring of plant health. Numerous questions remain concerning positioning the camera and locating a leaf, backlighting, frontlighting, camera resolution, and image processing. Edge detection could be improved by smoothing the edge coordinate data and possibly curve fitting. For potassium deficient detection, an edge-following routine should be developed for both green and red trichromatic values. Using the red curve as a base, deviations in the green and red curves would detect brown spots along the margin of a leaf. Section width is important, but was not studied here. The section must be wide enough to average pixel to pixel differences, but not so wide that significant data (such as brown spots) are averaged into insignificance. A study should be undertaken to determine the section width that provides the greatest information to noise ratio. Another area of study should be to determine the best wavelength(s) to use. Previous research (5) has shown that other nutrient deficiencies can be detected in the near infrared and red spectrums. An optical bandpass filter mounted over the camera lens would enable the same machine vision and image processing to cover at least the near infrared spectrum. These additional studies to prove the capabilities of multispectral video imaging to nondestructively monitor plant health are warranted by the results of this study.

REFERENCES

1. Peterson, R.F., Wheat, Interscience Publishers, Inc., New York. 1965.
2. Western Fertilizer Handbook. 5th Edition. The Interstate Printers & Publishers, Inc., Danville, IL 61832. 1975.
3. Marschner, Horst. Mineral Nutrition of Higher Plants. Academic Press, Harcourt Brace Jovanovich. New York. 1986.
4. Small Grains: Nutrient Deficiency, Disease and Insect Damage Symptoms. J.R. Simplot Company, Minerals and Chemical Division, Pocatello, ID 83201. 1977.
5. Al-Abbas, A.H., R. Barr, J.D. Hall, F.L. Crane, and M. F. Baumgardner. Spectra of Normal and Nutrient-Deficient Maize Leaves. Agronomy Journal Vol 66 pp 16-20. 1974.
6. Gonzales, Rafael C. and Paul Wintz. Digital Image Processing. Second Edition. Addison-Wesley Publishing Company. Reading, MA. 1987.
7. Castleman, Kenneth R. Digital Image Processing. Prentice-Hall, Inc., Englewood Cliffs, NJ 07632. 1979.
8. Ekstrom, Michael P. (Ed). Digital Image Processing Techniques. Academic Press, Inc. Harcourt Brace Jovanovich. Orlando, FL 32887. 1984.

Figure 1. Photograph of the Wheat Leaf used as a Control (Image A1WS1).

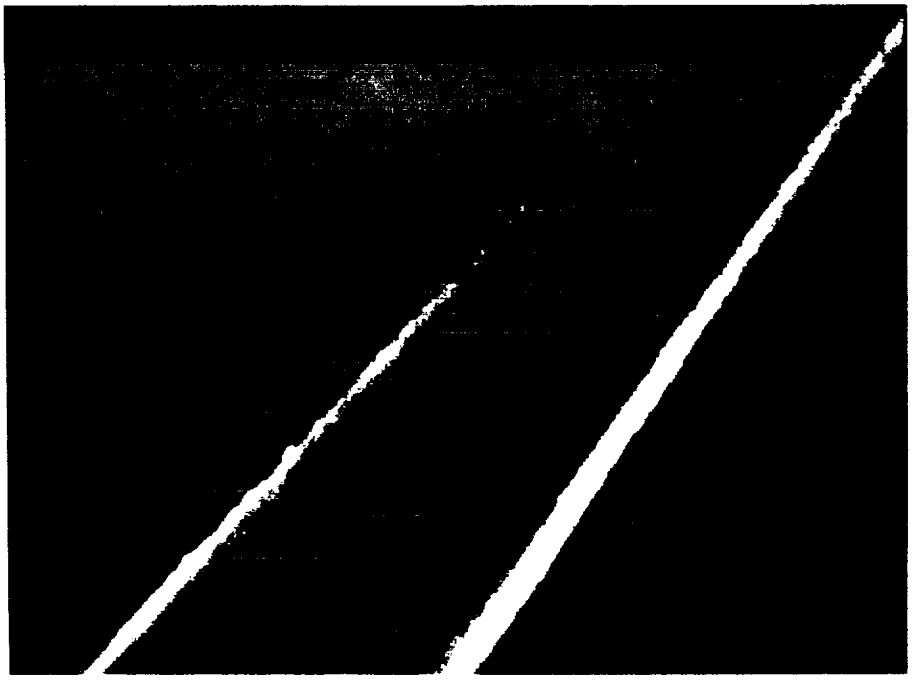


Figure 2. Photograph of a Wilted Wheat Leaf Image A1WS47), 7 Hours and 40 minutes after the Control Image.

ORIGINAL PAGE
BLACK AND WHITE PHOTOGRAPH

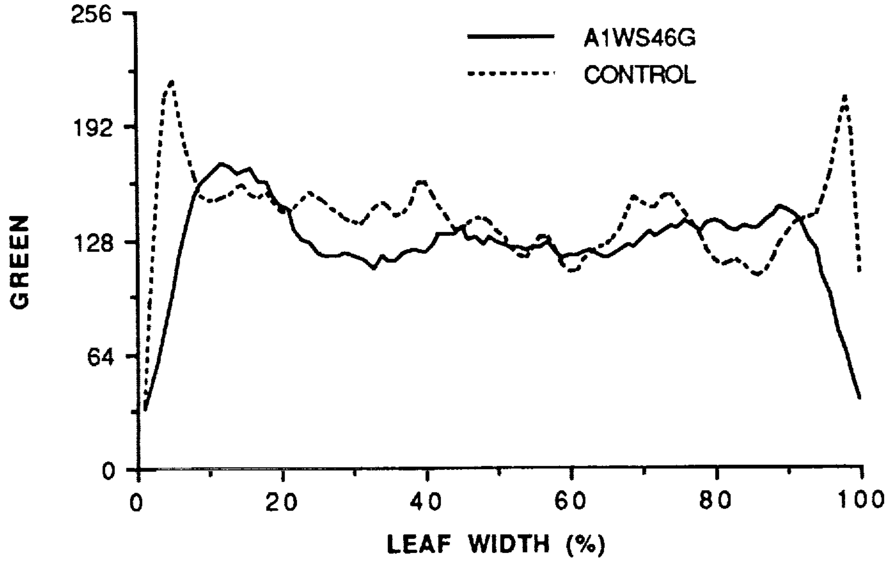


Figure 3. Absolute Measurements of Green for a Turgid (Control, A1WS1A) and Wilted (A1WS46G) Wheat Leaf.

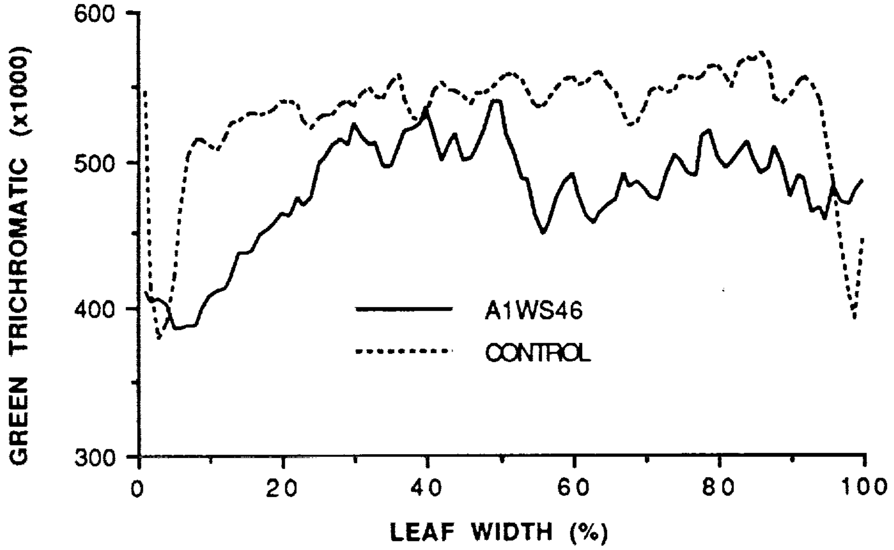


Figure 4. Relative Measurements of Green for a Turgid (Control, A1WS1A) and Wilted (A1WS46G) Wheat Leaf.

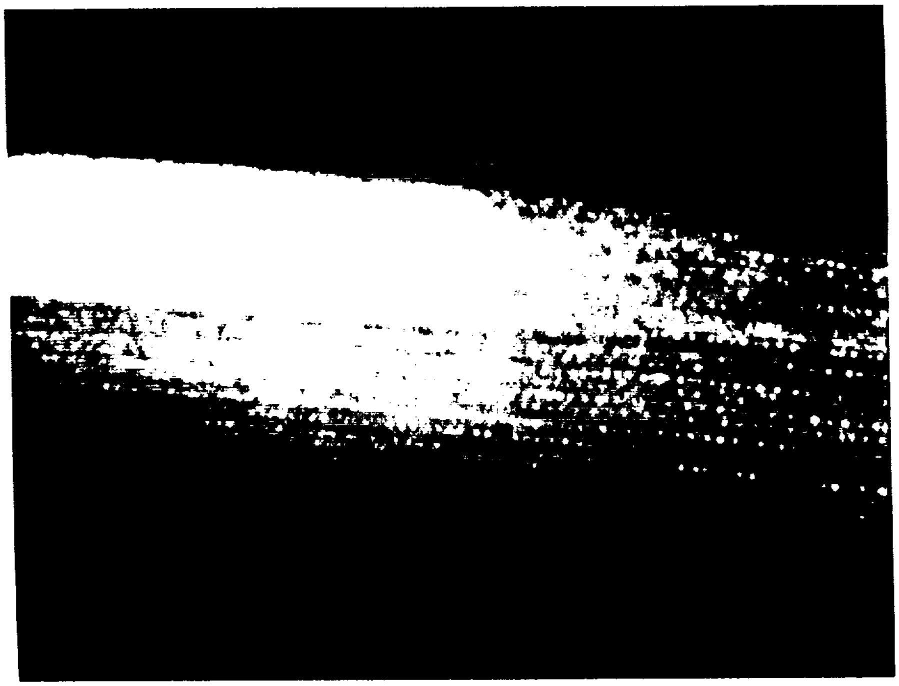


Figure 5. Photograph of a Nitrogen Deficient Wheat Leaf (N112A).

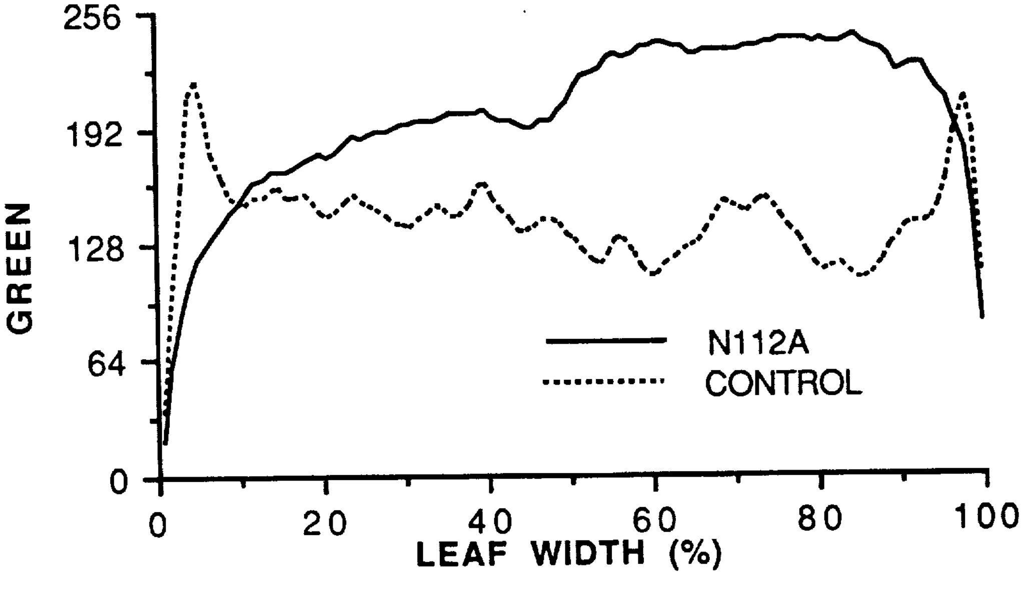


Figure 6. Absolute Measurements of Green for a Nitrogen Deficient Wheat Leaf (N112A).

ORIGINAL PAGE
BLACK AND WHITE PHOTOGRAPH

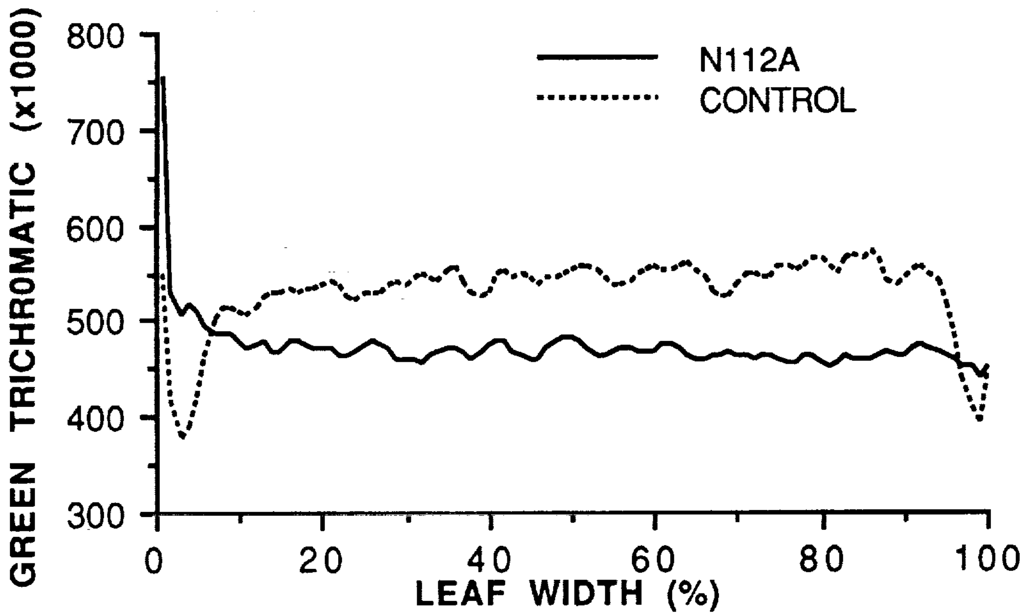


Figure 7. Relative Measurements of Green for a Nitrogen Deficient Wheat Leaf (N112A).

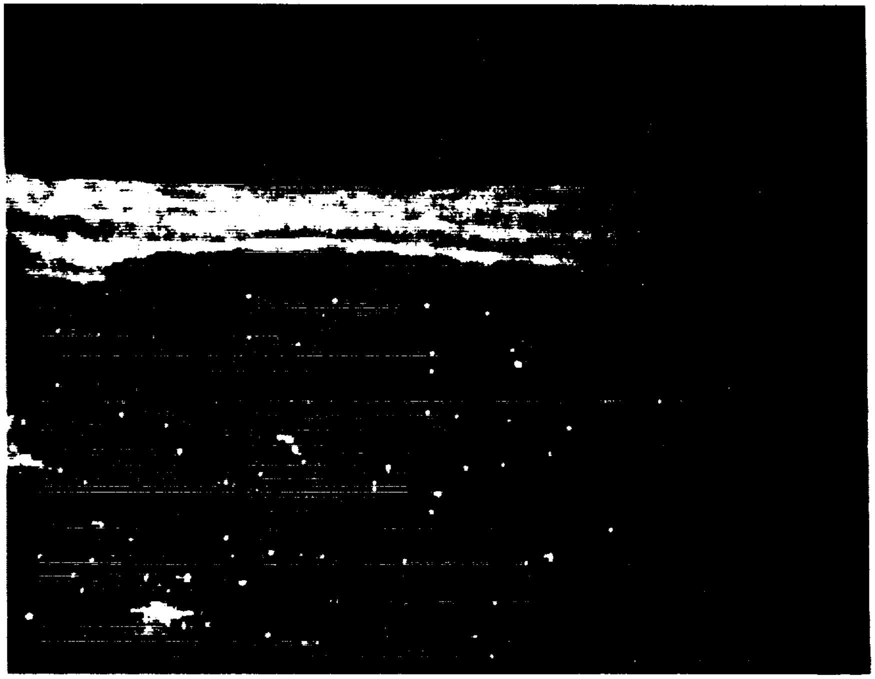


Figure 8. Photograph of a Potassium Deficient Wheat Leaf (K309A).

ORIGINAL PAGE
BLACK AND WHITE PHOTOGRAPH

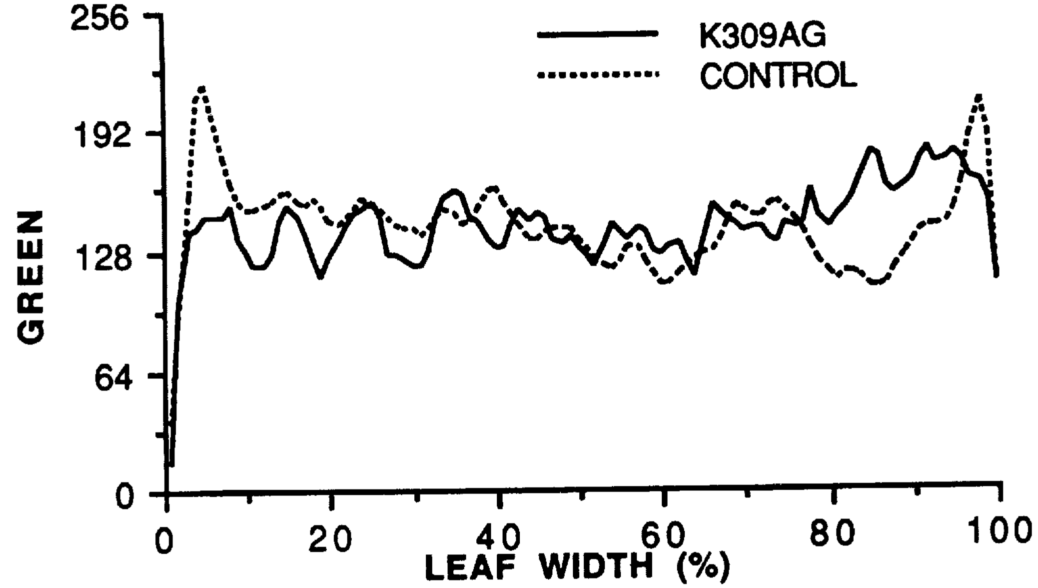


Figure 9. Absolute Measurements of Green for a Potassium Deficient Wheat Leaf (K309A).

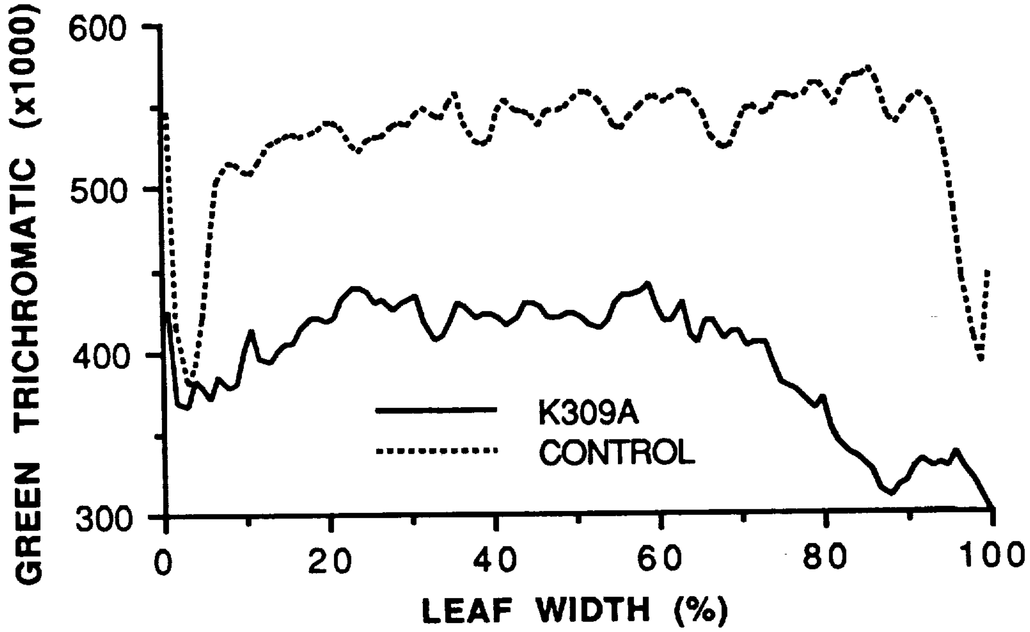


Figure 10. Relative Measurements of Green for a Potassium Deficient Wheat Leaf (K309A).

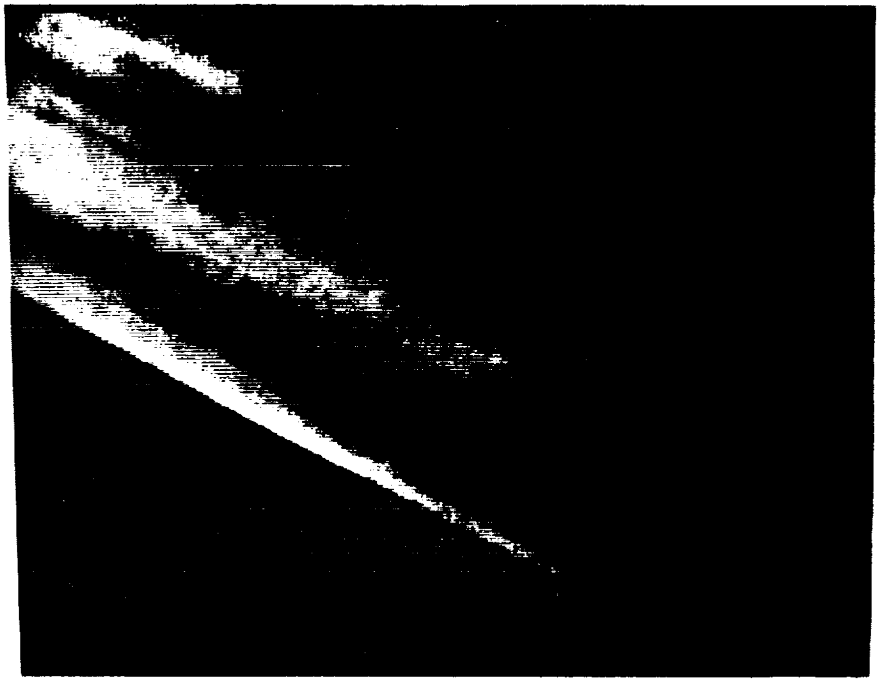


Figure 11. Photograph of an Iron Deficient Wheat Leaf (F312).

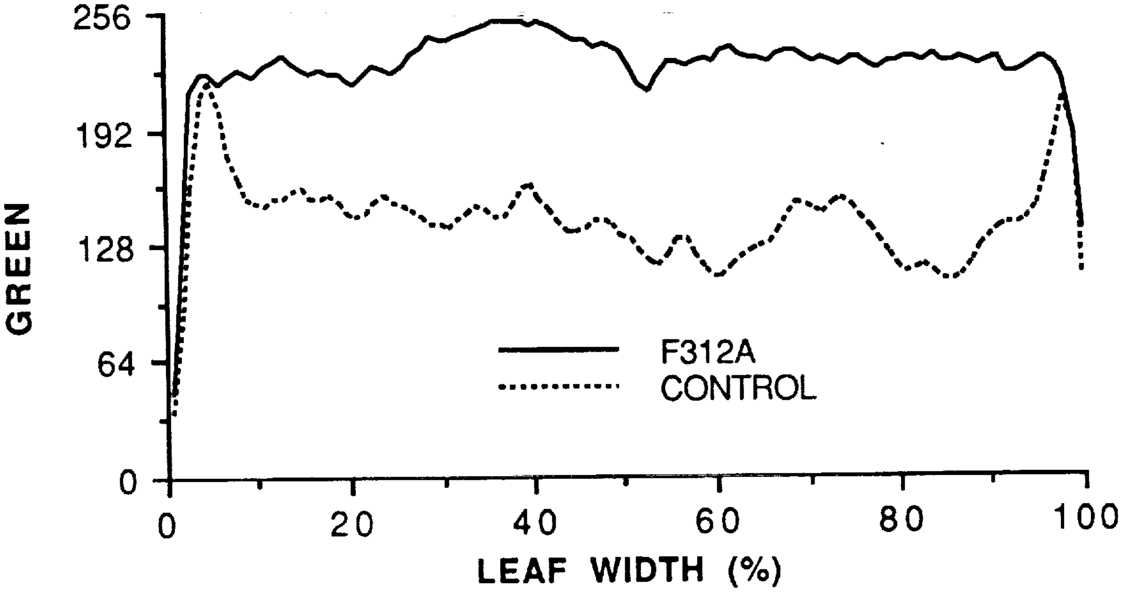


Figure 12. Absolute Measurements of Green for an Iron Deficient Wheat Leaf (F312A).

ORIGINAL PAGE
BLACK AND WHITE PHOTOGRAPH

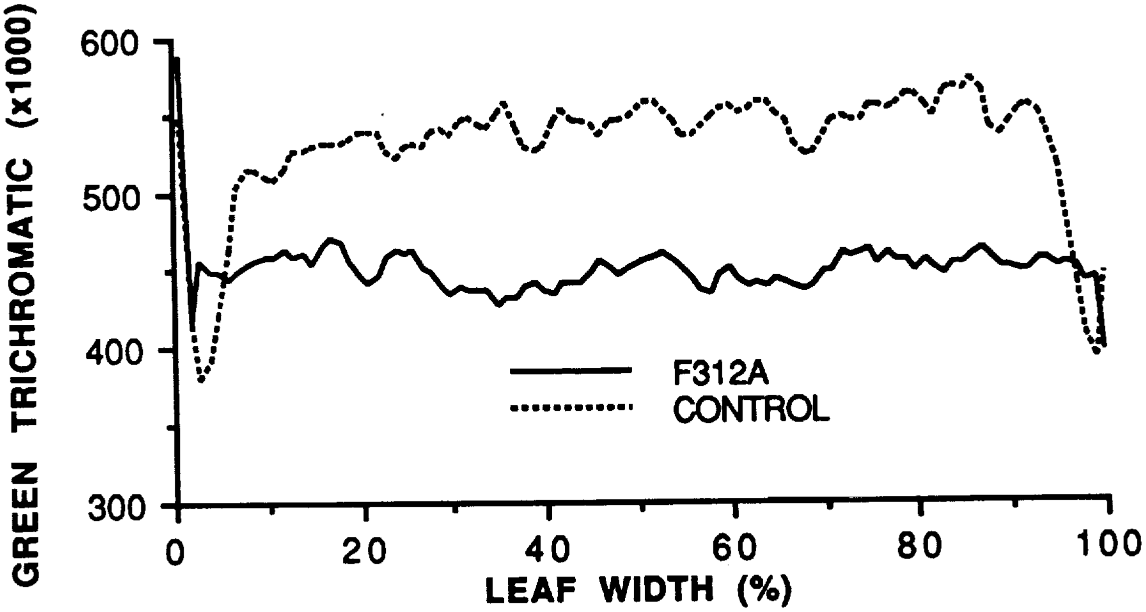


Figure 13. Relative Measurements of Green for an Iron Deficient Wheat Leaf (F312A).

Table 1. Nutrient Solution Formulation

	ADDITIONS			
Salt	**Normal** (ml/L)	**-N** (ml/L)	**-K** (ml/L)	**-Fe** (ml/L)
$CaCl_2 \cdot 4H_20$		2.50		
K_2SO_4		2.50		
KH_2PO_4	0.50	0.50		0.50
$Ca(NO_3)_2*4H_2O$	2.50		2.50	2.50
$(NH_4)_2HPO_4$			0.50	
NH_4NO_3			0.75	
KNO_3	2.50			2.5
$MgSO_4$	1.00	1.00	1.00	1.00
FeEDTA	11.15	11.15	11.15	11.15
micronutrients	1.10	1.10	1.10	1.10

ELEMENTAL CONCENTRATIONS (parts per million)

Element	**Normal**	**-N**	**-K**	**-Fe**
N	105	0	105	105
P	15.5	15.5	15.5	15.5
K	117	117	0	117
Ca	100	100	100	100
Mg	24	24	24	24
Fe	5.6	5.6	5.6	0
S	32	112	32	32
Zn	0.02	0.02	0.02	0.02
Cu	0.01	0.01	0.01	0.01
Mn	0.20	0.20	0.20	0.20
B	0.21	0.21	0.21	0.21
Mo	0.004	0.004	0.004	0.004

Table 2. Green (G) and Green Trichromatic (GT) Statistics

Treatment	Image	Mean	Std.Dev.
Control	A1WS1AG	140.0	16.2
	A1WS1AGT	530.0	39.5
WATER STRESS	A1WS46G	124.8	25.8
	A1WS46GT	478.2	37.3
IRON DEFICIENCY	F312AG	225.5	24.7
	F312AGT	449.1	17.7
POTASSIUM DEFICIENCY	K309AG	140.0	21.0
	K309AGT	393.5	39.1
NITROGEN DEFICIENCY	N112AG	198.0	42.8
	N112AGT	471.5	31.3

N90-16696

1989 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

ON THE SELECTION OF MATERIALS FOR CRYOGENIC SEALS AND THE TESTING OF THEIR PERFORMANCE

PREPARED BY:	Dr. John M. Russell
ACADEMIC RANK:	Associate Professor
UNIVERSITY AND DEPARTMENT:	Florida Institute of Technology Department of Mechanical and Aerospace Engineering
NASA/KSC	
DIVISION:	Mechanical Engineering
BRANCH:	Propellants and Gases
NASA COLLEAGUE:	Mr. W. I. Moore
DATE:	August 18, 1989
CONTRACT NUMBER:	University of Central Florida NASA-NGT-60002 Supplement: 2

ON THE SELECTION OF MATERIALS FOR CRYOGENIC SEALS AND THE TESTING OF THEIR PERFORMANCE

by

John M. Russell
Florida Institute of Technology

ABSTRACT

This report was prepared in support of efforts at the Kennedy Space Center to ensure that accidental leaks of fluids, especially cryogenic fluids, are reduced to a practical minimum. The report begins by addressing three questions: what mission must a cryogenic seal perform; what are the contrasts between desirable and available seal materials; and how realistic must test conditions be? The question of how to quantify the response of a material subject to large strains and which is susceptible to memory effects leads to a discussion of theoretical issues. Accordingly, the report summarizes some ideas from the rational mechanics of materials. The report ends with a list of recommendations and a conclusion.

Key words: Cryogenic seals, constitutive theory, seal assemblies

SUMMARY

A cryogenic seal must control the passage of cryogenic liquids through valves and couplings and must prevent mated parts from damaging each other. In doing so, it must accommodate large changes in absolute temperature. The design of any seal assembly is greatly expedited by the availability of a material that exhibits rubber-like behavior at all operating conditions, is inert to chemical attack, and does not present a fire hazard. No such material is available that meets all of the above requirements in cryogenic applications, so compromises and the diminutions of performance that attend them are difficult to avoid. Various design alternatives, such as fiber reinforced composite seals with a polymer matrix, foam polymers, metal spring-backed seal assemblies, and seals based on the idea of a inner tube filled with helium are discussed.

An example of how the routine installation of a TFE O-ring seal at KSC produced nonlinear strains and memory effects raises the question of how such effects can be quantified. There is a rich literature on the subject and this report includes a synopsis of some of the most important ideas proposed in it.

If the face of a polymer seal is scratched so as to allow a slender channel from the inside to the outside of the pipe it seals, then the rate of flow through that channel is of interest. The mathematical problem of finding the distribution of streamwise velocity across a channel cross section of arbitrary shape is formulated and solutions are presented for cross sections of triangular and semicircular shapes. Formulas for the rate of transport of fluid volume through such channels are then obtained by integration. For the purpose of comparison, these results are set alongside the velocity and flow rate formulas for a pipe of circular cross section.

The recommendations of the study are: (i) that future efforts be directed to the design of seal assemblies rather than the design of solid seals made of virgin plastic; (ii) the development of spring-backed seal assemblies, pneumatic tubes, foams, and fiber reinforced composites should be accelerated; (iii) the statically indeterminate problem associated with stretching of bolts beyond the elastic limit during cooldown of sealed joints should be incorporated into routine practice prior to testing and, if yielding is indicated, steps should be taken (through the installation of spring washers, extra long bolts, spacers, *etc.*) to eliminate it; (iv) tests should be undertaken to determine whether available foam TFE collapses at the temperature of liquid hydrogen, owing to the liquifaction of gases trapped in its cells; and, (v) some long range testing program (informed by appropriate theory) should be planned and undertaken with the aim of creating an archival literature on seal technology.

The conclusions of the present study are summarized as follows: the mechanical behavior of materials in general and seal materials in particular are captured (if at all) by its *constitutive functional*; if the material is susceptible

to memory effects and is subject to finite strains (as seal materials are in most designs of seal assemblies), then the constitutive functional can not be represented in terms of a few material constants (such as Young's modulus E, the shear modulus G, the coefficient of thermal expansion α, *etc.*). Rather, material functions and functionals are involved. Developers of quantitative prediction methods and experimentalists who provide data to support them must face this fact if they are to succeed.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

1. INTRODUCTION

Fuel for the space shuttle main engines is stored in the form of liquid oxygen and liquid hydrogen. These liquids are *cryogenic, i.e.* at atmospheric pressure, the absolute temperatures at which they boil are small compared to the absolute temperature of the coldest parts of the earth's atmosphere. The act of filling and draining the tanks to contain such liquids necessarily exposes pipes, hoses, valves, *etc.* through which such liquids are pumped to large changes in absolute temperature. The attendant changes in the geometry of mated parts and the mutual forces between them exacerbate all of the usual problems associated with the control of leaks.

Now the safety of personnel and equipment and the ability of machines to perform their functions can both be compromised by leaks of hydrogen, oxygen, and other fluids routinely pumped through pipes in a launch pad environment. Thus, while leaks of cryogenic materials are hard to avoid in space operations, the level of uncontrolled leaks that one can, in good conscience, tolerate must also be low.

There is little doubt that successful seal technology has been developed and employed in previous space operations undertaken by the United States. There is reason to suppose, however, that much of this technology was developed on an *ad-hoc* basis and is retained in the minds of experienced personnel, many of whom have left the space program or are planning to do so in the near future. While senior personnel can be encouraged to pass on an much of their best knowledge to junior personnel as they can, and while some of the rest can be developed by the junior personnel, one might hope that at some stage a more abstract, but also more durable, source of knowledge might be created. In this respect, there is a need for an *archival literature* on seal technology.

In chapter 2, various practical issues on seal technology are raised. Three key questions are posed, *i.e.* what mission must a cryogenic seal perform; (ii) what are the contrasts between *desirable* and *available* seal materials; and, (iii) how realistic must test conditions be if one is to have confidence that a component that performs well in leak tests will perform well in service?

In chapter 3, the discussion turns to theoretical issues. The manner in which a body composed of a given material deforms under load depends upon its history (including its deformation history), the nature of its supports (or, more abstractly, its surroundings) and the intrinsic properties of the material, to name three dependencies. Notwithstanding the availability of standardized tests, the intrinsic properties of a material can seldom, of ever, be measured directly. Normally, they are inferred by substituting measurable parameters (such as the length of a test specimen and its cross sectional area) which are *not* intrinsic to the material into one or more equations resulting from some theory of material behavior (such as Hooke's law) and solving for certain other parameters (such as Young's modulus) that characterize the material but may not be directly measurable. Simple theories of material behavior like Hooke's law do not suffice to describe the distortion of a cryogenic seal under typical service conditions. This insufficiency is due to several complications in the nature of that distortion. Thus, however one defines *strain*, the strains in real seals need not be small. Likewise, however one distinguishes between *elastic* and *inelastic* deformations, the deformations in real seals need not be elastic. Among the manifestations of inelastic deformation is the ability of some seal materials to extrude or *flow* at ambient conditions.

A literature on seal technology will not be archival if it is filled with vague terms. There is, therefore, a need for specific mathematical definitions of *deformation history, strain, elastic, solid, fluid, viscoelastic*, and other terms relevant to the technology of cryogenic seals. The first main part of chapter three, titled 'Aspects of the rational mechanics of materials', furnishes a primer on terms of this nature as they are used by Walter Noll and Clifford Truesdell in their comprehensive treatise *The Nonlinear Field Theories of Mechanics* (Reference 1). The second main part of chapter three, titled 'Flow through small holes', is relevant to the prediction of leak rates through a given hole across which a given pressure difference is applied.

Chapters four and five are titled 'Recommendations' and 'Conclusions', respectively. They are summaries in their own right, and so do need further synopsis here.

2. PRACTICAL ISSUES

2.1 GUIDING QUESTIONS

As was stated in the Introduction, this chapter is motivated by three questions, *i.e.*

Q1 What *mission* must a cryogenic seal perform?

Q2 What are the contrasts betweeen *desirable* and *avialable* seal materials?

Q3 *How realistic* must test conditions be if one is to have confidence that a component that performs well in leak tests will perform well in service?

Question Q1 is addressed under subheading 2.2 below. Question Q2 is addressed under subheadings 2.3, 2.4, and 2.5, and question Q3 is addressed under subheading 2.6.

2.2 MISSION OF A TYPICAL CRYOGENIC SEAL

A cryogenic seal must control the passage of cryogenic liquids through valves and couplings and must prevent mated parts from damaging each other. It must, moreover, perform this mission under a variety of constraints. Thus, the changes in the geometry and the mutual contact forces between separate parts of an assembly that attend large changes in absolute temperature must not induce inordinately large stresses within the assembly, nor must they allow gaps to open within it. The hardness of the seal material at cryogenic temperatures must not exceed that of the metal surfaces in touches. Enough toughness of the seal should be retained at cryogenic temperatures so that it does not become inordinately vulnerable to scuffing by adjacent metal parts with small surface defects or to brittle fracture.

2.3 PROPERTIES OF A HYPOTHETICALLY IDEAL SEAL MATERIAL

A seal material is desirable or undesirable accordingly to whether it does or does not expedite the design of valves and couplings. In this sense, the desirablility of a seal material depends upon how many of the following attributes it has:

A1 It can undergo large-strain deformations elastically over a large temperature range.

A2 It would be invulnerable to attack by fuels, oxidizers, and other fluids commonly pumped through pipes at KSC.

A3 It would not exacerbate fire hazards, even when placed in contact with liquid oxygen, nitrogen tetroxide, or other oxidizers.

A4 It would be soft enough not to scratch adjacent metal parts, but tough enough to resist scratching by them.

A5 It would be non-porus (*i.e.* it would not permit seepage of liquid through it).

A6 It would be inexpensive, easy to form into parts, and commercially available.

2.4 PROPERTIES OF AVAILABLE POLYMER PLASTICS

An extensive collection of physical property data on polymer plastics was compiled in the early 1970's and published as a monograph by the National Bureau of Standards (NBS Monograph 132, Reference 2). According to these data, available polymer plastics may readily be found which possess attributes 2, 3, 5, and 6 above at all temperatures and condition 4 at ambient temperatures.

Most cryogenic seals at KSC are made of Polytetrafluoroethylene (TFE) and its copolymer with Hexafluoropropylene (FEP). According to the authors of Reference 2, FEP has a lower crystalline melting temperature than TFE, which limits the highest service temperatures that parts made from FEP can stand. On the other hand, fabrication of parts by efficient techniques such as injection molding and extrusion is possible with FEP but not possible with TFE, owing to the lower melting temperature of the former. Fabrication of parts from TFE, by contrast, involves compression of powder. There is little difference between TFE and FEP in other respects, at least as regards the list of attributes cited above. If enough care is taken with respect to design and maintenance, seals made from FEP and TFE may perform satisfactorily under cryogenic conditions. Most of the leaks that do occur appear to be related to the inability of TFE and FEP to meet condition A1 above either at ambient conditions (when it is subject to stress relaxation) or at cryogenic conditions (when it fractures at small to modest strains). Figure 1 represents the temperature dependence of the ductility of three plastics. Here, 'ductility' is defined to be elongational strain at rupture of a test specimen loaded in uniaxial tension. 'Strain' is defined in the one-dimensional engineering sense, *i.e.* the ratio of the increase in length under load to the unloaded length. Figure 1 shows that at T = 20°K (the normal boiling point of liquid hydrogen) the ductility is less than 0.05. Figure 2 illustrates the results of a uniaxial tensile test of TFE specimens at T = 20°K. The three curves show the effects of radiation on the stress-strain diagram. The rightmost curve is the one of greatest interest to this report. It shows that rupture occurs at a tensile strain of 0.032. Regardless of how one classifies the behavior of TFE at this temperature, one can not regard its behavior as *rubberlike*.

The inability of TFE to spring back to its original shape after a compressive load is removed is illustrated in Figure 3. Each curve corresponds to a particular temperature and includes the stress-strain diagram for loading followed by unloading. The unloading curve does not retrace the loading curve, *i.e.* compressive loading of TFE exhibits *hysteresis*. Each time a valve containing a TFE seal is closed or two pipe sections are bolted together with a TFE seal between them, irreversibe work is done in deforming the seal. This circumstance limits the number of times that a valve with a TFE seal may be closed and then reopened without seriously altering the seal geometry. Figure 3 suggests, however, that the hysteresis loop can be reduced to an arbitrarily small size by placing a sufficiently low ceiling on the maximum compresive stress for which the valve seal is rated. Unfortunately, this solves one problem at the cost of creating another. It is a fact of experience that lightly closed valves are more prone to leakage that are tightly closed ones, so there is a conflict between the need to prevent leakage in any one cycle of closing and reopening and the need to extend the useful life of the valve over many such cycles.

The hysteresis effect shown in Figure 3 is one of at least two inelastic effects exhibited by TFE and FEP at ambient conditions that can be gleaned from a perusal of NBS monograph 132. The other effect is *stress relaxation*. A simple device for testing the stress relaxation of a gasket is illustrated in Figure 4. A metal bolt is presumed to behave according to linear elastic theory and its stress-strain curve is determined experimentally. Once cali-

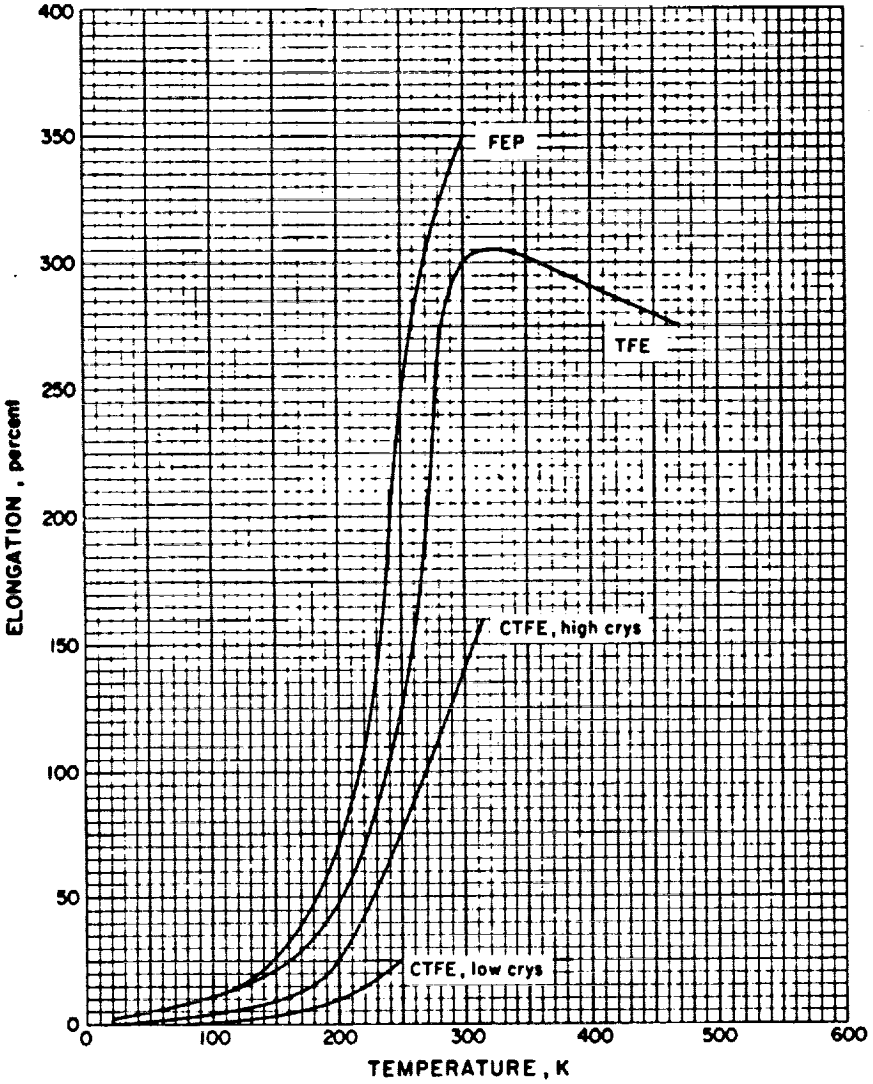


FIGURE 1. Temperature dependence of the ductility of three polymer plastics. Reproduced from Reference 2. Data represent average results from many different investigations.

ORIGINAL PAGE
BLACK AND WHITE PHOTOGRAPH

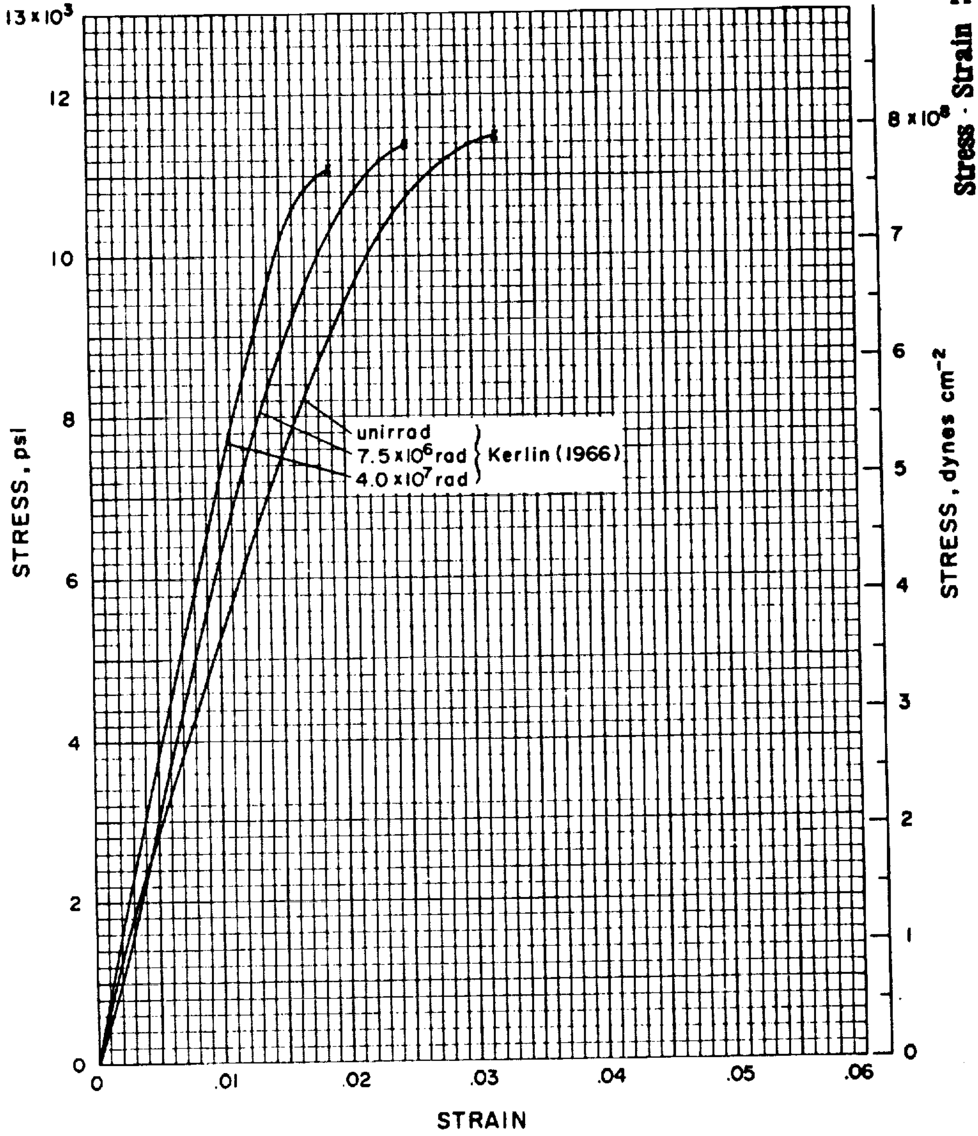


FIGURE 2. Behavior of TFE in a tensile test at 20°K. Data from Kerlin, E.E. and Smith, E.T. 1966 'Measured effects of the various combinations of nuclear radiation, vacuum, and cryotemperatures on engineering materials'. General Dynamics, Ft. Worth MSFC FZK-290. Reproduced from Reference 2.

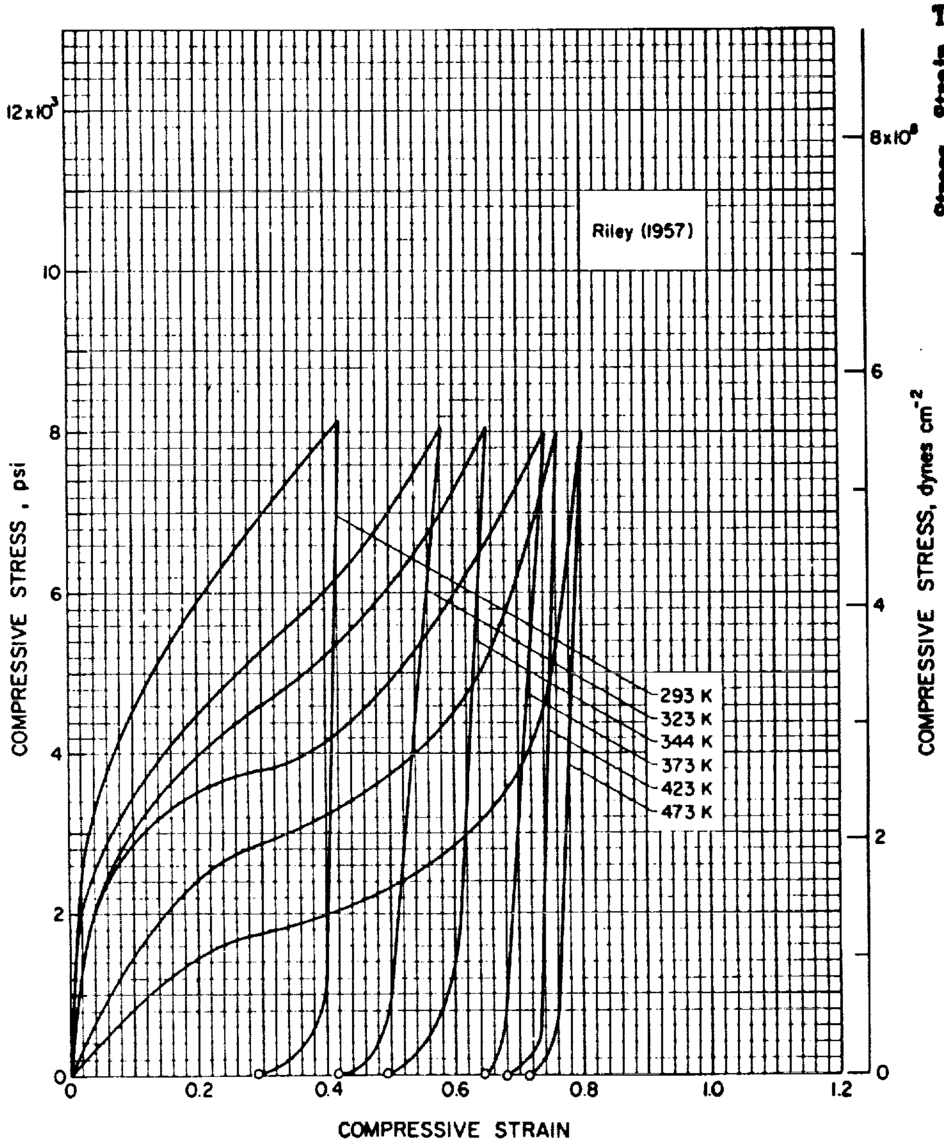


FIGURE 3. Plastic deformation and hysteresis of TFE in compression. Data from Riley, M.W. 1957 'Selection and design of fluorocarbon plastics.' *Materials and Methods*, Vol. 129. Figure from Reference 2.

F 38

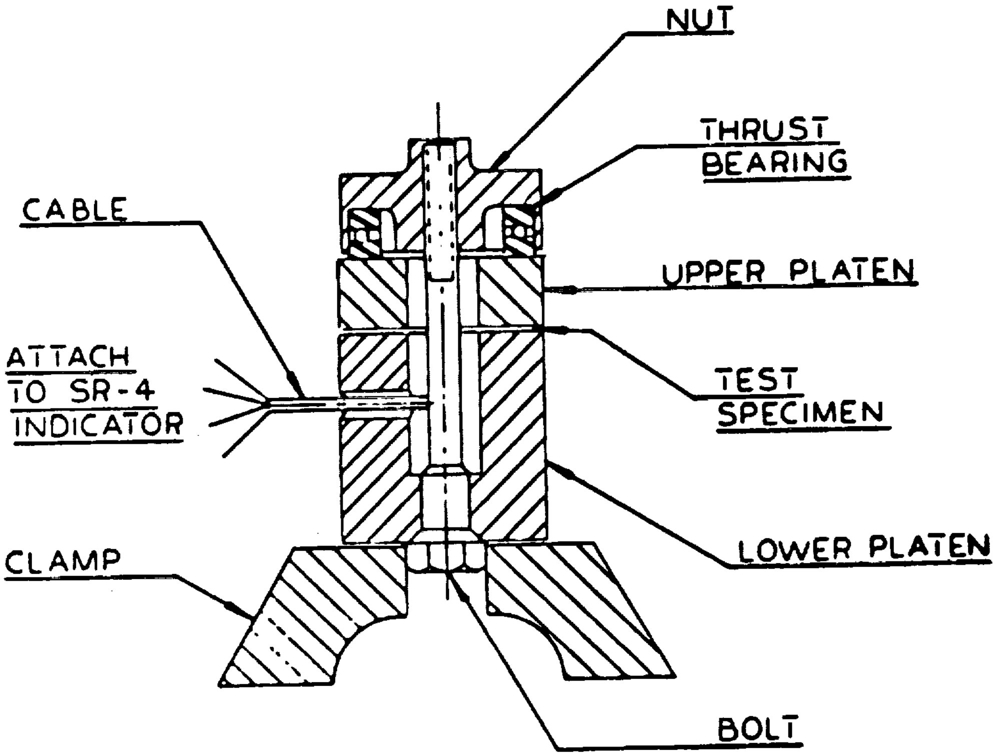


FIGURE 4. Sketch of a *Relaxometer*, a device for measuring long-term stress relaxation of gaskets. American Society of Testing and Materials, *Annual Book of ASTM Standards*, 1988.

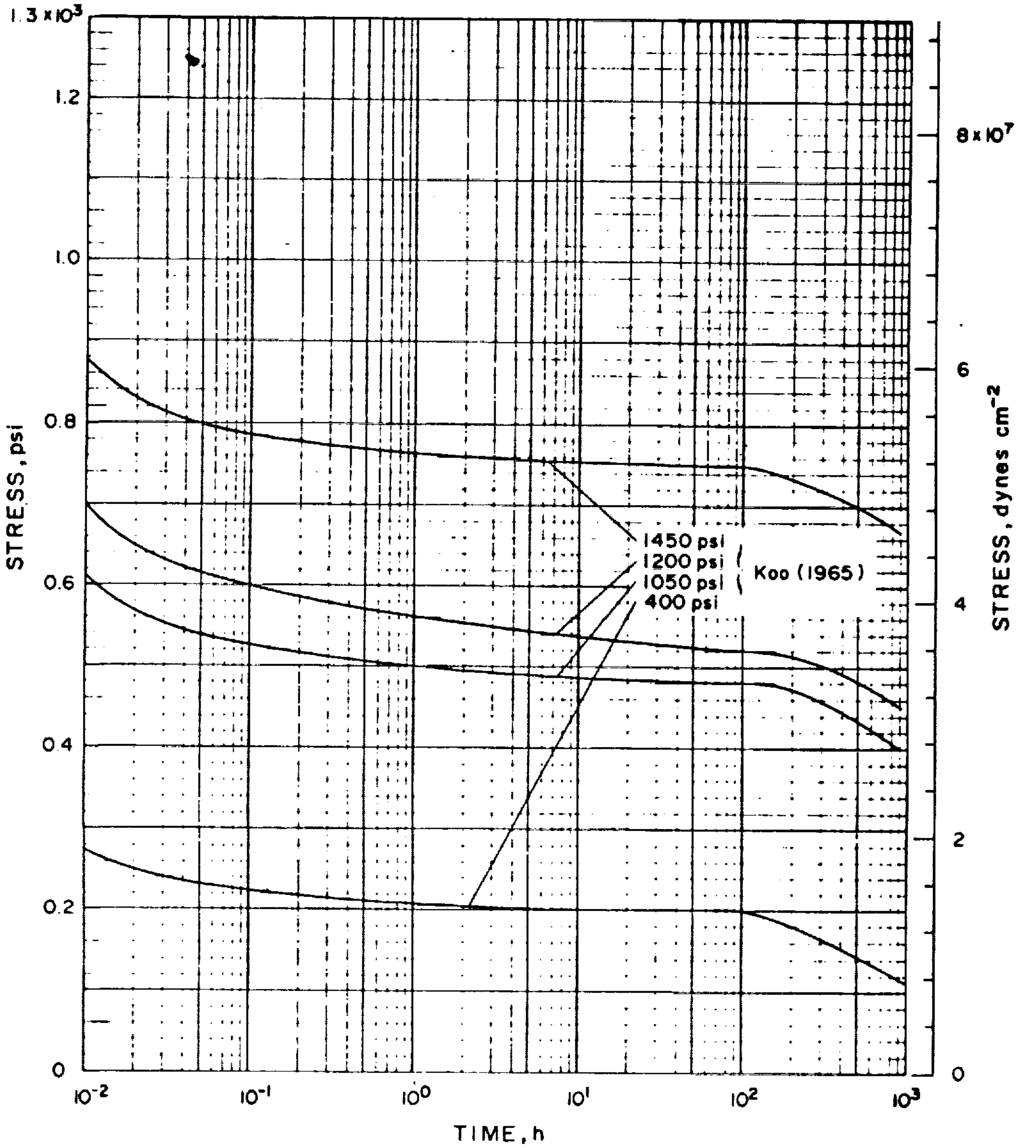


FIGURE 5. Relaxation of compressive stress in a TFE gasket held at constant compressive strain (curve labels indicate *initial* stress). Data from Koo, E.P., Jones, E.D. & O'Toole, J.L. 1965 'Polytetrafluoroethylene as a material for seal applications'. American National Conference on Fluid Power (21st), Illinois Institute of Technology, Chicago. Figure from Reference 2.

ORIGINAL PAGE
BLACK AND WHITE PHOTOGRAPH

brated, the bolt becomes a gauge for determining the compressive load on the gasket. A gasket may be loaded and the compressive load in it may be recorded at prescribed intervals of time over a period of hours, weeks, or months. Results of some tests of this kind are shown in Figure 5. It is an unfortunate fact that the gasket continues to relax in time after fourty days.

2.5 LIMITATIONS OF UNIFORM MATERIALS AND POSSIBLE ALTERNATIVES

The discussion under subheading 2.4 above indicates that TFE and related polymer plastics are not ideal materials for seal applications. Their most serious limitations, in the context of this report, are (i) hysteresis effects in large-strain compressive loading and unloading at ambient conditions; (ii) long-term stress relaxation at ambient conditions; and (iii) loss of ductility at cryogenic temperatures. The question of whether new compounds can be discovered and developed which exhibit true elastomeric behavior at ambient *and* cryogenic conditions and which have all of the other attributes listed in subsection 2.3 above can only be addressed by persons having expertise in the field of polymer chemistry, a field outside the present author's competence. There is, however, circumstancial evidence that suggests the degree of difficulty of the problem. Thus, one may argue that the need for a cryogenic elastomer was just as strong in the early days of the space program as it is today. Noting that no such material has appeared in the three decades since the space program began, one might reasonably surmise that the development of a cryogenic elastomer is a difficult task, to be accomplished, if at all, only by talented persons authorized to conduct basic research with few restrictions over a long period of time.

If the prospects of obtaining an ideal cryogenic elastomer are bleak, one is led to ask whether the need for them may be reduced by a design approach which separates functions. Thus, a cryogenic gasket must have *variable geometry* and must fulfill conditions A2-A6 listed under subheading 2.3 above. Metal springs may be designed to perform elastically over the full temperature range of interest here. Gasket assemblies consisting of thin portions of TFE backed by a metal spring may perform all of the tasks that one would assign to a solid part made from a cryogenic elastomer. Such spring backed gaskets assemblies are, in fact, used in some key couplings in conduits connected to the space shuttle orbiter and their service record (as far as I know) is satisfactory. Spring-backed gasket assemblies are, of course, less simple than gaskets cut out of uniform material and this extra complexity must be viewed as a disadvangage. One may also suppose that the ease with which a spring-backet gasket assembly may be fabricated is a strong function of its size. The smaller is the seal, the greater is the problem of miniaturization.

An alternative approach which fits in with the spring backed concept is one decribed to me by Mr. James Fesmire (DM-MED-4), namely an inflatable seal. The seal resembles an old fashioned inner tube and is filled with a gas like helium, which does not liquify at the intended service termerature. If the tube membrane is thin enough, then the overall shape of the tube may be altered much more readily than if it were solid (even if it is at cryogenic temperatures). Such ideas should be given serious consideration in the future.

The problem of stress relaxation in polymers may be mitigated, to some degree, through the use of composite materials. Thus, if TFE is reinforced by glass fibers in mat or whisker form, one might expect that slow flow of the polymer would result in slow loading of the glass fibers leading, at length, to a limiting equilibrium state in which the fibers eventually carry all of the load. To illustrate the point, suppose that a pad of fiberglass mat is dipped in oil and then clamped between vise jaws. In equilibrium, the portion of the vise load borne by the oil would be small compared to the part borne by the glass. The nature of this equilibrium might be similar if the matrix surrounding the fibers were TFE instead of oil. Glass reinforced TFE is, in fact, used as seal material at KSC. The present author was furnished with samples of such materals with the trade names Fluorogold® and Fluorogreen®. When handled, the most conspicuous feature of these materials is their stiffness. One may surmise that, while mitigating the problem of long-term stress relaxations, these materials exhibit the same problems as uniform polymers with respect to hysteresis in loading and unloading. Indeed, the greater stiffness of these composites may exacerbate the problem of designing joints that retain their tightness to leaks in the presence of large and rapid changes in absolute termperature.

The problems associated with low ductility of TFE at cryogenic temperatures may be mitigated, in part, through the use of foam TFE. The present author was shown samples of a foam TFE material marketed under the trade name GORE-TEX®. This material is delivered in tape form with an adhesive backing. It compresses noticably under finger pressure. To be used as a gasket, it must be formed by hand into a ring shape and placed so as to overlap itself. Gaps that result at the place of overlap or elsewhere must be wrung out by clamping pressure during installation. Such foam tape exhibits some rebound when subject to compression, but it is by no means elastic. The tightening of joints sealed with such tape results in some extrusion of the material. Of course, such a material would not be suitable as a bearing surface in a valve or other part that is subject to repeated loading and unloading. There is also reason to believe that foam TFE is susceptible to the same phenomenon of long-term stress relaxation as is uniform TFE.

If the gas that fills the microscopic bubbles in foam TFE is one which liquifies at temperatures above that of, say, liquid hydrogen, then exposure of foam TFE to such low temperatures may cause the microscopic bubbles to collapse. If such a collapse did take place on a microscopic scale its effect on a macroscopic scale might be a drastic shrinkage of the seal material. Such an effect would hardly serve the purpose of preventing leaks. One area in which laboratory testing might be especially illuminating is the response of such foam TFE to immersion in liquid hydrogen. Such tests would, of course, be complicated owing to the difficulty of taking measurements in a liquid hydrogen environment.

The liquifaction at the temperature of liquid hydrogen of the gas used to inflate the bubbles in foam polymers can be avoided by inflating them with gaseous helium. Unfortunately, helium diffuses through TFE quite readily (I am indebted to Mr. Cole Bryan for alerting me to this fact). If, therefore, a foam plastic consisting of microscopic helium balloons were to be formulated, it could not be made of TFE. The problem of finding (or formulating) a material with the requisite impermeability to helium may well be insurmountable, but the possible benefits suggest that the search might be worthwhile.

Seals made of cork are unsuitable for most applications at KSC owing to their vulnerability to chemical attack and to the fire hazard they present (especially in the presence of liquid oxygen). A closed TFE sheath with a cork core may have some features in common with spring-backed gasket assemblies, however, and may even be a realistic option in the case of small seals, in which miniaturization of spring-backed seal assemblies is problematic.

2.6 HOW REALISTIC MUST TEST CONDITIONS BE?

The degree to which one requires that test conditions be realistic depends upon the use one intends to make of data produced by those tests. Thus, if a new valve or coupling is meant to be an exact copy of a proven design, tests should address the question of whether the given test specimen is a faithful replica of its forbears. Checks to determine that the specimen is made of the proper materials and has the correct dimensions and surface smoothness followed by a routine test of leak tightness at ambient conditions may provide adequate confidence that the specimen is in conformity with required specifications. If, alternatively, the test specimen is a prototype of a completely new design, and if the purpose of the test program were to provide assurance that the component will perform well under a variety of critical conditions in service, then a much more exhaustive test program would be called for. The case when a seal assembly is redesigned and retrofitted to existing hardware falls in between these two extremes.

Considering the complicated interactions between changes in geometry and loading that attend large changes in absolute temperature, there is little justification for the hope that simple correlations between leak rates at ambient conditions and leak rates at cryogenic conditions would be reliable except in the most restricted circumstances, *i.e.* when a separate correlation is developed for each component assembly and history of temperature variations. Thus, cryogenic testing would seem to be an essential part of any component recertification program following any significant change in its design.

Some analytical tools for the prediction of leak rates are desirable, however, even if they are only used in the preparation of test apparatus and instrumentation. Some discussion of the problem of predicting the flow of fluid through small holes will be found in subsection 3 of chapter 3 below.

3. THEORETICAL ISSUES

3.1 BACKGROUND

During my first visit to the Propellants and Gases Prototype Laboratory this summer, I was shown the diassembled parts of a twelve-inch aluminum check valve whose purpose was to control the passage of liquid hydrogen. The two major parts of the housing mate along a stepped surface. The exterior corner of one of the steps is bevelled and an O-ring seal is installed in the gap that is formed between the (bevelled) exterior corner and its (unfilleted) interior mate. In a cross sectional view, this gap forms a right triangle. The nominal design at the time it was shown to me specified that the O-ring would be of virgin TFE. Its initial configuration would cause it to fit snugly into the interior corner of its seat. In a meridional plane, its cross section would be rectangular. During installation, then, this rectangular section O-ring is squeezed within an triangular gap.

I was shown a new O-ring and an old one that had been installed, subject to a cryogenic leak test, then removed after disassembly. Some features of the old O-ring were easy to interpret. Thus, the bevelled part of the housing that contains the exterior corner creates a conical surface that flattens one corner of the O-ring. The O-ring takes a permanent set which changes its shape in a meridional cross section in a obvious way. Far less obvious was the permanent set taken by the O-ring with regard to *azimuthal expansion*. The increase in its overall diameter was at least one half inch. It exhibited a conspicuous slack when placed onto the part of the housing containing the interior corner. Since it fit snugly when new and since installation of the O-ring into the housing did not subject it to any circumferential stretching, a question arises naturally as to what causes this slackening. Mr. Ken Ahmee, who escorted me to the Prototype lab, posed this question to me and I had no ready answer to it.

In studying the literature on the behavior of materials subject to finite strains, however, I did find some references to prior work that touches on similar issues. In 1909, J.H. Poynting (Reference 3) published the results of an experiment on the torsion of a straight wire. According to the equations of linear elasticity, a simple torsional strain of the wire, unaccompanied by any longitudinal compression or tension, should not cause any change in length of the wire. Poynting's results showed that the wire lengthened in response to the twist. He found that the lengthening is proportional to the square of the twist angle, an example of an effect that results from nonlinear straining of a material. In later years, Poynting found a way to measure a change in radius of the wire that attends its change in length. He also found that both effects were equally observable and more dramatic when the metal wire was replaced by a rubber rod.

The Poynting effect was independently rediscovered by R.S. Rivlin in 1947 by which time (as pointed out by Clifford Truesdell) Poynting's work was forgotten. The common denominator between the Poynting effect and the slackening of the O-ring described above is a coupling between shear straining in one plane and normal stresses both on the planes that receive the shear stresses and on the plane normal to the plane of the shear.

The number of people who know and care about the Poynting effect, nonlinear strains, normal stresses attendent to them, and the modeling of nonlinear material response in three dimensions with memory has never been large. Those people who have studied such effects have, however, produced a remarkable literature on it. In the hope of identifying those of their insights that are especially relevant to my work this summer, and of laying the foundations for future work, I spent a considerable portion of my time this summer studying that literature. The following section is a synopsis of what I learned.

3.2 ASPECTS OF THE RATIONAL MECHANICS OF MATERIALS

For want of a better term, I will refer to the people whom I described in the last paragraph (namely, those who have relaid the foundations of modern continuum mechanics) as the *foundationists*. The main structure of their theory is presented in (i) the treatise *The Nonlinear Field Theories of Mechanics* (Reference 1); (ii) *The Foundations of Mechanics and Thermodynamics, Selected Papers by Walter Noll* (Reference 4); and (iii) various specialized papers published primarily in *The Archive of Rational Mechanics and Analysis*. All of the ideas summerized in section 3.2 will be found in the above publications.

3.2.1 BODIES, CONFIGURATIONS, AND MOTIONS. In the formulation of a theory, some ideas must be treated as *primative notions, i.e.* notions that are not definable in terms of others more primative without engaging in circular reasoning. For present purposes, I will regard terms like *time, material point, geometric point, body, mass,* and a few others as terms that belong to this class.

Let each material point in a body B be distinguished from each other point in B by a unique *material point identifier* X. There is often an advantage in replacing this abstract identifier with a definite ordered triple of scalars (X_1,X_2,X_3), which may, for example, represent the cartesian position coordinates of the given material point at some reference time. In such a case, the scalars (X_1,X_2,X_3) are called the *material coordinates* of X.

The *configuration* χ of a body B is a function that maps particle identifiers X to their corresponding geometric positions x, *i.e.*

$$x = \chi(X) \qquad , \qquad X = \chi^{-1}(x) \quad .$$

The *motion* of a body B is a one-parameter family χ_t of configurations. The real parameter t here is the *time*. The above notation is then a shorthand for the more detailed notation

$$\mathcal{x} = \chi(X,t) \equiv \chi_t(X) \tag{3.1a}$$

$$X = \chi^{-1}(\mathcal{x},t) \equiv \chi^{-1}(\mathcal{x}) \quad . \tag{3.1b}$$

3.2.2 STRESS PRINCIPLE AND THE BASIC LAWS. Mechanics is endowed with certain basic laws, four of which may be stated in words as follows: (i) The mass of a body B is independent of time; (ii) there exists at least one reference frame, an *inertial reference frame*, relative to which the translational momentum of a body B is constant in time if and only if the resultant force exerted upon it by the surroundings is zero; (iii) the time rate of change of the translational momentum (with respect to an inertial reference frame) of a body B equals the resultant force exerted upon B by its surroundings; and (iv) the time rate of change of the rotational momentum (with respect to an inertial reference frame) of a body B equals the resultant torque exerted upon B by its surroundings. In popular terms, these laws are, respectively, (i) the law of conservation of mass; (ii) Newton's first law; (iii) Newton's second law; and, (iv) the law of moment-of-momentum.

The *stress principle* of Augustin-Louis Cauchy (1789-1857) asserts (in the paraphrase of Clifford Truesdell, reference 5) that 'upon any smooth orientable surface ∂V, be it an imagined surface in the body or the bounding surface of the body itself, there exists a field of *traction vectors* $\mathcal{t}_{\partial V}$, equipollent to the action of the exterior of ∂V and contiguous to it on that interior to ∂V.' Cauchy's principle, once accepted, permits the resultant force exerted on a body B by its surroundings to be expressed as a sum of two contributions, namely a *body force contribution* (of which gravitational force is the prime example) and a contribution from *contact forces*. Let $d\mathcal{A}$ represent a *differential directed surface area element*. Specifically, let $|d\mathcal{A}|$ be the geometric area of the surface it represents; let $d\mathcal{A}$ be perpendicular to that surface; and let one side of that surface be distinguished by requiring that $d\mathcal{A}$ point *away from* that distinguished side . Let $d\mathcal{f}_s$ be *the differential contact force exerted on* $d\mathcal{A}$ *by its surroundings*. Then, in the previous notation,

$$\mathcal{t}_{\partial V} = \frac{d\mathcal{f}_s}{|d\mathcal{A}|} \quad .$$

In general $d\mathcal{f}_s$ depends upon both the magnitude and the direction of $d\mathcal{A}$. Let the function T relate $d\mathcal{A}$ to $d\mathcal{f}_s$ at any given $\mathcal{x}$ and t, *i.e.*

$$d\mathcal{f}_s = T(d\mathcal{A}) = T(d\mathcal{A};\mathcal{x},t) \quad . \tag{3.2}$$

The laws listed as (ii)-(iv) above may be written, respectively, as follows:

$$\frac{d}{dt}\left(\iiint_{B} \rho \, dV\right) = 0 \quad , \tag{3.3}$$

$$\frac{d}{dt}\left(\iiint_{B} \rho \, \dot{\boldsymbol{x}} \, dV\right) = \iint_{\partial B} T(d\boldsymbol{A}) + \iiint_{B} \boldsymbol{b} \, dV \quad , \tag{3.4}$$

$$\frac{d}{dt}\left(\iiint_{B} \boldsymbol{r}\times(\rho\dot{\boldsymbol{x}})dV\right) = \iint_{\partial B} \boldsymbol{r}\times[T(d\boldsymbol{A})] + \iiint_{B} \boldsymbol{r}\times\boldsymbol{b} \, dV \quad , \tag{3.5}$$

in which

(i) ρ, $\dot{\boldsymbol{x}}$, and $\boldsymbol{b}$ are the local instantaneous values of the mass density, material velocity, and body force per unit volume exerted on the material by the surroundings

(ii) $\boldsymbol{r}$ is the *lever arm vector* whose head is situated at a material point and whose tail is at the *fulcrum* relative to which moments are taken.

(iii) The last of the three equations is restricted to the special case in which all torques exerted on B by the surroundings are moments of forces.

Two operations *commute* of they can be applied in either order with the same effect. Let $f(\ )$ be a function that maps vectors to vectors. Then the action that it exerts on its operand defines an operation. Now $f(\ )$ is called a *second rank tensor* (or *tensor*, for short) if it commutes with two other operations, namely (i) multiplication by a scalar and (ii) vector addition. Thus, $f(\ )$ is a tensor if and only if $f(\alpha\boldsymbol{x}) = \alpha f(\boldsymbol{x})$ and $f(\boldsymbol{x} + \boldsymbol{y}) = f(\boldsymbol{x}) + f(\boldsymbol{y})$ for all combinations of vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ and all scalars α. Given a tensor f, one can define a related tensor from it, denoted f^T and called the *transpose* of f such that the following identity is satisfied for all combinations of vectors $\boldsymbol{x}$ and $\boldsymbol{y}$:

$$\boldsymbol{x}\cdot[f(\boldsymbol{y})] = [f^T(\boldsymbol{x})]\cdot\boldsymbol{y} \quad . \tag{3.6}$$

A tensor f is called *symmetric* if $f^T = f$ and is called *skew* if $f^T = -f$.

In the 1820's Cauchy proved two fundamental theorems based upon the foregoing concepts which have become pillars of the rational mechanics of materials. They are, first, *Cauchy's first stress theorem*, which asserts that the function T in (3.2) is a tensor, and *Cauchy's second stress theorem*, which asserts that T is symmetric. Cauchy's second stress theorem is derived from (3.5) and replaces it in all subsequent analysis.

3.2.3 CONSTITUTIVE EQUATIONS. A *dynamical process* (according to the definition of the foundationist Walter Noll, *cf*. reference 6) is a combination $\{T,\chi_t\}$ of a stress tensor T and a motion χ_t which is compatible with the basic laws (3.3)-(3.5) (or, more accurately, with the equivalent statements of those laws as differential equations). In the absence of any further assumptions or definitions regarding material behavior, equation (3.4) (or its equivalent statement as a differential equation) could be taken as a formula for computing the body force density $\boldsymbol{b}$ for any given combination of present stress T and body motion χ_t. Note that the body motion χ_t describes, among other things the full history of the deformation of the body up to the present time. In this (artificial) sense, all combinations $\{T,\chi_t\}$ could be regarded as possible dynamic processes. If, however, $\boldsymbol{b}$ is imposed, say, by supposing that $\boldsymbol{b}$ equals the local gravitational force per unit volume $\rho\boldsymbol{g}$, then there must be other conditions specific to the choice of material of which the body is composed which restrict which combinations of $\{T,\chi_t\}$ can be dynamic processes. Such a restriction is called a *constitutive equation.*

There is a broad class of constitutive equations, each of which serves as a definition of a class of ideal materials, but the class of such constitutive equations is not arbitrary. There are several general principles that all constitutive equations must satisfy to be realistic. Though roots of these general principles may be found in the works of the savants of the previous centuries (like Cauchy), the modern statements of these principles is the work of the foundationists since World War Two (primarily Walter Noll) and it constitutes their primary contribution to our subject. These restrictions on constitutive equations (called axioms) will be presented in section 3.2.5 below after some preliminary remarks are made concerning changes in reference frames.

3.2.4 CHANGES IN REFERENCE FRAME. The representations of objects in mechanics like time, geometric position, direction, *etc.* are often specific to a choice of reference frame. Thus, if two observers simultaneously record the time at which a given event takes place and if the observers record that time by the unequal numbers t and t^*, this inequality need not imply that either observer is wrong in an absolute sense. The two observers may simply be reckoning time relative to different choices of *time origin.* A similar statement can be made about two different representations, say $\boldsymbol{x}$ and $\boldsymbol{x}^*$ that these observers assign to the geometric position of a given point in space. Inequality between $\boldsymbol{x}$ and $\boldsymbol{x}^*$ may simply reflect different choices of *place origin.* Finally, directional notions, such as *up, down, left, right, forward,* and *backward* are relative to a particular choice of *reference orientation.* Thus, if two observers stand beneath a hot air balloon and are asked to describe the direction of its velocity one minute after launch, one (who is facing north) might say that that velocity is 'upward and to the right'; the other (who is facing southward) might describe it as 'upward and to the left'. Both may be correct if due account is taken of the difference between their respective reference orientations.

Fortunately, the list of two origins and one reference orientation upon which representations of mechanical objects may depend is exhaustive. As pointed out by Truesdell, one may liken a reference frame to a combination of a rigid body and a clock. Reference frames may differ in the manner that a rigid body may be displaced (affecting its place origin and its reference orientation) and its clock may be reset (affecting its time origin), but they may not differ otherwise.

A tensor is called *orthogonal* (and is normally denoted by the symbol Q) if for all combinations of vectors $\mathbf{a}$ and $\mathbf{b}$, the following identity holds:

$$Q(\mathbf{a})\cdot Q(\mathbf{b}) = \mathbf{a}\cdot\mathbf{b} \quad . \tag{3.7}$$

One consequence of (3.7) follows from the special case $\mathbf{b} = \mathbf{a}$, *i.e.* $Q(\mathbf{a})\cdot Q(\mathbf{a}) = \mathbf{a}\cdot\mathbf{a}$. But the dot product of a vector with itself is just the square of its magnitude (distinguished from the original vector by the magnitude marks | |), from which it follows that

$$|Q(\mathbf{a})| = |\mathbf{a}|$$

for all choices of the vector $\mathbf{a}$ whever Q is an orthogonal tensor. In plainer language *the vector that serves as input to an orthogonal tensor and the vector that serves as output from it are vectors of the equal magnitude but (possible) different direction.*

If, for example, $\overrightarrow{XY}$ represents a vector drawn from a material point X to a material point Y, in a rigid body, then the distance between X and Y must not change as the body is rotated. If $\mathbf{r}$ and $\mathbf{r}^*$ denote the representations $\overrightarrow{XY}$ before and after such a rotation, respectively, then the relation

$$\mathbf{r}^* = Q(\mathbf{r})$$

implies that $|\mathbf{r}^*| = |\mathbf{r}|$ (which is consistent with the condition that the body be rigid) but is general in all other respects. In this way, an orthogonal tensor Q may represent a rigid rotation.

Truesdell's comparison of a reference frame to the combination of a rigid body and a clock relates to the feature of orthogonal tensors just described. Suppose, for example, that two observers adopt different representations, say $\mathbf{f}$ and $\mathbf{f}^*$ of the same force (say the force exerted by an eraser on a desk as it rests under the action of gravity). The representations $\mathbf{f}$ and $\mathbf{f}^*$ may differ owing to unequal reference frames adopted by the observers, but the magnitudes $|\mathbf{f}|$ and $|\mathbf{f}^*|$ must be equal. Such a relationship is captured by an equation of the form

$$\mathbf{f}^* = Q(\mathbf{f}) \quad , \tag{3.8}$$

in which the orthogonal tensor represents *the change in the reference orientation* between the two frames.

Now the term *force* is a primative notion in mechanics, *i.e.* one can not define it in terms of more primative notions without resorting to circular reasoning. Be that as it may, one may list properties always exhibited by anything one would call a force. Such a property may be called a *postulate* or *axiom*. One may note in passing that the foundationists have adopted (3.8) as on of their *axioms of forces*.

Let x and x^* be two representations of the same geometric point, each referred to a different reference frame. Representations of points, like representations of vectors, may differ under a change of frame owing to a difference between the reference orientations of those two frames. Unlike representations of vectors, however, representations of points may also differ owing to a difference of *place origin* of those two frames (vectors, remember, are 'free'—they are not specific to the choice of place origin). Following reasoning of this kind, the foundationists argue that the most general transformation between x and x^* under a change of frame may be expressed in the form

$$x^* = c + Q(x - q) \quad , \tag{3.9}$$

in which q and c represent the (possibly unequal) place origins in the 'unstarred' and 'starred' frames, respectively. The corresponding general transformation between the representations t and t^* of time in the two frames is

$$t^* = t - a \quad , \tag{3.10}$$

in which a is a constant.

The laws for transforming representations of velocity and acceleration under a change of frame may be derived by differentiating (3.9) with respect to time, taking due account of the defining features of orthogonal tensors and the possibilities that c, q, and Q may themselves be time dependent.

3.2.5 AXIOMS GOVERNING CONSTITUTIVE EQUATIONS. From the foregoing discussion, we may conclude that a change of reference frame $Fr \rightarrow Fr^*$ from an unstarred frame Fr to a starred frame Fr^* induces a change in the respective representations of many mechanical objects (*e.g.* time, force, velocity, *etc.*). The laws for converting these representations are, however, fixed once Q (the change in reference orientation) and the changes in place and time origin are specified.

Enough has now been said to allow us to state three axioms (due to Noll) which must be satisfied by all valid constitutive equations for materials. The first of these is the *axiom of determinism*:

> A1 The (present) stress (at a point) in a body is determined by the history of the motion of the body.

The second is *the axiom of local action*:

> A2 In determining the stress at a given material point X in a body, the motion outside an arbitrarily small neighborhood of that point may be disregarded.

Let $\psi(t)$ be any function of time. We introduce the shorthand notation

$$\psi^{(t)}(s) \equiv \psi(t - s) \quad , \quad s \geq 0 \quad . \tag{3.11}$$

Thus, $\psi^{(t)}(s)$ may be called *the history up to time t of the function* $\psi(t)$ and the parameter s may be regarded as the *time lapse*. If, as stated in equation (3.1) above, χ_t defines the placement of a body at time t, then we may write its *history of placements* up to time t as

$$\chi^{(t)}(s,X) \equiv \chi_{t-s}(X) \quad , \tag{3.12}$$

in which the function in the right member is the one introduced in (3.1a).

With this notation, the axiom of determinism may be expressed mathematically, *i.e.* in every kinematically possible process, the stress T at time t is related to the history of the motion $\chi^{(t)}(s,X)$ by an equation of the form

$$T(t) = \overset{\infty}{\underset{s=0}{\mathfrak{F}}} [\chi^{(t)}(s,X)] \quad . \tag{3.13}$$

The operator $\overset{\infty}{\underset{s=0}{\mathfrak{F}}}$ is called a *constitutive functional*. It relates the semi-infinite family of placements (each of which corresponds to a different value of the time lapse s) to the value of the stress T at time t and does so for every material point X in the body B.

Now the constitutive functional $\overset{\infty}{\underset{s=0}{\mathfrak{F}}}$ describes properties intrinsic to the material of which the body B is composed, and, indeed, all such intrinsic properties are included in it. Thus, for example, any attempt to give a mathematical criterion which distinguishes the fluid from the solid state is ultimately a way of partitioning the various kinds of constitutive functionals that apply to different groups of materials. Similar statements could be made about the difference between *elastic* and *inelastic* materials and between materials with memory and materials without it.

Once one accepts that all of the intrinsic features of a material are tied up in its constitutive functional, one is ready to accept *the axiom of material objectivity*:

A3 If $T \rightarrow T^*$ and $\chi^{(t)} \rightarrow (\chi^{(t)})^*$ are the changes in the representations of the present stress and the motion history attendent to a change $Fr \rightarrow Fr^*$ of reference frame, then a constitutive functional $\overset{\infty}{\underset{s=0}{\mathfrak{F}}}$ that correctly relates variables referred to the unstarred frame must hold equally well (without any change in $\overset{\infty}{\underset{s=0}{\mathfrak{F}}}$) when the variables are referred to the starred frame.

Thus, according to this axiom,

$$T = \Im_{s=0}^{\infty}\left(\chi^{(t)}(s,X)\right)$$

if and only if

$$T^{*} = \Im_{s=0}^{\infty}\left([\chi^{(t)}(s,X)]^{*}\right) \quad . \tag{3.14}$$

In words, axiom A3 states that the material is indifferent to the choice of reference frame a given analysis may make in trying to decribe its properties.

3.2.6 SIMPLE MATERIALS. At this point, the foundationists stop writing equations that hold for all materials and begin to define specific classes of materials. Of these, the first is the *simple material*. As a prerequisite to its discussion, we must introduce a new concept.

Let κ be function that relates material identifiers X in a body to geometric points, *i.e.*

$$\mathcal{X} = \kappa(X) \quad , \quad X = \kappa^{-1}(\mathcal{X}) \quad . \tag{3.15}$$

One may call κ a *reference placement* of a body. It may happen that a body actually assumes its reference placement at some reference time t_1, in which case, $\kappa(X) = \chi(X,t_1)$. For the moment, however, we do not introduce this assumption. Now

$$\mathcal{X} = \chi(X,t) = \chi[\kappa^{-1}(\mathcal{X}),t] \equiv \chi_{\kappa}(\mathcal{X},t) \quad , \tag{3.16}$$

in which χ_{κ} serves much the same purpose as χ, except that now a material point X is identified by its position $\mathcal{X}$ in the reference placement κ rather than by the material point identifier X .

Let $\mathcal{X}_0$ be a fixed point in the domain of χ_{κ}. A motion is called *homogeneous* (relative to the reference placement κ) if there exists a tensor $F_{\kappa}(t)[\ ]$ such

$$\chi_{\kappa}(\mathcal{X},t) = \chi_{\kappa}(\mathcal{X}_0,t) + F_{\kappa}(t)[\mathcal{X} - \mathcal{X}_0] \quad . \tag{3.17}$$

The above equation is a generalization of the equation of a straight line in the x-y plane, *viz.* $y(x) = y(x_0) + m(x - x_0)$. The constant m in the straight-line formula corresponds to the tensor $F_{\kappa}(t)[\ ]$ (which is a *linear* operator) in (3.17). In a homogeneous motion (as the foundationists point out) a chain of material points that lie on a straight line when the body is in its reference configuration remain on a straight line in all later configurations.

A material is called *simple* if the distinction between homogeneous and nonhomogeneous motion does not affect its constitutive equation. Now the operator

$F_\kappa(t)[\ ]$ in (3.17) is a generalization of the idea of a first derivative. It represents the first derivative of the function that maps X to x. In different terms, then, the constitutive equation of a simple material involves only the *first* derivative of the general motion function χ_κ.

If, in accordance with the rule introduced in (3.12), we denote the history of the gradient $F_\kappa(t)[\ ]$ up to time t by $F_\kappa^{(t)}(s)[\ ]$, we have

$$F_\kappa^{(t)}(s)[\ ] \equiv F_\kappa(t - s)[\ ] \quad . \tag{3.18}$$

Again, s is the time lapse. The constitutive relation (3.18) then becomes, for a simple material,

$$T(t) = \mathfrak{G}_{s=0}^{\infty} [F_\kappa^{(t)}(s)] \quad . \tag{3.19}$$

Now simple materials automatically satisfy the principles of determinism and of local action. Further reductions must be undertaken, however, to ensure that (3.19) satisfies the principle of material objectivity.

3.2.7 THE FLUID-SOLID DISTINCTION. The idea of the simple material has led the foundationists to many profound consequences. Limitations of space and scope (as well, of course, as limitations of my own understanding) prevent me from discussing more than two. These are (i) a sharp distunction between the fluid and solid states; and (ii) a framework for the general discussion of stress relaxation and other aspects of fading memory. The distinction between fluid and solid states seems to challenge an intutive notion implicit in the use of the term *viscoelasticity*, namely that a blurring of the fluid-solid distinction is to be expected when a material exhibits memory. Indeed, the foundationist Walter Noll titled his 1955 thesis 'On the continuity of solid and fluid states.' Three years later, however, he proposed a reorganization of mechanics (reference 6), which includes the fluid-solid distinction just referred to. This later version has permeated all subsequent thinking by the foundationists.

One may call the tensor $F_\kappa^{(t)}$ defined in (3.17)-(3.19) the *deformation history tensor*. It is specific to the choice of the reference placement κ [introduced in (3.15)]. There is a chain rule identity that relates the deformation history tensors $F_{\kappa_1}^{(t)}$ and $F_{\kappa_2}^{(t)}$ associated with two different reference reference placements κ_1 and κ_2, namely

$$F_{\kappa_1}^{(t)} = F_{\kappa_2}^{(t)} P \quad , \tag{3.20}$$

in which P is a tensor, namely the gradient of the function that maps position in κ_1 to position in κ_2. To appreciate what P does, one may imagine an ideal experiment for the determination of the constitutive functional in (3.19). Thus, one begins by preparing infinitely many identical test specimens, each of which is a body composed of the material whose behavior is to be captured by (3.19). Each specimen is subject to a different deformation history which begins with a common reference placement κ_1. At the end of each history, the stress tensor T of a material point X is found and an ordered pair is formed consisting of the deformation history of that particle and its final stress. The set of all such ordered pairs (which corresponds to the set of all such deformation histories) defined *the constitutive functional* $\overset{\infty}{\underset{s=0}{\mathfrak{G}}}$ in (3.19). If the whole sequence of hypothetical experiments were repeated but with a different choice of initial placement, say κ_2, then one could not assume, nor could one expect, that the resulting constitutive functional $\overset{\infty}{\underset{s=0}{\mathfrak{G}}}$ would be the same as the one that corresponds to the initial placement κ_1. There may, however, be special choices of P in (3.20) which represent changes from the reference placement that have no effect whatsoever on the stress. For them and only for them

$$\overset{\infty}{\underset{s=0}{\mathfrak{G}}}\left(F_{\kappa_1}{}^{(t)}\right) = \overset{\infty}{\underset{s=0}{\mathfrak{G}}}\left(F_{\kappa_2}{}^{(t)}\right) ,$$

or, equivalently,

$$\overset{\infty}{\underset{s=0}{\mathfrak{G}}}\left(F_{\kappa_2}{}^{(t)}H_{\kappa_1}\right) = \overset{\infty}{\underset{s=0}{\mathfrak{G}}}\left(F_{\kappa_2}{}^{(t)}\right) \cdot \qquad (3.21)$$

[by (3.20)]. A tensor H_{κ_1} that satisfies (3.21) is called a *material isomorphism* and the set of all such material isomorphisms that begin with the reference placement κ_1 is called the *peer group* (or *isotropy group*) belonging to κ_1. A deformation of material from one reference configuration to another is called *unimodular* if it does not involve any change in volume. One expects that a change in volume of a body is attended by a change in its internal stresses (especially its internal pressure) regardless of the material of which it is composed. Thus, no peer group contains tensors H_{κ_1} that are not unimodular.

Now the class of unimodular deformations may be divided into two subclasses, namely *orthogonal* and *nonorthogonal*. The orthogonal deformations are just those for which H_{κ_1} is an orthogonal tensor [the definition of an orthogonal tensor was given above in the paragraph containing (3.7)] . The orthogonal tensors preserve distances between distinct material points; the nonorthogonal ones do not. Thus, the orthogonal tensors consist of rigid rotations, reflections through a plane, or some combination of the two. The nonorthogonal tensors necessarily involve elongation or contraction of material lines in a body.

If a small sample of solid material is free of residual stresses (however defined), then any effort to subject it to a nonorthogonal unimodular deformation results in a net alteration of the stresses in it. Note that a material specimen that *has* residual stresses in it can be subject to a nonorthogonal unimodular deformation that first unloads then reloads the speciman without producing any *net* alteration of the stress. If, therefore, one is to characterize the solid state by a change in stress that *necessarily* attends nonorthogonal unimodular deformations, one must include the premise that such deformations begins from a *preferred* configuration (one that lacks residual stresses).

The foundationists thus propose the following definition of a simple solid: *A simple material is called a* simple solid *if there exists a reference placement (called an* undistorted placement *) such that only orthogonal tensors can be members of the peer group associated with that placement.*

Fluids, by contrast, have no such preferred configuration. the foundationists definition of a simple fluid amounts to the statement that *a simple material is a* simple fluid *if the peer groups belonging to any two reference placements are the same and are equivalent to the set of all unimodular tensors.* Simple fluids may or may not have memory. Indeed, they may have memory of the entire deformation history $F^{(t)}(s)$ that leads to the determination of the present stress. What they forget is the *initial* configuration κ_1 from which that deformation history $F^{(t)}(s)$, however long, commenced (provided, again, that the deformation history does not change the material volume —fluids *do* remember overall changes in volume).

3.2.8 FADING MEMORY. The stress tensor at a given material point depends, in general, upon the whole history of deformation of the material in its neighborhood. One may, however, decompose the stress at a point into an *equilibrium* part and a *transient* part. As the name suggests, the equilibrium part is the stress that would exist if the material had been in its present configuration forever. This decomposition makes sense only if one has some confidence that the transients die out in time. In somewhat more precise terms, the memory that the material retains of changes of configuration in the near past, as reflected in the present value of the transient part of the stress, is more significant than is the material's memory of changes in the distant past. At the same time, the equilibrium term may reflect changes in configuration that occurred prior to an arbitrarily long time lapse.

The foundationists (*cf.* reference 1, §§38-41) identify the principle of fading memory with an assumption that the constitutive functional for a simple material is a continuous functional of the deformation history in the neighborhood of the rest history (which corresponds to the history of a body that has been at rest forever).

Now the constitutive functional $\underset{s=0}{\overset{\infty}{\mathfrak{G}}}$ for a simple material does not operate on numbers or even on tensors. Rather, it operates upon *histories of* *tensors.* The mathematical idea of *continuity* must therefore be expressed in a rather abstract form. Fortunately, such a form is available, although it is technical. Readers interested in the details are referred to reference 1, pp 103-104.

Now continuity of a function imposes a weaker condition upon it than differentiability and differentiability to the order n imposes a weaker condition than differentiability to the order $n+1$. Thus, the foundationists define a *weak* principle and a sequence of *strong* principles of fading memory which impose the conditions of continuity and nth order differentiability, respectively, upon the constitutive functional. If the latter applies, then the constitutive functional may be expanded into a (suitably generalized) Taylor series expansion about the rest history. For a material with fading memory, changes in configuration that occurred in the near past (characterized by a small time lapse s) exert a greater influence upon the transient stress than do changes in configuration in the distant past. This focus on the case of small s leads to a Taylor series expansion of the deformation history $F^{(t)}(s)$ about $s = 0$. By applying these two Taylor series simultaneously (one of which expands the constitutive functional in a neighborhood of the rest history and the other of which expands the deformation history in the neighborhood of zero time lapse), one obtains a series representation for the transient stress whose terms involve only the present values and the present values of the time derivatives of the deformation history or expressions (such as the rate of strain) derived from it.

The expansion precedure just described is due to Bernard Coleman and Walter Noll (references 7 and 8). Now an expansion is valid in the mathematical sense only if one can estimate the error associated with the approximation that results if the expansion is stopped after n terms. Coleman and Noll keep the same number of terms in the Taylor series for the expansions of the constitutive functional and the deformation history. They show that the error left in stopping the expansion after n terms is proportional to the nth power of α, in which α is a real number in the range $0 \leq \alpha \leq 1$ and is called *the* α*-retardation* of the deformation history. The α-retardation of a given history of deformation $F^{(t)}(s)$ is denoted $F_{\alpha}^{(t)}(s)$ and is defined by

$$F_{\alpha}^{(t)}(s) \equiv F^{(t)}(\alpha t) \quad . \tag{3.22}$$

If one regards a given deformation history as analogous to a videotape played at a standard speed, then the modified deformation history that results when the given deformation history is played back at a speed slower by a factor α will be analogous to the α-retardation of the given history. The Coleman-Noll expansion represents an approximation within a family of retarded histories. The accuracy of the n-term expansion is greater the more retarded is its history (*i.e.* the smaller is the α). In the limit of vanishingly small α, the stress reduces to its equibrium value.

A material is *elastic* if its present stress T is determined by its present deformation $F^{(t)}(0)$ (*i.e.* the value of $F^{(t)}(s)$ with zero time lapse s). A deformation involves infinitesimal strain if $F^{(t)}(0)$ differs infintesimally from that of a rigid deformation. *Hooke's law of linear elasticity* (which, of course forms the starting point of most books on elasticity and most computer programs for stress analysis of three dimensional structures) results from restricting the equations for an elastic material to the case of infinitesimal strains in the Coleman-Noll expansion.

Returning to the case of finite strain, a feature of the Coleman-Noll theory is that the limit $\alpha \to 0$ (of infinitely slow deformations) corresponds to the stress distribution of an elastic material. If this elastic material is a fluid, then its stress distribution is that of a *frictionless compressible fluid.* If one applies the definition of a fluid to the case of slow flow and keeps the the first term in the Coleman-Noll expansion that involves a derivative with respect to the time lapse s of the deformation function $F^{(t)}(s)$ (and ignores the remaining terms, the net contribution from which vanishes with the second power of α), one obtains the constitutive relation for a *linearly viscous fluid.* This latter approximation is the one upon which the well-known Navier-Stokes equations of fluid mechanics are based (reference 12).

Any constitutive equation for a solid that retains at least one term for the rates [*i.e.* an s- derivative of $F^{(t)}(s)$] is an equation of *viscoelasticity.* Coleman and Noll use the term *finite linear viscoelasticity* to describe the special case of the constitutive equation of a simple material that results when the constitutve functional depends linearly upon the deformation history function. In this case, the deformation rates are not necessarily slow. This idealized constitutive equation may be expressed in integral form. Many books on viscoelasticity begin by assuming its validity (or the vaildity of a one dimensional version of it); see, for example, *Viscoelasticity of Polymers* by Ferry (reference 9, p 8).

Considering the abstractness of the Coleman-Noll theory, one is hardly surprised that few experimentalists learn enough about it to subject it to a fair test. An experimentalist named H. Markowitz has aquired the necessary knowledge and obtained quality experimental data. His data suggest that the theory of a simple fluid [one which includes all of the dependencies upon history implicit in (3.19) above] is the only theory general enough to permit quantitative agreement with experiments on solutions of polymer plastics in solvents [*cf.* the monograph *Viscometric Flows of Non-Newtonian Fluids* by Coleman, Markowitz, and Noll, reference 10, p2]. If a cryogenic seal made of virgin TFE is subject to finite amplitude strains (*e.g.* extrusion) during installation, one would be optimistic indeed to suppose that the time dependent stresses within it could be described with quantitative accuracy by any theory less general than the Coleman-Noll theory.

Unfortunately, as inclusive as the Coleman-Noll theory is (as of the time reference 1 was written), it is not all-inclusive. In a paper titled 'A New Mathematical Theory of Simple Materials' (reference 11), published in 1972, Noll listed what he called 'at least three severe defects' in his first theory (reference 6, 1958) of simple materials. One defect concerns the dependence of the present stress upon the infinite past (which is not knowable) or upon the recent past (which is not a valid assumption for all materials). A second defect is the failure of the original theory 'to give an adequate conceptual framework for the mathematical description of such phenomena as *plasticity*, yield, *and hysteresis*.' The third defect concerns Noll's original definition of *materials of the rate type* (those whose time dependent stresses satisfy a differential equation with respect to time). I have not had the opportunity yet to digest Noll's new theory of simple materials. Suffice it to say, however, that the earlier theory's lack of an adequate conceptual framework for plasticity, yield, and hystereses does not mean these effects are incompatible with it. Rather it means they are not identified explicitly when they occur.

3.3 FLOW THROUGH SMALL HOLES

So far, most of the discussion in this report has been devoted to the mechanics of seal materials. Since a valve or coupling controls the flow of *fluids*, any attempt to predict the rate of transport of fluid volume through a hole resulting from some imperfection or failure of a seal assembly must involve the mechanics of fluids.

An internal document titled 'Metrological Failure Analysis Investigation of the Liquid Hydrogen (LH_2) Leak at Complex 39B' MAB-128-88, dated September 26, 1988 contains photomicrographs of a TFE seal surface after it was removed from an assembly that had leaked. One such photomicrograph is reproduced as figure 6. The largest scratch shown on the figure appears to exhibit a semicircular cross section whose size and shape does not vary appreciably along the midline of the scratch. Such a scratch geometry suggests a fluid dynamics problem. Consider the channel formed between a scratched plastic sealing surface and a flat polished metal plate against which it is pressed. Suppose that this channel extends ultimately from the inside of the pipe to the outside. Let the size and shape of the cross section of the scratch be fixed and let s represent the total arc-length of its midline (defined, for definiteness, as the locus of centroids of the cross sections). Let the scratch be straight or only slightly curved, *i.e.* the smallest radius of curvature of its midline is large compared to its largest cross sectional diameter D. Let the total pressure drop from the upstream end to the downstream end be a given constant ΔP. Let the working fluid satisfy the Navier-Stokes equations and have the mass density ρ and shear viscosity μ. The problem, then, is to find the rate of transport $\dot{V}$ of fluid volume.

The problem of determining the distribution of streamwise velocity across the cross section may be solved mathematically in reasonably simple terms

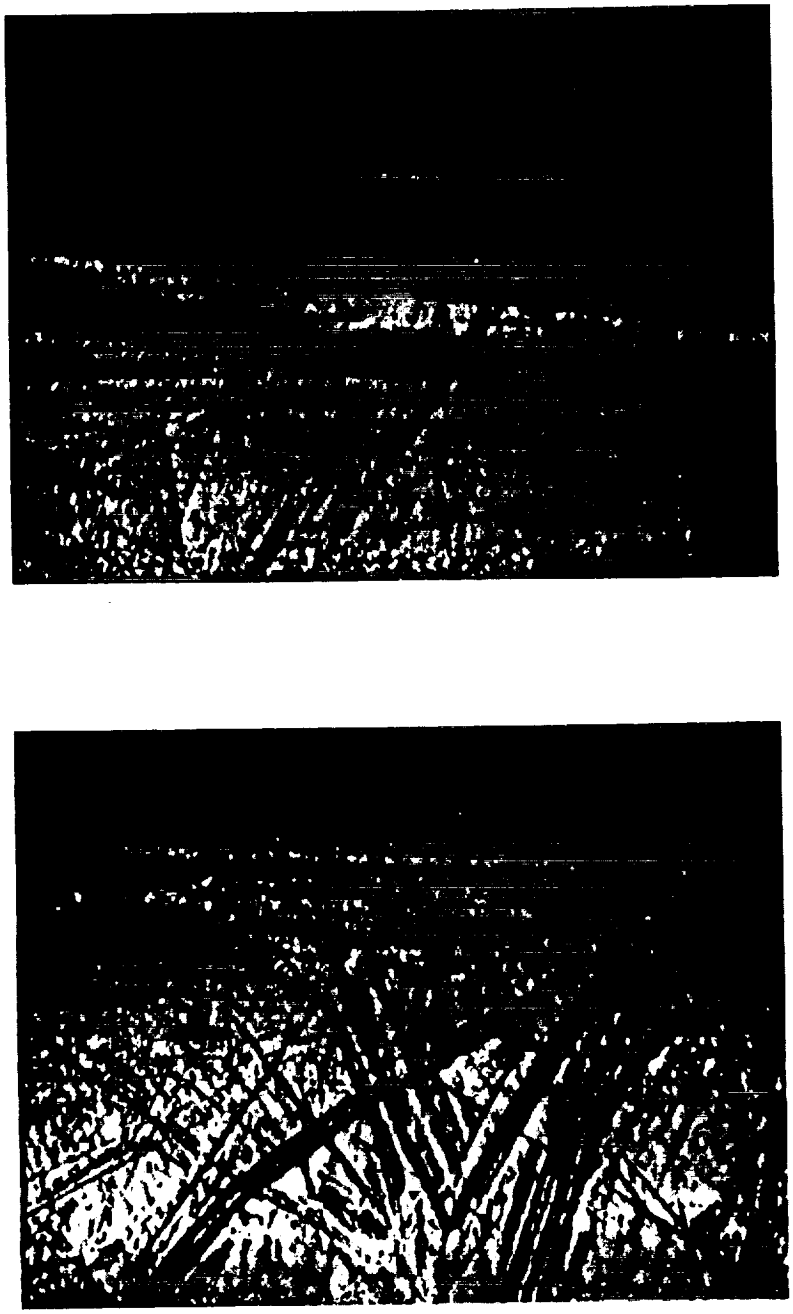


FIGURE 6. Scratches on the surface of a TFE seal (magnification 82X). From 'Metrological Failure Analysis Investigation of the Liquid Hydrogen (LH_2) Leak at Complex 39B', NASA/KSC Malfunction Analysis Branch Report MAB-128-88, September 26, 1988, p31.

if the components of the fluid velocity parallel to the cross sectional planes are uniformly equal to zero and the flow is steady. Under the assumption that all cross sections of the channel are the same, it then follows from the law of conservation of mass (for constant density flow) that the streamwise component u of the fluid velocity for a given point on the cross section is the same for all cross sections.

Let (x,y,z) represent local cartesian coordinates aligned so that the positive x-axis points in the direction of fluid flow. Let (g_x,g_y,g_z) be the three components of the vector $\boldsymbol{g}$ that represents the local gravitational force per unit mass. Since the scratch can be aligned in any direction, so also can the axes of the coordinates (x,y,z). Since no single coordinate axis in the system (x,y,z) necessarily points 'down', the three components of $\boldsymbol{g}$ relative to it may all be nontrivial. The three components of the velocity vector $\boldsymbol{v}$ relative to these axes are $(u,0,0)$ and u depends only on y and z. The three components of the equations for the rate of change of translational momentum reduce under these assumptions to

$$0 = -\frac{\partial p}{\partial x} + \mu\left(\frac{\partial^2 u}{\partial y^2} \quad \frac{\partial^2 u}{\partial z^2}\right) + \rho g_x$$

$$0 = -\frac{\partial p}{\partial y} + \rho g_y$$

$$0 = -\frac{\partial p}{\partial z} + \rho g_z$$

Taking ρ to be a constant and introducing the shorthand

$$p_{ex} \equiv p - \rho(g_x x + g_y y + g_z z) \quad ,$$

the above system becomes

$$0 = -\frac{\partial}{\partial x}(p_{ex}) + \mu\left(\frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}\right)$$

$$0 = -\frac{\partial}{\partial y}(p_{ex})$$

$$0 = -\frac{\partial}{\partial z}(p_{ex}) \quad .$$

The function p_{ex} (the *excess pressure*) represents the excess of the actual pressure at a point over what the pressure would be if the fluid were in static equilbrium under its own weight. From the assumption of steadiness and the last two equations, we conclude that p_{ex} depends only on x and that the first equation reduces to

$$\frac{d}{dx}(p_{ex}) = \mu\left(\frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}\right) \quad .$$

For reasons just given, the left member is independent of y, z, and t. Since, however, u is independent of x, the right member has this property. Thus, the common quantity to which both members of the above equation are equal is independent of x, y, z, and t, *i.e.* it must be a constant K. Thus, we may write

$$\frac{d}{dx}(p_{ex}) = K \tag{3.23}$$

$$\mu\left(\frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}\right) = K \quad . \tag{3.24}$$

If we assume that no slip occurs between the fluid and the channel wall that enclosed it, we arrive at the boundary condition

u vanishes on the boundary (3.25)

to which solutions of the partial differential equation (3.24) are subject.

The best known special case of the problem defined by (3.24) and (3.25) is the case of flow through a smooth-wall circular pipe, first solved by George Gabriel Stokes in 1845 (reference 12). Since the typical cross sectional shape of a scratch is *not* a circle (at least not a *full* circle), this well-known solution is useful, at best, only for purposes of comparison with results for more realistic cross sections. As it happens, not much effort is needed to derive the solutions of (3.24) and (3.25) for channels with the cross sectional shape of a semicircle or an equilateral triangle. The mathematical problem defined by (3.24) and (3.25) also arises in the calculation of the distribution of shear stress over the cross section of a prismatic bar subjected to small twist (according to the linear theory of elasticity). The solution for the bar of triangular cross section was found by Adhémar Saint Venant in the last century and is at our disposal for use in the present case.

Consider the triangle shown in figure 7. Let the three sides be numbered 1, 2, and 3 as shown. If O is the centroid of the triangle, then the perpendicular distances from O to sides 1, 2, and 3 are the same. Let this common distance be denoted ℓ. One may also think of ℓ as the radius of the largest circle that can be inscribed in the triangle. Let $\mathbf{n}_1$, $\mathbf{n}_2$, and $\mathbf{n}_3$ be outward unit normal vectors belonging to the three sides as shown. Let $\mathbf{y}$ be the position vector of an arbitrary point P in the plane of the triangle relative to its center. Let ℓ_n be the perpendicular distance from P to side n. Then

$$\ell_n = \ell - \mathbf{y}\cdot\mathbf{n}_n \,, \quad n \in \{1,2,3\} \quad . \tag{3.26}$$

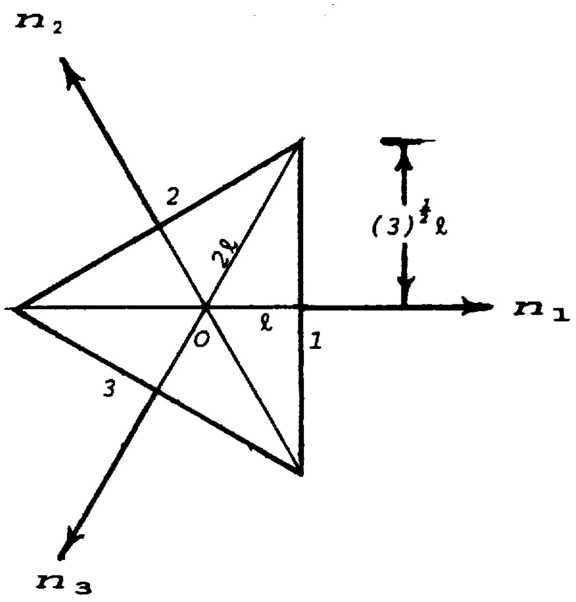


FIGURE 7. Nomenclature for the analysis of the flow through a channel with a triangular cross section.

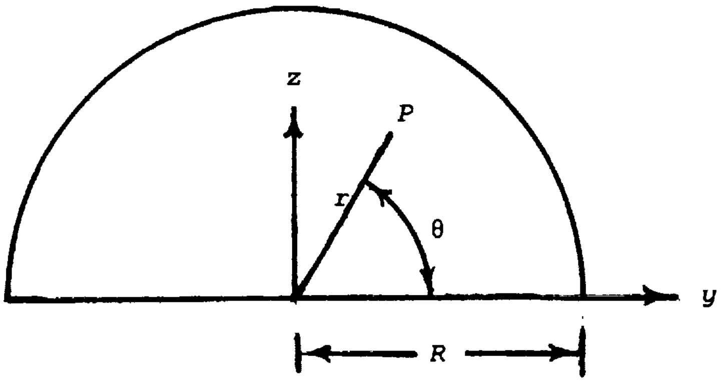


FIGURE 8. Nomenclature for the analysis of the flow through a channel with a semicircular cross section.

Let θ be the angle from $\boldsymbol{n}$ to $\boldsymbol{y}$ reckoned positive counterclockwise. Let $r \equiv |\boldsymbol{y}|$. Then

$$\ell_1 = \ell - r\,cos(\theta) \tag{3.27a}$$

$$\ell_2 = \ell - r\,cos(\theta + 2\pi/3) \tag{3.27b}$$

$$\ell_3 = \ell - r\,cos(\theta - 2\pi/3) \quad . \tag{3.27c}$$

One may relate the polar coordiantes (r,θ) to a set of cartesian coordinates (y,z) in the plane of the cross section by the equations

$$y = r\,cos(\theta) \qquad , \qquad z = r\,sin(\theta) \quad . \tag{3.28}$$

If P is a point on side n of the triangle, then ℓ_n is zero there. It follows that the product

$$\ell_1\ell_2\ell_3 \quad ,$$

taken as a general function of position on the cross section, vanishes on all three sides, a condition also satisfied by the function u in (3.24) and (3.25). It is a pleasant surprise to discover that the value of

$$\left(\frac{\partial}{\partial y^2} + \frac{\partial}{\partial z^2}\right)(\ell_1\ell_2\ell_3)$$

is independent of y and z. It follows that $u \propto \ell_1\ell_2\ell_3$ with an appropriate choice of the factor of proportionality, is a solution of the boundary value problem defined by (3.24) and (3.25). The solution, with the factor of proportionality included, may be written in the symmetric looking form

$$u = -\frac{1}{\mu}\frac{d}{dx}(p_{ex})\left(\frac{\ell_1\ell_2\ell_3}{\ell_1 + \ell_2 + \ell_3}\right) \quad . \tag{3.29}$$

In terms of polar and cartesian coordinates, the formulas for u take less symmetric looking forms, which, nevertheless, are useful. They are

$$u = -\frac{1}{\mu}\frac{d}{dx}(p_{ex})\left(\frac{\ell^2}{3} - \frac{r^2}{4} - \frac{r^3}{12\ell}\,cos(3\theta)\right) \tag{3.30}$$

and

$$u = -\frac{1}{\mu}\frac{d}{dx}(p_{ex})\left[\frac{\ell^2}{3} - \frac{(y^2+z^2)}{4} - \frac{1}{12\ell}\left(4y^3 - 3(y^2+z^2)y\right)\right] \quad . \tag{3.31}$$

Let $\dot{V}$ represent the rate of transport of fluid volume through the channel. Then $\dot{V}$ equals the integral of u across the cross section. To carry out this integral, the cartesian formula (3.30) for u is most convenient. The result is

$$\dot{V} = -\frac{1}{\mu}\frac{d}{dx}(p_{ex})\ell^4\,\frac{9(3)^{\frac{1}{2}}}{20} \quad . \qquad (3.32)$$

The negative signs in (3.29)-(3.31) reflect the fact that $p_{ex}(x)$ *decreases in the direction of increasing* x, *i.e.* the fluid flows from higher to lower excess pressure. Thus,

$$\frac{d}{dx}(p_{ex}) < 0 \; . \qquad (3.33)$$

It follows that the right members of (3.30) to (3.32) are all *positive valued*, as, of course, one would expect. In applying the above results, the following facts may be useful:

(i) The altitude of the triangle equals 3ℓ

(ii) The length of one side of the triangle equals $2(3)^{\frac{1}{2}}\ell$

(iii) The area of the triangle equals $(3)^{3/2}\ell^2$.

The foregoing results apply to laminar flow only. One may define a Reynolds number Re in terms of the fluid velocity u_0 on the centerline and the altitude 3ℓ of the triangle, *i.e.*

$$Re = \frac{\rho u_0(3\ell)}{\mu} \quad .$$

The centerline velocity may be evaluated by setting $r = 0$ in (3.30). The result is

$$u_0 = -\frac{1}{\mu}\frac{d}{dx}(p_{ex})\frac{\ell^2}{3} \quad .$$

Substituting this into our previous formula for the Reynolds number, we have

$$Re = -\frac{\rho\ell^3}{\mu^2}\frac{d}{dx}(p_{ex}) \quad . \qquad (3.34)$$

If Re is of the order of 10^3 or smaller, then one may surmise that the flow is laminar. The flow *may* be laminar if Re is an order of magnitude higher, but one should not commit onesself to a design decision based upon this assumption without knowledge of other relevant parameters, such as the quality of the inlet conditions, the surface roughness, the presence of acoustical noise and vibration, *etc.*

Now the channel formed between a polished metal plate and a scratched gasket pressed up against it may or may not have a triangular cross section. Indeed, the photomicrograph of figure 6 suggests that a cross sectional shape having the figure of a semicircle might be more appropriate in that case. The analysis leading up to the problem (3.24), (3.25) may, however, be applied without modification.

Consider the semicircle shown in figure 8. Let (y,z) be cartesian coordinates centered on the midpoint of the straight wall, as shown. Let (r,θ) be the polar coordinates of a point P in the cross section. Let r and θ be defined in terms of y and z by

$$y = r\cos(\theta) \qquad z = r\sin(\theta) \quad . \tag{3.35}$$

Let the inside radius of the semicircle be R. If one translates the boundary value problem defined by (3.24) and (3.25) into these coordinates, it requires one to solve the partial differential equation

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial u}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2} = \frac{1}{\mu}\frac{d}{dx}(p_{ex}) = constant \tag{3.36}$$

subject to the boundary conditions

$$(u)_{r=R} = 0 \quad , \quad (0 \leq \theta \leq \pi) \tag{3.37}$$

$$(u)_{\theta=0} = (u)_{\theta=\pi} = 0 \; , \; (0 \leq r \leq R \;) \quad . \tag{3.38}$$

One may find a solution of this problem in the form of a Fourier series expansion, *i.e.*

$$u(r,\theta) = -\frac{1}{\mu}\frac{d}{dx}(p_{ex})R^2\left\{\frac{4}{\pi}\sum_{\substack{n=1\\(n=odd)}}^{\infty}\frac{(-1)(r/R)^n\sin(n\theta)}{(n-2)(n)(n+2)} - \left(\frac{r}{R}\right)^2\frac{\sin^2(\theta)}{2}\right\} . \tag{3.39}$$

The rate of transport $\dot{V}$ of fluid volume corresponding to this u is

$$\dot{V} = \int_{r=0}^{R}\int_{\theta=0}^{\pi} u(r,\theta)\, r\, d\theta\, dr$$

or

$$\dot{V} = -\frac{1}{\mu}\frac{d}{dx}(p_{ex})R^4\left\{\left(\frac{8}{\pi}\right)\sum_{\substack{n=1\\(n=odd)}}^{\infty}\left(\frac{-1}{(n-2)(n)^2(n+2)^2}\right) - \frac{\pi}{16}\right\} \quad . \tag{3.40}$$

There is some arbitrariness in writing the efflux rate formula, particularly in the choice of length parameter. Some standardization can be achieved by expressing squares of the length parameter in terms of the cross sectional area. If one rewrites (3.32) and (3.40) with A as the common symbol to denote the cross sectional area, one obtains

$$\dot{V} = -\frac{1}{\mu}\frac{d}{dx}(p_{ex})A^2\, C_e \; , \tag{3.41}$$

in which the coefficient c_e depends upon the cross sectional shape. The following table furnishes the specifics:

Type of cross section	*Value of* C_e	
	exact	*numerical*
triangular	$(3)^{\frac{1}{2}}/60$	$(2.88675\ldots)10^{-2}$
semicircular	$\sum_{\substack{n=1 \\ n=odd}}^{\infty} \frac{-32/\pi^3}{(n-2)(n)^2(n+2)^2} - \frac{1}{4\pi}$	$(3.01488\ldots)10^{-2}$
circular	$\frac{1}{8\pi}$	$(3.97887\ldots)10^{-2}$

The efflux coefficient for the circular pipe is included for comparison. If $(p_{ex})_{in}$ and $(p_{ex})_{out}$ denote the inlet and the outlet values, respectively, of p_{ex}, then, under the assumptions implicit in the derivation of the above results, one may take

$$\frac{d}{dx}(p_{ex}) = [(p_{ex})_{out} - (p_{ex})_{in}]/s < 0 ,$$

in which s is the total arc-length of the channel centerline.

4. RECOMMENDATIONS

I began my work this summer by posing the three questions listed at the start of section 2.1 above, namely (i) what *mission* must a cryogenic seal perform; (ii) what are the contrasts between *desirable* and *available* seal materials; and (iii) *how realistic* must test conditions be? Later, the scope of my project shifted somewhat. I visited the Propellants and Gases Prototype Laboratory on three occasions, witnessed two cryogenic tests, and (to a modest degree) participated in one of them. After talking with Messers. Fesmire, Popper, and Fox and learning what purposes the tests were meant to serve and what methods were used in them, a fourth question arose, namely (iv) in installing seal assemblies in existing hardware, and, if necessary, redesigning some details, how can one best exploit the advantages and mitigate the disadvantages of available seal materials?

What recommendations I have to make in regard to the first three questions were discussed in section 2.5 above. Thus,

R1 There is much reason to doubt that an ideal elastomeric plastic having the compliance of room temperature rubber and the inertness of TFE at the temperature of liquid hydrogen can be found or developed. Thus, future efforts should be directed towards the development of seal *assemblies* rather than solid seal made from virgin plastic.

R2 In pursuing the implications of R1, the study and evaluation of seal assemblies involving, for example, hollow plastic shells with spring backing, composite seal materials with glass reinforcement, inflatable plastic inner tubes, foam plastic or other alternatives should be accelerated.

R3 If a bolt joins parts of a coupling made of the same material and if the torque used to tighten the bolt at ambient conditions does not cause it to yield, then the stresses in the bolt may or may not remain below the elastic limit as the assembly is cooled to cryogenic temperatures. Future analyses that precede cryogenic tests (or installation of hardward in service) should address this question. Modification of the installation, *e.g.* by adding extra long bolts with spring-washers, or spacers, if necessary, should be undertaken to avoid yield of the bolt during cooldown.

R4 Tests should be undertaken to determine whether the gases that reside within foam TFE (or FEP) do or do not liquify at the temperature of liquid hydrogen and, if so, whether the foam exhibits any tendancy to collapse.

R5 The tests I witnessed at the Prototype Laboratory emphasized the production of data to support operational decisions. Some future tests should be dedicated to the longer range goal of generating a literature on seal design.

This summer, I was invited to undertake tests of the sort envisaged in R5 above. I refrained from doing so right away, however, owing to the need to study the available literature on the subject, bring the issues into sharper focus, and plan a test program that was cost effective and had a clear sense of direction. Further research proposals by myself or others for support for experiments to be conducted at KSC or my home institution (FIT) may well address R1-R5 in detail.

5. CONCLUSION

Cryogenic and noncryogenic seals at KSC are often subject to finite amplitude strains either in the opening or the closing of valves or the installation of couplings. Polymer seal materials, such as TFE, are subject to various memory effects including hysteresis and stress relaxation. There is no basis for confidence that an analysis of seal mechanics predicated upon the equations of linear elasticity can be quantitative. Only constitutive equations that capture the effects of large strains, three dimensionality, and memory can succeed in the quantitative analysis of seal mechanics. Functions rather than numbers appear in the quantitative representations of such a constitutive equation. Thus, the attributes of two competing seal materials can not be described with any degree of completeness by a few numbers. To be reliable, future test programs on competing seal materials must face these facts.

REFERENCES

1. Truesdell, C.A. & Noll, W. *The Nonlinear Field Theories of Mechanics.* Volume III/3 of *The Encyclopedia of Physics* (S. Flügge, General Editor) Springer-Verlag, 1965.

2. Schramm, R.E., Clark, A.F. & Reed, R.P. *A Compilation and Evaluation of Mechanical, Thermal, and Electrical Properties of Certain Polymers.* U.S. Department of Commerce, National Bureau of Standards, NBS Monograph 132, September 1973.

3. Poynting, J.H. 'On pressure perpendicular to the shear-planes in finite pure shears, and on the lengthening of loaded wires when twisted.' *Proceedings of the Royal Society of London* (Series A) **82**, 1909, pp 546-559.

4. Noll, Walter *The Foundations of Mechanics and Thermodynamics, Selected Papers.* Springer-Verlag, 1974.

5. Truesdell, C.A. *Essays in the History of Mechanics.* Springer-Verlag 1969. (Article 'The Creation and Unfolding of the Theory of Stress.', pp 184- 238. Quotation from p 186).

6. Noll, W. 'A Mathematical Theory of the Mechanical Behavior of Continuous Media.' *Archive for Rational Mechanics and Analysis.* Vol. **2**, pp 197-226, 1958. (Reprinted in Reference 4, pp 1-30).

7. Coleman, B.D. & Noll, W. 'An Approximation Theorem for Functionals, with Applications in Continuum Mechanics.' *Archive for Rational Mechanics and Analysis,* **6**, 1960, pp 355-370. (Reprinted in Reference 4, pp 97-112).

8. Coleman, B.D. & Noll, W. 'Foundations of Linear Viscoelasticity.' *Reviews of Modern Physics.* **33**, 1961, pp 239-249. (Reprinted in Reference 4, pp 113-123).

9. Ferry, J.D. *Viscoelasticity of Polymers.* Second edition, John Wiley & Sons, 1970, p 8.

10. Coleman, B.D., Markovitz, H. & Noll, W. *Viscometric Flows of Non-Newtonian Fluids, Theory and Experiment.* Springer-Verlag, 1966.

11. Noll, W. 'A New Mathematical Theory of Simple Materials.' *Archive for Rational Mechanics and Analysis.* **48**, 1972, pp 1-50. (Reprinted in Reference 4, pp 243-292).

12. Stokes, G.G. 'On the Theories of Internal Friction of Fluids in Motion and of the Equilibrium of Elastic Solids.' *Transactions of the Cambridge Philosphical Society* **8**, 1945, pp 287-319. (Reprinted in *The Scientific Papers of Geroge Gabriel Stokes*, Cambridge University Press, 1903 and Johnson Reprint Corporation, 1966, pp 75-129).

N90-16697

1989 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

ARTIFICIAL INTELLIGENCE IN PROCESS CONTROL: KNOWLEDGE BASE FOR THE SHUTTLE ECS MODEL

PREPARED BY:	Dr. A. Kent Stiffler
ACADEMIC RANK:	Professor
UNIVERSITY AND DEPARTMENT:	Mississippi State University Mechanical Engineering Department
NASA/KSC	
DIVISION:	Data Systems
BRANCH:	Technical and Information Systems
NASA COLLEAGUE:	Ms. Carrie Belton
DATE:	August 12, 1989
CONTRACT NUMBER:	University of Central Florida NASA-NGT-60002 Supplement: 2

ABSTRACT

The general operation of KATE, an artificial intelligence controller, is outlined. A shuttle ECS demonstration system for KATE is explained. The knowledge base model for this system is derived. An experimental test procedure is given to verify parameters in the model.

SUMMARY

A scaled down version of the shuttle ECS system is being built to test KATE, an artificial intelligence expert system at KSC. KATE requires an accurate mathematical model of the ECS, called the knowledge base. This report gives the model derivation.

An explanation of how KATE works is given. Each component must be described and loaded into KATE as a LISP "frame". The description includes a functional equation and component relationships to aid fault diagnosis.

The ECS system consists of four ducts branching from a manifold which is supplied with chilled air. Each duct has a heater, flow meter, a butterfly control valve, and two manual (butterfly) valves. The air temperature and flow rate is to be controlled precisely by KATE. Component failure is to be initiated to test KATE's diagnostics.

The mathematical description relates flow rates to pressures. A major element is the determination of component loss coefficients. The heater loss coefficient and time constant were derived from the fundamentals of flow over fins. Butterfly valves show little control at angles less than 40° and too much control at angles greater than 60°. Manual valve settings in each duct are obtained which will optimize the valve control sensitivity.

An experimental procedure is given to verify the loss coefficients of key components. Loss coefficients are expressed in terms of pressure and flow rate readings.

TABLE OF CONTENTS

SYMBOLS AND SUBSCRIPTS

Symbols

a	admittance
A	duct area
C	constant
D	duct diameter
f	pipe friction factor, 0.017
g_r	32.17 lbm-ft/lbf - sec^2
$\dot{H}$	heat transfer rate
K	loss coefficient
L	duct length
$\dot{M}$	mass flow rate
P	pressure
T	temperature, Rankine
V	velocity
α	venturi loss, fraction of ΔP
B	D_2/D_1
ρ	air density
θ	butterfly valve setting or heater divergent wall angle to horizontal

Subscripts

a	aft
b	bend
c	cabin
d	dump
e	entrance
f	forward or fins
h	manual (hand) valve
L	line (duct)
m	motorized (control) valve
p	payload
r	reheat (heater)
s	supply
T	total
1	venturi entrance
2	venturi throat

I. INTRODUCTION

KATE (Knowledge-based Autonomous Test Engineer) is a knowledge-based expert system being developed at Kennedy Space Center. The basic concept of an expert system consists of two parts. The first part is a knowledge base which describes the process hardware. This allows for a simulation of real time behavior, and it can be used to gather information about the status of the system. The second part is the software or artificial intelligence (AI). AI represents a good design engineer, systems engineer, and operator in a single software package that operates on the knowledge base representation of the system. Software capabilities include graphical display generation, simulation, process control, redundancy management, constraint checking and diagnostics. In particular, unanticipated failures can be diagnosed, and instructions can be given to get the system to a new desired state, or correct it and return it to a normal state.

KATE has been under development at Kennedy Space Center since 1985. The first full test of the system will be carried out this Fall. The process chosen for the task is a scaled down version of the shuttle environmental control (ECS) ground system. Such a system consists of thermal, fluid and electrical components. In addition, the sensors and control actuators are both analog and digital. A mathematical description of these elements constitutes the process knowledge base. The summer task is to assist Kennedy engineers obtain this knowledge base.

C-4

II. HOW THE SOFTWARE WORKS

Artificial intelligence (AI) systems are composed of three parts: (1) input/output, (2) knowledge base, and (3) operating system. Input consists of the sensors (transducers), supplying the system with pressure, temperature, flow, etc. data in the form of voltages to an analog-to-digital converter. Outputs are the commands to actuators regulating the variables of temperature, flow, etc. The knowledge base is a mathematical description of the process. The operating system is the set of algorithms which interprets the data and knowledge base, controls the process and diagnoses faults. The latter is KATE. Her language is LISP although other AI languages include ADA and C.

2.1 KNOWLEDGE BASE

The operating system is sufficiently general that it can be applied to all processes. It is the knowledge base that changes with each process. There are two types: expert rules and model base.

With expert rules, the programmer sits down with the plant operator and grills him on how he controls the process. A plant is usually modified over the years, and the operator may be the only person who knows the system. He is aware of the system idiosyncrasies and often learns by experience why a certain procedure works. The programmer must root out these rules of intuitive behavior. These rules are of the form: if , then . For example, if valve one is open or valve two is closed and liquid level switch two is set, then open valve three. The rules are adaptable to Boolean algebra. Some large systems have been known to take five man-years and generate thousands of rules. Drawbacks to this approach are several.

a. It cannot react to unknown conditions.

b. There is no diagnosis when faults occur. The process must be shut down with a simple sensor failure.

c. The intelligence and system operation resides mainly within the knowledge base rules; and therefore, they change with each process.

KATE uses a "model base" knowledge base. It depends on a mathematical description of each component in the process. Thus, each measured variable can be calculated also. A comparison of the two is the basis for a complete fault diagnosis. At the same time the AI system can suggest alternative commands to circumvent the failed component. Intelligence resides in the operating system, and only the component mathematical descriptions need to be loaded when a new process is brought under AI control.

As an example of a component description, consider a venturi flow meter in an air duct. Normally, one calculates the volume flow rate from the differential pressure reading. In our case we want the mass flow rate which will include the duct temperature and pressure (in place of density) calculation. The equation may look like

$$DP = 6.2P\dot{M}^2/T$$

Notice that the equation is the inverse of the normal output $\dot{M}$, i.e. we want the differential pressure in terms of the flow rate. The reason is that the AI system must calculate *DP* to compare it with the measured *DP* for fault diagnosis.

All components and variables are prepared for programming by placing its information into a "frame". An example follows for the differential pressure above.

LISP FRAME

```
(deframe DPA
     (nomenclature aft duct differential pressure)
     (a i o PRESSURE)
     (source-path P
                  T
                  F)
     (in-path-of DPA TRANSDUCER)
     (status (/(*(6.2)(square Ṁ)(P))T)))
```

Line 1.	Variable symbol.
Line 2.	Description.
Line 3.	An instance of. What is the category? A variable: pressure, temperature, flow? An item that can fail: transducer, relay, heater?
Line 4.	Source-path. What variables are used to calculate DPA?
Line 5.	In-path-of. What component is in the path of DPA?
Line 6.	Status. The equation for DPA.

2.2 KATE'S OPERATING SYSTEM

KATE'S operating system consists of algorithms for fault diagnosis which function basically by forming lists. After a delay for command dynamics to settle out, the following procedure is used to identify failed components and determine alternate commands to circumvent the problem.

a. Get new measurements from sensors throughout the process. Compare these new measurements with previous readings stored in memory. Form a list of any measurements that change by more than a prescribed amount.

b. Go to the knowledge base and calculate the variable that is measured for each discrepancy in the above list. If the calculated and measured values disagree, place the deviants into a second list for the DIAGNOSER.

c. When a component fails, it is likely to generate many deviants for the second list above. The DIAGNOSER goes to the knowledge base and forms a list of all possible components that can be related to the deviant. Recall the frame for differential pressure. Line 5 identifies a transducer as a possibility. Line 4 has several variables that can be related to components upstream through other frames. The computer knows which categories (Line 3) in every frame that can fail. Every deviant measurement caused by the failed component will have a similar, though not identical, list of process components that factor into its reading. Every list will have several process components in common. KATE uses elementary set theory to identify those components with the most intersections and ranks them in the order of most likely to cause the deviations. Starting with the first component in the list, KATE simulates the component failure modes and calculates every measurement in the process from the knowledge base. The component mode can be ON/OFF for a relay to multiple command voltages for a motor control valve. KATE proceeds through the list until a component and its failure mode is found which will cause agreement between all process measured and calculated variables. If KATE is told to find a new command to circumvent the failure and to maintain variable settings, the procedure is the same as above. Command changes are simulated until the desired result is obtained. This is valuable at Kennedy Space Center where systems often have built in component redundancy.

III. SHUTTLE ECS DEMONSTRATION UNIT

A top view drawing of the scaled down shuttle environmental control system (ECS) is shown in Figure 1. It is built within a 18 by 13 foot open blockhouse. An external chiller (a mobile air force purge unit) supplies 100 lbm/min of cold air at 43° F, 3 to 4 psig through an 8 inch line. The line turns 180° into a 12 inch manifold which distributes the cold air to four ducts for the simulated shuttle payload, aft, forward and cabin areas. A motorized dump valve balances the out flow for the total of 100 lbm/min.

Each duct is identical and has the same components with one exception: the payload duct is a nominal 6 inch pipe (6.065ID) while the remaining ducts are a nominal 3 inch pipe (3.068ID). The components are as follows.

a. Keystone hand operated butterfly valve.

b. Chromalox air heater (6kw and 15kw, payload duct).

c. Leeds and Northrup venturi meter.

d. Keystone motorized butterfly valve.

e. Keystone hand operated butterfly valve.

Each duct also contains, as indicated in Figure 1, the following.

a. Two remote pressure sensors.

b. One remote temperature sensor.

c. Two pressure gauges.

d. One temperature gauge.

e. One remote differential pressure sensor (venturi).

Both the air temperature and flow rate are controlled in each of the four ducts. The temperature is controlled to within + 0.5° F over the range: 60 to 70° F. The flow rate is controlled to within 0.2 lbm/min where the range for each duct is as follows.

Payload duct:	28 to 55 lbm/min
Cabin duct:	4.6 to 12 lbm/min
Forward duct:	6 to 17.6 lbm/min
Aft duct:	11.4 to 14.2 lbm/min

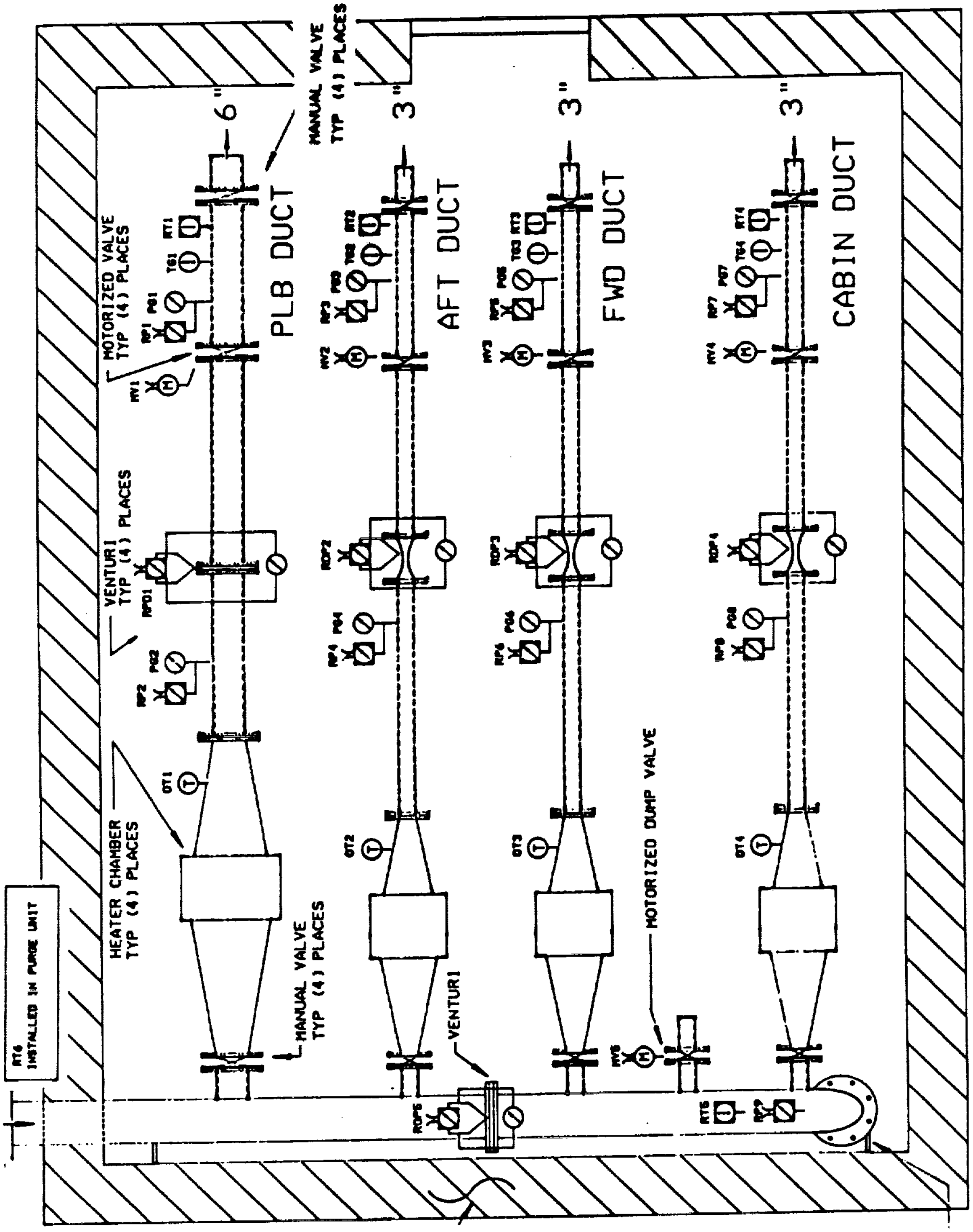


Figure 3-1. Shuttle ECS Test Unit

IV. ANALYTICAL DESCRIPTION

4.1 BASIC EQUATIONS

A mathematical description of the overall test unit consists of equations relating the flow rate and pressure drops for each duct in terms of admittance.

$$P_s = f(\dot{M}_s)$$

$$\dot{M}_s = a_s\sqrt{P_s - P_m}$$

$$\dot{M}_p = a_p\sqrt{P_m}$$

$$\dot{M}_a = a_a\sqrt{P_m}$$

$$\dot{M}_f = a_f\sqrt{P_m}$$

$$\dot{M}_c = a_c\sqrt{P_m}$$

$$\dot{M}_d = a_d\sqrt{P_m}$$

$$\dot{M}_s = a_t\sqrt{P_m}$$

$$a_t = a_p + a_a + a_f + a_c + a_d$$

$$P_m / P_s = a_s^2/(a_t^2 + a_s^2)$$

Each duct includes a number of components and elements in series. The duct admittance can be found in terms of the individual admittances. For example, the aft duct admittance is found from

$$a_a = \left[1 / \sum_{n=1}^{N} 1/a_{an}^2 \right]^{1/2}$$

where n is an individual admittance.

The admittances for each element can be expressed in terms of a loss coefficient K. The pressure drop is given by

$$\Delta P = K\rho V^2/2g_c$$

where $M = \rho A V$. Rearranging,

$$a = A\sqrt{2\rho g_c / K}$$

For a perfect gas,

$$\rho/\rho_o = \frac{P}{P_o}\frac{T_o}{T} \quad \text{(use absolute values)}$$

Then

$$a = \sqrt{2\rho_o g_c}\, A\sqrt{PT_o/P_oT} / \sqrt{K}$$

For air at 70° F, 14.7 psia,

$$a = 11.0A\sqrt{PT_o/P_oT} / \sqrt{K} \quad \text{lbm/min} \quad \sqrt{\text{psi}}$$

$$a = 2.09A\sqrt{PT_o/P_oT} / \sqrt{K} \quad \text{lbm/min} \quad \sqrt{\text{in.}H_2O}$$

and for elements in series

$$a = 11.0A\sqrt{PT_o/P_oT}\,[\textstyle\sum K_n]^{-1/2}$$

We will now review the individual components and elements in order, starting with the purge unit. The aft duct branch will be taken but the calculated variables are the same for each duct unless stated otherwise.

4.1.1 SUPPLY PIPE.

$$K_{SL} = fL/D = (0.017)(20)/(.666) = 0.5$$

4.1.2 SUPPLY FLOW METER.

$$\dot{M}_s = YC_dA_2\sqrt{\frac{2\rho g_c \Delta P_T}{(1 - B^4)}} \quad (\Delta P_T \text{ across taps})$$

$$Y \text{ (compressibility factor)} \approx 0.98 \text{ (neglect)}$$

$$\Delta P_v \text{ (loss across venturi)} = \alpha \Delta P_T$$

$$\alpha = 0.10 \text{ for } \theta = 7^\circ, \; B = 0.45$$

Flow Measurement Engineering Handbook REF TC 177

$$C_d = 0.96$$

$$K_{sv} = \left(\frac{1 - B^4}{B^4}\right)\left(\frac{\alpha}{C_d^2}\right) = 2.5$$

Loss between P_s and RP9,

$$\sum K = K_{sl} + K_{sv} = 3.0$$

4.1.3 90° BEND.

$$K_{sB} = 0.3$$

4.1.4 EXPANSION TO MANIFOLD.

$$K_{se} = \left[1 - \left(\frac{d}{D}\right)^2\right]^2 = .3$$

4.1.5 CONTRACTION: MANIFOLD TO DUCT.

$$K_{me} = 0.45$$

$$= 0.36 \text{ (PLB)}$$

4.1.6 BUTTERFLY VALVE (MANUAL 1).

$$K_{ah1} = \left(\frac{1.56}{1 - \sin\theta} - 1\right)^2 \qquad \theta > 25°$$

See Section 3.1.12

4.1.7 REHEAT CHAMBER.

Gradual enlargement

$$K_1 \approx \left[1 - \left(\frac{d}{D}\right)^2\right]^2 = .9$$

Fin drag

$$F_f = \frac{\rho V_e^2}{2g_c} A_f \times \frac{1.328}{\sqrt{\Re_{eL}}} \times 400 \text{ fins}$$

Cylinder drag

$$F_c = c_d \frac{\rho V_\infty^2}{2g_c} A_c \times 6 \text{ cyl.}$$

Loss coefficient due to drag

$$\Delta P A_R = (F_f + F_c)$$

also

$$\Delta P = K_2 \rho V_\infty^2 / 2g$$

$$K_2 = A_f \frac{1.328}{\sqrt{\Re_{eL}}} 400 + c_d A_c \times 6$$

$$= 76$$

correct to duct velocity at A

$$K_2 = 76(A_r/A)^2 = .25$$

total reheat chamber loss

$K_{ar} = 0.9 + 0.25 = 1.2$ (not calculated for PLB but scales up similarly)

reheat chamber time response

$\dot{H}_1 = MCdT_f/dt$ heat transfer to fin capacitance

$\dot{H}_2 = (T_f - T_o)h_f A_f$ heat transfer to film resistance

$$\dot{H} = \dot{H}_1 + \dot{H}_2 = MCdT_f/dt + h_f A_f (T_f - T_o)$$

But heat transfer to the film resistance is given to the air.

$$(T_f - T_o)h_f A_f = \dot{M}_a C_p (T_o - T_i)$$

Eliminating T_f and rearranging $\dot{H}$,

$$MCH(1 + (\dot{M}_a C_p / h_f A_f)) d(T_o - T_i)/dt + \dot{M}_a C_p (T_o - T_i) = \dot{H}$$

The time delay t_d is four time constants or

$$t_d = \frac{4MC}{\dot{M}_a C_p}(1 + (\dot{M}_a C_p / h_f A_f)) = 3.25 \text{ min}$$

$$= 4.0 \text{ (50\% fin efficiency)}$$

where

$$h_f = 0.664 \frac{k}{L} P_R^{1/3} (\mathfrak{R}_{eL})^{1/2} = 4 \times 10^{-3}$$

$$A_f = 840 \text{ in}^2 \text{ (est.)}$$

$$C = 0.11 \text{ Btu/lbm°F}$$

$$M = 11.5 \text{ lbm (est.)}$$

$$C_p = 0.24$$

reheat chamber sensitivity

$$\text{Power} = \dot{M}_a C_p (T_o - T_i)$$

at full power the temperature change of the air is

$$T_o - T_i = \frac{6000W \times 3.414 \frac{\text{Btu/hr}}{W} \times \frac{1 \text{ hr}}{60 \text{ min}}}{12 \text{ lbm/min} \times 0.24 \text{ Btu/lbm°F}} = 118°$$

For an 8 bit D/A converter, the temperature can be controlled to within 0.5° F at flow rate mid-range. Worse case: low cabin flow rate (4.6 lbm/min) where $T_o - T_i = 308°$ or control to within 1.2° F.

4.1.8 DUCT SECTION HEATER TO RP4 SENSOR.

$$K_{aL1} = fL/D = 0.18$$

$$= 0.09 \text{ (PLB)}$$

Loss between RP9 and RP4,

$$\sum K = K_{S8} + K_{se} + K_{me} + K_{ah1} + K_{ar} + K_{aL1}$$

$$= 2.4 + K_{ah1}$$

4.1.9 DUCT SECTION SENSOR TO FLOW METER.

$$K_{al2} = fL/d = 0.1$$

$$= 0.05 \text{ (PLB)}$$

4.1.10 FLOW METER.

Same equations as the supply flow meter.

$$3 \text{ in. duct} : B = 0.671, \ \theta = 5°$$

$$6 \text{ in. duct} : B = 0.45, \ \theta = 7°$$

Meters for ducts 4 inches in diameter or less have a slight sudden contraction from duct size A to area A_1, at the high pressure tap.

$$K_1 = \frac{(1 - B^4)}{B^4}\left(\frac{\alpha}{C^2}\right)\left(\frac{A}{A_1}\right)^2 = 1.6$$

$$K_2 = .28 \text{ (sudden contraction)}$$

$$K_{av} = K_1 + K_2 = 1.9$$

$$K_{al} = \frac{(1 - B^4)}{B^4}\left(\frac{\alpha}{C^2}\right) = 2.5 \text{ (PLB)}$$

4.1.11 DUCT SECTION TO MOTORIZED BUTTERFLY VALVE.

$$K_{al3} = fL/D = 0.1$$

$$= 0.05 \text{ (PLB)}$$

4.1.12 MOTORIZED BUTTERFLY VALVE

If both hand butterfly valves are full open, a simple calculation will show that the motorized valve must be set to nearly closed $\theta \approx 70°$ for the mid-flow range where its $K \approx 600$. A one degree change will change K by 150, much too sensitive for satisfactory flow rate control. Furthermore, at nearly open $\theta = 20°$, a one degree change will change K by 0.5. If the hand valves take too much of the loss to achieve the mid-flow rate setting, the motorized valve loses control. A logical solution is to select a minimum permissible setting for the maximum flow rate. This will maintain control, yet allow a maximum sensitivity for control. Selecting $\theta_{min} = 40°$

$$(K_{am})_{min} = 11$$

Now we can determine the total K by summing the Ks along a streamline from P_s (purge unit) to a selected duct end at $P = 0$ psig. The Ks in the larger supply duct should be corrected to the common velocity of a distributing duct or

$$K_{new} = K_{old} \times \left(\frac{\dot{M}_s / A_s}{\dot{M}_a / A_a} \right)^2 = 1.4 K_{old}$$

Then

$$\dot{M}_a = a\sqrt{P_s}$$

where $a = 11.8A[\sum K_n]^{-1/2}$ using an average value for the pressure/temperature correction. The butterfly setting for the two hand valves is found from the maximum duct flow with the control valve at $\theta = 40°$:

$$(\dot{M})_{max} = 11.8A[20 + 2K_h]^{-1/2}\sqrt{P_s}$$

solving for each duct ($P_s = 3.5$ psig)

,

$K_{ah1} = K_{ah2} = 50$ $\quad\theta = 53.8°$

$K_{fh1} = K_{fh2} = 29$ $\quad\theta = 49.1°$

$K_{ch1} = K_{ch2} = 75$ $\quad\theta = 57°$

$K_{ph1} = K_{ph2} = 54$ $\quad\theta = 54.4°$

Butterfly control valve settings for the average flow rate in each duct (hand valve settings above) are

$K_{am} = 40$ $\quad\theta = 51.9°$

$K_{fm} = 108$ $\quad\theta = 59.7°$

$K_{cm} = 194$ $\quad\theta = 63.6°$

$K_{pm} = 109$ $\quad\theta = 59.7°$

4.1.13 DUCT SECTION CONTROL VALVE TO SENSOR RP3.

$$K_{al3} = fL/D = 0.10$$

4.1.14 DUCT SECTION SENSOR RP3 TO HAND VALVE.

$$K_{al4} = fL/D = 0.10$$

4.1.15 BUTTERFLY VALVE (MANUAL 2).

$$K_{ah2} = \left(\frac{1.56}{1 - \sin\theta} - 1\right)^2 \qquad \theta > 25°$$

See Section 3.1.12

V. EXPERIMENTAL DESCRIPTION

An accurate model of the process begins with an analysis of the components and a calculation of the most probable values for the variables. Along the way certain engineering parameters must be estimated. The final link in the model development is an experimental program to verify the analysis and its associated parameters. For the ECS system there are several verifications that are most important.

(1) Characteristic curve of the purge unit: $P_s = f(\dot{M}_s)$

(2) Loss coefficient versus angle for the butterfly valves.

(3) Flow meter loss coefficient.

(4) Heater loss coefficient.

(5) Heater response time.

The ECS system is now being assembled and it should be completed for tests by the middle of August. The initial tests to verify the parameters will be run without KATE, using the various gauges that accompany the remote sensor units. I propose the following test procedure while recording all gauges.

5.1 OPEN ALL DUCT BUTTERFLY VALVES

a. Record minimum P_s on the characteristic curve along with its $\dot{M}_s$.

b. Find the open manual butter fly loss coefficient K_{ah2}. It should be the same for all valves.

$$\dot{M}_a = a\sqrt{PG3}$$

$$a = c[K_{aL4} + K_{ah2}(\text{open})]^{-1/2}$$

c. Find the venturi loss coefficient K_{av}

$$\dot{M}_a = a\sqrt{PG4 - PG3}$$

$$a = c[K_{aL2} + K_{av} + K_{aL3} + K_{am}(\text{open}) + K_{aL3}]^{-1/2}$$

d. Find the heater loss coefficient K_{ar}

$$\dot{m}_a = a\sqrt{PG9 - PG4}$$

$$a = c[((K_{SB} + K_{se})1.4 + K_{ah1}(\text{open}) + K_{ar} + K_{al1}]^{-1/2}$$

e. Check the supply admittance.

$$\dot{M}_s = a\sqrt{P_s - PG9}$$

$$a = c'[K_{Sl1} + K_{sv} + K_{sl2}]^{-1/2}$$

f. Check all ducts for identical gauge readings.

5.2 VARY CONTROL VALVE SETTINGS IN CABIN DUCT.

a. Obtain $K_{cm} = f(0)$

b. At K_{cm} (closed) check to see if manifold pressure given by PG8 agrees with P_s. This verifies that the supply line loss is negligible. Record P_s, M_s for another point on the characteristic curve.

5.3 CLOSE REMAINING DUCTS IN TURN

a. Completely close aft, forward and payload bay ducts, recording P_s and $\dot{M}_s$ after each closing.

b. Slowly close the dump valve. Continue recording P_s, $\dot{M}_s$ for the characteristic curve. Stop when P_s reaches a maximum.

5.4 DETERMINE HEATER TIME RESPONSE

Introduce a step change in the heater command. Measure the outlet air temperature every 10 seconds until a new temperature stabilizes.

VI. CONCLUDING REMARKS

KATE depends on an accurate knowledge base flow model. The ECS system is relatively simple with various loss coefficients the predominant determinate. Among the loss coefficients calculations for bends, contractions, expansions, smooth pipe lengths and even venturi flow meters are straight forward from handbooks. These values are small and very reliable. By subtracting these contributions between pressure gauges, the less certain loss coefficients for the heater and butterfly valves can be verified experimentally. The heater loss coefficient was calculated from aerodynamic drag principles. The formula for the butterfly valve loss coefficient versus angle was found in the "Handbook of Hydraulic Resistance", translated from Russian. There is no indication of the formula source, likely to be theoretical.

The single critical element in the knowledge base is the butterfly valve. The valve angle is set by activating the butterfly drive motor for a specified time. The multiple spur/worm gear drive is likely to suffer backlash. Thus, open loop positioning cannot be reliable. KATE should use the position potentiometer with some type of software feedback. The problem is compounded by the butterfly flow sensitivity. The mid-range flow rate valve settings for the three series (two manual plus one control) butterfly valves are approximately 55°. These settings were established to give the smallest allowable change in loss coefficient per angle of movement, thus the widest range for the angle over the flow range. Yet, the valve is so nonlinear that the high flow rates will change by 0.15 lbm/min per degree while the low flow rates will change by 0.55 lbm/min per degree over the angle span of 30° (forward duct example).

Finally, the heater time response was calculated to be approximately 4 minutes independent of the temperature command change magnitude. The length of time appears to be large. Thus, it represents an element of anticipation in the experimental study.

N90-16698

1989 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

THE EFFECTS OF HYDROGEN EMBRITTLEMENT OF TITANIUM

PREPARED BY:	Mr. Delbert J. Taylor
ACADEMIC RANK:	Instructor
UNIVERSITY AND DEPARTMENT:	Murray State College Engineering Technology
NASA/KSC	
DIVISION:	Materials Science Laboratory
BRANCH:	Malfunction Analysis Branch
NASA COLLEAGUE:	Mr. Cole Bryan
DATE:	August 8, 1989
CONTRACT NUMBER:	University of Central Florida NASA-NGT-60002 Supplement: 2

ABSTRACT

Titanium alloys, by virtue of their attractive strength to density ratio, fatigue, fracture toughness and corrosion resistance are now commonly used in various aerospace and marine applications. The cost, once very expensive, has been reduced making titanium even more of a competitive material today. Titanium and titanium alloys have a great affinity to several elements. Hydrogen, even in small amounts, can cause embrittlement, which in turn causes a reduction in strength and ductility. The reduction of strength and ductility is the subject of this investigation.

SUMMARY

Specimens of alpha-beta titanium were tensile tested to determine the effects of hydrogen embrittlement on room temperature mechanical properties, primarily the strain rate dependence of ductility. The combination of stress concentrations and hydrogen contamination decrease the strength of this alloy. A decrease of strength at the lower strain rates was observed.

TABLE OF CONTENTS

LIST OF FIGURES

I. INTRODUCTION AND BACKGROUND

Titanium (Ti), named after the Greek God Titan, was discovered in 1791, but was not commercially produced until the 1950s. Although expensive, the high strength-to-weight ratio of titanium and its corrosion resistance at room and elevated temperatures make it attractive for applications such as airframe structures, jet engines, missile and spacecraft parts, marine components, submarine hulls, and biomaterial such as prosthetic devices. Titanium alloys have been developed for service at 1000° F for long periods of time and up to 1400° F for shorter periods.

Titanium and titanium alloys are classified into three major categories according to the predominant phases present in the microstructure. Titanium has a hexagonal close packed crystal structure called alpha at room temperature. At approximately 1600° F, the alpha phase transforms to a body centered cubic structure called beta, which is stable up to approximately 3000° F (1).

Alloying elements favor either the alpha or the beta phase or are neutral. Alpha-beta titanium alloys are a compromise between the single phase alpha and beta alloys. Alpha phase stabilizing interstitials include aluminum, oxygen, nitrogen and carbon. The beta stabilizing interstitials include copper, chromium, columbium, iron, manganese, molybdenum, tantalum, vanadium and hydrogen (2).

Titanium alloys have a great affinity to the beta stabilizing interstitial hydrogen. Two types of hydrogen embrittlement will be exhibited; these have been designated impact embrittlement and low strain rate embrittlement. The type of hydrogen embrittlement that is most often encountered in high strength alpha-beta titanium alloys is the low strain rate type. Sensitivity of titanium alloys to low strain rate appears to increase with increasing tensile strength, notch severity, alpha grain size, continuity of the beta phase and the hydrogen content (3).

II. TEST PROCEDURES

2.1 MATERIALS AND EQUIPMENT

The testing methods selected for the investigation of hydrogen embrittlement were based on the slow strain rate sensitivity of alpha-beta titanium alloys. Specimens for the test were 3/8 inch diameter round gage alpha-beta titanium rods 7 inches in length. The specimens were machined to reduce the mid-section diameter to 0.250 inches. The ends were threaded to a 3/8-16 pitch thread (Figure 2-1). The large ratio of thread diameter to gage diameter was required to prevent brittle failure in the threads. All tensile tests were made using a Tinius Olsen tensile testing machine using an appropriate load scale.

2.2 TEST RESULTS

Two different strain rates were used. At a strain rate of 0.05 inches per minute from 0 load to failure, the ultimate tensile strength of the specimen was 191,900 psi. At a strain rate of .005 inches per minute from 0 load to failure, the ultimate tensile strength was 173,560 psi. The breaking strength was 161,950 psi and 135,670 psi respectively. The stress-strain curves are shown in Figure 2-2.

2.3 MICROSTRUCTURES AND HARDNESS

The microstructure of the titanium specimen (Figure 2-3) shows fine alpha grains (light) and intergranular beta. Microhardness tests revealed a case hardness of HK 876.4 (HRC 66.0) with a .002 inch depth, then rapidly decreasing to HK 380.2 (HRC 39). Figure 2-4 is the fracture surface of specimen 1 showing a flat break. The higher strain rate yielded a much higher tensile strength than the slow strain rate of specimen 2 (Figures 2-5 and 2-6) where a rough fracture surface with deep internal and surface cracking occurred.

REFERENCES

1. Donachie, M.J. Introduction to Titanium and Titanium Alloys. Source Book, ASM, 1982.
2. "A General Summary of the Physical Metallurgy of Titanium Alloys." TML Report No. 19, 1955.
3. Groeneveld, T.P., Fletcher, E.E., Elsea, A.R. A Study of Hydrogen Embrittlement of Various Alloys. Battelle Annual Report, 1966.
4. Metals Handbook, 9th Edition. Metallography and Microstructures. ASM, Vol. 9, 1985.
5. Metals Handbook, 9th Edition. Mechanical Testing. ASM, Vol. 8, 1985.
6. "Hydrogen Embrittlement Testing." ASM-STP543, 1974.
7. Private communication from C.R. (Bob) Denaburg, Chief, Malfunction Analysis Branch, Kennedy Space Center, Florida.
8. Olofson, C.T. Manual on the Machining and Grinding of Titanium and Titanium Alloys. Battelle, TML Report No. 80, 1957.

Figure 2-1. Alpha-beta Titanium Test Specimen

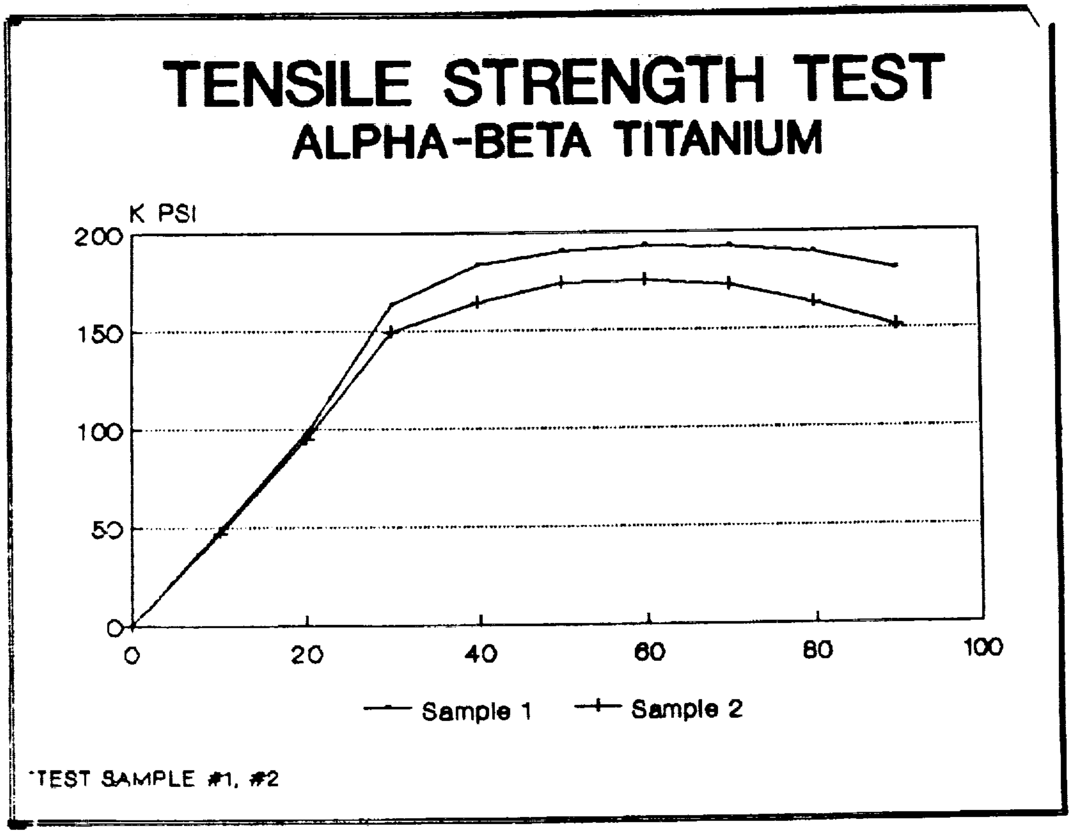


Figure 2-2. Stress Strain Curves of Alpha-beta Titanium Test Specimens

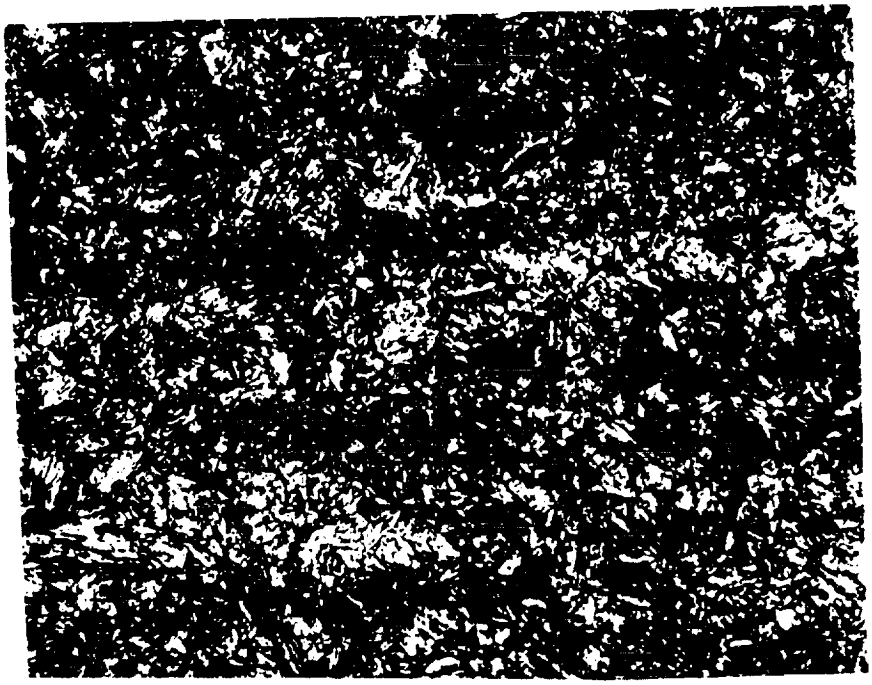


Figure 2-3. Microstructure of Titanium Alloy

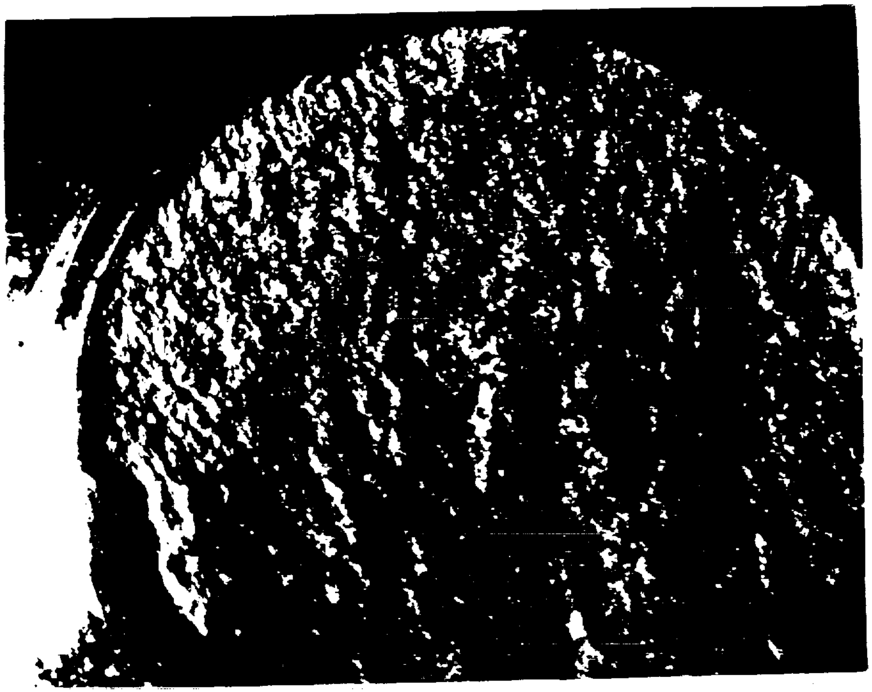


Figure 2-4. Fracture Surface of Specimen Number 1

ORIGINAL PAGE
BLACK AND WHITE PHOTOGRAPH

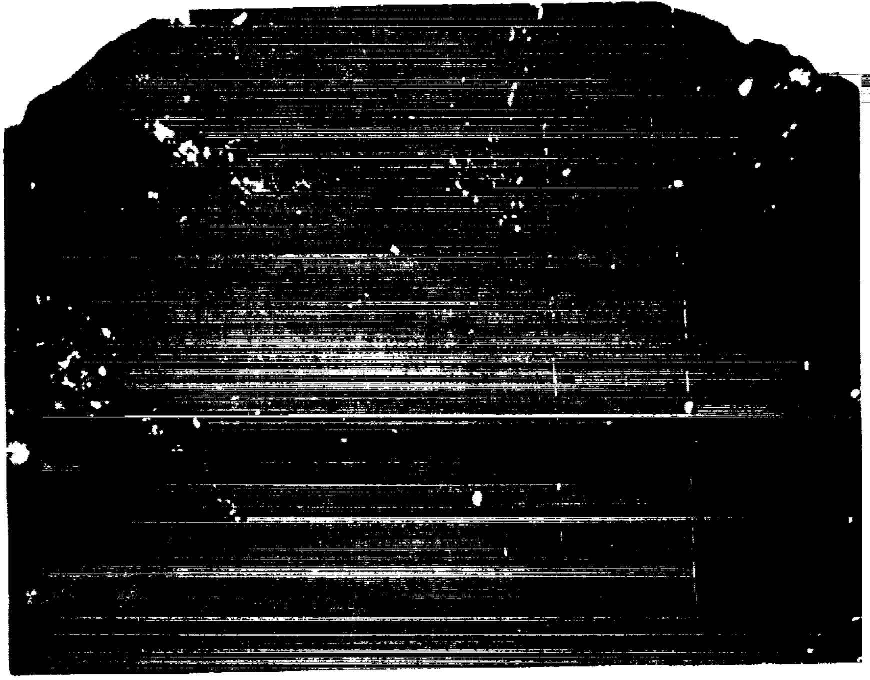


Figure 2-5. Fracture Surface of Alpha-beta Titanium Specimen Number 2

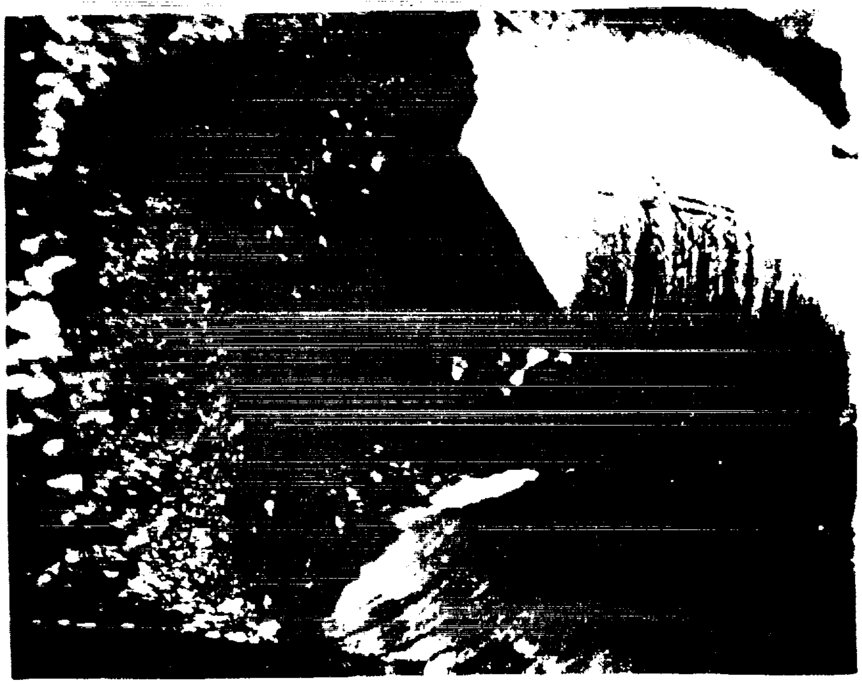


Figure 2-6. Fracture Surface of Alpha-beta Titanium Specimen Number 2

N90-16699

1989 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

OFF-LINE ROBOT PROGRAMMING AND GRAPHICAL VERIFICATION OF PATH PLANNING

PREPARED BY:	Dr. Gregory L. Tonkay
ACADEMIC RANK:	Assistant Professor
UNIVERSITY AND DEPARTMENT:	Lehigh University Department of Industrial Engineering
NASA/KSC	
DIVISION:	Mechanical Engineering
BRANCH:	Special Projects (RADL)
NASA COLLEAGUE:	Mr. V. Leon Davis
DATE:	August 8, 1989
CONTRACT NUMBER:	University of Central Florida NASA-NGT-60002 Supplement: 2

ACKNOWLEDGEMENTS

I would like to thank Leon Davis, my KSC colleague, for the opportunity to participate in this program, his guidance in this project, and his answers to my many questions about the how's and why's of life at KSC. I would also like to thank Ray Hosler for the time and effort he put forth in directing the program and making it such a success.

ABSTRACT

The objective of this project was to develop or specify an integrated environment for off-line programming, graphical path verification, and debugging for robotic systems. Two alternatives were compared. The first was the integration of the ASEA Off-line Programming package with ROBSIM, a robotic simulation program. The second alternative was the purchase of the commercial product IGRIP. The needs of the RADL (Robotics Applications Development Laboratory) were explored and the alternatives were evaluated based on these needs. As a result, IGRIP was proposed as the best solution to the problem.

SUMMARY

The RADL at KSC is experiencing competition for on-line time with the robots. This is because all of the programming, development, and debugging ties up the robots. To alleviate this problem, it was proposed that an off-line programming and debugging environment be developed. This project explored two alternatives:

1) the integration of two existing software packages, ASEA Off-line Programming and ROBSIM.

2) the purchase of commercially available software.

The commercially available software chosen was Deneb Robotic's IGRIP. This package was evaluated because it could run on the InterGraph workstations currently at KSC.

This report exams the types of projects the RADL is involved with and determines several features which would be desirable. Next, each of the alternatives was evaluated based on these and other criteria.

The ASEA Off-line Programming package was found to be easy to use except for the wrist orientation coordinates. The user interface on the ROBSIM package was difficult to use. The potential user had to understand joint transformation matrices, Euler angles, and dynamic parameters. In addition, the current version at KSC had several bugs.

The IGRIP package was found to be extremely easy to use and performed most of the functions required by the RADL personnel. The one capability it did not possess was dynamic simulation. However, this could be supplied by interfacing one of several commercial packages. The IGRIP package was superior in all respects to the other alternative except for price. Even in this category, it was unclear how much it would cost to integrate ASEA and ROBSIM, thus making a cost comparison difficult.

The final recommendation in this project was to purchase IGRIP for the InterGraph workstations that currently exist at KSC.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND ACRONYMS

ARLA - Asea Robotics LAnguage

CAD - Computer Aided Design

GSL - Graphical Simulation Language

IGRIP - Interactive Graphics Robot Instruction Program

KSC - Kennedy Space Center

LaRC - Langley Research Center

MSFC - Marshall Space Flight Center

OPF - Orbiter Processing Facility

PCR - Payload Changeout Room

RADL - Robotics Applications Development Laboratory

ROBSIM - ROBotic SIMulation program

TCP - Tool Center Point

I. INTRODUCTION

1.1 ROBOTICS AT KENNEDY SPACE CENTER

With the recent proposal by President Bush to establish a permanent lunar base and initiate a manned mission to Mars, there will be an increase in activity at KSC. Launches will occur more frequently and more payloads will be processed. In order to meet this goal, NASA will need to apply robots to tasks in space as well as ground preparation and servicing of spacecraft. Robots have replaced men performing dangerous or tedious tasks in the industrial and service sectors. It is only natural that space related tasks should be the next frontier for robotics. Several tasks at KSC are candidates for robot applications; for example, working with hazardous fuels and cryogenics, inspecting spacecraft and payloads, and performing last-second tasks at the launch pads.

1.2 MISSION OF THE ROBOTICS APPLICATIONS DEVELOPMENT LABORATORY

The Robotics Applications Development Laboratory (RADL) was established to explore the feasibility of applying robotic principles to the shuttle/payload ground processing activities at KSC. The robotic prototype system in the laboratory provides a testbed for projects dealing with many aspects of ground preparation. Furthermore, the laboratory provides a training environment in robotics for engineers. With the expected increase in activity, the laboratory will experience increasing competition for resources, especially programming time on the robots.

1.3 OBJECTIVE OF THIS RESEARCH PROJECT

The objective of this research project was to advise RADL personnel of the best way to proceed in order to alleviate the problem of limited availability of programming time and application time on the robots in the RADL. Furthermore, an analysis of current and future projects has shown that several types of tasks consistently reoccur. Tools that could be applied to these tasks have been evaluated and are discussed in greater detail in this paper. A list of these tasks includes the capability to:

program off-line which reduces the time actually spent using the robot graphically view the robot moving through its environment to detect many programming errors even before the robot is operated

detect collisions between objects in the environment

place various robots (or variations of a proposed design) in a graphical model of the environment to determine optimal configurations and limits

design and locate fixtures in the environment to minimize access problems

detect singularities in a program before it is actually run on the robot

view multiple devices moving within the environment and verify the communication signals between the devices

II. RADL FACILITIES

2.1 ROBOTS

In its current configuration, the RADL consists of two robots: an ASEA IRB 90/2 and a PUMA 560. Most of the development work to date has been performed on the ASEA. This robot has large reach and payload capabilities and is mounted on a 30 ft track to further increase the already large work envelope. It is an ideal candidate to work on the large pieces of equipment that exist at KSC. The ASEA robot is also equipped with an adaptive control option that allows it to dynamically alter its path planning based on outside signals.

2.2 PERIPHERAL EQUIPMENT

The robots in the RADL are interfaced to several other pieces of equipment which provide additional support [1]. A MicroVax II is the central computer in the laboratory. It communicates with the ASEA robot through a computer link that has the capability to upload/download programs and perform control functions. The MicroVax II is also interfaced to a DataCube vision system that performs complex vision calculations, a MasterPiece 280 PPC programmable logic controller that can control process outputs and monitor inputs, and a MasterView graphics presentation system.

III. APPLICATION AREAS

There are several projects at KSC, currently underway or proposed, that could be significantly enhanced by the findings of this research project. This section will briefly describe some of these projects and relate how off-line programming and graphical verification of path planning could enhance the projects.

3.1 ORBITER TILE INSPECTION

Each time an orbiter returns to earth, the protective heat tiles must be inspected for damage and misalignment. Of particular importance are the leading and trailing edges of the wings, the nose, and around the landing gear. Each of the tiles are individually inspected; a time consuming and tedious task that is ideally suited for a robot. Past projects in the RADL have shown that a robot can effectively inspect a mockup of a section of the orbiter. However, before a robot is used near a real orbiter, a graphical verification of the program would provide a substantially increased level of confidence.

If a decision were made to incorporate a robot in the inspection process, NASA would require specifications about the type of robot that should be purchased or designed. A state-of-the-art design environment could show the robot moving through its range of motion next to the orbiter. The number and location of positions required to inspect the orbiter could be determined without even turning the robot on, let alone moving it near the orbiter. If a robot was being designed to perform the task, the designer could experiment with various link lengths, joint limits, and joint configurations to determine the optimal configuration. Commercially available robots could be quickly and easily compared to determine the optimal robot for the inspection task.

3.2 INSPECTION AND PROCESSING OF ORBITER PAYLOADS

This task would employ a robot to inspect the payload of the shuttle prior to lift off. It would also involve tasks to bring experiments on-line just prior to lift off. Examples would include turning switches on, removing lens caps, verifying that pieces are in place, and inspecting for sharp edges that could catch and tear the space suits of the astronauts.

This robot would most likely be located in the PCR (Payload Changeout Room) at the launch pad. A graphical design environment could be used to model a robot in the PCR to determine the optimal configuration. Also, a model of the locations of the payloads in the orbiter cargo bay would allow off-line program generation of the path to perform the inspection tasks. Collision detection capabilities could verify that no collisions would occur.

3.3 ORBITER RADIATOR INSPECTION

Prior to each flight, the radiators on the orbiter must be inspected. These radiators are located on the inside of the cargo bay doors. The inspection would take place while the orbiter, with the cargo bay doors open, was in the OPF (Orbiter Processing Facility). Most likely the robot would be suspended vertically from an overhead track. This would cause minimal interference with existing hardware in the OPF.

This project would benefit from a graphical design environment by using a model of the OPF to determine the envelope requirements for the robot to operate efficiently. Collision detection and program generation would also be important in the later stages.

IV. CONSIDERATION OF ALTERNATIVE METHODS

The current method of robot programming in the RADL utilizes a teach pendant. While this is an adequate method for repetitive tasks, such as in a manufacturing environment, it is not sufficient for highly intelligent tasks where complex decisions must be made in a constantly changing environment.

In the past, robot specification and design has been performed in a trial and error manner. While this method can provide an adequate solution, it seldom approaches the optimal; primarily because the designer does not have time to try many different alternatives. A graphical design environment can provide the designer with tools to quickly make changes in the design and view the results.

4.1 ASEA OFF-LINE PROGRAMMING PACKAGE WITH ROBSIM

The first alternative explored was to integrate several pieces of software currently in the RADL. This was proposed to minimize the total cost of the project. The first piece of software was the ASEA Off-line Programming Package which uses the language ARLA. This software runs on the MicroVax II and communicates with the robot using the ASEA Computer Link hardware. It provides the capability to program without the teach pendant. Generally, the same functions are provided in ARLA as with the teach pendant [2].

Locations in the program can be entered using the coordinate system of the robot, registers, or a special record mode using the teach pendant. The biggest problem encountered while trying to program entirely off-line was using the ASEA scheme of representation for wrist coordinates. It is very difficult to visualize the map between the real world and the robot coordinate system taking into account the current TCP (Tool Center Point) definition.

Other limitations found in ARLA are the lack of arithmetic and trigonometric functions, the lack of data processing capabilities, and the failure to incorporate the robot track as an additional robot axis. Arithmetic and trigonometric functions are important to calculate positions and orientations of objects in the environment. Data processing capabilities are required to store data in files or access databases. Finally, the robot track is considered to be an external axis by the controller. When a position is entered using the keyboard, the option is not given to enter values for the external axes. Therefore, the calculation of the coordinates of the TCP are not affected by the track position. This makes it difficult to use the track in any mode other programming with the pendant.

Since the ASEA package does not include any kind of graphics, and hence no way to debug a program except to test it on the robot, the ROBSIM package was evaluated as the graphical display tool. ROBSIM was developed by Martin Marietta for LaRC [3,4]. It was designed to be a dynamic simulator, taking into account the physical properties and constants of the links and joints. ROBSIM can provide a graphical simulation of a robot in its environment if the appropriate hardware is available (Evans and Sutherland terminal). Otherwise, it must be run without graphics. There were several problems encountered in trying to model robots with ROBSIM. The following sections will describe some of these problems in more detail.

4.1.1 GRAPHICS TERMINAL REQUIREMENTS. ROBSIM requires an Evans and Sutherland terminal for proper graphics display. This type of terminal has a series of analog dials that can be used to change the perspective of the display. Without the capability to alter the perspective from the default side view, the user cannot determine where the robot is in three-dimensional space. Although the help files state that a VT240 terminal can be used, it only permits a two-dimensional side view. No capability exists to alter the perspective in software.

4.1.2 INTEGRATION WITH INTERGRAPH. It would be difficult and time consuming to rewrite the ROBSIM I/O routines to interface with the InterGraph family of workstations which are available throughout KSC. The hooks are not readily available, and more importantly are not documented in the current version of ROBSIM.

4.1.3 LACK OF UPDATED DOCUMENTATION. The documentation is different from the current version of ROBSIM that is running on the VAX. The documentation is for the version developed by Martin Marietta. The version of ROBSIM currently running on the VAX was modified by LaRC to reside in their environment.

4.1.4 INVERSE KINEMATIC DIFFERENCES. ROBSIM uses its own internal kinematic solutions to relate joint values to the TCP position. The user must be knowledgeable about joint and link transformation matrices and Euler angles to understand how to use the program. The ROBSIM solutions and displays would only be as good as the model which the user entered for the robot. However, since ROBSIM does not provide for collision detection, the lack of accuracy would not cause a significant problem.

4.1.5 DATA EXCHANGE FORMAT. The two packages in question do not store data in the same format. Conversion programs could be written to interface the two packages, but at the expense of user-friendliness and speed.

4.1.6 PACKAGES LOCATED ON DIFFERENT SYSTEMS. Currently the ASEA Off-Line Programming package is installed on the MicroVax II in the RADL and the ROBSIM package is installed on the Engineering VAX. Since the ASEA program must remain connected to the Computer Link, ROBSIM would ideally be ported to the same system. Unfortunately, the MicroVax II does do have enough disk space to store the ROBSIM. The two packages could be interfaced using DECNET, by sacrificing some speed and convenience.

4.2 IGRIP OFF-LINE ROBOT PROGRAMMING AND SIMULATION SYSTEM

IGRIP is a commercially available software package that combines many of the features required in the RADL. The software was written by Deneb Robotics, Inc. and has been on the market for several years. It is considered by many to be one of the best in its class. Since InterGraph has taken the IGRIP software and ported it to their hardware and since there are many InterGraph workstations already located at KSC, a cost effective solution exists: the purchase and installation of IGRIP on an existing system.

4.2.1 FEATURES OF IGRIP. Although a complete description of IGRIP is beyond the scope of this report, some of the highlights are mentioned in this section so that the various options can be compared. IGRIP integrates a CAD system with a simulation/animation system to provide high quality, shaded surface images of the environment. Multiple robots with unlimited degrees of freedom can move through the environment, manipulate objects, and communicate with other devices. Collision detection and near miss situations can be detected between any group of objects in the environment. The simulation can be recorded and played back at a later time.

The inverse kinematic solutions can be generated by generic algorithms or user written in the language C. Complex devices can be constructed which have joint limit dependencies. The path the robot is to traverse is defined using tag points. Unreachable points on the path can be easily detected. A special mode automatically places a robot so that a group of points can be accessed. This mode would be especially useful in the tile inspection task.

Using GSL, the user can construct descriptions of how a device will operate and communicate with other devices in the environment. Over 40 commercially available robots are predefined in IGRIP. The capability also exists, via supplied translators, to upload/download native robot code for 8 major robot manufacturers (including the ASEA and PUMA robots located in the RADL).

4.2.2 DISADVANTAGES OF IGRIP. There are few disadvantages to the choice of this alternative. The first disadvantage is the cost of the software, approximately $60,000. The same software, written to run on a different workstation, could probably be purchased directly from Deneb Robotics at a lower cost. However, a the workstation would also have to be purchased.

The second disadvantage is that currently there is no integrated dynamic modeling package. For certain applications, this may be critical. However, dynamic simulation packages can be used in conjunction with this package to provide dynamic simulations of the environment.

4.3 OTHER ALTERNATIVES

There are other software packages on the market which have features similar to IGRIP. However, none have been ported to use InterGraph hardware and CAD files. Since these systems would require the purchase of an additional workstation, these packages were not explored in greater detail.

V. COMPARISON OF ALTERNATIVES

This section will attempt to compare the two alternatives using criteria which are important to the RADL. A summary of the results of this section are listed in Table 1.

5.1 USER-FRIENDLINESS

This is probably the most important criterion in comparing the usefulness of the packages. If a system is difficult to use, no one will take the time to learn it or use it once they have learned it. IGRIP is by far the best choice in this category. It is a mature product that has a proven interface using sophisticated graphics and a mouse/menu system. It is easy to learn and provides many useful analyses. The ASEA/ROBSIM package is at the other end of the spectrum. While the user interface of the ASEA package is acceptable, the ROBSIM package is slow and tedious to use. The documentation does not agree with the code and several bugs exist which frustrate the user.

Table 1. Comparison of Alternatives

Feature	ASEA/ROBSIM	IGRIP
Ease of Use	Cumbersome, Difficult	Easy, Mouse-driven Menus
Cost	$ for Integration Man-Hours	$ 60,000
Training	None	5 Free People
Compatibility	No Other Centers Known	MSFC (Could Be Others)
Flexibility	Redo Integration for Each Robot	40 Commercial Robots Defined 8 Translators to Native Code
Support	None	Yes
Collision Detection	None	Exact Intersection of Surfaces
Graphics	Wireframe	Wireframe, Shaded Surface
Kinematic Solutions	Internal	Internal or User-defined C Program
Availability	2-3 Man-Months	Immediately
Hardware	Evans & Sutherland for Full Graphics, VT240 for Limited Graphics	Intergraph Workstation

5.2 COST

The ASEA/ROBSIM package is the least expensive alternative because both packages are already at KSC. However, there would be a cost associated with interfacing the two packages and defining a model of the ASEA robot in the ROBSIM package. For optimal use of graphics, an Evans and Sutherland terminal would be required at an additional cost. Furthermore, the current version of ROBSIM has several bugs which would need to be removed. A rough estimate of time/cost required to define the model and build the system would be 2 to 3 man-months. The IGRIP package, on the other hand, has a higher initial cost ($60,000), but this includes the cost to install the software and train the operators.

5.3 TRAINING

IGRIP has superior training because of the availability of vendor-supplied courses. According to the InterGraph representative, the cost of the software includes training for 5 people. To further reduce the training cost, it might be possible to negotiate for this training to take place at KSC rather than the Deneb school. ASEA/ROBSIM training would be totally self-directed. Other than the resident expert who performs the integration of the two packages, no one would be available to answer questions pertaining to the working environment.

5.4 COMPATIBILITY WITH OTHER NASA CENTERS

This is a difficult category to award because there are no official packages at other NASA centers. While it is doubtful that anyone will use the combination of ASEA/ROBSIM, some centers may be using ROBSIM to perform dynamic modeling. MSFC is currently using the IGRIP package and highly recommends it. Choosing this option would ensure compatibility between KSC and MSFC.

5.5 FLEXIBILITY

Flexibility is defined here as the ease to add new robots and/or alter existing models. In this category, IGRIP is far superior to ASEA/ROBSIM. IGRIP has over 40 commercial robots predefined, including the PUMA located in the RADL. This feature provides the user with a unique capability. Given that the application environment is already defined, the user can insert several different types of robots to determine which one is best-suited for the task. An estimate of the cycle time can also be determined. Eight of the most common off-line programming language translators are also included to allow the user to generate downloadable programs for the robot. Both the ASEA (ARLA) and PUMA (VAL II) translators are included.

With the ASEA/ROBSIM packages, each new robot would have to be kinematically modelled. Also a separate off-line programming package would have to be purchased and integrated with ROBSIM. This would be a labor intensive operation repeated each time a robot is purchased.

5.6 SUPPORT/UPDATES

IGRIP has the best support of the two alternatives. Support is available from InterGraph and Deneb Robotics. Updates are free for some specified time period (1 to 2 years).

On the other hand, the ASEA/ROBSIM combination offers little support. While ASEA will continue to support the ARLA language, ROBSIM is not currently supported and the likelihood of updates being released is low. Each time an update is received, the two packages must be combined again and the interface code rewritten.

5.7 COLLISION DETECTION

Since no collision detection is available in ROBSIM, IGRIP is superior in this category. IGRIP provides collision detection using an exact, surface to surface intersection calculation. Checking can be limited to any number of objects. A near miss mode and nearest distance between two objects mode are also available with the tradeoff of a reduction in processing speed.

5.8 TYPE OF GRAPHICS

IGRIP is also superior in this category. Images can be depicted using wireframe, shaded surface, or transparent modes. Calculations and screen updates are performed quickly, depending on the number of elements in the environment.

ROBSIM provides only wireframe images. These images are adequate when using the suggested Evans and Sutherland terminal (which is not available at KSC). With a VT240 terminal, only two-dimensional images are available. Furthermore, the point of reference cannot be changed.

Without the Evans and Sutherland terminal, the graphical analysis capabilities are severally limited.

5.9 KINEMATIC SOLUTIONS

IGRIP is the best choice in this category. The inverse kinematic solutions are implemented for all of the commercial robots included in the package. Furthermore, the user can write programs to calculate the kinematic solutions for any type of device. Thus, dynamic effects can be incorporated in the calculations.

The user has no control over the kinematic solutions used in the ROBSIM package. The program would have to be altered to add this feature, if it was required.

5.10 AVAILABILITY

IGRIP is available immediately. The ASEA/ROBSIM package would require several man-months for a useable version to be completed.

VI. RESULTS AND DISCUSSION

In comparing the two alternatives discussed in the previous section, it becomes obvious that in every aspect other than initial cost, IGRIP is better suited than the ASEA/ROBSIM combination for the needs of the RADL. The difference in cost is extremely small when compared to the additional capabilities that can be performed by IGRIP users.

IGRIP offers an additional capability not mentioned as a requirement by the RADL personnel: being able to create application scenarios quickly and easily to sell projects to upper levels of management and other funding bodies. It is true that a picture is worth a thousand words. If you can show the potential funding agency a video of a proposed robot, gripper, or fixture in operation, they will have more confidence and will be more likely to supply the funding.

CONCLUDING REMARKS

In this project two alternatives were compared to find the one which was best-suited for use in the RADL at KSC. It was desired that an integrated environment for off-line programming, debugging, and graphical verification of path planning be developed.

The first alternative, combining the ASEA Off-line Programming package with ROBSIM, had several disadvantages. It was awkward to learn and use, it did not provide collision detection, and it did not provide many of the extra features found in the second alternative. ROBSIM, in its current form, would not run on the Engineering VAX. Extensive modifications would be required to interface it with the ASEA package.

IGRIP, on the other, was found to be user-friendly. It performed all of the required functions except dynamic simulation. This feature could be achieved by purchasing additional software to analyze the dynamics. IGRIP provided better graphics, a modelling environment, and over 40 commercial robots already defined. In addition, translators were available for both the robots in the RADL.

With the additional features provided by IGRIP, it was easier to justify the additional cost. Since, workstations are available, the only additional cost would be that of the software. Therefore, in conclusion, it is recommended that the RADL purchase IGRIP for use as an integrated environment.

REFERENCES

[1] Davis, Virgil Leon, "Systems Integration for the Kennedy Space Center Robotics Applications Development Laboratory," MS87-482 SME Technical Report, 1987.

[2] Off-line Programming: DEC VAX/VMS User's Guide, CK 09-1422E, ASEA Robotics, 1988.

[3] Evaluation of Automated Decisionmaking Methodologies and Development of an Integrated Robotic System Simulation: Appendix A - ROBSIM User's Guide, NASA Contractor Report - 178051, March 1986.

[4] Evaluation of Automated Decisionmaking Methodologies and Development of an Integrated Robotic System Simulation: Appendix A - ROBSIM Programmers Guide, NASA Contractor Report - 178052, March 1986.

N90-16700

1989 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

LOW FLOW VORTEX SHEDDING FLOWMETER

PREPARED BY: Dr. Charles J. Waugaman

ACADEMIC RANK: Assistant Professor

UNIVERSITY AND DEPARTMENT: Pennsylvania State University - Erie
Department of Mechanical Engineering

NASA/KSC

DIVISION: Electronic Systems Support

BRANCH: Instrumentation - Transducer Section

NASA COLLEAGUE: Mr. Robert M. Howard

DATE: July 25, 1989

CONTRACT NUMBER: University of Central Florida
NASA-NGT-60002 Supplement: 2

ACKNOWLEDGEMENTS

I would like to thank all of my co-workers in the Transducer Section for their help and encouragement during the summer project. Special thanks goes to Bob Howard for his encouraging words which brought me to the Kennedy Space Center and for providing a good practical summer project.

ABSTRACT

The purpose of the summer project was to continue a development project on a no moving parts vortex shedding flowmeter used for flow measurement of hypergols. The project involved the design and construction of a test loop to evaluate the meter for flow of Freon which simulates the hypergol fluids. Results were obtained on the output frequency characteristics of the flow meter as a function of flow rate. A family of flow meters for larger size lines and ranges of flow was sized based on the results of the tested meter.

SUMMARY

Currently turbine type flow meters are used to meter the loading of hypergolics into the Space Shuttle Orbiter. Because of problems associated with refurbishment of these meters after each launch, NASA considered the development of a no moving internal parts vortex shedding flowmeter. The University of Florida developed such a flowmeter for 1/2 inch tubing. The objective of the current summer project was to test a modified version of the university prototype and to develope a family of vortex shedding flowmeters for larger line sizes and flow ranges.

In order to test the meter for flow of Freon, which is similar to hypergols, a flow test loop was designed and built. A series of tests were performed on the prototype to evaluate its output characteristics. An alternate pressure transducer was used when the one in the original design failed and could not be replaced or repaired in time. Results of the tests indicate a linear relationship between vortex shedding frequency, as indicated by counting pressure pulsations, and flow rate.

Results of the testing indicate promise for the use of this type of meter for the application mentioned previously. Difficulties were encountered in reliably counting frequencies which are addressed in the report. Recommendations are also made for improving the flow loop for larger flow rates.

Based on the test results obtained, projected characteristics of larger sized flow meters wrer determined and a 1 inch prototype was designed and is being fabricated for future testing.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

1.0 INTRODUCTION

1.1 PROJECT NEEDS.

During the loading of hypergolic fuels and oxidizers, flow meters are used to meter the amount of fluid. The current methods of metering these fluids involves turbine type meters and shuttle-ball type vortex shedding meters. One of the problems that occurs with these meters is that after each launch the meters have to be taken apart and refurbished then recalibrated. The reason for this process is that there are moving parts of the meters in contact with the flowing fluid. An illustration of a typical turbine meter is shown in Figure 1.1. The bushings and bearings of these meters are suceptible to wear especially during the purge phases of the fuel loading process when severe over speed of the roter can occur due to gas flow through the lines. The process of refurbishment of the meters is quite costly due to the techniques required to handle the very toxic hypergols. It is estimated that a savings of about $1000 per meter per launch can be made if the meters do not require this maintenance. The current project involves the development of a family of flow meters which have no moving parts subject to the problems mentioned previously. The next section describes the history of the project prior to the current period of study.

1.2 PROJECT HISTORY.

Several years ago the University of Florida was contracted by NASA to investigate alternative methods of flow measurement which involve no moving parts in the flow stream. The results of the study were presented in the Final Report for Flowmeter and Liquid Level Instrumentation Contract No. NAS 10-10932 dated April 15, 1985.(1) An extensive study of flow measuring techniques resulted in more detailed studies of the vortex shedding types of flow measurement. One of the techniques studied involved the use of fiber optics. The complexities of this method ruled it out as a viable solution. A method of measuring pressure pulses resulting

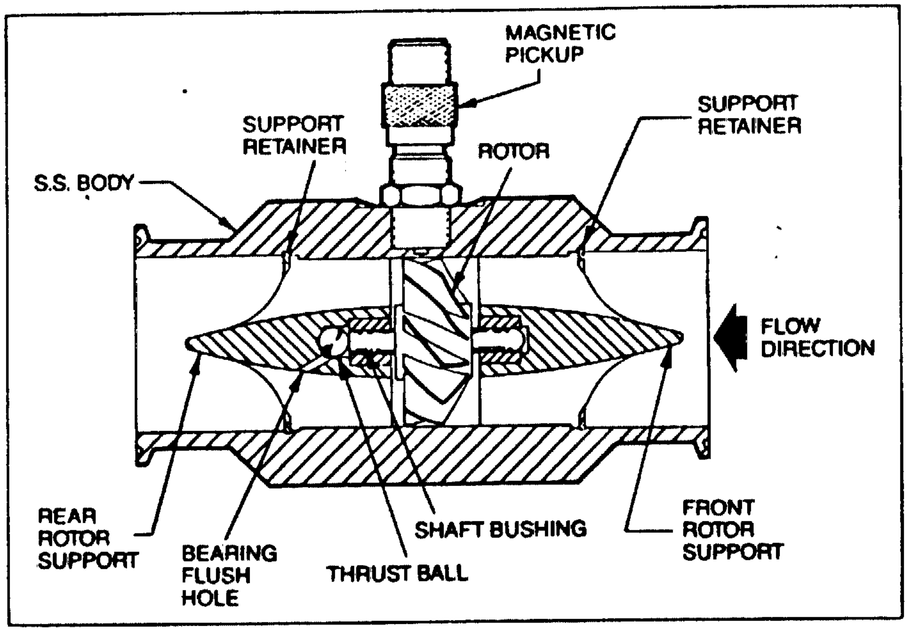


Figure 5: Turbine flowmeter consists of a multiple-bladed, free-spinning, permeable metal rotor housed in a non-magnetic stainless steel body. In operation, the rotating blades generate a frequency signal proportional to the liquid flow rate, which is sensed by the magnetic pickup and transferred to a read-out indicator.

Figure 1.1 Example of a Turbine Meter(Omega Engineering, Inc.)

from the vortex shedding phenomena was pursued and is described in Section 2 of this report.

2.0 BACKGROUND

2.1 VORTEX SHEDDING PHENOMENA.

The phenomena of vortices being shed from a surface in a flowing fluid is not new, and the application of this phenomena to the measurement of flowrate is well established. The 1985 NASA Final Report by the University of Florida (1) contains an extensive literature review of the use of vortex shedding in flow measurement. As a fluid flows over a surface placed in the flow stream vortices alternately spin off of the top and bottom surfaces of the body. This is illustrated in Figure 2.1. For a certain range of fluid velocities the rate or frequency of these vorticies is linear with velocity. The relationship between vortex shedding frequency and velocity is contained in the definition of the Strouhal Number (2).

$$St = \frac{f \times d}{u} \qquad \text{where}$$

f = vortex frequency
d = characteristic dimension
u = fluid velocity

The equation can be rewritten in terms of flow rate and pipe size and is

$$St = \frac{f \times D^2 \times d}{4.90 \times Q} \qquad \text{where}$$

D = pipe diameter (inches)

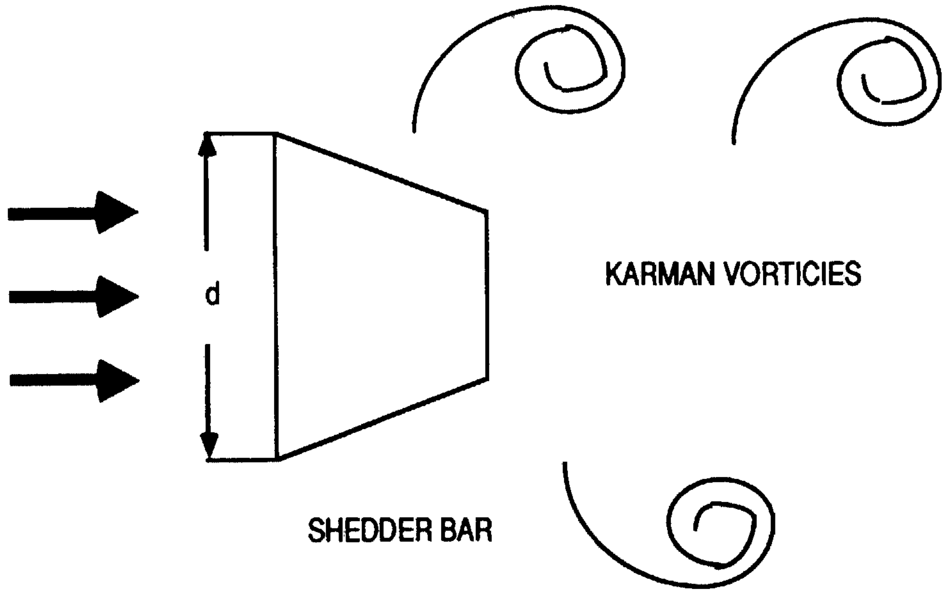


Figure 2.1 Vortex Shedding Phenomena

Q = flow rate (GPM)
d = characteristic dimension (inches)
f = frequency (Hz)

Current manufacturers of vortex shedding flow meters report a linear range of fluid velocities based on Reynold's Numbers in the range of 20,000 to 7,000,000 based on pipe diameter. The Reynolds number can be found using the following relationship.

$$Re = \frac{3160 \times Q \times G_t}{D \times CP} \quad \text{where}$$

Q = flow rate (GPM)
G_t = specific gravity
D = pipe inside diameter (inches)
CP = viscosity (cp)

Ranges of Reynolds numbers for the flow meter considered in this project and for the predicted family of meters are presented in Section 4 of this report.

The geometry of the shedder bar determines the characteristic dimension of the shedder which then determines the frequency of vorticies. Several shapes of the shedder bar were studied by the University of Florida in determining the best shape for the flow meter developed. A description of this shedder is presented in the next section. The most common shedder bar shape is the reversed wedge shown in Figure 2.1.

There are several methods used to detect and measure the vortex frequency. Currently NASA uses Eastech vortex meters which incorporate a shuttle ball in the shedder bar which oscillates with the alternating vorticies. A reluctance pick up emits a pulse rate proportional to shuttle ball oscillation which then correlates to flow rate. Another method is illustrated in Figure 2.2 which shows vibration of the shedder bar sensed by a piezoelectric element. The

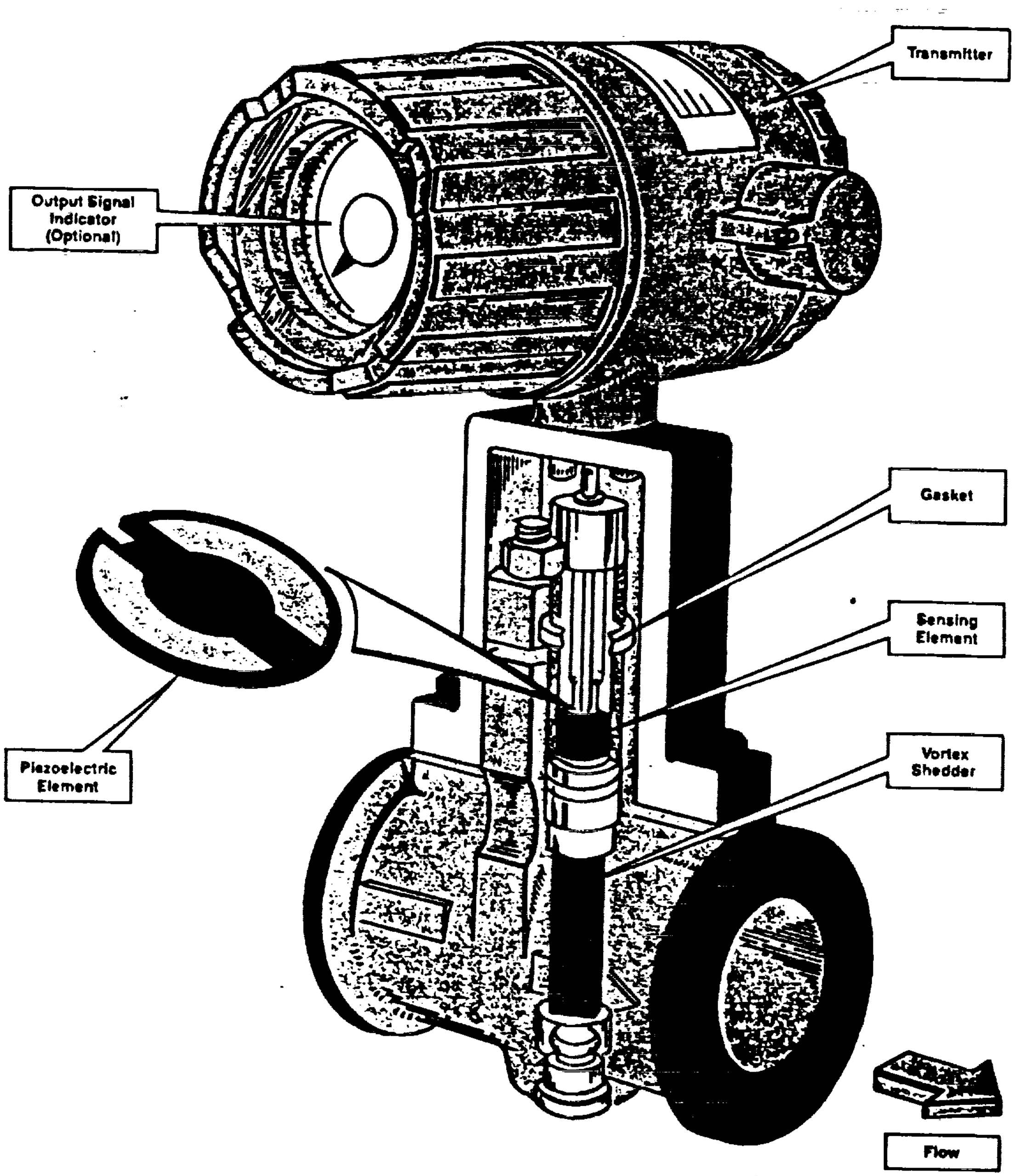


Figure 2.2 Example of a Vortex Shedding Flowmeter (Omega Engineering, Inc.)

ORIGINAL PAGE
BLACK AND WHITE PHOTOGRAPH

meter considered in this project uses a pressure transducer which detects pressure pulses associated with the vorticies. A description of this meter is included in the next section.

2.2 UNIVERSITY OF FLORIDA DESIGN.

Upon completion of the alternative flow study by the University of Florida, the university was contracted by NASA to develope a working prototype vortex shedding flow meter. The details of the university project are described in the Final Report Vortex Shedder Flow Meter NASA Contract No. 10-11230 December 31, 1986 (4) and in a technical paper (5). Three prototypes of the meter developed are the basis for the study reported herein.

The university design resulted in a flow meter suitable for 1/2 inch tubing having a usable range of 1.5 to 15 GPM. There was considerable effort by the university to optimize various aspects of the flow meter design. The final design is shown in Figure 2.3. The shedder bar has a rectangular cross-section shown in Section B-B. Pressure pulses resulting from the vorticies are are transmitted through the three holes below the shedder bar to a cavity in which is placed a Kistler Model 206 pressure transducer. The pressure signal is then conditioned with a Kistler Model 5116 Coupler and then sent to a frequency counter and an oscilloscope. The university prototype has male flare tubing ends for insertion into the flow lines. An identical meter was constructed at KSC for testing purposes. Also, a modified meter was built at KSC which incorporates female boss threads in the meter body which accepts a boss to male flare tube adapter. This modification protects the meter body in case of male tubing thread damage which did occur to the university built prototype.

Extensive testing was done by the University of Florida on their design using water as the working fluid. A comparison of the published university results with the tests conducted in this project are presented in Section 3 of this report.

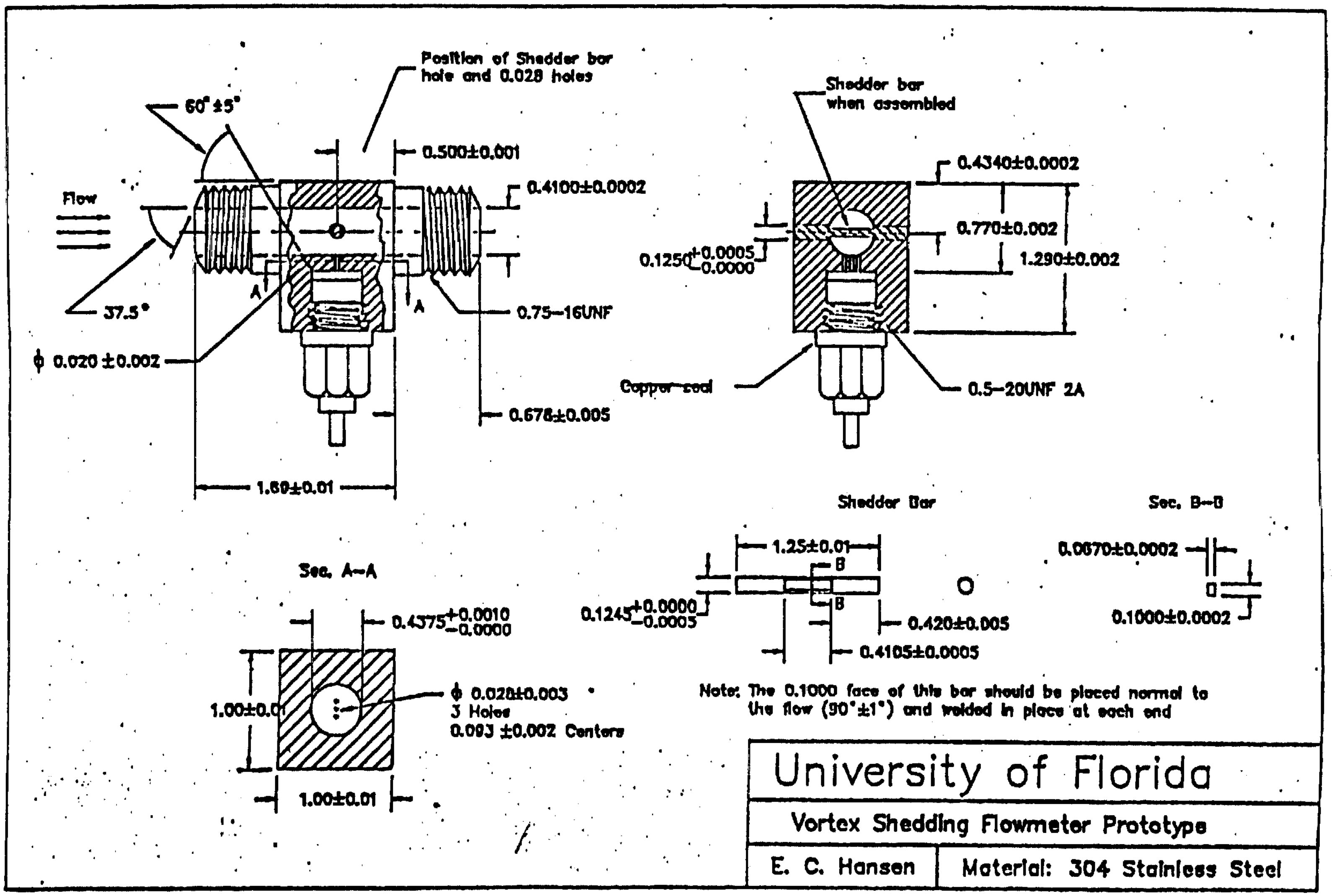


Figure 2.3 University of Florida 1/2 Inch Prototype

3.0 FLOW LOOP DESIGN AND TEST

3.1 INTRODUCTION.

The basic goals of the current project were to design and construct a flow bench to test the University of Florida design and to determine the appropriate size flowmeters for larger flow values. The flow meter sizing is discussed in Section 4 of this report. The test considered in this section is on the KSC modified prototype which is similar to the U of Fla. model and involves Freon 113 (Trichlorotrifluoroethane) as the working fluid. Freon was chosen as it has similar properties to the hypergolic fluids which will be metered by the flowmeter considered in this project.

3.2 FLOW BENCH DESIGN.

One of the goals of the summer project described in this report was to design and build a flow bench to test the 1/2 inch prototype flow meter. Due to the short time period involved, it was decided to utilize parts that were available on site. Pumps were the first consideration to circulate the fluid, however because of unavailable suitable pumps with desired pressure and flow characteristics an alternate method was selected. Results reported by the University of Florida indicated problems occured when pressure upstream of the vortex meter dropped below certain levels causing what was thought to be cavitation. Since pressure was a factor, a source of gaseous nitrogen was used as the driving pressure for flow. A similar technique was done at NASA-Johnson for flow measurement.(6)

The flow loop which was designed and built is shown schematically in Figure 3.1. Initially two liquid nitrogen dewars were used as containers for the freon fluid. One dewar was located outside of the laboratory window and was connected to the flow loop through the window penetration indicated Dewar #2. The second dewar was located inside the lab and was placed on a load platform used to

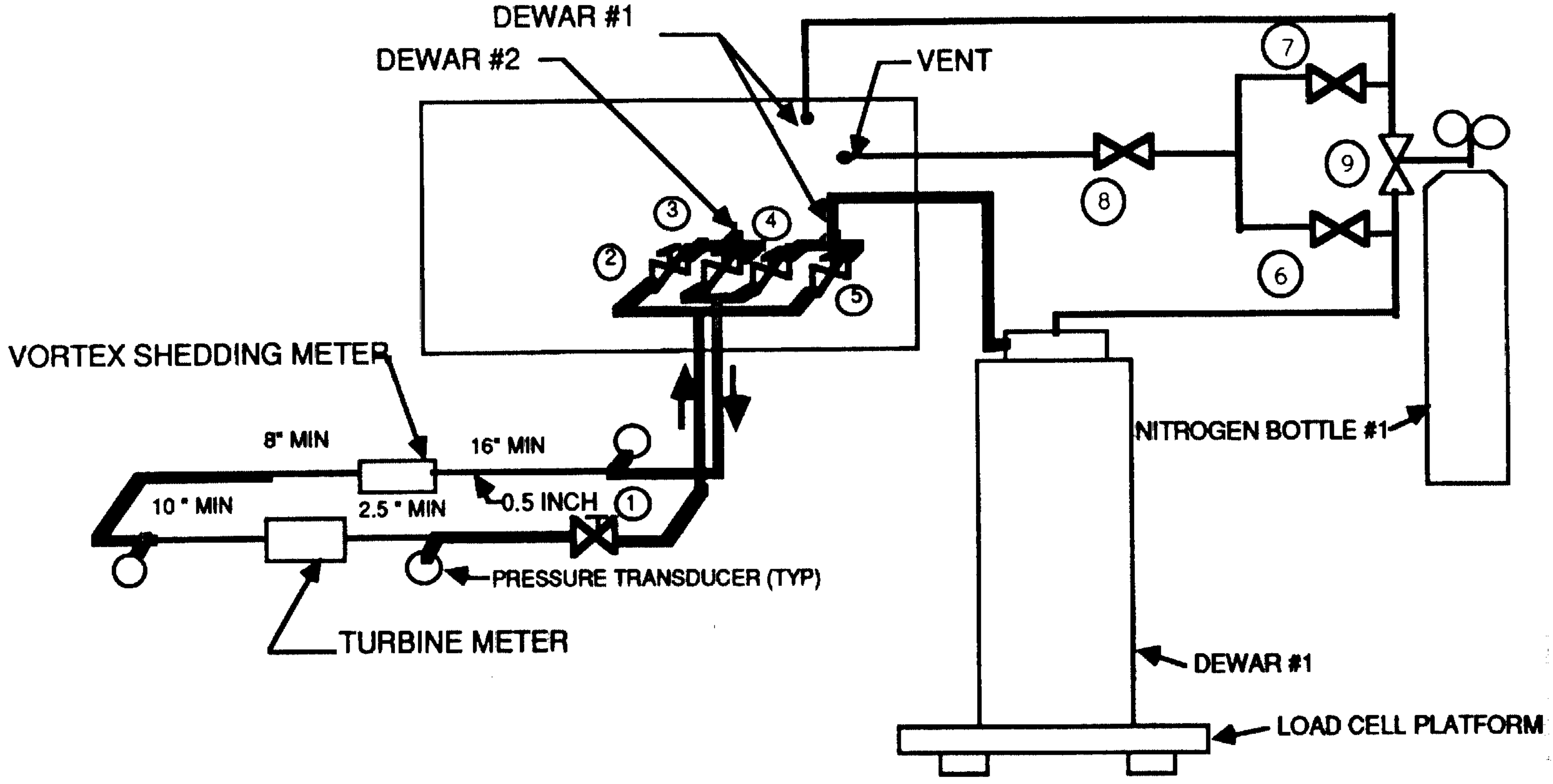


Figure 3.1 FLOW TEST SCHEMATIC

measure the weight change of the freon. By proper adjustment of the system valving fluid could flow through the loop from either dewar. If pumping from Dewar #1 to Dewar #2, Dewar #1 is pressurized with nitrogen gas and Dewar # 2 is vented as a receiver. A more detailed procedure is included in Section 3.3 and in the Appendix. The quantities measured in the loop include the outputs from the turbine meter used as a standard, from the vortex meter under test, from pressure transducers at various locations on the test loop, and from the load cell transducers located under Dewar #1. The vortex meter output signal is a pressure detected by a Kistler Model 206 pressure transducer as shown in Figure 2.3. The Kistler transducer is a piezoelectric type of device and comes with a signal coupler which produces an AC coupled millivolt output proportional to pressure fluctuations. During some of the modifications that were made to the prototypes the Kistler transducer was damaged beyond repair and could not be replaced before the conclusion of this summer project. As an interim solution to the problem of pressure measurement an Entran transducer was adapted to the flow meter to check out the flow loop. The Entran sensitivity was much less than that of the Kistler so considerable signal conditioning steps were taken to produce a measurable output. A schematic of the amplifier/filter circuit is shown in Figure 3.2. The basic experimental tests that were conducted are described in the following section.

3.3 FLOW METER TESTING.

The basic procedure used to produce the results obtained is outlined in more detail in the Appendix. Depending on whether flow was comming from Dewar #1 or Dewar #2 valves 2, 3, 4 and 5 were in the open or closed position.(See Figure 3.1) To vary the flow rate more or less pressure was applied to the full dewar using the three-way valve connected to Nitrogen Bottle #1. By opening gate valve #1 to the maximum open position allowing the flow rate to be governed by the nitrogen pressure. Steady flow was indicated by a steady output reading from the calibrated turbine meter in series with the

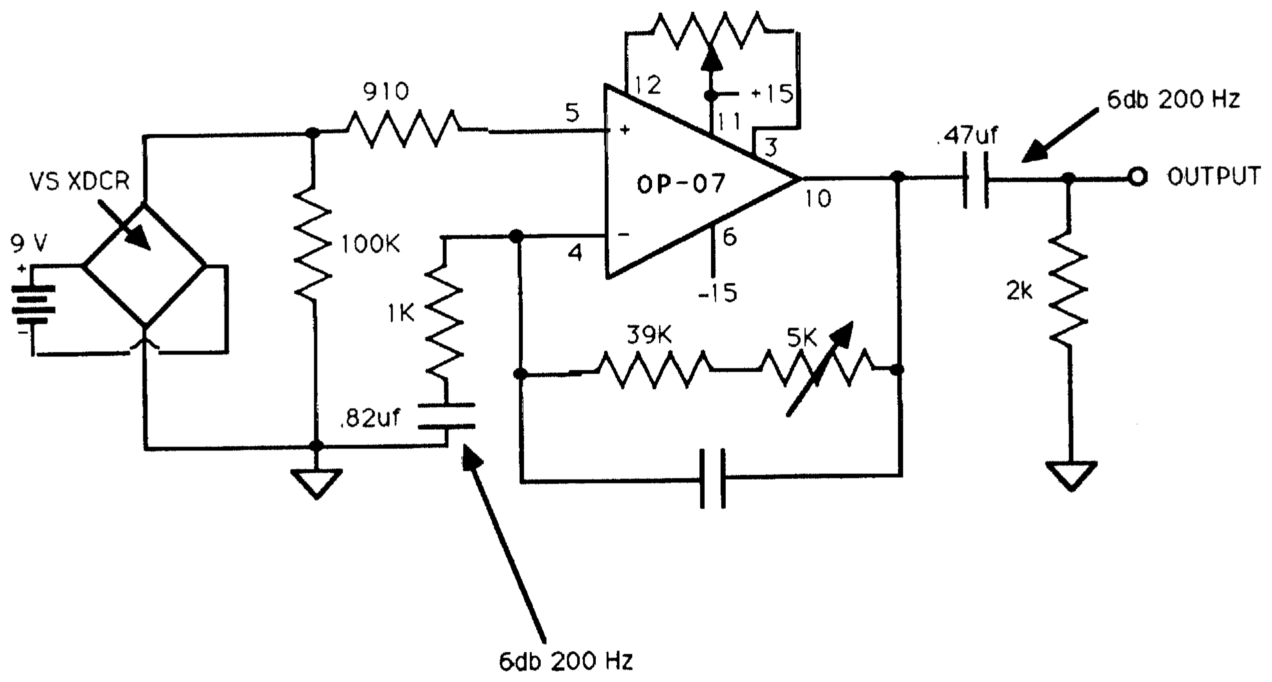


Figure 3.2 Signal Conditioning Amplifier for Temporary Pressure Transducer

vortex meter. The turbine meter output consists of a voltage from the signal conditioner which is linearily proportional to the flow rate. The calibration curve relating gallons per minute(GPM) to output voltage is presented in Figure 3.3. Calibration of the turbine meter was performed at the KSC flow calibration facility. Once the Dewar Load Platform data is incorporated into the data acquisition system, mass flow rate can be measured through an increase or decrease of the fluid in the dewar. As a rough check of the turbine output, a timed change in dewar weight was performed at one flow rate. Since load cell output was indicated at one pound increments determination of start time and stop time weights was suceptable to plus or minus one pound error on each of the three load cells. This uncertainty along with freon density uncertainty the result of this rough test compared favorably with the turbine meter output.

The output from the vortex shedding flow meter was measured using a frequency counter and an oscilloscope. Even with amplification, the output signal from the Entran pressure transducer was quite small although it was measurable with the frequency counter and visible on the oscilloscope. Up to 7 GPM the readings were consistent between the two methods, however above that flow rate flow noise or turbulence produced a wandering of the pressure pulsations which resulted in some pulses not being counted. Oscilloscope traces however could be counted and produced reasonable results. The oscilloscope used had a storage capability which permited freezing of the trace for easier pulse counting and frequency determination. Recommendations concerning this problem are presented in Section 5.

A problem that was encountered using the inside-outside dewar system was a limited flow capacity of the system. The dewar bottles have a small liquid line size which placed a restriction on the system due to flow out of one and into another. With the maximum desirable pressure in the outflow dewar of 150 psi flow was limited to approximately 10 GPM. A test was performed which permited flow directly from one dewar into the other without going through the flow loop. Flow was measured using a 3/4 inch

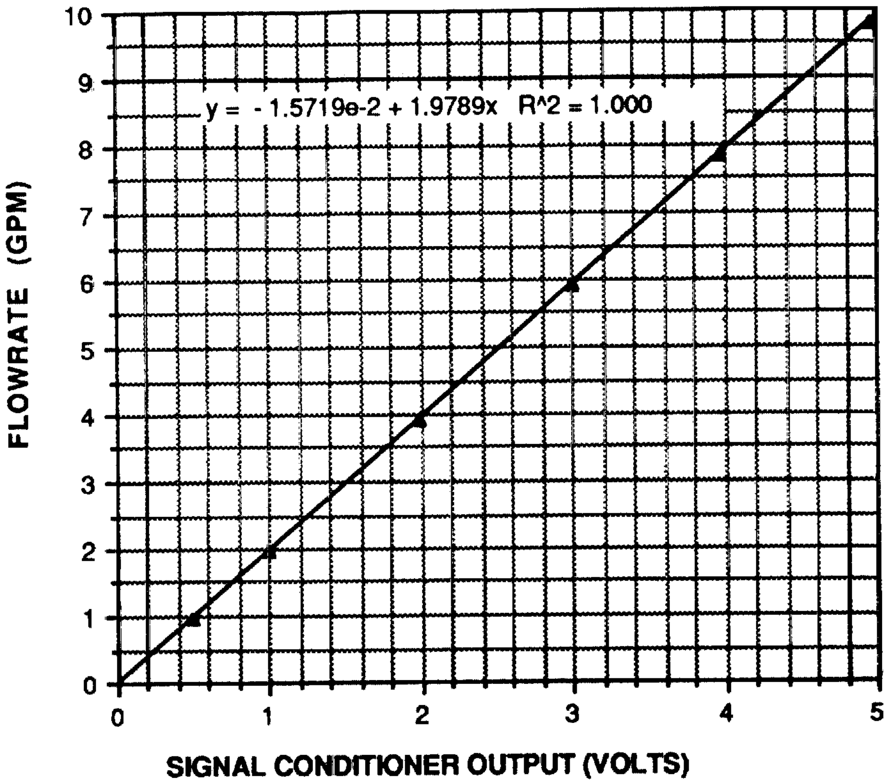


Figure 3.3 1/2 Inch Turbine Meter Calibration Results

calibrated turbine meter and was found to be not much more than the maximum found through the loop. In an attempt to increase the flow capacity, the outside dewar was brought inside and connected in parallel with the inside dewar. The outflow from the system was put into a vented 55 gallon drum which was placed outside. To refill the dewars the fluid was forced back inside using a small pressure applied to the drum. This limits the flow to one direction only, however even with the small driving pressure on the drum, flow could be circulated through the loop at a rate of 4 GPM. With the two dewars in parallel a flow rate of 12.5 GPM was achieved. Suggestions to improve the flow capacity are presented in Section 5.

Another test performed involved the measurement of pressure drops through the flow loop at higher flow rates. At a flow rate of approximately 11.5 GPM a pressure drop of about 36 psi occured through the 1/2 inch tubing including the vortex meter. Reduction of the length of 1/2 inch tubing upstream and downstream of the vortex meter will help. Also a similar pressure drop occured through the turbine meter and associated 1/2 inch tubing. This can be reduced by using the 3/4 inch meter and tubing.

3.4 RESULTS AND DISCUSSION.

The experimental results consists of a determination of vortex shedding frequency versus flow rate which is shown in Figure 3.4. As can be seen from the diagram, the relationship between vortex frequency and flow rate is linear. There is some scatter of the data at flow rates above six gpm which can be attributed to the method of determining the frequency by pulse counting an oscilloscope trace as discussed previously. The pressure signal produced from the Entran transducer required considerable amplification and filtering to make it measurable. At that, the output was very repeatable at lower flow rates where it could be counted reliably using a frequency counter. It is anticipated that the Kistler pressure transducer will produce a better output signal due to its greater sensitivity.

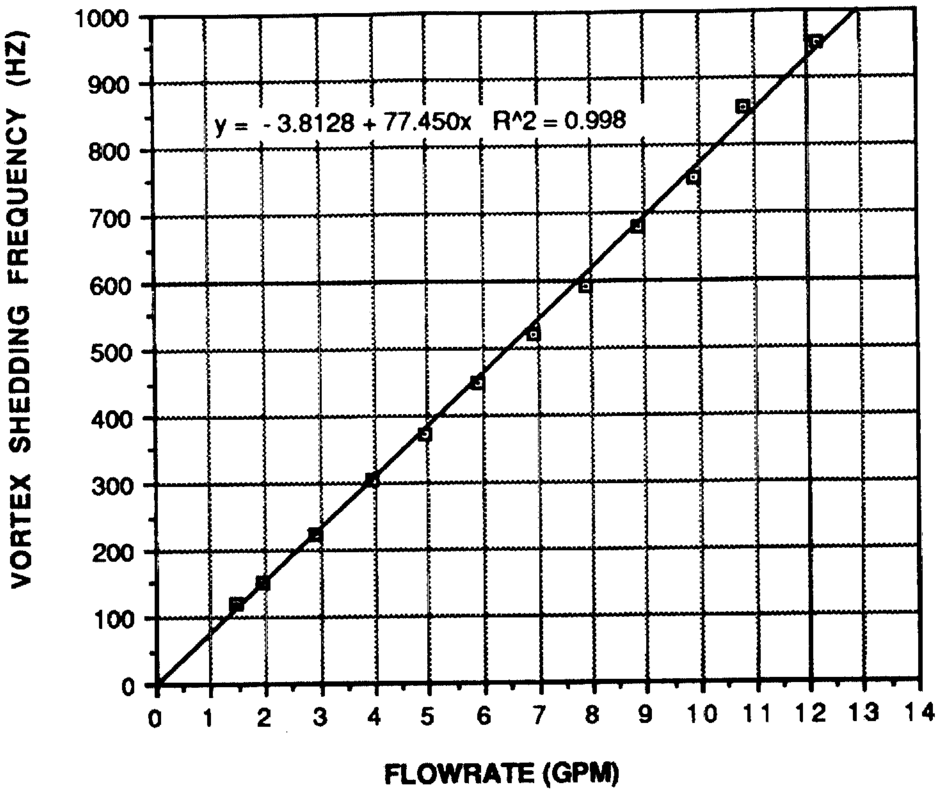


Figure 3.4 Test Results for KSC Modified 1/2 Inch Vortex Shedding Flowmeter Using Freon

Data from the tests performed by the University of Florida was plotted and is shown on Figure 3.5. The results also show a linear relationship between vortex frequency and flow rate. The university results are for water as the flowing medium. Table 3.1 shows a comparison of results from the university tests and those done under this project. A similar range of flow rates was used in both studies producing similar velocities. The Reynolds numbers involved were higher for the Freon test due to the increased density of Freon compared to water and due to the lower viscosity of Freon. The importance of this higher Reynolds number for a given velocity is that the useful range of vortex shedding can be extended since it is tied to a lower allowable value of Reynolds number.(3,1) A further discussion of this is found in Section 4 concerning vortex meters for larger flow rates. Another difference between the university results and those from the KSC test involved the Strouhal Number. In Section 2.1 it was shown that the dimensionless Strouhal number is proportional to vortex shedding frequency and characteristic dimension and is inversly proportional to fluid velocity. The university results have a Strouhal number less than that of the KSC results. In both cases the values of the Strouhal numbers are constant, again indicating the linear relationship between frequency and flow rate. It is not known why there is a difference in Strouhal numbers, however the tests were conducted on different meters and involved different fluids. The important point is that in either case the Strouhal number is a constant which permits a count of vortex frequency to be a linear measure of flow rate.

4.0 FLOW METER SIZING FOR LARGER FLOWS

The geometry of the shedder bar in a vortex shedding meter is related to the inside diameter of the pipe. It has been found empirically that the ratio of of characteristic dimension, A, to pipe diameter, D, should be in the range of 0.15 to 0.40(1). The university design uses an A/D ratio of 0.24. Also the ratio of the shedder bar dimension in line with the flow, B, to the characteristic dimension should be in the range of 1.0 to 2.0. The university design for this ratio is B/A = 0.67. Holding these ratios constant and using the

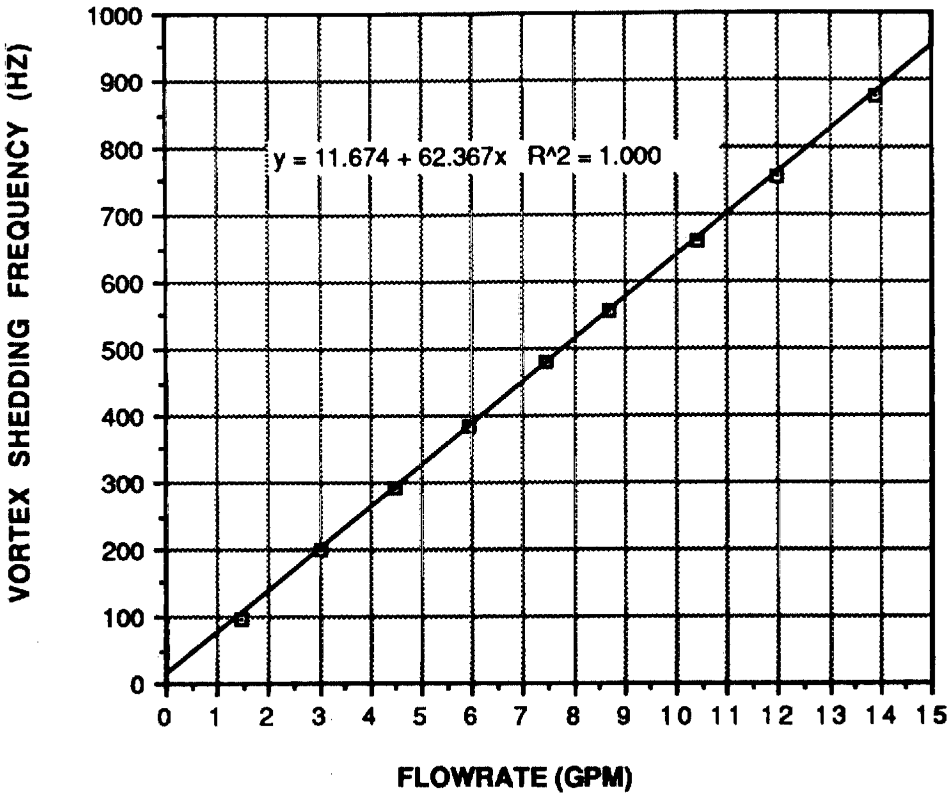


Figure 3.5 Test Results for 1/2 Inch Vortex Shedding Flow Meter Using Water (Univ of Florida)

Table 3.1 Comparison of Test Results of University of Florida and KSC Modified Prototype Vortex Flowmeters

	UNIV FLA PROTOTYPE	KSC MODIFIED PROTOTYPE
TYPE OF FLUID	WATER	FREON 113 (TRICHLOROTRIFLUOROETHANE)
RANGE OF FLOW (GPM)	1.5 - 13.9	1.5 - 12.2
RANGE OF VORTEX SHEDDING FREQUENCIES (HZ)	97 - 876	112 - 952
RANGE OF REYNOLDS NUMBERS	11,600 - 107,000	26,600 - 525,000
RANGE OF FLUID VELOCITIES (FT/SEC)	3.6 - 33.8	3.6 - 29.6
AVERAGE STROUHAL NUMBER	0.22	0.26

inside diameters for larger size tubing a table of flow meter properties for larger sizes is shown in Table 4.1. The table shows the anticipated flow range for several sizes larger than 1/2 inch. Also the scaled shedder bar size is indicated along with the projected vortex shedding frequency range. The frequency range is based on a Strouhal number of 0.26 which was found experimentally for the modified KSC prototype. The range of velocities and Reynolds numbers for Freon are included. Because the velocity range for all of the sizes considered is essentially the same, the fact that the characteristic dimension for the larger size meters is also larger results in a smaller range of vortex shedding frequencies. This does not present any difficulty in the counting of the frequencies, however it does mean that the measured frequency is closer to extraneous noise such as flow turbulence associated with piping fittings and also with electrical noise affecting the output signal. A further discussion of this is in Section 5.

A flow meter for a 1 inch application was designed and is shown in Figure 4.1. The meter is scaled up from the modified KSC prototype design which has internal boss thread on the meter body to accept boss to flare tubing end connections. The meter uses the same pressure transducer as that used in the 1/2 inch model. A prototype of this meter is currently being fabricated at KSC however it will be tested after this phase of the project is completed. The new prototype will be made from aluminum to ease in the construction however the final design will be stainless steel. The 1/2 inch meter has a hole drilled at an angle from the front of the pressure sensor port into the flow path.(See Figure 2.3) This hole was incorporated by the university to aid in purging trapped air from the pressure transducer port. The 1 inch design does not have this angled hole, for the purpose of determining whether it is really required. If necessary it can be readilly added. Another modification to the meter design which would allow investigation of the orientation of the shedder bar would be to flute the .250 inch round part of the shedder bar and also the hole through the meter body. This will allow rotation of the shedder bar within the flow stream. A means

Table 4.1 Projected Vortex Shedding Flowmeter Characteristics

	1/2 " D = 0.410	3/4 " D = 0.609	1 " D = 0.844	1 1/2 " D = 1.312	2 " D = 1.781
FLOW RANGE (GPM)	1.0 - 10.0	2.5 - 25.0	6.0 - 60.0	15.0 - 150.0	32.0 - 320.0
SHEDDER BAR SIZE (INCHES)	.100 x .067	.149 x .100	.206 x .138	.320 x .214	.435 x .291
VORTEX FREQUENCY RANGE (HZ) BASED ON ST = 0.26	76 - 758	58 - 576	52 - 520	35 - 347	30 - 295
VELOCITY RANGE (FT/SEC)	3.6 - 36.5	2.8 - 27.5	3.4 - 34.4	3.6 - 35.6	4.1 - 41.2
REYNOLDS NUMBER RANGE	26,600 - 266,000	29,900 - 299,000	51,700 - 517,000	83,200 - 832,000	131,000 - 1,310,000

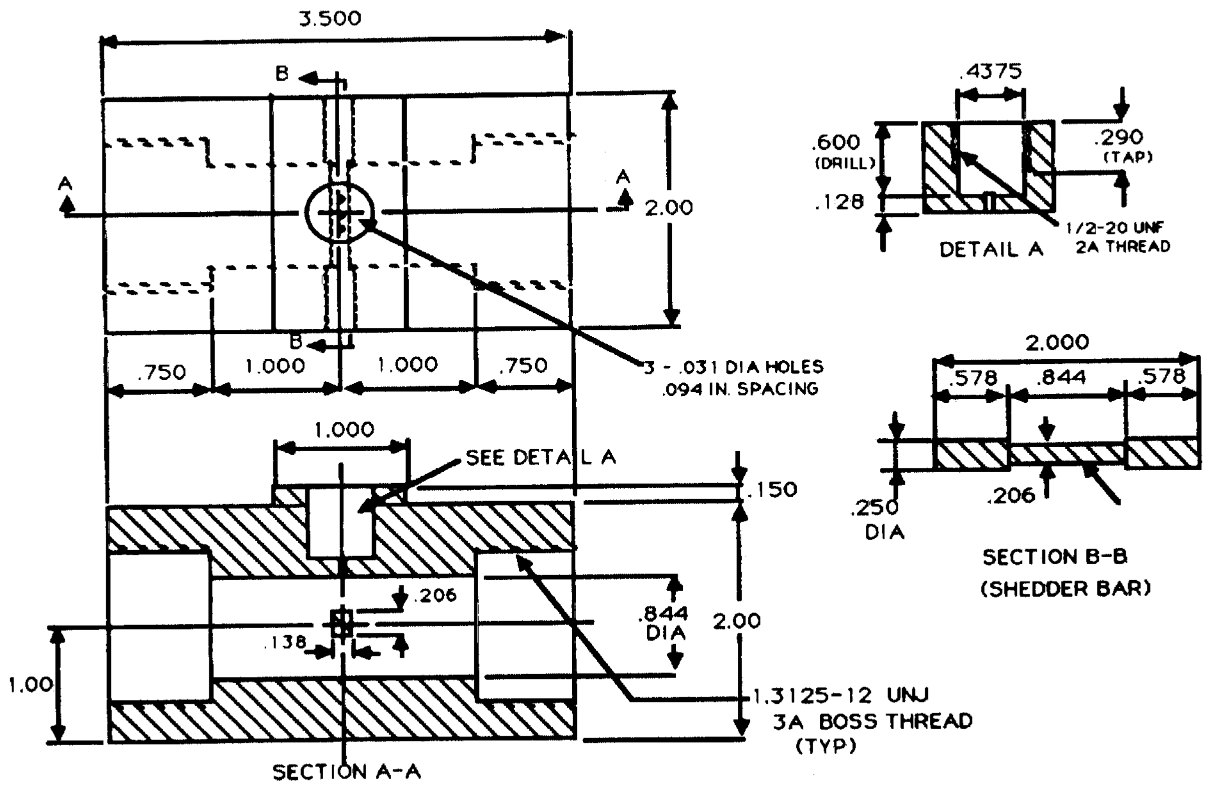


Figure 4.1 One Inch Vortex Shedding Flowmeter Prototype

of sealing the bar will be necessary such as recessing the shedder bar in the meter body and providing a sealed plug on the outside of each side. To test the 1 inch meter in the existing flow loop only the lower end of the 6 - 60 gpm flow range can be achieved. Some modifications to improve flow capacity are discussed in Section 5.

5.0 CONCLUSIONS

A few conclusions can be drawn as a result of the preliminary tests conducted on the KSC modified vortex shedding flowmeter prototype. These conclusions are generally in the form of recommendations for further study for the project.

Since the test data obtained for this project was for a flowmeter outfitted with a temporary pressure transducer in place of the Kistler Model 206, final judgement on the functionality of the meter should wait until tests can be run on the flowmeter with the Kistler installed. A similar set of flow rates can be achieved with the test flow loop as outlined in Section 3.3 and in the Appendix. Based on a preliminary test involving the Kistler transducer, the output should be much stronger than that from the Entran sensor which will result in better signal conditioning results.

Another difficulty encountered in the tests of this study involved flow noise in the forms of turbulence caused by line fittings and low frequency static pressure changes. Both of these occurances resulted in missed counts by the frequency counter. Overcomming this is critical to the use of this flowmeter due to the frequency being a measure of the flow rate. As part of the suggested solution a spectrum analysis of the output signal can be made which will identify frequency components of the signal. Hopefully the largest frequency component will be the vortex shedding frequency and the remaining frequencies will be noise. A first attempt at noise elimination should be through electronic filtering of the output signal. This may involve filters for each flowmeter size due to the different ranges of vortex shedding frequencies. Other solutions may involve flow line modifications. Restrictions on meter placement in the actual application may affect this solution. Flow line remedies may include some form of fluid accumulator similar to the one placed on the flow loop. Priliminary tests with this accumulator showed some promise in removal of some flow noise. Also, the use of flow straighteners in the meter body may improve the situation.

Some recommendations concerning the constructed flow loop are in order. As mentioned previously the loop was designed using available components rather than optimum ones. For a more permanent design certain parts should be changed. The two tank system with nitrogen gas as the driving force worked very well, however the restriction to flow rate placed on the system by the nitrogen dewars is too great. For flow rates larger than 15 gpm different tanks should be used. Tanks could be constructed capable of 200 psi internal pressure which have sufficient volume and flow inlet and outlet line sizes to permit the desired flow rate. Immediate solutions to the existing flow restriction problem involve the elimination of all 1/2 inch lines including upstream and downstream of the vortex meter, and replacement of the 1/2 inch turbine meter with a 3/4 inch meter which is calibrated over the range 1.5 - 32 gpm. The arrangement of the dewars into a parallel output arrangement helps to increase the capacity of the current system and should be used as such. More Freon should be ordered, and the outside storage drums should be connected in parallel for collection of the additional fluid.

A data acquisition system is proposed for the flow test loop. A Macintosh II computer using Labview software will monitor and analyze data from the loop. This will permit the recording an analysis of several simultaneous pieces of information from the loop including turbine meter output (voltage proportional to flow rate), vortex frequency count, pressures throughout the system (voltage proportional to psi), and change in dewar weight (voltage proportional to load or weight). The Labview software will permit plotting of results and perhaps frequency spectrum analysis.

A vortex shedding flowmeter which senses pressure pulsations produced by vorticies appears to be a viable solution to the no moving parts flowmeter. Further testing will determine whether the output signal count will present major problems.

APPENDIX

DATA TABLES

Data for KSC Prototype Test

Flow Rate GPM	Vortex Frequency HZ	Tank Press psi
1.460	120	xxxx
1.960	153	xxxx
2.950	226	xxxx
3.940	304	16.5
4.930	373	26.4
5.920	450	38.1
6.910	521	58.5
7.900	591	66.0
8.890	682	75.0
9.880	750	96.0
10.810	854	108.0
12.200	952	138.0

TEST PROCEDURES

Flow Out of Dewar #1 Into Dewar #2 Governed By Tank Pressure
Refer to Figure 3.1 in Section 3.2

1. Fill Dewar #1
2. Close all valves.
3. Open valves 1, 2, 7 and 8.
4. Open three way valve #9 to apply pressure to Dewar #1.
5 Adjust regulator on nitrogen bottle to desired tank pressure. See

data table above for approximate governing pressure for a given flow rate.

6. Open valve 4 to commence flow. Regulator valve can be adjusted to increase flow rate and vent valve 6 can be opened to decrease flow.
7. Steady flow is achieved when output voltage from turbine meter signal conditioner is constant.
8. Record sensor outputs and frequency counter output.
9. Close valve 4 to stop flow.
10. If additional fluid remains in Dewar #1 to run another flow rate, repeat steps 5 to 9. otherwise follow next procedure.

Flow Out of Dewar #2 Into Dewar #1

1. Close all valves.
2. Open valves 1, 5, 6, and 8.
3. Open three way valve 9 to apply pressur to Dewar #2.
4. Adjust regulator to desired tank pressure.
5. Open valve 3 to begin flow. Increase flow with regulator, decrease by opening vent valve 7.
6. Repeat steps 7 and 8 above.
7. Close valve 3 to stop flow.

Flow Out of Two Dewars in Parallel

1. With both dewars full close all valves.
2. Repeat steps 3 to 9 from Dewar #1 to Dewar #2 procedure.
3. When all fluid is out of the dewars refill from outside storage drums.

Flow From Storage Drums Into Dewars

1. Close all valves.
2. Open three way valve 9 to apply pressure to the storage drum.

Use no more than 20 psi on the drum.

3. Open valves 6 and 8 to vent dewars.

4. Once dewars have reached atmospheric pressure open valves 2 and 5 to begin the flow from the drum into the dewars.Flow can be directed through the flow loop by opening valves 1, 3, and 5 instead of 2 and 5.

3. Close all valves when dewars have been filled.

REFERENCES

1. Farber, E.A., et.al.,"Flowmeter and Liquid Level Instrumentation," Final Report for NASA Contract Number NAS 10-10932, April 15, 1985.

2. Holman, J.P., Experimental Methods for Engineers, Fifth Edition, McGraw Hill Book Co., New York, 1989.

3. "Omega Complete Flow and Level Measurement Handbook and Encyclopedia," Omega Engineering, Inc., Volume 26, 1989.

4. Farber, E.A. and Hansen, E.C.,"Vortex Shedder Flow Meter," Final Report for NASA Contract Number NAS 10-11230, December 31, 1986.

5. Hansen, E.C. and Restrepo, J.A.,"Development of a Small Vortex Shedding Flowmeter for Hypergolic Propellants," AIAA,ASME,SIAM, and APS First National Fluid Dynamics Congress, Cincinnati, OH, July 25-28, 1988, AIAA Paper 88-3602, p. 1500-1504.

6. Baird, R.S.,"Flowmeter Evaluation for On-Orbit Operations," NASA Technical Memorandum 100465, August, 1988.

N90-16701

1989 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

MATHEMATICAL MODEL FOR ADAPTIVE CONTROL SYSTEM OF ASEA ROBOT AT KENNEDY SPACE CENTER

PREPARED BY:	Dr. Omar Zia
ACADEMIC RANK:	Associate Professor
UNIVERSITY AND DEPARTMENT:	California Polytechnic State University Engineering Technology Department
NASA/KSC	
DIVISION:	Mechanical Engineering
BRANCH:	Special Projects (RADL)
NASA COLLEAGUE:	Mr. V. Leon Davis
DATE:	August 23, 1989
CONTRACT NUMBER:	University of Central Florida NASA-NGT-60002 Supplement: 2

ABSTRACT

This paper discusses the dynamic properties and determines the mathematical model for the adaptive control of the robotic system presently under investigation at Robotic Application and Development LAboratory at Kennedy Space Center.

NASA is currently investigating the use of robotic manipulators for mating and demating of fuel lines to the Space Shuttle Vehicle prior to launch. The Robotic system used as a testbed for this purpose is an ASEA IRB-90 industrial robot with adaptive control capabilities. THe system was tested and it's performance with respect to stability was improved by using an analogue force controller.

The objective of this research project is to determine the mathematical model of the system operating under force feedback control with varying dynamic internal perturbation in order to provide continuous stable operation under variable load conditions. A series of lumped parameter models are developed. The models include some effects of robot structural dynamics, sensor compliance, and workpiece dynamics.

SUMMARY

The Robot Application and Development Laboratory at Kennedy Space Center has been tasked to address the unique needs of the center in preparing, ground servicing and launching the nation's spacecraft.
Unlike industrial applications, these are not monotonous repetition of relatively simple tasks but occasional/intermittent performance of very sophisticated tasks. To achieve the goal, Robotic Application Laboratory has put together a state of the art robotic system which provides an excellent and easy to use testbed. The goal is to provide an experimental testbed to examine possible robotic solutions for a wide variety of tasks which might benefit the center in terms of safety, quality, reliability or cost saving.

Mating and demating of umbilical fuel lines for the main tank of the Space shuttle vehicle is one area that Robotic Application and Development Laboratory is working on. In order for a robot to accomplish the task of umbilical mating the following three distinct phases must occur.

o Vision tracking must take place to allow the robot to approach and track the umbilical socket.

o The second phase is the actual mating process to occur which require a combination of mechanical guidance, compliance and active force feedback .

o The last phase happens when a solid mating has occurred. This is the most critical part of the process where the random motions of the Space Shuttle Vehicle has to be duplicated by the robot using a force feedback approach to avoid large contact forces.

Initial experimental tests had indicated that the existing robotic system had tendency of becoming unstable while following the random motions of the Space Shuttle Vehicle simulator. This problem was investigated thoroughly in the summer of 1988.
The cause of the problem was traced (240 msec time delay in the adaptive control path). An alternate method of implementing force control to provide proof of concept to avoid time delay was developed. The goal in this research project is to determine the mathematical model of the system . The closed loop performance of the system has been observed in the laboratory to be stable and satisfactory for most applications. The particular properties of the system that can lead to instability and limit performance has been discussed. A series of lumped parameter models are developed in an effort to predict the closed loop dynamics of force controlled arm. While experimental tests indicated the computational time delay to be the main source of instability, qualitative analysis shows that the robot dynamics can have significant contribution to the system's instability.

TABLE OF CONTENTS

1. INTRODUCTION

Motion of robots can be accurately described by coupled sets of highly nonlinear ordinary differential equations. Closed form analytical solutions for these equation are not easily available. Physically the coupling terms represent gravitational torques, which depend on positions of the joints; reaction torques, due to acceleration of other joints; and of course coroilis and centrifugal torques. The magnitude of these interaction torques depends on the physical characteristics of the manipulator and the load it carries.

The effects mentioned above complicates the task of accurately determining model of the system. Therefore simple tasks like inserting a peg in the hole as well as complicated ones like following the random motions of flight simulator must be broken down into subtasks. Much work has been done by many researchers on the subject of force control for robotic manipulators [1], [2], [3], [4], [5]. One of the problems confronting anyone trying to assimilate this information is that there seem to be as many different techniques and models for force control as there are researchers in the field. After reviewing many of these results, I have attempted to come up with an approximate model for the system under investigation in Robotic Application and Development Laboratory at Kennedy Space Center.

While my main goal is to discuss force control models it should be noted that a force controller must always be used in conjunction with a position controller. Most commonly one wants to specify force control only along selected cartesian degrees of freedom while the remainder are controlled according to position trajectory.

2. FORCE CONTROL, GENERAL CONSIDERATIONS

In general if we put the issue of coordinate transformation aside for the moment, each axis of a force controlled arm can be viewed as a single input (the motor), dual output (position sensor and force sensor) system. The method by which the signals are processed and feedback to the motor determines the performance characteristics of the servo loop. Although it is impossible to make an unequivocal classification of all force servos, it is possible to group most algorithms into three broad categories: torque based, velocity based or position based. This classification is based upon the concept of successive loop closure, that is, closing an inner loop on one sensor and then closing an outer loop using another sensor.

In general the situation is illustrated in Figure 1, showing the sensor signal being processed along with command input, to form a corrective command for the manipulators motors. This model is appropriated for most electric arms where the basic control variable is motor torque. However, it is possible to have three different situations:

o The force sensor signal is processed to become a torque command.
o The force sensor signal is processed to become a velocity command to a inner velocity loop or
o the force signal is processed to become a position command to a inner position loop .

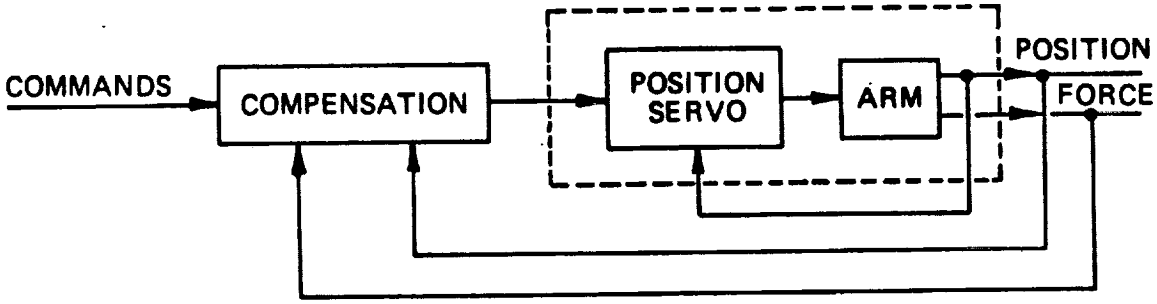


Fig. 1. Generalized force feedback servo with inner position loop

2.1 ADAPTIVE CONTROL FEATURE OF RADL SYSTEM

The problem of self-adjusting the parameters of a controller in order to stabilize the dynamic characteristics of a process, when the plant parameters undergo large and unpredictable variations, has led to the development of adaptive control techniques. Adaptation, in some sense, can be viewed as "combined identification and control of a particular system"
Since adaptive control has very extensive scope , therefore it is necessary to clarify what we have in mind by the term "Adaptive Control".
The role of adaptation mechanism can either be :
o A parametric adaptation, by adjusting the parameters of the simulated plant, or
o Signal-synthesis adaptation, by applying an appropriate signal to the input of the plant.
In the case of ASEA Robotics System which is used in the Robotic Application and Development Lab (RADL), the use of "Adaptive Control" implies the ability to adapt to real world changes as determined by sensory devices, by changing the input to the system. Since the sensory device (force/torque sensor) is sensing the force therefore it is also considered as force control.

The original intent of including "Adaptive Control feature on the ASEA robot was to allow external sensors to modify the trajectory of the robot to compensate for the irregularities and uncertainties in welding and gluing operations. Trajectory modifications through the adaptive control inputs allow real time adaptation of the path

2.2 FORCE CONTROL FEATURE OF RADL SYSTEM

AS was mentioned before the goal of RADL is to accomplish the mating of an umbilical fuel line to a moving target representing the external tank of the Space Shuttle Vehicle (SSV). To perform this a vision system is first used t o approach and track the target. This is followed by mating the robot-maneuvered umbilical plate to the SSV hardware. While it is mated the hard part of the process must take place namely the robot must dublicate the random motion of SSV to avoid any large contact forces and damage to the SSV. Finally, the force controller must allow the withdrawl of the mating plate and return control back to the vision system.

During contact between the robot and an external object in this case the random motion simulator (RMS) table, forces are generated. Since the system is typically quite stiff, relatively large forces can be created by small motions. The contact force can be modified by commanding small changes in the robot's position to adjust the force to desired value. Typically the desired contact force must be large enough to allow the robot to remain in contact with the object.

One very straightforward approach to force control is called damping Control. With this method the command velocity of the robot is proportional to and in the direction opposite the applied force. In effect, the robot moves so as to relieve the forces generated during elastic contact. this approach makes the robot appear as a viscous damper.

The proportional constant between the commanded velocity and the voltage signal representing the force is called the control gain. this value approximately determines the forces that are seen at a given speed. The proper selection of the controller gain will be a prime goal in th development of the force controller. Typically, the higher the control gain, the lower the apparent damping value of the robot. This results in lower contact force for a given tracking speed. However, the higher the control gain, the more prone a system is to instability.

2.3 FORCE CONTROL USING ASEA's ADAPTIVE APPROACH

The general configuration of RADL robotic system is depicted in Fig.2. This is a functional representation of ASEA controller with force feedback.
ASEA's controller is capable of operating the robot under force control by using the Adaptive Control software package. With this approach, a correction vector is programmed prior to operation of the robot. The velocity along the vector is set proportionally to an external input signal. The analogue output signals of the JR3 are able to work directly with these inputs. The adaptive control port operates in a damping control mode, as the resulting velocity is proportional to the input voltage (force signal).
Force control through the Adaptive Control software has been achieved in the lead-around demonstration. Further, force control when the robot is in contact with a rigid object can be achieved using the Adaptive Control software, provided that the controller gain is set low enough. At this value, the motion of the robot is extremely slow for a given force, and the force/velocity performance is far from the required values.
A significant point involved in the use of the ASEA robot with force feedback control is that only the terminal points can be programmed or downloaded from an external computer. The actual trajectory for the endpoint is generated internally by an interpolation routine , as diagrammed in Fig.2. The ramification of this observation is that only modifications of the trajectory endpoints can made using an external computer. The real-time trajectory as defined by the interpolation routine, can not be modified by this approach . the importance of this observation is dependent on the relative time scales involved. For the existing vision system , trajectory endpoints can be updated at a rate of between 7 and 10 hz. With a new trajectory determined at each interval and with the robot not being required to finish it's initial trajectory the robots dynamics are slow enough to smooth out these trajectory variations .

However for systems requiring rapid modifications , such as force/torque feedback control , the time delay associated with computer communication link is expected to be slow enough to cause instabilities in the control.

The adaptive control feature of ASEA robotic system provide a path for X, Y, and Z axis. This feature allows for the preprogrammed trajectories to be modified based on external inputs to the controller. The velocity of the generated trajectory can be modified by an analogue or digital input signal , allowing an integral force feedback control loop to be placed around the existing position control loop , as demonstrated in Fig.2

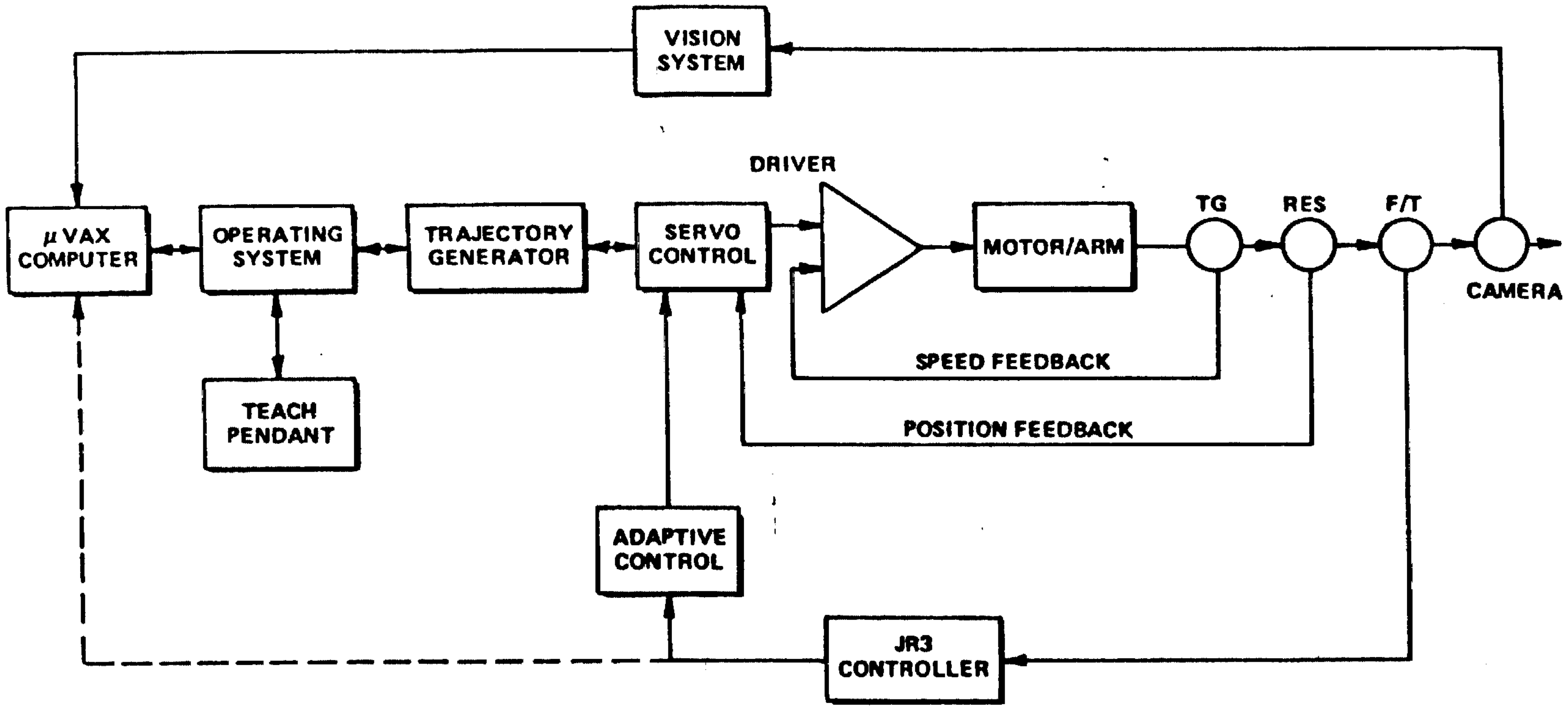


Fig.2: General configuration of RADL robotic control system

3. DYNAMIC MODELS OF FORCE-FEEDBACK ROBOT

3.1. CASE #.1. To begin with a simple case, let us consider the robot to be a rigid body with no vibrational modes. Let us also consider the workpiece (flight side) to be rigid , having no dynamics. The force sensor connects the two with some compliance as shown in Fig.3.

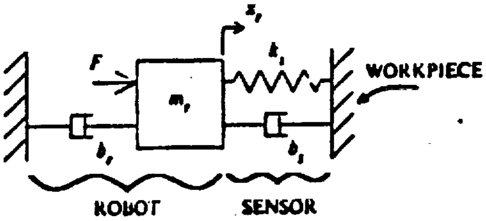


Fig.3: Robot model for case #.1

The robot has been modeled as a mass with a damper to ground. The mass m represents the effective moving mass of the arm. The viscous damper b is chosen to give the appropriate rigid body mode to the unattached robot. The sensor has stiffness k and damping b. The robot actuator is represented by the input force F and the state variable x measures the position of the robot mass.

The open-loop dynamics of this simple system are described by the following transfer function:

$$X(s)/F(s) = 1/[m_r s^2 +(b_r+ b_s)s + k_s]$$

Since this robot system is to be controlled to maintain a desired contact force, we must recognize that the closed loop system output variable is the force across the sensor, the contact force F

$$F_c = k_s x_r .$$

Implementing the simple proportional force control law :

$$F = k_f (F_d - F_c) \qquad k_f >= 0$$

which states that the actuator force should be some nonnegative force feed-back gain k_f times the difference between some desired contact force F_d and the actual contact force. This control law is embodied in the block diagram of Fig.4.

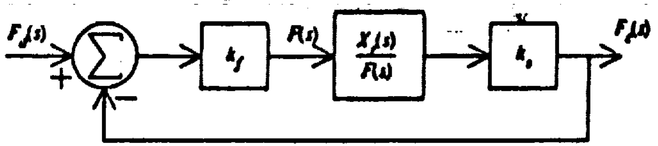


Fig.4 Block diagram for the system of case #.1

The closed loop transfer function then becomes

$$F_c\ (s)/F_d\ (s) = k_f * k_s / [m_r s^2 + (\ b_r + b_s\ )s + k_s(1 + k_f)]$$

The control loop modifies the the characteristic equation only in the stiffness term. The force control for this case works like a position servo system . This could have been predicted the model in Fig.5 by noting that the contact force depends solely upon the robot position x_r.

For completeness let us look at the root locus plot for this system.

Fig. 5 shows the positions in the s-plane of the roots of the closed loop characteristic equation as the force feedback gain k varies.

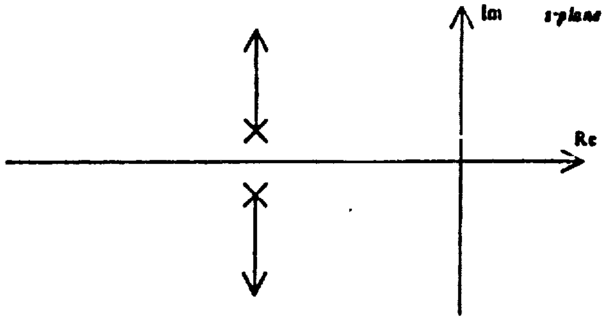


Fig.5 Root locus plot for system of case#.1

C-5

For $k_f = 0$, the roots are at the open loop poles. The loci show that as the gain is increased, the natural frequency increases, and the damping ratio decreases, but the system remains stable. In fact, k_f can be chosen to give the controlled system desirable response characteristic.

3.1.2 CASE #.2 Include flight side dynamics. The simple robot system of Fig.5 has been shown to be unconditionally stable for $k_f >= 0$. Force controlled systems, however, are not this simple and specially the neglecting of dynamics of the of the environment with which the robot is in contact plays an important role.

Fig.6 is representing the system in which the dynamics of the environment has been taken into consideration. The new state variable is now x_w measures the position .

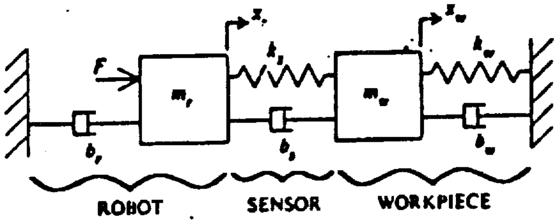


Fig.6: Dynamic model of robot described in case#.2

The open loop transfer function of this two degree of freedom system robot is :

$$X(s)/F(s) = [m_w s^2 + (b_w + b_s)s + (k_w + k_s)]/A$$

where $A = [m_r s^2 + (b_r + b_s)s + k_s]*[m_w s^2 + (b_s + b_w)s + (k_s + k_w)] - (b_s s + k_s)^2$

The output variable is again the contact force F , which is the force across the sensor, given by $F_c = k_s(x_r - x_w)$.

If we now implement the same simple force controller, the control law remains unchanged.

$$F = k_f(F_d - F_c)$$

The block diagram for this control system is shown in Fig.9.

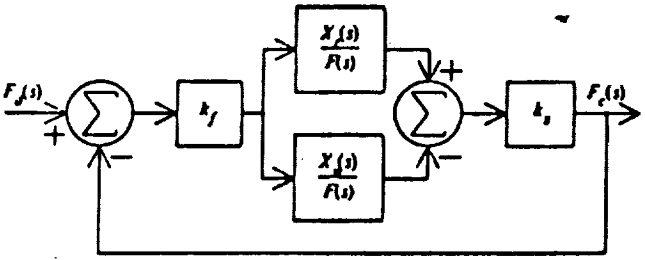


Fig.7 : Block diagram for the system of case #.2

Note that the feedforward path includes the difference between the two open loop transfer functions.

The root locus for this system is plotted in Fig.8 as the force feedback gain k_f is varied.

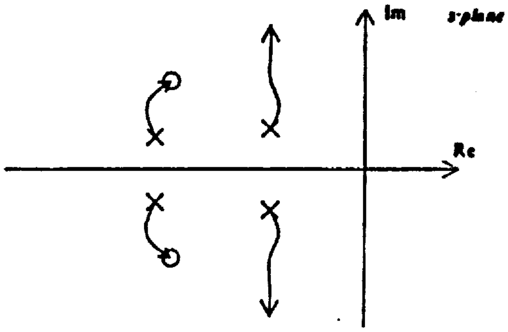


Fig.8 : Root locus plot of system of case #.2

As the root locus indicates there are four open loop poles and two two open loop zeros. The plot then still has two asymptotes at + 90 . The shape of the root locus plot tells us
that even for high values of gain, the system has stable roots Therefore, while the characteristic. of the workpiece affect the dynamics of the robot system, they do not cause unstable behavior.

3.1.3 CASE #.3. INCLUDE ROBOT DYNAMICS

Since the addition of the flight side dynamics to the simple robot system model did not result in the observed instability, we will consider a system with a more complex robot model. If we wish to include both the rigid-body and first vibratory modes of the arm, then the robot alone must be represented by two masses . Fig 9 shows the new system model.

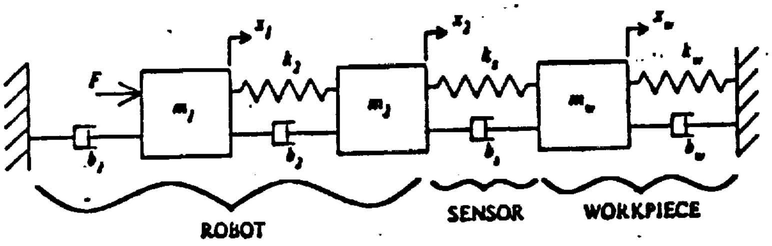


Fig.9 : Robot system model described in case #.3.

The total robot mass is now split between m_1 and m_2. The spring and the damper with values k_2 and b_2 set the frequency and damping of the robot's first mode, while the damper ground, b_1, primarily governs the rigid-body mode. The stiffness between the robot mass could be the drive train or transmission stiffness, or it could be the structural stiffness of a link. The masses m_1 and m_2 would then be chosen accordingly. The sensor and workpiece are modeled in the same manner as in case #.1 and case #.2. The three state variables x_1 x_2 and x_w measure the positions of the masses m_1 m_2 and m_w .

This-mass model has the following open-loop transfer function:

$X_1(s)/F(s) = A/Y$, $X_2(s)/F(s) = B/Y$ and $X_w(s)/F(s) = C/Y$

where

$$A = [m_2 s^2 + (b_2 + b_s)s + (k_2 + k_s)] * [m_w s^2 + (b_s + b_w)s + (k_s + k_w)] - (b_s s + k_s)^2$$

$$B = [m_w s^2 + (b_s + b_w)s + (k_s + k_w)][b_2 s + k_2]$$

$$C = [b_2 s + k_2][b_s s + k_s]$$

$$Y = [m_1 s^2 + (b_1 + b_2)s + k_2] * [m_2 s^2 + (b_2 + b_s)s + (k_2 + k_s)] * [m_w s^2 + (b_s + b_w) + (k_s + k_w)] -$$

$$-[m_w s^2 + (b_s + b_w)s + (k_s + k_w)][b_2 s + k_2] - [m_1 s^2 + (b_1 + b_2)s + k_2][b_s s + k_s]^2$$

The contact force is again the force across k ,

$$F_c = k_s(x_2 - x_N)$$

and the simple force control law is

$$F = k_f(F_d - F_c) \qquad (k >=0)$$

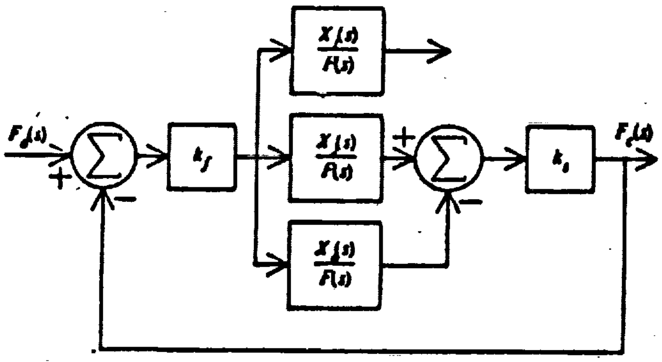


Fig.10: Block diagram of the system of case #.3

The block diagram for this controller, Fig.10, shows again that the feedforward path takes the difference between two open-loop transfer functions

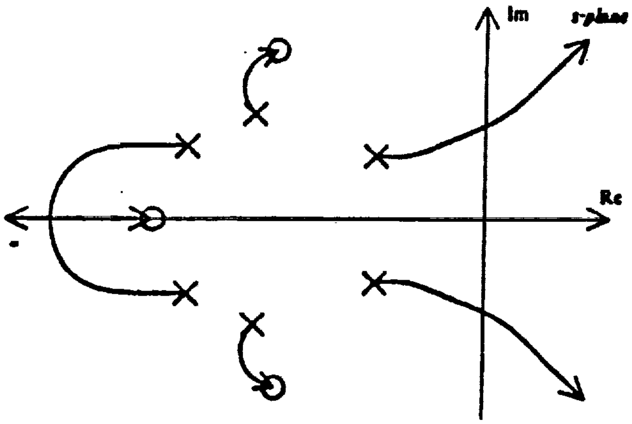


Fig.11: Root locus plot for the system of Fig.12

The root locus plot, Fig.11, shows a very interesting effect.

The system is only conditionally stable.
For low values of k, the system is stable; for high values of k , the system is unstable; and for some critical value of the force feedback gain, the system is only marginally stable.

The + 60 asymptotes result from the system's having six open loop poles, but only three open loop zeros. Inspection of the open-loop transfer function confirms this: the numerator of the transfer function ralating X (s) to F(s) is a third-order polynomial in s.

4. MATHEMATICAL MODEL OF FORCE FEEDBACK CONTROL , FOR (ASEA) ROBOT

4. 1 GENERAL DESCRIPTION OF ASEA ROBOT

The ASEA IRB 90 robot is a six axis manipulator coupled with a sophisticated controller. While fig. 2 provided functional representation of the ASEA robot , fig. 12 represents the control system for each axis of the robot.

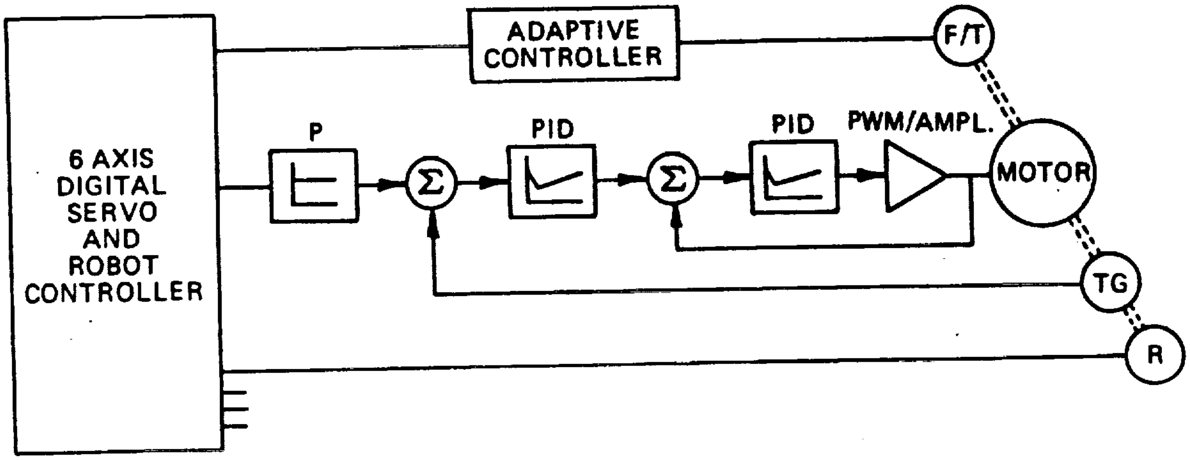


Fig.12 Block diagram of control system for each axis of the robot.

Each robot joint is driven by a DC motor through a redactor. The motors are powered by Pulse Width Modulation circuits using armature voltage control technique. The controller uses both velocity and position feedback signals in a conventional manner, with the PID inner velocity feedback loop surrounded by by a position control loop. In order to limit the armature current and to improve the linearity of the system a current feedback loop is also employed.

4.2 ACCURATE MODEL WITH GENERAL PARAMETERS

Based on the block diagram depicted above and the operation of the random motion simulator, it is clear that case #. 2 described in section 3 of this paper is most appropriate to be used as base model.
Using the force dependent voltage from the force/ torque sensor allows the ASEA's adaptive control software to generate a change in the velocity based on an error between the observed force and a bias value representing the force setpoint value.
Fig. 15 is block diagram representation of model of force feedback control structure . The equations governing the system is as following.

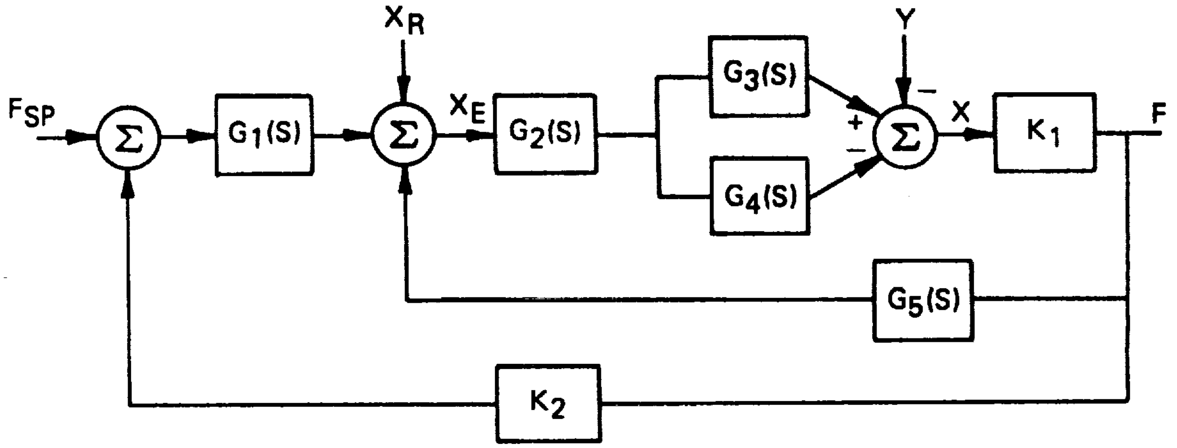


Fig. 13. Block diagram representing model of force feedback control

$$X_e = X_R + G_1(S)(F_{SET} - K_2 F) - G_5(S)F$$

$$X_e = X \,/\{[G_2(S)][G_3(S) - G_4(S)]\}$$

$$X_R + G_1(S)(F_{SET} - K_2 F) - G_5(S)F = X/\{[G_2(S)[G_3(S) - G_4(S)]\}$$

This would be the model that governs the behavior of the robot when operating in a lead around mode, because of the lack of coupling between the motion of the SSV and the generated force. When in contact with a rigid body the interaction between the robot and an external motion becomes coupled and is modeled as a stiff elastic member.

$$F = K(X - Y) \quad , \quad X = F/K + KY$$

$$[F/K + KY]/\{[G_2(S)][G_3(S) - G_4(S)]\} = X_R + G_1(S)\ (F_{SET} - K_2F) - G_5(S)F$$

$$[G_2(S)][G_3(S) - G_4(S)] = G_x(S)$$

$$[F/K + KY]/\ G_x(S) = X_R + G_1(S)(F_{SET} - K_2F) - G_5(S)F$$

$$[F/K + KY) = X_R G_x(S) + G_x(S)\ G_1(S)\ F_{SET} - G_x(S)G_1(S)\ K_2F - G_x(S)\ G_5(S)F$$

$$F/K = X_R(G_x(S)) + G_x(S)G_1(S)F_{SET} - G_x(S)G_1(S)K_2F - G_x(X)G_5(S)F$$

$$F[1/K + G_x(S)G_1(S)K_2 + G_x(S)G_5(S)] = X_R G_x(S) + G_x(S)G_1(S)F_{SET}$$

$$F = [X_R G_x(S) + G_x(S)G_1(S)F_{SET}]/[1/K + G_x(S)\ G_1(S)K_2 + G_x(S)G_5(S)]$$

To make the model practical, it is needed to determine the transfer functions of each block .

1. $G_2(S) = G_m(S)*G_c(S)$, where $G_m(S)$ is the well known transfer function for the torque output vs applied voltage for a DC motor is given :

$$G_m(S) = \frac{K_t}{K_t K_e + RB} * \frac{JS + B}{JL/[K_t K_e + RB]\ S^2 + [JR + LB]/[K_t K_e + RB]\ S + 1}$$

2. G (S) is the transfer function of the compensator = $K_p + K_I/S$

3. G (S) is the transfer function of the adaptive control path which was determined [3] to be equal to K /S without delay and K /S * $e^{-\tau S}$ with delay.

4. $G_3(S) = X_r(S)/F(S)$, $G_4(S) = X_w(S)/F(S)$, related to robot and flight simulator dynamics and were determined in section 3.

The model is still theoretical until the coefficients of the transfer function are determined. In order to obtain concrete parameters of the system , one can use two two different methods. Using catalogues and manufacturer's data or direct measurement. While manufacturer's data can often be accurate, accuracy of direct measurement obviously depends on precision in measurement. In our case unfortunately obtaining data from manufacturer was not possible, so the only alternative was to rely on direct measurement of frequency and time response of the system, which led to a simple single degree-of-freedom model as shown in Fig.14.

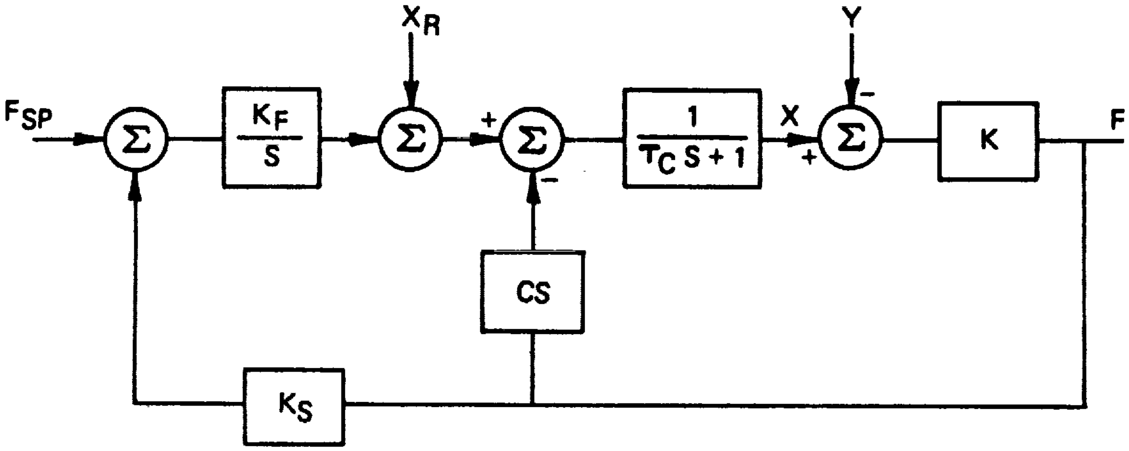


Fig. 14 Approximate model of the system

4.4 EXPERIMENTAL RESULTS

Fig. 15 is the time response of the system, figures 16 and 17 represent the frequency response of the system. Fig 18 demonstrates a significant time lag that exists in the adaptive control software for force control. It was determined by simultaneously plotting the reference voltage into the adaptive control port and the resulting motion .

This delay also could be identified via frequency response analysis.

Fig. 18 demonstrates the tremendous phase lag encountered at higher frequencies, as typically found in systems with a time delay. An approximate transfer function has been determined by [3] which provides a fairly good fit.

$$X(S)/V(S) = \frac{6.0\, e^{-.3S}}{S(0.1429\, S + 1)}$$

From this transfer function and the data obtained by [3], the following values may be assigned: T =0.14297, $K_f = 6$, K = 134lb/in, $K_s = 0.004$V/lb.

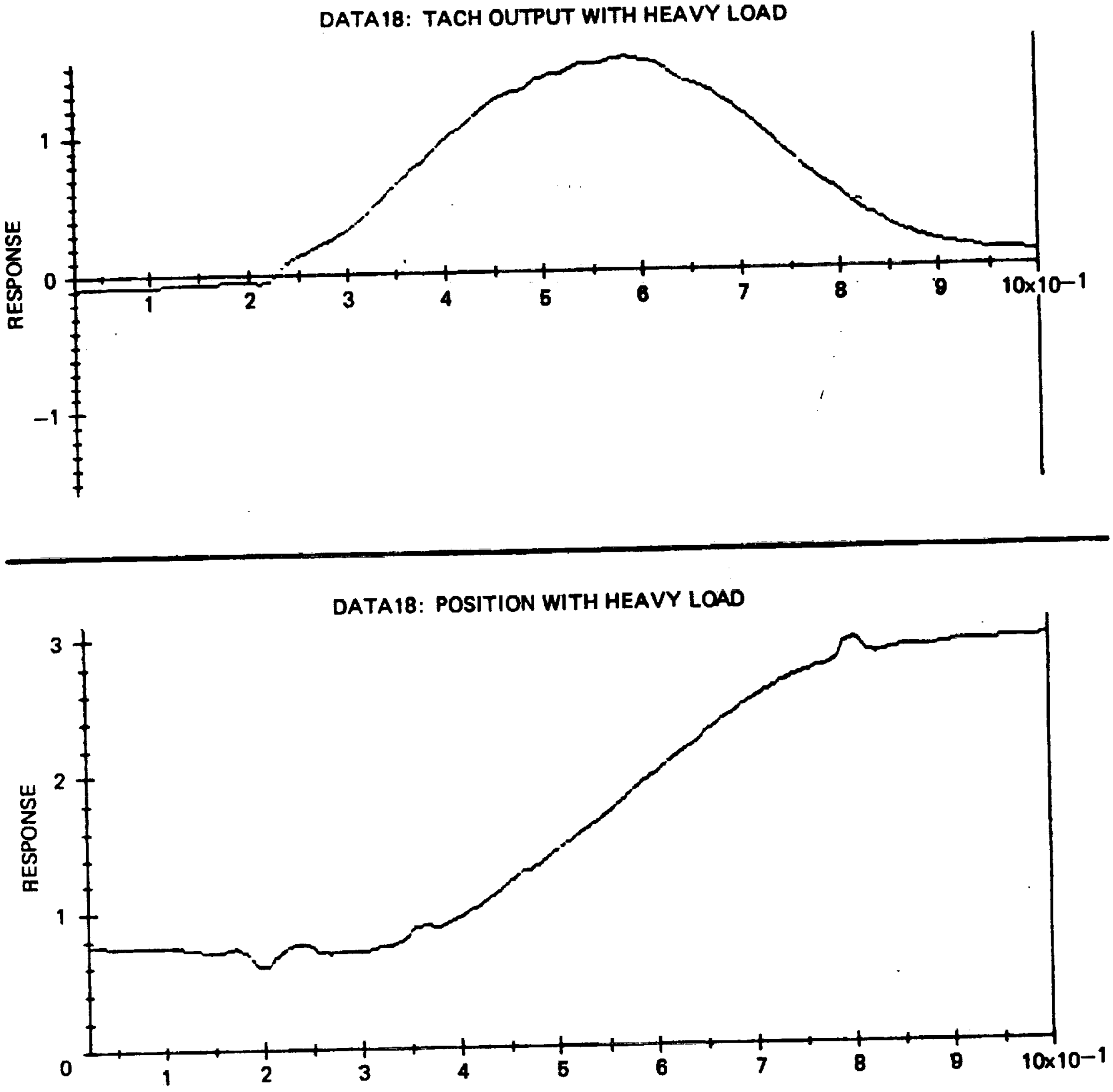


Fig. 15 Time response

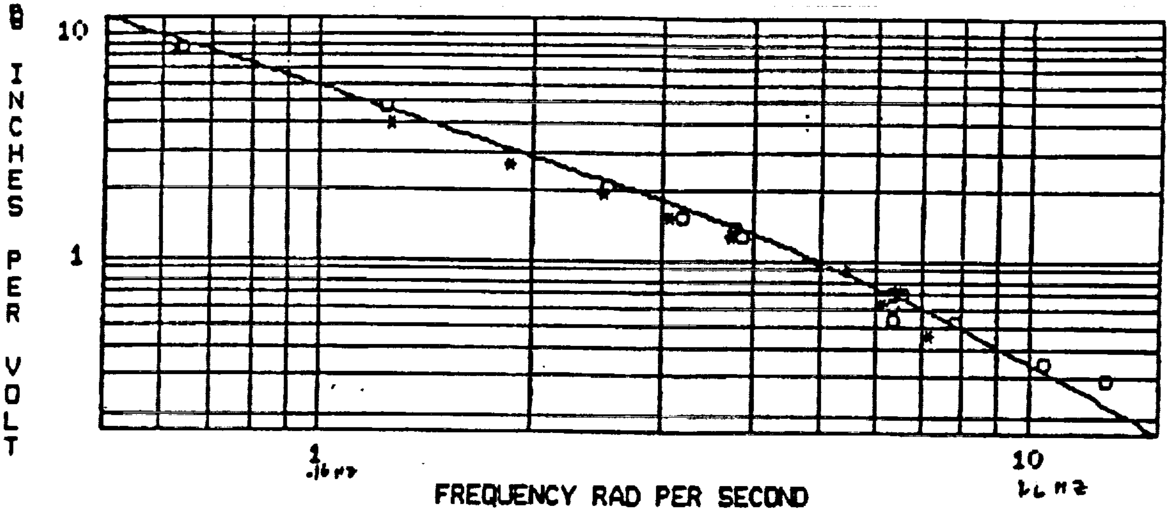


Fig. 16 Frequency response plot of base rotation axis

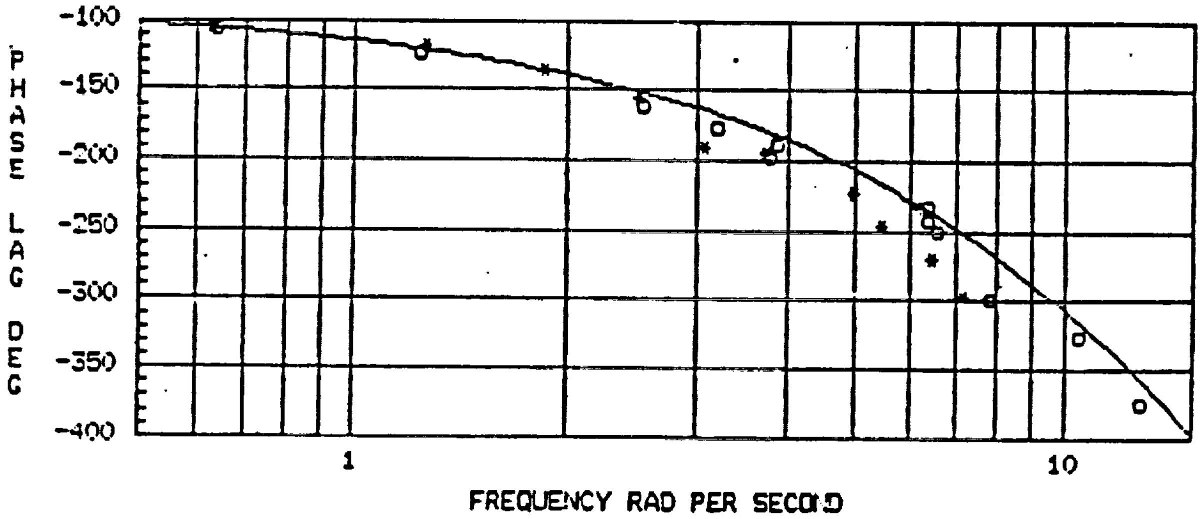
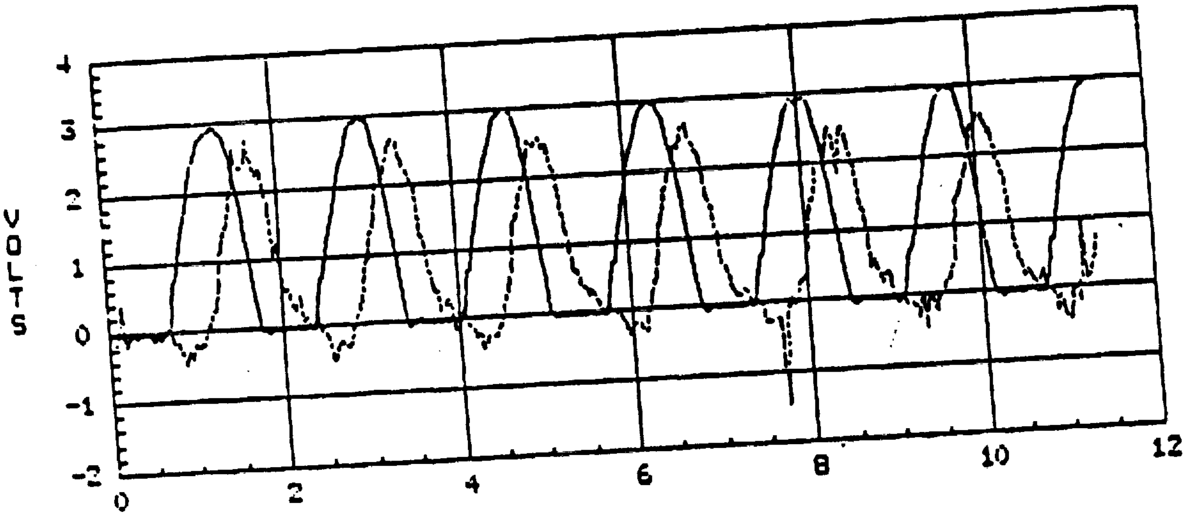


Fig. 17 Frequency response plot of base rotation axis



Delay in ASEA controller between
input voltage and resulting
servo command voltage

5. CONCLUSIONS

An accurate lumped parameter model for the control system of ASEA robot has been developed. Efforts to determine concrete values for the parameters has been unsuccessful. However, an approximate linear model with concrete parameters to replace the accurate model has been suggested.

Although no theoretical proof has been presented, practically it was found out that time delay between the output and input signals in ASEA controller can cause instability. Without the controller latency, stable force control during both tracking and mating can be achieved.

The combination of passive compliance and force control provide excellent performance when mated.

More over a series of lumped- parameter models has been developed in order to understand the effects of robot and workpiece dynamics on the stability of simple force controlled systems. An instability has been shown to exist for robot models that include representation of a first resonant mode for the arm. The effect of the workpiece dynamics remains unclear. It has been shown that when the workpiece is modeled as a rigid wall, the system can be unstable. Certainly if the workpiece were very complaint and extremely light there could be no force across the sensor, degenerating the closed loop system to the open loop case, which ofcourse is stable. The sensor and workpiece dynamics are therefore important and should be modeled. Limited actuator bandwith, filtering, and digital controller implementation can also cause instability. These performance limitations must also be included in the system model used for controller design.

6. REFERENCES

[1] S. D. Eppinger and W. P. Seering, "On Dynamic Models of Robot Force Control,"
Proc. Int. Conf. Robot. Autom., April 1986.

[2] O. Zia , "Adaptive Servo Control for Umbilical Mating. "
1988 NASA/ASEE Summer Faculty Report, University of Central Florida.

[3] R. Rees Fulmer, " The Development of Force Feedback Control for Umbilical Mating,"
Final Contract Report, RADL , Kennedy Space Center Florida.

[4] D. E. Whitney, " Force Control of Manipulator Fine Motions,"
J. Dyn. Syst. , Measurement and Control, Vol.99, June 1987.

[5] E. Carrizosa, V. Feliu, " Study of the Dynamical Behavior of an Assembly Robot,"
Robot Control (SYROCO'85),Proc. of 1st IFAC Symposium, Spain ,Nov. 1985.

[6] E. M. Onaga and L. L. Woodland, "Six-axis Digital Torque Servo "
Proc. Int. Symposium on Industrial Robot, April 1987.

N90-16702

1989 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

INVESTIGATION OF IGES FOR CAD/CAE DATA TRANSFER

PREPARED BY:	Dr. George W. Zobrist
ACADEMIC RANK:	Professor
UNIVERSITY AND DEPARTMENT:	University of Missouri - Rolla Department of Computer Science
NASA/KSC	
DIVISION:	Data Systems
BRANCH:	Technical/Information Systems CAD/CAE
NASA COLLEAGUE:	Mr. Jerry Barnes
DATE:	July 28, 1989
CONTRACT NUMBER:	University of Central Florida NASA-NGT-60002 Supplement: 2

ACKNOWLEDGMENTS

This is to acknowledge the support of Mr. Jerry E. Barnes and Mr. Hank Perkins of DL-DSD-22, Mr. Bill Higgs of EG&G, Mr. Ed Scarbourough and Mr. Don Chappell of Lockheed Space Operations, and Mr. Frank Kapr, TV-TIO. They were very helpful in the initial guidance of this research effort and in obtaining the necessary resources. Mr. L. Purves and Mr. S. Gordon, NASA/Goddard, provided an IGES test file and helpful comments.

Additionally, I wish to acknowledge the administrative support of Dr. R. Hosler, University of Central Florida and Mr. Dennis Armstrong, Training Section, NASA/KSC.

ABSTRACT

In a CAD/CAE facility there is always the possibility that one may want to transfer the design graphics database from the native system to a non-native system. This may occur because of dissimilar systems within an organization or a new CAD/CAE system is to be purchased. The Initial Graphics Exchange Specification (IGES) was developed in an attempt to solve this scenario. IGES is a neutral database format into which the CAD/CAE native database format can be translated to and from. Translating the native design database format to IGES requires a pre-processor and translating from IGES to the native database format requires a post-processor.

IGES is an artifice to represent CAD/CAE product data in a neutral environment to allow interfacing applications, archive the database, interchange of product data between dissimilar CAD/CAE systems, and other applications.

The intent of this paper is to present test data on translating design product data from a CAD/CAE system to itself and to translate data initially prepared in IGES format to various native design formats. This information can be utilized in planning potential procurement and developing a design discipline within the CAD/CAE community.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

I. INTRODUCTION

In a CAD/CAE facility there is always the possibility that one may want to transfer the design graphics database from the native system to a non-native system. This may occur because of dissimilar systems within an organization or a new CAD/CAE system is to be purchased. The Initial Graphics Exchange Specification (IGES) was developed in an attempt to solve this scenario. IGES is a neutral database format into which the CAD/CAE native database format can be translated to and from. Translating the native design database format to IGES requires a pre-processor and translating from IGES to the native database format requires a post-processor.

IGES was developed in 1979 under direction of the National Bureau of Standards and several industrial concerns. Version 1.0 of IGES was published as part of an ANSI standard in 1981, Version 2.0 in 1983, Version 3.0 in 1986, and Version 4.0 in 1988 (ref. 1,2,3,4).

Version 1.0 supported CAD/CAE geometries, annotation entities, wireframe entities and some surfaces, Version 2.0 additionally supported finite element modeling, printed circuit board models, more text fonts, and extended some of the geometrical entities, Version 3.0 added additional surfaces, clarification of view and drawing entities, enhanced MACRO capability, plant flow and ASCII compression, and Version 4.0 supports solid models, enhanced electrical and finite element applications, and introduction of architecture/engineering/construction applications.

IGES is an artifice to represent CAD/CAE product data in a neutral environment to allow interfacing applications, archive the database, interchange of product data between dissimilar CAD/CAE systems, and other applications.

Developers must write software to go from the native database format to the IGES neutral database, and vice versa, since IGES is a specification and not a product. Therefore the IGES file is only as good as the developer's effort in this regard. In general, IGES is a superset of a CAD/CAE systems entity menu.

The intent of this paper is to present test data on translating design product data from a CAD/CAE system to itself and to translate data initially prepared in IGES format to various native design formats. This information can be utilized in planning potential procurement and developing a design discipline within the CAD/CAE community.

II. USAGES OF NEUTRAL DATA FILE

The concept of the neutral data file was in usage before IGES was developed through the development of database interfaces by various vendors. These interfaces were normally used by application engineers to write programs of use to the design organizations. One example was the development of a Motor Control Center (MCC) placement and one-line diagram drawing by interfacing vendor catalog information, MCC module placement algorithms, and drawing commands through the host neutral data file (ref. 5).

This neutral datafile contained the drawing command structure to enable the application engineer to invoke various graphics design entities, such as lines, circles, points, text, etc.

This concept is useful as long as one is utilizing a single vendor for the applications and the system will not be changed in the forseeable future. Once the CAD/CAE system is changed then the application programs can not utilized since the graphics commands will not normally be recognized by a different vendor. To achieve an environment whereby the product design data and applications could become stable requires a standard product design data interface. This accomplishment is attempted by IGES.

The concept of the neutral datafile can be utilized in more scenarios than transferring product data between dissimilar systems. One example was illustrated in the preceding paragraphs.

Various uses of the neutral graphics database follows (ref. 6):

a. A means for transferring product graphics design data between dissimilar CAD/CAE systems. This in principle allows design data to be represented in a neutral file so that it can be translated to a future CAD/CAE systems native graphics database. Thereby design drawings need not be re-drawn each time a new system is purchased, or if one is required to transfer graphics design data to another system for integration of electrical/mechanical information, or for checking by a facility which has a non-compatible system, etc.

b. As mentioned earlier one can develop application programs that utilize the neutral database format. These applications are useful in the design/analysis mode and pre-preparation of various design commands.

c. It is also possible to edit CAD/CAE drawings from a terminal rather than at a design workstation. This reduces editing time and a possible reduction in cost, due to the cost differential of terminals versus workstations.

d. Possibly one of the more useful applications of the neutral file concept is to archive design drawings. If the design graphics is stored in the native graphics format, it is probable that in the future the product design database would not be compatible with the CAD/CAE system in usage at that time, even if it was from the same vendor. Once the graphics is in a neutral format, one can in principle write a post-processor to translate the neutral database to the present native design format. This translator can be utilized on all archived drawings that are to be installed on that particular system.

e. One can envision various artificial intelligence (AI) type applications utilizing an expert system that will operate upon the neutral database. Possible applications could be, rules that allow interference checking in electrical/mechanical/piping drawings, rules for printed circuit board physical layout, integrated diagnostics (ref. 7), etc. One could also envision development of an expert system that checked a drawing for completeness, i.e., a rectangle which is not closed, as a simple example. If the expert system is designed around the neutral file database, then if the native format changes this should not disturb the algorithms developed.

It should be noted that in practice most of these would be difficult to achieve with IGES in it's present form. This will be discussed in a later section.

III. GENERIC COMPONENTS OF PRODUCT DATABASE

The major components of a generic product database are the following (ref. 8):

3.1 FORMAT

Formatting refers to the various bit representations in a system, i.e., character, floating point, fixed point, and integer being the most common ones. This manifests itself in the basic accuracy of the drawing and the character set representation. There is an inherent problem in matching the accuracy of the model generated to the model being transferred to another CAD/CAE system.

3.2 REPRESENTATION

This refers to how the geometry of a part is represented. There are several different schemes for part representation. A part can be represented by **edge boundary** or, **wireframe**. This is where the part's extremes are represented by a collection of curves in space. Other representations are, **surface** and **hybrid edge-surface**. The surface representation is more precise, especially for points not on an edge boundary, and the hybrid edge-surface is a combination of the preceding representations.

The representation principally provides the collection of geometrical parameters that make up each data element. For example, the representation of a line is it's end-points versus an equation with initial and final points.

3.3 MEANING

The meaning conveys the design intent of the data elements. One may have four lines connected in a rectangular pattern. This could either represent four disjoint lines or could represent a plane. To convey meaning one needs the concept of associativity whereby the four lines can be associated together, or not. This is a subtle concept since many times the meaning can only be conveyed by the user, unless associativity attributes are given.

IV. IGES FILE STRUCTURE

The IGES file structure (ref. 4), illustrated in Figure 1, is composed of six sections and they must appear in the following order:

a. Start - this section provides a human-readable prologue to the file. There must be at least one start record.

b. Global - this section contains the information to be used by the pre- and post-processor to translate the file. A sampling of the items contained in this section are; parameter and record delimiters used, information about sending system, file name, data format information, model space scale, user intended resolution. Basically, this section provides a definition of the global conditions under which the model was generated.

c. Directory - there is a directory entry for each entity in the file. This entry is fixed in size and contains twenty fixed format fields. This section provides an index for the file and attribute information about each entity. Typical attributes would be, line font, view, level, transformation matrix used, line weight, color, and form number.

d. Data - this section contains parameter data associated with each entity. This section has a free format structure. Typically, items in this section enable the graphics system to place the entity in the drawing. Therefore, this section contains placement data, pointers to properties/attributes of the entity, and back-pointers to associativity instances. The Data and Directory section comprise the representation of the entity and are used together.

e. Terminate - this section contains only one line and is fixed format. This record is used to total up the number of entries in the previous sections.

The Directory/Data sections result in redundant data and forward/backward pointers. This results in voluminous file size and abortive results if pointers are omitted or corrupted.

The IGES structure is a fixed length record of up to 80 ASCII characters. This allows for universal file readability, but it also is quite cumbersome. Although, later versions have the option of a compressed ASCII and Binary format which can be utilized to reduce file size. The compressed ASCII and Binary formatting addresses the volume of data, but imposes a processing burden compared to ASCII.

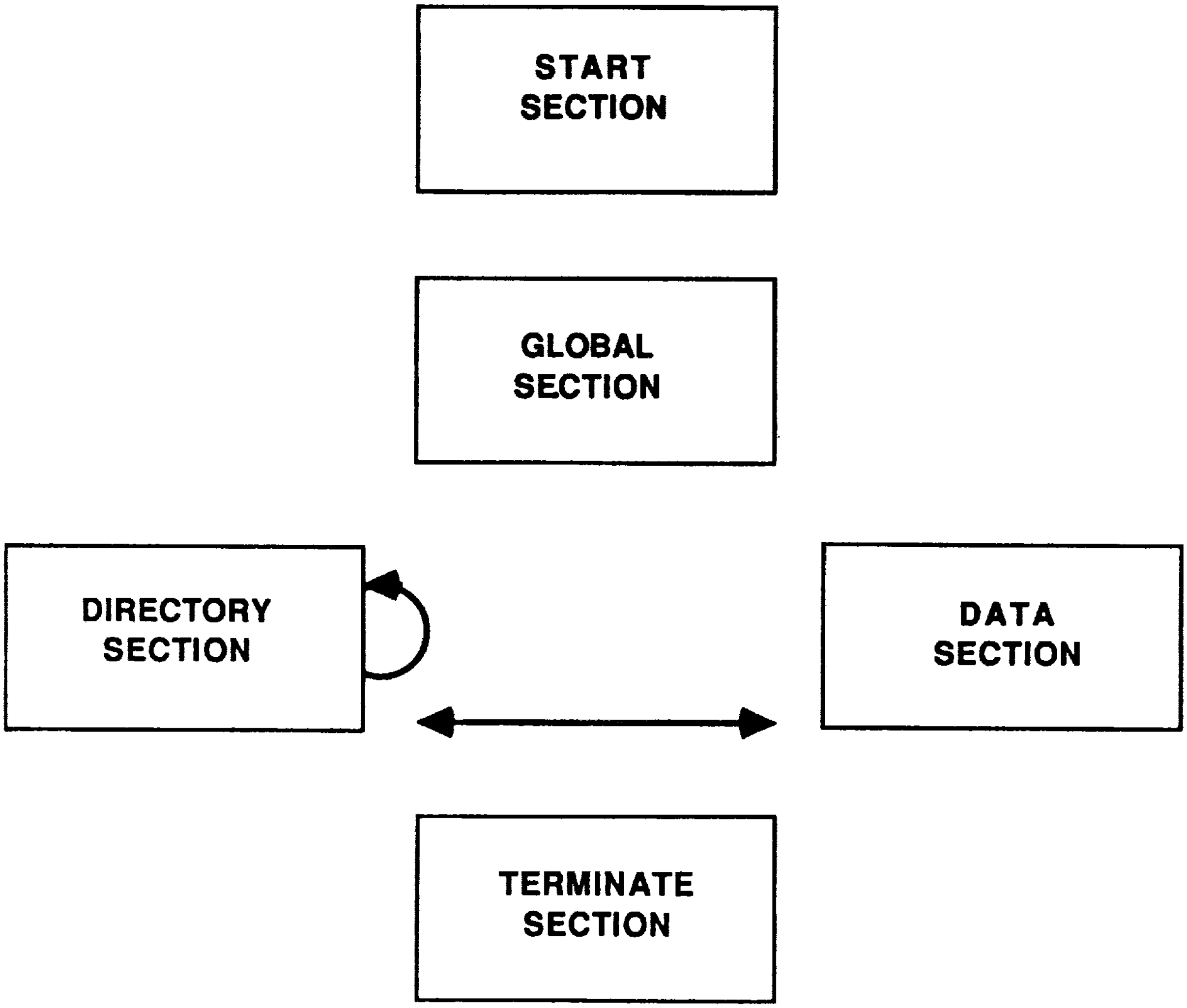


Figure 1 IGES File Structure Sections

V. IGES COMPONENTS OF PRODUCT DATA BASE

The format of the IGES file is 80 character ASCII records which detail the native system format. This also creates a large file structure.

There are four basic representations in IGES. They are, geometry, dimensioning and annotation, structure and properties. The main geometrical entities are the point, line, conic, arc, parametric spline, face, ruled surface, surface of revolution, and tabulated cylinder. These can be used to represent the basic graphical entities used in a drawing. Dimensioning and annotation are composed of text, arrowhead, and witness lines in various forms and styles. There are also special dimensioning entities such as, center lines. Splines are also represented, but different curves can result in the translation from a common set of input conditions. This is due to the host algorithm for representing a spline from it's input points and conditions. This is addressed in IGES by having a variety of spline forms.

IGES meaning is addressed through various structuring and property mechanisms. There are methods for assigning specific relationships between entities and also to convey meaning to these relationships. There are three important methods utilized. The associativity mechanism places specific entities into a group. An example, would be placing the four lines of a rectangle into a group representing a plane. Another mechanism is to place a group of entities into a view. This is a theoretical cube in which the entity group is placed and can be rotated and/or clipped. The final mechanism is the drawing which is a collection of view entities.

IGES also has the capability of the user defining properties.

VI. PROCESSING

The procedure for processing product design data is to translate from the native format to IGES by a pre-processor and a post-processor is utilized to translate from IGES to the native format. Many times this is done by developing pre- and post-processors to translate between the host's own neutral file. Intergraph translated IGES to/from the Standard Interchange Format (SIF) which is it's representation of the native graphics design database.

There are many difficulties associated with this task. There is not always a one-to-one mapping between the native graphics design format and IGES format. There can be one-to-many, many-to-one, and null translations. For example, the pre-processor must decide for a particular line font utilized by the host which IGES line font to use (one-to-many) and the post-processor must decide which native line font to use for a class of IGES line fonts (many-to-one). There is also the possibility that a particular native database entity has no IGES entity or vice-versa (null). For example, the native database may have only ellipse entities with the circle being a special case and IGES has both circles and elliptical entities, this will result in a null situation. There are also meaning conflicts; should a plane be represented as four connected line entities or a plane entity?

VII. IGES PROBLEMS

There are several problems which are typically encountered in utilizing the neutral database concept (ref. 9). They are; incomplete processors, poor choice of mapping, internal database organization has structural differences, and the user's choice of host drawing entities.

The incomplete processor problem must be addressed by the vendor since they are the one's who develop the translators between the native database format and IGES. Once this translator has been developed the user can not improve upon it. Although there may be some 'fine tuning' that could possibly be done through an expert system, if additional information could be obtained from the vendor on its native database structure.

The vendor has the responsibility for mapping choices. An example, would be whether a plane should be mapped into a separate entity, or mapped into it's constituent parts. Also, many times special symbols are preprocessed into a geometrical part, such as an ASCII character mapping into a particular arrowhead. Some of the mappings may be poor ones and hence difficult to recover through a re-translation.

Another problem is in how the host's internal data organization is represented. An example would be whether text should be free-standing or attached to the appropriate entity. The representation problem can result in unreadable drawings, caused by text overlapping, spacing problems, rotations, problems resulting from roundoff due to different numerical formats in vendor A and B. This is also, inherently, a result of how the vendor represents the model internally and little can be done by the user.

The last problem to be discussed is the user's choice of graphic entities. The entities that the user employs in the design process can result in efficient or inefficient translation of a drawing. If the user chooses and/or arranges entities that best suit the application and then when these are translated into IGES they may or may not be the best entities for re-translation to a design file. To address this problem the design organization can develop an IGES translation manual which lists host entities and their equivalent IGES entities, denoting if they are one-to-one, one-to-many, many-to-one, and null. This can result in user discipline in utilizing a set of host entities that are suited for translation. Of course, the problem is that user choice and innovation will be restricted.

VIII. OTHER NEUTRAL DESIGN FILES

Vendors have also developed their own neutral data files. Many of these formats are superior, in certain aspects, to IGES, but they are not industry standards and hence can normally only be utilized for the CAD/ CAE system for which they were developed. Typically they were designed as an application interface rather than for product data transfer. This section will briefly describe the format utilized by Intergraph and Autocad.

The Intergraph format is the Standard Interchange Format (SIF) and it has a relatively simple format, as shown in Figure 2. There are no forward and/or backward pointers, it is easily read and edited. It has only one entity record, as compared to the Directory/Data relationship in IGES. A disadvantage is that it is free format which requires a parser to read and interpret. Another disadvantage is that placement data is in UOR's rather than design units. A UOR is a drawing coordinate.

The Autocad format is called DXF and has a simple structure. Most of the file is fixed format and hence does not have to be completely parsed and interpreted. The format is simple enough, so that it can be edited from a terminal, as compared to IGES, although the files can be quite large. A disadvantage is that it doesn't support as many graphic entities as IGES.

One of the major advantages of IGES is that it accommodates most graphics entities that a design organization may require and does a reasonably good job with geometrical data. Disadvantages are; some translators are not fault tolerant, use forward/backward pointers in the Directory/Data section, errors in the pointer structure will destroy the entire drawing, difficult to edit, file transfer can be quite slow due to the large file size, e.g., a simple graphic line requires three entries in the Directory/Data section, see Figure 3, graphic entities may be transferred but their meaning lost, and at present very few translations are 100% correct.

The major difficulty with non-IGES neutral files is that they are not industry standards and typically not required by major industries and/or governmental agencies which utilize CAD/CAE services.

```
DID/NA=ZOBRIST2,DA=06/08/89,MO=3,RA=0,0,-1524000,4572000,0,1524000,DU=
     16909,1000,254,IN,ML,created by Intergraph SIF release 8.8 12-FEB-19
     88 REV2
LAC/LT=2
LAC/LC=3
BST/OP
LST/OP,0,0,762000,1524000,0,762000
ARC/CL,CE=2286000,0,762000,P1=1524000,0,762000,P2=3048000,0,762000,MA=
     1.,0.,0.,0.,0.,1.,0.,-1.,0.
LST/OP,3048000,0,762000,4572000,0,762000
EST/
BST/OP
LST/OP,0,0,-762000,1524000,0,-762000
ARC/CL,CE=2286000,0,-762000,P1=1524000,0,-762000,P2=3048000,0,-762000,
     MA=1.,0.,0.,0.,0.,-1.,0.,1.,0.
LST/OP,3048000,0,-762000,4572000,0,-762000
EST/
```

Figure 2   Intergraph Standard Interchange Format Structure

```
This IGES file was created by the INTERGRAPH IGES_OUT translator        S      1
1H,,1H;,21HINTERGRAPH 8.8.5 IGES,12HZOBRIST2.IGS,10H8.8.5 IGES,         G      1
3H1.0,32,08,24,08,56,,1.0,1,4HINCH,32,,13H890608.144640,                G      2
 0.000003937007932,,17HSTANDARD PRODUCTS,10HINTERGRAPH,;                G      3
     124       1       0       0       0       0       0   0 0 0 0 0D      1
     124       0       0       1       0              MATRIX        1D      2
     110       2       1              40       0       0     00010100D      3
     110       2       3       1                       LINE          1D      4
     124       3       1       1       0       0       0     00000000D      5
     124       0       0       2                      MATRIX         0D      6
     100       5       1              40       0       5     00010000D      7
     100       2       3       2                       ARC           1D      8
     110       7       1              40       0       0     00010100D      9
     110       2       3       1                       LINE          2D     10
     102       8       1              40       0       0             D     11
     102               0       1                       COMPCURV      1D     12
     110       9       1              40       0       0     00010100D     13
     110       2       0       1                       LINE          3D     14
     124      10       1       1       0       0       0     00000000D     15
     124       0       0       2                      MATRIX         0D     16
     100      12       1              40       0      15     00010000D     17
     100       2       0       2                       ARC           2D     18
     110      14       1              40       0       0     00010100D     19
     110       2       0       1                       LINE          4D     20
     102      15       1              40       0       0             D     21
     102               0       1                       COMPCURV      2D     22
124,1.0,0.0,0.0,0.0,0.0,1.0,0.0,0.0,0.0,0.0,1.0,0.0;                   1P      1
110,0.000000,0.000000,3.000000,6.000000,0.000000,3.000000;             3P      2
124,1.000000,0.000000,0.000000,9.000000,0.000000,0.000000,             5P      3
-1.000000,0.000000,0.000000,1.000000,0.000000,3.000000,0;              5P      4
100,0.000000,0.000000,0.000000,3.000000,0.000000,-3.000000,            7P      5
0.000000;                                                              7P      6
110,12.000000,0.000000,3.000000,18.000000,0.000000,3.000000;           9P      7
102,3,3,7,9,0;                                                        11P      8
110,0.000000,0.000000,-3.000000,6.000000,0.000000,-3.000000;          13P      9
124,1.000000,0.000000,0.000000,9.000000,0.000000,0.000000,            15P     10
1.000000,0.000000,0.000000,-1.000000,0.000000,-3.000000,0;            15P     11
100,0.000000,0.000000,0.000000,3.000000,0.000000,-3.000000,           17P     12
0.000000;                                                             17P     13
110,12.000000,0.000000,-3.000000,18.000000,0.000000,-3.000000;        19P     14
102,3,13,17,19,0;                                                     21P     15
S      1G      3D     22P     15                                        T      1
```

Figure 3   IGES File Structure.

IX. ALTERNATIVES

There are several alternatives to using IGES to transfer graphics data between dissimilar CAD/CAE systems.

One approach is to write a direct translator, i.e., one which translates the vendor A database directly to vendor B database. These translators usually are very efficient, since they address a particular problem. To build one of these translators requires knowledge about the data structure for each system for which there is to be a translator built. One possible technique is to utilize the vendors neutral file rather than IGES, such as, SIF or DXF.

Of course, if the CAD/CAE systems for which the direct translator is built is changed a new translator must be designed. This would require n(n-1) translators to be built, if graphics data is to be transferred between n dissimilar CAD/CAE systems, as illustrated in Figure 4. IGES only requires 2n pre/post-processors to be built for the same number of dissimilar systems, which is shown in Figure 5.

If a vendor changes the native database structure, then n-1 direct translators would have to be re-built, but only 2 pre/post-processors.

A new neutral file structure is being developed it is called Product Data Exchange Specification (PDES) (ref. 10). PDES is planned for release in the 1990's and defines a more conceptual model than IGES.

The model consists of an application layer, conceptual layer, and a physical layer. The application layer is concerned with the application, i.e., electrical, mechanical, architectural, etc. The conceptual layer is concerned with concepts such as, tolerance envelopes, solids with flanges, etc. and the physical layer is concerned with the manufacturing process, cost's, suppliers, numerical control tool paths, and layout drawings to mention a few.

If PDES is to replace IGES in the future it would have to be compatible, so that IGES files could be translated to PDES.

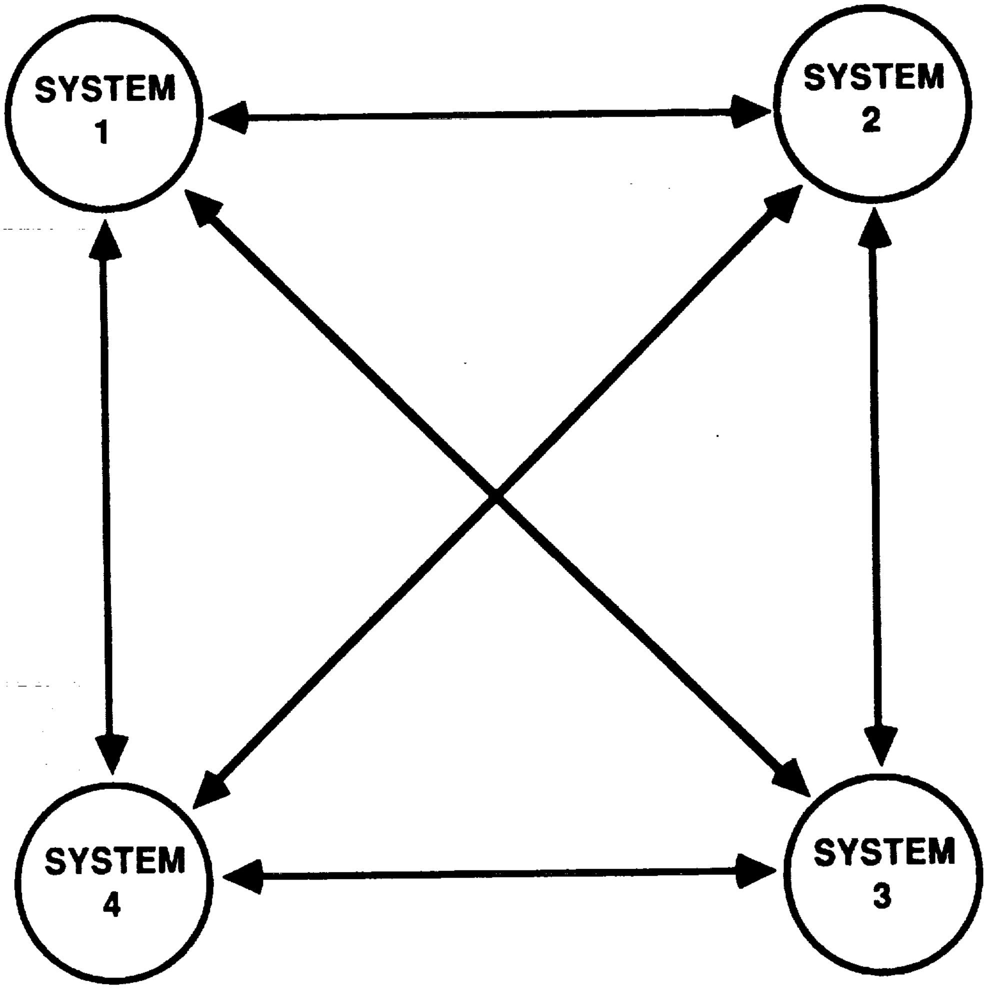


Figure 4 Direct Translation Process

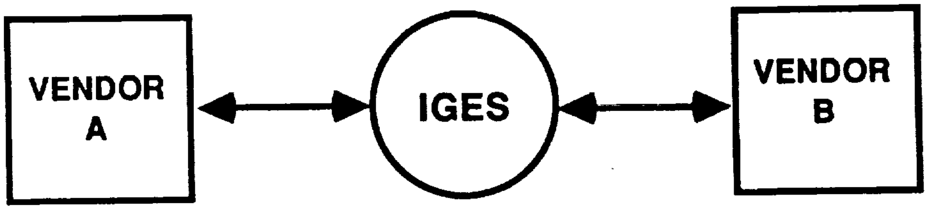


Figure 5 IGES Translation Process

X. TEST PROCEDURES

To evaluate an IGES translator one must perform several tests.

The IGES test files that are needed, are the following:

a. A test file that contains simple entities, mainly geometric to evaluate the basic translation process. These would be lines, points, circles, arcs, splines, etc and would provide a baseline.

b. Develop a test file with various entities, each enclosed in a box or separated. This would provide useful information on which entities transfer and also which native entity results.

c. A test file(s) that is a typical production part or schematic of a useful layout. This test file would be complex and give an indication of how reliable the translation process will be in the production environment. These file(s) should also include, if possible, a complex system that will be typical in the future.

There are various ways to evaluate the test results. One could compare plots through an overlay process, or a count of entities and their positions. This would give information as to how reliable the data is transferred and if in the same position. It is also important to see if the data elements can be manipulated. One could scale views, move geometric objects, place cells, edit text strips or dimensions, test to see if graphic groups are still graphic groups, etc.

More complex tests would be to test accuracy of curves, surfaces, and volume dimensions and postioning. This could be done for curves by creating a series of parallel lines through the curve and compare intersection points before and after translation. The same process could be used for surfaces and volumes by, respectively, using parallel planes and intersecting solids. These tests would be imperative if the drawing is used for analysis or direct measurements.

Any drawing that is translated will have to be verified that it corresponds to the original and validated, in the sense, that all functions will have to have been translated. This is no small task and has not been addressed thoroughly in this paper.

The quality of translation will most likely follow, in order of good to bad, the three tests outlined above.

XI. TEST RESULTS

Test translations were done with several CAD/CAE drawing packages with mixed results. The tests were performed with different levels of support and hence difficult to compare. Initially, simple geometrical parts were developed on the Intergraph CAD/CAE system and these were tested via a self-loop with success. Then a more complex part developed by an IGES test committee (ref. 11) was translated; as can be seen from Figure 6 and 7, the arrowheads and some attached text was lost, or mis-interpreted. Another self-loop test performed on the Intergraph system is shown in Figure 8. This test part was composed of lines and an arc mirrored. As can be seen from Figure 8, line fonts were mis-interpreted and a line was drawn through the arc endpoints. Figure 9 was a demonstration of how graphic group and cell entities were translated. In this case the graphic group was translated correctly but the cell capability was not translated. This was verified by bringing the design drawing up on the screen and then determining through menu commands if the circle and text was a graphic group or a cell.

The next suite of tests were for a drawing which contained 28 IGES entities, see Figure 10, and the Space Station. These IGES files were developed by NASA/Goddard, see ref. 12. The 28 entity file was translated by Intergraph (IGES version 8.8.5), AutoTrol series 7000 (on an Apollo platform), and the IBM CADAM package. The results of the 28 entity file are shown in Figures 11 and 12, for Intergraph and AutoTrol, respectively. The IBM CADAM system was unsuccessful in having the IGES file translated. The translation by Intergraph, see Figure 11, resulted in only one view, zero height text, and improper scaling. It should be noted that this was only accomplished after removing the B-spline entity from the design drawing, otherwise it killed the process. The translation by AutoTrol, see Figure 12, resulted in the four views being evident, but with some vector splash and certain entities missing, the main ones being surfaces of revolution. The AutoTrol drawings were translated with the help of an AutoTrol representative, while the Intergraph attempt was done by a design engineer. The 28 entity IGES file could not be translated by the IBM CADAM system.

The last IGES file translation attempted was for a very complex drawing. This is a drawing of the Space Station and the results of the translation by AutoTrol is shown in Figure 13. This translation is complete, since no translation errors were reported in the AutoTrol log. The translation by Intergraph resulted in only the border being displayed, and the IBM CADAM system was unable to translate. The translation of an IGES file containing solids entities was not attempted since the various CAD/CAE drawing packages either did not support solids, or could not translate the file (AutoTrol).

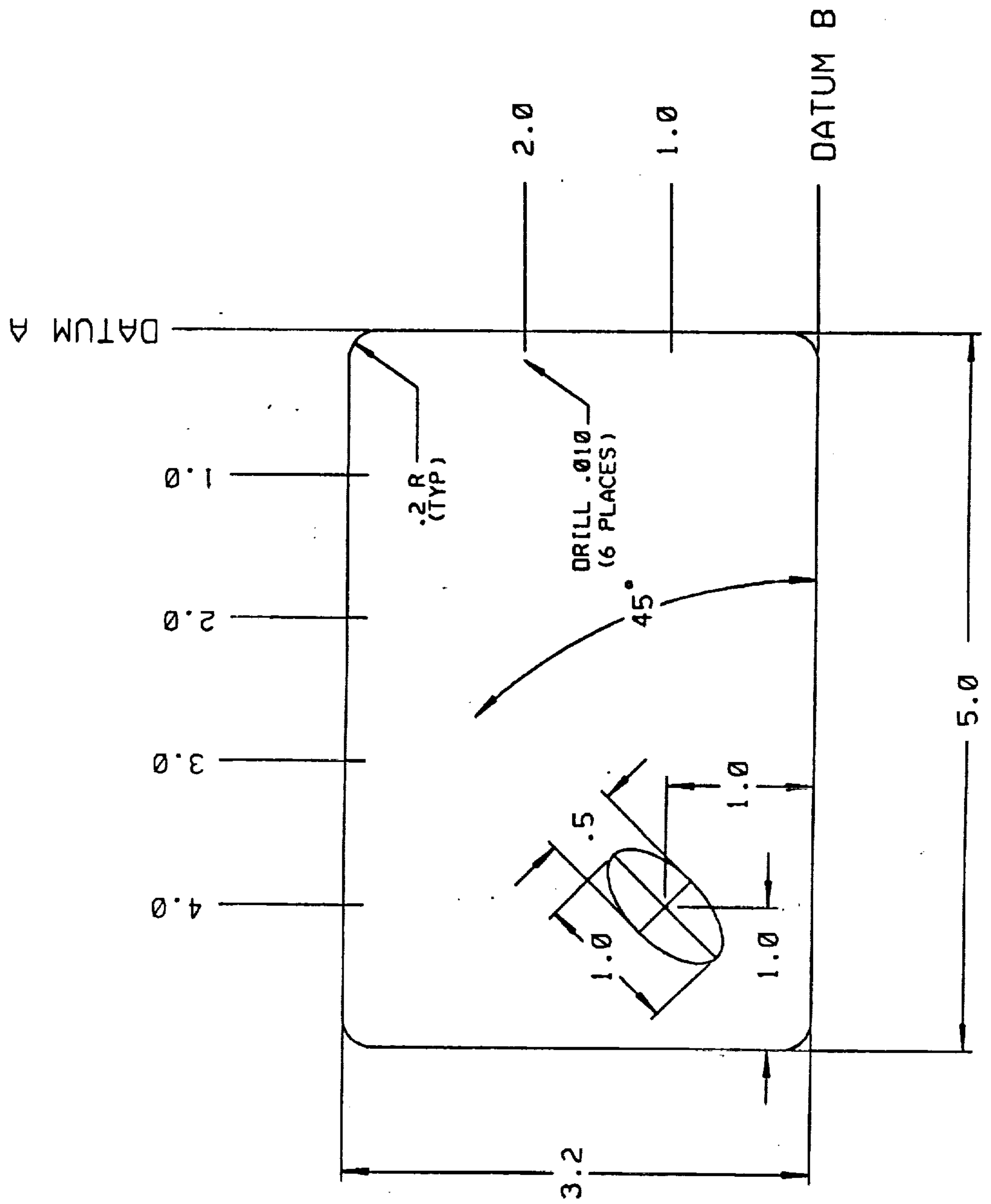


Figure 6 IGES Test Committee Part Geometry

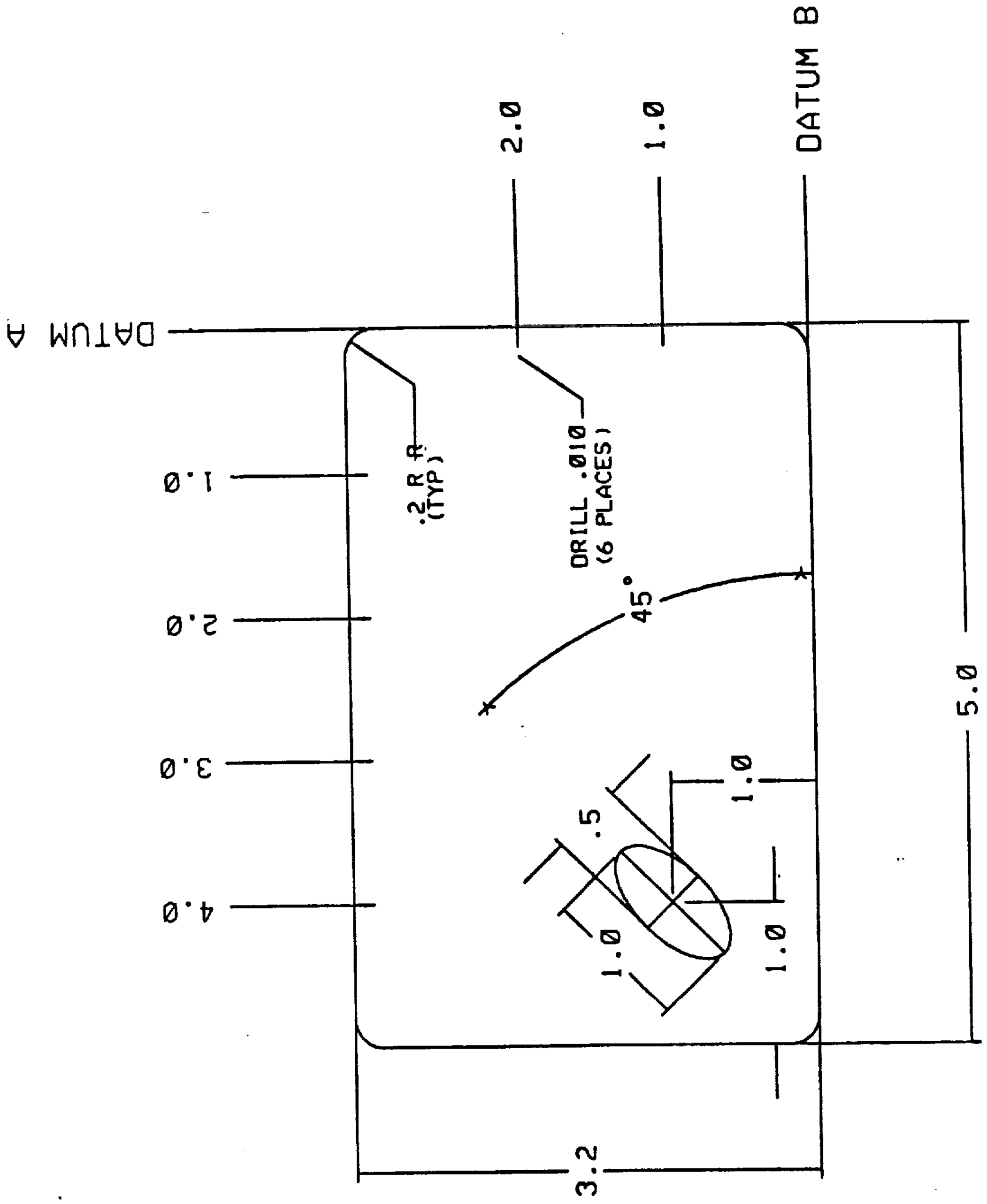
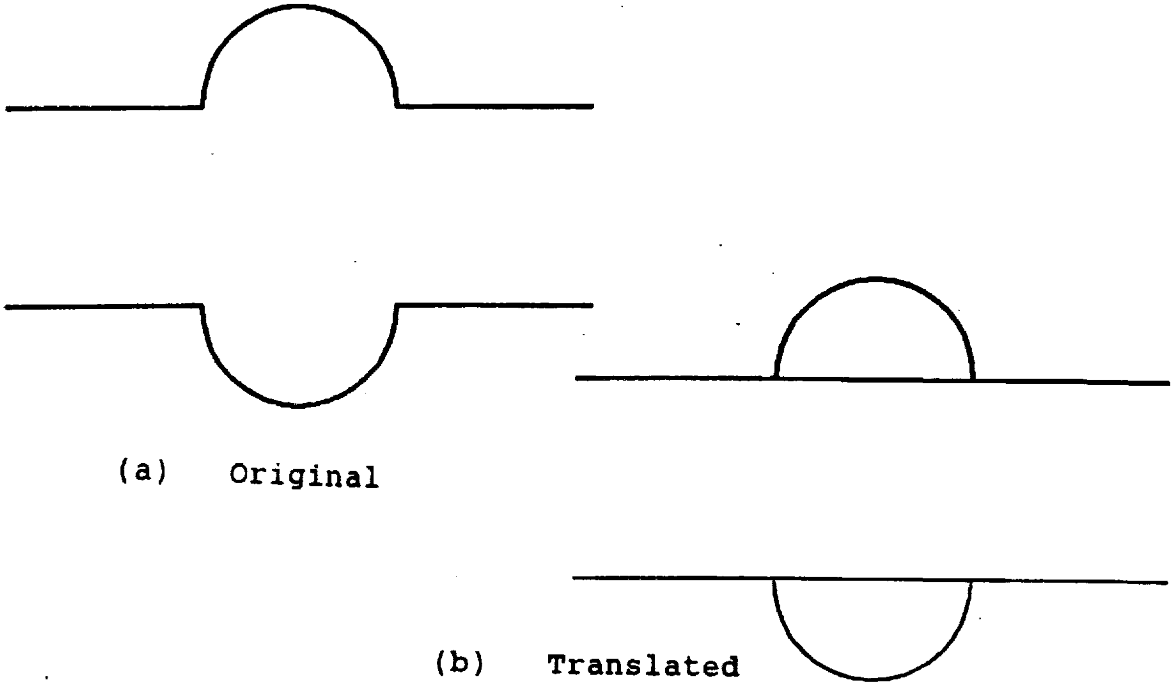


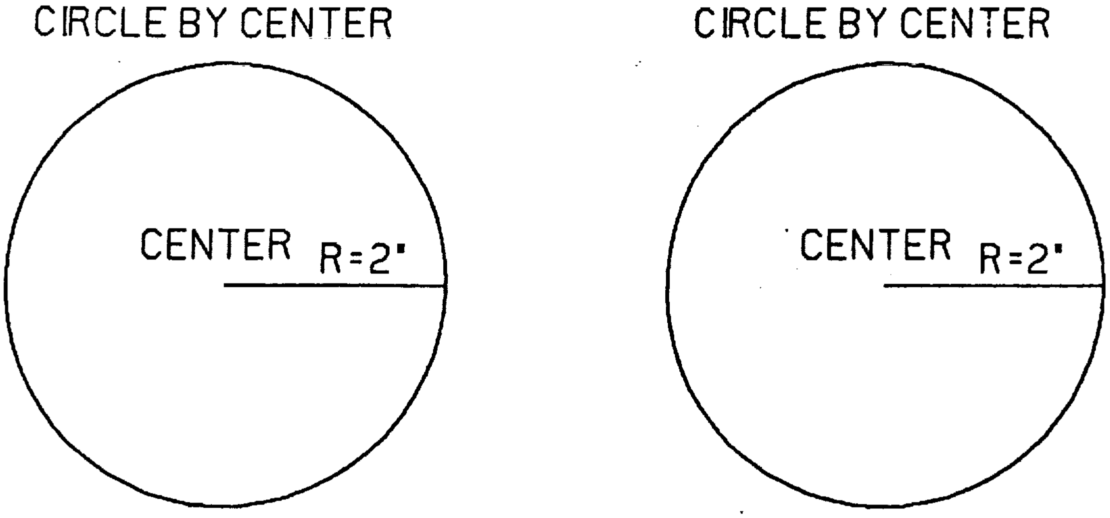
Figure 7 Translated Test Committee Part



(a) Original

(b) Translated

Figure 8 Mirrored Lines/Arcs Geometry



GRAPHIC GROUP

CELL

Figure 9 Translated Graphic Group/Cell Entities

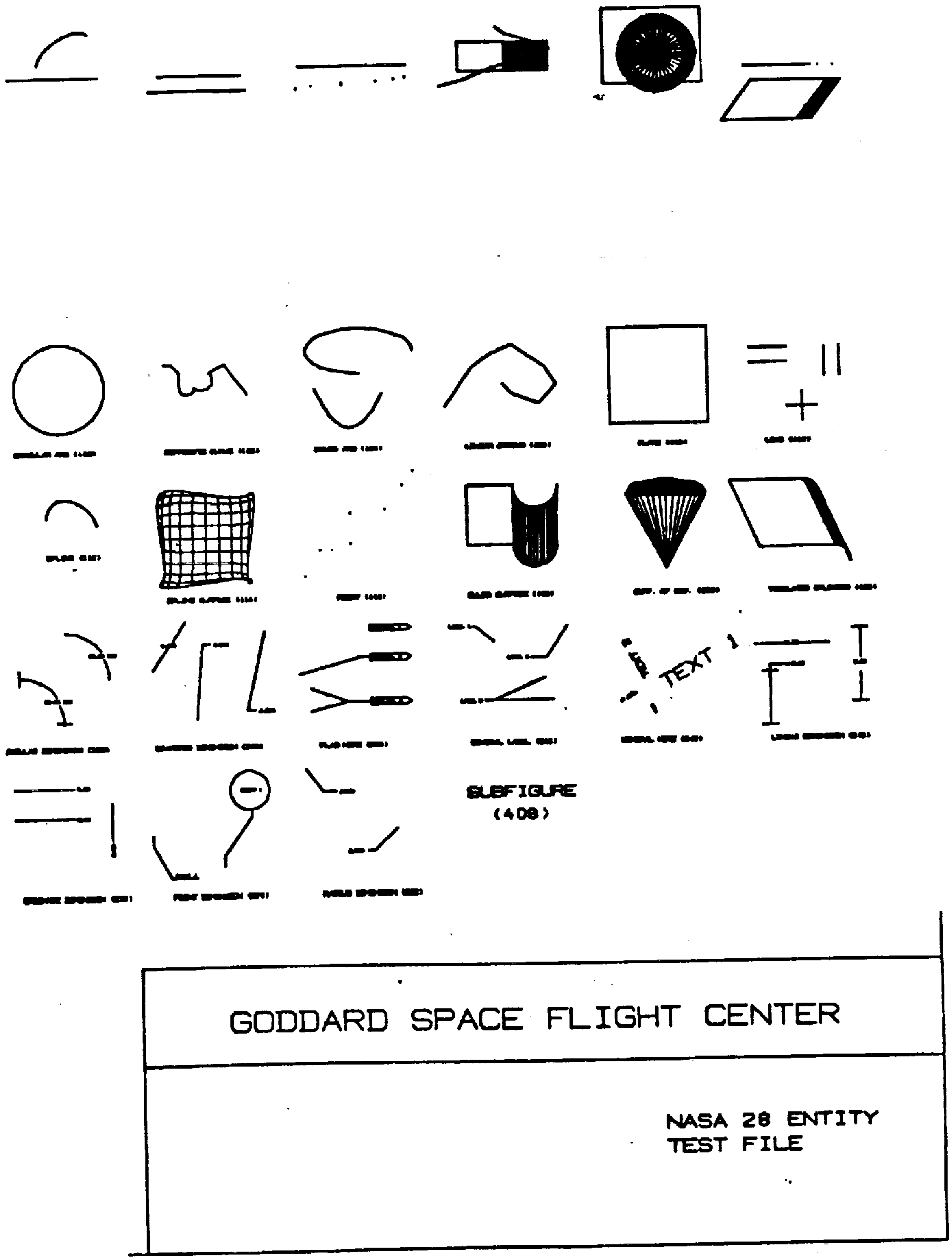


Figure 10 NASA/Goddard Four View 28 Entity File

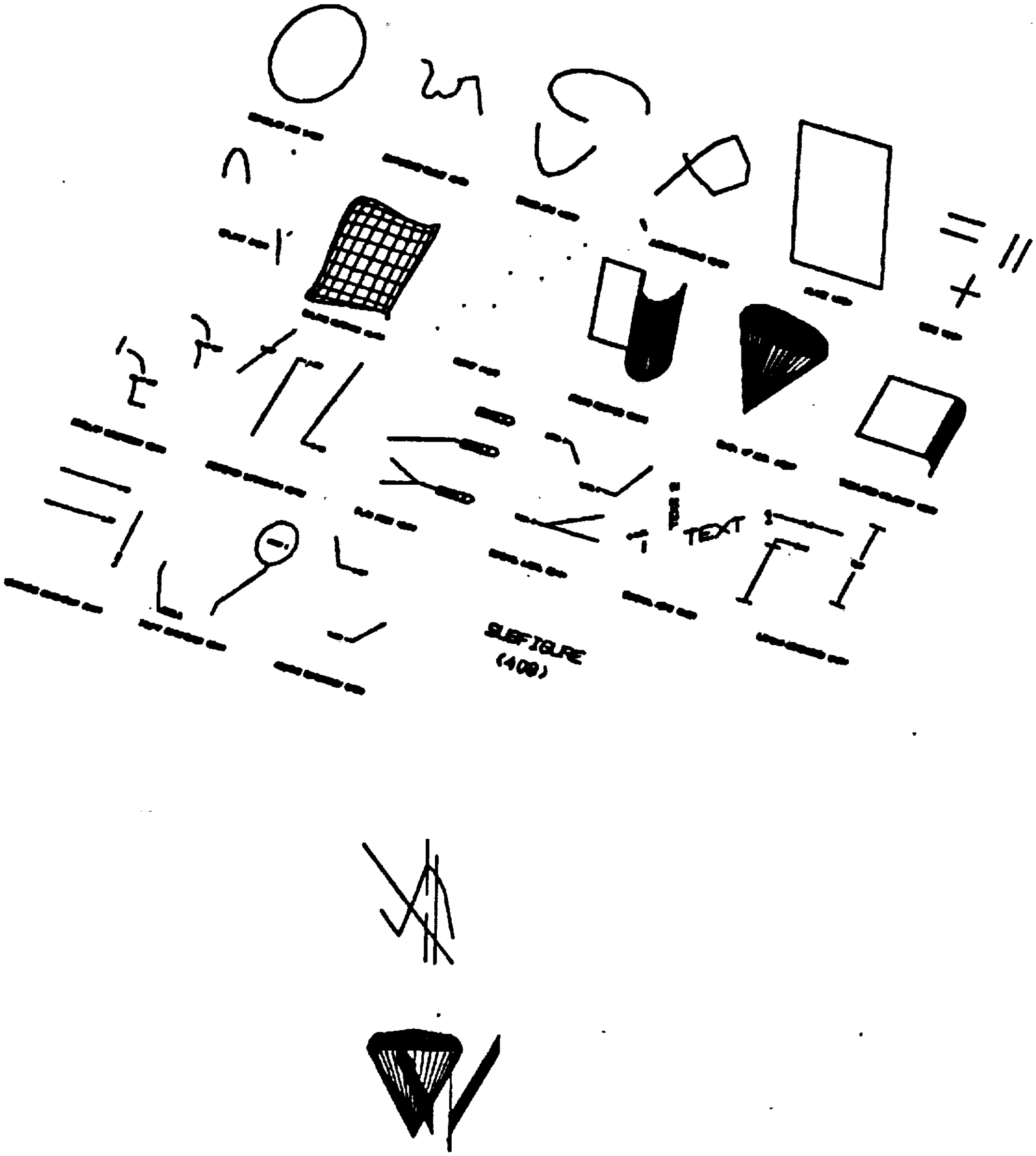


Figure 10 NASA/Goddard Four View 28 Entity File (Continued)

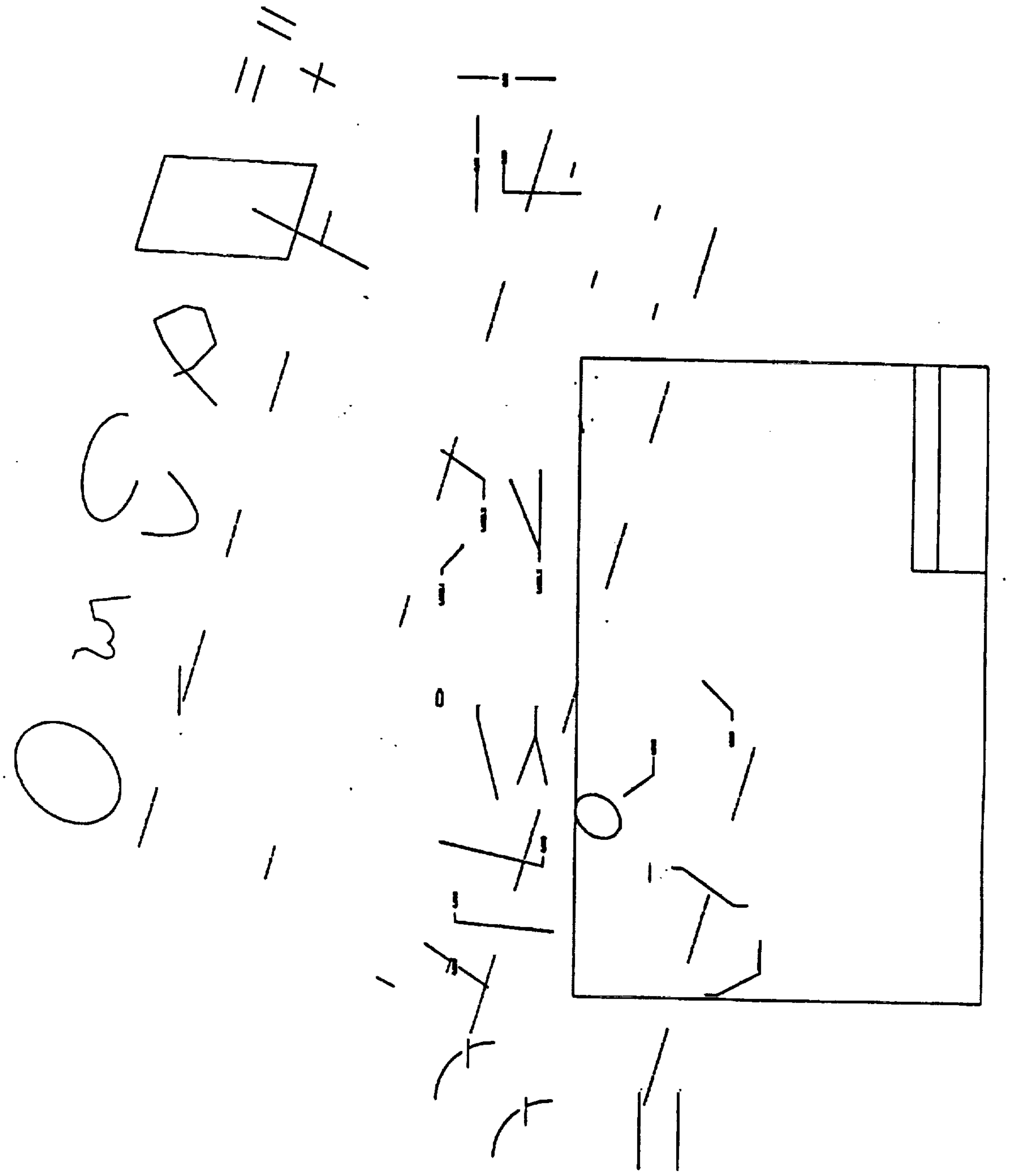


Figure 11 Translated NASA/Goddard 28 Entity File-Intergraph

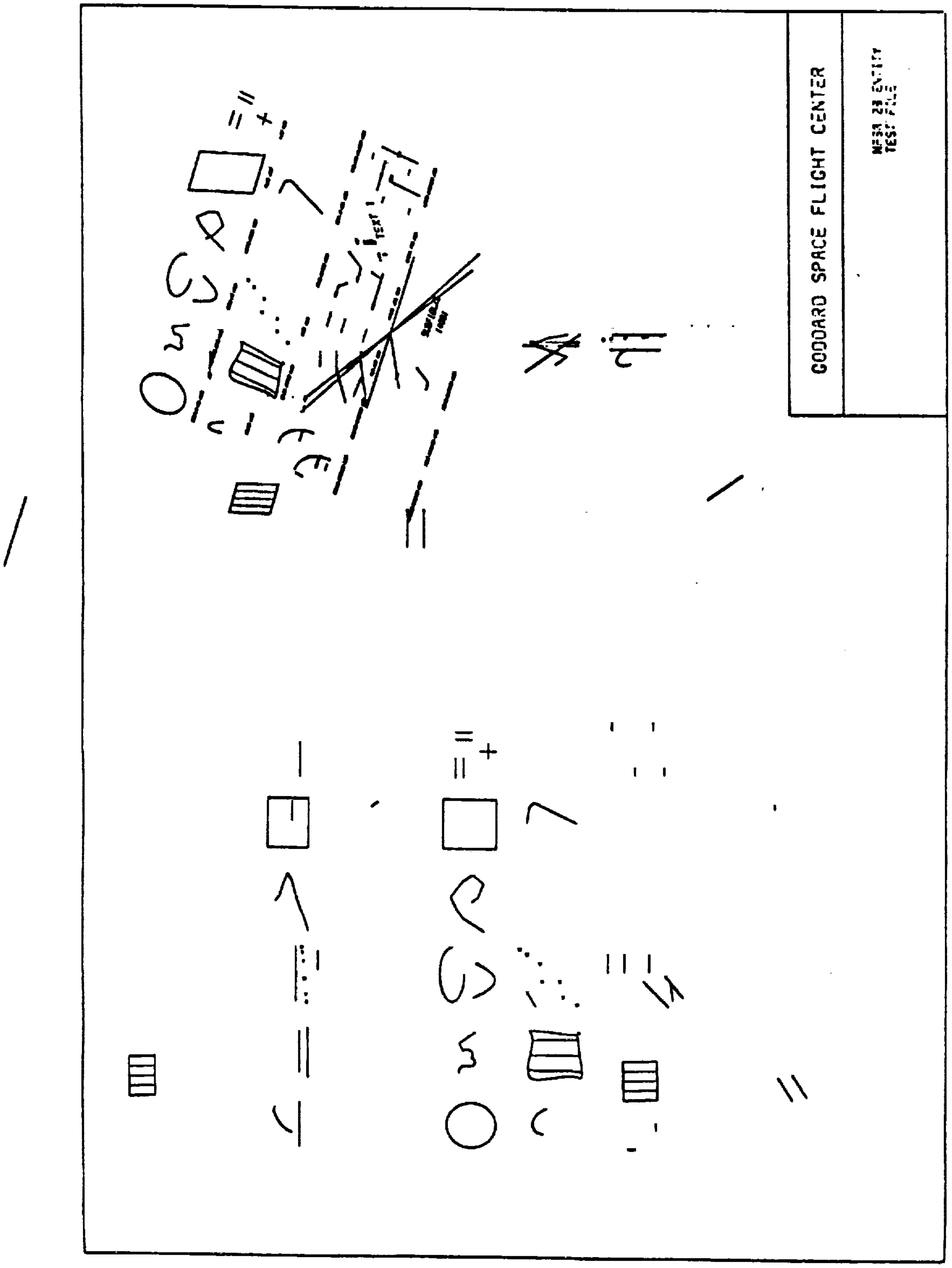


Figure 12   Translated NASA/Goddard 28 Entity File-AutoTrol

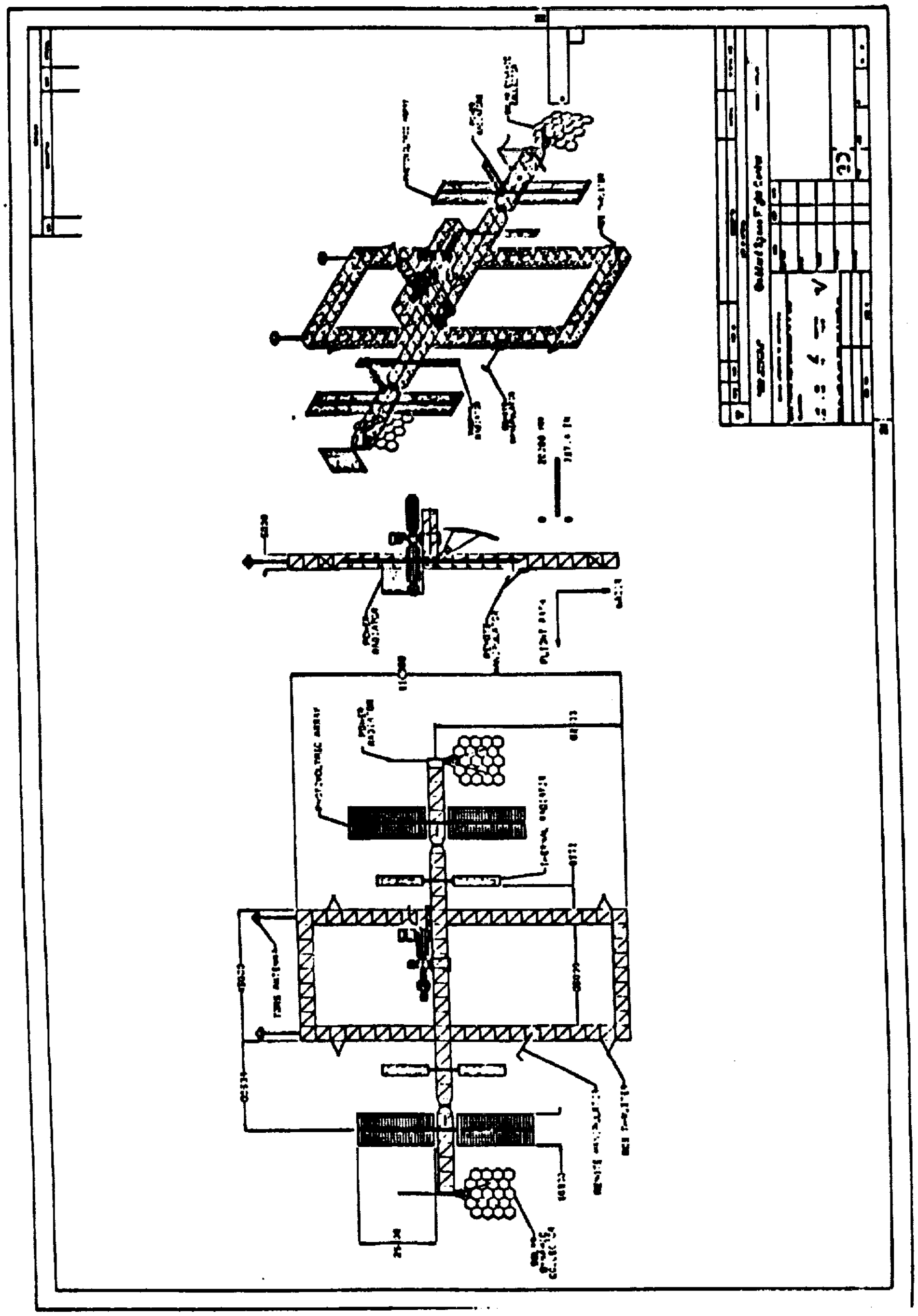


Figure 13 Translated NASA/Goddard Space Station-AutoTrol

ORIGINAL PAGE
BLACK AND WHITE PHOTOGRAPH

XII. POSSIBLE STRATEGIES

There are several strategies that can be implemented to increase the success rate of IGES translations.

One area is to develop user discipline in the design environment. The designer needs to understand the relationship between the product the end user perceives and the set of computing elements that represent that product. They should be disciplined to utilize neutral database standards. This is probably easier said then done, since by doing this one will restrict the user's inovativeness, efficiency, and interest. This would involve the development of an engineering IGES handbook (ref. 13). This handbook would include a primitive set of entities that could be used in a particular engineering discipline, restricted forms within that entity that should be used, and lists of native to IGES entity translations.

This disciplined approach could be rigidly enforced in certain applications, i.e., those that will require possible translation, now or in the future. These are files that must be maintained for many years and when re-used would probably be modified or appended.

One approach to this is to build an auxiliary procedure involving table look-up which will translate user commands into acceptable IGES entities. This would not normally be done on-line, but only if a translation is to be done. This approach has been taken by Sandia Laboratories in the concept they call vanilla deflavoring and reflavoring (ref. 9).

These flavor translators convert IGES data acceptable to the sending system into IGES suitable for the receiving system. This is a better approach than using ad hoc procedures for editing the drawing when translating from vendor A to vendor B. It also eliminates hand-editing the IGES file, although this would be very difficult due to the complexity of the file structure. One still has the problem that the flavor translator is only as good as the pre/post-processing done by the vendor.

An example of the flavoring concept is in the conversion of line fonts to system line fonts, or deflavoring. These can then be re-translated to the closest line font at the other end, reflavor. Another example would be to decompose a composite into it's component parts, if the other vendor does not support.

Another approach would be to employ the **bubble-up** technique. This would require one to translate a design file only when needed and then verify/validate file and re-do entities as needed to make it a workable drawing. Possibly, only re-working those portions that are needed. This approach is probably valid if one is moving from vendor A to vendor B and a large number of drawings are currently residing on vendor A. In conjunction with this, it might be acceptable to only scan in drawings and then modify the scanned drawing, as needed, at a workstation.

For the initial translation a viable alternative would be a one-to-one translator, especially, if it is a one time transfer and not one that is continually occurring between numerous dissimilar systems.

The most important strategy, if one is to purchase a system that is from a vendor different than the one presently available and if there are design files to be transferred to the new vedors system from the existing vendor, is to require that the vendor must successfully perform the tests in Section X. Only by requirements such as these will the vendors put more effort into developing efficient IGES translators. Although, it should be noted that translating files from an existing vendor's graphic design database is only as good as the available pre-processor.

Finally, one could absorb the cost of translation when going from vendor A to B. Until more vendors have efficient IGES translators one is probably doing this anyway through development cost pass-through, but as more procurements require efficient translators to be built this development cost should become less.

As a final note the Engineering Design organization should consider dedicating a person(s) to keeping up with and understanding the nuances of the translation process.

XIII. SUMMARY AND CONCLUSIONS

The translation process is a difficult one and should be planned for in any procurement process and on-going design environment.

One should view the IGES translation, or any automated translation process, as the first step in obtaining a viable design drawing. Probably, in practice one should be able to obtain 70 - 90% of the drawing transferred correctly. This assumes that the vendor has developed an efficient pre/post processor. If the vendor has not developed and maintained an efficient set of processor's there is little the user can do to enhance the translation process.

The experience gained from obtaining translated drawings for the different test classes follows what one might expect, i.e., the more complex the drawings are - the more difficult to translate, the more experienced technical resources that are available - the more successful the translation, and certain vendors have better pre/post processors than others.

The solution to the translation process is not easily solved since there are conflicting goals. The engineering design organization would like to have a homogeneous architecture, but this is impractical due to the following reasons; responsibility is normally distributed in a large design organization and competition among vendors results in enhanced products that are very attractive to the user. Therefore, one can assume that the design environment will be heterogeneous.

In conclusion, the design organization should make test translations part of the procurement, user's should be aware of IGES capabilities, design standards should incorporate IGES capabilities when drawings are to be maintained for many years or modified, and there should be a dedicated group (or, person(s)) involved in IGES translations and their nuances.

A final reminder, remember that an IGES translation environment is only as good as the pre/post processors developed by the vendor.

XIV. REFERENCES

1. Nagel, R. N., Braithwaite, P. R., and Kennicott, P. R., Initial Graphics Exchange Specification (IGES), Version 1.0, NBSIR 80-1978 (R), National Bureau of Standards, 1980.

2. Initial Graphics Exchange Specification (IGES), Version 2.0, NBSIR 82-2631 (AF), National Bureau of Standards, February 1983.

3. Smith, B. and Wellington, J., Initial Graphics Exchange Specification (IGES), Version 3.0, NBSIR 86-3359, National Bureau of Standards, April 1986.

4. Smith, B., Rinauldot, G. A., Reed, K.A., and Wright, T., Initial Graphics Exchange Specification (IGES), Version 4.0, NBSIR 88-3813, National Bureau of Standards, June 1988.

5. Zobrist, G. W., "CAD Program for Automated Motor Control Center Layout", Proc. of Autofact West-CAD/CAM VIII, Los Angeles, CA, November 1980.

6. Keller, C., "IGES gets an F", MICROCAD News, September / October 1988, PP. 60 - 65.

7. Beazley, W. G., "Expert System Rules from CAD Databases", Proc. Air Force Workshop on Artificial Intelligence for Integrated Diagnostics, University of Colorado, Boulder, CO, July 29 - 31, 1986, PP. 292 - 297.

8. Liewald, M. H. and Kennicott, P. R., "Intersystem Data Transfer via IGES", IEEE Computer Graphics and Applications, Vol. 2, May 1982, PP. 55 - 58.

9. Ames, A. L. and Fletcher, S. K., " Making IGES Work for CAD/CAM Data Exchange", Proc. Autofact 1985, Detroit, MI, November 4, 1985.

10. Kelly, J. C., "The Product Data Exchange Standard (PDES), Federal Computer Conference, September 9, 1985.

11. Smith, B. M. and Wellington, J., "IGES, A Key Interface for CAD/CAM System Integration", CAD/CAM Technology, Vol. 3, Spring 1984.

12. Purves, L. R., NEXUS/NASCAD - NASA Engineering Extendible Unified Software System with NASA Computer Aided Design, NASA Goddard Space Flight Center, MD, M86-1209, 1986.

13. Hammond, D. and Wesenberg, R., "Progress Report: Development of an IGES Handbook for KSC", Unpublished Technical Summary, Engineering Development Directorate, NASA/KSC, FL, September 1986.

NASA
National Aeronautics and Space Administration

Report Documentation Page

1. Report No. CR-166837	2. Government Accession No.	3. Recipient's Catalog No.

4. Title and Subtitle
1989 Research Reports
NASA/ASEE Summer Faculty Fellowship Program
Kennedy Space Center - University of Central Florida

5. Report Date
October 1989

6. Performing Organization Code

7. Author(s)
Dr. E. Ramon Hosler, Editor
Mr. Dennis W. Armstrong, Editor

8. Performing Organization Report No.

10. Work Unit No.

9. Performing Organization Name and Address
NASA - Kennedy Space Center
Systems Training and Employee Development Branch
Kennedy Space Center, Florida

11. Contract or Grant No.
NASA Grant NGT-60002
Supplement 2

12. Sponsoring Agency Name and Address
NASA Headquarters
University Programs Branch
Washington, D.C.

13. Type of Report and Period Covered
Final Contractor Report
6/89 - 8/89

14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

This contractor's report contains all sixteen final reports prepared by the participants in the 1989 Summer Faculty Fellowship Program. Reports describe research projects on a number of different topics.

17. Key Words (Suggested by Author(s)) Human-computer interface software, metal corrosion, rocket triggered lightning, automatic drawing, 60-Hertz power, carotid-cardiac baroreflex, acoustic field, robotics, AI, CAD/CAE, cryogenics, titanium, flow measurement

18. Distribution Statement
Unclassified - Unlimited

19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of pages 418	22. Price