

NASA
ANALYTIC
(1+25)
PRIMM7
165611

617959
166p.

The Telecommunications and Data Acquisition Progress Report 42-94

April-June 1988

E. C. Posner
Editor

(NASA-CR-183243) THE TELECOMMUNICATIONS AND
DATA ACQUISITION REPORT Progress Report,
Apr. - Jun. 1988 (Jet Propulsion Lab.)
312 p

CSCL 17B

G3/32

N89-10186
--THRU--
N89-10211
Unclass
0165611

August 15, 1988



National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

The Telecommunications and Data Acquisition Progress Report 42-94

April–June 1988

E. C. Posner
Editor

August 15, 1988



National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

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

Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement by the United States Government or the Jet Propulsion Laboratory, California Institute of Technology.

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 and radar 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 Operations (OSO).

In the search for extraterrestrial intelligence (SETI), the TDA Progress Report reports on implementation and operations for searching the microwave spectrum. In solar system radar, it reports on the uses of the Goldstone Solar System Radar for scientific exploration of the planets, their rings and satellites, asteroids, and comets. In radio astronomy, the areas of support include spectroscopy, very long baseline interferometry, and astrometry. These three programs are performed for NASA's Office of Space Science and Applications (OSSA), with support by the Office of Space Operations for the station support time.

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:

OSO 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 Existing Projects
 - Capabilities for New Projects
 - New Initiatives
 - Network Upgrade and Sustaining
- DSN Operations
 - Network Operations and Operations Support
 - Mission Interface and Support
 - TDA Program Management and Analysis
- Communications Implementation and Operations
- Data and Information Systems
- Flight-Ground Advanced Engineering

OSSA Tasks:

- Search for Extraterrestrial Intelligence
- Goldstone Solar System Radar
- Radio Astronomy

Discretionary Funded Tasks

Contents

OSO TASKS DSN Advanced Systems TRACKING AND GROUND-BASED NAVIGATION

Deep Space Tracking in Local Reference Frames	1
R. N. Treuhaft	
NASA Code 310-10-60-86-02	
Precise Orbit Determination for NASA's Earth Observing System Using GPS	16
B. G. Williams	
NASA Code 310-10-61-84-04	
PSA: A Program to Streamline Orbit Determination for Launch Support Operations	28
V. N. Legerton and N. A. Mottinger	
NASA Code 310-10-63-53-00	
Spain 31-GHz Observations of Sky Brightness Temperatures	42
B. L. Gary	
NASA Code 310-10-60-84-00	
Australia 31-GHz Brightness Temperature Exceedance Statistics	61
B. L. Gary	
NASA Code 310-10-60-84-00	
Gravitational Wave Searches Using the DSN	75
S. J. Nelson and J. W. Armstrong	
NASA Code 310-10-62-25-00	
Evaluation of the Characteristics of a Field Emission Cathode for Use in a Mercury Ion Trap Frequency Standard	86
J. M. Christman	
NASA Code 310-10-62-15-00	

COMMUNICATIONS, SPACECRAFT-GROUND

Simultaneous S- and X-Band Uplink-Downlink Performance at DSS 13	93
A. J. Freiley	
NASA Code 310-20-64-86-02	
X/X/Ka-Band Prime Focus Feed Antenna for the Mars Observer Beacon Spacecraft	103
P. Stanton, H. Reilly, and M. Esquivel	
NASA Code 310-20-64-86-02	
Dual Passband Dichroic Plate for X-Band	110
T. Y. Otoshi and M. M. Franco	
NASA Code 310-20-64-88-02	
Antenna Pointing Compensation Based on Precision Optical Measurement Techniques	135
L. L. Schumacher and H. C. Vivian	
NASA Code 310-20-65-67-00	
A 32-GHz Reflected-Wave Maser Amplifier With Wide Instantaneous Bandwidth	145
J. Shell and D. Neff	
NASA Code 310-20-66-87-02	

An 8.4-GHz Cryogenically Cooled HEMT Amplifier for DSS 13	163
L. Tanida	
NASA Code 310-20-66-09-00	
Design of an Optical PPM Communication Link in the Presence of Component Tolerances	170
C. -C. Chen	
NASA Code 310-20-67-59-00	
Atmospheric Propagation Effects Relevant to Optical Communications	180
K. S. Shaik	
NASA Code 310-20-67-88-03	
Node Synchronization of Viterbi Decoders Using State Metrics	201
U. Cheng	
NASA Code 310-30-71-83-02	
A Software Simulation Study of the Long Constraint Length VLSI Viterbi Decoder	210
S. Arnold and F. Pollara	
NASA Code 310-30-71-88-01	

STATION CONTROL AND SYSTEM TECHNOLOGY

An Automatic Frequency Control Loop Using Overlapping DFTs	222
S. Aguirre	
NASA Code 310-30-70-84-02	
Heat Exchanger Expert System Logic	232
R. Cormier	
NASA Code 310-10-62-25-00	
Storage and Computationally Efficient Permutations of Factorized Covariance and Square-Root Information Arrays	245
R. J. Muellerschoen	
NASA Code 310-10-61-81-10	

DSN Systems Implementation CAPABILITIES FOR EXISTING PROJECTS

Phasing the Antennas of the Very Large Array for Reception of Telemetry from Voyager 2 at Neptune Encounter	257
J. S. Ulvestad	
NASA Code 314-40-66-09-12	

CAPABILITIES FOR NEW PROJECTS

Phobos Lander Coding System: Software and Analysis	274
K. -M. Cheung and F. Pollara	
NASA Code 314-30-64-01-03	

NETWORK UPGRADE AND SUSTAINING

Relative Planetary Radar Sensitivities: Arecibo and Goldstone	287
N. A. Renzetti, T. W. Thompson, and M. A. Slade	
NASA Code 314-40-31-30-14	
DSN 64-Meter Antenna L-Band (1668 MHz) Microwave System Performance Overview ...	294
J. Withington	
NASA Code 314-40-56-44-14	

DSN Operations
MISSION INTERFACE AND SUPPORT

Summary of DSN Reimbursable Launch Support	301
N. A. Fanelli and M. E. Wyatt	
NASA Code 314-40-22-34-01	
Referees	306

1989000816

N89 - 10187

S1-32

~~165672~~

TDA Progress Report 42-94

April-June 1988

617964
15p.

JTB 24450

Deep Space Tracking in Local Reference Frames

R. N. Treuhaft

Tracking Systems and Applications Section

A self-calibrating deep space tracking technique is described which can potentially produce 2-nanoradian angular spacecraft determinations. The technique uses very long baseline interferometric observations of a spacecraft and several radio sources. This article first describes the currently employed single-source technique as a parameter estimation procedure. Extending the number of parameters and observations leads to the proposed local reference frame technique. Station clock, Earth rotation, and tropospheric parameters are estimated along with spacecraft position from the multisource observation sequence. The contributions to spacecraft angular uncertainty from system noise, tropospheric fluctuations, and uncalibrated radio source structure are evaluated. Of these experimental errors, radio source structure dominates the determination of the spacecraft position in the radio reference frame. It is shown, however, that the sensitivity of relative spacecraft position accuracies to time-invariant radio source structure effects may be on the order of 2 nanoradians.

I. Introduction

Deep space angular tracking is routinely performed on Voyager with 50- to 100-nanoradian accuracy [1] and will be performed on Galileo at the 50-nanoradian level. The analysis in this article suggests a tracking strategy for the Deep Space Network (DSN) which, in the 30-minute measurement time typically allotted for angular tracking, will potentially yield better than 2-nanoradian accuracy and will require virtually no calibration support external to the signal chains at the stations. The self-calibrating technique outlined below capitalizes on the extreme sensitivity of the DSN after it is equipped with Mark III observing bandwidths and recording hardware [2].

The proposed astrometric technique can potentially determine angular distances accurate to 10 km near Neptune, or

2 km near Jupiter. Such a capability would allow the determination of a planet's position in the radio frame by sensing gravitational forces long before encounter [3].¹ In addition to providing navigational benefits, the technique to be described would help the DSN realize its potential as a high-precision astrometric radio and planetary science tool. This astrometric technique would facilitate other tracking-related scientific measurement, such as the determination of planetary and satellite masses. Additional experiments, such as direct measurement of relativistic gravitational bending by Jupiter would also become possible.

¹ R. N. Treuhaft and J. S. Ulvestad, "Using Gravitational Signatures for Target-Relative Angular Tracking During Planetary Approach," JPL Interoffice Memorandum No. 335.3-88-76 (internal document), Jet Propulsion Laboratory, Pasadena, California, July 11, 1988.

Angular deep space navigation is currently achieved by tracking with Very Long Baseline Interferometry (VLBI) [4]–[6]. In a technique called Delta Differential One-Way Range (Δ DOR), the geometric delay of a radio source signal is differenced from that of a spacecraft signal [1]. Through this difference, the angular stability of the radio reference frame, determined by the accuracy of the radio source catalog, can be transferred to the spacecraft angular measurement. For the DSN astrometric radio catalog, measurement errors currently limit radio source astrometry to the 10-nanoradian level [7]. As measurement techniques and instrumentation improve, source structure variability will begin to limit astrometric accuracy at the 5-nanoradian level [8].

In practice, the current 10-nanoradian stability of the radio frame is not completely transferred to the spacecraft angular accuracy. Errors in the determination of the spacecraft and radio source delays at the time of the Δ DOR measurement limit tracking accuracy to the 20- to 50-nanoradian range, which is about 15 to 40 kilometers of projected error at Jupiter or 90 to 225 kilometers at Neptune [1].² It has been shown that 5- to 10-nanoradian accuracies could be achieved with future DSN hardware configurations and improved calibration support using the Δ DOR single radio source technique.³

The alternative measurement strategy outlined below entails interferometric observations of a spacecraft and a number of radio sources. This technique of navigation in a local reference frame of radio sources can potentially yield 2-nanoradian accuracy without clock synchronization, Earth orientation, or tropospheric calibration support. The thrust of the local reference frame technique is to use the known positions of several reference radio sources to determine the magnitude of systematic errors affecting the normal Δ DOR position measurement. Navigation in a local reference frame is especially attractive for outer planet missions in which the spacecraft moves by only a few degrees on the sky over a year. For example, Voyager will be in the same 15-square-degree piece of sky from late 1987 until Neptune encounter. An additional advantage of the local reference frame technique is that the radio sources do not have to be angularly close to the spacecraft; 30-degree separations still yield the high accuracy mentioned above. When a spacecraft drifts away from the radio source

to which it has been referenced with standard Δ DOR, it will sometimes be in a place where no other close sources are available. The local reference frame technique is ideal in situations in which a “hole” in the known extragalactic radio sky is encountered.

In the next section, the conventional single radio source Δ DOR strategy is reviewed as a parameter estimation process. The local reference frame technique is then developed in Section III as an extension of this process. Section IV contains covariance results for the local network technique as compared to conventional Δ DOR. Section V summarizes the proposed technique, outlines possible improvements, and reports the status of experimental efforts at validation.

II. Delta DOR as Parameter Estimation

In this section, the Δ DOR process will be formulated as parameter estimation from observed VLBI delays. The dependence of the VLBI delays on spacecraft position and clock parameters will be followed by the least squares result for estimating the parameters and their errors. Extending the dependence of observed VLBI delays to include station clock rates, Earth rotation, and tropospheric effects will then lead to the local reference frame technique in the next section.

The object of angular tracking is to determine the spacecraft unit vector, which points from the Earth to the spacecraft. The essential quantities derived from Δ DOR measurements are the components of the residual spacecraft unit vector.⁴ With DSN VLBI, these components are determined by observations on nearly orthogonal baselines. The following analysis will treat measurements on a single baseline only. In that case, the quantity to be determined, Δs_p , is the component of the residual spacecraft unit vector along the projected baseline, which is the baseline projected onto the plane of the sky. Given the length of the a priori projected baseline, B_p at the spacecraft observation epoch,

$$\Delta s_p = \frac{c\Delta\tau}{B_p} \quad (1)$$

In Eq. (1), c is the speed of light, and $\Delta\tau_g$ is the residual spacecraft geometric delay. It is given by

$$\Delta\tau_g = \frac{\vec{B}_m \cdot (\hat{s}_{sc} - \hat{s}_{m\ sc})}{c} \quad (2)$$

²J. B. Thomas, “An Error Analysis for Galileo Angular Position Measurements With the Block 1 Δ DOR System,” JPL Engineering Memorandum No. 335-26 (internal document), Jet Propulsion Laboratory, Pasadena, California, November 11, 1981.

³R. N. Treuhaft and L. J. Wood, “Revisions in the Differential VLBI Error Budget and Applications for Navigation in Future Missions,” JPL Interoffice Memorandum No. 335.4-601 (internal document), Jet Propulsion Laboratory, Pasadena, California, December 31, 1986.

⁴The term *residual* used in the analysis means the difference between the actual value of a quantity and its best known a priori value. Residuals are used both to linearize subsequent least squares procedures and to make computation less cumbersome.

where \vec{B}_m is the total baseline vector at the epoch of the spacecraft measurement, \hat{s}_{sc} is the actual unit vector pointing to the spacecraft, and $\hat{s}_{m,sc}$ is the a priori spacecraft unit vector. In much of the following analysis, $\Delta\tau_g$ will be regarded as the quantity to be determined from the interferometric observations. Geometric delays are related to projected spacecraft angles via Eq. (1). The rest of this section is devoted to the procedure for estimating $\Delta\tau_g$ and its error.

The extraction of $\Delta\tau_g$ from Δ DOR observations can be viewed as a simple example of parameter estimation. In Eqs. (3) and (4) below, the observed spacecraft and radio source residual delays, $\Delta\tau_{sc}$ and $\Delta\tau_{rs}$, are expressed in terms of $\Delta\tau_g$ and the residual difference between clock epochs at the two observing stations, $\Delta\tau_c$:

$$\Delta\tau_{sc} = \frac{\partial\tau_{sc}}{\partial\tau_g} \Delta\tau_g + \frac{\partial\tau_{sc}}{\partial\tau_c} \Delta\tau_c + \epsilon_{sc} \quad (3)$$

$$\Delta\tau_{rs} = \frac{\partial\tau_{rs}}{\partial\tau_c} \Delta\tau_c + \epsilon_{rs} \quad (4)$$

where ϵ_{sc} represents all contributions to observed spacecraft delay not included in the first two terms on the right side of Eq. (3). Similarly, ϵ_{rs} represents all effects not included in the first term on the right of Eq. (4). Since both $\Delta\tau_g$ and $\Delta\tau_c$ are delays, they map directly into observed residual delays, and all partial derivatives in Eqs. (3) and (4) are unity. Note that a residual radio source geometric delay, analogous to Eq. (2), does not explicitly appear in Eq. (4). This is because the residual delay effects induced by errors in the a priori baseline components and radio source coordinates are included in ϵ_{sc} and ϵ_{rs} . The only delay terms which explicitly appear in Eqs. (3) and (4) are those which contain parameters to be estimated. Generally, both random and systematic effects are included in ϵ_{sc} and ϵ_{rs} , but in the following discussion they are both assumed to have zero mean statistics.

Standard least squares procedures (e.g., [9]) give the minimum variance estimates of $\Delta\tau_g$ and $\Delta\tau_c$, $\hat{\Delta\tau}_g$ and $\hat{\Delta\tau}_c$, in terms of the observed residual delays:

$$\hat{\Delta\tau}_g = \Delta\tau_{sc} - \Delta\tau_{rs} \quad (5)$$

$$\hat{\Delta\tau}_c = \Delta\tau_{rs} \quad (6)$$

Equation (5) gives the familiar prescription of differencing the spacecraft and radio source observed residual delays to estimate $\Delta\tau_g$. The estimate of clock epoch given in Eq. (6) is available for correcting time tags in other radio metric data. The

parameter estimate variances due to white observable noise are also determined by standard least squares methods. For Δ DOR, with the dependence of observables on parameters given by Eqs. (3) and (4), the spacecraft delay and clock epoch parameter estimate variances are given by

$$\sigma_{\hat{\Delta\tau}_g, w}^2 = \sigma_{sc, w}^2 + \sigma_{rs, w}^2 \quad (7)$$

$$\sigma_{\hat{\Delta\tau}_c, w}^2 = \sigma_{rs, w}^2 \quad (8)$$

where $\sigma_{sc, w}^2$ and $\sigma_{rs, w}^2$ are the white components of $\langle\epsilon_{sc}^2\rangle$ and $\langle\epsilon_{rs}^2\rangle$, respectively. Note that since white observable noise is temporally uncorrelated, it is uncorrelated between the spacecraft and radio source observations.

The impact of all correlated errors affecting the delay measurements can be evaluated in a *consider* analysis (e.g., [10]). This is the formal method for determining how much a given error source cancels in the differencing of spacecraft and radio source delays prescribed by Eq. (5). The parameter estimate variances due to correlated error sources are

$$\sigma_{\hat{\Delta\tau}_g, c}^2 = \sigma_{sc, c}^2 + \sigma_{rs, c}^2 - 2\rho\sigma_{sc, c}\sigma_{rs, c} \quad (9)$$

$$\sigma_{\hat{\Delta\tau}_c, c}^2 = \sigma_{rs, c}^2 \quad (10)$$

where $\sigma_{sc, c}^2$ and $\sigma_{rs, c}^2$ are due to correlated components of $\langle\epsilon_{sc}^2\rangle$ and $\langle\epsilon_{rs}^2\rangle$, respectively, and ρ is the correlation coefficient between the spacecraft and radio source observations. Familiar Δ DOR rules of thumb, such as “ Δ DOR errors grow proportionally with spacecraft–radio source separation,” formally originate in Eq. (9). Examples of correlated error sources are clock rate, Earth rotation, and tropospheric effects.

III. Local Reference Frame Parameter Estimation

Equations (3) through (10) cast Δ DOR in the form of a parameter estimation procedure. The parameters estimated are the residual spacecraft geometric delay and residual differential clock epoch. It has been shown that clock rate, Earth orientation, and static tropospheric effects are the next dominant correlated error sources in ϵ_{sc} and ϵ_{rs} (see reference 3). (As will be assumed throughout this article, ionospheric errors are largely removed by dual-frequency calibration and contribute less than 100-picoradian observation errors for the typical 2300- and 8400-GHz observing frequencies.) The approach taken in the reference to reduce Δ DOR uncertainties from the 30- to the 5-nanoradian level was to improve calibration

support. Highly accurate earth orientation measurements and water vapor radiometry were suggested as the means to improving ΔDOR . An alternative to augmenting calibration support is to introduce more parameters into Eqs. (3) and (4) and make more radio source observations to solve for them. As implied in the introduction, the Mark III instrumentation now being installed in the DSN will allow on the order of six highly accurate (≈ 30 -picosecond) radio source group delay measurements to be made in the same time currently allotted for one.

In Eqs. (11) and (12) below, Eqs. (3) and (4) are rewritten with the extended parameter dependences. The first new parameter is $\dot{\tau}_c$, which is the delay due to the residual differential clock rate between the two stations. The next two parameters, θ and ϕ , are the magnitudes of rotations about two axes perpendicular to the baseline. The rotation axes are chosen to be $\hat{s}_{m,sc} \times \vec{B}_m$ and $\vec{B}_m \times \hat{s}_{m,sc} \times \vec{B}_m$. The next two parameters, z_1 and z_2 , are the magnitudes of the static zenith tropospheric delay over each of the two stations. The expressions for all partial derivatives of delay with respect to those parameters are given in Appendix A. The local-reference-frame extended parameterizations for observed delay are

$$\begin{aligned} \Delta\tau_{sc} = & \frac{\partial\tau_{sc}}{\partial\tau_g} \Delta\tau_g + \frac{\partial\tau_{sc}}{\partial\tau_c} \Delta\tau_c + \frac{\partial\tau_{sc}}{\partial\dot{\tau}_c} \Delta\dot{\tau}_c + \frac{\partial\tau_{sc}}{\partial\theta} \Delta\theta \\ & + \frac{\partial\tau_{sc}}{\partial\phi} \Delta\phi + \frac{\partial\tau_{sc}}{\partial z_1} \Delta z_1 + \frac{\partial\tau_{sc}}{\partial z_2} \Delta z_2 + \epsilon'_{sc} \end{aligned} \quad (11)$$

$$\begin{aligned} \Delta\tau_{rs} = & \frac{\partial\tau_{rs}}{\partial\tau_c} \Delta\tau_c + \frac{\partial\tau_{rs}}{\partial\dot{\tau}_c} \Delta\dot{\tau}_c + \frac{\partial\tau_{rs}}{\partial\theta} \Delta\theta + \frac{\partial\tau_{rs}}{\partial\phi} \Delta\phi \\ & + \frac{\partial\tau_{rs}}{\partial z_1} \Delta z_1 + \frac{\partial\tau_{rs}}{\partial z_2} \Delta z_2 + \epsilon'_{rs} \end{aligned} \quad (12)$$

where ϵ'_{sc} and ϵ'_{rs} include all observation errors not explicitly appearing on the right sides of Eqs. (11) and (12), respectively. There are now seven parameters to be estimated.

It should be noted that the VLBI phase delay rate, which is a measure of the short-term rate of change of VLBI delay, is extracted along with the delay in the VLBI fringe fitting procedure. It is not normally used in standard ΔDOR because it is believed to be dominated by atmospheric fluctuations rather than by the geometric signature of the rotating Earth [11]. It will be seen, however, that in the local reference frame technique, the phase delay rate can be used to constrain singularities in the least squares solution for the above seven parameters. The parameter dependences of the observed residual

spacecraft and radio source delay rates, $\Delta\dot{\tau}_{sc}$ and $\Delta\dot{\tau}_{rs}$, are expressed in Eqs. (13) and (14):

$$\begin{aligned} \Delta\dot{\tau}_{sc} = & \frac{\partial\dot{\tau}_{sc}}{\partial\dot{\tau}_g} \Delta\dot{\tau}_g + \frac{\partial\dot{\tau}_{sc}}{\partial\dot{\tau}_c} \Delta\dot{\tau}_c + \frac{\partial\dot{\tau}_{sc}}{\partial\dot{\tau}_c} \Delta\dot{\tau}_c + \frac{\partial\dot{\tau}_{sc}}{\partial\theta} \Delta\theta \\ & + \frac{\partial\dot{\tau}_{sc}}{\partial\phi} \Delta\phi + \frac{\partial\dot{\tau}_{sc}}{\partial z_1} \Delta z_1 + \frac{\partial\dot{\tau}_{sc}}{\partial z_2} \Delta z_2 + \epsilon''_{sc} \end{aligned} \quad (13)$$

$$\begin{aligned} \Delta\dot{\tau}_{rs} = & \frac{\partial\dot{\tau}_{rs}}{\partial\dot{\tau}_c} \Delta\dot{\tau}_c + \frac{\partial\dot{\tau}_{rs}}{\partial\dot{\tau}_c} \Delta\dot{\tau}_c + \frac{\partial\dot{\tau}_{rs}}{\partial\theta} \Delta\theta + \frac{\partial\dot{\tau}_{rs}}{\partial\phi} \Delta\phi \\ & + \frac{\partial\dot{\tau}_{rs}}{\partial z_1} \Delta z_1 + \frac{\partial\dot{\tau}_{rs}}{\partial z_2} \Delta z_2 + \epsilon''_{rs} \end{aligned} \quad (14)$$

The partial derivatives of delay rate with respect to the parameters are given in Appendix A in terms of baseline and source coordinates. The new parameter in Eq. (13) is the residual spacecraft geometric delay rate, $\Delta\dot{\tau}_g$. This parameter, inserted for completeness, represents the sum of two contributions to the spacecraft's delay rate: (1) the residual rate induced by unmodeled spacecraft motion; and (2) the residual rate induced by Earth rotation in the presence of a nonzero residual spacecraft position vector. In Eqs. (13) and (14), ϵ''_{sc} and ϵ''_{rs} are the contributions to the spacecraft and radio source delay rate not explicitly parameterized.

As will be described in the next section on covariance results, the seven parameters in Eq. (11) are largely determined by delay observations; approximately seven delay observations are therefore required in a local reference frame observing sequence with the above parameterizations. Only the $\dot{\tau}_g$ parameter is explicitly determined by the delay rate observation, and it is of minimal use in reducing angular position errors. In the next section, a data acquisition strategy is evaluated in which the spacecraft and five radio sources are observed, with one source observed twice. The error in determining residual spacecraft geometric delay due to system noise, tropospheric fluctuations, and source structure will be calculated.

IV. Covariance Results

A. Observation Configuration for Covariance Calculation

The radio source-spacecraft configuration used to calculate sample covariance results is shown in Fig. 1. The radio sources, which are denoted by R_s in the figure, are from the DSN source catalog. Their coordinates in the J2000 reference

frame are given in Table 1 [7].⁵ The S in Fig. 1 denotes the spacecraft angular position used for the covariance analyses; its coordinates are the last entry in Table 1. In these analyses, it was assumed that the spacecraft and four of the five radio sources were observed only once on the California–Australia baseline. One radio source was observed twice to provide the seven observations needed to determine the seven parameters in Eq. (11) above. The numbers below the letters in Fig. 1 indicate the temporal sequence of observations. Each observation was assumed to last one minute, with two minutes allowed for antenna slewing. In Section IVB, the covariance due to system noise and tropospheric fluctuations is presented. Section IVC contains the covariance due to source structure effects. Although the spacecraft observation was spatially and temporally centered in the sample observing sequence, the covariance results which follow do not critically depend on such a sequence. For example, observing the spacecraft before all radio source observations changes the results of the next section by only ± 1 nanoradian.

B. Covariance Due to System Noise and Tropospheric Fluctuations

In order to evaluate the spacecraft angular error due to white system noise in the local reference frame example of Fig. 1, white noise standard deviations of 30 picoseconds were assumed for both the radio source and spacecraft observations. For the selected radio sources, this assumption is consistent with typical Mark III group delay noise values for one-minute observations. If it were possible to use the higher-precision VLBI phase delays instead of group delays [12], then typical noise standard deviations would be below one picosecond. However, this analysis will not assume the availability of the phase delay data type. Using standard least squares techniques, an equation analogous to Eq. (7) can be derived for the spacecraft geometric delay error in terms of the system noise contributions of the seven observations described above.

The model described in Eqs. (11) through (14) assumes a static tropospheric delay over each station. Tropospheric fluctuations, predominantly due to the wet component, will cause an error in the spacecraft geometric delay. This error has been evaluated in a consider analysis using a covariance matrix, derived for average DSN conditions, from wet tropospheric fluctuation modeling ([11], Eq. [15]). The model in the reference assumes that a frozen spatial pattern of turbulence is blown across a site by the wind. The model, which is normalized by WVR data taken near Goldstone, is consistent with a daily wet zenith delay fluctuation of 1 cm.

Both system noise and tropospheric contributions to spacecraft angular error are shown in Fig. 2. The ordinate is the error in the estimation of Δs_p of Eq. (1) for the California–Australia baseline, which is approximately 10,600 km. This angular error is plotted versus hour angle, which is defined as the angle between the spacecraft and baseline vectors at the middle of the seven-observation sequence. The pattern of Fig. 1, therefore, moves across the sky to the west as the hour angle decreases. It can be seen that the system noise contribution is between 1 and 3 nanoradians. The tropospheric error is somewhat smaller. The peak on the right side of the plot is due largely to correlations between the first rotation parameter and the clock epoch. The location and strength of such peaks are completely dependent on the baseline vector and observation schedule used. In the example of Fig. 2, if delay rates had not been used to constrain the solution, the peak would have risen to about 20 nanoradians. The solution could also have been controlled by readily achieved a priori constraints on the Earth rotation and tropospheric parameters on the order of 500 nanoradians and 4 cm, respectively. In the covariances in this article, the delay rate constraints were chosen to avoid placing specific demands on a priori knowledge; but in an operational mode, some combination of delay rates and a priori constraints could be used to control singularities in the determination of geometric delay.

In order to compare the local reference frame technique to a standard ΔDOR observing sequence, in which Earth rotation and static tropospheric effects are not estimated, these effects have been evaluated for a spacecraft and single radio source observation sequence. The radio source used was the one closest to the spacecraft (about 10 degrees away) in Fig. 1, P0019+058. Earth rotation uncertainties of 50 nanoradians per component and tropospheric zenith delay uncertainties of 4 cm were used for the consider analysis. The above Earth rotation uncertainty is typical of what will be delivered to the DSN for the Galileo and Magellan missions. The tropospheric uncertainty of 4 cm reflects the error in wet zenith delay calibration derived from surface meteorology. Figure 3 shows the contributions of these two error sources as a function of hour angle. Comparison with the local reference frame system noise error of Fig. 2, shown in Fig. 3 as a broken line, shows that estimating Earth orientation and tropospheric parameters in the multisource observation strategy reduces the tracking error due to these effects by an approximate factor of 3 for most hour angles.

C. Covariance Due to Radio Source Structure

In this subsection, the effect of time-invariant radio source structure will be discussed. Time-varying source structure will be mentioned in Section V, but no detailed method for minimizing its effect is given in this article. If the absolute location of the spacecraft is required in the radio frame, radio source

⁵ The first source in the table was observed too recently to appear in the publication referenced, but its coordinates are derived from the same analysis as the other sources in the table.

structure at the 5-nanoradian level imposes a serious limitation. In the observation configuration of Section A, the absolute measurement of spacecraft location is degraded to the 10-nanoradian level due to 5-nanoradian, time-invariant source position errors caused by structure effects. Radio source maps could be made at regular intervals to calibrate the structure effects, but that effort would be costly and in contrast to the self-contained nature of the local reference frame technique. However, as will be shown below, uncalibrated source structure effects can be dramatically reduced for relative angular measurements. In many applications, measurement of the spacecraft's change in angular position over some period of time can provide valuable navigation information. The stability of typical radio source structures over 6-month to 1-year periods can then be exploited. Because the induced signature of constant source structure over these time periods is largely the same, it cancels in the relative position measurement. This statement is treated rigorously in Appendix B on stationary source structure, in which it is shown that maximum error canceling between observations is facilitated by using the same observation sequence.

Using the same observation sequence helps in two ways: (1) time-invariant source structure effects on each delay and delay rate observation will repeat because the projection of the source component vectors onto the baseline will repeat; and (2) as shown in Appendix B, observable errors which are completely correlated will largely cancel in the differencing of sequential spacecraft position measurements if the partial derivatives and system noise covariances are similar. The time-invariant structure effect will be identically zero if exactly the same observation sequence is used. Using similar observation sequences means that the spacecraft position cannot change substantially or the parameter estimation formalism, which leads to the multisource analog of Eq. (5), will propagate the structure effect differently from one angular determination to the next. This means that, in order to minimize the source structure effect and achieve the few-nanoradian relative angular accuracy suggested by Fig. 2, the spacecraft cannot move significantly in the sky between measurements, which is a typical geometry for an outer planet encounter.

In Fig. 4, the source structure effect on relative angular measurements is shown, assuming that the spacecraft moved from the *S* to the indicated points on the two circles between measurements. An apparent error of 5 nanoradians in radio source position due to structure was assigned for $\alpha \cos \delta$ and δ , where α is right ascension and δ is declination, for all radio sources. This error is random from source to source, but does not vary from epoch to epoch. For the smaller, 3-degree circle, which represents approximately 40 days of travel time for asymptotic speeds of 10 km/sec, 1- to 3-nanoradian errors in spacecraft position estimates are incurred due to source struc-

ture. For the larger, 6-degree circle, source structure errors grow to between 2 and 7 nanoradians. This error growth as the spacecraft moves from its initial position is a result of the changing least squares equations, as mentioned above.

V. Summary, Improvements, and Demonstrations

In the absence of radio source structure, the covariance analyses presented here show that VLBI observations of approximately five radio sources along with a spacecraft can yield 2- to 3-nanoradian tracking results. This accuracy, based on Mark III radio source measurement noise levels, is attainable through use of the radio source observations to solve for clock, Earth orientation, and static tropospheric effects. In addition to system noise, the error contributions from tropospheric fluctuations and radio source structure uncertainties have been evaluated in consider analyses. While the tropospheric fluctuation error contributions are smaller than errors due to the assumed VLBI system noise, radio source structure can contribute up to 10 nanoradians to the error in locating a spacecraft relative to the radio frame. However, this error source can be reduced to approximately 2 nanoradians in many applications where the relative spacecraft angle over small pieces of the sky is of interest. It has been shown elsewhere (see footnote 1) that using a 2-nanoradian differential measurement to sense Jupiter's gravitational field yields target relative tracking accuracies competitive with onboard optical navigation. However, subnanoradian accuracies will be necessary to use the same technique at Neptune, for example, and improvements to the technique should be considered.

Improvements to the technique could result from reducing the error contributions from system noise, tropospheric fluctuations, and source structure. As has been mentioned above, using the highly accurate phase delay will reduce system noise contributions almost two orders of magnitude. This would mean that system noise would contribute on the order of 20 picoradians to the local reference frame technique. Obtaining phase delay data is challenging in that it requires resolving cycle ambiguities at 8400 GHz or higher radio frequencies. Resolving cycle ambiguities is theoretically possible with the Mark III system and DSN instrumentation,⁶ but this capability has not yet been demonstrated on intercontinental baselines. Even if it is demonstrated, the accuracy inherent in the phase delay would be of marginal use if the other error sources considered above remained at the quoted levels.

⁶C. D. Edwards, K. M. Liewer, C. S. Jacobs, and R. N. Treuhaft, "An Algorithm for Connection of RF Phase in VLBI," JPL Interoffice Memorandum No. 335.3-475 (internal document), Jet Propulsion Laboratory, Pasadena, California, November 20, 1986.

The obvious means of reducing the tropospheric fluctuation error is by using water vapor radiometers (WVRs) [13]. Since typical water vapor fluctuations on 30-minute time scales are on the order of 5 mm [11], the WVR will have to track fluctuations at or below the millimeter level to substantially reduce the fluctuation contribution to the tracking error. WVR comparisons with spectral hygrometers, which have been conducted in the Advanced Systems Program, suggest the possibility of that precision. Note that apparent zenith delay biases in the WVR measurement, which have been suggested by some existing WVR data [14], might be tolerated as long as they could be parameterized and estimated in place of the zenith parameters appearing in Eqs. (11) through (14).

If WVRs were successfully applied to remove wet tropospheric fluctuation errors, the dry fluctuations, which are estimated to be about 25 percent of the wet on the time scale of the observation sequence, would then limit the angular measurement to approximately 300 picoradians. A possible albeit unexplored method of reducing the dry fluctuation effect is to measure unmodeled phase fluctuations on Earth-orbiting satellites to sense the dry component. Barometric arrays should also be considered. Another possibility for improving the angular measurement accuracy beyond 300 picoradians is to put an interferometric system on an orbiting platform or on the moon. In that case, the Earth rotation parameters in the system discussed here would be replaced by parameters describing the orientation of the platform or the moon. Since the variations of these parameters may be very different from Earth orientation parameters, the entire local reference frame strategy could conceivably be altered, perhaps ultimately reverting back to a single-radio-source Δ DOR strategy if platform orientation can be modeled at the 20-picoradian level.

It has been shown that tracking errors due to time-invariant radio source structure can be greatly reduced in relative measurements over small pieces of the sky. If tracking relative angular position is needed over long periods of time, the assumption that the source structure is stationary may not be valid. In these cases, it may be possible to parameterize source structure evolution⁷ and use the accumulated VLBI data to estimate and remove the structure effect. The feasibility of this technique will be explored as soon as a data base on structure effects can be accumulated.

If, some time in the future, the target ephemeris is known to the few-nanoradian level in the radio reference frame, it may be desirable to have absolute rather than the differential angu-

lar measurements discussed above. In other words, determinations of a spacecraft's position relative to the radio sources may be preferable to determinations of its position relative to previous positions. As already mentioned, 10-nanoradian absolute errors can be expected from source structure in the local reference frame technique. Parameterization of source structure may again be a possibility. Another approach for decreasing structure-related errors is to use the capability of the Mark III system to detect a much larger set of weaker radio sources and select only the most pointlike of those sources. The trade-off between source structure effects and system noise errors will have to be evaluated in this approach. It may also be possible to use the structure information obtained by observing programs on the Very Long Baseline Array [15] to calibrate the structure of sources used in local reference frames.

A potentially important error source which has not been considered in this analysis is that of dispersive phase effects across the bandpasses of the receiving instrumentation. Because the spacecraft signal is a tone and the radio sources are broadband noise, variations in the phase response across the bandpass will introduce tracking errors. There may be ways to include static bandpass error into the parameter estimation scheme. It is also possible that digital front-end electronics being explored in the Advanced Systems Program for the DSN will allow calibration or elimination of this effect.

Demonstrations of the local reference frame technique are under way to test the assumptions of the covariance analysis in this report. Three experiments using the Mark III system have been conducted between California and Australia. In the first experiment, conducted on March 21, 1988, the gravitational bend by Jupiter of the ray path of the radio source P0201+113 was investigated using the local reference frame technique. This gravitational shift in angular position of about 10 nanoradians simulates a spacecraft motion of about 8 km at Jupiter. On April 2 and May 16, when Jupiter was far from the ray path, P0201+113 was observed again in the same local reference frame of radio sources. About five other groups of radio sources were also observed to test the technique. At the time of this report, fringes have been found for the first two experiments with the Block II correlator at Caltech. However, the accuracy of the measurements and the resulting significance of the detection of gravitational bending are as yet undetermined.

If the technique described is validated in a non-real-time mode through experiment, an important practical operational issue will be the transport of large data volumes in real time. As opposed to the current VLBI operational system, which relays 500 kbits/sec from the stations, the Mark III system

⁷J. S. Ulvestad, "Possible Source Structure Effects in IRIS Data," JPL Interoffice Memorandum No. 335.3-88-15 (internal document), Jet Propulsion Laboratory, Pasadena, California, February 3, 1988.

needed to achieve the derived accuracies will require transportation of approximately 50 to 100 Mbits/sec. Although the Advanced Systems demonstrations of the local reference frame technique will not be done in real time, a real-time operational

system would have to employ relay satellites or optical fiber links. Data transport rather than analysis will probably be the most time-consuming step in producing tracking results for navigation.

Acknowledgments

I would like to thank the following people who made valuable contributions to the formulation presented in this article: C. D. Edwards for independent checks on some of the numerical results and comments on the manuscript; T. M. Eubanks for interpretations of local reference frame geometry; J. S. Border and J. S. Ulvestad for critical readings of the manuscript; G. E. Lanyi for clarifying relevant aspects of linear algebra; C. L. Thornton for suggesting techniques for understanding poorly determined parameter estimates; and J. L. Fanselow for initially encouraging investigation into a multisource observation strategy.

References

- [1] J. S. Border, F. F. Donovan, S. G. Finley, C. E. Hildebrand, B. Moultrie, and L. J. Skjerve, "Determining Spacecraft Angular Position With Delta VLBI: The Voyager Demonstration," *AIAA Paper 82-1471*, 1982.
- [2] A. E. E. Rogers et al., "Very-Long-Baseline Radio Interferometry: The Mark III System for Geodesy, Astrometry, and Aperture Synthesis," *Science*, vol. 219, pp. 51-54, 1983.
- [3] J. K. Miller and K. H. Rourke, "The Application of Differential VLBI to Planetary Approach Orbit Determination," *DSN Progress Report 42-40*, pp. 84-90, May 1977.
- [4] T. Gold, "Radio Method for the Precise Measurement of the Rotation Period of the Earth," *Science*, vol. 157, pp. 302-304, July 1967.
- [5] A. E. E. Rogers, "Very Long Baseline Interferometry With Large Effective Bandwidth for Phase-Delay Measurements," *Radio Science*, vol. 5, pp. 1239-1247, October 1970.
- [6] J. B. Thomas, "An Analysis of Long Baseline Radio Interferometry," *JPL Technical Report 32-1526*, vol. VII, Jet Propulsion Laboratory, Pasadena, California, pp. 37-50, 1972.
- [7] O. J. Sovers, C. D. Edwards, C. S. Jacobs, G. E. Lanyi, K. M. Liewer, and R. N. Treuhaft, "Astrometric Results of 1978-85 Deep Space Network Radio Interferometry: The JPL 1987-1 Extragalactic Source Catalog," *Astronomical Journal*, vol. 95, pp. 1647-1658, June 1988.
- [8] J. S. Ulvestad, "Effects of Source Structure on Astrometry and Geodesy," in *Proc. of IAU Symposium No. 129*, pp. 429-430, 1987.

- [9] W. C. Hamilton, *Statistics in Physical Science*, New York: Ronald Press, pp. 124–132, 1964.
- [10] G. J. Bierman, *Factorization Methods for Discrete Sequential Estimation*, New York: Academic Press, p. 169, 1977.
- [11] R. N. Treuhaft and G. E. Lanyi, “The Effect of the Dynamic Wet Troposphere on Radio Interferometric Measurements,” *Radio Science*, vol. 22, pp. 251–265, March 1987.
- [12] C. D. Edwards, “Short Baseline Phase Delay Interferometry,” *TDA Progress Report 42-91*, vol. July–September 1987, Jet Propulsion Laboratory, Pasadena, California, pp. 46–56, November 15, 1987.
- [13] G. M. Resch, D. E. Hogg, and P. J. Napier, “Radiometric Correction of Atmospheric Path Length Fluctuation in Interferometric Experiments,” *Radio Science*, vol. 19, pp. 411–422, 1984.
- [14] D. M. Tralli, E. C. Katsigiris, and T. H. Dixon, “Estimation of Wet Tropospheric Path Delays in GPS Baseline Solutions for the 1986 Caribbean Experiment,” *EOS*, vol. 69, p. 332, April 1988.
- [15] J. D. Romney, “The Very Long Baseline Array,” in *Proc. of IAU Symposium No. 129*, 1987.
- [16] G. E. Lanyi, “Tropospheric Calibration in Radio Interferometry,” in *Proc. Int. Symp. Space Tech. Geodyn., IAG/COSPAR*, vol. 2, pp. 184–195, 1984.
- [17] J. L. Davis, T. A. Herring, I. I. Shapiro, A. E. E. Rogers, and G. Elgered, “Geodesy by Radio Interferometry: Effects of Atmospheric Modeling Errors on Estimates of Baseline Length,” *Radio Science*, vol. 20, pp. 1593–1607, 1985.

Table 1. Radio source and spacecraft positions used for local reference frame covariance analysis

Radio source name	Right ascension, hours/minutes/seconds			Declination, degrees/minutes/seconds		
P0048-09	0	50	41.3180	-9	29	5.216
P0019+058	0	22	32.4413	6	8	4.272
P0106+01	1	8	38.7711	1	35	0.320
P2320-035	23	23	31.9538	-3	17	5.022
P2345-16	23	48	2.6085	-16	31	12.019
Spacecraft	0	20	0.0000	-5	0	0.000

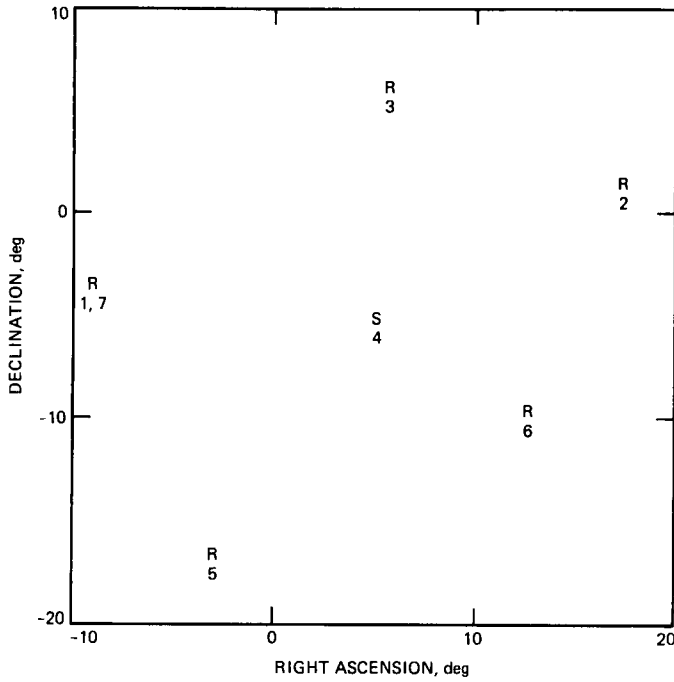


Fig. 1. The radio source and spacecraft angular distribution used in all covariance analysis results of Section III. In that section, only the California-Australia baseline is considered.

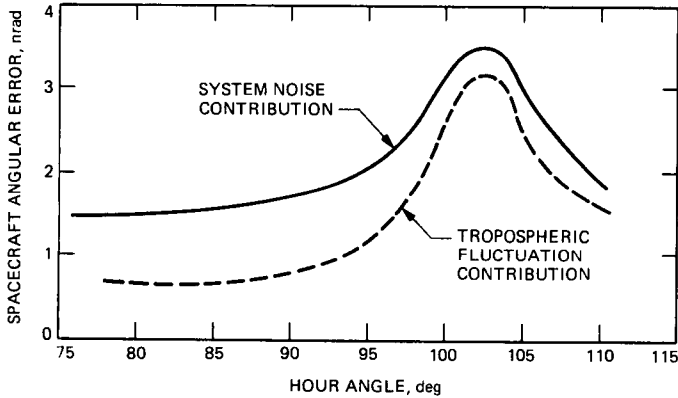


Fig. 2. Local reference frame covariance results: the spacecraft angular error (Δs_p of Eq. [1]) due to system noise of 30 picoseconds per observation and 1-cm daily zenith wet tropospheric fluctuations, as a function of hour angle. The hour angle is the angle between the spacecraft and the baseline at the middle of the 7-observation sequence.

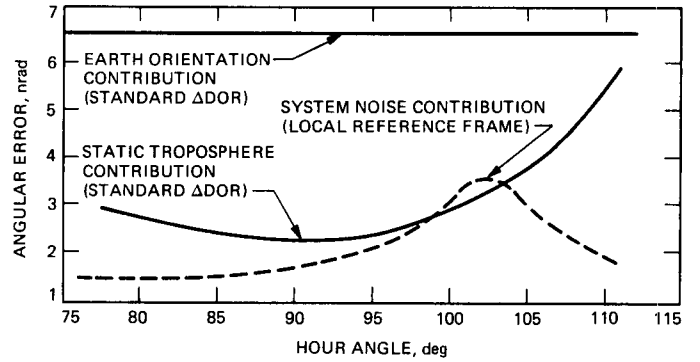


Fig. 3. Standard Δ DOR covariance results: the standard Δ DOR (single radio source) angular error due to 50-nanoradian Earth orientation uncertainty and 4-cm tropospheric delay uncertainty. For comparison, the dashed line is the local reference frame system noise covariance from Fig. 2.

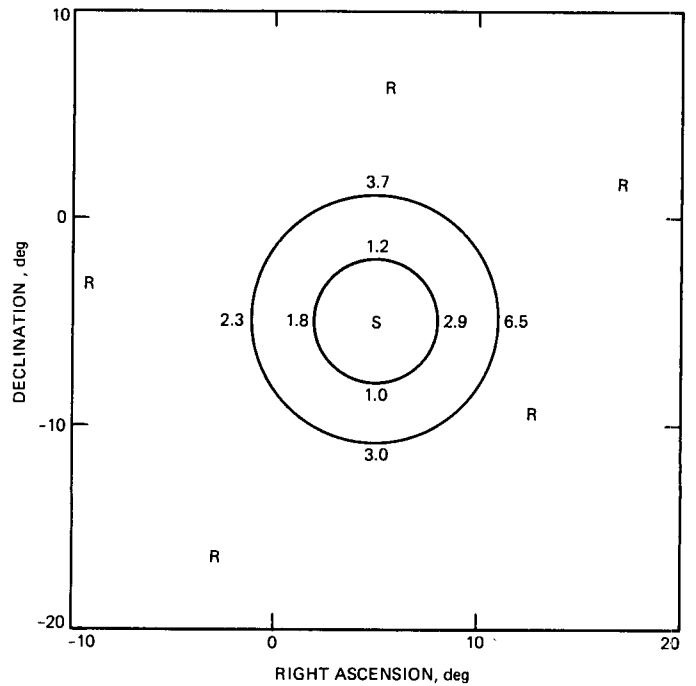


Fig. 4. Differential angular position error due to stationary source structure: the numbers indicate the error, in nanoradians, in a measurement of relative spacecraft position due to time-invariant source structure errors. The numbers apply to a differential position measurement between the S and the indicated point on the circle. All radio sources were assumed to have 5-nanoradian errors in $\alpha \cos \delta$ and in δ where α is right ascension and δ is declination.

Appendix A

Partial Derivatives of Delay and Rate

In the case of standard Δ DOR, the partial derivatives of observed delay with respect to the spacecraft and clock delay parameters, $\partial\tau_{sc}/\partial\tau_g$ and $\partial\tau_{sc}/\partial\tau_c$, are unity. Below, the partial derivatives with respect to all the additional parameters in Eq. (11) are expressed in terms of source and baseline components and observation epochs. The partial derivatives of delay rate are similarly treated. Since the only differences between a spacecraft and radio source observation in this treatment are the objects' angular positions and the observation epochs, the partial derivatives below apply to either type of observation. The subscripts *sc* and *rs* in Eqs. (11) and (12) will therefore be ignored. Five partial derivatives, one for each parameter added in the local reference frame approach, are given below. The derivatives are of delay with respect to station-differenced clock rate, two rotation angles, and two zenith tropospheric delays.

The partial derivative of observed delay with respect to station-differenced clock rate is

$$c \frac{\partial\tau}{\partial\dot{\tau}_c} = t - t_0 \quad (\text{A-1})$$

where t is the observation epoch and t_0 is the average of all spacecraft and radio source observation epochs.

The first Earth rotation partial derivative in Eq. (11) is as follows:

$$\begin{aligned} \frac{\partial\tau}{\partial\theta} = & B_{m\ x|t} (\cos\delta \sin\alpha \sin\delta_c - \sin\delta \cos\delta_c \sin\alpha_c) \\ & + B_{m\ y|t} (\sin\delta \cos\delta_c \cos\alpha_c - \cos\delta \cos\alpha \sin\delta_c) \\ & + B_{m\ z|t} (\cos\delta \cos\delta_c \sin(\alpha_c - \alpha)) \end{aligned} \quad (\text{A-2})$$

where, as defined in the text, θ is the magnitude of a rotation about the $\hat{s}_{m\ sc} \times \vec{B}_m$ axis, where \vec{B}_m is the a priori baseline vector at the spacecraft observation epoch. Note that $B_{m\ x|t}$, $B_{m\ y|t}$, and $B_{m\ z|t}$ are the a priori baseline components, in a space-fixed coordinate system, evaluated at the time t at which τ is measured. In Eq. (A-2), α , δ , α_c , and δ_c are the right ascensions and declinations of the observed source and the rotation axis, respectively. The partial derivative of delay with respect to ϕ , the magnitude of the rotation about $\vec{B}_m \times \hat{s}_{m\ sc} \times \vec{B}_m$, is given by an expression identical to Eq. (A-2) with α_c and δ_c replaced by the right ascension and declination of the $(\vec{B}_m \times \hat{s}_{m\ sc} \times \vec{B}_m)$ vector.

The partial derivative of observed delay with respect to observed station 1's zenith troposphere delay is given by

$$\frac{\partial\tau}{\partial z_1} = \frac{-1}{\sin\gamma_1} \quad (\text{A-3})$$

where γ_1 is the elevation angle at station 1. The error in covariance results incurred by not using a more realistic zenith partial derivative is about 0.1 nanoradian. When subnanoradian accuracies are attempted, the dry and wet tropospheres, each with its own mapping functions [16], [17], will need more careful treatment than implied by Eq. (A-3). An expression for the partial derivative of geometric delay with respect to station 2's zenith troposphere is obtained by substituting 2 for 1 in Eq. (A-3) and omitting the minus sign.

The partial derivatives of delay rate with respect to the model parameters of Eqs. (13) and (14) are given below. The first three derivatives of rate with respect to spacecraft geometric rate and clock parameters are

$$\begin{aligned} \frac{\partial\dot{\tau}}{\partial\dot{\tau}_g} &= 1 \\ \frac{\partial\dot{\tau}}{\partial\dot{\tau}_c} &= 0 \\ \frac{\partial\dot{\tau}}{\partial\dot{\tau}_c} &= 1 \end{aligned} \quad (\text{A-4})$$

where all symbols are defined in the text. The first line of Eq. (A-4) applies to the spacecraft rate observation only, as the partial of radio source observed rate with respect to spacecraft geometric rate is zero. Eq. (A-4) states that delay rate parameters map directly into delay rate observations, and the value of the clock epoch parameter does not affect the delay rate observations.

Because an important component of the modeled geometric delay rate is due to baseline rotation about the Earth's spin axis, baseline orientation errors cause residual observed delay rate. The partial derivatives are found by performing an infinitesimal rotation on the baseline and calculating the difference in delay rate between the rotated and original baseline. For a rotation about the $\hat{s}_{m\ sc} \times \vec{B}_m$ axis, with rotation axis right

ascension and declination α_c and δ_c , the delay rate partial derivative is

$$c \frac{\partial \tau}{\partial \theta} = \Omega \cos \delta \times \left[\cos \alpha (B_{mz|t} \cos \delta_c \cos \alpha_c - B_{mx|t} \sin \delta_c) + \sin \alpha (B_{mz|t} \cos \delta_c \sin \alpha_c - B_{my} \sin \delta_c) \right] \quad (\text{A-5})$$

where Ω is the Earth's rotation rate in radians/sec. A similar expression holds for $\partial \tau / \partial \phi$.

For the partial derivatives of delay rate with respect to zenith delay, the troposphere is taken to be static. That is, the rate signature is again induced by Earth rotation, which causes an elevation angle rate. The partial derivative is found numerically by calculating elevation angles at plus and minus two seconds from the observation epoch. The rate of change of Eq. (A-3) is thereby obtained.

Appendix B

The Effect of Stationary Radio Source Structure on Differential Angular Position Determinations

As noted in Section IVC, time-invariant source structure has less effect on differences in angular position than on absolute angular position in the radio frame. In this appendix, the effect of time-invariant source structure on the angular difference determination is evaluated. Figure 4 results from the derivation below. The treatment applies equally well to radio source position errors, because the source structure errors can be represented by unknown shifts in the effective radio source positions. For the source structure calculation, it will be assumed that the observing schedule for the radio sources is identical for the two local reference frame measurements. The spacecraft will be allowed to move. Its observation epoch will be taken to be identical for the two measurements, but its position on the sky (and therefore relative to the baseline) will be allowed to change. For the following calculation, the effect of constant but unknown errors on a general parameter estimate, \hat{X}_i , will be considered. This generalized notation is consistent with the index manipulations in the derivation. For the relative tracking problem in question, \hat{X}_i should be thought of as the residual spacecraft geometric delay parameter.

In the linear least squares formalism, the estimate of the i th parameter, \hat{X}_i , is given by

$$\hat{X}_i = \sum_{j=1, N} F_{i,j} \Delta\tau_j \quad (\text{B-1})$$

where N is the number of observations and the $\Delta\tau_j$ s are the residual observation delays due solely to source structure effects. The $F_{i,j}$ s are the least squares coefficients which describe the contribution of the j th observation to the i th parameter estimate; they are explicitly defined in Eq. (B-3) below. Equation (5) in the text is an example of Eq. (B-1) for the specific case of ΔDOR . An estimate of the same parameter, from observations at a later epoch, will be denoted by

$$\hat{X}'_i = \sum_{j=1, N} F'_{i,j} \Delta\tau_j \quad (\text{B-2})$$

The $\Delta\tau_j$ terms are not primed because it is assumed that they arise from the same source structure or uncertainty effects for both the unprimed and primed observations. Although the parameter estimates in this appendix are derived from $\Delta\tau_j$ delay measurements only, this treatment can be extended by

including delay rates in Eqs. (B-1) and (B-2). The $F_{i,j}$ and $F'_{i,j}$ coefficients are given by [9]

$$\begin{aligned} F_{i,j} &= [(A^T W A)^{-1} A^T W]_{i,j} \\ F'_{i,j} &= [(A'^T W A')^{-1} A'^T W]_{i,j} \end{aligned} \quad (\text{B-3})$$

The matrix A is defined by $A_{m,n} = \partial\tau_m / \partial X_n$. The A matrix includes the partial derivatives calculated in the previous appendix. If the spacecraft has moved between the unprimed and primed observations, then the partial derivatives of spacecraft delay with respect to model parameters in the A' matrix will be different from those in the A matrix; all other elements of A and A' will be identical to each other. In Eq. (B-3), W is the inverse of the modeled observable covariance matrix. It has been assumed that the modeled observable covariances, such as system noise, do not change between the unprimed and primed observations. Although this need not be assumed to complete the derivation, the modeled covariances would generally be nearly identical for each observation set.

In order to calculate the variance of relative \hat{X}_i determinations due to source structure, the $\Delta\tau_j$'s will be associated with geometric delay errors induced by position uncertainties. Below, $\Delta\tau_j$ is written in terms of parameter residuals ΔP_k :

$$\Delta\tau_j = \sum_k \frac{\partial\tau_j}{\partial P_k} \Delta P_k \quad (\text{B-4})$$

where the P_k parameters are the right ascensions and declinations of the observed radio sources. Obviously, the partial derivative in Eq. (B-4) vanishes unless P_k is the right ascension or declination of the radio source being observed in the j th VLBI delay measurement.

The variance of $(\hat{X}_i - \hat{X}'_i)$ in terms of the covariance of the P_k source coordinate parameters will now be calculated. Expressions for the variance of \hat{X}_i will be needed. From Eqs. (B-1) and (B-4), it is

$$\text{var}(\hat{X}_i) = \sum_{j,k,l,m} F_{i,j} \frac{\partial\tau_j}{\partial P_k} M_{p,k,l} \frac{\partial\tau_m}{\partial P_l} F_{i,m} \quad (\text{B-5})$$

where $M_{p,k,l}$ is the covariance of the k th and l th source coordinate errors. Expressions similar to Eq. (B-5) can be derived for the variance of \hat{X}_i and the covariance of \hat{X}_i and \hat{X}_i' , leading to the final result, shown below in matrix form:

$$\begin{aligned} \text{var}(\hat{X}_i - \hat{X}_i') &= (FDM_p D^T F^T)_{i,i} + (F'DM_p D^T F'^T)_{i,i} \\ &\quad - 2(FDM_p D^T F'^T)_{i,i} \end{aligned} \quad (\text{B-6})$$

where $D_{i,j} \equiv \partial \tau_i / \partial P_j$.

Equation (B-6) illustrates the parameter error cancellation which can arise when parameter estimates are differenced in

the presence of unknown but repeatable errors. If the observing schedules were exactly the same—that is, if the spacecraft did not move at all on the sky—then all primed quantities would equal all unprimed quantities and Eq. (B-6) would yield identically zero. Small changes in spacecraft position make small differences between the F and F' and A and A' matrices, yielding the results of Fig. 4.⁸ Note that the crucial step in realizing the cancellation was assuming that the $\Delta \tau_j$'s were identical for the unprimed and primed measurements. This is equivalent to saying that the source structure or position uncertainties did not change from one observation to the next.

⁸The actual calculation of the results for Fig. 4 employed the generalization of Eq. (B-6), which includes both delay and delay rate observations.

52-13

1989000817

N89-10188

TDA Progress Report 42-94

April-June 1988

Precise Orbit Determination for NASA's Earth Observing System Using GPS

B. G. Williams
Navigation Systems Section

An application of a precision orbit determination technique for NASA's Earth Observing System (EOS) using the Global Positioning System (GPS) is described. This technique allows the geometric information from measurements of GPS carrier phase and P-code pseudo-range to be exploited while minimizing requirements for precision dynamical modeling. Briefly, the method combines geometric and dynamic information to determine the spacecraft trajectory; the weight on the dynamic information is controlled by adjusting fictitious spacecraft accelerations in three dimensions which are treated as first-order exponentially time-correlated stochastic processes. By varying the time correlation and uncertainty of the stochastic accelerations, the technique can range from purely geometric (for zero time correlation, infinite uncertainty) to purely dynamic (for infinite time correlation, zero uncertainty). Performance estimates for this technique as applied to the orbit geometry planned for the EOS platforms indicate that decimeter accuracies for EOS orbit position may be obtainable. The sensitivity of the predicted orbit uncertainties to model errors for station locations, non-gravitational platform accelerations, and Earth gravity is also presented.

I. Introduction

NASA's Earth Observing System will be a primary user of the Space Station polar platforms planned for the 1990s. The EOS project requirements have helped define the payload requirements for the polar platforms and have resulted in preliminary plans for three platforms to be flown concurrently in order to provide a long-term (10-year) data base of Earth science information. The platforms will carry a variety of remote sensing instruments, several of which will require or will benefit from precise orbit determination. These include a precision radar altimeter (similar in performance to the TOPEX altimeter), the High Resolution Imaging Spectrometer (HIRIS), and the Thermal Infrared Imaging Spectrometer (TIMS) [1].

The three platforms, referred to as platforms 1, 2, and 3, will be in near-circular, sun-synchronous orbits at 824 km altitude. The orbit of each platform will have a sixteen-day ground track repeat and will be near-polar. The orbit plane of platforms 1 and 2 will have its ascending node at 1:30 p.m. (local solar time, LST) on the Earth's true equator of date, while platform 3 will have its descending node at 9:30 a.m. (LST). The orbits of the three platforms differ only in their

phasing relative to the GPS constellation and the ground receivers being simulated in this analysis, and hence are essentially the same for the purposes of this preliminary covariance analysis. Of course, a more detailed mission design analysis for each platform would be concerned with the day-to-day variability in orbit uncertainties due to these phasing differences. With this in mind, platform 3 orbit parameters were used in this study to simulate geometry typical of all three platforms. The orbit parameters assumed for platform 3 were based on the current mission baseline design¹ and are presented in Table 1. The epoch time shown in Table 1 is arbitrary and was chosen to coincide with the epoch of existing trajectory predictions for the GPS constellation. The orbit node relative to the Earth true equator of date shown in Table 1 was chosen to place the orbit descending node at 9:30 a.m. (LST) on the epoch date.

The GPS tracking analysis techniques applied in this study are among several that have been developed at JPL over the past several years to probe the ultimate on-orbit precision available from GPS [2], [3], [4]. The techniques all require a GPS receiver aboard the spacecraft to be tracked and a precisely known global network of GPS ground receivers. The basis of these sub-decimeter GPS tracking strategies is their ability to exploit the extreme precision of carrier phase tracking by using it to smooth the geometric solutions obtained from the less precise pseudo-range measurements [3], [4]. The application of two of these techniques, the "non-dynamic" and "reduced-dynamic," to the determination of EOS orbits is the subject of this article. The reduced-dynamic method is the more robust of the two techniques, as will be shown in the results to follow. The reduced-dynamic method has been described in general for Earth orbiters as a method for exploiting the redundant geometric information available from GPS measurements while minimizing requirements for precision dynamical models. This technique is less sensitive to momentary viewing geometry between the EOS receiver and GPS, as are the non-dynamic methods, which eliminate orbit dynamics entirely [2].

This analysis used the Orbit Analysis and Simulation Software (OASIS), which was developed at JPL especially for studying GPS tracking performance over a wide range of applications [5]. OASIS capabilities include simulation/covariance analysis features for a variety of model parameters, including multi-spacecraft states and dynamic parameters, tracking station location parameters, media delay parameters, and a host of clock modeling parameters for both spacecraft and ground receivers. Any of these model parameters (except

spacecraft states) may be treated as piecewise constant stochastic process noise in the filter. This flexibility of design allows the random clock behavior of GPS transmitters and receivers to be eliminated by modeling clock biases as uncorrelated stochastic processes at each measurement time. Also, the ability to model three-dimensional accelerations on EOS as exponentially time-correlated stochastic processes is the heart of the reduced-dynamic tracking technique. In this technique, the relative weighting of dynamics and geometry may be adjusted continuously by varying the a priori standard deviation and the correlation time of the stochastic acceleration. By selecting large a priori and steady-state standard deviations and a zero correlation time for these accelerations, the filter becomes the purely geometrical non-dynamic tracking technique.

The goal of this article is to demonstrate that if an advanced GPS receiver is flown on EOS (similar to the TOPEX GPS receiver) and if the data are processed correctly, then the orbit determination for EOS could achieve decimeter-level accuracies. This article outlines results from two differential GPS techniques and the system requirements that could be used to reach these decimeter-level accuracies. The study includes GPS tracking strategy for the EOS flight receiver as well as filter strategies and covariance results for short arcs of one and two orbits of EOS tracking.

II. Radio Metric Data Simulation

A. Simulation Assumptions

The data simulation assumed a fully operational GPS constellation consisting of eighteen GPS spacecraft, with three in each of six orbit planes. The orbit planes are inclined 55 degrees to the equator and are equally spaced 60 degrees apart in longitude. The three spacecraft in each orbit plane are equally spaced (120 degrees) and are phased 40 degrees from plane to plane to ensure global visibility (for ground sites) of at least four spacecraft. This is the anticipated configuration of the operational GPS assembly that is scheduled for completion sometime in 1990 [6].

As stated earlier, the orbit recovery methods used here require both a GPS receiver aboard EOS and a precisely known global network of ground GPS receivers. For this study, a worldwide network of ten GPS ground receivers was assumed with an a priori location uncertainty equal to 5 cm in each of three orthogonal directions. An elevation cutoff angle of ten degrees above the local horizon was also assumed for each ground receiver. A world plot showing the global distribution of these sites and the EOS orbit ground track for this study is presented in Fig. 1. It should be noted that the operation of the ground-based receiver network is completely independent of the EOS platform. These receivers make measurements only

¹H. N. Norton (ed.), "EOS Phase A Final Report," JPL Report D-4566 (internal document), Jet Propulsion Laboratory, Pasadena, California, April 20, 1988.

to the GPS, and the improved GPS orbits are used to determine the EOS orbit indirectly. In fact, once such a global network is established, any number of other user spacecraft which are carrying GPS receivers could make use of the system for precision orbit determination with no impact on the network operation or on each other (assuming that processing is distributed to different host computers).

The GPS receiver aboard EOS was assumed to have the capability to track continuous carrier phase and P-code pseudo-range from five GPS spacecraft simultaneously through an antenna with a hemispherical field of view centered at the zenith; ground-based GPS receivers were assumed able to track up to eight spacecraft simultaneously. The measurement and timing precision requirements for the receivers are summarized in Table 2, along with other pertinent simulation assumptions. The measurement precision assumed for all receivers was 0.5 cm for carrier phase and 5 cm for pseudo-range over a five-minute integration time. This implies that the dual-frequency measurements have been combined to remove the first-order ionospheric delay. These requirements are within the expected capabilities of the next generation of GPS receivers being developed both for NASA's Deep Space Network and for flight aboard TOPEX in the 1990s.

B. EOS GPS Receiver Scheduling

In order to extract the maximum information from the GPS carrier phase observable, it is important that the receiver be able to continuously track each of the GPS dual-band carrier signals at 1.6 GHz and 1.2 GHz. The continuous count phase observable is ambiguous to an integer number of cycles of the carrier, which requires that an initial phase bias be solved for in addition to other state parameters. If the receiver momentarily loses "lock" and the continuous count is interrupted, then additional phase bias parameters must be estimated at each phase break. In general, when the number of phase breaks decreases, the solution strength increases because there are fewer parameters to estimate and because dynamic information is accumulated from continuous phase observables.

To show the importance of satellite selection and its impact on the non-dynamic tracking technique, the EOS flight receiver assumed for this study was capable of tracking only five GPS spacecraft simultaneously. Since EOS is in a near-polar, retrograde orbit, there is a sparsity of spacecraft to be tracked during some parts of the orbit. The number of GPS spacecraft visible to the assumed EOS antenna for the geometries assumed in this study is presented in Fig. 2. As seen in the figure, the number of visible spacecraft ranges from five to eight. For optimal tracking, the selection algorithm may have to trade off the geometrical strength of a particular configuration in favor of minimizing switching between the spacecraft. Excessive

switching between GPS spacecraft could weaken the solution through the introduction of unnecessary phase breaks. Of course, there would be an advantage if the EOS receiver were assumed to be able to track all visible GPS, since then these scheduling trade-offs (and the corresponding receiver algorithms) would not be necessary.

The selection of which five GPS spacecraft to track at each measurement time over the two EOS orbits in this simulation was designed to maximize both the length of tracking for a given spacecraft and the geometric strength of the particular five tracked. In order to measure the geometric strength of a given GPS tracking selection, the Position Dilution of Precision [7], or PDOP, parameter was used. Small values of PDOP indicate good arrangements in the geometry and correspondingly small errors in position determination. Using tabular values of all possible PDOP values at each five-minute interval over the simulation, an iterative approach was used to select those five spacecraft at each measurement time which would give reasonably uninterrupted phase tracking with low or minimum PDOP values. The resulting tracks for each spacecraft were no shorter than ten minutes while some reached a maximum length of forty minutes; the average length was thirty minutes. The minimum values of PDOP that were available and the actual values for the tracking schedule used in this study are presented in Fig. 3. Overall, the tracking chosen here achieved the minimum PDOP values 30 percent of the time.

III. Performance Analysis

The simulated radio metric data described above were used to produce orbit covariance results for both the non-dynamic and reduced-dynamic filter strategies. In the sections that follow, the two techniques are compared for processing a single orbit of data, and then the processing is repeated for data arcs in which the geometric strength has been degraded. For this degraded case, it will be seen that the nondynamic solution degenerates sharply while the reduced-dynamic solution remains stable. Finally, some results for the reduced-dynamic filter strategy over arc lengths of two EOS orbits are presented.

A. Assumptions for Covariance Analysis

In each of the filter strategies presented here, all clock biases between receivers and GPS transmitters were solved for at each measurement time. This eliminates the effects of unstable oscillators which produce common systematic effects between receiver-GPS pairs, but the price that is paid is a loss of information and a reduction in geometric strength. Explicit double differencing of the data could be used to eliminate the clocks [2]; however, in the present analysis, this is done by

implicitly double differencing the data by modeling each clock bias as an uncorrelated stochastic process. Essentially, this modeling is equivalent to solving for a new clock bias at each measurement time. In this strategy, one clock is chosen as a reference and is not estimated; hence, all other clock offsets are relative to this master clock. Simulations and reductions of actual GPS radio metric data [8], [9] have demonstrated that these stochastic clock models give results comparable to explicitly double differenced data.

In both non-dynamic and reduced-dynamic strategies, the covariance of a pseudoePOCH filter state was computed for each of the time “batches” over which the stochastic process noise parameters were assumed constant [4]. In this study, the batch length, Δt , was chosen to coincide with the measurement interval of five minutes. In other words, each batch is the same length and contains only one set of measurements from a given GPS/ground station set. The batches for the stochastic clock parameters coincided with those for the stochastic forces. These stochastic processes were characterized by correlation time, τ , a priori variance, σ_0^2 , and steady-state variance, σ_p^2 , which controlled the propagation of the variance, P_j , of the stochastic parameters from batch to batch (i.e., from time t_j to t_{j+1}) as follows:

$$P_{j+1} = m^2 P_j + \sigma^2 \quad (1)$$

where σ^2 is given by

$$\sigma^2 = (1 - m^2) \sigma_p^2 \quad (2)$$

with

$$m = \exp(-\Delta t/\tau) \quad (3)$$

In the case where $\tau \rightarrow 0$, Eq. (1) represents a white noise process with variance σ_p^2 ; as $\tau \rightarrow \infty$ it approaches a random walk process if σ remains non-zero. These concepts will be useful in considering the results to follow. The nature of the stochastic process representing the fictitious forces on EOS determines whether the filter strategy is non-dynamic ($\tau \rightarrow 0$, $\sigma_p \rightarrow \infty$) or dynamic ($\tau \rightarrow \infty$, $\sigma_p \rightarrow 0$). The reduced-dynamic technique combines these two strategies by proper choice of τ , σ_0 , and σ_p . A mathematical description for this filter strategy in terms of a Kalman sequential filter formulation is presented in [4]. In the present study, both non-dynamic and reduced-dynamic techniques assume $\sigma_0 = \sigma_p$.

The filter states that were estimated for each of the strategies included the GPS spacecraft and EOS positions and velocities at epoch (114 parameters), the stochastic clock biases for each receiver (except the reference clock) and spacecraft

clock (28 parameters), the phase ambiguity parameters between GPS transmitters and receivers for each data arc (97 parameters), and the stochastic acceleration on EOS (3 parameters) for a total of 242 estimated parameters. Furthermore, the orbit uncertainties presented here have been adjusted for mismodeled parameters. These “consider” sigmas take into account the a priori errors due to the following:

- (1) Station location errors
- (2) Troposphere delay for each of the ten ground receivers
- (3) GM of Earth
- (4) Gravity harmonics through 4th degree and order
- (5) EOS solar pressure
- (6) EOS atmospheric drag.

The a priori standard deviations of each of the estimated and considered parameters are given in Table 2.

The processing steps used for both the techniques studied involved filtering, smoothing, and mapping of both the estimate state computed covariance and the sensitivity of the estimate state to the considered parameters. First, the filter was used to compute covariance and sensitivity results for each batch in the data arc being processed. Next, these were smoothed, taking into account measurement data from all batches processed. Then these smoothed pseudoePOCH results were mapped to each batch time in order to create “current state” results. These current state results are presented for both tracking techniques in the following sections.

B. Comparison of Two Filter Strategies

The results for non-dynamic processing of a single orbit of EOS tracking are presented in Fig. 4. In the figure, the consider uncertainty for each of three orthogonal components, the radial, cross-track, and down-track (or tangential), and the total Root-Sum-Square (RSS) of these components are plotted every ten minutes over the data span of one orbit (~100 min). The total number of measurements processed over this arc was 2,404, or 1,202 for each data type. This case assumed the stochastic acceleration on EOS had a steady state σ_p of 1.0 cm/sec² and a correlation time of zero. As will be demonstrated in the next section, the variations in uncertainty over time seen in Fig. 4 are the result of the momentary viewing geometry and the common visibility between EOS, GPS, and the ground receivers.

A reduced-dynamic strategy applied to the same one-orbit data arc produced the results shown in Fig. 5. In this case, the assumed stochastic acceleration on EOS had a $\sigma_p = 2.0 \mu\text{m/sec}^2$ and a correlation time of one day. Other reduced-dynamic

cases were run in which the steady state sigma on the EOS stochastic force was as small as $0.5 \mu\text{m}/\text{sec}^2$ with essentially the same result. The result in Fig. 5 shows much smoother uncertainty in each component over the arc than was seen for the non-dynamic case; this is due to time correlation of position fixes through dynamics which are included in the reduced-dynamic scheme but which are removed entirely in the non-dynamic case. The larger error in radial component evident at the final time in both plots is due almost entirely to the considered station location error and is most likely due to an end of data arc effect rather than poor geometry, as it does not appear in the longer arc fit presented later.

The average error contributions for the two cases above are presented in Fig. 6. Each error source and the total RSS are displayed in Radial, Cross-track, and Down-track components, which are labeled as R, C, and D, respectively, in Fig. 6. The errors in this figure represent the root mean square (RMS) of each error source over the single orbit of data in the fit. The errors labeled "Data" in Fig. 6 are the formal error contributions due to the assumed measurement precision. Note that the station location error is the dominant error source in both techniques. As expected, the non-dynamic scheme showed very little sensitivity to mismodeled dynamic parameters with computed perturbations of less than 1 mm in each component due to those parameters. The only exception was a 2-mm radial perturbation due to Earth GM. The reduced-dynamic strategy showed slightly higher sensitivity to dynamic mismodeling, although the largest of these perturbations were less than 2 cm. The perturbations due to dynamic mismodeling for the reduced-dynamic case are presented in Table 3. Note that the largest of the dynamic perturbations in this case are due to gravity harmonics at about a centimeter or less. Further note that the maximum effect of atmospheric drag occurs in the down-track component, although it is still insignificant at 7.2 mm. Since the perturbations due to solar pressure and atmospheric drag were so small for both techniques, they were excluded from Fig. 6.

C. Performance With Degraded Tracking

In this section, the two techniques are compared over the same single-orbit fit as before, but the data arc was degraded in this case by deleting EOS measurements to three GPS spacecraft at 60 and 65 minutes past the epoch time. These two times correspond to the instances when a maximum of five spacecraft are visible as indicated in Fig. 2 at 0.60 and 0.65 orbit periods past epoch. Hence, only two spacecraft are being tracked by the EOS receiver at these times. This was done to simulate conditions that might occur if the selection algorithm in the GPS receiver aboard EOS fails to properly schedule tracking.

The results for the non-dynamic technique, shown for this case in Fig. 7, indicate a dramatic loss of solution strength where the geometric strength has been weakened. The errors computed for this case at other times in the orbit are essentially the same as those shown in Fig. 4, but errors computed for 60 minutes past epoch have increased to over 362 m. The error components at this time were dominated by the formal estimate error (362.2 m due to data noise); all the consider error contributions remained less than 30 cm. The size of the orbit error at the geometric singularity is determined by the a priori uncertainties on the state and the stochastic force; otherwise, the orbit error becomes infinite at these times. This behavior along with the recovery of the orbit for all times except the degraded time shows the essential point positioning nature of the non-dynamic scheme. It should be noted that the rapid loss of solution accuracy seen here can also occur if the ground network lacks sufficient common visibility to produce strong geometric information [2].

The results for reduced-dynamic processing of the degraded data are shown in Fig. 8. Now the errors are hardly perturbed at the 60-minute time point, and in fact the RSS error there has increased to only 13.3 cm compared to the 12.0 cm computed for the corresponding good geometry case of Fig. 5. The smoothing effect of the orbit dynamics has "carried" the solution past the geometric singularity to produce a good overall solution. This robustness of the reduced-dynamic strategy in the presence of momentary loss of geometric strength in the radio metric data contrasts sharply with the non-dynamic results shown in Fig. 7, and makes the reduced-dynamic technique the method of choice for precise GPS tracking.

D. Performance Over Longer Data Spans

In this section, the reduced-dynamic filter technique was applied to two orbits (~ 200 min) of the simulated data. This measurement arc was an extension of the single-orbit data arc analyzed above and contained an extra 2,300 measurements for a total of 4,704, or 2,352 of each data type. This case estimated the same parameters as the single-orbit cases except for an additional 14 phase biases which were due to extra GPS tracks in the longer arc. The a priori parameter uncertainties are again those listed in Table 2. Also, the stochastic acceleration on EOS was assumed to have a steady state sigma, $\sigma_p = 2 \mu\text{m}/\text{sec}^2$, and a correlation time of one day to be consistent with the single-orbit reduced-dynamic cases already presented. The results obtained for position errors in the two-orbit fit are presented in Fig. 9. The corresponding velocity errors for this case are shown in Fig. 10.

The error contributions for two-orbit results were also averaged over time and are presented in Fig. 11. Comparing this figure to reduced-dynamic errors for the single-orbit fit

shown in Fig. 6, note that the optimal filter (i.e., data noise) contribution is smaller for the two-orbit fit (~ 1.3 cm vs. ~ 2.3 cm), and that the sub-optimal filter error (i.e., consider error) in the radial direction has decreased slightly (by 0.74 cm RMS) over that obtained in the single-orbit reduced-dynamic case. However, the total error has grown slightly for the other two components by 1.51 cm RMS in cross track and 0.32 cm RMS in down track. These increased total errors are due to increases in the station location errors and troposphere errors. This counter-intuitive result is due to the sub-optimal filter behavior. Perturbations due to station location errors still dominate the solution, although the particular perturbation at 100 min past epoch is much smaller for the two-orbit case than for the single-orbit fit (compare Figs. 5 and 9). The perturbations due to atmospheric drag and gravity harmonics were slightly less than those in the single-orbit case, while for solar pressure and GM they either remained essentially the same or grew slightly; all the dynamic perturbations for this case are summarized in Table 4.

The slight increase in cross-track and down-track errors is due to weighting of the dynamic information relative to the geometric information which is a result of the particular values for τ and σ_p used in this study. The particularly large value of $\tau = 1d$ relative to the batch size, $\Delta t = 5$ min, may also be at fault since it results in $m = 0.9965$ from Eq. (3), and hence the fictitious forces were being modeled approximately as random walks. These random walk forces had little cumulative effect on the short single-orbit cases, but they could have influenced the errors in the longer two-orbit fit. Although these considerations of optimality merit further study, the results for the one- or two-orbit cases presented here both yield reduced-

dynamic RMS errors of less than ten centimeters for each component.

IV. Conclusions

The EOS orbit can be determined to decimeter levels by GPS tracking if the EOS receiver takes full advantage of the precision inherent in the GPS signal. The GPS receiver aboard EOS should either be designed to receive signals from all GPS spacecraft in view, or be able to optimize selection for at least five spacecraft. In addition, this study assumed a ground network of ten GPS receivers distributed worldwide to ensure good common visibility of the GPS with the EOS receiver. This ground network was sufficient for the current analysis, but the optimal size and placement of such a network was not considered here. The establishment of any such network will certainly not fall entirely on EOS, however, since the tracking system will also be required by other precisely determined Earth orbiters in the 1990s (such as TOPEX). Sharing the network among many such users places no additional tracking burden on the system since the ground receivers track only the GPS.

Either of the data reduction techniques presented here can produce decimeter accuracy without precise dynamical modeling since they rely mainly on the geometric strength of the GPS measurements. However, the reduced-dynamic technique was the clear choice over the non-dynamic strategy whenever the geometric strength of the measurement data was degraded. Hence, this analysis presents reduced-dynamic GPS tracking as the best way to achieve EOS orbit accuracy without precise knowledge of orbit dynamical models.

Acknowledgments

The author wishes to acknowledge the help of P. J. Wolff, S. C. Wu, and T. P. Yunck in the preparation of this article.

References

- [1] R. Hartle and A. Tuyahov, "The Earth Observing System," *Astrodynamics 1985; Proc. Conf., Pt. I*, 1986, pp. 529-550.
- [2] T. P. Yunck and S. C. Wu, "Non-Dynamic Decimeter Tracking of Earth Satellites Using the Global Positioning System," paper AIAA-86-0404, AIAA 24th Aerospace Sciences Meeting, Reno, Nevada, January 1986.
- [3] T. P. Yunck, S. C. Wu, and J. T. Wu, "Strategies for Sub-Decimeter Satellite Tracking with GPS," *Proc. 1986 IEEE Position, Location and Navigation Symp.*, November 1986, pp. 122-128.
- [4] S. C. Wu, T. P. Yunck, and C. L. Thornton, "Reduced-Dynamic Technique for Precise Orbit Determination of Low Earth Satellites," AAS Paper 87-410, AAS/AIAA Astrodynamics Specialist Conf., Kalispell, Montana, August 1987.
- [5] S. C. Wu and C. L. Thornton, "OASIS - A New GPS Covariance and Simulation Analysis Software System," *Proc. First Int. Symp. Precise Positioning with the Global Positioning System*, Rockville, Maryland, May 1985, pp. 337-346.
- [6] B. W. Parkinson and S. W. Gilbert, "NAVSTAR: Global Positioning System - Ten Years Later," *Proc. IEEE*, vol. 71, no. 10, October 1983, pp. 1177-1186.
- [7] R. J. Milliken and C. J. Zoller, "Principle of Operation of NAVSTAR and System Characteristics," in *Global Positioning System*, Washington, D.C.: The Institute of Navigation, 1980, pp. 3-14.
- [8] S. C. Wu, S. M. Lichten, W. I. Bertiger, J. T. Wu, J. S. Border, B. G. Williams, and T. P. Yunck, "Precise Orbit Determination of GPS and Landsat-5," *Proc. Fourth Int. Geodetic Symp. Satellite Positioning*, Austin, Texas, 1986, pp. 275-288.
- [9] S. M. Lichten and J. S. Border, "Strategies for High Precision GPS Orbit Determination," *J. Geophys. Res.*, vol. 92, no. B12, November 10, 1987, pp. 12751-12762.

Table 1. EOS orbit characteristics used in this study

Semi-major axis	7202 km	Longitude of ascending node, ^a Ω	143.1 deg
Eccentricity	0.00114	Argument of perifocus, ω	90.0 deg
Inclination ^a	98.7 deg	True anomaly	0.0 deg
Orbit period	101.4 min		

^aAngles are referenced to Earth true equator and equinox of 21 March 1986, 14 h UTC.

Table 2. A priori standard deviations used in EOS study

EOS and ground GPS receiver performance	
Parameters	Standard deviations
Data noise	0.5 cm at 5-min interval, calibrated carrier phase 5.0 cm at 5-min interval, calibrated P-code pseudo-range
Phase bias	1 sec
Clock bias	1 sec for receivers and satellites
Station location	5 cm each component
Zenith troposphere	1 cm
Spacecraft orbit uncertainty	
Parameters	Standard deviations
GPS state	2 m (X, Y, Z); 0.2 mm/sec (DX, DY, DZ)
EOS state	2 km (X, Y, Z); 2 mm/sec (DX, DY, DZ)
EOS accelerations	
Parameters	Standard deviations
Earth GM	0.004 km ³ /sec ²
4 × 4 gravity	50% GEM10 – GEM10B
Atmospheric drag	20% uncertainty in C_D
Solar pressure	10% uncertainty in reflectivity

Table 3. Perturbations due to dynamic mismodeling for reduced-dynamic single-orbit case

Spacecraft position components	GM, cm	4 × 4 gravity, cm	Solar pressure, cm	Atmospheric drag, cm
Radial	0.18	1.22	0.08	0.20
Cross-track	0.10	0.73	0.04	0.18
Down-track	0.05	0.78	0.10	0.72

Table 4. Perturbations due to dynamic mismodeling for reduced-dynamic two-orbit case

Spacecraft position components	GM, cm	4 × 4 gravity, cm	Solar pressure, cm	Atmospheric drag, cm
Radial	0.31	1.03	0.09	0.10
Cross-track	0.14	0.72	0.06	0.14
Down-track	0.14	0.60	0.08	0.29

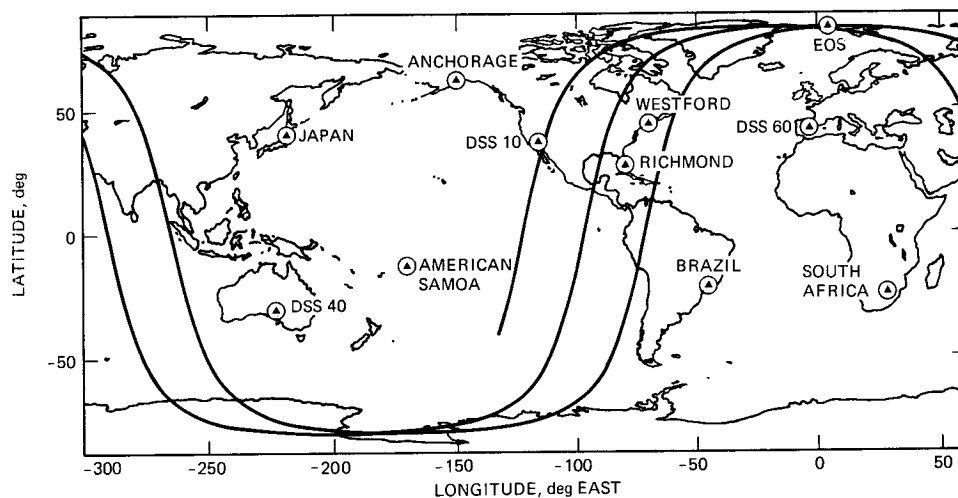


Fig. 1. Assumed worldwide network of GPS ground receiver locations and EOS orbit ground track

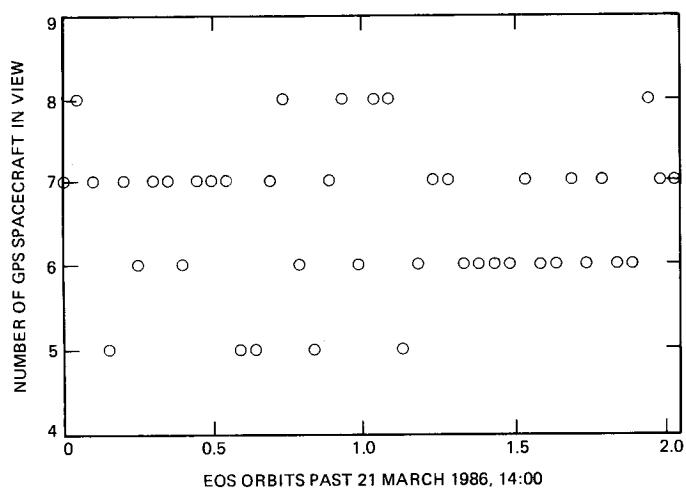


Fig. 2. Number of GPS spacecraft visible to EOS assuming a hemispherical field of view receiving antenna centered at zenith

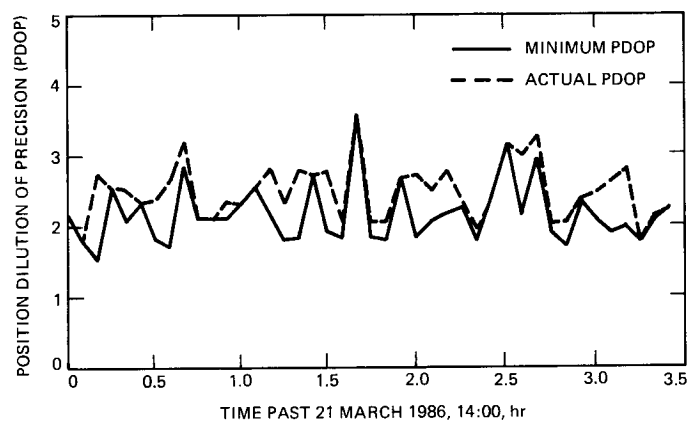


Fig. 3. Geometric strength of simulated EOS tracking; selection criteria maximizing length of track for a given GPS while minimizing PDOP

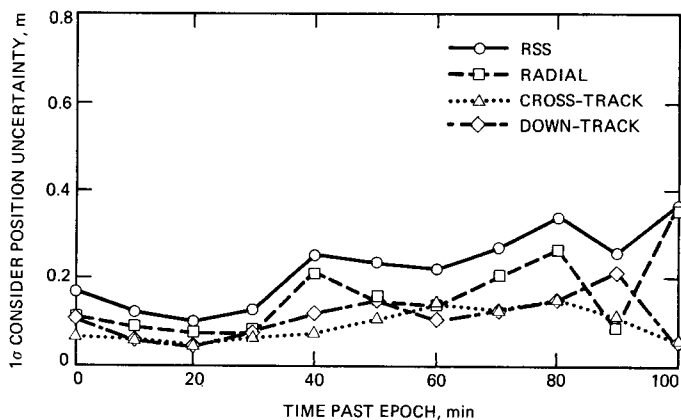


Fig. 4. EOS position uncertainty for non-dynamic tracking over one orbit

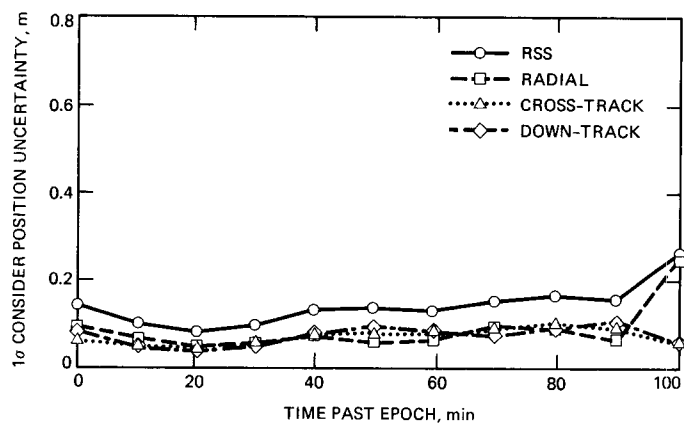


Fig. 5. EOS position uncertainty for reduced-dynamic tracking over one orbit

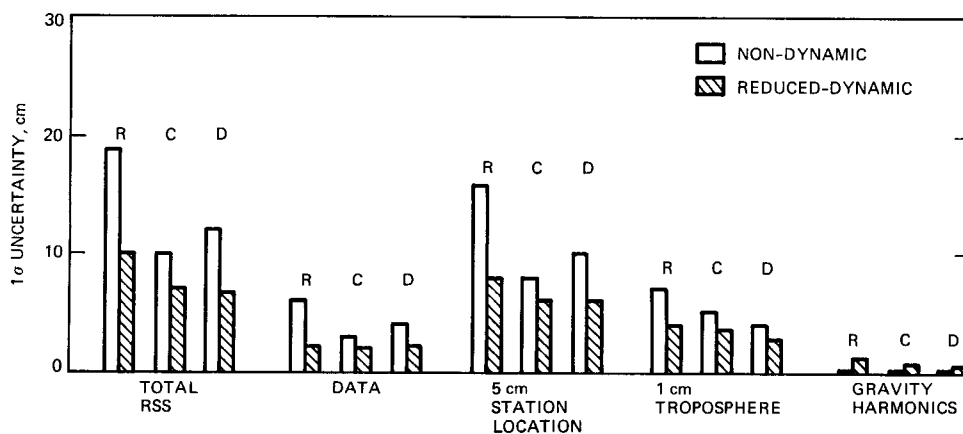


Fig. 6. EOS position error budget for both strategies; errors are RMS values over single-orbit cases of Figs. 4 and 5; R, C, D correspond to radial, cross-track, and down-track components

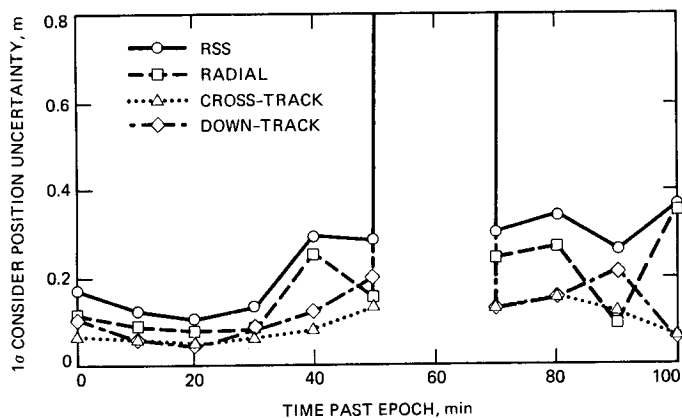


Fig. 7. EOS position uncertainty for non-dynamic tracking over one orbit; data deleted at 60 and 65 min to simulate bad GPS selection by EOS flight receiver

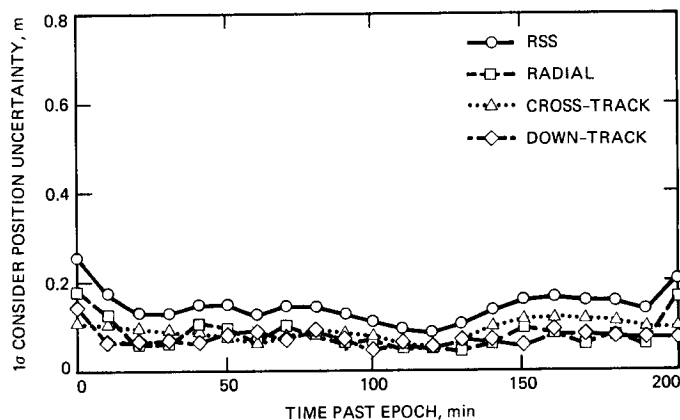


Fig. 9. EOS position uncertainty for reduced-dynamic tracking over two orbits

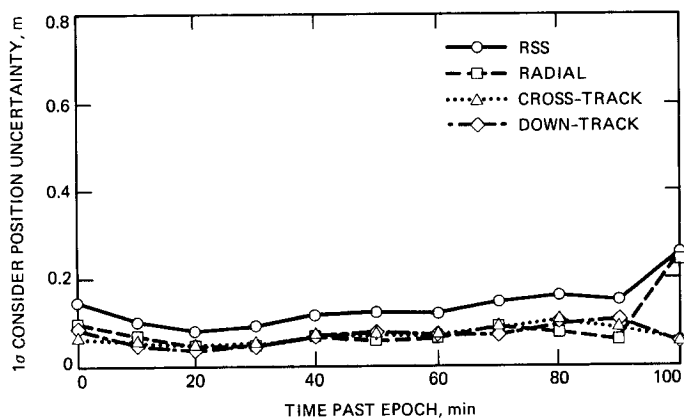


Fig. 8. EOS position uncertainty for reduced-dynamic tracking over one orbit; same measurement data as in Fig. 7

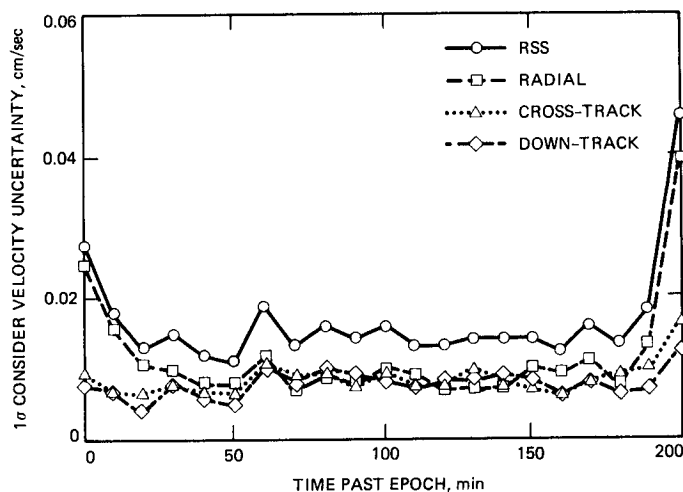


Fig. 10. EOS velocity uncertainty for reduced-dynamic tracking over two orbits

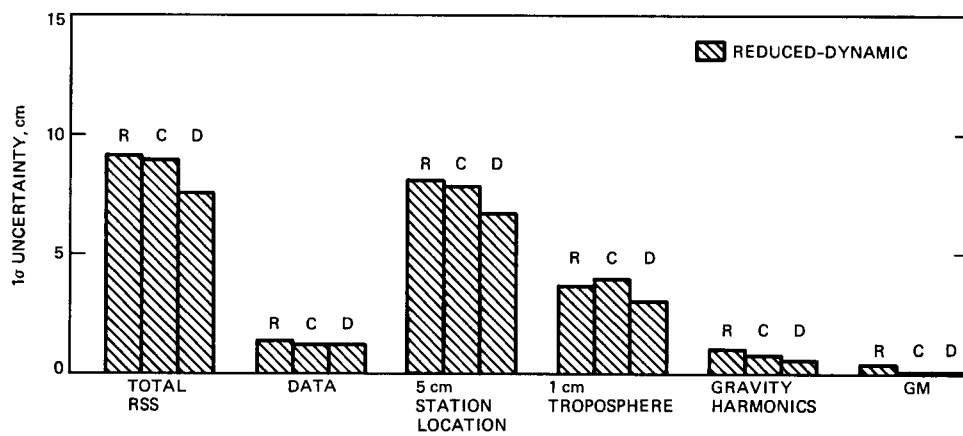


Fig. 11. EOS position error budget for reduced-dynamic tracking over two orbits; R, C, D correspond to radial, cross-track, and down-track components

53-13

617968

1989000818

N89 - 10189

14p.

TDA Progress Report 42-94

April-June 1988

JPV

PSA: A Program to Streamline Orbit Determination for Launch Support Operations

V. N. Legerton and N. A. Mottinger
Navigation Systems Section

An interactive, menu-driven computer program has been written to streamline the orbit determination process during the critical launch support phase of a mission. Residing on a virtual memory minicomputer, this program retains the quantities in-core needed to obtain a least squares estimate of the spacecraft trajectory with interactive displays to assist in rapid radio metric data evaluation. Menu-driven displays allow real-time filter and data strategy development. Graphical and tabular displays can be sent to a laser printer for analysis without exiting the program. Products generated by this program feed back to the main orbit determination program in order to further refine the estimate of the trajectory. The final estimate provides a spacecraft ephemeris which is transmitted to the mission control center and used for antenna pointing and frequency predict generation by the Deep Space Network. The development and implementation process of this program differed from that used for most other navigation software by allowing the users to check important operating features during development and have changes made as needed.

I. Introduction

Spacecraft radio metric tracking and telemetry support is provided by the Deep Space Network for a variety of international and domestic space agencies. One component of this tracking system, the High Earth Orbiter (HEO) Multimission Navigation Facility, uses the radio metric data to determine the spacecraft trajectory during the critical launch phase. For Earth-orbiting communications satellites destined for geostationary positioning or for out-bound interplanetary probes, an estimate of the trajectory is needed within hours after launch for planning maneuvers. These trajectory change maneuvers may be required to achieve geostationary status or to re-

fine the Earth departure trajectory. Another estimate of the spacecraft trajectory may be needed even sooner to update DSN Deep Space Station antenna pointing information in the event of non-nominal launch vehicle performance.

The functions of this Navigation Facility were described in a previous TDA Progress Report [1]. Significant in its design was the need to provide low-cost and efficient multimission navigation support using a dedicated minicomputer. The new computer program being described in this article has been designed by the HEO Navigation team and implemented by the Navigation Software Development Group to assist in the orbit determination process in this environment.

The fundamental computer program used for spacecraft navigation by the HEO team is the JPL Orbit Determination Program (ODP) [2]. This program is currently used in support of all JPL deep space missions, and it was implemented on a VAX 11/780 minicomputer for the HEO task. The ODP is not a single program but rather a set of software modules, each of which executes in a non-interactive mode and requires several mass storage files for inter-module communication. Experience with the ODP in the HEO launch environment revealed the need for a different operating mode in order to meet mission timeline requirements. This has been achieved by writing an easy-to-use, interactive, menu-driven program which combines five basic ODP functions of data selection, data editing, filter specification, parameter estimation, and residual display into a single program.

Processing radio metric data begins by running only the ODP modules needed to convert the data to a form which can be used to update the spacecraft trajectory. This new program is used to carefully evaluate those data, choosing which data types to use, deleting any bad data, and selecting the filter parameters necessary to properly estimate the spacecraft trajectory. When its task is completed, updated values of the spacecraft trajectory, information necessary to duplicate the set of radio metric data used, and filter specifications are passed back to the ODP for final refinement of the spacecraft trajectory. At this stage all the functions needed to process the data and update the trajectory estimate are performed by the ODP in the batch environment. Only solution summary information is monitored to determine when this iterative procedure has converged. When it has, the appropriate trajectory products are delivered to the mission control center for subsequent mission planning and to the DSN for updating Deep Space Station antenna pointing information.

The acronym PSA, chosen for this new program, is derived from the three basic functions the program performs: *P*acking information arrays of processed radio metric data, *S*olving for a set of parameters in a least squares sense, and *A*nalyzing the solution with graphic and tabular output. This article traces some of the significant steps in the program's evolution and shows some of the interactive features which have made it so useful. Future plans for its continued development and implementation will also be described.

This article assumes familiarity with the orbit determination process and some associated terms and quantities. It is not a tutorial introduction to the subject; nor is it a PSA user's guide. Illustrations of the menus and displays which PSA generates are included, and even though their contents cannot be described completely, it is hoped that the reader can appreciate how they help to streamline the orbit determination process in the launch support environment.

II. Program Development

In the traditional use and development of most navigation software, PSA represents a new approach, not because of new mathematical algorithms but because of the interactive capabilities it offers. The PSA mathematical requirements were easily satisfied with existing software libraries, but achieving easy-to-use human interfaces required a different approach to the implementation. In this section two significant steps leading to PSA will be described, followed by a general description of how it was implemented.

A. Virtual Memory Radio Metric Data Editing

The Flight Project Office-UNIVAC operating environment for the ODP consists of a large, fixed core, mainframe computer used primarily in a batch mode with no interactive capabilities. In preparation for this launch support activity with the DSN, the ODP was implemented on a VAX 11/780 minicomputer with virtual memory, but without any modification to its basic operating mode.

In the mid-1980s the potential use of virtual memory was first realized when a utility program to assist with radio metric data editing was written for the HEO Navigation team. This program retained pertinent information in virtual memory about the data processed by the ODP, sorting it by type, i.e., two-way Doppler, range, angles, etc., and using interactive menus to enable the user to study the different types. For any type selected, the display at the terminal screen showed basic information such as the time of each point, the Deep Space Station which received it, elevation angle, residual (difference between the value observed at the station and that computed by ODP), and a simple line printer plot of the residual. It was easy to see "bad" data from either the plot or the value of the residual. Commands similar to those associated with full-screen text editing made it possible to delete individual or sets of "bad" data. Inverse functions were also present to accept data previously deleted. Upon exiting the program, a status code for each radio metric data point was updated on a mass storage file to reflect the in-core status, and the updated file was passed back to the solution links of the ODP. An ASCII file was also written identifying the points which had been rejected so that a permanent record of "bad" data could be retained.

This program, called PTSCAN for *PoinT SCANing*, represented a significant improvement over other data editing techniques primarily because of the ease and speed it offered in displaying *any* of the radio metric data types being considered for use in determining the spacecraft trajectory. Other helpful attributes were the combination of numerical and "graphical" displays used to present the data. PTSCAN was used in support of several launches, but it did not provide all the capabilities needed, particularly that of being able to process the data

and determine an estimate of the spacecraft trajectory. It did suggest, however, testing the feasibility of retaining all the information needed in virtual memory.

B. Virtual Memory Least Squares Estimation

Following the favorable experience with PTSCAN, a program was written to test the feasibility of storing in virtual memory all the quantities needed to do the least squares solution. This represented an increase of more than an order of magnitude in memory storage requirements over PTSCAN.

This simple program, written by the HEO Navigation team, did not contain all the convenient user interfaces of PTSCAN or all that would ultimately be needed, but it definitely proved that the concept was a viable one. Trajectory estimates could be obtained very quickly without excessive demands on computer resources. An important part of determining the spacecraft trajectory involves verifying consistency among the solutions derived, using different combinations of the radio metric data available. Such changes could quickly and easily be made with this program and a new solution obtained without having to re-read the data from a mass storage file, as required when using existing ODP modules to do these types of studies. With all the data available in virtual memory, it is a simple matter to use only the types requested by the user. The speed with which the estimate was obtained was a real "eye opener" to the possibilities available. The thought of expanding the basic PTSCAN functions to include the estimation capability along with other necessary interactive features evolved into the current PSA. Continued use of the test program helped the Navigation team visualize the types of interactive menus and displays needed in PSA.

C. PSA User Interfaces: Design, Implementation, and Testing

The idea for PSA, an interactive program to enable the user to eliminate bad data, select different data types to use in a solution, tune the filter, generate graphic plots, and feed information back to the ODP, was becoming more clearly defined. The next step was to specify its design as completely as possible.

A variety of menus and displays were envisioned to assist in specifying general types of radio metric data to use in the filter, as well as the specific data points not to use, the parameters to estimate, the a priori uncertainty to use with each of them, and the display of the solution, to name a few. To clarify how the menus should appear, prototype text files were created for viewing at a computer terminal. It was easy to make changes until the files contained the desired information, displayed in a useful manner. After identifying all of them in this way, a detailed design document was written

describing how the menus would be activated in PSA and how each one would work.

It was possible that the menus might operate unacceptably and have to be changed even if they were implemented exactly as requested. To check for this, the first step in PSA implementation was to provide a version of the program in which only the menus operated. Since the files which would normally supply the menu contents were not read at this stage of the development, simulated numerical quantities were provided. The Navigation team ran this version of PSA to test how easy it was to access different menus and in turn to determine how easy they were to use. A few changes were made and then the implementation of the rest of the program began.

This development process represents an approach to ODP software implementation which is different from the customary one. That approach involves specifying the mathematical formulation, user inputs, and required output. The program is developed, checked by the programmer, and then delivered to the engineers working on a mission project. At this time, it may be impossible to change operating features which are different from those anticipated, and the users have no choice but to wait a long time (until the next mission) before it is possible to consider modifying the program. Even then it may be too late to make fundamental changes without major redesign and implementation. If the traditional process repeats, it still may be difficult to obtain the type of program desired.

D. PSA Computations: Implementation and Testing

Once the user interfaces were working well, implementation of the mathematical portions of PSA began. Much of this existed in other software, as would be expected, since PSA was combining functions already in existence and could be copied. For example, the filter had been implemented in the test program described in Section IIB and would serve as a model for PSA. It was important to ensure that the filter was working correctly, primarily verifying that only the intended data were actually used. Numerous cases were run in PSA varying the radio metric data set followed by duplication with the ODP. The flexibility offered in PSA for choosing different sets of data was especially apparent when compared with the time required to make the corresponding ODP run.

Graphical products generated by PSA were the same as in other ODP programs, so comparison could be accomplished merely by overlaying two plots and checking for any differences.

PSA debugging information was created to verify that quantities had been read correctly from the input files or to exam-

ine quantities computed by PSA to more digits than normally displayed. These data are written, when requested by the user, as ASCII files for study at the conclusion of a PSA execution.

The ultimate test of all PSA functions was the ability to duplicate a solution using the ODP. Certain products generated by PSA were best tested through use with the ODP to verify that the same number of data points, same solution, and same graphical output were generated. A binary file written by both PSA and the ODP could be checked using existing comparison software.

From the HEO Navigation team's experience with previous missions, it was initially decided to configure PSA to store information for a maximum of 5000 radio metric data points, with the capability to estimate a maximum of 25 parameters. Subsequent use of the program in mission operations indicated the need to increase the number of points to 20,000, with a maximum estimation capability of 30 parameters. No degradation in running time was observed after increasing the storage requirements by this amount.

III. Program Usage

This section will give a brief overview of the functions PSA can perform. The reader who has not experienced the real-time operations may not appreciate fully the conveniences afforded by PSA but will have an increased awareness of how PSA helps.

There are three basic functions performed in PSA:

- (1) Data EDITing
- (2) Data FILTERing
- (3) Data DISPLAY

Within these three functions are twelve subfunctions or options which the user may choose, in any order, to do the task. The first display to greet the PSA user shows these functions and options, and is reproduced in Fig. 1. The subfunctions or options are all preceded by a letter in "()"'s. To select the desired function, simply enter its code letter. This in turn will replace the screen with the menu or display associated with that option. Upon completion of any subfunction, a concise summary of the options available is displayed across the bottom of the terminal screen. This is shown in Fig. 2. This concise display contains the same information as the entry-level display but can be portrayed quickly on the screen to serve as a reminder of the options available. The help option, "h," will re-create the entry-level display any time the user needs it. Note that *no* carriage return is required when the options are entered.

The subfunctions are briefly described and illustrated in the following sections.

A. PSA EDITing Options

The EDITing options in PSA enable the user to control the radio metric data used in the solution, to decide how it should be weighted, and to generate text files which will pass this information back to the ODP.

1. Data selection. There are three options available to control the radio metric data used to estimate the spacecraft trajectory.

One high-level capability permits general radio metric data selection by type (two-way Doppler, range, angles, etc.) and by Deep Space Station (DSS) which received the data. This is referred to as Global Data selection and is controlled using an interactive menu as illustrated in Fig. 3. With this menu the user can easily choose to include (accept) or omit (reject) any of the radio metric data type/station combinations shown. With this menu, data consistency checks can easily be made by comparing the estimates of the spacecraft trajectory determined using different data types or data from different Deep Space Tracking stations.

The menu in Fig. 3 shows a configuration in which the radio metric data type F3 (three-way Doppler) from DSS 46 is rejected, as is the F2 (two-way Doppler) from DSS 24 at Guam. All of the other radio metric data will be used to estimate the spacecraft trajectory when a solution is requested.

A second high-level EDITing capability can be selected from the Global Data Selection Menu and is shown in Fig. 4. This menu is used to specify whether *all* radio metric data should be deleted (rejected) as a function of elevation angle or as a function of time, e.g., before, after, or between specified times. Another function deletes specific radio metric data points if the residual, the difference between the value received at the Deep Space Station and that computed by the ODP, exceeds a specified limit. The user may modify, remove, or add new delete commands to this menu at any time. The example in Fig. 4 illustrates the deletion (rejection) of any radio metric data recorded at less than 5 degrees of elevation.

A third EDITing option permits a point-by-point examination of the radio metric data in search of "bad" points. This is the "p" or PTSCAN option. Named for the PTSCAN program discussed in Section IIA because this function was copied from it, this option first produces a menu like that shown in Fig. 5. From this menu the user selects a radio metric data type for study, as shown in Fig. 6, where a portion of the interactive PTSCAN display for a range data type is shown. An individual data point Accepted for use in a solution is plotted with an

"A," while data Rejected are plotted with an "R" and flagged with an "*" at the left, as can be seen in Fig. 5. Using terminal keypad commands similar to those employed in full-screen text editing, the user may indicate to reject or accept individual or groups of radio metric data points. Other keypad commands automatically seek out "bad" data points which the user can then delete. Three methods for controlling the plot scale also exist. One technique uses a percentile method to minimize effects of extremely large "bad" residuals, while another uses the statistical mean and 3σ standard deviation to set the plot scale. The third allows the user to set whatever value is desired.

2. Data weighting. Another PSA EDITing function is the Weight or "w" option. The menu associated with this option enables the user to modify weights assigned to the different radio metric data types. Figure 7 shows an example of the Global Weighting Menu. The names of the radio metric data types are shown along with the default weight values which were read from a file also used by the ODP. Metric equivalent values are also entered computed from user-supplied conversion factors. Note that two "*"s replace a numerical quantity for the F2 data. This is because another Weight Menu has been used to specify weights as a function of the Deep Space Station which received the data. These are referred to as Menu Weights and are specified using the Menu Data Weighting menu shown in Fig. 8. Note that two-way Doppler radio metric data from DSS 24 are weighted one-tenth that from DSS 46, reflecting the higher noise level in the data received from this smaller-diameter antenna. Even though data from this station are currently rejected in the Global Data selection menu, Fig. 3, a weight is retained should these data ever be used in a solution.

3. Text files for the ODP. The fourth and final PSA EDITing function is the "c" option for writing ASCII files needed by the ODP to reproduce the set of accepted and deleted data defined in PSA. The information written in these files is based upon the contents of the Global Data Selection Menu, Fig. 3; the Menu Data Selection Menu, Fig. 4; and individual points deleted in the PTSCAN mode, Fig. 6. The ASCII file will concisely enable subsequent ODP runs to delete exactly the same set of data points as was deleted in PSA. In subsequent PSA runs, the ODP binary file created from this ASCII file will be read and used to set these menus in the same configuration and delete any individual points, again reproducing the same set of rejected data.

This is a vital function in PSA because as more and more radio metric data are received and processed by the ODP in the course of a mission, any data previously deleted with PSA should continue to be deleted. The user need be concerned only with new data when reentering PSA. It can be studied in

search of any "bad" points and rejected, and a new solution obtained quickly. It also permits flexibility in changing data rejection criteria. For example, it is easy to change the minimum elevation angle for data rejection, feed the information back to the ODP, and then have it automatically available in future PSA runs.

Other ASCII files written under the "c" option enable the information in the Weight menus, Figs. 7 and 8, to be transmitted back to the ODP just as the deleted data points were. This information is also retrieved by PSA in subsequent runs to maintain the current weighting values.

B. PSA FILTER Options

The FILTER function in PSA consists of five subfunctions or options which are used both to establish the correct estimation model depending on the types of radio metric data considered and to obtain estimates of the spacecraft trajectory. Other functions create files for passing the new estimate and associated filter model back to the ODP. This section will describe the menus and displays associated with this aspect of the program.

1. Estimate and a priori σ . The user may need to vary the set of parameters estimated in conjunction with the types of radio metric data used in the solution. For example, angle biases must be estimated when angular observables are in the solution, or a bias may be estimated in the two-way Doppler radio metric data for a spinning spacecraft with a circularly polarized antenna. A sample of the menu used to control which parameters are estimated is shown in Fig. 9. This menu can also be used to modify their a priori uncertainty. The menu shown in this figure is configured to estimate the six components of the spacecraft state, a two-way Doppler bias, and biases in the angle measurements received at DSS 24 and 46.

The nominal configuration of this menu is based upon a file maintained by the ODP. If changes are made in PSA, these can be communicated back to this ODP file so that future ODP and PSA runs will retain the new configuration. This is described in Section IIIB4, PSA FILTER: Solution Output to ODP.

2. Solution generation. The "s" option creates a least squares estimate for the parameters identified in the Estimate menu. Householder transformations are first used in a square root formulation to pack an information array of *all* regression quantities (partial derivatives) stored in the virtual memory PSA arrays. The estimate for the specified set of parameters is then obtained by extracting only those columns from this information array, repacking, and adding the a priori information. This implementation enables the user to vary the

set of estimated parameters or their a priori uncertainty without having to repack all the data, until a different data set is specified.

The radio metric data used in the solution is based upon the current information in the Global and Menu Data Selection menus and the point-by-point reject status set in the PTSCAN mode. A printed solution summary, as shown in Fig. 10(a), appears at the terminal screen each time the "s" option is entered. This summary contains a number of quantities needed to analyze the current solution and indicate any mistuning of the filter. Columnar data identify the parameters estimated, the least squares Correction, the computed σ (uncertainty), and the New value of the estimated parameter. Other information shows a summary of the data used to obtain the solution. The solution output concludes, as shown in Fig. 10(b), with a comparison of the classical orbital elements computed from the nominal trajectory of the spacecraft and the New value just estimated. Using these coordinates provides a valuable alternative assessment of the solution. All of this print can be sent to a laser printer, as will be described in Section IIIC, which discusses PSA DISPLAY functions.

3. Bias partials. The "b" option under the PSA FILTER Function enables the user to add Bias partials of a simple form to the regression data stored in virtual memory. The menu shown in Fig. 11 is used to make this addition. The user can request new partials, remove old ones, or modify some portion of an existing one. This capability enables the user to study the effect of possible biases in the radio metric data without having to exit PSA and compute them using other software. Doppler biases are often encountered in the missions which the HEO Navigation team supports, since many spacecraft are spinning with circularly polarized antennas. Modification to the spin rate results in a different Doppler bias. The bias option capability in PSA makes it easy to account for such changes.

4. Solution output to ODP. The FILTER function "g" writes binary and ASCII files for use in the ODP. The binary file contains the New values of the estimated parameters in a form which can be assimilated by the ODP to begin the next processing of the radio metric data. The ASCII file contains data which can be input to the ODP to duplicate the filter model set up in PSA.

5. Output regression file. The FILTER function "u" will create a new mass storage file duplicating the regression information contained in the PSA virtual memory arrays. This function was designed for the case in which this might be the easiest way to communicate the set of selected data back to the ODP for special processing, or continue with new bias partials.

C. PSA DISPLAY Options

There are four options which generate output to assist in assessing the solution. Some of them produce tabular output, two of them produce graphical output, and two of them send products to a laser printer. Any output directed to the laser printer can be received, without terminating PSA, and used for continued analysis.

1. Residual statistics. A summary showing the bias and 1σ standard deviation of the radio metric data residuals is controlled by the "r" option. A sample of the information is displayed in Fig. 12. The statistics shown here are computed from two sets of radio metric data residuals, those read from the input ODP file and the linearly predicted ones computed by PSA for the current solution. Optional conversion factors can be input to PSA to convert the statistics to metric units, as shown in Fig. 12. Also shown are the weights for each data type and the number of accepted and deleted points.

2. Residual statistics at laser printer. With the "i" option, tabular solution and data statistics are sent to a laser printer. The information is a combination of the output shown in Figs. 10 and 12, which is displayed at the terminal for the "s" and "r" options.

3. Residual statistics and plots at laser printer. With the "l" option it is possible to augment the information printed by the "i" option with graphic plots of the radio metric data residuals. Figure 13 shows representative graphical output generated by this option. Plot scaling is computed by PSA, using the mean and 3σ data statistics available for each radio metric data type. This entire summary output can be conveniently stored as a record of the analysis, providing both numerical and graphical information. This graphic output is available only at the laser printer, and not at the terminal.

4. PTSCAN mode. The PTSCAN Edit mode display can be accessed again to review radio metric data residuals. After a solution is requested, linearly predicted residuals are computed and available to the PTSCAN mode. This creates the same menus as shown in Figs. 5 and 6. The user may choose between these two sets of residuals, and data points may be deleted or accepted based upon examining either set. Sometimes additional "bad" data cannot be detected until a solution has been obtained and the linearly predicted residual computed. Such points are easy to detect by using the PTSCAN mode again; they can be deleted and a new solution obtained.

IV. Program Testing

A prototype version of PSA was available by the summer of 1987. It was tested and evaluated during prelaunch training

exercises and used for the launch of the Japanese NASDA ETS-5 mission. It proved a resounding success, interfacing with the ODP to provide the efficient means needed for quickly analyzing the radio metric data. It was possible to deliver an updated spacecraft trajectory for DSN support based upon an hour of radio metric data, within an hour after receipt of the data. At five hours into the mission, updated trajectory information was again available, this time for NASDA mission planning purposes. Since that time, PSA has been evaluated in support of another NASDA mission, CS-3a, and two European flights, the German TV-SAT and the French Telecom-1c. The current PSA prototype will be transferred to operations for support of numerous future missions.

In addition to direct navigation support, PSA was useful during the Telecom-1c mission for evaluating range data received from two different Deep Space Stations. The project reported receiving bad real-time range data, but these were quickly certified, utilizing PSA, as being valid data. PSA is also proving very useful for non-real-time parametric studies using radio metric data from deep space missions supported by the DSN.

V. Future Improvements

Following the evaluations of the prototype version of PSA, there are several new capabilities needed to enhance its usage. One is to provide graphic residual displays at the terminal. Creating plots like those shown in Fig. 13 at the terminal would make PSA operations much more efficient by eliminating the time required to generate copies on the laser printer. Often these plots are needed only for intermediate analysis

and are then discarded. Usually it is only the plots from the final solution which need to be sent to the laser printer to create copies for record-keeping purposes. This capability could reduce use of the laser printer by 50 to 75 percent, making more of that resource available to other members of the Navigation team.

Another useful graphic display is one which includes interactive data editing. In this mode, residual plots of only one or two data types would be displayed at the screen. The user would move a cursor to identify individual or sets of data to delete (or accept). This is similar to the data editing done in the PTSCAN mode in PSA, but would represent an improvement with the linear time scale and the ability to simultaneously view all the residuals for a particular data type. This technique exists in another program available to the HEO Navigation team and is known to be useful.

Another useful capability would be to simultaneously view various summaries such as the solution, Fig. 10, or statistics, Fig. 12, and residual plots, Fig. 13. Being able to correlate summaries and graphical output would improve PSA's effectiveness and efficiency.

Meeting these needs in PSA cannot be achieved with the terminals currently being used by the HEO Navigation team. Bigger screens with windowing capability and much faster graphic generation are required to make these features effective tools. Simultaneous display of tabular and graphical output also requires a large screen. The capabilities offered in a workstation-type environment are currently being evaluated to determine if they meet these needs.

Acknowledgments

The authors would like to thank the members of the HEO Navigation team, Jordan Ellis, Len Efron, Premkumar Menon, and Tim McElrath, for their active role in specifying the PSA operation mode. Special mention is also due Jim Collier for making the PTSCAN code available and other members of the Navigation Software Development Group and Computer Science and Applications Section who designed and implemented various subroutine libraries used in PSA. Peter Breckheimer is also recognized for making other resources available that were necessary to complete this project.

References

- [1] J. Ellis, "HEO Multimission Navigation Concept," *TDA Progress Report 42-86*, vol. April-June 1986, Jet Propulsion Laboratory, Pasadena, California, pp. 261-267, August 15, 1986.
- [2] T. D. Moyer, *Mathematical Formulation of the Double-Precision Orbit Determination Program (DPODP)*, Technical Report 32-1527, Jet Propulsion Laboratory, Pasadena, California, May 15, 1971.


```

                                PSA
EDIT:
(d) Data Selection
(p) PTSCAN
(w) Weighting
(c) Write CSP Text Files
FILTER:
(f) EST List, Apriori
(s) Solve
(b) Add Bias Partial
(g) Write Solution & Filter Text Files
(u) Update Regres File
DISPLAY:
(r) Residual Statistics
(l) Solution & Plots to Printer
(i) Solution to Printer
(p) PTSCAN Residuals (Before & After)
UTILITY:
(x) Exit
(h) Help

```

Fig. 1. Entry-level display

```

(d)Data   (w)Weights (f)Est list (b)Bias   (u)Regres (l)Printer (x)Exit
(p)PTSCAN (c)CSP    (s)Solve   (g)File 32 (r)Stats  (i)Prt Sol (h)Help

```

Fig. 2. Concise function display

Global Data Selection Menu							
#	Data	Rcvr	Band	First Point		Last Point	
182	X85	46	/	12-MAR-1988	00:33:30	12-MAR-1988	03:40:50
182	Y85	46	/	12-MAR-1988	00:33:30	12-MAR-1988	03:40:50
24	GSRAN	46	S/S	12-MAR-1988	00:35:20	12-MAR-1988	01:13:40
161	F2	46	S/S	12-MAR-1988	00:40:45	12-MAR-1988	01:35:45
8	X30	24	/	12-MAR-1988	01:17:30	12-MAR-1988	03:40:00
8	Y30	24	/	12-MAR-1988	01:17:30	12-MAR-1988	03:40:00
38	GSRAN	24	S/S	12-MAR-1988	01:17:50	12-MAR-1988	03:40:20
40 r	F2	24	S/S	12-MAR-1988	01:17:55	12-MAR-1988	03:40:35
42 r	F3	46	S/S	12-MAR-1988	01:17:55	12-MAR-1988	03:40:45
Functions: (M)enu, (P)age, (S)ummary, (E)xit							
Editing: (A)ccept, (C)lear all rejects, (R)eject							

Fig. 3. Global data selection menu

Menu Data Deletion Menu		
DELETE	For	Scope Limiter
ALL	ELVMIN	5.
a(R)esid, (E)lvmin, (B)efore, (A)fter, (F)rom-to		
Functions: (G)lobal, (P)age, (E)xit, (V)iolation clear		
Editing: (A)dd, (C)hange, (D)elete, (R)efresh		

Fig. 4. Menu data deletion menu

PTSCAN

X85
Y85
GSRAN / S
F2 / S
X30
Y30
F3 / S

Functions: (L)ook, (E)xit

Fig. 5. PTSCAN option menu

CASE 0		62 TOTAL	MEAN	--1.221E-03	SIGMA = 9.699E+00		
1988		59 ACCEPTED	CENTER	--1.995E+00	SCALE = 2.776E+01	%til	
REC	MMDD	HMMSS	TM	R1	ELM	RESID	GSRAN / S
* 1	0312	003520	46	46	11.5	-3.93E+02 R<
* 2	0312	003550	46	46	11.8	-3.80E+02 R<
* 3	0312	003740	46	46	13.0	-3.57E+02 R<
4	0312	011020	46	46	24.8	1.27E+01 A .
5	0312	011030	46	46	24.8	7.72E+00 .	. A . . .
6	0312	011040	46	46	24.9	-7.86E+00 A .
7	0312	011050	46	46	24.9	-8.85E+00 A .
8	0312	011100	46	46	24.9	-1.50E+01 .	. A . . .
9	0312	011110	46	46	25.0	-1.01E+01 A .
10	0312	011120	46	46	25.0	4.20E+00 A .
11	0312	011130	46	46	25.0	-2.00E+00 A .
12	0312	011140	46	46	25.0	-1.14E+01 A .
13	0312	011150	46	46	25.1	-3.81E+00 A .
14	0312	011200	46	46	25.1	-1.20E+01 A .
15	0312	011210	46	46	25.1	-3.69E+00 A .
16	0312	011220	46	46	25.2	7.26E+00 A .
17	0312	011230	46	46	25.2	7.09E+00 A .
18	0312	011240	46	46	25.2	3.01E+00 A .
19	0312	011250	46	46	25.3	4.23E+00 A .
20	0312	011300	46	46	25.3	-3.05E+00 A .
21	0312	011310	46	46	25.3	-4.62E+00 A .
22	0312	011320	46	46	25.3	1.70E+00 A .
23	0312	011330	46	46	25.4	1.61E+01 A .
24	0312	011340	46	46	25.4	1.88E+01 A .
25	0312	011750	24	24	34.5	1.71E+01 A .
26	0312	011800	24	24	34.5	1.16E+01 A .
27	0312	011810	24	24	34.6	1.99E+01 A .
28	0312	011820	24	24	34.6	5.13E+00 A .
29	0312	011830	24	24	34.6	5.19E-01 A .
30	0312	011840	24	24	34.7	1.32E+01 A .
31	0312	011850	24	24	34.7	-3.57E+00 A .
32	0312	011900	24	24	34.7	-1.70E+00 A .
33	0312	011910	24	24	34.8	-1.60E+01 A .
34	0312	011920	24	24	34.8	-4.30E+00 A .
35	0312	011930	24	24	34.8	1.56E+01 A .
36	0312	011940	24	24	34.9	7.30E-01 A .
37	0312	011950	24	24	34.9	-9.61E+00 A .
38	0312	012000	24	24	34.9	-1.13E+01 A .
39	0312	012010	24	24	35.0	-1.22E+01 A .
40	0312	012020	24	24	35.0	-3.21E+00 A .

Fig. 6. PTSCAN mode radio metric data editing

Global Data Weighting Menu			
Data	Band	Weight	Metric
X85	/	0.20000000E-01	0.34906580E+03
Y85	/	0.20000000E-01	0.34906580E+03
GSRAN	S/S	0.66670000E+02	0.10000433E+02
F2	S/S	**	
X30	/	0.20000000E-01	
Y30	/	0.20000000E-01	
F3	S/S	0.15000000E+00	0.98039999E+01
Functions: (M)enu, (P)age, (E)xit			
Editing: (C)hange, (D)elete menu, (S)et menu			

Fig. 7. Global data weighting menu

Menu Data Weighting					
Data	Band	T	R1	Weight	Metric
F2	S/S	46	46	0.15300000E-01	0.10000080E+01
F2	S/S	24	24	0.15300000E+00	0.10000080E+02
Functions: (G)lobal, (E)xit					
Editing: (C)hange					

Fig. 8. Menu data weighting menu

Epoch	12-MAR-1988 00:30:00.0000			ET	CENT - 3
Estimate & Apriori Menu					
Name	Apr	Sigma	Est	Nominal Value	
X	1.00D+05		X	-7.277844761887D+02	
Y	1.00D+05		Y	-1.624675317719D+04	
Z	1.00D+05		Z	-1.843971504740D+03	
DX	1.00D+02		DX	3.932390300891D+00	
DY	1.00D+02		DY	-4.098763063321D+00	
DZ	1.00D+02		DZ	-2.595886602461D-01	
BF2	1.00D+02		BF2	0.000000000000D+00	
BX16	5.00D-02			0.000000000000D+00	
BY16	5.00D-02			0.000000000000D+00	
BX17	5.00D-02			0.000000000000D+00	
BY17	5.00D-02			0.000000000000D+00	
BX66	5.00D-02			0.000000000000D+00	
BY66	5.00D-02			0.000000000000D+00	
BGS17	6.67D+03			0.000000000000D+00	
BGS16	6.67D+03			0.000000000000D+00	
BGS66	6.67D+03			0.000000000000D+00	
BPL61	7.04D+03			0.000000000000D+00	
BX46	5.00D-02		BX46	0.000000000000D+00	
BY46	5.00D-02		BY46	0.000000000000D+00	
BX24	5.00D-02		BX24	0.000000000000D+00	
BY24	5.00D-02		BY24	0.000000000000D+00	
Functions: (P)age, (E)xit					
Editing: (A)dd, (C)hange, (D)elete					

Fig. 9. Estimate list menu

ORIGINAL PAGE IS
OF POOR QUALITY

(a)

Enter PSA

29-MAR-1988 09:59:16.23 , linked 19-FEB-1988 10:30:46

TC1C.NAV.TC0601

SOLUTION	EPOCH	12-MAR-1988 00:30:00.0000	ET	CENT	-	3
S = 1.15022E+03	Q = 1.15022E+03	SOLNO = 1				
SL = 6.66446E+01	QL = 6.72302E+01	MDATA = 455				
RSS of correction	Position = 6.53446E-03	Velocity = 2.33642E-06				
RSS of sigma	Position = 1.36410E+00	Velocity = 4.84490E-04				

Name	Correction	Sigma	Sigma/Apriori	New Value	DX(Apriori)
X	-6.01130E-03	1.201E+00	1.201E-05	-7.277904874930D+02	-5.46110E+00
Y	2.52602E-03	5.104E-01	5.104E-06	-1.624675065117D+04	3.32638E+00
Z	4.27367E-04	3.962E-01	3.962E-06	-1.843971077373D+03	-8.70912E+00
DX	-1.28324E-06	2.584E-04	2.584E-06	3.932389017653D+00	-1.20960E-03
DY	-1.94028E-06	3.909E-04	3.909E-06	-4.098765003596D+00	1.49352E-03
DZ	2.17910E-07	1.232E-04	1.232E-06	-2.595884423359D-01	1.78506E-03
BF2	1.37433E-01	3.729E-01	3.729E-03	1.374333980741D-01	1.37433E-01
BX46	3.51083E-02	2.682E-03	5.364E-02	3.510830817475D-02	3.51083E-02
BY46	-7.69038E-03	4.429E-03	8.859E-02	-7.690375578295D-03	-7.69038E-03
BX24	-1.14663E-02	8.240E-03	1.648E-01	-1.146633349429D-02	-1.14663E-02
BY24	6.39953E-03	7.291E-03	1.458E-01	6.399528197851D-03	6.39953E-03

DATA IN SOLUTION -0-1-2-3-4-5-6-7-

#	Data	B	Rcvr	Span	First Point	Last Point
162	X85		46	02:56	12-MAR-88 00:44:30	12-MAR-88 03:40:50
162	Y85		46	02:56	12-MAR-88 00:44:30	12-MAR-88 03:40:50
45	F2	S	46	00:25	12-MAR-88 00:48:55	12-MAR-88 01:14:45
21	GSRAN	S	46	00:03	12-MAR-88 01:10:20	12-MAR-88 01:13:40
8	X30		24	02:22	12-MAR-88 01:17:30	12-MAR-88 03:40:00
8	Y30		24	02:22	12-MAR-88 01:17:30	12-MAR-88 03:40:00
38	GSRAN	S	24	02:22	12-MAR-88 01:17:50	12-MAR-88 03:40:20

(b)

Classical Orbital Element Summary					
Epoch	12-MAR-1988 00:30:00.0000	ET	Center	3	
GM	0.3986004480734463D+06		Radius		6378.140
Name	Units	New - Nominal	New Value		Nominal Value
a (km)		-0.002	24340.321		24340.323
e		0.00000	0.72954		0.72954
i (deg)		0.00000	6.95346		6.95346
APF (deg)		0.00003	176.81228		176.81226
Node (deg)		-0.00005	335.82047		335.82052
TFP (sec)		-0.001	2726.382		2726.382
PER	-	00:00:00.00	10:29:52.04		10:29:52.05
MA (deg)		0.000	25.971		25.971
TA (deg)		0.000	114.658		114.658
ALT a (km)		-0.002	35719.460		35719.462
Time a (sec)	-	00:00:00.00	12-MAR-1988		12-MAR-1988
			04:59:29.64		04:59:29.64
ALT p (km)		-0.001	204.903		204.904
Time p (sec)	-	00:00:00.00	12-MAR-1988		12-MAR-1988
			10:14:25.66		10:14:25.67

Fig. 10. Solution page

Bias Partial Addition Menu						
New	Data	Band	Station		End Time	
Partial	Type	U	D	T	R	
Editing: (A)dd, (C)hange, (D)elete, (E)xit Change: Name(CN), Data(CD), Bands(CB), Stations(CS), Time(CT)						

Fig. 11. Bias partial addition menu

DATA STATISTICS												F2 & F3: MM/S		GSRAN: M		X85 & Y85: MICRORAD	
Data	Bias				Sigma		Weight	Number of points		C							
Type	B	DSN	metric		DSN	metric		DSN	Acc		Rej						
X85		3.51E-02 4.87E-05	6.13E+02 8.51E-01	3.46E-03 3.46E-03	6.04E+01 6.04E+01		2.00E-02	162	20	P 1							
Y85		-7.69E-03 2.04E-05	-1.34E+02 3.57E-01	3.80E-03 3.80E-03	6.64E+01 6.64E+01		2.00E-02	162	20	P 1							
GSRAN	S	-1.22E-03 4.11E-03	-1.83E-04 6.16E-04	9.70E+00 9.70E+00	1.45E+00 1.46E+00		6.67E+01	59	3	P 1							
F2/10	S	1.39E-01 -2.76E-06	9.07E+00 -1.81E-04	4.25E-02 4.25E-02	2.78E+00 2.78E+00			45	156	P 1							
X30		-1.17E-02 -2.36E-04		4.83E-03 4.83E-03			2.00E-02	8	0	P 1							
Y30		6.54E-03 1.30E-04		2.99E-03 2.99E-03			2.00E-02	8	0	P 1							
F3/10	S	9.76E-01 9.76E-01	6.38E+01 6.38E+01	1.29E+00 1.29E+00	8.41E+01 8.42E+01		1.50E-01 (42)	0	P 1							
Menu Data Weighting																	
DATA	Band	T	Rl	Weight													
F2	S	46 24	46 24	1.53E-02 1.53E-01													

Fig. 12. Data statistics display

ORIGINAL PAGE IS
OF POOR QUALITY

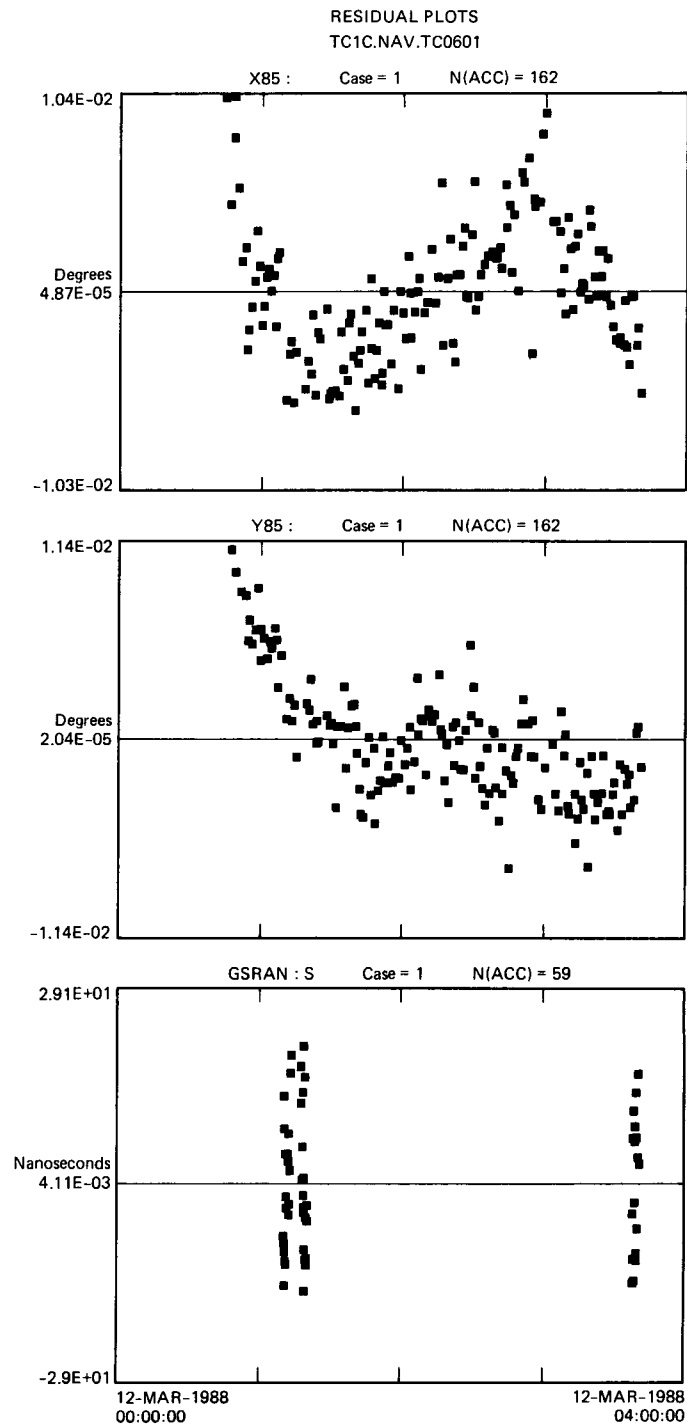


Fig. 13. Radio metric data residual plots

94-46

1989000819

617971

N89-10190

TDA Progress Report 42-94

April-June 1988

18p.

JPL

Spain 31-GHz Observations of Sky Brightness Temperatures

B. L. Gary

Microwave Observational Systems Section

A water vapor radiometer was deployed at DSS 63 for 3 months of sky brightness temperature measurements at 31 GHz. An exceedance plot has been derived from this data showing the fraction of time that 31-GHz 30-degree elevation angle brightness temperature exceeds specified values. The 5-percent exceedance statistic occurs at 75 K, compared with 70 K in Australia.

I. Introduction

The Deep Space Network is evaluating the merits of using 31 GHz for future spacecraft communications. One part of the Advanced Systems Program is to conduct observations at DSN sites to derive exceedance plots of 31-GHz sky brightness temperature. JPL's Ground-Based Microwave Applications Group was asked to provide these statistics for Spain using the group's SCAMS Water Vapor Radiometer (WVR). In [1], a background is given for acquiring Ka-band weather statistics; Australian results are given in [2].

The SCAMS WVR was deployed in Spain for observations during March 23 to June 28, 1984. The first 10 days of data were taken at the radiosonde launching site at Barajas, seven miles east of Madrid. The next three months were spent at DSS 63, 29 miles west of Madrid.

The Barajas data were intended for use in extrapolating results for the three months at DSS 63 to radiosonde archive-length time bases. (This part of the analysis has not yet been performed.) Another objective was to determine if there was anything unusual about Spain meteorology that would affect

the WVR's ability to convert sky brightness temperatures to vapor content and path delay. No differences were found.

II. Observing Setup

Figure 1 is a photograph of the SCAMS WVR. Figure 2 is a photograph of the WVR as it was deployed at DSS 63. The sensor box is on the roof and is tilted (22 degrees) in order to reduce the amount of water that is on the radome cover during (and after) rain. (DSS-63 personnel installed a blower to further assist in reducing "water on radome" problems, but it was rarely used.)

Figure 3 shows what SCAMS prints out in real time. The WVR was programmed to take 100-second tip curve measurements beginning every 10 minutes. Two printout modes are selectable; one shows a sky map of differences from a tip curve fit (lower part), and the other shows estimated brightness temperatures, vapor and liquid, using instrument physical temperature for deriving radiometer gain (upper part). Columns T1 and T2 show the zenith brightness temperature for 22.2 and 31.4 GHz. These are based on real-time tip curve reductions and are very reliable for clear sky conditions.

III. "Tip Curve" Reduction Method

Figure 4(a) is a plot of 22.2- and 31.4-GHz zenith brightness temperatures for a four-day period in April. The horizontal scale has two-hour markers. The vertical scale covers a range of zero to 110 K. The 31.4-GHz data are as low as 12 K (an all-time low for the 3-month observing period at DSS 63). Every day at 1600 UT, the sun passes through one of the tip curve pointing locations and causes a positive feature in this plot. The sky was clear most of this four-day period. Broken clouds were present on April 16, which accounts for some of the midday higher brightness temperatures.

Figure 4(b) is a plot of brightness temperatures for five days of the following week. Most of these data (as presented in this figure) is unusable because of cloud-produced sky brightness temperature "clumpiness." Tip curve reduction algorithms have not yet been developed for deriving good-quality radiometer gains (which are crucial for producing good-quality zenith brightness temperatures). The brightness temperatures that are "zero pegged" are clearly wrong, and they are caused by rainwater resting on the top of the radome (as will be shown below).

The entire data set from DSS 63 has been reduced by the procedure used in producing Figs. 4(a) and (b). A statistical analysis has also been performed on these data. However, since this reduction procedure has serious shortcomings in representing the rainy periods, these statistical results are a very pessimistic interpretation of the Spain measurements, and they will not be presented in this article. Fortunately, another reduction procedure has been developed which "rescues" the rainy periods.

IV. "Gain From Instrument Temperature" Reduction Method

Figure 4(c) is a re-reduction of the 31-GHz raw data used in creating the previous figure and is based on the assumption that knowledge of the SCAMS instrument physical temperature can be used to predict radiometer gain. By overlaying Figs. 4(b) and 4(c), it can be seen that the "well-behaved" parts of Fig. 4(b) agree with the corresponding parts of Fig. 4(c). This provides confidence in the quality of those parts of Fig. 4(c) corresponding to the "poorly behaved" periods of Fig. 4(b).

V. "Off-Zenith" Reduction Method

The reduction procedure that produced Fig. 4(c) still has shortcomings. According to this figure, the 31-GHz zenith

sky brightness temperature exceeds 110 K for a period of 3 hours on April 24. If this were a true sky brightness temperature, we could predict that at an elevation of 30 degrees we would expect to measure a brightness temperature exceeding 220 K for the same three-hour period.

Figure 5 is a plot of brightness temperature at one of the 30-degree elevation angle viewing directions (for the same period covered by the previous figure). The vertical scale is compressed by a factor of two so that the top of the plot is actually 220 K. The label for the vertical scale has not been changed, however. This means that the top of the scale, which is 220 K for 30-degree elevation data, can be read as predicting a 110-K brightness temperature at zenith. This is useful for the purpose of detecting non-sky contributions to the measured zenith brightness temperature. If there were no "water on the radome," for example, the two plots would overlap. (Strictly speaking, this is not true, because the 3-K cosmic background component does not double, and saturation effects reduce the doubling for large zenith brightness temperatures.)

Comparing Figs. 4(c) and 5 reveals that "water on radome" effects must have been present during the April 24 high brightness temperature three-hour period. Whereas the average zenith brightness temperature was 123 K between 0400 and 0500 UT, at 30 degrees of elevation it was not twice 123 K but actually lower than at zenith, being 61 K (which would predict a 34-K zenith value).

Figure 6 is a plot of brightness temperature versus zenith angle for the 1-hour period referred to in the previous paragraph. The solid line fits the measured brightness temperatures. The broken line fits the "zenith-derived" brightness temperature, according to the approximate formula: $T_{\text{zen equiv}} = 3 + (T_{\text{meas}} - 3)/\text{air mass}$. A dotted horizontal line is placed at 32 K to indicate the author's estimate of the zenith sky brightness temperature. This plot can be interpreted to mean that water on the radome contributed a component to brightness temperature represented by the difference between the broken line and the dotted line.

Referring to Fig. 6, the 32-K value for zenith brightness temperature is derived by extrapolating a hand-fitted line (fitted to the "zenith equivalent" data) to a zenith angle of 90 degrees which corresponds to the vertical edges of the radome, where, presumably, no water can remain standing. The lowest objectively determined zenith equivalent brightness temperature comes from the +60-degree zenith angle, and this would be a more conservative value to adopt (i.e., 34 K). The minimum zenith equivalent brightness temperature, wherever it occurs, would serve for determining an inferred sky component. This is the procedure that has been adopted for all subsequent data reduction that is described in this article.

Another example of the "off-zenith" method for reducing data is shown in Figs. 7(a) and 7(b). Figure 7(a) is a plot of the tip-curve-based 22- and 31-GHz brightness temperatures, showing that the sky was very "lumpy" for week number 6. Figure 7(b) is a plot of 31-GHz zenith brightness temperature based on radiometer gains that are determined from the instrument physical temperature. It contains both the "sky" and "water on radome" components. Figure 8, in addition to repeating the plot in the previous figure, shows the zenith equivalent brightness temperature derived from the -46- and -60-degree angles. The importance of using the "off-zenith" reduction method is most apparent during two 8-hour periods, one on May 2 and the other on May 3.

Figure 9 is very similar to Fig. 8, except that two-hour averages for the off-zenith data are used. The two-hour average data are obtained automatically by a program that accepts (from the user) gain adjustment information for the week in question, calculates two-hour average zenith-equivalent brightness temperatures for all zenith angles (for each two-hour period), and determines which viewing location was "driest," i.e., had the lowest persistent zenith-equivalent brightness temperature.

In using this automatic procedure for locating the "driest" part of the radome, it is implicitly assumed that any clear areas in the sky do not "hover" over a preferred viewing direction for periods as long as two hours. Indeed, the viewing direction which is found to be "driest" tends to persist for several two-hour periods, implying that the reduction algorithm is actually finding the part of the radome with the least amount of standing water, as opposed to a viewing direction which happened to have the least amount of cloud emission. Therefore, the lowest trace in Fig. 9 appears to be a fair upper-limit estimate of zenith brightness temperature.

Figure 10 is a printout produced by the automatic reducing program described in the previous paragraphs. It is a tabulation of data for week number 1. The dG column is a gain adjustment which the user entered, and will be described later. The right-most columns are average and maximum zenith-equivalent brightness temperatures for each of the five zenith angles. The column labeled "Best" is a duplicate of one entry of one of the five zenith angle columns. For example, the first row is for a two-hour period beginning March 29 at 10 hours UT, and the best viewing direction was +60 degrees. The next row shows a best viewing direction at -60 degrees.

Figure 11 is similar to the plot in Fig. 6, except that it is a plot of the "water on radome" component of measured sky brightness temperature for week number 5. The upper trace indicates maximum amount, and the other traces indicate 75th, 50th, and 25th percentiles. This plot is obtained by

blending individual two-hour traces with the zero level established by assuming that the driest viewing direction corresponds to "no water on radome."

Figure 12 is the same kind of plot for week number 6. The maximum "water on radome" effect is displaced toward the west. Presumably, the rain came from the west during this week.

It was mentioned earlier that the user had to enter information about gain adjustments for each particular week. This is accomplished by a three-step procedure. First, a plot of "tip curve derived zenith brightness temperature" is produced. Second, a plot of "instrument physical temperature-based zenith brightness temperature" is produced using the nominal "gain versus temperature" relation (unpublished). This plot can be for any desired offset in anticipation of an overall week's best offset. Third, the two plots are overlaid, and differences during the "well-behaved" parts of the tip curve data are noted. These differences are used to derive a table of gain adjustment values. For example, if the "gain from temperature" plot of zenith brightness temperatures is 1 K too high (compared to the tip curve temperatures), the gain needs to be adjusted by -1.5 units (where a unit is 0.01 count per degree of brightness temperature). This is a labor-intensive procedure and does not lend itself easily to automation. It has been done for all the Spain data. The success of the entire analysis described in this article depends on hand analysis derivations of gain adjustment tables.

Table 1 summarizes the results of the preceding reduction procedure for the entire 14-week observing period at DSS 63. The numbers on the left margin represent "zenith brightness temperature." These values include the components of brightness temperature attributable to the 2.9-K cosmic background, oxygen, water vapor, and liquid water. These results demonstrate that "water on radome" effects have been reduced to a negligible level through use of the elaborate reduction procedure described above.

The numbers in the body of the figure represent the number of occasions having average zenith brightness temperatures of the value indicated on the left, for the week number indicated at the top of the figure. The first column outside the box, on the right, is a summation across all weeks. The next column is a summation of the previous column, summing upward. It is an exceedance tabulation. The last column on the right is the same exceedance tabulation after conversion to percentages. It is to be read in the following manner: X percent of the time, two-hour averages of zenith brightness temperature are equal to, or less than, the brightness temperature value found at the left side of the figure. Thus, 53.5 percent of the time the two-hour averages of zenith brightness

temperature are equal to, or less than, 18 K. (Or, 53.3 percent of the time the two-hour averages of 31-GHz zenith brightness temperature are less than 19 K.)

VI. Results

Figure 13(a) summarizes the statistics on measured sky brightness temperatures for the 14-week observation period at DSS 63. This figure is an exceedance plot. The following example illustrates how this figure should be read: 75 percent of the two-hour averages of measured 31-GHz brightness temperatures correspond to brightness temperatures at a 30-degree elevation angle that are less than 37 K (after excluding the 2.9-K cosmic background component).

Another way to read Fig. 13(a) is illustrated by the following (different) example: 5 percent of the time, two-hour averages of 30-degree elevation angle brightness temperatures exceed 75 K. This brightness temperature level corresponds to an attenuation of about 0.8 dB.

Figure 13(b) is a plot similar to the previous one except that the averaging period is one hour (instead of two hours). The 75-K level is exceeded 5 percent of the time (which is the same as that for the two-hour averaging interval). Table 2 is a tabulation of the one-hour averaging period data.

Figure 14 combines exceedance information contained in the previous two exceedance plots and presents it in a slightly different format. Exceedance plots are shown for the 1-hour and 2-hour data for the high brightness temperatures (the plots are identical for lower brightness temperatures). This figure is to be read: 100-K brightness temperatures (at a 30-degree elevation angle) are exceeded 2.0 percent of the time for the two-hour averaged data and are exceeded 2.3 percent of the time for the one-hour averaged data.

It is not straightforward to produce similar exceedance plots for time intervals that are shorter than one hour. This is due to limitations of the complex reduction procedure used to remove "water on radome" effects. A great amount of time would be involved in extracting this information from the data, and at this time there are no plans to do this.

Figure 16 is an estimate of the family of curves relating exceedance plots to averaging period. The 1- and 2-hour plots are the measured plots from Fig. 14, while the other plot is the author's estimated extrapolation of the data as the observation time interval becomes increasingly small, or "instantaneous."

VII. Rainfall Adjustments

Figure 15 is a plot of rainfall versus season for Barajas. During the 1984 SCAMS observation period, the accumulated rainfall was 15.7 cm instead of the long-term average value of 10.7 cm. The observation period was therefore "wetter" than average. Thus, the exceedance plots presented in this article correspond to "wetter than average" conditions.

Rain-related weather defines the location and shape of the high brightness temperature end of the exceedance plots. There was 47 percent more rain during the observation period than during a typical April through June period. There must be many ways of adjusting the exceedance plots to produce a plot that corresponds to different weather conditions. The straightforward and intuitive assumption is that the "height" of the exceedance plot (at the high brightness temperature end) is directly proportional to the rate of rainfall for the interval in question. This method was used to produce Fig. 16, which is perhaps the best estimate that can be made from the SCAMS measurements of exceedance plots for average weather at DSS 63. This figure is the final product of this report.

Acknowledgment

The author wishes to thank Robert Clauss for suggesting that equivalent zenith brightness temperatures can be derived from off-zenith measurements, as described in Section V.

References

- [1] S. D. Slobin, "Models of Weather Effects on Noise Temperature and Attenuation for Ka- and X-Band Telemetry Performance Analysis," *TDA Progress Report 42-88*, vol. October–December 1986, Jet Propulsion Laboratory, Pasadena, California, pp. 135–140, February 15, 1987.
- [2] B. L. Gary, "Australia 31-GHz Observations of Sky Brightness Temperatures," *TDA Progress Report 42-94*, vol. April–June 1988, Jet Propulsion Laboratory, Pasadena, California, August 15, 1988.

Table 1. Statistical summary of Spain data after correction for radome water effects: 2-hour averages

Zenith brightness temperature, K	Week number ^a														Exceedance statistics		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Σ	$\Sigma\Sigma$	%
>80			1		2	-	-	-		-	1	2	-		6	994	100.0
61-80			-		2	-	2	2		3	-	1	2		12	988	99.4
51-60			5		2	-	6	3		1	-	-	-		17	976	98.2
41-50	3		7		1	3	1	7		2	3	2	1		30	959	96.5
31-40	6	2	6		3	7	7	8	7	12	3	1	4		66	929	93.5
26-30	4	2	7		3	3	4	13	7	3	3	3	6	1	59	863	86.8
23-25	11	-	2		5	2	6	8	3	6	9	2	6	2	62	804	81.0
21-22	4	-	2		4	2	4	5	5	3	5	5	27	15	81	742	74.7
20	5	3	2	1	-	2	4	4	6	2	7	3	17	5	61	661	66.6
19	4	-	4		5	3	2	4	6	6	12	4	15	4	69	600	60.4
18	2	4	5		6	4	2	9	7	3	15	20	7	8	92	531	53.5
17	2	8	31	1	7	3	-	9	5	11	13	10		1	101	439	44.2
16	1	26	6	10	10	1	8	12	16	14	2	9		2	117	338	34.0
15	2	20		15	28	4	14	6	8	15	7	17			136	221	22.3
14	2	6		15	-	1	16	3	2	2	2	4			53	85	8.6
13				10	5	2	4	1							22	32	3.2
12				6			1	3							10	10	1.0
Totals	46	71	78	58	83	37	81	97	72	83	82	83	85	38	994		

^aData show the number of occasions that 2-hour averages had the corresponding sky zenith brightness temperature.

Table 2. Statistical summary of Spain data with hourly averages

Zenith brightness temperature, K ^b	Week number ^a														Exceedance statistics		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Σ	ΣΣ	%
>80	—	—	1		3		1	3		3	1	4	—		16	1982	100.0
61–80	—	—	4		4		2	4		4	1	2	4		25	1966	99.2
51–60	1	—	9		2	1	10	5		2	2	2	—		34	1941	97.9
41–50	6	2	7		1	6	6	10	2	5	2	2	2		51	1907	96.2
31–40	9	—	16		7	15	11	21	12	13	10	3	5	1	123	1856	93.6
26–30	12	2	11		10	6	10	18	7	13	2	2	16	—	109	1733	87.4
23–25	12	1	6		14	5	6	10	9	12	12	9	13	1	110	1624	81.9
21–22	11	3	2		11	8	5	13	11	5	17	10	25	11	132	1514	76.4
20	8	2	2	2	9	6	4	6	11	5	14	4	42	16	131	1382	69.7
19	11	—	4	—	21	9	—	13	16	9	21	6	45	11	166	1251	63.1
18	7	6	9	—	27	5	2	18	13	11	34	39	13	14	198	1085	54.7
17	5	6	51	2	38	2	12	18	17	20	29	21	3	15	240	887	44.8
16	—	31	20	20	14	1	10	30	18	30	5	15		6	200	647	32.6
15	3	55	3	30	3	5	22	18	17	31	9	30		1	227	447	22.6
14	5	27		30		3	33	3	4	3	4	12		2	126	220	11.1
13	2	4		20			22	2	5	1	1	6			63	94	4.7
12	—	2		12			8		3		5	1			31	31	1.6
Totals	92	141	145	116	164	73	164	192	145	167	169	168	168	78	1982		

^aData show the number of occasions that 2-hour averages had the corresponding sky zenith brightness temperature.

^bIncludes 2.9 K.

ORIGINAL PAGE IS
OF POOR QUALITY

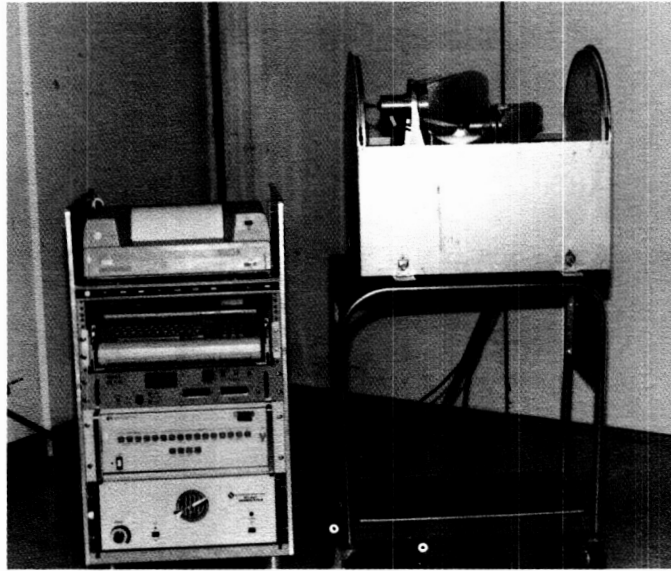


Fig. 1. Photograph of the SCAMS water vapor radiometer system prior to deployment

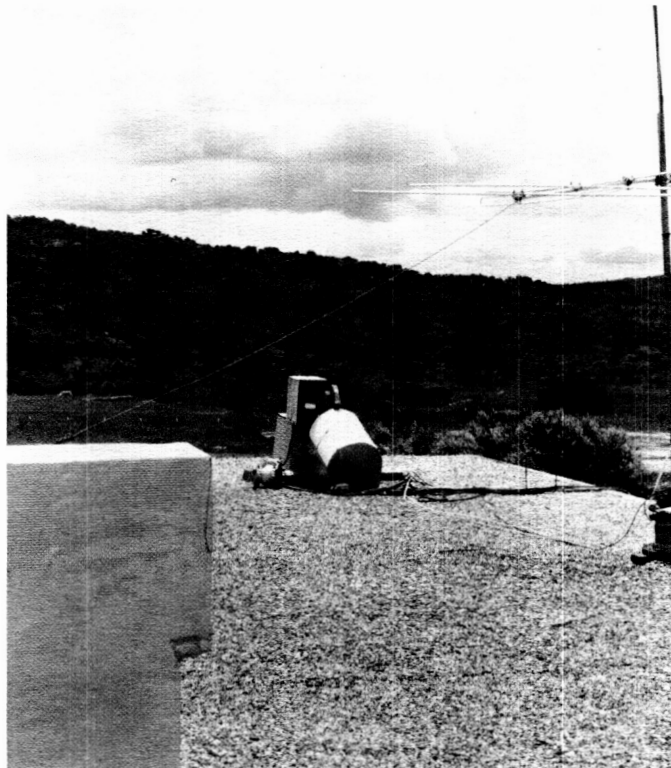


Fig. 2. Photograph of the SCAMS after deployment on the roof of the main control building at DSS 63. The view is to the south.

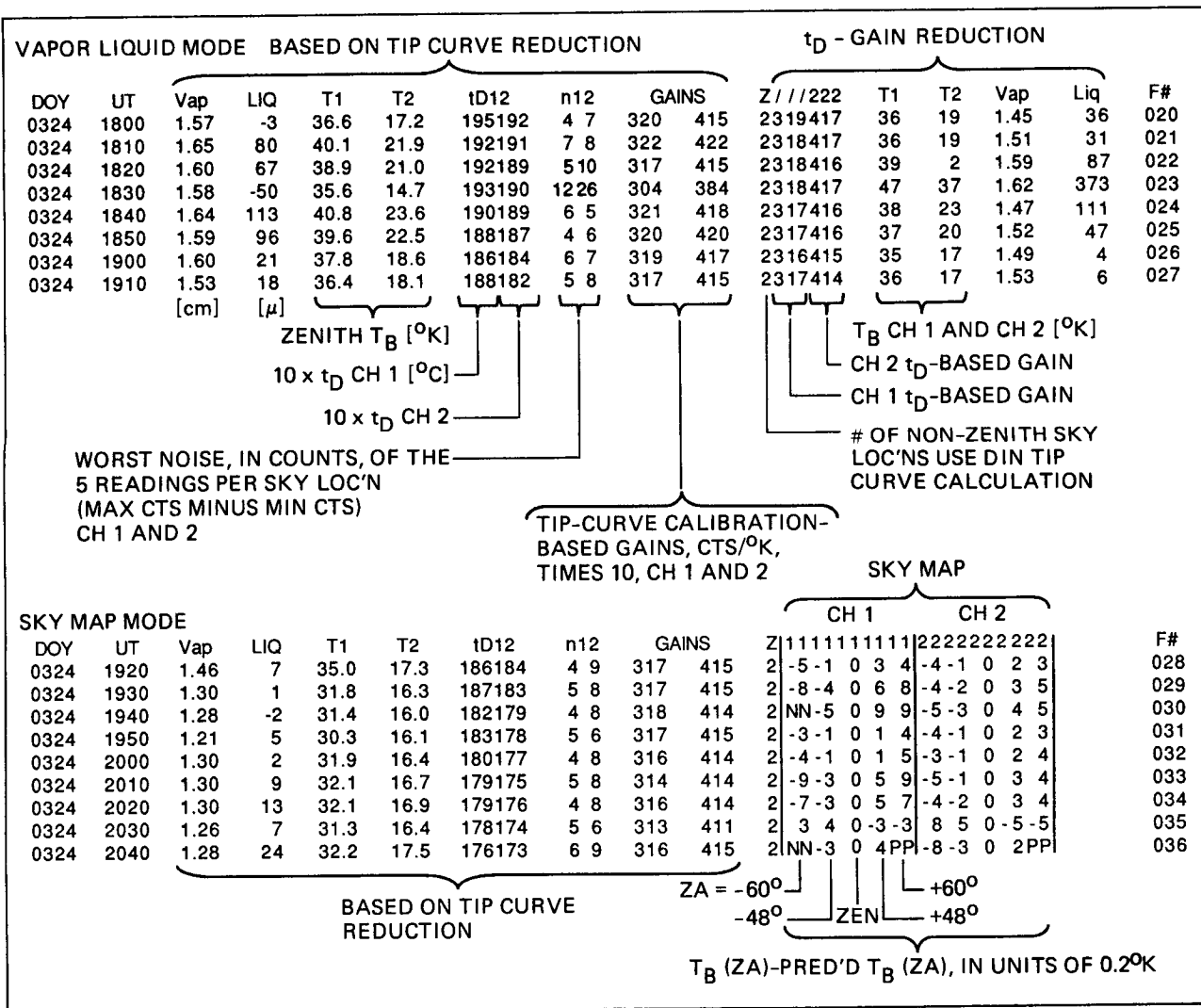


Fig. 3. Real-time printouts of the SCAMS observing program. Printouts for two data reduction modes are displayed.

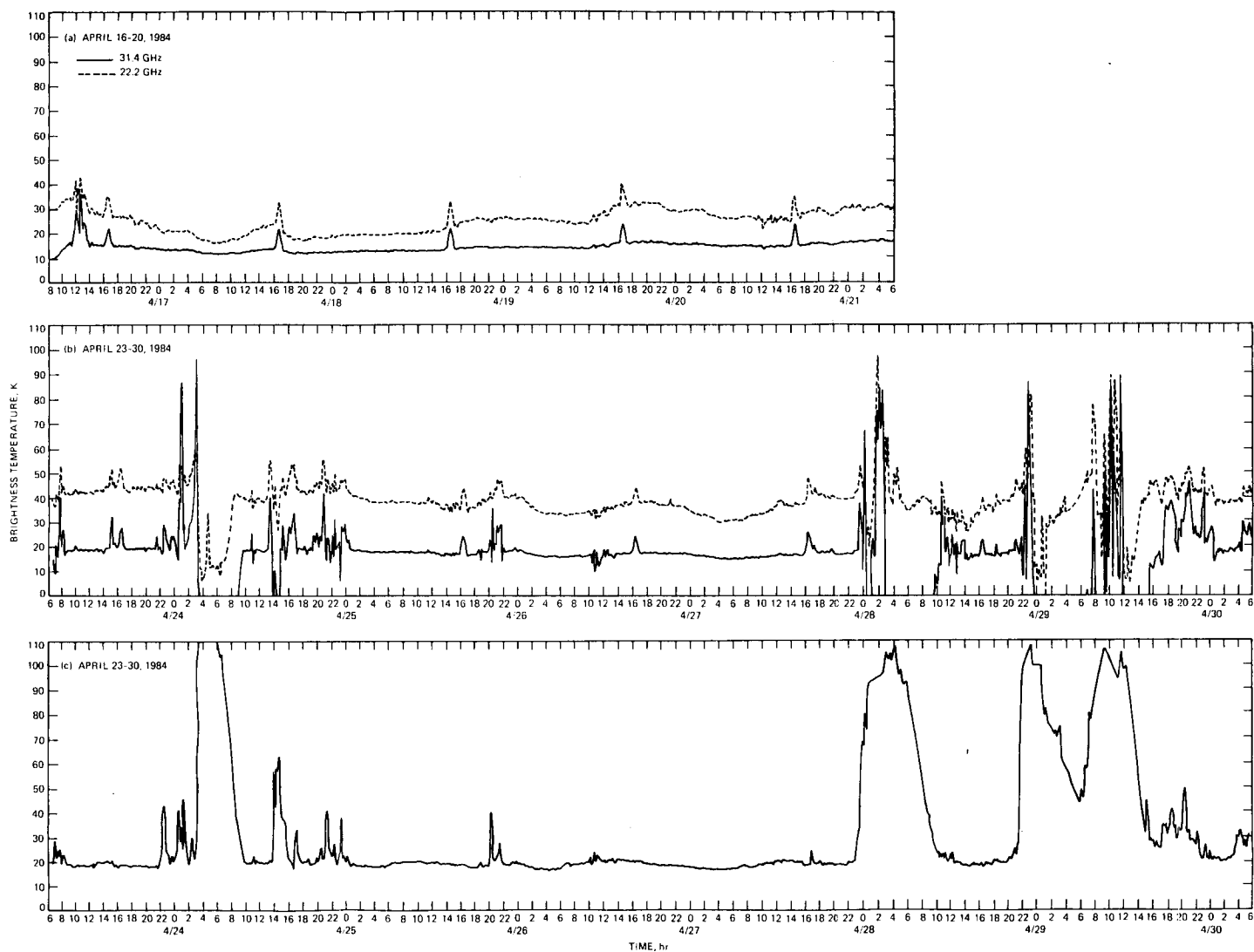


Fig. 4. Brightness temperatures for various dates and conditions at DSS 63: (a) plot of 22.2- and 31.4-GHz zenith brightness temperatures from April 16-21, 1984; (b) plot of brightness temperatures during cloudy conditions from April 23-30, 1984; (c) plot of the 31.4-GHz brightness temperature for the same period as shown in (b), using a data reduction algorithm that should give better results for cloudy conditions by deriving instrument gain from a gain versus instrument temperature relation, rather than deriving gain from tip curves.

FOLDOUT FRAME

FOLDOUT FRAME

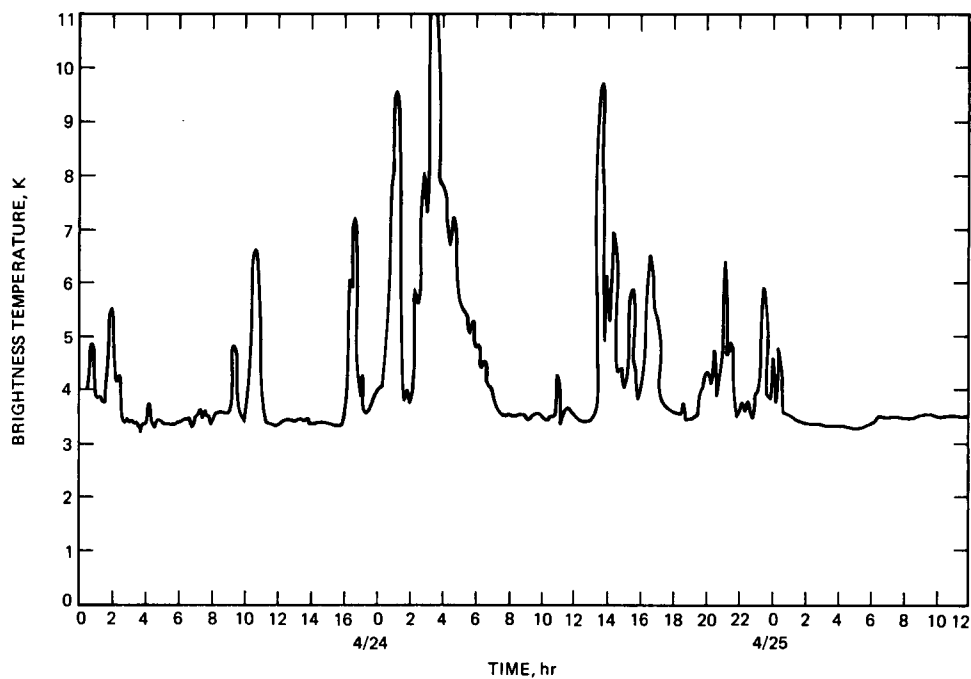


Fig. 5. The 31.4-GHz brightness temperatures for the viewing direction +60 zenith angle for the same period as Fig. 4(c). The Y-axis scale, shown extending from 0 to 110 K, is actually for 0 to 220 K. If the trace is thought of as a zenith equivalent brightness temperature, then the scale 0 to 110 K is appropriate.

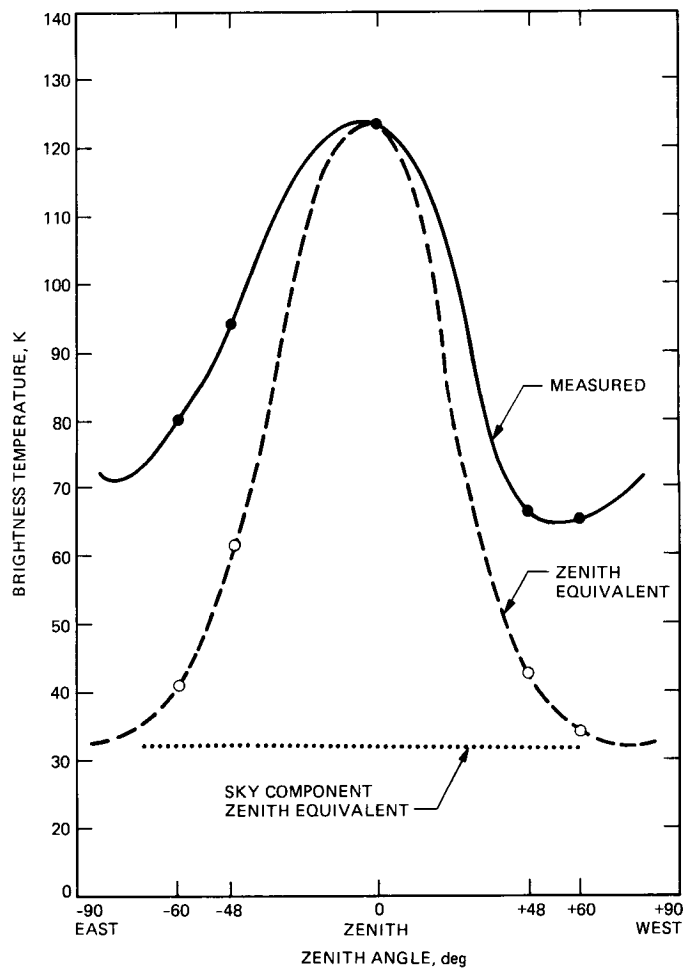


Fig. 6. Comparison of 31.4-GHz brightness temperatures at various zenith angles during a 1-hour period in which water is suspected of being on the radome. The dashed line is the zenith equivalent brightness temperature and the horizontal dotted line is the "no water" component attributable to the sky.

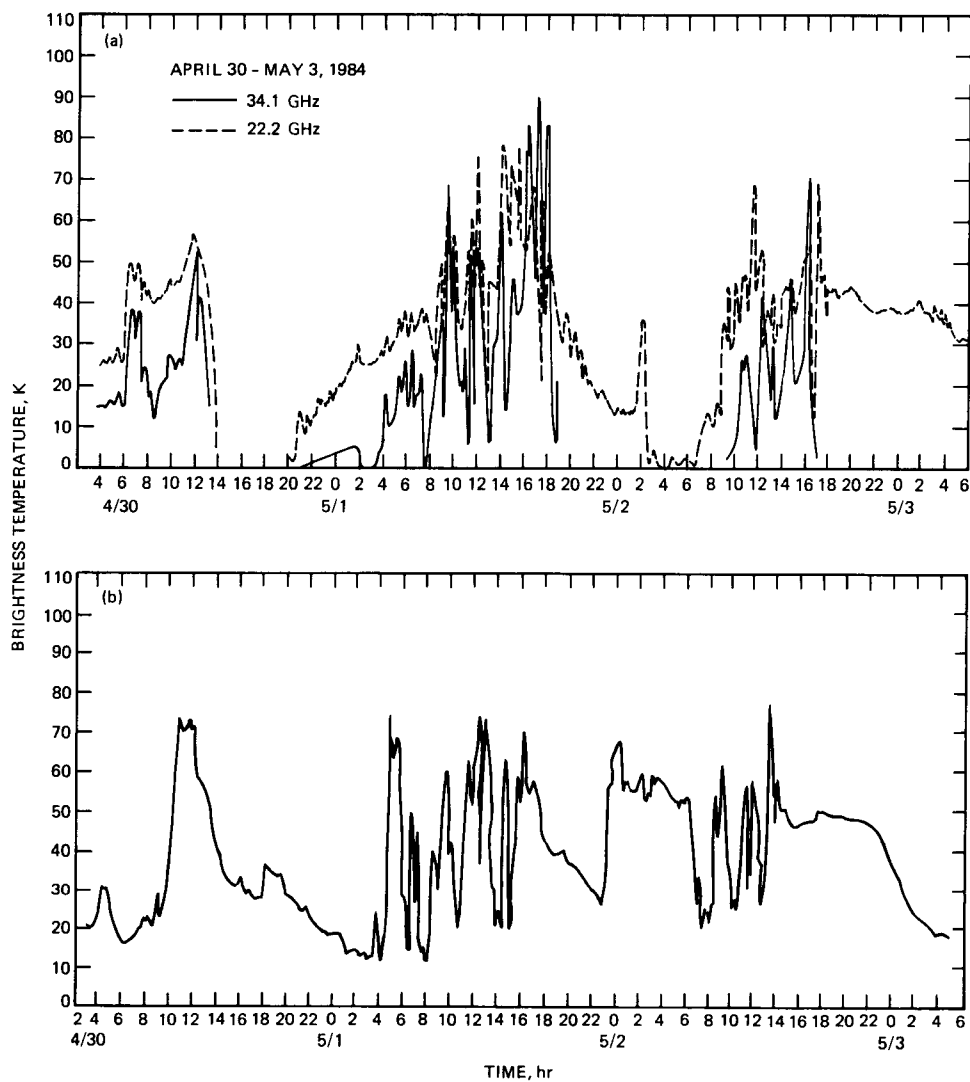


Fig. 7. Brightness temperatures for a second period during which water is suspected of being on the radome: (a) 22.2- and 31.4-GHz plot based on tip curve; (b) plot of zenith brightness temperatures based on the "instrument gain versus instrument physical temperature" relationship. Note that (b) is better behaved than (a).

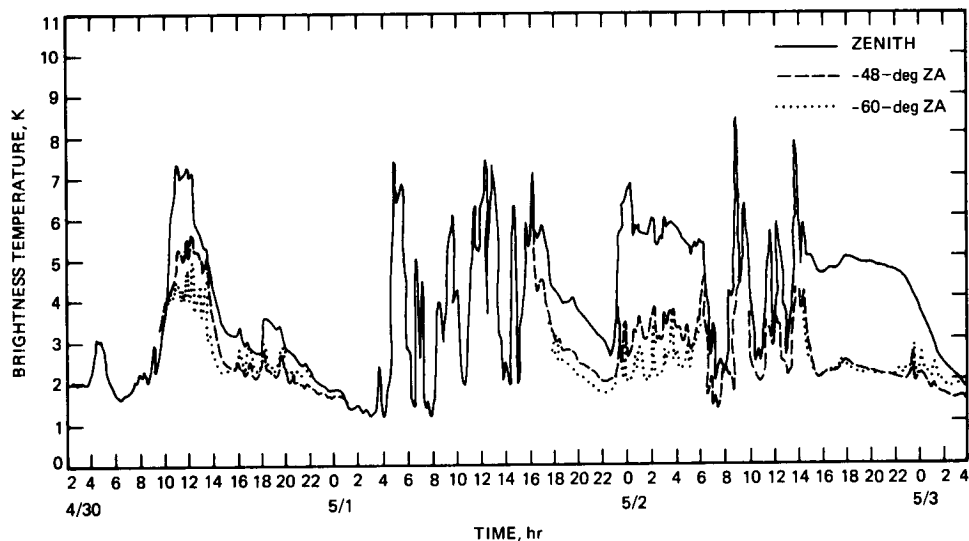


Fig. 8. Plot of zenith equivalent 31.4-GHz brightness temperatures derived from the observed brightness temperatures at 0, -48, and -60 degrees zenith angle. The observed zenith angle = 0 brightness temperature trace is taken from the previous figure.

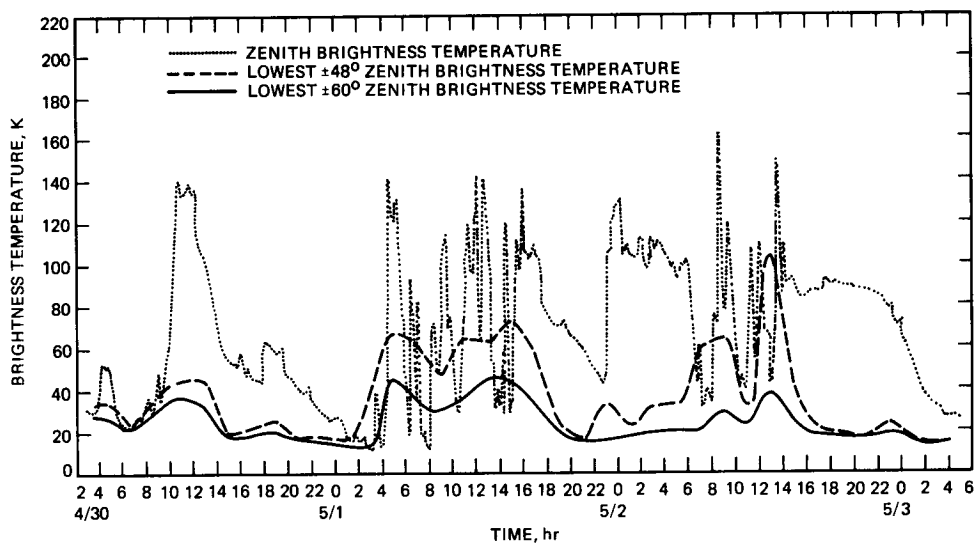


Fig. 9. The dotted line trace, taken from the previous figure, represents the zenith brightness temperature. The two bottom traces represent the lowest measured brightness temperatures during 2-hour observation periods at ± 48 and ± 60 zenith.

DATE/UT	dG	BEST		-60		-48		ZEN		+48		+60	
32910	4	23	30	23	29	25	32	28	40	25	32	23	30
32912	3	24	34	24	34	26	34	29	34	27	35	25	32
32914	2	31	57	31	57	36	70	60	109	46	74	39	61
32916	2	41	45	41	45	53	61	80	90	51	73	43	60
32918	1	41	62	41	62	46	68	95	114	62	87	51	68
32920	0	43	55	43	55	49	58	89	129	61	70	51	61
32922	-0	30	42	30	42	40	54	47	71	45	63	34	48
33000	-1	27	36	27	36	34	47	54	85	38	58	31	45
33002	-2	23	26	23	26	24	26	41	46	30	37	23	26
33004	-2	20	21	21	23	22	23	32	37	23	26	20	21
33006	-3	23	34	23	34	25	31	26	30	27	51	24	29
33009	-4	21	23	21	23	22	25	25	30	23	27	24	37
33012	-5	21	23	21	23	22	24	29	55	33	96	25	29
33015	-6	24	30	24	30	24	32	26	38	30	52	26	32
33017	-6	32	58	32	58	41	76	67	165	48	132	44	103
33019	-7	25	37	25	37	27	37	33	56	26	39	26	57
33021	-8	20	20	21	22	21	23	23	24	20	21	20	20
33023	-8	18	18	19	20	19	20	20	22	18	19	18	18
33100	-8	18	19	19	21	19	19	19	20	18	19	18	19
33102	-7	19	21	20	26	21	27	21	27	19	22	19	21
33104	-7	19	20	22	26	23	27	21	25	20	22	19	20
33106	-6	16	23	18	27	17	26	16	23	16	21	16	19
33108	-6	14	22	16	18	16	18	14	22	15	19	16	21
33110	-5	32	38	32	38	34	42	36	54	41	56	42	57
33112	-5	37	82	44	83	51	82	37	82	41	73	38	59
33114	-4	34	68	34	68	39	84	59	122	62	111	55	94
33116	-4	23	33	23	33	23	37	26	51	32	56	34	52
33118	-3	38	97	40	104	47	113	68	152	43	106	38	97
33120	-3	21	51	21	51	22	66	43	98	25	70	22	57
33122	-3	15	17	15	17	15	19	19	25	21	36	21	37
40100	-2	17	29	17	29	19	34	22	58	20	37	18	29
40102	-2	15	22	15	21	15	22	15	21	16	22	16	21
40104	-1	26	51	26	51	31	61	43	92	36	55	34	50
40106	-1	17	20	18	21	17	20	18	29	18	23	18	21
40108	-0	20	24	20	24	20	27	23	29	25	33	26	34
40110	0	28	41	28	41	30	42	36	46	37	47	34	44
40112	1	21	28	21	28	22	28	27	37	33	56	33	53
40114	1	23	26	23	26	26	31	30	37	38	50	32	45
40116	2	23	30	23	30	24	30	25	31	26	44	24	35
40118	2	24	28	26	35	24	28	25	29	25	33	24	29
40120	3	24	29	26	35	25	33	26	30	25	32	24	29
40122	3	20	20	20	20	20	21	21	22	20	21	20	20
40200	3	19	20	20	20	20	20	19	20	19	20	19	20
40202	3	19	21	19	20	20	22	20	24	19	22	19	21
40204	3	20	22	20	22	21	23	22	26	22	28	24	29

Fig. 10. Output of a program that automatically performs the "off-zenith" correction for water-on-radome effects

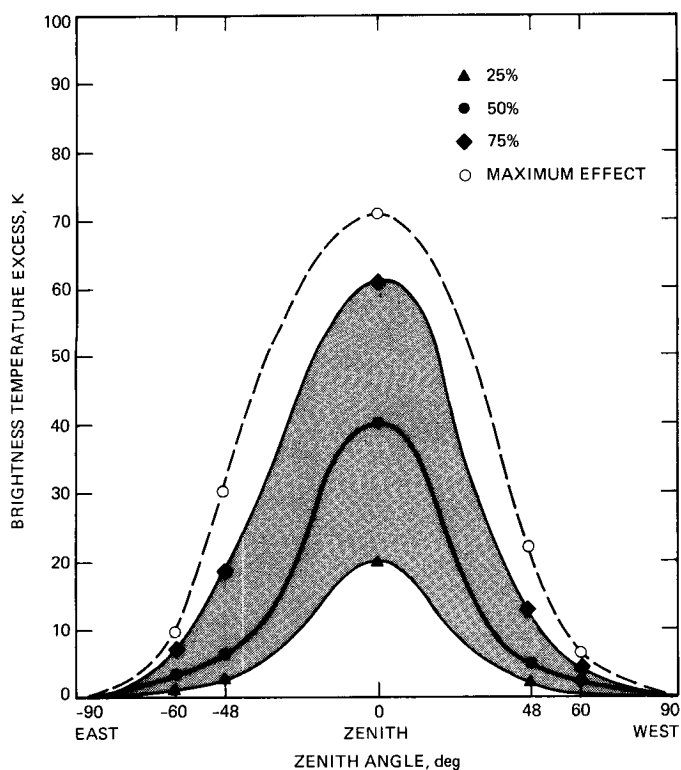


Fig. 11. Statistics for a week of data that are presumably affected by water on the radome, showing the component of 31.4-GHz brightness temperature versus zenith angle due to water on the radome. Percentile lines are shown for 25, 50, and 75 percent. A maximum effect trace is shown with a broken line.

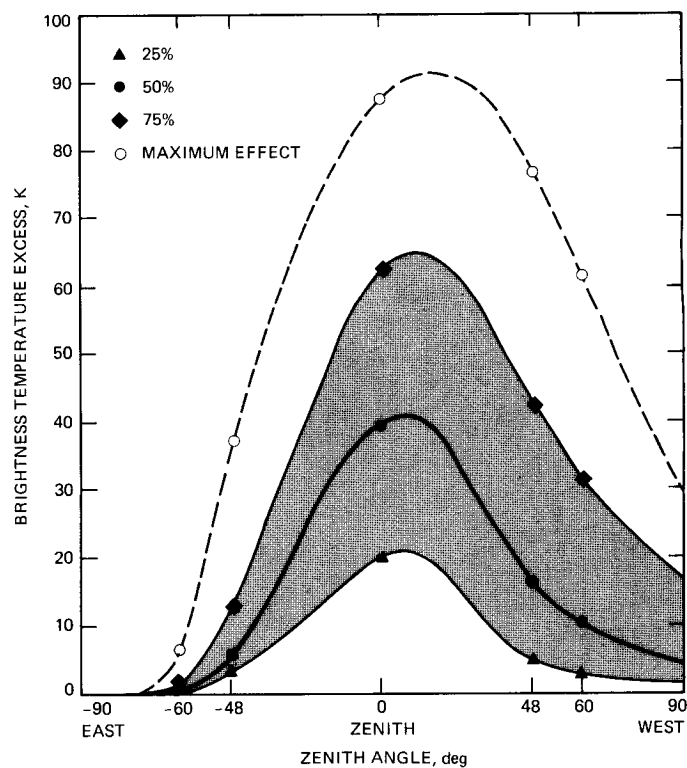


Fig. 12. Statistics analogous to those of Fig. 11 for a different week of data

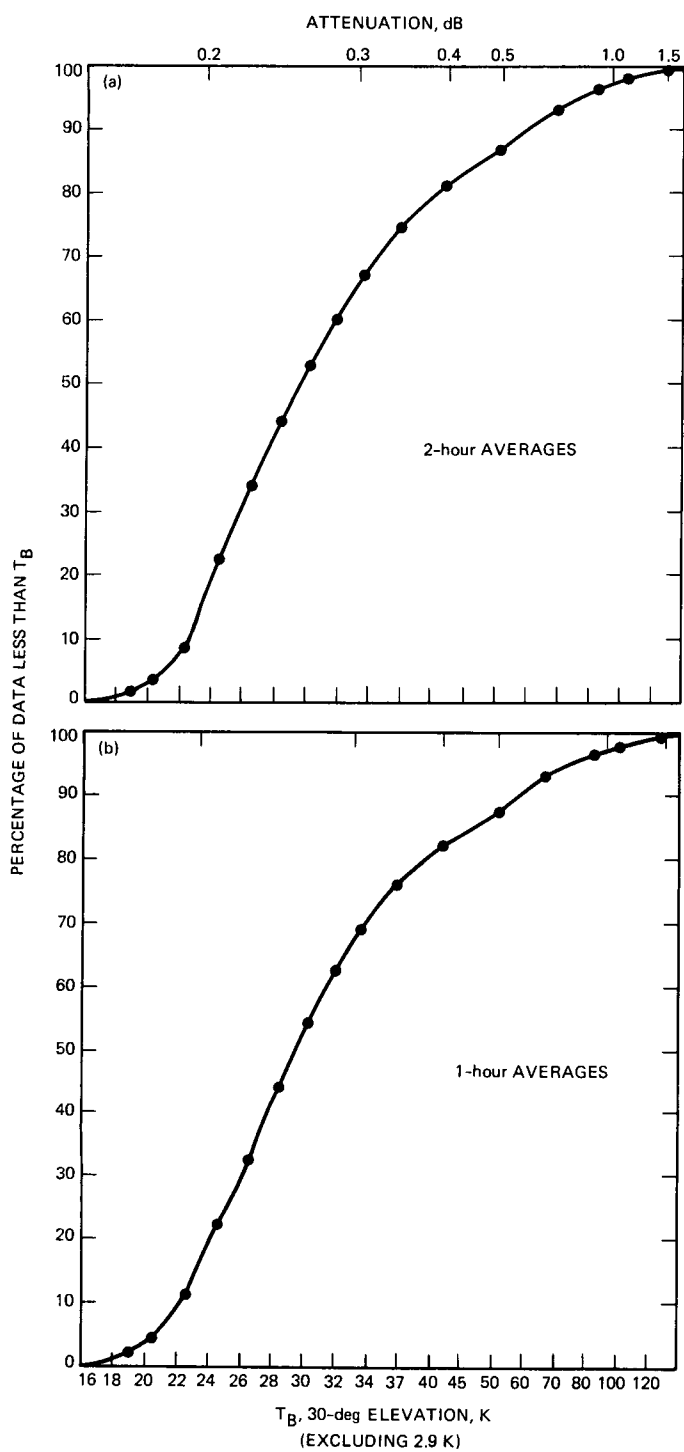


Fig. 13. Exceedance plots for 31.4-GHz sky brightness temperature for an elevation angle of 30 degrees: (a) data from 2-hour averages used; (b) data from 1-hour averages used.

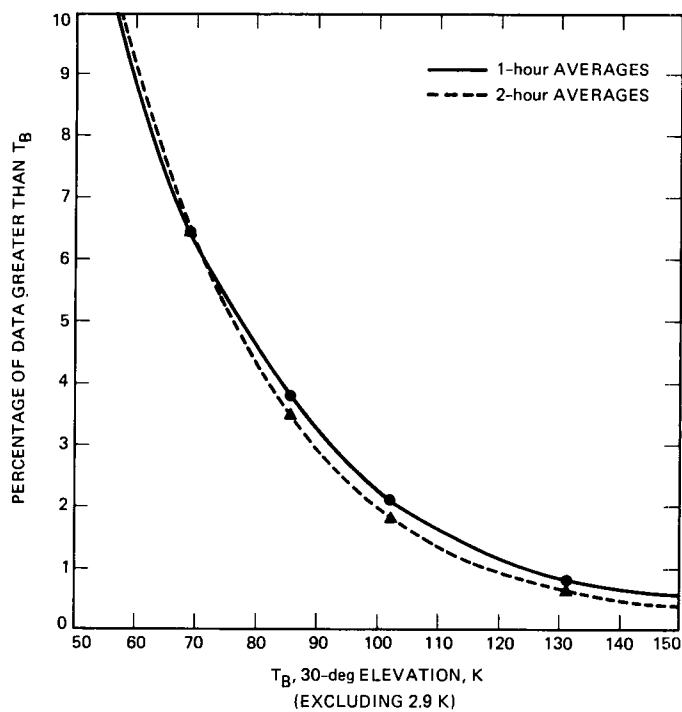


Fig. 14. Exceedance plots of 1- and 2-hour data averages for the "high end." The Y-axis is the percent of time that 30-degree elevation angle 31.4-GHz sky brightness temperature exceeds the X-axis value.

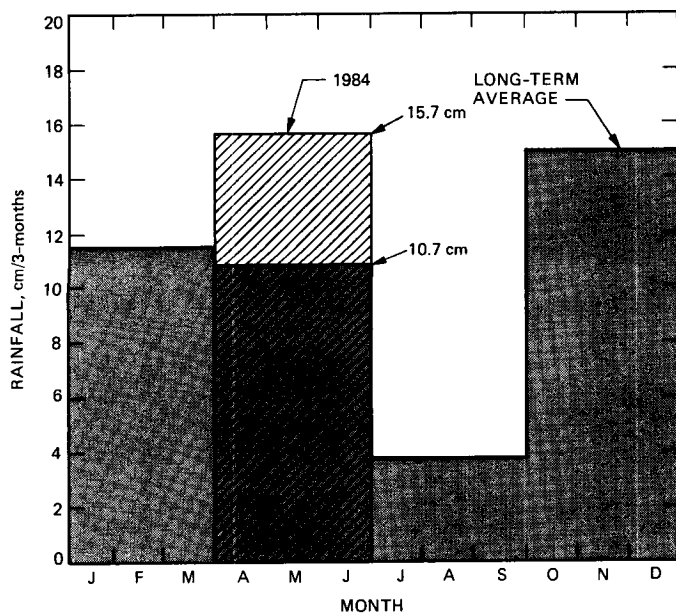


Fig. 15. Rainfall statistics for the 3-month observation period compared with historical averages. The observing period had 47 percent more rain than typical.

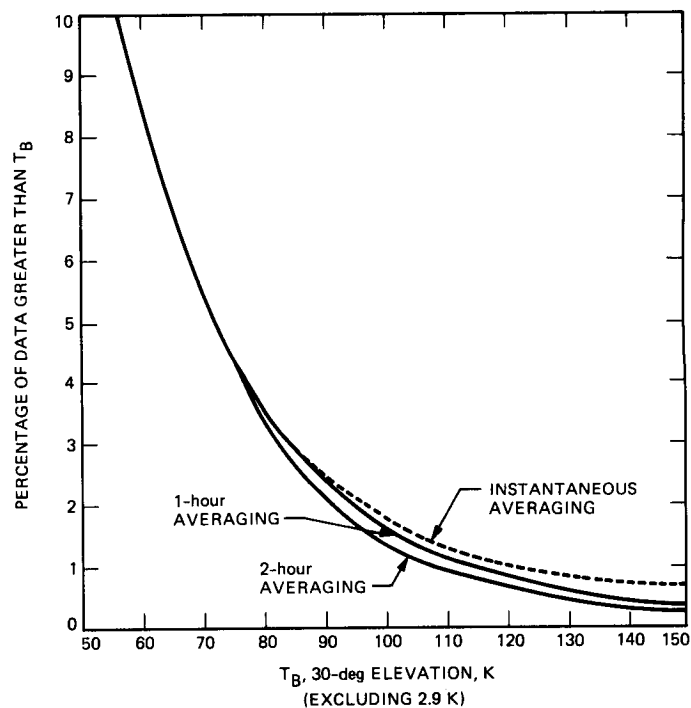


Fig. 16. Final plot of 31.4-GHz 30-degree elevation angle sky brightness temperature exceedances. The plots reflect a correction for the greater than average rainfall that occurred during the 14-week observation period. The 1- and 2-hour data averages are based on real data, whereas the broken line is an estimate for "instantaneous" statistics.

1989000820
617981
14p.55-46
~~165576~~
140
April-June 1988

N89 - 10191

JPL

Australia 31-GHz Brightness Temperature Exceedance Statistics

B. L. Gary

Microwave Observational Systems Section

Water vapor radiometer measurements were made at DSS 43 during an 18-month period. Brightness temperatures at 31 GHz were subjected to a statistical analysis which included correction for the effects of occasional water on the radiometer radome. An exceedance plot has been constructed, and the 1 percent exceedance statistic occurs at 120 K. The 5 percent exceedance statistic occurs at 70 K, compared with 75 K in Spain. These values are valid for all of the three-month data groupings that were studied.

I. Observations

In [1], models were derived for weather effects on Ka-band performance analysis. To use the models, measurements of noise temperature effects of a measured amount of water vapor are needed. Therefore, the RO6 water vapor radiometer (WVR) was deployed at DSS 43 in Australia for several years prior to early 1986. Beginning in the spring of 1984, RO6 was pointed at an elevation angle of 45 degrees so that line-of-sight data would be less affected by water on the radome cover. Tip curve data were taken for approximately 30 minutes once or twice weekly. All remaining hours were devoted to 45-degree line-of-sight data.

A line-of-sight observation consists of a 15-minute average of sky count data for the 20.7- and 31.4-GHz radiometer channels and a recording of physical temperatures within RO6. Hot load and base load readings were made. Data were recorded on a diskette, which was replaced twice weekly and mailed to JPL. The operation of RO6 was conducted by DSS 43 station personnel.

RO6 data from the 18-month period of July 1984 through December 1985 are the subject of the present article. Data from earlier epochs are subject to uncertain effects due to water remaining on the radome after precipitation, because the WVR was pointed at zenith at these earlier times. As will be explained below, radome water corrections are a crucial part of obtaining exceedance statistics at the high brightness temperature regime, and these corrections are much greater (and are subject to much greater uncertainties) for zenith data than for 45-degree data. See [2] for Spain measurements.

II. Data Reduction

In order to convert radiometer sky counts to brightness temperature, it is necessary to know the gain of the radiometer channel that is being used. Tip curves enable this measurement of gain to be made when sky conditions are clear. When it is cloudy, tip curve gains are "noisy." Plots of gain versus instrument temperature were made for several data groupings. Figure 1 is a plot of this relationship for both channels, and it

represents data from the 6-month period of July to December 1984. The solid line was determined from data from August 1985. The agreement of the plotted points in Fig. 1 with the lines indicates that the gain-temperature relationships did not change over a 13-month interval.

The gain-temperature relationship represented by the straight line in Fig. 1(b) was used to reduce all 31-GHz data that will be described in this report

$$\text{Gain 2} = 8.340 - 0.206 (T_{\text{amb}} - 40^{\circ}\text{C})$$

where the units for gain are counts/ $^{\circ}\text{C}$.

Figure 2 is a sample output from a computer program that reduces raw data files and calculates sky brightness temperatures. The first three rows correspond to tip curves, and they yielded measurements of the gain and tip curve correlation coefficient (R-squared) for each channel. The column of interest in this figure is labeled " TB_{zen} ." For the line-of-sight data cycles, TB_{zen} is the zenith-normalized brightness temperature. There is an entry in this column every 15 minutes. The channel 2 entries in this column-pair are "input" for the next data-reducing program.

The next program collects the 31-GHz zenith-normalized brightness temperatures for a user-specified time period and creates a histogram. Figure 3 is a histogram for the entire 18-month period. The left-most four columns are zenith-normalized brightness temperatures. The fifth column is the number of occasions that a measurement has a given brightness temperature value (designated by the first four columns).

Figures 4(a) and (b) are exceedance plots of 3-month groupings of the data in the previous figure. The figure labels use the term *wet radome* to remind the reader that the plots are based on data that have not been corrected for water that may have been on the RO6 radome when data were taken.

III. Radome Water Correction Procedure

High brightness temperature values can be produced by water drops remaining on the radome after rain or overnight dew. The SCAMS (Scanning Microwave Spectrometer) WVR has been used to investigate radome water effects. SCAMS measures sky brightness temperature at zenith angles that range from 0 (zenith) to ± 60 degrees (i.e., a 30-degree elevation angle at opposite azimuths). After rain, the zenith brightness temperatures are affected the most because more water remains at the top of the curved radome than on the sides. Measurements from each of the SCAMS viewing directions can be used to infer zenith brightness temperature, which will be

referred to as TB_z (for TB zenith-normalized). When the radome is dry and when there are no lumpy clouds, each of these TB_z values are the same. When different amounts of water are standing on the various parts of the radome, however, there are different TB_z values. The smallest value corresponds to the driest part of the radome. When the radome has water, the lowest value for TB_z usually comes from one of the 60-degree zenith angle viewing directions.

It is fortunate for the present investigation that SCAMS was used at DSS 43 in Spain for a 4-month study of 31-GHz brightness temperature statistics. During the course of analysis of these data, some interesting things were discovered about the effect of water on the SCAMS radome. These findings play a crucial role in deriving a radome water-correcting algorithm for the RO6 Australia data. The Spain data are described in [2]. Some of the Spain data that will be presented here are taken from this article.

Figure 5 is a plot of TB_z that is derived from the zenith, -48 -degree, and -60 -degree zenith angle measurements made in Spain. The vertical scale is labeled in 10-K increments, and the horizontal scale is labeled in 1-hour increments. The highest trace invariably corresponds to the zenith data, meaning that there is always more water at the top of the radome than on the sides. The three traces are together before high- TB_z events, and they approach each other during a "drying-out" period. Note that during an event there can be disparities as great as 24 K. While it is raining, the traces are "noisy." During the drying-out phase, the traces are smooth but separated. The last event has a drying-out period that lasts 13 hours.

An empirical relationship has been deduced for the amount of TB_z excess (attributable to radome water) and the measured (uncorrected) TB_z . This relationship has been derived for the ± 48 -degree SCAMS data. Figure 6 is a plot of the data that have enabled this relationship to be determined. The curves can be used to derive corrections, which the following example will illustrate. When measured TB_z is 80 K, 50 percent of the time the required correction is 27 K, 5 percent of the time the correction is 16 K, and another 25 percent of the time the correction is 35 K. This plot could have been used to correct ± 48 -degree measurements of TB_z to derive dry-radome-equivalent TB statistics. For SCAMS it was not necessary to use this procedure, since SCAMS measured TB_z for many angles, and the lowest TB_z value was adopted as corresponding to "dry."

Table 1 was derived from the plot of Fig. 6. Note that more categories are used for the redistribution of TBs. The following example will illustrate the use of this table. If there are 100 entries in the histogram of observed (uncorrected) TB_z values at the 100-K level, 10 percent of these entries (i.e.,

10 entries) would be moved to the histogram location 18 K below 100 K (i.e., $100 - 18 = 82$ K), 15 percent would be moved to the histogram location 24 K below 100 K, etc.

Before the SCAMS algorithm is used for correcting radome water effects that are present in the RO6 (Australia) data, it is necessary to demonstrate that the two WVRs respond to rain in the same way. That is to say, whenever TB_z for SCAMS is too large by a certain amount due to radome water, is TB_z for RO6 too large by the same amount? The answer is yes, and this was established by conducting side-by-side tests with a garden hose simulation of rain.¹

IV. Corrected Brightness Temperature Statistics

Figure 7 is a histogram of the "corrected" brightness temperatures. As in Fig. 3, the first four columns are zenith brightness temperatures and the fifth column is the number of entries in the histogram bin-range.

Table 2 outlines exceedance statistics for the observed (uncorrected) and "radome-water-corrected" brightness temperature histograms. Columns 1 and 2 correspond to "zenith" and "30-degree elevation angle" brightness temperatures (excluding the 3-K cosmic background component). Column 3 is an exceedance table for the uncorrected data (in Fig. 3), and column 4 is an exceedance table for the corrected data (in Fig. 7). Columns 5 through 10 correspond to 3-month group-

ings of the uncorrected data. Columns 11 through 16 correspond to 3-month groupings of the corrected data.

Figures 8(a) and (b) are exceedance plots of the uncorrected and corrected TBs (columns 3 and 4 of Table 2). The difference between the plotted line and "▲" symbols shows the result of correcting for radome water effects. The difference is "substantial," which accounts for the long explanation presented (in the previous section) of what was involved in correcting measured TBs for the effects of radome water.

Figure 9 shows exceedance plots for the 3-month groupings of corrected TB data. It is noteworthy that the exceedance statistics do not change with season in the 1 percent region.

The Spain exceedance statistics are shown in the next two figures. Figure 10 is for the 4-month observing period March to June 1984. (Remember that all Spain data shown here would correspond to "dry" even though the correcting table described above was not employed in reducing the Spain data.) The several lines correspond to different data-averaging times. The Australia data have an averaging time of 15 minutes per data point, so that is the appropriate time to use in selecting an exceedance trace in this figure. Figure 11 is based on the data in the previous figure, and it reflects an adjustment that was made for the difference in rainfall that occurred during the 4-month observing period and the climatic average for Madrid. This adjustment is described in the Spain exceedance statistic report [2].

Figure 12 is a repeat of Fig. 9, except that the Spain data of the previous figure are plotted for comparison. There is a remarkable similarity in the levels of the exceedance plots from both stations.

¹B. L. Gary, "Rain on Radome Effects," internal memorandum, Jet Propulsion Laboratory, Pasadena, California, November 3, 1985.

References

- [1] S. D. Slobin, "Models of Weather Effects on Noise Temperature and Attenuation for Ka- and X-Band Telemetry Performance Analysis," *TDA Progress Report 42-88*, vol. October–December 1986, Jet Propulsion Laboratory, Pasadena, California, pp. 135–140, February 15, 1987.
- [2] B. L. Gary, "Spain 31-GHz Brightness Temperature Exceedance Statistics," *TDA Progress Report 42-94*, vol. April–June 1988, Jet Propulsion Laboratory, Pasadena, California, August 15, 1988 (this issue).

Table 1. Correction table for radome water effects (corrections to brightness temperature versus uncorrected brightness temperature)

TB_z	Redistribution percentages					
	10	15	25	25	15	10
25	0	0	0	0	0	0
30	0	0	1	3	4	6
40	3	6	9	11	13	15
50	7	9	13	17	21	23
70	11	17	22	26	30	36
100	18	24	29	33	38	44
140	24	30	35	39	45	51
200	31	37	42	46	54	60
220	33	39	44	48	56	63

Table 2. Exceedance table for 18-months of 31-GHz brightness temperature data at DSS-43

TB	TB ₃₀	All wet	All dry	4JASW	4ONDW	5JFMW	5AMJW	5JASW	5ONDW	4JASD	4ONDD	5JFMD	5AMJD	5JASD	5ONDD
		%>	%>												
10.0	14.0	98.8	98.8	98.2	97.6	99.9	98.7	99.9	100.0	98.2	97.6	98.6	98.6	99.0	99.2
11.0	15.9	96.4	96.2	93.7	94.9	99.2	96.5	98.2	98.5	93.5	94.8	96.6	96.6	97.0	97.3
12.0	17.9	91.1	90.8	83.2	89.9	96.8	91.6	85.6	96.4	82.8	89.6	92.6	92.2	90.2	91.3
13.0	19.9	84.2	83.5	70.8	83.3	94.1	83.2	68.3	91.6	70.0	82.8	87.4	85.7	80.5	82.5
14.0	21.8	76.6	75.7	60.0	75.7	90.7	75.6	51.1	84.1	58.9	75.0	81.3	79.1	70.6	73.1
15.0	23.7	68.8	67.5	50.7	67.8	86.0	66.1	41.2	75.2	49.4	66.9	74.5	71.2	62.1	64.5
16.0	25.7	60.2	58.6	43.1	58.7	79.3	56.4	34.8	64.8	41.6	57.5	66.0	62.2	53.8	55.7
17.0	27.6	52.7	50.8	37.9	50.8	73.0	47.1	30.7	54.0	36.2	49.3	58.5	53.9	46.8	47.9
18.0	29.5	46.6	44.4	34.2	44.8	65.8	40.0	27.6	45.9	32.4	43.2	51.6	46.9	40.9	41.5
19.0	31.4	40.8	38.4	30.6	39.2	58.2	33.8	24.8	39.1	28.7	37.3	44.8	40.3	35.3	35.7
20.0	33.3	35.3	32.7	27.8	34.2	49.1	29.4	23.1	33.0	25.9	32.2	37.6	34.1	30.4	30.5
22.0	37.1	27.0	24.1	23.8	26.4	34.2	23.5	20.7	25.3	21.7	24.1	26.2	24.5	23.0	23.0
24.0	40.8	21.7	18.0	19.9	22.0	24.1	19.7	18.7	21.2	17.3	19.0	18.3	18.0	17.6	17.7
26.0	44.6	18.5	15.0	17.8	19.6	17.8	17.4	17.2	18.6	15.1	16.3	14.8	14.9	14.9	15.0
28.0	48.3	16.2	12.2	16.0	17.9	13.9	16.2	15.9	16.7	12.7	13.1	11.6	12.1	12.1	12.3
30.0	51.9	14.7	9.9	14.9	16.5	11.5	15.2	14.9	15.3	10.8	10.4	9.0	9.7	9.8	10.0
32.0	55.6	13.3	8.2	13.9	15.0	9.7	14.1	13.8	14.1	9.3	8.3	7.2	7.9	8.1	8.3
34.0	59.2	12.1	7.1	12.9	13.0	8.8	13.3	12.6	12.9	8.2	7.0	6.0	6.7	6.9	7.1
36.0	62.7	11.2	6.2	12.0	11.9	8.0	12.6	11.5	12.1	7.3	6.0	5.2	5.8	5.9	6.2
38.0	66.3	10.2	5.4	11.3	10.4	7.2	11.9	10.4	11.1	6.4	5.2	4.5	5.0	5.1	5.3
40.0	69.8	9.4	4.8	10.7	9.5	6.5	10.9	9.6	10.3	5.7	4.6	3.9	4.4	4.5	4.7
42.0	73.3	8.6	4.2	9.9	8.6	5.7	10.1	8.9	9.5	5.0	4.0	3.4	3.8	3.9	4.1
44.0	76.8	7.9	3.8	9.2	7.7	5.2	9.5	8.2	9.0	4.6	3.6	3.1	3.4	3.5	3.6
46.0	80.2	7.3	3.3	8.7	7.2	4.5	8.7	7.5	8.3	4.0	3.2	2.7	3.0	3.0	3.2
48.0	83.6	6.8	3.0	8.1	6.7	4.1	8.0	7.0	7.7	3.6	2.9	2.5	2.8	2.8	2.9
50.0	87.0	6.3	2.7	7.6	6.1	3.8	7.0	6.4	7.3	3.2	2.6	2.3	2.5	2.5	2.6
55.0	95.3	5.2	2.2	6.5	4.8	3.0	5.3	5.4	5.9	2.5	2.1	1.9	2.0	1.9	2.1
60.0	103.4	4.2	1.8	5.2	4.1	2.4	4.5	4.2	4.7	2.0	1.7	1.6	1.6	1.6	1.7
65.0	111.3	3.5	1.4	4.4	3.5	2.1	3.8	3.2	3.8	1.5	1.4	1.3	1.3	1.2	1.3
70.0	119.1	3.0	1.1	3.6	2.9	1.8	3.1	2.5	3.3	1.1	1.1	1.1	1.1	1.0	1.1
75.0	126.6	2.5	0.9	3.0	2.4	1.7	2.5	2.1	2.8	0.9	0.9	0.9	0.9	0.9	0.9
80.0	134.0	2.1	0.8	2.5	2.1	1.4	2.0	1.7	2.3	0.7	0.8	0.8	0.8	0.7	0.8
85.0	141.2	1.8	0.6	2.1	1.8	1.2	1.6	1.4	2.0	0.5	0.7	0.7	0.7	0.6	0.7
90.0	148.3	1.5	0.5	1.7	1.5	1.2	1.2	1.2	1.7	0.4	0.6	0.6	0.6	0.5	0.6
95.0	155.1	1.3	0.5	1.4	1.2	1.1	1.0	0.9	1.4	0.3	0.5	0.6	0.5	0.5	0.5
100.0	161.8	1.1	0.4	1.1	1.0	1.0	0.8	0.8	1.2	0.3	0.4	0.5	0.4	0.4	0.4
110.0	174.5	0.8	0.3	0.7	0.8	0.8	0.5	0.6	0.8	0.1	0.3	0.4	0.3	0.3	0.3
120.0	186.6	0.6	0.2	0.5	0.6	0.7	0.4	0.4	0.6	0.1	0.2	0.3	0.2	0.2	0.2
130.0	197.9	0.4	0.1	0.3	0.5	0.6	0.2	0.3	0.4	0.0	0.1	0.2	0.2	0.2	0.2
140.0	208.5	0.3	0.1	0.2	0.3	0.5	0.1	0.2	0.4	0.0	0.1	0.1	0.1	0.1	0.2
150.0	218.3	0.2	0.1	0.1	0.3	0.4	0.1	0.2	0.2	0.0	0.0	0.1	0.1	0.1	0.1
160.0	227.5	0.2	0.0	0.0	0.2	0.3	0.0	0.1	0.2	0.0	0.0	0.1	0.1	0.1	0.1
170.0	235.8	0.1	0.0	0.0	0.1	0.3	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0
180.0	243.5	0.1	0.0	0.0	0.1	0.2	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0
190.0	250.4	0.1	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
200.0	256.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

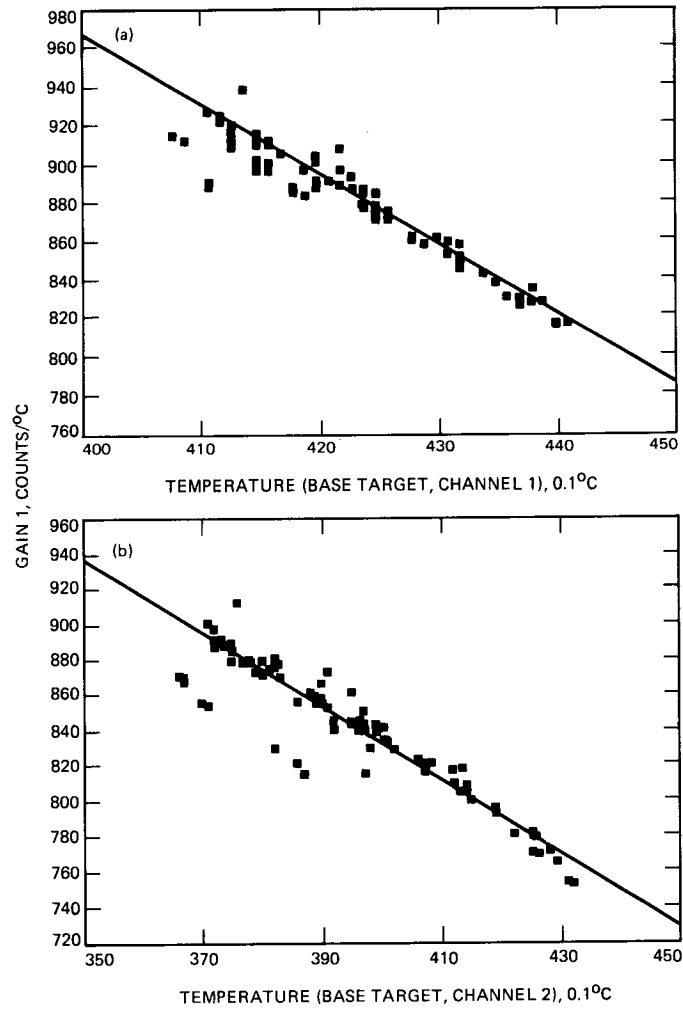


Fig. 1. WVR instrument gain versus mixer temperature from tip curves during a 6-month period in 1984. The straight line is based on data from August 1985.

01-10-1987 10:24:57 B:A5306.DAT to A:A5306.001
J-range = 2 to 14 TB' REJECT CRITERION= 100 degK

WILL DO TIP REDUCTION

NOTE: TIP data has entries up to TBzen columns; line-of-sight data has dots in some columns, & TBobs'd is for 45 deg elev'n

DAY	HR	MN	..TEMPS..	..GAINS...	..TB obs'd..	..R**2...	...RMS...	RJ	VAP	LIQ	..TBzen..	..good.tip.data..	G(t)
306	2	57	43.6 42.2	8.35 7.84	21.9 15.7	1.00 0.99	0.67 0.62	11	1.47	-26	21.9 15.7	43.6 8.35 42.2 7.84	835 789
306	3	2	43.6 42.2	8.38 7.87	22.8 16.4	1.00 1.00	0.61 0.52	00	1.55	-20	22.8 16.4	43.6 8.38 42.2 7.87	835 788
306	3	8	43.5 42.1	8.39 7.88	22.8 16.7	1.00 1.00	0.53 0.47	00	1.54	-11	22.8 16.7	43.5 8.39 42.1 7.88	841 789
306	3	20	43.7 42.2	30.1 23.3	0.32 0.29	00	1.45	25	22.5 17.5	834 787
306	3	35	43.6 42.3	31.1 23.2	0.25 0.20	00	1.54	9	23.2 17.4	835 785
306	3	50	43.7 42.3	31.5 23.9	0.22 0.13	00	1.55	22	23.5 18.0	834 785
306	4	5	43.7 42.4	30.9 23.7	0.13 0.06	00	1.50	24	23.0 17.8	832 784
306	4	20	43.7 42.3	31.3 23.7	0.30 0.27	00	1.54	18	23.3 17.8	834 784
306	4	35	43.7 42.3	31.3 23.6	0.17 0.08	00	1.54	17	23.3 17.7	834 786
306	4	50	43.7 42.3	31.7 23.8	0.24 0.18	00	1.57	17	23.6 17.9	834 785
306	5	5	43.6 42.2	32.8 24.2	0.16 0.14	00	1.64	14	24.4 18.2	835 787
306	5	20	43.5 42.1	33.8 24.4	0.20 0.21	00	1.72	6	25.1 18.3	839 788
306	5	35	43.5 42.1	33.6 24.2	0.19 0.35	00	1.71	3	25.0 18.2	839 790
306	5	50	43.6 42.1	33.0 24.6	0.47 0.39	00	1.65	21	24.6 18.5	835 790
306	6	5	43.7 42.2	35.1 25.2	2.39 1.39	00	1.81	8	26.1 18.9	834 788
306	6	20	43.4 42.0	39.6 27.1	0.35 0.30	00	2.13	-4	29.4 20.3	842 791
306	6	35	43.4 42.0	37.3 25.3	0.72 0.82	00	1.99	-19	27.7 18.9	842 792
306	6	50	43.4 41.7	35.8 23.5	0.53 0.71	00	1.92	-42	26.6 17.7	844 798
306	7	5	43.2 41.4	36.2 23.0	0.20 0.37	00	1.98	-61	26.9 17.3	850 804
306	7	20	43.0 41.1	38.1 24.3	0.23 0.21	00	2.10	-53	28.3 18.3	857 811
306	7	35	42.9 40.9	37.2 24.0	0.86 0.69	00	2.03	-49	27.6 18.0	861 815
306	7	50	42.8 40.5	37.1 24.3	0.38 0.47	00	2.01	-41	27.6 18.2	867 823
306	8	5	42.5 40.1	39.0 24.2	0.38 0.43	00	2.18	-68	29.0 18.2	878 830
306	8	20	42.3 39.8	40.2 26.9	0.31 0.67	00	2.19	-16	29.8 20.1	883 837
306	8	35	42.3 39.6	39.1 27.2	0.64 1.10	00	2.08	6	29.1 20.4	884 841
306	8	50	42.2 39.5	40.2 28.5	0.52 1.37	00	2.13	24	29.8 21.3	888 844
306	9	5	42.2 39.3	41.0 31.5	0.30 0.78	00	2.09	90	30.4 23.5	888 846
306	9	20	42.1 39.2	42.0 31.7	0.36 0.82	00	2.18	81	31.2 23.6	891 848
306	9	35	42.1 39.1	42.0 33.9	0.97 1.94	00	2.10	137	31.2 25.2	892 851
306	9	50	42.2 39.2	38.4 29.6	0.64 1.45	00	1.93	75	28.5 22.1	888 849
306	10	5	42.2 39.3	38.0 27.1	0.50 0.95	00	1.99	17	28.2 20.3	888 848
306	10	20	42.2 39.2	38.9 28.5	0.20 0.39	00	2.02	39	28.9 21.3	890 848
306	10	35	42.2 39.3	38.0 29.0	0.48 0.88	00	1.92	65	28.2 21.6	887 848
306	10	50	42.2 39.2	38.2 28.0	0.40 0.78	00	1.98	37	28.4 20.9	888 849
306	11	5	42.1 39.1	38.7 27.3	0.23 0.62	00	2.04	13	28.7 20.4	893 851
306	11	20	42.2 39.2	38.2 27.5	0.29 0.53	00	1.99	25	28.3 20.5	888 850
306	11	35	42.2 39.2	39.6 29.8	0.46 1.00	00	2.04	64	29.4 22.2	890 850
306	11	50	42.2 39.2	37.6 26.2	0.58 1.15	00	1.99	-1	28.0 19.6	888 850
306	12	5	42.2 39.2	38.0 27.1	0.91 1.90	00	1.99	18	28.2 20.3	890 849
306	12	20	42.2 39.2	40.4 33.8	0.73 1.59	00	1.95	155	30.0 25.1	887 848
306	12	35	42.2 39.2	40.3 32.0	0.59 1.21	00	2.02	109	30.0 23.8	888 848
306	12	50	42.2 39.3	39.2 29.2	0.65 1.26	00	2.02	54	29.1 21.8	887 847
306	13	5	42.3 39.3	40.1 33.2	1.18 2.50	00	1.96	144	29.8 24.7	883 846
306	13	20	42.3 39.4	43.3 39.2	1.24 2.95	00	2.01	258	32.1 29.1	885 845
306	13	35	42.3 39.3	42.3 38.8	0.91 1.78	00	1.94	261	31.4 28.8	885 846
306	13	50	42.2 39.3	40.6 35.2	0.66 1.40	00	1.92	189	30.1 26.1	887 846
306	14	5	42.2 39.3	39.7 33.7	0.99 2.05	00	1.90	163	29.5 25.1	887 847
306	14	20	42.2 39.3	40.8 34.4	1.04 2.24	00	1.96	168	30.3 25.6	890 847
306	14	35	42.2 39.3	40.8 35.3	1.00 2.02	00	1.94	188	30.3 26.2	888 847
306	14	50	42.2 39.3	39.6 33.2	0.29 0.59	00	1.91	150	29.4 24.7	888 847
306	15	5	42.1 39.2	41.0 34.8	0.67 1.38	00	1.97	174	30.5 25.9	891 849
306	15	20	42.2 39.3	38.4 32.0	0.43 0.72	00	1.84	138	28.5 23.9	887 848
306	15	35	42.2 39.2	39.4 34.1	0.50 1.01	00	1.85	177	29.2 25.4	888 848
306	15	50	42.1 39.2	40.5 35.3	0.24 0.37	00	1.91	195	30.1 26.3	892 850
306	16	5	42.2 39.2	39.3 34.2	0.42 0.90	00	1.84	181	29.2 25.4	890 850
306	16	20	42.2 39.2	38.0 32.7	0.38 0.80	00	1.79	159	28.3 24.4	888 850
306	16	35	42.2 39.2	37.7 31.1	0.98 1.91	00	1.82	123	28.0 23.2	890 850
306	16	50	42.2 39.2	38.7 33.4	0.65 1.32	00	1.83	167	28.8 24.8	890 849
306	17	5	42.2 39.2	39.2 34.3	0.51 1.03	00	1.84	183	29.1 25.5	890 850
306	17	20	42.3 39.2	37.4 33.1	0.64 1.39	00	1.72	178	27.8 24.7	885 848
306	17	35	42.2 39.2	38.1 32.5	0.80 1.48	00	1.80	154	28.3 24.2	888 849
306	17	50	42.1 39.2	38.0 31.6	0.17 0.34	00	1.83	130	28.2 23.5	891 849
306	18	5	42.2 39.2	37.6 31.8	0.67 1.45	00	1.78	142	27.9 23.7	888 848
306	18	20	42.3 39.3	36.6 31.3	0.82 1.70	00	1.72	142	27.2 23.3	885 847
306	18	35	42.2 39.2	36.9 29.8	0.30 0.56	00	1.80	99	27.4 22.2	888 848
306	18	50	42.2 39.2	37.8 32.3	0.55 1.06	00	1.79	151	28.1 24.0	887 848
306	19	5	42.2 39.2	39.0 34.1	0.39 0.71	00	1.83	181	29.0 25.3	888 848
306	19	20	42.2 39.3	39.3 33.0	0.64 1.29	00	1.89	150	29.2 24.6	890 847

Fig. 2. Sample output from a computer program that reduces raw data

18-MONTH UNCORRECTED (WET) 31-GHz BRIGHTNESS TEMPERATURE HISTOGRAM

	>140	100	:***		
137	138	139	140	12	:
133	134	135	136	12	:
129	130	131	132	10	:
125	126	127	128	18	:
121	122	123	124	19	:
117	118	119	120	20	:
113	114	115	116	29	:*
109	110	111	112	26	:
105	106	107	108	29	:*
101	102	103	104	42	:*
97	98	99	100	46	:*
93	94	95	96	51	:*
89	90	91	92	62	:**
85	86	87	88	77	:**
81	82	83	84	64	:**
77	78	79	80	82	:**
73	74	75	76	119	:****
69	70	71	72	112	:****
65	66	67	68	139	:*****
61	62	63	64	161	:*****
	58	59	60	148	:*****
	55	56	57	212	:*****
	52	53	54	170	:*****
	49	50	51	234	:*****
	46	47	48	242	:*****
	43	44	45	283	:*****
	40	41	42	359	:*****
		38	39	256	:*****
		36	37	272	:*****
		34	35	311	:*****
		32	33	410	:*****
		30	31	414	:*****
		28	29	547	:*****
		26	27	747	:*****
		24	25	1248	:*****
		22	23	1988	:*****
		20	21	2941	:*****
			19	1696	:*****
			18	1810	:*****
			17	2210	:*****
			16	2508	:*****
			15	2311	:*****
			14	2221	:*****
			13	2047	:*****
			12	1541	:*****
			11	732	:*****
			10	183	:*****
			<10	159	:*****

TOTAL # = 29430

NOTE: Above temperatures include 3°K cosmic background

Fig. 3. Histogram of zenith-normalized 31-GHz brightness temperature for an 18-month period. Data are not corrected for water-on-radome effects.

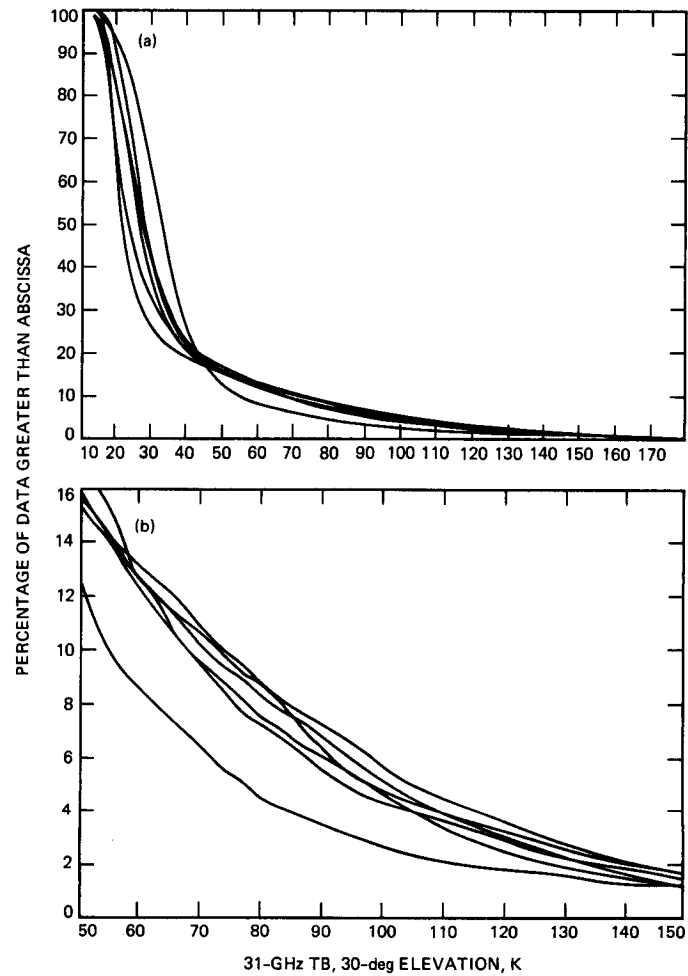


Fig. 4. Exceedance plots for 3-month groupings of 18 months' worth of 31-GHz zenith-normalized sky brightness temperature. Data are not corrected for water-on-radome effects.

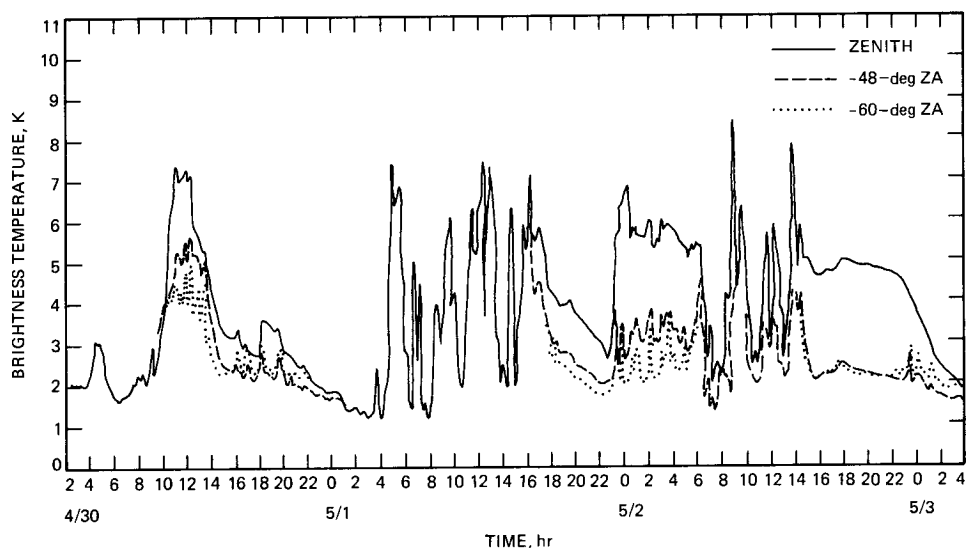


Fig. 5. Antenna temperature versus time for the elevation angles -60 degrees, -48 degrees, and zenith. The off-zenith data are normalized to zenith and are approximately the same when the radome is dry. Differences between these traces exist when water is on the radome.

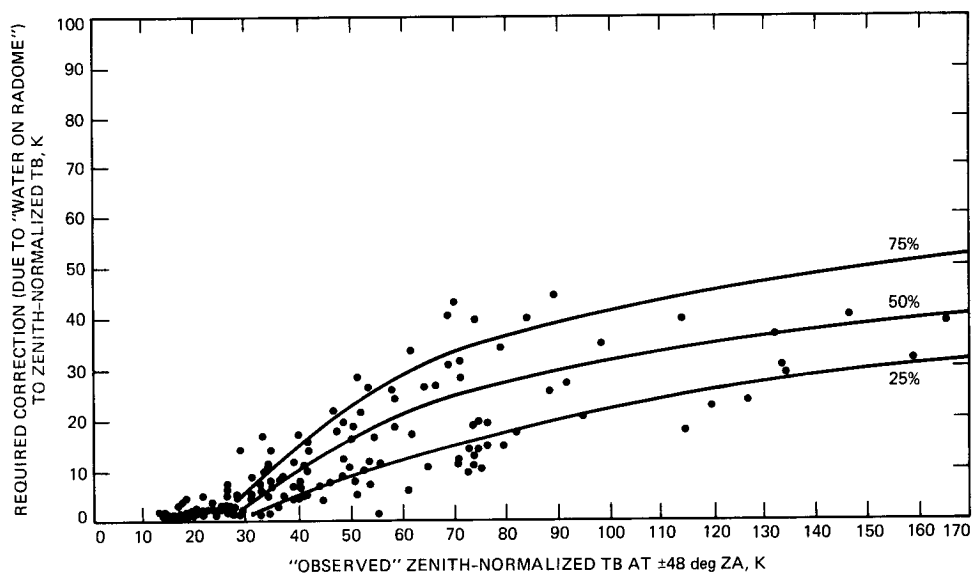


Fig. 6. An empirical relationship between observed zenith-normalized sky brightness temperature (X-coordinate) and the component of zenith-normalized sky brightness temperature that is due to water on the radome (Y-coordinate). Lines are hand-fit percentile estimators.

18-MONTH CORRECTED (DRY RADOME) 31-GHz BRIGHTNESS TEMPERATURE HISTOGRAM

	>140	24	:
137	138	139	140
133	134	135	136
129	130	131	132
125	126	127	128
121	122	123	124
117	118	119	120
113	114	115	116
109	110	111	112
105	106	107	108
101	102	103	104
97	98	99	100
93	94	95	96
89	90	91	92
85	86	87	88
81	82	83	84
77	78	79	80
73	74	75	76
69	70	71	72
65	66	67	68
61	62	63	64
	58	59	60
	55	56	57
	52	53	54
	49	50	51
	46	47	48
	43	44	45
	40	41	42
		38	39
		36	37
		34	35
		32	33
		30	31
		28	29
		26	27
		24	25
		22	23
		20	21
			19
			18
			17
			16
			15
			14
			13
			12
			11
			10
			<10

TOTAL # = 28299

NOTE: Above temperatures include 3°K cosmic background

Fig. 7. Histogram of wet-radome-corrected zenith sky brightness temperature for an 18-month period

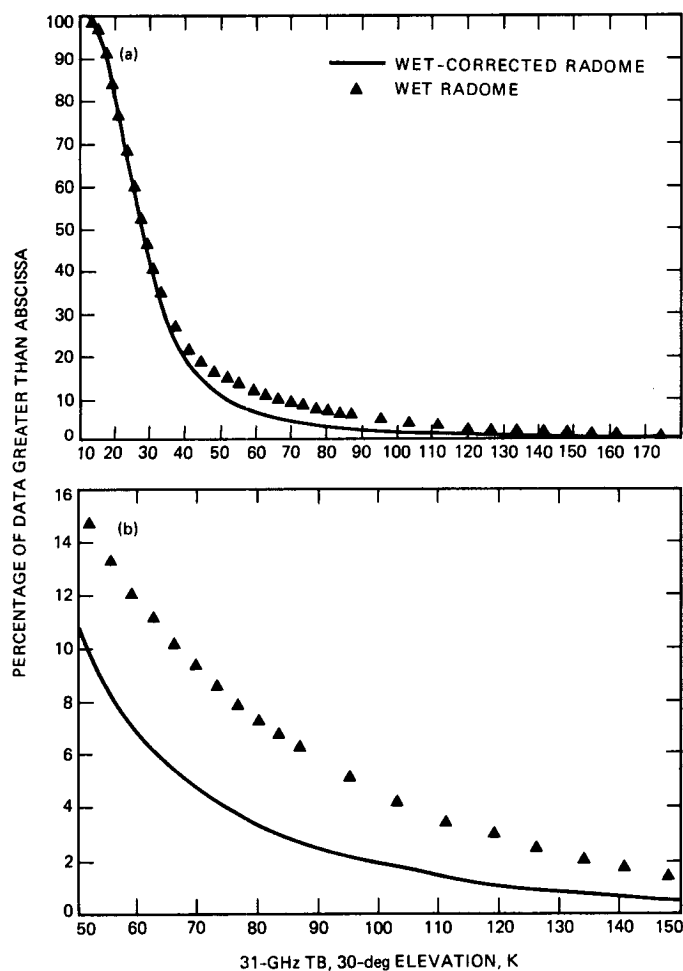


Fig. 8. The 31-GHz exceedance plots for 30-degree elevation for wet-radome data (Δ) and wet-radome-corrected data (lines)

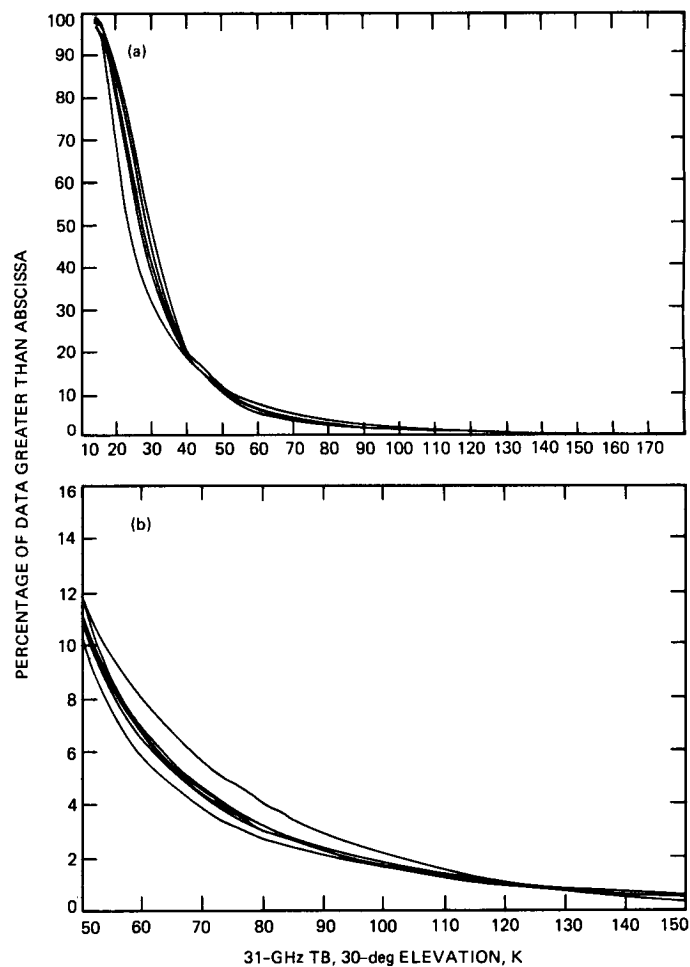


Fig. 9. The 31-GHz exceedance plots for 30-degree elevation for wet-radome-corrected-only data, grouped by season

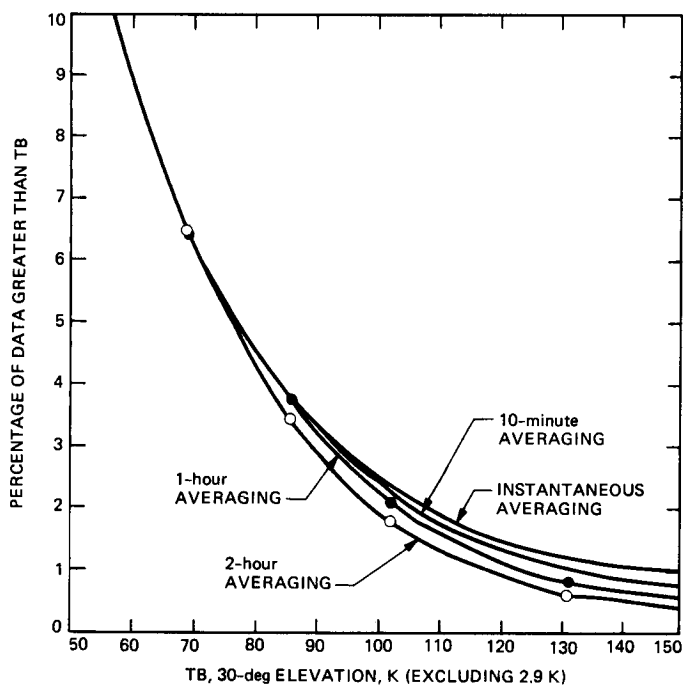


Fig. 10. Spain exceedance statistics [2]. Wet radome effects have been removed. Data are for a 4-month period in 1984 and are based on measurements from DSS 63.

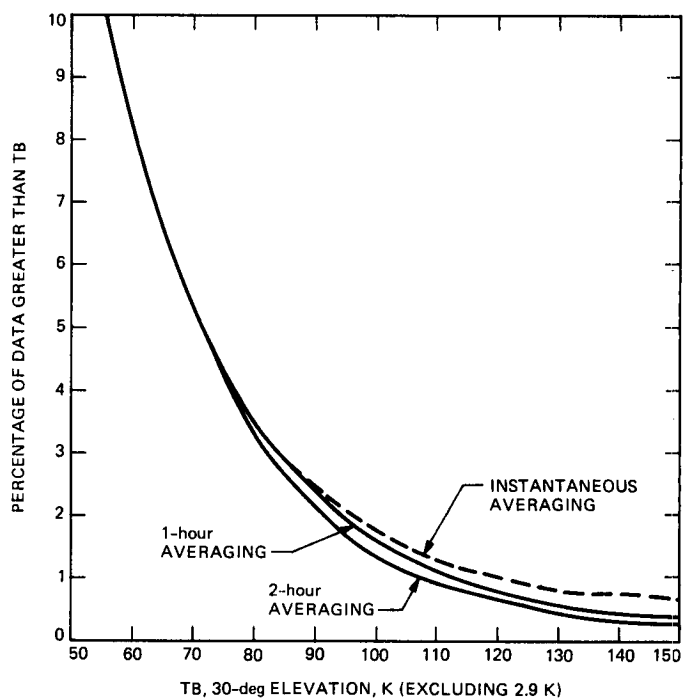


Fig. 11. Spain exceedance statistics with an adjustment made to correct for the larger-than-typical rainfall that occurred during the 4-month observing period in Spain (based on SCAMS WVR measurements)

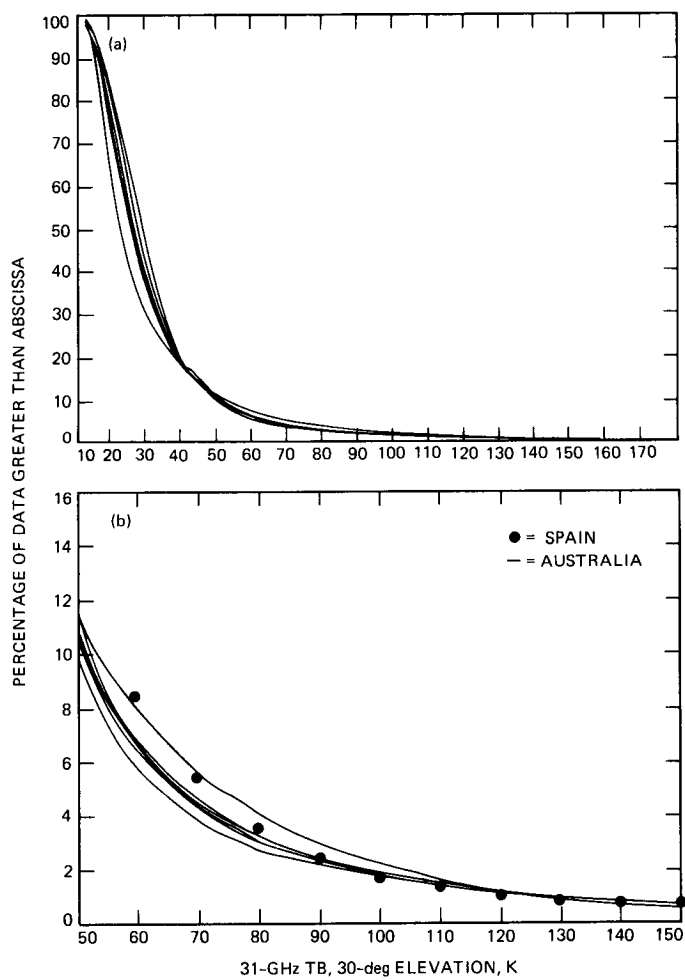


Fig. 12. Combination of Fig. 9 and Fig. 11 showing the similarity in exceedance plots for Spain and Australia

1989000821
617982
11p.
N89 - 10192
56-46
165617
April-June 1988
118.
JPL

Gravitational Wave Searches Using the DSN

S. J. Nelson

Documentation Section

J. W. Armstrong

Radio Frequency and Microwave Subsystems Section

The DSN doppler spacecraft link is currently the only method available for broadband gravitational wave searches in the 10^{-2} - to 10^{-4} -Hz frequency range. This report describes the DSN's role in the worldwide search for gravitational waves by first summarizing from the literature current theoretical estimates of gravitational wave strengths and time scales from various astrophysical sources. Current and future detection schemes for ground-based and space-based detectors are then discussed. Past, present, and future planned or proposed gravitational wave experiments using DSN doppler tracking are described. Lastly, some major technical challenges to improved gravitational wave sensitivities using the DSN are discussed.

I. Introduction

Gravitational waves are polarized gravitational fields that have escaped their sources and propagate independently [1]. These "ripples" in the curvature of space-time carry energy and momentum and move at finite speed. All relativistic theories of gravity agree on the existence of these waves, although the theories may differ in number of polarization states, propagation speed, efficiency of wave generation, etc. In General Relativity, gravitational waves are transverse, have two polarization states, and propagate at the speed of light. As a wave passes through space, it causes a strain—a dimensionless fractional change in the distances between massive objects and a similar fractional change in the rates at which separated clocks keep time. Figure 1 shows schematically the effect of a gravitational wave passing perpendicularly through a mass

(exaggerated by many powers of 10) according to General Relativity theory.

The search for gravitational waves has been going on since the 1960s using several detection approaches. In principle, gravitational waves can be detected through their effect on separated masses or clocks. All gravitational wave detectors attempt to sense some manifestation of these deviations in test mass separation (see Fig. 1 and Section III). However, gravitational waves are extremely weak, and they must be detected in a noisy environment; thus detection poses many experimental problems and provides profound challenges to instrument builders. Although gravitational waves are generated whenever there is time-varying, asymmetrical motion of any matter, they are generated at *appreciable* levels only by

extremely massive objects undergoing *extremely* violent dynamics. In other words, “strong” gravitational waves can be generated only by astrophysical sources. Because gravitational waves are so weak, Einstein—whose theory of General Relativity includes gravitational waves as a natural consequence—speculated that they could never be detected [2]. To date he is still right, although the search is gaining momentum worldwide, with ground- and space-based experiments being carried out or planned in Europe, Japan, China, the Soviet Union, Australia, and the United States.

There are many potential rewards of verified gravitational wave detection. Gravitational waves offer tests of fundamental physical law that may not be possible in any other way [3]. For example, comparison of arrival times of light and gravitational waves from, say, a supernova would test General Relativity’s prediction that light and gravitational waves travel at the same speed. Polarization measurements could verify or refute the prediction of General Relativity that the waves are transverse and characterized by traceless tensor waves. Also, detailed waveform analysis might well be the first unequivocal proof of the existence of black holes and General Relativity’s predictions of their dynamic behavior. Additional benefits of gravitational wave detection could include very accurate measurements of the Hubble constant and investigation of the dynamics of the Big Bang.

Another important consequence of a verified detection will be its implications for observational astronomy. Because gravitational waves interact very weakly with matter, they propagate unchanged from their sources. Thus detailed information about the time evolution of the source during violent events is preserved. This contrasts with electromagnetic waves, which can be absorbed or scattered by intervening matter. Even neutrinos produced in supernovae are scattered many times while leaving their sources. Relativistic motion of bulk matter and strong gravitational fields are central to most theoretical views about violent activity in supernovae, galactic nuclei, and quasars. When gravitational waves from these objects are detected, we will have the first observations of the interiors of strong-gravity, high-velocity regions. A new window will have opened for observational astronomy.

II. Gravitational Wave Sources and Anticipated Strengths and Frequencies

A. Gravitational Wave Sources

It is customary to classify gravitational waves and their sources based on their temporal waveforms [3].

Bursts (waves that are “on” for a few cycles) are expected to be produced by collisions of stars or black holes, the col-

lapse of a supernova to form a neutron star, the collapse of a star or star cluster to form a black hole, the coalescence of binary stars or black holes, and the fall of stars or small black holes into supermassive black holes.

Periodic waves, the prototypes of which are sinusoids, are expected to be produced by rotating neutron stars, binary stars, and binary black holes.

Stochastic waves (random fluctuations of long duration compared with observing times) are expected to be produced by ensembles of radiating binary stars, deaths of pregalactic massive stars, vibrations of cosmic strings, and the Big Bang.

Gravitational waves should be produced on a variety of time scales, depending on the masses of the objects involved. In general, stronger gravitational waves are expected to be produced on longer time scales. Figure 2 shows schematically some of the predicted waveforms from varying astrophysical events. Figure 3, adapted from [3], shows estimates of wave amplitudes at Earth as a function of the characteristic frequency of the waves.

B. Frequency Ranges

To detect gravitational waves in the full frequency range over which they are expected to occur, both ground-based and space-based technologies must be employed. Detector sensitivity depends on a number of factors, including test-mass separation and competing noise strengths. At present, doppler tracking of interplanetary spacecraft offers the only method for a broadband search in the 10^{-2} - to 10^{-4} -Hz frequency range. Ground-based detectors achieve good sensitivity at frequencies above about 10 Hz. (Below about 10 Hz, it becomes prohibitively difficult to isolate ground-based detectors from seismic and other noise.)

III. Gravitational Wave Detectors

A. Ground-Based Detectors

For very short period waves—on the order of 1/1000 second (e.g., from a supernova)—two ground-based technologies are currently employed: resonant bars and laser interferometers. Figure 4 shows schematically the two types of ground-based detectors.

Bar technology was pioneered in the 1960s by Joseph Weber of the University of Maryland. Gravitational waves excite mechanical oscillations in a cryogenically cooled bar which are read out by a transducer/amplifier system. Current-generation bars, like the one at Stanford University, now achieve sensitivities of better than 10^{-17} in strain amplitude.

This excellent sensitivity is comparable to the expected strain from a 1 percent efficient supernova collapse in our galaxy [3] and represents the best sensitivity achieved by any detector technology to date.

Laser interferometer detectors consist of three freely suspended test masses arranged at the corners of a right-angled "L." A gravitational wave incident on this system will "push" the test masses in one arm and "pull" the test masses in the other arm. This displacement will be reversed on the next half cycle of the wave (see Fig. 4). A laser light beam is split with a beam splitter, sent down both arms, bounced off mirrors on the test masses, and then recombined. The relative motions caused by a gravitational wave are detected as a fringe shift in the recombined light. This system is inherently broadband; many Fourier components are measured, and thus the gravitational waveform can be reconstructed. The prototype laser interferometers, such as Caltech's 40-meter system [4], currently have sensitivities comparable to those of the cryogenically cooled bars. A joint Caltech/MIT collaboration has been proposed to the National Science Foundation for construction of twin 4-km arm detectors that would yield more than three orders of magnitude better sensitivity [5]. The proposed joint facility has been named the Laser Interferometer Gravitational Wave Observatory (LIGO).

B. Space-Based Detectors

The current generation of space-based detectors uses the Earth and a distant spacecraft as free test masses. The DSN doppler tracking system, driven by an ultrastable time base on the ground, monitors a two-way coherent transponded microwave link with a distant spacecraft. In doing so, it continuously measures the relative dimensionless velocity ($\Delta v/c$) between the Earth and the spacecraft. A gravitational wave passing through the solar system produces strain ($\Delta l/l$) on the Earth-spacecraft system and similarly shifts the frequency of the clock driving the system. The result is that the gravitational waveform is replicated three times in the doppler tracking time series: once when the wave "buffets" the Earth, causing a small change in the difference between the transmitted and received doppler frequencies; once when the spacecraft is buffeted by the wave, causing the transponded signal to be different from the transmitted signal; and once when the initial Earth perturbation is transponded back to the Earth at "two-way light time" after the initial pulse (see Fig. 5). This three-pulse signal is an important signature of a gravitational wave and allows discrimination against noise sources in the doppler measurement system [6].

The sensitivity of a spacecraft-tracking gravitational wave search is limited by a variety of noise sources. The leading noise source for the current-generation (S-band radio link)

experiments is plasma scintillation noise. In the next generation of experiments (X-band radio link), the plasma noise will be reduced by more than an order of magnitude. At that point, very careful attention to other noise sources will be required. Instrumental contributions to the noise include, for example, station timekeeping, frequency-time distribution within the station, antenna mechanical changes with time, and spacecraft transponder instability. Tropospheric phase scintillations are also expected to be an important noise source for X-band experiments.

Other space-based detectors have been suggested that do not involve doppler tracking. These include orbiting optical interferometers that are conceptually similar to ground-based interferometers. See [3] for details of gravitational wave space-based experiments that are currently under consideration.

IV. Past, Current, and Future Low-Frequency Gravitational Wave Experiments

The first attempt to use spacecraft doppler data to search for gravitational waves was conducted at JPL by A. J. Anderson in 1971. Since that time, the DSN has obtained data from the Pioneer, Viking, and Voyager spacecraft that have been used to characterize noise sources, gain insight into experimental problems, and put upper limits on the strengths of the waves in the "low-frequency" band [7]–[11]. Both the Ulysses and Galileo missions have approved gravitational wave experiments and will be able to do experiments with better sensitivity than the current-generation prototype searches. Pulsar timing and precision celestial mechanics data allow searches for waves in the "very-low-frequency" band [11]–[13]. Finally, observation of excitations of the Earth, Moon, and Sun can also be used in principle [14]. Table 1 lists the published sensitivities of searches in the low-frequency and very-low-frequency bands.

V. Technology Needed for Future Experiments

As described previously, the current leading noise source in low-frequency experiments is plasma scintillation noise, i.e., random phase variations of the radio wave imposed as the signal propagates through the ionosphere and the solar wind plasma. Because of these variations, the apparent electrical distance to the spacecraft varies irregularly with time, acting as a noise source for gravitational wave detection. The magnitude of the effect depends inversely on the radio frequency of the tracking link squared; thus, higher frequency radio links are much less affected. X-band uplinks will produce more than an order of magnitude improvement in sensitivity over S-band experiments. The first of the X-band uplink observations will be made with the Galileo spacecraft during the

cruise to Jupiter, with an expected sensitivity of approximately 5×10^{-15} for bursts and approximately 10^{-16} for periodic waves (cf. Fig. 3). The Mariner Mark II and Mars Observer spacecraft will also have X-band uplinks and will be appropriate for use in gravitational wave searches. Observations with very-high-frequency links such as K-band (31,000 MHz), or those with multiple-frequency links, can provide very high immunity to plasma noise and very sensitive gravitational wave experiments (about 10^{-17} for K-band). Improved time-keeping and precision tropospheric monitoring would also be necessary to fully exploit the benefits of higher frequencies.

The remarkably good sensitivities of current and near-future experiments are due to the explosion in precision measurement technology that has occurred over the past few decades. For both space- and ground-based detectors, a continued push of the state of the art in relevant advanced instrumentation is required for further progress. For the doppler tracking link, technological improvements that would increase sensitivities include:

- (1) *Higher-precision timekeeping.* Timekeeping is at the heart of doppler gravitational wave experiments

(10^{-16} /1000-sec accuracy is currently the 1990 time-keeping goal for the DSN).

- (2) *Higher RF links and dual-frequency uplinks.* Higher-frequency links (e.g., K-band) or dual-frequency uplinks and downlinks can substantially reduce plasma scintillation noise, currently the leading noise source for gravitational wave experiments.
- (3) *Tropospheric monitoring technology.* The Earth's neutral atmosphere produces tropospheric scintillations that are at microwave frequencies independent of the radio frequency of the doppler link. As the plasma noise is reduced by higher-frequency links, tropospheric scintillation will become a leading noise source. Accurate monitoring of tropospheric scintillation with, for example, water vapor radiometry will be required to correct gravitational wave data for this effect.
- (4) *Further transponder development.* Instrumental stability on the spacecraft is also central to a high-sensitivity doppler tracking experiment.

Acknowledgment

The remarkable sensitivity of current and planned gravitational wave searches using doppler tracking is the result of the collective efforts of hundreds of people associated with the DSN, whose contributions are gratefully acknowledged. Parts of this article were adapted from a NASA presentation, "Relativistic Gravitation: NASA's Present and Future Programs," prepared by J. Anderson, J. Armstrong, F. Estabrook, R. Hellings, and H. Wahlquist.

References

- [1] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation*, New York: W. H. Freeman, 1973.
- [2] M. Bartusiak, "Sensing the Ripples in Space-Time," *Science* 85, vol. 6, no. 3, pp. 58-65, April 1985.
- [3] K. S. Thorne, "Gravitational Radiation," in S. W. Hawking and W. Israel (eds.), *Three Hundred Years of Gravitation*, Cambridge, England: Cambridge University Press, 1987.
- [4] R. W. P. Drever, "The Search for Gravitational Waves," *Engineering and Science*, vol. 66, no. 3, pp. 6-9 and 24-26, January 1983.
- [5] R. W. P. Drever, "Laser Interferometer Gravitational Wave Observatory," Caltech Public Information Office, February 3, 1986.
- [6] F. B. Estabrook and H. D. Wahlquist, "Response of Doppler Spacecraft Tracking to Gravitational Radiation," *GRG*, vol. 6, pp. 439-447, 1975.
- [7] R. W. Hellings, P. S. Callahan, and J. D. Anderson, "Spacecraft-Doppler Gravity-Wave Detection," *Phys. Rev.*, vol. D23, pp. 844-857, 1981.
- [8] J. W. Armstrong, R. Woo, and F. B. Estabrook, "Interplanetary Phase Scintillation and the Search for Very Low Frequency Gravitational Radiation," *Ap. J.*, vol. 230, pp. 570-574, 1979.
- [9] J. D. Anderson et al., "Pioneer 10 Search for Gravitational Waves—No Evidence for Coherent Radiation from Geminga," *Nature*, vol. 308, pp. 158-160, 1984.
- [10] J. W. Armstrong, F. B. Estabrook, and H. D. Wahlquist, "A Search for Sinusoidal Gravitational Radiation in the Period Range 30-2000 Seconds," *Ap. J.*, vol. 318, pp. 536-541, 1987.
- [11] R. W. Hellings and G. Downs, "Upper Limits on the Isotropic Gravitational Radiation Background from Pulsar Timing Analysis," *Ap. J. Lett.*, vol. 265, pp. L39-L42, 1983.
- [12] M. M. Davis, J. H. Taylor, J. M. Weisberg, and D. C. Backer, "High-Precision Timing Observations of the Millisecond Pulsar PSR 1937 + 21," *Nature*, vol. 315, pp. 547-550, June 13, 1987.
- [13] J. D. Anderson and B. Mashoon, "Pioneer 10 Search for Gravitational Waves—Limits on a Possible Isotropic Cosmic Background of Radiation in the Microhertz Region," *Ap. J.*, vol. 290, pp. 445-448, 1985.
- [14] S. P. Boughn and J. R. Kuhn, "Limits on a Stochastic Gravitational Background from Observations of Terrestrial and Solar Oscillations," *Ap. J.*, vol. 286, pp. 387-391, 1984.

Table 1. Gravitational waves: experimental limits in the LF and VLF bands

Burst Sources				
("Raw" = Allan variance without matched filtering in the postprocessing.)				
Frequency, Hz	Strain amplitude	Data type (date)	Data duration	Reference or comment
$\sim 10^{-3}$	$< 3 \times 10^{-14}$	Voyager (1979)	2 days	[7]
$\sim 10^{-3}$	$< 6 \times 10^{-14}$	Viking (1977)	2 days	[8]
$\sim 10^{-3}$	$< 2 \times 10^{-13}$	Pioneer 10 (1981)	21 days	JPL group (in preparation)
10^{-2} to 5×10^{-4}	$< 2 \times 10^{-13}$	Pioneer 11 (1983)	3 days	JPL group (in preparation)
10^{-1} to 10^{-4}	$\sim 3 \times 10^{-14}$	Ulysses (1990+)	28 days/year	Predicted performance
10^{-1} to 10^{-4}	$\sim 3 \times 10^{-15}$	Galileo (1990+)	40 days/year	Predicted (raw) performance
Sinusoidal Sources				
Frequency, Hz	Amplitude (1σ)	Data type (date)	Comment	Reference or comment
1.04×10^{-4}	$< 2 \times 10^{-14}$	Pioneer 10 (1981)	Geminga	[9]
0.005 to 0.033	$< 1.5 \times 10^{-14}$	Pioneer 11 (1983)	broadband	[10]
10^{-1} to 10^{-4}	$\sim 5 \times 10^{-15}$	Ulysses (1990+)	28 days/year	Predicted performance
10^{-1} to 10^{-4}	$\sim 3 \times 10^{-16}$	Galileo (1990+)	40 days/year	Predicted performance
Stochastic Background				
(Amplitude expressed as energy density in gravitational waves in bandwidth equal to the center frequency, except for the normal mode data, normalized by the energy density required to close the universe.)				
Frequency, Hz	Energy density/ closure density	Data type (date)	Reference or comment	
$\sim 10^{-8}$	$< 1.4 \times 10^{-4}$	pulsars	[11]	
$\sim 10^{-7}$	$< 5 \times 10^{-4}$	pulsars	[12]	
$\sim 10^{-6}$	< 18	Pioneer 10 (1981)	[13]	
$\sim 4 \times 10^{-6}$	< 38	Pioneer 10	[13]	
$\sim 7 \times 10^{-5}$	< 260	Pioneer 10	[13]	
$\sim 4 \times 10^{-4}$	< 40	Voyager (1979)	[7]	
$\sim 3 \times 10^{-4}$	< 100	Sun's normal modes	[14]	
$\sim 3.1 \times 10^{-4}$	< 1	Earth's normal modes	[14]	
1.2×10^{-4}	< 9	Pioneer 11 (1983)	JPL group (in preparation)	
10^{-1} to 10^{-4}	~ 5	Ulysses (1990+)	28 days/year, predicted performance	
10^{-1} to 10^{-4}	~ 0.01	Galileo (1990+)	40 days/year, predicted performance	

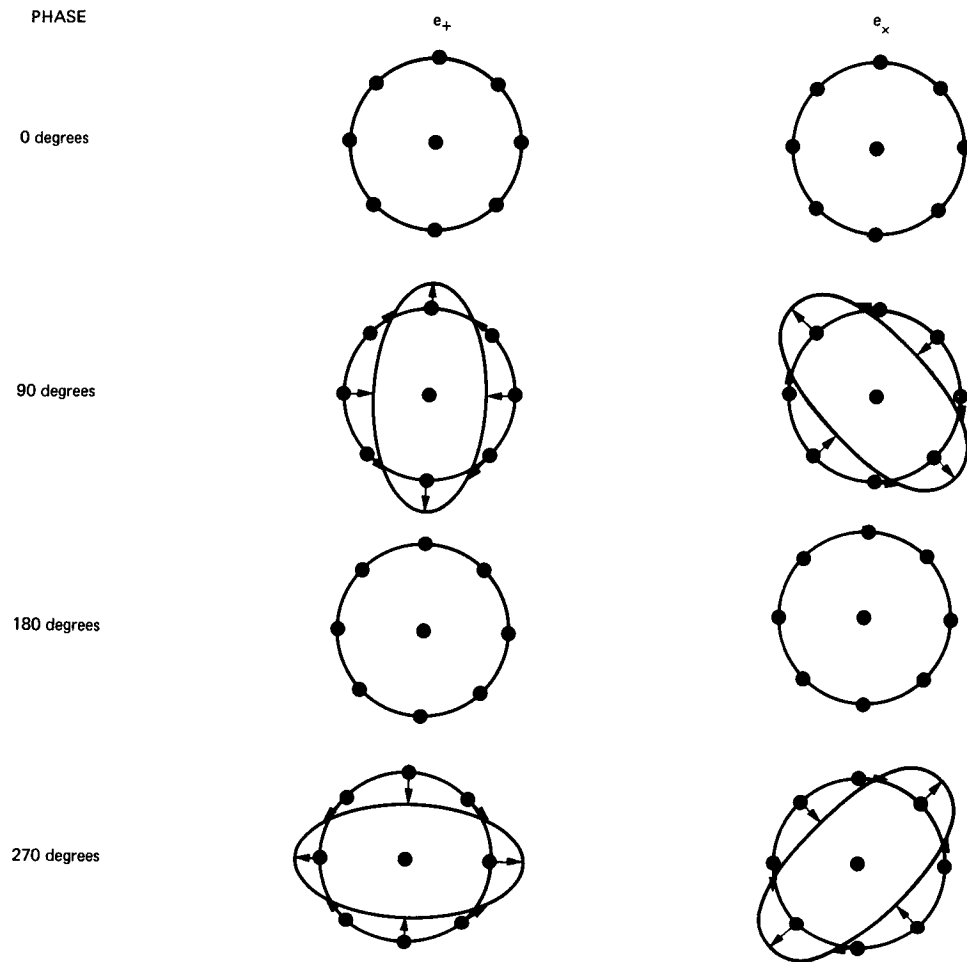


Fig. 1. A plane gravitational wave produces changes in the separation of test masses transverse to the direction of propagation. This figure, adapted from [1], shows schematically how an initially circular ring of test masses is distorted by the passage of waves in each of the two independent polarizations (“+” and “x” polarizations). Gravitational wave detectors attempt to sense some manifestation of these (very small) deviations in test particle separation.

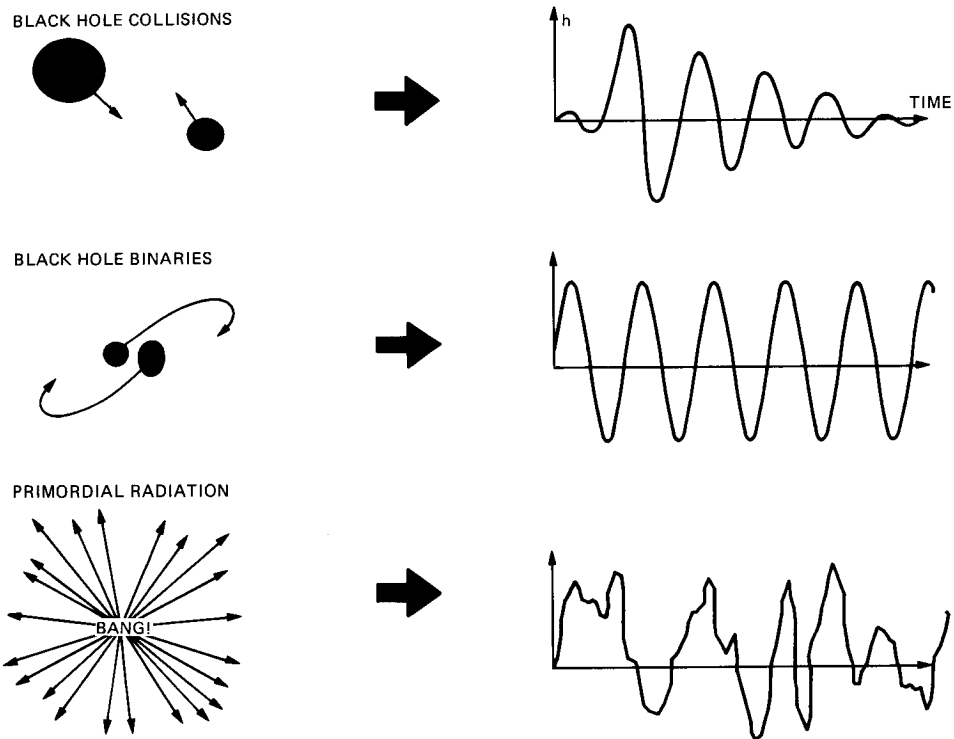


Fig. 2. Gravitational waves are conventionally classified by their temporal waveforms. Burst waves are “on” for at most a few cycles and might be produced by, for example, black hole collisions. Periodic waves, the prototype of which is the indicated sinusoid, could be generated by binary black holes. Stochastic waves, irregular variations in the wave strength on time scales that are long compared to an experimental time scale, might be generated by an ensemble of incoherent radiators or by the Big Bang.

ORIGINAL PAGE IS
OF POOR QUALITY

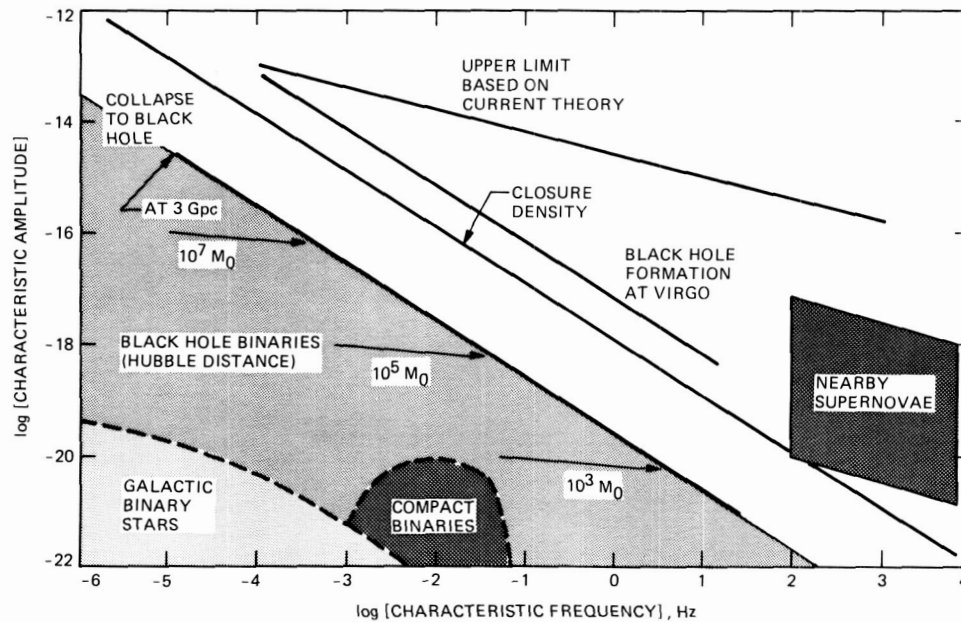


Fig. 3. Theoretical estimates of amplitudes and time scales for gravitational waves from some sources, adapted from [3]. It should be understood that these estimates are highly uncertain and many caveats apply; see [3]. "Upper limit based on current theory" is for the anticipated strength of burst radiation based on current understanding of the galaxy and general physical considerations. "Closure density" is the envelope of a family of curves having bandwidths equal to center frequency which represents the level of a gravitational wave background that would be sufficient to gravitationally close the universe. "Collapse to black hole" is the anticipated strength for burst radiation from black hole formation, assuming the mass of the hole (indicated) and assuming the source is at the Hubble distance (3 Gpc) or the Virgo cluster (where the amplitudes would be about 200 times larger). Shaded area labeled "black hole binaries (Hubble distance)" indicates amplitudes of periodic radiation from black hole binaries of indicated mass at 3 Gpc. The nearly horizontal arrows indicate evolutionary tracks. Amplitudes scale upward by about 200 if the source is in the Virgo cluster. "Galactic binary stars" shows the calculated strength of periodic waves from known binary stars in our galaxy. "Compact binaries" shows the calculated amplitudes for compact binaries, e.g., neutron star binaries, in our galaxy.

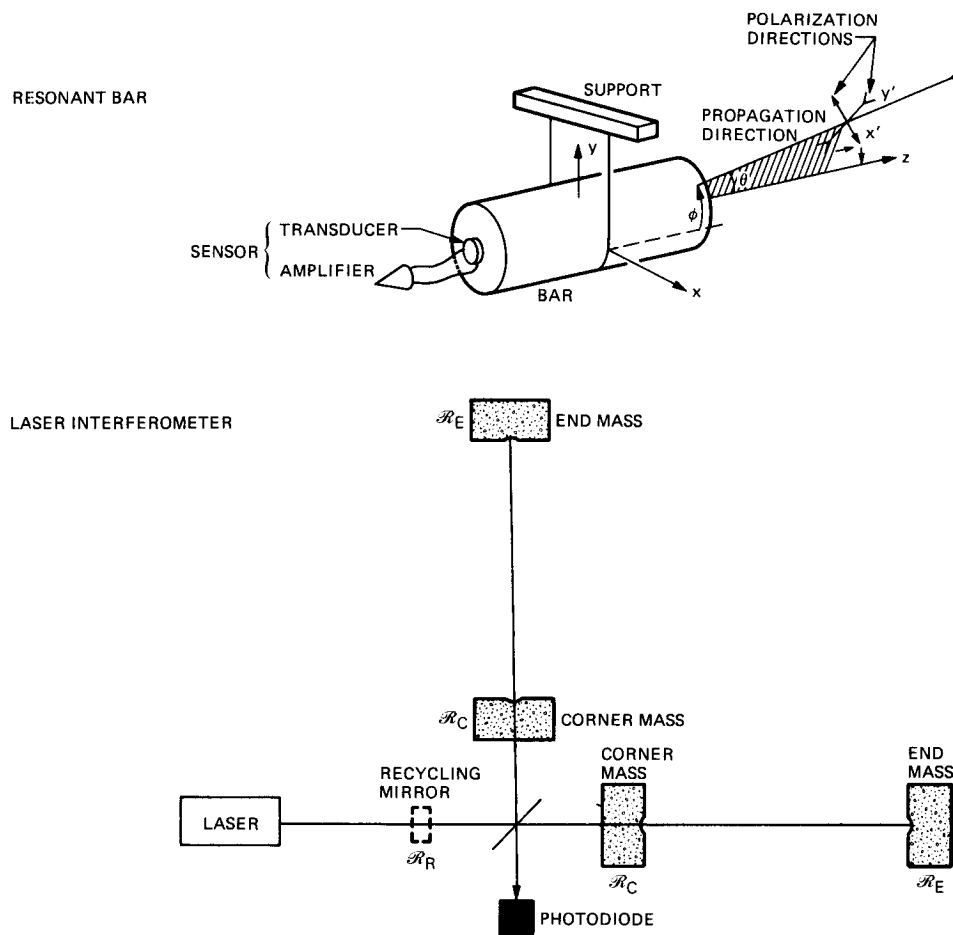


Fig. 4. Schematic diagrams of ground-based detectors—resonant bars and laser interferometers; see text for discussion.

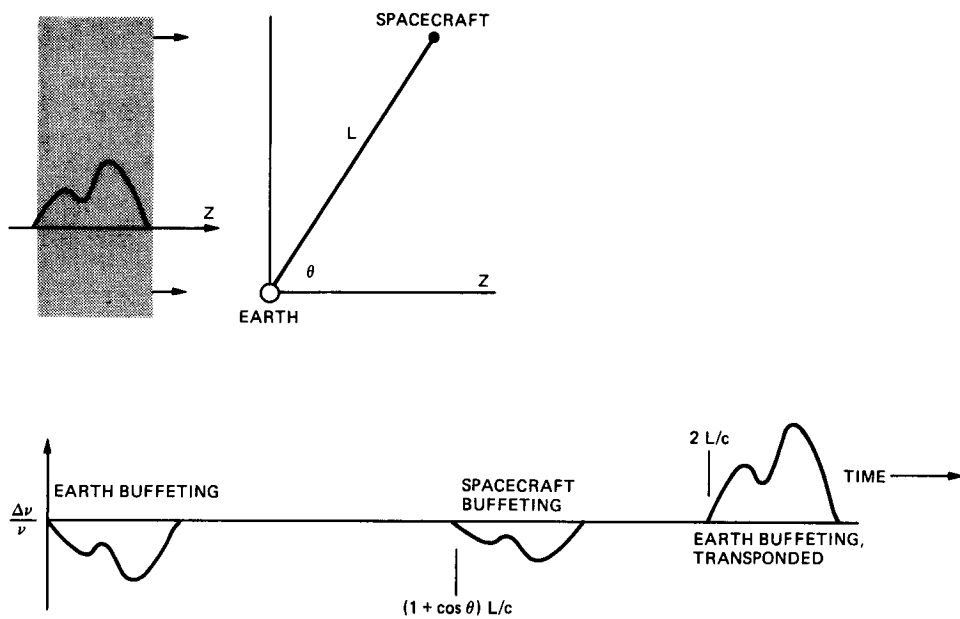


Fig. 5. Schematic diagram of spacecraft doppler tracking detector. Upper diagram shows a plane gravitational wave incident on the Earth-spacecraft system. Lower diagram shows "three-pulse" mapping of a gravitational waveform into the doppler tracking time series (see text and [6]).

57-33

165618

TDA Progress Report 42-94

7P.

1989030822
617781
7P.

N89 - 10193

April-June 1988

JPL

Evaluation of the Characteristics of a Field Emission Cathode for Use in a Mercury Ion Trap Frequency Standard

J. M. Christman

Communications Systems Research Section

This article reports on the performance of a field emission array characterized for the purpose of replacing the filament in a trapped ion frequency standard. This dark electron emitter eliminates the need for the interference filter currently used in the trapped ion standard. While reducing the filament's unwanted light, this filter causes a significant reduction in the signal. The magnetic field associated with the filament is also eliminated, thus potentially improving the present stability of the trapped ion standard. The operation of the filament in the present system is described, as well as the associated concerns. The cathode considered for the filament's replacement is then described along with the experimental system. Experimental results, observations, and conclusions are presented.

I. Introduction and Background

Stability requirements desired of modern frequency standards, in the range of one part in 10^{17} , are nearly two orders of magnitude beyond the capabilities of hydrogen masers currently employed in the DSN. Because of the increased potential of achieving higher frequency stability, extensive research has focused on the development of trapped ion frequency standards [1].

The inherent stability of atomic frequency standards is completely defined by two parameters: the signal-to-noise ratio (SNR) and the line Q. The SNR determines the efficiency of the atomic resonance measurement procedure. The

line Q represents the narrowness of the atomic resonant frequency.

Methods which increase the signal and suppress the noise are used to enhance the detection of the atomic resonance, increasing the SNR. The signal in the ion trap device is the fluorescence of the trapped ions, and the noise is due to stray light scattering from the optical pumping light and all other light sources. The principal source of noise is currently light emitted from the hot filament electron source. One way to improve the atomic resonance detection involves eliminating the interference filter required in the light detection system [1]. While suppressing unwanted light, this filter also signifi-

cantly reduces the signal. Removal of the interference filter is feasible only if the filament is replaced with a dark electron emitter.

Atoms with narrow hyperfine structures are used to improve the line Q. Energy levels are broadened as the magnetic field gradient at the position of the ion cloud is increased. Elimination of magnetic field gradients internal to the ion trap will result in an increased line Q. The electron current emitted by resistive devices is dependent on the applied current, which induces a magnetic field gradient. Replacement of the filament with a non-resistive electron emitter reduces the associated magnetic field.

This article reports on the performance of a field emission array characterized for the purpose of replacing the filament in a trapped ion frequency standard. The operation of the filament in the present system is described, as are the associated concerns. The cathode considered for the filament's replacement is described along with the experimental system. Experimental results, observations, and conclusions are also presented.

II. Present System

In the development of a mercury ion frequency standard, hot filaments are used in the vacuum chamber to ionize confined mercury atoms, which are then trapped and illuminated with ultraviolet and microwave radiation. During initial operation of the frequency standard, electrons are emitted into the trap for half a second and are then turned off for two and one-half seconds. Two power supplies are necessary to pulse the electrons. One provides the power through the filament, and the other is used to float the filament to a negative potential of 300 V, accelerating the electrons to a positively charged collector. Because the filament is a resistive element, the most effective method of pulsing the electrons is to switch the high-voltage supply on and off, leaving the filament supply constant.

The operating parameters of the filament are 2 V at 1.4 A. This continuous 2.8-watt level of power keeps the filament white hot. The bright filament makes an interference filter necessary in the optical collection portion of the vacuum system. While all filament light is efficiently blocked from the photon detection system, this filter also blocks 60 to 70 percent of the signal. The result is that twice as much collection time is required to obtain the equivalent signal-to-noise ratio with the filter in place than without.

The filament's continuous current also generates an unwanted magnetic field in the trap. Internal magnetic field gradients produce line broadening in the mercury ion hyperfine transition, thereby reducing the line Q.

An evaluation of the characteristics of a cold cathode was initiated to determine if these problems could be eliminated. Because the cathode is not a resistive element, it does not produce unwanted light and eliminates the need for the interference filter. As the maximum current between the base and the gate of the cathode is 200 μ A, four orders of magnitude less than the filament current, the resulting magnetic field would be greatly reduced. It is also possible to directly pulse the cathode, eliminating the magnetic field associated with the off time of the electrons.

III. Cathode Description

Field emission devices operate on the principle that applied electric fields at the surface of conductors cause electron emission. Efficiency of electron emission results from the choice of appropriate conducting materials and geometrically sharp conductor tips to achieve high electric fields. Limitations of these devices arise from the high potential necessary to generate these fields, typically 1000 to 10,000 volts. These high voltages generally cause current instability and limited device lifetime. Higher stability and longer lifetimes have been achieved by employing ultrahigh-vacuum environments, improving the sharpness of the pin used for emission, and decreasing the anode-emitting tip distance. Increasing the number of pins used for emission provides higher electron current, while reducing anode-emitting tip distance results in the achievement of the necessary high fields at much decreased applied voltages in the range of 50 to 200 volts.

The device used in this work was developed by Spindt et al. [2]. These field emission arrays are produced by a sandwich of conducting, insulating, and conducting materials. The manufacturing process consists of coating an oxidized silicon substrate with a molybdenum film, etching 10,000 holes through the various layers to the silicon, and forming molybdenum cones in the cavities. The bases of the cones are attached to the silicon substrate, and their tips lie in the plane of the molybdenum coating (Fig. 1). A negative potential applied between the molybdenum tips (base or cathode) and the molybdenum film (gate or anode) causes electron emission from the tips. A collector, positively biased with respect to the gate, attracts the electrons and allows for emission current measurements. The cathode is mounted on a TO-5 header with eight contact pins. Only two of these pins, the gate and the base, need exterior connections.

Manufacturer data indicate that emission current densities of 6600 mA/cm² are common with these cathodes, which have a packing density of 1.2E6 tips/cm². This density is strongly dependent on the geometry and biasing of the collector as well as on the proximity of the collector to the cathode. These arrays are capable of pulsed or continuous mode operation.

IV. Experimental Setup

For the initial turn-on of the cathode, it was necessary to achieve an ultrahigh-vacuum environment of about $1\text{E-}9$ torr. For this purpose, a stainless steel vacuum chamber was assembled using mechanical, turbo, and ion pumps. An electrical feedthrough served as the connection to the base and gate of the cathode and also provided four attachments to the isolated portions of the collector. The collector used was an actual stainless steel ion trap consisting of an extraction grid, two end caps, and a cylindrical ring electrode. The extraction grid and the end cap closest to the cathode are constructed of stainless steel mesh, while other components are solid stainless steel (Fig. 2). The mount for the electron emitter was secured to the collector, approximately 1 cm away. Because contamination causes cathode shorts, it was necessary to clean the collector thoroughly. Therefore, a tungsten filament was inserted into the vacuum chamber prior to the positioning of the cathode.

When the pressure was below $1\text{E-}9$ torr, the filament was operated and the collector was cleaned by electron bombardment. Before its use, the cathode is kept vacuum sealed to prevent destructive atmospheric contamination. Argon was introduced into the vacuum system to maintain the chamber's uncontaminated condition as the system was opened for the replacement of the filament with the cathode. The cathode is mounted in a standard transistor-type casing that is approximately the same size as the filament. Therefore, installing the cathode was merely a matter of removing the filament and inserting the cathode in its place, minimizing atmospheric contamination.

Two power supplies were used in series to extract the electron emission toward the collector held at ground potential. One provided power to the cathode, while the other established the cathode's negative biasing. A resistor was placed in series with the cathode in order to monitor the current provided by the power supplies. It is crucial to monitor the base-gate current of the cathode because destruction of the array can occur if this current is greater than $1\text{ }\mu\text{A}$. At high base-gate currents, the tips melt to the gate and short the cathode. Current from the various sections of the collector was combined and monitored using an electrometer. The current supplied by the power supplies is equal to the combination of the gate-base current and the emission current collected by the collector.

The base pressure of the vacuum system was $5\text{E-}10$ torr, lower than the manufacturer's suggested pressure of $1\text{E-}9$ torr for the initial turn-on of the cathode. The emission current was monitored as the voltage across the cathode was varied between 50 and 80 volts. The current between the base and

gate was determined by observing the total current supplied by the cathode's power supply and subtracting the collected emission current.

After the current-voltage relationship was determined, the cathode was isolated from the turbo and mechanical pumps by a valve. Helium was introduced into the vacuum system from an auxiliary port on the turbo pump to simulate the operating environment of the ion frequency standard. When the helium pressure had increased to $1\text{E-}5$ torr, the valve between the turbo pump and the cathode was opened. Current was again monitored to determine the effect of the helium pressure on the emission characteristics of the array.

The cathode has operated continuously for a four-week period. During the first week of operation, two shorts occurred from the base to the gate. This damage was reversed by the discharging of a $1\text{-}\mu\text{F}$ capacitor that was charged to 30 V. The array was again able to emit with no variation in the current.

V. Results

The emission characteristics of the cathode were determined by varying the applied voltage. The voltage was varied between 55 and 80 volts, and 15 values of the emission current were measured. The data were obtained after the emission current had stabilized. The emission current, the total current, and the base-gate current are plotted against the applied voltage in Fig. 3. Figure 4 shows the plot of the same data for the helium environment. As can be seen in these figures, 75 volts will produce more than $50\text{ }\mu\text{A}$ of collectible emission current in either environment. The electron current needed in an ion frequency standard is dependent on the trapping time of the standard, which corresponds to a few microamps in the present system. The desired level of emission collected in the simulated trap is $30\text{ }\mu\text{A}$. In both environments, the gate-base current was measured up to $20\text{ }\mu\text{A}$, much higher than the $1\text{ }\mu\text{A}$ suggested by the manufacturer.

Comparison of the current versus voltage graphs of the two environments indicates that the helium pressure does not affect the emitting characteristics of the cathode, confirming predictions based on the short-term effects of a background gas (private communication, SRI). The lifetime of an array under a substantial amount of background gas pressure has not been determined; yet the cathode has operated for a period of three weeks at a helium pressure of $1\text{E-}5$ torr without adverse effects on its performance.

Plots of $\log(I/V^2)$ versus $1000/V$ were generated to determine if the cathode followed the expected Fowler-Nordheim relationship of an emitting array [3]. The Fowler-Nordheim

plots of the current and voltage should produce a straight line if there is consistent geometry of the tips and the emitting area and if the work function is constant. The uniform geometry of the cathode depends on the manufacturing process. The degree of deviation of the work function from a constant value increases as the contamination level of the environment increases.

Figures 5 and 6 are the Fowler–Nordheim plots using the total current and the applied voltage of the cathode. For both the ultrahigh-vacuum (UHV) and the helium background pressure environments, the experimental points are graphed along with the best fit lines. The best fits were obtained using a least squares fit method. The slopes of the best fit lines are $-0.210 (\pm 0.005)$ and $-0.200 (\pm 0.005)$ for the UHV and helium environments, respectively. The slopes have a 2.5 percent experimental error margin, reflecting the precision of the measurement equipment. Within this margin the slopes have the same value, indicating that there is no appreciable effect on the emission current due to the helium pressure.

As indicated by comparison of the Fowler–Nordheim plots of the two environments, the cathode is adequately designed and manufactured for application in an ion frequency standard.

VI. Conclusion

An electron source has a vital function in an ion frequency standard. The existing filament offers sufficient electron current for the function of this standard but brings with it inherent disadvantages, including undesired light and a magnetic field. A novel method for producing electron current in an ion trap which functions without the production of light or a mag-

netic field has been presented in this article. An electron emission array has been placed in a simulated ion frequency standard environment to determine if it is suitable as a filament replacement.

The study reported here indicates the cathode's compatibility with the existing system. This device provides confident replacement of the filament in an ion frequency standard and offers many advantages.

The similar sizes of the cathode and the filament provide for the interchangeability of the two with minor modifications in the existing system. The amount of the cathode-collected emission current is sufficient for use as an electron source. Since no indication of adverse effects of helium on the performance of the cathode was found, the compatibility of the cathode with the present system is enhanced.

The cathode's ability to generate electrons without producing light allows for the elimination of the existing interference filter, thereby increasing the collected signal by a factor of two or three, which improves the SNR. The negligible amount of current required for cathode operation, together with the ability of the cathode to operate in an efficient pulsed mode, eliminates the associated magnetic field generated when the filament is utilized. Because of the dependence of the line Q on the magnetic field, improvements in the line Q follow the reduction of the magnetic field.

This article reported on an electron-emitting device which will increase the inherent stability of the ion frequency standard. The replacement of the filament with the emission array improves both of the stability-defining parameters, the SNR and the line Q.

Acknowledgments

Much appreciation is extended to Capp Spindt and his SRI co-workers for their suggestions and for the constant availability of their expertise. Gratitude is also expressed to Dr. L. Maleki for his technical advice and suggestions in the preparation of this article.

References

- [1] J. D. Prestage, G. J. Dick, and L. Maleki, "The JPL Trapped Mercury Ion Frequency Standard," *TDA Progress Report 42-92*, vol. October–December 1987, Jet Propulsion Laboratory, Pasadena, California, pp. 13–19, February 15, 1988.
- [2] C. A. Spindt, C. E. Holland, and R. D. Howell, "Field Emission Array Development for High-Current-Density Applications," *Applications of Surface Science*, vol. 16, pp. 268–276, 1983.
- [3] R. Gomer, *Field Emission and Field Ionization*, Cambridge, Massachusetts: Harvard University Press, 1961.

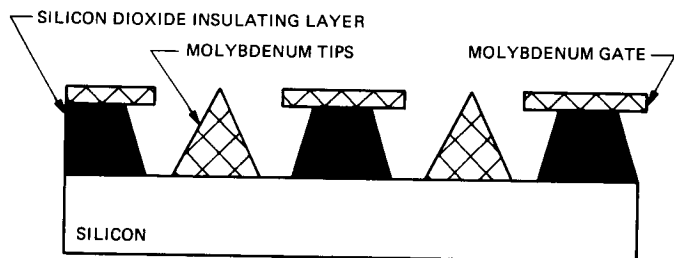


Fig. 1. Schematic diagram of a field emission array

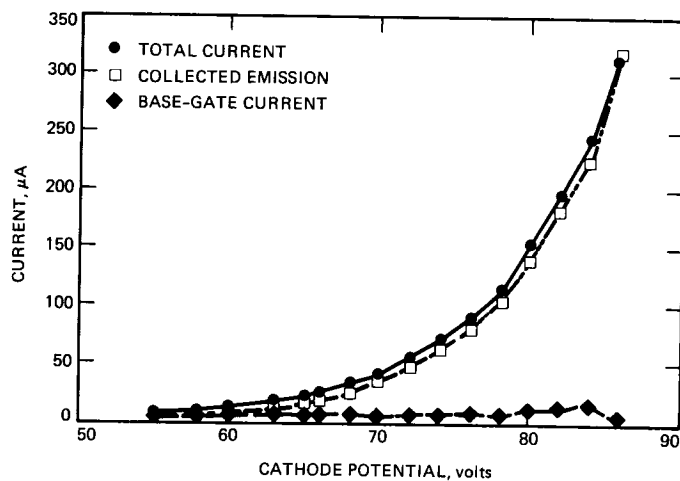


Fig. 3. Current versus voltage plot for the cathode with a background pressure of $1E-9$ torr

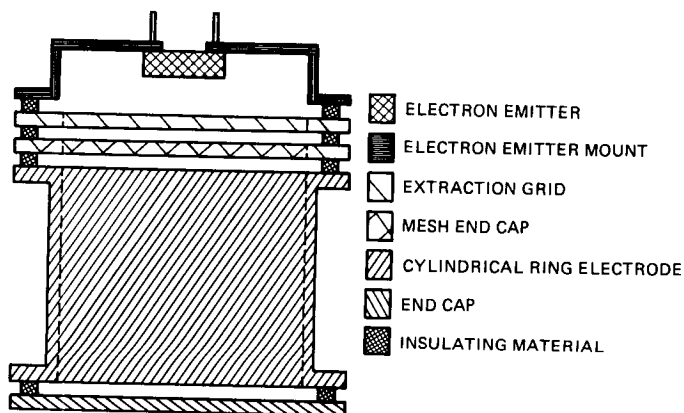


Fig. 2. Diagram of the electron emission collector; mount supports either a filament or a field emission array

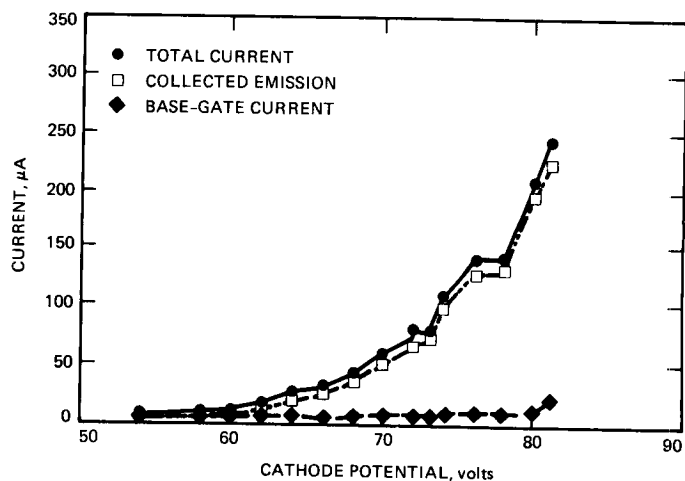


Fig. 4. Current versus voltage plot for the cathode with a helium background pressure of $1E-5$ torr

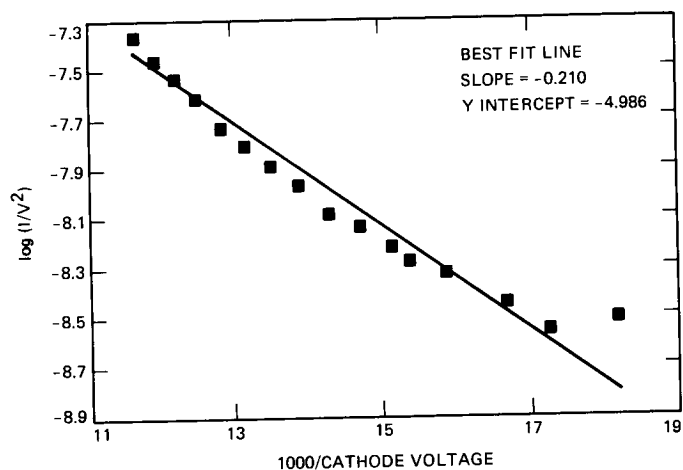


Fig. 5. Fowler-Nordheim plot of cathode total current and applied voltage (background pressure of 1E-9 torr)

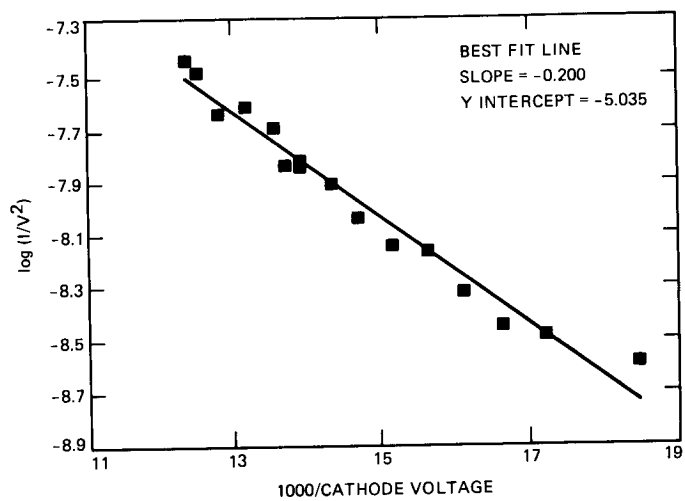


Fig. 6. Fowler-Nordheim plot of cathode total current and applied voltage (helium background pressure of 1E-5 torr)

1989000823
617990
10p.

S8-32

~~165679~~
158

April-June 1988

N89 - 10194

JPK

Simultaneous S- and X-Band Uplink-Downlink Performance at DSS 13

A. J. Freiley

Radio Frequency and Microwave Subsystems Section

The DSS 13 26-meter antenna with the second generation S/X feedcone was tested to determine the dual S- and X-band (2.1 to 2.3 GHz and 7.1 to 8.5 GHz, respectively) transmit and receive performance. Measurements were conducted using the 20-kW transmitters at S- and X-band while simultaneously receiving S- and X-band. This system proved to be very quiet compared with the other DSN antennas. Under normal tracking configurations, no noise burst or intermodulation product (IMP) activity was detectable to the -175-dBm level. To prove the instrumentation's ability to detect such phenomena, an IMP generator was introduced onto the system with positive, verifiable results. The IMP occurred at the -162-dBm level, accompanied by moderate noise burst activity, and was readily repeatable. The measurement also showed the possible need for additional fourth-channel filtering in the system to reduce the effect of the transmitter power on the LNAs.

I. Introduction

In support of the deep space missions, the Deep Space Network has been developing systems capable of providing multiple uplinks and downlinks from a single antenna. Since the mid-1970s, the DSN has experienced occasional difficulty with the high-power/low-noise systems that are used for deep space communications. These systems simultaneously transmit from 20 kW to 500 kW and receive signals at the -160- to -170-dBm level. These systems occasionally experience problems with low-level corona, or arcing, which causes a momentary increase in system temperature, a phenomenon called noise bursts. Considerable effort has been expended to improve or eliminate the problem [3]. With the development of a dual uplink capability, the problems of intermodulation products present additional challenges to the operation of high-power/low-noise systems.

Initially, the dual uplink was attempted by trying to modulate a single klystron transmitter with two carriers. The systems were plagued with noise bursts, as well as with a general steady-state increase in the system noise temperature and higher-order intermodulation products (IMPs) caused by the nonlinear mixing of the two carrier signals. In this early work, the carriers were only a few megahertz apart, so the mixing products were also a few megahertz apart throughout the receive spectrum. The IMP frequency is generally stable, but the amplitude fluctuates rapidly over a wide range [1], [2].

Noise abatement efforts in the DSN [1] have improved the performance of the antennas greatly over initial conditions [2]. The systems are currently able to transmit a single 400- to 500-kW carrier with little or no noise burst activity, provided good antenna housekeeping techniques are maintained. How-

ever, the IMP problems caused by the dual-carrier transmission have never been overcome.

Advanced Transmitter Development personnel tested the new common aperture feed at DSS 13, called the XSR, to demonstrate its ability to transmit and receive two widely spaced signals simultaneously. The goal was to measure the level of performance in terms of noise burst and IMP levels, and to identify the various noise and IMP contributors, both internal and external to the feed system. The frequency bands of primary interest are 2110 to 2120 and 2270 to 2300 MHz at S-band and 7145 to 7135 and 8400 to 8500 MHz at X-band. In each case, the lower frequencies are the transmit band. Noise bursts and IMP generation are much reduced as a result of the widely separated transmit signals used in this system.

II. XSR System

The DSS 13 antenna is a 26-meter az-el Cassegrainian center-feed reflector. The feedcone [3] houses the microwave feed components, the X-band transmitter power amplifier, and the receiver front ends. The feed provides simultaneous S- and X-band transmission and reception. The S-band transmitter power amplifier is located in the electronics room behind the main reflector. The system performance characteristics are given in Table 1. Figure 1 is a functional block diagram of the XSR system. The feedhorn is a common aperture multimode corrugated horn. The feedhorn is followed by the S/X combiner, which separates the S- and X-band signals. The S-band signals are fed in radially, while the X-band signals pass directly through, into another corrugated section below.

A. X-Band System

The X-band system consists of a quarter-wave plate polarizer surrounded by round rotary joints that provide switchable RCP or LCP polarization. Following the polarizer is a round to rectangular transition. Next is a 54-dB coupler, which is the point at which the test transmitter signal is injected into the X-band system. The next component is a rectangular waveguide switch that is used to switch the input of the low-noise amplifier (LNA) from the horn to the ambient load to provide the absolute temperature reference for the measurement of the antenna system noise temperature (T_{op}).

The X-band diplexer consists of three components. The first is the junction which combines the transmit and receive signals. In the transmit arm, a filter prevents the noise power in the receive band from reaching the LNA. In the receive arm, another filter prevents the transmit signal frequencies from reaching the LNA.

Continuing down the system along the receive path, the next component is the LNA input coupler, which is used to

inject test signals or the power from the noise diode assembly needed to measure T_{op} . Next is the cryogenically cooled X-band traveling wave maser. Its noise contribution to the system is 8 kelvins with an instantaneous bandwidth of 8 MHz; it is tunable through the range of 8400 to 8500 MHz. The amplified signal is then distributed or switched to one of three receivers. There are two modified Block III receivers, one of which can be configured to X-band. There is also an R&D receiver, which consists of a simple mixer/downconverter assembly, and a radar receiver used to condition the radar signals. The modified Block III receiver and the R&D receiver proved the most useful for supporting the measurements reported here. The output of each of the receivers is a standard 50-MHz signal that is distributed to various locations in the station.

The X-band transmitter portion of the system is driven by a synthesizer. The signal is multiplied by 336, amplified, and fed into the transmitter power amplifier. The RF signal passes through a harmonic filter to reduce levels of unwanted signals. Then the filtered signal is fed into the transmitter port of the diplexer and radiated out the horn.

B. S-Band System

The S-band signals emerge from the combiner into the labyrinth waveguide assembly. This assembly consists of four equal-length waveguide runs, two waveguide magic tees, and a short slot hybrid. This waveguide assembly functions as a polarizer. An S-band RCP signal is received into the horn and the energy is coupled into the waveguide. The labyrinth phases the four signals in such a way as to produce a linearly polarized fundamental mode in the rectangular waveguide feeding the first waveguide switch. This switch selects either the low-noise listen-only receive configuration or the somewhat higher noise diplexed configuration. For the purpose of the system tests, the diplexed configuration was required.

The S-band diplexer consists of a short slot hybrid to combine or separate the transmit and receive signals and a filter section of waveguide below cutoff to allow the receive signal to propagate through while attenuating the transmit signal (thus protecting the LNA from the high power). The receive signal is combined in another short slot hybrid and then propagates through another waveguide switch that selects between the diplexed configuration and the S-band ambient load. The next component is the WR430 X-band reject filter, which protects the S-band LNA from X-band transmitter power that may couple into the S-band waveguide system. The multiport LNA input coupler allows the injection of signal and noise diode power to calibrate the receive system. The S-band low-noise amplifier is a cryogenically cooled traveling wave maser that contributes 2.5 kelvins to the system. The amplified signal is distributed to the applicable receiver.

The S-band transmit signal is based on a synthesizer signal that is multiplied by 48 and fed into the transmitter power amplifier. The high-power RF signal then passes through a low-pass filter to reduce the levels at the receive frequencies, and through a harmonic filter to reduce the harmonic level to avoid interference in the X-band system. This filtered signal is then passed to the diplexer and out the horn.

III. Noise Instrumentation

The station noise instrumentation consists of three sets of equipment. The first is the conventional Y-factor instrumentation. The 50-MHz signal from the receiver passes through an AIL variable attenuator to the power meter detector. The power meter detector feeds a DC voltage to a stripchart recorder.

The second set of equipment is the noise-adding radiometer (NAR). The 50-MHz signal from the receiver is fed into the square-law detector, where the power is converted by a voltage-to-frequency converter. This proportional frequency is fed into a frequency-to-digital converter within the NAR to calculate the system temperature based on the strength of the noise diode power injected into the microwave system (see Fig. 1).

The third set of equipment is a subset of the NAR system. The 50 MHz are fed into the square law detector as in the previous set, but the detector also has a DC output that is proportional to the input power. This DC signal is displayed on a stripchart recorder. Through the use of the noise diodes, the stripchart can be calibrated in kelvins per inch.

Both the Y-factor instrumentation and the NAR instrumentation produced the same results when measuring the system temperature of the S-band and X-band systems, even though each set of instrumentation derives its results by a different method. This agreement indicates that the receiving system is linear. (The results of these measurements are given in Table 1.) However, each of these sets of instruments disrupts or masks the measurements of the noise burst activity and the IMP detection. The simplest instrumentation, the square-law detector and stripchart recorder, proved to be the best for monitoring the noise burst activity.

IV. IMP Predictions

The amplitude of the IMP signals is a quantity that is unpredictable with any accuracy. Previous measurements have encountered levels typically about -140 dBm for the order $N = 31$. For the measurements at hand, the order was much lower, and therefore one might expect the levels to be stronger

if all else were equal. The intermodulation product frequency (F_i) is predictable by the following equation:

$$F_i = mF_x + nF_s$$

where F_x is the frequency of the X-band transmitter signal, F_s is the frequency of the S-band transmitter signal, and m and n are the indication of the order of the intermodulation by the following equation:

$$\text{intermodulation order} = m + n$$

For the transmitter frequencies of 2110 and 7173 MHz, S-band and X-band respectively, and $m = 5$ and $n = 13$, the frequency of the IMP will be 8435 MHz.

V. Test Procedure

During the system tests, each session would generally follow the steps listed below. Some variations would result if some phenomenon occurred. Items 9 through 14 are some of the variations that were attempted.

- (1) Configure system.
- (2) Measure the S-band and X-band system noise temperatures.
- (3) Configure test signals for both S and X.
- (4) Calibrate test signal.
- (5) Measure or observe spectrum with and without signals to determine signal strength.
- (6) Radiate 20 kW from S-band transmitter only. Note any changes in the system.
- (7) Radiate 20 kW from X-band transmitter only. Note any changes in the system.
- (8) Simultaneously radiate 20 kW from each transmitter. Note any changes in the system.
- (9) Sweep S-band transmitter while X-band remains fixed. Note any changes.
- (10) Sweep X-band transmitter while S-band remains fixed.
- (11) Perform antenna tipping curve with both transmitters radiating.
- (12) Apply modulation signals.
- (13) Move the antenna in azimuth and/or elevation and apply brakes.
- (14) Install the "surefire" noise generator (described below).
- (15) Vary the amplitude of one transmitter or the other.

VI. Results

Each measurement session began by configuring the station to ensure consistency. The measurements of T_{op} , maser gain, and receiver follow-up temperature were the primary parameters used to determine if the system was configured consistently with previous sessions. The noise temperature was typically about 31 kelvins at S-band and 28 kelvins at X-band. The maser gain was set to 45 dB, and the follow-up temperature was usually under 0.1 kelvin for each receiving system, depending on the receiver used.

Early in the sessions, the S- and X-band R&D receivers were used with the spectrum analyzer installed in the maser instrumentation rack with no visible results such as noise burst or IMPs. Previous measurements as reported in [2] were visible with this instrumentation. No matter what was done, no noise bursts or IMPs were detected. The fourth harmonic of the S-band transmit signal was observed at 8440 MHz, which correlates with the 2110-MHz S-band transmitter signal. When the S-band frequency was tuned up 1 MHz, the harmonic would move up 4 MHz.

The next approach was to use the Block III receiver and slowly search through the range of frequencies where an IMP was predicted, based on the calculations, in an attempt to lock the receiver to an IMP and measure its amplitude and frequency. No positive results were obtained, although the fourth harmonic was easily detected. The test transmitter signal could also be easily detected. Much difficulty was encountered in maintaining a consistent system configuration for the duration of these tests. As is the nature of a research and development station such as DSS 13, equipment is changed or modified and the capabilities are changed over a period of time. Such was the case with the Block III receiver; some modified modules were installed, leaving the receiver unable to lock to a signal. Restoration of the capability could not be accomplished in a timely manner, so the Block III receiver testing was abandoned.

In the continuing search for some indication of noise burst and IMP activity, the R&D receiver was drafted back into service with the addition of the SETI digital spectrum analyzer [4] to increase the sensitivity of the receiving system. The digital spectrum analyzer bandwidth covers 20 MHz, divided into two windows of 10 MHz each. The analyzer divides that bandwidth into 2^{16} (65,536) channels, each channel representing a bandwidth of 305 Hz. The integration time for a single measurement is 3.3 msec. The total integration time is adjustable with input to the computer controlling the spectrum analyzer. With this system and about 6 seconds of integration time, the system threshold was -175 dBm. The spectrum display that provided the clearest indication of noise was developed by measuring the peak amplitude in each channel.

Figure 2 shows a sample of the spectrum from 8431 to 8441 MHz. Directly in the center is the "DC line" (an artifact of the instrument) and to the left of center is a signal that corresponds to the X-band test transmitter signal at 8433 MHz at an amplitude of approximately -162 dBm. During this test, there was no radiated RF power in the system. With each of the transmitters radiating 20 kW, S-band at 2110 MHz and X-band at 7173 MHz, the spectrum of Fig. 3 shows the addition of the fourth harmonic of the S-band at 8440 MHz (four times 2110). The amplitude was strong enough to saturate that channel of the spectrum analyzer, so no accurate amplitude measurement was obtained. During this period, no noise bursts were recorded and no IMPs detected. It was not the purpose of these tests to eliminate the fourth harmonic signal; should it prove troublesome in the future, it would yield to standard RFI prevention techniques.

In an attempt to find and measure some evidence of noise bursts and intermodulation activity, the "surefire" IMP generator was created. It consisted of a large aluminum pipe clamp, a rusty bearing and chain, aluminum shavings, and aluminum wire woven through the steel chain. This unusual collection of material was suspended from the quad leg about 4 feet above the dish. A rope was attached and fed down through the dish, allowing station personnel to mechanically disturb the IMP generator to increase its effectiveness in generating noise bursts and IMPs. Given the operating frequencies above, the IMP should have appeared at 8435 MHz, and the spectrum of Fig. 4 shows a signal at 8435. The test signal was adjusted in amplitude to match the IMP level and then measured to be -162 dBm. The S-band and X-band transmitters were radiating at 20 kW and the IMP generator was shaken during the integration period. The fourth harmonic signal also increased.

Figure 5 presents a clear record of the activity of the total noise power as the test progressed through the series just described. Area A shows the noise power with no RF power radiated. The level of the X-band trace was adjusted to provide a clear picture of the activities in each channel. Area B shows each transmitter coming up to full power with the accompanying momentary disturbance. Note the level changes that are most likely caused by a small amount of transmitter power getting into the LNAs, thus increasing the total system noise power and/or affecting the gain of the LNA. Note that the sensitivity of the recording at X-band is 2.3 kelvins per inch and at S-band is 3.5 kelvins per inch. Also, note that the general momentary response (trace jitter) of each trace did not change with or without the RF power. As the radiated power was shut off, the traces shifted back in the same direction as before the power was applied.

Area C shows the system response to the installation of the IMP generator on the dish. The X-band temperature increased

approximately 1.7 kelvins and the S-band temperature increased approximately 1.4 kelvins due to personnel on the reflector surface. Area D shows a dramatic increase in noise burst activity. Also detectable is the shift in the baseline level, as was seen before without the IMP generator installed. Areas E and F show the increase in activity as the IMP generator was disturbed mechanically. The level increases at both S- and X-band, denoting an increase in system noise temperature and noise burst activity. When the shaking ended, the traces returned to the same levels as before. During these shaking periods, the IMPs were measurable, as seen in Fig. 5.

Area G shows that the S-band system temperature returned to the same pre-transmitter condition when the S-band transmitter interlock automatically shut down the transmitter. When the transmitter came back to full power, the noise bursts also returned, as can be seen in Area H. Following Area H, one additional shaking period of the IMP generator occurred. Next, when the IMP generator was removed from the dish, the presence of personnel on the dish caused a corresponding increase in temperature (Area I). Area J shows performance with each transmitter back up to the full 20 kW but without the IMP generator. This compares well with the previous recordings of Area B.

VII. Conclusions

The common aperture feed at DSS 13 is free of noise bursts and of intermodulation product signals down to the -160- to -170-dBm levels. The instrumentation used for the measurements is capable of detecting activities at that level if they are present, as demonstrated by the artificial IMP generator tests.

Two factors contribute to this good performance. The first is the nature of the feed. The point at which both the S-band and X-band signals are combined is high in the feed, thus exposing few components to the RF power from both transmitters and limiting the opportunities for creating intermodulation products. The second factor is the excellent general "housekeeping" of this antenna. The feed is kept clean and free of foreign matter that contributes to arcing and mixing of signals. The reflecting surfaces of the antenna are free of loose items or marginally clamped items that could provide the nonlinear situations necessary for creating IMPs.

The system does demonstrate the possible need for additional RF filtering to reduce the levels of the fourth harmonic and to reduce the beam noise that apparently is seen by both LNAs. Additional measurements would be required to determine amplitudes and the corresponding filtering requirements.

References

- [1] D. A. Bathker, D. W. Brown, and S. M. Petty, *Single- and Dual-Carrier Microwave Noise Abatement in the Deep Space Network*, Technical Memorandum 33-733, Jet Propulsion Laboratory, Pasadena, California, August 1, 1975.
- [2] R. C. Chernoff and R. W. Hartop, "Noise and Intermodulation Interference in MSFN Back-up Tracking Systems During Transmission of Dual Uplink Carriers," *Deep Space Network Space Progress Summary* 37-57, vol. II, Jet Propulsion Laboratory, Pasadena, California, pp. 138-145, May 31, 1969.
- [3] J. R. Withington, "Second-Generation X/S Feedcone, Layout and Components," *TDA Progress Report* 42-63, vol. March-April 1981, Jet Propulsion Laboratory, Pasadena, California, pp. 97-103, June 15, 1981.
- [4] G. A. Morris, and H. C. Wilck, "JPL 2²⁰ Channel 300-MHz Digital Spectrum Analyzer," *Proc. IEEE International Conf. of Acoustic, Speech, and Signal Processing*, pp. 808-811, 1978.

Table 1. XSR system performance characteristics

Characteristics	S-band	X-band
System temperature	31 K	28 K
Receiving bandwidth	10 MHz	8 MHz
Tunable	2270–2300 MHz	8400–8500 MHz
IF frequency	50 MHz	50 MHz
Transmitter power	20 kW	20 kW
Tunable bandwidth	2110–2120 MHz	7145–7235 MHz

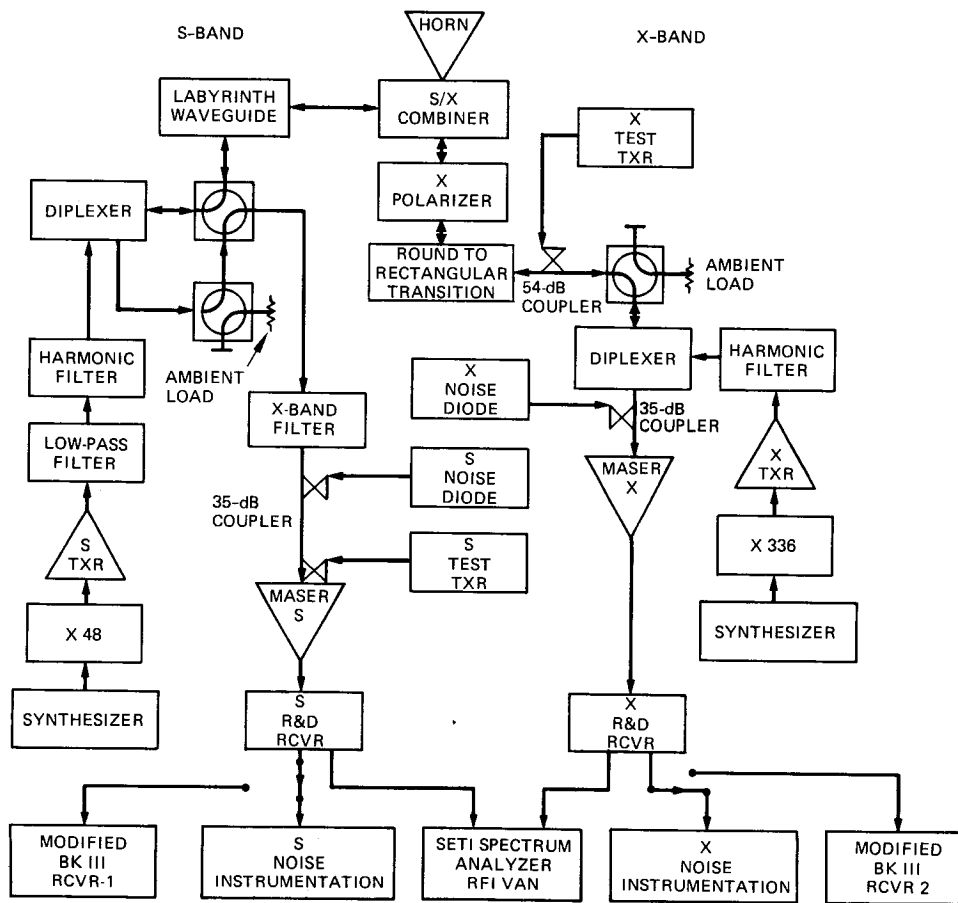


Fig. 1. XSR system functional block diagram

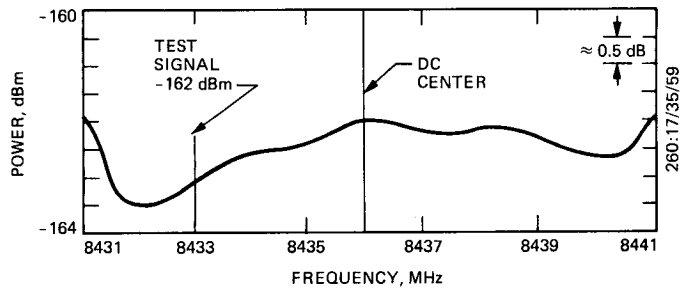


Fig. 2. X-band receive spectrum with locally injected test signal (no transmitter RF power)

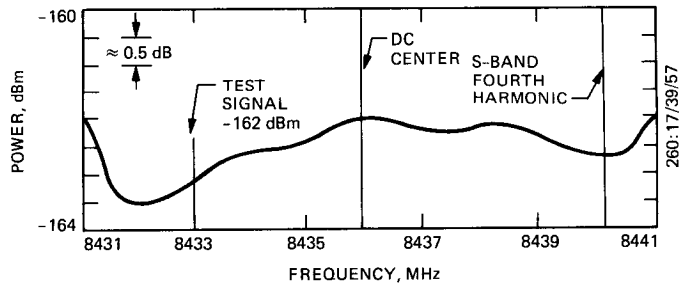


Fig. 3. X-band receive spectrum with S- and X-band transmitters each radiating 20 kW

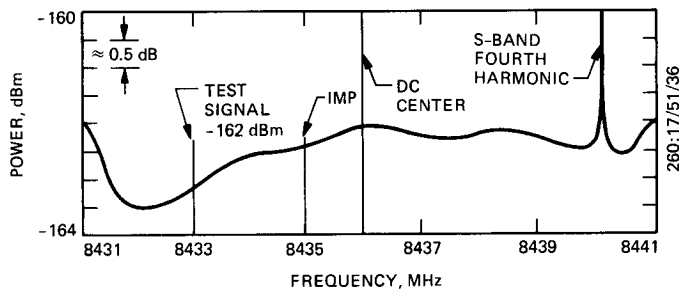
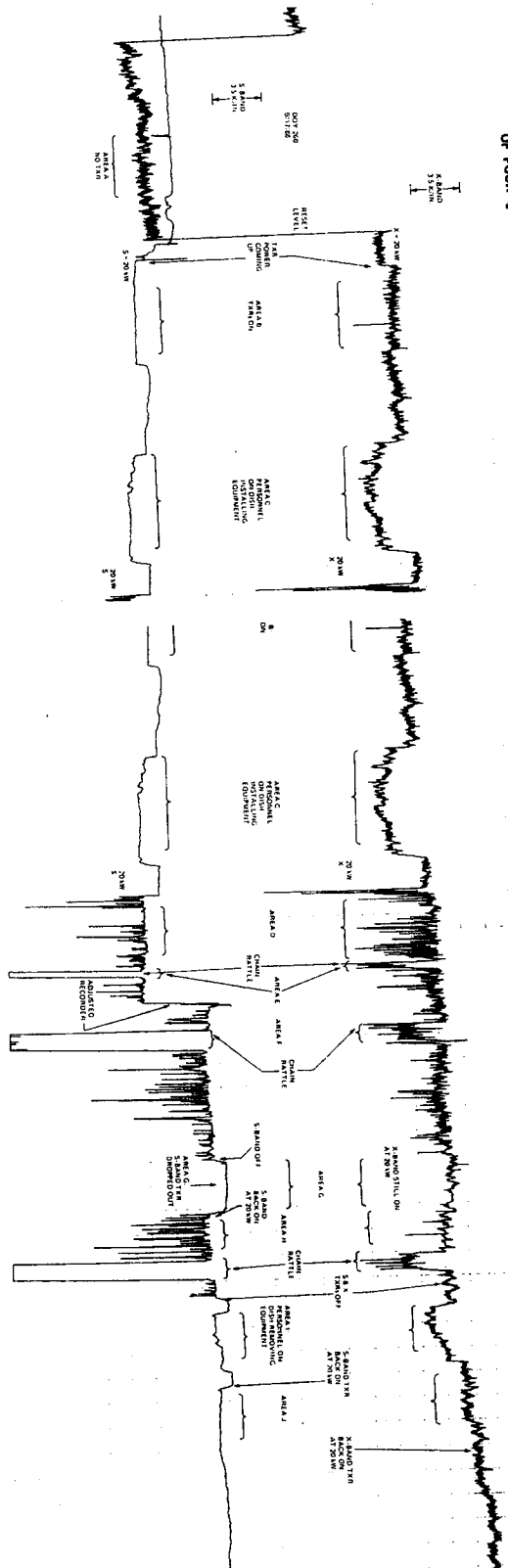


Fig. 4. X-band receive spectrum with artificial IMP generator on the dish surface

ORIGINAL PAGE IS
OF POOR QUALITY



ORIGINAL PAGE IS
OF POOR QUALITY

Fig. 5. Total system noise power spectrum recording

FOLDOUT PLANE

FOLDOUT PLANE

N89 - 10195

X/X/Ka-Band Prime Focus Feed Antenna for the Mars Observer Beacon Spacecraft

P. Stanton, H. Reilly, and M. Esquivel
Radio Frequency and Microwave Subsystems Section

The results of an X/X/Ka-band feed design concept demonstration are presented in this article. The purpose is to show the feasibility of adding a Ka-band beacon to the Mars Observer spacecraft. Scale model radiation patterns were made and analyzed.

I. Introduction

A concept demonstration of an X/X/Ka-band prime focus feed was made for the Mars Observer spacecraft's 1.5-meter reflector antenna. The purpose was to add a Ka-band capability to the spacecraft for downlink experiments [1]. Radiation pattern models of this feed were constructed and pattern measurements made. A computer analysis of these radiation measurements was performed. The analysis shows that further work is required to complete the design, especially in the areas of impedance matching and packaging.

II. Design

The X/X/Ka-band feed design consists of two types of antennas combined in a multifrequency, prime focus feed as shown in Fig. 1. Examples of similar combined feeds, but with much larger low-frequency apertures, are reported in [2] and [3]. The low-frequency (X-band) radiating element used in this case is an open-ended, circular waveguide with four choke rings as presented in [4] and shown in Fig. 2 (with the Ka-band dielectric rod coaxially mounted in the aperture). This radiator provides a suitable pattern over a bandwidth broad enough to include both the uplink (7.19 GHz) and downlink (8.42 GHz)

at X-band. The radiation pattern beamwidth is controlled by the open-ended waveguide aperture diameter.

The high-frequency (Ka-band, 33.7-GHz) element of this design is a dielectric rod surface wave antenna, shown in Fig. 2 (mounted coaxially in the aperture of the X-band antenna) and discussed in [5]. One of the main challenges of this design is to launch a sufficient percentage of the Ka-band power as a surface wave on the dielectric rod. If this is accomplished, the Ka-band radiation pattern is controlled by the characteristics of the dielectric rod and not by the X-band antenna which surrounds it. The launching efficiency of the surface wave is controlled by the profile of the dielectric rod and the flare of the horn at its base (see Fig. 1). The beamwidth of the radiation pattern is controlled mainly by the diameter of the dielectric rod at the antenna aperture.

The combination of the above into a single X/X/Ka-band feed would possibly take the form shown in Fig. 1. The main components of this feed are rectangular-to-circular waveguide transitions, linear-to-circular polarizers, a Ka-band tapered dielectric rod and launching horn, and an X-band aperture radiator with four concentric choke rings.

III. Results

The results presented here are from measurements of two scale models. The X-band model is full scale and is shown in Fig. 2. The Ka-band model is 4.4:1 scale and is shown in Fig. 3. Figures 4 through 6(a) and (b) display the measured E- and H-plane feed radiation patterns in both magnitude and phase. The X-band uplink and downlink patterns (Figs. 4 through 5[a] and [b]) are well matched in the two planes and result in feed efficiencies of approximately 74 percent, including spillover, illumination, cross polarization, and phase efficiencies. The Ka-band patterns (Fig. 6[a] and [b]) are well matched in the two planes but purposely underilluminate the reflector to lower the antenna pointing accuracy requirement, resulting in a feed efficiency of approximately 42 percent.

Figures 7, 8, and 9 show the co-polarized radiation pattern for the X-band uplink, X-band downlink, and Ka-band downlink, respectively. These co-polarized radiation patterns were calculated by the computer program documented in JPL Publication D-1203.¹ This program used the measured feed patterns

¹Y. Rahmat-Samii, "Offset Parabolic Reflector Computer Program for Analysis of Satellite Communications Antennas," JPL Publication D-1203 (internal document), Jet Propulsion Laboratory, Pasadena, California, December 1983.

(Figs. 4, 5, and 6) and the geometry of the Mars Observer spacecraft's 1.5-meter reflector antenna to predict the co-polarized radiation patterns.

Figures 7 and 8 (X-band uplink and downlink) show low sidelobe levels, which result from a larger than typical edge taper. Figure 9 (Ka-band downlink) shows no distinct sidelobes, which can be attributed to the use of an atypical feed taper.

Table 1 presents the X/X/Ka-band antenna performance in the form of feed efficiency plus overall antenna gain and 3-dB beamwidth. The above analysis assumes an ideal reflector and no aperture blockage.

IV. Conclusions

An X/X/Ka-band prime focus feed for the Mars Observer 1.5-meter reflector antenna seems practical based on the results of the scaled model radiation pattern measurements. The X-band and Ka-band feed patterns can be controlled individually. The present Ka-band design would require an antenna pointing accuracy of 0.26 degree to stay within its 3-dB beamwidth. Further work would be required to optimize these radiation patterns.

Acknowledgment

The authors would like to thank M. Gatti for his help in the use of the offset parabolic reflector computer program.

References

- [1] A. L. Riley, D. M. Hansen, A. Mileant, and R. W. Hartop, "A Ka-band (32-GHz) Beacon Link Experiment (Kable) With Mars Observer," *TDA Progress Report 42-88*, vol. October–December 1986, Jet Propulsion Laboratory, Pasadena, California, pp. 66–69, February 15, 1987.
- [2] M. S. Narasimhan and M. S. Sheshadri, "Propagation and Radiation Characteristics of Dielectric Loaded Corrugated Dual-Frequency Circular Waveguide Horn Feeds," *IEEE Trans. Antennas Propagation*, vol. AP-27, no. 6, pp. 858–860, November 1979.
- [3] J. C. Lee, "A Compact Q/K-Band Dual-Frequency Feed Horn," *IEEE Trans. Antennas Propagation*, vol. AP-32, no. 10, pp. 1108–1111, October 1984.
- [4] R. Wohlleben, H. Mattes, and O. Lochner, "Simple Small Primary Feed for Large Opening Angles and High Aperture Efficiency," *Electronics Letters*, vol. 8, pp. 474–476, September 21, 1972.
- [5] H. Jasik (ed.), *Antenna Engineering Handbook*, New York: McGraw-Hill, Chapter 16, 1961.

Table 1. X/X/Ka-band antenna performance (1.5-meter antenna)

Frequency, MHz	Feed efficiency	Gain, dB	3-dB beamwidth, deg
7190	0.741	39.8	2.1
8420	0.741	41.2	1.7
33,680	0.419	50.7	0.52

ORIGINAL PAGE IS
OF POOR QUALITY

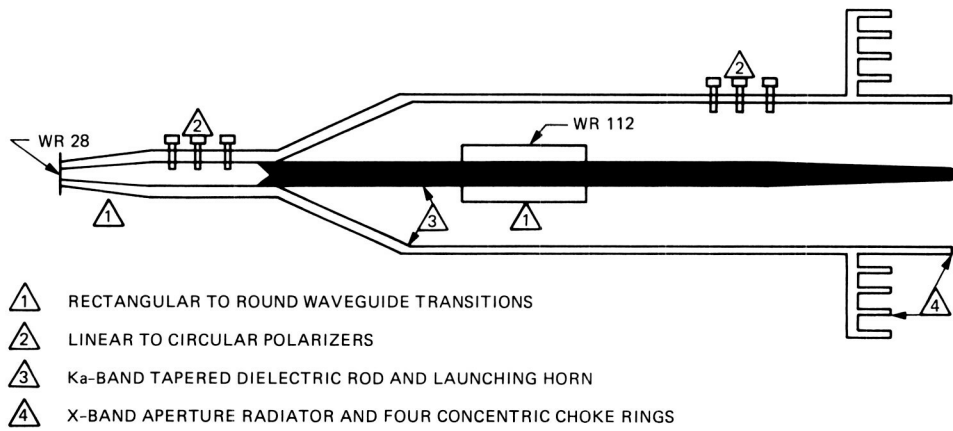


Fig. 1. X/X/Ka-band prime focus feed

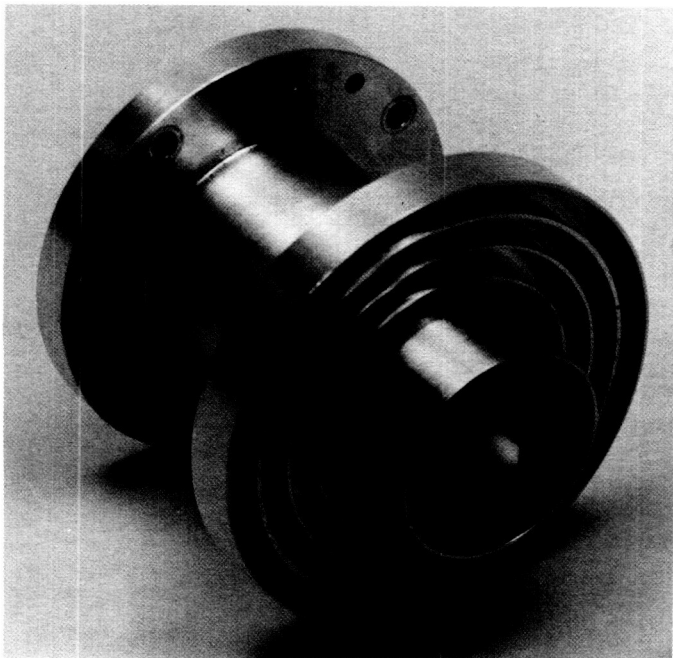


Fig. 2. X-band feed model

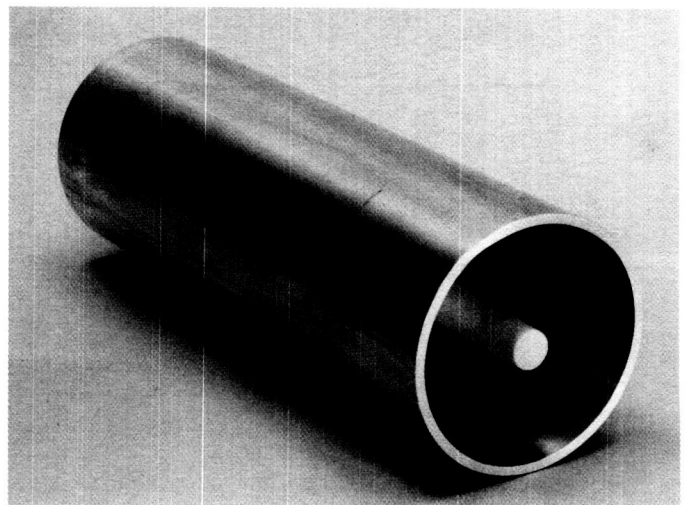


Fig. 3. Ka-band feed model

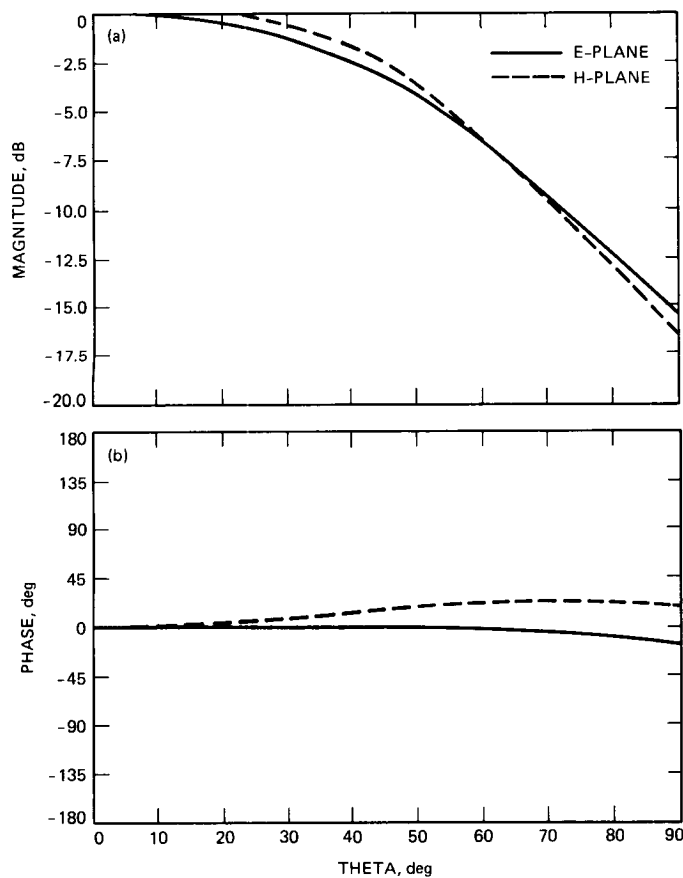


Fig. 4. X-band uplink (7190-MHz) feed: theta versus (a) magnitude and (b) phase

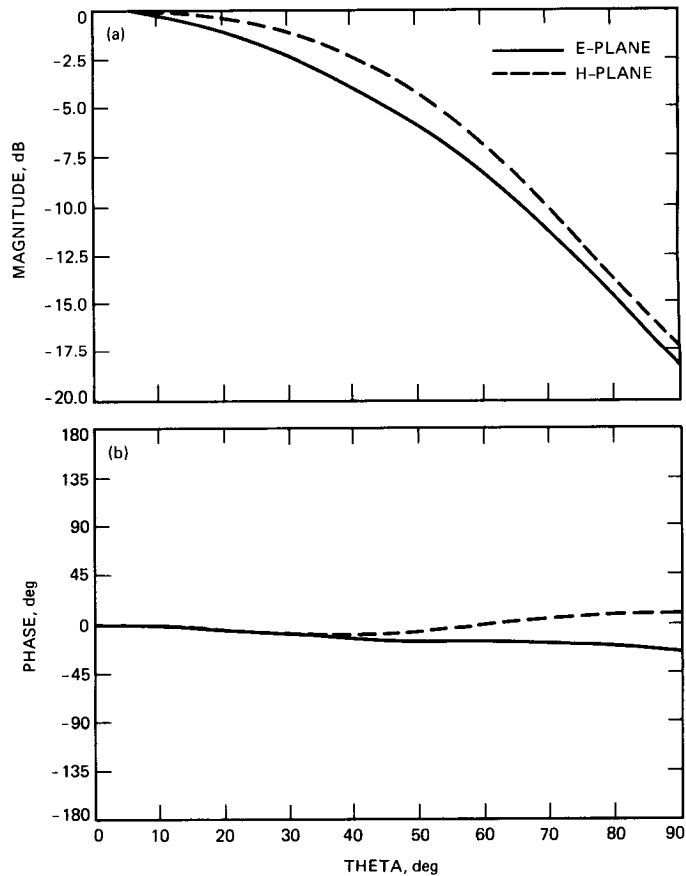


Fig. 5. X-band downlink (8420-MHz) feed: theta versus (a) magnitude and (b) phase

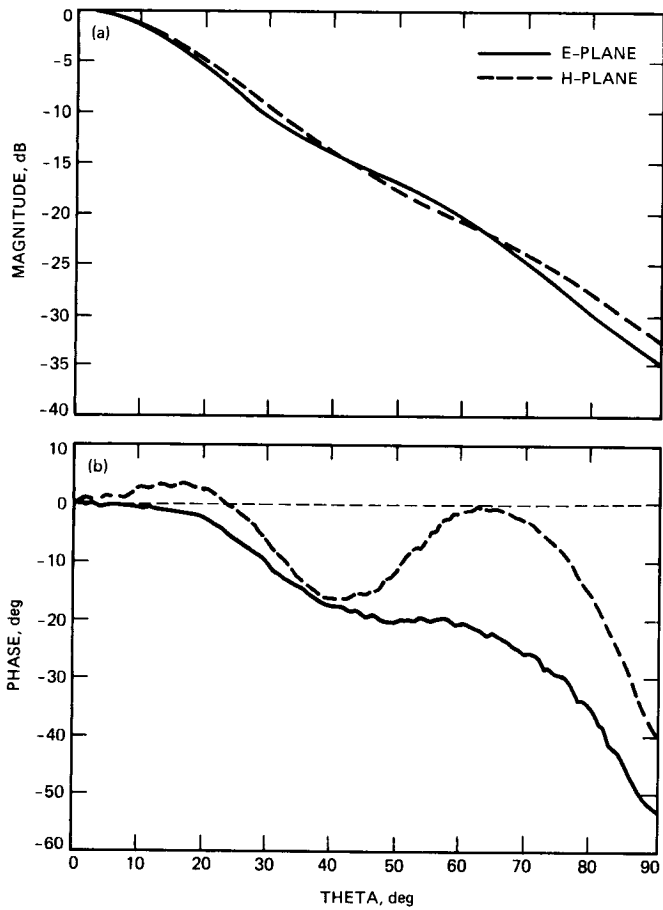


Fig. 6. Ka-band feed: theta versus (a) magnitude and (b) phase

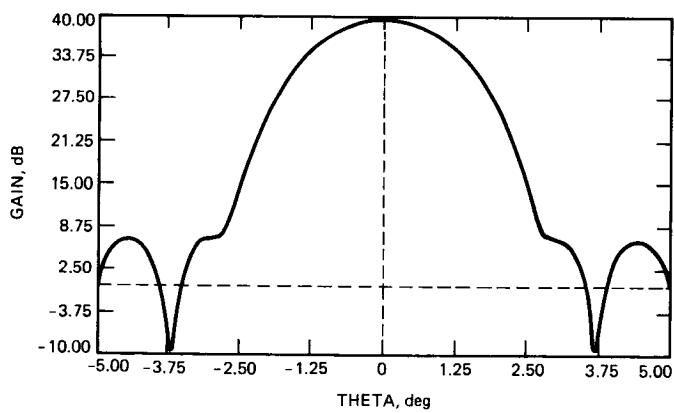


Fig. 7. X-band (7190-MHz) uplink: theta versus gain

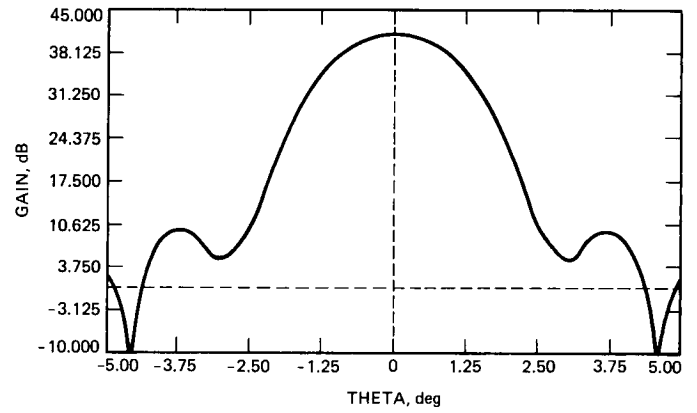


Fig. 8. X-band (8420-MHz) downlink: theta versus gain

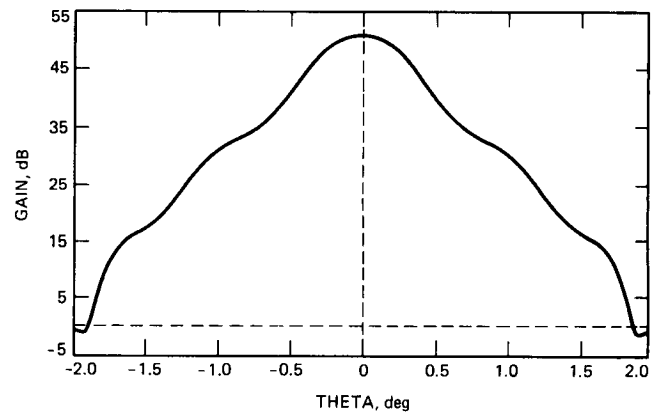


Fig. 9. Ka-band downlink: theta versus gain

510-32

~~145421~~

TDA Progress Report 42-94

~~145421~~

1989000825

618382

26p.

N89 - 10196

April-June 1988

JPL

Dual Passband Dichroic Plate for X-Band

T. Y. Otoshi and M. M. Franco

Radio Frequency and Microwave Subsystems Section

A need arose in the DSN for a dichroic plate that would simultaneously pass two desired frequency bands in the X-band region. In addition, the plate must be totally reflective for S-band frequencies. This article describes a dielectrically loaded dichroic plate that was developed to meet this need. The unique microwave properties that the new dichroic plate had to possess were (1) insertion losses of less than 0.04 dB for X-band uplink frequencies centered at 7167 MHz and for X-band downlink frequencies centered at about 8425 MHz; (2) insertion losses that met the low-loss requirements at 30-deg incidence angle simultaneously for both parallel and perpendicular polarizations; (3) total reflectivity at S-band frequencies; and (4) ability to maintain these electrical characteristics while passing 100 kW of CW power at X-band or while reflecting 100 kW of CW power at S-band. The dual-passband dichroic plate that was developed to meet the technical requirements is a thick metallic plate having an array of periodic round holes filled with Teflon plugs. Test results on an experimental prototype plate indicate that it is technically possible to design a dielectrically filled dichroic plate that meets all of the technical requirements.

I. Introduction

A need arose in the DSN to expand the antenna optics capabilities to include operation at X-band uplink frequencies in addition to maintaining current simultaneous operations at X-band downlink and S-band uplink and downlink frequencies. Although several options are available for converting the present DSN antenna systems to include X-band uplink operation, most of the options are expensive to develop and implement.

A relatively simple option was suggested by R. Clauss of the JPL TDA Technology Development Office. His suggestion involved the development of a dichroic plate of the same type currently being used in the DSN, with the exception that the

array of holes in the dichroic plate be filled with Teflon material. It was pointed out by Clauss that, based on a simple analysis, the proper choice of hole diameter, dielectric material, and plate thickness would result in a dichroic plate that would have passbands at the desired X-band uplink and X-band downlink center frequencies. It was also pointed out by Clauss that a dielectric loaded plate would permit hole diameters and hole-to-hole spacings to be smaller than those of an equivalent air-filled dichroic plate. If a design of an equivalent dual passband air-filled dichroic plate were attempted, a cursory analysis would indicate that grating lobes would appear at downlink frequencies when the plate is used at incidence angles in the region of 30 deg. A plate with closely packed, dielectric filled holes would tend to have fewer grating lobe problems at the higher incidence angles.

The approach of dielectrically filling the holes appeared to be technically promising and relatively easy to develop and implement. However, it was not certain that a more rigorous theoretical analysis would lead to the same conclusion as the simple analysis when taking into account multi-mode waveguide, incidence angle, and polarization considerations. Experimental measurements needed to be performed in order to accurately determine noise temperature contributions of the dielectrically loaded plate. Also, antenna patterns needed to be taken to verify the absence of potential grating lobes and the absence of unknown losses such as those that might be caused by surface wave propagation. It became necessary to initiate a theoretical study and an experimental verification program for dielectrically loaded plates.

A program was initiated to design, fabricate, and test a model plate to demonstrate that a dielectrically loaded dichroic plate could be developed to meet the specified technical requirements. Section II of this article presents descriptions of the dichroic plates that were designed and fabricated and additionally presents corresponding theoretical data. Section III presents results of insertion loss, noise temperature, and antenna pattern measurements. The project status is discussed in Section IV, and a summary and conclusions are presented in Section V.

II. Dichroic Plate Designs and Theoretical Data

Figure 1 shows the geometry of the dichroic plate. The angle θ is the angle of incidence of the incoming plane wave, and ϕ is the spherical coordinate azimuth angle defined with respect to the hole pattern geometry shown in Fig. 1. All dichroic plates currently used in the DSN are plates with air dielectric holes. The new proposed plates will have holes that are filled with a solid, low-loss dielectric material such as Teflon.

The requirements for new generation dichroic plates are that they be transparent to both X-band uplink and X-band downlink frequencies and simultaneously be (nearly 100 percent) reflective to S-band uplink and downlink frequencies. The dichroic plate must be able to transmit 100-kW X-band uplink power without arcing. The following are the microwave technical requirements that the plate must meet:

X-band uplink:

7145–7190 MHz (insertion loss <0.04 dB)

7145–7235 MHz (insertion loss <0.10 dB)

X-band downlink:

8400–8450 MHz (insertion loss <0.04 dB)
(dissipative loss <0.02 dB)

8400–8500 MHz (insertion loss <0.10 dB)
(dissipative loss <0.04 dB)

S-band uplink and downlink:

2090–2320 MHz (reflective insertion loss >45 dB)

The plate must have these microwave properties for both perpendicular and parallel polarizations (Fig. 1) at 30-deg incidence angle at a specified ϕ angle orientation. The requirement for optimization at these particular θ and ϕ angles was dictated by the antenna optical system [1] that the plate would be used in. It was also a goal to find a final plate design that would give the best performance over all incidence angles from 15 deg to 45 deg, and a particular ϕ angle plus and minus 5 deg.

A computer program based on equations derived by Chen [2], [3] was used to design the dielectrically loaded dichroic plate. The original computer program, which was written by Chen for an air dielectric hole medium, was modified by Otoshi to allow the medium in the hole to have any relative permittivity and loss tangent value. After many computer runs, data analyses, and numerous trial designs, it was concluded that all of the requirements could not be met simultaneously for both polarizations and both X-band passbands at 30-deg incidence angle if the aperture is restricted to a round hole. The lack of identical plate properties for both polarizations is a serious problem because it gives rise to elliptical polarization output when the incident wave is circularly polarized.

The only possible method for correcting this polarization problem is to go to a two-diameter (elliptical or two-cutoff wavelength waveguide) design of the type currently being used in the DSN [4]. However, a decision was made to try out the dielectrically filled round-hole design first, because a round-hole design could be handled by an existing computer program. A two-diameter loaded plate that is filled with dielectric plugs is also more expensive to fabricate than a dichroic plate with round holes.

Computer runs made of round-hole dichroic plates showed that if Teflon is used for the dielectric plug material, then the dielectric constant needs to be known to about 5 percent accuracy. The plates are expensive to fabricate. Computer runs with the Chen–Otoshi program showed that if the dielectric constant is in error by more than 5 percent, the plates will not have the passbands at the correct center frequencies. At the start of this project, the values of Teflon were not known well. References [5]–[7] give values for the dielectric constant of Teflon that varied from 2.01 to 2.08 over the X-band region.

A study program was initiated to determine the dielectric constant of the particular Teflon stock actually used in fabri-

cating the plugs for the plates. After studying various methods of measurement of dielectric constants, the Automatic Network Analyzer method [6], [7] was selected. The HP 8510A Automatic Network Analyzer equipment was readily available and data reduction software could be (and was) written in-house. The Teflon samples were made oversize by about 0.002 inch and press-fitted into WR112 waveguides. For the press-fitted waveguide Teflon samples made from the same stock material to be used to fabricate plugs for the dichroic plates, the measured dielectric constant was 2.04 ± 0.01 over the frequency region of 7.0 to 9.0 GHz. A value of 2.04 was used to design the plates. If a lower than actual dielectric constant were used to design the plate, the result would be a plate thickness that is too large. However, this situation might not be serious because it might be possible to later trim the plate thickness to cause the minimum insertion point to shift to the desired center frequency.

A total of three dichroic plates having Teflon plugs were fabricated. The plates had numerous (more than 1000) holes that were closely packed. To avoid tolerance buildup of the hole locations, the plates were fabricated with the aid of a computer-programmed milling machine. All of the coordinates for the holes were computer-programmed, and the coordinate of each hole was kept to within a 0.002-inch tolerance. The material for the plate was 6061-T4 aluminum, and the holes were filled with Teflon plugs made from Teflon rods conforming to Materials Specification MIL-P-19468. The method for holding the plugs in the holes was to make their diameters 0.005-inch oversize, shrink the plugs in liquid nitrogen, drop them into the holes, and allow the plugs to expand at room temperature. The plates were then machined to the proper thickness.

Figure 2 shows one of the fabricated plates with its holes filled with Teflon plugs. The theoretical performance of this plate, labeled Plate #1 for convenience, is shown in Fig. 3 for $\theta = 30$ deg for parallel and perpendicular polarizations. The theoretical values were based on a Teflon dielectric constant of 2.04 and a handbook loss tangent value of 0.00015 [5]. It is sufficient to state that the theoretical curves for Plate #1 show two passbands, a peak-to-peak insertion loss variation of about 2 dB over the band, and a resonance (due to higher-order TM modes) for parallel polarization. The range of the Y-axis is purposely made large to permit comparison with an insertion loss plot for an air-filled dichroic plate, which will be discussed later. In order to show low insertion loss values in the passband in more detail, the theoretical values are given in Table 1 along with the plate dimensions. It should be restated that Plate #1 is a round-hole design, and that a compromise had to be made to give the best performance for X-band downlink for both polarizations and to give an acceptable performance for X-band uplink frequencies.

Plate #2 differed from Plate #1 in that the plate was not as thick and the holes were drilled at a slant angle of 30 deg from the normal for the plate. The reason for drilling the holes 30 deg from normal was that since the plate is to be used for a 30-deg angle of incidence in the actual application, the direction of propagation of the incident ray would be exactly in line with the longitudinal axis of the holes. It was expected that the impedance match would be better and hence would make the passband of the plate broader and also less sensitive to changes of incidence angles in the region of 30 deg.

Plate #3 is the third plate having Teflon plugs as the medium filling the holes of the plate. This plate was originally the round-hole dichroic plate that was designed by Potter and is described in [4]. Although it was known that, when dielectrically loaded, this plate would not have the passbands at the desired uplink and downlink X-band frequencies, this plate was readily available and provided an opportunity to immediately verify the computer program used to design other dielectric filled dichroic plates (Plates #1 and #2). Although this plate is labeled Plate #3 (for convenience), this plate with Teflon plugs was fabricated and tested 2 months before the start of fabrication of Plates #1 and #2. Insertion loss measurements on Plate #3 revealed that theory and experiment were in good agreement to within an accuracy of ± 0.3 dB for the measurement method that was then employed. Based on the initial test results, a decision was made to proceed with the fabrication of Plates #1 and #2, which were already designed.

Figure 4 shows an air-filled dichroic plate, labeled Plate #4, which was used as a comparison standard for verifying the performance and integrity of insertion loss, noise temperature, and antenna pattern measurement test setups. This plate is the original Pyle waveguide dichroic plate, also designed by Potter and described in [4], and is the plate design currently being used in the DSN. To be useful as a comparison and verification standard, however, it was necessary to validate the performance of this plate as well.

Although some data on Plate #4 was available in lab notebooks and publications [4], there were no broadband theoretical or accurately measured insertion data that could be found on this plate. The original computer program used by Potter for the Pyle waveguide aperture also could not be located. Therefore, another approach had to be used to generate theoretical data for the Pyle waveguide dichroic plate. The approach used was to take the design dimensions of the Pyle waveguide on the dichroic plate and find the cutoff wavelengths from the Pyle waveguide curves given in [4]. Then the equivalent TE₁₁ mode circular waveguide diameter was found from these cutoff wavelengths for the applicable polarizations. The perpendicular polarization curve was computed using one round-hole diameter, and the parallel polarization curve was

computed using the other round-hole diameter. The theoretically calculated curves are shown in Fig. 5. Theoretical insertion loss values in the passband as well as plate dimensions for this plate are tabulated in Table 2. Figures 6 and 7 show experimentally obtained data that are in excellent agreement with the theoretical values. The acquisition of excellent experimental insertion loss data was made possible by making the insertion loss measurements in an anechoic chamber and using an automatic network analyzer. An interesting observation when examining the theoretical curves for the dielectrically filled dichroic plate (Fig. 3) and for the Pyle waveguide plate (Fig. 5) is that the air-filled plate appears to have low insertion losses over a wider frequency band.

All of the plates were fabricated to have the same outer diameter of about 33 inches. Each plate had the same mounting hole patterns to enable interchangeability in the test setups. By having all of the plates the same size, it was possible to make accurate comparisons to a standard reference plate whose test results were already known. It was also possible to make measurements of four plates in the same test setup without altering the setup once the integrity of the setup was verified. Test setup, calibration, and verification took up the most time in the measurement process. Once the test setups were verified, the interchangeability of plates enabled measurements of plates to be made relatively quickly.

Although all four plates were tested thoroughly, in order to keep this article to a reasonable length only the results of Plates #1 and #4 will be reported. Results on the other plates will be presented in future publications.

III. Experimental Results

A. Insertion Loss Tests

Low insertion loss in the passband is one of the more important properties that a dichroic plate must have when used in a deep space communications receiving system. Insertion loss includes losses due to both reflection and dissipation. The insertion loss of a plate can be determined from measurements made of the magnitude of the transmission coefficient S_{21} of the path between the transmit and receive horns with and without the dichroic plate inserted into the path. A block diagram of the test setup used to make insertion loss measurements on a dichroic plate is given in Fig. 8. An outdoor test setup with one of the plates being tested is shown in Fig. 9. Perpendicular-to-parallel polarization changes are made by a 90-deg rotation of the circular-to-rectangular waveguide transitions connected to the transmit and receive horns. Rapid 90-deg rotations of these waveguide transitions were facilitated through the use of low-loss, low-VSWR circular waveguide rotary joints.

To enable accurate insertion loss measurements to be made over a broad band of frequencies, the HP 8510A automatic network analyzer was used. The procedure was to first make a reference measurement of S_{21} , the transmission coefficient, for the total system without the plate. The total system without the plate consists of cables, coaxial-to-rectangular waveguide transitions, waveguide isolators, rectangular-to-circular waveguide transitions, the transmit and receive horns, the free-space propagation path, and the dichroic plate holder. The dichroic plate holder (without the plate) was first tilted at the desired incidence angle with the horns oriented for the desired polarization, and then measurements were made of S_{21} and recorded as a function of frequency. This first set of measurements provided the reference data. Then the test was repeated with the plate attached to the plate holder. The automatic network analyzer was then used to convert the S_{21} data into dB, and the first (or reference) measurement data were subtracted from the second (or test) measurement data, thus resulting in plate-only S_{21} data in decibels. Data were taken from 7.0 GHz to 9.0 GHz in 5-MHz steps and stored on diskette files. Insertion group delays of the plates were determined from S_{21} phase data, which were simultaneously recorded.

Figures 10 and 11, respectively, show measured insertion losses for perpendicular and parallel polarizations for the dielectrically loaded Plate #1 when the plate was mounted for a 30-deg incidence angle and $\phi = 90$ deg. The measurements were made in the outdoor setup, rather than in an anechoic chamber, to avoid having to pay the high rental cost (\$400 per day) of using the anechoic chamber. In Figs. 10 and 11, frequency markers show the desired uplink and downlink passbands and the approximate regions where the plates were theoretically supposed to have minimum insertion loss as predicted by the Chen-Otoshi program (see Table 1). It can be seen that the experimental and theoretical passbands agreed reasonably well for perpendicular polarization, but the experimental passband was about 55 MHz too high for the parallel polarization configuration.¹ Other measurements were also made at 0-, 15-, 45-, and 60-deg incidence angles as well as at other ϕ angles. In order to keep this article to a reasonable length, all test data presented will be restricted to the configuration of an incidence angle of 30 deg and $\phi = 90$ deg, which is the geometry that dichroic plates are in when installed on DSN antennas.

Figures 12 and 13 show the insertion loss of Plate #4 as measured in the outdoor test setup. This plate was used as the standard to check out the integrity of the outdoor insertion

¹Slightly better agreement between theory and experiment was obtained at a 30-deg incidence angle test configuration where the plate orientation was changed from a ϕ angle of 90 deg to a ϕ angle of 0 deg.

loss test setup because the insertion loss properties of this plate were known beforehand. Comparisons of the outdoor test setup data (Figs. 12 and 13) with theoretical data in Fig. 5 show that valid insertion loss measurements could be made with the outdoor test setup.

It should be mentioned that for measurements of insertion loss in the outdoor test setup, ripples on the insertion loss as a function of frequency were observed in the test data. These ripples are attributed to multipath signals interacting with the primary desired signal. Proper placement of absorbing material around the horns, frame holder, and flat reflecting surfaces helped reduce the magnitude of multipath ripples. In an outdoor setup such as the one that was used (Fig. 9), it is difficult to totally eliminate the multipath-caused ripples. When a limited number of tests were later made in an anechoic chamber, it was observed that multipath signal-caused ripples on measured insertion loss were almost nonexistent.

The cause of the discrepancy between theory and experiment for dielectrically loaded Plate #1 was originally thought to be an error in the dielectric constant value of 2.04 used to design the plate. However, use of different dielectric constant values ranging from 2.01 to 2.10 through the computer program failed to explain the discrepancy between the theoretical and experimental data for this plate. Hence, an incorrect dielectric constant was ruled out as the possible cause of the discrepancy.

Another cause of the discrepancy between theory and experiment for the dielectrically loaded plate was thought to be poor machining of the surfaces of Plate #1. Some of the Teflon plugs were recessed from the surface of the plates and some of the ends of the Teflon plugs had prominent machining marks and grooves. This uneven plate surface gave rise to the hypothesis that errors could be generated by unwanted surface waves. The plate surfaces were therefore remachined with an extra-sharp cutting tool. The final plate thickness was 0.017 inch less than the original 1.580 inch. However, the machining of the plate to a fine surface finish did not eliminate the discrepancy between theoretical values (based on the new thickness) and experimental values for the parallel polarization case. It should be pointed out that the plate was not remachined to the new thickness until all other data (noise temperature and antenna patterns) were first taken.

Another possible cause of the discrepancy between theory and experiment was thought to be an insufficient number of Floquet modes used in the Chen-Otoshi program, which generated the theoretical data. Changing the number of Floquet modes from $(p, q = 11, 5)$ to $(p, q = 51, 51)$ did not make any significant difference in the theoretical values.

The only other presently known explanation for the discrepancy between theory and experiment is that the theoretical data were computed on the basis of matching the free-space parallel polarization tangential E-field to 5 TM waveguide modes at the dichroic plate interface. When the plate is dielectrically loaded, the number of TM waveguide modes may have to be increased in order to compute the transmission coefficient more accurately when the plate is tilted at 30 deg. Adding more waveguide modes to the existing version of the Chen-Otoshi program turned out to be difficult. At this time, modifications of the Chen-Otoshi program to include additional waveguide modes are still in progress. No definite explanation exists at this time as to the cause of the discrepancy between theoretical and experimental results. The lack of a good theoretical computer program also makes it more difficult to design dielectrically loaded dichroic plates for other frequency bands.

B. Noise Temperature Tests

Noise temperature measurements were made because insertion loss measurements in the outdoor setup could not give results to accuracies of better than 0.1 dB. A good noise temperature setup, where mismatch errors are small, has the potential for providing dissipative loss information to accuracies of better than 0.1 K, corresponding to a dissipative loss of about 0.002 dB.

The method for determining noise temperature of the dichroic plate involves the measurement of effective receiver temperature with and without the plate inserted between the horn and sky. The change in measured effective receiver temperature is attributed to the dichroic plate dissipative loss, which includes plate surface resistivity loss, waveguide conductivity loss, and dielectric dissipative loss. The equations and description of the procedure are described elsewhere,² and hence will not be described further here.

Figures 14 and 15 show the test setups used for the noise temperature measurements of the plate for perpendicular and parallel polarizations, respectively. It was necessary to test the plates for both polarizations because the insertion loss test results showed that significant differences existed for the two polarizations. The basic equipment required for this measurement technique includes a calibrated low-noise horn of known noise temperature, a low-noise maser, a follow-up receiver, noise power measurement equipment, and an ambient load to place over the horn aperture. An attempt was made to reflect the power reflected by the dichroic plate back up to

²"X-Band Traveling Wave Maser Group Block IIA," Part I of II, Technical Manual Overhaul Instructions (internal document), Jet Propulsion Laboratory, Pasadena, California, pp. 113-119, August 1, 1983.

the sky. This was achieved by placing a secondary reflector plate parallel to the dichroic plate at a 30-deg angle from the horizontal (see Figs. 14 and 15). This secondary reflector was a solid aluminum sheet with a 33-inch diameter. Some of the open spaces between the horn aperture and the secondary reflector were covered with metal plates and aluminum foil. Test results later indicated that the 33-inch diameter secondary reflector was inadequate in size and should have been fabricated to have a larger diameter. It is valid to reflect the main reflection lobe from the dichroic plate back up to the sky with the secondary reflector because, in the actual application of the dichroic plate on the large antenna, the reflections from the dichroic plate are re-reflected to the sky by the large main reflector of the Cassegrainian antenna.

Table 3 shows the results of noise temperature tests for Plates #1 and #4. Plate #4 was used as the test setup verification plate because its noise temperature had been previously calibrated. The measured noise temperature value of 1.5 K in this current setup agrees well with expected values for Plate #4. The noise temperature of about 4.5 K for Plate #1 was higher than anticipated. Theoretically, Plate #1 should have a noise temperature of about 2.5 K, of which 1.5 K is due to plate surface resistivity and waveguide conductivity loss while the remaining 1.0 K is due to Teflon dissipation loss (as theoretically calculated from the Chen-Otoshi program when using a loss tangent of 0.00015 for Teflon). Therefore, the measured noise temperature of the dielectrically loaded plate was about 2.0 K higher than expected.

It was initially thought that the higher loss of Plate #1 might be due to the fact that the loss tangent of Teflon is 2 or 3 times higher than the published handbook values. Studies were made later to accurately determine the loss tangent of Teflon. Through the use of two independent measurement techniques, it was established that the Teflon material used for plugs in the dichroic plate had a loss tangent of 0.0002. When this value of loss tangent is used in the Chen-Otoshi modified computer program, a 1.5-K noise temperature due to dielectric loss is calculated. The two methods used to determine loss tangent were (1) an improved automatic network analyzer measurement technique; and (2) an improved noise temperature measurement method. These two improved methods and associated test results will be discussed in a separate report. It was important to determine the loss of Teflon accurately because Teflon is one of the less expensive readily available materials that have low loss. It was desirable to prevent erroneous conclusions from being drawn based upon the dichroic plate noise temperature measurements. It would be a serious mistake to conclude prematurely that dielectrically loaded dichroic plates are not useful because they have undesirably high noise temperatures generated by dielectric losses.

The poor results in Table 3 for Plate #1 might be explained by the following equations, which are applicable to the test setups shown in Figs. 14 and 15. The equations for system noise temperature for the measurement setups for the configuration with and without the dichroic plate are respectively given as:

$$T_{sys} = T_{sky} + T_{horn} + T_{maser} + T_{fr} \quad (1)$$

$$T'_{sys} = (1 - |\Gamma_{dp}|^2) T_{sky} + T_{dp} + |\Gamma_{dp}|^2 |\Gamma_{sp}|^2 T_{sky} + |\Gamma_{dp}|^2 (1 - |\Gamma_{sp}|^2) T_{pg} + T_{horn} + T_{maser} + T_{fr} \quad (2)$$

Then

$$T_{diff} = T'_{sys} - T_{sys} = T_{dp} + |\Gamma_{dp}|^2 (1 - |\Gamma_{sp}|^2) (T_{pg} - T_{sky}) \quad (3)$$

where

T_{sky} = sky noise temperature (approximately 5 K) due to cosmic background (2.7 K) and the zenith atmospheric noise temperature (nominally 2.3 K) at X-band, K

T_{horn} = horn noise temperature, K

T_{maser} = maser noise temperature, K

T_{fr} = follow-up receiver noise temperature, K

T_{dp} = dichroic plate noise temperature, K

T_{pg} = physical temperature of ground (nominally 290 K), K

$|\Gamma_{dp}|^2$ = square of the magnitude of the voltage reflection coefficient as seen looking at the dichroic plate from the horn aperture plane

$|\Gamma_{sp}|^2$ = ratio of the power reflected by the secondary plate to the total power radiated toward the secondary plate due to reflections from the dichroic plate

Equation (3) shows that the error in measuring the noise temperature of the dichroic plate will be proportional to the magnitude squared of the reflection coefficient of the dichroic plate times the ratio of power not re-reflected by the secondary plate. For example, if at the noise temperature measurement frequency, the insertion loss of a dichroic plate is 0.1 dB, corresponding to a voltage reflection coefficient magnitude

of 0.15, and the secondary plate captures and reflects only 90 percent of the power radiated toward it (as a result of reflection from the dichroic plate), then from Eq. (3) the error in measuring the noise temperature of the dichroic plate will be

$$0.15 \times 0.15 \times (1.0 - 0.90) \times (290 - 5) = 0.6 \text{ K}$$

If instead the insertion loss of the plate is 0.5 dB, corresponding to a voltage reflection coefficient magnitude of 0.33, the error in measurement will be 3.1 K. A study of this type is being made to correlate the insertion loss data measured for Plates #1 and #4 to noise temperature errors at the various measurement frequencies and the two polarizations. The results of this study will be reported later.

A good test would have been to cover the lower side of the dichroic plate with a metallic sheet having the same diameter as the dichroic plate. This would simulate a worst-case reflection coefficient of 1.0 from the dichroic plate. All of the power coming from the dichroic-plate reflection-lobe direction could be captured by a properly designed secondary plate, and this captured power could then be redirected towards the sky. According to Eq. (3), and taking into account resistive losses on the reflectors, a measurement of antenna temperature would reveal how effective the secondary plate was and how much noise temperature was due to ground absorption.

C. Antenna Patterns

As a final test, antenna patterns were taken of the dichroic plates to verify that there were no unknown effects such as unknown grating lobes, beam shifting, or scattering patterns not predicted by theory. Antenna patterns taken in different planes will give information about scattering losses in three dimensions, whereas insertion loss measurements basically give information only in one propagation direction between the transmit and receive horns.

The measurement of antenna patterns is more laborious and time consuming because, at a single frequency, a pattern must be recorded for 360 deg of angle rotation. There are main E-plane patterns (corresponding to parallel polarization) and main H-plane patterns (corresponding to perpendicular polarization) as well as cross-polarization patterns that have to be taken. Then the same types of patterns have to be taken with the plate rotated 90 deg so that the plate coordinate system is rotated 90 deg with respect to the axis of rotation for pattern recording.

Figure 16 shows the antenna pattern measurement setup for measurement of the horn only in the 60-ft anechoic chamber at the JPL Mesa Facility. The phase center of the horn was aligned with the axis of rotation of the table. The procedure

was to measure the pattern of the horn only, and then to measure patterns with the dichroic plate installed in place. Then the dichroic plate assembly was turned over on its side 90 deg and the pattern was again recorded. Figure 17 shows the test configuration where the plate holder and plate are in an upright position similar to that used for insertion loss measurements. In this configuration, the main reflection from the plate will be radiated toward the ceiling of the anechoic chamber and be absorbed. Figure 18 shows the test configuration where the plate is turned over on its side 90 deg. In this test configuration, the main reflection lobe will appear in the antenna pattern recording. The plate was mounted for a $\phi = 90$ -deg hole pattern orientation angle, and a 30-deg incidence angle position. The rotator angle reads 0 deg and the receive and transmit horns are boresighted to each other. Correct alignment and boresighting of the transmit and receive X-band horns were previously verified with an optical laser beam.

E- and H-plane patterns were taken of the horn only, horn plus plate holder, and horn plus plate holder and plate for both the upright and flopped 90-deg positions. All corresponding cross-polarization patterns according to Ludwig definition #1 [8] were also taken. All of these types of patterns were taken for Plate #1 at uplink frequencies of 7145 MHz and 7168 MHz and also at downlink frequencies of 8415 MHz, 8425 MHz, and 8450 MHz.

For the DSN plate (Plate #4), only E- and H-plane patterns at 8415 MHz were taken for verification purposes. The antenna patterns were virtually identical to those given in [4] and, therefore, measured patterns for Plate #4 will not be presented in this article.

Figures 19 and 20, respectively, show E- and H-plane patterns for the horn alone at a frequency of 7168 MHz, which corresponds approximately to the middle of the desired uplink passband. Figures 21 and 22, respectively, show the antenna patterns when dichroic Plate #1 and the plate holder were installed in the flopped 90-deg position. Figures 23 through 26 show similar sets of patterns taken at the downlink mid-passband frequency of 8425 MHz. Patterns taken with the frame assembly upright were virtually the same as the horn-alone patterns, with the exception of about 0.2 dB loss in the peak of the main beam relative to the horn-alone case. Therefore, patterns taken with the dichroic plate holder in the upright position (see Fig. 17) will not be shown. Cross-polarization patterns will not be shown because in all cases the patterns were at least 35 dB down from the peak of the main beam of the main polarization patterns.

An interesting correlation can be made of antenna pattern data and the insertion loss data given in Figs. 10 and 11. If the

level of the largest reflection lobe is subtracted from the peak level of the main beam of the horn-only pattern, and the difference of levels is treated as return loss, these data can be converted to insertion loss. The results of this analysis are given in Table 4. Reasonable agreement was obtained between insertion loss values computed from pattern data and insertion loss values (± 0.1 dB accuracy) in the outdoor test setup.

One final test was needed to show that Plate #1 behaves as a solid reflector at S-band frequencies. In order to perform the required test, it was necessary to replace the transmit and receive X-band horns in the anechoic chamber test setup with S-band standard gain horns. The particular type of standard gain horn used was a Scientific-Atlanta (Model 12-1.7) standard gain horn that has a WR430 waveguide input and a gain of about 16.4 dB at 2295 MHz. Antenna patterns were made at 2295 MHz with Plate #1 in the 90-deg flopped position, and then again after the front face of the plate was covered with a thin solid aluminum sheet of the same diameter. No differences were observed in the two patterns. It is safe to conclude that below cutoff, a dielectric loaded plate has properties similar to a perfect reflector at S-band. This conclusion is supported by theoretical data obtained from the Chen-Otoshi computer program.

The antenna pattern tests verified that there was no unusual behavior of the dielectrically loaded dichroic plate such as might be caused by unexpected surface wave radiations or grating lobes. The experimental work generally confirmed most of the theoretical predictions concerning the behavior of dielectrically loaded dichroic plates.

IV. Project Status

A. Accomplishments

The following is a list of accomplishments and spinoffs resulting from this project:

- (1) Numerous computer programs written in FORTRAN and BASIC for IBM-XT/AT, including:
 - (a) Waveguide conductivity loss for a circular or rectangular waveguide when the waveguide is filled with any dielectric material;
 - (b) Below cutoff attenuation for numerous TE and TM modes for a circular or rectangular waveguide of specified length and frequency;
 - (c) Generation of theoretical S-parameters for a dielectric sheet in either a circular or rectangular waveguide;

- (d) Derivation of complex permeability and complex permittivity values from measured S11 and S21 on a dielectric-filled waveguide sample;
- (e) Improvement of measurements of noise temperature when using the horn-maser Y-factor measurement technique. The improvement occurs in the data reduction that accounts for atmospheric noise temperature changes due to weather condition changes. (This improvement technique might be valuable for future Beam Waveguide system evaluations.);
- (f) Antenna computer programs related to dichroic plate measurements such as antenna range loss (Friis transmission loss formula), far-field criteria, and circular and rectangular aperture theoretical antenna gain programs; and
- (g) A computer program to compute group delay from network analyzer phase measurements.

- (2) Two methods were developed for measuring the permeability and permittivity of materials. The first method involves the measurement of S-parameters of the dielectric sample in a waveguide. The second method involves the measurement of the sample in a waveguide when terminated with a short-circuit. In addition, significant improvements were made on the current network analyzer technique to accurately determine the loss tangent of low-loss dielectric materials. Heretofore, the network analyzer technique could not be used to accurately determine loss tangents of low-loss dielectric materials.

- (3) A new noise temperature measurement technique was developed for the accurate determination of dissipative loss of dielectric materials from noise temperature measurements. This method uses the technique of making the waveguide sample an integer multiple of a half-waveguide wavelength at the test frequency. Substitution of this waveguide sample for an empty waveguide of identical electrical length (already in the system) enables cancellations of waveguide conductivity loss and minimization of mismatch errors.

Other new results obtained on dichroic plates not yet reported include the following:

- (1) Good agreement was obtained between theoretical and measured group delays of Plates #1 and #4;
- (2) Data were taken with various-thickness Teflon sheets placed on one or both sides of dichroic plates. The

passbands shifted, but as yet the experimental data have not been correlated to theory. The employment of dielectric sheets might be a good way to fine-tune the dichroic plate passband characteristics;

- (3) It was found that the insertion losses of a dichroic plate, as measured over a frequency range of 7.0 to 9.0 GHz with the test setup configurations shown in Figs. 8 and 9, did not change when the transmit X-band horn having a diameter of 7 inches was replaced by a pyramidal horn whose largest dimension was 5 inches. The multipath ripple magnitude increased for data taken with the smaller horn, but the average measured insertion loss curve (see Figs. 10 and 11) was the same for either transmit horn test setup;
- (4) Data obtained on the slant-drilled Teflon-loaded dichroic Plate #2 described in Section II showed that higher than expected insertion loss occurred for parallel polarization and 30-deg incidence angle at X-band downlink frequencies. The corresponding noise temperatures were also much higher than expected. No beam squint was observed on the antenna patterns for this plate;
- (5) Improved techniques for future insertion loss measurements of dichroic plates were learned. It was found that time-consuming calibrations could be skipped if isolators were used. Also, it was learned that a solid reflector should be used to check for leakage if very high insertion values are to be measured, such as for a dichroic plate below cutoff;
- (6) Improvements in noise temperature measurements of a dichroic plate can be obtained if a solid plate is placed in front of the dichroic plate. This measurement would enable determination of the integrity of the secondary reflector to determine how much noise temperature is coming from the ground (see Eq. (3)).

B. Problem Areas

An important part of the development of dielectrically loaded dichroic plates is the accurate determination of the properties of the dielectric material to be used for the plates. A dichroic plate with over 1000 holes and loaded with dielectric material is expensive to fabricate. If the dielectric constant is not closely known, then it may not be possible to fine-tune the plate by trimming the plate thickness. The dielectric constant needs to be determined for the density state that the plugs will be in after installation by whatever method is employed for locking the plugs in place. Accurate determination of the dielectric constant is a time-consuming task that is not generally appreciated or understood as being a necessary part of this developmental work.

Another technical challenge related to dichroic plate design lies in finding a suitable material that has low loss, but at the same time is mechanically durable, chemically inert, and commercially available at reasonable cost. The selection and evaluation of materials presents the same type of technical challenge that developers of radomes have had to face in previous and current decades.

Another technical problem that currently needs to be solved is the development of an improved computer program that accurately predicts the passbands for the dielectrically loaded plate. Currently, the Chen-Otoshi program version being used is a 10-waveguide mode program. Revising the program to include twice as many modes might result in a computer program tool that would be invaluable for designing dichroic plates. Also, the incorporation of a version that provides for arbitrary aperture shapes other than round holes might prove invaluable.

Another unsolved technical problem at this point of development is that of knowing the CW power at which arcing across the holes will occur. Some preliminary calculations show that breakdown will not occur at 100 kW of CW power, but experimental verifications need to be made.

V. Summary and Conclusions

The primary objectives of this study have been accomplished. It has been demonstrated that a dichroic plate having two passbands at the desired frequencies can be developed. Over the frequency range of 7.0 to 9.0 GHz, good agreement was obtained between theory and experiment for perpendicular polarization. For parallel polarization, the measured passband occurred at a center frequency that was about 55 MHz higher than theoretical predictions for both uplink and downlink passbands. The noise temperature of the Teflon-loaded plate is determined to be about 3.0 K if corrections are made for about 1.5 K error that is attributable to measurement setup problems. This 3.0 K noise temperature value includes a contribution of about 1.3 K due to dissipative losses of the Teflon plugs. Antenna pattern measurements showed that for the dielectrically loaded plate at 30-deg incidence, there was only one reflection lobe, as theoretically predicted, and that this lobe had a magnitude that correlated with the data obtained from insertion loss measurements.

Immediate technical challenges and tasks that remain for developing a dielectrically loaded dichroic plate for the DSN are (1) the development of a Pyle waveguide type dielectrically filled dichroic plate; (2) improvement in the computer program to theoretically design the plates; and (3) experimental

verification that arcing does not occur for a dielectrically loaded plate at high CW power levels at about 100 kW.

Long-term projects include (1) research into lower loss type dielectric materials; (2) research into different types of apertures (other than round holes and Pyle waveguide) for develop-

ing plates that operate over a broader bandwidth and over a greater range of incidence angles; (3) development of computer programs and theory to design new classes of improved dichroic plates; and (4) development of less expensive, less time-consuming techniques for experimental verifications of performance.

Acknowledgments

R. Clauss, of the TDA Technology Development Office, provided the initial idea of designing a dielectrically loaded plate for dual passband purposes. P. Parsons, of the TDA Engineering Office, provided technical requirements for the dichroic plate project.

The accomplishments of this project were made possible because of a cooperative team effort by various individuals of the Radio Frequency and Microwave Subsystems Section. Management support and encouragement came from Section Manager W. Imbriale, RTOP Manager A. Cha, and Group Supervisor D. Bathker. Technical and manpower assistance were provided by the Microwave Engineering Group on Automatic Network Analyzer measurements on Teflon dielectric samples, the Microwave Electronics Group on insertion loss and noise temperature measurements, and the Antenna and Microwave Development Group on all phases of the experimental work. Appreciation is expressed to J. Withington, of the Antenna and Microwave Development Group, who allowed temporary rearrangement of his work in order to permit antenna patterns to be taken of the dichroic plates in the 60-ft anechoic chamber. Invaluable computer-programming and data-reduction assistance was provided by E. Andres, JPL contractor. Appreciation is expressed to many unnamed technical and support people who directly or indirectly contributed to this project.

References

- [1] D. A. Bathker, "Dual Frequency Dichroic Feed Performance," *Antennas for Avionics*, AGARD-CP-139, 26th Meeting of Avionics Panel, pp. 29-1-29-10, June 1974.
- [2] C. C. Chen, "Diffraction of Electromagnetic Waves by a Conducting Screen Perforated Periodically with Circular Holes," *IEEE Trans. Microwave Theory and Techniques*, vol. MTT-19, pp. 475-481, May 1971.
- [3] C. C. Chen, "Transmission of Microwave Through Perforated Flat Plates of Finite Thickness," *IEEE Trans. Microwave Theory and Techniques*, vol. MTT-21, pp. 1-6, January 1973.
- [4] P. D. Potter, *Improved Dichroic Reflector Design for the 64-Meter Antenna S- and X-Band Feed Systems*, JPL Technical Report 32-1526, vol. XIX, Jet Propulsion Laboratory, Pasadena, California, pp. 55-62, February 15, 1974.
- [5] *Reference Data for Radio Engineers*, Fifth Edition, New York: International Telephone and Telegraph, pp. 4-28-4-29, 1968.
- [6] W. B. Weir, "Automatic Measurement of Complex Dielectric Constant and Permeability at Microwave Frequencies," *Proc. IEEE*, pp. 33-36, January 1974.
- [7] Hewlett-Packard, "Measuring the Dielectric Constant of Solids with the HP 8510 Network Analyzer," *Hewlett-Packard Product Note 8510-3*, August 1, 1985.
- [8] A. C. Ludwig, "The Definition of Cross Polarization," *IEEE Trans. Antenna Propagat.*, vol. AP-21, pp. 116-119, January 1973.

Table 1. Design data^a and theoretical passband insertion losses of dielectric filled hole plate (Plate #1) for $\phi = 90$ deg and incidence angle = 30 deg

Polarization	Band ^b	Passband insertion loss, dB			Comments
		Start frequency	Mid frequency	End frequency	
Perpendicular	Uplink	0.02	0.06	0.14	Not too good at mid-band and end of band
Perpendicular	Downlink	0.01	0.02	0.06	Not too good at end of band
Parallel	Uplink	0.13	0.06	0.02	Not too good at start and mid-band
Parallel	Downlink	0.04	0.02	0.01	

^aHole diameter = 0.803 in., hole-to-hole spacing = 0.850 in., plate thickness = 1.580 in.
Hole medium dielectric constant = 2.04.
Loss tangent = 0.00015.

^bUplink, downlink = 7170 \pm 30 MHz, 8430 \pm 30 MHz.

Notes:

- (1) The insertion loss values include dissipative loss due to the loss tangent of Teflon.
- (2) At some of the passband frequencies, the results are not good at 15- or 45-deg incidence angles.

Table 2. Design data^a and theoretical passband insertion losses of DSN air-filled hole plate (Plate #4) for $\phi = 90$ deg and incidence angle = 30 deg

Polarization	Band ^b	Passband insertion loss, dB		
		Start frequency	Mid frequency	End frequency
Perpendicular	Downlink	0.000	0.009	0.037
Parallel	Downlink	0.025	0.004	0.001

^aPyle waveguide dimensions = 0.866 and 0.900 in., respectively, for perpendicular and parallel polarizations. The corresponding cutoff wavelengths of the Pyle waveguide are 1.511 and 1.539 in. (see note 1 below for equivalent round-hole dimensions).

Hole-to-hole spacing = 0.940 in., plate thickness = 1.408 in.

Hole medium dielectric constant = 1.00.

Loss tangent = 0.00000.

^bDownlink = 8415 \pm 30 MHz.

Notes:

- (1) This DSN Plate has a Pyle waveguide aperture (see [4]) rather than a round-hole aperture. The equivalent round-hole diameters corresponding to the 0.866- and 0.900-inch Pyle waveguide dimension are respectively 0.886 and 0.902 inches. These round-hole dimensions were used to compute the theoretical insertion losses.
- (2) This plate has no dielectric dissipative loss term associated with insertion loss.
- (3) At some of the passband frequencies, the results were not too good at 15- or 45-deg incidence angles.

Table 3. Noise temperature test results for dichroic plates at $\phi = 90$ deg and $\theta = 30$ deg

Description	Perpendicular polarization noise temperature, K			Parallel polarization noise temperature, K		
	8400 MHz	8450 MHz	8500 MHz	8400 MHz	8450 MHz	8500 MHz
Plate #1 with Teflon plugs	4.7	4.3	4.0	11.0	5.2	4.0
Plate #4 air dielectric	1.5	3.2	1.6	1.6	1.6	1.7
Difference in noise temperature	3.2	1.1	2.4	9.4	3.6	2.3

Notes:

- (1) The measured mid-frequency of the insertion loss passbands are 8450 MHz for perpendicular polarization and 8500 MHz for parallel polarization for Plate #1.
- (2) The measured mid-frequency of the insertion loss passbands is 8415 MHz for both perpendicular and parallel polarizations for Plate #4.

Table 4. Correlation of the relative level of the reflection lobe to measured insertion loss for Plate #1

Pattern	Frequency, MHz	Relative level of reflection lobes, dB	Equivalent insertion loss, dB	Measured insertion loss	
				Polarization	dB
E-Plane	7168	-16.0	0.11	Parallel	0.2
H-Plane	7168	-21.0	0.03	Perpendicular	0.2
E-Plane	8425	-20.0	0.04	Parallel	0.2
H-Plane	8425	-21.0	0.03	Perpendicular	0.05

Notes:

- (1) Data for reflection lobe levels were obtained from patterns of Plate #1 and taken relative to the peak of the corresponding horn-only pattern.
- (2) Data for the measured insertion loss were obtained from Fig. 10 and 11, respectively, for perpendicular and parallel polarizations.

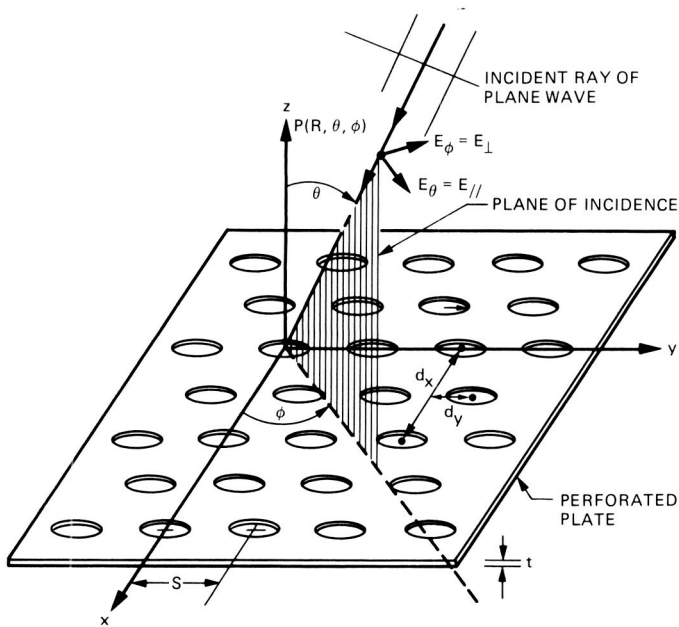


Fig. 1. Dichroic plate geometry

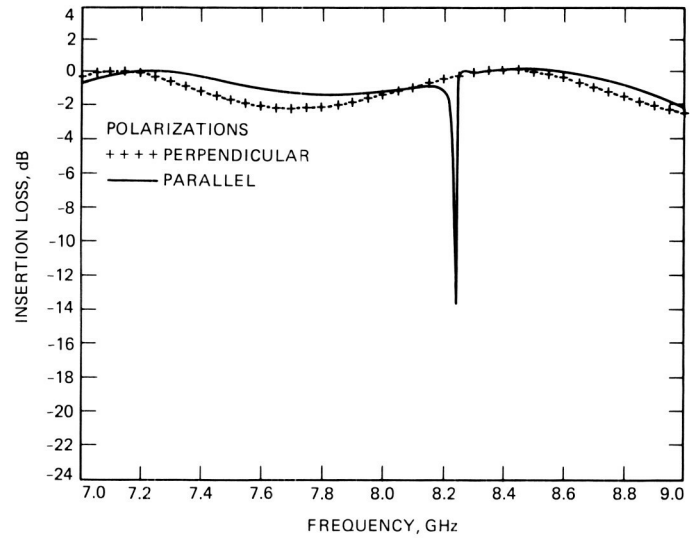


Fig. 3. Theoretical insertion loss versus frequency for the dielectrically loaded plate (Plate #1) for perpendicular and parallel polarizations

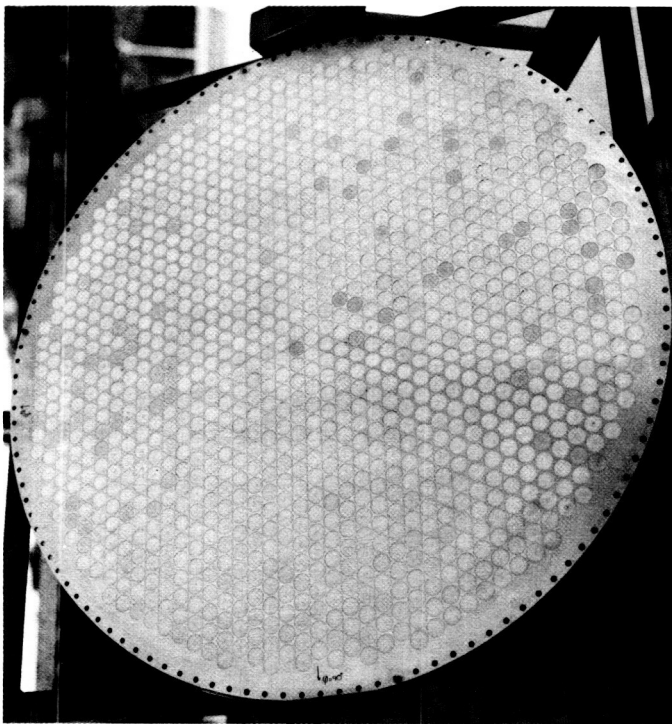


Fig. 2. New X-band dual passband dichroic plate with holes filled with Teflon plugs; this plate is referred to as Plate #1 in the text

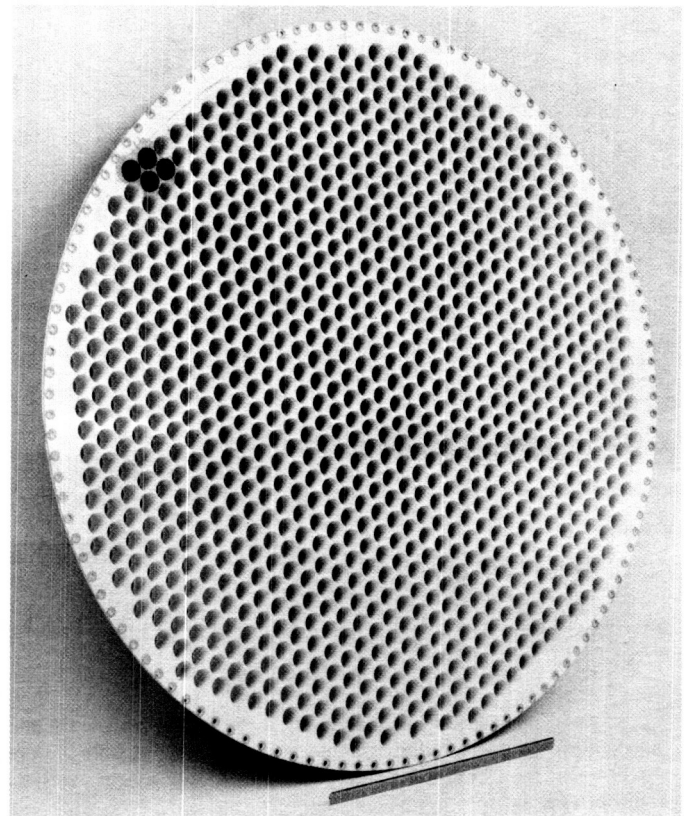


Fig. 4. Prototype DSN X-band dichroic plate; this plate is referred to as Plate #4 in the text

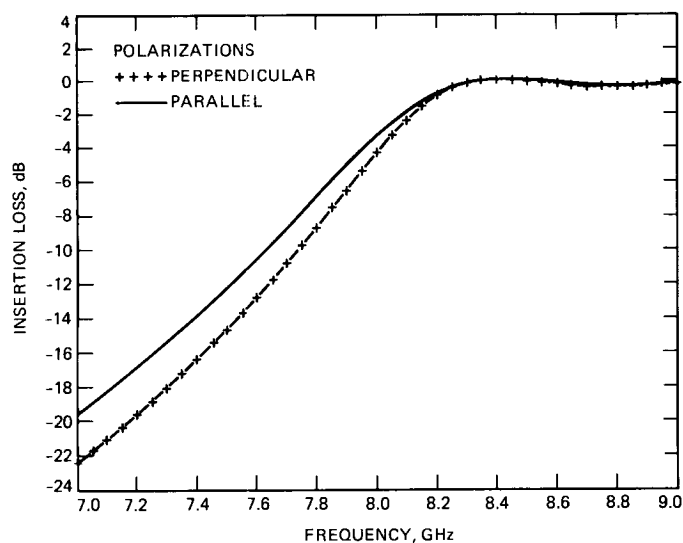


Fig. 5. Theoretical insertion loss versus frequency for the DSN plate (Plate #4) for perpendicular and parallel polarizations

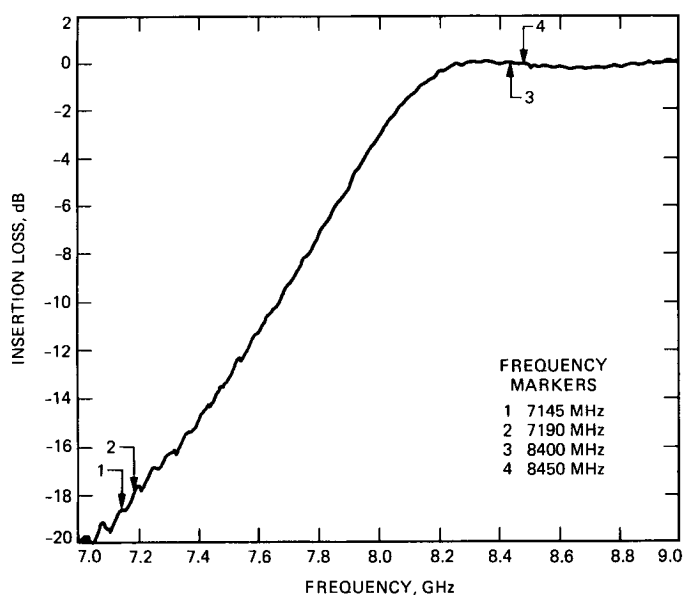


Fig. 6. Measured insertion losses of Plate #4 for perpendicular polarization, 30-deg incidence angle, and $\phi = 90$ -deg hole pattern orientation; measurements were made in the 60-ft anechoic chamber test setup

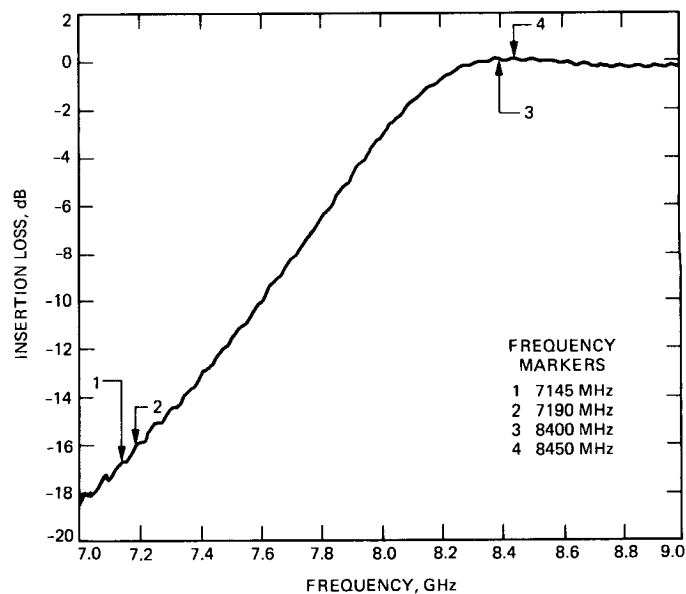


Fig. 7. Measured insertion losses of Plate #4 for parallel polarization, 30-deg incidence angle, and $\phi = 90$ -deg hole pattern orientation; measurements were made in the 60-ft anechoic chamber setup

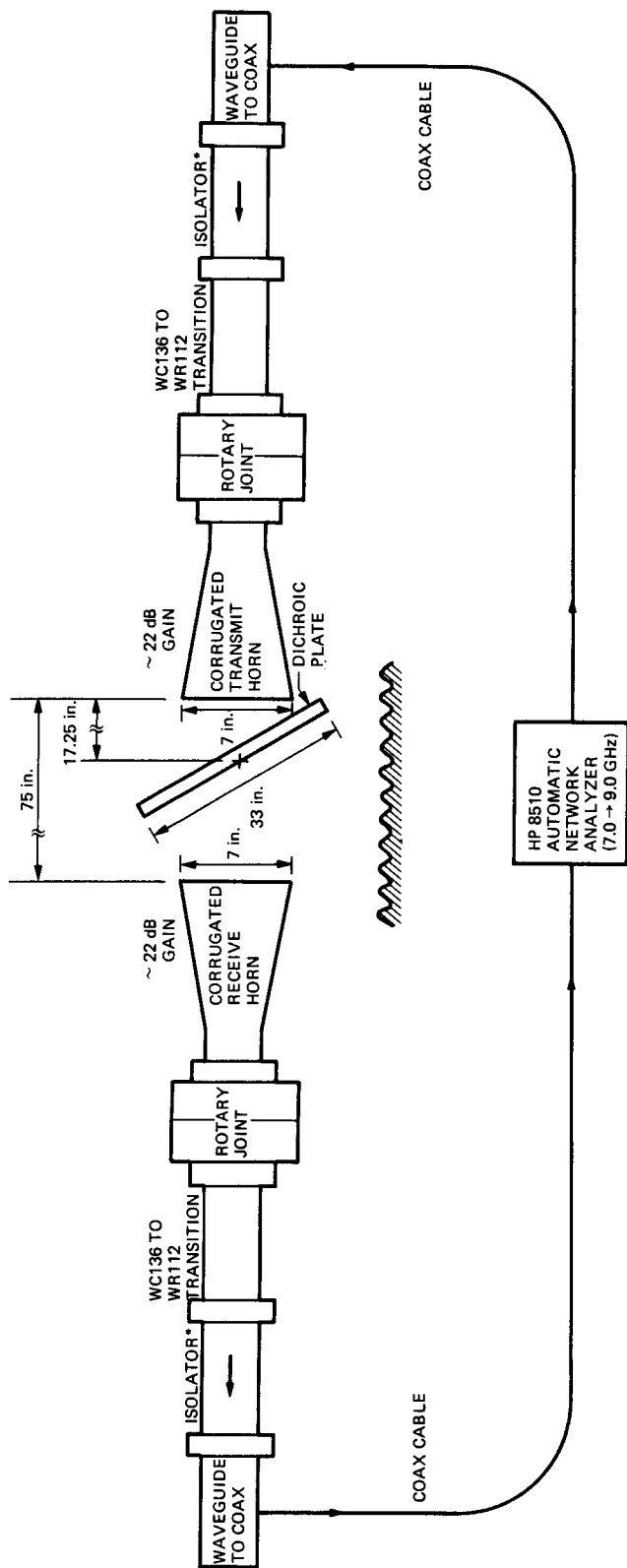


Fig. 8. Test setup block diagram for insertion loss measurements on dichroic plates at X-band

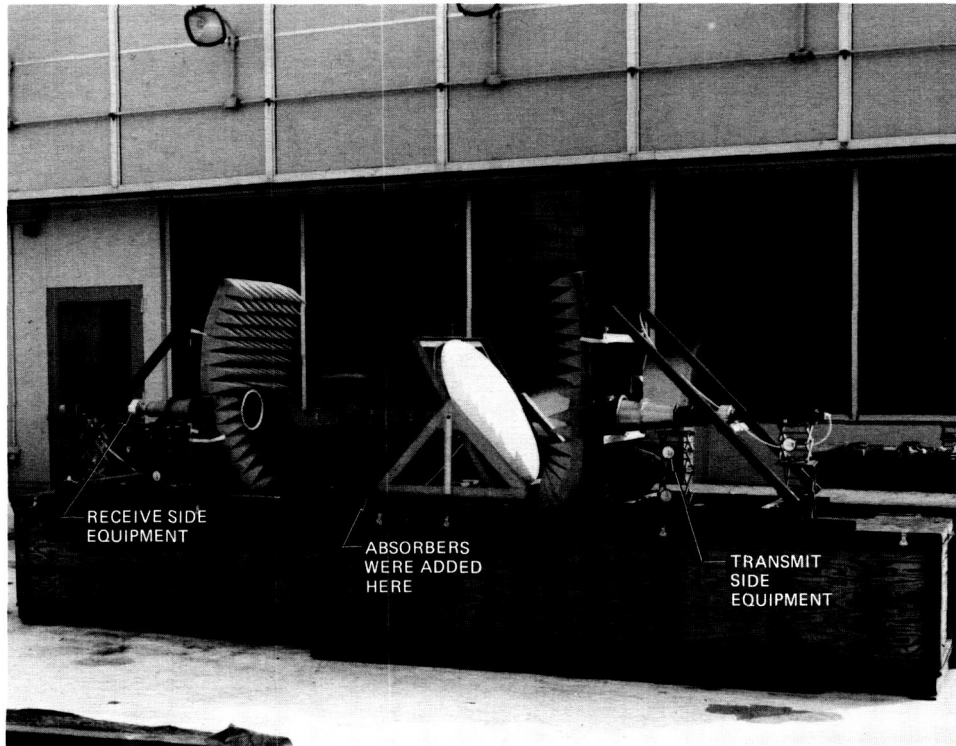


Fig. 9. Outdoor X-band test setup (roof, Building 238, JPL) for insertion loss measurements on dichroic plates; absorbers (not shown in the picture) were placed at the indicated area and on plate test fixture in order to minimize multipath signals reflecting into the receive horn

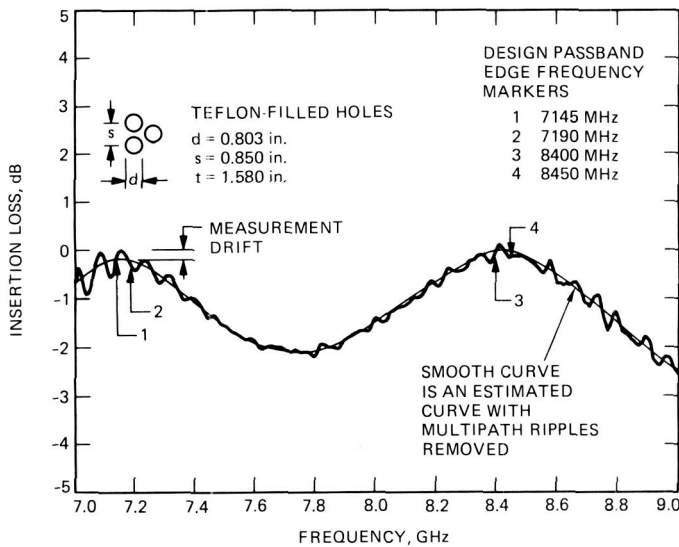


Fig. 10. Measured insertion loss of Plate #1 for perpendicular polarization, 30-deg incidence angle, and $\phi = 90$ -deg hole pattern orientation; measurements were made in the outdoor test setup (see Fig. 9)

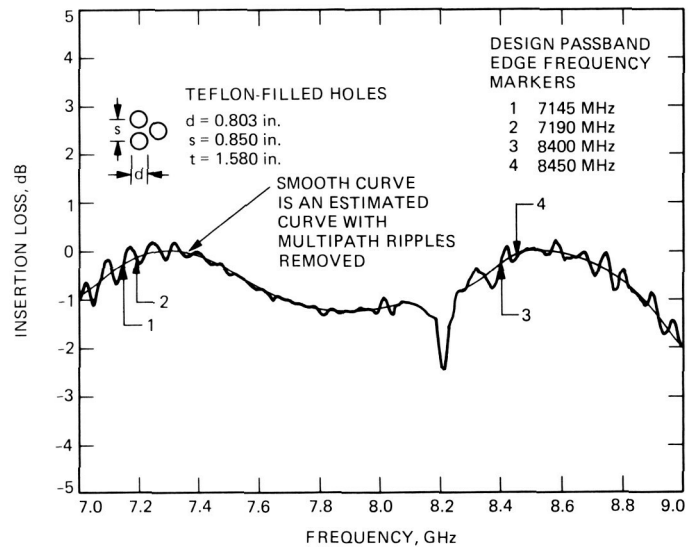


Fig. 11. Measured insertion losses of Plate #1 for parallel polarization, 30-deg incidence angle, and $\phi = 90$ -deg hole pattern orientation; measurements were made in the outdoor test setup

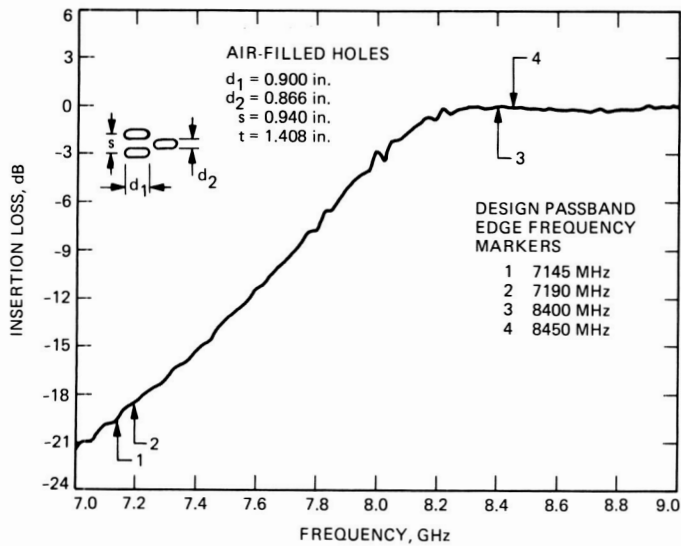


Fig. 12. Measured insertion losses of Plate #4 for perpendicular polarization, 30-deg incidence angle, and $\phi = 90$ -deg hole pattern orientation; measurements were made in the outdoor test setup

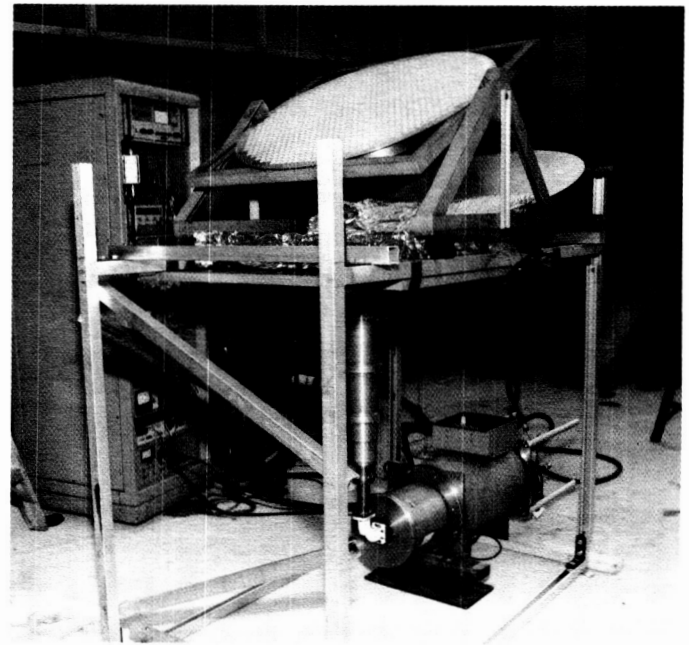


Fig. 14. Test setup for noise temperature measurements of Plate #1 for perpendicular polarization, 30-deg incidence angle, and $\phi = 90$ -deg hole pattern orientation (see Fig. 1)

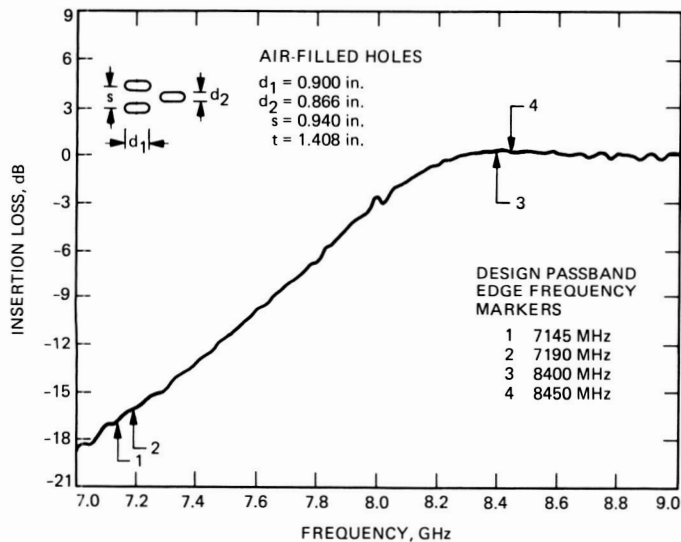


Fig. 13. Measured insertion losses of Plate #4 for parallel polarization, 30-deg incidence angle, and $\phi = 90$ -deg hole pattern orientation; measurements were made in the outdoor test setup

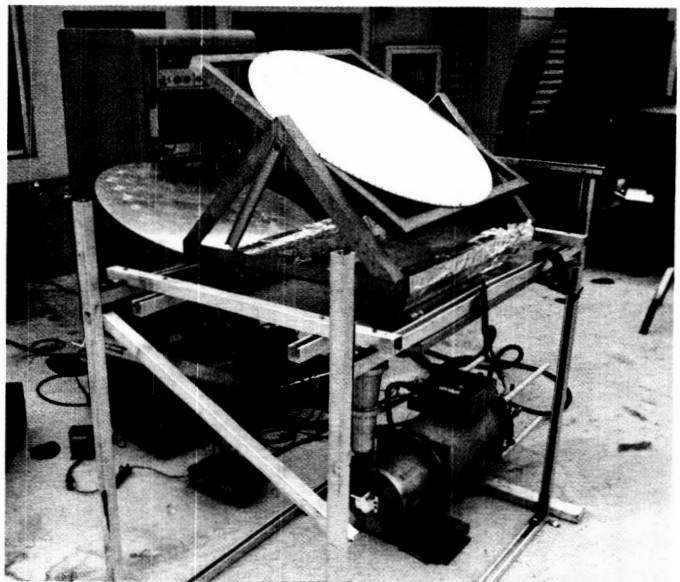


Fig. 15. Test setup for noise temperature measurements of Plate #1 for parallel polarization, 30-deg incidence angle, and $\phi = 90$ -deg hole pattern orientation (see Fig. 1)

ORIGINAL PAGE IS
OF POOR QUALITY

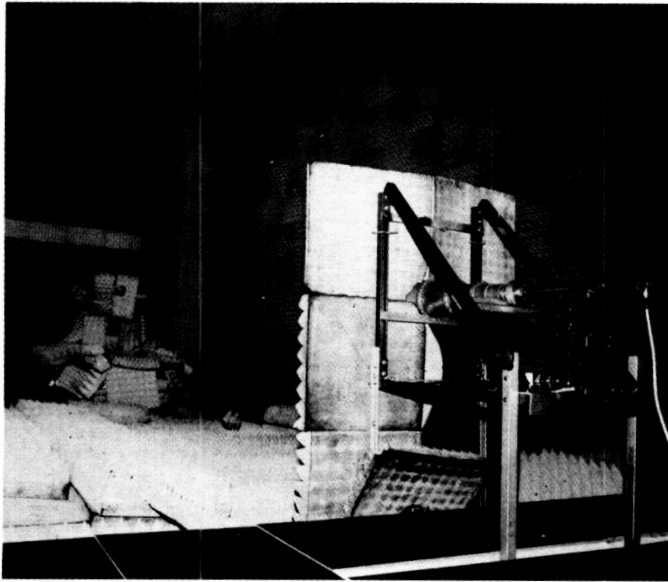


Fig. 16. Anechoic chamber antenna pattern measurement test setup for measurement of X-band horn only with axis of rotation aligned with the horn phase center



Fig. 17. Anechoic chamber antenna pattern measurement test setup for measurement of X-band horn and dichroic plate at 30-deg incidence angle with plate holder upright; in this configuration, the main reflection lobe is absorbed by the ceiling of the anechoic chamber, but scattering from plate in the orthogonal plane is recorded

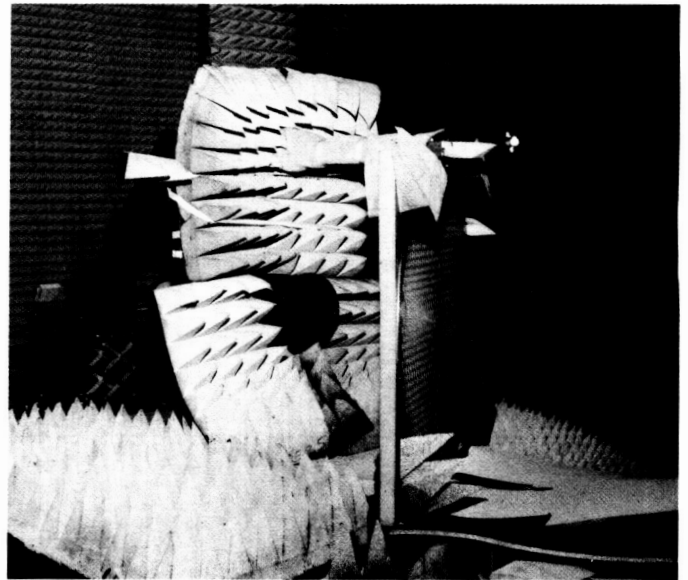


Fig. 18. Anechoic chamber antenna pattern measurement test setup for measurement of X-band horn and dichroic plate at 30-deg incidence angle with plate holder flopped 90 deg on its side; this configuration enables recording the main reflection lobe scattered from the dichroic plate

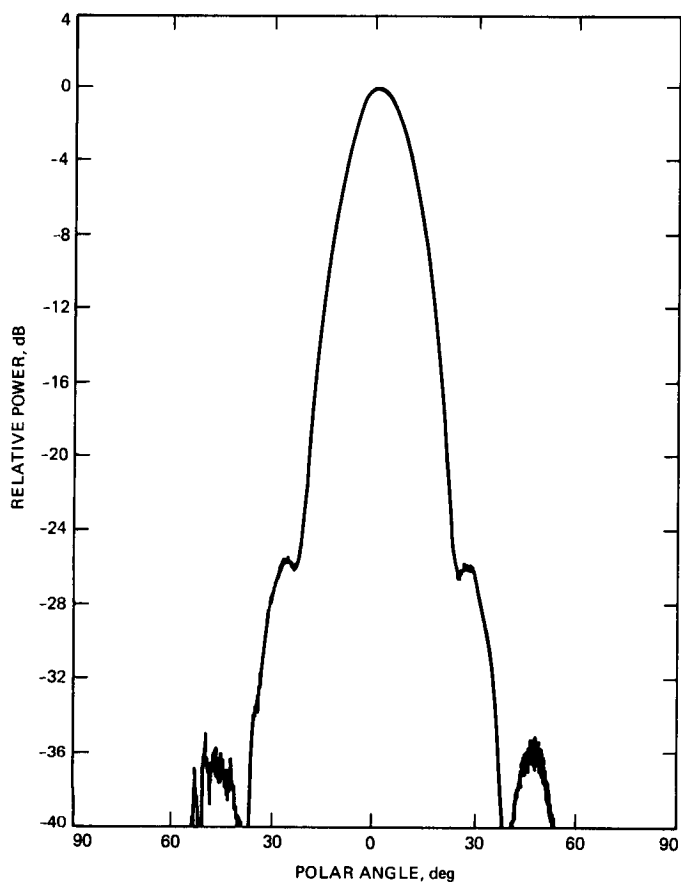


Fig. 19. E-plane pattern for X-band horn only at 7168 MHz

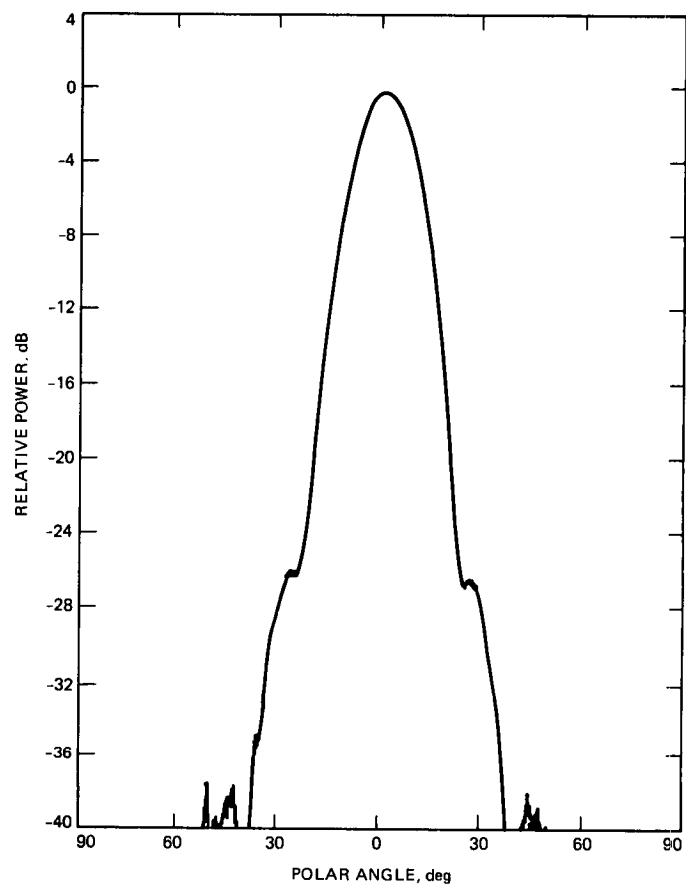


Fig. 20. H-plane pattern for X-band horn only at 7168 MHz

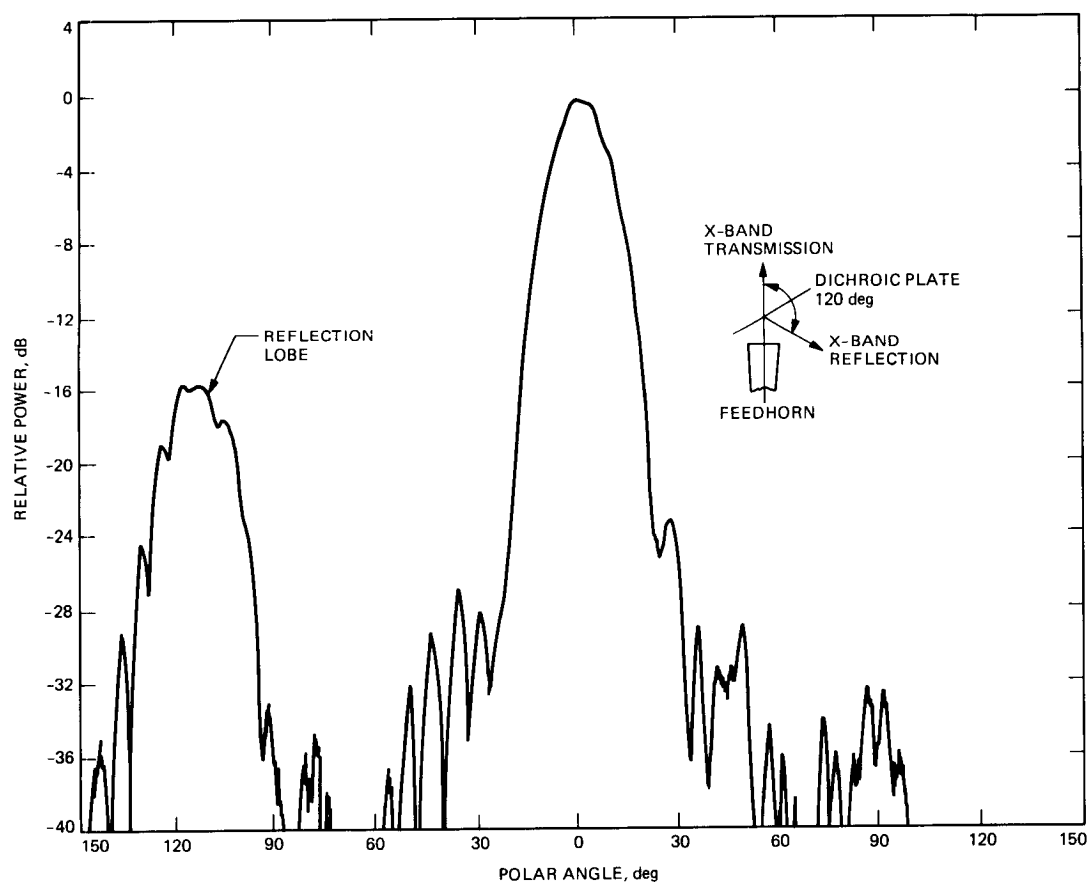


Fig. 21. E-plane pattern for X-band horn plus dichroic Plate #1 at 7168 MHz with plate holder flopped 90 deg on its side; plate is in the $\theta = 30$ deg, $\phi = 90$ deg geometry

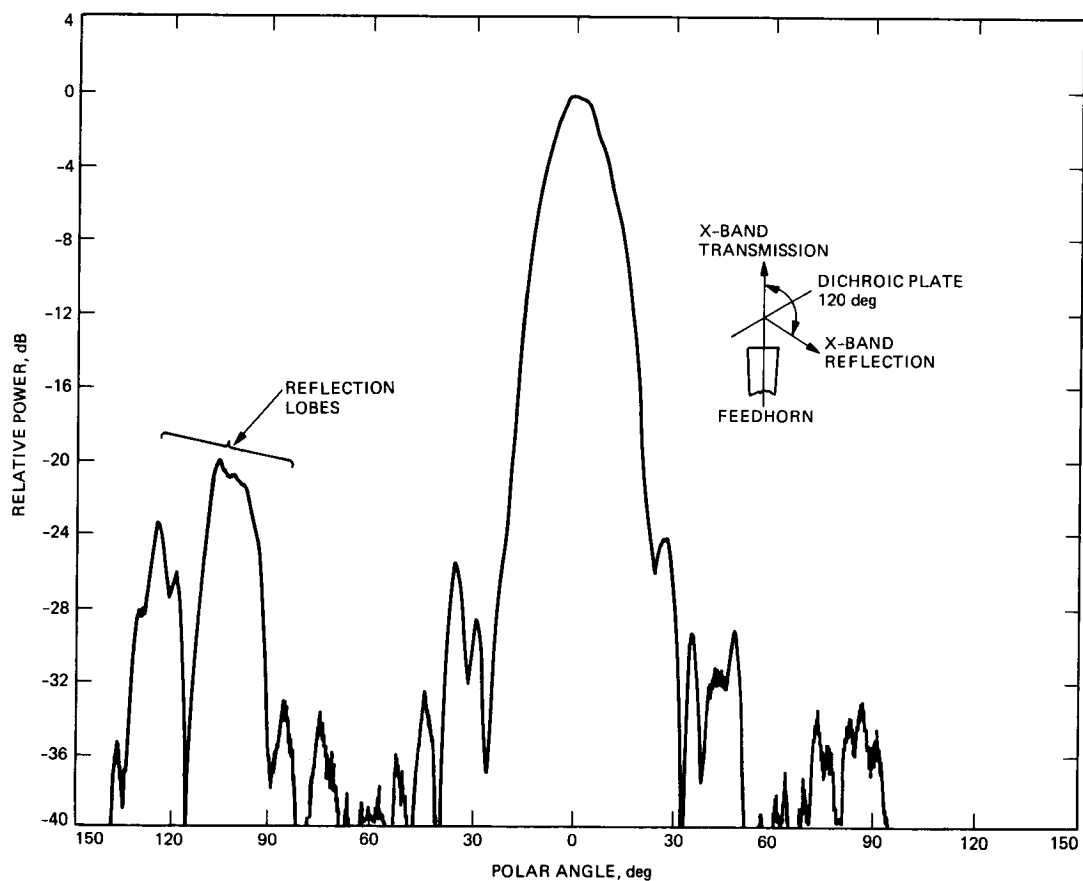


Fig. 22. H-plane pattern for X-band horn plus dichroic Plate #1 at 7168 MHz with plate holder flopped 90 deg on its side; plate is in the $\theta = 30$ deg, $\phi = 90$ deg geometry

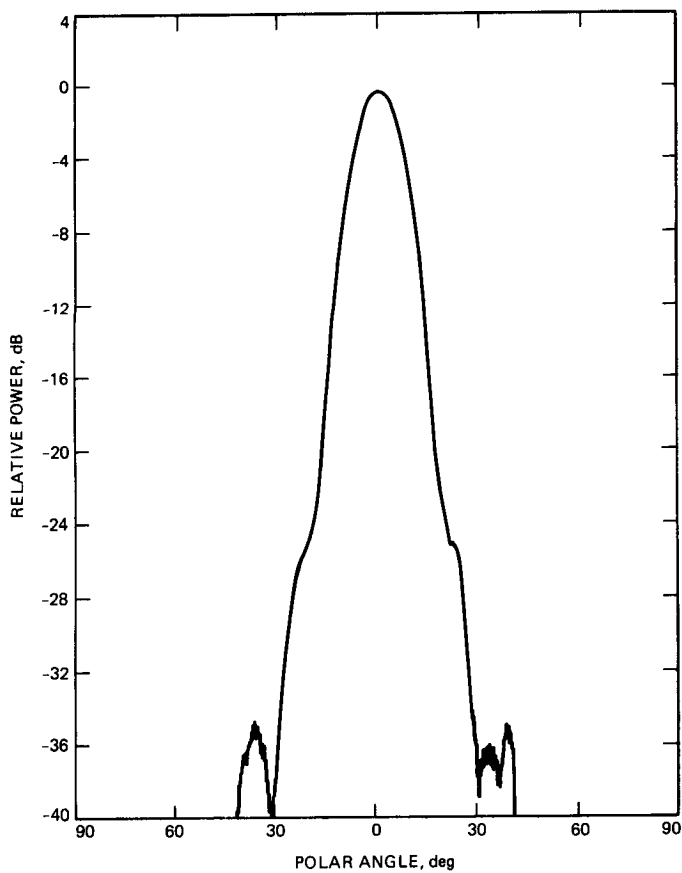


Fig. 23. E-plane pattern for X-band horn only at 8425 MHz

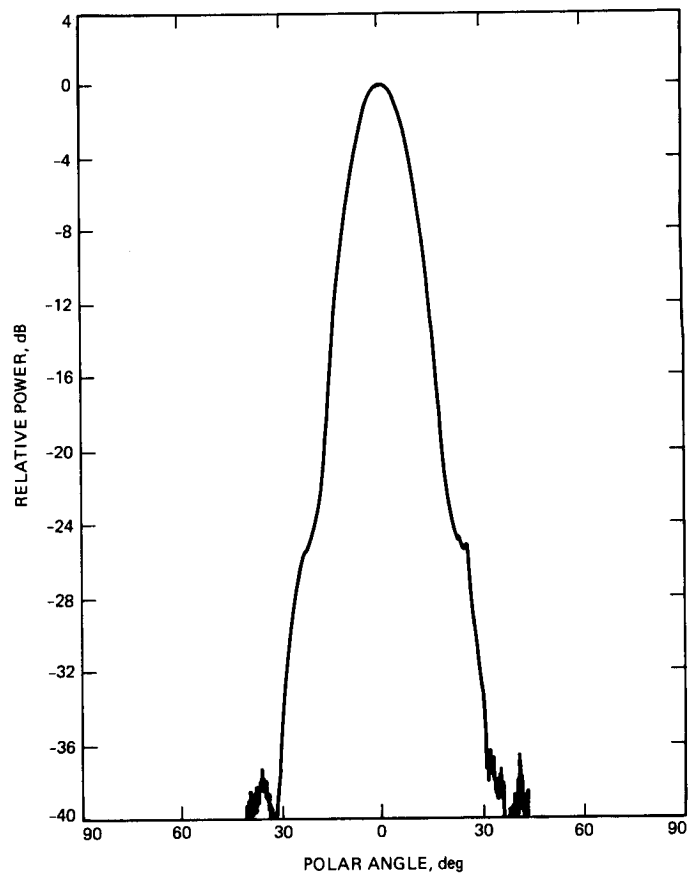


Fig. 24. H-plane pattern for X-band horn only at 8425 MHz

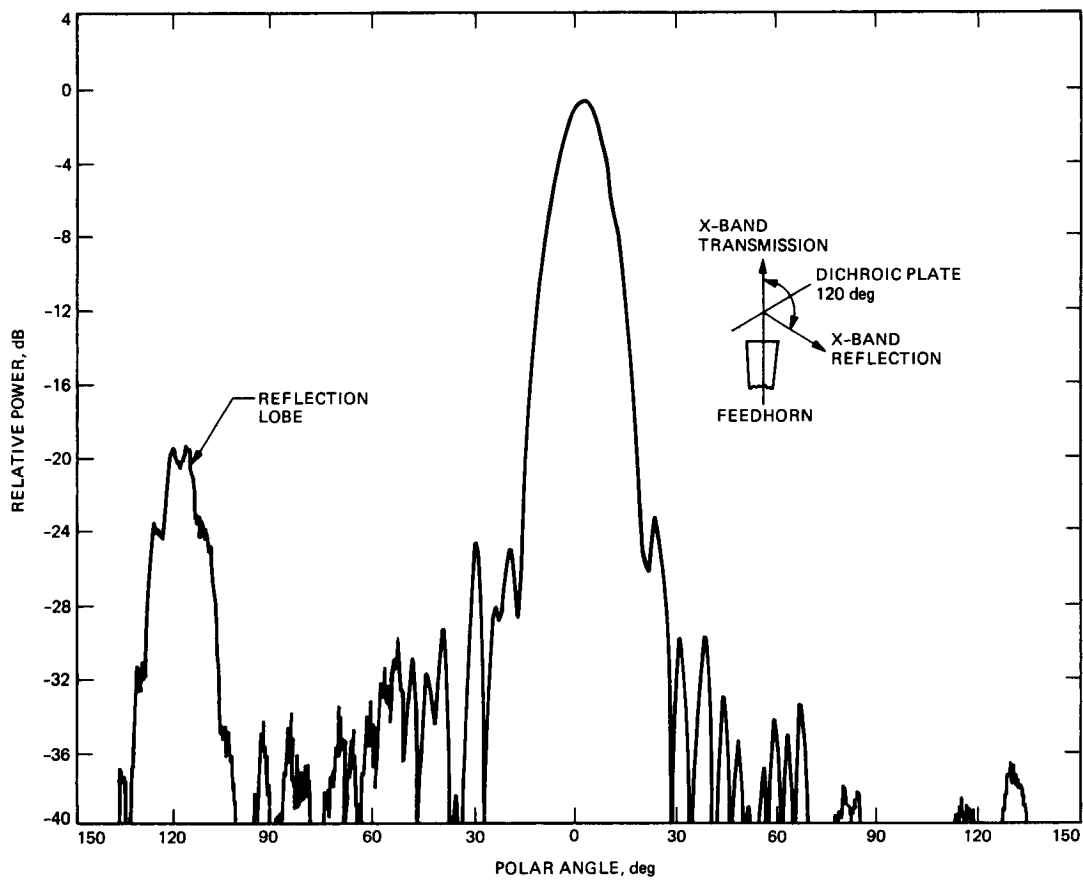


Fig. 25. E-plane pattern for X-band horn plus dichroic Plate #1 at 8425 MHz with plate holder flopped 90 deg on its side; plate is in the $\theta = 30$ deg, $\phi = 90$ deg geometry

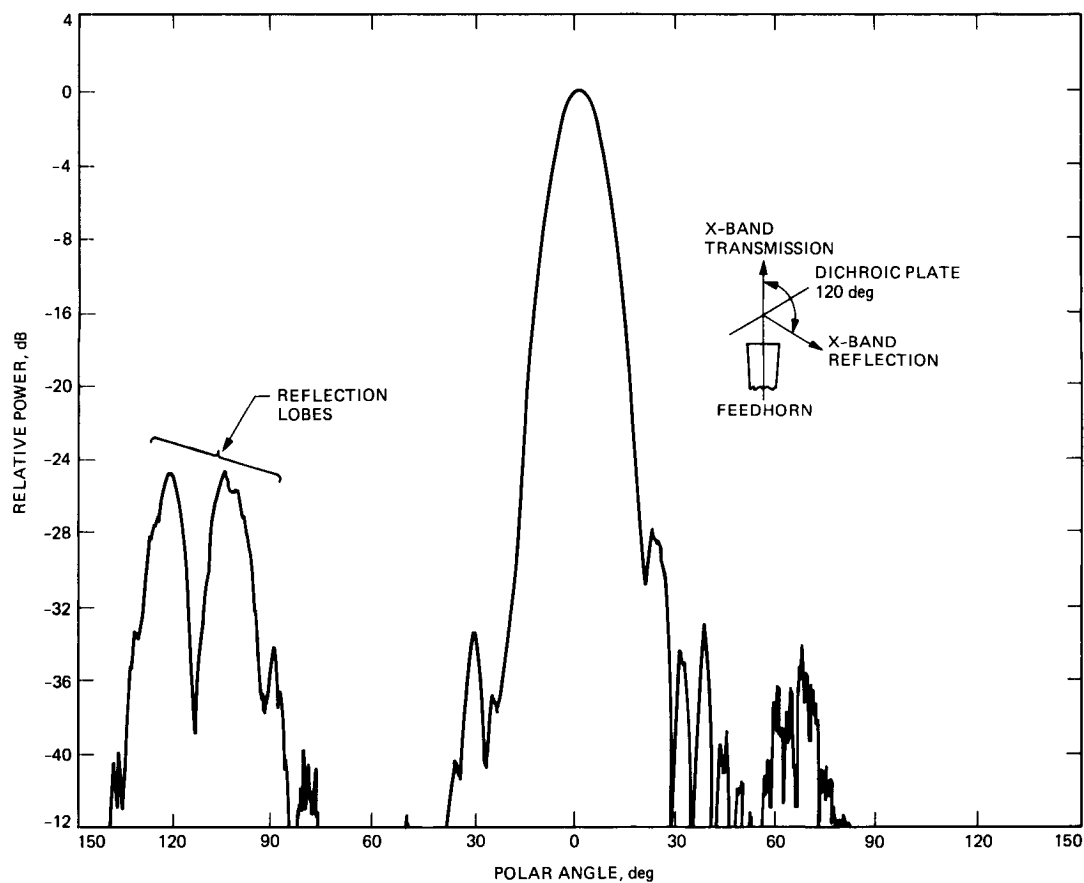


Fig. 26. H-plane pattern for X-band horn plus dichroic Plate #1 at 8425 MHz with plate holder flopped 90 deg on its side; plate is in the $\theta = 30$ deg, $\phi = 90$ deg geometry

1989000826
618387
10p.
N89 - 10197S/1-32
~~105022~~
105
April-June 1988

JPC

Antenna Pointing Compensation Based on Precision Optical Measurement Techniques

L. L. Schumacher and H. C. Vivian
Automated Systems Section

The pointing control loops of the DSN 70-meter antennas extend only to the Intermediate Reference Structure (IRS). Thus, distortion of the structure forward of the IRS due to unpredictable environmental loads can result in uncompensated boresight shifts which degrade blind pointing accuracy.

A system is described which can provide real-time bias commands to the pointing control system to compensate for environmental effects on blind pointing performance. The bias commands are computed in real time based on optical ranging measurements of the structure from the IRS to a number of selected points on the primary and secondary reflectors.

I. Background

The DSN 70-meter antenna pointing system derives the main Az-El pointing servo drive error signals from a two-axis autocollimator mounted on the Intermediate Reference Structure (IRS) [1]. The autocollimator projects a light beam to a precision mirror mounted on the Master Equatorial (ME) and detects elevation/cross-elevation errors (non-parallelism) between the two reference surfaces by measuring the angular displacement of the reflected beam. Thus, the pointing control loop extends only as far as the IRS, which is located aft of the main antenna structure near the intersection of the Az-El axes (see Fig. 1). Since the main antenna structure is outside the pointing control loop, distortion of the structure due to environmental loads can result in uncompensated RF boresight shift. While deformations and misalignments of the main paraboloid, subreflector, and feed cone can individually pro-

duce equivalent pointing shifts of nearly a degree due to environmental loads, self-compensating effects in the structural design of the antenna limit the net peak pointing offset to approximately 100 millidegrees.

The effect of structural distortion on pointing performance is illustrated in Fig. 2. Under design conditions, the main reflector paraboloid and subreflector hyperboloid axes lie on an axis normal to the IRS which passes through the paraboloid vertex. Environmental forces which distort the structure, as indicated by the broken line, cause rays reflected from different segments of the antenna to concentrate off the IRS reference axis. The objective of a pointing error compensation system design is to bias the pointing command so that the centroid of energy to and from the desired RF target falls along the IRS reference axis.

The following discussion presents a description and an analysis of a real-time optical measurement and processing concept which outputs real-time pointing system bias commands to compensate for the effects of environmentally induced structural distortions. The system optically measures the displacements of selected points on the main reflector and subreflector relative to the IRS. The discussion assumes that pointing of the IRS via the ME is error-free and deals only with the effects of environmentally induced distortions of the main paraboloid and displacements of the subreflector.

II. Environmental Effects

The principal environmental loads acting on the antenna are gravity, wind, and thermal effects. Gravity, the largest but most predictable load on the structure, causes sagging of the main paraboloid which varies as a function of elevation angle. Gravity also creates loads in the quadripod structure which displace the subreflector in the elevation direction. The effects of gravity are accounted for in the finite element models of the DSN antenna structures. Lookup tables constructed from finite element structural model data are currently used to provide an elevation pointing command bias signal to compensate for gravity-induced pointing error as a function of elevation angle.

Wind and thermal effects on the structure are less severe than gravity, but they are also much less predictable and can cause pointing errors in both the elevation and cross-elevation axes. These effects could also be included in the finite element structural models of the antennas, but in order to provide effective (real-time) blind pointing compensation, accurate and instantaneous knowledge of the true wind vector and thermal profile of the structures is required.

Through the use of lookup tables to bias the elevation pointing command to compensate for gravity effects, a blind pointing capability of 4 to 6 millidegrees rms is currently achievable in benign weather, with the accuracy degrading to about 12 millidegrees rms under moderately windy conditions. With the recently completed upgrades to enlarge and improve the shape of the large DSN tracking antennas in combination with future plans to quadruple the upper operating frequency, it will be necessary to improve blind pointing precision to 1 millidegree rms to fully realize the benefits of the upgrades for 32-GHz (Ka-band) operation.

III. Measurement Concept

A system of multiple point-to-point measurements using optical ranging techniques was conceived as a method for measuring and estimating the effects of structural distortions on

blind pointing accuracy in real time. It was recognized at the outset that a large number of measurement configurations are possible and that trade-offs exist between the number, location, and accuracy of measurements. To establish a baseline, a simple configuration was analyzed to determine the measurement accuracy required to meet a blind pointing accuracy of one millidegree rms. The analysis was divided into two parts: (1) a determination of the required measurement resolution; and (2) an evaluation of the effects of systematic (i.e., scale factor) errors on pointing offset determination. Also, because the main paraboloid boresight axis and subreflector positions can shift relative to each other, separate analyses of measurement resolution and systematic error requirements were made for these two major elements of the Cassegrain system. For analysis purposes, the measurement baseline was assumed to be located at the main paraboloid vertex. Since the IRS is actually located approximately 7 meters aft of the vertex, implementation of the system will require transfer of the IRS reference plane to the baseline measurement plane by optical or other methods. The main paraboloid and quadripod structures are considered to be flexible, whereas the subreflector itself is assumed to be a rigid assembly.

The SHAPES (Spatial High Accuracy Position Encoding Sensor) system under development at JPL [2] was taken as the baseline sensor for the measurement system since it is capable of multiple simultaneous high-speed (10 frames/sec) ranging measurements with accuracies at the submillimeter level. SHAPES is a time-of-flight laser ranging sensor which measures ranges from multiple sources to retroreflector targets placed at any desired (unobstructed) location throughout the structure. Each SHAPES "head" can accommodate up to 30 sources, whose actual locations can be remote from the head and optically connected through optical fibers. The measurement configuration employed for the analyses uses two-dimensional triangulation techniques to precisely determine shifts in the positions of preset targets relative to IRS coordinates due to structural distortions caused by environmental forces. From these, the "structural" boresight of the main paraboloid and the deviation of the subreflector relative to the IRS are determined by geometric calculations. These two quantities must be combined to determine the net shift of the boresight due to environmental forces. This determination is then used in computing pointing bias commands to compensate for structural distortions.

IV. Primary Reflector Structural Boresight Axis Determination

A. Measurement Configuration

The optical measurement configuration chosen as a test case for the analysis below is shown in Fig. 3. The configura-

tion was selected as one which maximizes the amount of useful boresight determination data for the number of measurements made. In this design, three SHAPES sources are placed in a precise pattern centered on the physical vertex of the main paraboloid at the base of the feed cone and aligned parallel to the elevation-bearing axis. They have unobstructed views of four retroreflectors equally spaced around the perimeter of the main dish. Two retroreflectors are placed on the antenna rim diametrically opposite each other such that the plane passing through the retros and the vertex is parallel to the elevation axis. (This is referred to as the elevation-boresight plane.) This retro placement is chosen to provide the greatest detection sensitivity to distortions caused by forces acting parallel to the surface of the Earth (i.e., crosswinds). Two other retroreflectors are located diametrically opposite each other such that the plane (e.g., the cross-elevation-boresight plane) passing through the retros and the vertex is orthogonal to the elevation axis. This set of retros provides the greatest detection sensitivity to distortions caused by gravity. The key assumption made is that the intersection of the planes containing the antenna vertex and the retroreflectors on the antenna rim represents the geometric boresight axis of the antenna dish surface distorted by environmental loads. A means of identifying that boresight axis using SHAPES measurements is described below.

B. Main Reflector Boresight Shift Analysis

The SHAPES sources and retroreflectors are identified in Fig. 4. The three SHAPES sources, *A*, *B*, and *C*, are located on the main dish surface in the RF shadow of the quadripod at the base of the feed cone. Four retroreflectors, 1, 2, 3, and 4, are placed on the antenna rim — numbers 1 and 2 in the cross-elevation-boresight plane and numbers 3 and 4 in the elevation-boresight plane. Measurements across the dish (e.g., from point 1 to point 2 or from point 1 to point 3) are made via relay mirrors placed at rim position numbers 1 and 3.

The analytic determination of the effective structural boresight of the deformed main paraboloid is divided into three parts:

- (1) Calculate the intersection of the elevation-boresight and cross-elevation-boresight planes in the antenna rim plane.
- (2) Calculate the antenna rim plane intersection coordinates in the pointing reference plane.
- (3) Calculate the angular orientation of the effective structural boresight axis in pointing reference coordinates.

1. Calculation of the rim-plane intersection. Figure 4 illustrates the range measurements in the antenna rim plane needed to calculate the coordinates of intersection *J* relative to retro-

reflectors 1, 2, 3, and 4. The objective is to compute R_{J3} and R_{J1} given range measurements R_{12} , R_{34} , R_{13} , R_{14} , and R_{23} . The procedure is outlined below.

Calculate angles θ_1 and θ_2 by the law of cosines:

$$\theta_1 = \cos^{-1} \left[\frac{R_{12}^2 + R_{13}^2 - R_{23}^2}{2 \cdot R_{12} \cdot R_{13}} \right] \quad (1)$$

$$\theta_2 = \cos^{-1} \left[\frac{R_{13}^2 + R_{34}^2 - R_{14}^2}{2 \cdot R_{13} \cdot R_{34}} \right] \quad (2)$$

Calculate angle θ_3 :

$$\theta_3 = 180 - (\theta_1 + \theta_2) \quad (3)$$

Calculate intersection *J* locations R_{J3} and R_{J1} :

$$R_{J3} = R_{13} \cdot \frac{\sin \theta_1}{\sin \theta_3} \quad (4)$$

$$R_{J1} = R_{13} \cdot \frac{\sin \theta_2}{\sin \theta_3} \quad (5)$$

2. Calculation of intersection coordinates in the pointing reference plane. Figure 5 illustrates the placement of SHAPES sources *A*, *B*, and *C* and the projections of retroreflectors 1, 2, 3, and 4 on the pointing reference plane. Given range measurements R_{AB} , R_{BC} , R_{A1} , R_{B1} , R_{A2} , R_{B2} , R_{B3} , R_{B4} , R_{C3} , R_{C4} , R_{12} , and R_{34} , it is possible to locate the intersection *J* in the pointing reference plane in coordinates orthogonal to R_{AB} and R_{BC} in the following manner:

Calculate angles θ_4 and θ_5 :

$$\theta_4 = \cos^{-1} \left[\frac{R_{BC}^2 + R_{B3}^2 - R_{C3}^2}{2 \cdot R_{BC} \cdot R_{B3}} \right] \quad (6)$$

$$\theta_5 = \cos^{-1} \left[\frac{R_{BC}^2 + R_{B4}^2 - R_{C4}^2}{2 \cdot R_{BC} \cdot R_{B4}} \right] \quad (7)$$

Calculate the projections of retros 3 and 4 from point *B* along line *BC*:

$$K_3 = R_{B3} \cdot \cos \theta_4 \quad (8)$$

$$K_4 = R_{B4} \cdot \cos \theta_5 \quad (9)$$

Calculate the projection of the elevation component (J_Y) of intersection J relative to the vertex in pointing reference plane coordinates:

$$J_Y = K_3 + \frac{C_1}{R_{34}} (K_3 - K_4) - \frac{R_{BC}}{2} \quad (10)$$

Calculate angles θ_6 and θ_7 :

$$\theta_6 = \cos^{-1} \left[\frac{R_{AB}^2 + R_{B1}^2 - R_{A1}^2}{2 \cdot R_{AB} \cdot R_{B1}} \right] \quad (11)$$

$$\theta_7 = \cos^{-1} \left[\frac{R_{AB}^2 + R_{B2}^2 - R_{A2}^2}{2 \cdot R_{AB} \cdot R_{B2}} \right] \quad (12)$$

Calculate the projections of retros 1 and 2 from point B along line AB in the pointing reference plane:

$$K_1 = R_{B1} \cdot \cos \theta_6 \quad (13)$$

$$K_2 = R_{B2} \cdot \cos \theta_7 \quad (14)$$

Calculate the projection of the cross-elevation component (J_X) of intersection J relative to the physical vertex in pointing reference plane coordinates:

$$J_X = K_1 + \frac{C_2}{R_{12}} \cdot (K_1 - K_2) - \frac{R_{AB}}{2} \quad (15)$$

3. Calculation of the structural boresight error angles.

There are now sufficient data to determine the angular deviation of the structural boresight axis relative to the reference boresight axis. Figure 6 illustrates the reference boresight axis in a pointing reference plane, fixed coordinate frame whose origin is at the physical vertex of the main paraboloid. The elevation and cross-elevation error angle components of the main paraboloid structural boresight axis relative to the reference boresight axis are

$$\alpha_{el} = \tan^{-1} \frac{J_Y}{H} \quad (16)$$

$$\alpha_{xel} = \tan^{-1} \frac{J_X}{H} \quad (17)$$

where H = the depth of the main paraboloid from the rim plane to the surface at the vertex.

V. Subreflector Displacement Determination

A. Measurement Configuration

Figure 7 shows the baseline measurement configuration for determining the orientation of the subreflector relative to the pointing reference coordinate system. Figure 8 is a view looking into the antenna boresight axis illustrating the placement of three retroreflectors on the subreflector backup structure as viewed from three SHAPES sources located on the pointing reference plane at the base of the feed cone. Two SHAPES sources and two retroreflectors define a plane. Hence, three SHAPES sources, A , B , and C , in combination with three retroreflectors, 5, 6, and 7, define two planes with one common SHAPES-retroreflector pair. A reference coordinate frame is defined with SHAPES defining the X - Y axes and with the Z axis orthogonal to both the X and Y axes. The analysis below solves for the subreflector displacements in the pointing reference coordinate frame.

B. Subreflector Displacement Analysis

Figure 9 is an elevation view of the antenna parallel to a plane defined by two SHAPES sources, A and B , located on the pointing reference plane, and by two retroreflectors, 5 and 6, mounted on the subreflector backup structure. The interior angles θ_9 , θ_{10} , θ_{11} , and θ_{12} can be determined as follows:

$$\theta_9 = \cos^{-1} \left[\frac{R_{A6}^2 + R_{AB}^2 - R_{B6}^2}{2 \cdot R_{A6} \cdot R_{AB}} \right] \quad (18)$$

$$\theta_{10} = \cos^{-1} \left[\frac{R_{A5}^2 + R_{A6}^2 + R_{56}^2}{2 \cdot R_{A5} \cdot R_{A6}} \right] \quad (19)$$

$$\theta_{11} = \cos^{-1} \left[\frac{R_{B6}^2 + R_{AB}^2 - R_{A6}^2}{2 \cdot R_{B6} \cdot R_{AB}} \right] \quad (20)$$

$$\theta_{12} = \cos^{-1} \left[\frac{R_{A6}^2 + R_{56}^2 - R_{A5}^2}{2 \cdot R_{A6} \cdot R_{56}} \right] \quad (21)$$

The lateral subreflector displacement parallel to the line between SHAPES sources in the pointing reference plane is

$$\beta_1 = 90 - \theta_{11} \quad (22)$$

$$\Delta Y = R_{B6} \cdot \sin \beta_1 \quad (23)$$

The rotational angle of the subreflector about an axis orthogonal to the plane of Fig. 9 is

$$\beta_2 = \theta_9 - \theta_{12} \quad (24)$$

The same procedure is used to determine the translational and rotational displacements in the other plane defined by SHAPES sources *B* and *C* and retroreflectors 6 and 7.

VI. Error Sources

Both random and systematic errors affect the accuracy in the determination of antenna pointing parameters. Random errors are attributed primarily to measurement instrument resolution but are also caused by environmental effects. Scale factor variations are representative of systematic errors which affect all measurements in proportion to their magnitude. Because the scale factor of SHAPES is determined by a precision crystal oscillator, the contribution of the instrument to systematic measurement errors is insignificant. However, as with any optical time-of-flight ranging system, the measurement scale factor is affected by the speed of light in the working medium (air). The speed of light in air is inversely proportional to the index of refraction, *n*, which depends on the temperature, pressure, humidity, and carbon dioxide content, as well as on other factors. The index of refraction of air is given in [3] as

$$n = 1 + \left[\left(\frac{78.6 + 42.4(y - 0.0003)P}{TZ} \right) - \left(0.00042 \cdot K \cdot es \cdot H \right) \right] 10^{-8} \quad (25)$$

where

n = refractive index of air

y = fractional concentration of carbon dioxide by volume

P = atmospheric pressure in pascals

T = temperature in kelvins

Z = compressibility factor

H = relative humidity

K = vapor pressure enhancement factor

es = saturation vapor pressure of water vapor over water

If the values of the environmental parameters above were known exactly, the uncertainty in the determination of the refractive index of air would be estimated to be between 0.05 and 0.1 ppm due to the empirical nature of the equation. Therefore, the key environmental parameters must be measured at the antenna site to minimize time-of-flight uncertainties. On-site measurement of parameters is expected to yield a global index-of-refraction determination accuracy of 0.1 to 1.0 ppm. In addition to global uncertainties, variations due to local turbulence can add errors to individual measurements.

The random uncertainties of SHAPES have been determined by laboratory measurements to be 25 μ m rms. Added to these are measurement uncertainties due to local index-of-refraction variations, which are also estimated to be approximately 25 μ m rms. Since the configuration analyzed requires 14 measurements to determine the structural boresight error of the main reflector plus 7 measurements to determine the orientation of the subreflector, computer programs were needed to individually evaluate the effects of both systematic and random measurement errors on the determination of geometric pointing parameters relative to the pointing reference coordinate frame. In the random error determination programs, a measurement error increment was individually added or subtracted to each range measurement, and all possible combinations were evaluated to determine the worst case pointing parameter determination error.

VII. Error Analysis Results

Figure 10 shows the worst case main reflector boresight determination error for random SHAPES measurement resolution errors from 1×10^{-6} to 100×10^{-6} meters. A plot of the boresight determination error due to systematic measurement errors ranging from 1 ppm to 100 ppm is shown in Fig. 11. Results are plotted in Figs. 12 and 13 for subreflector lateral displacement determination errors resulting from random and systematic measurement errors. Subreflector tilt determination errors due to random and systematic measurement errors are shown in Figs. 14 and 15.

Main reflector boresight determination errors due to random measurement errors have the most significant impact on the performance of the pointing compensation configuration analyzed. The effects of expected measurement error levels on other factors affecting pointing compensation system performance are one-tenth as great or less. Based on a ~ 35 - μ m-

rms overall measurement resolution (e.g., 25- μ m-rms SHAPES random error combined with 25 μ m rms of random noise due to local index-of-refraction variations), Fig. 10 shows that the main reflector structural boresight determination error is approximately 3 millidegrees. This represents the performance for the worst case combination of measurement errors and is considered to be conservative by a factor of about four.

VIII. Conclusions

This article covers work performed during fiscal year 1987. While the results are not fully conclusive, they do suggest that refinements of the technical approach have the potential to meet the 1-millidegree blind pointing accuracy required for full realization of 32-GHz (Ka-band) performance improvement.

References

- [1] M. D. Nelson, J. R. Schroeder, and E. F. Tubbs, "Optical Links in the Angle-Data Assembly of the 70-Meter Antennas," *TDA Progress Report 42-92*, vol. October-December 1987, Jet Propulsion Laboratory, Pasadena, California, pp. 154-165, February 15, 1988.
- [2] J. McLauchlan, W. Goss, and E. Tubbs, *SHAPES: A Spatial, High Accuracy, Position Encoding Sensor for Space System Control Applications*, American Astronautical Society Publication 82-032, January 20, 1983.
- [3] W. T. Estler, "High Accuracy Displacement Interferometry in Air," *Applied Optics*, vol. 24, no. 6, pp. 808-815, March 15, 1985.

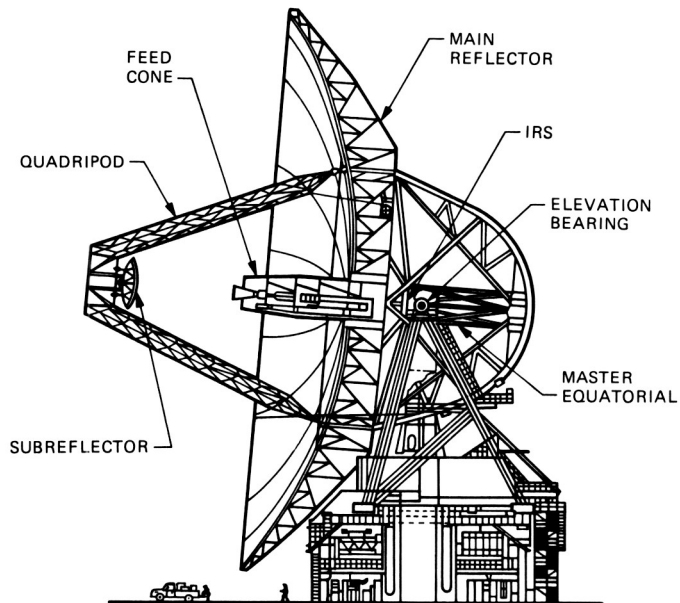


Fig. 1. The 70-meter DSN antenna

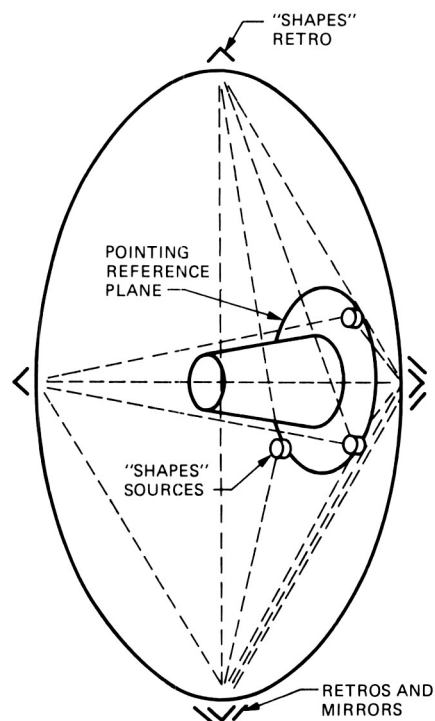


Fig. 3. Main reflector baseline measurement configuration

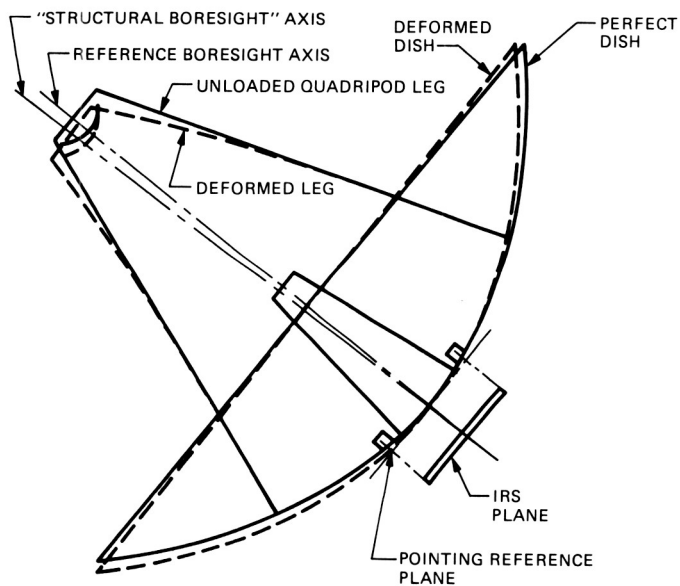


Fig. 2. Antenna deformation under environmental loads

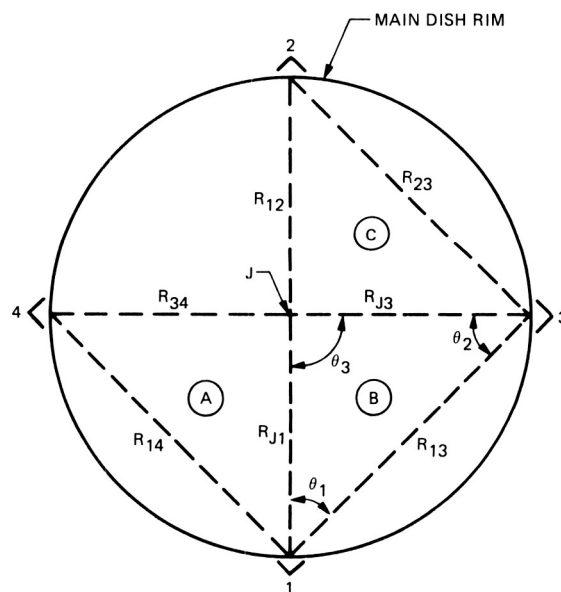


Fig. 4. Rim plane intersection (J) measurements

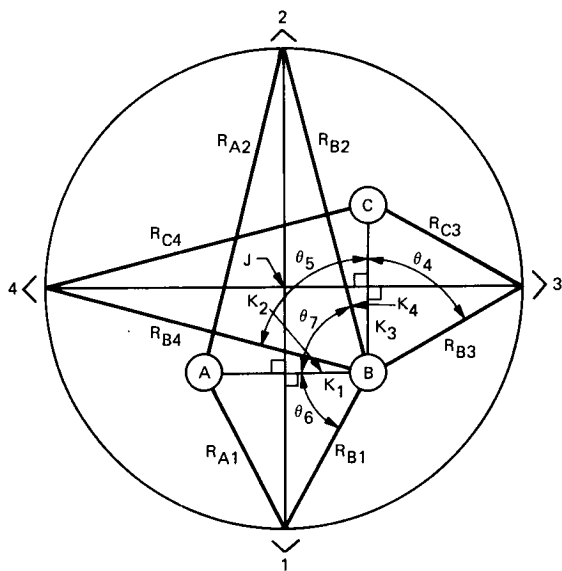


Fig. 5. Determination of J in pointing reference plane

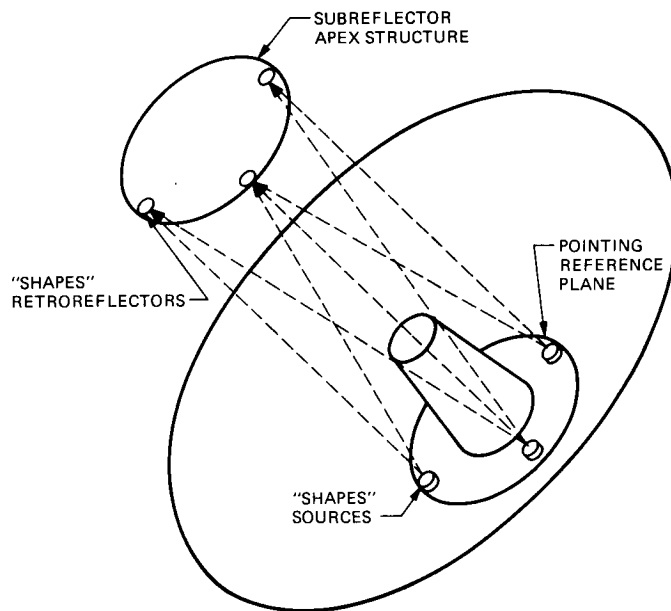


Fig. 7. Subreflector baseline measurement configuration

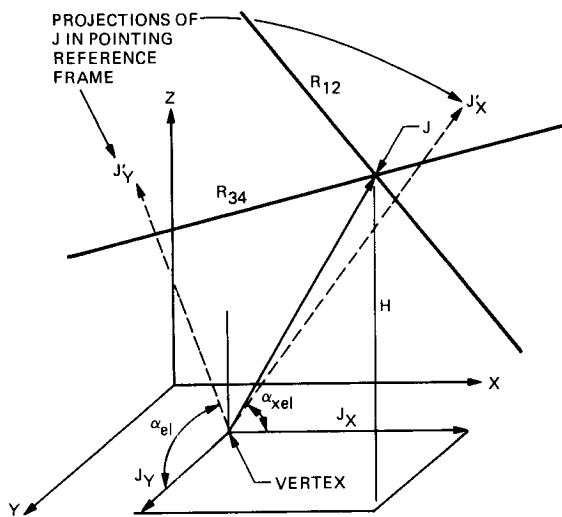


Fig. 6. Location of structural boresight axis in pointing reference coordinate frame

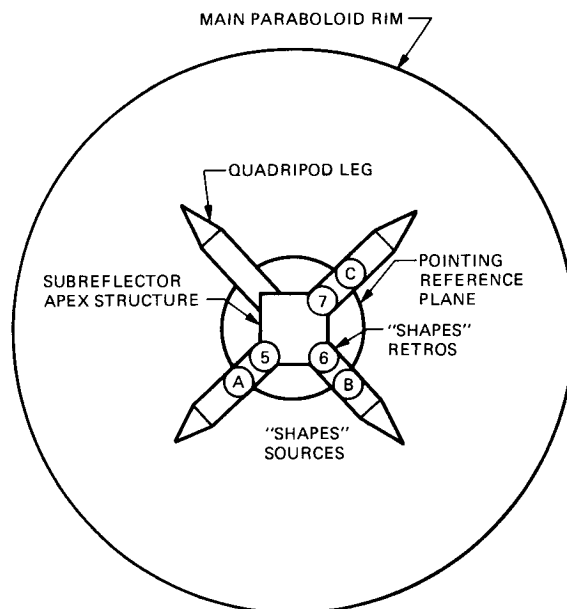


Fig. 8. Placement of SHAPES sources and retro-reflectors for subreflector measurements

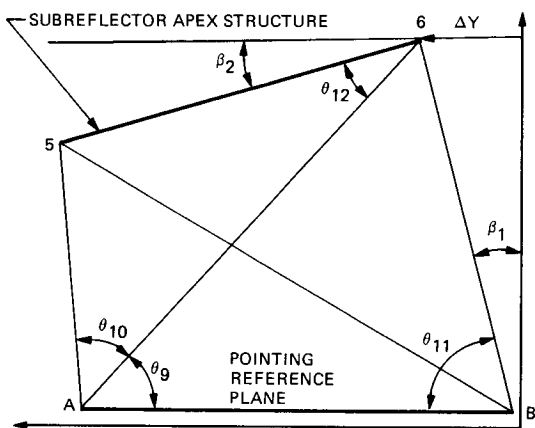


Fig. 9. Subreflector displacement measurement geometry

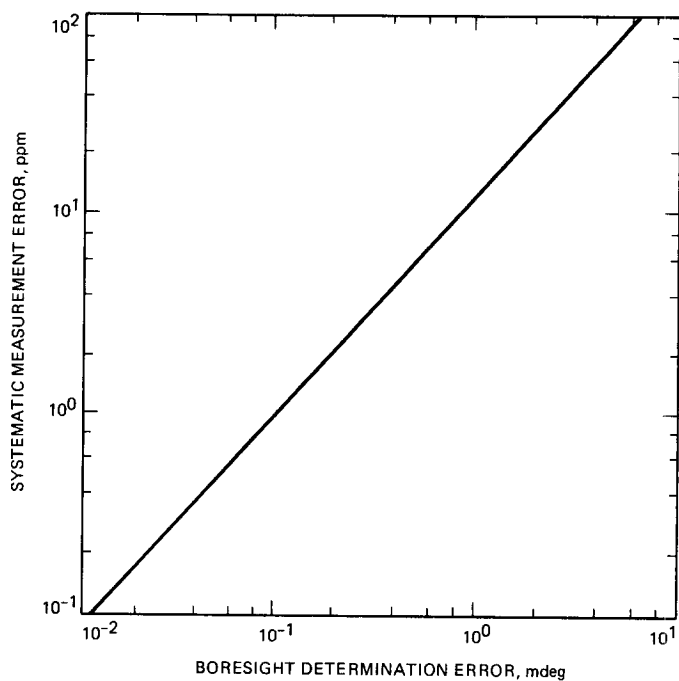


Fig. 11. Main reflector systematic boresight determination error

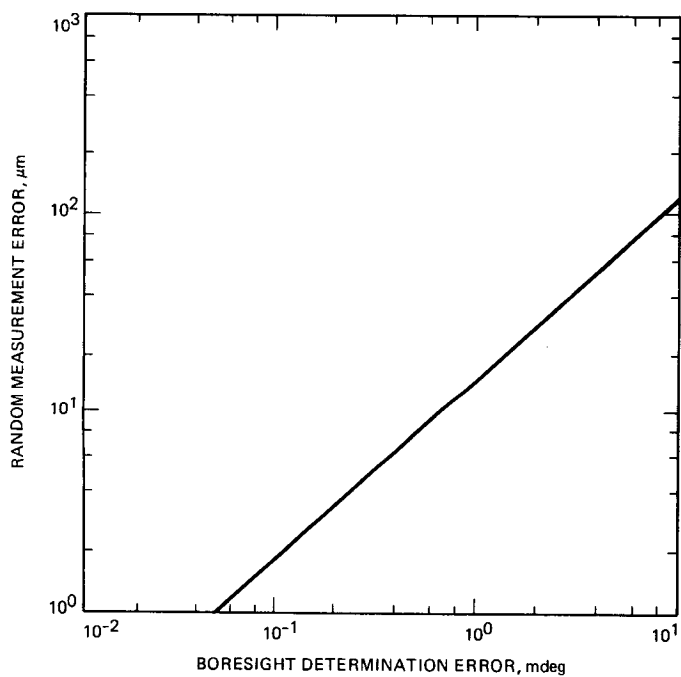


Fig. 10. Main reflector random boresight determination error

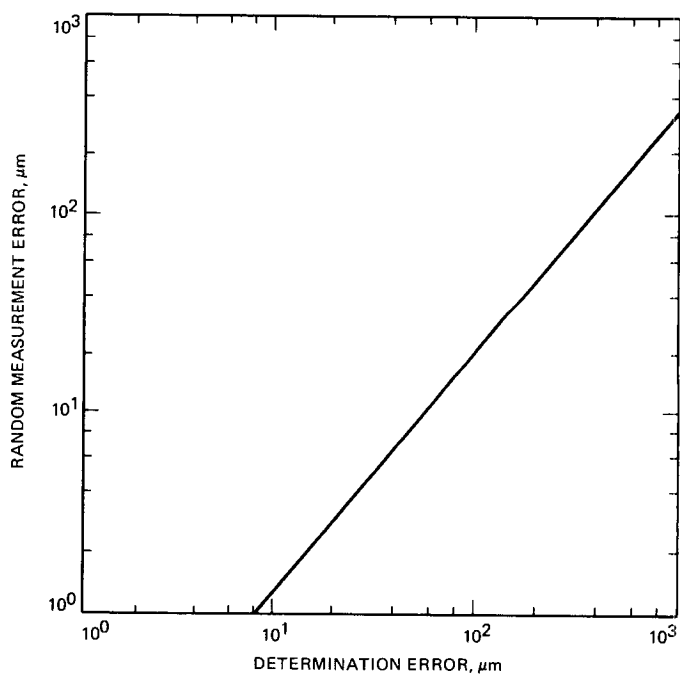


Fig. 12. Subreflector translation random determination error

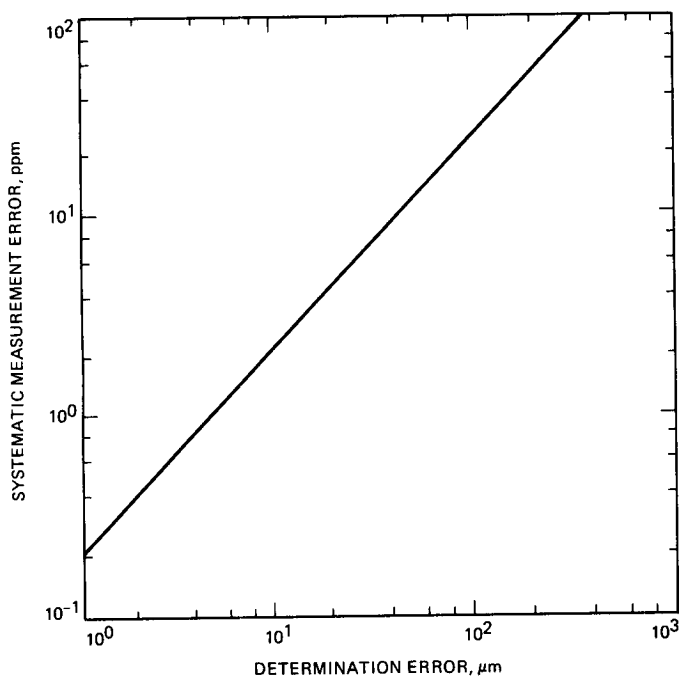


Fig. 13. Subreflector translation systematic determination error

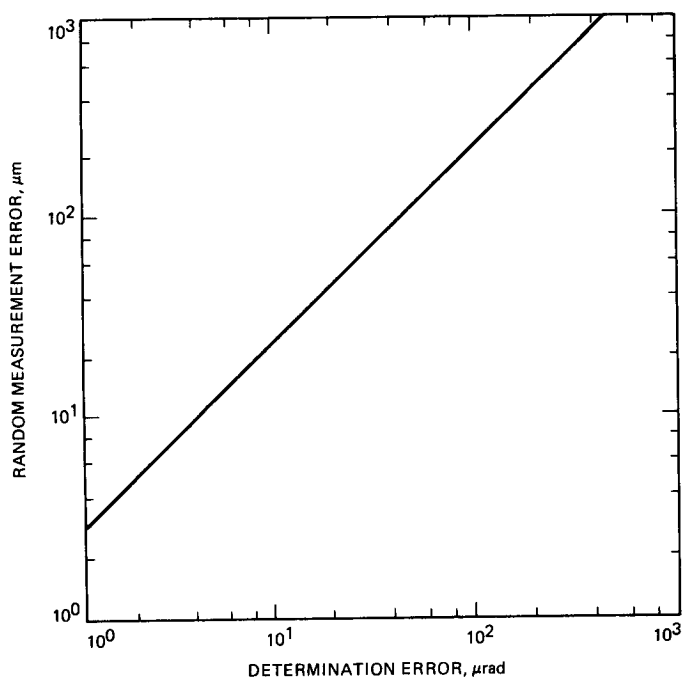


Fig. 14. Subreflector tilt random determination error

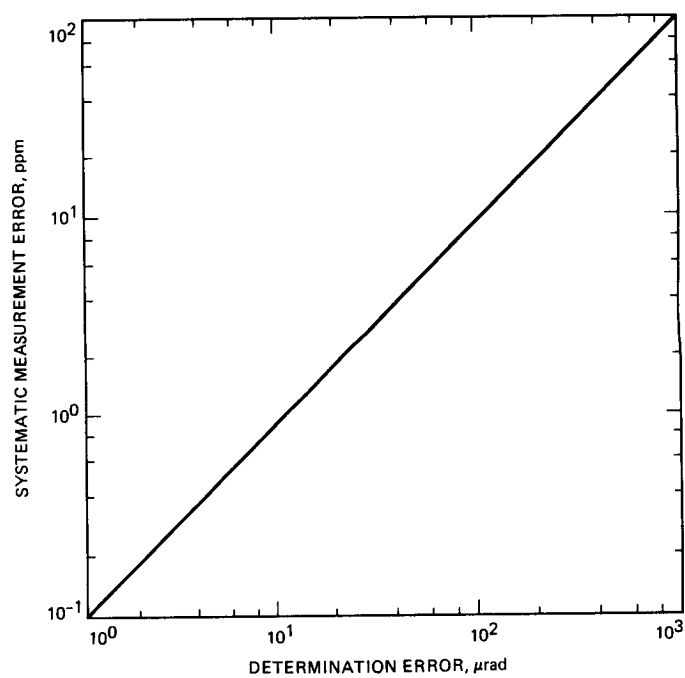


Fig. 15. Subreflector tilt systematic determination error

A 32-GHz Reflected-Wave Maser Amplifier With Wide Instantaneous Bandwidth

J. Shell and D. Neff

Radio Frequency and Microwave Subsystems Section

An eight-stage, 32-GHz reflected-wave ruby maser has been built. The maser operates in a 3-watt closed-cycle refrigerator (CCR) at 4.5 K and is capable of 21 dB of net gain with an instantaneous bandwidth of 400 MHz. The input noise temperature referred to the room-temperature flange is approximately 21 K.

I. Introduction

In recent years, NASA has been evaluating the relative advantages of 8.4-GHz, 32-GHz, and optical communication links for its future deep space communication needs [1]. As a result, the Deep Space Network is planning to upgrade its current capabilities by the addition of a 32-GHz downlink [2]. The 500-MHz band, from 31.8 to 32.3 GHz, is allocated for that purpose [3]. An important component in achieving that goal is the development of low-noise front-end amplifiers. One promising approach, drawing upon JPL's experience and success with masers, is to develop a 32-GHz ruby maser.

This article describes the first such device built at 32 GHz by the Cryo-Electronics Group at JPL. The design of this maser draws upon previous experience in building reflected-wave amplifiers at 22 GHz (K-band) [4]. The reflected-wave maser (a negative resistance amplifier utilizing a shorted ruby-filled transmission line) design is considered a more realizable approach for broadband 32-GHz performance than the traveling-wave maser (utilizing a periodic transmission line with distributed isolators) currently in use at 2 and 8 GHz.

This maser will initially be part of a receiver system for the Caltech Owens Valley Radio Observatory (OVRO) located near Big Pine, California. Provision has been made to accommodate either the 5.5-meter Cassegrain antenna with high-gain feedhorns or the 40-meter antenna with focal point feeds. Caltech personnel will operate and maintain the equipment for 32-GHz radio astronomy, specifically for use in measuring the microwave cosmic background (approximately 3 K). The benefit to JPL/NASA will derive from the receipt of weather data for the 32-GHz communications band, from the calibration of known radio stars, and from the acquisition of operational data on the maser's performance and experience in running the maser as part of a system.

II. Input Description

The design (see Figs. 1 and 2) uses two input feedhorns spaced 4.5 inches apart that will be symmetrically placed about the center line of the 5.5-meter antenna and asymmetrically placed about the center line of the 40-meter antenna. The

input lines are identical in structure, and each consists of the following:

- (1) A mica window.
- (2) A vacuum feedthrough waveguide.
- (3) A cryogenically coolable low-noise choked waveguide, nominally ambient to 4.5 K.¹
- (4) A round-to-square waveguide adapter (0.343 round to 0.234 square).
- (5) A polarizer with RCP and LCP outputs.²
- (6) A switching circulator to direct inputs to the maser from either the antenna or a variable-temperature cold load.³
- (7) A 30-dB cross-guide directional coupler for external signal calibration and noise diode inputs.

A third EMS switching circulator switches the maser input between the two antenna input lines at any rate up to 5 kHz. The signal input line is finally connected to the maser circulator block. (Operation of the maser will be fully described in the following sections.)

III. Maser Description

A block diagram of the maser is shown in Fig. 3. It consists of eight stages of ruby-filled waveguide connected in series with 18 junction circulators. This design was chosen to achieve our goal of low noise and wide instantaneous bandwidth. Low noise is possible because of lower dissipative losses than those typically achieved in a traveling-wave maser with slow-wave structures and isolators. The reflected-wave maser uses only the dielectric and waveguide slowing of the ruby-filled guide. Wide instantaneous bandwidth is achieved by using a long length of ruby (approximately 67 cm in total length) and artificially broadening the ruby linewidth with a magnetic field gradient.

The number of stages and the length of each stage are dictated by our goal of low-gain ripple, convenient magnet size, ease of fabrication of the ruby bars, and good noise performance across the band. The absence of a resonant slow-wave structure also allows for relatively large tunability. The tuning range is limited only by the frequency response of the junction circulators. This particular maser will tune from approximately

29 to 34 GHz. However, because of the large pump power required to invert the large volume of ruby, we have chosen to operate with fixed tuned "pump" oscillators, effectively making this a fixed-frequency amplifier at this time.

Pump power requirements are more stringent for the 32-GHz masers than for the 22-GHz masers. Waveguide and structure losses are higher, and the ruby is more difficult to invert due to smaller pump transition probabilities. More importantly, the pump power is divided eight ways rather than four ways as in the 22-GHz maser. The power is supplied by four independent oscillators. It is combined in pairs with two 3-dB hybrid couplers⁴ outside the vacuum vessel. The four waveguide outputs from the hybrids enter the closed-cycle refrigerator (CCR) and are split with four E-plane power splitters into eight channels.⁵ The power is then carried to each ruby separately in WR15 waveguide and impedance-matched to the ruby-filled guide. This ensures an even power distribution among all eight stages.

Almost all of the unfilled signal waveguide is a specially fabricated waveguide.⁶ The waveguide width for 32 GHz was chosen by frequency scaling the WR42 waveguide of the K-band maser from its best operating frequency of 23.3 GHz. The height was chosen to obtain broader-bandwidth circulator operation and to ease the matching into the ruby-filled guide.

The signal enters the circulator block and passes through two junctions. The first junction, configured as an isolator with a termination on the third port, reduces the level of reflected signals exiting the input. This function is performed by all the terminated junctions. After passing through the input isolator, the signal enters the input circulator and then the first ruby bar. The signal travels the full length of the ruby bar, growing exponentially with distance, and is reflected at the opposite end. It travels back up the bar, still being amplified, and is directed into the second stage. The gain of each stage depends on the chosen bandwidth. The signal makes a total of 26 passes through the junctions and 16 passes through the ruby. The forward insertion loss (at 4.5 K) is approximately 12 dB.

The ruby is single crystal and is grown by the Czochralski process by Union Carbide. The chromium-to-aluminum con-

¹Modified version of U.S. Patent #4,215,327.

²Model 2800, Atlantic Microwave Corporation, Boston, Massachusetts.

³Model 408-26, Electromagnetics Sciences, Inc. (EMS), Norcross, Georgia.

⁴Model 15HS12B, Microwave Development Laboratories, Inc., Natick, Massachusetts.

⁵The power splitters were manufactured at JPL. A commercial source is not known to the authors.

⁶Copper (OFHC) waveguide (0.051 inch by 0.306 inch), A. T. Wall Company, Warwick, Rhode Island.

centration ratio is approximately 0.05 percent. The ruby is cut and oriented so that the optic-axis is at an angle of 54.73 degrees to the dc magnetic field. The energy levels of the ground state orbital of the chromium spins are shown in Fig. 4 as a function of magnetic field for this orientation. Population inversion is achieved by push-pull pumping the 1-3 and the 2-4 transitions [5]. The signal transition is the 2-3 transition.

A. Maser Gain and Bandwidth

The natural linewidth of ruby is 60 MHz. In a maser with a uniform magnetic field and 30 dB of electronic gain, the 3-dB bandwidth is 20 MHz, about one-third of the linewidth. Therefore, to achieve masers with wider instantaneous bandwidths, it is common practice to inhomogeneously broaden the line shape by placing different parts of the ruby in different dc magnetic fields [5]. For large bandwidths, an efficient and convenient technique is to employ a linearly staggered magnetic field along the length of the ruby. This is the technique employed here. The frequency dependence of the gain is given by [5]

$$G(\nu) = \frac{g}{2} \frac{w}{d\nu} \left\{ \tan^{-1} \left[\frac{2(\nu - \nu_0)}{w} + \frac{d\nu}{w} \right] - \tan^{-1} \left[\frac{2(\nu - \nu_0)}{w} - \frac{d\nu}{w} \right] \right\} \quad (1)$$

where

$G(\nu)$ = electronic gain in decibels at frequency ν

ν_0 = center frequency

w = natural linewidth of ruby = 60 MHz

$d\nu$ = resonance separation (difference in resonant frequency between opposite ends of the ruby bars assuming a longitudinal magnetic field gradient) = 622 MHz

and g , the unbroadened gain (gain without artificial broadening), is given by

$$g = \frac{27SN}{Q_m}$$

where, in turn,

S = slowing factor = $(c/V_{\text{group}}) = 3.9$

N = length of ruby in free space wavelengths = 143

Q_m = magnetic Q

and

$$\begin{aligned} \frac{1}{Q_m} &= \frac{\gamma^2 h}{\pi \mu_0} \left(\frac{h\nu}{kT} \right) \frac{N}{n} \frac{I \sigma^2 \eta}{w} \\ &= 1.38 (10^{-18}) \frac{h\nu}{kT} \frac{N}{n} \frac{I \sigma^2 \eta}{w} \end{aligned}$$

where

γ = magnetogyric ratio = 2.8 MHz/gauss

h = Planck's constant/ 2π = $1.054\text{E}-27$ erg-sec

I = inversion ratio (gain with pumps on/absorption with pumps off) = 0.7

ν = frequency in hertz

T = absolute temperature = 4.5 K

k = Boltzmann's constant

N = spin density per cubic centimeter = $2.35\text{E}19$

n = number of energy levels = 4

σ^2 = signal transition probability = 0.96

η = filling factor = 0.5

μ_0 = permeability of free space

w = 60 MHz = natural linewidth of ruby in units of megahertz

In this instance, $1/Q_m = 2.29\text{E}-02$. Therefore $g = 232$ dB. Assuming a resonance separation of 622 MHz (corresponding to our tapered shim), the bandpass of the maser looks as shown in Fig. 5. This is the electronic gain (and does not include the maser structure insertion loss) and agrees quite well with what is observed. Obviously, gain can be traded for bandwidth and vice versa.

B. Noise Temperature

The noise temperature of a reflected-wave maser can be thought of as arising from two distinct noise sources. The first source is blackbody radiation from any dissipative losses in the system. The second source is spontaneous emission from the ruby. The theoretical noise temperature of the maser referred to the room temperature flange can be calculated if estimates of the losses and the state of the ruby are known. The results of these estimates are plotted in Fig. 6 and tabulated in Table 1. The noise temperature is a function of gain (and therefore bandwidth) for two reasons. First, as the gain per stage decreases (with increasing bandwidth), the ratio of loss per unit length to gain per unit length increases because the losses

are constant. This causes a large increase in noise temperature, especially for values of this ratio greater than 0.5 [6]. Second, as the gain per stage is decreased, noise from succeeding ruby stages makes a larger contribution to the total noise temperature. (The noise contribution of a given stage, when referred to the input of the complete amplifier, is given by the equivalent noise contribution at the input of the stage divided by the net gain which precedes that stage.)

C. Gain Ripple

There are several sources of gain ripple in this maser. One source is mismatch between the circulator junctions and the rubies. Another source is cross-talk between ruby stages where the structure mates with connecting waveguide flanges. A third is uneven pumping of the rubies due to a frequency-dependent coupling of pump power. Of these different sources of ripple, only the first is amenable to calculation. The VSWR of reduced-height guide to ruby-filled guide has been measured at room temperature and is typically 1.21. This is determined by measuring with a reflectometer the attenuation and ripple with the waveguide shorted at the pump end.⁷ The VSWR into the circulator junctions has also been measured at room temperature and is typically 1.32. This was measured by placing the entire circulator block in a 12-inch Varian electromagnet. Therefore, the equivalent circulator-ruby VSWR is somewhere between 1.09 and 1.60, depending on the frequency and path length between them.

D. Maser Structure

The copper structure that houses the eight ruby bars is constructed in three parts in a sandwich configuration (see Fig. 7). The center section is flat on both sides and forms one broad wall of the waveguide. The remaining three walls are formed by machining waveguide channels in the upper and lower sections of the sandwich. The dimensions of the ruby-filled guide are 0.100 by 0.050 inch. There are four waveguide channels in each section. The center section is made in two parts and contains channels through which liquid helium flows to cool the rubies.⁸ This was an attempt to improve the cooling, since some experiments had indicated that heating of the rubies may occur. This construction effectively eliminated all pressed contact thermal resistance between a number of metal-to-metal joints. These joints were located between the rubies and the main CCR 4.5-K station. Although there was a gain improvement of a few decibels, the improvement was not as large as had been hoped.

⁷T. Otoshi and R. C. Clauss, private communication.

⁸The two parts were vacuum furnace brazed by Scarrott Metallurgical, Los Angeles, California.

The ruby bars are firmly pushed against this center plate by spring-loaded copper pins. The rubies are flush at the pump end of the structure and tapered in the E-plane at the signal end of the structure. The tapered ruby sections are positioned in a six-step quarter-wave transformer to effect the transition from the 0.306-inch-wide waveguide of the circulator block to the 0.100-inch-wide ruby-filled waveguide.

One potentially serious problem with this design is the leakage of signal power from the waveguide in the top section to the waveguide in the bottom section at both ends of the structure. To reduce this problem, we have found it necessary to lap the structure completely assembled (minus the rubies) with a special fixture. After lapping, the structure is never disassembled completely. The rubies are installed without disassembly of the structure.

E. Circulator Block

The circulator block consists of 18 symmetrical Y-junction H-plane circulators. The circulators are also constructed in reduced-height waveguide. Each junction consists of two ferrite disks attached to the upper and lower waveguide broad walls. The disks are approximately 0.107 inch in diameter and 0.020 inch tall. The ferrite material used is type TT 2-111.⁹ Each junction also employs a Teflon matching element, a departure from the small alumina matching elements used in past designs. This circulator design is basically a frequency-scaled version of a design used at 22 GHz [4].

The circulator block construction is similar to that of the maser block (see Fig. 8). It consists of a sandwich of three pieces of copper. Nine of the circulators are between the top two layers and nine are between the bottom two layers. The top and bottom rows are connected in series by an external length of waveguide.

F. Superconducting Magnets

There are two separate superconducting magnets in this maser. Both magnets operate in a vacuum and are cooled by thermal conduction to the liquid helium within the 4.5-K CCR station. The main magnet is used to provide the 11.8-kilogauss dc field required to Zeeman-split the energy levels to achieve a 32-GHz frequency separation between levels 2 and 3. The second magnet is used to bias the ferrite disks in the circulator block assembly. First, the magnet for biasing the rubies will be discussed (see Fig. 9).

The magnetic field is created by a single superconducting racetrack coil completely enclosed in a Hipercob (iron/cobalt

⁹Trans-Tech, Inc., Gaithersburg, Maryland. The material is a nickel-zinc ferrite with a saturation magnetization near 7000 gauss at 4.2 K.

alloy) box.¹⁰ This magnet does not suffer from the fringing field problems associated with the open Cioffi-style magnets [7] used at K-band. The magnet coil form is rectangular with rounded corners and has inside dimensions of 7.2 by 3.57 inches. The coil form is 1.1 inches high and is constructed of copper. The magnet windings consist of 3056 turns of niobium-titanium (T48B) wire.¹¹ There are approximately 40 layers with 76 turns per layer. The coil windings are prevented from moving by vacuum-impregnating the coil with Scotchcast 235 epoxy.¹²

The top and bottom plates of the Hiperco are 9.6 inches by 6.0 inches by 0.64 inch. They are each constructed of two separate pieces. Slots are machined for the signal waveguides in one plate and for the pump waveguides in the opposite plate. The side pieces are 1.030 inch high, approximately 0.640 inch wide, and of varying lengths.

The magnetic field taper is achieved by inserting a tapered Hiperco shim between the top plate and the sides. There is also 0.016 inch of steel shim over the four center rubies. This corrects a lateral field gradient of approximately 180 gauss between the inner and outer rubies. The current required to achieve 11.8 kilogauss is 12.6 amps. The magnet is operated in the persistent mode. The thermal switch consists of a small light bulb illuminating a small section of the superconducting wire where the copper cladding has been etched away.

The magnet for the circulators is much smaller and constructed differently (see Fig. 10). The Hiperco yoke is C-shaped with additional cylindrical pole pieces to get as much Hiperco as close to the ferrite as possible. The superconducting wire¹³ is wound on the Kapton-covered Hiperco. To bias the ferrites correctly at 4.5 K, the coil is usually charged with 7.5 amps. It is also operated in the persistent mode.

G. Microwave Pumps

The question of how best to pump the ruby to obtain population inversion depends to some extent on the pump power required. It is well known that ruby becomes more difficult to pump at frequencies that are large compared to the zero field

splitting of 11.5 GHz. It is rather easy to estimate the minimum pump power required if one assumes a relaxation rate for the pump transitions. One can use

$$P_{\min} = hf_p w \alpha \left(\frac{N}{V} \right) V \quad (2)$$

where

f_p = pump frequency

w = relaxation rate

N/V = spin density

V = volume of maser material

α = a constant which depends on the population distribution because all the spins need not be flipped

In this case, $f_p = 66.4$ GHz, $\alpha = 0.17$, $V = 2.29$ cc, w is estimated to be 100/sec, and $N/V = 2.35 \times 10^{19}/\text{cc}$. Therefore, $P_{\min} = 40$ mW. However, Eq. (2) is a lower bound on the required pump power. It underestimates the required power for several reasons. First, it assumes that every photon incident on the ruby is successful in flipping a spin. This is a very poor assumption for the specific ruby operating point we are using because of the low transition probabilities. It is a particularly bad assumption when the ruby is located in an electromagnetic structure with a low Q at the pump frequency. Second, it ignores the fact that the pump RF magnetic field may not be properly polarized. Third, it ignores waveguide losses and losses due to mismatch.

In addition to the estimates, some experiments were carried out using a single bar of the ruby in liquid helium. We found that approximately 40 mW was required at the input to a single ruby channel in order to achieve an inversion ratio near 1.0 over 500 MHz of signal bandwidth. After adding in the waveguide losses and window losses, we estimated that each stage would require at least 50 mW. Therefore, the total requirement for eight stages would be 400 mW. Because of the large power requirement, it was decided to use four IMPATT diode oscillators, each supplying approximately 100 mW of CW power.¹⁴ The method by which the pump power was directed to the rubies was described earlier.

The coupling of the pump power to the rubies was accomplished with quarter-wave matching elements. The design of the quarter-wave matching elements was based on a wave impedance given by

¹⁰ Carpenter Steel, Reading, Pennsylvania.

¹¹ Supercon, Inc., Shrewsbury, Massachusetts. The wire is a single filament with a superconducting core diameter of 0.005 inch with 0.0025-inch copper cladding. The superconducting transition temperature is 9.5 K.

¹² 3M Corporation, Saint Paul, Minnesota. The epoxy was added by the JPL cable shop.

¹³ Also T48B wire with a 0.005-inch niobium-titanium core and 0.0015-inch copper cladding.

¹⁴ Model 47134H, Hughes Aircraft Company, Electron Dynamics Division, Torrance, California.

$$Z = \frac{b}{a} \frac{\lambda_g}{\lambda_0}$$

Then we use the matching condition for quarter-wave transformers, i.e., Z (matching section) = Z_m :

$$Z_m = \sqrt{Z_1 \cdot Z_2}$$

where Z_1 = impedance of the dielectric-filled guide and Z_2 = impedance of WR15. Then Z_m is given by

$$Z_m = \sqrt{\frac{b_1 b_2}{a_1 a_2} \frac{\sqrt{\lambda_g^{(1)} \lambda_g^{(2)}}}{\lambda_0}} = \sqrt{\frac{b_m}{a_m} \frac{\lambda_g^{(m)}}{\lambda_0}}$$

Thus $b_m = 0.061$ and $a_m = 0.122$. ϵ_m is chosen so that

$$\lambda_g^{(m)} = \sqrt{\lambda_g^{(1)} \lambda_g^{(2)}} = \frac{\lambda_0}{\sqrt{\epsilon_m - (\lambda/\lambda_{c0})^2}}$$

This requires that $\epsilon_m = 2.9$. We used a dielectric (polystyrene doped with titanium) with a dielectric constant ϵ of 3. The thickness of the elements is 0.029 inch. Experimentally, the return loss looking into the pump end of a ruby bar with the matching element was approximately 15 dB. The IMPATT diode oscillators are frequency modulated with a bandwidth roughly twice the desired signal bandwidth. The modulation rate is typically 4 kHz.

H. Cryogenics

A closed-cycle helium refrigerator (CCR) is used to provide a 4.5-K environment for the maser (see Fig. 11). This refrigerator uses a CTI-Cryogenics Model 1020 drive unit to precool a Joule-Thomson expansion circuit. The refrigerator is capable of more than 3 watts of cooling at 4.5 K. In addition, we have installed an externally adjustable Joule-Thomson valve that permits operating at reduced helium flow rates. This allows us to achieve temperatures lower than 4.5 K by reducing the vapor pressure over the helium with a vacuum pump.¹⁵ This produces an additional 10 dB of gain at the expense of reserve cooling capacity.

¹⁵A Leybold-Heraeus model S65B vacuum pump that has a pumping speed of 45.9 cubic feet per minute was used. This model pump will be used at OVRO.

I. Maser Package

The entire maser and refrigerator assembly (minus the drive unit) is contained within a cylindrical stainless steel vacuum housing approximately 12.5 inches in its inside diameter (see Fig. 12). The four IMPATT diode oscillators are mounted on a temperature-controlled copper plate attached to the vacuum housing.

The refrigerator and maser are both surrounded by a cylindrical copper radiation shield near 70 K, and the maser is further shielded by a cylindrical copper radiation shield near 5 K. The maser package weighs approximately 200 pounds.

For mounting in the antenna, the maser, the receiver, and various pieces of support equipment are housed in a circular container 18.75 inches in diameter and 84 inches long. The estimated total weight is approximately 400 pounds.

IV. Performance

The performance of the maser with the field tapered as described in Section IIA yielded a net gain of 21 dB and an instantaneous bandwidth of 400 MHz at 4.5 K (see Fig. 5). This is the maximum gain-bandwidth product we have achieved. Higher gains with reduced bandwidth can be obtained by decreasing the frequency modulation bandwidth of the "pump" oscillators and reducing the magnetic field gradient. The amount of ruby absorption is difficult to measure with our measurement setup (all components at 32 GHz). The power absorption is so large (because of the large volume of ruby) that the power level at the output of the maser is below the sensitivity of our diode detectors. However, the absorption certainly exceeds 36 dB. Therefore, the inversion ratio (gain with pumps on/absorption with pumps off) is less than 1.0 and probably closer to 0.7. The ruby is considered to be saturated when an inversion ratio of approximately 1.1 is achieved [8].

The noise temperature was measured as a function of gain by measuring output power with reference terminations at different temperatures connected to the maser. In this case, the "hot load" was an ambient target placed over the horn and the "cold load" was the sky. The maser output was downconverted to S-band with a superheterodyne mixer. Commercial S-band amplifiers, a tunable filter, and a power meter enabled us to compare the maser output power when the input was looking at the hot and cold loads. The measured maser noise temperatures (including the feedhorn) were as follows (see Fig. 6):

Net gain of 40 dB: $T = 17 \pm 2.5$ K

Net gain of 30 dB: $T = 20 \pm 2.5$ K

Net gain of 20 dB: $T = 25 \pm 2.5$ K

The error was estimated from the scatter in the measured values. Since these results were all obtained with a fixed magnetic field taper, in principle the noise temperature obtained with 40 dB of gain could be realized across a full 400-MHz bandwidth if there were sufficient pump power and if ruby heating was not a problem. The gain ripple can be as low as ± 1 dB with adjustment of the magnetic field, IMPATT center frequencies, and sweep bandwidths.

V. Applications

This wide-bandwidth maser will find immediate application in continuum radio astronomy as well as in deep space spacecraft communications. In addition to using this maser to search for spatial anisotropies in the cosmic background, other potential experiments include tracking the fourth harmonic of the Voyager or Galileo spacecraft X-band signal to study weather effects on a deep space communications link at 33.6 GHz.

Acknowledgments

The authors would like to thank several of their colleagues at JPL for very useful discussions. In particular, thanks are due to Bob Clauss, Sam Petty, and Javier Bautista for matters relating to the maser construction and testing. Also, thanks to Dan Hoppe for his assistance and for the use of his computer program for analyzing the pump matching as well as to Tom Otoshi for discussions relating to the VSWR measurements. Thanks also to Dr. Charles Lawrence of the Caltech Astronomy Department for his assistance in the final finishing stages of the maser and its installation at the Owens Valley Radio Observatory.

References

- [1] R. M. Dickinson, *A Comparison of 8.415, 32.0, and 565646-GHz Deep Space Telemetry Links*, JPL Publication 85-71, Jet Propulsion Laboratory, Pasadena, California, 1985.
- [2] J. G. Smith, "Ka-Band (32-GHz) Downlink Capability for Deep Space Communications," *TDA Progress Report 42-88*, vol. October–December 1986, Jet Propulsion Laboratory, Pasadena, California, pp. 96–103, February 15, 1987.
- [3] N. F. deGroot, "Ka-Band (32-GHz) Allocations for Deep Space," *TDA Progress Report 42-88*, vol. October–December 1986, Jet Propulsion Laboratory, Pasadena, California, pp. 104–109, February 15, 1987.
- [4] C. R. Moore and R. C. Clauss, "A Reflected-Wave Ruby Maser with K-Band Tuning Range and Large Instantaneous Bandwidth," *IEEE Trans.*, vol. MTT-27, pp. 249–256, 1979.
- [5] A. E. Siegman, *Microwave Solid State Masers*, New York: McGraw-Hill, 1964.
- [6] W. H. Higa, "Noise Performance of Traveling-Wave Masers," *IEEE Trans.*, vol. MTT-12, p. 139, 1964.
- [7] P. P. Cioffi, "Approach to the Ideal Magnetic Circuit Concept Through Superconductivity," *J. Appl. Phys.*, vol. 33, pp. 875–879, March 1962.
- [8] C. R. Moore and D. Neff, "Experimental Evaluation of a Ruby Maser at 43 GHz," *IEEE Trans.*, MTT-30, pp. 2013–2015, 1982.

Table 1. Estimated noise temperature contributions (in kelvins) from various components of the maser/receiver for three different values of net gain

Variable/stage	Net gain, dB		
	40	30	20
1	5.4	6.0	6.8
2	1.8	2.6	3.9
3	0.5	1.1	2.2
4	0.2	0.5	1.2
5	0.1	0.2	0.7
6	—	0.1	0.4
7	—	—	0.2
8	—	—	0.1
Zero point energy	0.8	0.8	0.8
T_{equiv} (4.5-K input)	8.8	11.3	16.3
Waveguide inputs (estimated)	3	3	3
T_{equiv} (ambient)	11.8	14.3	19.3
Feedhorn (estimated)	4	4	4
T_{total}	15.8	18.3	23.3

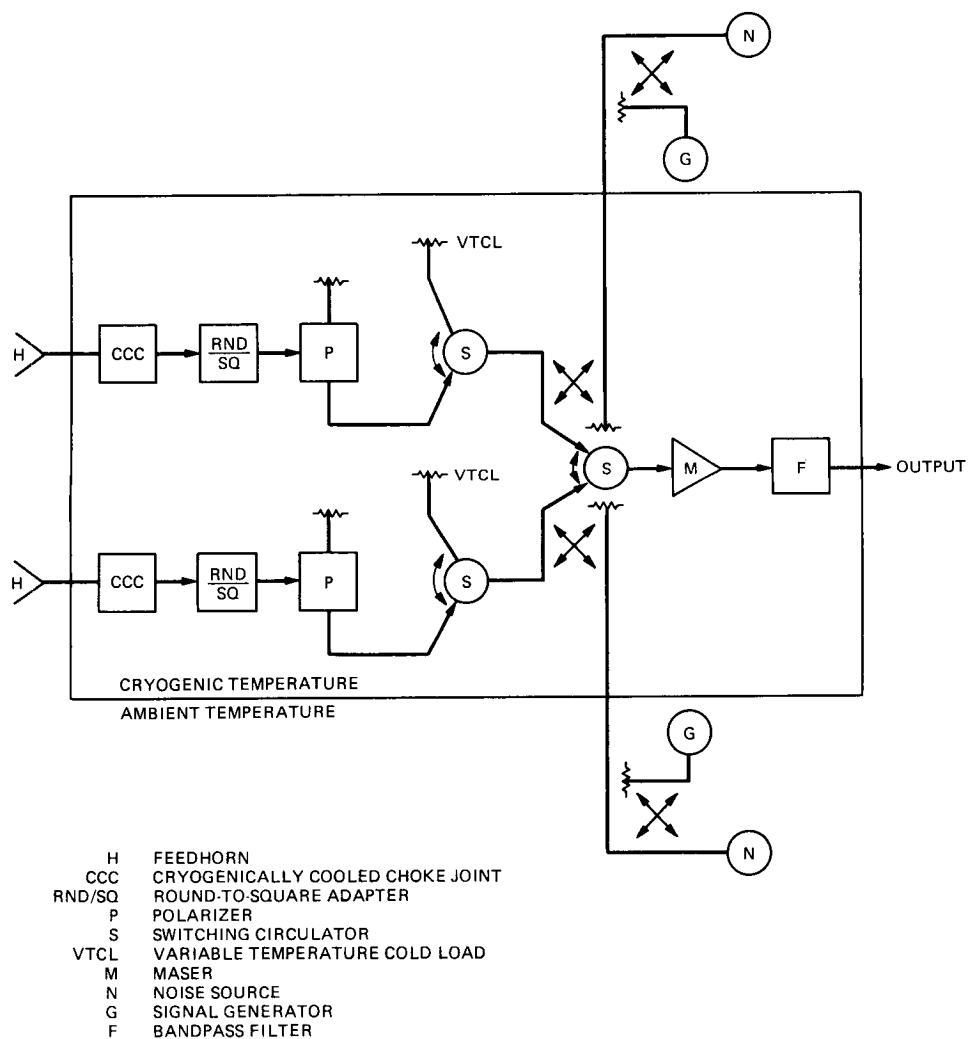


Fig. 1. Receiver input block diagram

ORIGINAL PAGE IS
OF POOR QUALITY

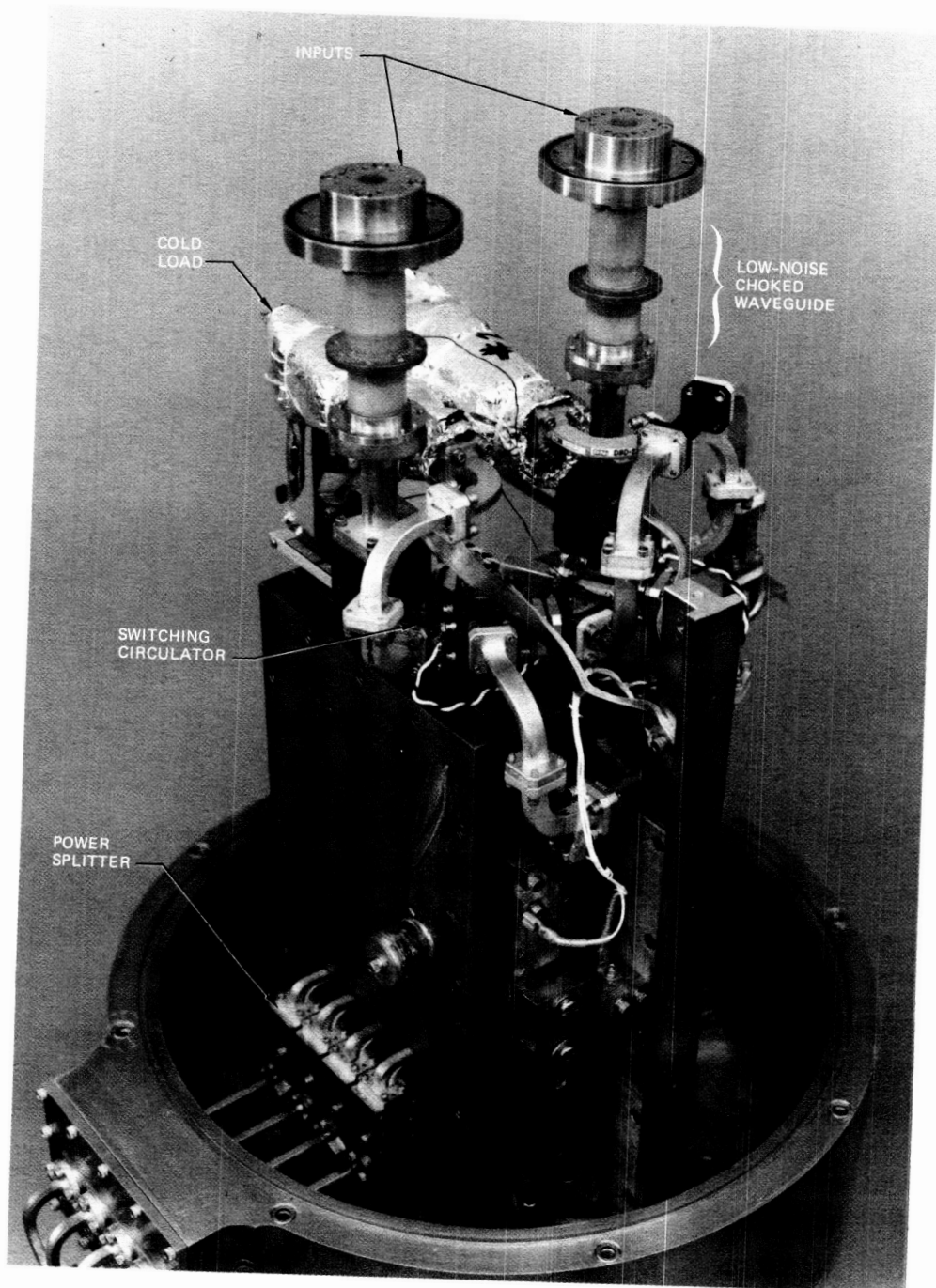


Fig. 2. View of the input feeds, waveguide plumbing, and maser

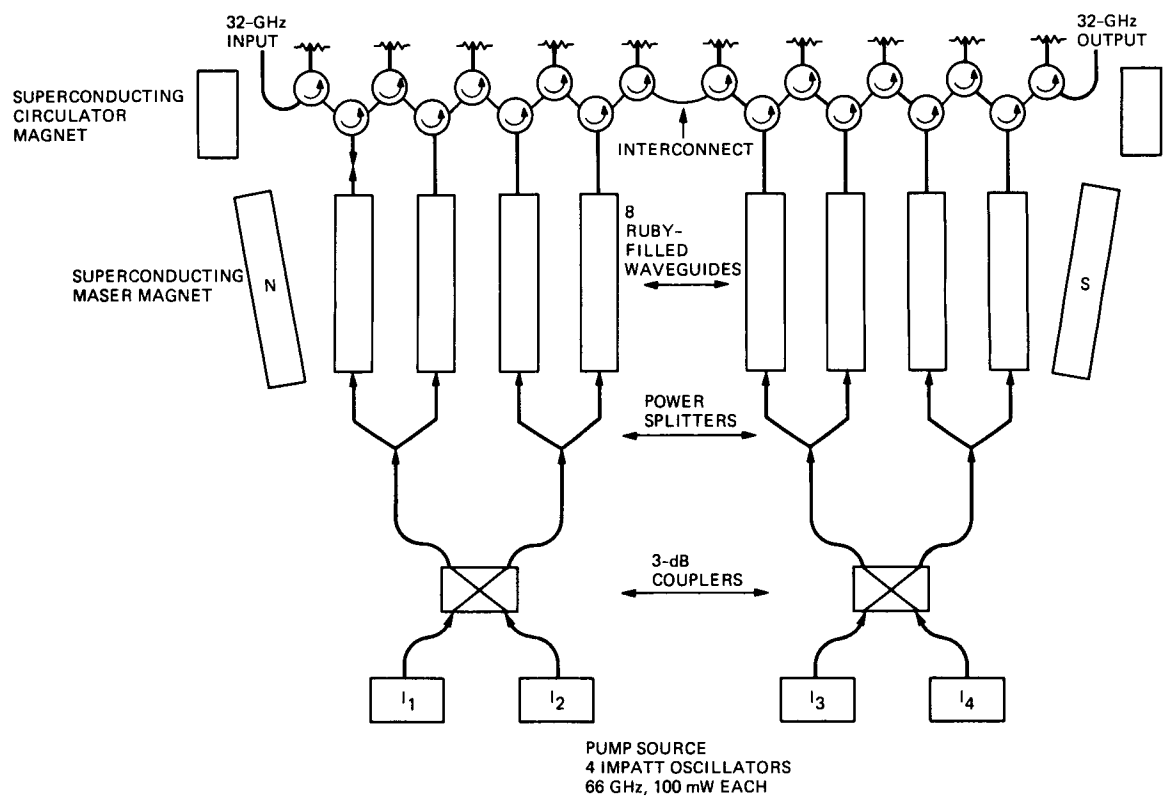


Fig. 3. 32-GHz reflected-wave maser block diagram

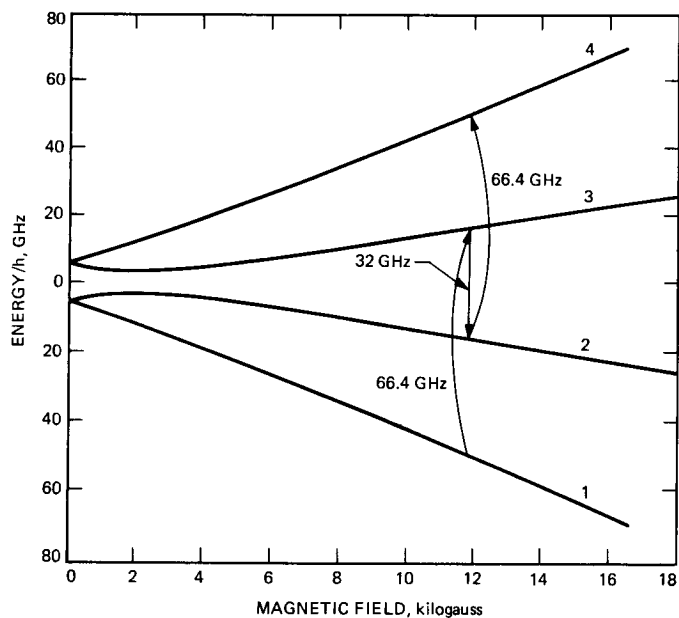


Fig. 4. Ground state energy levels of ruby with optic-axis orientation of 54 degrees 44 minutes. Upward pointing arrows denote pump transitions and downward pointing arrow denotes signal transition.

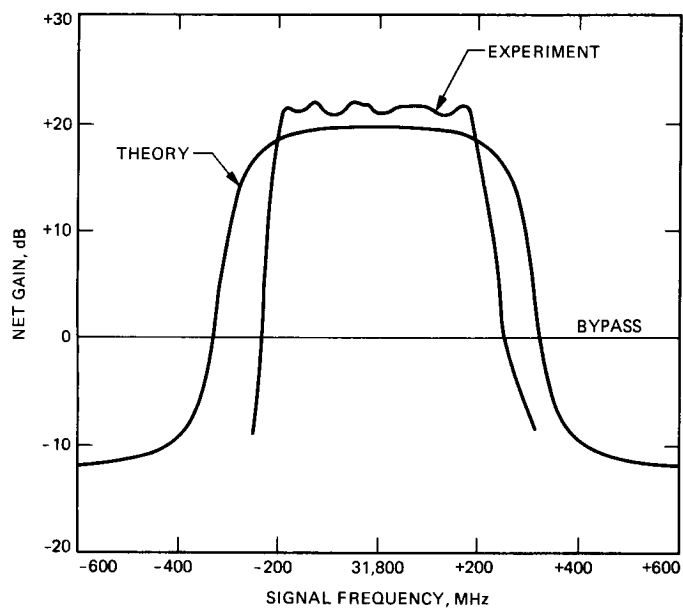


Fig. 5. Theoretical bandpass using equation (1) with $g = 232$ dB and $\Delta\nu = 622$ MHz superimposed on the experimentally measured bandpass

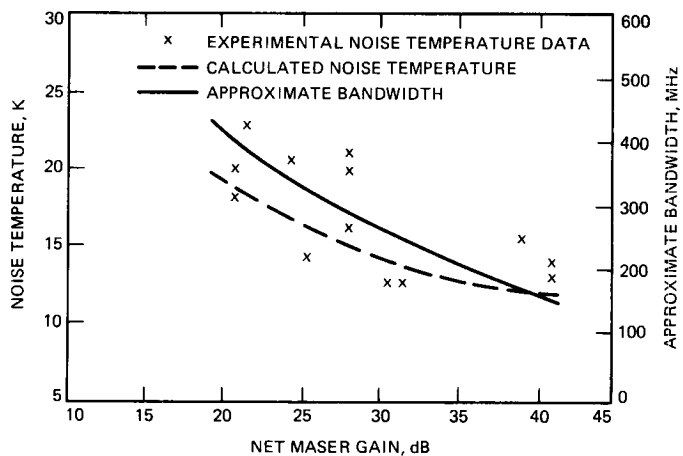


Fig. 6. Experimental and estimated noise temperatures as a function of maser gain and bandwidth

ORIGINAL PAGE IS
OF POOR QUALITY

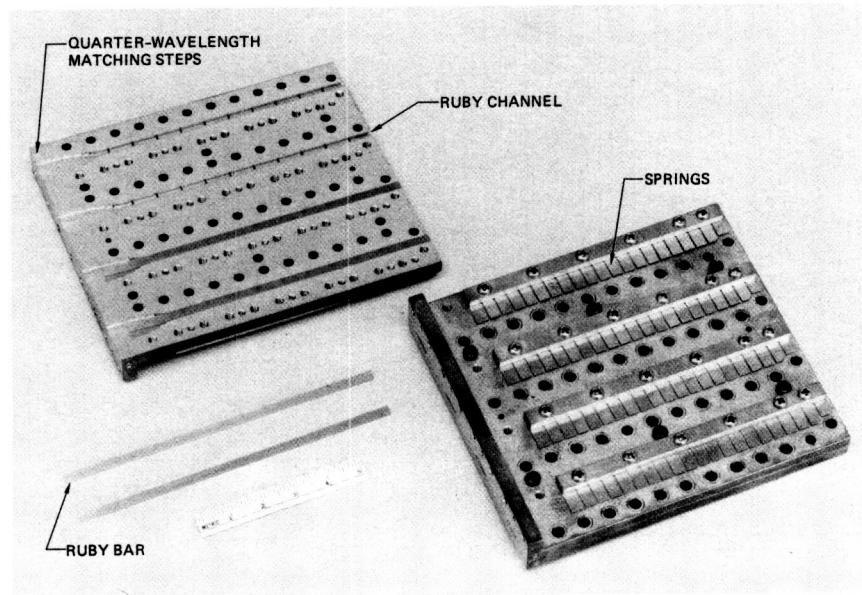


Fig. 7. Partially disassembled view of the ruby-filled structure. The helium lines to the center dividing plate are not shown.

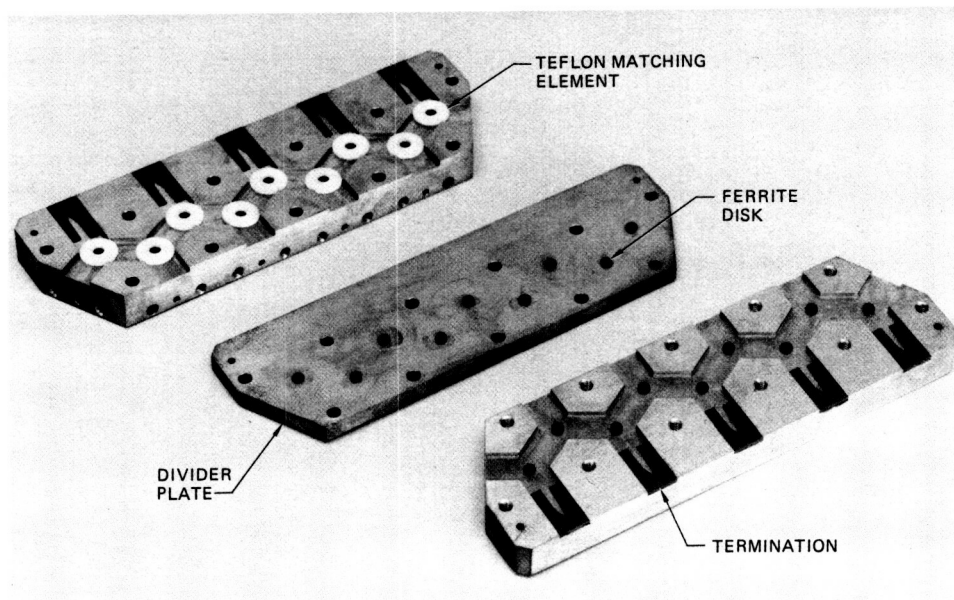


Fig. 8. Disassembled view of the circulator block

ORIGINAL PAGE IS
OF POOR QUALITY

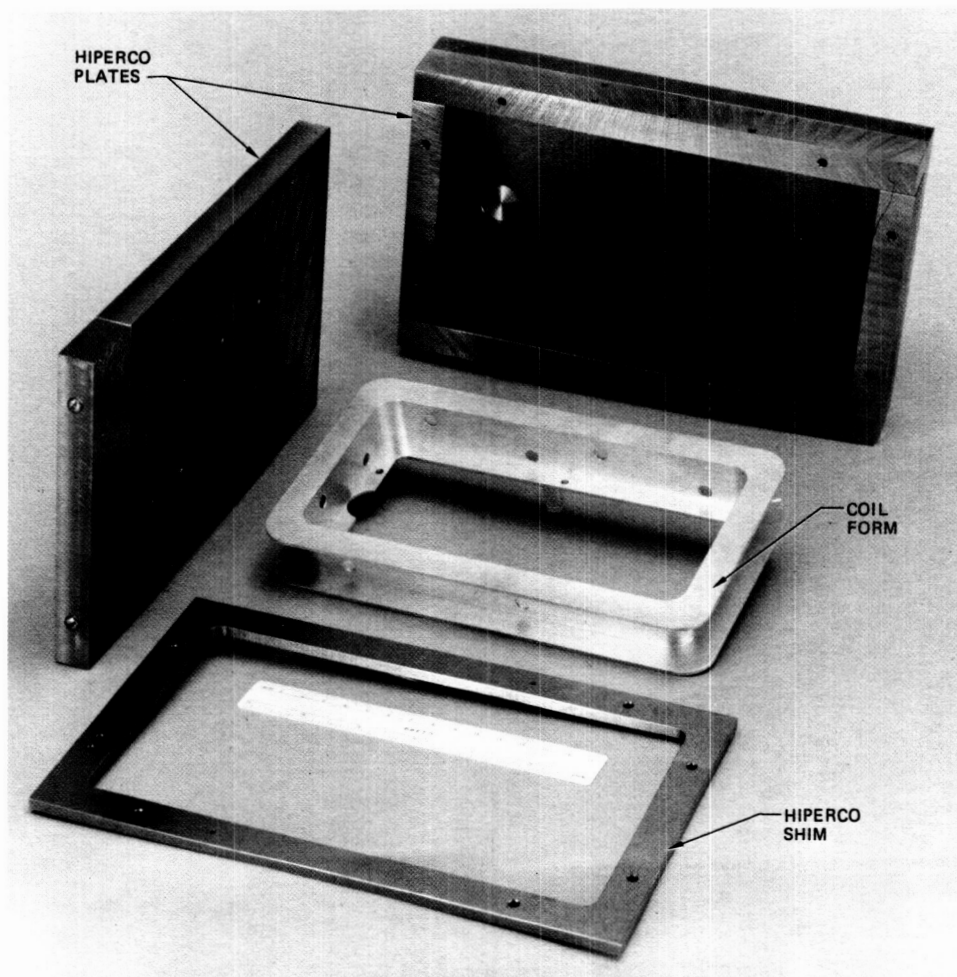


Fig. 9. View of the Hiperco box and copper coil form before the winding of the superconducting coil. The holes for the pump waveguide are not shown.

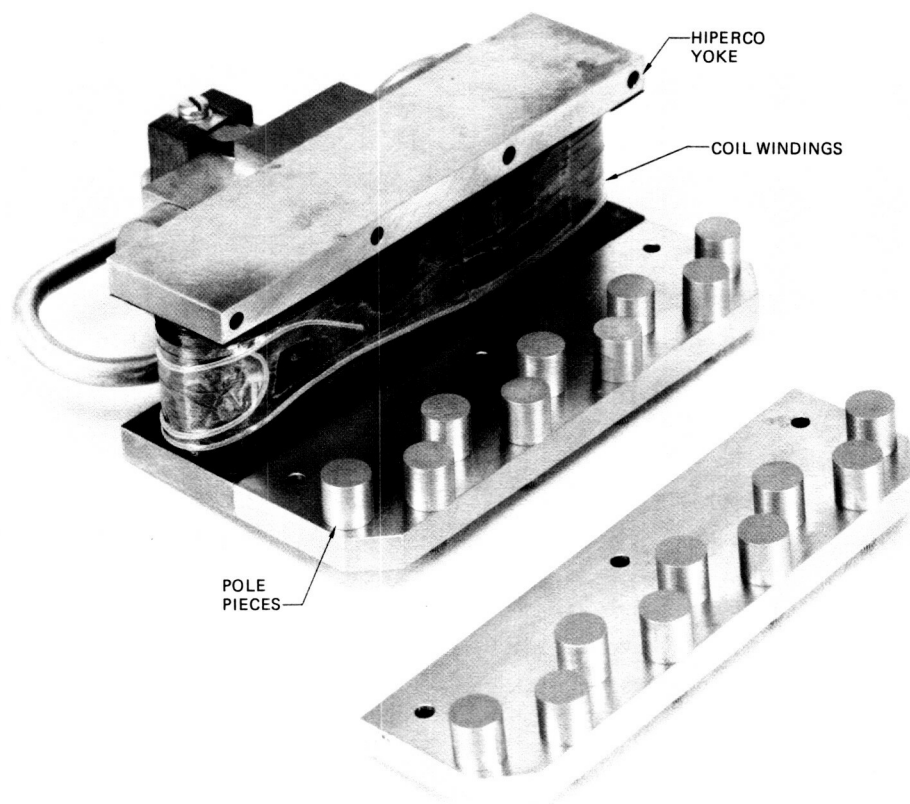


Fig. 10. Partially disassembled view of the circulator electromagnet

ORIGINAL PAGE IS
OF POOR QUALITY

ORIGINAL PAGE IS
OF POOR QUALITY

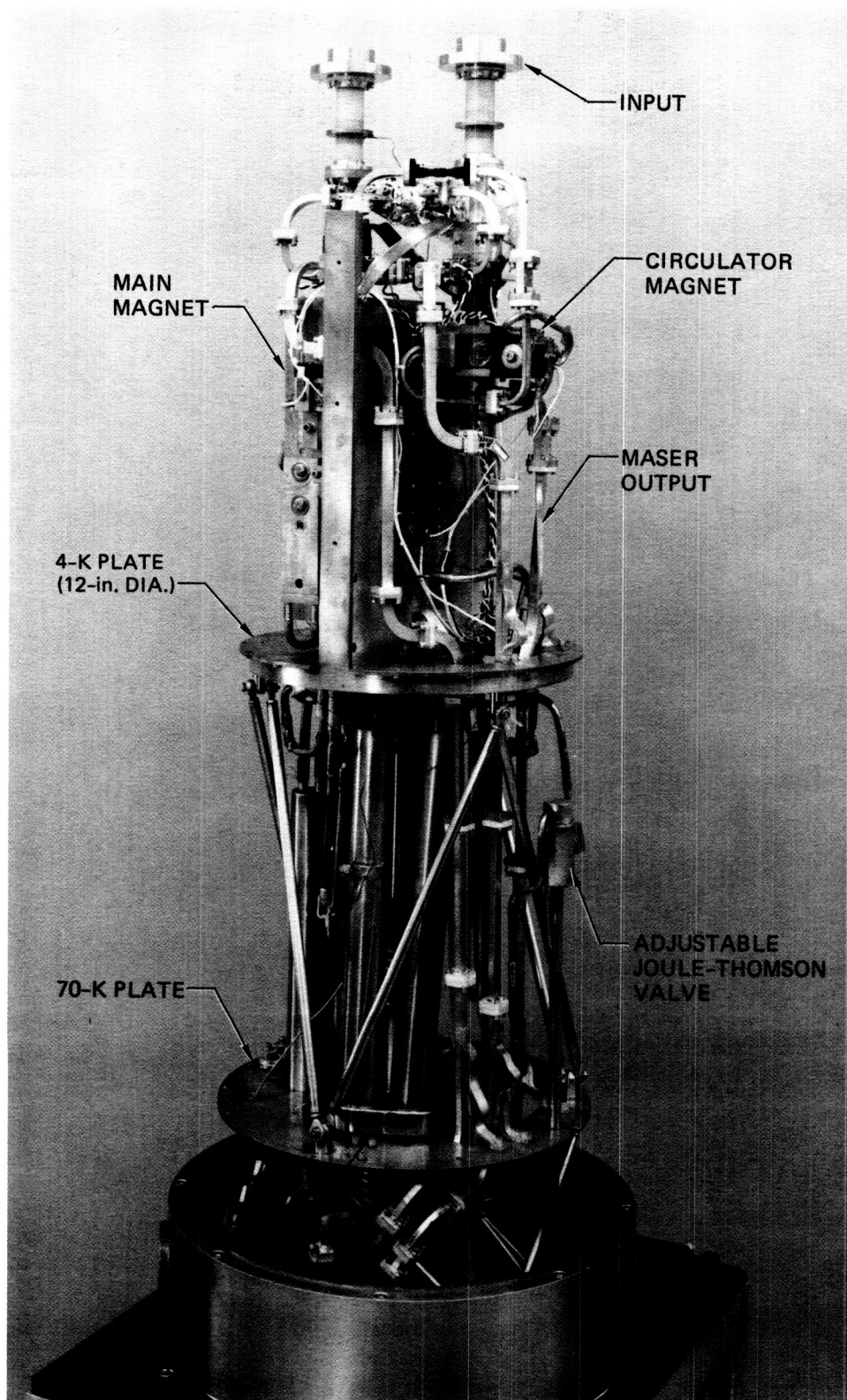


Fig. 11. Rear overview of the complete maser. The radiation shields and vacuum housings have been removed.

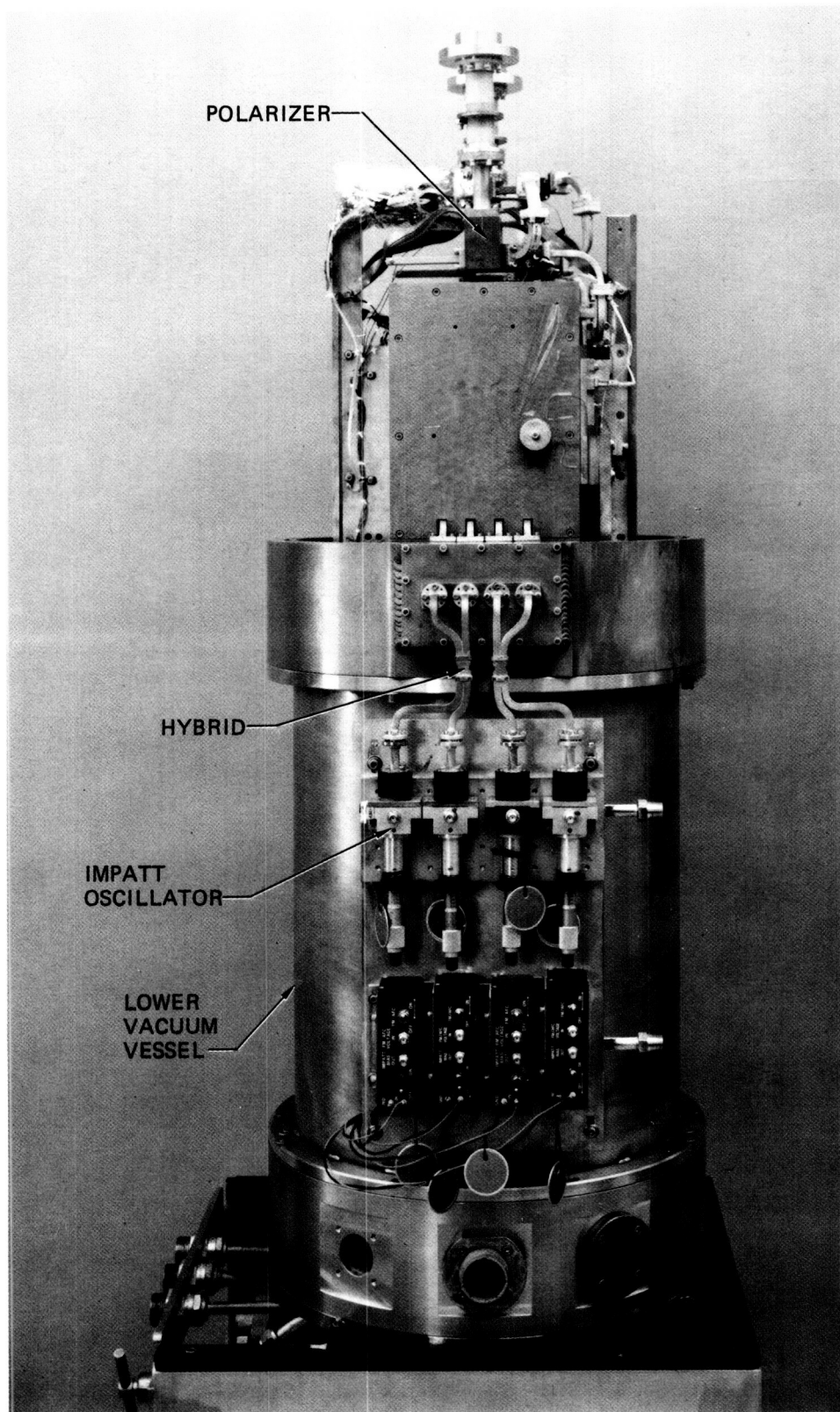


Fig. 12. Front overview of the complete maser. The upper vacuum housing and radiation shields have been removed.

N89 - 10199

S/3 -33
~~16824~~
April-June 1988
78
JPL1989000828
618390
7p.

An 8.4-GHz Cryogenically Cooled HEMT Amplifier for DSS 13

L. Tanida

Radio Frequency and Microwave Subsystems Section

A prototype 8.4-GHz (X-band) high electron mobility transistor (HEMT) amplifier/closed-cycle refrigerator (CCR) system was installed in the DSS 13 feedcone in August 1987. The amplifier is cryogenically cooled to a physical temperature of 12 K and provides 31-K antenna system noise temperature (zenith) and 35 dB of gain at a frequency of 8.2 to 8.6 GHz. Antenna system noise temperature is less than 50 K from 7.2 to 9.4 MHz. The low-noise HEMT amplifier system is intended for use as a radio astronomy or space communications receiver front end.

I. Introduction

This article describes the design, assembly, and performance of the first 8.4-GHz high electron mobility transistor (HEMT) amplifier system designed for the DSN. This system was installed on the DSS 13 antenna at Goldstone, California, in August 1987, where it is mounted beside the 2.3-GHz HEMT/CCR amplifier system [1]. A performance goal for this system was to provide the widest possible operating bandwidth for radio science applications while maintaining state-of-the-art noise performance in the DSN 8.2- to 8.7-GHz band. The antenna system noise temperature (looking at zenith, clear weather) measures less than 50 K across an instantaneous bandwidth of 7.2 to 9.4 GHz with the use of this HEMT/CCR system. At midband, antenna system noise is as low as 31 K. At the room temperature waveguide input flange at a frequency of 8.42 GHz, the amplifier input noise temperature measures 13 K. The operating frequency range of the HEMT/CCR system includes frequencies that operate very close to the DSS 13 transmitter frequency, and since no cryogenic transmitter reject filter has been provided, system operation is limited to

a listen-only mode.¹ The narrower-bandwidth 8.4-GHz traveling wave maser system remains available for use when diplexed operation is desired (see Fig. 1).

II. General Description

Figure 1 is a block diagram of the signal distribution for the DSS 13 8.4-GHz microwave front end, including the 8.4-GHz HEMT/CCR system. Figure 2 is a photograph of the HEMT/CCR package with vacuum housing and radiation shield removed to show the components mounted on the 12-K refrigerator station. The 12-K closed-cycle refrigerator contains a cryogenically cooled waveguide input transmission line,

¹This amplifier design typically experiences 1 dB of gain compression at -25 dBm input CW power. If diplexed operation is desired, a cryogenic filter (or, alternatively, an external waveguide filter) must be provided to attenuate transmitter leakage from the diplexer receiver port (up to +20 dBm in 1985) to a level that is at least 10 dB below that which causes 1-dB saturation (approximately -35 dBm).

a cryogenic input isolator, a three-stage amplifier that contains a HEMT in the first stage and GaAs FET devices in the second and third stages, and a cryogenic output isolator. Figure 3 shows the completed HEMT/CCR package.

III. Detailed Description

A. Cryogenic Input Transmission Line

The function of the input transmission line is to transmit the input RF signal from the room temperature WR 125 waveguide flange to the cryogenic low-noise amplifier while adding the least possible noise and an acceptably low heat load to the CCR system. A mica waveguide window on the top supports a vacuum in the CCR dewar. A thermal barrier separating the 300-K and 12-K portions of the input waveguide is achieved by a 0.005-inch gap in the waveguide wall. A radial choke prevents RF leakage from the gap. A waveguide-to-coaxial transition couples the input signal to a semi-rigid coaxial line at a 12-K physical temperature. The noise contribution of this assembly is estimated to be less than 0.5 K based on room temperature insertion loss measurements.

B. Cryogenic Isolator

A cryogenically cooled isolator is included at the input of the HEMT amplifier to provide good input match over the entire bandwidth. The cryogenic isolator, P&H Laboratories Model C1-X-13366, has a return loss exceeding 18.5 dB from 8.2 to 8.7 GHz and an insertion loss of 0.13 dB. The estimated noise contribution of this assembly is approximately 1 K or less.

C. HEMT Amplifier Module

The HEMT/FET/FET amplifier module is shown in Fig. 4. The amplifier is a three-stage microstrip circuit design now commercially available from Berkshire Technologies, Inc. The first-stage HEMT is a commercially available 0.5- μ m device, the Fujitsu FHR01FH. The second-stage and third-stage FETs are the Mitsubishi 1412 and 1402, respectively. Each stage is individually biased for optimum low-noise performance. Low-temperature HEMT performance is often complicated by a sensitivity to light, which is thought to be associated with the existence of traps in the AlGaAs layers [2], [3]. Minimum noise temperature is achieved with the Fujitsu HEMT device by application of a low level of light. An LED mounted in the cover of the amplifier provides this function. An output isolator is included to ensure >15 dB of output return loss. The minimum input noise temperature (at 8450 MHz) of the completed three-stage amplifier is 13 K with 35 dB of gain (includes the input and output isolators). It should be noted that the Fujitsu FHR01FH HEMT used in this amplifier is a very early device; the more recent shipments of this HEMT have not provided state-of-the-art noise performance.

D. Closed-Cycle Refrigerator

The HEMT/CCR package uses a Cryogenics Technology, Inc., Model 350 Cryodyne refrigerator and compressor. This is the same CCR used for all cryogenic FET and HEMT amplifiers in the DSN. The refrigerator is enclosed in a vacuum housing and radiation shield designed specifically for DSN applications. The first stage of the refrigerator operates at approximately 50 K and the second stage at 12 K. The refrigerator is designed to provide an unloaded cooling capacity of 15 watts at 70 K and 3 watts at 15 K. Cool-down time of the HEMT/CCR system is approximately 5 hours. The fully loaded reserve heat capacity was measured to be 2.7 watts at the 15-K second-stage station. The completed HEMT/CCR package is shown in Fig. 3.

IV. DSS 13 Installation and Performance Tests

Prior to installation, the input noise temperature of the assembled HEMT/CCR package was measured at JPL using the Y-factor method. A horn was attached to the amplifier input and Y-factors were obtained by switching between "cold sky" and a room temperature aperture load. The input noise temperature at the room temperature waveguide flange measured 13.0 to 13.2 K across the 8400- to 8500-MHz range. On July 10, 1987, the HEMT/CCR was installed at DSS 13, and antenna system noise temperature was subsequently measured at zenith during clear weather. These results are shown in Fig. 5.

Since the HEMT/CCR system replaced a maser/CCR system, it is appropriate to compare the characteristics of these two types of low-noise amplifier systems:

- (1) The HEMT amplifier provides excellent noise performance in a 12-K cryogenic refrigerator, whereas the maser requires a more complex and less reliable 4.6-K refrigerator.
- (2) In the event of a refrigerator and/or compressor failure, the noise temperature of a HEMT/CCR system degrades in a gradual "fail-soft" manner. Figure 6 shows the input noise temperature of the DSS 13 8.4-GHz HEMT/CCR as a function of time after the refrigerator and the compressor were turned off. Noise temperature, normally 13 K, degraded to 20 K after 45 minutes and reached approximately 25 K in 2 hours.
- (3) HEMT systems provide greater bandwidths than do maser systems. Masers, on the other hand, provide the lowest possible noise temperatures when ultimate noise performance is required, such as for Voyager and other spacecraft downlink applications.

- (4) HEMT amplifiers have no tuning controls that need to be adjusted in the field. Masers require tuning of the magnet and pump source (oscillators).
- (5) HEMT/CCR systems are smaller, less complex, and lighter in weight than maser systems.
- (6) HEMT amplifiers experience gain compression at approximately -25 dBm compared to -85 dBm for masers.
- (7) The HEMT amplifier will be permanently damaged when input power exceeds 100 to 500 mW, while a maser can survive many watts of input power (a distinct advantage during those few occasions when transmitter power has been accidentally applied to a maser-receiver system configured for listen-only).
- (8) The reliability of the DSS 13 X-band HEMT/CCR has not been as good as that of other FET/CCR systems because of a small vacuum leak that has not been considered serious enough to repair. This vacuum leak makes it necessary that the system be warmed and pumped down with a vacuum pump during a convenient time period approximately every 3 months. Otherwise, during the 14 months since installation, only one

failure occurred: a chip capacitor delaminated in the HEMT module. Berkshire Industries identified this capacitor as one that has failed in many of their commercial cryogenic amplifiers. A different capacitor was installed that has proven to be reliable in cryogenic use.

V. Conclusions

Since the time of installation, there have been numerous VLBI and other radio astronomy experiments utilizing the HEMT amplifier. The amplifier has proven to require less operator attention than have previous maser systems, and repairs are typically less time-consuming. The gain stability of this system, as well as that of the complete DSS 13 receiver subsystem, is very good. A 50-mK radio spectrum has been resolved with a noise level of 10 mK for a 45-minute integration time (T. Kuiper, private communication).

Since installation of this system, lower-noise HEMT devices have been developed by the General Electric Company. It is expected that the amplifier noise temperature of this system (at the room temperature waveguide input flange) could be reduced to less than 10 K at 8.2 to 8.7 GHz with the use of one of these GE devices.

Acknowledgments

The author wishes to acknowledge Jan Loreman, who supplied the refrigerator and input transmission line, and Juan Garnica and George Farner for an expeditious installation on the DSS-13 antenna.

References

- [1] L. Tanida, "A 2.3-GHz Cryogenically Cooled HEMT Amplifier for DSS-13," *TDA Progress Report 42-91*, vol. July-September 1987, Jet Propulsion Laboratory, Pasadena, California, pp. 94-101, November 15, 1987.
- [2] A. W. Swanson, "Pseudomorphic HEMT," *Microwaves & RF*, vol. 26, no. 3, pp. 139-150, March 1987.
- [3] J. Y. Chi, R. P. Holstrom, and J. P. Salerno, "Effect of Traps on Low Temperature, High Electron Mobility Characteristics," *IEEE Electron Device Letters*, vol. EDL-5, pp. 381-384, September 1984.

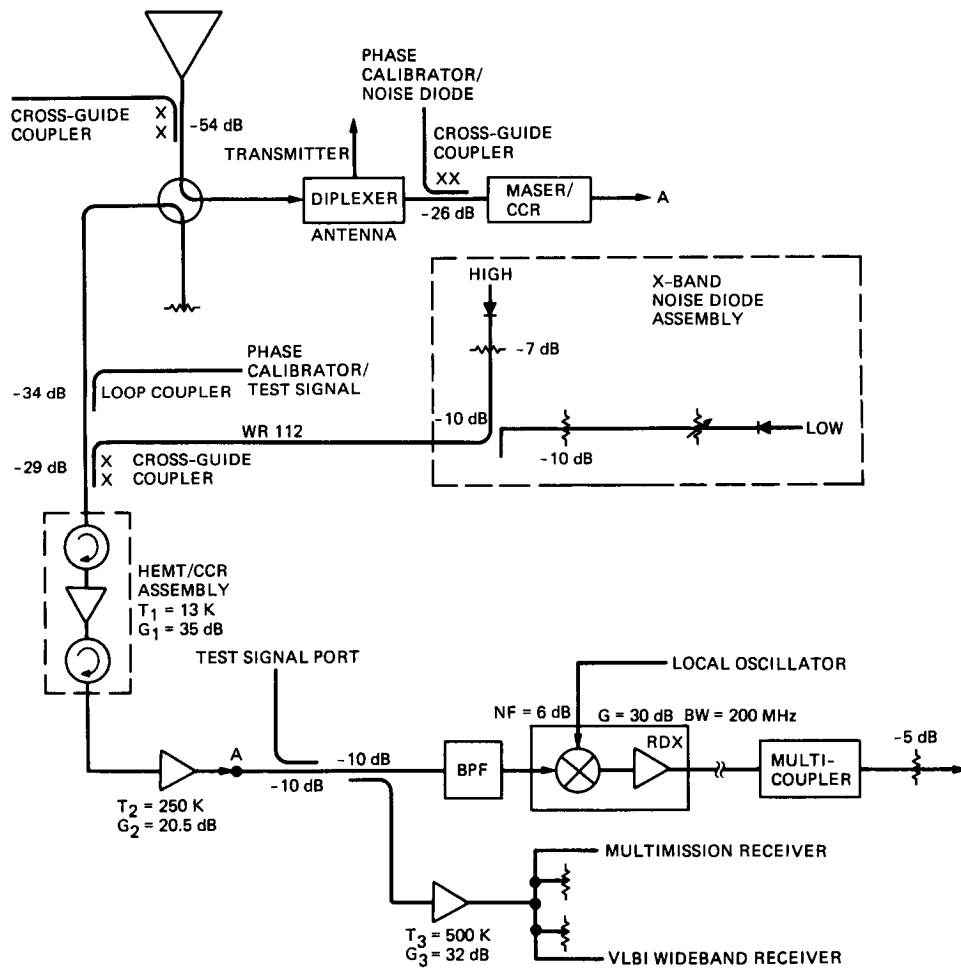


Fig. 1. DSS-13 X-band signal distribution

ORIGINAL PAGE IS
OF POOR QUALITY

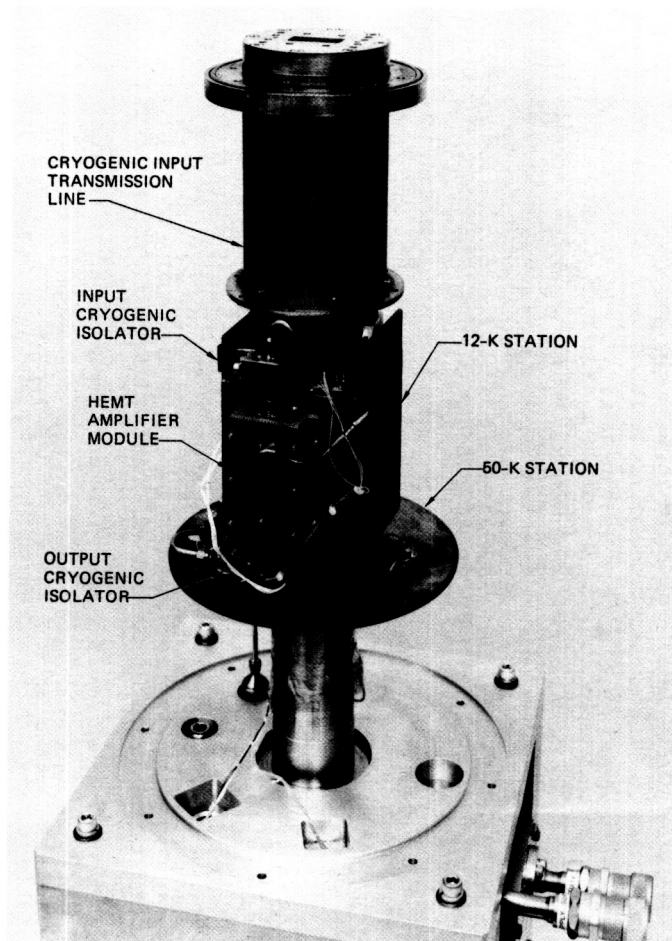


Fig. 2. HEMT/CCR package with vacuum housing removed

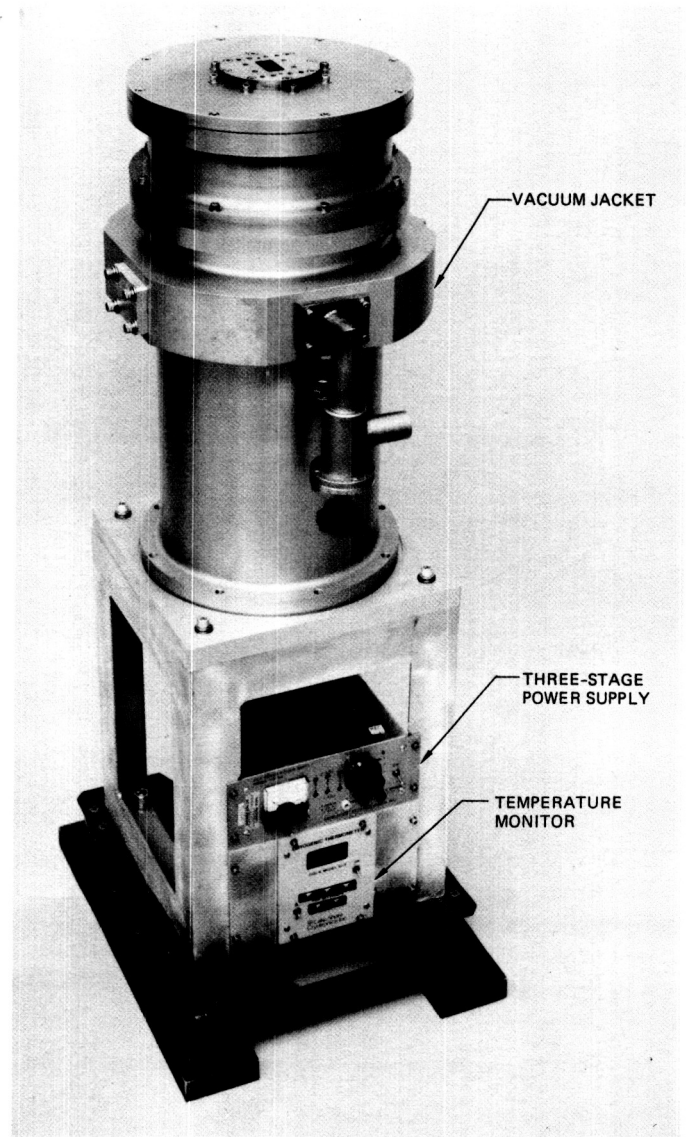


Fig. 3. HEMT/CCR assembly

ORIGINAL PAGE IS
OF POOR QUALITY

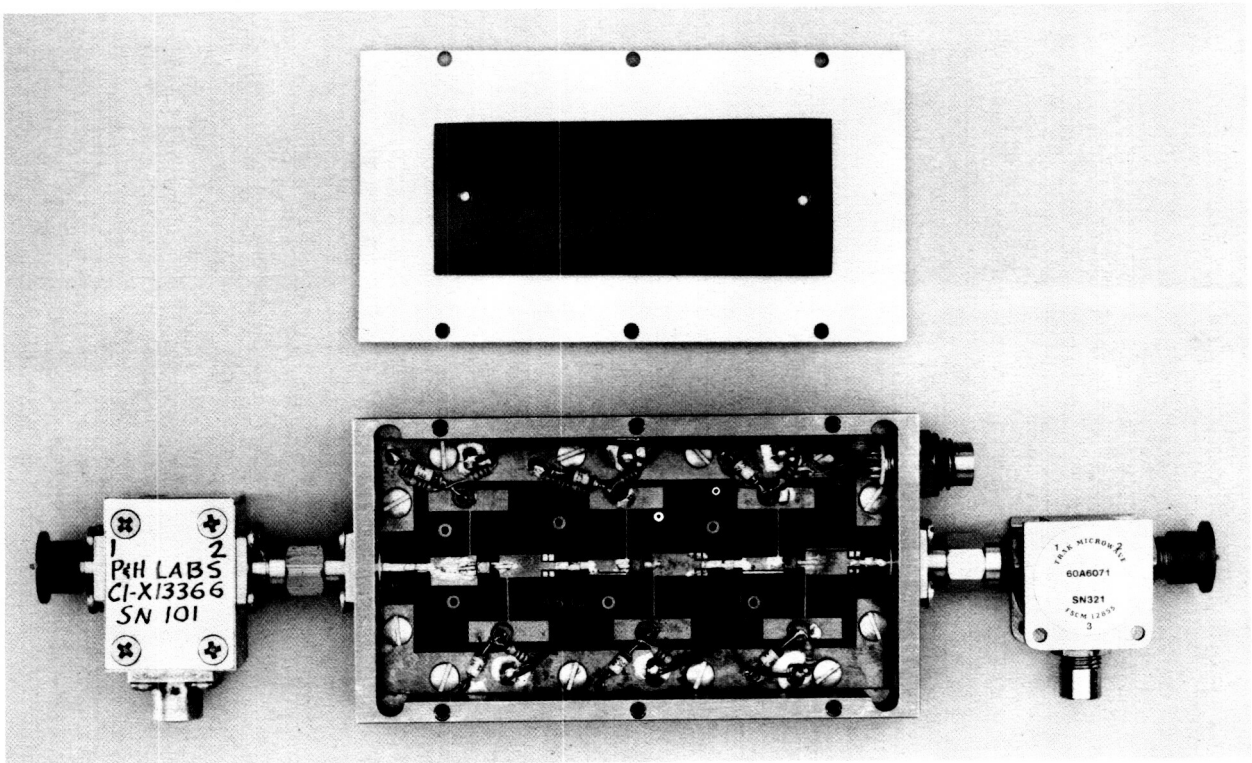


Fig. 4. Three-stage HEMT/FET/FET amplifier module

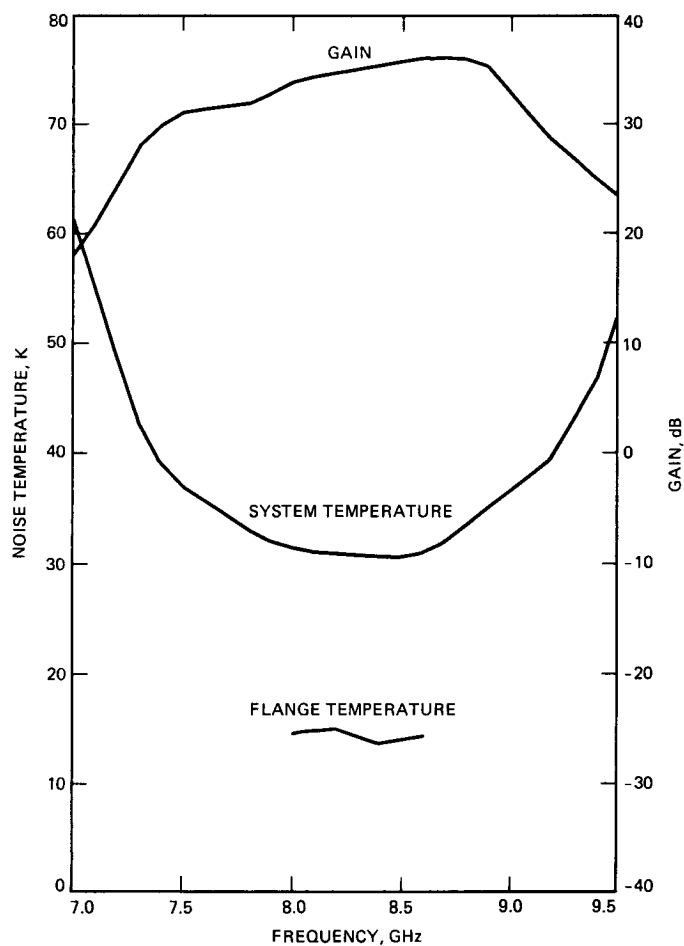


Fig. 5. System temperature, flange temperature, and gain measurements

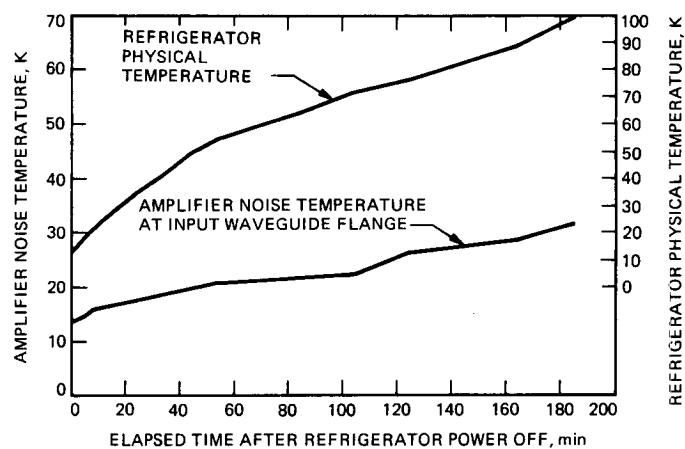


Fig. 6. Input noise temperature and refrigerator physical temperature as a function of refrigerator warmup

514-32

165625

TDA Progress Report 42-94

100

1989000529
616391
100

N89 - 10200

April-June 1988

JPL

Design of an Optical PPM Communication Link in the Presence of Component Tolerances

C.-C. Chen

Communications Systems Research Section

A systematic approach is described for estimating the performance of an optical direct-detection pulse-position modulation (PPM) communication link in the presence of parameter tolerances. This approach was incorporated into the JPL optical link analysis program to provide a useful tool for optical link design. Given a set of system parameters and their tolerance specifications, the program will calculate the nominal performance margin and its standard deviation. Through use of these values, the optical link can be designed to perform adequately even under adverse operating conditions.

I. Introduction

Uncertainties in system parameters have a strong impact on the design of deep-space communication links. Traditionally, the design practice for communication systems is to reserve sufficient power margin to account for the parameter uncertainties. However, for deep-space systems in which the system power is at a premium, how to trim the design margin and still maintain a sufficiently high confidence range in system performance is an important problem that can be solved only by a systematic approach.

For radio frequency (RF) systems, a rigorous and well-established design procedure has been identified [1] based on extensive experience with RF system design. In this procedure, each parameter in the link control table will be specified by its *design value*, *favorable tolerance*, and *adverse tolerance*. The *design value* is the best estimate of the parameter under normal operating conditions. The *adverse* and *favorable tolerances* are derived based on past experience with the par-

ticular system component. These tolerance values are determined so that the actual parameter value generally falls within the specified tolerances. Probability distribution models of these parameters are also constructed based on experience. From these specifications, the mean and variance of each link control table entry can be calculated. These entries are then tabulated so that the final link margin can be calculated.

A similar problem occurs in the design of optical deep-space links. Despite the fact that optical systems generally consume less power than comparable RF systems, the scarcity of prime system power still implies that the communication system must be designed with a tight performance margin. Unfortunately, analysis of the optical link is much more complicated than that of the typical RF system. This is because RF systems' performance depends only on the receiver signal-to-noise ratio (SNR), whereas the performance of the optical link depends not only on the SNR but also on the actual signal and noise powers [2]. Furthermore, in contrast to RF systems, in which

extensive design experience has been accumulated, comparatively little experience has been acquired for optical link design. As a result, larger uncertainties in parameter values can be expected for optical systems.

This article describes a systematic approach to estimate the performance of an optical direct-detection PPM communication link in the presence of uncertainties in component values. Section II outlines the standard procedure for calculating the performance of an optical communication link. Some shortcomings of this procedure are identified. The procedures and assumptions used to calculate the link control table in the presence of system parameter tolerances are then summarized and discussed in Section III.

II. Performance of an Optical Link

Given an optical system with source laser power P_T , the amount of signal power received by the detector is given by [3] as

$$P_S = P_T \eta_T G_T L_T \left(\frac{\lambda}{4\pi z} \right)^2 G_R \eta_R \eta_{\text{atm}} \eta_F \quad (1)$$

where η_T and η_R are the efficiencies of the transmitter and receiver optics, G_T and G_R are the transmitter and receiver antenna gains, λ is the optical wavelength, z is the link distance, L_T is the transmitter pointing loss factor, η_{atm} is the atmospheric transmission factor, and η_F is the narrowband filter transmission factor. The factor $(\lambda/4\pi z)^2$ is known as the space loss factor.

The transmitter antenna gain G_T is a function of the operating wavelength and the aperture diameter [4]. For a Gaussian input signal, G_T is given by [4] as

$$G_T = \left(\frac{\pi D_T}{\lambda} \right)^2 \frac{2}{\alpha_T^2} \left[e^{-\alpha_T^2} - e^{-\alpha_T^2 \gamma_T^2} \right]^2 \quad (2)$$

where D_T is the aperture diameter, d_T is the obscuration diameter, $\gamma_T = d_T/D_T$ is the obscuration ratio of the transmitter, and $\alpha_T \approx 1.12 - 1.30\gamma_T^2 + 2.12\gamma_T^4$ is the optimal truncation ratio of the Gaussian beam. Similarly, the receiver antenna gain can be related to the receiver optics and obscuration diameters by

$$G_R = \left(\frac{\pi D_R}{\lambda} \right)^2 (1 - \gamma_R^2) \quad (3)$$

where $\gamma_R = d_R/D_R$ is the receiver obscuration ratio.

The pointing loss factor L_T is a function of the transmitter antenna parameters and the transmitter pointing bias and jitter. Given an instantaneous pointing error, ϕ , the farfield intensity of the transmitted signal is reduced from its maximum by a factor [4] of

$$L_T(\phi) = \frac{\int_{\gamma_T^2}^1 e^{-\alpha_T^2 u^2} J_0(\pi d_T \phi u / \lambda) u du}{\int_{\gamma_T^2}^1 e^{-\alpha_T^2 u^2} u du} \quad (4)$$

The pointing efficiency of the transmitter, given in terms of the transmitter pointing bias error and the rms jitter, is simply the average of $L_T(\phi)$ over the probability distribution of ϕ . For a two-dimensional spatial tracking system, transmitter pointing errors in the azimuth and altitude directions can be modeled as independently distributed Gaussian random variables so that the resulting radial pointing error is Rician distributed. By averaging $L_T(\phi)$ over the distribution of ϕ , the pointing loss factor L_T can be written as

$$L_T = \int_0^\infty L_T(\phi) \frac{\phi}{\sigma_T^2} \exp\left(-\frac{\phi^2 + \epsilon_T^2}{2\sigma_T^2}\right) I_0\left(\frac{\phi \epsilon_T}{\sigma_T^2}\right) d\phi \quad (5)$$

where ϵ_T is the static pointing error and σ_T is the root-mean-square (rms) jitter in the transmitter line of sight.

In addition to the transmitted signal, the optical receiver also collects background radiation from other sources. Given the total irradiance of the noise source, W_N , the amount of noise power collected by the receiver can be written as

$$P_B = \eta_F \frac{\pi D_R^2}{4} (1 - \gamma_R^2) \frac{\pi \Theta^2}{4} W_N \Delta\lambda \quad (6)$$

where Θ is the receiver diameter field of view and $\Delta\lambda$ is the narrowband filter bandwidth.

As was previously stated, the performance of the optical link depends on both signal and background powers. Given P_S and P_B , the quantities of signal and background photons detected by the receiver are Poisson-distributed random variables with means

$$K_S = \left(\frac{\eta_D \lambda}{hc} \right) P_S T_w \quad (7)$$

$$K_B = \left(\frac{\eta_D \lambda}{hc} \right) P_B T_s \quad (8)$$

where η_D is the detector quantum efficiency and T_w and T_s are the PPM word and slot widths, respectively. The word width is related to the PPM order, M , and the data rate, R_b , by

$$T_w = \frac{\log_2 M}{R_b} \quad (9)$$

and the slot width is related to T_w and the dead time T_d by

$$T_w = MT_s + T_d \quad (10)$$

Since $T_w > T_s$, Eq. (7) shows that the peak signal power is much higher than the average power. This is because in a PPM signaling scheme, the laser is turned on only during the signal time slot, while during the rest of the word period, no signal is transmitted.

Given the expected photocounts, K_S and K_B , the bit error rate (BER) of an M -ary optical PPM link can be written [2] as

$$\begin{aligned} BER = & \frac{M}{2(M-1)} \left[1 - \frac{1}{M} e^{-(K_S + MK_B)} \right. \\ & - \sum_{k=1}^{\infty} \frac{(K_S + K_B)^k}{k!} e^{-(K_S + K_B)} \\ & \left. \times \left(\sum_{j=0}^{k-1} \frac{K_B^j}{j!} e^{-K_B} \right)^{M-1} \left(\frac{(1+a)^M - 1}{Ma} \right) \right] \quad (11) \end{aligned}$$

where

$$a = \frac{K_B^k}{k! \sum_{j=0}^{k-1} \frac{K_B^j}{j!}}$$

The procedure described above can be used to calculate effectively the expected BER of an optical channel. Given the required system BER, the signal power can also be iterated to achieve a desired power margin.

In some instances, having to repeat the calculation for several different links can be a tedious and time-consuming task. In order to ease the design of optical communication links, a simple yet elegant optical link analysis program was developed

by W. Marshall and B. Burk in 1986 [3], [5]. The objective of this program is to predict the performance of an optical communication link given a set of component and operational parameters as well as the noise source specification. A list of system parameters needed to specify the optical link is shown in Table 1. After all system parameters are entered, the program calculates and displays the link control table. A sample link control table for an Earth-Saturn link generated by this program is shown in Table 2.

III. Performance Estimate in the Presence of Parameter Uncertainties

The simple link analysis program is very useful in providing a preliminary estimate of the link performance. For systems in which all component and operational parameters can be precisely specified, the simple link analysis program is sufficient. In most systems, however, the parameters may not be specified precisely. For instance, the atmospheric transmission factor can vary from less than 2 dB on a clear day to over 200 dB in a thunderstorm. Components may degrade over time so that their performance specification cannot be given accurately. Accidents and interactions with interplanetary environments can also reduce the efficiency of the optical system. For these reasons, a systematic approach must be devised for the design of optical links in the presence of parameter uncertainties.

The optical link tolerance estimate program is designed to provide a simple analytic tool for estimating the performance of an optical link in the presence of uncertainties in component and operating parameters. Since most parameters are susceptible to time-dependent degradations, tolerance specifications must be given on these parameters. The parameters that must have their tolerances specified include the source power, the optics efficiencies, the detector quantum efficiency, the atmospheric transmission factor, and the narrow-band filter transmission factor. The probability distribution of these parameters must also be specified based on sample distributions. From the tolerance specification and the method of distribution, an estimate of the parameter variance can be derived. The procedure for determining the variance is similar to the one used in RF system design.

Some critical link parameters, however, can best be represented as functions of the basic physical quantities. These parameters include the transmitter and receiver antenna gains, the transmitter pointing loss, and the quantity of background photons received. The functional dependencies are in general very complicated so that it is infeasible to derive the tolerance specifications on these parameters based on the tolerance specifications of basic component parameters. Furthermore, the actual performance of these parameters can also depend on

factors not previously considered in the ideal link analysis. For instance, the transmitter antenna gain G_T depends not only on the aperture and obscuration diameters but also on the surface tolerance and the incoming beam quality. Deviations from the ideal surface and optical wavefront can result in a degraded antenna gain. Consequently, instead of specifying tolerances on the aperture and obscuration diameters, *adverse* and *favorable tolerances* will be specified directly for the transmitter and receiver antenna gains. The transmitter pointing efficiency is also a complex function of the component parameters. For simplicity, tolerance values will be specified directly on the transmitter pointing efficiency rather than on the static and rms pointing errors. Similarly, tolerance values will be specified for the noise photocount rather than for the receiver FOV and the narrowband filter bandwidth.

Finally, those parameters that can be specified exactly will be given no tolerance specifications. These parameters include the order of the PPM, the slot width and the modulation dead time, and the required bit error rate (BER). The link distance and the laser wavelength are also predetermined parameters. These parameters will be entered without tolerance specifications.

Once all the system parameters are properly specified, the amount of signal power needed to achieve the desired error performance (receiver sensitivity) can be calculated. In general, the receiver sensitivity is a function of the noise power and the modulation format. Since the noise power received by the detector varies for different values of the system parameters, the required signal power must vary accordingly. Unfortunately, the required signal power cannot be related to the noise power by a simple functional form. This can easily be seen from the complexity of the BER expression in Eq. (11). The lack of a simple functional dependence implies that the statistics of the required signal level cannot be deduced easily from the probability distribution of the noise count. Some simplifications must therefore be made before the tolerance on the required signal level can be calculated.

One such simplification is the functional dependence of the required signal level on the background strength. Under the condition of weak background, the BER can be approximated by the Union-Chernoff bound [2]:

$$BER \approx (M-1)e^{(\sqrt{K_S+K_B}-\sqrt{K_B})^2} \quad (12)$$

By fixing the BER and solving for the required signal level as a function of K_B , it is seen that

$$K_S \approx c + \sqrt{4cK_B} \quad (13)$$

where $c = \ln[BER/(M-1)]$. For a small fluctuation of background, $K_B = \bar{K}_B + \Delta K_B$, K_S varies as

$$K_S \approx \bar{K}_S + \sqrt{\frac{c}{\bar{K}_B}} \Delta K_B \quad (14)$$

Therefore, when the noise fluctuation is small compared to the average noise level, the required signal level can be assumed to have the same statistical dependence as the background level.

By using the above approximation, the required signal level can be calculated by iterating Eq. (11) given the expected noise photocount. The variance of the required signal power can then be calculated directly based on the variance of the noise power, or the favorable tolerance values of the noise power can be substituted into the BER expression to calculate the *favorable* and *adverse* required signal photocounts. From these values and the assumption that the required signal photocount has the same statistical dependence as the background photocounts, the variance on the required signal level can be calculated.

Once all the system parameters and the required signal power level have been calculated, the generation of the link control table is straightforward. Since the system power margin is generally expressed in terms of decibels, all parameters and their tolerances should be converted into decibels before they are entered into the link control table. This is done as follows: Given a parameter $x = \bar{x} + \Delta x$, where $E[(\Delta x)^2] = \sigma_x^2$, the value in decibels is given by

$$y = 10 \times \log x \approx 10 \times \left[\log \bar{x} + \frac{\Delta x}{(\ln 10)\bar{x}} \right] \quad (15)$$

The standard deviation of y can therefore be approximated as

$$\sigma_y \approx \frac{10\sigma_x}{(\ln 10)\bar{x}} \quad (16)$$

When all parameters and their variance values are computed, the final power margin can be calculated by algebraically summing all the loss factors. Given the independent parameter assumption, the final link performance variance is simply the sum of all parameter variances. Table 3 shows a typical output of the tolerance link control program for the same Earth-Saturn link given by Table 2. The program also calculates and displays the 3σ value of the link performance. Some parameter values are different because the link control table now displays the average values of the parameters instead of the design values. Note that the standard link analysis results in a 5.2-dB

margin. When component tolerances are considered, however, the 3σ margin is only 0.8 dB.

IV. Conclusions

The inclusion of tolerance calculations in the existing link analysis program gives the link designer a simple and effective tool for estimating the performance of a deep-space optical PPM communication link. The inclusion of tolerance calcula-

tions will permit the design of a deep-space link with sufficient power margin even under adverse operating conditions. At the same time, by minimizing the required system power while maintaining a confident operating margin, the cost of the system can be minimized without seriously affecting link performance. Finally, the use of rigorous design methodology allows critical link parameters to be identified. Improvements in these parameters can then be directly reflected in the reduction of performance uncertainty.

References

- [1] J. Yuen, *Deep Space Telecommunications Systems Engineering*, New York: Plenum Press, 1983.
- [2] R. Gagliardi and S. Karp, *Optical Communications*, New York: Wiley, 1975.
- [3] W. Marshall and B. Burk, "Received Optical Power Calculations for Optical Communications Link Performance Analysis," *TDA Progress Report 42-87*, vol. July-September 1986, Jet Propulsion Laboratory, Pasadena, California, pp. 32-40, November 15, 1986.
- [4] B. Klein and J. Degnan, "Optical Antenna Gain: I. Transmitting Antennas," *Appl. Opt.*, vol. 13, no. 10, pp. 2397-2401, October 1974.
- [5] J. Lesh, W. Marshall, and J. Katz, "A Simple Method for Designing or Analyzing an Optical Communication Link," *TDA Progress Report 42-85*, vol. January-March 1986, Jet Propulsion Laboratory, Pasadena, California, pp. 25-31, May 15, 1986.

Table 1. Component and operational parameters of a typical optical link

Component parameters:	
Operating wavelength of the transmitter laser, λ	
Average laser output power, P_T	
Diameter of the transmitter aperture, D_T	
Diameter of the transmitter center obscuration, d_T	
Transmitter optics efficiency, η_T	
Transmitter pointing bias error, ϵ_T	
Transmitter rms pointing jitter, σ_T	
Diameter of the receiver aperture, D_R	
Diameter of the receiver center obscuration, d_R	
Receiver optics efficiency, η_R	
Narrowband filter transmittance, η_F	
Narrowband filter spectral bandwidth, $\Delta\lambda$	
Detector quantum efficiency, η_D	
Detector field of view, Θ	
Operational parameters:	
Alphabet size, M	
System data rate, R_b	
Modulation slot width, T_s	
Link distance, z	
Atmospheric transmission factor, η_{atm}	
Desired/required link bit error rate, BER	

Table 2. Link control table of an Earth-Saturn link generated by the optical link analysis program

Link specifications			
Component parameters:			
Wavelength, μm	0.532		
Average laser output power, W	2.000		
Modulation extinction ratio	0.100E+06		
Diameter of transmitter aperture, m	0.300		
Obscuration diameter of transmitter, m	0.600E-01		
Transmitter optics efficiency	0.650		
Transmitter pointing bias error, μrad	0.100		
Transmitter rms pointing jitter, μrad	0.100		
Diameter of receiver aperture, m	10.000		
Obscuration diameter of receiver, m	4.280		
Receiver optics efficiency	0.380		
Narrowband filter transmission factor	0.500		
Filter spectral bandwidth, \AA	10.000		
Detector quantum efficiency	0.350		
Detector diameter field of view, μrad	100.000		
Receiver type (ideal = 0, APD-based = 1)	0.000		
Operational parameters:			
Alphabet size (M = ?)	256.000		
Data rate, kbps	114.350		
Dead time, μsec	67.401		
Slot width, nsec	10.000		
Distance between transmitter and receiver, AU	9.000		
Atmospheric transmission factor	0.500		
Required link bit error rate	0.200E-01		
Noise sources:			
Saturn receiver to source distance, AU	9.000		
		Factor	Decibels
Link control tables			
Laser output power, W		2.000	33.0 dBm
Minimum required peak power, W	0.130E+05		
Transmitter antenna gain		0.222E+13	123.5
Antenna diameter, m	0.300		
Obscuration diameter, m	0.060		
Beamwidth, μrad	3.068		
Transmitter optics efficiency		0.650	-1.9
Transmitter pointing efficiency		0.980	-0.1
Bias error, μrad	0.100		
RMS jitter, μrad	0.100		
Space loss (9.00 AU)		0.989E-39	-390.0
Atmospheric transmission factor		0.500	-3.0
Receiver antenna gain		0.285E+16	154.5
Antenna diameter, m	10.000		
Obscuration diameter, m	4.280		
Field of view, μrad	100.000		
Receiver optics efficiency		0.380	-4.2
Narrowband filter transmission		0.500	-3.0
Bandwidth, \AA	10.000		

Table 2. (contd)

		Factor	Decibels
Link control tables (continued)			
Received signal power, W	0.400E-09	0.758E-12	-91.2 dBm
Received background power, W			
Detector quantum efficiency		0.350	-4.6
Photons/joule		0.268E+19	154.3 dB/mJ
Detected signal PE/second		0.711E+06	58.5 dB/Hz
Symbol time, sec		0.700E-04	-41.6 dB/Hz
Detected signal PE/symbol		46.500	16.7
Required signal PE/symbol (ideal)		13.900	11.4
Detected background PE/slot	3.750		
Margin		3.340	5.2

Table 3. Link control table of an Earth-Saturn link when component value tolerances are included in the calculation

	Design value	Favorable tolerance	Adverse tolerance
Link specifications			
Component parameters:			
Wavelength, μm	0.5320		
Transmitter power, W	2.0000	0.5E-01	0.2
Transmitter aperture, m	0.3000		
Transmitter obstruction, m	0.6000E-01		
Transmitter antenna gain, dB	123.4700	0.0	1.0
Transmitter optics efficiency	0.6500	0.2E-01	0.5E-01
Transmitter pointing bias, μrad	0.1000	0.0	0.0
Transmitter pointing jitter, μrad	0.1000		
Transmitter pointing loss, dB	-0.8821E-01	0.3E-01	0.3E-01
Receiver aperture, m	10.0000		
Receiver obstruction, m	4.2800		
Receiver antenna gain, dB	154.5500	0.0	1.0
Receiver optics efficiency	0.3800	0.3E-01	0.4E-01
Filter transmission	0.5000	0.3E-01	0.3E-01
Filter bandwidth, \AA	10.0000		
Detector efficiency	0.3500	0.1	0.5E-01
Detector FOV, μrad	100.0000		
Operational parameters:			
Alphabet size ($M = ?$)	256.0000		
Data rate, kbps	114.3500		
Link length, AU	9.0000		
Dead time, μsec	67.4010		
Slot width, nsec	10.0000		
Atmospheric transmission factor	0.5000	0.2	0.2
Link BER	0.2000E-01		
Noise count/slot	3.9659	0.5	0.5
Noise sources:			
Saturn receiver to source distance, AU		9.0	
	Factor	Decibels	Variance
Link control tables			
Laser output power, W	1.9250	32.84 dBm	0.03
Minimum required peak power, W	0.1347E+05		
Transmitter antenna gain	0.1983E+13	122.97	0.03
Antenna diameter, m	0.3000		
Obscuration diameter, m	0.6000E-01		
Beamwidth, μrad	3.0680		
Transmitter optics efficiency	0.6350	-1.97	0.02
Transmitter pointing efficiency	0.9799	-0.09	0.00
Bias error, μrad	0.1000		
RMS jitter, μrad	0.1000		
Space loss (9.00 AU)	0.9887E-39	-390.05	
Atmospheric transmission factor	0.5000	-3.01	1.01
Receiver antenna gain	0.2539E+16	154.05	0.03
Antenna diameter, m	10.0000		
Obscuration diameter, m	4.2800		
Field of view, μrad	100.0000		
Receiver optics efficiency	0.3750	-4.26	0.05

Table 3. (contd)

Link control tables (continued)	Factor	Decibels	Variance
Narrowband filter transmission	0.5000	-3.01	0.02
Bandwidth, Å	10.0000		
Received signal power, W	0.5588E-12	-92.53 dBm	1.18
Background power, W	0.5015E-09		
Detector quantum efficiency	0.3750	-4.26	0.25
Photons/joule	0.2678E+19	154.28 dB/mJ	0.00
Detected signal PE/second	0.5612E+06	57.49 dB/Hz	1.44
Symbol time, sec	0.6996E-04	-41.55 dB/Hz	
Detected signal PE/symbol	39.2600	15.94	1.44
Detected background PE/slot	3.9660		
Required signal PE/symbol	14.2000	11.52	0.00
Margin	2.7660	4.42	1.44
3σ		± 3.60	

5/5-32

~~145626~~

TDA Progress Report 42-94

~~145626~~

1989000000
6/19/81
2/p.

N89 - 10201

April-June 1988

JT

Atmospheric Propagation Effects Relevant to Optical Communications

K. S. Shaik

Communications Systems Research Section

A number of atmospheric phenomena affect the propagation of light. This article reviews the effects of clear-air turbulence as well as atmospheric turbidity on optical communications. Among the phenomena considered are astronomical and random refraction, scintillation, beam broadening, spatial coherence, angle of arrival, aperture averaging, absorption and scattering, and the effect of opaque clouds. An extensive reference list is also provided for further study. Useful information on the atmospheric propagation of light in relation to optical deep-space communications to an earth-based receiving station is available, however, further data must be generated before such a link can be designed with committed performance.

I. Introduction

There is considerable interest in the development of optical communication systems for space applications. Use of optical frequencies will result in very high gain antennas as well as potentially enormous channel capacities. The interaction of electromagnetic waves with the atmosphere at optical frequencies is stronger than that at microwave frequencies. Hence, it is important to show that laser communication systems are capable of operating within the atmosphere with predictable statistics for availability and reliability.

There are several phenomena that affect the manner of light propagation through the atmosphere. A laser beam propagating through the atmosphere can quickly lose its energy due to molecular scattering, molecular absorption, and particulate scattering. Refractive turbulence may also contribute to

energy loss; however, it mainly degrades the beam quality, both by distorting the phase front and by randomly modulating the signal power. The presence of opaque clouds may occlude the signal completely, rendering the line-of-sight communication link useless. The problems described above are quite distinct from each other, and the difficulties presented by each of these obstacles need to be studied and understood independently.

The next section will identify and describe various ingredients of an atmospheric model that must be studied in some detail before a coherent view of optical communications through the atmosphere can be developed. Sections III and IV describe the optical consequences of the model atmosphere for the propagation of laser beams. The focus will be on those results that are especially relevant to optical communications.

Also suggested will be methods and strategies to minimize the adverse effects of the atmosphere on optical communication links. Section V then concludes the discussion by putting the future of optical communications research in perspective.

II. Description of the Atmosphere

The atmosphere is a dynamic system of considerable complexity. It is obvious that the entire set of phenomena that characterizes the atmosphere and its interaction with laser radiation needs to be understood before useful optical systems can be designed for operation in the atmosphere. Various components of the atmosphere will be considered in the subsections that follow.

The atmosphere is usually divided into a number of layers based on its mean temperature profile. From the ground to an altitude of 10 to 12 km, the mean temperature decreases steadily. This lowest layer is called the troposphere. For the next layer, which is called the stratosphere, the temperature increases with altitude. Unlike the troposphere, the air in this layer is very stable, and turbulent mixing is inhibited in the stratosphere due to the inverted temperature profile. Low turbulence and the absence of rainfall account for the long residence time of aerosols and other particulate matter in the stratosphere. Figure 1 shows the mean temperature of the atmosphere as a function of altitude at a 45-degree north latitude during July [1].

A. Chemical Composition

Table 1 gives a list of dry clear-air chemical components of the atmosphere near sea level [2]. The water vapor content of the atmosphere in the troposphere is highly variable and ranges from 1 to 3 percent in concentration. A number of minor constituents, all influencing turbidity at optical frequencies, are also found in the atmosphere in varying concentrations. Some of these constituents are aerosols, oxides of carbon, compounds of sulfur and nitrogen, hydrocarbons, and ozone.

B. Turbidity

For the present discussion, turbidity is defined to consist of all particulate matter which absorbs and scatters light. For the study of optical beam propagation, atmospheric turbidity can be roughly divided into two classes. The first includes gas molecules, aerosols, light fog and haze, and thin cirrus clouds. The attenuation or extinction loss of light energy from the laser beam for this category is generally due to Rayleigh and Mie scattering. The direct beam retains a fair percentage of its energy even after traveling through the entire atmosphere. The second class of turbidity consists of opaque clouds and of dense fog and haze. Scattering losses for this case are very high; most of the beam energy appears as diffused light. For

optical communications, the strategies to overcome the problems posed by the two categories are quite different and will be discussed later.

C. Astronomical Refraction

The index of refraction, which depends on the density of the atmosphere, decreases with height above ground. Light arriving at the top of the atmosphere on a slant path bends downward, which causes the observed zenith distance of the transmitter to be different from the true zenith. The angular distance between the true and the apparent zenith angle of the object, referred to as the terrestrial refraction angle, can be obtained by applying Fermat's principle to the atmospheric profile. Garfinkel [3] has prepared a computer program to compute the magnitude of the refraction angle for any atmospheric profile, including the *U.S. Standard Atmosphere* [4] for all apparent zenith angles. For a grazing ray at sea level, the terrestrial refraction angle can be as large as 10 mrad [5].

The refractivity, N , of the atmosphere for optical wavelengths can be approximated by the relation [5]

$$N \approx 79 \times 10^6 (n - 1) P/T \quad (1)$$

where n is the refractive index of the atmosphere, P is the atmospheric pressure in millibars, and T is temperature in kelvins. For a detailed analysis of optical refractive index and more accurate formulas, see [6] - [8]. N is about 290 at sea level [5]; it varies about 10 percent with wavelength over the visible range and by 0.5 percent with humidity [9].

D. Random Refraction

The average profile of the atmosphere, which follows from the meteorological conditions, determines the regular or mean refraction of light beams. Random or stochastic refraction is due to the motion of inhomogeneities in the air. It is this type of refraction that causes the light rays to wander in time and degrades their spatial as well as temporal coherence. Stochastic refraction may impose some special requirements on the pointing and tracking mechanisms of optical systems. It may also restrict the rate of data transmission on optical channels.

Systematic astronomical observations of the refraction angle have led to the discovery of several types of random oscillations. For most practical purposes, they can be put in two classes: (1) oscillations with frequencies of 1 Hz or less; and (2) rapid oscillations with frequencies of 10 Hz or higher. High-frequency oscillations will be discussed in detail in the following section.

Slow oscillations arise due to general shifts of weather patterns and air masses in time. Such oscillations cause the image

to drift slowly with amplitudes attaining several arc seconds [9]. Beckmann [10] and Hodara [11] have observed random refraction along horizontal paths, noting a slow drift of laser beams over paths of 5 to 15 km at a rate of several micro-radians per hour. Another study estimates that inhomogeneities in the atmosphere have a scale length of 10 to 40 km and that the fluctuations of the beam direction are as high as $75 \mu\text{rad}$ [9]. Lese [12], using a 0.9-m telescope, has measured angular deviations with a mean value of $3 \mu\text{rad}$ for zenith angles of less than a radian for the entire atmosphere.

E. Clear-Air Turbulence

Refractive turbulence of the atmosphere is caused by rapid, small-scale spatial and temporal fluctuations in temperature (on the order of 0.1 to 1.0 K). While the deviations of the refractive index from their average values are very small (a few parts per million), the cumulative effect of such inhomogeneities over large distances of practical interest can be quite significant.

Turbulence results from disordered mixing of air in the atmosphere. A flow becomes turbulent when the Reynolds number, $Re = \nu L / \mu$, for a flow process exceeds a critical value. Here, ν is a characteristic flow velocity, L is some scale size of the flow process, and μ is the kinetic viscosity of the fluid. For $L = 2 - 10 \text{ m}$, $\nu = 1 - 5 \text{ m/s}$, and $\mu = 15 \times 10^{-6} \text{ m}^2/\text{s}$, the Reynolds number is on the order of 10^6 . Such large Reynolds numbers, which are typical of the atmosphere, usually correspond to fully developed turbulent flow.

The kinetic energy of turbulence is usually introduced by wind shear or convection from solar heating at scale sizes of the inhomogeneities in the atmosphere, $L \geq L_0$, where L_0 is called the outer scale. The kinetic energy of large-scale motions characterized by the outer scale is transferred to increasingly small-scale inhomogeneities by turbulent means. When the Reynolds number, which depends on the scale size of motions, becomes small enough, the mechanism for dissipation of energy becomes predominantly viscous rather than turbulent. This transition takes place for a scale size l_0 , which will be called the inner scale of the turbulent flow. Typically, $l_0 \sim 1 - 10 \text{ mm}$ and $L_0 \sim 10 - 100 \text{ m}$ in the troposphere; close to the earth's surface, L_0 can be approximated by the height from the ground.

The flow for scale sizes L , where $l_0 \leq L \leq L_0$, is then strongly turbulent, with velocity gradients occurring in all possible directions randomly in time and space. For the present purposes, we may view the atmosphere to be composed of vortices or blobs of homogeneous fluid of sizes between l_0 and L_0 , which have dissimilar temperatures and pressures from their neighboring vortices, and which are mixing chaoti-

cally. It is necessary to use stochastic methods to explain and interpret atmospheric turbulence.

Modern understanding of atmospheric turbulence is based on the Kolmogorov-Obukhov theory. The range of applicability of their theory, referred to as the inertial range, is between the scale sizes l_0 and L_0 . Tatarski based his work on their theory to obtain results relevant to the propagation of electromagnetic waves through the turbulent atmosphere. The refractive index $n(r)$ of the atmosphere can be expressed as

$$n(r) = E[n(r)] + n_1(r) \quad (2)$$

where $E[\cdot]$ represents ensemble averaging, n_1 is the refractive index fluctuation with $E[n_1(r)] = 0$, and $E[n(r)] \approx 1$ for the atmosphere. The characteristics of the fluctuation may be expressed by a structure function which obeys the Kolmogorov-Obukhov $2/3$ law:

$$D_n(r) = E[|n_1(r + r_1) - n_1(r_1)|^2] = C_n^2 r^{-2/3} \quad (3)$$

where the structure constant C_n represents the strength of turbulence. Typically the values of C_n range from 10^{-9} for weak turbulence to 10^{-7} for strong turbulence. For a more complete account of turbulence spectrum, the refractive structure constant and its relation to the temperature structure constant, see [13] - [16].

Extensive experimental data available today confirm the validity of theoretical results fairly well for the atmospheric layers of altitudes higher than 50 m. For heights lower than 50 m during daytime, the shape of the structure constant is better approximated by $C_n(z) \sim r^{-4/3}$ [9].

For heights greater than 3 km above sea level, the Hufnagel model provides a good approximation for the refractive index structure constant [17] - [22]. According to the model, $C_n^2(z)$ may be expressed as

$$C_n^2(z) = 2.72 \times 10^{-16} \left\{ 3V^2 \left(\frac{z}{10} \right)^{10} \exp[-z] + \exp \left[-\frac{z}{1.5} \right] \right\} (m^{-2/3}) \quad (4)$$

where V is the wind speed in meters per second and z is the altitude in kilometers. Measurements by Barletti et al. [23] and Vernin et al. [24] show that for $z \geq 4 \text{ km}$, the data are nearly independent of the site location and agree quite well

with the Hufnagel model. The data also indicate little variation in the value of the structure constant with seasons.

III. Effects of Turbulence on Optical Beams

Most studies, after Tatarski, employ the hypothesis of "frozen" turbulence to model optical propagation through the atmosphere. The approximation consists of assuming that the temporal variations at any point result from a uniform, cross-beam motion of the atmosphere as a whole due to prevailing winds. The changes in the internal structure of the atmosphere due to evolution of turbulence in time are neglected.

Several mathematical techniques, including diagrammatic methods [25]–[31], coherence theory [32]–[34], Markov approximation [13], [35]–[37], and others [38], [39], have been used to solve the wave equation for the propagation of light in order to study the effects of turbulence. Most of these techniques are equivalent and yield comparable results, which will be reviewed in the following paragraphs.

A. Scintillation

Stellar scintillation is a well-known phenomenon. Turbulence causes fluctuations in the intensity of a light wave by redistributing its power spatially in time. The strength of scintillation can be measured in terms of the variance of the beam amplitude or its irradiance at a point. Theoretical investigations have led to the prediction that the log-amplitude, $\chi = \ln[A/A_0]$, where A is the amplitude and A_0 is a normalization factor, has a Gaussian distribution. Also Gaussian is the distribution for the log-intensity or the log-irradiance. Other methods point to a Rice-Nakagami distribution for the amplitude A [40]–[42]. For small variances in χ , the difference between the two is small. However, the solutions are valid when the variance of χ , $\sigma_\chi^2 \leq 0.5$; i.e., the solutions hold for weak turbulence only. The restriction is quite stringent: for horizontal paths near the ground, where the turbulence is strong, the limit may be reached over path lengths of about 1 km. Since C_n^2 decreases rapidly with height above the ground, the problem of optical propagation through the whole atmosphere may still remain amenable to weak turbulence methods for zenith angles of less than a radian. Experimental data seem to favor the log-normal distribution for the amplitude and the irradiance in the weak turbulence region. Theoretical results for strong turbulence are scarce and quite controversial [26]–[31], [33], [34], [37], [39], [43]–[61].

For weak turbulence, the variance in log irradiance, which is the usual parameter that is measured experimentally, can be simply related to the calculated log amplitude by

$$\sigma_{\ln I}^2 = 4\sigma_\chi^2 \quad (5)$$

It can be shown from Tatarski's work that

$$\sigma_{\ln I}^2 \approx 2.24 k^{7/6} (\sec \theta)^{11/6} \int_0^Z C_n^2(z) z^{5/6} dz \quad (6)$$

where $\theta \leq 1$ rad is the zenith angle, $k = 2\pi/\lambda$ is the optical wave number, and Z is the altitude of the source. For larger angles this result may not be very useful as scintillation effects move into the strong turbulence regime. Equation (6) above is insensitive to small values of z due to the $z^{5/6}$ factor in the integrand, and most of the intensity fluctuation effects come from higher altitudes. This justifies the use of the Hufnagel model, which Yura and McKinley [62] have employed to obtain the following result for log-irradiance variance:

$$\sigma_{\ln I}^2 \approx \left[7.41 \times 10^{-2} \left(\frac{V}{27} \right)^2 + 4.45 \times 10^{-3} \right] \lambda^{-7/6} (\sec \theta)^{11/6} \quad (7)$$

where λ is the wavelength in micrometers and V is wind speed in meters per second. Wind speeds at higher altitudes seem to be quite stable diurnally as well as seasonally and do not differ much from site to site. For Maryland, the wind speed appears to be normally distributed with a mean value of 27 m/s and a standard deviation of 9 m/s.

The wavelength dependence of scintillation is apparent from observations of stars near the horizon, where the turbulence is strong. The wavelengths at small elevation angles are decorrelated enough that the stars seem to scintillate with different colors at different times. The effect is much smaller for stars higher in the sky. Quantitative studies of this effect show that the correlation of intensity fluctuations drops to about 0.6 for wavelengths differing by 50 percent [5].

B. Beam Broadening

Consider a Gaussian beam of size W_0 and beam-axis intensity I_0 at the transmitter. Its intensity at distance z in free space is given by [63] as

$$I(z, \rho) = I_0 \left(\frac{W_0}{W_f} \right)^2 \exp \left[-\frac{2\rho^2}{W_f^2} \right] \quad (8)$$

where ρ is the transverse distance from the beam and W_f , the beam size at z for a collimated beam, is

$$W_f^2 = W_0^2 + \left(\frac{2z}{k W_0} \right)^2 \quad (9)$$

For large distances z , the beamwidth W_r in the turbulent medium is given by

$$W_r^2 = W_f^2 + 4.38 C_n^2 l_0^{-1/3} z^3 \quad (10)$$

and the average intensity, $E[I]$ on the beam axis is

$$E[I] = \frac{W_0^2 I_0}{W_r^2} \quad (11)$$

where it is assumed that the beam propagates without loss of power, i.e., the backscattering and absorption are negligible for clear air. A more complete description of both short- and long-term behavior of beam spreading can be found in [59] and [64].

C. Spatial Coherence

Loss of spatial coherence across a light beam is another important effect of clear-air turbulence. Refractive-index inhomogeneities of relatively larger scale sizes produce random phase fluctuations which degrade coherence of the propagating wavefront. Kon and Tatarski [65], Schmeltzer [66], and Ishimaru [67] have studied the problem in detail and have obtained expressions for the structure function of the phase fluctuations, $D_s(\rho_1, \rho_2)$, which is defined to be

$$D_s(\rho_1, \rho_2) = E[|s(\rho_1) - s(\rho_2)|^2] \quad (12)$$

where ρ_1 and ρ_2 are position vectors in the plane of observation across the beam and $s(\cdot)$ is the phase at that point. For weak turbulence, [9] gives a simple result for the phase structure function

$$D_s(0, \rho) = 2.91 b_1 C_n^2 k^2 z \rho^{5/3} \quad (13)$$

where the value of b_1 , a constant, ranges from 1.0 for a plane wave to 0.375 for a spherical wave.

The correlation in phase between two points ρ_1 and ρ_2 on the wavefront degrades with the distance $\rho = |\rho_1 - \rho_2|$. For a plane or a spherical wave, the degree of coherence can be expressed as

$$\gamma(z, \rho) = \exp \left[- \left(\frac{\rho}{\rho_0} \right)^{5/3} \right] \quad (14)$$

where ρ_0 is the phase coherence radius and is given by

$$\rho_0 = (b_2 C_n^2 k^2 z^2)^{-3/5} \quad (15)$$

where $b_2 = 1.45$ (0.55) for a plane (spherical) wave. When $\rho \geq \rho_0$, the random phase angle difference is larger than π , and the wavefront is assumed to have lost its spatial coherence.

D. Angle of Arrival

Fluctuations in the angle of arrival of a signal at the receiver aperture are a consequence of random phase distortions due to turbulence. The phase difference Δs across a receiver aperture of diameter d can be approximated by

$$\Delta s \approx kd \sin \alpha \approx k d \alpha \quad (16)$$

where α is the random angle of arrival. The variance of the angle of arrival can be written as

$$E[\alpha^2] = \sigma_\alpha^2 = \frac{E[(\Delta s)^2]}{k^2 d^2} = \frac{D_s(0, d)}{k^2 d^2} \quad (17)$$

Using Eq. (13), the variance in angle of arrival can be computed for the weak turbulence case.

E. Aperture Averaging

The scintillation statistics discussed above are true for a point observer. If the receiver has a non-zero aperture diameter, the observed effect of scintillation will be the spatially averaged value of σ_{in}^2 over the entire collecting surface. The strength of received intensity fluctuations is found to decrease with increasing aperture size. This effect, known as aperture smoothing, has been observed experimentally [68], [69]. For weak turbulence, the smoothing effect continues to be pronounced until the diameter of the aperture becomes as large as the Fresnel zone, i.e., $(\lambda z)^{1/2}$, after which it saturates. In the case of strong turbulence, the critical diameter is on the order of the transverse spatial coherence parameter of the incoming beam, which is usually much smaller than the Fresnel zone.

F. Other Effects

Refractive turbulence may also produce depolarization of light and temporal stretching of optical pulses. Calculations by Strohbehn and Clifford [70] show that average power in the depolarized component is about 160 dB smaller than that of the incident beam. Attempts to measure depolarization with an accuracy of 45 dB have yielded a negative result [71]. Indeed,

most theoretical studies of clear-air turbulence neglect the polarization term in the wave equation to simplify calculations.

Light pulses arrive at the receiver with variable path delays as they travel through spatially different parts of the turbulent atmosphere [72]–[77]. Calculations show that temporal stretching of pulses is typically about 0.01 picosecond for the entire height of the atmosphere.

We find that the magnitude of both of these effects is negligibly small, and consequently their effects on optical communications can be ignored. It must be noted, however, that the effects can be much stronger when scattering due to turbid constituents of the atmosphere is considered. This aspect of the problem will be discussed in a later section.

G. Optical Communications in Turbulence

For communications from an exoatmospheric laser source in deep space to an earth-bound receiver (downlink), the Fresnel zone size of the beam in the atmosphere will be much larger than the scale size L of the inhomogeneities. The main effects of turbulence on the signal for this configuration will be beam spreading, scintillation, and loss of spatial coherence. For earth-space optical communications (uplinks), with the optical transmitter residing inside the atmosphere, the Fresnel zone size will be much smaller than L , making beam wander and fluctuations in the angle of arrival the principal factors contributing to signal degradation.

Scintillation produces both temporal and spatial intensity fluctuations at the receiving aperture, resulting in power surges and fades. The typical duration of scintillation-induced temporal fades is on the order of a few milliseconds [78]. The probability of fade events for given fade levels, as well as the duration of such fades, is described in [62] and [78]. It is shown there that fade values of 10 dB or larger are observed 12 percent of the time for worst-case turbulence (1 percent of the time for weak turbulence). A brute force approach may be used to overcome fades produced by temporal scintillation. Yura and McKinley [62] provide a worst-case scintillation fade analysis. With this approach, a link margin of 10 dB will be necessary for the system to work properly 99 percent of the time. However, the situation in reality is not this bad for ground receivers [79], [80]. The results given above refer to a point receiver, whereas actual receivers have a non-zero size. Aperture averaging reduces the probability of fades. Yura and McKinley [80] have obtained an engineering approximation for the magnitude of this effect on ground-based receivers. The factor A by which the irradiance variance is reduced due to aperture averaging is given by

$$A \approx \left[1 + 1.1 (D^2 / \lambda h_0 \sec \theta)^{7/6} \right]^{-1} \quad (18)$$

where D is the aperture diameter, θ is the zenith angle, λ is the laser wavelength, and $h_0 \sim 10$ km is the scale height of the atmosphere. For $\lambda = 1 \mu\text{m}$, $D = 1$ m, and a scale height of the atmosphere $h_0 = 10.3$ km, the aperture averaging factor A becomes 4.3×10^{-3} . It should be noted here that the advantage of aperture averaging is not available for uplink applications, as the phase coherence radius in this case is much larger than the probable receiving aperture on a spacecraft.

A number of phase compensation techniques to remove turbulence-induced tilt have been discussed in the literature [81]–[88]. A rigorous calculation of tilt correction on axial beam intensity is provided in [89]–[91]. Dunphy and Kerr [86] have obtained a simplified approximate expression for this effect. For path length in turbulence, $z \leq kd^2$, where d is the transmitter aperture diameter, they find

$$E[I] = \left(\frac{d_e}{16z} \right)^2 \left\{ (kd_e)^{-2} + \left[1 - C \left(\frac{d_e}{r_0} \right)^{-1/3} (kr_0)^{-2} \right] \right\}^{-1} \quad (19)$$

where the effective diameter $d_e = d$ for a uniformly illuminated circular beam cross-section, but for a Gaussian profile d_e is set equal to $2d$. $C = 0$ if there is no tilt correction, and $C = 1.18$ when tilt correction is included. Also, $r_0 = 2.098 \rho_0$ is Fried's spatial coherence diameter. Furthermore, the above result is true when $d_e/r_0 > 2$. The largest improvement in the received intensity occurs when $2 \leq d/r_0 \leq 5$. Whereas theoretical calculations predict an improvement of about 5 dB, measured data show that the improvement can be as high as 8 dB [86].

IV. Turbidity Effects on Beam Propagation

Atmospheric constituents in the form of gases and particulates absorb and scatter light. Thus, the design of successful optical systems has to contend with atmospheric turbidity and must account for the diminished direct-beam energy as light travels through the atmosphere.

A. Absorption and Scattering

The only notable effect of molecular absorption is to take away some of the energy from the laser beam [92]. The beam irradiance I as the light travels a distance Z through the atmosphere can be written as

$$I = I_0 \exp \left[- \int_0^Z \gamma_a(z) dz \right] \quad (20)$$

where I_0 is the irradiance at $z = 0$, and $\gamma_a(z)$ is the absorption coefficient at position z . The argument of the exponential in the above equation is known as optical depth or optical thickness.

Light is absorbed when the quantum state of a molecule, characterized by its electronic, vibrational, or rotational energy, is excited from a lower to a higher state. The absorption cross section has a Lorentzian shape [93] that peaks at the molecular transition energy. For each transition line, one needs to know the peak frequency, the width, and its total absorption cross section. The widths of these lines are typically on the order of 10^{-5} nm. However, Doppler and pressure broadening, which result from the thermal motion of molecules and molecular collisions, respectively, lead to much larger Gaussian-shaped absorption lines. Therefore, to obtain the total absorption coefficient at a particular frequency, one must calculate the line shape factor, including temperature and pressure effects as well.

There are a number of texts on spectroscopy and catalogs of line parameters. The High Resolution Transmittance (HITRAN) program, one of the most complete compilations of such data, was developed at the Air Force Geophysics Laboratory (AFGL). The compilation gives various line parameters for almost 350,000 lines over a spectral region from the ultraviolet to millimeter waves [94]–[100].

For typical optical calculations, however, the source and receiver bandwidths will be much larger than the resolution provided by the HITRAN database. Another database, called the Low Resolution Transmission (LOWTRAN) program, which is suitable for optical communication needs, has also been developed by AFGL. The LOWTRAN codes calculate molecular absorption from 0.25 to $28.5 \mu\text{m}$ [101]–[106]. They also calculate extinction due to molecular and aerosol scattering. These codes have been used extensively, and a number of comments on their use have been published [107]–[114]. A typical LOWTRAN calculated plot, shown in Fig. 2, is presented in [114] for space-to-ground light transmission under hazy conditions. Note that it includes the effects of molecular and aerosol scattering in addition to molecular absorption.

Since the size of air molecules is much smaller than optical wavelengths, the scattering of light by molecules falls into the Rayleigh regime. The main effect on a beam of light, as in the case of molecular absorption, is extinction of the beam. A relation similar to Eq. (20) above can be defined by replacing the absorption coefficient by the Rayleigh scattering coefficient, $\gamma_{sr}(z)$. The argument of the exponential now gives the optical depth of the atmospheric path due to molecular scattering. The scattering coefficient for a gas with refractive index n is given by [115]

$$\gamma_{sr} = \frac{8\pi^3 (n^2 - 1)^2}{3N_m \lambda^4} \frac{6 + 3\delta}{6 - 7\delta} \quad (21)$$

where N_m is the number of molecules of the gas per unit volume, λ is the optical wavelength, and δ is the depolarization factor of the scattered radiation. According to recent measurements, $\delta = 0.035$ [116]. By adding contributions from the constituent gases in the air, the total molecular scattering coefficient for a given atmospheric profile can be computed. Tabulated values for various vertical paths are provided by Elterman [117]–[118]. As mentioned earlier, molecular scattering effects have been included in the LOWTRAN computer codes.

Scattering by aerosols, light fog and haze, and thin clouds generally falls into the Mie category. The size of the scatterers in this case is between 0.01 to $10.0 \mu\text{m}$, comparable to the optical wavelengths under consideration. A relation analogous to Eq. (20) defines the Mie scattering extinction coefficient, γ_{sm} , and the relevant optical depth. The calculation of the Mie scattering coefficient is not a simple task. The size, concentration, and shape of Mie particles in the atmosphere are not well defined and vary with time and height. A good sampling of numerous techniques used to determine Mie scattering coefficients is given in [119]–[155]. Tables for the scattering coefficient and the angular scattering function for various atmospheric particles are given in [156]–[164]. Some typical profiles of this type of scattering are also included in the LOWTRAN computer codes.

B. Opaque Clouds

As a rule of thumb, if the disk of the sun or moon can be seen, the clouds are considered thin, and their effect on light beams can be adequately explained in terms of Mie scattering as discussed above. Opaque clouds are a different matter altogether. Vertical attenuation of over 100 dB has been observed for cumulus clouds [165]. Calculated extinctions of over 1000 dB for realistic dense fog or clouds in the atmosphere are possible. The only viable strategy for optical system designers is to avoid such severe atmospheric conditions by employing spatial and temporal diversity.

C. Cloud Cover Studies

Satellite monitoring of the skies to build up a reasonable database to draw appropriate statistical conclusions about the suitability of a site seems promising, but for the present, imaging and direct visual observations dominate. The length of time for which GOES, GMS, and NOAA satellites have collected data is no more than a few years (1983–present). This duration is too short to obtain reliable, long-term cloud cover statistics.

Single-object monitoring can be very precise but must include many stars over the sky to be fully relevant. So far this method has not been used extensively. The expense of setting up such stations indiscriminately would alone be prohibitive.

A number of databases giving coarse information on cloud cover and visibility exist. Surface Airways Observations is one such database and is available from National Climatic Data Center in Asheville, North Carolina [166]. It provides information on cloud cover, visibility, and other parameters for over 1000 sites in the United States with a temporal resolution of 3 hours for the last 40 years. Data of this type are being used by the Air Force Geophysics Laboratory (AFGL) and other institutions to develop cloud cover models and computer programs with which to understand and design optical communication systems that would work under ambient weather conditions.

AFGL is also helping to set up whole sky imaging (WSI) stations. The WSI cameras use a fisheye lens to image the sky dome with a resolution of 512×512 pixels. There will be six such stations initially dispersed over the continental United States. This type of database can be quite useful for the work at hand but as yet is unavailable.

Almost all data and statistics currently available on cloud cover are not readily amenable to the study of optical propagation through the atmosphere. However, the available weather data may be used as a guide to develop computer models for the simulation of real-time dynamic cloud behavior. An early model for cloud cover was developed by scientists at SRI International. Work at AFGL, which is based on the SRI model, has produced considerably sophisticated computer programs suitable for modeling light propagation through the atmosphere. These models may be used to compute cloud-free line of sight (CFLOS) or cloud-free arc (CFARC) probabilities for any site. It is also possible to compute joint CFLOS and CFARC probabilities for two or more sites. These statistics, needless to say, are of great importance to the development of an optical space network (OSN).

D. Optical Communications in a Turbid Atmosphere

In the absence of opaque clouds, the only significant effect of the atmospheric scattering and absorption is described by Bouguer's law:

$$I = I_0 \exp \left[- \int_0^Z \gamma_t(z) dz \right] \quad (22)$$

where $\gamma_t(z)$ is the sum of all absorption and scattering coefficients due to gas molecules, aerosols, and other particulate

matter at position z , and Z is the propagation distance through the atmosphere. The law assumes that the extinction loss of beam power is independent of beam intensity, and that the absorbing and scattering events occur independently. The magnitude of the exponent in Eq. (22) is defined as the optical depth or thickness, τ , of the atmosphere, i.e.,

$$\tau = \int_0^Z \gamma_t(z) dz \quad (23)$$

Experiments with artificial fog and smoke and with diluted milk solution [9] show that the law holds well for optical thickness $\tau \leq 12$.

It is possible to describe the attenuation coefficient over horizontal paths in terms of meteorological range, R_v , commonly called "visibility" [167] - [169]. The approximate relationship is given by

$$\gamma_t(\nu, z) = \gamma_t = \frac{3.912}{R_v} \quad (24)$$

where we have disregarded the dependence of the attenuation coefficient on laser frequency and position, and R_v is measured in kilometers. Further approximate results may then be used to obtain the optical depth of the entire atmosphere along vertical paths. For example, it may be assumed that the scale height of the atmosphere is about 10 to 20 km along near-zenith paths; this assumption may then be used to obtain total attenuation loss. It may be necessary to devise direct measurement methods to obtain more accurate determinations of optical depths along vertical paths.

The presence of thick clouds, in general, will have a catastrophic effect on the availability of an optical communication link. Although scattered laser light will be available for communication, the system has to be designed to have (1) a wide field of view to collect enough power, which greatly increases the background noise; and (2) a low data rate to avoid intersymbol interference due to pulse spreading. Also, polarization coding of the signal cannot be used as the scattered light is depolarized. An optical communication system designed to employ the scattered beam through thick clouds, then quickly loses its advantages over conventional radio frequency systems.

The only reasonable strategy is to develop the OSN such that it avoids opaque clouds by employing diversity techniques. It will be necessary to identify sites for the installation of optical receiver and transmitter stations where the clouds have a low probability of occurrence. Several such sites with uncorrelated weather patterns may need to be operated simultaneously to obtain desired link availability.

Some of the possible configurations of an OSN which uses spatial diversity to get around the problem of opaque clouds is discussed in JPL IOM 331.6-88-491.¹ In [170], a first-order theoretical weather model is given for the estimation of the link budget for extinction loss through the atmosphere.

V. Conclusion

Various aspects of the light propagation problem through the atmosphere have been discussed in this article. Loss of beam energy due to absorption and scattering, degradation of beam quality due to scintillation and reduced spatial coherence from refractive turbulence, and link unavailability due to opaque clouds are some of the factors that cannot be overlooked while designing an optical communication system.

¹K. Shaik, interoffice memorandum to J. Lesh, IOM 331.6-88-491 (internal document), Jet Propulsion Laboratory, Pasadena, California, April 22, 1988.

A number of optical communication systems and techniques have been investigated to demonstrate optical communication links over horizontal as well as vertical paths through the lower atmosphere [171]–[178]. These include the Airborne Flight Test System (AFTS) developed by McDonnell Douglas, demonstrating a 1000-Mbit/s laser communication air-to-ground system [179]; a Laser Airborne Communications Experiment demonstrating a 50-km air-to-air and air-to-ground optical communication link using a 100-mW laser operating at 0.904 μm developed by GTE [180]; and SLCAIR demonstrations for submarine laser communications [181]. Detailed reviews of optical communications techniques and design procedures and considerations for atmospheric links can be found in [182]–[184].

Considerable information regarding the atmosphere and its effects on the propagation of light is available. To this end, an extensive reference list has been compiled. The work on optical communications through the atmosphere, however, is not yet complete. Further experiments and statistical studies will be necessary before reliable estimates on link availability and link budgets for optical communications through the atmosphere can be predicted with confidence.

References

- [1] *U.S. Standard Atmosphere Supplements*, Washington, D.C.: U.S. Government Printing Office, 1966.
- [2] S. H. Melfi, "Remote Sensing for Air Quality Management," in *Topics in Applied Physics*, vol. 14, New York: Springer-Verlag, 1976.
- [3] B. Garfinkel, "Astronomical Refraction in a Polytrropic Atmosphere," *Astron. J.*, vol. 72, no. 2, pp. 235–254, 1967.
- [4] *U.S. Standard Atmosphere*, Washington, D.C.: U.S. Government Printing Office, 1962.
- [5] R. S. Lawrence and J. W. Strohbehn, "A Survey of Clear-Air Propagation Effects Relevant to Optical Communications," *Proc. IEEE*, vol. 58, no. 10, pp. 1523–1545, 1970.
- [6] J. C. Owens, "Optical Refractive Index of Air: Dependence on Pressure, Temperature, and Composition," *Appl. Opt.*, vol. 6, pp. 51–60, 1967.
- [7] M. L. Wesley and E. C. Alcaraz, "Diurnal Cycles of the Refractive Index," *J. Geophys. Res.*, vol. 78, pp. 6224–6232, 1973.
- [8] C. A. Friehe et al., "Effects of Temperature and Humidity Fluctuations on the Optical Refractive Index in the Marine Boundary Layer," *J. Opt. Soc. Am.*, vol. 65, pp. 1502–1511, 1976.

- [9] V. E. Zuev, *Laser Beams in the Atmosphere*, New York: Consultants' Bureau, 1982.
- [10] P. Beckmann, "Signal Degeneration in Laser Beams Propagated Through a Turbulent Atmosphere," *J. Res. Natl. Bur. Stand.*, sec. D, vol. 69, p. 629, 1965.
- [11] H. Hodara, "Laser Wave Propagation Through the Atmosphere," *Proc. IEEE*, vol. 54, no. 3, pp. 368-375, 1966.
- [12] W. G. Lese, Jr., "Stellar Image Excursion Caused by Random Atmospheric Refraction," Memorandum Rep. 2014, Aberdeen Proving Ground, Maryland: Ballistic Research Laboratories, 1969.
- [13] V. I. Tatarski, *The Effects of the Turbulent Atmosphere on Wave Propagation*, Report No. TT-68-50464, Springfield, Virginia: National Technical Information Service, 1971.
- [14] L. A. Chernov, *Wave Propagation in Random Media*, New York: McGraw-Hill, 1960.
- [15] A. Ishimaru, *Wave Propagation and Scattering in Random Media*, vols. I and II, New York: Academic Press, 1978.
- [16] J. W. Strohbehn (ed.), *Laser Beam Propagation in the Atmosphere*, New York: Springer-Verlag, 1978.
- [17] R. E. Hufnagel, "Variations of Atmospheric Turbulence," in *Digest of the Topical Meeting on Optical Propagation Through Turbulence*, Washington, D.C.: Optical Society of America, 1974.
- [18] R. S. Lawrence et al., "Measurements of Atmospheric Turbulence Relevant to Optical Propagation," *J. Opt. Soc. Am.*, vol. 60, pp. 826-830, 1970.
- [19] G. R. Ochs and R. S. Lawrence, "Temperature and C_N^2 Profiles Measured Over Land and Ocean to 3 km Above the Surface," NOAA Report ERL 251-WPL 22, Boulder, Colorado: National Oceanic and Atmospheric Administration, 1972.
- [20] J. L. Bufton et al., "Measurements of Turbulence Profiles in the Troposphere," *J. Opt. Soc. Am.*, vol. 62, pp. 1068-1070, 1972.
- [21] J. L. Bufton, "Comparison of Vertical Profile Turbulence Structure With Stellar Observations," *Appl. Opt.*, vol. 12, pp. 1785-1793, 1973.
- [22] J. L. Bufton, "An Investigation of Atmospheric Turbulence by Stellar Observations," NASA TR R-369, Greenbelt, Maryland: Goddard Space Flight Center, 1971.
- [23] R. Barletti et al., "Optical Remote Sensing of Atmospheric Turbulence," *J. Opt. Soc. Am.*, vol. 66, p. 1380, 1976.
- [24] J. Vernin et al., "Optical Remote Sensing of Atmospheric Turbulence: A Comparison with Simultaneous Thermal Measurements," *Appl. Opt.*, vol. 18, pp. 243-247, 1979.
- [25] Y. Barabanenkov et al., "Status of the Theory of Propagation of Waves in Randomly Inhomogeneous Media," *Sov. Phys. Usp.*, vol. 13, pp. 551-580, 1971.
- [26] W. Brown, "Moment Equations for Waves Propagated in Random Media," *J. Opt. Soc. Am.*, vol. 62, pp. 45-54, January 1972.
- [27] W. Brown, "Fourth Moment of a Wave Propagating in a Random Medium," *J. Opt. Soc. Am.*, vol. 62, pp. 966-971, June 1972.

- [28] D. deWolf, "Strong Irradiance Fluctuations in Turbulent Air, Part I: Plane Waves," *J. Opt. Soc. Am.*, vol. 63, pp. 171-179, February 1973.
- [29] D. deWolf, "Strong Irradiance Fluctuations in Turbulent Air, Part II: Spherical Waves," *J. Opt. Soc. Am.*, vol. 63, pp. 1249-1253, October 1973.
- [30] D. deWolf, "Strong Irradiance Fluctuations in Turbulent Air, Part III: Diffraction Cutoff," *J. Opt. Soc. Am.*, vol. 64, pp. 360-365, March 1974.
- [31] M. Sancer and A. Varvatsis, "Saturation Calculation for Light Propagation in the Turbulent Atmosphere," *J. Opt. Soc. Am.*, vol. 60, pp. 654-659, May 1970.
- [32] M. Beran, "Propagation of a Finite Beam in a Random Medium," *J. Opt. Soc. Am.*, vol. 60, pp. 518-521, April 1970.
- [33] M. Beran and T. Ho, "Propagation of the Fourth-Order Coherence Function in a Random Medium," *J. Opt. Soc. Am.*, vol. 59, pp. 1134-1138, September 1969.
- [34] J. Molyneux, "Propagation of Nth Order Coherence Functions in a Random Medium," *J. Opt. Soc. Am.*, vol. 61, pp. 369-377, March 1971.
- [35] V. Tatarski, "Light Propagation in a Medium With Random Refractive Index Inhomogeneities in the Markov Approximation," *Sov. Phys.-JETP*, vol. 29, pp. 1133-1138, December 1969.
- [36] V. Klyatskin, "Applicability of the Approximation of a Markov Random Process in Problems Related to the Propagation of Light in a Medium With Random Inhomogeneities," *Sov. Phys.-JETP*, vol. 30, pp. 520-523, March 1970.
- [37] V. Klyatskin and V. Tatarski, "The Parabolic Equation Approximation for Propagation of Waves in a Medium With Random Inhomogeneities," *Sov. Phys.-JETP*, vol. 31, pp. 335-339, August 1970.
- [38] R. Lutomirski and H. Yura, "Propagation of a Finite Optical Beam in an Inhomogeneous Medium," *Appl. Opt.*, vol. 10, pp. 1652-1658, July 1971.
- [39] K. Furutsu, "Statistical Theory of Wave Propagation in a Random Medium and the Irradiance Distribution Function," *J. Opt. Soc. Am.*, vol. 62, pp. 240-254, February 1972.
- [40] W. P. Brown, Jr., "Validity of the Rytov Approximation in Optical Calculations," *J. Opt. Soc. Am.*, vol. 56, pp. 1045-1052, 1966.
- [41] D. A. deWolf, "Saturation of Irradiance Fluctuations due to Turbulent Atmosphere," *J. Opt. Soc. Am.*, vol. 58, pp. 461-466, 1968.
- [42] J. W. Strohbehn, "Line of Sight Wave Propagation Through the Turbulent Atmosphere," *Proc. IEEE*, vol. 56, pp. 1301-1318, 1968.
- [43] A. Prokhorov, F. Bunkin, K. Gochelashvili, and V. Shishov, "Laser Irradiance Propagation in Turbulent Media," *Proc. IEEE*, vol. 63, pp. 790-811, May 1975.
- [44] S. Clifford, G. Ochs, and R. Lawrence, "Saturation of Optical Scintillations by Strong Turbulence," *J. Opt. Soc. Am.*, vol. 64, pp. 148-154, February 1974.
- [45] H. Yura, "Physical Model for Strong Optical-Amplitude Fluctuations in a Turbulent Medium," *J. Opt. Soc. Am.*, vol. 64, pp. 59-67, January 1974.
- [46] D. deWolf, "Waves in Turbulent Air: A Phenomenological Model," *Proc. IEEE*, vol. 62, pp. 1523-1529, November 1974.
- [47] K. Gochelashvili, "Saturation of the Fluctuations of Focused Radiation in a Turbulent Medium," *Radiophys. Quant. Electron.*, vol. 14, pp. 470-473, April 1971.

- [48] I. Dagkesamanskaya and V. Shishov, "Strong Intensity Fluctuations During Wave Propagation in Statistically Homogeneous and Isotropic Media," *Radiophys. Quant. Electron.*, vol. 13, pp. 9-12, January 1970.
- [49] V. Shishov, "Strong Fluctuations of the Intensity of a Plane Wave Propagating in a Random Refractive Medium," *Sov. Phys.-JETP*, vol. 34, pp. 744-748, April 1972.
- [50] V. Shishov, "Strong Fluctuations of the Intensity of a Spherical Wave Propagating in a Randomly Refractive Medium," *Radiophys. Quant. Electron.*, vol. 15, pp. 689-695, June 1972.
- [51] R. Fante, "Some New Results on Propagation of Electromagnetic Waves in Strongly Turbulent Media," *IEEE Trans. Ant. Propag.*, vol. AP-23, pp. 382-385, May 1975.
- [52] R. Fante, "Electric Field Spectrum and Intensity Covariance of a Wave in a Random Medium," *Radio Sci.*, vol. 10, pp. 77-85, January 1975.
- [53] K. Gochelashvili and V. Shishov, "Saturated Intensity Fluctuations of Laser Radiation in a Turbulent Medium" (in Russian), *Zh. Eksp. Teor. Fiz.*, vol. 66, pp. 1237-1247, March 1974.
- [54] R. Fante, "Irradiance Scintillations: Comparison of Theory With Experiment," *J. Opt. Soc. Am.*, vol. 65, pp. 548-550, May 1975.
- [55] K. Gochelashvili and V. Shishov, "Laser Beam Scintillation Beyond a Turbulent Layer," *Opt. Acta*, vol. 18, pp. 313-320, April 1971.
- [56] R. Fante, "Some Results for the Variance of the Irradiance of a Finite Beam in a Random Medium," *J. Opt. Soc. Am.*, vol. 65, pp. 608-610, May 1975.
- [57] K. Gochelashvili and V. Shishov, "Focused Irradiance Fluctuations Beyond a Layer of Turbulent Atmosphere," *Opt. Acta*, vol. 19, pp. 327-332, April 1972.
- [58] V. Banakh, G. Krekov, V. Mironov, S. Khmelevtsov, and S. Tsvik, "Focused-Laser-Beam Scintillations in the Turbulent Atmosphere," *J. Opt. Soc. Am.*, vol. 64, pp. 516-518, April 1974.
- [59] R. L. Fante, "Electromagnetic Beam Propagation in Turbulent Media," *Proc. IEEE*, vol. 63, pp. 1669-1690, 1975.
- [60] V. A. Banakh and V. L. Mironov, "Phase Approximation of the Huygens-Kirchhoff Method in Problems of Laser-Beam Propagation in the Turbulent Atmosphere," *Opt. Lett.*, vol. 1, pp. 172-174, 1977.
- [61] R. L. Fante, "Electromagnetic Beam Propagation in Turbulent Media: An Update," *Proc. IEEE*, vol. 68, pp. 1424-1443, 1980.
- [62] H. T. Yura and W. G. McKinley, "Optical Scintillation Statistics for IR Ground-to-Space Laser Communication Systems," *Appl. Opt.*, vol. 22, no. 21, pp. 3353-3358, 1983.
- [63] A. Ishimaru, "Theory of Optical Propagation in the Atmosphere," *Opt. Eng.*, vol. 20, pp. 63-70, 1981.
- [64] V. E. Zuev, *Laser Beams in the Atmosphere*, Chapter 4, New York: Consultants' Bureau, 1982.
- [65] A. I. Kon and V. I. Tatarski, "Parameter Fluctuations in a Space-Limited Light Beam in a Turbulent Medium," *Izv. Vyssh. Uchebn. Zaved., Radiofiz.*, vol. 8, pp. 870-875, 1965.

- [66] R. A. Schmeltzer, "Means, Variances, and Covariances for Laser Beam Propagation Through a Random Medium," *Q. Appl. Math.*, vol. 24, pp. 339-354, 1967.
- [67] A. Ishimaru, "Fluctuations of a Beam Wave Propagating Through a Locally Homogeneous Medium," *Radio Sci.*, vol. 4, pp. 295-305, 1969.
- [68] G. Homstad, J. Strohbehn, R. Berger, and J. Heneghan, "Aperture-Averaging Effects for Weak Scintillations," *J. Opt. Soc. Am.*, vol. 64, pp. 162-165, February 1974.
- [69] J. Kerr and J. Dunphy, "Experimental Effects of Finite Transmitter Apertures on Scintillations," *J. Opt. Soc. Am.*, vol. 63, pp. 1-7, January 1973.
- [70] J. W. Strohbehn and S. F. Clifford, "Polarization and Angle-of-Arrival Fluctuations for a Plane Wave Propagated Through a Turbulent Medium," *IEEE Trans. Antennas Propag.*, vol. AP-15, pp. 416-421, 1967.
- [71] A. A. M. Saleh, "An Investigation of Laser Wave Depolarization due to Atmospheric Transmission," *IEEE J. Quant. Elect.*, vol. QE-3, pp. 540-543, 1967.
- [72] I. Sreenivasiah and A. Ishimaru, "Plane Wave Pulse Propagation Through Atmospheric Turbulence at Millimeter and Optical Wavelengths," AFCRL-TR-74-0205, Department of Electrical Engineering, University of Washington, Seattle, March 1974.
- [73] C. Liu, A. Wernik, and K. Yeh, "Propagation of Pulse Trains Through a Random Medium," *IEEE Trans. Antennas Propag.*, vol. AP-22, pp. 624-627, July 1974.
- [74] H. Su and M. Plonus, "Optical-Pulse Propagation in a Turbulent Medium," *J. Opt. Soc. Am.*, vol. 61, pp. 256-260, March 1971.
- [75] M. Plonus, H. Su, and C. Gardner, "Correlation and Structure Functions for Pulse Propagation in a Turbulent Atmosphere," *IEEE Trans. Antennas Propag.*, vol. AP-20, pp. 801-805, November 1972.
- [76] C. Gardner and M. Plonus, "Optical Pulses in Atmospheric Turbulence," *J. Opt. Soc. Am.*, vol. 64, pp. 68-77, January 1974.
- [77] A. Ishimaru, "Temporal Frequency Spectra of Multi-Frequency Waves in a Turbulent Atmosphere," *IEEE Trans. Ant. Propag.*, vol. AP-20, pp. 10-19, January 1972.
- [78] *The Technical Cooperation Program*, Volume V of the Laser Communications Workshop, Washington, D.C.: U.S. Government Printing Office, 1985.
- [79] D. L. Fried, "Aperture Averaging of Scintillation," *J. Opt. Soc. Am.*, vol. 57, pp. 169-174, 1967.
- [80] H. T. Yura and W. G. McKinley, "Aperture Averaging of Scintillation for Space-to-Ground Optical Communication Applications," *Appl. Opt.*, vol. 22, pp. 1608-1609, 1983.
- [81] J. Hardy, "Adaptive Optics: A New Technology for Control of Light," *Proc. IEEE*, vol. 66, pp. 651-697, 1978.
- [82] J. Wang and J. Markey, "Model Compensation of Atmospheric Turbulence Phase Distortion," *J. Opt. Soc. Am.*, vol. 63, pp. 78-87, 1978.
- [83] F. Gebhardt, "Experimental Demonstration of the Use of Phase Correction to Reduce Thermal Blooming," *Opt. Acta*, vol. 26, pp. 615-625, 1979.

- [84] C. Primmerman and D. Fouche, "Thermal Blooming Compensation: Experimental Observation of a Deformable Mirror System," *Appl. Opt.*, vol. 15, pp. 990-995, 1976.
- [85] D. Smith, "High Power Laser Propagation: Thermal Blooming," *Proc. IEEE*, vol. 65, pp. 1679-1714, 1977.
- [86] J. Dunphy and J. R. Kerr, "Turbulence Effects on Target Illumination by Laser Sources: Phenomenological Analysis and Experimental Results," *Appl. Opt.*, vol. 16, pp. 1345-1358, 1977.
- [87] J. Pearson, "Atmospheric Turbulence Compensation Using Coherent Optical Adaptive Techniques," *Appl. Opt.*, vol. 15, pp. 622-631, 1976.
- [88] M. Bertolotti et al., "Optical Propagation: Problems and Trends," *Opt. Acta*, vol. 26, pp. 507-529, 1979.
- [89] R. Lutomirski, W. Woodie, and R. Buser, "Turbulence-Degraded Beam Quality: Improvement Obtained With Tilt Correcting Aperture," *Appl. Opt.*, vol. 16, pp. 665-673, 1977.
- [90] M. Tavis and H. Yura, "Short Term Average Irradiance Profile of an Optical Beam in a Turbulent Medium," *Appl. Opt.*, vol. 15, pp. 2922-2931, 1976.
- [91] G. Valley, "Isoplanatic Degradation of Tilt Correction and Short Term Imaging Systems," *Appl. Opt.*, vol. 19, pp. 574-577, 1980.
- [92] E. J. McCartney, *Absorption and Emission by Atmospheric Cases*, New York: Wiley, 1983.
- [93] R. H. Pantell and H. E. Puthoff, *Fundamentals of Quantum Electronics*, New York: Wiley, 1969.
- [94] R. A. McClatchey, W. S. Benedict, S. A. Clough, D. E. Burch, R. F. Calfee, K. Fox, L. S. Rothman, and J. S. Garing, "AFCRL Atmospheric Absorption Line Parameters Compilation," AFCRL-TR-0096, Air Force Research Laboratory, Cambridge, Massachusetts, 1973.
- [95] L. S. Rothman, S. A. Clough, R. A. McClatchey, L. G. Young, D. E. Snider, and A. Goldman, "AFGL Trace Gas Compilation," *Appl. Opt.*, vol. 17, p. 507, 1978.
- [96] L. S. Rothman, "Update of the AFGL Atmospheric Absorption Line Parameters Compilation," *Appl. Opt.*, vol. 17, pp. 3517-3518, 1978.
- [97] L. S. Rothman, "AFGL Atmospheric Absorption Line Parameters Compilation: 1980 Version," *Appl. Opt.*, vol. 20, pp. 791-795, 1981.
- [98] L. S. Rothman, A. Goldman, J. R. Gillis, R. H. Tipping, L. R. Brown, J. S. Margolis, A. G. Maki, and L. D. G. Young, "AFGL Trace Gas Compilation: 1980 Version," *Appl. Opt.*, vol. 20, pp. 1323-1328, 1981.
- [99] L. S. Rothman, A. Goldman, J. R. Gillis, R. R. Gamache, H. M. Pichett, R. L. Poynter, N. Huson, and A. Chedin, "AFGL Trace Gas Compilation: 1982 Version," *Appl. Opt.*, vol. 22, pp. 1616-1627, 1983.
- [100] L. S. Rothman, R. R. Gamache, A. Barbe, A. Goldman, J. R. Gillis, L. R. Brown, R. A. Toth, J.-M. Flaud, and C. Camy-Peyret, "AFGL Atmospheric Absorption Line Parameters Calculation: 1982 Edition," *Appl. Opt.*, vol. 22, pp. 2247-2256, 1983.

- [101] J. E. A. Selby and R. A. McClatchey, "Atmospheric Transmittance From 0.25 to 28.5 μm : Computer Code LOWTRAN2," AFCRL-TR-0745, AD763-721, Air Force Research Laboratory, Cambridge, Massachusetts, 1972.
- [102] J. E. A. Selby and R. A. McClatchey, "Atmospheric Transmittance From 0.25 to 28.5 μm : Computer Code LOWTRAN3, AFCRL-TR-75-0255, ADA017-734, Air Force Research Laboratory, Cambridge, Massachusetts, 1975.
- [103] J. E. A. Selby, E. P. Shettle, and R. A. McClatchey, "Atmospheric Transmittance From 0.25 to 28.5 μm : Supplement LOWTRAN3B," AFGL-TR-76-0258, ADA040-701, Air Force Geophysics Laboratory, Hanscom Air Force Base, Massachusetts, 1976.
- [104] J. E. A. Selby, F. X. Kneizys, J. H. Chetwynd, and R. A. McClatchey, "Atmospheric Transmittance/Radiance: Computer Code LOWTRAN4," AFGL-TR-78-0053, ADA058, Air Force Geophysics Laboratory, Hanscom Air Force Base, Massachusetts, 1978.
- [105] F. X. Kneizys, E. P. Shettle, W. O. Gallery, J. H. Chetwynd, L. W. Abrew, J. E. A. Selby, R. W. Fenn, and R. A. McClatchey, "Atmospheric Transmittance/Radiance: Computer Code LOWTRAN5," AFGL-TR-80-067, Air Force Geophysics Laboratory, Hanscom Air Force Base, Massachusetts, 1980.
- [106] F. X. Kneizys et al., "Atmospheric Transmittance/Radiance: Computer Code LOWTRAN6," AFGL-TR-83-0187, Air Force Geophysics Laboratory, Hanscom Air Force Base, Massachusetts, 1983.
- [107] R. R. Gruenzel, "Mathematical Expressions for Molecular Absorption in LOWTRAN3B," *Appl. Opt.*, vol. 17, pp. 2591-2593, 1978.
- [108] J. H. Pierluissi, K. Tomiyama, and R. B. Gomez, "Analysis of the LOWTRAN Transmission Functions," *Appl. Opt.*, vol. 18, pp. 1607-1612, 1978.
- [109] W. M. Cornette, "Suggested Modification to the Total Volume Molecular Scattering Coefficient in LOWTRAN," *Appl. Opt.*, vol. 19, p. A182, 1980.
- [110] E. P. Shettle, F. X. Kneizys, and W. O. Gallery, "Suggested Modification to the Total Volume Molecular Scattering Coefficient in LOWTRAN: Comment," *Appl. Opt.*, vol. 19, pp. 2873-2874, 1980.
- [111] D. C. Robertson, L. S. Bernstein, R. Haimes, J. Wunderlich, and L. Vega, "5-cm⁻¹ Band Model Option to LOWTRAN5," *Appl. Opt.*, vol. 20, pp. 3218-3228, 1981.
- [112] J. H. Pierluissi and G. Peng, "New Molecular Transmission Band Models for LOWTRAN," *Opt. Eng.*, vol. 24, pp. 541-547, 1985.
- [113] J. H. Pierluissi and C.-M. Tsai, "New LOWTRAN Models for the Uniformly Mixed Gases," *Appl. Opt.*, vol. 26, pp. 616-618, 1987.
- [114] D. E. Novoseller, "Use of LOWTRAN in Transmission Calculations," *Appl. Opt.*, vol. 26, pp. 3185-3186, 1987.
- [115] E. J. McCartney, *Optics of the Atmosphere*, New York: Wiley, 1976.
- [116] L. Elterman, *Atmospheric Attenuation Model in the Ultraviolet, Visible, and Infrared Regions for Altitudes to 50 km*, AFCRL-64-740, ERPN 46, Air Force Research Laboratory, Cambridge, Massachusetts, 1964.
- [117] L. Elterman, "UV, Visible and IR Attenuation for Altitudes to 50 km," AFCRL-TR-68-0153, Air Force Research Laboratory, Cambridge, Massachusetts, 1968.

- [118] L. Elterman, "Vertical Attenuation Model With Eight Surface-Meteorological Ranges 2 to 13 km," AFCRL-TR-70-0200, Air Force Research Laboratory, Cambridge, Massachusetts, 1970.
- [119] D. Deirmendjian, "Scattering and Polarization Properties of Polydispersed Suspensions With Partial Absorption," in *Electromagnetic Scattering*, M. Kerker (ed.), New York: Macmillan, 1963.
- [120] D. Deirmendjian, "Scattering and Polarization Properties of Water Clouds and Hazes in the Visible and Near Infrared," *Appl. Opt.*, vol. 3, pp. 187-193, 1964.
- [121] D. Deirmendjian, *Electromagnetic Scattering on Spherical Polydispersions*, New York: Elsevier, 1969.
- [122] C. E. Junge, "Atmospheric Chemistry," in *Advances in Geophysics*, Volume 4, New York: Academic Press, 1958.
- [123] C. E. Junge, "Aerosols," in *Handbook of Geophysics*, C. F. Campen, et al. (eds.), New York: Macmillan, 1960.
- [124] C. E. Junge, *Chemistry and Radioactivity*, New York: Academic Press, 1963.
- [125] R. E. Pasceri and S. K. Friedlander, "Measurements of the Particle Size Distribution of the Atmospheric Aerosol. II: Experimental Results and Discussion," *J. Atmos. Sci.*, vol. 22, pp. 577-584, 1965.
- [126] W. E. Clark and K. T. Whitby, "Concentration and Size Distribution Measurements of Atmospheric Aerosols and a Test of the Theory of Self-Preserving Size Distributions," *J. Atmos. Sci.*, vol. 24, pp. 677-687, 1967.
- [127] R. F. Pueschel and K. E. Noll, "Visibility and Aerosol Size-Frequency Distribution," *J. Appl. Meteorol.*, vol. 6, pp. 1045-1052, 1967.
- [128] I. H. Blifford, Jr., and L. D. Ringer, "The Size and Number Distribution of Aerosols in the Continental Troposphere," *J. Atmos. Sci.*, vol. 26, pp. 716-726, 1969.
- [129] K. T. Whitby et al., "The Minnesota Aerosol Analyzing System Used in the Los Angeles Smog Project," *J. Colloid Interface Sci.*, vol. 39, pp. 136-164, 1972.
- [130] K. T. Whitby et al., "The Aerosol Size Distribution of Los Angeles Smog," *J. Colloid Interface Sci.*, vol. 39, pp. 177-204, 1972.
- [131] I. H. Blifford, Jr., "Tropospheric Aerosols," *J. Geophys. Res.*, vol. 35, pp. 3099-3103, 1970.
- [132] D. A. Gillette and I. H. Blifford, Jr., "Composition of Tropospheric Aerosols as a Function of Altitude," *J. Atmos. Sci.*, vol. 38, pp. 1199-1210, 1971.
- [133] D. J. Hofmann et al., "Stratospheric Aerosol Measurements. I: Time Variations at Northern Midlatitudes," *J. Atmos. Sci.*, vol. 32, pp. 1446-1456, 1975.
- [134] J. M. Rosen, "Simultaneous Dust and Ozone Soundings Over North and Central America," *J. Geophys. Res.*, vol. 73, pp. 479-486, 1968.
- [135] J. M. Rosen et al., "Stratospheric Aerosol Measurements. II: The Worldwide Distribution," *J. Atmos. Sci.*, vol. 32, pp. 1457-1462, 1975.
- [136] R. G. Pinnick et al., "Stratospheric Aerosol Measurements. III: Optical Model Calculations," *J. Atmos. Sci.*, vol. 33, pp. 304-314, 1976.
- [137] E. O. Hulbert, "Observations of a Searchlight Beam to an Altitude of 28 Kilometers," *J. Opt. Soc. Am.*, vol. 27, pp. 377-382, 1937.

- [138] E. O. Hulbert, "Optics of Searchlight Illumination," *J. Opt. Soc. Am.*, vol. 36, pp. 483-491, 1946.
- [139] E. A. Johnson et al., "The Measurement of Light Scattered by the Upper Atmosphere From a Search-Light Beam," *J. Opt. Soc. Am.*, vol. 29, pp. 512-517, 1939.
- [140] L. Elterman, "Aerosol Measurements in the Troposphere and Stratosphere," *Appl. Opt.*, vol. 5, pp. 1769-1776, 1966.
- [141] G. G. Goyer and R. D. Watson, "Laser Techniques for Observing the Upper Atmosphere," *Bull. Am. Meteorol. Soc.*, vol. 49, pp. 890-895, 1968.
- [142] R. T. Collis, "Lidar," *Appl. Opt.*, vol. 9, pp. 1782-1788, 1970.
- [143] W. E. Evans and R. T. Collis, "Meteorological Application of Lidar," *Soc. Photogr. Inst. Eng.*, vol. 8, pp. 38-45, 1970.
- [144] G. B. Northam et al., "Dustsonde and Lidar Measurements of Stratospheric Aerosols: A Comparison," *Appl. Opt.*, vol. 13, pp. 2416-2421, 1974.
- [145] A. Cohen and M. Graber, "Laser-Radar Polarization Measurements of the Lower Stratospheric Aerosol Layer Over Jerusalem," *J. Appl. Meteorol.*, vol. 14, pp. 400-406, 1975.
- [146] M. J. Post, F. F. Hall, R. A. Richter, and T. R. Lawrence, "Aerosol Backscattering Profiles at $\lambda = 10.6 \mu\text{m}$," *Appl. Opt.*, vol. 21, pp. 2442-2446, 1982.
- [147] M. J. Post, "Aerosol Backscattering Profiles at CO_2 Wavelengths: The NOAA Data Base," *Appl. Opt.*, vol. 23, pp. 2507-2509, 1984.
- [148] G. A. Newkirk, Jr., "Photometry of the Solar Aureole," *J. Opt. Soc. Am.*, vol. 46, pp. 1028-1037, 1956.
- [149] G. A. Newkirk, Jr., and J. A. Eddy, "Light Scattering by Particles in the Upper Atmosphere," *J. Atmos. Sci.*, vol. 21, pp. 35-60, 1964.
- [150] A. E. Green et al., "Interpretation of the Sun's Aureole Based on Atmospheric Aerosol Models," *Appl. Opt.*, vol. 10, pp. 1263-1279, 1971.
- [151] A. E. Green et al., "Light Scattering and the Size-Altitude Distribution of Atmospheric Aerosols," *J. Colloid Interface Sci.*, vol. 39, pp. 520-535, 1972.
- [152] G. Ward et al., "Atmospheric Aerosol Index of Refraction and Size-Altitude Distribution From Bistatic Laser Scattering and Solar Aureole Measurements," *Appl. Opt.*, vol. 12, pp. 2585-2592, 1973.
- [153] D. Deirmendjian, "A Survey of Light Scattering Techniques Used in the Remote Monitoring of Atmospheric Aerosols," *Rev. Geophys. Space Phys.*, vol. 18, pp. 341-360, 1980.
- [154] T. Nakajima, M. Tanaka, and T. Yamaguchi, "Retrieval of the Optical Properties of Aerosols From Aureole and Extinction Data," *Appl. Opt.*, vol. 22, pp. 2951-2959, 1983.
- [155] N. T. O'Neill and J. R. Miller, "Combined Solar Aureole and Solar Beam Extinction Measurements, Parts 1 and 2," *Appl. Opt.*, vol. 23, pp. 3691-3696, 1984.
- [156] A. N. Lowan, "Tables of Scattering Functions for Spherical Particles," *Appl. Math Series 4*, Washington, D.C.: National Bureau of Standards, 1949.
- [157] H. G. Houghton and W. R. Chalker, "Scattering Cross-Sections of Water Drops in Air for Visible Light," *J. Opt. Soc. Am.*, vol. 39, pp. 955-957, 1949.

- [158] R. O. Gumprecht and C. M. Sliepcevich, *Tables of Scattering Functions for Spherical Particles*, Ann Arbor: University of Michigan Press, 1951.
- [159] W. J. Pangonis et al., *Tables of Light Scattering Functions of Spherical Particles*, Detroit: Wayne State University Press, 1957.
- [160] J. B. Havard, "On the Radiation Characteristics of Water Clouds at Infrared Wavelengths," doctoral thesis, University of Washington, Seattle, 1960.
- [161] D. Deirmendjian, "Tables of Mie Scattering Cross Sections and Amplitudes," R-407-PR, Santa Monica, California: RAND Corporation, 1963.
- [162] R. Penndorf, "New Tables of Total Mie Scattering Coefficients for Spherical Particles of Real Refractive Index ($1.33 < i < 1.50$)," *J. Opt. Soc. Am.*, vol. 47, pp. 1010-1015, 1957.
- [163] W. M. Irvine and J. B. Pollack, "Infrared Optical Properties of Water and Ice Spheres," *Icarus*, vol. 8, pp. 324-360, 1968.
- [164] J. L. Zelmanovich and K. S. Shifrin, *Tables of Light Scattering. Part III: Coefficients of Extinction, Scattering, and Light Pressure*, Leningrad: Hydrometeorological Publishing House, 1968.
- [165] G. C. Mooradian, M. Geller, L. B. Stotts, D. H. Stephens, and R. A. Krautwald, "Blue-Green Pulsed Propagation Through Fog," *Appl. Opt.*, vol. 18, pp. 429-441, 1979.
- [166] W. Hatch, *Selective Guide to Climatic Data Sources*, Washington, D.C.: U.S. Department of Commerce, 1983.
- [167] W. L. Wolfe (ed.), *Handbook of Military Infrared Technology*, Washington, D.C.: U.S. Government Printing Office, 1965.
- [168] *RCA Electro-optics Handbook*, Technical Series EOH-11, RCA Corporation, Lancaster, Pennsylvania, 1974.
- [169] W. E. K. Middleton, *Vision Through the Atmosphere*, Toronto: University of Toronto Press, 1952.
- [170] K. S. Shaik, "A Preliminary Weather Model for Optical Communications Through the Atmosphere," to be published.
- [171] J. R. Kerr, "Microwave-Bandwidth Optical Receiver Systems," *Proc. IEEE*, vol. 55, pp. 1686-1700, 1967.
- [172] R. F. Lucy and K. Lang, "Optical Communications Experiments at 6328 Å and 10.6 Microns," *Appl. Opt.*, vol. 7, pp. 1965-1970, 1968.
- [173] F. E. Goodwin and T. A. Nussmeier, "Optical Heterodyne Communications Experiments at 10.6 Microns," *IEEE J. Quant. Electron.*, vol. QE-4, pp. 612-617, 1968.
- [174] J. W. Ward, "A Narrowbeam, Broad Bandwidth Optical Communications System," Report NAS8-20269, San Fernando, California: ITT, Inc., 1968.
- [175] H. W. Mocker, "A 10.6 Micron Optical Heterodyne Communication System," *Appl. Opt.*, vol. 8, pp. 677-684, 1969.
- [176] R. T. Denton and T. S. Kinsel, "Terminals for a High Speed Optical Pulse Code Modulation Communication System. I: 224-Mbit/s Channel," *Proc. IEEE*, vol. 56, pp. 140-145, 1968.

- [177] T. S. Kinsel and R. T. Denton, "Terminals for High Speed Optical Pulse Code Modulation Communication System. II: Optical Multiplexing and Demultiplexing," *Proc. IEEE*, vol. 56, pp. 146-154, 1968.
- [178] R. M. Montgomery, "Ultra-Fast Pulsed Laser Systems Study," AFWAL-TR-86-359, Mountain View, California: Sylvania Electronics Systems, 1969.
- [179] *Airborne Flight Test System*, SD-TR-82-2, McDonnell-Douglas Astronautics Co., St. Louis, Missouri, 1981.
- [180] C. G. Welles et al., "HAVE LACE," AFWAL-TR-86-1102, GTE Gov. Systems Corp., Mountain View, California, 1986.
- [181] *Submarine Laser Communications Program*, NOSC Tech. Document 473, vols. 1 and 2, Naval Ocean Systems Center, San Diego, California, 1981.
- [182] J. R. Kerr et al., "Atmospheric Optical Communications Systems," *Proc. IEEE*, vol. 58, pp. 1691-1709, 1970.
- [183] E. Brookner, "Atmospheric Propagation and Communication Channel Model for Laser Wavelengths," *IEEE Trans. Commun. Tech.*, vol. COM-18, pp. 369-416, 1970.
- [184] *Optical Communications Techniques*, AFWAL-TR-85-1049, Booz, Allen and Hamilton, Inc., Bethesda, Maryland, 1985.

Table 1. Composition of "clean" dry air near sea level¹

Component	Percent by Volume	Content, ppm
Nitrogen	78.09	780900
Oxygen	20.94	209400
Argon	0.93	9300
Carbon dioxide	0.0318	318
Neon	0.0018	18
Helium	0.00052	5.2
Krypton	0.0001	1
Xenon	0.000008	0.08
Nitrous oxide	0.000025	0.25
Hydrogen	0.00005	0.5
Methane	0.00015	1.5
Nitrogen dioxide	0.0000001	0.001
Ozone	0.000002	0.02
Sulfur dioxide	0.00000002	0.0002
Carbon monoxide	0.00001	0.1
Ammonia	0.000001	0.01

¹The concentrations of some of these gases may differ with time and place, and the data for some are open to question. Single values for concentrations, instead of ranges of concentrations, are given above to indicate order of magnitude, not specific and universally accepted concentrations.

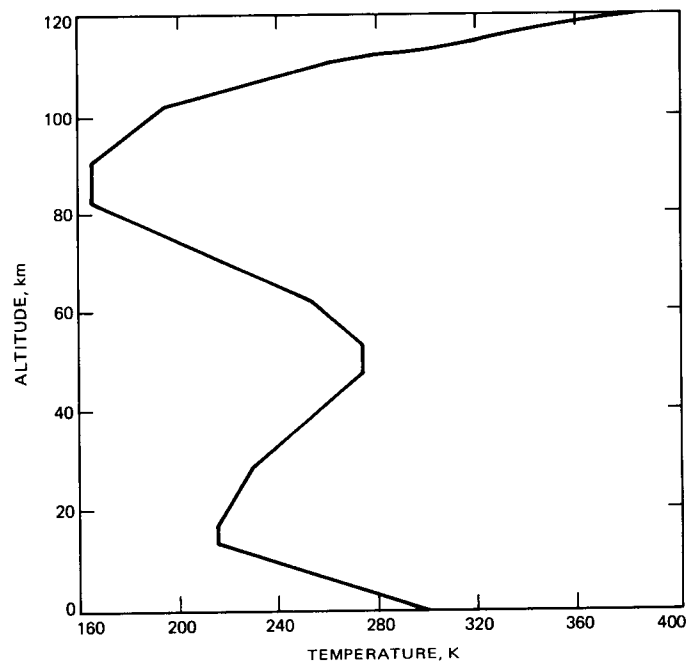


Fig. 1. Mean temperature as a function of altitude at 45 degree north latitude during July

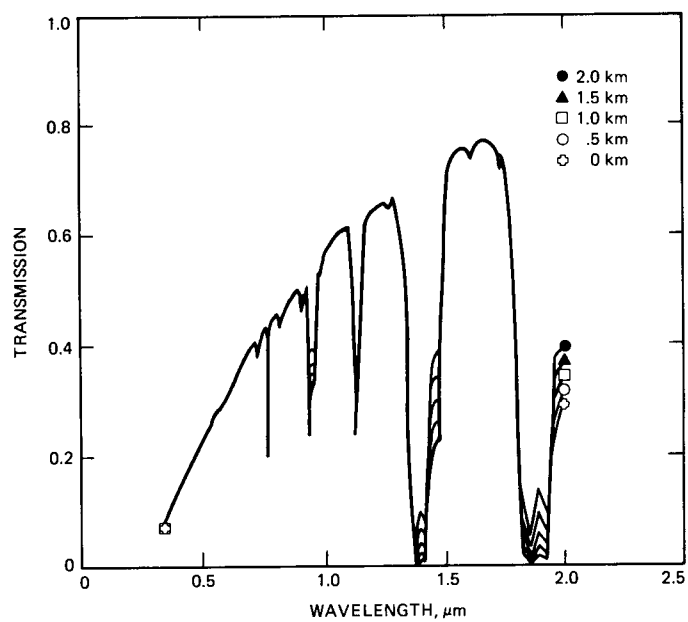


Fig. 2. LOWTRAN6 calculation of space-to-ground transmission as a function of wavelength in the presence of mid-latitude winter haze. The curves correspond to transmitter altitude.

1989000531
618418
9p.516-61
~~165027~~
April-June 1988
N89-10202
JPL

Node Synchronization of Viterbi Decoders Using State Metrics

U. Cheng

Communications Systems Research Section

This article investigates the concept of node synchronization using state metrics. In this study, the branch metrics are integrated over a fixed time interval and the results are compared to the detection threshold. If the threshold is exceeded, the out-of-sync hypothesis is accepted; otherwise, the in-sync hypothesis is accepted. It is shown that the detection threshold can be chosen independent of any particular convolutional code with fixed code rate and constraint length if the code has reasonably good bit error rate performance.

Three node synchronization schemes are compared in this article: (1) a scheme using the syndrome (Scheme 1); (2) a scheme using the frame-sync patterns (Scheme 2); and (3) a scheme using the state metrics (Scheme 3). At very low SNRs, Scheme 2 can be faster than Scheme 1. For Voyager's rate 1/2 and constraint length 7 convolutional code, this happens for SNRs of less than 0.75 dB. This result is obtained by assuming that the coded frame-sync pattern has good aperiodic autocorrelation properties. For a fixed false alarm probability, the sequential detection scheme based on the syndrome is faster than Scheme 3 with fixed integration time. A sequential detection technique is needed to improve the speed of Scheme 3.

I. Introduction

The NASA standard concatenated coding system uses a (7,1/2) convolutional code as its inner code and an 8-bit (255, 223) Reed-Solomon (RS) code as its outer code. This system achieves a bit error rate (BER) of 10^{-6} at a bit signal-to-noise ratio (SNR) of 2.53 dB. However, it is possible to improve this performance by 2 dB using the newly discovered (15,1/6) convolutional code concatenated with a 10-bit (1023,959) RS code [1], [2]. Toward the goal of demonstrating this performance improvement, a convolutional decoder is currently under development for codes with constraint lengths up to

15. Good node synchronization schemes are essential to the achievement of the aforementioned performance. Three node synchronization schemes are considered in this article. Emphasis is placed on the scheme using state metrics because this scheme can be integrated in the Viterbi algorithm. The (7,1/2) convolutional code is chosen as an example to illustrate the theory. The same principle is applicable to the convolutional codes of constraint length 15.

Many node synchronization schemes have been investigated [3], [4]. In [3], the syndrome, which is computed from the

hard-quantized received sequence, is used to detect the node synchronization. In this case, the syndrome $\{x_k\}$ is designed so that the distributions of sequences of $\{0,1\}$ in $\{x_k\}$ are different for the in-sync and out-of-sync hypotheses. Thus, by observing the distribution of $\{0,1\}$ in $\{x_k\}$, it is possible to determine which hypothesis is true. A sequential detection procedure based on this concept was described in [3]. This scheme is referred to as Scheme 1 in the subsequent discussion.

Another way to perform node synchronization is to use frame synchronization patterns. In this case, a fixed sync pattern is periodically inserted in the data stream. If the length of the sync pattern is longer than the memory length of the convolutional code, a fixed symbol pattern caused by the sync pattern periodically appears in the coded symbol stream. Therefore, a straightforward correlation process can be used to detect this pattern, and node synchronization is accomplished. The aperiodic autocorrelation property of the coded sync pattern is essential to node synchronization performance in this method. This scheme is referred to as Scheme 2 in the subsequent discussion.

The third method involves the use of state metrics in the Viterbi decoder to perform node synchronization. This is based on the observation that the growth rates of the state metrics are different under the in-sync and out-of-sync hypotheses. Many node synchronization techniques can be derived based on this observation. A sequential detection scheme was presented in [4]. In this article, a different approach is addressed, wherein the integration time is kept fixed and the result at the end of each integration period is compared to a preselected threshold. If the threshold is exceeded, the out-of-sync hypothesis is accepted (here we assume that the Viterbi decoder searches for the path with the smallest metric; see Section II for details); otherwise, the in-sync hypothesis is accepted. This scheme is referred to as Scheme 3 in the subsequent discussion. For this simple scheme, the "false alarm" and "missing" probabilities can be determined.

It will also be shown that the state metric growth rate under the out-of-sync hypothesis is independent of the choice of convolutional code with fixed rate and constraint length. This implies not only that the detection threshold can be the same for all convolutional codes with the same rate and constraint length, but that the resulting false alarm probability is also independent of the code. It turns out that the convolutional code having the best bit error rate (BER) performance may also give the best node synchronization performance. Suppose the detection threshold is determined for a given convolutional code. If another convolutional code with better BER performance is used in the future, the same detection threshold can still be used, and the missing probability is also reduced. It is also worth mentioning that the required integra-

tion time for a given false alarm probability is closely related to the BER of the convolutional code.

The concept of node synchronization using state metrics is explained in Section II. A performance analysis is given in Section III. Implementation considerations regarding Scheme 3 and a comparison of the three schemes are given in Section IV.

II. Node Synchronization Using State Metrics

Without loss of generality, we assume that the branch metrics are the square of the Euclidean distance between the received code words and the branch code words. Thus, the task of the Viterbi decoder is to search for the path with the smallest metric. For the in-sync hypothesis there is always a correct path, while for the out-of-sync hypothesis there is no correct path. If the decoder follows the correct path, the state metrics grow only because of channel noise. If the decoder follows the incorrect path, the state metrics grow at a rate determined by channel noise as well as by the mismatch between the received code word and the local reference branch code word. Thus, the state metric growth rates under the out-of-sync hypothesis are greater than those under the in-sync hypothesis. As the decoder event error probability decreases, the difference between the in-sync and out-of-sync state metric growth rates increases. Therefore, the performance of node synchronization is closely related to the BER performance of the convolutional decoder.

Note that the convolutional code improves BER performance over the uncoded system if the bit SNR is above a certain threshold, denoted by SNR_c , where the convolutional coded system has the same performance as the uncoded system. Simulation results indicate that if $SNR > SNR_c + 0.5$ dB, the node synchronization can be performed rapidly. If $SNR_c - 0.5$ dB $< SNR < SNR_c + 0.5$ dB, a long integration time may be needed. A straightforward way to use the state metrics for node synchronization is as follows: A fixed integration time is selected. The state metric growth at the end of each integration period is compared to the preselected detection threshold. If the threshold is exceeded, the out-of-sync hypothesis is accepted; otherwise, the in-sync hypothesis is accepted. The integration time and the detection threshold are selected to yield satisfactory false alarm and missing probabilities.

III. Performance Analysis

Let $D_{i\tau}$ (in-sync) and $D_{i\tau}$ (out-of-sync) denote the state metric growth of state i in τ branches under the in-sync and out-of-sync hypotheses, respectively. The false alarm and

missing probabilities based on the observations of D_{ir} are given by

$$P_{FA,i}(\tau) = \text{prob}\{D_{ir}(\text{out-of-sync}) < \gamma\} \quad (1)$$

and

$$P_{M,i}(\tau) = \text{prob}\{D_{ir}(\text{in-sync}) > \gamma\} \quad (2)$$

where γ is the detection threshold. At this moment, we assume that the state metric accumulators have infinite dynamic range. If the accumulators have a finite dynamic range, the actual implementation of the node synchronization scheme may depend on the renormalization procedure. This is explored in Section IV. The determination of the appropriate τ and γ can be made by simulation; however, it may be too time-consuming to be practical. Another approach is to generate the first $2N-1$ moments of D_{ir} by simulation. The false alarm and missing probabilities can then be upper bounded by the technique referred to as the moment method, described in [5] and [6].

A good choice of the detection threshold γ should balance the false alarm probability and the missing probability. A reasonable choice can be

$$\gamma = \frac{\sigma_1 E\{D_{ir}(\text{out-of-sync})\} + \sigma_2 E\{D_{ir}(\text{in-sync})\}}{\sigma_1 + \sigma_2} \quad (3)$$

where

$$\sigma_1 = [\text{var}\{D_{ir}(\text{in-sync})\}]^{1/2} \quad (4)$$

and

$$\sigma_2 = [\text{var}\{D_{ir}(\text{out-of-sync})\}]^{1/2} \quad (5)$$

The quantization scheme used in this simulation study is as follows: Let x be the demodulator output for each received coded symbol: namely, $x = s + n$ if "0" is sent and $x = -s + n$ if "1" is sent, where n is a normal random variable with zero mean and unit variance and

$$s = \left(\frac{2Sr}{N_0}\right)^{1/2} \quad (6)$$

In the above equation, S is the signal power, N_0 is the one-sided noise power spectrum density, and r is the code rate. The metric for each received coded symbol is given by

$$d = \begin{cases} (x - s)^2 & \text{if the local reference coded symbol is "0"} \\ (x + s)^2 & \text{if the local reference coded symbol is "1"} \end{cases} \quad (7)$$

If the desired dynamic range is α^2 and the number of quantization levels is Q , the quantized metric is given by

$$y = \begin{cases} \left\lceil \frac{Qd}{\alpha^2} + 0.5 \right\rceil & \text{if } d < \alpha^2 [1 - (1.5/Q)] \\ Q - 1 & \text{if } d > \alpha^2 [1 - (1.5/Q)] \end{cases} \quad (8)$$

where $\lceil \cdot \rceil$ represents the integer part. For the out-of-sync hypothesis, if the received coded symbols and the local reference coded symbols have equal probability of matching or mismatching, all branch metrics are identically distributed random variables. This observation indicates that D_{ir} (out-of-sync) are nearly identically distributed random variables for all convolutional codes with the same rate and constraint length. The observation is important because it implies that the false alarm probability performance is nearly independent of the selection of the convolutional code. We can see this by simulation. In Table 1, we compare $E\{D_{ir}(\text{out-of-sync})\}$ for two convolutional codes. The first convolutional code is Voyager's code with free distance 10. The second code was selected arbitrarily. It has free distance 7, and thus the BER performance may be much worse than that of the first code. We find that the first few moments of D_{ir} (out-of-sync) are roughly the same. Suppose that the detection threshold is determined for a given convolutional code (code A) and another convolutional code with the better BER performance (code B) is used in the future. Since $E\{D_{ir}(\text{out-of-sync})\}$ for the two codes are roughly the same and code B has a larger metric growth rate difference under two hypotheses (because of the better BER performance), the same detection threshold can still be used for code B, and the missing probability is also reduced.

The state metric growth rate under the out-of-sync hypothesis depends on three factors: SNR, the integration time, and the quantization scheme. For the in-sync hypothesis, in addition to the three factors just mentioned, the state metric growth rate also depends on the convolutional code. The integration time and the quantization scheme are usually determined in advance. Therefore, they do not create a problem in the selection of γ . The SNR dependence of γ must be handled with care since SNR is usually unknown in advance and can

be estimated only during operation. We know that the operation of the AGC/quantizer also needs a good SNR estimate. Therefore, a good estimate of SNR is essential to the operation of the Viterbi decoder.

Figures 1 and 2 show the state metric growth versus the integration time τ . It can be seen that the difference between $D_{i\tau}$ (out-of-sync) and $D_{i\tau}$ (in-sync) increases linearly with τ . In Fig. 2, $D_{i\tau}$ (in-sync) at an SNR of 2 dB is used as the reference in order to present all curves clearly. We see that although both $D_{i\tau}$ (in-sync) and $D_{i\tau}$ (out-of-sync) grow faster at the lower SNR, the separation between them grows more slowly. Figures 3 and 4 show the normalized metric growth difference versus the integration time. Note that the normalized metric growth difference determines the node synchronization performance. Figure 4 shows the need for long integration times at an SNR of 0.5 dB. In Figs. 5 and 6, the missing probability is shown. Figures 7 and 8 show the false alarm probability. All curves shown in Figs. 5 through 8 were computed using the upper bound based on the moment method, and in all cases the convolutional code generator is (171,133). The smoothness and accuracy of these curves depend on the amount of simulation data.

IV. Implementation and Comparison

The implementation of a node synchronization scheme based on state metrics depends to a large extent on the renormalization procedure. Suppose that the state metric accumulators have dynamic range 2^n , and 2^{n-1} is subtracted from all state metrics at each time of renormalization. If SNR = 2 dB, $\tau = 1600$, and $n = 18$, the expected numbers of times of renormalization are $E\{D_{i\tau} \text{ (in-sync)}\}/2^{17} = 0.8468$ and $E\{D_{i\tau} \text{ (out-of-sync)}\}/2^{17} = 0.9532$. Clearly, just looking at the number of times of renormalization is not enough to determine which hypothesis is true; the actual contents of the accumulators must be examined. On the other hand, if $n = 10$, the expected numbers of times of renormalization are $E\{D_{i\tau} \text{ (in-sync)}\}/2^9 = 216.78$ and $E\{D_{i\tau} \text{ (out-of-sync)}\}/2^9 = 244.02$. In this case, the number of times of renormalization can provide useful information about which hypothesis is true.

Since the statistics of the state metrics are roughly the same for all states, the implementation can be based on the growth of the minimum state metric. The node synchronization scheme by state metrics can be implemented outside the Viterbi

decoder if the minimum state metric is made available. The advantage of this external node synchronization installation is that many schemes can be implemented; the best scheme can be selected during the decoder operation.

For Scheme 3 (node synchronization using state metrics), if $P_{FA} = 10^{-6}$, the integration time is about 1500 branches at an SNR of 2 dB, 2800 branches at an SNR of 1 dB, and 12,000 branches at an SNR of 0.5 dB. For Scheme 2 (node synchronization using the frame sync patterns), the worst case detection delay is the number of bits between two consecutive sync patterns if one sync pattern is sufficient for detection; it is between 3000 and 8000 bits in our case. Thus, roughly speaking, Scheme 2 is faster when the SNR is less than 0.75 dB for Voyager's rate 1/2 and constraint length 7 convolutional code. Note that more than one sync pattern may be needed to get the satisfactory performance for the convolutional codes of constraint length 15.

The comparison between Scheme 1 (node synchronization using the syndrome) and Scheme 3 is based on two parameters: the short ARL and the long ARL, defined as follows:

Short ARL = average number of branches needed to accept the out-of-sync hypothesis if the out-of-sync hypothesis is true from the beginning.

Long ARL = average number of branches needed to accept the out-of-sync hypothesis if the in-sync hypothesis is true from the beginning.

For Scheme 3, the short ARL is determined by the integration time, and the long ARL is determined by the missing probability. Explicitly, the long ARL is equal to the integration time divided by P_M . Thus, the long ARL = 1.3×10^9 branches at SNR = 2 dB, and the required integration time is about 1300 branches. Compared to the results in [3], at an SNR of 2.3 dB, if the long ARL is 7.9×10^9 , the corresponding short ARL is 171.4 branches. Note that Figs. 5 and 6 are computed using the upper bound based on the moment method. This bound is usually not tight, although it does provide good insight into the performance. The actual integration time for Scheme 3 could be shorter than 1300 branches, and it may be determined by simulation. Based on the above results, we conclude that Scheme 1 could be faster than Scheme 3. To reduce the integration time of Scheme 3 further, a sequential detection procedure is needed [4].

Acknowledgment

The author would like to thank Dr. F. Pollara for his valuable comments during this study.

References

- [1] J. H. Yuen and Q. D. Vo, "In Search of a 2 dB Coding Gain," *TDA Progress Report 42-83*, vol. July–September 1985, Jet Propulsion Laboratory, Pasadena, California, pp. 26–33, November 15, 1985.
- [2] S. Dolinar, "A New Code for Galileo," *TDA Progress Report 42-93*, vol. January–March 1988, Jet Propulsion Laboratory, Pasadena, California, pp. 83–96, May 15, 1988.
- [3] G. Lorden, R. J. McEliece, and L. Swanson, "Node Synchronization for the Viterbi Decoder," *IEEE Transactions on Communications*, vol. COM-32, no. 5, pp. 524–531, May 1984.
- [4] G. K. Huth, *A Parametric Study of the Complexity of Sequential Decoders*, Axiomatix Report No. R7201-1, Axiomatix Corp., Los Angeles, California, January 1972.
- [5] J. K. Omura and M. K. Simon, *Satellite Communication Performance Evaluation: Computational Techniques Based on Moments*, JPL Publication 80-71, Jet Propulsion Laboratory, Pasadena, California, September 1980.
- [6] R. A. Scholtz, *The Moment Problem*, Department of Electrical Engineering, University of Southern California, 1981.

Table 1. Typical $E [D_{tr} \text{ (out-of-sync)}]$ for two convolutional codes

τ	Code #1 generator	Code #2 generator	Difference/ σ_2 (Code #1)
	(171,133)	(101,133)	
8000	695,476	697,437	0.76
16,000	1,391,302	1,394,790	1.02
24,000	2,086,506	2,090,037	0.83
32,000	2,781,027	2,787,637	1.15
40,000	3,477,955	3,482,137	0.82

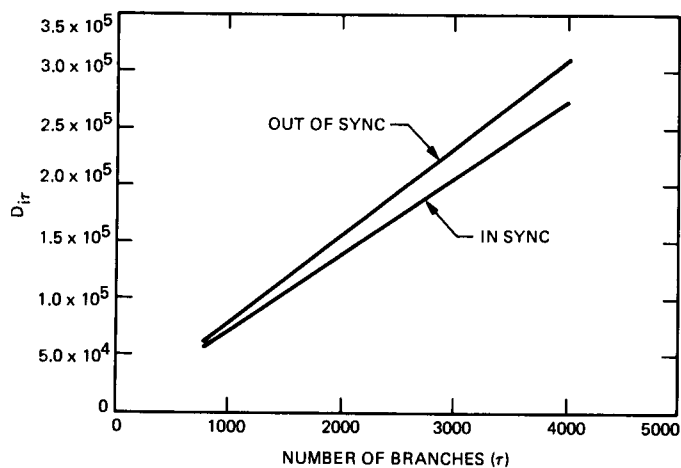


Fig. 1. Metric growth versus integration time; SNR = 2 dB, $Q = 1024$, and $\alpha = 5.52$

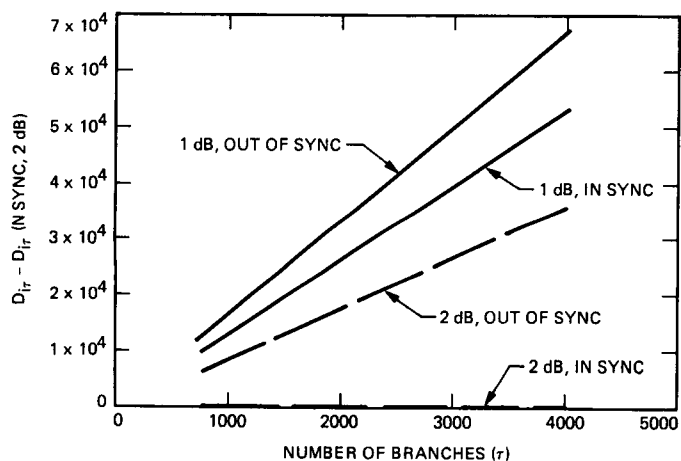


Fig. 2. Metric growth versus integration time with SNR as the parameter. $Q = 1024$, $\alpha = 5.52$ if SNR = 2 dB; $\alpha = 4.84$ if SNR = 1 dB.

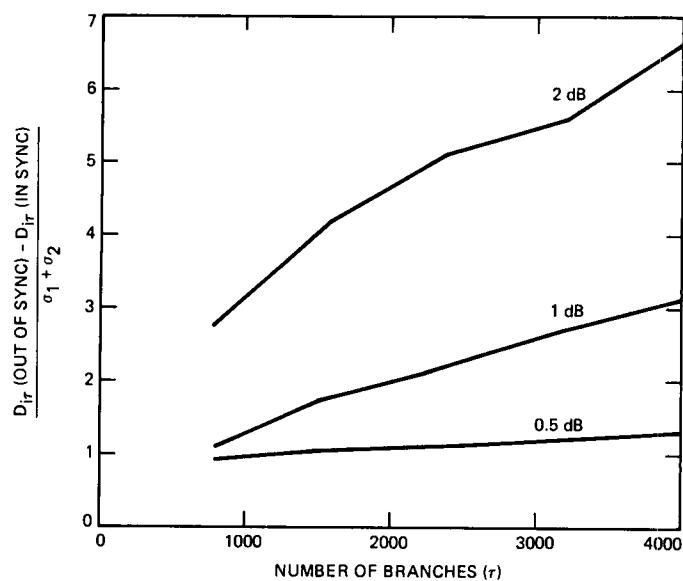


Fig. 3. Normalized metric growth difference versus integration time with SNR as the parameter. $Q = 1024$, $\alpha = 5.52$ if SNR = 2 dB; $\alpha = 4.84$ if SNR = 1 dB; $\alpha = 4.62$ if SNR = 0.5 dB.

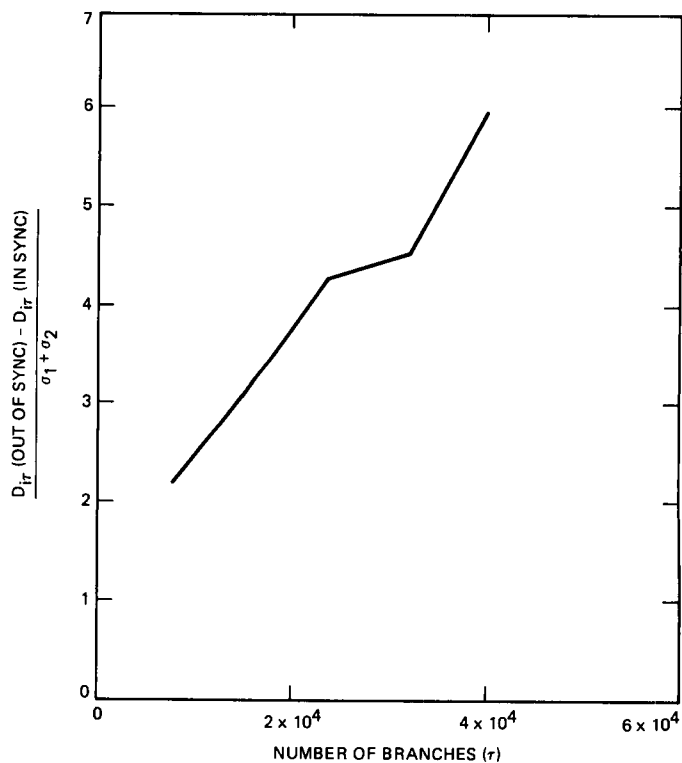


Fig. 4. Normalized metric growth difference versus integration time with SNR = 0.5 dB, $Q = 1024$, and $\alpha = 4.62$

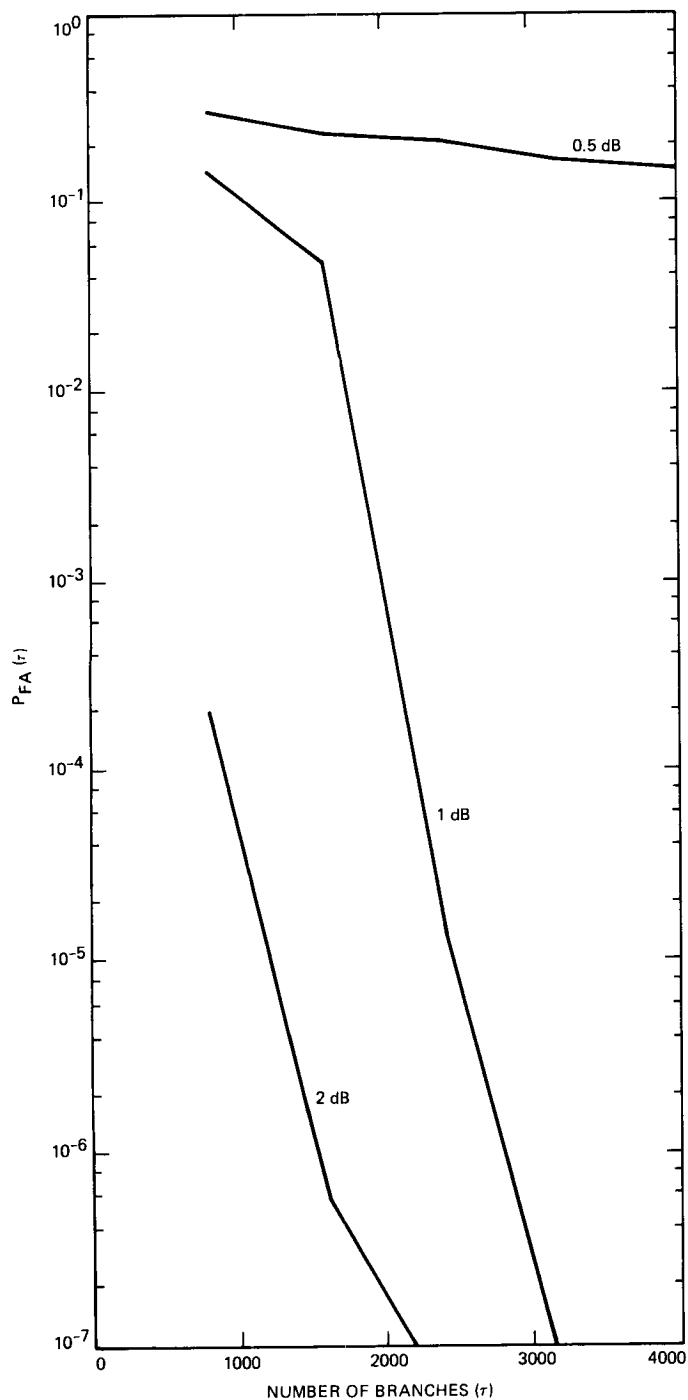


Fig. 5. Missing probability versus integration time with SNR as the parameter. $Q = 1024$; $\alpha = 5.52$ if SNR = 2 dB; $\alpha = 4.84$ if SNR = 1 dB; $\alpha = 4.62$ if SNR = 0.5 dB.

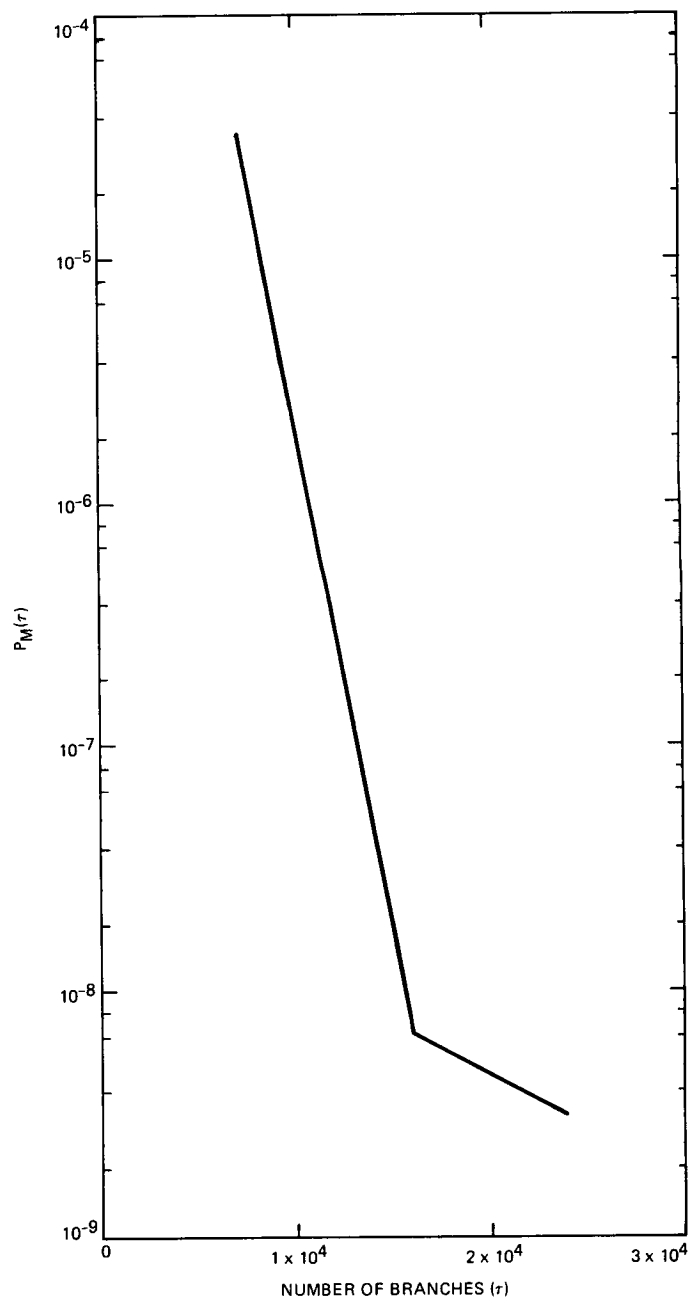


Fig. 6. Missing probability versus integration time with SNR = 0.5 dB. $Q = 1024$ and $\alpha = 4.62$.

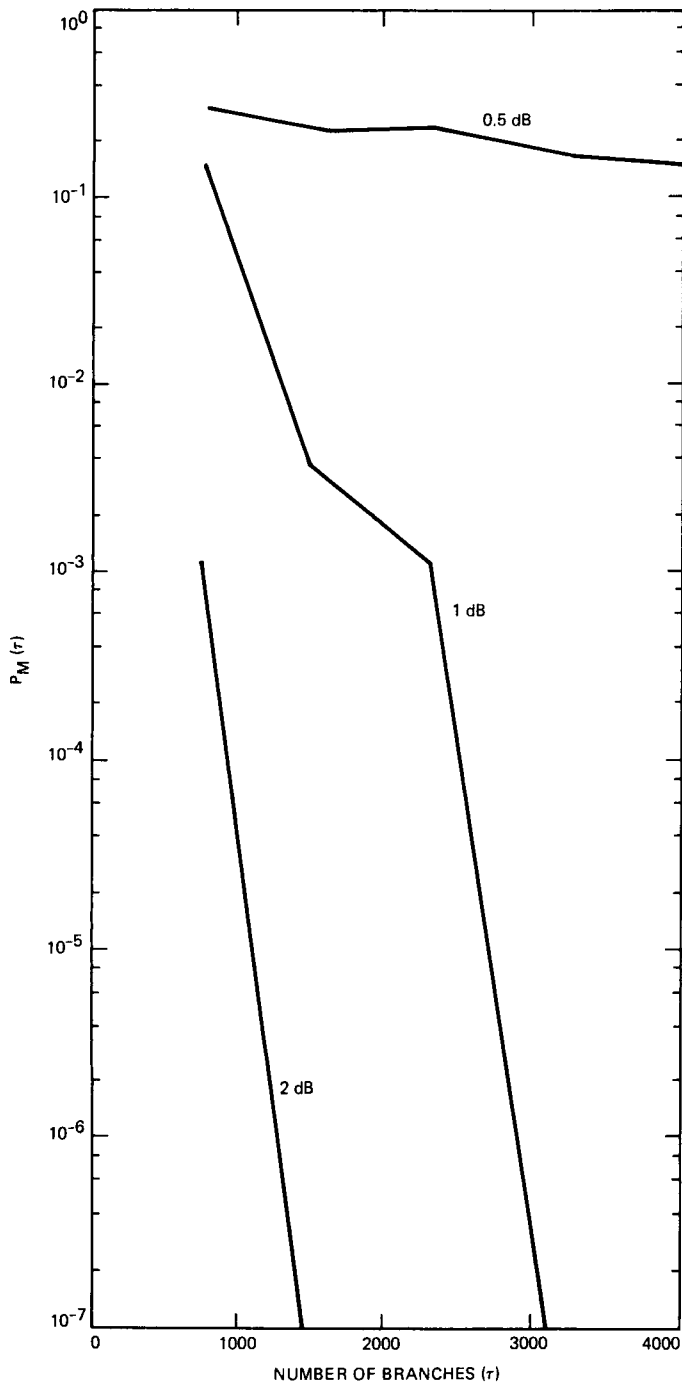


Fig. 7. False alarm probability versus integration time with SNR as the parameter. $Q = 1024$. $\alpha = 5.52$ if SNR = 2 dB; $\alpha = 4.84$ if SNR = 1 dB; $\alpha = 4.62$ if SNR = 0.5 dB.

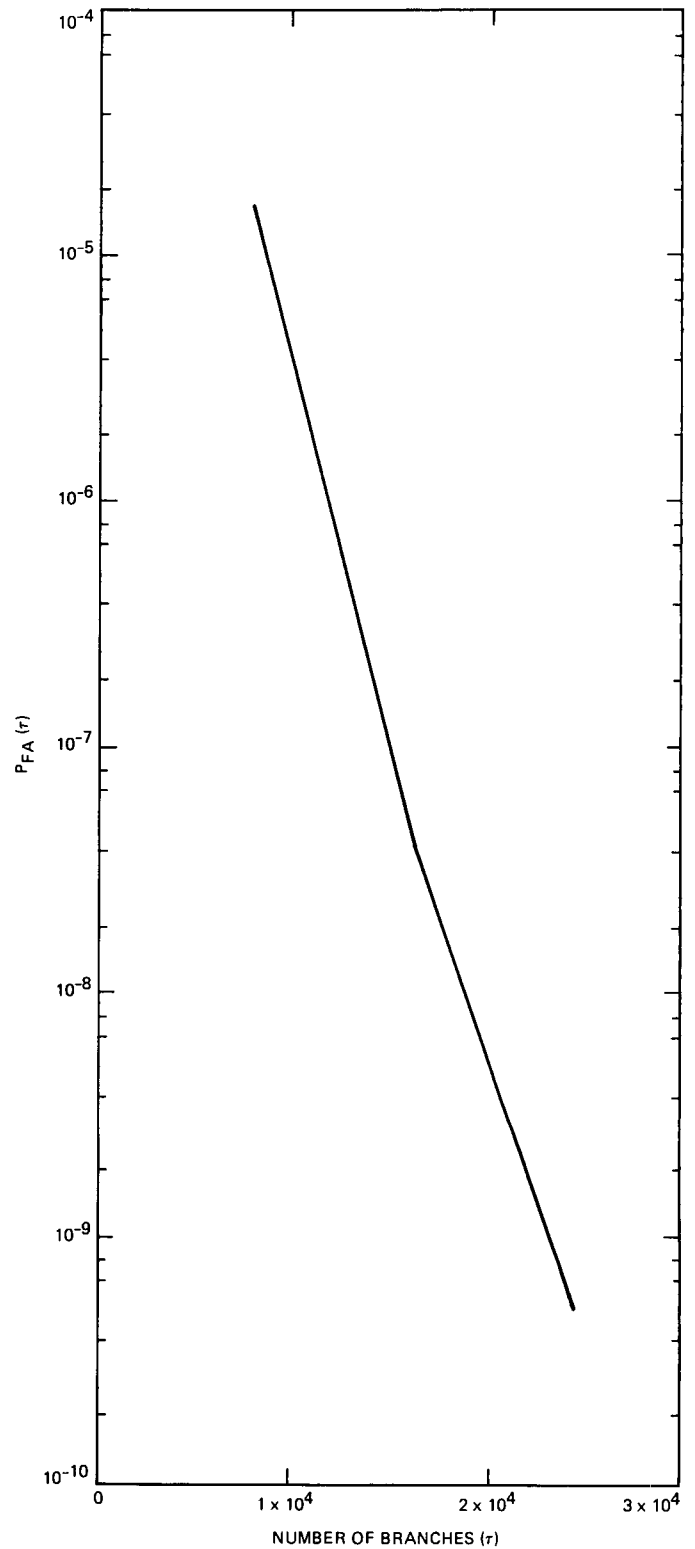


Fig. 8. False alarm probability versus integration time with SNR = 0.5 dB. $Q = 1024$ and $\alpha = 4.62$.

A Software Simulation Study of the Long Constraint Length VLSI Viterbi Decoder

S. Arnold and F. Pollara
Communications Systems Research Section

A software simulation of long constraint length Viterbi decoders has been developed. This software closely follows the hardware architecture that has been chosen for the VLSI implementation. The program is used to validate the design of the decoder and to generate test vectors for the VLSI circuits.

I. Introduction

Convolutional codes have been used on deep space probes for several years. During the last few years, TDA Advanced Systems undertook a research effort [1] to develop advanced coding techniques capable of gaining an additional 2 dB over the present performance of deep space missions. Current coding systems are based on a $K = 7$, $r = 1/2$ convolutional code concatenated with an 8-bit (255,223) Reed-Solomon (RS) code, where K is the constraint length and r is the code rate.

The main result of this research effort was the discovery of new convolutional codes with $K = 15$ and $r = 1/6$ which exceed the 2-dB goal when concatenated with a 10-bit (1023,959) RS code. Recently, the delay imposed on the Galileo mission introduced the possibility of including a $K = 15$, $r = 1/4$ code in this mission. This experimental code [2] will gain approximately 1.5 dB over the current NASA-standard code. The Galileo experiment, together with the potential offered by these coding gains for future missions, has led to an effort to build a VLSI-based Viterbi decoder capable of decoding codes

with K up to 15 and $r = 1/n$, $n = 2,3,4,5,6$, at speeds approaching 1 Mbit/s.¹

II. Decoder Architecture

The complexity of a Viterbi decoder depends mainly on the constraint length K , since the number of states is 2^{K-1} . The decoder for the new $K = 15$ codes is approximately 256 times more complex than the current MCD (Maximum-likelihood Convolutional Decoder) used in the DSN stations to decode $K = 7$ codes. The requirement on the information data rate forces the use of heavily parallel architectures.

After evaluation of several design alternatives, it was decided to use a fully parallel architecture consisting of $2^{K-2} = 8192$ physical butterflies operating in parallel. Each butterfly uses bit-serial arithmetic to perform the internal operations of add-

¹J. Statman, "Preliminary Design Review for Big Viterbi Decoder (BVD)," JPL internal document, March 31, 1988.

compare-select, since this is more suitable to fast VLSI circuits, and represents the metrics as 16-bit numbers. Each butterfly contains two states of the decoder and outputs two decision bits to the trace-back memory. The 8192 butterflies are organized in identical VLSI chips containing 32 butterflies each, and in 16 identical boards containing 16 chips each [3].

The concern was to develop a software simulation of the complete decoder so that (1) several new design ideas could be tested and validated; and (2) test vectors could be generated for signals at various key points in the decoder and then used to test the VLSI design. Given the complexity and the cost of this project, it was necessary to have a complete software decoder that closely emulated the hardware architecture and demonstrated the validity of the design.

III. Software Decoder

The software decoder consists of a program developed on a SUN 3/260 workstation and written in C-language. Since the program runs on a sequential computer, it scans through the butterflies in sequential order, while the hardware performs all these operations in parallel. The decoder is based on the hardware design summarized in Figs. 1 and 2.

Figure 1 represents the *metric computer* module present in each butterfly. It takes the received symbols in sign and magnitude representation and computes the two branch metrics, p and q , as two 16-bit numbers. The register denoted as $LABEL_i$ is initialized at startup time and contains an appropriate label for the i th butterfly. The value of this label is provided by the module *encoder*, whose operation is described by the flow diagram of Fig. 3. Here NB represents the total number of butterflies (8192), i is the index of the current butterfly, and j the index of encoded symbols e_j . First, the n encoded symbols e_j are computed for the current butterfly. Then $LABEL_i$ is just given by the decimal equivalent of the binary array (e_0, e_1, \dots, e_5) . The other input to the metric computer module, r_{max} , is just the sum of the magnitudes of the received symbols for each information bit time. Notice that the diagram in Fig. 1 shows six input received symbols, but it can be used for any code rate $r = 1/n$, $n = 2, 3, 4, 5, 6$, by setting the unused symbols to zero. Figure 4 shows a flow diagram representing the computations taking place in the software. The variable j counts the received symbols modulo n .

The *add-compare-select* circuit of Fig. 2 takes the branch metrics p and q just computed and the previously computed accumulated metrics m_{i0} and m_{i1} from states $i0$ and $i1$ and generates the updated metrics m_{j0} , m_{j1} and the decision bits bit_0 and bit_1 , which are stored in the trace-back memory. This memory is organized in three banks of L bits each, where L is the path truncation length. Decoded bits are given by the trace-back performed on the bank containing the "oldest" decision bits. The detailed operation of the add-compare-select module is shown in the flow diagram of Fig. 5. The test for overflow is performed on the output accumulated metric m_{j0} of butterfly number zero. Renormalization occurs if the two most significant bits of m_{j0} are both equal to one. In this case, the most significant bit of all accumulated metrics is reset to zero to prevent overflow of the metrics. The decoder described in this article and its future VLSI implementation can decode any code with connection vectors $G_i = (x_{i0}, x_{i1}, \dots, x_{i14})$, where $x_{ij} \in (0, 1)$ and $x_{i0} = x_{i14} = 1$. Code search results [1], [2] show that good codes always meet the constraint of having a leading and trailing "1" in the connection vectors. Because of this constraint, only two branch metrics, p and q , need to be computed. When $K < 15$, this constraint is no longer met, but it can be observed that in this case m_{j1} is always equal to m_{j0} , as shown in [3]. This is accomplished with the switch in Fig. 2 or the test ($K < 15$) in Fig. 5.

IV. Operation of the Software Decoder

Testing a large Viterbi decoder is a complex task, since some programming errors may be revealed only by particular input sequences or error patterns. This decoder has been tested first against an existing software decoder for $K = 7$ and $r = 1/2$, which has been extensively used in the past. After it was ascertained that the two programs had identical behavior, the new program was tested with various other codes, and it also performed according to expectations.

To run the program, which is reproduced in the Appendix, the user must enter K , the inverse n of the rate, the path truncation length L , and the generator polynomials in octal. Also, the names of the files used to get the input received symbols and to write the decoded bits must be provided. Currently, the output consists of the decoded information bits and is written to a disk file. The test signals to be used for future testing of the VLSI circuits can be obtained by inserting print statements anywhere desired in the program.

References

- [1] J. H. Yuen and Q. D. Vo, "In Search of a 2-dB Coding Gain," *TDA Progress Report 42-83*, vol. July–September 1985, Jet Propulsion Laboratory, Pasadena, California, pp. 26–33, November 15, 1985.
- [2] S. Dolinar, "A New Code for Galileo," *TDA Progress Report 42-93*, vol. January–March 1988, Jet Propulsion Laboratory, Pasadena, California, pp. 83–96, May 15, 1988.
- [3] O. Collins, "Techniques for Long Constraint Length Viterbi Decoders," Intern. Symposium on Information Theory, Kobe, Japan, p. 28, June 1988.

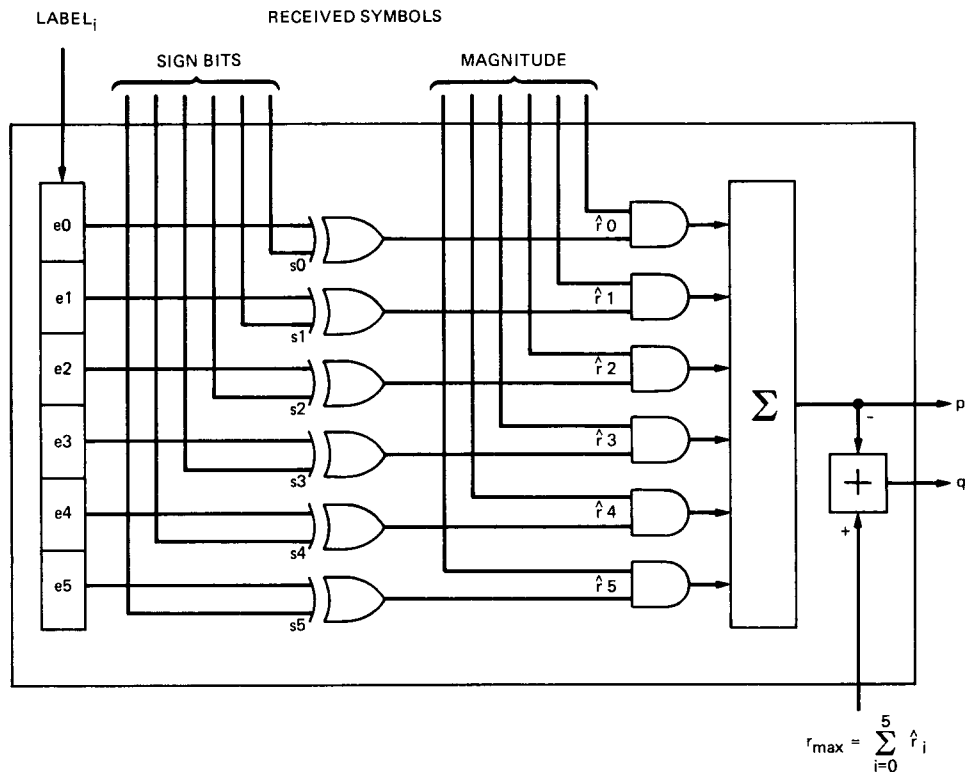
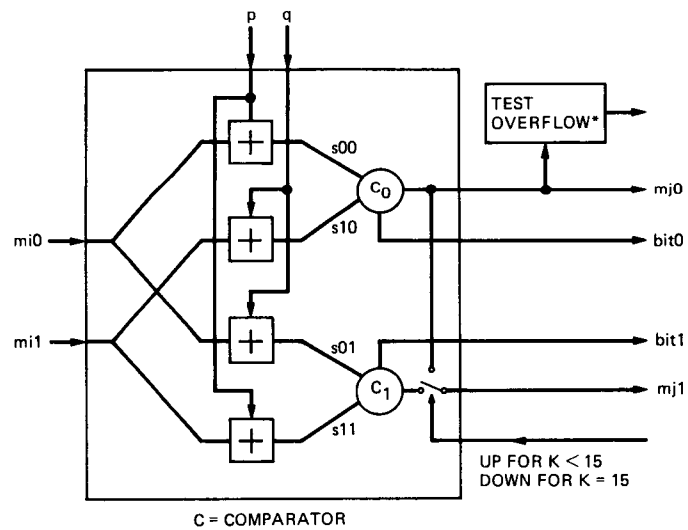


Fig. 1. Metric computer hardware diagram



*THIS IS USED ONLY FOR BUTTERFLY = 0

Fig. 2. Add-compare-select hardware diagram

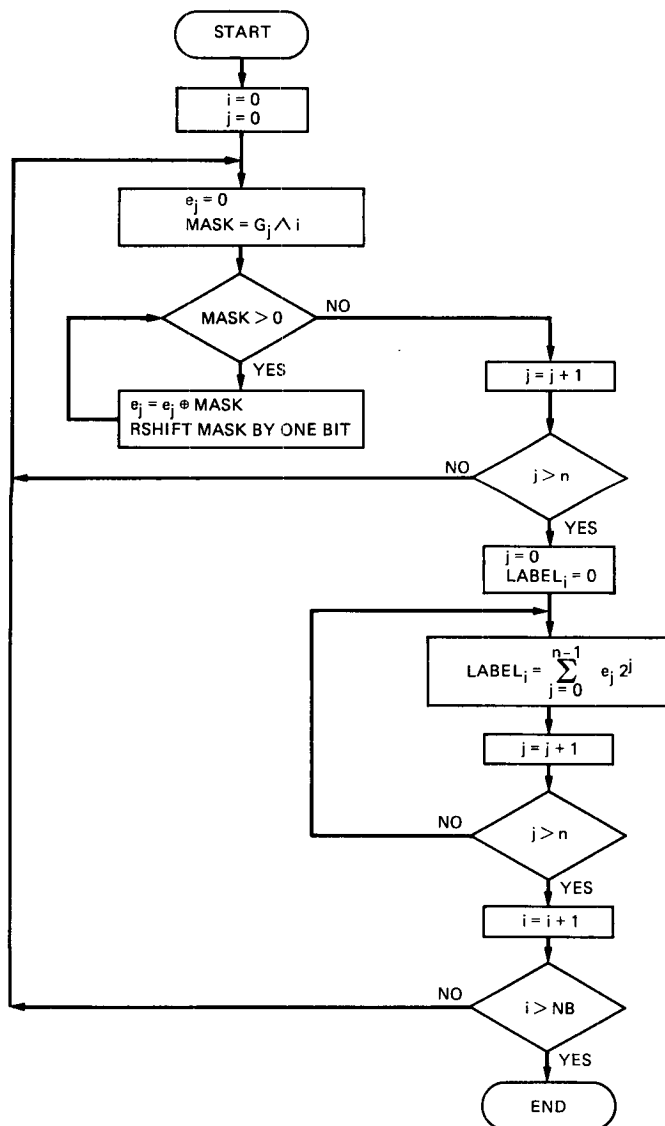


Fig. 3. Encoder flow diagram

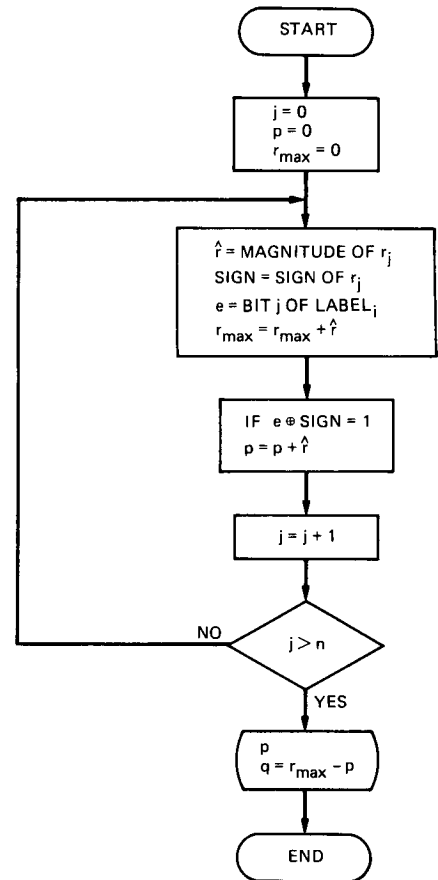


Fig. 4. Metric computer flow diagram

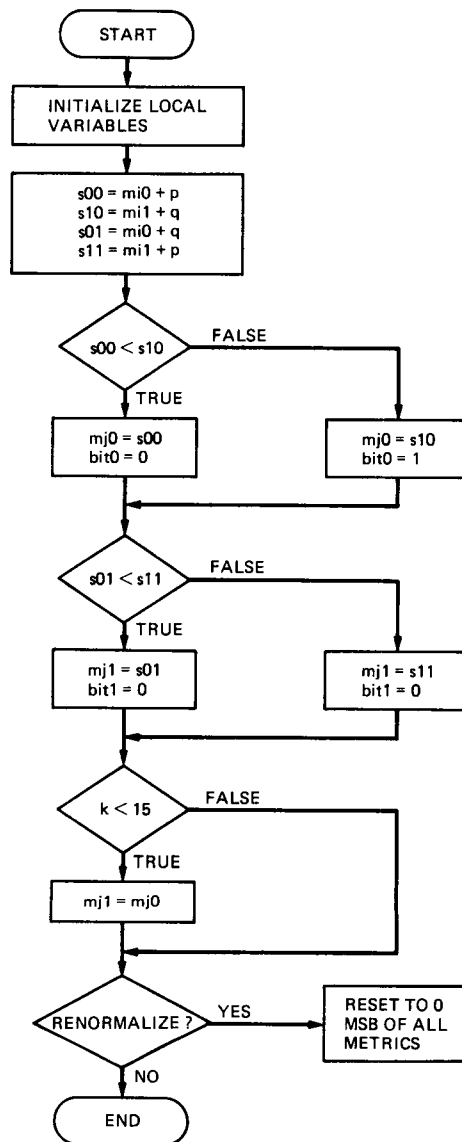


Fig. 5. Add-compare-select flow diagram

Appendix

```
# include <stdio.h>

/***** VLSI *****/
/*
/* This program simulates the long constraint length VLSI Viterbi decoder. */
/* It allows the user to decode convolutional codes with constraint length */
/* up to 15 and code rate 1/2 to 1/6. */
/*
/*****

int n;                /* rate= 1/n */
int L;                /* buffer length */
int p,q;              /* branch metrics */
int n_tb;             /* traceback addresses */
int time;             /* traceback time */
int butt;             /* loop counter */
int NS;               /* number of states */
int NB;               /* number of butterflies */
int k;                /* constraint length */
int dec;              /* decoding bank */
int mi0, mi1, mj0, mj1; /* accumulated metrics */
int blk_time;         /* time in traceback */
int bit_no;           /* number of symbols decoded */
int GP[6];            /* generator polynomials */
int out[100];         /* temp storage for decoded bits */
int outr[100];        /* storage for decoded bits */
int LABEL[8192];      /* butterfly labels */
int metric[16384], old_metric[16384]; /* accumulated metric storage */
char flag;            /* renormalization flag */
char bit0, bit1;      /* decision bits */
char RAM[16384][100][3]; /* traceback RAM */

main ()
{
    int tb;            /* traceback (tb) bank */
    int Mo;            /* parameter to calculate memory size */
    int bank;          /* loop counter */
    int blk_no;        /* number of blocks decoded */
    int blk_par;       /* block parity (0 or 1) */
    int n_dec;         /* addresses of decoded bits */
    int state;         /* loop counter */
    int symbol_no;     /* loop counter */
    int state0, state1; /* current states of butterfly */
    int prev_state0, prev_state1; /* previous states of butterfly */
    int recsym[6];     /* received symbols (8-bit) */
    char decinp[10], decout[10]; /* input and output files of decoder */
    FILE *fp1, *fp2;   /* input/output file pointers */

    printf ("The simulation can decode binary data with a constraint length");
    printf ("k <= 15 and code rate of 1/2 to 1/6");

    printf ("Enter constraint length k");
}
```

```

scanf ("%d",&k);
printf ("Enter number of symbols n (2-6)");
scanf ("%d",&n);
printf ("Enter length of traceback buffer L");
scanf ("%d",&L);

for (symbol_no= 0; symbol_no< n; symbol_no+ + ) {
    printf ("Enter generating polynomial GP[%d] in OCTAL > ",symbol_no);
    scanf ("%o",&GP[symbol_no]);
    if (k < 15) GP[symbol_no] < = (15 - k);
}

printf ("Enter binary input filename");
scanf ("%s",decinp);
printf ("Enter output filename that will contain decoded bits");
scanf ("%s",decout);

fp1 = fopen (decinp,"r");          /* open file of received symbols */
fp2 = fopen (decout,"w");          /* open file for decoder output */

bit_no = 0;                        /* set bit counter to zero */
symbol_no = 0;                     /* set symbol counter to zero */
flag = 0;                          /* set renormalization flag to zero */
n_tb = 0;                          /* set starting tb addr. to zero */

Mo = 14;
NS = 01 < M0;                     /* number of states */
NB = NS/2;                         /* number of butterflies */

/* set storage of decoded bits to zero */

for time = 0; time < L; time+ + ) out[time] = 0;

/* initialize metrics, accumulated metrics, and traceback RAM to zero*/

for (state = 0; state < NS; state+ + ) {
    metric[state] = 0;
    old_metric[state] = 0;
    for (time = 0; time < L; time+ + )
        for (bank = 0; bank < 3; bank+ + )
            RAM[state][time][bank] = 0;
}

/* generate the labels that are assigned to the butterfly */

encoder( );

/* receive data bits and enter decoder loop */

while ((recsym[symbol_no] =getc (fp1)) != EOF) {
    symbol_no+ + ;

    if (symbol_no == n) {
        symbol_no = 0;

```

```

/* check value of flag to determine to renormalize accumulated metrics */

    if (flag == 0)
        for (state = 0; state < NS; state++ )
            old_metric[state] = metric[state];
    else {
        for (state = 0; state < NS; state++ )
            old_metric[state] = metric[state]&077777; /* clear MSB */
        flag = 0;
    }

    blk_time = bit_no%L;

/* check to see if new traceback must be started */

    if (blk_time == 0) {
        blk_no = bit_no/L;
        blk_par = blk_no%2;
        tb = blk_no%3;
        dec = (tb+ 1)%3;
        n_dec = n_tb;
        n_tb = 0;
        for (time = 0; time < L; time++ ) outr[time] = out[time];
    }

/* determine whether to move left or right through traceback memory */

    if (blk_par == 0) time = L - blk_time - 1;
    else             time = blk_time;

/* generate the addresses for the decoded bits and traceback */

    n_dec = (n_dec >> 1) | (NB*RAM[n_dec][time][dec]);
    n_tb = (n_tb >> 1) | (NB*RAM[n_tb][time][tb]);

    out[blk_time] = (n_dec >> 5)&01; /* extract decoded bits */

/* Generate branch metrics associated with new received symbol, add to */
/* existing accumulated metrics, determine smallest accumulated metric */
/* at current state, and output decision bits to traceback memory */

    for (butt = 0; butt < NB; butt++ ) {

/* compute the two current states of butterfly and their associated previous states */

        state0 = butt << 1;
        state1 = state0 + 1;
        prev_state0 = butt;
        prev_state1 = prev_state0 | NB;

        metric_comp(recsym); /* call metric computer */

        mi0 = old_metric[prev_state0];
        mi1 = old_metric[prev_state1];

```

```

        add_comp_select();          /* call add, compare, and select */

        metric[state0] = mj0 ;
        metric[state1] = mj1 ;

/* write to traceback RAM, the bits at corresponding state of butterfly */

        RAM[state0][time][dec] = bit0;
        RAM[state1][time][dec] = bit1;

    }
    fprintf (fp2,"%d",outr[L-blk_time-1]); /* output decoded bits */
    fflush(fp2);
    bit_no+ + ;                        /* increment bit counter */
}
}

/***** METRIC COMPUTER *****/
/*
/* This subroutine computes the branch metrics from the "n" received
/* symbols.
/*
/*****

metric comp(recsym)
int *recsym;
{
    int symbol_no;          /* loop counter */
    int sum_recsym;          /* maximum branch metric */
    int encoded_bit;        /* one bit of branch label */
    int mag_recsym;          /* received symbol magnitude */
    int sign_recsym;        /* sign of received symbol */

    sum_recsym = 0;          /* set branch metric to zero */
    p = 0;

    for (symbol_no = 0; symbol_no < n; symbol_no+ + ) {
        mag_recsym = recsym[symbol_no] & 0177;      /* mask the first eight bits */
        sign_recsym = (recsym[symbol_no] >> 7) & 01; /* extract sign bit */
        encoded_bit = (LABEL[butt] >> symbol_no) & 01; /* strip label bits */
        sum_recsym += mag_recsym; /* sum all the received symbol magnitudes */
        if ((encoded_bit ^ sign_recsym) == 01) p += mag_recsym;
    }
    q = (sum_recsym -p);
}

/***** ADD, COMPARE, AND SELECT *****/
/*
/* Add branch metrics to accumulated metrics. The pair of sums at each
/* of the states is compared and the smallest is selected. The output
/* of each of these decisions is the smallest accumulated metric at each
/* state and the decision bits which are sent to the traceback memory.
/*
/*****

```

```

add_comp_select()

{
    int s00, s10, s01, s11;    /* the accumulated metrics */

    /* add branch metric to accumulated metric */

    s00 = (mi0 + p);
    s10 = (mi1 + q);
    s01 = (mi0 + q);
    s11 = (mi1 + p);

    /* determine smallest metric for present two states of butterfly */

    if (s00 < s10) { bit0 = 0; mj0 = s00; }
    else { bit0 = 1; mj0 = s10; }

    if (s01 < s11) { bit1 = 0; mj1 = s01; }
    else { bit1 = 1; mj1 = s11; }

    /* check constraint length and set output accumulated metrics respectively */

    if (k < 15) mj1 = mj0;

    /* determine if accumulated metrics must be renormalized, if so, set flag */

    if (but0 == 0 && (mj0 >> 14) == 3) flag = 1;
}

/***** ENCODER *****/
/*
/* This subroutine generates the labels for each butterfly by utilizing
/* the appropriate generating polynomials.
/*
/*****
encoder()
{
    int butt;                /* loop counter */
    int symbol_no;           /* loop counter */
    int encoded[6];          /* encoded symbols */
    unsigned int masked;     /* the masked state */

    /* encode butterfly labels and do appropriate shifting */

    for (butt = 0; butt < NB; butt++ ) {
        for (symbol_no = 0; symbol_no < n; symbol_no++ ) {
            encoded[symbol_no] = 0;
            masked = (butt < 1) & GP[symbol_no]; /* mask the butterfly */
            for ( ; masked > 0; masked >>= 1)
                encoded[symbol_no] ^= masked; /* sum the bits of butterfly */

```

```
    }  
    LABEL[butt] = 0;  
    for (symbol_no = 0; symbol_no < n; symbol_no+ + )  
        LABEL[butt] |= (encoded[symbol_no]&01) << symbol_no;  
    }  
}
```


518-32

165629

TDA Progress Report 42-94

101

1989000833

10p.

618432

N89 - 10204

April-June 1988

JIP

An Automatic Frequency Control Loop Using Overlapping DFTs

S. Aguirre

Communications Systems Research Section

An automatic frequency control (AFC) loop is introduced and analyzed in detail. The new scheme is a generalization of the well-known Cross-Product AFC loop that uses running overlapping discrete Fourier transforms (DFTs) to create a discriminator curve. Linear analysis is included and supported with computer simulations. The algorithm is tested in a low carrier to noise ratio (CNR) dynamic environment, and the probability of loss of lock is estimated via computer simulations. The algorithm discussed is a suboptimum tracking scheme with a larger frequency error variance compared to an optimum strategy, but offers simplicity of implementation and a very low operating threshold CNR. This technique can be applied during the carrier acquisition and re-acquisition process in the Advanced Receiver.

I. Introduction

Automatic frequency tracking is a subject of great importance in the fields of communications, control, and signal processing. For instance, it may provide initial rapid acquisition in coherent receivers, may supply information on the velocity of rapidly moving targets, and can be used in conjunction with channel doppler extractors, just to name a few applications.

This article presents a study of the fundamental problem of estimating the frequency of a sinusoidal wave embedded in noise, subject to severe dynamics. It is well known [1], [2] that the maximum likelihood estimator (MLE) of a constant (but unknown) frequency sine wave is equivalent in practice to a discrete Fourier transform (DFT) operation on the received discrete data, the MLE of frequency being the center frequency of the bin filter in the DFT with maximum output

power. For a stable frequency, arbitrarily good resolution is obtained by increasing the size of the DFT or, equivalently, by extending the observation interval. However, when the frequency of the incoming wave is time varying in a random manner, the implementation of a maximum likelihood approach may be computationally demanding, as discussed in [3], [4].

The object of this article is to propose and analyze a suboptimum frequency control loop based on short overlapping DFTs and suited to track the rapidly varying frequency of a sine wave. For brevity, the new scheme will be referred to here as ODAFC (Overlapping Discrete Fourier transform-based Automatic Frequency Control). This scheme can be thought of as a generalization of the so-called Quadri-Correlator [5]. The new scheme integrates ideas originating in the fields of classical spectral estimation and digital phase-locked loop theory. It will be shown that the ODAFC is capable of track-

ing the frequency of a carrier signal in a low carrier-to-noise (CNR) and high dynamic environment where a Phase-Locked Loop (PLL) is inoperative. Such a situation is foreseeable during planetary encounters. In these cases ODAFC will allow fast carrier re-acquisition and help the resumption of tracking itself. This is achieved with remarkably modest implementation complexity, since the algorithm may be implemented in software and does not require any modification to the Advanced Receiver (ARX) hardware.

First, as an indicator of the quality of the closed loop estimator, the variance of the loop is computed for a fixed frequency sine wave in the presence of additive white Gaussian noise (AWGN). Later, the loop is tested with severe dynamics and noise together, and the probability of loss of lock is estimated. Both analytical answers (whenever possible) and computer simulation results are included for greater confidence.

II. Description of the ODAFC Loop

A. System Block Diagram

A block diagram of the ODAFC loop is shown in Fig. 1(a). An equivalent baseband model is shown in Fig. 1(b). The discrete time in-phase and quadrature mixer outputs admit a representation of the form

$$I_n = A \cos(\phi_n) + n_{In} \quad (1a)$$

$$Q_n = A \sin(\phi_n) + n_{Qn} \quad (1b)$$

The amplitude A is related to the carrier power via $A^2 = P_c$, the instantaneous phase error is given by ϕ_n (rad), and the noise sequences $\{n_{In}\}$, $\{n_{Qn}\}$ are independent Gaussian random variables with zero mean and variance $\sigma^2 = N_o/2T_s$, where T_s is the loop update time (henceforth called the sampling interval). The carrier-to-noise ratio is defined as $\text{CNR} = P_c/N_o$. Every update time, N_s complex samples are Fourier transformed with N_s zeroes appended. At the tick of the clock, the data are shifted when another complex sample arrives; a new discrete Fourier transform is computed (of size $2N_s$), an error control signal is created and the loop filter is updated. This process is repeated every sampling interval T_s , each time incorporating just one more complex sample. A timing diagram of the underlying process is illustrated in Fig. 2.

B. On the Discriminator Characteristic

It is well known that the discrete Fourier transform can be seen as a bank of bandpass filters tuned at multiples of half the Nyquist rate [6]. In order to create a control signal proportional to the frequency error (discriminator characteristic), the powers in two adjacent filter bins around zero frequency

are subtracted every sampling interval (a weighted combination of the filter bin outputs was not selected to facilitate the implementation). The subtraction of the powers in the bin filters around zero frequency to create a discriminator curve as opposed to selecting the bin filter with maximum power was originally discussed in [7]. The idea of employing running overlapping DFTs has a connection with the use of the FFT for the estimation of power spectra based on averaged periodograms [8].

III. Linear Tracking of the ODAFC Loop in the Presence of Noise

A. Exact Computation of the Variance of the Frequency Error

This section focuses on the tracking performance of the ODAFC loop when the carrier-to-noise ratio is sufficiently high to justify a linear analysis. A sine wave with constant frequency is assumed for the remainder of the section. The goal is to compute analytically an expression that predicts the variance of the closed loop estimator. The results will be confirmed by computer simulations.

The closed loop estimator uses the received complex sample

$$x_n = I_n + j Q_n \quad (2)$$

and computes

$$\begin{aligned} X_{k,q} &= \frac{1}{N_s} \sum_{n=k-(N_s-1)}^k x_n e^{-j2\pi q n/2N_s} \\ &\triangleq R_{k,q} + j M_{k,q} \end{aligned} \quad (3)$$

The discriminator output at time $n = k$ is given by

$$\begin{aligned} P_k &= R_{k,1}^2 + M_{k,1}^2 - (R_{k,-1}^2 + M_{k,-1}^2) \\ &= \left(\frac{1}{N_s}\right)^2 \sum_{m=k-(N_s-1)}^k \sum_{n=k-(N_s-1)}^k 2(I_n Q_m - Q_n I_m) \\ &\quad \times \sin\left(\frac{\pi}{N_s}(m-n)\right) \end{aligned} \quad (4)$$

which upon expansion results in

$$P_k \triangleq S_k + N_{eq,k} \quad (5)$$

where

$$S_k = 2 \left(\frac{1}{N_s} \right)^2 \sum_{m=k-(N_s-1)}^k \sum_{n=k-(N_s-1)}^k A^2 \sin(\phi_m - \phi_n) \times \sin \left(\frac{\pi}{N_s} (m - n) \right) \quad (6)$$

$$N_{eq,k} = 2 \left(\frac{1}{N_s} \right)^2 \sum_{m=k-(N_s-1)}^k \sum_{n=k-(N_s-1)}^k \left\{ A n_{Qm} \cos(\phi_n) + A n_{Im} \sin(\phi_m) - A n_{Im} \sin(\phi_n) - A n_{Qn} \cos(\phi_m) + n_{Im} n_{Qm} - n_{Qn} n_{Im} \right\} \times \sin \left(\frac{\pi}{N_s} (m - n) \right) \quad (7)$$

for a fixed frequency error; $S_k(\cdot)$ denotes the discriminator characteristic for the ODAFC loop, and it is given more explicitly in Eq. (8).

$$S_k(\Delta\omega_k T_s) = A^2 \left\{ \frac{\left[\sin \left(\frac{\Delta\omega_k N_s T_s}{2} - \frac{\pi}{2} \right) \right]^2}{\left[\sin \left(\frac{\Delta\omega_k T_s}{2} - \frac{\pi}{2N_s} \right) \right]^2} - \frac{\left[\sin \left(\frac{\Delta\omega_k N_s T_s}{2} + \frac{\pi}{2} \right) \right]^2}{\left[\sin \left(\frac{\Delta\omega_k T_s}{2} + \frac{\pi}{2N_s} \right) \right]^2} \right\} \quad (8)$$

The slope at the origin is

$$S'_k(\Delta\omega_k T_s) \Big|_{\Delta\omega_k T_s=0} = A^2 \left(\frac{1}{N_s} \right)^2 \left\{ \frac{8\pi^2 \cos \left(\frac{\pi}{2N_s} \right)}{\sin^3 \left(\frac{\pi}{2N_s} \right)} \right\} \quad (9)$$

The discriminator characteristic is illustrated in Fig. 3 for a unit amplitude signal ($A = 1$) for $N_s = 2, 4$ corresponding to DFTs of sizes 4 and 8, respectively; notice that a larger slope for the discriminator at the origin is obtained with larger DFTs. For a fixed frequency sine wave in the presence of AWGN, the larger the size of the DFT, the smaller the variance of the estimator. However, in a dynamic medium, the maximum permissible size for the DFT is limited by how fast the

center frequency is moving, since larger DFTs have a smaller linear operating range.

The term $N_{eq,k}$ will be referred to as the additive equivalent noise. Unfortunately, the typical white noise assumption for the equivalent noise is not justified here, since there exists a significant amount of correlation between the noise samples. It is shown in Appendix A that the correlation function of the noise sequence $\{N_{eq,k}\}$ is given by

$$R_{N_{eq}}(\ell) = \mathbf{E}(N_{eq,k} N_{eq,k+\ell}) = 8 \left(\frac{1}{N_s} \right)^4 A^2 \sigma^2 \frac{(N_s - |\ell|) \cos \left(\frac{\pi \ell}{N_s} \right)}{\left[\sin \left(\frac{\pi}{2N_s} \right) \right]^2} - 8 \left(\frac{1}{N_s} \right)^4 A^2 \sigma^2 \frac{\sin \left(\frac{\pi |\ell|}{N_s} \right)}{\left[\sin \left(\frac{\pi}{2N_s} \right) \right]^2 \sin \left(\frac{\pi}{N_s} \right)} + 8 \left(\frac{1}{N_s} \right)^4 \sigma^4 (N_s - |\ell|)^2 - 8 \left(\frac{1}{N_s} \right)^4 \sigma^4 \left[\frac{\sin \left(\frac{\pi \ell}{N_s} \right)}{\sin \left(\frac{\pi}{N_s} \right)} \right]^2 \quad (10)$$

$$0 \leq |\ell| \leq N_s$$

otherwise

$$R_{N_{eq}}(\ell) = 0$$

With the aid of Fig. 1, it is a straightforward exercise to show that in operational notation the Z transform of the normalized radian frequency error is given by

$$\Delta\Omega(z) T_s = H(z) \Omega(z) T_s - \frac{H(z)}{S'_k(0)} N_{eq}(z) \quad (11)$$

where

$$H(z) \triangleq \frac{S'_k(0) F_2(z) NCO(z) k_N}{1 + S'_k(0) F_2(z) NCO(z) k_N} \quad (12)$$

represents the closed loop transfer function. The loop filter $F(z)$ considered here has the form

$$F_1(z) = \frac{F_2(z)}{1 - z^{-1}} = \left[\frac{k_1}{1 - z^{-1}} + \frac{k_2}{(1 - z^{-1})^2} \right] \frac{1}{S'_k(0) k_N T_s} \quad (13)$$

where

$$k_1 = \frac{r(4 B_A T_s)}{r + 1}$$

$$k_2 = \frac{k_1^2}{r}$$

$$r = 4 \xi^2$$

ξ = damping ratio

B_A = nominal bandwidth (Hz)

and $NCO(z)$ is the mathematical model of the numerically controlled oscillator and may include computational delays inherent in a digital implementation. Of particular interest is the NCO discussed in [9], which conceptually represents the cascading of an integrator using the trapezoidal rule and two extra delays.

$$NCO(z) = \frac{T_s(z+1)}{2z^2(z-1)} \quad (14)$$

Since the noise sequence $\{N_{eq,k}\}$ is stationary, the steady state variance of the frequency error (Hz) is given by [10], [11] as

$$\sigma_{\Delta f}^2 = \frac{1}{[S'_k(0)]^2} \left(\frac{1}{2\pi T_s} \right)^2 \frac{1}{2\pi} \int_{-\pi}^{\pi} |H(e^{j\omega})|^2 S_{N_{eq}}(e^{j\omega}) d\omega \quad (15)$$

where the equivalent noise spectral density has the form

$$S_{N_{eq}}(e^{j\omega}) = R_{N_{eq}}(0) + 2 \sum_{\ell=1}^{N_s-1} R_{N_{eq}}(\ell) \cos(\ell\omega) \quad (16)$$

Unfortunately, the previous integral does not reduce to a simple compact expression as in a standard phase-locked loop [5], and a numerical integration must be used. For illustration purposes, Fig. 4 shows numerical results, based on Eq. (15), for $N_s = 4$ and for a particular mechanization using $T_s = 2 \times 10^{-3}$ s with nominal bandwidth B_A (Hz) as a parameter. Computer simulation results are also included in the same graph to confirm the analysis. Notice that for high CNR, the analysis and the simulations are in excellent agreement.

B. A Useful Approximation for the Variance of the Frequency Error

A very useful approximation to the variance of the frequency error can be obtained for low to moderate CNR when the bandwidth of the loop is much smaller than the sampling rate. The approximation is obtained by expanding the cosinusoidal terms of the noise spectral density of Eq. (16) in a Taylor series and keeping only the significant terms of the expansion. The result is

$$\sigma_{\Delta f}^2 \approx \left(\frac{2}{N_s} \right)^2 \left(\frac{1}{2\pi} \right)^3 \left(\frac{1}{T_s} \right)^2 \left[a \Omega_L T_s - \frac{b}{3} (\Omega_L T_s)^3 \right] \quad (17)$$

where

$$a = R_{N_{eq}}(0) + 2 \sum_{\ell=1}^{N_s-1} R_{N_{eq}}(\ell) \quad (18)$$

$$b = \sum_{\ell=1}^{N_s-1} \ell^2 R_{N_{eq}}(\ell) \quad (19)$$

$$\Omega_L = 2\pi(2B_L)$$

B_L = one-sided loop noise bandwidth

The approximation is very tight for small bandwidths ($B_L \lesssim 10$ Hz) and for low CNRs ($\lesssim 30$ dB-Hz), and it loosens as these two parameters increase, until it becomes useless for $\text{CNR} \gtrsim 50$ dB-Hz, as shown in Fig. 4.

IV. Steady State Frequency Errors

The steady state frequency error due to acceleration in frequency can be calculated from Eq. (11) in the absence of noise using the final value theorem [12]. The input frequency in the Z domain is given by

$$\Omega(z) = z \left\{ \frac{1}{2} J_o t^2 \right\}$$

and the steady state error is

$$\Delta f_{ss}(\text{Hz}) = \lim_{z \rightarrow 1} \left(\frac{z-1}{z} \right) H(z) \Omega(z)$$

$$= \frac{J_o (r+1)^2}{16 B_A^2 r} \cdot \frac{1}{2\pi}$$

where the units of the jerk J_o are in rad/s^3 .

V. The ODAFC Loop in a Dynamic Environment

One of the main reasons for this study was the need to operate the frequency estimator under severe dynamic conditions at low CNR levels. It is natural to select the probability of loss of frequency lock as a measure of quality. Once cast into this form, computer simulations provide the most direct way to evaluate performance.

A. Dynamic Trajectory

A dynamic trajectory of duration 8 s was developed to evaluate loop performance. Initially, it is assumed that the phase, frequency, and frequency rate of the received signal are perfectly known, i.e., that the loop is in the tracking mode of operation. The transmitting vehicle then experiences a sudden maneuver resulting in an acceleration of 3 s duration and equivalent to a doppler rate of -1287 Hz/s . For the next 0.5 s, a frequency acceleration of 5150 Hz/s^2 occurs, followed by a 2 s duration doppler rate of 1287 Hz/s . A frequency acceleration of -5150 Hz/s^2 proceeds for the next 0.5 s, and the maneuver culminates with a 2-s-duration doppler rate of -1287 Hz/s .

B. Probability of Loss of Lock for the ODAFC Loop

The probability of losing frequency lock was estimated for a loop that is initially in-lock and then experiences the phase trajectory previously described. The simulations were carried out for various sizes of DFTs, but $N_s = 4$ was found to have the lowest possible threshold with this technique. Some simulation results are presented in Figs. 5 and 6 for $N_s = 4$ and 8, respectively, as a function of carrier to noise ratio (dB-Hz) with nominal loop bandwidth as a parameter. For a given CNR and nominal bandwidth, 250 different simulation runs were employed. Loss of lock was declared when the instantaneous frequency error exceeded the one-sided Nyquist bandwidth $1/2T_s$.

In order to put these results in proper perspective, from [4] a type-3 digital phase-locked loop (DPLL) with the same phase trajectory and an optimized bandwidth, has a probability of 0.1 of losing lock at approximately 26 dB-Hz; i.e., the threshold for a DPLL is about 3 dB higher than that of the ODAFC with $N_s = 4$. From the same reference, the threshold of a typical Cross-Product AFC loop is 2.3 dB-Hz higher than that of the ODAFC.

Estimation of the probability of loss of lock is only a partial performance measure. If in-lock, a phase-locked loop has typically a better frequency estimation ability than the ODAFC loop. In Figs. 7 and 8, the rms value for the ODAFC loop is estimated for the aforementioned trajectory. Notice that when $N_s = 8$, smaller tracking errors are achieved, but the threshold is higher. The same effect is noticed for larger values of N_s . This scheme then suffers from the usual trade-off between noise suppression and dynamic tracking: the effect of the noise increases as N_s decreases, while smaller dynamic errors are obtained as N_s decreases.

VI. Conclusions

A new automatic frequency control loop was introduced and discussed. The algorithm (OD AFC) is based on running overlapping DFTs, and can be viewed as a generalization of the well-known Cross-Product AFC (CPAFC) loop [7]. In fact, the CPAFC is a special case of the ODAFC loop when $N_s = 2$.

A detailed noise analysis was presented of the new algorithm for any desired DFT size, and the analysis supported with computer simulations. In addition, the algorithm was tested in a noisy dynamic medium, and the probability was estimated of loss of lock for various configurations for a specific phase trajectory. It was shown in the text that for this particular trajectory, a value $N_s = 4$ minimizes the operating threshold for this algorithm. In fact, the ODAFC with $N_s = 4$ has a threshold 3 dB-Hz lower than that of a DPLL and comparable to that of a maximum likelihood estimator [4], but with a larger error variance. The implementation complexity for the ODAFC is substantially simpler than for an optimum strategy, not even requiring fast Fourier transforms to compute the DFT when N_s is small.

In summary, the algorithm discussed in this article is a suboptimum tracking scheme with increased variance for the frequency error compared to an optimum strategy, but offers simplicity of implementation and a very low threshold.

Acknowledgments

The author wishes to thank Joseph I. Statman for providing both the dynamic trajectory used to test the loop and for helpful comments. Useful discussions with Sami Hinedi and Vic Vilmrotter are also appreciated.

References

- [1] S. L. Marple, Jr., *Digital Spectral Analysis with Applications*, Englewood Cliffs, New Jersey: Prentice-Hall, 1987.
- [2] M. S. Kay, *Modern Spectral Estimation*, Englewood Cliffs, New Jersey: Prentice-Hall, 1988.
- [3] W. J. Hurd, J. I. Statman, and V. A. Vilmrotter, "High Dynamic GPS Receiver Using Maximum Likelihood Estimation and Frequency Tracking," *IEEE Trans. on Aerosp. and Electr. Syst.*, vol. AES-23, no. 4, pp. 425-437, July 1987.
- [4] V. A. Vilmrotter, S. Hinedi, and R. Kumar, *A Comparison of Frequency Estimation Techniques for High Dynamic Trajectories*, JPL Publication 88-21, to be published.
- [5] F. M. Gardner, *Phase Lock Techniques*, 2nd ed., New York: John Wiley and Sons, 1979.
- [6] G. D. Bergland, "A Guided Tour of the Fast Fourier Transform," *IEEE Spectrum*, vol. 6, pp. 41-52, July 1969.
- [7] F. D. Natali, "AFC Tracking Algorithms," *IEEE Trans. on Commun.*, vol. COM-32, no. 8, pp. 935-947, August 1984.
- [8] P. D. Welch, "The Use of the Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms," *IEEE Trans. Audio Electroacoust.*, vol. AU-15, pp. 70-73, June 1967.
- [9] S. Aguirre, W. J. Hurd, R. Kumar, and J. I. Statman, "A Comparison of Methods for DPLL Loop Filter Design," *TDA Progress Report 42-87*, vol. July-September 1986, Jet Propulsion Laboratory, Pasadena, California, pp. 114-124, November 15, 1986.
- [10] A. V. Oppenheim and R. W. Schaffer, *Digital Signal Processing*, Englewood Cliffs, New Jersey: Prentice-Hall, 1975.
- [11] W. C. Lindsey and C. M. Chie, "A Survey of Digital Phase-Locked Loops," *Proceedings of the IEEE*, vol. 69, no. 4, pp. 410-430, April 1981.
- [12] C. L. Phillips and H. T. Nagle, Jr., *Digital Control System Analysis and Design*, Englewood Cliffs, New Jersey: Prentice-Hall, 1984.

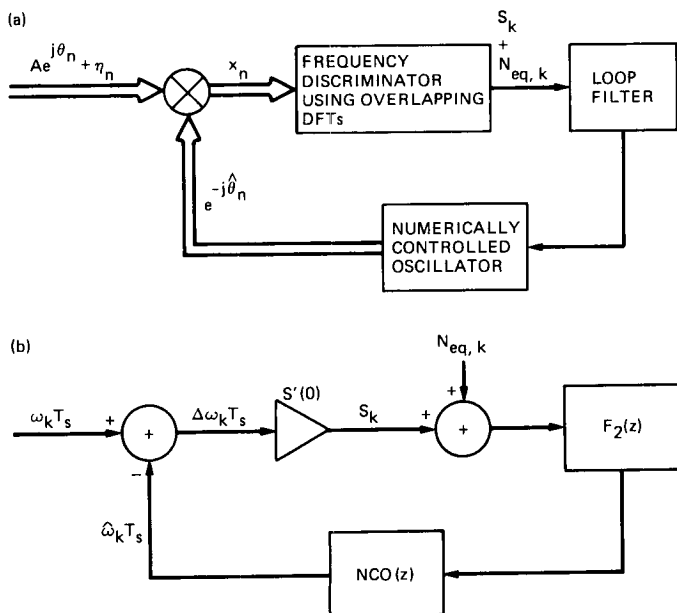


Fig. 1. Two loops: (a) an automatic frequency control loop using overlapping DFTs and (b) an equivalent linearized baseband model

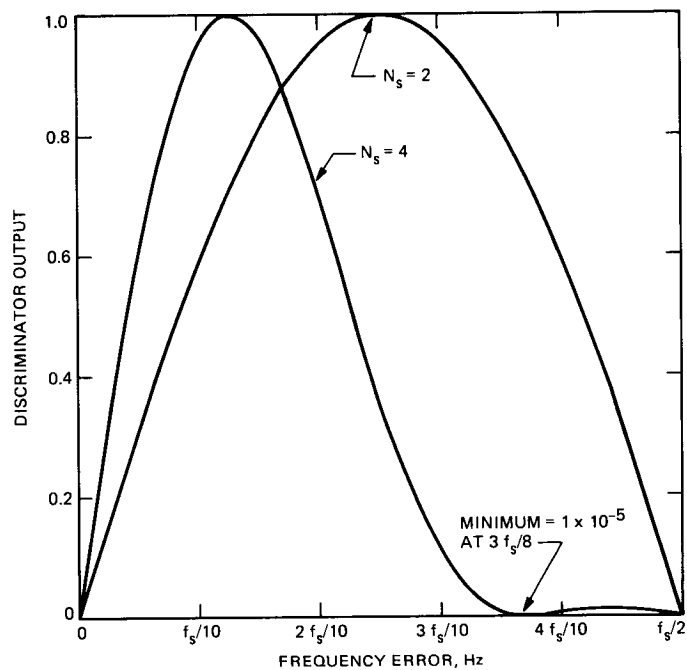


Fig. 3. Normalized discriminator curves for ODAFC loop (only positive frequency errors are shown)

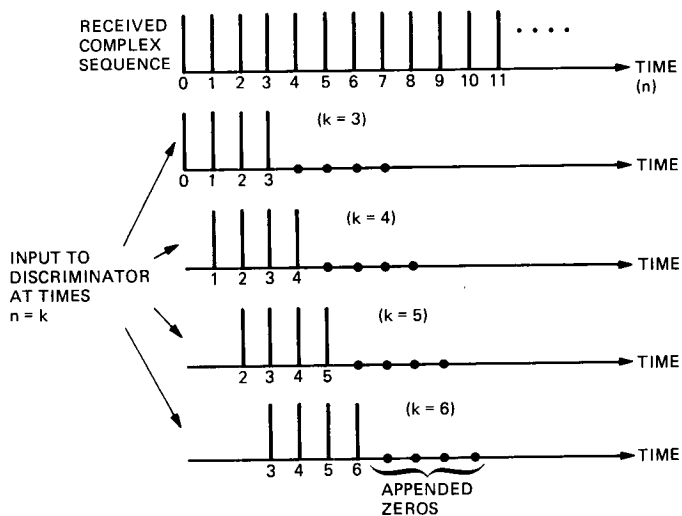


Fig. 2. Overlapping of the received signals to create a discriminator characteristic

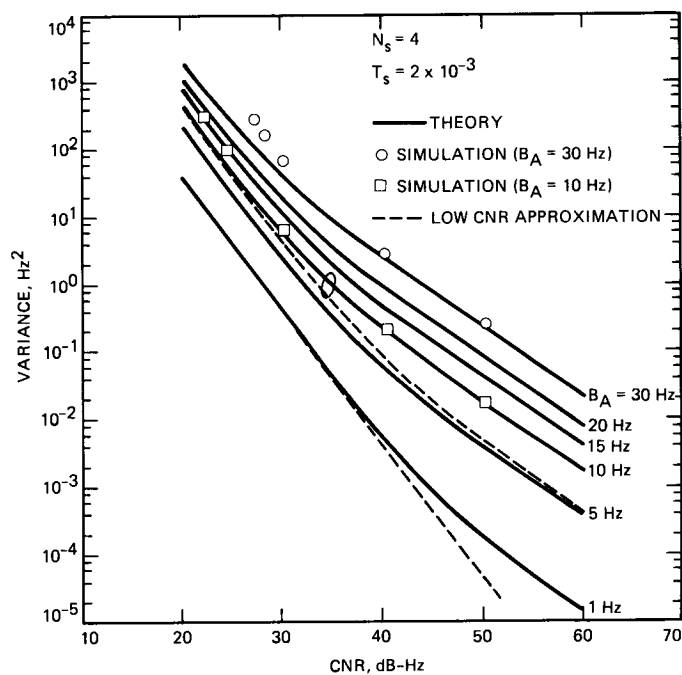


Fig. 4. Tracking performance of the ODAFC loop ($N_s = 4$) as a function of CNR with nominal bandwidth as a parameter

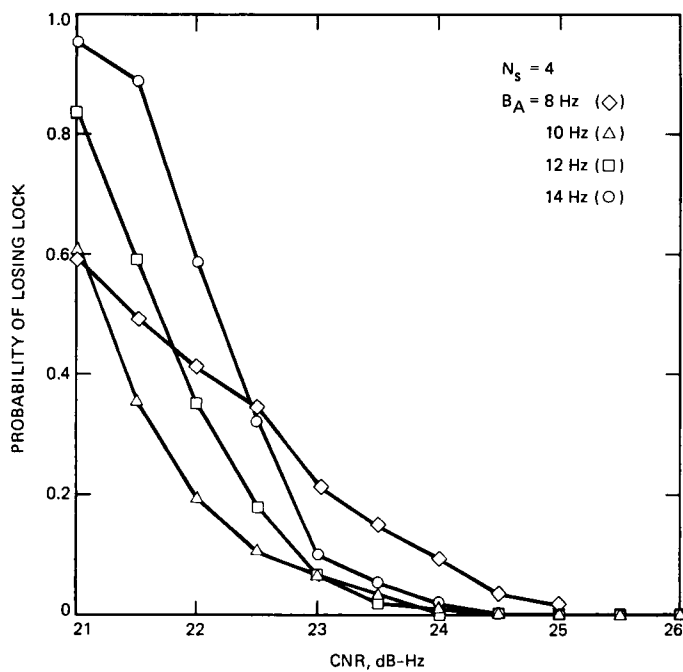


Fig. 5. Probability of losing frequency lock in a dynamic medium as a function of CNR with nominal bandwidth as a parameter ($N_s = 4$)

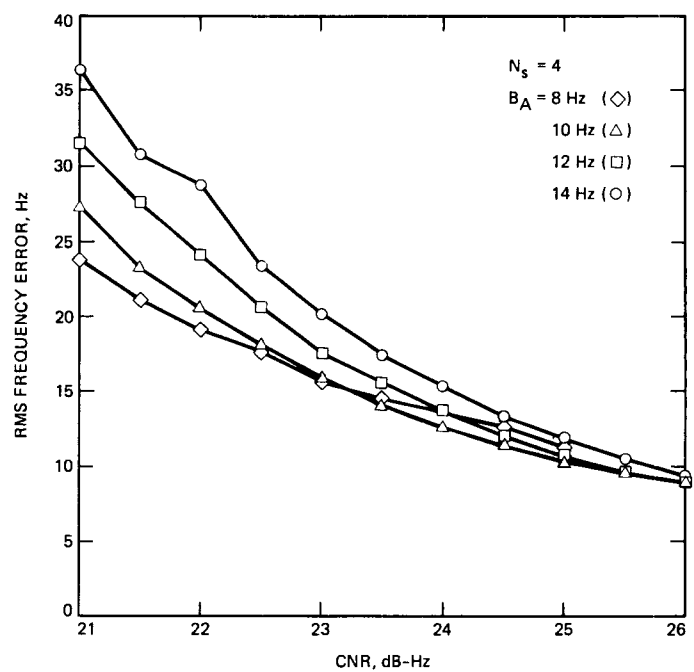


Fig. 7. Tracking performance of the ODAFC loop in a dynamic medium ($N_s = 4$)

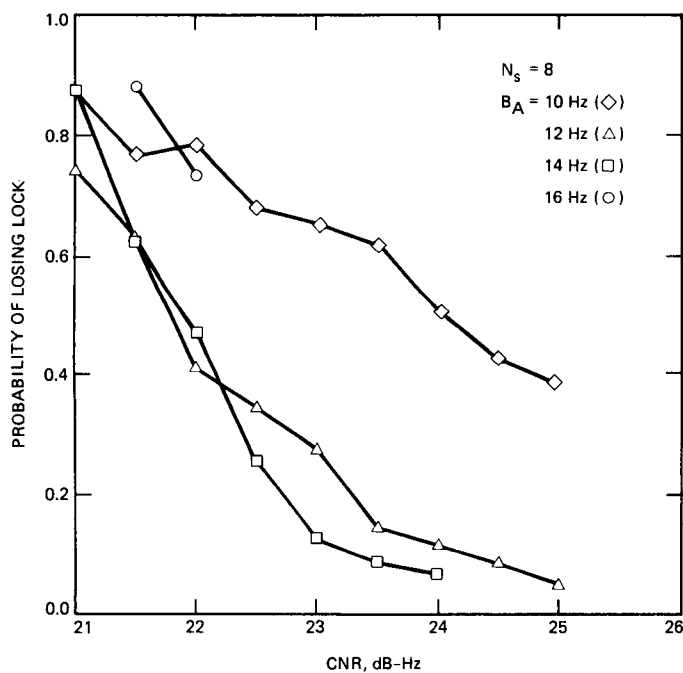


Fig. 6. Probability of losing frequency lock in a dynamic medium as a function of CNR with nominal bandwidth as a parameter ($N_s = 8$)

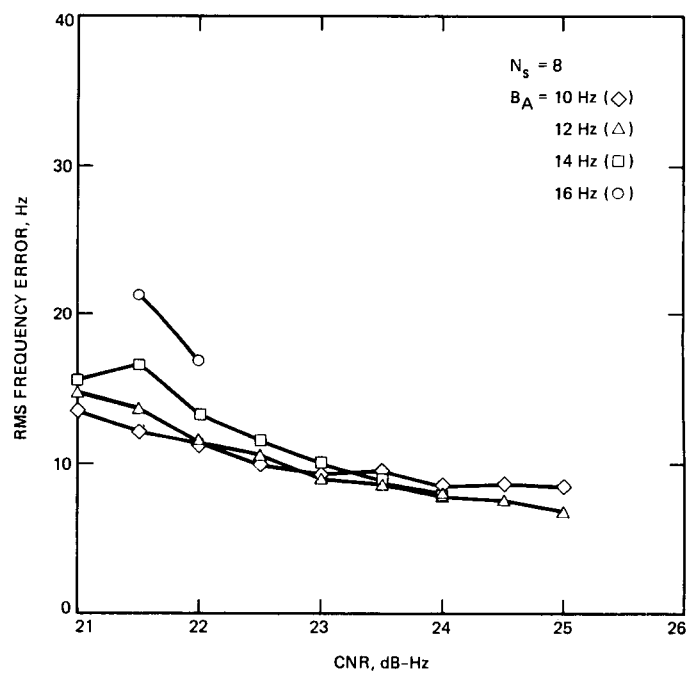


Fig. 8. Tracking performance of the ODAFC loop in a dynamic medium ($N_s = 8$)

Appendix A

Correlation Function of the Equivalent Additive Noise for the ODAFC Loop

In this appendix, the autocorrelation function of the equivalent additive noise for the ODAFC loop is computed for any lag. The starting point is Eq. (7). The autocorrelation is given by

$$\begin{aligned}
 R_{N_{eq}}(\ell) &= \mathbb{E}(N_{eq,k} N_{eq,k+\ell}) \\
 &= 4 \left(\frac{1}{N_s} \right)^4 \sum_{m=k}^k \sum_{n=k}^k \sum_{i=k+\ell}^{k+\ell} \sum_{j=k+\ell}^{k+\ell} \left\{ A^2 \sigma^2 \delta_{mi} \right. \\
 &\quad \times \cos(\phi_n - \phi_j) - A^2 \sigma^2 \delta_{mj} \cos(\phi_n - \phi_i) \\
 &\quad - A^2 \sigma^2 \delta_{ni} \cos(\phi_m - \phi_j) + A^2 \sigma^2 \delta_{nj} \cos(\phi_m - \phi_i) \\
 &\quad \left. + 2 \sigma^4 \delta_{nj} \delta_{mi} - 2 \sigma^4 \delta_{mj} \delta_{ni} \right\} C_{mn} C_{ij} \quad (A-1)
 \end{aligned}$$

where

$$\delta_{xy} = \begin{cases} 1 & x = y \\ 0 & x \neq y \end{cases} \quad (A-2)$$

$$C_{xy} = \sin\left(\frac{\pi}{N_s}(x-y)\right) \quad (A-3)$$

For linear analysis it is useful to make the approximation $\cos(\phi_x - \phi_y) = 1$ in Eq. (A-1); therefore, the correlation function becomes

$$\begin{aligned}
 R_{N_{eq}}(\ell) &\approx 4 \left(\frac{1}{N_s} \right)^4 \sum_{m=0}^{N_s-1} \sum_{n=0}^{N_s-1} \sum_{i=\ell}^{N_s-1+\ell} \sum_{j=\ell}^{N_s-1+\ell} \left\{ A^2 \sigma^2 \delta_{mi} \right. \\
 &\quad - A^2 \sigma^2 \delta_{mj} - A^2 \sigma^2 \delta_{ni} + A^2 \sigma^2 \delta_{nj} \\
 &\quad \left. + 2 \sigma^4 \delta_{mi} \delta_{nj} - 2 \sigma^4 \delta_{mj} \delta_{ni} \right\} C_{mn} C_{ij} \quad (A-4)
 \end{aligned}$$

Due to symmetry, the previous correlation reduces to

$$\begin{aligned}
 R_{N_{eq}}(\ell) &\approx 16 \left(\frac{1}{N_s} \right)^4 A^2 \sigma^2 \sum_{m=0}^{N_s-1} \sum_{n=0}^{N_s-1} \sum_{i=\ell}^{N_s-1+\ell} \sum_{j=\ell}^{N_s-1+\ell} \delta_{mi} C_{mn} \\
 &\quad \times C_{ij} + 16 \left(\frac{1}{N_s} \right)^4 \sigma^4 \sum_{m=0}^{N_s-1} \sum_{n=0}^{N_s-1} \sum_{i=\ell}^{N_s-1+\ell} \sum_{j=\ell}^{N_s-1+\ell} \delta_{mi} \\
 &\quad \times \delta_{nj} C_{mn} C_{ij} \quad (A-5)
 \end{aligned}$$

or, using short notation:

$$\begin{aligned}
 R_{N_{eq}}(\ell) &\approx C_1 f_{11} - C_1 f_{12} \\
 &\quad + C_2 f_{21} - C_2 f_{22} \quad (A-6)
 \end{aligned}$$

where

$$C_1 \triangleq 8 \left(\frac{1}{N_s} \right)^4 A^2 \sigma^2 \quad (A-7)$$

$$C_2 \triangleq 8 \left(\frac{1}{N_s} \right)^4 \sigma^4$$

and

$$\begin{aligned}
 f_{11} &= \sum \sum \sum \sum \delta_{mi} \cos\left[\frac{\pi}{N_s}(m-n+j-i)\right] \\
 f_{12} &= \sum \sum \sum \sum \delta_{mi} \cos\left[\frac{\pi}{N_s}(m-n+i-j)\right] \\
 f_{21} &= \sum \sum \sum \sum \delta_{mi} \delta_{nj} \cos\left[\frac{\pi}{N_s}(m-n+j-i)\right] \\
 f_{22} &= \sum \sum \sum \sum \delta_{mi} \delta_{nj} \cos\left[\frac{\pi}{N_s}(m-n+i-j)\right] \quad (A-8)
 \end{aligned}$$

519-34
~~165630~~
TDA Progress Report 42-94
+28.

1989000834
618443
14p

N89 - 10205

April-June 1988

JPV

Heat Exchanger Expert System Logic

R. Cormier

Radio Frequency and Microwave Subsystems Section

This article describes the reduction of the operation and fault diagnostics of a DSN heat exchanger to a rule base by the application of propositional calculus to a set of logic statements. The value of this approach lies in the ease of converting the logic and subsequently implementing it on a computer as an "expert system." The rule base was written in Process Intelligent Control (PICON) software.

I. Introduction

The application of formal systems analysis and propositional calculus has great value in the design of an "expert system" rule base, where the need is to model knowledge. Propositional calculus is a formal logic system from which conclusions are deduced independent of the structure of the propositions. The conclusions follow uniquely from the truth or falsity of the propositions. Once a knowledge base has been converted to propositions, assuming it has been done correctly, the process of deduction cannot lead to contradictions or to invalid conclusions. One of the consequences is that, given an appropriate model, every deduction is a tautology. A description of formal systems and propositional calculus can be found in [1]. The breadboard heat exchanger expert system is described in [2].

The task described in this article involved the conversion of the controls and operation of a DSN heat exchanger to a set of logic statements for an "expert system" computer operating system [2]. The heat exchanger elements and the sensors that were converted to logic statements are listed in Table 1. A flow diagram (Fig. 1) indicates the position of the sensors in the piping. The electrical schematic is shown in Fig. 2.

The controls and operation were first described in terms of hardware for all conditions of service. For example, the heat exchanger has dual pumps which are fully redundant. The failure of one pump does not affect the capability of the heat exchanger to cool the transmitter, except for the reliability factor. Two fans are provided, but these are only partially redundant. Under certain conditions, both fans are required. In addition, a number of other failures can be tolerated provided the heat exchanger is operated locally. The Pumps Alternating command can be bypassed with a local switch, in which case both pumps may be operated simultaneously. All of these conditions were described and incorporated in the rule base, as the expert system needed the capability to distinguish between and not be confused by these alternatives.

II. Operational Logic Statements

The word description of the heat exchanger operation was then converted to logic statements or schemata. The symbols used for the elements are defined in Table 1. The operand symbols are defined in the Appendix.

For simplicity, the limits in the summations have been omitted, but are equal to those of Eq. (A-3). After considerable algebra, the functions in Eq. (A-6) reduce to

$$f_{11} = \frac{(N_s - |\ell|) \cos \left[\frac{\pi}{N_s} \ell \right]}{\left[\sin \left(\frac{\pi}{2N_s} \right) \right]^2} \quad 0 \leq |\ell| \leq N_s \quad (\text{A-9})$$

$$f_{12} = \frac{\sin \left(\frac{\pi |\ell|}{N_s} \right)}{\sin \left(\frac{\pi}{N_s} \right)} \left[\frac{1}{\sin \left(\frac{\pi}{2N_s} \right)} \right]^2 \quad 0 \leq |\ell| \leq N_s \quad (\text{A-10})$$

$$f_{21} = (N_s - |\ell|)^2 \quad 0 \leq |\ell| \leq N_s \quad (\text{A-11})$$

$$f_{22} = \left[\frac{\sin \left(\frac{\pi \ell}{N_s} \right)}{\sin \left(\frac{\pi}{N_s} \right)} \right]^2 \quad 0 \leq |\ell| \leq N_s \quad (\text{A-12})$$

After collecting the terms in Eqs. (A-9)–(A-12) and using Eq. (A-6), the correlation function shown in the main text is obtained.

The heat exchanger is operational whenever it is providing cooling for the transmitter. Coolant must be flowing and heat must be removed from the coolant. This implies that at least one pump and at least one fan are operating, with the second fan available if required for additional heat transfer. The fan operation depends on the coolant temperature at the output of the core. The two conditions are (a) for a coolant temperature $<120^{\circ}\text{F}$; and (b) for a coolant temperature $>110^{\circ}\text{F}$.

Let

a denote condition a is true

b denote condition b is true

\bar{a} denote condition a is false

\bar{b} denote condition b is false.

Given this notation, it can be stated that the heat exchanger is operational when

$$v \vee w : \bar{a} \rightarrow xy : \bar{b} \rightarrow .xy \vee \bar{xy} : ab \rightarrow .x \vee y$$

The use of dots in the above statement indicates grouping as referenced in the appendix. This states that (1) for the heat exchanger to be operational, at least one pump (v or w) is on; and (2) if the coolant temperature is not less than 120°F , then two fans (x and y) are operating, if the coolant temperature is not greater than 110°F , then at least one fan is on but not both, and if the coolant temperature is between 110°F and 120°F , then either fan 1 is on or fan 2 is on or both. The advantage of schemata can be appreciated when it is desired to check the consequences of a statement like the one above. The heat exchanger is not operational when the negative of the above is made by the use of DeMorgan's law:

$$\bar{v}\bar{w} \vee : \bar{a} . \bar{x} \vee \bar{y} : \bar{b} : \bar{x} \vee y . x : \bar{y} : \bar{v} ab \bar{xy}$$

This states that for the heat exchanger to be not operational, both pumps (v and w) are off or the coolant temperature is not less than 120°F , and either fan 1 (x) is off or fan 2 (y) is off; or the coolant temperature is not greater than 110°F and either fan 1 is off or fan 2 is on, or vice versa; or the coolant temperature is between 110°F and 120°F and both fans are off.

In the process of taking the inverse, the "if-then" statements of the original schema were eliminated and only conjunctions and alternations remained. This form is known as the "alternative normal form." Using this form, it is simple to remove inconsistencies. If the schema was found to be valid, i.e., true under all conditions of its terms, then this would mean that the original schema before inversion was incon-

sistent. This is a useful check, and errors were found and eliminated by this method. Although the above may be logically true, from a practical point of view the heat exchanger would not be considered not operational if both fans were on with a coolant temperature of less than 110°F . In this case, the failure does not degrade the cooling capacity of the heat exchanger. This is an example of the necessity of analyzing each schema and its inverse. It is not always obvious what the consequences are, even for rather simple definitions.

Several assumptions and simplifications were made. It was assumed that the heat exchanger is commanded on either from the transmitter or from the local panel and that prime power is available. It was also assumed that with one pump operating, 20 gpm of coolant with a pressure differential of 125 psi is being supplied. Similarly, it was assumed that the operation of the fans moves a quantity of air adequate to maintain the coolant exhaust temperature at the transmitter output at less than 70°C . Obviously, however, these conditions are not always necessarily true; they are also addressed in Section III.

The heat exchanger is controlled from a three-position switch located on a panel on the heat exchanger. In the Local position (c_l), power is applied; in the Off position (c_o), no power is applied; and in the Remote position (c_r), power is applied if the transmitter control power (p_1) is on. With respect to the REMOTE/OFF/LOCAL switch S2, and again using the notation of the appendix, the heat exchanger is operational when

$$c_r p_1 \vee c_l \leftarrow \rightarrow .v \vee w . x \vee y$$

Associated with the pumps are time delay relays TDR1 and TDR2 and pressure switches S4 and S5. When the heat exchanger is turned on, the time delay relays inhibit the pressure switches during the time required for the motors to reach operating speed. If a pump does not deliver operating coolant pressure by the time the time delay relay has timed out, that pump will be turned off and the other backup pump will be turned on. This provides redundancy in case of a pump failure. The time delays of relays TDR1 and TDR2 are designated r_1 and r_2 , and the functions of pressure switches S4 and S5 are designated s_4 and s_5 .

In addition, there are two alternating relays for the pumps and the fans K9 and K14, respectively. To simplify which pump and which fan are being turned on at any given time, the times are designated t_1 for pump 1 and fan 1, and t_2 for pump 2 and fan 2. After t_2 , the time starts at t_1 again.

The hardware conditions for the operation of pump 1 are as follows:

- (1) The control circuit breaker (b_c) is closed.

- (2) The main circuit breaker (b_7) is closed.
- (3) The pump 1 circuit breaker (b_1) is closed.
- (4) The pump 1 contactor (k_1) is closed.
- (5) The tank low-level switch (s_3) is closed.
- (6) The 480-Vac three-phase power (p_2) is available.
- (7) The pump 1 thermal overload (o_1) is closed.

Similarly, for pump 2:

- (1) The control circuit breaker (b_6) is closed.
- (2) The main circuit breaker (b_7) is closed.
- (3) The pump 2 circuit breaker (b_2) is closed.
- (4) The pump 2 contactor (k_2) is closed.
- (5) The tank low-level switch (s_3) is closed.
- (6) The 480-Vac three-phase power (p_2) is available.
- (7) The pump 2 thermal overload (o_2) is closed.

The control switch is in either the Local or the Remote position. If it is in Remote, the 28-Vdc control power (p_1) is on from the transmitter and the tank low-pressure cutoff switch (s_1) is closed. In addition, either the alternating relay is in Pump 1 position (and when pump 1 was turned on, it reached a pressure of 60 psig before the pump 1 time delay relay closed) or pump 2 was turned on (and the alternating relay for pump 2 timed out before the pressure reached 60 psig or the pressure dropped while pump 2 was operating). This can be written in a condensed form using the dot notation in the appendix (with multiple dots representing a greater break) as follows:

For pump 1:

$$r_1 s_4 t_1 V \cdot t_2 \cdot \bar{r}_2 V \bar{s}_5 : c_1 V c_r s_1 p_1 \cdot b_6 b_7 p_2 b_1 k_1 s_3 o_1 \\ \therefore \longleftrightarrow v$$

Similarly, with appropriate change of components, pump 2 operation is:

$$r_2 s_5 t_2 V \cdot t_1 \cdot \bar{r}_1 V \bar{s}_4 : c_1 V c_r s_1 p_1 \cdot b_6 b_7 p_2 b_2 k_2 s_3 o_2 \\ \therefore \longleftrightarrow w$$

There are conditions when both pumps may be operating together. This involves either the use of the critical command (c_c) or the closure of the alternating relay disable switch (s_{10}). For pump 1 and pump 2 both operating,

$$c_1 V c_r s_1 p_1 \cdot c_c V s_{10} \cdot b_6 b_7 p_2 b_1 b_2 k_1 k_2 s_3 o_1 o_2 \\ \therefore \longleftrightarrow w v$$

The conditions for the fans are similar but not identical to those for the pumps. The appropriate circuit breakers, relays, overloads, switches, and voltages are required. The difference is that the fans are not fully redundant and do not back each other up. In addition, the coolant level, the alternating bypass switch, and the pressure cutoff switches are not in the fan circuit. A temperature switch monitors the coolant, and when the coolant temperature is greater than 110°F at the output to the coil, the second fan is commanded on.

The hardware conditions for the operation of fan 1 are:

- (1) The control circuit breaker (b_6) is closed.
- (2) The main circuit breaker (b_7) is closed.
- (3) The fan 1 circuit breaker (b_3) is closed.
- (4) The fan 1 contactor (k_3) is closed.
- (5) The 480-Vac three-phase power (p_2) is available.
- (6) The fan 1 thermal overload (o_3) is closed.

Similarly, for fan 2:

- (1) The control circuit breaker (b_6) is closed.
- (2) The main circuit breaker (b_7) is closed.
- (3) The fan 2 circuit breaker (b_4) is closed.
- (4) The fan 2 contactor (k_4) is closed.
- (5) The 480-Vac three-phase power (p_2) is available.
- (6) The fan 2 thermal overload (o_4) is closed.

For fan 1:

$$t_2 V t_1 k_{15} \cdot c_1 V c_r s_1 p_1 \cdot b_7 b_3 b_6 p_2 k_3 o_3 : \longleftrightarrow y$$

For fan 2:

$$t_1 V t_2 k_{15} \cdot c_1 V c_r s_1 p_1 \cdot b_7 b_4 b_6 p_2 k_4 o_4 : \longleftrightarrow x$$

For fan 1 and fan 2:

$$o_3 o_4 b_7 b_4 b_3 b_6 p_2 k_3 k_4 k_{15} \cdot c_1 V c_r s_1 p_1 : \longleftrightarrow x y$$

At this point, schemata have been written to describe the operation of the pumps and the fans. These are true when the antecedent and the consequent are both true or both false. The

schema is false when the antecedent and the consequent are not both true or false. Since the operation of the pumps or fans is monitored independent of the hardware used for the antecedent, a false schema indicates a problem with either the sensor or the data collection. If the biconditional is broken into two conditionals and each is tested independently, the conditionals may provide different information. This is the case with the pumps, and this procedure can be used as a diagnostic.

Part of knowing the operational status lies in keeping track of the various time durations. How long has the expert system been monitoring? How long has the heat exchanger been on? How many times has the heat exchanger been turned on and off? Other clocks were added as required. Time clock parameters are the following:

t_1 = elapsed time since PICON has been on line monitoring

t_r = continuous accumulated time since the heat exchanger was turned on

t_t = continuous accumulated time duration since the transmitter was turned on

t_c = number of times the heat exchanger has been turned on

III. Heat Exchanger Diagnostics Description

The next task was to list all components (least replaceable units) and failure modes. It was then necessary to identify the symptoms and the response of the sensors to each. After describing the various possible failures, the propositional logic statements were written using the available sensors. For this purpose, the phrase "heat exchanger on" was taken to mean that one pump and one fan were operating. A sensor was available for the coolant flow, but the sensor for the air flow proved to be so noisy as to be unusable. The fan rotation sensor was inoperative due to burned-out lamps. Consequently, no direct means of monitoring air flow was available, so the fan motor prime power loading was used instead for the rules. A sensor for the position of the local control switch S2 was not available, although such a sensor would have simplified the rules.

The heat exchanger is required to provide liquid cooling for the transmitter. In order to do this, it must deliver at least 20 gpm with a pressure differential of at least 125 psid. The coolant temperature is not to exceed 70°C. The head of the transmitter load from the output of the heat exchanger is 125 ft. Coolant purity is to be maintained with a resistivity of not less than 2 megohms per centimeter. These are the most critical requirements. Failure to meet these requirements may

be due either to parts failure in the heat exchanger or to failures in the transmitter or prime power; this project was limited in scope to the heat exchanger. External failures were identified whenever possible, but no procedures were developed to isolate such faults.

The diagnostics began with the heat exchanger off: In the Off state, is the heat exchanger available for operation? It is not possible to answer this on the basis of sufficient conditions, but it is possible to conclude some necessary conditions for the heat exchanger to be turned on. The "H.E. not available" rule defines the conditions that prevent the heat exchanger from being turned on. Fault localization depends on sensors. It is always possible that a sensor has failed and the equipment is nevertheless operational. Unfortunately, it is not easy to isolate sensor failures from equipment failures; only through gross sensor error or sensor redundancy can some sensor failures be isolated. This area has not been thoroughly analyzed, and rules have not been developed except for a few obvious ones.

The "H.E. not available" rule was then broken up into various rules that identified faults either to a component or, if that was not possible, to an area. For example, if a circuit breaker has tripped, the possibilities are endless. Reasonable and probable considerations must be used in the deductive process. The tripped circuit breaker could be the result of sabotage, but this is neither reasonable nor probable. The following basic assumptions were made:

- (1) Wiring has a negligible probability of failure.
- (2) Multiple failures of independent components involving the joint probability of two failures occurring simultaneously are negligible.
- (3) Problems that cannot be isolated to a component will require a technician to troubleshoot the fault with the use of test equipment and procedures.

For the category "H.E. not available," the choice of sensors was an adequate compromise. With the sensors available, it was not possible to identify the failure of a fuse in the control circuit. Sensors should be relocated in the future to identify this potential failure.

The next category is "H.E. available." In this category, warnings are included. In general, if the heat exchanger is capable of providing cooling for full power operation with no failures, no reporting is desired. If the heat exchanger is capable of providing cooling even though failures or potential failures exist, then that information is reported. The basic schema was deduced as an exclusive-OR of these two conditions, namely, "H.E. not available" and "no failures."

Some faults are detected only with the heat exchanger on. If the heat exchanger is not capable of cooling the transmitter as required, the condition is called "H.E. not working." Note that the conditions "H.E. not working" and "H.E. not available" differ only in that the heat exchanger must be commanded to operate to detect "H.E. not working" conditions. In the "H.E. not available" category are faults which are detected whether the heat exchanger is commanded on or off.

An interlock fault is indicated when a switch sensor located in the transmitter indicates a fault. A corresponding analog sensor provides the actual value for each interlock. Since these are independent of each other, small differences in calibration can be expected. The heat exchanger may be operating correctly and an interlock fault may indicate a problem in the transmitter.

When an interlock is activated, the transmitter beam power is turned off, reducing the heat load immediately and decreasing the coolant temperature drastically. The values of flow rate and temperature just before the activation of an interlock comprise very useful diagnostic information.

The failure of the pumps or the fans to cycle properly is an indication of equipment failure. Each time the heat exchanger is turned off and then turned on, the pumps and fans are alternately switched. A failure to switch indicates a failure that, if not repaired, will lead to a heat exchanger failure. In addition, the reliability factor is degraded by the loss of redundancy.

IV. Diagnostics Logic Statements

The next step in the task was to convert the diagnostic description to maintenance logic that is understood by a computer. One simplification made was that any fan failure is a heat exchanger failure.

A. "H.E. Not Available" Rules

In terms of hardware sensors:

- (1) If both pump thermal overloads (sensors $d_{6,100}$ and $d_{6,200}$) are open, or
- (2) If both pump circuit breakers (sensors $d_{3,80}$ and $d_{3,100}$) are open, or
- (3) If one pump circuit breaker and the other pump thermal overload are open, or
- (4) If either fan circuit breaker (sensors $d_{3,200}$ and $d_{3,400}$) is open, or
- (5) If either fan thermal overload (sensors $d_{6,400}$ and $d_{6,800}$) is open, or

- (6) If the control circuit breaker (sensor $d_{3,1}$) is open, or
- (7) If the main circuit breaker ($d_{3,8000}$) is open, or
- (8) If the phase-sensitive relay (sensor $d_{3,40}$) is open, or
- (9) If the 120-Vac control power (sensor $d_{6,8000}$) is not present, or
- (10) If low reservoir level switch ($d_{6,10}$) is open, or
- (11) If low nitrogen pressure switch ($d_{6,20}$) is open,

then the sensors indicate that if the heat exchanger were to be asked to operate, it would not be available to cool the transmitter. Indeed, if it *did* turn on, then the sensors or the data link have failed. This is written as:

$$\bar{d}_{3,1} \vee \bar{d}_{3,40} \vee \bar{d}_{6,10} \vee \bar{d}_{6,20} \vee \bar{d}_{3,80} \bar{d}_{6,200} \vee \bar{d}_{3,100} \bar{d}_{6,100} \vee \bar{d}_{6,800} \vee \bar{d}_{6,400} \vee \bar{d}_{3,80} \bar{d}_{3,100} \vee \bar{d}_{3,200} \vee \bar{d}_{3,400} \vee \bar{d}_{6,100} \bar{d}_{6,200} \vee \bar{d}_{3,8000} \vee \bar{d}_{6,8000} \rightarrow \text{Sensors indicate that the heat exchanger is not available}$$

where $d_{i,j}$ indicates a normal value from the sensor $D_{i,j}$ and $\bar{d}_{i,j}$ indicates a fault value from the sensor $D_{i,j}$.

When the rule "H.E. not available" is activated, the following sensor information is useful in localizing faults:

$$\bar{d}_{3,80} \bar{d}_{6,200} \vee \bar{d}_{3,100} \bar{d}_{6,100} \vee \bar{d}_{3,80} \bar{d}_{3,100} \vee \bar{d}_{6,100} \bar{d}_{6,200} \rightarrow \text{Sensors indicate that both pumps are not operational due to either tripped circuit breakers or tripped overloads.}$$

This can be divided further:

- (1) $d_{3,80} d_{3,100} \bar{d}_{6,100} d_{6,200} \rightarrow$ Sensors indicate a short in pump 1.
- (2) $\bar{d}_{3,80} d_{3,100} d_{6,100} d_{6,200} \rightarrow$ Sensors indicate a short circuit in primary wiring to pump 1.
- (3) $d_{3,80} d_{3,100} d_{6,100} \bar{d}_{6,200} \rightarrow$ Sensors indicate a short in pump 2.
- (4) $d_{3,80} \bar{d}_{3,100} d_{6,100} d_{6,200} \rightarrow$ Sensors indicate a short circuit in primary wiring to pump 2.
- (5) $\bar{d}_{6,10} \bar{a}_{1,1} \rightarrow$ Low reservoir level sensor indicates the heat exchanger is not available. Add coolant and check for leaks. (Note that an analog sensor always has a value because for a signal of zero a current of 4 mA is the output. A sensor failure or data link failure is detected by not receiving the 4 mA.)

- (6) $\bar{d}_{6,10} a_{1,1} \rightarrow$ Low reservoir level signal is not available.
- (7) $\bar{d}_{6,20} \bar{a}_{1,2000} \rightarrow$ Low nitrogen pressure indicates the heat exchanger is not available. Add nitrogen.
- (8) $\bar{d}_{6,20} a_{1,2000} \rightarrow$ Low-pressure cutoff switch signal is not available.

The next set of diagnostics involves control power:

- (1) $\bar{d}_{3,1} \rightarrow \bar{d}_{6,8000}$
- (2) $\bar{d}_{3,1} V \bar{d}_{6,8000} \cdot d_{3,8000} d_{3,40} \rightarrow$ Sensors indicate a heat exchanger control short circuit.
- (3) $\bar{d}_{3,40} V \bar{d}_{6,8000} \cdot d_{3,1} \rightarrow$ Sensors indicate a heat exchanger three-phase power fault.
- (4) $d_{3,200} d_{3,400} \bar{d}_{6,400} d_{6,800} \rightarrow$ Sensors indicate a short in fan 1.
- (5) $\bar{d}_{3,200} d_{3,400} d_{6,400} d_{6,800} \rightarrow$ Sensors indicate a short circuit in primary wiring to fan 1.
- (6) $d_{3,200} d_{3,400} d_{6,400} \bar{d}_{6,800} \rightarrow$ Sensors indicate a short in fan 2.
- (7) $d_{3,200} \bar{d}_{3,400} d_{6,400} d_{6,800} \rightarrow$ Sensors indicate a short circuit in primary wiring to fan 2.

B. "H.E. Available" Rules

The heat exchanger may be in need of maintenance and still be capable of cooling the transmitter. This condition is called "H.E. available." The following schema can be deduced from the "H.E. not available" rule:

$$\begin{aligned} &\bar{d}_{3,80} d_{6,200} V \bar{d}_{3,100} d_{6,100} V \bar{d}_{6,100} d_{6,200} V \\ &\bar{d}_{3,80} d_{3,100} V \bar{d}_{3,100} d_{3,80} V \bar{d}_{6,200} d_{6,100} V \\ &\bar{d}_{6,100} d_{3,100} V \bar{d}_{6,200} d_{3,80} \cdot \rightarrow \text{The heat} \\ &\text{exchanger is available for operation, but mainte-} \\ &\text{nance is to be performed as soon as possible.} \end{aligned}$$

When the "H.E. available" rule is activated, the following sensor information is useful in localizing faults. Included in this category are the warnings that require attention of maintenance personnel but still allow operation to continue:

- (1) $(a_{1,2000} < 30 \text{ psig}) d_{6,20} \rightarrow$ Add nitrogen to the heat exchanger tank.
- (2) $\bar{a}_{1,1} d_{6,10} \rightarrow$ Add coolant to the tank. Check for leaks in the heat exchanger and the transmitter cooling pipe assemblies.

Some faults are detected only with the heat exchanger on. If the heat exchanger is on and not capable of cooling the

transmitter as required, the condition is called "H.E. not working." Note that the condition "H.E. not working" and "H.E. not available" differ only by the ON/OFF command. All "H.E. not available" conditions are also included in the "H.E. not working" category, as these are not mutually exclusive. Only conditions that would not be detected with the heat exchanger off are listed below:

- (1) $\{(a_{1,800} - a_{1,2000}) < 125 \text{ psig}\} \rightarrow \bar{a}_{1,800}$
- (2) $\bar{a}_{1,20} \bar{a}_{1,800} (a_{1,20} > 2 \text{ gpm}) d_{6,4} \rightarrow$ The heat exchanger output flow is less than 20 gpm with a pressure of less than 125 psid as indicated by coolant flow sensors. Replace operating main pump.
- (3) $d_{6,4} \bar{d}_{6,2} d_{6,10} d_{6,20} \rightarrow$ Relay K7 defective or heat exchanger control switch S2 in the off position. The heat exchanger may be operated locally until K7 is replaced.
- (4) $\bar{a}_{2,4} \bar{a}_{2,8} d_{6,10} d_{6,20} d_{6,4} \rightarrow$ Fans are not operating properly. Both fans are off.
- (5) $\bar{a}_{1,40} a_{1,20} \cdot a_{2,4} \bar{a}_{2,8} V \bar{a}_{2,4} a_{2,8} \rightarrow$ Warning! Coolant temperature greater than 110°F and both fans are not on.
- (6) $\bar{d}_{6,1} a_{1,20} \rightarrow$ There is no flow in the purity loop. Replace the submicrometer filter (item 34 of the piping diagram).
- (7) $\bar{a}_{1,4} a_{1,20} (t_r > 10 \text{ min}) \rightarrow$ Replace the purity cartridge (item 24 of the piping diagram).
- (8) $\bar{a}_{1,2} a_{1,4} a_{1,20} (t_r > 1 \text{ hr}) \rightarrow$ Main coolant loop is being contaminated. Localize the cause of contamination. Replace defective component.
- (9) $(a_{1,20} < 2 \text{ gpm}) d_{3,1} d_{3,8000} d_{6,2} d_{6,4} d_{6,10} d_{3,80} d_{3,100} d_{6,20} d_{6,100} d_{6,200} d_{6,8000} \rightarrow$ Sensors indicate a heat exchanger control open circuit.

The transmitter and heat exchanger are usually operated with the heat exchanger controlled by the transmitter. It is also possible to operate the heat exchanger from local controls. This condition is detected only when the transmitter is off and the heat exchanger is on. If both systems are on, there is no way to tell that the heat exchanger is being operated locally:

$$a_{1,20} \bar{d}_{6,4} \rightarrow \text{Heat exchanger under local control.}$$

An interlock fault is indicated when a switch sensor located in the transmitter indicates a fault. A corresponding analog sensor provides the actual value for each interlock. Since these are independent of each other, small differences in calibration can be expected. The heat exchanger may be operating cor-

rectly and an interlock fault may indicate a problem in the transmitter:

$$\begin{aligned} & a_{1,20} \cdot a_{1,800} \cdot \bar{d}_{5,1} \vee \bar{d}_{5,2} \vee \bar{d}_{5,4} \vee \bar{d}_{5,8} \vee \bar{d}_{5,10} \vee \\ & \bar{d}_{5,20} \vee \bar{d}_{5,40} \vee \bar{d}_{5,80} : \longrightarrow \text{Interlock monitoring} \\ & \text{problem, not a heat exchanger problem.} \end{aligned}$$

The following sensor information is useful in isolating interlock faults:

- (1) $\bar{a}_{1,20} \cdot \bar{a}_{1,800} \cdot \bar{d}_{5,1} \vee \bar{d}_{5,2} \vee \bar{d}_{5,4} \vee \bar{d}_{5,8} \vee \bar{d}_{5,10} \vee \bar{d}_{5,20} \vee \bar{d}_{5,40} \vee \bar{d}_{5,80} \longrightarrow$ Sensors indicate the cause of the interlock is the heat exchanger.
- (2) $\bar{d}_{5,1} \vee \bar{d}_{5,4} \vee \bar{d}_{5,10} \vee \bar{d}_{5,40} \longrightarrow$ Flow interlock:
Magnet flow -----
Body flow -----
Collector flow -----
RF load flow -----
- (3) $\bar{d}_{5,2} \vee \bar{d}_{5,8} \vee \bar{d}_{5,20} \vee \bar{d}_{5,80} \longrightarrow$ Temperature interlock:
Magnet temperature -----
Body temperature -----
Collector temperature -----
RF load temperature -----

The values of flow and temperature would be given where the dashed lines appear. These values provide valuable information to a maintenance technician.

A condition for the cycling is that a different pump and a different fan be turned on each time the heat exchanger is turned on. Turn-on counter t_c is incremented each time the heat exchanger is turned on.

Pumps cycling:

- (1) $\bar{d}_{3,4} \bar{d}_{3,8} \longrightarrow$ Pump 1 is on.
- (2) $\bar{d}_{3,4} d_{3,8} \longrightarrow$ Pump 2 is on.
- (3) $d_{3,4} \bar{d}_{3,8} (t_c) \cdot \bar{d}_{3,4} d_{3,8} (t_c + 1) : \longrightarrow$ Pumps are cycling properly.
- (4) $\bar{d}_{3,4} d_{3,8} (t_c) \cdot d_{3,4} \bar{d}_{3,8} (t_c) : \longrightarrow$ Pumps are cycling properly.
- (5) $d_{3,4} (t_c) \cdot d_{3,4} (t_c + 1) \vee d_{3,8} (t_c) \cdot d_{3,8} (t_c + 1) : \therefore \longrightarrow$ Pumps are not cycling properly.

Fans cycling:

- (1) $\bar{d}_{3,10} \bar{d}_{3,20} \longrightarrow$ Fan 1 is on.
- (2) $\bar{d}_{3,10} d_{3,20} \longrightarrow$ Fan 2 is on.
- (3) $d_{3,10} \bar{d}_{3,20} (t_c) \cdot \bar{d}_{3,10} d_{3,20} (t_c + 1) : \longrightarrow$ Fans are cycling properly.
- (4) $\bar{d}_{3,10} d_{3,20} (t_c) \cdot d_{3,10} \bar{d}_{3,20} (t_c + 1) : \longrightarrow$ Fans are cycling properly.
- (5) $d_{3,10} (t_c) \cdot d_{3,10} (t_c + 1) \vee d_{3,20} (t_c) \cdot d_{3,20} (t_c + 1) : \therefore \longrightarrow$ Fans are not cycling properly.

This is not an exhaustive set of rules. Additional rules should be added by maintenance personnel as experience is obtained with actual failures and repairs. Ideally, each failure should be analyzed and new rules developed to improve the capability of the expert system.

V. Other Diagnostics

Analysis of heat exchanger efficiency and trends such as reservoir level changes can be implemented. This type of information alerts maintenance personnel to possible future problems. Preventive or scheduled maintenance can correct a potential problem before an actual failure occurs.

A decreasing reservoir level indicates a leak which, if not corrected, will lead to an equipment failure. Actually, detecting the tank level is somewhat complicated. The reservoir level changes when the pumps are turned on due to the expansion of the piping caused by the increased pressure at the output of the pumps. Another factor is the expansion of metals due to temperature. These complications can be taken into account in one of two ways: the variation of each measured parameter can be found and appropriate corrections made, or the measurement can be made only at times when conditions are identical. For example, the tank level will be measured only when the pumps are off and the temperature is within defined limits.

The heat exchanger efficiency was a different problem. Deteriorating efficiency indicates future problems. Originally, it was intended to calculate the U-factors in the following manner [3]:

$$q = UA \Delta T_m$$

where

U = overall heat-transfer coefficient

A = surface area for heat transfer consistent with the definition of U

ΔT_m = suitable mean temperature difference across the heat exchanger

However, this method proved to be impractical for a number of reasons. Some of the parameters needed are difficult to measure (e.g., mass air flow); some of the sensors proved to be unreliable; some of the sensors were not available (e.g., sensors to measure specific heat of the coolant); and the heat is dissipated in other places outside the heat exchanger core. (The dissipation of heat in the long piping lines between the heat exchanger and the transmitter depends on the elements, such as rain, snow, ambient temperature, and wind.) Actually, the choice of U -factors was a bad one. A better choice would have been the Effectiveness-NTU method, defined as

$$\text{effectiveness} = \frac{\text{actual heat transfer}}{\text{maximum possible heat transfer}}$$

This method has the advantage that the inlet and outlet air temperature do not have to be measured. The disadvantage of this method is the difficulty in obtaining the mass flow rate of air across the coil.

However, the problem can be reduced to a much simpler one. The heat taken out of the coolant by the coil can be measured without difficulty, and the efficiency of the heat transferred is a function of the difference in temperature between

the coolant temperature at the outlet of the coil and the air ambient temperature. The heat exchanger can be calibrated and a table of values stored for comparison. Corrections can be made for the specific heat of the coolant, air humidity, and barometric pressure. Allowing for data errors, a value of the efficiency can be used to estimate long-term degradation in cooling efficiency. The failure to measure U -factors appears to be a case of insufficient investigation of the problem during the design stage. As described above, however, a simple alternative exists that could provide the information needed by maintenance personnel.

VI. Conclusions

Based on the work carried out, the following conclusions were made:

- (1) The use of propositional logic is well suited to the development of expert maintenance systems.
- (2) The development of expert maintenance systems must involve on-site maintenance personnel.
- (3) The expert system must include not only past experience but present and future experience as well.
- (4) The initial effort need not be exhaustive, but rather should form a shell that can be expanded with new rules and new procedures. In this way it will be easily accepted by users and can be developed to reach its full potential.

References

- [1] J. P. Delahaye, *Formal Methods in Artificial Intelligence*, Chapters 3 and 4, New York: Wiley and Sons, 1987.
- [2] D. G. Bagby and R. Cormier, "Heat Exchanger Demonstration Expert System," *TDA Progress Report 42-93*, vol. January-March 1988, Jet Propulsion Laboratory, Pasadena, California, pp. 177-187, May 15, 1988.
- [3] J. P. Holman, *Heat Transfer*, second edition, Chapter 10, New York: McGraw-Hill, 1986.
- [4] W. V. Quine, *Methods of Logic*, fourth edition, Cambridge, Massachusetts: Harvard University Press, 1982.

Table 1. Heat exchanger sensors and elements

Sensor number	Description	Definition	Statement
A-01*0001	Reservoir level transducer	>1.5 ft → a _{1,1}	
A-01*0002	Resistivity inlet purity	>2 megohms/cm ² → a _{1,2}	
A-01*0004	Resistivity outlet purity	>5 megohms/cm ² → a _{1,4}	
A-01*0008	HE core flow	a _{1,8}	
A-01*0010	HE bypass flow	a _{1,10}	
A-01*0020	HE total flow	>20 gpm → a _{1,20}	
A-01*0040	HE core outlet temperature	<112°F → a _{1,40}	
A-01*0080	HE core inlet temperature	a _{1,80}	
A-01*0100	HE assy inlet temperature	not used	
A-01*0200	Ambient air temperature	a _{1,200}	
A-01*0400	Core air flow pressure delta	not usable; too noisy	
A-01*0800	Pump outlet pressure	a _{1,800}	
A-01*1000	Core outlet pressure	a _{1,1000}	
A-01*2000	Suction inlet pressure	>15 psig → a _{1,2000}	
A-01*4000	Pump inlet temperature	not used	
A-01*8000	Heater outlet temperature	not part of DSN HE	
A-02*0001	Pump 1 motor load	not used	
A-02*0002	Pump 2 motor load	not used	
A-02*0004	Fan #1 motor load	>1200 watts → a _{2,4}	
A-02*0008	Fan #2 motor load	>1200 watts → a _{2,8}	
A-02*0010	Fan #1 speed	not reliable	
A-02*0020	Fan #2 speed	not reliable	
A-02*0040	3-way valve position	not working	
A-02*0080	Spare		
A-02*0100	HE assy outlet temperature	a _{2,100}	
A-02*0200	Heater load	not part of DSN HE	
A-02*0400	Differential pressure across main line filter	<15 psid → a _{2,400}	
D-03*0001	Auxiliary contacts on CB6	d _{3,1}	b ₆
D-03*0002	Auxiliary contacts on K5	not used	
D-03*0004	Auxiliary contacts on K1	d _{3,4}	k ₁
D-02*0008	Auxiliary contacts on K2	d _{3,8}	k ₂
D-03*0010	Auxiliary contacts on K3	d _{3,10}	k ₃
D-03*0020	Auxiliary contacts on K4	d _{3,20}	k ₄
D-03*0040	Auxiliary contacts on K6	d _{3,40}	p ₂
D-03*0080	Auxiliary contacts on CB1	d _{3,80}	b ₁
D-03*0100	Auxiliary contacts on CB2	d _{3,100}	b ₂
D-03*0200	Auxiliary contacts on CB3	d _{3,200}	b ₃
D-03*0400	Auxiliary contacts on CB4	d _{3,400}	b ₄
D-03*0800	Auxiliary contacts on CB5	not used	
D-03*1000	Auxiliary contacts on K8	not working on test HE	c _c
D-03*2000	Limit SW #1, plenum door	not used	
D-03*4000	Limit SW #2, plenum door	not used	
D-03*8000	Auxiliary contacts on CB7	d _{3,8000}	b ₇
A-04*0001	Magnet flow	>1.5 gpm → a _{4,1}	
A-04*0002	Magnet temperature	<70°C → a _{4,2}	
A-04*0004	Body flow	>1.2 gpm → a _{4,2}	
A-04*0008	Body temperature	<70°C → a _{4,4}	
A-04*0010	Collector flow	>18 gpm → a _{4,10}	
A-04*0020	Collector temperature	<70°C → a _{4,20}	
A-04*0040	RF load flow	>6 gpm → a _{4,40}	
A-04*0080	RF load temperature	<70°C → a _{4,80}	
A-04*0100	Magnet voltage	not used	
A-04*0200	Magnet current	not used	

Table 1. (contd)

Sensor number	Description	Definition	Statement
A-04*0400	Spare		
A-04*0800	Body current	not used	
A-04*1000	RF load power	not used	
A-04*2000	Spare		
A-04*4000	Beam voltage	not used	
A-04*8000	Beam current	not used	
D-05*0001	Low magnet flow interlock	d _{5,1}	
D-05*0002	High magnet temperature interlock	d _{5,2}	
D-05*0004	Low body flow interlock	d _{5,4}	
D-05*0008	High body temperature interlock	d _{5,8}	
D-05*0010	Low collector flow interlock	d _{5,10}	
D-05*0020	High collector temperature interlock	d _{5,20}	
D-05*0040	Low RF load flow interlock	d _{5,40}	
D-05*0080	High RF load temperature interlock	d _{5,80}	
D-05*0100	RF load and RF power range	not used	
D-06*0001	Purity loop flow switch	d _{6,1}	
D-06*0002	Auxiliary contacts on K7	d _{6,2}	k ₇
D-06*0004	Monitor control power across connector pins A and B	d _{6,4}	p ₁
D-06*0008	Auxiliary contacts on S8	d _{6,8}	
D-06*0010	Low reservoir level switch	d _{6,10}	s ₃
D-06*0020	Low nitrogen pressure switch	d _{6,20}	s ₁
D-06*0040	PM1 resistivity meter status	d _{6,40}	
D-06*0080	PM2 resistivity meter status	d _{6,80}	
D-06*0100	OL1 K1 thermal O/L status	d _{6,100}	o ₁
D-06*0200	OL2 K2 thermal O/L status	d _{6,200}	o ₂
D-06*0400	OL3 K3 thermal O/L status	d _{6,400}	o ₃
D-06*0800	OL4 K4 thermal O/L status	d _{6,800}	o ₄
D-06*1000	OL5 K5 thermal O/L status	not used	
D-06*2000	S4 pressure switch status	not working in test HE	s ₄
D-06*4000	S5 pressure switch status	not working in test HE	s ₅
D-06*8000	120-Vac control power status	d _{6,8000}	
	K15 fan control relay		k ₁₅
	S2 panel control switch:		
	Remote		c _r
	Off		c _o
	Local		c _l
	TDR1 pump 1 time delay relay		r ₁
	TDR2 pump 2 time delay relay		r ₂
	S ₁₀ pump alternating disable switch		s ₁₀
	K9 pump alternating relay:		
	Pump 1		t ₁
	Pump 2		t ₂
	K14 fan alternating relay:		
	Fan 1		t ₁
	Fan 2		t ₂

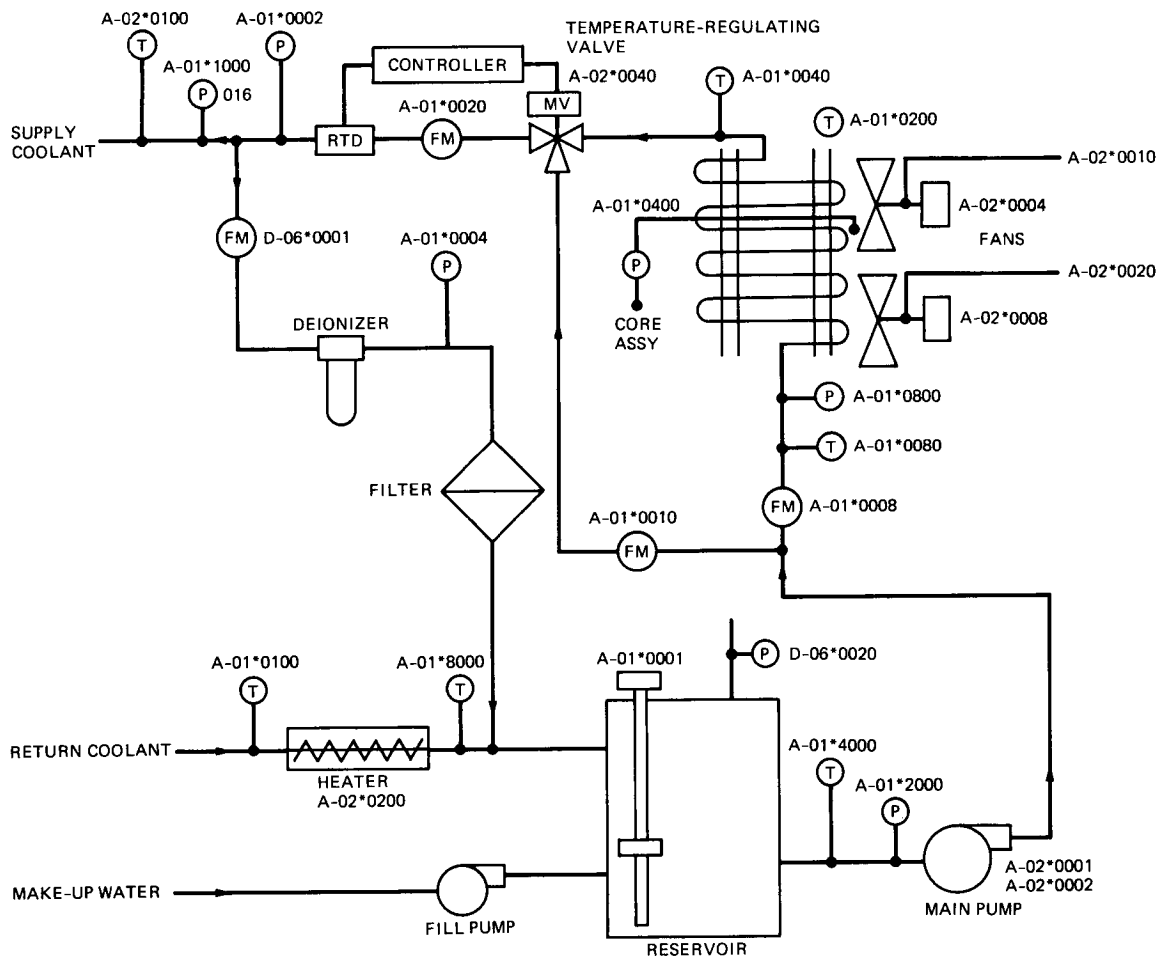


Fig. 1. Heat exchanger flow diagram

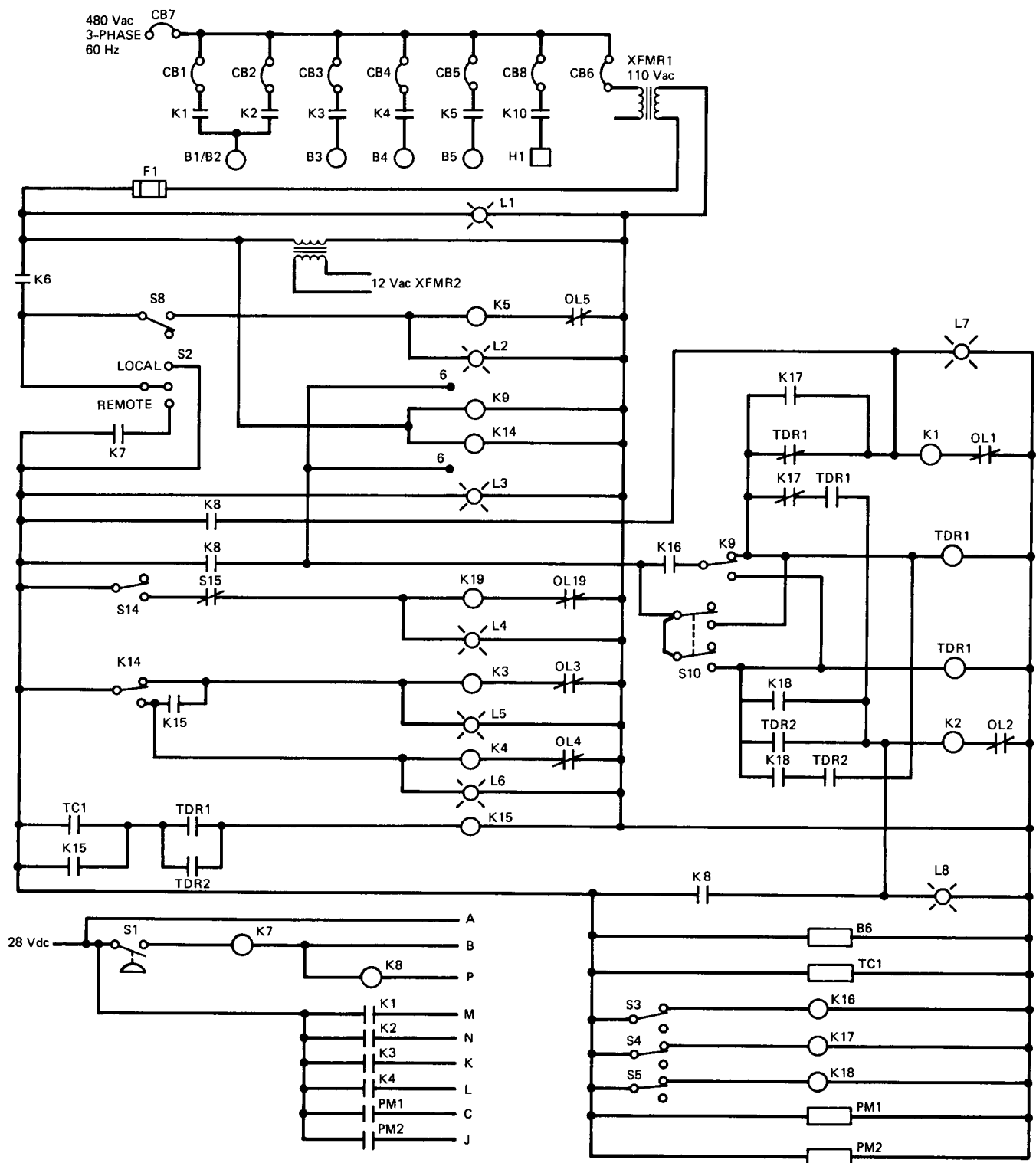


Fig. 2. Heat exchanger electrical schematic

Appendix

Logic Symbol Definitions

The symbols used in the description of the operation of the assemblies of the transmitter were taken from [4]. They are described below.

Negation. A bar above a letter represents negation of the statement that this letter represents. For example, \bar{p} means "not-p" or $\neg p$. The meaning of negation is summed up in "the negation of a true statement is false; the negation of a false statement is true."

Conjunction. The conjunction of two statements such as p and q is represented as pq . The meaning of conjunction is taken to mean "a conjunction of statements all of which are true is true, and a conjunction of statements not all of which are true is false."

Alternation. The alternation of two statements such as p or q is represented as $p \vee q$. The meaning of alternation is given by this rule: "An alternation is true if at least one of the components is true and otherwise false."

Conditional. A statement of the form "if p then q" is represented as $p \longrightarrow q$ and is called a conditional. The "p" component is called the antecedent and the "q" component is called the consequent. Note that the antecedent is not to be interpreted to be the cause of the consequent. If the antecedent is false, the conditional is taken as true regardless of whether the consequent is true or false. The relation is "if p then q" and "if q then p." It is represented as $p \longleftrightarrow q$ and is called biconditional.

Grouping. The dot convention is used in place of brackets. The insertion of a dot indicates a greater break at the point where it is inserted. For example, " $pq \vee r$ " is understood to mean " $(pq) \vee r$." If the intent is " $p(q \vee r)$," the representation is " $p.q \vee r$," indicating the break between "p and q." Greater emphasis is indicated by the use of multiple dots such as ":", ":", ":", etc.

Consistency and validity. A schema is called consistent if it comes out true under some interpretation of its letters. A schema is valid if it is true under every interpretation of its letters.

1989000835
618446
14p.S20-64
~~15551~~
April-June 1988
FEBN89-10206
JPR

Storage and Computationally Efficient Permutations of Factorized Covariance and Square-Root Information Arrays

R. J. Muellerschoen
Navigation Systems Section

A unified method to permute vector-stored Upper-triangular-Diagonal factorized covariance (UD) and vector-stored upper-triangular Square-Root Information (SRI) arrays is presented. The method involves cyclic permutation of the rows and columns of the arrays and retriangularization with fast (slow) Givens rotations (reflections). Minimal computation is performed, and a one-dimensional scratch array is required. To make the method efficient for large arrays on a virtual memory machine, computations are arranged so as to avoid expensive paging faults. This method is potentially important for processing large volumes of radio metric data in the DSN.

I. Introduction

In the reduction of observational data involving simultaneous least-squares estimation of many parameters, it is desirable to examine different modeling scenarios without reprocessing the data through the filter. Often the analyst will estimate a full contingent of system parameters in the initial filter pass. A reduced state estimate, reflecting a different modeling scenario, can then be obtained from this initial pass. Furthermore, the sensitivity of the reduced state to the excluded parameters is readily obtainable. Application of a priori information for the excluded parameters to the sensitivity and subsequent augmentation of the reduced state by these perturbations then yields a more conservative assessment of the reduced state errors. This type of analysis is particularly useful when the

reduced state is sensitive to the excluded parameters that are insensitive to data [1].

Upper-triangular data structures, such as those encountered in an Upper-triangular-Diagonal factorized covariance filter (UD filter) or in an upper-triangular Square-Root Information filter (SRI filter), are conducive to computing these reduced state estimates. In a UD filter, it is necessary to permute the rows of the UD array in order to compute the reduced estimate. The rows corresponding to the parameters that are to be excluded must be moved below those corresponding to the parameters that are to be included in the reduced estimate. In an SRI filter, it is necessary to permute the columns of the SRI array in order to compute the reduced estimate. The

columns corresponding to the parameters that are to be excluded must be moved to the right of the columns corresponding to those that are to be included in the reduced estimate.¹

A brute force method to accomplish these permutations reorders the rows (columns) of the UD (SRI) array into a two-dimensional work array. To retriangularize this two-dimensional array, one can then post(pre)-multiply it by an implicitly defined orthogonal transformation composed of Householder reflections.² This method is sound and numerically stable due to the orthogonality of the transformation. Furthermore, the commercially available Estimation Subroutine Library (ESL) has standard routines to conveniently perform these permutations and subsequent retriangularizations [2]. However, this method requires a two-dimensional work array to store the permuted UD (SRI) array. For large-scale systems, such as those encountered for Global Positioning System (GPS) studies, it is not at all uncommon to estimate on the order of 1000 parameters. This translates into four million extra bytes of storage. Additionally, over 3000 parameters are routinely solved when DSN VLBI data are processed with MASTERFIT software [3]. Moreover, for systems of this order, retriangularization is computationally expensive.

In 1986, Bierman (personal communication) suggested that the necessary reordering of UD arrays be performed as a series of pairwise permutations in such a way that the upper-triangular structure of the UD array is always maintained. Each pairwise permutation could be performed as a suboptimal noise-free measurement update:

$$P_{\text{update}} = (I - GH)P(I - GH)^t + RGG^t$$

where P_{update} is then the permuted covariance of the covariance P . The noise R is set equal to zero, and the gain G and design H are chosen so that $(I - GH)$ is a permutation operator. Such a permutation operator that permutes parameters i and j can be expressed as

$$I - GH = I - (e_i - e_j)(e_i - e_j)^t$$

¹ S. C. Wu et al., *Oasis Mathematical Description*, JPL Publication D-3139 (internal document), Jet Propulsion Laboratory, Pasadena, California, April 1, 1986.

² Before this retriangularization can be performed on the UD array, it is of course necessary to scale the columns by the square root of the corresponding diagonal. These square roots can be avoided by performing a modified weighted Gram-Schmidt UD triangularization algorithm.

where e_i is a column vector whose elements are zero except element i , which is 1. Trivially $H = (e_i - e_j)^t$ and $G = H^t$. In practice this computation is implemented with an optimal measurement update using the Bierman UD measurement update algorithm and a rank-1 update to include the effect of the suboptimality of the gain. This method eliminated the need for a two-dimensional scratch array. Software was eventually written that exploited the structure of the suboptimal gain G and design H . This method still required several scratch arrays and a multitude of pairwise permutations.³

In 1987, Wolff (personal communication) suggested that the same pairwise permutations could be performed on an SRI array with retriangularization accomplished after each column exchange with slow Givens reflections.⁴ This method also eliminated the need for a two-dimensional work array. However, in the retriangularization process, elements not consistent with the upper-triangular data structure of the SRI array were created. Each inconsistent element required another slow Givens reflection. Subsequently, this method required on the order of n^2 slow Givens reflections, where n is the column distance between the two parameters of the exchange. Furthermore, this method required two explicit scratch arrays. One of the scratch arrays stored the right column of the exchange while the other was used to store elements that were created in the retriangularization process.⁵

Also in 1987, Pombra suggested a technique that eliminated the need to perform a multitude of these pairwise permutations in reordering a UD array.⁶ His technique was to insert into the covariance an artificial parameter with zero variance and zero correlation with the other parameters. This is easily accomplished by inserting into the UD array a row and column of zeros. A parameter could then be moved directly into its proper place by performing a pairwise permutation of this artificial parameter and the parameter that is desired to reside in its place. After the permutation is performed, this artificial parameter can then be deleted from the UD array. This is also easily accomplished by removing from the UD array the appropriate row and column of zeros. This technique drastically

³ P. J. Wolff, "UD Permutations via ESL Subroutine UCON," JPL IOM 314.5-1040 (internal document), Jet Propulsion Laboratory, Pasadena, California, August 29, 1986.

⁴ Since the SRI array is not unique, either Givens reflections or rotations can be used.

⁵ P. J. Wolff, "Permuting R Matrices in Place," IOM 314.5-1091 (internal document), Jet Propulsion Laboratory, Pasadena, California, March 10, 1987.

⁶ S. A. Pombra, "Computationally Fast Version of UEDIT," IOM 335.1-87-160 (internal document), Jet Propulsion Laboratory, Pasadena, California, July 7, 1987.

reduced the number of permutations necessary to reorder the array. However, suboptimal noise-free measurement updates were still used to perform the pairwise permutations.

It was recognized that the insertion of a row and column of zeros into the UD array, together with the subsequent row exchange between the artificial parameter and the parameter of interest, created an array similar to a Morf-Kailath prearray with zero noise variance [4]. Retriangularization of this array into an upper-triangular Morf-Kailath postarray is easily accomplished with fast Givens rotations.⁷ This method requires a one-dimensional scratch vector which manifests itself as an augmentation of the UD array. An extra column of the UD array permits the insertion of a row and column of zeros. Furthermore, the method requires on the order of \tilde{n}_{row} fast Givens rotations, where \tilde{n}_{row} is the row distance between the parameter and the desired location of the parameter. Due to the nature of the fast Givens, square root computations are never required.

This technique can also be adopted to reordering SRI arrays. In a manner similar to the Wolff method, retriangulization is accomplished with slow Givens reflections. Unlike the Wolff method, however, no new elements are created. Hence, this method requires on the order of only $\tilde{n}_{\text{column}}$ slow Givens reflections, where $\tilde{n}_{\text{column}}$ is the column distance between the parameter and the desired location of the parameter.

Since the artificial parameter is only a convenient place holder, an alternative method could be constructed that does not incorporate the insertion/deletion step. Similar results can be achieved by cyclically permuting the rows and columns of the arrays. Retriangularization is then achieved as before. Furthermore, since the size of the array does not change, it is not necessary to augment the array with an extra column. However, a one-dimensional scratch vector is required.

Cyclic column permutations and retriangulization of SRI arrays have already appeared in the literature [5]. In fact, the discussed method to cyclically permute columns of an SRI array is similar to the left circular shift method used in the LINPACK routine SCHEX [6].⁸ Until now, however, these results have not been extended to cyclic row permutations and retriangulization of UD arrays. The permutations are performed in such a way that the only difference between permuting UD and SRI arrays lies in the retriangulization process.

⁷Since the UD array is unique, fast Givens reflections cannot be used.

⁸SCHEX requires two one-dimensional scratch arrays to store the transformation pairs and does not store the upper-triangular array as a vector.

II. Row Permutations of UD Arrays

To demonstrate the technique on UD arrays, let the factorization of a covariance P with parameters $U1, U2, U3, U4, U5$ be

$$P = UDU^t \quad (1)$$

where

$$U = \begin{bmatrix} u1_1 & u1_2 & u1_3 & u1_4 & u1_5 \\ 0 & u2_2 & u2_3 & u2_4 & u2_5 \\ 0 & 0 & u3_3 & u3_4 & u3_5 \\ 0 & 0 & 0 & u4_4 & u4_5 \\ 0 & 0 & 0 & 0 & u5_5 \end{bmatrix} \quad (2)$$

and

$$D = \begin{bmatrix} d_1 & 0 & 0 & 0 & 0 \\ 0 & d_2 & 0 & 0 & 0 \\ 0 & 0 & d_3 & 0 & 0 \\ 0 & 0 & 0 & d_4 & 0 \\ 0 & 0 & 0 & 0 & d_5 \end{bmatrix} \quad (3)$$

If we designate a weighted inner product as

$$x \cdot y \equiv \sum_i x_i d_i y_i \quad (4)$$

then the covariance can be written in terms of its factorized elements:

$$P = \begin{bmatrix} u1 \cdot u1 & u1 \cdot u2 & u1 \cdot u3 & u1 \cdot u4 & u1 \cdot u5 \\ u2 \cdot u1 & u2 \cdot u2 & u2 \cdot u3 & u2 \cdot u4 & u2 \cdot u5 \\ u3 \cdot u1 & u3 \cdot u2 & u3 \cdot u3 & u3 \cdot u4 & u3 \cdot u5 \\ u4 \cdot u1 & u4 \cdot u2 & u4 \cdot u3 & u4 \cdot u4 & u4 \cdot u5 \\ u5 \cdot u1 & u5 \cdot u2 & u5 \cdot u3 & u5 \cdot u4 & u5 \cdot u5 \end{bmatrix} \quad (5)$$

As usual, let $u_{i_i} = 1$ to obtain a unique factorization. With this convention, the diagonals of the D array can now be stored as the diagonals of the U array. This new array is referred to as the UD array:⁹

$$UD = \begin{bmatrix} d_1 & u_{1_2} & u_{1_3} & u_{1_4} & u_{1_5} \\ 0 & d_2 & u_{2_3} & u_{2_4} & u_{2_5} \\ 0 & 0 & d_3 & u_{3_4} & u_{3_5} \\ 0 & 0 & 0 & d_4 & u_{4_5} \\ 0 & 0 & 0 & 0 & d_5 \end{bmatrix} \quad (6)$$

Say it is desired to reorder the parameters as U_1, U_3, U_4, U_2, U_5 . The first step is to insert an artificial parameter, with zero variance and zero correlation with the other parameters, before parameter U_5 . This is accomplished by inserting into the UD array a row of zeros above row 5 and a column of zeros left of column 5:

$$\begin{bmatrix} d_1 & u_{1_2} & u_{1_3} & u_{1_4} & 0 & u_{1_5} \\ 0 & d_2 & u_{2_3} & u_{2_4} & 0 & u_{2_5} \\ 0 & 0 & d_3 & u_{3_4} & 0 & u_{3_5} \\ 0 & 0 & 0 & d_4 & 0 & u_{4_5} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & d_5 \end{bmatrix} \quad (7)$$

The corresponding covariance is

$$\begin{bmatrix} u_1 \cdot u_1 & u_1 \cdot u_2 & u_1 \cdot u_3 & u_1 \cdot u_4 & 0 & u_1 \cdot u_5 \\ u_2 \cdot u_1 & u_2 \cdot u_2 & u_2 \cdot u_3 & u_2 \cdot u_4 & 0 & u_2 \cdot u_5 \\ u_3 \cdot u_1 & u_3 \cdot u_2 & u_3 \cdot u_3 & u_3 \cdot u_4 & 0 & u_3 \cdot u_5 \\ u_4 \cdot u_1 & u_4 \cdot u_2 & u_4 \cdot u_3 & u_4 \cdot u_4 & 0 & u_4 \cdot u_5 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ u_5 \cdot u_1 & u_5 \cdot u_2 & u_5 \cdot u_3 & u_5 \cdot u_4 & 0 & u_5 \cdot u_5 \end{bmatrix} \quad (8)$$

⁹The upper-triangular elements of either a UD or an SRI array are stored contiguously in computer memory as a vector; the zero elements below the diagonal are obviously not needed to represent the array.

Rows 2 and 5 of Eq. (7) can then be exchanged:

$$\begin{bmatrix} d_1 & u_{1_2} & u_{1_3} & u_{1_4} & 0 & u_{1_5} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & d_3 & u_{3_4} & 0 & u_{3_5} \\ 0 & 0 & 0 & d_4 & 0 & u_{4_5} \\ 0 & d_2 & u_{2_3} & u_{2_4} & 0 & u_{2_5} \\ 0 & 0 & 0 & 0 & 0 & d_5 \end{bmatrix} \quad (9)$$

In practice, however, elements d_2, u_{2_3} , and u_{2_4} would not be moved, since a scratch array would be needed to store these elements. As will soon be evident, the second row can itself be used as a scratch array. The array in Eq. (9) implicitly corresponds to a nontriangular \tilde{U} array

$$\tilde{U} = \begin{bmatrix} u_{1_1} & u_{1_2} & u_{1_3} & u_{1_4} & 0 & u_{1_5} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & u_{3_3} & u_{3_4} & 0 & u_{3_5} \\ 0 & 0 & 0 & u_{4_4} & 0 & u_{4_5} \\ 0 & u_{2_2} & u_{2_3} & u_{2_4} & 0 & u_{2_5} \\ 0 & 0 & 0 & 0 & 0 & u_{5_5} \end{bmatrix} \quad (10)$$

a diagonal \tilde{D} array

$$\tilde{D} = \begin{bmatrix} d_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & d_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & d_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & d_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & d_5 \end{bmatrix} \quad (11)$$

and a covariance \tilde{P}

$$\tilde{P} = \tilde{U}\tilde{D}\tilde{U}^t$$

$$= \begin{bmatrix} u1 \cdot u1 & 0 & u1 \cdot u3 & u1 \cdot u4 & u1 \cdot u2 & u1 \cdot u5 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ u3 \cdot u1 & 0 & u3 \cdot u3 & u3 \cdot u4 & u3 \cdot u2 & u3 \cdot u5 \\ u4 \cdot u1 & 0 & u4 \cdot u3 & u4 \cdot u4 & u4 \cdot u2 & u4 \cdot u5 \\ u2 \cdot u1 & 0 & u2 \cdot u3 & u2 \cdot u4 & u2 \cdot u2 & u2 \cdot u5 \\ u5 \cdot u1 & 0 & u5 \cdot u3 & u5 \cdot u4 & u5 \cdot u2 & u5 \cdot u5 \end{bmatrix} \quad (12)$$

Except for the artificial parameter, this is the desired permuted covariance. The covariance factorization in Eq. (12) can be thought of as the product of an upper-triangular array and its transpose:

$$\tilde{P} = \tilde{U}\sqrt{\tilde{D}}(\tilde{U}\sqrt{\tilde{D}})^t \quad (13)$$

where

$$\sqrt{\tilde{D}} = \begin{bmatrix} \sqrt{d_1} & 0 & 0 & 0 & 0 & 0 \\ 0 & \sqrt{d_2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \sqrt{d_3} & 0 & 0 & 0 \\ 0 & 0 & 0 & \sqrt{d_4} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sqrt{d_5} \end{bmatrix} \quad (14)$$

It is desirable to construct an orthogonal transformation T such that $\tilde{U}\sqrt{\tilde{D}}T$ is an upper-triangular array with zeros in the second row and column:

$$\tilde{U}\sqrt{\tilde{D}}T =$$

$$\begin{bmatrix} u1_1\sqrt{d_1} & u1_2\sqrt{d_2} & u1_3\sqrt{d_3} & u1_4\sqrt{d_4} & 0 & u1_5\sqrt{d_5} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & u3_3\sqrt{d_3} & u3_4\sqrt{d_4} & 0 & u3_5\sqrt{d_5} \\ 0 & 0 & 0 & u4_4\sqrt{d_4} & 0 & u4_5\sqrt{d_5} \\ 0 & u2_2\sqrt{d_2} & u2_3\sqrt{d_3} & u2_4\sqrt{d_4} & 0 & u2_5\sqrt{d_5} \\ 0 & 0 & 0 & 0 & 0 & u5_5\sqrt{d_5} \end{bmatrix} T \quad (15)$$

$$\tilde{U}\sqrt{\tilde{D}}T = \begin{bmatrix} u1_1\sqrt{d_1} & 0 & x & x & x & u1_5\sqrt{d_5} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & x & x & x & u3_5\sqrt{d_5} \\ 0 & 0 & 0 & x & x & u4_5\sqrt{d_5} \\ 0 & 0 & 0 & 0 & x & u2_5\sqrt{d_5} \\ 0 & 0 & 0 & 0 & 0 & u5_5\sqrt{d_5} \end{bmatrix} \quad (16)$$

A portion of the array in Eq. (15) is similar to a Morf-Kailath prearray with zero noise variance. Reduction of this prearray to the postarray in Eq. (16) is easily accomplished with three fast Givens rotations. As a result of the zero noise variance, the first fast Givens T_0 is simply a permutation of columns 2 and 5:

$$\tilde{U}\sqrt{\tilde{D}}T_0 =$$

$$\begin{bmatrix} u1_1\sqrt{d_1} & u1_2\sqrt{d_2} & u1_3\sqrt{d_3} & u1_4\sqrt{d_4} & 0 & u1_5\sqrt{d_5} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & u3_3\sqrt{d_3} & u3_4\sqrt{d_4} & 0 & u3_5\sqrt{d_5} \\ 0 & 0 & 0 & u4_4\sqrt{d_4} & 0 & u4_5\sqrt{d_5} \\ 0 & u2_2\sqrt{d_2} & u2_3\sqrt{d_3} & u2_4\sqrt{d_4} & 0 & u2_5\sqrt{d_5} \\ 0 & 0 & 0 & 0 & 0 & u5_5\sqrt{d_5} \end{bmatrix} \times \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} u1_1\sqrt{d_1} & 0 & u1_3\sqrt{d_3} & u1_4\sqrt{d_4} & u1_2\sqrt{d_2} & u1_5\sqrt{d_5} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & u3_3\sqrt{d_3} & u3_4\sqrt{d_4} & 0 & u3_5\sqrt{d_5} \\ 0 & 0 & 0 & u4_4\sqrt{d_4} & 0 & u4_5\sqrt{d_5} \\ 0 & 0 & u2_3\sqrt{d_3} & u2_4\sqrt{d_4} & u2_2\sqrt{d_2} & u2_5\sqrt{d_5} \\ 0 & 0 & 0 & 0 & 0 & u5_5\sqrt{d_5} \end{bmatrix} \quad (17)$$

This is the required step to produce the zeros in the second row and column. The next fast Givens T_1 operates on columns 3 and 5 in such a way that the array element containing $u_{23}\sqrt{d_3}$ becomes zero. The final fast Givens T_2 operates on columns 4 and 5 such that the array element containing $u_{24}\sqrt{d_4}$ becomes zero. Each of these Givens transformations maps the elements in each row of the columns on which they operate into a linear combination of the same. Hence, elements not consistent with an upper-triangular structure are never created; a linear combination of zeros is still zero. For example, T_1 operates on Eq. (17) as follows:

$$\tilde{U}\sqrt{\tilde{D}}T_0T_1 =$$

$$\begin{bmatrix} u_{11}\sqrt{d_1} & 0 & u_{13}\sqrt{d_3} & u_{14}\sqrt{d_4} & u_{12}\sqrt{d_2} & u_{15}\sqrt{d_5} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & u_{33}\sqrt{d_3} & u_{34}\sqrt{d_4} & 0 & u_{35}\sqrt{d_5} \\ 0 & 0 & 0 & u_{44}\sqrt{d_4} & 0 & u_{45}\sqrt{d_5} \\ 0 & 0 & u_{23}\sqrt{d_3} & u_{24}\sqrt{d_4} & u_{22}\sqrt{d_2} & u_{25}\sqrt{d_5} \\ 0 & 0 & 0 & 0 & 0 & u_{55}\sqrt{d_5} \end{bmatrix}$$

$$\times \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & c & 0 & s & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -s & 0 & c & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} u_{11}\sqrt{d_1} & 0 & x & u_{14}\sqrt{d_4} & x & u_{15}\sqrt{d_5} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & x & u_{34}\sqrt{d_4} & x & u_{35}\sqrt{d_5} \\ 0 & 0 & 0 & u_{44}\sqrt{d_4} & 0 & u_{45}\sqrt{d_5} \\ 0 & 0 & 0 & u_{24}\sqrt{d_4} & x & u_{25}\sqrt{d_5} \\ 0 & 0 & 0 & 0 & 0 & u_{55}\sqrt{d_5} \end{bmatrix} \quad (18)$$

Likewise, T_2 operates on Eq. (18) as follows:

$$\tilde{U}\sqrt{\tilde{D}}T_0T_1T_2 = \begin{bmatrix} u_{11}\sqrt{d_1} & 0 & x & u_{14}\sqrt{d_4} & x & u_{15}\sqrt{d_5} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & x & u_{34}\sqrt{d_4} & x & u_{35}\sqrt{d_5} \\ 0 & 0 & 0 & u_{44}\sqrt{d_4} & 0 & u_{45}\sqrt{d_5} \\ 0 & 0 & 0 & u_{24}\sqrt{d_4} & x & u_{25}\sqrt{d_5} \\ 0 & 0 & 0 & 0 & 0 & u_{55}\sqrt{d_5} \end{bmatrix}$$

$$\times \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & c & s & 0 \\ 0 & 0 & 0 & -s & c & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} u_{11}\sqrt{d_1} & 0 & x & y & y & u_{15}\sqrt{d_5} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & x & y & y & u_{35}\sqrt{d_5} \\ 0 & 0 & 0 & y & y & u_{45}\sqrt{d_5} \\ 0 & 0 & 0 & 0 & y & u_{25}\sqrt{d_5} \\ 0 & 0 & 0 & 0 & 0 & u_{55}\sqrt{d_5} \end{bmatrix} \quad (19)$$

Trivially, since the transformations T_i are orthogonal, the covariance \tilde{P} in Eqs. (12) and (13) remains unchanged:

$$\begin{aligned} \tilde{P} &= \tilde{U}\sqrt{\tilde{D}}T_0T_1T_2(\tilde{U}\sqrt{\tilde{D}}T_0T_1T_2)^t \\ &= \tilde{U}\tilde{D}\tilde{U}^t \end{aligned} \quad (20)$$

The importance of the zeros in the second row and column of Eq. (16) is now readily apparent. Since the insertion of a row and column of zeros into the UD array in Eq. (7) did not affect the individual elements of the covariance in Eq. (8), deletion of the second row and column of zeros from Eq. (16) will have a similar null effect on the individual elements of the covariance \tilde{P} .

III. Column Permutations of SRI Arrays

To demonstrate the technique on SRI arrays, represent an augmented SRI array R with parameter order $R1, R2, R3, R4, R5$ as

$$(R|Z) = \left[\begin{array}{cccc|c} r1_1 & r2_1 & r3_1 & r4_1 & r5_1 & z_1 \\ 0 & r2_2 & r3_2 & r4_2 & r5_2 & z_2 \\ 0 & 0 & r3_3 & r4_3 & r5_3 & z_3 \\ 0 & 0 & 0 & r4_4 & r5_4 & z_4 \\ 0 & 0 & 0 & 0 & r5_5 & z_5 \end{array} \right] \quad (21)$$

where Z is the residual vector. The following identity suggests the justification of incorporating an artificial parameter into an SRI array:

$$\begin{pmatrix} R_x & 0 & R_{xy} \\ 0 & \epsilon & 0 \\ 0 & 0 & R_y \end{pmatrix}^{-1} = \begin{pmatrix} R_x^{-1} & 0 & -R_x^{-1}R_{xy}R_y^{-1} \\ 0 & \epsilon^{-1} & 0 \\ 0 & 0 & R_y^{-1} \end{pmatrix} \quad (22)$$

If there are no intentions of computing the inverse of the SRI array, it is permissible to allow $\epsilon = 0$. Hence, the insertion and subsequent deletion of an artificial row and column of zeros into an SRI array does not affect the information content of the array.

Say it is desired to change the parameter order to $R1, R3, R4, R2, R5$. The first step is to insert a row and column of zeros into the $(R|Z)$ array before the parameter $R5$:

$$\left[\begin{array}{cccc|c} r1_1 & r2_1 & r3_1 & r4_1 & 0 & r5_1 & z_1 \\ 0 & r2_2 & r3_2 & r4_2 & 0 & r5_2 & z_2 \\ 0 & 0 & r3_3 & r4_3 & 0 & r5_3 & z_3 \\ 0 & 0 & 0 & r4_4 & 0 & r5_4 & z_4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & r5_5 & z_5 \end{array} \right] \quad (23)$$

Columns 2 and 5 of Eq. (23) can then be exchanged. This is just a reordering of the design equation

$$\overline{(R|Z)} = \left[\begin{array}{cccc|c} r1_1 & 0 & r3_1 & r4_1 & r2_1 & r5_1 & z_1 \\ 0 & 0 & r3_2 & r4_2 & r2_2 & r5_2 & z_2 \\ 0 & 0 & r3_3 & r4_3 & 0 & r5_3 & z_3 \\ 0 & 0 & 0 & r4_4 & 0 & r5_4 & z_4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & r5_5 & z_5 \end{array} \right] \quad (24)$$

It is desirable to construct an orthogonal transformation T_0 such that $T_0(\overline{R|Z})$ has only zeros in the second row and column. Such a transformation is just a permutation of rows 2 and 5:

$$T_0(\overline{R|Z}) = \left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right]$$

$$\times \left[\begin{array}{cccc|c} r1_1 & 0 & r3_1 & r4_1 & r2_1 & r5_1 & z_1 \\ 0 & 0 & r3_2 & r4_2 & r2_2 & r5_2 & z_2 \\ 0 & 0 & r3_3 & r4_3 & 0 & r5_3 & z_3 \\ 0 & 0 & 0 & r4_4 & 0 & r5_4 & z_4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & r5_5 & z_5 \end{array} \right]$$

$$= \left[\begin{array}{cccc|c} r1_1 & 0 & r3_1 & r4_1 & r2_1 & r5_1 & z_1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & r3_3 & r4_3 & 0 & r5_3 & z_3 \\ 0 & 0 & 0 & r4_4 & 0 & r5_4 & z_4 \\ 0 & 0 & r3_2 & r4_2 & r2_2 & r5_2 & z_2 \\ 0 & 0 & 0 & 0 & 0 & r5_5 & z_5 \end{array} \right] \quad (25)$$

In practice, however, elements $r3_2$ and $r4_2$ would not be moved, since a scratch array would be needed to store these elements. As is evident, the second row can itself be used as a scratch array. It is now a simple matter of retriangularizing Eq. (25) with two slow Givens. The first Givens T_1 operates on rows 3 and 5 in such a way that the array element containing $r3_2$ becomes zero. The second Givens T_2 operates on rows 4 and 5 in such a way that the array element containing $r4_2$ becomes zero. Each of these Givens transformations maps the elements in each column of the rows on which they operate into a linear combination of the same. Elements not consistent with an upper-triangular structure are never created. For example, T_1 operates on Eq. (25) as follows:

$$T_1 T_0 \overline{(R|Z)} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & c & 0 & s & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & \pm s & 0 & \mp c & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\times \begin{bmatrix} r1_1 & 0 & r3_1 & r4_1 & r2_1 & r5_1 & z_1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & r3_3 & r4_3 & 0 & r5_3 & z_3 \\ 0 & 0 & 0 & r4_4 & 0 & r5_4 & z_4 \\ 0 & 0 & r3_2 & r4_2 & r2_2 & r5_2 & z_2 \\ 0 & 0 & 0 & 0 & 0 & r5_5 & z_5 \end{bmatrix}$$

$$= \begin{bmatrix} r1_1 & 0 & r3_1 & r4_1 & r2_1 & r5_1 & z_1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & x & x & x & x & x \\ 0 & 0 & 0 & r4_4 & 0 & r5_4 & z_4 \\ 0 & 0 & 0 & x & x & x & x \\ 0 & 0 & 0 & 0 & 0 & r5_5 & z_5 \end{bmatrix}$$

(26)

Likewise, T_2 operates on Eq. (26) as follows:

$$T_2 T_1 T_0 \overline{(R|Z)} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & c & s & 0 \\ 0 & 0 & 0 & \pm s & \mp c & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\times \begin{bmatrix} r1_1 & 0 & r3_1 & r4_1 & r2_1 & r5_1 & z_1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & x & x & x & x & x \\ 0 & 0 & 0 & r4_4 & 0 & r5_4 & z_4 \\ 0 & 0 & 0 & x & x & x & x \\ 0 & 0 & 0 & 0 & 0 & r5_5 & z_5 \end{bmatrix}$$

$$= \begin{bmatrix} r1_1 & 0 & r3_1 & r4_1 & r2_1 & r5_1 & z_1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & x & x & x & x & x \\ 0 & 0 & 0 & y & y & y & y \\ 0 & 0 & 0 & 0 & y & y & y \\ 0 & 0 & 0 & 0 & 0 & r5_5 & z_5 \end{bmatrix}$$

(27)

From Eq. (22) it is a trivial matter that the second row and column of zeros of Eq. (27) can then be deleted.

IV. Cyclic Permutations

Since the artificial parameter is only a convenient place holder, an alternative method can be constructed that does not incorporate the insertion/deletion step. By cyclically permuting the rows and columns of either the UD or SRI arrays, a data structure similar to Eq. (17) or Eq. (25), respectively, can be achieved. For example, cyclically left permuting columns 2 through 4 of Eq. (21) yields the following:

$$\begin{bmatrix} r1_1 & r3_1 & r4_1 & r2_1 & r5_1 & z_1 \\ 0 & r3_2 & r4_2 & r2_2 & r5_2 & z_2 \\ 0 & r3_3 & r4_3 & 0 & r5_3 & z_3 \\ 0 & 0 & r4_4 & 0 & r5_4 & z_4 \\ 0 & 0 & 0 & 0 & r5_5 & z_5 \end{bmatrix} \quad (28)$$

Cyclically upward permuting rows 2 through 4 of Eq. (28) then yields

$$\begin{bmatrix} r1_1 & r3_1 & r4_1 & r2_1 & r5_1 & z_1 \\ 0 & r3_3 & r4_3 & 0 & r5_3 & z_3 \\ 0 & 0 & r4_4 & 0 & r5_4 & z_4 \\ 0 & r3_2 & r4_2 & r2_2 & r5_2 & z_2 \\ 0 & 0 & 0 & 0 & r5_5 & z_5 \end{bmatrix} \quad (29)$$

This is the desired data structure of Eq. (25) without the artificial parameter. Slow Givens sweeps across rows 2,4 and rows 3,4 of Eq. (29) can then be applied to zero out the array elements containing $r3_2$ and $r4_2$, respectively. Since the size of the array never changes, this method would not require that the array be augmented with an extra column. However, one explicit scratch array is required to hold that part of Eq. (29) not consistent with an upper-triangular data structure.

A similar approach can also be taken toward permuting UD arrays. Cyclic permutations of the rows and columns would yield the data structure in Eq. (17) without the artificial parameter:

$$\begin{bmatrix} u1_1\sqrt{d_1} & u1_3\sqrt{d_3} & u1_4\sqrt{d_4} & u1_2\sqrt{d_2} & u1_5\sqrt{d_5} \\ 0 & u3_3\sqrt{d_3} & u3_4\sqrt{d_4} & 0 & u3_5\sqrt{d_5} \\ 0 & 0 & u4_4\sqrt{d_4} & 0 & u4_5\sqrt{d_5} \\ 0 & u2_3\sqrt{d_3} & u2_4\sqrt{d_4} & u2_2\sqrt{d_2} & u2_5\sqrt{d_5} \\ 0 & 0 & 0 & 0 & u5_5\sqrt{d_5} \end{bmatrix} \quad (30)$$

Fast Givens sweeps down columns 2,4 and columns 3,4 of Eq. (30) can then be applied to zero out the array elements containing $u2_3\sqrt{d_3}$ and $u2_4\sqrt{d_4}$, respectively.

V. Implementation

When dealing with large arrays on a virtual memory machine, computations should be arranged to minimize expensive page faulting. The Givens sweeps should access contiguous storage locations. The usual column ordering of a vector-stored upper-triangular data structure with n rows and n columns is as follows:

$$S_{\text{column}} = \begin{bmatrix} s(1) & s(2) & s(4) & \dots & s\left(\frac{n^2-n}{2} + 1\right) \\ 0 & s(3) & s(5) & \dots & s\left(\frac{n^2-n}{2} + 2\right) \\ 0 & 0 & s(6) & \dots & s\left(\frac{n^2-n}{2} + 3\right) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & s\left(\frac{n^2+n}{2}\right) \end{bmatrix} \quad (31)$$

Element $S_{\text{column}}(i, j)$ is easily referenced as $s[(j^2 - j/2) + i]$. For UD arrays, the Givens sweeps down the columns using the data storage in Eq. (31) would require minimal paging. For SRI arrays, the Givens sweeps across the rows with the data storage in Eq. (31) would be paging intensive. Hence, for SRI arrays, it proves to be more efficient to store the upper-triangular array with unnatural row ordering as follows:

$$S_{\text{row}} = \begin{bmatrix} s(1) & s(2) & s(3) & \dots & s(n) \\ 0 & s(n+1) & s(n+2) & \dots & s(2n-1) \\ 0 & 0 & s(2n) & \dots & s(3n-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & s\left(\frac{n^2+n}{2}\right) \end{bmatrix} \quad (32)$$

Element $S_{\text{row}}(i, j)$ can be referenced as $s[j + (i^2 - i/2) + (i-1)(n-i)]$.¹⁰ Since filtering software generally stores

¹⁰ $S_{\text{row}}(i, j) = S_{\text{column}}(i, j)$; this is just a relabeling of the elements.

upper-triangular arrays according to Eq. (31), it would be necessary to relabel from S_{column} to S_{row} before the permutations and then back to S_{column} afterward.

It is equally easy to perform cyclic permutations with either Eq. (31) or Eq. (32). In the former, movement is along the columns; in the latter, movement is across the rows. The only paging-intensive step in cyclically permuting the rows and columns with data storage S_{column} (S_{row}) is in storing the upper row (left column) into a scratch array.

VI. Performance

To assess the permutation methods on UD arrays, an upper-triangular array was constructed with 499 parameters and an estimate column. The following permutations were then performed on a VAX 11/780. First, the parameters were permuted randomly. In practice, of course, this would never happen. It is more realistic to move blocks of parameters to the bottom of the array. Therefore, the following systematic permutations were performed. A block of 10 parameters starting in rows 1, 167, and 334 were moved to the bottom of the array. These permutations were then repeated with a block of 100 parameters. CPU time and page faults were accumulated for three different methods. First, the ESL routine HHPOST was used to apply a post-Householder to a row-reordered column-scaled UD array. Second, pairwise permutations were performed with suboptimal noise-free measurement updates. This is equivalent to the ESL routine U2U. Third, the permutations were performed with the discussed method of cyclic permutations with fast Givens sweeps. The results are presented in Table 1.

The ESL method with HHPOST performs the same regardless of the number and type of permutations performed. The performance of the other methods depends on the number and type of permutations. The cyclic permutation method with fast Givens sweeps is shown to be superior to the ESL method in U2U. Furthermore, in the cyclic permutation method, as the starting row moves down there are fewer elements to retriangularize.

The ESL has standard routines to perform permutations on SRI arrays. The routine R2A copies and reorders the columns of an upper-triangular array into a two-dimensional work array. A Householder transformation can then be applied with routine TDHHT to retriangularize back to an upper-triangular array.

To assess the permutation methods on SRI arrays, ESL routines R2A/TDHHT, the discussed method of cyclic permutations with slow Givens sweeps using data storage S_{column} , and the same with data storage S_{row} were used to permute an upper-triangular array with 499 parameters and a residual column. CPU time and page faults were accumulated for permuting, retriangularizing, and any necessary relabeling. The results are presented in Table 2.

The ESL method performs the same regardless of the number and type of permutations performed. The performance of the cyclic permutations with slow Givens sweeps depends on the number and type of permutations. As the starting column is moved to the right, not only are there fewer elements to retriangularize, but the lengths of Givens sweeps also become smaller, and fewer computations are required. Since the VAX 11/780 is a virtual memory machine, the method with data storage S_{row} is generally superior to the method with data storage S_{column} . For fixed memory machines or small SRI arrays, the method with data storage S_{column} should be superior, since no relabeling or unlabeled is required.

VII. Conclusion

The most efficient method to permute UD (SRI) arrays has been shown to be cyclic permutation with fast (slow) Givens sweeps for retriangularization. Since the only difference is in the retriangularization process, it has been possible to combine the methods into one subroutine. This method has been incorporated into the OASIS software [7] and should prove to be advantageous for GPS, TOPEX, and DSN high-earth-orbiter data processing and analysis.

References

- [1] G. J. Bierman, *Factorization Methods for Discrete Sequential Estimation*, New York: Academic Press, pp. 139–140, 1977.
- [2] G. J. Bierman and K. H. Bierman, *Estimation Subroutine Library User Guide*, Studio City, California: Factorized Applications, Inc., August 1984 (revised August 1985).
- [3] J. L. Fanselow and O. J. Sovers, *Observation Model and Parameter Partial for the JPL VLBI Parameter Estimation Software MASTERFIT*, JPL Publication 83-89, Rev. 3, Jet Propulsion Laboratory, Pasadena, California, December 15, 1987.
- [4] M. Morf and T. Kailath, "Square-Root Algorithms for Least-Squares Estimation," *IEEE Trans. Automatic Control*, vol. AC-20, no. 4, pp. 487–497, August 1975.
- [5] J. J. Dongarra, J. R. Bunch, C. B. Moler, and G. W. Stewart, *LINPACK User's Guide*, Philadelphia: SIAM, pp. 10.15–10.16, 1979.
- [6] J. W. Daniel et al., "Reorthogonalization and Stable Algorithms for Updating the Gram-Schmidt QR Factorization," *Math. Comp.*, vol. 30, pp. 772–795, 1976.
- [7] S. C. Wu and C. L. Thornton, "OASIS – A New GPS Covariance and Simulation Analysis Software System," in *Proceedings of the First International Symposium on Precise Positioning With the Global Positioning System*, pp. 337–346, April 1985.

Table 1. CPU and paging faults in permuting UD arrays

Permutations	CPU time; thousands of page faults		
	ESL HHPOST	ESL U2U	Cyclic permutations and fast Givens
Random	0:33:06 1025	1:06:31 2051	0:17:04 612
Rows 1-10	0:32:43 948	0:48:55 1823	0:01:24 40
Rows 167-176	0:31:49 970	0:40:00 1338	0:01:14 35
Rows 334-343	0:31:12 970	0:24:26 728	0:00:45 21
Rows 1-100	0:32:44 968	0:59:25 1985	0:11:23 363
Rows 167-266	0:31:32 946	0:48:47 1433	0:08:46 275
Rows 334-433	0:30:59 973	0:27:29 747	0:02:20 52

Table 2. CPU and paging faults in permuting SRI arrays

Permutations	CPU time; thousands of page faults		
	ESL TDHHT	Cyclic permutations and slow Givens using data storage S_{column}	Cyclic permutations and slow Givens using data storage S_{row}
Random	0:22:28 729	0:41:09 3253	0:20:07 575
Columns 1-10	0:20:29 717	0:01:36 48	0:01:38 55
Columns 167-176	0:21:13 723	0:00:55 31	0:01:03 41
Columns 334-343	0:21:53 719	0:00:20 11	0:00:39 27
Columns 1-100	0:21:25 727	0:13:11 424	0:11:10 411
Columns 167-266	0:21:15 714	0:06:14 229	0:05:14 205
Columns 334-433	0:21:03 723	0:00:51 52	0:01:43 104

1989000836
618448
17p.

521-17-

~~165532~~

KLR

April-June 1988

N89-10207

JPL

Phasing the Antennas of the Very Large Array for Reception of Telemetry From Voyager 2 at Neptune Encounter

J. S. Ulvestad

Tracking Systems and Applications Section

The Very Large Array (VLA) radio telescope is being instrumented at 8.4 GHz to receive telemetry from Voyager 2 during its encounter with Neptune in 1989. This article examines the procedure in which the 27 antennas have their phases adjusted in near real time so that the signals from the individual elements of the array can be added coherently. Calculations of the expected signal-to-noise ratio, tests of the autophasing (or self-calibration) process at the VLA, and off-line simulations of that process are all presented in this article. Various possible procedures for adjusting the phases are considered. It is shown that the signal-to-noise ratio at the VLA is adequate for summing the signals from the individual antennas with less than 0.1 dB of loss caused by imperfect coherence among the antennas. Tropospheric variations during the summer of 1989 could cause enough loss of coherence to make the losses higher than 0.1 dB. Experiments show that the losses caused by the troposphere can probably be kept below 0.2 dB if the time delay inherent in the phase adjustment process is no longer than ~5 seconds. This relatively small combining loss meets the goal established in order to minimize the bit error rate in the Voyager telemetry and implies that the autophasing of the VLA should be adequate for reception of telemetry from Voyager 2 at its Neptune encounter.

I. Introduction

The Very Large Array (VLA) will be used in 1989 to receive telemetry from Voyager 2 when it encounters Neptune. The VLA, an array of 27 telescopes located on the Plains of San Agustin in New Mexico, is described in detail in [1]. Several years ago, it was identified as a potentially important contributor to the reception of telemetry from the Voyager 2 spacecraft at Neptune [2]. Work is currently proceeding to equip all

28 VLA antennas (27 in operation plus one in the maintenance cycle) for signal reception at 8.4 GHz and to design the optimal system for telemetry reception from Voyager. One of the crucial issues is the best procedure for keeping all 27 operational antennas in phase with one another so that the signals received by those antennas can be added coherently. The individual antenna phases will be adjusted in near real time so as to keep the summed signals coherent by maximizing the interferometric fringe amplitude of the Voyager signal. Preliminary

tests of this real-time self-calibration procedure (hereafter referred to as "autophasing") and a discussion of some of the issues involved can be found in [3]. Several key questions remain to be answered more completely. Specifically, those questions are as follows:

- (1) Is the signal-to-noise ratio on the Voyager signal sufficient to allow the autophasing process to work?
- (2) How well will the "global" autophasing procedure improve the signal-to-noise ratio in comparison with the previous method for determining antenna phases? (The difference between the two methods is described in Section IIA.)
- (3) To what extent can the summer troposphere corrupt the phases?
- (4) Given the constraints of both the troposphere and the expected signal-to-noise ratio, what is the optimal algorithm for phase correction that can be easily implemented to work in near real time?

For Voyager, the goal of the autophasing process has been to keep the combining loss for 27 VLA antennas below 0.2 dB. Such a loss would have minimal effect on the bit error rate of the Voyager telemetry. The purpose of the work described in this article is twofold: to address the above questions by means of real and simulated autophasing and to determine the best way to meet the goal of less than 0.2 dB of combining loss.

II. Autophasing Procedure

A. Description

The autophasing process is conceptually fairly simple. It will be described briefly here, with details about the implementation of models and corrections in VLA fringe rotators being largely ignored. A point-like celestial radio source, either a natural source or an artificial source such as the Voyager spacecraft, is observed by the individual VLA antennas. During the observations, standard phase corrections are implemented at each antenna based on a geometric model. For each pair of antennas, observed fringe phases after the application of that model are found from the cross-correlated data. The on-line computers then compute residual phases for each individual antenna. The residuals may be caused by a variety of problems, such as errors in the radio source position, imperfect refraction models, and instabilities in electronics. Most such problems give rise to slowly varying phases that are easily corrected, so the most stringent autophasing "problem" is that of trying to correct for a rapidly varying troposphere along each antenna's line of sight. The residual phases are used as error signals to derive phase corrections. These corrections are

applied to a local oscillator at each antenna in order to keep the phase of that antenna near zero relative to the other antennas. If the corrections are successful, the VLA will be "phased up," and the individual antenna signals will sum coherently.

In the past, the VLA phasing algorithm has used only one baseline in computing the residual phase for each antenna. The phases on the baselines from each antenna to a single reference antenna have been used to compute the individual antenna phases in an attempt to bring the phases together with that of the reference antenna. Since not all possible baselines are used in the phase computations, the highest signal-to-noise ratio (SNR) that might be available is not achieved. With the VLA being controlled by a new computer system that was brought on line in January 1988, the individual phases are computed "globally" via a least-squares algorithm that uses all available baselines to solve for the individual antenna phases. This improves the determination of the individual antenna phases, as described more fully below.

The determination of residual antenna phases is only part of the autophasing process. The other important aspect of the procedure is the correction algorithm by which these residual phases are "fed back" to the local oscillators of each antenna. It is important to recognize a key limitation of this feedback loop: namely, it takes a period of time for the residual phases to be computed and for phase corrections to be communicated to the individual antennas, *and the VLA must continue observing during this time*. Therefore, there is a delay before any correction can be applied to the antennas, and data taken during the delay interval will have a somewhat outdated phase correction applied. In previous implementations of autophasing at the VLA, a 20-second delay has existed between the end of one 10-second integration and the beginning of the 10-second integration in which the phase information from the earlier integration is used. Taking an average time at the middle of each integration period, the net delay in this procedure is approximately 30 seconds. (Also see Section IIC below.) Any correction or feedback loop acts as a filter that corrects for low-frequency tropospheric fluctuations but misses the highest-frequency fluctuations. Minimizing both the integration time and delays in the feedback loop should help correct more of the higher-frequency fluctuations and reduce the combining loss when a large number of antenna signals are added.

B. Signal-to-Noise Calculations

1. **General considerations.** One can calculate the SNR of a single interferometer pair in a variety of ways. An extensive description of such a computation for the VLA is given in [4], and the calculation made here largely follows that reference. The application to Voyager is fairly straightforward, with one

major exception. The applicable case for almost all of radio astronomy is that of a source with no significant circular polarization, whereas the signal from Voyager is 100 percent circularly polarized. Therefore, the antenna temperature (or signal strength) contributed by Voyager in a single circularly polarized channel is a factor of 2 higher than that for an unpolarized source having the same total power.

The noise power of a single antenna at the input to the correlator is

$$P_N = k_B T_{\text{sys}} \Delta\nu G \quad (1)$$

where k_B is Boltzmann's constant, T_{sys} is the system temperature of the antenna, $\Delta\nu$ is the bandwidth, and G is the power gain of the antenna. When two such antennas (whose noises are the results of independent, random processes) are correlated, the variance of the noise power coming out of the correlator is 1/2 the product of the powers divided by the number of independent samples in the correlator:

$$P_{\text{rms}}^2 = \frac{P_N^2}{2\Delta\nu\Delta t} \quad (2)$$

where Δt is the integration time. Therefore,

$$P_{\text{rms}} = \frac{k_B G T_{\text{sys}} \sqrt{\Delta\nu}}{\sqrt{2\Delta t}} \quad (3)$$

The correlated signal power from a source contributing an antenna temperature T_a is $P_a = \eta_c k_B G T_a \Delta\nu$, and the antenna temperature contributed by Voyager, a 100 percent circularly polarized source, is $T_a = \eta_a A P_V / (k_B \Delta\nu)$. Here, η_c is an efficiency factor including the effect of the correlator sampling and the duty cycle of the observations; η_a is the antenna aperture efficiency; A is the physical area of the antenna; and P_V is the spacecraft power per unit area at Earth. Combining all these expressions, we arrive at a formula for the SNR (R) for Voyager:

$$R \equiv \frac{P_a}{P_{\text{rms}}} = \frac{\eta_c \eta_a A P_V \sqrt{2\Delta t}}{k_B T_{\text{sys}} \sqrt{\Delta\nu}} \quad (4)$$

A small correction for atmospheric attenuation, which is dependent on the weather and on the elevation angle of the spacecraft during the observations, should also be applied but is assumed to be negligible in deriving Eq. (4).

2. Standard parameters and assumptions. We use the following standard parameters for the Voyager observations at the VLA:

$$\eta_c = 0.79$$

$$\eta_a = 0.62 \pm 0.03$$

$$T_{\text{sys}} = 35_{-4}^{+6} \text{ K at a 30-degree elevation angle}$$

$$P_V = 5.0 \pm 0.4 \times 10^{-21} \text{ W/m}^2 \text{ at Neptune}$$

$$\Delta\nu = 8 \text{ MHz}$$

$$\Delta t = 10 \text{ seconds}$$

$$\text{zenith opacity} = 0.010_{-0.0015}^{+0.008} \text{ neper}$$

Using the above values in Eq. (4), we find a Voyager SNR of $3.90_{-0.76}^{+0.58}$ for a single VLA baseline. Some of the errors and quantities quoted above are taken from [5], while others were estimated from unpublished data acquired by the author while performing tests at the VLA.

The probability distribution of the phase errors on each baseline has a well-known dependence on the SNR (see [4]). At a moderately high SNR, that distribution approaches a Gaussian distribution whose standard deviation is

$$\sigma_\phi = \frac{1}{R} \quad (5)$$

For $R \approx 3.9$, the actual probability of a given phase error differs from the value given by the Gaussian approximation by ≤ 3 percent for phase errors with a substantial probability of occurrence, so the use of Eq. (5) gives a "typical" phase error within a few percent of the value derived from the exact probability distribution. In the case of a perfect autophasing algorithm (or no tropospheric phase problems), the rms residual phase after autophasing should be equal to the rms phase errors predicted by the limiting SNR. For $R = 3.90_{-0.76}^{+0.58}$, the predicted rms residual phase for each antenna relative to a user-selected reference antenna would be $0.26_{-0.04}^{+0.05}$ radian, or $14.7_{-2.2}^{+2.9}$ degrees. In general, at the VLA, there is no absolute knowledge of the phase of any individual antenna. Therefore, the phase at the user-selected reference antenna is *defined* to be zero, and all other antenna phases are determined relative to that value.

3. Scaling with elevation angle (air mass). Using the results given in [2], a typical X-band atmospheric temperature at the VLA is found to be ~ 2.73 K per air mass. We will estimate errors of $+1.0$ K and -0.15 K per air mass (already incorporated in the T_{sys} uncertainty at 30-degree elevation that was given above). These errors are estimates based on the variations

seen in the data from tipping curves taken under various weather conditions at the VLA. Large positive excursions in atmospheric temperature can occur during the summer thunderstorm season, but there are not enough data to permit the statistical distribution of the summer atmospheric temperatures to be accurately determined. Relative to the 30-degree elevation value, the additional T_{sys} at 20-degree elevation will be $2.52^{+0.92}_{-0.14}$ K. At 10-degree elevation, the additional temperature will be $10.26^{+3.76}_{-0.56}$ K. The opacity caused by the atmosphere will be assumed to be as quoted above, $0.010^{+0.008}_{-0.0015}$ neper per air mass (again based on VLA tipping data). As with system temperature, the higher positive error in the opacity partially accounts for the possibility of thunderstorms at the VLA. However, the most severe thunderstorms may give greater increases in system temperature and opacity than are assumed here.

Crane¹ recently measured gain curves for VLA antennas at 22 GHz. He found significant changes at low elevations which may be caused by antenna deformations and/or failures of the antennas to track changing foci. From Fig. 2 of Crane,¹ we estimate median gain losses at 22 GHz of 5 percent in going from 30-degree to 20-degree elevation and 11 percent in going from 30-degree to 10-degree elevation. Assuming that the losses are dominated by antenna deformation or pointing errors, these losses should be reduced at X band. We estimate (conservatively) that the gain changes scale inversely with wavelength. Therefore, at X band, the expected gain losses in going to 20-degree and 10-degree elevation will be 2 percent and 4 percent, respectively. Errors of +50 percent and -25 percent are assigned to the degradation at each elevation angle, with the larger positive error caused by the increased spread in the antenna gain losses at the lowest elevation angles.

4. Changing integration times. Because an integral number of phase-switching cycles must be included in each unit VLA integration time, the allowed integration times are quantized in units of a single complete phase-switching cycle. This cycle time is $1\frac{2}{3}$ seconds (32 of the 19.2-Hz VLA waveguide cycles). Therefore, allowed integration times, in seconds, are $1\frac{2}{3}$, $3\frac{1}{3}$, 5, and so on. Integration times longer than 10 seconds will not be considered in this work. The SNR per baseline scales as the square root of the integration time.

5. Changing bandwidths. The SNR scales as $1/\sqrt{\Delta\nu}$. In addition to the nominal value of 8 MHz, a 6-MHz bandwidth will briefly be considered below. It will be assumed that all significant telemetry sideband power can be accommodated

in the 6-MHz bandwidth, so the Voyager signal power will not be reduced for the narrower bandwidth. At worst, this may give an error of ~ 0.1 dB, a value considerably smaller than the 0.3-dB uncertainty quoted for the Voyager power above.

6. Global autophasing. In the case of global autophasing, every baseline in the array is used to compute the phases of the individual antennas. Theoretical work (e.g., [6]) shows that the determination of individual antenna phases via a least-squares calculation will give rms errors that are reduced from the single-baseline values by $\sqrt{N-2}$, where N is the number of antennas. Simulations² and actual VLA observations (see below) show that this scaling may overestimate the accuracy of the phase determination by ~ 25 to 30 percent. It will be assumed that the rms phase improvement given by global autophasing is $(0.7 \pm 0.1)\sqrt{N-2}$.

7. Calculations of expected phase errors. Table 1 summarizes the various parameters and scaling laws described above. Using the numbers quoted in that table and assuming global autophasing, we find the following formulas for the rms phase errors at three different elevation angles:

$$\sigma_{\phi}(30^{\circ}) \approx \left(4.38^{+1.06}_{-0.90}\right) \times \sqrt{\frac{10}{\Delta t}} \sqrt{\frac{23}{N-2}} \quad (6a)$$

$$\sigma_{\phi}(20^{\circ}) \approx \left(4.88^{+1.23}_{-0.99}\right) \times \sqrt{\frac{10}{\Delta t}} \sqrt{\frac{23}{N-2}} \quad (6b)$$

$$\sigma_{\phi}(10^{\circ}) \approx \left(6.18^{+1.73}_{-1.23}\right) \times \sqrt{\frac{10}{\Delta t}} \sqrt{\frac{23}{N-2}} \quad (6c)$$

Table 2 gives the rms phases predicted *only from the SNR* for different elevations and integration times. For the global autophasing, it is assumed that 25 VLA antennas are operational. The right-hand column gives the expected values for each set of parameters. The predicted phase residuals for the highest SNRs in the table are not far above the rms phase noise of ~ 2 degrees on time scales of ~ 5 to 10 seconds for the VLA. (This phase noise is determined by factors such as the quantization in the lobe rotators.) All phases in the right-hand column of the table will give combining losses much less than 0.2 dB (see Section V), so the signal-to-noise ratio for short integrations on the Voyager spacecraft is adequate to give phase determinations of very high accuracy in the case of global autophasing. Narrowing the filter bandwidth to 6 MHz would give a slightly higher SNR and predicted rms phases about 15 percent lower than those in the table, but the calcu-

¹P. C. Crane, "Measurements of Flux Densities and Gain Corrections at 22485.1 MHz," VLA Test Memorandum No. 149 (internal document), Socorro, New Mexico: National Radio Astronomy Observatory, April 1987.

²F. R. Schwab, "Robust Solution for Antenna Gains," VLA Scientific Memorandum No. 136 (internal document), Socorro, New Mexico: National Radio Astronomy Observatory, September 1981.

lations show that the decreased observing flexibility of the narrower filters is not necessary to give an adequate SNR.

C. Tropospheric Fluctuations and the Autophasing Feedback Loop

The rms phase computed above is that which would be expected for observations of the Voyager spacecraft through a static troposphere. In fact, the troposphere over each antenna can be quite active. Water vapor cells of various sizes blow over the antennas, changing the effective path length, and hence the observed phase, at each antenna. (See [7] and [8] for more detailed theoretical and empirical treatments of this phenomenon.) The small-scale disturbances blow across an antenna quickly, giving rise to rapid phase fluctuations. These high-frequency fluctuations are the ones that are not adequately corrected by the autophasing feedback loop. Thus, the residual phases after the autophasing process can be much larger than would be predicted just from the signal power and the individual antenna characteristics.

When the VLA is being autophased, the following feedback loop has been implemented for correcting the residual antenna phases:³

$$\phi_{\text{corr}}^{t+1} = \phi_{\text{corr}}^t + \alpha \phi_{\text{res}}^{t-k} \quad (7)$$

where ϕ_{corr}^t is the phase correction applied during the t th integration period, ϕ_{corr}^{t+1} is the correction to be applied in the $(t+1)$ th integration interval, ϕ_{res}^{t-k} is the residual phase found in the $(t-k)$ th integration period, and α is the gain of the feedback loop. Strictly speaking, this formula is correct only for the "standard" VLA integration time of 10 seconds, because the exact nature of the time delays in the feedback loop is related to the internal data-handling procedures and to the frequency with which commands updating the local oscillator frequencies are sent to the antennas. For integration times of 10 seconds, the standard VLA phasing algorithm has used $k=2$ and $\alpha=0.25$. Note that this feedback loop is different from that described in the Appendix of [3], where the crucial built-in time delay is not taken into account.

Equation (7) can be rewritten in more generality by replacing ϕ_{corr}^t with the correction phase at any past integration period $(t-j)$. The index j can take on integer values 0, 1, 2, and so on. A more general expression for the feedback loop is then

$$\phi_{\text{corr}}^{t+1} = \phi_{\text{corr}}^{t-j} + \alpha \phi_{\text{res}}^{t-k} \quad (8)$$

³G. Hunt, interoffice memorandum to B. Brundage (internal document), National Radio Astronomy Observatory, Socorro, New Mexico, March 3, 1987.

In this expression, j and k need not take on the same values, although they may be the same if desired.

In [9] (e.g., see Fig. 9 of [9]), it is shown that the autocorrelation function of VLA phase data that are corrupted by the troposphere generally decreases with increasing time lag. As can be seen from the above formulas, there is a considerable time delay built into the feedback algorithm. If the time delay were reduced, more recent phase data from a higher part of the autocorrelation function could be used to generate the corrections of the autophasing process. That should reduce the residual phase substantially in cases where the tropospheric fluctuations dominate the phase errors. One of the major goals of the tests described below has been to determine how much the residual phases could be reduced when compared with the current feedback process.

III. Autophasing Observations of the Voyager Spacecraft

The VLA antennas are gradually being equipped with receivers at X band (8.4 GHz) in order to receive Voyager telemetry. Significant numbers of antennas were not instrumented until late 1986. As part of the monthly JPL tests during 1987, substantial effort was devoted to various tests related to autophasing. This section describes the portion of tests that involved the standard autophasing of the VLA in real time, accomplished by observing Voyager at X band through 8-MHz bandpass filters.

The observations in the spring of 1987 were made when Voyager was transmitting at low power, ~ 2.0 dB below its high-power mode. However, the spacecraft was also closer to Earth than it will be at Neptune. In March, April, and May, the spacecraft transmitter power should have been ~ 2.6 dB higher than the low-power values will be at Neptune. Therefore, the SNR on each baseline should have been ~ 4.4 , which would translate into an rms phase residual of 13 degrees. The respective measured values for the rms residual phase in March, April, and May 1987 were 15.0, 16.7, and 15.6 degrees, which would correspond to an average SNR of ~ 3.6 , ~ 0.8 dB less than the highest possible expected value.

Although the observed phase residuals were slightly higher than those expected based on simple SNR considerations, it is quite likely that they were dominated by the system noise. Several factors could have reduced the average SNR below ~ 4.4 : there is some uncertainty in the Voyager power; the pointing of the Voyager antenna is not updated as frequently during the cruise mode as it is near an encounter; the pointing of the VLA antennas was not completely optimized; several antennas had higher noise receivers than the production

receivers which use high-electron-mobility transistors (HEMTs); and two antennas were partially shadowed during some of the observations. Estimated errors in the Voyager power and the power losses due to spacecraft and ground antenna pointing were in the 0.1- to 0.5-dB range. The antennas lacking HEMT amplifiers were ~ 0.8 dB noisier than those with such amplifiers. The partial shadowing also caused the SNR to be reduced by ~ 0.2 to 0.3 dB on some baselines. Further, the fact that the autophasing feedback algorithm was not optimal for the SNR-limited case may have added as much as ~ 1 degree to the phase residuals (see Section IVA). The combination of these effects can easily account for the 0.8-dB (or ~ 3 degrees of phase) decrement relative to the optimal SNR. The closeness of the measured values to the predicted values, taken together with the factors that would serve to reduce the SNR slightly, implies that the SNR calculation models the actual observations quite well.

IV. Autophasing Simulations

Various off-line simulations have been performed to validate results obtained with the VLA autophasing process and to explore parameter space a bit more thoroughly. These simulations have used a variety of values for the gain and for the number of periods of delay (α , j , and k in Eq. [8]) in the feedback loop. Because of constraints set by the data-handling procedures at the VLA, not all of the feedback loop parameters considered are actually usable at the VLA. However, they serve to indicate the improvement that may be expected by decreasing the integration and delay times. The autophasing simulations have been applied to phase data obtained by the VLA in its "normal" interferometry mode—i.e., with no autophasing applied on line. This section describes the results of the off-line simulations.

A. Observations of Voyager: Single-Antenna-Referenced Phase Computations

The Voyager spacecraft was observed through the 8-MHz filters in normal interferometer mode on several occasions during 1987. The phases relative to a specified reference antenna, which were exactly those that would have gone into the feedback algorithm in real time during autophasing at the VLA, were fed into a simulation program instead. The data in March and May were taken with 10-second integration periods. Simulations were run for three different values of the gain in the feedback loop: 0.125, 0.25, and 0.5. Another set of simulations was run for the same three gains, but with a time delay of only one integration ($k = 1$ in Eq. [8]). All these simulations used the correction phase from the last integration period in the recursive formulation of Eq. (8) (i.e., $j = 0$). The rms residual phases for both sets of simulations are shown in Table 3.

Three important points should be noted from Table 3. First, the residual phases for a gain of 0.5 are very much lower for a time delay of only one integration than for a delay of two integrations. The high residuals for a feedback loop with a delay of two integration periods seem to be caused by the fact that the loop is marginally stable for a gain of 0.5; this is confirmed both by simulations using faked data and by analytical results found by J. W. Armstrong (private communication). Second, the results for a gain of 0.25 and a delay of two integration periods are very close to the values obtained on the same dates using the VLA in its autophasing mode (see Section III). This provides confidence that the simulations are working properly. Third, the simulations with a gain of 0.125 provided slight improvement over those with a gain of 0.25, since the lower gain gives a better performance in the SNR-limited case.

B. Observations of Voyager: Global Phase Computations

It is expected that the global phase computations will increase the SNR in the autophasing, as described in Section II. There were seven antennas used in March 1987 and nine in May 1987. Theoretical analysis implies that global phase computations should have improved the SNR by $\sqrt{5} = 2.24$ in March and by $\sqrt{7} = 2.65$ in May. As mentioned in Section II, these theoretical improvements should probably be reduced by ~ 25 percent, to ~ 1.7 and ~ 2.0 . Simulations of global autophasing were run on the same data used in the analysis described in the previous subsection, and with the same choices of parameters in the feedback loop. The VLA program ANTOSOL was first run in the off-line computers to compute the phases globally. Then, those phases were fed into the autophasing simulation programs. Results are shown in Table 4.

Several comments are in order regarding Table 4. For the March data, the rms phase residuals for the four cases with gains of 0.25 or less were reduced by factors between 1.71 and 1.80 when the global phase calculations were used. (The simulations with a gain of 0.5 have contributions from algorithm instability that are too large to allow for the reliable estimation of the improvement caused by better SNR.) These improvement factors are close to the predicted value of ~ 1.7 . The actual improvement in global autophasing indicated by the May data was only a factor of ~ 1.4 , substantially less than the predicted improvement. The lack of improvement in the May data is probably attributable to weather that was significantly worse than that in March, with the troposphere-induced phase problems becoming more important as a limiting factor when the SNR was improved by the global phase calculations.

In November 1987, data were again taken on the Voyager spacecraft in the normal interferometer mode. These data

were acquired using 14 X-band antennas, all equipped with HEMT receivers. Since the troposphere is relatively quiet during this time of year, these data were expected to be SNR-limited, as the March data had been. The increased number of antennas available meant that the global phase computation could be expected to give a larger improvement than it did with only seven or nine antennas. In fact, the expected improvement factor was ~ 2.6 , whereas the real improvement factors ranged from 1.7 to 2.0. Results are shown in Table 5.

Since the Voyager spacecraft was transmitting in its high-power mode (accounting for the generally lower phase residuals in November when compared to March), the SNR was higher on each baseline. This means that the residual phases were low enough so that the tropospheric "noise" would make a more important contribution and might create a "floor" above that expected from simple SNR considerations. There is some evidence for this effect in simulations using the last previous phase to predict the next phase (not actually possible in the real VLA because of finite computation times). As shown in the bottom line of Table 5, this gives an improvement factor of 2.4, much closer to that predicted. Since only the most recent phase point is used in this case, the effective delay is reduced, and the troposphere should be followed better. However, it is not yet clear whether the troposphere is really at fault or if the actual improvement from the global phase computations will be somewhat less than that expected. Therefore, an improvement factor of $(0.7 \pm 0.1)\sqrt{N-2}$ has been used in Section III and Table 2 rather than the factor of $0.8\sqrt{N-2}$, which is more in line with the results given in VLA Scientific Memorandum No. 136 (see footnote 2).

C. Summertime, Low-Elevation Observations at a High Signal-to-Noise Ratio

The worst tropospheric effects should occur during the summer, during observations of low-elevation sources. The tests run during July 1987 were designed for the observation of strong point radio sources at a variety of elevation angles that are characteristic of the Voyager spacecraft. Strong sources were observed in order to ensure that the phase errors were dominated by the troposphere rather than by limited SNR. For 5-second integrations on sources with several janskys of flux density ($1 \text{ jansky} = 10^{-26} \text{ W/m}^2/\text{Hz}$), observing through a 50-MHz bandwidth, the SNR was ~ 100 to 1 on each baseline. Normal interferometry observations were made of several sources during the 4 1/2-hour test. As stated previously, the fact that the autocorrelation of the phase fluctuations is larger for shorter time lags implies that the shorter integrations should help reduce the residual phases after autophasing corrections are applied.

Two strong radio sources, 0237-233 and 1730-130, were observed at X band at elevations ranging from 9 to 25 de-

grees. The phase data from these two sources were separated into three bins according to their elevation angle. A third source, 3C345, was observed at an elevation of ~ 25 degrees. This gave a total of seven different sets of interferometry data. All seven were treated with the autophasing simulation algorithm using various parameters for the integration time and the delay interval. The data were taken in the "A" configuration, in which the three arms of the VLA range from 19 to 21 km in length. There were nine X-band antennas operational.

Using an antenna near the center of the VLA as a reference, the simulations mimicked phase adjustments on the other eight antennas, which were at distances of 0.8 to 20.5 km from the reference antenna. The data were averaged for 10-second intervals, and autophasing was simulated using a feedback algorithm with delays of one or two integration periods. In addition, simulations were performed on the same data with their initial 5-second integration periods. A set of runs was also made with no time delay and with a gain of 1.0 for the 5-second-interval data. This simulation uses the last previous phase as the predictor of the next phase. It is not a physically realizable system, since a finite time is required to calculate the antenna phases before they can be applied to incoming data. However, it does provide a limiting phase residual; all possible systems having the same integration time should have phase residuals greater than those given by this method.

In addition to the feedback loop parameters described above, all of which used $j = 0$ in Eq. (8), simulations were run using a feedback loop in which $j \neq 0$. Specifically, the case where $j = 1$ and $k = 1$ has been considered. This choice of parameters has the effect of using a recent "observed" phase as the predictor of the next observed phase. In this special case, Eq. (8) becomes

$$\phi_{\text{corr}}^{t+1} = \phi_{\text{corr}}^{t-1} + \phi_{\text{res}}^{t-1} \quad (9)$$

The sum of the phase correction and the residual at time $(t-1)$ is the phase that would have been observed at time $(t-1)$ if there had been no autophasing correction at that time.

Table 6 gives the results of the autophasing simulations for the different source observations in July 1987. For each source and elevation angle, results for seven different sets of parameters are shown for the generalized feedback algorithm of Eq. (8). Figure 1 plots the residual phase for two different sources as a function of elevation angle, while Fig. 2 shows results for a single observation as a function of baseline length.

In each case, simulations with different sets of parameters for the autophasing are shown. Table 6 and these figures illustrate several very important points:

- (1) When compared with the current autophasing feedback algorithm, the residual phase is reduced by a factor of ~ 2.5 simply by reducing the integration time to 5 seconds, reducing the time delay in application of the residuals to a single integration period, and increasing the gain to 0.5. A further improvement of ~ 10 to 15 percent is possible if the same delay is used and if the last available phase for a given antenna is used to predict the next phase at that antenna—i.e., if the algorithm given in Eq. (9) is used.
- (2) A further factor of ~ 2 improvement could be achieved if it were possible to compute antenna phases in an infinitesimally short time and immediately use these phases as predictors for the following integration period. However, this is not a physically realizable system. The delay time could be reduced to some value closer to zero if there were more computing resources, but logistical and financial considerations make this action impractical. Since the integration time must be at least $1\frac{2}{3}$ seconds because of the VLA phase-switching characteristics, and the integration time should be even longer than that to achieve adequate autophasing SNR on the Voyager signal, there must be some delay inherent in the individual integrations (i.e., the data from the beginning of a 5-second integration are 5 seconds old by the end of the integration period).
- (3) The residual phase is independent of baseline length for baselines ranging from less than 1 km to over 20 km.
- (4) The residual phase depends on elevation angle, with the residuals near 12-degree elevation being ~ 1.5 to 2 times greater than those for observations of the same source near ~ 22 -degree elevation.
- (5) Although results are not displayed, the residuals for phases calculated only with reference to a single antenna are almost identical to those for phases computed via a global phase-determination procedure. Since the SNR on each baseline was well above 50 in 5-second integrations on the strong sources, this result simply confirms the expectation that system noise was not a limiting factor for the strong sources.

The result summarized in point 3 is not surprising given the character of tropospheric fluctuations. Fluctuations over a given antenna can be viewed as being caused by a tropospheric pattern blowing across that antenna (e.g., [7], [8]). For an

assumed tropospheric wind speed of 8 m/s, representative of that measured 1 km above Goldstone, California [7], fluctuations move no more than 80 meters in a time of 10 seconds or less. Therefore, phase fluctuations on such short time scales are caused only by atmospheric cells having sizes less than ~ 80 meters. These fluctuations are completely uncorrelated between antennas separated by more than ~ 80 meters, and larger (longer time scale) fluctuations are adequately corrected by the feedback loop if the total delay time is no greater than ~ 10 seconds. Hence the size of the VLA configuration seems to make little difference. Even the smallest possible VLA configuration, with each arm of the array having nine antennas along a line 600 meters long, has the majority of its baselines longer than 100 meters. Since there is considerable mutual antenna shadowing in this "D" configuration, the larger "C" configuration, with arm lengths of ~ 2 km, is preferred for Neptune encounter.

The apparent elevation dependence of the residual phase is expected because of the increased number of air masses; hence the increased chance of an antenna line of sight intercepting small tropospheric cells at the low elevation angles. It is clear that the shortest possible integration time that still provides adequate SNR is needed to correct the phases near the 8-degree elevation limit of the VLA.

D. High-SNR Observations During Thunderstorms

Several short observations of strong sources were obtained during thunderstorms in August and September 1987. The purpose of these observations was to determine the ability of the autophasing process to correct for the phase fluctuations during the worst weather expected for the Neptune encounter. These observations generally lasted for ~ 45 minutes and were made when there were thunderstorms and/or rain showers at the control building of the VLA. Since they were made with a small subarray while normal astronomical observing proceeded with the bulk of the VLA, only four antennas were available, and there was no direct control over the integration time. The interferometric data were analyzed via simulations similar to those described above. Table 7 gives the results from the three data sets acquired in August and September 1987.

Table 8 shows results of autophasing simulations on data taken on the radio source 1519-273 during a September thunderstorm (the same results were shown in Table 7) and on a November night. Although the antenna spacings were shorter in November than in September, they were still large enough so that the uncorrected fluctuations over different antennas were uncorrelated. The data during the two months were acquired at identical elevation angles and show a difference of a factor of 5 to 6 in residual phase. This illustrates the magnitude of the possible degradation caused by having the

Neptune encounter occur during the summer rather than the winter. Scaling the thunderstorm data to integration times and delay periods of 5 seconds, according to the trend shown in Table 6, yields an expectation for an rms phase residual of less than ~ 20 degrees in the worst summer weather encountered during these tests. However, it is clear that the previously employed feedback loop and integration time, which can yield residual phases in the 30- to 45-degree range, would have given substantial combining losses for the VLA antennas.

V. Combining Losses

Given a distribution of residual antenna phases, it is possible to compute the power loss expected when the signals from the individual VLA antennas are summed. At the VLA, the signal combination is done in an analog sum device. The power output of this device is

$$P = \sum_{j=1}^N E_j e^{i\phi_j} \sum_{k=1}^N E_k e^{-i\phi_k} \quad (10)$$

Here, P is the total power, E_j is the real electric field amplitude at antenna j , i is the complex number representing $\sqrt{-1}$, ϕ_j is the residual phase of the j th antenna, and N is the number of antennas.

Assuming that all antennas are identical, E_j and E_k can be set equal to a constant, E , and taken outside the sums. Then, Eq. (10) becomes

$$\begin{aligned} P &= E^2 \sum_{j=1}^N \sum_{k=1}^N e^{i(\phi_j - \phi_k)} \\ &= E^2 \sum_{j=1}^N \sum_{\substack{k=1 \\ k \neq j}}^N e^{i(\phi_j - \phi_k)} + NE^2 \end{aligned} \quad (11)$$

In Eq. (11), the terms in which $j = k$ have been evaluated and taken out of the sum. Next, assume that the ensemble average of the phase difference between any two antennas is independent of which antenna pair is chosen. If the phase difference between two antennas is $\Delta\phi$, the sum in Eq. (11) can be evaluated, and the average power coming out of the analog sum device is

$$\langle P \rangle = N(N-1)E^2 \langle e^{i\Delta\phi} \rangle + NE^2 \quad (12)$$

If the residual phases of the identical interferometer pairs are Gaussian distributed with variance σ_ϕ^2 , Eq. (12) becomes

$$\begin{aligned} \langle P \rangle &= \frac{N(N-1)E^2}{\sigma_\phi \sqrt{2\pi}} \int_{-\infty}^{\infty} e^{i\Delta\phi} \exp \left[\frac{-(\Delta\phi)^2}{2\sigma_\phi^2} \right] d(\Delta\phi) + NE^2 \\ &= N(N-1)E^2 \exp \left[\frac{-\sigma_\phi^2}{2} \right] + NE^2 \end{aligned} \quad (13)$$

The addition of N antennas with zero relative phase gives a total power ($\equiv P_{\max}$) of $N^2 E^2$, so the ratio of the summed power to the maximum possible signal power is simply given by

$$\begin{aligned} \frac{\langle P \rangle}{P_{\max}} &= \frac{(N-1)}{N} \exp \left[\frac{-\sigma_\phi^2}{2} \right] + \frac{1}{N} \\ &\approx \exp \left[\frac{-\sigma_\phi^2}{2} \right] \end{aligned} \quad (14)$$

The approximation in the second line of Eq. (14) slightly overestimates the combining loss. For 23 to 27 antennas, the overestimate of the loss is only ~ 0.02 dB for $\sigma_\phi = 25$ degrees and less for lower phase residuals. Therefore, $\exp(-\sigma_\phi^2)/2$ is a slightly conservative estimate of the expected combining loss at the VLA. Identifying σ_ϕ with the rms residual phase after the autophasing procedure, we can immediately see how the summed power will be reduced with increasing phase residual caused either by limited SNR or by tropospheric fluctuations. Figure 3 is a plot of the signal loss vs. rms phase residual. From this figure, note that the desired maximum combining loss of 0.2 dB occurs for a residual phase of ~ 17.4 degrees.

VI. Trade-offs Between Troposphere and SNR

In August 1989 (prime thunderstorm season), the first part of each Voyager pass will be in the late afternoon, when the atmospheric phases can be expected to fluctuate rapidly. It has been shown that the residual phases expected in the SNR-limited case are no more than ~ 10 degrees for the VLA with 25 operational antennas as long as the integration period is at least $3 \frac{1}{3}$ seconds. Phase errors this small are expected to give combining losses well below 0.2 dB. Since the shortest possible integration time is desired to correct for the high-frequency troposphere delays, this implies that an integration period of $3 \frac{1}{3}$ or 5 seconds is desirable during Voyager's Neptune encounter in 1989. An integration time of 5 seconds would provide some signal margin at low elevation angles and should serve to track the troposphere adequately in most cases.

The shortest integration time allowed, 1 2/3 seconds, would be even better for correcting the troposphere. However, there are two arguments against using such a short integration time. First, the SNR on the Voyager signal would be marginal, with predicted phase residuals of ~ 15 degrees at the lowest elevations. Second, the necessary phase computations, data handling, and communication to the individual antennas will take several seconds, long enough to reduce some of the benefits of the shortest integration period. In fact, the required computation time would probably be increased slightly because of the reduced SNR at the shortest integration time.

VII. Summary

VLA tests combined with computer simulations have been conducted to study the process of autophasing the antennas at the VLA. Investigations of the SNR of the autophasing process have been both theoretical and empirical, while the investigations of the troposphere and of various feedback loops have been largely empirical in nature. A study is needed to determine the statistical properties of the troposphere, the relation of the data to models such as that in [9], and the consequent expectations for various feedback loops.

The analysis reported here has yielded the following major conclusions:

- (1) The minimum rms phase achievable by the autophasing procedure can be predicted by considerations of the SNR derived from known properties of the antennas. Both real VLA autophasing and off-line simulations using VLA data taken during observations of Voyager 2 confirm the predictions. Formulas for the rms phase of the phasing process are given as functions

of elevation angle, system temperature, and integration time.

- (2) Also predicted is the improvement in the autophasing process that occurs when individual antenna phases are computed by means of a least-squares algorithm. The observed improvement for several 1987 tests is somewhat less than that predicted on pure theoretical grounds, but generally consistent with the results found in other simulations. Further tests with more antennas will be performed to see if the real improvement is close to that expected for the full VLA.
- (3) It is clear that tropospheric effects can greatly increase the residual phases of the VLA antennas in the autophasing process, especially at the lowest elevations. The residual tropospheric effects can be reduced substantially by decreasing the delay time inherent in the autophasing procedure. If the integration time can be made reasonably short, it appears that simple feedback loops are adequate to correct the VLA phases so that the loss in the process of combining the signals from the individual VLA antennas is less than 0.2 dB.
- (4) For short integration times, there is no apparent dependence of the rms residual phase on baseline length. This implies that the best tropospheric corrections can be made without choosing a VLA configuration that is so small that there is significant shadowing of antennas by one another.
- (5) Consideration of the trade-offs between tropospheric activity and SNR implies that a VLA integration time of 3 1/3 seconds or 5 seconds will be best for observing the Voyager spacecraft in 1989 and will be adequate to meet the goal of less than 0.2 dB of combining loss.

Acknowledgments

Thanks are owed to J. Armstrong, D. Brown, B. Brundage, J. Davidson, G. Resch, and R. Treuhaft for their comments on earlier versions of this article and for suggestions during various stages of this work.

References

- [1] P. J. Napier, A. R. Thompson, and R. D. Ekers, "The Very Large Array: Design and Performance of a Modern Synthesis Radio Telescope," *IEEE Proceedings*, pp. 1295–1320, November 1983.
- [2] J. W. Layland et al., "Interagency Array Study Report," *TDA Progress Report 42-74*, vol. April–June 1983, Jet Propulsion Laboratory, Pasadena, California, pp. 117–148, August 15, 1983.
- [3] J. W. Layland, P. J. Napier, and A. R. Thompson, "A VLA Experiment—Planning for Voyager at Neptune," *TDA Progress Report 42-82*, vol. April–June 1985, Jet Propulsion Laboratory, Pasadena, California, pp. 136–142, August 15, 1985.
- [4] P. C. Crane and P. J. Napier, "Sensitivity," in R. A. Perley, F. R. Schwab, and A. H. Bridle (eds.), *Synthesis Imaging*, Socorro, New Mexico: National Radio Astronomy Observatory, pp. 87–108, 1986.
- [5] J. S. Ulvestad, G. M. Resch, and W. D. Brundage, "X-Band System Performance of the Very Large Array," *TDA Progress Report 42-92*, vol. October–December 1987, Jet Propulsion Laboratory, Pasadena, California, pp. 123–137, February 15, 1988.
- [6] T. Cornwell, "Self-Calibration," in R. A. Perley, F. R. Schwab, and A. H. Bridle (eds.), *Synthesis Imaging*, Socorro, New Mexico: National Radio Astronomy Observatory, pp. 137–147, 1986.
- [7] R. N. Treuhaft and G. E. Lanyi, "The Effect of the Dynamic Wet Troposphere on Radio Interferometric Measurements," *Radio Science*, vol. 22, pp. 251–265, 1987.
- [8] J. W. Armstrong and R. A. Sramek, "Observations of Tropospheric Phase Scintillations at 5 GHz on Vertical Paths," *Radio Science*, vol. 17, pp. 1579–1586, 1982.
- [9] J. P. Basart and Y. Zheng, "Modeling Very Large Array Phase Data by the Box-Jenkins Method," *Radio Science*, vol. 21, pp. 863–881, 1986.

Table 1. Standard parameters for SNR calculations

Parameter	Value
System temperature (30-degree elevation)	35 ± 3 K
Atmospheric temperature	$2.73^{+1.0}_{-0.15}$ K per air mass
Aperture efficiency (30-degree elevation)	0.62 ± 0.03
Voyager signal strength	$(5.0 \pm 0.4) \times 10^{-21}$ W/m ²
Attenuation	$0.01^{+0.004}_{-0.0015}$ per air mass
Gain loss factor (20-degree elevation)	$0.02^{+0.01}_{-0.005}$
Gain loss factor (10-degree elevation)	$0.04^{+0.02}_{-0.01}$
Global autophase improvement	$(0.7 \pm 0.1)\sqrt{N-2}$

**Table 2. Predicted phase residual based only on SNR of Voyager observations with VLA
(errors in the rms phase residuals are $\sim +25$ percent and ~ -20 percent)**

Elevation, degrees	Integration time, seconds	SNR (one baseline)	$\sigma_\phi(\text{global})$, degrees
30	10	3.9	4.4
30	5	2.8	6.2
30	3 1/3	2.3	7.6
20	10	3.5	4.9
20	5	2.5	6.9
20	3 1/3	2.0	8.5
10	10	2.8	6.2
10	5	2.0	8.7
10	3 1/3	1.6	10.7

**Table 3. RMS phase residuals for Voyager observations after simulations of the old VLA
autophasing procedure with 10-second integrations**

Gain (α)	Number of periods delayed		RMS phase, degrees	
	j	k	March	May
0.125	0	2	14.7	14.8
0.25	0	2	15.5	15.8
0.5	0	2	24.7	24.1
0.125	0	1	14.8	14.3
0.25	0	1	15.1	14.6
0.5	0	1	17.1	17.4

Table 4. RMS phase residuals for Voyager observations after simulations of the global autophasing procedure

Gain (α)	Number of periods delayed		RMS phase, degrees	
	j	k	March	May
0.125	0	2	8.3	10.8
0.25	0	2	8.6	11.5
0.5	0	2	17.8	17.6
0.125	0	1	8.6	10.1
0.25	0	1	8.5	10.1
0.5	0	1	9.6	10.8

Table 5. Phase residuals for autophasing simulations during observations of the Voyager spacecraft in November 1987

Integration time, seconds	Gain (α)	Number of periods delayed		RMS phase, degrees	
		j	k	Reference antenna method	Global method
10	0.125	0	2	11.2	6.7
10	0.25	0	2	11.7	6.7
10	0.5	0	2	17.1	10.0
10	0.125	0	1	10.9	6.3
10	0.25	0	1	11.1	6.0
10	0.5	0	1	12.5	6.2
10	1.0	0	0	13.2	5.4

Table 6. RMS phase residuals for autophasing simulations on strong sources on July 22, 1987

Source	Integration time, seconds	Gain (α)	Number of periods delayed		RMS phase, degrees
			<i>j</i>	<i>k</i>	
0237-233, 9- to 14-degree elevation	10	0.25	0	2	24.0
	10	0.25	0	1	21.0
	5	0.25	0	2	15.5
	5	0.25	0	1	13.5
	5	0.5	0	1	10.6
	5	1.0	0	0	5.4
	5	1.0	1	1	9.2
0237-233, 15- to 19-degree elevation	10	0.25	0	2	14.1
	10	0.25	0	1	12.5
	5	0.25	0	2	8.8
	5	0.25	0	1	7.8
	5	0.5	0	1	6.1
	5	1.0	0	0	3.5
	5	1.0	1	1	5.6
0237-233, 20- to 24-degree elevation	10	0.25	0	2	10.2
	10	0.25	0	1	8.8
	5	0.25	0	2	6.6
	5	0.25	0	1	5.9
	5	0.5	0	1	5.0
	5	1.0	0	0	2.9
	5	1.0	1	1	4.5
3C345, 22- to 26-degree elevation	10	0.25	0	2	12.8
	10	0.25	0	1	11.5
	5	0.25	0	2	8.1
	5	0.25	0	1	7.3
	5	0.5	0	1	6.0
	5	1.0	0	0	3.5
	5	1.0	1	1	5.4
1730-130, 9- to 14-degree elevation	10	0.25	0	2	33.0
	10	0.25	0	1	27.2
	5	0.25	0	2	20.0
	5	0.25	0	1	17.1
	5	0.5	0	1	12.2
	5	1.0	0	0	6.1
	5	1.0	1	1	10.8
1730-130, 15- to 19-degree elevation	10	0.25	0	2	29.3
	10	0.25	0	1	23.3
	5	0.25	0	2	17.4
	5	0.25	0	1	15.0
	5	0.5	0	1	11.3
	5	1.0	0	0	5.7
	5	1.0	1	1	9.9
1730-130, 20- to 25-degree elevation	10	0.25	0	2	16.8
	10	0.25	0	1	14.4
	5	0.25	0	2	11.2
	5	0.25	0	1	9.6
	5	0.5	0	1	7.7
	5	1.0	0	0	4.1
	5	1.0	1	1	6.8

Table 7. Residual phases for simulations on data taken during thunderstorms and rain showers

Source	Integration time, seconds	Gain (α)	Number of periods delayed		RMS phase, degrees
			<i>j</i>	<i>k</i>	
1127-145,	8.33	0.25	0	2	25.9
21- to 26-degree	8.33	0.25	0	1	22.1
elevation,	8.33	0.5	0	1	17.7
August 11, 1987	8.33	1.0	1	1	16.3
	8.33	1.0	0	0	8.5
3C279,	8.33	0.25	0	2	17.5
50-degree elevation,	8.33	0.25	0	1	15.3
September 18, 1987	8.33	0.5	0	1	12.0
	8.33	1.0	1	1	10.6
	8.33	1.0	0	0	6.3
1519-273,	10	0.25	0	2	44.6
15- to 22-degree	10	0.25	0	1	37.6
elevation,	10	0.5	0	1	32.2
September 22, 1987	10	1.0	1	1	27.2
	10	1.0	0	0	16.0

Table 8. Comparison of data taken on 1519-273 at 15 to 22 degrees of elevation during a September thunderstorm and during a November night

Integration time, seconds	Gain (α)	Number of periods delayed		RMS phase, degrees	
		<i>j</i>	<i>k</i>	September	November
10	0.25	0	2	44.6	7.6
10	0.25	0	1	37.6	6.3
10	0.5	0	1	32.2	5.7
10	1.0	1	1	27.2	5.1
10	1.0	0	0	16.0	3.3

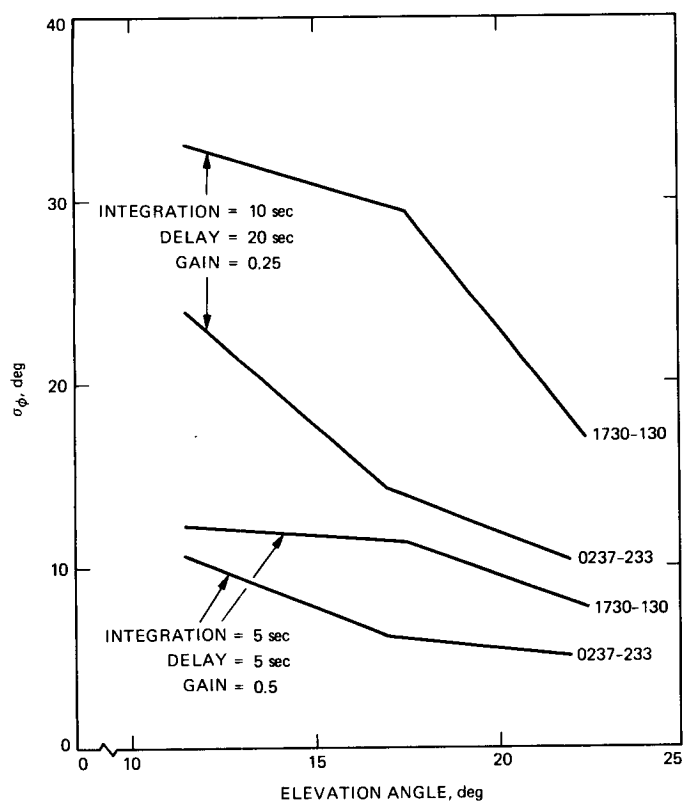


Fig. 1. Residual phase from autophasing simulations for two different sources for different autophasing parameters

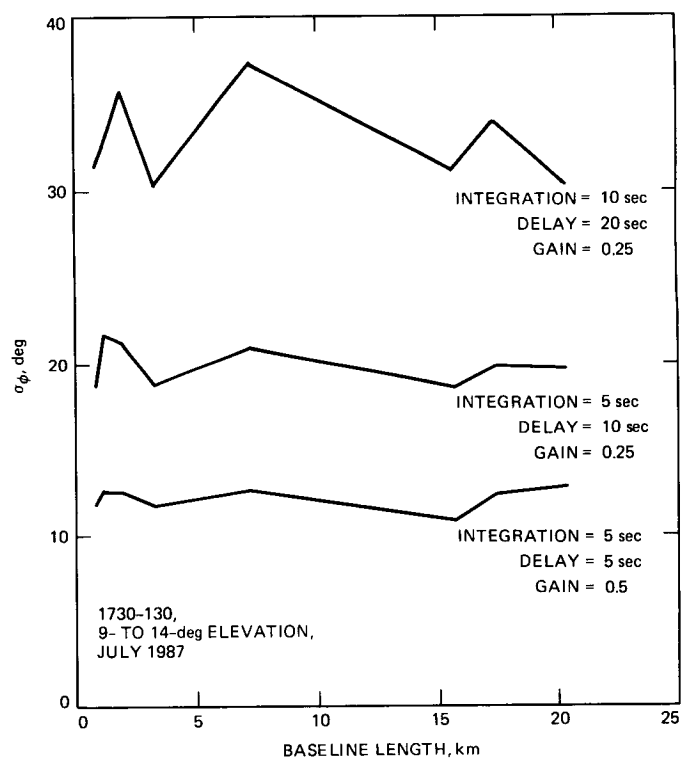


Fig. 2. Plot of residual phase versus baseline length for autophasing simulations on a single set of observations

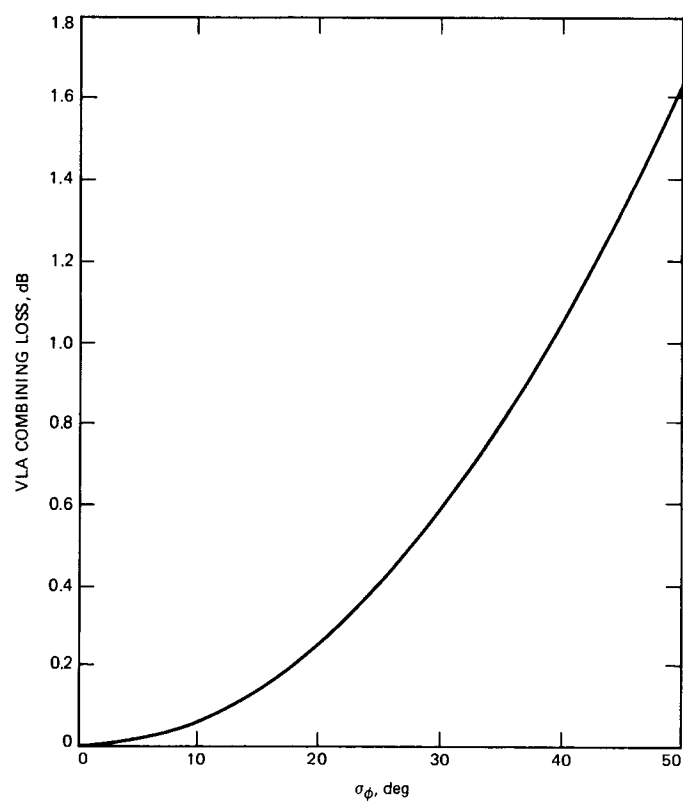


Fig. 3. Combining loss for the sum of many antennas as a function of rms residual phase for each antenna

522-61

165633

TDA Progress Report 42-94

138.

1989000837
618452
17p.

N89-10208

April-June 1988

JPL

Phobos Lander Coding System: Software and Analysis

K.-M. Cheung and F. Pollara
Communications Systems Research Section

This article describes the software developed for the decoding system used in the telemetry link of the Phobos Lander mission. Encoders and decoders are provided to cover the three possible telemetry configurations. The software can be used to decode actual data or to simulate the performance of the telemetry system. The theoretical properties of the codes chosen for this mission are analyzed and discussed.

I. Introduction

The Soviets plan to launch a mission to Mars in mid-July 1988. This mission consists of two identical spacecraft, each carrying a lander to be placed on Phobos, a moon of Mars. The insertion into Mars orbit will occur in mid-February 1989. The landing on Phobos will take place toward the end of April or early May 1989. This mission is designed to last one year with a possible extension to two years. NASA has agreed to provide some data acquisition support.

NASA/JPL's direct involvement in the Phobos project centers on the landers. These are complex spacecraft that are capable of receiving commands, transmitting telemetry, providing two-way range and doppler, and supporting Very Long Baseline Interferometry (VLBI) measurements.

The Phobos Lander coding system block diagram is shown in Fig. 1. The telemetry data are transmitted in frames of 2048 bits during a Phobos-Earth view period, which occurs approximately every 7 hours, 39 minutes (the Phobos orbital period) for a duration of approximately 20 to 40 minutes. The Phobos Lander telemetry system can be configured to operate in three different modes, as shown in Fig. 1. The first

mode involves the use of a $K = 6$, $r = 1/2$ convolutional code, where K is the constraint length and r the rate of the code. The second mode involves the use of a $K = 9$, $r = 1/2$ time-varying convolutional code with period 4. The third mode involves the use of a concatenation of a Bose-Chaudhuri-Hocquenghem (BCH) block code and a $K = 9$, $r = 1/2$ time-varying convolutional code.

Our work is concerned primarily with the analysis and implementation of the following software modules to support the Phobos Lander telemetry processing requirement:

- (1) A convolutional encoder ($K = 9$, $r = 1/2$, time-varying).
- (2) A BCH encoder (128,113).
- (3) A Viterbi decoder ($K = 9$, $r = 1/2$, time-varying).
- (4) A BCH decoder (128,113).
- (5) Convolutional code ($K = 9$, $r = 1/2$, time-varying) and BCH (128,113) test and simulation software.

In addition to the above software modules, an encoder module, a decoder module, and a simulation software module

for the $K = 6, r = 1/2$ convolutional code have been developed. The $K = 6, r = 1/2$ modules are used only to generate performance comparisons with the other operational modes.

All the software modules are written in FORTRAN-77 on an IBM-AT-compatible computer. They are contained in three programs:

- (1) *k6.f*. This program contains the encoder and decoder modules for the $K = 6, r = 1/2$ convolutional code. It determines the bit-error-rate performance of this convolutional code by simulation.
- (2) *k9.f*. This program contains the encoder and decoder modules for the $K = 9, r = 1/2$ time-varying convolutional code. It determines the bit-error-rate performance of this convolutional code by simulation.
- (3) *bchk9.f*. This program contains the encoder and decoder modules for the $K = 9, r = 1/2$ time-varying convolutional code as well as the encoder and decoder modules for the (128,113) extended BCH code. It determines the bit-error-rate performance of the concatenated system consisting of the extended BCH code as the outer code and the time-varying convolutional code as the inner code, with interleaving.

The software will be transported to a Modcomp 7845 computer and integrated with the Phobos Lander software package being developed by ICI of Spain.

In this article, theoretical and simulation results on the Phobos Lander coding system are presented. The $K = 6, r = 1/2$ convolutional code is discussed in Section II. The $K = 9, r = 1/2$ time-varying convolutional code is delineated in Section III. The concatenated system consisting of the extended BCH code and the time-varying convolutional code is outlined in Section IV. Finally, the detailed decoding algorithm of the extended BCH code and the weight enumerator and the decoder error probability of the extended BCH code are presented in Appendix A and Appendix B, respectively.

II. $K = 6, r = 1/2$ Convolutional Code

When operating in the first mode, the Phobos Lander spacecraft will convolutionally encode the telemetry data using a constraint length 6, rate $1/2$ convolutional encoder as shown in Fig. 2. The Phobos Lander telemetry processor software will provide the capability to decode convolutionally encoded telemetry data ($K = 6, r = 1/2$) using a maximum-likelihood decoding algorithm (Viterbi algorithm). This is an existing DSN capability which will be integrated into the Phobos Lander processor.

The $(6,1/2)$ convolutional code used has connection vectors $g_1 = 73$ and $g_2 = 61$ (in octal), is transparent, and has a free distance d_f equal to 8. The truncation length of the Viterbi decoder is a variable parameter and was set to 49 in our simulations. The choice of truncation length of the decoder is somewhat arbitrary as long as it is larger than approximately $5K$, which guarantees a small performance degradation due to path truncation. By assuming perfect synchronization on the biphasic modulated symbol stream on an additive white Gaussian noise (AWGN) channel, the bit-error-rate (BER) performance of the code is simulated as a function of the bit signal-to-noise ratio (E_b/N_0), and the result is given in Fig. 3. All the simulation results shown in this article assume that the received symbols are unquantized.

III. $K = 9, r = 1/2$ Time-Varying Convolutional Code

The second operational mode uses a constraint length 9, rate $1/2$, time-varying convolutional code. The connection vectors are $g_1 = 557, g_2 = 663, g_3 = 711, g_4 = 745$ (in octal) and are used in pairs according to the periodic sequence $(g_1, g_2), (g_2, g_3), (g_3, g_4), (g_4, g_1), \dots$, with period 4, as shown in Fig. 4. This code has $d_f = 10$, which is far from optimal, since the Plotkin upper bound applied to convolutional codes gives $d_f \leq 12$. For this code the minimum distance error events are always at distance 10 independent of their starting point within the period (phase), but the number of such events depends on the phase.

In fact, the best-known fixed convolutional code with $K = 9, r = 1/2$ has connection vectors 561,753 (in octal) and $d_f = 12$, which achieves the bound. Figure 5 shows the performance of this code compared to the code chosen for Phobos Lander on an AWGN channel, with perfect synchronization and a truncation length of 65 bits for both codes. The same two codes concatenated with the BCH code are compared in Fig. 6. It is well known that every periodic convolutional code with period T and rate k/n has an equivalent fixed code of rate Tk/Tn . In our case, the fixed code has rate $4/8$ and $d_f = 10$, but now error events can start only every four bits, since the encoder inputs four bits at a time.

IV. Concatenated Coding System

When operating in the third mode, the Phobos Lander spacecraft will first encode the telemetry data using a (128, 113) extended BCH code. Each data frame contains 19 BCH code words ($19 \times 113 = 2147$ bits) and is padded with 99 zeros to obtain the desired frame length of 2048 bits. The

actual frame structure is still undefined, but it is natural to envision that the 99 additional bits could be used as a frame marker. The BCH code words are then interleaved to depth 16 before they are convolutionally encoded by a $(9,1/2)$ time-varying convolutional code. The DSN Phobos Lander telemetry processor software will provide the capability to first decode the received data stream (convolutionally encoded) using a maximum-likelihood decoding algorithm. The output data of the Viterbi decoder will then be deinterleaved before they are block decoded using a simple but efficient BCH algebraic decoding algorithm. The detailed decoding algorithm for the extended BCH code is given in Appendix A, and the weight enumerator and the decoder error probability of the extended BCH code are given in Appendix B.

The $(128,113)$ extended BCH code consists of a $(127,113)$ BCH code plus an overall parity bit. The $(127,113)$ BCH code has a generator polynomial $g(x) = 1 + x + x^2 + x^4 + x^5 + x^6$

$+ x^8 + x^9 + x^{14}$ and is capable of correcting two errors. The schematic diagram of the encoder of the $(127,113)$ BCH code is given in Fig. 7. With the addition of a parity check bit to the BCH code word, the code is also capable of detecting three errors. The $(9,1/2)$ time-varying convolutional code is the same as that discussed in Section IV. Again, by assuming perfect synchronization on the symbol stream in a AWGN channel, the BER performance of the code is simulated and the result is given in Fig. 8. Figure 8 also shows a comparison of the three operating modes of Phobos Lander.

V. Conclusion

The results obtained in this article give a first performance evaluation of the Phobos Lander telemetry system. The software modules will be integrated in the Phobos Lander telemetry processor and used at the DSN stations to support this mission.

Acknowledgment

The authors would like to thank Dr. R. J. McEliece of Caltech's Electrical Engineering Department for his helpful suggestions on the decoding algorithms.

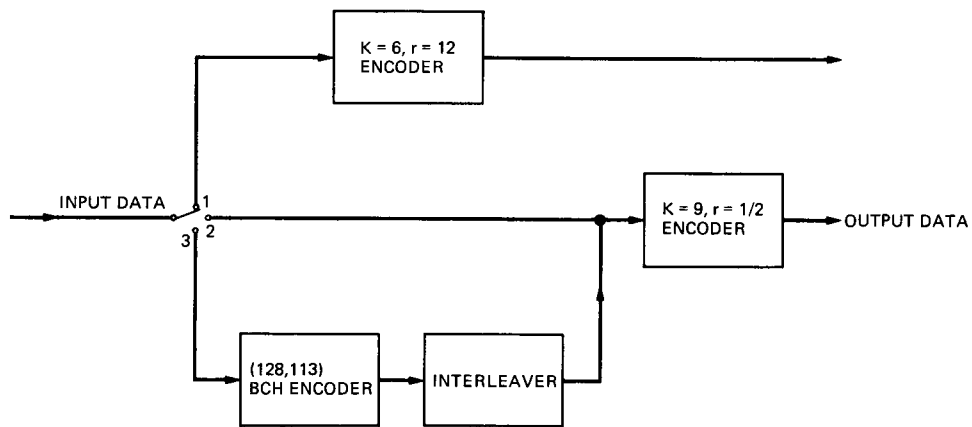


Fig. 1. Phobos Lander coding system

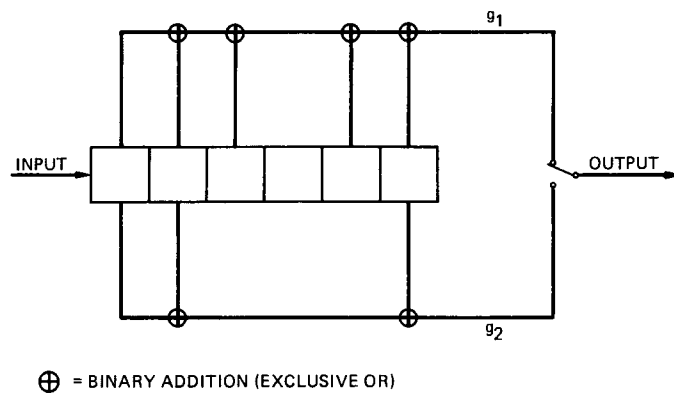


Fig. 2. The $K = 6, r = 1/2$ encoder

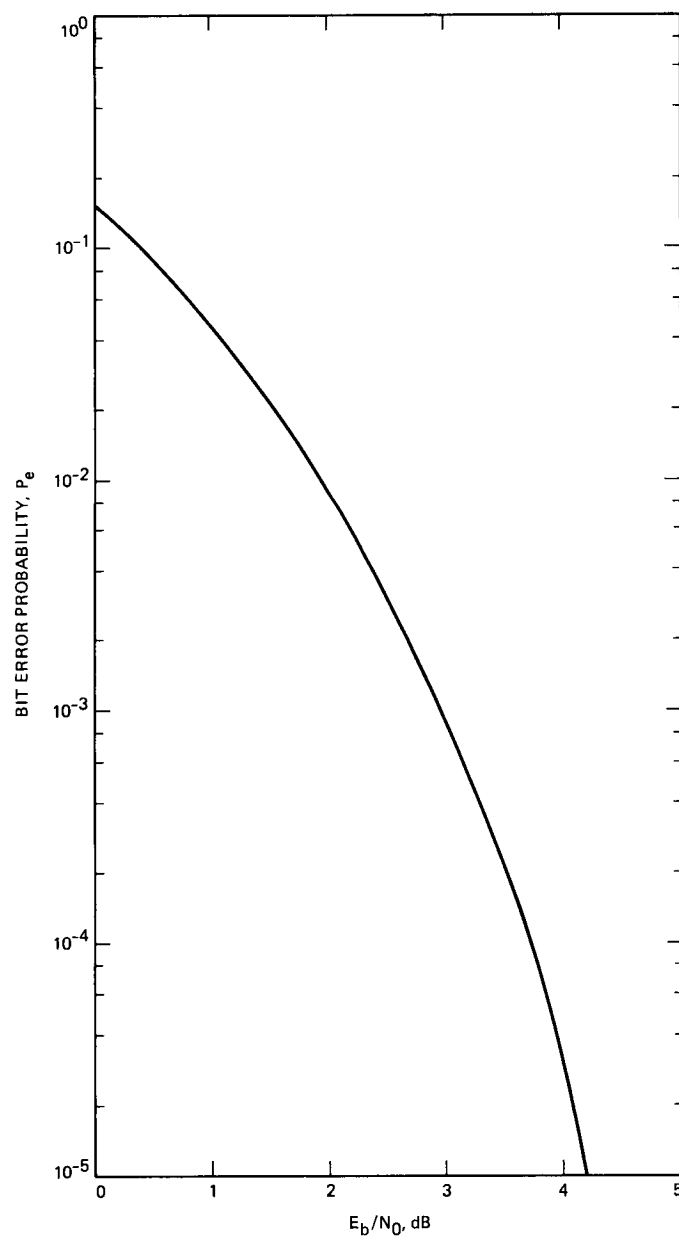
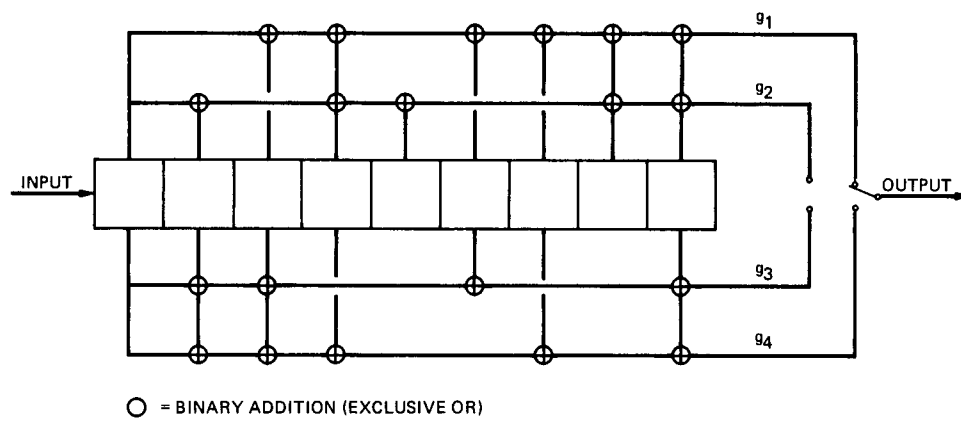


Fig. 3. Performance of the $K = 6, r = 1/2$ code



INPUT BITS	1		2		3		4	
OUTPUT SYMBOLS	g_1	g_2	g_2	g_3	g_3	g_4	g_4	g_1

Fig. 4. The $K = 9, r = 1/2$ encoder

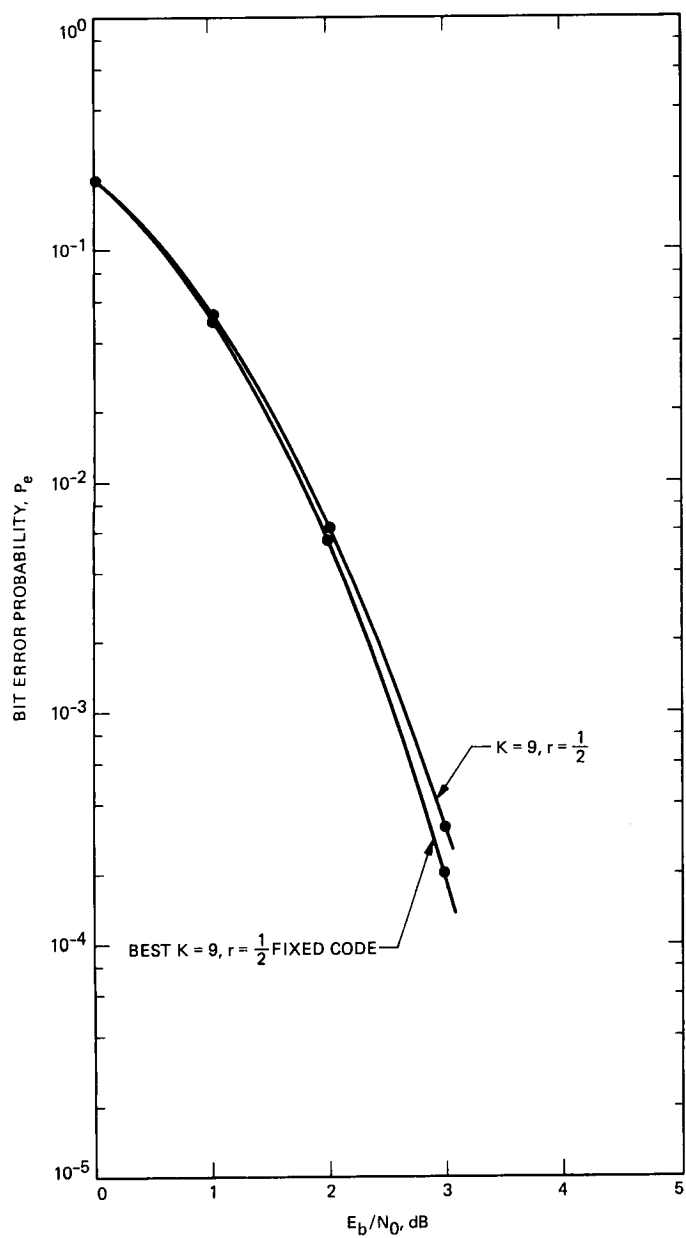


Fig. 5. Performance of two $K = 9, r = 1/2$ codes

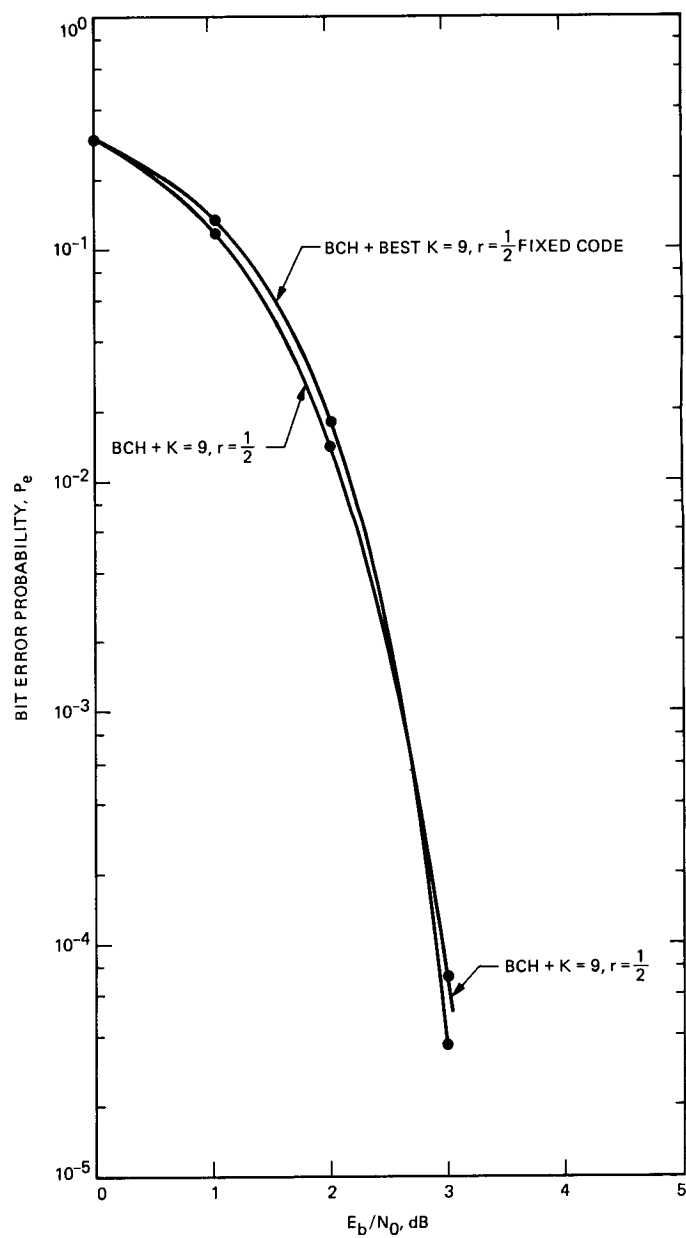


Fig. 6. Performance of two $K = 9, r = 1/2$ codes, concatenated with the BCH code

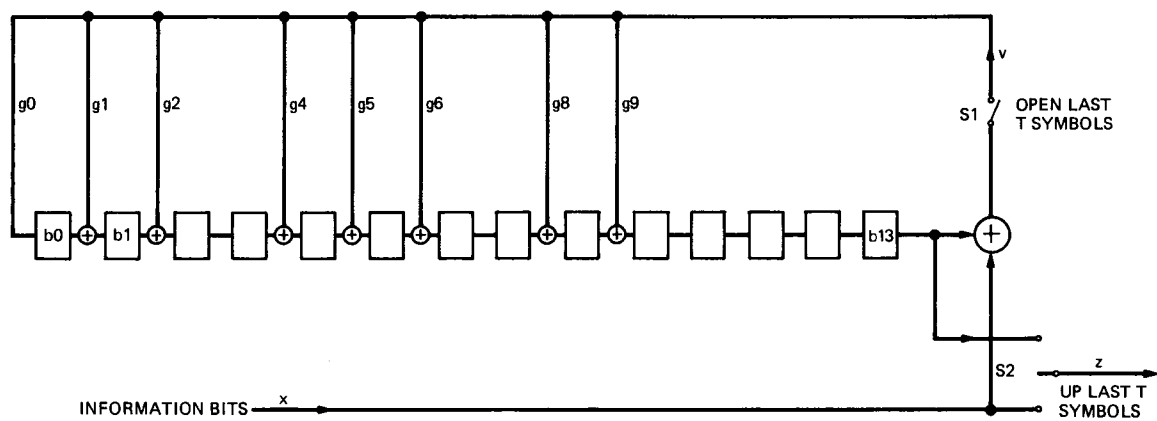


Fig. 7. Encoder for the (127,113) BCH code

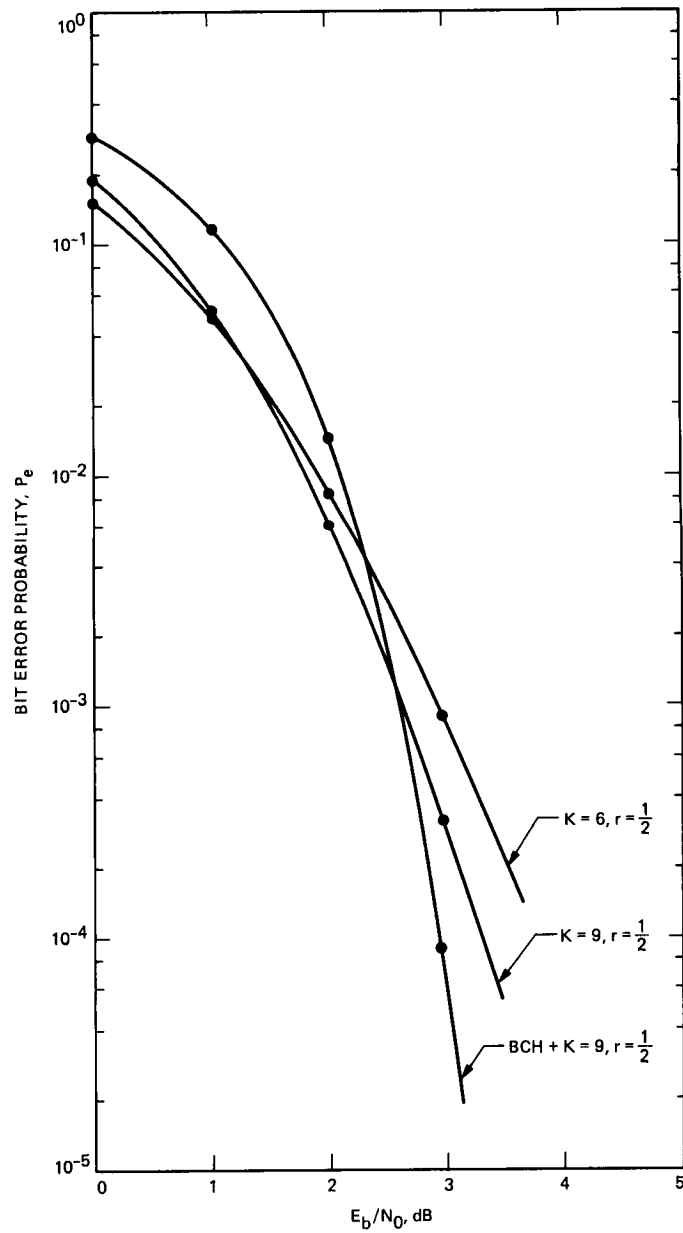


Fig. 8. Performance of a concatenated system

Appendix A

Decoding the (128,113) Extended BCH Code

The (128,113) extended BCH code consists of a (127,113) BCH code plus an overall parity bit. The generator polynomial for the (127,113) BCH code is $g(x) = 1 + x + x^2 + x^4 + x^5 + x^6 + x^8 + x^9 + x^{14}$. The (127,113) BCH code can correct two errors. With the introduction of an overall parity bit, the extended code can also detect three errors.

$$C'(\alpha) = S_1 \quad (\text{A-3})$$

and

$$C'(\alpha^3) = S_3 \quad (\text{A-4})$$

The decoding of the (127,113) BCH code requires finite field manipulations. A primitive irreducible polynomial $f(x) = x^7 + x^3 + 1$ of order 7 over $GF(2)$ is used to construct $GF(2^7)$. Let α be a root of $f(x)$ (i.e., $f(\alpha) = 0$). Then for all $\lambda \in GF(2^7)$, λ can be represented as a linear combination of $1, \alpha, \dots, \alpha^6$ over $GF(2^7)$.

where S_1, S_3 are known as the syndromes and $S_1, S_3 \in GF(2^7)$. It is shown in [A-1] that this code can correct two or fewer errors. The correction of a single error [A-1] is trivial. The correction of two errors, however, involves solving the following quadratic equation:

$$x^2 + S_1 x + \left(\frac{S_2}{S_1} + S_1^2 \right) = 0 \quad (S_1 \neq 0) \quad (\text{A-5})$$

Let $g_1(x)$ and $g_3(x)$ denote the minimal polynomials of α and α^3 , respectively (i.e., $g_1[\alpha] = 0$ and $g_3[\alpha^3] = 0$). It can be observed that $g(x) = g_1(x)g_3(x)$. Let $\underline{C} = [C_0, C_1, \dots, C_{126}]$ be a BCH code word, where $C_i \in GF(2)$. Let $C(x)$ denote the polynomial $C_0 + C_1 x + \dots + C_{126} x^{126}$. The following equations are then obtained [A-1]:

$$C(\alpha) = 0 \quad (\text{A-1})$$

and

$$C(\alpha^3) = 0 \quad (\text{A-2})$$

Let \underline{C}' be the received pattern and let \underline{E} be the error pattern. Then $\underline{C}' = \underline{C} + \underline{E}$. Let

This would normally require an exhaustive search for elements in $GF(2^7)$ which satisfy Eq. (A-5). This exhaustive search can be bypassed by transforming Eq. (A-5) to the form $x^2 + x = \beta$, where $\beta \in GF(2^7)$. By using the concept of the trace of an element in a finite field [A-2], the solvability of Eq. (A-5) can be determined by testing if $Tr(\beta) = 0$. If Eq. (A-5) is solvable, the two roots of that equation, which indicate the two error locations in the code word, can be found readily with simple algebra. The details of this algorithm are discussed in [A-1] and [A-2].

The detection of three errors is done by considering the overall parity of the 128 received symbols as well as the syndromes S_1 and S_3 . The complete decoding algorithm for the (128,113) extended BCH code is given in Fig. A-1.

References

- [A-1] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*, Amsterdam: North-Holland, 1977.
- [A-2] R. J. McEliece, *Finite Fields for Computer Scientists and Engineers*, Norwell, Massachusetts: Kluwer Academic Publishers, 1987.

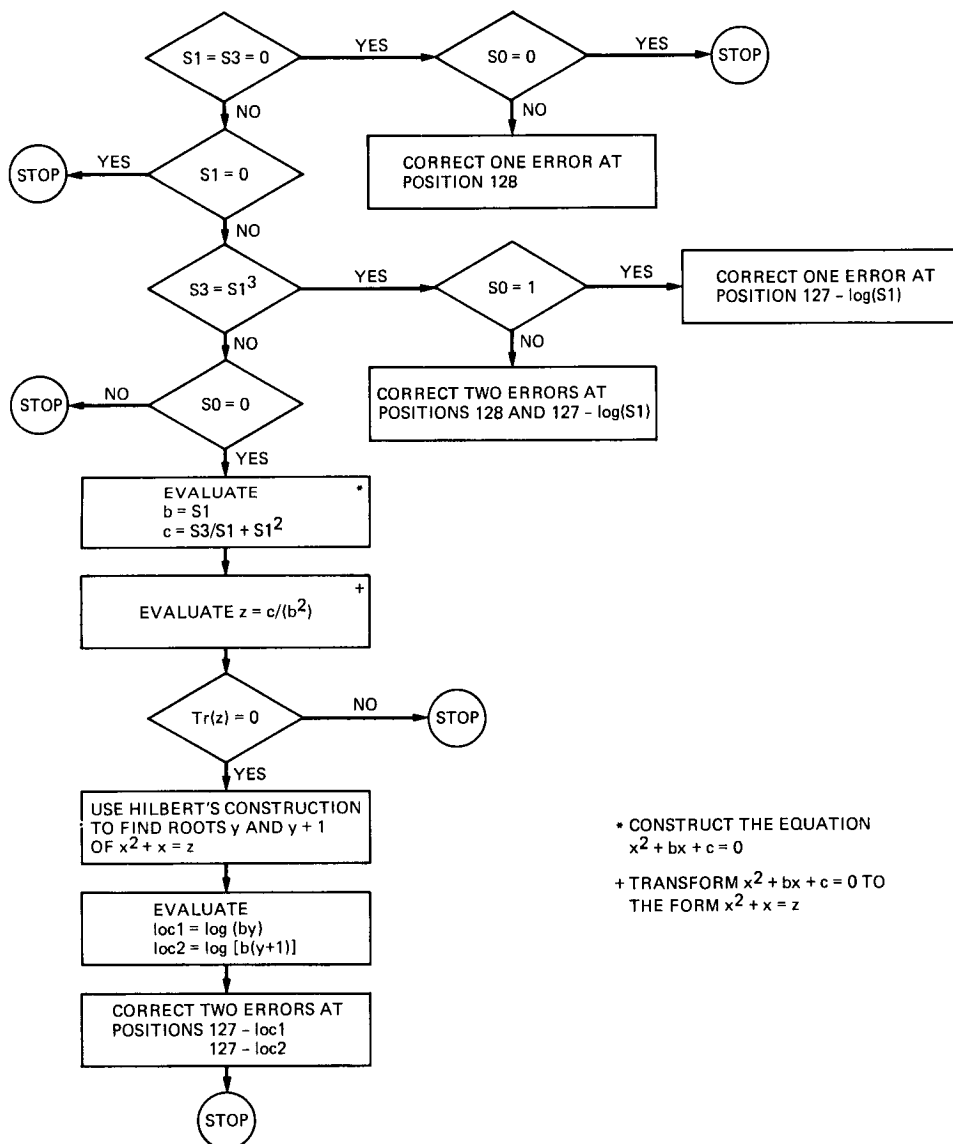


Fig. A-1. Decoding algorithm for the (128,113) BCH code

Appendix B

The Weight Enumerator and the Decoder Error Probability of the (128,113) Extended BCH Code

The decoder error probability $P_E(u)$ of a code is defined to be the conditional probability that a code word is being mis-decoded to another code word given that an error pattern of weight u occurs. If all error patterns of weight u are equally probable, then $P_E(u)$ is given by the following expression [B-1], [B-2] (assuming the code is binary):

$$P_E(u) = \frac{D_u}{\binom{n}{u}} \quad (\text{B-1})$$

where D_u denotes the number of decodable words of weight u .

In order to find D_u , the weight enumerator A_u of the code (the number of code words of weight u) must first be evaluated. The weight enumerator of the dual of the (128,113) extended BCH code, which is denoted by B_u , is given in [B-3]. A_u is then obtained by using MacWilliams's identity on binary codes [B-4]:

$$A(z) = 2^{-(n-k)} (1+z)^n B\left(\frac{1-z}{1+z}\right) \quad (\text{B-2})$$

where $A(z) = A_0 + A_1 z + \dots + A_n z^n$ and $B(z) = B_0 + B_1 z + \dots + B_n z^n$. For the dual of the (128,113) extended BCH code, $B_0 = 1$, $B_{56} = 8128$, $B_{64} = 16,510$, $B_{72} = 8128$, and $B_{128} = 1$. The A_i 's are then found exactly by using MacWilliams's identity, and their values are tabulated in

Fig. B-1 for even i , since $A_i = 0$ for odd i . The number of decodable words of weight u , D_u , is then calculated as follows [B-1]:

$$D_4 = \binom{6}{2} A_6$$

$$D_5 = \binom{6}{1} A_6$$

$$D_6 = A_6 + \binom{6}{1} \binom{122}{1} A_6 + \binom{8}{2} A_8$$

$$D_{2n-1} = \binom{130-2n}{1} A_{2n-2} + \binom{2n}{1} A_{2n}$$

$$D_{2n} = \binom{130-2n}{2} A_{2n-2} + A_{2n} + \binom{128-2n}{1} \binom{2n}{1} A_{2n} + \binom{2n+2}{2} A_{2n+2}$$

where $4 \leq n \leq 64$. The decoder error probability $P_E(u)$ is calculated using Eq. (B-1). The values of $P_E(u)$ are tabulated in Fig. B-2. It is observed that $P_E(u)$ "oscillates" between two values: 0.496 and 0.00781, depending on whether u is even or odd. This observation will be discussed in more detail in a forthcoming progress report.

References

- [B-1] R. J. McEliece and L. Swanson, "On the Decoder Error Probability for Reed-Solomon Codes," *IEEE Tran. Inform. Theory*, vol. IT-32, pp. 701-703, September 1986.
- [B-2] K.-M. Cheung, "More on the Decoder Error Probability for Reed-Solomon Codes," *TDA Progress Report 42-91*, vol. July-September 1986, Jet Propulsion Laboratory, Pasadena, California, pp. 213-223, November 15, 1987.
- [B-3] E. R. Berlekamp, *Algebraic Coding Theory*, Laguna Hills, California: Aegean Park Press, 1984.
- [B-4] S. Lin and D. J. Costello, *Error Control Coding: Fundamentals and Applications*, Englewood Cliffs, New Jersey: Prentice-Hall, 1983.

ORIGINAL PAGE IS
OF POOR QUALITY

A[0] = 1.000e+000	A[66] = 1.374e+033
A[2] = 0.000e+000	A[68] = 1.141e+033
A[4] = 0.000e+000	A[70] = 8.359e+032
A[6] = 3.414e+005	A[72] = 5.406e+032
A[8] = 8.729e+007	A[74] = 3.082e+032
A[10] = 1.384e+010	A[76] = 1.548e+032
A[12] = 1.448e+012	A[78] = 6.834e+031
A[14] = 1.061e+014	A[80] = 2.649e+031
A[16] = 5.697e+015	A[82] = 8.998e+030
A[18] = 2.315e+017	A[84] = 2.672e+030
A[20] = 7.303e+018	A[86] = 6.914e+029
A[22] = 1.827e+020	A[88] = 1.555e+029
A[24] = 3.683e+021	A[90] = 3.029e+028
A[26] = 6.070e+022	A[92] = 5.087e+027
A[28] = 8.272e+023	A[94] = 7.332e+026
A[30] = 9.413e+024	A[96] = 9.020e+025
A[32] = 9.020e+025	A[98] = 9.413e+024
A[34] = 7.332e+026	A[100] = 8.272e+023
A[36] = 5.087e+027	A[102] = 6.070e+022
A[38] = 3.029e+028	A[104] = 3.693e+021
A[40] = 1.555e+029	A[106] = 1.827e+020
A[42] = 6.914e+029	A[108] = 7.303e+018
A[44] = 2.672e+030	A[110] = 2.315e+017
A[46] = 8.998e+030	A[112] = 5.697e+015
A[48] = 2.649e+031	A[114] = 1.061e+014
A[50] = 6.834e+031	A[116] = 1.448e+012
A[52] = 1.548e+032	A[118] = 1.384e+010
A[54] = 3.082e+032	A[120] = 8.729e+007
A[56] = 5.406e+032	A[122] = 3.414e+005
A[58] = 8.359e+032	A[124] = 0.000e+000
A[60] = 1.141e+033	A[126] = 0.000e+000
A[62] = 1.374e+033	A[128] = 1.000e+000
A[64] = 1.462e+033	

Fig. B-1. Weight enumerator of the (128,113) BCH code

Pe(0) = 0.0000000000	Pe(65) = 0.0078130264
Pe(1) = 0.0000000000	Pe(66) = 0.4962243897
Pe(2) = 0.0000000000	Pe(67) = 0.0078141898
Pe(3) = 0.0000000000	Pe(68) = 0.4962602342
Pe(4) = 0.4800000000	Pe(69) = 0.0078141525
Pe(5) = 0.0077419355	Pe(70) = 0.4962101574
Pe(6) = 0.4967812973	Pe(71) = 0.0078124122
Pe(7) = 0.0078282110	Pe(72) = 0.4961433166
Pe(8) = 0.4960509691	Pe(73) = 0.0078122035
Pe(9) = 0.0078097428	Pe(74) = 0.4961540703
Pe(10) = 0.4960908022	Pe(75) = 0.0078128671
Pe(11) = 0.0078116229	Pe(76) = 0.4962010077
Pe(12) = 0.4959824226	Pe(77) = 0.0078137769
Pe(13) = 0.0078096133	Pe(78) = 0.4962063144
Pe(14) = 0.4961009553	Pe(79) = 0.0078125765
Pe(15) = 0.0078118830	Pe(80) = 0.4961524476
Pe(16) = 0.4962086132	Pe(81) = 0.0078122672
Pe(17) = 0.0078135469	Pe(82) = 0.4961573712
Pe(18) = 0.4961605096	Pe(83) = 0.0078130283
Pe(19) = 0.0078124440	Pe(84) = 0.4961930461
Pe(20) = 0.4961972575	Pe(85) = 0.0078132626
Pe(21) = 0.0078132943	Pe(86) = 0.4961724062
Pe(22) = 0.4961436700	Pe(87) = 0.0078117645
Pe(23) = 0.0078121348	Pe(88) = 0.4961092886
Pe(24) = 0.4961385234	Pe(89) = 0.0078118161
Pe(25) = 0.0078122653	Pe(90) = 0.4961297598
Pe(26) = 0.4961611465	Pe(91) = 0.0078128097
Pe(27) = 0.0078126833	Pe(92) = 0.4961759676
Pe(28) = 0.4961716826	Pe(93) = 0.0078128959
Pe(29) = 0.0078127893	Pe(94) = 0.4961790413
Pe(30) = 0.4961683532	Pe(95) = 0.0078128440
Pe(31) = 0.0078126913	Pe(96) = 0.4961742825
Pe(32) = 0.4961742825	Pe(97) = 0.0078126913
Pe(33) = 0.0078128440	Pe(98) = 0.4961683532
Pe(34) = 0.4961790413	Pe(99) = 0.0078127893
Pe(35) = 0.0078128959	Pe(100) = 0.4961716826
Pe(36) = 0.4961759676	Pe(101) = 0.0078126833
Pe(37) = 0.0078128097	Pe(102) = 0.4961611465
Pe(38) = 0.4961297598	Pe(103) = 0.0078122653
Pe(39) = 0.0078118161	Pe(104) = 0.4961385234
Pe(40) = 0.4961092886	Pe(105) = 0.0078121348
Pe(41) = 0.0078117645	Pe(106) = 0.4961436700
Pe(42) = 0.4961724062	Pe(107) = 0.0078132943
Pe(43) = 0.0078132626	Pe(108) = 0.4961972575
Pe(44) = 0.4961930461	Pe(109) = 0.0078124440
Pe(45) = 0.0078130283	Pe(110) = 0.4961605096
Pe(46) = 0.4961573712	Pe(111) = 0.0078135469
Pe(47) = 0.0078122672	Pe(112) = 0.4962086132
Pe(48) = 0.4961524476	Pe(113) = 0.0078118830
Pe(49) = 0.0078125765	Pe(114) = 0.4961009553
Pe(50) = 0.4962063144	Pe(115) = 0.0078096133
Pe(51) = 0.0078137769	Pe(116) = 0.4959824226
Pe(52) = 0.4962010077	Pe(117) = 0.0078116229
Pe(53) = 0.0078128671	Pe(118) = 0.4960908022
Pe(54) = 0.4961540703	Pe(119) = 0.0078097428
Pe(55) = 0.0078122035	Pe(120) = 0.4960509691
Pe(56) = 0.4961433166	Pe(121) = 0.0078282110
Pe(57) = 0.0078124122	Pe(122) = 0.4967812973
Pe(58) = 0.4962101574	Pe(123) = 0.0077419355
Pe(59) = 0.0078141525	Pe(124) = 0.4800000000
Pe(60) = 0.4962602342	Pe(125) = 0.0000000000
Pe(61) = 0.0078141898	Pe(126) = 1.0000000000
Pe(62) = 0.4962243897	Pe(127) = 1.0000000000
Pe(63) = 0.0078130264	Pe(128) = 1.0000000000
Pe(64) = 0.4961881147	

Fig. B-2. Decoder error probability of the (128,113) BCH code

1989000838
618455
7p.S23-32
~~16534~~
April-June 1988
JR
N89-10209
JPK

Relative Planetary Radar Sensitivities: Arecibo and Goldstone

N. A. Renzetti

TDA Science Office

T. W. Thompson

Radar Science and Engineering Section

M. A. Slade

Communications Systems Research Section

The increase of the Deep Space Network antennas from 64-meter to 70-meter diameter represents the first of several improvements that will be made over the next decade to enhance earth-based radar sensitivity to solar-system targets. The aperture increase at the Goldstone DSS-14 site, coupled with a proposed increase in transmitter power to 1000 kW, will improve the 3.5-cm radar by about one order of magnitude. Similarly, proposed Arecibo Observatory upgrades of a Gregorian feed structure and an increase of transmitter power to 1000 kW will increase the sensitivity of this radar about 20-fold. In addition, a Goldstone-to-Very Large Array bistatic observation with horizon-to-horizon tracking will have 3.5 times more sensitivity than will a Goldstone horizon-to-horizon monostatic observation. All of these improvements, which should be in place within the next decade, will enrich an already fertile field of planetary exploration.

I. Introduction

Between the mid-1980s and the late 1990s, several upgrades are planned to both the Arecibo (12.6-cm-wavelength) and the Goldstone (3.5-cm-wavelength) earth-based planetary radars. These upgrades, which are already under way with expansion of the DSS-14 (Mars) site aperture to 70 meters, have the potential to increase the sensitivity of earth-based radar observation by one order of magnitude within the next decade. Arecibo upgrades begin with a partial ground screen to reduce background noise. This step is scheduled for completion in the early 1990s. The Arecibo Observatory is also proposing a

Gregorian feed and increased transmitter capability for the 1990s. The Gregorian feed will yield higher and more uniform gains together with reduced background noise. The proposed transmitter upgrade at Arecibo will be 1000 kW, up from the current 450 kW. Goldstone radar upgrades include the completed antenna aperture increase to 70 meters as well as the proposed X-band (3.5-cm) transmitter upgrade from the current 350 kW to 1000 kW.

The purpose of this article is to compare the relative capabilities of these two planetary radars both for the existing con-

ditions and for the near-future conditions outlined above. We will generally ignore the 70-meter-wavelength capability of Arecibo and the 12.5-cm-wavelength capability of Goldstone, since the radars' sensitivities at these wavelengths are less than 10 percent of those associated with the shorter wavelengths. Also briefly considered will be a special Goldstone-to-Very Large Array (VLA) configuration proven in 1987 observations of Saturn. This is a bistatic configuration in which Goldstone transmits continuously at a 3.5-cm wavelength, and radar echoes are recorded at the VLA in Socorro, New Mexico. This configuration takes advantage of the Voyager project upgrading of the VLA to X-band capability for the Neptune encounter in 1989. However, this configuration is rarely available, since the VLA is a radio astronomy facility in high demand by that user community.

II. Signal-to-Noise Equations

The discussion of sensitivity will begin with a statement of the radar signal-to-noise (SNR) equation—the usual method for defining figures of merit for these types of earth-based radar observations [1]:

$$SNR \sim \frac{PG^2\lambda^2\sigma(NLB)^{0.5}}{R^4(KTB)} \quad (1)$$

where

P = peak transmitter gain

G = antenna gain

λ = radar wavelength

σ = target cross section

N = number of observations

R = earth–target range

K = Boltzmann's constant

T = system noise temperature

B = receiver bandwidth

L = post-detection integration time

To compare Arecibo and Goldstone on the same target, the radar cross section and the Earth–target range are assumed to be equal. Thus, the relative sensitivity of these two radars becomes

$$\frac{SNR'}{SNR} = \left(\frac{P'}{P}\right)\left(\frac{T}{T'}\right)\left(\frac{B}{B'}\right)^{0.5}\left(\frac{G'\lambda'}{G\lambda}\right)^2\left(\frac{N'}{N}\right)^{0.5} \quad (2)$$

To proceed further, it is noted that the various bodies of the solar system rotate. The radar astronomer often uses spectral analysis to increase the signal-to-noise ratio. Thus, the received bandwidth and the number of observations will vary with wavelength. In particular, we chose B to match the target's rotational frequency spreading so that

$$B = \frac{2\Omega D}{\lambda} \quad (3)$$

$$N = \tau \Delta f = \text{constant} \cdot (\tau/\lambda)$$

where

Ω = target rotation rate

D = target diameter

τ = observation time

Δf = frequency resolution needed to resolve a particular area

In Eq. (3), the product (ΩD) is the limb-to-limb velocity difference and $(2\Omega D/\lambda)$ is the Doppler frequency difference across the target. In Eq. (3), it should be noted that the frequency resolution needed to resolve a particular area (Δf) is proportional to λ^{-1} . Thus, observations at a higher frequency can be conducted with larger bandwidths, which require shorter times per spectrum. This, in turn, yields more observations for a fixed period of time.

Another factor which varies with observation wavelength is antenna gain, G , which is given by

$$G = \frac{4\pi\eta A}{\lambda^2} \quad (4)$$

where

η = antenna efficiency

A = antenna area

When these factors are combined by substituting Eqs. (3) and (4) into Eq. (2), the relative sensitivity of the two radars becomes

$$\begin{aligned} \frac{SNR'}{SNR} &= \left(\frac{P'T}{PT'}\right)\left(\frac{G'}{G}\right)^2\left(\frac{\tau'}{\tau}\right)^{0.5}\left(\frac{\lambda'}{\lambda}\right)^{2.5} \\ &= \left(\frac{P'T}{PT'}\right)\left(\frac{\eta'A'}{\eta A}\right)^2\left(\frac{\tau'}{\tau}\right)^{0.5}\left(\frac{\lambda'}{\lambda}\right)^{1.5} \end{aligned} \quad (5)$$

Several quantities in Eq. (5) vary with the declination of the target. In particular, the total time that a target is within the visibility limits of the Arecibo or Goldstone antennas is given by

$$\tau = 2 \cdot LHA$$

where τ = target visibility time. LHA (local hour angle at the tracking limits) is derived from the following equation:

$$\begin{aligned} \sin(EI) &= \cos(Zen) \\ &= \sin(\phi) \cdot \sin(\delta) + [\cos(\phi) \cdot \cos(\delta) \cdot \cos(LHA)] \end{aligned} \quad (6)$$

where

EI = elevation limit of the antenna

Zen = zenith limit of the antenna

ϕ = radar observatory latitude

δ = target declination

Total tracking times for Arecibo and Goldstone are shown in Fig. 1. The zenith angle limit for Arecibo is 20 degrees, so observations can be made only when the body lies within 20 degrees of a point directly overhead. The Arecibo tracking time for the planets varies from 2 to nearly 3 hours while a planet is at optimal northern declinations. The elevation limit for Goldstone is 15 degrees; this limit is set by the Federal Aviation Administration for radiation safety. The Goldstone tracking time for ecliptic objects varies from 6 to 12 hours. The Deep Space Network also operates another large 70-meter-diameter antenna at Tidbinbilla, Australia, which is located at a southern latitude within one degree of the northern latitude of Goldstone. Thus, a plot of the total tracking time for Tidbinbilla is simply the mirror image of the Goldstone curve in Fig. 1.

A second consideration for tracking time is that radar experiments consist of repeated cycles of transmission toward the observed body followed by reception of the echoes. These cycles are repeated as long as the body is within the tracking limits of the antenna. The time spent observing the echo is thus roughly half of the tracking time. However, when the special Goldstone-to-VLA bistatic configuration is considered, the Goldstone transmitter is assumed to be on for the entire tracking time, and the total echo reception time at the VLA becomes the entire tracking time as well. Thus, when the ratio (τ'/τ) is evaluated in Eq. (5), we used the ratio of tracking

times for the monostatic cases and multiplied this ratio by 1.414 for the bistatic case.

Other quantities in the relative sensitivity equation also vary with elevation-zenith angle. In particular, the Arecibo antenna gain and the receiver temperature vary with zenith angle as shown in Fig. 2.¹ The Goldstone antenna efficiency at 3.5 cm is expected to vary with elevation angle as shown in Fig. 3. These changes of the radar parameters with elevation-zenith angle must be accounted for in our computation of relative sensitivities (Eq. [5]). Several obvious approaches are (1) to replace antenna gain and temperature with an average; (2) to compute an average "sensitivity" for various elevation-zenith angles; or (3) to weigh the "sensitivities" for various elevation-zenith angles and then average these values. We used the third approach, as it simulates the procedures used in the data reduction of Arecibo asteroid observations (S. Ostro, personal communication). The equations for this approach are

$$RS = \frac{S_a}{S_g} \quad (7a)$$

where

RS = ratio of sensitivities

S_a = Arecibo sensitivities

S_g = Goldstone sensitivities

and

$$S = \frac{(W_1 S_1 + \dots + W_n S_n)}{(W_1 + \dots + W_n)} \quad (7b)$$

where

S = weighted sensitivity for Arecibo or Goldstone

W_i = weight of the i th observation

S_i = sensitivity of the i th observation

Furthermore, to relate the relative sensitivities given in Eqs. (5) and (7), we require

$$\begin{aligned} S_i &= P (G^2/T) (\tau/2)^{0.5} \lambda^{2.5} \\ W_i &= \frac{(S_i/S_{imax})}{T_{max}} \end{aligned} \quad (8)$$

¹D. B. Campbell, "Arecibo Observatory S-Band Radar Program," NASA Planetary Astronomy Management Committee memorandum (personal communication), 1986.

where P , G , T , and λ are defined in Eq. (1) and τ is the total track time given in Fig. 1 and Eq. (6). The weight is simply the ratio of S_i to the best S_i for the radar in question. Thus,

$$W_i = \text{constant} \cdot \left[\frac{(\text{antenna gain})^2}{\text{receiver temperature}} \right] \quad (9)$$

A summary of major system parameters is given in Table 1. For the VLA configuration, we assumed 27 antennas where each 25-meter antenna had an individual efficiency of 60 percent and a noise temperature of 35 K. We also assumed an array efficiency of 80 percent [2].

III. Relative Sensitivities

The equations for the relative sensitivities of earth-based radars were evaluated for various declinations, yielding the results plotted in Figs. 4, 5, and 6. Figure 4 shows the improvement in Goldstone sensitivity at the 3.5-cm wavelength in relation to both the aperture increase from 64 meters to 70 meters and the transmitter power increase from current (late 1980s) levels of 350 kW to the proposed 1000-kW level in the 1990s. Figure 5 shows the enhancement of Arecibo sensitivity at the 12.6-cm wavelength from current (late 1980s) capabilities to the proposed implementation of the Gregorian feed structure and with the transmitter power increase from current levels of 450 kW to 1000 kW in the 1990s. Figure 6 shows sensitivities for the mid-1980s versus those expected in the mid-1990s, when all Arecibo and Goldstone improvements are in place. The plots in Figs. 4, 5, and 6 are all scaled such that the maximum sensitivity in each figure is arbitrarily set to unity.

The improvements in Goldstone sensitivity at the 3.5-cm wavelength for the already completed aperture increase and

the proposed transmitter increase to 1000 kW (see Fig. 4) show an order-of-magnitude improvement between the mid-1980s and the mid-1990s. The improvements in Arecibo sensitivity at the 12.6-cm wavelength from the proposed Gregorian feed and the transmitter increase to 1000 kW (sketched in Fig. 5) show an increase of about a factor of 20. After these improvements, the Arecibo sensitivities have a broader peak in declination, indicating improved sensitivities for targets visible at Arecibo at zenith angles beyond ten degrees. A comparison of Goldstone and Arecibo before and after all of these improvements is sketched in Fig. 6. A horizon-to-horizon Goldstone-VLA observation (a rare experiment, since both antennas would simultaneously observe for around half a day) has about 3.5 times more sensitivity than the Goldstone monostatic observation, as shown in Fig. 6.

IV. Summary

The ability to perform radar astronomy experiments with the Arecibo and Goldstone radars will improve by a factor of ten in the next decade. The Arecibo and Goldstone radars are complementary in several ways. Goldstone can observe bodies with southern declinations beyond Arecibo tracking coverage. While bodies are in northern declinations, the Goldstone 3.5-cm and Arecibo 12.6-cm wavelengths provide dual-frequency observations of the same body.

The order-of-magnitude improvement in earth-based radar capability rests upon implementing a Gregorian feed and a 1000-kW S-band transmitter at Arecibo and/or implementing a 1000-kW X-band transmitter at Goldstone (with rare Goldstone-to-VLA experiments). All of these improvements are being pursued, promising a new era of planetary exploration with earth-based radars.

References

- [1] S. J. Ostro, "Planetary Radar Astronomy," in *Encyclopedia of Physical Sciences*, vol. 10, New York: Academic Press, pp. 613-614, 1987.
- [2] J. S. Ulvestad, G. M. Resch, and W. D. Brundage, "X-Band System Performance of the Very Large Array," *TDA Progress Report 42-92*, vol. October-December 1987, Jet Propulsion Laboratory, Pasadena, California, pp. 123-137, February 15, 1988.

Table 1. Radar parameters: Arecibo and Goldstone

Radar configuration	Arecibo		Goldstone	
	1980s	1990s	1980s	1990s
Transmitter power, kW	450	1000	350	1000
Peak antenna efficiency	0.55	0.60	0.47	0.63
Antenna diameter, meters	200	250	64	70
Peak antenna gain, dB	71.4	73.7	71.9	73.9
Wavelength, cm	12.6	12.6	3.5	3.5
Nominal receiver temperature, K	35	20	20	20

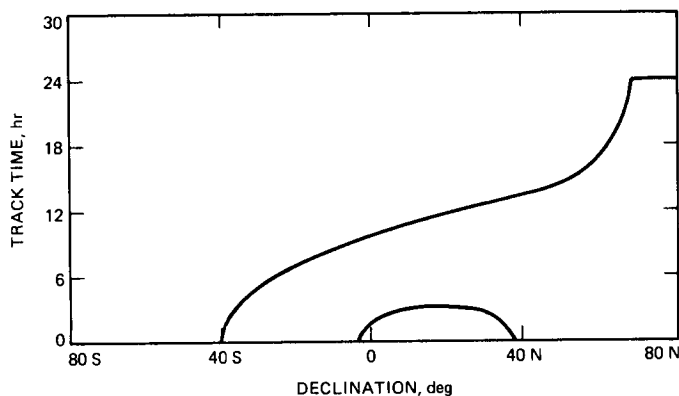


Fig. 1. Total tracking time for the Arecibo and Goldstone radars versus declination, assuming an Arecibo zenith angle limit of 20 degrees and a Goldstone elevation limit of 15 degrees

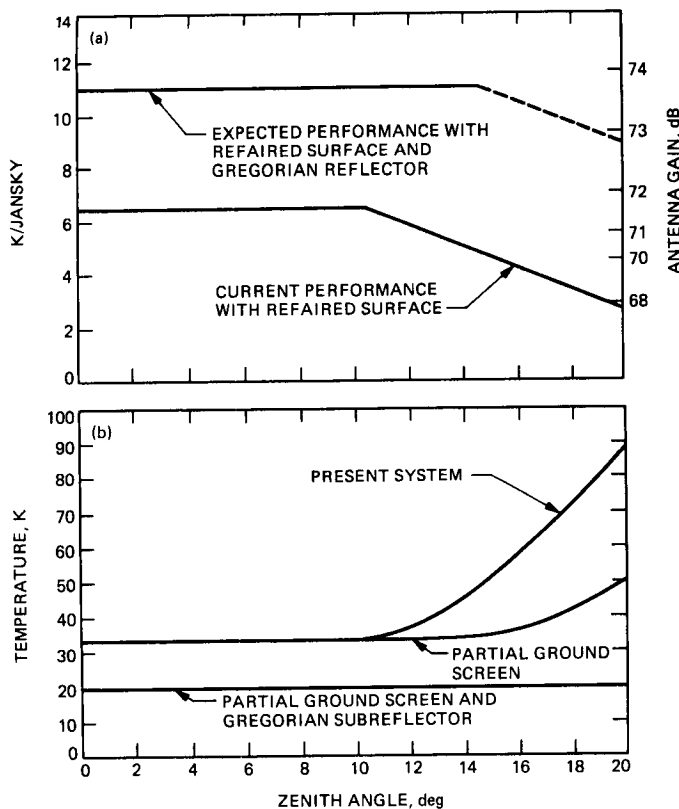


Fig. 2. Arecibo (a) antenna gain and (b) system temperature versus zenith angle for current conditions, the partial ground screen, and the Gregorian feed

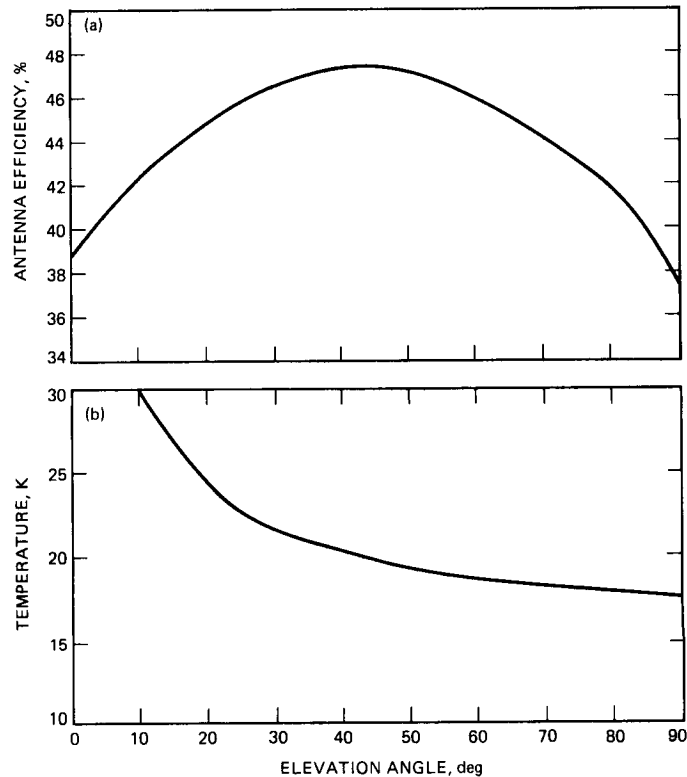


Fig. 3. Goldstone 3.5-cm (a) antenna efficiency and (b) system temperature versus elevation angle for current conditions. For the upgraded 70-meter antenna, a peak efficiency of 0.62 was assumed, with the same variation versus elevation angle as that shown above.

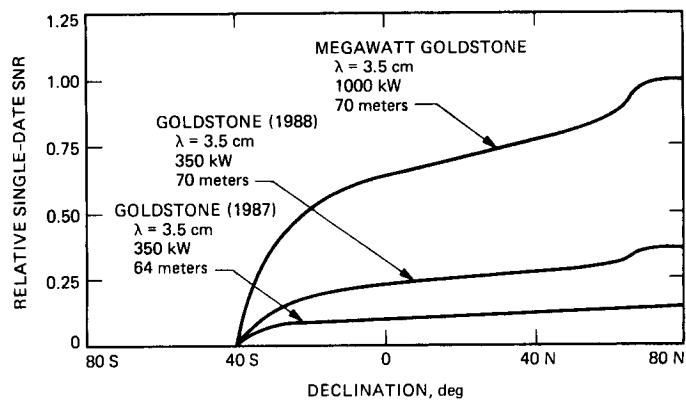


Fig. 4. Relative earth-based radar sensitivities for the Goldstone 3.5-cm radar, accounting for the already completed 64-meter to 70-meter aperture expansion and the proposed increase in transmitter power from 350 to 1000 kW

DSN 64-Meter Antenna L-Band (1668-MHz) Microwave System Performance Overview

J. Withington

Radio Frequency and Microwave Subsystems Section

In 1985, L-band (1668-MHz) receive-only feed systems were installed on the three DSN 64-meter antennas to provide tracking support for two non-NASA spacecraft. The specifications, design approach, and operational test results are presented in this article. The L-band microwave system met all of its tracking goals and is currently being upgraded to include a C-band (5000-MHz) uplink.

I. Introduction

In support of several international space exploration projects—the French/Soviet Vega mission to Venus (June 1985) and Halley's comet flybys (March 1986)—JPL was asked in late 1983 to modify the DSN to receive the L-band telemetry used by the Soviet space program. The major hardware implementation was undertaken by the JPL Radio Frequency and Microwave Subsystems Section, which was given the task of planning, designing, building, implementing, and documenting the microwave portions of fully operational (transferable) L-band receive systems.

II. System Requirements

The new L-band microwave system had to conform to and interface with the ongoing Mark IVA 64-meter antenna upgrade program and also had to be totally completed and operational within less than two years. Because of these constraints on time and resources, only the minimum microwave system necessary to support the immediate missions would be possible.

An extensive description of the Venus Balloon project and the L-band system requirements is given in [1]. Those requirements that affected the design of the microwave system are the following:

- (1) The antennas must receive 1668 ± 5 MHz.
- (2) Antenna gain must be at least 58 dBi, or 50 percent efficiency on a 64-meter antenna.
- (3) System noise temperature (T_{op}) must be <35 K at zenith.

Furthermore, the system required the ability to receive the LCP signal used by the Vega spacecraft and the Venus Balloon probe, and sensitivity needs required the use of refrigerated amplifiers.

III. Design Approach

The required 58-dBi gain precluded the use of all DSN antennas except the three 64-meter antennas. The design sequence of the microwave subsystem went through two itera-

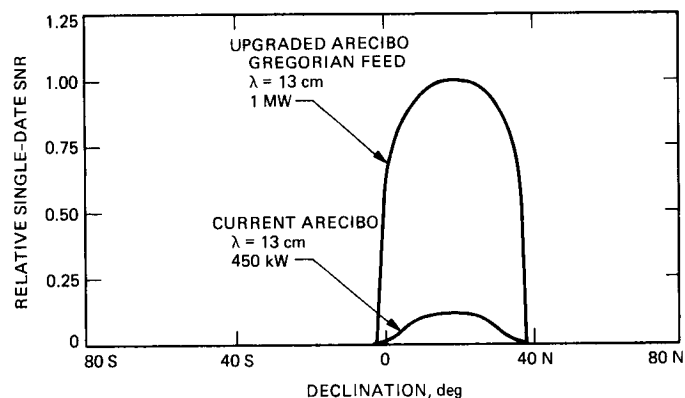


Fig. 5. Relative earth-based radar sensitivity for the Arecibo 12.6-cm radar for current and proposed Gregorian feed and transmitter improvements

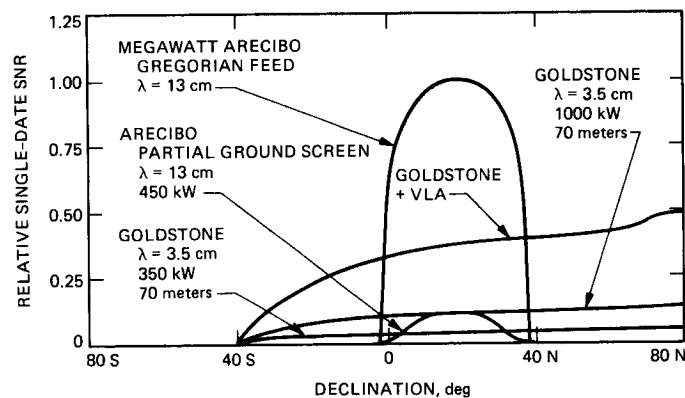


Fig. 6. Relative earth-based radar sensitivities for the Arecibo 12.6-cm radar in the 1980s and 1990s (before and after the Gregorian feed and increased transmitter power) as well as the Goldstone 3.5-cm radar in the 1990s (after the already completed aperture increase and the proposed increase in transmitter power)

tions. In the first analysis, it was envisioned that the "host country" feedcone position on the 64-meter (tricone-fed) antenna would be used. This would have involved the construction of three new feedcones outfitted with 22.4-dBi-gain L-band feedhorns. However, since this installation was to be permanent, a third operations feedcone on the 64-meter antenna was not acceptable to various science groups who use the third (host country) feedcone for radio science, radar, and other calibrations. Cost and implementation-time constraints also made this approach infeasible.

An alternative method, and the one that was implemented, was to suspend a feedhorn in the area between the XRO and the host country feedcone (Fig. 1). The idea was to cantilever a feedhorn on a bracket bolted to the top of the XRO feedcone. Since weight was then a critical concern, a sheet metal smooth-wall dual-mode horn 134 inches long, with a 38-inch aperture, was designed to be a light, low-loss, but narrowband (approximately ± 40 -MHz) solution with the illumination efficiency necessary to meet system requirements.

The outrigger horn suspension design presented one problem: the aperture of the feedhorn was too large to allow the phase center of the feedhorn to be placed on the focal ring of the asymmetrically fed 64-meter antenna. As a result, the phase center of the L-band feed lies some 24 inches radially outward from the focus ring (3.4 wavelengths), which produces a small scan loss in beam peak gain and beam pointing squint. Figure 2 shows the basic horn/antenna geometry used on the three 64-meter antennas. Initial analysis as reported in [2] predicted 60 percent efficiency, giving 58.8 dBi of gain. A simple scan angle equation using the antenna Cassegrainian magnification factor predicted a scan angle of 0.26 degree. A 58.8-dBi-gain antenna has a 0.2-degree half-power beamwidth.

A narrowband (1690 ± 50 MHz) quarter-wave plate polarizer was used to meet the circular polarization requirement. Spare DSN WR 430 waveguide components and a WR 430 switch completed the microwave feed system. No feedhorn pointing adjustments were provided because there was confidence that the fabrication techniques of machining the correct feedhorn orientations directly into the mounting brackets would provide the required accuracy. Since machining surfaces on the feedcones customarily provide alignment for their prospective feedhorns, it was felt that accurate mounting of the L-band feedhorn brackets on the feedcone surfaces would provide the necessary pointing alignment.

Two completely redundant, cryogenically cooled (physically cooled to 14 K) L-band FET LNAs provide the necessary preamplification. The FET amplifiers were designed with 38 dB of gain and a usable bandwidth of about 200 MHz. A bandpass filter in front of the amplifiers is used to limit the bandwidth

response of the FET amplifier to about 100 MHz (1668 ± 50 MHz). This was done to prevent out-of-band noise (at 2100 MHz) from the S-band transmitters of co-situated antennas in Spain and Australia and to prevent known RFI threats from saturating the preamplifiers. (A signal level approaching -40 dBm may be enough to saturate these FETs.)

The complete cooled FET system, with horn, polarizer, and waveguide, was assembled and tested to determine the microwave temperature contribution to the overall system T_{op} . The temperature contribution was determined to be approximately 10 K for the hardware and 14 K for the L-band FET LNA. Adding the temperature contributions and including 5 K for cosmic noise plus spillover and 1-K follow-on, the predicted system T_{op} was 30 K at zenith.

An L-band to S-band upconverter is used to convert the output of the L-band FETs to S-band. This allowed use of all station S-band receiver equipment necessary to meet the Vega telemetry processing requirements. In effect, the stations are transparent to the fact that L-band, not S-band, is being received. The upconverter further limits the bandwidth of the L-band system to 10 MHz, fixing the total bandwidth of the overall L-band receive system at 1668 ± 5 MHz.

A complete component-by-component description of the L-band microwave system can be found in the two L-band operation and maintenance manuals [3], [4]. Figure 3 is a block diagram of the complete system.

IV. Performance Measurements

The last of the L-band equipment was being installed on the 64-meter antennas at the time of the two Vega spacecraft launches. The remaining five-month period to the Venus encounter was used to track the Vega spacecraft, with very limited time to calibrate the L-band system. The minimal time available for antenna testing was used only to verify that the L-band microwave system met its design specification. This involved measuring the scan offset, system efficiency, and T_{op} . The 64-meter antenna 100 percent efficiency ratio of 1.166 K/jansky, along with the flux values of calibration sources (listed at 1665 MHz in the *Astronomical Almanac* [5]), was used to determine the antenna efficiency.

The measured values for gain, T_{op} , and scan offset on the three 64-meter antennas are shown in Table 1. As can be seen from the table, all efficiency values seem to be low by approximately 12 percent, and the T_{op} appears to be high by the same amount.

The possibility that more than 0.1 dB of loss was being caused by scan loss, beam broadening, or other antenna

anomalies was excluded when the measured scan offset and measured half-power beamwidth were considered. The predicted scan offset of about 0.26 degree was subsequently measured to be 0.26 degree, and the predicted half-power beamwidth of 0.20 degree was measured to be 0.19 degree. All measurements were felt to be within 5 percent.

Tracking azimuth and elevation offsets made necessary by the specific orientation of the L-band feedhorn were the following:

$$\begin{aligned} \text{El} &= -0.120 \text{ degree} \\ \text{Az} &= 0.232/\cos(\text{El}) \text{ degree} \end{aligned}$$

In subsequent tracking exercises, these values proved accurate enough to point on source to within 2 arc minutes over the 20- to 80-degree elevation range. This demonstrated the ability to blind point to less than 0.2 half-power beamwidth at L-band.

V. Analysis

Because of the 12 percent efficiency value differences obtained (see Table 1), a Physical Optics (PO) analysis was made of the L-band feed configuration on the 64-meter antenna. This analysis showed a spillover higher than that originally predicted. The PO analysis values at zenith, given in kelvins, are as follows:

Antenna temperature	8.5 (cosmic plus sky plus spillover)
Feed components	10.0
FET LNA	14.0
Station follow-on	1.0
Total	33.5

This total is compatible with measured data.

Similarly, the 0.6-dB loss difference between the measured and originally predicted values was resolved by PO analysis. The analysis predicted a scan offset angle of 0.26 degree, a half-power beamwidth of 0.19 ± 0.01 degree with a slightly elliptically shaped beam (0.01-degree difference), and an on-

scan axis gain of 59.46 dBi, or 70.6 percent efficiency (including a 0.03-dB scan loss and higher spillover) over a feedhorn placed on the focal ring of the 64-meter antenna. The 0.03-dB scan loss compares favorably with the approximate (0.05-dB) prediction published earlier [2]. The following additional antenna losses, expressed in decibels, must be subtracted from the PO result:

Surface RMS (97%)	0.13 dB
Spar and subreflector blockage (88%)	0.56
Feed dissipation losses (98%)	0.09
Feed mode losses (96%)	0.18
Total additional loss	0.96 dB

Adding all losses, the PO-based prediction is that the scan axis gain peak should be 58.5 dBi for an efficiency of 57 percent. This is compatible with the measured data. At least at DSS 14, another reason for the loss differential may be that these measurements were made before the full extent of FET saturation by RFI was understood. Some gain nonlinearity caused by saturation may account for the lower efficiencies reported here and by K. M. Liewer [6].

VI. Summary

This article includes all the measured data recorded during the L-band calibration sequence. Further work is needed to upgrade the L-band system to include a C-band uplink and to increase the bandwidth of the L-band to S-band upconverter. From what is currently understood of the RFI environment that exists at the Goldstone site, it is concluded that the radio science involved will not be degraded by RFI included in the wider bandwidths. It should be noted, however, that RFI spectrums of considerable power have been observed as close as 12 MHz from the Venus Balloon signal center frequency of 1668 MHz.

VII. Conclusion

The L-band microwave system met its design requirements, was successfully implemented in the short time allotted, and met all of its tracking goals.

Acknowledgments

The author would like to thank the scientists, engineers, and technicians who worked with and around him to make the microwave portion of this project a success. D. Jones provided information on blind pointing and efficiency testing at DSS 14. P. Parsons provided half-power beamwidth and offset measurements at DSS 14. A special thanks is extended to Art Freiley for doing the measurement at DSS 63 that enabled this article to be completed.

References

- [1] C. T. Stelzried, "The Venus Balloon Project," *TDA Progress Report 42-80*, vol. October–December 1984, Jet Propulsion Laboratory, Pasadena, California, pp. 195–201, February 15, 1985.
- [2] J. Withington, H. F. Reilly, and D. A. Bathker, "RF Performance of a Proposed L-Band Antenna System," *TDA Progress Report 42-75*, vol. July–September 1983, Jet Propulsion Laboratory, Pasadena, California, pp. 91–97, November 15, 1983.
- [3] *L-Band Receive Only Assembly*, JPL TM 514146, Jet Propulsion Laboratory, Pasadena, California, February 15, 1987.
- [4] *L-Band FET/CCR and Instrumentation*, JPL TM 13743, Jet Propulsion Laboratory, Pasadena, California, June 1, 1984.
- [5] *1987 Astronomical Almanac*, Washington, D. C.: U.S. Government Printing Office, p. H-62.
- [6] K. M. Liewer, "Selection of Radio Sources for Venus Balloon–Pathfinder Delta-DOR Navigation at 1.7 GHz," *TDA Progress Report 42-87*, vol. July–September 1986, Jet Propulsion Laboratory, Pasadena, California, pp. 279–284, November 15, 1986.

Table 1. Calculated and measured L-band microwave system efficiency and T_{op}

Source	Efficiency, % (gain) at approximately 45 degrees of elevation	T_{op} , K	Scan offset, degrees
Estimated [2]	60 (58.8 dBi)	30	0.260
Calculated (PO)	57 (58.5 dBi)	33	0.260
Measured at DSS 14	51 (58.2 dBi)	33 ^a	0.260
Measured at DSS 43	52 (58.2 dBi)	36 ^a	—
Measured at DSS 63 ^b	55 (58.4 dBi)	34 ^a	—

^a T_{op} determined using Y-factor methods.

^bMeasured by Art Freiley, Radio Frequency and Microwave Subsystems Section.

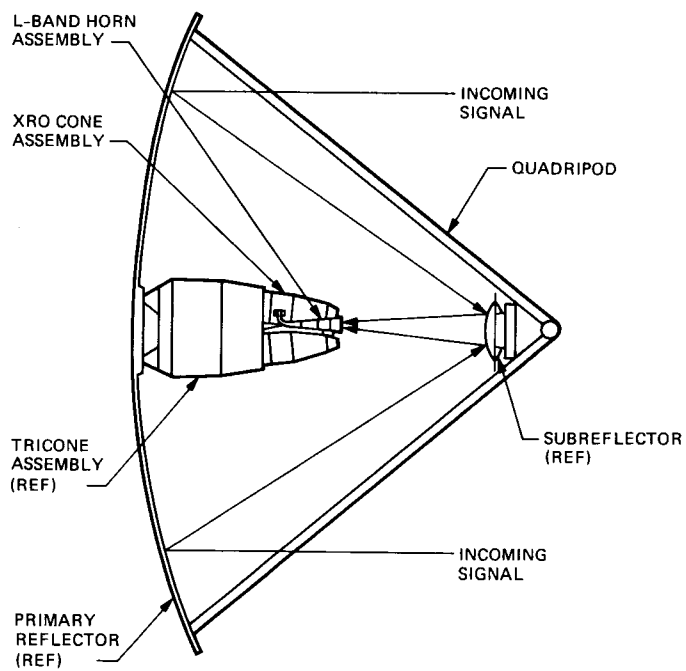


Fig. 1. L-band feedhorn antenna position

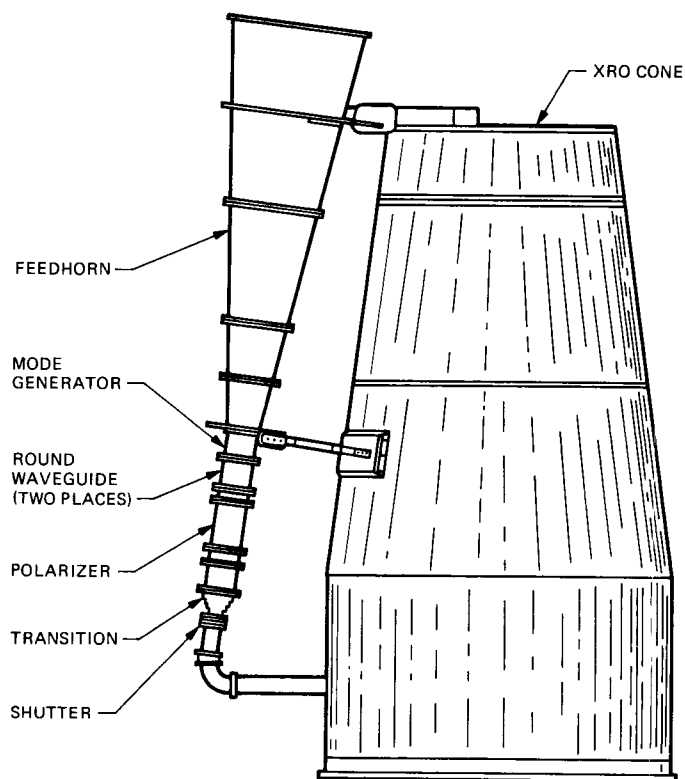


Fig. 2. L-band feedhorn geometry

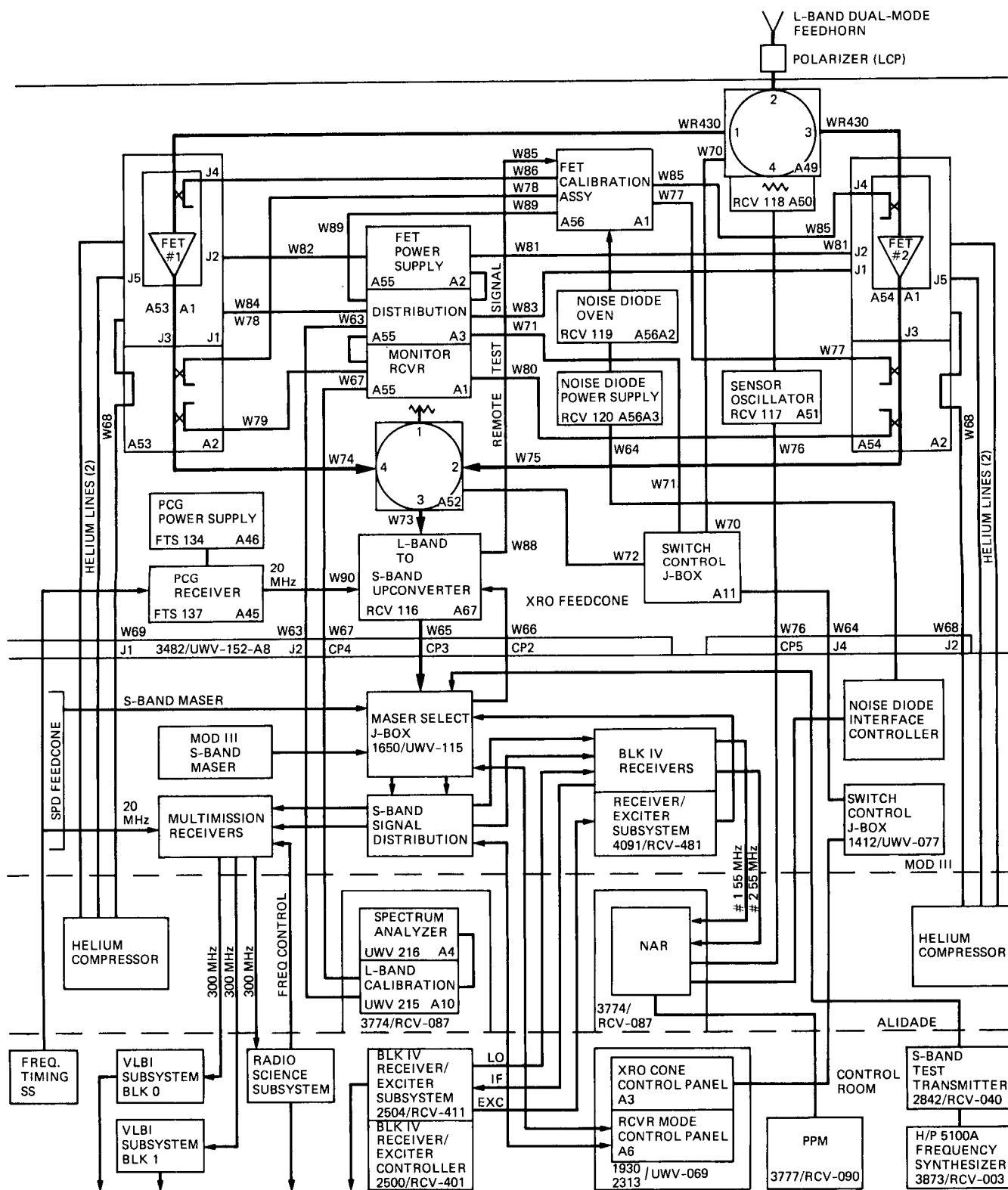


Fig. 3. Block diagram, L-band receive-only system

1989000840
618460
6p.525-14
~~45536~~
April-June 1988

N89 - 10211

JPL

Summary of DSN Reimbursable Launch Support

N. A. Fanelli and M. E. Wyatt
TDA Mission Support and DSN Operations

The DSN is providing ground support to space agencies of foreign governments as well as to NASA and other agencies of the federal government which are involved in space activities. DSN funding for support of missions other than NASA are on either a cooperative or a reimbursable basis. Cooperative funding and support are accomplished in the same manner as NASA-sponsored missions. Reimbursable launch funding and support methods are described.

I. Introduction

This article is a summary of how the DSN provides support for NASA reimbursable missions. Tables 1, 2, and 3 provide a summary of the missions that have been completed, approved missions to be supported in the near future, and the new proposed mission set. The tables list the missions, the responsible agency, the launch vehicle and launch site, the original and current launch dates, and the support period.

The process begins when other space agencies request NASA ground station support for the launch of their spacecraft. Normally the spacecraft are geostationary communications satellites, but they could range from low earth orbiters to deep space missions. Normally, a preliminary Support Instrumentation Requirements Document (SIRD) is sent to NASA by the requesting foreign agency. NASA tasks JPL to review the preliminary SIRD and to provide a cost estimate. NASA then attaches this estimate to the agreement with the foreign agency, and with that begins the DSN support commitment.

II. Prelaunch Preparation

A. Mission Support Documentation

1. **NASA Support Plan (NSP).** The Tracking and Data Systems Manager is responsible for the NSP. The NSP responds to the SIRD and represents NASA's commitment to the project.

2. **Network Operations Plan (NOP).** The Network Operations Project Engineer (NOPE) is responsible for the NOP. The NOP provides the operational procedures for project support.

3. **Interface Control Document (ICD).** The ICD is jointly prepared by the project and JPL. The project normally takes the lead and publishes the document. The ICD outlines the agreed-to data interfaces between the project and the DSN.

4. **Data Input Package (DIP).** The DIP is provided by the project. The DIP gives the DSN supporting organization de-

tailed information about the spacecraft, the network coverage profile, and the Sequence of Events (SOE).

5. Initial Acquisition Plan (IAP). The IAP is provided by the DSN. The IAP documents the initial acquisition procedure to be used by the DSN station or stations supporting the initial acquisition of signal after launch.

B. Tests and Training

Compatibility tests signal the start of the test and training phase, which begins about 12 to 14 months prior to launch and continues through the final prelaunch readiness test. Compatibility testing ensures that the spacecraft and the DSN ground stations are compatible. The project provides the compatibility test plan and the test criteria, while the DSN provides the compatibility test procedures. At the completion of the compatibility test, a test report is written documenting the test results.

Data flow tests are conducted at about launch minus 6 months. Their purpose is to verify that the project and DSN data interfaces are correct and work in the operational environment.

Ground data system tests are conducted at launch minus 2 months. These tests bring together all the supporting elements, which include the 26-m subnetwork, 34-m backup stations, DSN control center, project control center, NASCOM data/voice circuits, DSN operations control team, and project operations control team. The objectives of the ground data system tests are to test and verify all elements of the mission support systems and to provide operational training for all test participants.

Operational readiness tests are conducted at launch minus 1 month. They ensure that all supporting elements and teams are operationally ready to support the mission while also providing additional training for the mission support teams and a final check-out of operational procedures.

Initial acquisition tests are conducted at launch minus 2 weeks. The tests provide a check-out of the initial acquisition procedure, initial acquisition configuration, and operational team training.

Prelaunch data flow tests are conducted at launch minus 3 days and 1 day. These tests verify that the launch configuration for all elements of mission support are correct. The tested configuration is placed under configuration control to ensure that nothing changes prior to launch.

III. Working Group Meetings

A minimum of two working group meetings are held each year. One meeting is held at JPL, and one is held at the headquarters of the foreign agency. The meetings provide an excellent environment for the exchange of information between the project and the DSN operations and engineering personnel.

IV. Launch Support

On launch day, the worldwide communications circuits are in place at launch minus 3 hours. Status information is exchanged between the project and the DSN every half-hour.

The station prepass countdown is activated at launch minus 2 hours. The prepass preparations ensure that the stations are ready and properly configured for launch support. At launch plus 15 minutes, the stations are briefed on the upcoming acquisition support and are given any updates to the sequence of events or predict set to be used for initial signal acquisitions.

Most of the reimbursable missions supported by JPL are geostationary communications or broadcast satellites. A typical support period for launch, transfer, and drift orbit is 5 to 7 days. The project notifies JPL when mission support is completed and is to be terminated.

V. DSN Operations Launch Readiness Review

To ensure that all the necessary planning, testing, and training has been accomplished prior to a reimbursable mission launch, a formal Launch Readiness Review is held. The purpose of these reviews is to verify that any required implementation is complete and that an adequate test and training plan is in place and on schedule, with completion appropriately planned to accommodate the launch date.

VI. Special Precautions

The DSN takes special precautions for all critical activities to ensure readiness to provide the support required for mission success. It also investigates contingencies that could enhance the probability of mission success. The following paragraphs discuss several special precautions that are taken to ensure reimbursable mission success.

After the successful completion of the last launch rehearsal or launch operations readiness test (a few days before launch), the supporting DSN facilities (prime and backup stations and the control center) are placed under Modified Configuration

Control (MCC). MCC requires that DSN facilities obtain the approval of DSN Operations management before making configuration changes other than remove-and-replace in support of maintenance.

To provide maximum network reliability to the reimbursable missions, the DSN operates the supporting stations on diesel-generator power as opposed to commercial power. The Space Flight Operations Facility diesel generators are also operated at the request of the DSN to protect the Network Operations Control Center (NOCC) power integrity during the critical periods of the reimbursable mission support.

Station predicts, schedules, and sequences of events are created and transmitted by the Network Support Subsystem (NSS). During the launch phase, the ability to generate new predicts quickly based on updated launch parameters is critical to rapid and successful initial acquisition at the stations. Therefore, several precautions are taken to ensure that the NSS is available and not overloaded during the launch phase. Special on-call maintenance is contracted during the critical periods; the VAX System Manager or assistant is always on duty during critical periods; priorities of work are established and promulgated prior to the launch; nonessential work is

deferred; and dial-up utilities are disabled as part of the availability and reliability enhancement activities.

The DSN multimission navigation facility uses pointing vectors from various external agencies as well as DSN station radio metric data to generate predict files that are sent to the NSS for predict generation. Therefore, this critical resource is protected in much the same manner as the NSS, as described above.

Communications circuits between JPL and the stations are provided special contracted maintenance during the critical periods to ensure rapid recovery of circuits in case of failures.

At both the Control Center and the stations, technical specialists back up the operations crews on an advisory basis during critical activities to ensure rapid recognition of problems as well as rapid recovery.

Finally, to ensure that all personnel within the DSN as well as users of the DSN are aware of scheduled critical reimbursable mission launch activities, an awareness notification message is sent to all DSN facilities and to all projects using the DSN.

Table 1. Completed missions

Mission	Agency	Launch vehicle	Launch site	Original/ prepared launch date	Actual launch date	End support period
MS-T5	ISAS	M3S-2	KAG ^a	1/1985	1/7/85	1/17/85
MS-T5	Additional support					10/26/86– 12/10/86
Giotto	ESA	Ariane-1	KRU ^b	7/1985	7/2/85	3/18/86
PLANET-A	ISAS	M3S-2	KAG	8/1985	8/19/85	8/27/85
PLANET-A	Additional support					3/25/87– 4/15/87
BS-2B	NASDA	N-2	TAN ^c	8/1985	2/6/86	2/11/86
ASTRO-C	ISAS	MU-3S2	KAG	2/1987	2/7/87	2/8/87
ETS-V	NASDA	H-1	TAN	8/1987	8/27/87	9/1/87
TV-SAT	GSOC	Ariane-2	KRU	4/1985	11/20/87	11/23/87
CS-3A	NASDA	H-1	TAN	2/1988	2/19/88	2/22/88
TC1-C	CNES	Ariane-3	KRU	7/1986	3/11/88	3/13/88

^aKAG = Kagoshima, Japan^bKRU = Kourou, French Guiana^cTAN = Tanegashima, Japan**Table 2. Approved mission set**

Mission	Agency	Launch vehicle	Launch site	Original/ proposed launch date	Current launch date	Prime support period
CS-3B	NASDA	H-1	TAN ^a	8/1988	9/14/88	L + 5 days
TDF-1	CNES	Ariane-2	KRU ^b	8/1985	10/7/88	L + 8 days
DFS-1	GSOC	Ariane-3/4	KRU	9/1987	11/4/88	L + 8 days
TELE-X	CNES	Ariane-2/4	KRU	12/1987	3/1/89	L + 8 days
GMS-4	NASDA	H-1	TAN	8/1989	8/1/89	L + 5 days
EUTELSAT II-FI	GSOC	Ariane-3/4	KRU	10/15/89	1/15/90	L + 8 days

^aTAN = Tanegashima, Japan^bKRU = Kourou, French Guiana

Table 3. Proposed missions

Mission	Agency	Launch vehicle	Launch site	Proposed launch date	Current launch date	Prime support period
EXOS-D	ISAS	MU-3SH-4	KAG ^a (JKSC) ^b	2/1/89	2/1/89	L + 1 day
TDF-2	CNES	Ariane	KRU ^c	4/1989	9/15/89	L + 8 days
DFS-2	GSOC	Ariane	KRU	4/1989	9/15/89	L + 8 days
TV-SAT 2	GSOC	Ariane	KRU	10/1989	5/10/89	L + 8 days
BS-3A	NASDA	H-1	TAN ^d	10/1989	8/1/90	L + 5 days
BS-3B	NASDA	H-1	TAN	8/1991	8/1/91	L + 5 days
SFU	ISAS	H-2	KAG (JKSC)	1/1993		L + 7 days and STS recovery 7 days

^aKAG = Kagoshima, Japan
^bJKSC = Japanese Kagoshima Space Center
^cKRU = Kourou, French Guiana
^dTAN = Tanegashima, Japan

Referees

The following people have refereed articles for *The Telecommunications and Data Acquisition Progress Report*. By attesting to the technical and archival value of the articles, they have helped to maintain the excellence of this publication during the past year.

S. Aguirre	F. B. Estabrook	L. Maleki	M. K. Simon
G. Bagby	A. Freiley	W. L. Martin	J. R. Smith
D. Bathker	C. A. Greenhall	J. R. Matijevic	J. I. Statman
G. Blewitt	R. W. Hartop	G. Morris, Jr.	C. Stelzried
F. Borncamp	C. E. Hildebrand	J. W. Osman	R. Stevens
D. Brown	R. E. Hill	S. J. Ostro	L. Swanson
B. K. Chaney	S. D. Howard	T. Pen	R. L. Sydnor
B. Charny	R. S. Hughes	S. Petty	J. B. Thomas
C.-C. Chen	D. L. Johnson	F. Pollara	D. L. Trowbridge
U. Cheng	D. L. Jones	M. D. Rayman	T. K. Tucker
K. M. Cheung	R. F. Jurgens	R. L. Riggs	C. C. Wang
C. Chian	E. L. Kerr	M. Rodriquez	C. W. Wang
C. S. Christensen	M. J. Klein	D. H. Ragstad	M. Whang
R. Clauss	W. H. Kuehn	R. Sadr	W. M. Whitney
R. B. Crow	C. R. Lahmeyer	E. H. Satorius	C. P. Wiggins
J. M. Davidson	F. L. Lansing	B. L. Schumaker	B. G. Williams
R. M. Dickinson	G. Lanyi	J. A. Schwartz	G. Wood
D. Divsalar	J. W. Layland	M. Shahshahani	S.-C. Wu
S. Dolinar	J. R. Lesh	J. S. Shell	
J. Ellis	S. Lichten	D. L. Skinner	