https://ntrs.nasa.gov/search.jsp?R=19830011495 2020-03-21T04:11:04+00:00Z

NASA CR- 170,071

NASA-CR-170071
19830011495

The Telecommunications and Data Acquisition Progress Report 42-72

October - December 1982

E C. Posner
Editor

[illegible] COPY
[illegible] 1983
LANGLEY RESEARCH CENTER
LIBRARY, NASA
HAMPTON, VIRGINIA

February 15, 1983

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

All Blank Pages

Intentionally Left Blank

To Keep Document Continuity

The Telecommunications and Data Acquisition Progress Report 42-72

October - December 1982

E.C. Posner
Editor

February 15, 1983

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

N83-19766 #

The research described in this publication was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under 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 Project
  - New Initiatives
  - Network Sustaining
- DSN Operations
  - Network Operations and Operations Support
  - Mission Interface and Support
  - TDA Program Management and Analysis
- GCF Operations and Implementation
- Data and Information Systems

OSSA Tasks

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

Discretionary Funded Tasks

Contents

NETWORK SUSTAINING

DSN Operations

NETWORK OPERATIONS AND OPERATIONS SUPPORT

MISSION INTERFACE AND SUPPORT

GCF Operations and Implementation

CAPABILITIES TO MEET PROJECT REQUIREMENTS

OSSA TASKS

Geodynamics

ORION

Description and Overview of an Instrument Designed to Measure Line-of-Sight Delay Due to Water Vapor

G M Resch
Tracking Systems and Applications Section

M C Chavez and N I Yamane
Microwave Observational Systems Section

Eight dual-channel microwave radiometers have been constructed as a research and development effort for the Crustal Dynamics Project and the Deep Space Network These instruments, known as water vapor radiometers, are primarily intended to demonstrate that the variable path delay imposed by atmospheric water vapor can be calibrated in microwave tracking and distance measuring systems but could also be used in other applications involving moist air meteorology and propagation studies They are being deployed to various stations and observatories that participate in Very Long Baseline Interferometry (VLBI) experiments including the Deep Space Stations in Spain and Australia In this paper we review the development history of these instruments, outline the theory of operation and overall design considerations, and sketch the instrumental parameters and performance characteristics

I. Introduction

It takes longer for a radio wave to traverse an atmospheric path L relative to the time it would take to traverse the same path in a vacuum The electrical path length L_e is just the integral of the refractive index of the atmosphere along the path The difference ΔL between the electrical path length and the physical path length is simply,

$$\Delta L = 10^{-6}\int N\,ds \tag{1}$$

where N is the refractivity at the point s along the path The refractivity for the atmosphere as given by Smith and Weintraub (Ref 1) is composed of two terms The first and largest term is called the dry term and is proportional to the integrated dry air density For most purposes it is sufficient to "weigh" the atmosphere at the zenith with a barometer in order to estimate the "dry" delay ΔL_d and then scale this zenith quantity to the line-of-sight using a cosecant elevation law The second term contributing to the refractivity is a function of the atmospheric water vapor The excess path delay in centimeters for this term of the refractivity is of the form,

$$\Delta L_v = 0\,1723\int (\rho_v/T)\,ds \tag{2}$$

where ρ_v is the vapor density in gm/m^3, T is the temperature in kelvin, and s is in meters Unfortunately (for our applica-

tions) atmospheric water vapor is a highly variable and not well-mixed atmospheric constituent As a result it is impossible to estimate the path delay in Eq (2) with high accuracy from surface measurements alone

In interferometric systems our primary observable is the differential time-of-arrival of a radio wave at the stations comprising the interferometer In a ranging system the primary observable is the round-trip signal time For both systems atmospheric water vapor along the signal path will impose an additional delay between 3 to 60 cm (one way) depending on how much vapor is in the atmosphere and the elevation angle of the observations Typically, the baseline or range determination is derived from observations at several different elevation angles so that the mapping of propagation effect error into baseline or position is rather complex In general, if one requires a system accuracy better than 10-12 cm then the vapor along the signal path must be measured Of the several techniques that could be used to measure line-of-sight vapor delay, the most cost effective utilizes the techniques of passive remote sensing with a dual-channel microwave radiometer The water vapor molecule emits spectral radiation at a frequency of 22 235 GHz, whose intensity is approximately proportional to the number of molecules (or the delay) along the line of sight The presence of liquid water complicates the measurement problem as it contributes significantly to the intensity of the atmospheric emission but contributes very little to the excess path delay However, with a second microwave channel it is possible to separate the vapor and liquid effects in all but the most severe weather situations

II. History

At its inception, it was immediately recognized that the technique of Very Long Baseline Interferometry (VLBI) could be used to investigate a wide variety of geodetic and geophysical phenomena (e g , see Ref 2) The basic instrumentation appeared capable of centimeter-level measurement precision Shapiro and Knight (Ref 3) pointed out that the overall accuracy of VLBI would be limited by systematic effects In particular, the variable delay imposed by tropospheric water vapor seemed to be the limiting error source

In 1972 the Jet Propulsion Laboratory began a research task called ARIES (Astronomical Radio Interferometric Earth Surveying) under the leadership of P F MacDoran The objective of this task was to demonstrate that VLBI could be used to study the deformation of the earth's crust The idea was to use a transportable 9-m antenna as one end of an interferometer, with base stations at Goldstone and the Owens Valley Radio Observatory If the transportable station was moved around a network of geophysically interesting sites in southern California it would be possible to compile a time history of three-dimensional baseline vectors whose changes in time would tell us something about the crustal deformation around major faults in this tectonically complex region The accuracy goals for this task were beyond the capability of conventional surveying techniques Our aim was to measure three-dimensional vector baselines up to 1000 km long with 3- to 5-cm accuracy in each component and 1- to 2-cm accuracy in length Ong et al (Ref 4) demonstrated that the ARIES instrumentation was capable of 3-cm accuracy by measuring a 300-m baseline between the 9-m antenna and the 64-m antenna at Goldstone Progress toward the reduction or elimination of other major error sources was underway By 1974 it seemed that three-dimensional baseline accuracy of a few cm would be possible in just a few years – if only the water vapor problem could be solved

During the summer of 1974 J W Waters and R Longbothum compared a spacecraft prototype microwave radiometer with radiosondes launched at the El Monte Airport in an experiment to determine atmospheric delay using the techniques of passive remote sensing Their regression analysis against radiosonde data indicated a level of agreement better than 2 cm (units of excess path delay) Later work by Winn et al (Ref 5) and Moran and Penfield (Ref 6) extended this basic result It was clear that this device, which became known as a water vapor radiometer (WVR) could be used to calibrate vapor-induced delay along the line of sight of a nearby antenna If the water vapor delay along the line of sight at each station could be calibrated to 2 cm, then 3- to 5-cm VLBI system accuracy seemed possible During this same period a VLBI data acquisition system was being developed at the Haystack Observatory that was well suited for the small mobile VLBI antennas that we intended to use in ARIES It was clear that ARIES needed a WVR that would operate as an integral part of this much larger, fully automated VLBI data acquisition system The WVR would have to operate unattended and require a minimum of operator attention It must be well calibrated, stable, able to operate in all but the most severe weather conditions, and provide a vapor path correction that was accurate to better than 2 cm There was also a need to understand the random and systematic error budget of the WVR and to develop an inversion algorithm to convert the WVR observable (brightness temperature) to path delay correction

The technical aspects of demonstrating the performance of the WVR and specification of its system operation turned out to be the easy part of the problem Finding the funding to actually build the kind of WVR that ARIES needed was a much more complicated, time consuming, and discouraging task In late 1976, with the help of E J Johnston, ARIES was able to borrow the engineering model of a spacecraft radiometer and began a low key effort to modify and repackage the

device for VLBI support In 1977, the VLBI group at Goddard Space Flight Center (GSFC) purchased two microwave receivers from Sense Systems Inc (corresponding to the two channels of a WVR) with the intention of building their own WVR Of course everyone realized that WVR calibration was needed on both ends of the interferometer Both the ARIES and GSFC groups used the Owens Valley Radio Observatory (OVRO) as a base station and they simultaneously perceived a rare opportunity to act in their mutual self-interest and equip the observatory with a WVR Support was obtained from the Caltech President's Fund to purchase one of the microwave receivers necessary to make a WVR, and GSFC purchased the second channel as well as two positioners The ARIES group agreed to put all of these assorted pieces together to make three reasonably compatible WVRs Our intention was to use these WVRs as tools to demonstrate the validity of the technique as it applies to VLBI support and then build units that were an optimum design for our application In early 1978, with the support of the Goddard group, we received partial funding to package three WVRs Several months later we received directions to construct two additional WVRs for the Crustal Dynamics Project (which was then coming into existence)

The situation in July 1978 was very chaotic We had some of the parts for three WVRs, directions to put together five WVRs, no clear funding pattern, and were restricted from carrying over money from one fiscal year to the next We decided that it would be less expensive and faster to construct the two additional WVRs "in house" and immediately began the purchase of long-lead-time parts In the latter part of 1978 the Deep Space Network (Advanced Systems and Implementation Offices) added two more radiometers to what had now become a production line A year later Project ARIES requested another WVR (its second unit) These units were given several numbers, R01, R02, , R08 according to the order in which we began work on the hardware The "R" designation served to emphasize the R&D nature of the effort

We were directed by JPL management to construct these WVRs in a manner that was consistent with the needs of both JPL and non-JPL users This presented a problem in that we had already inherited the pieces of three radiometers and were constrained to use off-the-shelf components for the new units Our approach to this problem was dictated by our stated goal – to build research tools that were capable of demonstrating that vapor delay could be calibrated in VLBI measurements We did *not* attempt research in radiometer construction, but instead decided to adopt the "safe" design of a Dicke switched instrument taken largely from a spacecraft design It was decided to standardize on the two frequencies 20 7 and 31 4 GHz, use doubly stabilized Dicke radiometers, and discard the cold load calibration in the Sense Systems receivers The latter point was dictated by three factors (1) the load was inaccurate, (2) it was operationally cumbersome to keep filled with liquid nitrogen, and (3) it presented severe packaging problems

For three reasons we decided to mount the WVRs on their own positioner First, we assumed that if the WVR would be used for any application other than VLBI support it should not tie up the VLBI antenna Second, the procedure that we use to establish and check the temperature scale calibration (i e , the tipping curve) is rather awkward and time consuming when performed on a large antenna Third, preliminary analysis suggested that the only difference between the broad beam WVR and the smaller beam radio telescope would be in the short-period fluctuations of water vapor Since most VLBI observations are the result of incoherently averaging up to 15 minutes of data, it effectively smoothes the fast but small vapor fluctuations In order to simplify the interfacing we added a microcomputer controller to which the user can send simple commands and have the WVR point, take data, etc

There were several minor areas of design, assembly, and/or construction in the first three instruments that we felt could be improved The changes that were finally incorporated into units R04-R08 were dictated by the requirement that all WVRs appear identical to the user at the controller interface Unfortunately it proved to be impractical to remove all of the differences between the different units The most obvious remaining difference is that the vapor frequency of WVR-R02 and R03 is centered at 21 GHz (as compared to 20 7 GHz for all other units) and the slew rate is approximately a factor of 4 slower than the other radiometers Thus the eight radiometers that have been deployed consist of three different microwave packages, two different positioners, and three slightly different microcomputer controllers However, thanks to the software that resides in the microcomputer these differences are largely transparent to the user The current plan is to mount units R01 and R07 on the edge of the antenna they are intended to calibrate and to use the fast mounts from these units to retrofit units R02 and R03

In July 1979 the unit designated WVR-R03 was delivered to the Haystack Observatory According to our original agreement with GSFC (which now managed the Crustal Dynamics Project), the WVRs would be merely bench tested at JPL Each observatory would be responsible for engineering tests and calibration of the WVR it received This arrangement was found to be impractical, so in late 1979 at the request of the Crustal Dynamics Project we agreed to test and calibrate each WVR prior to shipment In addition, we would assist with installation of the WVR at those observatories that needed help, keep a stock of spare parts, assist with maintenance problems, provide a minimal level of documentation (i e ,

consistent with R&D instruments), and provide the algorithm necessary to convert WVR output into a delay correction In order to provide a backup capability for vapor calibration as well as a check on WVR operation, we also agreed to construct, test and calibrate a solar hygrometer to go along with each WVR These extensions of our responsibilities turned out to be more difficult and time consuming than was originally anticipated WVR-R02 was installed at OVRO in September 1980, unit R05 was installed at the Haystack Observatory in January 1981, unit R06 was installed on the ARIES electronics van (Mobile Van-2) also in January, unit R04 was installed at DSS 13 in February, unit R01 was delivered to ARIES (Mobile Van-1) in May, and unit R08 was installed at Ft Davis in June 1981 As per our agreement with the Crustal Dynamics Project, we have purchased a modest stock of spare parts, some of which have been used to solve problems at several installations We have installed the WVRs where it was necessary

This paper is the first of a series that will constitute the "minimal" level of documentation In later articles we will discuss the control module, the microwave package, the positioner, testing and calibration, the algorithms, and the solar hygrometers

III. Remote Sensing of Water Vapor

The physics of atmospheric water vapor and the techniques used to sense it are well developed in Refs 7 and 8 We shall merely summarize a few pertinent points in this paper In Fig 1 we see the apparent brightness temperature of the zenith atmosphere in the frequency interval 10 to 40 Ghz The lower curve illustrates the behavior of the brightness temperature in the complete absence of atmospheric water vapor The middle curve in Fig 1 is drawn for the case of 2 gm-cm^{-2} of precipitable water vapor M_v, and the upper curve represents the spectrum for an additional 0 1 gm-cm^2 of liquid M_L that is assumed to exist as small droplets This is a columnar measure, that is, if we were to take a column 1 cm^2 at the base that extended through the entire atmosphere and precipitated all of the vapor, we would wind up with 2 grams of water For our application a "handy dandy" relationship to remember is

$$\Delta L_v \simeq 6\,3\,M_v \tag{3}$$

or, the excess path delay ΔL_v (in cm) is about 6 times the precipitable vapor content M_v (expressed in gm/cm^2) Thus the middle curve of Fig 1 refers to a zenith vapor delay of 12 6 cm, representing roughly average conditions over much of the United States The reader should note that in the literature, precipitable vapor is often measured in centimeters (instead of gm/cm^2) This results in some confusion when instruments that measure precipitable vapor are compared with instruments that measure path delay

The bump in the brightness temperature curve of Fig 1 at the frequency of 22 235 GHz is due to an emission line from the water vapor molecule If there is not a lot of water vapor along the line of sight (i e , the line is not saturated), then the brightness temperature is roughly proportional to the columnar density plus the background radiation The background radiation is due to three sources (1) the blackbody radiation of the cosmos, which is known to be 2 9 K, (2) emission from the lower wings of higher frequency lines, in particular the band of oxygen lines near 60 GHz, and (3) emission and scattering of radiation from liquid water droplets that occur in clouds or rain The emission from oxygen is calculable and can be removed from the observations The presence of liquid water has little effect on the path delay but considerable effect on the brightness temperature, thereby confusing the vapor measurement With measurements at two frequencies we can separate the vapor and liquid contributions

If the average diameter of the water droplets is small compared to observing wavelength, then the spectrum of the radiation from the liquid component will vary as frequency squared (Ref 9) At some frequency f_1 (near the vapor emission line) the brightness temperature T_{b1} will be proportional to $k_1\ \Delta L_v + k_2 M_L$, where k_1 and k_2 are constants, and M_L is the integrated liquid content At some frequency f_2 (off the vapor emission line) the brightness temperature T_{b2}, will be proportional to $k_1'\ \Delta L_v + k_2' M_L$ If we observe these two brightness temperatures T_{b1} and T_{b2}, then we have two equations in two unknowns Given the frequency-squared dependence of the liquid emission, these equations can be solved simultaneously for ΔL_v and M_L We find that both the path delay and the precipitable liquid can be expressed as a linear combination of the observed brightness temperatures,

$$\Delta L_v \simeq a_0 + a_1 T_{b1} + a_2 T_{b2} \tag{4}$$

where a_0, a_1, and a_2 are "constants" This equation summarizes the basic measurement approach and is useful in considerations of performance of the measurement system In actual use where observations are made at low elevation angles and in cloudy conditions, Eq (4) is not a good approximation A slightly more complicated expression analogous to the approach used in Ref 10 will be derived in a later paper

IV. The Instrumentation

A water vapor radiometer is simply a device for measuring sky brightness temperature at two frequencies on and near the

emission line at 22 2 GHz It consists of two independent Dicke radiometers (see Ref 11 for an explanation of the operation of a Dicke radiometer), one tuned to 20 7 GHz and the other tuned to 31 4 GHz The RF electronics, switching circuitry, and power supplies comprise the microwave package The microwave package is mounted atop a positioner that can point in azimuth and elevation Both the microwave electronics and positioner are interfaced to the controller package that consists of a microcomputer and local control panel from which one can monitor any of the analog signals from either the microwave electronics or the positioner The ability to monitor and control all WVR functions from the local control panel is very useful for diagnostic purposes Under normal operating conditions, the microcomputer controls and acquires data from the microwave package and the positioner The microcomputer can be interfaced to another computer (i e , a host computer) or to a terminal via an RS-232 serial interface A set of simple instructions is available to command the microcomputer to point the WVR, switch to internal loads, take data, etc Figure 2 shows all three WVR modules, Figure 3 is a block diagram of the complete WVR

V. System Requirements

In VLBI the basic observable is a group delay (Ref 12), i e , the differential time of arrival of a radio wave at the stations comprising the interferometer Differing amounts of atmospheric water vapor surrounding the stations cause the observed group delay to differ from the calculated group delay (e g , calculated from the geometry). Of course there are many other sources of error that lead to differences between observed and calculated delay Our ability to either calibrate or "solve-for" these error sources sets the limit on the accuracy of the VLBI technique to measure vector distance As discussed in Ref 13, the analysis of VLBI error sources suggests that baseline accuracies of 3 to 5 cm are possible if the line-of-sight water vapor effects can be calibrated to ±2 cm At the 2-cm level there are several comparable error sources, so that it simply does not make sense to invest a large amount of resources in the reduction of any single error source Keep in mind that the 2-cm vapor delay accuracy refers to a differential measurement, so that the accuracy of an individual measurement should be $2/\sqrt{2}$ cm We would like the precision of the vapor measurement to be a factor of 3 or 4 better in the hope that the increased precision could be used to spot systematic effects Hence the overall system requirement for the WVR is that it measure vapor delay with accuracy of ±1 5 cm and precision of ±0 4 cm Our problem in 1978 was to convert this requirement on system performance into a general set of design specifications

As a response to a directive from the JPL Program Office for the Crustal Dynamics Project we formed a design team to review the design specifications for a water vapor radiometer The conclusions and recommendations of this team were issued as a JPL internal report, which was written in the very early stages of construction and became the engineering specifications document Table 1 summarizes the specifications of the water vapor radiometers

A. Physical Characteristics and Environment

Since we started with a stock of existing parts the physical characteristics of the WVR were somewhat predetermined The physical characteristics of units R01 and R04 to R08 are listed in Table 1 Units R02 and R03 have slightly larger volume but more or less the same physical characteristics The environmental factors were specified from consideration of the meteorological conditions for both the mobile and fixed antenna sites that were under consideration by the Crustal Dynamics Project At temperatures below -10°C it was felt that the atmospheric vapor content would be very small and hence there would be little need for the vapor correction The WVRs will survive temperatures as low as -40°C but below -10°C it becomes increasingly more difficult for the internal heaters to thermally stabilize the electronics Similarly, at wind speeds above 65 km/hr the VLBI antenna would probably shut down At wind speeds higher than 95 km/hr it is recommended that the WVR be stowed in the horizontal position in order to present the minimum surface to wind loading A particularly irksome problem that was not solved was the problem of dew accumulation on the cover of the horn antennas

B. Operating Frequencies

One of the critical specifications in the WVR design is the choice of operating frequencies This is because the water vapor line centered at 22 235 GHz is pressure broadened Thus the line profile will depend on the altitude distribution of the vapor A given amount of vapor that was concentrated at low altitude would produce a wide, flat profile relative to the same amount of vapor that was concentrated at high altitude Most of the vapor signal is produced at 22 235 GHz but this is also the frequency that is most sensitive to the shape of the line However, if we pick a frequency on the wing of the line, e g , between 20 and 21 GHz, we can minimize this sensitivity to vapor height (Ref 14 and 15)

The components of the NEMS (unit R01) radiometer operated at 22 2 and 31 4 GHz, and the two Sense Systems radiometers operated at 21 0 and 31 4 GHz Preliminary bench tests indicated that the NEMS components could possibly be used as low as 20 7 GHz There is a decided advantage of having all WVRs operate at identical frequency pairs, so we decided to standardize on 20 7 GHz as the vapor sensitive channel for all new WVRs and attempt to retune the existing components

C. Frequency Stability

The specification of frequency stability is important because the vapor channel operates on the part of emission line where the rate of change of brightness temperature with respect to frequency is largest The frequency will change because the Gunn diode local oscillator is temperature-sensitive and the WVR enclosure does not act as a perfect oven The sensitivity is approximately 3 MHz/°C and since temperature changes of ±4°C inside the WVR are possible, this implies frequency shifts of ± 012 GHz At a precipitable vapor content of 2 gm/cm^2 the rate of change of brightness temperature with respect to frequency is 3 3 K/GHz, so that 0 04 K errors are possible in the presence of moderate amounts of vapor These errors are below the quantization level of the A/D converter for most observing conditions

D. Bandwidth and Integration Time

The bandwidth specification, like the operating frequency of 20 7 GHz, was set by equipment in-hand Both the NEMS unit and the two Sense Systems units are double sideband receivers with an IF passband from 10 to 110 MHz The Dicke switch rate is 1000 Hz on all of the radiometers The minimum integration time is set by a lowpass filter on the output of the synchronous detector whose characteristics were based on three considerations First, the time constant should be long enough to span many cycles of the Dicke switch and thereby average out transients and any short term dithering in the switch reference frequency Second, the time constant should be large enough to reduce the random noise fluctuations to a value that is several times the quantization level of the analog-to-digital converter Third, the time constant should not be so large as to impose tedious wait states between changes of the input We chose a lowpass filter with an equivalent time constant of 100 millisec and thereby average 100 cycles of the Dicke switch The rms fluctuation level is approximately 0 4 K compared to a quantization level of 0 1 K The wait state after a mode change is determined by the time it takes the previous level to decay to less than the quantization level, which for our design is 9 time constants or 0 9 sec This delay is enforced in the controller software -- the WVR simply will not respond for 0 9 sec after a mode change The effective integration time can be lengthened by software averaging techniques

E. Antenna

Ideally, one would like the antenna beamwidth of the WVR to be comparable to the beamwidth of the radio telescope to be calibrated However, as indicated previously, this restricts the use of the WVR in other applications In addition, a comparable WVR beam would entail considerable cost either as hardware or in development, in that it would mean a WVR aperture whose diameter is comparable to the radio telescope with all of the attendant problems of pointing, spillover, etc These problems are solvable given enough time and money, but this is unnecessary in our application

We chose to use a horn antenna for the WVR that has a relatively broad beamwidth (7 degrees, full width at half maximum) for three reasons (1) it has virtually no sidelobes, (2) the dimensions could be easily scaled to give the same beamwidth at both operating frequencies, and (3) the temporal/spatial resolution of the broad beam was more than sufficient in our applications The beam pattern for this horn is shown in Fig 4 taken from Ref 16 In order to have a WVR beam that was comparable to that of the 64-m antenna operating at a 3 8-cm we would need a 16-m aperture We could either use a separate 16-m paraboloid for the WVR or we might consider mounting the WVR in such a way as to utilize some or all of the 64-m aperture In either case we would have to devote considerable engineering effort toward minimizing the problems of sidelobes and spillover that produce serious systematic error in the delay determination Furthermore, since we use a variety of radio telescopes in our application we would require a custom WVR aperture or feed for each installation Clearly this is impractical -- if we need eight WVRs perhaps a single design is most cost effective However, if one is to use a composite design, then the question that arises is how do we relate the output of the WVR to the output of the radio telescope when they have unequal beams?

Formally, there is an important distinction between the antenna temperature (which is what the WVR measures) and the sky brightness temperature (which is what we wish to measure) The antenna temperature T_a (neglecting losses) is given by a convolution of the brightness temperature distribution $T_b(\theta, \phi)$ with the antenna beam pattern $P(\theta, \phi)$ as,

$$T_a \sim \int T_b(\theta, \phi)\, P(\theta - \theta_0, \phi - \phi_0)\, d\Omega \qquad (5)$$

where θ, ϕ are spherical coordinates and (θ_0, ϕ_0) is the pointing direction If the beam pattern has unwanted responses in directions other than the pointing direction then we must correct for them A typical sky brightness temperature of 30 K would be considerably corrupted by a 100 K cloud or the 300 K earth in a sidelobe even if the sidelobe responses were 20 dB down from the main beam The virtue of having an antenna beam with no sidelobes (i e , very high main beam efficiency) is that we can equate the antenna temperature to the brightness temperature (except for a small pointing correction) and save a great deal of data "massaging "

When we ask how the vapor affects a large radio telescope we must keep in mind that the beam of the radio telescope

takes a relatively large distance to form The extent of the near field of an aperture is

$$L_{nf} \sim D^2/\lambda \tag{6}$$

where D is the diameter of the aperture and λ is the observing wavelength This quantity is tabulated in Table 2 for some of the antennas used in our applications Since the scale height of water vapor is typically 2 km, this means that most of the vapor is in the near field of a large radio telescope Thus fluctuations in phase for the radio telescope are determined by the vapor inhomogeneities that pass through a cylindrical column of length L_{nf} and diameter D

Let us assume that the water vapor exists in the form of "blobs" that are randomly distributed in size If we sample equal volumes of atmosphere containing $N_1, N_2,$ etc , then we would expect to find N blobs on the average with an rms variation of $\sqrt{N}$ We would obviously expect less variation if we sampled with a large volume as compared to sampling with a small volume If we normalize N to be the number of blobs per unit area, then the average value will be independent of sampling volume but the rms variations about the average will be a function of the sampling volume Thus we can roughly estimate the difference in fluctuation between the WVR and radio telescope by simply computing the average vapor contained in the two beams If we will assume that on the average the water vapor is exponentially distributed with scale height H_0, and the WVR beam is approximated by a cone of full-width β, then it is straightforward to calculate the ratio R of the volume of vapor in the WVR beam to the volume of vapor in the near field of the radio telescope,

$$R = V_{wvr}/V_{cyl} = 8[(H_0/D)\tan(\beta/2)]^2 \tag{7}$$

This ratio is shown in Table 2 using the 7-deg beamwidth of the WVR If we neglect instrumental noise then we would expect the ratio of the fluctuations for the WVR to the fluctuations from the radio telescope to vary as the square root of R The WVR, since it samples a larger volume, will vary less, but on the average it will determine the correct delay for the radio telescope

An alternate way to view this problem is to note that Eq (5) is a function of time The brightness temperature describes the two-dimensional projection of an assemblage of water vapor inhomogeneities that are constantly rearranged by winds aloft Thus the quantity T_a is really a time series If we had a pencil beam, i e , $P(\theta, \phi)$ were a delta function, then the time variations of T_a and T_b would be identical The fact that we have a beam of finite width that is convolved with T_b implies that the antenna acts as a lowpass filter by spatially averaging the details of the sky brightness distribution Therefore, we expect the output of the WVR to be a smoothed representation of the output from the radio telescope On very short time scales the radio telescope may exhibit fluctuations that the WVR will not detect

The calculation of the time scale that relates the WVR to radio telescope fluctuations is quite complex, involving the geometry and the three-dimensional spectrum of vapor inhomogeneities We shall resort to a more intuitive approach Consider that at the zenith the width of the WVR beam is approximately

$$D_w = 2H\tan(\beta/2) \tag{8}$$

where H is the height For a height of 2 km (the nominal scale height of water vapor) and a beam width of 7 deg we calculate D_w = 244 m or roughly 10 times the near field width from a 25-m radio telescope For a given blob diameter D and radio telescope diameter D_R, we can distinguish three different regimes Regime 1 $D < D_R$, the ratio of the time that the blob spends in the WVR beam to the time it spends in the radio telescope beam is $T \sim D_w/D_R$ Regime 2 $D_R < D < D_w$, the time scale is $T \sim D_w/D$ Regime 3 $D > D_w$, the ratio of the times becomes $T \sim 1$ Thus in this example the very small blobs of vapor will spend about 10 times longer in the WVR beam as they do in the radio telescope beam, but as the size of the blob grows the ratio of time spent in the two beams approaches unity If we further assume that the horizontal and vertical dimensions of the blob are roughly equivalent then it follows that the small blobs will cause small phase perturbations and large blobs will cause large perturbations The WVR will not "see" the short-period fluctuations in the radio telescope output that are due to water vapor but it can detect the larger long-term effects In reality, nature is not nearly so simple and all that can be safely said is that there is a rough equivalence between the spatial averaging of the WVR and temporal averaging from the radio telescope

F. Calibration

Two types of calibration are required for the WVR (1) a frequent calibration of the relative gain of the instrument, and (2) a less frequent calibration of the absolute gain The reference load of the Dicke switch is kept at the same temperature as the rest of the electronics This means that the Dicke receiver is not fully stabilized (e g , see Ref 11) and we are susceptible to gain changes Determination of the relative gain is accomplished by setting internal switches in the WVR so that the input of the Dicke receiver is connected to either of two waveguide terminations One termination is kept at an elevated temperature (100°C) and is called the hot load while the other is kept at the same temperature as the rest of the electronics (43°C) and is called the base load The physical

temperature of these loads is measured with thermistors and denoted by *TH* and *TB* respectively The output of the receiver is digitized and presented to the user as a number *N* that we shall term "counts" In order to determine an unknown antenna temperature *TA* we can (in principle) go through the observing sequences – hot load, base load, antenna – and observe the outputs *TH, TB, NH, NB*, and *NA* For an ideal radiometer the antenna temperature is simply

$$TA = TB + [(TH - TB)/(NH - NB)]\ (NA - NB) \qquad (9)$$

where the quantity in square brackets is called the relative gain or simply the gain of the instrument and is expressed in units of kelvin/count For all of the WVRs, the value of the gain is approximately 0 1 K/count and is determined by the product of the gains from all circuit elements extending from the Dicke switch to the analog-to-digital converter Figure 5 illustrates the typical fluctuations in the gain from one of the WVR channels We plot the relative gain (i e by subtracting out the average) vs time for a 40-hour period A diurnal variation of ±4% is quite obvious and is due primarily to temperature changes inside the microwave package Superimposed on the diurnal variation are short-term fluctuations due primarily to radiometer noise which can be reduced by averaging or smoothing

Equation (9) describes an ideal radiometer A real instrument suffers various shortcomings such as attenuation, less than perfect switch isolation, reflections, etc The calibration of absolute gain involves finding the correction factor that puts Eq (9) closer to the absolute temperature scale One of the procedures that can be used to find the instrumental correction factor is called a "tip curve", it is used to derive a correction to the hot load Let us assume that we can rewrite Eq (9) for a real radiometer as

$$TA = TB + [TH - TB + \Delta TH]\ N' \qquad (10)$$

where $N' = (NA - NB)/(NH - NB)$ is the normalized count in the antenna mode, and ΔTH is the hot load correction The brightness temperature of the sky can be written

$$T_b = T_c e^{-\tau} + T_m(1-e^{-\tau}) \qquad (11)$$

where T_c = 2 9 K is the cosmic blackbody background, T_m is the mean radiating temperature of the atmosphere, and τ is the opacity The mean radiating temperature exhibits some seasonal and site variations but with a reasonable degree of approximation we can assume that it is constant and equal to 275 K For a stratified atmosphere the opacity can be expressed as,

$$\tau = \tau_0(AM) \qquad (12)$$

where τ_0 is the zenith opacity and AM = 1/sin (elevation) is the air mass Suppose we make a series of observations, first observing the internal loads to get *TB, TH, NB*, and *NH*, and then moving to different elevation angles to get NA_1, NA_2, etc , at positions AM_1, AM_2, etc , and assume that $TA = T_b$ Equations (10), (11), and (12) can be combined to give,

$$N_i' = [T_M + T_B + (T_c - T_m)\exp(-\tau_0 AM_i)]/(TH - TB + \Delta TH) \qquad (13)$$

Thus we have expressed the observable from the WVR as the dependent variable in terms of the independent variable AM_i and the two parameters τ_0, and ΔTH Given two or more observations at different AM_i it is straightforward to solve for τ_0 and ΔTH that are "best" in the least squares sense Figure 6 shows tip curve data taken with WVR-R07 (31 4 GHz) located at Goldstone in May 1981 All WVRs are calibrated in this manner prior to shipment In addition, the temperature scale of each radiometer channel is checked by observing hot and cold aperture loads, which serves as a consistency check

The proof of WVR performance lies in a comparison of the WVR with some independent (and one would hope more accurate!) technique of measuring vapor path delay The most convenient comparison technique is to use a radiosonde to provide a vertical profile of temperature and relative humidity that can be integrated to give an estimate of the zenith path delay which we then compare with the delay estimate from the WVR that is observed at the zenith Figure 7 shows two such comparisons, the first done in May 1974 and the second from May 1975 that was reported in Ref 5 Figure 8 shows previously unpublished data taken at Pt Mugu, California in 1976, and Fig 9 shows the histogram of the residuals for this experiment During the Pt Mugu experiment we had an instrumented aircraft fly various slant paths up to a maximum altitude of 3 km, which provided a larger range of path delay estimates than are normally obtained from radiosondes This data is shown in Fig 10

The rms residuals from these comparisons are 1 6 cm from 1974, 1 1 cm from 1975, 1 5 cm from 1976 WVR/radiosonde, and 1 4 cm from 1976 WVR/aircraft (excluding data at 10 deg elevation) If the two techniques are truly independent then we expect the rms residual to be the root-sum-square of the error from each technique If we take the average rms residual to be 1 5 cm, and assume that the radiosonde has a 10% accuracy, the average path delay to be 1 0 cm, then we infer the accuracy of the WVR to be 1 1 cm

The WVR used in the experiments just described operates at 22 2 and 31 4 GHz Unfortunately none of the new WVRs

have undergone such comparison If anything, the older radiometers should be noisier due to the greater sensitivity to the vertical distribution of vapor Guiraud et al in Ref 17 report a WVR/radiosonde comparison for a WVR that operates at 20 6 and 31 6 GHz and infer a WVR accuracy of 0 5 cm We hope to demonstrate comparable performance with the new WVRs

The instrumental correction factor should be constant for long periods of time (e g , years) There are circumstances that could cause it to change slowly or dramatically For instance, it is possible that the waveguide walls could deteriorate slowly in time due to atmospheric constituents If the enclosure is leaking and there is a sudden change from warm, damp conditions to cold conditions, then liquid water or ice could form in the waveguide or switches or on the cover of the horn antenna In either case the performance of the instrument will change and the predetermined value of ΔTH will not be valid It is recommended that tip curves be performed periodically and the results kept as a history of the instrumental performance

G. Temperature Control

The quantity that correlates most strongly with gain variations is temperature, so that it is important that the WVR have enough thermal inertia to resist sudden changes in temperature Conversely, the gain should be monitored on time scales shorter than the time scale for real temperature changes The important specification is the restriction that the physical temperature of any load must not change by more than the quantization level during the time it takes to measure the gain Thus, the specification ±0 1°C/minute assumes that it will never take longer than one minute to measure the gain

Almost all of the microwave components are thermally connected to a large aluminum mounting plate A mercury thermostat controls strip heaters attached to this plate and regulates its temperature to 43°C A small fan circulates air around the few components that are not attached to the plate The entire microwave package is enclosed by a sheet metal box that is lined with styrofoam and provides a thermal attenuation of approximately 6-7 dB The box is attached to the positioner with a base plate made from a thermally insulating material Thermal baffles are mounted on standoffs on the four sides of the box that expose the largest area These baffles, which are covered with a paint that has high reflectivity in the infrared, were found to be necessary for summer operation in desert-type environments

Figures 11 and 12 show the warmup characteristics of the two waveguide terminations (i e the hot and base loads) Typically, the hot load requires 1 5 to 2 hours to thermally stabilize and will remain constant to ±0 2°C thereafter The base load, which is representative of the remainder of the electronics package, takes up to 3 hours to reach quasi-equilibrium Figure 13 shows the typical diurnal variation of the enclosure temperature, which correlates very well with the gain variations

H. Positioning Module

The specifications for the positioner follow from the requirement that the WVR point along the same line of sight as the radio telescope that is being calibrated This is most easily done with an Az/El mount and conversions from right ascension/declination done in software if necessary Due to the 7-deg bandwidth of the horn antennas, we have not included a capability for tracking – only pointing Also note that the pointing capability is ±1 0 deg but the readout precision is ±0 1 deg Analysis of the pointing sensitivity indicates that positioning errors of ±0 1 deg will give worst-case errors of 0 5 K in brightness temperature and 3 mm in the path delay in high opacity/low elevation angle conditions Using these instruments in a typical VLBI experiment one would point to within ±1 deg of the line of sight of the radio telescope (keeping 10 deg away from obstructions) Then one would request the data necessary to calculate the path delay, and note the actual position of the WVR so that the path delay could be mapped to the correct line of sight by assuming a cosecant elevation angle scaling

The slew rate of the positioner even for the two "slow" units is generally much faster than the associated radio telescope This allows the possibility of more complex observing strategy with the WVR than simply following along behind the radio telescope Details of the operation of the positioner will be given in a later paper of this series

I. Control and Interface Module

This module consists of a local control/monitor panel and microcomputer Using the local control panel the user can manually point the WVR, switch modes, or monitor any of the signal lines from either the microwave or positioner modules These signals are displayed on a built-in digital voltmeter and are available via BNC jacks on the panel in order to drive a chart recorder The control panel is intended primarily to assist in diagnosis of problems In normal operation the control panel is switched to the REMOTE position and all control and data acquisition is exercised by the microprocessor The interface to the microprocessor is set by the requirement of compatibility with the MK III data acquisition system In early talks with colleagues at Goddard Space Flight Center and at the Haystack Observatory it was decided that the WVR would simply be treated as one of several devices on a daisy-chain type of serial interface This interface has come to be known as the MAT (Microprocessor ASCII Transceiver) bus

Originally it was believed that the new RS422/423 interface standard would be necessary to implement the long lines to the WVR that we thought might occur at some installations Unfortunately, reliable components to implement the RS422/423 standard were not available in time to be implemented A simple three wire RS-232 is now in use with short-haul modems for the WVR interface lines

The baud rate of the WVR is set to any standard rate from 75 to 19,200 baud by a combination of software and jumper changes on the microcomputer board At the present time it is set to 9600 baud Implementation of the operating modes and command/respond sequence is primarily a matter of sufficient software in the microprocessor In the current version of this software only the OPERATE mode has been implemented along with a set of primitive commands Details of this software will be discussed in a later paper The microcomputer contains an analog I/O subsystem that is used to digitize the various analog signals from the mocrowave package and the positioner Our desire for 0 1 deg of position resolution dictated the choice of a 12-bit A/D converter (i e , $2^{-11} > 0\ 1/360 > 2^{-12}$) Similarly, the quantization of the physical and brightness temperature measurements is 0 1 K As indicated previously, the fluctuation level from the radiometer is on the order of 0 4 K It is possible to average many samples in order to reduce these fluctuations However, Clark (Ref 18), has pointed out this must be done with caution if one attempts to reduce the fluctuations below the quantization level

VI. In Retrospect

Table 3 summarizes the hardware costs and time expended as a result of our work on the WVRs Hardware costs are referred to epoch October 1981 and for the most part must be inflated to the current date An exception to this is the microprocessor hardware where many components are being reduced in price The hardware, assembly, testing, and calibration costs are listed on a per unit basis The development of the algorithm, microprocessor control software, and the data acquisition software is summarized as a total effort The final accounting of hardware cost and assembly time was very close to our original estimates for this task in spite of severe price inflation This was due largely to the fact that we were more efficient in assembly than originally estimated Having several WVRs under construction meant that some jobs could be done on a "production line" basis and there were always lots of little jobs that kept everyone busy

The software part of the WVR task cost a great deal more than we originally estimated There were essentially three reasons for this overrun First, purchase and delivery of the JPL MK III data acquisition system was delayed, which meant that we had to purchase additional hardware for host computers other than the MK III computer These additional microcomputers were intended to serve as "host" machines and were not in our original budget In order to save dollars we purchased lowest-bidder type of equipment, which was a serious mistake We saved a few dollars on the hardware but spent much more later in maintenance costs and schedules that had to be slipped because equipment was not working The second reason for the software overrun was due to the fact that we simply underestimated the amount of system-level support we would need to keep several development systems running The third and major reason for the overrun was due to the personnel turnover among the people working on the software Over the three years of the task there were eight people who worked on the software or algorithm development All worked part time and five left after several months on the job – an extremely inefficient way to develop software

Had we realized at the time we started building these instruments that we would construct eight of them, we probably would not have used the conventional Dicke design One can imagine a spectrum of design possibilities that range from a fully compensated Dicke radiometer utilizing full digital control and data acquisition to a single-horn total power instrument The goal in any redesign effort should be to lower the parts count, thereby decreasing the hardware cost, reducing assembly time, lowering weight, size and power requirements, and increasing instrumental stability At some point in the design spectrum one is inevitably faced with the possibility of cost/performance tradeoff that can limit the applicability of the instrument It is our educated guess that the cost of the WVR could be reduced by 15-30% by simply reengineering the current design but keeping the same overall characteristics of the current instruments Potential cost reductions of a factor of 2 are possible but entail some technical risk that would require a design study at the least and possibly the construction and testing of a prototype

Acknowledgments

It is a pleasure to acknowledge the assistance of colleagues who, at various times, have contributed to this task Pete MacDoran, Don Trask, and Jim Johnston provided support and encouragement at critical times Bruce Gary and Joe Waters have generously shared their insight into the techniques of remote sensing Bob Clauss has made valuable comments regarding both the real and the theoretical behavior of the instrumentation Al Banisch handled the engineering on the first WVR (Unit R01) Dave Peterson has helped with everything from fabrication to calibration Rick McKinney performed a valuable service with his part of the testing and the fix for some of the problems that were uncovered Bill Sinclair wrote the first two versions of the controller software and helped with some of the system support software Kwok Ong wrote the nucleus of the data acquisition software and helped with system support and data analysis At different times, Scott Claflin, Robin Hastings, and Henry Fliegel have provided much needed help with the algorithm development Bob Miller helped with software development and hardware testing and calibration

References

1 Smith, E K , and Weintraub, S , "The Constants in the Equation for Atmospheric Refractive Index at Radio Frequencies," *Proc IRE*, V41, pp 1035-1057, 1953

2 Gold, T , "Radio Method for the Precise Measurement of the Rotation Period of the Earth," *Science*, V167, p 302, 1967

3 Shapiro, I I , and Knight, C A , "Geophysical Applications of Long Baseline Interferometry," in *Earthquake Displacement Fields and the Rotation of the Earth* (Mansinha, Smylie, & Beck, eds), Reidel Publ , Holland, pp 284-301, 1970

4 Ong, K M , MacDoran, P F , Thomas, J B , Fliegel, H F , Skjerve, L J , Spitzmesser, D J , Batelaan, P D , and Paine, S R , "A Demonstration of a Transportable Radio Interferometric Surveying System with 3-cm Accuracy on a 307-m Baseline," *J Geophys Res* , V81, pp 3587-3593, 1976

5 Winn, F B , Wu, S C , Resch, G M , Chao, C C , von Roos, O H , and Lau, H , "Atmospheric Water Vapor Calibrations Radiometer Technique," *DSN Progress Report 42-32*, Jet Propulsion Laboratory, Pasadena, Calif , pp 38-49, Apr 15, 1976

6 Moran, J M , and Penfield, H , "Test and Evaluation of Water Vapor Radiometers and Determination of Their Capability to Measure Tropospheric Propagation Path Length," SAO final report on Contract NAS5-20975 (June 1976) available from National Technical Information Service, Springfield, Va , 1976

7 *Atmospheric Water Vapor*, ed Deepak, Wilkerson, and Ruhnke, Academic Press, New York, 1980

8 *Remote Sensing of the Troposphere*, ed V E Derr, U S Government Printing Office, Washington, D C , 1972

9 Staelin, D H , "Measurements and Interpretation of the Microwave Spectrum of the Terrestrial Atmosphere near 1-cm Wavelength," *J Geophys Res* , V71, pp 2875-2882, 1966

10 Westwater, E R , "The Accuracy of Water Vapor and Cloud Liquid Determination by Dual-Frequency Ground-Based Microwave Radiometry," *Radio Science*, V13, pp 677-685, 1978

11 Tiuri, M E , "Radio-Telescope Receivers," in *Radio Astronomy* (by J D Kraus), McGraw-Hill Book Co , New York, 1966

12 Rogers, A E E , "Very Long Baseline Interferometry with Large Effective Bandwidth for Phase-Delay Measurements," *Radio Science*, 5, pp 1239-1247, 1970

13 Resch, G M , "Water Vapor – The Wet Blanket of Microwave Interferometry," in *Atmospheric Water Vapor* (Deepak, Wilkerson, and Ruhnke, eds), pp 265-282 Academic Press, New York, 1980

14 Westwater, E R , "An Analysis of the Correction of Range Errors Due to Atmospheric Refraction by Microwave Radiometric Techniques," ESSA Technical Report IER 30-ITSA30, available from National Technical Information Service, Springfield, Va , 1967

15 Wu, S C , "Optimum Frequencies of a Passive Microwave Radiometer for Tropospheric Pathlength Correction," *IEEE Trans Ant & Prop* , AP-27, pp 233-239, 1979

16 Janssen, M A , Bednarczek, S M , Gulkis, S , Marlin, H W , and Smoot, G F , "Pattern Measurements of a Low-Sidelobe Horn Antenna," *IEEE Trans Ant & Prop* , AP-27, pp 551-555, 1979

17 Guiraud, F O , Howard, J , and Hogg, D C , "A Dual-Channel Microwave Radiometer for Measurement of Precipitable Waver Vapor and Liquid," *IEEE Trans Geoscience Elect , GE-17*, pp 129-136, 1979

18 Clark, B G , "The Effect of Digitization Errors on Detection of Weak Signals in Noise," *Proc IEEE*, V61, pp 1654-1655, 1973

Table 1. Water vapor radiometer specifications

(1) Physical characteristics	Weight, kgm	Volume, m^3	Power, W
Microwave package	45	0 12	300–600
Positioner	34	0 074	0–450
Control module	26	0 15	120
(2) Environmental			
Operating range			
Ambient temperature	–10 to +50°C		
Wind speed	up to 65 km/hr		
Relative humidity	0 to 95% or until liquid water precipitates onto the horn cover		
Survivability			
Ambient temperature	–40 to +60°C		
Wind speed	up to 160 km/hr in stowed position		
Relative humidity	100%, sealed from rain and dust		
(3) Microwave module			
Operating frequencies	20 7 and 31 4 GHz		
Frequency stability	±12 MHz over operating range		
RF bandwidth	200 MHz		
IF bandwidth	100 MHz		
Integration time	software selectable in steps of 0 1 sec		
Signal range	3 to 400 K		
Antenna beamwidth	7° (full width at half power)		
Beam efficiency	>99 9% for ±15° around beam		
Calibration	center supplied by two waveguide terminations held at fixed temperatures and supplemented with tip curve observations		
Base load	315 K		
Hot load	370 K		
Temperature stability	>±0 1 K/min		
(4) Positioner			
Coordinates	azimuth and elevation		
Range	0 to 355° in Az, 5 to 175° in El		
Slew rate	> 1 5 deg/sec (both axis)		
Position accuracy	1° both axes		
Readout accuracy	0 1° both axes		
(5) Control and interface module			
Operating modes			
(a) Local	all data and control lines available for monitoring by built-in digital voltmeter		
(b) Remote			
Interface	RS-232, twisted pair of cables		
Baud rate	110 to 19200 baud, selectable by software and hardware jumpers		
Protocol	compatible with MAT[a] bus		
Data storage	temporary storage only, mass storage is the responsibility of the host computer		

[a]MAT = Microprocessor ASCII Transceiver used as the control/communicator standard in the HaystackGSFC Mark III data acquisition system.

Table 2. Extent of the near-field and beam volume ratios for WVR and radio telescope

Diameter, m	Frequency, GHz	L_{nf}, km	R	$\sqrt{R}$
64	8 4	108 00	29 2	5 4
64	2 3	32 00		
40	8 4	42 00	74 8	8 6
40	2 3	12 30		
25	8 4	16 40	191.5	13 8
25	5 0	10 40		
25	2 3	4.80		
9	8 4	2 10	1477 9	38 4
9	2 3	0 60		
4	8 4	0 40		
4	2 3	0 01		

Table 3. Cost summary

A Cost per radiometer (as of Oct 1981)

Components	K$	Tasks	Man-month
Ferrite switches	7 3	Fabrication	3 0
LO/mixer/I F	8 5	Assemby and wiring	4 0
Horn antennas	3 1	Bench testing	3 0
Waveguide components	1 0	Engineering tests	0 8
Power supply	1 7	Calibration	2 0
Detectors	0 1		12 8
Temp sensors and control	0 4		
Enclosure (microwave)	0 6		
Misc components	1 2		
Control panel	0 8		
Microprocessor	4 5		
Positioner	2 8		
Cables and connectors	0 5		
	$32,500 0		

B Support tasks

Task	Man-years
Controller software (3 versions)	1 2
System support	0 8
Data acquisition S/W	1 0
Environmental test	0 2
Algorithm development	2 0
	5 2 MY

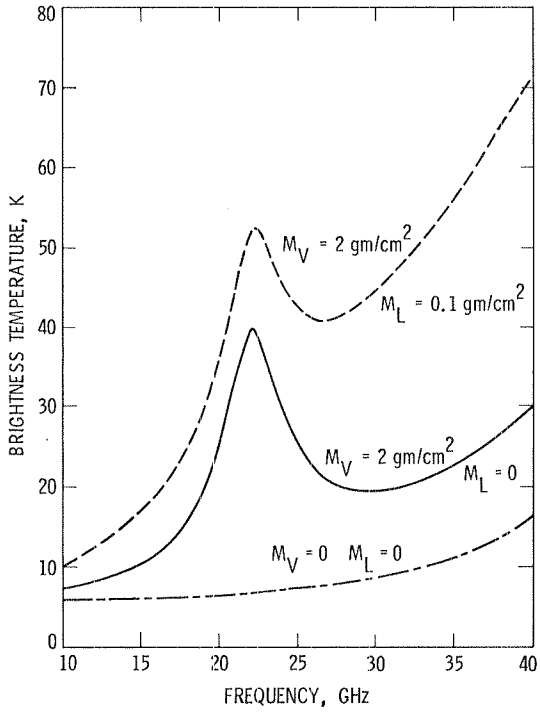


Fig. 1. Zenith brightness temperature of the atmosphere between 15 and 40 Ghz

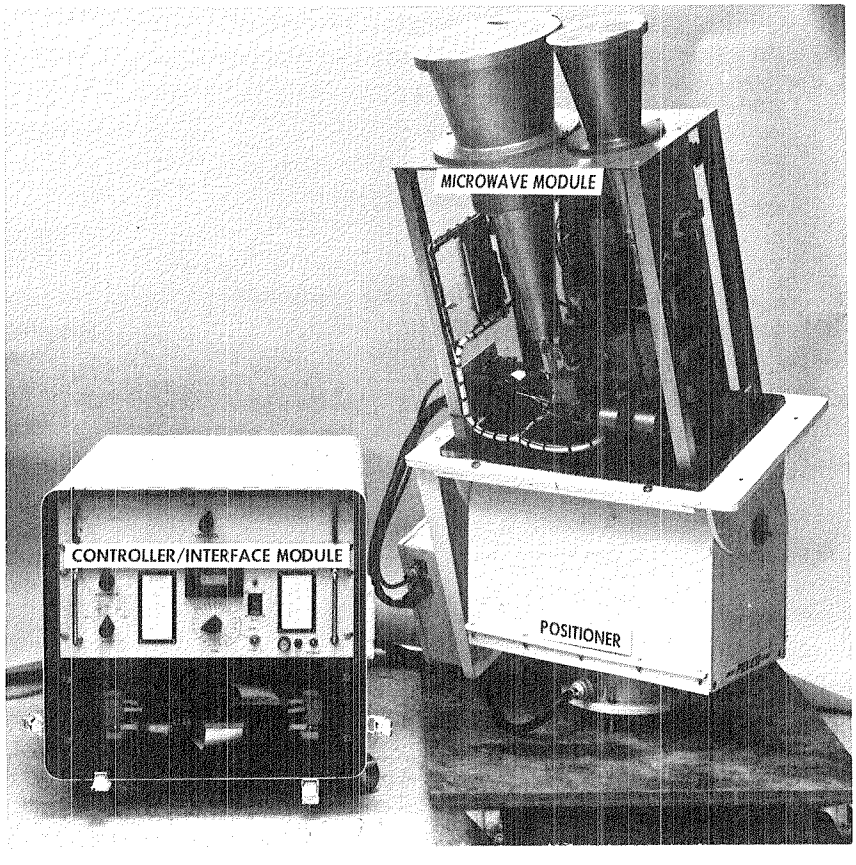


Fig. 2. Photograph of the three WVR modules (cover removed from the microwave package)

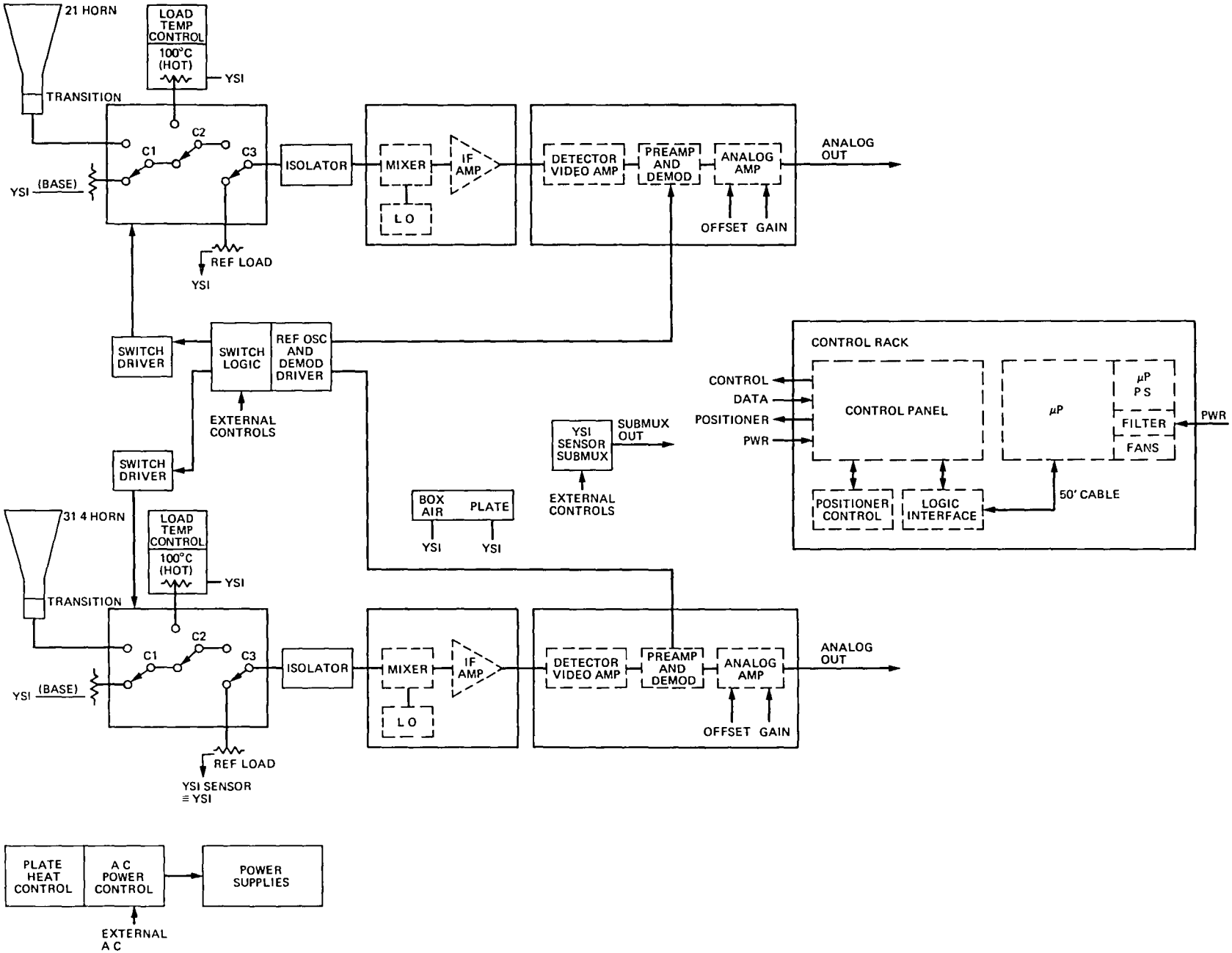


Fig. 3. Block diagram of the WVR

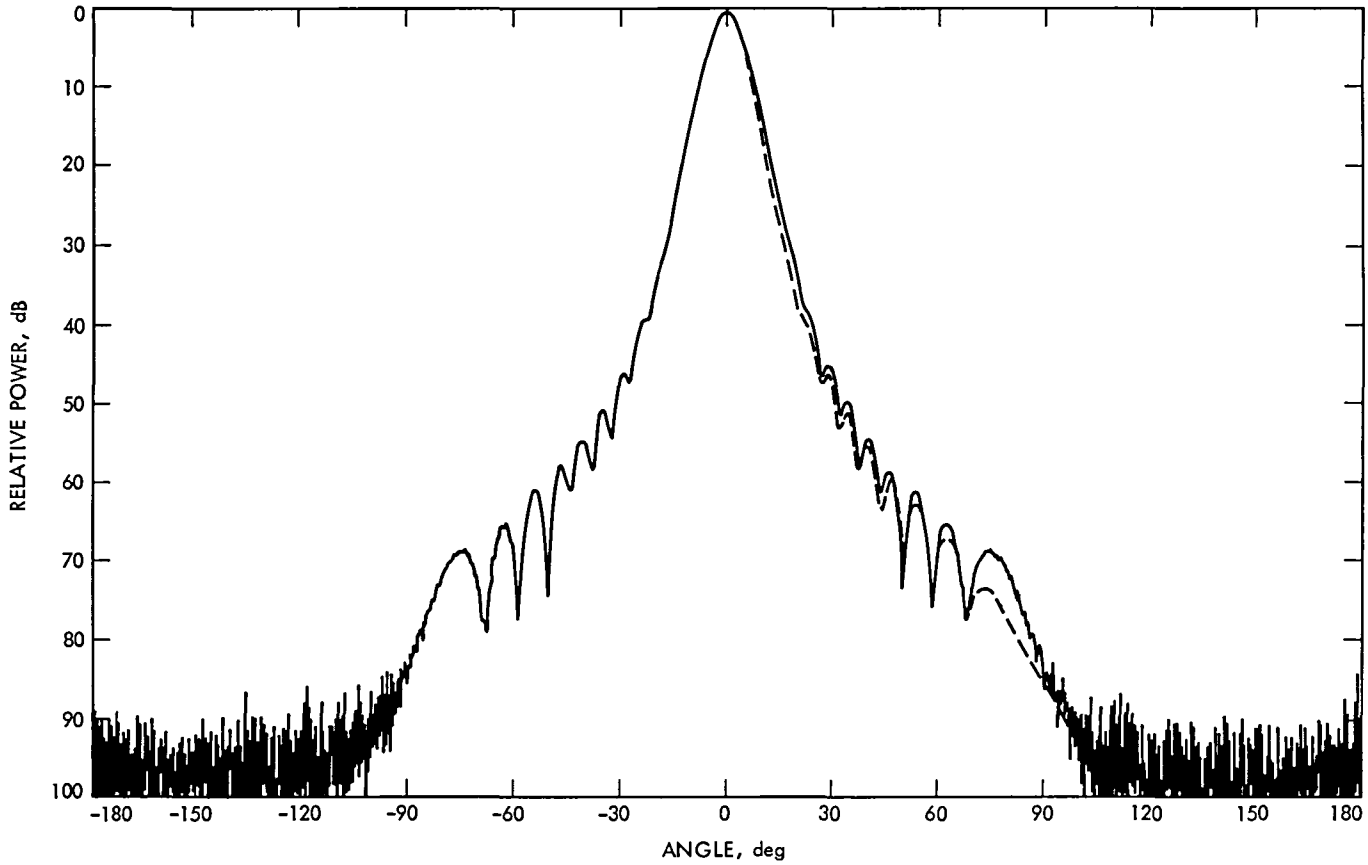


Fig 4 Antenna beam pattern of the WVR horn antenna

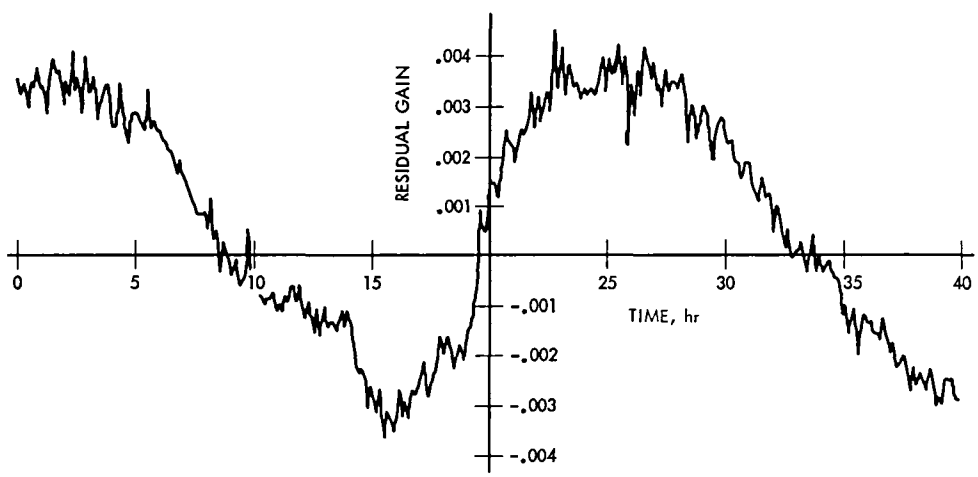


Fig 5 Gain of one of the WVR channels vs time

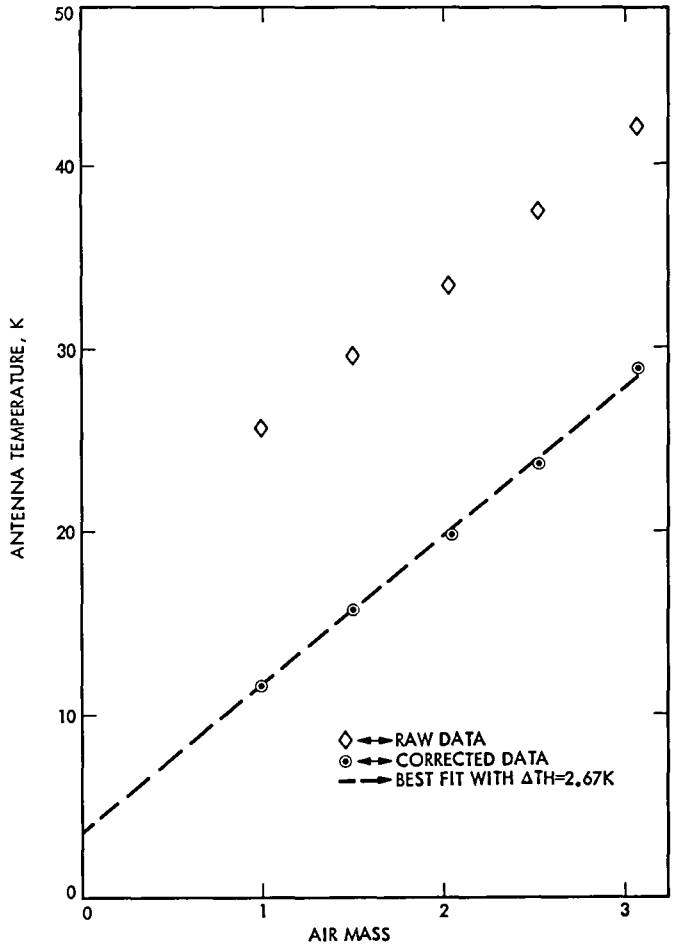


Fig. 6. Tip curve determination of the hot load correction

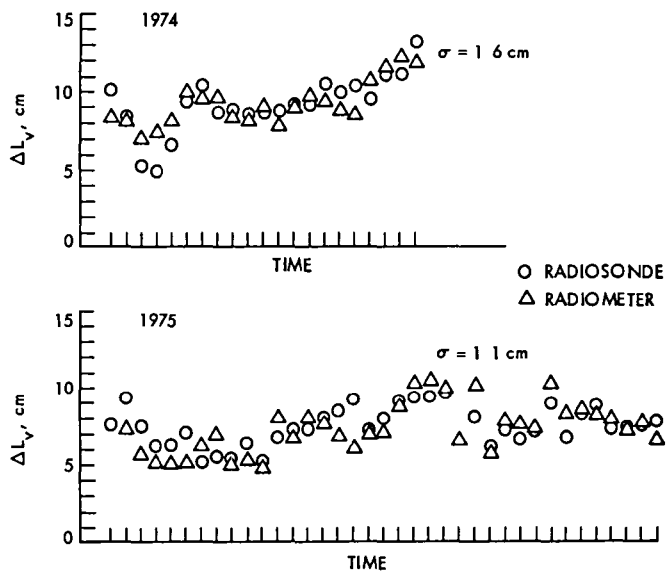


Fig 7. Comparison of path delay as determined by radiosonde and WVR (El Monte, 1974 and 1975)

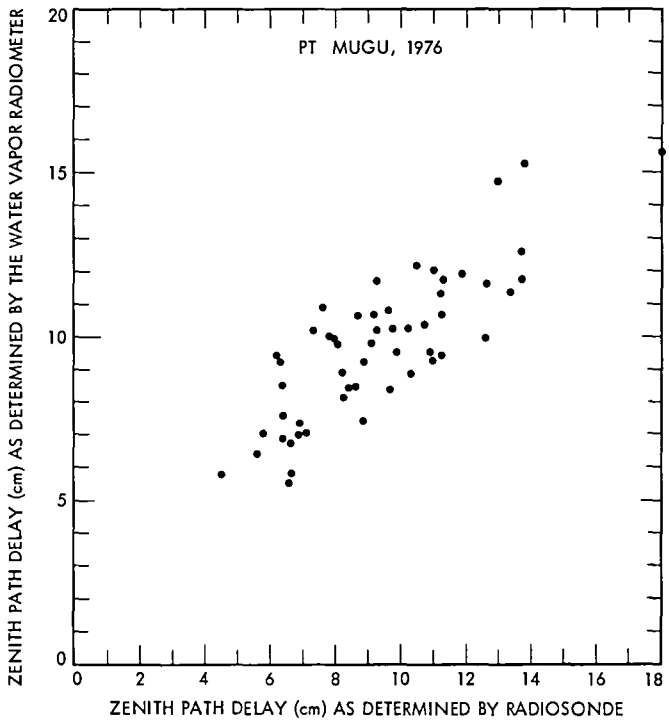


Fig. 8 Comparison of path delay as determined by radiosonde and WVR (Pt. Mugu, California, 1976)

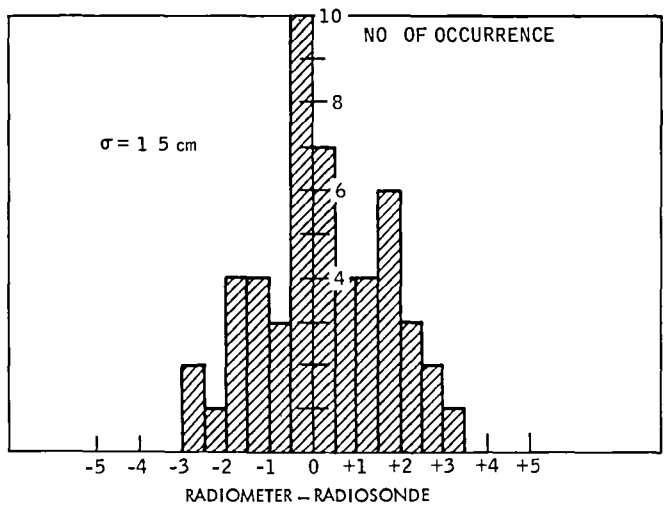


Fig 9. Histogram of path delay residuals from the Pt. Mugu data

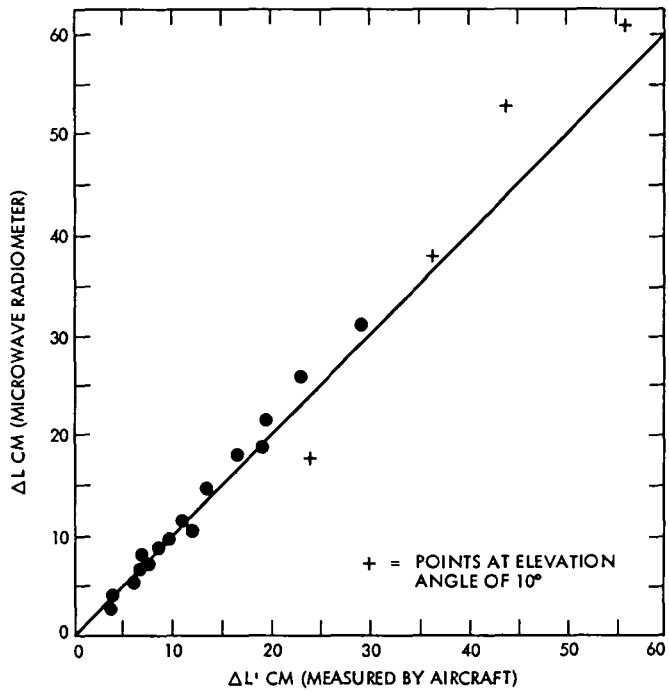


Fig. 10 Comparison of path delay as determined by an instrumented aircraft and the WVR (Pt. Mugu)

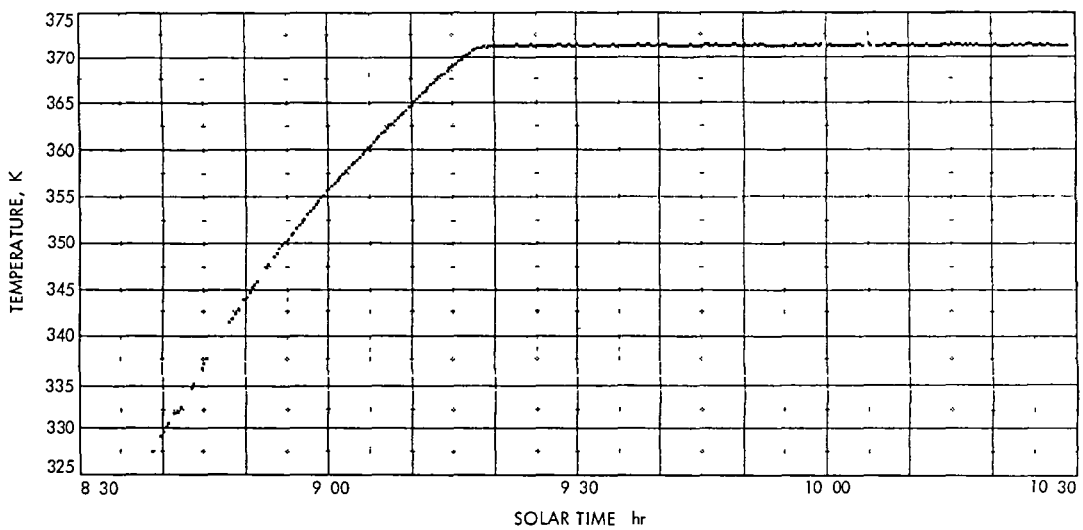


Fig. 11 Warmup characteristics of the hot load

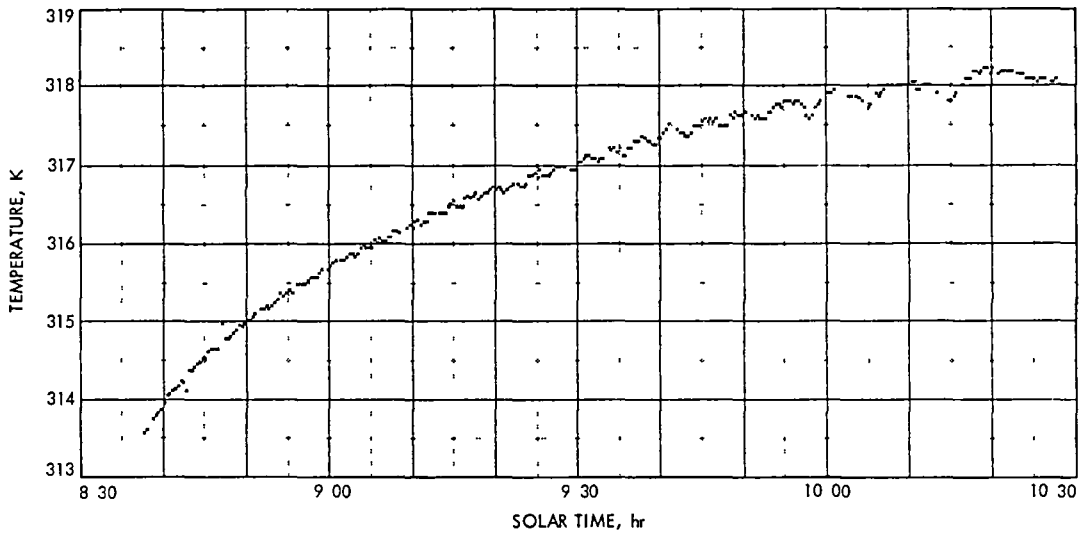


Fig. 12. Warmup characteristics of the base load

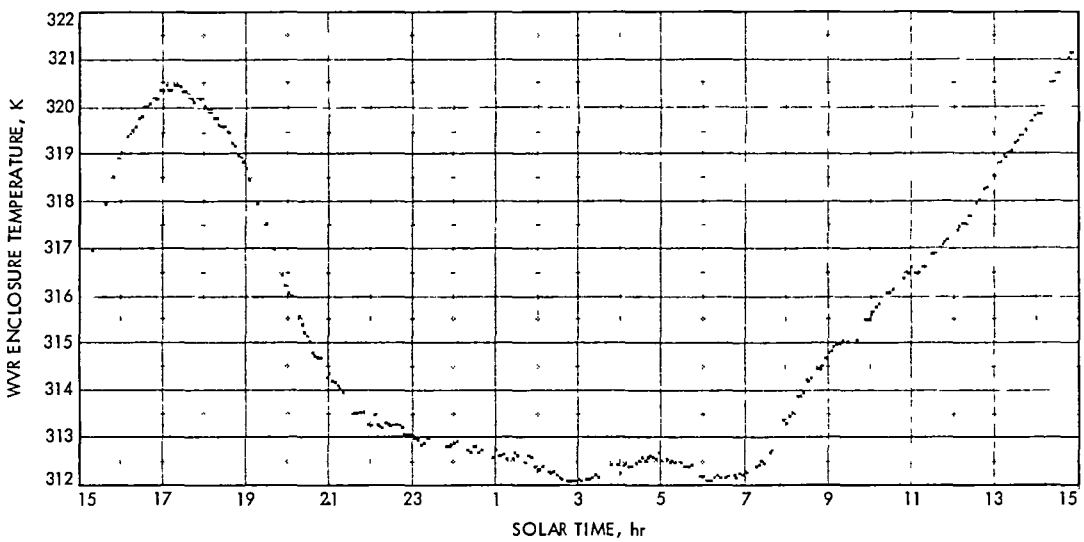


Fig 13. Enclosure temperature vs time

VLBI Collimation Tower Technique for Time-Delay Studies of a Large Ground Station Communications Antenna

T Y Otoshi
Radio Frequency and Microwave Subsystems Section

L E Young
Tracking Systems and Applications Section

W V T Rusch
Department of Electrical Engineering
University of Southern California

A need for an accurate but inexpensive method for measuring and evaluating time delays of large ground antennas for VLBI applications motivated the development of the collimation tower technique described in this article This article discusses supporting analytical work which was performed primarily to verify time delay measurement results obtained for a large antenna when the transmitter was at a collimation distance of 1/25 of the usual far-field criterion Comparisons of theoretical and experimental results are also given

I. Introduction

As accuracies of Very Long Baseline Interferometry (VLBI) systems for the measurement of time delay changes have continued to improve to current levels of less than 0.1 ns, a need has existed for studies to be made of accuracy-limiting errors associated with the microwave antenna itself Two types of VLBI errors attributed to the antenna are (1) errors produced by unwanted multiple reflections (multipath signals) within the antenna propagation media and (2) undesired phase delay changes produced when the antenna is mispointed from boresight of the radio source

The subject of antenna time delay and errors contributed by multipath signals has been studied in recent years both analytically and experimentally (Refs. 1-3) Results of studies made on some types of Deep Space Network (DSN) antennas tend to show that multipath problems can be mitigated by increasing the VLBI spanned bandwidth

The second type problem, which is concerned with phase delay errors due to antenna mispointing, is one to which little attention has been given in the past This subject is important because antenna mispointing can occur in practice due to intentional mispointing as in the case of conscanning VLBI systems, or unintentionally due to practical limitations on the accuracies of antenna pointing systems Although some far-field VLBI antenna mispointing experiments have been performed with a 64-m DSN antenna, poor reproducibility of the results and insufficient experimental data have thus far prevented any definite conclusion to

be drawn concerning the effects of antenna mispointing on VLBI time delay data There has been a strong need for quite some time now for a rigorous theoretical analysis to help predict the order of magnitude of phase delay changes that can occur for various VLBI type antennas when the mainbeam is pointed off the boresight coordinates of the radio source.

This article is concerned with supporting analytical work which was performed primarily to verify time delay measurement results obtained for a large antenna when the transmitter was at a collimation distance of 1/25 of the usual far-field criterion The unexpected spinoff from this work was that since a collimation tower geometry is basic to many experimental configurations, the same theoretical equations can also be used to study phase delay changes as a function of antenna pointing offset angles from boresight The analysis is applicable, not only for near-field but also for far-field studies

II. Theory

Application of the Lorentz Reciprocity Theorem yields the voltage coefficient of mutual coupling between two lossless, matched antennas to be (Ref. 4).

$$S_{12} = \frac{e^{j\Phi}}{4\sqrt{P_1 P_2}} \int_S (\bar{E}_1 \times \bar{H}_2 - \bar{E}_2 \times \bar{H}_1) \cdot \hat{n}\, dS \tag{1}$$

where Φ is an arbitrary phase reference, $(\bar{E}_1, \bar{H}_1)$ and $(\bar{E}_2, \bar{H}_2)$ are the transmitted fields of each antenna, P_1 and P_2 are the radiated powers, and S completely encloses one of the antennas This statement of Robieux's Theorem can be rewritten

$$S_{12} = \frac{e^{j\Phi} \int_S (\bar{E}_1 \times \bar{H}_2 - \bar{E}_2 \times \bar{H}_1) \cdot \hat{n}\, dS}{4\left[\frac{1}{2} Re \int_{A1} (\bar{E}_1 \times \bar{H}_1^*) \cdot \hat{n}\, dS\right]^{1/2} \left[\frac{1}{2} Re \int_{A2} (\bar{E}_2 \times \bar{H}_2^*) \cdot \hat{n}\, dS\right]^{1/2}} \tag{2}$$

which, when squared, yields

$$\frac{P_R}{P_T} = |S_{12}|^2 = \frac{\frac{1}{4}\left| \int_S (\bar{E}_1 \times \bar{H}_2 - \bar{E}_2 \times \bar{H}_1) \cdot \hat{n}\, dS \right|^2}{\left[Re \int_{A1} (\bar{E}_1 \times \bar{H}_1^*) \cdot \hat{n}\, dS\right]\left[Re \int_{A2} (\bar{E}_2 \times \bar{H}_2^*) \cdot \hat{n}\, dS\right]} \tag{3}$$

In the event that one or both of the antennas does not have a clearly defined aperture, Eq. (1) can be rewritten

$$S_{12} = \frac{e^{j\Phi}}{4\sqrt{P_1 P_2}} \int_V (-\bar{E}_2 \cdot \bar{J}_1 + \bar{H}_2 \cdot \bar{J}_{m1})\, dV \tag{4}$$

where $\bar{J}_1$ and $\bar{J}_{m1}$ are the source currents completely enclosed by S and $\hat{n}$ (in Eq 1) points *out* of V

As a relatively simple application of Eqs (1)-(4), consider a paraboloid and an on-axis electric dipole (Fig 1) The paraboloid is fed by a focused, point-source feed located at its focus The fields of the feed are

$$\bar{E}_f = V_0\, e^{-j\beta L}\, \frac{e^{-jk\rho}}{\rho} \left(\frac{2F}{1+\cos\theta_f}\right) (\hat{a}_{\theta_f} \cos\phi_f - \hat{a}_{\phi_f} \sin\phi_f) \tag{5}$$

Located at $P(0, 0, R)$ on the Z-axis is an electric dipole $\hat{a}I_0 d$ Application of Eq (4) over the paraboloid aperture then yields

$$\frac{P_R}{P_T} = \frac{G_d G_p}{\left(\frac{4\pi R}{\lambda}\right)^2} \tag{6}$$

where the dipole gain is

$$G_d = 1.5 \tag{7a}$$

and the Fresnel-region gain of the paraboloid is

$$G_p = \left(\frac{\pi D}{\lambda}\right)^2 \left\{\frac{\sin\left[k\,(D/2)^2/4R\right]}{\left[k\,(D/2)^2/4R\right]}\right\}^2 \tag{7b}$$

where D is the paraboloid diameter.

Equations (1)-(3) can be applied to the problem of coupling between two paraboloids (Fig 2)

(1) Test paraboloid #1 is a large reflector with a point-source feed located at its focus The feed pattern consists of arbitrary E- and H-plane patterns Central blocking is included, but strut blocking is not This reflector lies in the far-field of paraboloid #2

(2) Collimation paraboloid #2 is a smaller reflector with an axially-defocussed point-source feed Various feed patterns may be specified Central blocking is included, but strut blocking is not This reflector need not lie in the far-field of paraboloid #1, and it can be oriented arbitrarily with respect to #1

Voltage coupling between the two reflectors is computed by means of Eq (1) and the corresponding power coupling by means of Eq (3) The integration surface S is a large coaxial circle located in front of paraboloid #1 and perpendicular to its axis This surface is generally near to #1 and sufficiently large that the fields $\bar{E}_1$ and $\bar{H}_1$ are assumed to be negligible beyond the limits of the circle. In principle, S_{12} should be independent of the location of this surface, in practice the size and location of the circle must be determined empirically The fields $\bar{E}_1$ and $\bar{H}_1$ on S are determined by geometrical optics and geometrical edge diffraction (Refs 5, 6)

The integration surface S is assumed to lie in the far-field of collimation paraboloid #2 so that the fields of this antenna may be expressed

$$\bar{E}_2 = \frac{e^{-jkR}}{R}\left[E_{2a}(\theta_{2a})\cos\phi_{2a}\,\hat{a}_{\theta_{2a}} - H_{2a}(\theta_{2a})\sin\phi_{2a}\,\hat{a}_{\phi_{2a}}\right] \tag{8}$$

which can be decomposed into rectangular components E_{2x}, E_{2y}, H_{2x}, and H_{2y} for integration in Eqs (1) and (3) The above formulations are contained in an appropriate computer program.

Provision is also made to rotate paraboloid #1 (Fig 3) about an offset azimuth axis (Δz_{az}) and an offset elevation axis (Δx_{el}, Δz_{el}) In the rotated coordinate system the integration surface S rotates with #1 so that it remains symmetrically positioned on a plane normal to the axis of #1. However, it is no longer symmetric or normal with respect to the line between #1 and #2

III. Experimental Setup

Figure 4 shows a block diagram of the collimation tower experimental setup for time delay measurements using VLBI measurement techniques and associated equipment At the collimation tower, a simulated VLBI radio source signal, provided by a broadband noise source, is fed into a power divider One of the power divider outputs is a transmit-reference signal which is recorded

on a magnetic tape recorder simultaneously with clock signals furnished by a portable rubidium frequency standard The other power divider output is fed to the collimation tower transmitter antenna which is 1 83 m (6 ft) in diameter The signal is transmitted over a distance of about 1 6 km and illuminates both a 1 83-m (6-ft) antenna and a 26-m (85-ft) antenna The smaller antenna is a receive-reference antenna whose time delay (including feed and interconnecting transmission lines) has been precalibrated while the 26-m antenna is the antenna whose delay is being measured

For this test, the center frequency was 8 440 GHz and the VLBI spanned bandwidth 40 MHz When boresighted to each other, the distances between the focal points of the transmit and receive antennas were approximately 0 08 D^2/λ and 16 D^2/λ, respectively, for the 26-m and 1 83-m receive antenna paths where D corresponds to the applicable receive antenna diameter and λ is the free space wavelength Figure 5 shows the collimation tower at DSS 13 and the 1 83-m transmit antenna used in the test, while Fig 6 shows the 26-m antenna (being calibrated) and the receive-reference 1 83-m antenna referred to in this article

The experimental procedure was to first record the received signal for a duration of 2 minutes for the receive-reference antenna and then switch to the 26-m antenna under test and record its received signal for 2 minutes and then switch back to the reference antenna This procedure was repeated for a period of 1 hour during which the downconverted received signals were recorded on magnetic tape. Simultaneously recorded were clock signals provided by the station hydrogen maser frequency standard

The recorded data was later postprocessed on the Caltech-JPL VLBI correlator where received signal recordings were correlated against the transmit-reference signal data recorded at the collimation tower The final experimental value was determined after correcting the VLBI correlator output data for the precalibrated delays of the receive-reference antenna and all interconnecting transmission lines and waveguides of the receive-reference antenna and the 26-m antenna up to the common switching port.

The advantages of switching between a reference antenna and the antenna under test is that common modes changes such as propagation media changes, multipath, and system drifts tend to cancel out It is reasonable to expect that better accuracy and precision can be obtained from comparisons made with a calibrated delay standard than from direct absolute delay measurement over long distances

IV. Comparisons of Theoretical and Experimental Data

The basic theory of antenna time delay for Cassegrain antennas has been discussed in Ref. 1 To compute the theoretical value of antenna delay for the results of this paper, the output phase of integrated fields received by the antenna is computed at the two microwave frequencies involved in the VLBI experiment (resolving any multiples of 2 pi radian ambiguities) Then group delay is computed by the "VLBI bandwidth synthesis" equation (Ref 7) which is equivalent to the "phase-slope" method of determining group delay (Ref. 8)

One of the significant results of this experiment was that a measurement precision of 10 ps (corresponding to 0 3 cm) was achieved. A second important result was that a significant difference was found between the theoretical 26-m antenna delay determined from mutual coupling theory and the value calculated from the optical ray pathlength of a single ray traveling from the collimation tower antenna to the feed horn phase center of the 26-m antenna The near-field delay, based on mutual coupling theory described in Section II of this article, was calculated to be 0 6 ns longer than the value determined from the ray optical pathlength calculation A third significant result was that the experimentally determined delay of the 26-m antenna agreed with the mutual coupling theoretical value to within 0 2 ±0 2 ns probable error The agreement between mutual coupling near-field and experimental results is excellent when taking into consideration the relatively small VLBI spanned bandwidth of the 40 MHz that was used in the experiment.

It should be pointed out that the ray pathlength calculation method is only approximately valid for a focused antenna and at far-field distances, but is a method commonly used erroneously for near-field geometries as well The mutual coupling theoretical result is based on numerical integrations over the entire 26-m-dia aperture using subreflector pattern data (Ref 9) applicable to the X-band feed system that was installed on the 26-m Cassegrain antenna during the experiments Since the subreflector pattern is referenced to the focal point of the paraboloid, it is applicable to the theory and computer program described in this paper

Many collimation tower test setups are plagued with ground multipath problems However, a previous study made by the Howland Co of Atlanta, Ga , showed that for the particular ground terrain involved in this collimation tower setup and with the receive-reference antenna mounted on top of the 26-m antenna as shown in Fig 6, ground reflected multipath signals would not interfere with accurate measurements of the delays of the two receive antennas In addition, independent time delay measurements made on the 26-m antenna with a time domain technique similar to the FM/CW technique described in Ref 10 confirm the VLBI results discussed in this paper With the time domain technique, it was possible to identify and separate out the ground multipath signals and measure the delay of the primary signal accurately Although time domain techniques can also be used to measure time delays, the FM/CW instrumentation is primarily a diagnostics tool and is not part of the standard DSN instrumentation

Figure 7 shows a plot of power pattern for an elevation angle scan for a test frequency of 8 420 GHz and at the collimation tower geometry and distances described above for the transmit 1 83-m antenna and the receive 26-m antenna Both theoretical and experimental data are shown The theoretical pattern is based on uniform illumination of the 26-m antenna aperture Later comparisons will be made where the theoretical pattern is based on subreflector pattern data (Ref 9) The experimental data was interpolated from a linear power pattern recording which had dB calibration marks provided through the use of a rotary vane attenuator in series with the transmitter waveguide Since the data had to be extracted from the linear power recordings, it is estimated that the experimental data shown in Fig 7 is accurate to only ±0 5 dB above -6 dB, ±1 0 dB between -6 dB and -10 dB, and ±1 5 dB below -10 dB It is significant to note in the power pattern that both experimental and theoretical results show three bumps on the main beam Also note that the main beam beamwidths between the -10 dB points are close to 0 8 degree for both mutual coupling theory and experimental results at this near-field distance of 0 08 D^2/λ For comparison purposes, when the separation of the transmit antenna and the 26-m receive antenna changes from this near-field distance to far-field distances, the beamwidth between -10 dB points changes from the 0 8 deg value to about 0 15 deg This far-field beamwidth value was determined by performing numerical integration of the X-band subreflector pattern over the 26-m antenna aperture at 8 420 GHz and computing far-field patterns

Also of interest is the value of P_{R1}/P_{T2} calculated from mutual coupling theory (Eq 3) as compared to the value calculated from Friis transmission loss formula (Ref 11) For the near-field collimation tower distance involved in the experiment, the 26-m antenna received power level was calculated from mutual coupling theory (and use of Ref 9 subreflector data) to be smaller than the received power calculated from the Friis formula by 13 6 dB The experimentally measured received power was less than the Friis value by 13 5 ± 1 5 dB The large uncertainty associated with the measured value is due to uncertainty in the efficiency of the large antenna at 8 420 GHz during earlier installation periods, and uncertainty as to whether the receive and transmit antennas were optimally boresighted to each other at the particular time of this measurement Despite the large uncertainty in this particular experimental result, it is significant that the agreement with the mutual coupling theoretical value is still quite good

At the time that initial boresighting and power measurements were being made, it was thought that the large difference between actual received power and the value based on Friis formula was a major experimental leakage problem Had the theoretical results of this article been available prior to this measurement, a greater understanding of the near-field measurement procedure would have been available and greater care would have been exercised in boresighting procedures Termination of funding for this work and recent conversion to an updated feedcone at DSS 13 make it difficult to repeat this series of experiments in the near future

If at some time in the future, funding does become available, it would be of great interest and value to the DSN to perform additional experimental measurements with the collimation tower technique and make accurate comparisons to theory, especially with regards to studying phase delay change vs offset from boresight angles

V. Conclusion

This article has described the development of theoretical equations specifically developed for practical collimation tower test setup geometries A brief description has been given of the VLBI collimation tower technique that was used for the measurement of antenna delay An agreement of 0 2 ns was obtained between theoretical and experimental values of time delay for the DSS 13 26-m antenna at a collimation tower distance that was 1/25 the accepted $2D^2/\lambda$ far-field criterion These results apply to measurements performed with VLBI equipment at 8 440 GHz and 40 MHz spanned bandwidth

Based on the results presented in this article, one can conclude that the collimation tower technique has the potential of being very useful for checkout and diagnostics of the performance of a large ground station antenna Now that the necessary theoretical equations are available, theoretical values calculated for actual near-field geometries can be used as a criterion for expected performance Once the near-field performance of an antenna is verified experimentally, then transformations can be made on the near-field data to predict performance in the far-field

Acknowledgments

The VLBI collimation tower technique described in this article was suggested by Hugh Fosque of NASA. Measurements of received power vs elevation angle were made by Juan Garnica of Bendix Field Engineering Corporation.

References

1 Cha, A. G , Rusch, W. V. T., and Otoshi, T Y., "Microwave Delay Characteristics of Cassegrainian Antennas, "*IEEE Trans on Antennas and Propagation*," Vol AP-26, pp. 860-865, November 1978

2. Otoshi, T Y , and Rusch, W V. T , "Multipath Effects on the Time Delays of Microwave Cassegrainian Antennas, *IEEE 1980 International Symposium Digest – Antennas and Propagation*, Volume II, 80, CH 1557-8AP, pp 457-460, June 1980.

3. Otoshi, T Y , and Young, L E , "An Experimental Investigation of the Changes of VLBI Time Delays Due to Antenna Structural Deformations," *TDA Progress Report 42-68*, pp 8-16, Jet Propulsion Laboratory, Pasadena, Calif , April 15, 1982.

4 Ludwig, A C , "Fundamental Relationships of Electromagnetics," Technical Report NB 121, Electromagnetics Institute, Technical University, Denmark, pp 36-42, June 1979

5 Lee, S W , "Uniform Asymptotic Theory of Electromagnetic Edge Diffraction," *Electromagnetic Scattering*, P L E Uslenghi, editor, Academic Press, pp 67-119 1978

6 Rusch, W V T , "Antenna Notes," Technical Report NB84b, Electromagnetics Institute, Technical University, Denmark, pp 135-140, Aug 1974

7 Molinder, J. I , "A Tutorial Introduction to Very Long Baseline Interferometry (VLBI) Using Bandwidth Synthesis," *DSN Progress Report 42-46*, Jet Propulsion Laboratory, Pasadena, Calif , pp. 16-28, Aug 15, 1978

8 Otoshi, T Y , and Beatty, R. W , "Development and Evaluation of a Set of Group Delay Standards," *IEEE Trans on Instrumentation and Measurement*, Vol IM-25, No 4, pp. 335-342, December 1976

9 Williams, W., Nixon, D , Reilly, H , Withington, J., and Bathker, D , "A Prototype DSN X-S Band Feed DSS 13 Application Status (Third Report)," *DSN Progress Report 42-52*, Jet Propulsion Laboratory, Pasadena, Calif , pp. 51-60, Aug 15, 1979

10 Otoshi, T. Y , "An FM/CW Method for the Measurement of Time Delays of Large Cassegrainian Antennas," *TDA Progress Report 42-66*, Jet Propulsion Laboratory, Pasadena, Calif , pp 49-59, Dec 15, 1981

11. Friis, H T., "A Note on a Simple Transmission Formula," *Proc IRE*, Vol 34, pp 254-256, May 1946

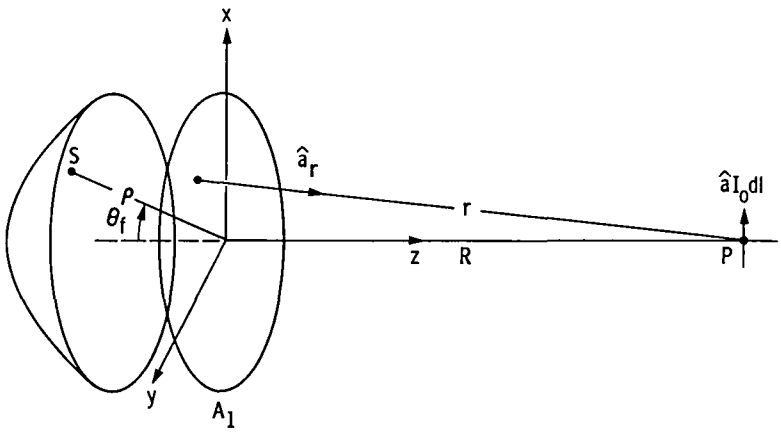


Fig 1. Coupling between paraboloid and on-axis electric dipole

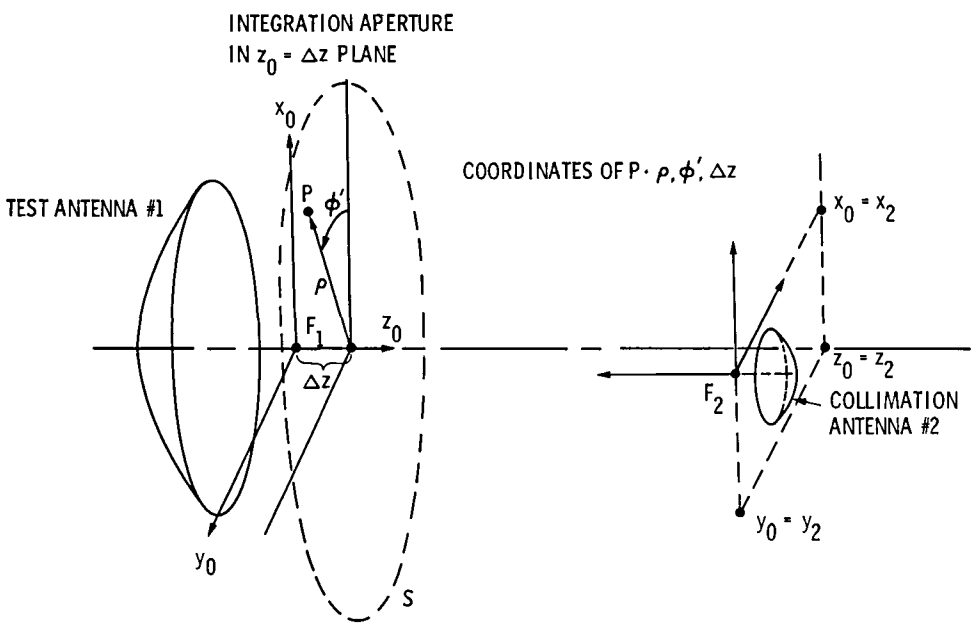


Fig 2. Coupling geometry for two paraboloids

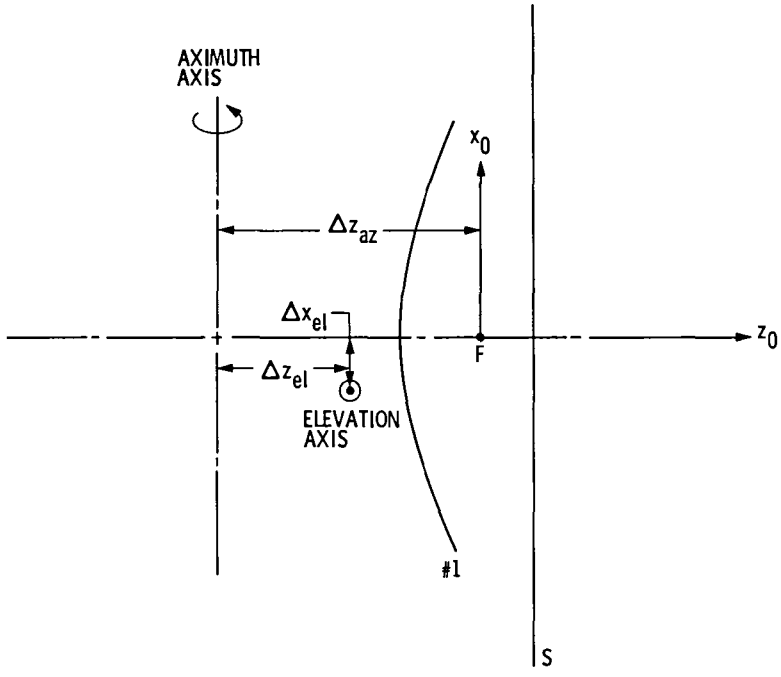


Fig. 3. Geometry of Az-El rotation

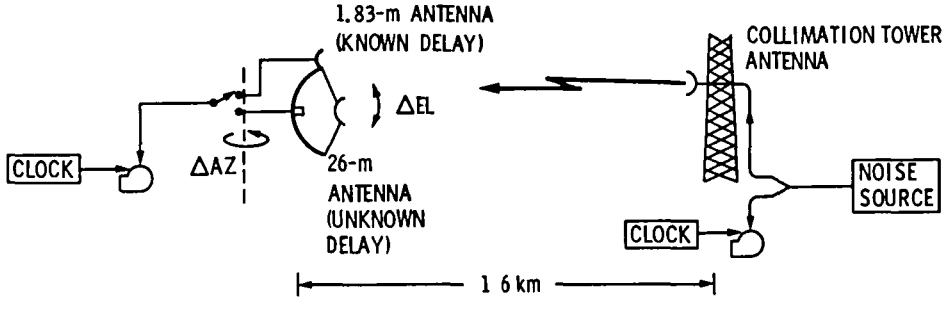


Fig. 4. Collimation tower geometry for the measurement of antenna time delay using VLBI technique and instrumentation

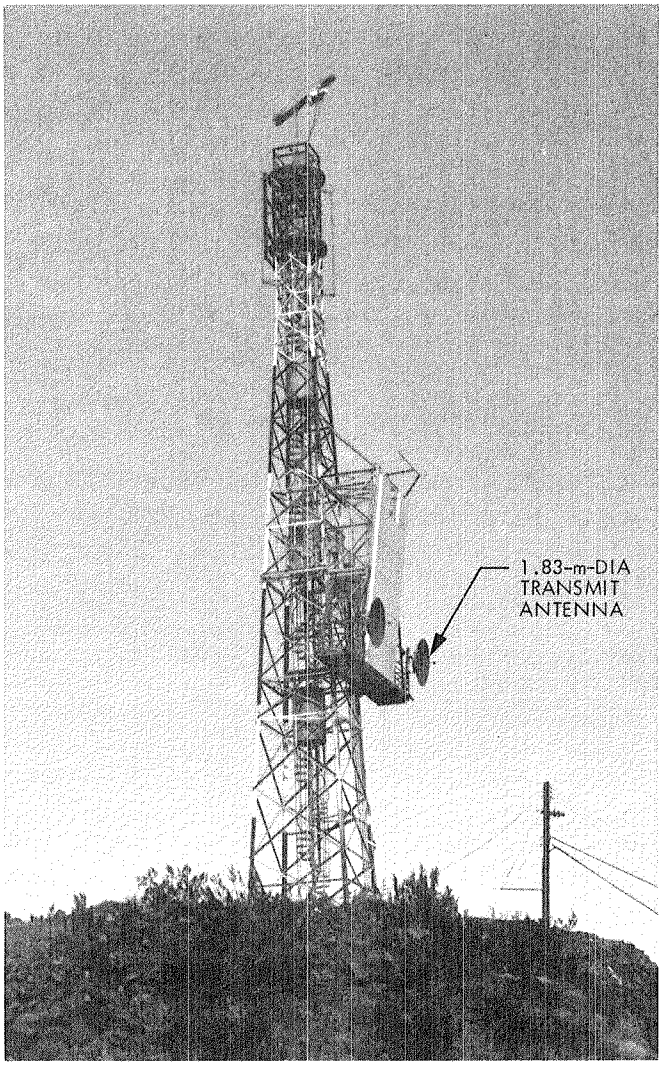


Fig. 5. DSS 13 collimation tower and 1.83-m (6-ft) diameter transmit antenna. The S-band feed shown in photograph was replaced by an X-band feed

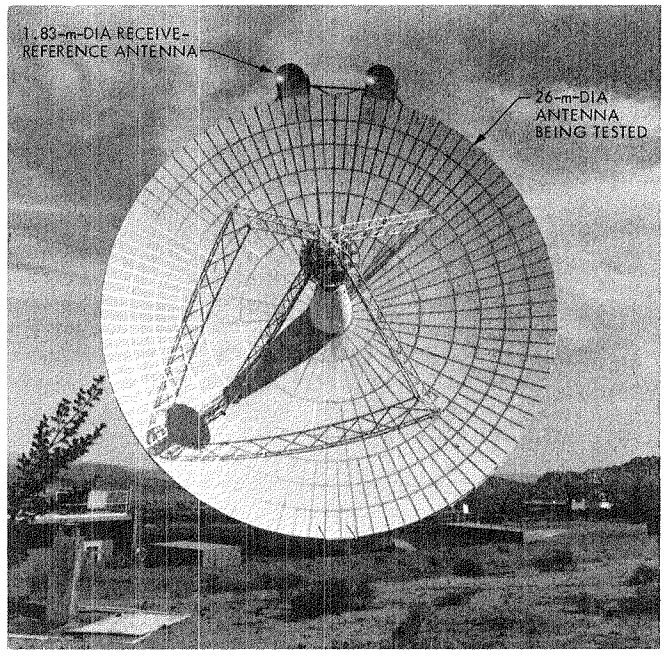


Fig. 6. 26-m-dia. antenna and 1.83-m-dia. receive-reference antenna at DSS 13. The S-band feed of the 1.83-m-dia. antenna shown in photograph was replaced by an X-band feed

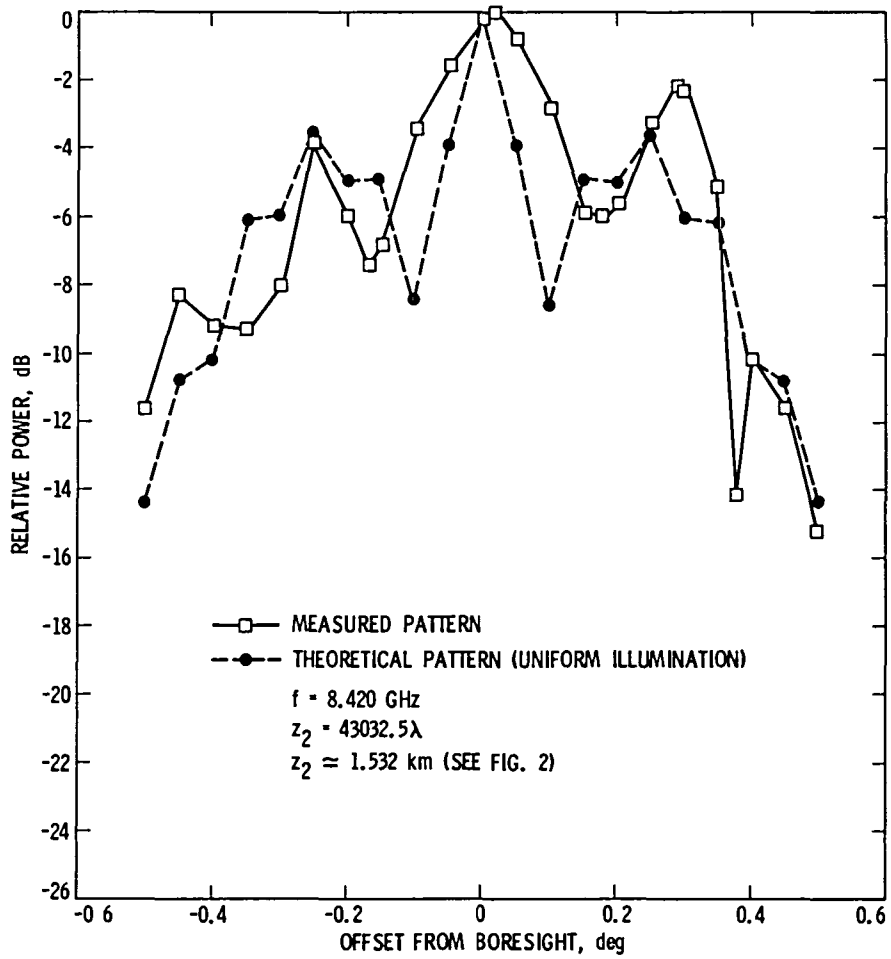


Fig. 7. Received power vs offset angle scanning in the elevation angle plane at 8 420 GHz when the 26-m paraboloidal antenna was located 1 53 km distance from the 1.83-m-dia transmit (collimation tower) antenna

A Near-Optimum Receiver Structure for the Detection of *M*-ary Optical PPM Signals

S J Dolinar, Jr
Communications Systems Research Section

A class of receivers called "conditionally nulling receivers" is defined for quantum noise limited optical communications. These receivers have the ability to decide at each moment in time whether or not to coherently combine a predetermined local oscillator field with the received optical field, prior to performing an energy measurement (photodetection) on the combined field. Conditionally nulling receivers are applicable to pulse position modulation (PPM) and related modulation schemes, which have the property that, at each moment in time, the transmitted signal is in one of only two states, ON or OFF. The local oscillator field which may or may not be added by the receiver is an exact replica of the negative of the received ON field, hence, the receiver can exactly null the ON signal if the ON signal is present and the receiver chooses to use the local field.

An ideal conditionally nulling receiver achieves very nearly the same error probability (within a multiplicative factor varying from 1 to 2.15) as the optimum quantum measurement for quantum noise limited detection of M*-ary PPM signals. In contrast, other known receiving methods, such as direct, heterodyne, and homodyne detection, are exponentially suboptimum.*

The performance of receivers which are only approximately conditionally nulling receivers, due to imperfect nulling, is also investigated for the quantum limited PPM detection problem. Imperfect nulling is assumed to be caused by a phase discrepancy between the received ON field and the local nulling field. The performance of the imperfect conditionally nulling receiver is found to degrade rapidly to that of a direct detection receiver in the presence of nonzero phase error.

I. Introduction

In many optical communications applications, the predominant "noise" that limits system performance is the fundamental uncertainty of measurement predicted by quantum theory. However, explicit calculation of the ultimate communication inaccuracy legislated by quantum noise has been accomplished for very few problems (Refs. 1-4), and the specification of physical receiving structures for achieving optimum performance is virtually nonexistent (except, see Ref. 5). Conventional receiver structures based on familiar physical devices exhibit peculiar quantum noise limited

behavior of their own, but their performance falls short of the fundamental quantum limit

In this article a new receiver structure is defined for quantum limited optical communications Receivers with this structure have the ability to decide at each moment in time whether or not to coherently combine a predetermined local oscillator field with the received optical field, prior to performing an energy measurement (photodetection) on the combined field These receivers are termed "conditionally nulling receivers," because the local oscillator field which may or may not be added by the receiver is an exact replica of the negative of one of the possible received fields An ideal conditionally nulling receiver achieves very nearly the same error probability (within a multiplicative factor varying from 1 to 2 15) as the optimum quantum measurement for quantum noise limited detection of M-ary PPM signals In contrast, other known receiving methods, such as direct, heterodyne, and homodyne detection, are exponentially suboptimum

The purposes of this article are threefold (1) to set forth a general definition of the class of conditionally nulling receivers, (2) to broaden the definition to include receivers which are only approximately conditionally nulling receivers, and (3) to compare the performance of ideal and imperfect conditionally nulling receivers applied to the quantum noise limited M-ary PPM detection problem

II. Applicable Modulation Schemes

The useful application of conditionally nulling receivers is restricted to PPM and related modulation schemes, which have the following characteristic at each moment in time, the transmitted signal is in one of only two states, ON or OFF The OFF state is assumed to correspond to a field of amplitude zero, and the ON state corresponds to a single nonzero field At any given instant in time, there may be m possible signals in the ON state and m_o possible signals in the OFF state, and m, m_o may vary with time Multiple amplitude levels to help distinguish among the m ON signals are not permitted, but the common ON state amplitude level may vary with time Any receiver attempting to decipher this type of signaling can be viewed as an assembler of *binary* information, favoring either the m ON signals or the m_o OFF signals collectively, extracted from the received signal at each instant in time

M-ary PPM signaling fits nearly in this category, because at every moment in time $(M - 1)$ of the M possible signals are OFF and only one possible signal is ON PPM is presented as the prototypical modulation scheme for conditionally nulling receivers, and all of the performance analysis in this article is based on PPM

Three examples of applicable modulation schemes are shown in Fig 1a, b, c Figure 1a shows standard 4-ary PPM, and Fig 1b shows a PPM-complementary modulation scheme, which is discussed at length in the next section Figure 1c shows a 4-slot coding scheme in which 6 possible messages are represented by 6 different pulse doublets

III. An Unconditional Role Reversal Strategy

It is well known that direct detection does not distinguish symmetrically between ON and OFF signals For example, under quantum noise limited conditions, the detection of a single quantum of energy is sufficient to positively identify the ON signal, but equally conclusive confirmation of the OFF signal is not possible It is reasonable to ask whether simple direct detection optimally utilizes this inherent asymmetry as it attempts to distinguish at each moment in time between m ON signals and m_o OFF signals Perhaps more useful information could be obtained if the m signals were OFF and the m_o signals were ON, rather than vice-versa

A simple, complete reversal of the roles of the ON and OFF signals does not necessarily lead to improvement For example, consider quantum limited minimum error probability detection of the "role-reversed" M-ary PPM signal set shown in Fig 1b With standard PPM signaling, if the jth hypothesis H_j is true, the signal is ON during the jth signaling slot and OFF during the remaining $(M - 1)$ slots On the other hand, with "role-reversed" PPM, the signal is OFF during the jth slot and ON during the remaining $(M - 1)$ slots, under hypothesis H_j A direct detection receiver is certain to record a zero energy measurement during at least one slot (the OFF period of the true signal), and it errs only if additional zero energy measurements are also obtained A performance expression for this problem is easy to write down, if equal a priori probabilities and equal ON signal slot energies are assumed

$$P_\epsilon' = \sum_{m=0}^{M-1} \frac{m}{m+1} \binom{M-1}{m} p^m (1-p)^{M-1-m} \qquad (1)$$

Here, P_ϵ' is the error probability, p is the probability of measuring zero energy from the ON signal during a single slot ($p = e^{-E}$, where E is the mean detected ON signal slot energy in units of photons), and m indexes the number of slots in addition to the true signal slot in which no photons are observed The sum in (1) may be collapsed to the form

$$P_\epsilon' = \frac{1}{Mp} [(1-p)^M - 1 + Mp] \qquad (2)$$

The corresponding performance expression for straightforward PPM detection is well known,

$$P_\epsilon^o = \frac{M-1}{M} \, p \tag{3}$$

and it can be shown that

$$P_\epsilon^o \leqslant P_\epsilon' \text{ , for all } p\epsilon\, [0, 1] \tag{4}$$

Thus, "role-reversed" PPM appears to be inferior to straightforward PPM, even with no consideration given to its apparently increased requirements for average transmitter power (requiring $(M - 1)E$ detected photons per symbol instead of E)

IV. The Ideal Conditionally Nulling Receiver

A. General Description

It turns out that the conclusion of the previous section is no longer valid if the ON-OFF role reversal can take place at the receiver and if the receiver is able to continually choose whether or not to perform the reversal, based on its prior observations Note that the receiver can accomplish the role reversal on uncorrupted ON-OFF signals by coherently adding the negative of the ON signal prior to detection, i e , by nulling the ON signal Exact role reversal at the receiver may prove impossible under two nonideal conditions (1) if the received signal has been corrupted by a noisy channel, or (2) if the receiver cannot produce an exact replica of the ON signal

Any receiver which has the power to precisely reverse the roles of the ON and OFF signals and, in addition, is able to decide at each instant in time t whether or not to perform this operation, based on its own observations prior to t, is termed an ideal conditionally nulling receiver, or ideal conditional nuller for short A receiver which tries to act like a conditional nuller, but falls short due to an inability to perform the ON-OFF role reversal exactly, is called a conditionally nulling receiver with imperfect nulling

A block diagram of the conditionally nulling receiver structure is shown in Fig 2 The received field at any given time t is either the ON signal, denoted $S(t)$, or else 0 At the receiver a replica of the negative of $S(t)$ is generated Provision is made to coherently combine this field with the received field, but the combining is subject to an on-off switch The combined field (or received field only) impinges upon a photodetector, and the output of the photodetector (with assumed conditionally Poisson statistics) is used both to control the position of the on-off switch and eventually to infer which message was sent

B. The Quantum Limited PPM Detection Problem

It has been shown elsewhere (Ref 6) that there exists a conditionally nulling receiver for quantum limited detection of M-ary PPM signals which approaches the error probability performance of the optimum quantum measurement within a *multiplicative* factor no greater than 2 23, over all ranges of M and slot energy E This result is significant because the performance of the direct detector in (3) is *exponentially* inferior to the optimum In this section, the structure of the ideal conditional nuller and its comparative performance are briefly described

1 Nulling strategy and decision algorithm The ideal conditionally nulling receiver for the quantum limited PPM problem is defined as follows During the first signaling slot, the ON signal is nulled and an energy measurement is performed If no photons are detected, this is considered a partial confirmation of the first hypothesis relative to the remaining $(M - 1)$, and the receiver will continue to believe this hypothesis unless it subsequently obtains sufficiently conclusive evidence in favor of another Given this state of affairs at the end of the first signaling slot, the receiver decides to forgo nulling the ON signal for the remaining $(M - 1)$ slots and henceforth to simply direct detect, so that any future evidence impeaching the first hypothesis will automatically confirm with certainty one of the others If, on the other hand, one or more photons are detected during the first signaling slot, the first hypothesis is completely contradicted and the receiver's task is reduced to the $(M - 1)$-ary version of the same problem, since no information discriminating among the remaining $(M - 1)$ hypotheses has yet been obtained In this case, the receiver proceeds to the second signaling slot and again nulls the ON signal, this time looking for partial confirmation of the second hypothesis if no photons are detected or complete denial of it otherwise In the former case the receiver discontinues nulling, while in the latter it proceeds to use nulling to test for the third hypothesis, and so forth The receiver simply continues to null the ON signal until such time as it obtains partial confirmation of a specific hypothesis by the measurement of zero energy throughout the corresponding PPM slot, and afterward it direct detects through all remaining time slots

The nulling strategy and decision algorithm for the ideal conditional nuller can be represented by a binary tree diagram, as illustrated in Fig 3 for the case $M = 4$ The four levels of branches in the tree correspond to the four PPM slots Each node in the tree is labeled "N" or "D" to denote whether nulling mode or direct detection mode is used during the next

slot The branches are labeled "$\geqslant 1$" or "0" to designate whether or not at least one photon is detected during the corresponding slot Some paths (such as those containing two "$\geqslant 1$" branches from "D" nodes) have probability zero under the assumed ideal conditions, and these are marked "impossible" The end nodes in the tree which are reached with nonzero probability are all marked with the corresponding optimum decision

2. Performance The performance of the ideal conditional nuller in quantum limited conditions is easy to calculate recursively, and a simple closed form expression is obtained for the case of equal a priori probabilities and equal slot energies E, for details, see Ref 6

$$P_\epsilon = \frac{1}{M}\left[(1-p)^M - 1 + Mp\right] \tag{5}$$

where $p = e^{-E}$ as before Note that the error probability in (5) is smaller by a factor of p than the error probability obtained for the unconditional nulling strategy in (2)

The error probability achievable by the optimum quantum measurement is also known for this problem, and it is given by (Ref 3)

$$P_\epsilon^* = \frac{M-1}{M^2}\left[\sqrt{1+(M-1)p} - \sqrt{1-p}\right]^2 \tag{6}$$

The three error probabilities P_ϵ, P_ϵ^*, P_ϵ^o are plotted in Fig 4a, b, c as a function of the average detected slot energy E, for $M = 2, 16, 256$ The conditional nuller's performance is seen to track the performance of the optimum quantum measurement very closely in all situations In fact, it can be demonstrated numerically that the deviation is never more than a multiplicative factor of 2 15, i e ,

$$1 \leqslant P_\epsilon / P_\epsilon^* \leqslant 2\ 15 \qquad \text{for all } E, M \tag{7}$$

On the other hand, the performance of the direct detection receiver is exponentially inferior to that of the optimum measurement This inferiority is most apparent in the $M = 2$ graph, in which the (logarithmic) slope of the near-optimum P_ϵ curve is exactly double the (logarithmic) slope of the P_ϵ^o curve As M increases, larger values of E are needed before the direct detection curve begins to diverge from the other two

V. Conditionally Nulling Receivers with Imperfect Nulling

The potential performance advantage of conditionally nulling receivers for quantum limited PPM detection has been demonstrated, but there remains a delicate practical problem in implementing the nulling operation Perfect amplitude and phase coherence is required in order to exactly null the ON signal, and such precision is probably impossible for the receiver to achieve Thus, it is important to investigate how the performance of a conditional nuller degrades with imperfect nulling.

A. The Effect of Imperfect Nulling

A general definition of the class of conditional nullers with imperfect nulling will not be given here As stated earlier, imperfect nulling at the receiver might be inevitable in principle due to random channel disturbances imparted to the transmitted signal, or else it may result simply from the receiver's inability to produce an exact replica of a deterministic ON signal Only the second case is considered in this article

A conditionally nulling receiver with imperfect nulling differs from an ideal conditional nuller in one important respect In the absence of nulling, both receivers continually obtain by direct detection information distinguishing an ON signal of mean slot energy[1] E from an OFF signal of slot energy 0 After perfect nulling, the ON signal is converted to slot energy 0 and the OFF signal to mean slot energy E, a perfect reversal of roles After imperfect nulling, on the other hand, the ON signal is reduced to some nonzero slot energy E_o', and the OFF signal is imparted slot energy E' not necessarily equal to E It is assumed that the nulling operation is at least accurate enough that $E_o' < E'$ However, the slot energies E', E_o' will always be less distinguishable via direct detection than the slot energies E, 0, and so the efficacy of attempting an imperfect nulling operation is drawn into question

B. The Quantum Limited PPM Detection Problem

The remainder of the article evaluates the impact of imperfect nulling on the performance of the ideal conditionally nulling receiver found to be near-optimum for the quantum limited PPM detection problem

1 A reasonable nulling strategy A slight generalization of the ideal conditional nuller's operation is required in order

[1]For simplicity, it is assumed that the mean slot energies E, E_o', E' are derived from mean conditional Poisson intensity functions which are constant over a slot time This assumption eliminates the need to worry about count data records over time intervals finer than the slot interval

to adapt it to imperfect nulling With imperfect nulling, the nulled signal can never be entirely contradicted, and in fact may actually be corroborated relative to the opposite signal by the detection of a small, nonzero number of photons Thus, a natural generalization of the conditionally nulling receiver described earlier involves setting a threshold θ to determine whether to continue or discontinue nulling after observations during each signaling slot The generalized receiver starts off in the nulling mode as before, only now the nulling operation is imperfect At the end of any signaling slot in which nulling was used, a decision is made to continue nulling if the number of photons detected during the slot exceeds θ and to discontinue nulling if the number does not exceed θ As with the ideal conditional nuller, nulling, once discontinued, is never resumed On the other hand, a downgraded hypothesis, corresponding to a signaling slot in which nulling was used and counts exceeded threshold, could later be revived and chosen as most likely by the receiver if more damaging negative evidence is ultimately obtained against all the other $(M-1)$ hypotheses

The conditional nuller with imperfect nulling just described is not necessarily the best conditional nuller subject to the same nulling inaccuracy It is simply a straightforward adaptation of the ideal conditional nuller with perfect nulling that was found to be near-optimum for the given detection problem

The description in this section has detailed only the basic *structure* of the imperfect conditional nuller, the rules by which it prescribes what type of measurement to perform, i e , to null or not to null prior to detection It still remains to specify the algorithm by which the receiver optimally combines the outcomes of all its measurements to arrive at a decision For the ideal conditional nuller, the decision rules and nulling rules went hand-in-hand, and no additional analysis was needed

2 The MAP decision rule For the nonideal conditional nuller, the derivation of the optimum decision rule is still somewhat simplified (for the quantum limited problem) by the hard decisions that take place if photons are ever detected in direct detection mode The receiver must combine the always inconclusive information obtained in nulling mode with the possibly conclusive evidence obtained in direct detection mode To minimize error probability, the receiver uses a maximum a posteriori probability (MAP) decision rule

The MAP rule for the nonideal conditional nuller is derived in the Appendix, under the assumptions of equal a priori probabilities, equal slot energies, constant nulling inaccuracy from slot to slot, and constant (but selectable) threshold level θ A slightly suboptimum version of the rule is used here It can be stated as follows

(1) If no switchover from nulling to direct detection occurs, choose the slot with the fewest counts (In case of a tie for fewest counts, select randomly from the tied slots)

(2) If a switchover occurs after slot k, and nonzero energy is later detected in slot i, $i > k$, choose H_i

(3) If a switchover occurs after slot k, and no further counts are recorded, choose H_k

This statement of the decision rule is exactly optimum if the nulling strategy threshold parameter θ satisfies

$$\theta \leqslant \theta_o \equiv \frac{E + E' - E_o'}{\ln (E'/E_o')} \tag{8}$$

If θ exceeds θ_o, the true MAP decision rule consists of rules 1 and 2 exactly as stated, together with a refined version of rule 3

(3*) If switchover occurs after slot k, and no further counts are recorded, choose H_k, unless $k < M$ and the number of counts during slot k exceeds θ_o In the latter case, randomly select from H_{k+1}, , H_M

The difference between rules 3 and 3* is minor, and it does not significantly affect the numerical results Rule 3 is used in this paper because it allows a much clearer presentation of the performance analysis than rule 3*

A tree diagram depicting the behavior of the nonideal conditional nuller is given in Fig 5 for the case $M = 4$ The notation conventions are the same as in Fig 3 The branches leaving the direct detection nodes are the same as before, but the branches leaving the nulling nodes are now marked "$> \theta$" and "$\leqslant \theta$" to indicate the nonzero threshold used to determine when to discontinue nulling The decisions taken at the end nodes are the same as those in Fig 3, with the exception that the top node on the tree is now reachable with nonzero probability and an optimum decision rule at that node is now well-defined

3 Performance The error probability $\tilde{P}_e$ achieved by the nonideal conditional nuller can now be calculated Use as a conditioning variable k, the slot which triggers a switchover from nulling to direct detection, and let n_i, $1 \leqslant i \leqslant M$, denote the number of counts recorded during the ith slot With the assumed nulling strategy, slot k triggers the switchover if and only if $n_k \leqslant \theta$ and $n_i > \theta$ for $1 \leqslant i \leqslant k-1$ Note that the possibility that $k = M$ is included in this definition, even

though in this case the switchover occurs too late to actually be implemented There is one remaining possibility, that of no switchover at all (i e , $n_i > \theta$ for $1 \leqslant i \leqslant M$), and this case is treated separately below

The error probability $\tilde{P}_\epsilon$ is expanded in the form

$$1 - \tilde{P}_\epsilon = \gamma + \frac{1}{M} \sum_{j=1}^{M} \sum_{k=1}^{M} Pr(k|H_j)\, Pr(C|k,H_j) \qquad (9)$$

where

$$\gamma = \frac{1}{M} \sum_{j=1}^{M} Pr(\text{correct detection and no switchover}|H_j) \qquad (10)$$

and $Pr(C|k,H_j)$ is the conditional probability of correct detection, given H_j and switchover after slot k The latter probability is easily calculated from decision rules 2 and 3

$$Pr(C|k,H_j) = \begin{cases} 1 - e^{-E} & , k < j \\ 1 & , k = j \\ 0 & , k > j \end{cases} \qquad (11)$$

The conditional probability for the switchover location, given H_j, is evaluated as

$$Pr(k|H_j) = \begin{cases} [1 - P(\theta,E')]^{k-1}\, P(\theta,E') & , k < j \\ [1 - P(\theta,E')]^{k-1}\, P(\theta,E'_o) & , k = j \\ [1 - P(\theta,E')]^{k-2}\, [1 - P(\theta,E'_o)]\, P(\theta,E') & , k > j \end{cases} \qquad (12)$$

where $P(N,x)$ denotes the cumulative Poisson distribution function with mean x

$$P(N,x) = \sum_{n=0}^{N} p(n,x) \qquad (13)$$

where

$$p(n,x) = \frac{x^n}{n!}\, e^{-x} \qquad (14)$$

After insertion of (11) and (12) into (9) and considerable algebraic manipulation, the expression for $\tilde{P}_\epsilon$ collapses to

$$\tilde{P}_\epsilon = (1 - P(\theta,E'_o) - e^{-E}) \frac{1 - [1 - P(\theta,E')]^M}{MP(\theta,E')} + e^{-E} - \gamma \qquad (15)$$

The final item needed for $\tilde{P}_\epsilon$ is the evaluation of the term γ, Expand this probability by conditioning on n_j, then apply decision rule 1 For an exact calculation, a second conditioning variable is also needed This variable, m, denotes the number of slot counts n_i, $i \neq j$, which exactly tie n_j for lowest count

$$\gamma = \frac{1}{M} \sum_{j=1}^{M} \sum_{n_j=\theta+1}^{\infty} p(n_j,E'_o) \sum_{m=0}^{M-1} \binom{M-1}{m} \frac{1}{m+1} \times [p(n_j,E')]^m\, [1 - P(n_j,E')]^{M-1-m} \qquad (16)$$

The sum over m is evaluated in the same manner as the one in (1) and the sum over n_j is independent of j Thus, (16) reduces to

$$\gamma = \frac{1}{M} \sum_{n=\theta+1}^{\infty} \frac{p(n,E'_o)}{p(n,E')} \{[1 - P(n-1,E')]^M - [1 - P(n,E')]^M\} \qquad (17)$$

Note that, for the ideal conditional nuller, $\theta = 0$, $P(\theta,E'_o) = 1$, $P(\theta,E') = e^{-E}$, and $p(n,E'_o) = 0$ for $n \geqslant \theta + 1$ Thus, $\gamma = 0$ and (15) reduces to (5)

4 Numerical results for imperfect nulling due to phase error only The error probability expression (15) is a function of the number of PPM slots M, the mean slot energy E, the two nulling mode slot energies E'_o, E', and the nulling strategy threshold parameter θ In this section, the dimensionality of this problem is reduced from five to three by (numerically) optimizing the nulling strategy threshold and by assuming that the nulling field is subject to phase inaccuracy only In this case, the nulling mode slot energies are related to the signal slot energy by

$$E'_o = E\,|1 - e^{j\phi}|^2 = 2E(1 - \cos\phi)$$
$$E' = E \qquad (18)$$

where ϕ is the phase discrepancy (assumed constant) between the nulling field generated at the receiver and the negative of the received ON field

Error probabilities were numerically evaluated for various values of phase error ϕ, with the nulling strategy threshold θ optimized Some representative curves are shown in Fig 6a, b, c, for M = 2, 16, 256, respectively The cusps evident in these curves occur at transitions between discrete values of the optimum threshold

Superimposed on the phase error performance curves in Fig 6a, b, c are the corresponding curves from Fig 4a, b, c for the direct detection receiver and the ideal conditionally nulling receiver It is seen from Fig 6a, b, c that small values of phase error are sufficient to cause much of the performance advantage of the ideal conditional nuller to disappear For example, approximately half the performance advantage (measured logarithmically) of the ideal conditional nuller is erased by phase errors of $\phi = 4°, 2°, 1/2°$, for $M = 2, 16, 256$, respectively

It is interesting to note that the error probability curves for the imperfect conditional nuller (with optimized threshold) are upper bounded by the error probability curves for the direct detection receiver This is a general characteristic of the assumed nulling strategy that also holds for larger amounts of nulling inaccuracy than those assumed in Fig 6a, b, c, even though for large nulling error the information extracted during nulling mode is very small This property results from the fact that the error probability performance of the direct detection receiver for the quantum limited PPM detection problem is unaffected if data from the first slot is unavailable An imperfect conditional nuller starting in nulling mode can always set its nulling strategy threshold high enough that direct detection mode is essentially guaranteed for the final $(M - 1)$ slots Therefore, it can always at least match the performance of the direct detection receiver, with optimized threshold, there will be some improvement Thus, it always pays, however slightly, for the imperfect conditional nuller to start out in nulling mode

VI. Conclusions

In this article a general class of optical receivers called "conditionally nulling receivers" was postulated for PPM and related modulation schemes to take advantage of the inherent asymmetry in the information obtained by direct detection of ON and OFF signals An ideal conditional nuller achieves essentially the same error probability as the optimum quantum measurement for quantum noise limited PPM detection Analysis of the conditionally nulling receiver structure was extended to assess the effects of imperfections in the nulling process It was found that the near-optimum performance of the conditionally nulling receiver degrades rapidly, albeit gracefully, to the performance of a direct detection receiver in the presence of small nulling field phase errors

References

1 Helstron, C W, *Quantum Detection and Estimation Theory*, Academic Press, New York, 1976

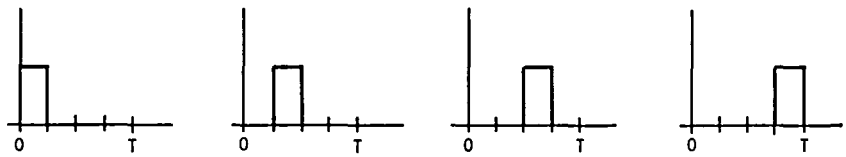
2 Helstrom, C W, "Detection Theory and Quantum Mechanics (II)," *Inform Contr*, Vol 13, pp 156-171, Aug 1968

3 Liu, J W S, "Reliability of Quantum-Mechanical Communication Systems," Tech Rep 468, Res Lab of Electronics, M I T, Cambridge, Mass, Dec 1968

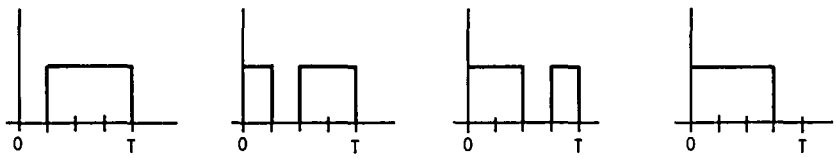
4 Personick, S D, "Efficient Analog Communication over Quantum Channels," Tech Rep 477, Res Lab of Electronics, M I T, Cambridge, Mass, May 1970

5 Dolinar, S J, "An Optimum Receiver for the Binary Coherent State Quantum Channel," *M I T Res Lab Electron Quart Prog Rep*, Vol 111, pp 115-120, Oct 15, 1973

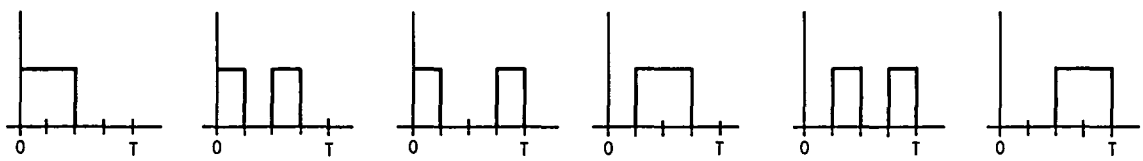
6 Dolinar, S J, "A Class of Optical Receivers Using Optical Feedback," Ph D thesis, Dept of Elec Eng and Comp Sci, M I T, Cambridge, Mass, June 1976



(a) 4-ARY PPM



(b) 4-ARY 'ROLE-REVERSED' PPM



(c) 6-ARY PULSE DOUBLET CODING ON 4 SLOTS

Fig 1 Examples of generalized ON – OFF modulation schemes

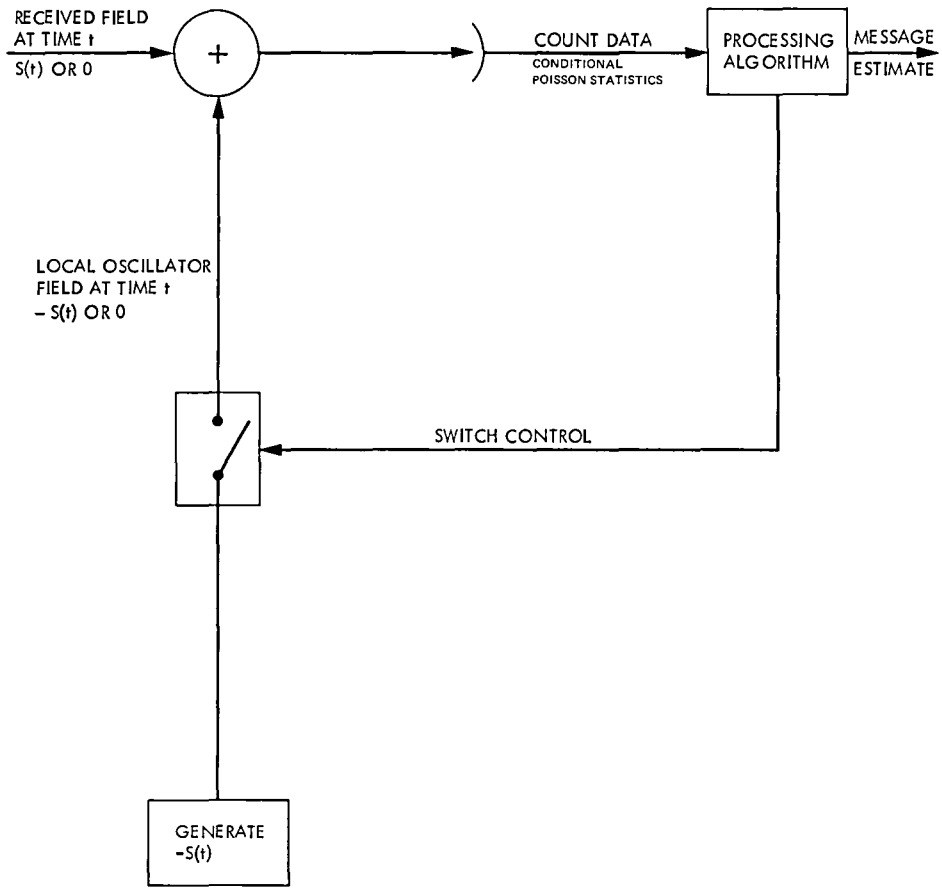


Fig. 2 Conditionally nulling receiver block diagram

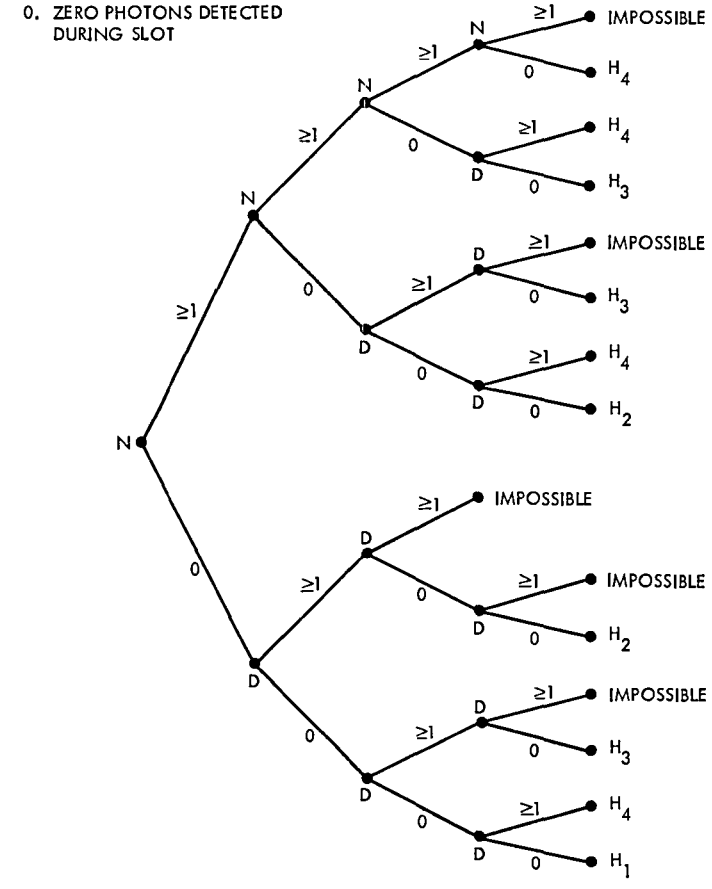


Fig. 3 Nulling strategy/decision algorithm of ideal conditional nuller for quantum limited PPM detection problem ($M = 4$)

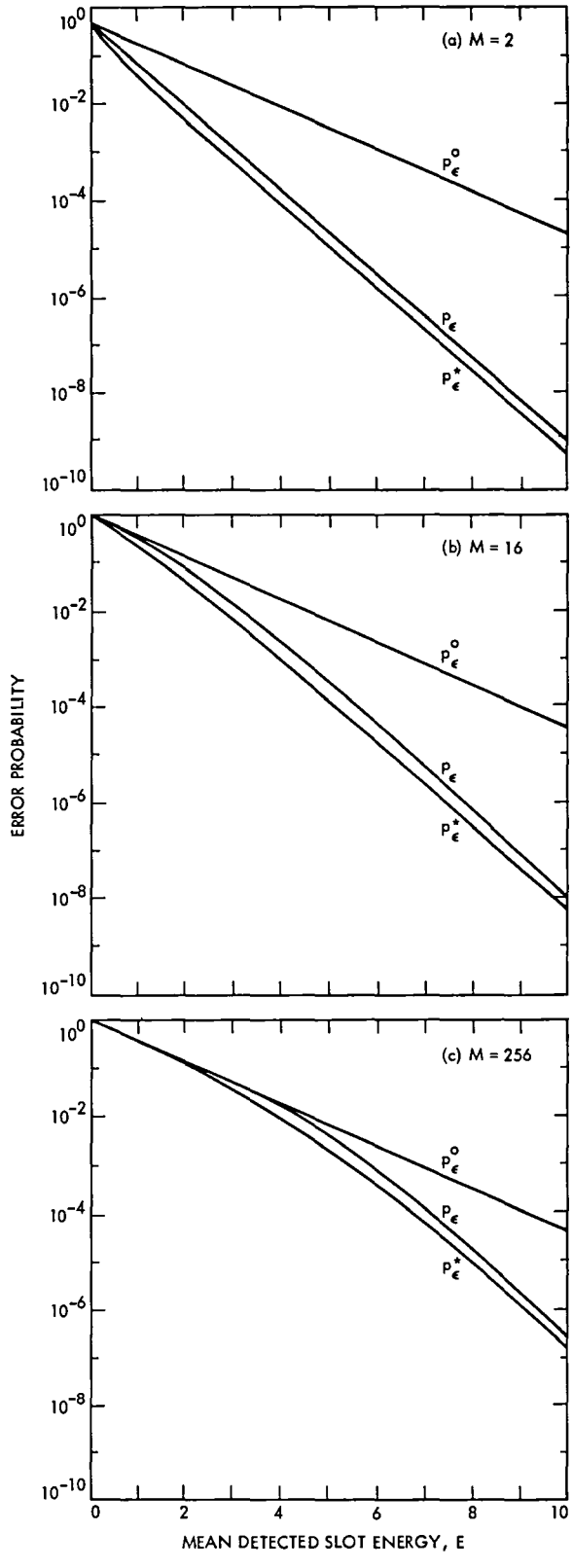


Fig 4 Comparison of P_ϵ, P_ϵ^*, and P_ϵ^o

N NULLING MODE FOR NEXT SLOT

D DIRECT DETECTION MODE FOR NEXT SLOT

n_i NUMBER OF COUNTS RECORDED DURING SLOT i

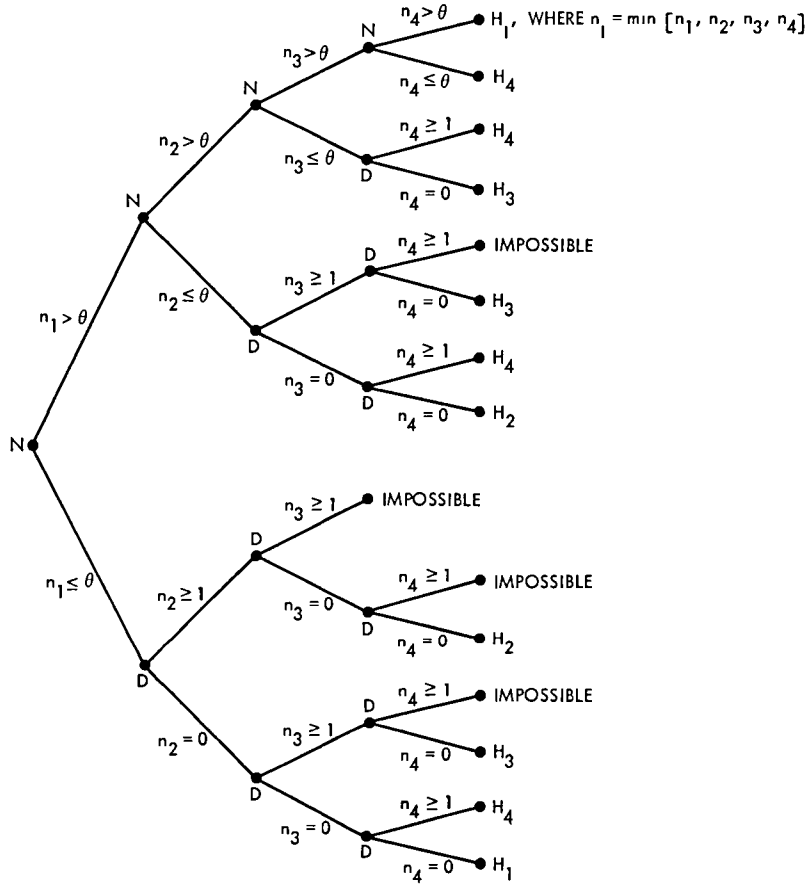


Fig 5 Nulling strategy/decision algorithm of conditional nuller with imperfect nulling for quantum limited PPM detection problem ($M = 4$)

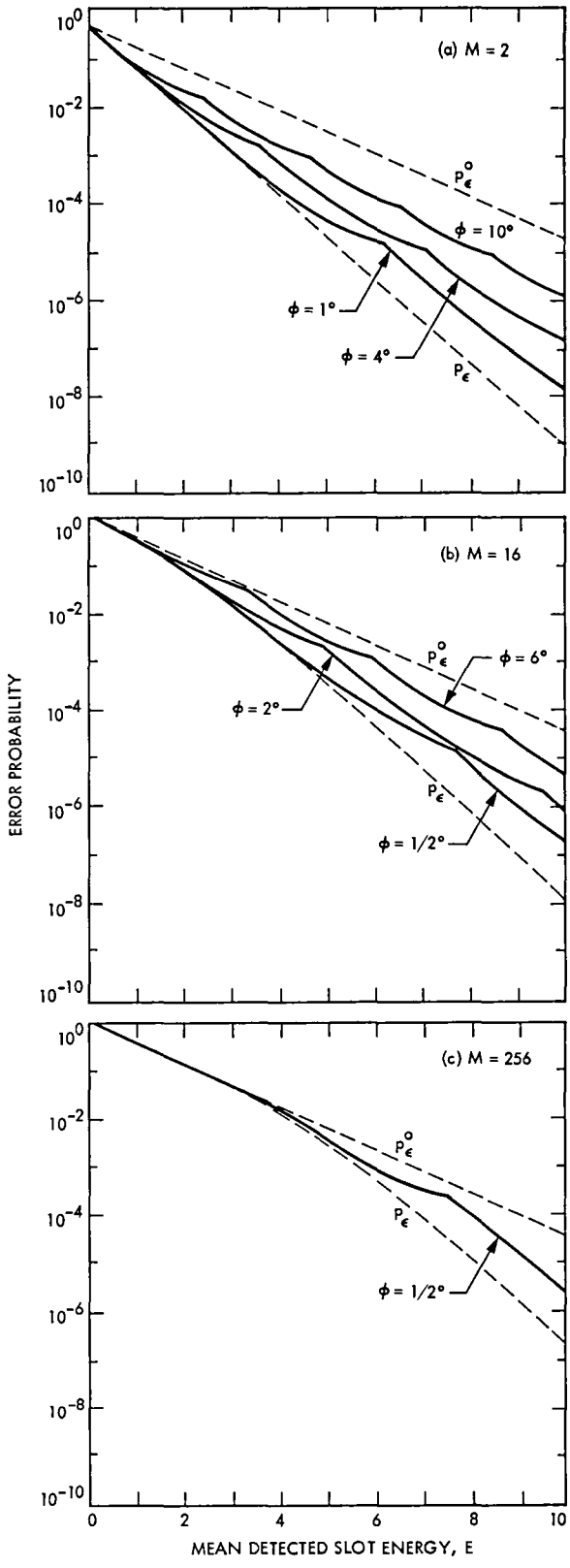


Fig. 6 Deterioration of the performance of the conditionally nulling receiver in the presence of nulling field phase error

Appendix

The following three observations lead directly to the optimum decision rule stated in the text

(1) If photons in any positive number are detected during a signaling slot in which nulling is *not* used, the corresponding hypothesis has a posteriori probability unity and is automatically selected If no photons are detected during the signaling slots in which nulling is not used, then all of the hypotheses corresponding to these slots have equal a posteriori probabilities

(2) Among the hypotheses corresponding to signaling slots in which nulling *is* used, the one with maximum a posteriori probability is the one whose signaling slot yielded the fewest detected photons In case of ties in count totals, the corresponding a posteriori probabilities are equal In particular, since the threshold θ is constant, this implies that any slot which triggers a switchover from nulling to direct detecting is "king of the hill" relative to all preceding slots

(3) If a switchover from nulling to direct detecting occurs after the kth slot and no photons are recorded in direct detection mode during the remaining $(M-k)$ slots, then the decision algorithm depends only on the number of detected photons in the kth slot, n_k There is a threshold

$$\theta_o = \frac{E + E' - E'_o}{\ln(E'/E'_o)} \tag{A-1}$$

such that H_k is selected if $n_k \leqslant \theta_o$, and a random choice among H_{k+1}, , H_M is made if $n_k > \theta_o$ In particular, this implies that H_k is always chosen in this situation if the switchover threshold θ is no greater than the decision threshold θ_o

Justification of the above statements is based on pairwise comparison of the a posteriori probabilities of hypotheses within two separate categories, corresponding to signaling slots in which nulling is or is not used Note that, for the PPM signal set, a pairwise comparison between the a posteriori probabilities of two hypotheses depends only on data obtained during the two corresponding signaling slots Thus, the most probable hypothesis within each category can be determined by successive application of binary PPM MAP decision principles The rule for the direct detection category (statement 1) is the usual rule for direct detection of PPM signals under noise-free conditions The rule for the nulling category (statement 2) is similar to the rule for PPM signals in background noise, except here the true signal slot is characterized by lowest energy instead of highest energy The final test to complete the MAP decision is to compare the a posteriori probabilities of the leaders from each category The likelihood ratio Λ_{ki} between hypotheses H_k and H_i, $k+1 \leqslant i \leqslant M$, is based entirely on the data from slots k and i Under the conditions of statement 3, during slot k nulling is used and n_k counts are observed, while during slot i nulling is not used and 0 counts are observed Thus

$$\Lambda_{ki} = \frac{Pr(n_k, 0|H_k)}{Pr(n_k, 0|H_i)} = \frac{\dfrac{(E'_o)^{n_k}}{n_k!} e^{-E'_o} \cdot 1}{\dfrac{(E')^{n_k}}{n_k!} e^{-E'} \cdot e^{-E}} \tag{A-2}$$

This expression leads to the threshold test

$$\Lambda_{ki} \geqslant 1 \quad \text{if and only if} \quad n_k \leqslant \theta_o \tag{A-3}$$

with θ_o defined in (A-1)

Capacity and Cutoff Rate of (*M*+1)-ary Decision Rules for Noisy *M*-ary Optical PPM Channel

P J Lee
Communications Systems Research Section

The channel capacity C *and the cutoff rate* R_0 *of two* (M + 1)*-ary decision rules for noisy* M *slots/symbol optical pulse position modulation (PPM) with ideal photon counting are computed and compared Also the values of the optimum thresholds needed to minimize the signal energy requirements are given. With a minor increase in hardware complexity, the symbol-by-symbol threshold decision rule is shown to be superior to the slot-by-slot threshold detection-and-decision rule in two aspects First, it saves more than 0 5 dB in signal energy for the very noisy cases of more than one noise photon per slot (for low noise cases it also saves signal energy, but a negligibly small amount) Second, it is more robust to variations in the noise level.*

I. Introduction

In this article, the theoretical performance of coded systems for the noisy optical PPM channel with two different (M + 1)-ary decision rules will be studied The system model is depicted in Fig 1 From the information bits the encoder selects channel input symbols x from an M-ary alphabet {1, · , m, , M} When $x = m$, the optical PPM modulator emits an optical pulse only at the mth time slot during a symbol time period which consists of M time slots Depending upon the numbers of the received noise and/or signal photons, the optical detector produces output $y_{m'}$ at the end of each m'th time slot, $m' = 1, 2, , M$ Ordinarily, hard decision M-ary symbols would be produced by the PPM demodulator, and passed to the decoder, which uses M-ary symbols as inputs However, if we provide, by an (M + 1)-ary decision rule, an additional symbol called "erasure" which denotes the fact that no confident decision can be made, and if we use a decoder which can use (M + 1)-ary symbols as inputs, then an additional gain can be obtained Hence, (M + 1)-ary decision (or decision with erasure) rules have been investigated

An (M + 1)-ary decision rule based on the slot-by-slot threshold detection, which requires very little complexity, has been proposed and used (Ref 1) In this case, each y_m, $m = 1, , M$, is compared to a given threshold γ, $\gamma \geq 0$, and the corresponding slot is declared active ($z_m = 1$) if $y_m > \gamma$, or inactive ($z_m = 0$) if $y_m \leq \gamma$ Then a nonerased decision is made if and only if there is only one active slot in a PPM symbol time, and the erasure decision is made for all the other cases This decision rule is shown in Fig 2(a)

More recently, the optimum and a near optimum (M + 1)-ary decision rules for the M-ary orthogonal input channel have been developed (Ref 2) The near optimum (M + 1)-ary decision rule, which uses symbol-by-symbol threshold decision, can be applied to the case of the noisy M-ary input optical PPM channel The decision rule produces a nonerased

symbol if and only if the maximum of y_m is greater than all the other $y_{m'}$ by a given threshold δ, $\delta \geq 0$ In all other cases, the erasure decision is made This decision rule is shown in Fig 2(b) We can see that the decision rule of Fig 2(b) is only slightly more complex than that of Fig 2(a)

The same $(M+1)$-ary decision rule with symbol-by-symbol threshold decision was also considered, independently, but without showing the near optimality, by Divsalar, et al (Ref 3) under the different name of "delta-max demodulation" They compared this decision rule to the slot-by-slot threshold detection-and-decision rule by calculating the decoded bit error rate with a Reed Solomon code in each case Their results, however, apply to the specific set of codes studied and cannot be easily extended to other situations

Notice that with any $(M+1)$-ary decision rule, the inside of the encoder-decoder pair (inside of the dashed line in Fig 1) becomes an M-ary input, $(M+1)$-ary output discrete memoryless coding channel The channel capacity C and the cutoff rate R_0 of a given coding channel have long been recognized as valuable performance indicators of the coding channel Also, these quantities are independent of the actual codes to be used over the coding channel Furthermore, their evaluation is usually much easier than computing the decoded bit error rates with specific complex codes

Therefore, in this article we will compute C's and R_0's with more variety of situations (noise levels and/or channel input alphabet sizes) than Ref 3 for more detailed comparisons of the two decision rules We will show that the performance gain of the symbol-by-symbol threshold decision rule, over the slot-by-slot threshold detection-and-decision rule, is larger than 0 5 dB under strong background noise conditions As a by-product of the evaluation we will, as in Ref 3, determine the optimum threshold settings We will also show that under weak background noise conditions (such as in Ref 1), the improvement is negligibly small

II. Evaluation of C and R_0

Let $G(a) = Pr\{y_m \leq a \mid x = m\}$ and $F(a) = Pr\{y_{m'} \leq a \mid x = m \text{ and } m' \neq m\}$ represent the cumulative probability distribution functions of the receiver output value for the slot with and without signal pulse, respectively Let $P_c(\epsilon)$, $P_s(\epsilon)$, and $P_e(\epsilon)$ be the probabilities of correct decision, decision error, and erasure decision, respectively, with a given threshold ϵ, which is either γ or δ depending on the decision rule Recall that γ is the threshold for the slot-by-slot detection-and-decision rule and δ corresponds to the symbol-by-symbol decision rule For the slot-by-slot threshold detection-and-decision, these probabilities are given by (Ref 1)

$$P_c(\gamma) = P_{ds}(\gamma) \cdot P_{dn}^{M-1}(\gamma),$$

$$P_s(\gamma) = (M-1) \cdot (1 - P_{ds}(\gamma)) \cdot (1 - P_{dn}(\gamma)) \cdot P_{dn}^{M-2}(\gamma),$$

and

$$P_e(\gamma) = 1 - P_c(\gamma) - P_s(\gamma),$$

where

$$P_{ds}(\gamma) = 1 - G(\gamma) \quad \text{and} \quad P_{dn}(\gamma) = F(\gamma)$$

For the symbol-by-symbol threshold decision (Refs 2, 3),

$$P_c(\delta) = \int F^{M-1}(t) \cdot dG(t+\delta),$$

$$P_s(\delta) = (M-1) \cdot \int G(t) \cdot F^{M-2}(t) \cdot dF(t+\delta),$$

if y_m takes on continuous values, or

$$P_c(\delta) = \sum_k F^{M-1}(k) \cdot g(k+\delta+1),$$

$$P_s(\delta) = (M-1) \cdot \sum_k G(k) \cdot F^{M-2}(k) \cdot f(k+\delta+1),$$

if y_m takes on integer values (in which case the threshold δ should also be an integer), with

$$g(i) = G(i) - G(i-1) \quad \text{and} \quad f(i) = F(i) - F(i-1)$$

Finally,

$$P_e(\delta) = 1 - P_c(\delta) - P_s(\delta)$$

The capacity C and cutoff rate R_0 in [infor bits/channel bit] for the resulting M-ary input, $(M+1)$-ary output coding channel, which can be described with $P_s(\epsilon)$, $P_e(\epsilon)$, and $P_c(\epsilon)$, are given by (Ref 2),

$$C = \max_{\epsilon \geq 0} C(\epsilon)$$

and

$$R_0 = \max_{\epsilon \geq 0} R_0(\epsilon)$$

where

$$C(\epsilon) = P_c(\epsilon) + P_s(\epsilon) + (P_c(\epsilon) \cdot \log [P_c(\epsilon)] + P_s(\epsilon) \cdot \log [P_s(\epsilon)] - [P_c(\epsilon) + P_s(\epsilon)] \cdot \log [P_c(\epsilon) + P_s(\epsilon)] - P_s(\epsilon) \cdot \log [M-1])/\log [M],$$

and

$$R_0(\epsilon) = 1 - \log [1 + (M-1) \cdot P_e(\epsilon) + (M-2) \cdot P_s(\epsilon) + \sqrt{4 \cdot (M-1) \cdot P_c(\epsilon) \cdot P_s(\epsilon)}\ ] / \log [M],$$

and ϵ is either γ or δ, depending on the decision rule (Note that the values of the optimum threshold need not be the same for each different criterion)

We assume here an ideal photon counting receiver That is, we neglect thermal noise and assume a constant gain PMT (photomultiplier tube) In this simple case, $f(i)$ and $g(i)$ are given by (Ref 3)

$$f(i) = \exp(-\bar{N}_n) \cdot \bar{N}_n^i / i!$$

$$g(i) = \exp(-\bar{N}_n - \bar{N}_s) \cdot (\bar{N}_n + \bar{N}_s)^i / i! \qquad i = 0, 1, 2, \ldots,$$

where $\bar{N}_n$ is the average number of the received background noise photons per PPM slot and $\bar{N}_s$ is the average number of the received signal photons per PPM symbol

When M and $\bar{N}_n$ are given, we can calculate C's and R_0's as functions of a single parameter of $\bar{N}_s$ Since C and R_0 are monotonically increasing functions of $\bar{N}_s$, they are invertable The coding theorem (Ref 4) says that there exists a code with code rate r [info bits/channel bit] which gives arbitrary small decoded error rate provided that

$$r < C(\bar{N}_s)$$

Or equivalently, since C is invertable, the condition can be rewritten as

$$\bar{N}_s > C^{-1}(r)$$

Since the parameter of interest is usually the average number of received signal photons per information bit $\bar{N}_b$ [photons/info bit], we can obtain it from $\bar{N}_s$ by the simple relation $\bar{N}_b = \bar{N}_s/(r \cdot \log_2 M)$ Therefore, the condition can again be rewritten as

$$\bar{N}_b > C^{-1}(r)/(r \cdot \log_2 M)$$

Hence, we call $C^{-1}(r)/(r \cdot \log_2 M)$ the "required average number of signal photons per information bit, or simply photons per information bit, to achieve capacity C with a given code rate r" Similarly, $R_0^{-1}(r)/(r \cdot \log_2 M)$ is called the "required average number of photons per info bit to achieve a cutoff rate R_0 with a given code rate r"

For M = 64, 256, and 1024 (or $\log_2 M$ = 6, 8, and 10), the required N_b [signal photons/info bit] to achieve (a) the channel capacity C and (b) the cutoff rate R_0 vs code rate r plots are given in Figs 3, 4, and 5 for various noise levels of N_n [noise photons/slot] for the two different (M + 1)-ary decision rules The curve marked N_n = 0 corresponds to the noiseless channel use

III. Discussions and Conclusions

First, we notice the nonsmoothness of the curves in the plots This is due to the discreteness of the observation values and threshold values Also the discreteness of the observation and threshold values gives large gaps in performance between the two decision rules near the nonsmooth points This phenomenon is more noticible for the low noise level cases, due to the small threshold values (Note that there is a much bigger difference when using threshold values of 0 or 1 than the difference corresponding to thresholds of 22 or 23) If, instead of using an ideal photon counting receiver assumption, we had considered a more realistic random gain PMT with additive thermal noise model (Ref 1), then the nonsmoothness and the large gaps for some rates would disappear and hence the differences in performance between the two decision strategies would be almost the same for all values of r (We can see this for very noisy cases where the curves tend to be much smoother, as well as for $\bar{N}_n = 10^{-2}$ of Fig 4(a) where accidently the optimum threshold levels do not vary for almost all values of r)

As was concluded in Ref 3, the symbol-by-symbol threshold decision rule saves more signal energy than the slot-by-slot threshold detection-and-decision rule as the noise level increases For the noiseless case there is no gain at all For the 2 5 bit/photon demonstration program (Ref 1), 256-ary PPM with rate 3/4 or 7/8 coding is used The typical value of background noise is 10^{-5} [noise photons/slot] In this case also, we calculated C's and R_0's The gain in signal energy of using the symbol-by-symbol threshold decision over using the slot-by-slot threshold detection-and-decision is less than 0 03 dB for almost all code rate, hence totally negligible With this noise level, the performance is about 0 2 dB poorer than the noiseless case However, for the very noisy case of more than

one noise photon per slot, the gain is more than 0 5 dB for all values of M considered We also observed as in Ref 3 that the symbol-by-symbol threshold decision rule is more robust to variations in the noise level

In conclusion, we have found that the symbol-by-symbol decision strategy always requires less signal energy than the simpler slot-by-slot detection-and-decision strategy, with greater benefits both in energy savings and in robustness to noise level variations for the high background noise cases In the weak background noise case, the differences in performance are negligibly small, and hence the symbol-by-symbol decision will not be attractive since it requires additional complexity

References

1 Lesh, J R , Katz, J , Tan, H H , and Zwillinger, D , "2 5 Bit/Detected Photon Demonstration Program Analysis and Phase I Results," *TDA Progress Report 42-66*, Jet Propulsion Laboratory, Pasadena, Calif , pp 115-132, Dec 1981

2 Lee, P J., and Lesh, J R , "On the Decisions with Erasure for M-ary Orthogonal Input Channel," in preparation

3 Divsalar, D , Gagliardi, R M , and Yuen, J H , "PPM Demodulation for Reed-Solomon Decoding for the Optical Space Channel," *TDA Progress Report 42-70*, Jet Propulsion Laboratory, Pasadena, Calif pp 47-59, Aug 1982

4 Gallager, R G , *Information Theory and Reliable Communication*, John Wiley and Sons, New York, 1968

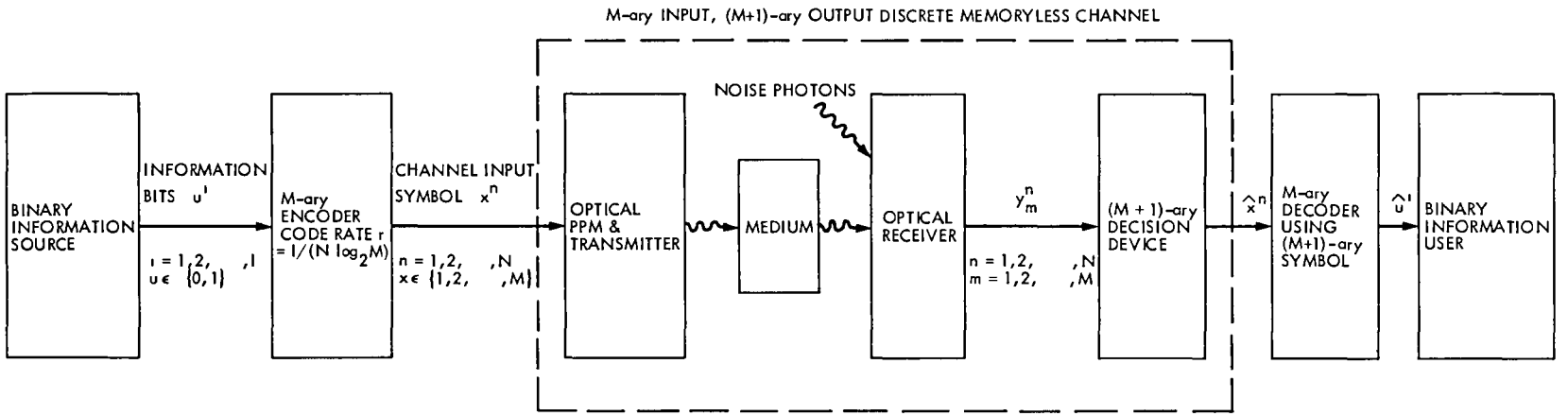


Fig. 1. Coding system for *M*-ary optical PPM channel with (*M*+1)-ary decision rule

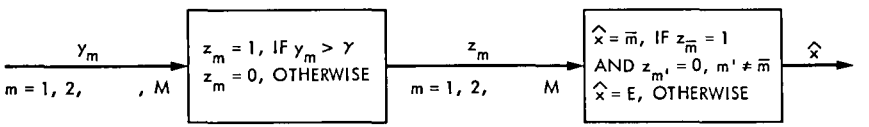


Fig. 2(a). Slot-by-slot threshold detection-and-decision device

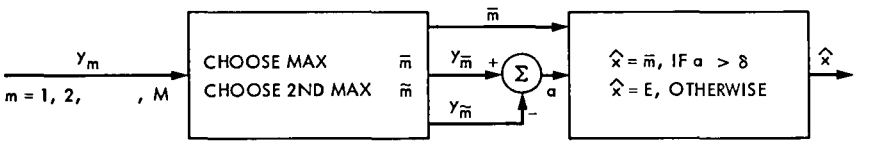


Fig. 2(b). Symbol-by-symbol threshold decision device

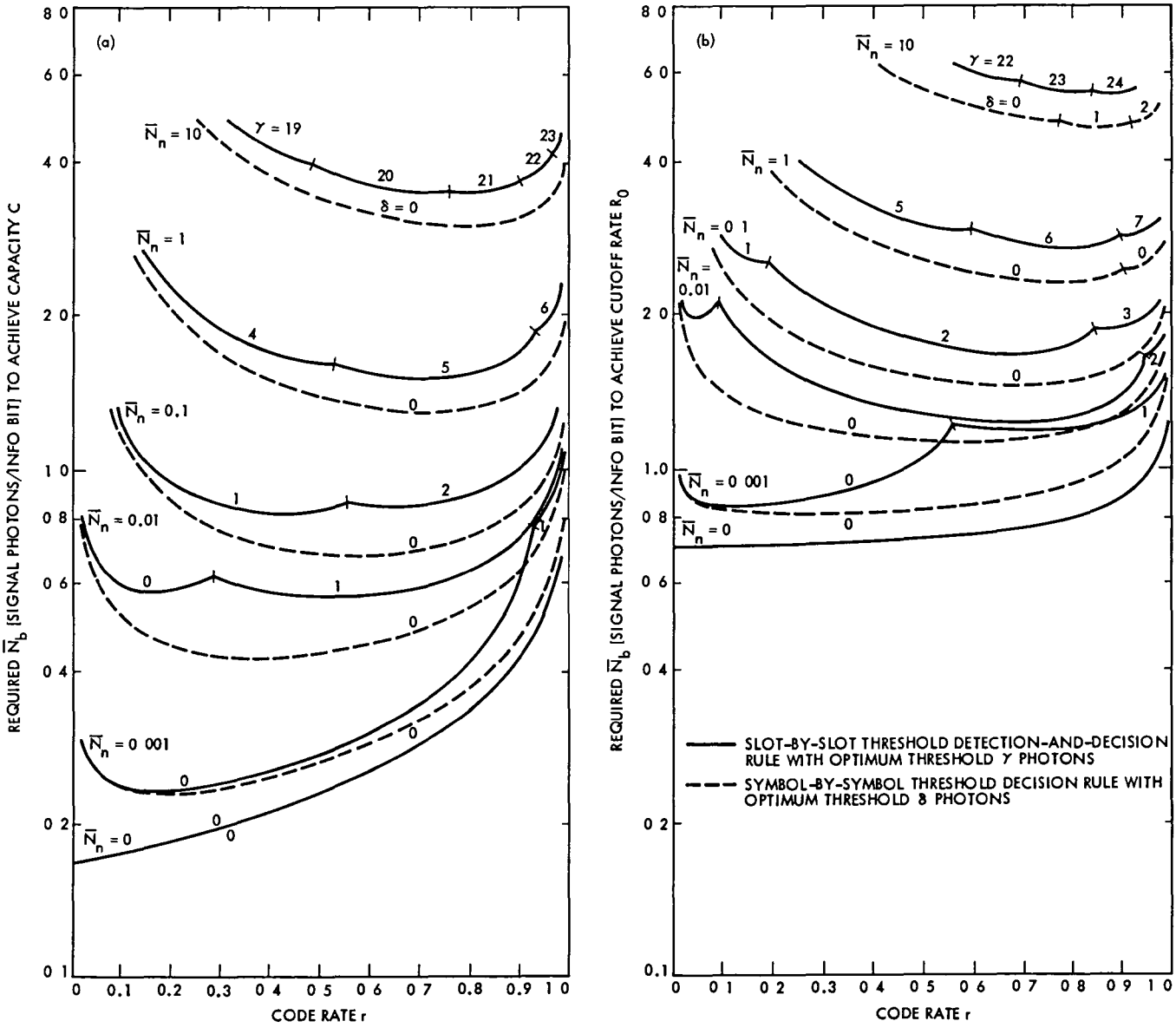


Fig. 3. Required N_b [signal photons/info bit] to achieve (a) channel capacity C and (b) cutoff rate R_0 vs code rate r for two different $(M+1)$-ary decision rules on ideally direct-detected $M = 64$-ary optical PPM channel with several background noise levels N_n [noise photons/PPM slot]

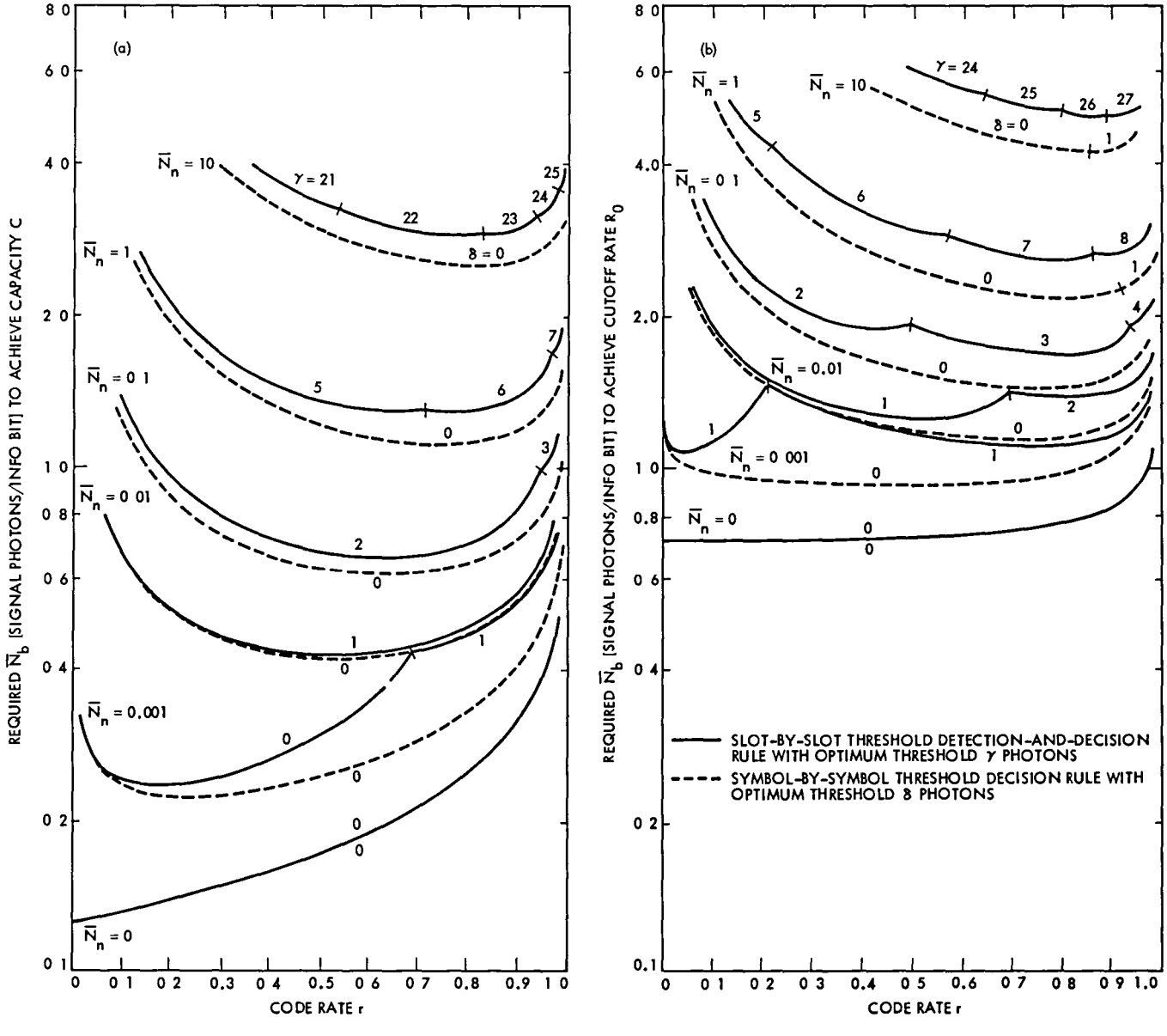


Fig. 4. Required N_b [signal photons/info. bit] to achieve (a) channel capacity C and (b) cutoff rate R_0 vs code rate r for two different (M+1)-ary decision rules on ideally direct-detected M = 256-ary optical PPM channel with several background noise levels N_n [noise photons/PPM slot]

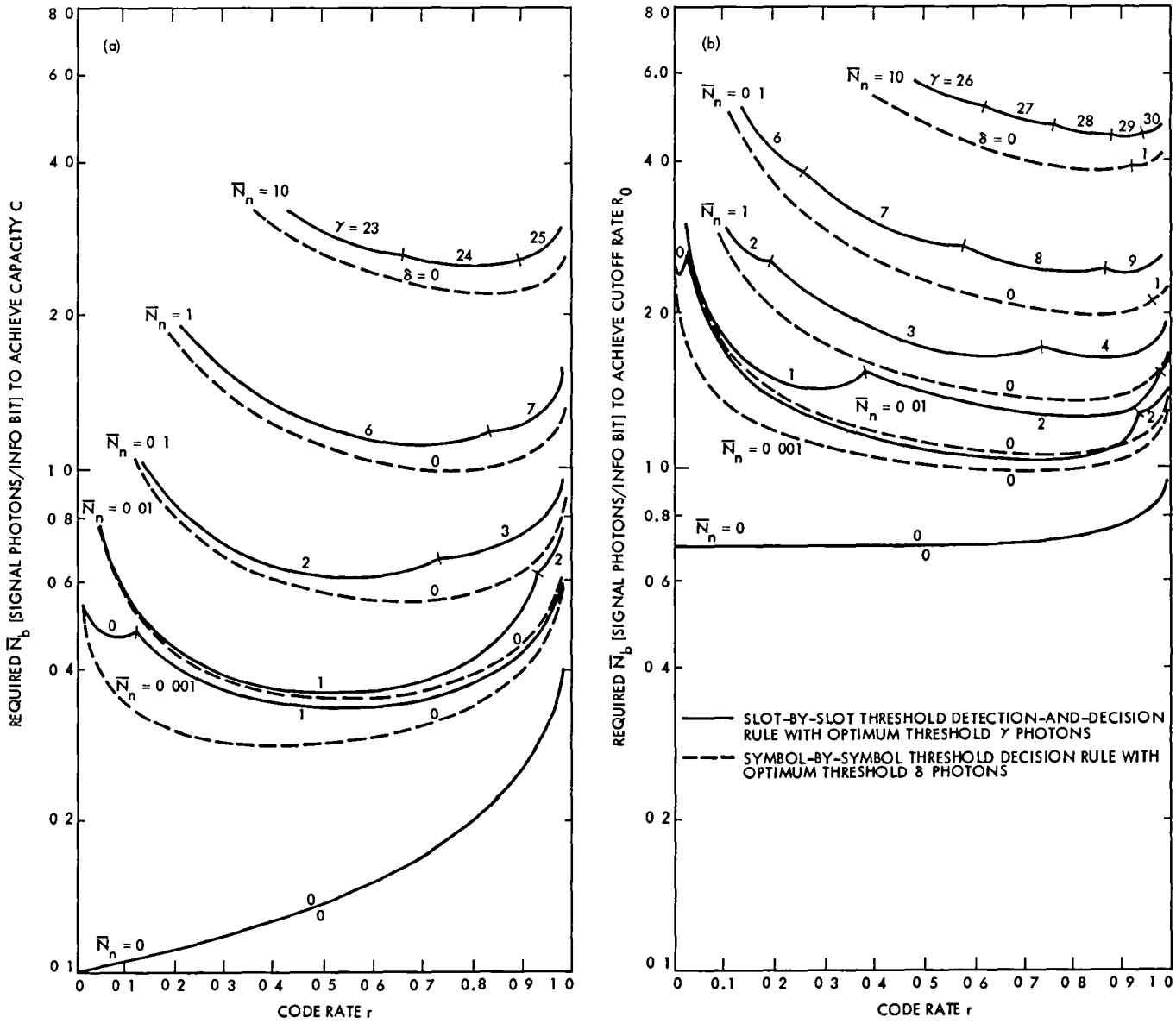


Fig 5. Required N_b [signal photons/info bit] to achieve (a) channel capacity C and (b) cutoff rate R_0 vs code rate r for two different $(M+1)$-ary decision rules on ideally direct-detected $M = 1024$-ary optical PPM channel with several background noise levels N_n [noise photons/PPM slot]

An Update on the Use of the VLA for Telemetry Reception

L. J Deutsch
Communications Systems Research Section

This article modifies a previous analysis to incorporate the actual structure of the command signal system of the Very Large Array (VLA). In particular, in addition to the 1-ms command signal of the previous article, there is a "data invalid" signal that is generated. The command signals are transmitted to the antennas during the period in which the "data invalid" signal is on This means that the gaps in the received data are really 1.6 ms long rather than 1 ms long Simulation results with this taken into account show that the VLA will not support (7, 1/2) convolutionally encoded telemetry at acceptable error rates at any of the Voyager telemetry data rates VLA will support Voyager encounters provided that either concatenated coding is implemented, VLA is arrayed with another receiving site (such as Goldstone), or VLA is reconfigured so that the gaps are rotated

I. Introduction

The performance of the Very Large Array (VLA) as a receiver for Voyager 2 coded telemetry data was recently determined (Ref 1). Since that time, however, more information concerning the operation of the VLA has come to light This additional knowledge does affect the previous results In fact, new simulations and modeling show that the VLA will not be capable of supporting convolutional coded Voyager data at any of the data rates that are being considered for Uranus or Neptune encounters

The VLA is a phased array consisting of 27 independently pointable parabolic dishes in the New Mexico desert Each dish measures 25 meters in diameter The array is configured in three linear arrays of nine antennas each that radiate from a common centerpoint The antennas may be moved along these lines by means of railroad-type tracks Communication between the antennas is by means of a connecting network of waveguides The total aperture of the VLA is about twice that of a DSN array consisting of a 64-m antenna and three 34-m antennas The VLA was designed for use as a radio astronomy observatory Some additional equipment will be needed in order for the VLA to receive the X-band Voyager telemetry transmissions

The main problem with using the VLA as a telemetry receiver is that signals are not transmitted continuously from the antennas to the processing facility Instead, the VLA operates in a duty cycle mode Every so often, at regular intervals, communications from the antennas are suspended momentarily so that calibration signals may be sent to the antennas In Ref 1 it was stated that the VLA collects data for 51 ms, followed by a 1-ms calibration period This turns out to be only partially true (Ref 2). There is indeed a 1-ms calibration signal However, data is only collected for 50 4 ms of the 52-ms cycle The relevant signals are illustrated in Fig 1 The 1-ms calibration signal is transmitted during the time that the "data invalid" signal is on This means that for the purposes of this study, the VLA must be treated as a system that collects data for 50 4 ms followed by a 1 6-ms gap.

If uncoded telemetry data were received by the VLA then, for data rates over a few thousand bits per second, the best possible error rate would be (1/2) (1 6/52) = 1 54 X 10^{-2} Coding must be used in order to achieve bit error rates below the 5 X 10^{-3} level required for good image quality In Ref 1 it was stated that VLA could support Voyager 2 convolutionally encoded transmissions at a data rate of 10 8 kbps at both Uranus and Neptune encounters In light of the new information mentioned above, this is no longer the case In order to achieve good quality imaging data from either encounter at the VLA, additional coding must be used or the VLA must be reconfigured to at least partially resolve the gap problem

II. Best Possible Performance of VLA with Convolutional Coding Only

In this section, the performance of the VLA for the reception of convolutionally encoded data is derived in the case of infinite bit SNR This represents the best possible performance of the system The formula that was derived in Ref 1 is still valid

$$P_b = 0\,5\,(R_bT_2 - k + 1)/[R_b(T_1 + T_2)]$$

As before, k is the constraint length of the convolutional code, R_b is the data rate in bits per second, T_1 is the time in seconds spent in that portion of the cycle representing a perfect channel, and T_2 is the time in seconds spent in each gap It is assumed that R_b is large enough so that the gaps are at least k bits long In the case of interest in this report, $k = 7$, $T_1 = 50\,4$ ms, and $T_2 = 1\,6$ ms A graph of P_b as a function of R_b is shown in Fig 2 This represents the best possible performance of the VLA with (7, 1/2) convolutional coding It is evident from the figure that in this scenario, the VLA will not support any of the planned Voyager data rates at either encounter

III. Software Simulations of VLA Performance

The performance of the VLA under various scenarios was determined by using the simulation software that was described in Ref 1 Five scenarios were simulated For each scenario (except the second) two data rates were considered 10 8 and 19 2 kbps These data rates are representative of those that can be expected at Voyager's future planetary encounters In addition, the performance of VLA was determined both for (7, 1/2) convolutionally encoded data and for (7, 1/2) coded data concatenated with a (255, 223) Reed-Solomon code This latter coding scheme, which will be referred to as *concatenated coding*, is an option for Voyager A detailed description of concatenated codes may be found in Ref 3 The scenarios that were considered are as follows

(1) *Ideal performance* In this scenario, it was assumed that no gaps were present in the data This represents a best possible case and it can be used to measure the magnitude of losses that occur in the other scenarios

(2) *Normal VLA* This scenario represents the VLA as described in the introduction The gaps are assumed to be periods of time in which the channel E_b/N_0 is arbitrarily small Since it is already known from the remarks in Section II that this scenario will not suffice for Voyager encounters, only the 10 8-kbps data rate was simulated The results give an indication of just how far off the mark the performance really is

(3) *VLA arrayed with an equal aperture* The gaps in this scenario are 3 0 dB lower in E_b/N_0 than the rest of the data This scenario can be interpreted as arraying half of the VLA with the Goldstone DSN complex The combining is assumed to be perfect combined carrier referencing (CCR) Two different methods of achieving nearly ideal CCR performance are described in Refs 4 and 5

(4) *Equally spaced rotating gaps* This is the first of two scenarios that involve a reconfiguring of the VLA In this scenario, each of the three arms of the VLA receives its 1 6-ms gaps at a different time These times are equally spaced so that the entire array has a duty cycle of 52/3 ms with 1 6-ms periods of 1 7-dB attenuation in the received data

(5) *Clumped rotating gaps* In this scenario, as in 4, each arm receives its 1 6 ms of control signals at a different time In this case, however, these individual arm gaps are consecutive in time This means that the VLA would still have a duty cycle of 52 ms but there would be a 3 X 1 6 = 4 8 ms part of each cycle that is attenuated by 1 7 dB It turns out that although one could expect better performance from scenario 4, this scenario is easier to implement in the present VLA hardware

For the purposes of displaying the results of the simulations, the expression E_b/N_0 represents the bit SNR during the good part of the VLA cycle at all times in the figures

The results of the simulations for the convolutional-only case are shown in Figs 3-7 The performance of the VLA in scenarios 2, 3, and 4 is slightly worse than that reported in Ref 1 due to the longer gap lengths Also, one can clearly see the slight advantage that scenario 4 has over scenario 5 This is because the Viterbi decoder can handle several short gaps better than a single long one

The performance of the VLA with concatenated coding was calculated from these simulations using the formulas described in Ref. 1. These results are shown in Figs. 8-12 In the case of concatenated coding, a bit error rate of 10^{-5} is considered necessary for good image quality

IV. VLA Throughput at Voyager Encounters

The most reasonable measure of performance for a telemetry receiving system such as the VLA is the amount of data throughput that it is capable of handling The particular measure of throughput used in this study is the number of "good bits per day " During each day of an encounter, Voyager will be "visible" from the VLA for some number of hours Depending on the data rate and coding that are implemented, only a fraction of this viewing time may support reception at the required bit error rate or better This is due to the fact that received power is a function of the elevation angle of the antennas

The raw data in this section comes from design control tables for Voyager 2 Uranus encounter The curve shown in Fig 13 represents the total power-to-noise-spectral-density ratio (P_T/N_0) that is expected to be incident at the Goldstone complex on day 34 of 1986 (Uranus encounter) This curve includes a 90% weather confidence level and the effects of antenna elevation angle Since the VLA is at approximately the same latitude as Goldstone, the values of P_T/N_0 for the VLA at Uranus encounter were derived from these by simply scaling them up by the ratio of the apertures involved For Neptune encounter, 3.5 dB was subtracted to account for the additional space loss It was assumed that the maximum viewing time is 8 3 hours

The throughputs for the various scenarios and data rates were calculated according to the formulas that were derived in Ref 1 and the data from the simulations of Section III The results are shown in Fig 14 A comparison with the corresponding figures in Ref 1 shows the extra loss from the 0 6-ms additional gap length

V. Conclusions

The main point of this study is that the VLA cannot be used as a telemetry receiver for Voyager 2 Uranus and Neptune encounters in its present configuration with only convolutional coding There are three workable solution to this problem The first and easiest is to use concatenated coding Voyager 2 does have this capability on board and there is already Reed-Solomon decoding hardware at JPL The second is to array the VLA with another receiver that does not have this gapped behavior The prime candidate for such an array would be the Goldstone complex due to its relatively close proximity The third solution is to reconfigure the VLA so that the gaps occur at different times in the different apertures Two such methods were described in Section III.

References

1 Deutsch, L J , "The Performance of VLA as a Telemetry Receiver for Voyager Planetary Encounters," *TDA Progress Report 42-71*, Jet Propulsion Laboratory, Pasadena, Calif , Nov. 15, 1982

2 Escoffier, R P , "Module L8 Digital Divider," VLA Technical Report No 12, National Radio Astronomy Observatory, Socorro, N. M , June 1976

3 Butman, S. A , Deutsch, L J , and Miller, R L , *On the Error Statistics of Viterbi Decoding and the Performance of Concatenated Codes*, Publication 81-9, Jet Propulsion Laboratory, Pasadena, Calif , Sept. 1, 1981

4 Deutsch, L J., Lipes, R. G , and Miller, R L , "Virtual Center Arraying," *TDA Progress Report 42-65*, Jet Propulsion Laboratory, Pasadena, Calif , Oct 15, 1981

5 Divsalar, D D , and Yuen, J H , "Improved Carrier Tracking Performance with Coupled Phase-Locked Loops," *TDA Progress Report 42-66*, Jet Propulsion Laboratory, Pasadena, Calif , Dec 15, 1981

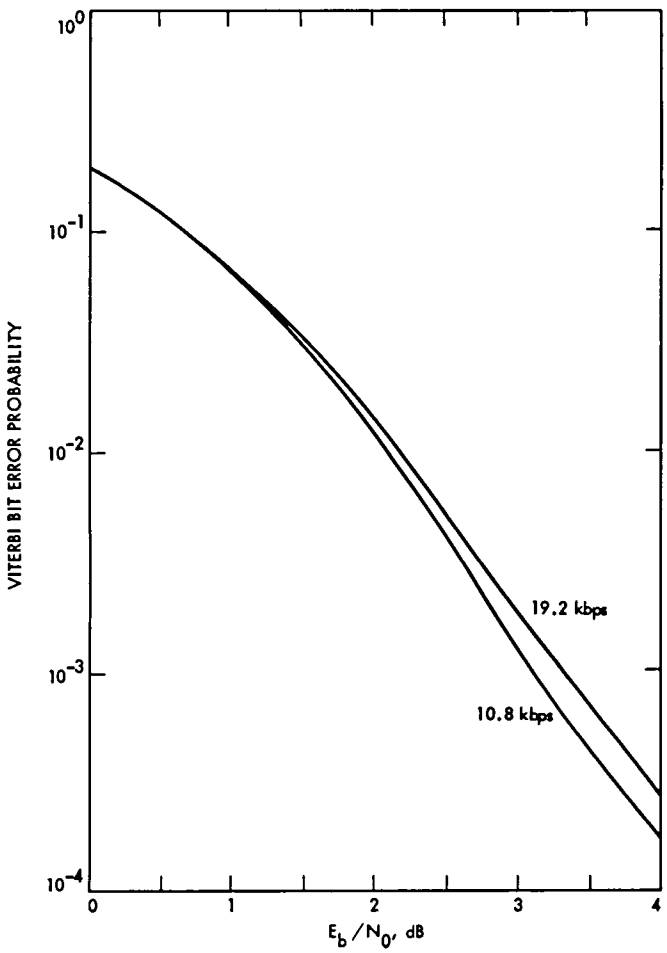


Fig 6. Simulated performance of VLA with equally spaced rotated gaps (Scenario 4) (7, 1/2) convolutional coding only

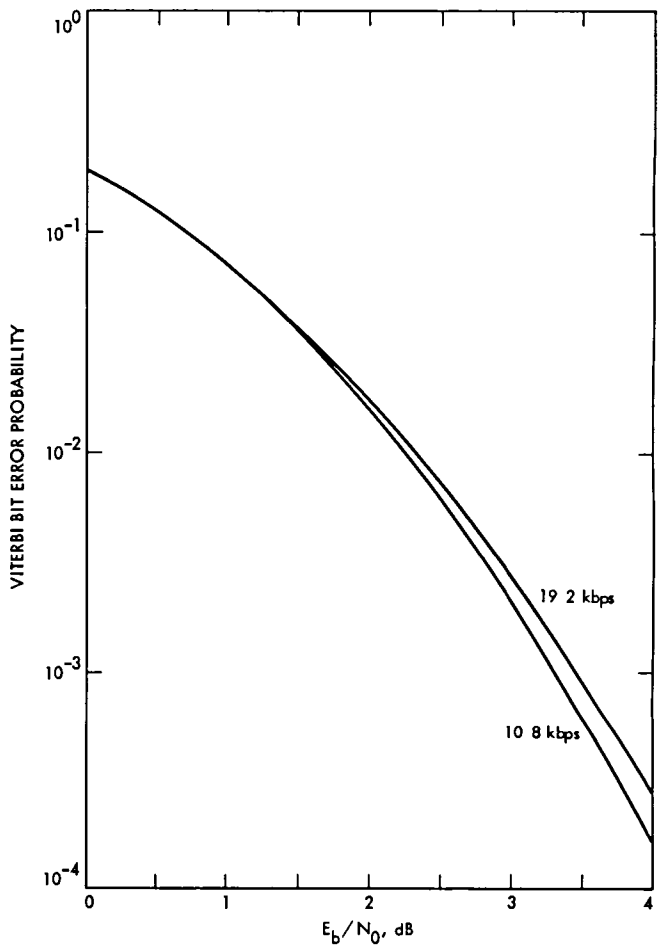


Fig 7. Simulated performance of VLA with clumped rotated gaps (Scenario 5) (7, 1/2) convolutional coding only

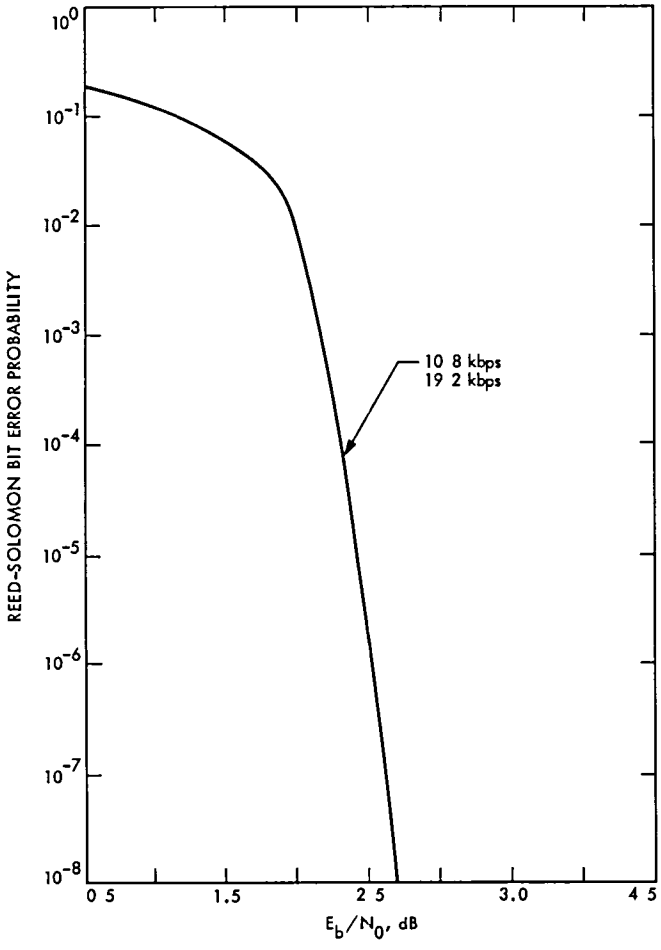


Fig 8 Simulated performance of an ungapped receiving system (Scenario 1) Concatenated coding

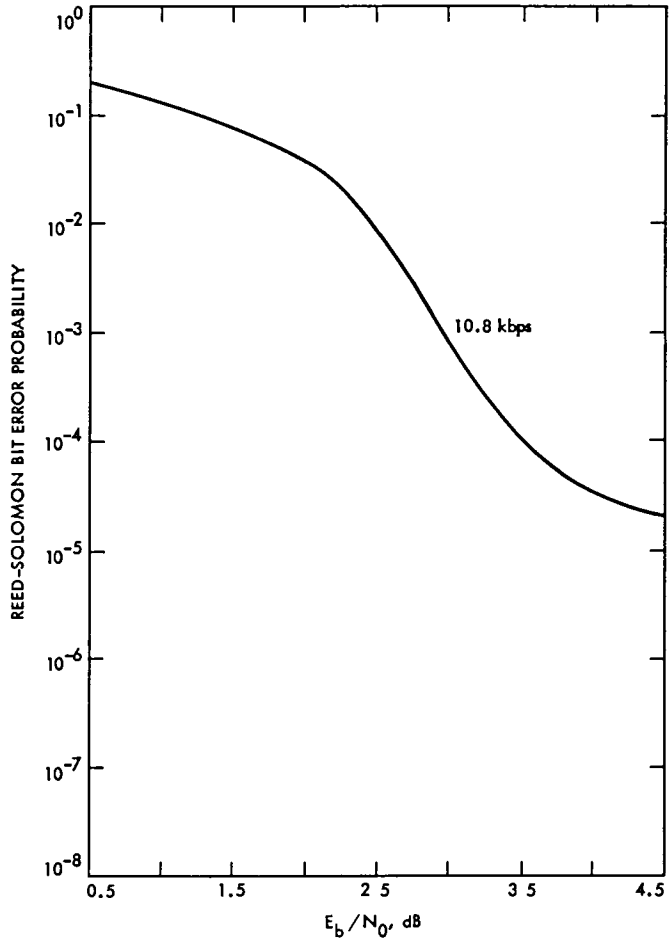


Fig. 9. Simulated performance of VLA with 1.6-ms gaps with no signal (Scenario 2) Concatenated coding

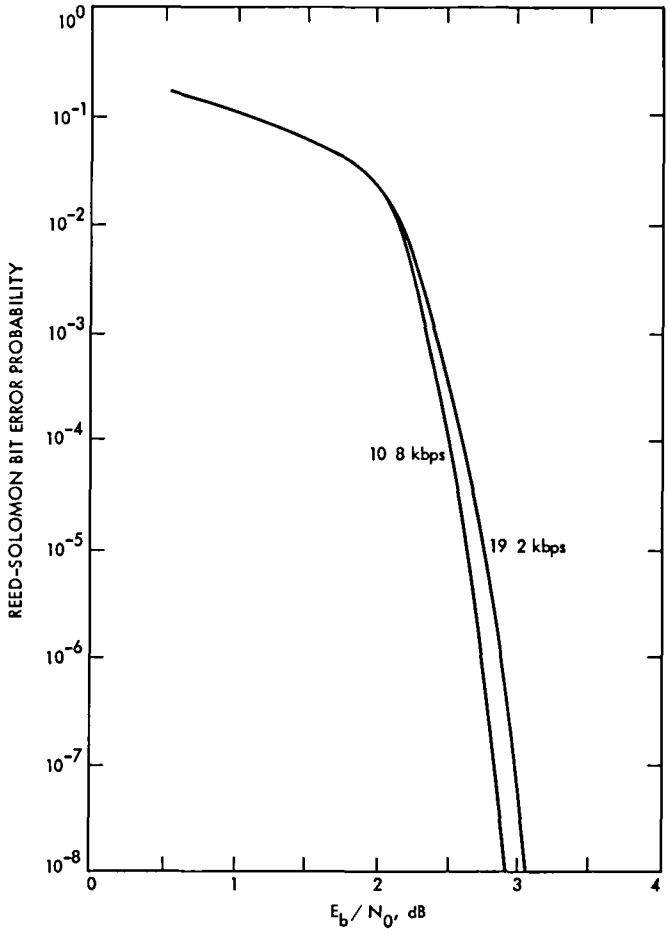


Fig. 10. Simulated performance of VLA with 1 6-ms gaps attenuated by 3 dB (Scenario 3) Concatenated coding

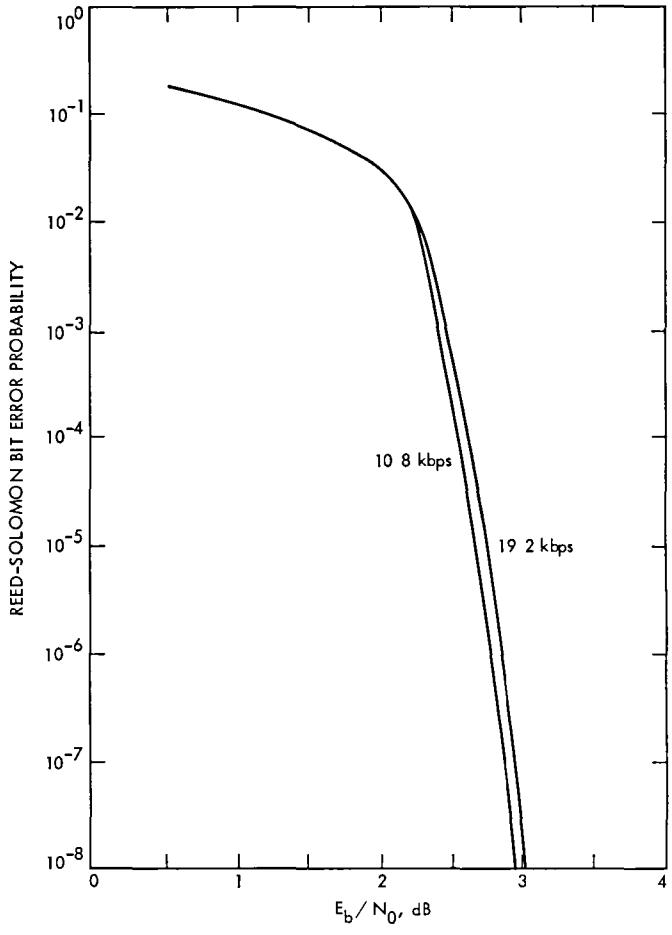


Fig 11 Simulated performance of VLA with equally spaced rotated gaps (Scenario 4) Concatenated coding

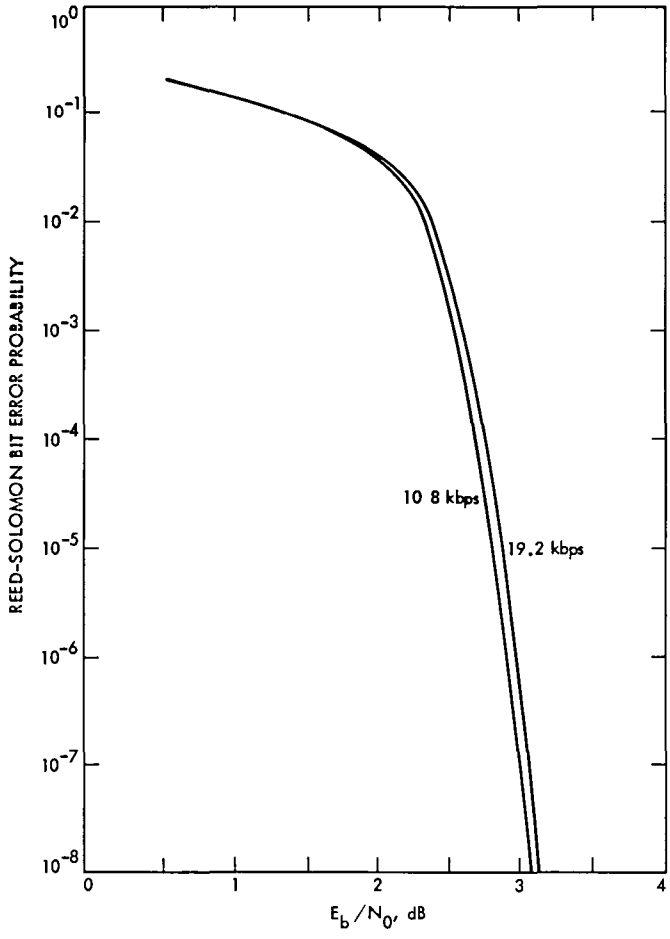


Fig. 12. Simulated performance of VLA with clumped rotated gaps (Scenario 5). Concatenated coding

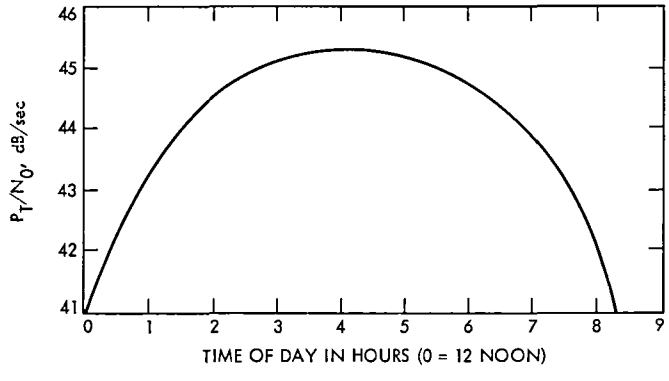


Fig. 13. Baseline performance of Goldstone array at Voyager 2 Uranus Encounter, array = 64 m + 34 m + 34 m, day 34, 1986, 90% weather

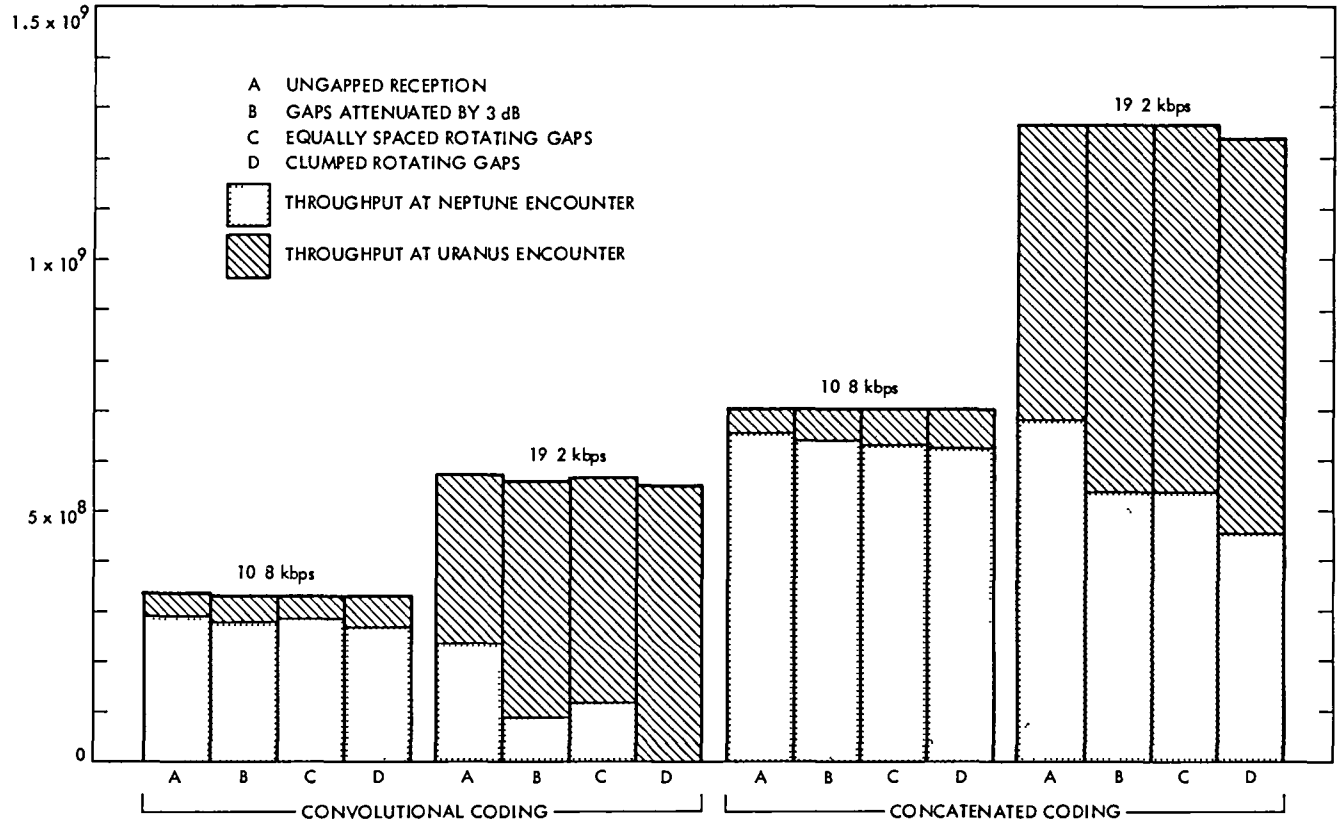


Fig. 14. VLA throughput at Voyager 2 Uranus and Neptune encounters

A VLSI Implementation of a Multicode Convolutional Encoder

L J Deutsch
Communications Systems Research Section

This article describes a VLSI architecture and layout for a convolutional encoder. This architecture allows a single chip implementation of an encoder that is capable of handling many different convolutional codes including all the convolutional codes that are presently used by NASA for deep space missions.

I. Introduction

Most present and all planned deep space missions make use of convolutional coding One reason for this is that the encoder for these codes is very simple and can be made relatively small and light even with off-the-shelf components A block diagram of a convolutional encoder is shown in Fig 1 The convolutional code implemented in the figure is the NASA standard (7, 1/2) code The input information bits $i_1 i_2 i_3$ are shifted into a register This register is used to generate two parity checks each time a new information bit is entered A commutator interleaves these parity checks to produce the coded output stream $a_1 b_1 a_2 b_2 a_3 b_3$

In a general convolutional decoder, the shift register is K bits long The number K is called the "constraint length" of the code In the most general situations, there can be more than one K-bit shift register in the encoder For the purposes of this article, however, it will be assumed that there is only a single register If there are P parity checks that are generated by the encoder, then the resulting code rate will be $1/P$ The code would then be referred to as a $(K, 1/P)$ convolutional code This is not enough to define the code uniquely The particular code is determined by the parity checks

Consider a single parity check that is generated in the encoder Let the positions in the shift register be labeled 1, 2, 3, with the position 1 being the one that the data bits enter first It follows that this parity check can be described by a K-bit binary number where a one in position i indicates the presence of a parity check tap after state i of the shift register The parity checks themselves can be ordered so that "parity check #1" is the first to be scanned by the commutator and the "parity check #P" is the last Using these conventions, any $(K, 1/P)$ convolutional code can be represented by an ordered set of P K-bit binary numbers For the (7, 1/2) code described above, the numbers are

1011011

1111001

If one reads the columns of this binary matrix as binary numbers, this can be reduced to simply 3233013 This sequence of integers, together with the more common (7, 1/2) notation, completely describes the convolutional code

Sometimes one or more of the parity checks are inverted before the multiplexing operation of the commutator This

is done so that additional channel symbol transitions are introduced in the case that the information bits consist of long strings of either all zeroes or all ones These transitions are needed in order to synchronize the symbols in the receiver before they can be sent to a decoder These inversions have no effect on the performance of the code itself but their presence in any usable encoder is essential

The performance of convolutional codes has been studied in Refs 1-3 The problem of node synchronization (determining what the order of the parity checks is in the decoder) was treated in Refs 3-5

II. A Functional Description of the Chip

The chip, as implemented, is capable of supporting three codes These are the standard NASA (7, 1/2) code 3233013, the standard NASA (7, 1/3) code 7376147, and a longer constraint length (10, 1/3) code 7461776427 For each of these codes, the capability of inverting any or all of the parity checks has been included in the design

There are 22 pins that will be connected to the chip These are

INPUTS

Vdd	The voltage connection to the chip – nominally -5 Vdc
Gnd	The ground connection to the chip
Sub	A connection to allow biassing of the chip substrate
1/3	Selects the code rate A "one" signifies that a rate 1/3 code has been selected, while a "zero" means the rate is 1/2
SEVEN	Selects the code constraint length A "one" indicates a constraint length seven code while a "zero" indicates a constraint length of ten
NRM1	This signal determines whether or not the first parity check is inverted A "one" indicates the noninverted state
NRM2	The same as NRM1 for the second parity check
NRM3	The same as NRM1 for the third parity check This signal has no effect if a rate 1/2 code has been selected
BITIN	This is the data input
PHI1IN	The input for the first phase of a two-phase clock
PHI2IN	The input for the second phase of a two-phase clock
CLOCKIN	The input for a one-phase clock
CLKSEL	Selects between a one- and two-phase clock A "one" indicates that a one-phase external clock has been selected

OUTPUTS

OUTPUT	The output for the encoded channel symbols
PHI1OUT	Output of first phase of the internal two-phase clock If a two-phase external clock has been selected (CLKSEL=0), then this is the same as PHI1IN
PHI2OUT	Similar to PHI1OUT for the second clock phase
S1	Indicates when the commutator is scanning the first parity check
S2	Indicates when the commutator is scanning the second parity check
S3	Indicates when the third parity check is being scanned In the case that a rate 1/2 code is selected, S3 is always zero
P1OUT	The first phase of the shift register advancing clock
P2OUT	The second phase of the shift register advancing clock
X#9	The output of the shift register This is simply BITIN delayed by 10 bit times

As implemented, there are two ways of getting clocking information into the chip A two-phase clock can be wired to the device by using PHI1IN and PHI2IN and grounding the input CLKSEL Also, a one-phase clock can be used on the CLOCKIN input In the latter case, the input pin CLKSEL must be wired to voltage This two-choice system is possible because a two-phase clock generator has been included in the chip itself

A timing diagram that shows the relationship between the time-varying signals is shown in Fig 2 It is assumed, for the moment, that the selected code rate is 1/3 Notice that the clock frequency must be P times the input data rate (recall that P is the inverse of the code rate) The data that is entered on pin BITIN is clocked through the shift register by the signals P1 and P2 P1 and P2 are simply the logical "and" of S1 and the signals PHI1 and PHI2 respectively The data in the register will remain stable during the three clock times that it takes to scan the three parity checks The particular parity checks that are scanned are also functions of the input signal SEVEN The signals S1, S2, and S3 represent the position of the commutator It is important that the input bits must be synchronized with the commutator so that a new bit is entered whenever P1 is high The output symbols on the pin OUTPUT are synchronized with PHI2 Since all the important timing signals are available as pins, it would be an easy task to interface this chip with data sources

In the case that the selected code rate is 1/2, then the clock frequency is twice that of the input data rate and the signal S3 is always low This case is indicated in the timing diagram of Fig 3

Although the actual clock speeds of the chip cannot be determined until the fabricated device is tested, it is likely that a maximum clock rate of at least 1-2 MHz will be achievable Such clock rates are nominal for the NMOS technology that is being used for this project

III. Chip Architecture

The VLSI multicode convolutional encoder logic is divided into two major sections The first is the timing signal generator and the second is the convolutional encoder array

The timing signals that were described in Section II are generated in the timing signal generator subblock A programmable logic array (PLA) (Ref 6) is used to implement a counter for generating the commutator signals S1, S2, and S3 If the input "1/3" is high, then the counter is set to count modulo three If "1/3" is slow, then it counts modulo two P1 and P2 are generated by ANDing S1 with PHI1 and PHI2 respectively This subblock also contains an output latch that synchronizes the output channel symbols with the signal PHI2 before they are sent to the output pad OUTPUT

The actual encoding takes place in the convolutional encoder array subblock It is shown conceptually in Fig 4 This figure also indicates the geometric layout of this section of the chip The input bits from BITIN are shifted through the ten-bit parallel output register shown along the bottom of the figure The output of each register cell is available at the top as well as the right end Above each cell are six blocks, each of which is either a PRTY (or "parity") cell or a NPRTY (or "no parity") cell Each row of parity cells corresponds to a particular parity check on the ten outputs of the register The presence of a PRTY cell in row i and column j means that the i^{th} parity check has a tap at position j in the register The top three parity checks are for the (10, 1/3) code The lower three are for the (7, 1/3) code The (7, 1/2) code uses the first two parity checks from the (7, 1/3) code

The parity checks are not implemented as exclusive OR operations as is done in a conventional encoder Instead, there are two signal paths through each row of the parity check matrix At the extreme left of the array, a "one" is connected to one of these and a "zero" to the other Each time this pair of signals passes through a PRTY cell, they exchange places if the output of the corresponding register cell is a "one " A NPRTY cell has no effect on these signals (circuit diagrams for these two cells appear in Fig 5) In this way, the pair of signals has gone through a number of path exchanges equal to the number of ones in the shift register at the taps that correspond to that particular parity check If the parity is even, then the signals are in the same place as when they started If the parity is odd, they have come out reversed Also, since a complementary pair of signals is used, the complement of the parity check, as well as the check itself, is available at this point

The signals NRM1, NRM2, and NRM3 are used to select either the parity information or its complement at the end of each row of the parity array The signal SEVEN is then used to select either the top three or bottom three parity checks The commutator is implemented as a multiplexer in which the signals S1, S2, and S3 are used to select the appropriate parity check

This architecture is a good example of how an algorithm that works well in a SSI system is not nearly optimal in VLSI In the conventional encoder, it would be impractical to provide all the interconnections that are necessary for implementing the above algorithm. This VLSI architecture is compact, fast, and easily extendable to longer constraint lengths and more parity checks In fact, it would be a relatively simple matter to write a silicon compiler (Ref 7) that would automatically lay out masks for similar encoders for arbitrary sets of convolutional codes

It should be mentioned that the overall size of this chip was determined by the number of pads rather than the size of the logic This means that there is some extra space available Some of this space is filled with a long array of "butting contacts " A butting contact is a structure for connecting the polysilicon and diffusion layers in NMOS There is some doubt as to whether this structure will work reliably in the 2 0-

micron NMOS technology By including this test array it will be possible to gather some additional data on this subject The two ends of the array of contacts are connected to the pins BUTT-TEST1 and BUTT-TEST2 It should be noted that the butting contact was not used in the remainder of the chip

IV. The Design Process

The multicode convolutional encoder chip was designed according to the NMOS design rules that appear in Ref 6 The design philosophy was influenced by the JPL VLSI design course as taught by Dr G Lewicki and Dr R Nixon

The actual chip layout was performed using the interactive graphics editor "CAESAR" that was developed at the University of California at Berkeley by Dr J Ousterhout (Ref 8) This software design system allows the user to develop a chip in a hierarchical fashion with full editing capabilities There are also a number of programs including a PLA generator and optimizer that can be used in conjunction with CAESAR A picture of the finished layout as produced with CAESAR appears in Fig 6 The UC Berkeley software also includes a circuit extractor called MEXTRA that works in conjunction with CAESAR to create a file that can be used with a number of simulation packages One very nice feature of the extractor is that it keeps track of labels that the user creates with CAESAR so they can be referred to during simulation A complete description of the Berkeley software can be found in Ref 9 The chip was completely simulated on a logical level using the program ESIM that was developed at the Massachusetts Institute of Technology Although it was not used for this project, the circuit level simulator SPICE (Ref 10) is also integrated with CAESAR and MEXTRA

V. Conclusions

It is evident, even in a simple project such as this one, that VLSI technologies allow the implementation of much more efficient algorithms than conventional circuit design This is because all intermediate signals that are generated in a circuit can be made available to other portions of the circuit without a proliferation of pins and drivers that can greatly reduce the computational speed of the system

The chip described here was sent out over the ARPANET for fabrication on December 2, 1982 When the completed chips return they will be evaluated and tested

These chips will probably be used as test function generators in research on convolutional decoding They will not be radiation hardened and so they cannot be flight qualified However, the architecture described in this article could be easily used to produce flight qualified chips if the need ever arises Also, the experience gained on this project will be invaluable in the design of the complex decoding systems that are planned for future projects

References

1 Layland, J W , "Information Systems Performance of Short Constraint Length Convolutional Codes and a Heuristic Code-Construction Algorithm," *Space Program Summary 37-64, Vol. II*, Jet Propulsion Laboratory, Pasadena, Calif , Aug 1970

2 Miller, R L , Deutsch, L J , and Butman, S A , *On the Error Statistics of Viterbi Decoding and the Performance of Concatenated Codes*, Publication 81-9, Jet Propulsion Laboratory, Pasadena, Calif , Sept 1, 1981

3 Liu, K Y , and Lee, J J , "An Experimental Study of the Concatenated Reed-Solomon/Viterbi Channel Coding System and Its Impact on Space Communications," Publication 81-58, Jet Propulsion Laboratory, Pasadena, Calif , Aug 15, 1981

4 Deutsch, L J , and Miller, R L , "The Effects of Viterbi Decoder Synchronization Losses on the Telemetry Receiving System," *TDA Progress Report 42-68*, Jet Propulsion Laboratory, Pasadena, Calif , Apr 15, 1982

5 Deutsch, L. J , and Miller, R L , "Viterbi Decoder Node Synchronization Losses in the Reed-Solomon/Viterbi Concatenated Channel," *TDA Progress Report 42-71*, Jet Propulsion Laboratory, Pasadena, Calif , Oct 15, 1982

6 Mead, C , and Conway, L , *Introduction to VLSI Systems*, Addison-Wesley Publishing Company, Reading, Mass , 1980

7 McNair, R , and Miller, M , "Bristle Blocks – Scrutinized and Analyzed," Computer Science Department Display File, California Institute of Technology, 1982

8 Ousterhout, J. K , "Caesar An Interactive Editor for VLSI Layouts," *VLSI Design*, Volume II, No 4, Fourth Quarter 1981

9 Fitzpatrick, D. T , et al , "A RISCy Approach to VLSI," *VLSI Design*, Volume II, No 4, Fourth Quarter 1981

10 Negal, L W , and Pederson, D O , "SPICE – Simulation Program with Integrated Circuit Emphasis," Memorandum No ERL-M382, Electronics Research Laboratory, University of California, Berkeley, April 12, 1973

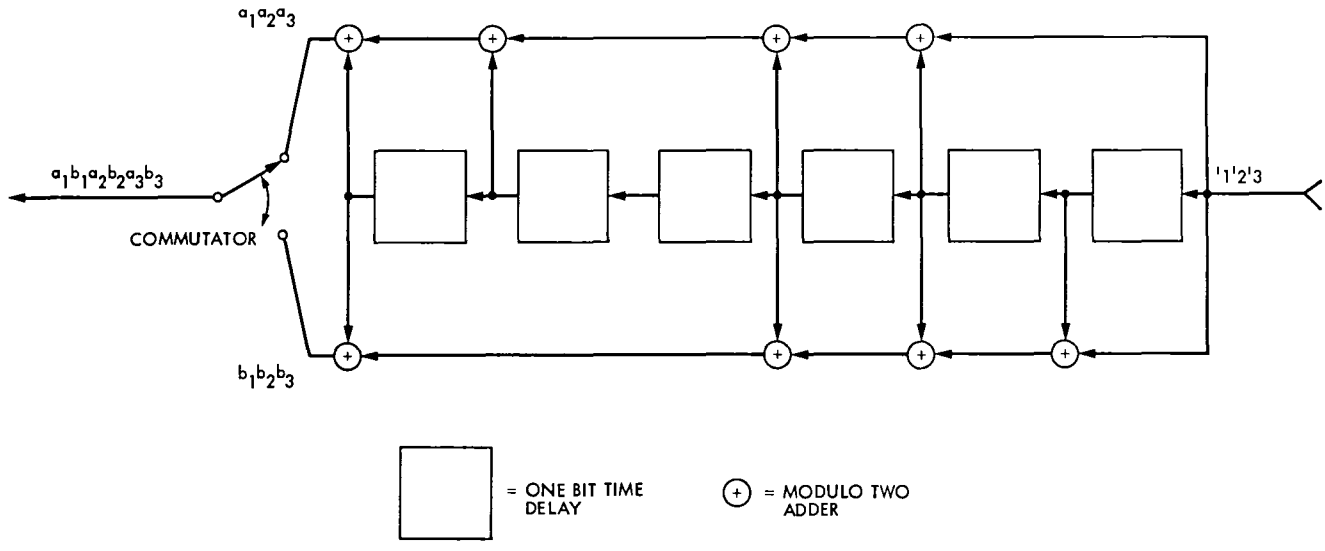


Fig 1 Conceptual diagram of a (7, 1/2) convolutional encoder

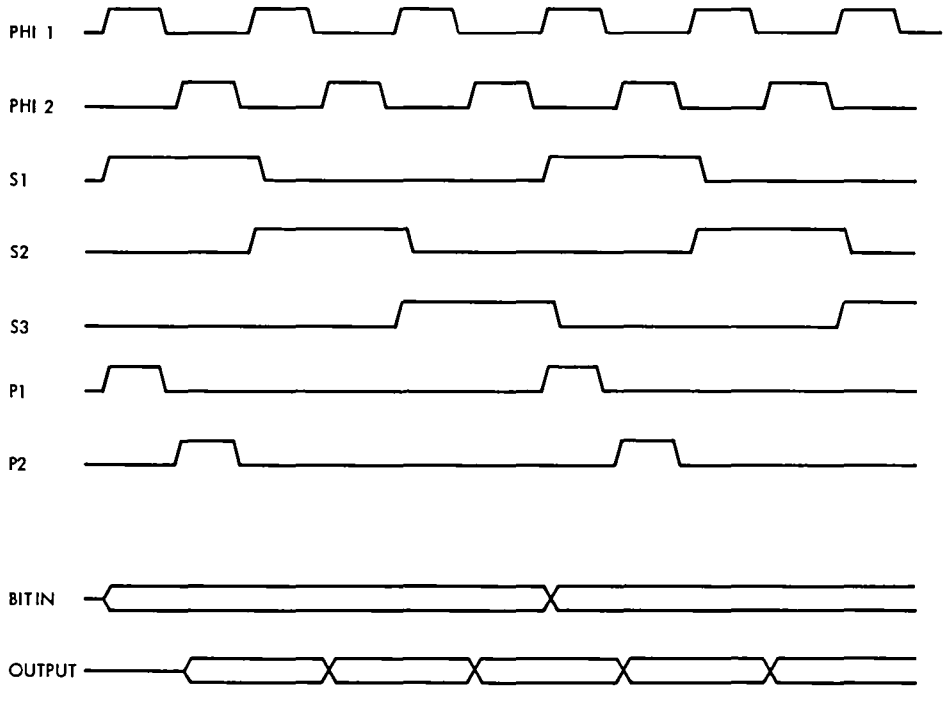


Fig 2 Timing signals for the convolutional encoder, rate 1/3 codes

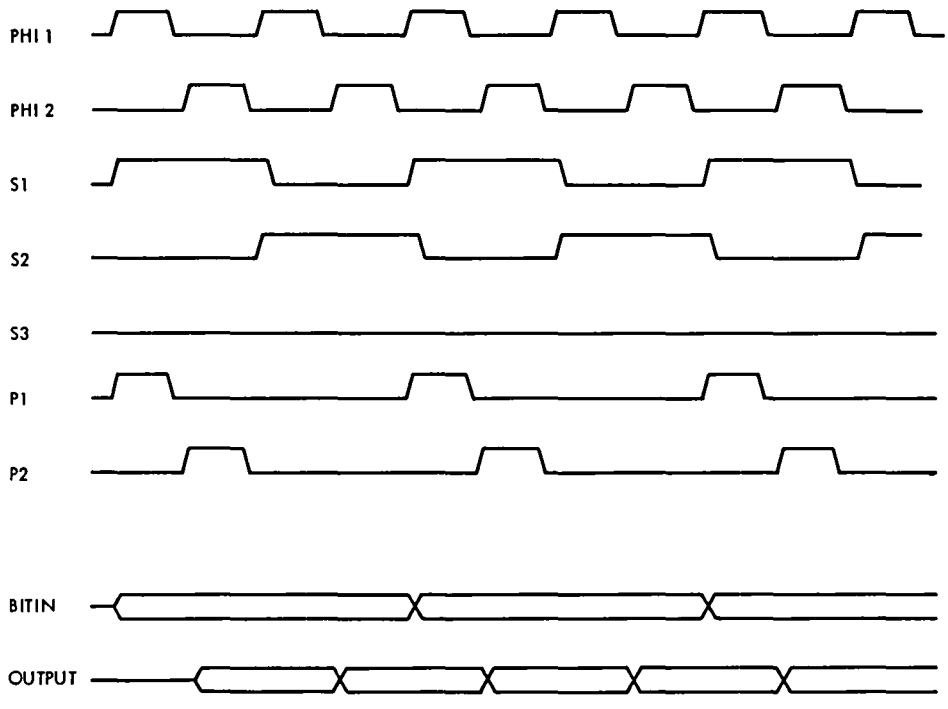


Fig 3 Timing signals for the convolutional encoder, rate 1/2 codes

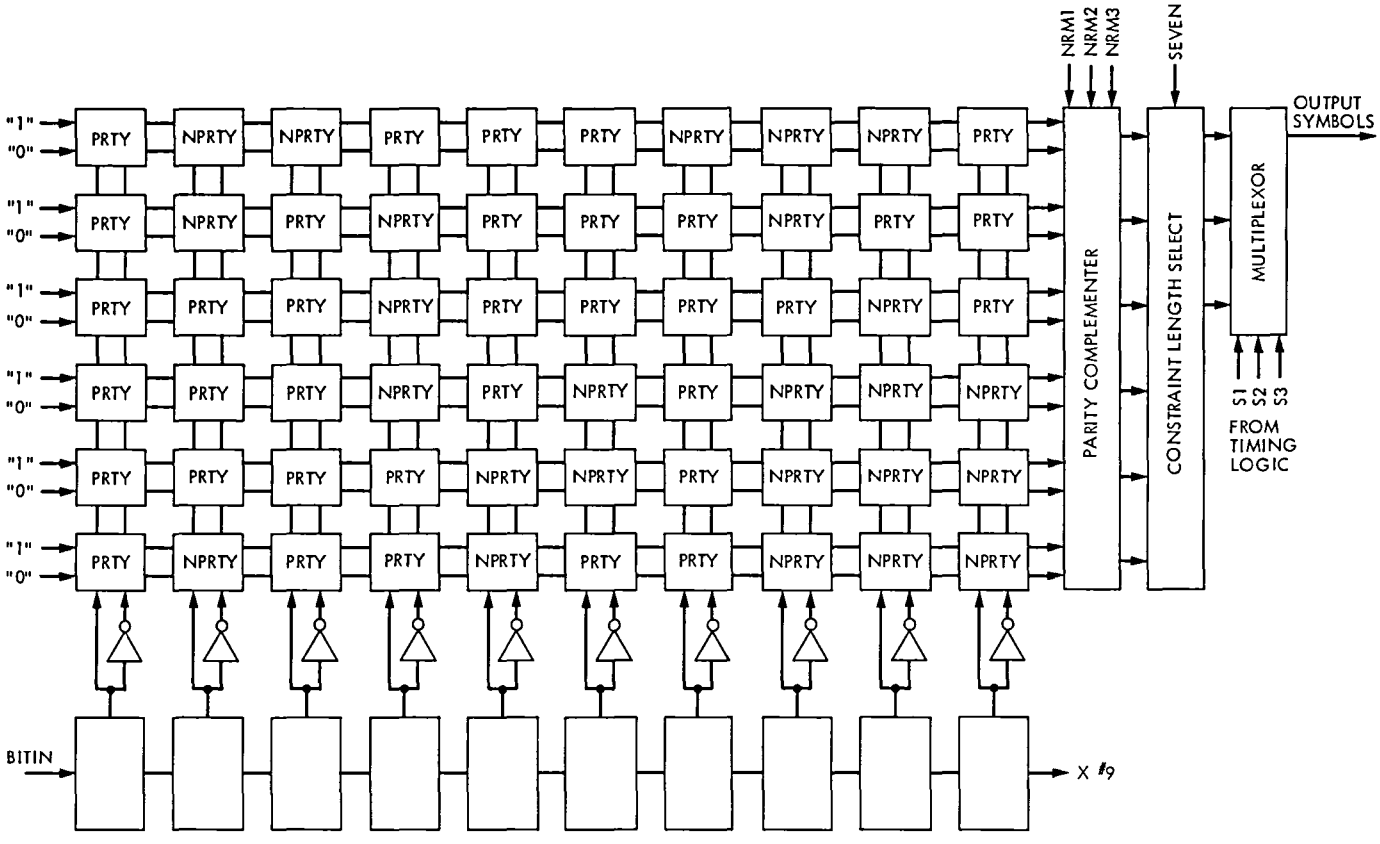


Fig 4 Conceptual diagram of the encoder logic array

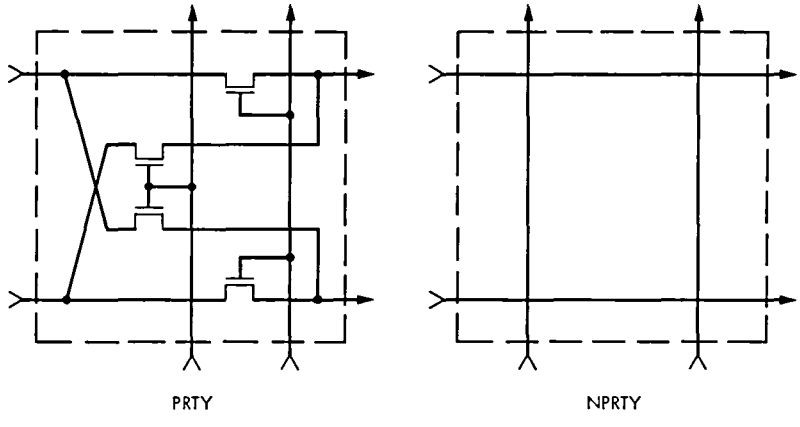


Fig 5 The PRTY and NPRTY cells

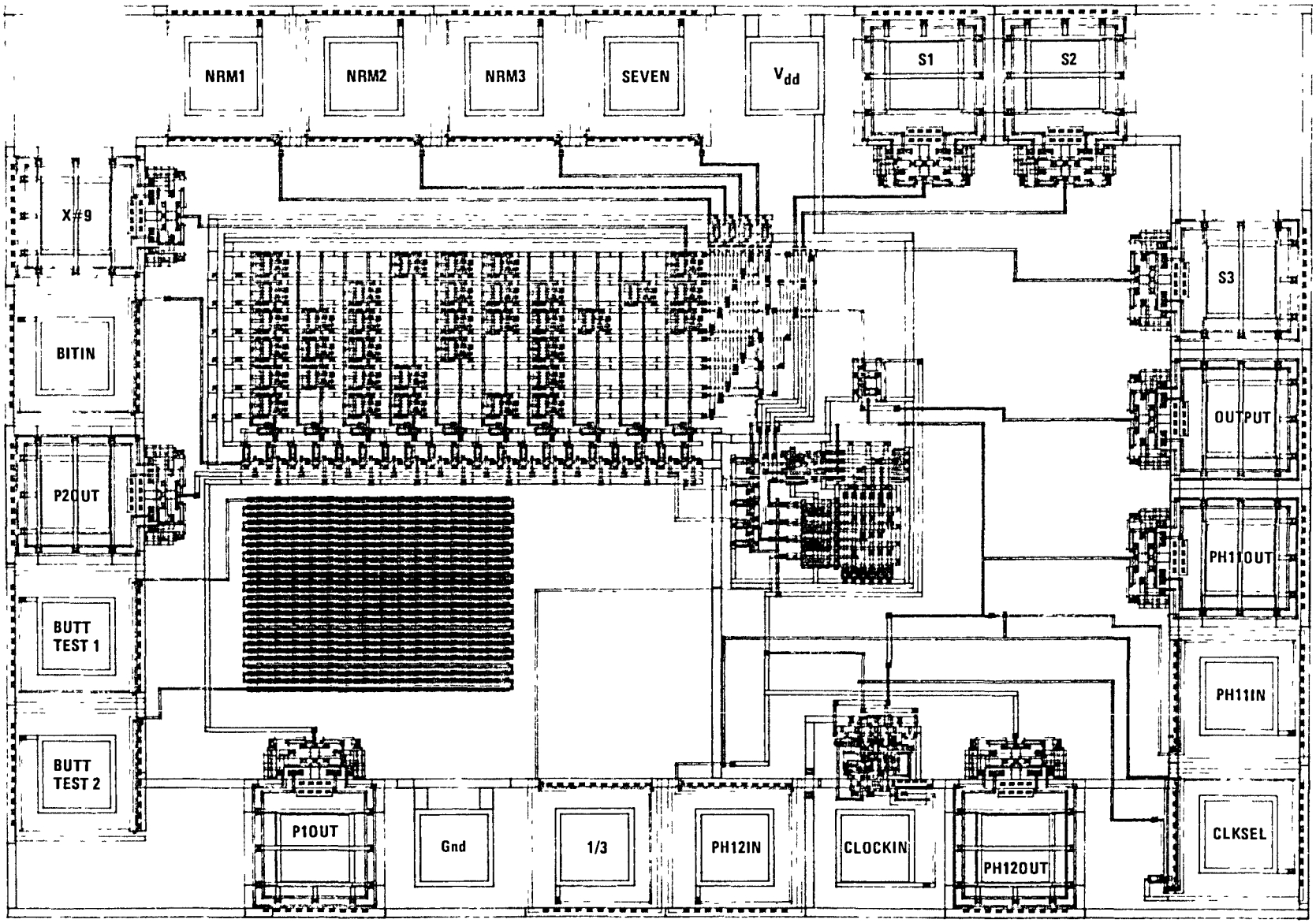


Fig. 6 Layout of the multicode convolutional encoder chip

Staffing Implications of Software Productivity Models

R C Tausworthe
DSN Data Systems Section

This article investigates the attributes of software project staffing and productivity implied by equating the effects of two popular software models in a small neighborhood of a given effort-duration point The first model, the "communications overhead" model, presupposes that organizational productivity decreases as a function of the project staff size, due to interfacing and intercommunication The second, the so-called "software equation," relates the product size to effort and duration through a power-law tradeoff formula The conclusions that may be reached by assuming that both of these describe project behavior, the former as a global phenomenon and the latter as a localized effect in a small neighborhood of a given effort-duration point, are that (1) there is a calculable maximum effective staff level, which, if exceeded, reduces the project production rate, (2) there is a calculable maximum extent to which effort and time may be traded effectively, (3) it becomes ineffective in a practical sense to expend more than an additional 25-50% of resources in order to reduce delivery time, (4) the team production efficiency can be computed directly from the staff level, the slope of the intercommunication loss function, and the ratio of exponents in the software equation, (5) the ratio of staff size to maximum effective staff size is directly related to the ratio of the exponents in the software equation, and therefore to the rate at which effort and duration can be traded in the chosen neighborhood, and (6) the project intercommunication overhead can be determined from the staff level and software equation exponents, and vice versa Several examples are given to illustrate and validate the results for use in DSN implementation

I. Introduction

Brooks (Ref 1), in *The Mythical Man-Month*, proposed a simple model of software project intercommunication to show that, if each task of a large project were required to interface with every other task, then the associated intercommunication overhead would quickly negate the believed advantage of partitioning a large task into subtasks While not meant to be an accurate portrayal of an actual project, the model effectively illustrated an increasing inefficiency symptomatic of projects too large to be performed by a single individual

Putnam (Ref 2), in a 1977 study of software projects undertaken by the US Army Computer Systems Command, discovered a statistical relationship among product *L*ines of code, *W*ork effort, and *T*ime duration for those projects whose best-fit formula was a power-law relationship, now referred to as the "software equation,"

$$L = c_k W^{0.33} T^{1.33}$$

(I have taken the liberty of changing Putnam's notation in order to be consistent with my notation in the remainder of the article)

One rather startling extrapolation one may make from the software equation is that in order to halve the duration of any

one of the projects studied, it would have taken 16 times the resources actually used! I say "extrapolation" because I suspect the software equation is more likely to be applicable incrementally—that is, if one were to require a 5% shortening of the schedule, then a 20% (actually 21 5%) increase in resources would be required

In this paper, I will generalize both of these models parametrically, and suppose that both do describe the statistical trends of software projects in small neighborhoods about a chosen project situation By equating the model behaviors in these neighborhoods, we shall be able to see how the parameters of one model relate to the parameters in the other In addition, we shall discover some rather interesting facts about some actual projects for which published data exists

II. A Generalized Intercommunication Overhead Model

Let us suppose that a software project is to develop L kilo-Lines of executable source language instructions, and that this number remains fixed over all our considerations of effort, duration, staffing, etc That is, we shall suppose that the product size is invariant over the neighborhood of variability in these parameters—a project utilizing greater effort attempting to shorten the schedule slightly would produce the same program as a smaller effort requiring somewhat more time

Let us denote by W the Work effort (in person-months) to be expended in the production of the L lines of code, and let the Time duration, in months, be denoted by T Then the average full-time equivalent Staff size S in persons, is

$$S = \frac{W}{T}$$

and the overall team productivity can be defined as the number

$$P = \frac{L}{W} \text{ (kilo-lines/person-month)}$$

Let us further suppose that the average fraction of time that each staff member spends in intercommunication overhead is dependent on the staff size alone, within a particular organizational structure and technology level, and let this fraction be denoted by $t(S)$

$$t(S) = \frac{\text{(intercommunication time/mo)}}{\text{(hours/mo worked)}}$$

Generally speaking, one intuitively expects $t(S)$ to increase monotonically in S due to the expanding number of potential interfaces that arise as staff is increased But the *individual average productivity* of the staff, defined as the individual productivity during nonintercommunication periods, P_t, is somewhat greater than P, being related to it by

$$P = P_t [1 - t(S)]$$

The relationship between the number of kilo-lines produced, the effort, and the staffing is

$$L = P_t W [1 - t(S)]$$

Let us denote by W_0 and T_0 the effort and time, respectively, that would be required by a single unencumbered individual to perform the entire software task (assuming also that it could be done entirely by this individual, no matter how long it took) Then, with respect to the actual W and T, there is the relationship

$$W_0 = \frac{L}{P_t} = W [1 - t(S)] = T_0$$

This W_0 represents the least effort that must be expended, and T_0 is the maximum time that will be required By substituting W/T for S, one obtains an effort-time tradeoff relationship

$$\omega = \frac{1}{[1 - t(\omega/\tau)]}$$

where $\omega = W/W_0$ and $\tau = T/T_0$ are "normalized" effort and duration, respectively

The rate at which an increase in staffing results in an increase in normalized work effort is then

$$\frac{\partial \omega}{\partial S} = \omega^2 t'(S) > 0$$

where $t'( )$ refers to the derivative of t with respect to S Because of the monotone character of $t(S)$, an increase in staff leads to an increase in effort

The overall staff production Rate R is the number of kilo-lines of code *per month* produced by the entire team of S persons,

$$R = P_t S [1 - t(S)]$$

The factor

$$\eta = [1 - t(S)]$$

is then the *team production efficiency* Note that the normalized task effort is the inverse of the production efficiency,

$$\omega = \frac{1}{\eta}$$

The maximum rate of software production will occur when the derivative of R with respect to S becomes zero, a condition requiring a value S_0 that will satisfy the relationship

$$t'(S_0) = \frac{[1 - t(S_0)]}{S_0}$$

We shall refer to this staffing level as the *maximum effective staff* Two particular examples of $t(S)$ will serve to illustrate the characteristics of the intercommunication overhead model

A. Linear Intercommunication Overhead

Let us assume first, as did Brooks, that the overhead is linear in staff,

$$t(S) = t_0(S - 1)$$

That is, there is no overhead for 1 person working alone, but when there are $S - 1$ other people, then each requires an average fraction t_0 of every other individual's time Under these assumptions, the maximum effective staff level is

$$S_0 = \frac{(1 + t_0)}{(2\, t_0)}$$

This value yields a maximum team production rate of

$$R_{max} = \frac{P_i S_0^2}{(2\, S_0 - 1)}$$

and team efficiency

$$\eta_0 = \frac{(1 + t_0)}{2 = S_0}\,(2\, S_0 - 1) \approx 0\,5$$

This perhaps alarming result states that a team producing at its maximum rate is burning up half its effort in intercommunication overhead! The behavior is illustrated in Fig 1

The normalized effort-duration tradeoff equation for this model takes the form

$$\tau = \frac{t_0\, \omega^2}{(1 + t_0)\, \omega - 1}$$

which has its minimum value at the maximum-production-rate point,

$$\tau_{min} = \frac{4\, t_0}{(1 + t_0)^2} \approx 4\, t_0$$

at which point the normalized effort is

$$\omega_0 = \frac{2}{(1 + t_0)} < 2$$

Figure 2 shows the characteristic of this tradeoff law at t_0 values of 0 1 and 0 2 for illustrative purposes

According to this model, it *never* pays to expend more than twice the single-individual effort Moreover, even though the ω producing the shortest schedule is less than 2, the cost-effective range is much less than this, as shown in the figure Effort can be traded for schedule time realistically only up to about 1 25 W_0, and a factor of 2 reduction in time can only come about if the individual intercommunication can be kept below about 15% per interface

B. Exponentially Decaying Intercommunication Overhead

One unsettling aspect of the linear intercommunication overhead model is that, at some staffing level, the production rate goes to zero, and beyond, unrealistically into negative values Perhaps a more realistic model is one which assumes that $t(S)$ tapers off, never exceeding unity, at a rate proportional to the remaining fraction of time available for intercommunication as staff increases, or

$$t'(S) = t_1\, [1 - t(S)]$$

Then we are led to the form

$$t(S) = 1 - \exp\,[-t_1(S - 1)]$$

The maximum effective staff in this case becomes

$$S_0 = \frac{1}{t_1}$$

and the maximum production rate is

$$R_{max} = P_i\, S \exp\left(-1 + \frac{1}{S}\right) \approx \frac{P_i S}{e}$$

The team effiency at this rate is

$$\eta_0 = \exp\left(-1 + \frac{1}{S}\right) \approx \frac{1}{e}$$

Now this is perhaps even more alarming a revelation than before, because it says that when producing software at the maximum team rate, that team is burning up 63% of its time in intercommunication! The consolation, as shown in Fig 1, is that the team performance under this assumed model is superior to that of the linear-time team model More staff can be applied before the maximum effective staff level is reached

The effort-duration tradeoff equation according to this model is

$$\tau = \frac{t_1 \omega}{[t_1 + \ln(\omega)]}$$

The minimum τ occurs at

$$\omega_0 = \exp(1 - t_1) < e$$

and the minimum value is

$$\tau_{min} = t_1 \exp(1 - t_1) \approx e\, t_1$$

The form of this tradeoff is shown in Fig 3 for t_1 values of 0 1 and 0 2, for illustrative purposes Note that the minimum τ is much broader in this model, so that, although the actual minimum occurs when ω is about e in value, the practical effective range for ω is less than about 1 5 That is, it is not cost-effective to expend more than about 1 5 times the single-individual effort W_0 in an attempt to reduce the schedule time A reduction in schedule by a factor of 2 is possible only when the individual intercommunication factor t_1 can be kept below 0 2

C. Conclusions from Intercommunication Overhead Models

Both of the examples of intercommunication overhead above bespeak a maximum effective staffing level at which the project is 37-50% efficient Beyond this point, further staffing is counterproductive Both examples conclude that the maximum practical extent to which added effort is effective in buying schedule time is limited to about 25-50% Significant schedule reduction factors are possible only when the intercommunication factors can be kept below 15-20%

III. Matching the Software Equation Model

Let us generalize the Putnam Software Equation as the form

$$L = c_k W^p T^q$$

and let is define $r = q/p$, the exponent ratio As in the previous section, L is held constant with respect to effort-duration tradeoff considerations The value of p is assuredly positive it generally requires more work at a given L to reduce T If q is positive, effort can be traded to decrease the schedule time required to deliver a given L The larger r is, the larger the increase in effort required to shorten the schedule, and the larger the team production inefficiency If q is zero, then L is a function of W alone, T is determined solely by the staffing level, $T = W/S$, and no additional effort is required to reduce schedule time (in the neighborhood in which the p and $q = 0$ are valid) If q were ever to be negative, then an increase in W would render an increase in T, a situation indicating overmanned projects

Substitution of $T = W/S$, differentiation with respect to S, and normalization of the software equation produces the result

$$\frac{\partial \omega}{\partial S} = \omega \frac{r}{[S(1+r)]} = \tau \frac{r}{(1+r)}$$

Let us now suppose that both the software equation and the intercommunications overhead model agree at the point (L, W, T) The two models can be equated by suitable choices of the "technology constant" c_k and individual productivity P_i Then, in addition, let us suppose that the derivatives of effort with respect to staff level for both models also agree at this point Such can only be attempted when $r > 0$, because the derivative in the intercommunication overhead model is always positive When this is the case, the two models may be said to agree in the neighborhood of the point (L, W, T)

Thus, by equating the derivatives, we arrive at a relationship between the parameters of the two models

$$\frac{S\, t'(S)}{[1 - t(S)]} = \frac{r}{1+r}$$

or

$$\eta = S\, t'(S)\, \frac{r+1}{r}$$

Let us now examine this relationship for the two examples of the interface overhead model

A. Linear Intercommunication Overhead

Substitution of the linear $t(S)$ form into the neighborhood agreement condition yields

$$S = \left[\frac{2r}{1+2r}\right]\left[\frac{1+t_0}{2t_0}\right] = S_0 \frac{r}{r+0.5}$$

This equation states that the staffing level is related to the maximum effective staff point through the software exponent ratio r. At the Putnam value, $r = 4$, the staffing level is 89% of the maximum effective level, and the team efficiency is

$$\eta = 0.55(1+t_0) \approx 55\text{-}65\%$$

$$\omega = \frac{1.8}{1+t_0} \approx 1.5 - 1.8$$

As seen in Fig. 2, projects having this high an ω are at the point that extra effort is very ineffective.

B. Exponentially Decaying Intercommunication Overhead

By substituting the exponential form for $t(S)$ into the neighborhood agreement condition, we find

$$S = \frac{r}{[t_1(1+r)]} = S_0 \frac{r}{(1+r)}$$

Again, we see that the staffing level is related to the maximum effective staff via the exponent ratio. The Putnam value $r = 4$ produces

$$S = 0.8 S_0$$

$$\eta = \exp\left[-\frac{(S-1)}{S_0}\right] = \exp[-0.8 + t_1] \approx 45\% - 55\%$$

$$\omega = \frac{1}{\eta} = \exp[0.8 - t_1] \approx 1.8 - 2.2$$

Although this example indicates a somewhat more comfortable margin below maximum effective staffing than did the linear model, it nevertheless shows an alarmingly low cost inefficiency.

IV. Examples Using Available Data

Several data sets of project resource statistics published in the literature readily show that Putnam's value of $r = 4$ is not universal. Specifically, Freburger and Basili (Ref. 3) publish data which yield the following 3-parameter best power-law fits

$$L_0 = 1.24\, W^{0.95}\, T^{-0.094} \qquad (r = -0.1)$$

$$L_1 = 0.22\, W^{0.78}\, T^{0.78} \qquad (r = 1.0)$$

in which L_0 is kilo-lines of delivered code, and L_1 is *developed* delivered code. It is interesting here to note that the former relationship is nearly independent of T, whereas the latter shows a definite beneficial $W - T$ tradeoff characteristic. The negative q in the former relationship indicates that, on a *delivered code basis*, added resources in one of the projects would have extended the schedule! An equivalence between the software equation and the intercommunication overhead model cannot be established when r is zero or negative.

This data set is not the only one to show a negative q. Boehm (Ref. 4), in his *Software Economics* book, has a data base used to calibrate his COCOMO software cost model. A 3-parameter best power-law fit to the adjusted data produces the relationship

$$L = 0.942\, W^{0.675}\, T^{-0.028} \qquad (r = -0.041)$$

and, on the unadjusted data,

$$L = 0.957\, W^{0.646}\, T^{0.0555} \qquad (r = 0.086)$$

Gaffney (Ref. 5), on the other hand, did a 3-parameter best power-law fit of IBM data (Federal Systems Division, Manassas) to arrive at the relationship

$$L = c_k\, W^{0.63}\, T^{0.56} \qquad (r = 0.88)$$

This last value of r aligns more closely with the Freburger-Basili value for developed delivered code.

Figure 4 shows plots of staffing normalized to maximum effective levels as a function of r, for both examples of the intercommunication overhead model. Efficiency curves are shown in Fig. 5.

V. Conclusion

This article has shown that when there is a positive effort-duration tradeoff relationship in a software project, it is possible to estimate the team production efficiency and proximity to maximum effective staffing. These figures can be

used to advantage by software managers who must judge the effectiveness of increasing resources in order to shorten schedules It points out the necessity of keeping accurate records of software project statistics, so that the parameters in the model can be estimated accurately

Low values of r in an organization are a mark to be proud of, showing efficiency in terms of structuring subtasks for clean interfaces High (or negative) values of r may be indicative of overall task complexity, volatility of requirements, organizational inefficiency, or any number of other traits that tend to hinder progress The value of r may thus be treated as a figure of merit—a measurable statistic indicative of the efficiency of a set of projects in performance of assigned tasks

The ratio S/S_0 is another indicator for management When low, it indicates that adding resources can potentially help a project in schedule trouble If closer to unity, it is a warning that adding resources may not help, will not appreciably shorten the schedule, will incur expense at a low return in productivity, and, if applied often in other projects, will thereby contribute to an organizational reputation for expensive software

References

1 Brooks, F P , *The Mythical Man-Month*, Addison-Wesley Publishing Co , Reading, Mass , 1975

2 Putnam, L H , "Progress in modeling the software life cycle in a phenomenological way to obtain engineering quality estimates and dynamic control of the process," *Second Software Life Cycle Management Workshop*, sponsored by U S Army Computer Systems Command and IEEE Computer Society, Atlanta, Ga , Aug 1978

3 Freburger, K , and Basili, V R , "The Software Engineering Laboratory, Relationship Equations," Report TR-764, University of Maryland Computer Science Center, College Park, Md, May 1979

4 Boehm, B W , *Software Economics*, Prentice-Hall Publishing Co , Englewood Cliffs, N J , 1982

5 Gaffney, J E , "An Approach to Software Cost and Schedule Estimation," submitted to *Journal of Defense Systems Acquisition Management* (pending)

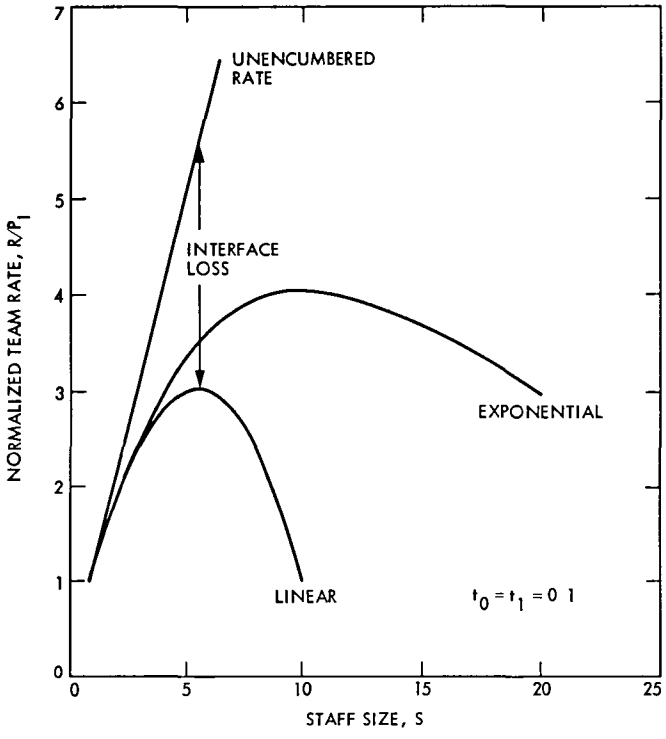


Fig 1. Software team production rate, intercommunications overhead model

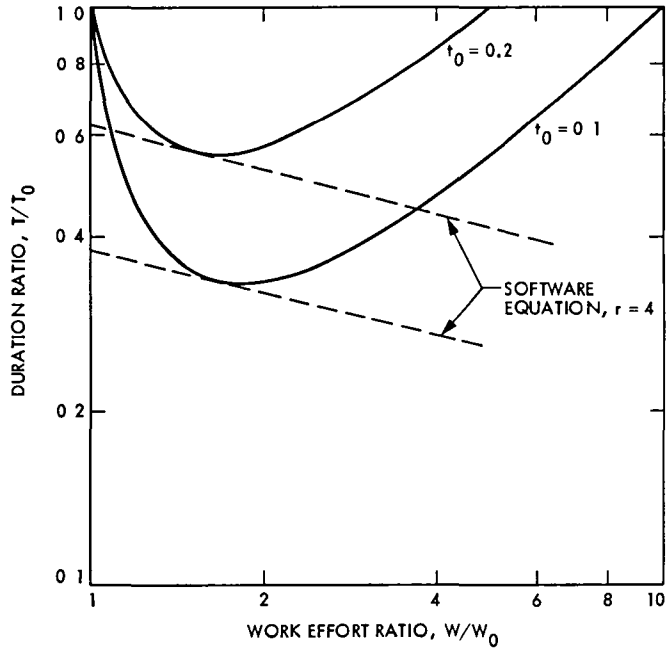


Fig. 2. Time duration vs effort tradeoff curve for linear intercommunication overhead model

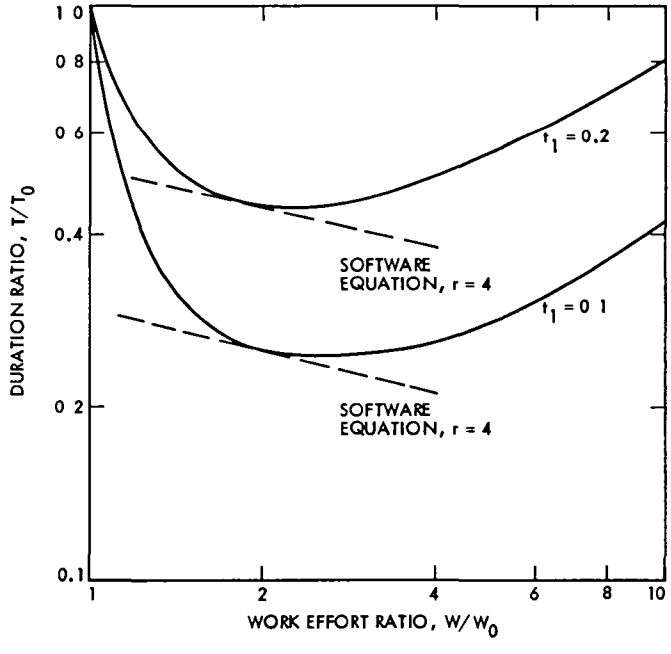


Fig 3. Time duration vs work effort tradeoff characteristic for the decaying exponential intercommunication overhead model

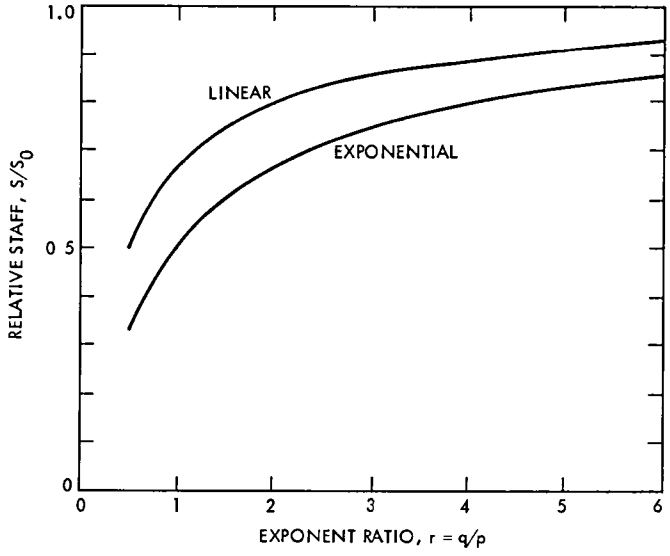


Fig. 4. Normalized staff levels for the intercommunication overhead model

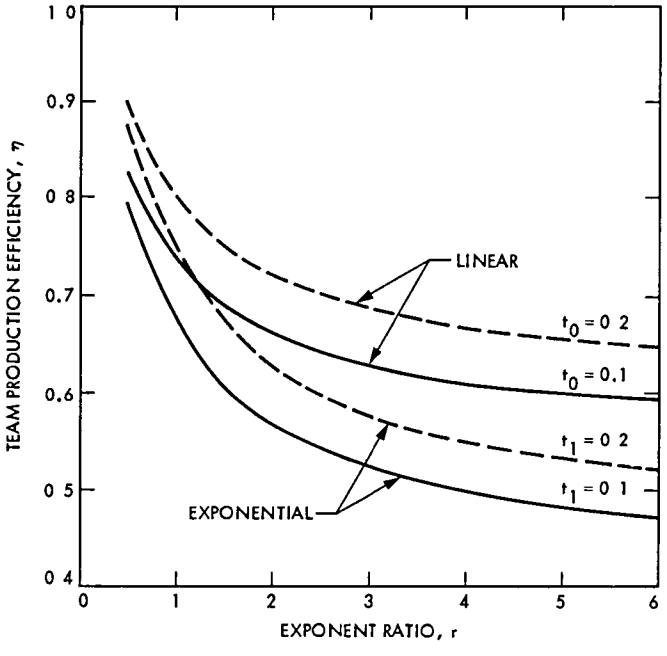


Fig 5 Team production efficiency of projects as a function of the exponent ratio and intercommunications overhead parameter

Applications of Telemetry Arraying in the DSN

R Stevens
Chief Engineer, Telecommunications and Data Acquisition

The individual large antennas at a DSN complex have been used in arrayed configurations to improve the telemetry return from very distant spacecraft The technique has been applied to support critical spacecraft events when good telemetry is most important An expanded array configuration, using the Parkes, Australia, Radio Astronomy Observatory will be provided to support the 1986 Voyager encounter with Uranus This report reviews the development and use of arraying by the DSN for spacecraft mission support Application of the experience to the 1986 design is discussed

I. Introduction

For support of the 1986 Voyager 1 encounter with Uranus, the DSN is planning to array telemetry signals from the Parkes, Australia, Radio Astronomy Observatory 64-m antenna with those of the 64-m and two 34-m antennas of the Tidbinbilla DSN complex During the process of assessing the array configuration system design for that application, the previous DSN operational experience with arrayed configurations was reviewed The purpose of the review was to ensure that lessons from experience were not neglected

The operational applications of arraying for telemetry reception by the DSN have been described in past TDA Progress Reports The reporting covers all applications, except for the Voyager 2 Saturn encounter, which will be reported in a future issue This material is an abstraction of those reports and of discussions with some of the authors It also includes discussion of anticipated new factors involved with the integrated participation of the Parkes Facility for critical scientific deep space mission support

II. Arraying Phase 0

In 1970, conceptual levels of combining were described and their basic performance analyzed The levels described were carrier, baseband, bit stream, and processed data Baseband and bit stream combining were configured with two 26-m antennas and demonstrated using the signals from the Pioneer 8 spacecraft The results showed approximately 2 5 dB (vs 3 dB maximum possible) improvement with baseband and bit stream combining Time delay tracking of the baseband signals was not required or used because of the low subcarrier frequency of the Pioneer signals The work was done by the personnel of the Spanish complex using available station equipment and the DSS 61/62 26-m antennas (Ref 1)

The next effort was demonstration of RT baseband combining of high-rate telemetry via microwave links of the signals from the DSS 12 and DSS 13 26-m antennas with the signal from DSS 14 (Ref 2) The array was used to enhance the telemetry signals from the MVM spacecraft during its September 1974 second Mercury encounter The system was

designed, built, and operated by Advanced Systems Program personnel with support from Operations personnel Installation and initial testing of the array system at Goldstone was straightforward Array performance was evaluated using spacecraft signals from far encounter TV sequences These tests indicated that the performance of the Subcarrier Demodulation Assembly (SDA) processing string, which was hard-wired to the array combiner signal stream, performed about 1/2 dB poorer than the string which was hard-wired to the 64-m-only signal stream In the time available, it was not practical to alter or improve the situation As a result, during the encounter pass, the arrayed signal stream was only 0 3–0 4 dB improved over the 64-m-only signal stream, rather than the ≈0 7 dB, which was in principle achievable The ≈0 7 dB (vs 0 8 theoretical) was demonstrated by deleting briefly the 26-m antenna signals from the combiner input and observing the increase in image data bit error rate

During a portion of the encounter pass over Goldstone, the spacecraft transmitted at maximum data rate (117 kbs vs 22 kbs), which was only supportable by the arrayed configuration Although the full 0 7 dB capability of the array could not be exploited, the 0 3-0.4 dB improvement did significantly enhance the mission imaging data yield The experience showed the difficulty in achieving full performance of an arrayed system, and the importance of a few tenths of a dB of performance

III. Arraying Phase I

The 1973-74 R&D demonstration project just described established technical feasibility of baseband arraying of very weak high-rate signals In 1977, the DSN started to develop an operational arraying capability for the network The first application was for Voyager encounters of Jupiter and Saturn The first phase was providing a prototype system at Goldstone to be used for enhancement of the Jupiter encounters (Refs 3, 4)

A prototype baseband real-time combiner (RTC) based on the analysis and design techniques developed by the earlier R&D activity was completed in the fall of 1978 It was designed to combine signals from DSS 14 and DSS 12 The signals from DSS 12 were sent to DSS 14 via the existing microwave link The RTC was installed at the Compatibility Test Area (CTA 21) for initial testing The tests were conducted in November and December 1978 The tests indicated that the RTC performed according to its analytical model Conducting the test program at CTA 21 was surprisingly difficult and time consuming, finally, the equipment in the telemetry strings had to be carefully realigned and recalibrated and the test data taken at night to avoid electrical transients Most of the problems encountered were external to the RTC—in the receiver/telemetry strings and test instrumentation

The RTC was installed at DSS 14 in mid-December 1978 After checkout and alignment, the system was tested by receiving Voyager signals. Performance on the initial Voyager tests in December was inconclusive After further off-line testing and alignment of the system, array gain on Voyager signals, ranging from 0 4 to 1 0 dB, was obtained in mid-February 1979 With intensive work by the design and operations personnel, the array system was brought on-line for the Voyager 1 Jupiter encounter during the first 10 days in March It was brought up and used again for the Voyager 2 Jupiter encounter for two weeks in early July 1979

The performance of the prototype array system during the Voyager Jupiter encounter operations was good It provided approximately 1 0 dB S/N improvement vs the 64-m-only, with an availability of ≈0 95. The Voyager Project used the data from the array in preference to the on-line alternate 64-m-only stream for most of each tracking pass

The prototype array system was next configured to provide enhanced support of the early September 1979 encounter of Saturn by the Pioneer 11 spacecraft (Ref 5) Pioneer 11 had two important data rate options for the encounter/near encounter period, 1024 and 512 bps The higher rate was greatly desired, but its support appeared marginal To enhance the support capability, the DSN installed ultra-low noise listen-only front ends and selected the best performing of normally equal telemetry strings at the 64-m stations, also, it adapted the DSS 14-12 RT array for Pioneer S-band parameters

At S-band, the S/N performance of the DSS 12 34-m system relative to the DSS 14 64-m system is poorer than at X-band As a result, the potential arraying gain was only 0 5 dB for Pioneer vs approximately 1 0 dB for Voyager Also, the DSS 12 Pioneer signals were so weak as to not support automatic tracking of their time delay by the RTC The varying time delay between the DSS 14 and DSS 12 signals during a pass was calculated and input to the RTC (that open-loop mode was provided for in the prototype RTC design, and it worked satisfactorily on the relatively narrow baseband Pioneer signal)

The initial tests in early August of the array on the Pioneer signals were not successful An intensive effort was required to resolve the performance problems with the array system On August 25, the first day of planned mission support by the array, it worked well It provided 0 4 to 0 5 dB performance improvement relative to the 64-m-only configuration and supported the 1024 bps rate On the next pass, it was no better than the 64 m's On the following pass, after

inconclusive testing and diagnosis, it was back working properly

The 10-day encounter sequence was hectic for the Pioneer Project and the DSN, unexpected solar noise and satellite RFI complicated the data rate strategy The efforts to maximize imaging data (by using 1024 bps) was impacting non-imaging experiments when the rate was unsupportable By two days before encounter, the array performance had stabilized fairly well This allowed use of the 1024 rate during mid-pass over the Goldstone array, 512 was used at other times and places

The experience on Voyager and Pioneer with the prototype RTC and array configuration showed the need for central control and monitoring of the array system, and again, the importance of having *all* elements of the array—receivers, antennas, instrumentation—operating completely correctly Also, array performance testing on live spacecraft signals was next to impossible if the spacecraft and the ground stations were not in stable configurations dedicated to supporting the array tests—array testing on a noninterference basis was usually unproductive This experience was applied to the design of the equipment and use procedures for the next phase of arraying system development

IV. Arraying Phase II

The objective of this phase was to provide operational arraying system configurations to combine the signals from the 64-m antenna and the 34-m antenna at each of the three DSN complexes The arrayed systems were committed for operational support of the Voyager 1 (fall, 1980) and Voyager 2 (summer, 1981) encounters with Saturn (Ref 6)

The design of the operational RTC's and arraying system was a direct upgrade of the Goldstone prototype just described Modifications to the system and its operating procedures were based on the experience with the Goldstone prototype The principal modifications to the system included

(1) Displays of 34-m station receiver in-lock and AGC were provided at the 64-m stations

(2) The Block IV receiver telemetry detectors were modified to improve performance and stability

(3) A routine procedure for measurement of microwave link path delay was developed (DSS 12-14 use)

(4) Pretrack countdown time for the RTC's was reduced

(5) The RTC was modified so that it could be restarted or accept new parameter inputs, without requiring reinitialization

(6) The procedures for testing the array performance on spacecraft signals were improved

(7) Simultaneous displays of symbol SNR for the 64-m-only, 34-m-only and the arrayed signals were provided at both the 64-m station and at the JPL network advisor's position

The first operational RTC was installed at Goldstone in mid-May 1980 Testing of the DSS 14-12 array on Voyager signals was started on May 19 Eight performance tests and training sessions were conducted in the following month's time Installation of the systems at the overseas stations was then done, followed by a similar test and training sequence The process of bringing up the operational systems was difficult, but much easier than the previous experiences with the prototype system

Beginning in mid-August 1980, the arrayed configurations were used at all stations at regular intervals for Voyager 1 Saturn far-encounter operations support, and for DSN test and training For the near-encounter period, the arrayed configuration was used for all passes (Ref 7)

During the first several days of the period the array performance of the complexes was only "fair" At that point (November 1, 1980, 12 days before encounter) a procedure for assessing array system performance at regular frequent intervals was established The procedure was at 1-hour intervals the JPL network advisor computed a 10-minute average of the symbol SNR from the 34-m, the 64-m, and the arrayed signal streams of the active DSN complex

The "observed" arraying gain was calculated as SNR (array)/SNR (64-m), the "theoretical" or "loss-free" arraying gain was calculated as 1+ SNR (34-m)/SNR (64-m) If the ratio of the two gain numbers exceeded 0 3 dB, the nominal combining loss plus tolerance, the DSN complex was alerted to look for possible system problems That procedure was used on all remaining passes through encounter, and good stable array performance was achieved The average array gain for the encounter period was 0 62 ±0 15 dB relative to the 64-m-only and the array availability was approximately 0 95 (Cf Ref 7)

In preparation for the late August 1981 Voyager 2 Saturn encounter, array system test and training was started by the Network in early April (Ref 8) The work went fairly smoothly It was supported by arrayed stream, 64-m-only stream, and 34-m-only stream symbol SNR performance monitoring in real-time at the complexes and at JPL

All complexes supported the beginning of the observatory phase on June 5, 1981, in the arrayed configuration, with

very good results Array training and performance validation tracks on the Voyager signals were conducted at regular intervals until the start of near-encounter operations

During the 30-day near-encounter period, the arrayed configuration was used extensively The procedure of hourly collection of 10-min average symbol SNRs and calculation of the combining loss was used throughout the 30-day period It was the basic procedure developed during the Voyager 1 Saturn encounter. To reduce operations staffing required, the 10-min averages were automatically calculated in the station Telemetry Processor Assemblies (TPAs) and reported verbally to the Network Analysis Team at JPL. Also, the near-real-time (within 10 minutes) image display at the stations was found to be useful in keeping the arraying performance near optimum at all times

The results during the Voyager 2 Saturn encounter were very satisfactory, clearly the best yet achieved. The *network* average of observed array gain relative to 64-m-only was 0 8 dB with a standard deviation of 0 2 dB The *network* average combining loss was 0 2 ±0 05 dB Those results are consonant with the design predictions/specifications Out of a total of 79 passes scheduled for arraying support during the near-encounter operations, satisfactory array capability was provided for 75 of them, giving an availability of ≃0 95 These above results are from a preliminary analysis of operations data

The experience with the operational RTC and 64-m–34-m arraying systems for the Voyager 1 and 2 Saturn near-encounter operations showed the following

(1) The array system works methodically according to its design principles

(2) Array availability data are sparse and subject to interpretation, but some conclusions can be drawn Downtimes are the sum of the 64-m string, the 34-m string and the combining system downtimes in that order of contribution, a logical result The availability of the total array from about 7 weeks of encounter operations is ≃0 95, not a super good result, but probably representative of the present system capability

(3) Extensive real-time visibility into the performance of the individual strings and the total system has been provided by the operations organization during the evolution of the use of the arrayed configurations It is concluded that was done because it was needed

(4) Bringing up and validating a new array configuration is difficult but can be systematically accomplished, culminating with 6-10 scheduled spacecraft live data exercises

V. Arraying Phase III and Beyond

The next application of DSN arraying will be support of the Voyager 2 Uranus encounter in January 1986 At the Australia and Goldstone complexes, the 64-m and two 34-m's will be arrayed, at Madrid, the 64-m and the single 34-m will be arrayed Also, at Australia, it is planned to array the Parkes 64-m with the DSN complex The Parkes facility is about 280 km distant from the DSN Tidbinbilla complex.

The design and the operational use planned for the 1985-86 arraying system for the DSN complexes is based on the previous configurations It will be baseband with RT-only combining at the DSN complexes The combining method, RT-only, NRT-only or hybrid, for the Parkes signals is being reviewed–the current baseline design includes both RT and NRT

A new Baseband Assembly (BBA) is being designed for the DSN complexes The BBA will provide the functions of the RTC, the SDA, and the Symbol Synchronizer Assembly (SSA) in a single integrated unit The new design is expected to improve the operational performance, stability, and monitor capability of the combiner/telemetry channel

Also, the Mark IVA DSN will incorporate improvements in the antenna pointing system and monitor and control of the station equipment Without these kinds of improvements, the availability of the planned three-element DSN arrays would clearly be less than the ≃0 95 of the present two-element configurations The design availability for the Mark IVA Australian DSN complex 64-m − 2 × 34-m array is ≃0 97, an ambitious objective

In the baseline design for the planned Parkes-Tidbinbilla array, both RT and NRT combining configurations are provided At critical times they would be used in parallel The basic function of the NRT capability is for backup or fail-soft protection to the RT microwave link failure node. The basic technical specification on the NRT configuration is to provide records from Parks and Tidbinbilla which can be played into the RTC with degradation ≤0 2 dB That is the same performance specification as for the RT link.

The DSN is developing a recording and playback system to meet the specification in a station environment It uses the Mark III data acquisition terminals (DATS) with special signal processing electronics to accommodate the weak telemetry signals The Mark III DATS is used by the DSN and many radio observatories for VLBI recording

DSN engineers have had experience in the development and application of sophisticated recording systems (e g , VLBI

recording and playback and the recording and playback systems for the Pioneer Venus multiprobe and DLBI wind experiment support), there is high confidence in the NRT telemetry system development A software simulation of the output spectrum reconstruction electronics has been successfully run That is an important accomplishment The recording system will be demonstrated in the fall of 1983 Output combining of two recorder systems through an RTC will be demonstrated in the spring of 1984, field demonstration at Goldstone will be in the summer of 1984.

The plan to include Parkes with the Tidbinbilla array presents new technical and managerial challenges to the DSN Compared to the Goldstone DSS 14-12 configuration, Parkes is truly remote from the Tidbinbilla complex Its expected contribution to the array performance is next in significance to the DSN 64 m. Its availability, as currently predicted, will be significantly less than for a DSN array element, so that its proper operational use is expected to be more difficult The use of Parkes will be the DSN's first experience with a non-NASA owned facility tightly coupled with critical mission support Hence, the operational procedures and interfaces will be relatively immature compared to those that have been brought to bear on past arrayed mission support operations

The GIOTTO Project will use the Parkes facility for prime support of its mission, and it will use the DSN Tidbinbilla complex for backup support The GIOTTO spacecraft encounter with Halley's Comet occurs in mid-March 1986, following the Voyager 2 Uranus encounter by less than two months This is perceived as an advantage (perhaps an enabler due to shared facility and equipment use at Parkes), and at the same time, an operational complication in the use of Parkes for Voyager support While there does not appear to be mutual conflict of critical project schedules at Parkes and Tidbinbilla during late 1985 and early 1986, they are interlaced without much contingency during January and February However, if the two spacecraft and the ground systems behave reasonably well during that period, the continuous usage of the ground facilities under tight configuration management could benefit both projects and the DSN

References

1 Urech, J M , "Telemetry Improvement Proposal for the 85-Foot Antenna Network," *Space Programs Summary 37-63, Vol II*, Jet Propulsion Laboratory, Pasadena, Calif , May 31, 1970.

2 Winkelstein, R A , "Analysis of the Signal Combiner for Multiple Antenna Arraying," *DSN Progress Report 42-26*, Jet Propulsion Laboratory, Pasadena, Calif , pp 102-118, April 15, 1975

3 Simon, N K , and Hoynes, C , "Preliminary Telemetry Operations Experience With the Real-Time Combiner 1 November 1978 to 1 November 1979," *DSN Progress Report 42-55*, Jet Propulsion Laboratory, Pasadena, Calif , Feb 15, 1980

4 Howard, L D , "Prototype Real-Time Baseband Signal Combiner," *TDA Progress Report 42-60*, Jet Propulsion Laboratory, Pasadena, Calif , Dec 15, 1980

5 Miller, R B , "Pioneer 11 Saturn Encounter Mission Support," *DSN Progress Report 42-54*, Jet Propulsion Laboratory, Pasadena, Calif , Dec 15, 1979

6 Fanelli, N , and Nance, H , "Voyager Mission Support (I)," *DSN Progress Report 42-60*, Jet Propulsion Laboratory, Pasadena, Calif , Dec 15, 1980

7 Bartok, C D , "Performance of the Real-Time Array Signal Combiner During the Voyager Mission," *DSN Progress Report 42-63*, Jet Propulsion Laboratory, Pasadena, Calif , June 15, 1981

8 Fanelli, N , and Nance, H , "Voyager Mission Support," *DSN Progress Report 42-64*, Jet Propulsion Laboratory, Pasadena, Calif , Aug 15, 1981

Antenna Arraying Performance for Deep Space Telecommunications Systems

C T Stelzried and A L Berman
TDA Mission Support Office

G K Noreen
Telecommunications Systems Section

Antenna arraying will be a crucial Deep Space Network technique in maximizing the science return of planetary and comet encounters in the 1980's This article develops the equations which describe the total figure of merit for a multiple system of arrayed antennas An example is given for three Canberra DSN antennas and the Parkes 64-m antenna to be arrayed for the Voyager 2 Uranus flyby

I. Introduction

As the Deep Space Network prepares for the decade of the 1980's, the science return for critical spacecraft encounters such as Voyager 2 at Uranus and Neptune and GIOTTO at Halley's Comet is expected to be significantly enhanced through the use of multiple antenna arraying Antenna arraying concepts were developed in the 1960's for radio astronomy and were successfully demonstrated for space telecommunications with Pioneer 8 (Ref 1) and during the Mariner 10 Mercury encounter in 1974 (Refs 2 and 3) and the Voyager 2 Saturn encounter in 1981 (Ref 4)

The figure of merit of a single aperture receiving system is given by (Ref 5)[1]

$$M = G_R/T_{op} \ , \ \mathrm{K}^{-1} \tag{1}$$

where

G_R = receiving antenna gain, ratio

T_{op} = system noise temperature, K

A profile of the figure of merit of the Deep Space Network (DSN) receiving systems from 1960 to 1982 is shown in Fig 1

The figure of merit required for a receiving system to support a given communication link is (Ref 6)

$$M \geq 4\pi kRLD^2 \, (E_b/N_0)_T/P_T A_T \tag{2}$$

where

D = distance between the transmitting and receiving antennas, m

R = data bit rate, bps

A_T = effective area of transmitting antenna, m^2

[1]For convenience, M is usually used in computations as shown with units K^{-1} and in discussion with units of dB where M (dB) = 10 log M, it is understood that this is relative to a reference figure of merit of 1 K^{-1}

L = total link losses (includes polarization loss, pointing loss, atmospheric loss, demodulation loss, etc), ratio

$(E_b/N_0)_T$ = threshold value (Ref 5) of (E_b/N_0) for a given bit error rate (BER), ratio

P_T = transmitter power, W

k = Boltzmann's constant, 1 3806 × 10^{-23}, J/K

For example, Voyager 2 at Saturn had the parameter values shown in Table 1 for transmitting imaging data The threshold figure of merit M_T from Eq (2) is calculated to be 55 dB This results in a communications link margin of 3 dB (Fig 1) This link margin is based on mean parameter values, adequate planned margin is required to allow for variations in the actual parameter values, particularly the change of T_{op} as a function of weather

To meet future requirements for Voyager 2 communications at Uranus and Neptune, it will be necessary to improve the figure of merit with a system of arrayed apertures This report reviews the equations which describe the figure of merit for a multiple array of antennas

II. Antenna Arraying

The figure of merit required for the receiving system can be obtained with an array of antennas each with a separate figure of merit ($M_i = G_i/T_i$) Refs 1, 2, and 3 developed the array performance analysis and optimum combiner strategy in terms of the receiver-detected SNR The array figure of merit could be deduced from this analysis or derived directly For completeness, an analysis for M in terms of M_i follows

Assume an array of antennas (Fig 2) where the voltage outputs of the individual channels are weighted by β_i and optimally combined[2] to maximize the output signal-to-noise ratio (SNR) Using subscript 1 for the "reference" channel, summing the coherent signal voltages ($\sqrt{G_i g_i}\ \beta_i$) and the assumed incoherent[3] noise powers ($T_i g_i \beta_i$),

$$\frac{S(\beta_i)}{S_1} = \frac{\left(\sum_{i=1}^{n} \sqrt{G_i g_i}\ \beta_i\right)^2}{G_1 g_1 \beta_1^2} \tag{3}$$

and

$$\frac{N(\beta_i)}{N_1} = \frac{\sum_{i=1}^{n} T_i g_i \beta_i^2}{T_1 g_1 \beta_1^2} \tag{4}$$

where

$S(\beta_i)/S_1$ = arrayed received signal power as a function of β_i, relative to the reference channel, ratio

$N(\beta_i)/N_1$ = arrayed noise power output as a function of β_i, relative to the reference channel, ratio

G_i = gain of ith receiving antenna, ratio

g_i = gain of ith receiver, ratio

T_i = operating system noise temperature of ith antenna, K

β_i = voltage weighting function of ith channel, ratio

Then the signal-to-noise ratio as a function of β_i is given by

$$SNR\ (\beta_i) = \frac{(SNR)_1}{M_1} \frac{\left(\sum_{i=1}^{n} \sqrt{G_i g_i}\ \beta_i\right)^2}{\sum_{i=1}^{n} T_i g_i \beta_i^2} \tag{5}$$

The value for β_i to maximize the SNR can now be obtained

Differentiating with respect to β_i and setting the result equal to zero,[4]

[2]This analysis assumes perfect combining and does not allow for unequal atmospheric losses, combiner losses, differential time delays, and other nonideal effects

[3]The assumption of incoherent noise cannot be made for that portion of received noise that is due to a hot body in view of all elements of the array

[4]The power weighting is given by

$$(\beta_i/\beta_1)^2 = \left(\frac{T_1}{T_i}\right)^2 \frac{G_i g_1}{G_1 g_i} = \frac{T_1 g_1}{T_i g_i} \frac{(SNR)_i}{(SNR)_1}$$

This can be obtained (Ref 7) in an operational system by first setting each channel receiver gain for equal noise level ($g_i = g_1 T_1/T_i$) and then weighting each channel by an additional factor $(SNR)_i/(SNR)_1$ Techniques are available (Ref 8) for monitoring $(SNR)_i$.

$$(\beta_i/\beta_1) = (T_1/T_i)\sqrt{G_i g_1/G_1 g_i} \quad (6)$$

and

$$SNR = \frac{(SNR)_1}{M_1} \sum_{i=1}^{n} M_i \quad (7)$$

Using $SNR/(SNR)_1 = M/M_1$ results in

$$M = \sum_{i=1}^{n} M_i \quad (8)$$

Expanding Eq (7) and using $(SNR)_i/(SNR)_1 = M_i/M_1$,

$$SNR = \sum_{i=1}^{n} (SNR)_i \quad (9)$$

In terms of antenna efficiency, ϵ and physical diameter D [using $G_i = \epsilon_i(\pi D_i/\lambda)^2$],

$$M = (\pi/\lambda)^2 \sum_{i=1}^{n} \epsilon_i D_i^2/T_i \quad (10)$$

Relative to the highest performance antenna with figure of merit M_1, using[5] $\Delta M_i(\text{dB}) = M_1(\text{dB}) - M_i(\text{dB})$,

$$\Delta M(\text{dB}) = 10 \log \sum_{i=1}^{n} 10^{-\Delta M_i(\text{dB})/10} \quad (11)$$

[5] $M(\text{dB}) = 10 \log M$

The following section presents some examples

III. Discussion

Consider an idealized array of one 64-m antenna and N 34-m antennas with equal antenna efficiencies and system noise temperatures Figure 3 shows the improvements of N 34-m antennas relative to the single 64-m antenna (upper curve) and relative to the single 64-m antenna plus $(N - 1)$ 34-m antennas (lower curve) This illustrates that adding multiple antennas to an array has a smaller percentage effect as N grows large

A successful DSN array configuration was used at the Voyager 1 and 2 Saturn flybys using a 64-m antenna and a 34-m antenna (Table 1 and Fig 1) This resulted in an average measured increase in signal-noise ratio of about 0 6 dB relative to the 64-m antenna only (Ref 4) Assuming 50% efficiency for both antennas and no array losses, the potential improvement was about 1 1 dB (Fig 1) Future refinements of the array techniques should reduce the implied $\simeq$0 5 dB average array loss

Finally, consider the optimum improvement at 8 42 GHz of the following array relative to the DSS 43 antenna DSS 43 (64-m), DSS 42 (34-m, G/T = -6 0 dB relative to DSS 43), DSS 45 (34-m, G/T = -4 5 dB relative to DSS 43), and Parkes (64-m, G/T = -1 1 dB relative to DSS 43) Using Eq (11), the estimated improvement in potential figure of merit is

$$\Delta M \simeq 10 \log (1 + 0\ 25 + 0\ 35 + 0\ 78)\ \text{dB}$$

$$\simeq 3\ 8\ \text{dB}$$

This is the array configuration and potential performance improvement presently planned for the Voyager 2 1986 Uranus encounter (Refs 9 and 10)

References

1 Urech, J M , "Telemetry Improvement Proposal for the 85-ft Antenna Network," *SPS 37-63 Vol II*, p 116, Jet Propulsion Laboratory, Pasadena, Calif , May 31, 1970

2 Wilck, H , "A Signal Combiner for Antenna Arraying," *DSN Progress Report 42-25*, Jet Propulsion Laboratory, Pasadena, Calif , Feb 15, 1975

3 Winkelstein, R A , "Analysis of the Signal Combiner for Multiple Antenna Arraying," *DSN Progress Report 42-26*, Jet Propulsion Laboratory, Pasadena, Calif , Apr 15, 1975

4 Bartok, C D , "Performance of the Real-Time Array Signal Combiner During the Voyager Mission," *TDA Progress Report 42-63*, Jet Propulsion Laboratory, Pasadena, Calif , June 15, 1981

5 Wait, D F , "Satellite Earth Terminal G/T Measurements," *Microwave Journal*, Vol 20, No 4, Apr 1977

6 Stelzried, C T , and Noreen, G K , "The System View," *The Deep Space Network – A Radio Communications Instrument for Deep Space Exploration*, N A Renzetti and C T Stelzried, eds , Publication 82-104, Jet Propulsion Laboratory, Pasadena, Calif , February 1983

7 Clauss, R C , private communication, Dec 1982

8 Howard, L D , "Split-Symbol Correlator Signal-to-Noise Ratio Detector," *New Technology Report*, JPL/NASA No 5116/15653, March 31, 1981

9 Brown, D W , private communication, Dec 1982

10 "Deep Space Network/Flight Project Interface Design Book," Vol 2 Proposed DSN Capability, 810-5, Rev D, TLM-10, Jet Propulsion Laboratory, Pasadena, Calif , Dec 1982

Table 1. Tabulated downlink parameters for Voyager 2 spacecraft, Saturn (August 1981) flyby, X-band (8 42 Ghz)

Parameter	Value
Transmitter power P_T, W	21 3
Spacecraft antenna effective area A_T, m^2	5 4
Distance D, m	$1\ 557 \times 10^{12}$
Total link loss L, ratio	1 1
Data rate R, bps	$4\ 48 \times 10^{4}$
Threshold signal-to-noise ratio for 5×10^{-3} BER $(E_b/N_0)_T$, ratio	1 8
Threshold figure of merit, calculated from Eq (2) M_T, dB	55

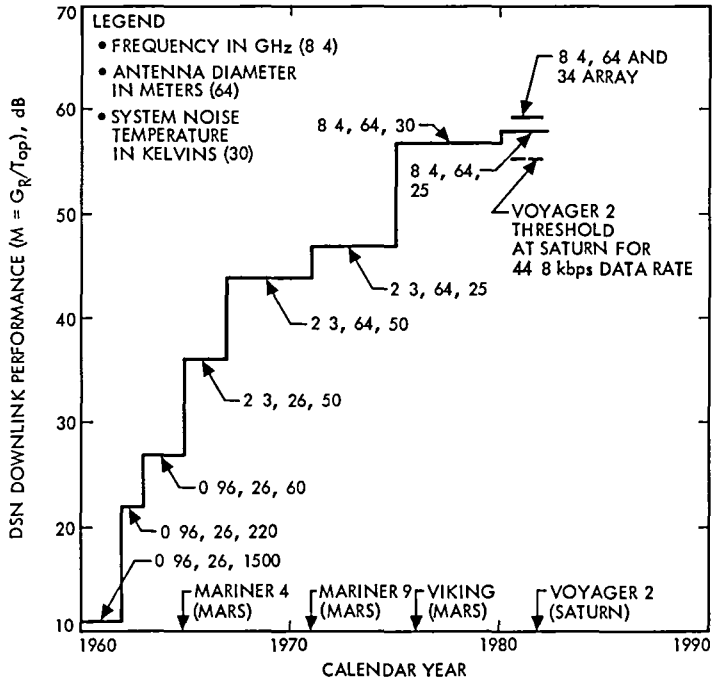


Fig 1 Profile of the DSN downlink performance ($M = G_R/T_{op}$) from 1960 to 1982

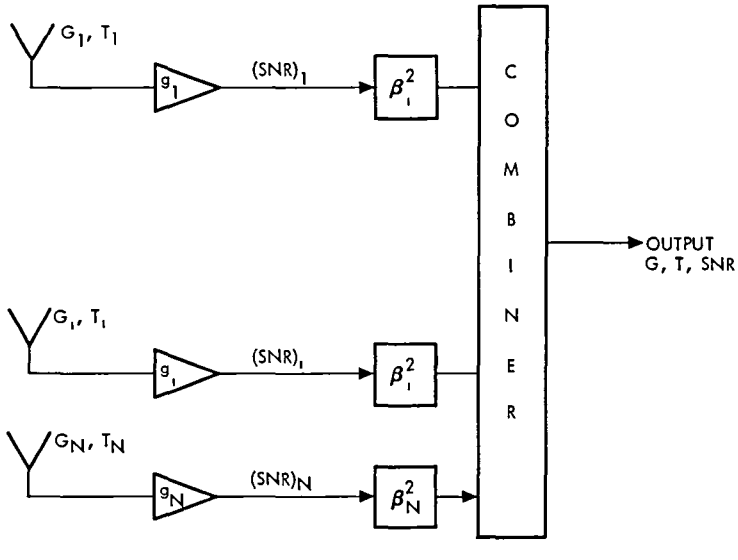


Fig. 2 Antenna array configuration

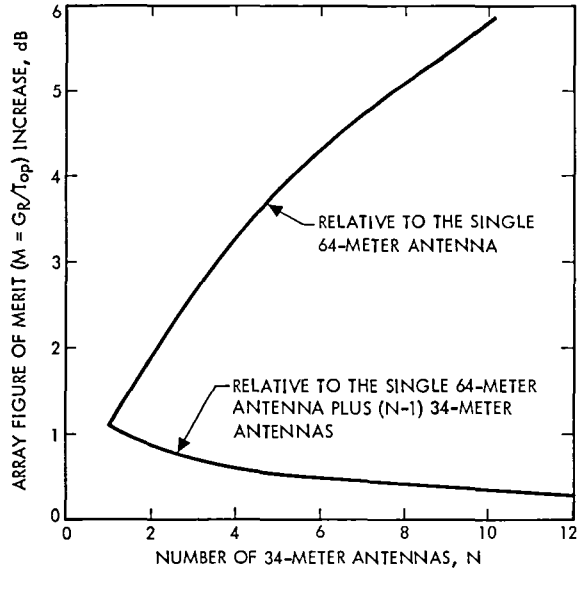


Fig 3 Ideal figure of merit increase of an antenna array consisting of one 64-m antenna and N 34-m antennas

S-Band Aperture Efficiency of the DSS 13 26-Meter Antenna with the S-X Common Aperture Feed

M J Klein
Atmospheric Sciences Section

Measurements of the S-band performance of the DSS 13 26-meter antenna at Goldstone, equipped with the second-generation S-X common aperture feed, are reported Aperture efficiency and overall antenna gain results are compared with predicted values The elevation dependence of the aperture efficiency is reported

I. Introduction

Precision measurements of astronomical radio sources have been analyzed to evaluate the gain and aperture efficiency of the 26-m antenna at DSS 13 The observations were made during January and February 1982 at 2270 MHz and 2295 MHz with the second generation S-X common aperture feed (Ref 1), which was installed in November 1981 The results have been used to compare the performance of the antenna before and after the installation of the new feed cone Slight changes in the elevation dependence of the aperture efficiency have been detected and the peak antenna gain was measured

This recalibration of the DSS 13 antenna was conducted to maintain the high precision capability required for performance evaluations of DSN antennas and for radio astronomy research programs An example of the latter is the Jupiter Patrol (Ref 2), by which variations in the radio emission from the Jovian radiation belts have been successfully monitored for the past decade The one-sigma precision of Jupiter's flux-density measurements has been ±2% (±0 1 dB) The effective gain of the antenna near the elevation of transit for the ISEE-3 spacecraft was also determined, and these results were used to evaluate carrier power measurements from the spacecraft

II. Observations

A selected list of five calibration sources were observed for several hours on January 26, January 29, and February 16, 1982 The observing sessions were planned so that the sources could be measured over a wide range of elevation angles The intensity of each source was measured several times each night using the drift-curve technique described in Ref 2

The one-sigma uncertainty of each source-temperature measurement was less than ±0 1 dB To achieve this precision, corrections were applied to account for residual pointing errors and a noise-adding radiometer (NAR) subsystem was used to provide receiver gain stabilization The feed subsystem was switched to the low-noise listen-only configuration (Ref 1) to optimize the system sensitivity

The absolute calibration is determined by the flux density scale published by Baars et al (Ref 4) This work is widely

used by the radio astronomy community The primary reference source is Virgo A (3C274), which unfortunately was not visible during the available observing time However, the flux densities of the five sources that were observed were calculated from the previously observed ratios of their intensities relative to Virgo A (Refs 3 and 5)

The flux density adapted for Virgo A at 2295 MHz is 140 0 Jy The flux densities of Virgo A and the calibration sources were calculated at 2270 MHz from the fluxes at 2295 MHz and the spectral indices given in Ref 4

The source list and the flux densities used at both frequencies are given in Table 1 The number, N, of independent measurements is given in the third column The correction factor C_R, listed in the fifth column, is applied to the measurements to account for the loss of intensity due to the angular resolution of the antenna beam, which partially resolves the spatial structure of some radio sources The effect is small, but not negligible, for the sources included in this report The source temperatures and the effective aperture efficiency η are discussed below

III. Effective Aperture Efficiency

The effective aperture efficiency is equal to the ratio of the observed source temperature and the theoretical source temperature for a 100% efficient antenna, $T(\eta = 100\%)$ The latter is given by the expression

$$T(\eta = 100\%) = \frac{\pi D^2}{4} \frac{S}{2k} \frac{1}{C_R}, \text{ kelvin} \qquad (1)$$

where D is the antenna diameter, S is the flux density of the source, $k = 1\ 3806 \times 10^{-23}\ Ws^{-1}\ K^{-1}$ is the Boltzmann constant and C_R is the resolution correction factor For D = 25 91 m (85 0 ft) and flux density S expressed in Jy (1 Jy = 10^{-26} Wm^{-2} Hz^{-1})

$$T(\eta = 100\%) = 0\ 19095 \frac{S}{C_R}, \text{ kelvin} \qquad (2)$$

The values of $T(\eta = 100\%)$ are given in the 6th column of the table, adjacent to the observed source temperatures The calculated effective aperture efficiencies are given in the last column The weighted averages of η for the two frequencies are also given beneath the entries in the last column The weighted average of the effective aperture efficiency for all seven entries is 0 594 ± 0 002 The uncertainty is the one-sigma standard error calculated from the scatter in the measurements

The measured efficiency is the effective aperture efficiency because the data are affected by systematic attenuation in the clear atmosphere as well as certain losses in the feed system With the NAR, all system temperatures are calibrated with respect to the temperature of the NAR reference diode, which is injected into the system via a waveguide coupler at the input to the maser Consequently, the radio source temperatures are referenced to the waveguide flange at the maser input

IV. Elevation Dependence of Aperture Efficiency and Gain

A. Experimental Results

The individual data that resulted in the average results given in Table 1 are plotted in Fig 1 as a function of source elevation angle The curve represents an empirically determined function of the elevation dependence The equation is given by

$$\eta(El) = \eta(37^\circ) A e^{-\tau \sec Z} [(1 - a) + a \cos(El - 30^\circ)] \qquad (3)$$

where $\eta(El)$ is the effective aperture efficiency at elevation El, A is a normalization factor, τ = 0 007 is the atmospheric loss factor at S-band frequencies, Z is the zenith angle of the source ($Z = 90^\circ - El$) Note that τ is the atmospheric attenuation in nepers at zenith and that 1 neper ≃ 4 343 dB The coefficient a = 0 098 was empirically determined from the data in Fig 1

Written in this form, Eq (3) is the expression for the elevation dependence of the effective efficiency as observed by an experimenter The equation is normalized to $El = 37^\circ$ to match the observed peak in the efficiency measurements The factor A = 1 0110 normalizes the curve to unity at $El = 37^\circ$ The last factor $[(1 - a) + a \cos(El - 30^\circ)]$ represents the loss in efficiency due to gravity loading The measured symmetry about $El = 30^\circ$ suggests that the peak efficiency of the antenna should occur near 30° elevation when atmospheric losses are removed

The average values of the effective aperture efficiency for each source (Table 1) are the average values of $\eta(37^\circ)$, which were found by solving Eq (3) for each measurement The effective aperture efficiency was converted to antenna gain for a frequency of 2295 MHz The equivalent antenna gain scale (in dB) is plotted along the right-hand ordinate Using this scale one can see in Fig 1 that the peak antenna gain, in the presence of atmospheric attenuation, occurs near 37° elevation The maximum value of the antenna gain, referred to the reference flange at the input to the maser, is 53 65 ±

0 16 dB at 2295 MHz The corresponding peak gain at 2270 MHz, not shown in the figure, is 53 56 ± 0 16 dB

The uncertainties quoted for the peak gain results are the one-sigma error estimates for the two frequencies The error budget includes the standard error on the measured values of η (±1%), the absolute uncertainty of the flux-density calibration scale (±3%), and the systematic error of the NAR temperature calibration (±2 0%) If these errors are uncorrelated, the resultant error is ±3 7%, or ±0 16 dB The peak effective aperture efficiency, with its total one-sigma error estimate, is $\eta(37^\circ) = 0\ 594 \pm 0\ 022$

B. Comparison with Calculated Predictions

In order to compare the effective aperture efficiency reported in this article with efficiency calculations performed for the S-X common aperture feed, corrections must be applied to account for attenuation through the clear atmosphere at Goldstone and for attenuation between the input flange of the maser and the aperture of the feed The first step is to write Eq (3) for $El = 30^\circ$ (the peak in the antenna gain curve without atmospheric losses)

$$\eta(30^\circ) = \eta(37^\circ) A\, e^{-\tau \sec(90^\circ - 30^\circ)} \qquad (4)$$

Next, multiply by $e^{+\tau \sec(90^\circ - 30^\circ)}$ to remove atmospheric attenuation and divide by f_L (the waveguide loss factor)

$$\eta^*(30^\circ) = (A/f_L)\,\eta(37^\circ) \qquad (5)$$

where $\eta^*(30)$ is the maximum aperture efficiency for this antenna referred to the phase center of the common aperture feed From J Withington (private communication) $f_L = 0\ 98 \pm 0\ 002$ (-0 09 dB) is the loss factor in the waveguide that includes two switches, three bends, and a rejection filter at the X-band transmitter fundamental frequency With $A = 1\ 0110$, the solution to Eq (5) is

$$\eta^*(30^\circ) = 1\ 0316\,\eta(37^\circ) = 0\ 613 \pm 0\ 022 \qquad (6)$$

$\eta^*(30)$ can be compared with calculations of the overall aperture efficiency for this feed, prepared by Williams and Reilly (Ref 6) Their result is 0 578

A review (private communication, D A Bathker) of the predicted aperture efficiency has been done Three factors have been identified as principal sources of uncertainty in the predictions The largest uncertainty is the spar (or quadripod) blockage, which was assumed to consist of the optical shadow area inflated by a factor of 1 2 to account for diffraction at S-band wavelengths A likely 3σ tolerance on the 0 861 spar blockage factor used in the predictions is $^{+0\ 04}_{-0\ 00}$ The next largest uncertainty is the forward spillover for this dual-band feed, which is higher than older designs at 11 1% A likely 3σ tolerance on this prediction is $^{+0\ 02}_{-0\ 00}$

Finally, the predictions accepted an estimate of 1 5 mm (0 06 in) rms for the surface tolerance of both reflectors This is conservative, and results in a prediction of 0 979 surface efficiency, a likely 3σ tolerance is $^{+0\ 01}_{-0\ 00}$ The net (conservative) 3σ tolerance on the overall predicted efficiency is thus $^{+0\ 07}_{-0\ 00}$, the one-sigma tolerance is $^{+0\ 023}_{-0\ 000}$

The results of the measurements reported here and the predicted values discussed above are summarized in Table 2 The ratio of the measured and predicted values of antenna-design aperture efficiency at S-band is (0 613/0 578) = 1 0606 The measured efficiency is 6 1% (0 25 dB) greater than that predicted from component evaluation calculations Although this difference is not statistically significant when the one-sigma error estimates are considered, the difference may be large enough to warrant additional investigation

V. Discussion and Conclusions

The elevation dependence of the S-band aperture efficiency and the peak gain of the DSS 13 antenna appear to have changed only slightly with the installation of the S-X common aperture feed The peak efficiency now occurs near 37° elevation, where the effective efficiency is 59 4% Similar calibration data taken on several occasions between 1972 and 1979 yielded curves with maxima near 50° elevation and peak effective efficiencies of 60 5% Independent measurements by Freiley, Batelaan, and Bathker (Ref 5) gave consistent results with a peak effective efficiency of 60 6%

The shift in the efficiency curve toward lower elevation angles is probably caused by a slight change in focus The Z-axis subreflector focus was measured soon after the new feed was installed A proper S-band focus was found when the Z-axis readout, which is a relative measure of distance between the subreflector and the feed, was set at 350 (arbitrary units) The previous value of the S-band focus was 330

Acknowledgments

I would like to thank D A Bathker, C T Stelzried, W F Williams, and J R Withington for reviewing this article and for contributing helpful suggestions and insights I also thank T J Thompson for assistance in processing the data

References

1 Withington, J R , and Williams, W F , "A Common-Aperture S- and X-Band Four-Function Feed Cone," *Proc 1981 Antenna Applications Symposium*, University of Illinois, Sept 23-25, 1981

2 Klein, M J , "The Variability of the Total Flux Density and Polarization of Jupiter's Decimetric Radio Emission," *J Geophys Res* , Vol 81, pp 3380-3382, July 1976

3 Klein, M J , and Stelzreid, C T , "Calibration Radio Sources for Radio Astronomy Precision Flux-Density Measurements at 2295 MHz," *Astron J* , Vol 81 pp 1078-1083, Dec 1976

4. Baars, J W. M , Genzel, R , Pauliny-Toth, I I. K , and Witzel, A , "The Absolute Spectrum of Cas A, An Accurate Flux Density Scale and a Set of Secondary Calibrators," *Astron Astrophys* , Vol 61, pp 99-106, 1977

5 Freiley, A J , Batelaan, P D , and Bathker, D A , *Absolute Flux Density Calibrations of Radio Sources 2.3 GHz,* Technical Memorandum 33-806, Jet Propulsion Laboratory, Pasadena, Calif , 1977

6 Williams, W , and Reilly, H , "A Prototype DSN S/X-Band Feed DSS 13 Application Status (Fourth Report)," *TDA Progress Report 42-60*, pp 77-88, Jet Propulsion Laboratory, Pasadena, Calif , Dec 15, 1980

Table 1. DSS 13 antenna aperture efficiency, common aperture feed: 1982 January and February data (calibrated with noise adding radiometer)

Source designation (IAU)	Source designation (Other)	N	Flux density, Jy	C_R (26-m ant)	Source temperature, T(η = 100%), K	Source temperature, T(observed), K	η (El = 37°)
				2295 MHz			
0433 + 295	3C123	10	31.94	1 000	6 10	3.61 ±0.01	0 592
1648 + 05	3C348	8	28 10	1 006	5 33	3.22 ±0 02	0 604
1957 + 40	Cygnus A	19	915 00	1 009	173.20	103.30 ±0.02	0 596
							(0 597) ±0.003
				2270 MHz			
0433 + 295	3C123	2	32 24	1.000	6.6	3 61 ±0.17	0.586
0134 + 329	3C48	3	10 80	1.000	2 06	1.23 ±0 06	0.596
1717 – 00	3C353	2	40.64	1 010	7.68	4 50 ±0 06	0 586
1957 + 40	Cygnus A	6	927.50	1 009	175 00	105.00 ±0 40	0.598
							(0 592) ±0 004

Table 2. DSS 13 aperture efficiency S-band results with S-X common aperture feed

Aperture efficiency type	Peak efficiency (±1σ)	Elevation of peak	Reference plane	Notes
Effective, measured	η = 0 594 ± 0 022	37°	Maser input	User effective aperture includes atmospheric and waveguide losses
Antenna design, inferred from measurements	η^* = 0 613 ± 0 022	30°	Phase center of feed	Does not include atmospheric or waveguide losses
Antenna design, predicted[1]	$\eta^* = 0\ 578\ ^{+0\ 023}_{-0\ 000}$	–	Phase center of feed	Does not include atmospheric or waveguide losses

[1]Includes feed pattern integration, subreflector and spar blockage, and surface roughness (Ref 6, and D A Bathker, private communication)

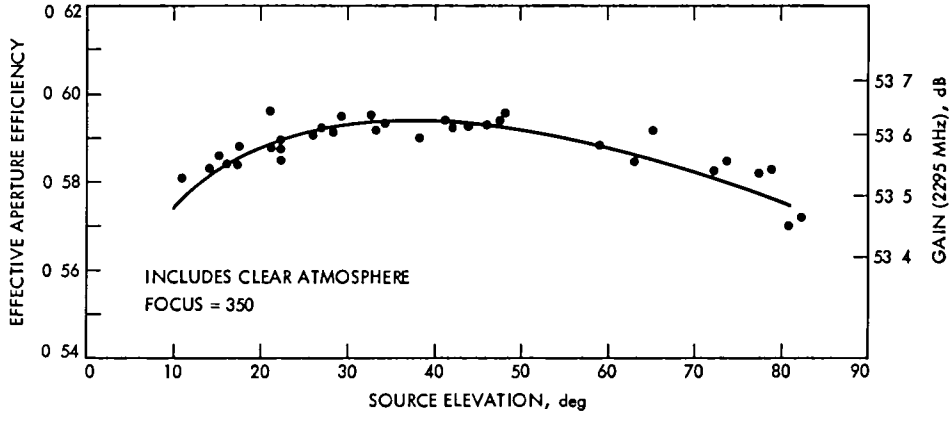


Fig 1. Elevation dependence of aperture efficiency of the DSS 13 26-m antenna, with S-X common aperture feed, 2295 MHz (1982 January and February data)

The Effect of Partial Coherence In Receiving System Noise Temperature on Array Gain for Telemetry and Radio Frequency Carrier Reception for Receiving Systems with Unequal Predetection Signal-to-Noise Ratios

M H Brockman
Telecommunications Science and Engineering Division

Signal-to-noise ratio improvement realized by arraying receiving systems or stations for coherent reception is reduced when some portion of the predetection noise is coherent in the array This article examines this effect for arrayed receiving systems with unequal apertures including expected performance for selected planets of the solar system within the array beamwidth.

I. Introduction

An earlier report (Ref 1) considered the effect of partial coherence in system noise temperature on predetection signal-to-noise ratio improvement and equivalent radio frequency (RF) carrier signal-to-noise ratio improvement for coherent reception and demodulation for an array of similar receiving systems with essentially equal predetection noise and equal size antenna apertures This report examines the effect of partial coherence in system noise temperature on predetection signal-to-noise ratio improvement and RF carrier margin improvement for coherent reception and demodulation for an array of receiving systems with unequal predetection noise and unequal antenna apertures The analytic expressions given represent the general case for the configuration examined here

Performance is presented in this report for the case which is representative of 34-meter-diameter antenna receiving systems arrayed with a 64-meter-diameter antenna receiving system which also has transmit capability For the performance shown here, the coherent portion of predetection noise (which is coherent periodically) results from a planet in the solar system which is within the beamwidth of the antennas of the array RF carrier array gain measurements made in the Laboratory for simulated 64-meter/34-meter antenna receiving system arrays are compared with theoretical results

This report also provides information on the expected periodic reduction in array gain at a reception frequency of 8420 MHz due to antenna feed spacing with the planets Jupiter or Venus in the array beamwidth The expected periodic reduction in array gain with these planets in the array beamwidth is less than 0 2 dB for a two-aperture array and less than 0 25 dB for a three-aperture array for the antenna spacings and orientation planned for the 64- and 34-meter antennas at the three DSN complexes Expected reduction in array gain with the planets Uranus or Neptune in the array beamwidth is negligible

II. Predetection Signal-to-Noise Ratio and Resultant Carrier Tracking Loop Phase Noise

Figure 1 illustrates the configuration considered herein which provides predetection signal-to-noise ratio improvement and equivalent RF carrier signal-to-noise ratio improvement (RF carrier margin improvement) for high rate telemetry with residual RF carrier The RF carrier margin improvement also provides a corresponding improvement in minimum RF carrier level for radio metric tracking Figure 1 and a modification of Fig 1 (so that much larger antenna separations for the array can be handled conveniently) were presented in Ref 1 with a discussion of predetection noise resulting from operating equivalent system noise temperature T_{op}

With receiving system 2 connected to the summing junction (see Fig 1), the RF carrier predetection signal-to-noise ratio in receiving system 1 at the output of the summing junction is (from Expression 3, Ref 1)

$$\frac{P_{c1_{\Sigma 1\,2}}}{P_{n1_{\Sigma 1\,2}}} = \frac{(A_1 \cos m_{pd} + \beta_2 A_2 \cos m_{pd})^2}{\left[NBW_{F_{A1}} (1-\epsilon_1) N_{o1} + \beta_2^2 NBW_{F_{A2}} (1-\epsilon_2) N_{o2}\right] + \left[(NBW_{F_{A1}} \cdot \epsilon_1 N_{o1})^{1/2} + \beta_2 NBW_{F_{A2}} \cdot \epsilon_2 N_{o2})^{1/2}\right]^2} \tag{1}$$

where $NBW_{F_{A1}}$ represents the noise bandwidth of second IF filter F_{A1}, $NBW_{F_{A2}}$ is equal to $NBW_{F_{A1}}$, N_{o1} is the one-sided noise spectral density for receiving system 1 related to T_{op1}, and N_{o2} is the noise spectral density of receiving system 2 related to T_{op2} The term β_2 is the voltage coupling of receiving system 2 relative to receiving system 1 at the summing junction, and γ_2^2 is the carrier power-to-noise spectral density ratio of receiving system 2 relative to receiving system 1 The statistically independent portion of predetection noise is $(1 - \epsilon_1)$ in receiving system 1 and $(1 - \epsilon_2)$ in receiving system 2 The corresponding portion of predetection noise that is coherent is ϵ_1 and ϵ_2 in the two receiving systems Expression 1 can be rewritten as

$$\frac{P_{c1_{\Sigma 1\,2}}}{P_{n1_{\Sigma 1\,2}}} = \frac{P_{c1}}{NBW_{F_{A1}} \cdot N_{o1}} \cdot \frac{(1+\beta_2\gamma_2)^2}{\left[(1-\epsilon_1) + (1-\epsilon_2)\frac{N_{o2}}{N_{o1}}\beta_2^2\right] + \left[\epsilon_1^{1/2} + \epsilon_2^{1/2}\left(\frac{N_{o2}}{N_{o1}}\right)^{1/2}\beta_2\right]^2} \tag{2}$$

Note that (from Ref 1) the predetection carrier signal-to-noise ratio in receiving system 1 (with receiving system 2 switched out of the summing junction) is $P_{c1}/(NBW_{F_{A1}} \cdot N_{o1})$ Consequently, the improvement in predetection signal-to-noise ratio for two receiving systems arrayed (n_2'') is

$$n_2'' = \frac{(1+\beta_2\gamma_2)^2}{\left[(1-\epsilon_1) + (1-\epsilon_2)\frac{N_{o2}}{N_{o1}}\beta_2^2\right] + \left[\epsilon_1^{1/2} + \epsilon_2^{1/2}\left(\frac{N_{o2}}{N_{o1}}\right)^{1/2}\beta_2\right]^2} \tag{3}$$

For N receiving systems arrayed, the predetection carrier signal-to-noise ratio in receiving system 1 becomes

$$\frac{P_{c1\Sigma 1,\ ,N}}{P_{n1\Sigma 1,\ ,N}} = \frac{P_{c1}}{NBW_{F_{A1}} \cdot N_{o1}} \cdot \frac{(1+\beta_2\gamma_2+\quad \beta_N\gamma_N)^2}{\left[(1-\epsilon_1)+(1-\epsilon_2)\frac{N_{o2}}{N_{o1}}\beta_2^2+\quad +(1-\epsilon_N)\frac{N_{oN}}{N_{o1}}\beta_N^2\right]+\left[\epsilon_1^{1/2}+\epsilon_2^{1/2}\left(\frac{N_{o2}}{N_{o1}}\right)^{1/2}\beta_2+\ \cdot\ +\epsilon_N^{1/2}\left(\frac{N_{oN}}{N_{o1}}\right)^{1/2}\beta_N\right]^2} \tag{4}$$

The improvement in predetection signal-to-noise ratio for N receiving systems arrayed is then

$$n_N'' = \frac{(1+\beta_2\gamma_2+\quad \beta_N\gamma_N)^2}{\left[(1-\epsilon_1)+(1-\epsilon_2)\frac{N_{o2}}{N_{o1}}\beta_2^2+\quad +(1-\epsilon_N)\frac{N_{oN}}{N_{o1}}\beta_N^2\right]+\left[\epsilon_1^{1/2}+\epsilon_2^{1/2}\left(\frac{N_{o2}}{N_{o1}}\right)^{1/2}\beta_2+\quad +\epsilon_N^{1/2}\left(\frac{N_{oN}}{N_{o1}}\right)^{1/2}\beta_N\right]^2} \tag{5}$$

It should be noted (as discussed in Ref 1) that for the condition where the varying group delay on the telemetry sidebands is tracked out among the various receiving systems of the array (which is accomplished at baseband for the configuration shown in Fig 1), the improvement in telemetry predetection signal-to-noise ratio that can be obtained is also represented by expressions (3) and (5) for the corresponding voltage coupling β in the signal spectrum combining process and ratio of signal-to-noise spectral density ratios γ^2 for telemetry However, the predetection signal-to-noise ratio improvement obtained for telemetry does not include loss in the telemetry signal spectrum combining process and the loss (radio loss) due to carrier tracking loop phase noise.

The resultant rms phase noise ($\sigma_{\Phi_{n1\Sigma 1,2}}$) at the output of the RF carrier tracking loop (i e , on the first local oscillator signal) in receiving system 1 due to the predetection signal-to-noise ratio within the closed-loop noise bandwidth of the RF carrier tracking loop becomes (see expression 7, Ref 1)

$$\sigma_{\phi_{n1\Sigma 1,2}} = \frac{\frac{N_{o1}}{2}\cdot 2B_{L1}}{P_{c1}} \cdot \frac{1}{n_2''}\left[\frac{1+\frac{P_{c1}\,n_2''}{NBW_{F_{A1}}\cdot N_{o1}}}{0\ 862+\frac{P_{c1}\cdot\eta_2''}{NBW_{F_{A1}}\cdot N_{o1}}}\cdot\frac{\exp\left(\frac{N_{o1}\cdot B_{L1}}{P_{c1}\cdot\eta_2''}\right)}{\sinh\left(\frac{N_{o1}\cdot B_{L1}}{P_{c1}\cdot\eta_2''}\right)}\right]^{1/2} \text{ rad, rms} \tag{6}$$

for two systems arrayed where n_2'' is defined in expression (3) above For N receiving systems arrayed, the resultant rms phase noise ($\sigma_{\phi_{n1\Sigma 1,\ \ ,N}}$) becomes

$$\sigma_{\phi_{n1\Sigma 1,\ \ ,N}} = \frac{\frac{N_{o_N}}{2}\cdot B_{L1}}{P_{c1}} \cdot \frac{1}{n_N''}\left[\frac{1+\frac{P_{c1}\cdot n_N''}{NBW_{F_{A1}}\cdot N_{o1}}}{0\ 862+\frac{P_{c1}\cdot n_N''}{NBW_{F_{A1}}\cdot N_{o1}}}\cdot\frac{\exp\left(\frac{N_{o1}\cdot B_{L1}}{P_{c1}\cdot n_N''}\right)}{\sinh\left(\frac{N_{o1}\cdot B_{L1}}{P_{c1}\cdot n_N''}\right)}\right]^{1/2} \text{ rad, rms} \tag{7}$$

where n_N'' is defined in expression (5) above The rms phase noise $\sigma_{\phi_{n1}}$ for receiving system 1 with receiving systems 2 through N switched out of the summing junction is shown in Ref 1 (expression 2)

Note that the total phase noise at the output of the principal carrier tracking loop (i e , on the first local oscillator) becomes

$$\left[\sigma^2_{\phi_{n1\Sigma 1,2}} + \left(\frac{\beta_2 \sigma_{\phi_{n2}}}{1+\beta_2}\right)^2\right]^{1/2} \tag{8}$$

for two receiving systems arrayed (see Ref 1) For N receiving systems arrayed, the total rms phase noise on the first local oscillator is

$$\left[\sigma^2_{\phi_{n1\Sigma 1,\ \ ,N}} + \left(\frac{\beta_2 \sigma_{\phi_{n2}}}{1+\beta_2}\right)^2 + \quad + \left(\frac{\beta_N \sigma_{\phi_{nN}}}{1+\beta_N}\right)^2\right]^{1/2} \tag{9}$$

Expressions for rms phase noise in systems 2 through N were developed in Ref 1

As developed in Ref 1, the total rms phase noise in expressions (8) and (9) can be considered as due to an equivalent carrier predetection signal-to-noise ratio within the closed-loop noise bandwidth of the RF carrier phase tracking loop Comparison of this equivalent carrier signal-to-noise ratio with the initial carrier predetection signal-to-noise ratio in a single receiving system (i e , system 1) alone provides the improvement in carrier margin due to radio frequency carrier arraying for the high rate telemetry configuration where some portion of the predetection noise is not statistically independent among the various receiving systems of the array

III. Performance

Expressions (3) and (5) represent improvement in predetection signal-to-noise ratio in receiving system 1 under the condition that the phase shift and group delay in the various receiving systems of the array are essentially matched in a predetection bandwidth that encompasses the signal of interest In addition for telemetry (as described in Ref 1), expressions (3) and (5) apply for the additional condition that the varying group delay on the signal modulation sidebands is tracked out among the receiving systems of the array by a telemetry signal spectrum combiner during a station pass Note again that the predetection signal-to-noise ratio improvement for telemetry does not include the telemetry signal combiner loss and the loss (radio loss) due to carrier tracking loop phase noise Reference 2 provides information on radio loss

Improvement in receiving system 1 predetection signal-to-noise ratio is shown in Fig 2 as a function of the statistically independent (noncoherent) portion of predetection noise $(1 - \epsilon_1)$ in receiving system 1 for an array consisting of a 64-meter-diameter antenna with a 50% aperture efficiency (system 1) and a 34-meter-diameter antenna listen-only, with a 60% aperture efficiency As pointed out in Section I, for the performance shown in this report, the coherent portion of predetection noise results from a planet in the solar system which is within the beamwidth of the antennas of the array Consequently, the coherent noise for the 34-meter-diameter listen antenna is scaled relative to the 64-meter-diameter antenna by $(34/64)^2 \times (0\ 60/0\ 50)$ Performance is shown for γ_2 values of 0 55, 0 61, and 0 66 for the case where $\beta_2 = \gamma_2$ As developed earlier in this report, γ_2^2 is the signal-power-to-noise-spectral-density ratio of receiving system 2 relative to the receiving system 1, and β_2 is the voltage coupling of receiving system 2 relative to receiving system 1 for the combining process Improvement in predetection signal-to-noise ratio is maximum for $\beta = \gamma$ when the predetection noise is all statistically independent or $1 - \epsilon_1 = 1$, which is representative for the DSN most of the time A γ_2 value of 0 61 represents the nominal value for this array and provides a 1 46-dB improvement in predetection signal-to-noise ratio at $\beta = \gamma$ for $1 - \epsilon_1 = 1$ A γ_2 value of 0 66 represents operation of the antennas at low (10-15 deg) and high (approximately 80 deg) elevation angles, whereas a γ_2 value of 0 55 represents a below nominal signal-to-noise ratio for the 34-m antenna (listen-only) The ratio of noise spectral densities (N_{o2}/N_{o1}) used in Fig 2 for noncoherent predetection noise $(1 - \epsilon_1 = 1)$ is 0 925 (18 5K/20 0K) A N_{o2}/N_{o1} value of 1 for $1 - \epsilon_1 = 1$ reduces the improvement in predetection signal-to-noise ratio shown in Fig 2 by about 0 1 dB and less than 0 1 dB for $1 - \epsilon_1$ less than 1 [1] The resultant decrease in predetection signal-to-noise ratio improvement is shown in Fig 2 as an increasing portion of predetection noise becomes coherent

[1] A N_{o2}/N_{o1} of 0 925 (18 5/20 0) represents operation at high elevation angles, while at low elevation angles, N_{o2}/N_{o1} approaches a value of 1

(ϵ_1) or as the statistically independent portion of predetection noise ($1 - \epsilon_1$) decrease for receiving system 1 An increase in aperture efficiency for the 34-meter-diameter listen antenna to 65% reduces the improvement in predetection signal-to-noise ratio shown in Fig 2 by about 0 1 dB at $1 - \epsilon_1 = 0\ 3$ and has no effect when $1 - \epsilon_1 = 1$ for a given γ_2 value

Figure 3 shows receiving system 1 predetection signal-to-noise ratio improvement as a function of the statistically independent portion of system 1 predetection noise ($1 - \epsilon_1$) for an array of two receiving systems utilizing a 64-meter-diameter antenna and a 34-meter-diameter antenna (transmit/receive) (with a 54% aperture efficiency), for γ_2 values of 0 48, 0 53, and 0 58 Coherent noise for the 34-meter-diameter transmit/receive antenna is scaled relative to the 64-meter-diameter antenna by $(34/64)^2 \times (0\ 54/0\ 50)$ A nominal γ_2 value of 0 53 provides a 1 0-dB improvement in predetection signal-to-noise ratio for $1 - \epsilon_1 = 1$ The preceding discussion for γ_2 values of 0 55, 0 61, and 0 66 for Fig 2 applies to the γ_2 values of 0 48, 0 53, and 0 58 respectively for Fig 3 The ratio of noise spectral densities (N_{o2}/N_{o1}) is 1 075 (21 5K/20 0K) for noncoherent predetection noise ($1 - \epsilon_1 = 1$) in Fig 3 In this case, a N_{o2}/N_{o1} value of 1 for $1 - \epsilon_1 = 1$ increases the improvement in predetection signal-to-noise ratio shown in Fig 3 by 0 1 dB (or less) A decrease in aperture efficiency for the 34-meter-diameter transmit/receive antenna to 50% increases the improvement in predetection signal-to-noise ratio shown in Fig 3 by about 0 1 dB at $1 - \epsilon_1 = 0\ 3$ and has no effect when $1 - \epsilon_1 = 1$ for a given γ_2 value

Figures 4 and 5 show improvement in receiving system 1 predetection signal-to-noise ratio for an array of three receiving systems as a function of the statistical independent portion of predetection noise ($1 - \epsilon_1$) in receiving system 1 for the case where $\beta_2 = \beta_3 = \gamma_2$ Figure 4 represents an array utilizing a 64-meter-diameter antenna (system 1), a 34-meter-diameter antenna listen-only (system 2) and a 34-meter-diameter antenna with transmit/receive capability (system 3) Nominal values of $\gamma_2 = 0\ 61$ and $\gamma_3 = 0\ 53$ provide a 2 17-dB improvement in predetection signal-to-noise ratio for $1 - \epsilon_1 = 1$ (Fig 4) For this figure, the ratio of noise spectral densities N_{o2}/N_{o1} and N_{o3}/N_{o1} are 0 925 and 1 075 respectively at $1 - \epsilon_1 = 1$ A value of 1 for N_{o2}/N_{o1} and N_{o3}/N_{o1} provides essentially the same performance as shown in Fig 4 Figure 4 represents performance for aperture efficiencies of 50, 60, and 54% for receiving systems 1, 2, and 3 respectively for the three sets of values for γ_2 and γ_3 shown in Fig 4 An increase in aperture efficiency to 65% for the 34-meter-diameter listen antenna and a decrease in aperture efficiency to 50% for the 34-meter-diameter transmit/receive antenna reduce the improvement in predetection signal-to-noise ratio shown in Fig 4 by about 0 05 dB at $1 - \epsilon_1 = 0\ 3$ and has no effect when $1 - \epsilon_1 = 1$ for a given set of γ_2 and γ_3 values Figure 5 represents an array utilizing a 64-meter-diameter antenna (system 1) and two 34-meter-diameter antennas listen-only (systems 2 and 3) with 60% aperture efficiencies Nominal values of $\gamma_2 = \gamma_3 = 0\ 61$ provide a 2 55-dB improvement in predetection signal-to-noise ratio for $1 - \epsilon_1 = 1$ and $N_{o2}/N_{o1} = N_{o3}/N_{o1} = 0\ 925$ A value for 1 for N_{o2}/N_{o1} and N_{o3}/N_{o1} decreases the improvement in predetection signal-to-noise ratio shown in Fig 5 by 0 15 dB (or less) An increase in aperture efficiency to 65% for systems 2 and 3 reduces the predetection signal-to-noise ratio improvement shown in Fig 5 by about 0 2 dB at $1 - \epsilon_1 = 0\ 3$ and has no effect when $1 - \epsilon_1 = 1$ for a given set of γ_2 and γ_3 values

Figures 6, 7, 8, and 9 show the effect on predetection signal-to-noise ratio improvement as the voltage coupling β in the signal combining or summing process is varied for two and three receiving systems arrayed The array parameters for Figs 6, 7, 8, and 9 are as described above for Figs 2, 3, 4, and 5 respectively Also shown in Figs 6, 7, 8, and 9 is the effect of predetection signal-to-noise ratio improvement when the portion of predetection noise that is statistically independent ($1 - \epsilon_1$) is 1 0, 0 7, or 0 3 in receiving system 1

Consider next the equivalent RF carrier predetection signal-to-noise ratio improvement (RF carrier margin improvement) as determined from the total rms phase noise on the first local oscillator (see expressions (6), (7), (8), and (9) and associated discussion in Section II) The following sets of design parameters apply for the performance presented in this report The sets of parameters for receiving system 1 are

Threshold two-sided noise bandwidth			
$2B_{L_{o1}}$	12	152	30 Hz
Predetection IF filter noise bandwidth			
$NBW_{F_{A1}}$	2200	2200	2000 Hz

while the corresponding sets of parameters for receiving system 2 through N are

Threshold two-sided noise bandwidth $2B_{L_{o2,\ \ ,N}}$	0 1	1 0	0 3 Hz
Predetection IF filter noise bandwidth $NBW_{F_{A2,\ \ ,N}}$	2200	2200	2000 Hz

Using the bandwidth parameters above and the array parameters described in conjunction with Fig 2, Fig 10 shows the RF carrier margin improvement as a function of summing junction voltage coupling β Figure 10 represents an array of two receiving systems with a 64-meter-diameter antenna (system 1) and a 34-meter-diameter antenna listen-only (system 2) Performance is shown for γ_2 values of 0 55, 0 61, and 0 66 with $1 - \epsilon_1$ values of 1 0, 0 7, and 0 3 as in Fig 6 The information in Fig 10 can be rearranged to show RF carrier margin improvement as a function of the statistically independent portion of predetection noise in system 1 $(1 - \epsilon_1)$ Figure 11 shows this characteristic for a voltage coupling $\beta_2 = \gamma_2$ for γ_2 values of 0 55, 0 61, and 0 66

Figures 12 and 13 show similar RF carrier margin improvement characteristics for an array of two receiving systems with a 64-meter-diameter antenna and a 34-meter-diameter antenna, transmit/receive Similar characteristics are shown in Figure 14 and 15 for an array of three receiving systems with a 64-meter-diameter antenna (system 1), a 34-meter-diameter antenna listen-only (system 2) and a 34-meter-diameter antenna (transmit/receive) (system 3) Figures 16 and 17 show RF carrier margin improvement characteristics for an array of three receiving systems with a 64-meter-diameter antenna (system 1) and two 34-meter-diameter antenna (listen-only)

Some initial measurements of RF carrier margin improvement have been made in the laboratory, for two and three receiving systems arrayed, by measurement of rms phase noise (σ_{ϕ_n}) on the first local oscillator signal in system 1 and utilizing this σ_{ϕ_n} in conjunction with expressions (6) and (7) in Section II For these measurements, γ_2 and γ_3 values of 0 61 were set up in the laboratory to simulate reception with 34-meter-diameter antenna(s) (listen-only) arrayed with a 64-meter-diameter antenna (system 1) Measurements were made using the following set of parameters The predetection IF filter noise bandwidth was 2200 Hz with a $2B_{L_o}$ of 152 Hz for receiving system 1 and $2B_{L_o}$ of 1 Hz for receiving systems 2 and 3

The statistically independent noise $(1 - \epsilon)$ coupled into the input of each receiving system from separate random noise sources was equal so that $N_{o2}/N_{o1} = N_{o3}/N_{o1} = 1$ for $1 - \epsilon_1 = 1$ Coherent noise (ϵ) coupled into the input of the simulated 34-meter receiving systems was scaled by the ratio $(34/64)^2 \times (0\ 6/0\ 5)$ relative to the simulated 64-meter receiving system and adjusted for correlation The values 0 6 and 0 5 represent the aperture efficiencies, discussed earlier, of the 34-meter-diameter antenna listen-only and the 64-meter-diameter antenna respectively Table 1 shows the experimental measurements relative to the calculated results

IV. Discussion

The situation where a planet in the solar system is within the beamwidth of an array with equal antenna apertures was examined in Ref 1 (Section V) The receiving systems were arrayed for coherent reception of a signal from a spacecraft in deep space The same situation is considered herein except that the antennas of the array have unequal apertures as presented in the earlier sections of this report

Consider the case (as in Ref 1) of two receiving systems arrayed for coherent reception of the spacecraft signal A two-aperture (antenna) array will see the hot body (noise) radiation from the planet as an interferometer operating at a reception frequency wavelength λ with an antenna spacing $B_{1,2}$ Designate the component of $B_{1,2}$ that is perpendicular to the line of sight to the planet as $B_{r_{1,2}}$, then the resultant fringe spacing of the interferometer is $\lambda/B_{r_{1,2}}$ radians

A large part of the following material was presented in Ref 1 and is included here for clarification and continuity of discussion Consider for the moment, a situation where the planet appears as a point noise source That is, the angular size of the planet as seen from Earth is very small compared to the fringe spacing ($\lambda/B_{r_{1,2}}$) Within the fringe spacing (as the Earth rotates), the noise radiation from the planet as seen in the predetection bandwidth of the array is in phase for the two receivers at the center of an angular segment $\lambda/2B_{r_{1,2}}$, orthogonal at each edge of the angular segment, and 180 deg out of phase at the center of the adjacent angular segments $\lambda/2B_{r_{1,2}}$ due to continuing change in path length to the two antennas This in-phase/out-of-phase situation continues alternately during a station pass as the array tracks the spacecraft and it is superimposed on the statistically independent portion of predetection noise It should be noted again that the varying group delay on the signal modulation sidebands is tracked out among the receiving systems of the array during a station pass (see Section III, first paragraph) For this situation of a planet appearing as a point noise source and representing a given fractional part (ϵ) of operating equivalent system noise temperature (T_{op}), the decrease in predetection signal-to-noise ratio improvement shown in Figs 2 and 3 (for the corresponding $(1 - \epsilon)$) corresponds to the in-phase predetection noise condition at the center of an angular segment $\lambda/2B_{r_{1,2}}$ The orthogonal predetection point noise source condition corresponds to a signal-to-noise ratio improvement for $(1 - \epsilon) = 1$, while the 180-deg out-of-phase point noise source condition corresponds to a signal-to-noise ratio improvement with a portion of the point source contributions removed (cancelled) The discussion above also applies to RF carrier margin improvement as shown in Figs. 11 and 13 for two systems arrayed. Cross plots of the performance shown in Figs 6 and 7 and in Figs 10 and 12 would provide similar information for other signal summing voltage couplings

A planet for the solar system may not appear as a point noise source for the array considered here In fact, it can have an angular size that is larger than the fringe spacing $\lambda/B_{r_{1,2}}$ The following development addresses this consideration In general, for the application presented in this report, the planets (as viewed from Earth) can be characterized as disks with a uniform brightness distribution D at the reception frequency The brightness distribution will be a function of the reception frequency Utilizing information in Ref. 3 and writing the brightness transform as a function of the angular size ν of the planet and the inverse of fringe spacing, the brightness transform R becomes

$$R_{1,2} = \frac{1}{\nu} \int_{-\nu/2}^{+\nu/2} D(\omega) e^{-2\pi i \left(\frac{B_{r_{1,2}}}{\lambda}\right) \times dx} \tag{10}$$

with integration from the centerpoint of the planet disk out to the edges Integration of expression (10) provides

$$R_{1,2} = D(\omega) \cdot \frac{\sin\left(\pi \frac{B_{r_{1,2}}}{\lambda} \nu\right)}{\pi \frac{B_{r_{1,2}}}{\lambda} \nu} \tag{11}$$

For the situation where the planet appears as a point noise source, the sin x/x type of expression above approaches 1 The brightness transform becomes the total power of the noise source (planet) as seen in the predetection bandwidth, and the preceding discussion relating to a point noise source applies Figure 18 shows expression (11) plotted as a function of the ratio of planet angular size to fringe spacing $[\nu/(\lambda/B_{r_{1,2}})]$ Note that the above integral is centered ($\nu = 0$) on the in-phase detection noise situation at the center of an angular segment $\lambda/2B_{r_{1,2}}$ that was discussed earlier in this section of the report Figure 18 is the resultant amplitude of the brightness transform for this particular path length situation to the two antennas for various ratios of planet angular size to fringe spacing. This figure shows the magnitude of the correlated portion of noise power relative to total noise power from the planet, which is defined as fringe visibility for radio interferometry measurements (Ref 3) Consequently, Fig 18 provides the information necessary to scale the coherent portion of predetection noise relative to a point noise source for any ratio of planet angular size to fringe spacing $[\nu/(\lambda/B_{r_{1,2}})]$ Any point on Fig 18 can be treated as an equivalent reduced-in-magnitude point noise source representing a reduced ϵ, and the resulting reduction in predetection signal-to-noise ratio and RF carrier margin improvement can be determined from Figs 2, 3, 6, 7, 10, 11, 12 and 13

Consider next three receiving systems (1, 2, and 3) arrayed for coherent reception of a spacecraft signal with a planet within the beamwidth of the three antennas For the moment, consider a situation where the planet appears as a point noise source which represents a given fractional part ϵ of the operating equivalent system noise temperature T_{op} At those times when the noise radiation from the planet as seen in the predetection bandwidth of the array is in phase from receivers 1 and 3 in an angular segment $\lambda/2B_{r_{1,3}}$ at essentially the same time as it is from receivers 1 and 2 in an angular segment $\lambda/2B_{r_{1,2}}$, the decrease in predetection signal-to-noise ratio improvement shown in Figs 4 and 5 applies for the corresponding $(1 - \epsilon)$ value At all other times, the decrease in predetection signal-to-noise ratio improvement will be less This discussion also applies to RF carrier margin improvement shown in Figs 15 and 17 for three systems arrayed Cross plots of the performance shown in Figs 8 and 9 and in Figs 14 and 16 would provide similar information for other signal summing voltage couplings

For the situation where the planet is not a point source, the normalized brightness transforms for three receiving systems become

$$R_{1,2,3} = \frac{D(\omega)}{2}\left(\frac{\sin\left(\pi\frac{B_{r_{1,2}}}{\lambda}\nu\right)}{\pi\frac{B_{r_{1,2}}}{\lambda}\nu} + \frac{\sin\left(\pi\frac{B_{r_{1,3}}}{\lambda}\nu\right)}{\pi\frac{B_{r_{1,3}}}{\lambda}\nu}\right) \tag{12}$$

Note again in this case that the above expression (12) represents a result centered ($\nu = 0$) on the in-phase predetection noise situation at the center of angular segment $\lambda/2B_{r1,2}$ and simultaneously at the center of angular segment $\lambda/2B_{r1,3}$ For this particular path length situation and given fringe spacings $\lambda/B_{r1,2}$ and $\lambda/B_{r1,3}$, expression (12) can be evaluated and the effect of the planet on decrease in predetection signal-to-noise ratio and RF carrier margin improvement can be determined from Figs 4, 5, 8, 9, 14, 15, 16, and 17

In order to illustrate the effect of a planet in the solar system on signal-to-noise ratio improvement for an array of receiving systems as described in Sections II and III of this report, consider reception at 8420 MHz ($\lambda = 3\ 56$ cm) Table 2 shows nominal radio disk temperatures T_R of the planets in the solar system at a reception wavelength of 3 55 to 3 6 cm, angular size as seen from Earth, and increase in system noise temperature due to the planet in the antenna beamwidth of the 64-meter-diameter antenna References 4, 5, 6, 7, and 8 provide the basis for the information shown in Table 2 Increase in system noise temperature ΔT_A due to a planet in the antenna beamwidth is obtained (except for Jupiter, due to its radiation belts) from the expression

$$\Delta T_A = T_R \times \frac{\text{planet (source) solid angle}}{\text{antenna beam solid angle}}$$

where the antenna beam solid angle is calculated from the expression (4/3)(half-power beamwidth)^2 (Ref 4) Increase in system noise temperature due to Jupiter is obtained from the normalized flux density at 4 04 AU using the expression

$$\Delta T_A = \frac{\text{Antenna Geometrical Area} \times \text{Aperture Efficiency} \times \text{Flux Density}}{2 \times (\text{Boltzmann's constant})}$$

and scaling ΔT_A for minimum and maximum distance from Earth by (ratio of distances)^2 (Ref 8) It is interesting to note that ΔT_A for the outer planets Saturn, Uranus, and Neptune can also be obtained from their normalized flux densities (as with Jupiter), utilizing the information in Ref 8 The ΔT_A for these three planets obtained by this method essentially agree with the ΔT_A obtained by using T_R and solid angles as described above It should be noted that expressions (11) and (12) above assume that the planet disk has a uniform brightness distribution $D(w)$ at the reception frequency Nonthermal radio emission from Jupiter's radiation belts contributes to Jupiter's brightness temperature at a wavelength of 3 56 cm (Refs 5 and 6), and in addition, the brightness temperature changes with time at this wavelength Consequently, the above assumption of uniform

brightness distribution is not strictly true for Jupiter and a small error results Note that the planets Jupiter and Venus produce the largest increase in system noise temperature

Consider two receiving systems (a 64- and 34-meter-diameter antenna) arrayed for reception of a spacecraft signal A planet is within the beamwidth of the antennas The following set of components of antenna feed spacing perpendicular to the line of sight to the planet ($B_{r_{1,2}}$) provides a corresponding set of fringe spacings ($\lambda/B_{r_{1,2}}$) as shown below

$B_{r_{1,2}}$, meters	Fringe spacing ($\lambda/B_{r_{1,2}}$), arc sec
100	73 50
200	36 75
300	24 50
400	18 37
500	14 70

Consider a situation where a $B_{r_{1,2}}$ of 200 meters represents operation of a 64-meter-diameter antenna receiving system (system 1) at an antenna elevation angle of about 60 deg (fringe spacing is 36 75 arcsec) The system noise temperature is 20 + 1 or 21 kelvins (excluding noise contribution from a planet) The planet Jupiter is in the beamwidth of each of the two antennas Jupiter is at closest approach to the Earth so that it subtends an angle of 50 arcsec and its contribution to the system noise temperature of the 64-meter-diameter antenna receiving system is 13 kelvins (Table 2) The ratio of 50/36 75 equals 1 36 which, from Fig 18, provides a reduction in relative magnitude of correlated noise power from one to 0.21, which represents 2.7 kelvins The operating system noise temperature (system 1) is 21 + 13 or 34 kelvins The statistically independent portion ($1 - \epsilon_1$) of predetection noise power is (34 - 2 7)/34 or 0 92 for system 1 For an array of two receiving systems utilizing a 34-meter-diameter transmit/receive antenna in conjunction with a 64-meter-diameter antenna (Fig 3), the improvement in predetection signal-to-noise ratio is reduced from 1 0 dB for $1 - \epsilon_1 = 1\ 0$ to 0 91 dB for $1 - \epsilon_1 = 0\ 92$ or 0 09 dB ($\gamma_2 = 0\ 53$) The reduction in RF carrier margin improvement (Fig 13) is also 0 09 dB This 0 09-dB reduction in array performance at $1 - \epsilon_1 = 0\ 92$ essentially applies over the range of γ_2 values shown in Figs 3 and 13 and also to Figs 7 and 12 over a large portion of the voltage coupling (β_2) range shown Note that, in light of the discussion above relative to expressions (10) and (11), this 0 09-dB reduction applies at those times when the path length to the two antennas represents an in-phase predetection noise situation (2 7 kelvins) due to the planet Jupiter At those times when the path lengths represent an out-of-phase predetection noise situation for the 2 7 kelvins, the predetection signal-to-noise ratio and carrier margin improvement will be larger than the values shown in Figs 3 and 13 for $1 - \epsilon_1 = 1\ 0$ The periodic reduction in predetection signal-to-noise ratio and RF carrier margin improvement for the situation discussed above is included as the first array configuration (first line) in Table 3 under the column titled Maximum Periodic Reduction in Array Gain

Table 3 shows periodic reduction in array gain (as defined in the discussion above) at a reception frequency of 8420 MHz (λ = 3 56 cm) for two receiving systems arrayed (64- and 34-meter-diameter antenna) with the planets Jupiter or Venus in the antenna beamwidth at either minimum or maximum distance from Earth The table includes a few selected values of $B_{r_{1,2}}$ (component of antenna feed spacing perpendicular to the line of sight to the planet) at antenna elevation angles of approximately 60 and 20 deg The $B_{r_{1,2}}$ value of 100 meters (shown with an asterisk) is not representative of planned antenna spacings for the DSN It is included to illustrate the effect on array gain reduction due to the planets for closer antenna feed spacings The planets Jupiter and Venus were chosen since their presence in the antenna beamwidth produces the largest effect on periodic array gain reduction.

The calculated performance shown in Table 3 is expanded in Figs 19 through 26 to provide a more graphic presentation of the maximum periodic reduction in array gain as a function of $B_{r_{1,2}}$ for Jupiter (Figs 19, 20, 23, and 24), or for Venus (Figs 21, 22, 25, and 26)) in the array beamwidth at their minimum and maximum distances from Earth Planned antenna spacings and orientation relative to the spin axis of the Earth provide a $B_{r_{1,2}}$ in the range from 195 to 290 meters for a 34-meter-diameter

transmit/receive antenna arrayed with a 64-meter-diameter antenna and a $B_{r_{1,2}}$ in the range from 445 to 550 meters for a 34-meter-diameter listen-only antenna arrayed with a 64-meter-diameter antenna

Note that for a 34-meter transmit/receive and 64-meter-diameter antenna array, a $B_{r_{1,2}}$ in the 195 to 290 meter range results in a maximum periodic reduction in array gain of 0 1 dB or less for Jupiter in the array beamwidth and 0 17 dB or less for Venus in the array beamwidth For a 34-meter listen-only and 64-meter-diameter antenna array, a $B_{r_{1,2}}$ in the range from 445 to 550 meters results in a maximum periodic reduction in array gain of 0 05 dB or less for Jupiter in the array beamwidth and 0 1 dB or less for Venus in the array beamwidth

Table 4 shows periodic reduction in array gain at a reception frequency of 8420 MHz for an array of three receiving systems with the planets Jupiter or Venus in the antenna beamwidth as in Table 3 The array utilizes a 64-meter-diameter antenna (system 1), a 34-meter-diameter listen-only antenna (system 2) and a 34-meter-diameter transmit/receive antenna (system 3) with antenna feed spacings (component perpendicular to the line of sight to the planet) of 500 meters ($B_{r_{1,2}}$) and 200 meters ($B_{r_{1,3}}$) This is representative of the planned three station array in Australia at an elevation angle of approximately 50 deg

It should be noted that with the planets Uranus or Neptune in the antenna beamwidth (see Table 2) their effect on array gain is negligible even though (for the planned antenna feed spacings discussed above) most of the noise contribution from Uranus or Neptune will be coherent periodically

References

1 Brockman, M H , "The Effect of Partial Coherence in Receiving System Noise Temperature on Array Gain for Telemetry and Radio Frequency Carrier Reception for Similar Receiving Systems," in *TDA Progress Report 42-66*, pp 219-235, Jet Propulsion Laboratory, Pasadena, Calif , Dec 15, 1981

2 Divsalar, D , and Yuen, J H , "Improved Carrier Tracking Performance with Coupled Phase-Locked Loops," in *TDA Progress Report 42-66*, pp 148-171, Jet Propulsion Laboratory, Pasadena, Calif , Dec 15, 1981

3 Thomas, J B , *An Analysis of Source Structure Effects in Radio Interferometry Measurements*, Publication 80-84, Jet Propulsion Laboratory, Pasadena, Calif , Dec 15, 1980

4 Kraus, J D , *Radio Astronomy*, McGraw-Hill Book Company, New York, 1966

5 Newburn, R L , Jr , and Gulkis, S , "A Survey of the Outer Planets Jupiter, Saturn, Uranus, Neptune, Pluto, and Their Satellites," *Space Science Reviews*, 3, pp 179-271, 1973

6 Gulkis, S , "Thermal Radio Emissions from the Major Planets," *Space Science Reviews*, 14, pp 497-510, 1973

7 Allen, C W , *Astrophysical Quantities*, 3rd ed , University of London, The Athlone Press, London, 1973

8 Turegano, J A , and Klein, M J , "Precision Flux Density Measurements of the Giant Planets at 8420 MHz," *Astronomy & Astrophysics*, 94, pp 91-94, 1981

Table 1. RF carrier array gain measurements

($\gamma_2 = \gamma_3 = 0.61, N_{o2}/N_{o1} = N_{o3}/N_{o1} = 1$ for $1 - \epsilon_1 = 1$)

Two receiving systems				Three receiving systems			
		Array gain, dB				Array gain, dB	
$1 - \epsilon_1$	β_2	Theory	Measured	$1 - \epsilon_1$	$\beta_2 = \beta_3$	Theory	Measured
1.0	0.5	1.3	1.3	1.0	1.0	1.9	1.8
1.0	0.43	1.23	1.2	1.0	0.97	1.94	1.9
0.73	0.5	0.92	0.9	1.0	0.61	2.28	2.3
0.73	0.43	0.88	0.8	0.73	1.0	1.33	1.2
0.21	0.5	0.28	0.2	0.73	0.61	1.55	1.6
0.21	0.41	0.24	0.2	0.21	0.59	0.46	0.4

Table 2 Increase in system noise temperature due to the planets; 64-meter-diameter antenna; λ = 3.55 to 3.6 km

Planet	Nominal radio disk temperature (T_R), K	Subtended angle from earth, arcsec		Increase in system noise temperature, K	
		max	min	max	min
Mercury	330	13	5	1.8	0.3
Venus	600	67	17	85.0	5.5
Mars	200	24	4	3.6	0.1
Jupiter	230	50	32	13.0	5.0
Saturn	165	20	14	1.4	0.7
Uranus	210	3.6		0.08	
Neptune	200	2.2		0.03	

Table 3. Maximum periodic reduction in array gain due to a planet within beamwidth of two-aperture array (two receiving systems arrayed for coherent reception, reception frequency 8420 MHz, λ = 3 56 cm)

Array configuration	Antenna elevation angle, deg	Planet in array beamwidth	Planet subtended angle from earth ν, arcsec	$B_{r1,2}$ m	$\nu/\left(\frac{\lambda}{B_{r1,2}}\right)$	Maximum periodic reduction in array gain, dB
64- and 34-m transmit/receive	~60	Jupiter	50	200	1 36	0 09
	~20		50	200	1 36	0 07
	~60		32	200	0 87	0 04
	~20		32	200	0 87	0 03
	~60		50	100[a]	0 68	0 18
	~20		50	100[a]	0 68	0 14
	~60		32	100[a]	0 44	0 16
	~20		32	100[a]	0 44	0 12
	~60	Venus	67	200	1 82	0 09
	~20		67	200	1 82	0 08
	~60		17	200	0 46	0 16
	~20		17	200	0 46	0 12
	~60		67	100[a]	0 91	0 09
	~20		67	100[a]	0 91	0 08
	~60		17	100[a]	0 23	0 22
	~20		17	100[a]	0 23	0 16
64- and 34-m listen-only	~60	Jupiter	50	200	1 36	0 13
	~20		50	200	1 36	0 09
	~60		32	200	0 87	0 05
	~20		32	200	0 87	0 04
	~60		50	500	3 40	0 05
	~20		50	500	3 40	0 04
	~60		32	500	2 18	0 03
	~20		32	500	2 18	0 02
	~60	Venus	67	200	1 82	0 13
	~20		67	200	1 82	0 11
	~60		17	200	0 46	0 22
	~20		17	200	0 46	0 15
	~60		67	500	4 56	0 09
	~20		67	500	4 56	0 08
	~60		17	500	1 16	0 04
	~20		17	500	1 16	0 03

[a]Not planned for the DSN

Table 4. Maximum periodic reduction in array gain due to a planet within beamwidth of three-aperture array (three receiving systems arrayed for coherent reception; reception frequency 8420 MHz; λ = 3.56 cm)

Array configuration	Antenna elevation angle, deg	Planet in array beamwidth	Planet subtended angle from earth ν, arc sec	$B_{r1,2}$, m	$\nu/\left(\frac{\lambda}{B_{r_{1,2}}}\right)$	$B_{r1,3}$, m	$\nu/\left(\frac{\lambda}{B_{r_{1,3}}}\right)$	Maximum periodic reduction in array gain, dB
64- and 34-m	~60	Jupiter	50	500	3 40	200	1 36	0 15
listen-only and	~20		50	500	3 40	200	1 36	0 12
34-m transmit/	~60		32	500	2 18	200	0 87	0 06
receive	~20		32	500	2 18	200	0 87	0 04
	~60	Venus	67	500	4 56	200	1 83	0 18
	~20		67	500	4 56	200	1 83	0 17
	~60		17	500	1 16	200	0 46	0 23
	~20		17	500	1 16	200	0 46	0 17

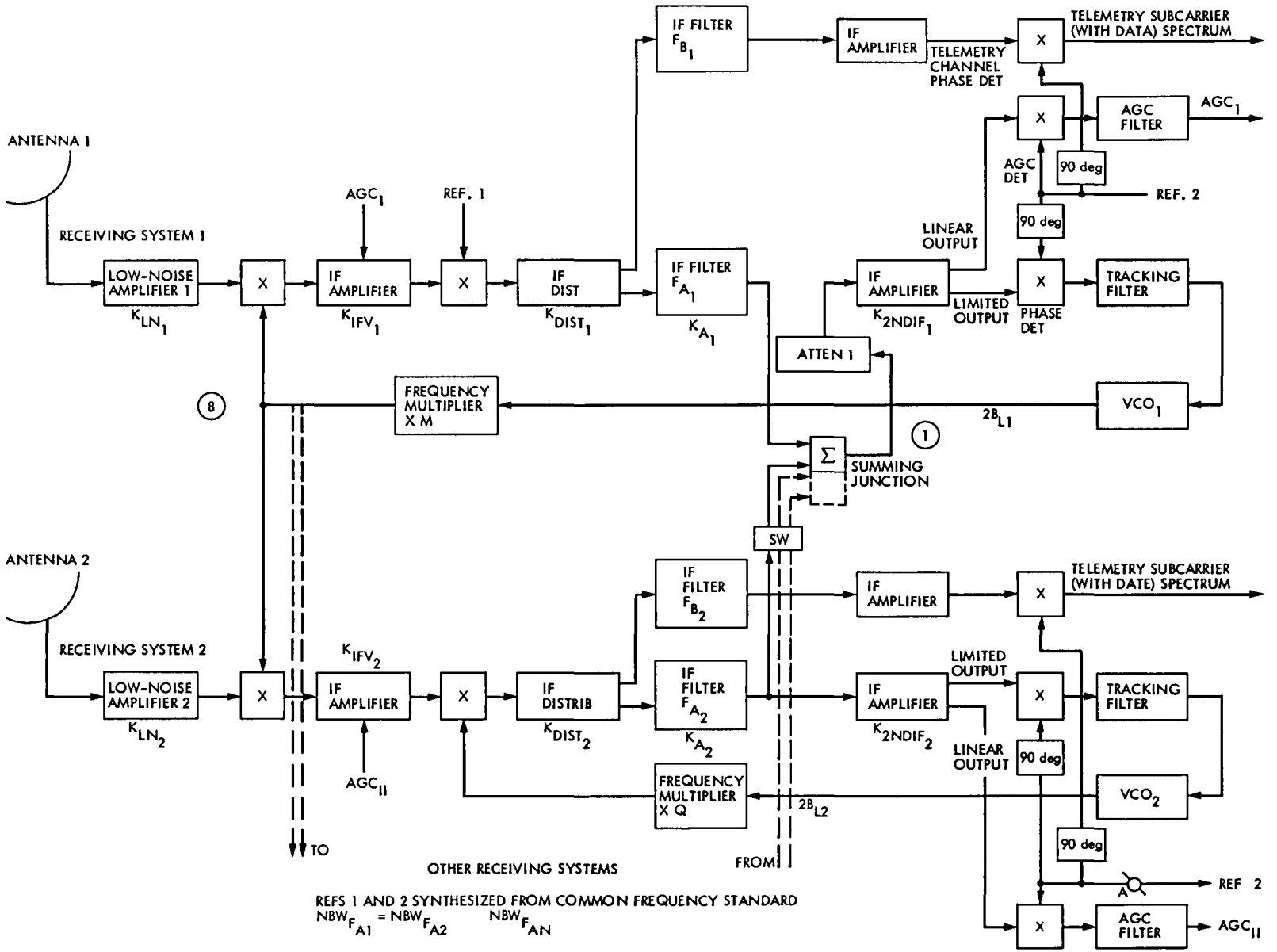


Fig. 1. Radio frequency carrier arraying for high-rate telemetry reception

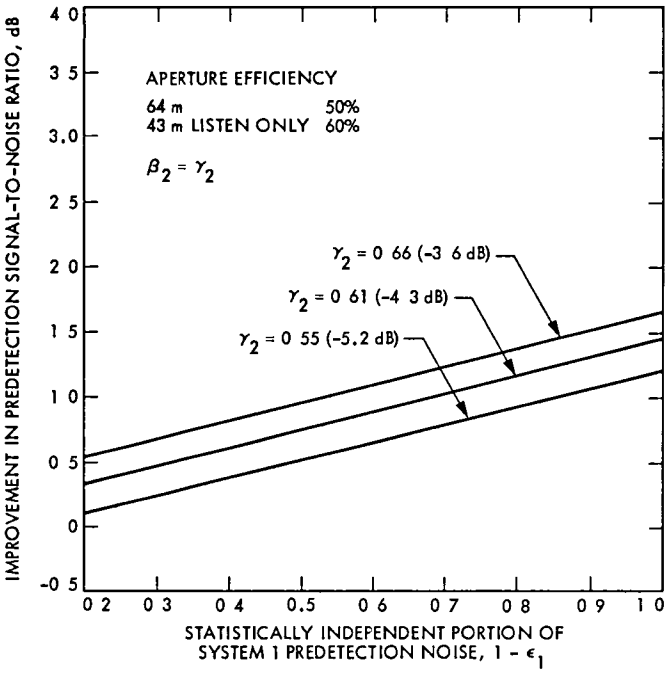


Fig. 2. Effect of partial coherence in system noise temperature on predetection signal-to-noise ratio improvement, two receiving systems arrayed, 64- and 34-m-diameter listen-only antennas

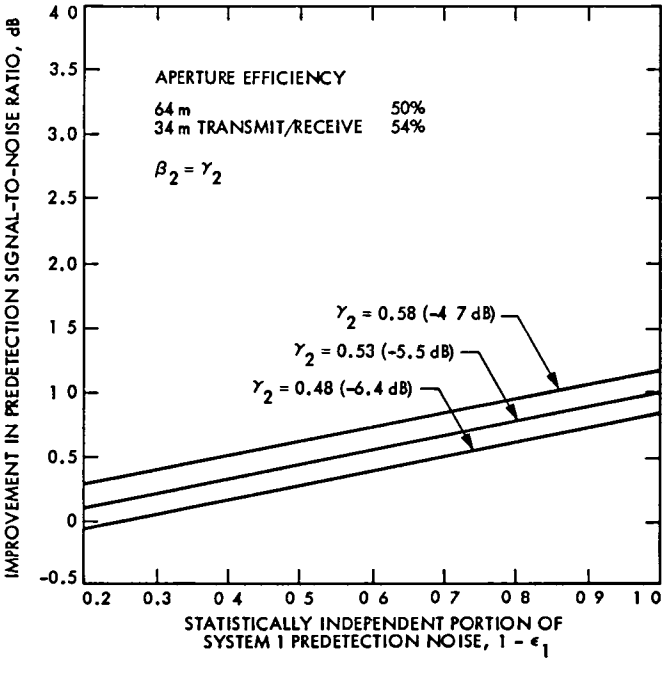


Fig. 3. Effect of partial coherence in system noise temperature on predetection signal-to-noise ratio improvement, two receiving systems arrayed, 64- and 34-m-diameter transmit/receive antennas

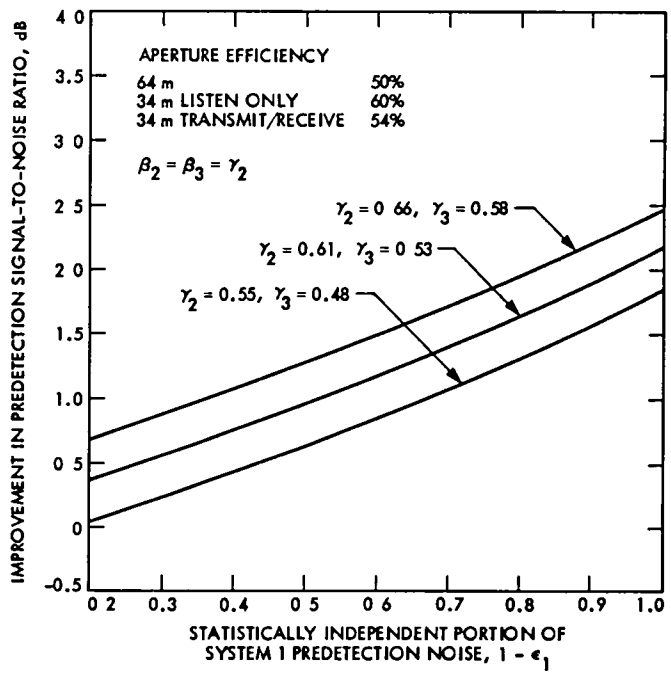


Fig. 4. Effect of partial coherence in system noise temperature on predetection signal-to-noise ratio improvement, three receiving systems arrayed; 64- and 34-m-diameter listen-only and 34-m-diameter transmit/receive antennas

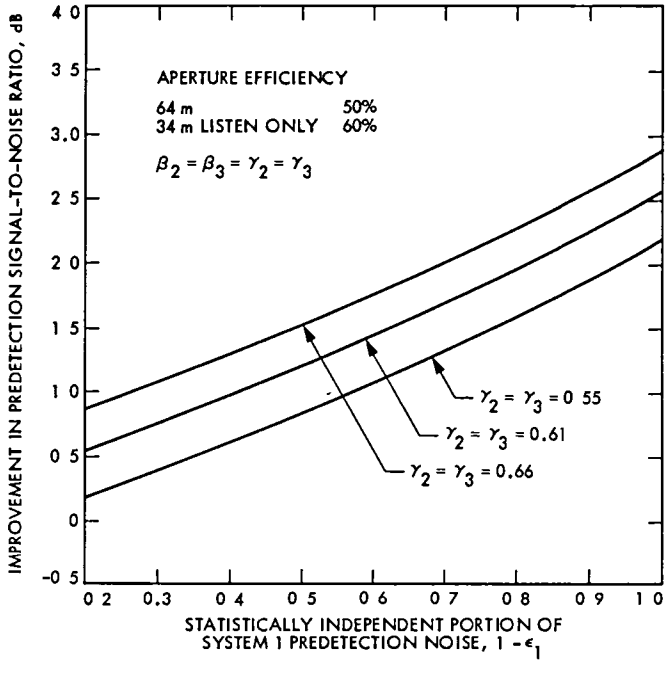


Fig. 5. Effect of partial coherence in system noise temperature on predetection signal-to-noise ratio improvement, three receiving systems arrayed, 64-m-diameter and two 34-m-diameter listen-only antennas

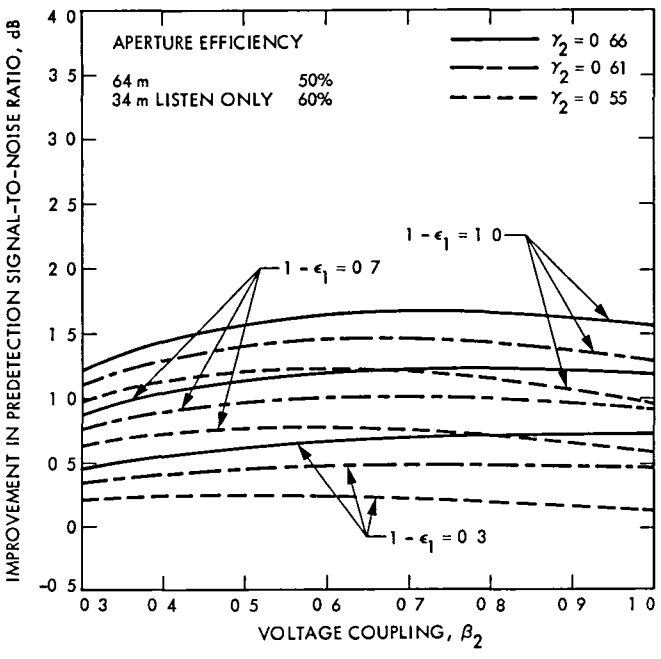


Fig. 6. Effect of signal summing voltage coupling on predetection signal-to-noise ratio improvement with partial coherence in system noise temperature, two receiving systems arrayed, 64- and 34-m-diameter listen-only antennas

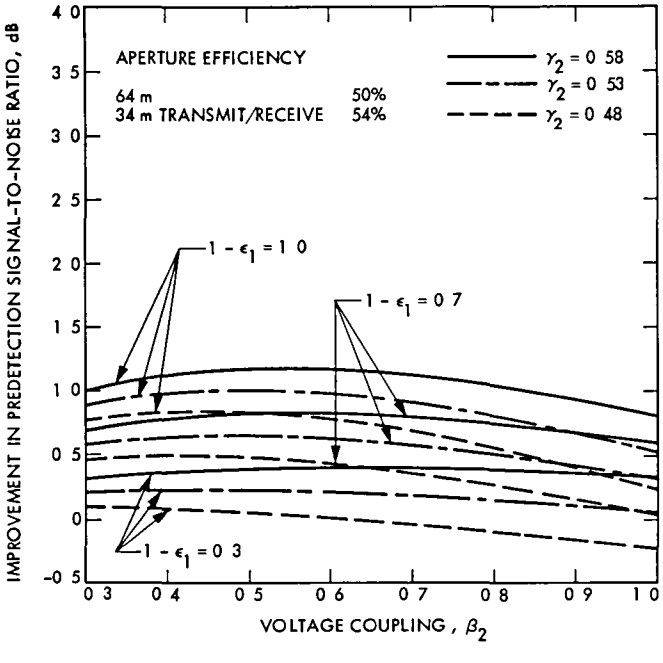


Fig. 7. Effect of signal summing voltage coupling on predetection signal-to-noise ratio improvement with partial coherence in system noise temperature, two receiving systems arrayed, 64- and 34-m-diameter transmit/receive antennas

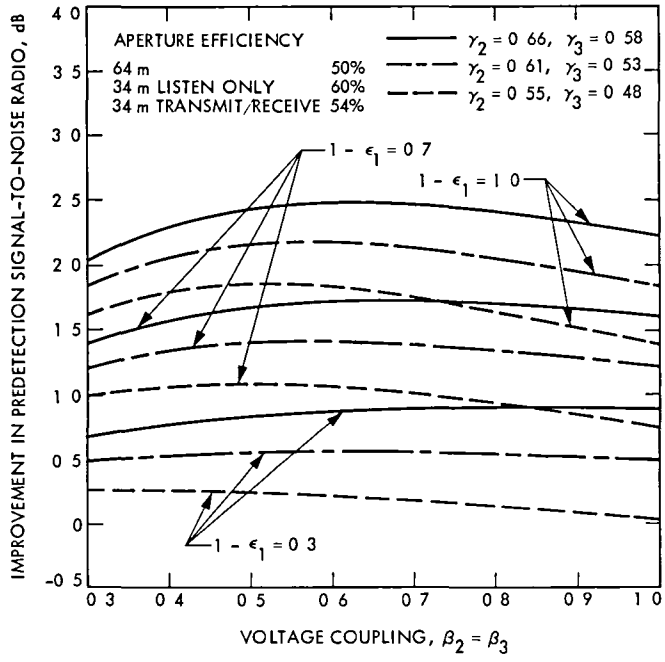


Fig. 8. Effect of signal summing voltage coupling on predetection signal-to-noise ratio improvement with partial coherence in system noise temperature, three receiving systems arrayed; 64-m-diameter, 34-m-diameter listen only, and 34-m-diameter transmit/receive antennas

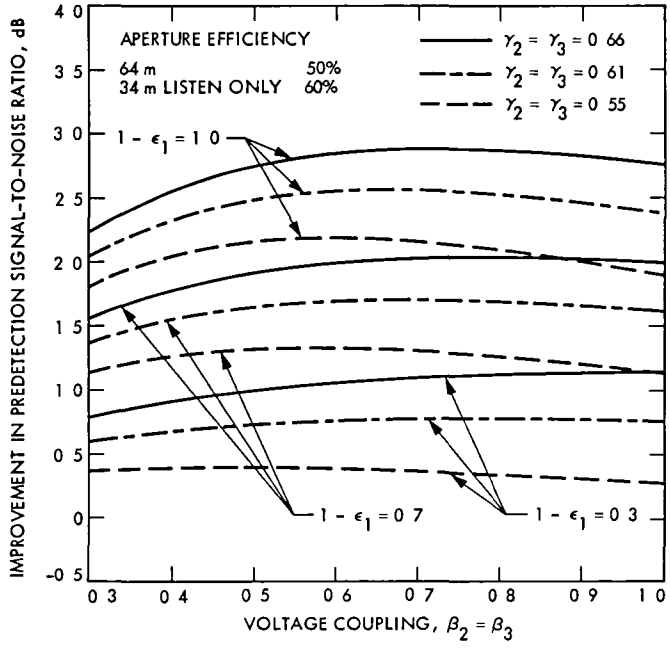


Fig 9 Effect of signal summing voltage coupling on predetection signal-to-noise ratio improvement with partial coherence in system noise temperature, three receiving systems arrayed, 64-m-diameter and two 34-m-diameter listen-only antennas

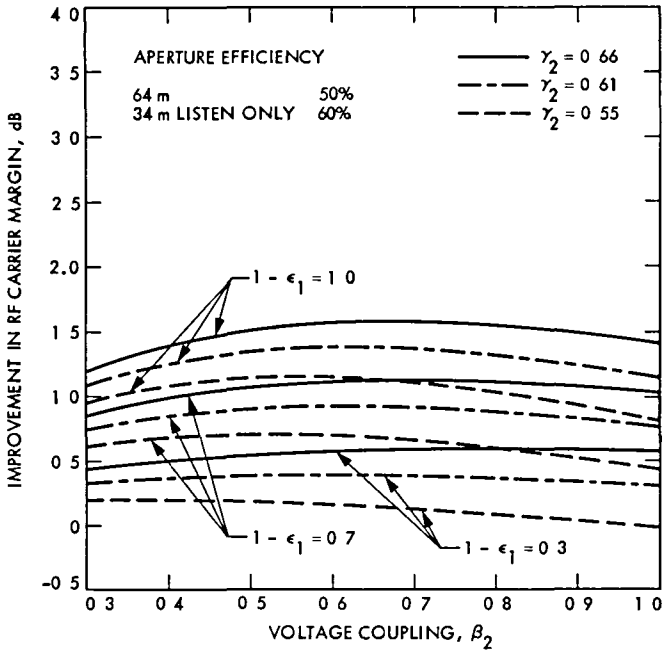


Fig. 10. Effect of summing junction voltage coupling on RF carrier margin improvement with partial coherence in system noise temperature, two receiving systems arrayed, 64- and 34-m-diameter listen-only antennas

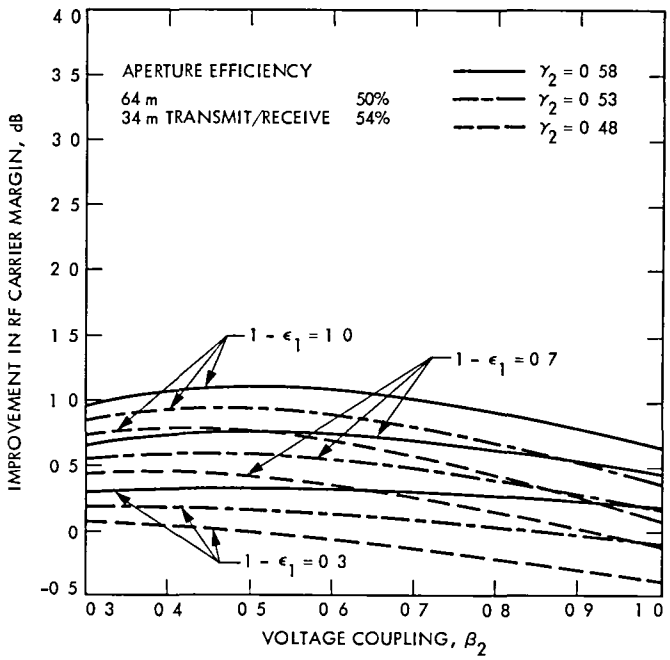


Fig. 12. Effect of summing junction voltage coupling on RF carrier margin improvement with partial coherence in system noise temperature, two receiving systems arrayed, 64- and 34-m-diameter transmit/receive antennas

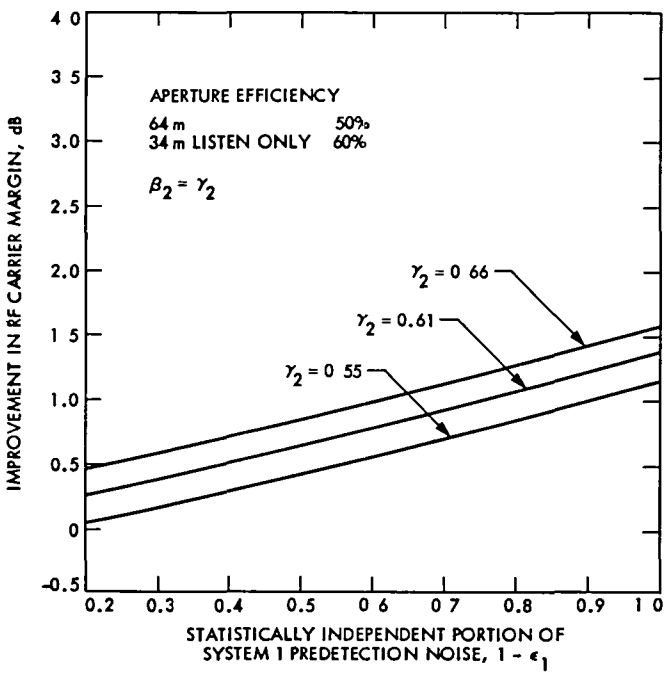


Fig. 11. Effect of partial coherence in system noise temperature on RF carrier margin improvement, two receiving systems arrayed, 64- and 34-m-diameter listen-only antennas

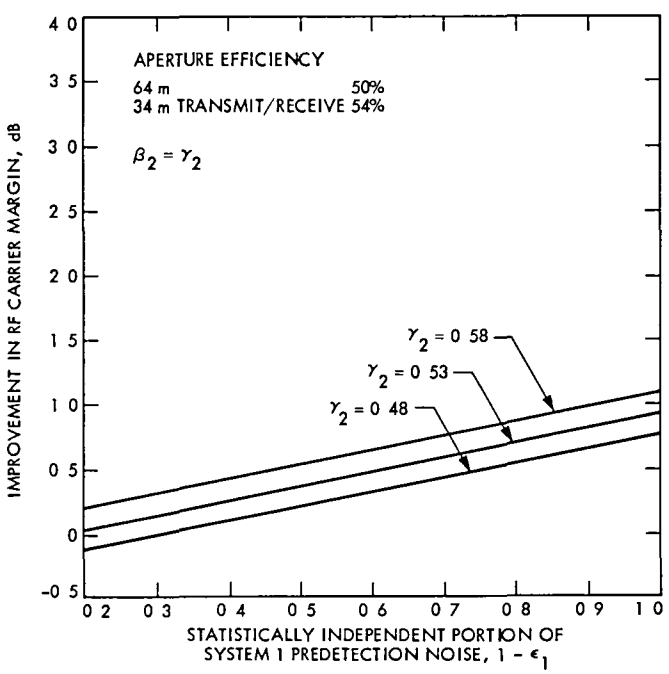


Fig 13 Effect of partial coherence in system noise temperature on RF carrier margin improvement, two receiving systems arrayed, 64- and 34-m-diameter transmit/receive antennas

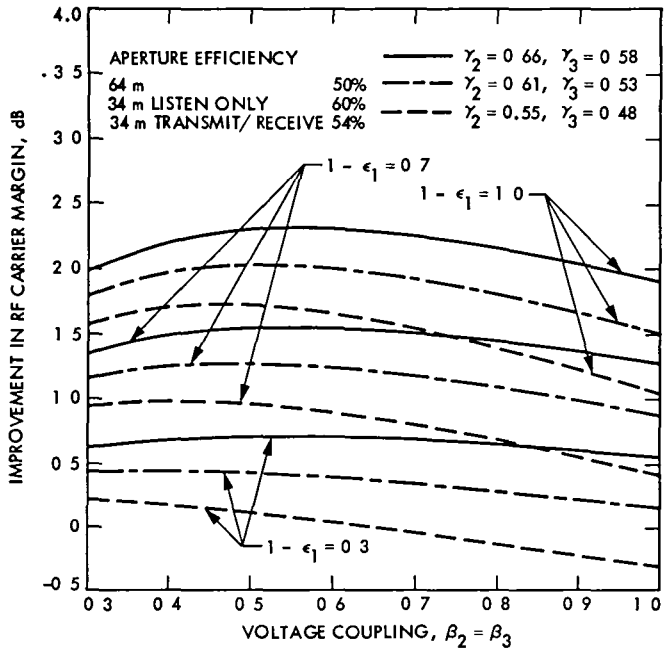


Fig. 14. Effect of summing junction voltage coupling on RF carrier margin improvement with partial coherence in system noise temperature, three receiving systems arrayed, 64-m-diameter, 34-m-diameter listen-only, and 34-m-diameter transmit/receive antennas

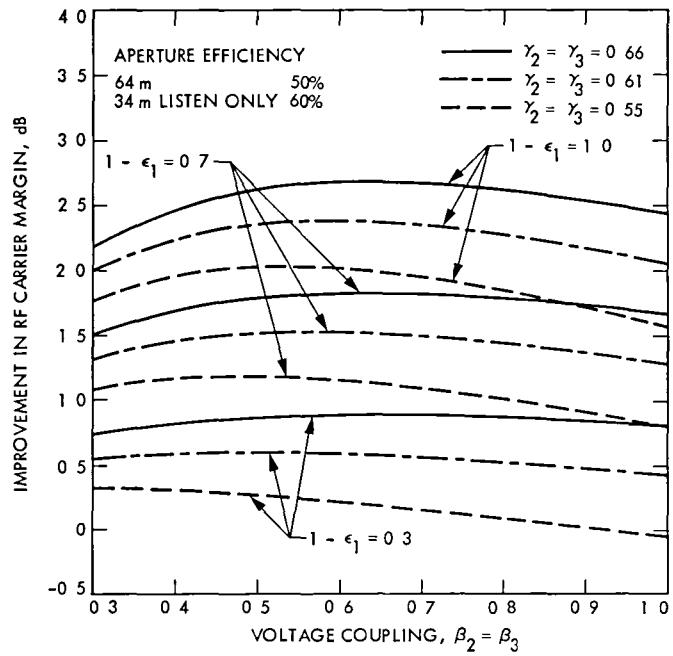


Fig. 16. Effect of summing junction voltage coupling on RF carrier margin improvement, with partial coherence in system noise temperature, three receiving systems arrayed; 64-m-diameter and two 34-m-diameter listen-only antennas

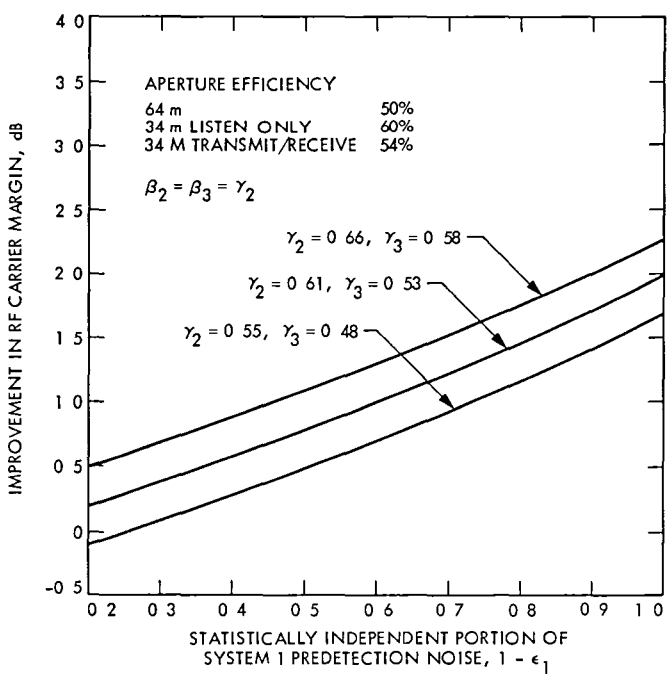


Fig. 15 Effect of partial coherence in system noise temperature on RF carrier margin improvement, three receiving systems arrayed; 64-m-diameter, 34-m-diameter listen-only, and 34-m-diameter transmit/receive antennas

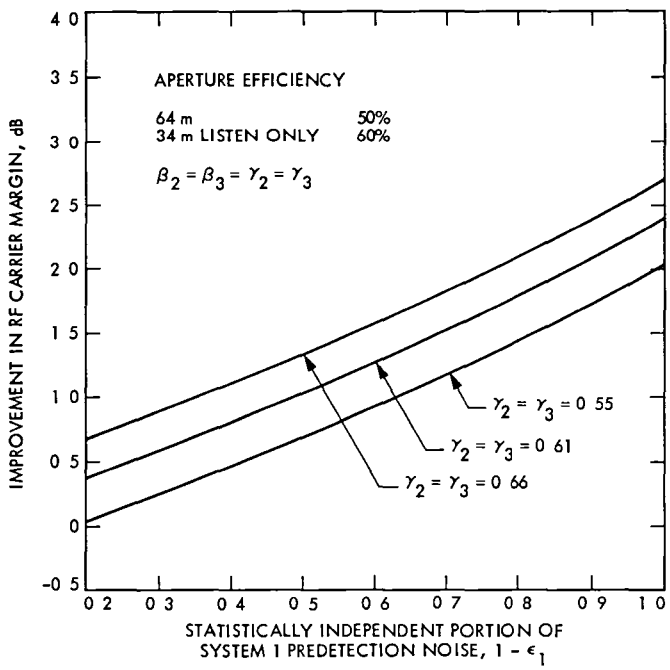


Fig. 17. Effect of partial coherence in system noise temperature on RF carrier margin improvement, three receiving systems arrayed; 64-m-diameter and two 34-m-diameter listen-only antennas

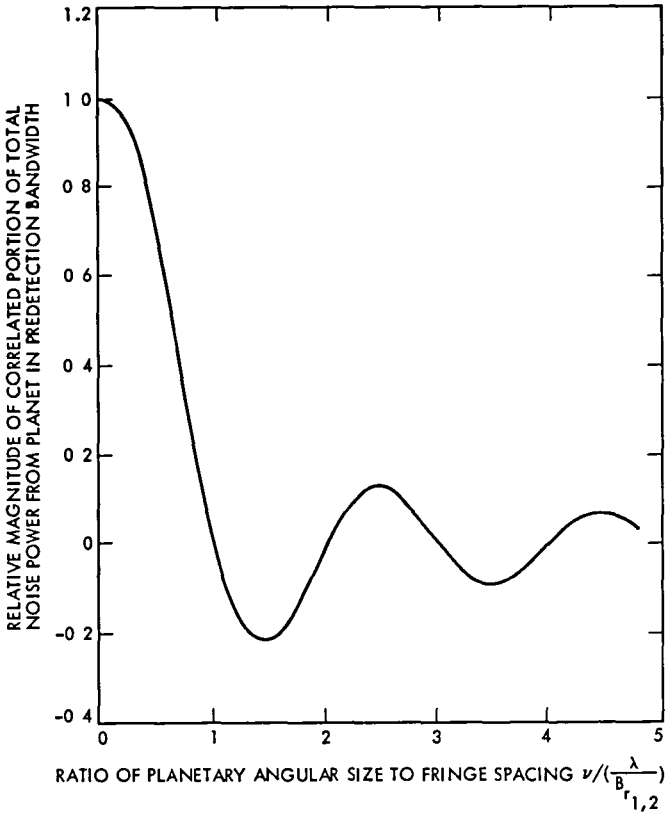


Fig. 18. Relative effect of ratio of planet angular size to fringe spacing on correlated predetection noise, two-antenna array

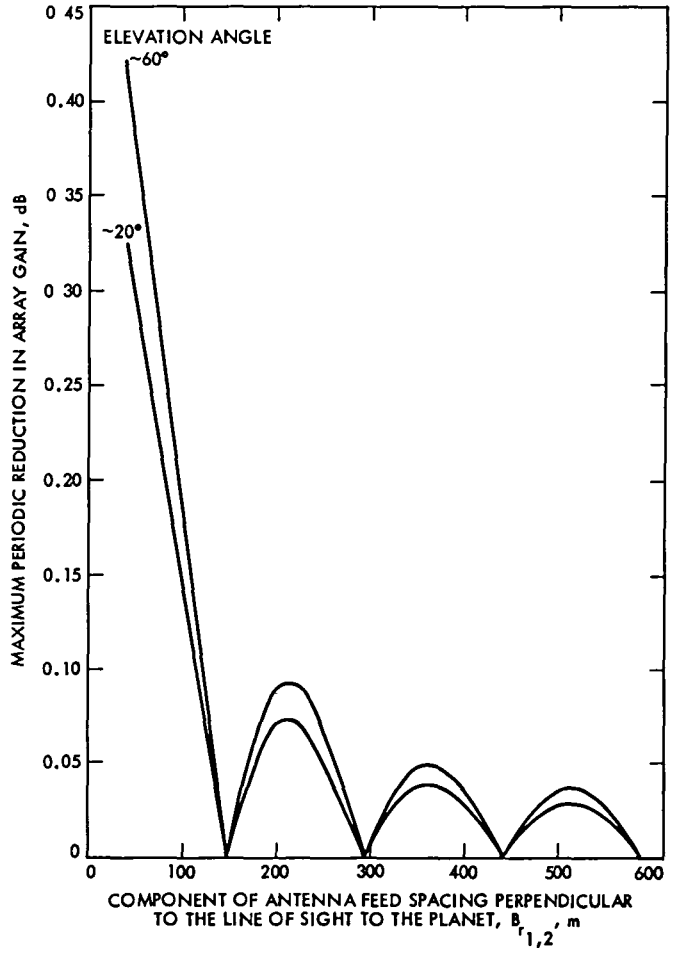


Fig. 19. Maximum periodic reduction in array gain due to Jupiter (at minimum distance) within beamwidth of two receiving systems arrayed for coherent reception vs effective feed spacing; 64- and 34-m-diameter (transmit/receive) antennas, reception at 8420 MHz

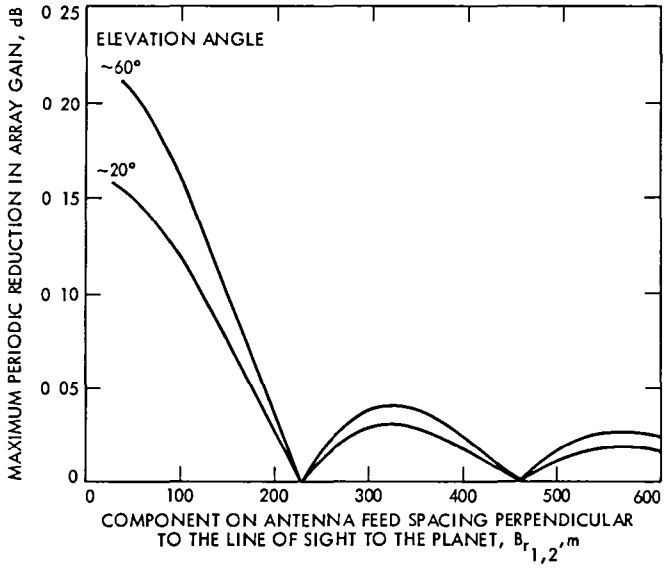


Fig. 20. Maximum periodic reduction in array gain due to Jupiter (at maximum distance) within beamwidth of two receiving systems arrayed for coherent reception vs effective feed spacing, 64- and 34-m-diameter (transmit/receive) antennas, reception at 8420 MHz

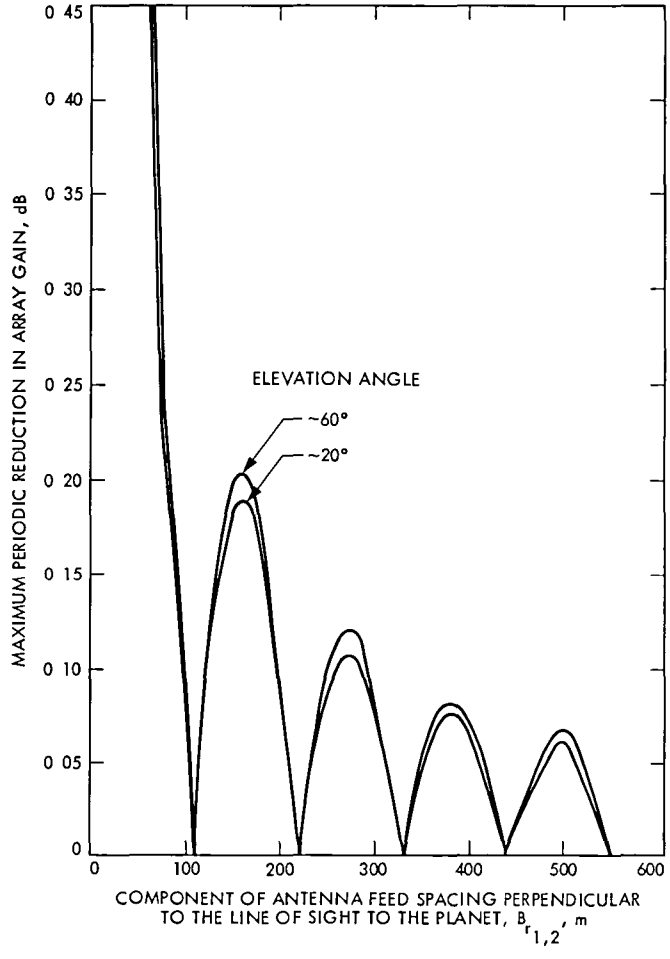


Fig 21 Maximum periodic reduction in array gain due to Venus (at minimum distance) within beamwidth of two receiving systems arrayed for coherent reception vs effective feed spacing, 64- and 34-m-diameter (transmit/receive) antennas, reception at 8420 MHz

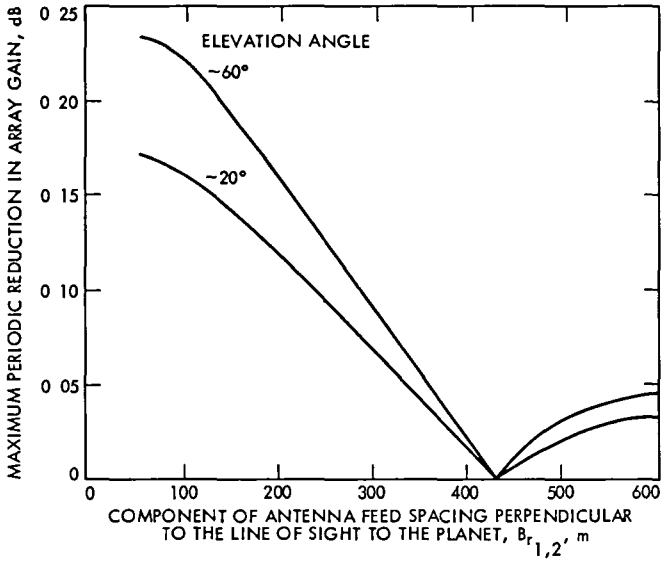


Fig. 22. Maximum periodic reduction in array gain due to Venus (at maximum distance) within beamwidth of two receiving systems arrayed for coherent reception vs effective feed spacing; 64- and 34-m-diameter (transmit/receive) antennas, reception at 8420 MHz

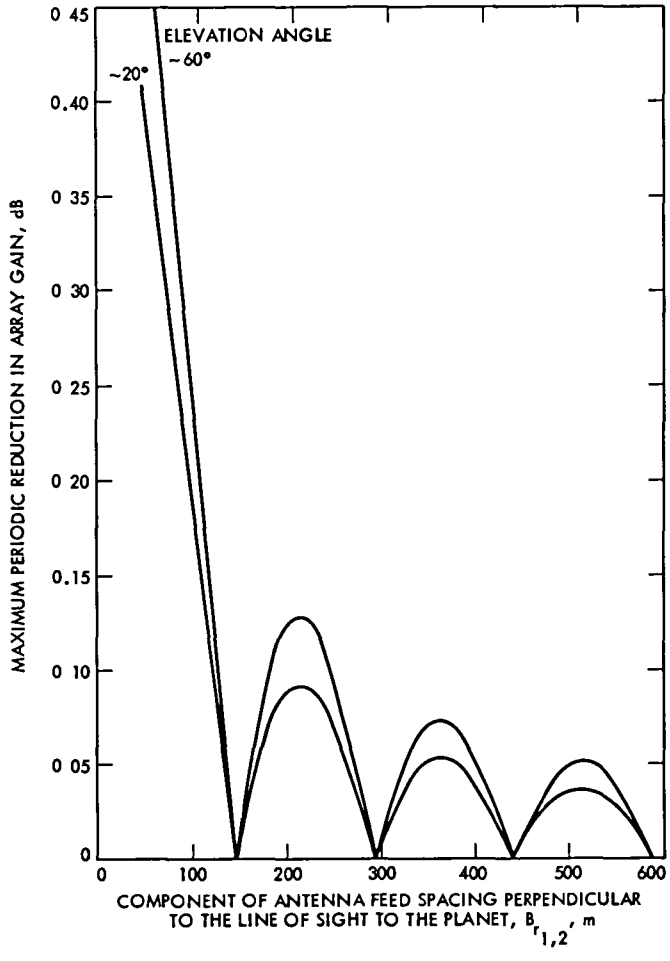


Fig. 23. Maximum periodic reduction in array gain due to Jupiter (at minimum distance) within beamwidth of two receiving systems arrayed for coherent reception vs effective feed spacing; 64- and 34-m-diameter (listen-only) antennas, reception at 8420 MHz

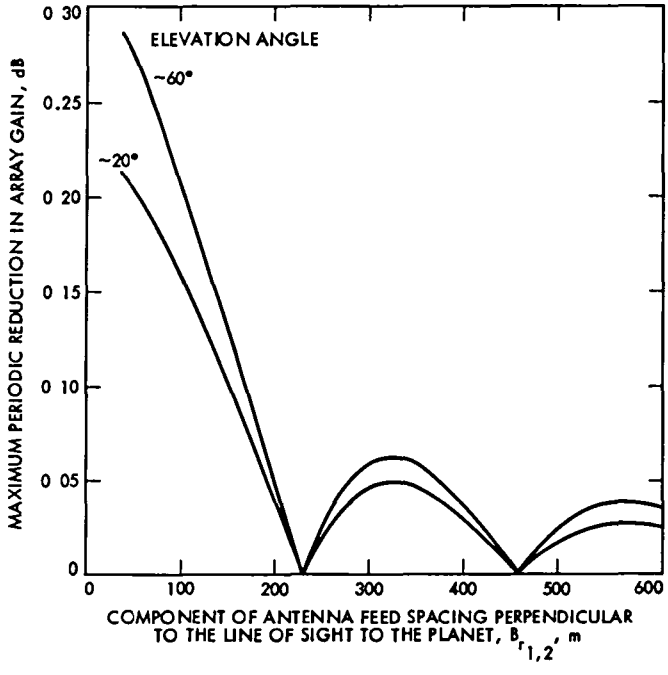


Fig 24. Maximum periodic reduction in array gain due to Jupiter (at maximum distance) within beamwidth of two receiving systems arrayed for coherent reception vs effective feed spacing; 64- and 34-m-diameter (listen-only) antennas, reception at 8420 MHz

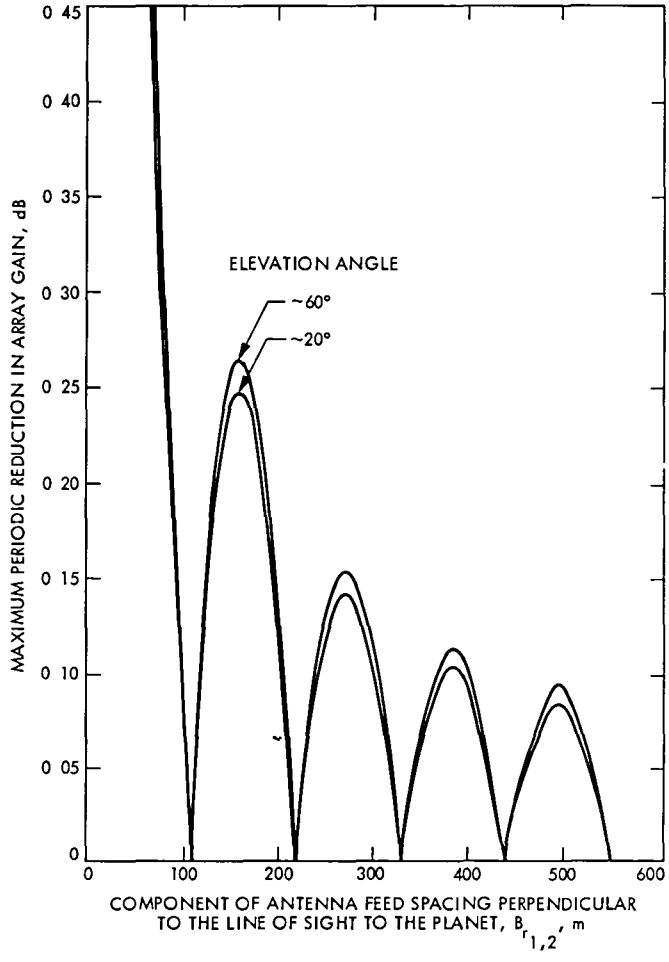


Fig. 25. Maximum periodic reduction in array gain due to Venus (at minimum distance) within beamwidth of two receiving systems arrayed for coherent reception vs effective feed spacing, 64- and 34-m-diameter (listen-only) antennas, reception at 8420 MHz

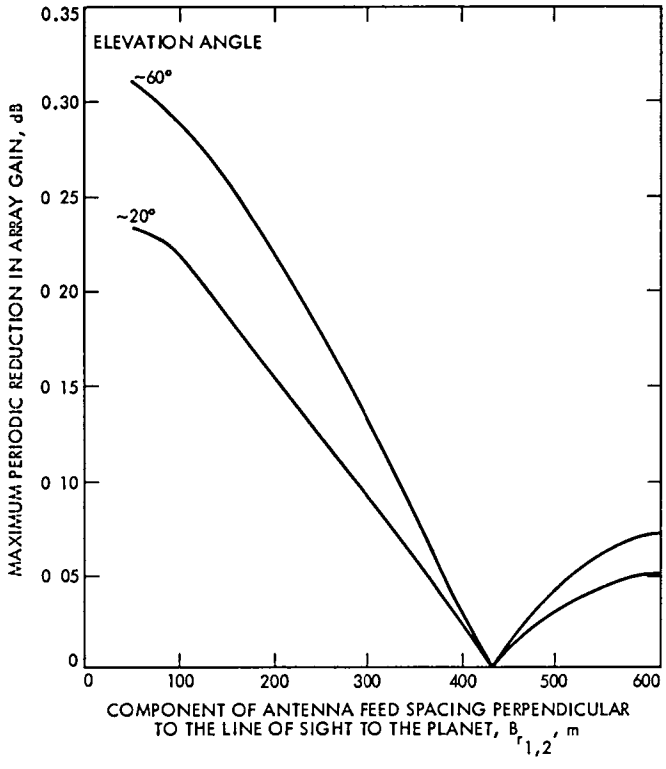


Fig 26 Maximum periodic reduction in array gain due to Venus (at maximum distance) within beamwidth of two receiving systems arrayed for coherent reception vs effective feed spacing; 64-and 34-m-diameter (listen-only) antennas, reception at 8420 MHz

A Two-Year History of Atomic Frequency Standards Syntonization in the Deep Space Network

S C Ward
Operations Sustaining Engineering Section

The NASA-JPL Frequency and Timing System (FTS) of the Deep Space Network (DSN) consists of a collection of three sets of clocks (located in Australia, California, U S A , and Spain) driven by independent atomic oscillators It is the synchronization of the output frequencies (syntonization) of these oscillators (reference frequency standards) that is reported here Prior to 1980 there was no DSN direct specification of syntonization. However, there was an implied specification of a ±5 5 X 10^{12} related to the DSN time synchronization specification of a ±100 microseconds Both the syntonization within the three sets and the syntonization of the sets to the international standard [International Atomic Time (TAI, UTC (BIH)] are considered

I. Establishing Syntonization

The original calibration was established during the third quarter of calendar 1980 by syntonizing a specially prepared transportable cesium atomic clock (Ref 1) to the standard frequency as maintained by the National Bureau of Standards at Boulder, Colorado This "calibrated transfer standard" was then transported to each of the three sets of atomic frequency standards, where it was used to measure their frequency offset with respect to the NBS standard (designated as UTC-NBS) In this manner each of the reference frequency standards were syntonized to UTC-NBS and, to a lesser degree, syntonized to each other

Each atomic reference frequency standard set consists of a Smithsonian Institution Astrophysical Observatory (SAO) model VLG-10B hydrogen maser (H2M) and two Hewlett-Packard model 5061-A option 004 cesium beam frequency standards The offset of each H2M's output frequency from the hydrogen hyperfine spectral line was established with respect to the transported transfer standard By means of phase comparators, the H2M calibration is transferred to the two cesium beam standards Thus, the syntonization between the three sets and between any set and UTC was established (Ref 2) within less than ±3 X 10^{-13}

II. Maintaining Syntonization

To make effective use of syntonization it must be maintained on a continual basis, e g , the tuning of an orchestra to A above middle C is a syntonization maintenance application The DSN makes use of four resource techniques for maintaining syntonization They are LORAN-C, traveling clocks, television vertical sync pulse, Very Long Baseline Interferometry (VLBI/ΔDOR[1]) Figure 1 illustrates how the DSN is con-

[1] A technique equivalent in accuracy to VLBI that makes use of an extragalactic radio source along with a spacecraft to gather data from which time sync information can be derived

figured to utilize these resources, making use of a combination of (1) Mediterranean chain LORAN, traveling clocks and VLBI/ΔDOR in Spain, (2) television, microwave links, traveling clocks and VLBI/ΔDOR in Australia, (3) microwave links, West Coast chain LORAN, traveling clocks[2] and VLBI/ΔDOR in California

Table 1 lists the comparative merits of the resources used for maintaining syntonization For all comparisons, it has been assumed that the resource has been in use for gathering data for periods longer than one year and that at least two out of three reference frequency standards in a set have been in continual operation over that period

III. Syntonization to UTC

Figure 2, a plot of $\Delta f/F$ vs time, presents a two-year history of the syntonization of the DSS 63[3] reference frequency standards with respect to UTC-USNO The "SPEC wrt UTC (NBS)" annotated dotted lines in the vicinity of UTC (USNO) plus and minus 10E-12 refers to System Functional Requirements Document 821-14 III B 2 The data were collected through the use of LORAN-C and have been adjusted through the use of USNO Bulletin 4 corrections With the exception of routine frequency standard maintenance (H2M tuning) periods, 98% of the data presented applies to H2M, SAO VLG-10B serial #07 The data indicate that the reference frequency standards were syntonized to UTC-USNO within less than 1 fractional part in 10^{12} over the entire two-year period.

Figure 3, a plot of $\Delta f/F$ vs time, presents a two-year history of the syntonization of the DSS 43[4] reference frequency standards with respect to UTC-AUS[5] The "SPEC wrt UTC (NBS)" annotated dotted lines in the vicinity of UTC(USNO) plus and minus 10E-12 refers to System Functional Requirements Document 821-14 III B 2[4] The data were collected through the daily use of the TV simultaneous view technique. The reference to which syntonization is made is the Department of National Mapping (NATMAP) clock ensemble (four clocks) in the local area custodian of UTC-AUS It is equivalent to and serves the same function for DSS 43 as the Deep Space Network Master Clock ensemble (NMC) does for DSS 14 at Goldstone, California, U S A The data have been adjusted per published NATMAP Bulletin E corrections With the exception of the routine frequency standard maintenance (H2M tuning) periods, 83% of the data applies to H2M, SAO VLG-10B serial #06 The data indicate that the reference frequency standards were syntonized to UTC-USNO within a part in 10^{12} over the entire two-year period

Figure 4 presents a two-year history of the syntonization of the DDS 14 reference frequency standards with respect to UTC-NMC[6] The data were adjusted per weekly NMC clock closure reports Over this two-year report period there were occasions when UTC-MC and/or UTC-DSN were not available[7] Continuity of the syntonization history of H2M SAO VLG-10B serial #05 was maintained at those times through the use of output frequency drift data from the hydrogen hyperfine line obtained during tuning maintenance Greater than 98% of the data presented applies to H2M SAO VLG-10B serial #05, and it indicates that the syntonization of the DSS 14 reference frequency standards was maintained within a part in 10^{12} of UTC(NBS) over the entire period Performance was borderline during April and May 1982

IV. Syntonization Within the DSN

Figure 5 presents a plot of $\Delta f/F$ vs time for a two-year history of the syntonization of the DSS 14 reference frequency standard with respect to the DSS 63 reference frequency standard The data were collected through the use of VLBI and ΔDOR time synchronization operations (Ref 3) These syntonization data were derived from sets of blunder-free time offset data points, where a set consists of a minimum of three "good" points[8] Further, where there was a lack of "good" VLBI data, ΔDOR data sets were used The data indicate that the DSS 63 reference frequency standards remained syntonized to the DSN master at DSS 14 within $\pm 3 \times 10^{-13}$ over the two-year report period

Figure 6 presents a plot of $\Delta f/F$ vs time for a two-year history of the syntonization of the DSS 14 reference frequency

---

[2]At Goldstone, California, traveling clocks serve a double role (1) to synchronize the Network Master Clock (NMC) ensemble of seven cesium clocks to UTC-NBS, and (2) to synchronize the DSS 14 clocks to the NMC ensemble on each VLBI time sync operation.

[3]Deep Space Tracking Station 63 (DSS 63) is located in the Madrid area of Spain (Fig 1)

[4]DSS 43 is located in the Australian Capitol Territory (A C T) near Canberra

[5]UTC-AUS is traceable to UTC-USNO within $\pm 5 \times 10^{-13}$ through the use of semiannual traveling clock trips between Canberra, A.C T , and Washington, D C

[6]UTC-NMC is syntonized to UTC-NBS, and to lesser degree to UTC-USNO, within a part in 10^{13} through the use of traveling clocks and West Coast LORAN-C UTC-NMC is syntonized to UTC-DSN within less than a part in 10^{13} through the use of weekly traveling clock trips and microwave links Thus UTC-DSN is syntonized to UTC-NBS within a part in 10^{12}

[7]Extended power outages and airconditioning failures in the building housing the NMC ensemble and extensive repair work on the DSS14 antenna caused the loss of UTC-DSN and UTC-NMC.

[8]A point is defined as "good" if it is within less than 3σ of the average rate computation ($\Delta f/f$), where the number od data points used is greater than 2

standard with respect to the DSS 43 reference frequency standard The data were collected through the use of VLBI and ΔDOR time synchronization operations (Refs 3, 4) These syntonization data were derived from sets of blunder-free time offset data points, where a set consists of a minimum of three "good" points Further, where there was a lack of "good" VLBI data, ΔDOR data sets were used The data indicate that the DSS 43 reference frequency standards remained syntonized to the DSN master of DSS 14 within $\pm 3 \times 10^{-13}$ over the two-year report period

References

1 Ward, S , "The Operational Performance of Hydrogen Masers in the DSN," *TDA Progress Report 42-63*, pp 203-218, Jet Propulsion Laboratory, Pasadena, Calif , June 15, 1981

2 Ward, S , "The Operational Performance of Hydrogen Masers in the DSN," Proceedings of the Twelfth Annual Precision Time and Time Interval (PTTI) Applications and Planning Meeting, pp 616-617, December 2-4, 1980, Goddard Spaceflight Center, Greenbelt, Md

3 Roth, M G , "International Clock Synchronization with the Block I VLBI System," Proceedings of the Thirteenth Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, pp 489-500, December 1-3, 1981, Goddard Spaceflight Center, Greenbelt, Md

Table 1 Comparative accuracy of syntonization maintenance resources

	Osc to osc (DSS to DSS)[a]	Osc to listed Ref[a]	Osc to UTC-USNO[a]	Osc to UTC-NBS[a]
LORAN-C (daily meas)	1 1E -12	±5E -13 LORAN	±7E -13	±7E -13
LORAN-C (simult view)	1 0E -12	±3 5E -13 LORAN	±5E -13	±6E -13
Traveling clocks (2 trips/yr)	1 4E -13[b]	±7 5E -14 Clock[b]	DNA	DNA
Traveling clocks (4 trips NBS/yr)	DNA	±7 8E -14 UTC-NMC	±2 5E -13	±1 3E -14
Television (daily meas)	1 05E -12	±2 3E -13 UTC-AUS	±5 5E -13	±7 4E -13
VLBI X-band (7 to 10 day)	5 4E -15	±3 8E -15 UTC-DSN	DNA	DNA
VLBI S-band (7 to 10 day)	1 4E -14	±9 9E -15 UTC-DSN	DNA	DNA
Microwave (3 meas /day)	2 46E -12	±2 2E -15 UTC-NMC (1)	DNA	DNA

[a]The number following E designates the exponent or power of 10 e g , ±7 5E -14 is $\pm 7.5 \times 10^{-14}$
[b]These are frequency or phase measurements All others are time measurements

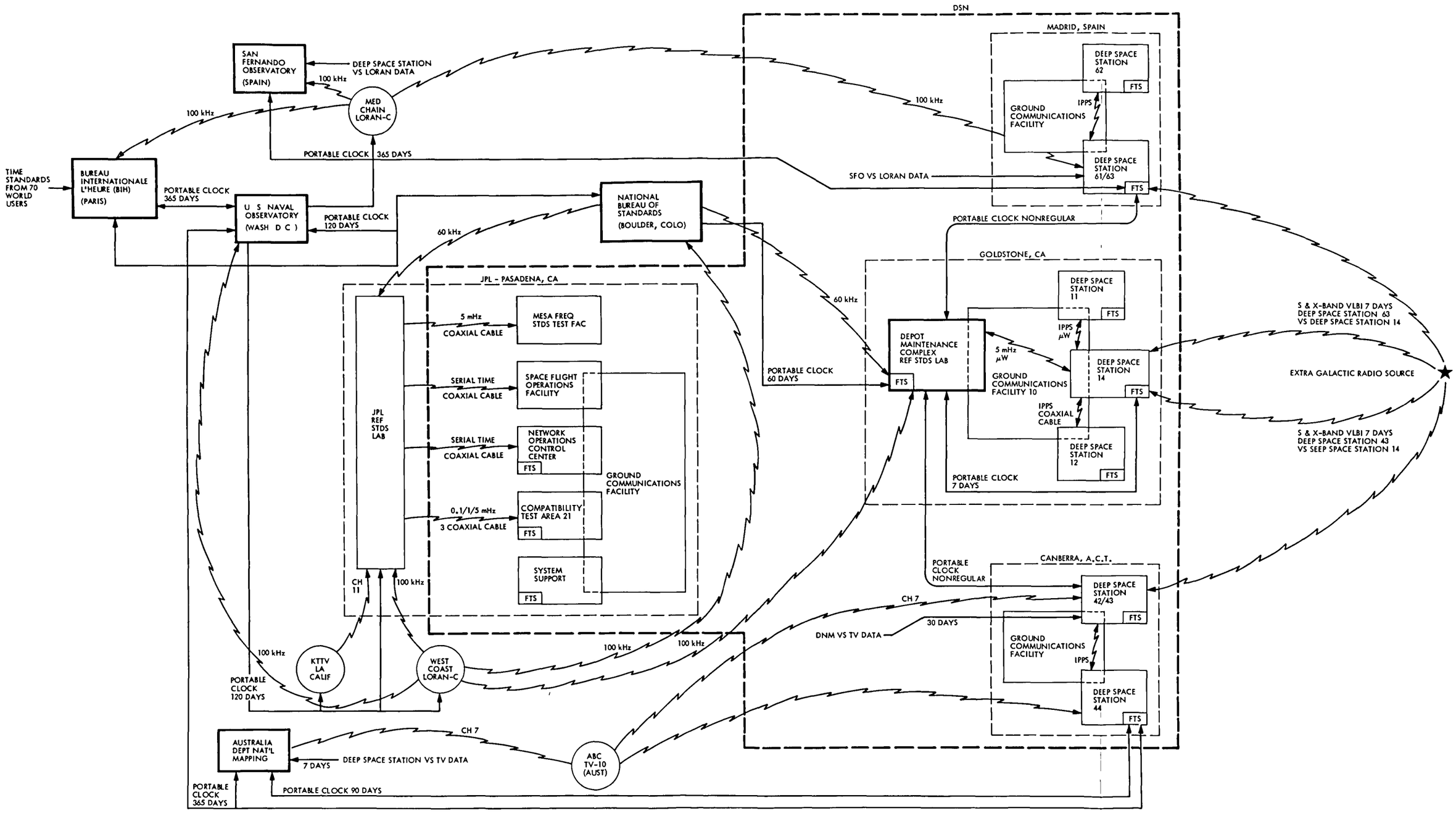


Fig 1 NASA/JPL intra/extra DSN frequency and time sync system

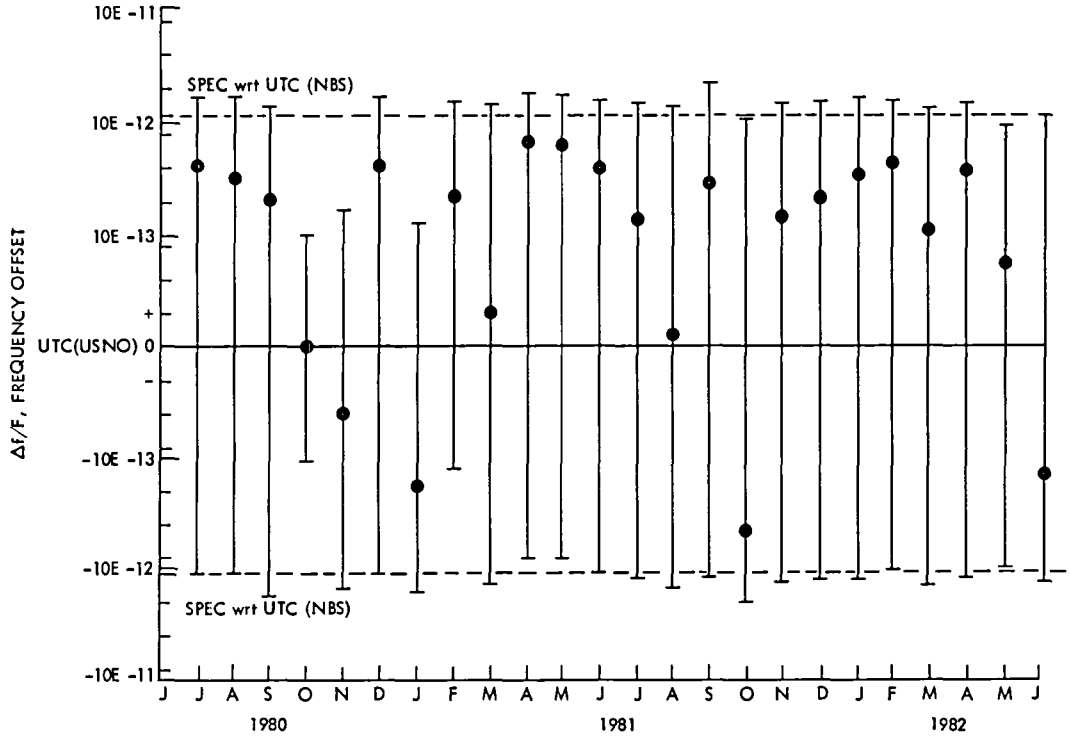


Fig 2. DSS 63 syntonization to UTC/(USNO) through the use of LORAN-C

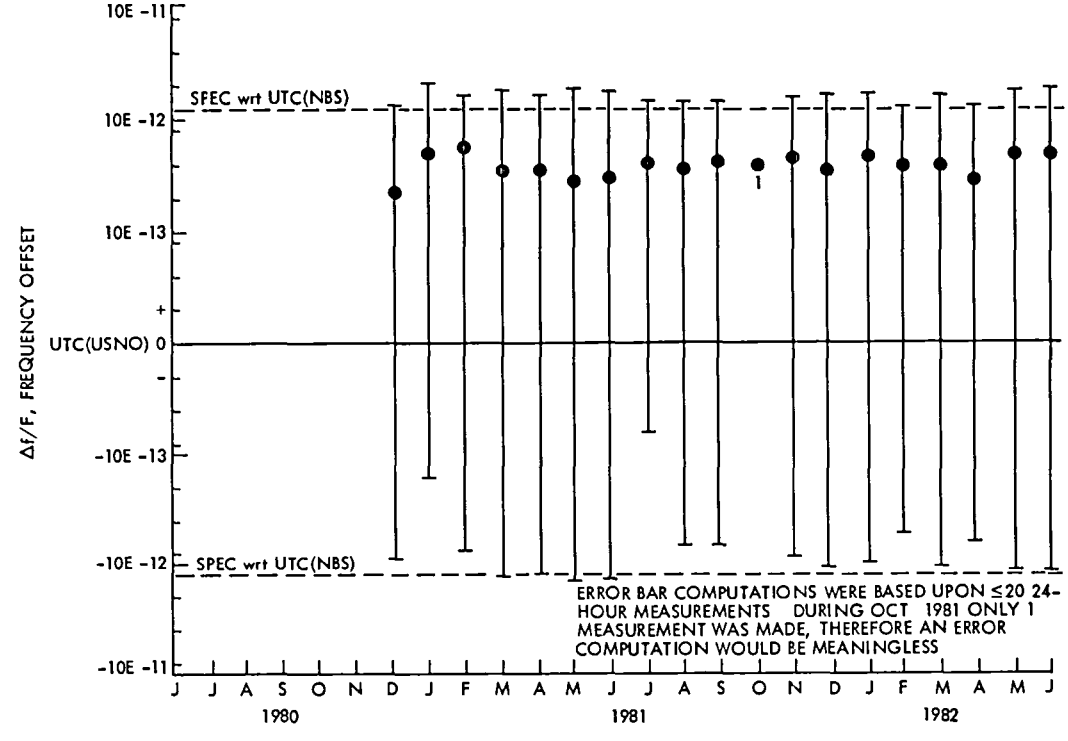


Fig 3. DSS 43 syntonization to UTC/(USNO) through the use of television

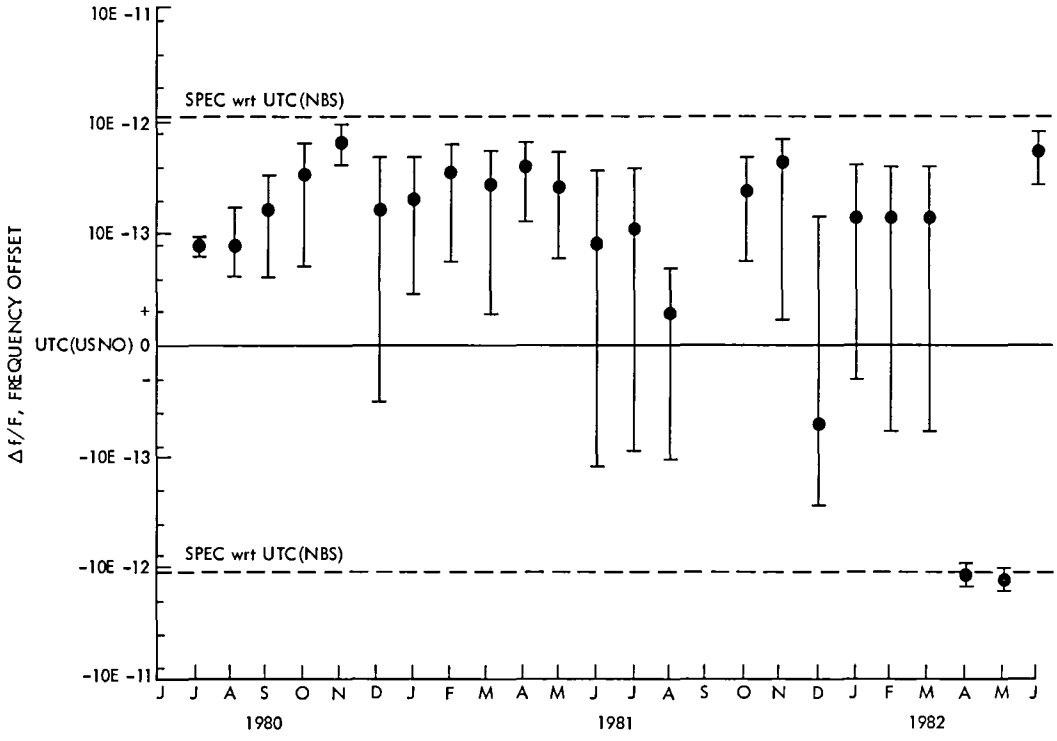


Fig. 4 DSS 14 syntonization to UTC/(USNO/NBS) through the use of microwave and traveling clocks

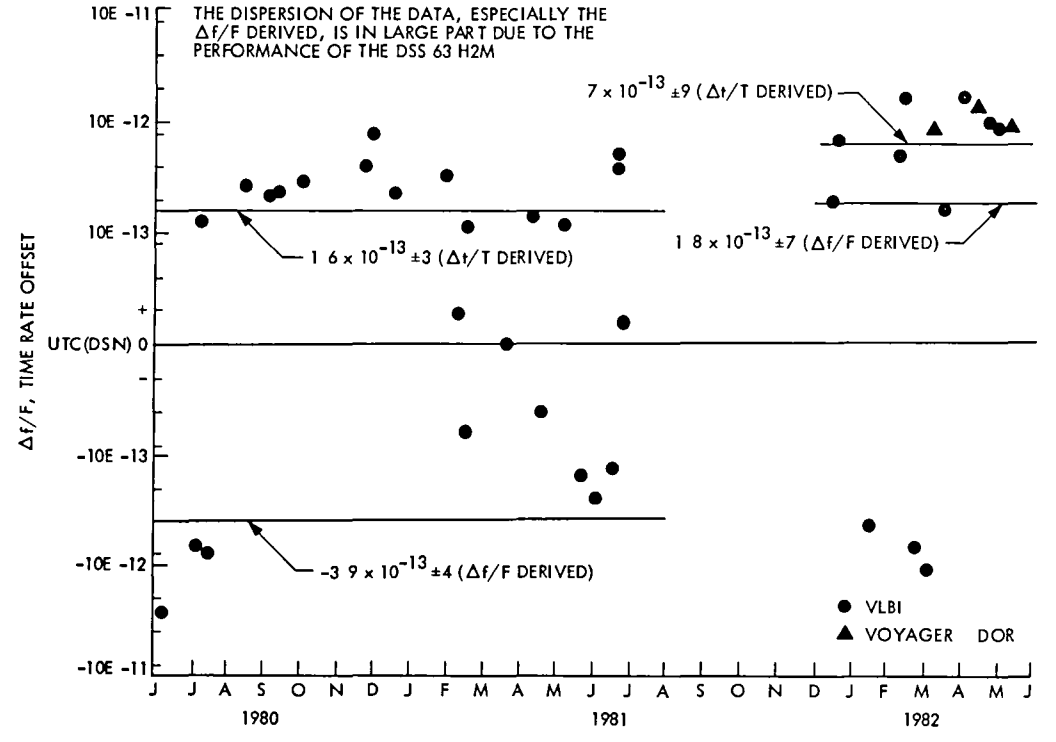


Fig 5 DSS 63 vs DSS 14 syntonization through the use of VLBI

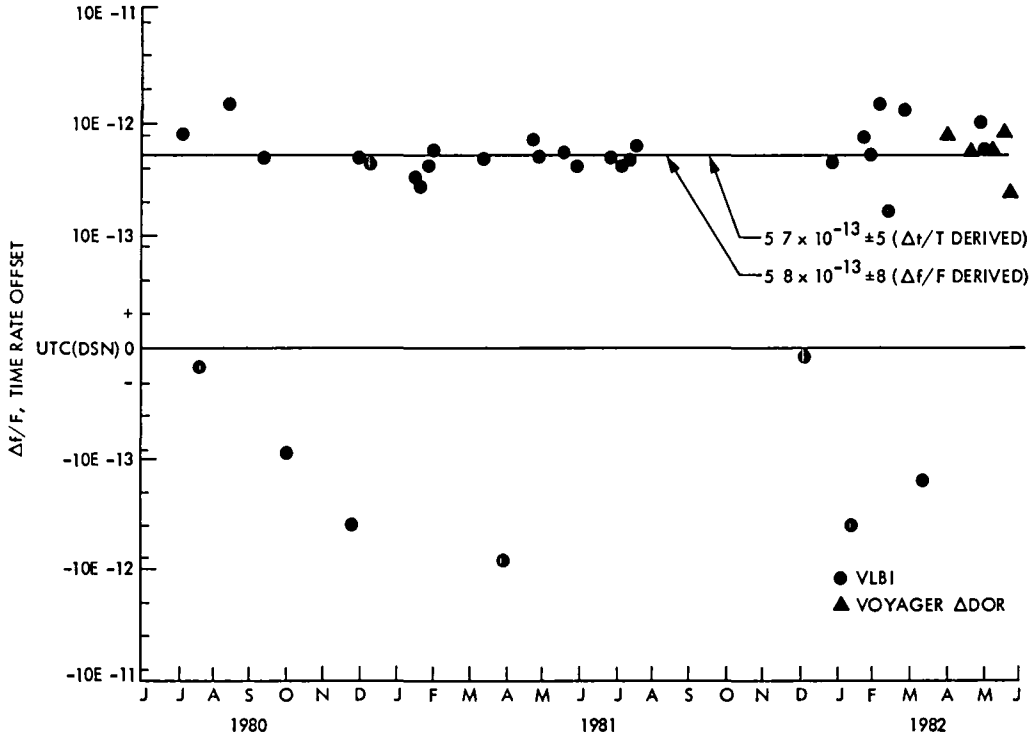


Fig. 6 DSS 43 vs DSS 14 syntonization through the use of VLBI

DSN Test Support System, Mark IV-85

B W Falin
TDA Engineering Office

As part of the Network Consolidation Program, the present DSN Test and Training System, Mark III-77, will be upgraded during the Mark IVA implementation planned for 1984-85 This new upgraded system will be referred to as the DSN Test Support System, Mark IV-85 This article provides descriptions and functional capabilities that exist or will be implemented in the subsystems which comprise the DSN Test Support System

I. System Definition

A. General

The DSN Test Support System, as part of the Network Consolidation Program (Ref 1), provides support capabilities for testing the DSN data systems at each of the three Deep Space Communications Complexes (DSCC) (Goldstone, California, Canberra, Australia, Madrid, Spain) and at the Network Operations Control Center (NOCC) The system also provides support to flight projects for testing and training activities with the DSN facilities The system includes capabilities for

(1) Facility-level system performance testing of each DSN system

(2) Maintenance and diagnostic testing for the DSCC subsystems

(3) End-to-end testing of each DSN system, including related DSCC, Ground Communications Facility (GCF), and NOCC functions

The system consists mainly of a number of hardware and software components that are normally parts of other DSN systems and subsystems However, it includes two subsystems that are dedicated to the test support DSCC Test Support Subsystem and NOCC Test Support Subsystem

Figure 1 describes the system's functions and interfaces This article updates the system description published in Ref 2

B. Key Characteristics

Design goal key characteristics of the DSN Test Support System are

(1) Centralization of test support capabilities at each DSCC

(2) Capability to support testing and maintenance activities concurrently with operational activities

(3) Capability to function without alteration of DSN operational configurations

(4) Accommodation of flight-project-supplied simulated data via GCF

(5) Capability to support network loading tests with a combination of actual and simulated data streams

(6) Testing functions to be controlled by Monitor and Control Subsystems

(7) Accommodation of other data sources, as follows

(a) Spacecraft test data via JPL Compatibility Test Area (CTA 21).

(b) Spacecraft prelaunch data via Merritt Island, Florida, Spacecraft Compatibility Monitor Station (STDN, MIL 71)

(8) DSCC Test Support Subsystem will use mission-independent hardware and software

C. System Usage

Since the system capabilities can be used to support various DSN activities, the following are provided

(1) Calibrations and prepass readiness verification

(2) Real-time diagnostics and fault isolation

(3) DSN implementation activities and performance testing of DSN systems, DSCC subsystems, and NOCC subsystems.

(4) DSN operational verification tests to prepare for mission support.

(5) Flight project ground data system tests, mission simulations, and operational readiness tests.

II. System Implementation

A. Description

The Mark IVA DSN plan establishes a Signal Processing Center (SPC) at each of the three Deep Space Communications Complexes. In each SPC, a new DSCC Test Support Subsystem will be implemented Included within this subsystem are two assemblies, Telemetry Simulation Assembly (TSA) and Maintenance Support Assembly (MSA), that provide major support capabilities for the DSN Test Support System Also, within the SPC, a new Monitor and Control Subsystem will be implemented that will provide control for operational activities The backup Monitor and Control Subsystem will be used to support new system performance test programs

In addition to the major implementation described in the above paragraph, additional changes will be implemented in other DSCC subsystems that support the DSN Test Support System The test support functions provided by all DSCC subsystem are described in Section III.

B. New Capabilities

This paragraph lists some of the new support features that are different from those provided by the Mark III-77 Test and Training System as described in Ref 2

(1) DSCC Test Support Subsystem In support of the telemetry simulation and conversions functions the DSCC Test Support Subsystem will

(a) Provide status to Mission Control and Computing Centers (MCCC) during remote simulation tests

(b) Provide capability to store and replay project-generated telemetry data

(c) Provide project tables that contain the data parameters for establishing unique spacecraft telemetry test signals

(d) Generate delayed-telemetry test signals

(e) Generate telemetry test signals to support the High Earth Orbiters mission that will be tracked by the Mark IVA DSN

(f) Provide test signal generation and distribution capability to support multiple telemetry equipment groups

(2) Maintenance support At the DSCCs, the Maintenance Support Assembly will provide maintenance support carts to interface with various subsystem controllers via maintenance interface connectors

(3) System performance test support With the implementation of a local area network as part of the DSCC Monitor and Control Subsystem, testing activities and operational activities can be supported concurrently. System performance test programs and procedures will be provided for each of the eight DSN systems, as part of the Mark IVA project

C. Schedule

The Mark IVA Network implementation plan is phased with the Goldstone facility, being completed in February 1985, followed by Canberra in May 1985 and Madrid in August 1985

III. Deep Space Communication Complex Functions

A. DSCC Test Support Subsystem (DTS)

As a major subsystem of the Test Support System, the following functions are provided

(1) Telemetry simulation and conversion functions are shown in Fig 2 The Telemetry Simulation Assembly (TSA) provides the functions that include the various digital and analog capabilities as itemized in Table 1

(2) Maintenance Support functions include the following·

(a) Accept equipment assignments for maintenance

(b) Maintain status of all equipment assigned for maintenance

(c) Provide portable maintenance support equipment

(d) Interface with subsystem-supplied diagnostics

B. Receiver-Exciter Subsystem (RCV)

The RCV provides the following test support functions

(1) Generation of simulated S-band and X-band downlink test carriers Specific capabilities are as follows

(a) 64-meter and 34-meter transmit-receive stations provide one X-band carrier and one S-band carrier.

(b) 34-meter receive-only stations provide a single X-band test carrier

(c) CTA-21 and MIL-71 provide the same capability as the 64-meter stations

(2) Modulation of telemetry subcarriers and data streams from the TDA onto simulated test carriers

(3) Variable attenuation of each test carrier signal level

(4) Translation of S-band exciter uplink frequencies to S- and X-band downlink frequencies

(5) Measurement of carrier power levels and modulation levels

C. Antenna Microwave Subsystem

The Antenna Microwave Subsystem provides the following test support functions

(1) Routing of simulated downlink carriers to masers and/or receivers

(2) Mixing of simulated S-band downlink carriers

D. Transmitter Subsystem

The Transmitter Subsystem includes provision for feeding the transmitter output into a dummy load to support Command System and Tracking System test operations

E. Frequency and Timing Subsystem

The Frequency and Timing Subsystem provides the following support functions

(1) Time code and reference frequencies

(2) Generation and distribution of a simulated time signal which can be substituted for the true GMT input to the various DSCC subsystems This capability is provided to support realistic mission simulations in flight project testing and training activities

F. DSCC Digital Communication Subsystem

The Digital Communications subsystem provides the following test support functions

(1) Transmission of data blocks of simulated telemetry data and DTS control messages from the NOCC or MOC

(2) Transmission of status and alarm messages from the DTS to NOCC and/or MOC

(3) Transmission of system performance test data blocks and simulated response data to the NOCC for DSN systems testing and project testing and training

(4) Provision of on-site loop-back of data blocks for testing

G. DSCC Monitor and Control Subsystem

The Monitor and Control Subsystem (DMC) provides the following test support functions

(1) Generating and transmitting control messages to the DTS and to all subsystem controllers during system performance testing and flight project training activities

(2) Displaying DTS alarm and status information for DSCC operations personnel

(3) Supporting system performance test programs for DSCC operations personnel

(4) Providing DSCC subsystem status data to DTS upon request

(5) Transferring control of DSCC subsystem controllers to DTS for maintenance support activities

H. DSCC Subsystem/Assembly Controllers

All DSCC subsystem controllers and/or assembly controllers provide the following functions in support of the DSCC maintenance support capability

(1) Diagnostics and test procedures to support the DSN maintenance program

(2) Standard maintenance interface connector

IV. Ground Communications Facility Functions

The DSN Test Support System utilizes the Ground Communications Facility Subsystems for communicating data and information between Network Operations Control Center (NOCC) or any Mission Operations Center (MOC) and the

Deep Space Communications Complexes, CTA-21, and MIL-71 The following capabilities are supported

(1) Transmission of control messages, simulated telemetry, and simulated command data to any DSCC from NOCC or from any MOC Simulated telemetry data rates will be less than 30 kbps

(2) On-site loop-back of test data for system performance testing

(3) Voice communications for purpose of test coordination and training

V. Network Operations Control Center Functions

Test support capabilities presently implemented in NOCC are as follows

(1) Off-line generation of recordings of high-speed data blocks for testing of the real-time monitors in the NOCC Tracking, Telemetry, Command, Radio Science, and Monitor and Control Subsystems

(2) Selection of stored data blocks and output of the data to the DSS for system readiness verification

References

1 Yeater, M L , and Herrman, D T , "Networks Consolidation Program," *TDA Progress Report 42-65*, pp 19-24, Jet Propulsion Laboratory, Pasadena, Calif , Oct 15, 1981

2 Herrman, D T , "DSN Test and Training System, Mark III-77," *DSN Progress Report 42-50*, pp 7-13, Jet Propulsion Laboratory, Pasadena, Calif , Feb 15, 1979

Table 1 DSCC Test Support Subsystem digital telemetry simulation capabilities

Functional capability	Description
Real-time data streams	6 independent channels
Variable data rates	6 SPS to 4 MSPS without subcarrier 6 SPS to 2 MSPS with subcarrier
Block-to-serial conversion	2 remote streams, combined rate 30 kbps
Block coding	1 channel, biorthogonal (32, 6)
Long constraint length convolutional codes	6 channels, K=32, r=1/2, 1/3 K=24, r=1/2, 1/3
Short constraint length convolutional codes	6 channels, K=7, r=1/2, 1/3
Selectable frame sizes	
Insert frame sync codes	Up to 48 bits
Increment data fields	Maximum of 3 fields per frame
Data patterns	Selectable per channel PN sequence, fixed formats, or project provided
Pulse coded modulation	Non-return-to-zero level (NRZL) or non-return-to-mark (NRZM) or bi-phase-L
Subcarrier frequency output	100 Hz to 2 MHz, 1/4 Hz resolution
Modulation index control	0 to 89 deg on each subcarrier
Subcarrier mixing	Maximum of two telemetry subcarriers
Delayed test signals	Five channels can be delayed from a reference channel, delays are selectable from 0 to 128 microsec

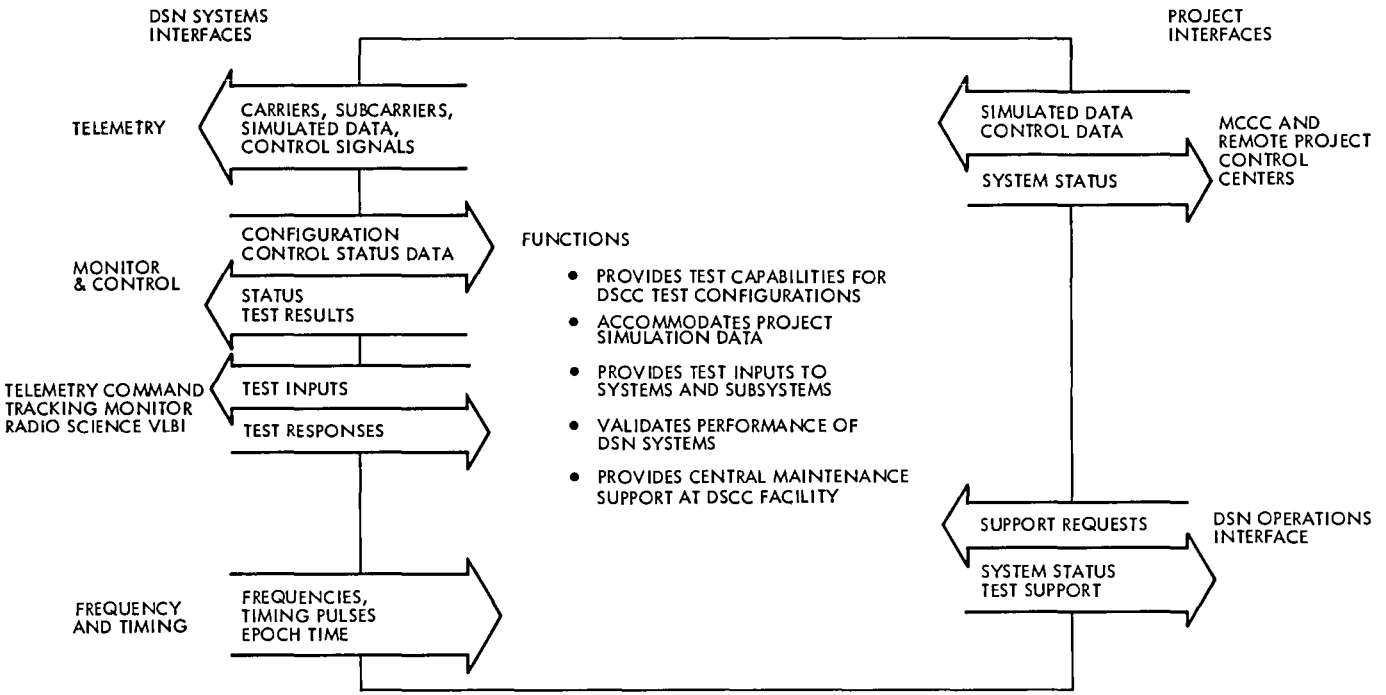


Fig 1. Test Support System, functiona and interfaces

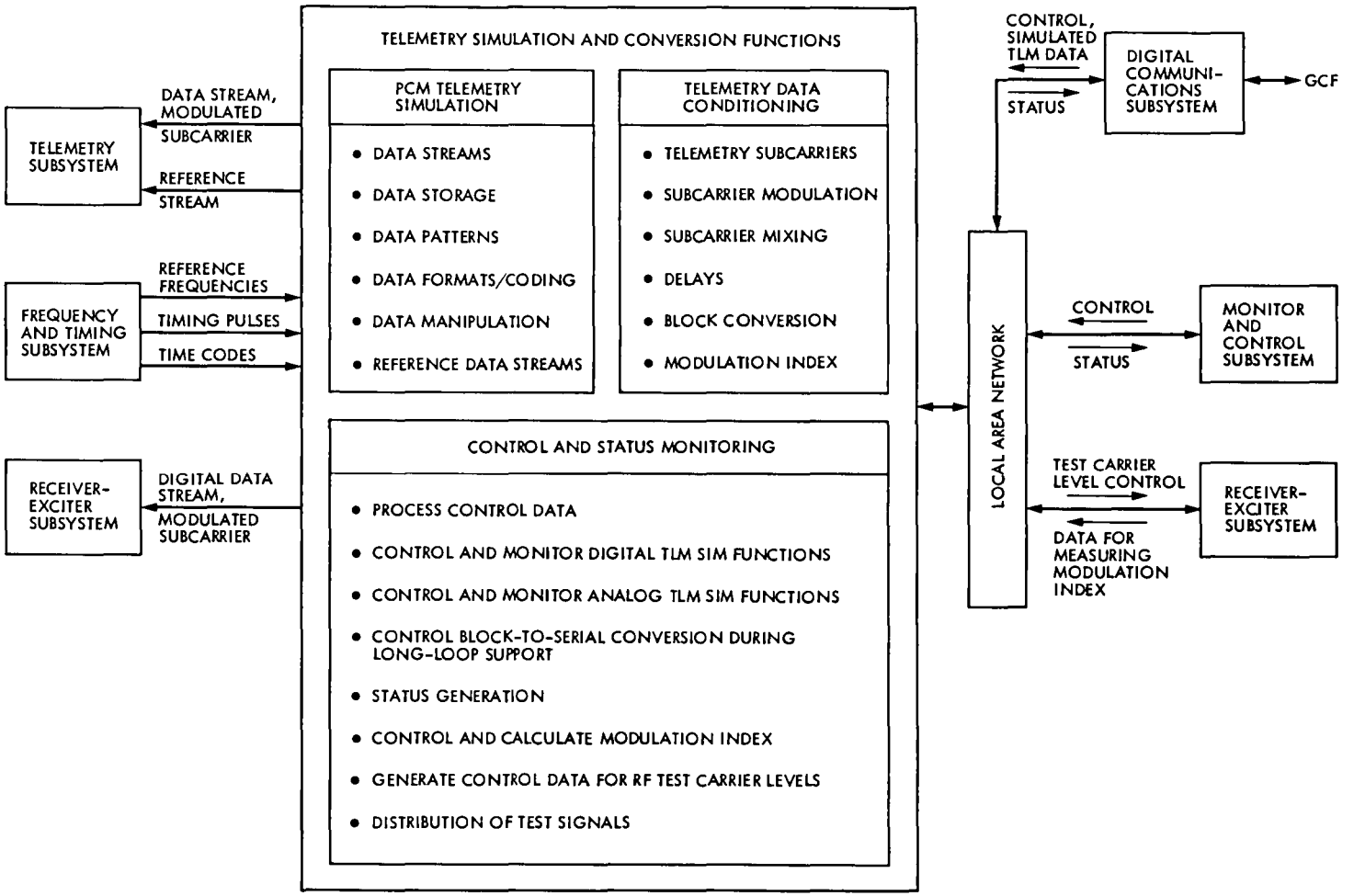


Fig 2 Telemetry functions, interfaces, and data flow

An Overview of the Goldstone Energy Systems Study

L S Rosenberg
Systems Engineering Section

The primary objective of the Goldstone Energy Systems Study (GESS) was to develop a systems planning methodology for analyzing and synthesizing Deep Space Network (DSN) energy systems The resultant product, the Deep Space Network Planning and Analysis Methodology (DESPAM), can support DSN energy planning evaluations It addresses a broad spectrum of tradeoff and dispatching scenarios It evaluates a variety of energy generation configurations and also includes a capability to evaluate conservation measures

I. Introduction

In recent years, energy costs have taken increasingly large portions of the Deep Space Network (DSN) annual operations budget There exists the possibility that energy supplied to the stations could be cut off or reduced during times of world crisis In response to these factors, several energy goals have been developed by NASA and the DSN. These include reducing energy consumption at all NASA installations by 50% from 1973 levels by 1985, minimizing life-cycle costs in balance with other costs through energy use reductions at the Goldstone Deep Space Communications Complex (DSCC), the attainment of a 90% non-utility, on-site energy generation capability, and an initiative to carry out demonstration projects in order to provide insight and data on existing low-risk technologies The Office of Telecommunications and Data Acquisition (TDA) implemented a number of studies which evaluated DSN energy consumption as well as future options for alternative energy use The Goldstone Energy Systems Study (GESS) was one of the TDA energy conservation projects

The primary objective of GESS was to develop a systems planning methodology for analyzing and synthesizing DSN energy systems The resultant product can support DSN short and long-term energy planning evaluations It addresses a broad spectrum of strategic tradeoff and tactical dispatching scenarios It evaluates a variety of energy generation configurations and also includes a capability to evaluate conservation measures By implementing the methodology, the user is able to address the following types of issues

(1) The definition of optimal energy dispatch strategies for current operations (optimal can be lowest cost or other user-defined attributes such as minimum grid reliance)

(2) The identification of conservation measures that would have the most favorable impact on operations and costs

(3) The determination of the value of new energy generation sources The identification of the best time for implementing new energy alternatives as well as the

dispatch strategy to be used and the quantity of the new source to use

(4) The calculation of the net cost benefits of one energy generation scenario vs another The calculation of the annual cash flows of varying scenarios

II. Approach

The tool for assessing the energy planning questions is the DSN Energy System Planning and Analysis Methodology (DESPAM). It consists of several key components The first is the characterization of new, alternative energy systems that have the potential to be implemented by the DSN in combination with existing generation sources such as grid and diesel This representation includes performance curves, cost equations, reliability factors, and the selection of a timeframe for implementation which will have the best chance for lowering system costs The next component relates to energy utilization by the DSN as well as physical conditions at the site Here, DSN energy loads are characterized, meteorological conditions are specified, and any desired conservation measure is defined The energy generation data and utilization parameters are then evaluated by implementing the DSN Energy System Simulation Model (DSNX) The next step of the DESPAM methodology is to use the DSNX model to evaluate actual test cases which replicate existing or potential energy generation scenarios of the DSN The last step is to select those strategies and options which the user deems best for actual implementation

The remainder of this article focuses on a description of the DSNX simulation model and on an end-to-end application of the DESPAM methodology. For a detailed discussion of the other components of the DESPAM methodology, see the Goldstone Energy Systems Study Final Report (an internal document)

III. DSNX Description

As stated earlier, the main purpose of DSNX is to serve within the overall DESPAM methodology as a tool for analyzing various operating policy scenarios and to perform multiyear generation mix analyses with respect to energy utilization for the DSN It does this by parametrically evaluating varying quantities of energy generation levels for different sources so that optimum configurations can be determined In addition, it is possible to evaluate the effects of different timeframes on each scenario evaluated by making multiple runs (the multiple run aspect of the code is covered in the next section) The model has the capability to address new, emerging technologies such as solar thermal, wind, photovoltaics, or solar ponds in conjunction with conventional utility grid or diesel-generated energy

When the model addresses these new, emerging technologies, it is usually necessary to know insolation and meteorological data pertinent to a specific geographical location This input can be provided to DSNX in the form of data tapes The hourly energy generation level of any new technology selected by the user is then estimated (A generalized flow chart of the entire model is shown in Fig 1)

The user then provides load data for the DSN facility in the form of data tapes, or he can create his own typical day load profile. It is also possible to adjust load data Additionally, sizing parameters for each energy generation source are required This includes any new energy technology to be implemented in addition to those sources currently available to the Goldstone DSCC, namely, utility grid and diesel (the diesel capability can also be used to model fuel cells) It is possible to address as many as four new energy technologies in addition to the utility grid and diesel Grid or diesel or any alternative can also be evaluated individually

The program then matches demand against energy availability for each source over a 1-year timeframe The user specifies the dispatch strategy Varying energy generation sizes and grid penetration levels can be evaluated within a single simulation run At the user's option, the utility grid maximum usage level can change in accordance with the billing period Only one diesel size per run is allowed

The program determines alternative source energy generation, excess alternative generation, grid generation, excess grid capacity, diesel generation, total energy generation, unmet demand, and total demand over the entire period simulated In addition, utility grid energy and power consumption levels are also determined on a monthly basis according to billing period All outputs are determined for all configurations evaluated

These outputs are then combined with user inputs of capital costs, maintenance and operations costs, fuel costs, subsystem replacement and/or rebuild, and utility billing charges, in order to determine one-time and recurring costs for each configuration under evaluation In addition, inputs of system life, escalation rates, cost of capital, tax parameters, and a time frame for new technologies enable DSNX to determine lifecycle cost and levelized busbar energy cost for each configuration The model also provides a complete yearly cash flow table Lastly, a summary table for all configurations is given and a data file is formed which can be accessed from a Hewlett Packard Graphics Terminal for use in creating output plots

IV. Implementation of DSN Energy System Planning and Analysis Methodology (DESPAM)

In the Approach Section it was explained that the overall methodology that covers all analysis components (characterization of generation systems, characterization of weather and load data, DSNX simulation, interpretation of results) is known as DESPAM. In this section, a typical end-to-end application of DESPAM is given

Let us suppose we want to evaluate a new technology (solar thermal) for use at the Mars site at the Goldstone DSCC We want to look at varying penetration levels from 100 kWe to 1500 kWe We want to integrate the new technology with the currently existing diesel-grid operation Our criteria for the "best" system will be lowest life cycle cost A secondary preference is to reduce reliance on grid-generated energy

The Goldstone Energy Systems Study Final Report provides the following necessary input information

(1) The solar thermal system will have a point focusing distributed receiver and a centrally located steam Rankine engine. This system will be assumed to need no major overhauls during a 24-year life and there will be no performance degradation over time

(2) The new system will be implemented in 1988 It will be assumed to meet DSN reliability standards at that time

(3) Insolation data for Barstow, California, for 1976 will be used

(4) The Southern California Edison demand tape for 1980 for the Mars site will be used to characterize the energy load During those periods when the diesels are in use (0 values on the tape), it will be assumed that 850 kWe of energy is used

(5) Diesel efficiency is 12 kWeh/gal There is no initial capital cost (it is assumed that the diesels are already in place) Diesel M&O cost is $ 003/kWeh Diesel fuel price is $ 915/gal Diesel subsystems are overhauled after every 20,000 hours of operation at a cost of $14,390

(6) The electric utility grid cost is based on actual 1982 bills ($0 069/kWeh) There will be no sell-back of any excess generation of solar thermal energy to the grid (this energy is lost)

(7) The nominal (including inflation) annual cost escalation rates are

Inflation rate[1]	10 0%
Capital cost	9.3
M&O	8 9
Discount rate[1]	10 0
Fuel	10 4
Purchased electricity	8 4

We now need to make assumptions regarding the dispatch strategy, baseline cases, and technology size

The dispatch strategy will be to use solar-generated energy first when available, the utility grid second, and diesel third In the baseline case (no solar, grid vs diesel only), five different diesel penetration levels will be tested

	Grid billing period		
Scenario	Off-peak	Mid-peak	On-peak
A	All grid	All grid	All grid
B	All grid	All grid	Grid up to 500 kWe, then diesel
C	All grid	Grid up to 1000 kWe, then diesel	All diesel
D	All grid	Grid up to 500 kWe, then diesel	All diesel
E	All grid	All diesel	All diesel

The current Goldstone DSCC operation of all grid, all grid, all diesel was not tested However, Scenario C is very similar These baseline scenarios will be simulated from 1982 to 2011 (Run 1)

The nominal solar thermal sizes tested are 100, 300, 1000, and 1500 kWe. These sizes will be tested in conjunction with the five grid-diesel combinations It will be assumed that the grid-diesel combinations are used without solar from 1982-1987, and that solar is added in 1988 (construction takes place in 1987). This second case consists of two simulation runs, 1982-1987 (Run 2), and 1988-2011 (Run 3) Each run evaluates performance for a single year separately.

[1]When making multiple DSNX runs with differing timeframes, which this case requires, the economic calculations are simplified when inflation equals discount rate

V. Results

The lowest cost combination in Run 1 (diesel vs grid for 30 years) was Scenario C Its life cycle cost in 1982 dollars is $11 64 million Its annual cash outflow is shown graphically in Fig 2 This scenario derives 82% of its energy from the grid and 18% from diesel.

In Run 3 the lowest configuration is 300 kWe of solar combined with Scenario C for diesel and grid Its life cycle cost is $8 69 million as compared to $8 98 million for the no-solar case over the same timeframe (1988-2011) [2] This best case only requires 72% of the energy to be generated by grid (16% comes from diesel and 12% is derived from solar thermal) At large solar penetration levels (1000 and 1500 kWe), reliance on the grid decreases but the monetary savings are lost now since large portions of solar-generated energy were wasted (It appears that a solar size of about 500 kWe would give the lowest cost) Figure 3 shows the results of Run 3 in terms of energy grid displacement as a function of life cycle cost The cash flow for the lowest cost configuration of Run 3 (300 kWe of solar) is shown in Fig 4 When the cash flow for the best scenario in Run 2 (Scenario C) is combined with the one from Run 3 and is compared to the baseline (Run 1), we get the net savings shown in Fig 5 From 1982 to 1987 the net change is 0 since for those six years the baseline is replicated The solar installation results in a negative net flow of about $130,000 annually from 1988 to 1992, but turns positive from 1993 to 2011 and results in an overall net life cycle cost savings of about $300,000. The energy generation breakdown for our optimal case from 1982 to 2011 is given in Fig 6 We now know that it might be a wise idea to investigate solar thermal systems more fully. Furthermore, if it is possible to acquire a solar system as described, and if the remaining assumptions hold, the DSN may be able to cut grid reliance and reduce net overall operating costs at the same time

[2]The Run 3 costs would have been somewhat lower had we assumed that excess solar generation was sold back to grid

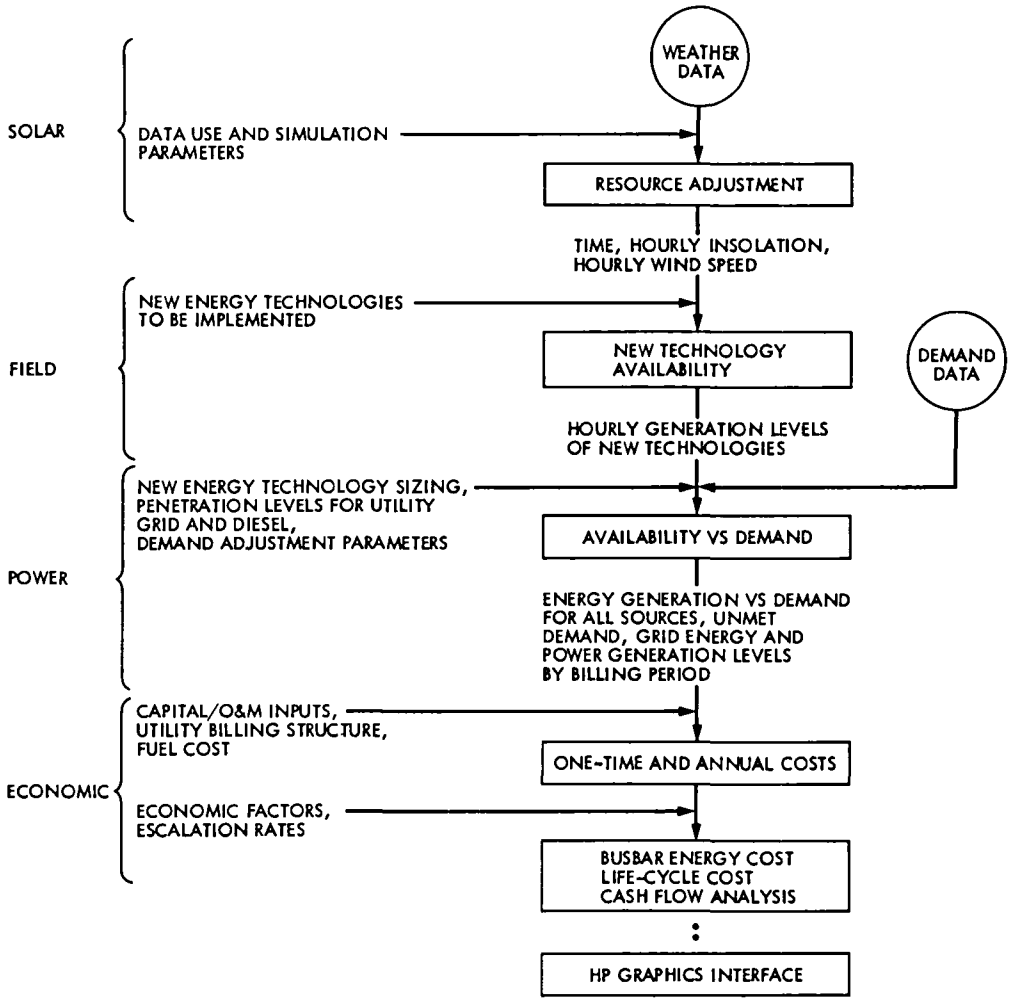


Fig 1 DSNX simulation sequence

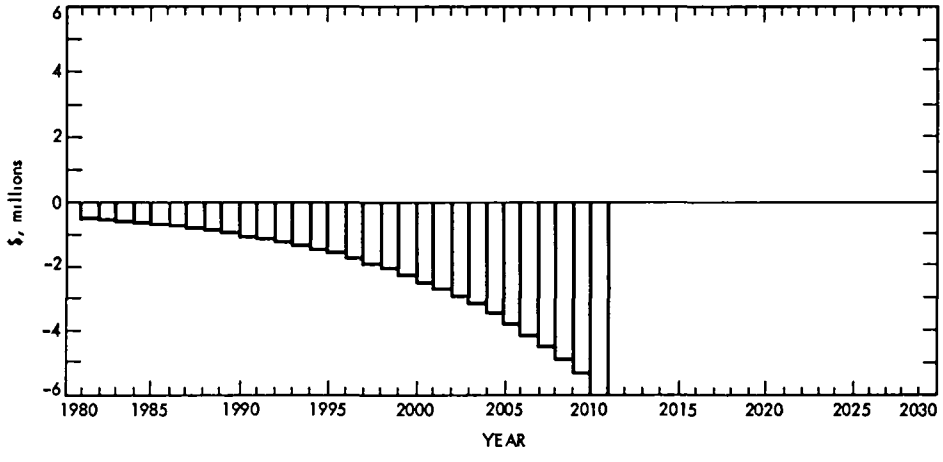


Fig 2 Net cash flow; diesel vs grid–Run 1, Scenario C

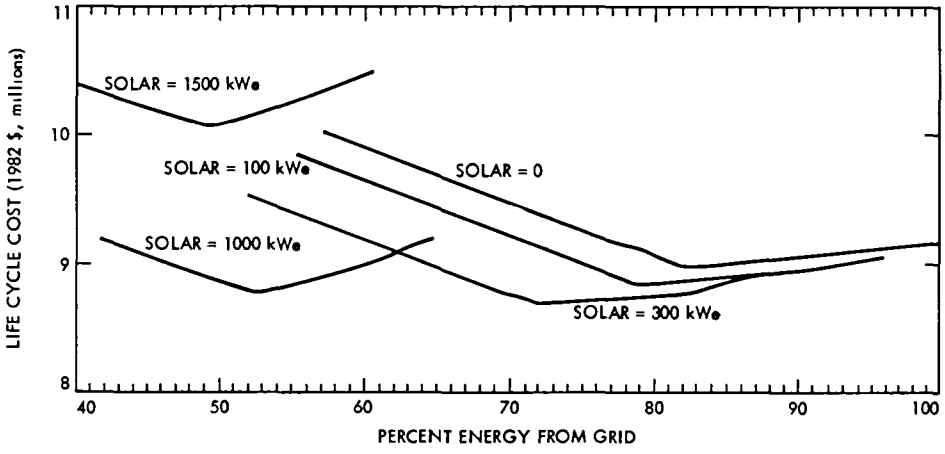


Fig 3 Grid displacement vs life cycle cost for various solar thermal/grid/diesel scenarios (Run 3)

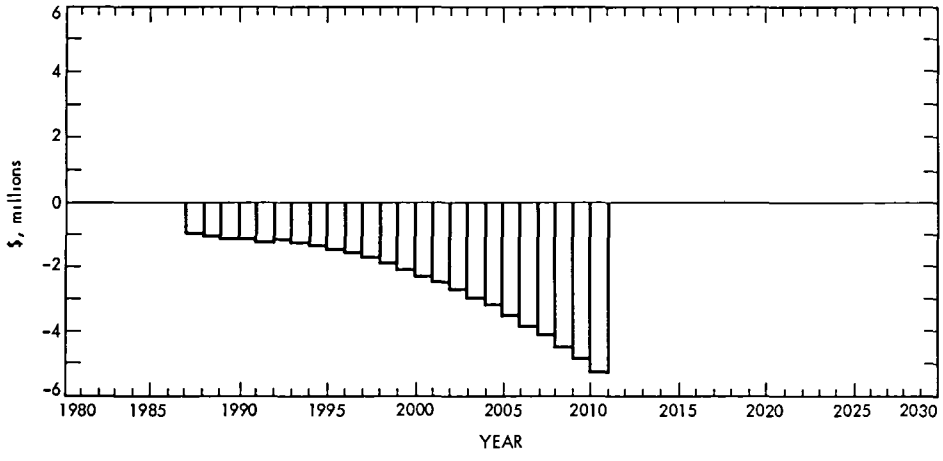


Fig 4 Net cash flow, Run 3–lowest cost configuration (solar power level = 300 kWe)

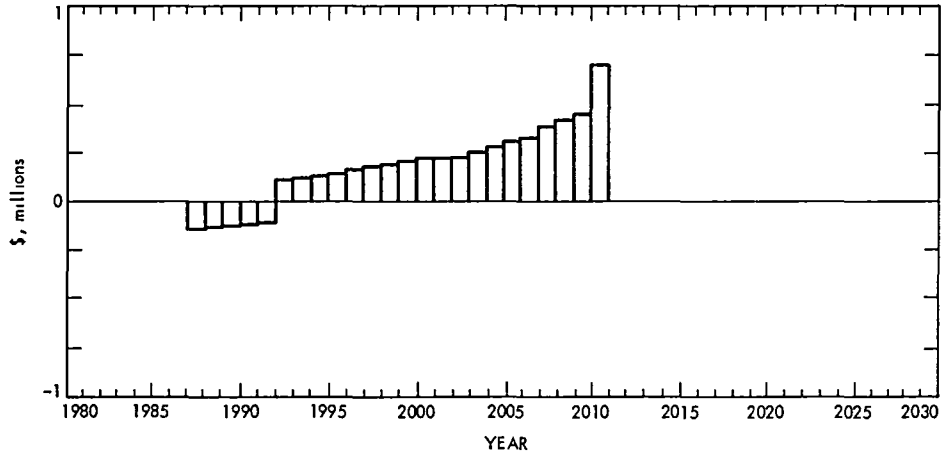


Fig 5 Net cash flow; solar strategy vs nonsolar

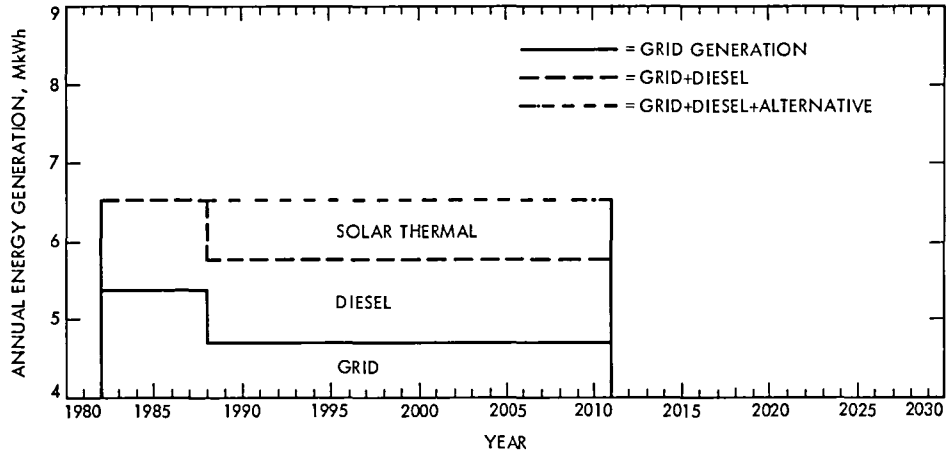


Fig. 6 Energy generation by source (addition of solar to grid and diesel in 1988)

Performance Simulation of the JPL Solar-Powered Distiller Part I. Quasi-Steady-State Conditions

C S Yung and F L Lansing
DSN Engineering

A 37 85-m^3 (10,000 gallons) per year (nominal) passive solar-powered water distillation system has been installed and is operational in the Venus Deep Space Station at Goldstone, California. The system replaced an old, electrically powered water distiller The distilled water produced with its high electrical resistivity is used to cool the sensitive microwave equipment A detailed thermal model has been developed to simulate the performance of the distiller and study its sensitivity under varying environment and load conditions The quasi-steady-state portion of the model is presented together with the formulas for heat and mass transfer coefficients used Initial results indicated that a daily water evaporation efficiency of 30% can be achieved A comparison made between a full day performance simulation and the actual field measurements gave good agreement between theory and experiment, which verified the model.

I. Introduction

A 37 85-m^3 nominal (10,000 gallons) per year passive solar-powered distillation (solar still) system has been operating successfully in the Venus station (DSS 13) of the Deep Space Network (DSN) tracking complex at Goldstone, California, since the initial installation in December 1981 Although this solar still system was intended to be an experimental system, only a few problems have been encountered Those problems were minor and were corrected without impairing the operation of the system Continuous observation of the installed system has indicated that it is highly productive, requiring only minimum maintenance The potential application of similar solar-powered distillers in other DSN stations is being examined to replace present electrically powered distillation systems

In order to complement the experimental investigation, a detailed analytical model of the solar still module is developed This report presents the quasi-steady-state portion of the model to simulate the module performance in a simple, yet accurate manner A computer program was written using the model characteristic equations to speed up the parameterization study

A literature search was made earlier (Refs 1-4) to gather available information on existing simulation tools for solar-distillation systems Dunkle (Ref 1) appeared to be a pioneer researcher on solar-stills The steady-state solution was tried by Dunkle (Ref 1) and Morse and Read (Ref 2) Sayigh (Ref 3) presented a gross steady-state model of solar still, neglecting the internal heat exchanges and the influence of

makeup water Cooper (Ref. 4) presented a transient model where heat flow to the ground beneath the still was strongly emphasized A finite-difference method was used to determine the temperature distribution below the ground surface The multiple reflections between the glass and the water were not considered in the model

In order to provide an accurate simulation of the JPL solar distiller, the model is developed based on the JPL design configuration (Ref 5) and includes all the influencing factors such as multiple reflections effect, internal heat exchange, and makeup water

II. Design Details

Figure 1 illustrates the design details (Ref 5) of the JPL solar distillation module The system consists of 8 modules The still module consists of a molded fiberglass tray with glass glazing on top sealed with special silicone rubber (RTV) The bottom of the still module is insulated with rigid foam and reinforced with marine plywood The base of the wetted tray is painted black to serve as the solar absorber for evaporation enhancement The remaining interior surface of the tray is colored white to promote reflection of sunlight into the water layer The solar still is mounted on a welded steel frame elevated 3 35 m (11 ft) above ground level to facilitate the gravity feed and to eliminate shading caused by nearby structures. The solar still is oriented due south for maximum solar exposure

Local well water with a high mineral content is supplied to the solar still by gravity from a storage tank located on a hill 91 m (300 ft) above the solar still inlet. The distilled water is collected in a 322-liter (85 gallon) temporary holding tank and then released to a large 5 7-m^3 (1500-gallon) underground storage tank for use in the microwave klystron-tube cooling system

III. Model Assumptions

The following assumptions are made for a one-dimensional thermal model

(1) The thermal losses from all four sides of the still module are neglected due to the presence of side air gaps The side surfaces are painted a light color which enhances the reflection of the incoming solar radiation toward the water body The sides temperature is assumed to be identical to the internal air temperature, with negligible heat loss to the ambient air

(2) The distillation mechanism is viewed as follows natural convection takes place while the water vapor is diffused uniformly from the water surface at T_w through the air space and condenses at the glass temperature T_g The bottom surface of the still is painted black and absorbs most solar energy received

(3) Due to the small inclination angle of the glass cover, the optical effect of the system is modeled as two parallel transparent media with an opaque bottom surface for the bottom medium

(4) The internal air space is assumed to be 100% transparent to solar energy No air leakage is assumed to take place to or from the solar still

IV. Quasi-Steady-State Model

Using the above assumptions, a quasi-steady-state model has been developed, the details of the heat balance equations are presented in Appendix A Appendix B gives the details of the heat transfer coefficients used in the model Appendix C gives the mass transfer and diffusion coefficients Appendix D illustrates the model use by a numerical example Since the heat capacitance of the solar still components is not included in a steady-state solution, only the glass cover, the water layer, and the bottom layer of the fiberglass tray are included From Eqs (A 8), (A 15), and (A 17), the temperatures of these three components are expressed as

$$T_w = [(B_1 + B_2 + m_w C_w) T_g - (E_1 + B_1 T_0)] / (B_2 + m_w C_w)$$

$$T_f = (E_3 + h_{cfw} T_w)/B_4 \qquad (1)$$

$$T_g = (X_0 + X_1 T_0 + X_2 T_m)/(X_1 + X_2)$$

where the B's are equivalent heat transfer coefficients and the X's are characteristic constants that depend on the evaporation rate, thermal conductance, and the solar absorptivity of various components Detailed descriptions of the B's and X's are given in Appendix A

The nonlinearities in the radiation energy and water diffusion terms make the equations difficult to solve explicitly An iteration method was used to obtain the final solution which uses a set of initially given temperatures T_g, T_w and T_f, for the glass cover, the water layer, and the fiberglass bottom, respectively The temperature-dependent terms such as the Grashof number Gr, saturated vapor pressure P_w, latent heat of evaporation of water L, heat convection coefficients for air and water, and the linearized radiation heat transfer coefficients can be calculated The new set of temperatures can be determined using Eq (1) By recomputing the temperature-dependent parameters using the new temperatures, a more

accurate set of temperatures can be determined, and the procedure is repeated until convergence is reached

V. Evaporation Rate

Water evaporates isothermally from the top surface of the water layer at temperature T_w and diffuses upward through the stagnant air layer to the glazing Water vapor condenses on the interior surface of the glazing at a lower temperature T_g This diffusion process is caused by the differential saturated water vapor pressures in the air layer, at the boundary of the water layer, and at the boundary of the glazing The evaporation rate of water, per unit still area m_w, is given in Appendix C as

$$m_w = \frac{DPM_w}{\bar{R}T_a\delta_a} \ln \frac{(P - P_{wg})}{(P - P_{ww})} \tag{2}$$

where D is the diffusion coefficient (m^2/hr), P is the system pressure (N/m^2), M_w is the molecular weight of water, $\bar{R}$ is the universal gas constant, δ_a is the average air spacing between the water surface and the glass cover, and P_{wg} and P_{ww} are the saturated water pressure at the glass temperature and the water temperature, respectively

If the leakage of air and water vapor is assumed negligible, the amount of water produced by evaporation-condensation would be completely replenished by the makeup water at an inlet temperature T_m Hence, Eq (2) represents the amount of water evaporation rate or the makeup water rate

VI. Distillation Efficiency

The instantaneous (or steady-state) distillation efficiency based on a constant solar radiancy on the projected area is defined by

$$\eta = \frac{m_w L}{I} \tag{3}$$

where L is the heat of vaporization of water In general, the evaporation efficiency increases with increasing the solar energy intensity where it reaches maximum efficiency at noon Although the instantaneous (or steady-state) efficiency gives an indication of the system performance under constant environmental conditions, it is not a complete characterization of the distiller performance over a full day or a year period. The other alternative to an accurate transient simulation is to have a sequence of 24-hour quasi-steady-state simulation for each day Accumulated daily efficiency can be obtained by dividing the accumulated distilled water production up to time t by the accumulated solar incident energy up to time t, i e ,

$$\eta(t) = \sum_0^t m_w(t) L \Big/ \sum_0^t I(t) \tag{4}$$

VII. Simulation Results

In order to facilitate the simulation process, a computer program was written to speed up the analysis of the system under many different operating conditions Appendix D gives a sample one-hour simulation When the sample solar still module operates at 299 85 K (~80°F) ambient temperature with 800 W/m^2 of solar intensity, it would yield 0 73 kg/hr (0 19 gal/hr) of distilled water The temperature of the glass cover, water layer, and the bottom surface would be 317 1 K (111°F), 343 8 K (159 2°F), and 347 2 K (165 30°F), respectively The steady-state or instantaneous efficiency would be 28 6%.

An experimental measurement program was conducted on the JPL solar still to determine its actual performance Under a solar intensity of about 800 W/m^2 and 32°C (89 6°F) ambient temperatures, the measured water temperature was approximately 71°C (159°F) With this actual field measurement in good agreement with quasi-steady-state simulations, we proceeded using the model to determine the effect of the major solar still performance parameters

In addition to an instantaneous performance simulation, a full-day simulation was performed for the month of June 1982 to predict the daily distilled water production rate In the full-day simulation, the hourly solar intensities are calculated by using the ASHRAE model (Ref 6) on an average day on June 21 at Goldstone, California (latitude is 35°N) The outdoor air temperature profile was measured and the wind speed is assumed to be constant for the day at 4 47 m/sec (10 miles/hr) Since in a full-day simulation, steady-state conditions are assumed to be reached at the end of each hour of operation, the summation of the hourly water production rate would give the accumulated daily water production The results of the full-day simulation of the JPL design indicated that the water production during 12 sunshine hours (from 6 00 a m to 6 00 p m) in June would yield 61 3 kg/day (16 2 gallons/day) of purified water vs the measured 64 kg/day (17 gallons/day), representing a 4 2% difference between the two values Part of this deviation could be due to applying the quasi-steady-state method instead of the true transient method Water production trails the solar intensity profile by a time-

lag Errors in estimating material optical and thermal properties are responsible for another part of the deviation The idealization and the assumptions listed in Section III contribute to the rest of the deviation

A graphical presentation of the daily performance of one solar-powered distillation module for the month of June is given in Fig 2. As shown in this graph, the water purification rate reached a peak value of 1 11 kg/hr (or 0 53 kg/(hr-m^2)) at noon, representing a maximum instantaneous efficiency of 35% Since the measured daily temperature profile was higher in the afternoon than in the morning, the module produced larger quantities of distilled water in the afternoon The accumulated efficiency increased asymptotically to 29 7% at 2 00 p m and then dropped slightly to give a daily efficiency of 29.1% as a typical daily efficiency in June

In addition to the above quasi-steady-state model, which provides accurate simulation, a transient model and a detailed sensitivity study are planned for the distiller. The results of the comparison between quasi-steady and transient simulations and parameterization will be presented in a future report

Acknowledgment

The authors are indebted to F W. Stoller, D. Kuma, R. Z Toukdarian, and F. Menninger for their technical management in the implementation of the JPL solar still at the Venus Deep Space Station. Acknowledgement is extended to R. Elder and J May for providing valuable experimental data on weather and water purification rates.

References

1. Dunkle, R. V , "Solar Water Distillation The Roof Type Still and a Multiple Effect Diffusion Still," paper presented at the 1961 International Heat Transfer Conference, University of Colorado, Boulder, Colo , Part 5, pp 895-902, 1961

2 Morse, R N , and Read, W. R W., "A Rotational Basis for the Engineering Development of a Solar Still," *Solar Energy*, Vol 12, pp 5-17, 1968

3 Sayigh, A A. M., Elsalam, E M. A , "Optimum Design of a Single Slope Solar Still in Riyadh, Saudi Arabia," AIAA Paper No A77-42956, Technical Information Service, New York, N Y , 1977.

4. Cooper, P I , "Digital Simulation of Transient Solar Still Processes," *Solar Energy*, Vol. 12, pp. 313-331, 1969.

5 Menninger, F J , and Elder, R J , "A Solar-Powered Water Purification System at Goldstone DSS 13," *TDA Progress Report 42-66*, Jet Propulsion Laboratory, Pasadena, Calif , pp 352-354, Dec 15, 1981

6 *ASHRAE Handbook of Fundamentals*, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc , New York, 1972.

7 Lansing, F L , and Yung, C S , "A Two-Dimensional Thermal Analysis of a New High Performance Tubular Solar Collector," *DSN Progress Report 42-49*, Jet Propulsion Laboratory, Pasadena, Calif , pp 116-131, Feb 15, 1979

8 Holman, J P , *Heat Transfer*, McGraw Hill Book Co , New York, N Y , 1981

9 Siegel, R , and Howell, J R , *Thermal Radiation Heat Transfer*, McGraw-Hill Book Co , New York, 1972

10 Eckert, E R G , and Drake, R M , *Analysis of Heat and Mass Transfer*, McGraw-Hill Book Co , New York, 1972.

11 Duffie, J. A , and Beckman, W A , *Solar Energy Thermal Processes*, Wiley Interscience Publication, New York, 1973

12 Keenan, J H , and Keys, G , *Thermodynamic Properties of Steam*, Wiley, New York, 1936

List of Symbols

A	surface area
C	specific heat
D	water vapor, air diffusion coefficient
F	radiation view factor
g	gravitational constant
Gr	Grashof number
h_c	convective heat transfer coefficient
h_d	mass transfer coefficient
H'	enthalpy of saturated water liquid
H''	enthalpy of saturated water vapor
I	solar radiation intensity
k	thermal conductivity
l	characteristic length
L	latent heat of vaporization
Le	Lewis number
m	rate of makeup water or distilled water per unit area
M	molecular weight
Nu	Nusselt number
P	pressure
Pr	Prandtl number
$\overline{R}$	universal gas constant
Sc	Schmidt number
Sh	Sherwood number
t	time
T	absolute temperature
V	wind speed
X	constants
α	absorptivity
$\overline{\alpha}$	thermal diffusivity
β	volumetric expansion coefficient
δ	thickness
ϵ	emissivity
η	efficiency
μ	dynamic viscosity
ν	kinematic viscosity
ρ	reflectivity
$\overline{\rho}$	density
σ	Stefan Boltzman constant
τ	transmissivity
ϕ	extinction coefficient

Subscripts

a	airspace between glass and water
c	convection
e	effective
f	bottom fiberglass
g	glass
i	bottom insulation
o	ambient air
r	radiation
s	bottom plywood support
w	water

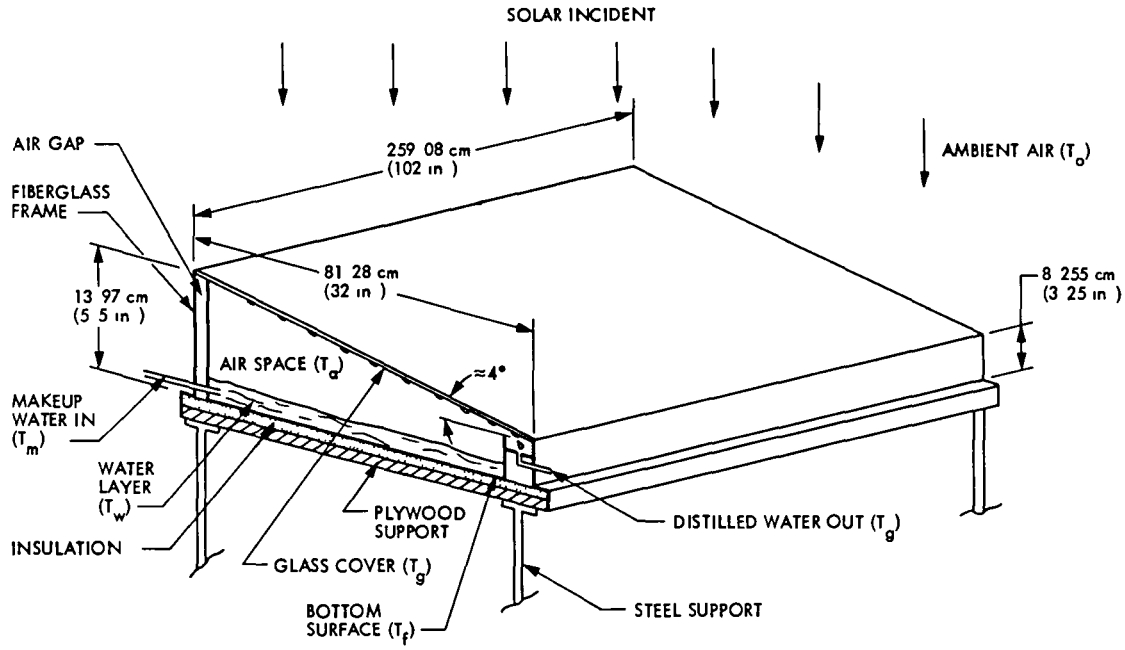


Fig 1 Layout of JPL solar-powered distillation module

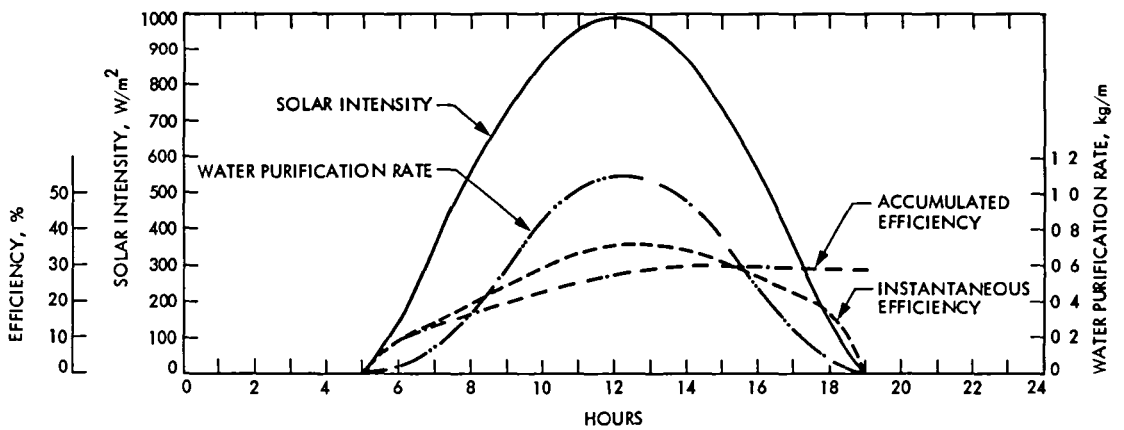


Fig 2 Daily performance of one module in June

Appendix A
Derivation of Quasi-Steady-State Equations

The derivations of the governing heat and mass transfer equations for the quasi-steady state conditions are made following the assumptions outlined in Section III. The heat transfer coefficients are given in Appendix B and the mass transfer coefficients are given in Appendix C Figure A-1 depicts the energy flow to and from the major components the glass cover, the water layer and the bottom of the still.

A. Optical Properties

The optical effect of the glass cover and the water layer of the solar still can be modeled as two parallel transparent media with an opaque bottom surface for the second medium representing the black interior bottom of the still The expressions of the "effective" absorptivity, reflectivity, and transmissivity due to multiple ray reflections are taken from Ref 7 as follows

For the glass cover, the effective absorptivity is

$$\left.\begin{aligned} \alpha_{g,e} &= \alpha_g + \frac{\alpha_g \tau_g \rho_w}{1-\rho_g \rho_w} \\ \text{for the water layer} \\ \alpha_{w,e} &= \alpha_w \tau_g/(1-\rho_g \rho_w) \\ \text{for the bottom surface} \\ \alpha_{f,e} &= \alpha_f \tau_g/(1-\rho_g \rho_w) \end{aligned}\right\} \quad \text{(A-1)}$$

where the single-layer absorptivities $\alpha_g, \alpha_w, \alpha_f$ for the glass cover, water layer, and the bottom surface, respectively, are given by the expressions

$$\left.\begin{aligned} \alpha_g &= (1-r_g)(1-a_g)/(1-a_g r_g) \\ \alpha_w &= (1-r_w)(1-a_w)(1+a_w b)/(1-a_w^2 b r_w) \\ \alpha_f &= a_w(1-b)(1-r_w)/(1-a_w^2 b r_w) \end{aligned}\right\} \quad \text{(A-2)}$$

where r_g and r_w are the ray reflection coefficient from the upper surface of glass and water respectively, b is the reflection coefficient of the opaque bottom surface of water and a is the ray absorption coefficient determined by

$$\left.\begin{aligned} a_g &= e^{-\phi_g \delta_g} \\ a_w &= e^{-\phi_w \delta_w} \end{aligned}\right\} \quad \text{(A-3)}$$

where ϕ is the extinction coefficient (in m^{-1}) Note that the reflectivities for the glass cover ρ_g and for the water layer ρ_w are computed from

$$\left.\begin{aligned} \rho_g &= r_g + \frac{r_g a_g^2 (1-r_g)^2}{(1-r_g^2 a_g^2)} \\ \rho_w &= r_w + [a_w^2 b (1-r_w)^2]/(1-a_w^2 b r_w) \end{aligned}\right\} \quad \text{(A-4)}$$

By using Eqs (A-2) and (A-4), the transmissivity of the glass cover τ_g must satisfy the energy equation

$$\tau_g = 1 - \rho_g - \alpha_g \quad \text{(A-5)}$$

B. Heat Balance Equations

Figure (A-1) depicts the energy flux to and from the still components At steady-state conditions the heat balance equations are divided as follows

1 The glass cover By selecting a control volume as shown in Fig A-2, the energy balance equation is

$$\begin{aligned} &\alpha_g I + m_w H''_{Tw} + (h_{c,wg} + h_{r,wg})(T_w - T_g) \\ &= (h_{c,go} + h_{r,go})(T_g - T_o)\left(\frac{A_g}{A_w}\right) + m_w H'_{Tg} \end{aligned} \quad \text{(A-6)}$$

where m_w is the mass flow rate of water per unit area, and I is the solar radiation intensity on the projected area of glass cover A_g (which is equal to the water surface area A_w) The enthalpy terms can be combined as

$$H''_{Tw} - H'_{Tg} = L_{Ta} + C_w (T_w - T_g) \quad \text{(A-7)}$$

where L_{Ta} is the latent heat of vaporization evaluated at the internal air space temperature T_a T_a is taken as the average

temperature between T_w and T_g Combining Eqs (A-6) and (A-7) gives

$$T_w = [(B_1 + B_2 + m_w C_w) T_g - (E_1 + B_1 T_o)]/(B_2 + M_w C_w) \tag{A-8}$$

where

$$\left.\begin{aligned} B_1 &= (h_{c,go} + h_{r,go})(A_g/A_w) \\ B_2 &= h_{c,wg} + h_{r,wg} \\ E_1 &= \alpha_g I + m_w L_{Ta} \end{aligned}\right\} \tag{A-9}$$

2 The water layer For the water layer, the energy equation can be derived by using the control volume as shown in Fig A-2 as

$$\alpha_w I + h_{c,fw}(T_f - T_w) + m_w H'_{Tm} = (h_{c,wg} + h_{r,wg})(T_w - T_g) + m_w H''_{Tw} \tag{A-10}$$

where H'_{Tm} is the enthalpy of the make up water at T_m, and H''_{Tw} is the enthalpy of the saturated water vapor at T_w. The difference in enthalpy $(H''_{Tw} - H'_{Tm})$ can be written as

$$H''_{Tw} - H'_{Tm} = L_{Tw} + C_w (T_w - T_m) \tag{A-11}$$

where L_{Tw} is the latent heat of vaporization of water at T_w (or taken at T_a approximately) Equations (A-10) and (A-11) give the temperature T_f as

$$T_f = (B_3 T_w - B_2 T_g - E_2)/h_{c,fw} \tag{A-12}$$

where

$$\left.\begin{aligned} B_3 &= h_{c,fw} + B_2 + m_w C_w \\ E_2 &= \alpha_w I + m_w C_w T_m - m_w L \end{aligned}\right\} \tag{A-13}$$

3 Bottom fiberglass surface For the bottom fiberglass surface

$$\alpha_f I = h_{c,fw}(T_f - T_w) + h_{fo}(T_f - T_o) \tag{A-14}$$

or

$$T_f = (E_3 + h_{c,fw} T_w)/B_4 \tag{A-15}$$

where

$$\left.\begin{aligned} B_4 &= h_{c,fw} + h_{fo} \\ E_3 &= \alpha_f I + h_{fo} T_o \end{aligned}\right\} \tag{A-16}$$

By manipulating Eqs (A-8), (A-12), and (A-15), the temperature T_g can be expressed as

$$T_g = (X_0 + X_1 T_0 + X_2 T_m)/(X_1 + X_2) \tag{A-17}$$

where

$$X_0 = (m_w L + \alpha_g I)\left[\frac{h_{c,fw} h_{fo}}{h_{c,fw} + h_{fo}}\right] + I(B_2 + m_w C_w)\left[\alpha_g + \alpha_w + \left(\frac{\alpha_f h_{c,fw}}{h_{c,fw} + h_{fo}}\right)\right] \tag{A-18}$$

$$X_1 = (B_1 + B_2 + m_w C_w)\left[\frac{h_{c,fw} h_{fo}}{h_{c,fw} + h_{fo}}\right] + B_1(B_2 + m_w C_w)$$

$$X_2 = m_w C_w (B_2 + m_w C_w)$$

where the latent heat L is taken at the air space temperature T_a

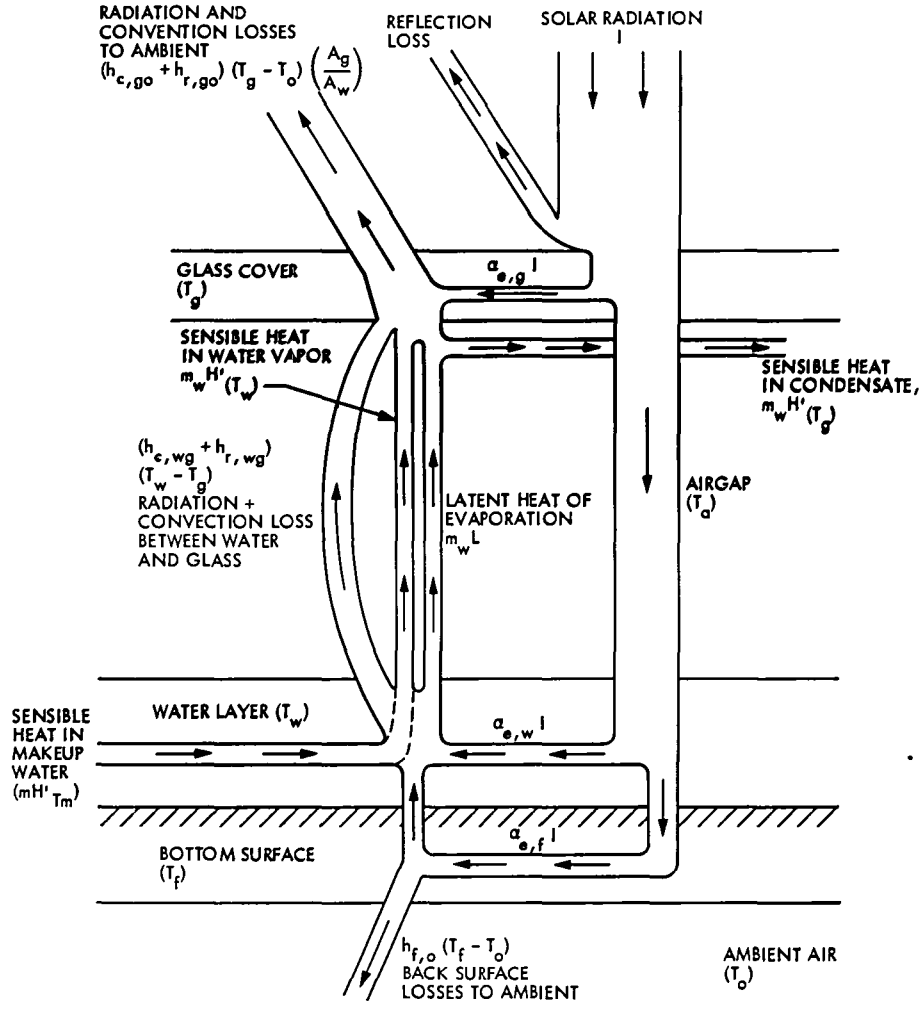


Fig A-1. Energy flux through still components

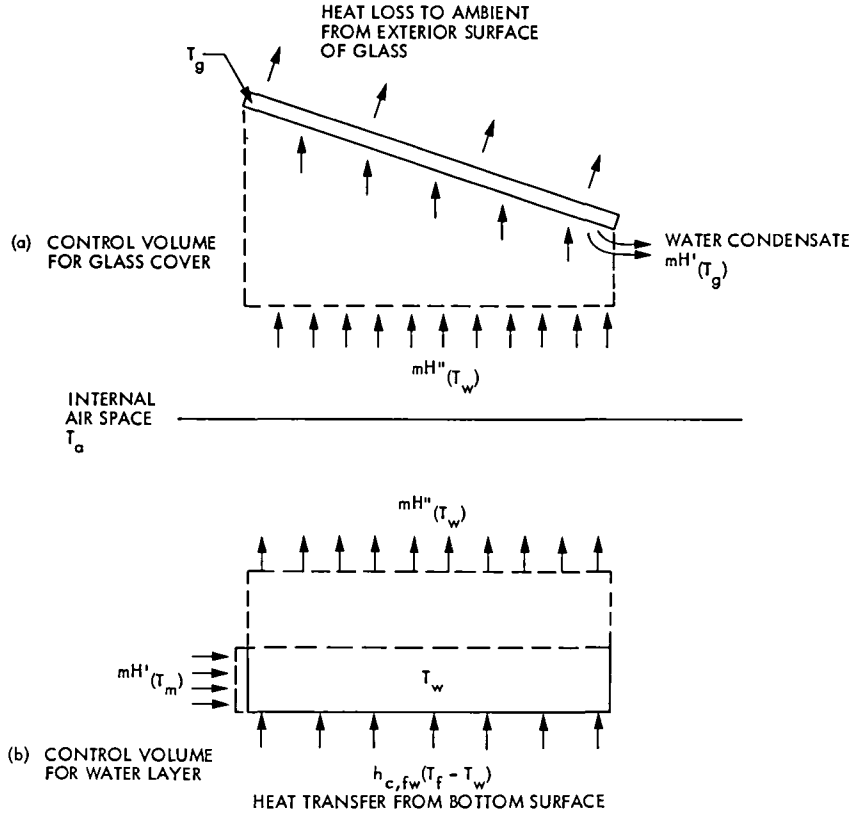


Fig A-2. Selected control volume for energy equations

Appendix B
Heat Transfer Coefficients

The development of the temperature distribution of the major solar still components as given in Appendix A uses a set of mass transfer, convection, and radiation heat transfer coefficients The heat transfer coefficients are grouped as follows

1 Radiation coefficient between glass cover and ambient air $h_{r,go}$ The coefficient $h_{r,go}$ is obtained by linearizing the expression of radiation exchange between the glass cover (taken as a gray surface with emissivity ϵ_g) and a black ambient air (treated as a black body), hence

$$h_{r,go} = \epsilon_g \sigma (T_g^4 - T_o^4)/(T_g - T_o)$$

or

$$h_{r,go} = \xi_g \sigma (T_g^2 + T_o^2)(T_g + T_o) \qquad \text{(B-1)}$$

where σ is Stefan-Boltzmann constant (= 5 729 X 10^{-8}W/($m^2 \cdot K^4$))

2 Radiation coefficient between glass cover and the water layer $h_{r,wg}$ The radiative coefficient $h_{r,wg}$ is obtained by linearizing the expression of radiation exchange between the glass cover and the water layer, treated as two gray surfaces

$$h_{r,wg} = \epsilon_{wg} \sigma (T_w^4 - T_g^4)/(T_w - T_g)$$

or

$$h_{r,wg} = \epsilon_{wg} \sigma (T_w^2 + T_g^2)(T_w + T_g) \qquad \text{(B-2)}$$

where the effective emissivity ϵ_{wg} is given by Holman (Ref 8)

$$\frac{1}{\epsilon_{wg}} = \left(\frac{1-\epsilon_g}{a_g}\right)\left(\frac{A_w}{A_g}\right) + \left(\frac{A_w}{A_g F_{wg}}\right) + \frac{1-a_w}{\epsilon_w} \qquad \text{(B-3)}$$

where F_{wg} is the radiation view factor between the water and glass surfaces Denoting the top, bottom, and side surfaces of the still module as shown in Fig (B-1) and making use of the view factors expressions between two perpendicular rectangles (Ref 9),

$$F_{2-4} = \frac{1}{2}\left(1 + \frac{h_4}{w_2} - \sqrt{1 - \left(\frac{h_4}{w_2}\right)^2}\right) \qquad \text{(B-4)}$$

$$F_{2-3} = \frac{1}{2}\left(1 + \frac{h_3}{w_2} - \sqrt{1 - \left(\frac{h_3}{w_2}\right)^2}\right) \qquad \text{(B-5)}$$

since surface 2 does not"see" itself,

$$F_{2-2} = 0$$

$$F_{2-1} = 1 - F_{2-3} - F_{2-4} = F_{wg} \qquad \text{(B-6)}$$

3 Convection coefficient between warm water surface and cold glass cover $h_{c,wg}$ For a horizontal air space enclosed between a warm plate at the bottom and a cold plate at the top, Holman (Ref 8) gave the natural convection coefficient as

$$h_{c,wg} = 0\ 195\ Gr_{\delta a}^{0\ 25}\ K_a/\delta_a \qquad 10{,}000 < Gr < 400{,}000 \qquad \text{(B-7)}$$

$$h_{c,wg} = 0\ 068\ Gr_{\delta a}^{0\ 333}\ K_a/\delta_a \qquad Gr > 400{,}000$$

where $Gr_{\delta a}$ is the Grashof number for the air, δ_a is the thickness of the air space, and k_a is the thermal conductivity of the air evaluated at temperature T_a k_a in W/m-K is determined by a curve-fitted expression using tabulated data in Ref 10 as

$$k_a = 0\ 0003623\ T_a^{0\ 7488} \qquad \text{(B-8)}$$

The Grashof number is computed with δ_a as the characteristic length by

$$Gr\delta_a = g\beta (T_w - T_g)\ \delta_a^3/\nu_a^2 \qquad \text{(B-9)}$$

where g is the gravitational constant (9 81 m/sec^2), β is the volumetric expansion coefficient (equals $1/T_a$ for an ideal gas at T_a), ν_a is the kinematic viscosity of air (in m^2/sec) evaluated at T_a An expression for ν_a is obtained, also, by curve-fitting tabulated data from Ref 10 as

$$\nu_a = 9\ 253 \times 10^{-10}\ T_a^{1\ 709} \tag{B-10}$$

4 Convection coefficient between the warm bottom surface and the relatively cold water layer $h_{c,fw}$ The average natural convection coefficient between a horizontal flat plate and a liquid is given by Holman (Ref 8) as

$$h_{c,fw} = 0\ 27\ k_w (Gr_l Pr_w)^{0\ 25} / l_w \tag{B-11}$$

where k_w (in W/m-K) is the thermal conductivity of water evaluated at T_w, l is the characteristic length taken as the mean of the two sides of the rectangular bottom surface of the still, and Pr_w is the Prandtl number for water By curve fitting tabulated data in Ref 10 one obtains

$$k_w = 0\ 557 + 0\ 0011799\ (T_w - 273\ 15) \tag{B-12}$$

and

$$Pr_w = -0\ 222365 + \frac{200\ 90503}{(T_w - 273\ 15)} - \frac{1115\ 8828}{(T_w - 273\ 15)^2} \tag{B-13}$$

The Grashof number of water is computed from

$$Gr_w = 0\ 00018 g\ (T_f - T_w)\ l^3 / \nu_w^2 \tag{B-14}$$

where the volumetric expansion coefficient for water β is $0\ 18 \times 10^{-3}\ k^{-1}$ The kinematic viscosity of water (ν_w) in m^2/sec is determined by curve-fitting tabulated data in Ref 10 as

$$\nu_w = 1\ 1853606 \times 10^{-10} + \frac{32\ 562357 \times 10^{-6}}{(T_w - 273\ 15)} - \frac{248\ 87299 \times 10^{-6}}{(T_w - 273\ 15)^2} \tag{B-15}$$

5. Convection coefficient between glass cover and the ambient air $h_{c,go}$ The forced convection heat transfer coefficient $h_{c,go}$ in W/m^2-K is given by Duffie and Beckman (Ref 11) as

$$h_{c,go} = 5\ 7 + 3\ 8V \tag{B-16}$$

where V is the wind speed in m/sec

6 Overall heat transfer coefficient from the bottom surface to the ambient air As shown in Fig 1, the bottom portion of solar still is composed of three layers a fiberglass layer, an insulation layer, and a plywood layer A general expression for the overall heat transfer coefficient h_{fo} through this composite wall can be written as

$$\frac{1}{h_{fo}} = \frac{\delta_i}{k_i} + \frac{\delta_s}{k_s} + \frac{1}{h_{c,fo}} \tag{B-17}$$

where δ_i and δ_s are the thickness of the insulation and the plywood support, respectively, and k_i and k_s are the thermal conductivity of the insulation and the plywood support, respectively The convection coefficient between the bottom plywood surface and the ambient air is $h_{c,fo}$

7. Latent heat of vaporization for water The latent heat of vaporization expression given by Cooper (Ref 4) in English units was converted to SI units as

$$L = 883 - 0\ 6687\ T \tag{B-18}$$

where T is the absolute water temperature in kelvin and L is given in Whr/kg

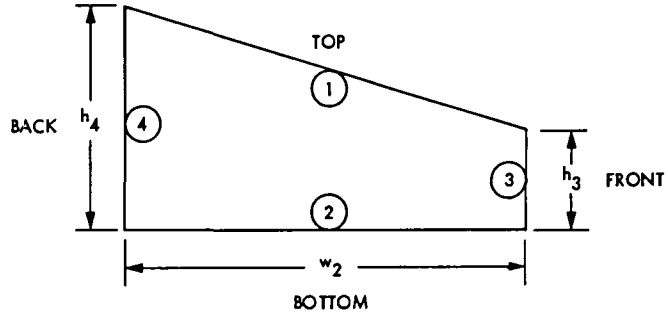


Fig B-1 Internal solar still radiation view factor

Appendix C
Mass Transfer and Diffusion

The mass transfer is significant in the model since it indicates the water purification rate and consequently the overall efficiency of the solar still Two approaches for evaluating the mass transfer rate, m_w, are presented as follows

I. First Approach: Stefan's Law of Diffusion

Following the Stefan's Law of Diffusion (Ref 8), the total mass flow of water vapor per unit area evaporating from the upper water surface to the still air above is

$$m_{evaporation} = \frac{DPM_w}{\bar{R}\,T_a(\delta_a/2)} \ln \frac{P - P_{wa}}{P - P_{ww}} \qquad \text{(C-1)}$$

where P is the total pressure for air-vapor mixture, $\bar{R}$ is the universal gas constant (8 315 kj/kmol-k), $(\delta_a/2)$ is half the air spacing between the water surface and glazing, D is the water-air diffusion coefficient, T_a is the temperature of air spacing, P_{wa} and P_{ww} are the saturation vapor pressures for water at temperatures T_a and T_w, respectively Also, for the condensation process at the glazing surface at T_g,

$$m_{condensation} = \frac{DPM_w}{\bar{R}T_a(\delta_a/2)} \ln \frac{P - P_{wg}}{P - P_{wa}} \qquad \text{(C-2)}$$

since at steady-state,

$$m_{condensation} = m_{evaporation} = m_w$$

By eliminating the pressure P_{wa} from Eqs (C-1) and (C-2) the mass flow m_w for evaporation or condensation at steady state is

$$m_w = \frac{DPM_w}{\bar{R}T_a\,\delta_a} \ln \frac{P - P_{wg}}{P - P_{ww}} \qquad \text{(C-3)}$$

where P_{wg} and P_{ww} are the saturated vapor pressure at temperature T_g and T_w, respectively An empirical equation for the saturated vapor pressure for water in KN/m^2 or (kP_a) is given by Kays (Ref 12) from 10°C to 150°C as

$$P_w = 22105\,8 \times 10^{-\left[\frac{X(a+bx+cx^3)}{T(1+dx)}\right]} \qquad \text{(C-4)}$$

where T is the saturated water temperature (K) and

$$\left.\begin{aligned} a &= 3\,2437814 \\ b &= 5\,86826 \times 10^{-3} \\ c &= 1\,1702379 \times 10^{-8} \\ d &= 2\,1878462 \times 10^{-3} \\ x &= 647\,27 - T \end{aligned}\right\} \qquad \text{(C-5)}$$

The diffusion coefficient D between water vapor and air in cm^2/sec is taken from Gilliland's semiempirical relation for any two gases A and B (Ref 8) as

$$D = 577\,7\,T_a^{3/2} \sqrt{\frac{1}{M_A} + \frac{1}{M_B}} \Big/ [P(\bar{V}_A^{1/3} + \bar{V}_B^{1/3})^2] \qquad \text{(C-6)}$$

where P is the total pressure of the air-vapor mixture in N/m^2, and $\bar{V}_A$ and $\bar{V}_B$ are the molecular volume of A and B For a mixture of air and water vapor the values of $\bar{V}$ are 29 9 and 18 8, respectively The molecular weights M for air and water vapor are 28 9 and 18, respectively At atmospheric pressure $P = 1\,0132 \times 10^5$ N/m^2, Eq (C-6) is reduced to

$$D = 0\,04977\,(T_a/100)^{3/2} \qquad \text{(C-7)}$$

Equations (C-3) and (C-7) are combined to give the distilled water production rate m_w in the text

II. Lewis Number Approach

The mass transfer (or diffusion) of water vapor is the isothermal evaporation of water at T_w and the subsequent diffusion through the stagnant air layer above at T_a By analogy to heat transfer, the mass transfer per unit surface area is expressed as

$$m_w = h_d(C_w - C_{wa}) \qquad \text{(C-8)}$$

where h_d is the mass transfer coefficient, C_w and C_{wa} are the water vapor concentration (or density $\bar{\rho}$) in saturated air-vapor mixture at temperature T_w and T_a, respectively

Assuming that water vapor at low pressure follows an ideal gas equation of state, the density $\bar{\rho}_w$ is given by

$$\bar{\rho}_w = M_w P_w/(\bar{R}T) \tag{C-9}$$

where M_w is the molecular weight of water, and P_w is the partial pressure of water vapor in the saturated air layer above, hence

$$m_{evaporation} = h'_d (M_w/\bar{R}T_a)(P_{ww} - P_{wa}) \tag{C-10}$$

where h_d is an evaporation mass transfer coefficient, and P_{ww} and P_{wa} are the saturation pressures of water vapor at the temperatures T_w and T_a, respectively Similarly, the condensation of water vapor taking place at the glass surface at temperature T_g is a result of the concentration difference between water vapor in two saturated air mixtures, one at temperature T_a and the other at a lower temperature T_g An analogous expression to Eq (C-10) can be written for the condensing mass rate, as

$$m_{condensation} = h''_d(M_w/\bar{R}T_a)(P_{wa} - P_{wg}) \tag{C-11}$$

where h''_d is the mass transfer coefficient during condensation At steady state, the make-up water will compensate for the condensate leaving the still, then

$$m_w = m_{evaporation} = m_{condensation}$$

Equating Eqs (C-10) and (C-11) and eliminating P_{wa}, one obtains the mass flow rate per unit surface area as

$$m_w = M_w(P_{ww} - P_{ww})/\left[\bar{R}T_a\left(\frac{1}{h'_d} + \frac{1}{h''_d}\right)\right] \tag{C-12}$$

Furthermore, if we assume that,

$$h'_d = h''_d = h_d \tag{C-13}$$

where h_d is the mass transfer coefficient at either the interface between the water and air space or between the air space and the glazing, then

$$m_w = (M_w h_d/2\,\bar{R}T_a)\,(P_{ww} - P_{wg}) \tag{C-14}$$

The mass transfer coefficient h_d is obtained following similar relationships between convective heat and mass transfer These are given by the dimensionaless numbers below

$$\left.\begin{array}{l}\text{(a) Schmidt number } Sc = \dfrac{\nu}{D} \text{ in mass transfer} \\ \text{Prandtl number } Pr = \dfrac{\nu}{\alpha} \text{ in heat transfer}\end{array}\right\} \tag{C-15}$$

and

$$\left.\begin{array}{l}\text{(b) Sherwood number } Sh = \dfrac{h_d \delta}{D} \text{ in mass transfer} \\ \text{Nusselt number } Nu = \dfrac{h_c \delta}{k} \text{ in heat transfer}\end{array}\right\} \tag{C-16}$$

where $\bar{\alpha}$ is the thermal diffusivity (= $k/\bar{\rho}\bar{c}$), ν is the kinematic viscosity (= μ/ρ), and D is the diffusion coefficient

If heat and mass transfer occur simultaneously as is the case in the solar still, the coefficients h_c and h_d are found to be related by Lewis number Le (Ref 8) as

$$\left.\begin{array}{l}h_c/h_d = \bar{\rho}_a C_a\, Le^{2/3} \\ Le = \dfrac{\bar{\alpha}_a}{D} = \dfrac{Sc}{Pr}\end{array}\right\} \tag{C-17}$$

Knowing the value of Lewis number, the physical properties of air-water vapor mixture and the convective heat transfer coefficient h_c will enable the mass transfer coefficient h_d to be computed from Eq (C-14) This approach was abandoned, however, in favor of the first, simpler, approach which uses direct calculation of the diffusion coefficient D

Appendix D

Numerical Example

A short computer program was developed to speed up the simulation of a solar still module using the equations presented in Appendices A, B, and C The total water production rate of the distillation system can be obtained by multiplying the output of one module by the total number of modules in the system

I. Input Variables

The variables used in the computer program are grouped as follows

A. Optical Properties

Glass reflection coefficient[1] r_g	0 043
Water reflection coefficient[2] r_w	0 0201
Glass absorption coefficient[3] a_g	0 9730
Water absorption coefficient[4] a_w	0 7065
Bottom surface reflection coefficient[5] b	0 05
Emissivity of water surface ϵ_w	0 9565
Emissivity of glass ϵ_g	0 925
View factor F_{wg}	0.8533

(1) Based on an index of refraction of 1 526

(2) Based on an index of refraction of 1 33

(3) Using an extinction coefficient of 6 85 m^{-1} and Eq (A-3)

(4) Using an extinction coefficient of 6 85 m^{-1} and Eq (A-3)

(5) Assuming an absorption coefficient of 0 95 for the opaque black-painted bottom

B. Module Geometry

Average spacing between water surface and the glass cover	6 00 cm
Length of the still base	259 08 cm
Width of the still base	81.28 cm
Height of the front side of the still	8 25 cm
Height of the back side of the still	13 97 cm
Thickness of glass X_g	0 4 cm
Thickness of water layer X_w	5 08 cm
Thickness of fiberglass layer X_f	0 3175 cm
Thickness of insulation X_t	1 905 cm
Thickness of support wood frame X_s	1 905 cm
Surface area of glass A_g	2 111 m^2
Surface area of water A_w	2 1059 m^2

C. Thermodynamic Properties

Prandtl number of air Pr_a	0 7
Conductivity of insulation k_t	0 151 w/m-K
Conductivity of wood k_s	0 2942 w/m-K
Conductivity of fiberglass k_f	0 433 w/m-K
Specific heat of water C_w	1 163 Wh/Kg-K
Total pressure P	97 7 KN/m^2 (KP_a) or 14 2 psi

D. Weather

Solar radiance I	800 W/m^2
Wind speed V	4 47 m/sec (10 mph)
Ambient temperature T_0	299 85 K (80 1°F)

E. Temperatures

Makeup water temperature T_m	299 85 K (80 1°F)

[1] Based on an index of refraction of 1 526
[2] Based on an index of refraction of 1 33
[3] Using an extinction coefficient of 6 85 m^{-1} and Eq (A-3)
[4] Using an extinction coefficient of 6 85 m^{-1} and Eq (A-3)
[5] Assuming an absorption coefficient of 0 95 for the opaque black-painted bottom

II. Sample Results

The following is a partial list of the results printed by the computer using the input data presented above and the equations given in Appendices A, B, and C

A. Optical Properties

Effective absorptivity of glass $\alpha_{e,g}$	0 02805
Effective absorptivity of water $\alpha_{e,w}$	0 26693
Effective absorptivity of bottom $\alpha_{e,f}$	0 58950

B. Thermal Properties

Conductivity of air k_a	0 02787 W/m-K
Conductivity of water k_w	0 6403 W/m-K
Kinematic viscosity of air ν_a	1 8664 X 10^{-5} m^2/sec
Kinematic viscosity of water ν_w	4 1132 X 10^{-7} m^2/sec
Prandtl number of water Pr_w	2 3988
Latent heat of evaporation at T_a L_{Ta}	662 15 Whr/kg
Saturated vapor pressure at T_g P_{wg}	8 91133 KN/m^2
Saturated vapor pressure at T_w P_{ww}	31.9833 KN/m^2

C. Heat Transfer Coefficients

Radiation coefficient between glass cover and ambient air $h_{r,go}$	6 2165 W/m^2-K
Radiation coefficient between water layer and glass cover $h_{r,wg}$	6 3690 W/m^2-K
Forced convection coefficient between glass cover and ambient air $h_{c,go} = h_{c,fo}$	22 686 W/m^2-K
Natural convection coefficient between water layer and glass cover $h_{c,wg}$	2 503 W/m^2-K
Overall heat transfer coefficient between the bottom surface and the ambient air h_{fo}	4 25549 W/m^2-K
Natural convection coefficient between the bottom surface and the water layer, $h_{c,fw}$	81 245 W/m^2-K
Combined heat transfer coefficients B_1	28 9725 W/m^2-K
B_2	8 8717 W/m^2-K
B_3	90 5190 W/m^2-K
Equivalent coefficient B_4	85 5011 W/m^2-K

D. Temperatures

Glass cover temperature T_g	317 1 K
Water layer temperature T_w	343 8 K
Bottom surface temperature T_f	347 2 K

E. Performance

Water production rate per module	0 727 kg/hr
Quasi-steady-state efficiency	28 6%

From the energy balance equations presented in Appendix A, an energy flux distribution of the solar still can be developed Figure D-1 shows the energy flux through the module. Assuming that 100 units of solar energy are radiated upon the still projected surface, 11.6 units of energy will be reflected back to the ambient air The glass cover absorbs 2 8 units of energy and transmits the remaining 85 6 units to the still interior The water layer absorbs 26 7 units of energy and the remainder is absorbed by the bottom surface The makeup water for this example contributes 0 9 units of sensible heat loss from the solar still The thermal loss of the top surface is more than twice that of the bottom The back loss could be reduced further if the still is better insulated

Water distillation rate depends mainly on incident solar radiation To investigate this effect, the ambient temperature T_0 was kept constant at 26 7°C (299 85 K) and several simulations were made using various solar intensities The results are illustrated in Fig D-2, showing a nonlinear relationship between the solar radiation intensity and the water purification rate At 200 W/m^2 solar intensity the water distillation rate is approximately 0 05 kg/(hr-m^2), which increases to 0 35 kg/(hr-m^2) at 800 W/m^2 The water distillation sensitivity is accelerated as the solar radiation becomes higher than 400 W/m^2 Since the average daily solar radiation is higher in summer than in winter, the water distillation rate will be high in summer and low in winter, as shown in Fig. D-2

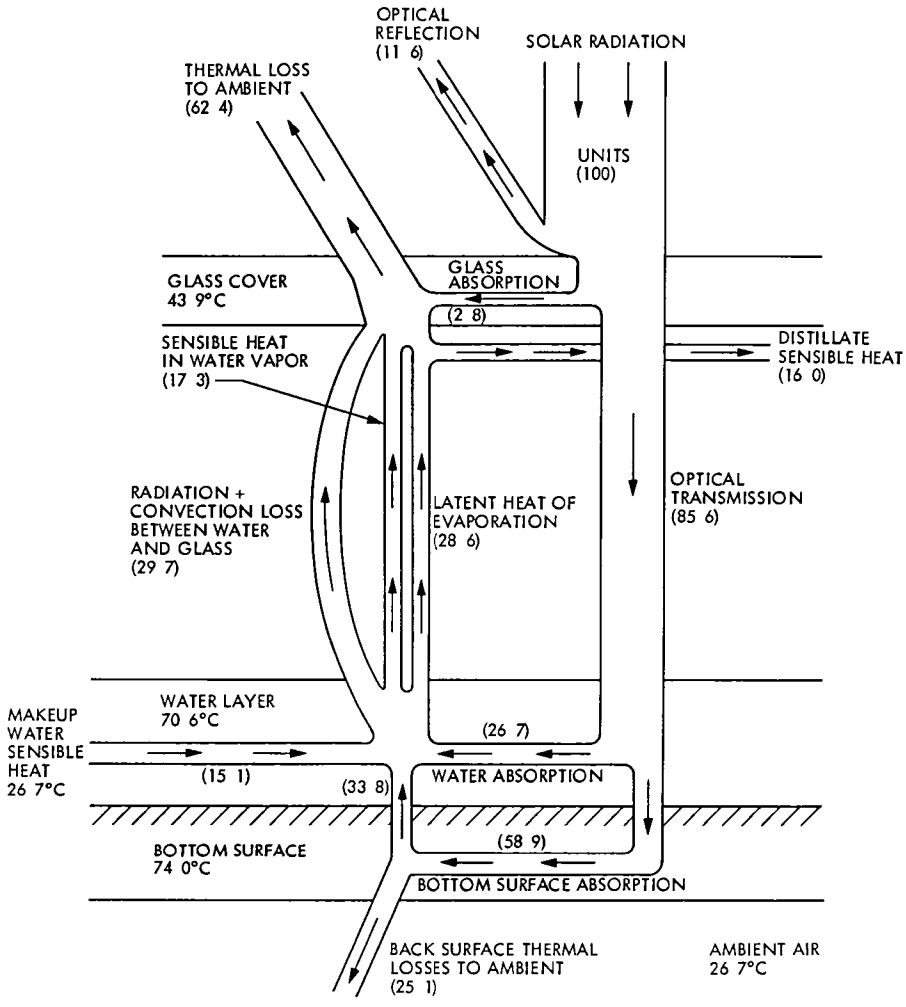


Fig D-1 Energy flux through still components for the sample 1-hour simulation

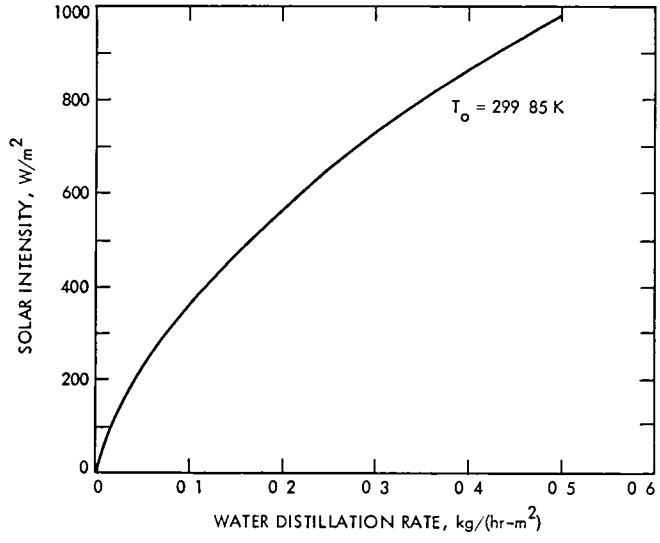


Fig D-2 Solar radiation versus water distillation rate

Efficient Computerized Model for Dynamic Analysis of Energy Conversion Systems

R D Hughes, F L Lansing, and I R Khan
DSN Engineering Section

In searching for the optimum parameters that minimize the total life cycle cost of an energy conversion system, one must examine various combinations of components and study the resulting system performance and associated economics The Systems Performance and Economics Simulation computer program (SPECS) was developed to fill this need The program simulates the fluid flow, thermal, and electrical characteristics of a system of components on a quasi-steady-state basis for a variety of energy conversion systems A unique approach is used in which the set of characteristic equations is solved by the Newton-Raphson technique. This approach eliminates the tedious iterative loops which are found in comparable programs such as TRNSYS or SOLTES-1 Several efficient features were also incorporated such as the centralized control and energy management scheme, the analogous treatment of energy flow in electrical and mechanical components, and the modeling of components of similar fundamental characteristics using generic subroutines Initial tests indicate that this model can be used effectively with a relatively small number of time steps and low computer cost

I. Introduction

The satisfactory development and design of efficient and economical energy conversion systems require a thorough understanding of the interactions of the energy sources, system components, exterior environment, load distribution and automatic control devices It is necessary to select components from among many proposed configurations that vary widely in performance characteristics and affect the overall system cost Since experiments on new configurations are expensive and time-consuming, properly formulated computer simulations can provide the same information at much less time and cost

As part of the continuing engineering efforts to upgrade the performance of the NASA-owned antenna facilities of the Deep Space Network (DSN) at JPL, it is desirable to obtain not only efficient "working" solutions but also "optimized" ones that minimize total system cost Two steps should be taken in order to obtain optimum solutions First, for a given energy conversion system configuration, the optimized performance parameters must be determined Because the design configurations of components are often extensive (Ref 1), some bounds and size limits should be set on these parameters Reliability and availability constraints are also added Energy conservation measures, which reduce the energy consumption of the facility and subsequently affect

its operation and maintenance costs, must also be taken into account Second, tradeoff studies of several optimized configurations must be made to select the optimum one

At the supply side of an energy conversion system, energy is generated from a set of power plants, at the demand side, the energy is distributed among facility loads It is essential, therefore, to match and balance, with minimum energy losses, the flow of energy between the generation and distribution sides With the above energy balance, reliability, and cost constraints in mind, a detailed dynamic (i e , time-varying) simulation of energy conversion systems is needed to analyze the system behavior and costs on a time-dependent basis

It is common in system design to select equipment by assigning single values of key properties (such as pressures, temperatures, flow rates, voltages, currents, etc) at critical network stations and select the appropriate size of components that meet given specifications The more efficient dynamic simulation model, in contrast, assumes that the system configuration is defined, that all component sizes are selected and that the performance characteristics of all components are known The simulation model then computes the station properties (pressures, temperatures, flow rates, voltages, and currents) at other than design loads The model hence suits the following purposes

(1) It can be used as an analytical tool for equipment design for large energy conversion systems where the cost of computer simulation represents a small portion of the total cost The program can assist the designer in avoiding oversized or undersized systems and in selecting optimum equipment size After completing a trial component selection based on one set of loadings, one can apply the model to other partial or overloading conditions to determine if the original design is adequate

(2) It serves as a basis for developing "short cut" techniques after making a large number of detailed simulations on many different designs These shorter solutions can be helpful in making initial assessments

(3) It provides the field operator with a learning and diagnostic aid for understanding the dynamic performance of the operating system and the interrelationships between components at off-design conditions

(4) It can be used to study the effects of add-on components and analyze the corrections sought for some operating problems if an existing system is expanded or reconfigured The benefits of any modification can be assessed by analyzing the simulation results before and after making a change

II. Survey of Existing Programs

A. Selection Criteria

The need to have such a time-varying simulation model has been encountered in many facility projects The criteria described below are used in selecting the proper computer model in our literature search

(1) The model should be a well-documented and low cost tool with an easily understood input, incorporating basic and fundamental analytical expressions, and thereby being available to a wide variety of users

(2) The model should be modular and open-ended to simulate either small or large problems The system size is determined by its number of components, and the model therefore should accommodate systems with arbitrary arrangements of series and parallel branches and have no limitations on these arrangements

(3) The model should have the capability to simulate steady-state conditions for a system without capacitive components as well as quasi-steady-state conditions for a system with capacitive components (such as storage tanks, batteries, etc) Quasi-steady-state simulation is done with time steps that are much larger than the largest time constant of all components A true transient simulation involves time steps that are smaller than the smallest characteristic time constant of all components A true transient simulation should be avoided since most of the components encountered in facility energy conversion systems, excluding capacitive components, respond to abrupt changes in local conditions in a relatively short time This time response varies from a few seconds to several minutes, which makes computations using large time steps (such as every half-hour or every hour) suitable for quasi-steady-state conditions The quasi-steady-state approach reduces the computation time tremendously without a significant loss of accuracy (Refs 2, 3)

(4) The model should treat mechanical and electrical components in an analogous fashion Examples are a fluid pump vs an electrical generator, a pipe vs a connecting cable, a fluid storage tank vs a battery, etc This analogous treatment will simplify the energy management and control schemes and reduce computer costs

(5) The model should have a central monitor and control software, as in real systems, to monitor changes in the system parameters and send the appropriate signals to respond to internal constraints as determined by the component's behavior or to external controls imposed by the user

The validity of any program depends primarily upon the formulation of the equations which it solves Fundamental laws and relations such as energy balances, momentum balances, mass balances, and properties of substances in combination with the performance characteristics of components must all be written in the proper form In general, one method of solving the equations is not necessarily as effective as another in saving time and computer cost Among the many programs written in system dynamics the following ones are examined

B. Available Programs

A detailed literature search showed that several computer programs which have been written for the dynamic simulation of solar-powered energy systems are relevant TRNSYS (Ref 4) for example, was developed by the University of Wisconsin to simulate solar-powered water heating and cooling systems Sandia Laboratories, Albuquerque, New Mexico, has developed SOLSYS (Ref 5), which is similar to TRNSYS but also includes simple Rankine-power cycles Sandia later on developed SOLTES-1 (Simulator of Large Thermal Energy Systems (Ref 6)), which is a combined version of SOLSYS and TRNSYS SOLTES-1 simulates, in a modular approach, Rankine-power cycles, solar powered heating and cooling systems Though flexible and modular, TRNSYS and SOLTES-1 are limited to solar-powered systems only

Other dynamic models such as LASL and CAL-ERDA were developed by Los Alamos Scientific Laboratory (LASL), Los Alamos, New Mexico (Ref 7) These are limited also to the dynamic description of a double-glazed flat-plate solar-collector, a hot water energy storage, an absorption water chiller, an air-handler, and a building shell and its interior loads for solar-heated and solar-cooled buildings

In computer hardware real-time simulation, the Utility Systems Group at JPL (Ref 8) has developed an Electric Power System Simulator (EPSS) for the U S Department of Energy (DOE) Office of Electric Energy Systems, which is coordinated with NASA's Energy Systems Division The simulator is a microprocessor-based device which is designed to illustrate the complexities of operating a real utility power system that consists of several combinations of electrical power generation units It is a portable, table-top device consisting of display and video monitors portraying the total system performance for different combinations of energy generation, energy storage, and load-varying effects Costs of energy are user-provided for each component according to the load curves and operating characteristics

W Stoecker at the University of Illinois, Urbana-Champaign, (Ref 2), has developed a dynamic simulation using a different approach for centralized chilled water circuits He has combined the Newton-Raphson technique for solving nonlinear pressure balance equations with a simple Gaussian elimination technique for solving linear heat-balance equations The extension of this approach to complex energy conversion systems employing phase-change fluids was not tried elsewhere, although it is worth investigating because of its simplicity and time-saving potential

In addition to the above programs, other models were developed by independent groups DYNSIM by Honeywell Energy Resources Center, Minnesota (Ref 9), SOLCOST by Martin Marietta Corporation, Colorado, HISPER by NASA's Marshall Space Flight Center (MSFC), Alabama, and STEPS by Colorado State University, Colorado (Ref 10) Differences in accuracies of modeling a given problem have varied from +2% to +20% (Ref 11), which shows the need for program unification and standardization

After comparing the above available computer programs with their differences regarding handling of equations, completeness of documentation, operating time and cost, complexity of input data, and ability to be reconfigured to accommodate additional components, it was decided that writing a new program instead of modifying an existing one was necessary in order to obtain an engineering tool which would have the desired capabilities and efficiency The Systems Performance and Economics Simulation Program (SPECS) was developed to fill this need This article describes the following aspects of the SPECS computer program the method used for modeling a system of components, the mathematical representation of these components, the numerical technique used in solving the resulting system of equations, and the program input and output

III. Modeling Theory

Since an energy conversion system is a set of components interconnected to accomplish a specific task, it is by definition modular It is possible to simulate its performance by collectively simulating the performance of its components A large problem can be reduced to a number of smaller problems, each of which can be solved independently Mathematically, the problem consists of constructing a set of nonlinear equations and solving these equations simultaneously using the Newton-Raphson technique instead of the commonly used consecutive approach This approach avoids iteration of configurations with cyclic (closed loops), acyclic (open ended) and multicyclic (joint loops) calculations The simultaneous solution process is repeated at each "selected" time step during a representative day of each month in the simulation period A detailed economic analysis of the system follows its performance evaluation to establish a base for further optimization study Additional features of the program are presented in the remainder of this section

A. Definitions

It is necessary to define the terms "station," "streamline," and "component," since each has a definite meaning in the program description

A "station" is a unique position on the electrical or fluid network A station is defined by 2-digit integers (e g , 27), a maximum of 99 stations can be defined

A "streamline" is the path of energy (or fluid) flux between two adjacent stations A streamline is identified by two stations, one upstream and one downstream

A "component" is the smallest constituent of the system which performs a well-defined physical function when assembled with other components Components as shown in Fig 1 can be simple, with one streamline (one inlet station and one outlet station) or compound, with two or more streamlines (one or more inlet stations and one or more outlet stations) Components are distinguished by two identifiers and described by a set of performance parameters The first identifier is for the component "type" which is a 2-digit integer (a maximum of 99 types could be identified) specifying whether the component is a pump, a pipe, a 2-way valve, a water tank, a heat exchanger, a wind turbine, etc The second identifier is a component "position" identifier giving the relative component location in the system and its associated links with other stations and streamlines No two components can have the same position identifier A particular component type can appear in more than one place as needed to complete the function of the system Additional component description parameters include size and flow characteristics and efficiencies or control strategies as described by the program inputs

B. Newton-Raphson Technique

The Newton-Raphson technique (Ref 2) used in the SPECS program is based on the linear approximation of the set of nonlinear functions $[F(X)] = 0$ around the root $[X^*]$ such that

$$\left.\begin{aligned} f_1(X) &= f_1(x_1, x_2, \quad , x_N) = 0 \\ f_N(X) &= f_N(x_1, x_2, \quad , x_N) = 0 \end{aligned}\right\} \tag{1}$$

or in matrix form

$$\begin{aligned} [F] &= [f_1, f_2, \quad , f_N] \\ [X] &= [x_1, x_2, \quad , x_N], \end{aligned} \tag{2}$$

where $[X]$ is the vector of unknown variables The Jacobian matrix $[J]$ is the "slope" matrix whose element J in the ith row and jth column is the derivative

$$J_{ij} = \partial F_i / \partial x_j \tag{3}$$

At the nth iteration the Taylor series expansion for $[F]^{(n)}$ around $[X]^{(n)}$ gives the following matrix equation

$$[F]^{(n+1)} - [F]^{(n)} = \left[\frac{\partial F}{\partial x}\right]^{(n)} \{[X]^{(n+1)} - [X]^{(n)}\}$$

$$+ 0\,[\Delta X]^2$$

By truncating the series to drop terms of the order ΔX^2 and higher and evaluating the derivative matrix at the nth iteration

$$[F]^{(n+1)} - [F]^{(n)} \cong [J]^{(n)} \{[X]^{(n+1)} - [X]^{(n)}\} \tag{4}$$

or

$$[X]^{(n+1)} \cong [X]^{(n)} + [J]^{-1(n)} \{[F]^{(n+1)} - [F]^{(n)}\} \tag{5}$$

At an $(n + 1)$th iteration, which is assumed to bring the trial $[X]^{(n+1)}$ close to the root $[X^*]$, the function $F^{(n+1)}$ approaches zero and the non-zero residual ($[F]^{(n)}$) is used to obtain the updated value of $[X]^n$ Equation (5) then gives

$$[X]^{(n+1)} \cong [X]^{(n)} - [J]^{-1(n)} [F]^{(n)} \tag{6}$$

Note that if $[X]^{(n)} \neq [X^*]$, the function set $[F]^{(n)}$ will not equal zero, rather, it equals a "residual" which tends to be close to zero as the iteration proceeds Instead of inverting the Jacobian matrix, a modified Gauss-Jordan elimination technique is used to solve Eq (6)

Convergence is accepted when the condition

$$| [X]^{(n+1)} - [X]^{(n)} | \leqslant [\epsilon] \tag{7}$$

is met or when n exceeds a user-specified upper limit The value of the variable $[X]$, computed at the end of one time step, is used as the initial guess for the iteration at the next time step and the above process is repeated until the simulation period is completed

Note that the variables sought at the end of each time step can be classified as (a) "equilibrium" variables that are free to take on any value during the time step and therefore they are determined by the program after satisfying all conservation

laws, and (b) externally controlled (or source/sink) variables that must be kept at a fixed value during each iteration step as determined by the user's input controls The structure of the Jacobian matrix includes only the first type of variables

C. Equations and Unknowns in a Fluid Network

The set of nonlinear equations needed to describe a fluid network can be obtained by characterizing each fluid station in the fluid network by a vector $[V]$ which is composed of seven elements

$$[V] = [\dot{m}, P, h, T, v, s, x] \tag{8}$$

where $\dot{m}$ is the fluid mass flow rate, P is the absolute pressure, h is the specific enthalpy, T is the absolute temperature, v is the specific volume, s is the specific entropy, and x is the quality (or dryness fraction) These seven fluid properties constitute $7S$ unknowns for a fluid network that consists of S number of fluid stations Energy is transferred across the boundary of each component in the network in three possible forms heat, $\dot{Q}$, electrical work, $\dot{W}_e$, and mechanical (or shaft) work, $\dot{W}_m$, thus adding $3C$ unknowns for C components The $(7S + 3C)$ equations needed to solve these $(7S + 3C)$ unknowns are grouped as follows

(1) For each fluid station where the fluid is a pure substance, only two independent thermodynamic properties are needed out of the six properties p, h, T, v, s, and x The thermodynamics-property subroutines are designed to compute the remaining four thermodynamic properties once two properties are specified The property subroutines thus act as $4S$ implicit equations for a network of S fluid stations

(2) The continuity and momentum equations form a set of $2(S - 1)$ equations for S fluid stations in a closed fluid loop These equations are written along streamlines only The number of equations depends upon the structure of the fluid network, two equations for each streamline minus two equations for closing the loop

(3) If the first law of thermodynamics is applied to each component boundary or control volume, a maximum set of C equations for C components can be obtained Additional component equations (equals $2C$) can be obtained from the second law of thermodynamics for the efficiency or effectiveness expressions and from the energy constraints for each component (such as no $\dot{Q}$, no $\dot{W}_e$, or no $\dot{W}_m$) A total of $3C$ equations are thus obtained

(4) The remaining $(S + 2)$ equations can be written to control the properties at certain fluid stations by assigning values to temperature limits (for two-phase flows), pressure limits, quality limits, or mass flow limits Control equations can be used to define load management as needed by the control software

D. Structure of the Jacobian Matrix and Component Equations

Although the Jacobian matrix $[J]$ could take on various arrangements, the one sketched in Fig 2 is one option The elements $(\partial F_i / \partial x_j)$ are grouped so that the seven variables of the first fluid station are followed by the variables of the second station and so on until the Sth station is reached Next to the station property variables, the derivatives of three variables (Q, W_e, W_m) of the first component are placed, followed by the three variables of the second component and so on until the Cth component is completed Other arrangements can be adopted if found more advantageous

Only components with a time constant much smaller than the iteration time step are included in the Jacobian structure since these components are treated as having Steady-State, Steady-Flow (SSSF) processes Any capacitive component with a large time constant (such as a fluid storage tank) will be dealt with separately at the end of each time step, where the component is treated as having Uniform-State, Uniform-Flow (USUF) processes

To avoid having discontinuous function derivatives, especially for the $4S$ implicit thermodynamic relations, the SPECS program includes a special treatment explained as follows Once two thermodynamic properties (p and v, say) are known at the nth iteration step, the other corresponding four properties (in this case h, s, T and x) are determined by the thermodynamics property subroutine These four properties are treated as "fixed" or constants throughout the $(n + 1)$th iteration step At the end of the $(n + 1)$th iteration, new values for the two properties (p, v) are obtained and the process is repeated

If a fluid property y is assigned a temporary constant value y_0, an equation can be written to reflect this process

$$y - y_0 = 0 \tag{9}$$

Accordingly, the element $J(i, j)$ of the Jacobian matrix corresponding to ith equation and the jth fluid property will be equal to 1 and all the other elements in the ith row become zero If desired, these fluid properties, which are assumed to be temporarily fixed, could be excluded from the set of unknown equilibrium variables in the Jacobian, hence reducing the size of the Jacobian matrix Additional tests, however, are

still required to determine whether these fixed variables in equations similar to Eq (9) may save computation time

As an illustration of the structure of the Jacobian matrix, consider the continuity, momentum and energy equations written for a compound stationary component (σ) with multiple inlet stations (i) and outlet stations (o) as sketched in Fig 3

For USUF conditions, the continuity equation becomes

$$\sum_i \dot{m}_i(t) - \sum_o \dot{m}_o(t) = \frac{m_\sigma(t+\Delta t) - m_\sigma(t)}{\Delta t} \quad (10)$$

The momentum - impulse vector equation at time t becomes

$$\sum_i \dot{m}_i [U_i] + \sum_i [P_i A_i] - \sum_o [P_o A_o] + [m_\sigma g] + [R] = \sum_o \dot{m}_o [U_o] \quad (11)$$

where [ ] indicates a vector of multiple components $[R]$ is a boundary-surface force which could be negative (as a friction force opposing the flow) or positive (as a result of forced agitation by a blower or a pump)

The energy equation is written as

$$\dot{Q} + \sum_i \dot{m}_i \left(h_i + \frac{U_i^2}{2} + gZ_i\right) - \dot{W}_m - \dot{W}_e - \sum_o \dot{m}_o \left(h_o + \frac{U_o^2}{2} + gZ_o\right) = \frac{E_\sigma(t+\Delta t) - E_\sigma(t)}{\Delta t} \quad (12)$$

where E_σ and m_σ are the internal energy and retained mass of the control volume σ, and t is the time Equations (10 through 12) are used for USUF components only The velocity U could be further expressed in terms of the mass flow rate as

$$U_i = v_i \dot{m}_i / A_i$$

$$U_o = v_o m_o / A_o$$

For SSSF components which appear in the Jacobian matrix, Eqs (10) through (12) are reduced to the following

Continuity equation

$$\sum_i \dot{m}_i - \sum_o \dot{m}_o = 0 \quad (13)$$

Momentum equation for a streamline

$$P_i - P_o - (a_0 + a_1 \dot{m} + a_2 \dot{m}^2) = 0 \quad (14)$$

where the pressure change (R/A) is expressed by a polynomial in $\dot{m}$, and a_0, a_1, a_2 are constants determined by the component type Energy equation for a SSSF component or a control volume

$$\dot{Q} + \Sigma_i m_i \left(h_i + \frac{U_i^2}{2} + gZ_i\right) - \dot{W}_m - \dot{W}_e - \Sigma_o \dot{m}_o \left(h_o + \frac{U_o^2}{2} + gZ_0\right) = 0 \quad (15)$$

The coefficients derivatives of Eq (13) appear in the Jacobian matrix as 1, corresponding to an inlet station mass flow rate, -1 to the outlet station mass flow rate, and zero otherwise Other simple relations can be derived from Eqs (14) and (15) in a similar manner

IV. Program Structure

The SPECS program is structured in a modular fashion which will allow future modification of any aspect of the program It consists of four blocks (1) an input block, (2) a library block, (3) a control block, and (4) an output block as shown in Fig 4 The input block contains the user-provided data The library block contains the subroutines which perform various analytical and numerical computations The library block computes all parameters for a given system configuration according to the data and specifications given in the input block The control block monitors and manages the flow of information, solves the system of nonlinear equations at each time step, and controls the computational stability requirements The output block prints for the user the final performance and economics results of a specific system Note that the output block is controlled by user options for short, intermediate, or detailed printout

The program language is FORTRAN IV Use of FORTRAN "NAMELIST" is made to classify the input data into groups of common feature Descriptions of program blocks and subroutines are as follows

A. Input Block

The input to the SPECS program is divided into seven major groups

(1) System configuration data

(2) Station properties data

(3) Components characteristics data

(4) Control data

(5) Outdoor environment/site data

(6) Energy loads data

(7) Economics data

Each group is further described below

1 System configuration data This group defines the number, the type and location of components, the number and location of inlet and outlet fluid stations, the number of streamlines, etc This information is used by the program to set limits of DO loops and determine the size of the problem under study

2 Station properties data This group includes the type of working fluid at each station, externally controlled properties, initial conditions and setpoints All the pressures are computed relative to a user-assigned pressure at a reference point The program also allows for differences in elevation of each station above some arbitrary datum

3 Component characteristics data This group includes the diameter and equivalent friction length for pipes and joints, the head-flow rate curve for a pump, the heat transfer coefficient times surface area product for a heat exchanger, the coefficient of velocity for a fully open valve, etc

4 General control data This group includes the case title, convergence criteria for termination of program, the time-step, the total simulation period, the type of results to be printed out for the user and options to read weather and load data

5 Outdoor environment/site data This group includes outside air dry-bulb temperature, outside air wet-bulb temperature, relative humidity, dewpoint temperature, wind speed and direction, site latitude, site elevation, and solar radiation components (direct normal, diffuse, and ground reflected) as measured by a pyrometer or tracking pyrheliometer A 24-hour profile for one representative day of each month is selected throughout a year's simulation The solar radiation data can be taken from field measurements or otherwise simulated using the deterministic ASHRAE model (Ref 12) which is built into the program

6 Energy loads data This group includes space and water heating load, space cooling load, lighting load and other electrical energy loads consumed by the facility Water consumption data can also be included Since it is impractical to run a large model for each hour of the 8760 hours of a year, an appropriate compactness is required to save computation time Therefore, loads data for two representative days for each month are satisfactory, the first is a load profile for weekdays and the second is a load profile for weekends and holidays Twenty-four hourly values for each representative day are required for energy loads

7 Economics data For detailed life cycle cost and cash flow analyses, this group includes the installation, operation, and maintenance costs of all components in the system, purchased energy costs, energy escalation rate, discount rate, and salvage values, in addition to preimplementation costs such as site preparation, surveying, study and design costs, etc

B. Library Block

This block contains five sets of subroutines

(1) Thermodynamic properties subroutines

(2) Component description subroutines

(3) Gauss-Jordan subroutine

(4) Solar-radiation simulation subroutine.

(5) Life cycle cost and cash flow subroutine

Each set is described as follows

1 Thermodynamic properties subroutines The driver routine for the thermodynamic properties computation is called by the Control Block to obtain values of fluid properties at any fluid station This routine forms, from a library file, an array containing tables of saturation data only for each fluid used in the network being analyzed The array does not contain data for all 16 fluids, listed in Table 1, which SPECS recognizes This subroutine computes the values of four properties (p, h, s and x), given the values of the two independent properties (T, v) for a fluid according to the flow chart in Fig 5

If the fluid is in the 2-phase saturated region ($v_f < v < v_g$) where the subscripts f and g denote saturated liquid and

saturated vapor, respectively, the driver routine computes the values of p, h, s, and x by interpolating according to T at constant v If the fluid is in the compressed region ($v < v_f$), the four properties are calculated based on approximations using the saturated liquid conditions

If the fluid is in the superheated region ($v > v_g$), general analytical expressions from Ref 13 are used to calculate p, h, s, and x as a function of T and v A set of fluid-specific subroutines calculates the two ideal gas properties (h^*, s^*) followed by the generalized expressions to compute the departure from ideal gas properties

After the computation of all four properties is complete, the values are returned to the Control Block Thus this subroutine acts as $4S$ implicit equations for a system having S number of fluid stations

In addition to the six thermodynamic properties analyzed in this subroutine, three other physical properties are stored for each of the 16 fluids the thermal conductivity, $k(T)$, dynamic viscosity, $\mu(T)$, and specific heat, $C_p(T)$ These three physical properties are given for both liquid and superheated vapor forms

2 **Component description subroutines** This set is a library of component model subroutines from which the user can draw to configure the specific system under study The components are divided into generic categories according to function Mass, momentum, energy conservation equations, and performance effectiveness relations can be written regardless of differences in design, operating conditions, and working fluids These differences are taken into account by assigning different values for similar parameters in each component This process eliminates superfluous component models of similar type

For example, a counterflow recuperator, a parallel flow recuperator, and a heat regenerator could be simply grouped under one component type a heat exchanger in which heat transfers from one medium to another The distinct behavior of each component, however, is accounted for by different analytical expressions for the effectiveness, heat capacities of exchanging media, number of heat transfer units, and the conductance heat transfer area product

Distinction between mechanical components is made based on the number of streamlines and inlet and outlet stations as shown in Fig 1 Electrical components are modeled by simple relations between $\dot{W}_m$, $\dot{W}_e$, and $\dot{Q}$ Some of the generic component types to be modeled include but are not limited to the following

(1) A heat exchanger (boiler, recuperator, and regenerator with phase change in one or two media)

(2) A fluid pump (rotary, screw, and positive displacement)

(3) A compressor (for air, refrigerant, etc)

(4) A turbine (for steam, organic fluid, gas, wind, water)

(5) A simple heater (for air or water) or a simple cooler without phase change of either medium

(6) A solar collector (flat plate, parabolic trough, paraboloid concentrator, etc)

(7) A fluid storage tank (compressed air, water, steam accumulator, etc)

(8) Electrical battery

(9) Electrical motor

(10) Electrical generator or alternator

(11) Photovoltaic cell (flat plate or concentrated type)

(12) A mixing valve

(13) A splitting valve

(14) A pressure reducing valve

(15) A fluid pipe (insulated or noninsulated)

(16) An underground cable or electrical transmission line

(17) Building structure

(18) Air-handler for air conditioning (dual duct, variable air volume type, etc)

(19) A thermoelectric cell for power generation or refrigeration

(20) A thermionic cell for power generation

(21) A fuel cell

(22) An absorption chiller (using lithium bromide-water, aqua-ammonia, etc)

3 **Newton-Raphson iteration subroutine** This subroutine solves the system of nonlinear equations using the Gauss-Jordan reduction method with diagonal pivoting instead of directly inverting the Jacobian matrix This subroutine is written in double precision arithmetic to overcome the problem of oscillatory convergence during iteration

4 **Solar radiation simulation subroutine** Based on the deterministic model set by ASHRAE (Ref 12), the total solar

radiation intensity (including the direct, diffuse, and ground-reflected components) falling on a surface is analytically simulated for the user's convenience The solar elevation, azimuth, and declination angles are computed for 24 hours for one representative day of each month The information available in this subroutine pertains to various solar collector types, selected as follows

(1) Tilted flat plate, with arbitrary tilt angles

(2) Horizontal flat plate

(3) Single-axis tracking surface, with east-west orientation

(4) Single-axis tracking surface with north-south orientation

(5) Single-axis tracking surface with polar mount, i e , the surface is oriented toward north-south line with a tilt angle to the horizontal which equals latitude

(6) Two-axis tracking surface, always perpendicular to the sun's direct rays

For northern latitudes, the surfaces are oriented toward the south, and for southern latitudes, toward the north Additional surface configurations can be added in this subroutine provided they follow the SPECS format

5 **Life cycle cost subroutine** This subroutine calculates, for the system under study, the yearly maintenance and operation costs and the life cycle cost and presents a detailed cash-flow analysis The methodology used in the subroutines follows the National Bureau of Standards' economic criteria as outlined in Ref 14

C. Control Block

This is the main routine which executes, monitors, controls and drives the whole program as sketched in Fig 6 It enters the input data and sends the desired output The main routine calls the thermodynamic properties subroutines in order to determine the properties of flow for each station It also calls the set of component description subroutines to compute equation residuals and partial derivatives, which compose the Jacobian matrix, and executes the Newton-Raphson software to solve the system of nonlinear equations

As in a real system, the control block is thought of as receiving input signals at the end of each iteration step, represented in the values of the variables (such as temperature, pressure, mass flow rate, etc) It compares these input signals to one another or to set points and transmits the resulting "logical" signal to a logic element where a transformation occurs to generate output signals to actuators (represented by a set of analytical expressions or tabular data) for the operation of dampers, valves, pumps, boilers, etc A "truth-table" is another way of representing the function of the controller-actuator in handling the input/output signals in the simulation

D. Output Block

The output of the program is controlled by several input variables The computation results consist of the value of the different thermal and electrical loads at that particular hour, the thermo-dynamic properties at each station, the energy flux across each component of the system, and the overall performance, efficiency, and economics of the system Several user options are introduced to obtain the desired level of printout detail

V. Simulation Steps

The philosophy of the program logic can be visualized as an attempt to perform in a stepwise manner what the real system does in a continuous mode of operation following a startup To simulate a particular system configuration, the major steps of program operation are as follows

Step 1 (user's input data) After all input data are entered into the program, the user can "echo" or print-back all input data for rechecking

Step 2 The dynamic simulation starts by the control block which groups the built-in characteristic equations of the components of the system being simulated Because the Newton-Raphson method is iterative, it is essential to choose initial trial values of the variables to start iteration at each time step Fluid valves can be initially set at their half-open position, and continuously adjusted as iteration proceeds At the end of one time step, the results are used as the initial guess for iteration at the following time step, and so on The above process is repeated until the simulation period has elapsed

Step 3 The check for convergence is based on the changes occurring between one iteration step and the preceding step If the changes in all the variables are less than a specified tolerance, $[\epsilon]$, the computation is considered convergent and the program proceeds to Step 5, otherwise the program proceeds to Step 4 for another iteration

Step 4 Revise the initial setting of temperature and pressure controllers, valves, set points, etc Diagnostic messages are printed if the variables attain unreasonable values or if convergence is not attained after a prescribed number of iterations The program returns to Step 2 with updated values of the parameters

Step 5 The program now prints out complete energy and flow rate profiles for the whole system under daily, monthly, and yearly operation Detailed life-cycle cost and cash flow analyses are made over a given number of years

Step 6 Additional optimization and parameter sensitivity will result from employing a search-type optimization technique at different design or operating conditions by repeating Steps 1 through 5

VI. Program Summary

The SPECS program in its present form is structured to follow the above block descriptions Current activities relate to the following

(1) Incorporating a detailed economics package for the system life cycle costs and cash flow analysis

(2) Improving and expanding the tabulated results in the output block.

(3) Expanding the components subroutines in the library block to include all the candidate components in Section IV-B-2

(4) Debugging, testing, and documenting the final version of the program in order to extend its use to other design projects

In summary, the SPECS model was developed to suit the following purposes

(1) To be used as an analytical tool for equipment design for large energy conversion systems where the cost of computer simulation represents a small portion of the total cost

(2) After completing a trial equipment selection based on one set of loadings, the designer can apply the model to other partial or overloading conditions to determine if the original design is adequate

(3) To serve the basis for developing "short cut" techniques after making a large number of detailed simulations on many different designs

(4) To provide the field operator with diagnostic aid for understanding the dynamics of the operating system and the relationships between components at off-design conditions

(5) To study the effects of add-on components and analyze the corrections sought for some operating problems if an existing system is expanded or reconfigured

The article has described these aspects of the SPECS computer program the method used for modeling a system of components, the mathematical representation of these components, the numerical technique used in solving the resulting system of equations, and the program input and output

References

1 Potter, A G , "Solar Assisted Power Systems," *Proceedings of the Second Annual UMR-MEC Conference on Energy*, University of Missouri, Columbia, Mo , Oct 1975 Published by Western Periodical, Hollywood, Calif , 1976 pp. 47-56

2 Stoecker, W F , "Methodology of Simulating the Performance of a Central-Chilled Water System," Paper presented at ASHRAE Meeting No 2458, Research Project 141

3 Nguyen, K H , "Solar Heating System Simulation Approaches Small Time Model Versus Large Time Model," Paper presented at the System Simulation and Economic Analysis for Solar Heating and Cooling Conference, San Diego, Calif , June 1978 (available from NTIS).

4 Klein, S A , et al , *TRNSYS, A Transient-Simulation Program*, Report 38-10, Solar Energy Laboratory, University of Wisconsin, Madision, Wis , June 1979

5 Edenburn, M W , and Grandjean, N R , "SOLSYS – Energy System Simulation Computer Program," SAND75-0048, Sandia Laboratories, Albuquerque, N M , June 1975

6 Fewell, M E , and Grandjean, N R , "User's Manual for Computer Code SOLTES-1 (Simulation of Large Energy Systems," SAND78-1315, Sandia Laboratories, Albuquerque, N M , June 1979

7 Murray, H S , and Kuckertz, T H , "Dynamic Simulation of a Solar Heated and Cooled Building," Paper presented at Summer Computer Simulation Conference, San Francisco, Calif , July 1975

8 JPL Electric Power System Simulator (EPSS) Display Panel, Publication 400-95/1, Jet Propulsion Laboratory, Pasadena, Calif , Nov 1980

9 Pejsa, J H , "The Dynamic Solar Heating and Control System Model and Simulation A New Design Tool," Paper presented at the System Simulation and Economic Analysis for Solar Heating and Cooling Conference, San Diego, Calif , June 1978 (available from NTIS)

10 Johnson, G R , and Elgabalawi, N , "Dynamic Simulations – Solar Thermal Electric Power Systems," Final Report NSF/RANN/SE/GI-37815/FR/74/3, Vol 3, Colorado State University, Fort Collins, Colo , pp J1-J59, 1974

11 Freeman, T L , Maybaum, M W , and Chandra, S , "A Comparison of Four Solar System Simulation Programs in Solving a Solar Heating Problem," Paper presented at the Systems Simulation and Economic Analysis for Solar Heating and Cooling Conference, San Diego, Calif , June 1978 (available from NTIS)

12 *Handbook of Fundamentals*, American Society for Heating, Refrigeration, and Air Conditioning Engineers, 1972

13 Reynolds, W C , "Thermodynamic Properties in SI, Graphs, Tables and Computational Equations for Forty Substances," Department of Mechanical Engineering, Stanford University, Stanford, Calif , Jan 1979

14 Hughes, R. D , "A Computerized Life Cycle Methodology for Engineering Analysis," *Telecommunications and Data Acquisition Progress Report 42-66*, Jet Propulsion Laboratory, Pasadena, Calif , pp 268-287, Dec 1981

Table 1. List of fluids used in SPECS

Type No	Fluid	Chemical symbol
1	Water	H_2O
2	Ammonia	NH_3
3	Helium	He
4	Hydrogen	H_2
5	Refrigerant 11 (trichlorofluoromethane)	CCl_3F
6	Refrigerant 12 (dichlorodifluoromethane)	CCl_2F_2
7	Refrigerant 13 (chlorotrifluoromethane)	$CClF_3$
8	Refrigerant 14 (carbontetrafluoride)	CF_4
9	Refrigerant 22 (chlorodifluoromethane)	$CHClF_2$
10	Refrigerant 23 (trifluoromethane)	CHF_3
11	Refrigerant 114 (dichlorotetrafluoroethane)	$CClF_2CClF_2$
12	Refrigerant 318 (octafluorocyclobutane)	C_4F_8
13	Refrigerant 500 (azeotropes (73 8/26 2))	CCl_2F_2/CH_3CHF_2
14	Refrigerant 502 (azeotropes (48 8/51 2))	$CHClF_2/CClF_2CF_3$
15	Refrigerant 503 (azeotropes (40 1/59 9))	$CHF_3/CClF_3$
16	Sodium	Na

List of Symbols

A	cross-section area
c_p	specific heat
C	number of SSSF components
f	function symbol
g	local acceleration of gravity
h	enthalpy
J	Jacobian matrix
k	thermal conductivity
l	streamline index
m	mass flow rate
n	iteration step index
N	number of unknown variables
P	absolute pressure
Q	heat rate
R	boundary force due to friction or forced agitation
s	entropy
S	number of fluid stations
T	absolute temperature
U	velocity
v	specific volume
V	vector of station properties
W	work energy
z	elevation above sea level (or arbitrary datum)
ϵ	error limit for convergence
μ	dynamic viscosity

Subscripts

i	inlet
o	outlet
m	mechanical
e	electrical

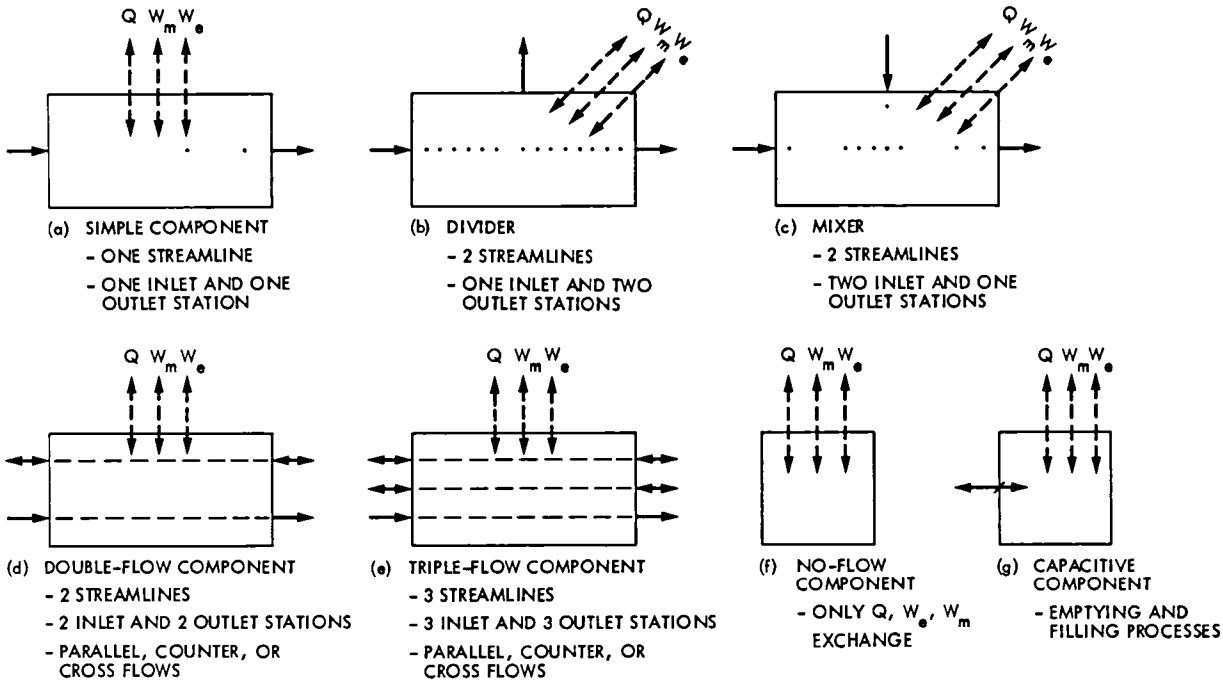


Fig 1 Simple and compound components

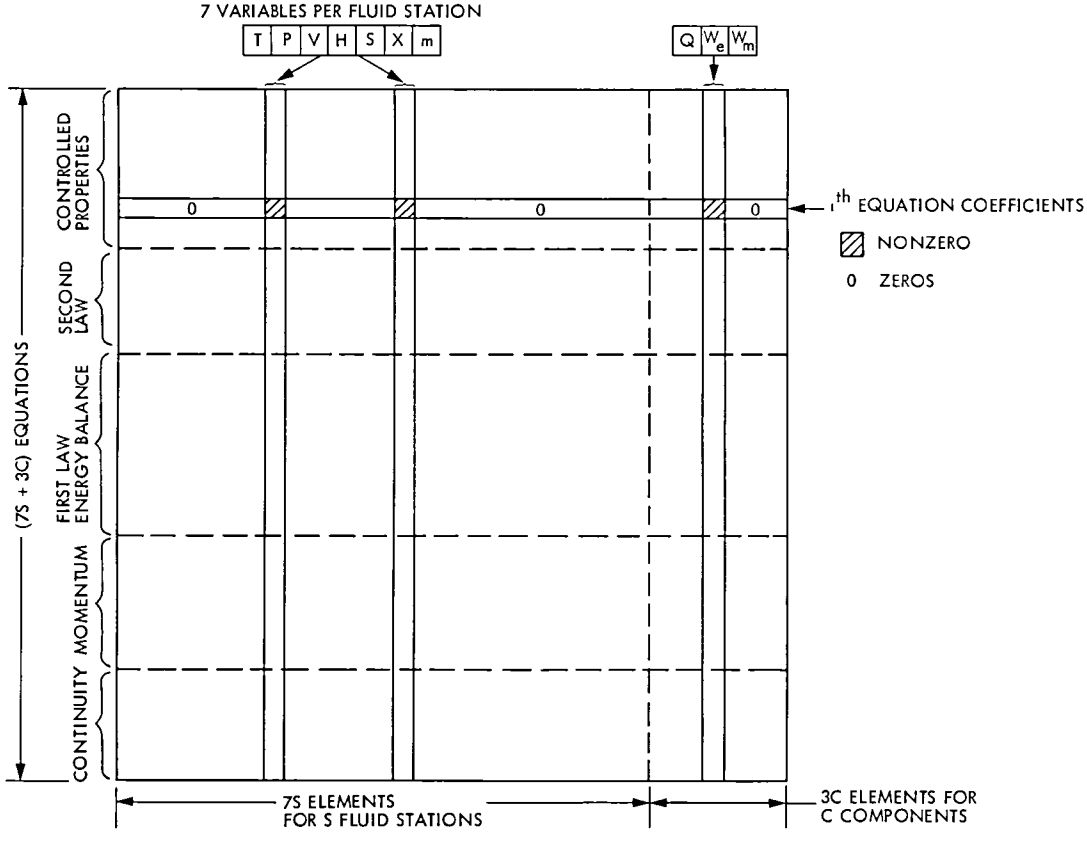


Fig 2 Details of the Jacobian matrix [J]

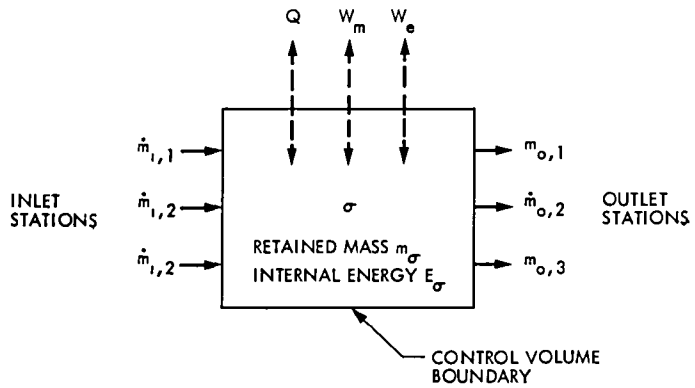


Fig 3 Compound component σ with multiple inlet and outlet ports

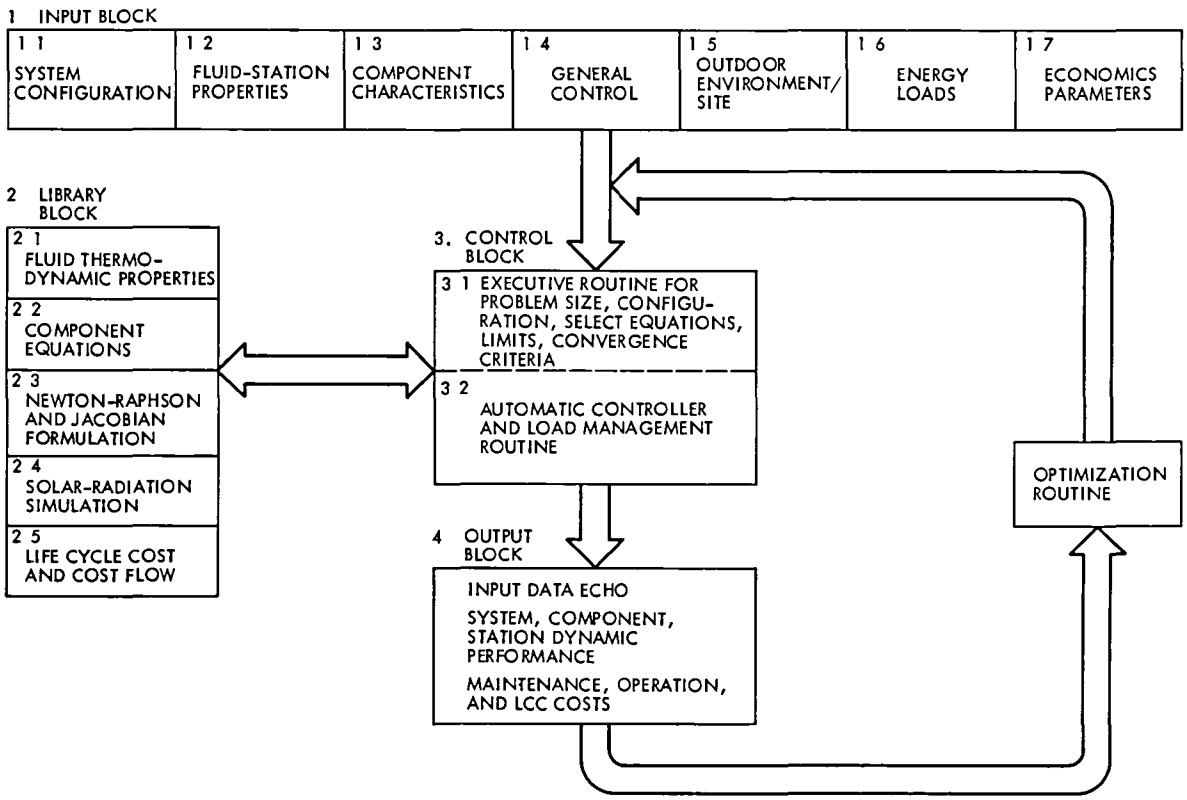


Fig 4 Overall structure of SPECS program

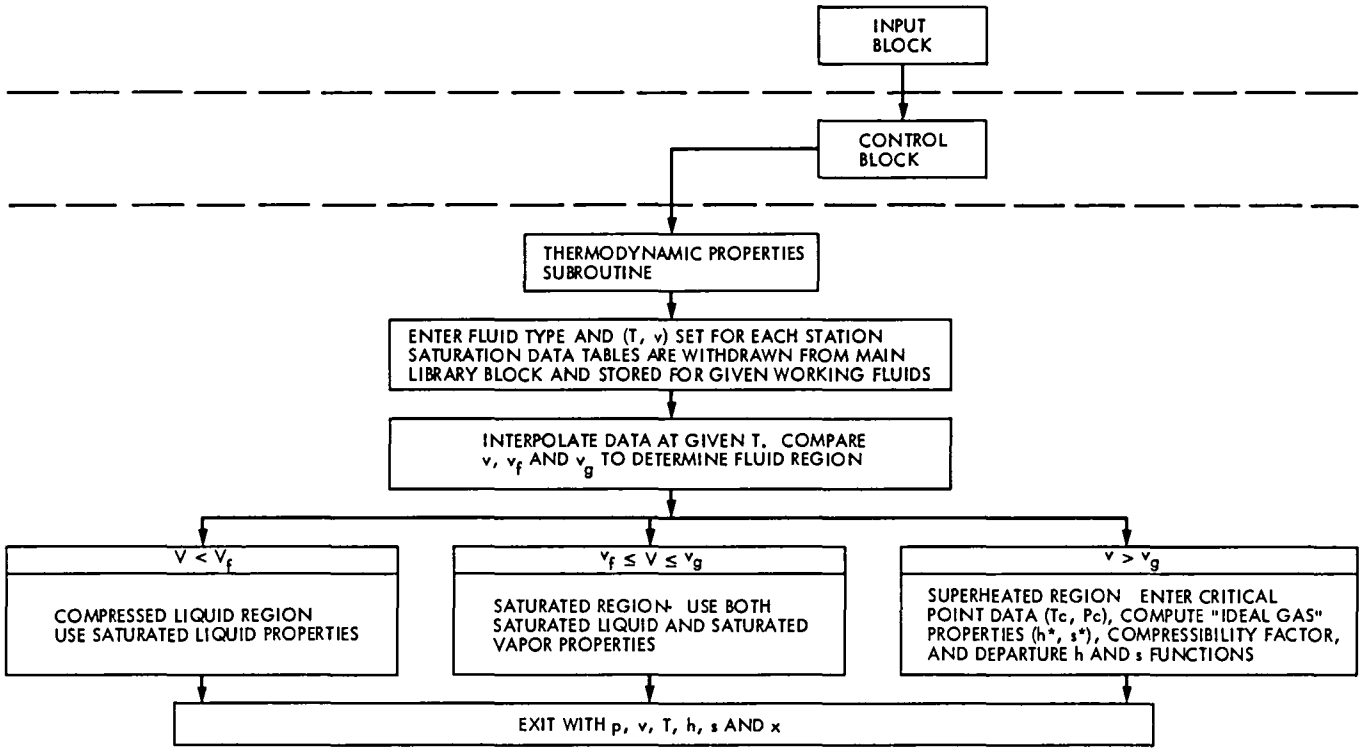


Fig 5 Flow chart for thermodynamic properties subroutine

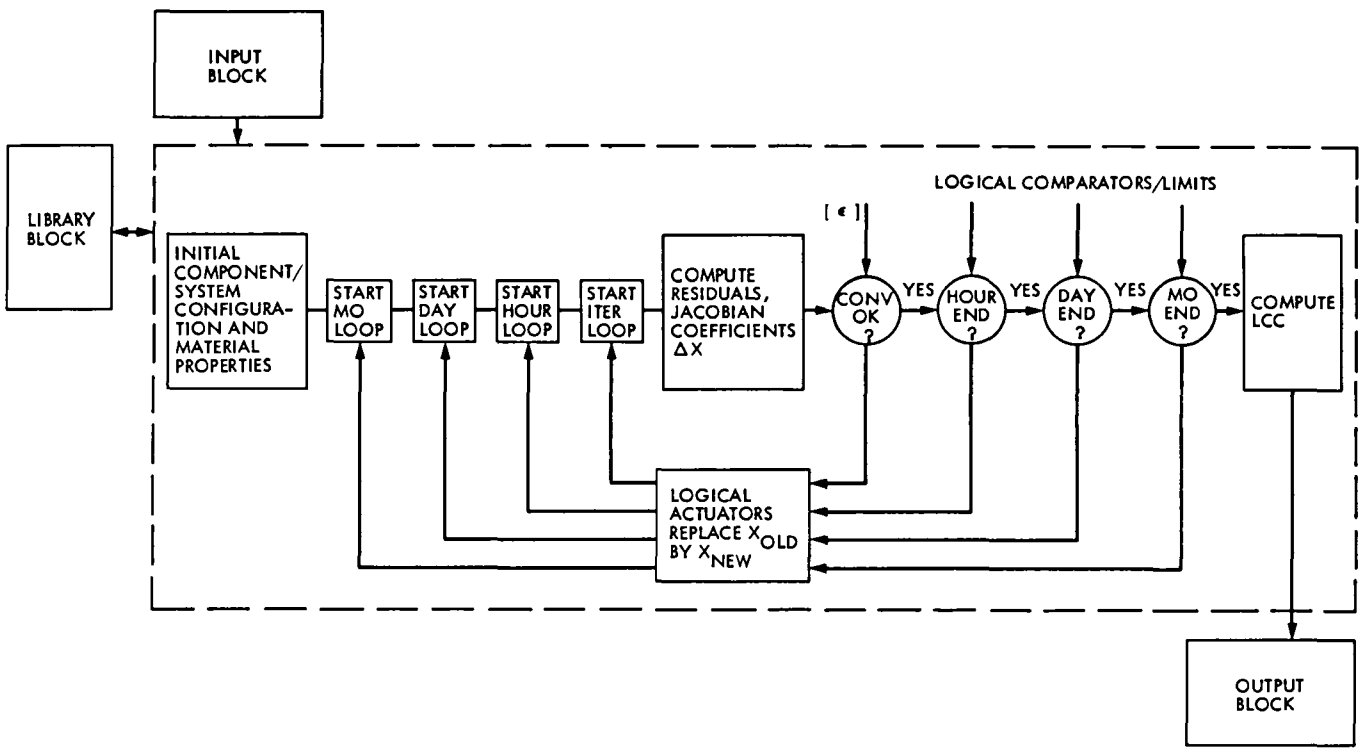


Fig 6 Information flow in control block

The Measurement of Operator Workload in the Mark IVA DSCC Monitor and Control Subsystem

M LeMay
Montclair State College, New Jersey

R L Chafin
Control Center Operations

E. E Hird
Operations Sustaining Engineering Section

An operator workload measurement methodology is presented which will be used in support of the Mark IVA Operational Test and Evaluation (OT&E) Plan Three operator workload measures are suggested operator ratings, primary task work measures, and information processing time measures A method of validating the workload measures using secondary task work measures is presented We can expect that operations testing using these measures will assist in establishing the time required to perform essential operational activities and will indicate high risk operations areas due to potential operator overload

I. Introduction

The goal of the Mark IVA DSN design is to provide a single cost-effective ground tracking and data acquisition network that will be capable of supporting both deep space and highly elliptical earth orbiter missions To achieve this goal, existing tracking stations at Canberra, Goldstone, and Madrid will each be centralized so that the necessary antennas will be collocated near the signal processing equipment, which will be housed within a single building Operating personnel will be reorganized and retrained to the extent necessary for them to be effective in their new roles

Typically each of the present day Mark III DSN or GSTDN tracking stations is operated separately by its own staff The Mark IVA DSN design will reduce the total amount of centralized equipment needed in conjunction with the antenna group to be developed at each 64-meter antenna site either by the relocation of an existing 34-meter antenna or by new construction

A station supervisor in a signal processing center will be able to assign tasks to a control room crew consisting of a Complex Monitor and Control (CMC) operator working together with three Link Monitor and Control (LMC) operators The centralized arrangement of the complex will permit a chosen level of tracking activities to be supported while the requirements for personnel and redundant equipment will have been reduced

The CMC operator will be responsible for allocating assignable equipment resources to a link and for monitoring and

controlling the global equipment that provides services to the complex An LMC operator will be responsible for the acquisition of tracking data during a scheduled project support activity

The LMC operator workload will be offset in the initial implementation by a combination of equipment automation and some additional roving operators who will perform specific tasks on manually operated equipment in the equipment room in response to LMC operator directions

This article reports on a project whose purpose was to develop a method of measuring control room operator workload This methodology is an essential part of the OT&E plan for Mark IVA operations testing In an effort to estimate operator working times on the Mark IVA DSCC Monitor and Control Subsystem (DMC) a series of operations sequence diagrams (OSDs) have been drawn up These are detailed task descriptions with estimates of the amount of time needed to complete each subtask by novice, experienced, and expert operators These estimates are added to obtain a prediction of the amount of time needed for completion of a whole task such as a VLBI clock sync prepass These times will be used as part of an effort to assess the actual workload on individual operators during task performance This is necessary since it is likely that the occurrence of some events, particularly alarms, will result in a sudden, significant increase in workload which may, in turn, increase operator error rate, with an attendant risk of data loss Minimizing the risk in the complex user-computer interface of the DMC is the purpose of the present effort

Although much attention in recent literature has been given to the user-computer interface (Refs 1, 2, and 3), most of it has not been aimed at gauging the effect of an increase in workload in an on-line, interactive system, with the possible exception of studies of pilot and aircrew performance, and this is a somewhat unique case of a system which is only partly, though increasingly, an information processing system Moreover, most of the studies cited in an extensive search (Ref 3) have been done in existing systems The present problem is to assess workload on a system which is still in the design stage and to predict its effect on system performance The measures developed to do this can be implemented in an operational test, which will show where human operator overload difficulties are likely to arise

There is extensive literature on the measurement of workload (Refs 3 and 4), and many possible measures have been proposed These have been classified into four general categories (Ref 3) operator ratings, spare mental capacity, primary task performance, and physiological measures

A method of evaluating workload assessment measures based on this classification, also presented in Ref 3, was applied to the operator's tasks as outlined in the OSDs, and a combination of operator ratings and primary task performance measures was chosen, with a spare mental capacity technique being used to appraise the validity of the measures Physiological measures were not considered because they are expensive to implement, may be intrusive on the operator, and are not practical for ordinary on-line application Some characteristics of the chosen measures are considered as follows

A. Operator Ratings

Previous studies have shown this to be a reliable measure which is consistently related to performance (Refs 5 and 6) It is relatively easy to obtain, and has the "face validity" of getting the operator's own estimate of the subjective phenomenon of mental workload With careful attention to design (Refs 7 and 8), rating scales yield quantitative results which may be used to gauge "crossover" points beyond which the chance for error may increase greatly or other system parameters may be affected

B. Primary Task Performance

This is perhaps the most direct way to assess workload, since it involves the measurement of the effect of workload on the performance of the task of interest The time needed to start and/or complete the task is the most relevant (Refs 8 and 9), although a number of errors may also be considered These measures must be carefully designed to reflect increasing workload, however, since human operators tend to adapt to the task and hold their performance constant over a broad range of conditions Thus, operators' initial starting times for a particular subtask are generally taken as a measure of their strategy for dealing with increased load The starting time reflects the speed at which incoming subtasks are processed mentally (Ref 10), and is a linear function of the amount of information processed by the operator

Additionally, successful completion of a particular subtask or set of tasks in a stated amount of time may itself be considered a measure of workload (Ref 11), and putting increasingly stringent time requirements on the operator until task completion becomes impossible may reveal operator strategies as well as workload limits for particular subtasks Therefore, both a measure of time to start (reaction time) and time to complete a subtask will be used to assess operator workload

C. Secondary Task Measures

Since human operators tend to hold their task performance constant under increasing workload, primary task performance measures may change very little in the normal course of opera-

tion If, however, a secondary task is introduced, so that the operator is required to divide attention between two tasks, then the secondary task may be considered a measure of "spare mental capacity" (Ref 4) and, as the demands of the primary task increase, performance on the secondary task decreases, so it is an indirect measure of operator workload This is the standard way of using a secondary task in workload measurement

For on-line, operational measurement, the imposition of a secondary task is impractical, since it could interfere with normal operations However, a variation of the secondary task technique can be used to validate the primary task measures in a trial or pilot study Such a study is necessary in order to demonstrate that the workload measures used do, in fact, increase with increasing objective workload and that this in turn results in performance degradation and possible data loss

For this purpose, the secondary task should be similar to tasks that would normally be performed by the operator, and it should be possible to make the task more or less time consuming Log keeping in writing and by voice is such a task It, in turn, will affect primary task performance as its requirements result in increased workload (Ref 9) Then the primary task workload measures should rise, and this will result in a decrement in system performance

Three specific ways of implementing rating and primary task workload measures are described below, along with a secondary task technique for validating the measures

II. The Measures

Three measures are proposed for use in gauging operator workload in the DMC They are interrelated, and should be collected simultaneously on a single task They are

(1) Sequential operator ratings of workload on subtasks

(2) The ratio of the time required to the time available to do a particular subtask

(3) Information processing time, or the time taken to initiate the physical (keystroke) portion of a subtask

The latter two measures are keyed to the operator rating measure, which is taken at the end of each subtask and is recorded with a short free-form input (FFI)

A. Operator Ratings

1 Procedure A single rating of workload, on a scale from 1 to 7 with a 7 meaning the highest workload, will be obtained from each operator during the performance of a task The instructions to the operator (reproduced in Fig 1) are given only once Thereafter, at the completion of each subtask, the words "difficulty rating" and "time pressure rating," will appear in the SOE, and the operator will respond with an FFI text message and a rating of from 1 to 7

A subtask is defined as a listed task with an estimated performance time in an operations sequence diagram (OSD) This will result in the operator's giving a rating at approximately 2-4 minute intervals, depending on operator experience

The operator rating will be recorded on the operations log, along with the time at which it was made This will provide a record of the time at which each subtask is finished (to be used in the other two measures described below), as well as of the operator's rating

2 Analysis The procedure will result in operator ratings on each of a number of subtasks, so that those points in the task at which overload occurs can be quickly located

The operator ratings are obtained from the operations log and classified according to subtask and operator experience Therefore, there should be some means of identifying the particular subtask (e g , "Display and Compare APA Predicts against VLBI Source Table") on the log The classification of operator experience (novice, intermediate, or expert) should be done by an experienced supervisor, with a set of guidelines taking into account the amount of time on the job and previous experience, if any

Then the mean rating can be obtained for each subtask and each operator classification, and plotted as a function of task, resulting in a graph similar to that illustrated in Fig 2 The peaks on such a graph will show the subtasks on which the operator is working at a higher subjective workload In conjunction with the measures described below, this will indicate tasks on which errors might be expected to occur in the event of a sudden increase in workload such as might result from an event notice or alarm

In this type of testing, the differences in ratings obtained for each subtask may be relatively small, the differences among the three types of operators may also be small but of interest In order to test the significance of the differences, the rating data should be subjected to an analysis of variance (ANOVA), which shows whether or not there are real differences among tasks and/or operators, and whether there is an interaction between them, i e , whether some tasks are perceived as having a higher workload by novice operators, and others by experienced operators, etc

In addition, after the completion of a task, such as a Delta DOR Prepass operation or 64-meter antenna configuration,

operators will rate several aspects of the job as a whole on a form like the one illustrated in Fig 3 This is a very simple form, and can be analyzed easily, whether it is done on the log or by hand Mean ratings on each of the first 10 questions should be obtained for each operator class These should be examined individually as indicative of problem sources perceived by operators The open-ended question 11 should also be used in the same way The questionnaire will complement the subtask ratings by providing a general operator rating of workload in question 1 This rating should be correlated with the mean of the subtask ratings for each subject (i e , x_i = mean of subtask ratings for S_i, Y_i = whole task workload rating from question 1 for S_i) It will also serve as an indication of problems perceived by operators in particular subsystems, as opposed to the problems of particular subtasks pointed out in the repeated workload ratings obtained after each subtask

B. Primary Task Workload Measure

1. Procedure This measure will be similar to standard system reliability workload measures, in that it will consist of a ratio of time required to time available Thus, workload is defined as

$$W = T_r/T_a$$

where

W = workload

T_r = time required by the operator to perform a specific subtask

T_a = time estimated in the OSDs for that particular subtask

Since we are estimating mental workload, there may be no obvious action recorded on the operations log to signal the start and end of a subtask Therefore, the operator will be asked to signal the start of each task with a FFI which will be used to time that task on the log

2. Analysis The time between the FFI messages which indicate the start of a subtask and the operator's rating of that task can be taken as the actual or required time to perform the subtask This time, T_r, will be available in printout form along with the operator ratings It can then be compared with the times estimated in the OSDs for that subtask

If subtasks are always performed in the proper sequence, particular tasks can be identified on this basis However, an identifying code for each subtask, either a mnemonic or a numeral, would insure that the proper task was being considered with its time and rating, and would aid data analysis

The resulting workload measure for each subtask will be analyzed in a fashion similar to that used for the ratings Mean workload will be plotted as a function of task, with separate curves for each operator classification, and an analysis of variance performed on the data to determine if different subtasks yield significantly different workloads In addition, the correlation between the workload measure and the operator rating for each subtask should be obtained

If the obtained mean workload for a particular task is greater than 1, then the time needed for the task is greater than that originally allowed in the estimate If it is above 0 7 or 0 8, it is approaching the operator's maximum capacity In either case, if the estimates are to be considered as the time allowed for the task, then some provision must be made to relieve operator workload on these subtasks, perhaps by task rearrangement, extra help from personnel or computer, or simply increasing the time allowed Otherwise, there will be an increased risk of operator error occasioned by sudden increases in workload, such as the occurrence of an event or alarm requiring operator action during an already overloaded subtask

C. Information Processing Time

1 Procedure The third value to be used does not measure workload directly, but is an attempt to measure the time required for mental processing of the information needed to perform a control function

Many of the subtasks in the OSDs require only monitoring functions, in this case, since there is no overt action, but only information processing (e g , comparing source time and ID against VLBI source table for compatibility), this information processing or "think" time will be the same as T_r (time required) in the second measure However, on most subtasks, an overt response is required The response may be part of the normal SOE or be contingent on the detection of some anomaly such as a high system temperature value However, the majority of tasks do require some sort of physical action, and the time to start such action is called a choice reaction time (CRT), since the operator must choose the action to be taken from among a number of possible actions It varies directly with the number of choices or actions, and is a measure of uncertainty, or the amount of information the operator must process The longer the CRT, the more complex the information processing task and the greater the workload Thus, information processing time, as measured by CRT, is also a measure of workload

One task follows another immediately in the OSDs However, this may not always be the case in practice, since there may be interruptions or operator recovery time from the previous task It is reasonable to include this extra time in the

direct workload measure (the second measure), since it becomes part of the workload and adds to it However, it is obviously not part of the information processing time for a particular subtask Therefore, it will be necessary, as for the second measure, for the operator to signal the start of each subtask with an FFI text message which can be used, in conjunction with the next action taken (keystroke or voice communication), in determining the information processing time

2 Analysis The time between the FFI text message which signals the start of a task and the next action is obtained from the operations log for each subtask for each operator The mean of these values is obtained for each class of operators This data is then subjected to the same type of analysis as before, i e , mean information processing time is plotted as a function of task for each class of operators, so that those tasks which carry a heavy information processing load can be identified The data is also subjected to a 3 (operator class) $\times$ n (number of tasks) $\times$ S (subjects) ANOVA, to determine the significance of the observed differences

In addition, when all three measures have been obtained, a grand mean across all three classes of operators should be calculated for each measure on each task If all three tasks are measuring workload, then there should be some degree of correlation among them In fact, since the information processing time is a fraction of the required time in the workload measure, these two will probably be highly correlated This makes the calculation of a multiple correlation coefficient inappropriate However, it is appropriate to calculate the correlation coefficients between the operator rating and each of the other measures These should be relatively high (above 0 5), which would indicate that all three measures are measuring different aspects of the same thing (workload) and would be evidence for the validity of the measures

III. Validation of the Measures

The three proposed measures seem logical, and they are based on measures that have been used for similar purposes in the literature (Refs 3 and 4) Nevertheless, their validity and reliability need to be evaluated for the use in the operational testing It is possible, for example, that one or two of the measures will show no variation, even when operator workload is obviously markedly increased, whereas a minor change in the data collection method would yield usable results This is a common occurrence in experimental trials with human operators The only way to determine whether or not a measure of human performance will reflect variations in external working conditions is to run a "pilot study", i e , try the measure out using a small sample of people performing a relevant task

In addition to demonstrating that the measure varies with task parameters, pilot studies can also be used to assess the effect of increasing workload on performance The purpose of measuring operator workload, after all, is first to detect tasks on which the workload is so high that the task becomes impossible (this would be relatively rare) and second to detect those tasks or parts of tasks in which the workload is high enough so that, if a further burden is added in the form of, say, an unforeseen event, operator errors are likely to occur If the measures are valid, i e , if workload is really being measured, then they should increase with increasing task demand, and this should result in increased operator errors The pilot studies, or trials, are designed to assess this

The trials must be run before full-scale operational testing, both to allow for changes in the measures and because they will involve "loading" the operator with extra, or secondary, tasks until the error rate starts to increase, a technique that would be undesirable during full-scale operations testing

1. Procedure From three to six people, preferably those with some experience on the Mark III or a similar system, should serve as subjects for the trials They need not be rated as novice, intermediate, or expert, since the purpose of the trials is simply to see whether the measures respond to increasing objective workload

A task of about 30-45 minutes duration will be designed for the trials so as to be as similar as possible to the Mark IVA LMC operator's task, i e , an interactive computer task Although the LMC operator's tasks may be intermittent, with periods of high workload alternating with relatively light workload, the trial task should occupy the operator continuously

The operators will perform the task under each one of three conditions In the first condition, the primary task alone will be performed by the operator (low workload), in the second condition, a moderate amount of a secondary task, to be done at the same time as the primary task, will be assigned (medium workload), in the third condition, a still heavier amount of the secondary task will be assigned the operator (high workload) The purpose of the secondary task is twofold First, it should demonstrate that as the secondary task increases the workload on the operator, the three workload measures rise, and second, performance on the secondary task should be negatively correlated with the workload measures, since, as primary workload goes up, the operator has less "spare mental capacity" to work on the secondary task Both these effects serve as checks on the validity of the workload measures

There will, of course, be no secondary task in the first condition In the second condition, the secondary task will consist in the operators' keeping a written log of all of the opera-

tions and parameters This should be done on a prepared form, similar to forms currently in use In the third condition, in addition to keeping the written log, the operator will keep a voice log by reporting the data over the voice circuit to a receiving operator, similar to the NOCC operator These two tasks have three characteristics (1) they are designed to overload the operator in two increasing steps so as to produce an increase in the workload measures and possibly an increase in procedural lapses (errors of a type that would not result in data loss) on the part of the operator, (2) they are designed to be similar to secondary tasks normally performed by operators, so they will not be perceived as simply "busy work" that can be ignored, and (3) the amount of work involved in them can be manipulated in terms of the amount of detail required Thus, if on first presentation the overloading tasks are found to be easily handled by the operator, and make no difference in the workload measures or in lapses, the amount of record keeping required should be increased until an effect is shown

The operators should practice on the simulated system before the trials until their performance is asymptotic in terms of the amount of time they require to do the job Then the trials with the three workload conditions can begin Each person who serves as a subject should experience all three conditions, so that performance differences among conditions can be attributed to the effects of the workload variaton and not to individual differences Each individual should experience the workload conditions in a different order chosen randomly from among the six possible orders of three conditions

2 Analysis The three workload measures described above will be taken for each individual for each subtask under each condition Means will be calculated for each condition, and an analysis of variance run separately for each measure This should demonstrate that the means of the workload measures increase significantly with the increase in imposed workload over the three conditions, thus providing evidence for the validity of the measures

It would also be of interest to gauge the effect of workload on operator performance Generally, operators compensate for increased workload by increased efficiency, so that errors which actually result in lost data are rare, and probably would not occur at all in a 30-45 minute trial However, anomalies such as an incorrect input, failure to detect a significant reading on a display, or failure to understand a verbal communication are fairly common They are usually corrected in a short time (so human operators are sometimes called "self-correcting") but they may significantly add to the workload itself and, more importantly, if the workload is already high, they may lead to further errors which may actually result in missed data Therefore, an increase in operator anomalies with increasing workload must be taken seriously, even though the anomalies are corrected

In the trials described here, then, all operator anomalies should be recorded in the course of data collection They can then be subjected to the same type of analysis as the workload measures, i e , an anlysis of variance which should demonstrate that anomalies increase as imposed workload increases Further, a detailed examination of the subtask in which anomalies occur, and their workload measures, should yield useful data on the possible effect of workload on the probability of operator anomalies, on which some predictions may be made about the probability of data loss given a particular level of workload

Some operator anomalies which should be considered are

(1) Failure to notice a signal change or update

(2) Hitting the wrong input key

(3) Failure to find the source of an error message

(4) Voice communication failure (incorrect reception or transmission), including requests to "say again"

(5) Incorrect information recorded on the log

(6) Incorrect information transmitted to the voice log

Furthermore, since such anomalies are inherently unpredictable, it is not possible to define them completely in advance, so that any anomaly that occurs during data collection should be recorded This will involve some on-the-spot judgment on the part of the test conductor The slightest anomaly should be recorded so as to increase the data pool

If this procedure is followed, a useful and meaningful (in terms of probability of data loss due to operator error) measure or measures of human operator workload can be obtained The measures can then be used in an operations test as a part of systems evaluation and periodically thereafter whenever system testing is needed

IV. Implementation

The operator workload measurement methodology would use specially developed SOEs for operations testing The SOEs would indicate when an operator rating is required The rating will be input to the system by means of a text message which is timed tagged and recorded on the operations log The analysis will use the data recorded on the operations log

Validating the operator workload measurement with the secondary task method would be accomplished before the formal Mark IVA operational workload testing is begun

V. Summary

Operator overload is identified as a potential problem in Mark IVA system operations Three methods for measuring operator overload have been presented The methods are operator ratings, primary task work measures, and information processing time measures A method of validating these three workload measures by secondary task performance measures is suggested

These operator workload measures would be an integral part of the OT&E plan for operations testing of the DMC Operations testing in this form would provide indications of where in the operations of the Mark IVA system we can expect system failures due to operator overload

The operator workload measurement methodology described here allows the problem of operator overload to be addressed directly It provides a general capability to measure operator workload in future systems

References

1 Benbosat, I , Dexter, A , and Masubi, P , "An experimental study of the human/ computer interface," *Communications of the ACM*, 1981, 24, 11, pp 752-762 Nov 1981

2 Schneiderman, B , *Software Psychology Human Factors in Computer and Information Systems*, Winthrop Publishers, Inc , Cambridge, Mass , 1980

3 Wierwill, W W , Williges, R C , and Schiflett, S G , "Aircrew workload assessment techniques," in Hartman, B O , and McKenzie, R E , *Survey of Methods to Assess Workload* AGARDograph No 246, Aug 1979

4 Moray, N , *Mental Workload Its Theory and Measurement* Plenum Press, New York, 1979

5 McConnell, J P , "An application of measurement methods to improve the quantitative nature of pilot ratings scales," *IEEE Trans Man-Machine Systems*, MMS-10, pp 81-92, 1969

6 Hicks, T , and Wierwille, W , "Comparison of five mental workload assessment procedures in a moving-base simulator," *Human Factors*, 21, pp 129-144, 1979

7 Dyer, W , Matthews, J J , Stalac, J F , Wright, C E , and Yudowitch, K , *Questionnaire Construction Manual Annex*, U S Army Research Institute for the Behavioral and Social Sciences, AD43012, July 1976

8 Adams, J A , "Issues in human reliability " *Human Factors*, 24, pp 1-10, 1982

9 Sternberg, S , "Memory-scanning mental process revealed by reaction-time experiments," *American Scientist*, 57, 4, pp 421-457, 1969

10 Chiles, W D , and Alluisi, E A , "On the specification of operator or occupational workload with performance-measurement methods," *Human Factors*, 21, pp 515-528, 1979

11 Albanese, R A , Mathematical analysis and computer simulation in military mission workload assessment," *Proceedings of the AGARD Conference on Methods to Assess Workload*, AGARD-CPP-216, AB-1-A13-6, April 1977

We are interested in obtaining information about specific points in the sequence of operations where the workload or amount of effort required by you, the operator, approaches the level where the time allowed is not sufficient to get the job done In order to do this, we will ask you for a series of ratings of the difficulty of the task and the amount of workload you experience on each of many of the small subtasks of a particular job At frequent intervals while you are doing a job, a request for two ratings will appear in the SOE At these times, please respond as fast as you can with the appropriate FFI text message which includes a value of from 1 through 7 to indicate your ratings of the difficulty of the task and the workload you have experienced since the last rating (or since starting the job) Use the following scale

The task was	There was
(1) Extremely easy to do	(1) Plenty of time to spare
(2) Very easy to do	(2) Some time to spare
(3) Easy to do	(3) A little time to spare
(4) Neither hard nor easy to do	(4) Time allowed is just enough to get the job done
(5) Hard to do	(5) Hard to finish in the time allowed
(6) Very hard to do	(6) Very hard to finish in the time allowed
(7) Extremely hard to do	(7) Impossible to finish in the time allowed
1 2 3 4 5 6 7	1 2 3 4 5 6 7
increasing difficulty –	increasing time pressure –
Difficulty rating	Time pressure rating

Fig. 1. Instructions and scale for operator workload ratings

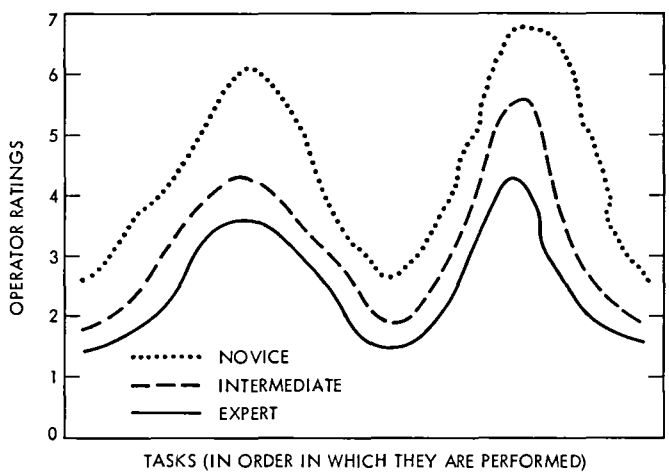


Fig 2. Mean workload ratings for each subtask

The following questionnaire should be given once after the completion of a trial and/or after a reasonable familiarization period during operational testing

Please indicate your rating of the following statements by writing a number from 1 to 7 in the space provided

1 – Very strongly disagree

2 – Strongly disagree

3 – Disagree

4 – Neither agree nor disagree

5 – Agree

6 – Strongly agree

7 – Very strongly agree

1	The workload required of the operators of this system is too high	___
2	The system provides all the displays that are needed for the job	___
3	In general, each display is adequate for its job	___
4	The working group was able to respond to the demands placed on it	___
5	Information can move freely and accurately among the LMC, roving, and NOCC operators	___
6	It is easy to find the information needed in the system documentation	___
7	System event reporting makes clear what action is to be taken	___
8	Error messages describe each fault in sufficient explanatory detail	___
9	Error messages make clear what action is to be taken	___
10	There are too many meaningless messages on the screen	___
11	There were other problems in operating the system (specify)	___

Fig 3 Operator workload questionnaire

The 1986 Launch of the Galileo Spacecraft via the Space Transportation System

A L Berman and D J Mudgway
TDA Mission Support Office

J C McKinney
Mission Information Systems Engineering Section

Beginning with the Galileo spacecraft launch in 1986, deep space payloads will be launched via the Space Shuttle This change from the previous use of expendable launch vehicles will introduce large changes in procedures and data flow configurations for both the flight project and the Deep Space Network during the launch period This article describes the planned Galileo launch period sequence of events and telemetry and command data flow configurations

I. Introduction

Starting with the launch of the Galileo spacecraft, now scheduled for late May 1986, deep space payloads will be launched via the Space Shuttle vehicle of the Space Transportation System (STS), in sharp contrast to all previous launches of deep space payloads via expendable launch vehicles This very significant change in the method of launch will result in large procedural changes for both flight projects and the Deep Space Network (DSN), in its capacity as lead support network for all deep space missions In a previous article (Ref 1), the Space Shuttle impact on the DSN initial acquisition was described, in this article, procedural differences for the flight project and DSN during the pre- and postlaunch periods are examined In particular, major subphases of the pre- and postlaunch period are identified, and telemetry and command data flow configurations are presented for each subphase. The Galileo mission sequence of events and data flow configurations here described can be considered typical of deep space launches planned for the Shuttle era

The information here described was first presented in Ref 2, under the assumption that the upper stage was to be the Inertial Upper Stage (IUS), this article now updates the information for the currently planned Centaur upper stage

Major differences for the flight project and DSN in the upcoming Space Shuttle era are the larger number of subphases in the launch period (e g , Shuttle on-orbit phase of several hours, for which there was no similar phase in the expendables era) and direct launch involvement of additional NASA centers (e g , Johnson Space Center and Lewis Research Center).

Section II describes the nominal Galileo sequence of events during the launch period Section III identifies major launch

period subphases, while Section IV identifies the telemetry and command data flow configurations for each of the launch period subphases

II. The Nominal Galileo Sequence of Events

As previously mentioned, the Galileo spacecraft is currently scheduled for a late May 1986 launch The overall launch period begins with the transporting of the spacecraft to the Kennedy Space Center (KSC), extends through liftoff, and terminates with a successful DSN initial acquisition Typical prelaunch (prior to liftoff) and postlaunch (following liftoff) Galileo sequences of events are described in detail below

A. Nominal Galileo Prelaunch Sequence of Events

The Galileo spacecraft is transported from the Jet Propulsion Laboratory (JPL) to the Kennedy Space Center (KSC) by a United States Air Force C-5A aircraft During all spacecraft testing at KSC, command and telemetry support will be provided via the DSN Facility (MIL 71) with operations control from the Mission Support Area (MSA) at JPL After being unpacked in the Spacecraft Assembly and Encapsulation Facility (SAEF), the spacecraft is inspected to determine if any damage was sustained during transit Baseline tests, including the use of S- and X-band radio frequency (RF) subsystems, are evaluated At the conclusion of these tests, Radioisotope Thermoelectric Generators (RTGs) are installed on the spacecraft and tests are rerun using the RTGs as the source of electrical power When these tests are completed, all power is removed from the spacecraft to ensure that the installation of pyrotechnic devices and the loading of consumables in the propulsion system can be carried out in a safe environment

At the conclusion of the propulsion loading and pyrotechnics installation, the spacecraft will again be activated for further testing and will be prepared for transfer to the Vertical Processing Facility (VPF) At the VPF the spacecraft will be mated to the Centaur Tests via the Merritt Island Launch Area (MILA) will be conducted to verify Space Shuttle end-to-end communications using the Cargo-Integration Test Equipment (CITE) to provide simulated Shuttle data interfaces A series of operational tests will also be conducted Upon satisfactory completion of these tests, the spacecraft will be placed into the storage mode to await shipment to the launch pad

About 30 days before launch, the spacraft will be removed from storage and transported to the launch pad, preceding the Shuttle arrival by two days After cargo preparation procedures are completed, the spacecraft will be installed in the Shuttle Orbiter Bay About four days before launch, the RTGs are re-installed in the spacecraft Final end-to-end communications tests using MILA and Space Transportation System (STS) Tracking Data Relay Satellite (TDRS) communications links will then be conducted These tests should last about two days At their conclusion, the launch countdown will commence

Figure 1 presents the nominal Galileo prelaunch sequence of events for a May 1986 launch

B. Nominal Galileo Postlaunch Sequence of Events

Galileo launch mode telemetry will be available in near-real-time via the STS-TDRS communications link, from liftoff to Shuttle-Centaur separation Based on the telemetry data, a decision to continue with the planned flight to Jupiter must be made by Launch plus 3 hours

If a "go" decision is made, the Shuttle-Centaur separation should occur during the third Shuttle orbit, at about four hours after launch During the more favorable periods of the launch window, the separation could be delayed until the fourth or the fifth orbit without jeopardizing the objectives of the Galileo mission After separation, a Centaur S-band link can be used to route data between the spacecraft and Shuttle or TDRS The maximum useful range of the Centaur to Shuttle link is 10 kilometers

About 45 minutes after Shuttle-Centaur separation, the Centaur main engine will burn for approximately 10 minutes After main engine cutoff (MECO), the Centaur will initiate a slow thermal roll of 0 1 rpm, and the Galileo spacecraft will start deployment of the RTG, science and magnetometer booms The spacecraft transmitters will then be turned on to provide a downlink through TDRS just prior to Centaur-spacecraft separation Centaur will turn the spacecraft to point 8 degrees to the earth side of the sun and will spin-up the spacecraft to 2 9 rpm The spacecraft will then separate from Centaur, which will maneuver to avoid the same trajectory path as the Galileo spacecraft The spacecraft transponder will now be the only means of exchanging data between the earth and the spacecraft

Immediately after separation, the high-gain antenna will be deployed and the retro propulsion module (RPM) pressurized During these events, DSN acquisition of the downlink will be established (approximately 15 minutes after Centaur-spacecraft separation) After RPM pressurization, the Galileo spacecraft will perform a sun acquisition (approximately one hour after Centaur-spacecraft separation)

Figure 2 presents the nominal Galileo postlaunch sequence of events

III. Major Launch Subphases During a Shuttle Deep Space Payload Launch

The overall launch period of a deep space payload launch is conveniently subdivided into three major categories, as follows

(1) Prelaunch phase

(2) Shuttle attached phase

(3) Shuttle detached phase

A. Prelaunch Phase

This phase starts with spacecraft assembly in the Spacecraft Assembly and Encapsulation Facility (SAEF), and includes that time the spacecraft spends in test at the Vertical Processing Facility, and finally the move to the launch pad This phase terminates with liftoff from the pad Subphases during the prelaunch phase are identified as

(1) Spacecraft assembly and encapsulation facility

(2) Vertical processign facility

(3) Launch pad

B. Shuttle Attached Phase

The phase starts at the moment of liftoff, and proceeds through the Shuttle ascent and the Shuttle on-orbit operations This phase is terminated when the Centaur spacecraft is placed outside the Shuttle Orbiter Subphases during the shuttle attached phase are identified as

(1) Shuttle ascent

(2) Shuttle on-orbit

C. Shuttle Detached Phase

This phase starts when the Centaur spacecraft is placed outside the Shuttle vehicle, and proceeds through the Centaur burn and spacecraft-Centaur separation This phase terminates with a successful DSN initial acquisition Subphases during the Shuttle detached phase are identified as

(1) Centaur burn

(2) Post Centaur separation

IV. Galileo Telemetry and Command Data Flow Configurations During the Launch Period

There are four major telemetry and two major command data flow configurations during the launch period, and these are detailed as follows

A. Galileo Spacecraft – DSN

This path exists for both telemetry and command data The link between the spacecraft and the Merritt Island Launch Area (MILA) DSN facility (MIL 71) is both radio frequency (RF) and hardline Communications from MIL 71 to the JPL Mission Control and Computing Center (MCCC) is via the JPL Ground Communications Facility (GCF)

B. Galileo Spacecraft – TDRSS

This path is for telemetry data only The link between the spacecraft and the MILA Ground Spacecraft Tracking and Data Network (GSTDN) station is RF From the GSTDN station an RF uplink is established to the Tracking and Data Relay Satellite (TDRS) Alternatively, an RF link can be established directly from the spacecraft to TDRS From TDRS, an RF downlink is established to the White Sands Ground Station (WSGS) From there, the data is transmitted via domestic satellite (DOMSAT) to the Goddard Space Flight Center (GSFC) NASA Communications (NASCOM) switching center, and thence through DOMSAT to JPL MCCC

C. Galileo Spacecraft – Centaur

This path is for telemetry data only Spacecraft telemetry is embedded in Centaur telemetry An RF link is established from the Centaur to the TDRS From TDRS, an RF downlink is established to WSGS From there, the data is transmitted via DOMSAT to the GSFC NASCOM switching center, and thence through DOMSAT to the Centaur Payload Operations Control Center (POCC), at the Lewis Research Center (LRC) From LRC POCC, Galileo telemetry is stripped out and transmitted via GSFC NASCOM to JPL MCCC

D. Galileo Spacecraft – Shuttle

This path is for both telemetry and command For telemetry the link can be direct from the Galileo spacecraft to the Shuttle Orbiter, or embedded in Centaur telemetry data to the Shuttle Orbiter From the Shuttle the link is RF to TDRS to WSGS From WSGS, the data is transmitted via GSFC NASCOM to Johnson Space Center (JSC) Mission Control Center (MCC) From JSC MCCC, Galileo telemetry is stripped out and transmitted via GSFC NASCOM to JPL MCCC Alternatively, Centaur Galileo telemetry is transmitted via GSFC NASCOM to LRC POCC, where Galileo telemetry is stripped out and provided through GSFC NASCOM to JPL MCCC

For command, the link begins with the LRC Centaur POCC, in conjunction with voice communication from JPL

MCCC From the LRC POCC, the link is first to JSC MCC through GSFC NASCOM, then to WSGS through GSFC NASCOM, and finally RF to TDRS to the Shuttle Orbiter From the Orbiter, the link is either hardline or RF (Centaur Shuttle Orbiter distance <20 km) to Centaur, and finally, to the Galileo spacecraft For this mode, only a limited series of commands ("discrete commands") is possible.

Figures 3 through 11 illustrate the above data flow paths for the various launch subphases

References

1. Khatib, A R , Berman, A L , and Wackley, J A , "Space Shuttle Launch Era Spacecraft Injection Errors and DSN Initial Acquisition," *TDA Progress Report 42-64*, Jet Propulsion Laboratory, Pasadena, Calif , pp 80-82, Aug. 15, 1981

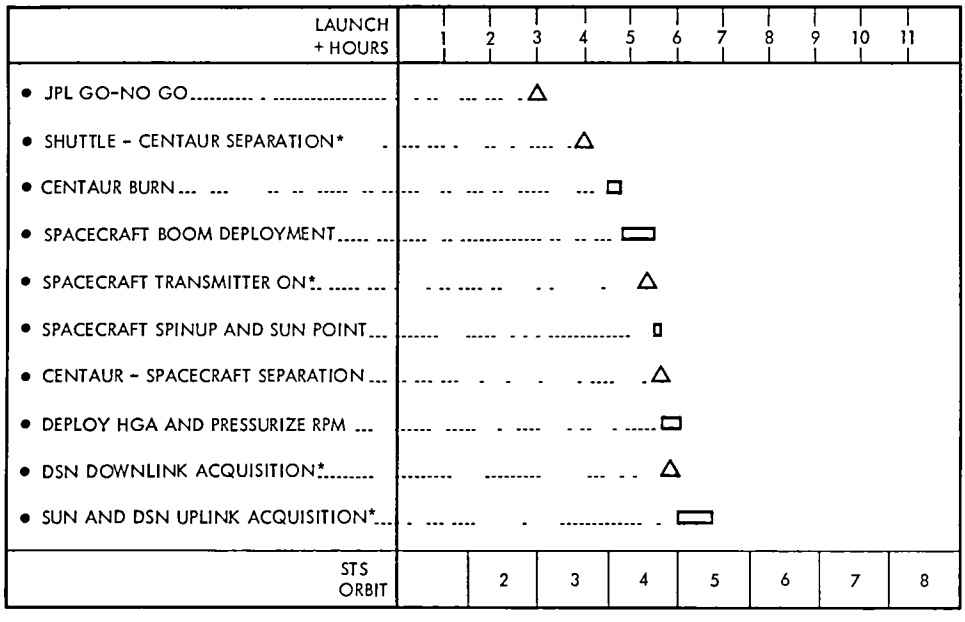
2 Berman, A L , Larkin, W E , and McKinney, J.C , "Deep Space Payload Launches via the Space Transportation System," *TDA Progress Report 42-70*, Jet Propulsion Laboratory, Pasadena, Calif , pp 21-36, August 15, 1982

Note Added During Publication

In December 1982, significant modifications were made to the Galileo launch support plan, primarily, these modifications involved locating the Centaur Payload Operations Control Center (POCC) at Kennedy Space Center (KSC) and performing the DSN initial acquisition prior to spacecraft-Centaur separation. The impact of these changes on the launch phase configurations and sequences of events will be described in a future article

SPACECRAFT EVENT	FACILITY	MIL 71	ACTIVITY	1986 FEBRUARY	MARCH	APRIL	MAY
• INSPECTION BASELINE TEST	S/C ASSEMBLY AND ENCAPSULATION	√	S- AND X-BAND T/R				
• INSTALL TEST RTG	S/C ASSEMBLY AND ENCAPSULATION	√	S- AND X-BAND T/R COAX TLM				
• PROPULSION PREPARATIONS	S/C ASSEMBLY AND ENCAPSULATION	–	NO POWER FOR SPACECRAFT				
• BASELINE TEST	S/C ASSEMBLY AND ENCAPSULATION AND VERTICAL PROCESSING	√	S-BAND RECEIVE COAX (TLM, TDRS, CENTAUR)				
• CENTAUR MATE AND TEST	VERTICAL PROCESSING	√	S-BAND RECEIVE COAX (TLM)				
• STORAGE AND BASELINE TEST	VERTICAL PROCESSING	√	S-BAND RECEIVE COAX (TLM)				
• INSTALL IN SHUTTLE CARGO BAY	LAUNCH PAD	–	NO POWER FOR SPACECRAFT				
• RTG FINAL INSTALL AND TEST	LAUNCH PAD	√	S-BAND RECEIVE COAX (TLM)				
• END-TO-END TEST	LAUNCH PAD	√	S-BAND RECEIVE COAX (TLM) CENTAUR, TDRS VIA SHUTTLE				
• LAUNCH	LAUNCH PAD	√	TRACKING AND TELEMETRY VIA CENTAUR AND TDRSS NETWORK				

Fig 1 Nominal Galileo prelaunch sequence of events for a May 1986 launch



* THIS SEQUENCE COULD BE DELAYED 1 OR 2 ORBITS

Fig 2 Nominal Galileo postlaunch sequence of events for a May 1986 launch

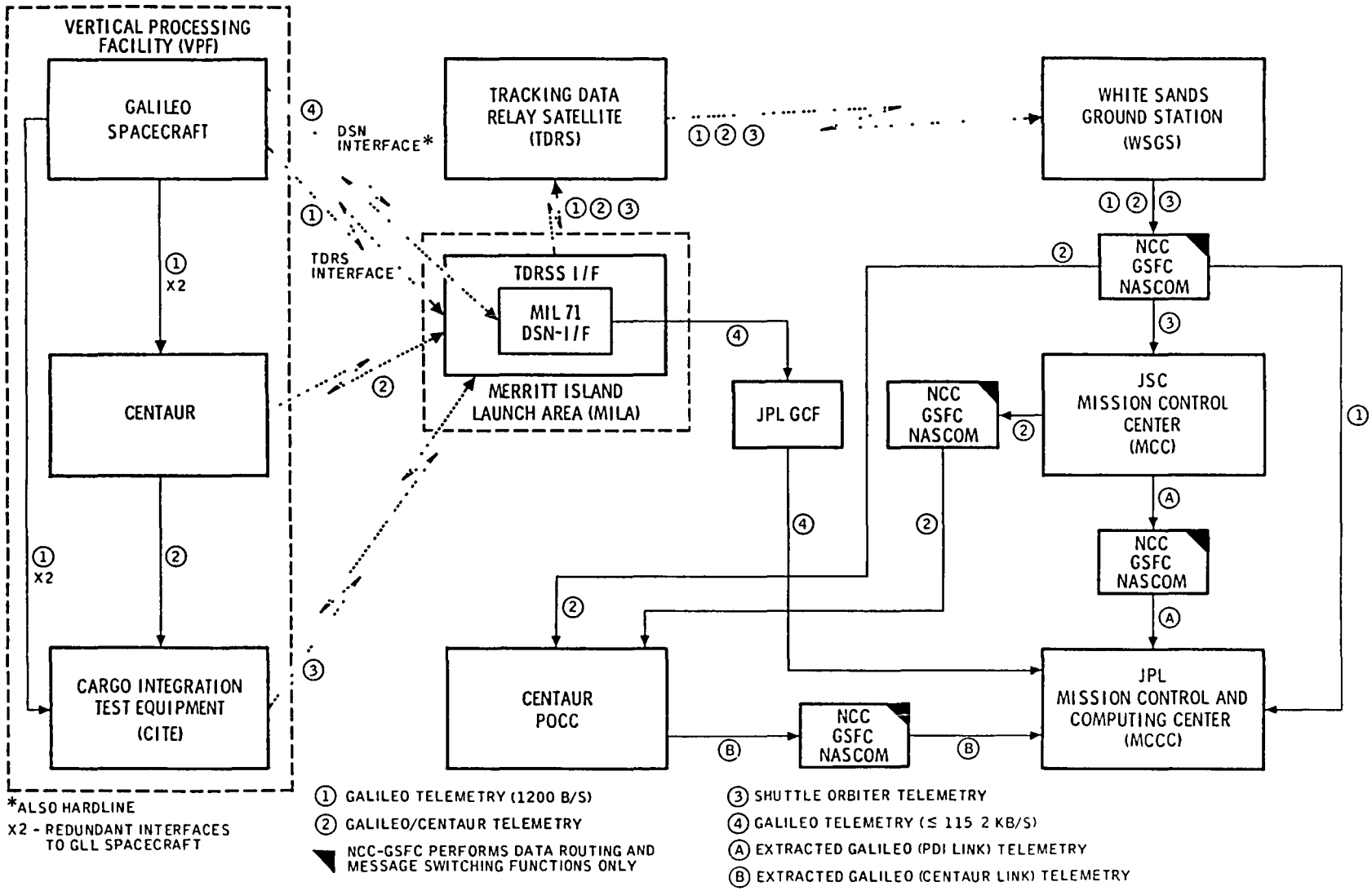


Fig 3 Telemetry data flow in the prelaunch phase – Vertical Processing Facility subphase

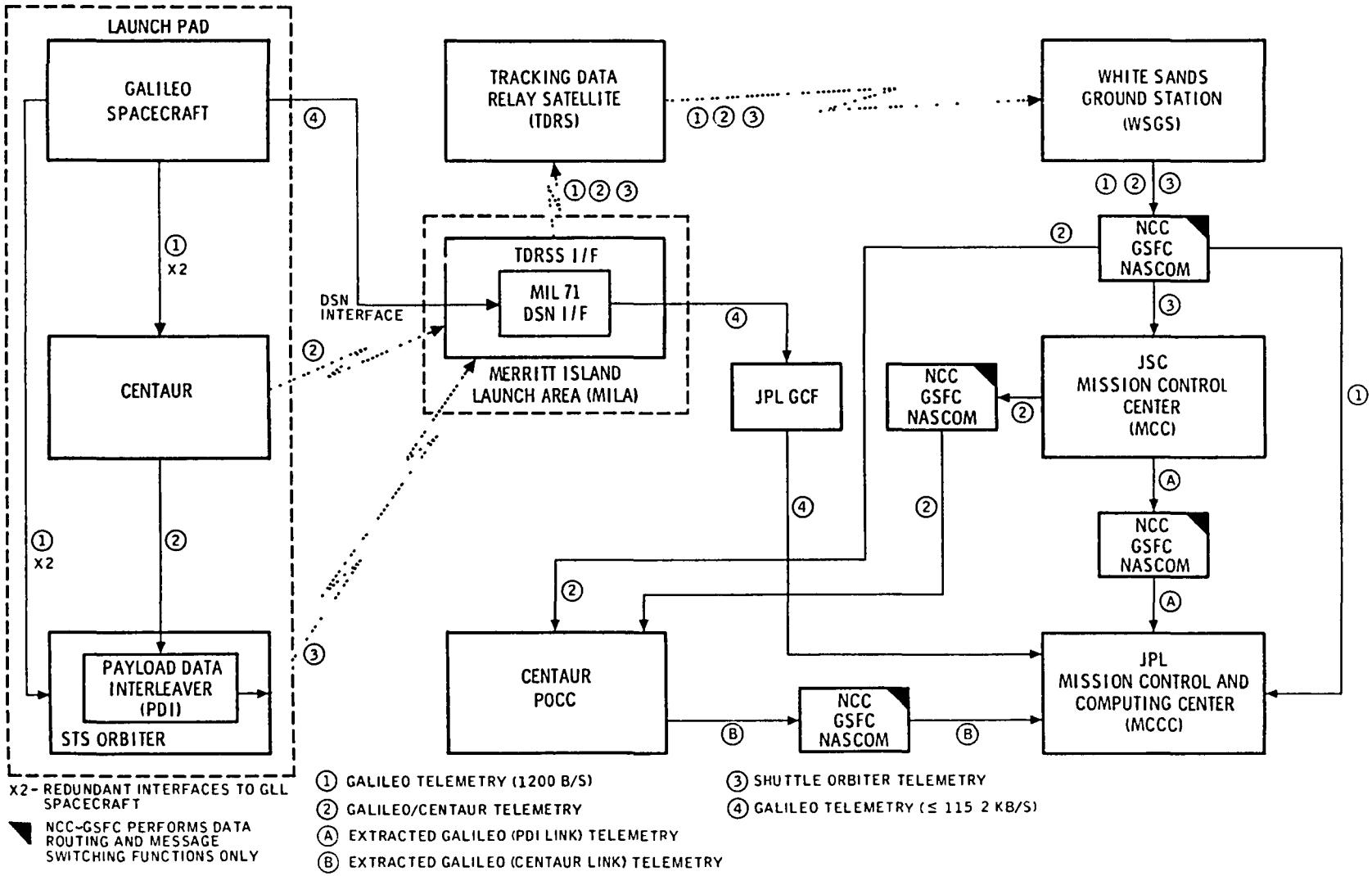


Fig 4 Telemetry data flow in the prelaunch phase – launch pad subphase

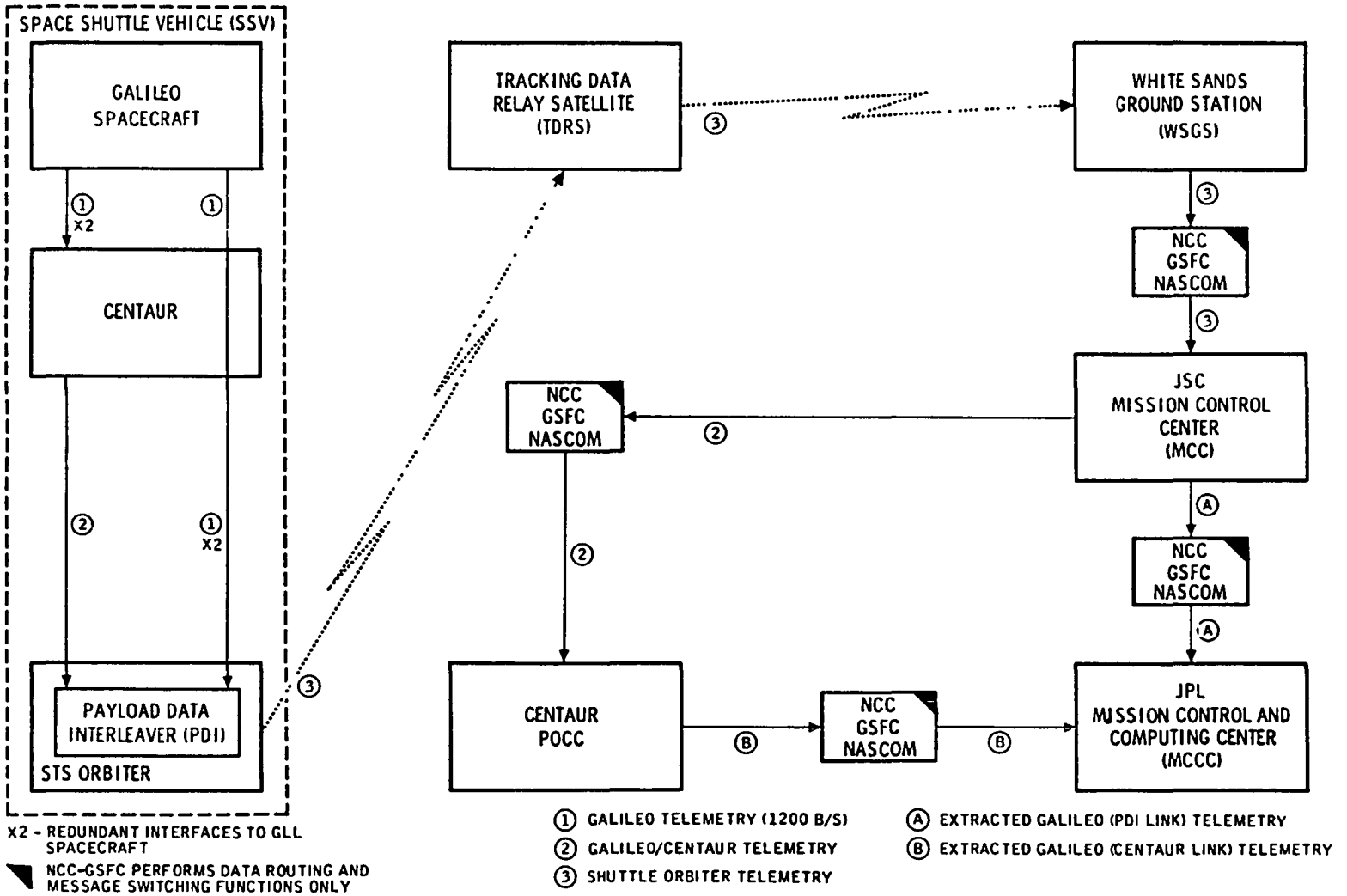


Fig. 5. Telemetry data flow in the Shuttle attached phase – ascent subphase

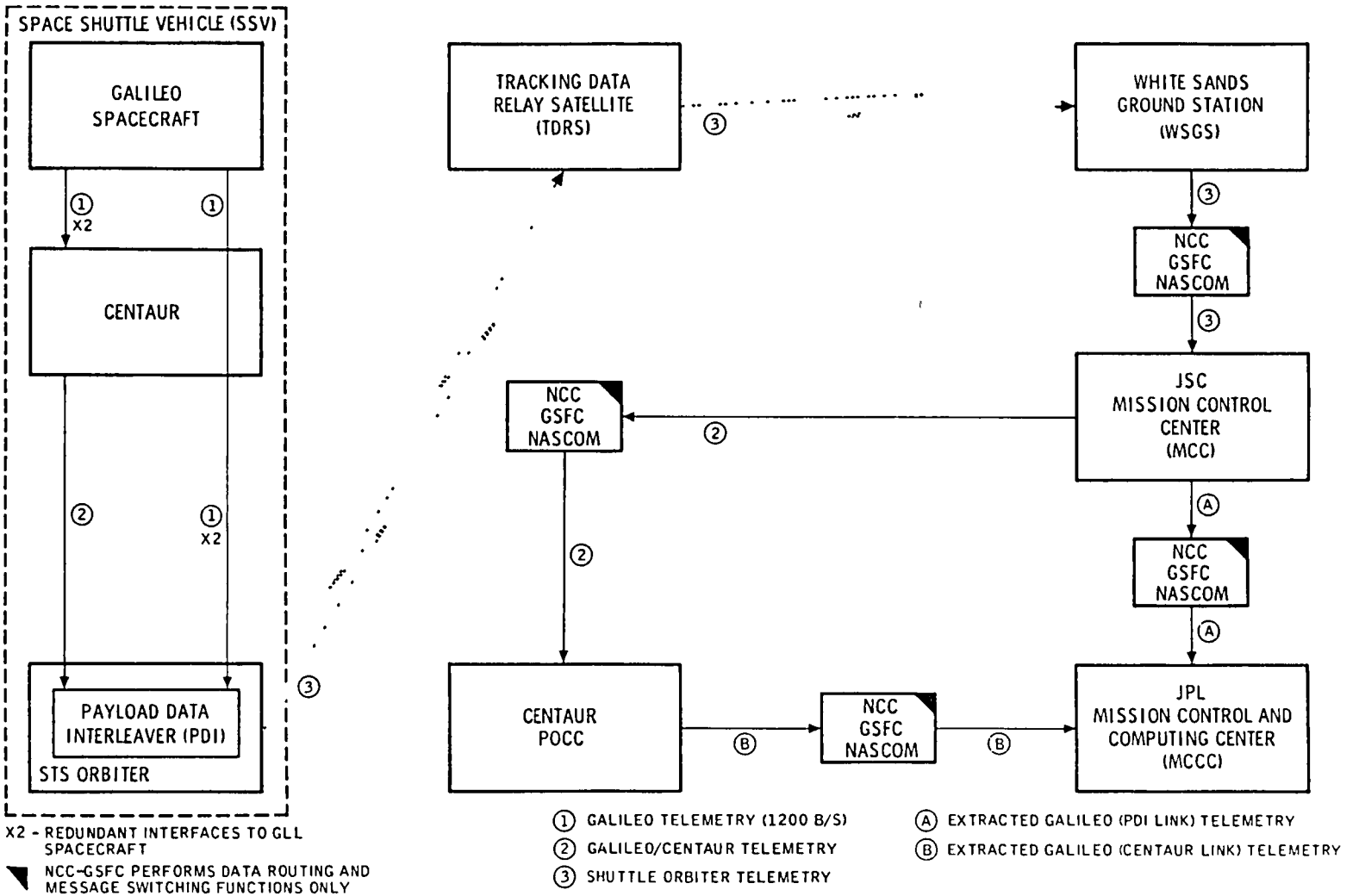


Fig. 6 Telemetry data flow in the Shuttle attached phase — Shuttle on-orbit subphase

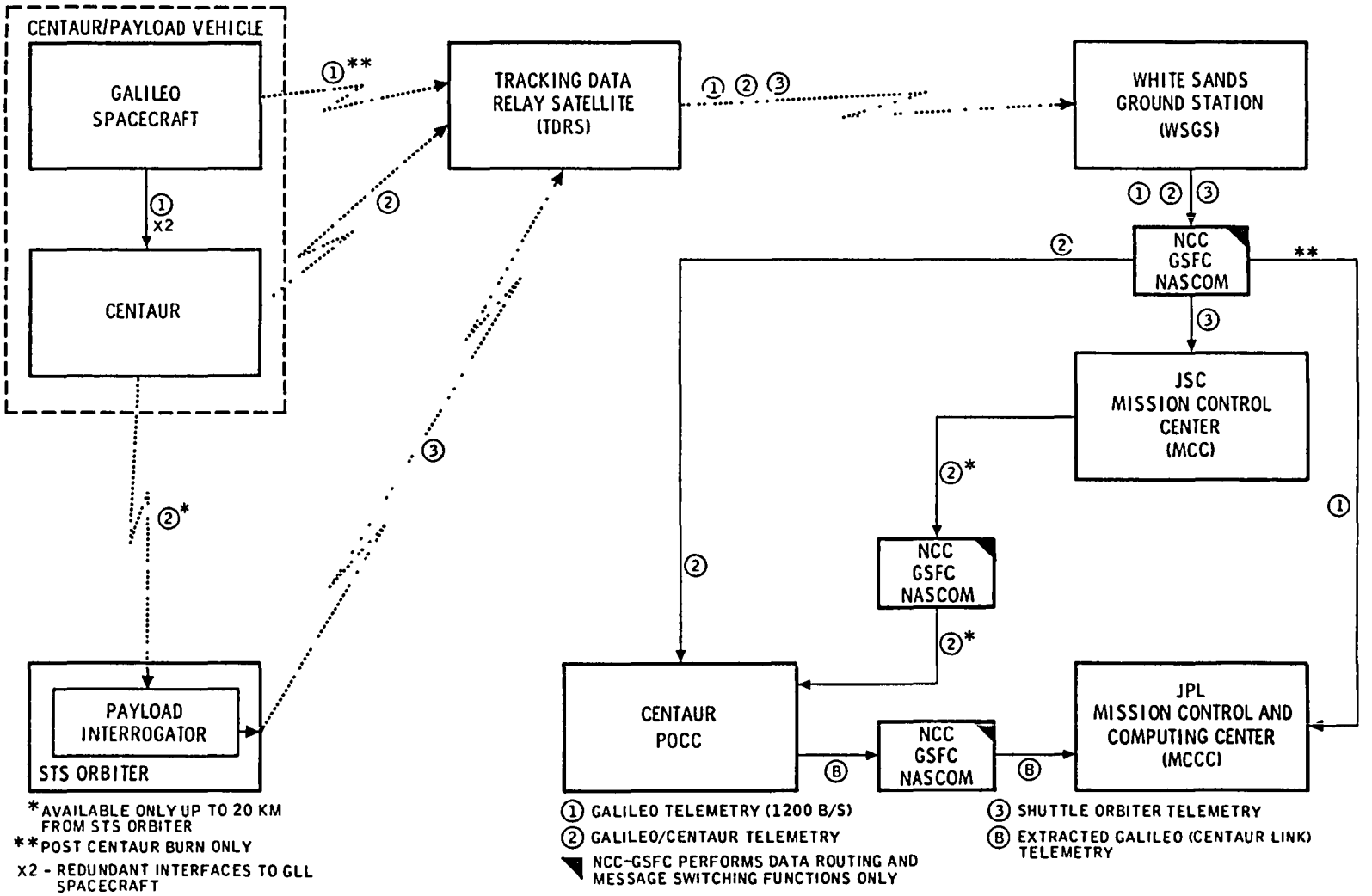


Fig 7 Telemetry data flow in the Shuttle detached phase

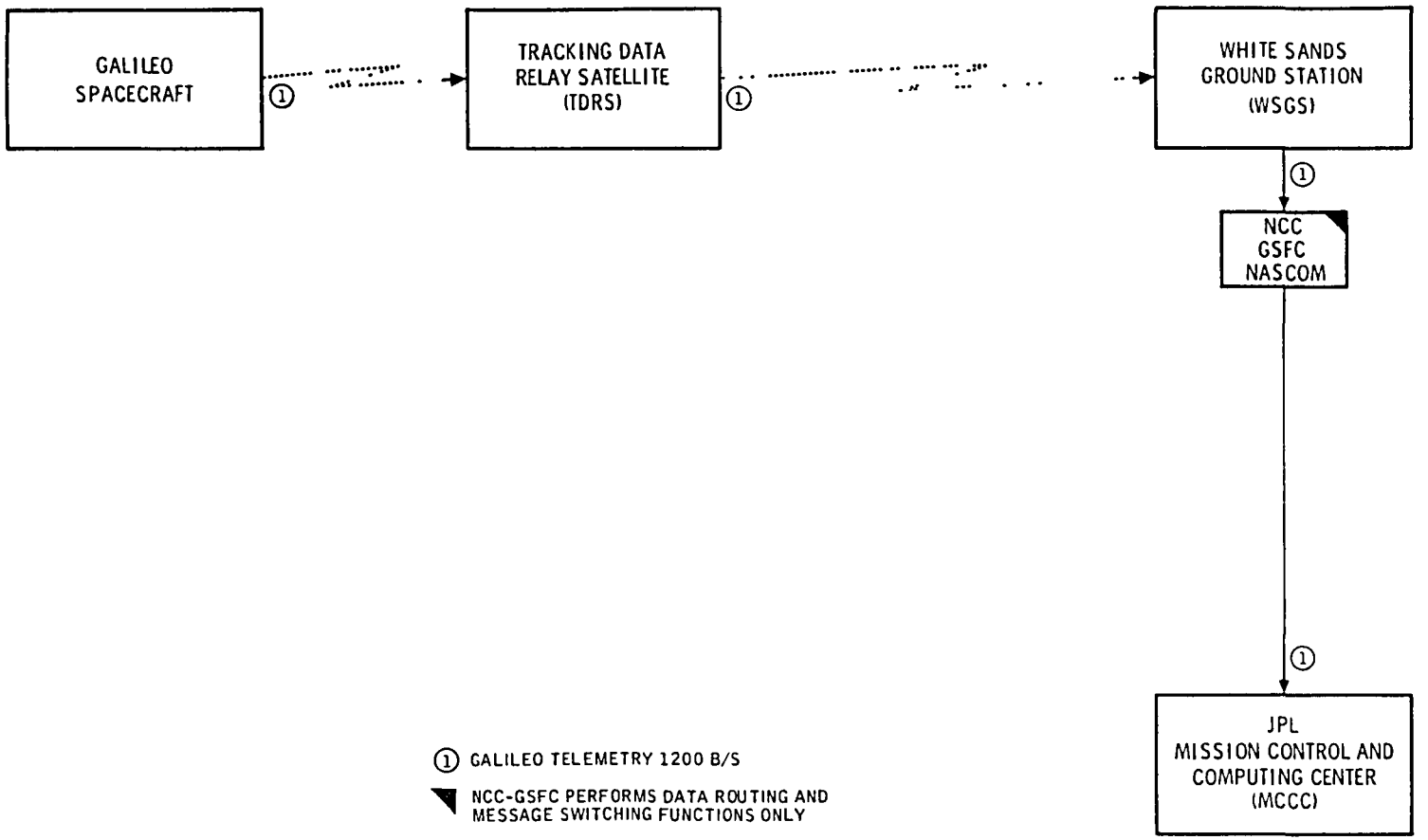


Fig. 8. Telemetry data flow in the Shuttle detached phase -- post Centaur separation phase

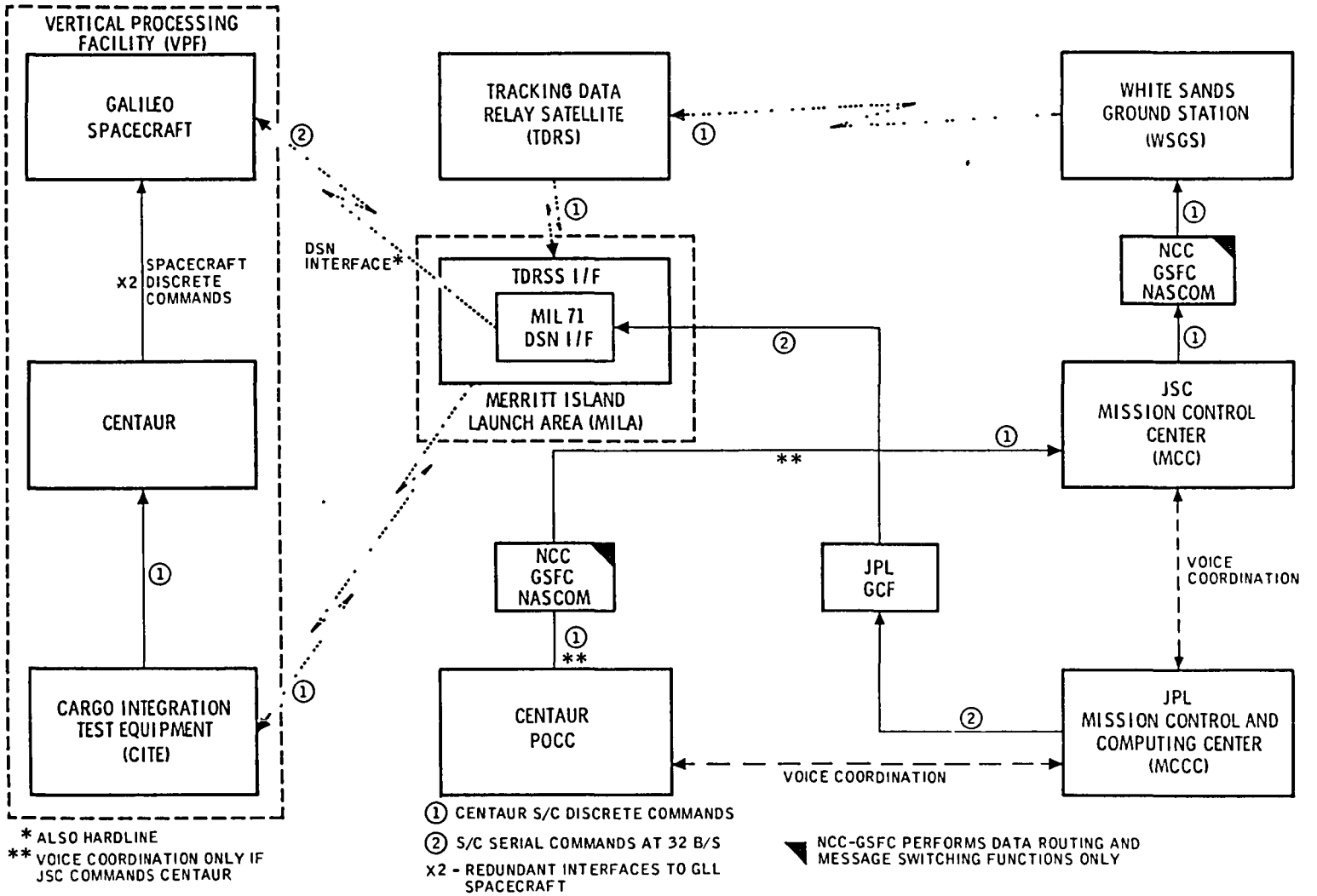


Fig. 9. Command data flow in the prelaunch phase — Vertical Processing Facility subphase

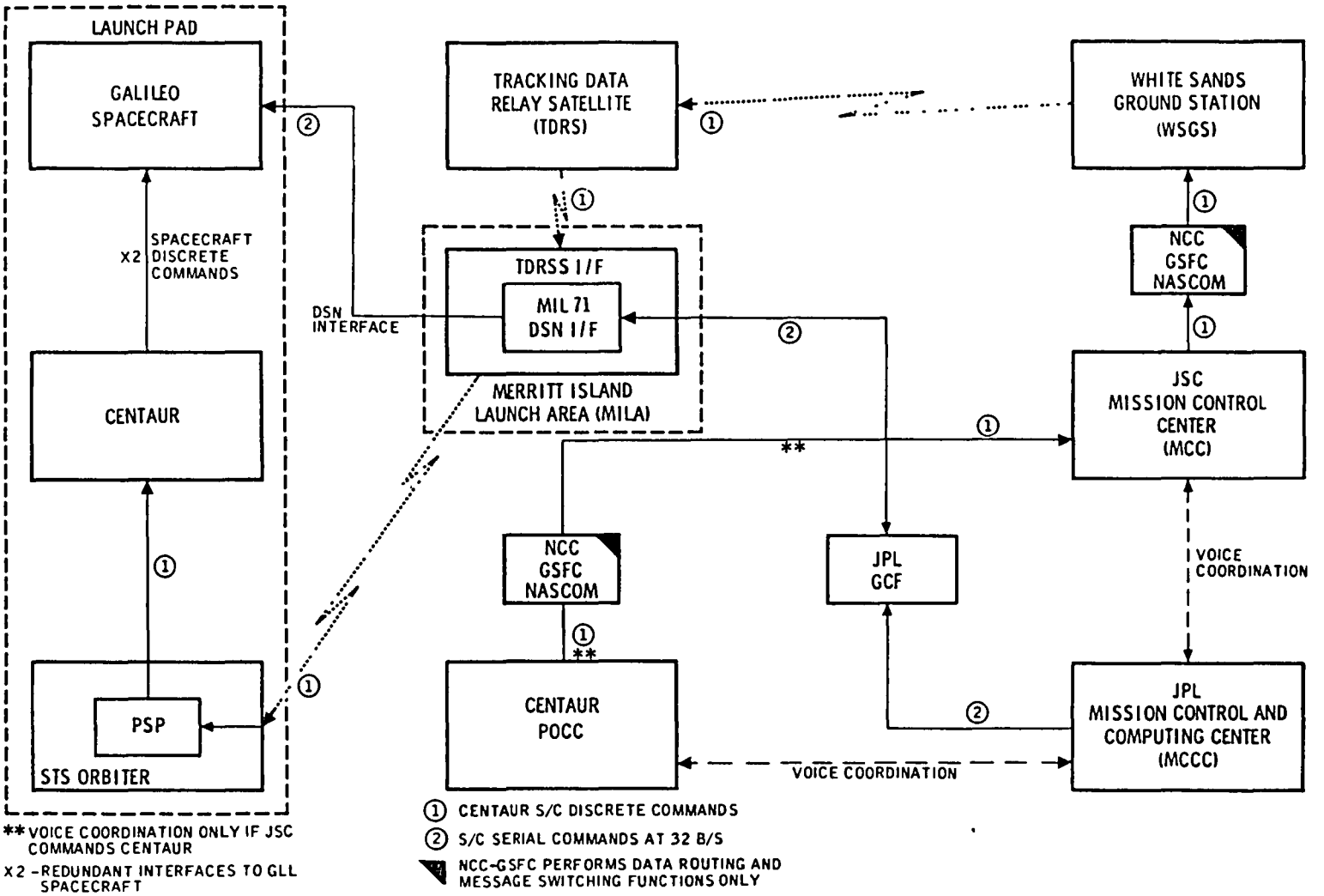


Fig 10. Command data flow in the prelaunch phase -- launch pad subphase

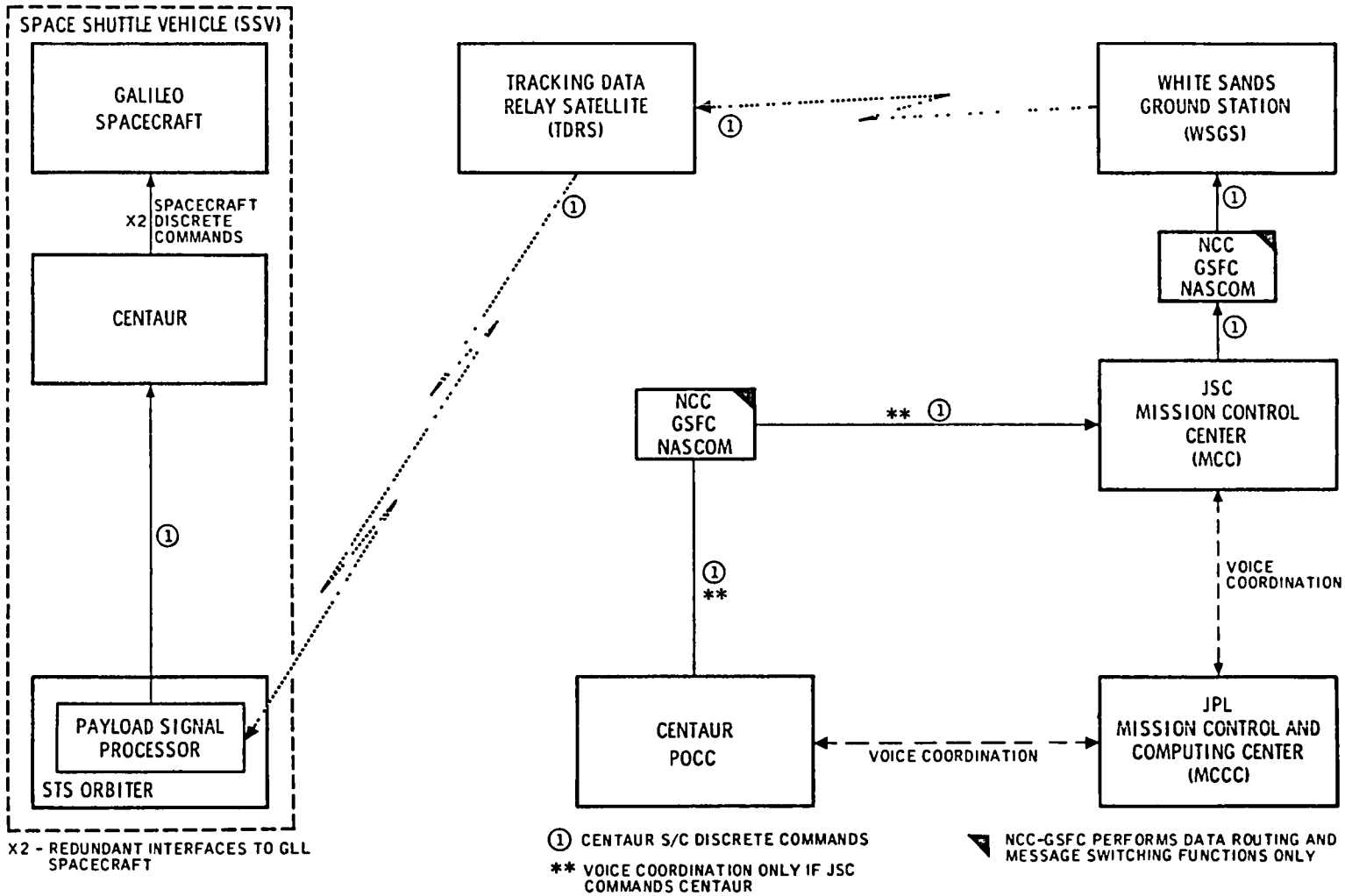


Fig 11. Command data flow in the Shuttle attached phase – Shuttle on-orbit subphase

DSN Ground Communications Facility

R H Evans
TDA Engineering Section

A functional description of the GCF and its relationships with other elements of the DSN and NASCOM is presented together with development objectives and goals and comments on implementation activities in support of flight projects

I. Introduction

The Ground Communications Facility (GCF) is one of the three elements of the DSN The GCF provides for transmission, reception, and monitoring of earth-based point-to-point communications between the Deep Space Stations (DSSs, one of the DSN elements), the Network Operations Control Center (NOCC, the other element) located at JPL, Pasadena, and the Mission Operations Control Center (MOC) at JPL Voice, teletype, high-speed data, and wideband circuits of the world wide NASA Communications Network (NASCOM) are utilized for all long-distance circuits, except those between JPL and the Goldstone Deep Space Communications Complex (GDSCC) Goddard Space Flight Center (GSFC) NASCOM Engineering has delegated the responsibilities for planning, budgeting, design, implementation, operation, and maintenance of the communications requirements between Goldstone and JPL to the DSN GCF. Additionally, the GCF provides communications services between the DSSs at the communications complex at Goldstone, California, via intersite microwave systems and between separated areas of the NOCC at JPL via 230 kbit/s wideband data channels. Also, voice communications are provided within the complexes, and within the NOCC The GCF is comprised of seven subsystems Voice, Teletype, High-Speed Data, Wideband Data, GCF Monitor and Control, Data Records, and Network Communications Equipment The DSN Telecommunications and Data Acquisition Engineering Office of JPL provides the technical direction and systems management of the GCF and acts as the representative of NASCOM for switching and interconnect functions on the west coast.

II. GCF-NASCOM Interrelationships

The interrelationships at the programmatic level between JPL's DSN GCF and the NASCOM Network, which is managed, engineered, and controlled at GSFC, are characterized as follows

NASCOM

(1) Provides long-haul operational ground communications in support of all NASA projects and mission activities including those supported by the DSN

(2) Accepts and supports communications requirements established by the DSN and validated through continuing consultation and review

(3) Establishes in consultation with the users the basic characteristics of the NASCOM systems, such as teletype line rate and block header formats for switching, and the user electrical interfaces

GCF

(1) Provides ground communications for all DSN missions and uses the services of NASCOM.

(2) Establishes additional characteristics of all GCF subsystems on an end-to-end basis such as block multiplexing, error correction, GCF monitoring and control, and data records capabilities

III. Objectives and Goals

The primary objectives of the GCF are to provide highest quality point-to-point transfer of operational data within the DSN and to provide simple user and NASCOM electrical and operational interfaces These objectives are being met by·

(1) Providing automatic message switching and routing

(2) Providing data transmission subsystems that are as transparent to the user as possible

(3) Minimizing project-dependent equipment within the GCF.

(4) Providing a centralized common user data records capability.

The goals of the GCF are to provide highly reliable and cost-effective data transmission while continuing an adequate capability balance for multiple mission users. These goals include the following

(1) Equipment and routing redundancy to minimize single-point-of-failure impact

(2) End-to-end performance which provides essentially block-error-free throughput

(3) Design coordinated and consistent with the NASCOM Development Program.

IV. Configuration and Functional Subsystem

The current GCF configuration, including the related NASCOM interfaces and functions, is illustrated in Fig 1. This configuration illustrates the long-haul communication circuit services external to JPL and Deep Space Communications Complexes which are the responsibility of NASCOM (except circuits between the Goldstone Complex and JPL) The point-to-point communications for Voice, Teletype, High-Speed Data, Wideband Data, and GCF Monitor and Control Subsystems are serviced by this Fig 1 configuration

A. High-Speed Data Subsystem

This subsystem consists of the GCF assemblies that switch, transmit, record, process, distribute, test, and monitor digital data and is used for transmission of

(1) All digital data for the DSN Command, Tracking, and Monitor and Control Systems.

(2) All low or medium rate data of the DSN Telemetry, Radio Science, Very Long Baseline Interferometry (VLBI), and the DSN Test Support System

The High-Speed Data Subsystem provides a capability for transmitting and receiving the serial bit stream block formatted data over a properly conditioned full duplex alternate voice/data channel having a 3 0-kHz bandwidth or over a time division multiplexed 56-kb/s satellite circuit This serial bit stream is impressed on communication circuits at a continuous line bit rate divided into message segments referred to as high-speed data blocks. Two types of data blocks are used

(1) Data blocks containing user data bits to be transmitted

(2) Filler blocks containing filler data bits provided by the GCF when the user data rate is insufficient to maintain contiguous blocks on line required for continuous line monitoring and error control.

Current capabilities for the GCF Mark III period provide the functional capabilities illustrated in Fig 2 The GCF High-Speed Data Subsystem is standarized on a 1200-bit block size (message segment) and a line bit rate of 7200 b/s The subsystem capabilities include a 22-bit error detection encoding/decoding polynomial code, two error status bits, and error control using two eight-bit fields The error control field facilitates numerical serialization and acknowledgement numbers for error correction by selective retransmission (The error correction capability significantly reduces the post-pass time required for non-real-time replay of blocks received in error) Figure 3 illustrates the High-Speed Data Subsystem and the GCF Subsystems located at the Central Communications Terminal

B. Wideband Data Subsystem

The Wideband Data Subsystem consists of assemblies that switch, transmit, receive, process, distribute, test and monitor data requiring the use of bandwidths greater than those provided by standard high-speed data channels The GCF Wideband Data Subsystem functionally illustrated in Fig. 4, together with a listing of functional capabilities provided, includes standard wideband circuits as well as all intersite

microwave (area microwave) capabilities The Wideband Data Subsystem is used for the transmission of

(1) All DSN Telemetry System high-rate data that exceed High-Speed Data Subsystem capabilities

(2) Data interchange between the NOCC and GCF Communications Terminal at JPL

(3) Data interchange between DSSs within a complex via intersite microwave, including critical timing signals and receiver baseband signals for antenna arraying and signal combining systems support

(4) Simulation System Data from the Mission Control and Computing Center/Mission Operations Center to the DSSs

(5) DSN Test Support System data from the Network Operations Control Center to the DSSs.

The wideband data circuits for interchange of data between the DSSs and JPL are impressed with serial bit streams at a continuous line rate, typically 56 or 230 4 kb/s, divided into 4800-bit message segments (data blocks) Similar to the High-Speed Data Subsystem, the blocks are either data blocks or filler blocks inserted when the user data load is insufficient to maintain contiguous data blocks

C. Voice Subsystem

The Voice Subsystem consists of GCF assemblies that switch, transmit, receive, distribute, test, and monitor transmissions originally generated in vocal form, and includes internal voice communications within the Deep Space Station Communications Complexes, DSSs, and the NOCC. The subsystem service provides capabilities between those areas and to non-DSN area interfaces as follows

(1) NOCC and DSS

(2) NOCC and MCCC/MOC (or remote MOC).

(3) MOC and DSS for Command System backup

The Voice Subsystem functional capabilities and key characteristics include

(1) Standard voice-data grade circuits for all traffic

(2) Conferencing capability on one intercontinental circuit during noncritical periods for all Deep Space Stations supporting a single project (individual circuits for each DSS during critical periods, resources permitting)

(3) User-controlled intercomm switching

(4) Circuits used for high-speed data transmission (backup) if required

(5) Voice traffic recording in the Central Communications Terminal upon request

D. Teletype Subsystem

The teletype (TTY) subsystem uses an eight-level ASCII (American Standard Code for Information Interchange) national standard. The subsystem consists of assemblies that switch, transmit, receive, distribute, test and monitor digital signals at a TTY line rate of 100 words per minute The operational use of teletype continues to be deemphasized and is used primarily for emergency backup operational transmissions and administrative communications Service functions and key characteristics include

(1) Handling Eastern Space and Missile Center (ESMC) generated predicts for DSN initial acquisition

(2) Transmitting nonoperational messages between the JPL Message Center and other locations

(3) Use of standard NASCOM format and the NASCOM communications processor for message switching

(4) Employment of time division multiplexing techniques to reduce trunk circuit costs

E. GCF Monitor and Control Subsystem

The GCF Monitor and Control Subsystem consists of assemblies that collect, process, and display the status and performance of the GCF Subsystems in real-time The GCF Monitor and Control Subsystem functional capabilities are illustrated in Fig. 5. Functions are implemented in minor subassemblies located at each DSS to interface station GCF status and performance indicators to the CMF for monitor block formatting and transfer to the Central Communications Monitor (CCM) Processor at JPL. The CCM also receives real-time status and performance information from local GCF subsystems. All real-time status and performance information received by the CCM is processed and displayed relative to preset standards and limits to facilitate operations monitoring and technical control Information and alarms are displayed on continuous line performance and data flow throughput including error control

F. Data Records Subsystem

The DSN requirements for the data record processing and production functions are implemented in the GCF Data Records Subsystem The Data Records Subsystem consists of assemblies in the CCT that log in real-time, monitor, identify gaps, provide for processing and editing of data gap lists, control data gap recalls from the DSSs and generate and account for Intermediate Data Records (IDRs) and selected fill-data records

The Data Record Subsystem maintains accountability of high-speed and wideband data, performs automatic recall (under operator control) of missing data, and generates near-real-time Intermediate Data Records for delivery to the Mission Control and Computing Center

G. Network Communications Equipment Subsystem

The Network Communications Equipment (NCE) Subsystem consists of GCF minicomputers and peripheral I/O assemblies that switch, transmit, and receive data The NCE Subsystem assemblies are located in the Network Data Processing Area (NDPA) The NCE assemblies comprise a GCF Area Communications Terminal located in the NDPA

The NCE interchanges multiplexed block formatted data with the ECS assembly located in the CCT over full-duplex 230-kb/s wideband data channels (see Fig 6). This GCF-NDPA interface function provides for

(1) Processing data for transmission to and accepting data from the GCF CCT

(2) Multiplexing/demultiplexing and buffering data for all NDPA processors

(3) Routing data to and from NDPA processors.

V. Typical Configuration

The DSN GCF is designed for multiple mission support Improvements and additions are integrated to meet new era and project requirements Support for the Voyager, Pioneer 6 through 12, Viking and Helios missions is continued The capability for West Coast Switching Center support of the IRAS Project has been completed Requirements for the AMPTE, ISPM and Galileo projects are being defined and included in ongoing plans Figure 7 illustrates, in general, the GCF configuration or support of these projects Additionally, remote information centers and other non-DSN NASCOM-serviced installations on the west coast are serviced through the NASCOM West Coast Switching Center, an integral part of the GCF 20/Central Communications Terminal at JPL.

VI. Implementation Activities

Implementation planning and design to meet the requirements of the DSN's Mark-IV long range plans were begun in late 1979. The priority design and implementation effort has been to consolidate the data communications interfaces to a single communications processor interface at each of the three Deep Space Communications Complexes (DSCCs) This single interface processor will accommodate the DSN's Networks Consolidation Program plans to simultaneously support multiple spacecraft from a single signal processing center located at each DSCC Other less significant GCF changes and additions required within the DSCCs along with changes at the CCT are included in this effort to be completed in 1985 With the consolidation of the networks, the GCF will see a significant increase in data communications bandwidth requirements, as both deep space and high elliptical near-earth orbiting spacecraft are tracked, and acquired data will pass through the single communications processor at the DSCC. Modifications will be made in the CCT to accommodate the new DSCC interfaces and the new type of remote operations control centers

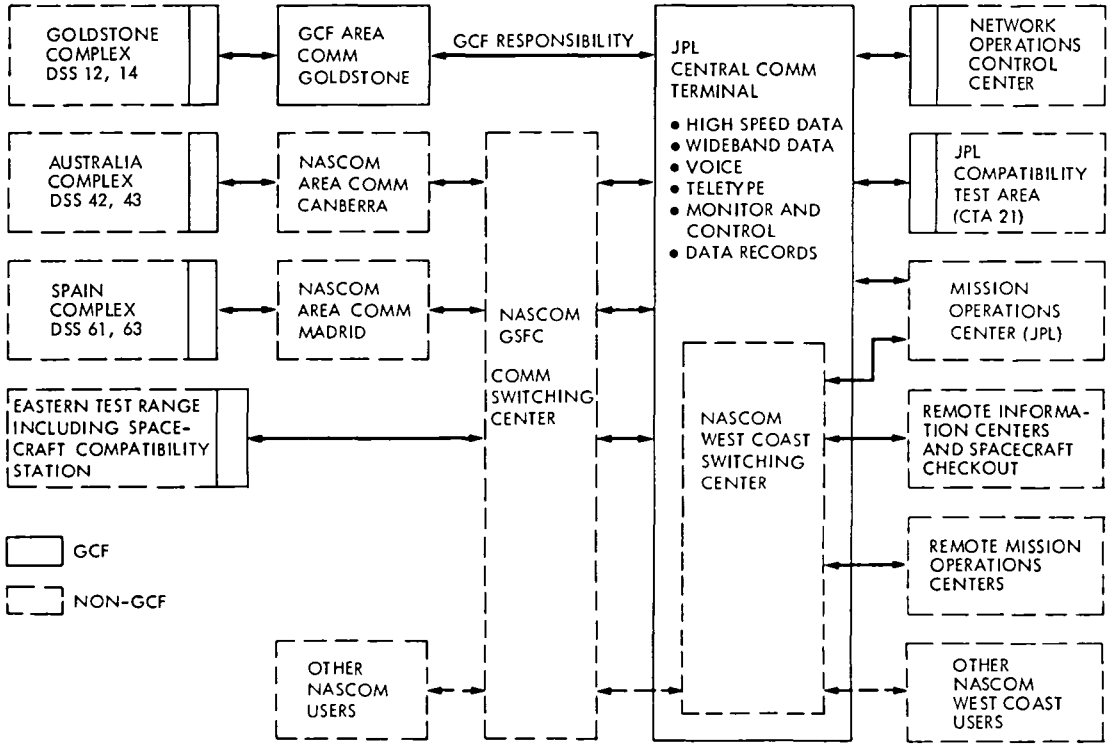


Fig 1. GCF configuration

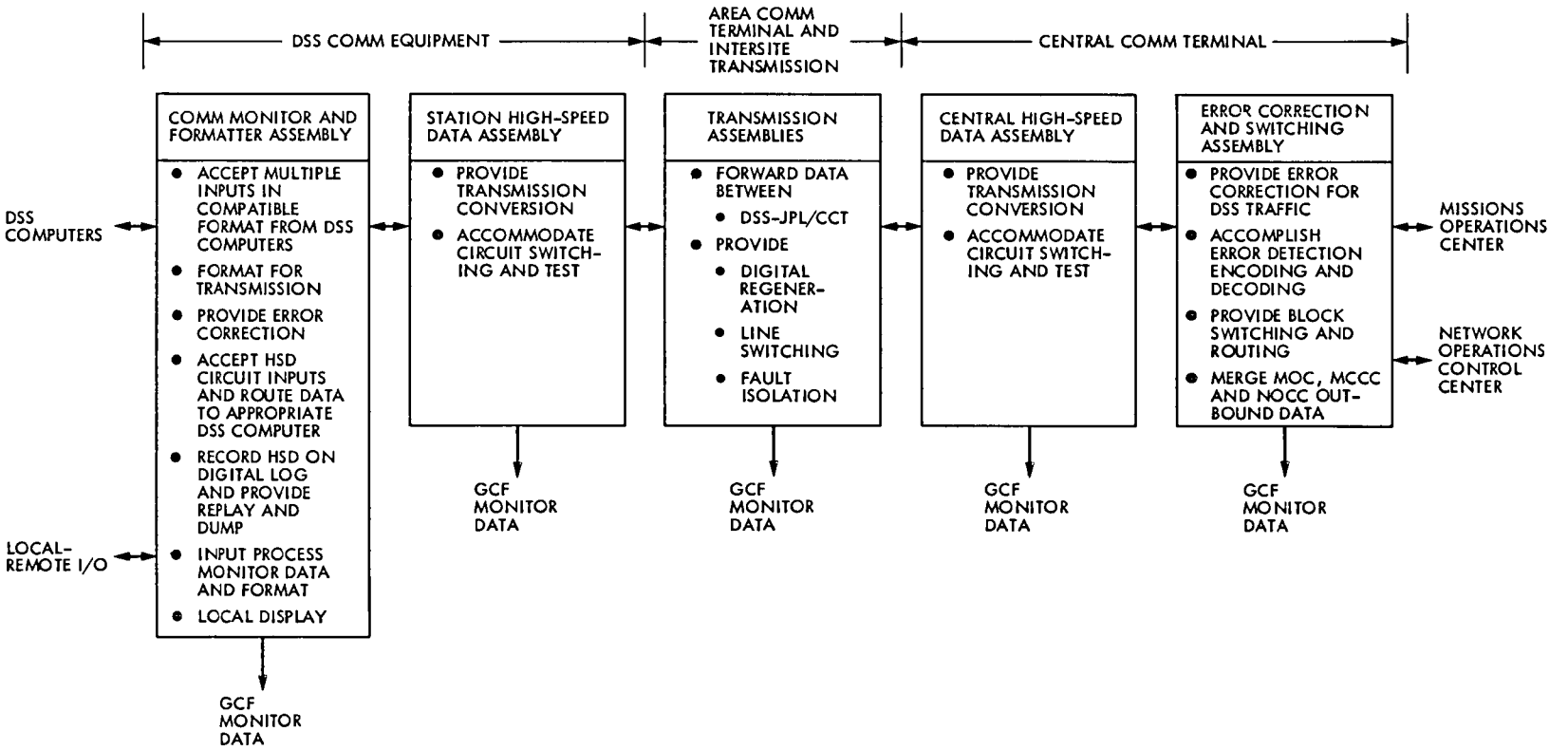


Fig. 2. GCF High-Speed Data Subsystem functional capabilities

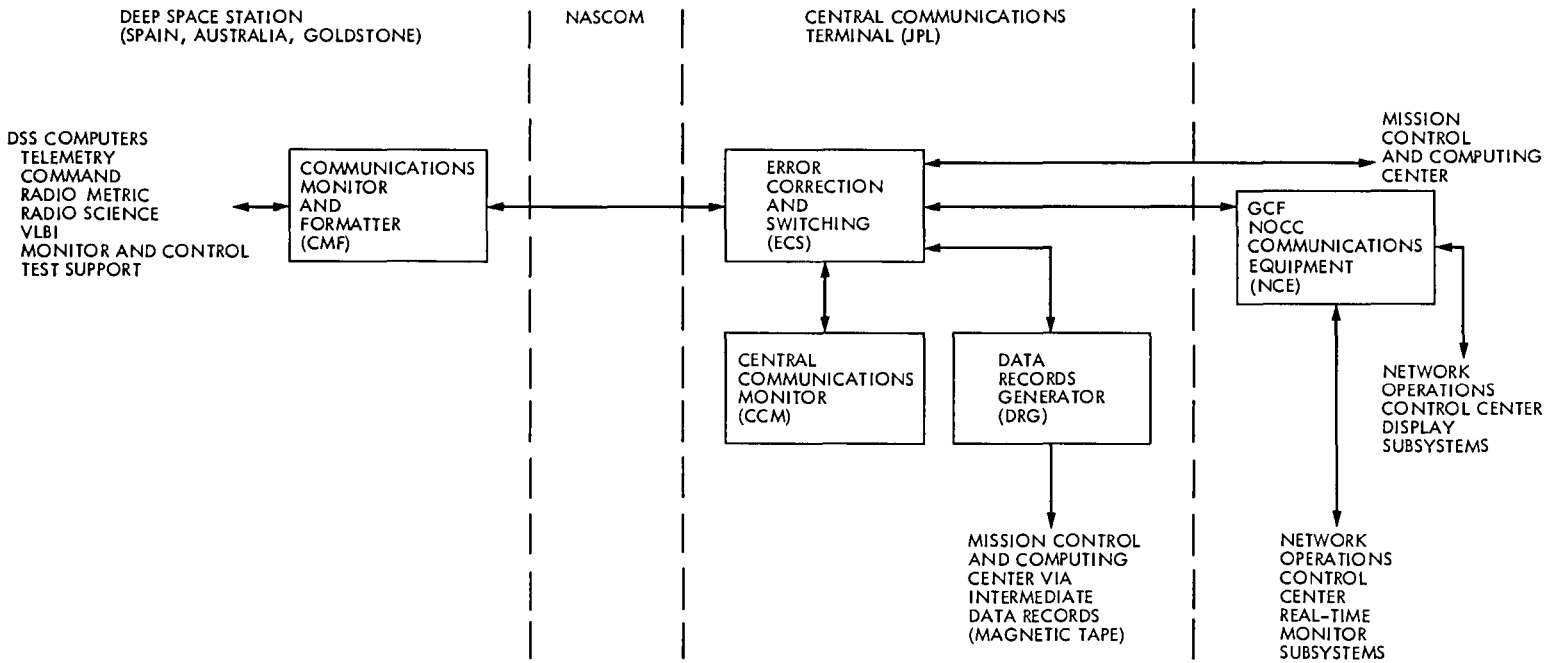


Fig. 3. GCF High-Speed Data Subsystem and Central Communications Terminal Subsystems and interfaces

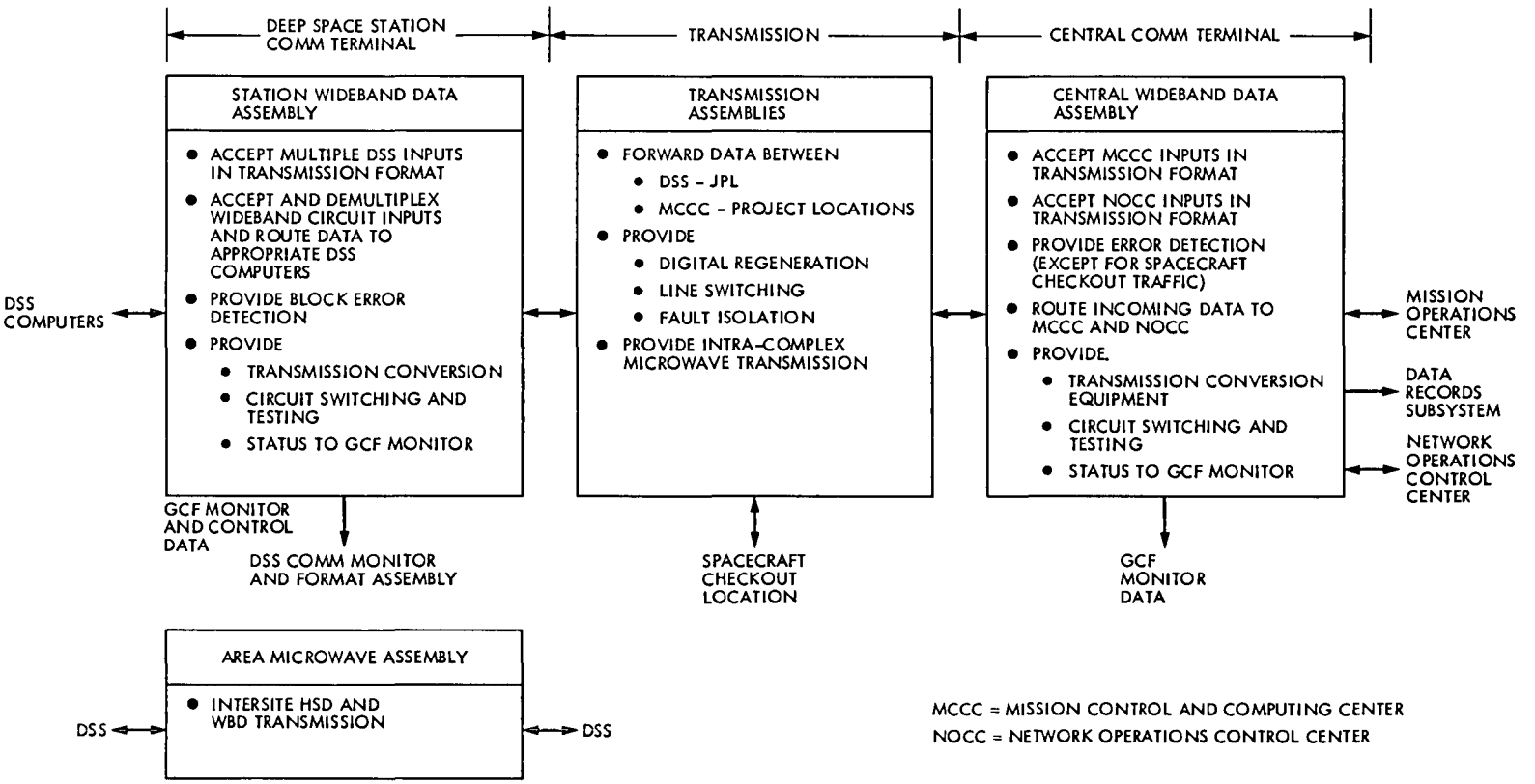


Fig 4 GCF wideband subsystem

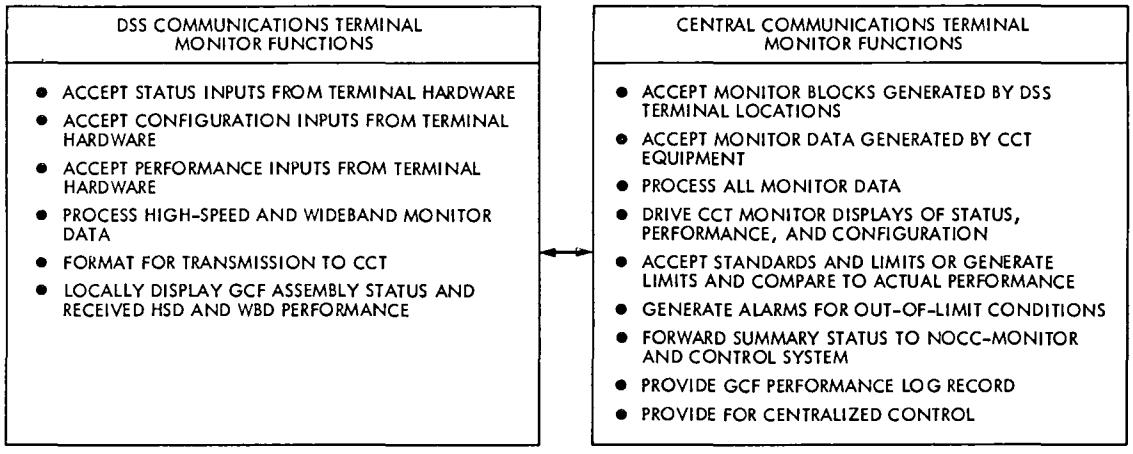


Fig. 5 GCF Monitor and Control Subsystem functional requirements

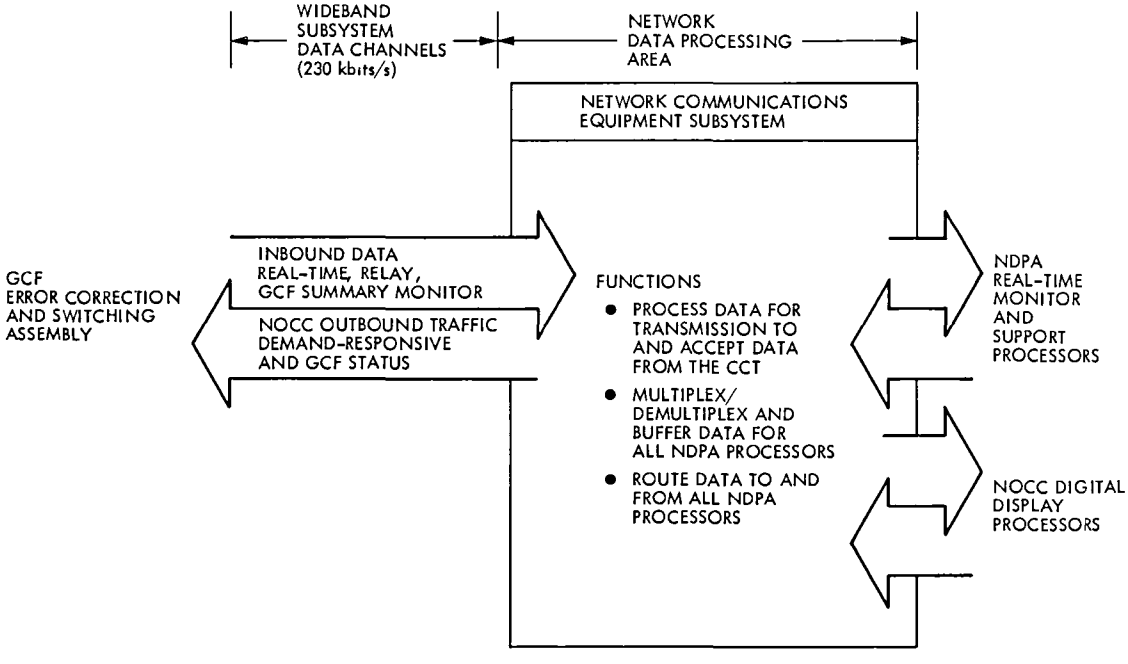


Fig 6 GCF Network Operation and Control Communications Equipment Subsystems

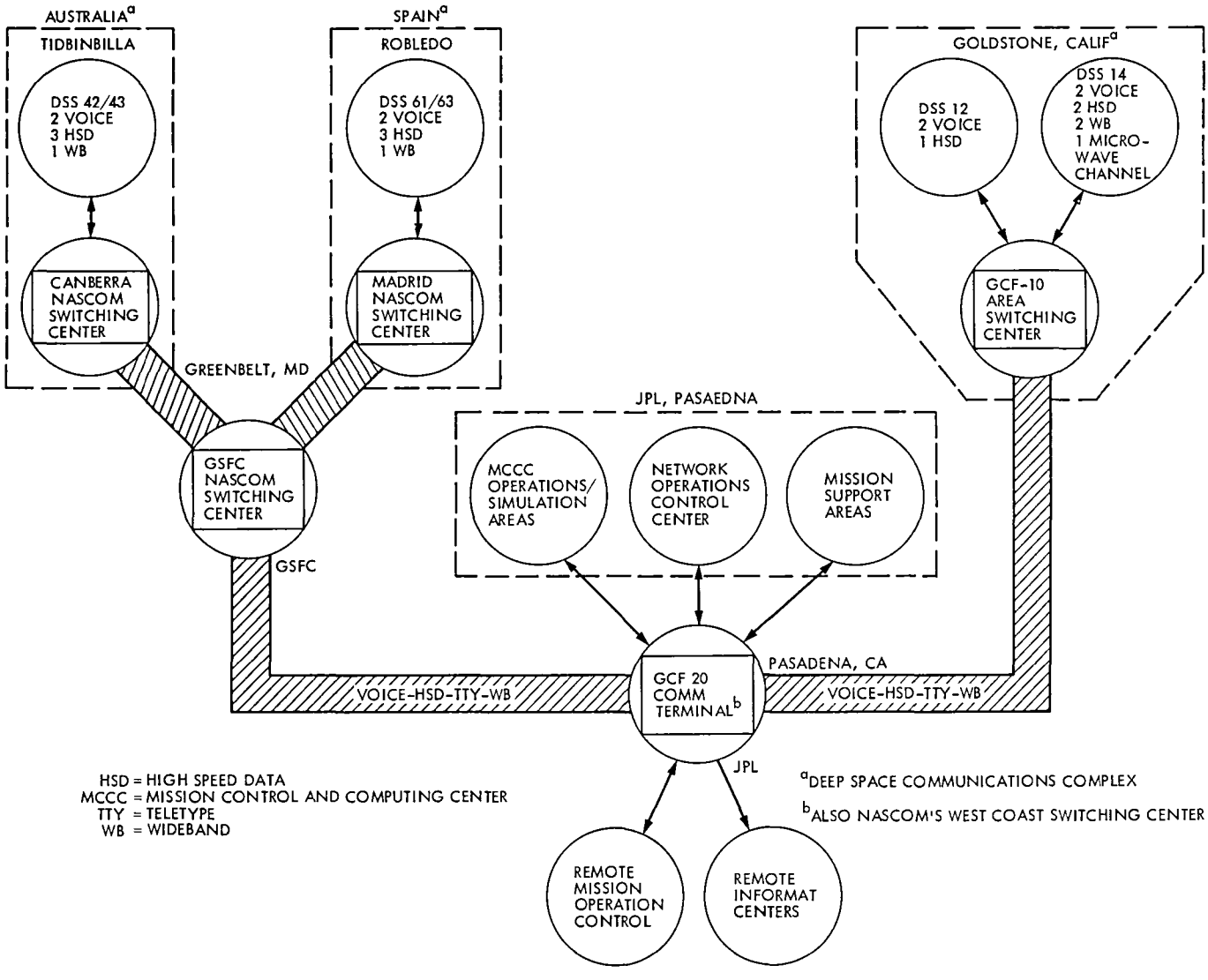


Fig. 7. DSN support locations and GCF-NASCOM circuit requirements

MV-3 Communication Subsystem

G S Parks, R S Markham, and T J Rowden
Tracking Systems and Applications Section

M L MacMedan
Mission Information Systems Engineering Section

The Operational Mobile VLBI Data Acquisition System (MV-3) being designed by JPL requires that a voice link be established between the mobile data system and a fixed site A communications subsystem was incorporated in the MV-3 design which consists of HF radio, VHF mobile radio telephone and conventional land line telephone The HF antenna design was optimized for short and long range transmission using both inverted V and yagi antennas mounted on a self-supporting telescoping mast

I. Introduction

The Operational Mobile VLBI Data Acquisition System (MV-3) is being implemented by JPL in support of the NASA Crustal Dynamics Project MV-3 is also known as Operational Radio Interferometry Observing Network (ORION) and consists of an electronics van and 5-meter dish antenna with transporter Operating in conjunction with one or more VLBI data acquisition systems, it will provide a means of measuring length and orientation of vectors between sites in areas of geophysical interest The technology used to accomplish this is Very Long Baseline Interferometry (VLBI, Ref 1)

A communications subsystem (COM) was incorporated to establish a voice link between the mobile data system and a fixed site This is necessary to coordinate experiment start times and help solve problems as they arise during setup or during data acquisition Reliable communications are also necessary for safety reasons while MV-3 is operating in remote locations.

A single control station will be used to coordinate all mobile and fixed data acquisition systems during an experiment Therefore, MV-3 need only communicate with this one station which can act as a relay for messages to other stations

II. Requirements

The MV-3 Design Requirements Document (an internal JPL document) forms the basis for COM requirements Generally, COM must include those transceivers and antennas necessary to provide 90% reliable voice contact communications to the control station from any place in the continental United States Practical reasons dictated the limitation to the continental U S in that it would be difficult to design a system around all possible foreign regulations and systems available Foreign operation may therefore require tailoring the COM capability to meet the requirements

Because MV-3 is a mobile system, COM must also meet the following requirements

(1) COM components must survive road vibrations and stress encountered while in transit

(2) Equipment mounted to the exterior of the electronics van must meet provisions set down by state vehicle codes The design requirements specifically prohibit the use of equipment which would require that MV-3 obtain special road permits.

(3) COM must meet its reliability requirement under a wide range of terrain conditions Mountains and valleys must be considered in the design.

(4) The setup and teardown of the COM subsystem should be held to a practical minimum Ease of operations can considerably lower cost to operate MV-3

III. Tradeoff Studies

The following seven modes of communication were considered for their merits in optimizing the MV-3 COM design.

(1) Land lines.

(2) Low and medium frequency radio

(3) High frequency radio

(4) Repeater communications

(5) VHF radio telephone.

(6) Satellite microwave

(7) ATS satellite.

A. Land Lines

The land line system considered here is that which is commercially available from the telephone company Once linked, it achieves better than the 90% reliability imposed on COM and is rarely affected by unpredictable environmental conditions

The land line system has two drawbacks First, it depends on the availability of the lines Some remote locations would require a special hookup, raising the cost of the system Second, it poses a logistics problem Arrangements must be made weeks in advance for the connection Also, a representative of the user must be available at the site to coordinate the placement of the lines The land line system is therefore practical only in cases where other systems cannot be used, or when the length of stay at a particular site warrants its installation

B. Low and Medium Frequency Radio

Radio communication is generally a practical alternative to land lines Each carrier frequency subdivision has its own main characteristics, as summarized in Table 1 Low frequency (LF) radio is generally good for medium and long range transmission, either by traveling close to the ground (ground wave) or reflecting off the ionosphere (sky wave). Medium frequency (MF) transmission is good for short range communications such as local AM broadcasting At night, MF can sometimes bounce off the ionosphere, which increases its range Picking up a faraway AM broadcasting station at night is a typical example of this

One serious disadvantage of LF or MF is that the antenna is extremely large The length of an efficient antenna is typically either 1/4 or 1/2 wavelength, which implies that the antenna would be tens or hundreds of meters long This is practical only in fixed locations where such an antenna could be permanently installed

C. High Frequencies

High frequency (HF) transmission is generally good for short distance ground wave and long distance sky wave communication Figure 1 shows the typical range for HF, which has a wavelength from 10 to 160 meters With multiple hops, HF will also transmit over intercontinental distances Ground wave transmission is usually useable up to 320 km, but depends greatly on terrain, while sky wave transmission works at distances greater than 160 km.

HF has two drawbacks First, it still requires a large antenna. Typical HF antennas are from 3 to 60 meters in length depending on the frequency Second, proper ionospheric conditions are required for reliable transmission. HF alone would not meet the reliability requirement imposed on COM

D. Repeater Communications

Frequencies above the HF band are generally good only for ground wave transmission and require repeaters to extend its range. Frequencies in this range include very high frequencies (VHF), ultra high frequencies (UHF), and microwave. Repeater stations are usually on mountain tops, and receive the incoming signal while retransmitting the same signal to a receiver that would otherwise be out of range of the first transmitter. The advantages of this system are First, these frequencies are only negligibly affected by atmospheric or environmental conditions Repeater communications are therefore very reliable as long as the user's transceiver is in range of a repeater Second, the antennas for this mode of transmission are small, typically a few feet, which simplifies design and operations The disadvantage of this scheme is the requirement for costly repeater systems throughout the area of anticipated use This system would only be practical for COM if existing repeater systems are used

E. VHF Radio Telephone

Mobile VHF radio telephones have all of the characteristics of a repeater communications system A critical advantage is

that radio telephones use a repeater network owned and operated by the telephone company and its contractors Through mobile telephone charges, repeater time is rented and shared by multiple users, bringing the cost down to an affordable value This system is designed around multiple channels shared by mobile telephone users in the immediate area Direct dial and automatic scanning features will be available this year, which will eliminate the requirement that channel number be coordinated before the call is placed

The disadvantages of radio telephone are First, the system is not designed to reliably cover remote areas, particularly off-highway Also, mountains in the radio path to the repeater can reduce its reliability considerably Second, the radio telephone channels tend to be overcrowded in populous areas Some waiting period could be encountered before contact is made

F. Satellite Microwaves

As the frequency spectrum has become crowded over the years, microwaves have become popular for transmitting information and messages Microwaves are negligibly affected by atmospheric conditions, but because of their short wavelength (< 0 3 m), transmission is in a strict line of sight path and thus requires repeaters to extend the range Satellite repeaters are practical for this purpose because of the large area of the earth covered by each satellite Numerous microwave transceivers can beam signals to a satellite repeater for reliable short or long range transmission

There are two considerable disadvantages of satellite microwave systems for MV-3 First, a large dish antenna (3 to 5 meters) must be pointed at the satellite for transmission MV-3 already incorporates a dish antenna, and to add another would raise the costs significantly and decrease its mobility and versatility Second, renting satellite time is costly

G. ATS Satellite

NASA-Goddard Space Flight Center is currently coordinating the use of two VHF communications satellites (ATS) which are available to government experimenters on a priority basis as determined by NASA When available, these satellites would provide reliable communication for MV-3 operations There are, however, three serious disadvantages First, use of the satellites must be coordinated well in advance of anticipated use This generates manpower requirements to coordinate the logistics and also disqualifies the system for emergency use Second, since ATS is shared by other users, there may be times when the system is unavailable to MV-3 Third, ATS is an experimental satellite with no guarantees that it will continue to function for an extended period of time

IV. Systems Chosen

From the characteristics of each system, it was decided that COM incorporate the following communication modes (1) land lines, (2) VHF mobile radio telephones, (3) HF radio Normally the VHF mobile radio telephone will be used where available because of its ease of use The HF radio will back up the radio telephone as necessary If a long stay at a site is anticipated, MV-3 will be configured with standard telephones so that a direct land line connection is possible

Both the mobile radio and land line telephones are standard telephone company installations A 1-m whip antenna will be installed on the van roof with a cable leading to the VHF transceiver mounted in an equipment rack MV-3 is equipped with a 12-Vdc battery system which will power the mobile telephone This system will therefore be useable before the MV-3 generators are started

The HF radio requires multiple frequencies on which it will operate to ensure reliable transmission over a wide range of distances (Fig 2) The lowest frequency (channel) is determined for its ability to transmit over short distances Higher frequency channels should progress in octaves to a channel with long range transmission characteristics. COM will use the frequencies issued to NASA (namely 3 395, 6 9825 and 14 455 MHz) which have the necessary qualifications

V. HF Antenna Tradeoffs

A well designed antenna is critical for reliable HF communication Because of its size, it also warrants the most attention to optimize ease of operations The following antenna types were considered for tradeoffs for the HF system (Fig 3) (1) whip, (2) horizontal dipole, (3) inverted V dipole, (4) yagi

A. Whip

The whip antenna has two advantages First, as can be seen from its radiation pattern (Fig 4), the whip transmits best on horizontal paths This quality is excellent for transmitting ground wave and long range (> 750 km) sky wave In both cases the radiation angle (angle between the main transmission lobe and horizon) is small Second, the whip transmits in all horizontal directions (omnidirectional), eliminating a need for antenna pointing

The whip has two disadvantages First, it does not transmit well at high angles required for short distance sky wave transmission (150-750 km) This mode is usually necessary in mountainous areas when ground wave communication is not possible Second, the whip requires a resonant ground plane for optimum performance which would be difficult to set up

The electronics van roof might be used as a substitute but would not provide optimum performance

B. Horizontal Dipole

The horizontal dipole transmits well both horizontally and vertically at all but very low radiation angles Normally one thinks of the horizontal dipole as having a figure-eight shaped radiation pattern, which is only true in free space. A computer-generated mathematical model (Ref 2) simulated the radiation patterns as a function of antenna height These are shown in Fig 5 and were confirmed by antenna range tests This antenna is generally good for short and medium range (0-1800 km) transmission using both ground and sky waves Another advantage of the dipole is that no ground plane is required for good performance

A disadvantage of the horizontal dipole is that it requires two long support towers, which would be difficult to erect In order to achieve the desired radiation angle, both towers must be more than 10 km high (This antenna height also helps match the antenna impedance to the 50 ohm transceiver as shown in Fig 6) Further difficulties arise in positioning the towers for cases in which the electronics van is parked near a hill

C. Inverted V Dipole

The inverted V dipole is different from the standard dipole only in that the antenna ends are brought down close to the ground A similar computer-generated antenna pattern (Fig 7) shows that the inverted V, like the horizontal dipole, transmits well horizontally and vertically Like the horizontal dipole, the inverted V does not require a ground plane and will transmit reliably up to distances of about 1800 km

An important advantage of the inverted V over the horizontal dipole is that it requires only one high support tower This tower can be mounted on the electronics van with the dipole ends brought down to near ground level Because the inverted V is nearly omnidirectional, any convenient orientation can be used The antenna ends can be held in place by tripods or sandbags, depending on terrain Also, because of the antenna shape, the inverted V requires less space than the horizontal dipole

D. Yagi

The yagi antenna is a very directional antenna which focuses almost all of the radiation energy in one direction (Fig 8) Because of the large gain in signal strength in the one direction, the yagi will exceed the performance of other antennas when aligned properly with the point of contact Unfortunately, the yagi is very large for HF frequencies Fixed-installation yagi antennas are rarely constructed for frequencies less than 6 MHz Because of the mobility requirements of MV-3, the yagi will only be practical with the 14 455-MHz channel Even at this frequency the boom length is about 25 feet, which is almost the limit of what the MV-3 crew could easily erect Fortunately, the 14 455-MHz channel will usually only be used for long distance communication when the extra antenna gain is needed Therefore the yagi is only practical in this case for long range transmission (> 1800 km)

VI. Antenna Selection and Implementation

Using the previous information it was decided that COM incorporate both the inverted V dipole and the yagi antennas For most applications of short to medium range contact (0-1800 km), the omnidirectional inverted V will be used The yagi, which will be more difficult to erect, will be used only for long range contact (> 1800 km), using the 14 455 MHz channel If conditions require that other channels be used for long range contact, the inverted V will be available

Tuned resonant traps in the inverted V will eliminate the need for separate antennas for each channel These traps are resonant L-C circuits which essentially "divorce" the wires beyond the traps at the appropriate frequency The traps are positioned in the antenna so that each operating frequency will be matched properly to the antenna The overall length of the antenna is 43 2 m, which is approximately 1/2 wavelength at 3 395 MHz

Both the inverted V and the yagi antennas are supported by a pneumatic telescoping mast as shown in Fig 9 The mast is available as a catalog component from Wil-Bert Company and is self supporting and easy to erect Mobile television systems are the prime user of the mast for transmitting microwave video information locally The mast is to be mounted to the van rear door with a separate air compressor mounted near the van suspension Activating the compressor extends the mast from 2 2 to 14 8 meters The entire mast is rotatable by hand for yagi antenna alignment

As can be seen by Fig 8, the yagi favors a high mounting position for long distance communication As the height of the yagi is increased, the radiation angle decreases, which is desirable to help overcome the barrier of the earth's curvature The 14 8 m mast plus 1 2 m of trailer suspension holds the yagi approximately 3/4 of a wavelength above ground

The inverted V antenna is supported on the ends by two tripods to reduce the capacitive effect of the ground near the antenna ends If the terrain makes this impractical, the antenna ends can be supported by rope tied to sandbags

References

1 Brunn, D L , McLaughlin, F D , Sweetser, B M , Thom, E H , and Wu, S C , "ORION Mobile Unit Design," in *TDA Progress Report 42-60*, Jet Propulsion Laboratory, Pasadena, Calif , December 1980

2 Covington D W "Radiation Patterns of V Dipoles over Perfect Ground," *QST*, April 1970

Bibliography

Bennington, T W , *Short-Wave Radio and Ionosphere*, W & J MacKay & Co , Ltd , Chatham, Great Britain, 1950

Sinnema, W *Electronic Transmission Technology*, Prentice-Hall, Inc , Englewood Cliffs, N J , 1979

The A R R L Antenna Book, American Radio Relay League, Inc Newington, CN

Table 1 Main transmission characteristics of frequency classes

Classes	Frequency subdivisions	Main characteristics	Principal uses
Very long waves and long waves	VLF and LF	Wave travels to considerable distances over earth's surface and to great distances by reflection from lower edge of ionosphere	Medium and long distance point-to-point communication Long-wave medium distance broadcasting
Medium waves	MF	Wave travels over earth's surface to relatively short distances during day, at night some energy comes from ionosphere and range increases	Local broadcasting, marine and aircraft communication, direction finding
Short waves	HF	Wave travels up to ionosphere whence it is reflected back to earth Conditions vary greatly with time of day and season, but great ranges obtained if conditions favorable	Long distance broadcasting, point-to-point communication Amateur communication, etc
Very short, ultra short, and microwaves	VHF, UHF, microwaves	No ionospheric reflection Wave travels directly through lower atmosphere from transmitter to receiver	Short distance communication, FM broadcasting, television, radar, aircraft guidance systems Amateur communication

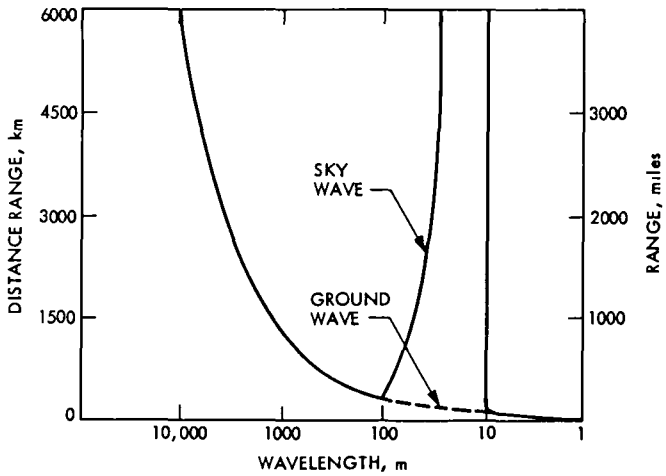


Fig. 1 Variation in range with wavelength

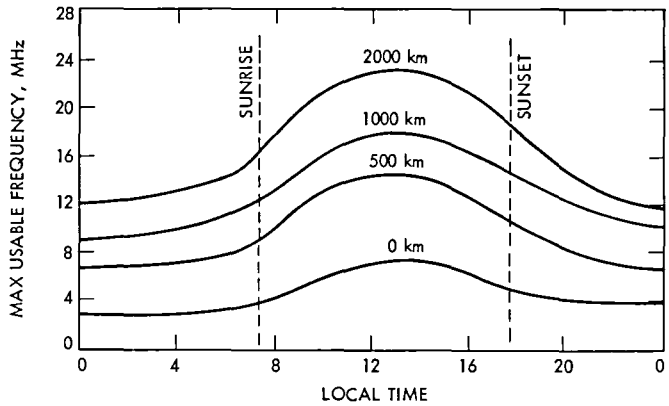
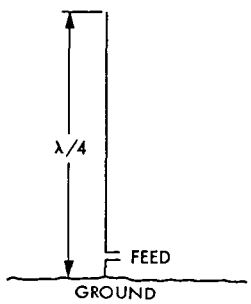
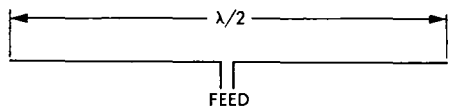


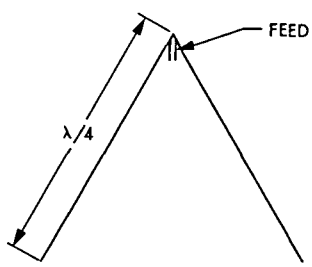
Fig 2 Maximum useable frequency chart



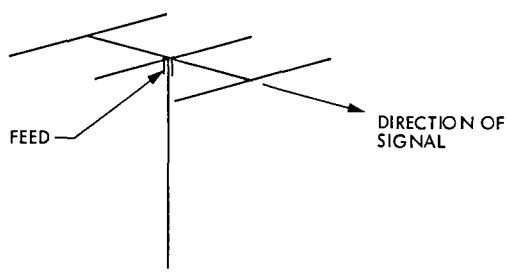
(a) WHIP



(b) HORIZONTAL DIPOLE

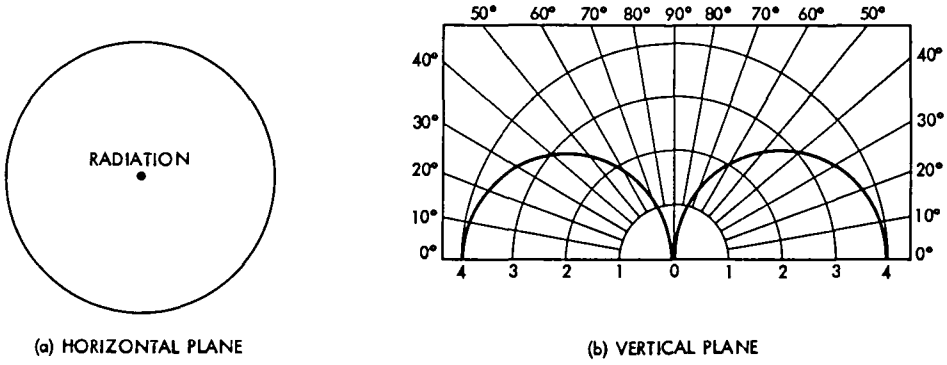


(c) INVERTED V DIPOLE



(d) YAGI

Fig 3. Antenna configurations



(a) HORIZONTAL PLANE

(b) VERTICAL PLANE

Fig. 4 Whip radiation patterns (assuming ideal ground plane)

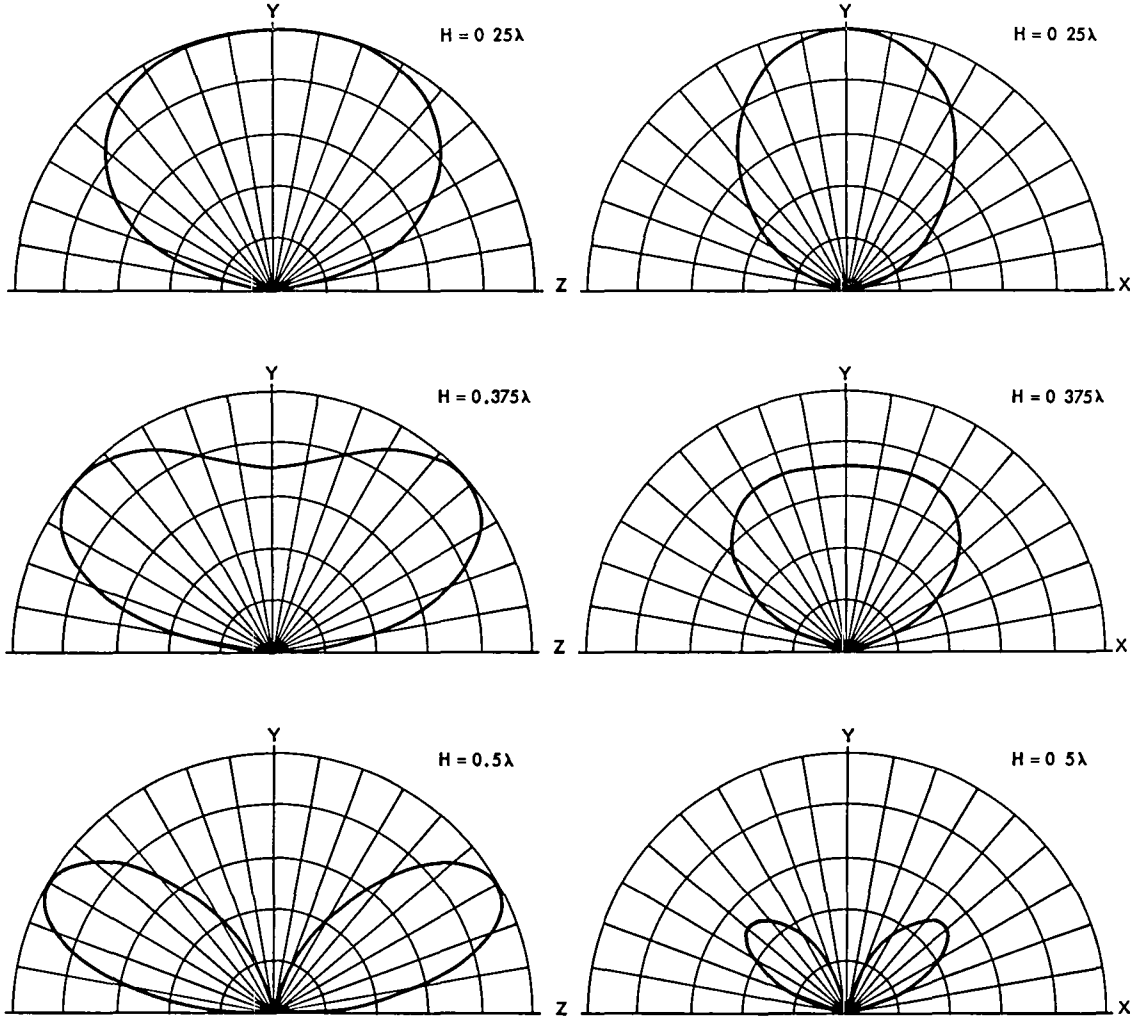


Fig 5 Antenna patterns off the end of (left) and broadside to (right) a horizontal dipole as a function of antenna height above ground (in wavelengths λ)

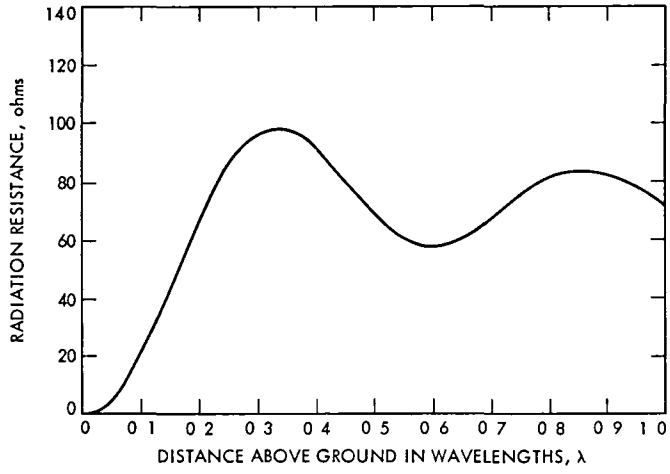


Fig. 6. Input resistance of a horizontal half-wavelength diple antenna at a height λ wavelengths above a ground plane

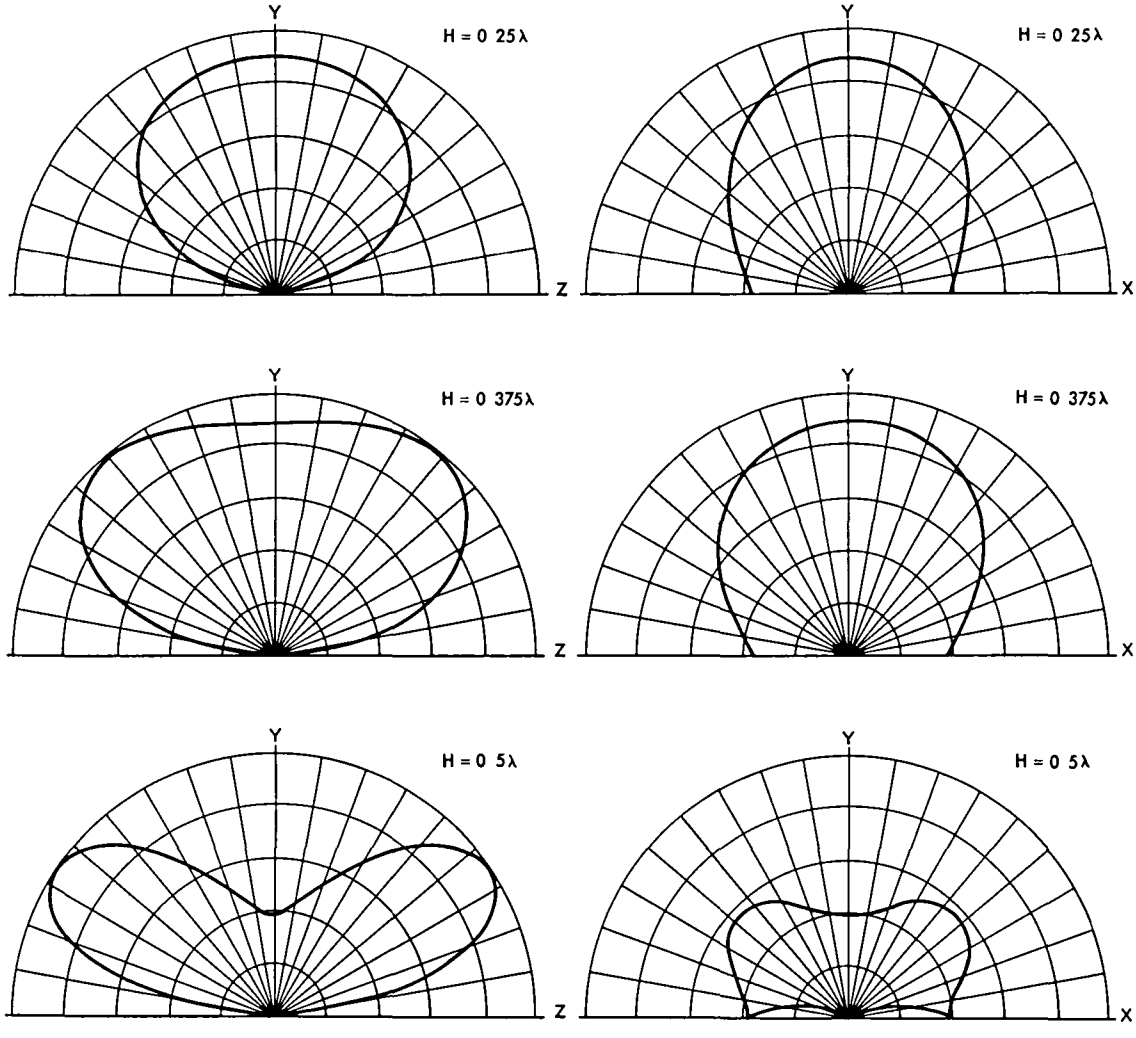


Fig. 7. Antenna patterns off the end of (left) and broadside to (right) an inverted V dipole as a function of antenna height above ground (in wavelengths λ)

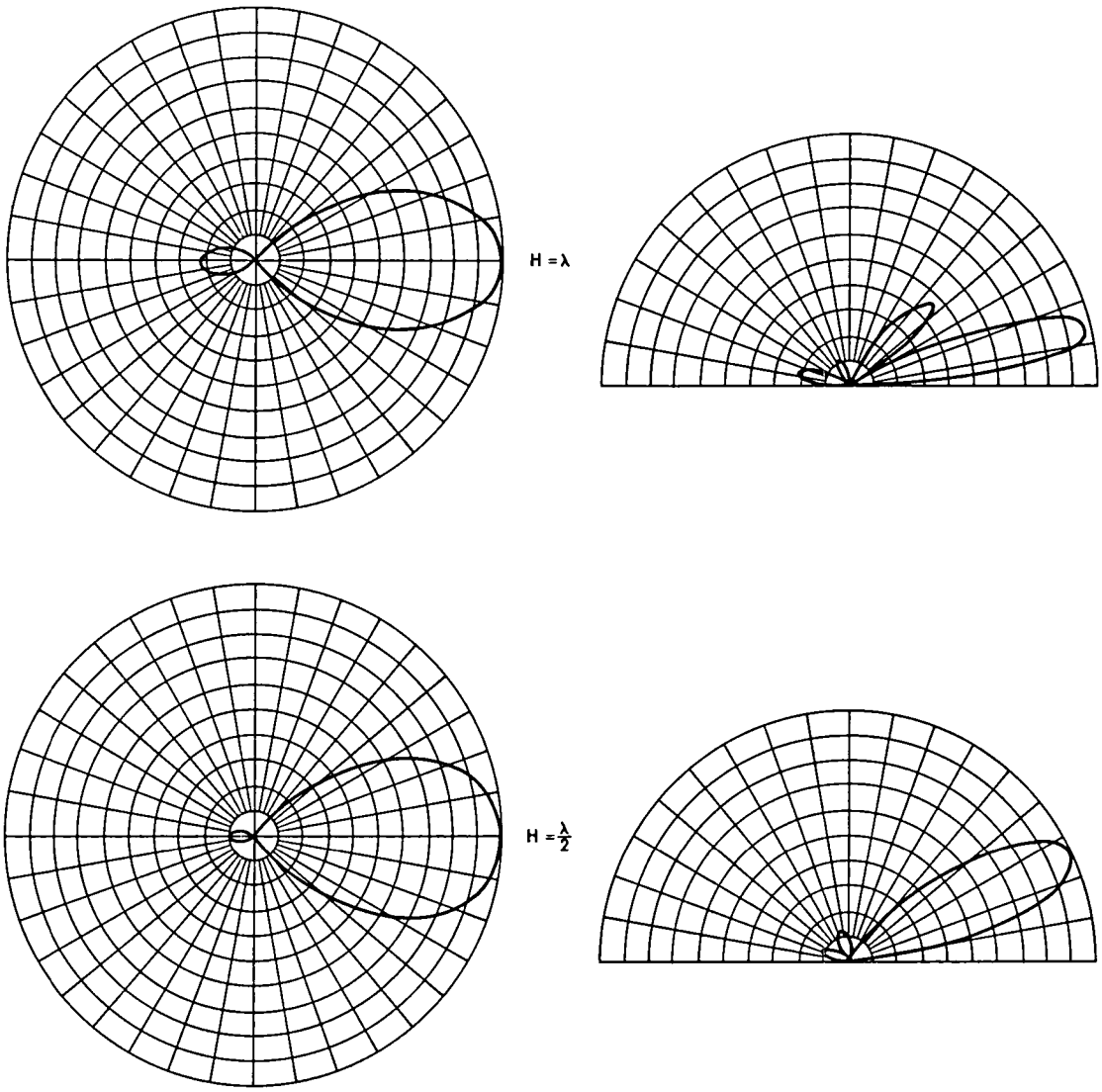


Fig 8 Yagi radiation patterns for horizontal plane (left) and vertical plane (right) as a function of antenna height (in wavelengths λ)

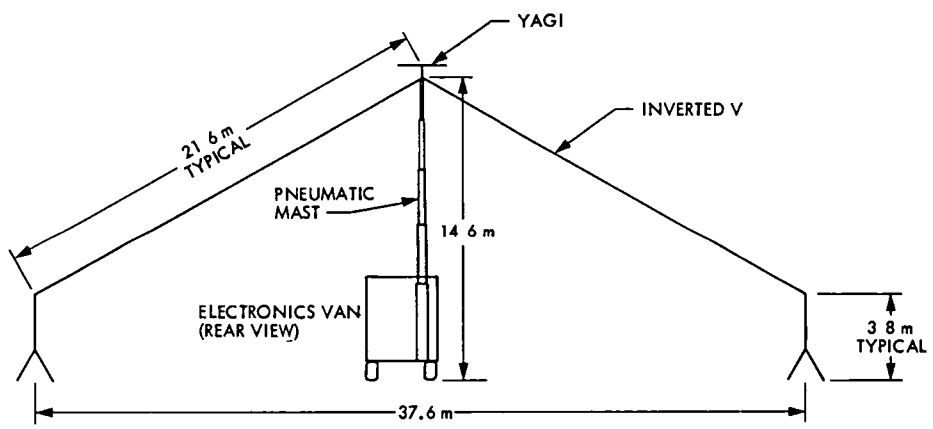


Fig 9. Antenna implementation

Bibliography

Ananda, M P, "Lunar Gravity A Mass Point Model," *J Geophys Res*, Vol 82, No 20, pp 3049-3064, July 10, 1977

Anderson, J D, et al, "Experimental Test of General Relativity Using Time-Delay Data From Mariner 6 and Mariner 7," *Astrophys J*, Vol 200, No 1 pp 221-233, Aug 15, 1975

Anderson, J D, et al, "Tests of General Relativity Using Astrometric and Radio Metric Observations of the Planets," *Acta Astronautica*, Vol 5, No 1-2, pp 43-61, Jan -Feb 1978

Barton, W R, and Miller, R B, *Tracking and Data System Support for the Pioneer Project Pioneer 11–Prelaunch Planning Through Second Trajectory Correction 'to May 1, 1973*, Technical Memorandum 33-584, Vol II, Jet Propulsion Laboratory, Pasadena, Calif, Mar 15, 1975.

Bartos, K P, et al, *Implementation of the 64-Meter-Diameter Antennas at the Deep Space Stations in Australia and Spain*, Technical Memorandum 33-692, Jet Propulsion Laboratory, Pasadena, Calif, Jan 15, 1975

Bathker, D A, Brown, D W, and Petty, S M, *Single- and Dual-Carrier Microwave Noise Abatement in the Deep Space Network*, Technical Memorandum 33-733, Jet Propulsion Laboratory, Pasadena, Calif, Aug 1, 1975

Bathker, D A, *Microwave Performance Characterization of Large Space Antennas*, Publication 77-21, Jet Propulsion Laboratory, Pasadena, Calif, May 15, 1977

Batty, M, Jauncey, D, and Gulkis, S, "Planetary Observations with the Tidbinbilla Interferometer," *Proc Astron. Soc Aust*, Vol 4, 72, 1980

Batty, M, Jauncey, D, Rayner, P, and Gulkis, S, "Evidence for Changes in the Microwave Brightness Temperature and Spectrum of Uranus," *The Astr J*, Vol 243, pp 1058-1061, 1981

Beatty, R W, and Otoshi, T Y, "Effect of Discontinuities on the Group Delay of a Microwave Transmission Line," *IEEE Trans Microwave Theor, Techniq*, Vol MTT-23, No 11, pp 919-923, Nov 1975

Benjauthrit, B, and Reed, I S, "Galois Switching Functions and Their Applications," *IEEE Trans Comput*, Vol C-25, No 1, pp 78-86, Jan 1976

Benjauthrit, B, and Reed, I S, "On the Fundamental Structure of Galois Switching Functions," *IEEE Trans Comput*, Vol C-27, No 8, pp 757-762, Aug 1978

Berlekamp, E R, et al, "On the Inherent Intractability of Certain Coding Problems," *IEEE Trans Inform Theor*, Vol IT-24, No 3, pp 384-386, May 1978

Berman, A L, and Ramos, R, "Pioneer Venus Occultation Radio Science Data Generation," *IEEE Trans Geosci and Remote Sensing*, Vol GE-18, No 1 pp 11-14, Jan 1980

Berman, A L, and Rockwell, S T, *New Optical and Radio Frequency Angular Tropospheric Refraction Models for Deep Space Applications*, Technical Report 32-1601, Jet Propulsion Laboratory, Pasadena, Calif, Nov 1, 1975

Berman, A L , *The Prediction of Zenith Range Refraction From Surface Measurements of Meteorological Parameters*, Technical Report 32-1602, Jet Propulsion Laboratory, Pasadena, Calif , July 15, 1976

Berman, A L , and Renzetti, N A , *The Deep Space Network–An Instrument for Radio Science Research*, Publication 80-93, Jet Propulsion Laboratory, Pasadena, Calif , Feb 15, 1981

Bird, M. K., "Coronal Investigations with Occulted Spacecraft Signals," *Space Sci. Rev* , Mar 1982

Born, G H , and Mohan, S. N , "Orbit Determination for Mariner 9 Using Radio and Optical Data," *J Spacecraft Rockets*, Vol 12, No 7, pp 439-441, July 1975

Butman, S A , "Linear Feedback Rate Bounds for Regressive Channels," *IEEE Trans Inform Theor* , Vol IT-22, No 3, pp 363-366, May 1976

Butman, S. A., et al , "Design Criteria for Noncoherent Gaussian Channels With MFSK Signaling and Coding," *IEEE Trans Commun* , Vol COM-24, No 10, pp 1078-1088, Oct 1976

Butman, S A , and Lesh, J R , "The Effects of Bandpass Limiters on *n*-Phase Tracking Systems," *IEEE Trans Commun* , Vol COM-25, No 6, pp 569-576, June 1977.

Chao, C.-C , "Interstation Frequency Offset Determination Using Differenced 2-Way/ 3-Way Doppler Data," paper presented at the 1978 Spring Meeting of the American Geophysical Union, Miami, Florida, Apr 17-20, 1978

Christensen, C S , and Reinold, S J , "Navigation of the Mariner 10 Spacecraft to Venus and Mercury," *J Spacecraft Rockets*, Vol 12, No 5, pp 280-286, May 1975

Christensen, C S , et al , "On Achieving Sufficient Dual Station Range Accuracy for Deep Space Navigation at Zero Declination," paper presented at AAS/AIAA Astrodynamics Specialist Conference, Jackson Hole, Wyo , Sept 7-9, 1977

Christensen, E J , et al , "The Mass of Phobos from Viking Flybys," *Geophys Res Lett* , Vol. 4, No 12, pp 555-557, Dec 1977

Clauss, R , Flesner, L. D , and Schultz, S , "Simple Waveguide Reflection Maser with Broad Tunability," *Rev Sci. Instrum* , Vol 48, No 8, pp 1104-1105, Aug 1977

A Collection of Articles on S/X-Band Experiment Zero Delay Ranging Tests, Technical Memorandum 33-747, Vol I, Jet Propulsion Laboratory, Pasadena, Calif , Nov 1, 1975.

Curkendall, D W , "Algorithms for Isolating Worst Case Systematic Data Errors," *J Guidance Contr* , Vol 1, No 1, pp 56-62, Jan -Feb 1978

Dickinson, R. M , "The Beamed Power Microwave Transmitting Antenna," *IEEE Trans Microwave Theor Tech.*, Vol MTT-26, No 5, pp 335-340, May 1978.

Dickinson, D F , et al , "Sensitive Ammonia Observations in the Southern Hemisphere," *Astron J* , in press (Aug 1982)

Downs, G S , and Reichley, P. E , *Techniques for Measuring Arrival Times of Pulsar Signals I DSN Observations from 1968 to 1980*, Publication 80-54, Jet Propulsion Laboratory, Pasadena, Calif , Aug 15, 1980

Downs, G S , "JPL Pulsar Timing Observations I The Vela Pulsar," *Ap J* , Vol 249, pp 687-697, Oct 15, 1981

Downs, G S, "JPL Pulsar Timing Observations Spinups in PSR 0525+21," *Ap J* (Letters), Vol 257, L67-L70, June 15, 1982

Duxbury, T C, Johnson, T V, and Matson, D L, "Galilean Satellite Mutual Occultation Data Processing," *Icarus*, Vol 25, No 4, pp 569-584, Aug 1975

Edelson, R E, "An Observational Program to Search for Radio Signals From Extraterrestrial Intelligence Through the Use of Existing Facilities," Preprint IAF-A-76-033, Int Astronaut Fed XXVII Congress, Anaheim, Calif, Oct 10-16, 1976

Edelson, R E, and Levy, G S, "The Search for Extraterrestrial Intelligence Telecommunications Technology," *Proceedings of the 1976 National Telecommunications Conference,* Vol I, Dallas, Tex, Nov 29-Dec 1, 1976

Edelson, R E, "An Experimental Protocol for a Search for Radio Signals of Extraterrestrial Intelligent Origin in the Presence of Man-Made Radio Frequency Sources," paper presented at the XXVIIIth International Astronautical Congress, Prague, Czechoslovakia, Sept 25-Oct 1, 1977

Estabrook, F B, and Wahlquist, H D, "Response of Doppler Spacecraft Tracking to Gravitational Radiation," *Gen. Relat Grav*, Vol 6, No 5, pp 439-447, Oct 1975

Estacion Espacial de Madrid Madrid Space Station, Special Publication 43-26, Jet Propulsion Laboratory, Pasadena, Calif, Aug 31, 1975

Evans, R H, Kent, S S, and Seidman, J B, *Satellite Remote Sensing Facility for Oceanographic Applications*, Publication 80-40, Jet Propulsion Laboratory, Pasadena, Calif, July 1, 1980

Ferrari, A J, and Ananda, M P, "Lunar Gravity A Long-Term Keplerian Rate Method," *J Geophys Res*, Vol 82, No 20, pp 3085-3097, July 10 1977

Fjeldbo, G, et al, "Viking Radio Occultation Measurements of the Martian Atmosphere and Topography Primary Mission Coverage," *J Geophys Res*, Vol 82, No 28, pp 4317-4324, Sept 30, 1977

Fliegel, H, *Session III of the VLBI/Laser Intercomparison Task of the NASA Crustal Dynamics Project*, Publication 81-96, Jet Propulsion Laboratory, Pasadena, Calif, Nov 1, 1981

Fortenberry, J W, Freeland, R E, and Moore, D M, *Five-Meter-Diameter Conical Furlable Antenna*, Technical Report 32-1604, Jet Propulsion Laboratory, Pasadena, Calif, July 15, 1976

Freiley, A J, Batelaan, P D, and Bathker, D A, *Absolute Flux Density Calibrations of Radio Sources 2 3 GHz,* Technical Memorandum 33-806, Jet Propulsion Laboratory, Pasadena, Calif, Dec 1, 1977

Gaghardi, R M, Vilnrotter, V A, and Dolinar, S J, *Optical Deep Space Communication via Relay Satellite*, Publication 81-40, Jet Propulsion Laboratory, Pasadena, Calif, Aug 15, 1981

Goldstein, R M, and Morris, G A, "Ganymede Observations by Radar," *Science,* Vol 188, No 4194, pp 1211-1212, June 20, 1975

Goldstein, R M, Green, R R, and Rumsey, H, Jr, "Venus Radar Images," *J Geophys Res*, Vol 81, No 26, pp 4807-4817, Sept 10, 1976

Goodwin, P S, et al, *Tracking and Data Systems Support for the Helios Project Project Development Through End of Mission Phase II*, Technical Memorandum 33-752, Vol I, Jet Propulsion Laboratory, Pasadena, Calif, July 1, 1976

Goodwin, P S, et al, *Tracking and Data Systems Support for the Helios Project DSN Support of Project Helios April 1975 Through May 1976*, Technical Memorandum 33-752, Vol II, Jet Propulsion Laboratory, Pasadena, Calif, Jan 15, 1977

Goodwin, P S, Jensen, W N, and Flanagan, F M, *Tracking and Data Systems Support for the Helios Project DSN Support of Project Helios May 1976 Through June 1977*, Technical Memorandum 33-752, Vol III, Jet Propulsion Laboratory, Pasadena, Calif, Mar 1, 1979

Gulkis, S, et al, "An All-Sky Search for Narrow-Band Radiation in the Frequency Range 1-25 GHz," paper presented at the 1976 U S National Commission, International Union of Radio Science, Amherst, Mass, Oct 10-15, 1976

Harris, A W, et al, "2290-MHz Flux Densities of 52 High-Declination Radio Sources," *Astron J*, Vol 81, No 4, pp 222-224, Apr 1976

Higa, W H, "Spurious Signals Generated by Electron Tunneling on Large Reflector Antennas," *Proc IEEE*, Vol 63, No 2, pp 306-313, Feb 1975

Higa, W H, *The Superconducting Cavity-Stabilized Maser Oscillator*, Technical Memorandum 33-805, Jet Propulsion Laboratory, Pasadena, Calif, Dec 15, 1976

Hunter, J A, "Orbiting Deep Space Relay Station, A Study Report," paper presented at AIAA Conference on Large Space Platforms, Future Needs and Capabilities, Los Angeles, Calif, Sept 27-29, 1978

Jacobson, R A, McDanell, J P, and Rinker, G C, "Use of Ballistics Arcs in Low Thrust Navigation," *J Spacecraft Rockets*, Vol 12, No 3, pp 138-145, Mar 1975

Jauncey, D F, et al, "18-25 GHz Low Noise Line Receiver for the Tidbinbilla 64M Antenna," *Proc Astr Soc Aust*, CSIRO preprint

Jauncey, D L, et al, "2 3 GHz Accurate Position and Optical Identifications for Selected Parkes Radio Sources," *Am J*, May 1982

Jurgens, R F, and Goldstein, R M, "Radar Observations at 3 5 and 12 6 cm Wavelength of Asteroid 433 Eros," *Icarus*, Vol 28, No 1, pp 1-15, May 1976

Jurgens, R F, and Bender, D F, "Radar Detectability of Asteroids," *Icarus*, Vol 31, No 4, pp 483-497, Aug 1977

Katow, M S, "Evaluating Computed Distortions of Parabolic Reflectors," *Record of IEEE 1977 Mechanical Engineering in Radar Symposium, Arlington, Virginia, Nov 8-10, 1977*, IEEE Publication 77CH 1250-0 AES, pp 91-93

Kliore, A J, Woiceshyn, P M, and Hubbard, W P, "Pioneer 10 and 11 Radio Occultations by Jupiter," *COSPAR Space Research*, Vol XVII, pp 703-710, Pergamon Press Ltd, Oxford, 1978

Kliore, A, et al, "The Polar Ionosphere of Venus Near the Terminator From Early Pioneer Venus Orbiter Radio Occultation," *Science*, Vol 203, No 4382, pp 765-768, Feb 23, 1979

Krishnamohan, S, and Downs, G S, "Intensity Dependence of the Pulse Profile and Polarization of the Vela Pulsar," *Ap J*, in press (Feb 1, 1983)

Kuiper, T B H, and Morris, M, "Searching for Extraterrestrial Civilizations," *Science*, Vol 196, pp 616-621, May 6, 1977

Lesh, J R, "Sequential Decoding in the Presence of a Noisy Carrier Reference," *IEEE Trans Commun*, Vol COM-23, No 11, pp 1292-1297, Nov 1975

Levitt, B K, "Long Frame Sync Words for Binary PSK Telemetry," *IEEE Trans Commun*, COM-23, No 11, pp 1365-1367, Nov 1975

Levy, G S., et al, "Helios-1 Faraday Rotation Experiment Results and Interpretations of the Solar Occultations in 1975," *J Geophys*, Vol 42, No 6, pp 659-672, 1977

Levy, R, "Computer-Aided Design of Antenna Structures and Components," *Comput Struc*, Vol 6, Nos 4/5, pp 419-428, Aug /Oct 1976

Levy, R, and McGinness, H, *Wind Power Prediction Models*, Technical Memorandum 33-802, Jet Propulsion Laboratory, Pasadena, Calif, Nov 15, 1976

Levy, R, and Katow, M S, "Implementation of Wind Performance Studies for Large Antenna Structures," *Record of IEEE 1977 Mechanical Engineering in Radar Symposium, Arlington, Virginia, Nov 8-10, 1977*, IEEE Publication 77CH 1250-0 AES, pp 27-33

Levy, R, "Antenna Bias Rigging for Performance Objective," *Record of IEEE 1977 Mechanical Engineering in Radar Symposium, Arlington, Virginia, Nov 8-10, 1977*, IEEE Publication 77CH 1250-0 AES, pp 94-97

Liu, A S, and Pease, G E, "Spacecraft Ranging From a Ground Digitally Controlled Oscillator," *J Spacecraft Rockets*, Vol 12, No 9, pp 528-532, Sept 1975

Liu, K Y, The Effects of Receiver Tracking Phase Error on the Performance of the Concatenated Reed-Solomon/Viterbi Channel Coding System, Publication 81-62, Jet Propulsion Laboratory, Pasadena, Calif, Sept 1, 1981

Martin, W L, and Zygielbaum, A I, *Mu-II Ranging*, Technical Memorandum 33-768, Jet Propulsion Laboratory, Pasadena, Calif, May 15, 1977

Melbourne, W G, "Navigation between the Planets," *Sci. Amer*, Vol 234, No 6, pp 58-74, June 1976

Miller, R B, *Tracking and Data System Support for the Pioneer Project Pioneer 10–From April 1, 1972, Through the Jupiter Encounter Period, January 1974*, Technical Memorandum 33-584, Vol III, Jet Propulsion Laboratory, Pasadena, Calif, June 15, 1975

Miller, R B, et al, *Tracking and Data System Support for the Pioneer Project Pioneer 10–From January 1974 to January 1975, Pioneer 11–From May 1, 1973 Through Jupiter Encounter Period, January 1975*, Technical Memorandum 33-584, Vol IV, Jet Propulsion Laboratory, Pasadena, Calif, Dec 1, 1975

Miller, R L, Deutsch, L J, and Butman, S A, *On the Error Statistics of Viterbi Decoding and the Performance of Concatenated Codes*, Publication 81-9, Jet Propulsion Laboratory, Pasadena, Calif, Sept 1, 1981

Mudgway, D J, *Tracking and Data System Support for the Viking 1975 Mission to Mars Prelaunch Planning, Implementation, and Testing*, Technical Memorandum 33-783, Vol I, Jet Propulsion Laboratory, Pasadena, Calif, Jan 15, 1977

Mudgway, D J, and Traxler, M R, *Tracking and Data System Support for the Viking 1975 Mission to Mars Launch Through Landing of Viking 1*, Technical Memorandum 33-783, Vol II, Jet Propulsion Laboratory, Pasadena, Calif, Mar 15, 1977

Mudgway, D J , *Tracking and Data System Support for the Viking 1975 Mission to Mars Planetary Operations*, Technical Memorandum 33-783, Vol III, Jet Propulsion Laboratory, Pasadena, Calif , Sept 1, 1977.

Mudgway, D. J , *Tracking and Data System Support for the Viking 1975 Mission to Mars Extended Mission Operations December 1976 to May 1978,* Technical Memorandum 33-783, Vol IV, Jet Propulsion Laboratory, Pasadena, Calif., Dec 15, 1978

Muhleman, D O., Esposito, P. B , and Anderson, J. D , "The Electron Density Profile of the Outer Corona and the Interplanetary Medium From Mariner-6 and Mariner-7 Time-Delay Measurements," *Astrophys J.,* No. 211, No 3, Part 1, pp 943-957, Feb 1, 1977

Murray, B C , Gulkis, S , and Edelson, R. E , "Extraterrestrial Intelligence An Observational Approach," *Science,* Vol 199, No 4328, pp 485-492, Feb 3, 1978

Ohlson, J E , "Polarization Tracking of a Partially Coherent Signal Using a Double Loop," *IEEE Trans Commun.*, Vol COM-23, No 9, pp 859-866, Sept 1975.

Ohlson, J E , and Reid, M S , *Conical-Scan Tracking With the 64-m-diameter Antenna at Goldstone,* Technical Report 32-1605, Jet Propulsion Laboratory, Pasadena, Calif , Oct 1, 1976.

Ong , K. M , et al , "A Demonstration of a Transportable Radio Interferometric Surveying System With 3-cm Accuracy on a 307-m Base Line," *J Geophys Res*, Vol. 81, No 20, pp 3587-3593, July 10, 1976

Otoshi, T Y , and Stelzried, C T , "Cosmic Background Noise Temperature Measurement at 13-cm Wavelength," *IEEE Trans Instr Meas.*, Vol IM-24, No 2, pp 174-179, June 1975

Phillips, R J , et al , "Simulation Gravity Modeling to Spacecraft-Tracking Data Analysis and Application," *J Geophys Res*, Vol 83, No B11, pp. 5455-5464, Nov. 10, 1978

Posner, E C , "Random Coding Strategies for Minimum Entropy," *IEEE Trans Inform Theor*, Vol IT-21, No 4, pp 388-391, July 1975

Posner, E C , "Life Cycle Costing with a Discount Rate," *Utilitas Mathematica,* Vol 13, pp 157-188, 1978

Powell, R V , and Hibbs, A R , "An Entree for Large Space Antennas," *Astronaut Aeronaut*, Vol 15, No 12, pp 58-64, Dec. 1977

Reed, I S , Truong, T K , and Benjauthrit, B , "On Decoding of Reed-Solomon Codes over GF(32) and GF(64) Using the Transform Techniques of Winograd," *Conference Record, 1978 National Telecommunications Conference*, Vol 2, Birmingham, Alabama, Dec 3-6, 1978

Simon, M K , "A Generalization of Minimum-Shift-Keying (MSK) Type Signaling Based Upon Input Data Symbol Pulse Shaping," *IEEE Trans Commun*, Vol COM-24, No 8, pp 845-856, Aug 1976.

Simon, M K., "An MSK Approach to Offset QASK," *IEEE Trans Commun.*, Vol COM-24, No 8, pp 921-923, Aug 1976

Simon, M K , "The False Lock Performance of Costas Loops with Hard-Limited In-Phase Channel," *IEEE Trans Commun*, Vol COM-26, No 1, pp 23-34, Jan 1978

Simon, M K, "On the Calculation of Squaring Loss in Costas Loops with Arbitrary Arm Filters," *IEEE Trans Commun*, Vol COM-26, No 1, pp 179-184, Jan 1978

Simon, M K, "Tracking Performance of Costas Loops With Hard-Limited In-Phase Channel," *IEEE Trans Commun.*, Vol COM-26, No 4, pp 420-432, Apr 1978

Sjogren, W L, et al, "Gravity Fields," *IEEE Trans Geosci. Electron.*, Vol GE-14, No 3, pp 172-183, July 1976

Slade, M A, et al, "Alsep-Quasar Differential VLBI," *The Moon*, Vol 17, pp 133-147, 1977

Smith, E K, and Edelson, R E, *Radio Propagation Through Solar and Other Extra-terrestrial Ionized Media*, Publication 79-117, Jet Propulsion Laboratory, Pasadena, Calif, Jan 15, 1980

Stelzried, C T, Seidel, B L, and Hollweg, J V, "Possible Evidence for Coronal Alfven Waves," *J Geophys Res*, Vol 87, No A1, pp 1-8, Jan 1, 1982

Sue, M K, *Radio Frequency Interference at the Geostationary Orbit Final Report*, Publication 81-34, Jet Propulsion Laboratory, Pasadena, Calif, June 15, 1981

Tausworthe, R C, *Standard Classifications of Software Documentation*, Technical Memorandum 33-756, Jet Propulsion Laboratory, Pasadena, Calif, Jan 15, 1976

Tausworthe, R C, *Deep Space Network Software Cost Estimation Model*, Publication 81-7, Jet Propulsion Laboratory, Pasadena, Calif, Apr 15, 1981

Thomas, J B, et al, "A Demonstration of an Independent-Station Radio Interferometry System With 4-cm Precision on a 16-km Base Line," *J Geophys Res*, Vol 81, No 5, pp 995-1005, Feb 10, 1976

Thomas, J B, *An Analysis of Source Structure Effects in Radio Interferometry Measurements*, Publication 80-84, Jet Propulsion Laboratory, Pasadena, Calif, Dec 15, 1980

Thomas, J B, *An Analysis of Radio Inteferometry with the Block O System*, Publication 81-49, Jet Propulsion Laboratory, Pasadena, Calif, Dec 15, 1981

Toukdarian, R Z, *Final Engineering Report for Goldstone Operations Support Radar*, Technical Memorandum 33-800, Jet Propulsion Laboratory, Pasadena, Calif, Nov 1, 1976

Truong, T K, and Reed, I S, "Convolutions Over Residue Classes of Quadratic Integers," *IEEE Trans Inform Theor*, Vol IT-22, No 4, pp 468-475, July 1976

Truong, T K, and Reed, I S, "Convolutions over Quartic Integer Residue Classes," *Proceedings of the International Conference on Information Sciences and Systems*, Patras, Greece, Aug 19-24, 1976

Truong, T K, Liu, K Y, and Reed, I S, "Fast Number-Theoretic Transforms for Digital Filtering," *Electron, Lett*, Vol 12, No 24, Nov 1976

Truong, T K, et al, "X-Ray Reconstruction by Finite Field Transforms," *IEEE Trans Nucl Sci*, Vol NS-24, No 1, pp 843-849, Feb 1977

Truong, T K, Golomb, S W, and Reed, I S, "Integer Convolutions Over the Finite Field $GF(3 \cdot 2^n + 1)$," *SIAM J Appl Math*, Vol 32, No 2, pp 356-365, Mar 1977

Truong, T K, Reed, I S, and Liu, K Y, "Fast Algorithm for Computing Complex Number-Theoretic Transforms," *Electron Lett*, Vol 13, No 10, pp 278-280, May 12, 1977

Weber, W J, III, "Performance of Phase-Locked Loops in the Presence of Fading Communication Channels," *IEEE Trans Commun.*, Vol COM-24, No 5, pp 487-499, May 1976.

Weber, W J, III, Ackerknecht, W. E, and Kollar, F J, *Viking X-Band Telemetry Experiment Final Report,* Technical Memorandum 33-794, Jet Propulsion Laboratory, Pasadena, California, Sept 1, 1976

Wong, J S L, et al, *Review of Finite Fields Applications to Discrete Fourier Transforms and Reed-Solomon Coding,* JPL Publication 77-23, Jet Propulsion Laboratory, Pasadena, Calif, July 15, 1977

Woo, R, "Measurements of the Solar Wind Using Spacecraft Radio Scattering Observations," in *Study of Traveling Inter-Planetary Phenomena,* pp 81-100, D. Reidel Publishing Co, Dordrecht, Holland/Boston, 1977

Woo, R, "Radial Dependence of Solar Wind Properties Deduced from Helios 1/2 and Pioneer 10/11 Radio Scattering Observations," *Astrophys J*, Vol 219, No 2, Part 1, pp 727-739, Jan 15, 1978

Woo, R T, "Observations of Turbulence in the Atmosphere of Venus Using Mariner 10 Radio Occultation Measurements," *J Atmos Sci.*, Vol 32, No 6, pp 1084-1090, June 1975

Yuen, J H, *A Practical Statistical Model for Telecommunications Performance Uncertainty,* Technical Memorandum 33-732, Jet Propulsion Laboratory, Pasadena, Calif, June 15, 1975

End of Document