(1+21)

(NASA-CR-173885) THE TELECOMMUNICATIONS AND DATA ACQUISITION REPORT Progress Report, Apr. - Jun. 1984 (Jet Propulsion Lab.) 230 p HC A11/MF A01 CSCL 17B G3/32

N84-32622 THRU N84-32643 Unclas 20349

The Telecommunications and Data Acquisition Progress Report 42-78

April–June 1984

E.C. Posner
Editor

August 15, 1984

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

The Telecommunications and Data Acquisition Progress Report 42-78

April–June 1984

E.C. Posner
Editor

August 15, 1984

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

The research described in this publication was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

Preface

This quarterly publication provides archival reports on developments in programs managed by JPL's Office of Telecommunications and Data Acquisition (TDA). In space communications, radio navigation, radio science, and ground-based radio astronomy, it reports on activities of the Deep Space Network (DSN) and its associated Ground Communications Facility (GCF) in planning, in supporting research and technology, in implementation, and in operations. Also included is TDA-funded activity at JPL on data and information systems and reimbursable DSN work performed for other space agencies through NASA. The preceding work is all performed for NASA's Office of Space Tracking and Data Systems (OSTDS).

In geodynamics, the publication reports on the application of radio interferometry at microwave frequencies for geodynamic measurements. In the search for extraterrestrial intelligence (SETI), it reports on implementation and operations for searching the microwave spectrum. The latter two programs are performed for NASA's Office of Space Science and Applications (OSSA).

Finally, tasks funded under the JPL Director's Discretionary Fund and the Caltech President's Fund which involve the TDA Office are included.

This and each succeeding issue of the TDA Progress Report will present material in some, but not necessarily all, of the following categories:

OSTDS Tasks:

- DSN Advanced Systems
  - Tracking and Ground-Based Navigation
  - Communications, Spacecraft-Ground
  - Station Control and System Technology
  - Network Data Processing and Productivity
- DSN Systems Implementation
  - Capabilities for New Projects
  - Networks Consolidation Program
  - New Initiatives
  - Network Sustaining
- DSN Operations
  - Network Operations and Operations Support
  - Mission Interface and Support
  - TDA Program Management and Analysis
- GCF Implementation and Operations
- Data and Information Systems

OSSA Tasks:

- Search for Extraterrestrial Intelligence
- Geodynamics
  - Geodetic Instrument Development
  - Geodynamic Science

Discretionary Funded Tasks

Contents

OSTDS TASKS
DSN Advanced Systems
TRACKING AND GROUND-BASED NAVIGATION

COMMUNICATIONS, SPACECRAFT-GROUND

STATION CONTROL AND SYSTEM TECHNOLOGY

DSN Systems Implementation
CAPABILITIES FOR NEW PROJECTS

GCF Implementation and Operations
CAPABILITIES TO MEET PROJECT REQUIREMENTS

OSSA TASKS
Geodynamics
GEODETIC INSTRUMENT DEVELOPMENT

A High-Performance Hybrid RF Isolation Amplifier

G. L. Stevens
Communications Systems Research Section

A high-performance hybrid RF Isolation Amplifier (Iso-Amp) has been developed at JPL. The circuit exhibits a unique combination of RF characteristics at performance levels exceeding those of any commercially available device. Recent improvements in the design have resulted in significantly higher reverse isolation, a four-fold increase in bandwidth and improved reliability. These devices are very useful in RF and IF signal conditioners, instrumentation, and signal generation and distribution equipment. These Iso-Amps should find wide application in future DSN and R&D RF systems.

I. Introduction

In 1976 a two-stage RF amplifier exhibiting very high reverse isolation and well-controlled impedance levels was designed at JPL.[1] The circuit was hybridized and several hundred of these hybrid Iso-Amps were used in JPL R&D systems including delay-stabilized microwave links for reference frequency distribution and various multiple output distribution amplifiers.

Recently, an additional sixty Iso-Amps were used in an Allan variance frequency stability analyzer where the high reverse isolation of this circuit was needed. Unfortunately, this group of devices, built under contract to JPL, had an unacceptably high failure rate. Within the first few months of operation, four of the sixty hybrid Iso-Amps had catastrophically failed and others were suspected of intermittent bursts of noisy operation. Failure analysis of these four Iso-Amps indicated that the failures were due to poor construction techniques, sloppy component mounting and wire bonding, and little or no quality control. A decision was made to replace all sixty hybrids in the frequency stability analyzer. The replacement parts would have to be built by a firm whose workmanship and quality assurance standards were better than those of the previous supplier.

Prior to interviewing manufacturers of RF hybrid circuits it was decided to see if it might be possible to extend the RF performance of the circuit. The result of this effort was a modified circuit with 10 dB higher reverse isolation and with the upper frequency range extended from 100 MHz to 400 MHz.

After interviewing several leading manufacturers of thick- and thin-film RF amplifiers, a fixed-price contract was let to a corporation in Florida to hybridize the modified Iso-Amp design and produce 300 units. Monitoring of the contractor was provided during the engineering and manufacturing phases by JPL engineering and quality assurance personnel.

[1] By J. MacConnell in JPL's Communication Systems Research Section.

Approximately 100 of the new hybrid Iso-Amps have been installed in the frequency stability analyzer and other R&D systems. No failures have occurred with these units and the improved performance provided by the modified design will benefit future DSN and R&D systems.

II. Functional Description and Applications

There are dozens of companies within the U.S. and abroad which manufacture monolithic and hybrid RF amplifiers. Hundreds of off-the-shelf devices are available which provide the RF engineer with a wide selection of easy to use gain blocks sporting various combinations of gain, bandwidth, noise figure, input and output standing wave ratio (SWR), and output compression level. Packaging options include flat packs, dual in-line devices, transistor outline (TO) cans, and connectorized modules. Recently, a number of wide bandwidth, monolithic silicon RF amplifiers have been introduced which are housed in tiny, 70 mil microwave transistor packages.

With this vast selection of commercially available devices, why would we want to design another RF amplifier? Could our requirements be so different than those of other RF engineers to justify the expense of designing and hybridizing a custom part? In what ways does the performance of the Iso-Amp excel over the commercially available parts?

The hybrid Iso-Amp was designed to have the following characteristics:

(1) Very high reverse isolation (greater than 50 dB at 400 MHz)

(2) Very good input impedance match in 50 ohm systems (input SWR less than 1.10:1)

(3) Output power level capability of at least +16 dBm

(4) Gain of approximately 7 dB

(5) Wide band operation (0.4 to 400 MHz)

(6) Single supply voltage operation with low power consumption (+15V at 45 mA).

(7) Easily inserted package for 50 ohm microstrip systems

(8) Exceptional RF performance at a moderate cost

High reverse isolation and accurate input impedance control were the primary Iso-Amp design goals. Wide bandwidth, large signal handling capability, single supply operation with low power dissipation, and moderate cost extend the usefulness of the part to a broad spectrum of applications.

Reverse isolation is a term describing the reverse transmission loss through a device. It is a measure of the "reverse leakage" of a signal applied at the output of the device and measured at its input. Although the RF designer is normally concerned with signals flowing unidirectionally from inputs toward outpus, there are numerous RF design situations requiring amplifiers which can effectively block those "wrong way" signals.

Multiple-output distribution amplifiers (DAs) accept a single RF or IF input which is amplified and then split to provide several identical outputs. These devices are used in situations where there are multiple users of a single signal. DAs are used, for example, in DSN tracking stations to supply reference frequency signals (originating in a single atomic frequency standard) to many separate pieces of equipment and systems. It is very desirable for the DA to have high isolation between its output ports. This will insure clean, stable outputs on all ports even if one output is corrupted by an unstable terminating impedance, or by signals or noise emanating from poorly designed or faulty equipment. An Iso-Amp placed in series with each of the DA's outputs will provide very high isolation between these ports.

Another application which sometimes requires high reverse isolation is in receiver or instrumentation IF strips. Local oscillator leakage through a mixer may contaminate the preceding RF or IF stages, adversely affecting other users of that signal. An Iso-Amp placed ahead of the mixer will block the LO leakage and maintain the purity of the input signal.

The use of high-isolation buffers is often necessary where RF signals must enter an exceptionally noisy environment. An example of this occurs in frequency-multiplying phase locked loops, where a sample of the voltage controlled oscillator's (VCO's) output signal must be supplied to a digital frequency counter. Switching transients and noise from the digital logic will not only add to the output of the VCO but may modulate the signal as well. This spectral contamination can be avoided by isolating the output of the VCO from the noisy digital circuitry.

Accurately controlled input impedance was also an important design goal in the development of the hybrid Iso-Amp. The Iso-Amp was designed to operate in 50 ohm systems and its input provides an exceptionally good terminating impedance over a wide bandwidth. This characteristic is useful when working with devices whose RF performance is sensitive to impedance levels, such as mixers, reactive filters, equalizers, directional couplers, power splitters, bridges, and transmission lines.

The versatility of the hybrid Iso-Amp stems from its unique combination of RF characteristics. There are, of course, alternate solutions to design problems requiring high isolation, excellent impedance matching, and high output level capability.

High reverse isolation can be achieved by using a cascade of several general purpose RF amplifiers with suitable interstage pads to control the overall gain. Care must be exercised during the design to insure an acceptable noise figure and dynamic range. With proper design, packaging, and construction, high isolation will be realized from the chain of devices. Compared to using a single device, the concatenation will occupy more board space and require more construction time. It will likely consume more power from DC power supplies, dissipate more heat, and have a higher overall parts cost.

Applications requiring a broadband RF amplifier with low input SWR can be satisfied by placing a fixed pad between the signal source and a high-SWR amplifier. If the required input SWR is very low, the pad's value may be large. The overall noise figure of the pad/amplifier combination is increased by the pad's value, and similarly, the gain is reduced by that amount. An additional stage of gain may be necessary to compensate for the lossy, impedance-matching pad. Again, the costs of using this approach include higher circuit complexity and increased design effort.

III. Circuit Description

A schematic diagram of the modified Iso-Amp circuit is shown in Figure 1. The circuit employs two complementary stages to achieve its high reverse isolation, good impedance matching, and large signal handling capability. Recent modifications to the circuit include:

(1) Frequency compensation (resulting in quadrupled bandwidth)

(2) Better transistors (resulting in higher reverse isolation and improved reliability)

(3) Changes in the output biasing (to eliminate a possible source of noise)

(4) Additional power supply decoupling (for improved isolation)

The input stage was designed to provide a good impedance match in 50 ohm systems and to exhibit very high reverse isolation. A pair of bipolar microwave transistors are connected in a common-base, complementary circuit. The input signal is fed through a 44 ohm resistor (R1) to both emitters. With a collector current of 10 mA, the AC impedance looking into the parallel emitters is 6 ohms. The total input impedance is

$$Z_{in} = 44 + 6 = 50 \text{ ohms}$$

DC bias voltages for the input transistor bases are derived from the voltage divider consisting of R3 through R6. The collector current of 10 mA is then established by the emitter resistors (R7 and R8) of 150 ohms each. R9 is the collector load resistor for the first stage. The total AC load presented to the first stage is the parallel combination of this resistor and the input impedance of the output amplifier. The net collector load impedance is 40 ohms and the common-base configuration of the circuit establishes a current gain of unity. This results in a voltage gain of approximately 0.8 (or about -2 dB) for the input stage.

PNP and NPN transistors are used in complementary configurations in both stages of the Iso-Amp. This allows the Iso-Amp to handle large RF signals with good linearity while consuming a modest current from its power supply.

The output structure developed for the Iso-Amp is best described by using separate DC and AC models.

In terms of DC biasing, transistor Q3 and its associated circuitry operate as a current source. Resistors R11 and R12 form a voltage divider across the supply voltage, holding Q3's base at about +11.5 volts. Approximately 2.75 volts appear across the series combination of R16 (68 ohms) and R17 (20 ohms) resulting in an emitter (and collector) current of 31 mA. This DC bias current is fed to the collector circuit of transistor Q4.

Approximately 5 mA of this current are diverted through the resistive divider consisting of R13 (390 ohms), R14 (820 ohms), and R15 (300 ohms). The DC voltage seen by the base of Q4 is proportional to the voltage appearing on its collector. In operation, the voltage at Q4's collector will rise to a level sufficiently high that its base voltage overcomes the base-emitter drop of 0.8 volts. At this point, Q4's increasing conduction slows the rise of collector voltage. Bias equilibrium is attained when the collector voltage reaches +7.5 volts. This circuit creates a "stiff" voltage source and the final collector voltage is quite insensitive to bias current changes.

The biasing arrangement of the output stage is, then, a current source driving a voltage source. This leads to a very stable DC operating point for the transistors. Collector current and collector voltage are independently controlled with this arrangement and each is accurately maintained over normal variations of power supply voltage and temperature.

Consider the AC model of the output stage. The two transistors operate in parallel. Examination of the schematic reveals that, for AC signals, both transistor bases are tied together, as are the emitters and collectors. The output stage can therefore be analyzed as if it were a single transistor amplifier employing emitter degeneration and collector voltage feedback. The voltage gain of this stage is given by:

$$A_v = \frac{-R_c}{R_e}$$

where

R_c = the collector load impedance

and

R_e = the total emitter impedance

The input impedance of this stage is

$$Z_{in} = \frac{R_f}{1 + A_v}$$

where

R_f = the unbypassed feedback resistance

and

A_v = the voltage gain of the stage

With the component values shown in Fig. 1, the second stage input impedance is 68 ohms (including the loading caused by bias resistors) and its gain is approximately −2.8 V/V, or 9 dB.

The net gain of the complete, two stage Iso-Amp is

$$G_{\text{Iso-Amp}} = -2\text{ dB} + 9\text{ dB} = +7\text{ dB}$$

IV. Breadboard Activities

Several changes in the Iso-Amp design were introduced during the recent modifications which resulted in improved RF performance. A number of breadboard circuits were built and tested during this activity. The first breadboards were built on copper-clad substrates using leaded components. Component leads and interconnecting wires were kept as short as possible to minimize signal radiation and parasitic inductance.

Once a candidate design was obtained, it was clear that a better breadboarding technique would have to be employed during the final circuit refinements and RF performance verification. The deleterious effects of component leads and the layout restrictions imposed by using packaged devices were clearly affecting the circuit's RF performance at frequencies beyond about 250 MHz. A better breadboarding technique was also desired to simplify replication of the circuit. Several identical breadboards had to be built and tested to determine the repeatability of the design's RF performance.

A two-sided printed circuit board was designed for the final breadboarding activities. Chip capacitors, 1/8 watt carbon composition resistors, and microwave transistors were surface mounted on the 4.0 X 2.8 cm (1.6 X 1.1 inch) boards. Plated through holes provide low inductance connections to the solid ground plane on the back of the boards. Several of these final breadboards were built and tested, confirming the repeatability of the modified design's RF performance. One of these breadboards is shown in Fig. 2.

V. Hybrid Circuit Implementation

The internal circuitry of one of the new hybrid Iso-Amps is shown in Fig. 3. The metal cap which would normally be welded in place over the circuitry was left off of this sample part. A 12-lead, circular TO-8 hybrid package measuring 1.52 cm (0.600 inch) in diameter provides the hermetically sealed environment for the thick film circuit.

All circuit elements within the hybrid circuit are either printed on or mounted on an alumina substrate measuring approximately 0.91 X 0.91 X 0.038 cm (0.36 X 0.36 X 0.015 inch). Initial processing of the substrates is performed on arrays of 9 or more substrates. Conductive circuit traces are added by the thick film process wherein gold-bearing ink is screened onto the substrate arrays in a pattern matching the desired metalization. Volatile solvents are subsequently removed in a drying process. The arrays are then fired at high temperature in an inert atmosphere, fusing the gold alloy conductors to the alumina substrates.

Thirteen of the nineteen resistors in the circuit are thick film and the balance are thin film devices. The thick film resistors are printed on the arrays with resistive inks. The ends of each resistor overlay the edges of conductors establishing the circuit connections. The substrate arrays are again dried and fired in a process similar to the one used for the circuit conductors. Initial resistance values are intentionally made slightly lower than desired since the thick film resistors are subsequently laser trimmed to their final values.

Chip capacitors, thin film resistor networks, and transistor dice are bonded to the substrate arrays with electrically conductive epoxy. The arrays are separated into individual substrates which are then bonded to the metal headers of the hybrid packages. Additional circuit connections are added with gold wirebonding. Due to the very small wirebond pads on the microwave transistors, 0.0018 cm (0.0007 inch) diameter gold bond wire is used to make base and emitter connections on all transistors. Other wirebonding, including connections to the header's base and pins, is performed with 0.0025 cm (0.001 inch) diameter gold wire.

Following inspection and initial electrical testing, the hybrid circuits enter an oxygen free environment where the metal covers are electrically welded to the headers, hermetically sealing the integrated circuit. Final electrical testing and case labeling are the last steps in the manufacturing process.

VI. Electrical Performance

Table 1 shows the typical RF performance of the hybrid Iso-Amps. Differences between the breadboard and hybrid Iso-Amps are insignificant with the notable exception of the reverse isolation. The reverse isolation of the hybrids is typically three or four dB lower than the performance obtained with the breadboard circuits. This may be due to slightly better grounding in the breadboards or it may be a consequence of shrinking the size of the circuit to a tenth of a square inch and mounting it in a package whose input and output pins are separated by only 1.0 cm (0.4 inch). A stray coupling capacity of only 0.05 picofarads between the RF input and output pins would, in fact, limit the device's isolation to about 62 dB at 100 MHz (a typical value for these hybrid Iso-Amps).

The two port scattering parameters of a typical hybrid Iso-Amp are given in Table 2.

VII. Applications Considerations

Since the hybrid Iso-Amp was designed to operate with relatively low power consumption, heat sinking is generally not a problem. In most applications the hybrid is soldered into a microstrip circuit board, with the base (header) of the circuit in intimate contact with a copper clad ground plane. This mounting configuration provides adequate heat sinking for the circuit without resorting to metal clamps or special heat sinks sometimes required with higher dissipation devices. Thermal resistance at the case/ground plane junction can be further reduced by adding a small amount of heat sink grease at this point when mounting the circuit. This precaution protects the Iso-Amps when they are operated in high temperature environments.

Some applications require higher reverse isolation than a single Iso-Amp can supply. Simply cascading a pair of these devices on a microstrip board does not double the reverse isolation. Signal leakage between output and input microstrip traces limits the isolation. Imperfect grounding and insufficient power supply decoupling can also adversely affect the cascade's ultimate isolation. Careful design, layout, and shielding can, however, yield Iso-Amp cascades with very high isolation.

Shown in Fig. 4 is an isolation amplifier module developed for the Allan Variance Frequency Stability Analyzer. Contained within this module are two independent cascades consisting of a concatenated pair of the hybrids with an interstage pad to control the overall gain. Careful attention to grounding, power supply decoupling and shielding (including beryllium-copper finger stock shields) was necessary to achieve reverse isolations in excess of 110 dB at 100 MHz.

VIII. Conclusion

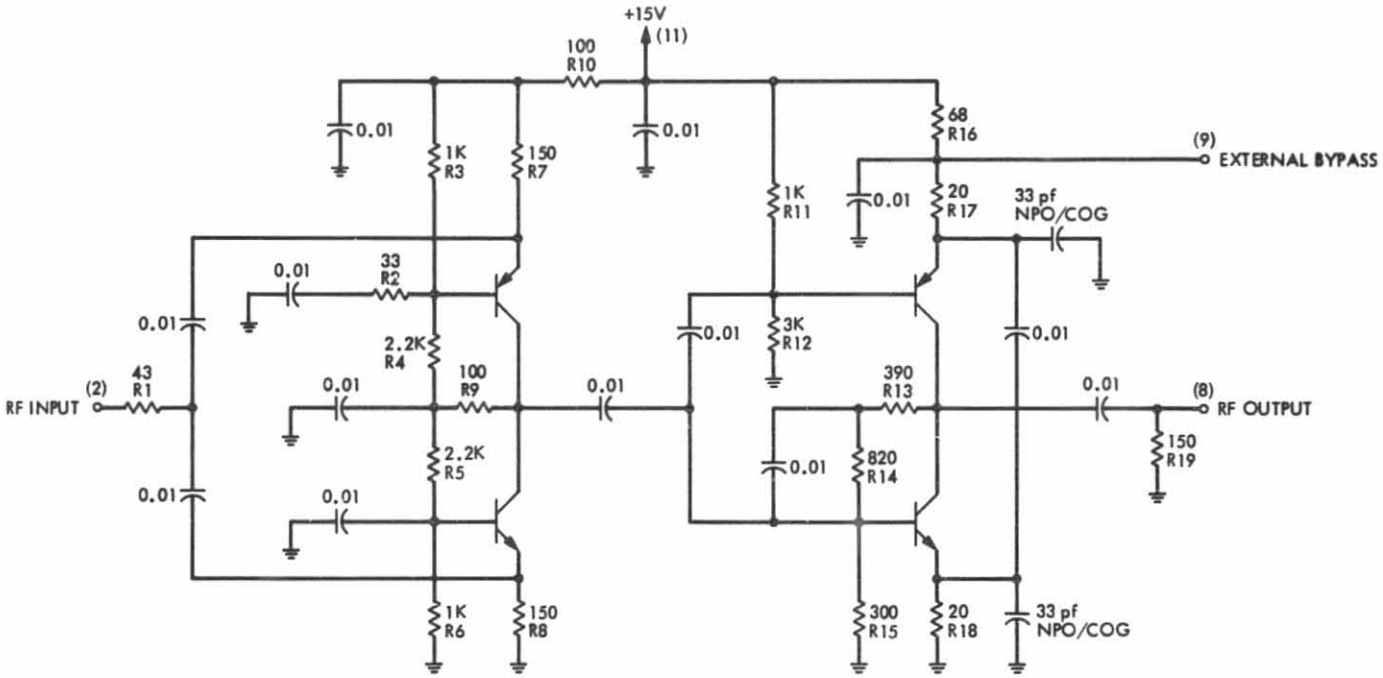
The hybrid RF isolation amplifier described in this article exhibits a unique combination of RF characteristics in a single easy to use package. This circuit provides an economical solution to many RF design problems requiring high reverse isolation, good impedance matching and high output level capability. Useful over a frequency range spanning ten octaves, these hybrid amplifiers should find wide application in future DSN and R&D RF systems.

Table 1. Hybrid Iso-Amp typical RF performance

Gain	7 dB
Frequency Response	Flat to within 0.5 dB from 2.0 to to 350 MHz -1.0 dB at 1.3 and 400 MHz -1.0 dB at 0.4 MHz with external 0.1 microfarad bypass capacitor from pin 9 to ground
Reverse Isolation	85 dB at 3 MHz 62 dB at 100 MHz 51 dB at 400 MHz
Input SWR	Less than 1.1 from 3 to 400 MHz
Output SWR	Less than 1.75 from 3 to 100 MHz
1 dB Output Compression Level	Greater than +16 dBm
Third Order Intercept Point	+27 dBm
Noise Figure	12.5 dB
DC Power	+15 V at 45 mA (Dissipation = 675 mW)
Package	1.52 cm (0.6 in.) diameter, 12 lead, hermetic TO-8

Table 2. Typical hybrid Iso-Amp scattering parameters

Frequency,	Input Return (S11)		Forward Gain (S21)		Reverse Isolation (S12)		Output Return (S22)	
MHz	dB	Angle	dB	Angle	dB	Angle	dB	Angle
3	-28.0	-138	7.06	-176	-87.3	16	-27.6	-145
20	-30.6	-178	7.00	175	-76.0	58	-27.5	-142
100	-31.1	146	7.29	155	-62.4	69	-17.5	-133
200	-31.0	117	7.56	129	-56.7	62	-12.5	-155
300	-30.2	96	7.46	100	-54.4	53	-10.3	178



NOTES: NUMBERS IN PARENTHESES ARE PACKAGE PIN NUMBERS
CAPACITOR VALUES ARE IN MICROFARADS UNLESS OTHERWISE NOTED
PNP TRANSISTORS ARE TYPE NEC 88900 (CHIP)
NPN TRANSISTORS ARE TYPE NEC 02100 (CHIP)
HYBRID PACKAGE IS 1.52 cm (0.600 inch) DIAMETER, 12 LEAD, HERMETIC TO-8
GROUND = PINS 1,3,4,5,6,7,10,12, AND CASE
ALL HYBRID GROUND PINS SHOULD BE CONNECTED TO GROUND PLANE

Fig. 1. RF Iso-Amp schematic diagram

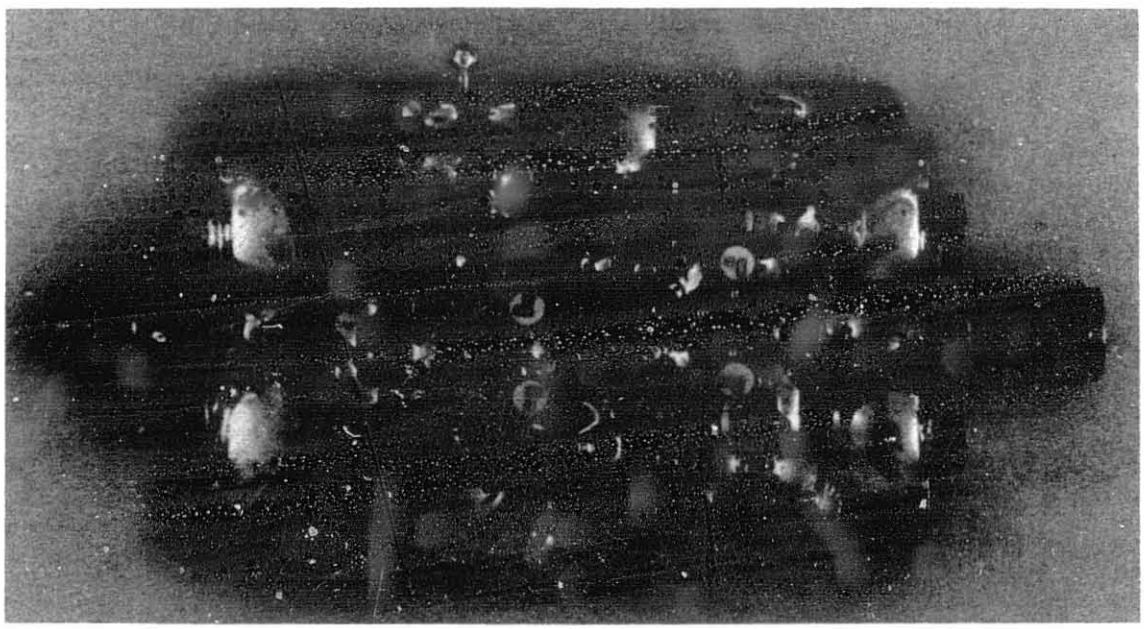


Fig. 2. RF Iso-Amp breadboard

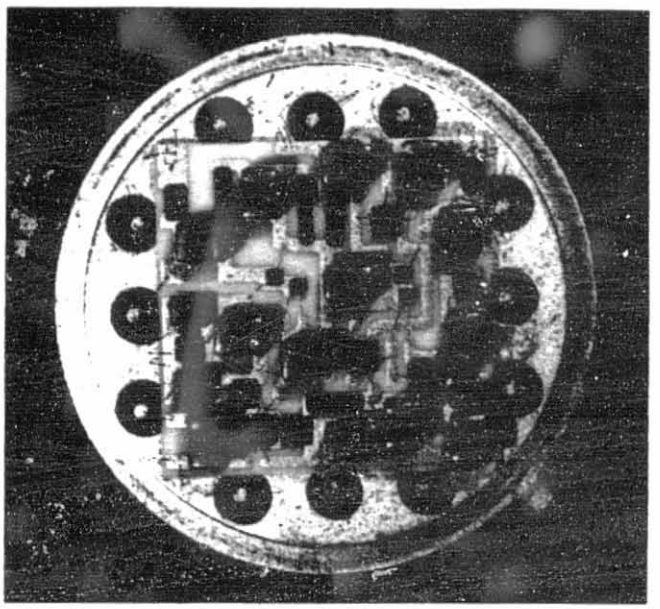


Fig. 3. Photograph of the hybrid Iso-Amp circuitry

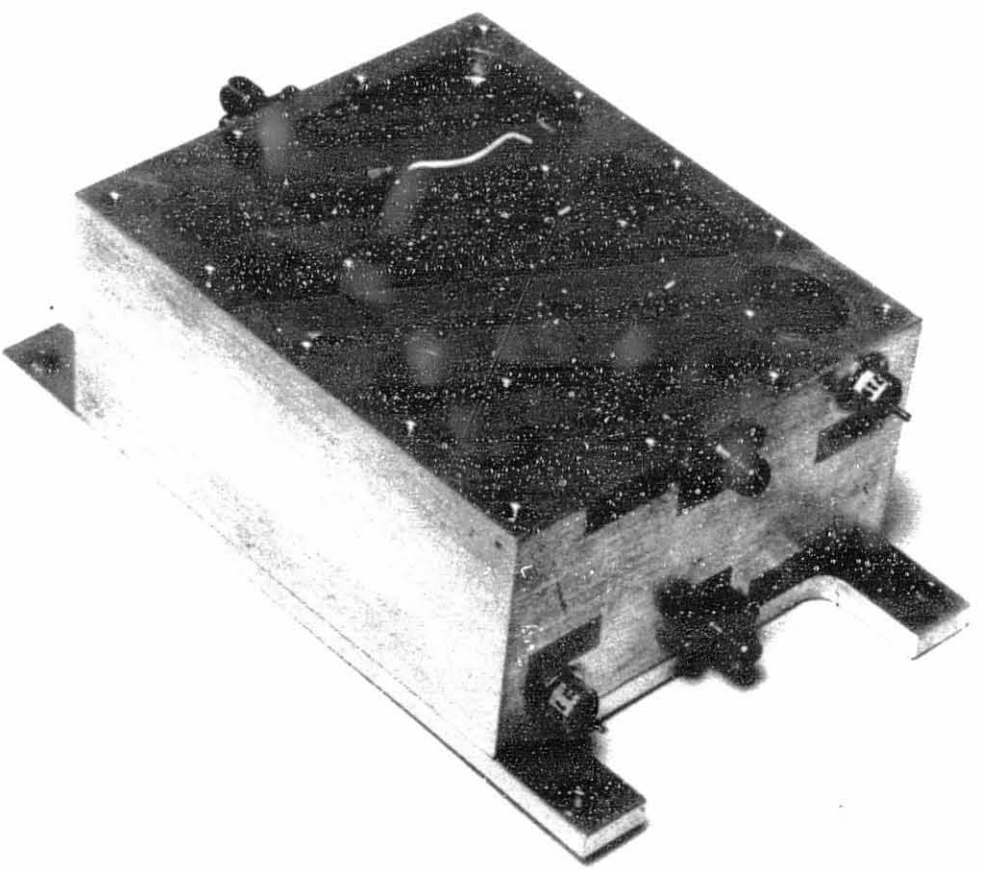


Fig. 4. Isolation Amplifier module for the Allan Variance Frequency Stability Analyzer

Optimization of Antenna Structure Design

R. Levy
Ground Antenna and Facilities Engineering Section

Optimality criteria design is applied for large antenna structures with multiple constraints on microwave performance. The constraints are on accuracy of the structure: restrictions on the microwave pathlength error, and the antenna pointing error. The examples given show convergence to low-weight feasible designs that satisfy the constraints. Truss-member sizes are automatically selected from tables of commercially available structural shapes, and approximations resulting from the method of selection are found moderate. The multiple constraint design is shown to be more effective in meeting constraints than the old envelope method. For practical structures, the new design method can be performed within reasonable core size and computation time.

I. Introduction

Structure design optimization theory and its application to a number of classical or academic problems have been described in the past two decades (14). Consequently, there are now theoretical design approaches to obtain low-weight structure designs that satisfy a variety of constraints on structure performance and safety. In particular, design approaches that use optimality criteria (Refs. 1, 6, 7, 10, and 15) are feasible for the design of structures with large numbers of degrees of freedom and/or design variables.

This article describes a specific application of the optimality criteria method to design structures that satisfies constraints on compliance, member tension, and buckling stresses. The model contains several thousand members and degrees of freedom and tens to hundreds of design variables, substantially more than encountered in design optimization models described in research literature. The emphasis here is on the design of a microwave antenna backup structure for which the design variables are the areas of the truss rod members or the thicknesses of membrane plates. Microwave performance constraints limit: (a) the root-mean-square of microwave pathlength deviations from a surface that is a best fitting alternative to the originally specified surface and (b) the antenna boresight pointing error. These constraints are functions of the displacements of the nodes of the structure that support reflective surface panels and must be satisfied for gravity and wind-pressure loadings at specified operational wind speeds. Stress and buckling constraints for gravity and operational wind-speed loadings tend to be benign, but can become significant at higher wind-speed survival conditions, or at other nonoperational environmental conditions.

The following discussions provide a brief review of the optimality criteria method and the pathlength error constraint. A new method that formulates pointing error constraints is introduced. A practical selection of structural members from tables of commercially available shapes is also described. Results from trial designs are given for the new 34-m (112-ft) diameter ground DSN antenna shown in Fig. 1. Histories of the design progress and comparisons of the efficiencies between the simple sequential envelope method designs and new designs that treat multiple constraints simultaneously are given. Another example compares the selection of design variables from commercial shape tables with the traditional assumption of a continuous spectrum of shapes.

II. Problem Formulation

Large ground antennas are essentially membrane-type structures in which the finite elements are rods (major) and plates (minor). Consequently, only the three translational degrees of freedom for displacements and forces at the structure nodes are needed. For brevity, only rod element areas (design variables) with identical material properties are discussed. The extensions to include the thicknesses of membrane plates as additional types of design variables and to treat members of different materials are straightforward (Refs. 7, 11, and 13).

A. Optimality Criteria Method

A brief formulation is provided for background. Further details, variations, and implementation strategies are available in the literature (Ref. 1, 7, 10, 13, and 15).

The key components of an optimization problem formulation are the objective function and the constraints. Here, the design objective is to minimize structure weight, or equivalently, volume, given as

$$V = \sum_{i=1}^{N} L_i a_i \tag{1}$$

in which a_i is the cross-sectional area, L_i is the length, and i is the index within a set of N design variables. Primary constraint equations are expressed as

$$G_j = \sum_{i=1}^{N} \frac{F_{ij} L_i}{a_i} - C_j^* \leqslant 0 \quad j = 1, \cdots, K \tag{2}$$

In Eq. (2), F_{ij} is a sensitivity coefficient such that the virtual work of the ith design variable for the jth constraint is $F_{ij}L_i/a_i$. When the design variable group consists of only a single member, the sensitivity coefficient is the product of the stress resultant (bar force) for a specified external loading and the stress resultant for an associated virtual (dummy) loading divided by Young's modulus. When the group contains more than one member, the length-weighted average product is used to compute the coefficient, and the length assigned to the group is the total length of members. C_j^* is a prespecified bound on the virtual work. In particular, when the displacement in a particular direction at a given node is the constraint, then C_j^* is that displacement bound and the virtual load is a unit load applied at that node and in that direction. Stress constraints for rod members can be converted to extension constraints, and the virtual loading vector consists of a pair of collinear self-equilibrating loads applied at the terminal nodes of the member. Antenna pathlength or pointing error constraint virtual loadings consist of particular vectors with components at all nodes that support surface panels.

Secondary constraints on the design variables can be imposed as

$$\underline{a}_i \leqslant a_i \leqslant \bar{a}_i \quad i = 1, \cdots, N \tag{3}$$

in which $\underline{a}_i$ and $\bar{a}_i$ are either user-established lower and upper bounds on the design variable a_i or dynamic adjustments constructed by algorithm. A possible replacement for a primary stress constraint is a lower bound side constraint developed by the stress ratio method. The replacement could produce significant computational savings and is valid for members that are not strongly affected by structural redundancy.

Equation (2) can be written as

$$G_j = C_j - C_j^* \tag{4}$$

where C_j is the realized value of the virtual work for the jth constraint and is given by the summation term of Eq. (2).

It is useful to define a constraint ratio, D_j, for which values greater than unity indicate an unsatisfied constraint. This can be computed as

$$D_j = C_j/C_j^* \tag{5}$$

The optimality criteria for the design variables for this problem are

$$a_i^2 = \sum_{j=1}^{K} F_{ij}\lambda_j \quad j = 1, \cdots, K \tag{6}$$

in which the λ_j are nonnegative Lagrangian multipliers to be determined from solution of the following auxiliary problem:

$$\lambda_j G_j = 0 \quad j = 1, \cdots, K \tag{7}$$

Strategies for solution of Eq. (7), typically either by Newton's method or a recursive approximation, have been discussed elsewhere (Refs. 5, 6, 10, and 15).

The sensitivity coefficients, F_{ij}, are assumed to be independent of the design variables, but structural redundancy weakens this assumption, so that the design must be approached iteratively through a number of cycles in which the sensitivity coefficients are recomputed. In the case of antenna gravity loading, the loads are a function of the design variables, which reinforces the need for iterative design. "Move" limits are often imposed upon the relative changes in design variables between adjacent cycles in an attempt to control the effects of redundancy or loading changes. These are invoked by adjustments to the upper and lower bounds on design variables as expressed in Eq. (3).

B. Antenna Pathlength Constraints

Microwave pathlength errors are determined from the original geometry of the surface and the deflection vectors of the surface nodes. Figure 2 shows the surface geometry relationships. The solid line through the antenna vertex represents the original ideal paraboloidal surface, and the broken line represents the deflected surface. The microwave pathlength is defined as the distance from the aperture plane to the surface plus the distance from the surface (after reflection) to the focal point and is shown by line *BCP*. In the figure, *GD* is the deflection vector and *dn* is the component of the deflection vector normal to the surface. It can be shown (Ref. 17) that the half-pathlength error R is given by

$$R = (BCP - ADP)/2 = \gamma_z \, dn \tag{8}$$

where γ_z is the direction cosine of the surface normal with respect to the focal (Z) axis.

In practice, the pathlength errors are computed from an alternative paraboloid that best fits the deflected surface. The best-fitting paraboloid is defined by at most six parameters, i.e., there are three independent shifts of the vertex parallel to the Cartesian X, Y, or Z coordinate axes, one relative change in focal length, and two independent rotations, one about the X (parallel to elevation) axis and the other about the Y (yaw) axis. Then the vector of pathlength errors **R** with respect to the best-fitting surface can be computed as follows:

$$\mathbf{R} = \mathbf{A}\,\mathbf{u} + \mathbf{B}\,\mathbf{h} \tag{9}$$

in which **u** is an external loading displacement vector for the nodes that support surface panels and the components of **u** are aligned parallel with the Cartesian axes, **h** is the vector of fitting parameters, and **A** and **B** are invariant matrices that contain the geometric relations to transform **u** and **h** as required by Eq. (8). Equation (9) is used as the basis of a least squares analysis (Ref. 17) to derive **h**. In particular, we find

$$\mathbf{h} = \mathbf{H}\,\mathbf{u} \tag{10}$$

in which **H** is again an invariant matrix derived from **A** and **B** and is equal to

$$\mathbf{H} = -(\mathbf{B}^t \mathbf{W} \mathbf{B})^{-1} \mathbf{B}^t \mathbf{W} \mathbf{A} \tag{11}$$

where **W** is a diagonal matrix of weighting factors usually taken to be proportional to the aperture area tributary to each surface node.

It has been shown (Ref. 8) for a pathlength error constraint that the virtual loading vector consists of loads directed normal to the antenna surface at each node that supports surface panels. The magnitudes of the loads are $\gamma_z^2 \, dn$, where *dn* is with respect to the best-fit paraboloid. When this loading is employed, the realized virtual work will become the sum of squares of half-pathlength errors and will be identical to the error computed by the usual geometric analysis of the deflections.

It is common to refer to the "rms error" of a paraboloidal antenna by the root-mean-square half-pathlength error from the best-fitting paraboloid. This is derived from the weighted sums of squares and the sums of the weighting factors.

In the cases of gravity loading, the pathlength errors at particular antenna elevation angles are computed for the difference in loading between that elevation and the "rigging" angle elevation. The "rigging" angle is the angle at which the panels are aligned in the field as accurately as possible to the ideal surface. Also, a single gravity loading can be constructed to define an appropriate constraint to represent either the maximum or a weighted average rms antenna pathlength error over the range of elevations from horizon to zenith. For either of these choices, there is an implicit computer algorithm (Ref. 9) to determine the rigging angle. Figure 3 illustrates different pathlength error distribution curves from horizon to zenith for different choices of rigging angle.

C. Antenna Pointing Constraints

Five independent additive components that contribute to the antenna boresight pointing error can be derived from geo-

metric optics theory (Ref. 12) for Cassegrain antennas. Definitions of the associated terms and the computational relationships are contained within Table 1. The first two rows of multiplying factors in the table apply to parameters of the best-fit paraboloid that can be extracted from the vector h that was defined in conjunction with Eqs. (9) and (10). However, as seen from Eq. (10), the inner product of each row of H with the displacement vector for external loading provides the corresponding term of h. Therefore, the virtual loading is derived from the appropriate rows of H multiplied by the factors of the table. As H is invariant, it needs to be computed only once for each structure, and it applies to all pointing constraints.

The virtual loads for the remaining pointing error contributions, which are in the last three rows of the table, are constructed from loads placed at the nodes that define the rotation and translations of subreflector and feed. The magnitudes of these loads are the tabulated factors times the coefficients of the equation used to define the rotation or translation. For example, if the feed translation is defined as the average displacement of two particular nodes in, say, the Y coordinate direction, the virtual loading consists of loads in the Y direction at these two nodes with magnitudes of 0.5 times the factor $k/(mf)$. Furthermore, if part of the pointing error is correctable by feedback from a sensor or encoder, subtractive virtual loading components can be included so that virtual work derived from this vector will provide the net pointing error. The virtual loading vector for each constraint is the union of the components constructed for each contribution to the pointing error.

III. Candidate Design Variables

The selection of structural members is usually restricted to a discrete set of commercially available cross-sectional shapes. Nevertheless, the optimality criteria (Eq. 6) assume a continuous spectrum for selection of design variables; consequently, a rigorous approach would employ a different design method (Refs. 2 and 16) to deal with the discrete spectrum. The approach used here, however, is to perform the design first for a continuous spectrum and then choose the nearest available size that also meets lower bound stress side constraints. This approach could theoretically lead to a nonoptimal selection, but if the available shapes provide a well-graduated set of properties without significant gaps, the departures from the optimal selection should become insignificant.

Figure 4 is a sample from one of several tables constructed for design within the JPL-IDEAS computer program. This particular table contains pipe and round-tube customary shapes. Since these tables are assembled from local warehouse stock lists, the data is in customary U.S. units. Under the heading HANDBOOK SHAPE, there is an abbreviated description of the section. For example, entry number 2 is a 1-in. (25.4-mm) schedule 10 pipe with 1.315-in (43.9-mm) outside diameter and 0.109-in. (2.7-mm) wall thickness. The cross-sectional area and radius of gyration are tabulated under the headings AREA and RAD. The load table contains allowable compression loads in kips (4,448 N) for the span lengths listed at the top. These loads are based upon the buckling stress formulas for ASTM A-36 steel in the ASCE design guide (Ref. 4). The design tension stress is also based upon the design guide, but the allowable loading can be reduced by a percentage to correct for reduced cross sections at connected ends.

The table is used to set the lower bound side constraint for compressive loads by using member length, maximum load for all loading cases, and the buckling stress formula. The nearest shorter-span-length column is searched for the first section with a larger allowable load, and the buckling formula is applied to determine the allowable load using the member length with the area and radius of gyration of the identified section. If the section is found to be inadequate, the section in the next lower row of the table is tested until an adequate shape is found. The shapes are listed in the table in the order of increasing area so that the search finds the lightest member to carry the load. Allowable loads tabulated as zero indicate that the slenderness ratio for that entry would be greater than an arbitrary limit of 200. The table is also used in each cycle after execution of the design algorithm to find the shape consistent with the lower bound side constraint that is closest in area to the one determined by optimality criteria.

IV. Example Applications

Structure models for the 34-m antenna examples that follow are extracted from design studies of NASA DSN antennae (Fig. 1) scheduled to be installed in California and Australia in 1985. Design optimization will be described only for the tipping structure, which consists of quadripod, reflector backup structure, and elevation wheel. Design of the alidade, which supports the tipping structure at the elevation bearings and at the elevation-wheel pinion, is relatively simpler and is not described here.

The tipping structure analytical model contains approximately 1145 nodes, 3400 unconstrained displacement degrees of freedom, 3900 rod members, and 90 membrane plates, and is redundant to about the 500th degree. The quadripod contains about 200 nodes and consists of four towerlike legs connected to an apex structure that supports the subreflector. The backup structure (Fig. 5) consists of 24 main and 24 secondary radial rib trusses, 12 circumferential hoop trusses, and other bracing. Parasitic reflective surface panels that support only their own weights and local surface loads are attached to

450 front nodes of the rib trusses. Eight of the main rib trusses are connected to and are supported by an octagonal frame of the elevation-wheel structure.

The elevation wheel (Fig. 6) contains about 50 nodes, provides the transition from backup structure to alidade, and supports the elevation-wheel bull gear. Backup structure and elevation-wheel structures are redundantly connected at eight points, which complicates their design. Design of the quadripod members is less affected by redundant coupling. Individual bar and plate member elements are grouped for fabricational convenience into 190 distinct design variable groups.

A. Design for Performance Constraints

Simultaneous rms pathlength and pointing error constraints are imposed for gravity and wind loadings. The antenna rigging angle is established to make the pathlength errors for gravity loading equal at horizon and zenith antenna elevations.

The design example here is subjected to the nine pathlength and pointing error constraints listed in Table 2. These nine constraints are known to be the most demanding for this model. No primary stress constraints are imposed; all stress requirements are treated as side constraints with minimum sizes established, as described previously, from commercial shape tables.

Although the performance constraints are specified for operational wind speeds of 13.4 m/s (30 mph), the structure must withstand wind speeds of 31.2 m/s (70 mph) at any orientation and 44.7 m/s (100 mph) when stowed at the zenith elevation. To satisfy these requirements, the wind-pressure loadings applied for design are at the higher wind speeds, and the allowable value for the constraint is increased accordingly.

The computer run was limited to six analysis cycles and five design cycles. Figure 7(a) shows the cyclic progression histories of structure weight and of the most severe constraint ratio. Note that the designs at cycles 1, 2, and 4 are not feasible because of constraint ratios greater than unity. Nevertheless, the discrepancy at cycle 4 between actual and predicted constraint ratios indicates that the design procedure operated correctly in cycle 3 by developing a design with a predicted constraint ratio of unity. Unfortunately, the effect of structural redundancy produces the unanticipated response in cycle 4. The move limit used to control redundancy is equal to 1.5, which requires each design variable to be at least 2/3 and not more than 3/2 of its value in the prior cycle. A smaller move limit could have overcome part of the excessive constraint ratio at cycle 4, but would have caused additional problems at cycle 2. The predicted constraint ratio here is greater than unity because the move limit activated a sufficient number of side constraints to prevent the free choices of the design variables necessary to satisfy the optimality criteria.

Figures 7(b) and (c) show the cyclic constraint ratio histories for the four pathlength error constraints and for the five pointing error constraints. These figures show the dichotomy of pointing and pathlength error performance. One example of this is seen at cycle 4, where the gravity pathlength error constraint is exceeded although the pointing error constraint is satisfied. Another example is the 0° elevation, 120° yaw wind-loading case; the pointing error constraint is always active, and yet, except for the first cycle, the pathlength error constraint never is. Nevertheless, the design in Fig. 7 is successful because there was no increase in structure weight from the first to the sixth design cycle, and all nine excessive constraint ratios at the first cycle became feasible at the sixth.

A more erratic design history is shown in Fig. 8. The structure model is analogous to but slightly larger than the previous and has two more constraints. The effect of redundancy is emphasized by the relatively large move limit of 2.0. This is assumed to be responsible for the significant constraint violations at cycles 2, 3, 5, and 6 since the predicted constraint ratios, which were close to unity at cycles 3 through 6, indicate successful execution of the design procedure. In particular, the design at cycle 4, which is 15% lighter than the initial, is feasible because the large initial constraint ratio has been reduced to unity.

B. Discrete Versus Continuous Design Variables

A design that chooses the nearest available discrete shape to approximate the assumed continuous spectrum for design variables is compared with a design based upon a continuous spectrum. The 16-cycle histories shown in Fig. 9 are for design of the backup structure of a hypothetical 40-m antenna. Sizes of the design variables for the continuous design case are determined by the optimality criteria whenever the lower bound side constraint does not control; but whenever stress or maximum slenderness ratio governs the selection, the appropriate discrete size is chosen. The small move limit of 1.25 was used in an attempt to obtain smooth convergence. The normalized structural weights plotted were obtained after feasibility scaling, which consisted of multiplying the structure weight by the largest constraint ratio for constraint ratios greater than unity. The figure shows that the small differences in weight in the early cycles tend to disappear as the designs approach convergence.

C. Multiple Constraint Versus Envelope Method Designs

The last example compares designs obtained by solving for the Lagrangian multipliers simultaneously to satisfy the mul-

tiple constraint conditions of Eq. (7) with the old envelope method (Refs. 1 and 3). The envelope method uses sequential explicit decoupled solutions for the Lagrangian multipliers and applies the optimality criteria to each solution. The final selection of design variables at each design cycle is the envelope of maximum values obtained from the decoupled solutions.

The example is for a 34-m antenna tipping structure model similar to the one described at the beginning of this section, but with only seven primary pathlength and pointing error constraints. The design history in Fig. 10(a) shows similar structural weights achieved for the fourth through sixth cycles. Figure 10(b) shows that both methods have overcome significant first-cycle violations of the first, fourth, and fifth constraints and that the two methods appear to be equivalent for the third through seventh constraint ratios. But at the sixth cycle, the envelope method is not feasible because of the first and second constraints, while the multiple constraint method is feasible for all seven constraints. Computer run times were about the same for both methods.

V. Computer Resources

Problem size and storage capacity for the JPL-IDEAS computer program is primarily limited by the requirement of keeping a triangular matrix of maximum wavefront size in core during stiffness matrix decomposition. The 34-m antenna problem has a maximum wavefront of 220 degrees of freedom, which requires core storage for about 55,000 36-bit words to contain the double-precision decomposition triangle and associated pointers. Program source code and other storage bring the total requirement to 110,000 words. The design program operates on a UNIVAC 1100/81 computer, which can provide at least twice this in-core storage. Therefore, problems about 50% larger than this could be processed by the present program.

One complete design cycle uses 348 central processing unit (cpu) seconds. Of these, 90 s are associated with the design problem: 83 s are used for constructing and solving the pathlength and pointing error virtual loading vectors, and the remaining 7 s are used to determine the Lagrangian multipliers and apply the optimality criteria. Furthermore, the 83 s for vector processing appears to be excessive and could be reduced significantly by improvement of the present program.

One cycle for statics analysis with no provisions for design would require 336 s. This time is derived by subtracting the 90 s used for design and adding 78 s used during a preamble phase of the program to read and provide initial processing of input data. The following times, in cpu seconds, are used for analysis: stiffness matrix decomposition, 209; load-displacement vector processing, 40; stress resultant recovery, 5; pathlength and pointing error analysis, 4. A total of 2157 cpu seconds is used for the six cycles of Fig. 7.

VI. Summary

The optimality criteria method is reviewed for design of large ground antennas with performance constraints on microwave pathlength errors. The formulation is extended to include constraints on antenna boresight pointing errors simultaneously with the pathlength constraints.

Examples drawn from practice show that significant performance improvements and low structure weight are achieved within as few as six cycles of computer analysis and redesign. Furthermore, optimization was found not to extend maximum core size requirements and adds only a moderate increase in the cpu time for a design cycle beyond that of a "standard" analysis cycle. This implies that computer-automated design optimization run times could be less than 10 times that for a single analysis cycle.

A practical method is described to automate selection of structural member design variables from a discrete table of commercially available structural shapes. An example comparison of a discrete shape design with a design that assumed a continuous spectrum shows no major discrepancies for the approximations of discrete selection.

An example comparison of the older envelope design method with the present simultaneous multiple constraint design method shows that, although the designed structure weights are almost the same, the envelope method violates two of the seven performance constraints.

References

1. Berke, L., and N. Khot, "Use of Optimality Criteria Methods for Large Scale Systems," AGARD Lecture Series No. 70, *On Structural Optimization*, Hampton, Va., Oct. 1974, pp. 1–29.

2. Cella, A., and R. Logcher, "Automated Optimum Design from Discrete Components," *Journal of the Structural Division*, Proc. ASCE 97 (ST–1), Proc. Paper 7845, Jan. 1971, pp. 175–189.

3. Gellatly, R., L. Berke, and W. Gibson, "The Use of Optimality Criteria in Automated Structural Design," 3rd Conference on Matrix Methods in Structural Mechanics, WPAFB, Ohio, Oct. 1971.

4. *Guide for Design of Steel Transmission Towers*, ASCE Manual and Report on Engineering Practice, No. 52, 1971.

5. Kamat, M., and R. Hayduk, "Recent Developments in Quasi-Newton Methods for Structural Analysis and Synthesis," AIAA/ASME/ASCE/ACS 22nd SDM Conference, Paper 0576, Atlanta, Ga., Apr. 6–8, 1981.

6. Khot, N., L. Berke, and V. Venkayya, "Comparison of Optimality Structures by the Optimality Criterion and Projection Method," AIAA/ASME/SAE 20th SDM Conference, Paper 79–0720, St. Louis, Mo., Apr. 4–6, 1979.

7. Khot, N., L. Berke, and V. Venkayya, "Comparison of Optimality Criteria Algorithms for Minimum Weight Design of Structures," *AIAA Journal*, Vol. 17, No. 2, Feb. 1979, pp. 182–190.

8. Levy, R., and R. J. Melosh, "Computer Design of Antenna Reflectors," *Journal of the Structural Division*, Proc. ASCE 99 (ST-11), Proc. Paper 10178, Nov. 1973, pp. 2269–2285.

9. Levy, R., "Antenna Bias Rigging for Performance Objective," IEEE Mechanical Engineering in Radar Symposium, Wash. D.C., Nov. 8–10, 1977.

10. Levy, R., and W. Parzynski, "Optimality Criteria Solution Strategies in Multiple Constraint Design Optimization," *AIAA Journal*, Vol. 20, No. 5, May 1982, pp. 708–715.

11. Levy, R., "Computer-Aided Design of Antenna Structures and Components," *Computers and Structures*, Vol. 6, 1976, pp. 419–428.

12. Rusch, W., and P. Potter, *Analysis of Reflector Antennas*. Academic Press, New York, London, 1970.

13. Sander, G., and C. Fleury, "A Mixed Method in Structural Optimization," *International Journal for Numerical Methods in Engineering*, Vol. 13, 1978, pp. 385–404.

14. Schmit, L., "Structural Synthesis: Its Genesis and Development," *AIAA Journal*, Vol. 19, No. 10, Oct. 1981, pp. 1249–1263.

15. Schmit, L. A., and C. Fleury, "Structural Synthesis by Combining Approximation Concepts and Dual Methods," *AIAA Journal*, Vol. 18, No. 10, Oct. 1980, pp. 1252–1260.

16. Schmit, L. A., and C. Fleury, "Discrete-Continuous Variable Structural Synthesis Using Dual Methods," *AIAA Journal*, Vol. 18, No. 12, Dec. 1980, pp. 1515–1524.

17. Utku, S., and S. M. Barondess, *Computation of Weighted Root-Mean-Square of Path Length Changes Caused by the Deformation and Imperfections of Rotational Paraboloidal Antennas*, Technical Memorandum 33–118. Jet Propulsion Laboratory, Pasadena, Calif., Mar. 1963.

List of Symbols

a	cross-sectional area
$\underline{a}$	lower limit on area
$\overline{a}$	upper limit on area
C^*	prespecified maximum value of constraint
C	computed value for constraint in current design
D	constraint ratio
dn	normal component of distortion vector
F	sensitivity coefficient
G	constraint equation
L	length
R	half-pathlength error
V	volume
γ	direction cosine of the surface
λ	Lagrangian multiplier

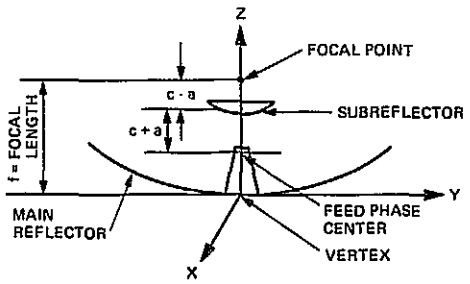
Vectors and Matrices

$\mathbf{u}$	displacement vector
$\mathbf{h}$	vector of fitting parameters
$\mathbf{A}$	geometric transformation matrix
$\mathbf{B}$	geometric transformation matrix
$\mathbf{H}$	invarient matrix as defined in Eq. (11)
$\mathbf{W}$	weighting matrix (diagonal)

Subscripts

i	design variable index
j	constraint index
z	reference to the Z axis

Table 1. Cassegrain antenna pointing error contributions



Definitions

k = beam deviation factor ($k \sim 0.75$)

m = magnification factor

$m = (c + a)/(c - a)$

($m \sim 5.0 - 10.0$)

Contributing displacement	Factor	
	X axis	Y axis
Best-fit paraboloid		
Vertex shift	$-k/f$	k/f
Rotation	$1 + k$	$1 + k$
Subreflector		
Lateral translation	$(1 - 1/m)k/f$	$-(1 - 1/m)k/f$
Rotation	$-2\,(c - a)k/f$	$-2\,(c - a)k/f$
Feed		
Lateral translation	$k/(mf)$	$-k/(mf)$

Pointing Error = Σ (Displacement X Factor)

Table 2. Performance constraints

Condition			Half-pathlength root-mean-square, mm (in.)	Pointing, arc seconds
Wind, m/s (mph)	Elevation, degrees	Yaw, degrees		
13.4 (30)	0	120	0.330 (0.0130)	30.0
13.4 (30)	60	180	0.330 (0.0130)	30.0
13.4 (30)	90	90	0.330 (0.0130)	30.0
13.4 (30)	90	180	None	30.0
Gravity-worst case: horizon to zenith			0.165 (0.0065)	75.0

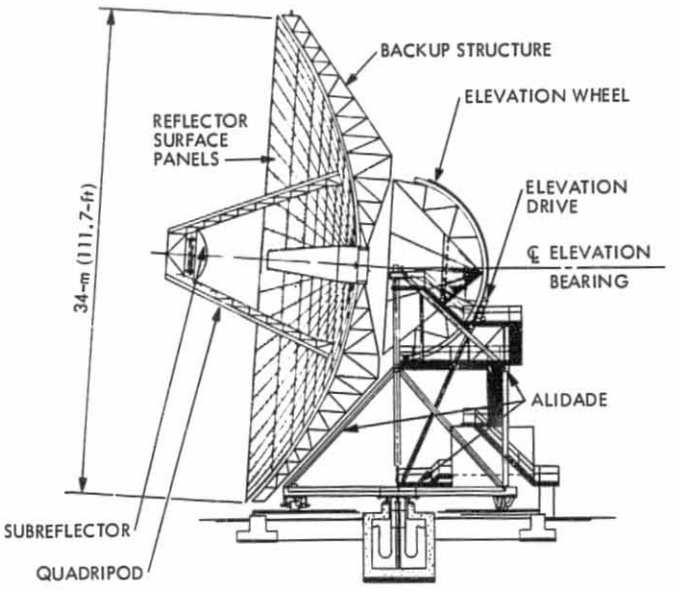


Fig. 1. 34-meter Deep Space Network antenna

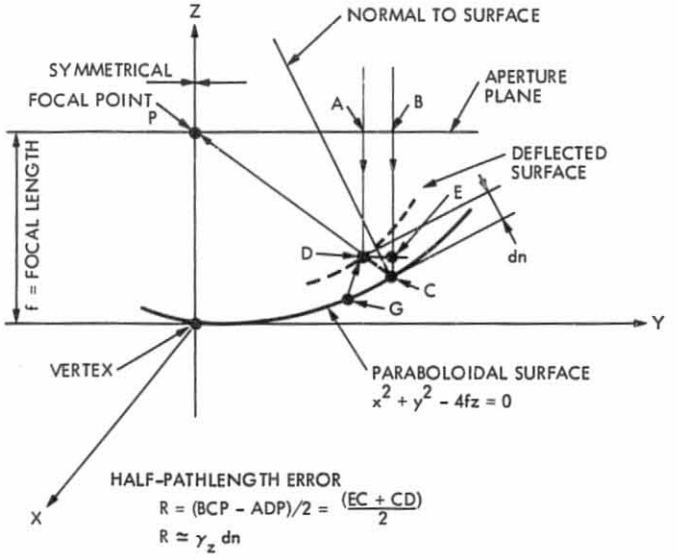


Fig. 2. Surface deflection geometry

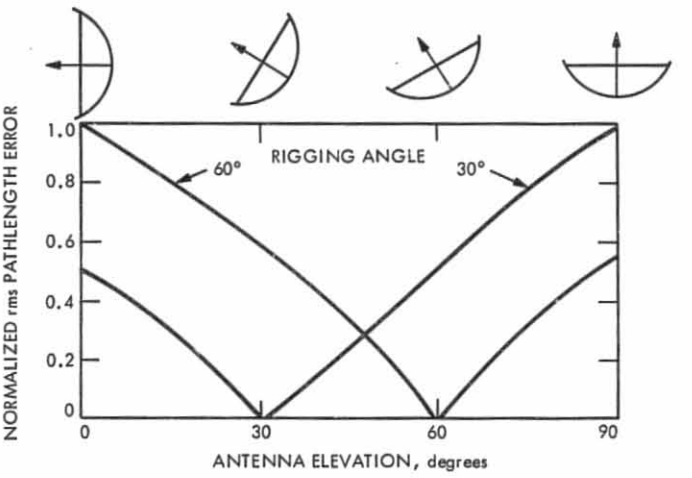


Fig. 3. Pathlength error distribution vs rigging angle

HANDBOOK PROPERTIES FOR PIPES

******LOAD TABLE******

HANDBOOK NO. SHAPE	AREA	RAD	SPAN LENGTHS 25.	50.	75.	100.	125.	150.	175.	200.
1).75STD, 1.05X.113	.333	.330	8.4	4.1	.0	.0	.0	.0	.0	.0
2)1.0-10, 1.315X.109	.413	.430	11.2	8.3	3.9	.0	.0	.0	.0	.0
3)1.0STD, 1.315X.133	.494	.420	13.3	9.8	4.4	.0	.0	.0	.0	.0
4)1.25-10, 1.66X.109	.531	.550	15.0	12.4	8.2	4.6	.0	.0	.0	.0
5)1.5-10,1.90X.109	.613	.630	17.7	15.2	12.2	7.0	4.5	.0	.0	.0
6)1.25STD, 1.66X.140	.669	.540	18.8	15.5	9.9	5.6	.0	.0	.0	.0
7)2.0-10, 2.375X.109	.776	.800	22.9	20.6	17.9	14.2	9.1	6.3	.0	.0

Fig. 4. Excerpt from common member shape table

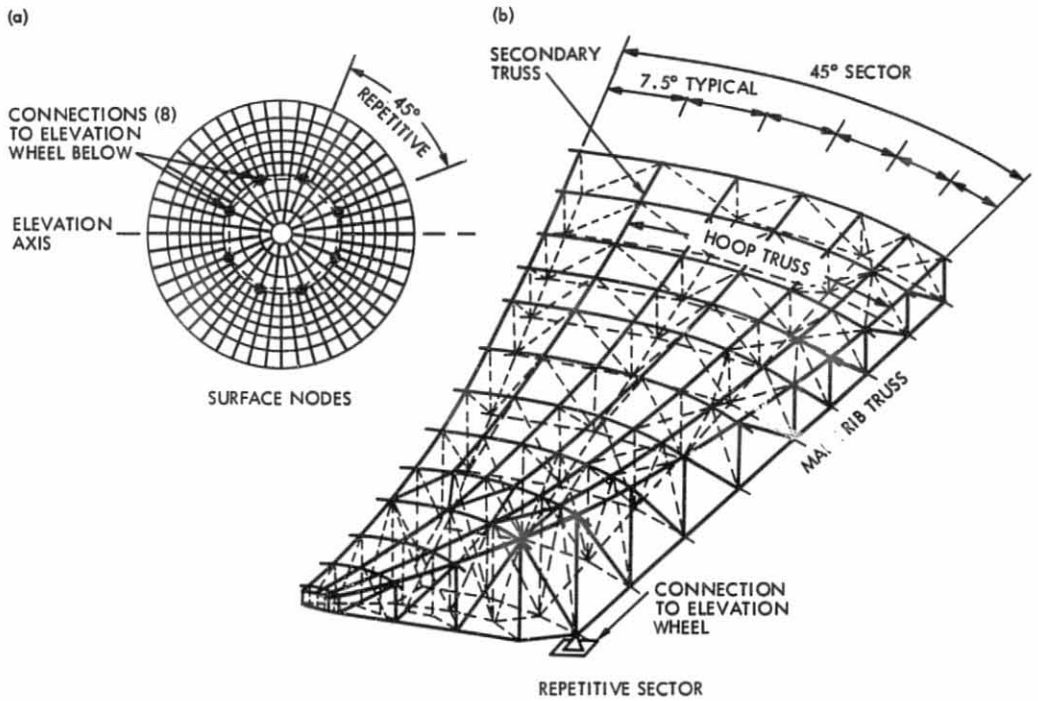


Fig. 5. 34-meter antenna backup structure: (a) surface nodes; (b) repetitive sector

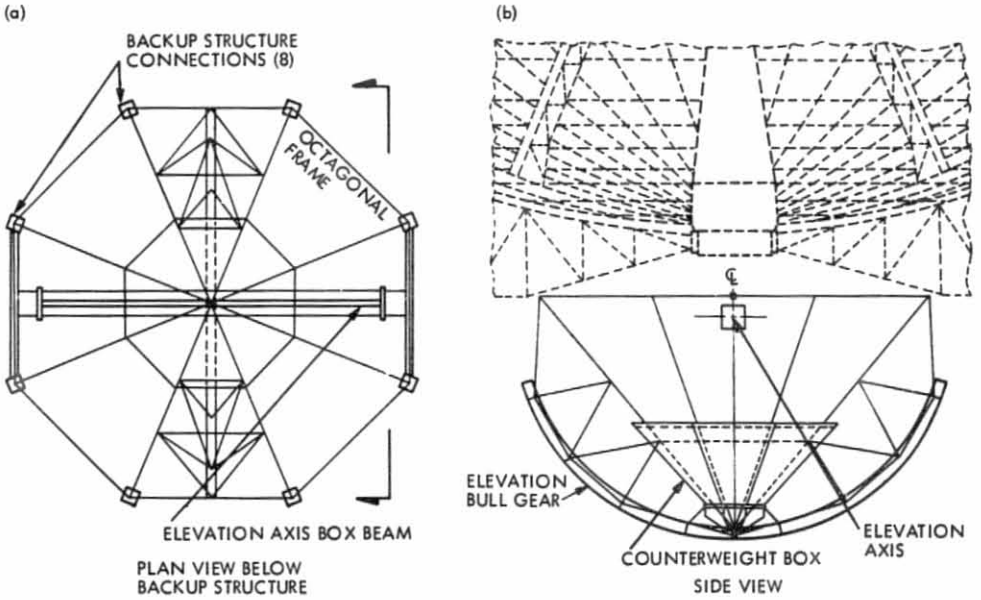


Fig. 6. 34-meter antenna elevation wheel: (a) plan view below backup structure; (b) side view

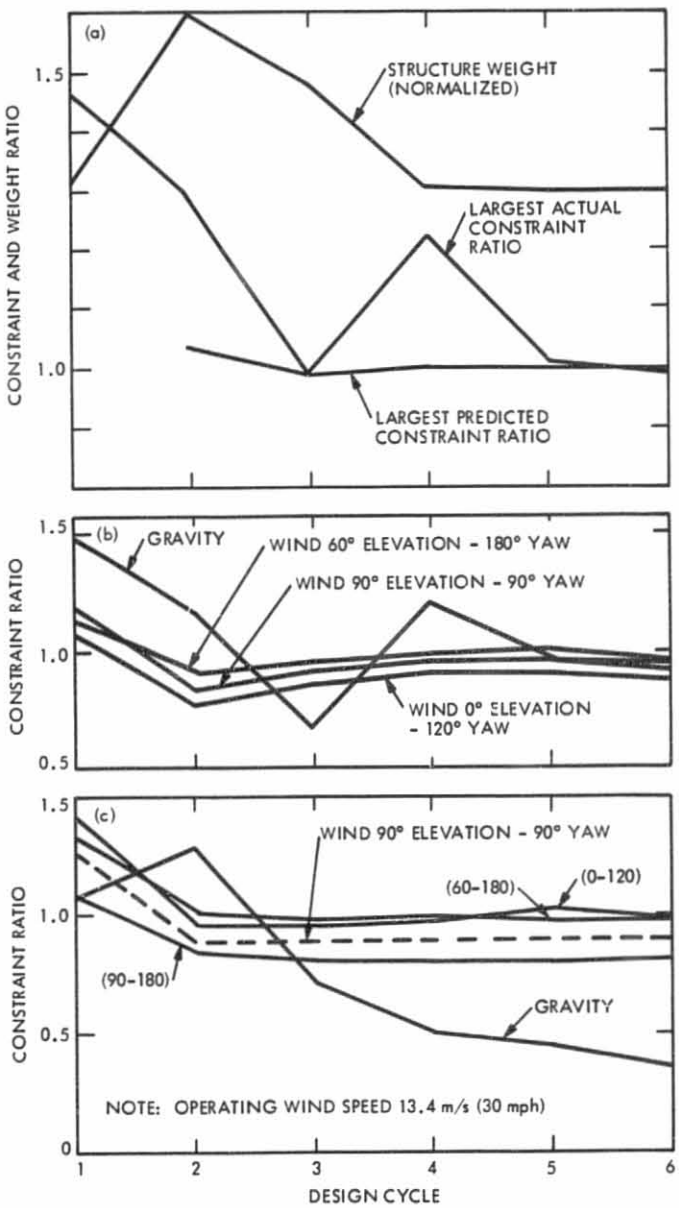


Fig. 7. 34-meter antenna tipping structure design history: (a) weight and controlling constraint ratios; (b) pathlength error constraint ratios; (c) pointing error constraint ratios

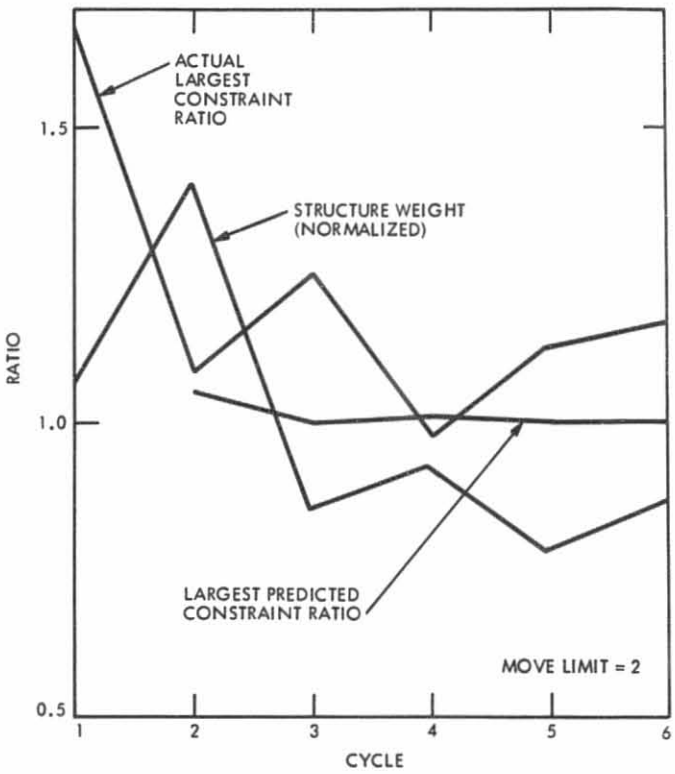


Fig. 8. 34-meter antenna cyclic weight and constraint ratio history, move limit = 2.0

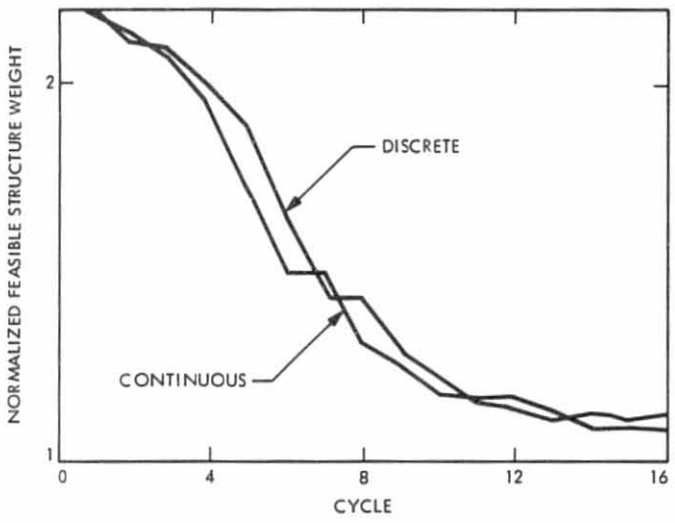


Fig. 9. 40-meter backup structure designs: discrete vs continuous design variables

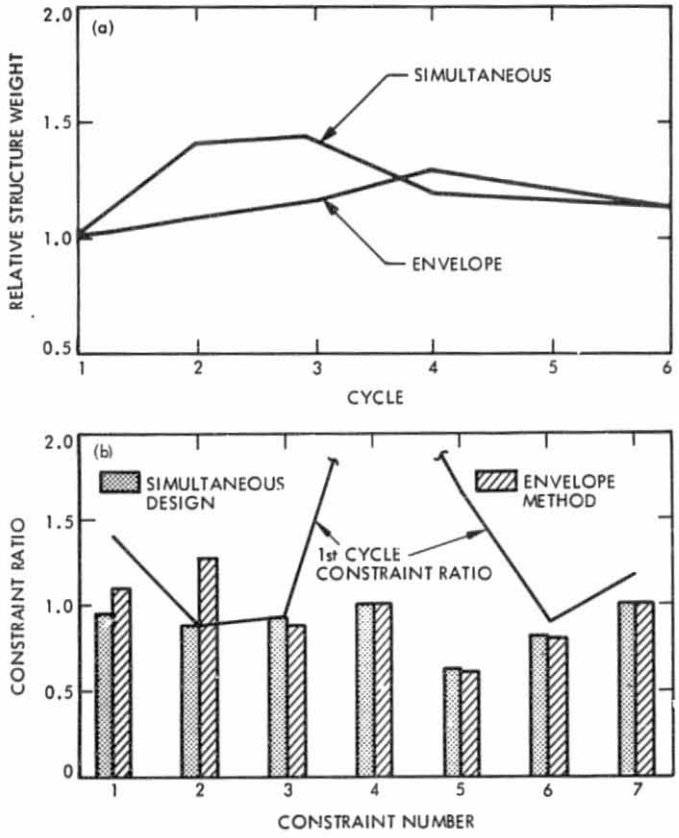


Fig. 10. 34-meter antenna simultaneous multiple constraint envelope method designs: (a) structure weight history; (b) sixth cycle primary constraint ratios

High Power K_a-Band Transmitter for Planetary Radar and Spacecraft Uplink

A. M. Bhanji, D. J. Hoppe, R. W. Hartop,
E. W. Stone, and W. A. Imbriale
Jet Propulsion Laboratory

D. Stone and M. Caplan
Varian Associates, Inc.

A proposed conceptual design of a 400 kW continuous wave (CW) K_a-band transmitter and associated microwave components to be used for planetary radar and serve as a prototype for future spacecraft uplinks is discussed. System requirements for such a transmitter are presented. Performance of the proposed high-power millimeter wave tube, the gyroklystron, is discussed. Parameters of the proposed power amplifier, beam supply, and monitor and control devices are also presented. Microwave transmission line components consisting of signal monitoring devices, signal filtering devices, and an overmoded corrugated feed are discussed. Finally, an assessment of the state of the art technology to meet the system requirements is given and possible areas of difficulty are summarized.

I. Introduction

Due to present user crowding of the S- and X-band microwave spectrums assigned to space exploration and to anticipated new user requirements that can be met by higher frequencies, the Jet Propulsion Laboratory has initiated a design and development program for a next generation K_a-band (34 GHz) continuous-wave (CW) transmitter that will first be used for planetary radar applications where shorter wavelength offers increased resolution of targets and improved ranging capabilities. The technology will be transferable to uplink communication with future spacecraft. The scientific value of a K_a-band radar can be assessed by breaking the criteria down into those that rely on the shorter wavelength to give new information and those that require greater two-way antenna gain. The shorter wavelength will be especially valuable in the investigation of Jupiter's moons, the rings of Saturn, and the detection of rain in the upper atmosphere of Venus. The number of detectable asteroids in a ten-year span will be enhanced by a factor of 10 due to higher antenna gain.

The experimental transmitter will be installed on the 64 meter Cassegrain antenna at Goldstone, California (Fig. 1) (Ref. 1), which is equipped with a rotatable, asymmetric hyperboloidal subreflector that permits the use of multiple feed systems at the Cassegrain focus. The subreflector can be precision indexed to a fixed number of positions that will allow each feed to properly illuminate the main reflector. The 64-meter antennas were designed in the early 1960's for S-band operation and subsequent modifications have resulted

in some 60 percent aperture efficiency at 8.5 GHz (Ref. 1). A block diagram of the existing transmitters and the proposed K_a-band transmitter to be installed on the 64 meter antenna is shown in Fig. 2. To achieve a reasonable antenna efficiency at the 34 GHz band, an upgrade of the antenna dish structure and surface panel accuracy is required and is part of a planned implementation project to extend the 64-meter diameter dish to 70 meters and incorporate shaped surfaces. The increased antenna diameter and shorter wavelength are expected to provide an antenna gain of 84 dB, and with the proposed 400 kW CW transmitter, the K_a-band radar system will have an effective radiated power of about 100 trillion watts.

Using an assessment of the state-of-the-art technology, this article will present a proposed conceptual design and calculated performances of a 400 kW CW K_a-band transmitter including overmoded transmission line components (e.g., taper, coupler and a mode filter) and an overmoded antenna feed system.

II. Transmitter System Requirements

The transmitters for radar astronomy systems differ from conventional radar systems in that they require a high average power, rather than high peak power, over the bandwidth required to handle the transmitted signal (Ref. 2). It is also important that these transmitters be coherent in order to determine the phase relationships of the returned signals, and they must have very high phase stability if measurements are to be made over long periods of time. The transmitter must also be capable of modulation by a variety of pulse programs while maintaining the phase and amplitude fidelity and pulse to pulse stability required for pulse compression systems incorporated in the radar.

The performance of a pulse compression system (Ref. 3) depends upon the transmitter stability, phase response, and amplitude response to preclude the creation of large time sidelobes (spurious amplitude responses on either side of the main compressed signal). By control of these characteristics and the use of weighting techniques, the compressed pulse system time sidelobes can be kept below −30 dB. If the transmitter phase pushing is not to degrade this level, spurious sidelobes due to the transmitter alone must be kept below −40 dB. The sidelobe level (Ref. 3) is given by:

$$SL = 20 \log \frac{\Delta P}{2} \qquad (1)$$

where ΔP is the peak phase ripple during the transmitter pulse. For a subsystem time sidelobe level of −40 dB, $\Delta P = 1.15°$.

The requirements for such a radar transmitter are then:

(1) High power

(2) Low incidental phase modulation (jitter)

(3) Phase stability

(4) Low incidental amplitude modulation

(5) Amplitude stability

(6) Frequency stability

(7) Bandwidth

(8) Low harmonics

(9) Low noise

(10) Phase modulation (phase code pulse compression (PN))

(11) Frequency shift keying (FSK)

(12) Linear capability

The above requirements illustrate that high power alone will not provide the desired CW radar transmitter capabilities. If this were the case, it might be more easily obtained with an oscillator instead of an amplifier. Besides the appeal of having dynamic control of amplitude and phase, the appeal for using an amplifier is that it eliminates the need for phase locking an oscillator to a control signal. The control signal itself could, after amplification, be the transmitted signal.

Based on the above requirements, the K_a-band radar transmitter specifications are given in Table 1.

III. The Transmitter

As shown in Fig. 3, the transmitter will include an existing power supply that converts 2400 V, 3 phase, 60 Hz line voltage to direct current at up to 90 kV with a power limitation of 1.1 MW for the gyroklystron (to be described later) amplifier beam. The frequency synthesizer and the exciter will provide an input signal to this 400 kW gyroklystron amplifier which will provide approximately a 50 dB power gain (goal). The automated transmitter control will furnish monitoring and control of all functions, while some 40 protective devices (interlocks) will prevent damage to equipment by removing voltage and in some cases drive power in event of a malfunction. The liquid to air 1.5 MW heat exchanger will be used to cool the amplifier, the power supply, various auxiliaries to the transmitter, and microwave components of the transmission line.

A. High Power Millimeter Amplifier Tube

The gyroklystron is a potential candidate.

1. Requirements background. The requirements of high power, high gain, ease of modulation and an output spectrum free from spurious signals and noise makes a klystron linear beam tube the natural choice for radar. However, at K_a-band frequencies, an examination of commercially available linear tubes shows there are no conventional klystrons, twystrons, or TWTs that can meet the specifications given in Table 2 (these specifications are derived from the K_a-band radar transmitter specifications given in Table 1). These conventional microwave tubes cannot be scaled to 34 GHz (K_a-band) and maintain high efficiency and high power because of heat transfer problems in the relatively small electron interaction volume. The result is that less conventional devices operating in higher order modes (thus larger in circuit area and dissipation capability) become attractive, and the usual $P = K/f^2$ (Ref. 4) scaling condition can be ignored.

The *gyrotron* (Ref. 5) or cyclotron resonance maser is such a device. The interaction mechanism responsible for microwave amplification in the so called fastwave device is azimuthal phase bunching due to the dependence of the electron cyclotron frequency on the relativistic electron mass and magnetic coupling to RF fields in the cavity/waveguide close to cut-off. Experimental research in this field has produced some exciting and impressive powers and efficiencies (Ref. 6).

However, these oscillators are not suitable for all applications (the primary use is for plasma heating) because they are essentially fixed frequency devices. There is considerable interest in gyroamplifiers for radar systems and satellite communications.

Early gyro-TWT amplifier development was promising but results did not greatly surpass the performance of conventional TWTs. The reason was that since the gyro-amplifier has a much longer interaction region than the gyro-oscillator, the destruction of phase bunching due to velocity spread in the electron beam is greater. However, improved gun design with a small velocity spread has led to some significant results. High efficiencies (25%) have been demonstrated in an experimental C-band gyroTWT at Varian Associates (Refs. 7 and 8). More recently, Varian has operated a 94 GHz gyroTWT with a 30 dB gain, 2% bandwidth, output power larger than 20 kW and an efficiency of 8%. The mode of operation is TE_{11}. There is also ongoing work at the Naval Research Laboratory (NRL) in Washington, D.C., to develop a wide bandwidth (10%) gyro-TWT (Ref. 9).

There continues to be interest in developing a gyroklystron amplifier consisting of resonant gyro-oscillator type cavities separated by drift tubes. An experimental gyroklystron is now being assembled at NRL with the following objectives (Ref. 10): frequency 4.5 GHz (the experiment is being performed at C-band due to the availability of a Magnetron Injection Gun (MIG) suitable for this frequency range); small signal gain >40 dB; efficiency >35%; output power >100 kW; and a bandwidth of about 1%. Measurements will be made to determine such characteristics as efficiency, gain, bandwidth as a function of magnetic field, and cavity tuning. An earlier ORNL K_a-band gyroklystron (Ref. 11) built at Varian did not meet design goals of power and efficiency and it was recognized that the principal difficulty with the 28 GHz gyroklystron was oscillation in the drift tubes, which were capable of supporting propogating modes because the buncher cavities themselves were designed to operate in the overmoded TE_{011} mode. However, a great deal of practical engineering information has been learned from this first gyroklystron work, and with the inclusion of Varian's design study of a gyroklystron for the NASA Lewis Research Center (Ref. 12) and work at NRL (described above), the gyroklystron appears to be the natural choice for a JPL K_a-band transmitter that will meet the tube specifications listed in Table 2 for high gain, high efficiency, high power, and narrow bandwidth. A paper design of the proposed 34 GHz gyroklystron and its characterization in terms of efficiency, instantaneous bandwidth, small signal and saturated gain, AM and PM sensitivity to operating parameters, spectral purity, phase stability and noise figure follows.

2. Paper design of JPL gyroklystron. The overall design considerations for gyroklystrons resemble in many ways the design procedures for conventional klystron amplifiers, even though the gain mechanism involves a beam interaction with cyclotron waves rather than space-charge waves. Caplan (Ref. 13) has shown that the same equivalent circuit used for many years in describing klystrons can be used to describe gyroklystrons if transadmittance, beam loaded Q's, and beam reactance are appropriately redefined for a gyroklystron.

The basic design approach is shown in Fig. 4. A Pierce gun produces a beam which is injected into a wiggler field to impart the rotational energy to a solid beam and then adiabatically compressed to the circuit. The circuit consists of 3 buncher cavities operating in the TE_{111} mode and an output cavity in the TE_{121} mode. A small, tapered transition is made to 0.787 in. diameter, at which point there exists a 7-inch-long mode converter to transform the TE_{12} circularly polarized mode to the TE_{11} circularly polarized output mode. The waveguide is then tapered optimally to the collector region where the spent beam is deposited. The down taper region then leads to a face-cooled double disc 2.5-inch diameter window.

The gyroklystron circuit consists of a number of cylindrical cavities operating in the transverse electric mode interconnected by cut-off drift tubes through which the rotating electron beam passes. As in conventional klystrons, the input

cavities and drift tubes serve to prebunch the beam while the actual rf energy interaction occurs in the output cavity, which in this case is an open irregular resonator. Designing a circuit consists of choosing the cavity lengths, drift tube lengths, and cavity Q's such that when an electron beam of desired voltage, current and rotational energy is made to pass through the circuit, the required gain, bandwidth and efficiency will be obtained, while at the same time circuit stability against spontaneous oscillations is maintained.

Because of power handling considerations and mode integrity, the rf power cannot be extracted conventionally through a coupling hole at right angles to the device, but must travel in line with the beam through an up taper and collector region.

As stated earlier, gyrotron-type devices can operate at high power and high frequencies; however, this introduces the problem of suppressing other unwanted modes in the cavity whose cutoff frequencies are close to the desired mode. Since the desired output mode in this case is TE_{11} circularly polarized, which is the dominant lowest order mode, the choice of this operating mode for the gyroklystron is desirable since it rules out mode competition. However, the requirement of 400 kW CW at 34 GHz demands a higher order mode in the output cavity if the power densities dissipated on the cavity walls are to be less than 1 kW/cm^2 (proven standard heat transfer technology at this time). Since there is no energy exchange in the input buncher cavities, these can operate in the TE_{111} dominant mode, thus eliminating the problem of unwanted interaction in the drift tubes. One can now have an overmoded output cavity, provided the symmetry of the output cavity mode is compatible with the bunching pattern imposed on the beam from the TE_{111} mode. This then requires the output cavity to be of the TE_{1n1} type; i.e., the azimuthal index 1 must be the same as that for TE_{111}. The power dissipation in such a cavity is given by Ref. 14 as:

$$P(W/cm^2) = 4.8377 \times 10^{-5} \frac{f_0^{5/2} P_0 Q_{ext}}{K_{mn}^2 (L/\lambda_0) \left(1 - \frac{m^2}{K_{mn}^2}\right)} \quad (2)$$

where

f_0 = frequency (GHz)

P_0 = output power (kilowatts)

Q_{ext} = external Q of output cavity

K_{mn} = transverse mode number for TE_{mn} mode

L/λ_0 = cavity length in units of free space wavelengths

The lowest order mode that will satisfy the criterion that $P < 1$ kW/cm^2 is the TE_{121} mode with $Q_{ext} \leq 400$ (a requirement in order to achieve bandwidth of 0.1%) and $(L/\lambda_0) \geq 3$ (for efficiency).

The overall stability of the gyroklystron amplifier (against the propensity to become an oscillator) is ensured by using sufficiently short cavities and low enough cavity Q's such that the threshold currents required for each cavity to self-oscillate are above the operating current. A sufficient number of cavities are then used to achieve the required gain of 50 dB.

As in all gyrotrons, one is required to create an electron beam with 70-80% of its energy in rotational motion. Two methods used for producing such a beam are the MIG and the Pierce gun/wiggler configuration. The wiggler configuration has not yet been used in oscillators as has the MIG, and thus has not yet had the opportunity to demonstrate, in a direct comparison, the capability of generating beams of a quality (low velocity spread) which equal or surpass MIG beam quality. One advantage of using the wiggler is that standard Pierce gun technology can be used, which ensures operation with a space charge limited beam (resulting in low noise, an important requirement for radar and communications tubes).

With the output cavity operating in the TE_{121} mode, power must be converted back into the desired TE_{11} mode and run into an overmoded waveguide in order for the output window and collector to handle the CW power requirement. A symmetric ripple wall mode converter will be used to convert TE_{12} to TE_{11} with > 99% conversion efficiency. The collector will have to be at least 4 inches in diameter for reliable power handling capability during operation with 1 MW CW of beam power. This requires tapering up the waveguide to 4 inches and then back down to the "standard" 2-1/2-inch output window. Non-linear gaussian tapers which can maintain 95% mode purity in these tapered sections will be used. A mode filter will be required in the external transmission line to remove this last 5% power in undesired modes, which consist mainly of the TE_{12}, TE_{13}, as well as TM_{11}, TM_{12}, TM_{13}, and TM_{14} modes. The next section will cover the calculated performances of a few of the tube components and their parameters, including operating characteristics.

3. Calculated performances and operating characteristics.

a. Circuit design. The calculated performance of the circuit design consists of a computer-simulated analysis of a large number of possible circuit configurations, using the small signal gyroklystron gain program for gyroklystron amplifiers developed at Varian by Caplan (Ref. 13) and a large signal efficiency program.

For a given set of input parameters, the small signal gain program calculates gain, beam-loaded Q's, beam reactance, transadmittances, cavity amplitudes, and phases, all as a function of frequency and magnetic field. These results are then used as inputs to a large signal code, which integrates a large number of particle trajectories through electric fields (calculated in the small signal code) to determine the amount of energy lost by the beam. This code uses an iteration procedure to determine saturated field amplitude in the last cavity. It is assumed that the field amplitudes and phases in the input cavities calculated from linear theory are valid even at saturation, since these cavities only serve to prebunch with little energy extraction. For a final check on the design and performance, an eleborate particle simulation code that uses up to 5000 particles will be used. Subtle effects such as beam loading on the mode structure can be studied and very accurate estimates of the effects of velocity spread on efficiency can be calculated.

A four cavity gyroklystron circuit design that meets the JPL specifications is given in Table 3. Figures 5 and 6 show calculated results of gain vs. frequency, and power out vs. power in (indicating a saturated output of 465 kW and 46.5% efficiency), respectively, for this gyroklystron. These calculations were based on the beam having zero axial velocity spread (cold beam). It is estimated that with a velocity spread (hot beam) of 5%, the efficiency would drop to 35%.

As stated earlier, a gyroklystron can be described by the same type of equivalent circuit as a regular klystron. The major difference is that the circuit elements such as beam loaded Q and transadmittance can have a strong dependence on magnetic field and frequency when compared with a conventional klystron. A small change in magnetic field can result in large changes in gain and efficiency. Figure 7 shows efficiency vs magnetic field for the above gyroklystron.

b. Mode converter. The power emitted from the output cavity will be in the TE_{12} mode, and therefore a mode converter must be designed to convert it to the desired TE_{11} mode. A relatively simple cylindrically symmetric mode converter (Ref. 15), which consists of a section of waveguide having a sinusoidal variation of the wall radius, $a(z)$, with distance, can be designed. For the converter, $a(z) = a_o + \delta_a \sin(2\pi z/\lambda_b)$, where a_o is the average wall radius, λ_b is the wavelength of the ripples, and δ_a is the amount of maximum wall perturbation. Mode conversion is accomplished by choosing the ripple wavelength equal to the beat wavelength between the TE_{11} and TE_{12} mode, $2/\lambda_b = 2/\lambda_{11} - 2/\lambda_{12}$. The optimum converter length is given approximately as $L_c = 0.638\lambda_b/(\delta_a/a_o)$. For a_o = 1 cm, which is about 25% above cutoff of the TE_{121} cavity, the beat wavelength is equal to 2.915 cm; choosing a ripple amplitude of 0.10626 cm makes the optimum converter length 6.9 inches. This design was checked using Varian's computer code which solves 20 coupled telegrapher's equations inside a waveguide of arbitrary wall profile. The code solves for the mode conversion occurring from the TE_{12} mode to all other modes with the same azimuthal symmetry, m = 1, which are the only conversions allowed in a symmetric converter. Figure 8 shows the calculated performance of such a converter. As expected, almost all power is converted to the desired TE_{11} mode, with the worst offender being TM_{12} at -23 dB (1/2%). Table 4 shows the calculated amount of power in each of the spurious modes at the end of the converter.

c. Beam collector. The collector has to satisfy two constraints which work directly against each other. It must dissipate high beam power at levels of no more than 1 kW/cm^2. This means the collector must be as large as possible, but at the same time it is desirable to keep the diameter of the output waveguide configuration (of which the collector is a part) as small as possible to ensure mode purity. Spurious modes will also be generated in the collector since the collector will contain tapers. For example, the up-taper will connect the 1" diameter waveguide at the end of the mode converter to the 4" diameter waveguide in the beam deposition region. Using non-linear taper designs, and the analysis by Sporleder and Unger (Ref. 16), a 4" taper design (optimized for a 60 GHz gyrotron) was examined for its mode conversion properties when used for the TE_{11} mode. A Varian program based on Sporleder and Unger's coupled mode theory, which solves for simultaneous coupling of the TE_{11} mode to as many TE_{1n} and TM_{1n} modes as are necessary to model the problem, was used. Figure 9 shows the total spurious mode level for all TE_{1n} modes versus axial distance. The taper described in Fig. 9 was not optimized to discriminate against TM_{1n} modes. With TM modes included in calculation, 10% of power was converted to TM_{11}. Further effort is being made to optimize the taper sections for the JPL gyroklystron to avoid conversion to TM modes.

d. The CW output window. The primary design approach for the CW output window is a double ceramic disc face-cooled with fluorocarbon liquid as shown in Fig. 10. Performance for a window of this type has already been demonstrated at K_a-band at power levels well above the 200 kW minimum (340 kW CW at 28 GHz). At 60 GHz the double disc concept has also proven capable of 200 kW CW (Ref. 17). These data prove that a 200 kW 34 GHz double disc window is feasible and indicate that a 400 kW CW 34 GHz double disc window can be designed to operate successfully.

Initial calculations for the 400 kW CW 34 GHz window, a design with sapphire window discs, offers ample bandwidth

(over 1.5 GHz for a Voltage Standing Wave Ratio (VSWR) ≤ 1.5:1 as shown in Fig. 11). Finite element heat transfer calculations predict a worst case peak temperature of 65.1°C at the center of the window face on the vacuum side. The resulting thermal stresses in the window material are well within conservative estimates of loads which cause sapphire breakage.

Having described the calculated performance of some of the components of the proposed gyroklystron, it is next appropriate to characterize the tube in terms of noise figure and AM and PM sensitivities.

e. Noise figure. The noise figure for the proposed gyroklystron configuration is calculated using the standard shot noise expression (Ref. 18).

$$P_{noise}\,(\mathrm{dB}) = 10 \log \left[\frac{2eI_o\,\Delta f\,\Gamma^2\,\beta^2\,G_{ss}\left(\frac{R_{shunt}}{Q_1}\right)Q_1}{P_{out}} \right] \quad (3)$$

where e is the electron charge, I_o is the beam current, Δf is the noise bandwidth, P_{out} is the output power, G_{ss} is the small signal gain, R_{shunt} is the shunt impedance of the first cavity, and Q_1 is the first cavity Q.

This expression has been found to be accurate for klystron amplifiers (Ref. 19) and it applies equally well to gyroklystrons. However, the parameters for the coupling cavity (β^2) and first cavity shunt impedance (R_{shunt}) must reflect the properties of the gyroklystron cavity design. Using the values, $R_{shunt}/Q_1 = 1433\Omega$, $Q_1 = 300$, $G_{ss} = 57$ dB, $\beta^2 = 0.9$, $\Gamma^2 = 2.39 \times 10^{-6}$, $I_o = 12.5$A, $\Delta f = 1$ MHz, and $P_{out} = 200$ kW, gives $P_{noise} = -110.3$ dB/MHz, which meets the tube specifications given in Table 2.

f. AM and PM sensitivities. In order to calculate AM and PM sensitivities with respect to various operating parameters of the proposed gyroklystron, responses of each section of the tube must be included in the calculation. For example, to determine the AM and PM sensitivity of the tube due to variation in the gun coil current, the following individual sensitivity factors must be multiplied together to provide overall sensitivity.

$$\left(\frac{\text{gauss}}{\text{amps}}\right)_{\substack{\text{gun coil}\\ \text{variation}}} \times \left(\frac{\%}{\text{gauss}}\right)_{\substack{\text{Pierce}\\ \text{gun}\\ \text{scalloping}}} \times \left(\frac{\%}{\%}\right)_{\substack{\text{wiggler}\\ \text{velocity}\\ \text{spread}}} \times \left(\frac{\%}{\%}\right)_{\substack{\text{adiabatic}\\ \text{compression}}} \times \left(\frac{\text{dB, deg}}{\%\ \ \%}\right)_{\substack{\text{circuit}\\ \text{sensitivity}}}$$

$$= \left(\frac{\text{dB}}{\text{amps}}, \frac{\text{deg}}{\text{amps}}\right)_{\substack{\text{Tube}\\ \text{sensitivity}}} \quad (4)$$

Tables 5 and 6 give very preliminary calculations of AM and PM pushing factors, respectively, with respect to cathode voltage, magnet coil current, and load VSWR, etc.

The results of a preliminary calculation of phase linearity are given in Fig. 5. Some parameters are yet to be calculated and will be given in a later report.

B. Power Amplifier

A functional block diagram of the proposed transmitter power amplifier is shown in Fig. 12. The power amplifier will be driven to a 400 kW CW output by a helix traveling wave tube preamplifier. Such a driver power preamplifier tube at 30 GHz with 40W output has been developed at the Nippon Electronics Corporation (NEC) in Japan (Ref. 20). The output of the preamplifier will be monitored through a 50 dB coupler and will be isolated from the gyroklystron power amplifier input by means of a circulator and attenuator. This is necessary to maintain the amplitude and phase response of the system over the 340 MHz bandwidth of the preamplifier, as the match of the gyroklystron power amplifier input will vary considerably over this band. In addition, the circulator-plus-attenuator must have a high (50 kW peak, 10 kW average) rating for reflected power in order to survive spurious emission from the gyroklystron input port which occurs when the gyroklystron is first installed and tuned up in the transmitter. The input of the preamplifier (TWTA) will be driven by an exciter which is required to generate the 34 GHz frequency and ±10 MHz phase modulation bandwidth for the radar. This exciter may have a switched output of 44 MHz which would then drive a multiplier (x 772) to obtain 34 GHz. The inexorable consequence of this frequency multiplication is the multiplication of phase modulation by $(772)^2$ (Ref. 21). This implies that the degradation by the multiplier has to be kept near the theoretical minimum. The specifications of such a multiplier will be very stringent, and none of these multipliers have been built yet.

Another area of crucial interest is the guiding magnet. This device is a solenoid which surrounds the interaction volume and keeps the electron beam focused in the tube length before the collector. A control of better than 1% must be exercised to maintain high efficiency (as shown in

Fig. 7) and, typically, a range of 600-700 A and 30-60 V would be required to maintain a 12.5 kG field if a water-cooled copper solenoid were to be used. At present, it has not been determined whether the solenoid should be a conventional magnet composed of a copper conductor with a hollow water channel for cooling, or a superconducting magnet composed of Nb_3S_n embedded in a copper matrix. Since power dissipation will be an important parameter, if the superconducting solenoid does not give a factor of 1000 to 1 lower power dissipation than that achievable at room temperature, its advantage over the conventional copper magnet is doubtful. Another factor to be considered is that for every watt of transfer in the cooled superconducting structure, about 400 watts of power must be supplied at the refrigeration system input (Ref. 22). The third factor to consider are the constraints on the design of the superconducting solenoid due to the fact that the magnet (and the gyroklystron) will have to tilt through 75° from zenith and rotate through 360° azimuth when tilted. However, tiltable superconducting designs have already been studied for other gyro device applications (Ref. 23). The main advantage of a superconducting magnet, besides its being small in size and weight, is that it provides a virtually ripple-free magnetic field profile in the gun, wiggler, and circuit. This ripple-free profile is essential in order to keep PM distortion down, as can be seen from Table 6.

The gyroklystron output, at 400 kW CW, will be fed via a waveguide arc detector, a forward and reverse power overmoded waveguide coupler, and a mode filter to an overmoded feed system for final antenna illumination. This overmoded transmission line with microwave components and an overmoded feed system is discussed later.

The gyroklystron body, collector, filaments, waveguide components, etc., will be cooled by distilled and deionized water, which in turn will be cooled by ambient air in an external existing heat exchanger. The details of the gyroklystron beam supply are described in the following paragraph.

C. Beam Supply

A block diagram of the existing beam supply is shown in Fig. 13. Power at 12,600 V, 3 phase, and 60 cycles per second is supplied to separate substations from a commercial line which is underground for the last mile. The 2400 V substation supplies the main motor generator only, while all auxiliaries are supplied from a 480 V substation. The output of the main motor generator at 400 Hz is stepped up in voltage in the transformer, rectified, and delivered to the load through a filter, crowbar, and series-limiter resistor at voltages adjustable up to 90 kV and 1.1 MW maximum. The output ripple under full load is less than 0.05%.

The use of a frequency converter (such as the motor generator) might seem unnecessary but actually provides worthwhile technical and economic advantages. It isolates the power line from a crowbar of the dc supply and greatly simplifies line protection problems. It also isolates the supply from short duration line voltage fluctuations and transients due to the large inertia of its rotating components. The change from 60 to 400 Hz reduces all transformer and filter sizes and costs.

The beam supply is required to provide 80 kV between the gyroklystron collector and cathode at a beam current of 12.5 amps during the long radar pulse (up to 4 hours).

Referring to Eq. (1) and Table 6, the intra-pulse ripple on the supply must be kept below 0.18% peak (144 volts peak for typical 80 KV for the gyroklystron) for pulse compression time sidelobes of -40 dB.

The ability of the beam supply to remain ripple-free during the long pulse depends upon the quality of the storage capacitor and wiring inductance between the tube and supply. It is also desirable to keep the storage capacitor as small as possible so as to limit the energy available to discharge in the tube during an arc.

The supply must be capable of withstanding the stress imposed on it when an arc occurs in the gyroklystron. The resultant firing of the crowbar will produce a peak current of 20,000 amps and a peak power of 1600 MW.

D. Monitor and Control

A transmitter monitor and control group (which will be comprised of the power amplifier monitor and control assembly in the antenna, transmitter control cabinet on the ground, and remote radar control in the operations room) will contain the control facilities and indicators necessary for an automated interface with the transmitter. This group will monitor the transmitter circuits and signals to determine the operational status. There will be some 20 major interlocks associated with the gyroklystron output stage alone and the status of each will be indicated. A memory circuit will be incorporated to "hold" an indication of intermittent fault to assist in fault diagnostics. Built in test equipment (BITE) and fault isolation test points will be provided for ease of maintenance and serviceability.

Faults demanding immediate protection such as a gyroklystron arc will result in the firing of the crowbar and the power supply will be shut down. The control unit will be programmed to run the system up again after a short delay, the length of which will depend upon the gas pressure within the gyroklystron envelope. This will be monitored by an ion pump. If the fault persists, the transmitter will remain in the

run down condition with appropriate faults indicated. A fault such as loss of coolant will cause the beam supply to be turned off, while a fault such as a waveguide arc will cause the drive and beam to be removed.

These monitoring and control assemblies will be shielded against both magnetic and electrostatic fields so that transients associated with the high power transmitter operation do not interfere with interlock logic circuit functions, and noise immunity will be achieved through the use of high threshold logic. The interlock functions will be connected via balanced lines and optical isolators as appropriate, to ensure reliable operation even in the presence of the large discharge currents associated with crowbar firing.

The supervision of this monitor and control group will be microprocessor based, which will allow the monitoring and displaying of a wide range of parameters, together with automatic run up and run down sequencing during normal operation and under fault conditions. The automatic shut down system will continuously monitor critical parameters and will take executive action under fault conditions that could lead to gyroklystron damage. A precision multiplexed analog–digital converter will monitor a large number of analog system parameters, all of which will be displayed on front panel meters. A serial interface (RS 232) will be provided to permit full monitoring and control of the transmitter including data logging via a remote radar control terminal in the operations room.

IV. Transmission Line System (Microwave Components)

In this section a description of the components comprising the transmission system will be given. A preliminary layout for an overmoded 400 kW CW transmission system is shown in Fig. 14. The components in the transmission system consist of signal monitoring devices, signal filtering devices and a circular waveguide taper. The monitoring devices include a waveguide arc detector, forward and reverse mode selective directional couplers, polarization monitoring and harmonic content monitoring devices. The only filtering device is the TE_{11} mode filter which serves three purposes: filtering unwanted spurious modes, ensuring the circularity of the TE_{11} mode, and filtering high harmonics.

The first component in the system immediately following the gyroklystron output window is a waveguide arc detector. This device allows the beam power and drive power to be disconnected should an arc be detected in the output waveguide, thus preventing permanent damage to the tube. Circular waveguide arc detectors of 2.5 inch diameter are designed routinely for use with 28 GHz and 60 GHz 200 kW CW gyrotrons.

Although the preferred diameter for the double disc window at the gyroklystron is 2.5", there are great advantages in using a smaller waveguide diameter for the remainder of the transmission system. Aside from the fact that fewer spurious modes can propagate at the smaller diameter, the overall lengths of the directional couplers and mode filter are strongly dependent on the waveguide diameter. Mode selective directional couplers are generally of the phase velocity type, obtaining mode discrimination from the fact that different modes propagate with different phase velocities. These differences in phase velocity increase with decreasing waveguide diameter. Similarly, the mode filter's operation is based upon the fact that different modes require different surface currents on the waveguide wall. These differences also increase with decreasing waveguide diameter.

For the reasons mentioned above, a smaller diameter of 1.75" is being considered for the transmission system. This diameter still allows for relatively low loss transmission and will handle the high CW power, but reduces the overall directional coupler and mode filter length. The allowable length for the gyroklystron and transmission system is determined by the height of the feed cone, which is approximately 16 feet. Additional length may be available, depending on the antenna feed length, since the phase center of the feed is located three feet above the top of the feed cone. The dimension of 1.75" is also preferred since this diameter is the approximate aperture diameter required for most efficiently illuminating the Cassegrain antenna. At present all of the JPL feeds have identical radiation patterns, and the scaled aperture diameter for 34 GHz is 1.75". Therefore with a 1.75" transmission system no flared horn is required. The feed will be discussed in more detail in a later section.

A circular waveguide taper will therefore be included as the next component in the transmission system, tapering down from the 2.5" window diameter to the preferred diameter of 1.75". A nonlinear taper, of the type described earlier, will be optimized for minimum length while the total spurious mode level will be kept below −15 dBc. The spurious mode most strongly coupled to the TE_{11} mode is the TM_{11} mode, which is, unfortunately, one of the most difficult modes to filter out.

Directional couplers are required for monitoring the forward and reflected TE_{11} wave. The couplers must not only be capable of distinguishing between forward and reverse traveling TE_{11} waves, but they must also be able to distinguish between the TE_{11} mode and each of the spurious modes. That is, the couplers must be both directive and mode selective. Mode selective directional couplers can be designed using the coupled transmission line analysis of Miller (Ref. 24). The physical layout of a dual mode selective directional coupler is

shown in Fig. 15. Here the overmoded circular waveguide is coupled to a rectangular waveguide operating in the dominant mode.

If the phase velocities of the desired TE_{11} mode in the circular waveguide and the dominant mode in the rectangular coupling guide are chosen to be equal, then the forward mode discrimination is defined as the forward coupled power for the desired mode divided by the forward coupled power for the spurious mode. For the n^{th} spurious mode, this ratio is given by

$$\text{Discrimination} = 20 \log \frac{C_o \int_{-L/2}^{L/2} \Phi(x)\, dx.}{C_n \int_{-L/2}^{L/2} \Phi(x)\, e^{i\Delta\beta x}\, dx} \quad (4)$$

where L is the overall coupling length,

C_o is the coupling coefficient for the TE_{11} mode,

C_n is the coupling coefficient for the n^{th} spurious mode,

$\Phi(x)$ is the coupling function,

$\Delta\beta$ equals $\beta_o - \beta_n$,

β_o is the phase constant of the desired mode, and

β_n is the phase constant of the n^{th} spurious mode.

Discrimination between forward and reverse traveling waves may be evaluated using an equivalent formula by appropriately redefining the quantity $\Delta\beta$.

Recently, mode selective couplers of the TE_{on} type have been designed and built by Felch et al. (Ref. 25), and Janzen and Stickel (Ref. 26). A coupler design using uniformly spaced round coupling holes with an axially tapered coupling profile capable of providing a coupling factor of -60 dB in the forward direction, -40 dB in the reverse direction, and 40 dB directivity will be investigated. Mode discrimination between the TE_{11} mode and each of the TE_{1n} and TM_{1n} modes will be greater than 40 dB.

Polarization monitoring devices may be included before and after the mode filter, providing a measurement of the ellipticity of the TE_{11} mode in the circular waveguide. The preferred configuration for the polarization monitor is a section of straight waveguide with two coupling pinholes located 90° apart about the waveguide center line. A combining bridge consisting of a hybrid ⁻, phase shifter, attenuator, and crystal detector in a standard WR-28 waveguide will then be capable of determining the ellipticity of the output signal.

The final component of the transmission system to be considered is the TE_{11} mode filter. The present estimate is that the worst case total spurious mode level at the output of the gyroklystron (including realistic effects of slight misalignments of the tube and component assemblies, insulating gaps in the tube output waveguide, etc., over the range of operating parameters) will be approximately -15 dBc. The exact distribution of the spurious mode power must be determined by actual measurements on the gyroklystron itself; however, the calculations from the paper design indicate that the primary spurious modes will be the TM_{11}, TE_{12}, and TM_{12} modes. Since the exact phase and amplitude distribution of these spurious modes is unknown, the TE_{11} mode filter will be included in order to provide a well defined signal at the input to the overmoded feed. An attempt will be made to design a mode filter capable of filtering each of the spurious modes to a level of -30 dBc. The mode filter will be of the helically loaded type, as described by Morgan and Young (Ref. 27). A schematic diagram of the TE_{11} mode filter is shown in Fig. 16. In this type of mode filter the smooth waveguide wall is replaced by a conducting helical winding which is surrounded by a lossy dielectric. The pitch of the helix is chosen so that the surface currents of the desired mode, in our case the TE_{11} RCP mode, follow the windings. Spurious modes, including the orthogonal TE_{11} polarization, suffer attenuation when passing through the filter. Elliptically polarized TE_{11} signals, which can be decomposed into TE_{11} RCP and LCP components, will therefore be purified by removing the LCP component and retaining only the desired RCP component. The amount of attenuation for a particular spurious mode depends upon the conductivity of the lossy jacket and the direction of the surface currents for the specific mode. Surface currents for the TE_{11} mode tend to become more longitudinal as the waveguide diameter increases and TM modes have purely longitudinal currents for any waveguide diameter. In order to obtain significant attenuation for TM modes, it is advisable to use as small a guide diameter as possible. Therefore the length of the filter section is critically dependent on the required attenuation for the TM_{11} mode. Tolerances on the helix pitch are determined by allowable attenuation for the TE_{11} RCP mode. The preferred filter configuration consists of a copper helix, bonded to a beryllia cylinder, backed by a water jacket.

Finally, requirements on the tolerances for alignment of the tube, waveguide components, and feed necessary to preserve the spurious mode level will be determined using the formulae reviewed by Quine (Ref. 28).

V. Antenna Feed System

The transmission system then terminates in an over-moded feed which will create a suitable radiation pattern for illuminating the Cassegrain subreflector. As was mentioned earlier, the 1.75" diameter waveguide provides essentially the correct aperture size for optimally illuminating the Cassegrain system; hence a flare angle horn is not necessary. However, equalization of the E and H plane radiation patterns is required for optimum overall antenna efficiency (Ref. 29). The over-moded corrugated feed section is shown in Fig. 17. The incident TE_{11} mode is transformed into the balanced HE_{11} mode along the feed section via a number of corrugations of varying depth. The balanced HE_{11} mode possesses a circularly symmetric radiation pattern with theoretically no cross-polarization. Similar feeds have been developed for use with linearly polarized plasma heating systems (Refs. 30 and 31). Detailed analysis of the corrugated section is performed using the mode matching and scattering matrix approach of James (Ref. 32). The analysis allows the determination of the number of corrugations and the required depth profile to give the optimum $TE_{11} - HE_{11}$ conversion. The analysis also predicts the resultant feed radiation pattern for an arbitrary set of input modes, i.e., TE_{11} plus any remaining spurious modes. This resultant feed pattern is then used to predict the overall antenna pattern sensitivity, in terms of gain, spillover, and cross polarization, with respect to spurious inputs to the feed. Corrugated feeds typically operate over a much wider bandwidth than that required for this application; therefore such a corrugated section should be capable of producing a suitable radiation pattern over the modest bandwidth of 0.1%.

VI. Discussions and Conclusion

In assessing the state of the art, the development of a 400 kW CW 34 GHz gyroklystron including the overmoded transmission system is subject to technical risks in several areas:

(1) Maintaining RF stability (preventing oscillation). The most challenging technical task is to configure and test an rf circuit which provides adequate gain and efficiency without allowing rf instabilities to occur in any portion of the rf structure. The preferred circuit design for preventing oscillations will include TE_{11} buncher cavities, rf cavity loss, mechanical tunability for the bunchers, cutoff drift tubes and an axial magnetic field profile which can be tailored to help prevent oscillations. Back up circuit design includes a TE_{01} mode buncher section.

(2) Achieving acceptable beam quality. The second most difficult problem is generating a one megawatt beam with adequate beam quality ($\Delta V_{11}/V_{11} \leqslant 5\%$). The beam optics configuration comprising the Pierce gun/wiggler is capable, on paper, of meeting the requirement and has other advantages as well, including space charge limited (low noise) operation. However, a magnetic injection gun (MIG), which has allowed oscillators to achieve efficiencies of over 50%, will be considered as a back-up.

(3) CW window with 400 kW rating. Double disc window technology demonstrated at 28 GHz has already come close to handling the 400 kW CW power level required for the JPL device. Nevertheless a back up approach of a "double-dish" window will also be considered. In this configuration the discs are dish-shaped and are arranged with their convex surfaces in contact with the fluorocarbon cooling channel. The dish shape allows much higher coolant pressure and coolant flow velocities and would therefore be capable of higher CW power levels.

(4) Mode filter. Most of the potential problems associated with the transmission system involve the mode filter and model purity requirement. Resistive wall filters which pass the TE_{on} modes are presently used with gyrotrons. However, such filters do not include a helical winding. Two problems associated with the TE_{11} filter and specifically its helical winding are a very stringent requirement for the tolerance of the pitch of the helical winding (in order to avoid significant attenuation of the TE_{11} mode), and the possibility of breakdown near the windings when transmitting the 400 kW CW power. Also, the modal purity requirement may dictate an excessive length for the mode filter, and a compromise between modal purity and filter length may need to be made.

(5) Antenna feed. With regard to the overmoded feed, corrugated sections have been used to obtain the more desirable HE_{11} radiation pattern from the TE_{11} mode in plasma heating experiments. However, the primary difference between the plasma heating application and the JPL application is the more stringent requirement for the feed radiation pattern which translates into stringent requirements for the modal purity. Due to the large waveguide diameter, an acceptable return loss should be obtainable over the very modest 0.1% bandwidth. Two problems dealing with the feed's response to spurious input modes are intimately related to the problem of the allowable spurious mode level and corresponding mode filter length. Spurious inputs will have an effect on the circular symmetry of the feed radiation pattern, and hence may cause a reduction in the overall efficiency of the Cassegrain

system. Secondly, spurious modes may increase the level of cross polarized radiation in the feed pattern and consequently in the overall antenna pattern.

(6) Power splitter. Finally, should the antenna structure upgrade be insufficient for high efficiency 34 GHz operation due to gravity deformation, a microwave solution using an array of properly phased feeds may be needed to compensate for these effects. The primary difficulty in the design of such a system would be developing a suitable power splitting device. One possible candidate would be a multiple arm coupler designed using tight coupling theory (Ref. 24). Due to significant losses in the dominant waveguide, these runs would need to be made as short as possible.

These then are the technical risks and developments needed in several areas of the K_a-band transmitter. A conceptual design for a 400 kW CW 34 GHz transmitter including overmoded microwave plumbing and an overmoded feed system has been presented. Upon completion of the future final paper design, hardware and implementation stages of the project, the K_a-band transmitter should prove to be a valuable instrument for planetary radar and also serve as a proving ground for new technology which will be transferable to future spacecraft uplinks.

References

1. M. S. Reid et al., "Low-Noise Microwave Receiving Systems in a Worldwide Network of Large Antennas," Proc. IEEE, Vol. 61, pp. 1330-1335, September 1973.
2. S. A. Hovanessian, *Radar Detection and Tracking Systems*, Artech House, Dedham, Massachusetts, Chapter 11, Section 5, pp. 11-22, 1973.
3. M. R. Barnett et al., "An Inverter Powered One Megawatt Power Amplifier for a Radar Transmitter," Proceedings from International Conference on RADAR, RADAR 77, London, England, pp. 349-363, October 1977.
4. D. D. P. King, "Millimeter-Wave Prospectus," Microwave Journal, Vol. 10, pp. 24-29, November 1967.
5. J. L. Hirshfield, "Gyrotron," in *Infrared and Millimeter Waves*, Vol. 1, Academic Press, New York, pp. 1-54, 1979.
6. V. L. Granatstein and S. Y. Park, "Survey of Recent Gyrotron Developments," Proceedings of the International Electronics Devices Meeting (IEDM), Washington, D.C., pp. 263-266, December 1983.
7. R. Symons, H. Jory, S. Heggi, and P. Ferguson, "An Experimental Gyro TWT," IEEE Trans. Microwave Theory Tech., Vol. MTT-29, pp. 181-184, March 1981.
8. P. Ferguson, G. Valier, and R. Symons, "Gyrotron-TWT Operating Characteristics," IEEE Trans. Microwave Theory Tech., Vol. MTT-29, pp. 794-799, August 1981.
9. L. A. Barnett et al., "An Experimental Wide-Band Gyrotron Traveling-Wave Amplifier," IEEE Trans. Electron Device, ED-28, pp. 872-875.
10. A. Arfin and A. K. Ganguly, "A Three-Cavity Gyroklystron Amplifier Experiment," Int. J. Elect., Vol. 53, No. 6, 709-714, 1982.
11. S. Hegji, H. Jory, and J. Shively, "Development Program for a 200 kW CW 28 GHz Gyroklystron," Varian Associates, Inc., Internal Document.
12. D. Stone, M. Chodorow, and A. Nordquist, "Fast Wave Amplifiers for Space Communications Phase II: Gyroklystron Design Study," Final Report, prepared by Varian Associates, Inc., for the NASA Lewis Research Center, April-September 1982.

13. M. Caplan, "Gain Characteristics of Stagger Tuned Multicavity Gyroklystron Amplifiers," IEEE International Conference on Infrared and Millimeter Waves, Miami Beach, Florida, Session W4.2, December 1983.

14. K. Felch, R. Bier, M. Caplan, and H. Jory, "100 GHz, 1 MW CW Gyrotron Study Program," Final Report, prepared by Varian Associates, Inc., under subcontract P.O. #53Y-21453C, for the Oak Ridge National Laboratory, Nashville, Tennessee, operated by Union Carbide Corp. for the U.S. Department of Energy, under contract #W-7405-eng-26.

15. C. Moeller, "Mode Converters Used in Doublet III ECH Microwave System," Int. J. Elect., Vol. 53, No. 6, pp. 573-585, 1982.

16. F. Sporleder and H. G. Unger, *Waveguide Tapers, Transitions and Couplers*, Institute of Electrical Engineers, London, England, Chapter 2, 1979.

17. H. Jory et al., "First 200 kW CW Operation of a 60 GHz Gyrotron," IEDM Technical Digest, Washington, D.C., pp. 267-270, December 1983.

18. K. R. Spangeberg, *Fundamentals of Electron Devices*, McGraw Hill, New York, pp. 427-428 and 464-465, 1957.

19. K. H. Sann, "The Measurement of Near-Carrier Noise in Microwave Amplifiers," IEEE Trans. Microwave Theory Tech., Vol. MTT-16, No. 9, pp. 761-766, September 1968.

20. H. Hashimoto et al., "A 30 GHz 40 Watt Helix Traveling-Wave Tube," Proc. IEDM, Washington, D.C., pp. 133-136, December 1983.

21. J. J. G. McCue and E. A. Crocker, "A Millimeter-Wave Lunar Radar," Microwave Journal, pp. 59-63, November 1968.

22. M. H. Cohen, *Superconductivity in Science and Technology*, Univ. of Chicago Press, Chicago, Illinois, p. 103, 1968.

23. H. T. Coffey, K. R. Efferson, and G. J. Svenconis, "A Study of the Requirements for Shipboard Use of Superconducting Magnet Systems Used for Gyrotrons," Naval Research Laboratory, Washington, D.C., Contract No. N00014-81-C-2586, American Magnetics, Inc., June 1982.

24. S. E. Miller, "Coupled Wave Theory and Waveguide Applications," Bell System Technical Journal, Vol. 33, pp. 661-719, 1954.

25. K. Felch, R. Bier, L. Fox, H. Huey, H. Jory, J. Manca, J. Shively, S. Spang, and C. Moeller, "Analysis of the Output Mode from 60 GHz, 200 kW Pulsed and CW Gyrotrons," IEEE International Conference on Infrared and Millimeter Waves, Miami Beach, Florida, Session T4.2, December 1983.

26. G. Janzen and H. Stickel, "Mode Selective Directional Couplers in Overmoded Circular Waveguides," IEEE International Conference on Infrared and Millimeter Waves, Miami Beach, Florida, Session TH4.6, December 1983.

27. S. P. Morgan and J. A. Young, "Helix Waveguide," Bell System Technical Journal, pp. 1347-1384, November 1956.

28. J. P. Quine, "Oversize Tubular Metallic Waveguides," in *Microwave Power Engineering*, Academic Press, New York, pp. 178-212, 1968.

29. P. D. Potter and A. Ludwig, "Antennas for Space Communications," JPL Space Programs Summary No. 37-26, Vol. IV, pp. 200-208, April 30, 1964.

30. M. Thumm, G. Janzen, G. Muller, P. G. Schuller, and R. Wilhelm, "Conversion of Gyrotron TE_{on}-Mode Mixtures into a Linearly Polarized HE_{11} Wave," IEEE International Conference on Infrared and Millimeter Waves, Miami Beach, Florida, Session TH4.7, December 1983.

31. J. L. Doane, "Mode Converters for Generating the HE_{11} (Gaussian-like) Mode from TE_{01} in Circular Waveguide," Int. J. Elect., Vol. 53, pp. 573-585, December 1982.

32. G. L. James, "Analysis and Design of TE_{11} to HE_{11}, Corrugated Cylindrical Waveguide Mode Converters," IEEE Trans. Microwave Theory Tech., Vol. MTT-29, pp. 1059-1066, October 1981.

Table 1. K_a-band transmitter specifications

Frequency	34 GHz
Bandwidth	0.1%
RF power output	400 kW CW (+ 86 dBm)
RF stability	0.1 dB over 1 transmit cycle
Incidental AM	60 dB below carrier at all modulating frequencies above 1 Hz
Phase stability	1.0×10^{-15} (1000 seconds) with goal of 1.0×10^{-16} (1000 sec)
Incidental PM (jitter)	< 1° peak to peak
Noise figure	-80 dB/MHz
Transmit pulse	20 sec minimum to a few hours maximum
Modulation	*Phase Modulation*: 0-100% carrier suppression 1 kHz to 2 MHz. This is accomplished by PN code modulation of length 2^n-1 where n = 6 to 15 and at baud length of 0.5 sec to 1000 sec.
	FSK: Shifting carrier frequencies separated from 1 Hz to 1 MHz and switching in less than 1 msec every 30 seconds.

Table 2. JPL gyroklystron specifications

f_0:	34 GHz
Bandwidth:	0.1% (1 dB points)
Output power:	400 kW CW saturated (goal)
Output mode:	Dominant TE_{11}° circularly polarized mode with very high modal purity (-30 dB for each extraneous mode) and circular polarization within 1 dB
Noise figure:	-80 dB/MHz
Orientation:	Gyroklystron and auxiliary components such as its magnets must be capable of operating through 75° of elevation motion (zenith to 15° above horizon) and simultaneously through 360° in azimuth when installed on the antenna
Efficiency:	40% (goal)
Saturated gain:	50 dB minimum (goal)

Table 3. Gyroklystron circuit design

a. Characteristics

Characteristic	Value
Voltage	80 kV
Current	12.5 amps
Beam radius	0.120 cm
Perpendicular velocity/parallel velocity	1.5
Magnetic field	12.5 kgauss
Number of cavities	4
Total length	7.84 cm
Input coupling Q_{coupl}	299
Output external Q_{ext}	120
Mode buncher cavities	TE_{111}
Mode output cavity	TE_{121}
Small signal gain	57 dB
Saturated gain	50 dB
Saturated efficiency	46.5%
Saturated bandwidth (-1 dB points)	0.3%

b. Cavity configuration

Cavity number	Length (cm)	Radius (cm)	Resonant frequency (GHz)	Cold Q_o	Beam Q_B
1	0.784	0.314	33.9	299	330
2	0.784	0.317	33.7	400	330
3	0.784	0.318	33.6	400	330
4	1.725	0.769	34.2	120	-220

c. Drift tubes

Drift tubes	Length (cm)
1	0.784
2	0.784
3	2.196

Table 4. Spurious mode level at converter output of mode converter TE_{12}–TE_{11} (99% efficient)

Spurious modes produced	Levels produced
TE_{13}	–27 dBc
TE_{14}	–34 dBc
TM_{11}	–38 dBc
TM_{12}	–23 dBc
TM_{13}	–37 dBc
TM_{14}	–35 dBc

Table 5. AM pushing factors

Parameter	dB/%
Cathode voltage	0.5 dB/%
Main magnet coil current	0.5 dB/%
Load VSWR	1.6 dB for VSWR 2:1 (oscillation at VSWR 2.4:1)
RF drive (input power)	0.007 dB/%
Filament voltage	0.005 dB/% (typical)
Wiggler coil current	0.6 dB/% (typical)
Gun coil current	0.5 dB/% (typical)

Table 6. PM pushing factors

Parameter	°/%
Cathod voltage	6.4°/%
Main magnet coil current	100°/%
Load VSWR	TBC[a]
RF drive (input power)	TBC
Filament voltage	TBC
Inlet coolant temperature	TBC
Wiggler coil current	TBC
Gun coil current	TBC

[a]To be calculated

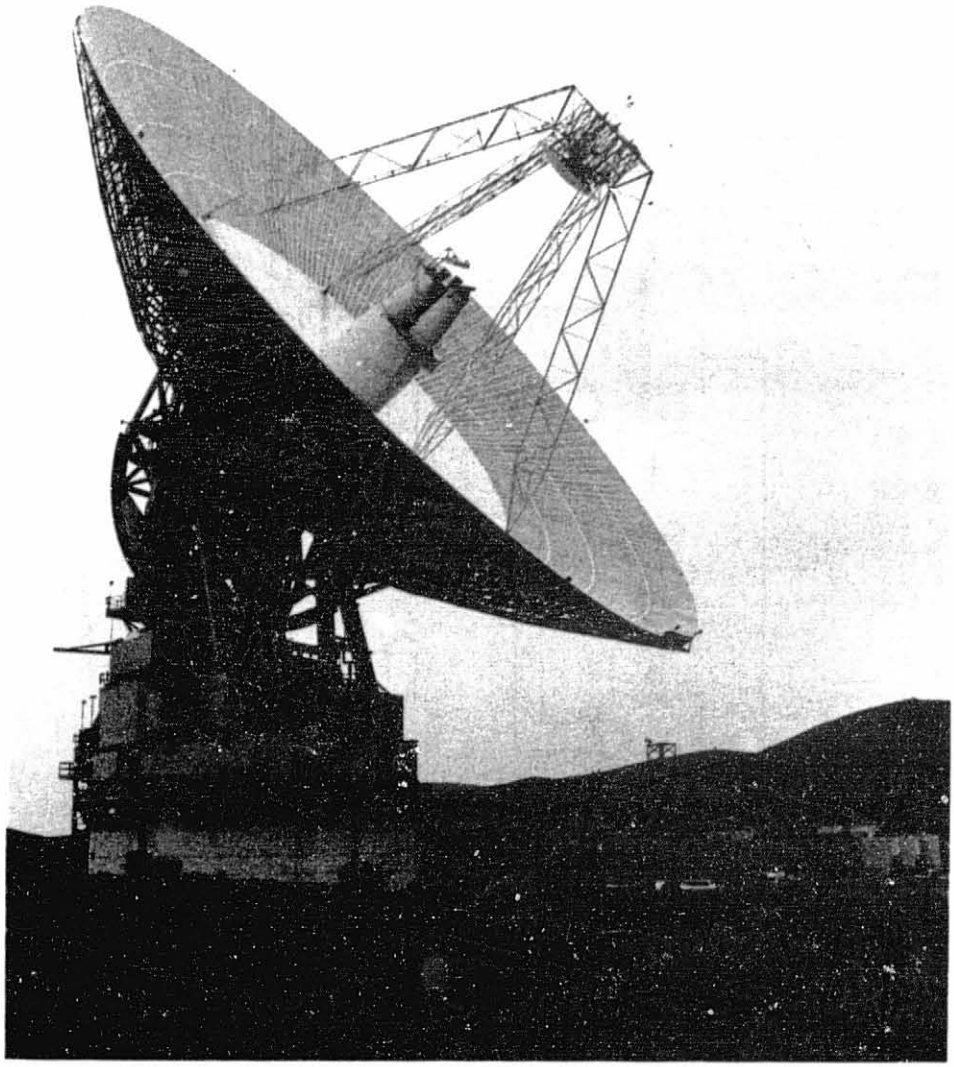


Fig. 1. 64 meter diameter antenna at Goldstone, California

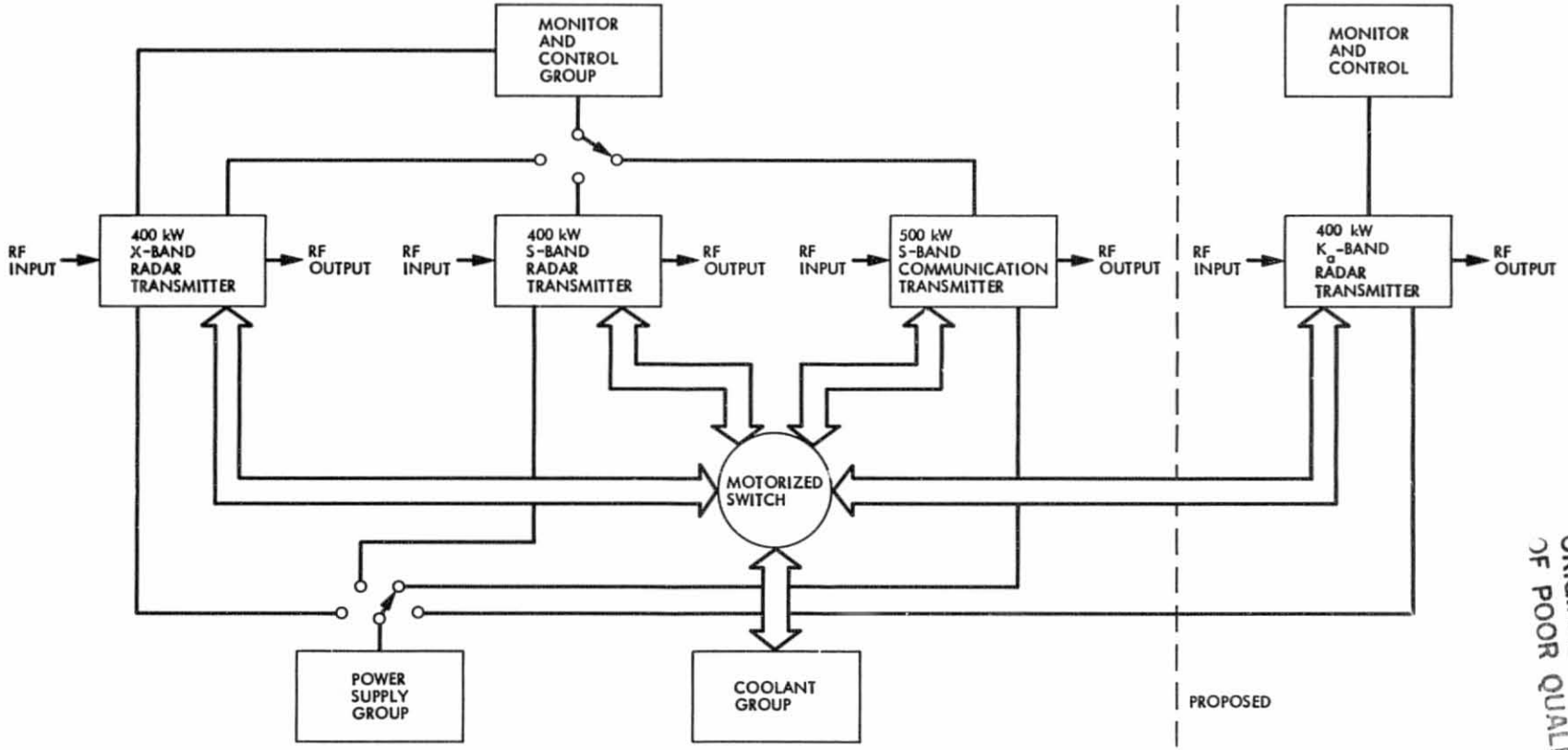


Fig. 2. Transmitters on the 64 meter antenna

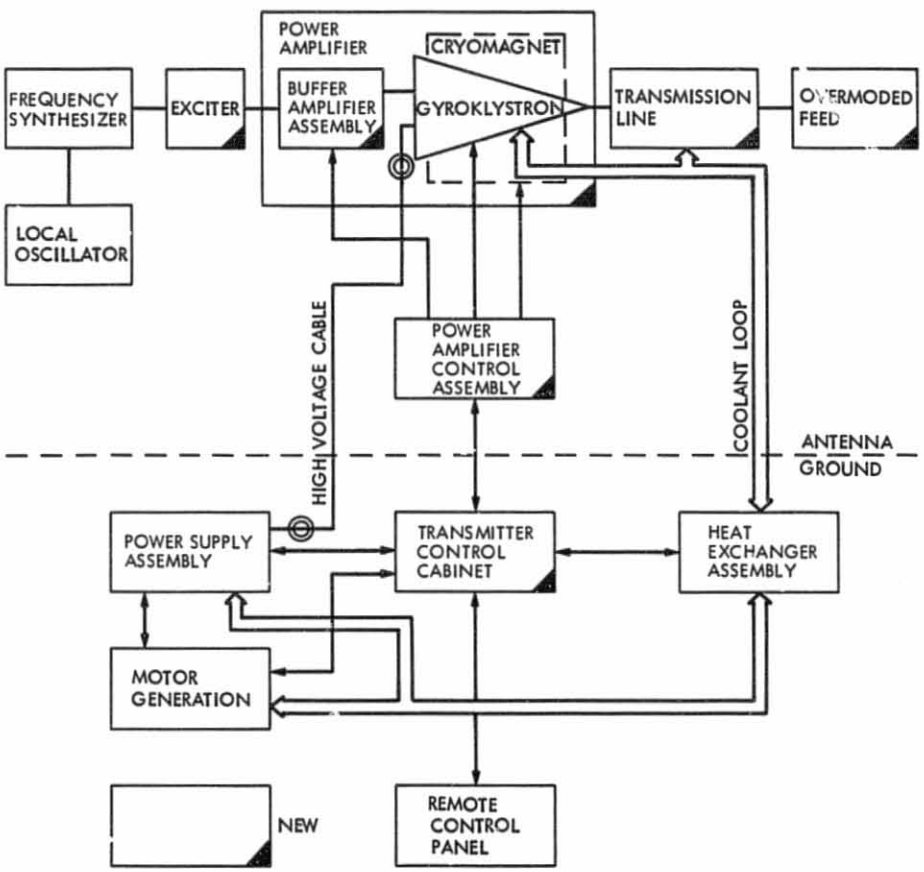


Fig. 3. K_a-band transmitter block diagram

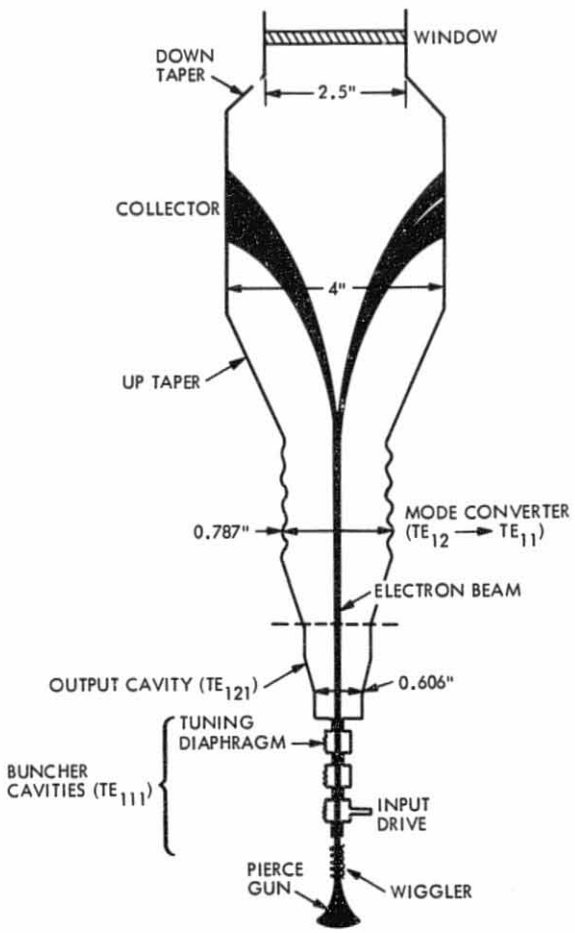


Fig. 4. Proposed 35 GHz gyroklystron schematic diagram

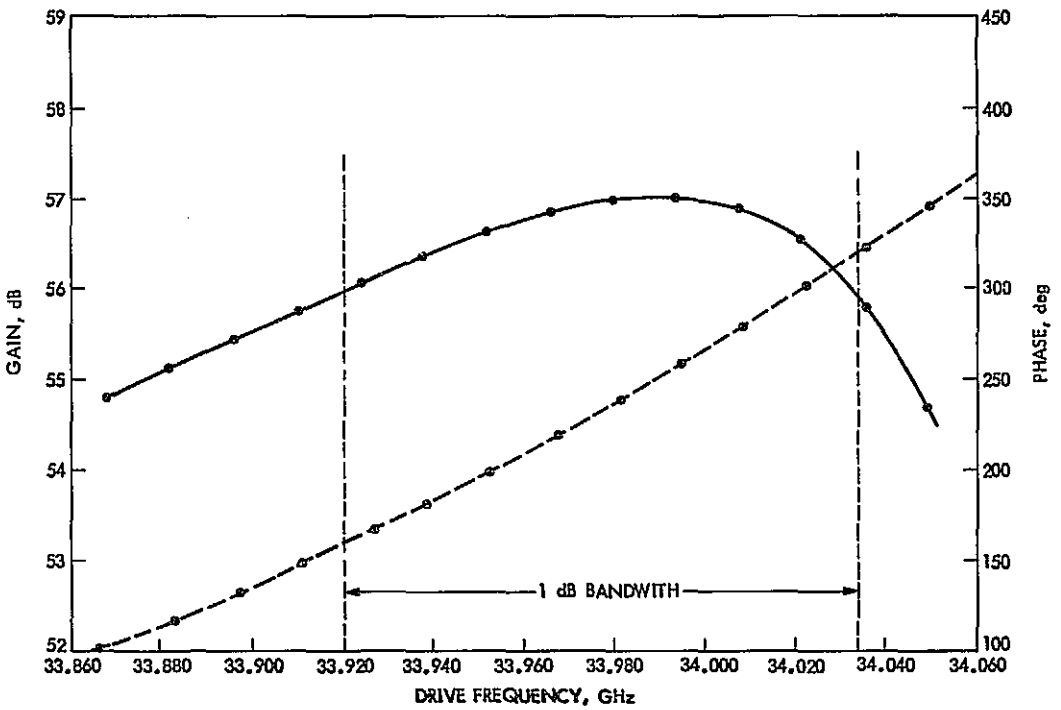


Fig. 5. Linear gain vs. frequency, and phase vs. frequency of the proposed 34 GHz gyroklystron

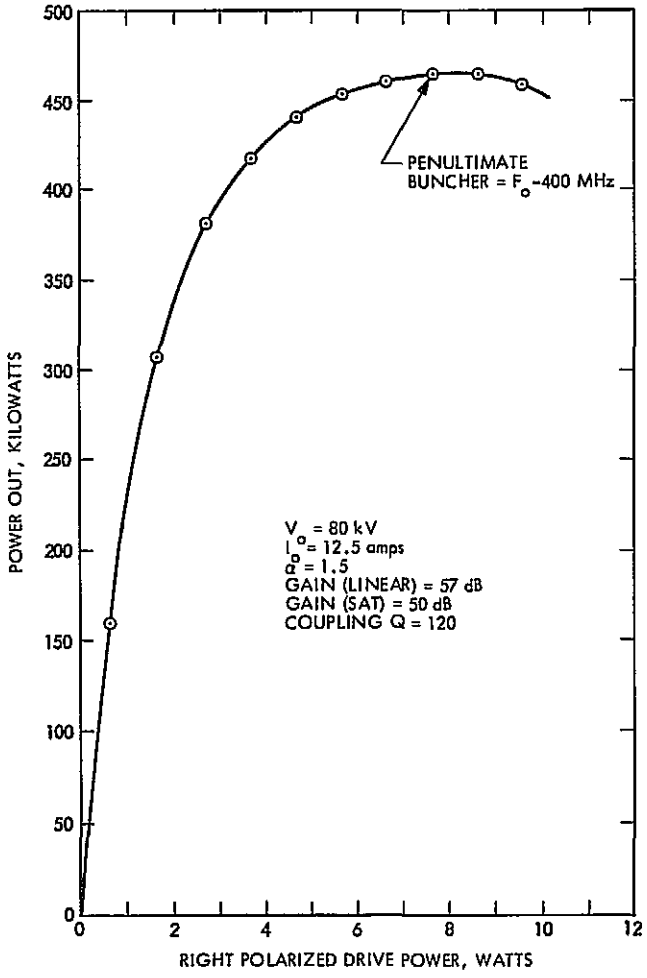


Fig. 6. Power out vs. drive power for the proposed 34 GHz gyroklystron

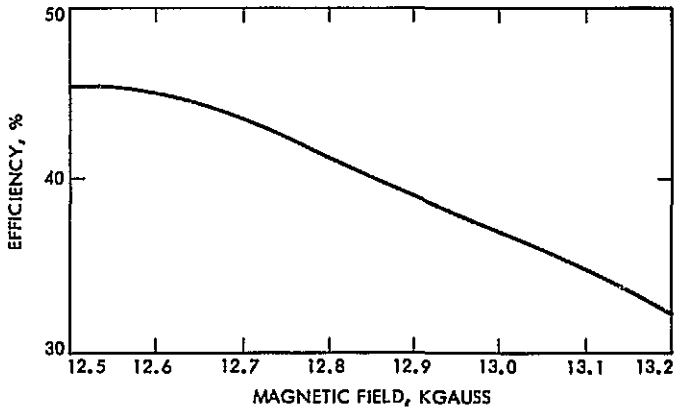


Fig. 7. Efficiency as a function of magnetic field for the 34 GHz gyroklystron

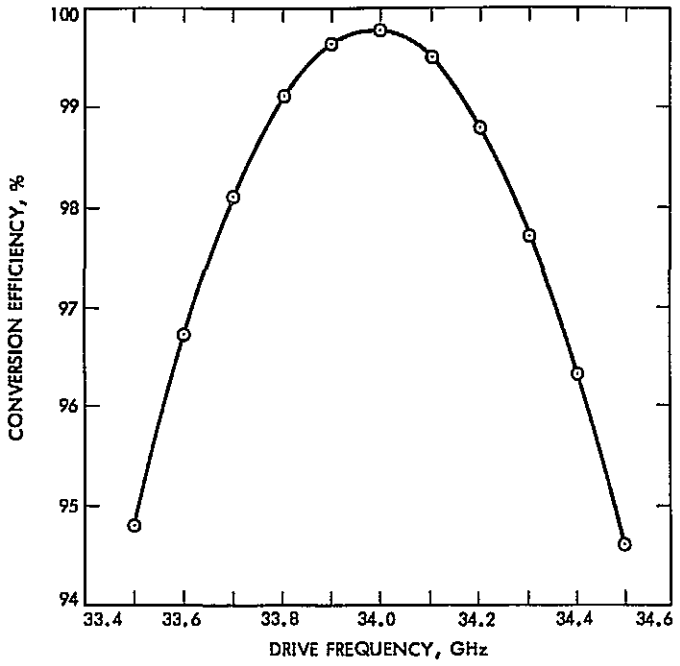


Fig. 8. Calculated conversion efficiency vs. drive frequency of the TE_{12} to TE_{11} converter

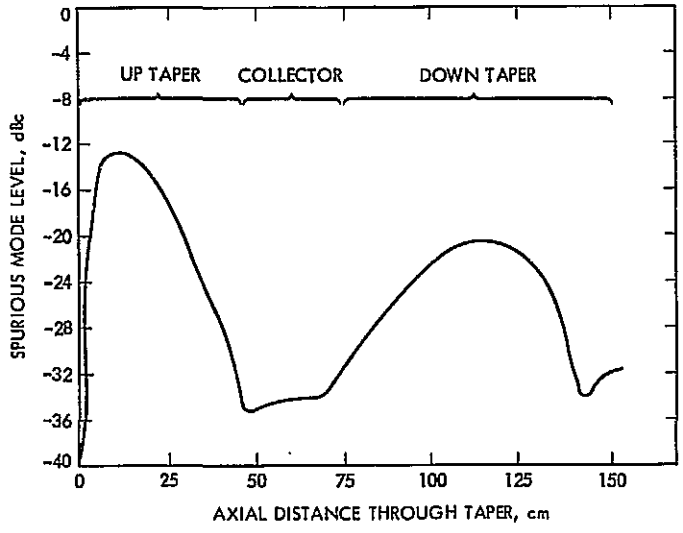


Fig. 9. Spurious TE_{1n} mode level in an optimized TE_{11} compound-tapered collector vs. axial distance. The final spurious mode level at the output is −32 dBc.

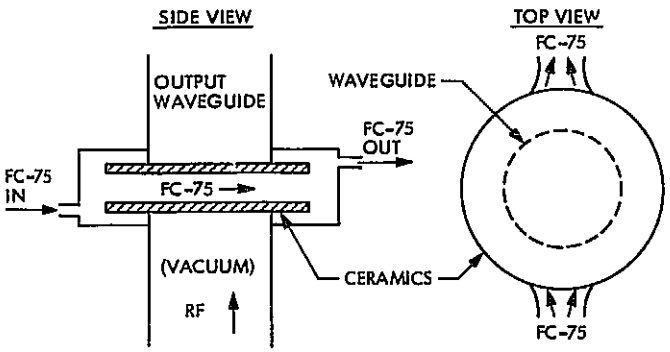


Fig. 10. Schematic diagram of a double disc window

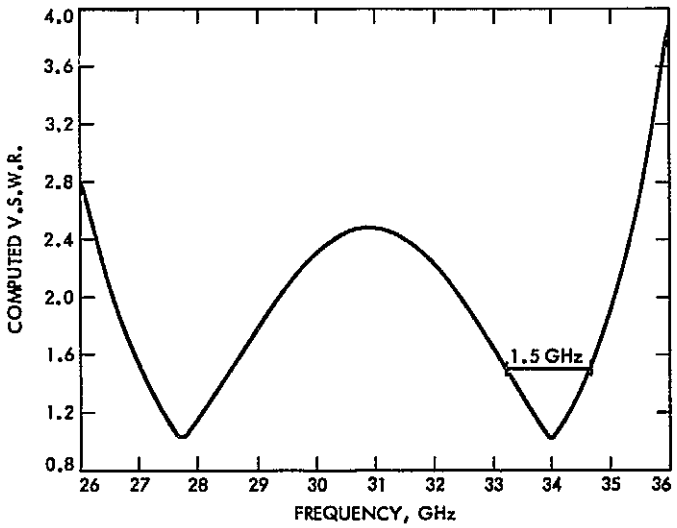


Fig. 11. Computed VSWR for a 34 GHz double disk window

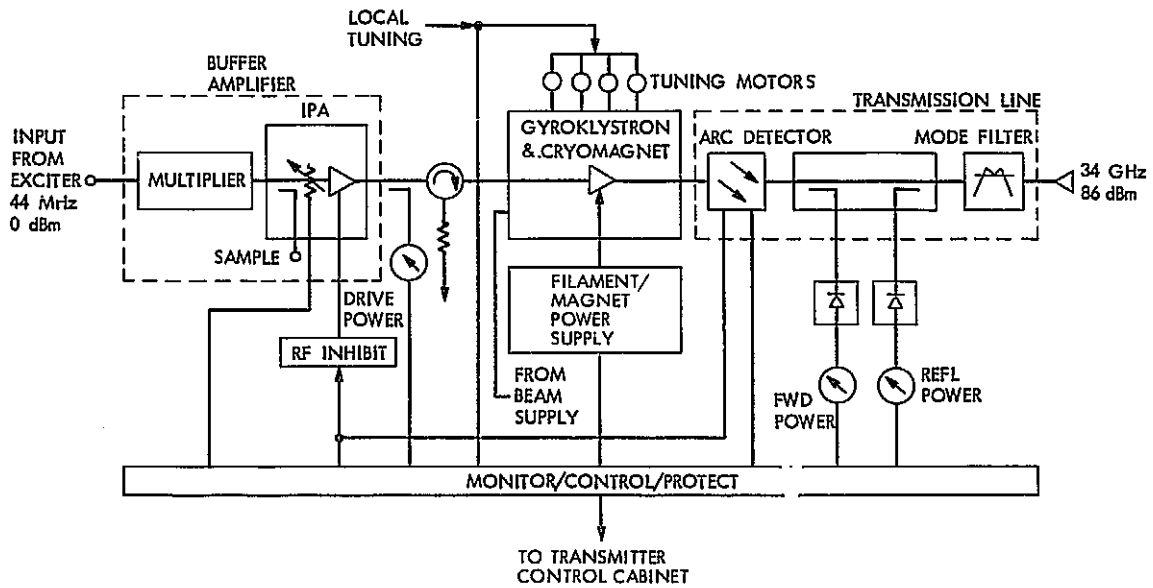


Fig. 12. Power amplifier block diagram

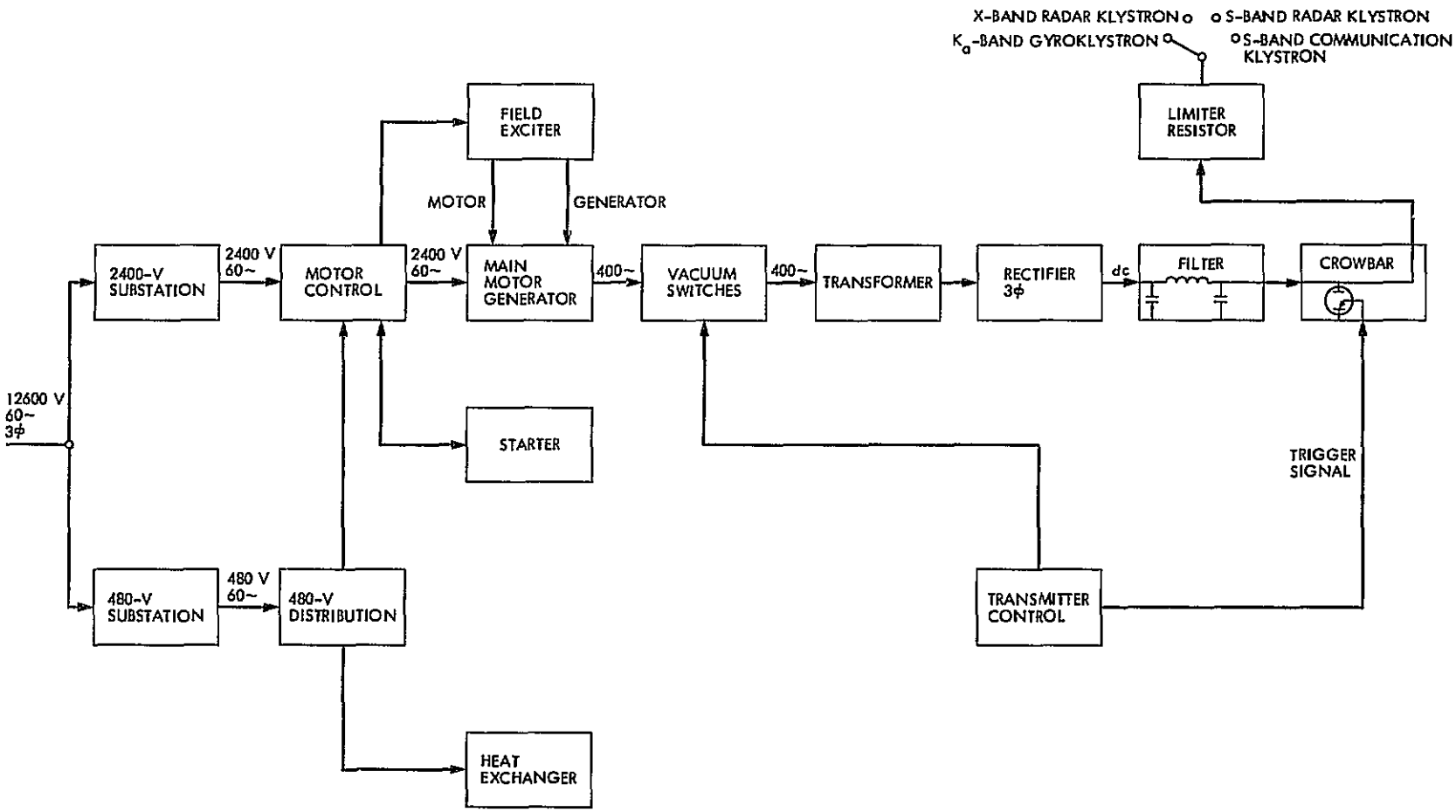


Fig. 13. Power supply block diagram

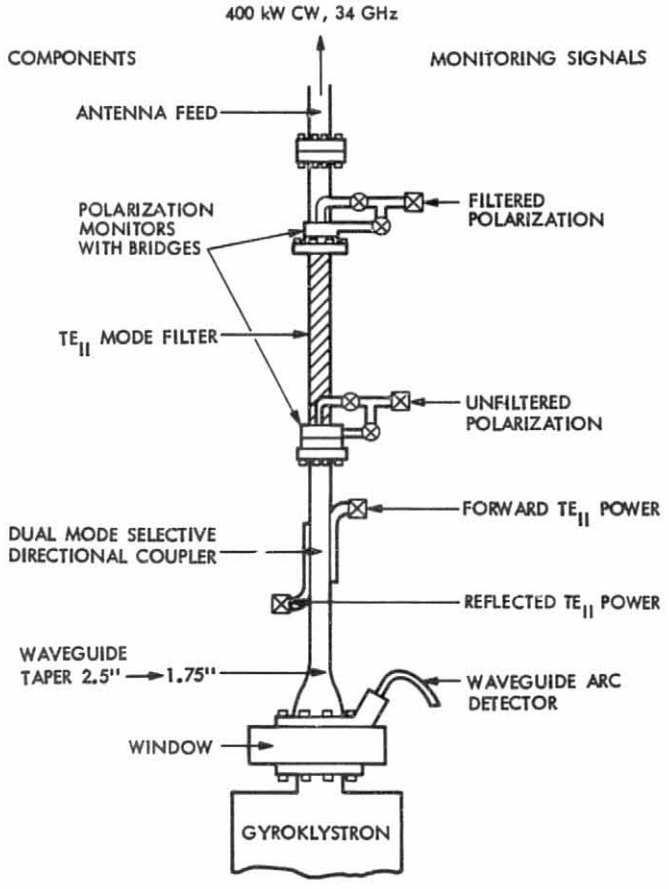


Fig. 14. Proposed 1.75 in. diameter 400 kW CW TE_{11} transmission line system

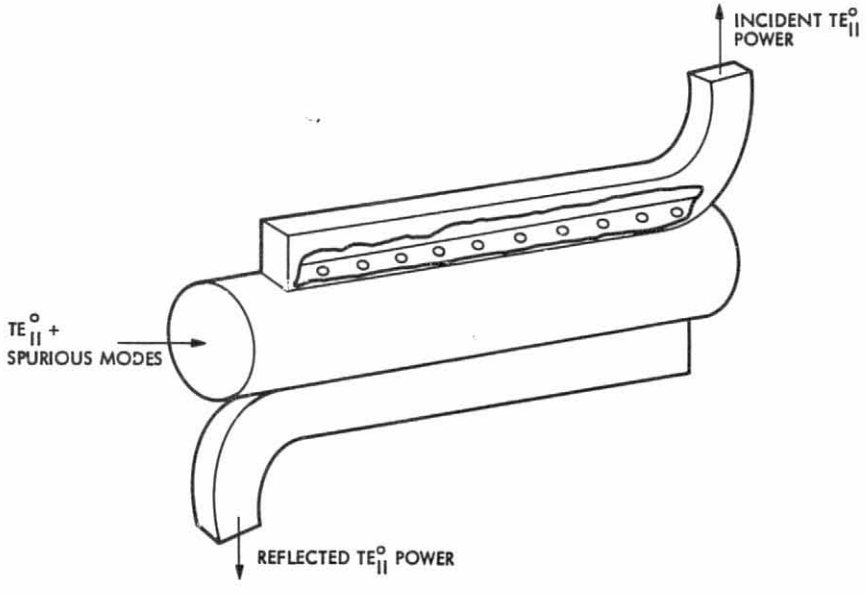


Fig. 15. A dual mode selective directional coupler

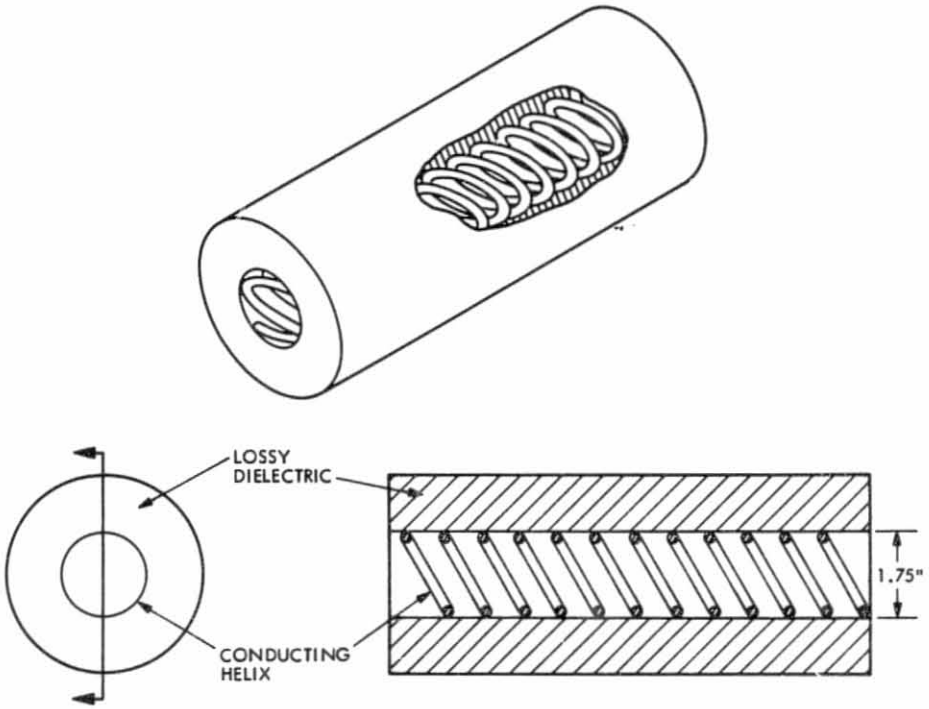
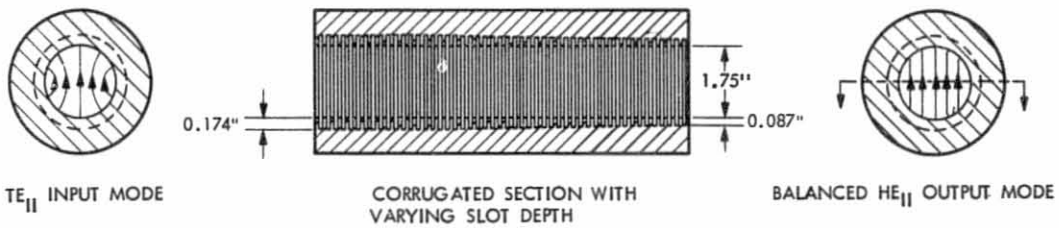


Fig. 16. The TE_{11} mode filter



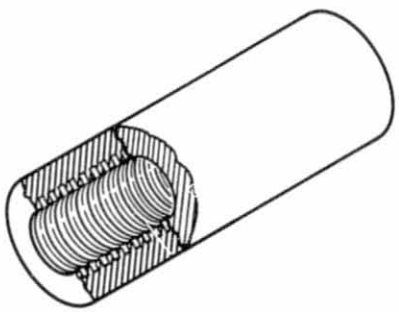


Fig. 17. The antenna feed

Magnetic Refrigeration Development

D. D. Deardorff and D. L. Johnson
Radio Frequency and Microwave Subsystems Section

Magnetic refrigeration is being developed to determine whether it may be used as an alternative to the Joule-Thomson circuit of a closed cycle refrigerator for providing 4 K refrigeration. An engineering model 4–15 K magnetic refirgerator has been designed and is being fabricated. This article describes the overall design of the magnetic refrigerator.

I. Introduction

Much interest has been generated in the past few years to develop adiabatic demagnetization into a reliable and efficient, continuous refrigeration stage for a closed cycle refrigerator. Until recently adiabatic demagnetization, or magnetic cooling, was basically regarded as a research tool, a one-shot device to producing extremely low temperatures for short periods of time. The pioneering work by Heer et al. (Ref. 1) in 1954 resulted in the first magnetic refrigerator to provide continuous low temperature (<1 K) refrigeration needed for physics research. Much of the present developmental work on continuous magnetic refrigerators centers on low temperature devices to cool infrared bolometers to below 0.3 K for spacecraft operation (Refs. 2, 3, 4), or to provide superfluid helium refrigeration (Refs. 5, 6, 7, 8) for enhancing the operation of superconducting devices, such as magnets, energy storage rings and transmission lines. All of these devices use liquid helium (^{3}He, ^{4}He or superfluid ^{4}He) as the high temperature heat reservoir.

Only very recently has the production of 4 K refrigeration using magnetic cooling been addressed. This temperature regime is generally reserved for the passively operating, but inherently inefficient, Joule-Thomson valve. A detailed analysis on a design for a 4–15 K magnetic refrigerator stage to complement a 15 K precooler was first presented in 1966 by Van Geuns (Ref. 9); however, any developmental work which may have followed has never been reported in the open literature. J. A. Barclay of the Los Alamos National Scientific Laboratory is presently pursuing the rotational magnetic wheel concept (private communication). His gadolinium gallium garnet (GGG) wheel is slowly rotated through a high magnetic field to achieve a 4–20 K temperature span. Helium gas is pumped through GGG matrix at the two temperature extremes to provide the heat exchange mechanism. Hashimoto et al. at the Tokyo Institute of Technology have recently presented experimental results obtained from a 4.2-20 K magnetic refrigerator they developed (Ref. 10). They have elected to ramp the magnetic field in order to keep their GGG matrix stationary. A helium thermosiphon extracts heat from the load; helium gas is used to transfer heat to the 20 K heat reservoir. Chinese workers are reportedly (Ref. 11) developing a 4–15 K magnetic stage to mount to a Gifford-McMahon precooler. No additional information about their work is known.

The Jet Propulsion Laboratory (JPL) has been using 1 Watt at 4.5 K closed cycle refrigerators (CCRs) since 1965 for cooling the low-noise maser amplifiers required to receive very weak signals from spacecraft in deep space. Up to 30 CCRs are in near continuous operation in the Deep Space Communications Network (DSN), logging approximately one quarter of a million hours annually. To meet the continuing requirement to increase both the reliability and efficiency of the CCR and to reduce life-cycle costs and achieve future technical objectives,

JPL has initiated the development of a 4-15 K magnetic refrigerator for use with a 15 K expansion engine. This new technology is being pursured to surpass the Joule-Thomson circuit in terms of efficiency, reliability, and achievable temperature span. The decision to develop a reciprocating device stems from the long experience JPL has with the reciprocating Gifford-McMahon expansion engine and its sliding seals, its relatively simpler fabrication requirements, and the greater ease with which the experimental tests results can be verified theoretically.

The concept of magnetic refrigeration was introduced in a recent TDA Progress Report (Ref. 12). That report presented a review of magnetic refrigerator designs which have either been conceptualized or built and tested. It is the objective of this article to describe the design of the engineering model 4-15 K magnetic refrigerator under development at JPL, the component test results and the status of the development effort.

II. Principle of Magnetic Refrigeration

The placement of a paramagnetic material in a magnetic field at low temperatures causes the material to warm up. Conversely, removal of the material from the magnetic field will cause the material to cool. If the paramagnetic material is held in contact with a constant temperature reservoir, the material will tend to expel or absorb heat from the reservoir as the changing magnetic field warms or cools the material beyond the temperature of the reservoir. This is the principle of operation for the magnetic refrigerator illustrated in Fig. 1. In this figure, the magnetic refrigerator operates ideally in a Carnot cycle. Panel 1 of the figure shows the paramagnetic material thermally isolated from the precooler (heat sink) and the load (heat reservoir). As the magnetic field is increased, the temperature of the material is increased. As the material's temperature reaches that of the precooler (T_H), contact is made between the material and the precooler so that the heat of magnetization created in the material during further magnetization is removed to the precooler (Panel 2). The paramagnetic material, now at T_H and in a strong magnetic field, is again isolated (Panel 3). A reduction in the magnetic field lowers the material's temperature until it reaches the temperature (T_C) of the load. Contact is then established with the load and, during further demagnetization, the cooling of the paramagnetic material draws heat from the load (Panel 4). Thermal contact is then broken and the cycle is started over again as in Panel 1. This cyclic operation for the paramagnetic material GGG is illustrated in the entropy-temperature diagram shown in Fig. 2.

III. Experimental Design

Choice of the magnetic refrigerator design must depend ultimately on the device it is to cool, in this case the maser, an ultrasensitive microwave signal amplifier whose performance depends critically on a stable DC magnetic field and a stable, low operating temperature. Operating in the DSN, the maser is located in the feedcone of a large antenna. The antenna may be oriented from zenith to horizon for tracking purposes. The overall cooling system requirements for the maser, listed in Table 1, are therefore quite stringent. The design of the engineering model magnetic refrigerator has addressed only the basic requirements of refrigeration capacity, DC field stability, reliability, and efficiency.

The schematic of the engineering model magnetic refrigerator design is shown in Fig. 3. The major components of the refrigerator include the piston and cylinder assembly for the paramagnetic material, the drive mechanism for the piston, the superconducting magnet, the gas pumps for the low and high temperature gas circuits, and the two stage CTI Model 1020 expansion engine. The CTI Model 1020 expansion engine provides the high temperature heat sink for the magnetic refrigerator and is capable of producing better than 9 W of refrigeration at 15 K. This refrigeration capacity is a major determining factor in the final 4 K cooling power of the magnetic refrigerator. The hydrogen heat switch is used during initial cooldowns to precool the helium dewar and magnet assembly to 20 K before liquid helium is transferred into the dewar. This design presently calls for the external transfer of helium; future designs call for the magnetic refrigerator stage to provide the parasitic refrigeration requirements of the magnet. The magnet assembly (superconducting magnet and Hiperco[1]) provide the large magnetic field needed for the paramagnetic material. The piston contains two chambers filled with porous matrices of the paramagnetic material. These matrices are alternately driven into the magnetic field in a reciprocating motion by the mechanical drive system (gearmotor and a "ball reverser"[2]). Coupling the gearmotor to the ball reverser is a rotary ferrofluidic seal[3] which functions as a vacuum feedthrough to prevent contamination of the helium gas. Gas pumps in the low and high temperature gas flow loops provide the gas flow needed for the heat exchange.

The 7 T magnetic field for the GGG piston is provided by a 10.2 cm NbTi solenoid having a 6.3 cm bore. The magnet is encased with a magnetically soft material, Hiperco, having a

[1] Hiperco is an iron-cobalt alloy available from Carpenter Steel.

[2] Ball Reverser is a trade name of a mechanical actuator patented by Norco, Inc.

[3] A vacuum rotary seal patented by Ferrofluidics Corporation.

maximum permeability of 10,000 and a saturation induction of 2.4 T (see Fig. 4). The Hiperco is used as a low reluctance path to entrap much of the magnetic flux exiting from the bore of the magnet. This provides a rapid transition between the high field and low field regions enabling a shortened stroke length for the GGG piston. Figure 5 compares the measured axial profile of the magnetic field with and without the Hiperco and shows the position of the piston at the end of the stroke. The figure shows the ability of the Hiperco to shape the field, enhancing the field fall-off rate outside the magnet while slowing the fall-off rate of the field inside the magnet, although the latter effect was not as pronounced as expected. Further field shaping can be obtained by varying the shape of the Hiperco material on the ends of the magnet.

The magnet was wound with single strand 0.254/0.406 mm NbTi wire around a copper coil former. The coil was wet-wound with GE 7031 varnish to prevent motion of the individual wires during magnet charging. After winding, the magnet was potted with Stycast 2850GT. The magnet required only a small amount of training to achieve 7 T field; however, with the addition of the Hiperco, the magnet required some retraining to again reach the 7 T field. A persistent switch for the magnet has a resistance of less than 0.2 μohms corresponding to a minimum five year decay time for the magnet. A resistive shunt made from a short length of stainless steel tubing is connected to the magnet coil leads in the 4.2 K bath. The shunt resistance is chosen to protect the coil during quench while slowly dumping the 10 kJ of stored energy into the liquid helium bath.

In the cylinder assembly of the 4–15 K magnetic refrigerator, two chambers containing porous matrices of a paramagnetic material are located in tandem on a single reciprocating piston machined from phenolic (Fig. 6). In this design, each matrix volume is 33 mm long and 38 mm in diameter and is filled to about a 40% porosity with 160 grams of 1.1 mm diameter $Gd_3Ga_5O_{12}$ (GGG) spheres. The use of the two matrices effectively doubles the heat removal capabilities per cycle of the piston and reduces the temperature fluctuations by providing for a more continuous removal of energy from the heat source. The cooling power at 4.2 K for this refrigerator operating ideally in the Carnot cycle can be given as

$$\dot{Q}_C = (T_C/T_H)\,\dot{Q}_H\,\eta = 1.76\,\mathrm{W}$$

where T_C is the refrigeration temperature, T_H is the sink temperature, $\dot{Q}_H$ is the rate of heat rejection, η is the fraction of carnot efficiency at which the magnetic refrigeration stage operates, and where the CTI 1020 limits the heat expelled at 15 K to 9 W. The efficiency (assumed to be 70%) is determined by factors such as the thermal heat leaks along the cylinder and drive shaft walls, the heat exchange between the gas and the matrices, the heat capacity of the gas entrained in the matrices, and gas leakage by the seals, as well as other factors. The factors contributing to the loss of cooling power of the refrigerator will be identified and minimized during refrigerator testing.

The GGG matrices in the piston are separated sufficiently so that at either end of the stroke one matrix is in the high field region while the other matrix is in the low field region. The placement of the GGG matrices on either side of the magnet's center provides force compensation to reduce the overall force exerted on the piston drive shaft to move the piston. The magnetic interaction force that attracts the GGG to the magnet is substantial (an estimated force of 1550 newtons [350 pounds] is required to move one of the 160 gram GGG matrices through the 7 T field produced by this superconducting magnet); thus careful consideration of the separation distance between the matrices is required to greatly reduce the net magnetic force. The basic equation for the magnetic force is

$$\mathbf{F} = (\mathbf{M} \cdot \nabla)\mathbf{B}$$

where M is the field and temperature dependent magnetization of the paramagnetic material, and B is the magnetic field. Thus as a first order guesstimate, the separation of the matrices should coincide with the separation distance between the maxima in the field gradient on either side of the magnet. Figure 7 shows an initial measurement wherein a 880 N force was required to move the piston through a 7 T field (similar to the profile shown in Fig. 5 produced by the Hiperco-encased magnet). The curve represents the magnitude of the force on the piston throughout the length of the stroke. A reduction in the magnitude of the net force to less than 450 N (100 lb) is desired to ensure smooth operation of the piston's drive mechanism. This is being pursued through force compensation methods which include changing the separation distance between the GGG matrices and by reshaping the field profile by changing the shape of the Hiperco end pieces of the magnet assembly. If required, an additional force compensation method, involving the placement of small slugs of GGG between the two matrices but thermally isolated so as not to become part of the refrigeration process, will be implemented.

The GGG piston is driven with a speed-controllable gearmotor having a maximum rotation rate of 10 rad/s. This rotational motion is converted to reciprocating motion by means of a commercially available "ball reverser," a nut with ball bearings that run in a cross-hatched track cut into the drive shaft. The track has a set stroke length of 9.2 cm and the angle of the track is set to provide a displacement 3.175 cm/2π rad. This permits a maximum linear speed of 5.1 cm/s for the GGG piston. A turn-around in the ends of the track automatically

reverses the direction of travel of the nut to provide smooth reciprocating motion without changing direction of rotation of the drive motor.

Heat exchange with either matrix is accomplished with helium gas provided by a separate bidirectional displacer (gas pump) in both the low and high temperature gas circuits. When the GGG piston is positioned to be adjacent to a port in the cylinder, a gas flow loop occurs. The indents in the outer surface of the piston in the area of the helium flow apertures allow the helium gas to flow through the porous matrices while the displacer is still in motion so that the gas flow need not occur only when the displacer is stalled at the ends of the stroke. The outer ridges of the piston form close tolerance seals to help prevent gas leakage along the cylinder wall. The seals further insure that gas leakage is minimized between the two gas loops. The design of the piston and gas circuitry is such that no mechanical cryogenic valves are required.

The two gas pumps (Fig. 8) are driven electromagnetically in phase relation to the motion of the GGG piston. Samarium cobalt permanent magnets are inserted in each end of the phenolic rod extending axially from the displacer. The coils are then energized with DC current in switched alternate directions to drive the displacer back and forth. The coil designs are being optimized to minimize the I^2R resistive heating in the coils. Superconducting NbTi coils are being tested for use with the low temperature gas pump. The volume displacement required of each pump was determined by

$$\dot{V} = \frac{\dot{Q}}{(\rho C_p \Delta T)}$$

where $\dot{V}$ is the volume flow rate of helium in the gas loop, ρ is the helium gas density, $\dot{Q}$ is the quantity of heat to be removed, C_p is the helium specific heat at constant pressure, and ΔT is the temperature change of the gas as it passes through the matrix. Volume displacements of 100 cc and 350 cc for the cold and warm temperature pumps were chosen assuming ΔT_C = 0.5 K and ΔT_H = 1.0 K and adding the dead volume for half of the corresponding gas loop.

The magnetic refrigerator has been designed to acheive high reliability. The magnetic refrigerator stage is a closed gas loop system; the gas circuit is sealed after the initial charge of helium gas. Internal gas displacers provide the movement of the gas through the circuitry, eliminating the need for an external compressor to provide the gas flow. The external and internal portions of the piston drive train are coupled together through a rotary seal to prevent gas contamination through the housing along the drive shaft. The magnetic refrigerator requires no small orifices as needed in the conventional Joule-Thomson valve, further minimizing the problems associated with gas contamination. Finally, the magnetic refrigerator will operate at slow reciprocating speeds, minimizing the wear rate of the low-temperature sliding seals.

IV. Conclusions

The design of a reciprocating magnetic refrigerator to pump heat from 4-15 K has been presented. The individual components have been designed and have been fabricated. Tests are underway to optimize the field profile and the placement of the GGG matrices within the piston. The assembly of the magnetic refrigerator has been initiated. With the experimental results that will be forthcoming, a careful analysis of this magnetic refrigerator concept can be used to design an efficient magnetic refrigerator usable for cooling maser amplifiers.

References

1. C. V. Heer, C. B. Barnes, and J. G. Daunt, The design and operation of a magnetic refrigerator for maintaining temperatures below 1 K, *Rev. Sci. Instrum.* 25:1088 (1954).

2. R. D. Britt and P. L. Richards, An adiabatic demagnetization refrigerator for infrared bolometers, *Int. J. Infrared Millimeter Waves* 2:1083 (1981).

3. S. Castles, Design of an adiabatic demagnetization refrigerator for studies in astrophysics, NASA/Goddard Space Flight Center Report No. X-732-80-9 (1980).

4. P. Kittel, Temperature stabilized adiabatic demagnetization for space applications, *Cryogenics* 20:599 (1980).

5. J. A. Barclay, O. Moze, and L. Paterson, A reciprocating magnetic refrigerator for 2-4 K operation: Initial results, *J. Appl. Phys.* 50:5870 (1979).

6. W. P. Pratt, Jr., S. S. Rosenblum, W. A. Steyert, and J. A. Barclay, A continuous demagnetization refrigerator operating near 2 K and a study of magnetic refrigerants, *Cryogenics* 17:689 (1977).

7. R. Beranger, G. Bon Mardion, G. Claudet, C. Delpeuch, A. F. Lacaze, and A. A. Lacaze, A gadolinium gallium garnet double acting reciprocating magnetic refrigerator, in *Advances in Cryogenic Engineering, Vol. 27*, Plenum Press, New York (1982).

8. A. F. Lacaze, R. Beranger, G. Bon Mardion, G. Claudet, and A. A. Lacaze, Double acting reciprocating magnetic refrigerator: recent improvements, Proceedings of CEC and ICMC: BA-2, Colorado Springs (1983).

9. J. R. Van Geuns, A study of a new magnetic refrigerating cycle, *Philips Res. Rep. Suppl.* 6:1 (1966).

10. H. Nakagome, N. Tanji, O. Horigami, T. Numazawa, Y. Watanabe, and T. Hashimoto, The helium magnetic refrigerator I: Development and experimental results, *Proceedings of CEC and ICMC: FC-10*, Colorado Springs (1983).

11. R. Longsworth, Cryocooler activities in China, in *Refrigeration for Cryogenic Sensors and Electronic Systems*, National Bureau of Standards Special Publication (1983).

12. D. L. Johnson, Magnetic refrigeration for maser amplifier cooling, in *The Telecommunications and Data Acquisition, Progress Report 42-67*, pp. 29-38, Jet Propulsion Laboratory, Pasadena, CA, Feb. 1982.

Table 1. CCR system requirements for maser cooling

Reliable
Efficient
Multiyear lifetime
Unattended operation
Rapid cooldowns
1–4 W cooling capacity
Compact
Magnetic field isolation of maser package
Low microphonics
mK temperature stability
Orientation independence
Continued operation during power failures

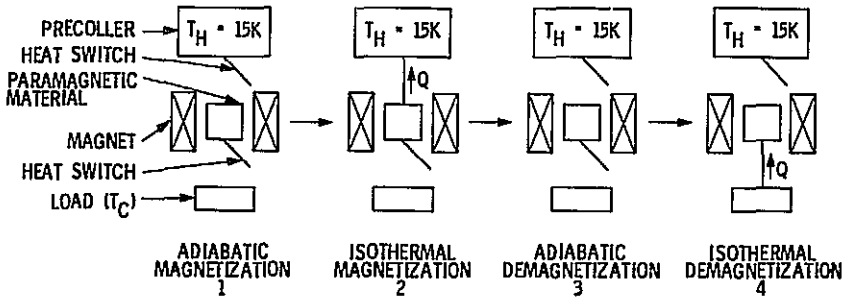


Fig. 1. Schematic of Cyclic Magnetic Refrigerator showing the refrigerator operating in the four stages of a Carnot cycle

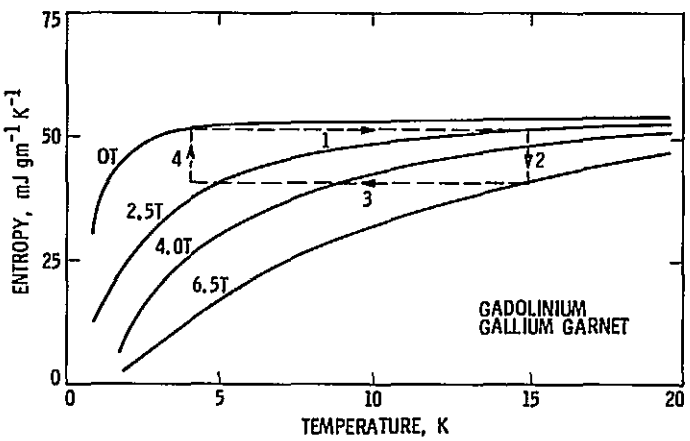


Fig. 2. S,T-diagram depicting a gadolinium gallium garnet magnet refrigerator operating in a Carnot cycle. The different stages of the Carnot cycle correspond to the illustrations of Fig. 1.

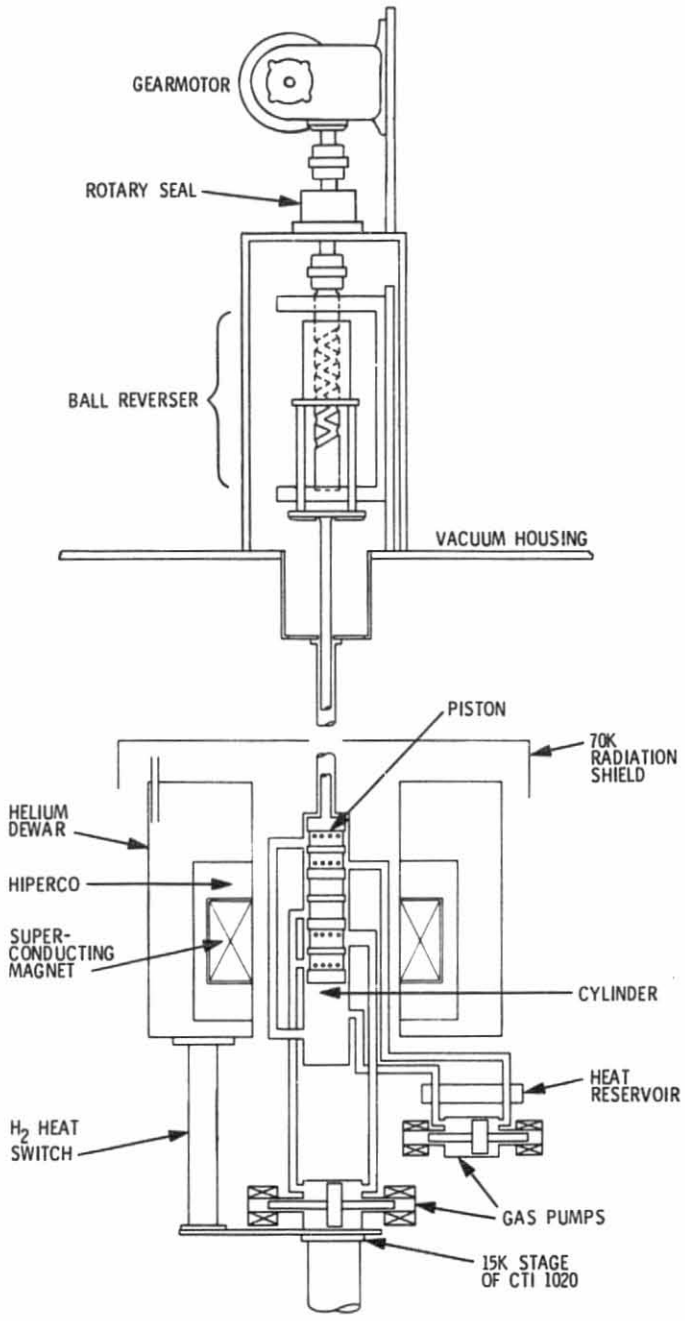


Fig. 3. Schematic of the engineering model magnetic refrigerator design

Fig. 4. Superconducting magnet/Hiperco assembly

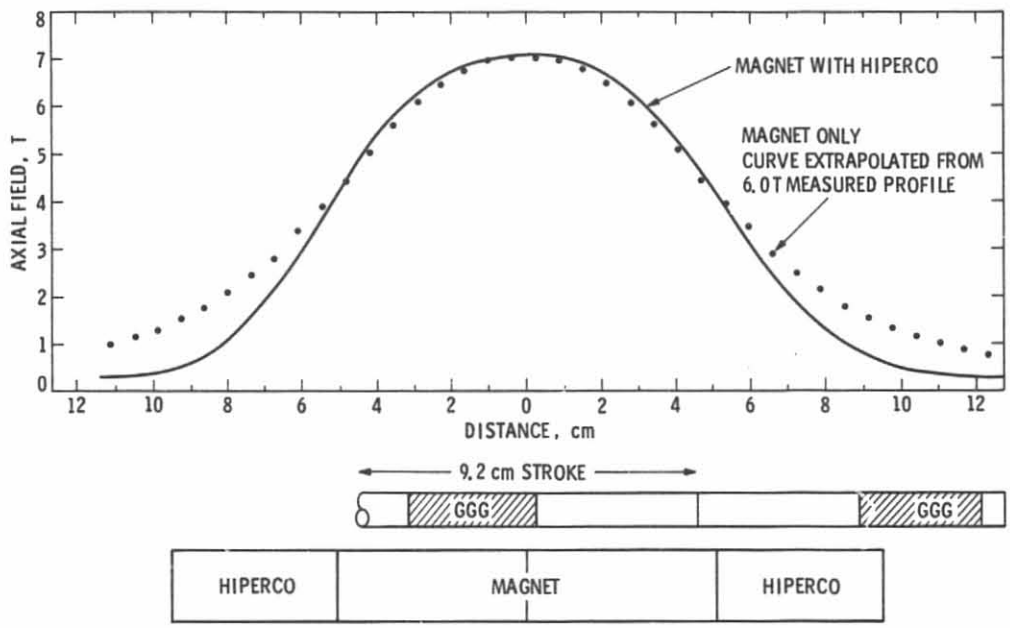


Fig. 5. Magnetic field profiles

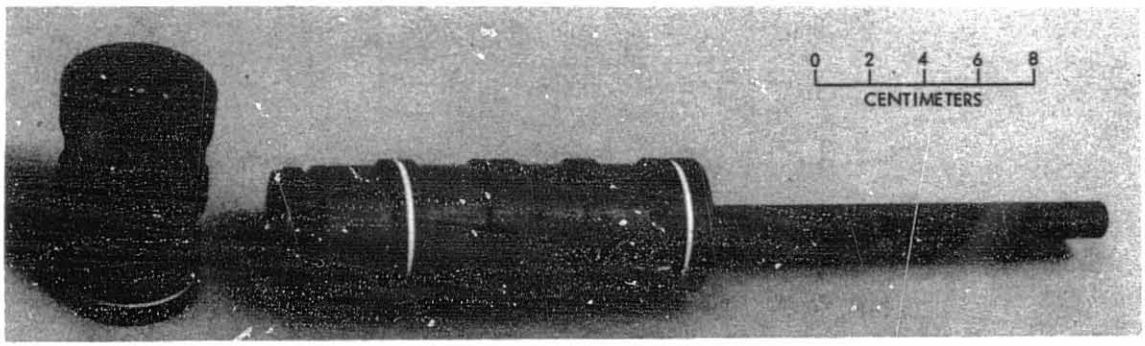


Fig. 6. Phenolic piston containing the paramagnetic material matrices

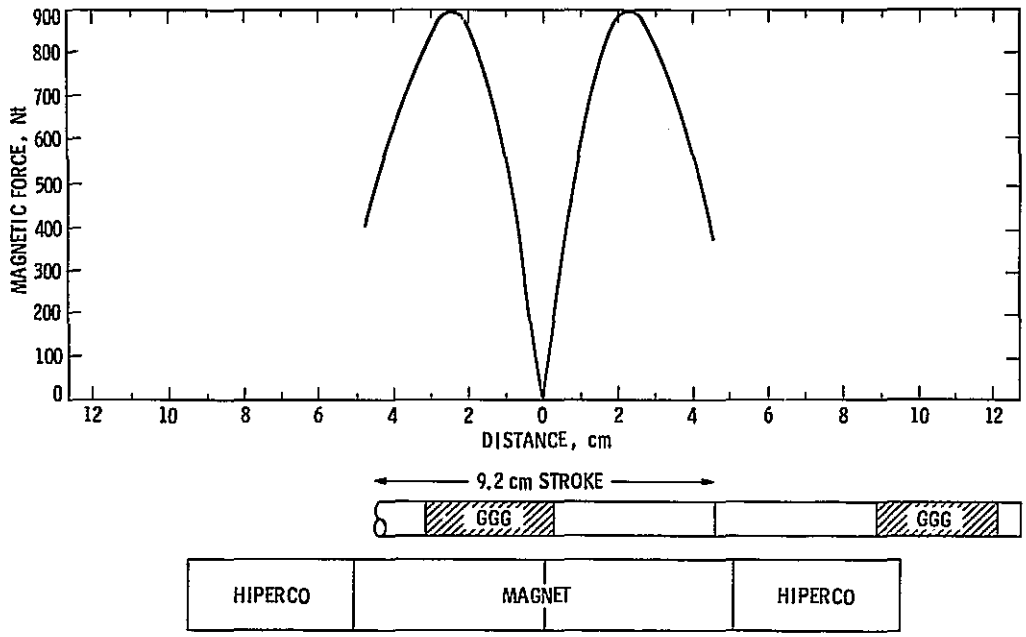


Fig. 7. Magnitude of the magnetic force acting on the piston throughout the stroke length

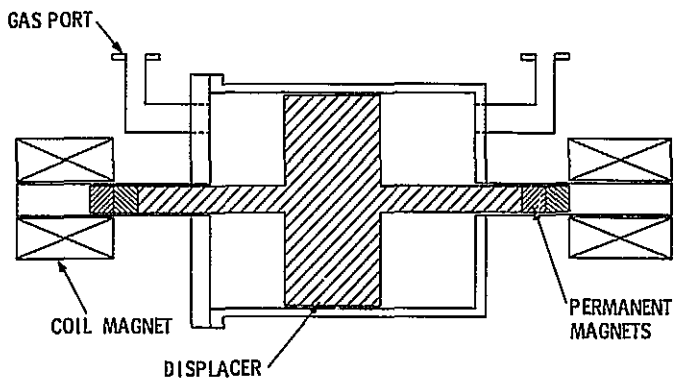


Fig. 8. Schematic of the gas pumps

Potential Surface Improvements by Bump Removal for 64-m Antenna

S. Katow and C. N. Guiar
Ground Antenna and Facilities Engineering Section

The surface panels of the main reflector of the 64-m antenna are initially set at an elevation angle of 45 deg, where most tracking occurs, to ideally match a prescribed paraboloid. As the antenna is rotated about the elevation axis, distortions are introduced at the surface panel's supporting nodes as well as at the main reflector backup structure by changes in the direction of the gravity forces relative to the reflector symmetric axis. Major bump displacements could be corrected by controlling the position of the surface-panel corners using adjustable mechanical jacks that change in length with the antenna elevation angle. The analysis of two bump-removal configurations is presented and one unique adjustment mechanism is proposed. A gain recovery of 0.2 dB at X-band would be available if the reflector structure distortion rms were reduced from 0.63 mm (0.025 in.) to 0.15 mm (0.006 in.).

I. Introduction

Antenna surface-panel distortions and deflections caused by changes in gravity loading are introduced by antenna rotation about the elevation axis. These gravity-induced deflections result in differences in the radio-frequency (RF) pathlengths, thus contributing to the RF gain losses of the antenna. The gain loss is a function of, among other parameters, the root-mean-square (rms) of the distortions of the main reflector and of the operating frequency being used.

As the planetary exploration program continues to grow, the need for an efficient antenna system that provides increased gain, performance, and productivity becomes evident. Upgrading the present 64-m antenna network to improve performance and gain proves to be economically more practical than building a new replacement system. Viable modifications, presently under investigation, include extending the reflector aperture to 70-m, improving the surface panel fabrication accuracy and setting precision, reducing the gravity-induced distortions by stiffening braces, and increasing the operating frequency (Ref. 1).

This article describes one additional possible modification in upgrading the 64-m antenna. It calls for the correction of major displacements by controlling, on a real-time basis, the height of the surface panel's corners using mechanical means, with corrections changing with the elevation angle changes.

A preliminary study of the potential benefits (Ref. 2) resulted in an improved surface with less distortions (rms value), where simplified assumptions of displacement in the major bumpy areas were made. This report, a follow-up to the prior study, describes an improved computational technique used to determine the nodal corrections necessary to reduce the surface tolerance. A conceptual mechanical design to implement the corrections is also presented in this report.

II. Analysis

Since the surface panels of the 64-m antenna are initially set at an elevation angle of 45 deg to match a prescribed paraboloid, gravity-induced distortions are introduced at the horizon, zenith attitude, or any position in between while the antenna rotates about the elevation axis.

The distortion vectors of the joints (or nodes) in the structure supporting the reflector panels are first computed for unit gravity loading (one g where g is the acceleration of gravity) in each of the symmetric and antisymmetric directions as shown in Fig. 1. Either NASTRAN or IDEAS (in-house) structural analysis computer programs can be used for this purpose. Using the relationship of the unit gravity load vectors and their components in the symmetric and antisymmetric directions, as shown in Fig. 1, the three-dimensional distortion vectors at any elevation angle are given by:

$$\mathbf{d}_\theta = (\sin\theta - \sin\theta_s)\,\mathbf{U}_{sym} + (\cos\theta - \cos\theta_s)\,\mathbf{U}_{antisym} \quad (1)$$

where

$\mathbf{U}_{sym}$ = symmetric, unit gravity distortion vector for gravity off to on at θ_s

$\mathbf{U}_{antisym}$ = unit gravity distortion vector for gravity off to on at θ_s

$\mathbf{d}_\theta$ = gravity distortion vector at angle θ

θ = an. nna elevation angle

θ_s = elevation angle . which panels are set (usually 45 deg)

Th. RMS program (Ref. 3) is used to compute the "normal" surface errors (perpendicular to reflector surface) at a specific elevation angle (θ) using the two sets of deflections d_{sym} and $d_{antisym}$. Only two sets of deflection data (for the first and fourth quadrants of the antenna) are supplied to the RMS program. Data for the second and third quadrants are generated by symmetry. The RMS program multiplies both sets of deflection data by the appropriate angular functions in Eq. (1), adds the resulting weighted deflections, and computes the surface normal errors after making a paraboloid best fit that minimizes the pathlength errors. For the antenna horizon-look case ($\theta = 0°$) at $\theta_s = 45$ deg, the distortion vector ($\mathbf{d}_0$) is formed as:

$$\mathbf{d}_0 = -0.707\,\mathbf{U}_{sym} + 0.293\,\mathbf{U}_{antisym} \quad (2)$$

and for the zenith-look case ($\theta = 90$ deg) at $\theta_s = 45$ deg,

$$\mathbf{d}_{90} = 0.293\,\mathbf{U}_{sym} - 0.707\,\mathbf{U}_{antisym} \quad (3)$$

The output of the rms program provides two types of normal errors: (1) no-fit error and (b) best-fit error. One file outputs the no-fit normal errors and another file outputs the best-fit normal errors. The no-fit normal error is the total distortion vector due to gravity loading changes as the antenna rotates from the setting elevation angle (θ_s) to the horizon or zenith configuration. This means that the no-fit normal errors are the three-component distortion vectors normalized to the given paraboloid surface.

The RMS program also outputs the best-fit root-mean-square value of the ½-pathlength error data together with a contour map of the normal errors. Contour maps of the gravity off/on distortions for the 64-m antenna measured normal to the surface of the best-fit paraboloid for the antisymmetric horizon-look ($\theta = 0$ deg) and symmetric zenith-look ($\theta = 90$ deg) cases are shown in Figs. 2 and 3, respectively.

The surface distortion contour maps for the unit gravity (off to on) loading vector provide important information. Inspection of the contour maps in Figs. 2 and 3 for the U vectors show major displacements, hereafter referred to as "bumps." Removing these bumps would minimize the overall distortion (rms) of the main reflector, thus increasing the gain of the antenna. Since the bumps, from Figs. 2 and 3, are on distinctly different parts of the paraboloid surface, i.e., independent, it can be assumed that bump nodes for the zenith-look case ($\theta = 90$ deg) are affected only by symmetric gravity loading, and bump nodes for the horizon-look case ($\theta = 0$ deg) are affected only by antisymmetric gravity loading. This independence of bumps at these two extreme positions allows the use of eccentric rollers (driven by the rotation of the elevation-axis shaft) to reduce gravity distortions. The eccentric rollers produce sine and cosine linear functions of the elevation angle, θ, that match, by proper design, the desired distortion compensation due to gravity loading changes.

Note that the distortion (in rms) for the unit symmetric loading is 0.863 mm, which is larger than the 0.616 mm for the unit antisymmetric loading vector. Therefore, the horizon-look case was selected in this study for corrections at selected

nodes since 0.7 part of the symmetric loading change occurs from 45-deg elevation to horizon-look. Another reason for that selection was the fact that the DSN 64-m antenna is used most frequently between the 45-deg to horizontal position range. The analytical technique used to determine the most effective node corrections is described in the following section.

III. Trial Computations

The analysis for determining the bump corrections for the 64-m antenna was done in two parts. The first part determined the necessary bump-node corrections for the horizon-look case. The second part of the correction analysis determined the relationship between the horizon-look and zenith-look corrections, which generated the necessary bump-node corrections and the new overall rms where the zenith-look gravity distortions were initially assumed to be due to the symmetric gravity loading vector only. Therefore, the initial zenith-look surface corrections are a function of the correction made at horizon-look.

The RMS program, using the three-component distortion data from the structural programs, produced at each node a listing of both the no-fit normal errors and the best-fit normal errors (after best fitting of the paraboloid). By examining the best-fit contour map of the horizon-look case, with the surface panels set at 45-deg elevation, a number of large, bump-displacement nodes were observed. Data for 157 nodes were entered in the program with zero displacements as a baseline from which corrections could start.

An initial set of corrected nodes was generated after subtracting the best-fit normal errors from the no-fit normal errors for the large displacement nodes. This correction corresponds to the distance from the actual paraboloid to the best-fit paraboloid and results in a substantially improved rms of 0.27 mm (0.0107 in.). The best-fit results associated with this new rms represent the normal errors to the corrections provided by the new deflection data.

Additional corrections to the selected set of bump nodes were attempted to improve further the overall rms. In the first node-correction attempt, (option-1 design) the best-fit normal errors were subtracted directly from the no-fit normal errors. Subtraction was necessary to determine the distance between the actual paraboloid and the best-fit paraboloid. Since the new best-fit data produced a best-fit paraboloid for which the vertex is shifted in the Z direction as the bump-node errors are removed, the new best-fit paraboloid must be shifted so that its vertex coincides with that of the no-fit paraboloid. An over-correction factor of 1.8 was used to compensate for this offset. A new set of corrected nodes can therefore be generated by subtracting the new best-fit data, multiplied by the over-correction factor, from the first set of corrected node data. This new set of corrected nodes resulted in an improved overall rms of 0.18 mm (0.007 in.).

Some surface nodes originally gave a best-fit rms near zero, thus misleadingly eliminating the need for any further corrections. However, repeated corrections of other nodes shifted the best-fit paraboloid in the Z direction by approximately 0.50 mm (0.020 in.), giving the originally "good" nodes a poor rms. To fine-tune the corrections made thus far, 63 of these nodes were given a correction designed to compensate for the shift in the Z direction. The final rms for (157 + 63) or 210 corrected nodes (with the surface panels set at 45 deg) is 0.15 mm (0.006 in.). The structural model for the selected nodes of this case is shown in Fig. 4. Figures 5 and 6 show the contour maps of the distortions measured normal to the surface of the best-fit paraboloid before and after the bumps are removed for the horizon-look case.

Corrections for the zenith-look case are made for the same nodes as for the horizon-look case and are initially assumed to be affected by symmetric gravity loading only. Therefore the zenith-look case (θ_Z) corrections are functions of the horizon-look (θ_H) case corrections and are given as:

$$\text{sym-gravity corrections} = (\text{hor-look corrections}) \frac{(\sin\theta_Z - \sin\theta_s)}{(\sin\theta_H - \sin\theta_s)}$$

$$= (\text{hor-look corrections}) \frac{0.3}{0.7} \qquad (4)$$

and for antisymmetric distortions:

$$\text{antisym-gravity corrections} = (\text{hor-look corrections}) \frac{(\cos\theta_Z - \cos\theta_s)}{(\cos\theta_H - \cos\theta_s)} \qquad (5)$$

where θ_Z is 90 deg, θ_H is 0 deg, and θ_s = 45 deg.

The initial set of corrected nodes for the zenith-look case was generated using Eqs. (4) and (5) after determining the rms using the initial set of corrected nodes. It was found that the prior assumption that the zenith-look case has only symmetric gravity loading is not quite accurate. Most of the nodes, however, agree with that assumption, but a few have distortions due to both symmetric and antisymmetric gravity loadings. The multiplication factors for the nodes exhibiting these traits were scaled accordingly and produced a final rms for the 220 corrected nodes (with the surface panels set at 45-deg elevation) of 0.23 mm (0.009 inches).

P. Potter (Ref. 1) designated a minimum acceptable surface tolerance for the 64-m antenna that allows efficient operation

at the 32-Ghz (Ka-band) frequency. The overall surface rms determined in this study for the zenith-look and horizon-look cases are below these tolerances. Reducing the number of nodes selected for this study will increase the overall rms to meet the minimum requirements of Ref. 1 and would result in a cost effective modification. The reduction effort, represented as option-2 design, results in an overall rms of 0.23 mm (0.009 in.) for the horizon-look case and 0.38 mm (0.015 in.) for the zenith-look case.

IV. Proposed Panel Adjustment Mechanism

The surface panels are supported from the reflector structure as shown in Fig. 7 by the inverted U-pads, which are now weld-connected to the top chords of the ribs. Figure 8 illustrates the proposed changes in the mounting of the inverted U-pads using flextures. The inverted U-pads can be pivoted to vary their heights up to the top chords of the reflector's rib trusses.

The proposed mechanism for raising or lowering the U-pad is illustrated in Fig. 9, and the mechanism is bolted to the U-pad and the top chord. An eccentrically mounted roller, which rotates in unison with the elevation axis, can provide the sine (θ) function as well as the amplitude change of the surface-panel position.

The driving flexible cables, connecting the antenna elevation shaft motion and the eccentric roller motion, can be driven in unison by a specially designed gearbox that has a large spur gear driving multiple smaller pinion gears placed around the periphery of the main gear. If there is excess torque capacity in the elevation drive, this special gearbox can be mounted on the alidade next to the elevation shaft protruding from the elevation bearing housing; the spur gear would be driven directly by the rotating elevation axis. The gearbox may also be driven by its own power unit using a simple switching device to maintain synchronous rotation with the elevation axis.

The angle of twist in the flexible driving cables between the special gearbox and the eccentric roller will determine the accuracy of the corrections imposed on the surface panels. The final design may require the use of a worm-gear reduction instead of the spur gear shown in Fig. 9. Initial cost estimates, however, show that further development is still needed.

V. Summary

This article has considered only the distortions of the main reflector surface resulting from gravity loading. Additional distortions result from both wind and temperature-difference loadings on the reflector structure, the surface panels, and the subreflector; these loadings become dominant for antenna operation at higher than X-band frequencies. Reflector nodes having major bump displacements could be corrected by controlling the position of the surface panel corners using adjustable mechanical jacks, which change position with changes in the antenna elevation angle. The analysis of two bump-removal configurations is presented in Table 1. The potential gain improvement is about 0.2 dB at X band (8.4 GHz) as the reflector surface distortion is reduced from 0.63 mm (0.025 in.) to 0.15 mm (0.006 in.). Although the proposed panel-adjustment mechanism can neutralize the nodal displacements at the selected corrected nodes, it requires future development effort and cost trade-off studies among other antenna upgrade options presently being implemented.

References

1. Potter, P. D., "64-Meter Antenna Operation at Ka-Band," in the *Deep Space Network Progress Report 42-57*, March and April 1980. Jet Propulsion Laboratory, Pasadena, Calif., pp. 65-70.

2. Katow, M. S., "A Proposed Method of Reducing the Gravity Distortions of the 64-Meter Antenna Main Reflector," in the *Deep Space Network Progress Report 42-23*, Oct. 1974. Jet Propulsion Laboratory, Pasadena, Calif., pp. 92-97.

3. Katow, M. S., and Schmele, L. W., "Antenna Structures: Evaluation Techniques of Reflector Distortions," in *Supporting Research and Advanced Development, Space Programs Summary 37-40*, Vol. IV, Sept. 30, 1968. Jet Propulsion Laboratory, Pasadena, California, pp. 73-76.

Table 1. Surface distortion (rms) due to gravity off/on loading

	Horizon position, mm (in.)	Zenith position, mm (in.)
Original 64-m antenna (structural only)	0.63 (0.025)	0.50 (0.020)
Upgrade option – 1 (220 corrected nodes)	0.15 (0.006)	0.23 (0.009)
Upgrade option – 2 (116 corrected nodes)	0.23 (0.009)	0.38 (0.015)

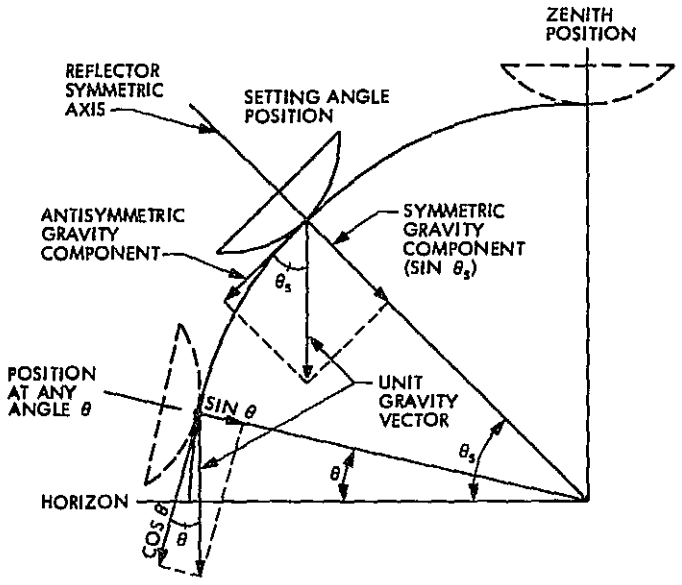
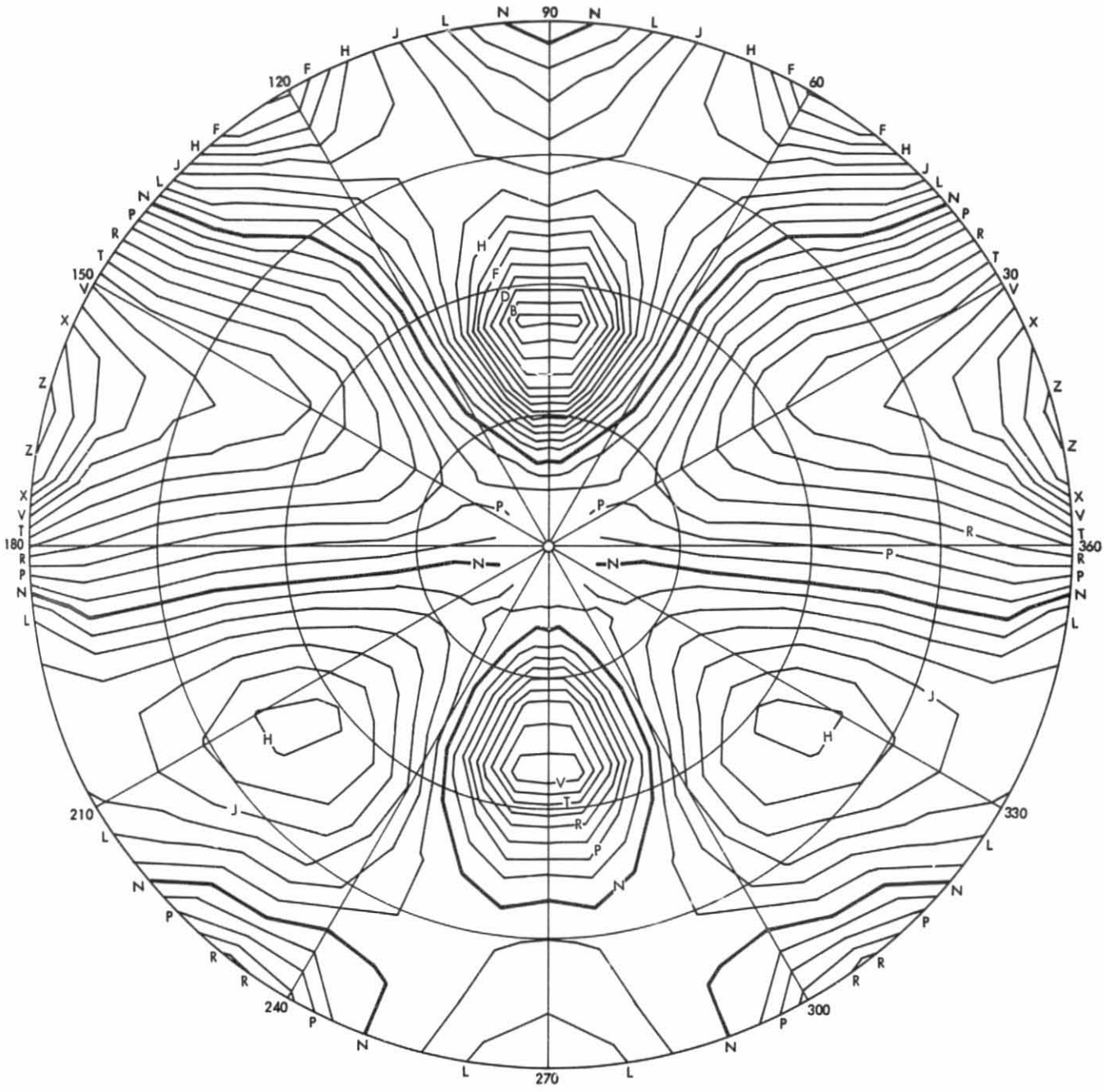


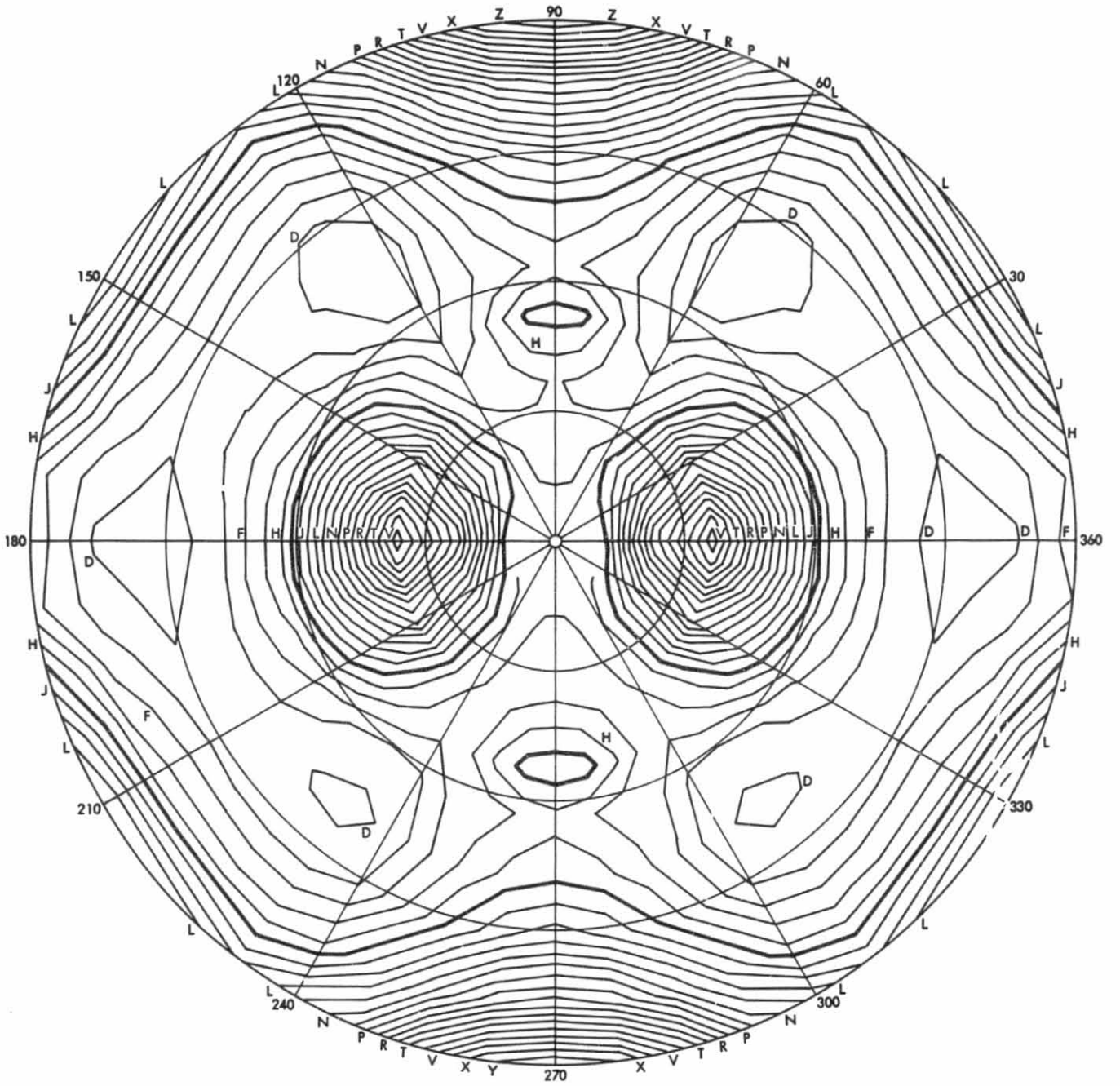
Fig. 1. Components of the gravity force vector



CONTOUR DEFINITIONS

NORMAL ERROR, mm	LABEL	NORMAL ERROR, mm	LABEL	NORMAL ERROR, mm	LABEL
-1.95	A	-0.60	J	0.75	S
-1.80	B	-0.45	K	0.90	T
-1.65	C	-0.30	L	1.05	U
-1.50	D	-0.15	M	1.20	V
-1.35	E	0.00	N	1.35	W
-1.20	F	0.15	O	1.50	X
-1.05	G	0.30	P	1.65	Y
-0.90	H	0.45	Q	1.80	Z
-0.75	I	0.60	R		

Fig. 2. Horizon-look (unit gravity, antisymmetric loading off/on) normal errors to best-fit paraboloid for existing 64-m antenna



CONTOUR DEFINITIONS

NORMAL ERROR, mm	LABEL	NORMAL ERROR, mm	LABEL	NORMAL ERROR, mm	LABEL
-1.60	A	0.20	J	2.00	S
-1.40	B	0.40	K	2.20	T
-1.20	C	0.60	L	2.40	U
-1.00	D	0.80	M	2.60	V
-0.80	E	1.00	N	2.80	W
-0.60	F	1.20	O	3.00	X
-0.40	G	1.40	P	3.20	Y
-0.20	H	1.60	Q	3.40	Z
0.00	I	1.80	R		

Fig. 3. Zenith-look (unit gravity, symmetric loading off/on) normal errors to best-fit paraboloid for existing 64-m antenna

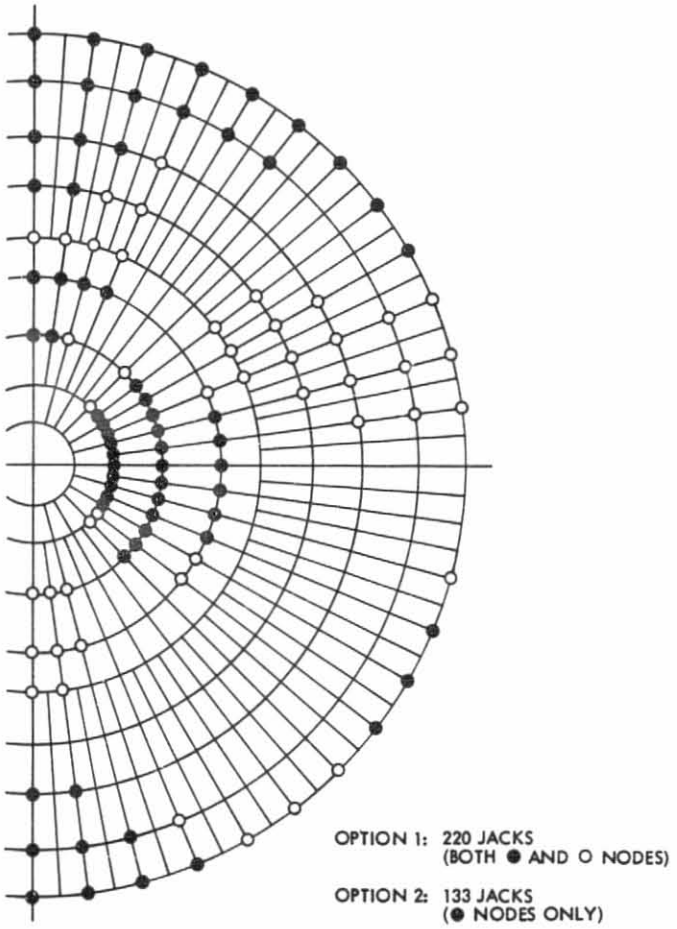
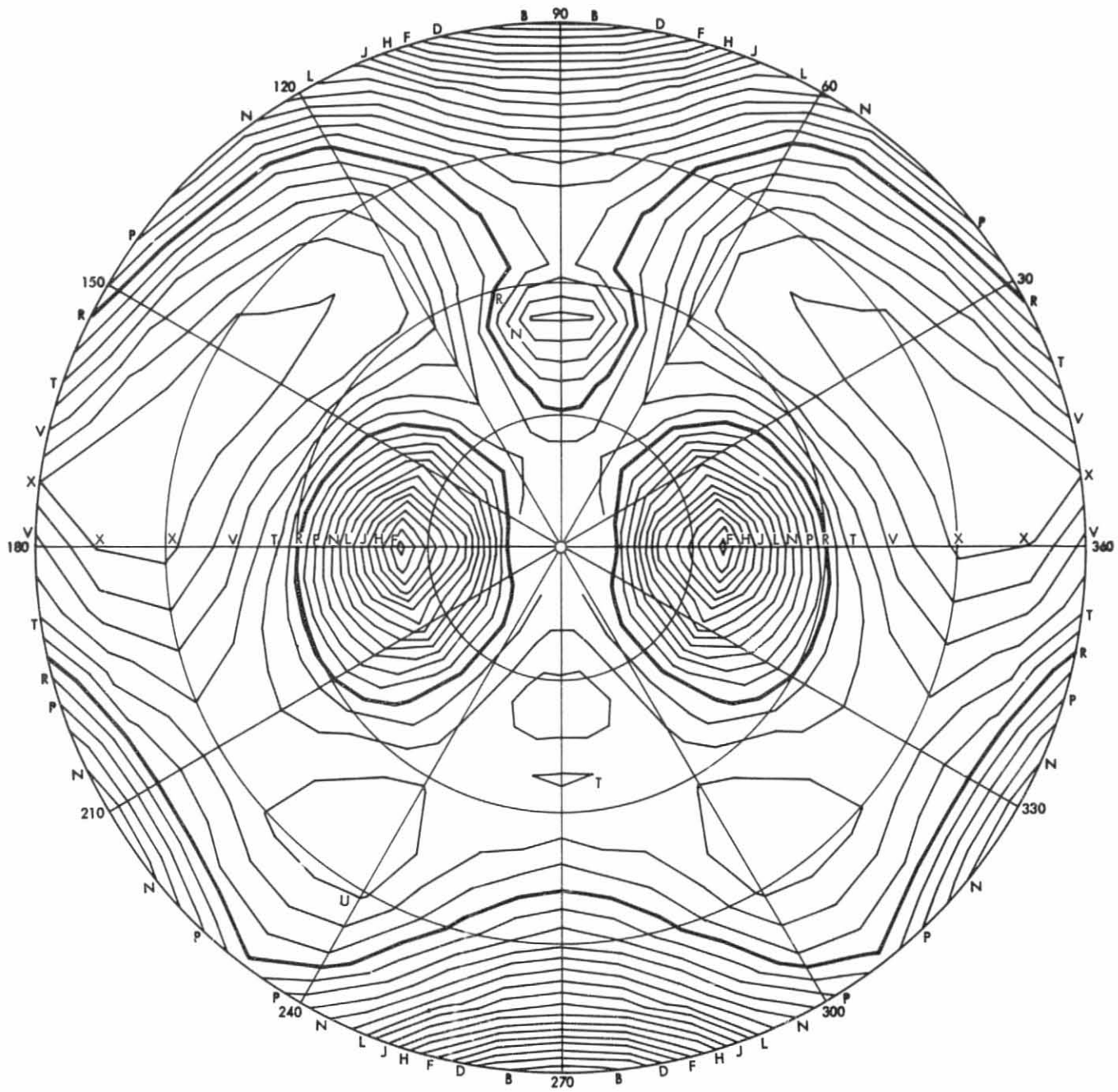


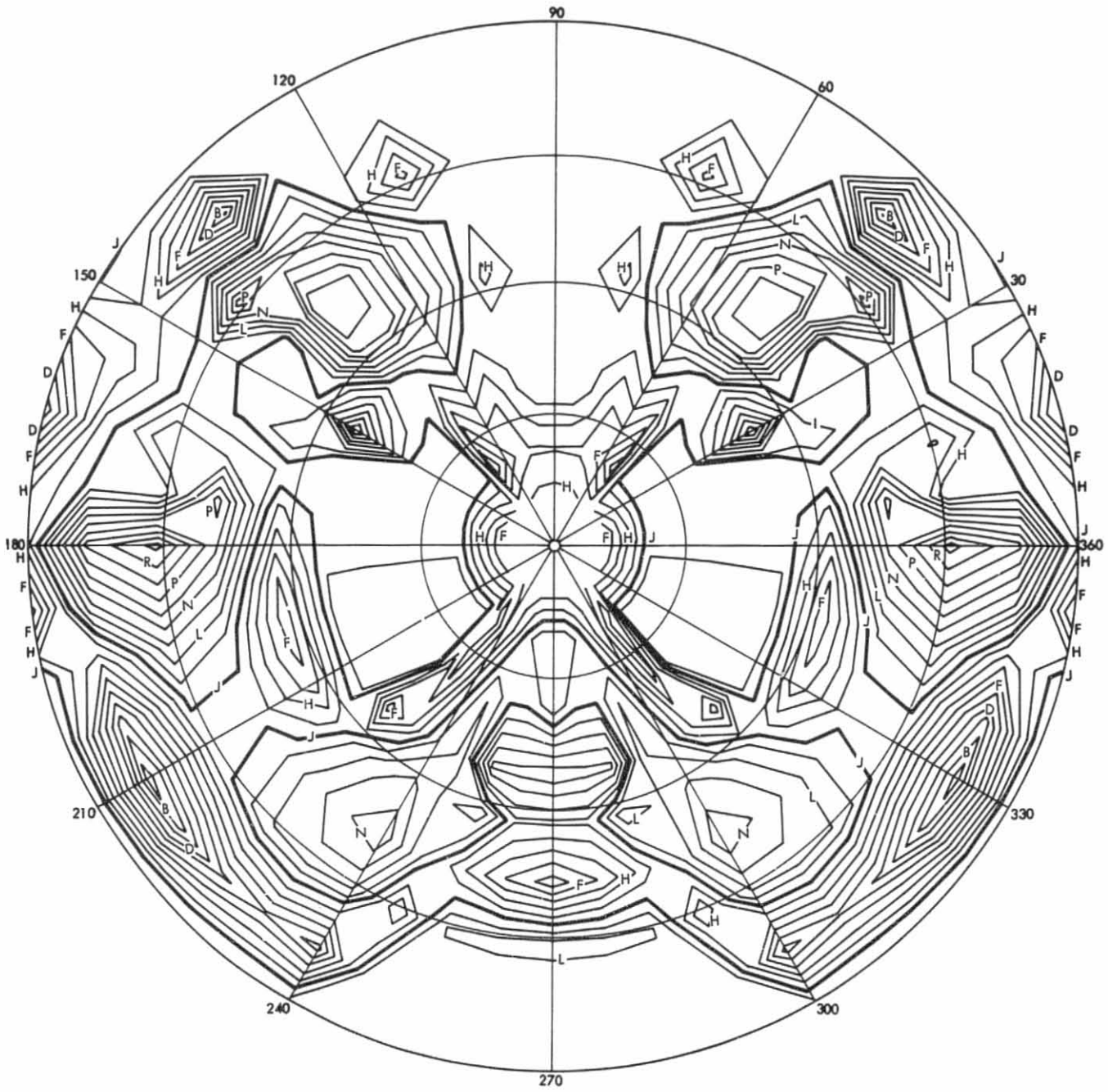
Fig. 4. Bump-removing adjustable jack stations for 64-m antenna



CONTOUR DEFINITIONS

NORMAL ERROR, mm	LABEL	NORMAL ERROR, mm	LABEL	NORMAL ERROR, mm	LABEL
-2.55	A	-1.20	J	0.15	S
-2.40	B	-1.05	K	0.30	T
-2.25	C	-0.90	L	0.45	U
-2.10	D	-0.75	M	0.60	V
-1.95	E	-0.60	N	0.75	W
-1.80	F	-0.45	O	0.90	X
-1.65	G	-0.30	P	1.05	Y
-1.50	H	-0.15	Q	1.20	Z
-1.35	I	0.00	R		

Fig. 5. Horizon-look (panels set at 45-deg elevation) normal errors to best-fit paraboloid for existing 64-m antenna; rms = 0.63 mm (0.025 in.)



CONTOUR DEFINITIONS

NORMAL ERROR, mm	LABEL	NORMAL ERROR, mm	LABEL	NORMAL ERROR, mm	LABEL
-0.45	A	-0.10	H	0.25	O
-0.40	B	-0.05	I	0.30	P
-0.35	C	0.00	J	0.35	Q
-0.30	D	0.05	K	0.40	R
-0.25	E	0.10	L	0.45	S
-0.20	F	0.15	M		
-0.15	G	0.20	N		

Fig. 6. Horizon-look (panels set at 45-deg elevation) normal errors to best-fit paraboloid with option-1 corrections; rms = 0.15 mm (0.006 in.)

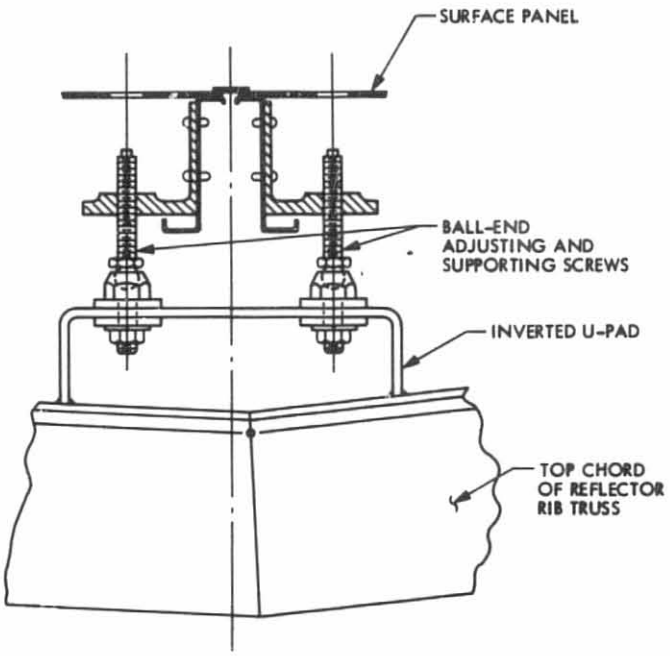


Fig. 7. Typical surface panel support

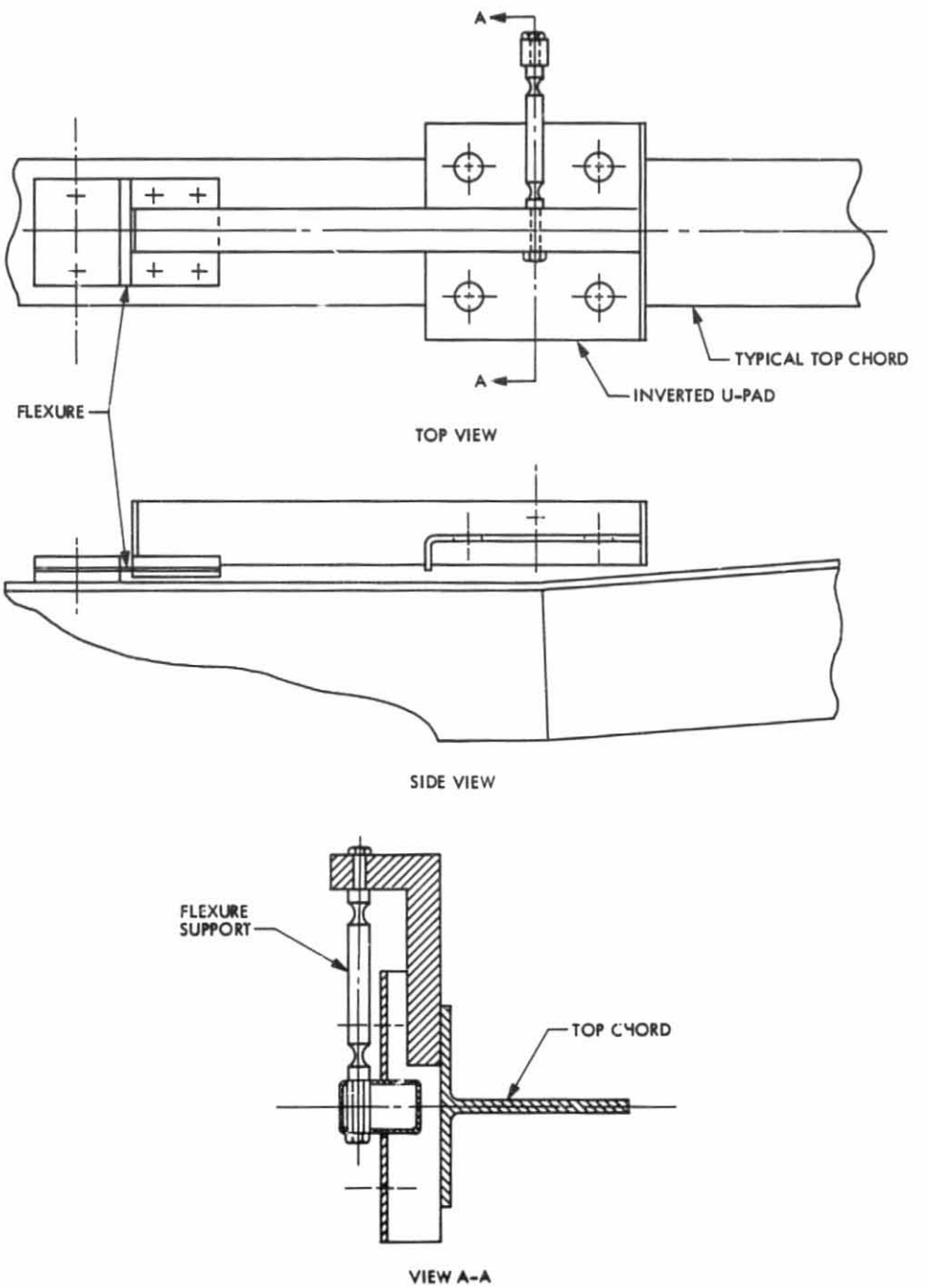


Fig. 8. Proposed pivoted inverted U-pad

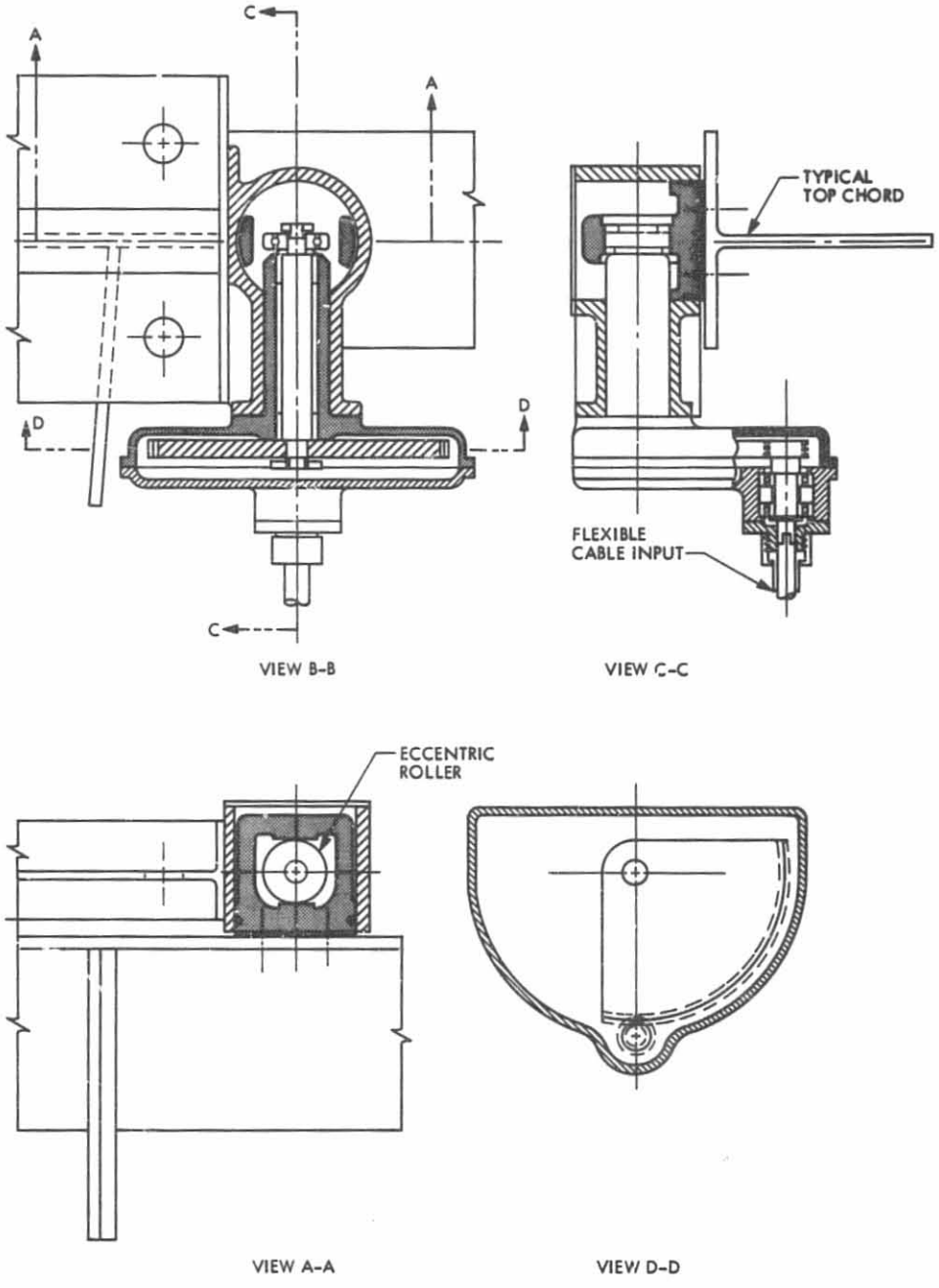


Fig. 9. Proposed eccentric roller assembly

64-m Antenna Automatic Subreflector Focusing Controller

C. N. Guiar and L. W. Duff
Ground Antennas and Facility Engineering Section

Defocussing of the radio frequency beam arises from gravity-induced structural deformations as the antenna rotates about the elevation axis. The new Subreflector Controller generates the axial (z) ana lateral (y) offset corrections necessary to move the subreflector, thus minimizing the gain losses due to this defocussing. This article discusses the technique used to determine these offset errors and presents a description of the new Subreflector Controller.

I. Introduction

With the progressive needs to communicate with farther depths of space comes the increased need to improve the efficiency of the existing NASA - JPL Deep Space Network (DSN) ground antennas. The DSN, incorporating both mechanical and microwave engineering efforts, has initiated this task, with a main goal to increase the large antenna (64-m diameter) network performance by approximately 1.9 dB at X-band.

Some of the modifications being developed include the following:

(1) Fabrication of precise main and subreflector surfaces (0.5 dB)

(2) The use of optimally shaped single or dual reflectors (0.3 dB)

(3) The extension of the main reflector diameter to 70 m (0.8 dB) with several structural and optical pointing changes (0.1 dB)

(4) The subreflector focusing automation and upgrade (0.2 dB) described in this report

The present 64-m antenna subreflector consists of a 6.4-m (21-ft) diameter, asymmetrical, hyperboloidal surface with a moveable vertex plate, a 0.3-m (1-ft) high solid skirt attached at a fixed tilt angle about the perimeter of the hyperboloid, a hub, a backup space frame structure, and four independently adjusted motion mechanisms comprising electric motors and jack screws. The four subreflector motion (focusing) mechanisms are located to allow linear travel along the x- (cross elevation), y- (elevation), and z- (axial or microwave beam) axes in addition to rotational travel to select any of several feed horns for use. These focus adjustments permit the subreflector to optimize the radio frequency (RF) alignment and maximize gain for any one of the five feed positions on the tricone assembly. Defocussing of the RF beam arises from gravity-induced structural deformations as the antenna rotates about the elevation axis. The subreflector controller generates the corrective signals which move the subreflector to minimize the gain loss due to defocussing. The axial (z) and y-axes corrections are automated, while any X-axis corrections (minor) can be adjusted manually, if needed.

This report discusses the technique used to determine the focusing offsets or Δy and Δz corrections necessary to reduce gain losses due to gravity-induced structural deformations.

Also included is a description of the new subreflector controller designed to enforce these corrections. Operating instructions, theory of operation, and associated software for the SRC belong in an operation and maintenance manual (in preparation).

II. Focus Offsets

The surface panels of the 64-m antenna main reflector are initially set to represent ideal cassegrain conditions when the antenna is oriented at an elevation angle of 45 degrees (where most tracking occurs). As the antenna rotates about the elevation axis, gravity distortions are introduced due to changes in the direction of the gravity force vectors with respect to the antenna symmetric (z) axis (Ref. 1). Structure deformations contribute to a reduced RF performance since these displace the focus as demonstrated in Fig. 1 and increase antenna gain losses.

Displacements of the focus in axial (z) and lateral (y) directions have major effects on gain loss because they correspond to the focus of the best-fit paraboloid and the virtual secondary focus of the hyperboloid. These offsets can be compensated for using the new subreflector controller which automatically generates the correct signals to move the subreflector in both the axial and y directions.

A series of tests on the large antenna at DSS 14 were run to determine the optimum subreflector z-axis focus position vs the elevation angle. These tests consisted of a series of conical scanning (CONSCAN) boresight and subreflector focus measurements using an astronomical radio source. A polynomial least squares curve fit was applied with a correction function of the form

$$\Delta z = A + BX + CX^2 + DX^3 \qquad (1)$$

where Δz is the Z-axis focus offset and X is the elevation angle complement. The constants are found as follows: A = 14.2 mm (0.560 in.), B = -0.068 mm/deg (-0.00271 in./deg), C = -0.0069 mm/deg (-0.000275 in./deg), D = 0.000033 mm/deg (0.0000013 in./deg) and X is (90-E) where E is the elevation angle in degrees. This polynomial provided the required correction of the subreflector Z- position, plotted as shown in Fig. 2. A maximum correction of about 25 mm (1 in.) is needed therefore at E = 0 (horizon position).

The y-axis focus offset (Δy) was determined analytically using the NASTRAN structural analysis computer program. The results included the offset of the best-fit paraboloid focus from the virtual focus of the hyperboloid (relative to setting elevation angle of 45 degrees). After curve-fitting, a y-axis correction equation is obtained as

$$\Delta y = 3.41\ [\sqrt{2}\cos E - 1] \qquad (2)$$

where Δy is the y-axis focus offset in inches. Equation (2) is plotted as shown in Fig. 2. At extreme elevation angles (90°) the y-axis offset can reach approximately 100 mm (4 in.).

After determining the Δy and Δz offsets, the Radiation Pattern Computer program (Ref. 2) was used to determine the expected gain losses that would result without subreflector corrections. As an example, Fig. 3 shows a plot of the gain loss given as its equivalent surface distortion (in mm) vs focus offsets for 8.45 GHz.

The new subreflector Controller software incorporates Eqs. (1) and (2) and is described in detail below.

III. Controller Description

The subreflector axial (z) and y positioning are controlled by a closed loop as shown by the block diagram in Fig. 4. The loop is closed using the subreflector axial (or y) synchro position encoder as the feedback element.

The Subreflector Controller (SRC) and Interface are sketched in Fig. 5. The SRC receives the antenna elevation position data from the Antenna Servo Controller (ASC) at a rate of one sample per second. The ASC transmits this data to the SRC through the 534 Serial Communications Board. The maximum elevation the antenna is able to change is 0.25 deg/s. The antenna elevation angle is used as an index into a look-up table stored in the memory of the SRC where the subreflector offset values are located (as defined by Eqs. [1] and [2]). The actual subreflector position is read from the X, Y, and Z position synchros on the subreflector and entered into the SRC through synchro-to-digital converter (12-bit). If the actual position of the subreflector is not within 1.27 mm (0.050 in.) of the desired position, then a rate command will be generated through the 12-bit digital/analog converter (D/A), which will engage the motor drives of the subreflector at a fixed rate. Once the motor drives start, they will continue to move the subreflector until it has reached its desired position. At that time all commands to the subreflector motor drives will stop. The final position will be displayed on the SRC.

IV. Positional Accuracy

The new SRC must be capable of meeting the existing positional accuracy requirements summarized in Table 1. Positional accuracy of the 64-m antenna subreflector is based on these capabilities and the calculated value of offset, due to gravity deformations. The calculated offset values reside in a look-up table which is part of the SRC software. Axial (z) offsets are calculated at 0.088-degree elevation angle intervals and lateral (y) offsets at 0.022-degree elevation angle intervals with an error tolerance of 0.25 mm (0.010 in.). The actual accuracy for z-axis and y-axis positioning was chosen to be ±1.25 mm (±0.050 in.), thus providing a limiting 0.5 dB gain loss. The "deadband" tolerance is wide compared to the positioning accuracy, and once the subreflector stops, it would take several seconds for the elevation angle change to require another subreflector position adjustment.

V. Summary

The Subreflector Controller (SRC) provides a method for correcting error offsets of the RF beam due to gravity-induced structural deformations. The SRC unit has been designed and tested at DSS 14, demonstrating the potential for increased large antenna performance.

References

1. Katow, M. S., "34-meter Antenna-Subreflector Translations to Maximize RF Gain," *TDA Progress Report 42-62*, Jet Propulsion Laboratory, Pasadena, California, January and February 1981, pp. 112-120.

2. Guiar, C. N., and Hughes, R. D., "NASA Tech. Brief on JPL Antenna Radiation Pattern Computer Program (ANRAD)," 1983.

Table 1. Performance capabilities of the subreflector and antenna drive electronics

Parameter	*Y* Axis	*Z* Axis
Maximum Axis Travel, mm (in.)	152.4 (6)	203.2 (8)
Maximum Rate of Travel, mm/s (in./s)	0.20 (0.008)	1.27 (0.050)
Maximum Error for 0.05 dB Gain Loss, mm (in.)	2.54 (0.100)	1.27 (0.050)
Maximum Movement/Degree Elevation, mm (in.)	20.3 (0.80)	5.60 (0.22)
Maximum Required Tracking Rate, mm/s (in./s)	0.50 (0.020)	0.15 (0.006)
Maximum Position Resolution, mm (in.)	0.062 (0.00244)	0.062 (0.00244)

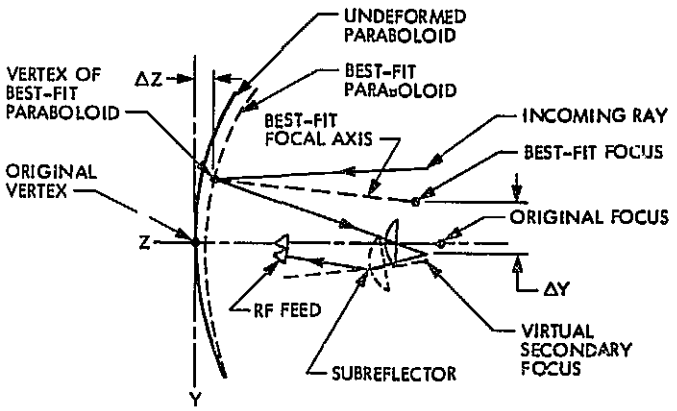


Fig. 1. RF center ray tracing and hyperboloid offsets with gravity distortions

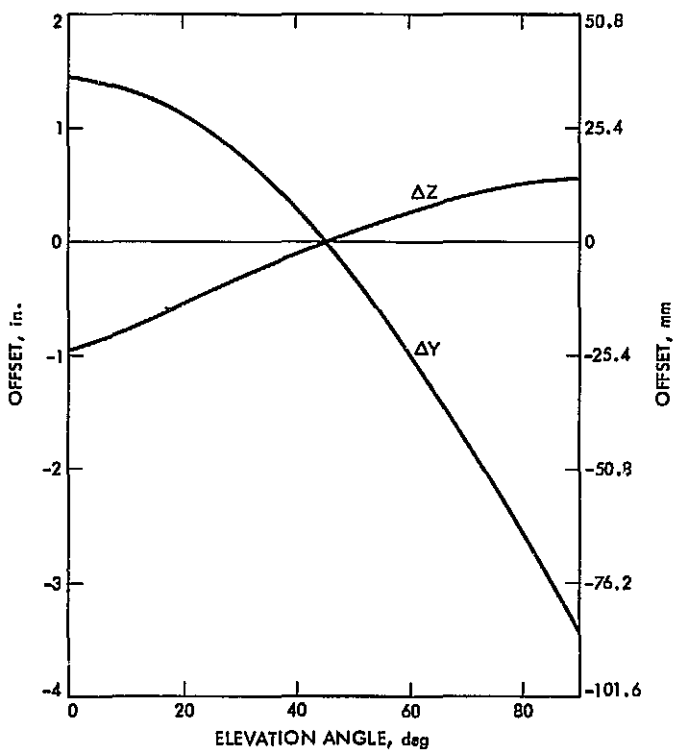


Fig. 2. Focus offset vs elevation angle

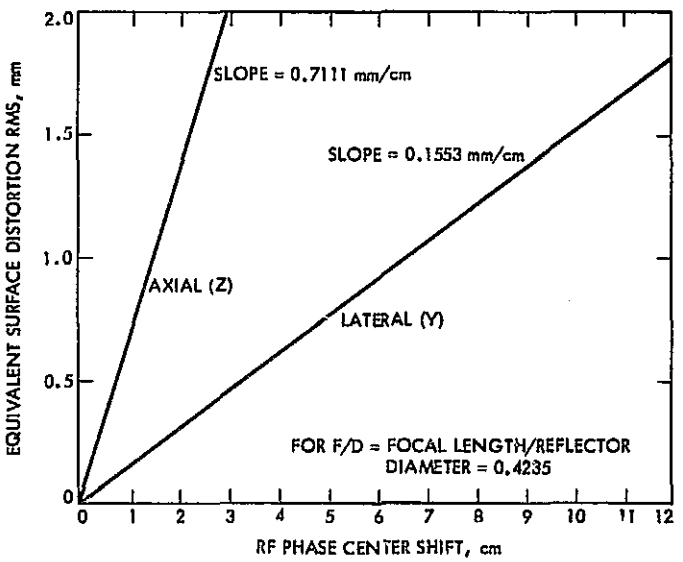


Fig. 3. Gain loss (Ruze equivalent surface distortion RMS) vs focus offsets for 64-m antenna at X-band

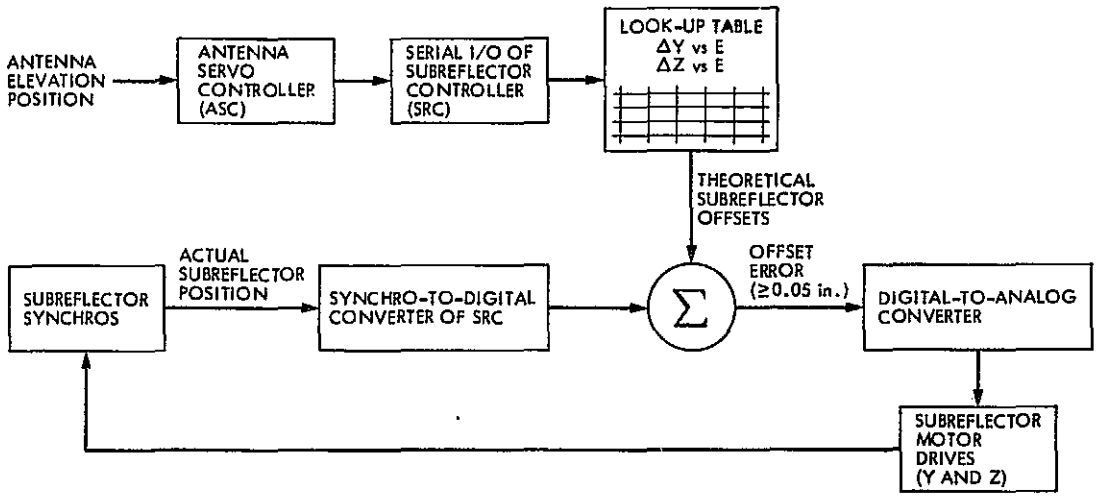


Fig. 4. Subreflector position servo block diagram

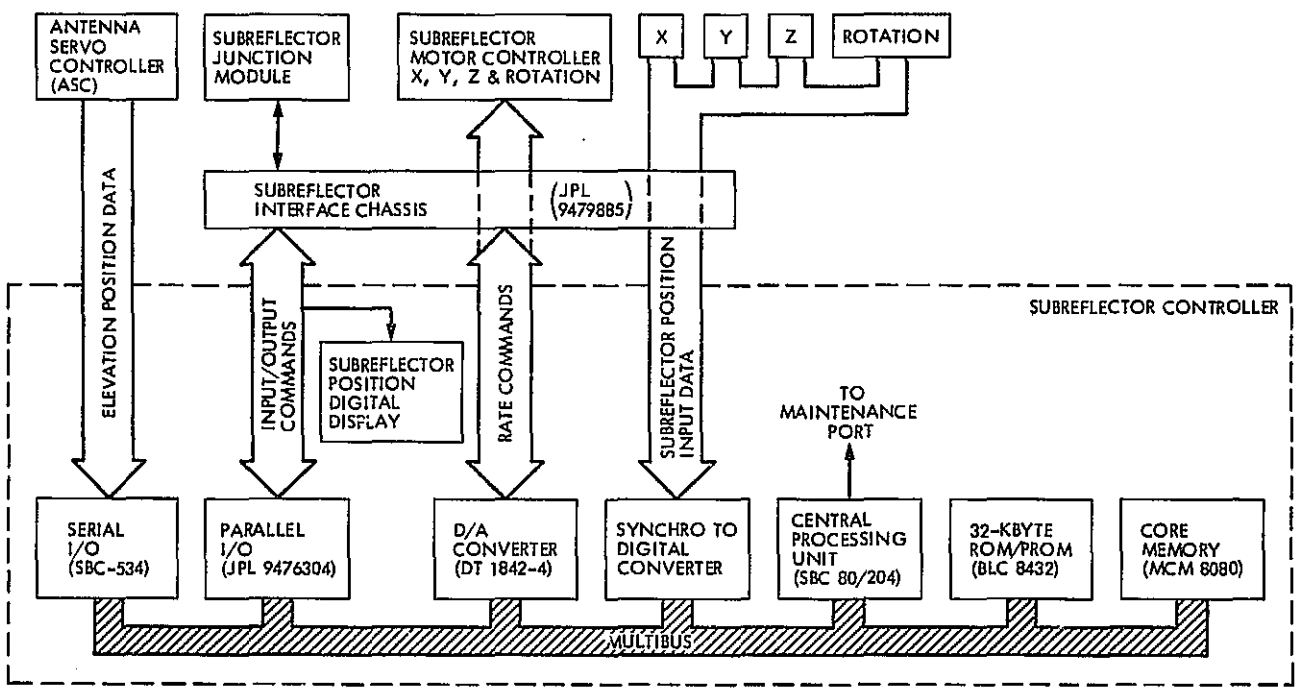


Fig. 5. Subreflector controller (SRC) and interface block diagram

Intermodulation Product Levels in Flame-Sprayed Materials

C. S. Yung, F. W. Stoller, and F. L. Lansing
Ground Antenna and Facilities Engineering Section

S. Brazil
Ford Aerospace and Communications Corporation, Palo Alto, California

Ford Aerospace (under contract with the Jet Propulsion Laboratory) completed a preliminary investigation on intermodulation product (IMP) levels of fiberglass-backed flame-sprayed surfaces. The purpose was to demonstrate the use of modified techniques and materials in combustion flame spraying of formed surfaces in reducing intermodulation products. The approach used improved metal wire stock without impurities or with smaller droplet sizes, used new high-temperature release agents, used wire stock with lower electrical resistance, and used variations in spraying distances, and intense buffing processes which amalgamate the material gaps and droplets in an effort to fabricate IMP-free light-weight and low-cost subreflectors. The study revealed positive material candidates with an IMP level around -150 dB, which is comparable to solid aluminum surfaces used as a reference.

I. Introduction

Ford Aerospace and Communications Corporation (FACC) (under contract with JPL) performed a study examining the intermodulation product (IMP) response characteristics of certain flame-sprayed materials and processing procedures. The purpose for this study was to demonstrate the feasibility of applying certain modified combustion flame-spray (metallizing) techniques and a number of materials for fiberglass-backed subreflector surfaces, used for the cassegrain microwave antennas, while avoiding the generation of impairing IMPs. The aim is to provide a method of producing at low cost, low IMP, and high efficiency, small-tolerance (RMS) and light-weight subreflectors. Material candidates were sought with characteristic IMP levels of -40 dB or below the antenna third sidelobe radiation level when radiated with multiple carriers at a power density approximating the antenna operating conditions. The study objective was to perform Lab tests on sample materials and fabrication procedures and to compare results versus conventional solid reflecting surfaces.

Past experience (Ref. 1) has shown that conventional flame-spray techniques using standard aluminum wire feed stock produce unacceptable IMP and noise problems when a multiple carrier uplink RF transmitter is used. Hence, a better approach is developed. This report describes the materials that were investigated, their method of fabrication, the test facility, the test procedure, and the test results and gives recommendations for future work.

II. IMP Mechanisms

The frequency F_i of a given intermodulation product resulting from two carriers of frequencies F_1 and F_2 ($F_1 < F_2$) is determined by the expression

$$F_i = NF_1 - MF_2 \tag{1}$$

where N and M are integers. The IMP order is defined as $N + M$. For example, the third, fifth, and seventh IMP orders can be written as follows.

$$\left.\begin{aligned} F_i(3) &= 2F_1 - F_2 \\ F_i(5) &= 3F_1 - 2F_2 \\ F_i(7) &= 4F_1 - 3F_2 \end{aligned}\right\} \tag{2}$$

Intermodulation product effects have been investigated since multiple frequency carriers were first introduced in space communications systems. Passive intermodulation products occur because some microwave components in radio-frequency systems, presumed to be linear, are in reality very slightly nonlinear. Transmit-to-receive isolations on the order of -150 to -200 dB are typically needed for high-power transmitter systems with sensitive receivers. For such systems, nonlinearities as little as 1 part/10^{10} may present a problem.

Three of the most predominant mechanisms for producing nonlinearties and intermodulation products are (1) electronic tunneling (a semiconductor action) through thin oxide layers separating metallic conductors at metallic junctions; (2) microdischarge between microcracks, whiskers, or across voids in metal structures; and (3) nonlinearities associated with dirt, metal particles, and carbonization on metal surfaces. Each of these different mechanisms manifests itself in identical power laws and in nearly equal levels of IMP generation.

The nonlinearities responsible for the IMP are a result of the summation of many different microcurrent conduction processes. Microscopically, all surfaces are highly irregular and have a surface oxide layer between particles or droplets several angstroms thick. When two or more surfaces (particles) come in contact, rupture spots through the oxide coatings are formed, and very thin oxide layers separate the metals.

The nonlinearity will depend on the proportion of the conductive and displacement currents. For metal surfaces separated by thin oxide layers, less than 50 Å (50×10^{-10} m) nonlinear electron tunneling occurs. For thicker oxide layers, semiconductor current flow can take place. At high-power levels, low-level water vapor, weak gaseous plasma, and nonlinear processes in the material come into play. The observed IMP currents are a result of the statistical summation of the microcurrents from many different nonlinear contacts.

A nonlinear device is one that does not obey Ohm's law, and the relation of current and voltage for such a device is a curve that can be represented by a polynominal of a degree higher than one over a finite interval. Analysis made by FACC shows that third order intermodulation power is the predominant contributor to IMP interferences. Relationships are derived to show the magnitude of the third order IMP as a function of the magnitude of the power of two or more signals applied to the nonlinear device. Where all signal voltages are constant, the IMP power, P_i, is shown to vary with the carrier power ratio R as

$$P_i = \frac{R}{(R+1)^3} \tag{3}$$

In summary, IMPs result from nonlinear junctions where two or more carriers of given power ratio exist simultaneously and where the power of third order frequencies of the interference is sufficient to cause interfering sources.

III. Background

The detrimental effects of IMPs to JPL radio telescopes was described during the Voyager space program (Ref. 1). Past experience on flame-sprayed surfaces has shown that the unacceptable intermodulation and noise products, produced by conventional flame spray techniques, have resulted from one or all of the following conditions:

(1) Some *impurities* in the wire feed stock were included, typically, standard aluminum wire which contains up to 10% silicon impurities. These impurities cause undesirable coating of the sprayed aluminum particles (flakes), which create multiple resistive cells with local eddy currents and noise effects.

(2) The spray particles may have significant *oxide coatings*, developed during flight from the spray gun to the desired surface, resulting in noise generation.

(3) The low-temperature oxide-forming and *nonuniform spray pattern* of spray particles create a porous surface that, when power illuminated, results in IMP generation.

(4) The use of *incorrect spraying distances* which affects the porosity and oxide coatings referred to in (3), above.

(5) The use of a low-temperature "release agent" (a coating used for separating surfaces in the simulated female molds) causes excessive outgassing and thereby creates

local resistive cells with their accompanying eddy currents and noise effects.

(6) The selection of wire materials relative to the size of the sprayed particles is important. Other wire materials with good electrical characteristics, such as copper, silver, tin, zinc-tin, etc., have finer particle diameters – which at S- and X-band frequencies have a significant improved effect on IM products.

(7) Conventional techniques do not use intense *buffing*. Since flame-spray particles are laid down in semiflat flakes and are oxide coated in random (creating resistive cells and a porous surface), the application of intensive buffing of the finished surface may amalgamate the surface flakes together. Buffing will break down the interparticle oxide barriers and minimize the amount of porosity – especially in softer metals.

IV. Study Approach

The study approach is to avoid the past pitfalls mentioned above and to investigate the feasibility of significantly reducing the IMPs in the combustion flame-spray process by using a combination of pure wire-feed materials, high-temperature release agents, finer particle sprays, proper spray distances, different types of wire feed, and the amalgamating effects of intense buffing. In the future, the plasma flame-spray technique may be investigated to determine whether it can be adapted to NASA and DSN needs and is not included in this study.

V. Description of Material Samples

A number of flat flame-spray samples were fabricated as an approximation to the curved female mold used for forming the subreflector surface. The sample size was 30.5 cm X 30.5 cm with a 0.15-cm thick fiberglass backing. The samples were prepared by Antenna Systems Inc. (ASI), San Jose, California. The wire-feed materials were flame sprayed against a flat mold whose surface had been prepared with a high-temperature release agent. After the simulated mold had been metal-sprayed to a thickness of approximately 0.025 cm, the fiberglass backing was applied to the flame sprayed surface. The backing material was bonded and cured, and the sample was then removed from the mold. The release agent was removed using acetone.

A total of thirteen material samples was prepared as described in Table 1. Following the examination, processing, and testing of the first seven samples, certain handling procedures, test methods, and test criteria evolved which suggested that more meaningful results could be obtained by altering the original methods for the remaining samples. Pure tin was deleted from the list since flame spraying of this material constituted a health hazard.

VI. Test Conditions

The test conditions were set to approximate the operating power level of JPL antennas. The peak power density incident on JPL subreflectors was given as 5.4 W/cm^2. A test was configured to provide that power density as a minimum plus any margin the test facility would provide. Since IMPs from the flame-spray samples were the primary concern, the test configuration was designed to minimize or eliminate IMP contributions from the facility itself.

A means for mounting and supporting the test samples was constructed that would not in itself contribute to the observed IMP level. The holding fixture was a polyvinyl chloride (PVC) frame which supports the sample at 0.635 cm from the radiating aperture. PVC is used to eliminate metallic surfaces which are known IMP generators. The sample was held in proximity to the horn aperture to reduce radiation into the anechoic absorber which covers the interior of the test facility. This too was to reduce the background IMP of the measurement system. Figure 1 shows the sample holding fixture located in place within the test facility.

The radiating aperture was a 12.7-cm diameter conical horn which connects to the remaining IMP test facility providing the radiation and monitoring system. The transmitter, receiver, and recording/monitoring system used were at FACC IMP test facility located in Palo Alto, California. This facility was built to test IMPs having a level of -170 dBm from components radiated at multiple carrier power levels of +63 dBm into a free space environment. An RF-shielded anechoic test facility is also included that permits measurements in excess of -150 dBm from those power sources. The transmitter is a high-power amplifier capable of delivering 5 kW of carrier power. This level of power is necessary to provide a minimum of 2 kW at the feed interface following distribution losses through the facility waveguide and monitoring equipments. The frequency bandwidth covers 7.90 to 8.40 GHz. Most of the tests were run at a power level of 1 kW. This provided a power density level of 6.6 W/cm^2, exceeding the minimum required for the study. Attempts were made to increase the power level; however, the level of reflected energy (caused by the waveguide short presented by the flame-spray sample) into the receiver bandpass filter causes excessive heating of that component.

The receiving system has a noise figure of 1.5 dBm (105 K). This is achieved by including a Field Effect Transistor (FET) before the downconverter and spectrum analyzer. Since

thermal noise power is -174 dBW/Hz at room temperature (290 K), a 10 Hz predetection bandwidth on the spectrum analyzer provided a theoretical noise power of -160 dBm. Also included in the test setup are a spectrum display and an equipment controller. These provided a swept display of all IMPs generated in the 7.25 to 7.75 GHz band. Those data were stored in computer memory, recorded on tape and by an X - Y recorder. The modulated carriers which are used to drive the high power amplifier (HPA), the down converter local oscillators, and the spectrum analyzer are stabilized to 5×10^{-11} by an oscillator. This provided long term or swept measurement stability. Between 7.75 to 7.90 GHz at least 100 dB of rejection was provided by a separate bandstop filter. This prevented energy from the traveling wave tube (TWT) HPA passing through this window that would generate an IMP in the field effect transistor (FET) amplifier. Procedures are included to identify and isolate IMPs generated within the HPA.

Separate screen rooms having greater than -80 dB isolation were used between the receive test area and transmit area. These rooms prevent "floating" signals of the transmitter from influencing the low-noise receiver information. Figure 2 shows a block diagram of the FACC test facility. All tests were performed at room (ambient) conditions.

The recorded IMP level observed at each setting of the carrier frequencies was taken from a statistical average of ten separate readings. This was done to eliminate peaks, nulls, and equipment variations. As a result of this averaging process, each frequency measurement occurs over a 10-minute time period, and each test sample measurement occurs over a 90-minute time period.

No special facilities were required to perform the sample buffing, surface resistance tests, or porosity tests. Conventional equipment was used for each of these. A Kelvin Resistance Bridge was used for the simple surface resistance measurements. A 30X power microscope and light source were used in judging the sample porosity.

VII. Test Results

The flame-spray samples were sequentially tested for surface resistance, checked for porosity, IMP tested, buffed, and then retested. Initially the samples were tested, then buffed using various buffing techniques, and then retested. The data herein are grouped according to the sample identification in Table 1.

A. Buffing Procedures

One of the study objectives was to determine whether the observed IMP level could be reduced by intense buffing which causes amalgamation of the metal. If the IMP is a result of the material porosity or if it is a result of oxidation between globules of metal formed during the flame-spray process, the IMP level should be reduced following the amalgamation process.

Several buffing methods were examined. The first used a horse-hair brush wheel operated at very high speed. It was found that the heat developed from the brush was not adequate to cause distortion of the metal surface and no amalgamation occurred. The second attempt was to place the metal surface against a granite lapping plate. An orbital sanding machine was modified to accept the material sample during this process. Again, sufficient heat to cause amalgamation could not be developed, probably due to the heat sink effect of the granite block.

As a third attempt at buffing the flame-spray samples, a hard leather disk was fabricated and used on edge, similar to the horse-hair brush. This method proved to be too harsh and abrasive. The disk edge, which was approximately 1.2-cm thick, removed some metal from the sample, thus leaving large, unacceptable voids. The leather disk was then applied flat, using an automotive-type polishing machine. This method was marginally acceptable for certain metal surfaces. The pure aluminum (sample number 2, Table 1) was partially amalgamated by the leather-disk buffing process. Materials having a coarse surface (copper and silver, sample numbers 4 and 5) tended to load up the leather disk, thus significantly impairing the process. The pure zinc and tin/zinc materials (sample numbers 4 and 6) exhibited hard surfaces and large areas flaked off during the buffing process.

On examination of the above three buffing processes it was recognized that none of them yielded the desired result and a new trial was needed. A dry-lubricant Silicon Carbide polishing disk was tried, and it was found that grade-80 polishing agent provided significant improvement to the surface smoothness and its porosity and seemed to approach amalgamation somewhat more than the leather disk. The processing time was also significantly reduced. This buffing method was used on the remaining six samples (8 through 13).

B. Porosity Check

Samples 8 through 13 were checked before and after buffing for granularity and porosity. This examination was made by viewing the metallic surface in a darkened room while holding a high-intensity constrained light source to its fiberglass surface. The degree of porosity was a subjective judgment made by the test conductor with ranking from 1 to 5 (1 = no light showing through; 5 = significant light showing through). Porosity was helpful in determining the effectiveness of the buffing process. The results are shown in Table 2.

C. Surface Resistance Tests

Each sample was measured to determine its DC surface resistance before and following the buffing process. A Kelvin Resistance Bridge was used for these tests by placing two electrodes on the surface of the sample, spaced apart by a constant distance. The results are tabulated in Table 3 for comparative evaluation before and after buffing only. In general, materials having low resistance contribute lower system noise temperature. However, these measurements do not represent the actual dissipative loss associated with the metallic surface as described in Refs. 2 and 3[1]. A more accurate measurement technique such as the cavity resonator technique was developed by R. Clauss and P. Potter (Ref. 3).

D. Reflection Tests

One area of concern was the reflective quality of the samples following the buffing process. A test was added to determine the surface reflection to RF energy. A microwave reflectometer using an Automatic Network Analyzer was used to monitor reflected energy from an open-end waveguide placed flush to the sample surface. Table 4 lists the results of samples 8 to 13 before and after buffing for comparative evaluation only. Note that for a perfect reflector, the reflection coefficient should be zero. A different reflection measurement technique wherein each sample is placed at 45° slope from the radiating horn is contemplated for future work.

E. Intermodulation Products (IMP) Tests

The IMP results acquired from the measurements described in Section VI are plotted as a function of the receive frequency in Figs. 3-5. Each graph includes a frequency distribution. Fifteen carrier pairs were tested with carrier 1 frequencies ranging from 7.900 to 8.075 GHz and carrier 2 between 8.050 to 8.400 GHz. IMP 3rd order frequencies ranged from 7.40 to 7.75 GHz. An analysis of the results is given below.

1. System calibration. A reference calibration demonstrated the background IMP of the facility, holding fixture, and test equipment. The source horn was radiating into the anechoic room without a test sample plate in position. The calibration demonstrated (Fig. 3[a]) a background IMP level averaging -171 dBm over the receive frequency band.

Second, reference calibrations were made to demonstrate the inherent IMP level of the measurement system when a reference test plate (of known IMP purity) was substituted for the test samples. Two aluminum and copper plates, with dimensions identical to the flame-spray test samples, were mounted in the holding fixture and their IMP responses were measured. This calibration revealed (Figs. 3[b] and 3[c]) a nominal increase in the baseline IMP level. The increase is attributed to two factors: first, a concentration of incident energy on the chamber absorber in the proximity of the horn; and second, a significantly increased reflected energy into the monitoring microwave components (horn, orthomode junction, filters). Since IMPs increase at a rate of 3 dB per dB of incident power (Ref. 4), the reflected power level through the monitoring RF components is increased several fold when the calibration plates and the flame-spray sample plates are positioned over the horn aperture. Thus, the observed system IMP level increase in the calibration level was expected. The calibration data shown in Fig. 3 are accurate within ±5 dB of the indicated values. Thus, the two copper and aluminum plates exhibited generally comparable IMP signatures (~ -160 dBm).

2. Test samples. The results of the first seven flame-spray samples are shown in Fig. 4 where the general IMP level is observed between -100 and -170 dBm. Standard aluminum flame spray (sample 1, Fig. 4[a]) shows the largest IMP (-114 dBm); however the high-temperature release agent improves its performance significantly. The IMP performance of standard aluminum flame spray with high temperature release agent (Fig. 4[g]) is comparable (at -137 dBm) to the general trend of the other samples. IMP performance was plotted before and after buffing. When the 10 dB tolerance window (±5 dB) is considered, it was apparent that the samples performed generally the same except for the "standard" aluminum with low-temperature release agent (Fig. 4[a]). Pure zinc (Fig. 4[f]) and zinc-tin (Fig. 4[c]) appear as slight favorites within each group (~ -132 dBm).

The results of samples 8 through 13 are shown in Fig. 5. During the tests on these samples, the transmitter/receiver control software was modified (wherein each frequency was repeated ten separate times), and the transmitter power output was set at a constant 1 kW. The results show a closer correlation between sample tests than was observed from the previous seven samples and show consistent improvement in the measured IMP level when the buffing process was used although no significant change can be attributed to the double buffing process.

The measured tolerance was reduced to ±2.5 dB on samples 8 through 13 due to improved repeatability of measurements. This also is attributed to the modified measurement method.

The results of Fig. 5(a) also show that the 70/30 tin-zinc (at -134 dBm) performs better following the buffing procedure. Buffing apparently affected the tin-zinc (improves IMP

[1]See also C. W. Choi and G. S. Kirkpatrick, "Surface Resistivity Measurements for the JPL 34-m X-Band Antenna," Harris Corporation, Melborne, Fla., September 9, 1980.

from -134 dBm to -155 dBm) far more than any of the other samples tested. The pure aluminum sample in Fig. 5(b) measured low initial IMP (-156 dBm) and exhibited a lower value (average -158 dBm) following the second buffing procedure (with dry-lubricant silicon followed by leather disk).

The METCO Babbit™ sample in Fig. 5(c) and the pure aluminum with a 70/30 tin-zinc subsurface sample in Fig. 5(d) show generally the same mixed trend as the pure aluminum.

The two remaining samples (No. 12 and 13 in Figs. 5[e] and 5[f]) having overspray and material buildup to double thickness exhibited poor IMP response after buffing. No particular reason could be found for their higher IMP measurement.

VIII. Summary

It was learned that as a general class of materials, flame-sprayed fiberglass-backed laminates can be considered viable candidates as reflector surfaces in high-intensity microwave multiple carrier systems without excessive interference due to secondary emissions from intermodulation products. This rather board conclusion is supported by the overall low level (between -150 and -160 dBm) of IMPs observed from the present tests.

The flame-spray materials were exposed to a radiation density in excess of 6.6 W/cm^2 with selected tests exceeding 13 W/cm^2. Some materials (tin-zinc and pure aluminum) exhibited IMPs of -150 dBm when measured over a 6.3 percent frequency band. These levels are generally considered satisfactory for most receiving systems yet may be marginal for stringent JPL systems.

The study shows that buffing the flame-spray surface for some samples improved IMP performance; however, the results were mixed. General improvements of 10 dB were noted and may reach 20 dB (such as using tin-zinc in Fig. 5[a]). The "best" buffing process found was light polishing with a dry-lubricant silicon fine grit paper. Further buffing with hard leather provided only nominal improvement. Microscopic examination of the flame-spray material before and after the buffing processes revealed no conclusive evidence that the material had undergone amalgamation; the improvement may result from rupture of the oxide coating around the spray particles. Porosity tended to increase only slightly following the buffing process.

Taking the norm of multiple recordings of the IMP level at each frequency set improved the repeatability and thus the accuracy of the data. Setting the power at a constant level removed the 3 dB/dB slope from the data curve.

The study results showed a definite feasibility for some flame-spray materials for DSN antenna subreflectors. From an IMP consideration, the tested materials reveal positive candidates for that application. The selection of a specific material, however, should not be made without further technical and economical investigations. Until the observed IMP level from a sample material can be maintained at a value below the measurement system noise level, there is room for improvement. Indeed, as the material technology is enhanced, as was experienced during the current study, improvements can be made in the measurement methods to demonstrate even lower levels of performance. The areas envisioned for further work include the following:

(1) Plasma spray metallizing techniques result in surfaces whose characteristics may avoid many of the known sources of IMP generation. The process uses a combination of high-energy, high-velocity gas with an inert gas carrier to develop high-density, oxide-free, nonporous metallic surfaces.

(2) In fabricating a scaled subreflector, with a hyperbolic surface and testing for its IMP performance, material candidates should include tin-zinc and pure aluminum. In addition, comparative microwave reflective efficiency measurements should be made on the subreflector using a standard solid aluminum unit as a reference.

(3) The effect of local environment on selected candidate materials should also be considered prior to a final material selection. Weather effects, such as rain, ice, hail, humidity, dust, high- and low-temperature cycling, should be determined.

References

1. Bathker, D. A., Brown, D. W., and Patty, S. "Single and Dual Carrier Microwave Noise Abatement in the Deep Space Network," Jet Propulsion Laboratory Technical Memorandum 33-373. August 1975.

2. Thom, E. H., and Otoshi, T. Y., "Surface Resistivity Measurements of Candidate Subreflector Surfaces," *TDA Progress Report 42-65*, Jet Propulsion Laboratory, July/August 1981, pp. 142-150.

3. Clauss, R., and Potter, P. D., "Improved RF Calibration Techniques: A Practical Technique for Accurate Determination of Microwave Surface Resistivity," *Technical Report 32-1526*, Vol. XII, Jet Propulsion Laboratory, Dec. 1972, pp. 59-67.

4. Chapman, R. C., Rootsey, J. V., Polidi, L., and Davison, W. W., "Hidden Treat: Multicarrier Passive Component IM Generation," AIAA/CASI 6th Communications Satellite Systems Conference, April, 1976. Montreal, Canada, Paper No. 76-296.

Table 1. Flame-spray samples

Sample Number	Description
1	Wire spray using "standard" aluminum wire (with 10% impurities) and a low-temperature release agent
2	Wire spray using pure aluminum wire, high-temperature release agent and no buffing
2(A)	Same as Sample 2 with buffing
3	Wire spray using pure zinc–tin wire, with high-temperature release agent
3(A)	Same as Sample 3
4	Wire spray using pure copper wire with high-temperature release agent
4(A)	Same as Sample 4 with buffing
5	Wire spray using pure silver, with high-temperature release agent
5(A)	Same as Sample 5 with buffing
6	Wire spray using pure zinc wire with high-temperature release agent
6(A)	Same as Sample 6 with buffing
7	Same as Sample 1 except using a high-temperature release agent
7(A)	Same as Sample 7 with buffing
8	A composite of tin/zinc in a 70/30 mix, with high-temperature release agent
8(A)	Same as Sample 8 after buffing
9	Repeat of Sample 2 above (pure aluminum) for repeat of measurements
9(A)	Repeat of Sample 2(A) above (pure aluminum) after buffing for repeat of measurements
10	Wire spray using METCO Babbit™ with high-temperature release agent
10(A)	Same as Sample 10 after buffing
11	Wire spray of pure aluminum as the surface area with a 70/30 tin–zinc as a subsurface backup material, with high-temperature release agent
11(A)	Same as 11 after buffing
12	Same as Sample 8 composite (70/30 tin–zinc) except flame-spray surface material thickness was increased to approximately 0.05 cm using an overspray
12(A)	Same as Sample 12 after buffing
13	Same as Sample 2 except flame-spray surface material thickness was increased to approximately 0.05 cm using an overspray
13(A)	Same as Sample 13 after buffing

Table 2. Porosity of some samples

Sample number	Material	Porosity	
		Before buffing	After buffing
8	70/30 tin/zinc	3	5
9	Pure aluminum	4	4
10	METCO Babbit™	5	4
11	Pure aluminum (with 70/30 backup)	2	2
12	70/30 (overspray)	1	1
13	Pure aluminum (overspray)	1.5	4

Table 3. Bridge Resistance of samples

Sample Number	Material	Resistance, ohms	
		Before buffing	After buffing
1	"Standard" aluminum (low-temp. agent)	0.0102	0.0114
2	Pure aluminum (high-temp. agent)	0.0127	0.0133
3	70/30 tin/zinc (high-temp. agent)	0.0038	0.0037
4	Copper (high-temp. agent)	0.0010	0.0010
5	Silver	0.0029	0.0028
6	Zinc (high-temp. agent)	0.0061	0.0065
7	"Standard" aluminum (high-temp. agent)	0.0150	0.0154
8	70/30 tin-zinc	0.0113	0.0105
9	Pure aluminum	0.0063	0.0065
10	METCO Babbit ™	0.0053	0.0051
11	Pure aluminum	0.0055	0.0060
12	70/30 (with overspray)	0.0033	0.0034
13	Pure aluminum (with overspray)	0.0004	0.0041

Table 4. Surface Reflection of flame-spray samples

Sample number	Material	Frequency, GHz	Reflection loss coefficient, dB	
			Before buffing	After buffing
–	Copper reflection plate	8.1	0.06	–
		8.4	0.01	–
8	70/30 tin/zinc	8.1	0.11	0.03
		8.4	0.11	0.01
9	Pure aluminum	8.1	0.43	0.08
		8.4	0.40	0.07
10	METCO Babbit ™	8.1	0.24	0.08
		8.4	0.36	0.04
12	70/30 tin/zinc with overspray	8.1	0.17	0.08
		8.4	0.09	0.08
13	Pure aluminum with overspray	8.1	0.23	0.02
		8.4	0.42	0.01

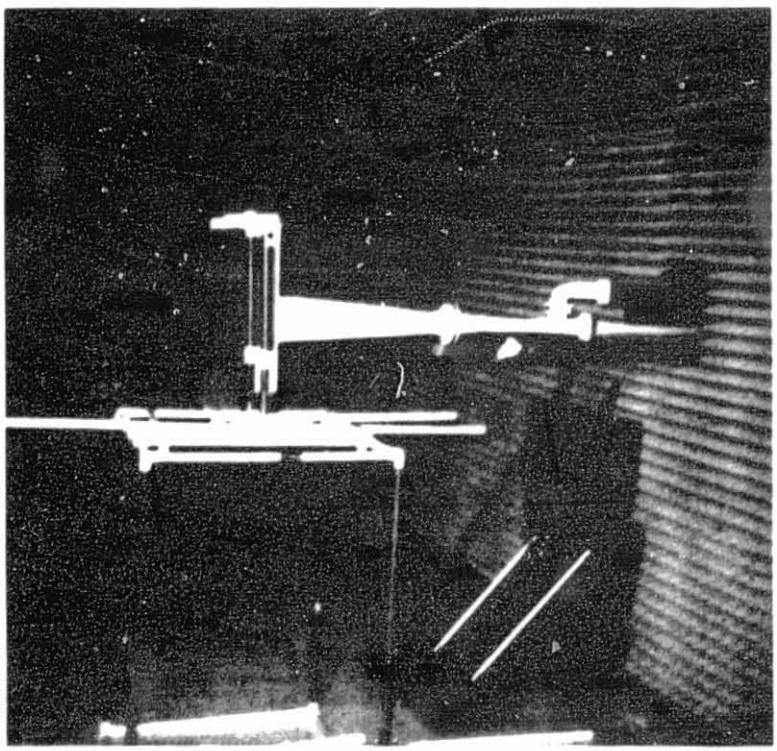
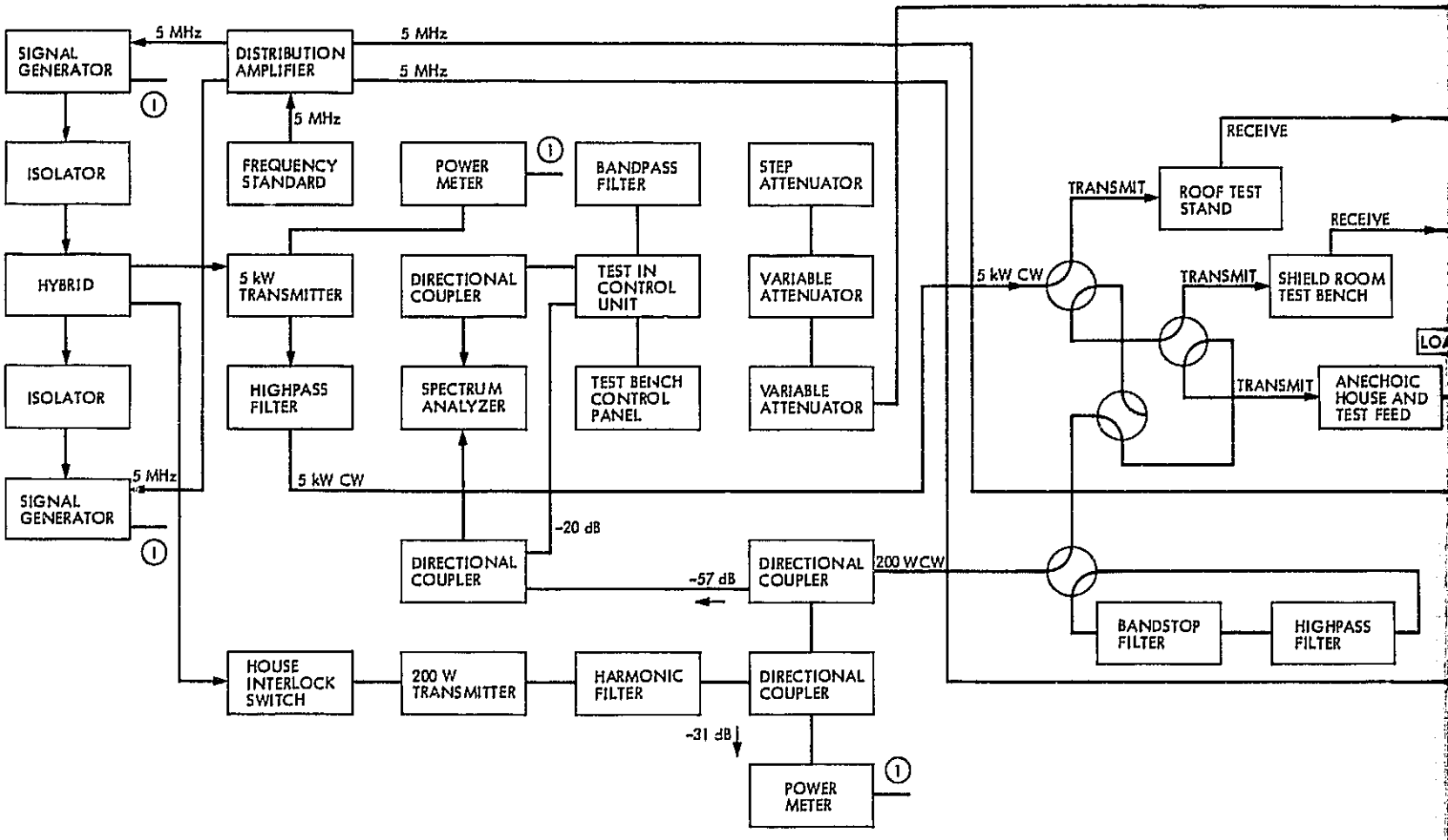


Fig. 1. Flame-spray test sample in anechoic chamber



FOLDOUT FRAME

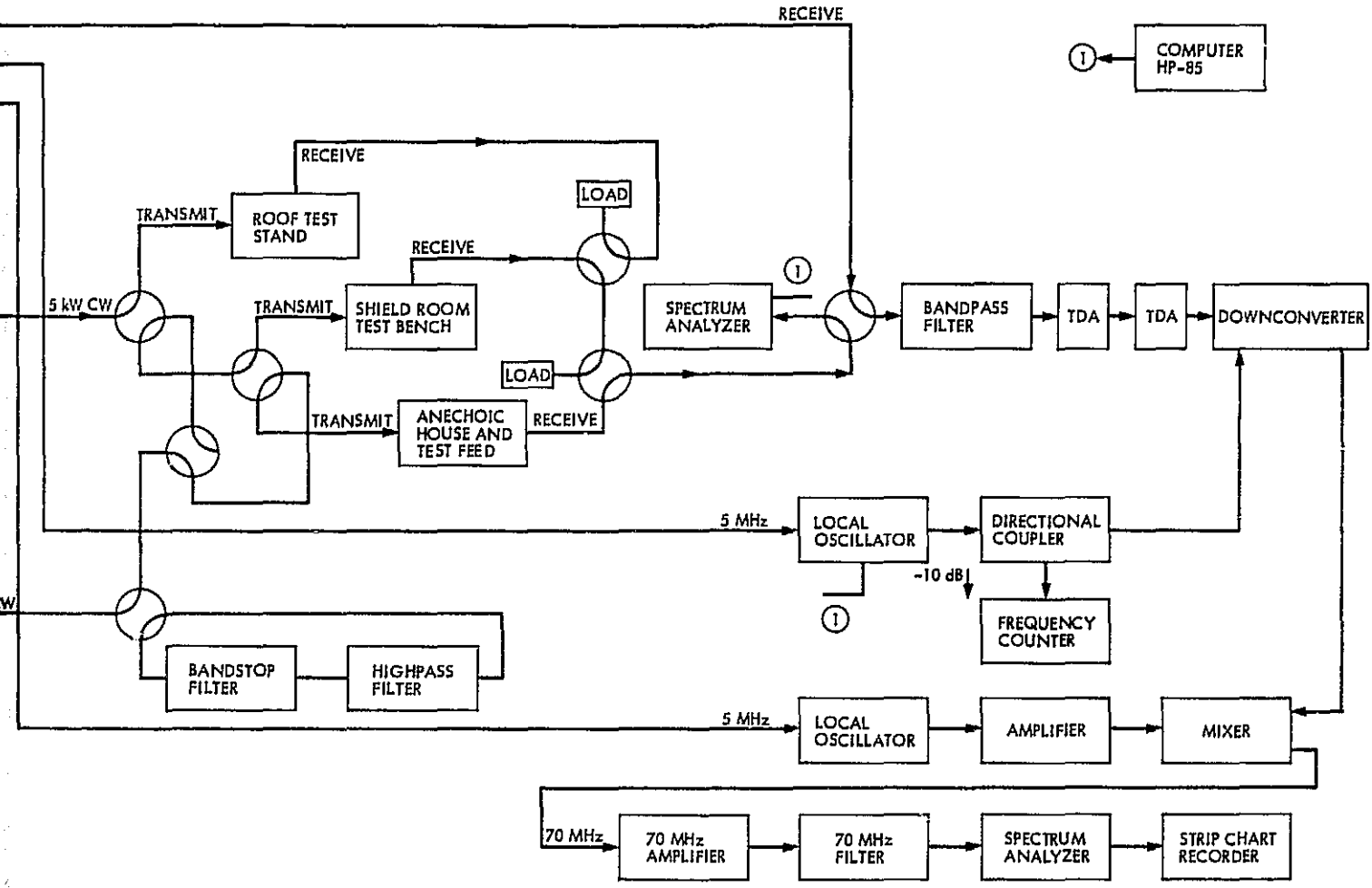


Fig. 2. IMP test facility at FACC

FOLDOUT FRAME

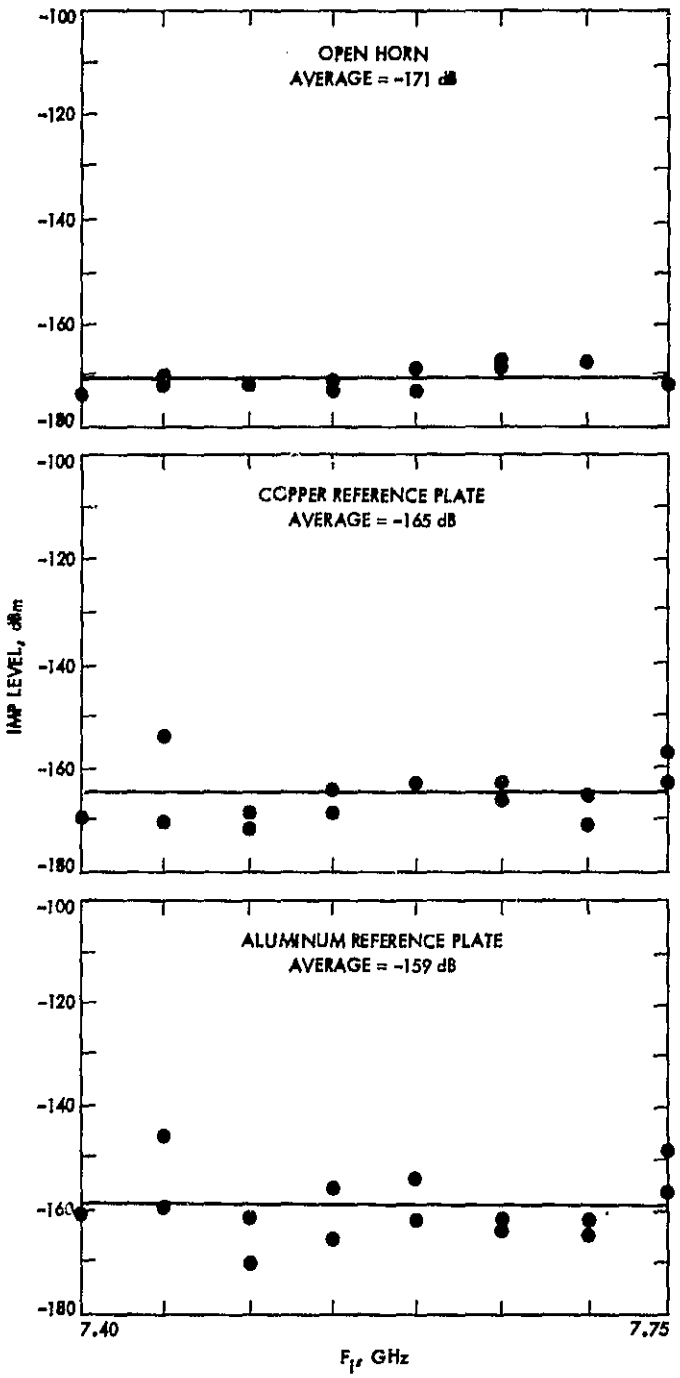


Fig. 3. IMP levels for facility and calibration plates

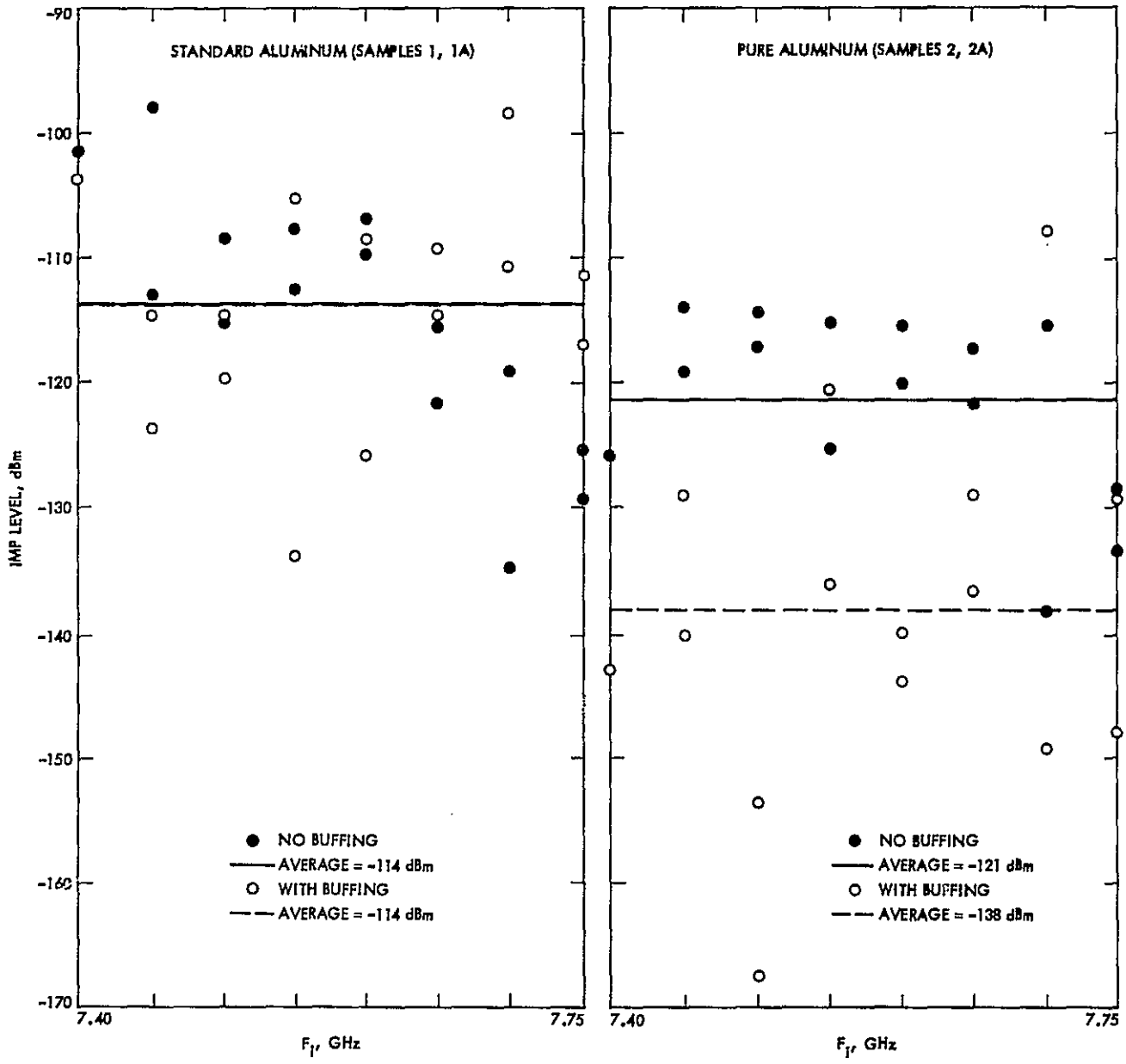


Fig. 4. IMP levels for samples 1 through 7 before and after buffing

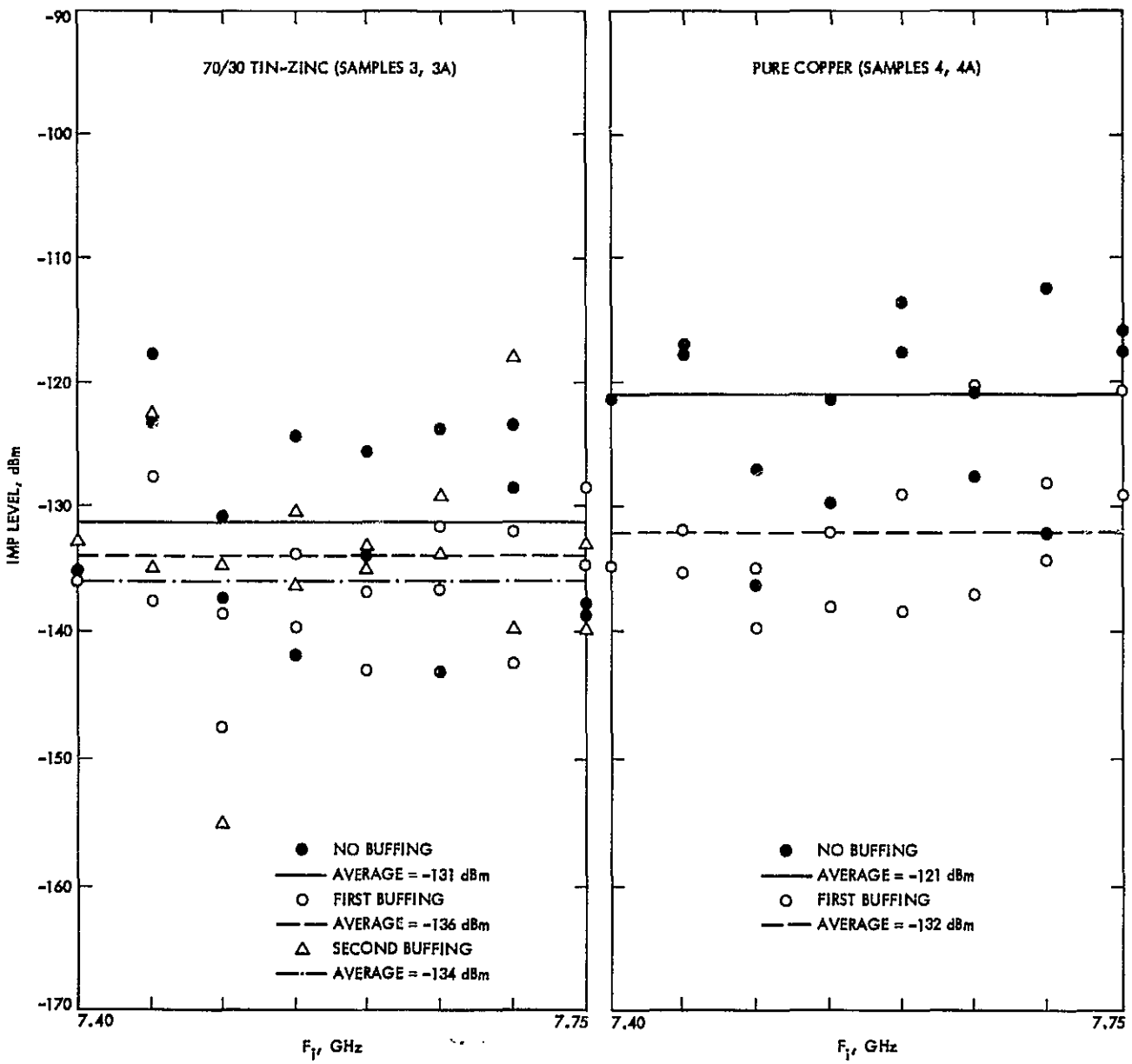


Fig. 4 (contd)

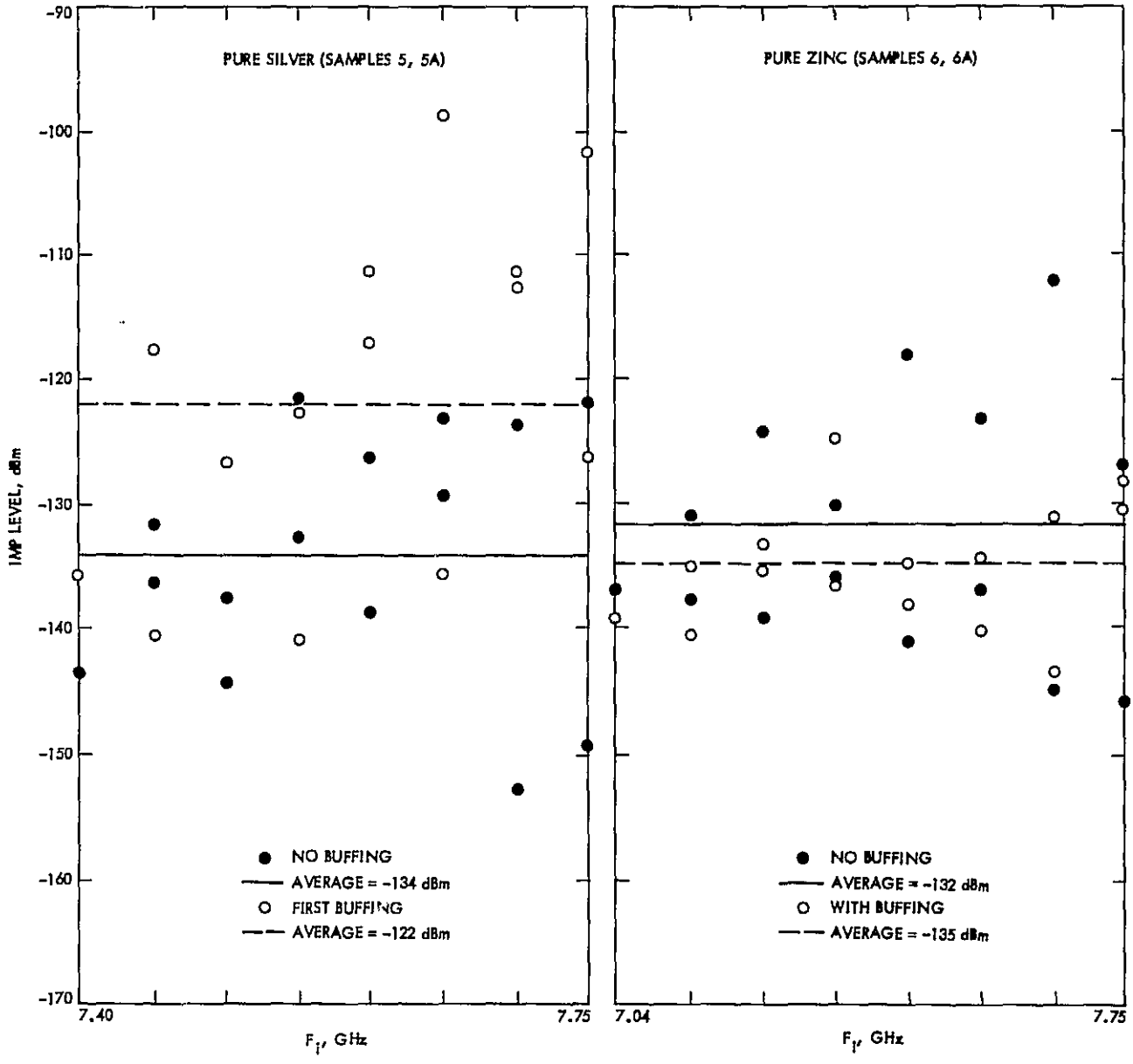


Fig. 4 (contd)

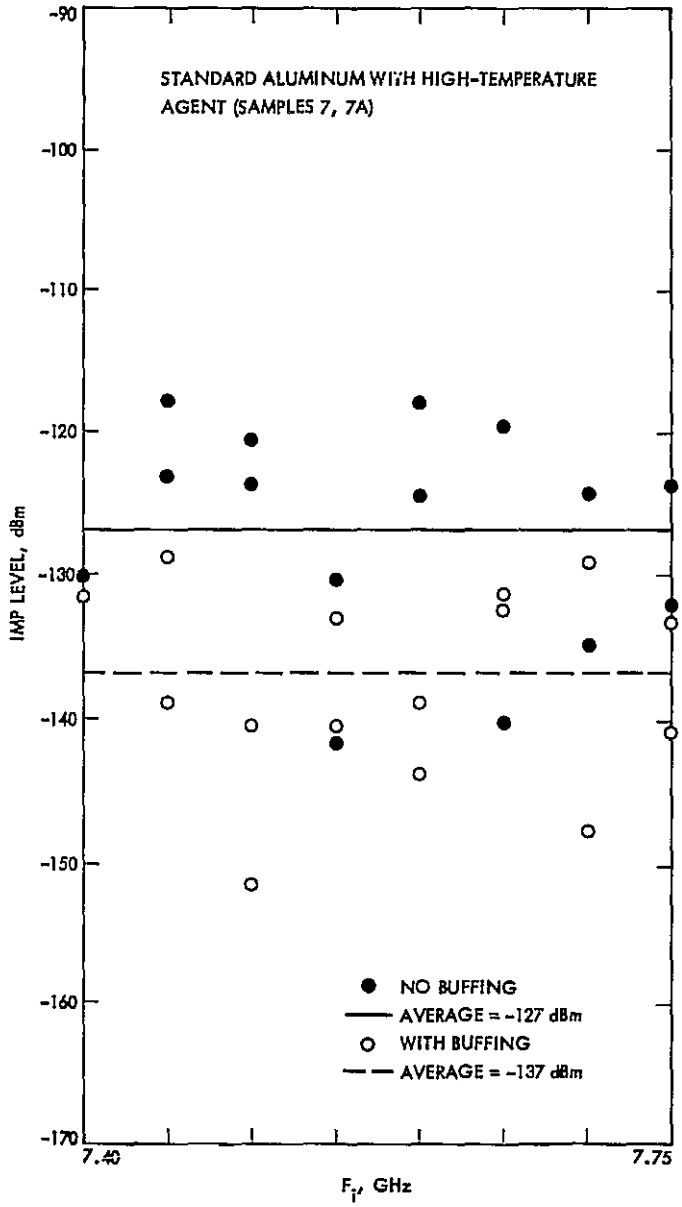


Fig. 4 (contd)

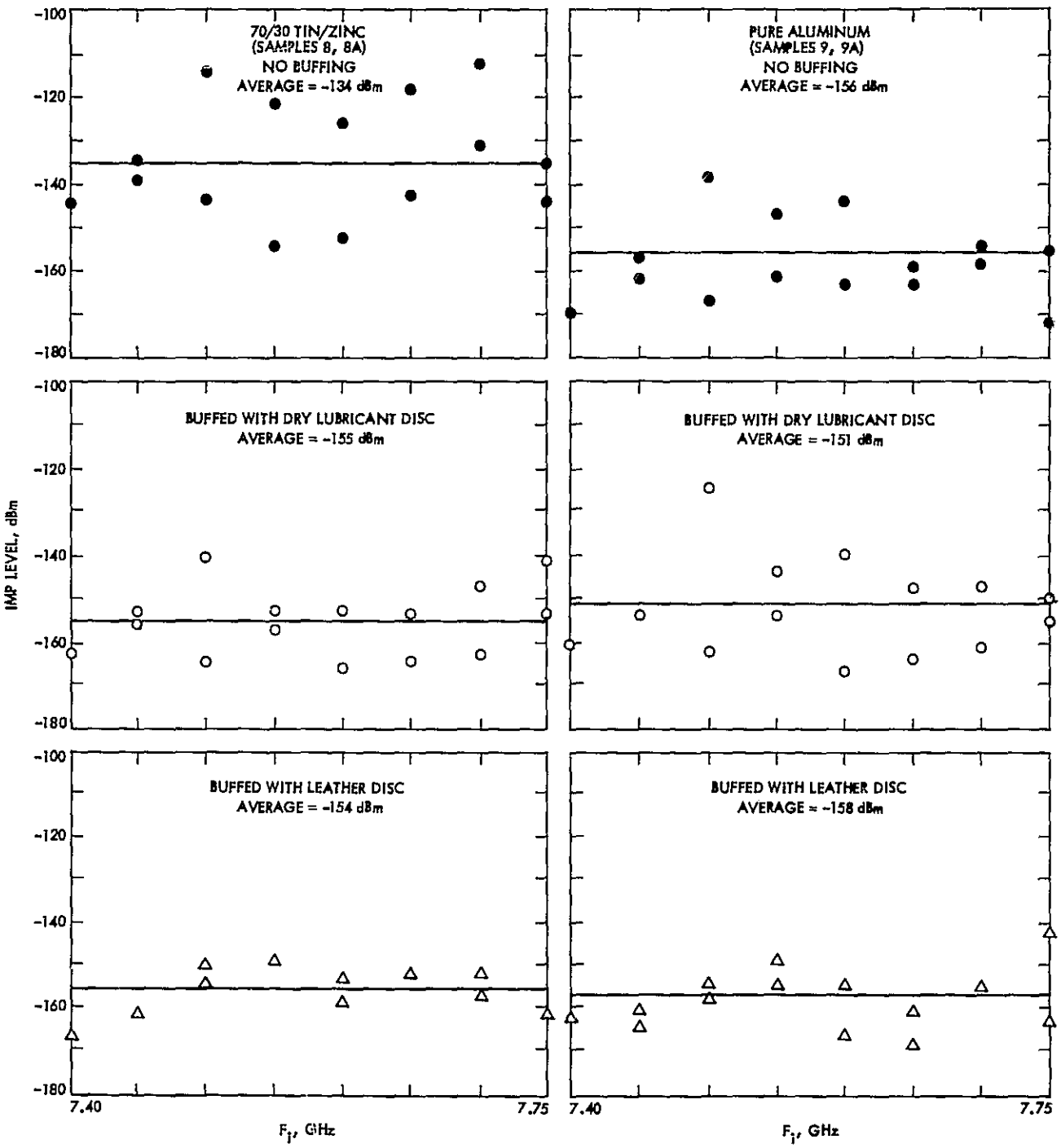


Fig. 5. IMP levels for samples 8 through 13 before and after buffing

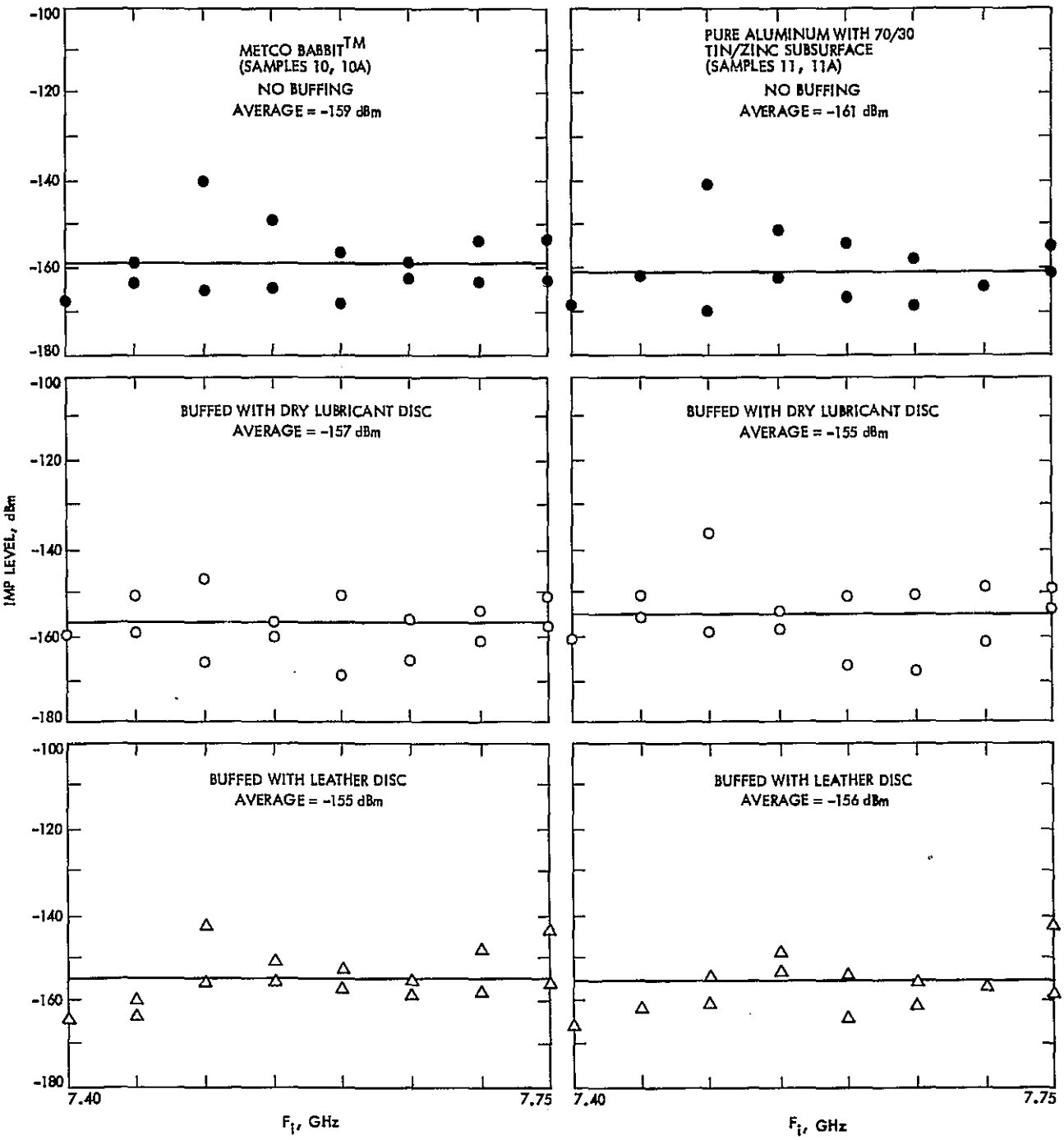


Fig. 5 (contd)

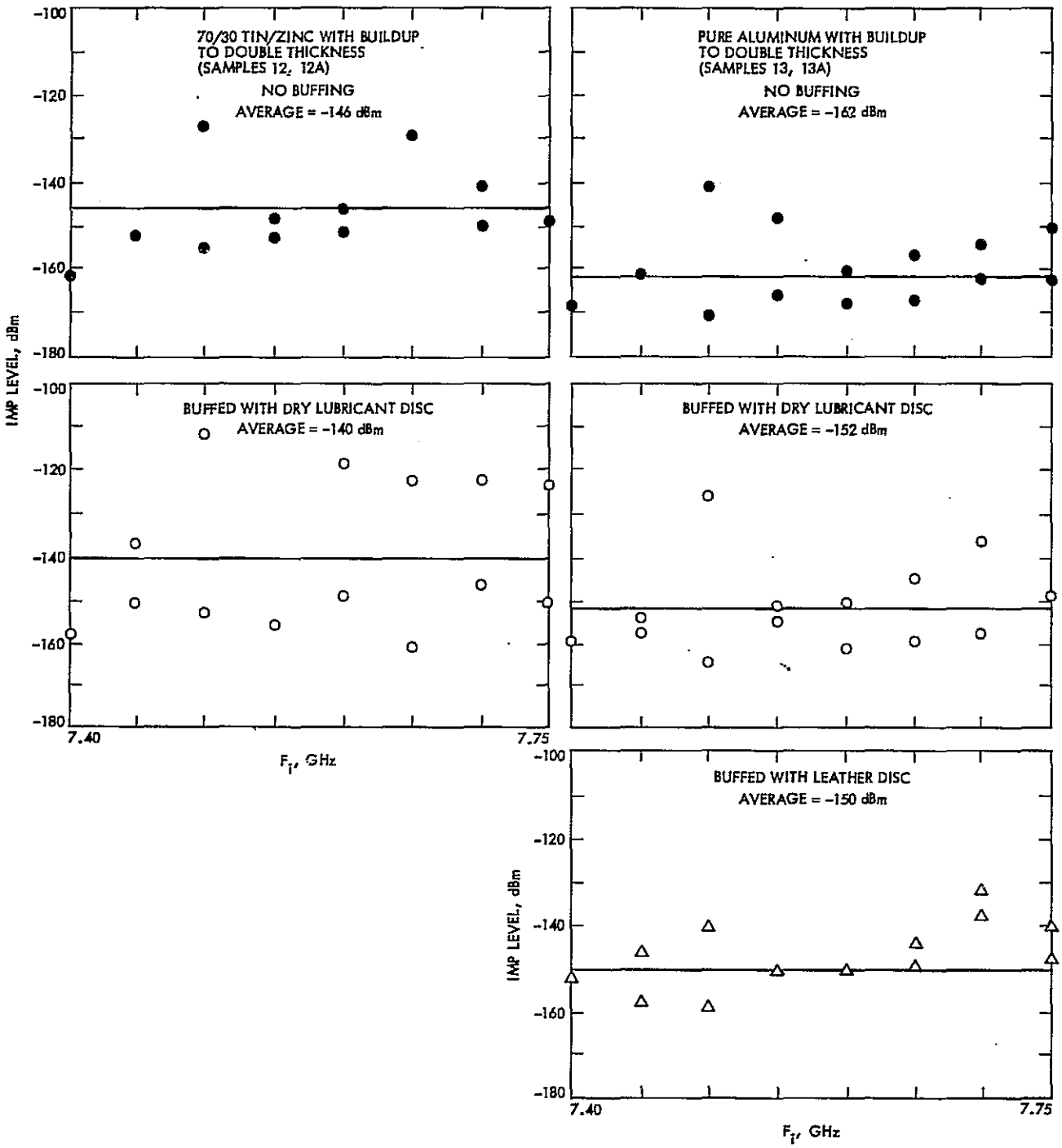


Fig. 5 (contd)

Performance Simulation for Unit-Memory Convolutional Codes With Byte-Oriented Viterbi Decoding Algorithm

Q. D. Vo
Communications Systems Research Section

This article describes a software package developed to simulate the performance of the byte-oriented Viterbi decoding algorithm for unit-memory (UM) codes on both 3-bit and 4-bit quantized AWGN channels. The simulation was shown to require negligible memory and less time than that for the RTMBEP algorithm, although they both provide similar performance in terms of symbol-error probability. This makes it possible to compute the symbol-error probability of large codes and to determine the signal-to-noise ratio required to achieve a bit error rate (BER) of 10^{-6} for corresponding concatenated systems. A (7, 10/48) UM code, 10-bit Reed-Solomon code combination was found to achieve the required BER at 1.08 dB for a 3-bit quantized channel and at 0.91 dB for a 4-bit quantized channel.

I. Introduction

A general $(l_0, k_0/n_0)$ unit-memory (UM) convolutional encoder is shown in Fig. 1. Let $\mathbf{a}_t$ be the k_0-bit byte of input to be encoded at time t, $\hat{\mathbf{a}}_{t-1}$ be the l_0-bit byte of delayed input, and $\mathbf{b}_t$ be the corresponding n_0-bit byte of encoded output. Let G_0 and G_1 be encoding matrices with dimensions $k_0 \times n_0$ and $l_0 \times n_0$ respectively, then the encoding equation may be written as

$$\mathbf{b}_t = \mathbf{a}_t G_0 + \hat{\mathbf{a}}_{t-1} G_1; \qquad t = 1, 2, \cdots$$

There are two different decoding algorithms that exhibit similar performance: the RTMBEP and the byte-oriented Viterbi. The RTMBEP decoding rule, which has been previously simulated (Ref. 1), has as its estimate $\mathbf{a}_t^0$ the value of $\mathbf{a}_t$ that maximizes $P(\mathbf{a}_t | \mathbf{r}_{[1, t+\Delta]})$ where $\mathbf{r}[1, t + \Delta]$ is the observed sequence with delay Δ. To speed up the simulation, we set up probability matrices $P(\mathbf{r}_{t+i} | \mathbf{b}_{t+i})$, where $i = 0, 1, \cdots, \Delta$. As a result, the required memory is at least $(\Delta + 1)2^{l_0+k_0}$, which is quite large for big codes. For example, for $l_0 = 9$, $k_0 = 10$, and $\Delta = 8$, at least 4,718,592 real numbers are needed. On the other hand, the simulation for the byte-oriented Viterbi decoding algorithm requires practically no memory (too small to count) since it does not have to store the matrix $P(\mathbf{r}_t | \mathbf{b}_t)$. Furthermore, the Viterbi algorithm itself is much simpler and hence runs faster than the RTMBEP algorithm.

II. Byte-Oriented Viterbi Decoding Algorithm

The byte-oriented Viterbi decoding rule chooses its estimated sequence $\mathbf{a}_1^0, \mathbf{a}_2^0, \cdots, \mathbf{a}_I^0$ to be the value of $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_I$, which maximizes

$$P(\mathbf{a}_1, \cdots, \mathbf{a}_I | \mathbf{r}_1, \cdots, \mathbf{r}_I) = \frac{P(\mathbf{r}_1, \cdots, \mathbf{r}_I | \mathbf{a}_1, \cdots, \mathbf{a}_I) P(\mathbf{a}_1, \cdots, \mathbf{a}_I)}{P(\mathbf{r}_1, \cdots, \mathbf{r}_I)}$$

Here we assume that all information sequences are equally likely (i.e., $P(\mathbf{a}_t) = 2^{-k_0}$) so that the algorithm is the same as maximizing

$$P(\mathbf{r}_1, \cdots, \mathbf{r}_I | \mathbf{a}_1, \cdots, \mathbf{a}_I)$$

or

$$P(\mathbf{r}_1, \cdots, \mathbf{r}_I | \mathbf{b}_1, \cdots, \mathbf{b}_I (\mathbf{a}_1, \cdots, \mathbf{a}_I))$$

A recursive method is developed as follows. Let $\mathbf{a}_t = \tilde{\mathbf{a}}_t, \hat{\mathbf{a}}_t$ where the comma denotes concatenation of $(k_0 - l_0)$-bit byte $\tilde{\mathbf{a}}_t$ with l_0-bit byte $\hat{\mathbf{a}}_t$. $\hat{\mathbf{a}}_t$ is called the state at time t since it affects the output at time $t+1$. Let

$$f(\hat{\mathbf{a}}_{t-1}) = \max_{\mathbf{a}_1, \cdots, \tilde{\mathbf{a}}_{t-1}} P(\mathbf{r}_1, \cdots, \mathbf{r}_{t-1} | \mathbf{a}_1, \cdots, \tilde{\mathbf{a}}_{t-1}, \hat{\mathbf{a}}_{t-1})$$

Then

$$f(\hat{\mathbf{a}}_t) = \max_{\mathbf{a}_1, \cdots, \tilde{\mathbf{a}}_{t-1}, \hat{\mathbf{a}}_{t-1}, \tilde{\mathbf{a}}_t} P(\mathbf{r}_1, \cdots, \mathbf{r}_t | \mathbf{a}_1, \cdots, \tilde{\mathbf{a}}_{t-1}, \hat{\mathbf{a}}_{t-1}, \tilde{\mathbf{a}}_t, \hat{\mathbf{a}}_t)$$

$$= \max_{\mathbf{a}_1, \cdots, \tilde{\mathbf{a}}_{t-1}, \hat{\mathbf{a}}_{t-1}, \tilde{\mathbf{a}}_t} \left\{ \begin{array}{l} P(\mathbf{r}_t | \mathbf{r}_1, \cdots, \mathbf{r}_{t-1}, \mathbf{a}_1, \cdots, \tilde{\mathbf{a}}_{t-1}, \hat{\mathbf{a}}_{t-1}, \tilde{\mathbf{a}}_t, \hat{\mathbf{a}}_t) \\ \times P(\mathbf{r}_1, \cdots, \mathbf{r}_{t-1} | \mathbf{a}_1, \cdots, \tilde{\mathbf{a}}_{t-1}, \hat{\mathbf{a}}_{t-1}, \tilde{\mathbf{a}}_t, \hat{\mathbf{a}}_t) \end{array} \right\}$$

Since the code has unit memory we can write

$$f(\hat{\mathbf{a}}_t) = \max_{\hat{\mathbf{a}}_{t-1}, \tilde{\mathbf{a}}_t} \left\{ P(\mathbf{r}_t | \hat{\mathbf{a}}_{t-1}, \tilde{\mathbf{a}}_t, \hat{\mathbf{a}}_t) \left\{ \max_{\mathbf{a}_1, \cdots, \tilde{\mathbf{a}}_{t-1}} P(\mathbf{r}_1, \cdots, \mathbf{r}_{t-1} | \mathbf{a}_1, \cdots, \tilde{\mathbf{a}}_{t-1}, \hat{\mathbf{a}}_{t-1}) \right\} \right\}$$

$$= \max_{\hat{\mathbf{a}}_{t-1}, \tilde{\mathbf{a}}_t} \left\{ P(\mathbf{r}_t | \mathbf{b}_t (\hat{\mathbf{a}}_{t-1}, \tilde{\mathbf{a}}_t, \hat{\mathbf{a}}_t)) f(\hat{\mathbf{a}}_{t-1}) \right\}$$

The process can be described by a trellis diagram (Fig. 2). At time t we have to compute and store the metric at each state (i.e., $f(\hat{\mathbf{a}}_t)$). Also, we need to store the corresponding optimal path leading to that state. The performance simulation software package for the byte-oriented Viterbi decoding algorithm is summarized as follows:

(1) Initialization ($t = 0$)

(a) Set up coder matrix that gives $\mathbf{b}_t$ for each $\mathbf{a}_t$ and $\hat{\mathbf{a}}_{t-1}$.

(b) Set up 3-bit quantized AWGN channel probability matrix.

(c) Set

$$f(\hat{\mathbf{a}}_0 = \mathbf{0}) = 1 \text{ and } f(\hat{\mathbf{a}}_0 \neq \mathbf{0}) = 0$$

$$\text{IERR}(\hat{\mathbf{a}}_0) = 0$$

Note that since we are interested just in symbol error probability, in the simulation we need to store only the accumulated number of errors of the optimal path that leads to a particular state (i.e., IERR $(\hat{\mathbf{a}}_t)$).

(2) At time t (main loop; $t = 1, 2 \cdots$), the following steps are taken:

(a) Simulate current observed byte $\mathbf{r}_t$.

(b) For each state $\hat{\mathbf{a}}_t$, compute

$$f(\hat{\mathbf{a}}_t) = \max_{\hat{\mathbf{a}}_{t-1}, \tilde{\mathbf{a}}_t} \left\{ P(\mathbf{r}_t \mid \mathbf{b}_t(\hat{\mathbf{a}}_{t-1}, \tilde{\mathbf{a}}_t, \hat{\mathbf{a}}_t)) f(\hat{\mathbf{a}}_{t-1}) \right\}$$

and count the corresponding number of errors IERR $(\hat{\mathbf{a}}_t)$. Since $f(\hat{\mathbf{a}}_t)$ will be very small after many iterations, it is necessary to normalize it:

$$f(\hat{\mathbf{a}}_t) = f(\hat{\mathbf{a}}_t)/f(\mathbf{0})$$

(3) Finally, the estimate $\mathbf{a}_1^0, \mathbf{a}_2^0, \cdots, \mathbf{a}_I^0$ is chosen to be the path that leads to the state $\hat{\mathbf{a}}_I^0$ such that

$$f(\hat{\mathbf{a}}_I^0) \geq f(\hat{\mathbf{a}}_I), \quad \text{for all } \hat{\mathbf{a}}_I\text{'s}$$

with its corresponding number of errors IERR $(\hat{\mathbf{a}}_I^0)$.

III. Performance

The above simulation software package requires little memory (in the order of 2^{k_0}) compared to the one for RTMBEP algorithm (in the order of $2^{k_0+l_0}$). This occurs since in the RTMBEP algorithm we need to store the probability matrices $P(\mathbf{r}_{t+l} \mid \mathbf{b}_{t+l})$ where $i = 0, 1, \cdots, \Delta$ so that not all of these have to be recalculated in the next iteration. Even so, the RTMBEP algorithm is still slower due to complicated recursive procedures (see Ref. 1). The Viterbi decoder runs about four times faster for small codes ($k_0 = 4$) and about twice as fast for big codes ($k_0 = 9$). Amazingly enough, with all these advantages, the Viterbi algorithm still achieves similar performance. For comparison, results based on 8000-byte decoding simulation are shown in Table 1 for a (4,4/8) code and a (6,6/30) code. The Viterbi algorithm simulation is run for a (6,9/36) code and a (7,10/48) code found by Pil Lee. The symbol-error probabilities based on 4000-byte decoding simulation for the (6,9/36) code and 2000-byte decoding simulation for the (7,10/48) code are given in Table 2 and plotted in Fig. 3 for both 3-bit and 4-bit quantized AWGN channels (see the Appendix). These codes are concatenated with various matching symbol-size Reed-Solomon codes. The required E_b/N_0 (i.e., outer code signal-to-noise ratio) to achieve a bit-error-rate (BER) of 10^{-6} is shown in Table 3 and plotted in Fig. 4. With 4-bit channel output quantization, the (7,10/48) unit-memory code, (1023, 927) Reed-Solomon code combination requires only 0.91 dB in E_b/N_0. This represents an improvement of 1.62 dB over the proposed NASA standard (i.e., (7,1/2) convolutional code, (255,223) Reed-Solomon code combination).

IV. Conclusion

A software package was developed to simulate the performance of the byte-oriented Viterbi decoding algorithm for unit-memory codes. This simulation requires negligible memory compared to that for the RTMBEP algorithm. It also runs faster because of its simplicity. As a result, it is possible to determine the symbol-error probability for large byte-oriented codes. Then the required E_b/N_0 to achieve a BER of 10^{-6} can be evaluated for concatenated systems. A (7,10/48) code, (1023,927) Reed-Solomon code combination is found to achieve the required BER at 0.91 dB, which is a 1.62-dB improvement over the proposed NASA standard.

Acknowledgment

The author would like to thank Pil J. Lee who provided extensive code search for unit memory codes and Dr. James Lesh for his suggestions.

Reference

1. Vo, Q. D., "Simulations for Full Unit-Memory and Partial Unit-Memory Convolutional Codes with Real-Time Minimal-Byte-Error Probability Decoding Algorithm," *The Telecommunications and Data Acquisition Progress Report 42-76*, pp. 77-81, Jet Propulsion Laboratory, Pasadena, California, February 15, 1984.

Table 1. Performance comparison between Viterbi and RTMBEP decoding algorithms

Code	G_0	G_1	Viterbi decoding				RTMBEP decoding			
(4,4/8)	87 4B 2D 1E	8B E2 B8 D1	E'_b/N_0 (dB) P_s	1.5 0.0313	2.0 0.0157	2.5 0.0063	E'_b/N_0 (dB) P_s	1.5 0.0317	2.0 0.0154	2.5 0.0065
(6,6/30)	20FBAC1C 107DD60E 08BEE307 04DD71A3 02EEB0F1 01F75878	0F14B4C1 1E296982 3C52C344 39A19688 33472D10 278A5A60	E'_b/N_0 (dB) P_s	0.75 0.0252	1.00 0.0122	1.25 0.0077	E'_b/N_0 (dB) P_s	0.75 0.0246	1.00 0.0125	1.25 0.0079

Table 2. Simulated symbol-error probability for a (6, 9/36) code and a (7, 10/48) code—Viterbi decoding

Code	G_0	G_1	Symbol-error probability, P_s			
(6,9/36)	FFFFF0000 FFC00FFE0 F83E0FC1F E4210C3D8 07398B018 D71062816 909EC4234 6064DA924 6A0956DB0	C144DAA24 92168221E 973860692 072155018 60A50EACA 7008D69B1	E'_b/N_0 (dB) 4-bit channel 3-bit channel	0.3 0.0765 0.0841	0.5 0.0567 0.0605	0.7 0.0328 0.0446
(7,10/48)	FFFFFFF00000 FFFC000FFFC0 FE03F80FE03F C183870E183C 3D7B44C9DF22 BAE2B7AFB4FB 415CF745D496 3846FE7B6C28 B101CDE50AB4 73F328165182	AA84C7D08C3B A4DF8474F71D C11A5A2916B4 A65295EC8A17 DE156BCAEA0B 7EE41D2591E3 486C6ECAD964	E'_b/N_0 (dB) 4-bit channel 3-bit channel	0.0 0.0850 0.0965	0.25 0.0460 0.0590	0.50 0.0245 0.0380

Table 3. Required E_b/N_0 to achieve a BER of 10^{-6}

Code	(6,9/36)				(7,10/48)			
R-S code rate	399/511	415/511	431/511	447/511	831/1023	863/1023	895/1023	927/1023
Required P_s	0.06214	0.05080	0.03986	0.02942	0.06143	0.04914	0.03722	0.02580
Required E'_b/N_0 (dB) 3-bit	0.482	0.618	0.77	0.987	0.23	0.354	0.512	0.716
Required E'_b/N_0 (dB) 4-bit		0.557	0.65	0.724		0.222	0.335	0.482
Required E_b/N_0 (dB) 3-bit	1.56	1.52	1.51	1.57	1.13	1.09	1.09	1.14
Required E_b/N_0 (dB) 4-bit		1.46	1.39	1.31		0.96	0.92	0.91

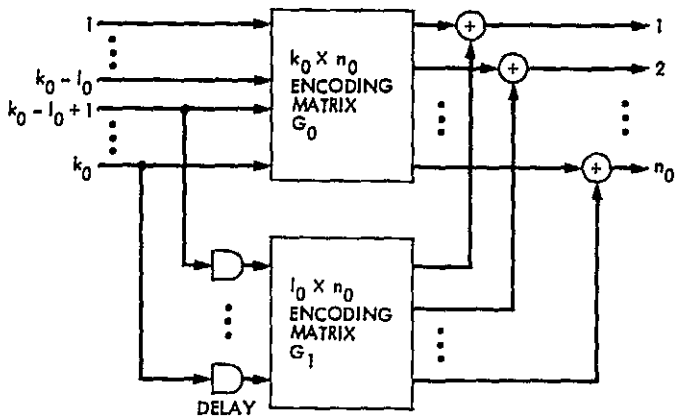


Fig. 1. A general $(l_o, k_o/n_o)$ unit-memory convolutional encoder

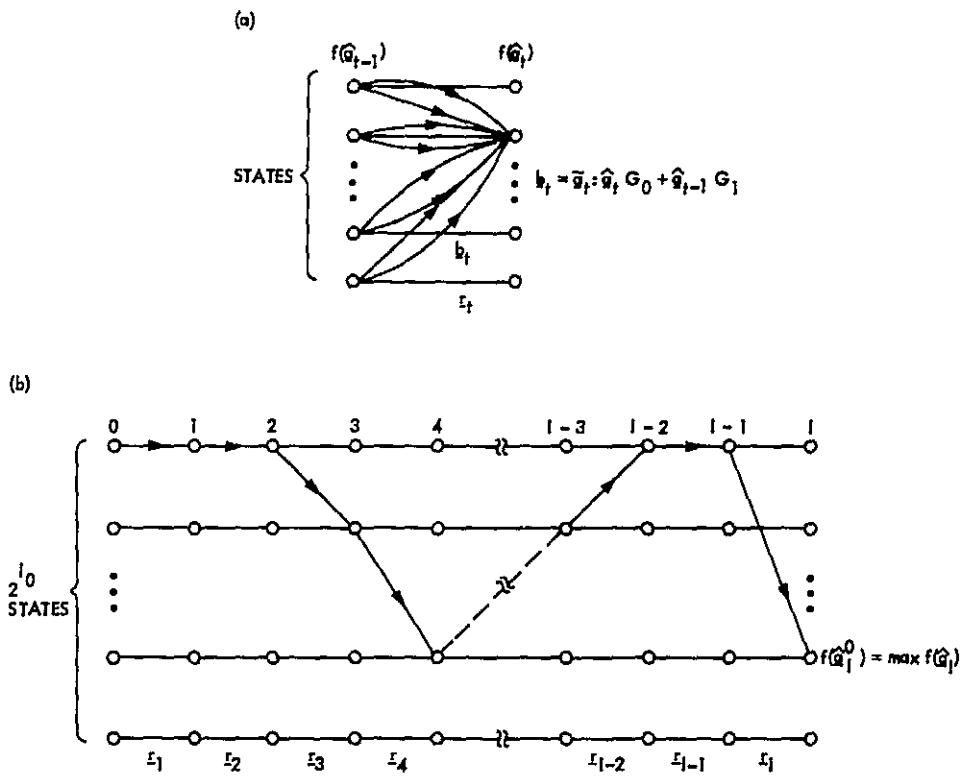


Fig. 2. Trellis diagram: (a) byte-oriented Viterbi decoding algorithm; (b) optimal path tracing

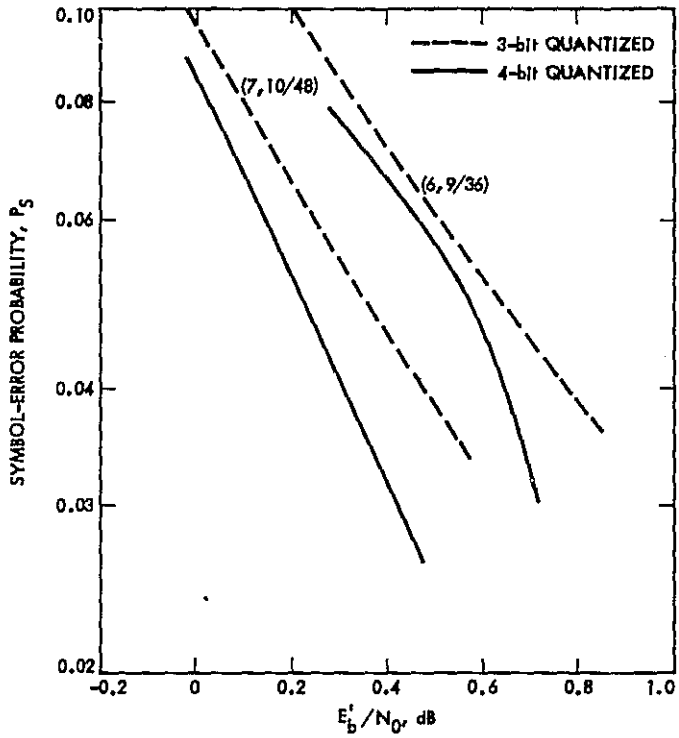


Fig. 3. Simulated symbol-error probability for a (6, 9/36) code and a (7, 10/48) code

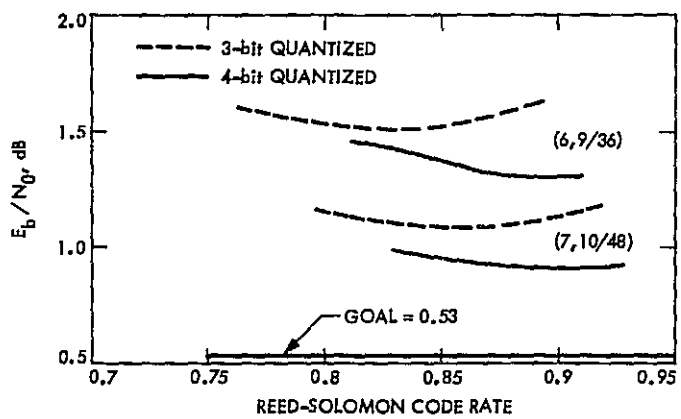


Fig. 4. Required E_b/N_0 to achieve a BER of 10^{-6}

Appendix

4-bit vs 3-bit Channel Output Quantization

The quantization of the output to one of J levels simply transforms the AWGN channel to a finite-input, finite-output alphabet channel (Ref. A-1). For our case, a biphase modulated AWGN channel with output quantized to eight levels (3-bit quantizer) is shown in Fig. A-1. The channel conditional probabilities can be computed as follows:

$$P(1|0) = Q(3a - x)$$

$$P(M|0) = Q((4 - M)a - x) - Q((5 - M)a - x); \qquad M = 2, \cdots, 7$$

$$P(8|0) = 1 - Q(-3a - x)$$

$$P(M|1) = P(9 - M|0)$$

where

$$a = \text{quantizer step size}$$

$$x = \sqrt{2E_s/N_0}$$

The quantizer step size is chosen to maximize the Bhattacharyya distance:

$$d = -\ln \sum_{M=1}^{8} \sqrt{P(M|0)P(M|1)}$$

From this, the channel cutoff rate R_0 can be easily computed:

$$R_0 = \max_q \left\{ -\ln \sum_y \left[\sum_x q(x)\sqrt{P(y|x)} \right]^2 \right\}$$

$$= \max_q \left\{ -\ln \sum_x \sum_{x'} q(x)\,q(x') \sum_y \sqrt{P(y|x)P(y|x')} \right\}$$

For binary input, this becomes

$$R_0 = -\ln \left[\frac{1}{2} + \frac{1}{2} \sum_y \sqrt{P(y|0)P(y|1)} \right] = -\ln \left(\frac{1 + e^{-d}}{2} \right)$$

The 4-bit quantized channel model is shown in Fig. A-2 with the channel conditional probabilities given by

$$P(1|0) = Q(7a - x)$$

$$P(M|0) = Q((8 - M)a - x) - Q((9 - M)a - x); \qquad M = 2, \cdots, 15$$

$$P(16|0) = 1 - Q(-7a - x)$$

$$P(M|1) = P(17 - M|0)$$

where the quantizer step size a is chosen to maximize

$$d = -\ln \sum_{M=1}^{16} \sqrt{P(M|0)P(M|1)}$$

The cutoff rate is plotted for both channels over the interested range of E_s/N_0 in Fig. A-3. We see that, to achieve the same cutoff rate, we can save approximately 0.11 dB (in the range around $E_s/N_0 = -6$ dB) in required signal-to-noise ratio by using a 4-bit quantizer instead of a 3-bit quantizer. This fact was originally suggested by Pil Lee (personal communication).

Reference

A-1. Viterbi, A. J., and J. K. Omura, *Principles of Digital Communication and Coding*, pp. 78-82, McGraw-Hill, New York, N.Y., 1979.

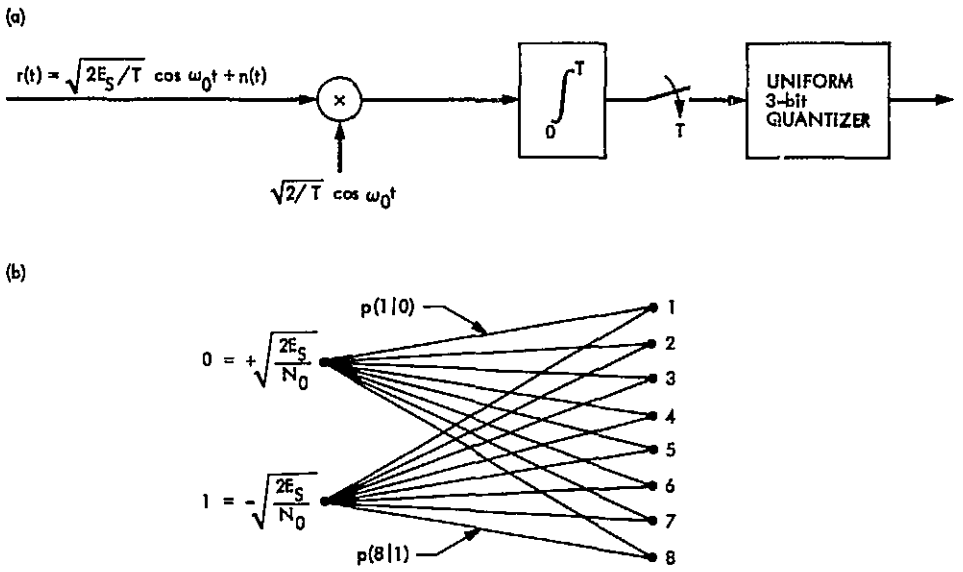


Fig. A-1. 3-bit quantized demodulator and channel model: (a) demodulator for BPSK signals; (b) channel model

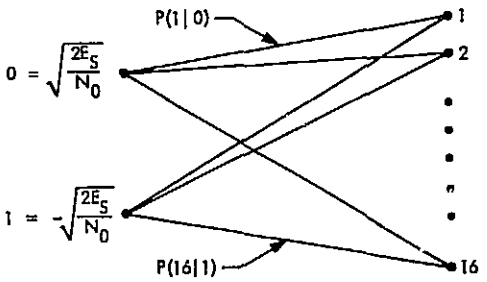


Fig. A-2. 4-bit quantized channel model

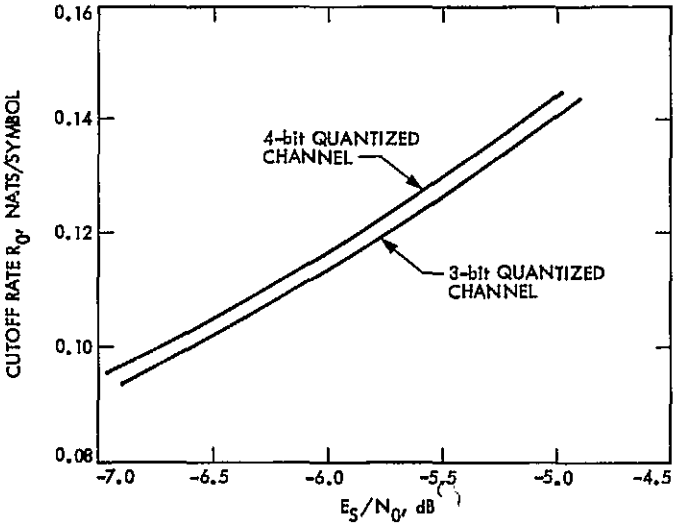


Fig. A-3. Channel cutoff rate

A Simple Algorithm for the Metric Traveling Salesman Problem

M. J. Grimm
Communications Systems Research Section

An algorithm was designed for a wire list net sort problem. A branch and bound algorithm for the metric traveling salesman problem is presented for this. The algorithm is a best bound first recursive descent where the bound is based on the triangle inequality. The bounded subsets are defined by the relative order of the first K of the N cities (i.e., a K city subtour). When K equals N, the bound is the length of the tour. The algorithm is implemented as a one page subroutine written in the C programming language for the VAX 11/750. Average execution times for randomly selected planar points using the Euclidean metric are 0.01, 0.05, 0.42, and 3.13 seconds for ten, fifteen, twenty, and twenty-five cities, respectively. Maximum execution times for a hundred cases are less than eleven times the averages. The speed of the algorithm is due to an initial ordering algorithm that is a N squared operation.

The algorithm also solves the related problem where the tour does not return to the starting city and the starting and/or ending cities may be specified. The algorithm can easily be extended to solve a nonsymmetric problem satisfying the triangle inequality.

I. Introduction

The Digital Projects Group uses an in-house development aid program (Ref. 3) for specifying the interconnections of the I/O pins of integrated circuits placed on wire wrap boards. The program produces nets of (x, y) coordinates of points that must be interconnected. The coordinates are the locations of socket pins upon which at most two wires may be placed. An insulated wire electrically connects two socket pins. The length of a net connecting the N socket pins is the sum of the effective lengths of the $N-1$ wires connecting pairs of points in that net. When a wire wrap machine connects the point $(x1, y1)$ with the point $(x2, y2)$, the effective length of the wire is $|x1 - x2| + |y1 - y2|$. It is desirable to minimize the length of the nets.

For a typical logic design, a board has two thousand nets, with each net containing an average of three points. Approximately eight percent of the nets have ten or more points, and a few exceed fifteen points. With such small nets, it would seem that a simple algorithm could be specified that finds minimum length nets using a reasonable amount of computer time. The algorithm to be presented satisfies the requirement, and also

solves the metric traveling salesman problem (TSP). It also allows one or both end points of the net to be specified, facilitating the routing of electrically terminated nets.

II. The Combinatorial Problem

The TSP can be defined as follows. Given an N by N matrix **D** of non-negative integers; find an order vector **O** such that

$$\mathbf{D}(\mathbf{O}[N], \mathbf{O}[1]) + \sum_{I=1}^{N-1} \mathbf{D}(\mathbf{O}[I], \mathbf{O}[I+1])$$

is minimal. The order vector may be any permutation of the numbers one through N. If $\mathbf{D}[I, J] = \mathbf{D}[J, I]$ then the problem is called the symmetric TSP. If, additionally, **D** has a zero diagonal and satisfies the triangle inequality ($\mathbf{D}[I, J] \leqslant \mathbf{D}[I, K] + \mathbf{D}[K, J]$) the problem is called the metric TSP. If the first term in the minimization equation is omitted, the salesman does not have to return home, and the problem becomes the net sort problem. The distance defined in the introduction is a metric; so that problem is a metric net sort problem.

The algorithm to be presented solves both the metric TSP and the net sort problem. It can easily be extended to solve a nonsymmetric problem given that the triangle inequality holds.

III. The Algorithm

Bently (Ref. 2) describes approximate solutions to the TSP that are feasible for $N = 1000$. Smith (Ref. 4) defines the branch and bound algorithm but does not give execution times. Bellmore and Malone (Ref. 1) give execution times for random Euclidean problems that are three orders of magnitude inferior to the algorithm to be presented. The algorithm is a simple application of branch and bound using the triangle inequality. The novelty of the algorithm is a worst possible ordering presort which enhances execution times by about three orders of magnitude for $N = 20$.

The branch and bound algorithm (Refs. 1 and 4) can be defined recursively as follows. Given a set of nets and an upper bound (current minimum) on the absolute minimum length net: Partition the set into subsets and compute lower bounds on the lengths of the nets in each subset. If the lower bound of a subset is not less than the current minimum, discard that subset because it cannot contain a net shorter than the current minimum. If the lower bound of a subset is less than the current minimum and the subset contains more than one net, explore that subset. Otherwise, if the set has only one net, then the length of that net becomes the current minimum, and that net is saved as a potential minimum length net. The algorithm terminates when the set has been explored, and the potential minimum length net is then the minimum length net.

The algorithm to be presented is a best bound first algorithm; that is, when a set is partitioned, the first subset to be explored is the subset with the smallest lower bound. The algorithm also has the property that subsets are mutually exclusive and collectively exhaustive and that the subsetting process ultimately yields a subset with just one net in it, whose length is the lower bound.

The algorithm computes an initial ordering for the points, and it labels them one through N. The initial ordering will be described in the next paragraph; this one describes the subsets and bounds used by the algorithm. The subsets are defined by the relative order of the first K points in the net, and the lower bounds are computed directly from the triangle inequality. The depth of a subset is defined to be the number of points considered (K). A convenient label for a depth K subset is its K point order vector. For example, the depth three subset 1-3-2 is the set of all nets for which point three is between points one and two. If N is greater than three, 1-3-2 can be partitioned into the four depth four subsets: 4-1-3-2, 1-4-3-2, 1-3-4-2, and 1-3-2-4. For the net sort problem, the lower bound for the set 1-3-2 is $\mathbf{D}[1, 3] + \mathbf{D}[3, 2]$. For the TSP the lower bound is $\mathbf{D}[1, 3] + \mathbf{D}[3, 2] + \mathbf{D}[2, 1]$. The triangle inequality guarantees that no net in the subset is shorter than the lower bound. Moreover a depth N subset contains just one net whose length is the lower bound.

The algorithm consists of recursively partitioning depth K subsets into depth $K + 1$ subsets until either the lower bound exceeds the current minimum or a new current minimum is found. The speed of the algorithm is found to be extremely data sensitive, and an initial ordering of the points is required. In order to maximize the lower bounds of depth K subsets, and hence tend to eliminate subsets without having to partition them, the first K points should be chosen to be maximally separated. This is accomplished with the following presort algorithm. For the net sort problem, if both end points are specified, they are labeled one and two. If one end point is specified, it is labeled one, and point two is the one farthest from it. Otherwise, and for the TSP, the first two points are chosen as the ones farthest apart. Point $P[J + 1]$ is chosen as a point not already chosen which is the farthest distance from all of the J points already chosen. That is, point $P[J + 1]$ is the point K such that: $\mathrm{MIN}(\mathbf{D}[P[I], K]: I = 1, \ldots, J)$ is maximal for $1 \leqslant K \leqslant N$ and K not in $P[S]: S = 1, \ldots, J$.

This initial ordering is a N squared operation and significantly enhances the execution speed of the algorithm.

The branch and bound algorithm is applied to the presorted points and is as follows. Start with the set of all nets to be considered, the lower bound for the set, and a current minimum of infinity. From the set's lower bound, compute lower bounds for each one deeper subset. Record the two best (smallest) lower bounds. If the best lower bound is not less than the current minimum, discard the entire set because it cannot contain a net shorter than the current minimum. Otherwise if the subset depth is N, record the single net in the best subset and set the current minimum to be the subset's lower bound (length); and the set has been explored. Otherwise if the best subset's lower bound is less than the current minimum, explore it. Having explored the best subset, if the second best subset's lower bound is not less than the current minimum, the entire set has been explored. Otherwise, successively explore each subset whose lower bound is less than the current minimum. The algorithm terminates when the initial set has been explored. At this time, the current minimum is the length of an absolute minimum length net, and the recorded net is one of the absolute minimum length nets.

IV. The Computer Realization of the Branch and Bound Algorithm

The recursive algorithm is easily implemented as a recursive subroutine (SUBSET) whose arguments are as follows:

LP1	The depth (L + 1) of the resulting one deeper subsets
LEN	The lower bound of the depth L subset to be explored
D	Row LP1 of the distance matrix

The subset being explored is globally defined by the singly linked list LINK. For the TSP, the initial set to be explored is 1-2-3 which is defined as a circular list (LINK[1] = 2; LINK[2] = 3; LINK[3] = 1). Other global variables used by SUBSET are as follows:

N	The number of points in a net
MIN	The current minimum
WIN	The linked representation of the current minimum net.

The increase in lower bound of the depth LP1 subset formed by placing point LP1 between points I and LINK[I] is

$$D[I, LP1] + D[LP1, LINK[I]] - D[I,LINK[I]]$$

To facilitate the recursive computation of lower bounds in the net sort problem, the distance matrix is augmented with a column of zeros so that D[J, 0] = 0 for all J. For the net sort problem, the initial set to be explored is 1-2 which is defined by

$$LINK[0] = 2; LINK[2] = 1; LINK[1] = 0$$

Notice that the lower bound equation is now also valid for end points of the net. The valid indices (I) for the net sort problem with S preselected end points are S, . . . , L. The valid indices for the TSP are 1, . . . , L.

For a symmetric distance matrix, only one row of D is required to compute the increase in lower bound provided that D[I,LINK[I]] is maintained. DL[I] is defined to be D[I,LINK[I]]. Now the lower bound equation becomes

$$D[I] + D[LINK[I]] - DL[I]$$

where D[J] is D[LP1, J]; i.e., D is now just a row of the distance matrix.

To explore the depth L + 1 subset formed by placing L + 1 between I and LINK[I], the required updates to LINK and DL are

```
LK = LINK[I]
LINK[I] = LP1
LINK[LP1] = LK
DT = DL[I]
DL[I] = D[I]
DL[LP1] = D[LK]
```

To restore DL and LINK to the depth L subset,

```
DL[I] = DT
LINK[I] = LK
```

The C language realization of this algorithm is given in Fig. 1.

V. Example of Algorithm Execution

Consider the five-point net with its corresponding distance matrix in Fig. 2. The problem to solve is the net sort with no preselected points. The initial ordering algorithm gives the points labeled as shown. The initial set 2-1 is partitioned and lower bounds are calculated until a new current minimum of 7 is generated from 5-2-3-1-4. Sets 2-3-1-4 and 2-3-1 both have second best lower bounds not greater than the current

minimum, so they have been explored. The second best subset of 2-1 yields a new current minimum of 6 for 3-2-5-4-1. The rest of the subsets fail the second best test and the algorithm terminates. In this example, only twenty-one lower bounds were computed to determine the minimum of sixty (5!/2) possible nets.

VI. Execution Times

Table 1 gives average and maximum execution times for randomly selected planar points using the Euclidean metric for the TSP. These measurements were made on a VAX 11/750 running UNIX.

References

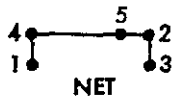
1. Bellmore, M. and Malone, J. C., Pathology of Traveling-Salesman Subtour Elimination Algorithms. *Oper. Res.* 19(1971), 278–307.
2. Bently, J. L., A Case Study in Applied Algorithm Design. *IEEE Computer* 17:2, Feb. 1984, 75–87.
3. Lushbaugh, W. A., Quicklist – The Basis for a Computer Aided Design System. *DSN Progress Report 42-58*, 67–71, May-June 1980, Jet Propulsion Laboratory, Pasadena, California.
4. Smith, D. R., Random Trees and the Branch and Bound Procedures. *J. ACM 31* (Jan. 1984) 163–188.

Table 1. Execution times

Number of Points	Number of Cases	Average, s	Maximum, s
10	100	0.01	0.07
15	100	0.05	0.23
20	100	0.42	3.87
25	100	3.13	33.55
30	25	55.36	400.37
35	14	263.63	1786.93

```
subset(lp1,len,d)
int lp1,len,*d;
{
        int dd[SIZ1],i,lk,dt,lim,lim2;
        for(lim=lim2=30000000,lk=s;lk<lp1;lk++)
                if((dd[lk]=d[lk]+d[link[lk]]-dl[lk])<lim2)
                        if(dd[lk]<lim){
                                i=lk;
                                lim2=lim;
                                lim= dd[lk];
                        }
                        else
                                lim2=dd[lk];
        if((lim+=len)>=min)
                return;
        lk=link[i];
        link[i]=lp1;
        link[lp1]=lk;
        if(lp1==n){
                for(dt=0;dt<=n;dt++)
                        win[dt]=link[dt];
                link[i]=lk;
                min=lim;
                return;
        }
        dt=dl[i];
        dl[i]=d[i];
        dl[lp1]=d[lk];
        subset(lp1+1,lim,d+n+1);
        dl[i]=dt;
        link[i]=lk;
        if((lim=min-len)<=lim2)
                return;
        for(dd[i]=30000000,i=s;i<lp1;i++)
                if(dd[i]<lim){
                        lk=link[i];
                        link[i]=lp1;
                        link[lp1]=lk;
                        dt=dl[i];
                        dl[i]=d[i];
                        dl[lp1]=d[lk];
                        subset(lp1+1,len+dd[i],d+n+1);
                        dl[i]=dt;
                        link[i]=lk;
                }
}
```

Fig. 1. The C Language realization of branch and bound algorithm



NET

	0	1	2	3	4	5
1	0	0	5	4	1	4
2	0	5	0	1	4	1
3	0	4	1	0	5	2
4	0	1	4	5	0	3
5	0	4	1	2	3	0

AUGMENTED DISTANCE MATRIX (D)

SUBSET	LOWER BOUND	CURRENT MINIMUM	DELTA LOWER BOUND 0	1	2	3	4
2–1	5		1	4	0		
2–3–1	5		4	1	8	2	
2–3–1–4	6		1	6	2	2	3
5–2–3–1–4	7	7					
3–2–1	6		5	1	0	8	
3–2–4–1	6		2	4	0	2	6
3–2–5–4–1	6	6					

Fig. 2. Example of algorithm execution for a five-point net and its corresponding distance matrix

Symbol-Stream Combiner: Description and Demonstration Plans

W. J. Hurd, L. J. Reder, and M. D. Russell
Communications Systems Research Section

A system is described and demonstration plans presented for antenna arraying by Symbol Stream Combining. This system can be used to enhance the signal-to-noise ratio of spacecraft signals by combining the detected symbol streams from two or more receiving stations. Symbol Stream Combining has both cost and performance advantages over other arraying methods. Demonstrations are planned on Voyager II both prior to and during Uranus encounter. Operational use is possible for Interagency Arraying of non-DSN stations at Neptune encounter.

I. Introduction

Symbol-Stream Combining (SSC) is a method of combining the received signals from two or more antenna-receiver systems in order to achieve a signal-to-noise ratio (SNR) which is approximately equal to the sum of the signal-to-noise ratios at each antenna-receiver. The primary motivation for SSC is interagency arraying of non-DSN and DSN stations for the Voyager Neptune encounter.

This article describes the symbol-stream combining system concept, the hardware and software implementation, and the preliminary demonstration plans.

In Symbol-Stream Combining, the received signals at each antenna station are processed through symbol detection. Then the detected symbols from the two or more stations are brought together, aligned, and combined with proper weighting. This is as opposed to Baseband Combining (BBC), in which the signals are brought together, aligned, and combined as broadband baseband signals prior to subcarrier demodulation.

Symbol-Stream Combining has three major advantages over Baseband Combining. First, the ground communication bandwidth for SSC is less than one-tenth of that for BBC. This is a major cost reduction, especially when international satellite links are required. Second, the SSC combining and backup recording equipment is less complicated and less expensive. Third, the SNR performance is approximately 0.35 dB better for SSC than for BBC, as shown by Divsalar (Ref. 1). The main disadvantage of SSC is that the subcarrier demodulators and symbol synchronizers must operate at lower SNRs than for BBC; this disadvantage will be overcome by use of synchronization techniques being developed for the DSN Advanced Receiver (Ref. 2).

Figure 1 shows the planned configuration for the first demonstration of SSC, using DSS-13 and DSS-14. Voyager 2 will be tracked by both stations. The symbol-stream combiner will be located at DSS-14. The signal at DSS-14 will be processed by a standard telemetry stream through symbol detection. Then the detected symbol stream will be input to the Symbol-Stream Combiner (SSC). The signal at DSS-13

will be synchronized and the symbols detected using a partial breadboard of the Advanced Receiver signal processing unit. These detected symbols will be transmitted to DSS-14 on the existing microwave link and input to the SSC. The combined symbols at the output of the SSC will be input to a decoder, and the symbol stream from DSS-14 only will also be decoded. Either two decoders will be used simultaneously, or one will be time shared. The decoder performance will be monitored to establish the SNR improvement achieved by SSC.

II. System Description

One of the main advantages to the symbol stream combining method of antenna arraying is the simplicity of the hardware necessary for the implementation of the system. Because SSC is performed on the quantized symbols after subcarrier demodulation and symbol synchronization, the data rate into the SSC is equal to the symbol rate, which is under 60 kHz for Voyager Uranus encounter. This is much lower than the 15 MHz sampling rate of the Baseband Assembly (BBA), which implements Baseband Combining. Thus, functions performed in hardware in the BBA can be done in software in the SSC. These software functions include cross correlation for alignment and SNR estimation. The SSC also performs SNR estimation, in software, on the input symbol streams to determine the weighting constants to be used in the hardware for the combining.

The basic functions that must be carried out in SSC are alignment of the quantized symbol streams, weighting of each symbol, and combining or summing the weighted symbols. The combining function is as follows:

$$Z(n) = a_0 X_0(n) + a_1 X_1(n + k)$$

where

0, 1 = station index

X_i = symbols from station i

Z = combined symbols

n = time index for nth symbol

k = delay to station 1 from station 0

a_i = weighting constant for station i

The weighting constant is computed in the SSC using the algorithms analyzed by Q. D. Vo (Ref. 3). The above relation is implemented in hardware with software control and computation of values for a_0, a_1, and k. Alignment is accomplished by delaying the $X_1(n)$ stream by k samples with respect to the $X_0(n)$ stream. The value of k is determined by software cross correlation of the two streams and by utilizing delay calculations based on antenna pointing predicts and knowledge of the vector baseline between the antennas. Once the value of k is determined, weighting and combining is performed in special purpose hardware.

III. Hardware Description

The special hardware is a multibus board which was developed for the BBA for the combining and weighting of sampled data and a special delay card developed for this project. A block diagram of the hardware configuration is shown in Fig. 2. The hardware consists of a multibus card cage with seven standard and two special purpose cards, two disk drives, two function generators, and an interface assembly. The standard cards are:

1 INTEL SBC-86/14 Single Board Computer with an 8087 Multimodule

1 Chrislin Industries CI-8086 512 KByte Dynamic RAM Card

1 National Semiconductor 8222 Disk Controller

1 INTEL SBC-534 I/O Expansion Card

2 SBX-488 Multimodule interface cards for controlling the function generators

1 Digital Pathways TCU-410 Clock Calendar Card

The two special multibus cards are a Programmable Digital Delay Card and a Multiplier-Adder Board for combining. Two eight-inch Shugart Disk Drives are used also.

A custom interface assembly to interface to the microwave link at the stations and generate test sequences for system self-test is also shown. The two function generators supply clock signals. One supplies a times-four clock, phase locked to the channel-0 symbol clock. The other supplies a test signal clock for channel 1 during self-test.

A. Delay Card

The function of the Delay Card is to input two symbol streams, buffer them, delay one stream with respect to the other so as to align the corresponding symbols from the two sources, and to output the two aligned symbol streams to the Multiplier-Adder Board.

A diagram of Delay Card signal flow is shown in Fig. 3. The card consists of two independent channels so two streams can be delayed simultaneously. Each channel has a first-in-first-out (FIFO) circular buffer memory, a read address

counter and a write address counter. Channel 0 is set at a fixed delay of k_0 symbols, depending on the particular stations being arrayed. This fixed delay is greater than the maximum allowance (negative k) for the baseline. The channel-1 data are delayed $k_0 + k$, which is always positive. The differential delay is k, as required.

The main complexity in the Delay Card arises because the symbol rates from the two stations are slightly different due to Doppler shift. This Doppler shift or changing delay is caused by Earth rotation. The differential fractional Doppler between stations is on the order of 10^{-9} for two Goldstone stations, and up to 2×10^{-6} for widely separated stations. Thus, the delay changes by one symbol time in 10^6 to 10^9 or more symbols. In each channel-0 symbol time, there is almost always one channel-1 symbol, but occasionally there are zero or two channel-1 symbols.

The Delay Card operates synchronously with the symbol-stream clock from channel-0. A clock at four times the symbol-0 clock is input to the Delay Card. This clock is divided into four phases, a read phase, two write phases (W0 and W1), and a do-nothing phase.

The operation for channel-0 symbols is straightforward. During each channel-0 symbol time, the newly arriving symbol is written into the buffer on write phase W0, and one symbol is read out on the read cycle.

The situation for channel-1 input is more complicated. The channel-1 symbols arrive asynchronously to the times-four clock. They are first synchronized to that clock using a shift register. As the channel-1 symbols are synchronized, a Data-Ready signal is generated for each symbol. Then, on the next write phase, either W0 or W1, the symbol is written into the channel-1 buffer and the Data-Ready signal is turned off. Thus, there are two opportunities to write into the channel-1 buffer during each channel-0 symbol time, but 0, 1, or 2 writes occur depending on the number of channel-1 symbols arriving.

One symbol is read out of the channel-1 buffer on each read phase, i.e., at the same instant as the corresponding channel-0 symbol is read. The synchronization delay for channel-1 is denoted by ϵ. The FIFO buffer delay is $k_0 + k - \epsilon$. Finally, the differential delay between the two channels is k, as desired.

B. Multiplier–Adder Board

The Multiplier-Adder Board performs the weighting of the symbol streams and summation of the weighted streams to produce the combined output. This card is similar to the BBA Multiplier–Adder Board (MAB), with a minor modification to enable 8086 addressing rather than 8080 addressing. The MAB can combine up to four symbol streams and can be used with an additional MAB to combine up to eight streams. The card performs 8-bit multiplication on input data streams with weighting function values input from the 8086 over the multibus. The card then performs sums on these data to produce the combined results. In addition, the MAB provides circuitry for capturing either input data to the card or data at a selected point in the computation and transferring these data over the multibus to the computer. This enables the software to acquire the symbols for use in the cross correlation and SNR estimation computations. The weighting coefficients are returned to the MAB from the software.

C. Interface Assembly

The interface assembly provides two interfaces between the SSC Delay Card and the microwave link or a local symbol synchronizer assembly (SSA). The interface assemblies also generate psuedorandom test patterns for self-test of the data paths, under computer control.

For flexibility and reliability, both interfaces can input either SSA or microwave link data. The interfaces use Zilog Z8530 Serial Communication Controllers for parallel-to-serial and serial-to-parallel conversion, and Computrol modems for FSK modulation and demodulation of the microwave link signal. The Z8530 is used to send and receive the streams using a standard SDLC protocol.

IV. Software Description

The SSC software is composed of modules that perform operator interfacing, calculation, and control functions. The software organization is shown in Fig. 4.

The operator communicates with the system through a series of menus and prompts. After the operator has made the necessary key-ins and has selected the option to align the incoming data streams, the program takes over. First, the SNR Estimation Module computes SNR estimates and combiner weight values. Next, the Delay Calculation Module computes delay values from the input antenna predict information. These values are used to initialize the boards. Then control passes to the Alignment Module, which consists of the Correlation and Control Submodules. Control passes between these submodules until the streams are aligned or the operator intervenes. During the alignment process, operator displays including SNR estimates, correlation values, and delay estimates are constantly updated.

A. Environment

The program operates under the CP/M-86 operating system. As much software as possible is written in Pascal MT+86, with low-level interface routines written in 8086 assembly language. In critical areas where bottlenecks might occur, assembly language is used for speed.

B. Operator Interface

The operator has the option of using built-in default values or entering his own values for most system variables. Variables such as SNR estimates or antenna pointing predicts can be entered through the keyboard or from a data file that the operator specifies. Upon power up, the operator can enter either a test mode or the Real Time Combiner (RTC) mode. From the main menu the operator can also display system date and time. The SNR Estimation and Delay Calculation Modules can also be run individually with results displayed to the operator, for testing. Operator displays take the form of changing real time displays that can be viewed during the test mode or the RTC mode. Operator control is maintained via the keyboard.

The ability to write test data to disk is available in either the test mode or the RTC mode. Data written to disk consist of system variables, raw data, combined data, computed results, time, date, test identification, such as test name or number, and a small field for comments.

C. Mathematical Modules

There are three mathematical modules, the Delay Calculation Module, the SNR Estimation Module, and the Alignment Module.

1. Delay Calculation Module. Due to the geographical separation of the receiving stations, spacecraft signals arrive at one station later than the other. That delay and the delay associated with transmitting the received data stream from one station to the station at which the SSC equipment is located require that one stream be delayed to allow proper alignment with the other stream. The Delay Calculation Module takes as input two DSS station names and antenna pointing predicts associated with each station. The geometrical delay is calculated using the predicts and the baseline vector between the two stations. It is necessary to perform this calculation only for acquisition because the hardware maintains delay synchronization after acquisition. Transmission delay time between stations and cabling delay times for each station are also accounted for. A database containing this information is maintained on disk.

2. SNR Estimation Module. The SNR Estimation Module computes an estimate of SNR for each incoming data stream. Using a number of raw data samples from the MAB, it computes an estimated SNR and then weighting constants for each data stream. It also computes the output SNR. The algorithms used are those developed by Q. D. Vo (Ref. 3).

3. Alignment Module. The Alignment Module consists of the Correlation Module and the Control Module that aligns the data streams. It accepts delay estimates from the Delay Calculation Module for initial alignment. It then uses the Correlation Module to determine the exact delay, which will be close to the delay input from the Delay Calculation Module, and aligns the symbol streams accordingly.

The Correlation Module reads the two symbol streams from the MAB, correlates them, and decides if the streams are aligned. This decision is based on a correlation threshold value which can be changed by the operator. If the streams are aligned, then this program is left and the tracking/combining mode is entered. If the streams are not aligned, then the delay is changed by a software algorithm, and the process is repeated until alignment is achieved.

The Control Module takes delay values from the Correlation Module and computes values for the read and write counters for the registers on the Delay Card. It stops the delay function on the card, resets the counters to the desired values, restarts the delay function on the card, and returns to the Correlation Module. It also accepts new weight values from the Correlation Module and loads them to the MAB. During the alignment process, this module forms the one control point for access to both the MAB and the delay card.

D. Interface Modules

The two main interface modules are the modules that communicate with the MAB and the Delay Card. They perform the low-level control functions to pass data and commands between the high-level PASCAL software and the hardware. The MAB can operate in a test mode or a combiner mode. In its test mode, outputs to the board consist of start and stop test commands and values for the on-board test and weight registers. Input from the MAB in its test mode consists of intermediate and final results from the adder trees and status information. Output in the combiner mode to the board consists of multiplier constants for the weight registers, and input consists of the raw data symbols, intermediate adder tree results, and combined data symbols.

The Delay Card is controlled by outputs to the card consisting of values loaded into the read and write counters for each delay buffer and commands to enable or disable the delaying function of the card.

Testing of both the interface modules and the cards themselves is controlled by low-level menus built into the software

modules. The Delay Card test menu allows the operator to load the read and write counters for each channel directly and to enable or disable the delaying function for each channel. The MAB has the circuitry on board to generate its own input to the adder tree. A static test or a dynamic test can be conducted. Intermediate results can be read from seven different points in the adder tree.

There are two other smaller hardware interface modules. One allows the operator to view and change the system date and time which are maintained on the Timing and Control Unit (TCU) board. The other interfaces the two HP 3314 Function Generators via two SBX488 GPIB modules. This allows the function generators to be under program control.

V. Preliminary Demonstration Plans

Initial demonstration of SSC is planned for the first quarter of calendar year 1985, using DSS-13 and DSS-14. Prior to the first combining demonstration, the SSC will be laboratory-tested, and the microwave link interfaces and other station interfaces will be checked out. Several SSC demonstrations are planned for the first half of 1985.

The desired operating conditions are that the symbol SNR at DSS-14 be approximately -2 to +3 dB, and that the symbol SNR at DSS-13 be approximately -10 to -5 dB. Thus, the breadboard of the advanced receiver telemetry processor is required at DSS-13 in order to achieve subcarrier and symbol synchronization at low SNR with low degradation. If the advanced telemetry processor breadboard is not ready when desired, an alternate configuration using DSS-12 and its standard telemetry processing system will be used for a first demonstration. This would demonstrate the SSC technique, but not the ability to use a very low SNR station. The low SNR capability would be demonstrated when the advanced telemetry processor breadboard becomes available.

Tentative plans are being developed to use the SCC during Voyager Uranus encounter in January, 1986. This usage could be in a backup mode to BBC between Goldstone stations or could be to array Owens Valley Radio Observatory or DSS-13 into the Goldstone array. Either configuration would be on a nonoperational basis.

A major goal of these demonstrations is to make symbol-stream combining a viable option for implementation of Interagency Arraying for the Voyager Neptune encounter.

References

1. Divsalar, D., "Symbol Stream Combining Versus Baseband Combining for Telemetry Arraying," *TDA Progress Report 42-74*, Jet Propulsion Laboratory, Pasadena, Calif., August 15, 1983, pp. 13-28.

2. Sfeir, R., Aguirre, S., and Hurd, W. J., "Coherent Digital Demodulation of a Residual Carrier Signal Using IF Sampling," *TDA Progress Report 42-78*, Jet Propulsion Laboratory, Pasadena, Calif., August 15, 1984.

3. Vo, Q., "Signal-To-Noise Ratio and Combiner Weight Estimation for Symbol Stream Combining," *TDA Progress Report 42-76*, Jet Propulsion Laboratory, Pasadena, Calif., February 15, 1984, pp. 86-98.

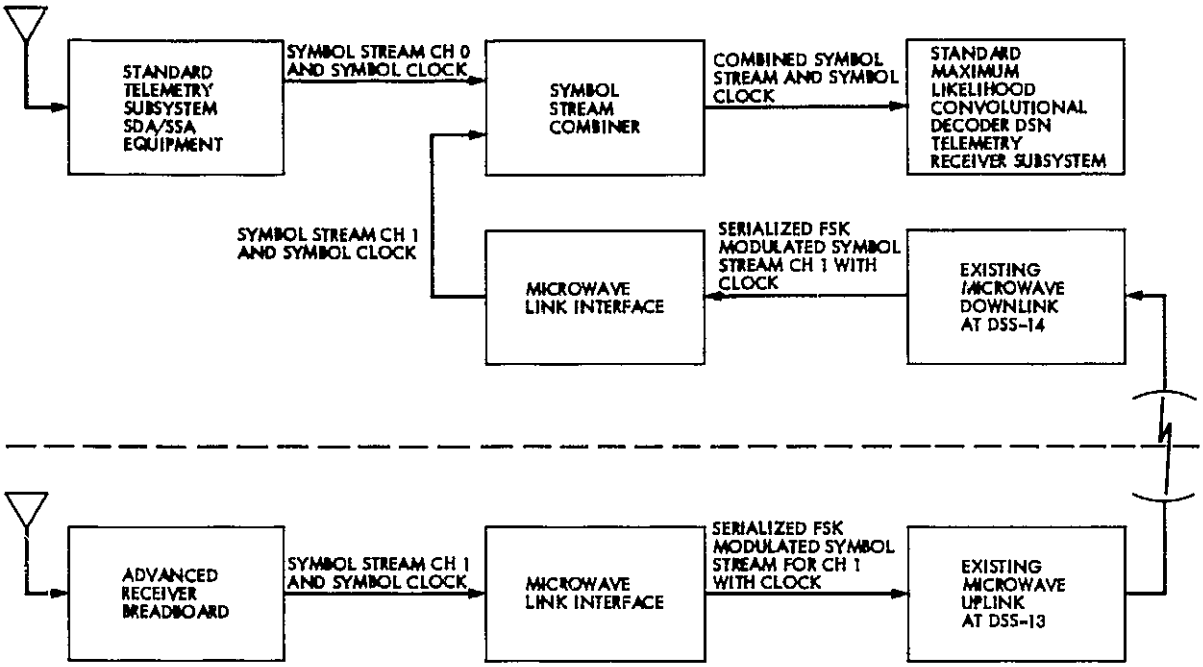


Fig. 1. Symbol-stream combiner demonstration station configuration block diagram for DSS-14 and DSS-13

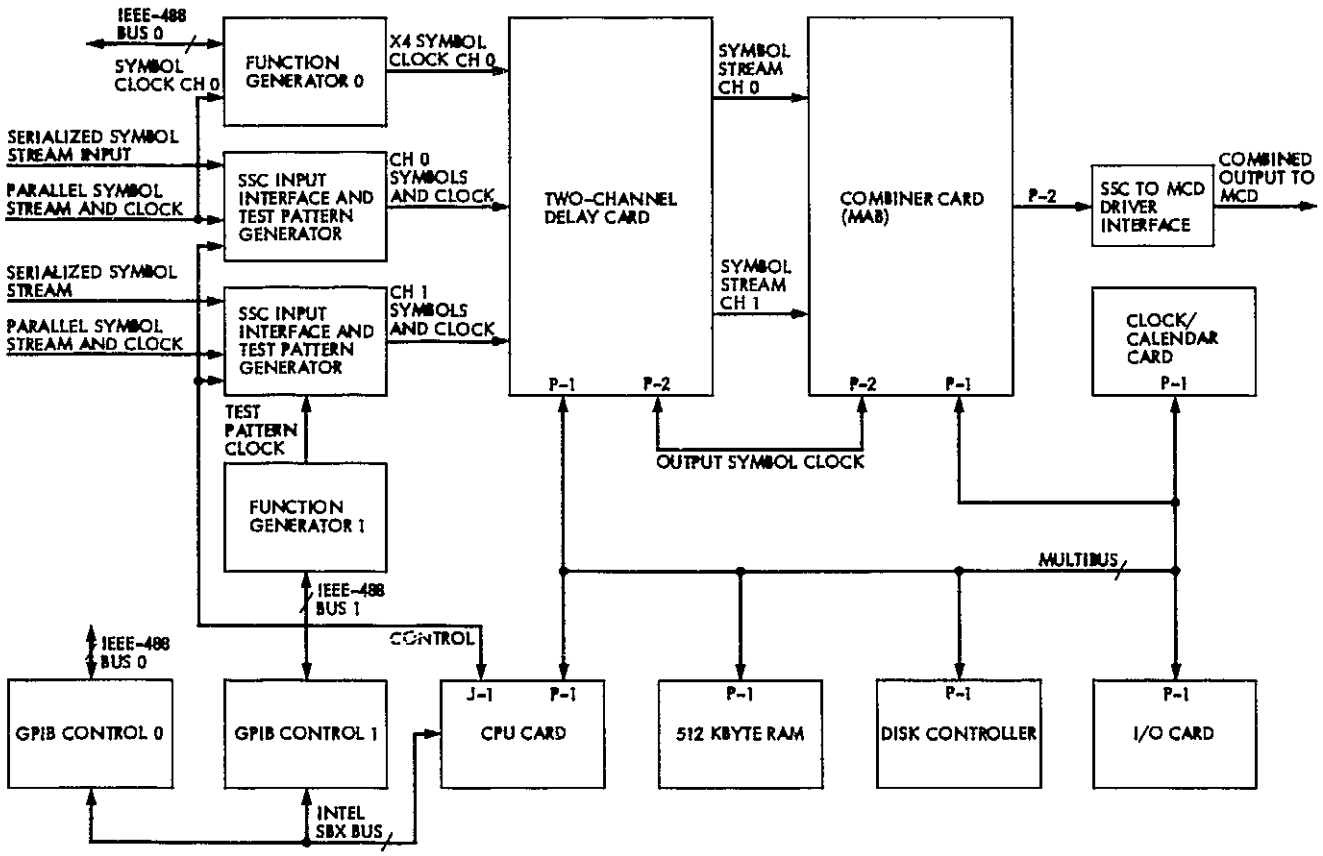


Fig. 2. Symbol-stream combiner block diagram for two-station configuration

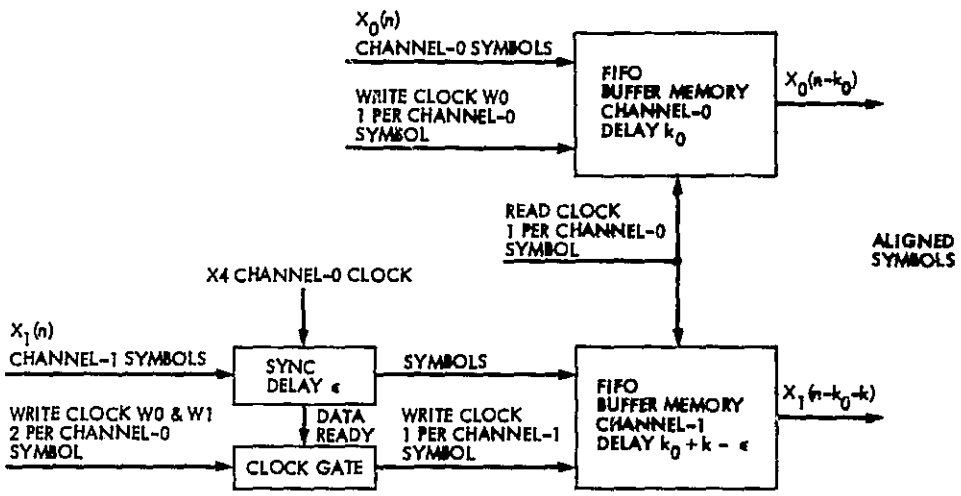


Fig. 3. Delay Card signal flow

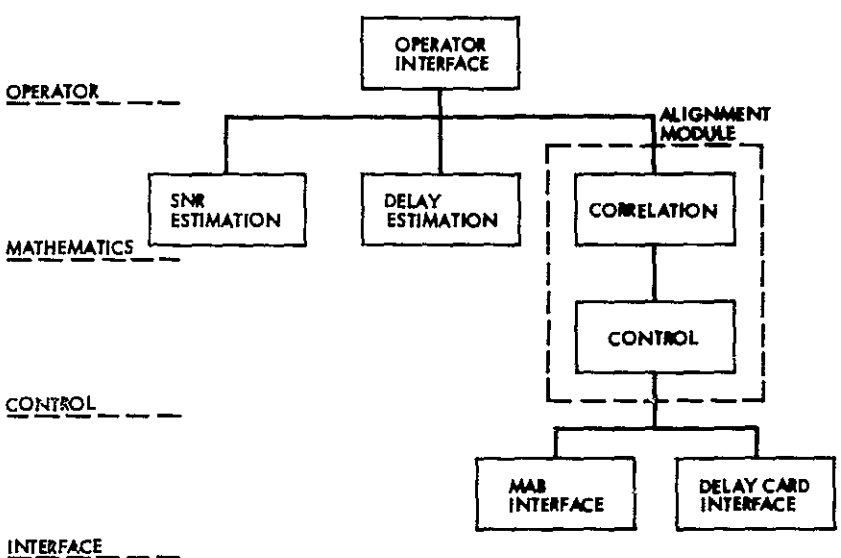


Fig. 4. Software organization

Error-Trellis Syndrome Decoding Techniques for Convolutional Codes

I. S. Reed
University of Southern California

T. K. Truong
Communication Systems Research Section

In this paper, a new error-trellis syndrome decoding technique for convolutional codes is developed. This algorithm is specialized then to the entire class of systematic convolutional codes. Finally, this algorithm is applied to the high-rate Wyner–Ash convolutional codes. A special example of the one-error-correcting Wyner–Ash code, a rate 3/4 code, is treated in this paper. The error-trellis syndrome decoding method applied to this example shows in detail how much more efficient syndrome decoding is than, say, Viterbi decoding, if applied to the same problem. For standard Viterbi decoding, 64 states would be required, whereas in the example only 7 states are needed. Also, within the 7 states required for decoding, many fewer transitions are needed between the states.

I. Introduction

This paper outlines a simplification of previous syndrome decoding methods (Refs. 1, 2) for convolutional codes (CCs). The new method involves finding minimum error paths in what is called an error tree, or its more compact equivalent, an error trellis. As will be shown, the computation of the error trellis is accomplished by finding the solution of the syndrome equations explicitly in terms of the received coded sequence. The error trellis is a graph of all path solutions of the syndrome equations. This new procedure for finding the error trellis differs from previous methods in that it does not involve an explicit computation of the syndrome.

After the error trellis has been computed, the minimum weight path in the error trellis is found by any one of many minimization techniques, including the Viterbi and sequential minimum-path-finding techniques. The minimum error path that is found by such a minimization of the path weights in the error trellis is shown to be a best estimate of the correction factor needed to correct the "noisy" message. Such a noisy message is obtained by the Massey and Sain method (Ref. 3) of applying the right inverse of the generator matrix to the received coded message.

Development of the new error trellis syndrome decoding scheme is followed by a discussion of its application to high-rate systematic convolutional codes. This application to high-rate CCs shows the real advantage of syndrome decoding over Viterbi decoding of CCs in terms of reduced complexity.

II. Syndrome Decoding With the Error Trellis

This section provides a brief development of the concepts of a convolutional code (CC) needed for systematically constructing an error trellis for minimum-error-path decoding. Here, only a brief synopsis of these concepts is given, enough

to systematically construct an error tree or trellis without resorting to the intermediate step of computing the syndrome.

The inputs and outputs of an (n, k) CC can be represented, respectively, as D-transforms,

$$x(D) = \sum_{j=0}^{\infty} x_j D^j \tag{1}$$

and

$$y(D) = \sum_{j=0}^{\infty} y_j D^j \tag{2}$$

of the input sequence of k-vectors of form $x_j = [x_{1j}, x_{2j}, \ldots, x_{kj}]$ and the output sequence of n-vectors of form $y_j = [y_{1j}, y_{2j}, \ldots, y_{nj}]$, where x_{ij} and y_{ij} belong to a finite Galois field $F = G(q)$ usually restricted to the binary field $GF(2)$ of two elements, and D is the delay operator. The input $x(D)$ and the output $y(D)$ are linearly related by means of a $k \times n$ generator matrix $G(D)$ as follows:

$$y(D) = x(D)\, G(D) \tag{3}$$

where the elements of $G(D)$ are assumed usually to be polynomials over the finite field $GF(q)$, where q is the power of a prime integer. The maximum degree M of the polynomial elements of $G(D)$ is called the memory delay of the code, and the constraint length of the code is $k = M + 1$.

In order to avoid catastrophic error propagation, the encoder matrix $G(D)$ is assumed to be *basic* (Ref. 3). This means that the Smith normal form of $G(D)$ is

$$G = A\, [I_k, 0]\, B \tag{4}$$

where $A = A(D)$ is a $k \times k$ invertible matrix with elements in $F[D]$, the ring of polynomials in D over F, and $B = B(D)$ is an $n \times n$ invertible matrix with elements in $F[D]$. The elements of the inverses A^{-1} and B^{-1} of matrices A and B, respectively, are polynomials in $F[D]$ (Ref. 4).

By definition, the parity check matrix associated with $G = G(D)$ is any full-rank $(n - k) \times n$ matrix with polynomial elements in $F[D]$ which satisfies

$$G(D)\, H^T(D) = 0 \tag{5}$$

where T denotes matrix transpose. A modification of the method of Forney (Ref. 4) is used to find H. The method involves a partitioning of matrix B in Eq. (4), as well as its inverse B^{-1}. That is, let

$$B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \tag{6}$$

and

$$B^{-1} = [\bar{B}_1, \bar{B}_2] \tag{7}$$

where the first k *rows* of B constitute the submatrix B_1 and the remaining $(n - k)$ rows are the matrix B_2, and where, likewise, the first k *columns* of B^{-1} constitute the submatrix B_1 and the remaining $(n - k)$ columns are the matrix $\bar{B}_2$.

Since B times its inverse B^{-1} is the $n \times n$ identity matrix, the following identities evidently hold:

$$\left.\begin{aligned} B_1 \bar{B}_1 &= I_k \\ B_1 \bar{B}_2 &= 0 \\ B_2 \bar{B}_1 &= 0 \\ B_2 \bar{B}_2 &= I_{n-k} \end{aligned}\right\} \tag{8}$$

In terms of the partition in Eq. (7), the Forney parity-check matrix is defined by

$$H = \bar{B}_2^T \tag{9}$$

It is readily verified using Eq. (4) and the identities of Eq. (8) that Eq. (9), in fact, satisfies Eq. (5), the requirement for H to be a parity-check matrix. It should be noted that the parity-check matrix is not unique. For example, it can be shown that $H = C\, B_2^T$ is a parity-check matrix where C is any $(n - k) \times (n - k)$ invertible matrix with elements in $F[D]$.

For an input message $x(D)$, as defined in Eq. (1), the encoded message or code sequence is $y(D)$ as generated by Eq. (3). Suppose that $y = y(D)$ is transmitted and $z = z(D)$ is received. Then, the transmitted and received sequences are related by

$$z(D) = y(D) + e(D) \tag{10}$$

where $e(D)$ is the D-transform of the error sequence. The *syndrome* of the received code $z(D)$ is

$$s(D) = z(D) \times H^T(D) \tag{11}$$

If $y(D)$ in Eq. (3) is substituted in Eq. (10), then the syndrome, computed in Eq. (11), satisfies, by Eq. (5),

$$\begin{aligned} s &= z H^T \\ &= (x G + e) H^T \\ &= e H^T \end{aligned} \tag{12}$$

This is the syndrome equation for the error sequence $e = e(D)$. The syndrome equation, Eq. (12), shows that the syndrome computed in Eq. (11) is functionally independent of the original transmitted code $y(D)$ as well as the original message $x(D)$.

The problem of syndrome decoding of convolutional codes is, as for block codes, to solve the syndrome equation, Eq. (12), for the set of all possible solutions $e = e(D)$. It has been shown (Ref. 1) that this set of solutions is a coset of the set of all codewords.

To explicitly solve the syndrome equation, Eq. (12), substitute H as given by Eq. (9) in Eq. (12), thereby obtaining

$$s = e\bar{B}_2 = e B^{-1} \begin{bmatrix} 0 \\ \\ I_{n-k} \end{bmatrix} \tag{13}$$

where I_{n-k} is the identify matrix of $(n-k)$ rows. In Eq. (13), let

$$\epsilon = e B^{-1} \tag{14}$$

so that Eq. (13) becomes the simple equation

$$s = \epsilon \begin{bmatrix} 0 \\ \\ I_{n-k} \end{bmatrix} \tag{15}$$

where $s = [s_1, s_2, \ldots, s_{n-k}]$ and $\epsilon = [\epsilon_1, \epsilon_2, \ldots, \epsilon_n]$. The *general* solution of Eq. (15) over the ring $F[D]$ is given evidently by

$$\begin{aligned} [\epsilon_1, \epsilon_2, \ldots, \epsilon_k] &= [\tau_1, \tau_2, \ldots, \tau_k] \equiv \tau \\ [\epsilon_{k+1}, \epsilon_{k+2}, \ldots, \epsilon_n] &= [s_1, s_2, \ldots, s_{n-k}] \\ &= s \end{aligned} \tag{16}$$

where $\tau_i = \tau_i(D)$ are *arbitrary* elements in $F[D]$. Thus, more compactly, the general solution of Eq. (14) is

$$\begin{aligned} \epsilon &= [\tau, s] \\ &= e B^{-1} \end{aligned} \tag{17}$$

where τ, as in Eq. (16), is an arbitrary k-vector with elements in the ring $F[D]$. Finally, a multiplication of both sides of Eq. (17) by B yields

$$\begin{aligned} e &= \epsilon B \\ &= [\tau, s] \begin{bmatrix} B_1 \\ \\ B_2 \end{bmatrix} \\ &= \tau B_1 + s B_2 \end{aligned} \tag{18}$$

in terms of submatrices B_1 and B_2 in Eq. (6) as the most general solution of the syndrome equation, Eq. (12).

The general solution, Eq. (18), of the syndrome equation can be expressed in a number of different forms. For example, it can be put into canonical form originally found heuristically by Vinck, De Paepe, and Schalkwijk (Ref. 4). Towards this end, note from the identities in Eqs. (8) and (9) that B_2^T is the left inverse, denoted by H^{-1}, of the parity-check matrix H. Hence,

$$B_2 = (H^{-1})^T \tag{19}$$

Next, note from the Smith normal form in Eq. (4) of a basic encoder that

$$\begin{aligned} A^{-1} G &= [I_k, 0] B \\ &= B_1 \end{aligned} \tag{20}$$

A substitution of B_1 in Eq. (20) and B_2 in Eq. (19) into Eq. (18) yields

$$e = \tau A^{-1} G + s(H^{-1})^T \tag{21}$$

Since τ is an arbitrary k-vector of elements in $F[D]$,

$$t = \tau A^{-1} \tag{22}$$

is also an arbitrary vector of polynomials in $F[D]$. Finally, substituting t in Eq. (22) into Eq. (21) yields,

$$e = tG + s(H^{-1})^T \tag{23}$$

as the general solution of the syndrome equation, Eq. (12), where G is the $k \times n$ generator matrix, H^{-1} is the left inverse

of the parity-check matrix, s is the $(n - k)$ component syndrome computed by Eq. (11), and t is an arbitrary k-vector with elements in $F[D]$. The above proof is a simplification of a more general version, given in Ref. 2, of the Vinck, de Paepe, and Schalkwijk identity heuristically established in Ref. 4. Herein, it is desired to put Eq. (23) in a form which makes it possible in the syndrome decoding process to bypass the explicit computation of the syndrome $s(D)$.

Towards this end, substitute Eq. (19) into Eq. (23) and, by Eqs. (9) and (11), the quantity $z\bar{B}_2$ for the syndrome s. These substitutions yield

$$e = tG + z(\bar{B}_2 B_2) \tag{24}$$

in terms of received sequence z as the general solution of the syndrome equation.

In Eq. (24), let R be the $n \times n$ matrix $B_2 B_2$, since B_2 and $\bar{B}_2$ have ranks $(n - k)$, it can be shown that the matrix $R = \bar{B}_2 B_2$, where B_2 and $\bar{B}_2$ are defined in Eqs. (6) and (7), respectively, also has rank $(n - k)$. Substituting R into Eq. (24) yields

$$e = tG + zR \tag{25}$$

as the general solution of the syndrome equation. Here, R is the $n \times n$, rank $(n - k)$ matrix

$$R = \bar{B}_2 B_2 \tag{26}$$

t is an arbitrary k-vector of elements in $F[D]$, and z is the D-transform of the received sequence.

Let $z(D)$ be any finite-length received sequence. By the maximum likelihood principle, the most likely error sequence is the one with minimum Hamming weight. Given $z(D)$, the sequence $e(D)$ with minimum Hamming weight is found by minimizing the weight of the right side of Eq. (25) over all polynomials $t(D)$ in $F[D]$. That is,

$$\min \| e \| = \min \| tG + zR \|, \qquad t \in F[D] \tag{27}$$

where $z = z(D)$ is the D-transform or polynomial of any finite-length received sequence and $\| x \|$ denotes the Hamming weight or "norm" of an element $x = x(D)$ in $F[D]$.

The minimization required in Eq. (27) is analogous to certain optimum nulling techniques in control theory. The sequence $r(D) = z(D)\, R(D)$ is the error sequence for $t(D) = 0$. What one attempts to do in Eq. (27) is to find that sequence $\hat{t}$ which, when encoded as $\hat{t}\, G$ and subtracted from $r(D)$, yields the sequence $\hat{e}$ of minimum Hamming weight. That is, if $\hat{t} = \hat{t}(D)$ is the D-transform for which $\| e \| = \| tG + zR \|$ is minimum, then

$$\hat{e} = \hat{t}G + zR \tag{28}$$

is the D-transform of the minimum-weight-possible error sequence.

By Eq. (4), the right inverse G^{-1} of the generating matrix G is

$$G^{-1} = B^{-1} \begin{bmatrix} I_k \\ 0 \end{bmatrix} A^{-1} \tag{29}$$

This is verified by multiplying G in Eq. (4) on the right by G^{-1} in Eq. (29). Multiplying both sides of Eq. (28) on the right by G^{-1} in Eq. (29) yields, by Eqs. (7) and (8), the identity

$$\begin{aligned} \hat{e}G^{-1} &= [\hat{t}G + z\bar{B}_2 B_2]\, G^{-1} \\ &= \hat{t} + z\bar{B}_2 B_2 [\bar{B}_1, B_2] \begin{bmatrix} I_k \\ 0 \end{bmatrix} A^{-1} \\ &= \hat{t} + z\bar{B}_2 [0, I_{n-k}] \begin{bmatrix} I_k \\ 0 \end{bmatrix} A^{-1} \\ &= \hat{t} \end{aligned} \tag{30}$$

By Eq. (10), the subtraction of $\hat{e}$ from z produces a best estimate $\hat{y}$ of the transmitted code, i.e.,

$$\hat{y} = z - \hat{e} \tag{31}$$

The best estimate $\hat{y}$ of the code, if multiplied on the right by G, yields

$$\hat{x} = \hat{y}G^{-1} \tag{32}$$

which is the best estimate of the original message. Hence, substituting Eq. (31) in Eq. (32) and using Eq. (30) produces

$$\begin{aligned} \hat{x} &= (z - \hat{e})\, G^{-1} \\ &= z\, G^{-1} - \hat{t} \end{aligned} \tag{33}$$

This important identity shows that $\hat{t} = \hat{t}(D)$, obtained by the minimization in Eq. (27), is a *correction factor* to the standard method of recovering the message from $z = z(D)$ if z were noise-free.

In the following section, the techniques of performing the minimization in Eq. (27) for finding $\hat{e}$ and $\hat{t}$ are discussed. Among these methods are the Viterbi dynamic programming algorithm and some of the sequential decoding techniques. Then, the syndrome-decoding algorithm described above is applied to systematic high-rate CCs and, in particular, to the one-error-correcting CC developed originally by Wyner and Ash (Ref. 5).

III. Syndrome Decoding of Systematic Convolutional Codes

The results of the preceding section are now applied to systematic convolutional codes. The generator matrix for a systematic CC has form

$$G(D) = [I_k, P(D)] \tag{34}$$

where I_k is the $k \times k$ identity matrix and $P(D)$ is a $k \times (n-k)$ matrix of polynomials over $GF(q)$ in the delay operator D. Again, as in the general case, the maximum degree M of the polynomials in $P(D)$ is called the memory of the code and $K = M + 1$ is the constraint length.

A parity-check matrix associated with $G(D)$ in Eq. (34) is the $(n-k) \times n$ matrix,

$$H(D) = [-P^T(D), I_{n-k}] \tag{35}$$

This follows from the fact that $H(D)$ has rank $n-k$ and that it satisfies Eq. (5).

The Smith formal form of Eq. (34) is, by Eq. (4),

$$\begin{aligned} G &= A\,[I_k, 0]\,B \\ &= [I_k, 0] \begin{bmatrix} I_k, P \\ 0, I_{n-k} \end{bmatrix} \end{aligned} \tag{36}$$

where $P = P(D)$, the matrix of polynomials in the generator matrix $G(D)$ in Eq. (34). Hence, for a systematic code, $A = I_k$ and

$$B = \begin{bmatrix} I_k, P \\ 0, I_{n-k} \end{bmatrix} \tag{37}$$

Because of the triangular form of B, the inverse is readily found to be

$$B^{-1} = \begin{bmatrix} I_k, -P \\ 0, I_{n-k} \end{bmatrix} \tag{38}$$

which actually equals B when the field of coefficients is the binary field $GF(2)$.

The partitions, given in Eqs. (6) and (7), of B and B^{-1}, respectively, are, for a systematic CC,

$$B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}$$

where

$$\begin{aligned} B_1 &= [I_k, P(D)] \\ B_2 &= [0, I_{n-k}] \end{aligned} \tag{39}$$

and

$$B^{-1} = [\bar{B}_1, \bar{B}_2]$$

where

$$\begin{aligned} \bar{B}_1 &= \begin{bmatrix} I_k \\ 0 \end{bmatrix} \\ \bar{B}_2 &= \begin{bmatrix} -P \\ I_{n-k} \end{bmatrix} \end{aligned} \tag{40}$$

Note that the parity-check matrix found by Eq. (9) from $\bar{B}_2$ in Eq. (40) actually equals the parity-check matrix found already in Eq. (35) by satisfying Eq. (5). As a consequence, for a systematic CC, the syndrome s in Eq. (12) is

$$\begin{aligned} s &= z H^T \\ &= z \begin{bmatrix} -P \\ I_{n-k} \end{bmatrix} \\ &= [z_m, z_p] \begin{bmatrix} -P \\ I_{n-k} \end{bmatrix} \\ &= -z_m(D) P(D) + z_p(D) \end{aligned} \tag{41}$$

where $z_m(D)$ is the message code vector of k components, possibly corrupted by noise, and $z_p(D)$ is an $(n-k)$ component vector of parity symbols, also possibly changed by channel noise.

Next, by Eqs. (39) and (40), the matrix R in Eq. (26) is given by

$$R = \bar{B}_2 B_2$$

$$= \begin{bmatrix} -P \\ I_{n-k} \end{bmatrix} [0, I_{n-k}]$$

$$= \begin{bmatrix} 0, -P \\ 0, I_{n-k} \end{bmatrix} \tag{42}$$

Thus, for a systematic CC, the general solution, Eq. (25), of the syndrome equation, Eq. (12), is, by substituting Eqs. (34) and (42) into Eq. (25),

$$e(D) = t G + z R$$

$$= t [I_k, P] + z \begin{bmatrix} 0, -P \\ 0, I_{n-k} \end{bmatrix}$$

$$= [t I_k, t P(D)] + \left[0, z \begin{bmatrix} -P \\ I_{n-k} \end{bmatrix} \right]$$

$$= [t(D), (t(D) - z_m(D)) P(D) + z_p(D)] \tag{43}$$

where $z_m(D)$ is the received message sequence "in the clear," $z_p(D)$ is the received parity sequence of the CC, and $t(D)$ is an element of $F[D]$. By Eq. (41), the above general solution, Eq. (43), of the syndrome equation for a systematic CC can be expressed in the alternate form

$$e(D) = [t(D), t(D) P(D) + s(D)] \tag{44}$$

where $s(D)$ is the syndrome, computed by Eq. (41) in terms of $z_m(D)$ and $z_p(D)$.

The "best" correction factor $\hat{t}(D)$ for all systematic CCs is found, as in Eq. (27), by minimizing the Hamming weight of $e(D)$, given in Eqs. (43) of (44). For low-rate systematic CCs, this minimization can be taken over $F[D]$, whereas for high-rate systematic CCs, this minimization need only be accomplished over a small subset, call it E, of $F[D]$, defined by error-bound constraints of the particular CC. This latter fact for high-rate systematic CCs will be demonstrated for the one-error correcting Wyner-Ash CC (Ref. 6). It is the very small size of the set E compared to the set $F[D]$ which makes syndrome decoding more efficient than the classical maximum likelihood method for decoding CCs.

Let $\hat{e}$ denote the error sequence of the solution, Eq. (44), of minimum Hamming weight, and let $\hat{t}$ be the element $t(D) \in F(D)$, for which the Hamming weight of $e(D)$ in Eq. (43) or Eq. (44) is minimum. Then, by Eqs. (43) and (44), as in Eq. (28), $\hat{e}$ and $\hat{t}$ are related by

$$\hat{e} = [\hat{t}, (\hat{t} - z_m) P + z_p]$$

$$= [\hat{t}, \hat{t} P + s] \tag{45}$$

By Eqs. (29), (36), and (38), the right inverse of the generator matrix G in Eq. (34) is

$$G^{-1} = B^{-1} \begin{bmatrix} I_k \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} I_k, & -P \\ 0, & I_{n-k} \end{bmatrix} \begin{bmatrix} I_k \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} I_k \\ 0 \end{bmatrix} \tag{46}$$

Hence, by Eqs. (45) and (46), the relation

$$\hat{e} G^{-1} = \hat{t}$$

given in Eq. (30), also holds for systematic CCs. Again, the subtraction of $\hat{e}$ from z produces

$$\hat{y} = z - \hat{e}$$

as the best estimate of transmitted code, so that

$$\begin{aligned}\hat{x} &= \hat{y}\,G^{-1} \\ &= (z - \hat{e})\,G^{-1} \\ &= z\,G^{-1} - \hat{t} \\ &= [z_m, z_p]\begin{bmatrix} I_k \\ 0 \end{bmatrix}\hat{t} \\ &= z_m - \hat{t} \end{aligned} \tag{47}$$

as the best estimate of the received message in terms of z_m, the received message "in the clear," and the correction factor, $\hat{t}$.

IV. Error Trellis Syndrome Decoding of Wyner-Ash Convolutional Code

The Wyner-Ash one-error correction codes were first defined in Ref. 5. A more modern and understandable development can be found in Blahut's recent book (Ref. 7). Instead of defining the CC only in terms of its infinite generator or parity-check matrix, as is done in Ref. 7, here the infinite matrices are converted first into compact matrices in terms of the delay operator D. If

$$G(D) = G_0 + G_1 D + \ldots, + G_m D^m \tag{48}$$

is a generator matrix of a CC of memory $M = m$, as defined in Eq. (3), then evidently

$$G = \begin{bmatrix} G_0 & G_1 & G_2 & \ldots & G_m & 0 & 0 & \ldots \\ 0 & G_0 & G_1 & G_2 & \ldots & G_m & 0 & \ldots \\ 0 & 0 & G_0 & G_1 & G_2 & \ldots & G_m & \ldots \\ \cdot & \cdot & \cdot & \cdot & \cdot & & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \cdot & & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \cdot & & \cdot & \end{bmatrix} \tag{49}$$

is the infinite generator matrix associated with $G(D)$. Thus, a systematic code with generator matrix $G(D) = [I_k, P(D)]$ has

$$G = \begin{bmatrix} I_k & P_0 & 0 & P_1 & 0 & P_2 & \ldots & 0 & P_m & & \\ & I_k & P_0 & 0 & P_1 & 0 & P_2 & \ldots & 0 & P_m & \\ & & I_k & P_0 & 0 & P_1 & 0 & P_2 & \ldots & 0 & P_m \\ & & & & \cdot & \cdot & \cdot & & & & \\ & & & & & \cdot & \cdot & \cdot & & & \\ & & & & & & \cdot & \cdot & \cdot & & \end{bmatrix} \tag{50}$$

as its companion infinite generator matrix, where

$$P(D) = P_0 + P_1 D + \ldots, P_m D^m \tag{51}$$

where 0 is the $k \times k$ all-zero matrix and P_i are $k \times k\,(n - k)$ matrices. Since, by Eq. (35), $H(D) = [-P^T(D), I_{n-k}]$ is a parity-check matrix of $G(D)$, given above, the associated infinite parity-check matrix is

$$H = \begin{bmatrix} P_0^T & I & & & & & & \\ P_1^T & 0 & P_0^T & I & & & & \\ P_2^T & 0 & P_1^T & 0 & P_0^T & I & & \\ \cdot & \cdot & & & & & & \\ \cdot & \cdot & P_2^T & 0 & P_1^T & 0 & \cdot & \\ \cdot & \cdot & & & & & & \cdot \\ P_m^T & 0 & \cdot & \cdot & & & & \cdot \\ & & \cdot & \cdot & & & & \\ & & \cdot & \cdot & \cdot & \cdot & & \\ & & P_m^t & 0 & \cdot & \cdot & & \\ & & & & \cdot & \cdot & & \\ & & & & P_m^T & 0 & & \\ & & & & & & \cdot & \\ & & & & & & & \cdot \\ & & & & & & & \cdot \end{bmatrix} \tag{52}$$

The results in Eqs. (50) and (52) are given in Ref. 6 in the same notation.

In terms of Eqs. (51) and (52), Blahut defines an $(n, k) = (2^m, 2^m - 1)$ Wyner-Ash code as follows: Let H^1 be the parity-check matrix of the binary $(2^m - 1, 2^m - 1 - m)$ Hamming one-error-correcting block code. Choose matrices $P_1^T, P_2^T, \ldots, P_m^T$ to be the m rows of the parity-check matrix H^1, i.e.,

$$H^1 = \begin{bmatrix} P_1^T \\ P_2^T \\ \cdot \\ \cdot \\ \cdot \\ P_m^T \end{bmatrix} = [P_1, P_2, \ldots, P_m]^T \qquad (53)$$

Finally, let P_0^T be a vector of $2^m - 1$ ones, i.e.,

$$P_0^T = \underbrace{[1, 1, 1, \ldots, 1]}_{2^m - 1} \qquad (54)$$

Blahut shows (Ref. 7, Theorem 12.5.1) that the minimum distance of the Wyner-Ash code is 3 and, as a consequence, it will correct at least one error. To understand this code in more detail and to apply the decoding technique developed in the last section to it, consider now an example for $m = 2$.

Example: The $m = 2$, the parity-check matrix of the Hamming code, is

$$H^1 = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

so that, by Eqs. (53) and (54), $P_0^T = [1\ 1\ 1]$, $P_1^T = [1\ 1\ 0]$, and $P_2^T = [1\ 0\ 1]$. Thus, by Eqs. (51),

$$P(D) = \begin{bmatrix} 1 + D + D^2 \\ 1 + D \\ 1 \qquad + D^2 \end{bmatrix}$$

and, by Eqs. (34) and (35),

$$G(D) = \begin{bmatrix} 1\ 0\ 0, & 1 + D + D^2 \\ 0\ 1\ 0, & 1 + D \\ 0\ 0\ 1, & 1 \qquad + D^2 \end{bmatrix}$$

and

$$H(D) = [1 + D + D^2, 1 + D, 1 + D^2, 1] \qquad (55)$$

are the generator and parity-check matrices of the (4, 3) Wyner-Ash CC, respectively. Also, by Eqs. (37) and (38)

$$B = B^{-1} = \begin{bmatrix} I_k, P(D) \\ 0, I_{n-k} \end{bmatrix} = \begin{bmatrix} 1\ 0\ 0, & 1 + D + D^2 \\ 0\ 1\ 0, & 1 + D \\ 0\ 0\ 1, & 1 \qquad + D^2 \\ 0\ 0\ 0, & 1 \end{bmatrix}$$

so that, by Eqs. (39) and (40), $B_2 = [0\ 0\ 0\ 1]$ and $\bar{B}_2 = H^T$ and, finally, by Eq. (42),

$$R = \bar{B}_2 B_2 \equiv \begin{bmatrix} 1 + D + D^2 \\ 1 + D \\ 1 \qquad + D^2 \\ 1 \end{bmatrix} [0\ 0\ 0\ 1] = \begin{bmatrix} 0\ 0\ 0, & 1 + D + D^2 \\ 0\ 0\ 0, & 1 + D \\ 0\ 0\ 0, & 1 \qquad + D^2 \\ 0\ 0\ 0, & 1 \end{bmatrix} \qquad (56)$$

The above results for this 3/4 rate CC can now be used to explicitly obtain the general solution $e(D)$ in Eq. (43) of the syndrome equation. This is accomplished by substituting Eqs. (55) and (56) into Eq. (25) or directly from Eq. (43). The result is

$$\begin{aligned} e(D) \equiv e &= [e_1, e_2, e_3, e_4] \\ &= [t, (t_1 + z_1)(1 + D + D^2) \\ &\quad + (t_2 + z_2)(1 + D) \\ &\quad + (t_3 + z_3)(1 + D^2) + z_4] \end{aligned} \qquad (57)$$

where

$$t(D) \equiv t = [t_1, t_2, t_3] \tag{58}$$

By Eqs. (41) and (44), e in Eq. (57) can also be expressed more compactly as

$$e = [t, r+s] \tag{59}$$

where s is the syndrome,

$$s(D) \equiv s = z_1(1+D+D^2) + z_2(1+D) + z_3(1+D^2) + z_4 \tag{60}$$

and

$$r(D) \equiv r = t_1(1+D+D^2) + t_2(1+D) + t_3(1+D^2) \tag{61}$$

Note that the term $r(D) \equiv r$ in Eqs. (59) and (61), in order to minimize the Hamming weight of $e(D)$, must be chosen to "cancel $s(D)$" in Eq. (59). For this reason, one might call $r(D)$ the *regulator* needed to cancel the syndrome $s(D)$.

Now, the formal power series for $e(D)$ in the delay operator is explicitly

$$e(D) = [e_1(D), e_2(D), \ldots, e_n(D)] = \sum_{j=0}^{\infty} [e_{1j}, e_{2j}, \ldots, e_{nj}]\, D^j \tag{62}$$

Define the truncation of $e(D)$ at stage or frame time N in terms of Eq. (62) as

$$[e(D)]_N = \sum_{j=0}^{N} [e_{1j}, e_{2j}, \ldots, e_{nj}]\, D^j \tag{63}$$

Thus, the Hamming weight of the sequence of possible errors in N frames is

$$\| [e(D)]_N \| = \sum_{j=0}^{N} \| [e_{1j}, e_{2j}, \ldots, e_{nj}] \| = \sum_{j=0}^{N} \| \underset{D^j}{\text{coef}} [e(D)] \| \tag{64}$$

where the latter expression under the summation is the Hamming weight of the coefficient of the jth power of D.

By Eqs. (57) and (64) for this particular example of convolutional code,

$$\underset{D^j}{\text{coef}} [e(D)] = [t_{1j}, t_{2j}, t_{3j}, r_j + s_j] \tag{65}$$

where

$$r_j = t_{1,j} + t_{1,j-1} + t_{1,j-2} + t_{2,j} + t_{2,j-1} + t_{3,j} + t_{3,j-2} \tag{66}$$

is the regulator function at frame j, and

$$s_j = z_{1j} + z_{1,j-1} + z_{1,j-2} + z_{2j} + z_{2,j-1} + z_{3j} + z_{3,j-2} + z_{4j} \tag{67}$$

is the syndrome function at frame j in terms of binary variables t_{ij} and z_{ij}, defined by,

$$r(D) = \sum_{i=0}^{\infty} [t_{1j}, t_{2j}, \ldots, t_{kj}] D^j = \sum_{i=0}^{\infty} \underline{t}_j D^j$$

$$z(D) = \sum_{i=0}^{\infty} [z_{1j}, z_{2j}, \ldots, z_{nj}] D^j$$

Note in Eq. (66) that r_j is a function of $\underline{t}_j = [t_{1j}, t_{2j}, t_{3j}]$, $\underline{t}_{j-1} = [t_{1,j-1}, t_{2,j-1}, t_{3,j-1}]$ and $\underline{t}_{j-2} = [t_{1,j-2}, t_{2,j-2}, t_{3,j-2}]$. That is, at frame j, r_j is a function of $\underline{t}_j$ at frame j; a function of $\underline{t}_{j-1}$ at frame $j-1$; and a function of $\underline{t}_{j-2}$ at frame $j-2$; i.e.,

$$r_j = r(\underline{t}_j, \underline{t}_{j-1}, \underline{t}_{j-2}) \tag{68}$$

If the values of the regulator function r_j at frame j are imagined to be generated by a sequential circuit, then the pair

$$\sigma_j = (\underline{t}_{j-1}, \underline{t}_{j-2}) \tag{69}$$

constitutes the values of the internal state of the circuit and vector $\underline{t}_j$ is the jth input to the circuit.

Let the sequential circuit with output

$$u_j = [\underline{t}_j, r(\underline{t}_j, \sigma_j)] \tag{70}$$

be the regulator circuit of the decoder, where σ_j is the internal state defined by Eq. (69). Also, call the set of all allowable paths generated by Eq. (70) the regulator tree or trellis. Finally, by Eq. (59), the error trellis of the code is, for all paths generated,

$$v_j = [\underline{t}_j, s_j + r(\underline{t}_j, \sigma_j)] \tag{71}$$

To illustrate the above concepts, let the input to the present example of the (4, 3) CC be

$$x = [1\ 1\ 1,\ 0\ 0\ 0,\ 1\ 1\ 1,\ 0\ 0\ 0,\ 1\ 1\ 1]$$

i.e., $x_1 = [1\ 0\ 1\ 0\ 1] = x_2 = x_3$. By the generating matrix given in Eq. (55), the output $y = [y_1, y_2, y_3, y_4]$ is obtained in what follows: $y_1 = y_2 = y_3 = x_1 = [1\ 0\ 1\ 0\ 1]$, and $y_4 = (1 + D + D^2)\,x_1 + (1 + D)\,x_2 + (1 + D^2)\,x_3$. Explicitly, y_4 is computed from this relation as follows:

$$\begin{aligned}
x_1 &: \ 1\ 0\ 1\ 0\ 1 \\
Dx_1 &: \ \ 1\ 0\ 1\ 0\ 1 \\
D^2x_1 &: \ \ \ 1\ 0\ 1\ 0\ 1 \\
x_2 &: \ 1\ 0\ 1\ 0\ 1 \\
Dx_2 &: \ \ 1\ 0\ 1\ 0\ 1 \\
x_3 &: \ 1\ 0\ 1\ 0\ 1 \\
D^2x_3 &: \ \ \ 1\ 0\ 1\ 0\ 1 \\
y_4 &= [1\ 0\ 1\ 0\ 1\ 0\ 0]
\end{aligned}$$

Thus, the output of the encoder is

$$y = [1\ 1\ 1\ 1, 0\ 0\ 0\ 0, 1\ 1\ 1\ 1, \\ 0\ 0\ 0\ 0, 1\ 1\ 1\ 1] \tag{72}$$

Assume y, given in Eq. (72), is transmitted over a binary symmetric channel with probability of error somewhat less than 1/12 = 0.0833 .... Then, suppose that the received coded sequence is

$$z = [1\ 1\ 0\ 1, 0\ 0\ 0\ 0, 1\ 1\ 1\ 1, \\ 0\ 0\ 0\ 0, 0\ 1\ 1\ 1] \tag{73}$$

i.e., $z_1 = [1\ 0\ 1\ 0\ 0]$, $z_2 = [1\ 0\ 1\ 0\ 1]$, $z_3 = [0\ 0\ 1\ 0\ 1]$ and $z_4 = [1\ 0\ 1\ 0\ 1]$. By Eq. (60), the syndrome sequence for this value of received sequence is computed to be

$$s = [1\ 0\ 1\ 0\ 1\ 1\ 1] \tag{74}$$

by the same method used above to obtain y_4.

It is shown in Ref. 7 (p. 366) that the rate 3/4 code of this example can correct one error in every 3 frame times or code length of 12. As a consequence, one needs only to correct one error every 3 frames. This limits the number of values of $t = [t_1, t_2, t_3]$ to 4, namely the values

$$[0\ 0\ 0] \equiv 0, \quad [1\ 0\ 0] \equiv 1 \\ [0\ 1\ 0] \equiv 2, \quad [0\ 0\ 1] \equiv 3 \tag{75}$$

Note that the four values of t in Eq. (75) allow for, at most, one error, and that these four values are conveniently labeled by the integers $t = 0, 1, 2$, or 3.

Figure 1 shows a constrained regulator trellis with outputs $[t, r]$. In Fig. 1, note that, because of the limited error-correction capability of the code, the number of internal states $\sigma = (Dt, D^2t)$ of the regulator circuit can be limited to 7 out of a possible 64. Moreover, the number of state transitions can be limited to those shown in Fig. 1 for the regulator trellis. The branches of the regulator trellis are labeled with the value $[t, r]$. For example, the branch from state $\sigma = [0\ 0]$ to $\sigma = [3\ 0]$ is labeled by $[t, r] = [3, 1] \equiv [0, 0, 1, 1]$, which means $t_1 = 0$, $t_2 = 0$, $t_3 = 1$, and $r = 1$.

To decode the message in Eq. (73), by Eq. (70) an error trellis is created by adding the vector $[0, s]$ to all labels in the regulator trellis where s is the syndrome value. Thus, in Fig. 2, the values of $[0, s]$, where s is the syndrome value in Eq. (74), appear on all possible transitions $\sigma = [0\ 0]$ to $\sigma = [0\ 0]$ on the top line of the error trellis. At each node, the

cumulative Hamming weight of the path, passing through that node, is written.

The Hamming weight at each node, plus the weight of a possible branching from that to the next node, is used to eliminate branches. The technique is similar to the method in Viterbi decoding for eliminating branches. To illustrate, in Fig. 2 there are four branches at Frame 2 which could go to state or node $\sigma = [0\ 0]$. The transition is chosen as the branch from $\sigma = [0\ 3]$ to $\sigma = [0\ 0]$ since the node weight 2 plus branch weight 0 is 2, the *minimum* of the 4 possible transitions.

The minimum overall path weight of the error trellis in Fig. 2 is

$$[0\ 0, 3\ 0, 0\ 3, 0\ 0, 0\ 0, 1\ 0, 0\ 1, 0\ 0, 0\ 0]$$

in terms of state values $\sigma = Dt, D^2t$. Hence, based on the criterion of Eq. (27), the best estimate of t is

$$\begin{aligned}\hat{t} &= [3, 0, 0, 0, 1, 0, 0, 0] \\ &= [0\ 0\ 1, 0\ 0\ 0, 0\ 0\ 0, 0\ 0\ 0, \\ &\quad 1\ 0\ 0, 0\ 0\ 0]\end{aligned}$$

If this vector is added component-wise to z in Eq. (73), the message is corrected to yield $\hat{x} = x$, the original message.

References

1. Reed, I. S. and Truong, T. K., "New Syndrome Decoding for $(n, 1)$ Convolutional Codes," *Electronic Letters*, Vol. 19, No. 9, April 1983, pp. 344-346.
2. Reed, I. S. and Truong, T. K., "New Syndrome Decoding Techniques for Convolutional Codes Over $GF(q)$," to be published in *Proceedings IEE.*
3. Massey, J. L. and Sain, M. K., "Inverses of Linear Sequential Circuits," *IEEE Trans. Comput. C-17*, pp. 330-337, April 1968.
4. Forney, C. D., Jr., "Convolutional Codes I: Algebraic Structure," *IEEE Trans. Info. Theor. IT-9*, 1963, pp. 64-74.
5. Vinck, A. J., de Paepe, A. J. P., and Schalkwijk, J. P. M., "A Class of Binary Rate One-Half Convolutional Codes that Allows an Improved Stack Decoder," *IEEE Trans. Info. Theor. IT-26*, No 4, 1980, pp. 389-392.
6. Wyner, A. D. and Ash, R. B., "Analysis of Recurrent Codes," *IEEE Trans. Info. Theor. IT-9*, 1963, pp. 143-156.
7. Blahut, R. E., *Theory and Practice of Error Control Codes*, Addison-Wesley, London, 1983.

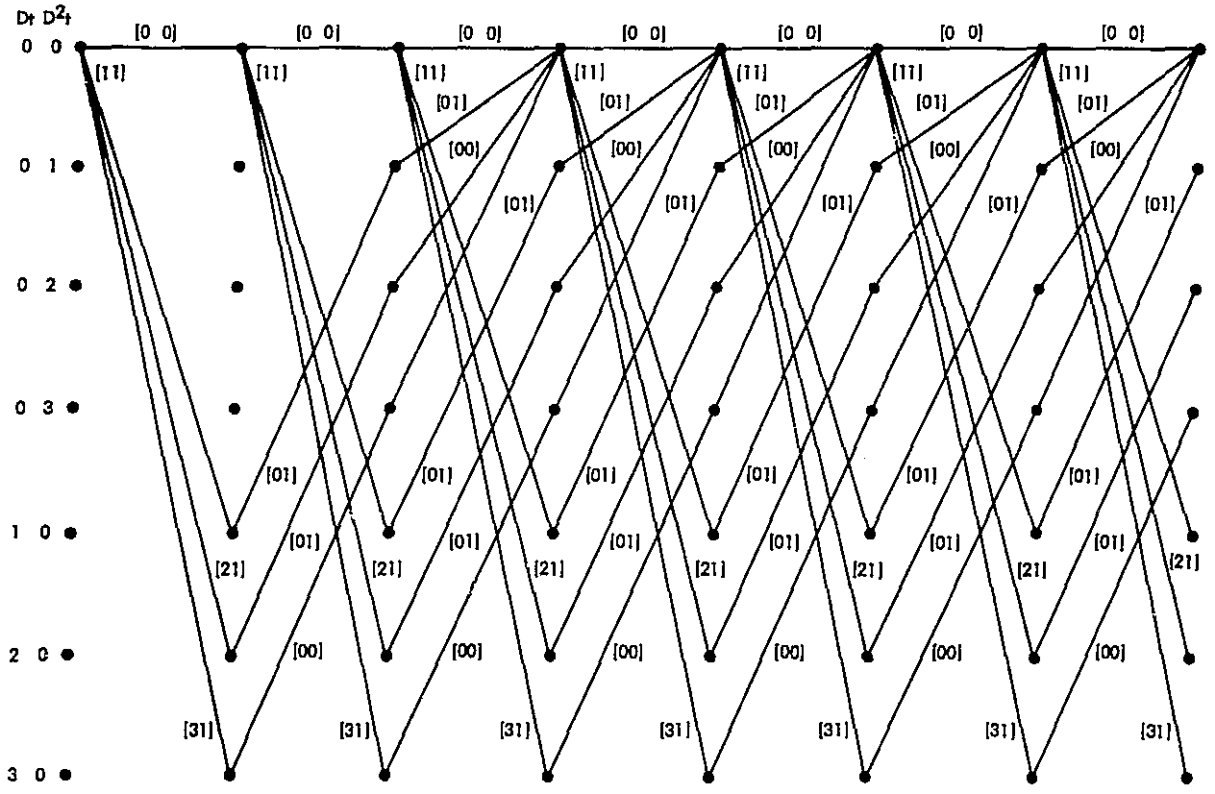


Fig. 1. Constrained regulator trellis with outputs $[t, r]$, where $r = t_1(1 + D + D^2) + t_2(1 + D) + t_3(1 + D^2) \equiv r(t, \sigma)$ and where $t = [t_1, t_2, t_3] = [0\ 0\ 0] \equiv 0$, $t = [1\ 0\ 0] \equiv 1$, $t = [0\ 1\ 0] \equiv 2$, and $t = [0\ 0\ 1] \equiv 3$

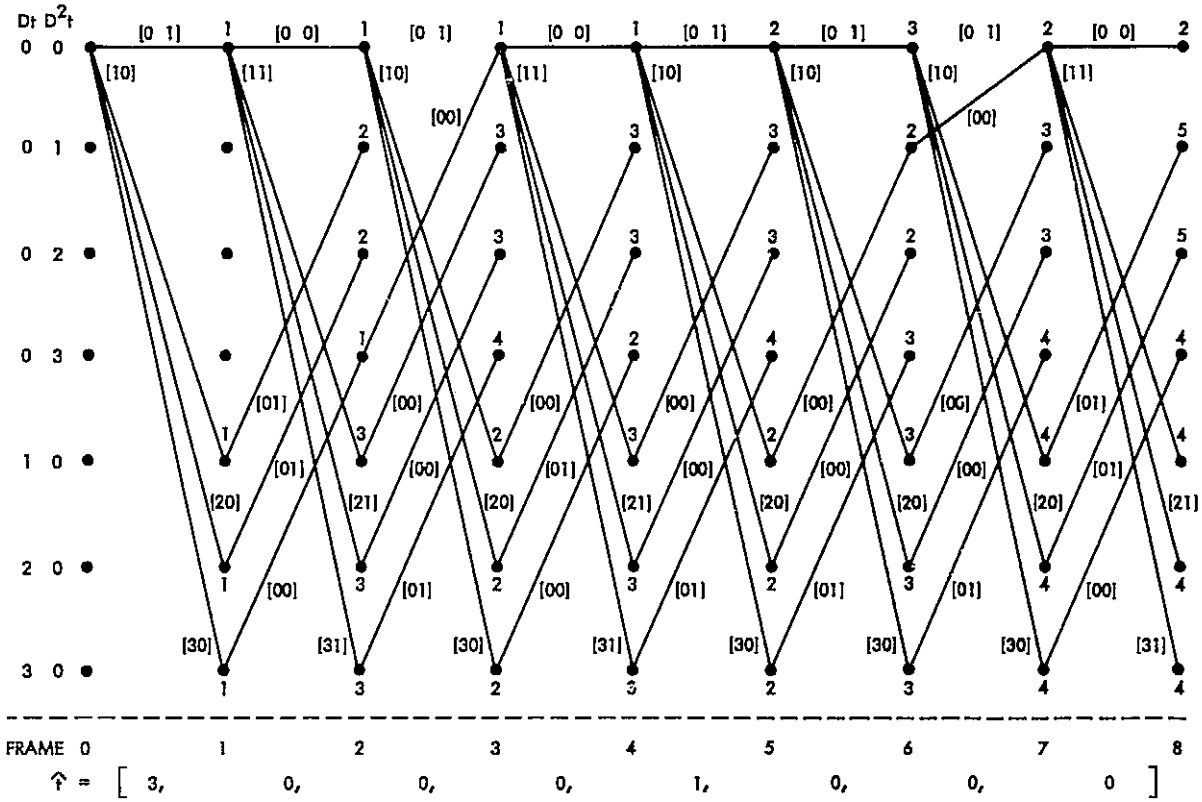


Fig. 2. Error trellis with input and state-transition constraints for one-error-correcting Wyner-Ash convolutional code

Coherent Digital Demodulation of a Residual Carrier Signal Using IF Sampling

R. Sfeir, S. Aguirre, and W. J. Hurd
Communications Systems Research Section

Analysis is presented of an all-digital technique for the coherent demodulation of a residual carrier signal with a biphase modulated square-wave subcarrier. The processing technique, proposed for use in the DSN advanced receiver, employs the concept of IF sampling. It also uses an optimum Costas loop for subcarrier demodulation and data-aided carrier tracking, i.e., combined Costas and residual carrier tracking. It is shown that the loops perform essentially the same as the corresponding analog loops in terms of signal-to-noise ratio and loop bandwidth. Furthermore, the sampling does not introduce biases or other significant effects on the loops provided that the loop bandwidth is very small compared to the symbol rate, and that the number of samples per symbol is large compared to inverse loop bandwidth.

I. Introduction

The ever increasing advances in LSI and VLSI technology make possible the use of digital integrated circuits in real-time processing areas that were previously implemented with analog circuitry. This paper describes a digital implementation proposed for the DSN Advanced Receiver to avoid the inherent problems associated with analog systems such as dc offsets in the mixers and amplifiers, the need for calibration and adjustments, and less reliability, versatility, and flexibility than digital systems.

Two families of configurations have been considered for the design of the proposed advanced receiver: one associated with I-Q baseband sampling and another associated with IF sampling. In the first one, the sampled I and Q channels are generated first by demodulating by two in-quadrature reference signals and then sampling with two in-phase samplers; in the second, we first sample at IF and then demultiplex into baseband I and Q samples. The IF sampling configuration was chosen for three main reasons. First, it overcomes the dc offset problem of baseband systems, which is a serious problem in phase detection at low signal-to-noise ratios. Second, the hardware is simpler and dc amplifiers are eliminated. Third, the implementation concept is not proven, and is thus deserving of analysis and demonstration. One complexity and potential disadvantage of IF sampling is that the samples are offset from each other by one-half sampling period. It is shown that this I-Q offset inherent to IF sampling has no significant effect on the loops' performances.

The digital system employs synchronous bandpass sampling for the coherent demodulation of a residual carrier signal biphase modulated by a square-wave subcarrier. The receiver uses a PLL to track the residual carrier phase, an optimum Costas loop with integrate and dump filters for subcarrier demodulation, and a data-aided loop for improved tracking performance of the carrier phase.

II. Functional Block Diagram and Description

Figure 1 shows the functional block diagram of the proposed receiver. The received signal is at an intermediate frequency f_i and contains data that biphase modulates a square-wave subcarrier. The modulated subcarrier phase modulates the carrier with a modulation index Δ. The signal at the output of the bandpass filter is:

$$r(t) = \sqrt{2P}\sin(\omega_i t + \Delta D(t) + \theta_c) + n(t)$$

where

$r(t)$ is the received signal in volts (V)

P is the average signal power in V^2

Δ is the modulation index

$D(t) = d(t) \; \mathbf{S}\sin(\omega_{sc} t + \theta_{sc})$ with $\mathbf{S}\sin x = \mathrm{sgn}(\sin x)$

$$d(t) = \sum_{\ell=-\infty}^{+\infty} a_\ell \, p(t - \ell T), \; a_\ell = \pm 1 \text{ with equal probability}$$

ω_i is the received IF frequency in rad/s

θ_c is the carrier phase in rad

ω_{sc} is the subcarrier frequency in rad/s

θ_{sc} is the subcarrier phase in rad

$n(t) = \sqrt{2}\, n_c(t)\cos(\omega_i t + \theta_c) - \sqrt{2}\, n_s(t)\sin(\omega_i t + \theta_c)$ is a narrowband white Gaussian noise process with $n_c(t)$ and $n_s(t)$ being statistically independent, stationary band-limited white Gaussian noise processes with one-sided spectral density N_0 in V^2/Hz and one-sided bandwidth W.

T is the symbol time

The bandpass filtering effect on the subcarrier is neglected, i.e., the subcarrier waveform is assumed to be an ideal square wave.

The reference signal for carrier lock is a combination of the sampling clock and the carrier-loop digitally controlled oscillator (DCO) output signal. The sampling frequency is derived from the symbol synchronization loop and is denoted by $4f_0$. The symbol synchronization loop, not considered here, is assumed to maintain perfect symbol sync, and f_0 is a multiple of the symbol rate.

The input signal, $r(t)$, is mixed to frequency f_0 using the carrier-loop DCO output signal, whose instantaneous phase is $\omega_1 t + \hat{\theta}_c$, with $\omega_1 = \omega_i - \omega_0$ and $\hat{\theta}_c$ denoting the estimate of θ_c. Thus the instantaneous residual carrier phase of $s(t)$ is denoted by $2\pi f_0 t + \phi_c$, where $\phi_c = \theta_c - \hat{\theta}_c$.

The IF signal $s(t)$ is sampled at frequency $f_s = 4f_0$, at times $t_n = n/4f_0$. The mth output of the sampler is multiplied by $\exp(jm\pi/2)$ and demultiplexed into real and imaginary parts with the real or m-even samples becoming the Q channel, and the imaginary or m-odd samples becoming the I channel.

The Q samples q_n are input to the residual carrier tracking loop and to the subcarrier Costas loop. The I samples i_n are input to the carrier Costas I arm. Note that the carrier Costas Q arm is the same as the subcarrier I arm. Sideband aided carrier tracking is accomplished by properly combining the outputs of the residual carrier and the carrier Costas phase detectors in order to maximize the total loop SNR.

III. Residual Carrier Phase Detection and Tracking

Residual carrier phase detection is accomplished using the Q-channel samples. The bandpass filtered signal $r(t)$ is mixed against the carrier DCO output signal $\sqrt{2}\cos((\omega_i - \omega_0)t + \hat{\theta}_c)$ and lowpass filtered to give:

$$s(t) = \sqrt{P}\sin(\omega_0 t + \Delta D(t) + \phi_c) + n_c(t)\cos(\omega_0 t + \phi_c) - n_s(t)\sin(\omega_0 t + \phi_c)$$

where $\phi_c = \theta_c - \hat{\theta}_c$ is the carrier phase error.

From the above equation, it follows that the quadrature sampled term is:

$$\begin{aligned} q_n &= Re\{s(m/(4f_0)) \cdot \exp(jm\pi/2)\}, \quad m = 2n \\ &= \sqrt{P}\sin(\Delta D(nT_s) + \phi_c(nT_s)) + n_c(nT_s)\cos(\phi_c(nT_s)) + n_s(nT_s)\sin(\phi_c(nT_s)) \end{aligned} \tag{1}$$

and the in-phase sampled term is:

$$\begin{aligned} i_n &= Im\{s(m/(4f_0)) \cdot \exp(jm\pi/2)\} \quad m = 2n+1 \\ &= \sqrt{P}\cos(\Delta D(nT_s + T_s/2) + \phi_c(nT_s + T_s/2)) \\ &\quad - n_c(nT_s + T_s/2)\sin(\phi_c(nT_s + T_s/2)) \\ &\quad - n_s(nT_s + T_s/2)\cos(\phi_c(nT_s + T_s/2)) \end{aligned} \tag{2}$$

where $T_s = 1/(2f_0)$.

The continuous counterpart of the noise portion of Eqs. (1) and (2) is essentially white noise with zero mean and one-sided spectral density N_0.

The S curve for the residual carrier-loop phase detector, i.e., the average output of the phase detector as a function of the phase, is obtained by plotting $C(\phi_c) = E\,[q_n\,|\phi_c]$ versus ϕ_c, where $E[\cdot]$ denotes the statistical expectation. From Eq. (1), and since the noise is zero mean,

$$C(\phi_c) = E\,[q_n\,|\phi_c] = \sqrt{P}\cos\Delta\sin\phi_c$$

which is the expected sinusoidal S curve, and is plotted in Fig. 2.

From Holmes (Ref. 1, Chapter 5), the rms phase error of a phase-locked loop due to random noise is

$$\sigma^2_{\phi_c} = N_0\,B_L/A^2$$

where N_0 is the one-sided spectral density at zero frequency at the phase detector output, B_L is the one-sided loop bandwidth, and A is the slope of the S curve at $\phi = 0$. It is also necessary to assume that the noise bandwidth is wide compared to B_L.

For the residual carrier phase detector, $A = \sqrt{P}\cos\Delta$ and the noise spectral density is the input noise spectral density or N_0. Thus

$$\sigma^2_{\phi_c} = N_0\,B_L/(P\cos^2\Delta) \tag{3}$$

This is the same as for an analog loop with the same bandwidth.

IV. Subcarrier Phase Detection and Tracking

The subcarrier phase detection and synchronization is done by means of an optimum Costas loop. The analysis assumes perfect symbol synchronization and that the residual carrier loop is locked so that $\phi_c = 0$.

The carrier Q samples q_n are multiplied in the two arms of the Costas loop by phase quadrature square-wave reference signals at the subcarrier frequency. The products in the two channels are summed over the symbol times to implement the matched filters of the optimum Costas loop. The channel with the reference in phase with the subcarrier is the data channel.

Let $I_q(i)$ and $Q_q(i)$ denote the subcarrier I- and Q-channel outputs for symbol i, with subscript q denoting the carrier Q channel. Then the outputs of the I and Q arm filters are respectively

$$\begin{aligned} I_q(i) &= \frac{1}{N}\sum_{n=1}^{N} q_n\,\mathrm{Sin}\,(\omega_{sc}nT_s + \hat{\theta}_{sc}) \\ &= \sqrt{P_d}\,d(i)\cdot F_I(\phi_{sc}(i))\cdot\cos(\phi_c(i)) + N_i \\ &\quad + \text{periodic terms.} \end{aligned} \tag{4}$$

$$\begin{aligned} Q_q(i) &= \frac{1}{N}\sum_{n=1}^{N} q_n\,\mathrm{Cos}\,(\omega_{sc}nT_s + \hat{\theta}_{sc}) \\ &= \sqrt{P_d}\,d(i)\cdot F_q(\phi_{sc}(i))\cdot\cos(\phi_c(i)) + N_q \\ &\quad + \text{periodic terms} \end{aligned}$$

where the signal terms are:

$$\sqrt{P_d} = \sqrt{P}\sin\Delta = (\text{signal power})^{1/2}$$

$$d(i) = \frac{1}{N}\sum_{n=1}^{N} d(nT_s) \tag{5}$$

$$F_I(\phi_{sc}) = 1 - \frac{2}{\pi}\,|\phi_{sc}|, \qquad 0 \leqslant |\phi_{sc}| \leqslant \pi$$

$$F_q(\phi_{sc}) = \begin{cases} -\dfrac{2}{\pi}\,\phi_{sc}, & 0 \leqslant |\phi_{sc}| \leqslant \dfrac{\pi}{2} \\ -\dfrac{2}{\pi}\,(\pi - \phi_{sc}), & \dfrac{\pi}{2} \leqslant |\phi_{sc}| \leqslant \pi \end{cases}$$

and the noise terms are:

$$\begin{aligned} N_i &= \frac{1}{N}\sum_{n=1}^{N}\,[n_c(nT_s)\cos\phi_c - n_s(nT_s)\sin\phi_c] \\ &\quad \times \mathrm{Sin}\,(\omega_{sc}nT_s + \hat{\theta}_{sc}) \end{aligned}$$

$$\begin{aligned} N_q &= -\frac{1}{N}\sum_{n=1}^{N}\,[n_c(nT_s)\cos\phi_c - n_s(nT_s)\sin\phi_c] \\ &\quad \times \mathrm{Cos}\,(\omega_{sc}nT_s + \hat{\theta}_{sc}) \end{aligned}$$

with means:

$$E[N_i] = E[N_q] = E[N_i \cdot N_q] = 0$$

$$E[N_i^2] = E[N_q^2] = N_0/(2T)$$

Note that in the previous expressions we have dropped the variable i denoting the ith symbol.

The phase detector output signal is given by:

$$Z = -I \cdot Q$$

$$= P_d \cdot F_i(\phi_{sc}) \cdot F_q(\phi_{sc}) + N_i \cdot N_q + \sqrt{P_d} \cdot F_i(\phi_{sc}) \cdot d \cdot N_q + \sqrt{P_d} \cdot F_q(\phi_{sc}) \cdot d \cdot N_i + \text{periodic terms}$$

The periodic terms arise from end effects when there is not an integer number of subcarrier cycles in a symbol time. Thus these terms have frequency depending on the relationship between symbol rate and subcarrier frequency. When the subcarrier frequency is close to a multiple of the symbol rate, the frequency of the periodic terms is low, but the amplitude is very small. When the subcarrier frequency is not close to a multiple of the symbol rate, the amplitude is higher, but the frequency is on the order of half the symbol rate. In this case, the frequency is assumed to be well outside the loop bandwidth. In either case, the effect of the periodic terms on the loop phase error is negligible. There are also small effects due to the sampling but they are negligible. The detailed analysis regarding these periodic terms will be published in a future report.

For ϕ_{sc} small, $F_q(\phi_{sc}) \approx 0$ and $F_i(\phi_{SC}) \approx 1$. With this approximation and neglecting the periodic terms, the phase detector output becomes

$$Z = P_d \cdot F_i(\phi_{sc}) \cdot F_q(\phi_{sc}) + N_{sc}$$

where $N_{sc} = \sqrt{P_d}\, d \cdot N_q + N_i \cdot N_q$ represents the signal X noise plus the noise X noise terms.

Now if we fix ϕ_{sc} and take the statistical expectation of Z we get the S curve, $S(\phi_{sc})$, of the subcarrier loop-phase detector. This is

$$S(\phi_{sc}) = E[Z \mid \phi_{sc}] = (2/\pi) \cdot P_d \cdot F_i(\phi_{sc}) \cdot F_q(\phi_{sc})$$

Notice that $S(\phi_{sc})$ is periodic in ϕ_{sc} with period π; in the interval $[-\pi/2, \pi/2]$, $S(\phi_{sc})$ has the following expression:

$$S(\phi_{sc}) = (2/\pi) \cdot P_d \cdot \phi_{sc} \cdot [1 - (2/\pi)\,|\phi_{sc}|], \quad |\phi_{sc}| \leq \pi/2$$

Figure 3 shows a plot of the normalized S curve $Sn(\phi_{sc})$ for $|\phi_{sc}| \leq \pi/2$; the S curve is defined as:

$$Sn(\phi_{sc}) = \pi/(2P_d) \cdot S(\phi_{sc})$$

Also notice that we have stable lock points at $\phi_{sc} = \pm k\pi$, which is to be expected in a Costas-loop S curve.

To get the phase error variance $\sigma^2_{\phi_{sc}}$, we assume that the noise term N_{sc} is white compared to the closed-loop bandwidth B_L. The one-sided spectral density of the above noise process, N_0', is given by

$$N_0'/(2T) = E[N_{sc}^2] = P_d \cdot N_0/(2T) + (N_0/2T)^2 \qquad (6)$$

In deriving the above expression, $\phi_c = 0$ is used.

As for the residual carrier loop, the variance of the phase error can be expressed as (Ref. 1):

$$\sigma^2_{\phi_{sc}} = N_0' B_L / [S'(0)]^2$$

where B_L is now the one-sided bandwidth of the subcarrier loop and

$$S'(0) = \left. \frac{dS(\phi_{sc})}{d\phi_{sc}} \right|_{\phi_{sc} = 0}$$

Substituting N_0' and $S'(0)$ by their expressions we get:

$$\sigma^2_{\phi_{sc}} = \left(\frac{\pi}{2}\right)^2 \frac{N_0 B_L}{P_d} \left[1 + \frac{1/2}{E_s/N_0}\right]$$

Except for a factor of $(\pi/2)^2$, this is the same expression as for a Costas-loop tracking a sine-wave subcarrier. The factor $(\pi/2)^2$ is the same as that used for tracking unmodulated square waves with square-wave references. The factor $[1 + (1/2)/(E_s/N_0)]$ represents the squaring loss due to signal X noise and noise X noise terms, and is the same as for an optimum analog Costas loop, (Ref. 2). Thus, in this method of analysis, there is no theoretical degradation due to the digital implementation.

V. The Data-Aided Loop

A data-aided carrier tracking loop is a loop that combines residual carrier tracking and Costas-loop tracking to reduce carrier phase error. To accomplish this, a Costas data-aiding-

type phase detector is implemented for the carrier by adding a third arm, with carrier I and subcarrier I. The output of the carrier Costas phase detector is added to the output of the residual carrier phase detector, with appropriate weighting. Minimum phase error for a given loop bandwidth is achieved by maximizing A^2/N_0 for the composite phase detector, where A is the slope of the S curve at zero phase and N_0 is the spectral density at zero frequency.

In the following analysis, it is assumed that the subcarrier loop is locked so that $\phi_{sc} = 0$, that ϕ_c is a slowly varying process, and that there is perfect symbol synchronization.

The Q arm of this loop is the same as the I arm of the subcarrier loop, and the output of its filter is given by Eq. (4):

$$I_q = \sqrt{P_d} \cdot d \cdot \cos(\phi_c) + N_i$$

The above expression neglects the periodic terms that get filtered out by the loop filter and the NCO; the variable i denoting the ith symbol is also omitted, and it is assumed that $\phi_{sc} = 0$; hence, $F_i(\phi_{sc}) = 1$.

Similarly, the output of the I-arm filter is given by:

$$I_i = \frac{1}{N} \sum_{n=1}^{N} i_n \, \text{Sin}\left(\omega_{sc} nT_s + \omega_{sc} \frac{T_s}{2} + \hat{\theta}_{sc}\right)$$

$$= -\sqrt{P_d}\, d' \sin(\phi_c) + N_i'$$

where

$$d' = \frac{1}{N} \sum_{n=1}^{N} d(nT_s + T_s/2)$$

is also equal to d defined in Eq. (5) since we have a multiple of two samples per symbol and we have perfect symbol synchronization. The noise term is given by

$$N_i' = \frac{1}{N} \sum_{n=1}^{N} [n_c(nT_s + T_s/2) \sin\phi_c + n_s(nT_s + T_s/2) \cos\phi_c]$$

$$\times \quad \text{Sin}(\omega_{sc} nT_s + \omega_{sc} T_s/2 + \hat{\theta}_{sc})$$

with $E[N_i'] = 0$, $E[N_i'^2] = N_0/(2T)$, and $E[N_i' \cdot N_i] = 0$ since N_i' and N_i are two orthogonal noise processes.

The carrier Costas phase detector output is:

$$U = -I_i \cdot I_q$$

$$= (1/2) P_d \sin(2\phi_c) + N_{cc}$$

where $N_{cc} = \sqrt{P_d} \cdot d \cdot [N_i \cdot \sin(\phi_c) + N_i' \cdot \cos(\phi_c)] + N_i \cdot N_i'$ represents the signal X noise and the noise X noise terms.

The S curve, $S1(\phi_c)$, of the carrier Costas-loop phase detector is found to be the regular Costas S curve, i.e., a sine wave in $2\phi_c$. Its expression is given by:

$$S1(\phi_c) = E[U \mid \phi_c] = (1/2) P_d \sin(2\phi_c)$$

The spectral density of N_{cc} is the same as that of N_{sc}, which was given by Eq. (6):

$$N_0' = P_d \cdot N_0 + N_0^2/(2T)$$

For only Costas tracking of the carrier, the phase-error variance is given by:

$$\sigma^2_{1\phi_c} = N_0' B_L / [S'1(0)]^2 = \frac{N_0 B_L}{P_d} \left[1 + \frac{1/2}{E_s/N_0}\right] \quad (7)$$

which is the same as that for any optimum Costas loop.

Finally, the output F_a of the residual carrier phase detector is combined with the output F_b of the carrier Costas phase detector. Let F_c be the combined phase-error signal with weights a and b; we have

$$F_c = a \cdot F_a + b \cdot F_b$$

If we want to maximize the total loop SNR then the weights a and b should be chosen such that:

$$a/b = (\sqrt{SNR_a} / \sqrt{SNR_b}) \cdot (\sigma_b/\sigma_a)$$

where

$SNR = 1/\sigma^2_{\phi_c}$ is the residual carrier loop SNR: Eq. (3)

$SNR = 1/\sigma^2_{1\phi_c}$ is the carrier Costas-loop SNR: Eq. (7)

$\sigma_a^2 = N_0 B_L$ is the variance of the noise process in the residual carrier loop

$\sigma_b^2 = N_0' B_L$ is the variance of the noise process in the carrier Costas loop

After substitution we get

$$a/b = \frac{\cos \Delta}{\sin \Delta} \sqrt{P_d + \frac{N_0}{2T}}$$

and the sideband aided loop SNR is: $SNR_c = SNR_a + SNR_b$

VI. Summary and Conclusions

For a telemetry processing system utilizing IF sampling, the random noise performances of residual carrier loops, optimum subcarrier Costas loops, and data-aided carrier tracking loops have been shown to be the same as the corresponding loops implemented by traditional analog means, or by digital means after analog demodulation to I-Q baseband. Some periodic effects are introduced by end effects when there is not an integer number of subcarrier cycles in a symbol time, but these effects are negligible for narrow-loop bandwidths because the frequency of the periodic effect is then outside the loop bandwidth. It is thus concluded that IF sampling is useful for avoiding the problems inherent with analog implementations, such as dc offsets in mixers and amplifiers, the need for calibration and adjustments, and less reliability, versatility, and flexibility than digital systems.

References

1. Holmes, J. K., *Coherent Spread Spectrum Systems*, John Wiley & Sons, New York, N.Y., 1982.
2. Yuen, J. H., *Deep Space Telecommunications Systems Engineering*, Publication 82-76. Jet Propulsion Laboratory, Pasadena, California, July 1982.

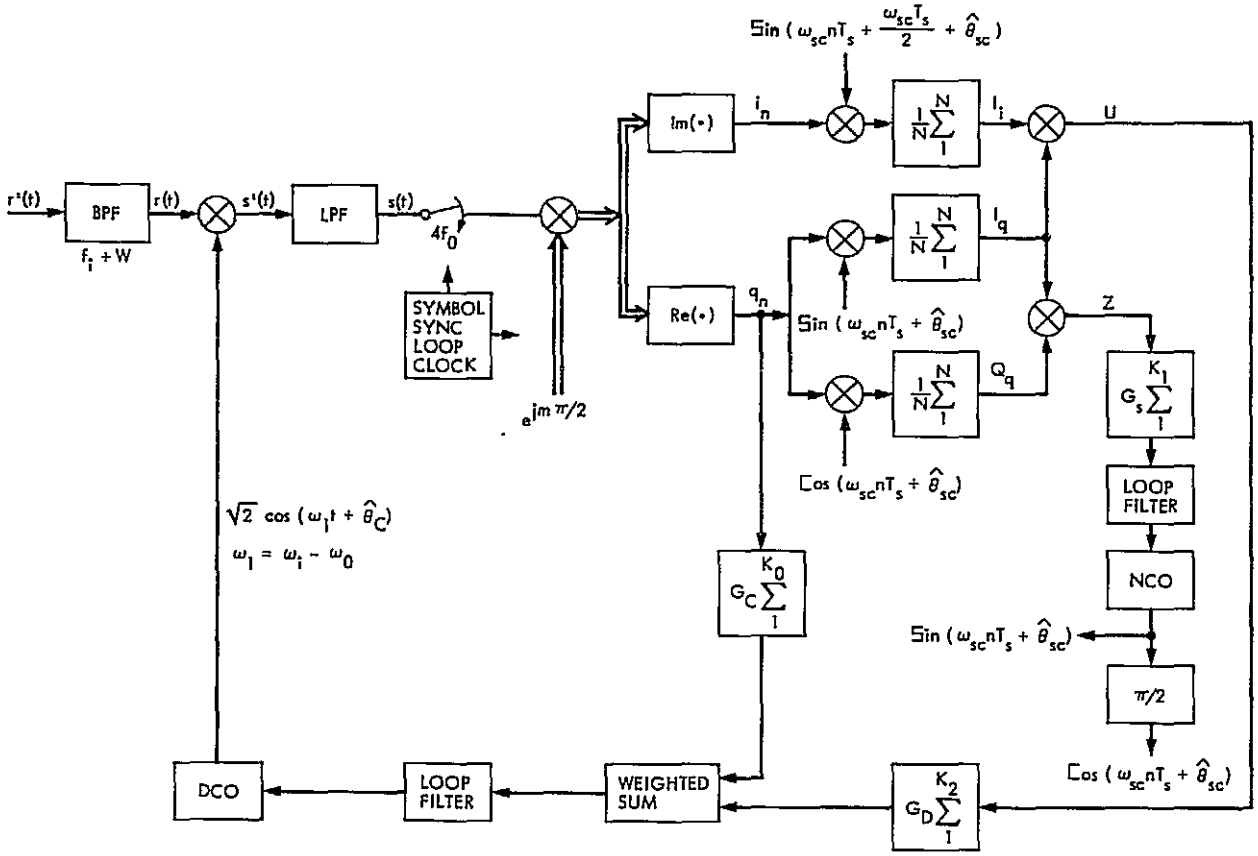


Fig. 1. Functional block diagram of proposed advanced receiver

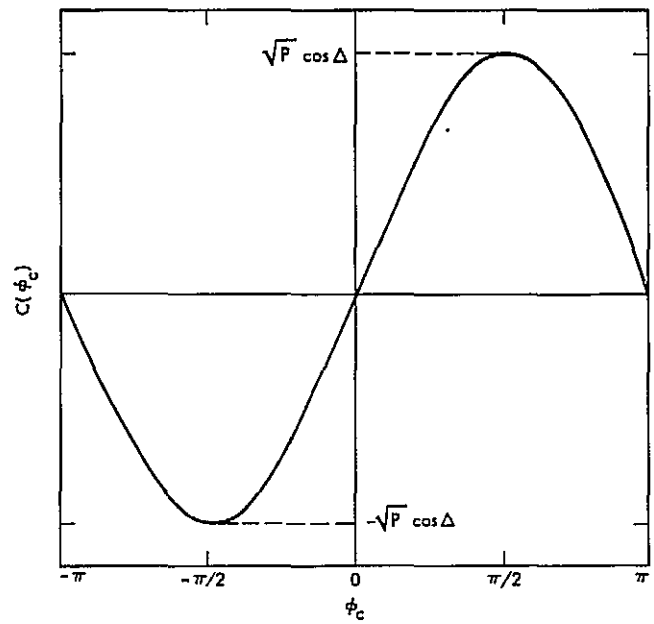


Fig. 2. S curve of residual carrier loop-phase detector; $C(\phi_c) = \sqrt{P}\cos\Delta \sin\phi_c$, $-\pi \leq \phi_c \leq \pi$

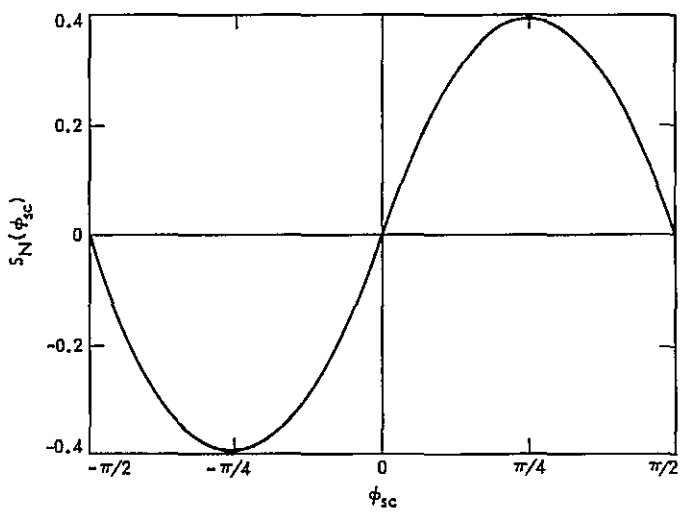


Fig. 3. S curve of subcarrier loop-phase detector; $S_N(\phi_{sc}) = \phi_{sc}(1 - 2/\pi|\phi_{sc}|)$, $-\pi/2 \leq \phi_{sc} \leq \pi/2$

A Strategy for Successful Deep Space Information Transmission in Bad Weather

L. Swanson and J. H. Yuen
Communications Systems Research Section

To retrieve data during bad weather, most missions slow the data rate to accommodate a certain amount of attenuation, or to allow for all but a very small percentage of all-weather conditions. This system has two well-known and balancing disadvantages: no data is received reliably during very bad weather, and the data rate is slowed during good weather. We propose a system of processing that encodes the most critical data more heavily, allowing it to be retrieved under bad conditions, while at the same time allowing most of the data to be sent at a higher data rate.

I. Introduction

A brief look at past progress in data transmission from space shows that a lot of improvement has come from frequency increases and associated technology. From 108 MHz, through L-band and S-band to our current X-band of the Voyager spacecraft, telemetry capability measured in bits per second at a given range has improved by a factor of 10^{18} since the first free-world satellite, Explorer I in 1958 (Ref. 1).

In the near future, use of Ka-band (32 GHz) will undoubtedly allow even further improvements in data rate. But as frequency increases, data rate increases must be weighed against performance degradations due to weather and atmospheric effects. Clouds are almost transparent at S-band; weather causes virtually no degradation. At the current X-band, data rate must be chosen for a weather trade-off. By lowering the data rate, a mission can be more nearly certain of clear reception, but the cost is less telemetry. For example, as shown in Table 1, total weather degradation for X-band data at the Madrid 64-meter antenna, 90% weather, 30° elevation angle, is 1.1 dB. This means that to ensure reception under 90% of all weather conditions, data must be slowed to a factor of 0.78 compared to what it would be if we assumed clear, dry weather. To ensure reception under 99% of all weather conditions (at the same station and elevation), a mission would have to accept a loss of 4.2 dB, or a data rate factor of 0.38, compared to clear, dry weather.

At Ka-band, while increased data rate is available in clear, dry weather, degradation due to bad weather is worse than at X-band. There is little experimental data currently available to quantify the losses, but in any case the same principle applies.

The current Voyager and Galileo communication systems employ the concatenated Reed-Solomon/convolutional code,

which is very sensitive to even a small degradation in signal-to-noise ratio. For example, bit error rate rises from 10^{-5} to 10^{-2} as E_b/N_0 drops from 2.3 dB to 1.9 dB (Ref. 2). This means that 0.4 dB can be the difference between successful data transmission and failure. Future deep-space channel coding schemes are likely to be just as sensitive.

II. Protecting Critical Data

One way to design a reliable telemetry system is to provide extra margin. For example, at the Madrid 64-meter antenna 30° elevation, one could adjust the data rate to allow E_b/N_0 = 3.4 dB, allowing a margin of 1.1 dB over the 2.3 dB required for a bit error rate of 10^{-5}. This would mean that, during 90% of all weather conditions, E_b/N_0 is sufficient to expect a decoded bit error rate of 10^{-5}. In fact, even in the absence of weather degradation, we must provide some margin for other system uncertainty (Ref. 1). The Voyager and Galileo missions provide about 2.0 dB for these nonweather effects. Thus a total of 5.4 dB of E_b/N_0 is needed to assure reliable communication 90% of the time.

This system has two obvious drawbacks. One is that no useful data is received 10% of the time. The other is that the 1.1 dB is unnecessarily conservative most of the time, and so some data rate is being sacrificed. These drawbacks play off against each other: the higher data rate during good weather means a higher likelihood that the reception of any useful data is precluded during bad weather.

We propose an information transmission system that will allow certain data, viewed by the mission as critical or "must receive," to be retrieved under the worst of circumstances (e.g., 99% weather) while allowing a reasonably high data rate for all data during most weather. Another goal, such as maximizing the expected total data return, might lead to a different coding scheme.

Each mission has different scientific and mission objectives, and so different critical data. Our scheme is to encode this critical data separately so that it is recoverable under very bad circumstances. We call the rest of the data "normal" data. (E. C. Posner (Ref. 3) refers to the two types as "base" and "bonus" data, respectively.)

Our coding scheme does not require a completely new deep-space telemetry system, but builds upon the existing concatenated Reed-Solomon/convolutional code, requiring only simple additional equipment. Hence it is an efficient and economically effective way to enhance data reception capability.

The current deep-space telemetry coding scheme, as used on JPL missions and adopted as the guideline of the International Consultative Committee on Space Data Standards, can be seen in Fig. 1.

We propose adding a repetition code to the critical data, yielding the system depicted in Fig. 2. On the ground, critical data is identified and the repetition code decoded. During good weather, all data is Viterbi and Reed-Solomon decoded. During bad weather, only critical data can be decoded. (Critical data must be sent in whole frames in order that the outer decoders can work during bad weather.)

The performance of our system is a parametric function of the repetition code rate and of the amount of critical data. Of course, the extra power given to the critical data means that there is less power available to the normal data; again the amount of power lost to the normal data is a function of the repetition code rate and of the amount of critical data. Table 2 shows the loss of power in overall data (for critical and normal data combined) in a system using our scheme, compared to the current concatenated convolutional/Reed-Solomon system, as a function of repetition code rate and the amount of critical data. When x of the data is repeated n times, the loss of power in overall data is just $(n-1)x$, which is shown on Table 2 in dB.

The fact that a concatenated Reed-Solomon/convolutional/repetition code is a good low-rate code for low symbol signal-to-noise ratios was first called to our attention by Pil Lee during a technical discussion. Indeed, this seemingly almost trivial repetition code works quite well. We compared it to other low-rate coding schemes (orthogonal and biorthogonal codes); this comparison is shown in Appendix A.

The scheme described above is different from the one proposed by E. C. Posner (Ref. 3), and first suggested by T. M. Cover (Ref. 4), which uses a single "cloud" code to protect some data more than others. The Cover-Posner scheme is a theoretical one, giving bounds on the data rates for the two kinds of data, while ours is a concrete, easy-to-implement system based on a very minor addition to the existing proven deep-space coding system. A comparison of the performance of a time-multiplexed system like ours to the optimum is given in Appendix B.

Aside from the obvious critical data protection, a scheme like ours offers several advantages. Of course, a mission can determine what data is critical. For example, some science data and some highly compressed imaging data might be the critical data on a mission. Also, a mission can determine how heavily to encode the critical data, and how much critical data to send, trading these off against power for the normal data. There could even be different levels of critical data: five repetitions for very critical data, three repetitions for less critical

data, etc. Also, the amount of redundancy can be changed in flight, in case of hardware changes on the ground or even on account of short-term weather predictions.

What does our protection cost in terms of data rate during good weather? The fact that the critical data has been expanded means that overall data rate must suffer in some way. Three ways to deal with this loss present themselves: (1) compressing the rest of the data, (2) sacrificing the least desirable data, and (3) lowering the probability that normal data will be received reliably. We examine each in the paragraphs below.

Current missions use data compression for imaging data.[1] Depending on the amount of redundancy in the original data, data compression can allow a large increase in the amount of information communicated at a given data rate. If the normal data being sent are redundant, further data compression would present little problem. If all possible data compression has already been done, source coding could be done to code the information bits. This, however, adds substantially to the bit error rate. So data compression should be used only if redundancy exists in the normal data.

The simplest idea is just to sacrifice (never transmit) the least desirable data, lowering the real data rate. This gives each transmitted normal bit exactly the same power it had before the critical bits were heavily encoded.

The last possibility is to speed channel symbols, increasing their rate. This means that each channel symbol carries slightly less power. Thus the normal data is received under slightly more restrictive weather conditions, in exchange for the critical data being received under less restrictive conditions.

[1]E. Hilbert et al., *BARC Data Compression for Galileo Imaging*. Publication GLL-625-301, Jet Propulsion Laboratory, Pasadena, California, 1979 (internal document).

III. Conclusions

We have proposed a coding system to protect a mission's critical data against very low signal-to-noise ratio conditions. This system is simple to implement, easy to change, and is based on a proven, reliable, existing coding system. It allows a small amount of data to be protected against very bad attenuation, while allowing all of the data to be sent at a higher data rate than would be the case if all data were protected against such bad attenuation. Critical data is heavily encoded and then embedded in the normal data. If only a small amount of data is critical, the effect on the power available for the rest of the data is minimal. Besides protecting critical data against bad weather, another goal might be to maximize expected total data return. This goal might lead to a different coding scheme.

There are other means of dealing with weather effects, which are operational in nature. Our method does not in any way preclude the use of these. It does offer additional protection to any link. Historically, this protection has been offered by the use of a lower frequency link, like S-band, which is virtually independent of weather, for critical data. In the future, such weather-transparent links may not be available, but in any case our scheme can be viewed as additional protection.

We have assumed that synchronization will not be a problem. This and the ability of the deep-space telemetry system to recover channel symbols under very bad weather conditions are questions that still need to be addressed.

References

1. Yuen, Joseph H. (Editor), *Deep Space Telecommunications Systems Engineering*, Plenum, New York, 1983.

2. Miller, R. L., L. J. Deutsch, and S. A. Butman, *On the Error Statistics of Viterbi Decoding and the Performance of Concatenated Codes*, JPL Publication 81-9, Jet Propulsion Laboratory, Pasadena, California 1981.

3. Posner, E. C., "Strategies for Weather Dependent Data Acquisition," *TDA Progress Report 42-65*, Jet Propulsion Laboratory, Pasadena, California, October 15, 1981, pp. 34-46.

4. Cover, T. M., "Broadcast Channels," *IEEE Transactions on Information Theory*, IT-18, January 1972, pp. 2-13.

5. Golomb, Solomon W., et al., *Digital Communications with Space Applications*, Chapter 7 and Appendix 4 (Andrew J. Viterbi), Prentice-Hall, Inc., Englewood Cliffs, 1964.

Table 1. Total weather degradation (including effects of both increased atmospheric attenuation and increased system noise temperature) in decibels; Madrid X-band 64-meter[a]

Percent weather	Elevation angle, degrees														
	10	12	14	16	18	20	25	30	35	40	50	60	70	80	90
10.0	0.2	0.2	0.2	0.2	0.2	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
20.0	0.3	0.3	0.2	0.2	0.2	0.2	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
30.0	0.4	0.4	0.3	0.3	0.3	0.3	0.2	0.2	0.2	0.2	0.2	0.1	0.1	0.1	0.1
40.0	0.5	0.5	0.4	0.4	0.4	0.3	0.3	0.3	0.2	0.2	0.2	0.2	0.2	0.2	0.2
50.0	0.6	0.5	0.5	0.4	0.4	0.4	0.3	0.3	0.3	0.3	0.2	0.2	0.2	0.2	0.2
60.0	0.9	0.8	0.7	0.6	0.6	0.5	0.4	0.4	0.3	0.3	0.3	0.2	0.2	0.2	0.2
70.0	1.3	1.1	1.0	0.9	0.8	0.7	0.6	0.5	0.4	0.4	0.3	0.3	0.3	0.3	0.3
75.0	1.6	1.4	1.2	1.1	1.0	0.9	0.7	0.6	0.5	0.5	0.4	0.3	0.3	0.3	0.3
80.0	2.1	1.7	1.5	1.3	1.2	1.1	0.9	0.7	0.6	0.5	0.5	0.4	0.4	0.3	0.3
85.0	2.6	2.2	1.9	1.7	1.5	1.3	1.0	0.9	0.7	0.7	0.5	0.5	0.4	0.4	0.4
90.0	3.4	2.9	2.4	2.1	1.9	1.7	1.3	1.1	0.9	0.8	0.7	0.6	0.5	0.5	0.5
95.0	5.0	4.2	3.6	3.2	2.8	2.5	2.0	1.6	1.4	1.2	1.0	0.8	0.8	0.7	0.7
98.0	7.7	6.6	5.8	5.2	4.6	4.2	3.4	2.9	2.5	2.2	1.8	1.6	1.4	1.3	1.3
99.0	10.3	9.0	7.9	7.1	6.5	5.9	4.9	4.2	3.7	3.3	2.7	2.4	2.2	2.1	2.0
99.5	14.6	12.6	11.2	10.1	9.3	8.6	7.3	6.3	5.6	5.1	4.3	3.9	3.5	3.4	3.3

[a]This table is one of a set of tables, describing weather attenuation under many circumstances at many Deep Space Network antennae, prepared by P. Kinman of the Jet Propulsion Laboratory in an internal memorandum to N. Burow.

Table 2. Loss of power (in decibels) to data because of repetition of critical data

Attenuation allowed critical data, dB / Critical data, % / Code repetition	2	3	5	10
	3	4.7	7	10
1	0.04	0.09	0.18	0.41
2	0.09	0.18	0.36	0.86
5	0.22	0.46	0.97	2.60
10	0.46	0.97	2.20	10.00

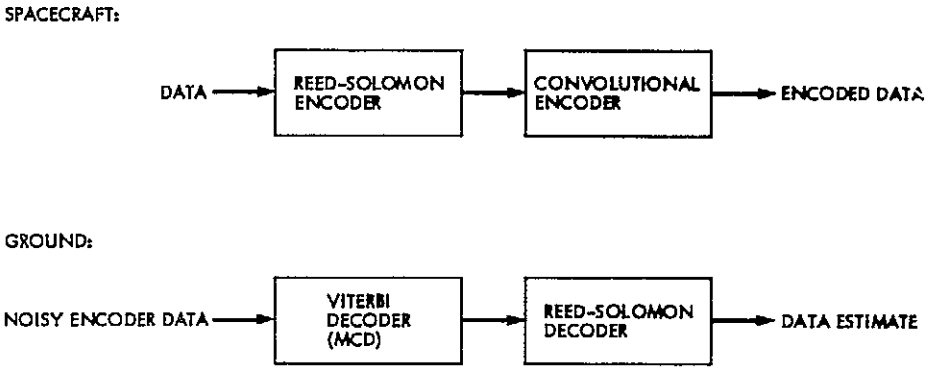


Fig. 1. The current deep-space telemetry coding system

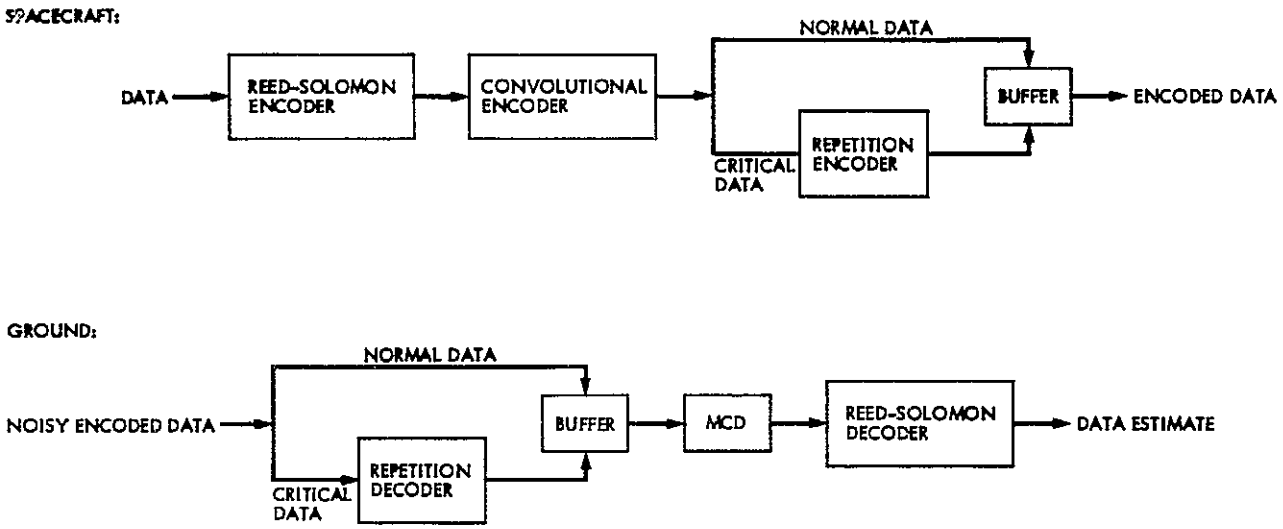


Fig. 2. Proposed new deep-space telemetry coding system

Appendix A

Choice of Code for Critical Data

The codes suggested in this article are simple to implement in a mission using concatenated Reed-Solomon/convolutional coding. Other low-rate codes could be considered for the critical data. Prominent examples of low-rate codes are orthogonal and bi-orthogonal codes (Ref. 5). Bi-orthogonal and orthogonal codes of appropriate rates are considered, along with concatenated Reed-Solomon/convolutional/repetition codes, in Table A-1.

For each of these three types of codes, we have chosen several specific codes to examine. Of course, rate balances against performance at low symbol signal-to-noise ratio.

Table A-1 gives these attributes for several codes of each type. In addition to code rate, expansion 7/(16 X rate) is given. This is the ratio of the number of channel symbols required for transmission of one critical bit under a given low-rate code to that required for a normal bit under Reed-Solomon/convolutional coding.

Table A-1 also gives performance, in terms of decoded bit error rate, of each code at two symbol signal-to-noise ratios: -4.4 dB and -6.5 dB. These were chosen because they represent 99% weather and 99.5% weather (X-band, Madrid 64-meter antenna, 30° elevation) when normal data is reliable in 90% weather. A choice among these codes for critical data would then depend on a balance between the expansion of the code and its performance. Of course, this balance gets more difficult if there is much critical data (which makes expansion touchier), or if a very low symbol signal-to-noise ratio must be considered. Performance for orthogonal and bi-orthogonal codes was obtained by interpolation from tables in Appendix 4 of Ref. 5; performance for the concatenated code was obtained from Ref. 2.

Comparison of codes for these purposes is different from normal comparison for error correcting codes. One usually compares codes by considering the bit signal-to-noise ratio (E_b/N_0) required for a given decoded bit-error rate (or conversely, the decoded bit-error rate for a given E_b/N_0). When one can adjust the data rate in order to control E_b/N_0, this is the logical way to compare codes. But we are assuming that the symbol rate has been adjusted to control E_b/N_0 for Reed-Solomon/convolutionally encoded normal data under certain weather conditions, so that symbol signal-to-noise ratio E_s/N_0 is determined entirely by weather conditions. Once we decided how much attenuation we wish to accommodate, we have determined the *symbol* signal-to-noise ratio. We can then choose a code, weighing the probability of error at that symbol signal-to-noise ratio vs code rate or expansion. The lower the code rate, of course, the more expansion, or the more channel space will be used by the critical data, lowering the power available for normal data. Notice that the error rates in Table A-1 show that the concatenated Reed-Solomon/convolutional/repetition codes perform better than the orthogonal and bi-orthogonal codes. (All error rates assume perfect carrier and subcarrier tracking, and perfect code synthronization.)

Table A-1. Performance of several low rate codes at low symbol signal-to-noise ratios

Orthogonal codes

k	Rate	Expansion	Bit-error rate at symbol SNR −4.4 dB	Bit-error rate at symbol SNR −6.5 dB
4	0.25	1.75	0.04	0.11
5	0.15625	2.8	0.004	0.03
6	0.09375	4.7	10^{-5}	0.002

Biorthogonal codes

k	Rate	Expansion	Bit-error rate at symbol SNR −4.4 dB	Bit-error rate at symbol SNR −6.5 dB
5	0.325	1.4	0.05	0.12
6	0.1875	2.3	0.005	0.04
7	0.109375	4.0	3×10^{-5}	0.003

Repetition codes – inside Reed-Solomon/convolutional

k	Rate	Expansion	Bit-error rate at symbol SNR −4.4 dB	Bit-error rate at symbol SNR −6.5 dB
2	0.21875	2.0	5×10^{-4}	>0.05
3	0.145833	3.0	$<10^{-5}$	>0.05
4	0.109375	4.0	$<10^{-5}$	$<10^{-5}$
5	0.0875	5.0	$<10^{-5}$	$<10^{-5}$

Appendix B

Time-Multiplexing vs Cloud Coding

We wish to compare a system like ours, which time-multiplexes critical data with normal data, to a cloud coding system (Ref. 4), which is optimal. To avoid comparing apples with oranges, we assume that we have optimal codes in all cases, that is, we are time-multiplexing an optimal code for the critical data with an optimal code for the normal data, and comparing this to an optimal cloud code.

For bandwidth B, signal-to-noise ratio P/N for normal data, signal-to-noise ratio P/AN for critical data, and critical data rate x_2, the normal data rate is bounded by

$$x_1 = B \log_2 \left[1 + A\left(2^{-x_2/B} - 1\right) + 2^{-x_2/B} P/NB\right]$$

in a cloud coded system Ref. 3.

Using time-multiplexing, the critical data rate x_2 uses

$$J = \frac{x_2}{B \log_2 (1 + P/ANB)}$$

of the channel time, leaving $(1 - J)$ of the channel time, at data rate $B\log_2 (1 + P/NB)$, for normal data. Thus the largest possible normal data rate $\hat{x}_1$, for time-multiplexing is

$$\hat{x}_1 = \left(1 - \frac{x_2}{B \log_2 (1 + P/ANB)}\right) B \log_2 (1 + P/NB)$$

We wish to bound the ratio $x_1/\hat{x}_1$.

The ratio $x_1/\hat{x}_1$ approaches 1 as x_2 approaches zero or $B \log_2 (1 + P/ANB)$; so, for fixed B, P/N, and A, it is largest when $d(x_1/\hat{x}_1)/dx_2$ is zero. This occurs when

$$\log_2 \left[1 + A\left(2^{-x_2/B} - 1\right) + 2^{-x_2/B} P/NB\right]$$

$$= \frac{\log_2 (1 + P/ANB) - x_2/B}{1 + A\left(2^{-x_2/B} - 1\right) + 2^{-x_2/B} P/NB} (A + P/NB) 2^{-x_2/B}$$

Substituting this into $x_1/\hat{x}_1$, we find that

$$\frac{x_1}{\hat{x}_1} = \frac{\log_2 (1 + P/ANB)}{\log_2 (1 + P/NB)} \cdot \frac{(A + P/NB) 2^{-x_2/B}}{1 + A\left(2^{-x_2/B} - 1\right) + 2^{-x_2/B} P/NB}$$

at the x_2 for which $x_1/\hat{x}_1$ is maximized. Since

$$x_2 \leq B \log_2 (1 + P/ANB)$$

we get

$$\frac{x_1}{\hat{x}_1} \leq \frac{A \log_2 (1 + P/ANB)}{\log_2 (1 + P/NB)}$$

Letting $D = B\log_2 (1 + P/NB)$ represent the largest supportable normal data rate, we get

$$\frac{x_1}{\hat{x}_1} \leq \frac{AB}{D} \log_2 \left(1 + \frac{2^{D/B} - 1}{A}\right)$$

For example, if B = 2,000,000, A = 5, and D = 115,000, we find that $(x_1/\hat{x}_1) \leq 1.016$, or the maximum loss due to using time-multiplexing (with optimal codes) compared to cloud coding (with optimal codes) is 0.07 dB.

Tropospheric Delay Effects in Radio Interferometry

G. Lanyi
Tracking Systems and Applications Section

A new tropospheric mapping function is derived which is more accurate than previous mapping functions above elevations of 4 degrees. The error due to the given analytic approximation is estimated to be less than 0.02% for elevation angles larger than 6 degrees, (less than 0.4 cm at 6 degrees, and approximately 0.004% or 0.03 cm at 20 degrees). The mathematical expansion used in the derivation is valid for any laterally homogeneous atmospheric model of refractivity. The new mapping function, computer generated ray tracing tables and other mapping functions are compared. The results can be used in correcting for tropospheric delays of radio signals.

I. Introduction

Radio waves traversing the atmosphere of the earth are delayed due to electromagnetic refraction. The delay depends on the length of the path; consequently, it is a function of the incident elevation angle of the radio wave. Usually, since satellite and stellar radio sources are observed only above 6° elevation, we set our precision requirements at 6°. The primary causes for this low elevation angle limit are: antenna pointing limits, ground obstructions, system noise and signal multipathing. In addition, modeling errors of the tropospheric delay are relatively large at very low elevation angles.

The tropospheric delay at an (unrefracted) elevation angle E can be determined by the mapping function $s(E)$. If the Earth's atmosphere is approximated by a laterally homogeneous plane air layer and the refractive bending of the ray is neglected, then $s(E) = Z/\sin E$, where Z is the zenith delay.

The mean sea level value for the dry tropospheric delay is ~2.3 m at zenith and ~20 m at 6° elevation. The delay of 20 m is a mean delay value for a laterally homogeneous dry atmosphere. Spherical geometry and refraction bending effects are included in the delay. For $E = 6°$, $1/\sin E \sim 9.6$; thus the difference between the mean and the approximated (plane air layer) delay is ~−2 m and it is smaller at higher elevation angles, approximately given by $-\epsilon Z/(\sin E \tan^2 E)$. This effect is primarily due to the curvature of the air layer. The mean value for ϵ is ~0.00122. This functional form for the delay correction was first derived by Saastamoinen (Ref. 1). The proportionality factor ϵ is a function of the atmospheric model of refractivity. Thus a 10% modeling error in ϵ results in an ~20 cm delay error at 6° elevation. Consequently, mismodelings of the tropospheric mapping can cause significant errors at low elevation angles.

In addition to the errors resulting from inherent zenith delay uncertainties, there are errors due to mismodelings, inhomogeneities and large-scale temporal variations. The corresponding major delay errors at low elevation angles are the following:

(1) Refractive bending.

(2) Dry modeling errors.

(3) Water vapor variations.

The refractive bending effect is ~13 cm at 6° elevation and can be completely removed by proper modeling. The variation of the bending effect due to zenith delay changes is ~20% or 2-3 cm at 6°. The dry and wet errors are about 10 and 11 cm, respectively, and these errors correspond to 1-2 σ estimates. The dry errors are due primarily to variations of the temperature profile of the troposphere. The water vapor errors are due primarily to large-scale temporal variations (e.g., diurnal variations) and inhomogeneities. In the estimate of the wet variations, a 6 hour observation period is assumed.

In the following, the effect of these errors in radio interferometric measurements will be discussed and we will describe a relatively accurate tropospheric delay function, which aids in removing some of these errors.

II. Tropospheric Delay Errors in Radio Interferometry

The delay errors described in Section I may occur in radio interferometric measurements. The present formal accuracy level of various interferometric measurements varies between 0.1 and 10 cm, depending on the baseline length and the number of observations. These formal accuracy figures do not account for all the systematic errors. Systematic errors can be incorrectly absorbed into parameter estimates. The magnitude of the error for estimates of intercontinental baselines is about the size of the delay errors estimated in Section I. Observations at low elevation angles are, however, necessary on long baselines for the following reasons:

(1) The visibility of radio sources by two stations is limited.

(2) Good geometry requires well separated multiple observations of radio sources.

(3) Such observations enable better determinations of zenith delay or other model parameters.

Direct experimental verifications of tropospheric delay errors are often difficult. A statistical evaluation of elevation dependent post-fit residuals may be used for indirect verification. However, elevation dependent systematics can be caused by other effects also. The major sources of error are ionospheric delays and antenna cable delays. Dual band ionospheric calibration results in small calibration errors. Antenna cable delays may have a different elevation dependence, and we do not have well established magnitude estimates for the Deep Space Network antennas at present. There are indications, however, that the effect does not exceed 15 cm in magnitude (L. E. Young, private communication, 1980) and antenna cable delay errors, in principle, are removed by proper calibration techniques. Consequently, a large portion of elevation dependent systematic errors may be caused by tropospheric delay errors. Elevation dependent systematics were reported by Treuhaft[1], Treuhaft et al. (Ref. 2), Treuhaft, Lanyi, and Sovers (Ref. 3), Shapiro et al. (Ref. 4), and Davis, Herring and Shapiro (Ref. 5).

Treuhaft[1] (see also Ref. 3) introduced a statistical measure for testing the elevation-angle dependence of the post-fit residuals for the 1978-1983 Deep Space Network Very Long Baseline Interferometry (VLBI) data. This test indicated a statistically significant elevation dependence for the data set. Treuhaft's other results indicate that the systematics persist for different tropospheric mapping functions with constant temperature model parameters. Additional results showed that elevation dependent systematics can be induced by simulations of temporal tropospheric variations.

III. Computational Approach

The tropospheric delay expression has to be a function of all major atmospheric model parameters if precise corrections for atmospheric changes are necessary. The first task is to establish an atmospheric model of the refractivity. Refractivity models of dry air can be relatively accurate. However, there is no accurate model for the highly variable water vapor.

For a given refractivity model one can evaluate the actual tropospheric delay by two different computational approaches:

(1) Computerized numerical calculation, i.e., ray tracing.

(2) Analytic formulation and approximation.

The first approach is relatively straightforward but time-consuming. This feature is particularly amplified when tropospheric parameters are varied. The second approach results in relatively fast computation, and, as a byproduct, it gives more insight into the problem than the previous approach. Analytic formulations were given by Hopfield (Refs. 6 and 7), Saastamoinen (Ref. 1), Chao (Ref. 8), Marini and Murray (Ref. 9), Black (Ref. 10), Black and Eisner (Ref. 11), and Lanyi[2] (see also Ref. 12).

For the dry air, the assumptions of static equilibrium and the ideal gas law are basically sufficient to describe the dry

[1]Treuhaft, R. N., "Time Variation of Intercontinental Baselines Using VLBI: Analysis and Validation," IOM 335.1-176, January 1984, private communication.

[2]Lanyi, G., "Tropospheric Propagation Delay Effects for Radio Waves," IOM 335.1-156, November 15, 1983, private communication.

refractivity profile if the temperature profile of the atmosphere is known. We use a temperature profile consisting of three linear sections. In former work one- or two-section profiles were used. For the water vapor, Saastamoinen's semi-empirical model is applied (Ref. 1).

In the analytic approximation we expand the tropospheric delay up to the third order in refractivity. The second and third order terms describe the refractive bending effect. These higher order terms were not explicitly included in former tropospheric mapping functions. Other expansions are carried out in parameters related to the curvature of the Earth's surface, e.g., the scale height divided by the local radius of curvature. Using these approximations we integrate over the curved path of the ray (see Fig. 1) and obtain the formula given in Appendix B. The derivation is not given here, but detailed calculations and other considerations will be presented in another publication.

The tropospheric delay formula given in Appendix B can be easily computer coded. We performed computerized ray tracing also for comparing various analytic mapping functions. Using the mean atmospheric parameters listed in Appendix A, we obtained comparison plots (see Figs. 2 through 4). These plots show that for the given atmospheric parameters the new mapping function is very close to the ray tracing result. For a different set of atmospheric parameters the other mapping functions might approach more closely the ray tracing in certain elevation regions, but the fundamental discrepancies cannot be removed.

IV. Summary

The tropospheric delay formula presented in Appendix B is a relatively accurate expression as far as the model and the derivation are concerned. The zenith delays and temperature profile parameters may be obtained from meteorological data. Alternatively, under some experimental circumstances, certain model parameters can be estimated from the data itself. In radio interferometric experiments on long baselines, total or individual (dry or wet) tropospheric zenith delays can be relatively well estimated if observations with low elevation angles are included in the experiment.

The major error, which cannot be estimated from the experiment, is due to water vapor inhomogeneities. This error is ~ 6 cm at 6° elevation. The temporal and azimuthal variations of the water vapor are also difficult to estimate from the experiment itself. Consequently, precise line-of-sight measurements by water vapor radiometers may be necessary to determine the delay due to water vapor. For this purpose, we assume that the total error for the water vapor radiometer is smaller than the errors we seek to eliminate. However, water vapor radiometers, in general, are not designed to function accurately at low elevation angles; thus the path toward high precision could be a difficult one.

As far as experimental verification of the new mapping function is concerned, two initial results should be mentioned:

(1) The new mapping function with fixed mean temperature profile parameters is statistically preferred by our data (Ref. 3). This is the same data set as was mentioned in Section II for antenna stations in Australia, California and Spain for the full 5 year period. The station in Australia has some bias, though, and sites at other geographical locations might have also somewhat different mean temperature profile parameters and correspondingly different mapping functions.

(2) Elevation dependent systematics can be partially removed by using variable atmospheric model parameters (Ref. 3).

Both findings are initial results.

In conclusion, it would seem that the best approach for determining the tropospheric delay function parameters lies in the combined use of surface and radiosonde meteorological data, water vapor radiometer data and statistical model parameter estimates.

References

1. Saastamoinen, J., Atmospheric Correction for the Troposphere and Stratosphere in Radio Ranging of Satellites, The Use of Artificial Satellites for Geodesy, Geophysical Monograph 15, American Geophysical Union, Washington, D. C., 1972.

2. Treuhaft, R. N., Liewer, K. M., Niell, A. E., Sovers, O. J., Thomas, J. B., and Wallace, K. S., The Time Variation of Intercontinental Baselines Using VLBI: Analysis and Validation, EOS Transactions, American Geophysical Union, Washington, D. C., 64, 678, 1983.

3. Treuhaft, R. N., Lanyi, G. E., and Sovers, O. J., Empirical Troposphere Modeling from DSN Intercontinental VLBI Data, EOS Transactions, American Geophysical Union, 65, 191, 1984.

4. Shapiro, I. I., et al., Geodesy by Radio Interferometry: Interpretation of Intercontinental Distance Measurements, EOS Transactions, American Geophysical Union, 64, 677, 1983.

5. Davis, J. L., Herring, T. A., and Shapiro, I. I., Geodesy by Radio Interferometry: Effects of Errors in Modeling the Troposphere, EOS Transactions, American Geophysical Union, 65, 191, 1984.

6. Hopfield, H. S., Two-quartic Troposphere Refractivity Profile for Correcting Satellite Data, J. Geophys. Res., 74, 4487-4499, 1969.

7. Hopfield, H. S., Tropospheric Effect on Electromagnetically Measured Range: Prediction from Surface Weather Data, Radio Sci., 6, 357-367, 1971.

8. Chao, C. C., The Tropospheric Calibration Model for Mariner Mars 1971, Tracking System Analytic Calibration Activities for the Mariner Mars 1971 Mission, Technical Report 32-1587, 61-76, Jet Propulsion Laboratory, Pasadena, California, March 1974.

9. Marini, J. W., and Murray, C. W., Correction of Laser Range Tracking Data for Atmospheric Refraction at Elevations above 10 degrees, NASA/GSFC X-591-73-351, 1973.

10. Black, H. D., An Easily Implemented Algorithm for the Tropospheric Range Correction, J. Geophys. Res., 83(B4), 1825-1828, 1978.

11. Black, H. D., and Eisner, A., Correcting Satellite Doppler Data for Tropospheric Effects, J. Geophys. Res., 89(D2), 2616-2626, 1984.

12. Lanyi, G. E., Tropospheric Propagation Delay Effects in Radio Interferometric Measurements, EOS Transactions, American Geophysical Union, 64, 210, 1983.

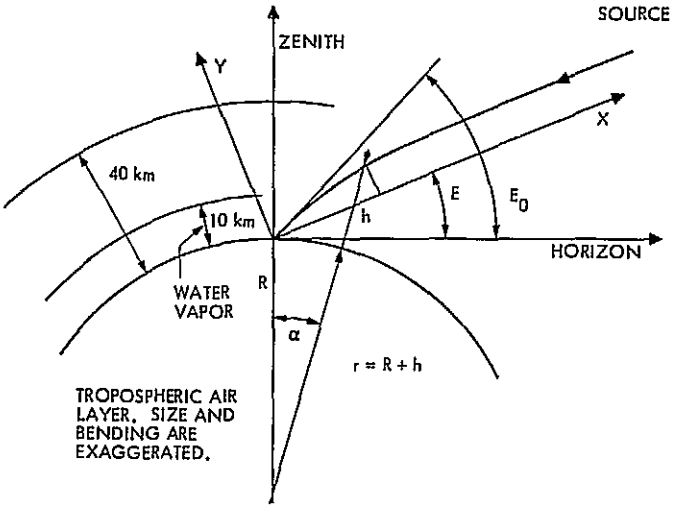


Fig. 1. Tropospheric refraction

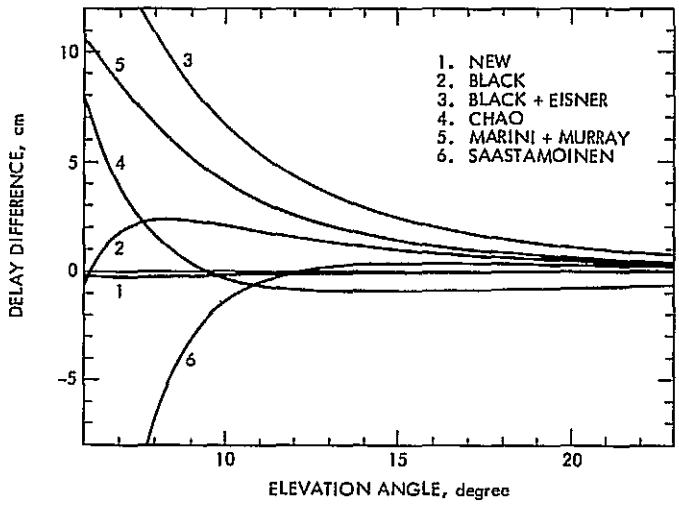


Fig. 3. Difference between the mapping functions and ray tracing for elevation angles of 6°–20°

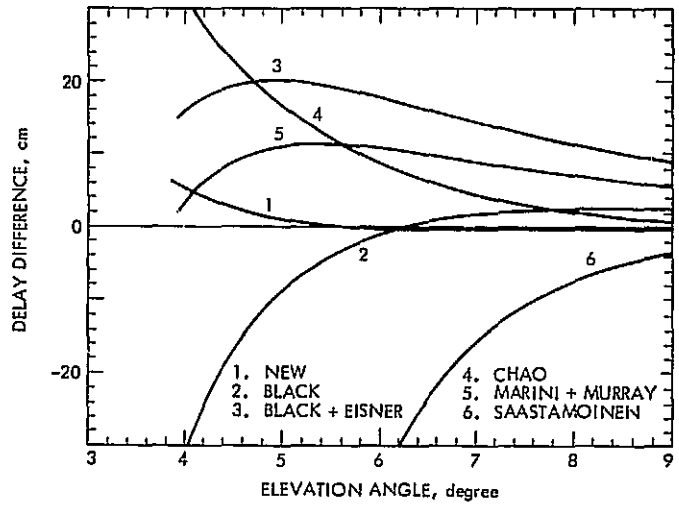


Fig. 2. Difference between the mapping functions and ray tracing for elevation angles of 4°–9°

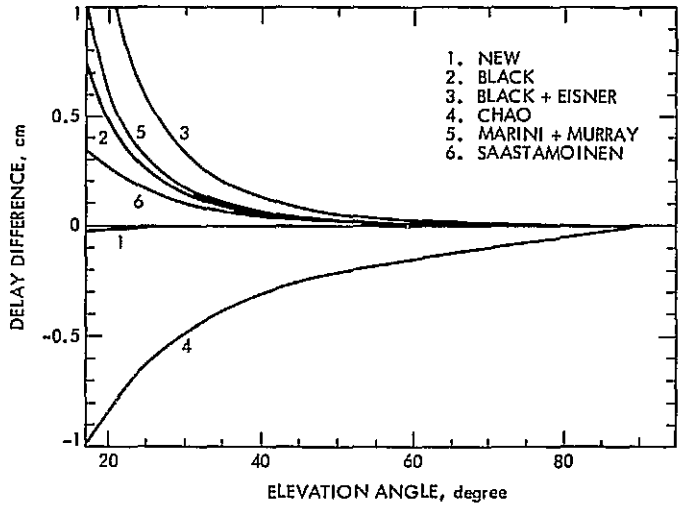


Fig. 4. Difference between the mapping functions and ray tracing for elevation angles of 20°–90°

Appendix A
List of Symbols

This is a list of symbols used in the tropospheric delay formula (see Appendix B). The values given here were used for calculating the plots in Figs. 2 through 4. The values of model parameters are chosen such that different models would match each other as closely as possible. Since the models can be matched only in an approximate manner, the values given here are the results of a compromise.

E	(unrefracted)	elevation angle
$s(E)$		tropospheric delay mapping function
p_0	= 1013.25 mbar (sea level)	surface pressure
T_0	= 292K	surface temperature
R	= 6371 km (for matching models)	radius of curvature of Earth
θ	= 45°	latitude
m	= $4.8097\ 10^{-23}$g (dry air)	mean molecular mass
k	= $1.38066\ 10^{-16}$erg/K	Boltzmann's constant
k/m	= $2.8706\ 10^{6}$ erg/gK	a gas constant of dry air
g_c	= 978.37 erg/g · cm	gravity at the center of gravity of air column
w	= 6.8165 K/km (a mean value)	temperature lapse rate
α	= 5 ($\alpha = mg_c/kw$, Hopfield's and Chao's value)	dry model parameter
β	= 3.5 (a mean value)	wet model parameter
h_1	= 0 (for matching models)	inversion altitude
h_2	= 12.2 km (Chao's value)	tropopause altitude
Z_{wet}	= 0 (for matching models)	wet zenith delay
λ	= 1 for $E > 10°$, = 3 for $E < 10°$	a scale factor

From the values above we obtain:

Δ	= 8.567 km ($\Delta = kT_0/mg_c$)	scale height
σ	= $1.345\ 10^{-3}$ ($\sigma = \Delta/R$)	a curvature measure
q_1	= 0 (scale height-normalized)	inversion altitude
q_2	= 1.424 (scale height-normalized)	tropopause altitude
Z_{dry}	= 230.70 cm	dry zenith delay

Appendix B

Tropospheric Delay

The tropospheric delay is the difference between the actual and a hypothetical vacuum (straight line) propagation time of the radio signal. In principle, by applying Fermat's principle for a laterally homogeneous spherical air layer, the path of the radio signal can be derived. The path is a function of the constant of integration. After expressing the constant of integration as a function of the unrefracted elevation angle E, an integration of the refractivity over the path is carried out. This gives the total propagation delay. After subtracting the vacuum delay, the tropospheric delay $s(E)$ is obtained.

These steps cannot be carried out exactly in an analytic fashion. The tropospheric delay formula presented in this appendix is a result of numerous approximations. The resultant tropospheric delay is a function of dry and wet zenith delays, Z_{dry} and Z_{wet}. The dry and wet zenith delays can be predicted from meteorological data (see, e.g., Ref. 1) or from the experiment itself. The tropospheric delay is also a function of other temperature profile related parameters: Δ, σ, α, q_1, q_2. The parameter β is a semi-empirical model parameter for the refractivity of the water vapor. The definition of the symbols and a set of mean values are given in Appendix A. These parameters can be adjusted according to local meteorological conditions.

In the following equations we use the indices d and w for referring to certain integrals of dry and wet surface-normalized refractivities, respectively. The index dw refers to the integrals of the product of dry and wet surface-normalized refractivities.

The tropospheric delay can be written as:

$$s(E) = F(E)/\sin E \tag{1}$$

Expansions in the refractivity and other quantities result in:

$$\begin{aligned} F(E) = {} & Z_{dry}F_{dry}(E) + Z_{wet}F_{wet}(E) + (Z_{dry}^2/\Delta)\,F_{bend1}(E) \\ & + 2\,(Z_{dry}Z_{wet}/\Delta)\,F_{bend2}(E) + (Z_{wet}^2/\Delta)\,F_{bend3}(E) \\ & + (Z_{dry}^3/\Delta^2)\,F_{bend4}(E) \end{aligned} \tag{2}$$

where

$$F_{dry}(E) = G(\lambda\langle q\rangle_d, u)A_d(E) + (3/4)G^3(\langle q\rangle_d, u)\langle q^2\rangle_d u\sigma \tag{3}$$

$$F_{wet}(E) = G(\lambda\langle q\rangle_w/\langle q^0\rangle_w, u)A_w(E)/\langle q^0\rangle_w \tag{4}$$

and

$$\left.\begin{aligned} F_{bend1}(E) = {} & -\frac{1}{2\tan^2 E}\left[G^3\left(\frac{\{q\}_d}{\{q^0\}_d}, u\right)\{q^0\}_d - G^3(\langle q\rangle_d, u)\,\frac{\sigma}{\sin^2 E}\right] \\ F_{bend2}(E) = {} & -\frac{1}{2\tan^2 E}\,G^3\left(\frac{\{q\}_{dw}}{\{q^0\}_{dw}}, u\right)\frac{\{q^0\}_{dw}}{\langle q^0\rangle_w} \\ F_{bend3}(E) = {} & -\frac{1}{2\tan^2 E}\,G^3\left(\frac{\{q\}_w}{\{q^0\}_w}, u\right)\frac{\{q^0\}_w}{\langle q^0\rangle_w^2} \\ F_{bend4}(E) = {} & -\frac{1}{2\tan^4 E}\,G^3\left(\frac{\{\{q\}\}}{\{\{q^0\}\}}, u\right)\{\{q^0\}\}_d \end{aligned}\right\} \tag{5}$$

The quantity $G(q,u)$ is a geometric factor related to the curvature of the Earth's surface and given by

$$G(q,u) = (1+qu)^{-1/2} \tag{6}$$

where

$$u = 2\sigma/\tan^2 E \tag{7}$$

The quantity $A(E)$ is given by

$$A(E) = \langle q^0\rangle + \sum_{n=1}^{10} ((2n-1)!!/n!)(-1/2)^n\,[u/1+\lambda\langle q\rangle u]^n\,\langle (q-\lambda\langle q\rangle)^n\rangle \tag{8}$$

The indices d and w in $A_d(E)$ and $A_w(E)$ refer to dry and wet moments in Eq. (8). The moments $\langle (q-\lambda\langle q\rangle)^n\rangle$ can be evaluated in terms of the moments $\langle q^n\rangle$ by the use of the binomial theorem. The quantities $\langle q^n\rangle$, $\{q^n\}$ and $\{\{q^n\}\}$ are the n-th order moments of the surface-normalized dry/wet refractivity $f(q)$, $f^2(q)$ and $f^3(q)$,

$$\left.\begin{aligned}
\langle q^n \rangle &= \int_0^\infty dq\, q^n f(q) \\
\{q^n\} &= \int_0^\infty dq\, q^n f^2(q) \\
\{q^n\}_{dw} &= \int_0^\infty dq\, q^n f_{\text{dry}}(q) f_{\text{wet}}(q) \\
\{\{q^n\}\} &= \int_0^\infty dq\, q^n f^3(q)
\end{aligned}\right\} \quad (9)$$

Denoting all the six types of dry, wet, and bend moments by $[q^n]$, for the particular three-section temperature profile model, we have

$$\begin{aligned}
[q^n] = n!\, [(1/a)^{n+1} (1 - \exp(-aq_1)) \\
+ \exp(-aq_1) \left(\prod_{i=0}^{n} (\alpha/(b + i + 1)) \right) (1 - \hat{T}_2^{b+n+1}(q_2)) \\
+ (1/a)^{n+1} \exp(-aq_1)\, \hat{T}_2^{b+n+1}(q_2)]
\end{aligned} \quad (10)$$

where a and b are listed in the following table:

a	b		$[q]$
1	$\alpha - 1$	dry	$\langle q \rangle_d$
β	$\beta\alpha - 2$	wet	$\langle q \rangle_w$
2	$2\alpha - 2$	dry squared	$\{q\}_d$
$\beta + 1$	$\beta(\alpha + 1) - 3$	product of dry and wet	$\{q\}_{dw}$
2β	$2\beta\alpha - 4$	wet squared	$\{q\}_w$
3	$3\alpha - 3$	dry cubed	$\{\{q\}\}_d$

and

$$\hat{T}_2(q_2) = 1 - (q_2 - q_1)/\alpha \quad (11)$$

If we set $q_1 = 0$ and neglect the bending terms and the second term in the dry expression, set $A(E) = 1$ and $\lambda = 1$ in the dry and wet formulas, and expand $G(\langle q \rangle, u)$, retaining the first order term for the dry and zeroth order term for the wet mapping function, then Saastamoinen's mapping function is obtained.

If we set $q_1 = 0$, $q_2 = 5$, and $\alpha = 5$, neglect the bending terms, set $A(E) = 1$ and $\lambda = 1$ in both the dry and wet formulas, and ignore the wet effects, then we obtain Black's single-term mapping function to a good degree of approximation.

Investigation and Rehabilitation to Extend Service Life of DSS-13 Antenna Concrete Foundation

A. A. Riewe, Jr.
Ground Antennas and Facilities Engineering Section

It was noticed in 1977 that exposed surfaces of the reinforced concrete foundation of the DSS-13 26-meter antenna were exhibiting relatively light cracking. By 1980 the cracking had worsened to the point were it was decided that an investigation should be undertaken to establish the cause and, as needed, devise a repair technique that would maintain the serviceability of the antenna. Core samples were obtained from the concrete and various laboratory tests conducted. In-place nondestructive type tests were also performed. The tests established that the concrete was deteriorating because of alkali-aggregate reactivity. This is a phenomenon wherein certain siliceous constituents present in some aggregates react with alkalies in the portland cement to produce a silica gel which, in turn, imbibes water, swells, and cracks the concrete. After a thorough structural analysis, a rehabilitation scheme was devised and installed. The scheme consisted of a supplemental steel frame and friction pile anchored grade beam encircling the existing foundation. This system provides adequate bracing against base shear and overturning due to seismic loading. Larger cracks were sealed using a pressure injected two-component epoxy.

I. Introduction

An extensive investigation was undertaken by the Ground Antennas and Facilities Engineering Section to determine the cause of severe cracking observed in the DSS-13 (Venus-Station) antenna reinforced concrete foundation and to design structural steel bracing to extend the useful life of the antenna. The antenna is one of several based at the Deep Space Communications Complex, Goldstone, California, which is owned by NASA and operated by the California Institute of Technology's Jet Propulsion Laboratory.

The investigation involved examination of the original concrete design, in-place testing and testing of cores taken from the interior and top surfaces of the foundation. The tests included ultrasonic pulse readings, petrographic examination, standard compression tests, standard tensile splitting tests, surface condition survey, determination of rebound number, pull-out tests of anchor bolts, and monitoring of crack widths under dynamic and wind loads. The rehabilitation found necessary to extend the useful life of the antenna foundation was established and has been successfully completed.

II. Structure Description

The DSS 13 antenna is a 26-meter-diameter parabolic reflector with an azimuth-elevation mount. The base frame consists of eight steel columns that taper outward as they descend to the concrete foundation where they are at a radius of 3.98 m (13 ft, ½ in.) at the top of the bearing plates. Each column is anchored into the foundation with four 57.15-mm-(2¼-in.-) diameter anchor bolts that extend 2.13 m (7 ft, 0 in.) into the concrete.

The foundation is 10.67 m (35 ft, 0 in.) square in plan. Figure 1 shows a partial plan and section. The top of concrete is slightly above adjacent grade and extends to a depth of 3.05 m (10 ft, 0 in.) as measured around its perimeter. In the center is a 3.66-m- (12-ft-0-in.-) diameter pit, 3.96 m (13 ft, 0 in.) deep. The footing under this pit is 0.76 m (2 ft, 6 in.) thick. The main reinforcing steel in the 3.05-m- (10-ft-0-in.-) thick section consists of No. 8 (25.4-mm- (1-in.-) diameter) bars at 152.4 mm (6 in.) on center each way, top and bottom. The reinforcing steel in the 0.76-m- (2-ft-6-in.-) thick section under the pit is also No. 8 bars at 152.4 mm (6 in.) on center each way on the bottom including the sloping transition, but the top bars in this area are No. 4 bars (12.7-mm- (½-in.-) diameter) at 0.305 m (12 in.) on center.

The only vertical reinforcing steel in the main 3.05-m- (10-ft-) thick section of the footing consists of No. 4 bars at 0.305 m (12 in.) on center each way in the outer face and in the face of the 3.66-m- (12-ft-) diameter pit.

III. Foundation Concrete Materials

The concrete foundation was placed during May and June of 1962. The mix design (Table 1) called for a 25.0-mm (1-in.) maximum aggregate, 334.6 kg/m^3 (6.0 sacks per yd^3) of cement and a maximum allowable water content of 5.33 X 10^{-4} m^3/kg (6.0 gal per sack) of cement. The aggregate source (Table 1) was Barstow, California; no records are available indicating the cement brand or type used. No admixtures were used.

The mix design was for a 24.13-MPa (3500-psi) compressive strength although the structural design was based on 20.68 MPa (3000 psi). Only five field test cylinders were taken during construction. This was inadequate. The volume of foundation concrete is on the order of 436 m^3 (570 yd^3) and the American Concrete Institute Building Code requires that a minimum of 16 field cylinders should be tested for this volume.

Three of the 152-mm- (6-in.-) diameter by 304.8-mm (12-in.) high field test cylinders had an average compressive strength of 17.31 MPa (2490 psi) at age 7 days. Two cylinders at age 28 days had an average compressive strength of 24.75 MPa (3590 psi).

IV. Structural Loading

The horizontal load, due to wind, on the main reflector is 311,375 N (70 kips) and on the steel base, 44,482 N (10 kips). The dead load of the main reflector is 1,334,467 N (300 kips) and of the steel base, 756,198 N (170 kips). A design check by a licensed structural engineer indicated there were no deficiencies in the original as-built structure.

V. Condition of the Structure

Cracks of various types have developed on almost all exposed surfaces of the foundation. "Map" or "pattern" cracking has developed over all exposed surfaces. The "map" cracks are most prominent on the upper horizontal surface of the foundation, with little or no difference between the exterior area and the area enclosed within the base. The "map" cracking, although moderate to very slight, appears on all surfaces within the pit area. The major system of cracks is a series of generally horizontal cracks, located on the vertical surfaces in the pit area. These cracks have a vertical spacing ranging from about 0.305 m (12 in.) to 0.457 m (18 in.). A majority of these cracks have a width on the order of 0.25 to 1.27 mm (0.010 to 0.050 in.), but two cracks have widened to approximately 12.7 mm (0.5 in.). Figure 2 is a photo taken of one of these cracks. An inspection pit was dug along the westerly side of the foundation exterior extending to its full depth of 3.05 m (10 ft). The crack pattern there was generally similar to that exposed in the interior pit described above.

VI. Investigation Program

Investigation of the foundation began in January 1980, when it was becoming apparent that areas of structural distress were developing. Cracking was first reported in 1977 by station personnel, but no untoward problem was thought to exist at that time because the cracking was relatively light and surficial in nature, and cracks often form in normal concrete due to thermal changes or drying shrinkage. The cracks, however, progressed to a point where, in 1980, an investigative program was initiated. The investigation included visual inspection, measurement of crack widths at selected locations, pull-out tests on two anchor bolts, monitoring of crack width during periods of high winds, and testing of concrete cores.

The most important information in concrete investigative work is obtained from the testing of cores. Tests on core samples provide a direct determination of absolute strength and elastic properties. Specimens obtained from cores permit petrographic study that can possibly establish the cause of cracking. Horizontal 152-mm- (6-in.-) diameter cores were taken in August, 1980, from the vertical wall in the pit area, which extended 0.61 to 0.91 m (2 to 3 ft) into the footing. Vertical 50.8-mm- (2-in.-) diameter cores were also obtained from the outdoor top of the foundation at the locations shown in Fig. 1. They extended to a depth on the order of

2.90 m (9 ft, 6 in.). The core locations were generally established by visual inspection to provide information on areas showing greatest distress as well as those of less distress.

VII. Description of Concrete Test Cores

The cores contained fine cracks of random orientation and slightly larger cracks of a generally horizontal orientation. Photographs of the two vertical cores in core storage boxes are shown in Fig. 3 and 4. Reinforcing steel was cut in three cores. No corrosion was observed on the steel and the bond between the steel and the concrete was tight. No rust stains have been observed on concrete surfaces near the cracks. This indicates that no corrosion of the steel is occurring.

VIII. Description of Concrete Tests

To determine the general condition of the concrete, the following tests were made:

(1) Ultrasonic pulse velocity tests to determine depth of major surface cracks.

(2) Standard compression tests.

(3) Petrographic (microscopic analysis) examination of cores to establish aggregate mineral composition.

(4) Standard tensile splitting test.

(5) Determination of rebound number.

Pulse velocity tests were performed in accordance with Ref. 1. The test, in this project, was used to measure the depth of the crack having the largest surface width. The test was done in accordance with procedures set forth in Ref. 2.

The basic principle of crack detection by pulse velocity tests is: If a crack is of appreciable width and is of considerable depth perpendicular to the test path, the path of the pulse will be blocked and no signal will be received at the receiving transducer. If the depth of the crack is small compared to the distance between the transducers, that is, the path length, the pulse will pass around the end of the crack and a signal will be received at the transducer. However, in doing so it will have traveled a distance longer than the straight line path upon which the pulse velocity computations are based. The resulting calculated pulse velocity will then be low in comparison with that through uncracked concrete in the same vicinity. The difference in the pulse velocity is then used to estimate the path length and hence the crack depth. Thus if the transducers are equidistant (x) from each edge of the crack, as shown in Fig. 5, the depth h of the crack can be obtained as follows:

$$\text{longitudinal pulse velocity} = \alpha$$

$$\text{distance traveled in concrete without crack} = 2x$$

$$\text{distance traveled in concrete with crack} = 2\sqrt{x^2 + h^2}$$

$$T_c^2 \text{ (in concrete with crack)} = \frac{4h^2 + 4x^2}{\alpha^2}$$

$$T_s^2 \text{ (in concrete without crack)} = \frac{4x^2}{\alpha^2}$$

$$h = x\sqrt{\frac{T_c^2}{T_s^2} - 1}$$

where

T_c = travel time around the crack

T_s = travel time along the surface of the same type of concrete without cracks

Compression tests were performed in accordance with Ref. 3. All cores were tested dry although ASTM C 42 provides that cores be tested promptly after being stored in lime-saturated water for 40 hours. Reference 4 states that cores shall be tested dry if service conditions in the structure are dry.

A petrographic study of random segments of the test cores was made microscopially to establish the types of minerals present in the aggregates and to determine if alkali-aggregate reactivity was occurring.

Splitting tensile strength was determined by testing core samples in accordance with Ref. 5.

Rebound number of the in-place concrete was measured on the interior face of the foundation in the pit area in accordance with Ref. 6. Three representative areas were tested.

IX. Test Results

A. Sonic Tests

The sonic test was used to measure the depth of the largest surface crack exposed in the interior pit face. The depth was determined to be on the order of 0.69 m (27 in.). Values obtained using this method were supplemented with other observations made on actual cores on cracks because it has not yet been established how wide a crack must be to significantly increase the transmission time.

B. Standard Compression Tests

Six 101.6-mm- (4-in.-) diameter cores were obtained in the pit area walls and delivered to an independent materials testing laboratory for testing. Due to the friability of the samples it was possible to cut and trim only one of the cores for compression testing. The ultimate compressive strength of the one core tested was 22.03 MPa (3195 psi). This one test result is obviously very selective and can be considered to represent an approximate upper limit for compressive strength in that it was the only one segment out of six cores sound enough to enable fabrication of a test sample. Indeed, portions of some cores could be crumbled by finger pressure. The one compressive strength test value obtained is below the design mix concrete strength of 24.13 MPa (3500 psi), but slightly greater than the structural design strength of 20.68 MPa (3000 psi). It must be assumed that the strength of the concrete in much of the mass might be well below the design strength.

C. Petrographic Examination

Core segments were petrographically examined by Dr. Richard Merriam, consulting engineering geologist, in September 1980. His examination revealed the aggregate consists largely of volcanic rocks, many of which are of approximately andesitic composition and have partly glassy groundmasses. Such rocks are known to be reactive with cement alkalies. Dark "reaction rims" were observed around the periphery of some of the broken andesitic aggregate particles as well as deposits of white silica gel in air voids and within cracks. Both are features of concrete experiencing alkali-aggregate reactivity, and Dr. Merriam concluded that the deterioration was due to alkali-aggregate reaction. About one year later, in October 1981, Mr. David Stark, Principal Research Petrographer of The Portland Cement Association, Skokie, Illinois, visited the Venus antenna to investigate our alkali-aggregate reactivity because none had previously been reported in the Barstow area. His on-site inspection supplemented by his petrographic study of JPL-supplied core segments led him to agree with Dr. Merriam that the Venus antenna foundation concrete was indeed experiencing alkali-aggregate reactivity.

D. Standard Tensile Splitting Test

The splitting tensile test was performed on two test cores by Twining Laboratories of Long Beach, California. The values obtained were 2.55 and 2.96 MPa (370 and 430 psi). The tests were selected to determine the influence of the reactivity microcracking on tensile strength. The values obtained were considered somewhat low by Twining Laboratories. Further, the percentage of broken aggregate exposed on the split sample faces was very small, which indicates that the bond of the aggregate to the paste matrix was low.

E. Rebound Number

The compressive strengths obtained using the rebound method were above 35.85 MPa (5200 psi). The rebound hammer, in this case the Schmidt type, is principally a surface hardness tester and there is little apparent theoretical relationship between the strength of concrete and the rebound number obtained using the hammer. Within limits, however, empirical correlations have been established between strength and the rebound number but there is a wide degree of disagreement among various researchers concerning the accuracy of the estimation of strength from rebound readings. The consensus among users is that it is useful only as a rough indication of concrete strength. In the case of the Venus foundation concrete, the hammer yielded much higher values of compressive strength than those obtained from tests of core samples taken immediately adjacent to hammer test areas. It is therefore believed that no conclusions can be drawn from the hammer data.

X. General Condition of the Concrete

The test results indicated the following condition of the concrete:

(1) The concrete is of questionable to poor quality.

(2) Compressive strength is below the design compressive strength.

(3) The aggregate is very reactive.

XI. Cause of Cracking

When all the test results were reviewed, the cracking was established to be principally the result of the alkali-aggregate reactivity. No corrosion of reinforcing steel was observed in the cores and no rust stains were observed on the concrete near cracks. Slight corrosion may be present but not in sufficient amounts to be disruptive. Corrosion is a major concern because the corrosion products occupy 2.2 times as much volume as the original metal and may develop pressures up to 32.41 MPa (4700 psi). This is considerably greater than the tensile strength of concrete which is generally less than 3.45 MPa (500 psi), and disruption of the concrete ensues. Once the corrosion begins, it continues as long as oxygen and moisture can reach the reinforcing steel. The dry desert atmosphere has no doubt helped in keeping corrosion to less than that needed to noticeably crack the concrete.

XII. Alkali-Aggregate Reactivity

The phenomenon of alkali-aggregate reactivity involves a chemical interaction produced by certain siliceous constituents

present in the aggregates with alkalies in the cement. The reaction produces silica gel, which then imbibes water from the surrounding concrete and swells. The resulting volumetric expansion damages the concrete by causing intense internal microfracturing with an attendant reduction in the strength of the concrete and a diminution of its elastic properties.

T.E. Stanton of the California Division of Highways first recognized the serious cracking and deterioration of a concrete pavement in the Salinas Valley, California, in 1938, and in 1940 he published a paper (Ref. 7) on the influence of cement alkalies on certain aggregates. Based on his work, the Division of Highways changed its cement specifications to include a top limit of 0.50 percent total alkali content (percent $Na_2O + 0.658x$ percent K_2O) for regions where reactive aggregates were used in making concrete. Subsequently, an alkali content of 0.6 percent of equivalent Na_2O was accepted in the United States as an upper limit for cement when used with reactive aggregates. No records exist to indicate that the cement used for construction of the Venus antenna foundation was certified as being a "low-alkali" type. Manufacturers in the Mojave Desert area have reported verbally that all of their cement production, regardless of type, has met the 0.6-percent limit since the early 1950s, which would cover the period of construction of DSS 13. Thus the best preventative practice of the time was followed, consciously or not, during construction. There are indeed no indications that the Barstow area concrete industry was aware, at the time of construction, that local aggregates were reactive.

In recent years, some investigators have come to realize that the use of low-alkali cement is not always effective in controlling deleterious expansion and in many cases only slows the reaction.

David Stark of the Portland Cement Association first reported in 1979 (Ref. 8) that field and laboratory observations indicated that certain glassy volcanic aggregates used with low-alkali cements can react deleteriously. Stark identifies the reactive volcanic materials as being of andesitic to rhyolitic composition. These material types were found in significant amounts in core samples taken from the Venus foundation.

XIII. Rehabilitation Program

The results of the investigation indicated that the useful life of the Venus antenna foundation could be successfully extended if a rehabilitation program were developed and implemented. The design approach assumed that the foundation concrete was satisfactory to carry the pedestal's vertical dead and live loads even though cracked, but not its base shear and/or uplift loads.

Structural analysis determined that the antenna pedestal base does not develop uplift from wind loading, but from seismic loading described in Ref. 9. To restore the effectiveness of the existing foundation, it was considered necessary to provide only enough additional structure to resist uplift and base shear loads. The foundation is still capable of performing its function of spreading the concentrated antenna base dead and live loads at its top surface over the larger area of the foundation bottom and thereby distribute the load to a bearing pressure allowable for the bearing or founding soils. The required resistance to uplift was provided by installing a structural system composed of steel-framed braces attached to each of the eight sloped steel columns in the antenna base core. The outer lower end of each brace is anchored to a large reinforced concrete grade beam, which, in turn, is anchored by a system of cast-in-place concrete friction piles. A plan and partial section of the structural system is shown in Fig. 6. Construction started July 5, 1983, and was completed, August 4, 1983. Figure 7 shows the installed system.

In addition to the installation of the braces, all accessible cracks, greater than approximately 1.59 mm ($^1/_{16}$ in.) wide, were pressure injected with a two-component epoxy adhesive to seal the surface and stop moisture intrusion into the reinforcing steel.

All aggregate used in the new concrete work was obtained from the Owl Rock Co. pit located near San Bernardino, California. Petrographic study of this material showed it to be completely free of minerals known to be reactive.

XIV. Monitoring Program

Although the useful life of the Venus Antenna has been extended, periodic inspections will be required to observe the formation of new cracks or widening of the injected cracks. Further cracking is to be expected in that the alkali-aggregate reaction can continue for long periods of time before stopping. As new cracks occur, or old ones widen, additional epoxy adhesives will be injected. Thus the useful life of the antenna can be prolonged.

References

1. *Standard Method of Test for Pulse Velocity Through Concrete.* ASTM C 597, American Society for Testing and Materials, Philadelphia, Pennsylvania, 1982.

2. Malhotra, V. M., *Testing Hardened Concrete: Nondestructive Methods.* Monograph No. 9, American Concrete Institute, Detroit, Michigan, 1976.

3. *Standard Method of Obtaining and Testing Drilled Cores and Sawed Beams of Concrete.* ASTM C 42, American Society for Testing and Materials, Philadelphia, Pennsylvania, 1982.

4. ACI Committee 318, *Building Code Requirements for Reinforced Concrete.* ACI Standard 318, 8th Printing, American Concrete Institute, Detroit, Michigan, 1976.

5. *Standard Test Method for Splitting Tensile Strength of Cylindrical Concrete Specimens.* ASTM C 496, American Society for Testing and Materials, Philadelphia, Pennsylvania, 1982.

6. *Test for Rebound Number of Hardened Concrete.* ASTM C 805, American Society for Testing and Materials, Philadelphia, Pennsylvania, 1982.

7. Stanton, T. E., "Expansion of Concrete Through Reaction Between Cement and Aggregate," *Proceedings of the American Society of Civil Engineers*, Vol. 66, 1940, pp. 1781–1811.

8. Stark, D., "Alkali-Silica Reactivity: Some Reconsiderations," *Cement, Concrete and Aggregates*, CCAGDP, Vol. 2, No. 2, Winter 1980, pp. 92–94.

9. *Uniform Building Code, Seismic Zone IV.* International Conference of Building Officials, Whittier, California (latest issue).

Table 1. Supporting contractors

Name	Service
Pacific Materials Laboratory, Inc., Bloomington, Calif.	Concrete mix design
Concrete Materials Co., Barstow, Calif.	Source of aggregate

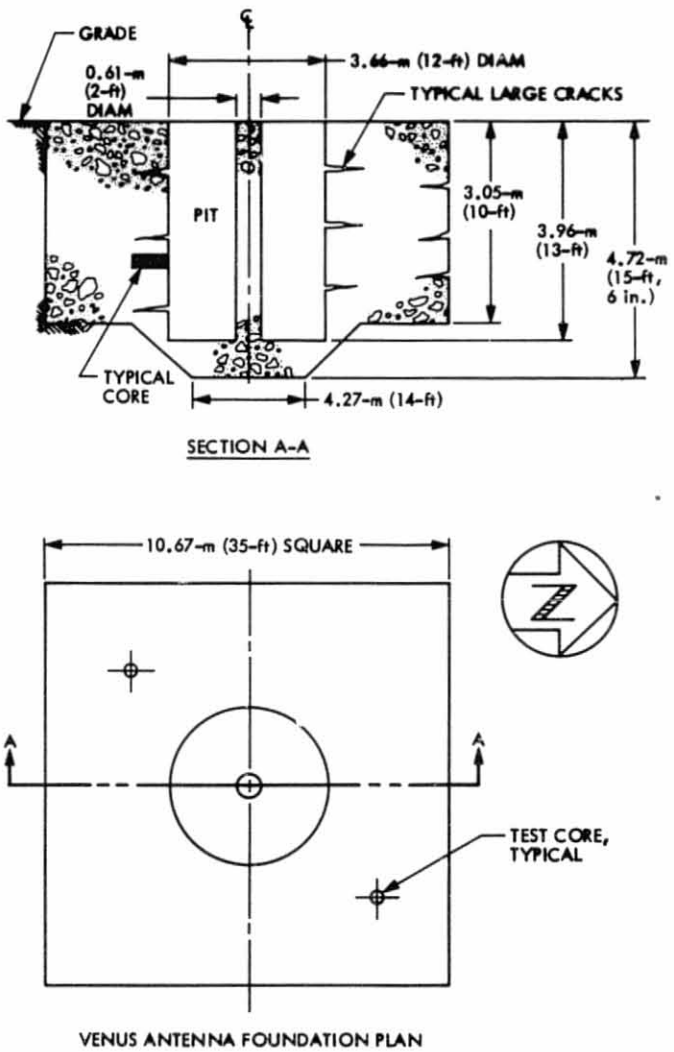


Fig. 1. Partial plan and section of original foundation

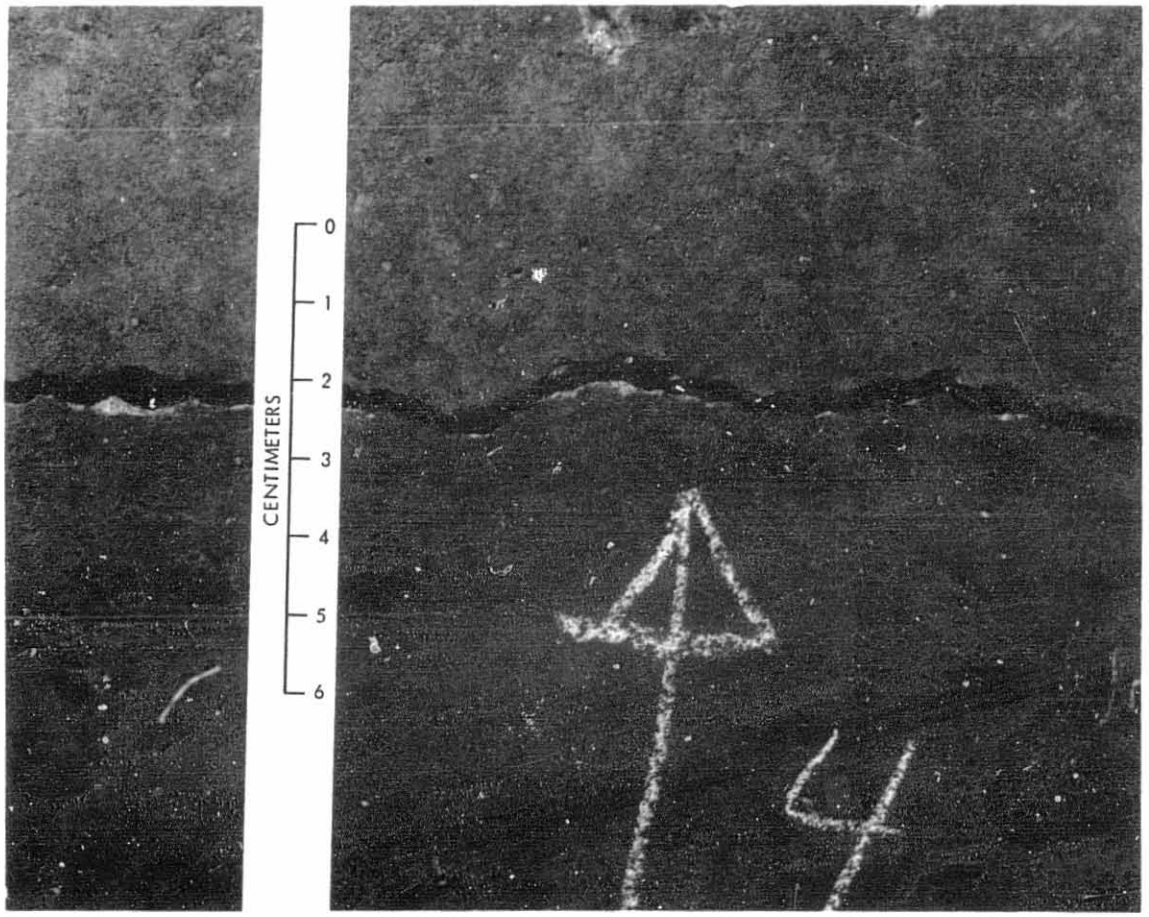


Fig. 2. Crack in pit wall

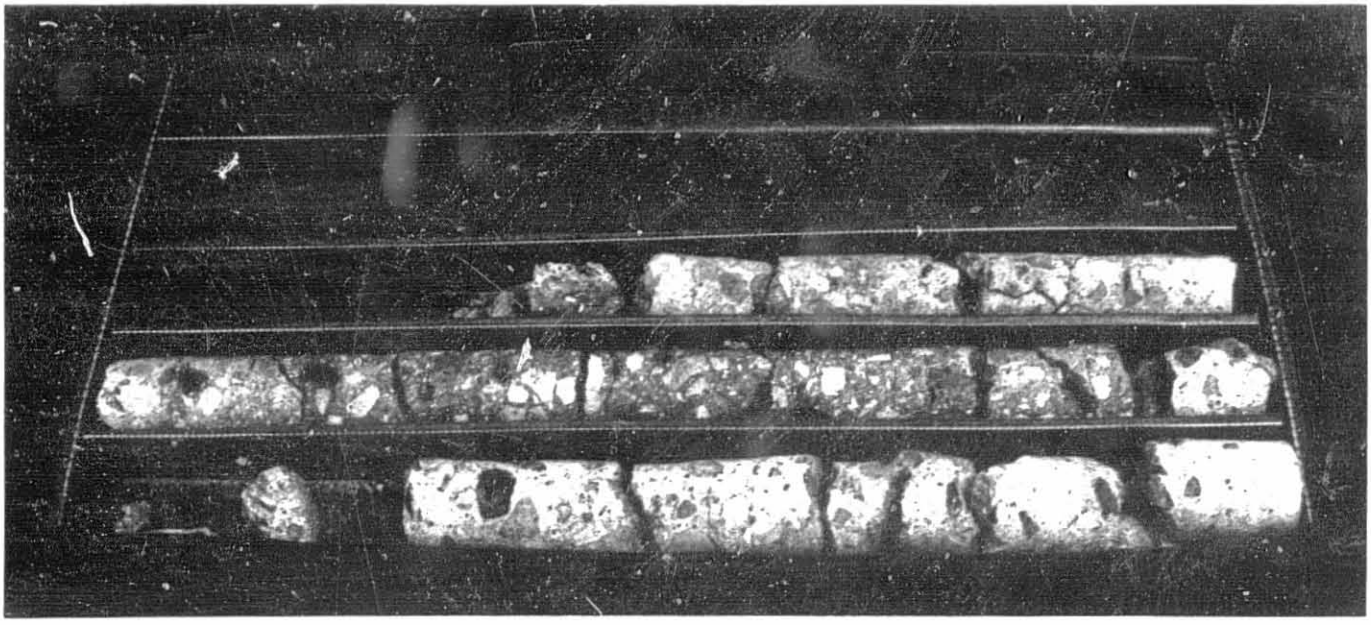


Fig. 3. Vertical core No. 1

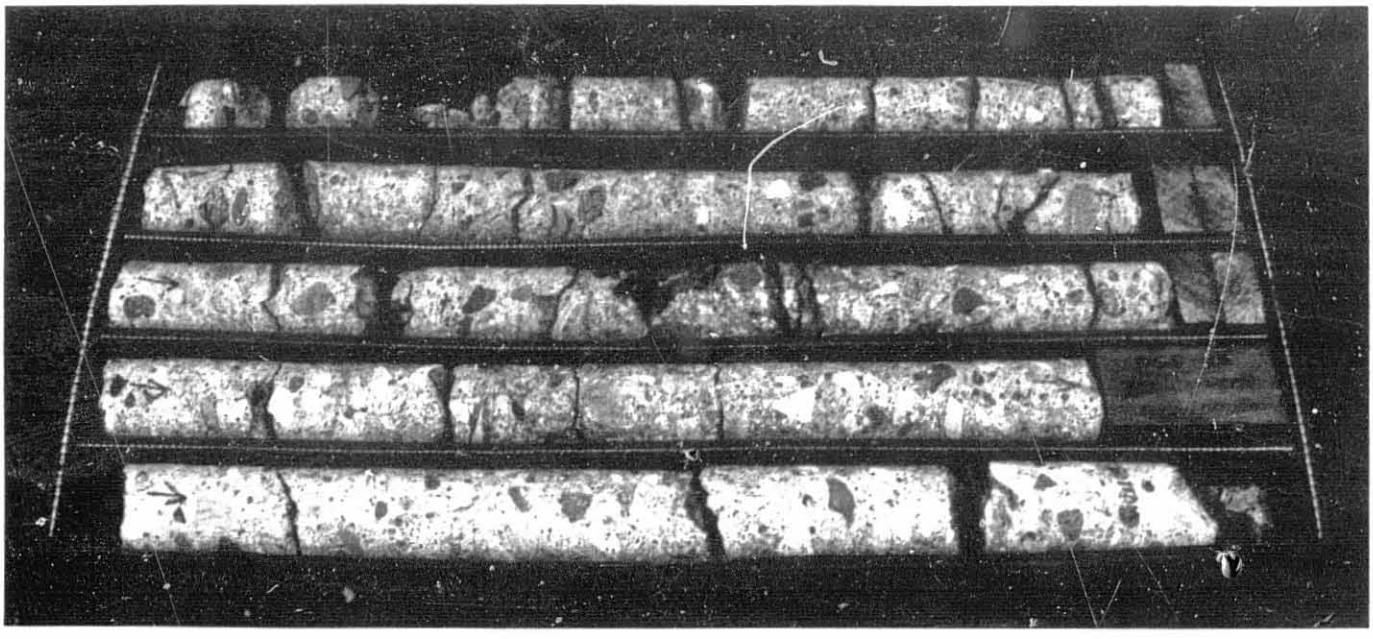


Fig. 4. Vertical core No. 2

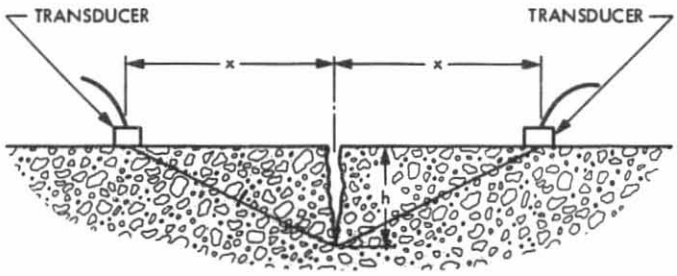
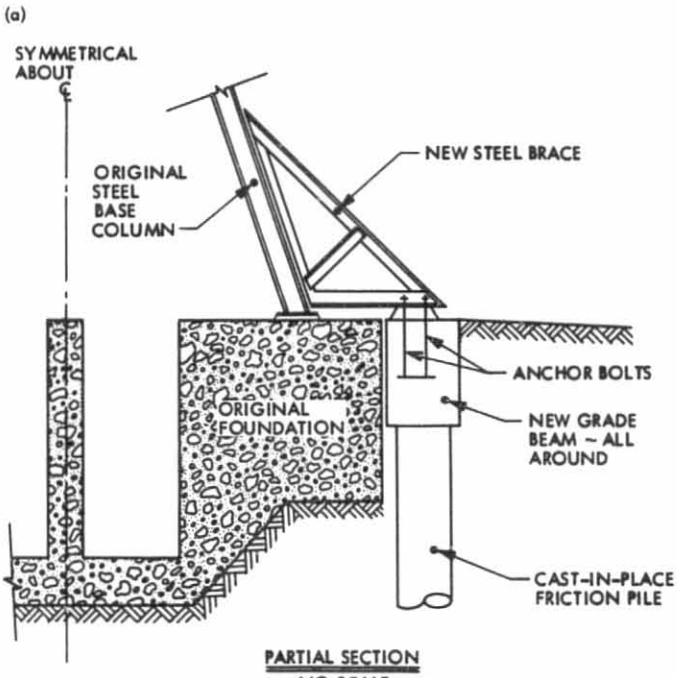


Fig. 5. Measurement of crack depth



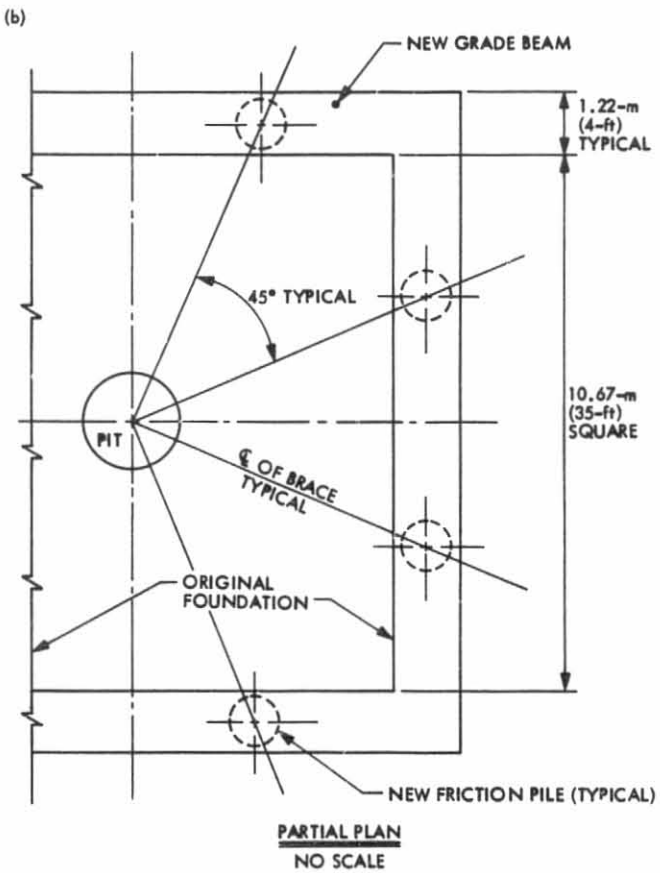


Fig. 6. Repair scheme (no scale): (a) partial section; (b) partial plan

Fig. 7. The installed system

NASTRAN Structural Model for the Large 64-Meter Antenna Pedestal Part III – Applications to Hydrostatic Bearing Oil Film

C. T. Chian and D. Schonfeld
Ground Antennas and Facilities Engineering Section

Investigations were conducted on the 64-meter antenna hydrostatic bearing oil film thickness under a variety of loads and elastic moduli. These parametric studies used a NASTRAN pedestal structural model to determine the deflections under the hydrostatic bearing pad. The deflections formed the input for a computer program developed by the Franklin Institute to determine the hydrostatic bearing oil film thickness. For the future 64-meter to 70-meter antenna extension and for the 2.2-meter (86-in.) haunch concrete replacement cases, the program predicted safe oil film thickness (greater than 0.13 mm (0.005 in.) at the corners of the pad). The effects of varying moduli of elasticity for different sections of the pedestal and the film height under distressed runner conditions were also studied.

I. Introduction

This is the third and final article in a series of reports on the static analysis and computer modeling for the large 64-meter antenna pedestal. The pedestal structural model previously reported in Refs. 1 and 2 was developed using the MSC version of the NASTRAN Program.

The top surface deflection of the pedestal obtained from the NASTRAN model was used as an input to the hydrostatic bearing oil film program (Ref. 3) to determine the minimum oil film thickness between the hydrostatic bearing pad and the runner. The knowledge of the oil film thickness was necessary to conduct a variety of hydrostatic bearing rehabilitation studies. A minimum oil film thickness of 0.13 mm (0.005 in.) is considered necessary for safe operation, to avoid any metal-to-metal contact between the hydrostatic bearing, and to accommodate a variety of runner malfunctions and placement tolerances.

A cross-sectional diagram of the hydrostatic bearing is shown in Fig. 1. Deflected shapes of the hydrostatic bearing pad and runner surface are illustrated in Fig. 2.

Three parametric studies were conducted to evaluate the performance of the hydrostatic bearing system. Effects on the oil film thickness due to the following factors were considered in each of the three parametric studies:

(1) The height of the new concrete in the pedestal haunch area.

(2) The different moduli of elasticity of the concrete in the pedestal wall and haunch area.

(3) The hydrostatic bearing pad load increase due to the planned antenna aperture extension from 64-meter to 70-meter.

The results of these parametric studies will be presented in Section III.

II. Description of the Oil Film Computer Program

The hydrostatic bearing computer program was developed by the Franklin Institute (Philadelphia, Pa.) under contract with the Jet Propulsion Laboratory (Ref. 3). The program's original use was to support the design of the 64-m antenna. At that time, the program included a number of capabilities previously unavailable in other models. Among these capabilities was the ability to analyze nonuniform film thickness, tilting moments, and various lubricant supply modes. This program is used in conjunction with the NASTRAN pedestal model to predict hydrostatic bearing oil film thicknesses under a variety of load conditions. In this section, a description of the Franklin program is presented to indicate how it was used for the 64-m antenna pedestal studies.

The derivation of the mathematical equations used in the program is based on a number of assumptions. These assumptions are:

(1) Incompressible fluid.

(2) Two-dimensional laminar flow.

(3) Steady-state conditions.

For bearings with a negligible amount of relative motion, the Reynolds Equation becomes

$$\frac{\partial}{\partial x}\left(h^3\frac{\partial P}{\partial x}\right)+\frac{\partial}{\partial y}\left(h^3\frac{\partial P}{\partial y}\right)=0 \tag{1}$$

The program solves Eq. (1) in its dimensionless form. The dimensionless form is obtained using a characteristic length L, a characteristic oil film thickness c, and a characteristic pressure P_r.

The result is:

$$\left(\frac{\partial^2 P}{\partial X^2}+\frac{\partial^2 P}{\partial Y^2}\right)+\frac{3}{H}\left(\frac{\partial H}{\partial X}\frac{\partial P}{\partial X}+\frac{\partial H}{\partial Y}\frac{\partial P}{\partial Y}\right)=0 \tag{2}$$

$$P \equiv \frac{P-P_{atm}}{P_r-P_{atm}}$$

$$H = h/c$$

$$X=\frac{x}{L} \qquad Y=\frac{y}{L}$$

where P_{atm} is the atmospheric pressure.

The program makes use of the fact that Eq. (2) is linear in pressure, P, in order to obtain a solution by superposition. Superposition is obtained by assuming that each recess is pressurized in turn with the pressure at other recesses equal to zero (Fig. 3). The final pressure distribution can then be expressed as a linear combination of the individual solutions:

$$P(X, Y)=\sum_j \alpha_j P_j(X, Y) \tag{3}$$

where α_j is a dimensionless pressure weighting factor for each recess:

$$\alpha_j=\frac{(P_r)_j-P_{atm}}{P_{ref}-P_{atm}} \tag{4}$$

The program has the ability to solve for various lubricant supply schemes, with the following options:

(1) Separate pumps feed each recess.

(2) Separate pumps feed opposite pairs of recesses with capillary compensation.

(3) A common manifold feeds all recesses with capillary compensation.

For the work presented here, option (1) was used throughout, since that is the supply scheme on the 64-m antennas. Once the program obtains a pressure distribution map $P(X, Y)$, other quantities can be computed as well. In particular we are interested in the clearance distribution $H(X, Y)$.

The general shape of the clearance distribution can be obtained either by the evaluation of an appropriate analytic function or by point-by-point input. The program contains a number of coefficients, $A_1, A_2, \cdots A_{23}$ that are used to specify possible functions for $H(X, Y)$. The general formula is:

$$\begin{aligned}H(X, Y) = {} & A_1+A_2 s+A_3 t+A_4 s^2+A_5 t^2+A_6 st \\ & +A_7 s^3+A_8 t^3+A_9 s^2 t+A_{10} st^2 \\ & +A_{11}\sqrt{A_{12}+A_{13}+A_{14}t^2} \\ & +A_{15}\cos(A_{16}s)+A_{17}\cos(A_{18}t) \\ & +A_{19}\cos(A_{16}s)\cos(A_{18}t) \\ & +A_{20}\left\{e^{-A_{21}X}\cos(A_{21}X)\right. \\ & +e^{-A_{21}(1-X)}\cos[A_{21}(1-X)] \\ & \left.-2e^{-A_{21}/2}\cos(A_{21}/2)\right\}\end{aligned} \tag{5}$$

C-3

In this formula, the dimensionless coordinates s and t are given by:

$$s = X - X_0 = X - A_{22}$$
$$t = Y - Y_0 = Y - A_{23} \quad (6)$$

with X_0 and Y_0 denoting the geometric center of the pad. As an example, for a uniform clearance distribution A_1 = constant and all A_i $(i \neq 1) = 0$. A doubly-parabolic shape was assumed to approximate the actual runner deflection pattern and hence the clearance distribution, and has provided acceptable results. In terms of the formula shown above, the resulting equation is:

$$H(X, Y) = A_1 + A_4 s^2 + A_5 t^2 \quad (7)$$

The tie-in between the input to the Franklin program and the NASTRAN pedestal model output is the runner deflection map. Given a set of pedestal/bearing loads and elastic properties, the NASTRAN model predicts a deflection map for the runner area under the pad (Fig. 4). The deflections at the end points of the center line $(s, t) = (0, 1/3), (0, -1/3), (0.5, 0)$ and at the center point $(0, 0)$ are used to find the coefficients A_i in Eq. (5). Fitting these points gives a representation of the whole surface to within 5% of the actual deflections, which is considered sufficiently accurate.

To obtain the actual oil film thickness, the A_1 coefficient is varied until the pad load computed by the Franklin program is equal to the actual load on the pad. For most of our work, we determined oil film thickness based on the pad 3 load of 1.1×10^6 kg (2.4×10^6 lb). For the planned 70-meter extension load analysis, we varied this load to a maximum of 1.6×10^6 kg (3.6×10^6 lb).

The flexibility of Eq. 5 allowed various other deflection configurations to be tried, in particular, the case of twist in the runner. The clearance equation for this case is given by

$$H(X, Y) = A_1 + A_4 s^2 + A_5 t^2 + A_6 st \quad (8)$$

or, for a simpler planar twist:

$$H(X, Y) = A_1 + A_6 st \quad (9)$$

Note that this is taken as an extreme case of runner distress, reflecting actual profile measurements in the past.

Another advantage of Eq. 5 is that it allows us to consider pad deflection. By reducing the s^2 term in Eq. 5, for example, one can account for a parabolic deflection in the s-direction of the pad. The t-direction deflection was not considered because most pad deflection occurred in the circumferential direction (s). The results showed that minimum oil film thickness occurred at the corners of the pad. Based on previous operational experience, a 0.13-mm (0.005-in.) oil film thickness at the pad corners is considered a minimum acceptable value. The flow chart of the Franklin computer program is given in the Appendix and Ref. 3.

III. Parametric Studies

Three parametric studies were conducted to evaluate the operability of the large 64-meter antenna:

(1) Effect on the oil film thickness due to the height variation of the new concrete in the pedestal haunch.

(2) Effect on the oil film thickness due to the variation of concrete elastic moduli in the pedestal wall and haunch area.

(3) Effect on the oil film thickness due to the pad load increase for an antenna aperture extension from 64 meters to 70 meters.

A NASTRAN pedestal model was used to obtain the pedestal top surface deflections. These in turn served as the input to the hydrostatic bearing computer program for determining the oil film thickness between the hydrostatic bearing pad and the runner.

Two design characteristics are used to evaluate the sensitivity of the hydrostatic bearing pad operation to the modulus of elasticity. The first characteristic is the maximum pad out-of-flatness. Deflected shapes of the hydrostatic bearing pad and runner surface are illustrated in Fig. 2. Relative deflections within the hydrostatic bearing pad and within the runner surface (from centerline to edge of pad) are shown as Δ_p and Δ_r, respectively.

Design criteria (Ref. 4) require that the mismatch of deflected surfaces, $\Delta\delta$, be within 0.101 mm (0.004 in.). (This is the variation of the film height between the pad and the runner.) Out of this a maximum mismatch of deflected shapes of 0.076 mm (0.003 in.) was established as the allowance for creep during construction before the bearing pads could be moved. The remaining 0.025 mm (0.001 in.) was the design criteria for mismatch of elastic deformations (Ref. 4). Since creep strains have been compensated for by releveling of the runner, the maximum pad out-of-flatness, a $\Delta\delta$ of 0.101 mm (0.004 in.), can now all be accounted for by elastic deformations. These elastic deformations are part of the NASTRAN output.

The second characteristic used to evaluate the operability of the hydrostatic bearing is the minimum oil film thickness between the pad and the runner. Based on previous operational experience, a minimum oil film thickness, h, of 0.127 mm (0.005 in.) is considered necessary for safe operation. Each of the three parametric studies is explained below.

A. Height of New Concrete in the Pedestal Haunch

In Ref. (1) we used the original pedestal model without including the pedestal haunch lip. The pedestal concrete, with an initial modulus of elasticity E of 2.1×10^{10} N/m^2 (3×10^6 psi), was replaced by a new concrete with the modulus of elasticity of 3.5×10^{10} N/m^2 (5×10^6 psi) at various heights from the top.

In Ref. (2) we used the improved pedestal model, which included the pedestal haunch lip. As before, the pedestal concrete with an initial modulus of elasticity E of 2.1×10^{10} N/m^2 (3×10^6 psi) was replaced by a new concrete with the modulus of elasticity of 3.5×10^{10} N/m^2 (5×10^6 psi) at different heights from the top. Results of this parametric study are shown in Tables 1 and 2, as well as in Figs. 5 and 6.

B. Variation of Concrete Elastic Moduli in the Pedestal Wall and Haunch Area:

The severity of the concrete deterioration with accompanying reduction in compressive strength and modulus of elasticity varies widely throughout the pedestal mass. Studies to date have shown that the most serious damage was in the haunch area. A height of 2.2 m (86 in.) of the concrete in the haunch area has been replaced as part of the rehabilitation efforts.

Portions of the remaining pedestal concrete not replaced have experienced moderate damage and are expected to drop further in strength and modulus of elasticity in the future since the alkali-aggregate reaction (the main reason of deteriorations) is continuous, and not fully understood. Therefore, this study was made to evaluate the operability of the hydrostatic bearing under these continuous deteriorations. The moduli of elasticity of the concrete in the pedestal wall and the haunch area were varied. This study was further subdivided into two parts:

(1) The new haunch area down to a depth 2.2 m (86 in.) was assigned a fixed modulus of elasticity of 3.5×10^{10} N/m^2 (5×10^6 psi), while the modulus of elasticity of the remaining wall was taken to be 2.1×10^{10} N/m^2 (3×10^6 psi), 1.4×10^{10} N/m^2 (2×10^6 psi), and 0.7×10^{10} N/m^2 (1×10^6 psi), to simulate time deteriorations. Note that tests made on replaced concrete showed $E > 3.5 \times 10^{10}$ N/m^2 (5×10^{16} psi).

(2) The pedestal wall was assumed to have a fixed modulus of elasticity of 1.4×10^{10} N/m^2 (2×10^6 psi), while the new haunch area was assigned a modulus of elasticity of 3.5×10^{10} N/m^2 (5×10^6 psi), 3.15×10^{10} N/m^2 (4.5×10^6 psi) and 2.8×10^{10} N/m^2 (4×10^6 psi) to simulate different values of the replaced concrete.

Results of this parametric study showing the effect on the oil film thickness due to the variation of concrete elastic moduli are summarized in Tables 3 and 4. Figures 7 and 8 also give the results of this study.

C. Pad Load Increase With an Antenna Aperture Extension From 64-meters to 70-meters

This study investigates the effects of the increased pad load of the antenna with an aperture extension from 64 meters to 70 meters on the pedestal deflection and the oil film thickness. Pad 3 was assumed to have a load of 1.1×10^6 kg (2.4×10^6 lb). In this study, four loads of 1.1×10^6 kg (2.4×10^6 lb), 1.3×10^6 kg (2.8×10^6 lb), 1.45×10^6 kg (3.2×10^6 lb), and 1.6×10^6 kg (3.6×10^6 lb) were considered for pad 3, which correspond to load factors of 1.00; 1.17; 1.33; and 1.50, respectively, relative to the estimated present 64-meter pad 3 load. This load was based on integration of recess pressure readings. The modulus of elasticity was assumed to be 3.5×10^{10} N/m^2 (5×10^6 psi) for both the pedestal wall and the haunch area. The maximum film height variation, $\Delta\delta$, and the minimum film thickness, h, are given in Table 5 for the four loads considered. The results are also shown in Fig. 9.

IV. Conclusions

In this study we reported on applications of the NASTRAN pedestal model to the hydrostatic bearing oil film for the large 64-meter antenna. The NASTRAN model gave as one result the top surface deflections of the pedestal. These deflections formed the input for the hydrostatic bearing oil film computer program to determine the minimum oil film thickness.

The knowledge of the minimum oil film thickness between the hydrostatic bearing pad and the runner was required to conduct a variety of hydrostatic bearing rehabilitation studies.

Based on results presented in this study, a height of 2.2 meters (86 in.) of concrete in the top-most pedestal haunch area has been replaced in the DSS 14 as part of the rehabilitation efforts. For a new concrete with the modulus of

elasticity of 3.5 X 10^{10} N/m^2 (5 X 10^6 psi), the study predicted a safe oil film thickness of more than 0.13 mm (0.005 in.).

The effect on the oil film thickness due to the pad load increase for an antenna aperture extension from 64-meters to 70-meters was also investigated. For a pad load increase of up to 20%, the study predicted a safe oil film thickness.

The techniques developed in this study will also be applicable to future rehabilitation studies of the large 64-meter antennas in the DSSs 43 and 63.

Acknowledgments

The authors acknowledge the assistance given by H. Phillips, A. Riewe, F. Lansing, D. McClure, F. McLaughlin, M. Pompa, D. Wells, H. McGinness, M. S. Katow, and S. Rocci during the various steps of this work. We also gratefully acknowledge the assistance of Dr. V. Castelli (Xerox Palo Alto Research Center) with the Franklin Institute program.

References

1. Chian, C. T., M. S. Katow, and H. McGiness, "NASTRAN Structural Model for the Large 64-m Antenna Pedestal, Part I", *TDA Progress Report 42-74*. Jet Propulsion Laboratory, Pasadena, California, August 15, 1983.

2. Chian, C. T., "NASTRAN Structural Model for the Large 64-Meter Antenna Pedestal, Part II – Improved Model", *TDA Progress Report 42-75*. Jet Propulsion Laboratory, Pasadena, California, November 15, 1983.

3. Hinkle, J. G., and V. Castelli, *A Computer Solution for Hydrostatic Bearings with Variable Film Thickness*. The Franklin Institute, Philadelphia, Pennsylvania, Report No. F-B2015, Jan. 11, 1963. JPL reorder No. 63-615, NASA CR 56898.

4. TDA Technical Staff, *The NASA/JPL 64-Meter-Diameter Antenna at Goldstone California: Project Report*. JPL Technical Memorandum 33-671, Jet Propulsion Laboratory, Pasadena, California, July 15, 1974.

Table 1. Effect on the oil film thickness due to the height variation of the new concrete in the pedestal haunch[a,b]

Description, N/m^2 (psi)	Film height variation $\Delta\delta$, mm (in.)	Minimum oil film thickness h, mm (in.)
Entire pedestal: $E = 2.1 \times 10^{10}$ (3×10^6)	0.236 (0.0093)	0.100 (0.0042)
Top 1.4 m (56 in.): $E = 3.5 \times 10^{10}$ (5×10^6) Remaining pedestal: $E = 2.1 \times 10^{10}$ (3×10^6)	0.158 (0.0062)	0.190 (0.0075)
Top 2.2 m (86 in.): $E = 3.5 \times 10^{10}$ (5×10^6) Remaining pedestal: $E = 2.1 \times 10^{10}$ (3×10^6)	0.150 (0.0059)	0.193 (0.0076)

[a]See Section III.A.
[b]Based on the original pedestal model (Ref. 1).

Table 2. Effect on the oil film thickness due to the height variation of the new concrete in the pedestal haunch[a,b]

Description, N/m^2 (psi)	Film height variation $\Delta\delta$, mm (in.)	Minimum oil film thickness h, mm (in.)
Entire pedestal: $E = 2.1 \times 10^{10}$ (3×10^6)	0.147 (0.0058)	0.132 (0.0052)
Top 1.4 m (56 in.): $E = 3.5 \times 10^{10}$ (5×10^6) Remaining pedestal: $E = 2.1 \times 10^{10}$ (3×10^6)	0.102 (0.0040)	0.196 (0.0077)
Top 2.2 m (86 in.): $E = 3.5 \times 10^{10}$ (5×10^6) Remaining pedestal: $E = 2.1 \times 10^{10}$ (3×10^6)	0.097 (0.0038)	0.191 (0.0075)

[a]See Section III.A.
[b]Based on the improved pedestal model (Ref. 2).

Table 3. Effect of varying the modulus of elasticity of the pedestal wall[a,b]

Modulus of elasticity of the pedestal wall, N/m^2 (psi)	Film height variation $\Delta\delta$, mm (in.)	Minimum oil film thickness h, mm (in.)
2.1×10^{10} (3×10^6)	0.097 (0.0038)	0.193 (0.0076)
1.4×10^{10} (2×10^6)	0.102 (0.0040)	0.191 (0.0075)
0.7×10^{10} (1×10^6)	0.119 (0.0047)	0.178 (0.0070)

[a]See Section III.B.
[b]The modulus of elasticity of the top 2.2 m (86 in.) in the haunch is considered to be fixed at 3.5×10^{10} N/m^2 (5×10^6 psi).

Table 4. Effect of varying the modulus of elasticity of the haunch area[a,b]

Modulus of elasticity of the top 2.2 m (86 in.) in the haunch, N/m^2 (psi)	Film height variation $\Delta\delta$, mm (in.)	Minimum oil film thickness h, mm (in.)
3.5×10^{10} (5×10^6)	0.102 (0.0040)	0.191 (0.0075)
3.15×10^{10} (4.5×10^6)	0.112 (0.0044)	0.152 (0.0060)
2.8×10^{10} (4×10^6)	0.125 (0.0049)	0.152 (0.0060)

[a]See Section III.B.
[b]The modulus of elasticity of the pedestal wall is assumed to be fixed at 1.4×10^{10} N/m^2 (2×10^6 psi).

Table 5. Effect of the pad load increase due to the antenna extension[a,b]

Pad (No. 3) load, kg (lb)	Load factor	Film height variation, $\Delta\delta$, mm (in.)	Minimum oil film thickness, h, mm (in.)
1.09×10^6 (2.4×10^6)	1.00	0.089 (0.0035)	0.185 (0.0073)
1.27×10^6 (2.8×10^6)	1.17	0.104 (0.0041)	0.152 (0.0060)
1.45×10^6 (3.2×10^6)	1.33	0.119 (0.0047)	0.122 (0.0048)
1.63×10^6 (3.6×10^6)	1.50	0.135 (0.0053)	0.086 (0.0034)

[a]See Section III.C.
[b]The entire pedestal is assumed to have a modulus of elasticity of 3.5×10^{10} N/m^2 (5×10^6 psi) in all cases.

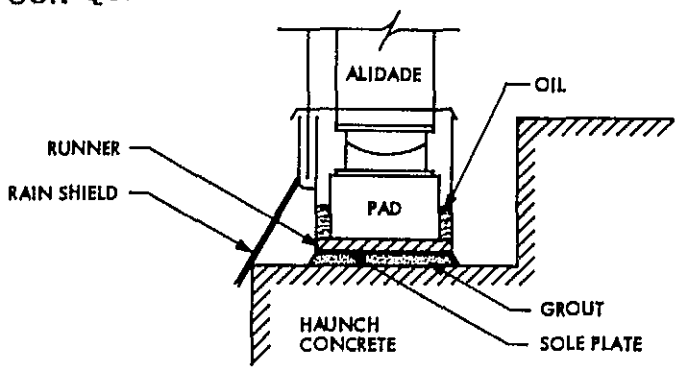
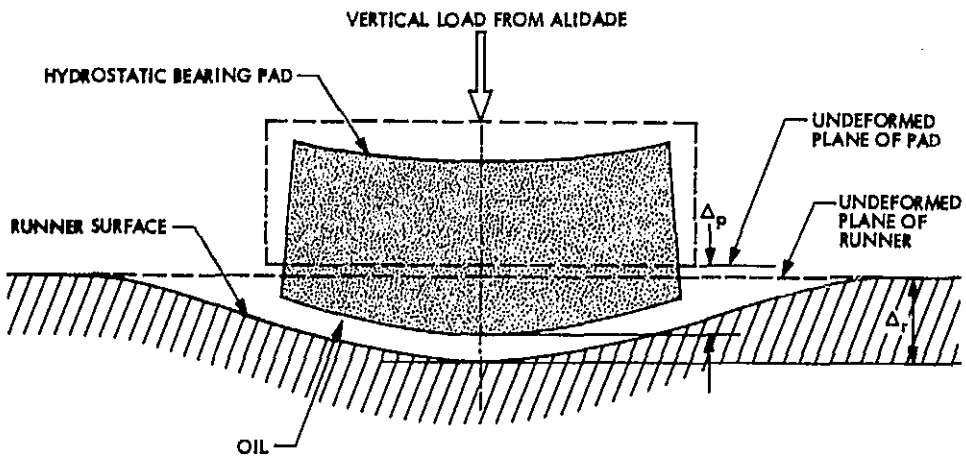


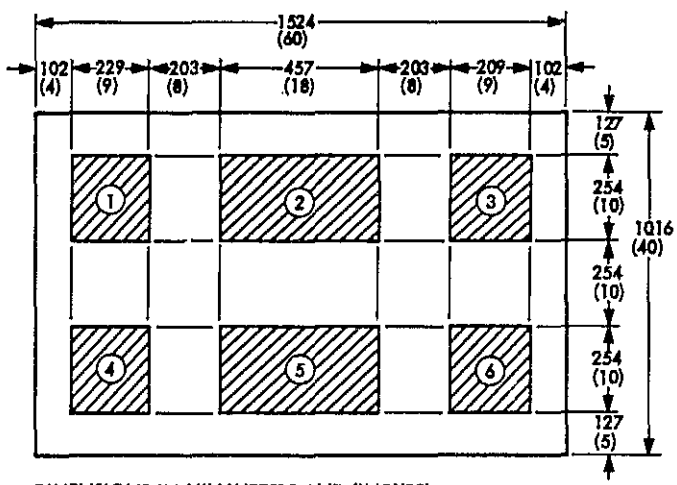
Fig. 1. Cross section of hydrostatic bearing system



δ = MISMATCH OF DEFLECTED SURFACES

$\delta = \Delta_r - \Delta_p$

Fig. 2. Deflections of hydrostatic bearing pad and runner surface



DIMENSIONS IN MILLIMETERS AND (INCHES)

Fig. 3. Recess pattern of hydrostatic bearing pad

ORIGINAL PAGE IS
OF POOR QUALITY

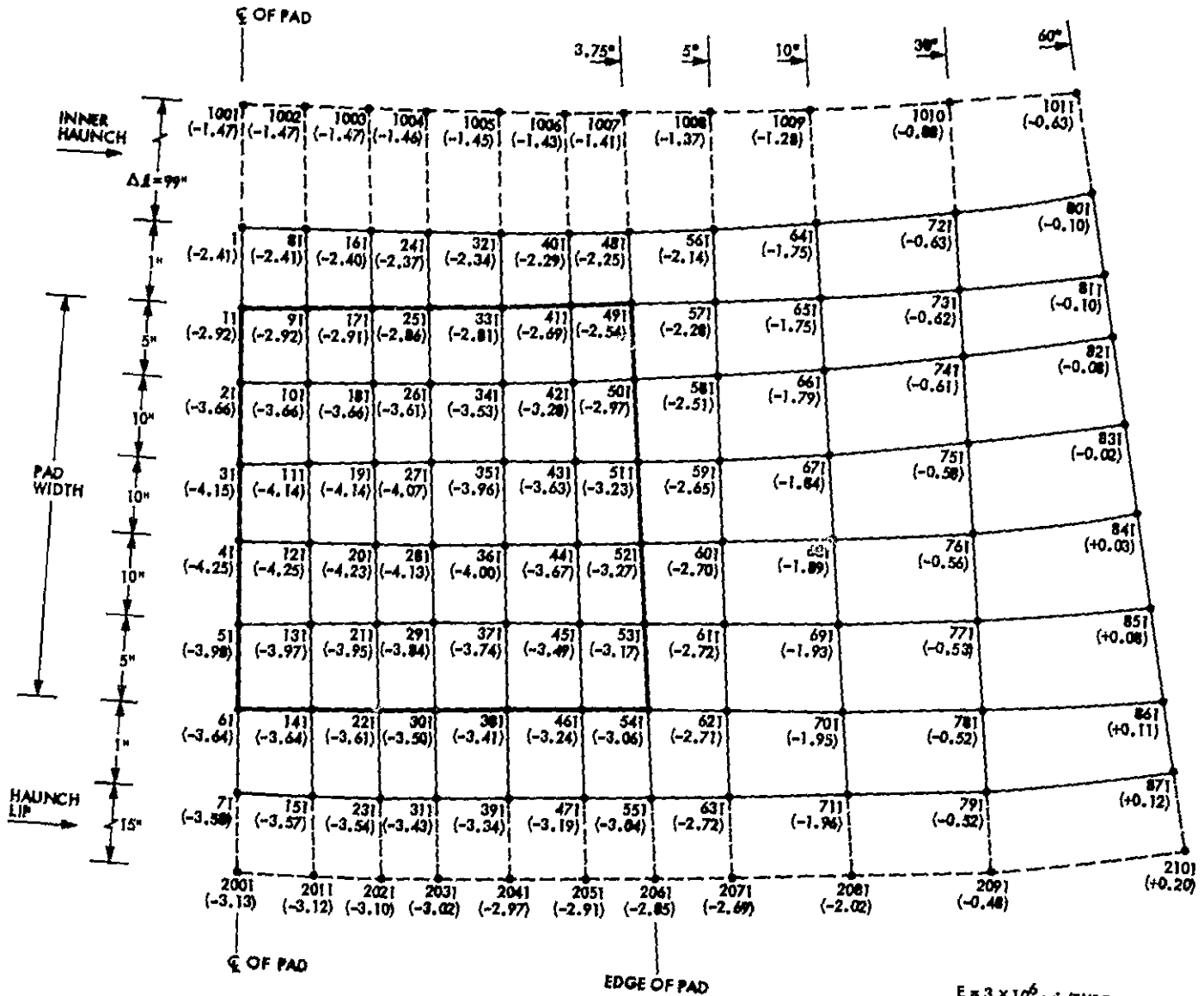


Fig. 4. Typical deflection map of top pedestal surface

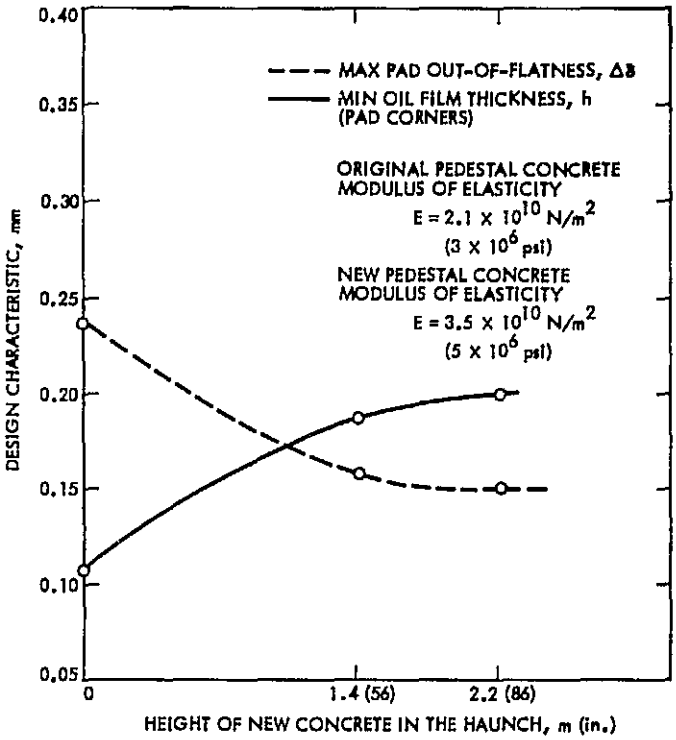


Fig. 5. Effect on the oil film thickness due to the height variation of the new concrete in the pedestal haunch (original pedestal model, Ref. 1)

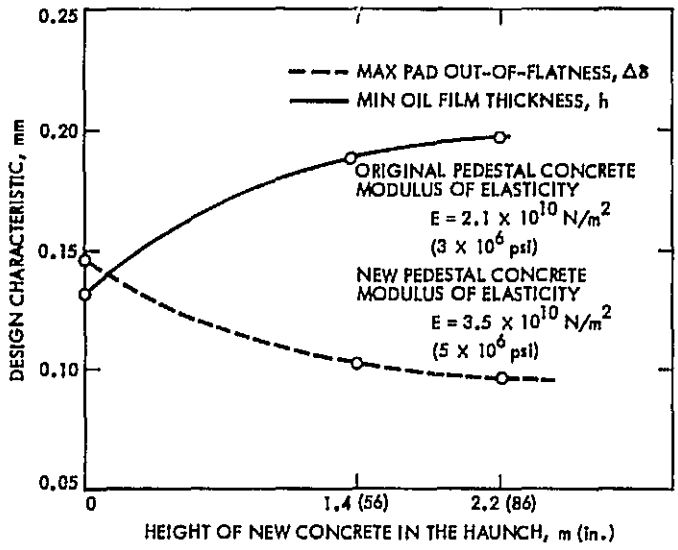


Fig. 6. Effect on the oil film thickness due to the height variation of the new concrete in the pedestal haunch (improved pedestal model, Ref. 2)

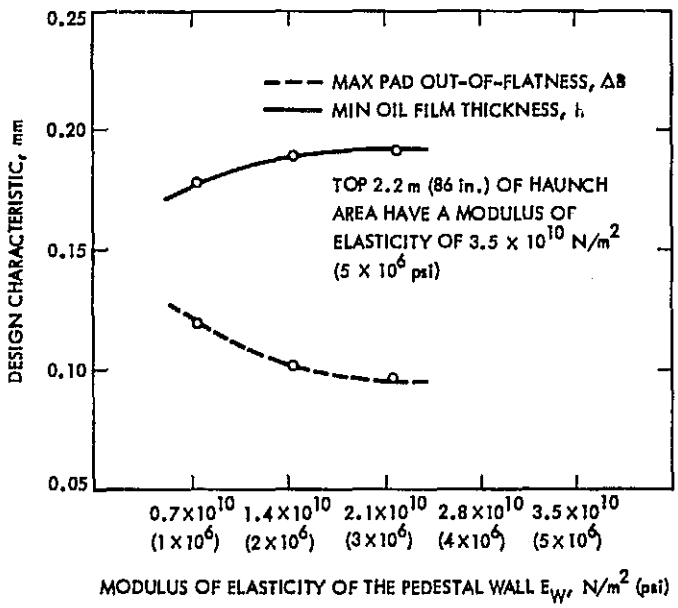


Fig. 7 Effect of varying the modulus of elasticity of the pedestal wall

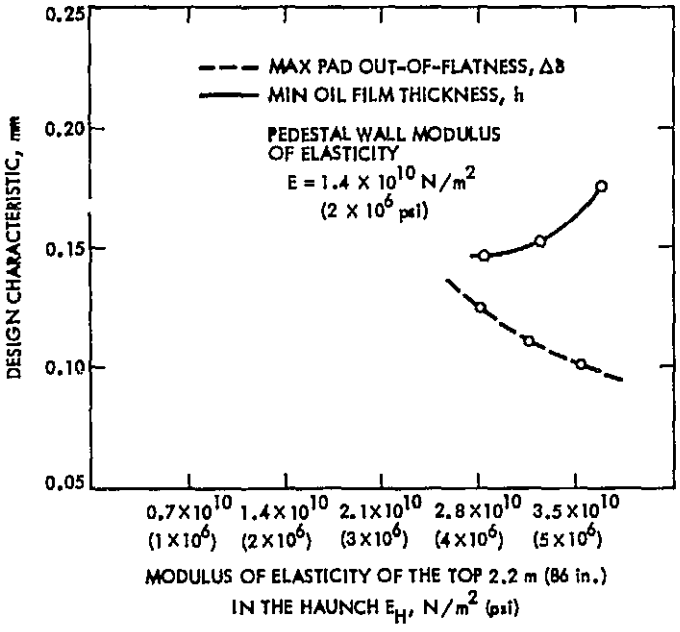


Fig. 8. Effect of varying the modulus of elasticity of the haunch area

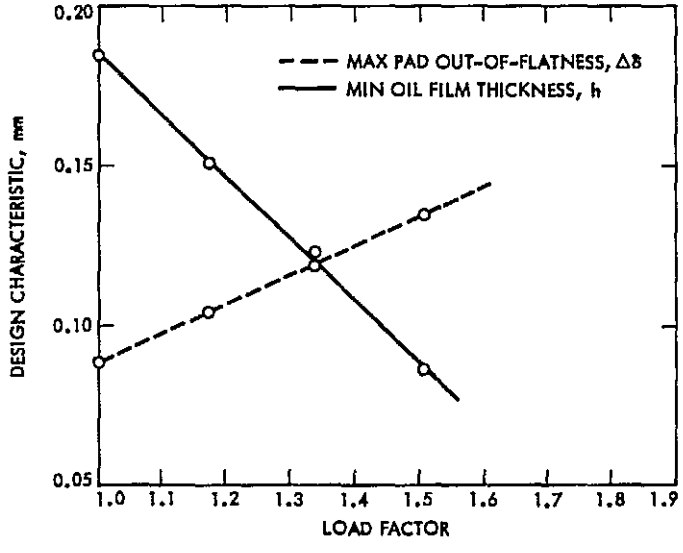


Fig. 9. Effect of the pad load increase due to the antenna extension

Appendix

Oil Film Computer Program Flow Chart

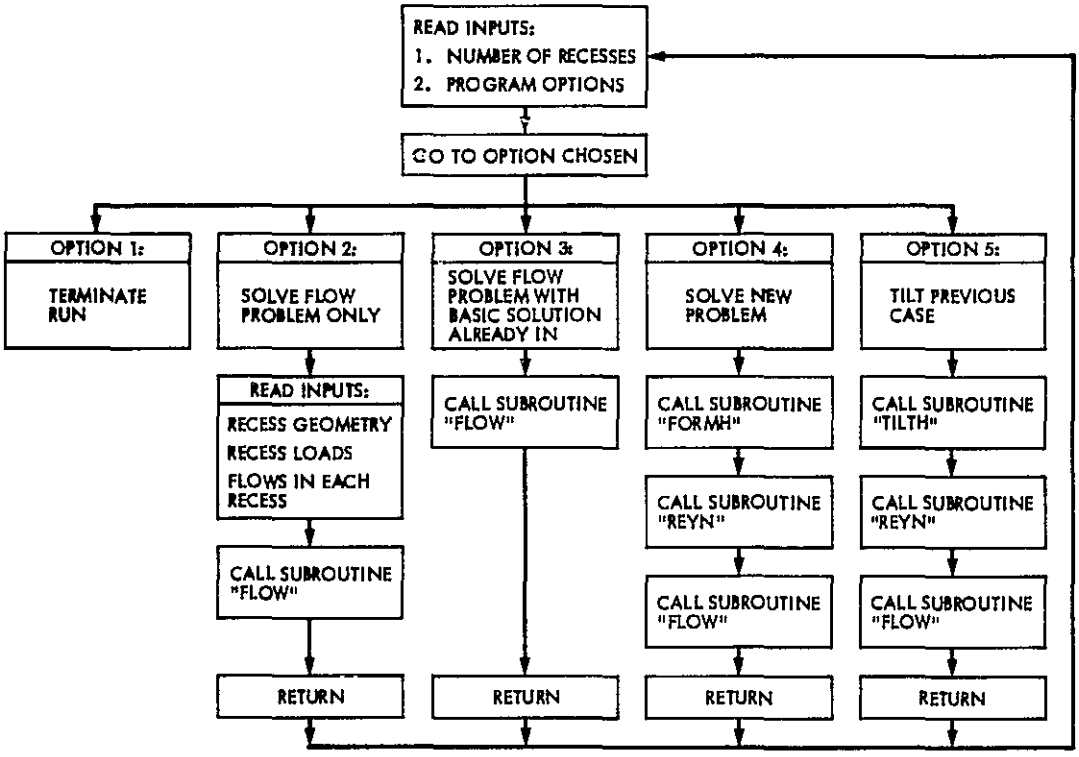


Fig. A-1. Main program

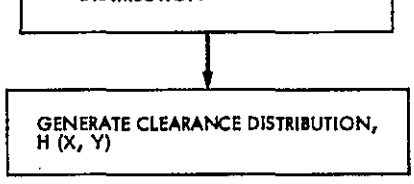


Fig. A-2. Subroutine "FORM H"

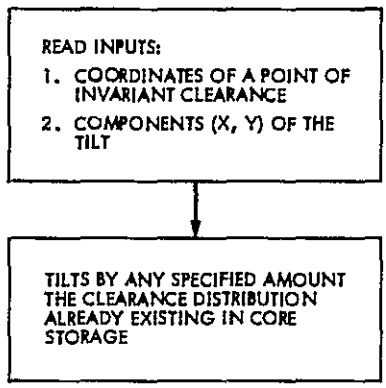


Fig. A-3. Subroutine "TILTH"

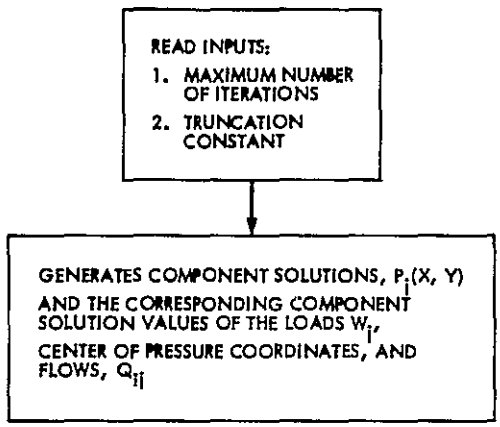


Fig. A-4. Subroutine "REYN"

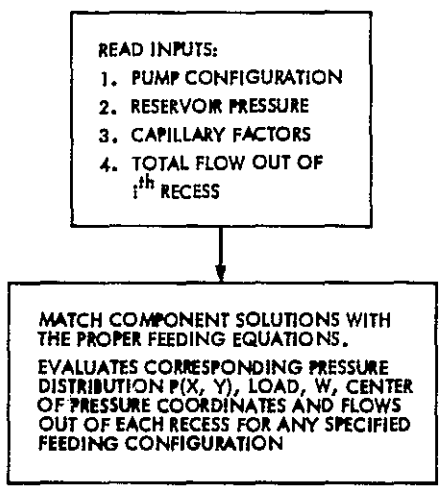


Fig. A-5. Subroutine "FLOW"

Availability of the DSN Telemetry Data System and Its Major Elements, Including the TWM Assemblies

R. Stevens
Telecommunications and Data Acquisition Office

The DSN Discrepancy Report System records all outages of DSN Data Systems that occur during mission support operations. The recorded outages of the Telemetry Data System for 1981 through 1983 were tabulated and analyzed. The analysis developed availability characteristics of the Telemetry Data System and several of its major elements, including, in particular, the Traveling Wave Maser (TWM) Assemblies. The principal objective of the work was to provide a comparison of availability characteristics of the TWM Assemblies with those of other DSN subsystems and assemblies.

For the three-year period, the Availability of the Telemetry Data System is 99.03%, and its Mean Time to Restore Service is 0.9 hours; the Availability of the TWM Assemblies is 99.83%, and their Mean Time to Restore Service is 2.5 hours.

I. Introduction and Summary

This article presents Mean Time Between Failures (MTBF), Mean Time to Restore Service (MTTRS), and Availability data of the Telemetry Data System and certain of its critical elements, including the Traveling Wave Maser (TWM) Assemblies. The results are based on analysis of Telemetry Data System outages recorded in the DSN Discrepancy Reports for 1981, 1982, and 1983.

The purpose of the article is to support a comprehensive study of TWM Assembly reliability. That study is reported in Ref. 1.

Table 1 contains the most significant results of the analysis.

More than 100,000 station support hours are represented in the data of the table. The Antenna + Pointing Subsystems and the TWM Assembly are the two leading equipment contributors to Telemetry Data System outage time.

The most significant problem with the TWMs is their time-to-restore-service characteristics. Table 1 shows a relatively very long MTTRS (2.5 hours) for the TWMs. The results of a detailed analysis of the 1982 and 1983 DR Monthly Reports are more descriptive of the problem. That sample covers approximately 60K station hours. In that sample, there is a

total of 22 times during the two years when telemetry was unavailable for most or all of a pass. Ten of those times were caused by TWM unavailability. The TWM Assembly is currently by far the worst offender in causing long telemetry outages.

Averaged over the last three years, the performance of the Telemetry Data System meets the future MK IVA Availability requirements; it does not meet the future MK IVA service restoration requirements.

The time trajectory of the Telemetry Data System Availability and the related outage hours for 10,000 station support hours are shown in Fig. 1. Station support hours represented by the points of Fig. 1 are: 1981 – 50K hours, 1982 – 30K hours, 1983 – 30K hours. As an example, the planned Voyager Uranus encounter operations require approximately 10,000 individual station support hours. Thus, if the 1986 MK IVA performance is no better than the 1981 - 83 MK III performance, about 100 hours of Uranus encounter telemetry will be lost or significantly degraded by DSN Telemetry Data System outages.

II. Discussion of Analysis

A. Data Source

The raw data were taken from the monthly DSN Discrepancy Report (DR) System. During spacecraft mission support, any interruption in a DSN Data System service is documented by a DR. The DR identifies the faulty DSN Data System and its faulty subsystem and assembly. The DR also gives the duration of service outage.

Data were taken from available monthly DRs for 1982-83 and from the DR data archives for 1981. The 1981 through 1983 period includes support of the Voyager 2 Saturn encounter and subsequent S/C testing and preparations for the 1986 Uranus encounter. The majority of other support was of S/C in extended mission cruise. Overall, the period must be characterized as much less demanding on DSN mission support than what lies ahead after mid-1985.

The prime objective of this study is to compare operability characteristics of TWMs with operability characteristics of other subsystems or assemblies. That objective was met by tabulating and analyzing the outages of the Telemetry Data System. The Telemetry Data System includes the TWM Assembly and other major subsystems and assemblies.

B. Results of Analysis

The summary tabulation of the DR data and its analysis are in Table 2. Footnotes below the table explain the entries.

Failures in the five Telemetry Data System elements that were analyzed, items 2-6 of Table 2, caused 80% of the total telemetry outage time during the three-year period (excluding that caused by RFI). The remaining 20% of the outage time was caused by failures in several other subsystems/assemblies, the effects of adverse weather and procedural errors. No detail analysis of their individual contributions was done.[1]

Figure 2 is a plot of the Table 2 MTBF and MTTRS data for the Telemetry Data System and the five of its system elements. Data for the individual years and for the entire period are plotted. Contours of constant availability are shown. MTBF/MTTRS time-trajectories that move to higher availabilities are generally good, and, conversely, those that move to lower ones are generally not good.

C. Discussion of TWM MTBF and MTTRS

The TWMs occupy a lonely position on Fig. 1. They have a relatively high MTBF, but also a high MTTRS. Actually, the 2.5 hour MTTRS shown is the average of a number of brief outages when backup systems restore service quickly and a number of very long outages when backup is not available and the failed maser has to be fixed.

That characteristic is further illustrated in Fig. 3 by plots of the percentage of telemetry function outage events (ordinate) for which service was restored in less than the indicated time (abscissa). Figure 2 shows the character of the Telemetry Data System and three of its major elements: the Antenna + Pointing Subsystems, the TWM Assembly, and the Telemetry Subsystem. The relatively very high percentage of long outages of the TWMs is evident.

Incidentally, Fig. 2 shows that the median time to restore Telemetry Data System service is approximately 10 minutes. It also shows that 10% of the outages are of 2 hours or more duration.

D. Outages from RFI

Telemetry outages from RFI were considerable, but they did not impact high priority mission support. During the three-year period, RFI caused 251 Telemetry System outages totaling 141 hours – 8 hours in 1981, 23 hours in 1982, 110 hours in 1983 (cf. Table 2).

Detailed analysis of the 1982 and 1983 DRs showed the following: Only missions in extended phase were affected

[1]The weather effects were analyzed for 1982 and 1983. The outages were: Total – 25.2 hrs; from wind – 19.4 hrs; from rain – 3.0 hrs (all X-band S/N degradation); from snow and ice – 2.8 hrs.

(Pioneers 10, 11, 12); there were no X-band outages; essentially all RFI was from external sources (only 17 minutes were identified as station internal); approximately 80% of the events were predicted in advance.

III. Availability, MTBF, and MTTRS: MK IVA Future Requirements vs Current Performance

A. Requirements

The DSN has two basic mission support regimes: mission critical phase and mission cruise phase. An example of the former is Voyager 2 Uranus encounter; an example of the latter is Pioneer 10 extra-solar-system cruise.

The Availability and MTTRS requirements for the MK IVA Telemetry Data System recognize the two support regimes.[2] The basic requirements are:

Spacecraft support regime	Tele. funct. avail.	Tele. funct. restore time
Critical activity	99.0%	30 mins max; 15 mins mean
Normal (cruise) activity	96.0%	30 mins max; 15 mins mean

The requirements recognize that, realistically, performance may be impacted by critical system elements that sometimes cannot be restored to service in less than 30 minutes. Those are elements having inherently long repair times and lacking ready redundancy. The recognized inclusion of any such elements requires justification on a cost vs performance basis during the system design phase.

B. Discussion of Requirements

1. Cruise support. The Availability requirement of 96% is probably as high as can be justified. Actually, the dominant factor in the support availability is the limited station tracking time that can be allocated to cruise missions. For the lower priority missions, that results in an average network support availability vs their SIRD requirements of about 50-80%. Providing a Data System Availability greater than the 96% specified would scarcely be felt by those users. Also, informal studies by the TDA Mission Support Office have shown that, usually, the economical approach to providing a significant percentage increase in availability is via more station facilities, rather than via increased system functional availability.

Regarding restoration time – except for emergencies and scheduled special events support – the acquisition of cruise mission data is basically not time critical. Probably the requirement shown is more rigorous than need be, although it is a commendable goal for a smoothly operating network.

2. Critical support. The Availability requirement of 99% appears realistic. It has a historical foundation with deep space missions, because it appears that that is about what has been provided in the recent past. The figure applies to a single stream, and, in an arrayed configuration of N independent antennas, the Availability implicitly is approximately 0.99^N. For example, the availability of the full capability of a three-antenna array would be approximately 97%. Because the array configurations are applied to the most critical support activities, this suggests that a more ambitious Availability goal may be appropriate in the future.

The Time-To-Restore-Service requirement appears appropriate from a mission support perspective. However, it will be very difficult to meet. If it is met, and the present MTBF is maintained, the single antenna system Availability will be approximately 99.7%.

C. Discussion of Performance

The analysis did not differentiate between critical and cruise S/C support periods. During the time covered by the analysis, the majority of the support was for cruise phase missions.

The average Telemetry Data System Availability determined for the three-year period is 99.0%. That meets the stated MK IVA requirement for critical support.

The average MTTRS determined is 0.9 hours. The 1982-1983 data show a significant number of long outages: e.g., 10% of the outages are greater than 2 hours. These service restoration characteristics are far off of the MK IVA mark.

D. Discussion of the Findings

The leverage on improving the Availability and service restoration performance of the Telemetry Data System is with the Antenna + Pointing Subsystems and the TWM Assemblies. They cause 31% and 18%, respectively, of the total system outage time. The TWMs cause almost half of the system outages longer than 6 hours.

[2]Deep Space Network System Requirements, MK IVA Telemetry System (1984 through 1986), JPL document 832-16 (preliminary), March 15, 1982. (Internal)

If the MTTRS of the TWMs were improved to the average of the other Telemetry Data System elements examined (0.8 hrs), the System MTTRS would be reduced from 0.9 to 0.8 hours, and its Availability would be increased from 99.0% to 99.2%. Achieving that TWM MTTRS will require full redundancy and exceptionally vigilant maintenance and use procedures.

The current MTBF of the TWMs is very good, relative to other station equipment. It is believed to be practical to double it. Achieving that is especially important for critical event support when TWM redundancy is not available.

Finally, the large antennas and the ability to point them correctly are in a special category. An outage of the antenna causes an outage of all DSN Data Systems that are in use for mission support, not only Telemetry. The subsystems that provide the function are large and complex, have mechanical and structural elements that are time consuming to replace or repair, and full and effective redundancy is very expensive. Nonetheless, their contribution to system outage is prominent, and any improvement that could be made in their availability characteristics would produce a significant improvement in the Telemetry Data System Availability. That is probably also true for the Availability of the other DSN Data Systems.

Reference

1. Stevens, R., and C. P. Wiggins, A Study of DSN Traveling Wave Maser Reliability, *The Telecommunications and Data Acquisition Progress Report 42-78,* Jet Propulsion Laboratory, Pasadena, California, August 15, 1984.

Table 1. Analysis of Telemetry Data System outages

Syst/SS/assy	No. of outages	Outage hours	MTBF (hrs)	MTTRS (hrs)	Availability[a] %
Tele Data Syst	1182	1044	90	0.9	99.03
Antenna + Pointing SSs	354	322	300	0.9	99.70
TWM Assy	75	186	1420	2.5	99.83

[a]Availability = MTBF/(MTBF+MTTRS)

Table 2. Summary of telemetry outage events and resulting MTBFs, MTTRSs, and Availabilities[a]

	Syst/Subsyst/Assy	Time period	Number of outage events[b]	Total outage (hours)	Station hours	MTBF (hours)[c]	MTTRS (hours)[d]	Availability for telemetry (%)[e]
1.	Telemetry Data System (without RFI outages)	1981	500	414	47,700	95	0.83	99.13
		1982	362	294	30,100	83	0.81	99.03
		1983	320	336	29,000	91	1.05	98.86
		All Years	1182	1044	106,800	90	0.88	99.03
2.	Antenna + Pointing SSs	1981	138	107	"	346	0.78	99.78
		1982	129	109	"	233	0.84	99.64
		1983	87	106	"	334	1.22	99.64
		All Years	354	322	"	302	0.91	99.70
3.	TWM Assy	1981	29	65	"	1645	2.24	99.86
		1982	20	55	"	1500	2.75	99.82
		1983	26	66	"	1120	2.54	99.77
		All Years	75	186	"	1424	2.48	99.83
4.	Receiver SS	1981	80	77	"	596	0.96	99.84
		1982	52	29	"	578	0.56	99.90
		1983	58	59	"	501	1.02	99.80
		All Years	190	165	"	562	0.87	99.85
5.	Telemetry SS	1981	66	58	"	723	0.88	99.88
		1982	36	42	"	835	1.17	99.86
		1983	47	13	"	618	0.28	99.95
		All Years	149	113	"	717	0.76	99.89
6.	Facility SS	1981	32	20	"	1491	0.63	99.96
		1982	14	14	"	2210	1.00	99.95
		1983	18	9	"	1610	0.50	99.97
		All Years	64	43	"	1669	0.67	99.96
7a.	Telemetry Data System (with RFI outages)	1981	516	422	"	92	0.82	99.12
		1982	413	317	"	73	0.77	98.96
		1983	504	446	"	58	0.88	98.51
		All Years	1433	1185	"	75	0.83	98.91
7b.	RFI Outages	1981	16	8	"	2981	0.50	99.98
		1982	51	23	"	589	0.45	99.92
		1983	184	110	"	158	0.60	99.62
		All Years	251	141	"	425	0.56	99.87

[a]From DSN DRs for 1981–83. Data from 35 monthly DR reports are included in Table 2 – the September 1982 report was not available.

[b]Only DRs that caused a recorded outage of the Telemetry Data System are included in the tabulation. The outages are recorded to the closest minute, so outages less than 1/2 minute are uncounted. For example, during 1982–83, there was a total of 46 DRs against the TWM Assy that resulted in telemetry outages – those 46 were included in the count. Also, there was a total of 16 Telemetry System DRs against the TWM Assy that did not cause recorded outages – those 16 were not included in the count.

[c]The MTBFs shown are calculated as the number of station tracking hours divided by the number of telemetry outage events for the period considered. That approach implicitly assumes that station hours=telemetry hours, which is not quite true. For example, a small amount of VLBI for mission support does not provide simultaneous telemetry. That results in telemetry hours less than station hours. Conversely, some telemetry is dual channel, which results in telemetry hours greater than station hours. The assumption was used for simplicity; it probably doesn't bias the MTBFs by more than a few percent.

[d]The MTTRSs shown are calculated as the total outage time divided by the number of recorded outages for a period.

[e]The Telemetry Data System Availabilities are calculated as the MTBF divided by the sum of the MTBF plus the MTTRS.

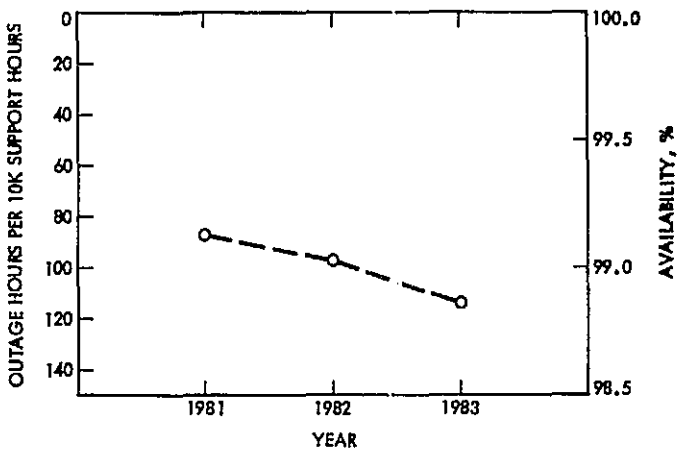


Fig. 1. Availability and outage hours per 10K support hours for DSN Telemetry Data System

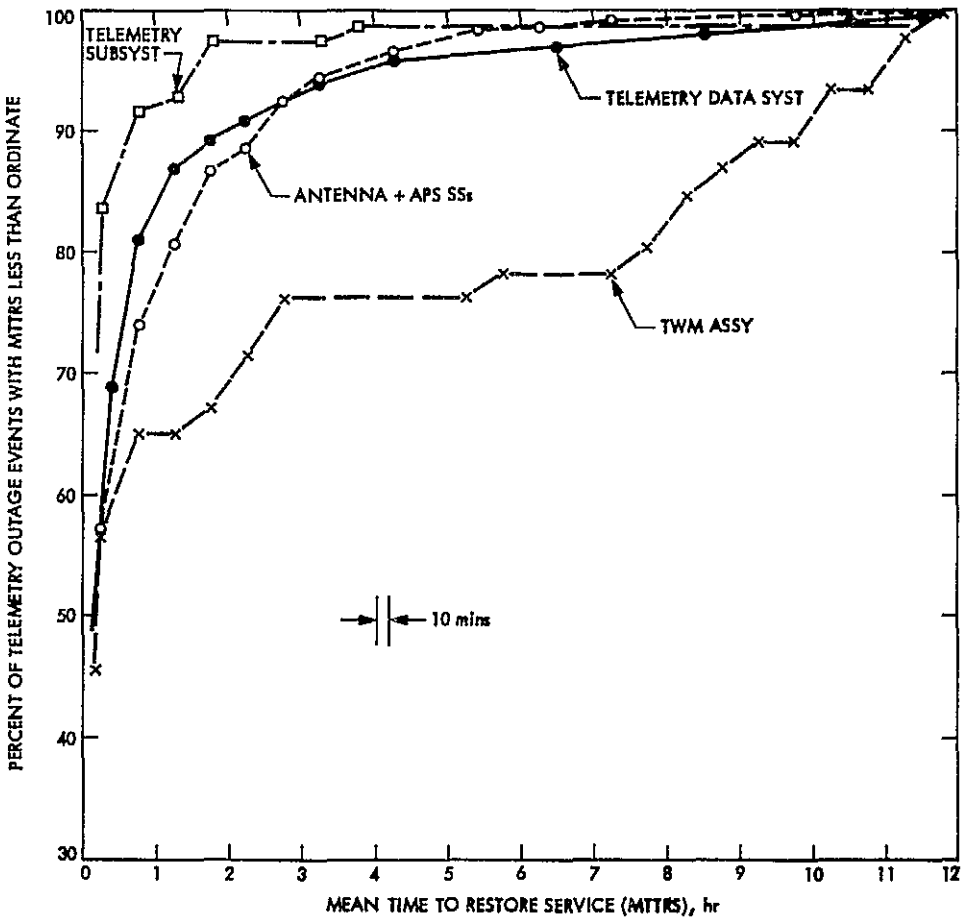


Fig. 2. Mean Time Between Failures, Mean Time to Restore Service, and Availability for DSN Telemetry Data System and its major elements

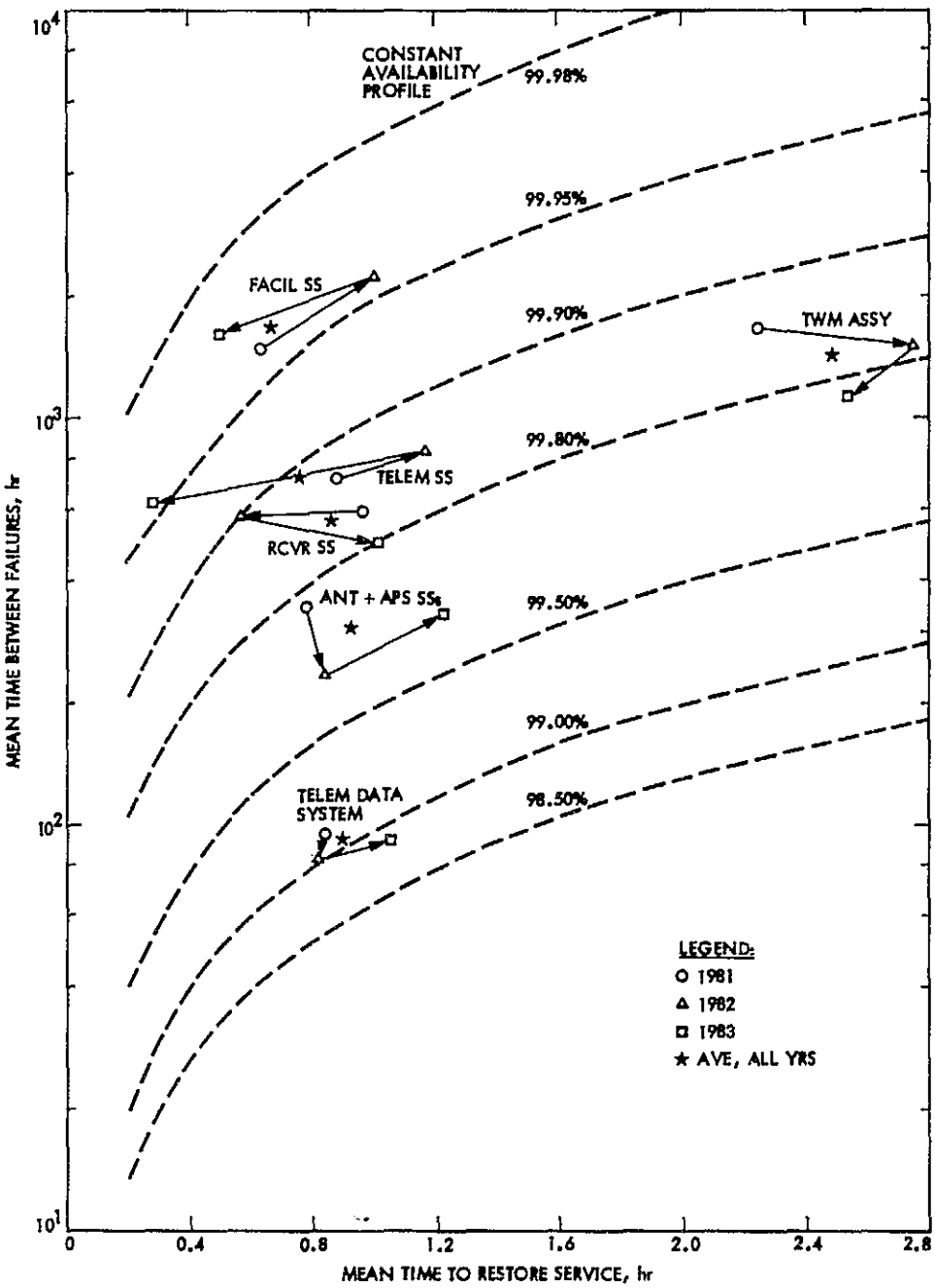


Fig. 3. Percent of telemetry system outage events with Mean Time to Restore Service less than a given time

D18

N84 32640

A Study of DSN Traveling Wave Maser System Reliability

R. Stevens
Telecommunications and Data Acquisition Office

C. P. Wiggins
Telecommunications Science and Engineering Division

This article reports on past and present reliability and availability characteristics of the DSN Traveling Wave Maser (TWM) Assemblies. For the years 1981 through 1983, the characteristics determined are: Mean Time Between Failures (MTBF) – 1200 hours; Mean Time to Restore Service (MTTRS) – 2.5 hours; and Availability – 99.83%. The TWM MTBF currently is very good relative to other DSN subsystems and assemblies; however, it has been significantly better in the past – in the late 1970s it was 3000 hours. The TWM MTTRS is currently about three times as long as the average of other DSN subsystems.

The dominant cause of TWM failures is contamination of the helium gas in the closed cycle refrigerators. Station configurations that do not provide TWM redundancy are subject to having reception outages for long periods of time.

A number of recommendations are made to improve the TWM Assembly availability characteristics for future mission support operations. Many of the recommendations result from a network-wide workshop of TWM experts that was recently conducted at the Canberra Complex.

I. Introduction and Summary

A. Introduction

The Traveling Wave Maser (TWM) Assemblies used on the large DSN antennas are truly remarkable devices. Their technical performance in the field is essentially equivalent to the best attainable in a laboratory environment.

The unique and critical attribute of the TWMs for deep space communications is the very small noise they introduce – the latest models add no more noise than the cosmic background (3K). That has enabled total system temperatures of 13 to 20K, increasing the reception capability of the DSN's large antennas by about a factor of two relative to that currently obtainable with any other amplifier approach. Other

important attributes of the TWMs are: 40-45 dB gain over the S- and X-band frequencies assigned for deep space reception, excellent phase and amplitude stability, large dynamic range (greater than 100 dB), and immunity from damage by any reasonable input overload.

However, achieving a high availability of the TWM Assemblies is a very demanding task. And high availability is essential, because a TWM outage results in loss of DSN reception capability until the function is restored. In recent years, the availability of the TWM Assemblies apparently has been degrading. This study was done to assess the situation.

B. Summary of Study Findings

The current Mean Time Between Failures (MTBF) of the TWM Assemblies is very good relative to other DSN station equipment. The MTBF has been, and can be, significantly better than it is now. The dominant cause of TWM Assembly failure is gas contamination in the helium refrigerator.

The Mean Time to Restore Service (MTTRS) of the TWM Assemblies is about three times longer than the average of other station subsystems. The TWMs are the dominant cause of very long outages of station reception capability. Unreliable single point of failure mechanisms must be reduced at the stations to improve the service restoration characteristics of the TWM Assemblies.

The Availability[1] of the TWM Assemblies is approximately 99.8%. The TWMs account for about 20% of the total outage time of the Telemetry Data System, and almost one-half of its outages that are longer than six hours. It is practical to improve this performance significantly in the future.

A DSN Maser Reliability Workshop was conducted. Maser experts from all of the complexes and from JPL participated. The Workshop report gives a prioritized set of recommendations to provide improved TWM Assembly maintenance and repair capability at the complexes. The report also gives recommendations for engineering development effort to provide improved TWM Assembly availability characteristics.

C. Summary of Recommendations

The recommendations from the study are, in order of priority:

1. Implement the First Priority items as recommended in the Workshop report.

The items pertain to use of best available field maintenance and repair techniques and to providing needed instrumentation and logistic support at the complexes. Implementation will lead to significantly improved TWM MTBF. It will also minimize the risk of failures from maintenance or repair processes that have unperceived faults. Such faults can result in concurrent failures of both prime and back-up TWM Assemblies.

2. Provide on-line back-up configurations at supporting stations for all periods of critical S/C activity support.

This will greatly improve the service restoration characteristics of those Front End Areas (FEAs) currently lacking redundant TWM Assembly configurations. Specifically, that includes redundant X-band TWMs for FEAs 12 and 42 by early 1985, and for FEAs 15, 45, and 65 by 1988.

3. Emphasize the care lavished on TWM Assemblies at FEAs that will be used for critical support but that will lack redundancy.

This will maximize the MTBF of the non-redundant TWMs, and thus will decrease the likelihood of long reception outages. Specifically, that applies to FEAs 15 and 45 at X-band for Voyager at Uranus. It also applies to FEAs 14, 43, and 63 at S-band for International Cometary Explorer (ICE) critical telemetry and for Galileo critical radio science – both require simultaneous use of the prime and the back-up TWMs.

4. Implement the distributed ground fault protection in the TWM AC input power circuits at FEAs 14 and 12.

Currently at the Goldstone Complex, AC input power ground fault protection circuitry is common to all individual TWM Assemblies. Thus, a single ground fault can cause all TWMs, primes and backups, to fail. Implementation of distributed protection will lower the risk of long outages at Goldstone.

5. Undertake the High Priority Engineering Development Tasks, and implement the Second Priority Maintenance and Sustaining items cited in the Workshop report.

This will further improve MTBF and MTTRS and will reduce the effort required at the complexes to maintain a very high level of TWM Assembly performance.

II. Study Background

The basic objectives of this study are: (1) to understand the availability characteristics, past and present, of the DSN TWM Assemblies, and (2) to identify actions that would improve their availability characteristics for future mission support. The study was started in June 1983. This report covers the complete study.

[1] Availability = MTBF/(MTBF+MTTRS)

III. Availability Characteristics of TWM Assemblies

The term, "Availability Characteristics," as used in this article, includes the MTBF; the service restoration properties, including the MTTRS; and the formal Availability that was previously defined.

A. MTBF History

Analysis of failure and DSN Discrepancy Report (DR) data gave the MTBF history shown in Table 1.

Approximately 1.8 million TWM operating hours are represented in the table.

B. Availability of the TWM Assembly vs Other Station Equipment

An analysis was done of the DSN Telemetry Data System outages for 1981 through 1983 (Ref. 1). The outage data were obtained from the DSN DR System reports. The Telemetry Data System includes the TWM Assembly as well as other major assemblies and subsystems. By tabulating the specific causes of the System outages, a comparison of the availability characteristics of the TWMs vs other major System elements was obtained.

The analysis of the Telemetry Data System outages gave the results shown in Table 2.

Approximately 107,000 hours of scheduled Telemetry Data System support are represented in the table.

The MTBF of the TWMs is relatively very good. However, the MTTRS of the TWMs is relatively very poor. Also, the data showed that the TWMs caused almost half of the Telemetry Data System outages that were longer than six hours.

C. Causes of TWM Assembly Failures

Analysis of the 1981-1983 DSN DRs to identify causes of TWM Assembly failures gave the results shown in Table 3.

The DR data show that refrigerator contamination is the dominant cause of TWM Assembly failure. That is supported by analysis of failure data samples collected from the stations during the past ten years: Of a total of 376 failures, 218, or 58%, were attributed to refrigerator contamination.

IV. Maser Reliability Workshop

A. Workshop Background

A Maser Reliability Workshop was conducted in December 1983, hosted by the Australian Complex. An agenda was prepared in advance by the JPL Cognizant Operations Engineer (COE) and the JPL Cognizant Development Engineer (CDE) with input and review by personnel from all of the complexes. A preliminary maser reliability improvement engineering program plan was prepared by the JPL Section 333 Microwave Electronics Group, including the CDE, and supplied to the Workshop participants for assessment. Maser experts from all complexes and maintenance facilities and the JPL COE and CDE participated in the Workshop.

B. Workshop Results

A formal report of the Workshop activities was issued. The report gives recommendations in three categories:

I. First Priority Implement FY84

II. Second Priority Implement FY85

III. High Priority Development Tasks

The Workshop attendees recommended that the Category I items be implemented prior to the Voyager encounter at Uranus. The Category I items abstracted from the Workshop report are:

1. Implement new and upgraded maser operation and maintenance procedures (developed at the Workshop);

2. Supply to each complex the necessary support equipment identified by the Workshop as required to perform Workshop recommended procedures;

3. Establish a special training course at each Complex Maintenance Facility for maser operation, maintenance, and repair personnel. The course should embody the latest and best engineering and operating knowledge of the DSN TWM Assemblies;

4. Implement the computer-based TWM-CCR monitoring and analysis system proposed in the preliminary reliability improvement engineering plan, previously noted, and include additional monitoring points as recommended by the Workshop; and

5. There are several engineering changes to the TWM Assemblies currently proposed. The following Engineering Change Orders (ECOs) are recommended by the Workshop as being top priority:

a. Provide new extended range flow meters for proper instrumentation of the Joule-Thompson valve return flow.

b. Provide for constant flow through the storage tank of the helium compressor, and add a pressure relief valve to the helium storage tank.

c. Provide a second stage oil injection line filter in the helium compressor.

There are thirteen Category II (Second Priority Implementation) items. They are considered by the Workshop attendees as being straightforward, of lower priority than Category I items, and they should be implemented as soon as practical after the Voyager Uranus encounter operations. The top four items on the list are:

1. Provide a mechanism to bypass in-service flow meters so that they may be calibrated or replaced without interrupting service.

2. Provide a mechanism to replace in-service adsorbers without affecting maser operation.

3. Refrigerator stage temperature monitoring sensors should be made available as soon as possible and added to all CCRs during routine repair cycling.

4. A service loop should be put in the compressor plumbing between the first stage oil separator and the orifice block to overcome the vibration failure of this line.

There are thirteen Category III (High Priority Development) items that the Workshop attendees recommended should be investigated as soon as possible. The top three items on the list are:

1. Compressor upgrade to a 5hp motor. The compressors currently use a 3hp motor that is overloaded by the new TWMs.

2. Develop field techniques for measurement of gaseous impurities in helium.

3. Develop a thorough understanding of the behavior of the compressor oil at increased temperature.

V. Goldstone Power Distribution Study

A. TWM Assembly Outages Due to Power Failure

Analysis of the 1981 through 1983 DSN DR records for TWM outages due to AC power failure gave the results shown in Table 4.

Several of the outages at Goldstone were long. DSS 44 has been deactivated since the reported failures.

B. Study of Goldstone Power for TWMs

A consultant was engaged to assess the suitability of the power system that supplies the TWMs at Goldstone. The following is extracted from the consultant's report:

> Four of the (Goldstone) outages were caused by an accidental ground contact and eleven of the outages were caused by the undesired placement of ground fault protective equipment. The result was that a ground fault in any single compressor would cause loss of power to all other compressors served by the same "cryogenic" power panel. This condition has been recognized and ground fault protective equipment for the individual compressors was purchased and portions have been received. The study RECOMMENDED that installation of this equipment be given high priority.
>
> . . . Other methods of improving reliability of electric service were studied which included by-passing busses and panels to reduce exposure to equipment failure and the addition of "clean" or dedicated circuits between the power source and the cryogenic panels. Costs ranged from $3,250 to $68,000 for the possible incremental improvements to a basically reliable system. NO additions were RECOMMENDED.
>
> The study also examined the reliability of commercial power and the first level of redundancy which is the local station power. The advisability of installing a second level of redundancy consisting of emergency generation located adjacent to cryogenic power panels was also examined. Costs ranged from $85,000 at DSS 12 & 15 to $143,000 at DSS 14. It was determined that the multiplicity of local station generators and the good record of automatic transfer from commercial power to station generation during commercial power outages provide adequate reliability of power service to the maser compressors. NO RECOMMENDATIONS were made. The transfer process should be given periodic and realistic tests.

VI. TWM Assembly Redundancy and Spares

A. Planned Operational Configurations

The Mk IVA DSN will have TWMs configured as shown in Table 5.

The "2" TWM configurations provide redundancy for all conventional support operations. Some non-conventional

support operations will be discussed in the next section. In summary, Table 5 shows that if current plans are implemented, there will be TWM redundancy at all FEAs except 15 and 45 by 1985; by 1988 there will be no exceptions.

B. Consideration of Support Operations When Redundancy is Not Available

When a TWM Assembly fails, it is almost sure to be out of service for 12 hours or more. Therefore, if there is not on-line redundancy available, a long reception outage is inevitable.

Here are three cases where on-line TWM redundancy is not available:

1. The FEA has a single TWM. An example is FEAs 15 and 45 used for X-Band support of Voyager at Uranus. All FEAs at the complexes will be simultaneously in use for Voyager support, so that functional back-up will not exist.

2. The mission support requires use of both prime and back-up TWMs. An example is support of ICE for a year beginning in March 1985 with a one-day critical encounter in September. ICE requires use of both prime and back-up S-Band TWMs to enable combining of two carrier channels. The support will be provided by FEAs 14, 43, and 63.

Another example is support of polarization measurement of the Galileo S-Band signal. That measurement requires use of both prime and back-up TWMs. FEAs 14, 43, and 63 will be used. The measurements are critical to Jupiter environment radio science experiment support. They will be made for about one week in August of 1988, and intermittently on a predefined schedule until April of 1990.

3. A common failure node brings down both prime and back-up TWMs. Recent examples of this are faulty charcoal filter stock unknowingly used in maintenance of multiple TWMs, and extended power outage to the TWMs.

Without TWM redundancy, lowering the risk of having an extended reception outage requires increasing the TWM MTBF. Although the TWM MTBF is good, it is believed that application of the use, maintenance, and repair practices recommended by the recent TWM Workshop will provide significant improvement. Also, healthy TWMs have the best chance of surviving AC power interruptions.

Efforts to provide improvements should focus especially on preparations for critical S/C support events.

VII. Improved Instrumentation of TWM Refrigerator Operating Parameters

We believe that observation and analysis of the operating parameters of the TWM Closed Cycle Refrigerators (CCRs) can provide insight into CCR condition and can identify impending CCR failures before they occur.

A. Previous Work on Instrumentation

An early result of this study was the recommendation to develop and demonstrate a computer-based data collection and analysis system for the CCRs. The concept was endorsed by the TWM Workshop – it is Recommendation 4 of the Priority I list (see Section IV.B.) Work was started but has been temporarily set aside on the design of a prototype.

Meanwhile, attempts were made to analyze CCR parameter data previously collected by hand. The results could not validate the belief that reliable predictions of incipient failures could be obtained from the data. One problem with all of the past data available is that the critical measurement of the Joule-Thompson stage return flow is spoiled by a flowmeter that is saturated almost all of the time. That problem can be corrected for the future by the implementation of a planned ECO (refer to IVB, item 5a).

B. Current Work on Instrumentation

The TWM/CCRs at DSS 12 will be properly instrumented, and their operating parameter data will be systematically collected for a period of several months. The data will be manually collected and analyzed. Definition of the data collection regimen and the analyses of the data are under way.

If the effort is successful, as we believe it must be, the results will be used to guide the future development of a computer-based prototype system. Also, the results will provide techniques for manual data collection and analysis that can be used during the interim until the computer-based system is available to the network.

VIII. Concluding Observations

Reliable performance of the TWM Assemblies is essential, especially during support of critical S/C activities.

Both the MTBF and the MTTRS of the TWMs can be improved significantly by practical measures. The MTTRS probably needs the most attention, but when redundancy is lacking, improved MTBF is the only road to lowering the risk of long outage of reception capability.

Acknowledgment

We have appreciated the help of many people in collecting and analyzing information for this report. These people include: R. Clauss, G. Davis (consultant), W. Higa (consultant), E. Jackson, R. Latham, M. Loria, F. McCrea, J. McNeil, P. Parsons, S. Petty, C. Quach, D. Trowbridge, the maser experts from all three DSN complexes, and especially T. Reid and the Canberra complex personnel for hosting the Workshop.

Reference

1. Stevens, R., Availability of the DSN Telemetry Data System and Its Major Elements, Including the TWM Assemblies. *The Telecommunications and Data Acquisition Progress Report 42-78*, Jet Propulsion Laboratory, Pasadena, California, August 15, 1984.

Table 1. MTBF history

Time Period	Approximate MTBF (hrs)
Late 1960s	1300
Late 1970s	3000
Early 1980s	1200

Table 2. Analysis of Telemetry Data System outages

Syst/SS/Assy	Total Outage (hrs)	MTBF (hrs)	MTTRS (hrs)	Availability (%)
Tele Data Syst	1044	90	0.9	99.03
Antenna+APS SSs	322	300	0.9	99.70
TWM Assy	186	1420	2.5	99.83
Receiver SS	165	560	0.9	99.85
Telemetry SS	113	720	0.8	99.89
Facility SS	43	1670	0.7	99.96

Table 3. Causes of TWM Assembly failures

Area of Failure	No. of Failures	% of Total No.
Helium Refrigerator	120	64
Contamination	(109)	(58)
Other (drive unit, etc.)	(11)	(6)
Helium Compressor	14	7
Miscellaneous (pump, MWV, electr.)	30	16
AC Input Power	19	10
Unidentified (DR not definitive)	6	3
Total	189	100

Table 4. TWM outages due to AC power failure

Complex–Station	No. of TWM Outages
Goldstone–DSS 12	3
Goldstone–DSS 14	12
Canberra–DSS 44	4
Total (refer to Table 3)	19

Table 5. TWMS configurations

Complex–FEA	X-Band	S-Band	Comments
Gold–12(34m)	1(2)[a]	2	(2) X by 1985 if 1 SPC-10 spare instld as B/U.
Gold–14(64m)	2	2	
Gold–15(34mHEF)	1(2)	-	(2) planned by 1988. S-Band FET.
Gold–10(spares)	2(1)	4	(1) X if spare instld as on-line B/U at FEA-12.
Aust–42(34m)	1(2)	2	(2) X by 1985 if 1 SPC-40 spare instld as B/U.
Aust–43(64m)	2	2	
Aust–45(34mHEF)	1(2)	-	(2) planned by 1988. S-Band FET.
Aust–40(spares)	2(1)	3	(1) X if spare instld as on-line B/U at FEA-42.
Spain–61(34m)	2	2	Spare X has been instld as on-line B/U.
Spain–63(64m)	2	2	
Spain–65(34mHEF)	(2)	(-)	FEA on-line 1988. Gets 2 X TWMs and S FET.
Spain–60(spares)	1	2	

[a]Numbers in parentheses are changes to present configurations.

GCF Compatibility With Packets and Data Compression

E. C. Posner
Telecommunications and Data Acquisition Office

P. Merkey
California Institute of Technology, Graduate Student

Some missions using packets and/or data compression may want an undetected GCF block error rate of 10^{-6}. Here we show that the present GCF meets this requirement.

I. Introduction

The DSN GCF High-Speed Data Lines and Wideband Data Lines use blocks of length 4800 bits, 22 of which are parity bits generated by the NASCOM polynomial (Ref. 1):

$$G(X) = X^{22} + X^{20} + X^{14} + X^{13} + X^{12} + X^{11} + X^8 + X^7 + X^5 + X^3 + X + 1 \quad (1)$$

or, in factored form,

$$G(X) = (X+1)^2 (X^{10} + X^3 + 1)(X^{10} + X^3 + X^2 + X + 1)$$

A *codeword* or code block in this code is a block of 4800 bits of 0 or 1 which is identified with a polynomial in X with 0 or 1 coefficients and modulo-2 addition. A codeword then is polynomial of degree at most 4799 divisible by $G(X)$.

If an error is detected by virtue of the fact that the received block, regarded as a polynomial, is not exactly divisible by $G(X)$, an error is detected, and a retransmission *may* be requested depending on mission requirements (Ref. 2). Measurements conducted for the TDA Engineering Office show that at most one block in 200 contains at least one bit error, so, without retransmissions, a throughput of 99.5% (or more) is obtained. The actual probability that a block contains an error will of course be slightly greater than this because of undetected errors.

The undetected error rate is the probability that a block passes the divisibility test but nonetheless contains an error. For packet telemetry systems (and all the more for packet telecommand) as well as for missions with image or other data compression, it may be necessary to keep this undetected block error probability below 10^{-6} (Ref. 3). Is the block error probability below 10^{-6}? It is hard to measure directly with current instrumentation, but this article shows that the requirement is met anyway.

II. Code Structure

Reference 4, plus a little calculation, shows that a length-4800 binary code with 22 check bits is at most single-error-correcting. Since $X + 1$ is a divisor of $G(X)$, the weights of the codewords are all even. Hence our code has minimum distance 2 or 4.

Actually, the minimum distance is only 2. This is because of the following argument. From Ref. 5, define $M^{(1)}(X) = X^{10} + X^3 + 1$, a primitive polynomial, with root, say, a. Then

$M^{(3)}(X)$, defined as the irreducible polynomial with root a^3, also has degree 10 and is the second degree -10 factor in $G(X)$ of Eq. (1). Thus:

$$G(X) = (X+1)^2 M^{(1)}(X) M^{(3)}(X)$$

Here $M^{(1)}(X)$ divides $X^{1023} + 1$ and no smaller binomial, by primitivity. The theory of equations shows that $M^{(3)}(X)$ divides $X^{341} + 1$ and no smaller (because $1023 = 3 \times 341$). So $H(X) = (X+1) M^{(1)}(X) M^{(3)}(X)$ also divides $X^{1023} + 1$ and no smaller binomial. Thus, $G(X) = (X+1) H(X)$ divides $(X^{1023} + 1)^2 = X^{2046} + 1$. But $X^{2046} + 1$ is a codeword, being a multiple of $G(X)$. Hence there are codewords of weight 2, and the code is of distance 2, not 4.

We note in passing that the code generated by $M^{(1)}(X) \cdot M^{(3)}(X)$ alone, of degree 20 (20 check bits) generates a distance-5 BCH code of length 1023 (Ref. 6). The NASCOM code however uses $(X+1)^2$ times this, of degree 22, but, more importantly, out to length 4800. How many codewords of weight 2 are there in the NASCOM code? We need to know this to estimate error probabilities.

If $C(X)$ is a codeword of weight 2, then

$$C(X) = (X^i + 1) X^j$$

for some non-negative integers with $i + j$ at most 4799. Now $G(X)$ divides $C(X)$ since $C(X)$ is a codeword, so $G(X)$ divides $X^i + 1$. All the more, $M^{(1)}(X)$ divides $X^i + 1$, so i is a multiple of 1023. Since $(X+1)^2$ must divide $X^i + 1$, i must be even. Since $i + j$ is at most 4800, $i = 2046$ or 4092.

If $i = 2046$, j can range from 0 to $4799 - 2046 = 2753$. There are 2754 codewords $X^{i+j} + X^j$ with $i = 2046$. If $i = 4092$, j can range from 0 to $4799 - 4092 = 707$, so there are 708 more codewords of weight 2. Altogether, the code has $2754 + 708 = 3462$ codewords of weight 2.

III. Independent Errors

First suppose bit errors in a block occur independently. This is not necessarily the case. But, if it were, the input bit error probability is derivable from the fact that at most one block in 200 contains at least one error. If p is the bit error probability, it will be small. So the *block* error probability r is about $4800p$, and

$$4800p = \frac{1}{200}$$

$$p = 0.96 \times 10^{-6}$$

What then is the undetected block error probability of this code? We will upper-bound it as the probability of an undetected double error plus the probability of *all* quadruple (or higher) errors. We can ignore triple errors because they are of odd weight, hence detectable. The probability of a *particular* double error is

$$b = p^2 (1-p)^{4798}$$

$$b = 9.174 \times 10^{-13}$$

There are 3462 double errors which are codewords, so the probability of an undetected double error is 3462 times b, or 3.176×10^{-9}.

We can upper-bound the probability of quadruple or higher errors in the independent-error case by the probability of quadruple errors alone, because p is so small. This can be made quantitative using the "tail estimate" for the binomial distribution (Ref. 5, App. [A.5], p. 467), but we omit it. The probability of quadruple error is $(4800 \times 4799 \times 4798 \times 4797)/24$ times $p^4 (1-p)^{4796} = 1.868 \times 10^{-11}$. Adding the previously derived probability of undetected double error, we find that an upper bound to the undetected block error probability when using the NASCOM 22-bit polynomial with independent errors is

$$3.195 \times 10^{-9}$$

This more than meets the undetected block error probability requirements.

IV. Arbitrary Error Structure

Now suppose we know nothing of the error patterns, just that one block in 200 contains at least one detected error. This is almost the same as one block in 200 containing an error, detected or not, and it is this that we shall actually assume. Let us even assume, as the worst case, that every block with an error contains at least two errors. In fact, a little thought shows that the worst case is when *all* the errors are double errors. This is essentially because higher error patterns are so numerous that they tend to distribute themselves randomly with respect to the code. Thus, approximately $2^{-22} = 2.5 \times 10^{-7}$ of the higher error patterns can be expected to be codewords. What is the exact fraction for double errors?

There are $(4800 \times 4799)/2 = 1.152 \times 10^7$ error patterns of weight 2, but, as we have seen in Sec. II, only 3462 codewords of weight 2. If we assume, as we do, that all error patterns of

a given weight, in particular of weight 2, are equally likely to occur, the probability that a double error pattern is a codeword is

$$3462/1.152 \times 10^{7} = 3.006 \times 10^{-4}$$

This is almost 1261 times as large as the 2^{-22} probability we would get if double error patterns distributed themselves randomly with respect to the entire code.

The probability of undetected error in these strange circumstances can be found as follows. We start with 1/200, the probability that there is at least one (and so, by our assumptions, exactly two) errors in the block. We multiply this by the probability that the error pattern is a codeword and, hence, is undetected. Thus, the undetected block error probability is

$$\frac{1}{200} \times 3.006 \times 10^{-4} = 1.503 \times 10^{-6}$$

This slightly exceeds the 10^{-6} undetected block error probability requirement. However, the assumptions under which we derived this high error probability are so extreme that we can consider that we do meet the 10^{-6} requirement. For example, if half the blocks in error contain a single error (which is, of course, detected) and half contain a double error, the above estimate drops by a factor of 2 to 7.5×10^{-7}, and the requirement is met.

We restate here that if we monitor the links to make sure we are getting the 99.5% throughput, then we will also be confirming the 10^{-6} or less undetected block-error probability as well. The GCF with error detection by the NASCOM 22-bit polynomial is compatible with the extremely low undetected GCF block error probabilities that some missions may want in the packet era.

Acknowledgments

We are indebted to the referee for sharpening the results of the last section.

References

1. NASA Goddard Space Flight Center, *NASCOM Error Detection*, NASA Goddard Space Flight Center Document GSFC-844-71-09, Sec. 3.1.6.2, (i), NASA Goddard Space Flight Center, Greenbelt, Md. (no date given), pg. 9.

2. Nightingale, D., "High-Speed System Design, Mark IIIA," Part B (pp. 103-105) of "VIII. GCF Development," *JPL Space Program Summary 37-66*, Vol. II, The Deep Space Network, Jet Propulsion Laboratory, Pasadena, Calif., Nov. 30, 1970, pp. 99-110.

3. Posner, E. C., and R. Stevens, "TDA Assessment of Recommendations for Space Data System Standards," *TDA Progress Report 42-77* (May 15, 1984), Jet Propulsion Laboratory, Pasadena, Calif., pp. 75-85.

4. Merkey, P., and E. C. Posner, "Optimum Cyclic Redundancy Codes for Noisier Channels," *TDA Progress Report 42-76* (Feb. 15, 1984), Jet Propulsion Laboratory, Pasadena, Calif., pp. 189-195.

5. Peterson, W. Wesley, and E. J. Weldon, Jr., *Error-Correcting Codes*, Second Edition, MIT Press, Cambridge, MA (1972), App. C, pp. 472-492.

6. Mac Williams, F. J., and N. J. A. Sloane, *The Theory of Error-Correcting Codes*, North Holland, Amsterdam (1977), Chap. 7, para. 6, p. 201.

D20

N84 32642

Earth Orientation Effects on Mobile VLBI Baselines

S. L. Allen
Tracking Systems and Applications Section

Improvements in data quality for the Mobile VLBI systems have placed higher accuracy requirements on earth orientation calibrations. Errors in these calibrations may give rise to systematic effects in the nonlength components of the baselines. In this work, various sources of earth orientation data were investigated for calibration of Mobile VLBI baselines. Significant differences in quality were found between the several available sources of UT1-UTC. The JPL Kalman-filtered space-technology data were found to be at least as good as any other and adequate to the needs of current Mobile VLBI systems and observing plans. For polar motion, the values from all services suffice. In addition, the effect of earth-orientation errors on the accuracy of differenced baselines (i.e., baselines between Mobile VLBI sites which were not simultaneously occupied) was investigated. This effect was found to be negligible for the current mobile systems and observing plan.

I. Introduction

The Mobile VLBI systems developed at JPL[1] have been producing high quality data since the beginning of 1980 (Ref. 1). Hardware upgrades made during this interval include conversion to wideband receivers, the Mark III data acquisition system, and dual (S- and X-band) frequency capability. These improvements have resulted in single measurements of baseline length with formal errors of less than 1 cm and repeatability of 1 to 2 cm.

To achieve similar accuracy in the measurement of nonlength components requires an additional calibration. This is due to the fact that the VLBI technique provides a very accurate measurement of the baseline within the reference frame of the quasi-stellar radio sources (Ref. 2). However, measurement of the various components of these baselines in an earth-fixed frame requires very precise knowledge of the orientation of the earth in space.

II. Earth Orientation Calibration of Baselines

By using a worldwide network of VLBI stations located in regions where local earth motion is not common, it would be possible to estimate the baselines and the earth orientation parameters from the same data set (Ref. 3). However, the Mobile VLBI baselines do not satisfy either of these criteria since they are measured between stations on the western U.S. coast. Furthermore, solving for earth orientation in an absolute sense requires placing some kind of constraint or model on the motion of the baselines. This risks the contamination of earth motions by imposing a possibly incorrect model. Thus, it is necessary to use an external source for the values of UT1-UTC and polar motion (UTPM).

[1]Allen, S. L., et al., "Current Mobile VLBI Data Base," Submitted to the NASA Crustal Dynamics Data Information System, Goddard Space Flight Center, Greenbelt, Maryland, May 1984.

A. Services Providing Earth Orientation Data

In order to calibrate the Mobile VLBI data it is desirable to have a self-consistent, high-precision UTPM data set which spans the entire history of the mobile experiments. The highest precision techniques for measuring UTPM are lunar laser ranging (LLR) and VLBI. Such measurements are presently carried out by several independent groups. Unfortunately, neither of these techniques has a complete data span overlapping with Mobile VLBI. It is necessary to use UTPM data from a service which combines the raw data from several techniques. Four services provide such combined data: the U.S. Naval Observatory, the International Time Bureau (BIH), Dr. Robert W. King (MIT Combination Solution; Ref. 4), and T. M. Eubanks (JPL Kalman-filtered space-technology data; Ref. 5). In the present work the latter three types have been applied to the Mobile VLBI data in order to evaluate their usefulness in removing systematic trends.

Since a combination solution for UT and PM is typically only as accurate as its input data, it is instructive to consider the expected accuracies of several different services of UT and PM. These figures are given in Table 1.

The values for the first three services are derived from past performance (Refs. 4, 5). The values of the last service are predicted from a covariance analysis (Ref. 6). These represent the accuracy of UTPM measurements which can be expected within the next two years.

The frequency of observation of the earth orientation services is also of importance to the calculation of differenced baselines. Astrometric UT and polar motion are derived from many daily observations of lower precision than LLR or VLBI. Although the frequency of these observations is greater than that of LLR or VLBI, the smoothing which is required removes much of the high-frequency signal. However, the relatively quick data reduction of the astrometric data makes them available for services such as the BIH Circular D. VLBI makes high-precision observations of UT and PM, but these are generally obtained at intervals of one week. LLR comes closest to providing daily, high-precision values of earth orientation, but even so, LLR cannot make observations for a week around new moon. To recover a value of UT and PM at the epoch of a Mobile VLBI experiement, it is necessary to interpolate or filter.

For polar motion the interpolation can be done simply and accurately because there are no large amplitude components with frequencies less than 2 weeks. This is not the case for UT1-UTC. Recent investigations indicate that even if UT were perfectly measured at 5-day intervals, the one-sigma uncertainy in interpolating midway between the measured points would be at least 0.5 ms (Ref. 7). As discussed in Section III, this has implications for the computation of differenced baselines. It requires appropriate deployment of mobile units and the use of base stations to minimize differencing errors.

B. Results of Earth-Orientation Calibrations

The most often measured Mobile VLBI baselines are those between JPL in Pasadena, Calif., Owens Valley Radio Observatory (OVRO), near Lone Pine, Calif., and DSS-13, at the NASA Goldstone complex (the JOG triangle). The baselines vary in length from 171 km to 336 km. There are also many measurements of the Owens Valley to Ft. Davis, Texas (HRAS) baseline. This baseline is 1508-km long; thus, its sensitivity to rotations is a factor of 5 greater than any JOG baseline.

Deficiencies in UTPM calibration show up as scatter in the nonlength components between one experiment solution and another. On these relatively short, regional baselines, the vertical direction is nearly coincident for each station. This allows nonlength error sources to be relatively easily separated into two perpendicular components: transverse and vertical.

The scatter of points in the vertical direction is caused by unmodeled day-to-day changes in the tropospheric path delay. The uncertainties in the troposphere calibration commonly cause 10-cm variations in the baseline vertical component. This effect is larger than the deficiencies in UTPM, and it masks any rotations along this direction. Fortunately, the transverse baseline components are not affected by any large systematic errors except UTPM; they provide a sensitive probe of UTPM.

The Mobile VLBI data for the JOG triangle are presented with three different calibrations in Figs. 1, 2 and 3. These are plots of the transverse component of the baseline vs experiment date. For each plot, the RMS deviation from a fitted line and the chi-squared per degree of freedom are given.

With the BIH Circular D values of UT1-UTC, the points which are closely spaced in time cluster quite well. These closely spaced points were obtained during the same mobile field exercise or "burst." However, from one cluster to another, shifts in the transverse component are evident. These shifts are especially noticeable on the OVRO/DSS-13 baseline. The other UT1-UTC series greatly reduces the shift from one mobile burst to another.

Figure 4 contains a similar comparison plot for the OVRO/HRAS baseline. The effect of deficiencies in UTPM is much greater on this baseline. The first two points in this baseline are known to contain systematic errors unrelated to UTPM. (These points are labelled H and I, and they are unimportant to this discussion.) Again, on this baseline the transverse

scatter is reduced by using the MIT and JPL values of UT1-UTC.

Figure 5 presents the OVRO/HRAS baseline in another form. The ellipses are the projection onto the transverse-vertical plane of the 3-D error ellipsoids obtained from each solution. Since this plane is perpendicular to the baseline, rotations of the baseline naturally show up as deflections on the plot. Of course, vertical deflections may also be caused by troposphere calibration errors. The inset vectors in each plot show the magnitude and direction of typical displacements caused by variations in UTPM. Point C in the JPL plot (Fig. 5[b]) has a very large uncertainty in UT1-UTC due to gaps in the Kalman filter's input data. The BIH Circular D calibration is not shown because its scatter exceeds the boundaries of this plot.

From Figs. 1, 2, and 3 it is clear that the MIT and JPL UT1-UTC values give less scatter than the BIH Circular D. The two series produce results which are comparable in transverse scatter. The JPL Kalman-filtered space data produce a slightly narrower vertical column of ellipses in Fig. 5(b); however, further data will be required to draw definite conclusions about the relative merits of the JPL and MIT combination solutions.

A similar investigation was carried out on these data using various sources of polar motion. In no case was the result of one polar motion calibration significantly better than any other. More data, preferably with longer baselines, will be required before Mobile VLBI data can be used to make a significant evaluation of polar motion data services. For present purposes, BIH Circular D suffices for the calibration of polar motion.

III. Differenced Baselines

A differenced baseline is defined as a baseline between two mobile VLBI sites which were not occupied simultaneously. Unlike the simultaneously measured "direct" baselines, the length of a differenced baseline can be affected by deficiencies in UTPM calibration. Doubt regarding the size of this error and the model necessary to best remove the error has prevented the report of any differenced baselines before this time. The capability of producing differenced baselines would greatly increase the data yield of Mobile VLBI experiments, and it would permit a rethinking of the deployment strategy for the Mobile VLBI units.

A. Error Sources

Typically, a differenced baseline would be formed from two separate solutions involving VLBI networks operated on different days. The mobile sites would not be the same for the two experiments, but at least one of the base stations would be identical. Differenced baselines would be trivial to compute given two assumptions: first, that no crustal motion occurs during the interval between occupations of the different sites; second, the earth orientation is known exactly for the occupations. The calculation would simply involve taking the difference of the vectors to each of the sites (hence the term "differenced baseline") and adding the covariance matrices to compute the errors.

The possibility of actual motion of the sites between occupations degrades the accuracy of a differenced baseline. Hence, we restrict calculation of differenced baselines to occupations in the same burst, with the maximum invervals between occupations of the endpoints of a differenced baseline being about 3 weeks. Even a baseline moving 6 cm per year would only move 5 mm in 3 weeks, and most measured baselines appear to be changing by less than 2 cm per year. Barring earthquakes, the uncertainty in the troposphere and ocean-loading corrections is larger than this motion; hence, earth motion will be ignored here. Errors in earth orientation for the two days would cause a misalignment of the two networks. More precisely, the solution on each day would be expressed in a different, slightly rotated coordinate system. This would cause a systematic error in a differenced baseline produced from these solutions.

B. Computation and Error Modeling

The geometry involved in the computation of a differenced baseline is depicted in Fig. 6. Figure 6(a) shows a scenario with one base station and a mobile station which occupies two sites on two different days. The uncertainty in the source of earth orientation on each day is represented as an error arc perpendicular to the baseline from base station to mobile site. The differenced baseline receives error contribution from each mobile site. Figure 6(b) depicts the geometry of a three-base-station experiment. All three base stations are active on both observing days, and the mobile moves from one site to another. In this case the base-station network can be used to solve for the difference in earth-orientation parameters between the two days, using the values obtained from an earth orientation service as *a priori* information. Only the baseline from each mobile site to the nearest base station is shown, though all are measured. This is done in order to emphasize that when earth orientation is not known exactly, the shortest baseline to the mobile units serves as the tightest constraint on their positions.

The results of a differenced baseline calculation should be independent of the choice of reference station in the solution. Figure 6(c) shows another three-base-station scenario where

both mobile sites are closest to the same base station. If the correlations between all stations were not included, then choosing different reference stations would produce different error estimates for the differenced baseline. This would be physically unreasonable. JPL's multiparameter least-squares fitting program MASTERFIT (Ref. 8) does handle all correlations properly; differenced baselines produced by MASTERFIT are independent of the choice of reference station.

The results of a differenced baseline calculation should also be independent of arbitrary choices used to define UT and PM. Consider the effect in Fig. 6(a) if day 1 of the burst were chosen to have UT and PM fixed with zero error. Since the baseline on day 2 is much longer than the baseline on day 1, its uncertainty is much larger. Although the post-fit error in the length of the differenced baseline is the same with either day chosen as reference, the error in the transverse component is much larger if day 1 is fixed. This is also a nonphysical result caused by the assumption of perfect knowledge of the coordinate system on a particular day.

The correct strategy for producing post-fit errors which are not dependent on arbitrary choices of a reference day is as follows. Assume that earth-orientation services provide values which are correct but lacking in the high-frequency components (i.e., periods of less than ~5 days). Thus, the data from consecutive experiments with common base stations can still be quite useful for tracking the day-to-day differences which were not resolved by the UTPM service. For each experiment set UT and PM to the values from the service, and assign *a priori* constraints equal to the errors quoted by the service. If possible, the day-to-day correlations of the UTPM values should also be used. This will force the post-fit values of earth orientation to agree in the mean with the service supplying UT and PM, but it will also allow the VLBI data to adjust the differences between one day and the next in order to remove misalignment of the networks. This use of both *a priori* and VLBI data correctly utilizes the available information, and it is simple to implement in MASTERFIT.

C. Estimated Errors

The errors which may be expected on differenced baselines are described below. These errors were calculated using the assumption that an individual baseline is measured with an uncertainty of 2 cm in its length and transverse components and 9 cm in the vertical component. The values of the earth-orientation services' uncertainties were set *a priori* to the values given in Table 1 and constrained using the VLBI network to determine the post-fit uncertainty in UT and PM. These uncertainties were propagated to the mobile stations along the shortest baseline from mobile to base station. Although this is not a full covariance analysis, it serves to set upper limits on the errors which will be encountered when actually reducing data.

The tables and figures give the estimated errors in the "rational" coordinate system (in which the principle axes are along the baseline length, vertical, and transverse directions) for various arrangements of base stations and mobile sites. Note that when there are three or more base stations, the VLBI data alone are sufficient in principle to determine earth orientation. However, due to the large vertical uncertainties on the baselines caused by troposphere, the earth-orientation value obtained from a service always serves as the tighter constraint in the vertical direction.

The additional error accrued quadratically to a differenced baseline due to UT and PM uncertainties is given in Tables 2 through 5. Table 2 details the effects when four base stations are operated on both days of the experiment, including the antenna in Ft. Davis, Texas (HRAS). This network allows for very precise determination of earth-orientation differences using the Mobile VLBI data. The additional errors are tabulated for several baselines in California using *a priori* errors for UT and PM from the various services described above. These errors are expressed in the rational coordinate system of components: the length, the component transverse to length, the vertical component, and the root sum of squares of all three components.

Table 3 gives the same figures when calculated without the presence of HRAS, and Tables 4 and 5 reduce the number of base stations to 2 and 1, respectively. It is clear that the presence of more and longer base-station baselines reduces the amount of error due to uncertainty in earth orientation. Plots of these errors are given in Figs. 7, 8, and 9 for a cross section of the data in the tables.

The interpretation of these tables and plots must be made with the understanding that this is the extra error due only to UT and PM uncertainty. These must be combined with the other errors inherent in the VLBI system to produce the full error. Figures 9 and 10 illustrate this point. Figure 9 shows the worst tabulated case of differenced baseline error. There is only one reference station (OVRO). Figure 10 shows the very same errors with a VLBI system noise of 2 cm in the length and transverse added. Only in the case of the astrometric UT and PM is there a significant degradation of accuracy for the differenced baseline.

Thus, differenced baselines within California do not suffer significant degradation due to uncertainties in earth orientation. However, the above analysis should not be taken to mean that only one base station is sufficient for the calculation of differenced baselines. It is very important that there be a base

station relatively near the mobile locations; e.g., a base station in Massachusetts with mobile units in California gives very large errors for a differenced baseline.

When more than one mobile unit is available and base stations are nearby, each mobile unit can move between each experiment. In the simple case of one base station and two mobile stations, differencing increases the number of baselines produced per burst of n site occupations from $3n$ to $(n + 1)(n + 2)/2$. When base stations are far away, two mobile units can work together to serve as base stations for each other. In this case each mobile unit remains at each site for 2 experiments while the other moves. Each extra base station serves to decrease the VLBI system noise by providing more data. More importantly, redundancy in the base stations helps prevent errors such as those which can be seen in Figs. 11 and 12.

The February 1983 burst provided an opportunity to calculate differenced baselines using real data. Experiments H83A and H83B shared the base stations OVRO, JPL, and Vandenberg. MV3 occupied Pearblossom and Pinyon Flat on the two days, respectively. These experiments allowed for a solution of a differenced baseline using MASTERFIT. Figure 11 shows the results of the solution when all three base stations were used and when only pairs of the base stations were used. All of the solutions agree, although the solution with all three base stations has slightly smaller errors. Figure 12 shows the same experiment when only one base station is used as a reference. The JPL-and-OVRO-only solutions still agree well with the full network solution; however, the Vandenberg-only solution differs by several sigma.

IV. Conclusion

For Mobile VLBI baselines with lengths less than 1000 km, it now is possible to remove all significant errors caused by uncertainties in earth orientation. This can be achieved using either the JPL or MIT combination solutions. The continued efforts to develop VLBI and laser systems for UTPM measurement should insure that this high accuracy is improved in the future. Thus, for regional work, earth orientation will no longer pose a problem, at least until other system errors are reduced below a level of 1 cm.

The occasionally measured longer baselines from the Mobile VLBI systems will continue to serve as indicators by which UTPM data sets can be evaluated. Calibration of the long baseline data will provide information on the systematic errors remaining in the UTPM combinations.

The computation of differenced baselines is a simple and natural way to take full advantage of the data produced by the Mobile VLBI project. Their inclusion approximately doubles the number of baselines produced under the current deployment plan. The additional error encountered on these differenced baselines will usually be small in comparison to the other errors in the VLBI error budget. The base-station baselines are also enhanced by the inclusion of more data into a single coherent solution.

References

1. Trask, D. W., et al., Mobile VLBI Surveying: Instrumentation, Operating Procedures, and Survey Results of ARIES, NOAA Tech. Report NOS 95 NGS 24, 1982, pp. 120-133.

2. Thomas, J. B., An Analysis of Long Baseline Radio Interferometry, Technical Report 32-1526; Vol. VII, 37-50; Vol. VIII, 29-38,; Vol. XVI, 47-64, Jet Propulsion Laboratory, Pasadena, Calif., 1972.

3. Bock, Y., and Zhu, S.-Y., On the Establishment and Maintenance of a Modern Conventional Terrestrial Reference System, NOAA Tech. Report NOS 95 NGS 24, 1982, pp. 36-45.

4. R. W. King, Improved Representation of Length-of-Day Variations from Combination of Different Types of Observations, abstract G22-07, Trans. Amer. Geophys. Union, Vol. 64, No. 18, May 1983.

5. Eubanks, T. M., Steppe, J. A., Spieth, M. A., The Accuracy of Radio Interferometric Measurements of the Earth Rotation, submitted to *Nature*.

6. Stolz, A., and Larden, D., "Accuracy Obtainable for Universal Time and Polar Motion During the EROLD Campaign," in *Scientific Applications of Lunar Laser Ranging*, pp. 201-216, ed. J. D. Mulholland, D. Reidel, Dordrecht, Holland, 1977.

7. Eubanks, T. M., Steppe, J. A. Dickey, J. O., Callahan, P. S., A Spectral Analysis of the Earth's Angular Momentum Budget, *Journal of Geophysical Research*, in press, 1984.

8. Fanselow, J. L., Observation Model and Parameter Partials for the JPL VLBI Parameter Estimation Software "Masterfit-V1.0," JPL Publication 83-39, December 1983.

Table 1. Typical errors for various sources of UT and PM

Earth Orientation Component	Source			
	Astrometric	Lunar Laser Ranging (LLR)	VLBI	3 Station "Super" LLR (Predicted)
σ UT1, ms	1.25	0.4	0.3	0.07
σ PMX, marcsec	7	7	3	2.0
σ PMY, marcsec	7	5	7	1.5

Table 2. Differenced baseline error: Additional error (in cm) for differenced baselines due to UT and PM OVRO, Mojave, Vandenberg, and HRAS as fixed base stations

Baseline	Source of Earth Orientation			
	Astrometric	LLR	VLBI	"Super" LLR
Monument Peak – Quincy	σ L = 0.12 σ T = 0.41 σ V = 1.38 RSS = 1.45	σ L = 0.10 σ T = 0.37 σ V = 1.06 RSS = 1.13	σ L = 0.11 σ T = 0.37 σ V = 1.22 RSS = 1.28	σ L = 0.06 σ T = 0.25 σ V = 0.33 RSS = 0.42
Pt. Reyes – Yuma	σ L = 0.16 σ T = 0.46 σ V = 1.59 RSS = 1.66	σ L = 0.15 σ T = 0.41 σ V = 1.21 RSS = 1.29	σ L = 0.16 σ T = 0.42 σ V = 1.22 RSS = 1.30	σ L = 0.10 σ T = 0.29 σ V = 0.35 RSS = 0.46
Sta. Paula – Pearblossom	σ L = 0.08 σ T = 0.14 σ V = 0.63 RSS = 0.65	σ L = 0.07 σ T = 0.13 σ V = 0.48 RSS = 0.50	σ L = 0.07 σ T = 0.14 σ V = 0.37 RSS = 0.40	σ L = 0.05 σ T = 0.09 σ V = 0.13 RSS = 0.16

Table 3. Differenced baseline error: Additional error (in cm) for differenced baselines due to UT and PM OVRO, Mojave, Vandenberg used as fixed base stations

Baseline	Source of Earth Orientation			
	Astrometric	LLR	VLBI	"Super" LLR
Monument Peak – Quincy	σL = 0.30 σT = 1.45 σV = 1.93 RSS = 2.43	σL = 0.24 σT = 1.08 σV = 1.20 RSS = 1.63	σL = 0.21 σT = 0.84 σV = 1.44 RSS = 1.68	σL = 0.08 σT = 0.36 σV = 0.33 RSS = 0.50
Pt. Reyes – Yuma	σL = 0.49 σT = 1.62 σV = 2.90 RSS = 3.36	σL = 0.41 σT = 1.24 σV = 1.51 RSS = 2.00	σL = 0.33 σT = 0.96 σV = 1.49 RSS = 1.80	σL = 0.14 σT = 0.41 σV = 0.36 RSS = 0.57
Sta. Paula – Pearblossom	σL = 0.27 σT = 0.51 σV = 1.01 RSS = 1.17	σL = 0.21 σT = 0.40 σV = 0.57 RSS = 0.72	σL = 0.16 σT = 0.31 σV = 0.42 RSS = 0.55	σL = 0.07 σT = 0.13 σV = 0.14 RSS = 0.20

Table 4. Differenced baseline error: Additional error (in cm) for differenced baselines due to UT and PM Mojave and Vandenberg used as fixed base stations

Baseline	Source of Earth Orientation			
	Astrometric	LLR	VLBI	"Super" LLR
Monument Peak – Quincy	σL = 1.13 σT = 2.57 σV = 2.26 RSS = 3.61	σL = 0.73 σT = 1.66 σV = 1.74 RSS = 2.51	σL = 0.52 σT = 1.21 σV = 2.02 RSS = 2.41	σL = 0.21 σT = 0.48 σV = 0.52 RSS = 0.73
Pt. Reyes – Yuma	σL = 0.91 σT = 2.23 σV = 2.61 RSS = 3.55	σL = 0.58 σT = 1.47 σV = 1.46 RSS = 2.15	σL = 0.41 σT = 1.06 σV = 1.79 RSS = 2.12	σL = 0.17 σT = 0.43 σV = 0.40 RSS = 0.61
Sta. Paula – Pearblossom	σL = 0.37 σT = 0.69 σV = 1.12 RSS = 1.37	σL = 0.24 σT = 0.47 σV = 0.57 RSS = 0.77	σL = 0.18 σT = 0.34 σV = 0.43 RSS = 0.57	σL = 0.07 σT = 0.14 σV = 0.14 RSS = 0.21

Table 5. Differenced baseline error: Additional error (in cm) for differenced baselines due to UT and PM OVRO 30 used as fixed base station

Baseline	Source of Earth Orientation			
	Astrometric	LLR	VLBI	"Super" LLR
Monument Peak - Quincy	σ L = 0.49 σ T = 3.81 σ V = 2.92 RSS = 4.82	σ L = 0.27 σ T = 1.96 σ V = 1.62 RSS = 2.56	σ L = 0.22 σ T = 1.33 σ V = 2.03 RSS = 2.44	σ L = 0.07 σ T = 0.51 σ V = 0.46 RSS = 0.69
Pt. Reyes - Yuma	σ L = 1.23 σ T = 4.14 σ V = 4.53 RSS = 6.26	σ L = 0.69 σ T = 2.17 σ V = 1.96 RSS = 3.00	σ L = 0.50 σ T = 1.46 σ V = 2.14 RSS = 2.64	σ L = 0.19 σ T = 0.56 σ V = 0.49 RSS = 0.77
Sta. Paula - Pearblossom	σ L = 2.57 σ T = 0.64 σ V = 1.58 RSS = 3.08	σ L = 1.33 σ T = 0.34 σ V = 1.20 RSS = 1.82	σ L = 0.91 σ T = 0.22 σ V = 1.33 RSS = 1.62	σ L = 0.34 σ T = 0.09 σ V = 0.35 RSS = 0.50

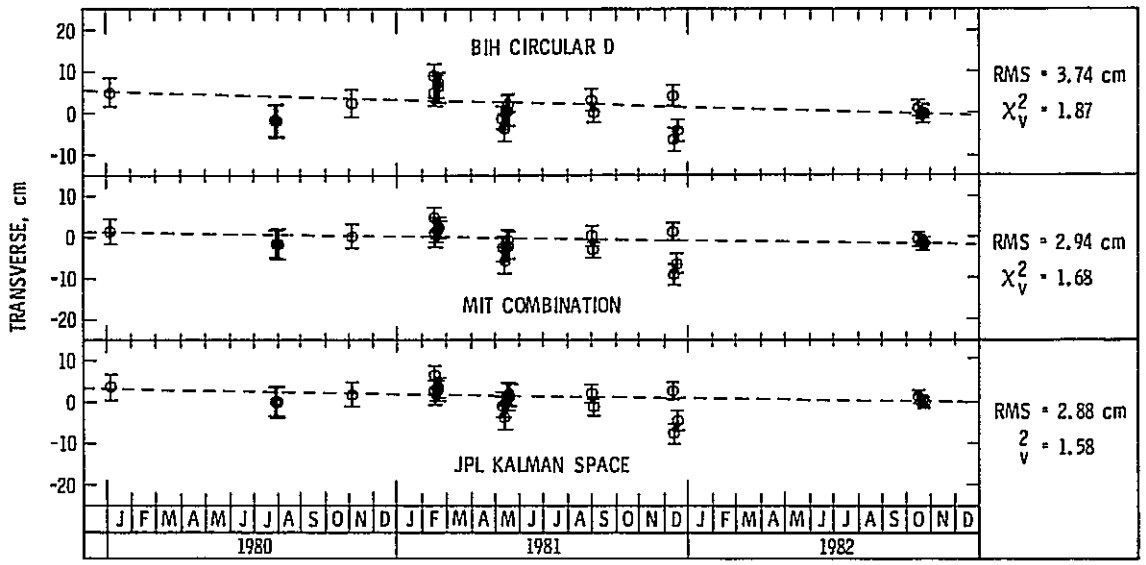


Fig. 1. OVRO 130/DSS 13 baseline, transverse component vs time

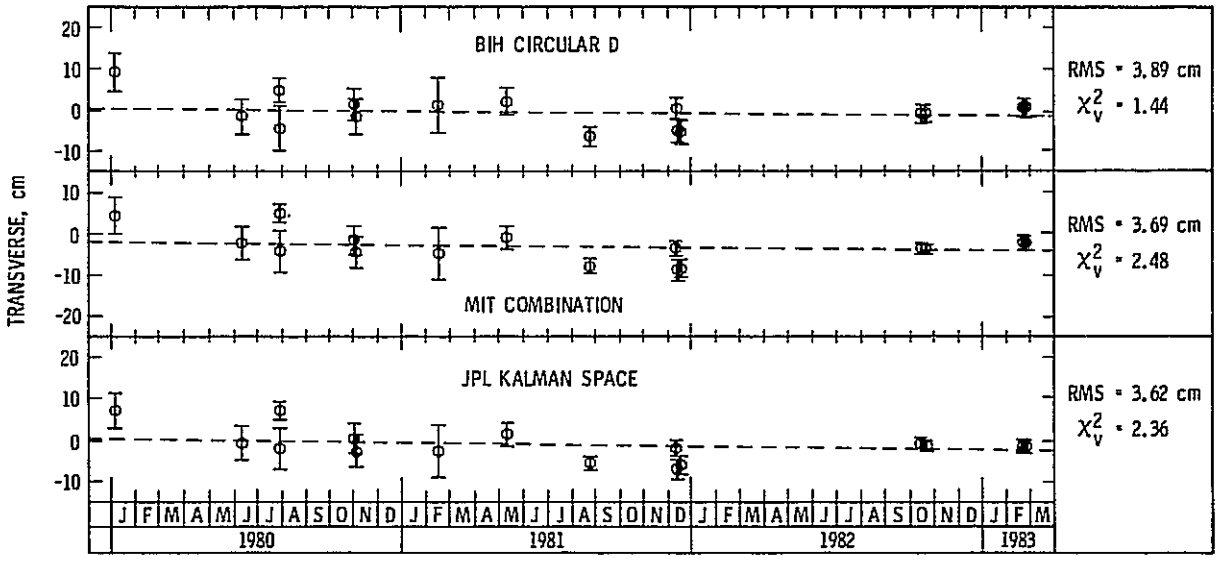


Fig. 2. OVRO 130/JPL baseline, transverse component vs time

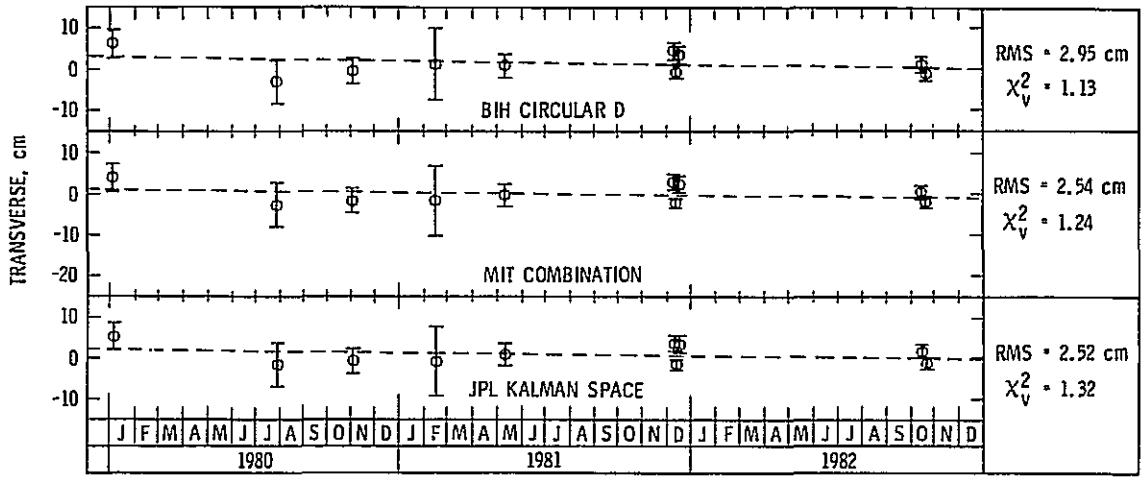


Fig. 3. DSS 13/JPL baseline, transverse component vs time

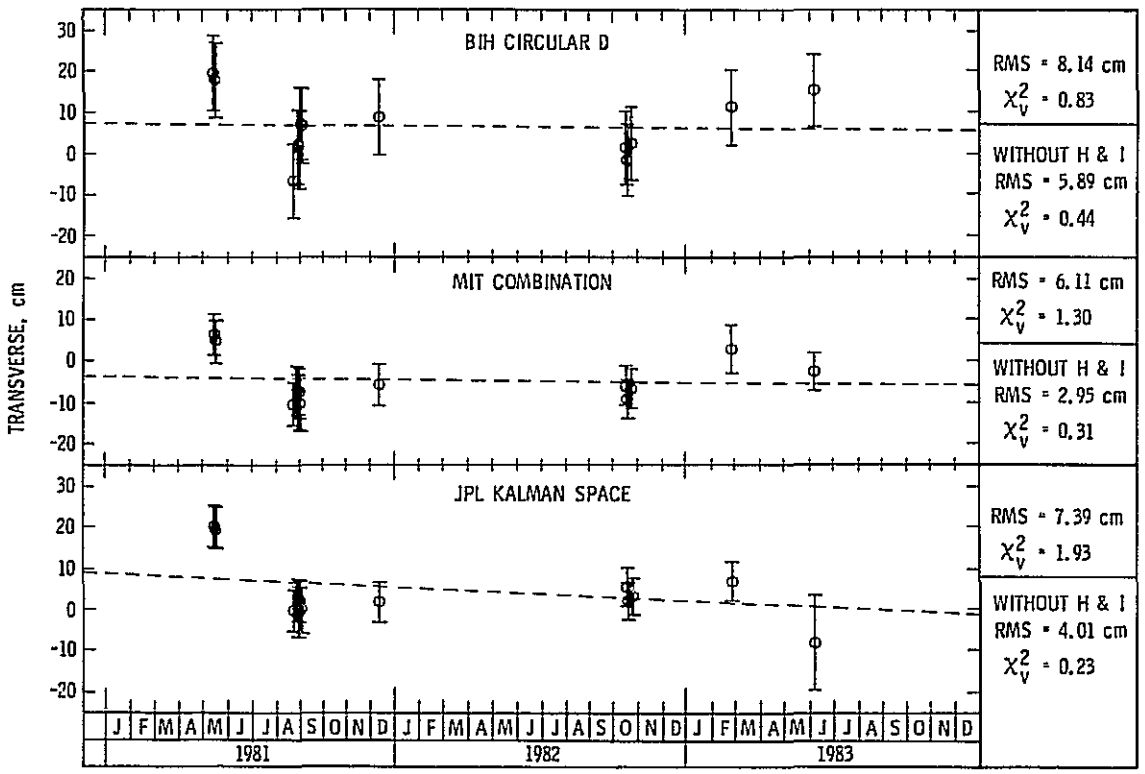


Fig. 4. OVRO 130/Ft. Davis (HRAS) baseline, transverse component vs time

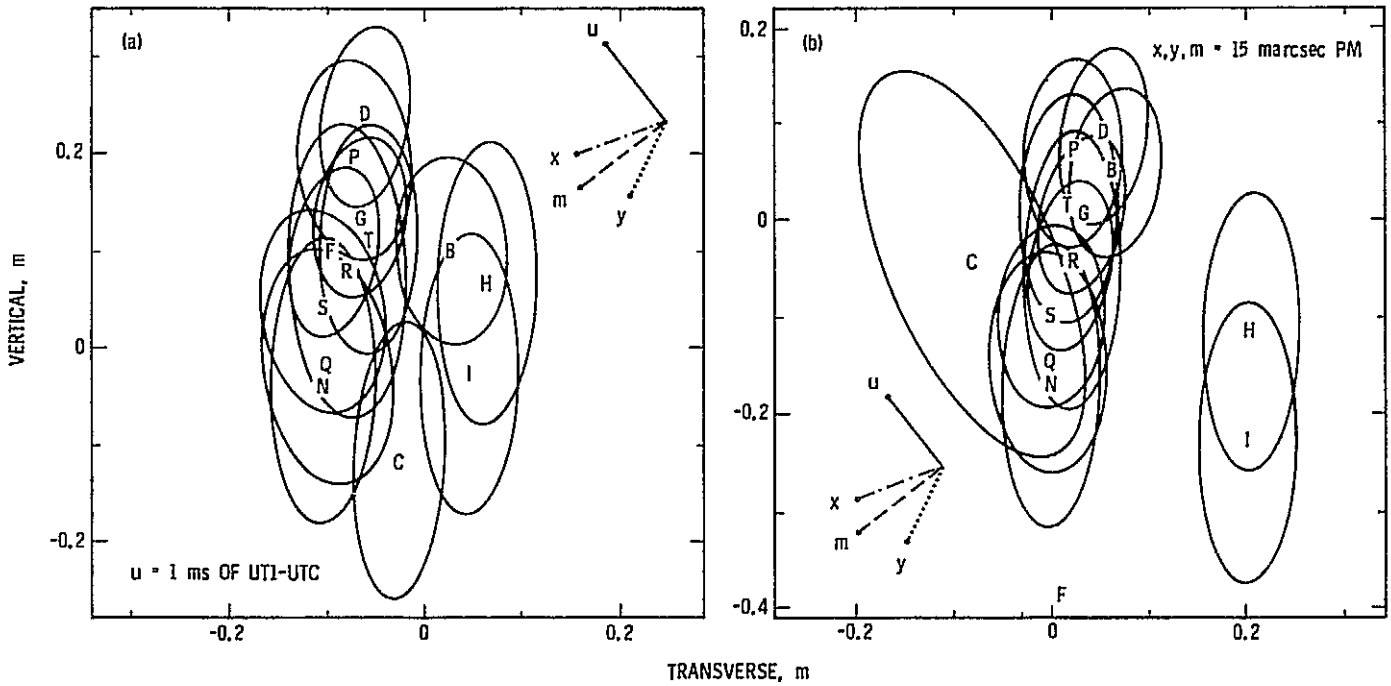


Fig. 5. OVRO 130/HRAS baseline, nonlength components. Inset vectors show directions and typical sizes of UT and PM shifts: (a) UT1−UTC: MIT combination solution; (b) UT1−UTC: JPL-Kalman filtered space

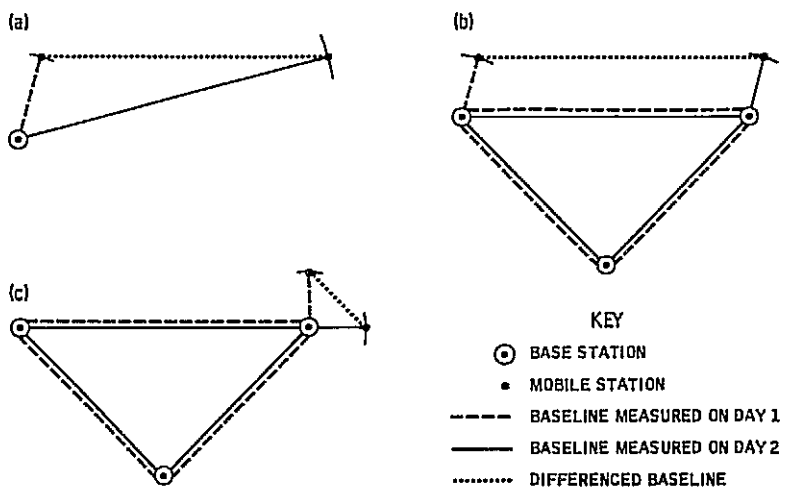


Fig. 6. Geometry of differenced baselines: (a) one base station, one mobile station; (b) three base stations, two mobile stations, long, differenced baseline; (c) three base stations, two mobile stations, short, differenced baseline

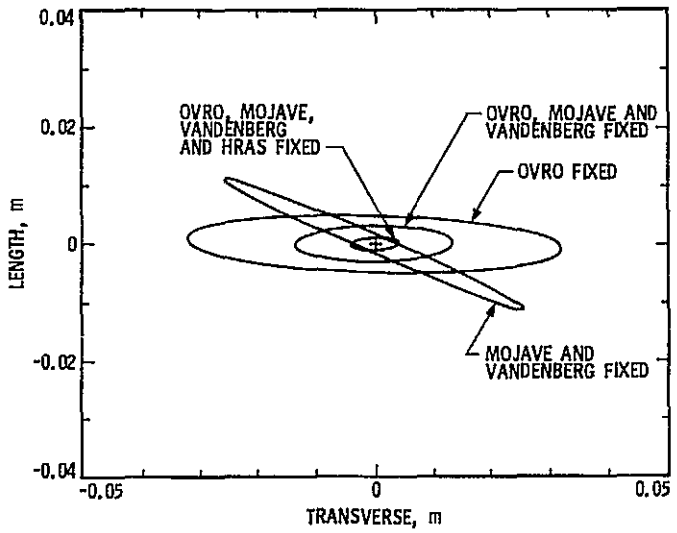


Fig. 7. Additional error from UT and PM, Monument Peak/Quincy baseline; errors with astrometric data and various base station groups

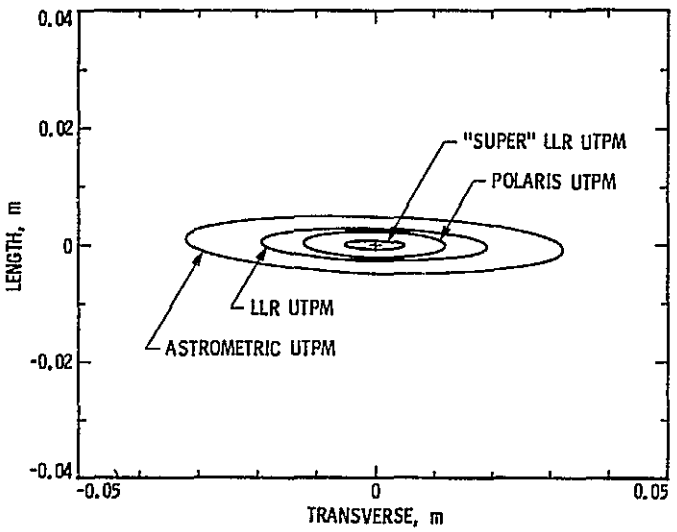


Fig. 9. Additional error from UT and PM, Monument Peak/Quincy baseline; errors with only OVRO 130 as base station and various UTPM services

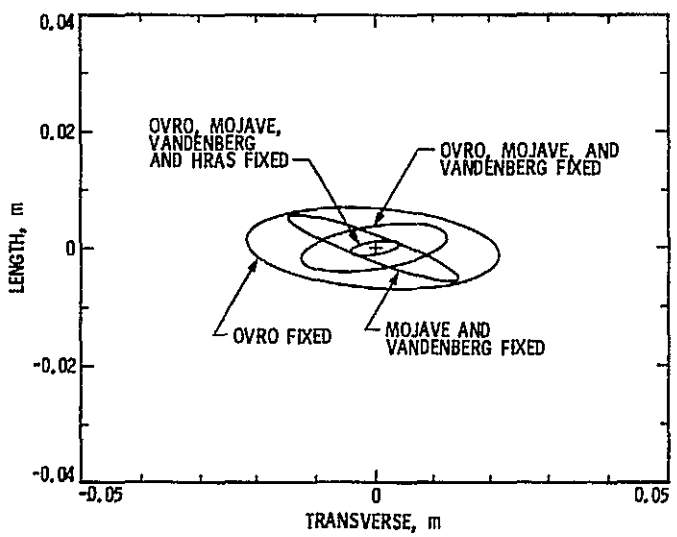


Fig. 8. Additional error from UT and PM, Pt. Reyes/Yuma baseline; errors with LLR data and various base station groups

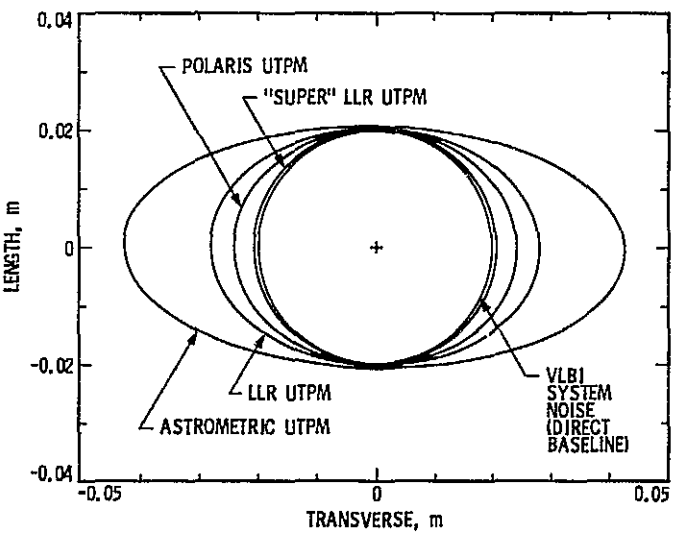


Fig. 10. Total differenced baseline error, Monument Peak/Quincy baseline; errors with only OVRO 130 as base station and various UTPM services

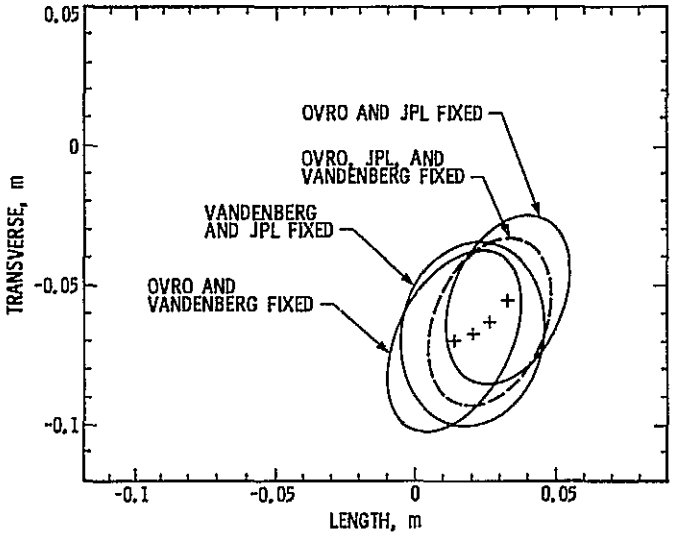


Fig. 11. Pearblossom/Pinyon baseline; MASTERFIT solutions of a differenced baseline with various pairs of base stations

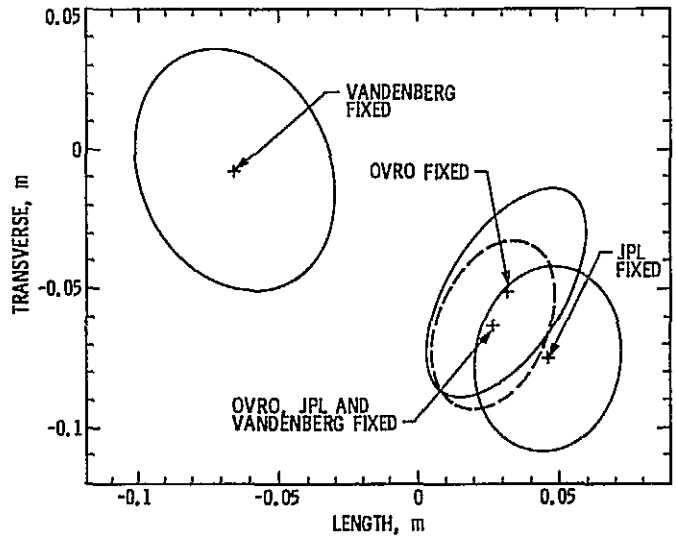


Fig. 12. Pearblossom/Pinyon Baseline; MASTERFIT solutions of a differenced baseline with various single base stations

D21

N84 32643

Mojave Base Station Implementation

C. G. Koscielski
Mission Support and Operations Division

A 12.2 meter (40 ft.) diameter X–Y mount antenna has been reconditioned for use by the Crustal Dynamics Project as a fixed base station. System capabilities and characteristics are presented, as well as key performance parameters for subsystems. The implementation is complete and transfer to National Geodetic Survey/National Oceanic and Atmospheric Administration (NGS/NOAA) is under way.

I. Introduction

Since the late 1960's, NASA has been using space technology to develop new methods for making geodetic measurements over large areas with high precision, high mobility, and minimum observation time. One of these technologies is microwave Very Long Baseline Interferometry (VLBI), which uses extragalactic radio sources (quasars) as measurement reference points (Ref. 1).

VLBI is used by NASA's Deep Space Network in the navigation of interplanetary spacecraft (Ref. 2). Its application to mobile geodetic systems is being accomplished with the collaboration of the radio astronomy community.

A VLBI geodetic system consists of a pair of radio telescopes at separate locations. Quasars emit large amounts of intense microwave radiation. The two stations simultaneously receive and record these random quasar signals, which arrive slightly sooner at one of the antennas, depending on which station is closer to the radio wave front. By using ultraprecise atomic clocks, it is possible to measure the slight difference in the signal arrival times. The time difference and the pointing angles of the two antennas are used geometrically to determine the baseline distance between the two stations (Ref. 3).

If one of the stations is mobile, it can be positioned on the opposite side of an earthquake fault from a fixed station, and measurements can be made over a period of time. Such measurements can determine changes in the baseline in all three dimensions (longitude, latitude, and elevation) with subdecimeter precision (Ref. 4).

The Mojave Base Station is the newest fixed station to join the network of fixed and mobile antennas. The 12.2 meter X–Y mount antenna was built in 1962 to support the Relay I communication satellite. The facility, operated by the Goddard Space Flight Center (GSFC), along with 7 other X–Y type meter antennas located around the globe, formed the Data Acquisition Facility Network. These stations were a prototype of modern satellite "earth stations." In 1976, after program completion, the station was placed in standby status by GSFC.

As early as 1978, JPL recognized the need for a dedicated VLBI base station facility at Goldstone. The Venus station had

served as the original base station, but many other commitments of the station caused serious schedule conflicts with the Crustal Dynamics Project (CDP) requirements. An agreement was reached within NASA which provided for transfer of the 12.2 meter antenna and the 12,000 square foot control building to the CDP.

II. System Requirements

A study team of JPL and GSFC engineers was formed to establish system requirements. See Table 1 for key system requirements. See Fig. 1 for a Mojave Base Station detail block diagram.

III. Implementation Plan

The Bendix Field Engineering Corporation (BFEC) of Columbia, MD, was contracted to plan and carry out the rehabilitation of the antenna and the control building. A detailed study was completed in December 1980 which called for sandblasting and painting the antenna and modifying and repainting a portion of the control building. The antenna work included stripping old cables and equipment from the antenna and removing the hydraulic subsystem which was no longer required. The antenna RF cassegrain housing was also removed.

The design of the antenna dish surface does not allow for alignment of the surface tolerance, but the surface accuracy was measured at ±0.020 inches RMS as set by the original manufacturer.

Antenna bearings and gears were inspected and lubricated. The orthogonality of the X and Y axes was checked. The alignment of the X axis, parallel to the North–South meridian, was found to be in error by 30 arc minutes. It was determined that this error could be accommodated by the antenna pointing software. Therefore, no attempt was made to realign the X axis.

Implementation of antenna electronic equipment began in November 1982 and continued into the Spring of 1983.

The Receiver RF feed focus assembly was installed, the antenna electric drive servo motors and electronics were installed and the antenna position encoder assemblies were installed along with antenna position limit hardware. All associated cables were installed on the antenna and between the antenna and control room.

After completion of rehabilitation of the control building in December 1982, the installation of electronic equipment racks began. The antenna controller, Hydrogen Maser frequency standard, Mark III data acquisition terminal, and Hewlett-Packard 1000 computer were installed and cabled.

IV. Control Building Rehabilitation

The main control building, M-8, on the Mojave Site was selected as the building to house the electronic equipment for the VLBI system. The building had been vacant or used for storage since 1976, when it was last used to support the ATS-6 satellite. Aside from dirt which had collected over the ensuing 6 years, the building was in good condition. The fire protection system had been maintained regularly, the air-conditioning was operable and plumbing was functional.

Modifications to the air-conditioning were contracted to a local vendor. Air ducts to unused areas were blocked and a bypass was installed to reduce the air flow through the system. These changes were made to reduce energy usage. The system was balanced and tested: temperature varies ±3 degrees F over a 24 hour period. The electric power distribution system was modified to meet the system requirements. Power meters were added to monitor the utility, technical and antenna buses. All rehabilitation work was completed by February 1983 and the building was occupied at that time.

V. Electronic Equipment

The electronic equipment for the VLBI system was provided by GSFC. Several commercial vendors and Haystack Radio Observatory were responsible for the fabrication of various subsystems. The equipment was installed by a team of JPL, contractor, and GSFC personnel.

A. Microwave and Receiver

The Microwave and Receiver Assemblies, along with the phase calibration equipment, are located in a single housing and mounted on the Quadripod Structure of the antenna parabolic dish. The Monitor and Control Assembly monitors microwave and receiver functions and controls operation. Key characteristics of the Microwave and Receiver Assembly are shown in Table 2.

B. Data Acquisition Terminal

The Mark III VLBI Data Acquisition Terminal (DAT) is an integrated computer-controlled electronics/recording system which takes a broadband analog I.F. signal, converts selected frequency windows to video (baseband), separately clips, samples and formats each video signal, and records the resulting time-tagged Mark III serial data streams in parallel on magnetic tape. The DAT includes the control computer and necessary

communication systems for complete computer control. Phase and cable calibration, data storage, tape read after write and reference frequency distribution facilities are included in the DAT. See Reference 5 for a detailed explanation of the Mark III DAT.

C. Frequency and Timing

A Hydrogen Maser frequency standard and clock assembly have been manufactured by the Applied Physics Laboratory of Johns Hopkins University in Baltimore, Maryland. The maser, Serial No. NR-7, embodies the latest technology. Frequency stability requirements are $<1 \times 10^{-14}$ parts for time periods >1 minute up to 1 day (24 hours).[1] The temperature environment of the unit is required to be stable within ±3 degrees F. Measurement of the clock offset to a known standard such as NBS is accomplished by the Goldstone Mobile timing standard. The Hydrogen Maser subsystem supplies 5 MHz and 1 pps signals to other subsystems.

D. Monitor and Control

The Monitor and Control Subsystem (MCS) consists of the HP-1000 computer and software. The RS-232 and IEEE-488 communications interfaces provide a central control point for station operation and the display of all station status and alarm messages. The monitor and control subsystem performs the following functions:

(1) Provides a central control point for the station

(2) Provides schedule input via floppy disk

(3) Prepares the ancillary data record

(4) Provides operator input–output and command via CRT terminal

(5) Generates antenna pointing commands

(6) Monitors subsystem performance

E. Water Vapor Radiometer

The Water Vapor Radiometer provides the sky apparent brightness temperature, which is used to calibrate delays due to water vapor content along the ray path from the radio source to the receiver. A thorough discussion of Water Vapor Radiometer requirements is provided in Reference 4.

F. Antenna Pointing Control and Drive

The antenna pointing subsystem is comprised of several units mounted in various locations throughout the antenna structure and control room.

[1]Chiu, M., personal communication, January 1983.

1. Antenna Control Unit (ACU). The central assembly of the subsystem is the ACU which houses all the electronics to generate motor drive commands in all modes of operation. Motor commands are produced by the ACU as a result of input commands from the front panel rate controls (manual mode), the internal tracking routines (tracking mode), or the position commands (position designate and manual modes). The closed loop output of the ACU is a low voltage analog signal proportional to the antenna position error for each axis. The X and Y axis motor controllers provide the interface between the command output of the antenna control unit and the high current requirements of the antenna drive electric motors.

2. Antenna interlocks. Antenna protection and personnel safety are accomplished by the use of travel limiting switches and safety interlocks located at critical positions throughout the antenna.

3. Angular Position Unit (APU). The angular position of the antenna is reported by the use of optical position encoder–transducers mounted on each axis and an interface–decoder within the ACU. The transducer outputs Gray code which the ACU converts into binary digits. The binary position information is used in all tracking modes.

Table 3 provides key characteristics of the Antenna Control Subsystem.

G. Meteorology Data

The Meteorology Data Subsystem (MDS) provides data which aids in the analysis of VLBI data. The atmospheric conditions at a VLBI station present a source of error in measurements which can be removed in final data processing.

Functions of the Meteorology Data Subsystem are to:

(1) Measure outside air temperature.

(2) Measure atmospheric pressure.

(3) Measure relative humidity.

(4) Record meteorology parameters on the MCS.

Table 4 presents the key characteristics of the MDS.

VI. System Tests

After installation of equipment and initial power-up tests, system tests were performed to verify system operation. Tests were performed in the following functional areas:

(1) Receiver–Mark III Terminal Checkout

(2) Antenna Pointing Tests

(3) S- and X-band Performance Tests

A. Receiver and Mark III Terminal

Except for initial cool-down problems with the cryogenically cooled GaAs–FET Low Noise Amplifiers, receiver tests were completed without trouble. Compatibility of the receiver unit which was fabricated at Haystack Radio Observatory and the air-cooler unit fabricated by a local vendor in Barstow, was a challenge. After minor difficulties, the unit operated and temperature tests indicated proper operation was obtained. Performance results are shown in Table 2.

B. Determination of Pointing Corrections

An antenna must be accurately pointed at quasi-stellar radio noise sources for successful VLBI experiments. Most pointing errors are caused by imperfections of the mechanical structure of the antenna and mounting of the RF feed assembly.

D. B. Shaffer[2] of Interferometrics, Inc., provided a methodology for determining pointing corrections. The procedure describes some of the considerations for good pointing determinations, contains a list of the best sources to be used for pointing observations and gives some suggested observing plans. Pointing accuracy of the 12.2 meter antenna was determined by Shaffer to be .02 degrees at X-band. Performance results are shown in Table 3.

C. S- and X-Band Performance

During June 1983 RF performance of the 12.2 meter (40 foot) antenna was determined. Aperture efficiency and system temperature were measured by tracking radio star noise sources (see Table 5). Observations were performed with two different front-end configurations: one configuration included uncooled amplifiers, while the other configuration used cooled GaAs–FET amplifiers.[3]

VII. Activities

The Mojave Base Station (MBS) began VLBI observations in support of the Crustal Dynamics Project in June of 1983 and since then has participated in each observation session with the mobile vehicle antennas. The MBS has participated in several other special observations including the first baseline measurements between the VLBI station at Kashima, Japan and the U.S.

The original NASA Crustal Dynamics Project plan called for NASA to transfer a dedicated base station to the National Geodetic Survey/NOAA Branch of the Department of Commerce. The transfer of the Mojave Base Station will take place in 1984. GSFC is responsible for the transfer, but JPL has the task of operating the facility until transfer and of assisting GSFC with the transfer. Actual station operation is performed by Bendix Field Engineering Corporation personnel.

The Mobile Vehicle (MV) stations will also be transferred to NOAA as planned beginning in 1984 with the transfer of the 5-meter MV-3 antenna and electronic trailer van. The 4-meter (MV-2) and 9-meter (MV-1) facilities will be transferred in 1985. MV 2 and 3 will be located and operated out of the Mojave Base Station. MV-1 is permanently located at Vandenberg Air Force Base.

A depot level maintenance facility has been located at the Mojave Base Station to support the MV's and to perform special engineering and testing functions.

[2]Shaffer, D. B., personal communication, September 1983.

[3]Shaffer, D. B., personal communication, August 1983.

References

1. MacDoran, P. F., "Radio Interferometry for International Study of the Earthquake Mechanism," *Acta Astronautica*, Vol. 1, pp. 1427–1444. March 1974.

2. Renzetti, N. A., J. F. Jordan, A. L. Berman, J. A. Wackley, and T. P. Yunck, *The Deep Space Network – An Instrument for Radio Navigation of Deep Space Probes*, JPL Publication 82-102, Jet Propulsion Laboratory, Pasadena, California, pp. 68–71. December 1982.

3. Niell, A. E., K. M. Ong, P. F. MacDoran, G. M. Resch, D. D. Morabito, E. S. Claflin, and J. F. Dracup, "Comparison of a Radio-Interferometric Differential Baseline Measurement with Conventional Geodesy," *Technophysics*, Vol. 52, pp. 49–58. 1979.

4. Winn, F. B., S. C. Wu, G. M. Resch, C. C. Chao, O. H. von Roos, and H. Lau, "Atmospheric Water Vapor Calibrations: Radiometer Technique," *The Deep Space Network Progress Report 42-32*, Jet Propulsion Laboratory, Pasadena, California, pp. 38–49, April 15, 1976.

5. Rogers, A. E. E., R. J. Cappallo, H. F. Hinteregger, J. I. Livine, E. F. Nesman, J. C. Webber, A. R. Whitney, T. A. Clark, C. Ma, J. Ryan, B. E. Corey, C. C. Counselman, T. A. Herring, I. I. Shapiro, C. A. Knight, D. B. Shaffer, N. R. Vandenberg, R. Lacasse, R. Mauzy, B. Rayher, B. R. Schuper, and J. C. Pigg, "Very-Long Baseline Radio Interferometry: The Mark III System for Geodesy, Astrometry and Aperture Synthesis," *Science*, Vol. 219, pp. 51–54. January 1983.

Table 1. Key system requirements

Simultaneous Reception Bands:	8180–8580 MHz 2235–2335 MHz
Polarization:	RCP Both Bands
Operating System Temperature:	150°K max. (X-band) 100°K max. (S-band)
System Instantaneous Bandwidth (1 dB):	400 MHz (X-band) 100 MHz (S-band)
IF Frequency:	100–500 MHz (X-band) 215–315 MHz (S-band)
Signal Cable Calibration:	Measure Delay to 10 pico sec.
Data Acquisition Terminal:	Mark III Compatible
Data Record Rate:	112 MB/s
Channels:	28
Antenna Blind Point Accuracy:	0.1 Antenna Beamwidth at X-band
Antenna Slew Rate:	1.0 degree/sec.
Frequency Standard:	Hydrogen Maser
Water Vapor Radiometer:	2 Frequencies
Meteorology:	Measure Temperature, Humidity and Barometric Pressure

Table 2. Microwave and receiver key characteristics

Input Frequencies:	X-band 8180–8580 MHz (–1 dB) S-band 2210–2350 MHz (–1 dB)
Aperture Efficiency:	≥ .45 X-band ≥ .47 S-band
Antenna Beamwidth:	.205 degrees, X-band .745 degrees, S-band
System Temperature: (at Zenith)	60°K X-band (Cooled FET) 72°K S-band (Cooled FET)
Intermediate Frequency: Bandwidth	X-band 400 MHz Wide at –1 dB S-band 140 MHz Wide at –1 dB
Phase Calibration Input:	–30 dB coupler

Table 3. Antenna pointing control key characteristics

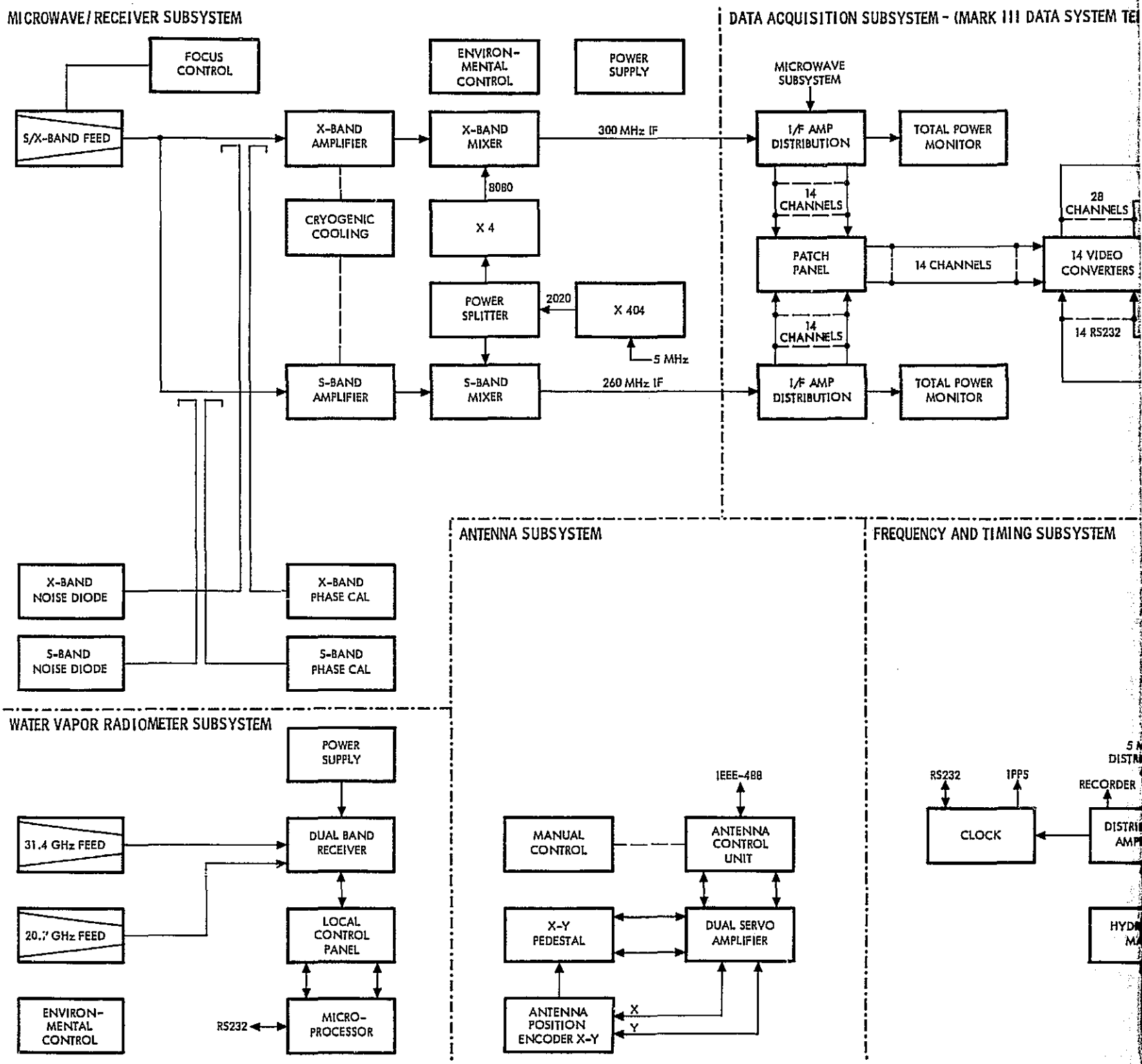
Prime Mover:	Electric Drive Motors, 2.5 Horsepower, 2 Per Axis
Elevator Limits:	>80° from Zenith in all Directions
Maximum Velocity:	1.0 degrees/second Both Axes
Pointing Accuracy:	0.02 degrees Both Axes
Angle Readout:	
Accuracy	0.01 degrees Both Axes
Resolution	0.00275 degrees Both Axes

Table 4. Key characteristics of the Meteorology Data Subsystem

Measure:	
1. Temperature	1°C
2. Relative Humidity	5%
3. Barometric Pressure	≤ 1 millibar
Readout Resolution	0.1 millibar
Measurement Accuracy at Sea Level	.01%

Table 5. Aperture efficiency and system temperature

	Cooled FET Tsys °K	Aperture Efficiency
S-band	72	.47
X-band	60	.45



FOLDOUT FRAME

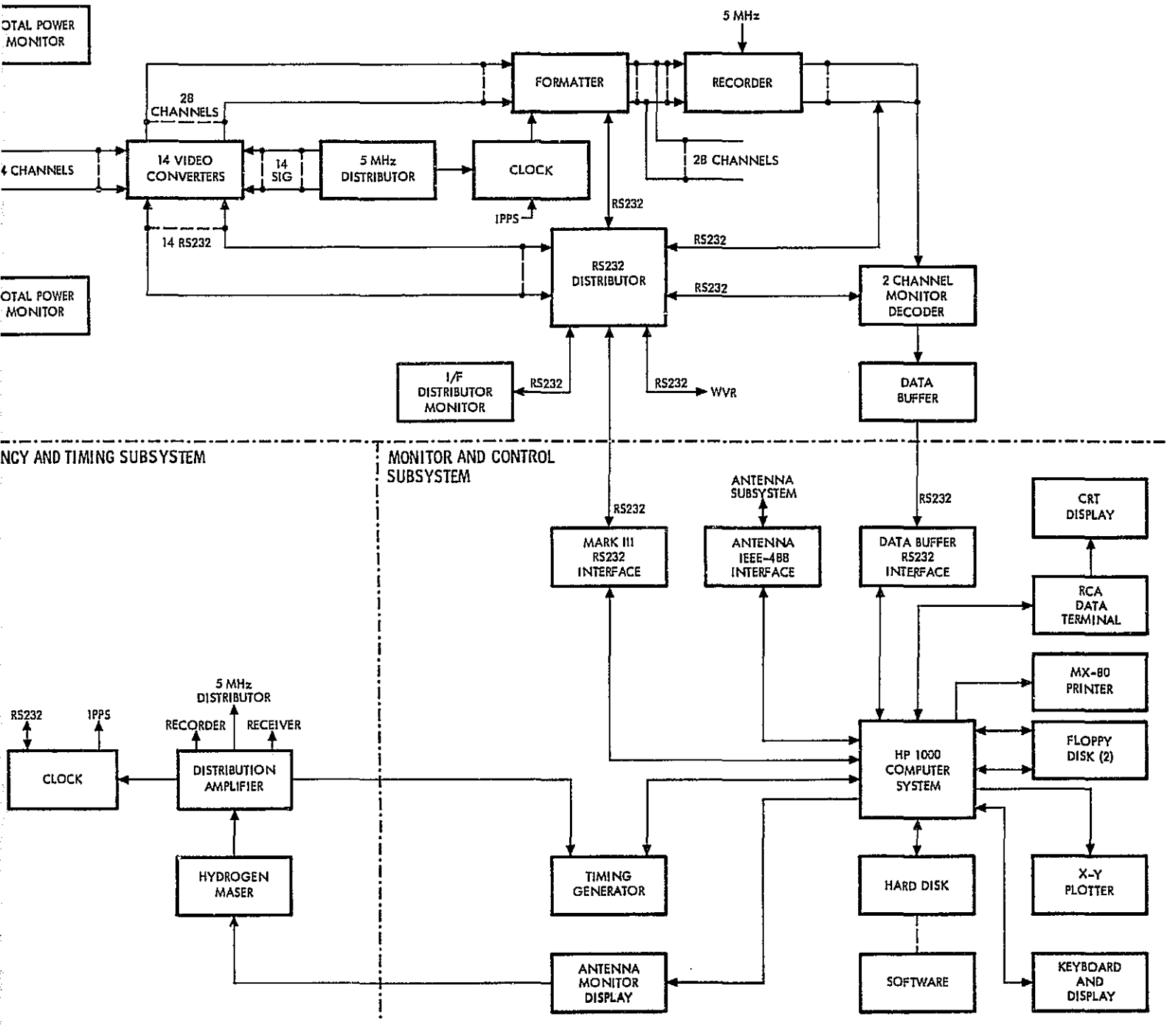


Fig. 1. Fixed station block diagram

PRECEDING PAGE BLANK NOT FILMED

2 FOLDOUT FRAME

OMIT

Referees

The following people have refereed articles for the *Telecommunications and Data Acquisition Progress Report*. By attesting to the technical and archival value of the articles, they have helped to maintain the excellence of this publication.

R. J. Amorose	J. L. Fanselow	P. F. Kuhnle	W. G. Stinnett
T. H. Bird	C. A. Greenhall	F. L. Lansing	L. Swanson
D. W. Brown	R. W. Hartop	J. W. Layland	B. M. Sweetser
P. S. Callahan	R. L. Horttor	J. R. Lesh	R. L. Sydnor
C. T. Chian	R. S. Hughes	P. L. Parsons	D. W. Trask
J. M. Davidson	W. J. Hurd	S. M. Petty	Q. D. Vo
L. Deutsch	W. A. Imbriale	H. P. Phillips	C. C. Wang
D. Divsalar	C. W. Johnson	F. Pollara	R. A. Winkelstein
H. Donnelly	R. F. Jurgens	N. A. Renzetti	G. H. Winn
G. S. Downs	M. J. Klein	E. R. Rodemich	S. Wu
R. F. Emerson	P. H. Knowlton	S. D. Slobin	