NASA-CR-169195
19820022361

The Telecommunications and Data Acquisition Progress Report 42-69

March and April 1982

N.A. Renzetti
Editor

June 15, 1982

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Preface

This publication provides reports on developments in Earth-based radio technology with applications to several programs. In space communications it reports on activities of the Deep Space Network. In geodynamics it reports on the application of radio interferometry at microwave frequencies for geodynamic measurements. In the field of astrophysics the Deep Space Stations individually and in pairs as an interferometer have been applied to direct observation of celestial radio sources.

Each succeeding issue of this report will present material in some, but not necessarily all, of the following categories:

Radio Astronomy
Radio Interferometry at Microwave Frequencies

- Goedetic Techniques Development
- Spacecraft Navigation
- Orbiting Very Long Baseline Interferometry

Deep Space Network

- Description
- Program Planning
- Advanced Systems
- Network and Facility Engineering and Implementation
- Operations
- Spacecraft Radio Science
- Planetary Radar

In each issue, there will be a report on the current configuration of one of the seven DSN systems (Tracking, Telemetry, Command, Monitor and Control, Test Support, Radio Science, and Very Long Baseline Interferometry).

The work described in this report series is either performed or managed by the Telecommunications and Data Acquisition organization of JPL.

Contents

Deep Space Stations

OPERATIONS

Network Operations

Deep Space Stations

ENERGY

Radio Astronomy

R. D. Shaffer
Mission Control Center Section

S. Gulkis
Planetary Atmospheres Section

This article reports on one Radio Astronomy activity supported by the Deep Space Network: namely, use of the Tidbinbilla Interferometer to refine the source positions in the Parkes 2.7-GHz survey of the southern sky. A result of the first phase of this work was the identification of a quasi-stellar object which appears to be the most remote object yet observed.

The 34- and 64-m antennas at the Tidbinbilla Deep Space Communications Complex (DSCC) can be configured as a real-time interferometer operating in the 13-cm band. The system parameters are as follows:

Operating frequency:	2.3 GHz
Predetection bandwidth:	12.0 MHz
Baseline:	200.0 m (north-south)
Lobe separation:	2.3 arc min
IF frequency:	70.0 MHz
Sensitivity:	1.6 mJy/hr (5 X rms)

Funding for the equipment necessary to create the Tidbinbilla Interferometer was granted by the Laboratory Director's Discretionary Fund (DDF) in 1972. Once the system was built and its capability demonstrated, funding for its operation and for the reduction of data was provided by the Office of Space Science and Application (OSSA 188-41-55-16). The Tidbinbilla Interferometer is used primarily to accurately measure the positions of southern hemisphere radio sources. The accurate positions thus obtained then allow astronomers to identify the optical counterparts of the radio sources.

The primary catalog of radio sources used at Tidbinbilla is the Parkes Radio Observatory's survey of the southern sky at 2.7 GHz. The source positions in the Parkes Survey are accurate to within 15 arc seconds. The use of interferometric techniques allows more accurate position determinations than can be made with a single antenna. With the Tidbinbilla Interferometer, the source positions from the Parkes Survey can be determined within 1-2 arcseconds. Thus, when an astronomer inspects a patrol plate taken by a widefield camera, he or she must inspect only 1% as much of the plate with the Tidbinbilla results as would be necessary with the unrefined Parkes Survey data. This allows for unambiguous source identification in approximately 85% of all cases.

The April 1982 issue of *Astrophysical Journal* contains an article entitled "2.3 GHz Accurate Positions and Optical Identifications for Selected Parkes Radio Sources," by S. Gulkis (JPL), D. L. Jauncey and M. J. Batty (CSIRO), and A. Savage (Royal Observatory). This article contains the

refined positions of 74 sources from the Parkes Survey. Even before publication, A. Savage, A. Wright (CSIRO), and B. Peterson (Australian National Observatory) were making optical observations of many of the objects identified in the article.

One object, PKS 2000-330, was of particular interest because it is approximately 15 times as bright in the red as it is in blue light. On 25 March 1982, a spectrogram of this quasi-stellar object (QSO) revealed that the normal positions of spectral lines are red-shifted by the largest amount ever recorded.

The red shift is given by:

$$Z = \frac{\lambda - \lambda_0}{\lambda_0}$$

where λ is the observed wavelength of the spectral lines, and λ_0 is the wavelength of the same spectral lines in a laboratory frame of reference. The value of Z for PKS 2000-330 is 3.78, surpassing the previously recorded 3.53 by a wide margin.

Most astonomers believe that the large red shifts exhibited by QSOs are caused by enormous velocities of recession, i.e., that the radiation from QSOs is being Doppler-shifted toward the red by the fact that the QSOs are receding from us at velocities which are a large fraction of the speed of light. In the case of PKS 2000-330, its velocity of recession is equal to 91% of that of light. On a cosmological scale, an object's distance from the observer is thought to be proportional to its velocity of recession. This is known as Hubble's law. Using a proportionality constant of the 70 km/sec/10^6 parsec, a widely used value of the Hubble constant, PKS 2000-330 appears to be 12 billion light-years away. In any case, PKS 2000-330 appears to be the most remote object yet observed.

ORION Downconverter and Power Supply

H. G. Nishimura
Radio Frequency and Microwave Subsystems Section

The receiver subsystem supplies the front end assembly (downconverter) and power supply for the ORION Mobile Station. These assemblies are designed to withstand severe environmental conditions. This article discusses the mechanical, electronic, environmental and maintenance design considerations encountered during the design phase of this project. The two channel S/X downconverter has a 400-MHz bandwidth channel. Phase stability of 2 and 7 deg at S- and X-bands, respectively, has been achieved with a temperature stabilized first local oscillator.

I. Introduction

The ORION Receiver Subsystem is a part of the Crustal Dynamics Project's ORION Mobile Station. The ORION Mobile Station is the mobile portion of the ORION network which uses VLBI techniques for determining earth crustal deformations. Radio energy from extragalactic star sources is simultaneously recorded at two sites to determine through interferometry the vector length and direction between the sites.

This article discusses the design of the Receiver Subsystem's frequency downconverter. (The downconverter is a low phase drift device which converts incoming frequencies at S- and X-bands to intermediate frequencies.) The downconverter is capable of operating in desert as well as arctic environments. During nonoperational transport, the receiver must survive air shipment and road-induced vibrations. An ORION downconverter is currently being fabricated that will meet these requirements (see Fig. 1).

II. General Description

The receiver subsystem consists of two separate assemblies: (1) the downconverter which contains the electronics necessary for converting the incoming S- and X-bands frequencies to the intermediate frequencies, and (2) the power supply, which provides all of the ac and dc requirements of the downconverter. The power supply also supplies dc voltages to the Phase Calibration Subsystem.

The downconverter and power supply are designed for outdoor mounting with protection from the elements; the equipment package is bulkier than the usual electronic assemblies used in a protected environment. Both assemblies use wall-box types of enclosures. They are part of the antenna transporter mobile unit. During operation they are mounted above the elevation bearings of the microwave dish antenna. However, while being transported between measuring sites, the equipment is stowed on the transporter bed.

Phase stability is one of the important requirements of the downconverter. To achieve good stability, the first local oscillators operate within a temperature-controlled environment. Controlled monitoring of the vital functions is provided to the ORION system computer. The maintenance concept is simplified through the use of appropriate LED failure indicators within the enclosures. Components subject to field replacement are readily accessible and have been designed for ease of replacement.

To overcome the effects of desert temperatures and solar loading, thermoelectric (T/E) coolers are used in the downconverter.

III. Design Discussion

A. Layout and Packaging

The packagings for the downconverter and power supply are similar in that they both utilize Hoffman wallbox enclosures and rear-mounted heat dissipating fins (see Figs. 2 and 3). Beyond these similarities, the layouts and packaging methods are quite different.

In the power supply the individual power supply modules are mounted directly to the flat side of the heat dissipating fins which form the floor of the enclosure. This provides a low center of gravity as well as a good thermal conductive path from the power modules into the fins.

The downconverter layout is based upon the following considerations: serviceability, component density, temperature-sensitive components and rf grounds. The resultant layout consists of three oven assemblies in one sector of the enclosure, all mounted upon a common subplate. The multipliers within the ovens are temperature-stabilized by T/E coolers. The balance of the enclosure floor area has three vertically mounted component plate assemblies which contain the rf components for accomplishing the downconversion. The amplifiers and passive components, which are not as susceptible to temperature as the multipliers, are mounted on these plate assemblies. The vertically mounted plates provide adequate thermal conduction and convection cooling for the heat from the mounted components.

B. Electronic Considerations

The electronic specifications of the downconverter are listed in Table 1. One of the design objectives was the utilization whenever possible of easily obtainable, commercial off-the-shelf electronic components, thereby minimizing in-house development and in-house manufacturing costs. As a result, the downconverter contains only three JPL-developed low-cost components. The usual specially manufactured semirigid hardlines were replaced with commercial Gortex flexible cable assemblies which are superior in phase stability and attenuation characteristics to their conventional semirigid counterparts. Despite their higher initial cost, the new cables eliminate costly documentation and fabrication and are much easier to handle in the field.

Phase instability is primarily introduced into the downconverter by the first local oscillator multipliers, the times twenty (X20) and the times eighty-one (X81) assemblies. The other components, such as amplifiers, mixers and cables contribute little to the overall instability. In order to achieve required overall stability, component ovens in conjunction with T/E coolers are used to control the temperature of the X20 and X81 multipliers. Temperature design is discussed in more detail in the Environmental section of this report.

A wide bandwidth (400 MHz) system when used with conventional coaxial cable suffers from gradual amplitude decrease at the high frequency end of the band due to the attenuation characteristics of the cable. In the Orion project, this cable is about 200 feet long. A compensating network of the same but opposite slope as the cable is required to maintain a constant amplitude IF input signal at the data acquisition assembly at the far end of the cable. A cable equalizer was designed for operation between 100 and 500 MHz, using a single section of a high pass network with an attenuation slope of −5.5 dB per 400 MHz. The circuit maintains a constant resistance of 50 ohms at both ports across the band.

Phase calibration tones are injected into the receiving channels in the microwave subsystem. These tones are sent through together with the noise signals and appear within the downconverter channels as a picket fence spectrum. The tones periodically produce a peak power that is much greater than the noise power of the signal. For this reason it is necessary to maintain the noise level well below the 1-dB saturation point of the downconverter. In order to minimize the intermodulation products caused by the amplitude distortion near saturation, it is also desirable to maintain the noise power at a safe margin below the 1-dB saturation point. Because of their combined effects, the noise level in the S-band channel has been designed to remain at 27 dB below the 1-dB point, and 32 dB below in the X-band channel.

C. Monitor and Control

The downconverter and the power supply provide analog data indicating the operating parameters, such as output levels, phase lock and internal temperatures. This data is hardwired to

the Monitor and Control subsystem. The following list identifies all the monitored data:

Monitored data

Downconverter

IF output levels, S- and X-bands

×20, ×81 and ×5 multiplier output levels

Internal temperatures of enclosure and ovens

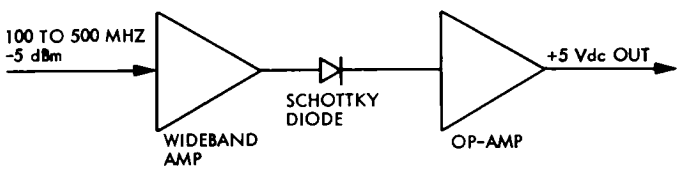
Multiplier phase lock

Power supply

Output voltages of individual power modules

Internal box temperature

Commercial power level packaged detectors usually require about a +7 dBm input level in order to obtain a +0.4 Vdc output level. For continuous IF level monitoring, a more sensitive detector was required. A Schottky zero bias diode was used in a new design to provide the necessary level to the Monitor and Control Subsystem. In the design, the diode is preceded by a wideband IF amplifier for additional sensitivity. The diode output was connected to a dc op-amp for level amplification. Sensitivity is better than -30 dBm, although -5 dBm is all that is required.



Internal temperatures are monitored within the downconverter and power supply by mercury thermostats which are preset by the manufacturer. At the predetermined temperature the thermostat contacts either open or close. The contacts are connected to TTL level pull-up resistors. These levels are sent through the Monitor and Control Subsystem to the system computer as warning indicators. In the power supply one of the sensors automatically shuts down the entire power supply when the internal temperature reaches a danger level. A power relay is used for this purpose.

D. Environmental Considerations

Equipment is designed to operate under the following environmental conditions:

Conditions	Operational specifications
Temperature	-29 to 45°C (-20 to 113°F)
Wind speed	48 km/hr (20 mph)
Relative humidity, maximum	100% at 46°C (115°F)

The downconverter and power supply require protection from the environment. The Hoffman aluminum enclosures provide adequate protection from blowing sand and rain. Unlike the downconverter, the power supply can withstand the temperature extremes without excessive degradation of its output parameters. Therefore, the power supply design is, relatively speaking, more straightforward than the downconverter.

The basic mode of heat removal on the power supply is by conduction and forced air convection. The floor of the enclosure was replaced with a finned heat dissipator. To prevent the external fins from absorbing solar load, a shroud was built over the fins to act as a sun shield. The shroud also helps to distribute the air from a blower mounted at one end of the shroud. The heat dissipator is mounted to the enclosure with screws instead of weld in order to prevent the welding heat from distorting the machined inner mounting surface of the aluminum dissipator. The individual power modules are mounted on the inner or smooth surface of the dissipator. The heat is conducted from the base of the power modules into the heat dissipator. Forced air convection removes this heat from the fins into the atmosphere.

The downconverter design utilizes, in addition to the cooling methods of the power supply, thermoelectric (T/E) coolers and a special subplate for the purpose of mounting the multiplier ovens. The subplate, which is smaller than the main baseplate, requires fewer T/E coolers than a bulky baseplate. The three frequency multipliers (two for the downconverter and one for the Phase Calibration Subsystem) require precision temperature control. Each is contained within a component oven slightly larger than the multiplier. Individual ovens are preferred from a maintenance viewpoint. The oven cavity temperature is maintained to within ±0.1°C of the 48°C set temperature by a proportional controller energized by dc. All three ovens are mounted on the common subplate, which is thermally isolated from the baseplate or any part of the enclosure. Six T/E coolers are sandwiched between the subplate and the external finned heat dissipator. With the T/E coolers connected in the cooling mode, the heat from the subplate is transmitted to the fins. Like the power supply and its shroud, a blower on the downconverter blows external air across the fins to dissipate the heat to the atmosphere. The heat dissipating efficiency is preserved by mounting the downconverter on

the movable antenna structure in such a way that the fins are never below the main enclosure. This prevents the warm air coming off the fins from reheating the enclosure.

Figure 4 shows the main elements of the oven design. The oven is made from a standard component oven which is rated for a cavity load of 2.5 watts.

In order to keep the oven volume as small as possible, the smallest available oven was used. The same oven was selected for the S- and X-band and Phase Calibration Subsystem multipliers. The design approach consisted of sufficiently modifying the standard ovens so that their cavities would be capable of dissipating from 4.5 to 6.5 watts of load. As discussed in the previous paragraph, T/E coolers, heat dissipating metal fins and forced air cooling methods were used. The subplate on which the ovens are mounted is maintained at 40 ±3°C. The thermal gradient from the subplate to the multiplier is accomplished with a metal shim of low thermal conductivity such as stainless steel.

Heat removal from the downconverter is a concern only at the higher ambient temperatures, e.g., 15°C (59°F) to 45°C (113°F). At temperatures below 15°C heat must be added to maintain the oven temperatures. Two methods of heat addition are employed: (1) polarity reversal of the T/E coolers, and (2) ac-powered path heaters. Although the T/E coolers function well as heaters with essentially the same efficiency as in the cooling mode, their heat adding capability is insufficient as the ambient temperature approaches the colder end of the operating range. The patch heaters turn on when the subplate can no longer be maintained within the 40 ±3°C range. The patch heaters are equipped with a proportional controller. All other controls are accomplished with preset mercury thermostats and relays.

Condensation is expected to become a problem when the temperature of the trapped humid air within the downconverter is lowered to its dew point. The resulting water droplets can cause rapid oxidation, resulting in poor contacts, especially at the rf connectors. For protection against condensation, a dessicant of silica gel is mounted in a pocket on the inside of the downconverter cover. The silica gel crystals are individually contained in cloth bags.

E. Maintenance

Maintenance philosophy was tailored for mobile requirements. Equipment servicing while on the road consists of no more than simple replacement of modules or assemblies. Although no time limits were specified for servicing, replacement time of 30 minutes was the design goal. Any maintenance requiring more than 30 minutes would be performed at the ORION depot. There is no soldering during field servicing. Connectors and screw terminals were extensively designed into the equipment to facilitate servicing. Commercial components with solder terminals were redesigned to accept screw terminals and connectors. LED-type indicators are employed to quickly pinpoint the problem areas. Desiccant was purchased in cloth bags to simplify replacement. Terminal board junction points were eliminated in favor of internal wire splices to reduce the package size. Troubleshooting time was decreased by wiring all critical internal test points to a 61-pin test connector located on the assembly bulkhead. At this connector measurements can easily be made with a simple external test fixture.

The power supply and downconverter are mounted during operation on the antenna structure. The enclosure covers and sun shield are easily removable for servicing these assemblies, which may at times be at difficult angles. On both the power supply and downconverter, with the removal of the covers, the internal components are readily accessible for servicing. Each oven is designed in a single assembly, called an oven cover, which is removed from the subplate by thumb-release latches. The desiccants, which may require replacement on a once-a-year basis, are equipped with a visual color type of indicator to monitor the moisture content within the downconverter enclosure. The frequency of desiccant replacement depends on how often the covers are removed and the humidity of the operating environment. The environment within the downconverter is a closed system with no outlets to the atmosphere.

IV. Testing and Implementation

The receiver tests will commence in May 1982. The multitude and length of some of the anticipated tests are expected to make the testing phase a very important part of the overall program. Because of the very short time allowed before the receiver is integrated with the rest of the ORION system, some of the tests will be postponed until later. Tests will be conducted across the ambient temperature of -29 to 45°C and across the frequency band of each IF. Each channel of the downconverter will be tested independently. Thermal stability will be checked at the multipliers and the subplate on which they are mounted. Phase stability, phase linearity and phase jitter will be checked as a function of temperature and/or frequency.

The power supply will be tested for input step voltage variations, output ripple, and output stability with load variations.

Implementation of the receiver subsystem is expected to commence in July 1982. Manufacturing documentation will be updated at this time to reflect the final tested model.

Table 1. Downconverter electronic specifications

Input characteristics	Performance
Frequency range	S-band: 2220 to 2320 MHz X-band: 8200 to 8600 MHz
Signal level	
S-Band	-65.2 dBm ±0.4 dB
X-band	-66.8 dBm ±1.2 dB
Reference signal	100 MHz, +5 dBm
Noise figure	S-band: 280 K (2.9 dB) X-band: 400 k (3.8 dB)
AC voltage	117 Vac, 60 Hz, single phase
AC power	800 W, max.
Output characteristics	**Performance**
S-band IF	300 ±50 MHz
X-band IF	300 ±200 MHz
Monitor and control interface	RS-232C
Phase stability	S-band: $< 2°$ drift, temp stabilized
	X-band: $< 7°$ drift, temp stabilized
IF amplitude stability	S-band: ±1.0 dB from 250 to 350 MHz
	X-band: ±1.5 dB from 100 to 500 MHz

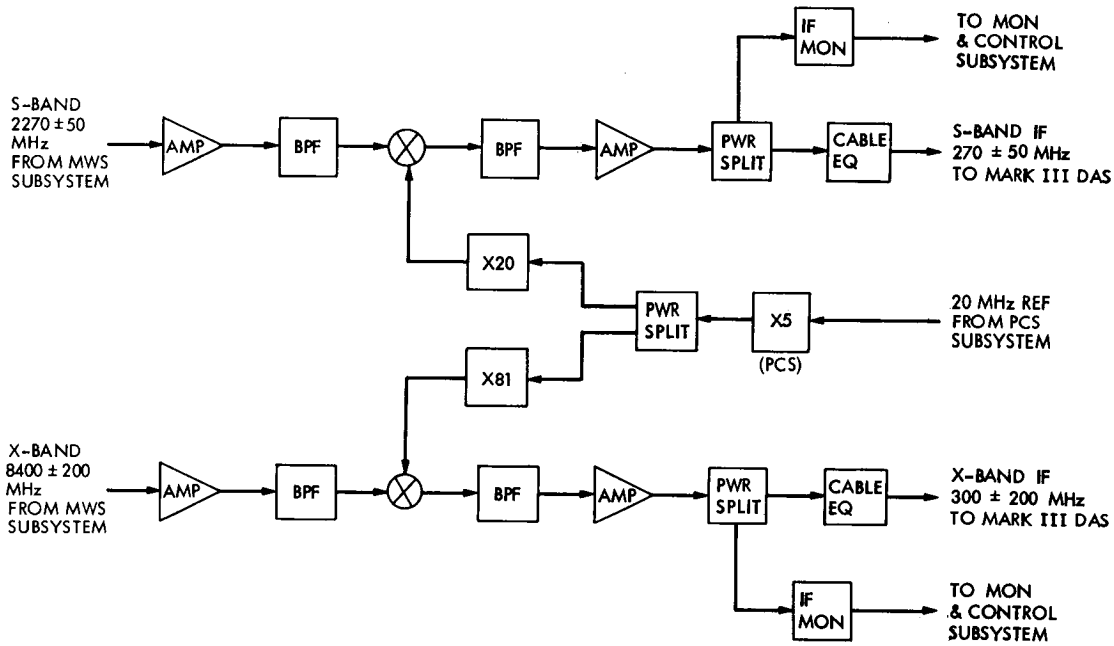


Fig. 1. Downconverter block diagram (simplified)

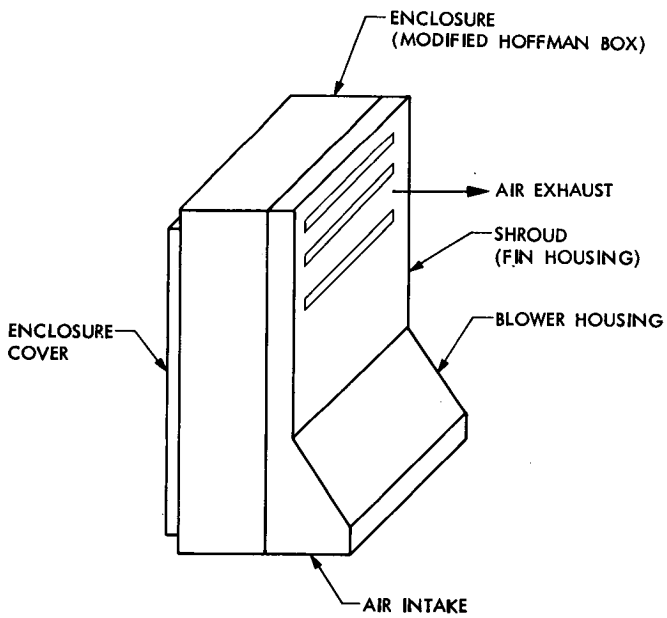


Fig. 2. Power supply enclosure

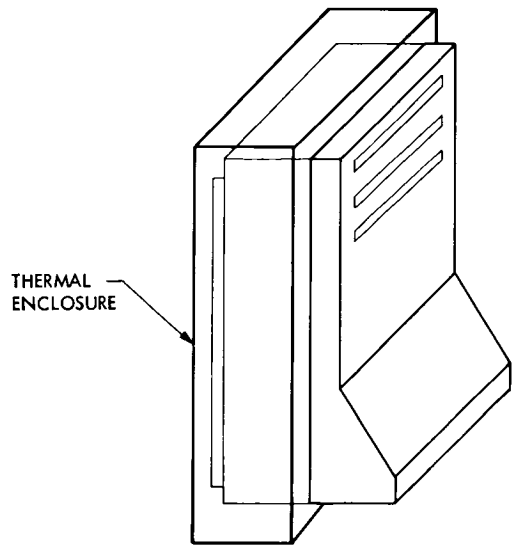


Fig. 3. Downconverter enclosure (power supply enclosure modified by adding thermal enclosure)

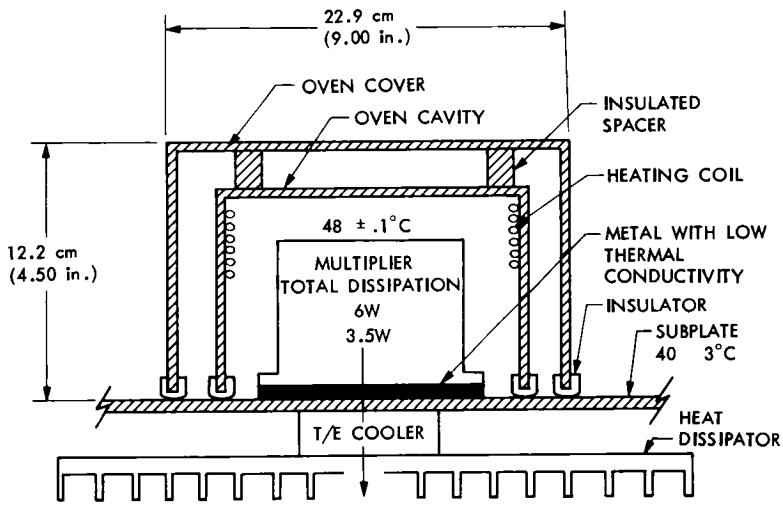


Fig. 4. Multiplier oven

Networks Consolidation Program

M. L. Yeater and D. T. Herrman
Office of Telecommunications and Data Acquisition

E. B. Luers
Bendix Field Engineering Corporation

The Networks Consolidation Program (NCP) was approved by the NASA Administrator in the fall of 1979 to combine the resources of the two NASA ground spacecraft tracking networks (the DSN, operated by JPL, and the GSTDN, operated by GSFC) into one consolidated network. This consolidated network will provide support to both deep space and high earth-orbiting spacecraft which cannot be supported by the TDRSS.

Previous TDA Progress Reports traced the history of activities and events that led to the decision to consolidate the NASA ground tracking and data networks into a single network. This report shows progress of the NCP from the previous report to the present, with special emphasis on planning and budgeting activities which have occurred.

I. Introduction

Initial planning and guidelines for the Network Consolidation Program (NCP) were presented in Ref. 1. Prior reports of progress have appeared in volumes 42-59 and 42-65 of the Telecommunications and Data Acquisition Progress Report (Refs. 2 and 3). They included a history of development of the NCP from initial meetings of the Networks Planning Working Group through activities completed in August 1981. This report addresses management, design, and implementation activities that have occurred from the time of the previous article through April 1982.

II. Program Accomplishments

A. NCP Organization

The Planning Team continues its supporting role to the project, which has had a succession of program guideline, mission support requirement, and budget changes. Also, the System Design Team, which functions as a subteam to the Implementation Planning Team, continues to provide engineering support for necessary design changes. Steering Committee meetings have been discontinued. The group provides support to the NCP Manager on an as-needed basis.

B. Formal Reviews

A fifth formal review was held in January 1982. Formal reviews were established to report on planning, design, and implementation activities of the project. Two reviews were reported in Ref. 2 and two in Ref. 3.

The fifth review described changes to program plans that occurred after January 1981, and presented a revised implementation plan and status report of objectives, budgets, and schedules. The review board recommended that the program proceed according to the implementation plan presented at

the review. Board members noted areas where changes in requirements and budgets may require further revisions to the implementation plan and recommended that another review be scheduled approximately one year hence to assure JPL and NASA management that the program is proceeding in an orderly and appropriate manner.

C. Budgets

Since the last report, changes in program budget levels have occurred as a result of changes in NASA management plans and congressional and executive budget decisions. Some of these changes have occurred because of changes in mission schedules, while others are due to reallocation of funds. The overall effect on the NCP has been to defer some planned capabilities until the post-NCP era and to reduce levels of capability in other areas. However, the program and basic network configuration remain intact. Overall, total NCP costs remain within the boundaries estimated in July 1980, but the expenditure profile has changed.

D. Antenna Array Alternatives

The previous report (Ref. 3) noted that the two existing 26-meter antennas at each consolidated antenna complex would be relocated and enlarged to 34 meters and include an arraying capability with the existing 64-meter and 34-meter antennas.

The comparative antenna analysis, noted in Ref. 3, provided a cost analysis which showed that construction of six new 34-meter antennas was more cost-effective than relocating, enlarging, and adding X-band capability to the six existing 26-meter antennas. Subsequent budget guidelines from NASA reduced the quantity of new 34-meter antennas to two, one each at Goldstone and Canberra. The final configuration is:

Complex	Antenna no. and size	Frequency band	Transmit/Receive or listen-only
Goldstone and Australia	1 64-m	S/X	T/R
	1 34-m	S/X	T/R
	1 34-m	S/X	LO
	1 9-m	S	LO
Spain	1 64-m	S/X	T/R
	1 34-m	S/X	T/R
	1 9-m	S	LO

E. Mission Set

References 2 and 3 noted that initial planning for the NCP included a requirement for support of a set of deep space missions, plus three high earth orbiter missions: ISEE-3, AMPTE and OPEN. After the third formal review, the NCP Parallel Study Team was formed to study effects on the NCP Baseline Design of including requirements in the NCP guidelines to support a number of additional missions. The additional missions were separated into three sets:

Set 1: Space Shuttle (STS), Tracking Data and Relay Satellite System (TDRSS), and Geosynchronous Satellite Transfer Orbits (e.g., GOES).

Set 2: ISEE-1 and -2, Dynamic Explorer-A, and San Marco D/M.

Set 3: Nimbus 7, GOES-D, E, and F, and Solar Maximum Mission.

The study team identified four options to provide support to some or all sets:

Option 1: Set 1 only, starting May 1984.

Option 2: Set 1 only, starting August 1985.

Option 3: All three sets, starting August 1985.

Option 4: Set 1 only, starting 1987.

The final report of the Parallel Study Team recommended that Option 4 be selected. This recommendation was forwarded to NASA OSTDS, but was not approved. A study of a fifth option, which included support for TDRSS, GOES, and STS was requested by OSTDS.

As a result, the mission set to be supported by the consolidated network has been revised to include GOES and emergency backup to the TDRSS. A decision about emergency support for STS has been delayed pending the results of tests that are included on STS-3. This represents a modest increase in support requirements levied on the network. A copy of the latest approved NCP Mission Set is included as Fig. 1 of this report. However, several changes have been made in overall mission planning within NASA which will be reflected in the next issue of this mission set.

At this time it is known that Solar Polar-S (line 9 of Fig. 1) has been removed from the mission set as have all of the Deep Space Missions Under Study (lines 13 through 20). In addition, the AMPTE IRM (line 24) now reflects the addition of the UKS (United Kingdom Satellite) on the same time line and should read AMPTE IRM/UKS.

One other change reflects the delay of the launch dates of the OPEN missions as follows:

OPEN IPL to February 1989

OPEN GTL to February 1989

OPEN PPL to August 1989

OPEN EML to February 1990

F. Project Dependencies

Since the last report several changes have been made in program guidelines due to changes in budget directives from NASA Headquarters, support requirements for ongoing missions, and support requirements for new missions. As the system design process progressed, some design changes have resulted in reductions in the amount of equipment required from GSTDN stations scheduled for closure.

There still is the question of whether the program will use 9-meter antennas or 26-meter antennas to track the residual high earth-orbiting spacecraft. This question will be resolved within the next six months.

Mission schedules, including launches of the Space Shuttle and the TDRSS, have settled to the point where the program can depend on them, and develop implementation and mission support schedules in an orderly manner with a high degree of confidence.

The NCP Baseline Mission Set (Fig. 1) identifies several mission critical events, including the Voyager-Uranus encounter, and the launches of several spacecraft such as AMPTE, Galileo and GOES which continue to represent major constraints on the development of the consolidated network.

Modifications in funding profiles have continued to require changes in the sequence and timing of tasks. These changes will be reflected in the level of mission support that the DSN can provide during some implementation periods, and the ground aperture available to support Voyager Uranus will be less than originally planned. In addition, some network improvements, particularly in the area of monitor and control, will be delayed until after the NCP/MARK IVA implementation.

G. Documentation

The NCP Management Plan, Transition Plan, and Mission Support Plan were published and distributed in accordance with initial program plans. Publication of the one remaining planned document, the NCP Implementation Plan, was delayed due to the large number of changes in program plans that were required in response to budget and mission support directives. The document was published in April 1982.

References

1. Layland, J. W., OSTDS Networks Planning for the TDRSS Era, OSTDS Report, January 1980.

2. Yeater, M. L., Herrman, D. T., and Sanner, G. E., "Networks Consolidation Program," *TDA Progress Report 42-59, July and August 1980*, p. 107, Jet Propulsion Laboratory, Pasadena, Calif., Oct. 15, 1980.

3. Yeater, M. L., and Herrman, D. T., "Networks Consolidation Program," *TDA Progress Report 42-65, July and August 1981*, p. 19, Jet Propulsion Laboratory, Pasadena, Calif., Oct. 15, 1981.

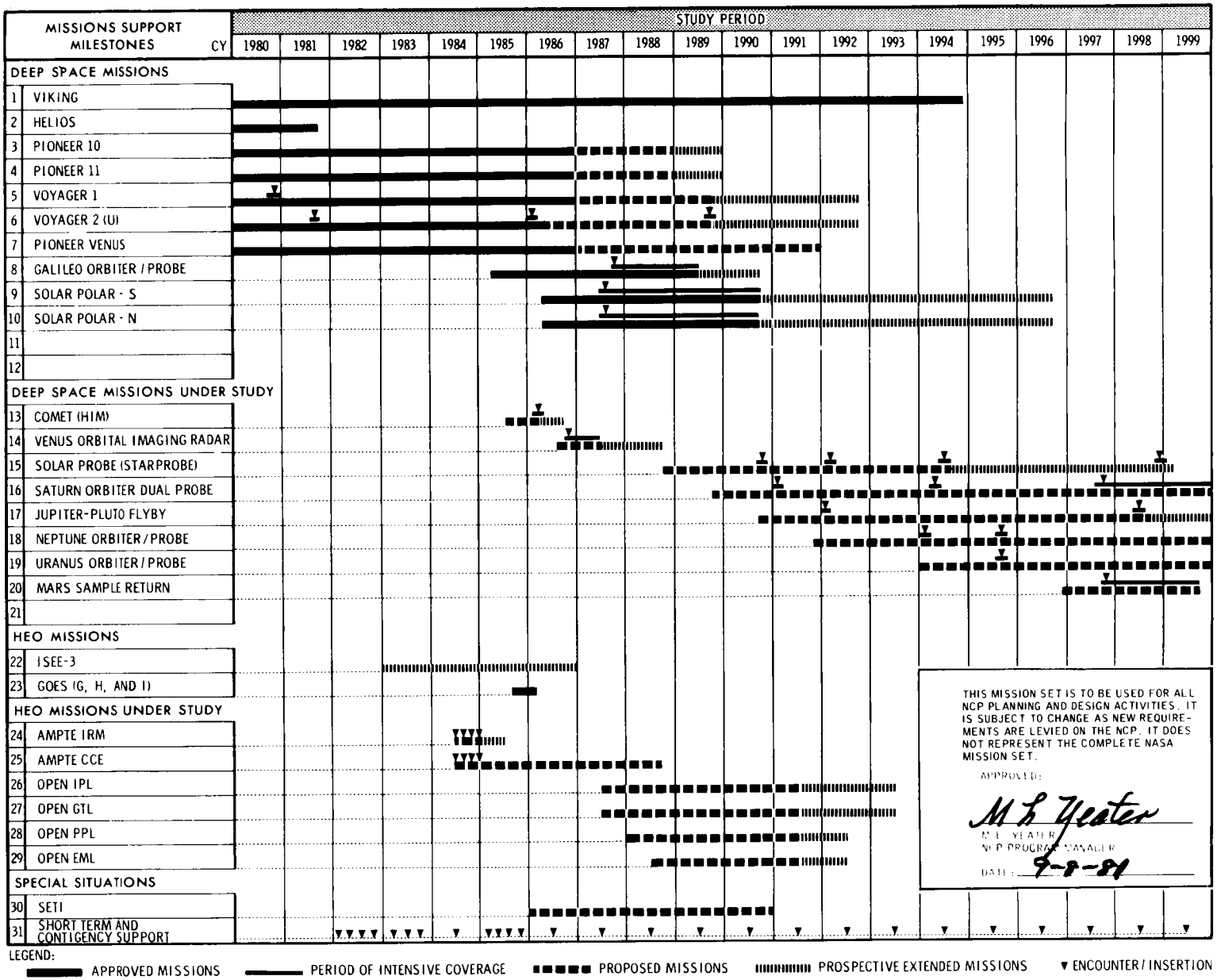


Fig. 1. NCP flight mission set

Management and Development Local Area Network Upgrade Prototype

T. J. Fouser
DSN Data Systems Section

Given the situation of having management and development users accessing a central computing facility and given the fact that these same users have the need for local computation and storage, the utilization of a commercially available networking system such as CP/NET from Digital Research provides the building blocks for communicating intelligent microsystems to file and print services. The major problems to be overcome in the implementation of such a network are the dearth of intelligent communication front-ends for the microcomputers and the lack of a rich set of management and software development tools.

I. Introduction

The purpose of this paper is to report on the progress of a research effort to study local area networks and the application of networking to administrative and developmental needs. A local area network of communicating, intelligent workstations provides the users with enough localized computing power to perform tasks such as word processing, program development, and other applicable implementation engineering work. The network with a particular node designated as the server, equipped with extra disk storage and one or more printers of different capabilities, provides the users with print server and file server functions. The research effort is directed toward investigation of commercially available microcomputer operating systems, CP/NET, CP/M, and MP/M.[1] CP/M has become the industry-wide standard microcomputer operating system. MP/M is the multi-user, multiprogramming version of CP/M, and CP/NET is the networking interface between a user and his CP/M system and the server node operating with MP/M. CP/M, CP/NET, and MP/M are products and trademarks of Digital Research.

[1] CP/NET, CP/M, and MP/M are trademarks of Digital Research Inc., Pacific Grove, Calif.

II. Background

The current configuration of the Management and Development Network (MADNET) is illustrated by the top portion of Fig. 1. The system centers around dual MODCOMP 7870 computers, one of which supports the HAL/S software data base, three printers, and up to 26 users of the HAL/S software development effort. The other supports the work breakdown structure (WBS) data base, two printers and 28 management users. Each of the users has a MADNET standard, DEC VT-100 terminal and runs the WBS software on the 7870, generating reports from the data base that are output to the printers. The VT-100 terminals were chosen since they could display 132 columns of output when generating reports and

are considered to be a constant in the upgrade process. With the existence of many VT-100 terminals which can be upgraded to CP/M personal computers with a commercially available product and with the existence of many AODC computers, networking these systems together was primarily concerned with bringing up a server node with the desired peripherals and fine tuning the network software for optimum performance.

III. Implementation Effort

The lower portion of Fig. 1 indicates the added CP/NET system. The dumb VT-100 terminals on the MODCOMP are upgraded with the addition of a Z80A cpu, 64-kbytes RAM, floppy disk controller, parallel printer port and four serial ports. These components are on two cards that are housed inside the VT-100 cabinet and connected to one or more floppy disk drives in a separate enclosure. The upgraded terminal runs CP/M at 4 MHz and allows for dumb terminal emulation when needed for operating with the MODCOMP. The upgraded terminal can have an optional printer that serves as the CP/M list device. Communication to the network is provided through the additional serial I/O ports.

The central CP/NET node, the server, in the current prototype studies, consists of a Z80-based microcomputer running MP/M along with the CP/NET server software. The hardware is a Z80A running at 4 MHz with 64-kbytes RAM, seven serial I/O ports, two single-sided single density floppy disk drives, and one Morrow 26MB hard disk. The seven serial ports support two local consoles for MP/M operations, one serial printer port, and four serial ports for network communication. The communication with the nodes can occur at various baud rates. Currently, direct connections run at 9600 baud and connections through modems run at 1200 baud.

The software on the CP/NET server, as shown in Fig. 2, consists of the standard MP/M system software and utilities with additional CP/NET utilities and a customized extended I/O system (XIOS). The MP/M multiprogramming monitor control program provides a microcomputer environment with multiple consoles, each with multiprogramming capabilities. The standard MP/M features include spooling of print files to the printer, scheduling of programs to be run by date and time, and setting and viewing the current date and time. Each user at a console has the ability to start a process, detach the console from that process, and initiate another process. The ability of MP/M to support detached processes allows processes that are always in memory acting as a resource. The CP/NET software executes in this manner, providing support for node operations such as printer and disk accesses. The CP/NET server software adds capability to receive messages from the nodes and send messages to the nodes, with the ability to broadcast to all nodes. The XIOS provides the custom interface between the standard MP/M and CP/NET software and the hardware the system is running on.

The software on the CP/NET node consists of the standard CP/M and the network disk operating system (NDOS), which intercepts operating system calls that need to be redirected onto the network. The customized slave network I/O system (SNIOS) interfaces the standard software to the particular hardware configuration and provides the communication over the network.

The mapping of the CP/NET configuration to the International Organization for Standardization's Reference Model of Open Systems Interconnection (ISO OSI) is shown in Fig. 3. The physical layer is the hardware that transfers the bits from one node to another without regard for whether or not the collections of bits constitute a valid packet. The data layer picks the checksum out of the transmitted bitstream and judges the integrity of the packet. These ISO Reference Model layers are implemented in the hardware and the customized XIOS in the server and SNIOS in the nodes. The network, transport, session, and presentation layers of the Model are implemented in the standard CP/NET software, the NDOS in the nodes and the node support processes in the server.

IV. Operation of Current Prototype

The operation of the CP/NET local area network provides the user with print and file server functions. A typical session might be retrieving a file from the file server, performing the desired operations on the file such as editing or other computing, perhaps printing the file or processing it with printed report output, then restoring the file to the file server. Once the print file is generated, it is then sent to the print server for spooling and printing.

The applications of these functions include electronic mail, common software, off-loading of the editing function from the MODCOMP, and WBS off-loading and viewing. The utilization of data base management systems on the server provides for operations such as SRM, ECM, WAD and action item accessing.

The use of the network as a print server allows operations to be performed with as little as possible impact on the user. The typical node has a slow and possibly noisy local printer or, more likely, has no printer at all. The local printer is used for print jobs that are short or do not require a printer with special features not found on the local printer. To use the print server, the local user first equates the CP/M print device with the desired printer on the server, then uses the print

device as usual. A program such as Wordstar[2], generating printer output, would have that output intercepted by the NDOS and rerouted to the network and on to the server where it would be automatically spooled for printing. When the printing is complete, the end-of-printfile message is sent to the server, triggering the despooling of the print file. The local user also uses the CP/M file transfer utility to send previously generated print files to the print server, where they are automatically spooled for printing.

The additional capabilities of the CP/NET must be stated in terms of utility to the users. For use as a file server, the network must be able to store the user's files such that they can be retrieved without undue delays. Speeds of the current configuration do not approach those of using the local disk, discouraging the use of the file server for temporary storage. The ease of this operation to the user also depends on the configuration of the user's hardware. If the node has a local printer that is letter quality and/or slow, sending large print files to the server is faster than printing the file at an effective baud rate of 300 to 1200 bps. If the local printer is fast and can produce the required quality, use of the print server would be of questionable value. The cost of equipping each node with such a printer may be prohibitive.

To determine the actual performance of the network file services, a "typical" file size was selected. The file is slightly larger than 6 kbytes, which represents approximately four pages of typewritten material. Since a typical operation would include transferring from server to node, then node to server, both times were investigated. The nature of the environments of the server and the node made a significant difference in the times recorded. The critical times comprising the transfer consist of the time to get the data from the disk and transmit the data block (typically a 128-byte CP/M sector), and the time to store the block on the destination disk.

In the transmission from the server to the node, the transfer utility in the node, operating at the application layer in the ISO Model, starts the transfer by creating a new destination file on the local disk. The source file is requested to be opened, but this command is intercepted by the NDOS and redirected to the server over the network. The node then issues a series of "read next sector" commands which are also intercepted and passed along to the server. With each command, the server goes to the disk, reads the sector and transmits the CP/M logical sector to the node. While the server disk sector is being read, the node processor, having nothing else to do, can take the data at almost a full 9600 baud. Once the 128 bytes of data has been received, the file transfer utility stores it in memory, from which it is written to the disk upon a full memory condition or end of file.

In the transmission from the node to the server, the operation is the reverse. The file transfer utility reads the file (or as much of the file as will fit) into memory. The utility then issues a series of commands that contain the 128-byte sectors and request a "write next sector" operation. These commands are intercepted by the NDOS and sent to the server over the network. Since the server must support activities of other nodes and from the local consoles, it cannot receive the data at a full 9600 baud. The reason for this is that characters arriving at a serial I/O port at 9600 baud arrive at 100-microsecond intervals. With each character, an interrupt is generated to the server processor. The processor must identify the interrupting hardware, input the character, store it in the appropriate buffer, perform other housekeeping, then return to the interrupted process. This procedure consumes approximately 50 microseconds. It is easy to see that two nodes inputting data to the server could easily overrun the processor. To maintain reliable data transmission without losing data or having to retransmit many packets, a delay of approximately 8 milliseconds is inserted between each character that is transmitted. As a result, the best case transmission time for the "standard" 6-kbyte file from the server to the node is approximately 40 seconds and the node to server time is 85 seconds. The best case is when the server has no other node or local demands on it. With another node placing a similar demand on the server and with a local console actively listing a file, the server-to-node time goes from 40 up to 60 seconds and the node-to-server time goes from 85 to 95 seconds. These times indicate some elasticity of the server in being able to accept the additional loading. With the 8-millisecond-per-character delay decreased to 6 milliseconds, the best-case server-to-node time decreases from 40 to 35 seconds and the best-case node-to-server time decreases from 85 to 65 seconds. But when the load is added to the server, the communication breaks down due to lost messages. In an environment where there are up to 16 nodes making demands on the server, larger delay times for transmission to the server will be necessary.

V. Conclusions Based on Current Status

The current implementation state of the prototype local area network has shown that the use of CP/NET software provides an acceptable networking environment in all respects save one: speed of transmission time. The CP/NET software provides a standard, commercially available implementation of the ISO Model layers 3 through 6, providing full networking support for the application programs running on ISO Model layer seven. The customized implementations of layers 1 and 2 are solely dependent upon the available hardware for the physical layer and the implementations of the XIOS and SNIOS for

[2]Wordstar is a trademark of MicroPro International Corp., San Rafael, Calif.

the data link and network layers. The fact that the processor, particularly on the server but also on the nodes, spends so much time doing the byte-at-a-time I/O in these layers not only places a physical limitation on the network transfer speeds but also deprives application processes of valuable compute time.

To make a CP/NET implementation of a local area network a viable alternative, the low level communications tasks must be removed from the processor through the use of an intelligent or semi-intelligent communications front end.

VI. Future Directions

Communication with the nodes, which currently runs at an effective rate of less than 9600 baud, would be greatly enhanced by the addition of an intelligent data communications front end. The key problem is the fact that I/O to the network is performed through serial ports in a byte-at-a-time fashion. By performing the network I/O through a DMA operation to a board connected to Ethernet or some other high-speed medium such as broadband or fiberoptics, the communicating processes are relieved of much of the most time-consuming work. The processor, operating in the XIOS or SNIOS, no longer does the character I/O but instead simply indicates to the communications front end the memory location of the message. The communications front end takes the message through a DMA operation, sends it, receives the reply and places it in memory, only then notifying the processor that the reply has arrived. The processor during this time is free to perform other tasks. Operation of nodes with no local disks, previously unworkable due to the speed of the network, would now be more plausible.

The new version of MP/M, MP/M II provides additional features and capacities for the server node of the network. This version of the software supports up to 16 disk drives of up to 512 megabytes each, up to 16 printers, and up to 16-character I/O devices. Of the 16-character I/O devices, up to eight can be local consoles to MP/M and up to 15 can be CP/NET nodes. This version of MP/M provides file passwords, the ability to lock files and/or records within files for multiple user protection, automatic archive tagging for backing up files, and time/date stamping of files.

The development of software tools and managerial tools resident on the system will enhance the productivity of the users. The use of MP/M II, a new version of MP/M, allows for up to 400 kbytes of RAM, some of which could be used for a memory-resident data base manager and memory-resident virtual disk storage area. This arrangement would allow for the implementation of powerful data base management techniques.

New versions of CP/NET software will consist of ROM-able system modules that allow the operation of network nodes that have no disk drives at all. This software, coupled with the higher-capacity communication links, will allow bringing needed managerial and developmental tools to users that have only a terminal to upgrade and have no need for local disk storage.

Preliminary tests have indicated the plausibility of file transfer software that would interface the MODCOMP 7870 to the CP/NET server. This facility would provide needed offloading of file editing, archiving and software development version control. Interfacing of CP/NET to the VAX 11/780 could be accomplished with the possible procurement of commercially available software packages that meet the corresponding ISO Model layers of the CP/NET software. Hardware and software drivers already exist that meet ISO Model layers 1 and 2 of the Ethernet specification for the VAX 11/780, MODCOMP, multibus, and S-100 bus.

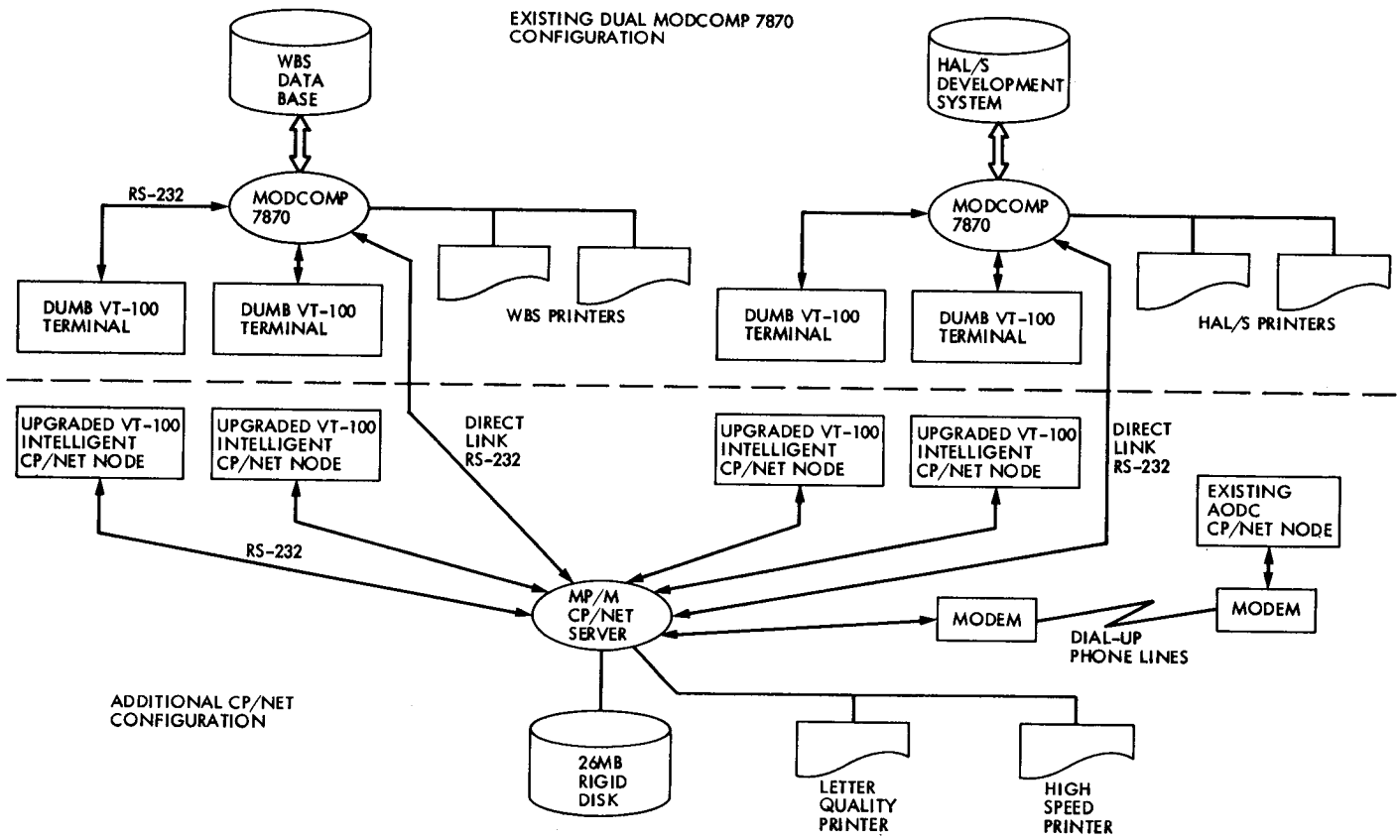


Fig. 1. MADNET initial upgrade configuration

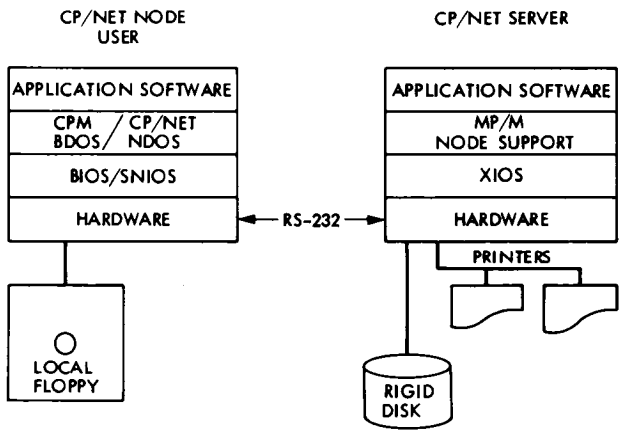


Fig. 2. Software/hardware layering

7 - APPLICATION LAYER	- MANAGEMENT AND DEVELOPMENT TOOLS	DBMS SYSTEMS	
6 - PRESENTATION LAYER	- CP/NET NDOS	NODE SUPPORT	STANDARD CP/NET SOFTWARE
5 - SESSION LAYER	- CP/NET NDOS	NODE SUPPORT	
4 - TRANSPORT LAYER	- CP/NET NDOS	NODE SUPPORT	
3 - NETWORK LAYER	- CP/NET NDOS	NODE SUPPORT	
2 - DATA LINK LAYER	- SNIOS	XIOS	CUSTOMIZED NETWORK SOFTWARE AND HARDWARE
1 - PHYSICAL LAYER	- HARDWARE	HARDWARE	
ISO MODEL LAYERING	CP/NET NODE	CP/NET SERVER	

Fig. 3. ISO Model mapping to CP/NET

Algorithms for Software Development Version Control and Change Detection

R. C. Tausworthe
DSN Data Systems Section

This article describes simple computer algorithms for processing source program and text data files in order to extract change detection, version control, version history, and current status information easily. These algorithms presuppose that it is possible to attach to each record of the source files a 6-character code, placed within delimiters that will cause the compiler, or other using program, to ignore this code field. The code contains a 2-character code for a character-by-character position-sensitive checksum of the record, another for the record number in the file, and a third for the date on which the encoding took place. Once the source file has been thus encoded, it is possible to detect the following transactions on the file since the most recent version coding: (a) addition of new records (having no version code); (b) modification of existing records; (c) deletion of a number of records; (d) movement and/or duplication of existing records; and (e) modification and duplication of records. In addition, it is possible to extract a version history of the number of records created or modified by date. A special file listing program is described which prints the file records without showing the version codes, but places a "change bar" at the right margin whenever a change is detected. The program also provides a list of changed pages and a version history.

I. Introduction

During the development and maintenance phases of software tasks, it is extremely convenient to have some way of distinguishing in what way a particular version of the product differs from its ancestor versions. In this way, those persons making and monitoring the changes in the product can more readily focus on the differences made since a particular proven benchmark. When one is given a new release of a large document or program for review, for example, one must be able to locate what is different about *this* version, as opposed to the last one, in order to avoid the needless expense of rereading the entire document.

Managers concerned with implementation status may be interested in knowing the rate at which the product is being developed, as well as other statistics of the process, such as the amounts of new code, modified code, and reused, unmodified code contributing to a particular effort.

It is often important to control the issuance of new versions, so that only a known, prescribed, and contracted capability is released to customers and utilized by them.

Such change identification, version history, status, and version control processes are virtually infeasible for large com-

puter programming projects without computerized aid. Some techniques that have been employed in the past include compile-time dating of the listing and source file, usage of file comparison programs, and special annotations of the source made by a text editor while producing the program text. All of these techniques are useful, but have their limitations, as does the technique described in this article. Compile-time dating, for instance, suffers by only telling when the program was last compiled, not what is different about the program. File comparison programs are generally slow and can only identify relatively small changes between the two versions of the source. Annotations placed in the source by a text editor, usually limited to sequence numbers of the records, or perhaps a version number or date of last update of the entire file, do not provide for identification of modifications made on individual records, nor do they distinguish which lines have appeared in which versions of the software.

This article describes a combination of techniques yielding simple computer algorithms for change identification and version history that are efficient in file storage requirements, execution speed, programmer overhead, and operations. These algorithms can be applied whenever it is possible to attach to each record of the source files a 6-character code, placed within delimiters that will cause the compiler, or other using program, to ignore the code field. The code contains a 2-character code for a character-by-character position-sensitive checksum of the record, another for the record number in the file, and a third for the date on which the encoding took place. Once the source file has been thus encoded, it is possible to detect the following transactions on the file since the most recent version coding: (a) addition of new records (having no version code); (b) modification of existing records; (c) deletion of a number of records; (d) movement and/or duplication of existing records; and (e) modification and duplication of records. In addition, it is possible to extract a version history of the number of records created or modified by date. A special file listing program is described which prints the file records without showing the version codes, but places a "change bar" at the right margin wherever a change is detected. The program also provides a list of changed pages and version history.

The scenario of operations is as follows (see Fig. 1): The principal activities associated with software development and sustaining involve the programmer interacting with a text editor building and modifying program and documentation source files. At regular intervals, these source files are submitted to a version encoding program that checks each line for correct checksum and then resequences the file. When the checksum is correct (i.e., matches the checksum in the code supplied in an earlier run), the date is not changed. If the checksum is in error, the current date is inserted in the date code. Thus, after each run of the encoding program, the source file has a correct checksum code, correct sequence number code, and applicable date for each line in the file.

Subsequent editing of the file may create new lines, delete old ones, modify characters, insert characters, move lines, or copy (reuse) existing lines. The combination of the three elements of the line code permits all of these transactions to be detected, including identification of changes made since a given status date.

Each time the source file is checksummed, information as to which lines have been added, which modified, how many were deleted, etc., is lost. However, version history information is retained, as it is possible to tell which lines were redated at each checksumming.

II. Source File Structure

The source file structure after the version coding operation is described in the following schematic logic structure:

```
Source file
  *Physical line
      *Data segment
       Version code
         Left delimiter
         Checksum code
            2 characters
         Sequence number code
            2 characters
         Date code
            2 characters
         o Right delimiter
```

The annotations (* and o) in this logic refer to iteration and alternative structures, respectively. Indentation corresponds to refinement of the item definition. This structure thus defines the source file as an iteration of physical lines, each made up of a number (perhaps zero) of data segments, followed by a version code. The version code, in turn, consists of a left delimiter, checksum, sequence number, date, and perhaps a right delimiter (depending on the compiler or program using the file).

The delimiters will generally need to be file-dependent. For program source files, these delimiters will be chosen to make the code appear as a comment to the compiler, perhaps with a character or two thrown in for easy recognition that this comment is the version code, not an ordinary comment.

Other files may require special handling procedures, such as special input subroutines to fetch input records and strip out

the version codes, or a preprocessing pass of the source file to create a dummy file without the codes for access by the program.

III. Generating the Two-Character Codes

The two characters used in the three code elements could be any printable members of the ASCII set, i.e., ranging from space (character 32) through the tilde (character 126). Control characters (0 through 31) are apt to cause sensitivities within the using program(s), especially compilers, and are thus omitted here. Also, characters following "z" in the ASCII set may be set aside for use only in the delimiters, leaving 91 characters for the code elements.

Care must be taken in generating the version code not to generate character sequences sensitive to the using program. For example, if a Pascal program were being version encoded using comment delimiters "(*"and"*)" around the code, then care would have to be taken to ensure that the characters "*)" do not appear within the version code. Pascal has alternate comment delimiters "{"and"}", ASCII characters 123 and 125, which may be used to obviate this problem. However, PL/I programs will have problems with "/*"and"*/". If the algorithms herein are to be applied to Pascal programs not permitting the alternate delimiters or to PL/I programs, then the version code should omit the offending combinations. For the remainder of this article, we shall assume that the full 91-character span can be used.

The combination of two such characters is sufficient to span a range of 91 X 91 = 8281 alternatives. The checksum of characters on the line, discussed later, will lie randomly within the interval (0, 8281). The probability that a randomly modified line will have the same checksum as was recorded before the change will be 1/8281 = 0.00012; this low figure is probably sufficient for most change identification purposes. The particular checksumming algorithm presented is one which makes almost all of the usual types of modifications to a line detectable within the probability stated above.

The limit of the sequence number to 8281 limits the file size to as many records; this probably does not pose much of a problem, as programs this large are usually segmented for compilation purposes, and documents this large are similarly segmented for ease in processing.

The correspondence between a number N and its representation by the two code characters C1 and C2 is defined by

$$C1 = \text{chr}(\text{ORD}(' ') + N/91)$$

$$C2 = \text{chr}(\text{ORD}(' ') + (N \text{ modulo } 91))$$

where chr(n) is the ASCII character having collating sequence number n and ORD(C) is the collating sequence number for the character C. The ASCII collating sequence number for space is 32.

Similarly, given a code with the two characters C1 and C2, the corresponding number N is

$$N = 91 * [\text{ORD}(C1) - \text{ORD}(' ')] + \text{ORD}(C2) - \text{ORD}(' ')$$

$$= 91 * \text{ORD}(C1) + \text{ORD}(C2) - 92 * \text{ORD}(' ')$$

A. The Date Code

The conversion from a date having day number d in month m of year 1980+y into the date code number D is defined by

$$D = (d-1) + 31 * (m-1) + 372 * y$$

This date number can be encoded into two characters as described above.

When the two-character date code has been retrieved from a line and converted to a date number D, the day of the month will be computed as 1+(D modulo 31), the month as 1+(D/31 modulo 12), and the year as 1980 + (D/372). In these computations, and indeed, throughout the remainder of this article, integer division is assumed (remainder discarded) whenever "/" appears in an expression. Although this date encoding appears to use up 372 days per year (i.e., 6 or 7 unused numbers each year), it is still sufficient to carry the version code through 8281/372=22.26 years, or until March 5, 2002.

B. The Checksum Code

The line characters to be checksummed, call them $C_1, \ldots, C_N$, are combined by the following algorithm:

Each character C_i is combined with the line checksum L accumulated so far to obtain a new value by bit-wise "exclusive-or" (xor) and "and" operations. First set $L = 0$; then for each of the C_i,

$$\text{set } L = L \text{ xor } 2*(L \text{ and } 16383),$$

and then

$$\text{set } L = L \text{ xor } (L/16384) \text{ xor } \text{ORD}(C_i))$$

Finally,

$$\text{set } L = (L \text{ modulo } 8281)$$

This computation simulates the driving of a maximum-period linear feedback 15-element finite-state machine with the char-

acter values from the input line. Values of the checksum lie in the range 0-8281. The iterations above, when given only a single non-null character and nulls thereafter, generate a pseudo-random sequence which repeats only after $2^{15}-1$ = 37,767 such characters have been input. The combination when characters are non-null is a convolution of the input sequence with the linear pseudo-random response of the finite-state machine. This method tends to randomize the checksum values in an extremely character-sensitive way, so that the chances of any modification to an input line falling in the same equivalence class of input lines having the same checksum is 1/8281 = 0.00012.

The recursion relation above corresponds to the maximum-period linear shift-register device having the primitive characteristic polynomial over GF(2)

$$f(x) = x^{15} + x + 1$$

The algorithm for utilization of this polynomial to generate pseudo-random numbers is given in [Ref. 1]. The operations are to shift the checksum left one bit, filling a 0 on the right, and xor with the unshifted value. Then xor this intermediate result with a right shift of itself by (15-1)=14, filling 0's in on the left. For further information regarding the properties of shift-register sequences, see Ref. 2.

C. Sequence Number Code

The sequence number code is a straight translation of the record sequence number, starting at zero, into the two-character code, and vice-versa.

IV. The Checksum Program

This section describes the design of a prototype program written to produce the updated version-encoded source file. For convenience, it will be referred to herein as "CHECKSUM".

Some basic assumptions about the files processed by the program and the processing done by the program are:

- Three types of version codes will be accommodated:
  - files requiring only a left version code delimiter
  - files requiring both left and right delimiter
  - files in which version codes are restricted to certain columns on the input line
- It will be possible to detect which of these version code types is to be used from the file name, say by the file name extension; alternately the program could accept this information from the user of the program.
- Source (program) files may have embedded comment lines containing design information expressed in a Program Design Language (PDL). Such lines are distinguished by the presence of an "extraction code" at the beginning of the line.
- PDL lines and non-PDL lines are to have separate, independent sequence numbering so that if the PDL is extracted, the numbering is still correct.
- PDL and non-PDL lines may have different delimiters, if accommodated by the using program.
- Placement of the version code on an output line may be designated to appear in a certain column on the line if desired.

The main procedure of the program is described in the program design language of Fig. 2; the structure of the version code parameter lookup table is presented in Fig. 3; the extraction algorithm for retrieving version code information from the input line is given in Fig. 4; and the method for checksumming and reformatting of the output line is shown in Figs. 5 and 6.

V. The Version Code Removal Program

The CHECKSUM program must detect whether an input line has a version code on it, and whether it is a PDL variety or not, so that the code can be removed and another reattached in its place. However, in some applications, it may be necessary to strip out the version codes from an entire file before it can be properly processed by its using program, or when a file is to be completely reencoded.

The algorithm given in Fig. 4 has been used in such a program, called STRIP, which removes the version codes from an entire file. The algorithm has been only slightly modified, shortened because the BEGINNING, SEQUENCE_NUMBER, and SEQUENCE_CODE variables are not needed.

VI. The File Listing Program

The file listing program, called PRINT, is similar to the CHECKSUM program in overall structure, except that a "change bar" is reattached to the output line, rather than the version code. Also, the PRINT program keeps track of what sequence numbers have appeared in the file, and therefore knows when deletions and duplications have occurred, as well as the number of lines which have undergone such transactions with the text editor since the most recent processing by CHECKSUM.

Besides the name of the file to be processed, a status date is given to the PRINT program; all lines of the input file having a date code after this status date will be identified on the listing. The change bar applied to the output is chosen to distinguish what kind of change has been detected:

	Change bar symbol	Detected change		
		Checksum	Sequence number	Date
No change	(None)	OK	OK	= < Status date
Change since status date	[ ]	OK	OK	> Status date
Added line	[A]	None	None	None
Sequence number went backwards	[B]	–	< Previous	–
Deleted line(s)	[D]	–	> Previous + 1	–
Modified line	[M]	Bad	OK	–
Reused line	[R]	OK	Duplicate	–
Reused and Modified	[RM]	Bad	Duplicate	–

Whenever [B] or [D] situations are detected, a blank line is printed with these change bars; then the input line checksum and date codes are checked as in the table above for possible application of either (none), [ ], or [M]. The PRINT program records the page number each time a change is detected. A list of changed pages is then output to assist readers of the documentation. Separate counts of the numbers of input lines falling into each of these categories (except [B] and [D]) are maintained as the input file is processed. The total number of deletions is determined after the entire file has been processed.

In addition, each time an input line is processed, the version history is updated, as follows: The date on the line (or current date if the line has none) is found (or inserted, if not found) in the VERSION_DATE field of the VERSION_HISTORY table, and the corresponding NUMBER_OF_TIMES entry in the table is incremented. After the entire file has been processed, the version history of dates vs. number of lines is printed.

The algorithm for detection of deletions and duplications of records makes use of a list of pairs of numbers corresponding to intervals of sequence numbers *not* yet found up to the current input line. Initially, the PRINT program starts with the list containing only the pair (0, INFINITY). At any time the list will contain a number of such pairs signifiying ranges of sequence numbers not yet seen. For example, if at one point the list contained the pairs (5, 8), (10, 10) and (12, INFINITY), then only the sequence numbers 0, 1, 2, 3, 4, 9, and 11 will have been seen so far.

Whenever this list is checked against the current input line sequence code (converted to a number), the sequence number may be found not to appear anywhere within the intervals of any of the pairs in the list; in this case, the sequence number is a duplication, it has already been seen. However, in the usual case, the sequence number has not been seen yet, having been determined to lie within the interval specified by a particular pair in the list. Then one of three things takes place:

(1) If the sequence number is the lower limit of the pair, this lower limit is incremented by 1

(2) If the sequence number is the upper limit of the pair, this upper limit is decremented by 1

(if in either of these two cases the lower limit afterwards exceeds the upper limit, the pair is deleted from the list)

(3) If the sequence number is between the limits (low, high) of the pair, the pair is split into two pairs:

(low, seq_no–1), (seq_no+1, high)

When all the records of the input file have been processed, the number of deletions may be computed as follows: the last list pair (last+1, INFINITY) is discarded; the remaining pairs identify the gaps in sequence numbers of the input file. These intervals may be printed out, if desired (PRINT does not, however). The total number of deletions is the (high-low)+1 value of each list pair summed over all remaining list pairs.

Since the input file may be a source program containing embedded PDL statements sequence numbered separately, it is necessary for the PRINT program to maintain separate SEQUENCE_NUMBER counters and occurrence lists for the two types of records.

VII. Application Problems

Several concerns may occur to those considering the application of version codes to their source files. These concerns include

(1) Expansion of file size due to version code overhead.

(2) Sensitivity of the using program(s) to the version code.

(3) Effort demands and discipline required of user to update the version codes periodically.

(4) Effort and nuisance of having to remove the codes before submitting the file to a sensitive using program.

(5) Nuisance in seeing meaningless version codes on compiler output listings, even when compiler is insensitive to the codes.

(6) Impracticality of applying the version codes to some kinds of packed text files, such as commonly created by some word processors, that are free-form and not line-oriented.

The applicability of the technique described in this article is therefore not universal, as it stands. Some of the difficulties may be slaked by the advantages of a version control and change identification capability, such as a slight file size overhead, the discipline of update and archival, etc. Others, such as the sensitivity of the using program(s) to the appearance of the version code, present real problems.

The technique, as it stands, *is* very useful in the following cases:

(1) For input text files of any kind whenever the using program(s) may be adapted, or written from the outset, to ignore the version codes. Programmers can be supplied with standard library subroutines to handle input of fundamental data types.

(2) For program source code when the programming language permits a comment to end the line, signalled by a comment left delimiter only.

(3) For program source code when the programming language permits a comment to end the line, using both left and right comment delimiters. This usually requires that the programmer not put comments in the program that extend across physical line boundaries.

The algorithms herein may be modified to accommodate the last of these cases when comments do extend over physical line boundaries. Such modification would demand only that the detection and reattachment procedures for the version code be extended to sense whenever the version code was being inserted within a comment or not, and to choose alternate delimiters as appropriate.

Should the method described above prove altogether infeasible because of the limitations listed, it may prove worthwhile to investigate the reasonability of modifying the operating system to reserve positions within each file record for version code usage, with special access functions to retrieve and deposit these codes. Existing programs using a file would then be unaware of the version code altogether, as it would never be delivered by the operating system when file data is requested. There are some obvious problems that one immediately encounters in such a proposed solution, such as how the operating system accommodates an existing, unmodified text editor in maintaining the correspondence of input file codes with rewritten output file codes. Such an investigation seems premature at this point, however.

VIII. Prototype Demonstration Results

To demonstrate the methods of this article, the program text of Fig. 2 was chosen for example. It was checksummed, modified, and rechecksummed repeatedly to simulate the growth and correction of a software design. Then it was modified without rechecksumming, so that additions, deletions, modifications, and duplications would be made more evident. Then it was printed out using PRINT, so that changes and version history would be shown.

Figure 2 shows the original text; Fig. 7 shows the encoded text, using left and right delimiters "{" "and"}", respectively; and Figs. 8 and 9 show the output of the PRINT program, which illustrates the appearance of change bars, the change statistics, and the version history printout. The list of changed pages is not shown, since the example is but one page long.

IX. Conclusion

The technique given here is a useful and feasible method for providing source file change detection and version control measures for many types of program, documentation, and configuration data files. As the application of such techniques becomes more widespread, one will find that the sensitivities to version encoding will certainly decrease. Operating systems will accommodate the codes automatically; compilers and other system software will expect such codes to be present; programmers will come to expect the system to provide version statistics services without incursion on their productivity; and managers and maintenance personnel will be able to perform their functions more effectively because of the better visibility into exactly what the extent and content of changes to a software package have been in any given release.

References

1. Tausworthe, R. C., "Random Numbers Generated by Linear Recurrence Modulo 2," *Math. Comp.*, Vol. XIX, No. 90, pp. 201-208, 1965.

2. Golomb, S. W., et al., *Shift-Register Sequences*, Holden-Day, Inc., San Francisco, Calif., 1967.

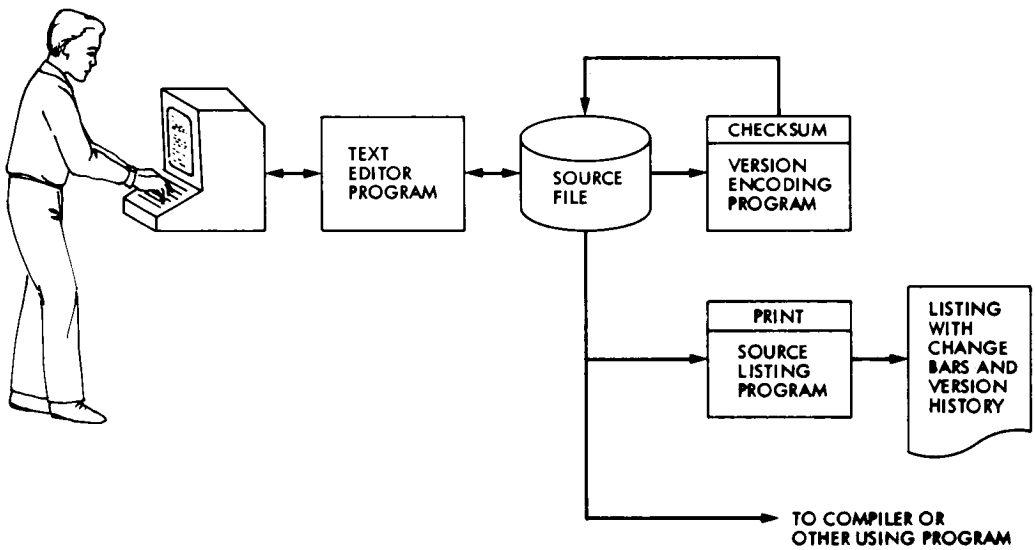


Fig. 1. Software development version control and change identification scenario

```
Program: CHECKSUM
/*********************************************************************/
/*                                                                   */
/* This program inserts a version code on each physical line         */
/* of a source file for use in applications where version            */
/* statistics of such files are to be extracted and analyzed.        */
/*                                                                   */
/*********************************************************************/

.1      define global_constants
.2      declare global_variables
.3      initialize the program
.4      announce the program title and version
.5      get_todays_date /* encode it into date_code */
.6      loop /* for at least one file, and maybe more */
.7        !  get_file_names(IN_FILE, OUT_FILE)
          !    /* OUT_FILE is a dummy file name only */
.8        !  open both files
.9        !  set_up_version_delimiters /* using extension from */
          !    /*  IN_FILE name to identify the particular     */
          !    /*  delimiters.                                 */
.10       !  reset_parameters /* to initial values, in case    */
          !    /* more than one file is processed.            */
.11       !  loop while not end_of_file on input
.12       !    !  input a line /* check for ASCII characters   */
          !    !    /* only and set WARNING switch if not     */
          !    !    /* already set and file is not ASCII      */
.13       !    !  extract_version_code from input line
.14       !    !  compute_checksum of input line
.15       !    !  reattach_new_version_code
          !    !    /* add tabs to code_column, if indicated
          !    !    /* also, set WARNING if line would be too long
.16       !    !  output the line
          !    !..repeat
.17       !  close files
.18       !  if WARNING not set
.19       !  :    rename the OUT_FILE as IN_FILE
          !  :->(else)
.20       !  :    inform the user that the checksummed file has
          !  :      the dummy file name
          !  :..end if
.21       !..repeat if user desires another run
        end program
```

Fig. 2. The main program design of the CHECKSUM program

```
Data GLOBAL_CONSTANTS

     ...

Let FILE_INFO be a table containing delimiter information vs file type
  Each "record" of the table has the following format:
   .  FILE_EXTENSION, the search key which determines the
          following:
   .  KIND, integer, a composite made up of SourceMode and PDLmode:
       SourceMode = KIND modulo 3
                  = 0 for left delimiter required only in source file
                  = 1 for left and right delimiters both required
                  = 2 for version code located by column only
       PDLmode    = (KIND/3) - 1
                  = -1 when no CRISP extraction is to take place
                  = 0 for left delimiter required only in source file
                  = 1 for left and right delimiters both required
                  = 2 for version code located by column only
   .  EXTRACT_SIGNAL, string prefixing lines that are embedded PDL
   .  CODE_COLUMN, integer, for placement of code on output line, when >0
   .  LEFT_DELIMITER[0]  for source lines that are not PDL
   .  RIGHT_DELIMITER[0]  "    "      "     "   "   "   "
   .  LEFT_DELIMITER[1]  for source lines that are PDL comments
   .  RIGHT_DELIMITER[1]  "    "      "     "   "   "   "
                         /* right delimiter may be null string */

     ...

     end data
```

Fig. 3. Excerpt of data definition showing FILE-INFORMATION table format

```
To EXTRACT_VERSION_CODE

.1      set BEGINNING to the first non-blank character after
          the line number field on the input line (if any)
.2      if (the line begins with the EXTRACT_SIGNAL)
        :    set EXTRACT = 1
        :->(else)
        :    set EXTRACT = 0
        :..end if
.3      increment SEQUENCE_NUMBER[EXTRACT] /* This is the     */
          /* sequence number that should be found in the      */
          /* version code of the incoming line                */
.4      set SEQUENCE_CODE = VERSION_CODE(SEQUENCE_NUMBER[EXTRACT])
          /* this is the sequence number code that will be    */
          /* put on the output line                           */
.5      compute where version code would begin on input line,
          if it were present
.6      if (EXTRACT > 0)
.7      :    set MODE = PDLmode
        :->(else)
.8      :    set MODE = SourceMode
        :..end if
.9      case (MODE)
        :->(0) /* Left delimiter only in version code         */
.10     :    if (LEFT_DELIMITER[EXTRACT] is at proper place
        :    :  on input line) then
.11     :    :    extract 2-character CHECK_CODE and OLD_DATE_CODE
        :    :      from input line
        :    :->(else)
        :    :    set LINE_CODE = NULL
        :    :..end if
        :->(1) /* Both left and right delimiters in version code /*
.12     :    if (LEFT_DELIMITER[EXTRACT] and RIGHT_DELIMITER[EXTRACT]
        :    :  at proper place on input line) then
.13     :    :    extract 2-character LINE_CODE and OLD_DATE_CODE
        :    :      from input line
        :    :->(else)
.14     :    :    set LINE_CODE = NULL
        :    :..end if
        :->(2) /* Version code starts in CODE_COLUMN of input line */
        :    extract 2-character LINE_CODE and OLD_DATE_CODE
        :      from input line
        :..end case
      end to
```

Fig. 4. The procedure for extracting the CHECK-CODE and OLD-DATE-CODE from the input line and computing the proper sequence number for the line

```
To COMPUTE_CHECKSUM

.1      if (EXTRACT > 0) /* i. e., if there is an                */
        :  /* EXTRACT_SIGNAL on the line                          */
.2      :    set BEGINNING = BEGINNING + length(EXTRACT_SIGNAL)
        :      /* so as to jump over the extraction signal in  */
        :      /* the checksum count.                          */
        :..end
.4      set EXPANSION COUNT = 0
        LINE_SUM = 0
.5      loop for i = 1 to length of input line
.6        !  set CH to the i-th character on the input line
.7        !  if (i >= BEGINNING)
.8        !  :    set LINE_SUM = LINE_SUM XOR 2*(LINE_SUM AND 16383);
          !  :        LINE_SUM = LINE_SUM XOR (LINE_SUM/16384) XOR CH
          !  :      /* XOR and AND operate on the binary          */
          !  :      /* equivalents of each operand.               */
          !  :..end if
.9        !  if CH is a tab (control-I)
.10       !  :    increment EXPANSION_COUNT by
          !  :      TAB_WIDTH-[(EXPANSION_COUNT-1) modulo TAB_WIDTH]
          !  :      /* to count the equivalent column that the */
          !  :      /* output line will be in at this point    */
          !  :->(else)
.11       !  :    increment EXPANSION_COUNT only by 1
          !  :..end if
          !..repeat
.12      set LINE_SUM = (LINE_SUM MOD 8281);
             CHECK_CODE = VERSION_CODE(LINE_SUM)
         end to
```

Fig. 5. The procedure for computing the checksum code and column at end of input line

```
To REATTACH_NEW_VERSION_CODE

.1   print "." on the terminal to indicate a line has been processed
.2   if (CODE_COLUMN >0) /* i. e., the code is supposed to be */
     :  /* put in a certain column                              */
.3   :    NumberTabsNeeded = max[(CODE_COLUMN + TAB_WIDTH -2
     :                          - EXPANSION_COUNT)/TAB_WIDTH, 0]
     :->(else)
.4   :    NumberTabsNeeded =0
     :..end if
.5   lop the old version code and any trailing spaces and tab
       characters off of the input line
.6   if (the line would be too long if a version code were appended)
.7   :    print a warning message
.8   :    set WARNING = TRUE
     :->(else)
.9   :    make output line by concatenating input line with
     :      NumberTabsNeeded tab characters, LEFT_DELIMITER[EXTRACT],
     :      CHECK_CODE, and SEQUENCE_CODE
.10  :    if (LINE_CODE = CHECK_CODE)
.11  :    :    append OLD_DATE_CODE to output line
     :    :->(else)
.12  :    :    append TODAY_CODE to output line
     :    :..end if
.13  :    finally, append RIGHT_DELIMITER[EXTRACT] to output line
     :      /* It may be null */
.14  :    if (LINE_CODE <> CHECK_CODE)
.15  :    :    print the output line on the terminal
     :    :..end if
     :..end if
     end to
```

Fig. 6. Procedures for reattaching the version code to the output line

```
Program: CHECKSUM                                                    {hT  (r}
/*******************************************************************/ {RI !(r}
/*                                                                 */ {u# "(r}
/* This program puts a version code on each physical line          */ {f% #(y}
/* of a source file for use in applications where version          */ {7e $(r}
/* statistics of such files are to be extracted and analyzed.      */ {VQ %(r}
/*                                                                 */ {u# &(r}
/*******************************************************************/ {RI '(r}
                                                                     {   ((r}

.1      define program_constants                                     {CF ))3}
.2      declare program_variables                                    {JU *)%}
.3      set_initial_values                                           {N/ +(y}
.4      announce the program title and version identifier            {"V ,(y}
.5      get_todays_date /* encode it into date_code */               {$n -(r}
.6      loop /* for at least one file, and maybe more, as will */    {&` .)3}
          !  /* be determined later by the user                */    {=Q /)3}
.7        !  get_file_names(IN_FILE, OUT_FILE, KIND)                {p: 0(r}
          !    /* KIND may be input or determined from IN_FILE */
          !    /* OUT_FILE is a dummy file name only             */ {-u 1(y}
.8        !  open both files                                        {SI 3(r}
.9        !  set_up_version_delimiters /* using search key from*/  {kb 4(y}
          !    /*  IN_FILE name to identify the particular       */ {v? 5(y}
          !    /*  delimiters to be used by the program.         */ {U3 6)%}
.10       !  reset_parameters /* to initial values, so that     */ {]P 7)%}
          !    /* more than 1 file can be processed, as will     */ {Y? 8)3}
          !    /* be determined later by the user               */ {=Q /)3}
.11       !  loop while (not end_of_file on input)                  {8W :)3}
.12       !    !  input a line /* check for ASCII characters     */ {b( ;(r}
          !    !    /* only and set WARNING switch TRUE if      */ {K4 <(y}
          !    !    /* file contains non-ASCII characters       */ {"e =)%}
.13       !    !  extract_version_code from input line             {oB >(r}
.14       !    !  compute_checksum of input line                   {c$ ?(r}
.15       !    !  reattach_new_version_code                        {,L @(r}
          !    !    /* add tabs to code_column, if indicated.   */ {"\ A(y}
          !    !    /* also, set WARNING to true if line would  */ {0u B(y}
          !    !    /* be too long if the code were appended.   */ {N^ C(y}
.16       !    !  output the line                                  {'B D(r}
          !    !..repeat                                          {X) E(r}
.17       !  close both files                                     {L\ F(y}
.18       !  if WARNING not set to TRUE                           {Jj G(y}
.19       !  :    rename the OUT_FILE as IN_FILE                  {3d H(r}
          !  :->(else)                                           {eJ I(r}
.20       !  :    inform the user that there has been a problem,  {v5 J(y}
          !  :      and that the dummy is the checksummed file    {\[ K(y}
          !  :..end if                                           {%H M(r}
.21       !  prompt user for "what next", and accept answer       {UE N(y}
.22       !....repeat if user desires another run                 {>a O(y}
        end program                                                {rJ P(r}
```

Fig. 7. The CHECKSUM program design language file after execution of the CHECKSUM program on the file

```
                    FILE:  CHECKSUM.PDL,  STATUS:  26MAR82
Program: CHECKSUM
/*****************************************************************/
/*                                                               */
/* This program puts a version code on each physical line        */
/* of a source file for use in applications where version        */
/* statistics of such files are to be extracted and analyzed.    */
/*                                                               */
/*****************************************************************/

.1      define program_constants                                    []
.2      declare program_variables
.3      set_initial_values
.4      announce the program title and version identifier
.5      get_todays_date /* encode it into date_code */
.6      loop /* for at least one file, and maybe more, as will */    []
          !  /* be determined later by the user               */    []
.7        !  get_file_names(IN_FILE, OUT_FILE, KIND)              [M]
          !    /* KIND may be input or determined from IN_FILE */  [A]
          !    /* OUT_FILE is a dummy file name only           */
                                                                   [D]
.8        !  open both files
.9        !  set_up_version_delimiters /* using search key from*/
          !    /*  IN_FILE name to identify the particular     */
          !    /*  delimiters to be used by the program.       */
.10       !  reset_parameters /* to initial values, so that    */
          !    /* more than 1 file can be processed, as will   */  [M]
          !    /* be determined later by the user              */  [RM]
.11       !  loop while (not end_of_file on input)                  []
.12       !    !  input a line /* check for ASCII characters   */
          !    !    /* only and set WARNING switch TRUE if     */
          !    !    /* file contains non-ASCII characters      */
.13       !         !  extract_version_code from input line
.14       !    !       compute_checksum of input line
.15       !    !  reattach_new_version_code
          !    !    /* add tabs to code_column, if indicated.  */
          !    !    /* also, set WARNING to true if line would */
          !    !    /* be too long if the code were appended.  */
.16       !    !  output the line
          !    !..repeat
.17       !  close both files
.18       !  if WARNING not set to TRUE
.19       !  :    rename the OUT_FILE as IN_FILE
          !  :->(else)
.20       !  :    inform the user that there has been a problem,
          !  :      and that the dummy is the checksummed file       [M]
                                                                   [D]
          !  :..end if
.21       !  prompt user for "what next", and accept answer
.22       !....repeat if user desires another run
        end program
```

Fig. 8. The output of the PRINT program, showing changes made in the CHECKSUM program design since March 26, 1982

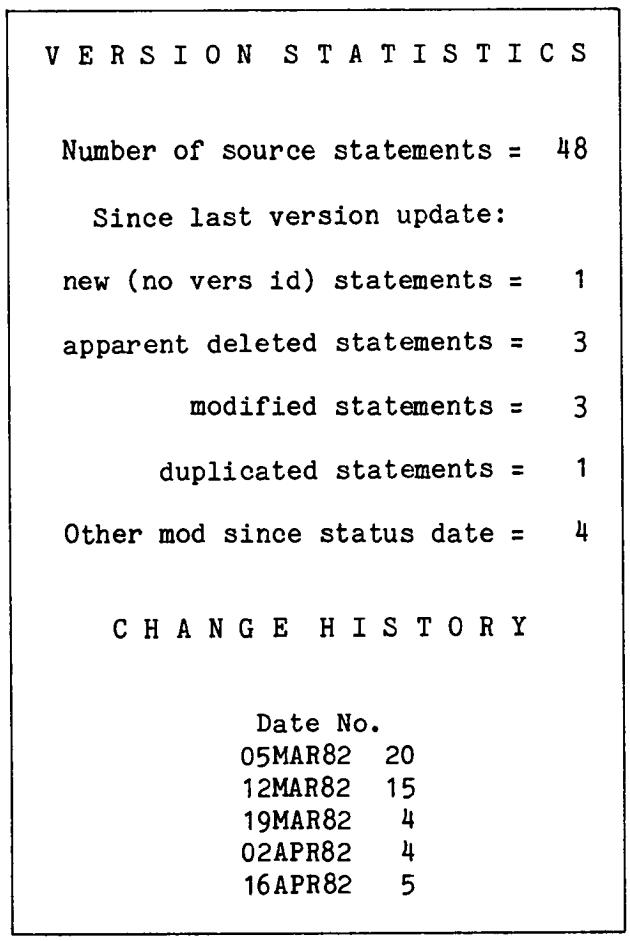


Fig. 9. The Version Statistics and Change History printout for the CHECKSUM program design

Performance Degradation of the Block IV Telemetry System Due to the Presence of a CW Interference

M. K. Sue
Telecommunications Systems Section

The presence of an in-band CW interference can seriously degrade the performance of a telemetry system. Degradation effects for a phase-shift keying (PSK) system can be found in Refs. 1 and 2. The telemetry system employed for deep-space communications is a binary phase-shift keying system (BPSK) with squarewave subcarriers. The use of squarewave subcarriers makes the system less sensitive to in-band interference than a system using sinusoidal subcarriers. A model that allows one to predict the telemetry degradation for the deep-space telemetry system is presented in this article, backed with experimental data.

I. Introduction

This study is part of a continuing effort to investigate the adverse effects that a radio frequency interference (RFI) may have on the Deep Space Receiving System. Depending on the frequency and power level, an RFI can have various effects on the receiving system such as saturating receiver components, generating harmonics and degrading system performance. The saturation effects and the performance degradation of the carrier tracking loop have been documented (Refs. 3, 4) for the Block IV Receiving System. The purpose of this report is to examine the performance degradation of the Block IV telemetry channel due to in-band as well as out-of-band interference. An in-band interference, often called co-channel interference, is one whose frequency falls in the telemetry channel which is assumed to have a bandwidth on the order of twice the subcarrier frequency. An out-of-band interference is one that is not a co-channel one. This includes adjacent channel interference and interference that is remote from the desired channel frequency. An interference outside the telemetry channel can affect the telemetry channel performance; but the effects of an in-band interference are most severe.

An interference to the telemetry channel can produce two kinds of degradation effects. A strong interference can cause the telemetry Subcarrier Demodulation Assembly (SDA) and/or the Symbol Synchronization Assembly (SSA) to lose lock, resulting in a total loss of data. At a weaker level, an interference can increase the telemetry bit error rate. This report analyzes only the latter. The former will be examined in future articles.

II. Analysis

The effect of an in-band interference is first to be analyzed. Figure 1 shows a simplified functional block diagram of the Block IV telemetry receiving system. If we let $S(t)$ and $I(t)$ denote the received signal and interference respectively, then

$$S(t) = \sqrt{2}A \sin(\omega_c t + m\, p(t)\, d(t) + \theta) \quad (1)$$

and

$$I(t) = \sqrt{2}B \sin(\omega_c t + \omega_{sc} t + \Delta\omega t + \phi) \quad (2)$$

where A is the rms voltage of $S(t)$, B is the rms voltage of $I(t)$, ω_c is the carrier frequency in radians per second, $p(t)$ is the subcarrier, $d(t)$ is a binary data stream, $\Delta\omega$ is the RFI offset frequency in radians per second relative to the telemetry subcarrier frequency (ω_{sc}), θ is the phase angle of $S(t)$ m is modulation index and ϕ is the phase angle of $I(t)$, assumed to be a random variable uniformly distributed over $(0,2\pi)$. It is noted that $\Delta\omega$ is positive when the interference frequency is larger than ($\omega_c + \omega_{sc}$). The angular frequency of $I(t)$, ω_i is equal to the sum of ω_c, ω_{sc} and $\Delta\omega$.

The subcarrier $p(t)$ used for deep-space communications is a squarewave subcarrier and can be represented by Fourier series expansion,

$$p(t) = \sum_{k=1}^{\infty} \frac{4}{\pi(2k-1)} \sin((2k-1)\omega_{sc} t) \quad (3)$$

If we let $X_I(t)$ denote the input to the receiving system, then we have:

$$X_1(t) = \sqrt{2}A \sin(\omega_c t + mp(t)\, d(t) + \theta) + \sqrt{2}B \sin(\omega_i t + \phi) + n_1(t) \quad (4)$$

where $n_I(t)$ denotes the white Gaussian noise process with one-sided spectral density N_0.

The interference-to-signal ratio (ISR), which is an important parameter in determining the interference effects, is defined as follows:

$$ISR = P_I/P_D \quad (5)$$

where P_I denotes the power of the interference and P_D denotes the power of the telemetry. From Eqs. (1) and (2), we have:

$$P_I = B^2 \quad (6)$$

$$P_D = A^2 \sin^2(m) \quad (7)$$

Substituting Eqs. (6) and (7) into Eq. (5), we have:

$$ISR = \left(\frac{B}{A \sin(m)}\right)^2 \quad (8)$$

It is noted that ISR as defined in Eqs. (5) and (8) is referenced at the receiver input. Another parameter that is of importance is the energy per symbol to noise spectral density ratio, ST_s/N_0, defined as follows:

$$\frac{ST_s}{N_0} \triangleq \rho_s = \frac{P_D T_s}{N_0} = \frac{\left(A \sin(m)\right)^2 T_s}{N_0} \quad (9)$$

The first operation that the receiving system performs is carrier tracking, where the carrier phase is tracked by a phase locked loop and a locally generated reference signal $r_1(t)$ is created. Assuming perfect tracking, $r_1(t)$ can be written as follows:

$$r_1(t) = 2\cos[(\omega_c - \omega_{IF})\, t + \theta] \quad (10)$$

where ω_{IF} represents the IF frequency in radians per second.

This reference signal is used to downconvert $X_1(t)$ to IF frequency (see Fig. 1). Denoting the downconverted signal by $X_2(t)$ and ignoring double frequency terms and expressing the interference frequency in terms of ω_{IF}, ω_{sc} and $\Delta\omega$, we have

$$X_2(t) = \sqrt{2}A \cos(m) \sin(\omega_{IF} t) + \sqrt{2}A \sin(m) \cos(\omega_{IF} t)\, p(t)\, d(t) + \sqrt{2}B \sin\left((\omega_{IF} + \omega_{sc} + \Delta\omega)\, t + \phi - \theta\right) + n_2(t) \quad (11)$$

where $n_2(t)$ is the filtered noise process at IF frequency with one-sided spectral density N_0.

The subcarrier tracking loop then tracks the subcarrier and generates a subcarrier reference signal. Ignoring subcarrier

tracking error, then the subcarrier reference $r_2(t)$ is identical to $p(t)$; i.e.,

$$r_2(t) = p(t) = \sum_{k=1}^{\infty} \frac{4}{\pi(2k-1)} \sin((2k-1)\,\omega_{sc}t). \tag{12}$$

The subcarrier reference signal is then used to demodulate $X_2(t)$.

Denoting the demodulated signal by $X_3(t)$, i.e.,

$$X_3(t) = X_2(t)\, r_2(t) \tag{13}$$

and substituting Eqs. (11) and (12) into Eq. (13), we have

$$\begin{aligned} X_3(t) = {} & \sqrt{2}A \cos(m) \sin(\omega_{IF}t) \cdot \\ & \sum_{k=1}^{\infty} \frac{4}{\pi(2k-1)} \sin[(2k-1)\omega_{sc}t] \\ & + \sqrt{2}A \sin(m) \cos(\omega_{IF}t)\, d(t) \cdot \\ & \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \frac{16}{\pi^2(2k-1)(2j-1)} \begin{Bmatrix} \sin[(2k-1)\omega_{sc}t] \cdot \\ \sin[(2j-1)\omega_{sc}t] \end{Bmatrix} \\ & + \sqrt{2}B \sin((\omega_{IF} + \omega_{sc} + \Delta\omega)\, t + \phi - \theta) \cdot \\ & \sum_{k=1}^{\infty} \frac{4}{\pi(2k-1)} \sin[(2k-1)\omega_{sc}t] \\ & + n_3(t) \end{aligned} \tag{14}$$

The second term in Eq. (14) is the desired telemetry signal and can be reduced to $\sqrt{2}A \sin(m) \cos(\omega_{IF}t)\, d(t)$ because the result of the double summation is unity. The noise process is represented by $n_3(t)$. The output of the mixer $X_3(t)$ is then filtered by a predetection filter (PDF). This filter is assumed to be an ideal filter that passes all in-band signals and completely rejects out-of-band signals. Ignoring interference and carrier components more than $\omega_{sc}/2\pi$ hertz from the IF frequency, the predection filter output is:

$$\begin{aligned} X_4(t) = {} & \sqrt{2}A \sin(m)\, d(t) \cos(\omega_{IF}t) \\ & + \frac{2\sqrt{2}}{\pi} B \cos[(\omega_{IF} + \Delta\omega)\, t + \phi - \theta] \\ & + n_4(t) \end{aligned} \tag{15}$$

where $n_4(t)$ is the filtered noise process. If we ignore the broadening of the noise spectrum due to the multiplication of $n_2(t)$ by $r_2(t)$ in Eq. (13), then the one-sided noise spectral density of $n_4(t)$ is approximately equal to N_0.

We have assumed in Eq. (15) that the interference component at $\omega_{IF} + \Delta\omega$ is inside the passband of the filter so that there is no attenuation caused by filtering. This assumption requires that $\Delta\omega$ be small compared to the bandwidth of the predetection filter and it represents a worst case condition.

After subcarrier demodulation the signal is multiplied by a coherent 10 MHz reference signal to downconvert $X_4(t)$ to baseband. Denoting this reference signal by $r_3(t)$, we have:

$$r_3(t) = \sqrt{2} \cos(\omega_{IF}t) \tag{16}$$

Ignoring the double frequency terms, the output of the multiplier denoted by $X_5(t)$ is:

$$X_5(t) = A \sin(m)\, d(t) + \frac{2B}{\pi} \cos(|\Delta\omega|\, t + \phi - \theta) + n_5(t) \tag{17}$$

where $|\ |$ denotes the absolute value and $n_5(t)$ is the noise process at baseband with one-sided spectral density N_0.

The baseband signal including interference and noise components is then fed to the Symbol Synchronization Assembly (SSA) for symbol synchronization and symbol detection. Symbol detection is achieved using an integrate and dump circuitry.

Ignoring symbol synchronization error, the probability of symbol error P_s has been derived in Appendix A and is given in Eq. (18).

$$P_s = \frac{1}{2\pi} \int_0^{\pi} \operatorname{erfc}\left[\rho_c^{1/2} \left(1 + \frac{2}{\pi} R \cos\phi\right)\right] d\phi \tag{18}$$

where

$$\operatorname{erfc}(x) = \frac{2}{\pi} \int_X^{\infty} \exp(-y^2)\, dy.$$

and

$$R^2 = ISR.$$

Using a different approach, Rosenbaum (Ref. 2) has obtained an expression for the probability of symbol error for a PSK system under the influence of a CW in-band interference. The results of Ref. 2 differ from Eq. (18) by the multiplication factor $2/\pi$ in front of $R \cos \phi$. This factor is due to the use of squarewave subcarrier and it makes the telemetry system less sensitive to in-band CW interference by approximately 4 dB. Equation (18) indicates that the performance degradation in terms of increasing symbol error rate is independent of the symbol rate for a given ST_s/N_0 and a given R.

In the absence of interference, R equals zero and Eq. (18) becomes

$$P_s(R = 0) = \frac{1}{2} \operatorname{erfc}(\rho_s^{1/2}) \tag{19}$$

which is the baseband error probability for coherent reception of PSK signals.

RFI degradation effects can be expressed in terms of an equivalent loss in ST_s/N_0. If we let $P_s(\rho_s, R)$ denote the probability of symbol error in the presence of an interference having an ISR equal to R^2 and a ST_s/N_0 equal to ρ_s, then the equivalent loss in ST_s/N_0 denoted by γ is

$$\gamma = \{\operatorname{erfc}^{-1}[2P_s(\rho_s, R)]\}^2/P_s \tag{20}$$

where erfc^{-1} denotes the inverse of erfc. Equations (18) and (20) represent our model for the prediction of uncoded telemetry performance degradation due to in-band CW RFI.

III. Experimental Verification

Equation (18) gives the symbol error performance as a function of ISR and ST_s/N_0 for a system with perfect carrier, subcarrier and symbol synchronization. A set of curves have been generated using Eqs. (18) and (20) and are shown in Figs. 2 and 3. Experimental verification of error rate curves was performed in TDL. A comparison of analytical and experimental data is shown in Fig. 4 after the observed ST_s/N_0 has been compensated for TDL system loss. The system loss, which is mainly caused by noise, includes losses due to errors in carrier tracking, subcarrier tracking and symbol synchronization. It was measured in the absence of interference by comparing the ST_s/N_0 required to obtain a given symbol error rate in TDL to the theoretical ST_s/N_0 corresponding to same error rate. The difference is taken as the system loss. As indicated in Fig. 4, TDL data agrees very well with theoretical data except at strong interference level (ISR = 0 dB) where the actual degradation is slightly more severe than predicted, by approximately 1 dB in interference to signal ratio. When the interference exceeds 0 dB, ISR total loss of data occurs due to loss of lock in SSA or SDA, or both.

The TDL measurements shown in Fig. 4 were based on a data rate of 115.2 kbps. Other data rates (10 and 50 kbps) had been tested and there was no noticeable difference in performance degradation. This is in agreement with the analysis (Eq. 18).

IV. Degradation on the Coded System

The effect of an in-band CW interference on the performance of a system using convolutional codes is difficult to analyze mathematically. Our derivation of the bit error expression under the influence of a CW interference is partly based on experimental observations. For the MCD (Maximum Likelihood Convolutional Decoding) implemented in the DSN, the probability of bit error in the absence of interference can be approximated by Eq. (21) (Ref. 5).

$$P_E(\text{No RFI}) \approx A \exp(B\rho) \tag{21}$$

where $A = 85.7501$, $B = -5.7230$ and ρ is the signal energy per bit to noise spectral density ratio (E_b/N_0). Equation (21) is applicable for $K = 7$, rate 1/2 convolutional codes with 3-bit quantization and $\rho = 2\rho_s$.

If we can estimate the effective signal energy per bit to noise spectral density ratio ρ_e, then Eq. (21) can be applied to estimate the probability of error under the influence of an interference. From Eq. (20), the effective signal to noise ratio for a given R and ρ_s can be modelled as

$$\rho_e = 2\{\operatorname{erfc}^{-1}[2P_s(\rho_s, R)]\}^2. \tag{22}$$

Substituting Eq. (22) for ρ in Eq. (21), the probability of bit error under the influence of a CW RFI is

$$P_E \approx A \exp(2B\{\operatorname{erfc}^{-1}[2P_s(\rho_s, R)]\}^2) \tag{23}$$

for a given R and ρ_s.

Since Eq. (23) is based on the analytical model for the uncoded case, i.e., Eqs. (18) and (20), the accuracy of Eq. (33) depends on the accuracy of Eq. (18) or (20). As previously mentioned, the analytical model for the uncoded case works

very well for ISR less than 0 dB. At 0 dB, the analytical model is about 1 dB less sensitive to RFI than the actual system. To account for this deficiency, it is necessary to modify Eq. (23) slightly by placing restrictions on its applicability.

$$P_E = \begin{cases} A \exp\left(2B\{\text{erfc}^{-1}\ [2P_s(\rho_s, R)]\}^2\right) & ISR < 0 \text{ dB} \\ A \exp\left(2B\{\text{erfc}^{-1}\ [2P_s(\rho_s, R^*)]\}^2\right) & ISR = 0 \text{ dB} \end{cases} \tag{24}$$

where 20 log R^* = 1 dB. Equation (24) represents our model for the prediction of telemetry performance degradation for the coded case due to in-band CW RFI.

A set of curves have been generated using Eq. (24) and they are shown in Fig. 5. Experimental verification was performed in TDL. To avoid the need for excessive testing time, the range of E_b/N_0 was limited to from 2 to 5 dB, excluding TDL system losses. This range of E_b/N_0 corresponds to a bit error rate of 10^{-2} to 10^{-6}, which approximately covers the normal DSN operating range. A comparison of TDL data to those predicted by Eq. (24) is shown in Fig. 6 and they are in good agreement.

V. Translation Effects

In the preceding analysis, we have assumed that the interference is an in-band interference; i.e., the frequency separation between the interference and the subcarrier frequency is on the order of the subcarrier frequency or less. Although the telemetry system is most sensitive to in-band interference, out-of-band interference can degrade telemetry performance through a process often termed as translation where an out-of-band interference gets multiplied by a subcarrier reference during the subcarrier demodulation process and creates an in-band interfering component that falls inside the passband of the predetection filter. If we let $I_i(t)$ denote the interference at the ith position of the receiver (see Fig. 1), then after subcarrier demodulation the interference becomes (Eq. 14):

$$I_3(t) = \sqrt{2}B|H(\omega_i)| \sin\left[(\omega_{IF} + \omega_{sc} + \Delta\omega)\, t + \phi - \theta + \phi_H(\Delta\omega)\right] \sum_{k=1}^{\infty} \frac{4}{\pi(2k-1)} \sin\left[(2k-1)\omega_{sc} t\right] \tag{25}$$

where $|H(\omega_i)|$ and $\phi_H(\omega_i)$ are the amplitude and phase of $H(j\omega)$ evaluated at ω_i and $H(j\omega)$ represents the transfer function of the first part of the receiver; i.e.,

$$H(j\omega) = \frac{x_2(j\omega)}{x_1(j\omega)}$$

where $x_1(j\omega)$ and $x_2(j\omega)$ are the Fourier transform of $x_1(t)$ and $x_2(t)$ respectively. In Eq. (14), $|H(\omega_i)|$ has been assumed to be 1.

For a given ω_i, the phase angle $\phi_H(\omega_i)$ is a constant and does not affect the outcome and thus can be ignored. The factor $|H(\omega_i)|$ in Eq. (25) represents the attenuation on the out-of-band interference. A plot of $|H(\omega)|$ as a function of frequency is shown in Fig. 7 for the Block IV receiver. The frequency response was measured in TDL.

It is obvious from Eq. (25) that a single CW RFI has been, through the process of subcarrier demodulation, decomposed into a set of interfering components each having a radian frequency ω_k and power P_k; $k = 1, 2, 3 \ldots$, where

$$\omega_k = \begin{cases} \omega_{IF} + \Delta\omega + 2k\omega_{sc} \\ \omega_{IF} + \Delta\omega - 2(k-1)\omega_{sc} \end{cases} \qquad k = 1, 2, 3 \ldots \tag{26}$$

and

$$P_k = \left[\frac{2B|H(\omega_i)|}{\pi(2k-1)}\right]^2 \qquad k = 1, 2, 3 \ldots \tag{27}$$

These interfering components are equally spaced at $2\omega_{sc}$ radians/sec apart. If we assume that the bandwidth of the predetection filter is on the order of $2\omega_{sc}$, then one of these components would fall inside the predetection bandwidth and would be passed practically without attenuation. The power of this interfering component at the predetection filter output is given by Eq. (27), with k being related to the interference frequency ω_i as follows

$$CPOI\left(\frac{\omega_i - \omega_C}{\omega_{sc}}\right) = 2k - 1 \tag{28}$$

where $CPOI(x)$ denotes the positive odd integer that is closest to x. If x is equidistant from two odd integers, the larger integer is arbitrarily assigned to $CPOI(x)$.

In order for out-of-band CW interference to produce the same degradation effect on the telemetry system as an in-band interference, it is obvious from Eqs. (27) and (28) that an

additional amount of interference power is needed to compensate for the attenuation due to the frequency response of the telemetry channel and the spreading effects due to multiplication by the subcarrier reference. If we let ΔP_K denote the additional power and P_1 denote the power level of an in-band interference at the predetection filter output, then

$$\Delta P_k \text{ (dB)} = 10 \log \frac{P_1}{P_k} \tag{29}$$

The power of an in-band interference at the filter output can be obtained from Eq. (15) or from Eq. (27) by setting $k = 1$ and $|H(j\omega)| = 1$,

$$P_1 = (2B/\pi)^2 \tag{30}$$

Substituting Eq. (30) and Eq. (27) into (29), we have

$$\Delta P_k \text{(dB)} = 20 \log [(2k - 1)/|H(\omega_i)|] \tag{31}$$

Eq. (31) when used in conjunction with Eqs. (18), (20) and (24) provides the necessary models to account for the effects of out-of-band interference.

Eq. (3) has been evaluated for selected frequencies as shown in Table 1. Experimental verification of the translation mechanism was performed in TDL using the Block IV receiver. The power level of an in-band interference required to produce a symbol error rate of 2.5×10^{-3} was first determined and used as a reference. (The symbol error rate was approximately 1.0×10^{-3} in the absence of interference.) The power level of an out-of-band interference was then adjusted to produce the same symbol error rate (2.5×10^{-3}). The additional power level was obtained by subtracting the reference power level from the power level of the out-of-band interference. TDL data are also included in Table 1; they are in good agreement with analytical data.

Table 1 reveals a potential RFI problem associated with the unique feature of the receiving system. The wide bandwidth of the telemetry channel offers good performance in a noisy environment, but leaves the telemetry channel open for out-of-band interference. As indicated in Table 1, a CW RFI with a moderate power level and at a frequency relatively remote from the frequency of the desired signal can produce significant performance degradation through the translation mechanism. At 2.5 MHz away, which corresponds to a frequency separation of about seven S-band channels in the DSN channel plan, a CW RFI needs to be only about 17 dB stronger than a co-channel interference in order to produce the same degradation. The protection against out-of-band interference is therefore considered to be inadequate with the existing system. Improvement in RFI immunity should be considered in future designs. Significant improvement would ease the problem of channel selection for future missions and allow more efficient use of band allocations.

VI. Conclusion

Telemetry performance degradation of the Block IV receiving system due to CW interference has been examined. The use of a square wave subcarrier instead of a sine wave has both advantage and disadvantage in terms of RFI protection. On one hand, it makes the system less sensitive to in-band CW interference by about 4.0 dB. On the other hand, it makes the system susceptible to out-of-band interference due to the translation mechanism. Despite simplifying assumptions in the analytical model, experimental data obtained in TDL indicate that the model is sufficiently accurate throughout nominal operating range. The telemetry degradation models are provided by Eqs. (18), (20) and (24) for in-band and out-of-band interference and for the coded and uncoded systems.

Acknowledgment

The author wishes to express his appreciation to J. F. Weese, who performed the experimental measurements in TDL, and D. R. Hersey for technical consultations.

References

1. Koerner, M. A., *Effect of Interference on a Binary Communication Channel Using Known Signals*, Technical Report 32-1281, Jet Propulsion Laboratory, Pasadena, Calif., Dec. 1, 1968.

2. Rosenbaum, A. S., "PSK Error Performance with Gaussian Noise and Interference," *Bell System Technical Journal*, Feb. 1969.

3. Hersey, D. R., and Sue, M. K., "Maximum CW RFI Power Levels for Linear Operation of the DSN Block IV Receiver at S-Band Frequencies," *DSN Progress Report 42-56, January and February 1980*, Jet Propulsion Laboratory, Pasadena, Calif. April 15, 1980.

4. Sue, M. K., "Block IV Receiver Tracking Loop Performance in the Presence of a CW RFI," *DSN Progress Report 42-60, September and October 1980*, Jet Propulsion Laboratory, Pasadena, Calif., Dec. 15, 1980.

5. Webster, L., "Maximum Likelihood Convolutional Decoding (MCD) Performance Due to System Losses," *DSN Progress Report 42-34*, Jet Propulsion Laboratory, Pasadena, Calif., Aug. 15, 1976.

Table 1. Additional interference power level required for an out-of-band CW interference to produce the same effect as an in-band interference (experimental results are based on a symbol error rate of 2.5×10^{-3} in the presence of an interference or 1.0×10^{-3} in the absence of it)

Interference frequency relative to carrier frequency $(\omega_i-\omega_c)/2\pi$ kHz	Value of k[1]	$\lvert H(\omega_i)\rvert^2$, dB	20 log(2k−1), dB	ΔP_K, dB	
				Predicted	Measured[2]
360	1	0	0	0	0
1080	2	0	9.5	9.5	10.1
1800	3	0	14.0	14.0	14.2
2520	4	0	17.0	17.0	17.2
3240	5	0	19.1	19.1	20.2
3960	6	−2.0	20.8	22.8	23.7
4690	7	−3.0	22.3	25.3	25.4
5400	8	−6.0	23.5	29.5	30.9
6120	9	−20.0	24.6	44.6	46.2
6840	10	−42.0	25.6	67.6	70.2

[1]Equation (28).
[2]Block IV receiver.

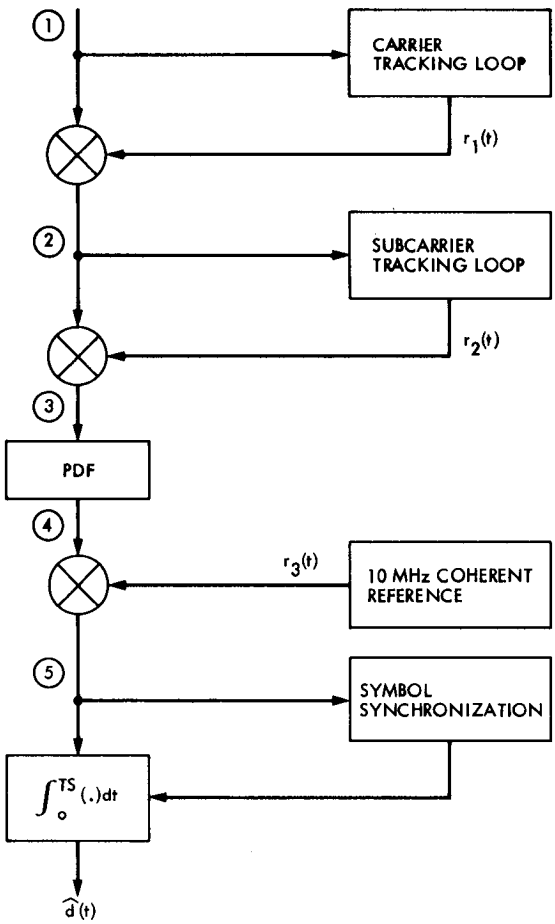


Fig. 1. Functional block diagram of the Block IV Telemetry Receiving System

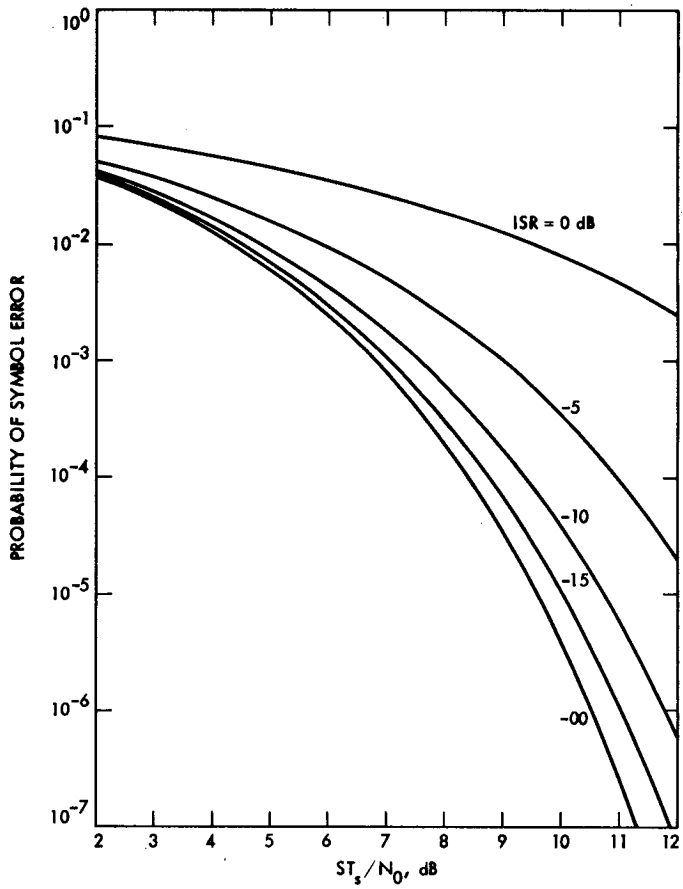


Fig. 2. Probability of symbol error vs ST_s/N_0

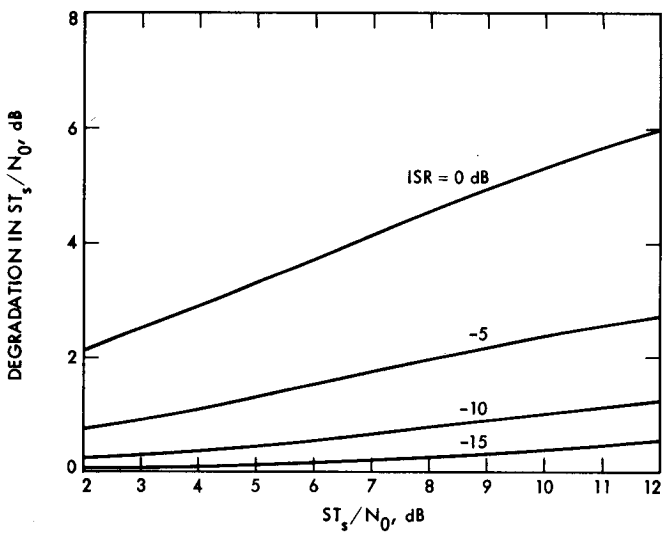


Fig. 3. Degradation in ST_s/N_0 vs ST_s/N_0, dB

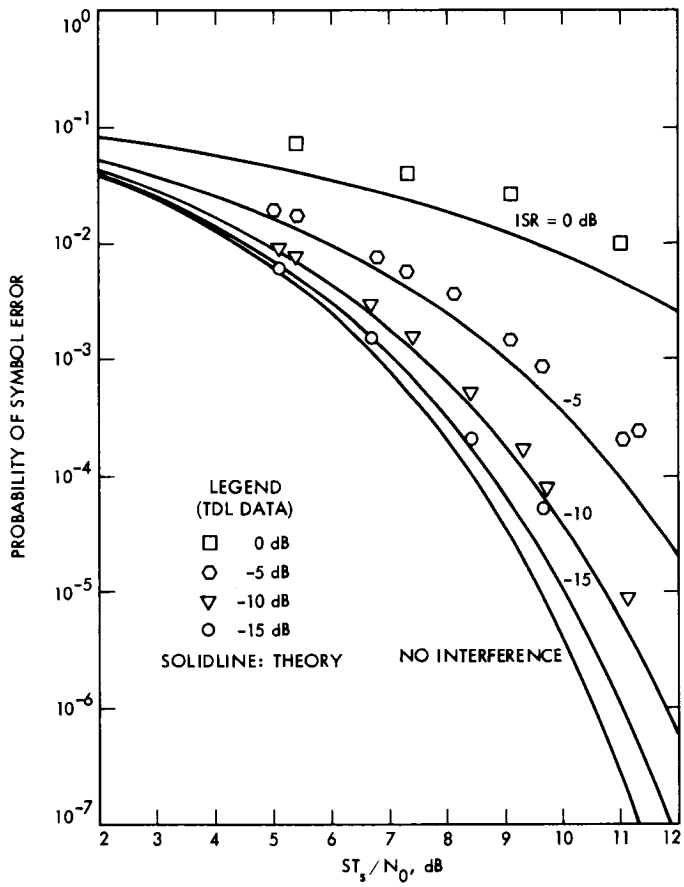


Fig. 4. Comparison of experimental and analytical results

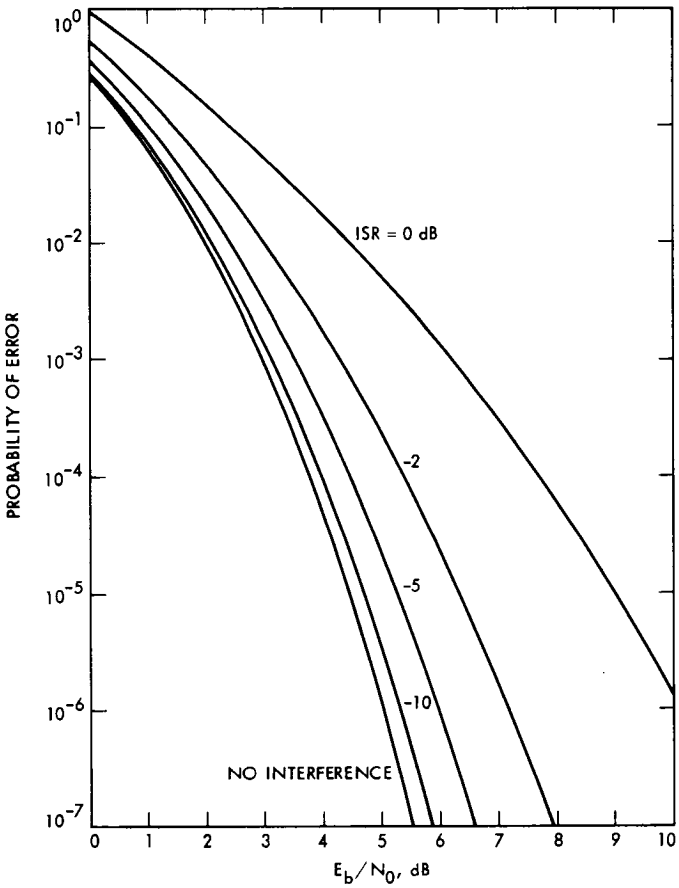


Fig. 5. Probability of bit error vs E_b/N_0 for selected values of ISR (Eq. 24)

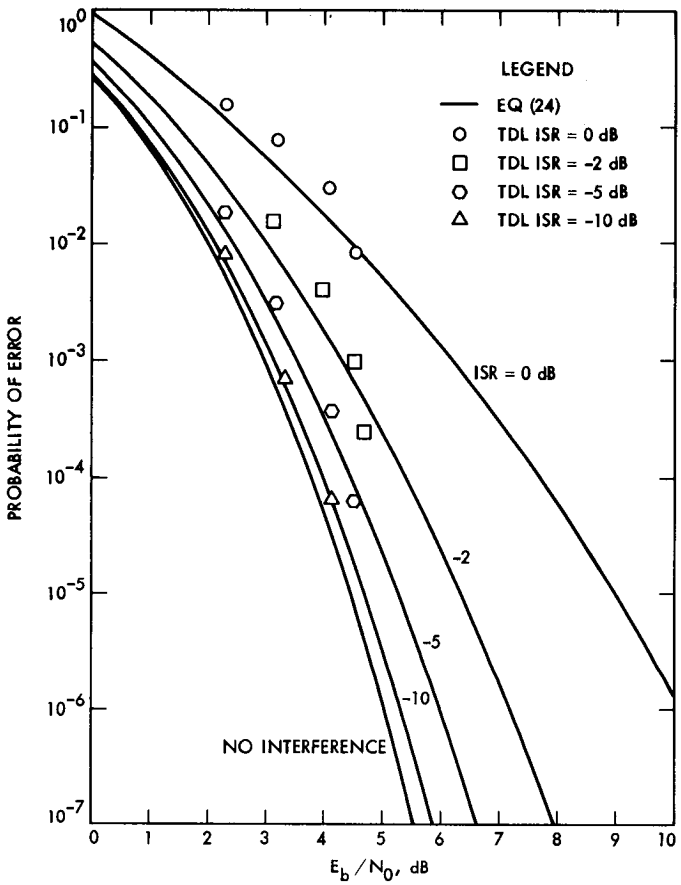


Fig. 6. Comparison of TDL and Eq. (24) results

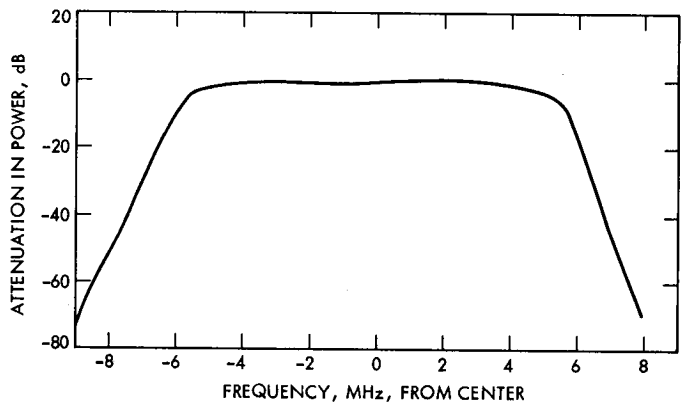


Fig. 7. TDL TLM channel frequency response vs frequency, Block IV receiver S-band (from receiver input to SDA input)

Appendix A

Derivation of Probability of Symbol Error

The probability of symbol error under the influence of a CW interference is derived in this appendix. The input to the data detector consists of three components: signal, interference and noise. Denoting the input by $y(t)$, we have from Eq. (17) the following expression:

$$y(t) = a\,d(t) + b\cos(|\Delta\omega|\,t + \phi - \theta) + n_5(t) \qquad \text{(A-1)}$$

where we have replaced $A \sin(m)$ in Eq. (17) by a in Eq. (A-1) and $2B/\pi$ by b. The noise process $n_5(t)$ is the same noise process as in Eq. (17) and is assumed to be Gaussian with a correlation function $R(\tau) = N_0/2\,\delta(\tau)$.

The data detector performs data detection by integrating $y(t)$ over one symbol period T_s and announces its decision $\hat{d}$ according to the following rule:

$$\hat{d} = \begin{cases} d_1 = 1 & \text{if } q \geqslant 0 \\ d_2 = -1 & \text{if } q < 0 \end{cases} \qquad \text{(A-2)}$$

where

$$q = \int_0^{T_s} y(t)\,dt \qquad \text{(A-3)}$$

The interference in Eq. (A-1) can be modeled as $b\cos\psi$ with ψ being a random variable uniformly distributed over $(0, 2\pi)$. To determine the probability of symbol error P_s, it is necessary to first determine the conditional statistics of q. It can be shown that for a given $\psi = \psi_0$ and $d(t) = d_1 = 1$, we have the following conditional statistics:

$$E[q|\psi_0, d_1] = (a + b\cos\psi_0)\,T_s,$$

$$VAR[q|\psi_0, d_1] = \frac{N_0}{2}T_s$$

and

$$p(q|\psi_0, d_1) = \frac{1}{\sqrt{\pi N_0 T_s}}\exp\left(-(q - (a + b\cos\psi_0)\,T_s)^2/N_0T_s\right) \qquad \text{(A-4)}$$

where $E[\ ]$, $VAR[\ ]$, $p[\ ]$ and T_s denote the expectation, variance, probability density function and symbol duration respectively.

The probability of making an error given that d_1 was transmitted is equal to the probability that q is less than 0; i.e.,

$$P_s(s|\psi_0, d_1) = \text{prob}(q < 0|\psi_0, d_1) = \int_{-\infty}^{0} p(q|\psi_0, d_1)\,dq \qquad \text{(A-5)}$$

where $P_s[\ ]$ denotes the symbol error probability conditioned on ψ_0 and d_1. After simplification Eq. (A-5) becomes:

$$P_s(s|\psi_0, d_1) = \frac{1}{2}\,\text{erfc}\left(\frac{(a + b\cos\psi_0)\,T_s}{\sqrt{N_0T_s}}\right) \qquad \text{(A-6)}$$

where

$$\text{erfc}(x) \overset{\Delta}{=} \frac{2}{\sqrt{\pi}}\int_x^{\infty} exp(-y^2)\,dy$$

Similarly, the probability of making an error given that $d_2 = -1$ was sent is equal to the probability that q is greater than or equal to 0 and is equal to:

$$P_s(s|\psi_0, d_2) = \text{prob}(q \geqslant 0|\psi_0, d_2)$$

$$= \frac{1}{2}\,\text{erfc}\left(\frac{(a - b\cos\psi_0)\,T_s}{\sqrt{N_0T_s}}\right) \qquad \text{(A-7)}$$

If we assume equal probability of occurrence for d_1 and d_2, then the probability of symbol error averaged over $d(t)$ and conditioned on $\psi = \psi_0$ can be calculated as follows:

$$P_s(s|\psi_0) = \text{prob}(d(t) = d_1)\,P_s(s|\psi_0, d_1) + \text{prob}(d(t) = d_2)\,P_s(s|\psi_0, d_2)$$

$$= \frac{1}{4}\,\text{erfc}\left(\frac{(a + b\cos\psi_0)\,T_s}{\sqrt{N_0T_s}}\right) + \frac{1}{4}\,\text{erfc}\left(\frac{(a - b\cos\psi_0)\,T_s}{\sqrt{N_0T_s}}\right) \qquad \text{(A-8)}$$

The probability of symbol error can be obtained from $P_s(s|\psi_0)$ by averaging over ψ_0.

$$P_s(s) = \int_{\psi_0} P_s(s|\psi_0)\, p_\psi(\psi_0)\, d\psi_0.$$

$$= \frac{1}{8\pi}\int_0^{2\pi} \text{erfc}\left(\frac{(a + b\cos\psi_0)\,T_s}{\sqrt{N_0 T_s}}\right) d\psi_0$$

$$+ \frac{1}{8\pi}\int_0^{2\pi} \text{erfc}\left(\frac{(a - b\cos\psi_0)\,T_s}{\sqrt{N_0 T_s}}\right) d\psi_0 \quad \text{(A-9)}$$

where $P_\psi(\ )$ is the probability density function of ψ.

The two terms on the right-hand side of the above equation are equal and can be combined to yield:

$$P_s(s) = \frac{1}{2\pi}\int_0^{2\pi} \text{erfc}\left(\frac{(a + b\cos\psi_0)\,T_s}{\sqrt{N_0 T_s}}\right) d\psi_0 \quad \text{(A-10)}$$

Expressing $P_s(s)$ in terms of A and B and substituting Eqs. (5) – (8) into the above equation, we have after simplification of the following:

$$P_s(s) = \frac{1}{2\pi}\int_0^{\pi} \text{erfc}\left(\rho_s^{1/2}\left(1 + \frac{2}{\pi} R\cos\psi_0\right)\right) d\psi_0 \quad \text{(A-11)}$$

where $R^2 = ISR$.

Selection of Frequencies for Deep-Space Telecommunications

N. F. de Groot and M. K. Sue
Telecommunications Systems Section

This report explains the procedure used by the Jet Propulsion Laboratory to select and recommend frequencies to be used for deep-space telecommunications. Recommendations are made for missions conducted by the United States and also for other countries or organizations upon request. The report was prepared for the October 1981 meeting of the Space Frequency Coordination Group, held at Oberpfaffenhofen, Federal Republic of Germany. The group included representatives from ten countries and three international organizations, all interested in frequency management issues related to space research.

I. Introduction

As more nations plan to make use of deep-space allocations it becomes increasingly important to cooperate in the selection and assignment of radio frequencies for their new missions. This cooperation is needed to avoid or minimize the possibility of radio frequency interference (RFI) between the telecommunications links of the several missions using each band.

The process used by the Jet Propulsion Laboratory to select frequencies for deep-space missions is described in this report. The process has been used for U.S. missions. It has also been used in response to requests by Japan and the European Space Agency. For appropriate frequency selections to occur, all existing and planned deep-space missions must be considered.

The frequency selection process described in this report deals only with the potential for RFI between deep-space telecommunications links. It does not consider band sharing with other services. It also does not deal with adjacent or harmonic band interference, or with RFI from other spurious emissions.

II. Allocated Bands

The pairs of bands allocated for deep-space telecommunications are shown in Table 1.

III. Channel Plans and Coherence

To provide for orderly selection and assignment of frequencies for deep-space missions, channel plans have been developed within the United States. The plans were based on bandwidth, hardware implementation and frequency ratio considerations. All U.S. deep-space missions utilize frequencies included in the channel plans.

To provide for spacecraft navigation and some types of scientific measurement it is necessary that uplink (Earth-to-space) and downlink (space-to-Earth) transmissions be phase-coherent. This means that the frequency received by the spacecraft transponder must be translated by a fixed ratio and used to control the downlink frequency from the spacecraft.

The requirement for coherence applies between the uplink and downlink frequencies of a band pair. The requirement can also apply to simultaneous transmissions in more than one band pair. Of the four band pairs listed in Table 1, channel plans have been developed for the first two. These plans are shown in Table 2. Table 3 shows the frequency ratios associated with the channel plans.

In Table 2 we see that the channel center frequencies in column 3 lie within the 2290-2300 MHz downlink band. Channel center frequencies in the other bands are a necessary result of the frequency ratios shown in Table 3. A particular channel number, for example channel 17, specifies frequencies in all four bands.

Because of the spacing between allocations and the frequency ratios embodied in the channel plans, some channels in each band are not usable. When complete coherence between the two uplink and two downlink bands is required, the frequencies selected can only be chosen from channels 5-27. Similar channel planning has been proposed for the higher frequency bands. The details of this work are beyond the scope of present frequency selection studies and this report.

IV. The Process of Channel Selection

The selection process is based on calculations and analysis of interference-to-signal power ratios (ISR) as a function of time for each possible pair of missions. The initial calculation assumes that both spacecraft are using the same channel. Co-channel operation is often possible because of the very narrow beams of Earth station antennas and the diverse position and motions of spacecraft engaged in deep space missions.

The worst case ISR is compared to a criterion of acceptable interference. If the ISR meets this criterion for all spacecraft, any channel may be selected. This is true because the co-channel condition was used for the calculation. If the criterion of acceptable interference is not met for the co-channel case, alternatives must be examined. A separate, unused channel may be required. Another possibility is that interference may be acceptable at certain times during each mission.

The ISR calculation is made for each of the two spacecraft in a pair. First, one is assumed to be the desired spacecraft and the other is considered a potential source of interference. The calculation is then repeated for the opposite situation. When separate channels are required, the selection is based on more detailed analysis of the degree of interference, the spectral characteristics of the interfering signals, and the doppler shift caused by relative spacecraft motion.

V. Modes of Interference

The radio frequency signal for deep-space communications normally includes a carrier and one or more sets of data sidebands. Examples of two sets of data sidebands are combined telemetry and ranging signals on a downlink, or combined command and ranging signals on the uplink. When separate data streams are carried on individual subcarriers, there are additional intermodulation products.

Given the necessary frequency relationships, the signals to or from two spacecraft may interfere in one or more of the following ways:

(1) Carrier-to-carrier

(2) Data-to-carrier

(3) Intermodulation product-to-carrier

(4) Carrier-to-data

(5) Data-to-data

(6) Intermodulation product-to-data

These interference modes can occur between uplinks or between downlinks.

In addition, there is a quite different interference mode for the uplink case. Consider the situation where one uplink signal is being transmitted and intended for a particular spacecraft. If this signal is received by another spacecraft and has sufficient strength and the necessary frequency components, it is possible for the receiver in the unintended spacecraft to lock to the uplink signal. This must be avoided if independent operation of the two spacecraft is to be maintained. This interference mode is called one way uplink interference.

In practice, the carrier and the telemetry sidebands are usually the most susceptible to interference between downlinks. The predominant mode for uplinks is the one-way uplink interference.

VI. Interference Protection Ratio

The acceptable ratio of interference to signal power is called the interference protection ratio. The protection ratio used for channel selection purposes is –15 dB, i.e., the maximum allowable effective interference power is 15 dB below the signal power. The signal power refers to the portion of the received signal that is related to a particular function: carrier tracking, telemetry, command, or ranging. A –15 dB ratio will produce negligible effect on carrier tracking performance (Ref. 1, 2), 0.4 dB degradation of telemetry performance and 1.0 dB degradation of command performance, assuming that both the

telemetry and command are operating at a symbol error rate of 10^{-5}.

The protection ratio may seem conservative. It is justified by the fact that the frequency selection study usually is performed in a very early phase of a mission when uncertainties exist about many parameters which can affect the interference situation. Additionally, it is a goal of channel selection to provide the safest environment so that mission planners will have maximum flexibility.

The maximum acceptable uplink one-way interference is called the uplink one-way threshold. It is equal to the spacecraft receiver sensitivity. This sensitivity is usually in the range from -155 to -165 dBm. The maximum allowable interference power level as determined by either the protection ratio or the uplink one-way threshold is referred to as the interference threshold. Interference is said to exist whenever this value is exceeded.

VII. Interference Analysis

There are four steps in the interference analysis and channel selection process:

(1) Determine the interference and the signal power levels.

(2) Determine the likely modes of interference: carrier-to-carrier interference, carrier-to-data interference, etc.

(3) Determine the time and duration of potential interference.

(4) Select one or more channels to minimize the potential interference. Avoid potential interference during critical mission events.

To achieve these, it is necessary to have the following information for all the missions involved:

(1) Characteristics of the telecommunication system.

(2) Expected modes of operation as a function of time.

(3) Dates of important mission events.

(4) Orbital elements that specify the mission trajectory.

A complete list and description of the above items are provided in Appendix A.

A computer program has been developed to assist in the calculations needed for interference analysis. This program examines two missions at a time. It accepts as its input a set of orbital elements that completely specify the trajectories of the missions being examined. Based on these orbital elements, it computes for both missions the spacecraft-to-Earth range, the doppler rate, and the angular separation between the two spacecraft. From these data and an assumed fixed e.i.r.p., it then computes the ratio of the total received signal power to the total received interference power (TSIR)[1] and the uplink interference power level as received by the spacecraft. The calculations are made for selected intervals of time during the period of operation that is common for a particular pair of missions.

A. Downlink Interference Analysis

The Earth station receiving antenna is assumed to be pointing at the desired spacecraft and receiving data from it (Fig. 1). Downlink interference occurs when the signal from the other spacecraft exceeds the protection criterion. The downlink TSIR is calculated by the computer program for both missions assuming an equal e.i.r.p. The equations for the computation are:

$$TSIR_1 = G_{MAX} - G(\theta) - 20 \log_{10} (R_1/R_2)$$

$$TSIR_2 = G_{MAX} - G(\theta) + 20 \log_{10} (R_1/R_2)$$

where G_{MAX} is the gain of the receiving antenna in dBi, R is the spacecraft-to-Earth range, $G(\theta)$ is the gain of the receiving antenna in the direction of θ, and the subscripts 1 and 2 refer to the two missions, with 1 being arbitrarily assigned to one mission and 2 the other. The off-axis antenna gain $G(\theta)$ is modeled by the following expressions:

$$G(\theta) = G_{MAX} - (G_{MAX} - 32) \cdot \theta \text{ dB} \quad \text{for } 0^\circ \leqslant \theta < 1^\circ,$$

$$= 32 - 25 \log_{10} (\theta) \text{ dB} \quad \text{for } 1^\circ \leqslant \theta < 48^\circ,$$

$$= -10 \text{ dB} \quad \text{for } 48^\circ \leqslant \theta$$

The next step is to determine the potential modes of interference based on the spectra of the desired signal and the interference. Knowledge of the interference effects on various parts of the system is necessary in this step.

Basically, the interference modes can be determined by finding out which part of the unwanted signal can spectrally interfere with which part of the signal; carrier-to-carrier interference, data-to-carrier interference, etc. An unwanted signal can spectrally interfere with a wanted signal if they are close in frequency. Since the doppler frequency can move the two spectra closer to or farther from each other, it may be necessary

[1] The program calculates TSIR; the protection ratio is usually specified inversely, that is, interference-to-signal.

to consider the doppler effects in determining the modes of interference, particularly when the doppler rate is significant.

The computer program calculates the total interference power. The analyst must consider the amount of interference power that applies to a particular interference mode. The effective interference power of an in-band CW interference is simply equal to the power of the interference. The effective interference power of an interference having a dense spectrum is equal to the power of the interference reduced by a factor equal to the bandwidth conversion factor. The bandwidth conversion factor is defined as the ratio of the bandwidth of the interfered channel to the bandwidth of the interference. The maximum value of the bandwidth conversion factor is unity.

Having determined the interference modes, it is then necessary to calculate the effective interference-to-signal ratio. This ratio can be derived from the total signal-to-interference ratio as follows:

$$EISR = P_W - P_I - TSIR$$

where P_W denotes the power of the wanted signal being interfered with and P_I denotes the effective power of the interference. For example, if the interference mode is carrier-to-data, then the effective interference power is the carrier power of the unwanted signal and the signal power is the power in the data sidebands of the wanted signal.

The effective interference-to-signal ratio can then be compared with the protection ratio. If the protection ratio is exceeded, interference exists. The total number of days for which interference exists is calculated and used as a measure of the amount of potential interference between a given mission pair. This process is repeated for all mission pairs of interest. A detailed step-by-step analysis of the downlink potential interference is provided in Appendix B for a hypothetical system.

B. Uplink Interference Analysis

There are two situations in which an unwanted uplink signal constitutes an interference to a spacecraft. The first is when the effective interference-to-signal ratio exceeds the protection ratio. The second is when the level of an uplink signal as received by a spacecraft exceeds the receiver threshold of that spacecraft for which the uplink signal is not intended. The analysis for the first situation is very similar to the analysis of downlink interference with the exception that the total signal-to-interference ratio is given by a different expression and is the same for both missions. The expression is:

$$TSIR = G_{MAX} - G(\theta)$$

To evaluate the amount of potential interference for the second situation, the power of an unintended uplink as received by a spacecraft located in a direction θ degrees from the main axis of the DSN transmitting antenna is calculated:

$$PU_1 = PT + G(\theta) + 20 \log_{10} \left(\frac{\lambda}{4\pi}\right) - 20 \log_{10} (R_1),$$

and

$$PU_2 = PT + G(\theta) + 20 \log_{10} \left(\frac{\lambda}{4\pi}\right) - 20 \log_{10} (R_2)$$

where PU is the power of the unintended signal, PT is the DSN transmitter output power, $G(\theta)$ is the gain of the transmitting antenna in the direction of θ, λ is the wavelength of the uplink signal, R and the subscripts 1 and 2 are as defined before. The interference power as calculated by the above expressions assumes an isotropic receiving antenna. It is therefore necessary to increase the interference power by an amount equal to the receiving antenna gain before it can be compared to the interference threshold to determine the amount of potential interference.

VIII. Additional Selection Considerations

In the future it may not be possible to select a channel that is completely free of potential interference. This might be the result of many missions sharing a particular band. It will then be necessary to select a channel where the time and severity of interference is such that the most important parts of the affected missions are protected.

Deep-space missions usually have periods of intense activity separated by longer periods of relatively low activity. Interference during the quiet periods will often be acceptable. The potential interference may be strong and virtually destroy successful telecommunications, or it may only cause a slight degradation of performance. The severity of interference therefore affects the choice of channel.

Often, more than one frequency band is used on a mission. The potential interference is usually different for different frequency bands. It may be necessary to examine the potential interference for all frequency bands involved before a selection can be made. Since these frequencies are related to each other by a fixed translation ratio, the resulting channels may be optimal in one band, but not in the other.

IX. Conclusion

The trend of deep-space exploration is toward higher data rates and longer mission life times. Higher data rates tend to

require additional bandwidth. Missions with a life time of 10 to 20 years are currently under study. The combination of high data and long life times will increase the opportunity for mutual interference among deep-space missions. The interference analysis and channel selection process described in this paper is a means to minimize such interference and to assure the efficient use of the deep-space frequency spectrum. While the process does not guarantee that a channel or channels completely free of interference can always be found, it does lead to the selection of the best available channel and, equally important, it provides information regarding the resulting potential interference. This information can help mission planners to minimize the effects of the anticipated interference.

References

1. Sue, M. K., "Block IV Receiver Tracking Loop Performance in the Presence of a CW RFI," *TDA Progress Report 42-60, September and October, 1980*, Jet Propulsion Laboratory, Pasadena, Calif., Dec. 15, 1980.

2. Sue, M. K., "Telemetry Degradation Due to a CW RFI Inducted Carrier Tracking Error for the Block IV Receiving System with Maximum Likelihood Convolutional Decoding," *TDA Progress Report 42-61, November and December, 1980*, Jet Propulsion Laboratory, Pasadena, Calif., Feb. 15, 1981.

Table 1. Allocations for deep-space telecommunications

Earth-to-Space	Space-to-Earth
2110-2120 MHz	2290-2300 MHz
7145-7190 MHz	8400-8450 MHz
16.6-17.1 GHz	12.75-13.25 GHz
34.2-34.7 GHz	31.8-32.3 GHz

(There are some qualifications affecting these allocations; see the Radio Regulations.)

Table 2. Channel center frequencies

Channel	2110-2120 MHz uplink channel center frequency, MHz	2290-2300 MHz downlink channel center frequency, MHz	7145-7190 MHz uplink channel center frequency, MHz	8400-8450 MHz downlink channel center frequency, MHz	Remarks
1		2290.185185	7147.286265		
2		2290.555556	7148.442131		
3		2290.925926	7149.597994		
4		2291.296296	7150.753857		
5	2110.243056	2291.666667	7151.909724	8402.777780	↑
6	2110.584105	2292.037037	7153.065587	8404.135803	
7	2110.925154	2292.407407	7154.221450	8405.493826	
8	2111.266204	2292.777778	7155.377316	8406.851853	Channels
9	2111.607253	2293.148148	7156.533179	8408.209876	
10	2111.948303	2293.518519	7157.689045	8409.567903	5-27
11	2112.289352	2293.888889	7158.844908	8410.925927	
12	2112.630401	2294.259259	7160.000771	8412.283950	are
13	2112.971451	2294.629630	7161.156637	8413.641977	
14	2113.312500	2295.000000	7162.312500	8415.000000	fully
15	2113.653549	2295.370370	7163.468363	8416.358023	
16	2113.994599	2295.740741	7164.624229	8417.716050	coherent
17	2114.335648	2296.111111	7165.780092	8419.074073	
18	2114.676697	2296.481481	7166.935955	8420.432097	in
19	2115.017747	2296.851852	7168.091821	8421.790124	
20	2115.358796	2297.222222	7169.247684	8423.148147	all
21	2115.699846	2297.592593	7170.403550	8424.506174	
22	2116.040895	2297.962963	7171.559413	8425.864197	four
23	2116.381944	2298.333333	7172.715276	8427.222220	
24	2116.722994	2298.703704	7173.871143	8428.580248	bands
25	2117.064043	2299.074074	7175.027006	8429.938271	↓
26	2117.405092	2299.444444	7176.182868	8431.296294	
27	2117.746142	2299.814815	7177.338735	8432.654321	
28	2118.087191		7178.494597	8434.012344	
29	2118.428241		7179.650464	8435.370371	
30	2118.769290		7180.814838	8436.738395	
31	2119.110339		7181.962190	8438.086418	
32	2119.451389		7183.118056	8439.444445	
33	2119.792438		7184.273919	8440.802468	
34			7185.429783	8442.160493	Channels
35			7186.585617	8443.518517	34-39
36			7187.741511	8444.876542	frequencies
37			7188.897375	8446.234566	are estimates
38				8447.592591	only
39				8448.950616	

Table 3. Channel frequency ratios

Band pair	Channel frequency ratio
2110-2120 MHz, 2290-2300 MHz	$\frac{221}{240}$
7145-7190 MHz, 8400-8450 MHz	$\frac{749}{880}$
2290-2300 MHz, 8400-8450 MHz	$\frac{3}{11}$

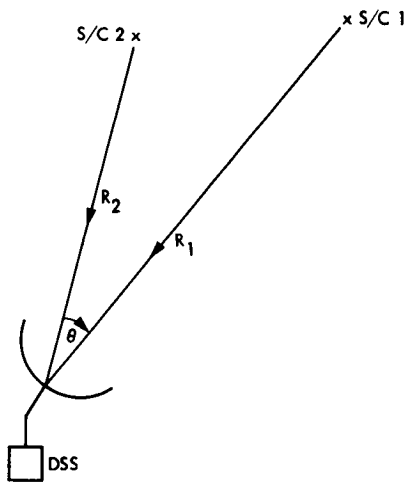


Fig. 1. Downlink interference situation with spacecraft 2 as the interferer

Appendix A

Parameters Needed for Interference Analysis and Channel Selection

I. Information Needed for Detailed Study

Information needed from flight projects in order to perform an interference analysis and subsequent channel selection is detailed in the following paragraphs.

A. Information About the Characteristics of the Telecommunication Systems

(1) Spacecraft antenna gain (for all antennas at their operating frequencies).

(2) Spacecraft transmitter power output (for all possible transmitter modes).

(3) Modulation schemes (including carrier modulation, subcarrier modulation, subcarrier frequencies, number of subcarriers, type of subcarriers, i.e., squarewaves or sinusoids, modulation indices, coding schemes.)

(4) Transmitted data rate or symbol rate.

(5) Number of uplink carrier frequencies, downlink carrier frequencies and their interrelationship.

(6) Frequencies preferred for reasons independent of interference considerations.

(7) Earth receiving station antenna gain.

(8) Spacecraft receiver sensitivity.

B. Expected Modes of Operation as a Function of Mission Phase

These modes are, for example, different combinations of antenna, transmitter power, frequency band, etc.

C. Dates of Important Mission Phases

Dates such as encounter, maneuver, landing, etc., where interference is less tolerable.

D. Mission Trajectory Data

Orbital elements that specify the trajectory of a mission are needed. It is preferred that the orbital elements be provided in the form of classical orbital elements using "Sun Centered, Earth Equator and Equinox of 1950" as a reference frame. The following parameters are needed:

(1) Six classical orbital elements:

Semimajor axis a	Node angle Ω
Eccentricity e	Argument of periapsis ω
Inclination i	Mean anomaly M

(2) Epoch for the above set of elements.

(3) Reference frame used for the above orbital elements. (It is preferred to use "Sun Centered, Earth Equator and Equinox of 1950" as a reference frame.)

(4) Time period for which the above set of elements are applicable.

(5) Launch date.

(6) Arrival date.

(7) Destination.

(8) Expected date for end of mission (EOM).

II. Information Needed for Rough Estimation

Parameters listed in preceding paragraphs constitute a complete set of information needed for frequency selection study. Some of these data may not be available at the early phase of mission design. It is possible, even with an incomplete set of data, to perform a frequency selection study by using typical parameter values or worst-case values, whichever is appropriate. This, however, may place unnecessary restrictions on channel selection. As a minimum, the following parameters are needed for a crude estimation:

A. Characteristics of the Telecommunication Systems

(1) Spacecraft antenna gain.

(2) Spacecraft transmitter power output.

(3) Earth station antenna gain.

(4) Number of uplink and downlink carrier frequencies and their interrelationship.

(5) Frequency preferred for reasons independent of interference consideration.

B. Mission Trajectory Data

In general, all trajectory parameters listed in the previous paragraph are needed. In some special cases where the trajectory of a spacecraft does not consist of any breaks, it is possible to estimate the trajectory by specifying the following parameters:

(1) Launch date.

(2) Arrival date.

(3) Destination.

(4) End of mission date.

Appendix B
Determination of Potential Interference

I. Introduction

This appendix illustrates the necessary steps to determine if potential downlink interference exists. The two spacecraft examined are designated as 1 and 2. Spacecraft 1 is arbitrarily chosen as the wanted spacecraft and the other as the source of interference. Thus, this appendix examines only interference to spacecraft 1 from spacecraft 2. The interference from spacecraft 1 to spacecraft 2 can be obtained by following the same procedures.

II. Assumptions

The hypothetical system used here has the following characteristics:

	Spacecraft 1	Spacecraft 2
Downlink e.i.r.p., dBw	35	30
Antenna gain, dBi	30	25
Telemetry symbol rate, bps	10k	20k
Telemetry subcarrier frequency, kHz	300	300
Telemetry modulation angle, deg	80	40
Number of subcarrier channels	1	1

The doppler frequency and the total signal-to-interference ratio computed by the program are assumed to have the following values:

	Spacecraft 1	Spacecraft 2
Downlink TSIR, dB	0	40
Doppler frequency, kHz	2	1

Both spacecraft are assumed to occupy the same frequency channel.

III. Interference Modes

To determine the interference modes, it is necessary to examine the spectra of the signal and the interference. A sketch of the spectra of these two signals is shown in Fig. B-1. From the sketch, it can be seen that there are two interference modes: carrier-to-carrier and data-to-data. The potential interference for both modes is discussed in the following sections.

IV. Carrier-To-Carrier Interference

To determine if potential interference exists for this interference mode, it is necessary to calculate the power of the wanted signal and the power of the unwanted signal in the carrier channel. The power of the wanted signal in the carrier channel is simply the carrier power of the wanted signal, i.e., the carrier power of spacecraft 1. The unwanted power is the carrier power of spacecraft 2. The wanted and unwanted power can be calculated as follows:

Spacecraft 2 downlink e.i.r.p.	35.0 dBw
Modulation loss (20 log (cos (80°))	-15.2 dB
Wanted power level	19.8 dBw

Spacecraft 2 downlink e.i.r.p.	30.0 dBw
Modulation loss (20 log (cos (40°))	-2.3 dB
Unwanted power level	27.7 dBw

The effective interference-to-signal power is thus equal to:

$$EISR = PW - PI - TSIR$$

$$= 19.8 \text{ dBw} - 27.7 \text{ dBw} - 0 \text{ dB} = -7.9 \text{ dB}$$

which exceeds the protection ratio by about 7 dB. Potential interference to the carrier of spacecraft 1 thus exists.

V. Data-To-Data Interference

To determine if this mode of interference exists it is necessary to calculate the data power of spacecraft 1 and spacecraft 2. These power levels can be calculated as follows:

Spacecraft 1 downlink e.i.r.p.	35.0 dBw
Modulation loss (20 log (sin (80°))	-0.1 dB
Wanted power level	34.9 dBw

Spacecraft 2 downlink e.i.r.p.	30.0 dBw
Modulation loss (20 log (sin (40°))	-3.8 dB
Unwanted power level	26.2 dBw

Since the unwanted signal has a dense spectrum, it is necessary to adjust the unwanted signal power according to the

bandwidth conversion factor. The bandwidth conversion factor, B_F, is defined as follows:

$$B_F = \begin{cases} B_w/B_I & B_w < B_I \\ 1 & B_w \geq B_I \end{cases}$$

where B_w is the bandwidth of the signal and B_I is the bandwidth of the interference. Since the bandwidth of a data channel is proportional to the data rate, the bandwidth conversion factor can thus be calculated from the following equation.

$$B_F = \frac{(Data\ rate)_1}{(Data\ rate)_2} = \frac{10k}{20k} = 1/2 \text{ or } -3 \text{ dB}$$

where the subscripts 1 and 2 refer to spacecraft 1 and spacecraft 2 respectively.

The effective power is thus equal to 26.2 dBw –3.0 dB = 23.2 dBw and the effective interference-to-signal ratio is 23.2 dBw – 34.9 dBw = –11.7 dB. Since this value exceeds the protection ratio, interference to the data channel of spacecraft 1 is said to exist.

VI. Conclusion

The foregoing analysis of a simple hypothetical case serves only as an illustration of the techniques used for interference analysis. Real spacecraft often employ more than one subcarrier channel on the downlink and their spectra are thus more complicated. As a result, interference analysis of an actual system is more involved.

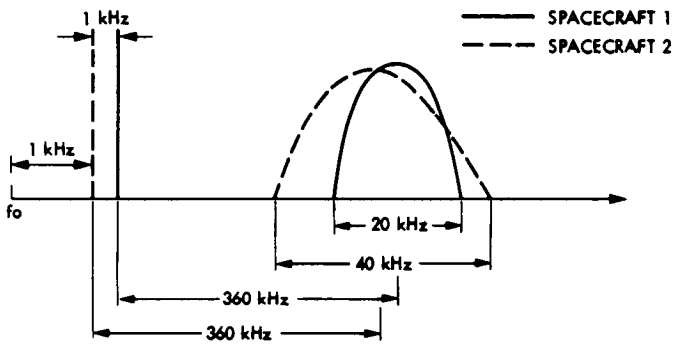


Fig. B-1. Sketch of the spectra of the signal and the interference

New CCIR Papers on Telecommunications for Deep Space Research

N. F. de Groot
Telecommunications Systems Section

Eight JPL papers dealing with telecommunications for deep space research were recently adopted by Study Group 2 of the International Radio Consultative Committee (CCIR). A brief description of the preparation and adoption of the papers was given in TDA Progress Report 42-68, along with the paper on preferred frequency bands in the 20-120 GHz range. In this article we present three more papers.

Study Group 2 of the Comité Consultatif International des Radiocommunications (CCIR) met in Geneva last October to consider papers for inclusion in the 1982 edition of Reports and Recommendations of the CCIR. Study Group 2 deals with radio communication aspects of space research and radio astronomy. The adopted ensemble of papers included eight that were based on input documents prepared by the JPL Spectrum Engineering Group and are:

Doc. 5007 New Recommendation	Protection Criteria and Sharing Considerations Relating to Deep-Space Research.
Doc. 5008 New Recommendation	Preferred Frequencies and Bandwidths for Deep-Space Research.
Doc. 5010 New Report	Frequency Bands in the 20-120 GHz Range that are Preferred for Deep-Space Research.
Doc. 5015 Revision of Report 683	Frequency Bands in the 1-20 GHz Range that are Preferred for Deep-Space Research.
Doc. 5023 Modification of Question 14	Feasibility of Frequency Sharing Within and Among Space Research Systems.
Doc. 5024 Revision of Report 685	Protection Criteria and Sharing Considerations Relating to Deep-Space Research.
Doc. 5025 New Report	Potential Interference Between Deep-Space Telecommunications and the Fixed-Satellite and Broadcasting Satellite Services in Harmonically Related Bands.
Doc. 5026 Modification of Study Program 14B	Frequency Sharing Between Deep-Space and Other Space Research Systems.

In this issue of the TDA Progress Report we present three papers. The first, Doc. 5025, considers the potential for interference between deep-space links and some satellite links in harmonically related bands. Harmonic bands are those with frequencies that are an integer multiple of frequencies that lie within the band in question. Radio transmitters typically produce unwanted harmonic signals, and it is these signals that can cause interference.

In the past, problems of harmonic band interference have been minimal and generally ignored in the context of international frequency management. The spatial and propagation characteristics of the rapidly growing number of satellite links in several bands are, however, creating a significant new potential for interference with space systems in harmonically related bands. The new paper treats a part of the problem as it affects deep space research.

The other two papers, Docs. 5007 and 5008, are recommendations. A CCIR Recommendation (Ref. 1) is "A statement issued in response to a Question or Study Programme which the CCIR considers to be sufficiently complete to serve as a basis for international cooperation." (A Question or a Study Programme is a formal CCIR statement of the scope and content of a technical area to be studied.)

Recommendations are based on CCIR reports and generally represent an official position concerning some aspect of radio frequency use and regulation. Recommendations serve to guide the administrative conferences that lead to the international treaty that is expressed in the Radio Regulations (Ref. 2).

The reports presented in this article are drafts. They may have editorial modifications when formally published.

References

1. *Recommendations and Reports of the CCIR*, Vol. XIII-1, p. 248, International Telecommunication Union, Geneva 1978.

2. *Radio Regulations*, Edition of 1982, International Telecommunication Union, Geneva, 1982.

Documents
CCIR Study Groups
Period 1978-1982

Doc. 2/5025-E
(Ref. 2/210)
19 October 1981
Original : English/
French/
Spanish

Working Group 2-B

DRAFT NEW REPORT*

POTENTIAL INTERFERENCE BETWEEN DEEP-SPACE TELECOMMUNICATIONS AND THE FIXED-SATELLITE AND BROADCASTING-SATELLITE SERVICES IN HARMONICALLY RELATED BANDS
(Question 19/2)

1. Introduction

The possibility of harmful interference resulting from unwanted emissions to and from space stations in the geostationary orbit is of increasing concern to deep-space research. This concern is prompted by the expected increase in utilization of the geostationary satellite orbit in frequency bands adjacent to and harmonically related to the operational frequencies used for deep-space telecommunications.

This Report presents the results of an analysis of potential mutual interference caused by unwanted emissions from the United States Deep-Space Network (DSN) of the space research service, and from the fixed satellite and broadcasting-satellite services. The study also includes an assessment of potential interference to a future deep-space relay station on the geostationary satellite orbit.

2. Interference analysis

Frequency bands harmonically related to the DSN operating frequencies are shown in Table I. The harmonically related bands are direct multiples and submultiples of the deep-space bands. Fixed satellite and broadcasting-satellite service allocations which fall within these bands are indicated in the table.

Of the many potential interference situations resulting from the relationships shown in Table I, certain typical cases have been examined and analyzed parametrically in this Report. Since the levels of unwanted emissions from transmitters in harmonically related bands are not specifically known, this analysis treats the problem parametrically by using the out-of-band emissions of the interferor as an independent variable. In this Report that variable is called "level of suppression". It is expressed with respect to the energy contained in the fundamental frequency. The results of the analysis can then be used to determine suppression levels of unwanted emissions necessary to meet the interference criteria of the services involved.

---

* This Report should be brought to the attention of Study Group 4 and JWG 10-11S.

TABLE I

DSN EARTH STATION TRANSMIT FREQUENCIES	HARMONIC BANDS			
	2nd	3rd	4th	5th
2110 - 2120 MHz	4220-4240 MHz	6330-6360 MHz All regions: Fixed-Satellite Service (E-to-S)	8440-8480 MHz	10.55-10.6 GHz
7145 - 7190 MHz	14.29-14.38 GHz All regions: Fixed-Satellite Service (E-to-S)	21.435-21.57 GHz	28.58-28.76 GHz All regions: Fixed-Satellite Service (E-to-S)	35.725-35.95 GHz
16.6 - 17.1 GHz	33.2-34.2 GHz	49.8-51.3 GHz All regions: Fixed-Satellite Service (E-to-S)	66.4-68.4 GHz	83.0-85.5 GHz All regions: Fixed-Satellite (S-to-E) Broadcasting-Satellite Service
DSN EARTH STATION RECEIVE FREQUENCIES	SUBHARMONIC BANDS			
	1/2	1/3	1/4	1/5
2290 - 2300 MHz	1145-1150 MHz	763-767 MHz	573-575 MHz	458-460 MHz
8400 - 8450 MHz	4200-4225 MHz	2800-2817 MHz	2100-2113 MHz	1680-1690 MHz
12.75 - 13.25 GHz	6375-6625 MHz All regions: Fixed-Satellite Service (E-to-S)	4250-4417 MHz	3188-3313 MHz	2550-2650 MHz All regions: Broadcasting-Satellite Service. In region 2: Fixed-Satellite Service (S-to-E)

Pointing statistics for a given DSN Earth station have been obtained from a computer study of the trajectories of 40 realized and potential deep-space missions. Figure 1 presents pointing angle toward the geostationary arc as a function of time for one DSN Earth station. It should be noted that some deep-space missions do not require antenna pointing to within 5-10 degrees of the geostationary arc for many months or even years of mission duration.

3. Harmonic Interference

3.1 Interference to a DSN Earth Station From a Fixed-Satellite Earth Station

The 2nd harmonic of an Earth station in the Fixed Satellite Service transmitting in the frequency range 6.375-6.625 GHz has the potential for interference to a DSN Earth station receiving 12.75-13.25 GHz.

The coordination distance between a transmitting Earth station in the Fixed Satellite Service and a receiving DSN Earth station was computed for two modes of propagation according to the procedure described in Appendix 28 of the Radio Regulations.

Those modes are:

- Clear air propagation mode 'a'.
- Rain scatter propagation mode 'c', in Zone 4.

The following assumptions were made for both modes of propagation:

- Permissible level of interference to a DSN Earth station is -220 dB (W/Hz).
- Elevation angle of transmitting and receiving antennae is taken to be 5° above the horizon.

Assumed Fixed Satellite Earth station e.i.r.p.'s* at 5° off-main beam axis are:

Case 1.	21.0	dB(W/4 KHz)	for satellite network with a large Earth station antenna
Case 2.	14.5	dB(W/4 KHz)	for FM-TV or SCPC global systems
Case 3.	8.5	dB(W/4 KHz)	for FDM-FM systems.

For various suppression levels of spurious emissions from a fixed satellite Earth station the coordination distance for propagation modes 'a' and 'c' is shown graphically in Figures 2 and 3.

*See Report 453-2.

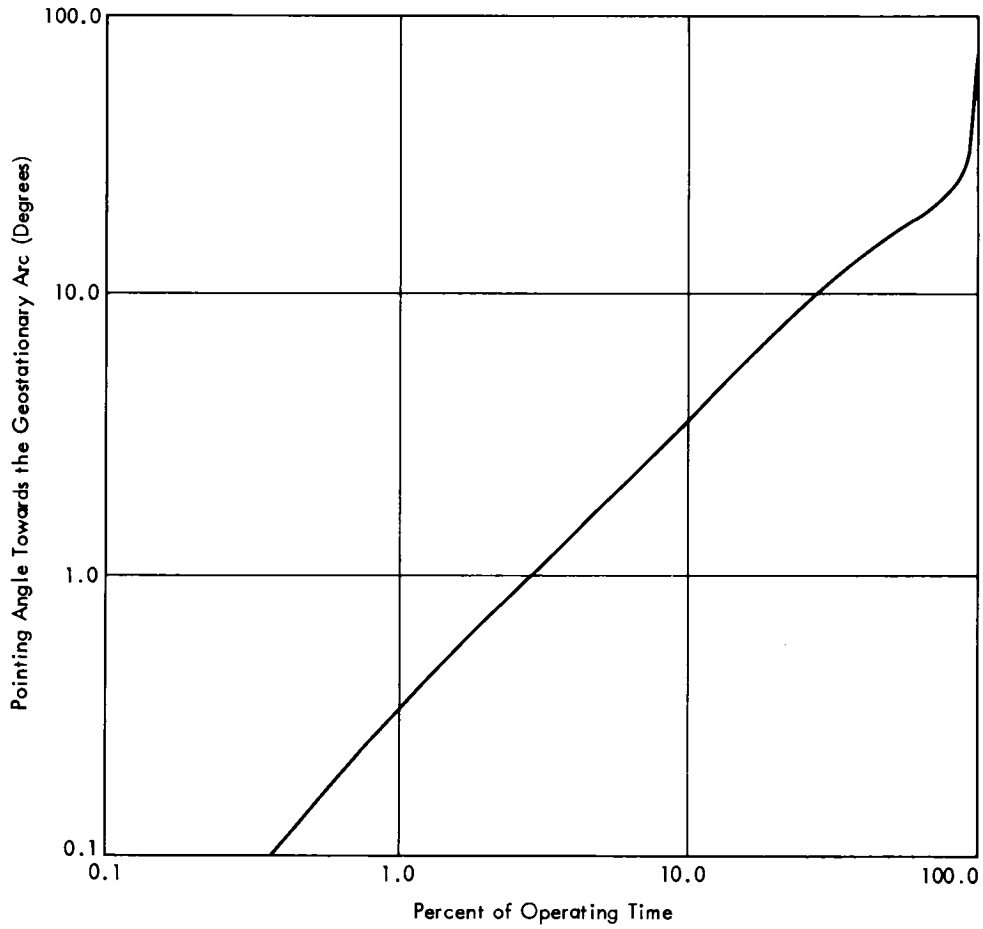
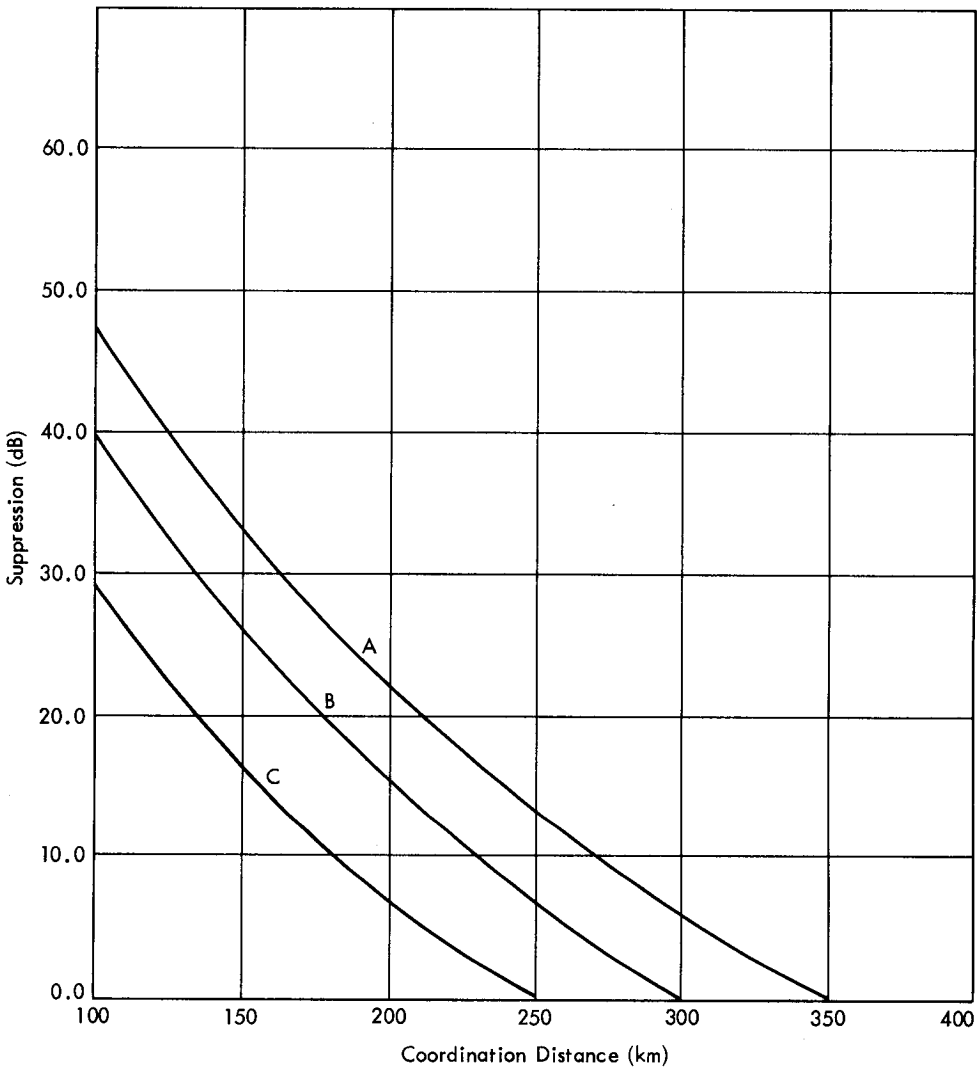


FIGURE 1

Pointing statistics of a DSN earth station

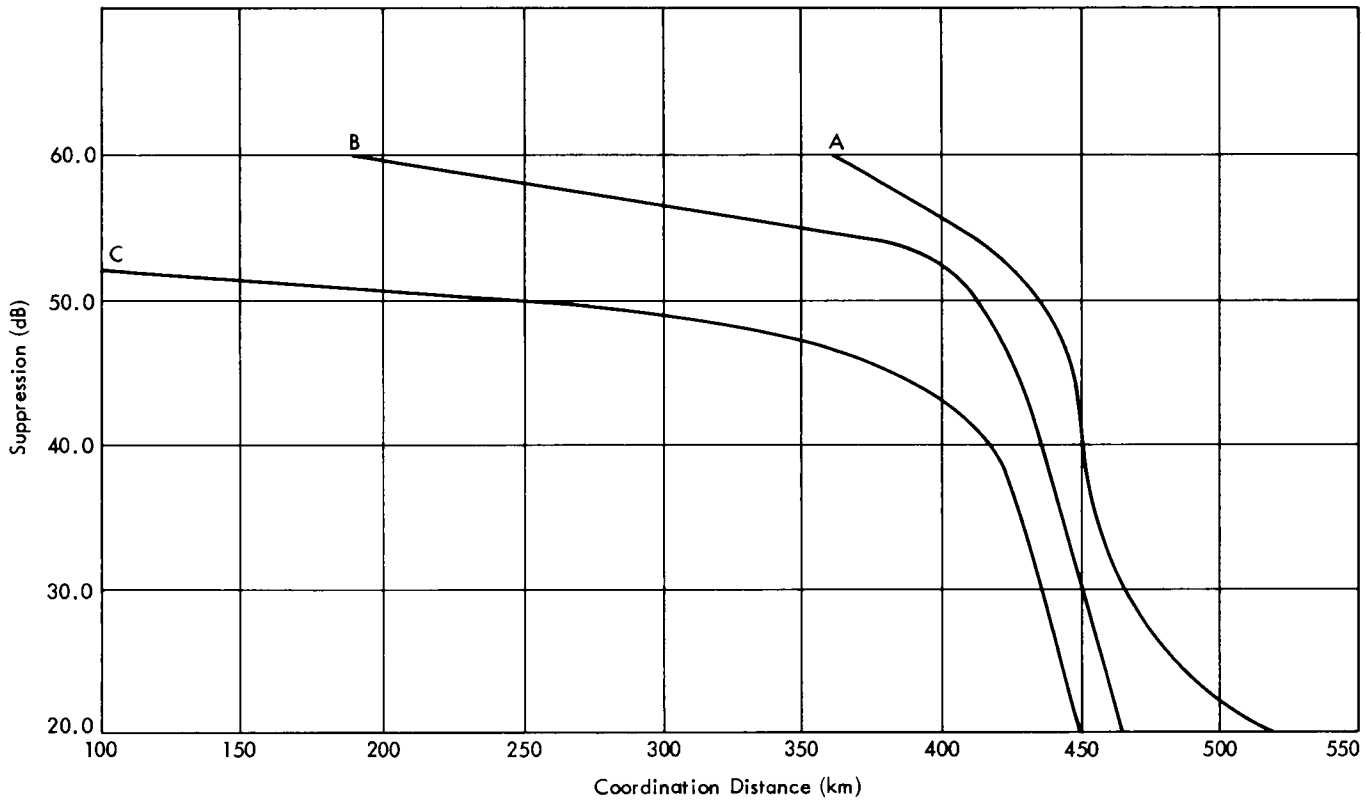


LEGEND

A = Case 1
B = Case 2
C = Case 3

FIGURE 2

Coordination distance between a transmitting FSS earth station and a receiving DSN earth station for clear air "Mode A"



LEGEND

A = Case 1
B = Case 2
C = Case 3

FIGURE 3

Coordination distance between a transmitting FSS earth station and a receiving DSN earth station for rain scatter "Mode C"

3.2 Interference to a Satellite of the Fixed Satellite Service from a DSN Earth Station

Potential interference exists to a satellite of the Fixed Satellite Service receiving in the 6330-6360 MHz range from the 3rd harmonic of a DSN Earth station transmitting in the 2110-2120 MHz band.

The assumed DSN Earth Station transmitting characteristics are:

- Frequency 2.1 GHz
- R.F. Power 50 dBW
- Antenna Gain 62 dBi
- R.F. Bandwidth:

  Ranging 10 MHz

  Telemetry 3 MHz

  Command 0.3 MHz

It is also assumed that the DSN Earth station is transmitting 100% of the time.

Typical receiver noise temperature and antenna gain of a Fixed Satellite are 3000°K and 23.0 dBi, respectively (Report 207-4). The interference criterion is taken to be 4%* of the noise power of the satellite receiver. It should be noted that this 4% criterion is used only as an indicator, to determine those situations in which a more detailed analysis should be performed.

Figure 4 shows the percent of time that a DSN Earth station will cause a 4% increase in Fixed Satellite noise power for various levels of DSN spurious energy suppression.

3.3 Interference to a DSN Earth Station from a Fixed-Satellite

The 5th harmonic of a Broadcasting-Satellite or Fixed Satellite transmitting in the 2550-2650 MHz frequency range has the potential for interference to a DSN Earth station receiving at 12.75-13.25 GHz.

According to Article N26 of the Radio Regulations, the maximum permitted power flux density limit on the surface of the Earth is -137 dB (W/m^2) in any 4 KHz band for a Broadcasting-Satellite operating in the 2-3 GHz band.

---

*Modified Appendix 29 of the Radio Regulations as per WARC 1979

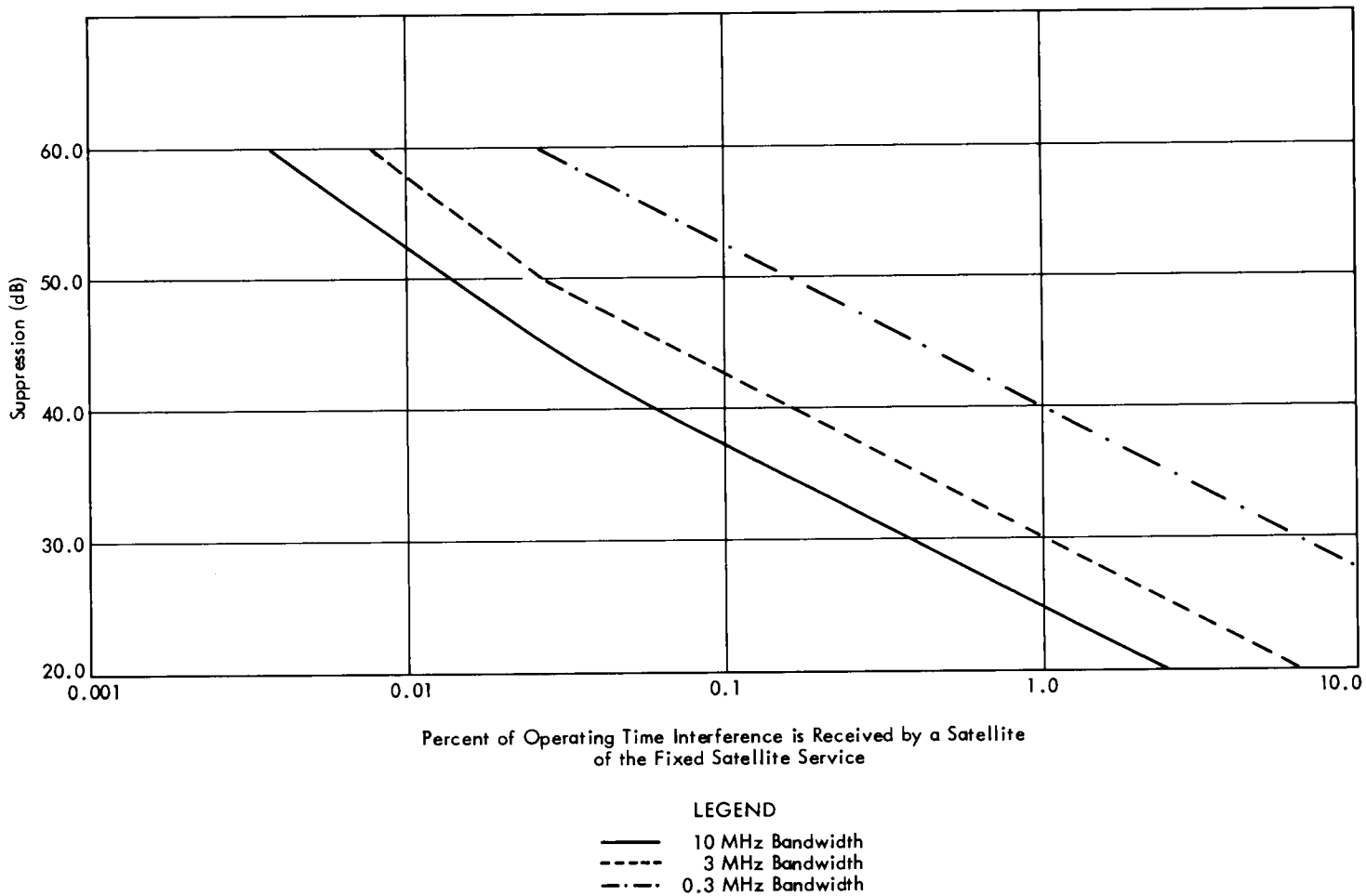


FIGURE 4

Interference from the third harmonic of a DSN earth station to a satellite of the fixed-satellite service

Further assumptions made in this part of the analysis are:

- Broadcasting-Satellite transmitting 100% of the time.
- DSN Interference criterion = -220 dB (W/Hz).

The percentage of time that the interference criterion is met for various levels of suppression of the interfering harmonic signal is shown in Figure 5. This is based upon the pointing statistics of the DSN as described in Figure 1.

Significant interference to a DSN Earth Station from more than one satellite is not considered likely since Broadcasting-Satellite systems using spot beams will probably not illuminate the same service areas on the Earth at the same frequency, due to the problem of mutual interference.

3.4 Interference to an Earth Station of the Fixed Satellite or Broadcasting-Satellite Service From a DSN Earth Station

The potential interference to an Earth station of the Fixed Satellite or Broadcasting-Satellite Service from the 5th harmonic of a DSN Earth station transmitting in the band 16.6-17.1 GHz is not considered in this report due to lack of information on 80 GHz space systems.

3.5 Interference to DSN Geostationary Relay Satellite

A satellite in the geostationary satellite orbit may be used in the future to relay signals from deep-space research spacecraft to Earth. Although this relay spacecraft may employ other frequencies than those currently used for deep-space research (especially in the geostationary satellite-to-earth links) this portion of the analysis is directed at a brief assessment of interference potential at harmonics of the same frequencies as those analyzed in the previous sections of this paper.

3.5.1 Interference to a DSN Relay Satellite from Fixed Satellite Earth Station Transmissions

The 2nd harmonic of an Earth station in the Fixed Satellite Service transmitting in the frequency range 6.375-6.625 GHz has the potential for interference to a DSN relay satellite receiving from deep-space at 12.75-13.25 GHz.

For this analysis it is assumed that the DSN relay satellite has a 45 m receiving antenna and a gain of 0 dBi toward the transmitting FSS Earth station. Additionally, the same values of harmful interference (-220 dB (W/Hz) and FSS Earth station transmitting characteristics, as assumed in section 3.1, are used. For various levels of spurious harmonics, the required geocentric separations of the receiving fixed-satellite and the DSN relay satellite have been calculated. Even on a worst case basis of 0 dB suppression the required spacing is small, as demonstrated below:

FSS Earth station transmit power:		-36 dB (W/Hz)
Free space loss	:	-207 dB
Received power density	:	-243 dB (W/Hz)

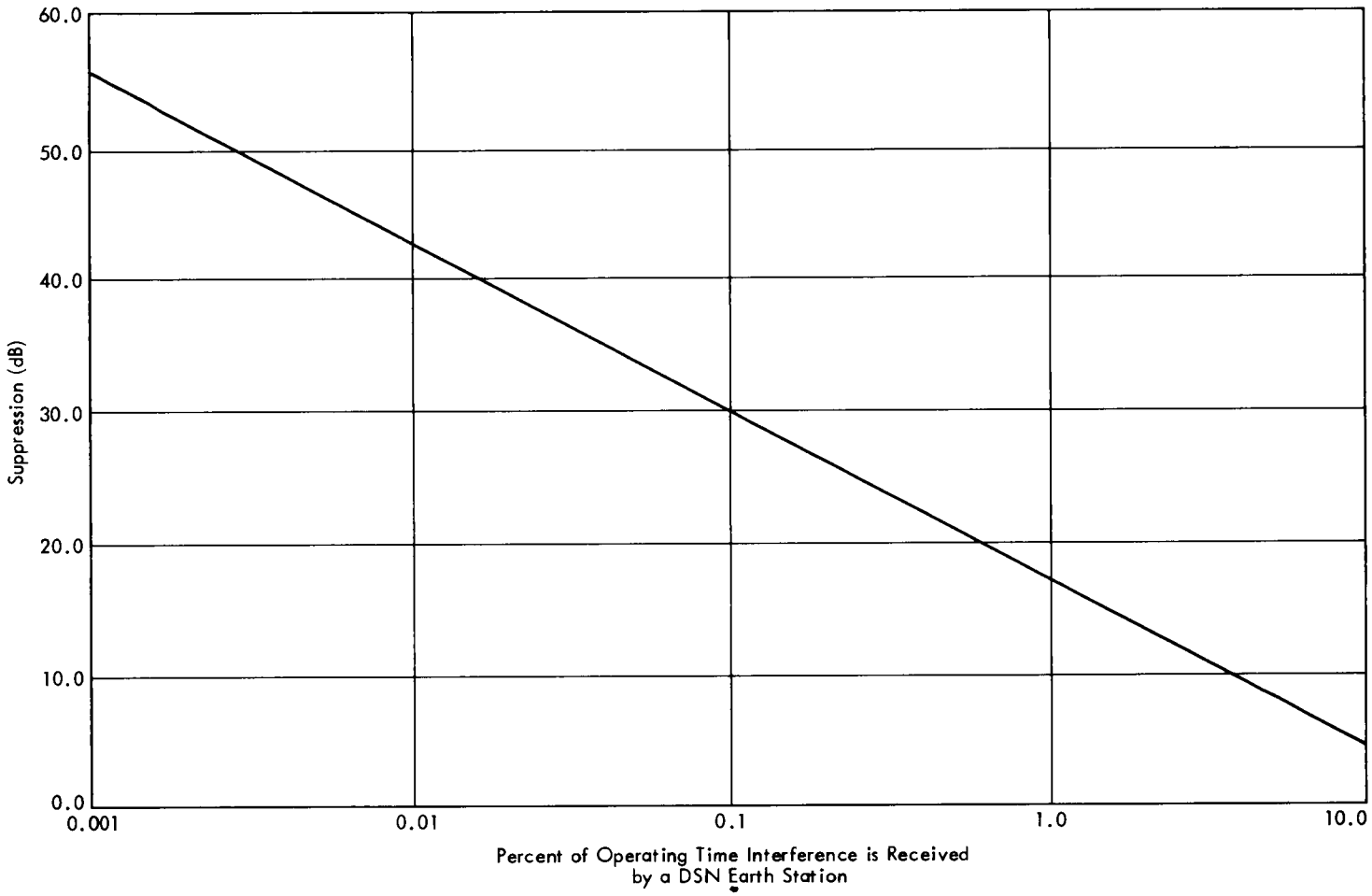


FIGURE 5

Interference to a DSN earth station from the fifth harmonic of a broadcasting satellite with -137 dB(W/m^2) power flux density at the surface of the Earth

Using -220 dB (W/Hz) as the interference criterion would mean that the gain of the FSS Earth station toward the DSN relay satellite could be as high as 23 dB. This translates via the reference antenna pattern of 32-25 log (θ) to a spacing on the order of 2.3°.

At higher levels of harmonic suppression, the required separation is correspondingly less.

3.5.2 Interference to DSN Relay Satellite From FSS Satellite Transmissions

The 5th harmonic of a Broadcasting or Fixed Satellite transmitting in the 2550-2650 MHz frequency range has the potential for interference to a DSN relay satellite receiving from deep-space in the 12.75-13.25 GHz band.

In this analysis it is assumed that:

- The gain of the fixed satellite toward the DSN relay satellite is 0 dBi.
- The maximum value of PFD allowed on the Earth's surface is -137 dB ($(W/cm^2)(4\ kHz)$).
- The DSN relay satellite interference criterion is -220 dB (W/Hz).

Figure 6 presents the angle that the DSN relay satellite must point away from a Fixed or Broadcasting Satellite as a function of the geocentric spacing between the two. These curves are shown parametrically for various levels of energy suppression.

Although pointing statistics of a DSN relay satellite are not currently known, it can be assumed that a DSN relay satellite would point toward the geostationary orbit (for a given mission) for the same or less length of time as an Earth based DSN station.

If a DSN relay satellite is receiving only when the shortest distance between the line of sight to a DSN probe and the surface of the earth is greater than or equal to 200 km, the assumed pointing statistics of the relay satellite can be used, together with the geostationary arc length visible to the relay satellite, to give some indication of the percentage of time a single satellite may cause interference to the relay satellite.

When this is done for a single FSS satellite spaced 1° away from the relay satellite, it is found that for energy suppression levels of 50, 40, 30 and 20 dB, the percentages of time interference is received by a DSN relay satellite are 0.003%, 0.017%, 0.104% and 0.35% respectively. For the same levels of suppression, the percent of time a relay satellite receives interference from a single satellite located at greater distances from the relay satellite is found to be correspondingly less.

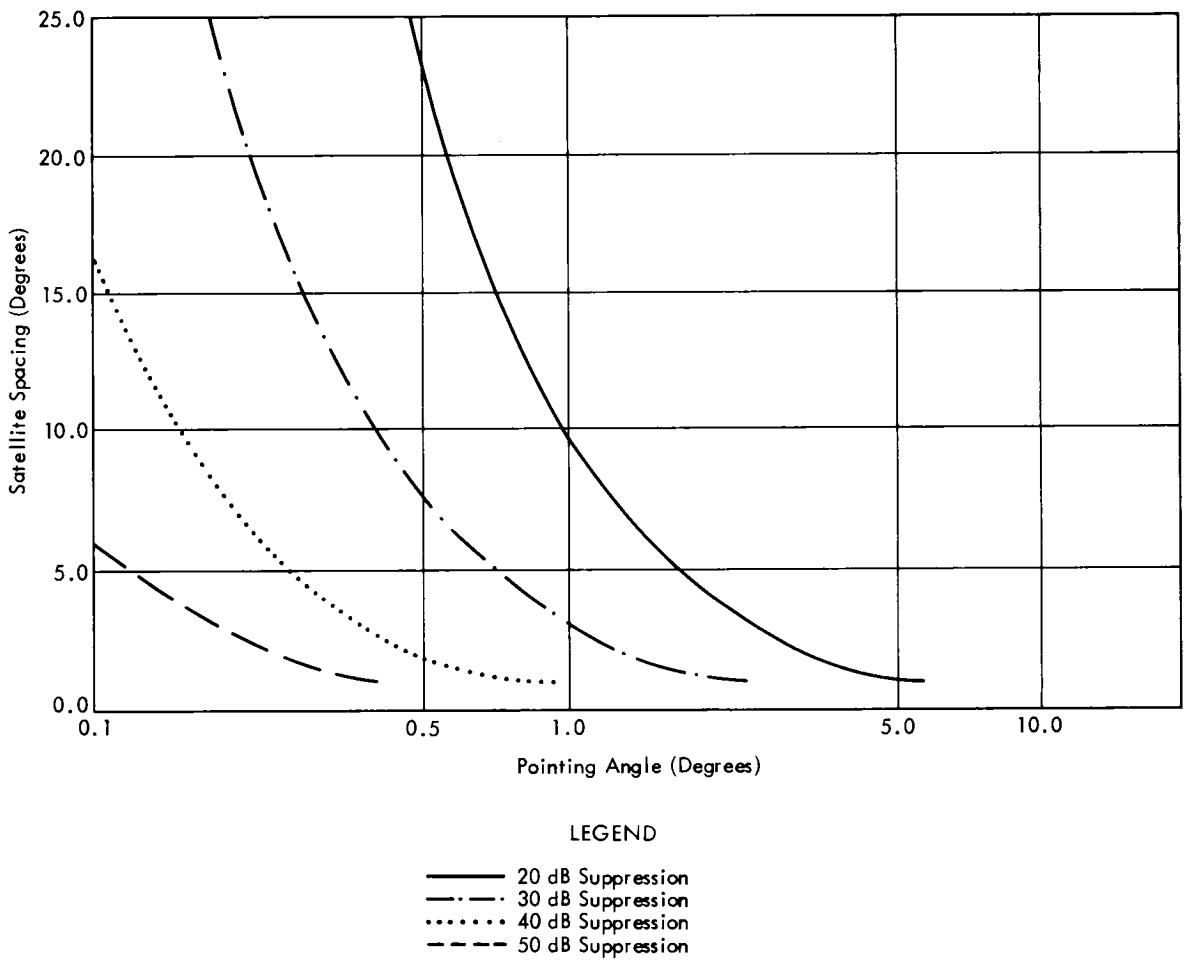


FIGURE 6

Interference from the second harmonic of a fixed or broadcasting satellite to a DSN relay satellite

4. Adjacent Band Considerations

In addition to the potential for interference from unwanted emissions in harmonically related bands, there is the possibility of interference from unwanted emissions of service in adjacent bands. Specific analysis of this possibility has not yet been accomplished with respect to the services considered in this report.

5. Conclusions

Because of the high gain antennas, high power transmitters, and extremely sensitive receivers employed by DSN earth stations, deep-space research and the other satellite services in harmonically related bands may be subjected to interference from unwanted emissions unless measures are taken to suppress the energy of these emissions to acceptable levels.

Precise levels of unwanted emissions that are acceptable for all services cannot be determined in this Report as this requires a detailed interference analysis of the services involved and their individual characteristics. However, from the parametric analysis of "worst case" situations, it appears that significant interference would be avoided if harmonic emissions were suppressed by at least 50 dB.

In most cases the severity of interference between earth stations can be reduced by proper separation and site shielding. In the case of interference to and from space stations, additional filtering may be required to suppress the power in unwanted emissions to an acceptable level. In the case of harmonic band radiation, extra filtering is easily achieved at frequencies far removed from the fundamental.

Analysis of interference from unwanted emissions from services in adjacent bands remains to be done.

Documents
CCIR Study Groups
Period 1978-1982

Doc. 2/5007-E
(Ref. 2/191)
19 October 1981
Original : English/
French/
Spanish

Working Group 2-B (ED)

DRAFT RECOMMENDATION AE/2*

PROTECTION CRITERIA AND SHARING CONSIDERATIONS RELATING TO DEEP-SPACE RESEARCH

(Question 1-1/2)

The C.C.I.R.,

CONSIDERING

(a) that manned deep-space research has unique requirements for extreme reliability of telecommunications so as to ensure safety of life;

(b) that both manned and unmanned deep-space research have unique requirements for extreme reliability of telecommunications so as to ensure successful reception of valuable scientific data collected at particular critical times, and that repeat transmission of these data is often not possible;

(c) that the extreme sensitivity of deep-space earth stations results in unusually low levels of permissible interference;

(d) that some terrestrial and earth stations in other services have sufficient e.i.r.p. to cause interference to stations in deep-space;

(e) that sharing studies and protection criteria have been presented in Report 685 (MOD F) for deep-space research earth stations and for stations in deep-space;

(f) that protection criteria for relay stations in earth orbit, used for deep-space research, have not yet been determined and are not considered in Report 685 (MOD F),

RECOMMENDS

1. that protection criteria for deep-space research earth stations be established as follows: -222 dB (W/Hz) in the 2 GHz region, -220 dB (W/Hz) in the 8 GHz region, -220 dB (W/Hz) in the 13 GHz region, and -216 dB (W/Hz) in the 32 GHz region;

*This Recommendation is brought to the attention of Study Groups 4, 8, 9 and JWG 10-11S.

2. that protection criteria for stations in deep-space be established as follows: -191 dB (W/20 Hz) in the 2 GHz region, -189 dB (W/20 Hz) in the 7 GHz region, -186 dB (W/20 Hz) in the 17 GHz region, and -184 dB (W/20 Hz) in the 34 GHz region;

3. that calculation of interference that may result from atmospheric and precipitation effects be based on weather statistics for 0.001% of the time;

4. that with coordination, deep-space research can share Earth-to-space bands with stations in other services except:

   - receiving aeronautical mobile stations, receiving satellite stations, and microwave sensor satellites, when any of these may come within line-of-sight, and

   - receiving mobile stations that come within the separation distance required for interference protection, and

   - transmitting terrestrial stations having an average e.i.r.p. exceeding 81 dBW in the 2 GHz region and 84 dBW in the 7 GHz region.

5. that with coordination, deep-space research can share space-to-Earth bands with stations in other services except:

   - the radio astronomy service, and

   - transmitting aeronautical mobile stations, transmitting satellite stations, and active microwave sensor satellites, when any of these may come within line-of-sight, and

   - transmitting mobile stations that come within the separation distance required for interference protection.

Documents
CCIR Study Groups
Period 1978-1982

Doc. 2/5008-E
(Ref. 2/196)
19 October 1981
Original : English/
French/
Spanish

Working Group 2-B (ED)

DRAFT RECOMMENDATION AF/2

PREFERRED FREQUENCIES AND BANDWIDTHS FOR DEEP-SPACE RESEARCH

(Question 22/2, Study Programme 22A/2)

The C.C.I.R.,

CONSIDERING

(a) that frequencies most suited for telecommunications between the Earth and Spacecraft in deep-space are determined partly by atmospheric and interplanetary propagation phenomena;

(b) that technology also influences the selection of preferred frequencies;

(c) that requirements for telecommunication reliability must be satisfied during periods of adverse atmospheric effects;

(d) that the same frequency may be used for spacecraft at different celestial coordinates, but that different spacecraft in the vicinity of the same coordinates and within the beamwidth of an earth station antenna will usually require different frequencies;

(e) that it is practical and desirable to effect telemetering and tracking functions on the same space-to-Earth link, and telecommand and tracking functions on the same Earth-to-space link;

(f) that to effect precision tracking, a pair of coherently-related Earth-to-space and space-to-Earth frequencies is desirable;

(g) that for more accurate calibration of the effects of charged particles on the velocity of propagation, simultaneous use of links with coherent frequencies in two or more widely separated bands is required;

(h) that voice and video links associated with manned spacecraft in deep-space could use frequency bands allocated for telemetering, telecommand and tracking functions;

(i) that Report 683 (MOD F) considers the selection of preferred frequencies for deep-space research in the 1 to 20 GHz range;

(j) that Report ... (Doc. 2/192) considers the selection of preferred frequencies for deep-space research in the 20 to 120 GHz range.

RECOMMENDS

1. that frequency bands for deep-space research in the 1 to 20 GHz range be located, with due regard to the feasibility of sharing, in the preferred frequency regions listed in Table III of Report 683 (MOD F);

2. that frequency bands for deep-space research in the 20 to 120 GHz range be located, where sharing is feasible, in the preferred frequency regions listed in Table I of Report ... (Doc. 2/192);

3. that allocation widths at the preferred frequencies be in harmony with the bandwidth requirements discussed in Report 536-1 (MOD I) in order to provide for present and future deep-space telecommunications in a multi-spacecraft, multi-mission environment.

Analysis, Prediction and Control of Radio Frequency Interference with Respect to the DSN

N. F. de Groot
Telecommunications Systems Section

The objective of this report is to inform the reader about four aspects of analysis, prediction, and control of RFI with respect to the DSN. The four aspects are: susceptibility modelling, prediction of RFI from satellites, operational RFI control and international regulations. Special emphasis is given to the existing satellite interference prediction program called DSIP2. This report is intended to be a tutorial for those not familiar with all of the covered topics and also to provide a summary status evaluation from the author's point of view. The report is also expected to serve as a catalyst in guiding future work.

Principal conclusions of the report are:

(1) Analytic modelling and laboratory verification of DSN receiver susceptibility to RFI should continue, this being useful for RFI analysis and prediction, and the setting of regulatory protection criteria.

(2) The performance of the satellite interference prediction program needs to be tested and assessed to determine the need for and kind of refinement that would be effective.

(3) Existing operational management organizations and methods are effective in controlling Goldstone RFI and should continue. The Electromagnetic Compatibility Analysis Center has been independently modelling DSN susceptibility to RFI. Continuation of this effort is not recommended.

(4) Continued participation in organizations that influence and control international use of radio frequency bands is deemed essential with respect to the setting of protection criteria for deep-space telecommunications.

I. Background

In the mid 1970's the DSN experienced an increase in radio frequency interference (RFI). Some of the interference caused degradation or outage of data from deep-space missions. The trend of RFI events suggested the need for:

(1) A better understanding of the susceptibility of DSN earth stations to RFI.

(2) Development of a capability to predict RFI occurrences from known signal sources.

(3) The creation of operational and organizational methods of controlling the interference environment.

(4) The adoption of satisfactory protection criteria in the international Radio Regulations.

At the beginning of 1982 we have the following situation regarding the four needs:

(1) The RFI susceptibility of the Block IV receiver has been modelled for most of the CW cases. Models for the pulse interference (radar) cases do not yet exist.

(2) There is an operating computer program (DSIP2) to predict satellite interference. The program appears to provide the needed DSN protection in the sense that mission data is not seriously compromised by unexpected RFI from known satellite sources. The quantitative accuracy of predictions has not been determined.

(3) An effective Mojave Coordinating Group is managing the Goldstone environment. A similar arrangement does not exist at the overseas sites. A capability to predict RFI incidental to Fort Irwin operations near Goldstone is under development by an agency of the Department of Defense.

(4) The international Radio Regulations specify permissible levels of interference to deep-space downlinks. These levels were developed by JPL. They are not known to be inadequate although they are based on old analyses that do not consider current methods of coded telemetry.

In the ensuing sections of this report the reader will find discussion of each of these four needs and the status of efforts to satisfy them.

II. Susceptibility of DSN Earth Stations to RFI

In the development of DSN earth station receivers, the traditional emphasis has been on maximum sensitivity in an environment free of RFI. Indeed, the station sites were chosen to provide a quiet environment. With the increase in interference episodes, it became necessary to understand how the receiver performance is degraded by the presence of unwanted signals. In this section we consider JPL modelling of RFI susceptibility. Similar work by the Electromagnetic Compatibility Analysis Center (ECAC), a DOD agency, is being done in connection with Army operations at Fort Irwin near Goldstone. This work will be discussed in Section IV.

A. JPL Susceptibility Models

A susceptibility model is a mathematical expression that describes the response of a receiver to an interfering signal. For example, a particular interfering signal may result in an increased bit error rate for the desired telemetry signal from deep space. The model allows calculation of the change in bit error rate, provided the characteristics of the interference are known.

Because DSN receivers are complex, theoretical analysis of their behavior usually involves mathematical simplifications or idealizations. Independent verification of the analytic expressions is therefore necessary. This may be accomplished by laboratory or field testing under controlled conditions. The verified susceptibility models may be used:

(1) To analyze potential and experienced RFI.

(2) To enable the creation of operational RFI prediction capability.

(3) To provide the basis for interference protection criteria to the international Radio Regulations.

B. Status of JPL Modelling

The DSN Block IV receivers have been chosen for analysis primarily because they are associated with the 64-m antennas that are used when maximum sensitivity is required. Verification of the models has been accomplished by means of testing in the Telecommunications Development Laboratory. The tests involve measurement of RFI effects under controlled conditions of desired and interfering signals. Results of these tests are compared with effects predicted by the theoretical analysis. The analytic modelling has so far been accurate enough so that numerical adjustment (curve fitting) has not been necessary and is not done.

Modelling efforts to date have concentrated on CW interference. References 1-4 discuss this work. Current status of modelling is described in detail in Ref. 5. The CW susceptibility models provide a base for future study and modelling of the effects of pulse and noiselike interference. The relationship between the JPL models and the existing computer program for predicting RFI from satellites will be discussed in Section III.

An important motivation for the study of RFI susceptibility is the determination of acceptable levels of interference. These levels are specified in the Radio Regulations, as discussed further in Section V. The specification of levels of permissible interference implies an acceptable degradation with respect to the performance in the absence of interference. Permissible interference has been taken to be the amount that results in 1 dB receiver gain compression due to saturation, 15 deg phase jitter in the carrier tracking loop, and/or 1 dB degradation of telemetry performance. These values are somewhat arbitrary. The gain compression limitation is particularly

tenuous in that the resulting effects on data quality have not been quantified. In addition, signals strong enough to cause gain compression have also been observed to generate spurious signals, but this effect has not yet been modelled.

III. Satellite RFI Prediction

An important class of interference is that coming from earth-orbiting satellites. Some of these operate in bands used for deep-space communications downlinks, or in adjacent bands. These satellites come within view of DSN earth stations in a regular, predictable way. It is therefore possible to determine in advance the times at which RFI may occur. To make RFI predictions, it is necessary to know the satellite orbits, deep-space mission trajectories, characteristics of the satellite signals, and the susceptibility of the DSN receivers.

The recurring nature of satellite interference suggests that an automated analysis program could be used to predict RFI. A computer program called Deep-Space Interference Prediction Program (DSIP2) has been developed for this purpose. The prediction program is in regular use to warn of potential RFI episodes. Corresponding action by the DSN depends on the consequences of the RFI that could occur. In some cases, satellite operators have been successfully requested to cease transmission during critical mission times. A general description of DSIP2 may be found in Ref. 6. Reference 7 is the DSIP2 Users Guide.

A. Development of DSIP2

The Deep-Space Interference Program, DSIP2, predicts the time when satellite RFI will occur, and it predicts the effect of the RFI: degradation of telemetry performance, telemetry drop lock, or receiver (carrier tracking) drop lock. A necessary condition for interference to occur is that a satellite is in view of the station that is tracking a deep-space spacecraft.

The periods of potential interference are determined by DSIP2 from satellite and spacecraft time and position data computed by other separate programs. Tests for interference are then made for these periods. Six tests are made:

(1) Receiver interference: received power above a fixed threshold and frequency separation less than a fixed value.

(2) Receiver drop-lock, Type 1.

(3) Receiver drop-lock, Type 2.

(4) Telemetry drop-lock, Type 1.

(5) Telemetry drop-lock, Type 2.

(6) Telemetry SNR degradation.

A description of these tests may be found in Appendix A. Appendix B lists the equations that are in the RFI subroutine of program DSIP2. The mathematical models embodied in the several tests are based on the effects of CW interference. DSIP2 presumes that the effects of more complex signals may be related to the CW models by characterizing the interference in terms of a family of spectral lines, each of which has a particular frequency and amplitude. Each line is considered to be a CW signal, and DSIP2 calculates the RFI signal effect of each line. For RFI predictions it is therefore necessary to know the spectral nature of the interfering satellite signal. Data on these parameters is obtained from the operator of the satellite, or by measurement of the received signal. Where military classification is an issue, other special arrangements are made for RFI prediction.

Because verified analytic models were not available at the time of DSIP development, receiver susceptibility was characterized by means of empirical testing of the effect of CW interference. These tests and the associated expression of results in the form of mathematical equations were done circa 1977 and are described in Refs. 8 and 9. In essence, some general theoretical analysis was developed and expressed in terms of equations and curves. Experimental data was then compared to the curves. Where necessary, the equations were modified to more nearly match the test data by means of a curve fitting process. This curve fitting results in the sometimes strange numerical constants included in the equations.

The theoretical basis for the prediction equations in DSIP2 is particularly unsatisfying from an analyst's point of view. In connection with the telemetry tests, there is an allusion in Ref. 4 to an analysis of the phase lock loop (PLL) jump phenomenon (Ref. 10). The phenomenon refers to a PLL dropping lock to a desired signal and locking to a stronger interfering signal. Extensive examination of the applicability of the Ref. 10 analysis to DSN receivers reveals considerable theoretical difficulty.

The origins of the DSIP2 models for receiver drop lock are obscure and apparently unpublished. (The term "receiver drop-lock" refers to the phase lock loop that tracks the received carrier signal). The general form of the equations is like those used for telemetry, but the corresponding test data has not been documented, if it exists.

Although a case can be made that DSIP2 can't work very well because of its simplistic modelling, it has not been demonstrated that more elaborate modelling is required for DSN protection. What is specifically missing at the present time is a useful determination of the detail strengths and weaknesses of DSIP2 performance. Also missing is a specification of required performance of the prediction capability: time accuracy, false

alarm rate, and the accuracy of predicted effects such as drop-lock or SNR degradation.

B. DSIP2 Evaluation

In September 1977 there was a review of DSIP program status. Statistics were presented that described the ability of DSIP2 to predict interference to the Viking Orbiter 2 from the ESA GEOS satellite. When no RFI was predicted, no RFI was observed. For 31 predictions of RFI there were 13 episodes observed. The predictions were thus quite conservative for this particular satellite-deep-space mission pair.

More recently (Ref. 6) it was reported that nine instances of Voyager telemetry drop-lock caused by a Cosmos satellite were correctly predicted by DSIP2. Out of 15 instances of Pioneer 10 telemetry or receiver drop-lock, 11 were correctly predicted by DSIP2. The incorrect predictions for the other 4 instances were said to be a result of incomplete signal characterization of the interfering Cosmos.

These statistics suggest that DSIP2 is conservative: there are many false alarms and (apparently) no missing alarms. It is also clear that the quality of signal characterization is a determinant of prediction accuracy. Although the satellite interference prediction program has been used for several years, only limited anecdotal evidence of performance exists. For meaningful evaluation that can lead to validation, improvement, or simplification of prediction, it will be necessary to test two aspects of DSIP2: (1) the ability to correctly determine the amplitude and frequency of interfering spectral lines, and (2) the ability to correctly determine the effect of the received interference.

The existing RFI reporting system does not meet the needs of DSIP2 validation. The reports refer primarily to observed RFI events. There is no explicit reference to predictions, and there is no characterization and reporting of received signals that may be present when predicted RFI is or is not observed. A well planned test of DSIP2 accuracy is needed. The requirements and methods of such a test are being examined.

C. JPL RFI Models and DSIP2

Modelling of RFI susceptibility and the development of DSIP2 began at approximately the same time. The modelling effort was based on the premise of developing an analytic expression that would directly relate input signals to RFI effects. The DSIP2 approach uses a spectral line characterization of input signals and depends upon tests of each line against CW susceptibility expressions. Because of these different approaches, the verified analytic models are not always directly and easily incorporated into DSIP2. It has been judged that such incorporation should wait for a detailed evaluation of DSIP2 performance.

IV. Operational and Organizational Methods

The rising trend of RFI incidents at Goldstone was related to rising military activity in the surrounding area. Goldstone is situated within an Army training area and surrounded by other military reservations. The growing complexity of the electromagnetic environment was seen to be detrimental to DSN interests and of growing concern to other spectrum users in the extended Goldstone area. In recognition of the need for environmental control, a memorandum of understanding (Ref. 11) between NASA and the Department of Defense was written. The MOU included provisions for an operational coordination group and for the exchange of necessary technical information. In this regard, JPL is obligated to provide data suitable for interference calculations by the Electromagnetic Compatibility Analysis Center (ECAC), a contractor-operated facility for the Department of Defense.

The Mojave Coordinating Group has been operating for the last few years and is probably the single most effective mechanism for protecting the DSN stations at Goldstone. Since the group came into being, the trend of actual RFI to the DSN has been dramatically reversed. Incidental to the greatly expanded training activity being planned for the Army's Fort Irwin, ECAC undertook the development of RFI prediction capability, aimed in part at protecting Goldstone operations from the more intense potential for RFI.

A. ECAC Modelling

At the time that ECAC began the development of RFI prediction capability for Fort Irwin, a suitable mathematical model of DSN susceptibility did not exist. ECAC has therefore been working on model development based on detailed circuit and specification data on the Block IV receivers. This data was supplied by JPL. The ECAC effort parallels in time the JPL work on RFI models but has followed a somewhat different analytic approach.

From time to time JPL has been invited to comment on the ECAC reports describing their model development. The difficulty is that the analytic models require either laboratory or field verification. ECAC cannot do the laboratory testing since they do not have the requisite DSN equipment. In the absence of such testing, the only alternative is a comparison of trial predictions made by ECAC with predictions using the verified JPL models and/or DSIP2. JPL requested early in 1981 that ECAC provide trial predictions for selected CW cases, but these have not yet been received.

B. ECAC/Fort Irwin RFI Prediction

ECAC expects to manage the Fort Irwin environment by means of various models and analysis programs resident in an on-site computer. Utilizing predicted data on DSN operations (mission criticality, antenna pointing, signal level, etc.) the corresponding RFI potential of Fort Irwin exercises using many types of radio equipment would be analyzed.

The ultimate use of RFI predictions by ECAC would be to control the scheduling of Fort Irwin operations, to influence the selection and use of radio equipment in the area, and to influence DSN operations where possible. Experience will show whether elaborate or simple models are needed for RFI effective control vis-a-vis the Goldstone stations. One simple approach would be to cease operation of certain equipment during critical DSN mission phases at times when DSN elevation angles are below a selected value. This kind of control needs only operational data since the necessary criteria can be determined by analysis done in advance. ECAC tends to think of more elaborate schemes involving detailed real-time information transfer from Goldstone via hardline or microwave: received signal strength, antenna pointing, data mode, etc. The objective of this complicated approach is to maximize the freedom for Fort Irwin operations.

V. International Protection

The international Radio Regulations (Ref. 12) govern the use of the radio frequency spectrum. The Regulations specify methods and procedures by which the potential for interference is avoided or managed. To provide interference protection for various radio services, maximum permissible levels of interference are listed. The current values for deep-space receiving earth stations are based on analysis done at JPL circa 1968 (Ref. 13). The internationally adopted values of permissible interference power also serve JPL and NASA frequency managers in negotiations with other domestic agencies.

Current values of permissible interference as listed in the Regulations are based on an old analysis that does not consider coded telemetry. A more modern determination of permissible interference power is lacking, and hence the possibility of inadequacy of existing values of permissible interference cannot be determined. The protection afforded by the current international Radio Regulations is not known to cause problems for deep-space downlinks.

VI. Summary Assessment and Expected Future Work

In the foregoing sections we have discussed some history and status of four aspects of RFI analysis and prediction:

(1) Modeling

(2) Prediction

(3) Operational management

(4) Regulations

A. Model Development

The objectives of modelling remain valid:

(1) Support the detailed analysis of specific RFI situations.

(2) Provide for RFI prediction, both automatic and manual.

(3) Establish levels of permissible interference.

Continued JPL modelling is applicable to providing capability for RFI analysis and for providing improvements to the prediction program DSIP2. Although much of this report has dealt with DSIP2, it should be remembered that there is a more general class of analysis regarding actual and potential interference from a wide variety of sources other than earth-orbiting satellites.

The largely completed work on CW modeling sets the stage for consideration of pulse interference. Several approaches to analysis and modelling are possible, and the results of DSIP2 performance evaluation (Sec. VI-B) may affect the final selection. A likely approach is a continuation of the philosophy of developing models that use direct mathematical expression of input signals rather than the spectral line approach of DSIP2. The direct expression method is usually more general and can be expected to handle signals not amenable to the CW spectral line approach.

If DSIP2 evaluation shows that the empirical CW models themselves need improvement, the JPL analytic CW models should be utilized. It is presently intended that several approaches to pulse modelling will be explored. DSIP2 evaluation will provide additional information to guide future work.

B. RFI Prediction

Incorporated in the satellite RFI prediction program, DSIP2, are a number of assumptions:

(1) It is assumed that an interfering signal may be characterized by a set of spectral lines, each line being equivalent to a CW signal. Implicit in this assumption is signal duration that is long with respect to the various time constants in the receiving system. For some kinds of pulse signals this condition may not be met.

(2) It is assumed that the effect of each spectral line may be determined by models based on analysis and test of

receiver response to CW signals. Implicit in this assumption is a relatively stationary spectrum, in terms of amplitude and frequency, with respect to the receiver time constants. Frequency hopping and chirp signals as well as other complex modulation do not fit this assumption.

(3) It is assumed that drop-lock predictions may be based on tests for each spectral line; the cumulative effect of several lines need not be considered. The basis for this assumption has not been explained.

(4) It is assumed that the prediction program is operationally appropriate. The effect of high false alarm rate has not been determined. It is not known if the RFI criteria could be adjusted to reduce the false alarm rate without creating a missed alarm rate.

As discussed in the section on DSIP2 evaluation, it is timely and important to assess the actual performance of the program. A simple statistical summary of prediction vs experience is interesting but not sufficient. Several questions need to be answered:

(1) Are the predicted interference spectra actually present in the receiver?

(2) Do the RFI effects actually encountered reflect the prediction that would be made on the basis of the actual interference spectra?

(3) Is the basic approach of spectral analysis with CW models the correct one for further development?

It is planned that a field test at a selected station will be designed and completed, utilizing sufficient instrumentation to acquire the needed data. The results of these tests would be used to guide further model development, program improvement, or the adoption of different RFI threshold criteria.

The accuracy of DSIP2 predictions is critically dependent on knowledge of trajectory parameters (position, velocity, time) of each satellite and spacecraft pair being considered. It is also necessary to know the particular spacecraft communications mode that will be used during the period of potential RFI. This is because the interference prediction is related to the strength and modulation characteristics of the received spacecraft signal. The telemetry data rate and the presence or absence of ranging modulation affect the calculations.

An intrinsic difficulty of RFI prediction is that the circumstances assumed for calculation purposes may not actually exist at the time that the RFI is predicted to occur. The timeliness of prediction, the number of different calculations that should be made in order to include all likely or possible spacecraft modes, and the criteria concerning the list of satellites to be considered are important factors in the effectiveness of DSIP2. In addition to the assessment of the RFI calculations themselves, these additional factors deserve consideration when deciding the course of future development or prediction capability.

C. Operational Management

The operational management of the DSN RFI environment is believed to be well in hand. A detailed description of the organizations and methods of management is beyond the intended scope of this report and will not be treated further.

The DSN susceptibility modelling by ECAC, done in support of their Fort Irwin RFI management responsibility, does deserve comment. Continuation of this effort is not recommended. It is the author's opinion that the actual RFI environment that will attend expanded Fort Irwin operations is not well enough understood to justify an elaborate, computerized, near-real-time analysis capability. (It is of course possible to postulate an impossibly difficult environment.) What is needed is the ability to locate and characterize those sources that actually cause interference to the DSN. Experience with the environment will confirm or deny the possibility of simple control procedures, such as minimum elevation angle constraints, that will permit economical control of interference. It may be that computerization can be limited to simple a priori total power and path loss calculation with respect to a level of permissible interference.

D. Radio Regulations

To continue the protection of deep-space telecommunications links it is essential that levels of permissible interference be accurately stated in the Radio Regulation. It is therefore planned that a report be prepared to modernize the analysis that supports the regulatory process. Of particular importance is the inclusion of coded telemetry susceptibility analysis. The report would be submitted to the International Radio Consultative Committee (CCIR) for adoption.

References

1. Levitt, B. K., "Carrier Tracking Loop Performance in the Presence of Strong CW Interference," *The Deep Space Network Progress Report 42-51*, Jet Propulsion Laboratory, Pasadena, Calif., June 15, 1979.

2. Hersey, D. R., and Sue, M. K., "Maximum CW RFI Power Levels for Linear Operation of the DSN Block IV Receiver at S-band Frequencies," *Deep Space Network Progress Report 42-56*, Jet Propulsion Laboratory, Pasadena, Calif., April 15, 1980.

3. Sue, M. K., "Block IV Receiver Tracking Loop Performance in the Presence of a CW RFI," *TDA Progress Report 42-60*, Jet Propulsion Laboratory, Pasadena, Calif., December 15, 1980.

4. Sue, M. K., "Telemetry Degradation Due to a CW RFI Induced Tracking Error for the Block IV Receiving System with Maximum Likelihood Convolutional Decoding," *TDA Progress Report 42-61*, Jet Propulsion Laboratory, Pasadena, Calif., February 15, 1981.

5. Sue, M. K., "DSN RFI Susceptibility Models, Development Program Overview, *TDA Progress Report 42-68*, Jet Propulsion Laboratory, Pasadena, Calif.

6. Cain, D. L., and Beyer, P. E., "Radio Frequency Interference by Earth Orbiting Satellites: Deep Space Interference Prediction Program," *TDA Progress Report 42-66*, Jet Propulsion Laboratory, Pasadena, Calif., December 15, 1981.

7. Gallagher, J. F., "Users Guide For Deep Space Interference Program (DSIP2)," Jet Propulsion Laboratory, Pasadena, Calif., November 30, 1977.

8. Low, P. W., "Radio Frequency Interference Effects of Continuous Sinewave Signals on Telemetry Data," *DSN Progress Report 42-40*, Jet Propulsion Laboratory, Pasadena, Calif., August 15, 1977.

9. Low, P. W., "Radio Frequency Interference Effects of Continuous Wave Signals on Telemetry Data: Part II," *DSN Progress Report 42-51*, Jet Propulsion Laboratory, Pasadena, Calif., June 15, 1979.

10. Kliger, I. E., and Olenberger, C. F., "PLL Jump Phenomenon in the Presence of Two Signals," IEEE Transactions on Aerospace Electronic Systems, Vol. AES-12, January 1976, IEEE, New York, NY.

11. Memorandum of Understanding Between Department of Defense and National Aeronautics and Space Administration for Compatible Operations in the Mojave Desert Area, February 1979.

12. Radio Regulations, Edition of 1982, Vols. 1 and 2, International Telecommunication Union, Geneva, 1982.

13. Koerner, M. A., *Effects of Interference on a Binary Communication Channel Using Known Signals*, Technical Report 32-1281, Jet Propulsion Laboratory, Pasadena, Calif., December 1, 1968.

Appendix A

Description of DSIP2 RFI Models[1]

I. Introduction

In this appendix we examine the equations and tests for RFI as they are implemented in DSIP2. Appendix B is taken from a JPL internal document and is a concise listing of these, from the point of view of a software engineer.

Subroutine CIANSP in DSIP2 makes six tests for interference:

(1) Receiver interference: received power above a fixed threshold and frequency separation less than a fixed value.

(2) Receiver drop-lock, Type 1.

(3) Receiver drop-lock, Type 2.

(4) Telemetry drop-lock, Type 1.

(5) Telemetry drop-lock, Type 2.

(6) Telemetry SNR degradation.

For each of these tests there are equations that are used to make the necessary calculations and decisions that lead to a prediction to RFI.

The calculations and tests are made with respect to the individual spectral lines that characterize the received signals. The power and frequency of each spike (spectral line) are determined and used in the various tests for RFI.

II. The Received Signal

It is important to remember that the accuracy of DSIP2 predictions is totally dependent on the knowledge and assumptions made regarding the strength of the interfering signal. The radiated amplitude and frequency of each spectral line must be specified for input to DSIP2. The amplitude is normally expressed as an e.i.r.p. that accounts for the transmitted power and main beam antenna gain. Calculations using these values of e.i.r.p. are therefore worst case in the sense that the satellite antenna may not always point directly at the earth station.

[1]The material presented in this appendix is based on a more extensive analysis by Ali Salmasi and his contribution is gratefully acknowledged.

III. Receiver Interference

Receiver interference is judged to occur if two conditions are met as shown in formula[2] F13:

$$\Delta f^*_{c_i} - 1000 \text{ Hz} \leqslant 0,$$

and

$$I_{e_i} \geqslant -175 \text{ dBm}$$

$\Delta f^*_{c_i}$ is the frequency separation between an interfering spike and the carrier frequency of the desired signal, taking into account doppler shifts, as computed by F11. The 1000-Hz criterion is the result of experimental tests (Ref. 8).

The −175 dBm criterion is related to the sensitivity of the Block IV receiving system. For a desired signal that results in normal threshold conditions, a CW signal less than −175 dBm is predicted to be free of RFI effects.

The power of a received spike is calculated from F8:

$$I_{e_i} = P_{SAT_i} - L_{S_{SAT}} + G_{A_{SAT}} - MGR$$

The transmitted power level of the ith spike, P_{SAT_i}, is the e.i.r.p in the direction of the earth station. This value actually varies as a function of the pointing direction of the satellite antenna. For DSIP2, the worst-case main beam e.i.r.p. is used. $L_{S_{SAT}}$ is found by F2 and gives a value for free space loss. It does not include atmosphere loss effects and hence is conservative in the sense of giving maximum interference power. $G_{A_{SAT}}$ is the earth station antenna gain in the direction of the satellite and is determined by F1, which gives a simple envelope relationship between gain and angle off boresight. The envelope method is conservative.

MGR is an estimate of the reduction in maser gain as a function of received power. *MGR* is found from an empirical formula, F5, which was derived from test data and which depends on the total received power level I_e. I_e is given by F4 and once again is e.i.r.p. in the direction of the DSS. Notice that $MGR = 0$ for interference power less than −90 dBm.

[2]Formulas referred to in this appendix may be found in Appendix B.

Notice also that *MGR* does not depend on frequency. More recent studies have shown that maser/receiver saturation *is* a function of frequency. The DSIP2 modelling of saturation could be modernized.

IV. Receiver Drop-Lock

The carrier tracking loop in a DSS receiver normally is locked to a desired signal. If interference causes the loop to no longer track the desired signal, drop-lock is said to have occurred. DSIP2 predicts two kinds of drop-lock: jump to the interfering signals, and loss of desired signal by suppression.

A. Receiver Drop-Lock, Type 1 (Jump)

Type 1 drop-lock is based on the jump phenomenon observed in some phase-locked loop (PLL) receiver systems. If such a PLL is initially tracking a desired signal, the loop will jump to an interfering signal that is sufficiently near in frequency and strong in power. In DSIP2, the tests for Type 1 drop-lock are given by F15:

When $\Delta f^*_{c_i} < B_e$, the test for drop-lock is

$$I_{e_i} - (P_c - MGR) \geq 0$$

When $\Delta f^*_{c_i} \geq B_e$, the test for drop-lock is

$$I_{e_i} - (P_c - MGR) - 20 \log\left(\frac{\Delta f^*_{c_i}}{B_e}\right) \geq 0.$$

where B_e is the noise bandwidth of the carrier tracking loop. In DSIP2, B_e is set at a fixed value of 12 Hz, which is a typical value for DSN receivers operating near threshold. The loop bandwidth is actually a function of signal level, becoming wider as the level increases. P_c is calculated by F14 and is the carrier component of the total power of the desired signal; P_c depends upon the modulation index and is related to the telemetry mode.

For the case where the interfering spectral line is separated from the desired carrier by an amount less than the loop bandwidth, drop-lock is predicted when the line is equal to or stronger through the desired carrier signal. For the case where the interfering line is separated from the desired carrier by an amount equal to or greater than the loop bandwidth, drop lock is predicted when the line less an amount proportional to the frequency separation is equal to or greater than the desired carrier signal.

These equations were developed from empirical data and, perhaps, an interpretation of the analysis presented in Ref. 7.

B. Receiver Drop-Lock, Type 2 (Maser Saturation)

If a Type 1 receiver drop-lock is not predicted, the test of F16 is made. In this test, Type 2 drop-lock is predicted if the carrier power is reduced by maser gain reduction (*MGR*) to a value less than the noise power in the tracking loop bandwidth. This test is independent of frequency separation and applies to the case where the total interference power exceeds -90 dBm; at lower power there is no *MGR*.

The Type 2 model does not consider a more general analysis of receiver saturation and corresponding effects on performance. More complete modelling is now possible.

V. Telemetry Drop-Lock

The telemetry demodulation and detection process in DSS receivers makes use of phase-locked loops. These track the subcarrier frequency and the symbol and bit rates. Based on experimental data, empirical expressions for predicting telemetry drop-lock were developed. Two types of telemetry drop-lock are predicted by DSIP2: Type 1, jump, and Type 2, SNR degradation.

A. Telemetry Drop-Lock, Type 1 (Jump)

The tests for telemetry Type 1 drop-lock are given by F30:

When $\Delta_{f_{nsc_i}} < SR$, the test for drop-lock is

$$1.3\ [P_{I_i} - (P_D - MGR) - 3] \geq 0.$$

When $\Delta f_{n_{sc_i}} \geq SR$, the test for drop-lock is

$$1.3\ [P_{I_i} - (P_D - MGR) - 3] - 20 \log\left(\frac{\Delta f^*_{n_{sc_i}}}{SR}\right) \geq 0.$$

where $\Delta f_{n_{sc_i}}$ is the frequency difference between a particular spike and a particular subcarrier harmonic in the desired signal. It has been shown that a CW signal at or near a subcarrier harmonic can cause telemetry drop-lock. The effective inter-

ference power of the CW line is reduced for the higher order harmonics and P_D is computed by F19 to account for this. P_D is the data power in the desired signal and is calculated by F29.

These expressions were developed by assuming a mathematical model and then modifying it to fit a set of test data points (Ref. 8). The assumed model is of the same general form presented in Ref. 10, which given an analysis of the PLL jump phenomenon for a particular set of conditions. Comparison of Ref. 10 with equations in F30 shows that the DSIP2 implementation differs in detail. The numerical constants in F30 are the result of curve fitting.

B. Telemetry Drop-Lock, Type 2 (Maser Saturation)

If a Type 1 telemetry drop-lock is not predicted, the following test is made:

$$SNR_{OUT} - \Delta SNR_T + 5 \text{ dB} \geqslant 0$$

where SNR_{OUT} is found by Formulas 32, 31 and 28. This test is based on the performance of the symbol synchronizer assembly (SSA), wherein a *SNR* greater than −5 dB must be present to achieve an in-lock condition.

VI. Telemetry SNR Degradation

Interference that is not strong enough to cause drop-lock may nevertheless result in degraded telemetry performance. The effect of degraded performance is an increase in the data error rate. For a given telemetry system there is a relationship between signal-to-noise ratio (*SNR*) and the error rate. DSIP2 predicts the change in *SNR* due to interference. The predicted change is given by F28:

$$\Delta SNR_T = \Delta SNR + MGR$$

where

ΔSNR_T is the total change in *SNR*

ΔSNR reflects the noise added by the interference

MGR is the reduction in signal caused by maser saturation.

The degradation in *SNR* caused by the interference is given by F27:

$$\Delta SNR = 10 \log \left(\frac{T_R + T_s}{T_s} \right)$$

where

T_R is the cumulative noise temperature for all interfering spikes.

T_s is the system noise temperature in the absence of interference.

The determining concept in this formulation is that the effect of a set of CW spikes can be represented in terms of a change in system noise temperature. References 8 and 9 present an analytic development using this premise. The relationship between CW interference and the corresponding change in noise temperature was determined experimentally. Using theoretical curves of bit error rate as a function of *SNR*, and experimental data on bit error rate as a function of interference, the expressions of F24 and F25 were derived by curve fitting.

The first experimental test data was obtained for a bit rate of 2000 bps and for the case where the interfering spike was coincident with the frequency of the telemetry subcarrier. The relationship between interference power and noise temperature was then determined.

F25 gives the noise temperature for a single spectral line:

$$T_{R_i} = \left[\left(821\, e^{0.0421 P^*_{I_i}} \right)^2 + 40^2 \right]^{1/2} - 39.5$$

where

T = is the noise temperature due to the ith spike

P_I is the power in the ith spike

An observed effect is that a CW spike at or near the subcarrier frequency, or its harmonics, will cause interference related to the frequency separation and the harmonic number. To account for this mechanism, as well as to accommodate symbol rates other than 2000 bps, additional test data and curve fitting gave Formula 24, which relates several factors:

P_I, the interference power to be used in Formula 25

I_e, the interference power of a spike found by Formula 8

N_i, the subcarrier harmonic number found by Formula 18

P_D, the total data power of the desired signal as found by Formula 29

SR, the symbol rate

$\Delta f^*_{nSC_I}$, the frequency separation between the subcarrier harmonic and interfering spike as found by Formula 21

The reader is urged to examine F25 and its subordinate equations in Appendix B. They are excellent examples of the result of curve fitting that forces a mathematical expression to fit experimental data by means of peculiar arithmetic constants.

The telemetry *SNR* degradation model described above can be expected to predict interference for the conditions implicit in the experimental test and subsequent curve fitting. Regardless of the analytic assumptions that were made, the several formulas were forced to fit these data. The problem is that other conditions may require different values for the constants in the formulas, or different formulas. For example, the test data used to develop the existing model was taken with an uncoded data stream. For each type of coded data, there is a different curve relating error rate and *SNR*. DSIP2 does not account for this difference.

Finally, there is a general test for telemetry degradation. Interference is judged to be present when

$$\Delta SNR_T \geqslant 0.5 \text{ dB}.$$

Appendix B

Interference Tests and RFI Formulas

I. Interference Tests

The following six interference functions are the basic tests made by subroutine CIANSP to determine if radio frequency interference exists, the form it takes and the telemetry modes impacted at a given point in time:

1. Receiver Interference Function (Formula F13, see Part II for this and other formulas)

If for one or more satellite spectrum spikes it is true that:

$$\Delta f^*_{c_i} - 1000 \text{ Hz} \leqslant 0 \quad .\text{AND.} \; I_{e_i} \geqslant -175 \text{ in dBm}$$

then receiver interference is judged to have occurred and the flag BRI = .TRUE. is set at line 212 of CIANSP. These conditions are a necessary part of the requirements for receiver drop-lock of the first type. The event of receiver interference occurring is recorded in the bits of array IISW for the subcarrier bit rate mode being tested if receiver drop-lock does not occur.

$\Delta f^*_{c_i}$ = frequency separation between a spike and the spacecraft carrier as received at the antenna and adjusted for worst-case trajectory errors (Formula F11) in Hz and

I_{e_i} = effective power level of a spike (Formula F8) in dBm

2. Receiver Drop-Lock Function for the Jump Phenomenon (First Type) (Formula F15)

If receiver interference occurs for a subcarrier bit rate mode, its spikes are checked for the following conditions:

$$I_{e_i} - (P_c - MGR) - 20 \log \left(\frac{\Delta f^*_{c_i}}{B_e} \right) \geqslant 0 \text{ for } \Delta f^*_{c_i} \geqslant B_e \text{ in Hz}$$

or

$$I_{e_i} - (P_c - MGR) \geqslant 0 \text{ for } \Delta f^*_{c_i} < B_e \text{ in Hz}$$

If these conditions also hold for one or more spikes, receiver drop-lock of the first type has occurred, the flag BRDL = .TRUE. is set at line 313 of CIANSP and the appropriate bits of array IISW are set for this bit rate mode.

P_c = downlink carrier signal power level (Formula F14) in dBm

MGR = maser gain reduction (Formula F5) in dB and

B_e = receiver RF loop noise bandwidth for the station = 12 Hz

3. Receiver Drop-Lock Function for Maser Saturation (Second Type) (Formula F16)

If receiver drop-lock of the first type does *not* occur for a subcarrier bit rate mode, the following test is made:

$$P_c - MGR + 198.6 - 10 \log B^*_e - 10 \log T_s \leqslant 0.$$

If this condition holds, receiver drop-lock of the second type has occurred, the flag BDRL = .TRUE. is set at line 323 of CIANSP and the appropriate bits of array IISW are set for this bit rate mode. The bit setting is the same in IISW for both types of receiver drop lock.

T_s = antenna cold sky temperature (K) for the station being checked and

B_c^* = receiver RF loop noise threshold for the station = 12 Hz.

4. Telemetry Drop-Lock Function of the First Type (Formula F30)

If receiver drop lock of either type does not occur for a subcarrier bit rate mode, the following test is made for each satellite spectrum spike:

$$P_{I_i} \geqslant -175 \text{ dBm .AND. } 1.3\ [P_{I_i} - (P_D - MGR) - 3]\ -20 \log \left(\frac{\Delta f^*_{n_{SC_i}}}{SR} \right) \geqslant 0 \text{ for } \Delta f^*_{n_{SC_i}} \geqslant SR$$

or

$$P_{I_i} \geqslant -175 \text{ dBm .AND. } 1.3\ [P_{I_i} - (P_D - MGR) - 3]$$

$$\text{for } \Delta f^*_{n_{SC_i}} < SR.$$

If this condition holds for one or more spikes, telemetry drop-lock of the first type has occurred, the flag BTDL = .TRUE. is set at line 386 of CIANSP and the appropriate bits of array IISW are set for this bit rate mode.

P_{I_i} = power quantity for a spike (Formula F19) in dBm,

P_D = total data power (Formula F29) in dBm,

$\Delta f^*_{n_{SC_i}}$ = frequency separation between the ith spike and the N_ith subcarrier harmonic as adjusted for worst-case trajectory errors (Formula F21) in Hz,

SR = bit rate mode symbol rate (Formula F23) in bits per second.

5. Telemetry Drop-Lock Function of the Second Type

If neither receiver drop-lock nor telemetry drop-lock of the first type occurs for a subcarrier bit rate mode, the following test is made:

$$SNR_{OUT} - \Delta SNR_T + 5 \text{ dB} \leqslant 0.$$

If this condition holds, telemetry drop lock of the second type has occurred, the flag BTDL = .TRUE. is set at line 410 of CIANSP and the appropriate bits of array IISW are set for this bit rate mode. The bit setting is the same in IISW for both types of telemetry drop-lock.

SNR_{OUT} = output signal-to-noise ratio (Formula F32) in dB and

ΔSNR_T = total signal-to-noise degradation (Formula F28) in dB.

6. Signal-to-Noise Ratio Degradation Function

If none of the above types of drop-lock occurs for a subcarrier bit rate mode, the following test is made:

$$\Delta SNR_T \geqslant 0.5 \text{ dB}.$$

If this condition holds, signal-to-noise ratio degradation has occurred, the flag BDSNI = .TRUE. is set at line 414 of CIANSP and the appropriate bits of array IISW are set for this bit rate mode.

II. RFI Formulas

The RFI formulas referred to in the preceding discussion are specified below as they are implemented in the subroutine CIANSP code:

1. F1: Antenna Gain, G_A,

   For 26 meter S-band stations 11, 12, 13, 42, 44, 61 and 62:

$$G_A = \begin{cases} 53.3 \text{ dB} \\ 32 - 25 \log \Delta \\ -10 \end{cases} \text{ for } \begin{cases} 0° \leqslant \Delta \leqslant 0.14°, \text{ or} \\ 0.14° < \Delta \leqslant 45°, \text{ or} \\ 45° < \Delta . \end{cases}$$

   For 64 meter S-band stations 14, 43 and 63:

$$G_A = \begin{cases} 61.7 \text{ dB} \\ 32 - 25 \log \Delta \\ -10 \end{cases} \text{ for } \begin{cases} 0° \leqslant \Delta \leqslant 0.065°, \text{ or} \\ 0.065° < \Delta \leqslant 45°, \text{ or} \\ 45° < \Delta . \end{cases}$$

   Δ = cone angle (°) between the antenna/spacecraft downlink direction and the antenna/satellite direction

2. F2: Space Loss, L_S in dB,

   $L_S = 32.45 + 20 \log f + 20 \log \rho$,

   f = spike frequency or carrier frequency in MHz and

   ρ = station-satellite or station-spacecraft distance in km.

3. F4: Total Effective Interference Power I_e in dBm,

$$I_e = P_{OUT_{SAT}} - L_{S_{SAT}} + G_{A_{SAT}},$$

   $P_{OUT_{SAT}}$ = satellite total power in dBm,

   L_{SAT} = space loss for satellite using frequency of first spike as the input frequency in dB and

   $G_{A_{SAT}}$ = antenna gain with respect to the spacecraft-to-satellite cone angle separation in dB.

4. F5: Maser Gain Reduction, MGR, in dB,

$$MGR = \begin{cases} 25.106 * \left[1 + \dfrac{(I_e + 90)^2}{441.378}\right]^{1/2} - 0.131 * I_e - 36.644, \\ \text{or} \qquad \text{for } I_e > -90 \text{ dBm} \\ 0 \qquad \text{for } I_e \leqslant -90 \text{ dBm.} \end{cases}$$

5. F6: Transmitter Frequency, TX, in Hz,

TX = 96 * TSF and

TSF = transmitter synthesizer frequency in Hz,

6. F7: Carrier Frequency Received at Antenna, f_c', in Hz,

$$f_c' = TX \left(1 - \frac{\dot{\rho}_u}{c}\right) \left(\frac{240}{221}\right) \left(1 - \frac{\dot{\rho}_d}{c}\right),$$

$\dot{\rho}_u$ = uplink range rate in km/s,

$\dot{\rho}_d$ = downlink range rate in km/s and

c = speed of light in km/s.

7. F8: Effective Power Level of a Spike, I_{e_i}, in dBm,

$$I_{e_i} = P_{SAT_i} - L_{S_{SAT}} + G_{A_{SAT}} - MGR \text{ and}$$

P_{SAT_i} = transmitted power level of the ith spike in dBm.

8. F9: Frequency of a Spike as Received at Antenna, f_{I_i}', in Hz,

$$f_{I_i}' = f_{I_i} \left(1 - \frac{\dot{\rho}_{SAT}}{c}\right),$$

f_{I_i} = satellite spike frequency in Hz and

$\dot{\rho}_{SAT}$ = antenna/satellite slant range rate in km/s.

9. F10: Frequency Separation Between a Spike and the Carrier as Received at Antenna, Δf_{c_i}, in Hz,

$$\Delta f_{c_i} = \left\| f_{I_i}' - f_c' \right\| .$$

10. F11: Spike/Carrier Frequency Separation as Adjusted for Worst-Case Trajectory Errors, $\Delta f_{c_i}^*$, in Hz,

$$\Delta f_{c_i}^* = \begin{cases} \Delta f_{c_i} - \epsilon_{\Delta f_c} & \text{for } \epsilon_{\Delta f_c} < \Delta f_{c_i} \\ \text{or} & \\ 10^{-100} & \text{for } \epsilon_{\Delta f_c} \geq \Delta f_{c_i} \end{cases}$$

$\epsilon_{\Delta f_c}$ = frequency separation adjustment for worst-case trajectory errors in Hz

$$\epsilon\Delta f_c = f'_{I_i} * \frac{\dot{\rho}_{WC}}{c} + TX * \left(\frac{240}{221}\right) * \left[\left(1 - \frac{\dot{\rho}_u}{c}\right)\frac{\dot{\rho}_{WC}}{c} + \left(1 - \frac{\dot{\rho}_d}{c}\right)\frac{\dot{\rho}_{WC}}{c}\right]$$

and

$\dot{\rho}_{WC}$ = worst-case antenna/spacecraft range rate error in km/s.

11. F12: Receiver Peak Value Function for a Spike, P_{r_i}, in dBm,

$$P_{r_i} = \begin{cases} I_{e_i} - 20 \log \Delta f^*_{c_i} & \text{for } \Delta f^*_{c_i} \geq 12 \text{ Hz.} \\ \text{or} & \\ I_{e_i} & \text{for } \Delta f^*_{c_i} < 12 \text{ Hz.} \end{cases}$$

Note: p_{r_i} is not required as part of input to any essential interference computations but is computed as information for the user as regards the most potent spike at this time point and station.

12. F13: Receiver Interference Function

The BRI receiver interference flag is set = .TRUE. when for one or more spikes *both* of the following are true:

$$\Delta f^*_{c_i} - 1000 \text{ Hz} \leq 0 \text{ .AND. } I_{e_i} \geq -175 \text{ dBm.}$$

13. F14: Downlink Carrier Signal Level, P_c, in dBm,

$$P_c = \begin{cases} P_{OUT_{SC}} + 20 \log (\cos \theta) + 20 \log (\cos \theta^*) - L_{S_{SC}} + G_{A_{SC}} - L_0 & \\ \text{for a dual subcarrier} & \\ \text{or} & \\ P_{OUT_{SC}} + 20 \log (\cos \theta) - L_{S_{SC}} + G_{A_{SC}} - L_0 & \\ \text{for a solo subcarrier.} & \end{cases}$$

$P_{OUT_{SC}}$ = spacecraft transmitter output power level in dBm,

0 = modulation index for current bit rate mode in deg,

θ^* = modulation index for dual subcarrier in deg,

$L_{S_{SC}}$ = space loss for the spacecraft from F2 using carrier frequency f_c' in dB,

$G_{A_{SC}}$ = antenna gain for the spacecraft from F1 with $\Delta = 0$ in dB and

L_0 = all other losses = 0.5 dB.

14. F15: Receiver Drop-Lock Function for the Jump Phenomenon (First Type)

The BDRL receiver drop-lock flag is set = .TRUE. when for one or more spikes and the current bit rate mode:

$$\begin{cases} BRI = \text{.TRUE. .AND. } I_{e_i} - (P_c - MGR) - 20 \log \left(\dfrac{\Delta f_{c_i}^*}{B_e} \right) \geqslant 0 \\ \text{for } \Delta f_{c_i}^* \geqslant B_e \text{ in Hz} \\ \text{or} \\ BRI = \text{.TRUE. .AND. } I_{e_i} - (P_c - MGR) \\ \text{for } \Delta f_{c_i}^* < B_e \text{ in Hz,} \end{cases}$$

B_e = receiver RF loop noise bandwidth for the station = 12 Hz.

15. F16: Receiver Drop-Lock Function for Maser Saturation (Second Type)

If receiver drop-lock of the first type does not occur for the current bit rate mode, nevertheless *BDRL* = .TRUE. is set if:

$P_c - MGR - (-198.6 + 10 \log B_e^* + 10 \log T_s) \leqslant 0$ in dBm,

T_s = antenna cold sky temperature for the current station in K and

B_e^* = receiver RF loop noise threshold for the station = 12 Hz.

16. F17: Frequency Ratio for a Spike, w_i

$$w_i = \frac{\Delta f_{c_i}}{f_{SUBC}} \quad ,$$

f_{SUBC} = subcarrier frequency for the current bit rate mode in Hz.

17. F18: Subcarrier Harmonic Number for a Spike, N_i

$$N_i = \begin{cases} 1 \text{ for } 0 \leqslant w_i \leqslant 2 \text{ and} \\ w - 1 \text{ for } W - 2 < w_i \leqslant W \end{cases}$$

where $W = 4, 6, 8 \ldots$ any positive even integer above 2;

N_i = number of the subcarrier harmonic most affected by the ith spike.

18. F19: Power Quantity for a Spike, P_{I_i}, in dBm,

$$P_{I_i} = I_{e_i} - 0.94 * 20 \log N_i .$$

19. F20: Frequency Separation Between the ith Spike and the N_ith Subcarrier Harmonic, $\Delta f_{n_{SC_i}}$, in Hz,

$$\Delta f_{n_{SC_i}} = \begin{cases} \left\| f'_{I_i} - (f'_c + N_i * f_{SUBC}) \right\| & \text{for } f'_{I_i} \geq f'_c \\ \text{or} & \\ \left\| f'_{I_i} - (f_c - N_i * f_{SUBC}) \right\| & \text{for } f'_{I_i} < f'_c . \end{cases}$$

20. F21: Spike/Subcarrier Harmonic Frequency Separation as Adjusted for Worst-Case Trajectory Errors, $\Delta f_{n_{SC_i}}$, in Hz

$$\Delta f^*_{n_{SC_i}} = \begin{cases} \Delta f_{n_{SC_i}} - \epsilon_{\Delta f_c} & \text{for } \epsilon_{\Delta f_c} < \Delta f_{n_{SC_i}} \\ \text{or} & \\ 10^{-100} & \text{for } \epsilon_{\Delta f_c} \geq \Delta f_{n_{SC_i}} . \end{cases}$$

21. F22: Telemetry Peak Value Function for a Spike, p_{t_i}, in dBm,

$$p_{t_i} = P_{I_i} - 20 \log \Delta f^*_{n_{SC_i}} .$$

Note: P_{t_i} is not required as part of input to any essential interference computations but is computed as information for the user as regards the most potent spike at this time point, bit rate mode and station

22. F23: Symbol Rate of the Current Bit Mode, SR, in bits per second,

$$SR = BR * m_{CODE} ,$$

BR = bit rate of the current bit rate mode and

m_{CODE} = telemetry code multiplier for the current bit rate mode.

23. F24: Quantity $P^*_{I_i}$ for a Spike in dBm,

$$P^*_{I_i} = \begin{cases} I_{e_i} - 0.94 * 20 * \log - 0.10 * (P_D - MGR + 141.) - 0.90 * 10 * \log\left(\frac{SR}{2000}\right) & \text{for } \frac{\Delta f^*_{n_{SC_i}}}{SR} \leqslant 1, \\ \text{or} & \\ I_{e_i} - 0.94 * 20 * \log N_i - 0.10 * (P_D - MGR + 141.) - 0.90 * 10 * \log\left(\frac{SR}{2000}\right) & \\ \quad - 0.90 * 20 * \log \left\{\left[\text{Int}\left(\frac{\Delta f^*_{n_{SC_i}}}{SR}\right) + 0.5\right] * \Pi\right\} & \text{for } \frac{\Delta f^*_{n_{SC_i}}}{SR} > 1, \end{cases}$$

P_D = total data power (F29) in dBm.

24. F25: Noise Temperature for a Spike, T_{R_i}, in K,

$$T_{R_i} = \begin{cases} \left[\left(10^{821\, e^{0.421\, P^*_{I_i}}}\right)^2 + 40^2\right]^{1/2} - 39.5 & \\ \text{or} & \text{for } P_{I_i} \geqslant -175 \\ 0 & P_{I_i} < -175. \end{cases}$$

25. F26: Noise Temperature for All Spikes, T_R, in K,

$T_R = \sum_i T_{R_i}$ where $\sum_i$ indicates summation over all of the spikes for the current satellite.

26. F27: Signal-to-Noise Ratio Degradation, ΔSNR, in dB,

$$\Delta SNR = 10 \log\left(\frac{T_R + T_S}{T_S}\right)$$

27. F28: Total Signal-to-Noise Ratio Degradation, ΔSNR_T in dB,

$$\Delta SNR_T = \Delta SNR + MGR.$$

28. F29: Total Data Power, P_D, in dBm,

$$P_D = P_C + 20 \log (\tan \theta)$$

29. F30: Telemetry Drop-Lock Function (First Type)

If receiver drop-lock at either type does not occur for a bit rate mode but *both* of the following conditions hold for one or more spikes:

$$P_{I_i} \geq -175 \text{ dBm .AND. } 1.3\ [P_{I_i} - (P_D - MGR) - 3] - 20 \log \left(\frac{\Delta f^*_{n_{SC_i}}}{SR} \right) \geq 0 \quad \text{for } \Delta f^*_{n_{SC_i}} \geq SR,$$

or

$$P_{I_i} \geq -175 \text{ dBm .AND. } 1.3\ [P_{I_i} - P_D - MGR - 3] \quad \text{for } \Delta f^*_{n_{SC_i}} < SR,$$

then *BTDL* = .TRUE. is set for this bit rate mode.

30. F31: Input Signal-to-Noise Ratio, SNR_{IN}, in dB,

$$SNR_{IN} = P_D - 10 \log SR - 10 \log T_S + 198.6.$$

31. F32: Output Signal-to-Noise Ratio, SNR_{OUT} in dB,

$$SNR_{OUT} = SNR_{IN} - SL,$$

where SL = system losses for the station = 0.5 dB.

32. Telemetry Drop-Lock Function for Signal-to-Noise Degradation

If neither receiver drop-lock nor telemetry drop-lock of the first type occurs for a bit rate mode, the following test is made:

$$SNR_{OUT} - \Delta SNR_T + 5 \text{ dB} \geq 0,$$

If all these conditions hold, *BTDL* = .TRUE. is set for this bit rate mode.

33. Signal-to-Noise Ratio Degradation Function

If none of the several types of drop-lock occur for a bit rate mode the following condition is tested:

$$\Delta SNR_T \geq 0.5 \text{ dB}.$$

If this holds, *BDSNI* = .TRUE. is set for this bit rate mode.

Field Interface Module Software Description

H. Valtier
Software Development Section

The software for the Field Interface Module (FIM) is designed to fully optimize the hardware capabilities of the FIM. Software routines exist which enable the FIM to monitor data, control components, communicate with an uplink computer, and perform real-time and local self-diagnostics. The FIM program is not dependent upon any particular uplink computer and is adaptable to various applications with a minimum of modifications.

I. Introduction

The Field Interface Module (FIM) is a device that functions as a standalone multipurpose controller and data collector. The FIM can control various binary signals and collect (monitor) various digital and analog signals for analysis or interpretation. Furthermore, the FIM can perform real-time self-diagnostics to ensure its own operational integrity. The FIM is designed to interface and communicate with a computer from which it receives real-time commands and to which it can transmit monitored data. The FIM can also perform local self-diagnostics, independent of the uplink computer, by interfacing with a diagnostic tool called the Field Interface Diagnostic Assembly (FIDA).

The topics that will be discussed will provide a description of the architecture and the general capabilities of the FIM from a software standpoint, a brief description of the FIM program, its general I/O characteristics, and some typical applications of the FIM.

II. Architecture

A diagram of the functions and interfaces of the FIM is illustrated in Fig. 1. A system that contains analog sensors (pressure transducers, temperature gages, displacement transducers, etc.) and/or control components (on-off switches, relays, solenoids, etc.) can be monitored and controlled by the FIM. The FIM accepts as its inputs a maximum of 32 digital signals and 32 differential analog signals (three of which are dedicated to internal testing). Furthermore, the FIM will control as its outputs a maximum of 32 digital output lines. A serial I/O communication link is provided for communicating with an uplink computer from which the FIM can receive control commands and to which the FIM can transmit monitored data and its diagnostic status. Finally, the FIM can perform local (as well as real-time) self-diagnostics utilizing the FIDA by temporarily terminating the communication link with the uplink computer.

The FIM is centered around a National Semiconductor BLC 80/204 single board computer (INS8080A-2 CPU). The single board computer contains: (1) 8k of programmable read-only memory (PROM), (2) 4k of random-access memory (RAM), (3) an 8259 programmable interrupt controller (PIC), (4) an H2818 system bus controller (multibus), (5) an 8251 serial I/O interface (USART), (6) two 8255A parallel I/O interfaces, and (7) an 8253 programmable interval timer (Ref. 1). Other boards that comprise the FIM include a Data Translation DT1742 analog-to-digital converter, an Ampex (MCM-8080)

16k expansion memory board, and a multibus-compatible digital I/O board to provide expandable I/O capability.

Local self-diagnostics, independent of the uplink computer, are accomplished by utilizing the FIDA. Basically, the FIDA will allow the FIM to perform the same diagnostics that are performed while the FIM is operating in real-time. However, the FIDA contains a 10-character alphanumeric display for conveying information regarding the status of the PROM, the A/D converter, and the expandable digital I/O board. Local, channel by channel, data monitoring of all digital and analog signals can be performed simply by keying in the desired channel type and number. Routines in the FIM program will enable the FIM to automatically sample the desired channel and dynamically display the data with the proper sign and annotation.

III. Capabilities

The software for the FIM is designed in such a manner as to fully optimize the hardware capabilities of the FIM. The software (approximately 4k bytes) resides in PROM and is automatically activated when power is applied to the FIM. Routines exist which enable the FIM to monitor data, control components, communicate with the uplink computer, and perform real-time and local self-diagnostics.

A. Monitoring Data

The data monitoring portion of the FIM software provides the capability of sampling 16 ac (binary) channels, 16 dc (binary) channels and 32 differential analog channels (12 bits/channel). The software allows the FIM to transmit a full data buffer upon request.

B. Controlling Components

The FIM can be commanded by the uplink computer to control various components by utilizing functions built in to the FIM program to initialize, control, set (ON), and reset (OFF) components. An arming function for the control commands is provided to prevent spurious signals from being issued to the control components. Thus an "arm/set" or "arm/reset" command sequence is necessary to physically turn a control component on or off.

C. Communications

The FIM has the additional capability of communicating with an external computer via the RS232C serial I/O interface. All commands enter the FIM through this interface, which is tied to the interrupt structure of the FIM software. Thus any command that is transmitted by the uplink computer will cause the CPU of the FIM to temporarily suspend all other processing and to recognize, interpret, and process the command. Conversely, all monitored data and diagnostic status information is transmitted by the FIM to the external computer via the RS232C interface. The programmable serial I/O interface is programmed for 7-bit ASCII asynchronous transmission (9600 baud) with single start/stop bits, and odd parity. Normal control commands (initialize, control, set, reset, and arm) will take approximately 60 msec to be interpreted and acted upon. The command to send the entire data buffer will take approximately 245 msec.

D. Diagnostics

Real-time self-diagnostics are performed while the FIM is operating in the REMOTE state and will continuously update its latest diagnostic status. Local self-diagnostics are possible when the FIM is programmed into the LOCAL state and is directly interfaced to the FIDA. The following set of self-tests is performed in either state:

(1) Program (PROM) checksum.

(2) A/D converter channel increment capability.

(3) A/D converter reference voltage (ground, +2.5, and +5.0 volts).

(4) Parallel I/O.

Future enhancements will enable the FIM to perform additional diagnostics regarding:

(5) CPU integrity testing.

(6) RAM bit pattern and latency testing.

IV. Program Description

The overall program flow is diagrammed in Fig. 2, which emphasizes the major control routines. Upon power-up, the FIM initializes the BLC 80/204 single board computer (PWRUPR), then proceeds to initialize the rest of the hardware such as the multibus, interrupt controller, RS232C interface, and parallel I/O interfaces. Finally, the program stack is defined, the RAM is cleared, the operating state of the FIM is set to REMOTE SELF TEST, and the program executes the main program (executive) loop.

In the executive loop (EXEC), the FIM continuously calculates and compares the checksum of the program stored in the PROM. The monitor data (ac, dc, and analog) are also read and updated in a large data buffer in RAM for eventual transmission to the uplink computer. Self-diagnostics are performed and the resulting error status conditions are stored in the large data buffer. The state of the FIM is changed to REMOTE and the next pass through the executive loop is

repeated. Only while the FIM is in the REMOTE state are the interrupts enabled. Thus the uplink computer can communicate with the FIM and transmit commands only in this state. All valid command requests to control, arm/set, and arm/reset any ac and dc control components are executed while the FIM is in the REMOTE state (INTRUP). If the state of the FIM is switched to LOCAL via the FIDA keyboard, the FIDA is directly interfaced to the FIM and local self-diagnostics can be performed. All interrupts are disabled while the FIM is in the LOCAL state and, consequently, commands transmitted by the uplink computer will not be recognized. Switching the state of the FIM to REMOTE via the FIDA keyboard will reestablish communications with the uplink computer.

V. I/O Characteristics

A. Input

All commands that are input to the FIM from the uplink computer must be properly formatted in order to achieve the desired results. The command must commence with a Beginning of Transmission (BOT) character and terminate with an End of Transmission (EOT) character, while the body of the actual command is comprised of 13 bytes which contains information such as the FIM number, request number, request type, command source, and input data. All input commands also contain a command checksum, calculated by the uplink computer, with which the FIM can verify that the command was properly transmitted.

The input request buffer requires 13 bytes to define the body of the request, which is represented in 26 ASCII characters (2 characters/byte). Thus 28 characters (BOT+26+EOT) are required to transmit a proper command to the FIM via the RS232C interface. The desired command is dependent upon the request number, request type, and the input data.

B. Output

Whenever any command is communicated to the FIM from the uplink computer, a response from the FIM is immediately transmitted back (after the requested command is processed) to inform the uplink computer of the outcome of its request. In the case of an invalid command request, the FIM will transmit a Negative Acknowledge (NAK) to signal this condition. For most valid command requests, the appropriate response is governed by a format identical to the input request format with the exception of an additional byte (for a total of 14 bytes) that specifies the current operational state of the FIM. Another exception to the above format concerns the request for the FIM to transmit all of its monitored data and diagnostic status. The response that will be transmitted back is comprised of 103 bytes and will contain the latest diagnostic status information of the FIM (0 = all tests passed, non-0 = last test that failed) as well as the entire set of analog and digital data.

The body of the response message requires either 14 or 103 bytes (as described above), which are represented in 28 or 206 ASCII characters (2 characters/byte). A BOT and an EOT character envelop the response message and thus a total of 30 or 208 characters are transmitted back to the uplink computer to convey the outcome of the input request.

VI. Applications

The FIM program is not dependent upon any particular uplink computer and is adaptable to various applications with a minimum of modifications. Because the FIM functions as a standalone multipurpose controller and data collector, there are potentially several applications for the FIM. Two possible applications will be discussed with regard to (1) The Antenna Control and Monitor (ACM) Subassembly and (2) the Technical Facilities Controller (TFC).

A. ACM/FIM Application

The ACM is planned as a subassembly of the Antenna Control Assembly (ACA), which is in turn a part of the overall MARK IVA DSCC Antenna Mechanical Subsystem (ANT) of the Networks Consolidation Project (NCP). The ACM will provide the capability to monitor and control the safety and integrity of the antenna by the utilization of various sensors and controls dispersed throughout the ANT. The ACM will be configured with a dedicated DSN microcomputer operating in a real-time environment. It will interface with a number of FIMs, via RS232C, to monitor and control the ANT subassemblies. Figure 3 shows the interface block diagram of the ACM and its relationship with the FIM. For the sake of brevity, the other subassemblies that interface with the ACM will not be discussed.

Basically, the FIMs will provide the ACM (uplink computer) with the latest monitor data that is continuously being sampled. Enhancements to the software will enable the FIM to perform comparisons of the monitored data with its applicable operational limits (which will be downloaded by the ACM) and to report any out-of-limit conditions to the ACM. Also, the FIMs will accept commands from the ACM to control certain components on the ANT such as the precharge pump selection switch on the servo hydraulics, the flow switch for the alidade pad recess oil flow rate on the hydrostatic bearing, and the brake set/release limit switch on the gear drive. The monitor data will consist of information such as the oil conditioner outlet temperature, the alidade pad oil film thickness and the accumulator pump inlet pressure of the hydrostatic bearing.

B. TFC/FIM Application

The TFC will perform utility control functions at each of the Goldstone Tracking Stations on a real-time basis based upon programming and upon monitor data accumulated from the sensors distributed throughout the stations. The basic units of the TFC will be small microprocessor units located at each of the tracking stations. These units will collect and perform minimal analysis of data from sensors in the Air Conditioning Group (ACG), Meteorological Monitoring Assembly (MMA), Power Distribution Group (PDG), Civil Structures Group (CSG), Power Generation Group (PGG), Lighting Group (LTHG), Energy Conservation Group (ECG), and the Site Protection Group (SPG). These units will also process data, detect variations from desired conditions, and issue alarms and configuration commands based on monitor data, supplied by the FIMs. A simplified block diagram which illustrates the relationship of the TFC with the FIM is shown in Fig. 4. Again, all other TFC interfaces will not be discussed.

The TFC (uplink computer) provides the capability of scanning all FIMs on a periodic basis for acquiring initial data from the various groups and any subsequent changes to the data. Also, the TFC forwards control commands (received from another processor) to the FIMs and receives messages (monitored data, out-of-limit conditions, or responses to commands) from the FIMs. Typical components that are controlled by the FIMs are such points as the starter battery voltage relay, diesel engine speed relay, and the substation circuit breaker relay of the PGG. Other components controlled by the FIMs include the run status relay of the air conditioning unit, the speed control relay of the air handling unit, and the start/stop control relay of the evaporative cooler of the ACG. The monitor data might consist of information such as the space temperatures of the evaporative coolers, the monitor temperature of the chilled water storage of the ACG, and the fuel level of the main tank of the PGG.

VII. Summary

The Field Interface Module (FIM) is a standalone multipurpose controller and data collector which is designed to interface and communicate with a computer from which it receives commands and to which it can transmit data. The software for the FIM enables it to monitor data, control components, communicate with an uplink computer, and perform real-time and local self-diagnostics. The software for the FIM resides in approximately 4k bytes of PROM which is contained on a BLC 80/204 single board computer. The FIM program provides the capability of sampling 32 digital signals and 32 differential analog signals whose data can be transmitted to an uplink computer upon request. Also, the FIM can be commanded by the uplink computer to control (ON/OFF) up to 32 digital output lines. Communications with an uplink computer is via an RS232C serial I/O interface programmed for asynchronous ASCII transmission at 9600 baud. The FIM software is independent of any particular uplink computer and, with a minimum of modifications, can be utilized in a number of applications.

Acknowledgments

The author wishes to express his appreciation to Tammy Rimmer of the Mini/Microcomputer Applications Group of Section 366 for her initial efforts in developing the FIM Prototype software and related documentation.

Reference

1. Hardware Reference Manual, Pub. No. 420305521-001C, BLC 80/204 Board Level Computer, National Semiconductor Corp., April 1979.

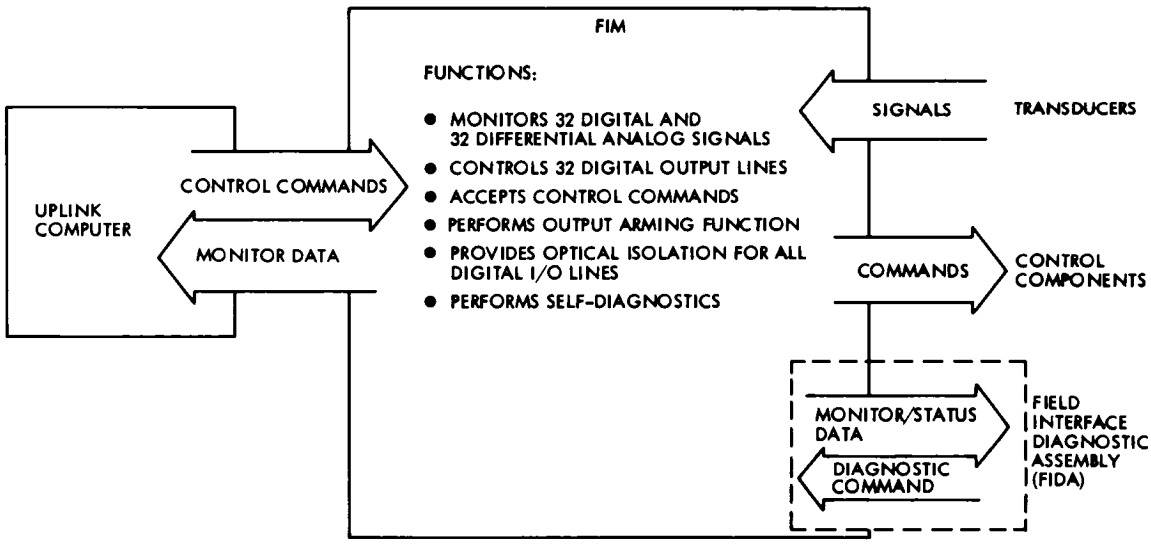


Fig. 1. FIM functions and interfaces

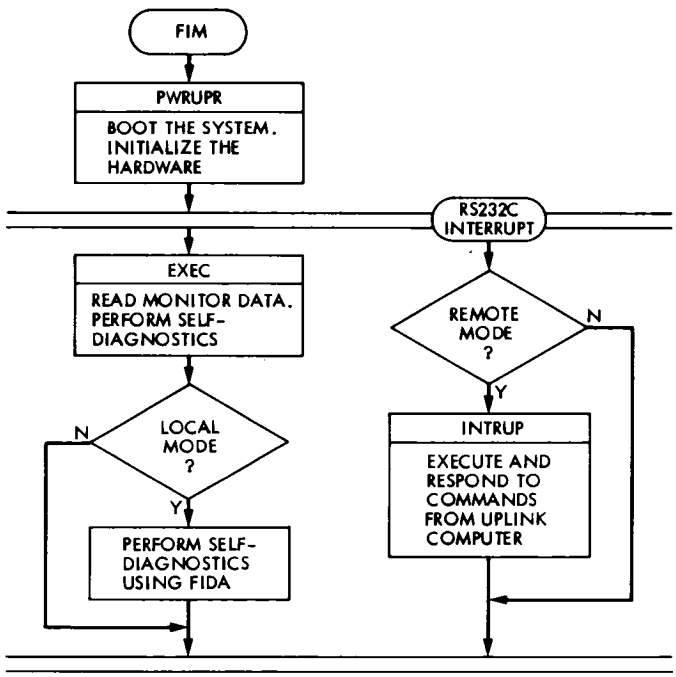


Fig. 2. Control routline flow

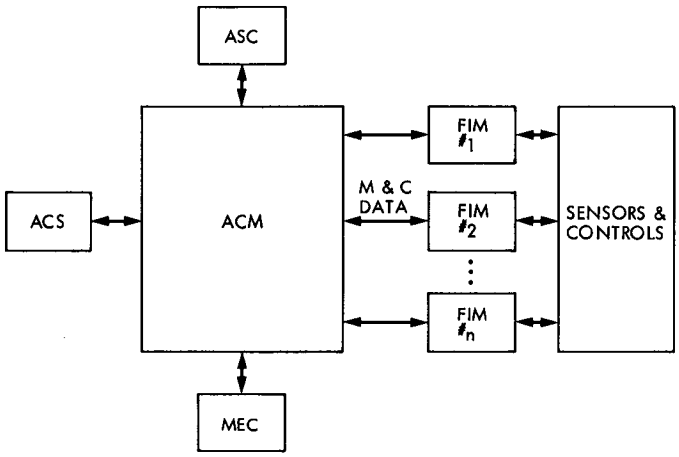


Fig. 3. ACM interface block diagram

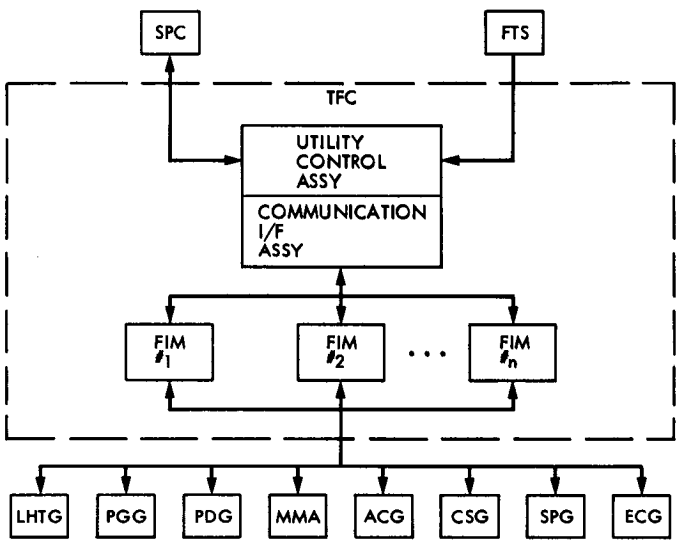


Fig. 4. TFC interface block diagram

Field Interface Module for Antenna Control Assembly

F. Baher
DSN Engineering Section

This paper describes the hardware aspects and capabilities of the Field Interface Module developed for monitor and control function in the Antenna Mechanical Subsystems of the DSN and in the Technical Facilities Controllers for the various complexes. The FIM is capable of monitoring and responding to a range of analog and digital inputs and controlling external elements. The flexibility of the design makes it applicable to other control needs, using software developed for those specific applications.

I. Introduction

The Field Interface Module (FIM) is a packaged, microcomputer-based controller designed to provide controls for and monitoring of the antenna mechanical functions of antennas in the Deep Space Net. The flexibility of the monitoring capabilities, software, and communications makes the FIM adaptable to monitor and control functions in other subsystems.

The FIM uses control and computational modules (CCM) standard components, is modular in construction, and is connectorized for quick replacement of the "Least Replaceable Element" (LRE). It is designed to minimize efforts for system hookup, diagnostics and operation.

II. General Characteristics

The FIM has the following monitor/control capabilities:

(1) Input channels for monitoring functions.
   (a) 32 differential analog input channels of which three are used for internal tests.
   (b) 16 isolated input ac binary channels.
   (c) 16 isolated dc input binary channels.

(2) Output channels for control functions.
   (a) 16 isolated ac output channels.
   (b) 16 isolated dc output channels.

(3) Communications
   (a) An RS232C serial communication link.

(4) Internal memory.
   (a) 4K RAM (on the central processor board).
   (b) 8K PROM (on the central processor board).
   (c) 16K nonvolatile memory.

A. Self-Test

Self-test diagnostics are performed for the following purposes:

(1) To validate the status of hardware unique to the FIM.

(2) To isolate or identify faults or failures to the LRE level.

(3) To verify that the H/W and application S/W of the FIM are functioning correctly.

The diagnostics are performed in two independent modes: Remote and Local. The diagnostics included at the present are the following:

(1) A checksum of the program.

(2) A/D board reference voltages (ground, 2.5 volts and 5.0 volts).

(3) A/D board channel sequence check.

(4) Parallel I/O board check.

The following tests, which are to be standardized for CCM controllers, will be included when available:

(1) CPU test.

(2) RAM integrity test.

When the FIM is in the Remote mode, diagnostics are performed as part of the FIM's operation automatically.

A Local diagnostics capability is provided to allow operation and test of the necessary functions and the communication link of the FIM. This is accomplished through a Field Interface Diagnostics Assembly (FIDA), a device that enables the FIM to perform self-diagnostics in the Local mode, independent of the uplink computer. Basically, the FIDA will allow the FIM to perform the same diagnostics that are performed while the FIM is operating in real-time, with the status displayed on the FIDA.

In order to go into Local mode, the FIM must be powered and the Local button pressed. When the FIM is in the Local mode, a message is sent to the uplink computer notifying that the state is changed and the FIM is no longer in touch with the uplink computer. The FIM will remain out of the communication link until the operator presses the Remote button.

B. Housing

The FIM assembly is contained in a Hoffman EMI/RFI shielded NEMA type 12 enclosure which is oil-tight, dust-tight, and has a strong and rigid construction. The size of the enclosure is 36 X 30 X 10 inches.

III. Architecture

The block diagram of the Field Interface Module is illustrated in Fig. 1. The primary blocks are as follows:

(1) Multibus microcomputer including card cage and circuit cards.

(2) Power supply.

(3) DC/AC interface board.

(4) AC power panel.

A. Multibus Microcomputer

1. **Card cage.** The card cage is an AM95/6440 standard cage with six multibus compatible card-slots mounted in a rigid metal enclosure. The cooling fan is mounted so that it blows air into the card cage but can also be reinstalled to function as an exhaust fan. This card cage contains the CPU board, I/O expansion board, expansion memory board and A/D converter board. As presently configured, five of the six slots are used in the final assembly, with the sixth slot reserved for a floppy disc controller used in system testing.

2. **The CPU board.** The heart of the FIM is the National Semiconductor BLC 80/204 single board computer (Intel 8080A-2). The single board computer contains 8K of RAM, 4K of ROM, Serial I/O interface, parallel I/O interface, interrupt controller, H2818 system bus controller and programmable interval timer.

3. **The I/O expansion board.** The parallel I/O expansion board uses four Intel 8255A programmable peripheral interface chips. Together, these devices provide 32-channel input and 32-channel output. This board has an arming function capability to ensure that spurious data will not get in to the data lines. This can be done by entering a specific code into the data line, which will then be compared with the data set in the comparator. If both data are the same, the output of the comparator will be enabled, which in turn will enable the arming function logic. This board is currently wire wrapped and uses two card case slots. If the demand justifies the expense, a printed circuit card will be developed and this will free one card slot for other possible uses.

4. **Expansion memory board.** The expansion memory board is an MCM-8080 microcomputer core memory module made by Ampex Corporation. It has a storage capacity of 16K words with word lengths of 8 bits and is used as expansion memory for the BLC 80/204 single board computer.

5. **The A/D converter.** Data Translation DT1742 analog-to-digital converter card is used in the FIM to provide expandable I/O capability. This board can process up to 32 differential channels of analog data at a throughput rate of 24,000 channels per second and 12 bit resolution.

B. The A/D Signal Conditioning Board

The A/D signal conditioning board is designed to interface the DT1742 A/D converter board to analog sensors (pressure transducers, temperature transducers, etc.). This board can accept up to 32 differential channels. (Three channels are used to check ground, 5 volts, and 2.5 volts reference voltage; 29 channels are available for monitoring analog signals.) This board has a built-in 2.5-volt reference diode, plus compensating circuitry which serves as the reference voltage for the A/D converter card. This voltage reference is monitored by A/D converter via channel 1 of the signal conditioning board. Channels 0 and 2 are devoted to 0 volts and +5 volts respectively.

C. The DC/AC Interface Board

The DC/AC interface board is an iCS 930 ac signal conditioning/termination board made by Intel. It can be configured as input or output for both ac and dc by the type of opto-isolator modules used on the board.

D. The AC Power Panel

The power panel consists of an ac plug, an ac lamp, a 3-A circuit breaker and a power line interference filter (EMI filter).

E. Power Supply

The power supply is a multiple output switching supply having +12, -12 V and +5, -5 V outputs.

IV. Inputs

A. Analog Monitor

The A/D signal conditioning board which is interfaced between the sensors and A/D converter board has a passive single pole filter network for each channel. The A/D converter accepts analog voltages of ±5 V full-scale. Nominal resolution of the A/D converter is 12-bits binary weighted with accuracy of +0.03% and linearity of ±1/2 least significant bit (LSB). The first three channels are used for the reference voltages (0.0 V, 2.5 V, 5.0 V).

B. DC/AC Monitor

The iCS 930 signal conditioning/termination board is used for easy configuration for dc or ac signals. It can be arranged as dc or ac by selection of the opto-isolator modules. For dc use the opto-isolators are Motorola OPTO 22 series. These opto-isolators are capable of accepting digital signals of 4.5 V to 32 V dc and dry contacts inputs. The ac modules (Motorola OPTO 22) can accept voltages of 115 V or 220 V depending on the opto-isolators used in the iCS 930 board. The isolation between input and output of the module is 2500 V rms.

C. Communications

The communication between the FIM and outside world is accomplished by utilizing a standard interface RS 232. All uplink communication will be in 7-bit ASCII and will be used with asynchronous transmission, odd parity, single start/stop bits and transmitting data at a rate of 9600 baud (10 bits per character is 960 characters/second).

V. Output

The interface between the FIM output logic and control components is accomplished by using Intel iCS 930 ac signal conditioning board and opto-isolator modules (Motorola OPTO 22).

VI. External Connections

The following external connections are provided on the FIM housings:

(1) Power connection. Power connection will be through wire in conduit off the antenna 120-V distributions system. Where desired, a power connector cord can be fitted.

(2) Communications. An RS 232C connector is provided for communication to the uplink computer.

(3) I/O ports. The following connectors are provided for I/O functions:

  (a) One 32-pin connector for ac signals at 120 or 240 V.

  (b) One 32-pin connector for ac controls at 120 or 240 V.

  (c) One 32-pin connector for 5 to 24 V dc signals going to the DC-AC interface board opto-isolators.

  (d) One 32-pin connector for 5 to 24 V dc control signals.

  (e) Two 32-pin connectors for ±5 V dc signals to the A/D processing board. One connects to channels 0-15; one connects to channels 16-31.

The I/O connections will require external junction boxes to collect cables from individual sensors or control points for

concentration and transmission to the FIM. This arrangement will vary for each application.

VII. Typical Implementations

A. 64-M Antenna Monitor and Control

In the 64-m Antenna Monitor and Control (ACM) (Fig. 2) a single microcomputer is used as a central point for monitor and control of all Antenna Mechanical Subsystem functions. A number of FIMS will be placed in specific locations on the antenna to gather data and to perform control functions. This arrangement will control the start/stop functions of the servo system precharge and high pressure pumps, the hydrostatic bearing precharge, high pressure and accumulator pumps, and the gear drive lubrication pumps. It will also monitor operational parameters for each assembly such as oil temperatures, oil levels, pressures, and motor controller status. The FIMs serve as data collection points and as points for evaluation of performance parameters in terms of standards and limits downloaded from the supervising computers.

B. Typical Hydraulic System Controller

Figures 3 and 4 show a possible application in which a FIM is used to control and monitor a typical hydraulic system, including an oil reservoir, precharge and high pressure pumps and actuators. This system can, on command, validate oil reservoir level, start the low pressure and then the high pressure pump, cause actuator movement, sense completion of actuator movements, and monitor oil pressures and temperatures and differential pressures across filters and take appropriate defensive actions as required. Note that the system as shown uses only five analog input channels, four digital binary input channels and four digital output channels. The same FIM could be used to monitor and control other similar assemblies or could be arranged to perform additional data processing of input or operational parameters, based on the specific software developed for the application.

References

1. Hardware Reference Manual, Pub. No. 420305521-001C, BLC 80/204 Board Level Computer, National Semiconductor Corp., April, 1979.
2. Hardware Reference Manual, Pub. No. 9800802A, iCS 930 AC Signal Conditioning/ Terminator Board, Intel Corp., 1978.
3. User Manual for Data Translation, DT1742 Analog to Digital Converter, Data Translation, Inc. 1978.

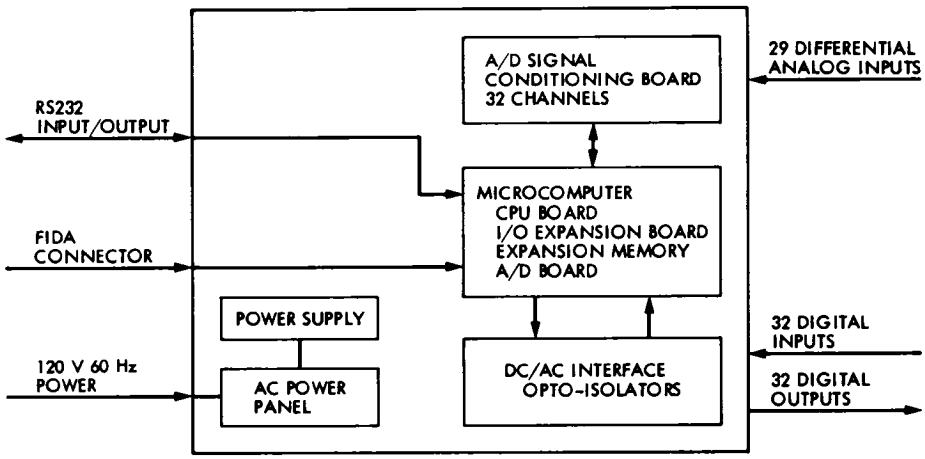


Fig. 1. Field interface module block diagram

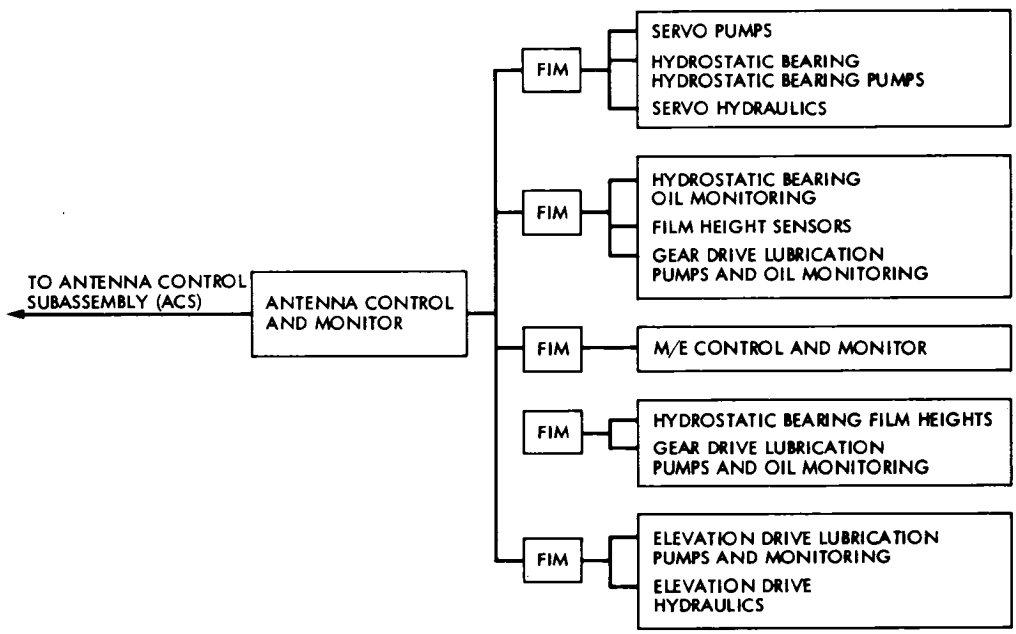


Fig. 2. 64-m antenna monitor and control block diagram (proposed)

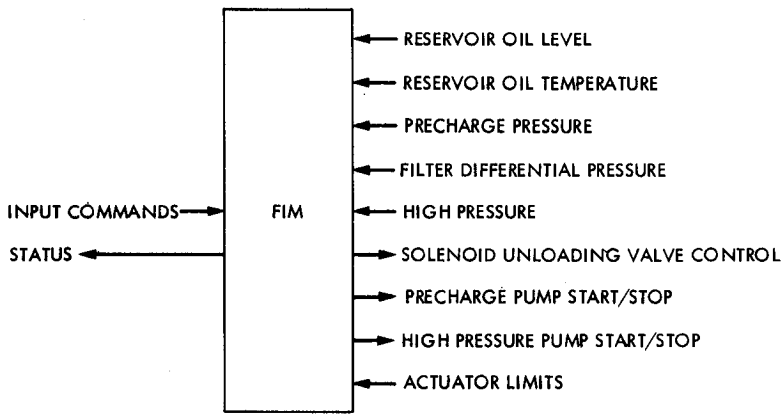


Fig. 3. Typical hydraulic system control and monitor functions

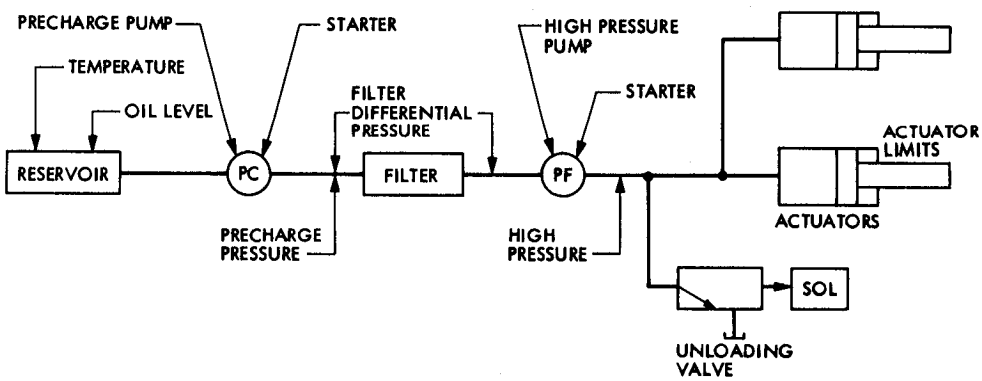


Fig. 4. Typical hydraulic system

The Development Version Control and Visibility Subsystem

L. R. Hawley
DSN Data Systems Section

This article describes a prototype Development Version Control and Visibility Subsystem (DVCS). DVCS provides an implementation/management interface serving both the implementor and management. For management, DVCS monitors the production of design and source code. DVCS provides the implementor listings annotated with change bars, detects errors in the block structure of the design and indicates when standards requiring the use of structured programming constructs in the design of software are violated. It is operated by the Software Production and Management Control (SPMC) group of the Deep Space Network at the Jet Propulsion Laboratory in Pasadena, California. The DVCS will be used by the DSN to monitor the implementation of the Network Consolidation Project, a multiyear project at JPL.

I. Introduction

Modern software production methodologies promote the idea of an all-inclusive programming environment. Such an environment provides not only the tools required by programmers to implement a given software project, but includes mechanisms to monitor and manage the data base of code and documentation produced by the implementors. With regard to these management features of a programming environment, the DVCS addresses three main areas of current concerns: (1) monitoring the output of the implementation staff, (2) adherence to standard practices, and (3) the production of high-quality as-built software specification documentation.

Within the DSN organization is a Software Production and Management Control (SPMC) group. This group is chartered to:

(1) Monitor the production of software code and documentation in terms of quantity (lines of code).

(2) Produce final versions of software documentation describing implemented code.

(3) Provide management with reports on the status of software production.

(4) Other activities not directly connected with the DVCS such as archiving release versions, distribution, etc.

The Development Version Control and Visibility System (DVCS) is intended to automate, to some extent, items (1) through (3) above. The remainder of this paper is in four sections. Section II enumerates the goals and objectives of the DVCS. Sections III and IV describe its implementation and

current use. Section V provides conclusions and indicates future plans.

II. Goals and Objectives

A brief description of the software implementation process follows; goals and objectives of DVCS are then related to the functions that occur within the software generation process. First, a software engineer analyzes a Software Requirements Document (SRD) and produces a Software Design Document (SDD) describing the architecture of a software module to implement the specified functions. After review and approval, the implementation phase begins. The major deliverables of the implementation phase are:

(1) The source, object and other (e.g., job control) code necessary to generate a working system.

(2) A Software Specification Document (SSD) describing the as-built system. Within the SSD there is a detailed design section comprised of flowcharts or flowchart equivalents describing the software. This detailed design section may comprise roughly 90% of an SSD; this is clearly an area that can benefit from automation.

(3) Other documentation related to the use, test, and interface to other software.

As major goals, DVCS will benefit the following:

(1) Implementors: DVCS will foster the production of design prior to coding. A PDL language (flowchart equivalent) will generate design and derived documentation (e.g., variable cross reference) useful in the implementation process.

(2) SPMC: DVCS will extract PDL pseudocode from files containing both PDL and assembler/compiler statements and automate the production of design documentation. DVCS will measure the progress of an implementation by monitoring and reporting lines of code, lines of PDL, changed lines, etc. In addition, DVCS will verify that the design conforms to DSN standards for operational software.

(3) Management: DVCS provides a means of measuring the progress of a project. Automation of the measuring process ensures that management receives timely information (as supplied by SPMC, which does the measuring and reporting of a project's progress).

(4) End-Users: End use functions consist mainly of operations, sustaining and maintenance. These users should receive indirect benefits resulting from the consistent and orderly production of software code and documentation. The sustaining and maintenance functions derive the same direct benefits useful to the implementors when making changes to a software module. The database of code and documentation generated in the implementation phase is available to the software personnel involved with the sustaining and maintenance functions.

A major goal in the implementation of DVCS is to make a tool that is easily operable by SPMC personnel and will not require the use of a skilled programmer for its operation. The DVCS should be adaptable to different situations and flexible in its use. Lastly, it was desired to code the DVCS in the HAL/S programming language to gain experience using the DSN's standard real-time high level language.

III. Implementation

The implementation of the DVCS software utilizes design techniques to maximize the functionality of the component parts while reducing the interfaces between modules to a minimum. Essentially, this consists of separating the requirements for DVCS into modules that do one function completely, and nothing else. The data interfaces between modules are then checked for simplicity. The design is iterated until an acceptable level of module functionality and simplicity of module interfaces is achieved. The use of "bubble charts" allows the designer to describe a program as a set of functions (inside circles known as "bubbles") interconnected by lines representing data interfaces. Figure 1 shows a high level "bubble chart" for the DVCS. In concept the "bubbles" are separate tasks that may operate in parallel in a multi-tasking operating system. In practice, such a scheme is practical only if some form of pipelining is available to implement the data interface; that is, the output of one task is the input to the next task. The concept of "pipelining" can be visualized as a mechanism by which data flows through a pipe from one task to another. Unfortunately DVCS was not implemented on a computer that provided pipelining; it is, however, a useful conceptual tool. In place of a "pipeline" the current DVCS utilizes intermediate files.

In order to achieve maximum flexibility, DVCS is designed as a set of separate functions that may be combined in a desired order by the job control language that initiates execution of tasks on the computer. The major functions incorporated into the current prototype version of DVCS are:

(1) Initialize DVCS: This function creates an empty file for the summary data produced by DVCS78.

(2) DVCS78: This is the main function provided. It inputs the current and previous versions of a specified module's source code and outputs (1) a summary of the module including lines of code, lines of comments and

lines of PDL statements, (2) a listing of the input source annotated with "!" characters to indicate changes from one version to another, (3) PDL statements for a function that validates the design adheres to standards, and (4) writes extracted PDL statements onto a file consisting solely of PDL statements for processing by a PDL processor (a commercial product).

(3) Summarize Statistics: This function creates a totals summary of all module statistics that have been written to the summary file by DVCS78. Its output is a listing used by SPMC to report weekly progress.

(4) Validate Design: This function inputs PDL statements and checks that the design does not violate standard structured programming constructs. In addition, this function checks for errors in the block structure of the design. The output of this function is a Design Structure Analysis listing of violations of design structure standards. Typical design errors detected are the use of GOTO statements, loop exits from the middle of a loop, IFs without ENDIFs, and DOs without ENDs. Currently there is only one Validate Design module; in the future there may be other validation modules for other PDL languages.

IV. Current Usage

The prototype DVCS has recently been released for use by SPMC for a three-month evaluation period. Currently, SPMC personnel are being trained in the use of DVCS and specification of the job control statements required to initiate execution of DVCS. Eventually, the production version of DVCS will be utilized by the DSN in the implementation of the Network Consolidation Project (NCP), a multiyear project currently in the design stage at JPL.

It was previously mentioned that DVCS is coded in HAL/S, the DSN's standard high-level real-time language. There was some risk involved in selecting HAL/S since the DVCS is not a real-time application and the needs of real-time programs are significantly different from a "businesslike" program such as DVCS. On the positive side, the execution speed of DVCS is extremely fast. This was to be expected since the HAL/S compiler must produce highly optimized code for real-time programs. Other features of HAL/S that benefited the design process were (1) the standard structured programming constructs are implemented as primitives of the language, (2) HAL/S provides useful data structures such as strings, records, etc., (3) variable cross reference and other information derived from the source listing, (4) formatted "pretty printed" listings using indentation to indicate levels of nesting of conditional and iterative structured programming constructs. On the negative side, HAL/S does not provide for file I/O as primitives of the language (the programmer must implement code to open, read, write, position and close a file). Taken as a whole, HAL/S was significantly superior to the alternative languages normally available on the Modcomp computer (assembly and FORTRAN IV).

V. Conclusions

As indicated in Section I, the DVCS is one component of a total programming environment serving the needs of the implementors, management staff, and subsequent operations. DVCS is, in many ways, an implementation/management interface. It monitors the production of design and source code for reporting to management. DVCS also provides the implementor with listings annotated with change bars, detects errors in the block structure of the design and indicates when DSN standards require the use of structured programming constructs in the design of software. Future plans involve making the prototype into a production version and including additional functions to:

(1) Produce a magnetic tape to interface with a high-speed laser printer on another computer.

(2) Generate a change page listing indicating which pages of PDL listings changed (as a result of changes to the PDL).

(3) Additional functions determined desirable in the current evaluation phase.

References

1. Constantine, L, and Yourdon, E., *Structured Design*, Yourdon Press, New York, 1975.
2. Stenning, V., Froggatt, T., Gilbert, R., and Thomas, E., "The Ada Programming Environment: A Perspective," *Computer Magazine*, June 1981.
3. Kernighan, B., and Masinter, L, "The Unix Programming Environment," *Computer Magazine*, April 1981.

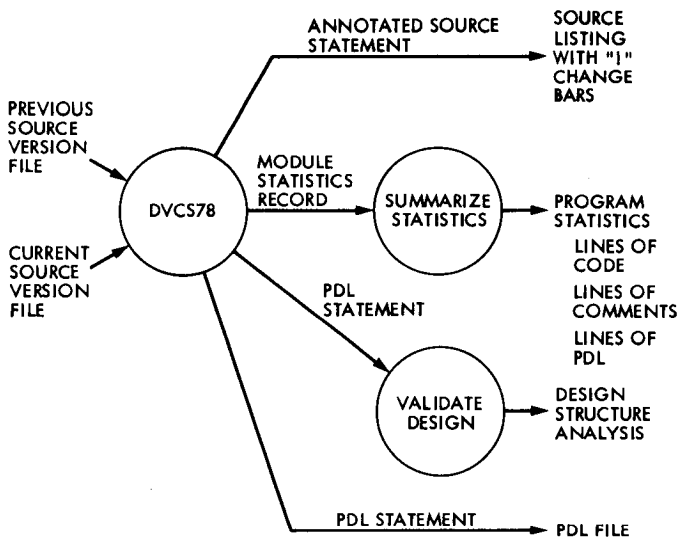


Fig. 1. DVCS bubble chart

Receiver-Exciter Controller Design

P. A. Jansma
Radio Frequency and Microwave Subsystems Section

A description of the general design of both the Block III and Block IV Receiver-Exciter Controllers for the Deep Space Network (DSN) Mark IV-A System is presented along with the design approach.

I. Introduction

Two Receiver-Exciter Controllers (RECs) are being developed as part of the DSN Mark IV-A System implementation for the Receiver-Exciter Subsystem (RCV). One controller is designed to control DSS Block III RCV hardware and is known as the Block III REC. The other controller is designed to control DSS Block IV RCV hardware and is known as the Block IV REC. The Block III REC is designed for use at the following three types of antenna sites:

(1) 9-meter track and telemetry (S-band)

(2) 34-meter track and telemetry (S- or X-band)

(3) 34-meter listen-only (S- or X-band)

The Block IV REC is designed for use at the 64-meter track and telemetry antenna sites. Associated with each Block III REC and Block IV REC is an instrumentation controller and various supporting hardware assemblies.

In accordance with established general DSN requirements, these controllers are designed to enable the Receiver-Exciter Subsystem (RCV) to be configured, calibrated, initialized and operated from a central location via high-level instructions. This is being accomplished by designing the RECs to be operated under the control of the DMC subsystem. The high-level instructions are in the form of standard subsystem blocks (SSBs) received via the local area network (LAN). The centralized control provided by these RECs and other DSCC controllers in Mark IV-A is intended to reduce DSN operations costs from the Mark III era.

II. Design Approach and Guidelines for Implementation

The design approach being used on the RECs as regards to commonality, interfaces and standards is discussed below.

The RECs are designed so as to provide maximum commonality in both hardware and software. Both RECs use a similar chassis, boards, peripherals and instrumentation controller. Both RECs share the majority of software modules. Industry standard interfaces are used except where existing hardware design prevents it and commercially available products are used extensively.

All software is designed to comply with standards established for (a) general data flow, (b) data interchange, and (c) man-machine interface. The software functional capabilities are based on the Consolidated Requirements for DSCC Subsystems Interfacing with DMC in addition to the RCV functional requirements document and general DSN requirements. The software is being developed using strict software engineering practices and a top-down approach.

III. Design

The design of both the Block III REC and the Block IV REC incorporates the following functional capabilities:

(1) RCV and REC initialization.

(2) Interfaces to the following subsystems via the LAN:

  (a) DSCC Monitor and Control Subsystem (DMC).

  (b) Antenna Pointing Subsystem (APS).

  (c) Precision Power Monitor (PPM).

  (d) Tracking Subsystem (TRK).

(3) Automated control of RCV configuration.

(4) Automated control of acquisition and tracking of downlink signal based on predicts.

(5) Automated control of frequency generation of uplink signal.

(6) Automated control of pretrack and posttrack calibration.

(7) Status monitoring of all controlled functions resulting in spontaneous generation of messages indicating alarm conditions and periodic monitor data blocks to DMC.

(8) Diagnostic procedures.

(9) Controller self-testing and monitoring.

(10) An interface to the Maintenance Support Assembly (MSA).

Details of the exact implementation of these functions will be left to future progress reports. This report will discuss only the implementation of configuration in any detail.

A. Hardware Design

1. Block IV REC. In addition to the REC itself, other special-purpose hardware is required to fully automate the Block IV RCV. Fortunately, in the Block IV RCV much of this hardware already exists. A block diagram of the Block IV REC and its peripherals is shown in Fig. 1. A brief description of each item in the block diagram follows.

The Block IV REC hardware design is based on the Computational and Control Modules (CCM) standard chassis designed by R. Reynolds. This chasis features a 12-slot multibus card cage and a power supply. The front panel contains only an ON/OFF switch, a RESET button and an RS-232C port for the MSA. The port is used only for testing and diagnostic purposes. The back panel contains ports for the following four types of interfaces: RS-232C, IEEE-488, SIA and, TTL.

A partial list of boards used within the Block IV REC includes the following:

(1) Two 8086-based CPU boards.

(2) A serial communications board.

(3) A programmable I/O board.

(4) An IEEE-488 interface board.

(5) An SIA interface board.

(6) A core memory board (nonvolatile).

(7) An expansion PROM board.

Current plans provide for 96K bytes of PROM, 32K bytes of core memory and 64K bytes of on-board RAM.

The decision to use two CPU boards within the RECs was based on several factors. First, the Exciter Synthesizer Controller (DCO) transmits data whenever it wants and it must have an "immediate listening ear" to prevent loss of data. Second, communication with the DMC subsystem and other DSCC subsystems via the LAN requires considerable processor time and high responsiveness. Lastly, too much processor loading can cause timing problems, inability to respond in a timely manner, and degradation of RCV performance. The two CPU boards have been designated as the Control Processor (CP) and the I/O Processor (IOP), respectively. The IOP serves as a "slave" to the "master" CP. The two boards communicate with each other using their respective parallel ports (8255 chips).

The Instrumentation Controller (IC) is based upon the use of commercial equipment. It uses IEEE-488 interfaces throughout. A detailed block diagram of the IC and its peripherals is shown in Fig. 2. There are three digital voltmeters (DVM's), two of which are dedicated to measuring automatic gain control (AGC) from Receiver A and Receiver B respectively. The third DVM is used together with a multiplexer to measure 60 parameters in the Block III RCV such as static phase error (SPE), dynamic phase error (DPE), dynamic AGC, power monitor signals, etc. Nine different voltages are measured for the Block IV REC and are shown in the diagram. In addition, there is one frequency counter in the IC which is used to measure receiver VCO, doppler and synthesizer frequencies, etc.

The simulation (doppler) synthesizer is a programmable synthesizer which provides reference frequency for the extraction of doppler.

The Exciter Synthesizer Controller is known as the Digitally Controlled Oscillator (DCO). The DCO is an existing piece of hardware which was designed to monitor and control a Dana Frequency Synthesizer. The DCO accepts ASCII commands across an RS-232C data link and transmits parallel BCD frequency settings to the synthesizer. The DCO is capable of performing a variety of functions such as loading a series of up to 100 frequency rate and time pairs, generating status reports, and performing diagnostic tests.

The Exciter Configuration Control and Status Assembly (CCSA) is an existing piece of hardware which monitors and controls the Block IV exciter configuration. The Exciter CCSA communicates with the REC using an SIA interface and receives commands to establish various switch settings such as exciter drive, ranging and command modulation, doppler bias, etc.

The Receiver CCSA is also an existing piece of hardware and is used to monitor and control the configuration and status of both Receiver A and Receiver B in the Block IV RCV. The Receiver CCSA also communicates with the REC using an SIA interface. The Receiver CCSA receives commands to establish various switch settings such as receiver band, loop bandwidth, AGC bandwidth, telemetry bandwidth etc. It also provides receiver lock status to the REC.

The Receiver Programmed Oscillator Control Assemblies (POCAs) are identical, existing pieces of hardware which are designed to monitor and control a Dana Frequency Synthesizer in Receiver A and Receiver B respectively. The POCAs communicate with the REC using an SIA interface and transmit parallel BCD frequency settings to the synthesizers. Each POCA is capable of storing up to four frequency rate and time pairs. It can also perform triangular sweeps when given a center frequency and corresponding upper and lower limits.

The time code translator provides a 27-bit binary data stream containing date and time from the Frequency and Timing Subsystem (FTS).

The Maintenance Support Assembly (MSA) is connected to the REC using an RS-232C interface and is used for maintenance and development purposes only. It has an ASCII keyboard and a CRT.

2. **Block III REC.** The Block III REC also requires additional special-purpose hardware to fully automate the Block III RCV. Some of this hardware is common to the Block IV REC. A block diagram of the Block III REC and its peripherals is shown in Fig. 3. A brief description of the hardware unique to the Block III REC follows.

The Block III REC also uses the CCM-standard chassis and has a front panel similar to the Block IV REC. The Block III REC back panel contains several more ports to accommodate the three types of configurations. There is extensive commonality in the use of boards within the controller with the following exceptions. The Block III REC does not require an SIA interface board but does require two additional programmable I/O boards and an optically isolated I/O board (for handling relay drivers).

The Exciter Configuration and Status Assembly (CSA) is a new piece of equipment designed to monitor and control the Block III exciter configuration. It has a special-purpose TTL interface with the REC. The exciter CSA is used to establish various switch settings such as exciter drive and ranging, test and command modulation.

Each receiver in the Block III RCV requires a receiver CSA to monitor and control its configuration. The receiver CSAs also have special-purpose TTL interfaces with the REC. The Receiver CSAs are used to establish various switch settings such as loop bandwidth, AGC bandwidth, telemetry bandwidth, etc. It also provides receiver lock status to the REC.

The S-X controller is an existing piece of hardware which is used to control various switches associated with S-band and X-band such as the S-X translator, receiver bandwidth (S or X), translator input, doppler bias, etc.

The phase shifters are unique to the 9-meter configuration of the Block III REC. They shift the phase of the reference to the angle channel phase detectors with respect to the reference (sum-channel) receiver. The REC sends an 8-bit value indicating how much shift is desired.

The output of the 10-MHz acquisition detector is a bit called acquisition trigger at zero crossing (ATZ). This bit indicates when the incoming 10-MHz IF signal is at zero-beat with the 10-MHz reference frequency. This bit is used in the REC to stop the receiver synthesizer sweep unshort the receiver loop filter and allow the receiver to be phase-locked to the signal.

B. Software Design

The software design is based on the use of two 8086-based CPU boards and a Real-Time Multi-tasking Executive. Again, the two boards have been designated as the control processor and the I/O processor, respectively. The I/O processor serves as a "slave" to the "master" control processor.

Tasks have been formulated and assigned to one or both of the boards based on ports required as well as functional

grouping. A list of tasks assigned to the control processor and a brief description of each is given in Table 1 in order of highest to lowest priority. Similar information for the I/O processor is given in Table 2, again in order of highest to lowest priority. Some details of the implementation of the configuration task are discussed later in this section.

In addition to the tasks defined in Tables 1 and 2 there are several functions which must be interrupt-driven. Interrupt service routines for these types of functions are listed below for each processor board in order of highest to lowest priority.

Control processor

Power-on, reset.

Functionally Independent Data flow Module (FIDM).

FTS date/time read.

Executive time-out timer.

I/O processor

Power-on, reset.

DCO service request.

MSA service request.

I/O executive.

Instrumentation controller communications.

The FIDM is used to implement Level 3 of the LAN protocol.

The configuration task is executed partly in the control processor and partly in the I/O processor. A functional flow for this task is given in Fig. 4. This task begins with the receipt of a controller directive SSB from the DMC subsystem containing a configuration directive. At this point, the directive must have been correctly transmitted across the LAN and received in the REC, using FIDM. The checksum for the SSB is computed and verified in FIDM. Processing continues as the SSB is decoded and the directive is parsed and validated. If the directive is valid, the directive response SSB indicates that the directive was accepted. Otherwise, the directive response indicates that the directive was rejected. All of the processing up to this point is done in the control processor. Then the specified configuration parameters are passed to the I/O processor (IOP). The IOP then builds and sends commands to the indicated hardware assembly or assemblies. The IOP reads the status of the hardware assemblies and verifies that the specified configuration has been reached. Then the IOP notifies the control processor of the status of the directive – either normal completion or error condition. Finally, the control processor builds and sends an event notification SSB to the DMC subsystem to indicate the results of the directive execution. The SSB is sent across the LAN using FIDM again, of course.

IV. Future Plans

The requirements documents have been written and the general design phase of the Block III and Block IV Receiver-Exciter Controllers is nearing completion. The detailed design phase is about to begin and the writing of the SDDs is in progress. Coding and unit testing of all software modules is scheduled for completion in January 1983. A three-phased integration testing period is planned with final delivery of the controllers including firmware and associated documentation at the end of June 1983.

Acknowledgments

The design of the Receiver-Exciter Controllers is the result of the work of several key individuals. The Block III RCV CDE is R. Weller and the Block IV RCV CDE is E. Serhal. Members of the Receiver-Exciter Controller development team whose contributions are acknowledged are J. Child, J. Estrada, G. Klein, P. Knowlton, K. Krauter, J. Nelson, and S. Ritchie.

Table 1. Tasks assigned to the control processor

Task	Description
Decode SSBs	Validate: source process code, SSB type, directive code and format (if applicable); activate appropriate task
Initialize controller	Check ports, boards, memory, registers; activate real-time executive
Monitor health	Perform continuous checkout of controller
Real-time processes (displays, MDTL)	Provide continuously updated data to all active displays and monitor data transfer lists (MDTLs)
Configure	Send configure commands to I/O processor
Acquire	Generate command stream for uplink and downlink acquisition and send to I/O processor
Display graphics	Build display graphics SSBs, activate real-time processes
Monitor data request	Send requested monitor data segment to requestor
Update data base	Place latest data from hardware assemblies into MDTL
Diagnostics	Execute requested diagnostic routine(s)
Halt/Idle	Halt controller; forward command to I/O controller

Table 2. Tasks assigned to the I/O processor

Task	Description
Halt/idle	Halt designated process or assembly
Initialize controller	Check ports, boards, memory, registers; wait for control processor commands
Initialize hardware assemblies	Place designated assembly into INIT configuration
Calibrate	Coordinate test signal and instrumentation controller to perform calibration
Monitor health	Perform periodic checks of processor and hardware assembly status
Real-time process (performance)	Check performance of subsystem (periodic)
Configure	Command and verify hardware assembly configuration
Acquire	Command and verify uplink and downlink acquisition
Diagnostics	Execute requested diagnostic routines

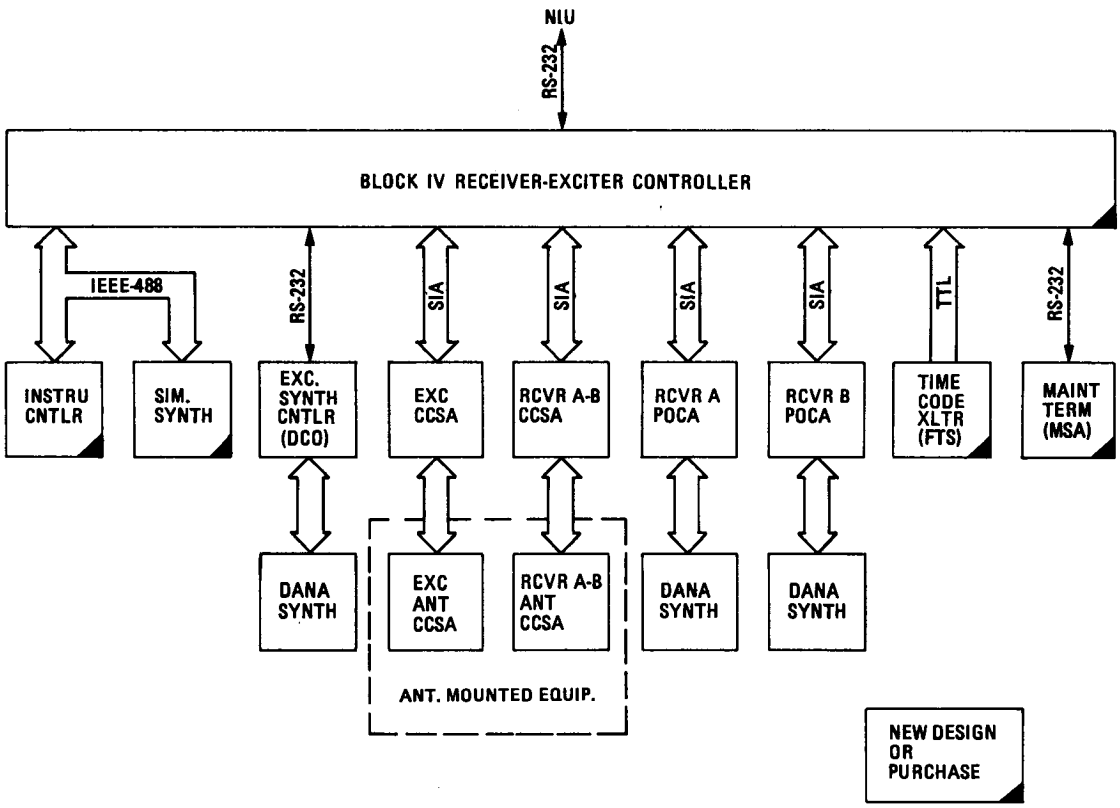


Fig. 1. Block IV Receiver-Exciter Controller block diagram

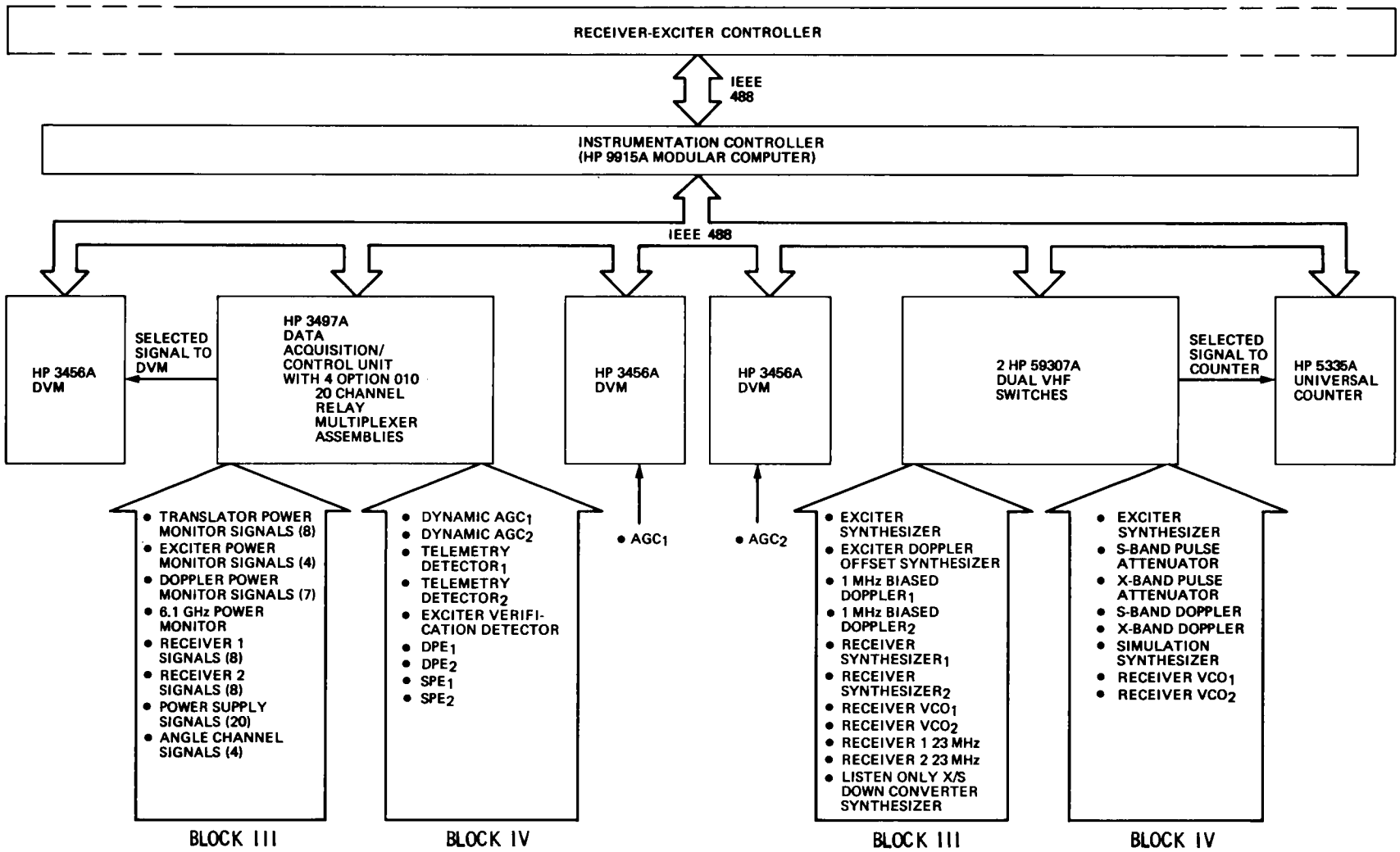


Fig. 2. Instrumentation Controller block diagram

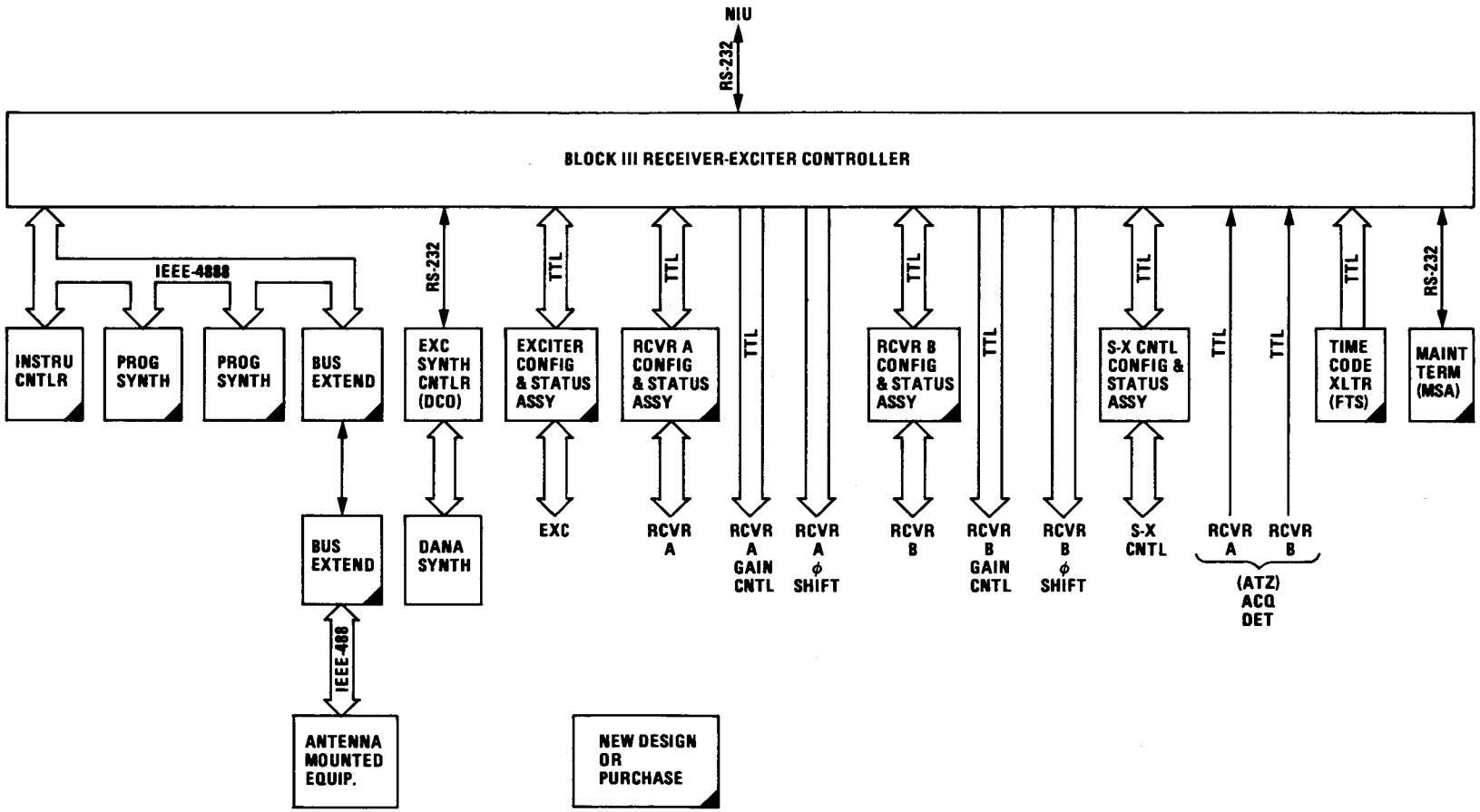


Fig. 3. Block III Receiver-Exciter Controller block diagram

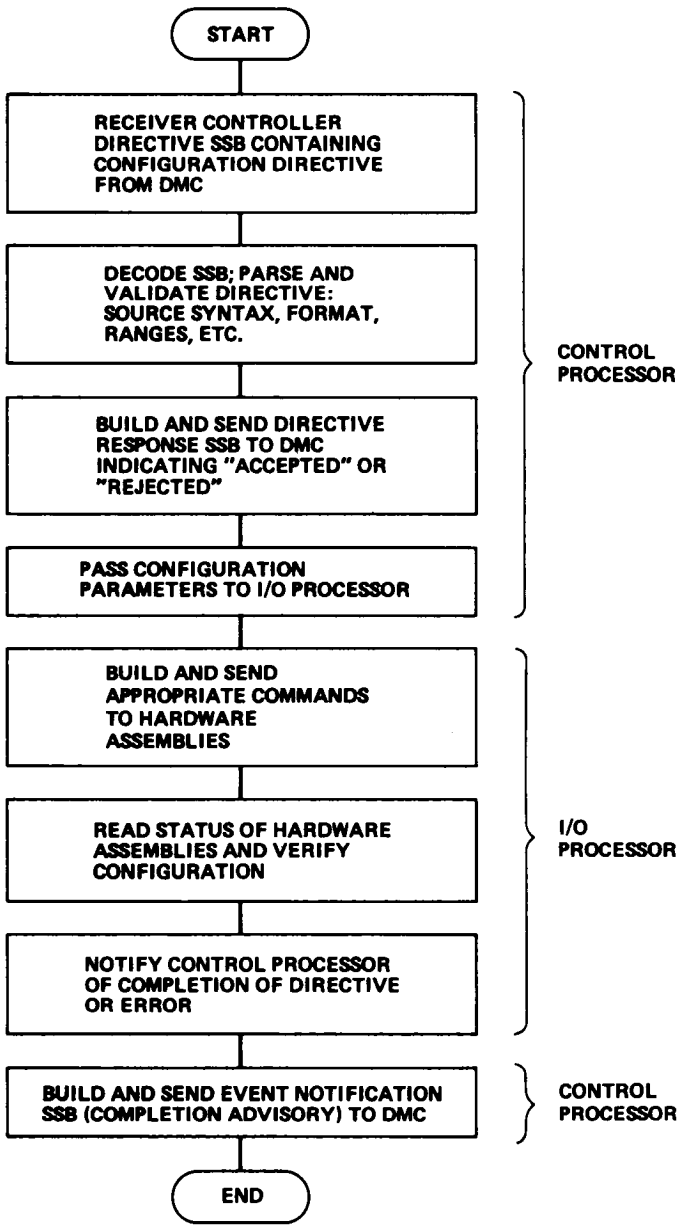


Fig. 4. Functional flow of configuration task

Analysis of Capacitive Heat Exchangers Part II

D. Schonfeld
DSN Engineering Section

The performance of a simple capacitive heat exchanger is analyzed. The shooting method for solving the governing equations is presented in detail. Simplified linear and nonlinear models are analyzed using this method and it is shown that capacitive heat exchangers are more efficient than conventional ones.

I. Introduction

Heat exchangers are important components of most industrial facilities. In the Deep Space Network they have many uses ranging from the heat exchangers that serve the hydrostatic oil bearings to those employed in the heating, ventilating and air conditioning system. Because of the large number of heat exchangers that exist throughout the DSN and because of the vital functions that they perform, it is desirable to seek improvements in their performance. To this end, the Advanced Engineering Group of the DSN Engineering Section is continuously investigating advanced concepts in heat exchangers.

The focus of one such investigation has been to analyze the theoretical performance of capacitive heat exchangers. These are two-phase, counter flow units, consisting of a stream of initially "hot" particles falling through an ascending stream of "cold" gas. The particles may be either solid or they can be liquid droplets. The potential advantages of such an arrangement are threefold. First, the two streams are in direct contact with each other. This eliminates the resistance to heat transfer which occurs when the hot and cold streams are separated by a tube wall. Second, the contact area between the two streams is greatly magnified because the heat exchange takes place on the entire surface of each particle. Finally, friction between the two streams is less than when a solid interface is present.

The equations that govern the heat transfer in such a heat exchanger have been laid out in Part I of this article (Ref. 1). It was pointed out that the equations were based on some simplifying assumptions, assumptions which were necessary because of the complexity of the physical phenomena that occur in capacitive heat exchangers. In particular it was assumed that:

(1) The flow can be considered as one-dimensional and steady.

(2) The discrete phase is composed of identical spherical particles of volume V_D.[1] The volume fraction occupied by the particles, κ, is assumed to be small and therefore the interactions between particles can be neglected.

[1] A list of symbols appears at the end of the article.

(3) The force interaction between the two streams is due to the drag between them and to gravitational and buoyancy effects. Electric and spin forces are neglected.

(4) Radiative heat transfer can be neglected, convection being the main mode of heat transfer between the spheres and the gas.

Following these assumptions, the governing laws for a capacitive heat exchanger were shown to be represented by a set of 13 equations and 13 unknowns. This set is reproduced in Table 1.

In this second part of the report, the main purpose is to present methods for solving the equations and to discuss the results obtained. The methods of solution are numerical in nature and their application necessitates a careful look at the equations. For example, these numerical methods require parameterization and taking the derivative of complicated functional relationships. Part II also presents a sensitivity analysis of the equations. In addition some clarifications are made to Part I of the paper. Finally, the performance of capacitive heat exchangers is compared to that of conventional counterflow exchangers and areas of future investigation are sketched out.

II. The Boundary Value Problem and Methods of Solution

A. The Normalized Equations

The equation set given in Table 1 consists mainly of conservation equations for mass, momentum and energy. In each of these pairs, the first equation represents conservation of the continuous phase and the second represents conservation of the dispersed phase. In addition, functional relationships are needed for the viscosity coefficient, the heat capacities of the two phases, and for the dispersed phase density. All these relationships are functions of the respective phase temperature. The state equation and two empirical correlations which are needed for the drag coefficient and for the Nusselt numbers complete the set. The last two equations are based on experiments reported in the literature (Refs. 2, 3). The initial equations chosen are:

$$N_{Nu} = (1.14 \cdot 10^{-3}) * (N_D V_D)^{-0.5984} (N_{Re})^{0.8159} \tag{1}$$

$$\overline{C_d} = 27.0 N_{Re}^{-0.84} \tag{2}$$

The coefficients in these relations, especially for the drag equation have been known to vary greatly from one investigation to another.

It was pointed out in Part I that in order to solve the system of equatations there must not be any derivative implicitness. This means that one must be able to write the equations such that only one derivative occurs in each equation. In Part I it was suggested that this can be accomplished by assuming that

$$\frac{dN_D}{dy} = \text{constant}$$

which is equvalent to

$$N_D = N_D \Big|_{y=0} \left(1 + \epsilon \frac{y}{H}\right)$$

However, this assumption is not the only one that can eliminate derivative implicitness. Instead, it is simpler to assume that ρ_D is a constant. Physically this is plausible because the density of the particle material will not change substantially within the temperature ranges of the problem. This is especially so when the particle material is a silicate such as the glass used in the model. From a mathematical point of view, the assumption that ρ_D is a constant reduces the number of unknowns and therefore the number of equations in the set. At the same time, the derivative implicitness is also eliminated.

Before the equations are solved, it is first necessary to normalize them. This is important when comparisons are made with a standard heat exchanger and when sensitivity analyses are to be performed. In these cases, normalization can unify the results by relating them to common and dimensionless scales. For the problem at hand, one must normalize with respect to known quantities. These known quantities are the values that the variables take at the boundaries of the heat exchanger. For the counter flow configuration studied here, it seems reasonable to assume that the magnitude of some of the variables is known only at one end of the heat exchanger, while the magnitude of the other variables is known at the other end. For example, it is logical to assume that one knows the entering air temperature but not the exit temperature. Conversely, the temperature of the beads may be known at the top, but not at the bottom of the heat exchanger (Fig. 1).

The normalized variables are given in Table 2. Note that in the case of velocities and temperatures, only the dispersed phase is used for normalization, even though the values that the continuous phase variables take at the bottom of the heat exchanger are also known, e.g., T_c at $y = 0$. However, in these cases, by using only the values of the dispersed phase, the normalized equations become simpler. Note also that, together with ρ_D a constant, it was also assumed that c_{p_C} and c_{p_D} are also constant; the physical justification for all these assumptions is the same. The gravitational acceleration is normalized by its value at sea level g_0, and for all practical purposes $g^* = 1$. The density and the pressure of the continuous phase can be normalized differently than shown in Table 2. For example, one can set $p^* = p/(1/2\ \rho_C v_C^2)$ and $\rho_C^* = \rho_C/\rho_D$. These formulations, when tried out, yield very complicated algebraic equations and create serious scaling problems in the boundary conditions. Scaling problems such as these can prevent the convergence of the numerical solution. Finally, by using correlations of the form

$$N_{Nu} = m\,(N_D V_D)^n\, N_{Re}^q \tag{3}$$

and

$$\overline{C_d} = a\, N_{Re}^b \tag{4}$$

one obtains the following set of dimensionless equations:

$$\frac{dv_C^*}{dy^*}\left(\rho_C^*\left(1-\sigma_1 N_D^*\right)\right)+\frac{d\rho_C^*}{dy^*}\left(v_C^*\left(1-\sigma_1 N_D^*\right)\right)-\frac{dN_D^*}{dy^*}\left(\sigma_1\,\rho_C^*\,v_C^*\right)=0 \tag{5}$$

$$v_D^*\frac{dN_D^*}{dy^*}+N_D^*\frac{dv_D^*}{dy^*}=0 \tag{6}$$

$$v_C^*\frac{dv_C^*}{dy^*}+\sigma_7\frac{dp^*}{dy^*}\left(\frac{1}{\rho_C^*\left(1-\sigma_1 N_D^*\right)}\right)=-g^*\,\sigma_2-\frac{3}{8}\,a\,\sigma_1\,\sigma_3^b\left(\frac{\rho_C^*\,|\,v_C^*-v_D^*\,|\,r_D^*}{\mu_C^*}\right)^b\frac{N_D^*}{r_D^*}(v_C^*-v_D^*)\,|\,v_C^*-v_D^*| \tag{7}$$

$$v_D^*\frac{dv_D^*}{dy^*}=\frac{3}{8}\,a\,\sigma_3^b\left[\frac{\rho_C^*\,|\,v_C^*-v_D^*\,|\,r_D^*}{\mu_C^*}\right]^b\left(1-\sigma_1 N_D^*\right)\sigma_8\,\frac{\rho_C^*\,(v_C^*-v_D^*)\,|\,v_C^*-v_D^*|}{r_D^*}-\sigma_2\left(1-\sigma_8\,\rho_C^*\right) \tag{8}$$

$$\sigma_4\,\rho_C^*\,v_C^*\,c_{P_C}^*\frac{dT_C^*}{dy^*}=\frac{3}{2}\frac{k_C^*}{r_D^{*2}}\,m\,\frac{\left(\sigma_1 N_D^*\right)^{n+1}}{1-\sigma_1 N_D^*}\,\sigma_3^q\left(\frac{\rho_C^*\,|\,v_C^*-v_D^*\,|\,r_D^*}{\mu_C^*}\right)^q\left(T_D^*-T_C^*\right) \tag{9}$$

$$\sigma_4\,v_D^*\frac{dT_D^*}{dy^*}=-\frac{3}{2}\frac{k_C^*}{r_D^{*2}}\,m\left(\sigma_1 N_D^*\right)^n\sigma_3^q\left(\frac{\rho_C^*\,|\,v_C^*-v_D^*\,|\,r_D^*}{\mu_C^*}\right)^q\left(T_C^*-T_D^*\right) \tag{10}$$

$$\sigma_9 \frac{dp^*}{dy^*} = \frac{d\rho_C^*}{dy^*} + \frac{dT_C^*}{dy^*} \tag{11}$$

$$\mu_C^* = \beta_1 T_C^* + \beta_2 \tag{12}$$

$$k_C^* = \beta_3 T_C^* + \beta_4 \tag{13}$$

The σ's are dimensionless quantities that are made up of constants and boundary values. They are defined as follows

$$\sigma_1 = \frac{4}{3} \pi r_D^{*3} H^3 N_D\big|_{y=H} \tag{14}$$

$$\sigma_2 = \frac{g_0 H}{v_D^2\big|_{y=H}} \tag{15}$$

$$\sigma_3 = \frac{2\rho_C|_{y=0}\, v_D|_{y=H} H}{\mu_C|_{y=0}} \tag{16}$$

$$\sigma_4 = \frac{\rho_C|_{y=0}\, v_D|_{y=H} c_{P_D} H}{k_C|_{T_D(H)}} \tag{17}$$

$$\sigma_7 = \frac{p_C|_{y=0}}{\rho_C|_{y=0}\, v_D^2|_{y=H}} \tag{18}$$

$$\sigma_8 = \frac{\rho_C|_{y=0}}{\rho_D} \tag{19}$$

$$\sigma_9 = \frac{\rho_C|_{y=0}}{\rho_C|_{y=0} R T_D|_{y=H}} \tag{20}$$

Two of these quantities are easily recognizable, σ_3 is in the form of a Reynolds number and σ_2 in the form of a reciprocal Froude number. The first quantity, σ_1, represents the volume fraction κ, occupied initially, i.e., at $y = H$ by the particles. σ_4 and σ_5 result from the way the variables are normalized. The last two equations, (19) and (20), are approximations to the actual functional relationships. The data for these relationships is illustrated in Figs. 2 and 3 (Ref. 4). The values for the constants are

$$\beta_1 = 2.004 \qquad \beta_2 = 0.3710$$

$$\beta_3 = 0.8994 \qquad \beta_4 = 0.1133$$

and, to simplify the notation

$$v_D^*|_{y=H} = v_D^*(H); T_D^*|_{y=H} = T_D^*(H); N_D^*|_{y=H} = N_D^*(H) \tag{21}$$

$$v_C^*|_{y=0} = v_C^*(0); T_C^*|_{y=0} = T_C^*(0); \rho_C^*|_{y=0} = \rho_C(0); \rho_C^*|_{y=0} = \rho_C(0) \tag{22}$$

B. The Shooting Method

Equations (5) through (13), together with boundary conditions (21) and (22) form a two-point boundary value problem. When dealing with a system of n first order differential equations, the classical form of a two-point boundary value problem is

$$\dot{y}_i = f_i\ (y_1, y_2, \ldots, y_n, t), \quad i = 1, 2, \ldots, n \tag{23}$$

$$y_i\,(t_0) = C_i, \quad i = 1, 2, \ldots, r \tag{24}$$

$$y_{i_m}\,(t_f) = C_{i_m}, \quad m = 1, 2, \ldots, n - r \tag{25}$$

In this system, $y_1, y_2, \ldots, y_n$ are the dependent variables and t is the independent variable; f_i can be either a linear or a nonlinear function. The boundary conditions are separated here into initial conditions (24) and final conditions (25).

For a linear system of equations, analytical solutions can be found relatively easily. Such simple systems can be written as

$$\dot{\mathbf{y}} = A\,\mathbf{y} \tag{26}$$

where $\mathbf{y}$ is the vector $y_1, y_2, \ldots, y_n$. The analytical solution consists of finding the eigenvalues and eigenvectors of A and then of constructing the matrix e^{tA}. The boundary conditions are then applied to yield the complete solution. When the equations are nonlinear, such as in the present case, the system can no longer be written as in Eq. (26) and, generally, numerical methods must be used to solve the problem. These methods are of two types: finite differences and shooting methods.

Finite difference methods basically replace the derivatives by finite differences; this has two important practical results. First, the ordinary differential equations are converted into a set of algebraic equations. Second, at the boundaries of the domain of interest, the boundary conditions are directly incorporated into the equations. Thus, finite difference methods have the advantage that they "hold" the entire solution at once. When the original differential equations are nonlinear, the resulting algebraic finite difference equations are also nonlinear. Theoretically, this does not present a problem as these algebraic equations can be solved by standard techniques such as the Newton-Raphson method. However, finite difference solutions of nonlinear boundary value problems do have certain disadvantages. The main problem with the Newton-Raphson method is to find a good starting guess. Therefore, when a reasonably good initial estimate for the y vector is not available or when the equations are strongly nonlinear, the finite differences method may not converge or may not give a unique solution (Refs. 5, 6).

An alternative to finite differences is the so-called shooting method (Refs. 6, 7, 8). The main idea behind the method is illustrated in Fig. 4. The missing conditions at one of the boundary points (Point "a" in the figure) are assumed and the equations are then integrated forward toward the other boundary point (Point "b" in the figure). When the solution reaches this "end" boundary point, it is compared to the actual conditions there, the initial guesses are adjusted and the problem is solved again. If the solution of the equation is well behaved, it is possible to use a linear interpolation for the missing initial conditions. Thus, a systematic computational scheme can be devised so that after an initial bracketing of the "target," the next "shots" land closer and closer to it until a specified tolerance is reached. Therefore the shooting method reduces the boundary value problem to a repeated iteration of an initial value problem. In general, such shooting solutions are faster than finite ones.

The method described above is called simple shooting and it too can have certain disadvantages. The main problem is that, quite often, "the differential equations are so unstable that they "blow up" before the initial value problem can be completely integrated. This can occur even in the face of extremely accurate guesses for the initial values" (Ref. 9). For such cases, an alternative procedure has been devised which is a sort of compromise between the simple shooting and the finite differences method. This is the multiple shootings method and will be the one used in the present case.

Multiple shooting is often successful in solving nonlinear problems because it limits the "blowing up" of the equations. The method starts by subdividing the integration interval by M nodes. For example, if the starting boundary point is at t_i and the end boundary point is at t_f, the new subdivisions will be at $t_1, t_2, \ldots, t_{M-1}, t_M, t_{M+1}, \ldots, t_{M-1}, t_M$, where $t_1 = t_i$ and $t_M = t_f$ (see Fig. 5). Consider now the initial value problem given by the original equation

$$\dot{\mathbf{y}} = f(\mathbf{y}, t), \mathbf{y} = (y_1, y_2, \ldots, y_N) \tag{27}$$

and by the set of initial conditions:

$$\mathbf{y}(t_K) = \mathbf{S}_K, t \in [t_K, t_{K+1}], K = 1, 2, \ldots, M-1 \tag{28}$$

Let the solution of this initial value problem be denoted by $\mathbf{y}(t_K, \mathbf{S}_K)$. Note that S_K is unknown. The original boundary value problem can be recovered if a $\mathbf{S}_K$ vector is found such that it satisfies the following two sets of conditions:

(1) Continuity conditions. For each adjacent pair of intervals, the solution must agree at their common node, t_M. These conditions can be written as:

$$F_K(S_K, S_{K+1}) = y(t_{K+1}; t_K, S_K) - S_{K+1} = 0 \quad (K = 1, 2, \ldots, M-2) \tag{29}$$

(2) Boundary conditions. At the ends of the interval the initial value problem must satisfy the boundary conditions of the original boundary value problem. These boundary conditions can be written as

$$r[y(a), y(b)] = 0 \tag{30}$$

and therefore the equations are:

$$F_{M-1}(S_1, S_{M-1}) = r[S_1, y(t_M; t_{M-1}, S_{M-1})] = 0 \tag{31}$$

The continuity and boundary conditions result in the following set of $N(M-1)$ nonlinear equations

$$F(S) = \begin{pmatrix} F_1(S_1, S_2) \\ \vdots \\ F_{M-2}(S_{M-2}, S_{M-1}) \\ F_{M-1}(S_1, S_{M-1}) \end{pmatrix} = 0 \quad ; \quad S = \begin{pmatrix} S_1 \\ \vdots \\ S_{M-1} \end{pmatrix} \tag{32}$$

These equations can be solved iteratively by the Newton-Raphson method. What has been accomplished so far is that the original problem has been broken down into a number of smaller interval problems. The advantage is that these smaller intervals "can be made short enough so that the intergration will end before any instability takes over" (Refs. 9, 11).

Stoer and Bulirsch (Ref. 10) show by means of examples that "... even for simple linear separated boundary-value problems (difference) methods and variational methods are feasible ... only if the solution need not be computed very accurately. For the

treatment of nonlinear boundary-value problems for ordinary differential equations, the only feasible methods, effectively, are the multiple shooting method and its modifications."

The multiple shooting method used here is incorporated into a computer code available from IMSL.[2] This routine is called DTPTB. The use of the Newton-Raphson method to solve (32) requires the computation of a Jacobian. In particular, the subroutine FCNJ (see Appendix A) computes $\partial Y_j'/\partial Y_i$. Because the expressions for Y_j' are algebraically quite complicated, analytical calculations of the derivatives are prone to error. The author was fortunate to have made available to him a new Fortran code, PDGEN, a partial derivative generator.[3] The input code of PDGEN consists of arithmetic assignment statements which define the functions to be differentiated. If the problem includes functions not in the PDGEN library, further defining statements must be used. For example, in the present case the function $|v_C^* - v_D^*|$ appears very often. The ABS function is not built in in PDGEN; therefore it is necessary to use

$$\text{SGNFUN (U)} = \text{SIGN (1., U)}$$

$$\text{ABS}'\text{ (Z)} = \text{SGNFUN (Z)}$$

(The first line is needed because the code does not accept a definition of a derivative as a function of two or more arguments.) PDGEN proceeds to break down the given function into temporary variables $T1, T2, \ldots$ and to give the results as

$$D(i, j) = f(T1, T2, \ldots)$$

There are some limitations on how many T statements can be generated; this was not found to be a serious impediment in the present problem.

It turns out that even the multiple shootings method requires some help in ensuring its convergence. Again, the problem has to do with initial choices: on how to choose the intermediate points t_K for the set of initial value problems (27-28). The IMSL subroutine employs a parameterization process based on introducing into the problem a parameter α. This parameter is introduced so that for $\alpha = \alpha_0 = 0$, the problem is relatively simple; usually this means a linear problem. The solution of the simpler problem furnishes an adequate guess for the next initial value problem. For this new problem, $\alpha = \alpha_0 + \epsilon$ where ϵ is a "small" number dependent on the number of iterations. The iterations procedure continues until $\alpha = 1$, at which point the original nonlinear problem has been recovered.

In the present case, the following scheme was considered.[4] Let the system of equations be written as

$$\frac{dY_i}{dy} = f_i(\mathbf{Y}) \quad , \quad \mathbf{Y} = Y_i, \ldots, Y_N \tag{33}$$

Consider a first approximation for this equation. Let this first approximation be based on expanding the right-hand side of (33) into a Taylor series and retaining only the linear terms of the series. Then this first approximation can be written as:

$$\frac{dY_i^{(0)}}{dy} = \left.\frac{\partial f}{\partial Y}\right|_{y=0} (\mathbf{Y}^{(0)}(y) - \mathbf{Y}^{(0)}(0)) + f_i(\mathbf{Y}(0)) \tag{34}$$

---

[2]IMSL Inc. (International Mathematical and Statistical Library, 7500 Bellaire Blvd., Houston, TX 77036). A library reference manual, available to IMSL subscribers, details the various subroutines used in the solution of two-point boundary value problems.

[3]The PDGEN code was designed and implemented by Dr. Webb Miller, UC Santa Barbara. The program has been implemented for JPL use by K. Stewart, E. Ng and C. Lawson for the JPL Software Development Section.

[4]The author is indebted to F. Krosh for suggesting this scheme.

where the superscript (0) denotes the "zeroth" value in a series of parameterizations. Using the parameter α, this series can be written as

$$\frac{dY_i^{(\alpha)}}{dy} = (1-\alpha)\left[\left.\frac{\partial f^{(\alpha)}}{\partial Y}\right|_{y=0}(\mathbf{Y}^{(\alpha)}(y) - \mathbf{Y}^{(\alpha)}(0)) + f_i(\mathbf{Y}^{(\alpha)}(0))\right] + \alpha f_i(\mathbf{Y}^{(\alpha)}) \tag{35}$$

where

$$\frac{\partial f^{(\alpha)}}{\partial y} = \left[(1-\alpha)\left.\frac{\partial f}{\partial Y}\right|_{y=0} + \alpha \left.\frac{\partial f}{\partial Y}\right|_{y=y}\right] \tag{36}$$

Note that when $\alpha = 0$, the first approximation (Eq. 34) is obtained, and when $\alpha = 1$, the original nonlinear problem (Eq. 33) is recovered. Therefore, for each $0 < \alpha < 1$, the code has a starting guess based on the previous α.

III. Analysis of Results for a Simplified Problem

A. Linear Case

Kays and London (Ref. 12) have introduced the concept of heat exchanger effectiveness. This is defined as the ratio between the actual heat transferred and the theoretical maximum that could be transferred. The effectiveness is a function of the hot and cold fluid capacity rates and of the "heat transfer size" of the exchanger. The heat transfer size is often labeled as the number of heat transfer units (NTU) and is given by the formula

$$NTU = \frac{1}{C_{min}}\int_0^A U\,dA$$

where U is the overall conductance and A is the same transfer area as used in the definition of U. C_{min} is the smaller of the hot and cold capacity rates, i.e., of $(\dot{m}C_P)_{hot}$ and $(\dot{m}C_P)_{cold}$ respectively. For a counterflow heat exchanger such as the one considered here:

$$\epsilon = \left\{\begin{array}{ll} \dfrac{T_{hot_{in}} - T_{hot_{out}}}{T_{hot_{in}} - T_{cold_{in}}} & \text{if } C_{hot} = C_{min} \\ \dfrac{T_{cold_{out}} - T_{cold_{in}}}{T_{hot_{in}} - T_{cold_{in}}} & \text{if } C_{cold} = C_{min} \end{array}\right\} = \frac{1 - e^{-NTU\left(1 - \frac{C_{min}}{C_{max}}\right)}}{1 - \left(\frac{C_{min}}{C_{max}}\right)e^{-NTU\left(1 - \frac{C_{min}}{C_{max}}\right)}} \tag{37}$$

Therefore, the efficiency ϵ can be readily determined from a knowledge of the temperatures at the entrance and exit of the exchanger. Indeed, if the quantities describing the capacity rates and the heat transfer coefficient are approximated as constants (i.e., not functions of temperature), only two equations are needed to solve for the efficiency; the two equations are the temperature equations ($T1.5$) and ($T1.6$) Table 1.

Let

$$C_1 \equiv \frac{h}{\rho_c v_c c_{P_c}}\left(\frac{4\pi N_D r_D^2}{1 - N_D V_D}\right) = N_{ST_c} \cdot \frac{N_D V_D}{1 - N_D V_D} \cdot \frac{3}{r_D}, (C_1 > 0) \tag{38}$$

and

$$C_2 \equiv \frac{3h}{\rho_D v_D c_{P_D} r_D} = N_{ST_D} \cdot \frac{3}{r_D}, (C_2 > 0) \tag{39}$$

where C_1 and C_2 are constants independent of temperatures. Therefore, in this case, the performance of the heat exchanger can be analyzed by solving the much simpler boundary value problem:

$$\begin{pmatrix} T_c \\ T_D \end{pmatrix} = \begin{pmatrix} -C_1 & C_1 \\ -C_2 & C_2 \end{pmatrix} \begin{pmatrix} T_c \\ T_D \end{pmatrix} \tag{40}$$

subject to the boundary conditions:

$$\left.\begin{aligned} y = 0,\ T_c &= T_c(0) \\ y = H,\ T_D &= T_D(H) \end{aligned}\right\} \tag{41}$$

Based on the eigenvalues

$$\lambda_1 = 0;\ \lambda_2 = C_2 - C_1$$

the solutions are:

$$T_c = T_c(0) - \frac{T_c(0) - T_D(H)}{1 - \frac{C_2}{C_1} e^{(C_2 - C_1)H}} \left(1 - e^{(C_2 - C_1)y}\right) \tag{42}$$

$$T_D = T_c(0) - \frac{T_c(0) - T_D(H)}{1 - \frac{C_2}{C_1} e^{(C_2 - C_1)H}} \left(1 - \frac{C_2}{C_1} e^{(C_2 - C_1)y}\right) \tag{43}$$

The physical meaning of the quantities C_1 and C_2 can be readily obtained by constructing the ratio C_2/C_1. This ratio is:

$$\frac{C_2}{C_1} = \frac{\rho_c v_c C_{P_C}}{\rho_D v_D C_{P_D}} * \frac{1 - N_D V_D}{N_D V_D} \tag{44}$$

and it represents the ratio between the heat capacities of the two streams. The magnitude of this ratio, C_{cold}/C_{hot}, can be less than, equal to or greater than 1. Changes in this ratio about a value of 1 affect the curvature of T_c and T_D variations with respect to distance. This is illustrated in Fig. 6. When $C_2/C_1 = 1$, the capacities are said to be "matched." In this case the temperatures form straight parallel lines with respect to distance; i.e., the difference $T_D - T_c$ is a constant.

It was mentioned previously that C_1 and C_2 are assumed to be independent of temperature. For the cases analyzed here, the physical properties of glass and air were evaluated at 1000 K. The performance of the heat exchanger does not change significantly if the properties are evaluated at a different temperature. This is illustrated in Fig. 7, where the axial variations of temperatures are plotted first for physical properties evaluated at 1000 K and then at 1600 K. Therefore, for the rest of the linear analysis the physical properties are kept constant and evaluated at 1000 K (Ref. 13). Furthermore, in most standard heat exchangers as well as in capacitive ones (Refs. 14, 15), the velocity ratio and the heat capacity ratio are both of order 1. For the present linear analysis the following values were chosen as fixed:

$$v_C = 2.25 \text{ m/sec} \qquad v_D = 2.0 \text{ m/sec}$$

$$c_{P_C} = 0.27 \text{ cal/g, K} \qquad c_{P_D} = 0.30 \text{ cal/g, K}$$

Under these conditions, the only change in the C_2/C_1 ratio can occur due to a change in $N_D V_D$ only. This is illustrated in Fig. 8, where $\kappa = N_D V_D$ is plotted vs C_2/C_1. From this figure, note the unique value $\bar{\kappa} = 1.6 * 10^{-4}$, which for the physical properties chosen gives the matched capacity rates condition.

To determine the efficiency ϵ of a heat exchanger, Eq. (37), the heat capacity ratio $C_R = C_{min}/C_{max}$ and the NTU must be computed. For various values of κ, this is illustrated in Fig. 9. As predicted from the previous figure, C_R reaches a value of 1.0 at $\bar{\kappa}$ and then falls off for both increasing and decreasing κ. The NTU can be computed as

$$NTU = \frac{(h)\, 4\pi r_D^2 N_D}{C_{min}} \tag{45}$$

The dependency of NTU on κ is illustrated in Fig. 9 for two typical particle radii, $r_D = 0.1$ cm and $r_D = 0.2$ cm. The shape of this curve is due mainly to the changes in the convective film coefficient h with respect to κ; for various ranges of κ one can choose from different correlations of the Nusselt number. (These correlations are given in Ref. 2). Given either C_R and NTU or C_{min}, C_{max} and the boundary temperatures, one can determine the efficiency of the heat exchanger as given in Eq. (37). The efficiency as a function of κ is plotted in Fig. 10. As expected, at $\bar{\kappa}$ the value of the efficiency is the lowest, $\epsilon = 0.79$. On either side of $\bar{\kappa}$, the efficiency rises rapidly, reaching 99% for $\kappa \geq 4 \cdot 10^{-4}$ and $\kappa \leq 2. \cdot 10^{-5}$. It is instructive to compare the capacitive heat exchanger efficiency with that of a conventional counterflow heat exchanger. One way to do this is to look at Fig. 11, which is adapted from Ref. 12. From the previous two figures, for a capacitive heat exchanger at matched capacity rates, the efficiency occurs at NTU = 1.5 and is 79%. From Fig. 11 one can see that it takes more than twice the number of NTUs (NTU = 4.0) for a conventional heat exchanger to equal this efficiency at the same capacity rate ratio. Therefore, the linear case illustrates that capacitive heat exchangers are more efficient than conventional ones.

B. Nonlinear Case

If the physical properties of the gas are strongly dependent on temperature, as they are in this case for the temperatures considered, then it is more accurate to input these relationships into the respective temperature equation. In this case, the appearance of a function of T_c in Eqs. (38) and (39) creates a nonlinear boundary value problem that can be solved by using the shooting method. In particular, the following formulations can be used for the gas

$$\left.\begin{aligned} \rho_c &= A_1 (1 + B_1 T_c) \\ \mu_c &= A_2 (1 + B_2 T_c) \\ K_c &= A_3 (1 + B_3 T_c) \\ C_{P_C} &= A_4 (1 + B_4 T_c) \end{aligned}\right\} \tag{46}$$

Some of these relationships are also incorporated in the correlation for h, as given by Eq. (3). A great deal of simplification can be obtained by keeping only the linear terms of the binomial expansion; i.e.,

$$(1 + B_i T_c)^m \cong 1 + mB_i T_c$$

After some algebra, the following set of equations is obtained:

$$\frac{dT_c}{dy} = -\xi_1 \left[\frac{1 + \xi_2 T_c + \xi_3 T_c^2}{1 + \xi_4 T_c + \xi_5 T_c^2}\right] \left(\frac{3}{r_D}\right) \left(\frac{\kappa}{1 - \kappa}\right) (T_c - T_D) \tag{47}$$

$$\frac{dT_D}{dy} = \xi_6 \left[\frac{1 + \xi_7 T_c + \xi_8 T_c^2}{1 + \xi_9 T_c}\right] \left(\frac{3}{r_D}\right) (T_D - T_c) \tag{48}$$

with the same boundary conditions as before (Eq. 41). The results appear in Fig. 12. For this nonlinear case the efficiency is 98% as compared to 92% for the linear case; this gives an indication of the increased accuracy of the nonlinear formulation.

IV. Conclusion

This report has illustrated by simple examples the nature of the performance of a capacitive heat exchanger. In particular, it was shown that capacitive heat exchangers are more efficient than conventional ones from a thermodynamic point of view. It also pointed out the important role played by the particle volumetric concentration parameter κ in describing the performance of a capacitive heat exchanger. Although the complete problem (Eqs. 5-13) was not attacked, a restricted case (Eqs. 47-48) has shown that the shooting method is capable of obtaining accurate results for such complex, nonlinear boundary value problems.

Acknowledgments

The author is grateful to F. Krogh of the Jet Propulsion Laboratory and G. Sewell of the International Mathematical and Statistical Library for many helpful discussions regarding the solution of boundary value problems by shooting methods.

References

1. Schonfeld, D. "Analysis of Capacitive Heat Exchangers. Part 1, *Telecommunications and Data Acquisition Progress Report 42-64, May-June 1981*, p. 207, Jet Propulsion Laboratory, Pasadena, Calif., Aug. 15, 1981.

2. Bandrowski, J. and Kaczmarzyk, G., "Gas-to-Particle Heat Transfer in Vertical Pneumatic Conveying of Granular Materials," *Chem. Eng. Sci., 33*, pp. 1303-1310, 1978.

3. Ingebo, R. D., "Drag Coefficients for Droplets and Solid Spheres in Clouds Accelerating in Air Streams," NACA TN 3762, 1956.

4. Svehla, R. A., "Estimated Viscosities and Thermal Conductivity of Gases at High Temperatures," NASA TR R-132, 1962.

5. Gerald, G. E., *Applied Numerical Analysis*, 2nd ed., Addison-Wesley, 1978.

6. Carnahan, B., Luther, H. A., and Wilkes, J. O., *Applied Numerical Methods*, J. Wiley, N.Y., 1969.

7. Keller, H. B., *Numerical Methods for Two-Point Boundary Value Problems*, Blaisdell, London, 1968.

8. Roberts, S. M., and Shipman, J. S., *Two-Point Boundary Value Problems, Shooting Methods*, Elsevier, N.Y., 1972.

9. Morrison, D. D., Riley, J. D., and Zancanaro, J. F., "Multiple Shootings Method for Two-Point Boundary Value Problems," *Communications of the ACM, 5*, pp. 613-614, 1962.

10. Stoer, J., and Bulirsch, R., *Introduction to Numerical Analysis*, Springer-Verlay, N.Y., 1980.

11. Diekhoff, H. J., et al., "Comparing Routines for the Numerical Solution of Initial Value Problems of Ordinary Differential Equations in Multiple Shooting," *Numer. Math.*, 27, pp. 449-469, 1977.

12. Kays, W. M., and London, A. L., *Compact Heat Exchangers*, 2nd ed., McGraw-Hill, Book Co., N.Y., 1964.

13. Handbook of Chemistry and Physics, 57th ed., R. C. Weast, ed., McGraw-Hill, N.Y., 1979.

14. Decher, R., "The Falling Bead Dry Cooling Tower," Paper 809476, 15th Intersociety Energy Conversion Engineering Conference (IECEC), 1980.

15. Shaw, D. Bruckner, A. P., and Hertzberg, A., "A New Method of Efficient Heat Transfer and Storage at Very High Temperatures," Paper 809029, 15th IECEC, 1980.

Definition of Symbols

A cross-sectional area of the conduit

c_p specific heat at constant pressure

c_v specific heat at constant volume

C_d, $\overline{C_d}$ drag coefficient

$\mathscr{D}$ kinetic energy dissipation rate

E internal energy

g gravitational acceleration

H maximum length (height) of the heat exchanger

h convective heat transfer coefficient

k conductive heat transfer coefficient

$\dot{m}$ mass flow rate

N_D number of particles per unit volume of mixture

p pressure

Q energy sources per unit volume

q heat per unit volume

R gas constant

r radius

T temperature

t time

V volume

v velocity

y axial distance for the heat exchanger

β heat transfer parameter

η loading ratio, equal to $\lambda_D v_D / \lambda_c v_c$

κ volume fraction occupied by particles

λ species density, equal to volume fraction times material density

μ coefficient of viscosity

ρ material density

τ_{ij} stress tensor

Suffixes

c pertains to continuous phase (gas)

D pertains to dispersed phase (particles)

Table 1. Summary of equations and unknowns

Equation type	Equation		Unknowns
Continuity (continuous phase)	$\frac{d}{dy}\left[(1-N_D V_D)\,\rho_c\, v_c\right] = 0$	(T1.1)	N_D, ρ_c, v_c
Continuity (dispersed phase)	$\frac{d}{dy}\left(\rho_D N_D v_D\right) = 0$	(T1.2)	ρ_D, v_D
Momentum (continuous phase)	$v_c \frac{dv_c}{dy} = -g - \frac{1}{\rho_c (1-N_D V_D)\, dy} + \frac{1}{2}\,\overline{C}_d\, v_{net}^2 N_D \pi r_D^2$	(T1.3)	$\rho, \overline{C}_d$
Momentum (dispersed phase)	$v_D \frac{dv_D}{dy} = \frac{3}{8}\,\overline{C}_d (1-N_D V_D) \frac{\rho_c}{\rho_D} \frac{v_{net}^2}{r_D} - g\left(1-\frac{\rho_c}{\rho_D}\right)$	(T1.4)	–
Energy (continuous phase)	$\rho_c v_c c_{p_c} \frac{dT_c}{dy} = +h\left(\frac{4\pi N_D r_D^2}{1-N_D V_D}\right)(T_D - T_c)$	(T1.5)	T_D, T_c, h, c_{P_c}
Energy (dispersed phase)	$\rho_D v_D c_{p_D} \frac{dT_D}{dy} = \frac{-3h}{r_D}\,(T_c - T_D)$	(T1.6)	c_{P_D}
State	$\rho = \rho_c R T_c$	(T1.7)	–
Drag coefficient correlation	$C_d = f_1\,(N_{Re})$	(T1.8)	μ_c
Nusselt nr. correlation	$h = f_2\,(N_{Re})$	(T1.9)	–
Viscosity correlation	$\mu = f_3\,(T_c)$	(T1.10)	–
c_{p_c} correlation	$c_{p_c} = f_4\,(T_c)$	(T1.11)	–
c_{p_D} correlation	$c_{p_D} = f_5\,(T_D)$	(T1.12)	–
Density correlation	$\rho_D = f_6\,(T_D)$	(T1.13)	–

Table 2. Dimensionless variables

$y^* = \frac{y}{H}$	$c_{p_c}^* = \frac{c_{p_c}}{c_{p_D}}$
$r_D^* = \frac{r_D}{H}$	$c_{p_D}^* = 1$
$v_D^* = \frac{v_D}{v_D(H)}$	$k_c^* = \frac{k_c}{k_c(H)}$
$v_c^* = \frac{v_c}{v_D(H)}$	$\rho_c^* = \frac{\rho_c}{\rho_c(0)}$
$T_c^* = \frac{T_c}{T_D(H)}$	$\rho_D^* = 1$
$T_D^* = \frac{T_D}{T_D(H)}$	$N_D^* = \frac{N_D}{N_D(H)}$
$p^* = \frac{p}{p(0)}$	

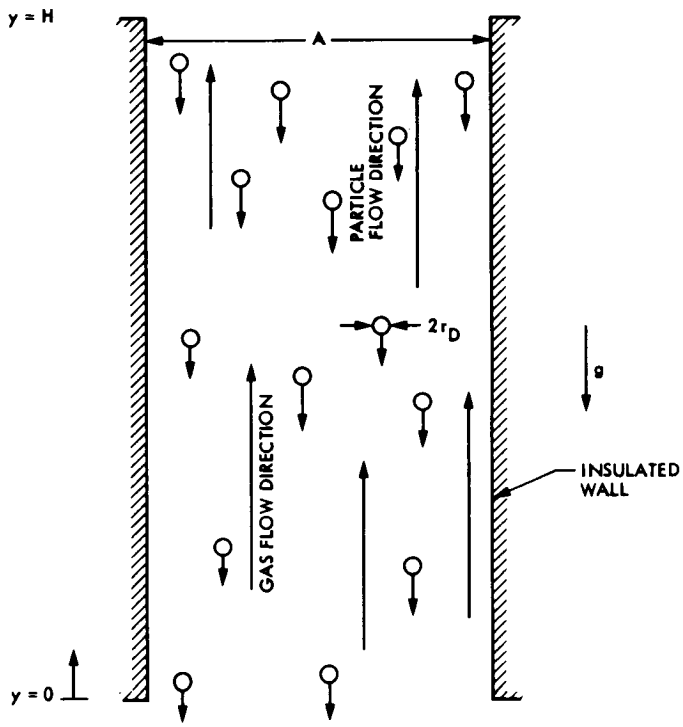


Fig. 1. Geometry of a capacitive heat exchanger

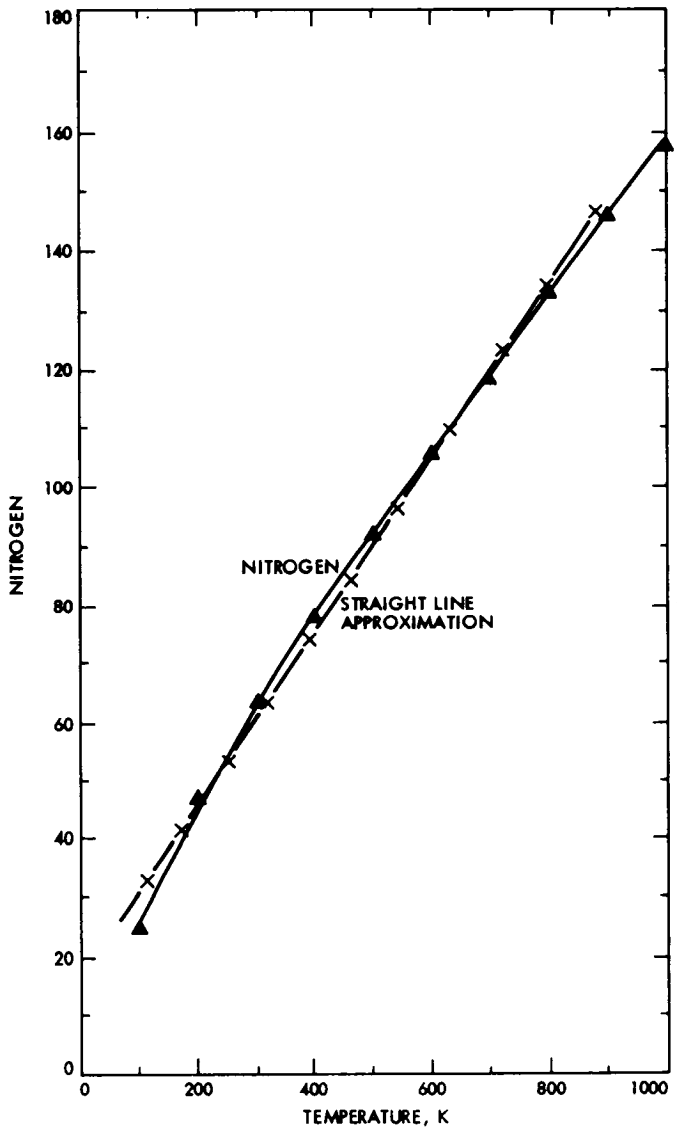


Fig. 2. Thermal conductivity coefficient vs temperature for various gases

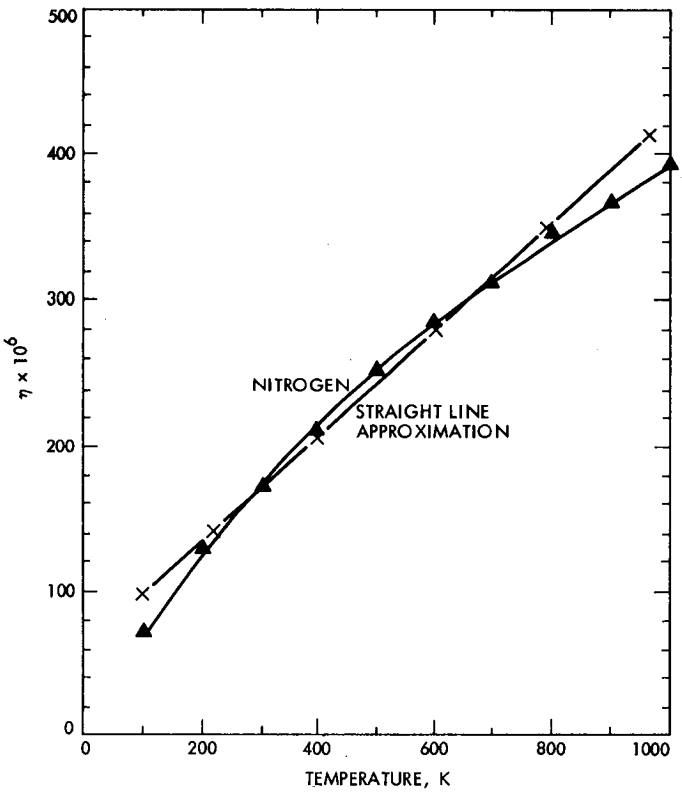


Fig. 3. Viscosity coefficient vs temperature for various gases

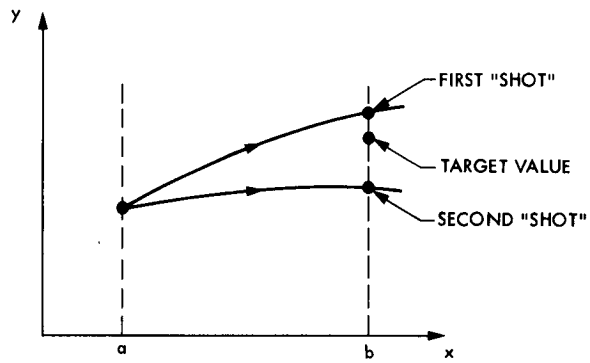


Fig. 4. General principles of the simple shooting method

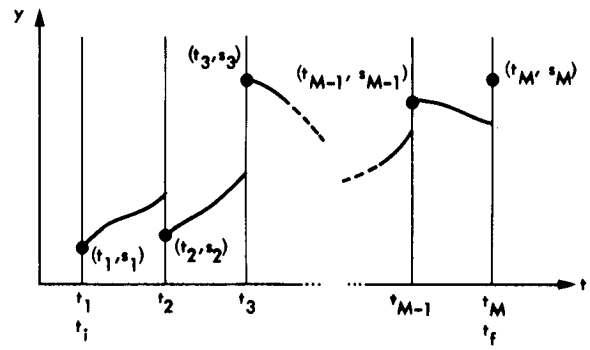


Fig. 5. Multiple shooting method

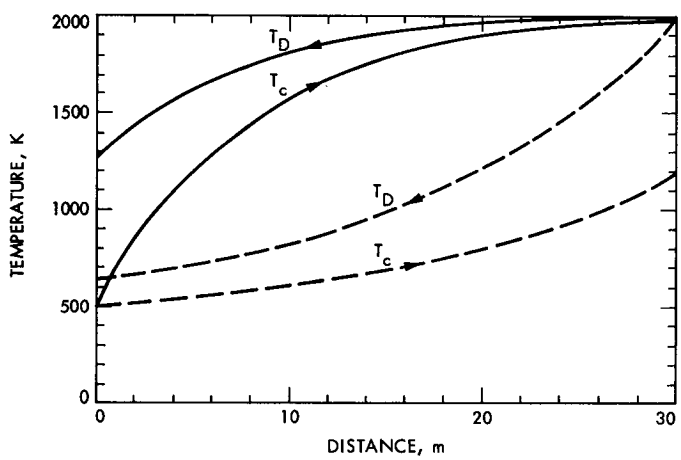


Fig. 6. Variations of temperatures with distance

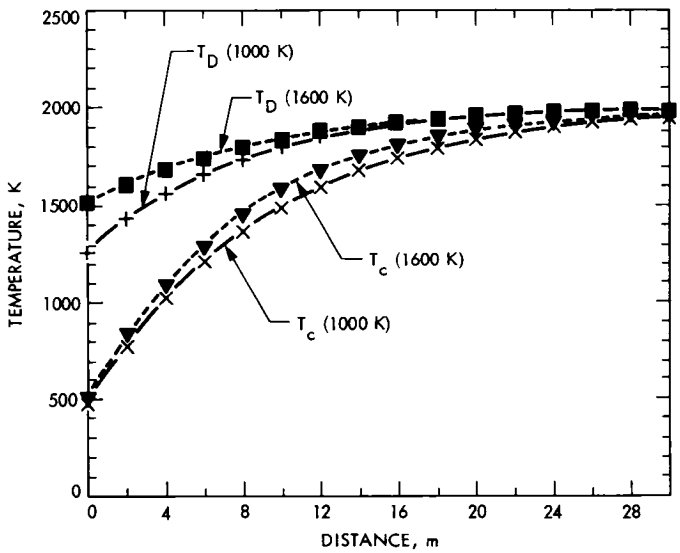


Fig. 7. Effect of physical properties and axial temperature variations

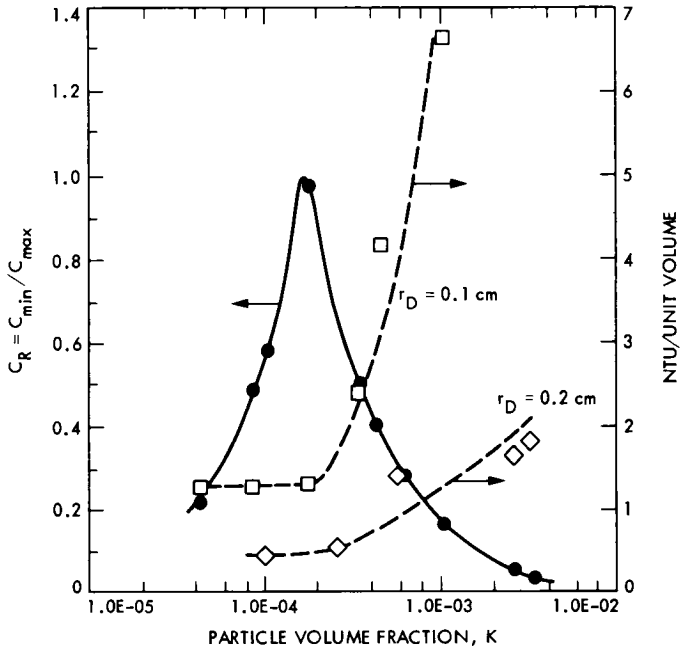


Fig. 9. C_R, NTU vs κ

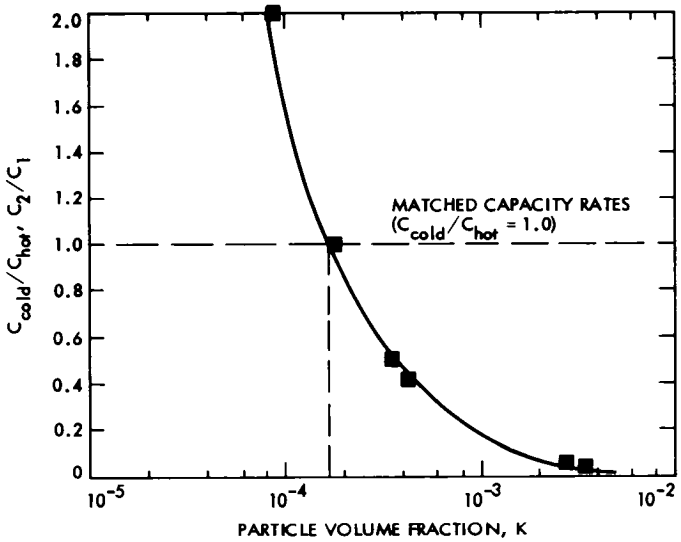


Fig. 8. C_2/C_1 vs κ

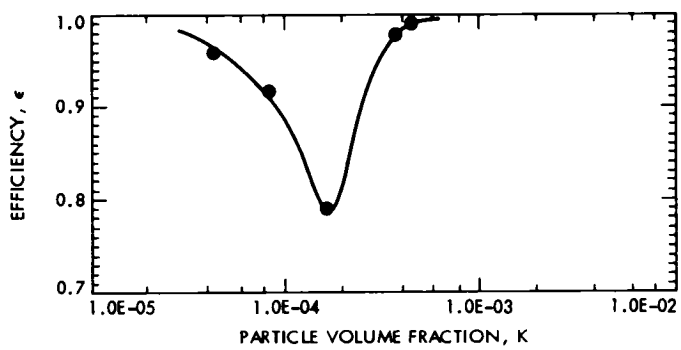


Fig. 10. ϵ vs κ

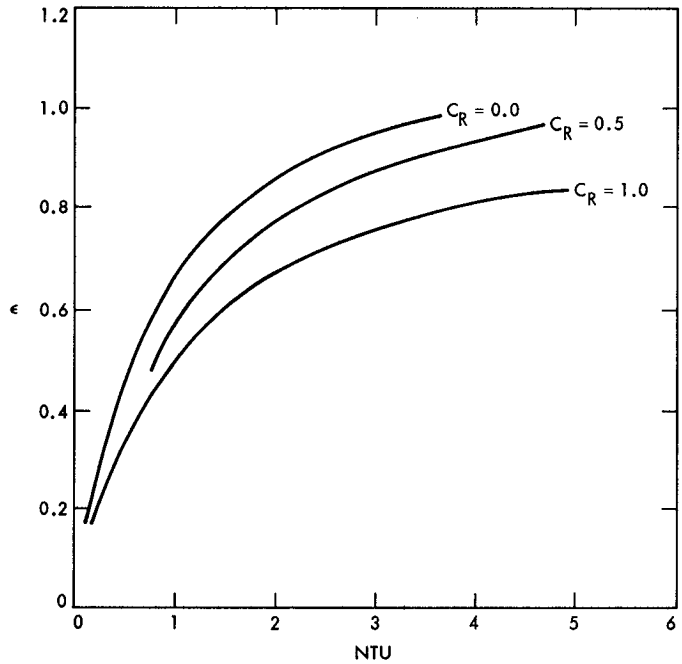


Fig. 11. ϵ-NTU relationships for a counterflow heat exchanger

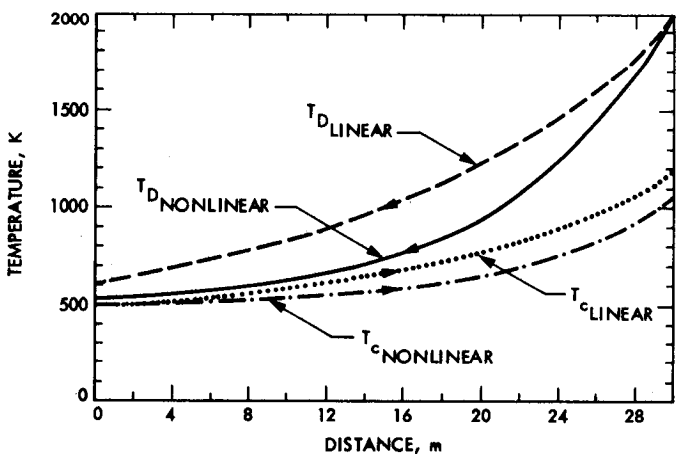


Fig. 12. Axial variations of temperatures in the nonlinear case

Review of Corrosion Causes and Corrosion Control In a Technical Facility

T. Charng and F. Lansing
DSN Engineering Section

This report summarizes a general review of causes of corrosion of metals and their alloys. The corrosion mechanism is explained using the concept of electrochemical reaction theory. The causes and methods of controlling of both physicochemical corrosion and biological corrosion are presented in detail. Factors which influence the rate of corrosion are also discussed.

I. Introduction

In a complex facility such as the Deep Space Network, the problems of corrosion and fluid conditioning have been long recognized by engineers and operation and maintenance personnel. A corrosion combatting program influences the economic aspects of the operating equipment; it helps to insure a satisfactory, efficient and continuous operation of the air conditioning system, it lowers power and maintenance costs, and increases equipment life. An extensive review of corrosion causes in metals and alloys and of methods of controlling corrosion and applications of such methods to the DSN environment is required before a successful preventive maintenance program can be started. This report reviews the corrosion field in detail and highlights key areas of concern.

Corrosion is a destructive phenomenon that, besides its economic effects, is detrimental to the appearance of metals and in some cases can cause equipment failure. It occurs in practically all environments (Refs. 1, 2). Corrosion of metals takes several forms. First, an overall surface attack slowly reduces the thickness or the weight of the metal. Second, instead of an overall surface attack, only isolated areas may be affected, producing the familiar localized corrosion. Third, it also occurs along grain boundaries or other lines of weakness because of a difference in resistance to corrosive destruction.

Metals and their alloys tend to enter into chemical union with the elements of a corrosive medium to form stable compounds similar to those found in nature. When metal loss occurs this way, the compound formed is referred to as the corrosion product. Corrosion prevention may involve the use of corrosion-resistant materials, the application of protective coatings, or control of the environment. The selection of materials or methods of protection must be determined for each environmental condition and within prescribed economic limits. Past experience and laboratory testing can serve as a guide in this selection, but exposure under actual conditions is necessary.

Since corrosion is the destruction of metal or alloy by chemical or electrochemical change, it is apparently preceded by a wide variety of mechanisms. Usually this destruction process is associated with the formation of tarnish or oxide films

when directly combined with gases or liquids. The mechanisms of corrosion attack have never been fully understood by researchers in this field. Past experiences have shown several theories to be reasonable, although without complete answers for all types of corrosion. In the literature, it is generally agreed that most corrosion mechanisms involve an electrochemical reaction.

The mechanisms of the galvanic interaction relating the corrosion of ferrous as well as nonferrous metals are discussed in Section II. The various types of metallic corrosion and biological corrosion commonly encountered together with the methods of control or prevention are presented in Sections III and IV, respectively. Physical and chemical factors that affect the corrosion process are discussed in Section V. A summary of the general guidelines on how to combat corrosion is presented at the end of this phase of the review.

II. Corrosion Mechanism

According to electrochemistry, the corrosion reaction can be considered as taking place by two simultaneous reactions: the oxidation of a metal at an anode (a corroded end releasing electrons) and the reduction of a substance at a cathode (a protected end receiving electrons). In order for the reaction to occur, the following conditions must exist:

(1) A chemical potential difference must exist between adjacent sites on a metal surface (or between alloys of a different composition).

(2) An electrolyte must be present to provide solution conductivity and as a source of material to be reduced at the cathode.

(3) An electrical path through the metal or between metals must be available to permit electron flow.

Figure 1 illustrates a typical electrochemical corrosion of iron in contact with water, which is an example case that can be used to describe the electrochemical reactions. In a nearly neutral or slightly acid environment, the water is dissociated into hydrogen ions (H^+) and hydroxyl ions (OH^-) as:

$$H_2O = H^+ + OH^- \quad (1)$$

When metal is placed in contact with a liquid, surface ionization occurs because of the electric charge difference at the solid-liquid interface. For example. iron dissolves in water in the form of positively charged ferrous ions (Fe^{++}), where

$$Fe = Fe^{++} + 2e^- \quad (2)$$

Electrochemically, a chemical substance is "oxidized" when it loses electrons to a second substance. The electrode at which oxidation takes place is called "anode." A chemical substance is "reduced" when it acquires electrons. The electrode at which reduction takes place is called "cathode." Hence, oxidation reaction results in the formation of positive charge ferrous ions at the anode. Ferrous ions moving away from the metal surface are further oxidized to ferric ions (Fe^{+++}) as follows:

$$Fe^{++} = Fe^{+++} + e^- \quad (3)$$

The positively charged ferric ions are attracted to the negatively charged hydroxyl ions and form the corrosion product $Fe(OH)_3$

$$Fe^{+++} + 3(OH)^- = Fe(OH) \quad (4)$$

The rust consists of iron hydroxide or iron oxide hydrates in various states, depending on the degree of oxidation and dehydration.

The reduction reaction at the cathode must take place concurrently in order to continue the corrosion process. Several reactions are possible and the one that occurs is determined by the environment. Without the presence of air or oxygen, hydrogen ions can be reduced by the excess of electrons at the cathode surface and evolve as molecular hydrogen by

$$2H^+ + 2e^- = H_2 \quad (5)$$

If hydrogen is not removed from the surface, the cathodic reaction decreases, and the corrosion rate is reduced. With the presence of air, the more likely reaction is the reduction of oxygen. Two possible reactions occur:

$$O_2 + 4H^+ + 4e^- = 2H_2O \quad (6)$$

$$O_2 + 2H_2O + 4e^- = 4(OH)^- \quad (7)$$

Hydrogen evolution, (Eq. 5), or oxygen reduction with the formation of water, (Eq. 6), is likely to occur in acid media. On the other hand, oxygen reduction with the formation of hydroxyl ions, (Eq. 7), is more dominant in a neutral or alkaline environment. In either case, there is an increase in the alkalinity of the solution at the cathode.

In summary, corrosion occurs when metal atoms detach themselves from the metal surface at the anode and enter the

solution as ions, leaving behind negatively charged electrons in the metal. The electrons flow through the metal to the cathode and neutralize positively charged hydrogen ions that collect at the surface. The neutral hydrogen atoms combine to form hydrogen gas. In solutions where hydrogen tends to evolve too slowly, oxygen is reduced and combines with hydrogen ions or water to form water or hydroxyl ions, respectively.

Corrosion can be attacking the overall surface or be a local phenomenon, depending on the relative proportion of anodic and cathodic areas. When the areas are approximately equal, corrosion is usually uniform over the whole surface. However, when the cathodic area is large compared to the anodic area, the localized attack at anodic sites can be intensified.

Commonly encountered corrosion may be classified into two main groups: physicochemical corrosion and biological corrosion. There are many categories of corrosion in each group. Each is associated with a specific environment. The major types of corrosion attacks, together with the causes and the methods of prevention, are discussed below.

III. Physicochemical Corrosion

The physicochemical corrosion group may be divided into the following seven categories: (a) galvanic corrosion, (b) concentration-cell corrosion, (c) pitting corrosion, (d) intergranular corrosion, (e) stress corrosion, (f) dezincification, and (g) erosion-impingement-cavitation corrosion. Physicochemical corrosion is usually caused by the combination of electrochemical attack, chemical attack, suspended solids or air bubbles, fluid velocity, solution acidity, dissolved salts, temperature, and fabrication. Each category is explained below:

A. Galvanic Corrosion

Galvanic corrosion takes place when the metallic surface is exposed to an aqueous solution which contains some aggressive ionic substances such as NaCl, $MgCl_2$, HCl, H_2S, etc.

According to electrochemistry, the galvanic corrosion reaction is caused by an anodic portion and a cathodic portion occurring simultaneously at discrete points on the metallic surface. Flow of electricity from the anodic to the cathodic areas may be generated by local sites either on a single metallic surface because of local point-to-point chemical potential differences on the surface or between dissimilar metals. The driving force to cause galvanic corrosion resulted from a difference in the electromotive force (e.m.f.). The e.m.f. (in volts) can be calculated as,

$$e.m.f. = -\frac{\Delta G}{nF} \tag{8}$$

where ΔG is the Gibbs free energy of the reaction, n is the number of moles of electrons transferred in the reaction, and F is the Faraday constant (23.06 kcal/(g-mole, volt)).

Table 1 is a list of the galvanic series of metals and alloys (Ref. 3). Metals listed toward the top of the table are more active and therefore anodic to those below; the metal higher in the list will corrode. The magnitude of the electrochemical potential difference of some metals is indicated by their relative position in Table 2. The farther apart metals are in the galvanic series, the greater will be the accelerated corrosion of the least noble metal.

The galvanic corrosion is also affected by other factors such as the distance from the bimetallic junction and the relative areas of two metals. When a bimetallic couple cannot be avoided or the metals insulated from one another, metals should be selected such that they are close to each other in the galvanic series. The smaller of the two metals should be fabricated of the more noble metal.

The anodic and cathodic areas on a metal surface are formed by such variables as inhomogeneities in metal composition, differential surface conditions, metal stresses, or variation in solution concentration as illustrated in Fig. 2. The effects of area on galvanic corrosion are very important. As the ratio of cathode-to-anode area increases, the current density at the anode increases and the cathode is more effectively depolarized. Thus, large cathode areas together with small anode areas should be avoided. Corrosion of the anode in this case may be 100 to 1000 times more than if the two areas were the same.

B. Concentration-Cell Corrosion

This type of localized corrosion is caused by variation in the concentration of environment at various locations on the metal surface such as a deficiency of oxygen, acidity change, buildup of ions, or depletion of an inhibitor. Common forms of this attack are referred to as "oxygen concentration-cell corrosion," or "crevice corrosion." The oxygen concentration cell is an electrolytic cell in which the driving force causing corrosion results from a difference in the amount of oxygen in solution at one point as compared with other locations. Crevice corrosion occurs within or adjacent to a crevice formed by contact with another piece of the same metal, another metal, or with a nonmetallic material.

Concentration-cell corrosion is associated with gaskets, joints, scale, debris, loose protective films, etc. Corrosion

attack is accelerated where the oxygen concentration is least. Metal at the area of low oxygen availability becomes anodic to other areas. Because the cathodic area is large compared to the anodic area, the intensity of attack is usually more severe than on surrounding areas of the same surface.

As mentioned above, in environments where the oxygen concentration is variable, oxygen-derived areas become anodic to oxygen-rich areas. Thus, at the air-water interface in a water tank, fully areated cathodic areas drive the corrosion of deeper, air-deprived areas. The resulting concentration cells can cause corrosion of the steel if the water level stays constant. The oxygen deprivation promotes attack in the anodic areas of the crevices.

However, in equipment handling high sodium chloride concentrate water, crevice corrosion is attributed to the combination of depletion of oxygen and the accumulation of chloride and hydrogen ions in the crevice. As the oxygen in the crevice is depleted, ferrous ions from the corrosion reaction are accumulated, and negatively charged chloride ions must enter the crevice to maintain electrical neutrality. This phenomenon accelerates the following hydrolysis,

$$Fe^{++}(C1^{-})_2 + 2H_2O = Fe(OH)_2 + 2H^{+} + 2C1^{-} \quad (9)$$

The presence of both H^{+} and $C1^{-}$ ions accelerates the corrosion attack.

C. Pitting Corrosion

Pitting is generally associated with stagnant pools of liquid. Most pits develop on horizontal steel surfaces and grow downward. Pitting occurs as small areas of localized corrosion which vary in size, frequency of occurrence, and depth. Rapid penetration of the metal may occur, leading to metal perforation. Pitting is generally not a distinctly identifiable problem for steel because the environment that induces pitting also tends to cause a rapid overall corrosion.

Pits are often initiated because of inhomogeneity of the metal surface, deposits or scale on the surface, or breaks in a protective film. These result in the development of cavities or pits. Once a pit is initiated, a concentration-cell is developed since the base of the pit is less accessible to oxygen than is the metal surface.

Halide ions such as chlorides often stimulate pitting corrosion. The mechanism is similar to that described for crevice corrosion. The depletion of oxygen in the pit slows down the generation of hydroxyl ions. The accumulation of positive charges in the form of Fe^{++} then attracts negatively charged chloride ions. The resulting ferrous chloride hydrolyzes to produce insoluble ferrous hydroxide plus excess hydrogen and chloride ions; both of these ions then accelerate the corrosion at the bottom of the pit.

Inhibitors are sometimes helpful in preventing pitting. Also, a clean, smooth metal surface will be more resistant to pitting.

D. Intergranular Corrosion

Localized corrosion at grain boundaries of the metal or alloy without appreciable attack on the grains or crystals themselves is called intergranular corrosion. It is the result of a difference in potential between anodic grain boundaries and the grains. When severe, this attack causes a loss of strength and ductility to the extent that the metal is actually destroyed by corrosion.

Intergranular corrosion can result in complete metal failure even though only a small portion of the metal is affected. This type of corrosion on stainless steel and some aluminum-copper alloys is thought to be the result of improper heat treatment which causes the precipitation of intergranular compounds. The most effective means of prevention is the proper selection of alloy and/or suitable heat treatment.

E. Stress Corrosion

This form of corrosion can be accelerated by either residual internal stress in the metal or externally applied stress. Internal stress may be produced by nonuniform deformation during cold working (bending, shearing, punching, etc.), by unequal cooling from high temperature, and by internal structure rearrangements involving volume change. Stress induced by rivets and bolts, and by press and shrink fits can also be classified as internal stresses. Stress may simply affect the corrosion behavior of the metal and, if combined with an electrochemical reaction, will produce cracks at an accelerated rate.

Cyclic or repeated stress results in a fatigue failure which is accelerated by corrosion attack. The time required to produce stress-corrosion cracking varies from minutes to years. The severity of stress corrosion for a given material depends, in part, on the degree of stress concentration and the nature of environment and temperature. Typical examples are: cracking of cold formed brass in ammonia environments, cracking of austenitic stainless steel in the presence of chlorides, cracking of monel in hydrofluosilicic acids, and the caustic embrittlement of steel.

Some instances of caustic embrittlement took place in steam boilers made from carbon steel. Sodium hydroxide added for water treatment became concentrated around the welding where there are residual tensile stresses from the

welding operation. The combination effect initiated numerous fine cracks and led to material failure and explosion.

The stress-corrosion cracking can best be avoided by using appropriate heat treatment, selecting the proper alloy for a given environment, putting the equipment in service in a stress free condition, or using suitable protective coatings. Stresses developed during fabrication, particularly during welding, are frequently the main sources of trouble; stress-relieving or annealing should always be considered.

F. Dezincification

Dezincification is recognized by the formation of a pronounced copper color, rather than the yellow color of brass, of copper-zinc alloys. The corrosion may occur as plug filling pits or as continuous layers surrounding an unattacked core of brass. The plug-type dezincification occurs more frequently on low-zinc brasses and the uniform type on high-zinc brasses. The mechanism may involve selective removal of zinc, leaving the copper behind, or dissolution of the alloy followed by redeposition of the copper from the corrosion products.

The tendency to dezincification can be reduced by the addition of small amounts of arsenic, antimony, or phosphorous to the alloy. Rate of attack generally increases with an increase in temperature, increase in solution conductivity, or decrease in solution flow.

This category of corrosion is similar to graphitization of cast iron where iron is removed, leaving a graphite mass with a porous structure. Also, similar types of corrosion have been observed for other alloys, and the tendency is to identify them by similar nomenclature, e.g., dealuminumification, denickelification, etc.

G. Erosion-Impingement-Cavitation

Erosion is the destruction of a metal by the combined action of corrosion and abrasion or attrition resulting from the flow of liquid or gas. When the fluid contains solid particles that are harder than the metal surface affected, erosion will occur by the combined action of corrosion and abrasion. When the fluid does not contain suspended solids, or contains particles that are softer than the metal, erosion will be by corrosion and attrition. Visual evidence of such attack usually takes the form of directional grooves, gullies, waves, rounded holes, or valleys.

Impingement attack is corrosion associated with turbulent flow of a liquid such as at the entrance of a condenser tube or around bends in a pipeline. The high velocity impingement of a liquid can strip away the metal surface's protective film. Solids and gas bubbles can aggravate impingement attack.

Air bubbles in a liquid stream are especially harmful. Formation of transient voids or vacuum bubbles in a liquid stream passing over a surface is referred to as cavitation. This is often encountered around propellers, in pumps and agitators. Cavitation corrosion is often blamed for the damage resulting from the collapse of such vacuum bubbles. More correctly, this should be called cavitation erosion. When bubbles collapse on a metal surface, there is a severe impact or explosive effect that can cause considerable mechanical damage, and corrosion can be greatly accelerated because of the destruction of protective films.

Harder materials, lower velocity, streamlined flowpaths, extra thickness for vulnerable areas, or replaceable impingement plates are general methods used to avert equipment failures due to erosion-impingement attack.

IV. Biological Corrosion

Fungi, algae, and bacteria are the major three causes of biological corrosion which can cause rapid and severe wood rot and metal corrosion. Some important types of these harmful organisms together with methods of prevention and/or control (Refs. 4, 5) are reviewed as follows:

A. Fungi

Fungi are yeasts and molds. Fungal growths are found on cooling-tower wood, basin walls, and in heat exchangers. In wood decay, the cellulose in wood is consumed by organisms until it loses its strength. Fungi are not directly corrosive to metals. However, the deposits of fungi on metal surfaces produce differential concentration cells and interfere with the action of corrosion inhibitors by shielding the metal surfaces from it, causing corrosion. Wood deterioration can be prevented by impregnation with toxic salts that inhibit fungal growth. Fungi can also be controlled by the periodic application of fungicides such as pentachlorophenol salts or tributyl tin compounds. Chlorine is not effective against these organisms.

B. Algae

Algae, like fungi, are relatively large organisms. Algae commonly cause slimy deposits in cooling towers where sunlight and water are present. Deposits of dead algae provide food for bacteria and fungi. Algae are not known to cause corrosion directly, except for occasional occurrences under their deposits. Control can be effected by covering cooling-tower decks to prevent sunlight from reaching the tower water, or with chemicals such as chlorine, quaternary ammonium compounds and copper salts.

C. Bacteria

Bacteria associated with corrosion are of two types: aerobic and anaerobic. Aerobic microorganisms readily grow in an environment containing oxygen, while anaerobic species thrive in an environment virtually devoid of atmospheric oxygen. Each type has a specific action and often is referred to by its effect on materials. Deterioration and corrosion of materials due to the metabolic activity of microorganisms is quite complicated and in some cases not fully understood. Corrosion can be explained by any one or more of the following: (1) producing a corrosive environment, (2) creating electrolytic concentration cells on the metal surface, (3) alternating the surface protecting film, (4) influencing the rate of anodic-cathodic reaction, and (5) changing the environment composition.

Sulfate-reducing bacteria convert water-soluble sulfur compounds to hydrogen sulfide in a slightly acid to alkaline environment. Corrosion of mild steel, stainless steel, and copper alloys is due to conversion of iron to iron sulfide. Nickel and nickel-based alloys are severely pitted under a combination of low pH, sulfides, and reducing conditions. Using chlorine to control the bacteria is not effective because the microorganisms are usually covered by slime masses that prevent the chlorine from reaching the bacteria. In addition, the sulfides surrounding these microorganisms react with chlorine to form chloride salts that negate the effectiveness of chlorine. Special toxicants such as long-chain fatty-acid amine salts and organic-sulfur compounds (for example, methylene bisthiocynate) are effective in controlling these bacteria.

Another group of microorganisms are the nitrifying (or acid-producing) bacteria, which produce nitric acid from ammonia. This results in corrosion of mild steel, copper, and aluminum by chemical corrosion in low pH conditions. These microorganisms are not affected by oxygen and do not neutralize corrosion inhibitors such as chromate or zinc. Chlorine, as well as many nonoxidizing biocides, is very effective in controlling these bacteria. However, with any appreciable amount of ammonia, chlorine may not appear effective; chlorine is neutralized and is unavailable for bacteria control.

A third group of these corrosive microorganisms converts soluble ferrous iron salts into insoluble ferrous oxide. These iron deposits shield metal surfaces from corrosion inhibitors, also promoting corrosion because of concentration-cell. Because the bacteria remove the iron from the area of corrosion, the reaction is accelerated. The iron-depositing bacteria can be controlled with chlorine or many nonoxidizing biocides such as quaternary ammonium compounds.

A fourth group of bacteria feeds on nitrites used as corrosion inhibitors. The growth of this particular group of microorganisms must be kept under control when sodium nitrite inhibitor is used.

Chemical treatment, rather than removal by mechanical cleaning, is a more satisfactory method of combating organisms because of the inaccessibility of corroded areas in some equipment. Fungi and bacteria live in dark areas such as in heat exchangers and piping, as well as in light areas such as in cooling towers or evaporative condensers. If slime and algae are allowed to form an appreciable deposit, it should be removed where practical by mechanical means, and the system should be flushed before chemical treatment.

Chemical treatment using chlorinated phenols, copper sulfate, potassium permanganate, and chlorine has proved to be effective in controlling biological corrosion. Frequently, microorganisms can build up an immunity to a particular algaecide, but not to chlorine. This characteristic makes it necessary to switch to other chemical compounds periodically.

V. Corrosion Parameters

Atmospheric corrosion of carbon steel proceeds at rates up to 0.1 mm per year in environments free of strong chemical splash, spillage or fumes. This corrosion resulted from condensing moisture, sulfur dioxide from fuel combustion, dust bearing corrosives, and the remoteness of structures and equipment from the washing effects of rain water. There is often rapid and severe corrosion that cannot be stopped even when inhibitors are used. Such corrosion may be caused by microbiological organisms. Solution acidity, oxidizing agents, temperature, film deposition, dissolved salts, fluid velocity, and impurities are some of the major corrosion contributors. The seven major corrosion parameters are listed as: (a) solution acidity, (b) oxidizing agents, (c) temperature, (d) films, (e) dissolved salts, (f) fluid velocity, and (g) impurities. Each parameter is explained as follows:

A. Solution Acidity

Solution acidity is represented by the concentration of hydrogen ions with the relation:

$$pH = -\log [H^+] \qquad (10)$$

The tendency for metals to corrode by displacing hydrogen ions from solutions is indicated in the electromotive series such as Table 2 (Ref. 3). Metals above hydrogen in the series displace hydrogen more readily than do those below hydrogen. Whether or not hydrogen evolution will occur is affected by additional factors such as the rate of the corrosion reaction.

Since the discharge of hydrogen ions takes place in most corrosion reactions, acidity of a solution is one of the most important factors in corrosion combatting. A survey of 944 cases involving carbon steel (Ref. 6) showed that 71 cases are related to corrosive acids. As a general rule, acid ($pH < 7$) solutions are more corrosive than neutral ($pH = 7$) solutions or alkaline ($pH > 7$) solutions. In the case of ordinary iron or steel, the dividing line between rapid corrosion in neutral or alkaline solutions occurs at about $pH = 4.5$ (Refs. 2, 10). For example, Fig. 3 shows the corrosion rate of carbon steel in low-velocity room-temperature water with pH range 4 to 10 is about 0.3 mm per year. In the acidic environment (HCl addition) whereof $pH = 2.9$, the corrosion rate is above 0.8 mm per year. In an alkaline environment ($pH > 10$), the corrosion of carbon steel is below 0.3 mm per year. Exceptions are the amphoteric metals such as aluminum and zinc in highly alkaline solutions which cause even more corrosion than acid solutions.

B. Oxidizing Agents

In some corrosion processes, such as the solution of zinc in hydrochloric acid, hydrogen evolves as gas. In the solution of copper in sodium chloride, the removal of hydrogen is effected by a reaction between hydrogen and some oxidizing chemical, such as oxygen, to form water. For this reason, oxidizing agents are often powerful accelerators of corrosion. In many cases, the oxidizing power of a solution is its most important single property.

Oxidizing agents may accelerate the corrosion of one class of materials and retard the corrosion of another class. In the latter case, the behavior of the material usually represents the surface formation of oxides or layers of absorbed oxygen which make the material more resistant to further chemical attacks. This property of chromium is responsible for the principal corrosion-resisting characteristics of stainless steels.

At room temperature, the corrosion rates of carbon steel in a slowly moving, air-saturated (dissolved oxygen 6 ml/l) water containing 165 parts per million (ppm) of $CaCl_2$ range between a negligible amount to about 0.5 mm per year (Ref. 7). The corrosion rate is almost linearly proportional to the concentration of dissolved oxygen. The effect of oxygen concentration on the corrosion of carbon steel in slowly moving distilled water at 25°C is shown in Fig. 4. As indicated in Fig. 4, the corrosion rate decreases as the oxygen level is raised from 12 to 25 ml/l. This is due to passivation formation of a protective oxide film or the chemisorption of excess oxygen on iron. However, the decrease in corrosion rate at high oxygen content will not occur with the presence of appreciable chloride content. The protective film breaks down locally in the presence of chloride. Also, oxygen-lean areas occurring in crevices cannot be protected.

Destructive effects of high oxygen levels may justify deaeration to lessen the rate of corrosion. In general, the expected rate of attack for air-saturated water at low fluid velocities and ambient temperature is about 0.3 mm per year.

C. Temperature

Rate of corrosion tends to increase with rising temperature. Higher temperatures accelerate the diffusion of oxygen through cathodic layers of protective oxide film. Temperature also has a secondary effect through its influence on the solubility of air (or oxygen).

Experimental results indicate that a temperature rise of 18 to 20°C will double the corrosion rate (Ref. 8). Figure 5 presents the influence of temperature on corrosion of iron in water containing dissolved oxygen (Ref. 9). In a closed system where the oxygen cannot escape the corrosion rate increases with temperature until all of the oxygen is consumed. In an open system where the oxygen is free to escape the corrosion rate increases with temperature up to about 80°C. Further increase in temperature results in a marked decrease in corrosion. This is due to a drop in the oxygen solubility in water above 80°C. On the other hand, the corrosion rate of stainless steels will increase considerably through the loss of the oxidizing substance (dissolved oxygen) which is essential to maintain its protective film.

D. Films

There are films of metal oxide, oil and grease that may protect a material from direct contact with corrosive substances. Such oil films may be applied intentionally or may occur naturally as in the case of metals submerged in sewage or equipment used for the processing of oily substances. Once corrosion starts, its further progress often is controlled by the nature of films that may form or accumulate on the metallic surface. One common example is the $PbSO_4$ film on the lead container in contact with sulfuric acid. Another example is the thin oxide film that forms on stainless steel surface.

Insoluble corrosion products may be completely impervious to the corroding environment, hence completely protective, or they may be permeable and allow local or general corrosion to proceed unhindered. Nonuniform or discontinuous film may tend to localize corrosion at certain points by initiating electrolytic effects of the concentration-cell type. Films may tend to retain or absorb moisture and thus, by delaying the time of drying, increase the extent of corrosion resulting from exposure to the atmosphere or to corrosive vapors. It is generally agreed that the rust films formed on low-alloyed steels are more protective than those formed on unalloyed steels.

E. Dissolved Salts

A survey of 180 inorganic salt-aqueous systems (Ref. 6) has indicated that 51% of the salts are corrosive to carbon steel at rates greater than 1.3 mm per year. Acid salts, such as aluminum chloride, ferrous chloride, ammonium chloride, etc., hydrolyze to form acid solutions. Acid salts have a low pH, which will accelerate corrosion simply because of their acidic effect.

Alkaline salts hydrolyze to increase solution pH that may sometimes act as corrosion inhibitors. Examples of these salts are trisodium phosphate, sodium tetraborate, sodium silicate, and sodium carbonate.

Oxidizing salts such as ferric chloride, cupric chloride, and sodium hypochlorite are especially corrosive to carbon steel. Examples of oxidizing salts that are inhibitors include Na_2CrO_4, $NaNO_2$, and $KMnO_4$.

Hard water is less corrosive than soft water. Deposition of calcium carbonate provides a protective film which retards corrosion by shielding oxygen from the cathodic areas. However, protection by $CaCO_3$ precipitation may prove undesirable or unfeasible, since it can clog equipment or reduce heat transfer.

In summary, the presence of acid or neutral salts may increase corrosion rate; whereas the presence of alkaline salts may decrease the corrosion rate.

F. Fluid Velocity

An increase in the relative velocity between a corrosive fluid and a metallic surface tends to accelerate the corrosion rate. This effect is due to the higher rate at which corrosive chemicals, including oxidizing substances such as air, are brought to the corroding surface. Whereas, corrosion resistance results from the accumulation of layers of insoluble corrosion products on the metallic surface, the effect of high velocity may be either to prevent their normal formation, or to remove them after they are formed. The higher the velocity, the thinner will be the films through which corroding substances must penetrate, and through which soluble corrosion products must diffuse. Either effect allows corrosion to proceed unhindered. Similar effects are associated with cavitation-erosion corrosion.

Corrosion rate influenced by fluid velocity occurs frequently in small-diameter tubes or pipes at high velocity such as in condenser and evaporation tubes, in the vicinity of bends and contraction of pipelines, on propellers of agitators, and in centrifugal pumps.

Velocity effects on corrosion rate of carbon steel in natural water, as shown in Fig. 6, can be divided into four ranges according to the velocity magnitude (Ref. 6):

(1) Slight motion (less than 0.3 m/sec) may stop localized attack such as pitting.

(2) At about 0.3 m/sec, the flow may increase the oxygen supply to a level that raises the corrosion rate to as much as 1.0 mm per year.

(3) At a velocity range from 2.4 to 3.0 m/sec, corrosion rate will be about 0.3 to 0.8 mm per year, depending on the surface roughness.

(4) At velocities over 4.5 m/sec, turbulence may greatly accelerate the corrosion rate up to 5 mm per year.

G. Impurities

Impurities in a corrodent can be good or bad. The chloride ion is a good example; the presence of a small amount of chloride in a fluid can break down the passive oxide film on stainless steels. Some impurities may act as inhibitors to retard corrosion. For instance, inorganic oxidizers such as chromates are used as a corrosion inhibitor in cooling water systems. However, if the impurity is removed, a marked increase in corrosion rates may result.

The effects of impurities are varied and complex. One should be aware of the type, quantity, causes, and location before implementation.

VI. Summary

Several ways to prevent or retard corrosion on equipment are reviewed and discussed for their importance in assuring proper engineering function, contributing to extended service life, maintaining appearance, and saving operation and maintenance costs. The "best" method to avoid corrosion is the proper choice of materials that are inert toward the fluids handled and the environment. A knowledge of the reactions, characteristics and general behavior of materials when exposed to certain environments is essential. Obtaining reliable field corrosion data and service history is highly valuable in support of an accurate analysis and satisfactory material and method selection for long equipment life. In the absence of actual corrosion information for a given application, information on other similar applications could be used as a starting point. Materials selected should be further studied in the real system under actual operating conditions.

Corrosion allowances should be considered in the design stage of all new equipment; therefore, it is important that

corrosion rates be determined as accurately as possible. Corrosion influences the choice of materials in design. If the corrosion rate is high, higher alloyed materials, though more expensive, might be more economic. In addition to selecting the proper material of construction, other ways or methods to reduce or prevent corrosion are:

(1) Protective paints.

(2) Protective metal or organic coatings.

(3) Protective films produced on surfaces by chemical reactions.

(4) Control of aqueous solution pH values toward slightly alkaline.

(5) Application of electric potential to equipment.

The choice of cost-effective corrosion-prevention methods is greatly influenced by the interior environment and the exterior equipment environment such as moisture content, sunlight, outdoor temperature, chemical reaction, salt, galvanic reaction, wear, abrasion, and stress. The key to successful corrosion control is the thorough understanding of its causes. Corrosive environment, service conditions, economics and life expectancy of the equipment should be investigated. As the environmental conditions become more severe, or as the standard of acceptability is elevated, changes to higher grade construction material are dictated. Additional work in this area is now being pursued and will be the subject of a follow-on report.

References

1. Fontana, M. G., and Greene, N. D., *Corrosion Engineering*, McGraw-Hill Book Co., New York, 1967.
2. Uhlig, H. H., *Corrosion and Corrosion Control*, 2nd. ed., John Wiley & Sons, New York, 1971.
3. Perry, R. H., and Chilton, C. H., *Chemical Engineers' Handbook*, 5th ed., McGraw-Hill Book Co., New York, 1973.
4. *System Design Manual, Part 5, Water Conditioning,* Carrier Air Conditioning Company, 1972.
5. Puckorius, P. R., "Controlling Corrosive Microorganisms in Cooling-Water Systems," *Chem. Eng.*, Oct. 1978, p. 171.
6. *Corrosion Data Survey – Metals Sections*, 5th ed., National Association of Corrosion Engineers, Houston, Texas, 1974.
7. Uhlig, H. H., Triadis, D. N., and Stern, M., "Effect of Oxygen, Chlorides, and Calcium Ion on Corrosion Inhibition of Iron by Polyphosphates," *J. Electrochem. Soc.*, Feb. 1955. p. 59.
8. Kirby, G. N., "Corrosion Performance of Carbon Steel," *Chem. Eng.*, Mar. 1979, p. 72.
9. Speller, F. N., *Corrosion, Causes and Prevention*, McGraw-Hill Book Co., New York, 1951.
10. Leckie, H. P., "Iron, Carbon Steel, and Low-Alloy Steels in the Process Industries," *Process Industries Corrosion*, National Association of Corrosion Engineers, Houston, Texas, 1975.

Table 1. Galvanic series of metals and alloys

Corroded end (anodic, or least noble)

- Magnesium
- Magnesium alloys
- Zinc
- Galvanized steel or galvanized wrought iron
- Aluminum 6053
- Aluminum 3003
- Aluminum 2024
- Aluminum
- Alclada
- Cadmium
- Mild steel
- Wrought iron
- Cast iron
- Ni-Resist
- 13% chromium stainless (active)
- 50-50 lead-tin solder
- 18-8 stainless type 304 (active)
- 18-8-3 stainless type 316 (active)
- Lead
- Tin
- Muntz metal
- Naval brass
- Nickel (active)
- Inconel (active)
- Monel (active)
- Yellow brass
- Admiralty brass
- Aluminum bronze
- 70-30 cupronickel
- Nickel (passive)
- Inconel (passive)
- Monel (passive)
- 18-8 stainless type 304 (passive)
- 18-8-3 stainless type 316 (passive)
- Silver
- Titanium
- Graphite
- Gold
- Platinum

Protected end (cathodic, or most noble)

Table 2. Electromotive potential of some metals

Metal	Ion	Molal electrode potential[a] at 25 C° Volt
Magnesium	Mg^{++}	−2.34
Aluminum	Al^{+++}	−1.67
Zinc	Zn^{++}	−0.76
Chromium	Cr^{+++}	−0.71
Iron	Fe^{++}	−0.44
Cadmium	Cd^{++}	−0.40
Nickel	Ni^{++}	−0.25
Tin	Sn^{++}	−0.14
Lead	Pb^{++}	−0.13
Hydrogen	H^{+}	Arbitrary zero point
Copper	Cu^{++}	+0.34
Silver	Ag^{+}	+0.80
Palladium	Pd^{++}	+0.83
Mercury	Hg^{++}	+0.85
Platinum	Pt^{++}	+1.20
Gold	Ag^{+++}	+1.42

[a]The potential values given in this table apply only to the conditions where the metal is in contact with a solution in which the activity of the ion indicated is 1.0 mol/100 gr water. In any other solution, different values for the potentials would be developed.

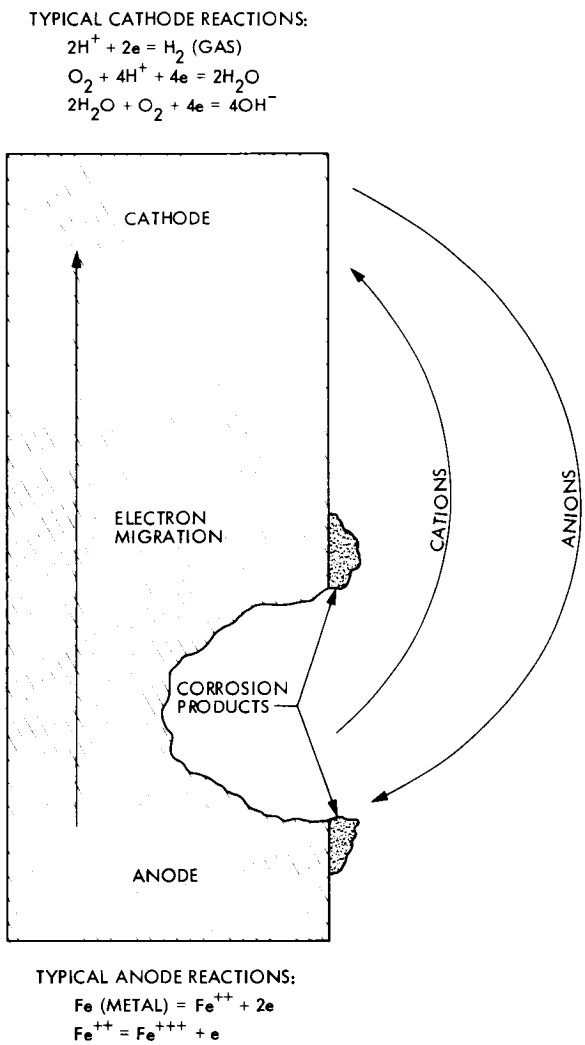


Fig. 1. Electrolytic corrosion cell

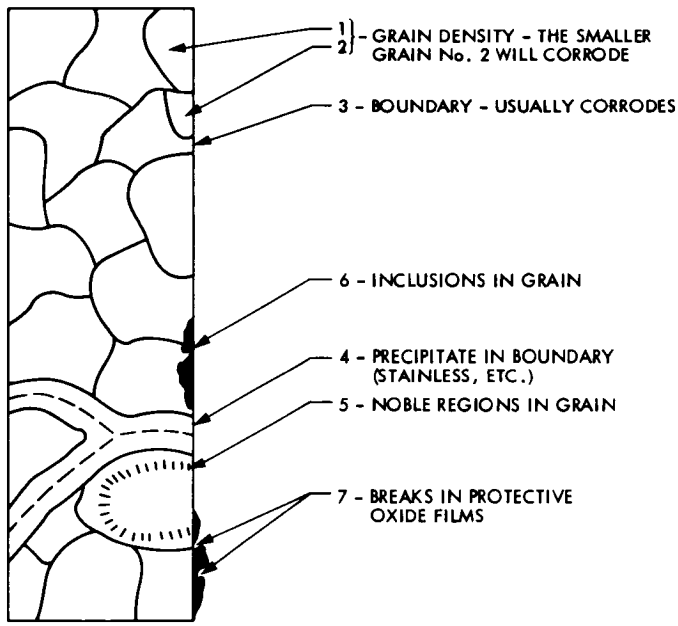


Fig. 2. Grain influence on corrosion

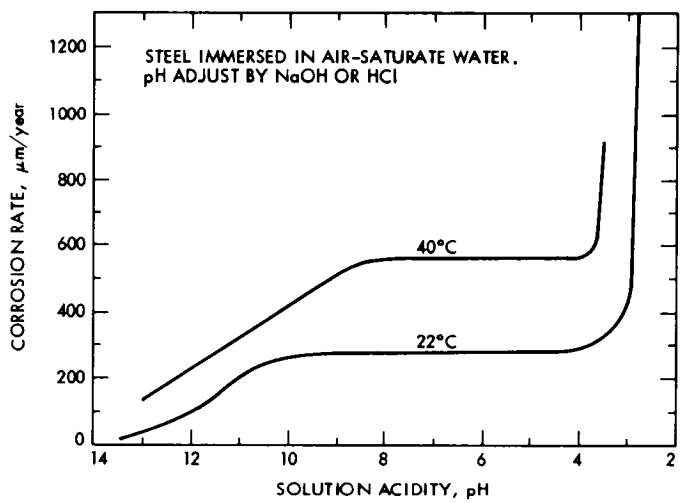


Fig. 3. Effect of solution acidity on corrosion rate

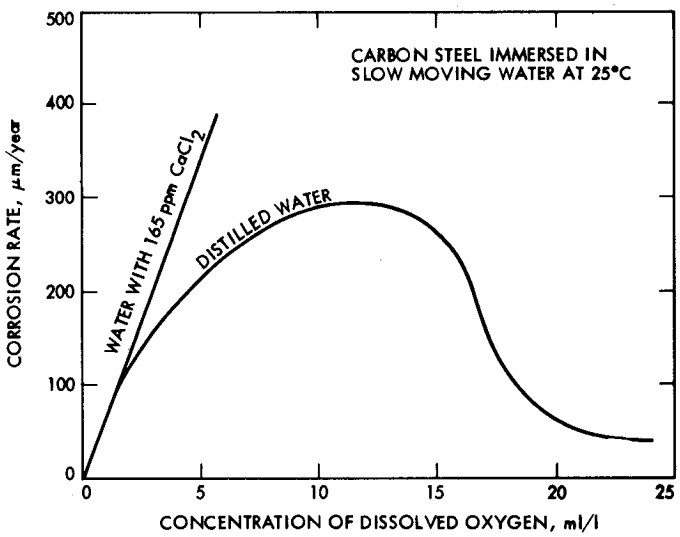


Fig. 4. Effect of dissolved oxygen on corrosion rate

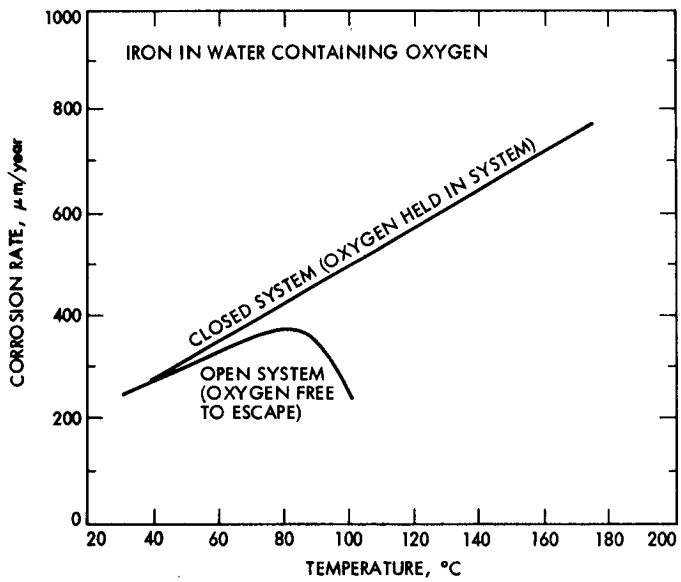


Fig. 5. Effect of temperature on corrosion rate

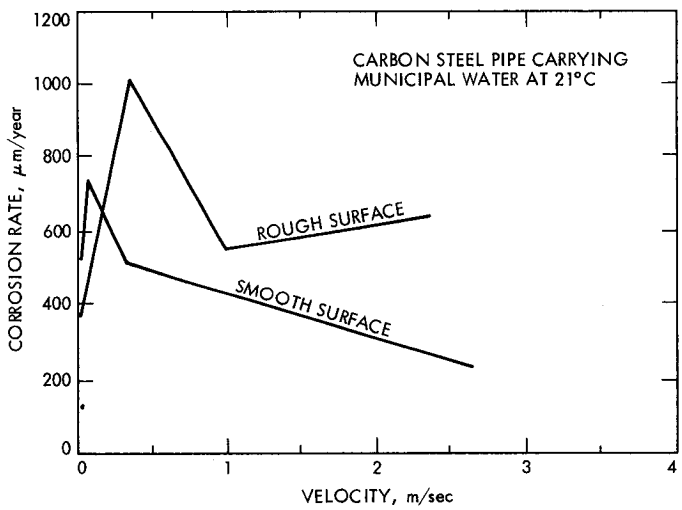


Fig. 6. Effect of fluid velocity on corrosion rate

A New High-Power Klystron for the DSN

A. Goldfinger
Varian Associates, Inc.

M. A. Gregg and R. Hartop
Radio Frequency and Microwave Subsystems Section

In order to obtain a very-high-reliability 100-kW CW klystron for the DSN high-power transmitters in support of spacecraft to the distant planets, a study contract was awarded to Varian Associates in 1977. Results of that study have been reported previously. This report continues with the last phases of the contract, which resulted in the delivery of a prototype klystron meeting all requirements.

I. Introduction

In 1977, JPL initiated a study program with Varian Associates, Palo Alto, to develop an improved, high-reliability 100-kW CW klystron for the DSN. The program was divided into four phases:

Phase I	Study Definition
Phase II	Design Improvement
Phase III	Electron Gun Fabrication and Evaluation
Phase IV	Prototype Klystron Fabrication and Delivery

The results of Phases I and II were reported in DSN Progress Report 42-50 (pp. 196-205). Phases III and IV are now complete and the results are presented here. Included are data measured on the prototype klystron which met or exceeded all requirements.

II. Electron Gun Fabrication and Beam Analyzer Evaluation

A. Background and New Information

At the conclusion of the Phase II study, it was clear that removal of the modulating anode from the existing X-3060 design klystron would substantially reduce the voltage gradients within the electron gun structure and thereby reduce any tendency toward arcing during high-voltage operation. Computer design techniques were used to modify the focus electrode and anode design to compensate for the removal of the modulating anode. The correct electron trajectories and electron beam size were provided by this computer design.

A logical progression would dictate that the computer design be evaluated in the Varian beam analyzer to verify its performance. However, coincident with the Phase II study, another very-high-power development program (VKS-8269; 450 kW CW @ 2450 MHz) had produced an electron gun with such excellent performance that the decision was made to scale this newly proven design to fit the new VKS-8274 JPL klystron and test this scaled model in the beam analyzer. Beam analyzer testing at this point would be used then to verify the performance of the scaled model as opposed to the more common usage of iterative correction to the Phase II computer design.

This new scaled structure was first evaluated to see that the voltage gradient improvement established in the Phase II computer design would not be compromised. Quite the opposite was true. Because of the new design's larger physical

dimensions and more generously radiused electrodes, a significant gain was made. When compared to the Phase II computer design, the highest gradient was reduced by 81%. Compared to the original X-3060's highest gradient of 312 kV/inch, the new design's highest gradient of 143 kV/inch represents an improvement of 312/143 = 2.18X.

Figure 1 presents a comparison of the three designs under discussion. The 250-kW/inch reference line shown represents a generally accepted value for conservative gradient design in a dc application. By any standards, the new VKS-8274 JPL design can be considered ultraconservative.

One additional gain has been made in adopting the new design. Although the new design was scaled "down" from the 450-kW VKS-8269, it represents a scale "up" from the Phase II computer design and the X-3060 structure. The new cathode diameter is 1.263 inches compared to the 1.0-inch diameter of both the X-3060 and Phase II computer design. The increased diameter represents an increased area of 1.595 times and reduces the already conservative cathode loading of 1.375 A/cm^2 to 0.862 A/cm^2. The 0.862 A/cm^2 cathode loading is also in the ultraconservative region, permitting operation of the cathode at reduced temperatures and providing long life reliability.

B. Fabrication Details

To take full advantage of both the electrostatic and magnetic beam analyzer, two versions of the new electron gun were fabricated. The two guns used many common parts, but were different in that one version used no magnetic components whatsoever, and the other used the proposed magnetic pole piece and magnetic field shaping cylinder. The construction of these two models is shown diagrammatically in Figs. 2 and 3.

C. Beam Analyzer Results

The three major goals in the development of an electron gun for the VKS-8274 JPL are (1) correct design perveance, (2) correct design size electron beam, and (3) minimum longitudinal scalloping in the presence of the design magnetic field.

(1) The rf design of the VKS-8274 JPL was established around a perveance 1.0 X 10^{-6} electron gun. This perveance assures compatibility with existing X-3060 installations and represents a conservative value in terms of good klystron performance and electron optics.

(2) The desired beam size is determined primarily by the rf design parameters. The filling factor (beam size/tunnel size) is a compromise between the desire to have the best coupling (largest beam) for high efficiency and the requirement that the beam be small enough so that interception by the drift tubes will be minimal, thus insuring thermal stability and reliability. Experience has shown that a filling factor of 65% is a desirable compromise. The design tunnel size for the VKS-8274 JPL is 0.480 in.; therefore, the target beam size is 0.65 in. X 0.480 in. = 0.312 in.

(3) Minimum beam scalloping (size change) is an obvious design requirement, for without a consistent beam size, consistent performance is not possible. A badly scalloping beam will change size markedly with variations in both beam voltage and magnetic field, with subsequent changes in power output, gain and phase response. A reasonable target value for beam scallop is less than 10%, the scallop percentage being defined by:

$$\frac{\text{Beam maximum - Beam minimum}}{\text{Beam maximum + Beam minimum}} \times 100.$$

1. Electrostatic test. Testing of any new electron gun design is performed first in the electrostatic beam analyzer to determine its true characteristics without the presence of magnetic fields. Figure 4 shows the electrostatic beam profile in the region of the beam minimum. The region beyond the beam minimum is of no importance in this test since it will later be completely controlled by the confining magnetic field. The beam is cross-sectionally scanned every 0.050 in. from just below the gun anode (0.025 in.) to a point 0.700 in. from the anode. The beam minimum can be seen at a point 0.400 in. from the anode and shows a beam diameter on paper of 0.370 in. The presented curves are always in error (too wide) by the diameter of the pinhole in the scanning target. In this case, the target pinhole is 0.010 in. and when subtracted from the apparent width of 0.370 in. gives a true beam diameter of 0.360 in. This 0.360 in. beam, just 15% larger than the target value of 0.312 in., is considered very suitable for further testing in the magnetic field without alteration of the gun electrodes.

In addition, from the data presented, an excellent beam profile can be seen. The perveance is also on target at 1.008 X 10^{-6}.

2. Magnetic test. After the electrostatic test above was completed, the electron gun was disassembled and rebuilt to include the magnetic elements. It was then installed in the magnetic beam analyzer for vacuum processing and testing.

Figure 5 shows the beam profile in the presence of all the magnetic elements of the electron gun and operating in the design magnetic field. As in the electrostatic tests, the beam profiles are taken from just below the anode (0.050 in.). In the magnetic tester, however, profiles are taken over an

extended range, in this case, 3.5 in., to view more critically the confining effect of the magnetic field. This 3.5 in. scan (not shown) represents a distance in the actual klystron which extends well past the input cavity of the klystron. From this point on, the confining magnetic field is uniform and no further change in the beam profile will occur.

Results of the magnetic beam tester are as follows:

(1) The diameter of the electron beam at the former beam minimum (Z = 0.400 in.) has been compressed from 0.360 in. to 0.305 in.).

(2) The first inward scallop occurs at Z = 1.0 in. and the beam diameter is measured at 0.280 in.

(3) The first outward scallop occurs at Z = 1.9 in. and measured 0.300 in.

(4) The scallop pattern is essentially repeated for the remainder of the Z axis traverse.

Comparison of the largest beam diameter to the design target = 0.312 in./0.300 in. = 1.04 in., just 4% from the design goal.

The beam scallop as defined by

$$\frac{\text{Beam maximum - Beam minimum}}{\text{Beam maximum + Beam minimum}} \times 100\%$$

calculates to be:

$$\frac{0.300 - 0.280}{0.300 + 0.280} \times 100\% = 3.45\%, \text{ an insignificant ripple.}$$

The current density profile is quite uniform and compares favorably with well-tested electron beams. The minor variations in the beam profiles, as they are scanned on the Z axis, have been previously traced to slight misalignment of the electron and magnetic axis. This is not expected in the actual klystron.

In summary, the beam analyzer tests have shown the production of an excellent electron optical system, and verify that the excellent performance of VKS-8269 (450 kW @ 2450 MHz) is to a large degree dependent on an excellent electron beam. It permits the prediction of comparable performance in the VKS-8274 JPL.

III. Klystron Prototype Fabrication

A. Mechanical Design Improvements

With only one minor exception, all mechanical improvements proposed in the Phase II Design Improvement final report were incorporated in the VKS-8274 JPL klystron. That one minor exception concerned the lengthening of the electron gun insulator to permit operation of the VKS-8274 JPL in air. This suggestion was found unacceptable with regard to operational safety, and therefore this modification was not incorporated. The insulator's length was changed to permit compatibility with existing oil-filled socket tanks.

The mechanical improvements proposed by the Phase II Design Improvements final report and also incorporated in the VKS-8274 JPL prototype klystron are reviewed below.

1. **Diaphragm trim tuner.** The proposed trim tuner used in the VKS-8274 JPL, and the wide-range tuner now used in the X-3060, are shown in Fig. 6. Casual observation will show the complexity of the old tuner design as compared to the new trim tuner. Not so obvious in the old design are the six vacuum-to-water and three vacuum-to-air brazing joints used in the present X-3060 tuner. In addition to the large number, many of the brazing joints are blind; that is, they can neither be inspected nor repaired after the final braze pass. This type of assembly severely compromises reliability and rebuildability. Compared to this, the trim-tuner design has only *two* vacuum-to-air brazing joints and *no* vacuum-to-water joints. Neither of the vacuum-to-air joints is blind, and both are easily repairable.

In addition to these mechanical advantages, the trim-tuner diaphragm is relatively far removed from the high rf fields at the drift tube gap center and is subjected to far less rf heating than the present X-3060 capacitive paddle. The diaphragm is thermally coupled by large cross-sectional areas of copper to the massive water-cooled copper cavity walls. In Fig. 6, a water-cooled post is shown joined to the diaphragm face. This was a contingency plan only; no diaphragm cooling was required.

2. **Extended tail pipe elimination.** The extended tail pipe configuration shown in Fig. 7 was first introduced in the X-3060 klystron in 1965. Its purpose was to minimize the asymmetry normally caused by the exit path of the output waveguide through the focusing magnet, and to reduce to an absolute minimum body current interception caused by the magnetic asymmetry. Mechanical considerations, however, created an extended tail pipe section that is exposed to a rapidly expanding electron beam just beyond the output gap. This extended section has created both directly and indirectly related failures in the X-3060 klystron. The problem is caused by the fact that the electron beam expands in that region faster than was predicted at the time of the design's introduction, and, in addition, secondary electrons return from the collector to impinge on the tail pipe surfaces. For these reasons, and because the tail pipe is electrically part of the

klystron body, the body current readings for the X-3060 have been recorded as high as 10% of the total beam current (700 mA). This tail pipe current (reading as body current) completely masks body current interception in any other part of the rf structure and is unacceptable because it requires that body current protective-circuit trip levels be set too high (1.0 A). This removes the protection required in the remainder of the rf structure, which may suffer damage from relatively low-quantity (0.050 A), but high-velocity electrons.

Under the circumstances just described, protection *cannot* be provided for the following common field conditions:

(1) Low or incorrectly adjusted magnetic field,

(2) Incorrect tuning,

(3) Overdrive,

(4) Stray magnetic fields (gun region),

(5) Disturbance of the main magnetic field by accidental introduction of ferrous materials (screwdrivers, wrenches, etc.).

Elimination of the extended tail pipe design was a mandatory condition for reliability. Figure 8 shows the new design to be employed in the VKS-8274 JPL klystrons. As shown, the new design provides adequate clearance for the expanding beam and transfers tail pipe interception current to the collector where it is properly metered as collector current. With this configuration, body current readings will return to normal values in the order of 0.050 A, and the rf structure can be properly protected with body current protective trips set at 0.100 A.

One additional design change has been made in conjunction with the elimination of the extended tail pipe. A reentrant output pole piece has been introduced to create a peak in the magnetic field near the output gap. Recent designs have shown that this peaked magnetic field is very beneficial in terms of reducing beam interception in the output region of the klystron. This reentrant design is shown in Fig. 8.

3. **Cavity/body construction.** During the Phase I investigation of the X-3060, it was found that the rigidity of the rf structure was somewhat lacking, and structural stiffeners were promised for proposed new designs. That is the case, but in addition, recent tests of an X-3060 (May 1978) showed thermal drift attributable to cavity detuning, indicating a need for additional cavity cooling.

The newly designed VKS-8274 JPL incorporates relatively massive copper cavities which have a wall thickness in the order of three times the wall thickness of the present X-3060. Not only does this create a much more rugged rf structure, it permits the passage of water through the cavity walls to insure greater thermal stability. Comparative views of the X-3060 cavity structures and the new design are shown in Figs. 9 and 10. In addition to increased cavity wall thickness and cooling, cavity and walls and drift tubes have increased thermal cross section to provide the best possible thermal stability. This can be seen clearly in Figs. 10 and 11.

In short, the new rf structure has construction closely paralleling a 450-kW klystron (VKS-8270) known to be operationally stable at an even higher frequency; i.e., 2380 MHz.

B. Electrical Design Improvements

Without exception, all electrical design improvements proposed in the Phase II Design Improvement final report were incorporated in the VKS-8274 JPL. Those design improvements are reviewed below.

1. **Electron gun redesign.** Findings in the Phase II Design Improvement study showed that a substantial improvement could be made in electron gun voltage gradients by removing the modulating anode. This change, plus incorporation of a new optical design, have been incorporated and successfully tested in the VKS-8274 JPL. The improvements and results are fully discussed in Sections I and II.

2. **RF structure redesign.** During the Phase I investigation of the present X-3060 rf design, a number of deficiencies were noted. These deficiencies were a natural result of designing a klystron with a wide tuning range; i.e., 2114 to 2388 MHz. In particular, drift tube gap spacing and gap-to-gap drift tube distances were compromised to satisfy operation at both ends of the frequency tuning range. In fact, the output cavity drift tube gap in the present X-3060 design is far too long to provide optimum efficiency at either 2114 or 2388 MHz, but it was designed that way to provide low gap capacitance, a requirement of wide tuning range cavities. It was suggested in the Phase II final report that the wide tuning range requirement be eliminated and that two tubes be developed, each electrically optimized for its particular frequency, one at 2114 MHz and the other at 2388 MHz. These tubes would have narrow-range trim tuners only, to permit initial factory tune-up.

JPL's decision at the conclusion of the Phase II final report was to proceed with the development of an optimized design at 2114 MHz. The design chosen and implemented in the VKS-8274 JPL is nearly electrically identical to the 5K70SG klystron. The 5K70SG was used as the basic model because it has demonstrated high efficiency performance, with reliability

in numerous field applications for a period exceeding ten years.

The test results shown in Figs. 11-13 and Tables 1 and 2 verify the validity of this approach. Substantial gains were made in efficiency, gain, and bandwidth.

IV. Conclusion

The development of a new high power klystron for the DSN has been very successful as demonstrated by the prototype results. A new contract for four production klystrons has been awarded to Varian Associates, with deliveries planned for early 1983.

Table 1. Certified test report VKS-8274 klystron

	Broad band tuned at 2114 MHz
Power output	146 kW
Beam voltage	37 kV dc
Beam current	7.44 A dc
Beam power input	275 kW
Efficiency	53%
Drive power	0.15 W
Gain	59.8 dB
Bandwidth (- dB)	1.9 MHz
Body current, no rf drive	6.5 MA dc
Body current, sat. rf drive	14 mA dc
Focusing magnet current	12 A dc
Focusing magnet voltage	120 V dc
Heater voltage	10.5 V ac
Heater current	11.35 A ac
Coolant water	
Collector	
Flow	65 gpm
Pressure drop	32 psi
Body	
Flow	7 gpm
Pressure drop	65 psi
Electro-magnet	
Flow	2.5 gpm
Pressure drop	67 psi

Table 2. Comparative data

Parameter	Units	X-3060 performance data	Phase II improvement predictions	VKS-8274 JPL performance data	JPL specification requirements
Frequency	MHz	2114	2114	2114	2114
Beam voltage	kV	36.0	36.0	36.0	40 max
Beam current	Adc	7.7	6.96	7.15	10 max
Power output	kW CW	112.0	132.0	146.0[a]	110 min
Efficiency	%	40.4	52.7	52.8[a]	40 min
Saturation gain	dB	54.7	59.3	59.2	50 min
Saturation BW	MHz	10.4	16.7	18.9[a]	14 min
Body current no rf	MA dc	30.0	–	6.1[a]	–
Body current sat. rf	MA dc	540.0	50.75	11.1[a]	100 max
Electron gun high voltage gradients (worst case)	kV/inch	312.0	254.0	143.0[a]	–
Cathode loading	amps/cm^2	1.375	1.375	0.862[a]	–

[a]Notable improvement over X-3060 performance.

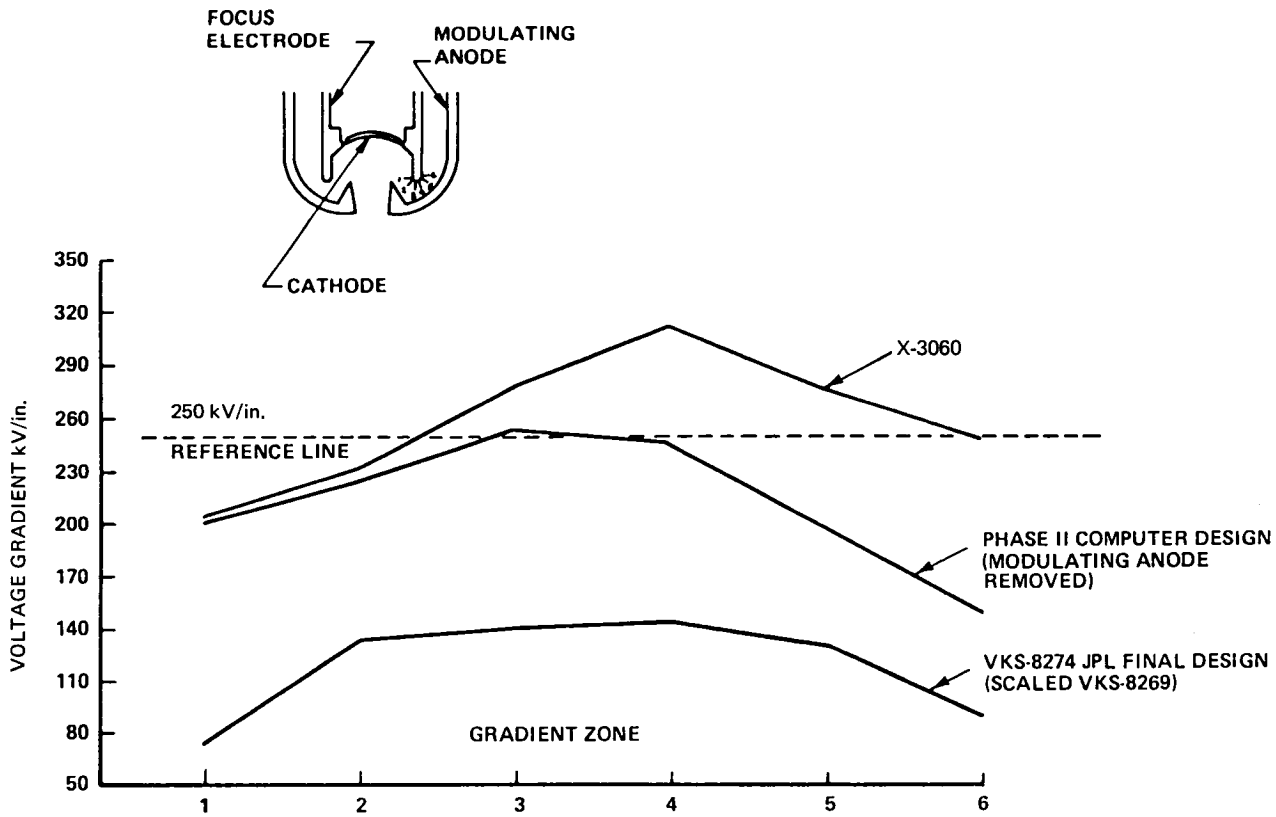


Fig. 1. High voltage gradient comparison: three designs

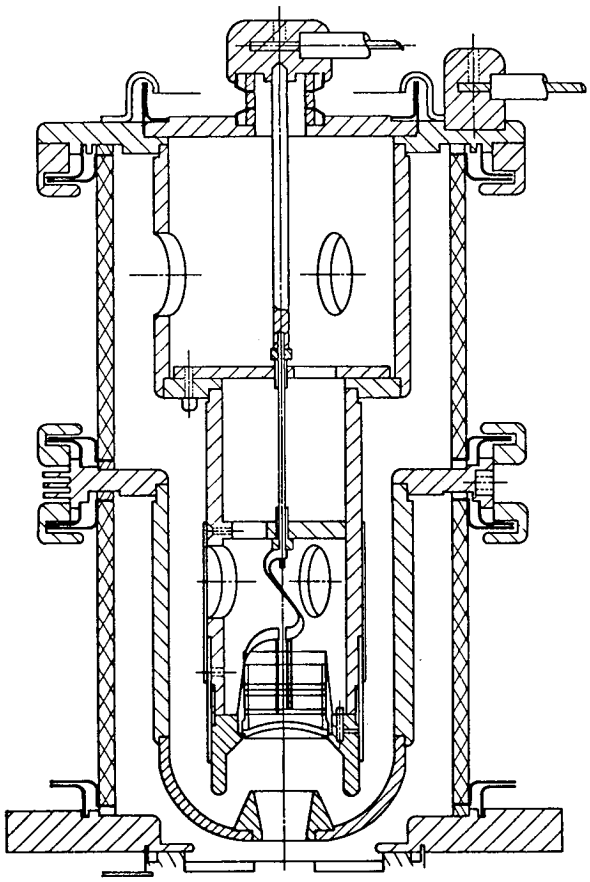


Fig. 2. X-3060 electron gun layout

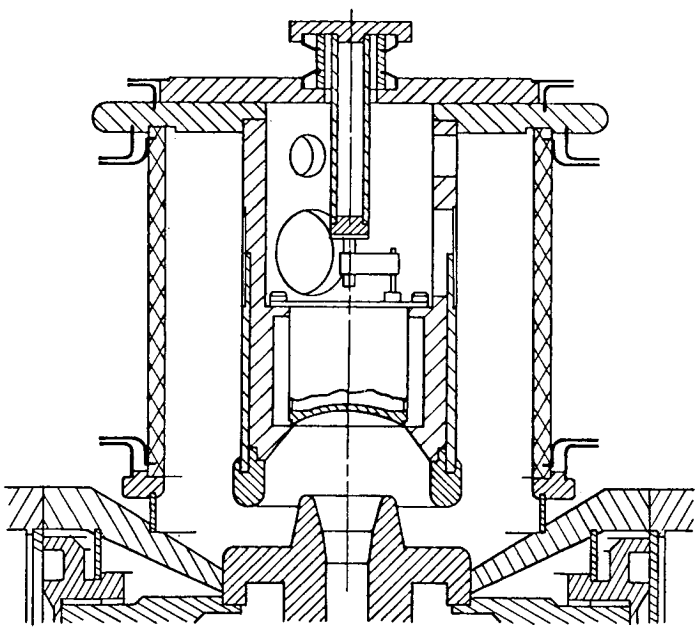


Fig. 3. New design layout VKS-8274 JPL

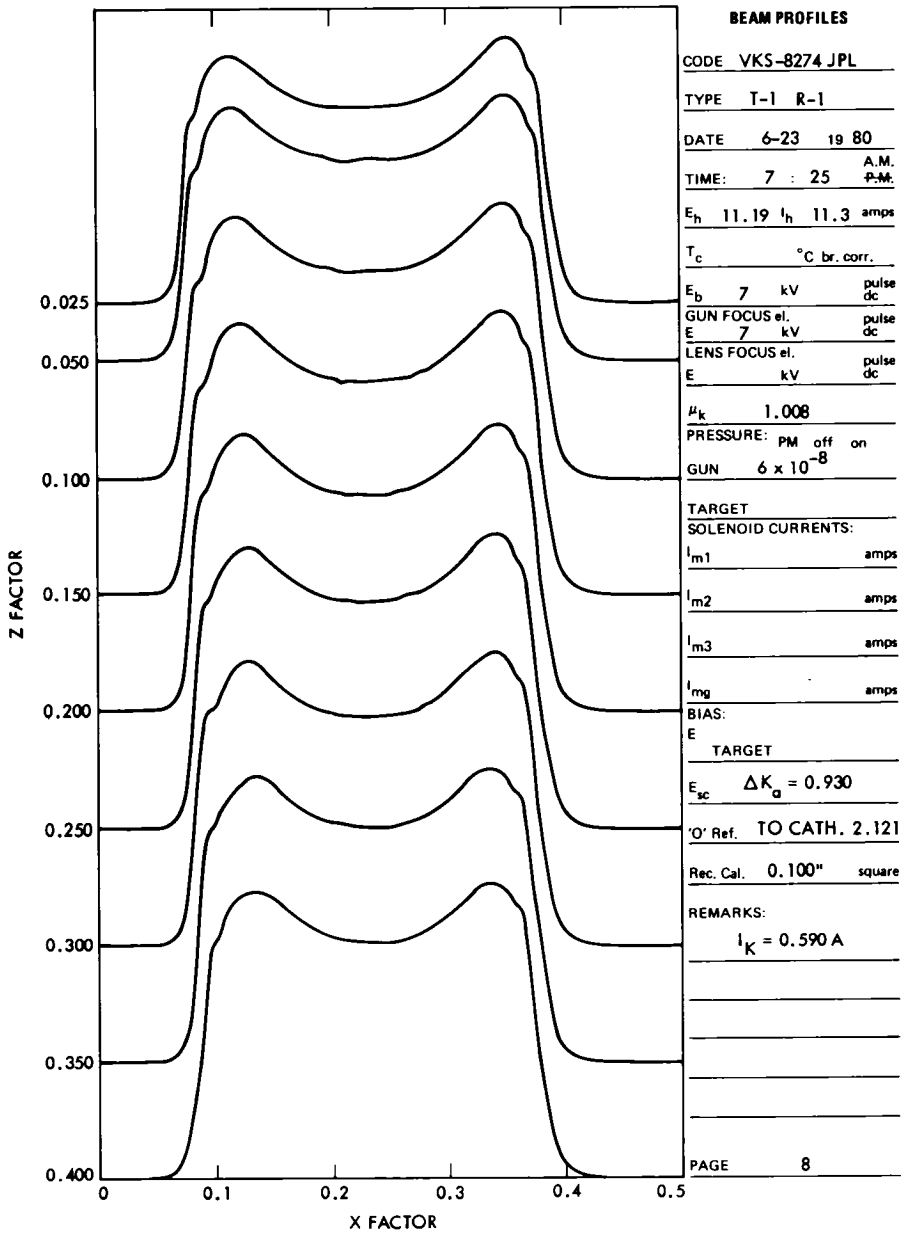


Fig. 4. Electrostatic beam profiles

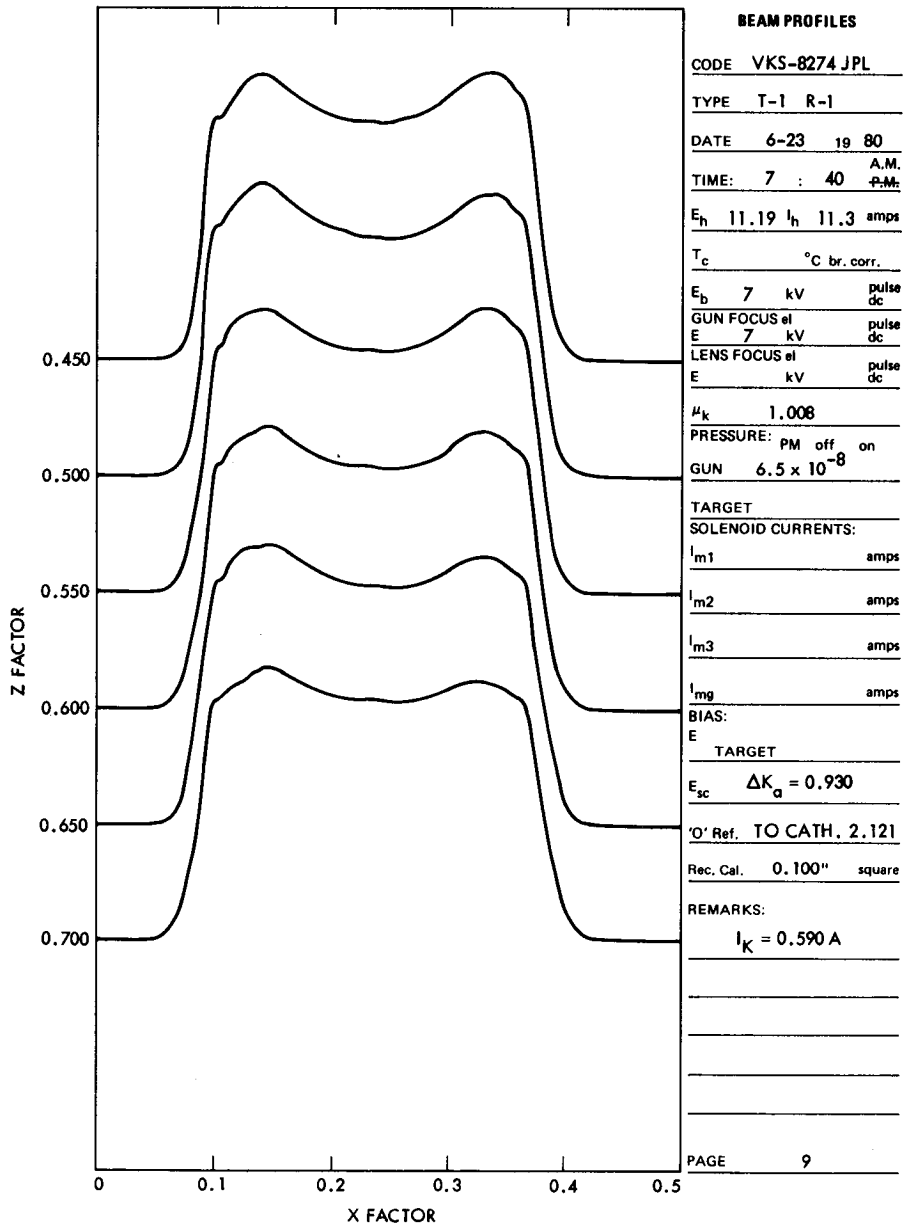

BEAM PROFILES
CODE VKS-8274 JPL
TYPE T-1 R-1
DATE 6-23 19 80
TIME: 7 : 40 A.M. P.M.
E_h 11.19 I_h 11.3 amps
T_c °C br. corr.
E_b 7 kV pulse dc
GUN FOCUS el
E 7 kV pulse dc
LENS FOCUS el
E kV pulse dc
μ_k 1.008
PRESSURE: PM off on
GUN 6.5 x 10^{-8}
TARGET
SOLENOID CURRENTS:
I_{m1} amps
I_{m2} amps
I_{m3} amps
I_{mg} amps
BIAS:
E
TARGET
E_{sc} ΔK_a = 0.930
'O' Ref. TO CATH. 2.121
Rec. Cal. 0.100" square
REMARKS:
I_K = 0.590 A
PAGE 9
Z FACTOR
0.450
0.500
0.550
0.600
0.650
0.700
0
0.1
0.2
0.3
0.4
0.5
X FACTOR

Fig. 4 (contd)

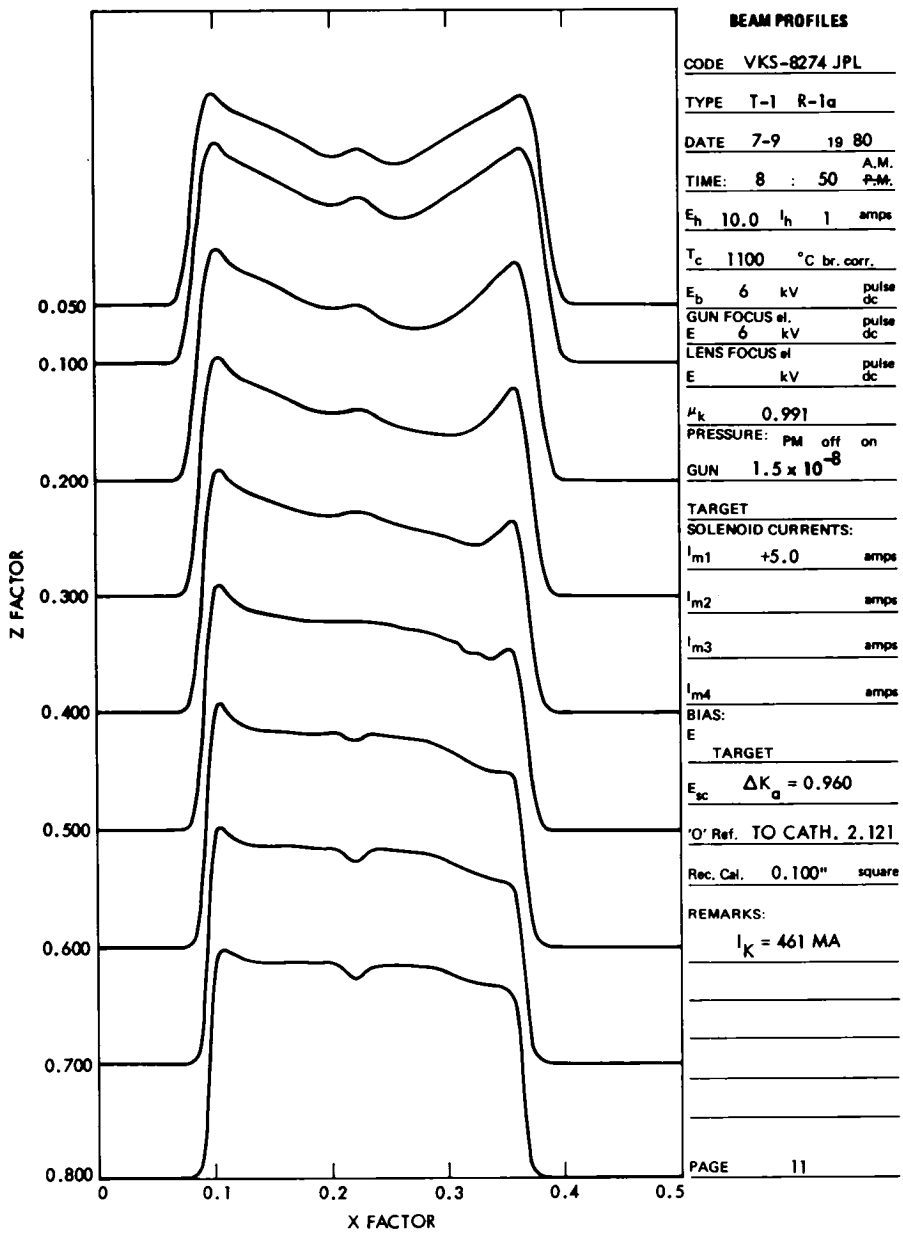


Fig. 5. Magnetic beam profiles

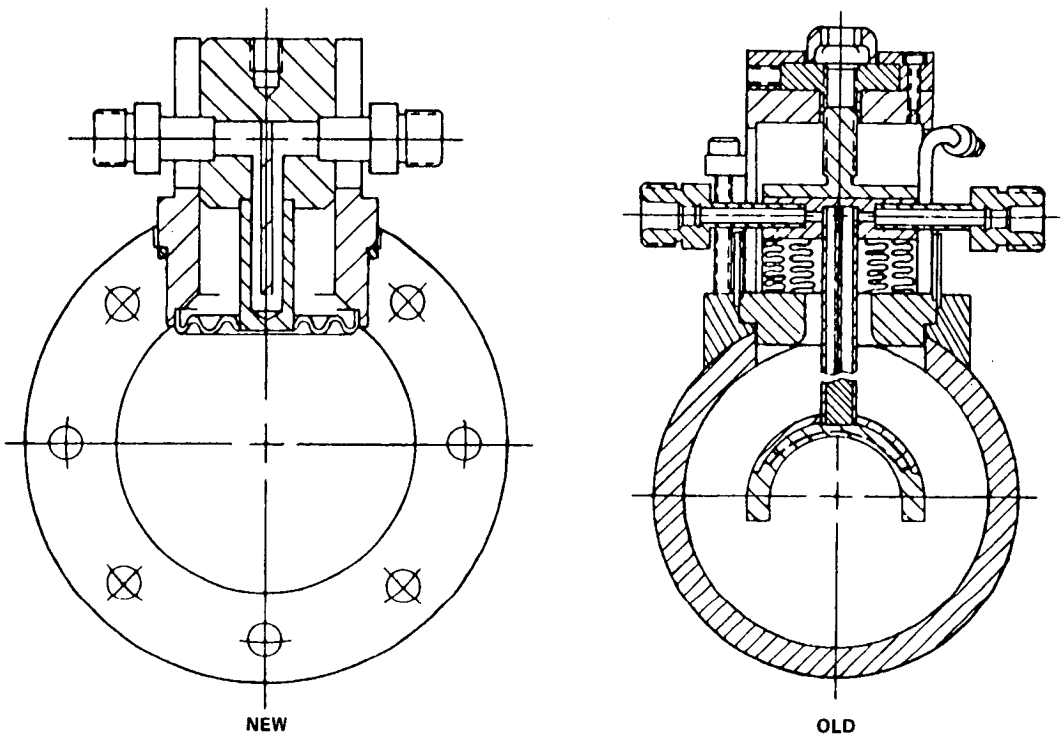


Fig. 6. Tuner designs compared

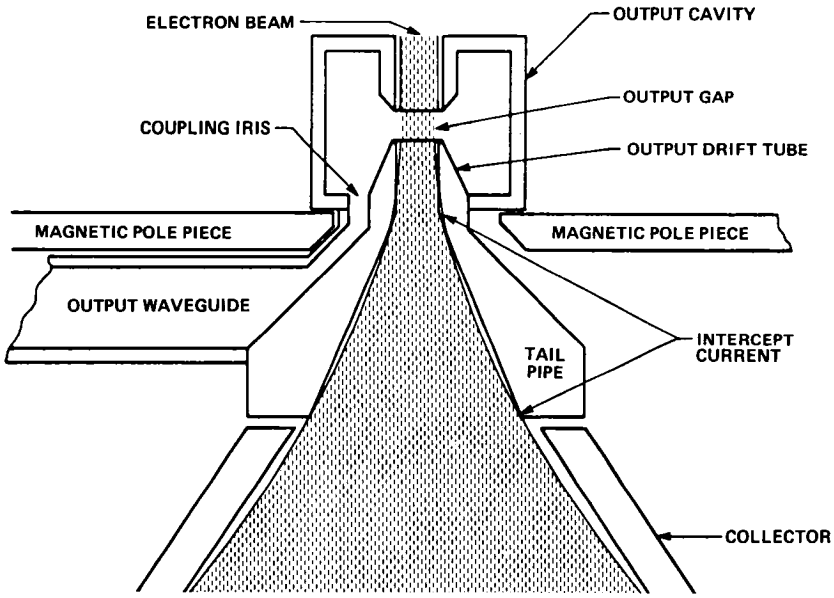


Fig. 7. X-3060 extended tail pipe

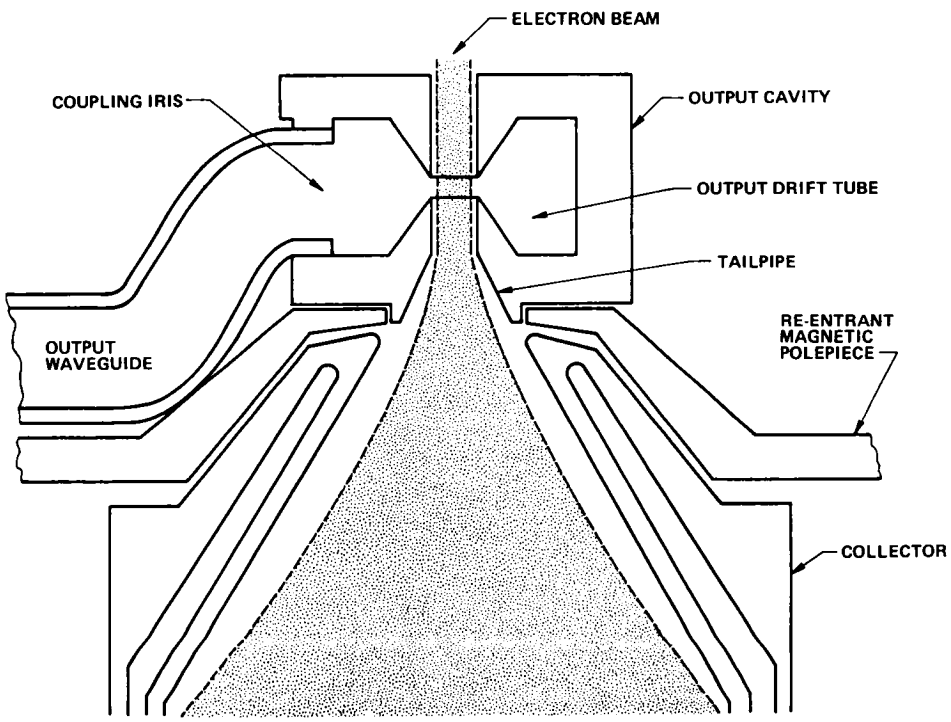


Fig. 8. VKS-8274 JPL output circuit

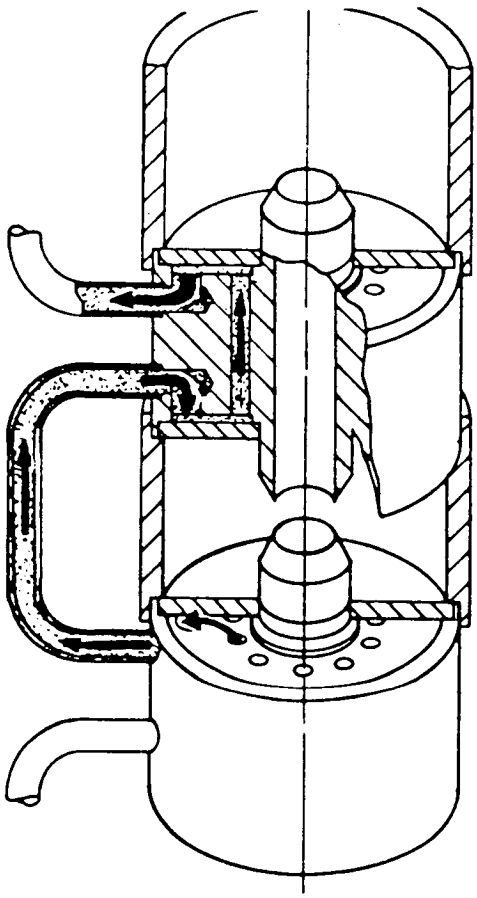


Fig. 9. Old cavity structure (X-3060)

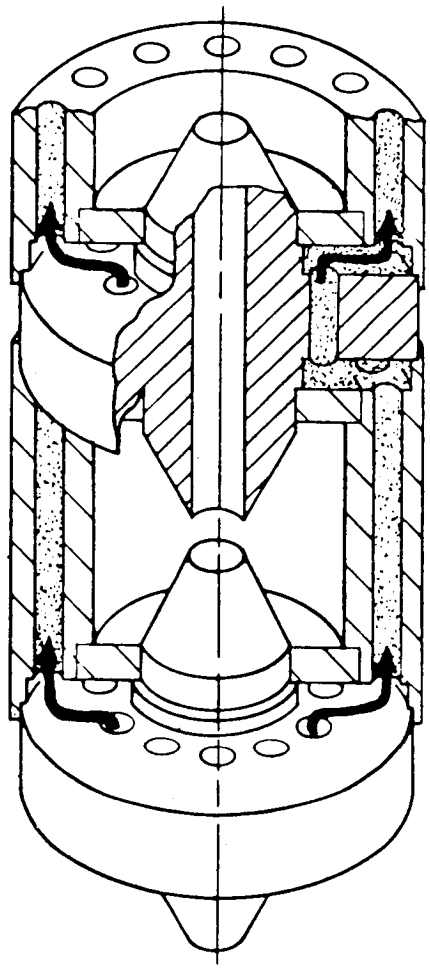


Fig. 10. New cavity structure (VKS-8274 JPL)

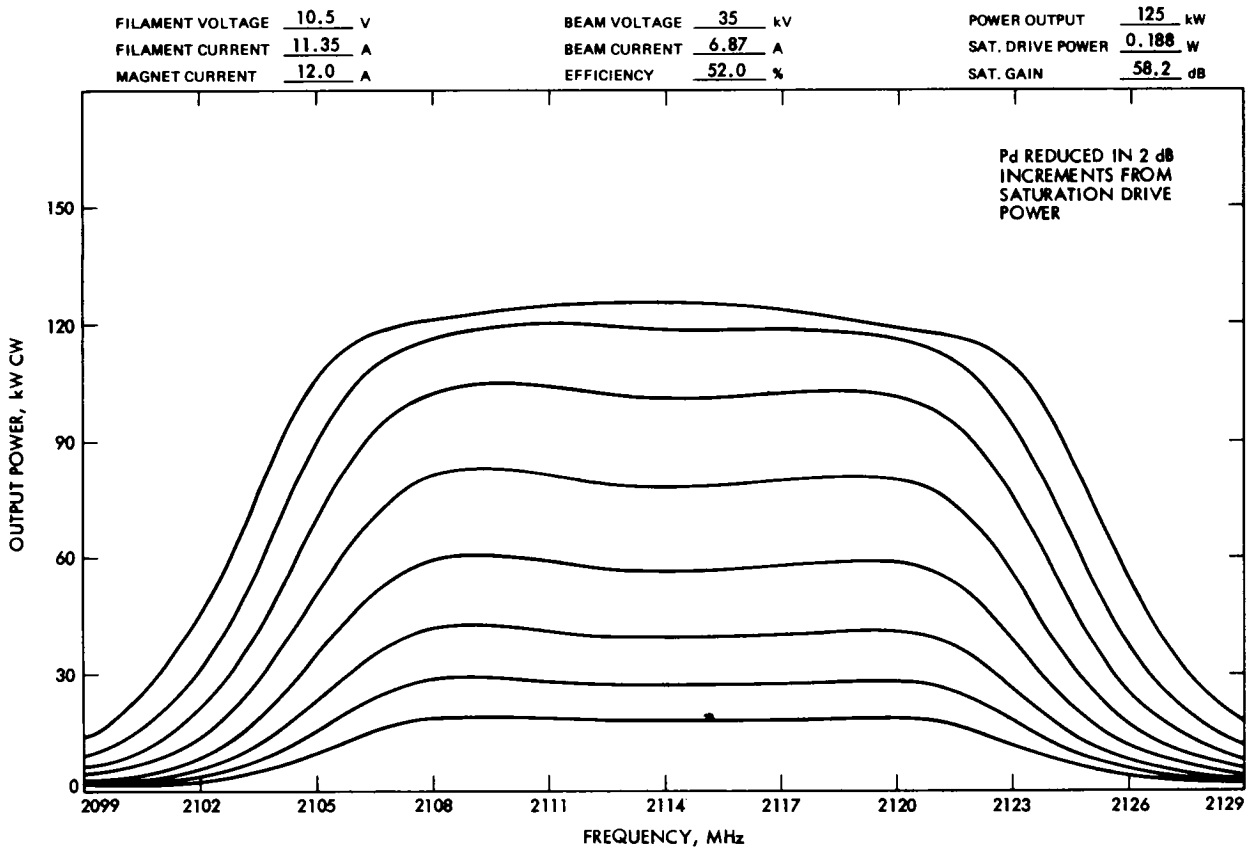


Fig. 11. Output power vs frequency

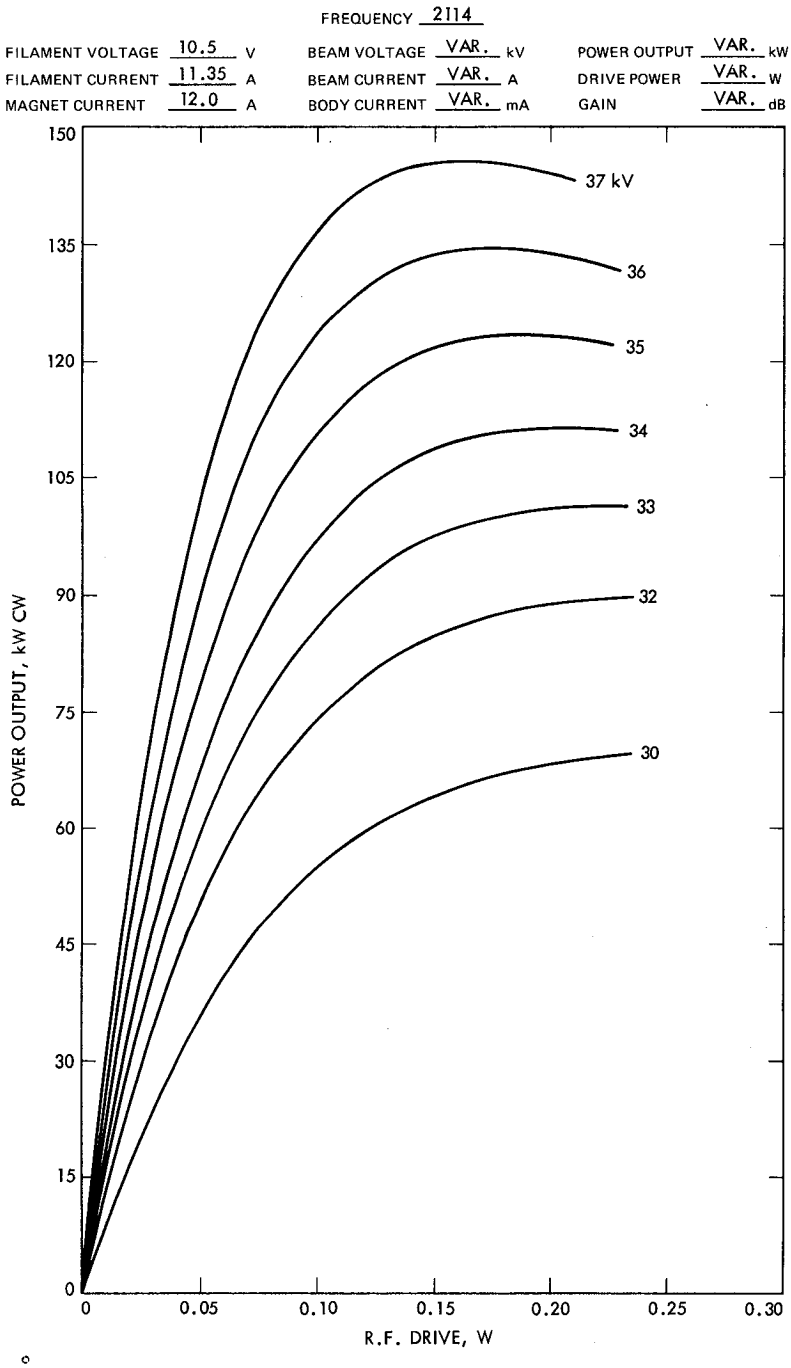


Fig. 12. Output power vs drive power

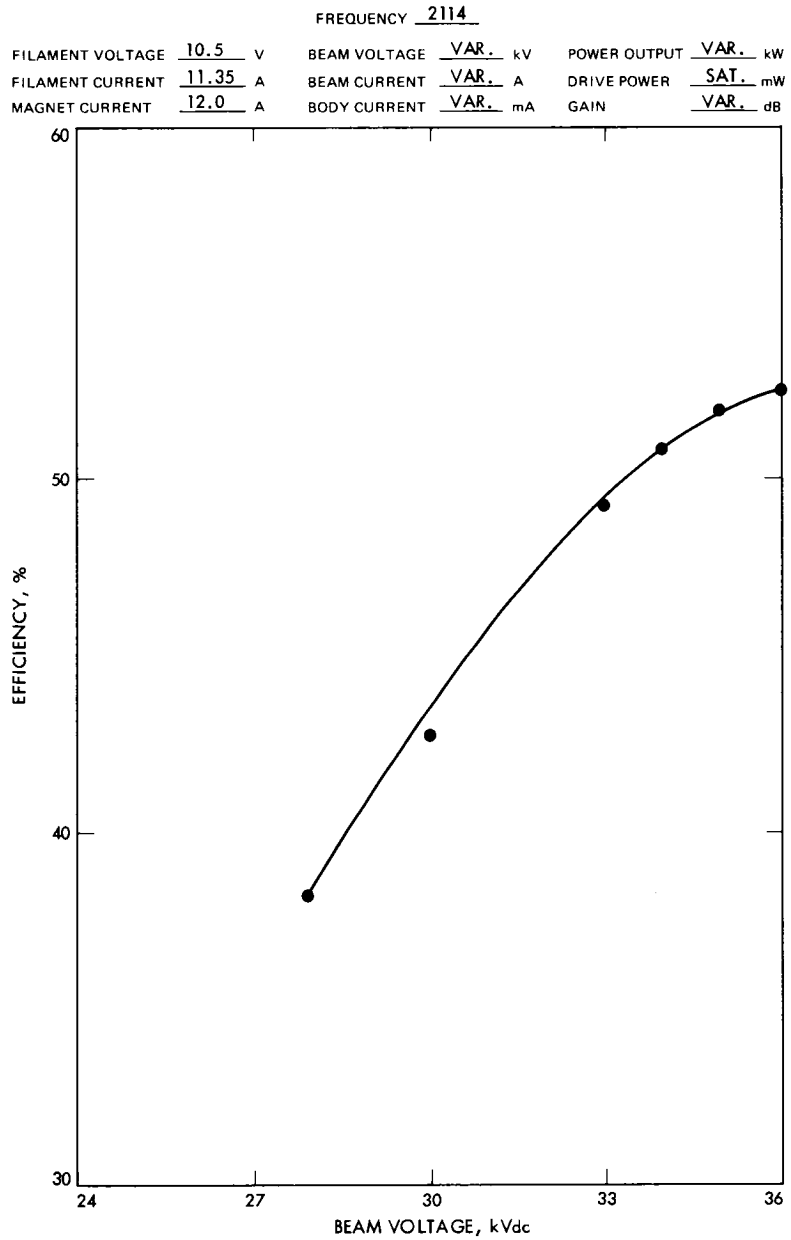


Fig. 13. VKS-8274 efficiency

Lateral and Drag Forces on Misaligned Cylindrical Rollers

H. McGinness
DSN Engineering Section

Formulas are presented for estimating the drag and lateral forces on misaligned cylindrical rollers.

I. Introduction

In DSN Progress Report 42-45 (Ref. 1), a concept was presented which accounted for the maximum lateral forces induced by a misaligned steel cylindrical wheel rolling on a flat track. The concept may be understood by referring to Fig. 1, which depicts the displacement as being comprised of two components, namely, a frictionless rolling perpendicular to the axis of the roller followed by a sliding parallel to the roller axis. The concept is that the two actions occur simultaneously. Since the rolling action is assumed to be frictionless, the resultant frictional force acting on the roller is directed along the roller axis in the sense indicated in Fig. 1. The components of the resultant frictional force $W\mu$ acting on the roller along the x and y axes are:

$$F_x = -W\mu \sin \alpha \quad (1)$$

$$F_y = W\mu \cos \alpha \quad (2)$$

where W is the normal force between the roller and track and μ is the effective coefficient of sliding friction.

Although the foregoing concept produces a lateral force which is approximately the same as the maximum values obtained from tests, it is not valid for very small values of the misalignment angle α. For example, Eq. 2 indicates that the maximum lateral force occurs when the misalignment angle is zero, whereas tests show that the lateral force is zero for zero misalignment, but increases to the full value of $W\mu$ within one degree of misalignment.

The present report is a description of a concept which yields formulas for the lateral force and the drag force in terms of the misalignment angle α, the coefficient of sliding friction μ, and the rolling friction coefficient f, which pertains to the condition of zero misalignment. From tests on various roller bearings, good estimates of f can be obtained. Values of the sliding coefficient of friction μ can be obtained from many sources.

II. Description of the Concept

The drag resistance of a roller comes from several independent sources, among which are material hysteresis, viscosity of lubricant, and sliding friction caused by a variation in the rolling radius. The last effect, for the case of a cylindrical roller on a track of greater width, may be understood by realizing that the contact area between the loaded roller and track is not flat but curved, such that the edge of the roller has a smaller rolling radius than that of the center of the roller. The total drag coefficient for a perfectly aligned roller f is some combination of these factors, and its range and values can be obtained from numerous published reports.

The concept used here is that the drag resistance of an aligned roller is caused only by differential sliding of two different roller radii. The sliding on the smaller radius is opposite to the roller motion and the sliding on the larger radius is in the direction of the roller motion, and the effect of their difference is equal to Wf. If the roller undergoes an angular misalignment α, a sliding along the roller axis occurs. The resultant sliding direction is obtained by geometrically adding the two sliding components, thus obtaining one direction for the small radius and another direction for the large radius. The resultant frictional forces must be coincident with these resultant sliding directions. The total frictional force acting on the roller is assumed to be the vector sum of the resultant from the small radius and the resultant from the large radius. The side force F_S and the drag force F_D are components of this total frictional force.

In the following analysis of the forces it is assumed that the bearing between the roller and its axle is frictionless.

III. Analysis of the Forces

Figure 2 shows two positions of the wheel or roller. The wheel is composed of radii r and R where $r < R$. Let the mean of these two radii be ρ and assume that this ficticious ρ rolls without sliding along the line α. As the wheel travels the distance U along the X axis, the r radius slips along line α by the amount $U[1 - (r/\rho)] \cos \alpha$ in the sense shown in the figure. From the figure it is clear that the radius r also slips perpendicularly to line α by the amount $U \sin \alpha$ in the sense indicated in the figure. Similarly radius R slips along line α by the amount $U[(R/\rho) - 1] \cos \alpha$ and slips perpendicularly to line α by the amount $U \sin \alpha$ in the senses shown in Fig. 2. The sliding frictional forces associated with radii r and R, namely, $Wa\mu$ and $W(1 - a)\mu$ respectively, where a is the fraction of the total normal force applied to radius r, will have the same directions as the resultant slip lines for radii r and R. Considering the wheel as a free body, the senses of the forces $Wa\mu$ and $W(1 - a)\mu$ are as shown in Fig. 2.

By referring to Fig. 2, two equations of equilibrium may be written by summing forces along line α and by summing moments about the wheel center, as follows:

$$Wa\mu \cos \gamma - W(1 - a)\mu \cos \beta = Wf \tag{3}$$

$$Wa\mu r \cos \gamma - W(1 - a)\mu R \cos \beta = 0 \tag{4}$$

When α is zero both γ and β are zero. For this condition, Eqs. (3) and (4) become:

$$Wa\mu - W(1 - a)\mu = Wf \tag{5}$$

$$Wa\mu r = W\mu(1 - a)R \tag{6}$$

From (5)

$$a = (1 + f/\mu)/2 \tag{7}$$

It is assumed that this value of a holds for any value of α. By substituting (7) into (6) the following is obtained:

$$\frac{R}{r} = \frac{a}{1 - a} = \frac{1 + f/\mu}{1 - f/\mu} \tag{8}$$

The radius ρ was defined as the mean of r and R; thus from (8),

$$\rho = \frac{r + R}{2} = \frac{r}{2}\left[1 + \frac{1 + f/\mu}{1 - f/\mu}\right]$$

$$= \frac{R}{2}\left[1 + \frac{1 - f/\mu}{1 + f/\mu}\right] \tag{9}$$

$$\rho = \frac{r}{1 - f/\mu} = \frac{R}{1 + f/\mu} \tag{10}$$

$$1 - \frac{r}{\rho} = \frac{R}{\rho} - 1 = f/\mu \tag{11}$$

Using the expressions for $1 - (r/\rho)$ and $(R/\rho) - 1$ as given by Eq. (11), namely f/μ, the angle γ and β of Fig. 2 may be evaluated thus:

$$\tan \gamma = \frac{U \sin \alpha}{U\left(\frac{f}{\mu}\right) \cos \alpha} = \frac{\mu}{f} \tan \alpha = \tan \beta \tag{12}$$

From (12) the following expressions for $\sin \gamma$ and $\cos \gamma$ may be obtained:

$$\sin \gamma = \frac{\mu \tan \alpha}{\sqrt{\mu^2 \tan^2 \alpha + f^2}} = \sin \beta \tag{13}$$

$$\cos \gamma = \frac{f}{\sqrt{\mu^2 \tan^2 \alpha + f^2}} = \cos \beta \tag{14}$$

The drag force F_D and the side force F_S are obtained respectively by summing the x and y components of the forces μWa and $\mu W(1 - a)$, shown in Fig. 2, obtaining the following:

$$F_D = -[aW\mu \cos\gamma - (1-a)\, W\mu \cos\beta]\cos\alpha$$

$$-[aW\mu \sin\gamma + (1-a)\, W\mu \sin\beta]\sin\alpha \qquad (15)$$

$$F_S = [aW\mu \sin\gamma + (1-a)\, W\mu \sin\beta]\cos\alpha$$

$$+[(1-\alpha)\, W\mu \cos\beta - aW\mu \cos\gamma]\sin\alpha \qquad (16)$$

Substituting Eqs. (7) (13) and (14) into (15) and (16), there are obtained:

$$\frac{F_D}{W} = \frac{\mu^2 \tan\alpha \sin\alpha + f^2 \cos\alpha}{\sqrt{\mu^2 \tan^2\alpha + f^2}} \qquad (17)$$

$$\frac{F_S}{W} = \frac{\mu^2 \tan\alpha \cos\alpha - f^2 \sin\alpha}{\sqrt{\mu^2 \tan^2\alpha + f^2}} \qquad (18)$$

Equations (17) and (18) express the drag force and side force in terms of the sliding coefficient of friction μ and the aligned roller coefficient f for any value of the misalignment angle α.

Some of the test results from Ref. 1 are presented in Fig. 3. The solid line pertains to the conditions of a clean track, whereas the dashed line represents the average of all the tests made. Two sets of μ and f were evaluated by Eq. (18) and compared to the test results.

IV. Conclusion

A simple analysis has yielded a formula for the lateral force coefficient of a misaligned cylindrical roller on a flat track. Reasonable values for the sliding coefficient of friction μ and for the rolling drag coefficient f when inserted into the formula produce values which agree quite well with the lateral force test results of Ref. 2.

The drag force coefficient was calculated for a misalignment value of 13.7 arc minutes and compared with the measured values obtained on a model of an antenna wheel and track system. The agreement as reported in Ref. 3 was good, but very few tests were made.

A useful application of Eqs. (17) and (18) is the estimation of drag torque of an antenna employing a wheel and track bearing system, and the estimation of the lateral forces induced on the wheel by a slight misalignment.

The present analysis gives no explanation of the "stick-slip" behavior of certain misaligned rollers reported in Ref. 2, and observed on the radial bearing of the 64-meter antenna at DSS 14 when the roller misalignment exceeds a certain value. It is hoped that further study will bring forth a quantitative explanation of this phenomenon.

References

1. McGinness, H., "Lateral Forces Induced by a Misaligned Roller," *DSN Progress Report 42-45, March and April 1978*, Jet Propulsion Laboratory, Pasadena, Calif., June 15, 1978.

2. Cross, H. A., and Talbert, S. A., "Experimental Determination of Lateral Forces Developed by a Misaligned Steel Roller and Steel Rail," Machine Design Divisions of ASME, Oct. 2, 1963.

3. McGinness, H., "Antenna Azimuth Bearing Model Experiment," *DSN Progress Report 42-53, July and August 1979*, Jet Propulsion Laboratory, Pasadena, Calif., Oct. 15, 1979.

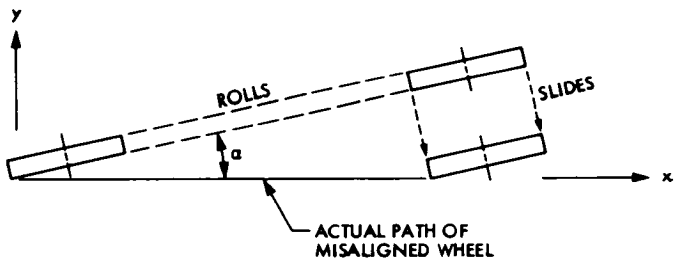


Fig. 1. Model of induced lateral force for frictionless wheel

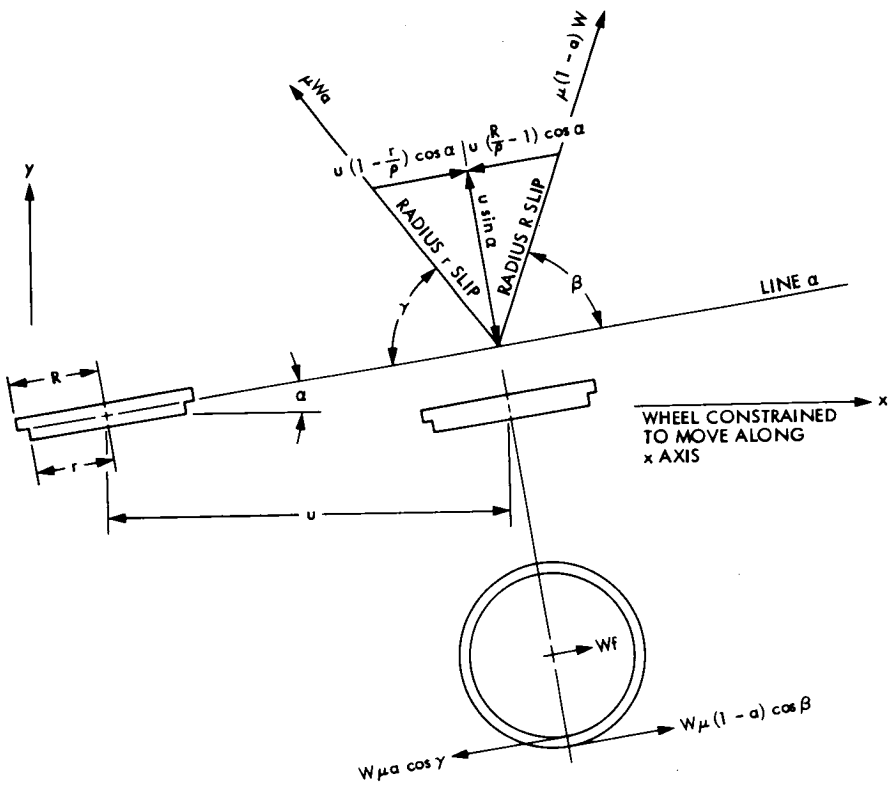


Fig. 2. Directions of slip of radius *r* and radius *R*

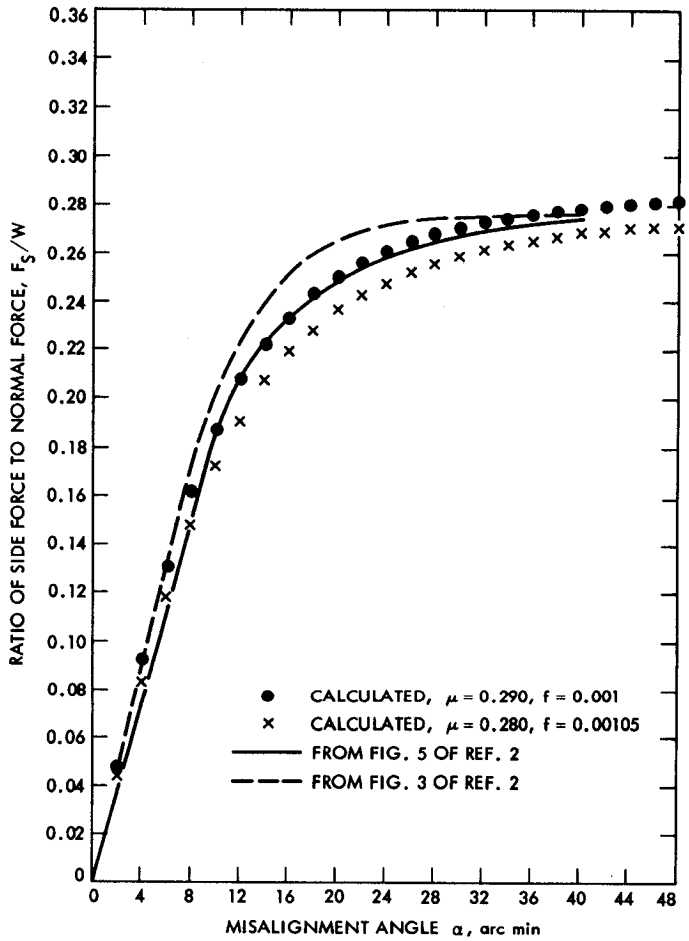


Fig. 3. Comparison of calculated lateral force coefficients with test results

Application of Radiative Transfer Theory to Microwave Transmission Medium Calibrations

C. T. Stelzried
TDA Mission Support

Precise determinations of the transmission medium loss and noise temperature contribution are important to the performance characterization of low noise microwave receiving systems and thermal noise standards. Tropospheric loss is frequently inferred from microwave radiometer noise temperature measurements. Interpretation of these measurements requires an inversion of the radiative transfer integral equation. This is inconvenient even with computer techniques. Although there have been many published studies of the radiative transfer equation, this article provides a very accurate solution in terms of a rapidly convergent power series. This solution is applicable to a low-loss medium with either uniform or nonuniform loss distributions. A four-layer atmosphere model is investigated to demonstrate the accuracy of the solution relative to the model. These solutions for total loss range in accuracy from about 0.001 dB (≃0.06 K) at 0.1 dB loss to about 0.01 dB (≃0.6 K) at 1 dB total loss. Applications include thermal noise standards and single- and dual-frequency water vapor radiometers.

I. Introduction

Precise determinations of the transmission medium loss and noise temperature contribution are important to the performance characterization of low noise receiving systems (Ref. 1) and thermal noise standards (Ref. 2). Tropospheric loss is frequently inferred from microwave radiometer noise temperature measurements (Ref. 3). Interpretation of these measurements requires an inversion of the radiative transfer integral equation (Refs. 4, 5, 6). This is inconvenient even with computer techniques. Although there have been many published studies (Refs. 7, 8, 9) of the radiative transfer equations, this article clearly details the solution and applications, including thermal noise standards and single- and dual-frequency water vapor radiometers. The solutions are applicable to a low loss medium with either uniform or nonuniform loss distributions described analytically or in discrete steps.

II. Theory

The total loss through a transmission medium (Figs. 1 or 2) is given by (neglecting scattering[1])

$$L = e^{\tau}, \text{ ratio } (\geqslant 1) \tag{1}$$

[1] Scattering should be considered for the troposphere for rain above ≃10 GHz and for clouds above ≃100 GHz (Ref. 10).

where

τ = total attenuation (optical depth), nepers

$$[\tau\,(\text{nepers})] = [L\,(\text{dB})/10 \log e] \simeq [L\,(\text{dB})/4.343]$$

$$= \int_0^{\ell} \alpha(x)\, dx$$

$= \alpha_0 \ell$ for $\alpha(x) = \alpha_0$, uniform loss

$\alpha(x)$ = absorption coefficient of the medium at x, nepers/m

x = distance along the propagation path, m ($x = 0$ at surface, $= \ell$ at "top")[2]

ℓ = total path length, m

From the theory of radiative transfer (neglecting scattering),

$$T = (T_s/L) + \int_0^{\ell} T(x)\, \alpha(x)\, e^{-\tau(x)}\, dx, \text{K} \qquad (2)$$

where

T_s = source temperature, K

$T(x)$ = physical temperature of medium at x, K

$\tau(x)$ = attenuation between 0 and x, nepers

$$= \int_0^{x} \alpha(x')\, dx'$$

This can be integrated directly or solved stepwise (Ref. 12). For small transmission loss (expanding $e^{-\tau}$ and $e^{-\tau(x)}$ in power series)

$$T \simeq T_s + A\tau + B\tau^2 + C\tau^3 + \cdots \qquad (3)$$

where

$A = T_p - T_s$

T_p = weighted physical temperature of the transmission medium, K

$$= \int_0^{\ell} T(x)\, [\alpha(x)/\tau]\, dx$$

$$B = -\left\{ \int_0^{\ell} T(x)\, [\alpha(x)/\tau]\, [\tau(x)/\tau]\, dx - (T_s/2) \right\}$$

$$C = (1/2) \int_0^{\ell} T(x)\, [\alpha(x)/\tau]\, [\tau(x)/\tau]^2\, dx - (T_s/6)$$

Solving for τ in a power series,

$$\left.\begin{aligned} \tau = [L\,(\text{dB})/4.343] &\simeq [(T - T_s)/(T_p - T_s)] \\ &- (B/A)\, [(T - T_s)/(T_p - T_s)]^2 \\ &+ [2\,(B/A)^2 - (C/A)]\, [(T - T_s)/(T_p - T_s)]^3 + \cdots \end{aligned}\right\} \qquad (4)$$

A, B and C can be solved and treated as "constants," assuming the physical temperature and loss distribution of the medium are known. For a transmission medium composed of discrete sections ($\ell = n\Delta x$ and $x = i\Delta x$)

$$T_p = \sum_{i=1}^{n} T_i\, \Delta k_i$$

$$B = -\left[\sum_{i=1}^{n} T_i\, \Delta k_i k_i - (T_s/2) \right] \qquad (5)$$

$$C = (1/2) \sum_{i=1}^{n} T_i\, \Delta k_i k_i^2 - (T_s/6)$$

where

$$\Delta k_i = \alpha_i\, \Delta x/\tau \simeq \Delta L_i(\text{dB})/L\,(\text{dB})$$

$$k_i = (\tau_i/\tau) = \sum_{j=1}^{i} \Delta k_j$$

$$\sum_{i=1}^{n} \Delta k_i = 1, (0 \leqslant k_i \leqslant 1)$$

The last terms of Eqs. (3) and (4) are small and provide an indication of the number of terms required and the accuracy

[2] It is sometimes convenient to integrate from the "top" to the surface (Ref. 11).

of the power series expansion. Only one or two terms are required for most applications. For many applications, it is suitable to use $(B/A) \simeq -(1/2)$ and $[2(B/A)^2 - (C/A)] \simeq (1/3)$ as obtained from the exact solution for a uniform loss and temperature distribution. Then

$$T \simeq T_s + (T_p - T_s)\,[\tau - (1/2)\,\tau^2 + (1/6)\,\tau^3 + \cdots] \quad (6)$$

and

$$\tau \simeq [(T - T_s)/(T_p - T_s)] + (1/2)\,[(T - T_s)/(T_p - T_s)]^2 + (1/3)\,[(T - T_s)/(T_p - T_s)]^3 + \cdots \quad (7)$$

III. Nonuniform Loss

Example 1:

Consider a two-layer nonuniform loss stratified transmission medium with $\Delta k_1 = \Delta k$ for the first (surface) layer. This implies that $\Delta k_2 = 1 - \Delta k$, $k_1 = \Delta k$ and $k_2 = 1$.

Then

$$\left.\begin{aligned} T_p &= T_1\,\Delta k + T_2\,(1 - \Delta k) \\ B &= -\,[T_1(\Delta k)^2 + T_2\,(1 - \Delta k) - T_s/2] \\ C &= (1/2)\,[T_1(\Delta k)^3 + T_2\,(1 - \Delta k)] - T_s/6 \end{aligned}\right\} \quad (8)$$

For a numerical atmospheric example, assume that $T_1 = 300$ K, $T_2 = 200$ K, $T_s = 3$ K, (cosmic background), $\Delta k = 2/3$, so that from Eq. (5)

$$T_p = 266.667 \text{ K}$$

$$B = -198.50 \text{ K}$$

$$C = 72.80 \text{ K}$$

From Eq. (3) using $L\,(\text{dB}) = 0.10$,

$$T \simeq 3 + 6.071 - 0.105 + 0.001 = 8.967 \text{ K}$$

and from Eq. (4) using $T = 8.97$ K,

$$L\,(\text{dB}) \simeq 0.0983 + 0.0017 + 0.0000 = 0.10$$

Example 2:

Consider (Fig. 3) a four-layer nonuniform loss stratified transmission medium example with T (measured) of 13 K. Assume $\Delta k_1 = 0.4$, $\Delta k_2 = 0.3$, $\Delta k_3 = 0.2$, $\Delta k_4 = 0.1$, $T_1 = 280$ K, $T_2 = 240$ K, $T_3 = 220$ K, $T_4 = 200$ K and $T_s = 3$ K, so that

$$T_p = 248.0 \text{ K}$$

$$B = -153.3 \text{ K}$$

$$C = 53.9 \text{ K}$$

and from Eq. (4)

$$L\,(\text{dB}) \simeq 0.177 + 0.004 + 0.000 = 0.181$$

This can be verified within approximately 0.001 dB ($\simeq$0.06 K), solving for T with direct stepwise calculations. Eqs. (3) and (4) have been verified with this model to within 0.01 dB ($\simeq$0.6 K) up to 1 dB overall loss. Increased accuracy requires additional terms in the expansions of Eqs. (3) and (4).

Example 3:

It is frequently convenient to perform an antenna "tipping" measurement to evaluate atmospheric loss. In this measurement technique, the difference in antenna temperature between two elevation angles is measured. This simplifies the calibration requirement from an absolute to a relative temperature measurement. Assuming an infinitely narrow antenna beamwidth and stratified atmosphere, the difference temperature between antenna angle Z (Fig. 2) and zenith is (using Eq. (3) with $T(z = 0^\circ) = T_z$ and $\tau(Z = 0^\circ) = \tau_z$)

$$\begin{aligned} T &= T - T_z \\ &\simeq A\,(\tau - \tau_z) + B\,(\tau^2 - \tau_z^2) + C\,(\tau^3 - \tau_z^3) + \cdots \end{aligned} \quad (9)$$

By analogy with Eq. (4),

$$\begin{aligned} (\tau - \tau_z) &= [L(dB) - L_z(\text{dB})]/4.343 \\ &= [\Delta T/(T_p - T_s)] - (B/A)\,[\Delta T/(T_p - T_s)]^2 \\ &\quad + [2(B/A)^2 - (C/A)]\,[\Delta T/(T_p - T_s)]^3 + \cdots \end{aligned} \quad (10)$$

If T is measured at $Z = 60^\circ$, $\tau = 2\tau_z$ and then (using $T(Z = 60^\circ) - T(Z = 0^\circ)) = \Delta T_z$, $(B/A) \simeq -(1/2)$, $[2(B/A)^2 - (C/A)] \simeq (1/3)$,

$$\tau_z \simeq [L_z(\text{dB})/4.343] \simeq [\Delta T_z/(T_p - T_s)] + (1/2)\,[\Delta T_z/(T_p - T_s)]^2 + (1/3)\,[\Delta T_z/(T_p - T_s)]^3 + \cdots \quad (11)$$

If this measurement is performed near the water vapor resonance frequency of ≃22 GHz, the integrated water vapor content in the line of sight is proportional to τ_z. Measurements of τ_z with a water vapor radiometer allow monitoring of the integrated water vapor when proper calibration constants are obtained (from weather balloon calibrations, etc.). Usually only one or two terms of Eq. 11 with $(B/A) \simeq -(1/2)$ are required. These equations can be suitably expanded to include two frequencies (Ref. 13) so that two equations with two unknowns are obtained. It is then possible to monitor both liquid and water vapor in the line of sight. A simple single-frequency "fair weather" water vapor radiometer may be of interest for some applications. Cloudy weather is "handled" by simply discarding data taken under these conditions. One- and two-frequency water vapor radiometers are analyzed in Appendices A and B.

IV. Uniform Loss

For uniform loss, $\alpha(x)\ell = \tau$ and $\tau(x) = x\tau/\ell$ (usually applicable to thermal noise standards), so that the average physical temperature of the transmission line is given by, from Eq. (3),

$$T_p = \frac{1}{\ell}\int_0^{\ell} T(x)dx = \frac{1}{n}\sum_{i=1}^{n} T_i$$

$$B = -\left[\frac{1}{\ell^2}\int_0^{\ell} xT(x)dx - (T_s/2)\right] = -\left[\frac{1}{n^2}\sum_{i=1}^{n} iT_i - (T_s/2)\right] \quad (12)$$

$$C = \frac{1}{2\ell^3}\int_0^{\ell} x^2 T(x)dx - (T_s/6) = \frac{1}{2n^3}\sum_{i=1}^{n} i^2 T_i - (T_s/6)$$

These solutions for T_p, B and C for uniform loss are directly applicable to the solutions for T and L(dB), Eqs. (3) and (4).

Example 4:

Consider the solution of Eq. (3) for a linear physical temperature distribution, $T(x) = T_1 + (T_2 - T_1)x/\ell$ (applicable to a thermal noise standard consisting of a source at temperature T_s and a transmission line with a linear temperature distribution between T_1 and T_2). We have

$$\left.\begin{aligned} T_p &= (T_1 + T_2)/2 \\ B &= -(T_1 + 2T_2 - 3T_s)/6 \\ C &= (T_1 + 3T_2 - 4T_s)/24 \end{aligned}\right\} \quad (13)$$

resulting in

$$\left.\begin{aligned} T \simeq\ & T_s + [L(\text{dB})/4.343]\,(T_1 + T_2 - 2T_s)/2 \\ & - [L(\text{dB})/4.343]^2\,(T_1 + 2T_2 - 3T_s)/6 \\ & + [L(\text{dB})/4.343]^3\,(T_1 + 3T_2 - 4T_s)/24 + \cdots \end{aligned}\right\} \quad (14)$$

in agreement with Ref. 2, p. 648. Also

$$[L(\text{dB})/4.343] \simeq [2(T - T_s)/(T_1 + T_2 - 2T_s)] + 4\,(T - T_s)^2\,(T_1 + 2T_2 - 3T_s)/3\,(T_1 + T_2 - 2T_s)^3 + \cdots \quad (15)$$

V. Conclusion

The radiative transfer equation has been evaluated assuming known transmission medium temperature and loss distributions. A rapidly convergent power series solution is given for the total transmission medium loss in terms of these distributions and the measured radiometric noise temperature. A four-layer atmospheric model is investigated to demonstrate the accuracy of the solution relative to the model. These solutions for total loss range in accuracy from about 0.001 dB at 0.1 dB to about 0.01 dB at 1 dB total loss. Applications include thermal noise standards and single and dual frequency water vapor radiometers.

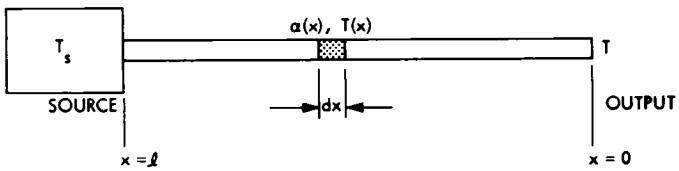


Fig. 1. Representation of a thermal noise standard consisting of a source and a lossy transmission line

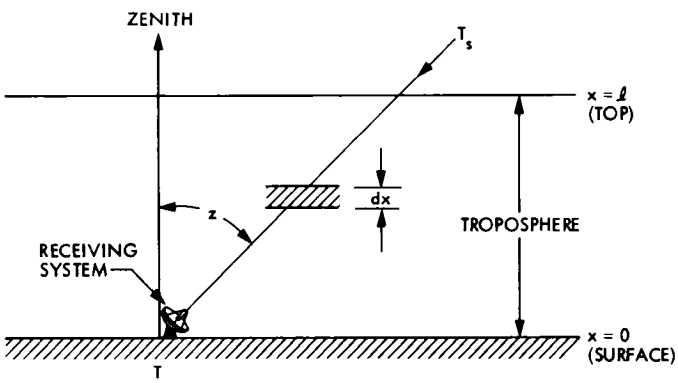


Fig. 2. Representation of a receiving system with signal propagating through a lossy medium

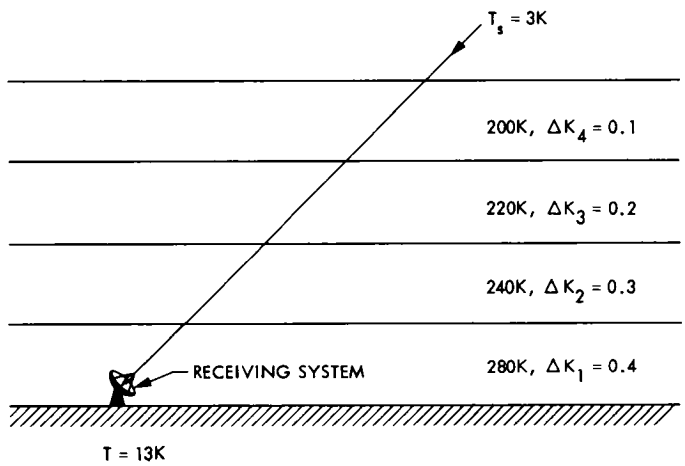


Fig. 3. Representation of a four-layer nonuniform loss atmospheric model for example calculation

Appendix A

Single-Frequency Water Vapor Radiometer

The single-frequency water vapor radiometer is useful for clear weather tropospheric water vapor density calibrations. A stratified troposphere and pencil beam antenna pattern with negligible sidelobes is assumed. The total attenuation of zenith angle Z in terms of the increased antenna noise temperature due to the troposphere is

$$\tau = [(T - T_s)/(T_p - T_s)] - (B/A)\,[(T - T_s)\,(T_p - T_s)]^2 + \cdots \qquad \text{(A-1)}$$

This loss consists of a stable oxygen component τ_0 [τ_0(zenith) = τ_{0z}] assumed known, ≃0.124 nepers at 20.6 GHz (Ref. 13, p. 38) assuming 5.4 km scale height and an unknown water vapor component τ_V (assuming no clouds or rain)

$$\tau \simeq \tau_V + \tau_0 \qquad \text{(A-2)}$$

The precipitable water vapor along the line of sight (for a 2-km scale height troposphere with 7.5 gm/m^3 water vapor density, M_V = 1.5 cm at zenith) is given by

$$M_V = \tau_V / k_V \qquad \text{(A-3)}$$

where

k_V = normalized water vapor attenuation coefficient per unit precipitable water, nepers/cm ($k_V \simeq 0.0426$ at 20.6 GHz; k_V has greater uncertainty at 22.2 GHz, Ref. 4)

Combining Eqs. (A-2) and (A-3),

$$M_V = (\tau - \tau_0)/k_V \qquad \text{(A-4)}$$

In terms of the measured antenna temperature from Eq. (A-1),

$$M_V = a_1(T - T_s) + a_2(T - T_s)^2 + a_3, \text{ cm} \qquad \text{(A-5)}$$

where (assuming $(A/B) \simeq -(1/2)$, as in Eq. 7)

$$a_1 \simeq 1/k_V\,(T_p - T_s)$$

$$a_2 \simeq a_1/2\,(T_p - T_s)$$

$$a_3 \simeq -\tau_0/k_V$$

Using the "tipping" measurement technique of Example 3,

$$M_V \simeq (\cos Z/k_V)\,\{[\Delta T_z/(T_p - T_s)] + [\Delta T_z/(T_p - T_s)]^2 - \tau_{0z}\}, \text{ cm} \qquad \text{(A-6)}$$

As an example, at 20.6 GHz, $(T_p - T_s) \simeq 280$ K, $k_V \simeq 0.0426$ (Ref. 14, p. 479), $\tau_{0z} \simeq 0.0124$ (Ref. 13, p. 38, assuming 5.4-km scale height), so that

$$M_V \simeq \cos Z\,[0.084\,\Delta T_z + 0.00015\,(\Delta T_z)^2 - 0.29], \text{ cm} \qquad \text{(A-7)}$$

The precipitable water vapor content at zenith angle Z is monitored by repeated "tipping" measurements or by switching between horns mounted at suitable zenith angles. For clear sky conditions, this technique has the simplicity and accuracy of a relative measurement of ΔT_z as compared to the difficulty of the absolute measurements required for the two-frequency instruments described in Appendix B.

The measurement error for M_V (neglecting the effects of a nonstratified trospheric pointing error, etc.) is

$$\delta(M_V) \simeq [\cos Z\,0.084\,\delta(\Delta T_z)^2 + (\delta a_3)^2]^{1/2} \qquad \text{(A-8)}$$

Using $Z = 60°$, $\delta(\Delta T_z) \simeq 1$ K and $\delta a_3 \simeq 0.12$ (≃20%), results in $\delta(M_V) \simeq 0.2$ cm. The measurement error for neglecting the $(\Delta T_z)^2$ term for $\Delta T_z \simeq 30$ K is ≃0.27 cm out of ≃2.3 cm or ≃12%.

The increase in propagation path length due to the total precipitable water vapor is (Ref. 15)

$$\Delta \ell_V \simeq 6.48\,M_V, \text{ cm} \qquad \text{(A-9)}$$

or, in terms of ΔT_z, from Eq. (A-7),

$$\Delta \ell_V \simeq \cos Z\,[0.54\,\Delta T_z + 0.00010\,(\Delta T_z)^2 - 1.9], \text{ cm} \qquad \text{(A-10)}$$

Although the single-frequency water vapor radiometer will perform well in clear sky conditions, serious performance degradation occurs during cloudy weather. Some applications allow the selection of good data, while discarding poor data.

Appendix B

Dual-Frequency Water Vapor Radiometer

The dual-frequency water vapor (Refs. 16, 17) radiometer is useful during cloudy weather to determine both tropospheric water vapor and liquid content. A pencil beam antenna pattern with negligible sidelobes is assumed. The equations for determination of the total attenuation obtained from radiometer noise temperature calibrations for frequencies f_1 and f_2 are,

$$\tau_1 \simeq [(T_1 - T_s)/(T_p - T_s)] - (B/A)\,[(T_1 - T_s)/(T_p - T_s)]^2 + \cdots, \text{nepers} \tag{B-1}$$

and

$$\tau_2 \simeq [(T_2 - T_s)/(T_p - T_s)] - (B/A)\,[(T_2 - T_s)/(T_p - T_s)]^2 + \cdots, \text{nepers} \tag{B-2}$$

Using subscripts V, L and 0 for water vapor, water liquid, and oxygen,

$$\tau_1 \simeq \tau_{V1} + \tau_{L1} + \tau_{01}, \text{nepers} \tag{B-3}$$

and

$$\tau_2 \simeq \tau_{V2} + \tau_{L2} + \tau_{02}, \text{nepers} \tag{B-4}$$

The total precipitable water vapor through the tropospheric line of sight is

$$\left.\begin{aligned} M_V &= \tau_{V1}/k_{V1} \\ &= \tau_{V2}/k_{V2} \end{aligned}\right\} \text{ cm} \tag{B-5}$$

and the total precipitable water liquid is

$$\left.\begin{aligned} M_L &= \tau_{L1}/k_{L1} \\ &= \tau_{L2}/k_{L2} \end{aligned}\right\} \text{ cm} \tag{B-6}$$

where

K_V, k_L = proportionality constants relating precipitable water to attenuation, nepers/cm

Also,

$$\tau_{V2}/\tau_{V1} = k_{V2}/k_{V1} = K_V \tag{B-7}$$

and

$$\tau_{L2}/\tau_{L1} = k_{L2}/k_{L1} = K_L \tag{B-8}$$

Combining and solving for M_V and M_L, in terms of T_1 and T_2, assuming all constants are "known,"

$$M_V \simeq a_1 (T_1 - T_s) + a_2 (T_1 - T_s)^2 + a_3 (T_2 - T_s) + a_4 (T_2 - T_s)^2 + a_5, \text{cm} \tag{B-9}$$

and

$$M_L \simeq b_1 (T_1 - T_s) + b_2 (T_1 - T_s)^2 + b_3 (T_2 - T_s) + b_4 (T_2 - T_s)^2 + b_5, \text{cm} \tag{B-10}$$

where (assuming $(B/A) \simeq -(1/2)$ as in Eq. 7),

$$a_1 = -k_L/kk_{V1}\,(T_p - T_s) \qquad b_1 = k_V/kk_{L1}\,(T_p - T_s)$$

$$a_2 = a_1/2(T_p - T_s) \qquad b_2 = b_1/2(T_p - T_s)$$

$$a_3 = 1/kk_{V1}\,(T_p - T_s) \qquad b_3 = -1/kk_{L1}\,(T_p - T_s)$$

$$a_4 = a_3/2(T_p - T_s) \qquad b_4 = b_3/2\,(T_p - T_s)$$

$$a_5 = -(\tau_{02} - k_L\,\tau_{01})/kk_{V1} \qquad b_5 = (\tau_{02} - k_V\,\tau_{01})/kk_{L1}$$

$$k = k_V - k_L$$

This allows monitoring of M_V and M_L from measurements of T_1 and T_2. S. C. Wu (Ref. 17) has investigated optimum frequency selection. The constants in Eqs. (B-9) and (B-10) can either be evaluated from the definitions above or from direct tropospheric calibrations (from radiosonde balloons, etc.). For estimates of these constants, use the same values as in the example (or Appendix A for f_1 = 20.6 GHz and for f_2 = 31.6 GHz, $k_{V2} \simeq 0.0256$ (Ref. 14), $k_{L1} \simeq 0.743$ (Ref. 16), $k_{L2} \simeq 1.75$ (Ref. 16), $\tau_{02z} \simeq 0.0224$ (Ref. 13)) to obtain[3]

$$M_V \simeq \left.\begin{array}{l} 0.11\,(T_1 - T_s) + 0.00020\,(T_1 - T_s)^2 \\ -0.048\,(T_2 - T_s) - 0.000086\,(T_2 - T_s)^2 \\ -0.064\cos Z \end{array}\right\} \text{cm} \qquad \text{(B-11)}$$

[3]Hogg (Ref. 18) has $M_V \simeq 0.11\ T_1 - 0.053\ T_2 - 0.18$ and $M_L \simeq -0.0011\ T_1 + 0.0027\ T_2 - 0.17$ appropriate for the zenith climatology of Denver, Colorado. The biggest difference between these expressions is the constants term for M_L (accounting for $T_s \simeq 2.7$ K).

and

$$M_L \simeq \left.\begin{array}{l} -0.0016\,(T_1 - T_s) - 0.0000029\,(T_1 - T_s)^2 \\ +0.0027\,(T_2 - T_s) + 0.0000048\,(T_2 - T_s)^2 \\ -0.013\cos Z \end{array}\right\} \text{cm} \qquad \text{(B-12)}$$

The increase in propagation path length due to both water vapor and liquid water is (Ref. 15)

$$\Delta\ell \simeq 6.48\,M_V + 1.45\,M_L, \text{ cm} \qquad \text{(B-13)}$$

or, in terms of T_1 and T_2, for this example,

$$\Delta\ell \simeq \left.\begin{array}{l} 0.71\,(T_1 - T_s) + 0.0013\,(T_1 - T_s)^2 \\ -0.31\,(T_2 - T_s) - 0.00056\,(T_2 - T_s)^2 \\ -0.43\cos Z \end{array}\right\} \text{cm} \qquad \text{(B-14)}$$

Inspection of the above equations indicates that most of the tropospheric delay is due to the water vapor and very little from the liquid water. The primary effect of the liquid water is to alter the noise temperature measurements.

If it is required that the constants of Eqs. (B-13) and (B-14) be determined from direct radiometer calibrations, constants a_2, a_4, b_2 and b_4 might best be determined analytically as in the example. This is suggested due to the difficulty of direct calibration of these small second-order terms.

Acknowledgment

S. Slobin provided assistance and encouragement.

References

1. Reid, M. S., et al., "Low-Noise Microwave Receiving Systems in a Worldwide Network of Large Antennas," *Proc. of the IEEE,* Vol. 61, No. 9, Sept. 1973.

2. Stelzried, C. T., "Microwave Thermal Noise Standards," *IEEE Trans. Microwave Theory and Techniques*, Vol. MTT-16, No. 9, Sept. 1968, p. 646.

3. Slobin, S. D., "Microwave Noise Temperature and Attenuation of Clouds at Frequencies Below 50 GHz," *Publication 81-46*, Jet Propulsion Laboratory, Pasadena, Calif., July 1, 1981.

4. Chandrasekhar, S., *Radiative Transfer*, Dover Publications, N.Y. 1960.

5. Stelzried, C. T., "Calculation of Atmospheric Loss From Microwave Radiometric Noise Temperature Measurements," *TDA Progress Report 42-62*, Jet Propulsion Laboratory, Pasadena, Calif., April 15, 1981, pp. 73-80.

6. Stelzried, C. T., "Atmospheric Noise Temperature Measurements," *TDA Progress Report 42-63*, Jet Propulsion Laboratory, Pasadena, Calif., June, 1981, pp. 87-96.

7. Staelin, D. H., "Passive Remote Sensing at Microwave Wavelengths," *Proc. IEEE*, Vol. 57, No. 4, April 1969, p. 427.

8. Chahine, M. T., "Inverse Problems in Radiative Transfer: Determination of Atmospheric Parameters," *Atmospheric Sciences*, Vol. 27, No. 6, Sept. 1970, pp. 960-967.

9. Rosenkranz, P. W., et al., "Microwave Radiometeric Measurements of Atmospheric Temperature and Water from an Aircraft," *J. Geophys. Res.*, Vol. 77, No. 20, Oct. 1972, p. 5833.

10. Tsang, L., et al., "Theory of Microwave Emission From a Layer of Cloud or Rain," *IEEE Trans. Antennas and Propagation*, AP-25, No. 5, Sept. 1977, pp. 640-657.

11. Waters, J. W., "Absorption and Emission by Atmospheric Gases," *Methods of Experimental Physics*, Vol. 12 Academic Press, N.Y., 1976.

12. Stelzried, C. T., "A Liquid Helium Cooled Coaxial Termination," *Proc. IRE*, Vol. 49, No. 7, July 1961, p. 1224.

13. Smith, E. K., and Waters, J. W., "Microwave Attenuation and Brightness Temperature Due to the Gaseous Atmosphere," *Publication 81-81*, Jet Propulsion Laboratory, Pasadena, Calif., August 1981.

14. Westwater, E. R., "The Accuracy of Water Vapor and Cloud Liquid Determination by Dual-Frequency Ground-Based Microwave Radiometry," *Radio Science*, Vol. 13, No. 4, July-August 1978, pp. 677-685.

15. Flock. W. L., "Effects of the Gaseous and Liquid Water Content of the Atmosphere on Range Delay and Doppler Frequency," *TDA Progress Report 42-63*, Jet Propulsion Laboratory, Pasadena, Calif., June 15, 1981, pp. 71-86.

16. Slobin, S. D., "DSN Water Vapor Radiometer – Tropospheric Range Delay Calibration," *TDA Progress Report 42-49*, Jet Propulsion Laboratory, Pasadena, Calif., Feb. 15, 1979, p. 136.

17. Wu, S. C., "Frequency Selection and Calibration of a Water Vapor Radiometer," *DSN Progress Report 42-43*, Jet Propulsion Laboratory, Pasadena, Calif., Feb. 15, 1978, pp. 67-81.

18. Hogg, D. C., "Ground-Based Remote Sensing and Profiling of the Lower Atmosphere Using Radio Wavelengths," *IEEE Trans. Antennas and Propagation*, Vol. AP-28, No. 2, March 1980, pp. 281-283.

Provision of 9.6-kbps Wideband Data Rate Capability in the DSN

G. J. Brunder
DSN Data Systems Section

This article provides an overview of the new 9.6-kbps wideband data rate capability in the DSN. A functional description of the completed implementation is presented, together with a plan to upgrade the JPL Central Communications Terminal for additional 9.6 kbps operational flexibility.

I. Introduction

In the latter part of 1981, budget constraints caused the NASA Communications Network (NASCOM) to reevaluate the service it provides to the NASCOM users. JPL, being a major user of the NASCOM network, was approached by NASCOM to evaluate the feasibility of replacing the overseas 56-kbps wideband (WBD) data circuits with 9.6-kbps WBD data circuits. JPL recognized that NASCOM would achieve major cost savings if this could be accomplished. JPL therefore conducted studies to evaluate whether its ongoing projects could be adequately served by utilizing 9.6-kbps WBD service.

JPL completed its studies in January 1982. JPL concluded that its projects could be adequately supported by the new 9.6-kbps WBD data rate with occasional supplementary use of overseas NASCOM 56-kbps circuits. The 56-kbps circuits would be time-shared by both NASCOM and JPL on a scheduled basis.

II. Results of JPL's Project Assessments to Replace 56-kbps Circuits with 9.6-kbps Circuits

The Voyager Project and the DSN's Very Long Baseline Interferometry (VLBI) differential one-way range (DELTA DOR) data transfer were the only users affected by the WBD rate change.

A. Voyager

The Voyager Project concluded that 56-kbps WBD service could be replaced on a day-to-day basis by 9.6-kbps WBD service while Voyager is in its post-Saturn cruise mode in the 1982-85 era. Eighty percent of its telemetry requirements are for 7.2-kbps spacecraft data rates. Real-time data requirements in excess of 7.2 kbps could be forecast one to two months in advance to aid in scheduling available overseas NASCOM lines on a shared basis. The high-volume Voyager navigation DELTA DOR data could be scheduled in the same manner.

B. VLBI

Both the VLBI Mark I System and the VLBI Mark II System concluded that they could schedule the NASCOM overseas 56-kbps WBD circuits on a shared basis for their high-volume data. The VLBI Mark I System provides for time synchronization, Universal Time One, and polar motion data requirements. The VLBI Mark II System provides delta differential one-way range (DELTA DOR) capability data for flight project navigation support.

III. Significant Functional Differences Between Old and New Overseas DSN Wideband Subsystems

Refer to Figs. 1 and 2, which depict the DSN/NASCOM 9.6-kbps wideband configurations.

A. Old Overseas DSS 42/43 and DSS 61/63 56-kbps Wideband Data Assembly (WBDA)

The WBDA received block-formatted, serial data from the on-site computers (OSC) and conditioned the data for transmission by WBD equipment. The block-encoded data was applied to a 303-type data set which converted it to a baseband signal and inputted it to an LWM-6 type modem. The modem translated the baseband signal into vestigial sideband (VSB) signal in the range of 60-108 kHz. This form of signal is known as group-band. The group-band signals were then applied to a 56-kbps NASCOM line to the Madrid or Canberra Switching Centers for retransmission over satellite communication links to the Goddard Space Flight Center (GSFC). From GSFC, the signals were retransmitted via a 56-kbps satellite link to the JPL Central Communications Terminal (CCT) WBD assembly terminal equipment.

B. New DSS 42/43 9.6-kbps Wideband Equipment

The old 56-kbps equipment remains unchanged. NASCOM has provided new 9.6-kbps 209-type data sets and current-to-voltage (I/V) signal converters to serve the 9.6-kbps circuits described in Section III-A above.

The block-encoded data is now applied to the I/V converters for conversion to a voltage-mode signal that can be interfaced to the 9.6-kbps 209-type data sets. The data set converts it to an analog signal which is routed, via voice/data switching equipment, onto a NASCOM voice/data circuit for transmission to the Canberra Switching Center. The Canberra Switching Center has been configured by NASCOM to allow for two methods to route the signal to GSFC. It can either regenerate the signals with 9.6-kbps 209-type data sets for transmission to GSFC via voice/data circuits, or it can multiplex the signals onto a 56-kbps Time Division Multiplex (TDM) System for transmission via a satellite link to GSFC. GFSC transmits the signals to JPL via a 56-kbps satellite link or by discreet D-1 conditioned voice/data lines using 9.6-kbps 2096 data sets.

C. New DSS 61/63 9.6-kbps Wideband Data Equipment

The old 56-kbps equipment remains unchanged. NASCOM has provided line drivers capable of driving the new 9.6-kbps signals to the Madrid Switching Center. The line drivers accept the block-encoded data described in Section III-A and converts the data into balanced current-mode signals for transmission, via existing NASCOM cable facilities, to the Madrid Switching Center. The Madrid Switching Center has the same capabilities as the Canberra Switching Center for alternate data transmission to GSFC as described in Section III-B.

IV. 9.6-kbps Implementation Completed in the JPL CCT Central Wideband (CWB) Assembly

NASCOM has provided three new 9.6-kbps 2096 data sets, on a leased basis, and I/V converters. The analog sides of the 2096 data sets have been interfaced via the CCT voice/data switcher to two D-1 conditioned lines to receive transmission from GSFC. GSFC receives the signals from the overseas switching centers and stations as previously described.

The RS-232 digital sides of the data sets at JPL are applied to I/V converters where the signals are converted to balanced current mode signals for input to the CWB terminal equipment, and thence to the CCT computers and Mission Control and Computer Center (MCCC) users.

V. Plan for Upgrading the 9.6-kbps Wideband Capability in the JPL CCT

The following plan will provide the capability to either multiplex 9.6-kbps signals onto a 56-kbps line, or to route them directly to 9.6-kbps 2096 data sets for transmission over discreet D-1 conditioned Voice/Data lines.

(1) Convert the existing NASCOM-provided WBD prime and backup TDMs to enable them to multiplex 9.6-kbps circuits. Presently, the TDMs multiplex seven 7.2-kbps high speed data (HSD) channels onto a 56-kbps circuit. The planned conversion will allow multiplexing three 7.2-kbps HSD channels and three 9.6-kbps channels. All six of these channels will be interfaced to the line side (data communications equipment side) of the Error Correction and Switching (ECS) Assembly HSD switch.

(2) Interface the existing 9.6-kbps 2096 data sets to the line side of the HSD switch.

(3) Interface the equipment side (data terminal equipment side) of the HSD switch directly to the existing unused RS-232 ports of the WBD Network Encoders/Decoders (NEDs) which, in turn, are cabled directly to the existing WBD terminal equipment. The WBD terminal equipment interfaces both the CCT computers and MCCC customer equipment.

VI. Summary and Conclusions

JPL, being a major user of the worldwide NASCOM Communications Network, is successfully supporting the data requirements of its projects at a new nominal data rate of 9.6 kbps. JPL and NASCOM have planned the time-shared use of the NASCOM overseas 56-kbps circuits when high-volume Voyager Project or VLBI data require their use.

To sum up, NASCOM and JPL have achieved major cost savings in an era of increased budgetary restraints.

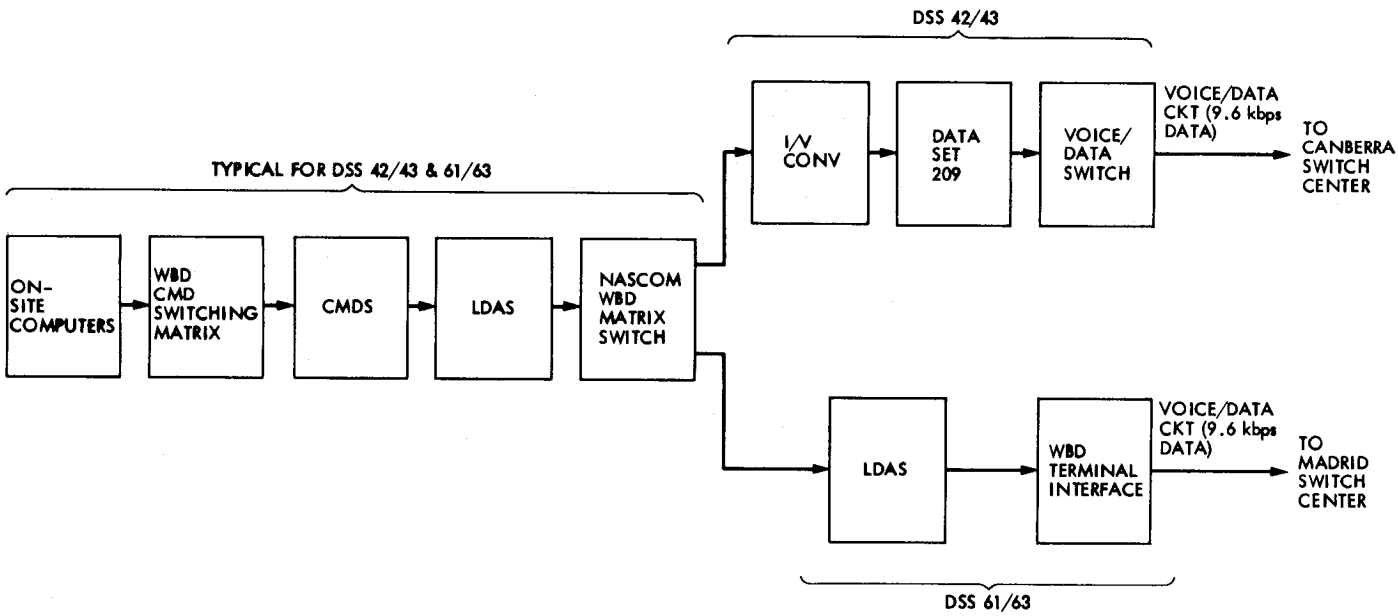


Fig. 1. Overseas station 9.6-kbps wideband configuration

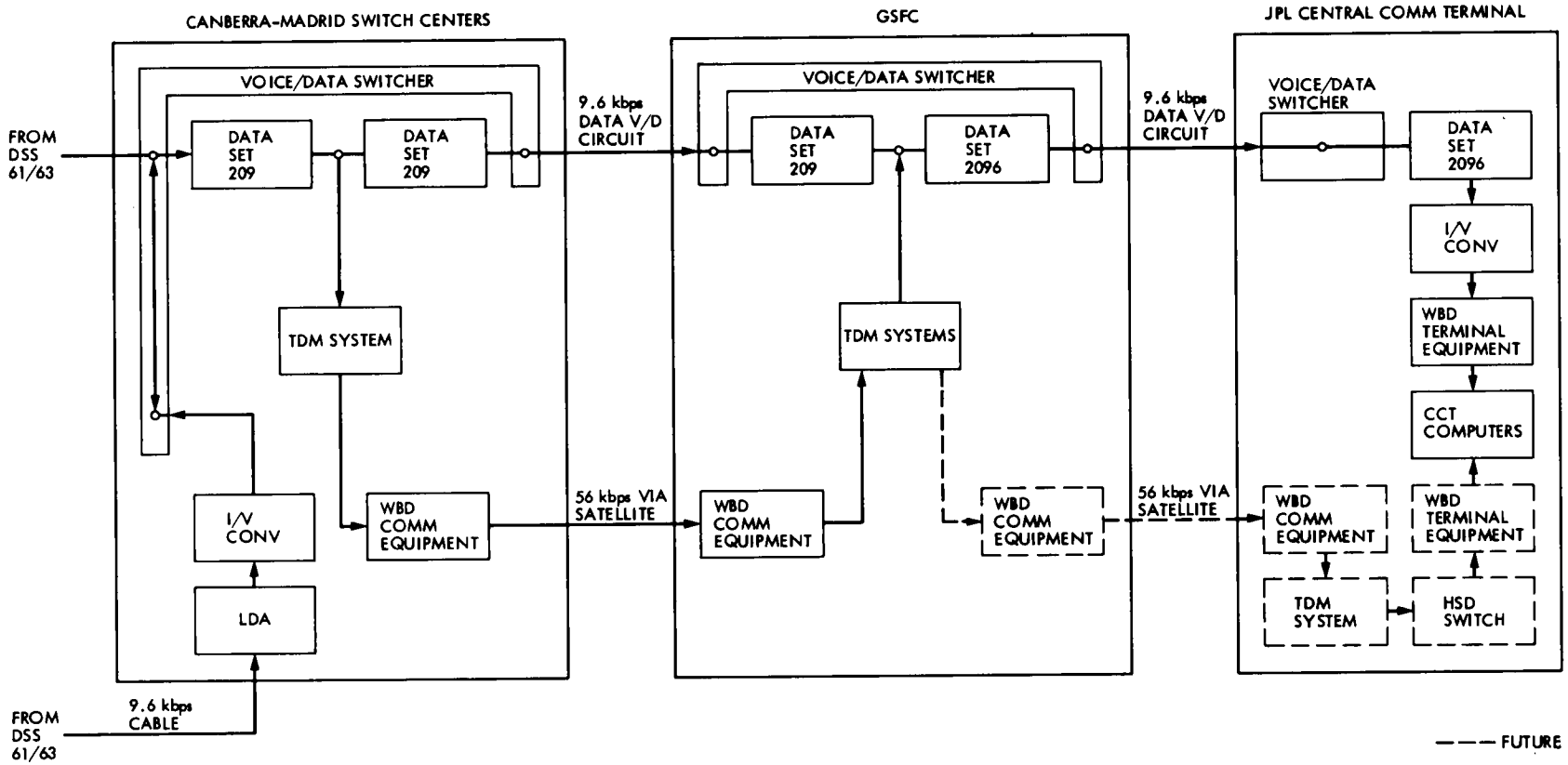


Fig. 2. 9.6-kbps wideband configuration switching centers/GSFC/JPL

Field Repair of Microprocessor-Based Equipment

A. Burford
Deep Space Network Support Section

This article describes the methods in the Deep Space Network to support the maintenance of microprocessor based devices. This effort is focused on testing by in-circuit emulation and the ability to generate executable microdiagnostics in the field. A further enhancement of this effort is the generation of more comprehensive diagnostics which may be down-loaded from a remote location. This article also points out the need for a change in past training philosophy when dealing with microprocessors.

I. Introduction

In 1971 the Intel Corporation introduced the first microprocessor, the 4004, a 4-bit processor. One year later Intel Corporation introduced the 8008, an 8-bit processor. A major improvement on these two devices resulted in the 4040 and the 8080, which provided external interrupt capabilities necessary for real-time processing. In 1978 Intel Corporation introduced the 8086, a 16-bit processor offering many features and capabilities of minicomputers.

The first microprocessors to appear in the Deep Space Station Tracking Network were (1) a 4040 resident in the Block III Subcarrier Demodulator Assembly Controller and (2), an 8080 resident in the Block IV Receiver/Exciter Controller. Microprocessors currently in use in the Deep Space Network are resident in the following equipments:

(1) Real-time combiner 8080
(2) Precision power monitor 8080
(3) Spectral signal indicator 8080
(4) Configuration control group 8080
(5) Hydrostatic bearing monitor 8048
(6) Digital tone extractor 8085
(7) Subreflector control assembly Z80
(8) Block III subcarrier demodulator assembly 4040
(9) Digital controlled oscillator 8086

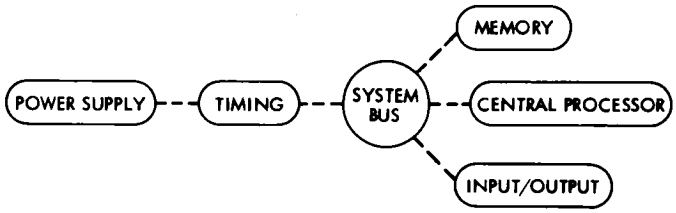
Projected estimates for the Network Consolidation Era include both 8- and 16-bit microprocessors with a combined total usage in excess of 100 controllers at each complex.

II. Background

Prior to the introduction of microprocessors, the Deep Space Station equipment complement included mini and mainframe computers for the purpose of station monitor, telemetry tracking, command processing and antenna pointing. The typical maintenance methodology in these systems is to load diagnostics from an external media and receive reports on an input/output device. Detailed micro diagnostic routines are entered via the system control panel. The system control panel provides maintenance personnel with an in-depth visibility into the system. Training on these processors invariably consists of detailed instruction on specific equipment design, yielding a specialist on a single piece of equipment.

III. Microprocessors

Microprocessor-based equipments present a special maintenance problem because of the lack of visibility into the system caused by the absence of a control panel. When a failure occurs, internally resident diagnostics report to resident indicators or to a remote device such as an input/output terminal. This system of fault isolation may be satisfactory if the kernel of the device is operational. (The kernel is defined as that part of a system that must be functional for a diagnostic to provide useful results.) The operational procedure to establish a usable kernel is depicted as follows, in order from left to right:



What is desirable for troubleshooting the kernel is a control panel. The Deep Space Network Maintenance Program currently uses a microsystem analyzer which qualifies as an "intelligent control panel" in its simplest application.

The microsystem analyzer requires in-circuit emulation (removing the central processor chip in the unit under test and inserting the analyzer in its place) to take full advantage of its capabilities. This instrument provides front panel visibility into the system, plus the following attributes:

(1) Signature analysis.

(2) Time domain analysis.

(3) Frequency and events counting.

(4) The ability to execute specially prepared diagnostics nonresident in the unit under test. These nonresident diagnostics may be either in read-only memory or they may be in random access memory which is resident in the test instruments.

Diagnostics resident in a program development system may be down-loaded into the test instrument's random access memory and executed to exercise the unit under test.

Support for repair of microprocessor-based assemblies in the Deep Space Network is focused on several pieces of equipment as follows:

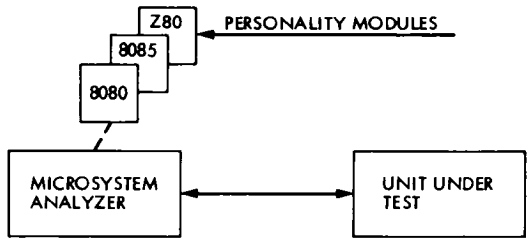
(1) Diagnostic Program Development System:
Micro/sys DS22
Tektronix LP8200 line printer
Tektronix 4025 display terminal

(2) Programmable read-only memory programmer
Data I/O system 19

(3) Portable analyzer
Millenium microsystem analyzer

(4) Test fixture
Special-purpose to accommodate JPL-designed wire wrap modules.

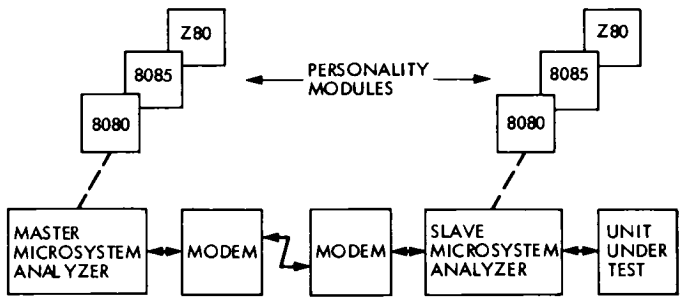
The following diagrams illustrate the various configurations of the available equipment for the support of microprocessor testing. It should be noted that these devices do not necessarily preclude the need for standard test equipment.



Configuration 1. Portable instrument (local control)

A. Diagnostics for unit under test are resident in replaceable, programmable read-only memory on the microsystem analyzer.

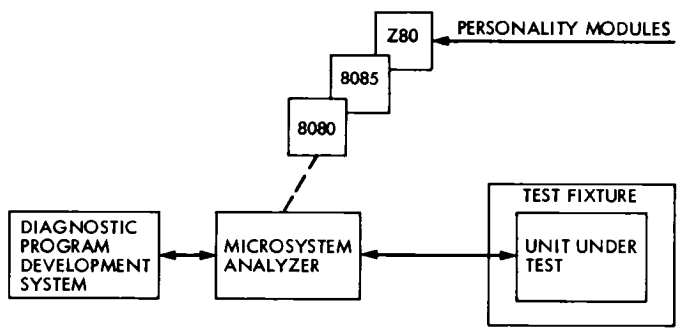
B. Diagnostics may be entered manually into the analyzer and executed from either the analyzer or the unit under test.



Configuration 2. Portable instrument (remote control)

A. Same features as Configuration 1 except all control is from the master unit.

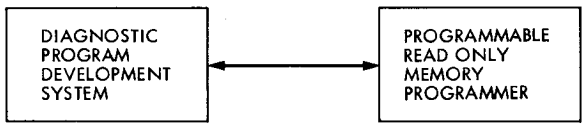
B. Master unit may be resident at some centralized maintenance facility.



Configuration 3. Laboratory instrument

A. Develop diagnostics which are down-loaded to the microsystem analyzer for exercising the unit under test (may be defective unit returned from field service).

B. General repair of defective field units.



Configuration 4. Diagnostic development

A. Down-load diagnostics into PROM Programmer for programming PROMS used in microsystem analyzer.

B. Generate diagnostic file on floppy disks for distribution to other users.

IV. Conclusions

Current microprocessor support capabilities through in-circuit emulation exist in the Deep Space Network for the following processor chips: 8080, 8085, Z80; this type of testing is thus limited to 8-bit microprocessors. Similar capabilities are being investigated for the support of 16-bit microprocessors. Initial emphasis will be on the support of the 8086 based devices.

Constant training of maintenance personnel will be required to keep abreast of the proliferation of microprocessors. An overall system understanding is essential, with training in the technology of microprocessor-based systems rather than specific detailed equipment design.

Goldstone Radio Spectrum Signal Identification March 1980–March 1982

B.A. Gaudian
Goldstone Operations Section

The Goldstone radio spectrum environment contains countless signals that are a potential source of electromagnetic interference to the Goldstone tracking receivers. The identification of these signals is accomplished by the use of signal parameters and environment parameters. This article describes the signal identification process and provides statistical data on the Goldstone radio spectrum environment from 2285 to 2305 MHz.

I. Introduction

Goldstone is located within one of the largest military training areas in the United States. In addition there are several military installations very near to Goldstone. Most of the military installations operate ground and airborne equipment that may be potential sources of electromagnetic interference (EMI). To assist in the identification of the potential sources of EMI, the Goldstone Radio Spectrum Surveillance Station provides data on the signals detected above its threshold in the frequency band being monitored.

The Goldstone Radio Spectrum Surveillance Station is a wideband signal presence indicator used to monitor the Goldstone radio frequency environment for the presence of signals in and near the deep space frequency band. The surveillance station was upgraded in March 1980 to provide real-time radio spectrum monitor capability. Subsequent to the upgrade, additional improvements were made to increase the sensitivity of the system during the relocation of the equipment to the Goldstone Radio Spectrum Surveillance Site. Since the upgrade, thousands of signals have been detected and identified with the aid of the information gathered by the Radio Spectrum Surveillance Station. This report is intended to explain in general terms portions of the signal identification process developed by the Goldstone Radio Spectrum Office and to present statistical data on the Goldstone radio frequency environment in the 2285 to 2305 MHz band.

II. Signal Identification Process

The actual signal identification process ranges from immediate identification to "unable to identify." The priority with which one attempts to identify a detected signal also varies from immediate, in the case of a signal in the DSN band (2290-2300 MHz) that is interfering with or has the potential to interfere with a Goldstone station, to a lower priority for those signals which are near the DSN band (2285-2290 and 2300-2305 MHz).

The signal identification process for the signals detected by the Radio Spectrum Surveillance Station relies on human interaction to correlate and interrogate signal parameters and environmental parameters. The most significant of the signal parameters are frequency, time, bandwidth, azimuth and signal magnitude. The most useful environmental parameters are known scheduled activities, previously identified sources and unscheduled airborne activity within line of sight of Goldstone.

It is important in the signal identification process to make an early determination to classify the signal into one of the following categories:

(1) Satellite (SAT)

(2) Airborne telemetry (TM)

(3) Airborne electronic countermeasure (ECM)

(4) Fixed location

To determine one of these classes of signals may require considerable study and in many cases the easiest method is by eliminating the categories that are easy to identify. Experience indicates that there are certain leading characteristics about each category that will aid in the identification process. These characteristics (see Table 1) will generally hold true.

As you can see in Table 1 there are many things in common; for example, a satellite could look similar to airborne electronic countermeasure if the countermeasure signal had a magnitude of -115 to -123 dBm. This is when the environment parameters become important. A check of the satellite schedules for the appropriate frequency could validate the presence of a satellite during the time in question. A check of the scheduled airborne activity at the azimuth and time in question could be helpful.

The most difficult signals to identify are those on a single frequency that are not reflected in any frequency management records as an assigned frequency. In these cases when there are no scheduled or potential users, the identification follows two paths. First, attempts are made to determine if activity may have been taking place at the azimuth in question by calling potential user locations. The parameters taking precedence in these cases are time and azimuth. Operations personnel will usually have some information concerning time and location but will seldom have information concerning the emission characteristics of equipment. Secondly, additional effort is expended on the documentation of the identification effort. If the process of the two paths fails to identify the origin of the signal, the signal is classified as unidentified and carried in an abeyance file for possible future identification or correlation.

To assist in the signal identification process, selected signals that are frequently detected by the surveillance station are identified as they are detected. This is a first level cull with frequency being used to identify the signal. This method can produce some errors in signal identification, but to date has proven to be adequate and correct over 99% of the time.

III. Signal Identification Statistics

The information in Table 2 represents a breakdown of the various categories of the signals sighted by the surveillance station. The total sightings in each category provide an indication of their relative duration. The number of incidents is an indicator of either a number of sightings in a satellite orbit, a large number of sightings in a 2-hour shuttle test or a single sighting for an activity in which an aircraft may be within line of sight of the surveillance station for an instant.

The quantity and hence the duration of sightings range in time from one sighting for an incident to a maximum of 4 hours. The variance in the duration may be due to emitter on and off times, emitter antenna altitude and line of sight conditions.

IV. Conclusion

The surveillance station described is a highly useful tool to aid in the identification and confirmation of signals that are potential EMI threats to the Goldstone stations. As each new source is identified, the appropriate coordination channels are established to preclude or minimize EMI. The ideal situation would be that the knowledge of the RF environment will be maximized to the extent that coordination would eliminate unidentified EMI except for occasional equipments with unknown emission characteristics or human error.

Table 1. Signal identification categories

Category	Frequency	Bandwidth	Azimuth	Duration	Magnitude, dBm
Satellite	Discrete	Narrow	Moving	2-15 min	-110-123
	Discrete	Narrow	Fixed	1-50 min	-115-123
Airborne TM	Discrete	Narrow	Moving	5-50 min	-110-123
Airborne ECM	Various	Broad	Moving	1-10 min	-90-123
	Discrete	Narrow	Fixed	0-2 min	-90-123
Fixed azimuth	Discrete	Narrow	Fixed	Various	-115-123

Table 2. Radio spectrum surveillance station sightings April 1980 to December 31 1981

Sighting category	1980		1981	
	Total sightings	Number of incidents	Total sightings	Number of incidents
All sightings	8121	1198	30,583	3285
NASA satellites	6082	1010	22,906	2860
Identified	1821	19	7,318	291
DOD ECM	88	7	1,501	66
DOD TM	331	12	1,714	28
DOD SAT	7	4	37	5
Foreign SAT	439	73	463	123
Amateur radio	971	23	2,174	48
Shuttle	–	–	1,055	13
Lightning	–	–	364	8
Unidentified	218	69	369	134
ECM	84	13	147	26
TM	52	4	91	7
SAT	3	1	13	3
Unknown	76	51	118	98

Energy Consumption Analysis for the Mars Deep Space Station

N. V. Hayes
DSN Engineering Section

This article is the second in a series of energy consumption analysis and verification reports for the Goldstone Deep Space Communications Complex. Results for the energy consumption analysis at the Mars Deep Space Station are presented. It is shown that the major energy consumers are the 64-Meter Antenna Building and the Operations Support Building. Verification of the antenna's energy consumption is highly dependent on an accurate knowledge of the tracking operations. The report also indicates the importance of a regular maintenance schedule for the watt-hour meters installed at the station.

I. Introduction

In 1973, the Goldstone Deep Space Communications Complex (GDSCC) initiated an energy management program in order to reduce energy consumption. The specific objective was to reduce the consumption of purchased energy by 50% by the end of 1985, using the consumption level of 1973 as a base.

A computer model was developed which simulates energy loads in buildings. This model, the Energy Consumption Program (ECP, described in Refs. 1 and 2) uses parameters of building construction, weather conditions, and mechanical/electrical components usage to calculate the energy consumption requirements of the building and the cost of the energy and to suggest which configurations and procedures would expend less energy upon modification.

The ECP model simulates the energy requirements for a building in four major steps. First, the heat loss or heat gain to the space under observation is computed. Second, the heating or cooling loads imposed on the air handlers are determined. Third, the energy input to all of the primary equipment or components constituting the air-conditioning system, such as compressors, heat pumps, boilers, etc., is calculated. Fourth, the model provides an economic analysis of existing configurations through which the cost-effectiveness of a specific modification may be obtained. The simulation yields data that describes daily, monthly, and yearly total consumption for one or more buildings. The computer model is later verified by comparing its predictions to the actual watt-hour meter data.

The GDSCC is composed of a number of Deep Space Stations. A verification analysis of the Echo Station (DSS 12) was presented earlier (Ref. 3). This report analyzes the energy consumption at the Mars Station (DSS 14).

II. Energy Consumption Analysis

The Mars Station is composed of eleven support, storage and control buildings dominated by the 64-Meter Antenna Building. The buildings are listed by their function in Table 1. The ECP divides the energy consumers for each building interior into the five following groups: (1) electrical equip-

ment, (2) mechanical equipment, (3) heating, ventilation, air-conditioning (HVAC) equipment, (4) accessories, and (5) lights. The interior electrical equipment includes computers, electronic racks, and other electronics not related to HVAC. Mechanical equipment includes those heat generating machines inside the space such as machine shop equipment, air compressors and oil pumps. Other mechanical and electrical equipments associated with the HVAC operation are listed separately under accessories. The accessories include equipment necessary for building operation but which do not affect interior heating and cooling load calculations. The accessories are divided into thermal-powered and electrical-powered types. Air handler fans, condenser fans (for air-cooled type), boiler pumps, and building external lights are classified as electrical-powered accessories. Thermal-powered accessories include fuel-consuming equipment not located within air-conditioned zones, such as domestic hot water boilers. The lighting equipment is classified into incandescent and fluorescent types. A month-by-month listing of energy consumption by each of the above five groups is given in Table 2. Monthly heating and cooling loads for all the buildings are presented in Tables 3 and 4, and depicted in Figs. 1 and 2. A discussion of these simulated results is given below.

A. Electrical Consumption Analysis via the ECP Program

Figure 3 illustrates the distribution of electrical energy consumption between electrical equipment, mechanical equipment, HVAC, accessories, and lights. The largest consumers of energy are the accessories. Figure 4 gives a graphical representation of the electrical energy consumption for the entire Mars Site on a monthly basis.

Table 5 presents the yearly electrical consumption for site buildings as calculated by the ECP. Buildings G-85, G-87, G-89, and G-90 are not included in this listing because they are small consumers of energy. Figures 5, 6, 7, and 8 indicate which buildings are the major consumers in four categories. G-80, the 64-m antenna, is the largest consumer for HVAC and accessories. The Operations Support Building, G-86, houses data acquisition/processing equipment for the station (computers, printers, etc.) and is the primary consumer in the electrical equipment category (70.7%). It accounts for 23% of the HVAC consumption, because such equipment is necessary to the operation of computing devices. The relative distribution of building electrical consumption at Mars Station is shown in Fig. 9.

B. The ECP vs Watt-Hour Meter Verification

Electrical consumption at GDSCC is monitored by watt-hour meters. The stations are supplied with commercial (Southern California Edison) and site-generated power. All meters are read once a month. A review of the meter readings was done for the years 1978-1980. The ECP vs building meter readings have generally been accurate to within 10%.

The 64-meter antenna consumes power in accordance with the nature of the tracking assignment. For example, the antenna uses a 20-kW transmitter for routine signals to spacecraft, and a 100-kW transmitter when more power is necessary. For research and development assignments such as radar mapping and planetary radio astronomy, the antenna uses a 400-kW transmitter. The other buildings at the site perform support functions to the radio antenna; their energy consumption reflects the antenna's schedule. The tracking schedule at Mars Station has been erratic in recent years: the 64-meter antenna was down 15 March – 19 May 1980 for alignment. For this reason, the 1980 data had to be excluded from the ECP-hour meter verification.

During the course of this study, it was discovered that a number of the meters monitoring the two generators (500 and 750 kW) at the site were either out of service or malfunctioning. It can be deduced, however, from a comparison of the SCE meter (number 15) and the total reading of the meters monitoring individual buildings (see Fig. 10) that the generators provided DSS 14 with 31% of its energy in 1978, 41% of its energy in 1979, and 30% of its energy in 1980. The total ECP vs individual meter readings agree to 1.2% in 1978 and 3.5% in 1979. The only thermal consumer at DSS 14 is G-82/83, the Pump House/Cooling Tower System. There is no LPG consumption at the site.

At the beginning of 1980 it was discovered that the G-82/83 Pump House/Cooling Tower System would operate adequately with only one of its two cooled water loops in operation. This modification resulted in a 45% reduction in energy consumption by that building.

III. Summary

The ECP program allows a detailed analysis of energy consumption for a complex of buildings that includes a categorization of energy consumption. The present verification study disclosed that G-86, the Operations Support Building, is the largest consumer in the electrical equipment category. This study also showed the irregular tracking schedule during the years 1978-1980, a schedule that necessitated the taking of averages and the exclusion of some data. The irregular tracking schedule during 1978-1980 precludes a close year by year agreement between the ECP results and the meter data. For this reason the electrical consumption data was

averaged over this period and then compared with the meter readings.

It was found during the course of this study that several of the meters monitoring the two generators supplying the site were malfunctioning. It is suggested that the meters be inspected regularly and that meter systems be regularly checked for consistency (the sum of the readings of generator meters numbers 44 and 46, for example, should equal the reading of meter number 70).

Acknowledgment

The author wishes to acknowledge the assistance of P. Stelzmuller, F. Menninger, C. S. Yung and D. Schonfeld of DSN Engineering in providing data for this report. The author would also like to thank R. Kelley of GDSCC for information regarding Mars Station tracking operations.

References

1. Stoller, F. W., et al., "Energy Consumption Program – A Computer Model Simulating Energy Loads in Buildings," *DSN Progress Report, 42-45*, Jet Propulsion Laboratory, Pasadena, Calif., pp. 288-293, June 15, 1978.

2. Lansing, F. L., et al., "The Updated Algorithm of the Energy Consumption Program (ECP)," *DSN Progress Report 42-49*, Jet Propulsion Laboratory, Pasadena, Calif., pp. 107-115, Feb. 15, 1979.

3. Guiar, C. N., and Schonfeld, D., "Energy Consumption for the Echo Station (DSS 12)" *TDA Progress Report 42-66*, Jet Propulsion Laboratory, Pasadena, Calif., pp. 355-363, Dec. 15, 1981.

Table 1. Mars Station buildings

Building no.	Building description
G-80	Large radio antenna, 64-m (210-ft) diameter
G-81	Generator building
G-82	Pump house
G-83	Cooling tower
G-84	Training and office building
G-85	Flammable storage building
G-86	Operation support building
G-87	Security building
G-88	Transmitter-rectifier building
G-89	Reverse osmosis building
G-90	Storage building
G-91	Switch gear building

Table 2. Simulated energy consumption for Mars Station using ECP program

Month	Accessories, electric, MWhe	Lights		Electronic equipment, MWhe	Mechanical equipment MWhe	HVAC equipment		Monthly Total, MWhe
		Incandescent, MWhe	Flourescent, MWhe			Thermal, MWht	Electric, MWhe	
January	485	1.1	23.3	89.9	22.6	0.67	141	762.9
February	437	1.0	21.1	81.2	20.4	0	132	692.7
March	500	1.1	23.3	89.9	22.6	0	230	866.9
April	483	1.0	22.5	87.0	22.0	0	148	763.5
May	499	1.1	23.3	89.9	22.6	0	177	812.9
June	500	1.0	22.5	87.0	22.0	0	192	824.5
July	517	1.1	23.3	89.9	22.6	0	409	1062.9
August	517	1.1	23.3	89.9	22.6	0	214	867.9
September	484	1.0	22.5	87.0	22.0	0	185	801.5
October	500	1.1	23.3	89.9	22.6	0	161	797.9
November	484	1.0	22.5	87.0	22.0	0	142	758.5
December	485	1.1	23.1	89.9	22.6	0.97	139	760.0
Year total	5891	12.7	274.0	1058.5	266.6	1.64	2270	

Table 3. Heating load (in MWh) for the Mars Station

	January	February	March	April	May	June	July	August	September	October	November	December	Total (year)
G-80	63.6	54.6	60.0	57.0	49.4	45.9	46.7	46.8	46.3	57.4	58.8	64.3	650.8
G-81	8.6	6.2	6.3	4.7	11.5	9.1	9.6	10.0	9.8	4.9	6.6	8.4	95.7
G-82/83	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
G-84	3.1	1.3	1.2	0.7	0.6	0.1	0.0	0.0	0.14	0.14	1.4	3.6	12.3
G-86	1.0	0.4	0.3	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.4	1.2	3.5

1 Btu = 0.293×10^{-6} MWh

Table 4. Cooling load (in MHh) for the Mars Station

	January	February	March	April	May	June	July	August	September	October	November	December	Total (year)
G-80	83.4	78.7	88.1	87.1	125.7	127.9	136.1	135.4	126.0	93.0	84.1	82.5	124.8
G-81	5.8	5.7	6.3	5.6	11.9	12.3	14.4	14.6	12.3	7.4	6.0	5.2	107.5
G-82/83	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0
G-84	0.04	0.75	1.25	1.9	1.6	3.6	5.2	4.9	2.7	2.6	0.6	0.003	25
G-86	128.2	120.0	134.1	131.8	142.5	146.8	157.3	156.4	144.0	140.0	128.4	126.8	165.6

1 ton-hr = 3.516×10^{-3} MWh

Table 5. Simulated yearly electrical consumption for major Mars Station buildings

Building no.	Electrical consumption, MWhe
G-80	5913.621
G-81	777.619
G-82/83	885.210
G-84	64.928
G-86	1856.944

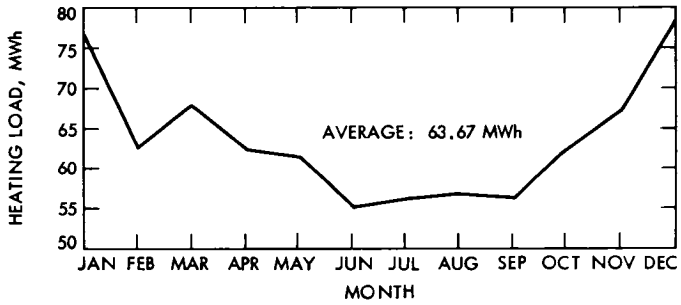


Fig. 1. Monthly variations in the heating load of the Mars Station

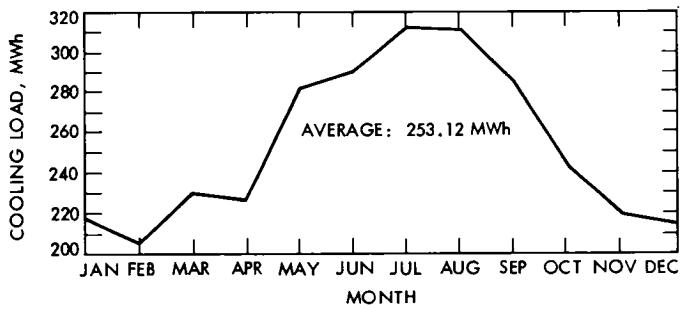


Fig. 2. Monthly variations in the cooling load of the Mars Station

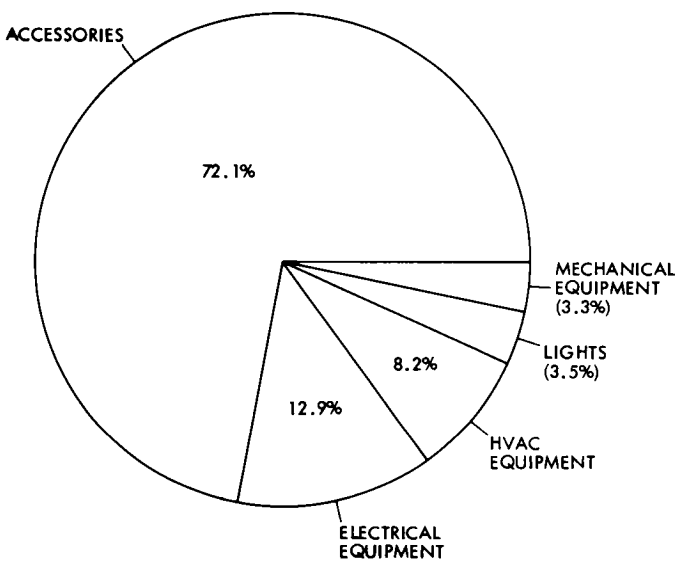


Fig. 3. Itemization of annual electrical consumption at the Mars Station

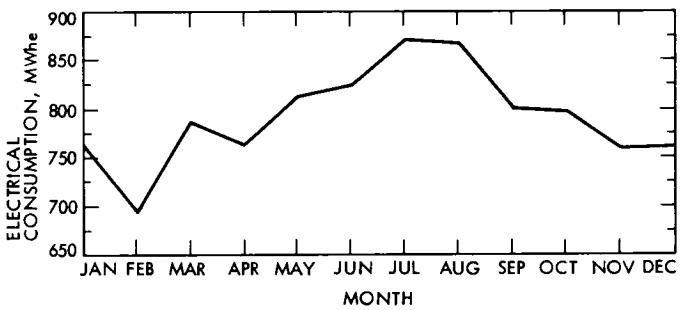


Fig. 4. Electrical consumption on a monthly basis, Mars Station

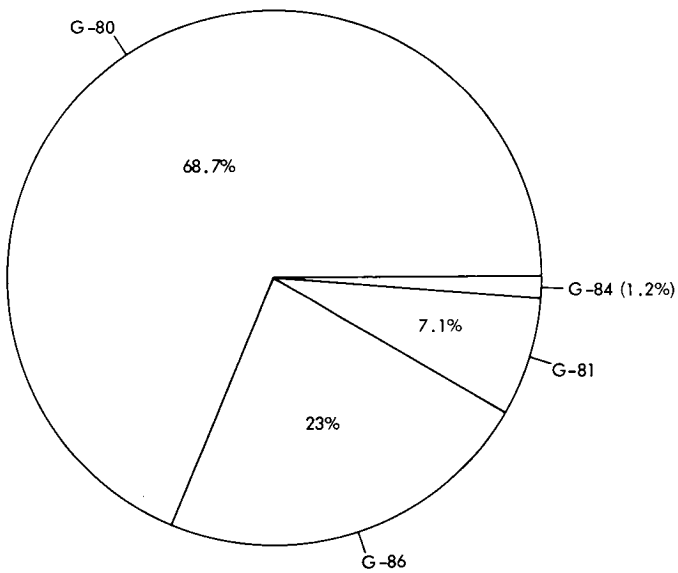


Fig. 5. Electrical consumption of HVAC equipment for buildings at the Mars Station

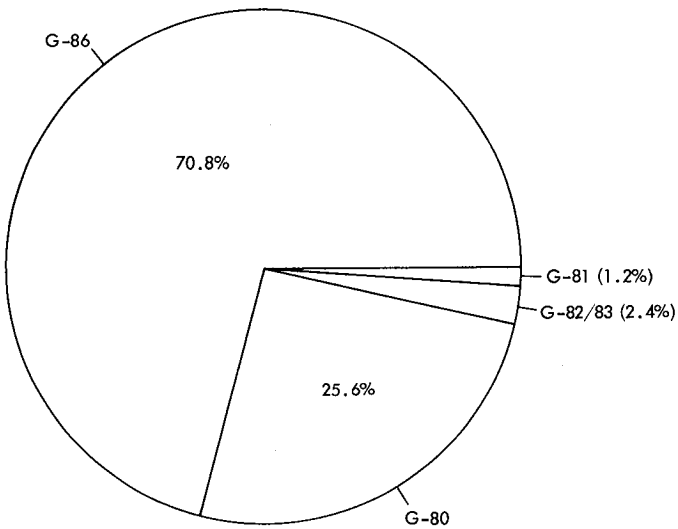


Fig. 6. Electrical equipment consumption for major buildings at the Mars Station

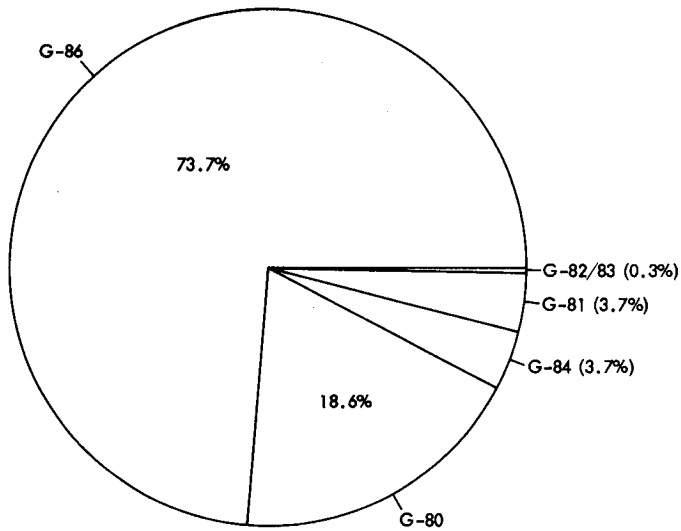


Fig. 8. Light consumption for major buildings at the Mars Station

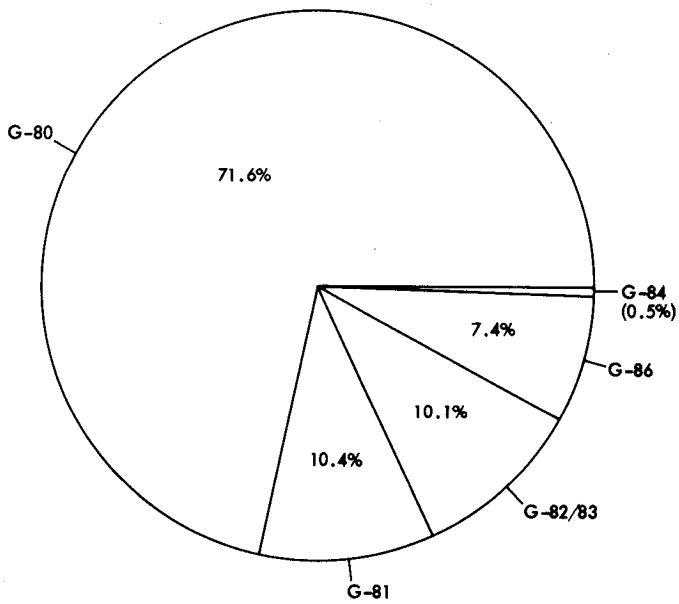


Fig. 7. Accessories electrical consumption of major buildings at the Mars Station

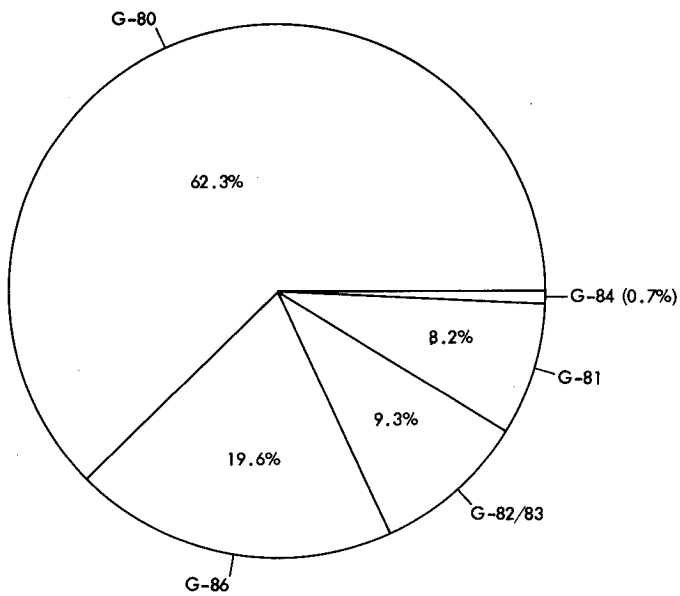


Fig. 9. Yearly electrical consumption for major buildings at the Mars Station

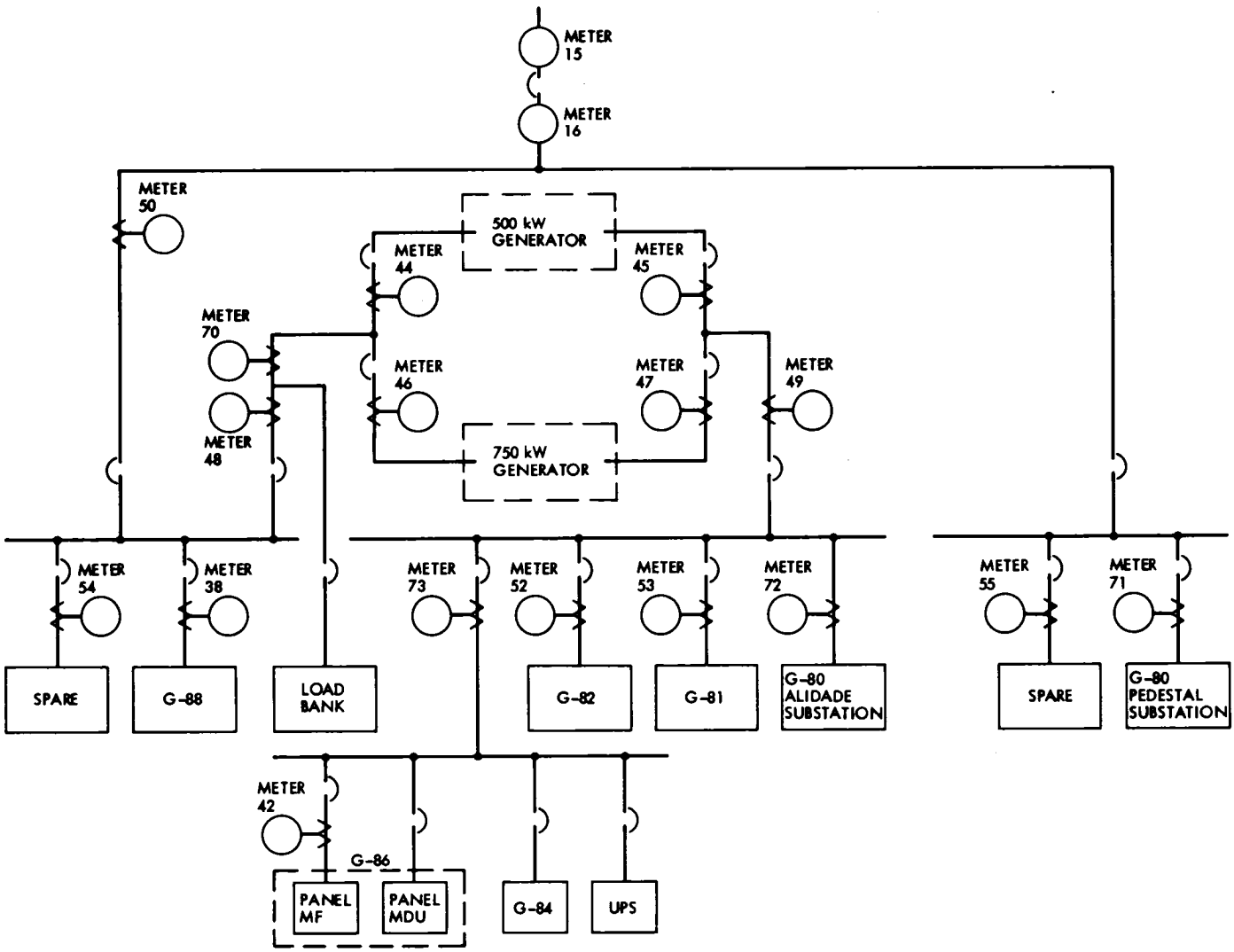


Fig. 10. Layout of watt-hour meters for the Mars Station