The Telecommunications and Data Acquisition Progress Report 42-109

January–March 1992

E. C. Posner
Editor

May 15, 1992

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

The Telecommunications and Data Acquisition Progress Report 42-109

January–March 1992

E. C. Posner
Editor

May 15, 1992

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

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

Preface

This quarterly publication provides archival reports on developments in programs managed by JPL's Office 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) in planning, in supporting research and technology, in implementation, and in operations. Also included is standards activity at JPL for space 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). The TDA Office also performs work funded by two other NASA program offices through and with the cooperation of the Office of Space Operations. These are the Orbital Debris Radar Program (with the Office of Space Station) and 21st Century Communication Studies (with the Office of Aeronautics and Exploration Technology).

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 the Office of Space Operations for funding DSN operational support.

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, for 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
- Ground Communications Implementation and Operations
- Data and Information Systems
- Flight-Ground Advanced Engineering
- Long-Range Program Planning

OSO Cooperative Tasks:
- Orbital Debris Radar Program
- 21st Century Communication Studies

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

COMMUNICATIONS, SPACECRAFT–GROUND

STATION CONTROL AND SYSTEM TECHNOLOGY

NETWORK DATA PROCESSING AND PRODUCTIVITY

DSN Systems Implementation

NETWORK UPGRADE AND SUSTAINING

DSN Operations

MISSION INTERFACE AND SUPPORT

TDA PROGRAM MANAGEMENT AND ANALYSIS

Ground Communication Implementation and Operations

Flight–Ground Advanced Engineering

OSSA TASKS
Goldstone Solar System Radar

A Determination of the Radio-Planetary Frame Tie From Comparison of Earth Orientation Parameters

M. H. Finger[1] and W. M. Folkner
Tracking Systems and Applications Section

The orientation of the reference frame of radio source catalogs relative to that of planetary ephemerides, or "frame tie," can be a major systematic error source for interplanetary spacecraft orbit determination. This work presents a method of determining the radio-planetary frame tie from a comparison of very long baseline interferometry (VLBI) and lunar laser ranging (LLR) station coordinate and Earth orientation parameter estimates. A frame tie result is presented with an accuracy of 25 nrad.

I. Introduction

Very long baseline interferometry (VLBI) offers interplanetary spacecraft navigation a highly accurate data type for orbit determination. The most commonly used data type, delta differential one-way ranging (ΔDOR), provides 50-nrad (or better) information about the angular position of a spacecraft relative to a nearby radio source [1]. The positions of radio sources within the inertial reference frame defined by extragalactic radio sources are typically known to 3–5 nrad [2]. The planetary ephemeris defines a separate inertial reference frame [3]. Knowledge of the relative global rotation, or frame tie, between these two inertial reference frames is necessary to take full advantage of VLBI tracking of interplanetary spacecraft. The uncertainty in the frame-tie calibration can be the dominant orbit determination error for inner planet approach navigation [4].

In addition to providing a navigational data type, VLBI plays another important role in spacecraft orbit determination. VLBI radio source observations are used to monitor, with 10-nrad accuracy, the orientation of the Earth with respect to inertial space. The time dependent transformation from terrestrial to celestial coordinates is expressed in terms of universal time and polar motion (UTPM) parameters and precession and nutation corrections that result from fitting the VLBI data. Since these observations refer the orientation of Earth to the radio reference frame, the radio-planetary frame tie will affect interplanetary orbit determination even when no VLBI observations of the spacecraft are employed.

Lunar laser ranging (LLR) is an alternate technique for monitoring the orientation of the Earth with respect to inertial space. The LLR observations refer the orientation of the Earth to the lunar ephemeris, which gives LLR tracking station locations in the lunar ephemeris frame. Through the effect of solar perturbations on the lunar or-

[1] Now with Astronomy Programs, Computer Sciences Corporation.

bit, the LLR data are sensitive to the ecliptic plane and the direction to the sun. The orientation of the planetary ephemeris system, as defined by the Earth's orbit, can therefore be tied to the lunar ephemeris system with about 10-nrad accuracy [3].

Since both LLR and VLBI measure the orientation of the Earth with 10-nrad accuracy, it should be possible to determine the frame tie from a comparison of these measurements. An earlier attempt to do this by Niell[2] was limited by a restricted ability to determine the orientation of the terrestial frames for VLBI and LLR. Now that there are three well-determined LLR station locations, and because of recent efforts to unify terrestrial coordinate systems [5], it is possible to determine the frame tie with 15–25 nrad accuracy from a comparison of LLR and VLBI Earth orientation series.

The theoretical foundation for this comparison is established in Sections II and III. Section II examines in detail the nature of the time dependent terrestrial–celestial ties mentioned above. In Section III an expression is derived for the radio-planetary frame tie in terms of the LLR to VLBI station-coordinate system tie and the parameters of the two terrestrial–celestial ties. The LLR and VLBI solutions compared in this analysis are presented in sections IV and V, respectively.

In Section VI, a determination of the LLR to VLBI station coordinate transformation is presented. Since the VLBI stations and LLR sites are widely separated, it is necessary to use other data to bridge the two coordinate systems. A recent Crustal Dynamics Project (CDP) VLBI station location set, which includes mobile VLBI observations at LLR sites, is used to connect the LLR and VLBI terrestrial coordinate systems.

With this terrestrial tie determined, the planetary ephemeris to radio source catalog frame tie can be determined from an intercomparison of the VLBI and LLR nutation and Earth orientation parameters. This comparison and the resulting frame tie are presented in Section VII. The derived frame tie is compared with other available results in Section VIII. Some comments on how to best use the results of this work for spacecraft navigation are included in Section IX.

[2] A. E. Niell, "Absolute Geocentric DSN Station Locations and the Radio-Planetary Frame Tie," JPL Interoffice Memorandum 335.2-159 (internal document), Jet Propulsion Laboratory, Pasadena, California, March 21, 1984.

Much of this work parallels the current efforts of the I ternational Earth Rotation Service (IERS). IERS is in t process of comparing and unifying terrestrial VLBI, LL and satellite laser ranging (SLR) coordinate systems ar unifying different VLBI celestial coordinate systems wi the goal of testing the consistency of various Earth rot tion parameter series [6–8]. Where possible, the notatic used here is consistent with that of the IERS.

II. Ties Between Celestial and Terrestrial Frames

In the process of reducing LLR or VLBI data, a time d pendent transformation between implicitly defined cele tial and terrestrial coordinate systems is established. Th transformation represents a dynamic tie between Eart fixed and inertial frames and includes estimated and a sumed precession, nutation, and Earth orientation paran eters. To employ this transformation, it must be unde stood in some detail. To this end, the standard represent tion of the orientation of the Earth with respect to inerti space is presented here. Particular attention is paid t the quantities that are commonly estimated and how the affect this representation.

Let $\vec{X}$ represent the Earth-fixed coordinate vector a station in an equatorial coordinate system with the axis nominally aligned with the Greenwich meridian.[3] Th station's instantaneous (J2000) celestial coordinate vect $\vec{C}$ at time t is calculated as

$$\vec{C} = \mathbf{P}\,\mathbf{N}\,\mathbf{S}\,\mathbf{O}\,\vec{X} \qquad (1$$

The (polar motion) rotation **O** corrects for the offset b tween the Earth-fixed coordinate pole and the Celesti Ephemeris Pole (CEP). The CEP is conceptually define as the axis which, in the theory of the rotation of th Earth, has no forced daily or semi-daily nutations [9,10 **S** models the rotation of the Earth about the CEP, **N** a counts for the quasi-periodic nutation of the CEP abo the "mean pole of date," and **P** models the precessio or secular drift, of the "mean pole of date" and "mea equinox of date" with respect to the celestial fixed po and equinox of J2000. Each of these rotations is discusse in detail below.

All together, eight or more angles are used to describ the rotation between terrestrial and celestial coordinate

[3] The time dependence of the Earth-fixed location due to plate m tion is ignored throughout this section.

(including three precession angles, two nutation angles, two polar motion angles, and UT1–UTC). The conceptual definitions of these angles are largely a product of the historical development of the theory of the Earth's rotation. While the overall rotation can be experimentally determined, many of the intermediate rotation angles have no precise empirical definition, and thus cannot be uniquely measured. The philosophy that will be adopted here is that in the final analysis, only the total rotation matrix has a well-defined physical meaning.

A. Notation

In order to discuss modeling of Earth's rotation in detail, a notation for positive rotations $\mathbf{R}_X$, $\mathbf{R}_Y$, and $\mathbf{R}_Z$ about the x-, y-, and z-axis, respectively, is introduced, where $\mathbf{R}_X$, $\mathbf{R}_Y$, and $\mathbf{R}_Z$ are defined by

$$\mathbf{R}_X(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{pmatrix}$$

$$\mathbf{R}_Y(\theta) = \begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix} \tag{2}$$

$$\mathbf{R}_Z(\theta) = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

These rotations are positive in the sense that they represent a transformation between two coordinate systems with the final coordinate system's basis vectors being rotated from the initial system's basis vectors by a right-handed rotation of angle θ about the designated axis.

A rotation about an arbitrary axis will be defined by

$$\mathbf{R}(\vec{\Theta})\vec{r} = \vec{r} - \sin\theta\,\hat{\Theta}\times\vec{r} + (1-\cos\theta)\hat{\Theta}\times(\hat{\Theta}\times\vec{r}) \tag{3}$$

where $\theta = |\vec{\Theta}|$ is the angle of rotation, $\hat{\Theta} = \Theta/\theta$ is the rotation axis, and $\vec{r}$ is an arbitrary coordinate vector. For example, with this notation $\mathbf{R}_X(\theta) = \mathbf{R}(\theta\hat{e}_X)$, $\mathbf{R}_Y(\theta) = \mathbf{R}(\theta\hat{e}_Y)$, and $\mathbf{R}_Z(\theta) = \mathbf{R}(\theta\hat{e}_Z)$, where $\hat{e}_X$, $\hat{e}_Y$, and $\hat{e}_Z$ are the x- , y- , and z-unit vectors, respectively. Two results will prove useful in connection with this notation. First, if $\mathbf{M}$ is a rotation matrix, then it can be shown that

$$\mathbf{M}\mathbf{R}(\vec{\Theta})\mathbf{M}^{-1} = \mathbf{R}(\mathbf{M}\vec{\Theta}) \tag{4}$$

Equation (4) follows from Eq. (3) and the invariance of the cross product under orthonormal coordinate transformations ($\mathbf{M}[\vec{A}\times\vec{B}] = [\mathbf{M}\vec{A}]\times[\mathbf{M}\vec{B}]$). A second result is the approximation rule for small rotation vectors,

$$\mathbf{R}(\vec{B})\mathbf{R}(\vec{A}) \approx \mathbf{R}(\vec{A}+\vec{B}+\vec{A}\times\vec{B}/2) \tag{5}$$

which is accurate through second order.

B. Terrestrial Pole Orientation

The first rotation applied in the transformation from terrestrial to celestial coordinates is the orientation matrix $\mathbf{O}$, which accounts for polar motion, the offset of the CEP from the terrestrial coordinate system pole

$$\mathbf{O}(t) = \mathbf{R}\left(\begin{pmatrix} y \\ x \\ 0 \end{pmatrix}\right) \tag{6}$$

The angles x and $-y$ are approximately the x and y coordinates of the CEP in the Earth-fixed system.

C. Rotation About the Pole

The vast majority of the rotational velocity of the Earth is modeled in the spin matrix

$$\mathbf{S}(t) = \mathbf{R}_Z(-\theta_G) \tag{7}$$

where θ_G represents Greenwich Mean Sidereal Time, the hour angle between the meridian containing both the terrestrial x-axis and the CEP and the meridian containing both this pole and the mean equinox of date. The equation of the equinoxes, which is normally included with the spin rotation $\mathbf{S}$, will be incorporated below in the nutation matrix $\mathbf{N}$. It should be noted that small rotational velocities occur due to precession, nutation, and polar motion, and therefore the CEP is not the rotation axis of the Earth's crust. By definition, the Earth-rotation-based time scale UT1 is directly related to θ_G, with the explicit relationship given by [11]. When UT1 is estimated, the spin matrix may be represented as

$$\mathbf{S}(t) = \mathbf{S}_0(t)\mathbf{R}_Z(-\Omega[\mathrm{UT1}-\mathrm{UTC}]) \tag{8}$$

where Ω is the mean rotation rate of the Earth and $\mathbf{S}_0(t)$ is the value of $\mathbf{S}(t)$ obtained by assuming UT1 = UTC.

D. Nutation

Nutation describes the short-term quasi-periodic variations in the CEP. The longest period terms have an 18.6-year period with an amplitude of 43 μrad. The standard model for nutation is given by

$$\mathbf{N}(t) = \mathbf{R}_X(-\epsilon)\mathbf{R}_Z(\Delta\psi)\mathbf{R}_X(\epsilon+\Delta\epsilon)\mathbf{R}_Z(-\alpha_E) \quad (9)$$

where $\Delta\psi$ is the nutation in the ecliptic longitude of the pole, and $\Delta\epsilon$ is its nutation in obliquity (the angle from the ecliptic pole). Here, the equation of the equinoxes

$$\alpha_E = \Delta\psi\cos(\epsilon+\Delta\epsilon) \quad (10)$$

has been included with the nutation since the equation of the equinoxes depends only on nutation parameters. Fourier series for $\Delta\psi$ and $\Delta\epsilon$ for the standard (IAU 1980) model are given by Wahr [12] or Seidelmann [9].

To provide a better understanding of the nutation matrix, an approximate formula for it will be derived using Eqs. (4) and (5). First, using Eq. (5), the nutations can be grouped together:

$$\mathbf{N}(t) \approx \mathbf{R}_X(-\epsilon)\mathbf{R}\left(\begin{pmatrix}\Delta\epsilon\\0\\\Delta\psi\end{pmatrix}\right)\mathbf{R}_X(\epsilon)\mathbf{R}_Z(-\alpha_E)$$

The nutations $\Delta\psi$ and $\Delta\epsilon$ are applied in ecliptic coordinates, with $\mathbf{R}_X(\epsilon)$ representing a transformation into ecliptic coordinates, and $\mathbf{R}_X(-\epsilon)$ representing a transformation back to equatorial coordinates. By using Eq. (4) with $\mathbf{R}_X(-\epsilon)$ as $\mathbf{M}$, the nutation matrix may be reduced to a series of small rotations:

$$\mathbf{N}(t) \approx \mathbf{R}\left(\begin{pmatrix}\Delta\epsilon\\-\Delta\psi\sin\epsilon\\\Delta\psi\cos\epsilon\end{pmatrix}\right)\mathbf{R}_Z(-\alpha_E)$$

By using the explicit form [Eq. (10)] for the equation of the equinoxes, and by collecting together small rotations by using Eq. (5), the approximation

$$\mathbf{N}(t) \approx \mathbf{R}\left(\begin{pmatrix}\Delta\epsilon\\-\Delta\psi\sin\epsilon\\0\end{pmatrix}\right) \quad (11)$$

is obtained, which is accurate to a few nrad.

When nutation is estimated, the corrected nutation matrix can be represented as

$$\mathbf{N}(t) = \mathbf{N}_{IAU}(t)\mathbf{R}\left(\begin{pmatrix}\delta\epsilon\\-\delta\psi\sin\epsilon\\0\end{pmatrix}\right) \quad (12)$$

where $\mathbf{N}_{IAU}$ is the (unapproximated) standard model and $\delta\epsilon$ and $\delta\psi$ are corrections to $\Delta\epsilon$ and $\Delta\psi$.

E. Precession

Precession describes the long-term drift of the mean pole of date and mean equinox of date. The mean pole drifts in declination by $n \approx 97.16$ μrad per Julian year, and the mean equinox drifts in right ascension by $m \approx 223.60$ μrad per Julian year [13]. The standard model for the precession is given in the form

$$\mathbf{P}(t) = \mathbf{R}_Z(\zeta_A)\mathbf{R}_Y(-\theta_A)\mathbf{R}_Z(z_A) \quad (13)$$

Polynomial expressions for the angles ζ_A, θ_A, and z_A as a function of t are given by Lieske [13,14]. These are, however, physically uninformative. A clearer vectorial formulation of precession is given by Fabri [15].

When corrections to the standard precession model are estimated, the corrected precession matrix may by represented as [16]

$$\mathbf{P}(t) = \mathbf{P}_{IAU}(t)\mathbf{R}\left(\begin{pmatrix}0\\-\delta n\\\delta m\end{pmatrix}(t-t_D)\right) \quad (14)$$

where $\mathbf{P}_{IAU}$ is the standard model [Eq. (13) with Lieske's polynomials]; δn and δm are corrections to the precession rates in declination and right ascension, respectively; and t_D is a reference epoch, which is preferably near the mean data epoch. The corrections to the general precession in

declination and right ascension, δn and δm, may be expressed as

$$\begin{aligned} \delta n &= \delta p_1 \sin\epsilon \\ \delta m &= \delta p_1 \cos\epsilon - \delta\dot{\chi} \end{aligned} \tag{15}$$

with ϵ representing the mean obliquity of the ecliptic, δp_1 representing the correction to the luni-solar precession in longitude, and $\delta\dot{\chi}$ representing the correction to the planetary precession in right ascension.

F. The Total Effect of Estimated Quantities

By combining the above results and using Eqs. (4) and (5), one finds that the transformation from terrestrial to celestial coordinates, including estimated quantities, may be expressed as

$$\mathbf{PNSO} = \mathbf{P}_{IAU}\mathbf{N}_{IAU}\mathbf{S}_0\mathbf{R}(\vec{\Theta}) \tag{16}$$

where

$$\vec{\Theta} = \mathbf{S}_0^{-1}\mathbf{N}_{IAU}^{-1}\begin{pmatrix} 0 \\ -\delta n \\ \delta m \end{pmatrix}(t - t_D) + \mathbf{S}_0^{-1}\begin{pmatrix} \delta\epsilon \\ -\delta\psi\sin\epsilon \\ 0 \end{pmatrix} + \begin{pmatrix} y \\ x \\ \Omega(\mathrm{UTC} - \mathrm{UT1}) \end{pmatrix} \tag{17}$$

or, neglecting the effect of the nutation matrix on the precession corrections (this is less than 0.1 nrad),

$$\vec{\Theta} = \begin{pmatrix} \delta\epsilon\cos\theta_G - (\delta n(t - t_D) + \delta\psi\sin\epsilon)\sin\theta_G + y \\ -\delta\epsilon\sin\theta_G - (\delta n(t - t_D) + \delta\psi\sin\epsilon)\cos\theta_G + x \\ \delta m(t - t_D) + \Omega(\mathrm{UTC} - \mathrm{UT1}) \end{pmatrix} \tag{18}$$

Several points should be noted from this relationship. First, estimation of precession in declination δn is equivalent to estimating a linear trend in the nutation in longitude $\delta\psi$. It can therefore be neglected in the following analysis without loss of generality. Second, estimation of precession in right ascension δm is equivalent to estimating a trend in UT1–UTC. As has been discussed by Williams and Melbourne [17], when corrections to precession are adopted in the future, the definition of UT1 should be altered so that the UT1 series is continuous. Guinot has gone further by suggesting that rather than referring UT1 to the meridian of the mean equinox, it should be referred to a "nonrotating origin," which is defined on the instantaneous Earth equator so as to be largely independent of precession and nutation models [18]. In the light of this thinking it makes no sense to estimate the precession in right ascension δm. Finally, it should be noted that on time scales short compared to a day, it is impossible to distinguish between nutation and polar motion; only three angles are needed to describe a general rotation. In fact, if nutations were allowed to have rapid variations with nearly daily periods, there would be no need for the polar angles x and y. Thus, whatever the conceptual definition of the CEP, its actual implementation results from fitting data to slowly varying nutation and polar motion models.

III. Relating the VLBI, LLR, and Planetary Ephemeris Reference Systems

In this section, the relationships between the celestial and terrestrial reference frames for VLBI and LLR will be discussed. The end result will be an expression relating the radio-planetary frame tie to quantities available from the LLR and VLBI data reductions and the tie between LLR and VLBI terrestrial coordinate systems. The

radio-planetary frame tie will be represented by a rotation vector $\vec{A}$ that relates VLBI celestial coordinates $\vec{C}_{VLBI}$ and planetary ephemeris coordinates $\vec{C}_{PE}$ by

$$\vec{C}_{VLBI} = \mathbf{R}(\vec{A})\vec{C}_{PE} \tag{19}$$

The derivation of the frame tie starts with the planetary ephemeris, as represented by the ephemeris of the Earth (which is the celestial reference frame for LLR), and proceeds in steps to the LLR terrestrial coordinate system, the VLBI terrestrial system, and finally to the VLBI celestial system.

For the LLR data reduction, terrestrial coordinates $\vec{X}_{LLR}$ and celestial coordinates $\vec{C}_{PE}$ are related by

$$\vec{C}_{PE} = \mathbf{P}_{IAU}\mathbf{N}_{IAU}\mathbf{S}_0\mathbf{R}(\vec{\Theta}_{LLR})\vec{X}_{LLR} \tag{20}$$

where $\vec{\Theta}_{LLR}$ has the form of Eq. (18). It should be noted here that many LLR data reductions estimate corrections to the planetary ephemeris. This derivation is limited to the case where no such corrections are estimated and overall orientation variations are absorbed by estimated precession and nutation parameters.

The complete transformation from LLR to VLBI terrestrial coordinates must account for a rotation, translation, and a possible difference of scale. This will be discussed in Section VI. Here the main concern is with the relative orientation of the two coordinate systems. Therefore, in this section the relationship between direction coordinates in the two systems will be represented as

$$\vec{X}_{VLBI} = \mathbf{R}(\vec{R})\vec{X}_{LLR} \tag{21}$$

where the vector $\vec{R}$ parameterizes the rotation between the LLR and VLBI terrestrial coordinate systems.

In the VLBI data analysis, terrestrial coordinates $\vec{X}_{VLBI}$ and celestial coordinates $\vec{C}_{VLBI}$ are related by

$$\vec{C}_{VLBI} = \mathbf{P}_{IAU}\mathbf{N}_{IAU}\mathbf{S}_0\mathbf{R}(\vec{\Theta}_{VLBI})\vec{X}_{VLBI} \tag{22}$$

where $\vec{\Theta}_{VLBI}$ has the form of Eq. (18). The parameterization of $\vec{\Theta}_{VLBI}$ estimated in the VLBI data reduction will be discussed in Section V.

By tracing the coordinate transformation

$$\vec{C}_{VLBI} \leftarrow \vec{X}_{VLBI} \leftarrow \vec{X}_{LLR} \leftarrow \vec{C}_{PE}$$

the planetary-radio frame tie is found to be given by

$$\mathbf{R}(\vec{A}) = \mathbf{P}_{IAU}\mathbf{N}_{IAU}\mathbf{S}_0\mathbf{R}(\vec{\Theta}_{VLBI})\mathbf{R}(\vec{R}) \times \mathbf{R}(-\vec{\Theta}_{LLR})\mathbf{S}_0^{-1}\mathbf{N}_{IAU}^{-1}\mathbf{P}_{IAU}^{-1} \tag{23}$$

By using Eq. (4) and Eq. (5), this reduces to

$$\vec{A} = \mathbf{P}_{IAU}\mathbf{N}_{IAU}\mathbf{S}_0[\vec{\Theta}_{VLBI} + \vec{R} - \vec{\Theta}_{LLR}] \tag{24}$$

By neglecting the effect of precession and nutation on small quantities (this is less than 0.5 nrad), the components of this equation are given by

$$\left.\begin{aligned} \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} &= \begin{pmatrix} \delta\epsilon_{VLBI} - \delta\epsilon_{LLR} \\ -(\delta\psi_{VLBI} - \delta\psi_{LLR})\sin\epsilon \end{pmatrix} + \begin{pmatrix} \cos\theta_G & -\sin\theta_G \\ \sin\theta_G & \cos\theta_G \end{pmatrix} \begin{pmatrix} R_1 + y_{VLBI} - y_{LLR} \\ R_2 + x_{VLBI} - x_{LLR} \end{pmatrix} \\ A_3 &= R_3 - \Omega(\mathrm{UT1}_{VLBI} - \mathrm{UT1}_{LLR}) \end{aligned}\right\} \tag{25}$$

where precession in declination is included as a trend in $\delta\psi$. As argued earlier, nutation and precession can be separated from polar motion only by requiring that each be slowly varying. Therefore, the terms in Eq. (25) that are modulated by sinusoids in θ_G, the Greenwich Mean Sidereal Time, must be separately zero. This gives the

bias between the two polar motion series in terms of the terrestrial transformation parameters R_1 and R_2, which correspond to a displacement of the coordinate pole:

$$\begin{aligned} y_{LLR} - y_{VLBI} &= R_1 \\ x_{LLR} - x_{VLBI} &= R_2 \end{aligned} \tag{26}$$

The frame-tie rotation vector is then given by

$$\vec{A} = \begin{pmatrix} \delta\epsilon_{VLBI} - \delta\epsilon_{LLR} \\ -(\delta\psi_{VLBI} - \delta\psi_{LLR})\sin\epsilon \\ R_3 - \Omega(\mathrm{UT1}_{VLBI} - \mathrm{UT1}_{LLR}) \end{pmatrix} \tag{27}$$

Equation (27) shows how a full three-dimensional radio-planetary frame tie may be deduced. Its use requires a comparison of LLR and VLBI nutation and UT1 estimates, and a determination of the transformation between VLBI and LLR terrestrial coordinate systems.

IV. The LLR Solution Set

The LLR solution employed here was provided by Newhall, Williams, and Dickey, and is similar to results published by IERS as solution JPL 90 M 01 [19]. However, for this particular solution, no corrections to the planetary ephemeris were estimated. The solution included twenty years of data from August 1969 to January 1989. Station locations, reflector locations, lunar gravity, lunar ephemeris, nutation, precession, and UT0 parameters were estimated from the LLR data. The planetary ephemeris used in the fit was DE200 [20,21] with an updated lunar ephemeris.

Coefficients were estimated for in-phase corrections for the 9-year and annual nutation terms and both in-phase and out-of-phase 18-year nutation terms. The sum total nutation corrections, including the linear trend to account for precession, may be written as

$$\begin{aligned} \delta\epsilon_{LLR} &= 8.73\cos l' - 0.97\cos(2\Omega_n) \\ &\quad + 15.13\cos\Omega_n + 6.88\sin\Omega_n \,\mathrm{nrad} \\ \delta\psi_{LLR}\sin\epsilon &= -156.98 - 5.14\,T \\ &\quad + 8.73\sin l' - 2.18\sin(2\Omega_n) \\ &\quad - 16.68\sin\Omega_n + 5.09\cos\Omega_n \ \mathrm{nrad} \end{aligned} \tag{28}$$

where l' is the mean anomaly of the sun, Ω_n is the mean longitude of the ascending lunar node, and T is the time measured from the epoch J2000 in years.

Plate motion was applied to the stations by using the AM0-2 model of Minster and Jordan [22]. This model is based on geological data, and consists of rotation rates for the crustal plates. The model imposes a global condition of "no net rotation" to define absolute site velocities. In the data reduction, the station locations for the epoch 1988.0 were estimated. The resulting estimates are given in Table 1.

V. The VLBI Solution Set

The JPL VLBI software [23] was used to analyze a selected set of NASA's Deep Space Network (DSN) Catalog Maintenance and Enhancement (CM&E) data and Time and Earth Motion Precision Observations (TEMPO) data. CM&E passes are long observation sessions (12–24 hr) used for the determination of radio source positions, while TEMPO passes are shorter sessions (2–4 hr) used to update Earth orientation. With only three station complexes in the DSN (California, Spain, and Australia), measurements are generally only made on one intercontinental baseline at a time. Two problems occur when measurements are made on one baseline only. First, the component of the total rotation $\vec{\Theta}_{VLBI}$ of Eq. (18) that is along the baseline direction is unobservable. The second problem has to do with determining the angle between the Spain–California and California–Australia baselines. If the observations on these baselines are independent, then this angle cannot be determined. By adding a constant bias to x, y, or UT1 for the sessions on the Spain–California baseline, but not to those for the California–Australia baseline, this angle may be arbitrarily changed.

The strategy adopted here to solve these two problems was to work entirely with pairs of back-to-back Spain–California and California–Australia baseline sessions and to constrain the changes in nutation, UT1, and polar motion between the sessions of any pair to physically reasonable levels. All catalog session pairs with fewer than 24 hours of separation were included. TEMPO session pairs with fewer than 24 hours of separation were included only if they coincided with LLR measurements. From these data, a set of epoch 1988 DSN station locations and a series of nutation corrections, UT1 corrections, and polar motion corrections were estimated. Radio source positions were taken from the JPL radio source catalog 1989-5, which agrees to 5 nrad with IERS celestial reference frame

RSC 89 C 01 [6]. No adjustments were made to the source positions.

For each catalog development session, nutation corrections $\delta\psi$ and $\delta\epsilon$, UT1 corrections δUT1, and polar motion corrections δx and δy were estimated. The nutation corrections were relative to the standard IAU 1980 series, while the UT1 and polar motion corrections were relative to an a priori series. Changes in these corrections between sessions in a back-to-back pair were constrained (in a least squares sense) to 5 nrad in nutation, 5 nrad in polar motion and 0.2 ms in UT1, which corresponds to the level of random fluctuations of these parameters over one day [24]. The TEMPO sessions are too short to separate nutation from polar motion. An initial solution that did not include TEMPO sessions showed that the nutation corrections for the catalog sessions were 25 nrad or less. In the final solution, therefore, the nutation offsets for the TEMPO sessions were constrained to be zero with 25-nrad sigmas. In all other aspects the TEMPO sessions were modeled identically to the catalog development sessions.

Epoch 1988.0 locations were estimated for all the DSN stations involved in the observations, with constraints from short baseline experiments applied to intracomplex vectors. The motion of the stations was described by the AM0-2 plate motion model. A priori epoch station locations for DSS 14, DSS 43, and DSS 63 were taken from the IERS station set ITRF88 [6]. In order to specify the coordinate system for the adjusted station location set, a rotation and a translation vector between the a priori and adjusted station coordinates were defined. The rotation vector $\vec{R}$, the translation vector $\vec{T}$, and a scale change D were defined in terms of an unweighted least-squares fit between the a priori station locations $\vec{X}^i$ and the (as yet uncalculated) adjusted station locations $\vec{X}^i + \delta\vec{X}^i$. This fit results from minimizing

$$J = \sum_{i=14,43,63} \left| \vec{T} - \vec{R} \times \vec{X}^i + D\vec{X}^i - \delta\vec{X}^i \right|^2 \qquad (29)$$

Minimization of J with respect to $\vec{T}$, $\vec{R}$, and D resulted in a set of linear equations that gives these fit parameters in terms of the station coordinate adjustments $\delta\vec{X}^i$. The translation $\vec{T}$, and rotation $\vec{R}$, defined in this manner, were constrained to be zero. The scale change D was left unconstrained. The estimated scale change value was $-9 \pm 4 \times 10^{-9}$.

The station location set resulting from this estimation is given in Table 2. The nutation, UT1, and polar motion estimates are presented in Table 3.

VI. Determination of the DSN VLBI–LLR Station Coordinate Transformation

Since the DSN stations and LLR sites are widely separated, it is necessary to use other data sets to compare the two coordinate systems. The CDP has been performing a number of collocations of SLR and VLBI instruments in order to be able to compare and unify terrestrial reference frames. The results of their comparison indicate agreement between the CDP VLBI terrestrial system and the SLR terrestrial system at the 2-cm level for relative station locations [5]. SLR data are sensitive to the locations of stations with respect to the Earth's center of mass while VLBI data are insensitive to the geocenter. The collocations of VLBI and SLR instruments allow the SLR geocenter determination to be applied to the VLBI terrestrial frame. Three of the SLR sites used in the VLBI–SLR collocation study are also the LLR sites listed in Table 1. The CDP VLBI solution, therefore, includes the LLR sites as well as the DSN sites. Thus, one can find the relative orientation of the LLR and DSN VLBI terrestrial frame by comparing them with the CDP VLBI solution.

In fitting the station sets it was assumed that each station set (LLR, DSN VLBI, and CDP VLBI) is internally consistent but expressed in a different coordinate system. By using the least-squares procedure described below, a seven-parameter transformation was estimated to map the LLR, DSN VLBI, and CDP VLBI terrestrial systems to a unified terrestrial frame constrained to agree with the CDP VLBI frame in orientation, scale, and translation. The transformation estimates were based on the station coordinates in the station location sets and on ground tie information.

Since there are only three LLR and three DSN VLBI sites used to estimate a seven-parameter transformation there is more susceptibility to systematic errors than desirable. However, given the good agreement of the SLR and CDP VLBI station sets at the centimeter level and the good agreement in modeling for station locations in the CDP VLBI, SLR, LLR, and DSN VLBI software, the authors do not expect any significant systematic error at the 5- to 10-cm level, which is the accuracy of the LLR station location determination. As more LLR sites become active (recently Wettzell started taking LLR data), it will be possible to strengthen the terrestrial comparison.

The transformations between the station coordinate systems and the unifying coordinate system were assumed to be linear. The transformations included possible offsets of origins and possible rotations. In addition they included possible differences in scale. Scale differences can arise due

to differing treatments of general relativistic corrections. In particular, both the CDP VLBI and DSN VLBI station sets were adjusted to the geocentric metric preferred by IERS [25], while the LLR station locations are expressed with respect to a heliocentric metric [26]. Thus, the LLR terrestrial frame is expected to be different in scale by about 1.5×10^{-8}.

In the fit, the coordinate vector of the ith station in the jth station set (either LLR, DSN VLBI, or CDP VLBI) $\vec{X}^i_j$ is given in terms of $\vec{X}^i_{UCS}$, the coordinate vector in the unified coordinate system, by

$$\vec{X}^i_j = \vec{T}_j + (1 + D_j)\vec{X}^i_{UCS} - \vec{R}_j \times \vec{X}^i_{UCS} + \vec{W}^i_j \quad (30)$$

where $\vec{T}_j$ is the origin offset of coordinate system j, D_j is the scale offset of coordinate system j, and $\vec{R}_j$ is the rotation offset vector of the jth station set coordinate system. $\vec{W}^i_j$ is the measurement noise on the coordinate vector, which was assumed to be independent for each Cartesian component. The unified coordinate system was defined by constraining $\vec{T}_{CDP}$, D_{CDP}, and $\vec{R}_{CDP}$ to be zero.

Ground ties are measurements of the displacement between nearby sites, where "nearby" means that the distance between sites is short enough that the errors caused by differences in orientation and scale between coordinate systems are smaller then measurement errors. In this analysis, ground ties were incorporated as measurements of differences between the coordinates of stations in the unified coordinate system. A tie $\vec{E}^{ik}$ between station i and station k was modeled as

$$\vec{E}^{ik} = \vec{X}^i_{UCS} - \vec{X}^k_{UCS} + \vec{V}^{ik} \quad (31)$$

where $\vec{X}^i_{UCS}$ and $\vec{X}^k_{UCS}$ are the station coordinates in the unified system, and $\vec{V}^{ik}$ is the measurement noise, which was assumed to be independent for each Cartesian component.

The estimation procedure used the ground ties and the Cartesian coordinates for each station location set to solve for the transformation parameters. Only diagonal errors were used for coordinates and ties since full covariances were not available for each station set and ground tie. The station locations and errors for the LLR solution are given in Table 1. The DSN VLBI solution station locations and errors are given in Table 2. The station location and formal errors for the CDP station locations used are given in Table 4. Table 5 shows the ground ties used and their assumed errors. A priori values for transformation parameters for the LLR and DSN VLBI frames were taken as zero with a priori errors set to 100 km for the translation, 1.0 for the scale, and 1 radian for the rotation.

The transformation parameters estimated between the CDP VLBI terrestrial system and the LLR and DSN VLBI terrestrial systems are given in Table 6. Note that the rotations for both LLR and DSN VLBI terrestrial systems are smaller than 10-nrad. This indicates that each terrestrial system is in agreement with the IERS system at the 10-nrad level. The origin of the DSN VLBI system, which was set to agree with the IERS terrestrial reference system ITRF88, is consistent with the CDP VLBI origin, constrained to the ITRF89 terrestrial system, at the few-cm level, which is about the level of uncertainty of the DSN VLBI station location determination. Unlike the VLBI measurements, LLR data are sensitive to the geocenter. The LLR translation offset at the 5- to 10-cm level thus shows agreement of the determination of the geocenter for the LLR and IERS systems at about the LLR uncertainty level.

The overall fit had a χ^2 per degree of freedom of approximately 0.7, indicating that station sets are well fit by the seven-parameter transformation. The fit residuals for the station locations and the ground ties are given in Tables 7 and 8.

The rotation transformation parameters in Table 6 give the rotation from the CDP terrestrial system to either the LLR or DSN VLBI system. For the frame tie one needs the rotation from the LLR terrestrial system to the DSN VLBI system, which is given by

$$\vec{R} = \begin{pmatrix} -\ 11 \pm 22 \\ +\ 4 \pm 16 \\ -\ 3 \pm 7 \end{pmatrix} \text{ nrad} \quad (32)$$

VII. The Planetary-Radio Frame Tie

Comparisons of the LLR and DSN VLBI nutation corrections and UT1 series are presented in this section. In accordance with the analysis in Section III, the planetary-radio frame tie is then synthesized from the derived biases between the LLR and DSN VLBI nutation and UT1 series, and results that have been presented above.

A. Comparison of LLR and DSN VLBI Nutation Corrections

Equation (27) shows that the x and y components of the frame tie rotation vector $\vec{A}$, which correspond to an offset between the planetary ephemeris and radio catalog coordinate poles, are related to the bias between the LLR and DSN VLBI nutation corrections. The DSN VLBI and LLR corrections to the IAU nutation theory are compared in Figs. 1 and 2, with Fig. 1 showing corrections to $\Delta\epsilon$ and Fig. 2 showing corrections to $\Delta\psi \sin\epsilon$. The points with error bars are the DSN VLBI estimates. They come in pairs, one for each session in a back-to-back session pair. The points with large errors in the late 1980s are the TEMPO sessions. The solid lines represent the LLR nutation corrections as given by Eq. (28). For ease of comparison, the best-fit VLBI–LLR offset has been added to each of the LLR curves. The biases were found to be

$$\begin{aligned} \delta\epsilon_{VLBI} - \delta\epsilon_{LLR} &= 5 \pm 10 \text{ nrad} \\ (\delta\psi_{VLBI} - \delta\psi_{LLR})\sin\epsilon &= -49 \pm 10 \text{ nrad} \end{aligned} \tag{33}$$

The fits, which were performed by neglecting the errors in the LLR corrections, resulted in formal bias errors near 1 nrad. Correct treatment of the LLR nutation errors is problematic. Williams, Newhall, and Dickey [27] quote separate uncertainties of 10 nrad for the 18.6-year nutation amplitudes and 1.2 nrad/year for the precession in declination as measured by LLR. However, what matters here is the accuracy of the total nutation correction at epochs within the data span. This total should be better determined than any of its component parts. The errors given in Eq. (33) reflect the authors' estimate of the uncertainty of the LLR corrections.

B. Comparison of LLR and DSN VLBI UT1 Series

As was shown in Section III, the planetary-radio frame tie in right ascension consists of two offsets. The first offset is R_3, the difference in longitude origin of the two terrestrial coordinate systems. The second offset arises from a bias between the VLBI and LLR UT1 series and is considered here.

The DSN VLBI solutions for UT1 and polar motion have already been presented in Table 3. Table 9 presents those LLR measurements that occured near one of the pairs of VLBI session pairs. In some instances no LLR measurements occured near a session pair, and in others several measurements occured. For the LLR reduction, only UT0 was estimated. Table 9 shows the time a date, estimated UT0−UTC and its error, the polar moti angles x_o and y_o assumed in the reduction, the sensitivi S_x and S_y of UT0 to the polar motion angles, and the LL station name.

A corrected value of UT0 for the LLR measuremen to include updated polar motion values, was modeled a

$$\begin{aligned} \mathrm{UT0}^{corr}_{LLR} &= \mathrm{UT0}_{LLR} + [\mathrm{S}_x(x_{LLR} - x_0) \\ &\quad + \mathrm{S}_y(y_{LLR} - y_0)]/\Omega \end{aligned} \tag{3}$$

where $\mathrm{UT0}_{LLR}$ is the value estimated in the LLR reducti by using polar motion values x_0 and y_0, $\mathrm{UT0}^{corr}_{LLR}$ is t value that would result if the LLR reduction had us updated polar motion values x_{LLR} and y_{LLR}, and Ω is t mean rotation rate of the Earth. The updated LLR pol motion angles were obtained from the DSN VLBI analys with corrections given by Eq. (26):

$$x_{LLR} = x_{VLBI} + R_2\,, \quad y_{LLR} = y_{VLBI} + R_1$$

where x_{VLBI} and y_{VLBI} are the polar motion values fro the DSN VLBI fit interpolated to the time of the LL measurement, and R_1 and R_2 are VLBI–LLR terrestri coordinate rotations about the x- and y-axis presented i Section VI.

UT1 is given in terms of UT0 and polar motion by

$$\begin{aligned} \mathrm{UT1}_{LLR} &= \mathrm{UT0}^{corr}_{LLR} - \tan\phi\,[x_{LLR}\sin\lambda \\ &\quad + y_{LLR}\cos\lambda]/\Omega \end{aligned} \tag{35}$$

where ϕ and λ are the latitude and longitude of the LLI station.

Using all DSN VLBI session pairs with coincident LLI measurements, a least-squares fit was performed to esti mate the bias, ΔUT1, between DSN VLBI and the LLI UT1, as defined by

$$\mathrm{UT1}_{LLR} = \mathrm{UT1}_{VLBI} - \Delta\mathrm{UT1} \tag{36}$$

where $\mathrm{UT1}_{VLBI}$ is the DSN VLBI UT1 determination in terpolated to the time of the LLR measurement.

The least-squares fit using the LLR measurements for all back-to-back DSN VLBI session pairs incorporated the full six-by-six information matrix for the two sets of VLBI UT/PM estimates in each back-to-back session pair. The offset ΔUT1 was estimated, and the sensitivity of this estimate to the rotations $R1$ and $R2$ was calculated. The resulting offset measurements are shown in Fig. 3. The weighted mean offset was found to be

$$\mathrm{UT1}_{VLBI} - \mathrm{UT1}_{LLR} = 0.22 \pm 0.12 \text{ msec} \qquad (37)$$

where the error given here includes the considered effect of the errors in the rotations R_1 and R_2.

C. The Frame Tie

By returning to Eq. (27) and substituting the terrestrial rotation, nutation, and UT1 biases given above, the following estimate of the frame tie from the ephemeris DE200 to the IERS celestial reference frame results:

$$\vec{A} = \begin{pmatrix} +\ 5 \pm 15 \\ -\ 49 \pm 15 \\ -\ 19 \pm 25 \end{pmatrix} \text{ nrad} \qquad (38)$$

The errors from the comparison process include the nutation bias uncertainties (10 nrad), the UT bias uncertainty (9 nrad), the R_3 uncertainty (7 nrad), and the uncertainty in the alignment of the radio source catalog with the IERS celestial reference frame (5 nrad in each component). The errors given in Eq. (38) include estimates of systematic errors. Since many steps in the comparison procedure rely on separate data reductions, each with correlated parameters, to derive formal errors would require combining full covariance matrices, not all of which are available. The largest potential source of systematic error is thought to lie in the comparison of the UT1 series where the number of points included is small for a measurement known to be fairly noisy. In future work, more comparison points can be obtained by including a more extensive VLBI data set. Other improvements could come from combining the LLR and VLBI information matrices [28], along with the full covariance for the CDP station determination, to allow an estimate of the errors including correlations of parameters. However, the presented frame tie result is apparently the most global and accurate determination yet available.

The frame tie result given in Eq. (38) represents the rotation (and uncertainty) between the IERS radio source frame and the reference frame determined by the tabulated orbit of the Earth within the ephemeris DE200. The orbit of the Earth is tabulated with respect to a projected dynamical equator and equinox for the year 2000 [20]. There is significant uncertainty in the determination of the equinox of 2000 since it depends upon predictions using estimated precession and nutation constants, which are quantites that (aside from data reduction) do not affect the orbits of the planets. The authors believe that the physical content of the ephemeris can be referred to the orbit of the Earth for the definition of the reference system. The orbits of the other planets do not define different reference systems but can instead be referred to the orbit of the Earth with an uncertainty characteristic of the internal consistency of the ephemeris.

VIII. Comparison With Other Frame Tie Determinations

Other methods of determining the radio-planetary frame tie include VLBI observation of spacecraft at other planets and comparison of positions of millisecond pulsars based on VLBI and timing measurements. The result presented above in Eq. (38) can be compared with other results by examining the offset in right ascension and declination in the part of the sky where the other measurements exist.

There have been a number of VLBI observations of spacecraft at other planets. A planetary orbiter, or a spacecraft making a planetary encounter, has a position determined with respect to the planet from the gravitational signature on the spacecraft Doppler data. VLBI measurements between the spacecraft and one or more angularly nearby radio sources can be used to estimate the radio source coordinates in the planetary reference frame. Newhall, Preston, and Esposito [29] reported average right ascension and declination offsets consistent with zero, with uncertainty of 40–60 nrad, based on the the results of VLBI measurements for the Viking and Pioneer Venus orbiters. McElreath and Bhat [30] derived a position of the radio source P 0202 + 14 in the planetary ephemeris frame from observations of the Soviet Vega 1 and Vega 2 spacecraft as they flew by Venus in 1985. The Vega measurements resulted in right ascension and declination offsets consistent with Eq. (38) within their errors of 50 nrad in each component. In 1989, VLBI observations of the Soviet Phobos spacecraft at Mars were made for frame tie determination in nearly the same part of the sky as the Vega observations. Preliminary results of the Phobos data[4] are consis-

[4] B. A. Iijima and C. E. Hildebrand, personal communication, Jet Propulsion Laboratory, Pasadena, California, 1991.

tent with the results in Eq. (38) at the 1-sigma level. An observation sequence for the Magellan spacecraft at Venus is being pursued to extend the data set of VLBI spacecraft observations.

Timing of millisecond pulsars gives positions with few-nrad accuracy based on the orbit of the Earth [31]. VLBI observations of these sources are difficult since the pulsars are weak radio sources. Two groups[5,6] have made VLBI observations of the pulsar PSR 1937+21. Preliminary results from one group[6] give right ascension and declination offsets in agreement with Eq. (38) within their errors.

In the future more spacecraft VLBI measurements and refinements of the technique presented here, as well as results from other methods, should combine to produce a consistent frame tie determination at the 5-nrad level. In the meantime, the radio-planetary frame tie result presented here with 15–25 nrad accuracy is a useful reference point.

IX. Application to Interplanetary Spacecraft Navigation

In order to apply this (or any other) frame tie result to spacecraft navigation with full accuracy, it will be necessary to have well-defined standards for reference frame definition. Guidelines for standards and implementation are proposed below.

For orbit determination, UT and polar motion are usually read in from an external service and not adjusted (or estimated) in the navigation process. The importance of this external UTPM information in defining the reference frame for the spacecraft is often overlooked. The adoption of a station set and an Earth orientation series essentially defines a celestial reference frame that may be different from the desired reference frame (often the planetary ephemeris reference frame). This inconsistency is most important for interplanetary missions tracked mainly by Doppler. In this case, the signature of the Earth on the Doppler data is tied most strongly to a celestial reference frame that is defined by the station set and Earth orientation series. When there are data directly sensitive to the planetary ephemeris, such as range (which is sensitive to the orbit of the Earth) or onboard optical data, there can be a systematic discrepancy among the various data types which will be resolved by the amount and weighting of the data.

It is preferable to clearly define the reference frames in use rather than having them defined implicitly and/or inconsistently. The most practical choice would be to use the IERS definitions for station locations, Earth orientation, and quasar locations while allowing the planetary ephemeris to define its own reference system. Each ephemeris would be related to the standard celestial coordinate system by three rotation angles, such as those given in Eq. (38). This choice of standards would simplify matters by minimizing the number of parameters that would vary from mission to mission or from ephemeris to ephemeris. Each mission could use a standard station set, Earth orientation series, and quasar catalog, regardless of the ephemeris used. A priori values for the three rotation angles could be adopted from an external determination (such as reported in this work). If it is desired and the necessary partials exist, corrections to the frame tie could be estimated for a particular orbit determination analysis.

This work has utilized the high accuracy of terrestrial station locations in determining the frame tie. Geocentric station locations have been determined with accuracy better than 10 cm in all components. Each aspect of the station location determination has been checked by independent data sets and reduction software. VLBI and SLR each produce relative station locations with 2- to 3-cm accuracy. SLR and LLR independently determine the geocenter to 10-cm accuracy or better. The time dependence of the DSN station locations (i.e., plate motion) and the rotation of the station locations into inertial space are also well known. There is no reason that interplanetary spacecraft navigation cannot take advantage of these high-accuracy station locations.

X. Conclusion

A determination of the rotational offset between the planetary ephemeris DE200 and the radio reference frame has been presented based on a comparison of VLBI and LLR Earth orientation measurements. The accuracy of the frame tie is about 25 nrad. The frame tie result is substantiated by comparison with determinations from other techniques. The frame tie result is made possible by the ability to determine the location of DSN tracking stations with accuracy better than 10 cm. The frame tie result, combined with proper use of accurate station locations, will enable more accurate interplanetary spacecraft navigation.

[5] N. Bartel, personal communication, Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusettes, April 9, 1991.

[6] R. J. Dewey and D. L. Jones, "Millisecond Pulsar Frame Tie," JPL IOM 335-6-91-006, (internal document) Jet Propulsion Laboratory, Pasadena, California, April 17, 1991.

Acknowledgments

The authors would like to thank Skip Newhall, Allan Steppe, Ojars Sovers, and Jim Williams for many hours of useful discussion. They also would like to acknowledge the careful and very useful review of this work provided by Roger Linfield. Jim Ray graciously provided results on terrestrial frame comparisons at many stages of this work.

References

[1] J. S. Border, F. F. Donivan, S. G. Finley, C. E. Hildebrand, B. Moultrie, and L. J. Skjerve, "Determining Spacecraft Angular Position with Delta VLBI: The Voyager Demonstration," paper AIAA-82-1471, presented at the AIAA/AAS Astrodynamics Conference, San Diego, California, August 9–11, 1982.

[2] O. J. Sovers, "JPL 1990-3: A 5-nrad Extragalactic Source Catalog Based on Combined Radio Interferometric Observations," *TDA Progress Report 42-106*, vol. April–June 1991, Jet Propulsion Laboratory, Pasadena, California, pp. 364–383, August 15, 1991.

[3] J. G. Williams and E. M. Standish, "Dynamical Reference Frames in the Planetary and Earth-Moon Systems," in *Reference Frames in Astronomy and Geophysics*, edited by J. Kovalevsky et al., Boston: Kluwer Academic Publishers, pp. 67–90, 1989.

[4] C. S. Christensen, S. W. Thurman, J. M. Davidson, M. H. Finger, and W. M. Folkner, "High-Precision Radiometric Tracking for Planetary Approach and Encounter in the Inner Solar System," *TDA Progress Report 42-97*, vol. January–March 1989, Jet Propulsion Laboratory, Pasadena, California, pp. 21–46, May 15, 1990.

[5] J. R. Ray, C. Ma, T. A. Clark, J. W. Ryan, R. J. Eanes, M. M. Watkins, B. E. Schutz, and B. D. Tapley, "Comparison of VLBI and SLR Geocentric Site Coordinates," *Geophys. Res. Let.*, vol. 18, pp. 231–234, 1991.

[6] *International Earth Rotation Service, Annual Report for 1988*, Observatoire de Paris, France, 1989.

[7] Z. Altamimi, E. F. Arias, C. Boucher, and M. Feissel, "Earth Orientation Determinations: Some Tests of Consistency," paper presented at the 105th Symposium on Earth Rotation and Coordinate Reference Frames, Edinburgh, Scotland, August 10–11, 1989.

[8] C. Boucher and Z. Altamimi, "The Initial IERS Terrestrial Reference Frame," IERS Technical Note 1, Observatoire de Paris, France, 1989.

[9] P. K. Seidelmann, "1980 IAU Theory of Nutation: The Final Report of the IAU Working Group on Nutation," *Celestial Mechanics*, vol. 27, pp. 79–106, 1982.

[10] N. Capitaine, J. G. Williams, and P. K. Seidelmann, "Clarifications Concerning the Definition of the Celestial Ephemeris Pole," *Astron. Astrophys.*, vol. 146, pp. 381–383, 1985.

[11] S. Aoki, B. Guinot, G. H. Kaplan, H. Kinoshita, D. D. McCarthy, and P. K. Seidelmann, "The New Definition of Universal Time," *Astron. Astrophys.*, vol. 105, pp. 359–361, 1982.

[12] J. M. Wahr, "The Forced Nutations of an Elliptical, Rotating, Elastic and Oceanless Earth," *Geophys. J. R. Astr. Soc.*, vol. 64, pp. 705–727, 1981.

[13] J. H. Lieske, T. Lederle, W. Fricke, B. Morando, "Expressions for the Precession Quantities Based upon the IAU (1976) System of Astronomical Constants," *Astron. Astrophys.*, vol. 58, pp. 1–16, 1977.

[14] J. H. Lieske, "Precession Matrix Based on IAU (1976) System of Astronomical Constants," *Astron. Astrophys.*, vol. 73, pp. 283–284, 1979.

[15] E. Fabri,"Advocating the Use of Vector-Matrix Notation in Precession Theory," *Astron. Astrophys.*, vol. 82, pp. 123–128, 1980.

[16] S. Y. Zhu and I. I. Mueller, "Effects of Adopting New Precession, Nutation and Equinox Corrections on the Terrestrial Reference Frame," *Bulletin Geodesique*, vol. 57, pp. 29–42, 1983.

[17] J. G. Williams and W. G. Melbourne, "Comments on the Effect of Adopting New Precession and Equinox Corrections," in *High-Precision Earth Rotation and Earth-Moon Dynamics*, edited by O. Calame, Boston: D. Reidel, pp. 293–303, 1982.

[18] N. Capitaine, B. Guinot, and J. Souchay, "A Non-Rotating Origin on the Instantaneous Equator: Definition, Properties, and Use," *Celestial Mechanics*, vol. 39, pp. 283–307, 1986.

[19] X X Newhall, J. G. Williams, and J. O. Dickey, "Earth Rotation (UT0) From Lunar Laser Ranging," IERS Technical Note 5, Observatoire de Paris, France, pp. 41–44, 1990.

[20] E. M. Standish, Jr., "Orientation of the JPL Ephemerides, DE200/LE200, to the Dynamical Equinox of J2000," *Astron. Astrophys.*, vol. 114, pp. 297–302, 1982.

[21] E. M. Standish, Jr., "The Observational Basis for JPL's DE200, the Planetary Ephemeris of the Astronomical Almanac," *Astron. Astrophys.*, vol. 233, pp. 252–271, 1990.

[22] J. B. Minster and T. H. Jordan, "Present Day Plate Motions," *J. Geophys. Res.*, vol. 83, pp. 5331–5354, 1978.

[23] O. J. Sovers, *Observation Model and Parameter Partials for the JPL VLBI Parameter Estimation Software MODEST—1991*, JPL Publication 83-39, Rev. 4, Jet Propulsion Laboratory, Pasadena, California, August 1, 1991.

[24] D. D. Morabito, T. M. Eubanks, and J. A. Steppe, "Kalman Filtering of Earth Orientation Changes," in *Earth's Rotation and Reference Frames for Geodesy and Geodynamics*, edited by A. H. Babcock and G. A. Wilkens, Dordreecht, The Netherlands: D. Reidel, pp. 257–267, 1988.

[25] D. D. McCarthy, "IERS Standards (1989)," IERS Technical Note 3, Observatoire de Paris, France, 1989.

[26] R. W. Hellings, "Relativistic Effects in Astronomical Timing Measurements," *Astronomical J.*, vol. 91, pp. 650–659, 1986.

[27] J. G. Williams, X X Newhall, and J. O. Dickey, "Luni-Solar Precession: Determination from Lunar Laser Ranges," *Astron. Astrophys.*, vol. 241, pp. L9–L12, 1991.

[28] P. Charlot, O. J. Sovers, J. G. Williams, and X X Newhall, "A Global VLBI/LLR Analysis for the Determination of Precession and Nutation Constants," in *Reference Systems: Proceedings of the 127th Colloquium of the International Astronomical Union*, edited by J. A. Hughes, C. A. Smith, and G. H. Kaplan, Washington, DC: U.S. Naval Observatory, pp. 228–233, 1991.

[29] X X Newhall, R. A. Preston, and P. B. Esposito, "Relating the JPL VLBI Reference Frame and the Planetary Ephemerides," *Proceedings of the 109th Symposium of the IAU on Astrometric Techniques*, edited by H. K. Eichorn and R. J. Leacock, Boston: Reidel, pp. 789–794, 1986.

[30] T. P. McElrath and R. S. Bhat, "Determination of the Inner Planet Frame Tie Using VLBI Data," paper 88-4234, presented at the AIAA/AAS Astrodynamics Conference, Minneapolis, Minnesota, 1988.

[31] L. A. Rawley, J. H. Taylor, and M. M. Davis, "Fundamental Astrometry and Millisecond Pulsars," *Astrophys. J.*, vol. 326, pp. 947–953, 1988.

Table 1. Lunar Laser Ranging station coordinates.

Station	No.	x, m	y, m	z, m	σ_x, m	σ_y, m	σ_z, m
McDonald 107-in.	7206	−1330781.1660	−5328755.6310	3235697.6320	0.0300	0.0300	0.1000
McDonald MLRS[a]	0108	−1330120.9160	−5328532.2060	3236146.6410	0.0300	0.0300	0.1000
Haleakala	0210	−5466006.9080	−2404428.1360	2242188.5400	0.0300	0.0300	0.1000
Grasse	7845	4581692.2540	556195.8208	4389354.8430	0.0300	0.0300	0.1000

[a] MLRS = McDonald Laser Ranging System.

Table 2. DSN VLBI station coordinates.

Station	No.	x, m	y, m	z, m	σ_x, m	σ_y, m	σ_z, m
DSS 14	1514	−2353621.0830	−4641341.5930	3677052.3000	0.0312	0.0318	0.0315
DSS 43	1543	−4460894.4630	2682361.6260	−3674748.7600	0.0341	0.0311	0.0317
DSS 63	1563	4849092.7130	−360180.6860	4115108.9730	0.0324	0.0320	0.0338

Table 3. VLBI nutation, polar motion, and UT1 estimates.

Mean date	Time	$\Delta\epsilon$, mas	$\sigma(\Delta\epsilon)$, mas	$\Delta\psi$, mas	$\sigma(\Delta\psi)$, mas	UT1−UTC, ms	σ(UT1−UTC) ms	x mas	$\sigma(x)$, mas	y, mas	$\sigma(y)$, mas
Dec. 20, 1979	13:39:24	0.64	0.48	2.40	1.56	−324.83	0.16	150.16	2.23	273.76	0.77
Dec. 21, 1979	16:56:57	0.50	1.05	1.55	2.85	−327.65	0.17	150.18	2.22	271.91	1.20
Jan. 25, 1980	23:37:29	0.87	2.32	2.46	3.44	583.53	0.15	123.28	2.55	214.17	2.69
Jan. 27, 1980	14:44:25	1.01	2.47	1.52	3.42	579.54	0.18	120.90	2.56	211.90	2.52
Feb. 14, 1980	05:02:33	−1.36	0.96	1.86	2.57	539.85	0.09	83.77	1.61	189.25	1.50
Feb. 14, 1980	22:04:27	−1.34	1.16	1.59	2.71	538.03	0.15	82.80	1.71	188.50	1.23
Feb. 23, 1980	18:32:42	0.52	0.56	4.32	1.94	514.35	0.15	73.39	2.16	184.22	0.78
Feb. 24, 1980	12:47:20	0.95	1.09	5.24	3.09	512.45	0.12	72.80	2.14	183.70	1.20
Dec. 8, 1981	16:03:38	2.91	1.26	10.70	5.28	69.45	0.15	−110.76	2.15	324.21	2.01
Dec. 9, 1981	09:45:29	2.51	1.03	10.33	5.01	67.72	0.17	−110.51	2.18	325.74	1.80
May 20, 1983	20:10:50	0.03	0.24	−0.29	0.55	−164.44	0.10	143.35	1.19	542.34	0.42
May 22, 1983	09:05:59	0.08	0.54	−3.42	1.54	−168.12	0.06	150.14	1.00	539.78	0.96
Nov. 18, 1983	10:55:03	2.79	0.49	−7.94	1.71	489.12	0.14	19.05	1.28	19.99	0.81
Nov. 19, 1983	10:09:52	2.56	0.51	−6.53	1.49	486.73	0.09	14.88	1.09	20.67	1.18
Dec. 17, 1983	11:12:20	4.67	0.63	−1.79	1.38	427.56	0.09	−89.02	1.10	59.67	1.15
Dec. 18, 1983	10:00:46	5.33	0.55	−5.61	1.39	425.85	0.13	−92.03	1.27	61.98	0.76
Mar. 24, 1984	22:55:12	0.50	0.56	−0.86	1.29	255.08	0.08	−225.02	1.04	389.58	1.07
Mar. 25, 1984	22:01:27	0.82	0.25	−1.16	0.64	253.06	0.11	−223.82	1.21	393.11	0.63
May 12, 1984	12:02:11	−1.19	0.74	0.02	1.72	161.00	0.10	−84.21	1.34	540.76	1.60
May 13, 1984	04:09:50	−1.00	1.13	−0.49	2.88	159.83	0.18	−81.94	1.52	541.95	1.37
Jul. 14, 1984	14:44:35	0.48	0.83	−7.05	1.55	90.22	0.10	175.04	1.19	520.60	1.24
Jul. 15, 1984	14:23:34	0.70	0.91	−8.32	1.58	89.50	0.13	179.03	1.34	518.19	0.89
Sep. 28, 1985	18:08:45	−0.04	0.78	−14.46	2.03	472.50	0.09	213.43	1.30	411.71	1.06
Sep. 29, 1985	20:52:30	0.61	0.55	−13.42	1.30	470.64	0.13	214.63	1.43	408.48	0.60
Mar. 22, 1986	23:46:10	4.09	3.35	−8.31	8.91	200.13	0.18	−16.50	4.73	125.40	3.62
Mar. 23, 1986	07:44:37	4.23	3.36	−8.77	9.00	199.71	0.22	−17.15	4.70	125.81	3.52
Apr. 19, 1986	12:47:01	1.86	3.74	−2.03	9.58	155.94	0.23	−69.80	4.69	173.06	4.26
Apr. 20, 1986	12:31:53	2.01	3.86	−1.91	9.78	154.45	0.21	−71.56	4.69	174.99	4.36
Jun. 28, 1986	19:00:33	1.53	1.04	−4.07	2.74	75.10	0.12	−70.61	1.96	319.24	1.16
Jun. 29, 1986	20:51:35	1.69	0.52	−4.15	1.49	74.39	0.14	−69.09	1.97	321.68	0.72
Apr. 18, 1987	22:17:11	3.61	3.45	−8.79	10.87	−305.34	0.25	74.90	3.73	206.40	5.54
Apr. 19, 1987	15:30:21	3.52	3.50	−9.08	11.15	−306.40	0.27	73.98	3.77	206.10	5.46
May 9, 1987	09:49:51	−0.61	0.64	−2.26	1.31	−336.87	0.12	54.37	1.26	203.11	0.82
May 10, 1987	19:22:27	−0.77	0.52	−2.81	1.22	−338.75	0.07	52.68	1.06	202.90	1.18
Jan. 9, 1988	10:59:49	3.36	1.62	−8.73	6.99	353.65	0.20	−5.78	3.61	423.26	2.19
Jan. 10, 1988	02:56:59	3.48	1.71	−9.15	7.12	352.66	0.12	−4.69	3.65	423.58	2.31
Oct. 1, 1988	18:10:17	−0.03	2.81	−16.21	7.11	24.15	0.11	9.17	4.63	131.47	2.45
Oct. 1, 1988	22:20:34	0.04	2.93	−15.88	7.14	24.03	0.18	8.41	4.60	131.58	2.33

Table 4. CDP VLBI station coordinates.

Station	No.	x, m	y, m	z, m	σ_x, m	σ_y, m	σ_z, m
DSS 13	1513	−2351128.9948	−4655477.0834	3660956.8432	0.0027	0.0053	0.0047
DSS 45	1645	−4460935.0547	2682765.7891	−3674381.6368	0.0195	0.0130	0.0117
Haleakala	7120	−5465998.3924	−2404408.5665	2242228.4099	0.0142	0.0074	0.0071
DSS 65	1665	4849336.7823	−360488.9205	4114748.5369	0.0125	0.0049	0.0132
Grasse	7605	4581697.8187	556125.6265	4389351.2470	0.0067	0.0026	0.0075
DSS 15	1615	−2353538.6234	−4641649.5275	3676669.9431	0.0044	0.0088	0.0075
McDonald	7850	−1330008.0136	−5328391.5430	3236502.6372	0.0028	0.0102	0.0063

Table 5. Ground tie vectors.

Site	From	To	x, m	y, m	z, m	σ_x, m	σ_y, m	σ_z, m
McDonald[a]	7206	7086	655.9005	229.0262	452.6222	0.0100	0.0100	0.0100
McDonald[b]	7086	0108	4.3560	−5.6310	−3.5570	0.0100	0.0100	0.0100
McDonald[a]	7086	7850	117.2017	135.0619	352.4838	0.0100	0.0100	0.0100
Haleakala[c]	7210	0210	−0.4830	−0.2120	1.0030	0.0100	0.0100	0.0100
Haleakala[a]	7120	7210	−8.0140	−19.4100	−40.9270	0.0100	0.0100	0.0100
Grasse[a]	7835	7845	0.5780	36.4730	−4.4390	0.0100	0.0100	0.0100
Grasse[d]	7605	7835	−6.0150	33.7350	8.1020	0.0100	0.0100	0.0100
Goldstone[e]	1513	1514	−2492.0800	14135.5350	16095.4150	0.0100	0.0100	0.0100
Goldstone[e]	1513	1615	−2409.6220	13827.5670	15713.0940	0.0200	0.0200	0.0210
Madrid[e]	1665	1563	−244.1140	308.2930	360.3200	0.0200	0.0200	0.0200
Canberra[e]	1645	1543	40.6580	−404.1520	−367.1850	0.0100	0.0100	0.0100

[a] Crustal Dynamics Project Site Catalog, Goddard Space Flight Center, Greenbelt, Maryland, May, 1989.

[b] T. M. Sager and J. L. Long, internal memorandum, Bendix Aerospace Corporation, Columbia, Maryland, April 11, 1985.

[c] L. S. Baker, internal memorandum, National Geodetic Survey, Rockville, Maryland, October 24, 1975.

[d] C. Boucher, personal communication, Institut Geophysique, St. Mande, France, April 1990.

[e] C. S. Jacobs, personal communication, Jet Propulsion Laboratory, Pasadena, California, December 1989.

Table 6. Transformation parameters from CDP VLBI coordinate system to DSN or LLR system.

System	T_1,cm	T_2,cm	T_3,cm	$D(10^{-9})$	R_1, nrad	R_2, nrad	R_3, nrad
DSN	−0.7±2.3	−2.4±2.3	6.6±2.3	5.5±3.5	−5.2± 4.9	−0.5± 4.6	−1.5±5.0
LLR	−8.0±6.2	−3.2±7.7	4.3±8.0	−9.8±4.4	6.0±21.5	−4.1±15.0	1.8±5.3

Table 7. Input coordinate residuals.

Site	System	No.	x, mm	y, mm	z, mm
Goldstone	DSN	1514	3.2	25.1	−18.9
Canberra	DSN	1543	−20.2	−11.6	0.4
Madrid	DSN	1563	14.7	−13.1	21.3
McDonald	LLR	7206	10.3	−30.5	51.1
McDonald	LLR	0108	4.5	−3.7	−4.5
Haleakala	LLR	0210	2.2	29.5	−3.2
Grasse	LLR	7845	−17.1	4.7	−43.4
DSS 13	CDP	1513	0.1	0.0	0.2
DSS 45	CDP	1645	6.6	2.0	−0.1
Haleakala	CDP	7120	−0.5	−1.8	0.0
DSS 65	CDP	1665	−2.2	0.3	−3.3
Grasse	CDP	7605	0.8	0.0	0.2
DSS 15	CDP	1615	−0.3	−1.8	0.7
McDonald	CDP	7850	−0.1	4.0	−0.2

Table 8. Ground tie residuals.

Site	From	To	x, mm	y, mm	z, mm
McDonald	7206	7086	1.1	−3.4	0.5
McDonald	7086	0108	−0.5	0.4	0.0
McDonald	7086	7850	1.6	−3.8	0.5
Haleakala	7210	0210	−0.2	−3.3	0.0
Haleakala	7120	7210	−0.2	−3.3	0.0
Grasse	7835	7845	1.9	−0.5	0.4
Grasse	7605	7835	1.9	−0.5	0.4
Goldstone	1513	1514	−0.3	−2.5	1.9
Goldstone	1513	1615	6.2	9.3	−5.3
Madrid	1665	1663	−5.6	5.1	−7.5
Canberra	1645	1543	1.7	1.2	0.0

Table 9. Lunar Laser Ranging estimated UT0.

Date	Time	UT0–UTC ms	UT0–UTC σ, ms	x, mas	y, mas	S_x, mas/mas	S_y, mas/mas	Site
Jan. 26, 1980	03:16:19	573.695	0.304	114.042	213.383	−0.03837	0.15363	McDonald
Jan. 27, 1980	01:00:23	571.010	0.589	112.599	212.143	0.04666	−0.18683	McDonald
Jan. 28, 1980	04:02:20	569.085	0.449	110.804	210.599	0.00398	−0.01595	McDonald
Feb. 24, 1980	02:54:08	506.952	0.469	67.877	181.884	−0.03957	0.15844	McDonald
Feb. 25, 1980	02:28:28	503.540	0.350	66.969	181.103	0.00981	−0.03928	McDonald
Dec. 8, 1981	05:42:44	71.082	0.424	−115.685	324.240	−0.07837	0.31382	McDonald
Dec. 9, 1981	06:09:18	66.559	0.457	−115.550	326.493	−0.05714	0.22880	McDonald
Dec. 10, 1981	04:41:58	65.207	0.266	−115.426	328.571	−0.00842	0.03371	McDonald
Nov. 17, 1983	04:12:40	489.83	0.367	20.977	16.719	0.00295	−0.01182	McDonald
Dec. 17, 1983	05:01:36	430.525	0.598	−92.780	57.207	−0.00951	0.03808	McDonald
May 11, 1984	23:23:32	193.926	0.355	−88.808	536.713	0.62278	−0.07560	Grasse
Jul. 16, 1984	08:37:45	78.004	0.698	178.951	513.587	0.01670	−0.06688	McDonald
Mar. 22, 1986	02:13:30	200.782	0.221	−19.178	124.652	0.07114	−0.28499	MLRS[a]
Mar. 23, 1986	03:01:29	199.584	0.117	−21.201	125.789	0.04375	−0.17528	MLRS
Apr. 18, 1986	20:10:57	167.546	0.151	−71.066	170.577	0.27160	−0.03297	Grasse
Apr. 19, 1986	03:47:27	157.666	0.132	−71.532	171.207	−0.03970	0.15902	MLRS
Apr. 19, 1986	20:58:15	165.807	0.230	−72.629	172.553	0.26490	−0.03216	Grasse
Jun. 29, 1986	14:19:11	65.176	0.253	−74.057	319.475	0.26536	−0.11673	Haleakala
Apr. 19, 1987	14:36:20	313.090	0.385	74.779	204.107	0.10398	−0.04574	Haleakala
Apr. 20, 1987	15:02:45	313.976	0.283	73.923	203.901	0.14778	−0.06501	Haleakala
May 8, 1987	03:57:04	339.732	0.148	55.129	200.679	−0.04213	0.16879	MLRS
May 8, 1987	08:39:43	341.488	0.094	54.920	200.634	−0.13585	0.05976	Haleakala
May 9, 1987	04:27:15	341.479	0.084	53.998	200.431	−0.04549	0.18223	MLRS
May 9, 1987	09:11:16	343.184	0.286	53.742	200.368	−0.19244	0.08465	Haleakala
May 10, 1987	04:54:11	343.459	0.101	52.672	200.107	−0.02686	0.10761	MLRS
Jan. 9, 1988	14:34:02	342.844	0.088	−7.259	420.476	0.01294	−0.00569	Haleakala
Oct. 1, 1988	13:02:10	19.006	0.121	9.764	129.043	0.26353	−0.11592	Haleakala
Oct. 2, 1988	13:55:35	18.314	0.093	6.340	129.105	0.27412	−0.12058	Haleakala

[a] MLRS = McDonald Laser Ranging System.

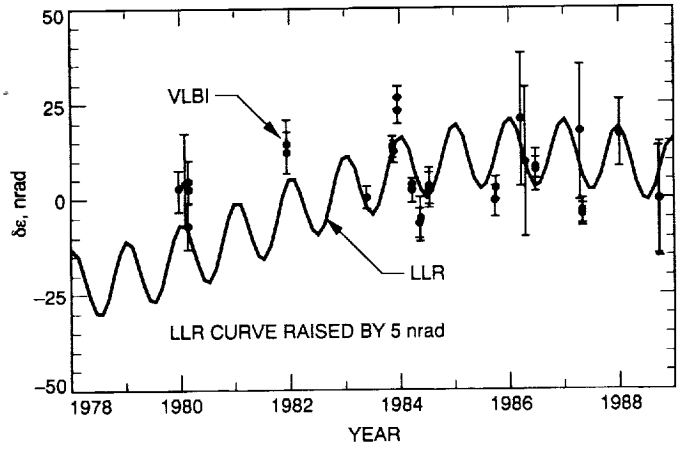


Fig. 1. Comparison of the nutation correction $\delta\epsilon$ for VLBI and LLR.

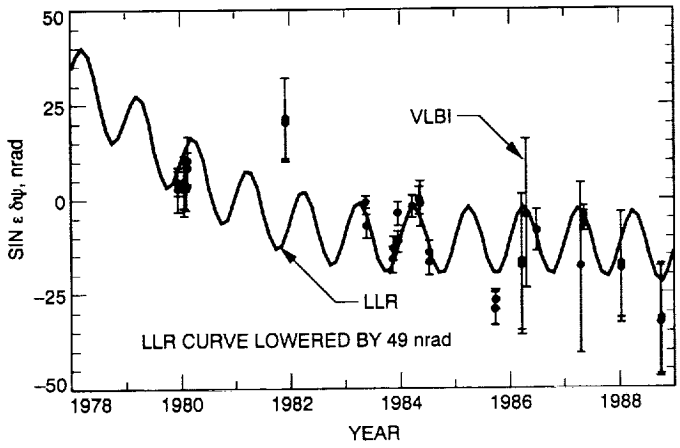


Fig. 2. Comparison of nutation correction $\delta\psi$ sin ϵ for VLBI and LLR.

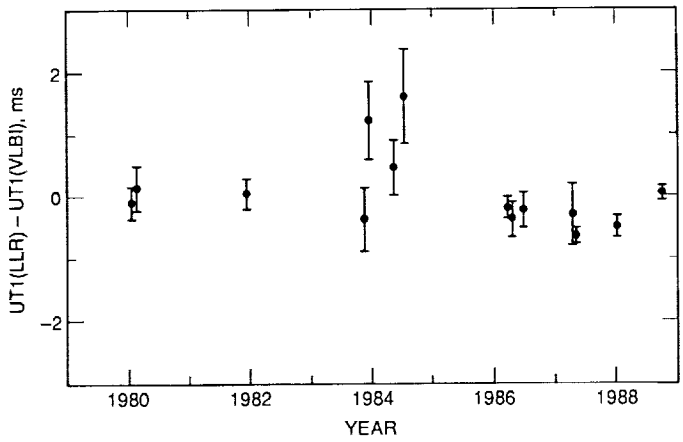


Fig. 3. Offset between the VLBI and LLR UT1 series.

Occultation of a Compact Radio Source by Venus

R. Linfield
Tracking Systems and Applications Section

An occultation of the compact radio source P 0507+17 by Venus on July 19, 1988, was observed in Tidbinbilla, Australia at a frequency of 2.3 GHz. The purpose of this observation was to measure the position of Venus in the radio reference frame. When data from both ingress (Venus dayside) and egress (Venus nightside) were used to solve for the position of Venus in ecliptic longitude and latitude, the results were consistent with zero offsets from the nominal values, with an uncertainty of approximately 0.2 arcsec in both coordinates. By using the nightside data alone, a value of -0.026 ± 0.04 arcsec was obtained for the linear combination $\Delta\lambda + 0.51\Delta\beta$, where $\Delta\lambda$ and $\Delta\beta$ were the offsets from their nominal values of the ecliptic longitude and latitude of Venus.

Distortion of a vacuum Fresnel fringe pattern by the Venus troposphere, and especially by the Venus ionosphere, was observed. The dayside ionosphere of Venus caused very large distortions; the amplitude of the first Fresnel fringe in the ingress data was eight times larger than had been expected for an airless planet. The observed fringe patterns were modeled by using plausible ionospheres (i.e., consistent with spacecraft measurements of the Venus ionosphere and with solar extreme ultraviolet flux and solar wind pressure measurements at the occultation epoch). However, the range of Venus ionospheric profiles (electron density as a function of altitude) allowed by a priori constraints and by the occultation data was large (e.g., the ionopause height on the dayside was uncertain by a factor of two). This ionospheric uncertainty (particularly on the dayside) translated into a large position uncertainty (0.2 arcsec for the dayside and 0.04 arcsec for the nightside). If it had been possible to calibrate the Venus ionosphere by some external means, the accuracy in $\Delta\lambda$ and $\Delta\beta$ would have been 0.01 arcsec or better.

I. Introduction

Very long baseline interferometry (VLBI) observations of compact extragalactic radio sources have established a stable, inertial reference frame (the radio reference frame), with source positions known to approximately 1 mas [1,2]. The planetary reference frame is defined by the orbits of the Moon and the planets. These orbits are known to a variable accuracy (relative to the orbit of the Earth), ranging from 5 mas for Mars to 500 mas for Pluto [3]. The orientation of the orbit of the Earth relative to the radio reference frame is known to 5–10 mas [4]; the positions of

the planets in the radio frame will have errors at least that large. Furthermore, due to modeling uncertainties, such as unknown asteroid masses [5], the planetary reference frame has an unknown net rotation rate. The errors of the planetary orbits in the radio frame are therefore time-variable.

Accurate measurements of the positions of planets relative to compact radio sources would allow the planetary orbits to be calculated in the stable radio reference frame. The orbits determined in that frame would allow improved studies of solar system dynamics. In addition, improved knowledge of the positions of the planets in the radio frame would lead to better navigation of planetary spacecraft. VLBI observations can locate the position of a spacecraft on the sky in the radio frame [6]. The current accuracy of such measurements is in the 1- to 10-mas range. For most navigation purposes, the position of a spacecraft relative to a planet or natural satellite is needed. Determining the relationship between the planetary and radio reference frames would have immediate benefits for spacecraft navigation.

The epoch and duration of an occultation of a compact radio source by a solar system object are sensitive to the location of that object in the radio frame. The epoch of the midpoint of occultation (specifically, the mean of the epoch of geometric ingress and egress) is the time when that solar system object and the radio source have the same position along the direction of motion of the object (approximately the ecliptic longitude). The duration of the occultation determines their relative position perpendicular to the direction of motion (i.e., approximately the ecliptic latitude).

The Moon sweeps out a solid angle on the sky, which grows at a rapid rate (approximately 3 square degrees per day as seen from any given point on the Earth, or 30 square degrees per day when integrated over the surface of the Earth), and it therefore frequently occults strong compact radio sources. Lunar occultations were used extensively in the 1960s to measure the structure and positions of celestial radio sources [7,8]. However, the topography of the lunar surface introduces variations as large as 3 arcsec in the limb of the Moon. These variations are known to an accuracy of approximately 0.2 arcsec [9]. Astrometric measurements from lunar occultations (e.g., a measurement of 3C 273 [10]) are limited to this accuracy.

Planetary occultations are much less sensitive than lunar occultations to errors from topography because planets are so much farther from the Earth than the Moon. The potential astrometric accuracy is therefore better. However, because planets subtend a much smaller angular diameter than the Moon and move more slowly on the sky, they sweep across a much smaller solid angle of sky per unit time than does the Moon (e.g., this rate of solid angle coverage is a factor of approximately 3000 smaller for Mercury and Venus than for the Moon) and occult many fewer strong radio sources.

II. The Occultation of P 0507+17 by Venus: Event Parameters and Observations

The search for planetary occultation events has been described elsewhere [11]. It is briefly summarized here. The JPL Planetary Ephemeris DE200 [12] was used to search through several catalogs of compact radio sources to identify planetary occultation events. The radio source catalogs that were searched include the JPL Astrometric VLBI Catalog [1], a catalog of compact sources within 10 deg of the ecliptic [13], and the Very Large Array (VLA) Calibrator Catalog. The period from January 1, 1988, to January 1, 2000, was searched.

Based on the strength of the occulted radio source and the sensitivity of radio telescopes in the occultation region on the Earth, the most favorable event discovered in this search (by a substantial margin) was an occultation of P 0507+17 by Venus on July 19, 1988. The region on the Earth where this event was visible was quite restricted: southeastern Australia, New Zealand, and a portion of the South Pacific Ocean extending eastward to 110 deg W longitude. Outside this region, one or both of the following constraints were violated. The first constraint was that Venus must be above the horizon at the time of the event, as seen from a given location on the Earth. The second constraint was that this location on the Earth must pass inside the umbra (i.e., the source must pass inside the limb of Venus). The finite distance between the Earth and Venus makes the relative position of Venus and a background radio source a function of location on the Earth. For the Earth–Venus distance at the time of this occultation, the position of Venus on the sky varied by 38 arcsec across the surface of the Earth.

There were two large radio telescopes within the occultation region: a 64-m diameter antenna at Parkes, Australia (operated by the Commonwealth Scientific and Industrial Research Organization) and a 70-m diameter antenna (DSS 43) in Tidbinbilla, Australia (operated by the JPL Deep Space Network). Due to a minimum elevation limit of 30 deg for this antenna, the Parkes antenna could not observe this event, but the event was observed with DSS 43. Parameters of this event, as seen from Tidbinbilla, are given in Table 1.

Total power measurements were made in two channels of 12-MHz bandwidth (BW) each, centered at 2272 and 2284 MHz. The integration time for each measurement was 0.050 sec. The data for the 2284-MHz channel are shown in Figs. 1(a) and (b), for ingress and egress, respectively. The observing frequency was chosen to maximize the ratio of the flux density of P 0507+17 to the flux density of Venus. At 8.4 GHz (the other observing frequency band available at DSS 43), that ratio was lower by a factor of 12. When combined with the smaller primary antenna beamwidth at the higher frequency (which causes time-dependent pointing errors to be more serious), the ratio of time variations in the total system temperature (due to causes other than the occultation) to the antenna temperature of P 0507+17 was expected to be much larger at 8.4 GHz than at 2.3 GHz. As an example, if the mean direction of antenna pointing were correct, drifts of 10 arcsec would cause system temperature variations of ~0.6 K at 8.4 GHz, but only ~0.003 K at 2.3 GHz. If the mean direction of antenna pointing were offset from the source direction, the system temperature variations would be larger at both frequencies. The antenna temperature of P 0507+17 was approximately 0.8 K at both frequencies. The total system temperature (including Venus and P 0507+17) was approximately 36 K at 2.3 GHz.

The BW was chosen as a trade-off between maximizing the signal-to-noise ratio (SNR), which argued for high BW, and minimizing the attenuation of high-order Fresnel fringes (discussed in the next section), which argued for a small BW. The final astrometric accuracy was not limited by either SNR or attenuation of high-order Fresnel fringes.

The weather at Tidbinbilla at the time of the occultation was changing rapidly. During ingress, mostly clear weather resulted in a stable system temperature. The data shown in Fig. 1(a) have not been modified. However, during egress, heavy clouds caused large, nonlinear system-temperature drifts. Polynomials of system temperature versus time were fit to the midpoints of the fringes in the raw data (i.e., halfway between maxima and minima) and subtracted to give the results plotted in Fig. 1(b). Two quartic polynomials, one for each half of the time span in Fig. 1(b), were used. Because of concern that this polynomial system temperature subtraction process had introduced errors into the measured epochs of fringe maxima and minima, the following test was performed: A cubic polynomial was added to the curve plotted in Fig. 1(b). This polynomial was zero at both ends of the data span, with a peak–peak span of 1 K in between [larger, by a factor >10, than any expected deviation of the true light curve from Fig. 1(b)]. The changes in the epochs of the fitted fringe maxima and minima resulting from the addition of this cubic polynomial had a mean value of 0.06 sec, with a standard deviation of 0.10 sec. The mean time offset of 0.06 sec corresponded to an offset in ecliptic longitude of 0.001 arcsec.

III. Occultation Theory

The theory for lunar occultations of celestial radio sources has been well developed [14,8]. Fresnel diffraction theory (e.g., [15]) expresses the received power as an integral across the two-dimensional impact plane (geometry shown in Fig. 2). For the case where the curvature of the limb can be neglected (a good approximation for lunar occultations), the received monochromatic flux density $I(p,\nu)$ at frequency ν for an impact parameter p is

$$
\begin{aligned}
&I(p,\nu) \\
&= \frac{I_\infty(\nu)}{2}\left[\int_0^\infty du \int_{-\infty}^{\infty} dv \cos\left\{\frac{\pi}{2}\left[(u-\psi)^2+v^2\right]\right\}\right]^2 \\
&+ \frac{I_\infty(\nu)}{2}\left[\int_0^\infty du \int_{-\infty}^{\infty} dv \sin\left\{\frac{\pi}{2}\left[(u-\psi)^2+v^2\right]\right\}\right]^2
\end{aligned}
\tag{1}
$$

The parameters u and v are normalized coordinates (Fresnel units) in the impact plane: $u = x\sqrt{2\nu/cd}$ and $v = y\sqrt{2\nu/cd}$, where x and y are physical displacements (in the impact plane) perpendicular and parallel to the limb, respectively; c is the velocity of light; d is the distance from the Earth antenna to the impact plane; and ψ is the impact parameter, in Fresnel units ($\psi = p\sqrt{2\nu/cd}$). The phase term $(\pi/2)[(u-\psi)^2+v^2]$ in the sine and cosine functions is the excess geometric path length (in radians) from that point in the impact plane to the Earth antenna, relative to the length of the most direct ray path. The monochromatic flux density as $p \to \infty$ is $I_\infty(\nu)$ (i.e., the total unobscured flux density of the occulted source). The limits of integration can be changed to take into account the curvature of the limb.

For the case of negligible curvature of the limb and no atmosphere on the occulting object, the identity

$$
\int_0^\infty \cos\left(\frac{\pi}{2}x^2\right) dx = \int_0^\infty \sin\left(\frac{\pi}{2}x^2\right) dx = \frac{1}{2}
$$

can be used to simplify Eq. (1).

$$I(p,\nu) = \frac{I_\infty(\nu)}{2}\left[\left(\frac{1}{2}+C(\psi)\right)^2+\left(\frac{1}{2}+S(\psi)\right)^2\right] \quad (2)$$

where $C(\psi)$ and $S(\psi)$ are the standard Fresnel integrals:

$$C(\psi) \equiv \int_0^\psi \cos\left(\frac{\pi}{2}x^2\right) dx$$

$$S(\psi) \equiv \int_0^\psi \sin\left(\frac{\pi}{2}x^2\right) dx$$

Equation (2) represents the vacuum Fresnel light curve for a point source. For an observed lunar occultation of a very compact source, the observed light curve can be fitted to a vacuum Fresnel light curve with two source coordinates as adjustable parameters. For a resolved source, the one-dimensional brightness distribution (integrated along the direction parallel to the limb of the Moon) can be derived from the light curve [14].

For an occultation by a planet with an atmosphere, the calculation of the intensity at the Earth becomes more complex. The integration in Eq. (1) assumes a constant amplitude and phase across the impact plane, but an atmosphere will, in general, modify both the amplitude and the phase of an incoming wave front. A thin screen in the impact plane, which changes the phase and the amplitude of the incoming radiation, can be incorporated into the expression for $I(p,\nu)$

$$\begin{aligned} I(p,\nu) &= \frac{I_\infty(\nu)}{2}\left[\int_0^\infty du \int_{-\infty}^\infty dv\, A(u,v) \right. \\ &\quad \left. \times \cos\left\{\frac{\pi}{2}\left[(u-\psi)^2+v^2\right]+\phi(u,v)\right\}\right]^2 \\ &\quad + \frac{I_\infty(\nu)}{2}\left[\int_0^\infty du \int_{-\infty}^\infty dv\, A(u,v) \right. \\ &\quad \left. \times \sin\left\{\frac{\pi}{2}\left[(u-\psi)^2+v^2\right]+\phi(u,v)\right\}\right]^2 \end{aligned} \quad (3)$$

where $A(u,v)$ and $\phi(u,v)$ are the amplitude and phase of the screen at the location (u,v) in the impact plane. The normalization of $A(u,v)$ is such that $A(u,v) = 1$ in the absence of a screen.

Note that because the conversion from physical coordinates to Fresnel units is a function of ν, the intensity at a given physical impact parameter p will depend on ν, *independent of any spectral variations in the source.* The observed flux $I(p)$ for a bandpass response with shape $B(\nu)$ is

$$I(p) = \int_0^\infty I(p,\nu)B(\nu)d\nu \quad (4)$$

The fringe amplitude exhibited by $I(p)$ will be less than the monochromatic fringe amplitude, due to imperfect coherence (i.e., nonzero phase spread) across the passband. As p increases, the fringes at the ends of the passband get increasingly out of phase, and the fringe amplitude is damped more severely.

The attenuation factor for this data set (0.5-percent fractional bandwidth) reached 0.59 for the highest observed fringe (no. 83) on the ingress light curve and 0.94 for the highest observed fringe (no. 31) on the egress light curve. These attenuation factors were calculated by integrating $I(p,\nu)$ across the passband (assumed to be rectangular) for one each of the best-fitting ingress and egress light curves. The attenuation factor is defined as the ratio of fringe amplitude for the actual bandpass to the monochromatic fringe amplitude.

The distance between P 0507+17 and Venus was $\sim 10^{15}$ times larger than the Earth–Venus distance. Therefore, the path length from P 0507+17 was essentially constant across the impact plane. The geometric phase terms in the Fresnel integrals, $(\pi/2)\ [(u-\psi)^2+v^2]$, were determined entirely by the Earth–Venus distance.

For an occultation of a spacecraft that is either in orbit about Venus or making a close flyby, a very different situation occurs. In this case the path length from the spacecraft to different locations in the impact plane varies dramatically. This variation is so large that only a very small region in the impact plane contributes to the Fresnel integrals, and geometric optics are a good approximation. This use of geometric optics, combined with tracking the frequency of the spacecraft carrier signal (Doppler shifts due to changing refraction angles can be measured), allows spacecraft occultation light curves to be inverted [16]. The variation of refractivity N [$N \equiv (n-1)$, where n is the index of refraction] with altitude is thereby obtained. Dual-frequency observations of spacecraft occultations allow a

separation of the refractivity of the neutral and ionized atmosphere, and permit a determination of electron density as a function of altitude.

Reflection of radio waves occurs only for electron densities large enough that the plasma frequency is comparable to the frequency of radiation. The maximum electron density in the dayside ionosphere of Venus is typically $5 - 10 \times 10^5 \mathrm{cm}^{-3}$ [17], giving a plasma frequency of ~10 MHz, far below the 2.3-GHz observation frequency used for this experiment. Absorption at 2.3 GHz in the Venus atmosphere is significant at altitudes of 55 km and lower [18]. Absorption can be incorporated into Eq. (3) by using a value of $A(u, v)$ less than 1.0.

Refraction in the atmosphere is more of a problem. A low-refractivity atmosphere, in which ray paths experience very little lateral displacement (i.e., shift in position perpendicular to their original direction of propagation) during their passage, can be treated as a thin screen. The ionosphere of Venus (maximum displacement of 2.3-GHz ray paths of ~5 m) fulfills this criterion. However, the neutral atmosphere of Venus causes large displacements of radio waves. Below 35 km altitude, the radial refractivity gradient becomes so large that entering radio waves do not escape [19]. Ray paths that pass within a few kilometers of this altitude are displaced by tens or hundreds of kilometers in the impact plane before emerging from the troposphere.

In order to incorporate these large tropospheric effects into the Fresnel integrals, a hybrid geometric/physical optics approach was used. Geometric ray-tracing calculations were made for the Venus troposphere for a one-dimensional grid of impact parameter p: 47–110 km in 20-m increments (rays entering with an impact parameter less than 47 km are refracted inside the 35-km boundary and ultimately reach the surface of the planet). The ray tracing was performed from the point of entry into the troposphere (chosen as 110 km altitude, where the refractivity is less than 10^{-7}), past the true impact plane, to a shifted impact plane, (i.e., closer to the Earth in the occultation geometry and outside the Venus atmosphere for nearly all ray paths). A sample displaced impact plane is shown in Fig. 2(a). Two separate ray-tracing runs, with impact plane shifts of 1000 and 4000 km were performed. Altitude profiles of refractivity [20] and the 2.3-GHz absorption coefficient [18] were used in these ray tracings.

The amplitude on this shifted impact plane was less than the amplitude on the true impact plane due to two effects. The first was absorption in the troposphere; the amplitude reduction due to this effect was $e^{-\tau(p)}$, where $\tau(p)$ was the integrated optical depth for a ray with impa parameter p. The second effect was refractive amplitu loss. Because the refractivity of the Venus troposphe increases with depth, rays that were parallel when th entered the troposphere diverged. In particular, the se aration $\Delta x'$ in the shifted impact plane was greater tha the difference Δp in impact parameters, diluting the r diation (i.e., reducing its amplitude). The refractive an plitude loss was dp/dx', and this factor was multiplied b the absorptive amplitude loss $e^{-\tau(p)}$ to give the amplitu A at location x' in the shifted impact plane. The "scre phase" ϕ for use in Eq. (3) was the excess phase for th ray path: the difference between its actual phase and th phase it would have had at the shifted impact plane i the absence of an atmosphere. This phase was the sum a dielectric term $\int 2\pi\nu(n-1)ds/c$ and a geometric ter $\int 2\pi\nu(\sec\theta - 1)ds/c$, where ν is the frequency of radiatio (2.3 GHz) and θ is the angle between the ray path at tha point and its initial direction. Both integrals were calcu lated over the entire path length from the entry into th troposphere until the arrival at the shifted impact plan (Note that the dielectric term remained constant after th ray exited from the troposphere.) The exit angle fro the troposphere was also calculated in order to allow th ionospheric contribution to be added later.

For sufficiently small impact parameters ($p < 53$ km and $p < 47.4$ km for impact plane shifts of 1000 km an 4000 km, respectively), the ray paths had not emerge from the troposphere when they reached the shifted impac plane. Ray tracing for these ray paths continued until the exited from the troposphere. The following procedure wa used to derive values of amplitude A and phase ϕ to us in the Fresnel integrals for these impact parameters. Th value of A was the product of the refractive and absorp tive losses at the shifted impact plane. The value of ϕ wa the excess phase (the sum of the geometric and dielectri integrals) at the shifted impact plane, with one extra ter added. This extra term was the dielectric integral fro the shifted impact plane until exit from the troposphere (i.e., the integrated refractivity encountered by that ray after the shifted impact plane). As discussed below, use of 1000- and 4000-km shifts in the impact plane gave negligible differences in the theoretical occultation light curves. Therefore, it is believed that this hybrid geometric and physical optics approach was adequate.

IV. Data Reduction and Modeling

The analysis of the occultation data fell into three general tasks. The first task was to calculate $I(p)$ in the presence of the atmosphere of Venus. The second task was to

determine $p(t, \Delta\lambda, \Delta\beta)$, where t represents time and $\Delta\lambda$ and $\Delta\beta$ represent the Venus ephemeris/radio frame offsets in ecliptic longitude and ecliptic latitude. The third task was to combine these results to obtain a frame tie. The first task is described in Subsections IV.A–IV.D. The numerical integration technique is described in Subsection IV.A. The properties of the Venus ionosphere are summarized in Subsection IV.B. Modeling of the Venus ionosphere at the occultation epoch, by using the occultation data, is presented in Subsection IV.C. The Venus troposphere and its effect on the occultation data are described in Subsection IV.D. Subsection IV.E describes the calculation of $p(t, \Delta\lambda, \Delta\beta)$. The third task (solving for the frame tie) is presented in Section V.

A. Numerical Integrations

Equation (3) was used to calculate $I(p, \nu)$, and the bandpass attenuation factor (see Section III) was used to convert $I(p, \nu)$ to $I(p)$. The sine and cosine expressions were factored so that the integration over v was a function of u but not of ψ.

$$\int_0^\infty du \int_{-\infty}^{+\infty} dv A(u,v) \cos\left\{\frac{\pi}{2}\left[(u-\psi)^2 + v^2\right] + \phi(u,v)\right\}$$

$$= \int_0^\infty du \cos\left\{\frac{\pi}{2}(u-\psi)^2\right\}$$

$$\times \int_{-\infty}^{+\infty} dv A(u,v) \cos\left\{\frac{\pi}{2}v^2 + \phi(u,v)\right\}$$

$$- \int_0^\infty du \sin\left\{\frac{\pi}{2}(u-\psi)^2\right\}$$

$$\times \int_{-\infty}^{+\infty} dv A(u,v) \sin\left\{\frac{\pi}{2}v^2 + \phi(u,v)\right\}$$

An integration over v was then performed for a grid of 1000–2000 u values. A Simpson's rule numerical integration was used. The integration step size was chosen by performing trial runs in which the step size was reduced until the results converged. The numerical integration was truncated at $\psi = 5$ or at the ionopause, whichever was higher. Analytic expressions for large arguments (≥ 5) of the Fresnel integrals [21] were then used to complete the integrations. This procedure reduced the calculation of I to a one-dimensional numerical integration for each value of ψ.

The values of $A(u, v)$ and $\phi(u, v)$ in Eq. (3) were determined by the Venus troposphere and ionosphere. With negligible absorption or ray path displacement by the ionosphere, $A(u, v)$ was a function only of the troposphere. The ray-tracing results described above were used to specify $A(u, v)$ on the shifted impact plane. Due to refraction in the troposphere, $A(u, v)$ did not drop abruptly to zero at the physical limb of Venus, but instead remained positive (with decreasing magnitude) for a considerable distance "inside" the limb. For impact parameters small enough that the ray paths entered the troposphere, $\phi(u, v)$ was the sum of three terms. The first was the contribution from the ionosphere due to entry with the impact parameter p (the value of p corresponding to a given u and v was obtained from the troposphere ray tracing). The second was the contribution from the troposphere, and the third was the contribution from the ionosphere due to an exit from the troposphere at an angle derived from ray tracing. For larger impact parameters, $\phi(u, v)$ was derived solely from the ionosphere.

B. General Properties of the Venus Ionosphere

Both $A(u, v)$ and $\phi(u, v)$ were entirely determined by p, the variation of the refractivity N with altitude h in the troposphere, and the variation of electron number density n_e with altitude in the ionosphere (n_e was assumed to be a function only of h, although that function was very different on the dayside and nightside of Venus). The function $N(h)$ was assumed to be known and not time variable, although two modified $N(h)$ functions were briefly examined (discussed below). The function $n_e(h)$ has been measured many times from spacecraft occultations (e.g., [22]). It is known to depend strongly upon time and upon the solar zenith angle (SZA) on Venus (the SZA is the angle between the zenith and the Sun, as seen from a given location on Venus). Due to the very slow rotation of Venus, the dependence of $n_e(h)$ on the solar hour angle (other than that predicted from the SZA dependence) is very weak [23]. There were no spacecraft occultation measurements at or near the epoch of the P 0507+17 event, so little more than statistical information on the ionosphere was available. The SZA's on Venus at the points of geometric occultation were 58 deg at ingress and 122 deg at egress. Typical integrated electron densities for "horizontal" lines of sight through the ionosphere (i.e., those which do not intersect the planet) are $\sim 2\text{–}5 \times 10^{13}\text{cm}^{-2}$ for 58-deg SZA and $\sim 1\text{–}2 \times 10^{12}\text{cm}^{-2}$ for 122-deg SZA, which corresponds to 10–30 and 0.5–1.0 cycles of phase at 2.3 GHz, respectively. The effect of the day ionosphere on 2.3-GHz radio waves is therefore much larger than that of the night ionosphere, as is very evident by comparing Figs. 1(a) and 1(b).

As is the case for the ionosphere of the Earth, the Venus ionosphere is strongly time variable. The time scale for large variations of the night ionosphere is at least as short as 24 hours, and perhaps as short as 1 hour [24]. The nightside ionosphere has a peak electron density that is typically around 15,000 cm^{-3}, with variations of a factor >2 in either direction [25]. These variations are weakly correlated with solar wind pressure and almost uncorrelated with solar extreme ultraviolet (EUV) flux. The ionopause height varies from 200 to over 3500 km. The height of maximum electron density is remarkably constant: 142.2 ±4.1 km [26]. Transport of energetic ions from the dayside is believed to drive the night ionosphere [27]; this process is highly dynamic.

The dayside ionosphere is less variable than that of the nightside. Kliore and Mullin showed that the peak electron density can be predicted from the SZA and the solar EUV flux, with a scatter of only approximately 5 percent [28]. They found that the altitude of the peak electron density is 140 ±2 km. Brace et al. [29] have studied the height of the dayside ionopause. The ionopause height varies inversely with solar wind pressure. However, the scatter is large, especially at large SZA (>50 deg) and low solar wind pressure ($<3 \times 10^{-8}$ dynes/cm^2).

C. Modeling of the Venus Ionosphere by Using the Occultation Data

At the epoch of the P 0507+17 occultation, measurements from the Pioneer Venus Orbiter spacecraft gave a solar EUV flux of 1.12×10^{12} photons/cm^2/sec and a solar wind pressure of 1.6×10^{-8} dynes/cm^2.[1] Using the results from [28] and [29], the predicted maximum n_e at an SZA of 58 deg was 3.85×10^5 cm^{-3}, with an uncertainty of 2×10^4 to 3×10^4 cm^{-3}. The predicted ionopause height at an SZA of 58 deg was 950 km, with an uncertainty of 400–500 km. The a priori model for the dayside ionosphere used these values, along with the functional shape for $n_e(h)$ shown in [23]. The a priori uncertainties in the nightside ionosphere were so large that large ranges in ionospheric parameters were examined in the modeling. Once $n_e(h)$ was specified, the contribution of $\phi(u,v)$ from the ionosphere was calculated by a numerical integration perpendicular to the impact plane, for a grid of 1000–2000 p values.

These a priori ionospheric $n_e(h)$ profiles were then adjusted in order to optimize the agreement between the model and the fringe amplitudes of the observed light curves. Initially, the agreement between the model and the observed light curve was judged visually. After an approximate fit was achieved, a numerical agreement factor was used to fine-tune the models. The absolute antenna temperature was difficult to determine from the observations due to weather-induced system temperature variations. Therefore, peak–peak fringe amplitudes were used to guide the modeling process. The weather-induced system temperature variations were substantially slower than the fringe frequency, so that the peak–peak amplitudes appeared to be nearly unaffected. The amplitudes at the fringe peaks were determined with parabolic fits to the peak regions. This procedure effectively discarded data away from the vicinity of the peaks. The SNR was degraded as a result, but still contributed negligibly to the final frame tie uncertainty. The amplitudes between most pairs of adjacent peaks could be well represented by a sinusoid, so that very little information was lost by saving only the peak amplitudes and their epochs.

The numerical integration yielded I for a grid of p. Parabolic fits to the fringe maxima and minima from these integrations gave values of I and p at the peaks. The χ^2 agreement between the model and observed light curves was not used to adjust the models. The errors in the observed peak–peak dayside amplitudes, especially in the lowest order fringes, were much smaller than the disagreement between the model and observed amplitudes. The χ^2 agreement would have been dominated by the modeling error on the largest two or three fringes. Instead, an agreement factor AF was used, which gave less weight to the largest fringes:

$$AF \equiv \sum_i \frac{(amp_{obs} - amp_{mod})^2}{amp_{obs}}$$

In this equation amp_{obs} and amp_{mod} represent the observed and model (i.e., calculated) peak–peak fringe amplitudes. These fringe amplitudes are expressed as fractions of the total source antenna temperature (i.e., when the source is not near Venus on the sky), and therefore they are dimensionless. The sum in AF was performed over the peaks of all the observed fringes (83 on the dayside or 31 on the nightside, where one fringe is defined to include both a maximum and a minimum). In the sum, both peak–peak combinations were included (i.e., the differences between a minimum and *both* adjacent maxima were included). This dual sum was performed to help drive the model dayside light curve toward the observed ingress curve. The values of amp_{mod} were derived from $I(p)$ curves and were therefore independent of $\Delta\lambda$ and $\Delta\beta$. Only the epochs of the fringes depended on these two offsets.

[1] L. Brace, personal communication, Goddard Space Flight Center, Greenbelt, Maryland, 1991.

The optimum weighting scheme for calculating the agreement factor depended on the detailed nature of the ionospheric $n_e(h)$ profile errors, which were unknown. It also depended on the manner in which these $n_e(h)$ errors projected into the values of amp_{mod}, which was highly nonlinear. An alternate agreement factor AF_{alt} was therefore also calculated by using a different weighting scheme:

$$AF_{alt} \equiv \sum_i \frac{(amp_{obs} - amp_{mod})^2}{amp_{obs}^2}$$

AF and AF_{alt} gave similar results as to which of two model light curves was a better fit to the data, although AF gave greater weight to the strongest fringes than did AF_{alt}. The process of model adjustment (described below) used AF.

1. Egress: Night Ionosphere of Venus. For the night ionosphere, simple models were used. The primary ionospheric component was modeled with seven parameters: h_{min}, $n_{e_{max}}$, h_{peak}, n_{e_2}, h_2, l, and h_{max} ($h_{min} < h_{peak} < h_2 \leq h_{max}$). In this model, $n_e = 0$ for $h < h_{min}$, $n_e = n_{e_{max}}$ at $h = h_{peak}$, and $n_e = n_{e_2}$ at $h = h_2$. Linear interpolation was used for the electron density in the region $h_{min} < h \leq h_2$ (i.e., the electron density rose linearly from zero to a peak value at $h = h_{peak}$ and then fell linearly to a value n_{e_2} at $h = h_2$). For $h_2 < h < h_{max}$, $n_e = n_{e_2} e^{-(h-h_2)/l}$. For $h \geq h_{max}$, $n_e = 0$. For all models, $n_{e_2} < 0.3 n_{e_{max}}$. A second component, with a peak electron density <10 percent of that of the primary component, consisted of only four parameters, as h_{max} was constrained to equal h_2, with $n_{e_2} = 0$. For the primary component h_{peak} was chosen as 142 km (the mean value observed from spacecraft occultations) for most models. The second component was located at an altitude well above the main peak of the second component [i.e., $h_{min}(2) > h_2(1) + 50$ km, where $h_{min}(2)$ was the value of h_{min} for component 2 and $h_2(1)$ was the value of h_2 for component 1]. Varying the ten parameters [other than $h_{peak}(1)$] produced several model ionospheres which all fit the observed amplitudes equally well. The qualitative agreement with the observed light curve was quite good (Fig. 3), but there were differences in detail. Ionospheric models with h_{peak} for the primary component of 147 and 152 km produced equally good fits. The observed egress light curve does not appear to constrain this parameter.

2. Ingress: Day Ionosphere of Venus. None of these eleven-parameter ionosphere models gave even a qualitative agreement with the observed ingress (dayside) light curve, so a more complex ionosphere model was used. In this model, a table of electron densities n_{e_i} at 20–30 altitudes was used, with a cubic spline interpolation in the value of n_e between these altitudes. For both types of models (the simple nightside model and the spline dayside model) the fitting process was automated. The parameters (the ten parameters other than h_{peak} for the nightside, or the 20–30 n_{e_i} values for the dayside) were adjusted one at a time through a series of 2–6 values relative to the nominal value (e.g., −5 percent, −1 percent, +1 percent, and +5 percent) and the agreement factor was calculated each time. If any of these adjusted values reduced the agreement factor AF, the nominal value for that parameter was then changed. Typically, this process was repeated for 5–10 loops through the complete set of parameters. The calculation of the light curve (which was needed in order to calculate AF) for one set of parameters required approximately six minutes of CPU time on a SUN 4/390 computer. This strongly limited the possible ways of adjusting parameters and precluded a full least-squares solution. The fit to the dayside curve could probably have been improved if multiple n_{e_i} values had been adjusted simultaneously. The large spread in frame-tie offsets among models which all fit the observed light curves fairly well suggests that an improved fitting procedure would not have significantly improved the frame-tie result.

For most fits, h_{peak} (the altitude of maximum n_e) was fixed at 140 km, although two fits each were run with h_{peak} of 135 and 144 km. The value of $n_{e_{max}}$ was constrained to lie within $3 \times 10^4 \text{cm}^{-3}$ of $3.85 \times 10^5 \text{cm}^{-3}$. The major differences among different models were the ionopause height (a range of 500–1300 km) and the value of n_e at the ionopause (a range of 2×10^3 to 10^4cm^{-3}). A good (but not excellent) fit to the observed dayside light curve was obtained for several combinations of the two parameters. As for the nightside modeling, the quality of the fit did not depend on h_{peak} (at least within the range of values that are consistent with spacecraft measurements of this parameter). Two of the best fits are shown in Fig. 4.

The structure of the ingress light curve shortly before the flux density dropped to zero was complex. Some models of the dayside ionosphere yielded one or more fringes in the region with amplitude much smaller than that of any observed fringe. When matching fringes between the model and observed light curves in order to calculate AF, any fringes among the first ten with an amplitude less than 20 percent of the total flux density [$I_\infty(\nu)$] were ignored (this limit is equal to 0.16 K in antenna temperature). Such fringes are indicated with small arrows in Fig. 4. These fringes are believed to be due to insufficiencies in

the model, probably as a result of a limited set of parameters.

D. The Venus Troposphere

For both the dayside and nightside model fitting, the sensitivity to troposphere changes was studied. Replacing the entire region of the troposphere below 80 km altitude with an opaque absorber [i.e., setting $A(u, v)$ in Eq. (3) to zero] did not change the results at all. Below 80 km altitude, the tropospheric phase contribution changed so rapidly with the impact parameter that contributions from adjacent regions canceled nearly perfectly. Using the opaque absorber in the analysis reduced or eliminated two sources of uncertainty from the overall modeling procedure. First, uncertainties in the absorption coefficient were unimportant because measurable absorption occurs only below 55 km. Second, concerns about the accuracy of the hybrid geometric/physical optics approach to the troposphere were greatly reduced. Above 80 km altitude, the deflections of ray paths passing through the troposphere are small (100 m at 80 km, 10 m at 90 km). The deviations of the fully modeled troposphere from that of a phase screen approach were negligible.

Above approximately 85 km altitude, the refractivity of the troposphere is poorly known [20]. The refractivity at these altitudes is so low that the Doppler shift of a carrier signal during spacecraft occultations is too small to measure accurately. However, the integrated phase of the troposphere was sufficiently large in this region to affect the occultation light curve of P 0507+17.

A nominal troposphere for light-curve modeling [in the form of a polynomial $\log(N(h))$, where $N(h)$ was the refractivity as a function of altitude] was adopted from [20]. In addition, two alternate troposphere polynomials were each used for one additional model of both the dayside and nightside light curves. These two alternate polynomials were chosen to reflect the range of possible errors in the nominal curve. Table 2 lists refractivity values from these three polynomials.

The effect of these alternate troposphere models was small. Small adjustments in the ionosphere parameters (much smaller than their a priori uncertainties) brought the agreement factor down to nearly the same value as for the nominal troposphere. For the nightside ionosphere, the nominal and the thin tropospheres gave better fits (20–40 percent lower AF) than did the thick troposphere. For the dayside ionosphere, all three tropospheres gave equally good fits. The uncertainty in the ionosphere was too large to allow significant constraints on the troposphere fro this occultation data set.

The magnitude of the effects of the different atmc spheric components is illustrated by their effective angul shift in the position of Venus. This shift in the light cur was approximately 0.25 arcsec for the night ionosphere approximately 0.8 arcsec for the day ionosphere, and ap proximately 0.3 arcsec for the troposphere. Of these thre the tropospheric shift was the most accurately known.

E. Impact Parameter as a Function of Both Time an Frame-Tie Offsets

The calculation of the impact parameter p as a func tion of time and frame-tie offsets required knowledge o the orbits of the Earth and Venus, and of the Earth' rotation. The Planetary Ephemeris DE200 was used t calculate the relative geocentric positions of Venus an P 0507+17 for zero frame-tie offset. Because both DE20 and the JPL Astrometric Catalog are expressed in J2000 no precession corrections were needed. Because P 0507+1 and Venus were very close together on the sky, no aber ration corrections were needed. However, a general rela tivistic correction was needed due to the bending of ra paths in the solar potential from P 0507+17 to Venu (from Venus to the Earth, ray paths from Venus an from P 0507+17 were deflected by the same amount, s that no further shift in the relative position of Venus an P 0507+17 occurred). The magnitude of this correctio was 4.07 mas/$[d \tan(\omega/2)]$ = 8.8 mas. Here d is the dis tance in astronomical units between Venus and the Su and ω is the angle between the Sun and P 0507+17, a seen from Venus. The direction of this bending lay en tirely in ecliptic longitude (in the sense that the apparen position of P 0507+17 was shifted to lower longitudes).

The topocentric parallax (i.e., the shift in the appar ent relative positions of P 0507+17 and Venus relative to their geocentric positions) was 19.6 arcsec $\cos el$, where el represented the elevation of the two objects as seen from DSS 43. It was therefore necessary to know the Earth's orientation and rotation and the station position to a frac tional accuracy better than 5×10^{-5} (320 meters or 10 arcsec) to ensure an error of <1 mas in the parallax cor rection. The station vector was precessed from coordinates of date to J2000.

The impact parameter p was calculated for every 1 sec of time during the ingress and egress intervals displayed in Fig. 1. These calculations were performed for several small frame-tie offsets $\Delta\lambda$ and $\Delta\beta$, in addition to $\Delta\lambda = \Delta\beta = 0$.

Polynomial fits to the tabulated difference in impact parameters for different frame-tie offsets gave $\partial p/\partial\Delta\lambda(t)$ and $\partial p/\partial\Delta\beta(t)$, where t represents time.

For frame-tie offsets that are small as compared with the angular size of Venus, the change $\Delta T_{N_\pm}$ in epoch of a Fresnel fringe (maximum: + or minimum: −) was

$$\Delta T_{N_\pm} = -\frac{\frac{\partial p}{\partial\Delta\lambda}\Delta\lambda + \frac{\partial p}{\partial\Delta\beta}\Delta\beta}{\frac{\partial p}{\partial t}}$$

$$= \frac{\partial T_{N_\pm}}{\partial\Delta\lambda}\Delta\lambda + \frac{\partial T_{N_\pm}}{\partial\Delta\beta}\Delta\beta \qquad (5)$$

All derivatives in Eq. (5) were evaluated at $\Delta\lambda = \Delta\beta = 0$ and $t = T_{N_+}$ or $t = T_{N_-}$. The expressions for $\partial T_{N_\pm}/\partial\Delta\lambda$ and $\partial T_{N_\pm}/\partial\Delta\beta$ for both ingress and egress are given in Table 3. Note that $\partial T_{N_\pm}/\partial\Delta\lambda$ was nearly independent of time, because the path of Venus on the sky was nearly along the ecliptic.

There were a large number of measured fringe maxima or minima (166 during ingress and 62 during egress). It was possible to perform a least-squares solution for both $\Delta\lambda$ and $\Delta\beta$ from any subset of two or more residual values $\Delta T_{N_\pm} \equiv T_{N_{\pm observed}} - T_{N_{\pm model}}$. However, the resulting values of $\Delta\beta$ and $\Delta\lambda$ were highly correlated for time spans shorter than a few hundred seconds. When residuals from both ingress and egress were used together in one fit, the correlation between $\Delta\beta$ and $\Delta\lambda$ was much lower. In particular, if data from both ingress and egress were used and distributed roughly symmetrically in time about the occultation midpoint, the correlation dropped to nearly zero. For such a data set, $\Delta\beta$ was determined by the event duration and $\Delta\lambda$ was determined by its mean epoch.

V. Frame-Tie Results

Residual values ΔT from the two best-fitting models for each of the day and night ionospheres are given in Table 4. Labels of "A" and "B" for the day ionosphere models and "1" and "2" for the night ionosphere models are used for later reference. All four of these models used the nominal Venus troposphere. The values of ΔT are given at intervals of 50 sec, with respect to the reference epochs listed in Table 3 (these reference epochs were within five seconds of the times of geometric occultation for $\Delta\lambda = \Delta\beta = 0$). By using epochs from ingress and egress of $t = \pm 150$ sec and ± 200 sec, a symmetrical solution set was obtained. These four epochs were therefore used to make least-squares frame-tie estimates, in which time residuals at all four epochs were given equal weight. The results are given in Table 5.

There is a large scatter among the four entries in Table 5. The mean values from these four solutions are $\Delta\lambda = +0.03$ arcsec and $\Delta\beta = +0.09$ arcsec, with standard deviations of 0.11 arcsec for both parameters. These standard deviations are clearly only a very crude estimate of the true error. The range of allowed solutions includes $\Delta\lambda = \Delta\beta = 0$, with an error of approximately 0.2 arcsec.

The scatter in $\Delta\lambda$ and $\Delta\beta$ was dominated by uncertainty in the dayside ionosphere. By using egress data alone (this is equivalent to giving very low weight to the ingress data), a more accurate result was obtained. Because $\Delta\lambda$ and $\Delta\beta$ were highly correlated for solutions using just the egress data, a linear combination was used. The combination $\Delta\lambda + 0.51\Delta\beta$ was chosen because $\partial T_{N_\pm}/\partial\Delta\lambda$ and $0.51\,(\partial T_{N_\pm}/\partial\Delta\beta)$ were equal at the midpoint epoch (+150 sec) of the egress data set. A least-squares solution for $\Delta\lambda + 0.51\Delta\beta$ from the two sets of night residuals given in Table 4 yielded -0.026 ± 0.030 arcsec. As this estimate was based on only two model ionospheres, a more conservative error of 0.040 arcsec was adopted.

Model fits with the thin or thick troposphere models gave residual times ΔT that differed from nominal troposphere results by approximately 1 sec for ingress and <0.5 sec for egress. If it had been possible to somehow calibrate the Venus ionosphere, the errors in the frame-tie offsets due to tropospheric uncertainty would have been at least as small as those calculated for the night ionosphere (i.e., $\leq$0.01 arcsec, equivalent to 0.5 sec).

VI. Conclusions

"Light curves" of 2.3 GHz from both ingress and egress were obtained for the occultation of P 0507+17 by Venus. By comparing model light curves with the observed light curves for both ingress and egress, the frame-tie offsets $\Delta\lambda$ and $\Delta\beta$ (ecliptic longitude and latitude) were determined. The shape and strength of the Venus ionosphere were constrained by a priori information and by the requirement that the observed and model light curves have the same shape. Despite these constraints, uncertainty in the Venus ionosphere was the dominant error source for the frame-tie measurement. The next-largest error was due to uncertainty in the upper troposphere (altitude >85 km) of

Venus; the contribution from this error source was approximately a factor of 10 smaller than that from ionospheric uncertainty.

The solution for both $\Delta\lambda$ and $\Delta\beta$ gave results consistent with zero offsets, but with an error of approximately 0.2 arcsec. Using the egress (nightside) light curve alone (i.e., discarding data with large systematic errors) yielded higher accuracy for a linear combination of $\Delta\lambda$ and $\Delta\beta$: $\Delta\lambda + 0.51\Delta\beta = -0.026 \pm 0.04$ arcsec. Techniques other than occultations promise higher accuracy. VLBI observations of the Magellan spacecraft, now in orbit about Venus, have been performed, and are expected to yield a frame-tie accuracy of better than 0.01arcsec.[2] Radio occultations would appear to have useful astrometric potential only for planetary bodies with negligible ionospheres.

[2] C. Hildebrand, personal communication, Jet Propulsion Laboratory, Pasadena, California, 1991.

Acknowledgments

The author thanks J. Reynolds for extensive help with the data taking, and R. Treuhaft and X X Newhall, who made many useful suggestions based on a reading of a draft of this article.

References

[1] O. J. Sovers, C. D. Edwards, C. S. Jacobs, G. E. Lanyi, K. M. Liewer, and R. N. Treuhaft, "Astrometric Results of 1978–1985 Deep Space Network Radio Interferometry: The JPL 1987-1 Extragalactic Source Catalog," *The Astronomical Journal*, vol. 95, no. 6, pp. 1647–1658, 1988.

[2] C. Ma, D. B. Schaffer, C. De Vegt, K. J. Johnston, and J. L. Russell, "A Radio Optical Reference Frame, Part 1: Precise Radio Source Positions Determined by Mark III VLBI, Observations From 1979 to 1988 and a Tie to the FK5," *Astron. J*, vol. 99, no. 4, pp. 1284–1298, April 1990.

[3] E. M. Standish, "Celestial Reference Frames: Definitions and Accuracies," in *The Impact of VLBI on Astrophysics and Geophysics*, IAU Symposium no. 129, edited by M. J. Reid and J. M. Moran, Boston: Kluwer Academic Publishers, pp. 309–315, 1988.

[4] M. H. Finger and W. M. Folkner, "A Determination of the Radio-Planetary Frame Tie and the DSN Tracking Station Locations," AIAA paper 90–2905, presented at the AIAA/AAS Astrodynamics Conference, Portland, Oregon, August 20–22, 1990.

[5] J. G. Williams, "Determining Asteroid Masses from Perturbations on Mars," *Icarus*, vol. 57, p. 1, 1984.

[6] J. S. Border, F. F. Donivan, 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, presented at the 1982 AIAA Conference, San Diego, California, August 1982.

[7] M. H. Cohen, "High-Resolution Observations of Radio Sources," *Annual Reviews of Astronomy and Astrophysics*, vol. 7, pp. 619–664, 1969.

[8] C. Hazard, "Lunar Occultation Measurements," in *Methods of Experimental Physics*, vol. 12C, edited by M. L. Meeks, New York: Academic Press, pp. 92–117, 1976.

[9] C. B. Watts, *Astronomical Papers of the American Ephemeris and Nautical Almanac*, vol. 17, Washington, DC; Government Printing Office, 1963.

[10] C. Hazard, J. Sutton, A. N. Argue, C. M. Kenworthy, L. V. Morrison, and C. A. Murray, "Accurate Radio and Optical Positions of 3C273B," *Nature Physical Science*, vol. 233, pp. 89–91, October 4, 1971.

[11] R. P. Linfield, "Using Planetary Occultations of Radio Sources for Frame-Tie Measurements, Part 1: Motivation and Search for Events," *TDA Progress Report 42-103*, vol. July–September 1990, Jet Propulsion Laboratory, Pasadena, California, pp. 1–13, November 15, 1990.

[12] E. M. Standish, "Orientation of the JPL Ephemerides, DE 200/LE 200, to the Dynamical Equinox of J 2000," *Astronomy and Astrophysics*, vol. 114, pp. 297–302, 1982.

[13] A. E. Wehrle, D. D. Morabito, and R. A. Preston, "Very Long Baseline Interferometry Observations of 257 Extragalactic Radio Sources in the Ecliptic Region," *The Astronomical Journal*, vol. 89, pp. 314–336, 1984.

[14] P. A. G. Scheuer, "On the Use of Lunar Occultations for Investigating the Angular Structure of Radio Sources," *Australian Journal of Physics*, vol. 15, no. 2, pp. 332–343, 1962.

[15] M. Born and E. Wolf, *Principles of Optics*, Elmsford, New York: Pergamon Press, pp. 370–434, 1964.

[16] V. R. Eshleman, G. Fjeldbo, and A. Kliore, "The Neutral Atmosphere of Venus as Studied with the Mariner V Radio Occultation Experiments," *Astronomical Journal*, vol. 76, pp. 123–140, 1971.

[17] S. J. Bauer, L. M. Brace, H. A. Taylor, Jr., T. K. Breus, A. J. Kliore, W. C. Knudsen, A. F. Nagy, C. T. Russell, and N. A. Savich, "The Venus Ionosphere," *Advances in Space Research*, vol. 5, no. 11, pp. 233–267, 1985.

[18] J. M. Jenkins and P. G. Steffes, "Results for 13-cm Absorptivity and H_2SO_4 Abundance Profiles From the Season-10 (1986) Pioneer Venus Orbiter Radio Occultation Experiment," *Icarus*, vol. 90, pp. 129–138, 1991.

[19] V. S. Avduevsky, M. Y. Marov, and M. K. Rozhdestvensky, "Model of the Atmosphere of the Planet Venus Based on Results of Measurements Made by the Soviet Automatic Interplanetary Station Venera 4," *Journal of The Atmospheric Sciences*, vol. 25, pp. 537–545, 1968.

[20] A. Seiff, J. T. Schofield, A. J. Kliore, F. W. Taylor, and S. S. Limaye, "Models of the Structure of the Atmosphere of Venus from the Surface to 100 Kilometers Altitude," in *The Venus International Reference Atmosphere, Advances in Space Research*, vol. 5, no. 11, pp. 3–58, 1985.

[21] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, New York: Dover Publications, p. 322, 1985.

[22] T. E. Cravens, A. J. Kliore, J. U. Kozyra, and A. F. Nagy, "The Ionospheric Peak on the Venus Dayside," *Journal of Geophysical Research*, vol. 86, pp. 11,323–11,329, 1981.

[23] R. F. Theis, L. H. Brace, and H. G. Mayr, "Empirical Models of the Electron Temperature and Density in the Venus Atmosphere," *Journal of Geophysical Research*, vol. 85, pp. 7787–7794, December 30, 1980.

[24] L. H. Brace, T. I. Gombosi, A. J. Kliore, W. C. Knudsen, A. F. Nagy, and H. A. Taylor, "The Ionosphere of Venus Observations and Their Interpretation," in *Venus*, edited by D. M. Hunten, L. Colin, T. M. Donahue, and V. I. Moroz, Tucson: University of Arizona Press, pp. 779–840, 1983.

[25] L. H. Brace, R. F. Theis, and J. D. Mihalov, "Response of Nightside Ionosphere and Ionotail of Venus to Variations in Solar EUV and Solar Wind Dynamic Pressure," *Journal of Geophysical Research*, vol. 95, pp. 4075–4084, April 1, 1990.

[26] A. J. Kliore, I. R. Patel, A. F. Nagy, T. E. Cravens, and T. I. Gombosi, "Initial Observations of the Nightside Ionosphere of Venus from Pioneer Venus Orbiter Radio Occultations," *Science*, vol. 205, pp. 99–102, July 6, 1979.

[27] L. H. Brace, R. F. Theis, H. B. Niemann, H. G. Mayr, W. R. Hoegy, and A. F. Nagy, "Empirical Models of the Electron Temperature and Density in the Nightside Venus Ionosphere," *Science*, vol. 205, pp. 102–105, July 6, 1979.

[28] A. J. Kliore, J. Arvydas, and L. F. Mullen, "The Long-Term Behavior of the Main Peak of the Dayside Ionosphere of Venus During Solar Cycle 21 and Its Implications on the Effect of the Solar Cycle Upon the Electron Temperature in the Main Peak Region," *Journal of Geophysical Research*, vol. 94, pp. 13,339–13,351, October 1, 1989.

[29] L. H. Brace, R. F. Theis, W. R. Hoegy, J. H. Wolfe, J. D. Mihalov, C. T. Russell, R. C. Elphic, and A. F. Nagy, "The Dynamic Behavior of the Venus Ionosphere in Response to Solar Wind Interactions," *Journal of Geophysical Research*, vol. 85, pp. 7663–7678, December 30, 1980.

Table 1. Parameters of July 19, 1988, occultation by Venus as seen from Tidbinbilla, Australia.

Parameter	Value
P 0507+17 position	
Right ascension, J2000	05 10 02.3691
Declination, J2000	18 00 41.581
Flux density of P 0507+17	
2.3 GHz	1.0 Jy
8.4 GHz	1.0 Jy
Flux density of Venus	
2.3 GHz	2.4 Jy
8.4 GHz	30 Jy
Approximate midpoint of occultation	UT 19:58
Approximate duration of occultation	19 min
Elevation of Venus at midpoint	18 deg
Angle between Venus and the Sun (as seen from the Earth)	40 deg
Distance from Earth to Venus	67.5×10^6 km
Angular diameter of Venus	37 arcsec
Angular velocity of Venus	1.21 arcsec/min

Table 2. Troposphere refractivity values used in light-curve modeling.

Altitude, km	$\log_{10} N$		
	Nominal	Thin	Thick
80	−5.4	−5.4	−5.4
85	−6.0	−6.0	−6.0
90	−6.6	−6.9	−6.4
95	−7.1	−8.0	−6.6
100	−7.7	−9.1	−6.8
105	−8.2	−10.1	−7.1
110	−8.8	−10.8	−7.4

Table 3. Dependence of fringe epochs upon frame-tie offsets.

Quantity	Constant coefficient	Linear coefficient	Quadratic coefficient
Ingress $\partial T_{N_\pm}/\partial\Delta\lambda$	−48.4	+0.003	
Ingress $\partial T_{N_\pm}/\partial\Delta\beta$	−71.7	−0.124	-2.1×10^{-4}
Egress $\partial T_{N_\pm}/\partial\Delta\lambda$	−53.4	+0.002	
Egress $\partial T_{N_\pm}/\partial\Delta\beta$	+67.2	−0.121	$+2.0 \times 10^{-4}$

Note: Reference times for the polynomials are 19:48:50 for ingress and 20:07:42 for egress. Units for $\Delta\lambda$ and $\Delta\beta$ are arcseconds, and units for T_N are seconds.

Table 4. Fringe epoch residuals for the best-fitting models.

Model	−150 sec	−200 sec	−250 sec	−300 sec
Day A	−18 sec	−14 sec	−14.5 sec	−16 sec
Day B	+ 7 sec	0 sec	− 5.5 sec	− 9 sec

Model	+100 sec	+150 sec	+200 sec
Night 1	−1.3 sec	+1 sec	+1.3 sec
Night 2	+4.4 sec	+4.8 sec	+4.6 sec

Table 5. Frame-tie offsets.

Model combination	$\Delta\lambda$, arcsec	$\Delta\beta$, arcsec
A1	+0.14	+0.16
A2	+0.10	+0.19
B1	−0.04	−0.02
B2	−0.08	−0.01

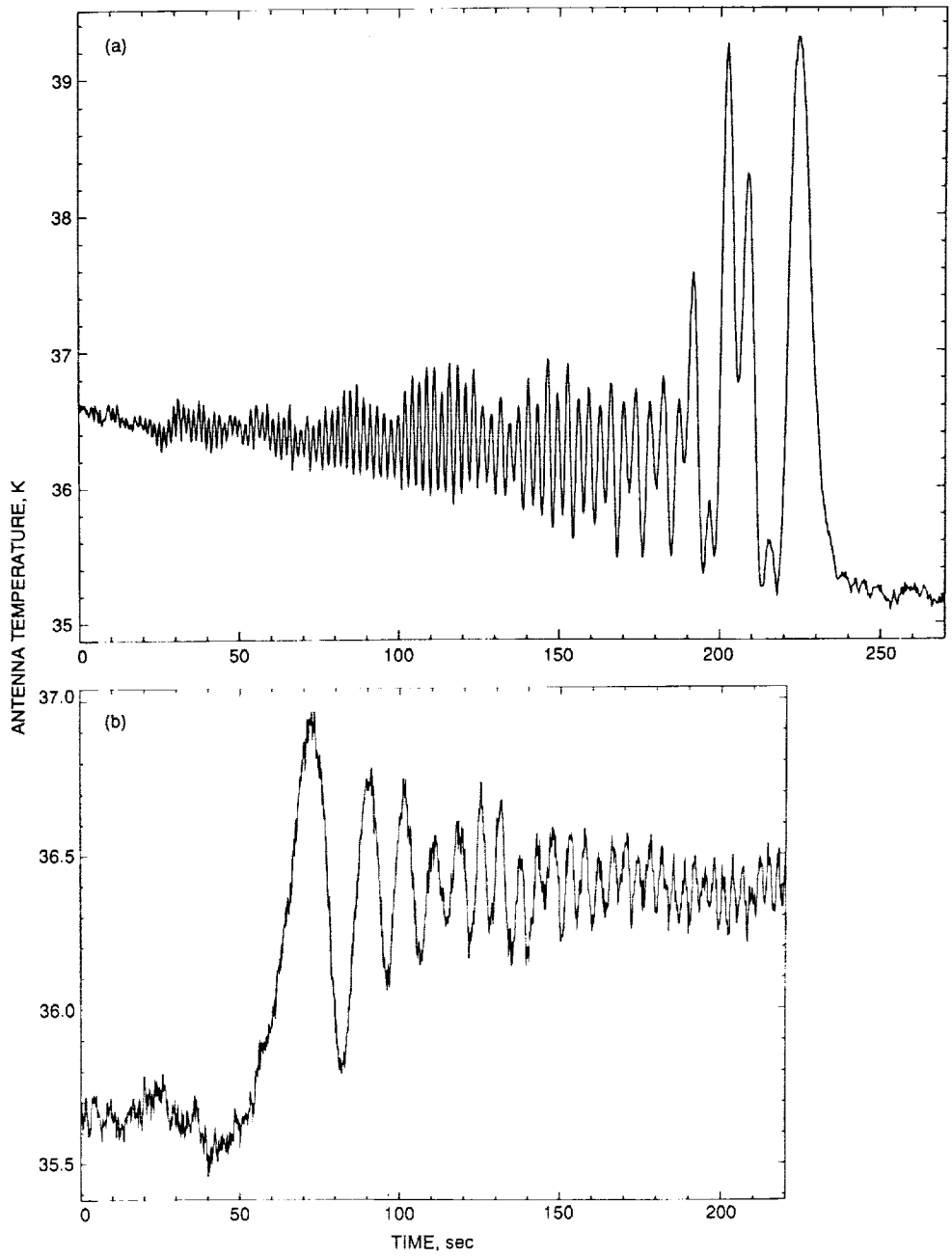


Fig. 1. Observed light curves for the occultation of P 0507+17 by Venus. The data were taken with the 70-m DSN antenna at Tidbinbilla, at a frequency of 2.3 GHz: (a) ingress light curve (dayside of Venus) and (b) egress light curve (nightside of Venus).

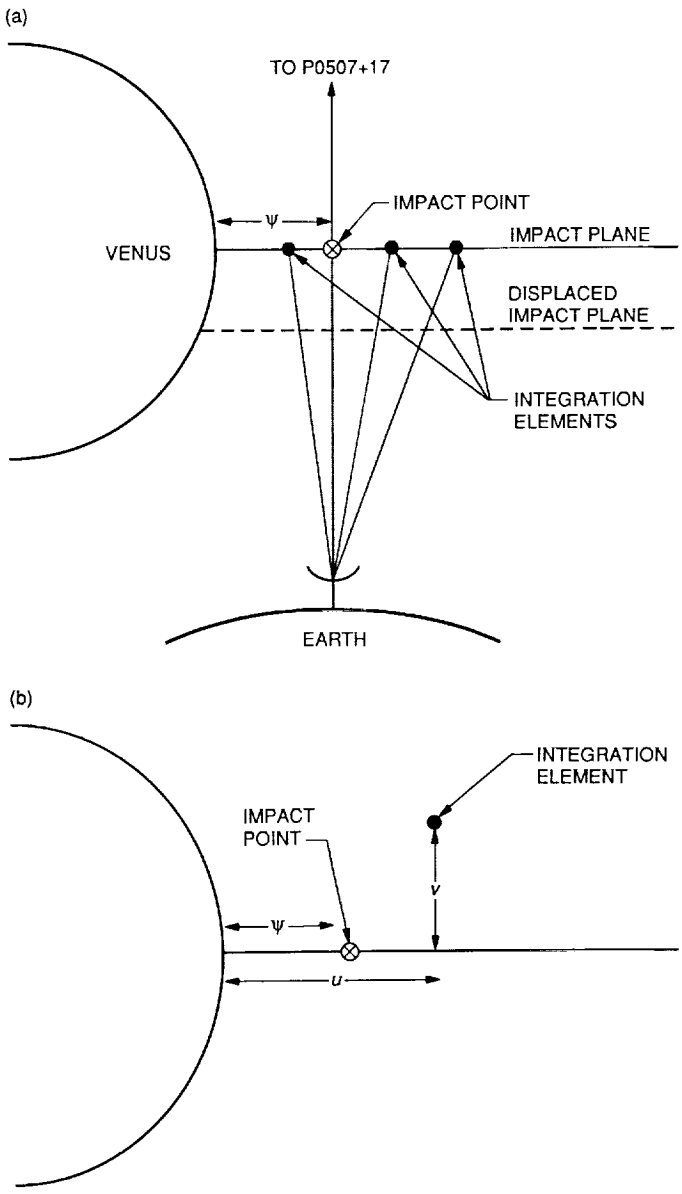


Fig. 2. Geometry for occultation calculations for (a) the observer-source dimension and one dimension in the impact plane. Note the impact point, the three integration elements (for Fresnel integrals) in the impact plane, and a sample displaced impact plane (discussed in the text). (b) The two-dimensional impact plane. Both u and v represent the coefficients (in Fresnel units) of the integration element, and ψ represents the impact parameter.

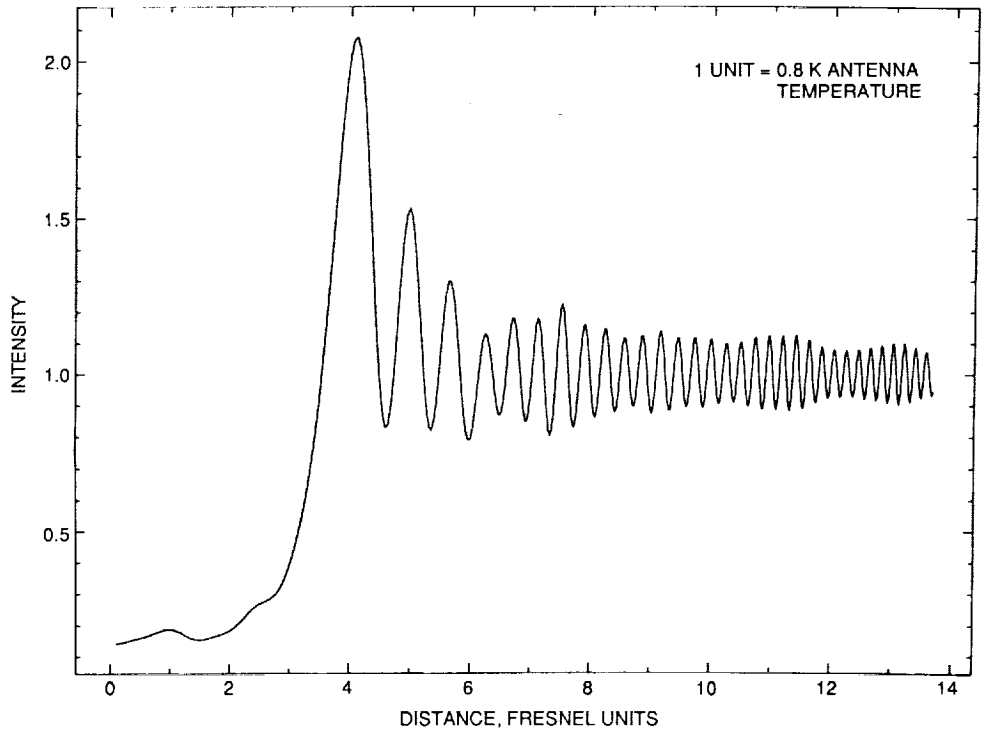


Fig. 3. One of the best-fitting model light curves for the egress (nightside) data.

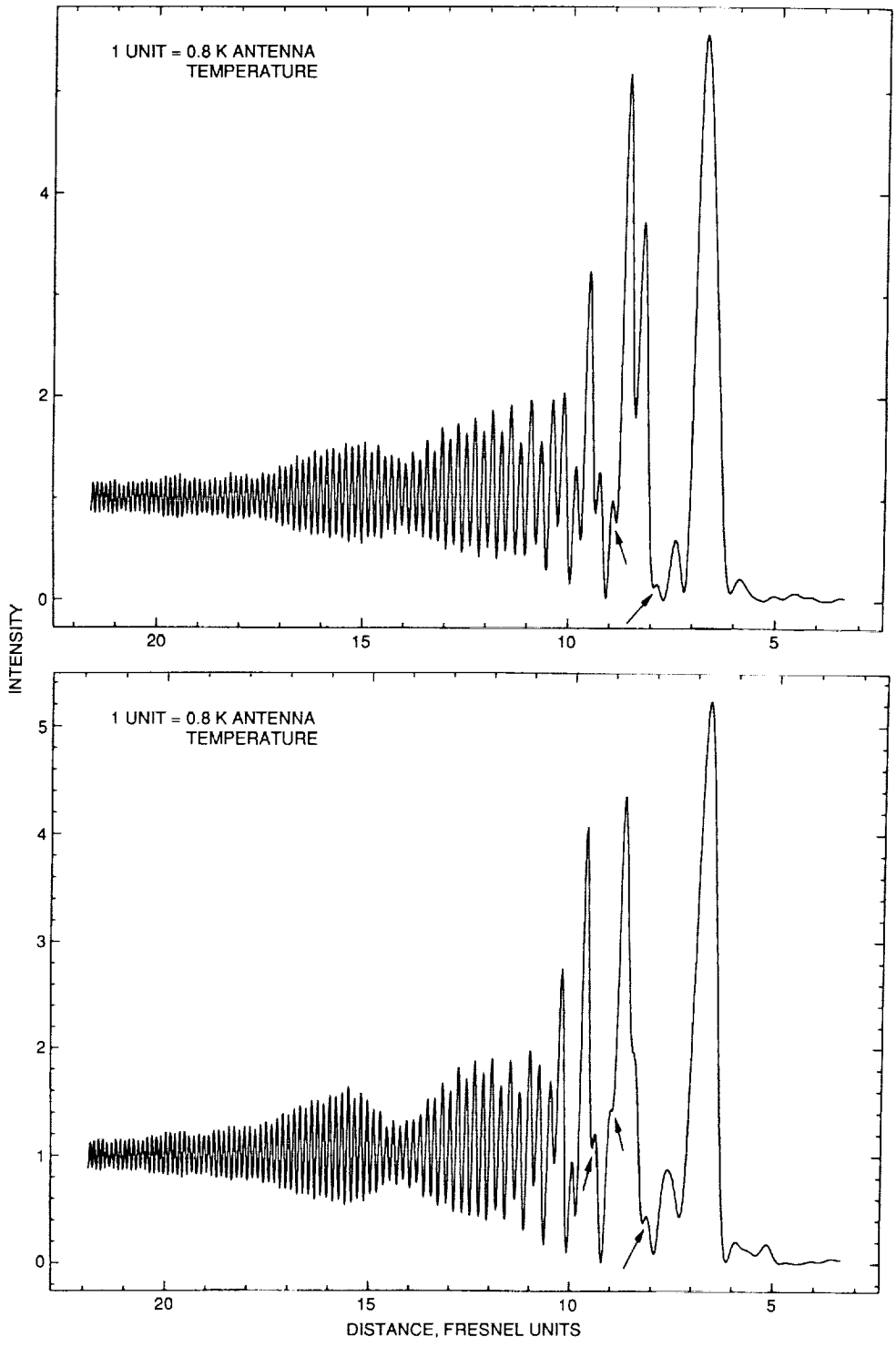


Fig. 4. Two of the best-fitting model light curves for the ingress (dayside) data. The tiny fringes marked with arrows were ignored when calculating residual fringe epochs, as discussed in the text.

A Nanoradian Differential VLBI Tracking Demonstration

R. N. Treuhaft and S. T. Lowe
Tracking Systems and Applications Section

The shift due to Jovian gravitational deflection in the apparent angular position of the radio source P 0201+113 was measured with very long baseline interferometry (VLBI) to demonstrate a differential angular tracking technique with nanoradian accuracy. The raypath of the radio source P 0201+113 passed within 1 mrad of Jupiter (approximately 10 Jovian radii) on March 21, 1988. Its angular position was measured 10 times over 4 hours on that date, with a similar measurement set on April 2, 1988, to track the differential angular gravitational deflection of the raypath. According to general relativity, the expected gravitational bend of the raypath averaged over the duration of the March experiment was approximately 1.45 nrad projected onto the two California–Australia baselines over which it was measured. Measurement accuracies on the order of 0.78 nrad were obtained for each of the ten differential measurements. The χ^2 per degree of freedom of the data for the hypothesis of general relativity was 0.6, which suggests that the modeled dominant errors due to system noise and tropospheric fluctuations fully accounted for the scatter in the measured angular deflections. The χ^2 per degree of freedom for the hypothesis of no gravitational deflection by Jupiter was 4.1, which rejects the no-deflection hypothesis with greater than 99.999-percent confidence. The system noise contributed about 0.34 nrad per combined-baseline differential measurement and tropospheric fluctuations contributed about 0.70 nrad. Unmodeled errors were assessed, which could potentially increase the 0.78-nrad error by about 8 percent. The above χ^2 values, which result from the full accounting of errors, suggest that the nanoradian gravitational deflection signature was successfully tracked.

I. Introduction

This article describes the first demonstration of a multi-source, wide-field very long baseline interferometry (VLBI) tracking technique, with temporally differential accuracies on the order of 1 nrad. In the standard mode of spacecraft angular tracking, called Delta Differential One-Way Ranging (ΔDOR), the VLBI delays from a target spacecraft and one reference radio source are differenced in order to cancel common mode errors. An extension of this technique to include observations of multiple reference sources along with that of the target was proposed to improve on ΔDOR performance [1]. As compared with typical ΔDOR errors of 10–30 nrad, the multiple source, or "local reference frame," approach yields 1-nrad performance; the demonstration reported here was the first test of the mul-

tiple source approach. The sensitivity of the DSN, coupled with wide recorded and spanned bandwidths for the VLBI data, enables a single angular determination with the multisource observation strategy in approximately the same amount of time (approximately 30 minutes) required for a single ΔDOR measurement with the narrow-bandwidth operational system.

In this demonstration, a natural source rather than a spacecraft was used as the target. In the first phase of demonstrating advanced angular tracking techniques, natural radio sources have frequently been used as targets instead of spacecraft [2,3]. The principal reasons for using natural sources are that (1) they exist in sufficient number and strength that a variety of target–reference source geometries are available at any time and (2) their "trajectories" are well known. In most natural source demonstrations, the goal is to see how closely a stationary target can be tracked. In this demonstration, the goal was to track the angular shift of the target source, resulting from planetary gravitational deflection, between two epochs. The angular position of the radio source P 0201+113 was measured ten times during each of two VLBI sessions, on March 21, 1988, and April 2, 1988. According to general relativity, the proximity of the target's raypath to Jupiter (1 mrad, or 10 Jovian radii), on March 21, 1988, produced an average gravitational deflection of about 1.45 nrad. On April 2, 1988, the raypath passed within about 3 degrees of Jupiter, which produced an expected deflection of less than 0.10 nrad, and the same observation schedule was repeated to attempt to track the differential gravitational signature. This differential signature was equivalent, for example, to a spacecraft motion of about 1 km at Jupiter between the March and April sessions. It has been shown, for example [4], that tracking at this level on approach to Jupiter would improve the determination of time of arrival and altitude above Io for Galileo.

During the period of mutual visibility (about 4 hours), the near-occultation event was observed ten times over two California–Australia (approximately 10,600-km) DSN baselines, with DSS 13, DSS 15, and DSS 43. Along with the target source, P 0201+113, several other sources were observed to estimate parameters characterizing clock, Earth rotation, and tropospheric effects [1]. Because the session-to-session *differences* in the angular positions of P 0201+113 were inferred from the VLBI data, the results were largely insensitive to stationary radio source position or structure uncertainties. The results were also largely insensitive to any other error source, such as antenna deformation, which is a function of antenna position, and therefore repeats with the same sidereal schedule. Details of the experimental procedure are presented in Section II following a description of high-precision astrometric VLBI tracking below.

In astrometric VLBI tracking measurements, the angular position shift of a radio source from its expected or a priori value is inferred from the residual geometric delay.[1] The geometric delay is the difference between the arrival times, at each station of a baseline, of an electromagnetic wavefront from a radio source, which can be either a spacecraft or natural radio source. In this article, the geometric delay is defined to be positive if the wavefront arrives at Station 2 later than Station 1. The residual geometric delay for a single observation, $\Delta\tau_g$, is defined here to mean the delay due only to a shift Δs_p of the apparent source coordinate from its expected value, projected onto the baseline. The quantities Δs_p and $\Delta\tau_g$ are related by (see, for example, [1]):

$$\Delta s_p = -\frac{c\Delta\tau_g}{B_p} \tag{1}$$

where B_p is the projection of the baseline onto the plane of the sky and c is the speed of light. The baseline vector, of which B_p is a component, points from Station 1 to Station 2. In the absence of measurement or modeling errors, Δs_p includes contributions from gravitational deflections induced by masses close to the raypath, and from radio source position and structure uncertainties. For spacecraft measurements, Δs_p also contains position departures from those given by a priori trajectories. In this analysis, solar deflection has been modeled in the a priori estimates of the geometric delay. The gravitational deflection signatures contributing to $\Delta\tau_g$, and therefore to Δs_p, are due only to Jupiter. The accuracy of the solar deflection modeling will be discussed in Subsection V.B. Special relativistic effects have also been modeled. In the differences between Δs_p determined from two sessions at the same sidereal time, stationary position and structure errors largely cancel, while the changes in gravitational signatures between the two sessions remain. In this natural source demonstration, the differences in gravitational signatures between the March and April sessions mimic unmodeled differential spacecraft motion. The set of measured differences in Δs_p are the final result of this demonstration. It will be shown below that these differences, $\delta\Delta s_p$, arise primarily from (1) the difference in the strength of Jovian gravitational deflection between the two sessions and (2) modeled stochastic errors.

[1] The term *residual* used throughout this article means the difference between the measured value of a quantity and an a priori estimate.

According to general relativity, the contribution to the terrestrially measured $\Delta\tau_g$ due to the curvature of spacetime around a spherically symmetric body of mass M is approximately

$$\Delta\tau_{g,rel} = \frac{2MG}{c^3} \ln\frac{r_1 + \vec{r}_1 \cdot \hat{k}}{r_2 + \vec{r}_2 \cdot \hat{k}} \quad (2)$$

where G is the gravitational constant equal to 6.67×10^{-8} $cm^3/gm{\cdot}s^2$, $\vec{r}_i$ is the vector pointing from the center of the gravitating body to the ith station of the baseline, and $\hat{k}$ is the unit vector pointing from the center of the gravitating body to the radio source. A term which differentially cancels between two observations at equal sidereal times has been dropped. This term, which compensates for the coordinate time of flight increment due to the body's field at the Earth (e.g., [5,6]), is less than one picosecond for Jupiter, even without differential cancellation.

In the next section, the experimental approach and details will be given. Sections III and IV describe the analysis procedures and the results, respectively. Section V enumerates unmodeled error sources, and Section VI contains conclusions and future directions.

II. Experimental Approach

A. The Local Reference Frame VLBI Technique

In the local reference frame technique, the residual geometric delay of a target radio source is inferred from VLBI observations of the target and several reference sources. For the target observation, the *measured* residual delay between the two antennas, $\Delta\tau$, contains the desired geometric contribution, $\Delta\tau_g$, along with other unwanted delay effects caused by errors in clock, Earth rotation, and troposphere characterizations. The basic idea of the VLBI technique used here is, for each of the target observations in a single experiment, to first separate the $\Delta\tau_g$ component of $\Delta\tau$ from other contributions to the measured target interferometric delay. From the set of $\Delta\tau_g$'s, a set of apparent angular deflections, Δs_p's, are inferred, as indicated by Eq. (1). This procedure is then repeated for a later experiment and the Δs_p determinations at equal sidereal times are differenced to form the final result, a set of differential apparent positions, $\delta\Delta s_p$'s. In order to understand the extraction of the target residual geometric delay from the measured residual delay, the latter is expressed in terms of the residual geometric delay and other residual delays due to the effects mentioned above:

$$\Delta\tau = \Delta\tau_g + \Delta\tau_{epoch} + \Delta\tau_{rate} + \Delta\tau_{rot1} + \Delta\tau_{rot2} + \Delta\tau_{trop1} + \Delta\tau_{trop2} + \epsilon \quad (3$$

In Eq. (3), $\Delta\tau_{epoch}$ and $\Delta\tau_{rate}$ are the delay errors associated with offsets in the clock epochs and rates between the two stations of the VLBI baseline. Delay errors due to uncalibrated Earth rotations about each of two axes orthogonal to the baseline are represented by $\Delta\tau_{rot1}$ and $\Delta\tau_{rot2}$. The errors due to static tropospheric delays at each station are represented by $\Delta\tau_{trop1}$ and $\Delta\tau_{trop2}$ for Station 1 and Station 2, respectively. The quantity ϵ represents all other errors not explicitly included. Associating one parameter with each explicit term in Eq. (3) means that a minimum of seven observations, one target and six reference scans are needed to estimate all indicated error parameters as well as the actual source shift. An expression similar to Eq. (3) can be written for each reference source scan, with $\Delta\tau_g$ equal to zero. The geocentric delays are set to zero anticipating that stationary reference radio source position and structure uncertainties, which can each be as large as 5 nrad [7,8], will have identical contributions to reference residual geometric delays at corresponding epochs for each session. As with target observations, solar deflection effects have been modeled and removed from reference source delays (see Subsection V.B). By using Eq. (3) and the analogous equations for the reference source observations, the target $\Delta\tau_g$ is extracted from the measured target and reference $\Delta\tau$'s. As noted above, repeating the observation sequence in a later session enables the extraction of $\delta\Delta\tau_{g,rel}$, the change in relativistic delay between observation sessions. This is actually accomplished by differencing Δs_p determinations extracted for the target during each session. An equation analogous to Eq. (3) can be written for the residual delay rates [1]. The temporally differential local reference frame technique can be regarded as an extension of the technique used to measure gravitational bending in [9] with one important distinction: Spatially differential observation errors are parameterized in terms of explicit physical effects (e.g., Earth rotation), as opposed to a parameterization linear in arc length separations on the sky.

B. Experimental Details

As indicated by Eq. (3), multiple reference sources were used to estimate systematic delay effects for each of two experiments. The near-occultation event was observed from March 20, 1988, 23:30 UT, to March 21, 1988, 03:28 UT, and the entire observing schedule was repeated from

April 2, 1988, 22:39 UT to April 3, 1988, 02:37 UT. The Mark III VLBI data acquisition system [10] was operated at each station in mode A, which entails receiving and recording data from 28 2-MHz channels spread over the rf band. The centroids of the 2-MHz channels spanned two bands (S-band and X-band) for charged particle calibration. The spanned bands were approximately 2285 ±19 MHz and 8450 ±53 MHz. The characteristics of the three antennas used are summarized in Table 1.[2,3]

The target and reference radio sources observed are listed in Table 2. The table shows source coordinates and formal errors,[4] but the errors are preliminary and should be used to gauge relative measurement uncertainties of source coordinates. For a given source, these errors depend on its strength, the number of times it was observed, the baseline–source geometries, and the parameterization of the VLBI delays from which they were estimated in the global reference frame analysis. The true source position accuracy is probably never better than about approximately 2 nrad due to a number of possible systematic effects at that level. The baseline components used in the analysis are given in Table 3.

The observing sequence was divided into ten subsequences, one for each target measurement. In each subsequence, reference sources preceded and followed the target. The subsequences included target observations at the epochs given in Table 5 on results. The target observations at the first, fourth, and eighth epochs were surrounded by six other reference observations. The target observations for all other epochs were preceded and followed each by a single observation of the closest source, P 0202+14. As emphasized in Subsection V.D, in order to minimize error contributions due to radio source position or structure uncertainties, observation schedules were duplicated as much as possible between the March and April sessions. Therefore, if a reference scan was missed on an individual baseline for one epoch of a session, it was deleted from the analysis, for that baseline, at the corresponding epoch in the other session. All target scans were successful. One to two reference source scans were deleted from each baseline. The determination of VLBI delays, rates, target positions, and modeled error parameters are described in the next section.

III. Analysis Procedures

The standard procedures for analyzing VLBI data are discussed in [11,12,13]. In this section, the analysis of the gravitational deflection VLBI data will be discussed, with emphasis on the departures from the standard procedures. The extraction of interferometric delays and target source coordinates is followed by a discussion of the errors modeled in the analysis.

A. Extracting Interferometric Delays

From Eqs. (1), (2), and (3), it can be seen that the interferometric delays $\Delta\tau$ are required to infer the measured $\Delta\tau_{g,rel}$ and the resulting angular deflections Δs_p for each session. The first step in the VLBI processing was to obtain the interferometric group delays for each few-minute scan from the signals recorded at each station for each baseline. Since the integer cycle ambiguity associated with the group delay is much larger than that associated with the phase delay, group delays were used because they require less-accurate a priori information. The signals recorded at the stations included the broadband noise from the radio sources, background noise, and phase calibration tones to measure instrumental stability. Interferometric group delays were determined by cross correlation of the radio source data at the JPL/Caltech Block II processor [14]. Phase calibration signals were also extracted at the correlator. Both interferometric and phase calibration delays were refined with post-correlation fringe-fitting procedures.[5] A combination of phase calibration delays and short-baseline interferometric delays from the DSS 13–15 baseline were used to detect and correct significant instrumental effects in the long-baseline data. The corrected group delays were then used to estimate target source coordinates for the March and April sessions, as described below.

B. Extracting Source Coordinate Shifts by Parameter Estimation

The ultimate product of the data analysis is the set of projected, differenced source coordinate shifts of P 0201+113, $\delta\Delta s_p$, due to $\delta\Delta\tau_{g,rel}$ at each of the ten observation epochs. The projected coordinate shift at each

[2] J. C. Breidenthal, *DSN/Flight Interface Design Handbook 810-5, Rev. D, vol. I, Module VLBI-10*, Table 2 (internal document), Jet Propulsion Laboratory, Pasadena, California, June 1, 1990. See also S. D. Slobin, *Module TCI-10, Rev. E*; R. W. Sniffen, *Module TCI-20, Rev. C*; and S. D. Slobin, *TCI-30, Rev. D* in the same document.

[3] L. J. Skjerve, personal communication regarding the characteristics of DSS 13, Tracking Systems and Applications Section, Jet Propulsion Laboratory, Pasadena, California, February 1991.

[4] O. J. Sovers, personal communication regarding radio source catalog derived from JPL International Radio Interferometric Surveying and Crustal Dynamics Project data, Tracking Systems and Applications Section, Jet Propulsion Laboratory, Pasadena, California, February 1991.

[5] S. T. Lowe, *Theory of Post-Block II VLBI Observable Extraction* (internal document), Jet Propulsion Laboratory, Pasadena, California, March 1992.

epoch was determined by fitting the group delays and phase delay rates for all target and reference source observations to a set of projected source coordinate parameters for P 0201+113, as well as other parameters describing the errors in Eq. (3). The group delays were more important than the phase delay rates in the fit because, as discussed in the next section, typical delay errors were equivalent to about a 0.78-nrad error in the source coordinate shift per differential observation, while rate errors were equivalent to about a 30-nrad error.

Since VLBI delays are sensitive only to the component of the residual source position that lies along the baseline vector, this projection was estimated at each observation epoch for each session. The projection of the angular shift of P 0201+113 onto the baseline was determined by estimating the shifts of right ascension ($\Delta\alpha$) and declination ($\Delta\delta$) from a priori values. In order to constrain the source coordinate shifts to lie along the baseline projection, $\Delta\alpha$ and $\Delta\delta$ were forced to obey the following equation:

$$\Delta\alpha(t)(\hat{B}_z(t) - \hat{B}(t)\cdot\hat{s}\sin\delta) + \frac{\Delta\delta(t)}{\cos\delta}(\hat{B}(t)\times\hat{s})_z = 0 \quad (4)$$

where the residual right ascension and declination shifts are at epoch t, and $\hat{s}$ is a unit vector in the direction of P 0201+113. The unit baseline vector $\hat{B}$ and its z-component $\hat{B}_z$ must be evaluated in the same celestial frame as $\Delta\alpha$ and $\Delta\delta$. The constraint of Eq. (4) was used in the analysis because MODEST [13], the parameter estimation software used, was set up to estimate $\Delta\alpha$ and $\Delta\delta$. An equivalent approach, which would have required writing new parameter estimation code, would have been to estimate the projected coordinate shifts, $\Delta s_p(t)$, directly, without using the intermediary shifts $\Delta\alpha$ and $\Delta\delta$. Using the existing code, the values of $\Delta s_p(t)$ were then calculated from $\Delta\alpha$ and $\Delta\delta$ according to

$$\Delta s_p(t) = \sqrt{\Delta\alpha^2(t)\cos^2\delta + \Delta\delta^2(t)} \quad (5)$$

In addition to a projected residual source coordinate per epoch per session, at least one parameter per term in Eq. (3) was also estimated for each session. The equations for the partial derivatives of the delay and delay rate with respect to these parameters and source coordinate shifts are described in [13]. A more schematic treatment of the partial derivatives is given in the appendix of [1]. Four sets of station-differential clock epochs and rates were estimated for each session, over intervals of approximately one hour. Earth rotation signatures were assumed to b linear over each session, during which a single set of thre Earth rotation parameters was estimated. The rates c change of the Earth rotation parameters over the 4-hr ses sions were inferred from 5-day differences in Earth rotatio tables [15]. In MODEST, the rotation parameters wer the magnitudes of the standard UT1, polar motion-x, an polar motion-y rotations. Since a single baseline is sensi tive to only two components of Earth rotation, the thre parameters were constrained as follows:

$$\hat{b}_x\Delta\theta_y + \hat{b}_y\Delta\theta_x - \hat{b}_z\Delta\theta_z = 0 \quad (6$$

where $\hat{b}_x$, $\hat{b}_y$, and $\hat{b}_z$ are the components of the unit base line vector in an Earth-fixed frame, and $\Delta\theta_x$, $\Delta\theta_y$, an $\Delta\theta_z$ are the x-pole, y-pole, and UT1 residual rotation magnitudes. The constraint of Eq. (6) ensures that th magnitude of the rotation about the baseline vector tc which the data are insensitive is not estimated. The othe estimated errors indicated by Eq. (3) are the static tropospheric delays at each station. A single zenith delay parameter per station per session was estimated from the data. The stochastic nature of the troposphere was included in the analysis with delay and rate troposphere covariance matrices described in the next subsection.

C. Modeled Stochastic Errors

The errors assigned to the difference between projected source coordinate shifts for the March and April experiments were derived from the standard least-squares formalism (e.g., [16]). These errors in the final results are solely a function of the modeled covariance of the ϵ error term in Eq. (3) and of the partial derivatives relating delay and rate observations to estimated parameters. Modeled stochastic observational errors, assumed to be independent between the two sessions, included only the white VLBI system noise and correlated tropospheric noise. The system noise error for each group delay and phase delay rate was calculated during the delay and rate extraction procedure [17],[6] and is based solely on the number of independent data samples in each scan and the observed correlated amplitude of the radio source over the baseline. Typical system noise errors were about 10–20 picoseconds per observation (0.34–0.68 nrad) for the delay and 1.5×10^{-15} to 3×10^{-15} sec/sec (0.58–1.16 nrad) for the delay rate. The system noise errors for intercontinental baselines using DSS 13 were about twice as large as those using DSS 15, which is consistent with the station characteristics of Table 1.

[6] Ibid.

The delay and rate covariance due to wet tropospheric fluctuations was calculated using the model of Treuhaft and Lanyi [18]. Their model accounts for spatial and temporal correlations of the tropospheric fluctuations. Standard wind speeds of 8 m/sec and wet tropospheric scale heights of 2 km were adopted. Because the model predicts that the tropospheric contribution to the delay rate data far outweighs the system noise contribution, the rate data were used to estimate the level of tropospheric fluctuations for each session. The key model parameter derived from the VLBI rates was the structure constant of the refractivity structure function. This structure function, D_χ, for two points in the atmosphere separated by a distance r, is given by

$$D_\chi(r) = C^2 r^{2/3} \tag{7}$$

For each experiment, normalizing structure constants C, for each station, were chosen to make the reduced χ^2 of a rate-only fit equal to unity. Since the rate data were fit for each experiment separately, and since these data have a very small effect on the final astrometric result, as compared with the delays, this method of normalizing the troposphere covariance is essentially independent of the final differential angular result. It is therefore an a priori method of estimating the error due to tropospheric fluctuations and does not use the consistency of the final result to scale the assigned errors. If the rms rate scatters induced by the troposphere over each of the California stations are assumed to be equal, then the short (approximately 21-km) baseline can be used to infer the level of fluctuation in California. The long baseline rates can then be used to determine the Australia fluctuation contribution. The values of the derived single-station rms delay rates σ_{dr} and structure constants C are given in Table 4 for the two locations.

These structure constants were used to calculate troposphere observation covariance matrices for each baseline, for each session. These matrices were added to the diagonal system noise matrices to form the total observation covariance in the estimation of the Δs_p's. In combining the baselines from a single session, the delay and delay-rate system noise errors from each baseline can be considered independent, but the tropospheric noise is correlated between baselines. The correlation arises because the two baselines share the common Australia station, and to a lesser degree, because the California stations are separated by 21 kilometers. The tropospheric correlations between observations on different baselines were explicitly included in the analysis. Assuming that tropospheric fluctuations in California and Australia are uncorrelated, the troposphere covariance matrix element between the ith VLBI delay of the DSS 13–43 baseline τ_{13-43_i} and the jth delay of the DSS 15–43 baseline τ_{15-43_j} is given by

$$\mathrm{cov}(\tau_{13-43_i}, \tau_{15-43_j}) = \mathrm{cov}(\tau_{43_i}, \tau_{43_j}) + \mathrm{cov}(\tau_{13_i}, \tau_{15_j}) \tag{8}$$

where cov indicates the covariance between the two tropospheric delays in the argument, and τ_{43_i}, for example, is the tropospheric delay at DSS 43 for the ith scan.

The level of error induced by the tropospheric fluctuations, assuming the structure constants in Table 4, was approximately 0.70 nrad for each differential measurement. This combines with an average system noise of about 0.34 nrad to give a total error of about 0.78 nrad per differential observation for the combined baseline system. Because the modeled troposphere covariance downweights delay rate measurements in accordance with the σ_{dr} values in Table 4, the delay rates are much less powerful in determining the final differential angles than the delays. Delay rates alone would have determined angular deflections with accuracies worse than 30 nrad per differential measurement.

IV. Results

The results of the gravitational deflection measurements are shown in Figs. 1(a), 1(b), and 1(c). Figures 1(a) and 1(b) are plots of the measured angular differences, $\delta\Delta s_p$, between the March and April experiments on the baselines between DSS 13 and DSS 43, and between DSS 15 and DSS 43, respectively, as a function of Universal Time for the March experiment. Figure 1(c) shows the combined result for the two baselines, accounting for tropospheric correlations between baselines as described by Eq. (8). The curved lines of Fig. 1 show the prediction of general relativity for Jovian deflection, which can be obtained from Eqs. (1) and (2), and the horizontal, zero-deflection line is shown for reference. The general relativity curve decreases with time largely because the *projection* of the apparent coordinate shift on March 21 decreased as the baseline vector rotated. The unprojected effect decreased by approximately 10 percent over the 4-hour session, due to the decreasing colinearity of P 0201+113, Jupiter, and the Earth. The larger error bars at earlier epochs are due to low-elevation observations in Australia, where tropospheric fluctuations were strongest, as can be seen from Table 4. In order to produce Figs. 1(a) and 1(b), the parameters discussed after Eq. (5) were estimated along with the values of Δs_p. In order to produce

the combined-baseline solution of Fig. 1(c), the identical set of parameters was estimated, with an additional set of clock parameters to account for the independence of the clocks at DSS 13 and DSS 15.

Table 5 shows the results in numerical form. The values of the reduced χ^2, χ^2_ν, of the data about the general relativity and zero-deflection curves are given in Table 6 for each baseline and the combined result. In these χ^2_ν calculations, correlations are included between the values of Δs_p estimated at different epochs within each experiment. The χ^2_ν for the general relativity curves are all within one standard deviation of unity for the χ^2_ν distribution with 10 degrees of freedom. Based on the combined no-Jovian-deflection χ^2_ν of 4.1, that hypothesis is rejected with greater than 99.999-percent confidence, given the error modeling discussed in the previous section. The rms scatter of the combined result about the hypothesis of general relativity was 0.76 nrad. This is a measure of the demonstrated tracking accuracy per data point, or per 25- to 30-minute time interval. The χ^2_ν values for general relativity and no-Jovian deflection were subject to shifts of 0.2 and 0.4, respectively, when plausible modeling variations were tried. An example of one such variation was the alteration of the wind vector direction in the troposphere covariance calculation.

V. Unmodeled Error Contributions

In the next few subsections, the levels of possible errors not included in the estimated or modeled error terms of Eq. (3) will be explored.

A. Stochastic Behavior of Clocks and Earth Rotation

As described in Section III.B, both clock and Earth rotation effects were characterized with linear trends. In the case of the clocks, both the slope and intercept of the linear trend were estimated. Clock epochs and rates were estimated over approximately one-hour periods to remove the stochastic wandering of the hydrogen maser standards used at the DSN stations. This characterization is only an approximation of the actual stochastic clock behavior. Typical DSN clocks have Allan standard deviations between 10^{-15} sec/sec and 10^{-14} sec/sec [19], over time spans of a few thousand seconds. The clock rate parameters, $\Delta\tau_{rate}$ of Eq. (3), estimated from the short-baseline analyses, implied an approximate single-station Allan standard deviation of 5×10^{-15} sec/sec. Simulations were performed to determine the effect of clock behavior characterized by Allan standard deviations with temporal dependences,[7] normalized to the short-baseline results. The simulations showed that deviations from the piecewise linear clock behavior contributes about 0.05 nrad of additional error to angular deflection estimates of Fig. 1(c). This angular error is much smaller than that expected from either system noise or the troposphere, which is why stochastic clock behavior was not modeled.

Earth rotation stochastic behavior was also modeled as linear due to similar reasoning. The variation of Earth rotation offsets during the 4-hour sessions was assumed to be equal to that derived from the published rate over 5-day periods surrounding each epoch [15]. Earth rotation rates over 4-hour periods can differ from those nominal 5-day rates by less than 5 nrad per 4-hour observing schedule. A sensitivity analysis showed that such a departure in Earth rotation rate would contribute less than 0.10 nrad of error to the estimated source coordinate shifts in Fig. 1(c).

B. Solar Deflection Modeling

One possible contribution to the results of Fig. 1 is the incomplete removal of solar gravitational bending. If the solar gravitational deflections from each day, of both target and reference sources, were not removed, the apparent differential angular deflections of P 0201+113 between sessions would have been on the order of 50 nrad. Therefore, accurate solar gravitational deflection modeling is needed to separate the Jovian effect from that of the Sun. In addition to analytically confirming the general relativity calculation in the modeling segment of MODEST [13],[8] an empirical test of the solar deflection model was performed by considering another source in the schedule, P 0202+14, as the target. This source is 4 degrees away from P 0201+113 and from Jupiter and should therefore show no planetary gravitational deflection signature between the two sessions. This test, although consistent with zero deflection for P 0202+14, was a very weak one. The correlations between the deflection results for P 0202+14 and those for P 0201+113 were very high because both results came from essentially the same data. The P 0202+14 solution is mentioned to report that this obvious test was tried, and that it was consistent with accurate solar modeling. However, the analytic check of the MODEST modeling is a much stronger reason to doubt solar gravitational modeling errors. The analytic check showed that the solar modeling was accurate to better than about 3 picoseconds,

[7] J. B. Thomas and R. N. Treuhaft, personal communication regarding the treatment of relativity by MODEST software, Tracking Systems and Applications Section, Jet Propulsion Laboratory, Pasadena, California, October 1990.

[8] Ibid.

or about 0.10 nrad, for a single epoch, which should be regarded as an upper bound for the differential result. The calculated Jovian deflection in Fig. 1 is accurate to a much higher level.

The level of solar deflection for both target and reference sources mentioned above suggests that these data could have been analyzed to determine solar deflection to an accuracy on the order of one percent. As a consistency check, the data set was reanalyzed to solve for the γ parameter of relativistic gravitational theories [20]. Based on the differential solar deflection signatures between the March and April sessions, due to the motion of the Earth about the Sun between those sessions, γ was determined to be equal to 1.011 $\pm$0.036. This value is consistent with general relativity's prediction that γ be equal to unity, and the error is equivalent to a 1.8-percent measurement of solar deflection.

C. Propagation Through Jupiter's Magnetosphere

Another possible mechanism contributing to the measured deflections of Fig. 1 is refractive bending through Jupiter's magnetosphere. Although this error source would not be a concern for demonstrations with sources far from Jupiter, it is described here because it is important in this experiment and could conceivably be of operational importance for a spacecraft in orbit about Jupiter. As noted in Section II.B, charged particle effects were calibrated by observing at S-band (2285 MHz) and X-band (8450 MHz). Magnetic fields or electron columnar contents (electron densities integrated along the direction of propagation) much larger than those found near Earth would require higher order corrections than the simple dual frequency method used here. From Voyager 1 measurements, the magnetic field at 10 Jovian radii, near the equatorial plane, is approximately 4 mG [21]. This field strength is about 100 times smaller than that of the field at the surface of the Earth. The differential field across the baseline is a few tenths of a milligauss. Voyagers 1 and 2 plasma wave measurements indicate that the equatorial electron content, differenced between the ends of a 10,000-km baseline, for a ray at 10 Jovian radii is about 5×10^{15} electrons/m^2 [22]. The presence of these baseline differential magnetic fields and electron columnar contents will cause delay errors equivalent to less than 0.001 nrad in the angular measurement, if the dual frequency approach to charged particle calibration is used.[9]

An empirical determination of the electron content along the line of sight through Jupiter's magnetosphere can be obtained by using the VLBI data from the March and April sessions. In order to investigate the electron columnar content along the line of sight to P 0201+113, Fig. 2 shows the difference in electron columnar content between adjacent scans of P 0201+113 and P 0202+14 versus time, differenced between the March and April sessions. This double difference was formed to look for an increased charged particle columnar content along the line of sight to P 0201+113 during the March session. Differencing the electron content between sources compares the line of sight near Jupiter with one far (4 degrees) from Jupiter, and differencing between sessions reduces geometric effects due to the different slant angles through the ionosphere of the two raypaths. The electron contents of Fig. 2 are derived from the DSS-15 to DSS-43 S- and X-band delays. The electron content at time t, $TEC(t)$, is calculated from the residual group delays at S-band and X-band, $\Delta\tau_S(t)$ and $\Delta\tau_X(t)$ as follows:

$$TEC(t) = \frac{mc}{2\pi e^2}\frac{\omega_S^2\omega_X^2}{\omega_X^2 - \omega_S^2}\left(\Delta\tau_S(t) - \Delta\tau_X(t)\right) \quad (9)$$

where m is the electron mass in grams, c is the vacuum speed of light in centimeters per second, e is the electron charge in statcoulombs, and ω_S and ω_X are the S- and X-band centroid frequencies [23]. From the figure, it can be seen that the temporally differential columnar content differs between the two sources by less than 10^{15} electrons/m^2, or about an order of magnitude lower than the number derived from the literature above. Since the raypath of the target passed 10 Jovian radii to the north of the planet, the baseline differential electron content derived from the equatorial Voyager data should be regarded as an upper bound. In Fig. 2, there may be some ionospheric contribution to the observed electron contents due to the changes in Sun–radio source angles between the two sessions, and a possible change of ionospheric activity between the two sessions. Both of these effects would cause an imperfect cancellation of the geometric ionospheric effect mentioned above and would further lower the differential electron content ascribed to the Jovian magnetosphere. This empirical study of the magnetospheric electron content suggests that the 10-picoradian error derived from the Voyager data is probably an overestimate and not a concern for the error budget of this experiment.

D. Radio Source Position and Structure Errors

There are two classes of radio source position and structure errors: stationary and fluctuating. As has been mentioned, stationary reference radio source position or structure errors identically cancel between sessions in which the

[9] S. T. Lowe, op. cit.

observation sequence is exactly repeated, as long as the equations that relate measured residual delay to parameter shifts are also identical for each session. This cancellation occurs because any delay or delay rate effect that exactly repeats will affect the apparent coordinates of P 0201+113 identically in both sessions. If the observation sequence is changed, or if the analysis equations are changed, even delay and delay-rate position or structure effects which sidereally repeat between schedules will cause errors in the final results [1]. Stationary radio source structure errors can therefore affect the differential results presented here because: (1) Beginning and ending sidereal times for each scan differed slightly between the March and April epochs, and, as indicated in Table 4, (2) a different ratio of tropospheric error to system noise error was ascribed to each session. The effect of stationary reference source position errors on the measured angular deflection of P 0201+113 was calculated by assuming stationary position errors and determining their effect on the final result for each epoch. It was found that reference sources temporally close to the target observation in question were most important and that the incurred error in the target position was on the order of 10 percent of the average stationary errors in those reference sources. By conservatively considering possible systematic errors, an average accuracy of 2.42 nrad was ascribed to the radio source coordinates in Table 2. It follows that errors on the order of 0.24 nrad could result from typical stationary reference source uncertainties. While data editing could have reduced this error further by making the observation sequences more similar to each other, it did not seem warranted for an error of this size. It is also important to note that had the resulting sensitivity to stationary errors been higher, the inequality of the observable troposphere covariance between sessions could have been adjusted. In this set of experiments, using a suboptimal covariance matrix for the troposphere of one experiment to make it more equal to that of the other was deemed unnecessary. In other differential experiments, the error incurred by the inequality of the tropospheric covariance matrix between sessions may be large. In that case, suboptimal tropospheric matrices may be adopted in order to desensitize the differential result to stationary reference source uncertainties; a trade-off between these two error sources must be considered.

The second type of radio source uncertainty is due to fluctuations of apparent source position due to a changing radio source structure, which results in an effective proper motion. From [24], it can be seen that average radio source position shifts due to structure changes are on the order of 5 nanoradians per year. Assuming that the apparent change in position is linear with time, over 13 days, 0.15 nrad of position change per source would be expected. Covariance studies show that these errors would propagate into about 0.10-nrad errors in the P 0201+113 coordinates. Thus, the combined stationary and fluctuating radio source coordinate errors could add about 6 percent to the nominal 0.78-nrad formal errors in this experiment.

E. Geophysical Effects

The geophysical model used to calculate the residual delays of Eq. (3) contains many components which could conceivably contribute to a differential error between the two observation epochs. Delays due to the tides, baseline length uncertainties, ocean loading, and tropospheric mapping functions could change between epochs and must be removed in order to avoid aliasing into the gravitational deflection signature. Each of these effects was estimated with a sensitivity analysis, and they evidenced typical signatures of less than 0.05 nrad apiece. An overall error of 0.10 nrad is assigned to geophysical effects.

F. Summary of Unmodeled Errors

Table 7 summarizes the unmodeled error effects. Added in quadrature, the unmodeled effects could contribute up to 0.32 nrad. The total unmodeled error could therefore add 8 percent to the 0.78-nrad modeled error. However, there is no indication from the χ^2_ν values of Table 6 that unmodeled errors were important in the analysis of these data.

It is worth noting that the accuracy of the ephemeris of Jupiter is not a concern for the analysis of this experiment. Errors in the Jovian ephemeris would change the general relativity curve of Fig. 1, which is obtained with Eqs. (1) and (2). The current ephemeris accuracy of Jupiter is about 200 nrad [25], which causes uncertainties in the theoretical curve of Fig. 1 at the level of less than 0.005 nrad. Conversely, the position of Jupiter in the radio reference frame is very poorly determined by this experiment, relative to the current ephemeris accuracy.

VI. Conclusions

The technique of differential VLBI, over two DSN California–Australia baselines, was used to track the angular deflection of the raypath of P 0201+113 when it passed within 200 arcseconds (approximately 10 Jovian radii) of Jupiter. Two experiments were performed: one at the time of near-occultation and one 13 days later, when the raypath was about 3 deg from Jupiter. The results of Fig. 1(c) yield a χ^2_ν about the hypothesis of Jovian deflection of 0.6,

and are therefore consistent with Jovian gravitational deflection of the raypath. The χ^2_ν about the hypothesis of zero Jovian deflection is 4.1, which rejects that hypothesis at greater than the 99.999-percent confidence level, and suggests that the deflection was successfully tracked at the nanoradian level. The tracking demonstration involved estimating clock, Earth rotation, and tropospheric parameters from observations of a local reference frame of radio sources surrounding P 0201+113. Because the reference sources were many degrees from the target to be tracked, the experiments reported in this article demonstrate a wide-field differential astrometric tracking technique with 0.78-nrad accuracy.

The dominant errors in the measurements were white system noise due to sky and instrument background and correlated tropospheric noise. The effect of the white system noise can be calculated from the number of samples, or bits, cross-correlated in the VLBI processing and the observed correlated amplitude. The effect of the correlated tropospheric errors was assessed with a statistical model of tropospheric fluctuations. This model was normalized separately for each experiment and station by using the phase delay rate data from both the short and long baselines formed by the three stations. The system noise and tropospheric errors contributed about 0.34 nrad and 0.70 nrad to the final 0.78-nrad uncertainty per differential measurement. The values of χ^2_ν derived from the hypothesis of general relativistic, Jovian deflection suggest that these two modeled error sources fully account for observed errors in the final results. However, unmodeled effects due to the stochastic behavior of clocks and Earth rotation, solar deflection mismodeling, charged particles in Jupiter's magnetosphere, radio source position and structure errors, and geophysical effects could add approximately 8 percent to the 0.78-nrad uncertainty.

There are several improvements in observation and calibration strategies which could result in higher measurement accuracies for future wide-field astrometric experiments (for a more complete discussion, see [1]). If the highly accurate phase delay could be used instead of the group delay, the system noise error would be reduced by about two orders of magnitude to about 10 prad. Efforts to resolve phase-delay ambiguities on intercontinental baselines are currently under way. The ultimate precision of this astrometric technique is determined by the system noise level of the phase delay measurement. It is therefore worth considering means for reducing other errors to that level. It is possible that water vapor radiometry [26,27] could be used to calibrate the wet tropospheric fluctuations, the other dominant error source. Refractivity fluctuations in the dry atmosphere will also contribute to the astrometric error. The exact level of the dry fluctuation contribution has not been measured, but it is probably about a factor of 5–10 smaller than the wet fluctuation contributions. Barometric arrays or other radio metric techniques might be useful in calibrating the dry fluctuations. Finally, as noted in Section V, radio source structure fluctuations are important for differential measurements, made within a few weeks of each other, at the level of 0.10 nrad. For differential measurements made over longer time periods, the structure errors will exceed the 0.10-nrad effects reported in this article. If all other error sources have been reduced to the approximately 10-prad level, it may become necessary to account for radio source structure fluctuations, even over periods as short as a few days. In addition to mapping reference and target radio sources over time, another possibility, as yet unexplored, is to add structure parameters to Eq. (3) and estimate time-varying structure effects directly from the astrometric VLBI observations. This possibility will be explored in the Advanced System Program next fiscal year.

An interesting byproduct of this tracking technique is the measurement of solar gravitational deflection. Because the solar gravitational effect on the data presented here was on the order of 50 nrad (the target was about 25 degrees from the Sun), this experiment constituted a 2-percent solar gravitational measurement. If the target were a few degrees from the Sun, the solar deflection could be measured, with better than 0.1-percent accuracy, with two experiments of the duration of those reported here. Solar plasma fluctuations would be a possible obstacle, but with sufficient signal strength, it may be possible to freeze the solar plasma fluctuations on short time scales in conjunction with dual frequency calibration. If successful, repeated measurements near the Sun could begin to improve on the state-of-the-art measurement of the post-Newtonian γ parameter of gravity theories [28]. Proof-of-concept solar deflection experiments are currently supported by the TDA Science Office.

This demonstration was done with a natural radio source target, P 0201+113. A nanoradian-tracking demonstration on Galileo had been planned before the high-gain antenna availability became an issue. Unless high-gain transmission is restored, spacecraft system noise and charged particle errors incurred with the low-gain S-band downlink will make a 1-nrad Galileo demonstration virtually impossible. A key difference between the spacecraft and natural source targets is in the bandwidth of the received signals; spacecraft transmit tones, while natural sources are broadband in nature. Dispersive phase effects in the receiving electronics may therefore affect the target spacecraft and natural reference source signals differ-

ently. Studies of the dispersive nature of the DSN receiving electronics[10] will be necessary before demonstrating nanoradian accuracy on future spacecraft, such as Cassini.

Applications of nanoradian accuracy include Jovian ephemeris development on approach for Galileo, ring analysis at Saturn for Cassini, and aerocapture approach trajectory optimization for missions to Mars. Temporally differential results have been presented in this article. While temporally differential measurements frequently yield useful navigation products (for example, Jovian ephemeris development on approach), absolute positions relative to solar system bodies are also needed. It should be noted that radio source structure at the 5-nrad level and planetary position errors in the radio frame as high as 200 nrad limit some classes of body-relative measurements. Temporally differential high-accuracy results in the radio frame, however, are necessary precursors to nanoradian body-relative demonstrations. Temporally differential demonstrations of the sort reported here address a large subset of the body-relative errors, namely system noise and atmospheric fluctuations; in fact, analysis of high-accuracy temporally differential measurements of planetary orbiters can locate those bodies in the radio frame.

[10] C. Edwards and K. Zukor, "Video Converter Local Oscillator Stability for Block I and Block II VLBI," JPL Interoffice Memorandum 335.1-90-055 (internal document), Jet Propulsion Laboratory, Pasadena, California, October 30, 1990.

References

[1] R. N. Treuhaft, "Deep Space Tracking in Local Reference Frames," *TDA Progress Report 42-94*, vol. April–June 1988, Jet Propulsion Laboratory, Pasadena, California, pp. 1–15, August 15, 1988.

[2] B. K. Trinkle and S. M. Lichten, "Differential Very Long Baseline Interferometry for 50-nanoradian Deep Space Navigation," *Astrodynamics 1985*, Proceedings of the Conference, Part 2 (A86-43201 20-12), edited by Kaufman et al., Univelt, Inc.: San Diego, California, pp. 1257–1276, 1986.

[3] R. N. Treuhaft and B. K. Trinkle, "High Precision Navigation Demonstration With Extragalactic Radio Sources," *JPL Highlights*, Jet Propulsion Laboratory, Pasadena, California, pp. 62–63, 1986.

[4] W. M. Folkner, "Navigational Utility of High-Precision Radio Interferometry for Galileo's Approach to Jupiter," *TDA Progress Report 42-102*, vol. April–June 1990, Jet Propulsion Laboratory, Pasadena, California, pp. 34–46, August 15, 1990.

[5] M. H. Soffel, J. Muller, X. Wu, and C. Xu, "Consistent Relativistic VLBI Theory with Picosecond Accuracy," *Astronomical Journal*, vol. 101, pp. 2306–2310, June 1991.

[6] B. Shahid-Saless, R. W. Hellings, and N. Ashby, "A Picosecond Accuracy Relativistic VLBI Model via Fermi Normal Coordinates," *Geophysical Research Letters*, vol. 18, pp. 1139–1142, June 1991.

[7] O. J. Sovers, C. D. Edwards, C. S. Jacobs, G. E. Lanyi, K. M. Liewer, and R. N. Treuhaft,"Astrometric Results of the 1978–1985 Deep Space Network Radio Interferometry—The JPL 1987-1 Extragalactic Source Catalog," *Astronomical Journal*, vol. 95, pp. 1647–1658, June 1988.

[8] P. Charlot, "Radio-Source Structure in Astrometric and Geodetic Very Long Baseline Interferometry," *Astronomical Journal*, vol. 99, pp. 1309–1326, April 1990.

[9] E. B. Fomalont and R. A. Sramek, "A Confirmation of Einstein's General Theory of Relativity by Measuring the Bending of Microwave Radiation in the Gravitational Field of the Sun," *Astrophysical Journal*, vol. 199, pp. 749–755, August 1, 1975.

[10] A. E. E. Rogers, R. J. Cappallo, H. F. Hinteregger, J. I. Levine, E. F. Nesman, J. C. Webber, A. R. Whitney, T. A. Clark, C. Ma, J. W. Ryan, B. E. Corey, C. C. Counselman, T. A. Herring, I. I. Shapiro, C. A. Knight, D. B. Shaffer, N. R. Vandenberg, R. Lacasse, R. Mauzy, B. Rayhrer, B. R. Schupler, and J. C. Pigg, "Very-Long Baseline Radio Interferometry—The Mark III System for Geodesy, Astrometry, and Aperture Synthesis," *Science*, vol. 219, pp. 51–54, January 7, 1983.

[11] I. I. Shapiro, "Estimation of Astrometric and Geodetic Parameters," in *Methods of Experimental Physics, Part C*, edited by M. L. Meeks, Orlando, Florida: Academic Publishers, pp. 261–276, 1976.

[12] T. A. Herring, "Precision and Accuracy of Intercontinental Distance Determinations Using Radio Interferometry," Ph.D. thesis, Hanscom Air Force Base, Massachusetts: Air Force Research Cambridge Laboratories, July 1983.

[13] O. J. Sovers, *Observation Model and Parameter Partials for the JPL VLBI Parameter Estimation Software MODEST/1991*, JPL Publication 83-39, Rev. 4, Jet Propulsion Laboratory, Pasadena, California, August 1, 1991.

[14] J. B. Thomas, *Interferometry Theory for the Block 2 Processor*, JPL Publication 87-29, Jet Propulsion Laboratory, Pasadena, California, October 15, 1987.

[15] "IRIS Bulletin A (International Radio Interferometric Surveying)," *Earth Orientation Bulletin A*, vol. 53, 1988.

[16] W. C. Hamilton, *Statistics in Physical Science*, New York: Ronald Press, p. 124, 1964.

[17] A. R. Thompson, J. M. Moran, and G. W. Swenson, *Interferometry and Synthesis in Radio Astronomy*, New York: Wiley and Sons, 1986.

[18] R. N. Treuhaft and G. E. Lanyi, "The Effect of the Dynamic Wet Troposphere on Radio Interferometric Measurements," *Radio Science*, vol. 22, March–April, pp. 251–265, 1987.

[19] P. F. Kuhnle, "NASA/JPL Deep Space Network Frequency and Timing," *Proceedings of the Twenty-First Annual Precise Time and Time Interval Applications and Planning Meeting*, Redondo Beach, California, pp. 479–489, 1989.

[20] C. M. Will and K. Nordtvedt, Jr., "Conservation Laws and Preferred Frames in Relativistic Gravity, Part I: Preferred-Frame Theories and an Extended PPN Formalism and Part II: Experimental Evidence To Rule Out Preferred-Frame Theories of Gravity," *Astrophysical Journal*, vol. 177, pp. 757–792, 1972.

[21] N. F. Ness, M. H. Acuna, R. P. Lepping, L. F. Burlaga, K. W. Behannon, and F. M. Neubauer, "Magnetic Field Studies at Jupiter by Voyager 1—Preliminary Results," *Science*, vol. 204, pp. 982–987, June 1, 1979.

[22] D. A. Gurnett, F. L. Scarf, W. S. Kurth, R. R. Shaw, and R. L. Poynter, "Determination of Jupiter's Electron Density Profile From Plasma Wave Observations," *Journal of Geophysical Research*, vol. 86, no. A10, pp. 8199–8212, 1981.

[23] J. D. Jackson, *Classical Electrodynamics*, New York: Wiley and Sons, pp. 288–289, 1975.

[24] R. W. Porcas, "Summary of Known Superluminal Sources," in *Superluminal Radio Sources*, edited by J. A. Zensus and T. J. Pearson, New York: Cambridge University Press, pp. 12–25, 1987.

[25] A. E. Niell, X X Newhall, R. A. Preston, G. L. Berge, D. O. Muhleman, D. J. Rudy, J. K. Campbell, P. B. Esposito, and E. M. Standish, "Relating the Planetary Ephemerides and the Radio Reference Frame," *TDA Progress Report 42-81*, vol. January–March 1985, Jet Propulsion Laboratory, Pasadena, California, pp. 1–8, May 15, 1985.

[26] G. M. Resch, D. E. Hogg, and P. J. Napier, "Radiometric Correction of Atmospheric Path Length Fluctuations in Interferometric Experiments in Radio Astronomy," *Radio Science*, vol. 19, January–February, pp. 411–422, 1984.

[27] G. Elgered, J. L. Davis, T. A. Herring, and I. I. Shapiro, "Geodesy by Radio Interferometry—Water Vapor Radiometry for Estimation of the Wet Delay," *Journal of Geophysical Research*, vol. 96, pp. 6541–6555, April 10, 1990.

[28] R. D. Reasenberg, I. I. Shapiro, P. E. MacNeil, R. B. Goldstein, J. C. Breidenthal, J. P. Brenkle, D. L. Cain, T. M. Kaufman, T. A. Komarek, and A. I. Zygielbaum, "Viking Relativity Experiment—Verification of Signal Retardation by Solar Gravity," *Astrophysical Journal, Part 2—Letters to the Editor*, vol. 234, pp. L219–l221, December 15, 1979.

Table 1. Characteristics of the DSN antennas used in the gravitational deflection measurements.

Antenna	Location	Diameter, m	System temperature, K	Efficiency, percent
DSS 13	Goldstone	26	30	45
DSS 15	Goldstone	34	20	72
DSS 43	Australia	70	20	66

Table 2. Target and reference radio sources used in the gravitational deflection measurements.

Source name	Right ascension			Declination			Right ascension error, msec	Declination error, mas
	hr	min	sec	deg	min	sec		
P 0201+113	02	03	46.65701	11	34	45.4107	0.03	0.6
P 0019+058	00	22	32.44122	06	08	04.2692	0.03	0.9
P 0106+01	01	08	38.77107	01	35	0.3179	<0.01	0.1
GC 0119+04	01	21	56.86167	04	22	24.7347	<0.01	0.2
CTD 20	02	37	52.40567	28	48	08.9904	<0.01	<0.1
GC 0235+16	02	38	38.93011	16	36	59.2750	0.01	0.1
OD 166	02	42	29.17090	11	01	00.7275	0.03	0.4
3C 454.3	22	53	57.74793	16	08	53.5610	<0.01	0.1

Table 3. California–Australia DSN baseline vectors.

Baseline	X, m	Y, m	Z, m	Length, m
DSS 13–43	−2109765.511	7337838.348	−7335705.773	10588085.819
DSS 15–43	−2107355.861	7324010.785	−7351418.891	10588930.183

Table 4. Single-station delay rates and tropospheric refractivity structure function constants.

Session	σ_{dr} – California, psec/sec	σ_{dr} – Australia, psec/sec	C – California, $10^{-7}m^{-1/3}$	C – Australia, $10^{-7}m^{-1/3}$
March 21, 1988	0.023	0.086	0.46	1.70
April 2, 1988	0.025	0.060	0.41	0.99

Table 5. Measured angular deflections of P 0201 + 113 between March 21 and April 2, 1988.

Observation time past March 21, 1988			DSS 13–43 measured deflection, nrad	DSS 15–43 measured deflection, nrad	Combined measured deflection, nrad	Expected deflection, nrad
hr	min	sec				
−00	19	40	2.63 ± 2.09	−0.42 ± 2.39	1.44 ± 1.69	3.21
00	22	14	2.29 ± 1.27	3.72 ± 0.90	3.52 ± 0.84	2.49
00	39	14	3.25 ± 1.18	2.85 ± 0.82	3.01 ± 0.78	2.21
00	56	17	1.00 ± 1.06	2.10 ± 0.76	1.86 ± 0.72	1.95
01	36	40	3.25 ± 1.28	0.99 ± 0.90	1.67 ± 0.84	1.38
01	54	05	3.10 ± 1.33	−0.05 ± 0.66	0.49 ± 0.63	1.18
02	11	06	0.75 ± 0.94	1.01 ± 0.65	1.06 ± 0.61	1.00
02	28	05	1.62 ± 0.99	−0.16 ± 0.65	0.35 ± 0.61	0.85
03	07	05	−0.79 ± 1.11	0.44 ± 0.66	0.28 ± 0.62	0.61
03	24	04	−1.51 ± 1.41	0.79 ± 0.77	0.47 ± 0.71	0.55

Table 6. The χ^2_ν values for general relativity and no Jovian deflection hypotheses.

Hypothesis	DSS 13–43 χ^2_ν	DSS 15–43 χ^2_ν	Combined χ^2_ν
General relativity	1.0	1.1	0.6
No Jovian deflection	2.8	3.7	4.1

Table 7. Unmodeled error contributions to $\delta\Delta s_p$.

Unmodeled effect	Error, nrad
Clock stochastics	0.05
Earth rotation stochastics	0.10
Solar deflection	0.10
Magnetosphere propagation	<0.01
Stationary source position/structure	0.24
Fluctuating source structure	0.10
Geophysical effects	0.10
Root-Sum-Square	0.32

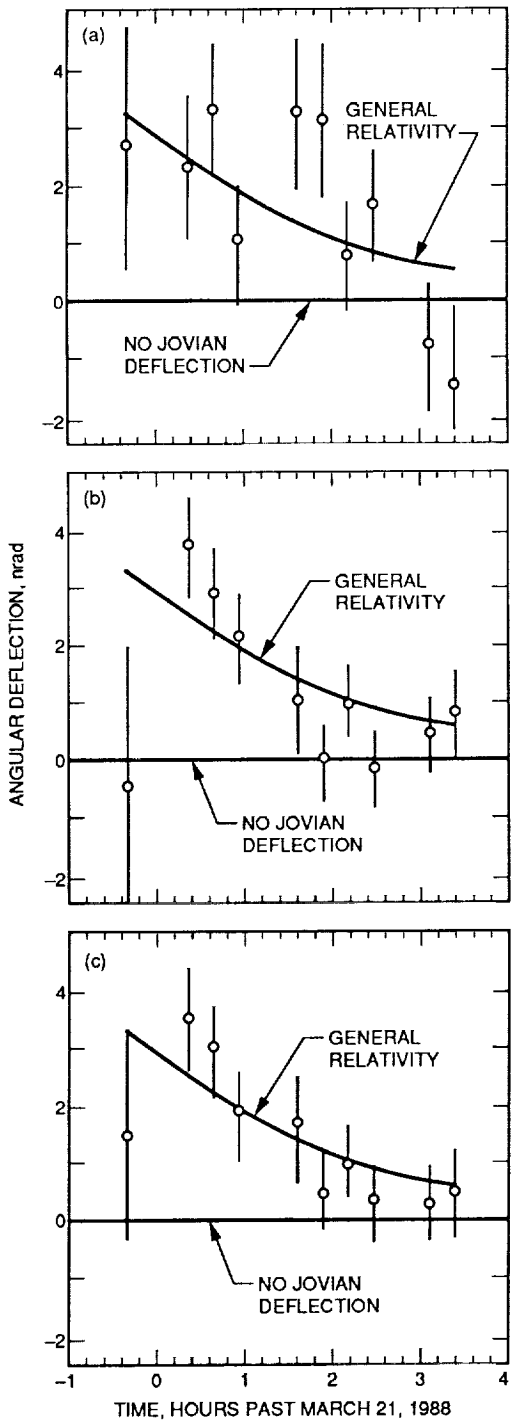


Fig. 1. The measured angular deflections, versus time, of the target radio source P 0201+113 between the March 21, 1988, and April 2, 1988, sessions for the DSS 13–43, DSS 15–43, and combined baselines, respectively. The curve in each figure is the baseline-projected deflection, versus time, caused by the changes in the target raypaths' proximity to Jupiter between sessions, according to general relativity.

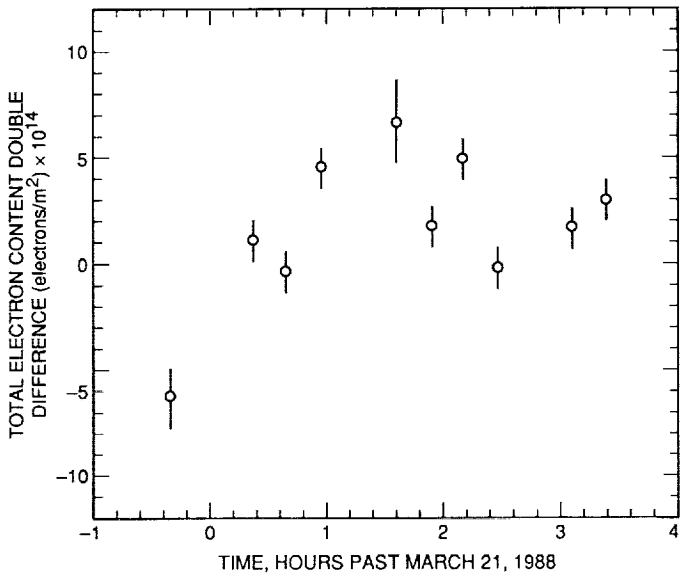


Fig. 2. The difference in measured electron columnar content between adjacent scans of the target, P 0201+113, and the closest reference source, P 0202+14, versus time, differenced between the March and April sessions. The columnar contents were inferred from the dual frequency VLBI data to investigate charged particle propagation effects through Jupiter's magnetosphere.

Deep-Space Navigation With Differenced Data Types Part III: An Expanded Information Content and Sensitivity Analysis

J. A. Estefan and S. W. Thurman
Navigation Systems Section

An approximate six-parameter analytic model for Earth-based differenced range measurements is presented and is used to derive a representative analytic approximation for differenced Doppler measurements. The analytical models are tasked to investigate the ability of these data types to estimate spacecraft geocentric angular motion, Deep Space Network station oscillator (clock/frequency) offsets, and signal-path calibration errors over a period of a few days, in the presence of systematic station location and transmission media calibration errors. Quantitative results indicate that a few differenced Doppler plus ranging passes yield angular position estimates with a precision on the order of 0.1 to 0.4 μrad, and angular rate precision on the order of 10 to 25 × 10^{-12} rad/sec, assuming no a priori information on the coordinate parameters. Sensitivity analyses suggest that troposphere zenith delay calibration error is the dominant systematic error source in most of the tracking scenarios investigated; as expected, the differenced Doppler data were found to be much more sensitive to troposphere calibration errors than differenced range. By comparison, results computed using wideband and narrowband ΔVLBI under similar circumstances yielded angular precisions of 0.07 to 0.4 μrad, and angular rate precisions of 0.5 to 1.0 × 10^{-12} rad/sec.

I. Introduction

The analysis described herein represents a follow-up study to two recent work efforts, each separately describing the information content of differenced (two-way minus three-way) range and Doppler radio metric data [1,2]. In these earlier studies, systematic measurement errors induced by observing platform uncertainties, such as baseline coordinate and Earth orientation errors, were not considered. Also excluded from those analyses was an assessment of the effects of transmission media (ionosphere and troposphere) calibration errors on the data. Furthermore, the mathematical models for approximating the differenced range and Doppler measurements were based on the assumption that spacecraft geocentric angular coordinates remained constant over time—a reasonable assump

tion given that the performance characteristics of these data types were investigated for a single tracking pass alone.

In this analysis, the information content of several tracking passes is investigated, with the spacecraft angular coordinates assumed to vary linearly with time. What follows is a detailed derivation of a six-parameter differenced range and Doppler observable model, which is used to assess the performance of these data types under a variety of tracking scenarios. Despite the fact that realistic navigation operations scenarios are not investigated here, due to the relatively short data arc lengths assumed, the selected station (baseline) combinations, and the absence of line-of-sight data such as two-way Doppler or range, the resulting analysis does provide some useful insight into the merit and potential of the differenced data types for navigational purposes. Recall that these "quasi-VLBI" techniques have some operational advantages over the wideband and narrowband ΔVLBI techniques of delta differenced one-way range (ΔDOR) and delta differenced one-way Doppler (ΔDOD) in that the differenced *quasar-less* data can be acquired without interruption of spacecraft command and telemetry activities—a characteristic that may prove invaluable during periods of the approach phase preceding planetary encounters or spacecraft maneuvers. Despite the operational shortcomings of ΔDOR and ΔDOD, it must be acknowledged that they are, for the most part, self-calibrating data types and are therefore less dependent upon accurate externally supplied calibrations of various potential error sources.

II. Observable Models

The mathematical models presented here account for effects due to observing platform and transmission media errors on the differenced data types. As stated previously, several tracking passes are assumed to be acquired, therefore, the angular coordinates are taken to vary over time. Over a period of a few days, the angular motion of an interplanetary spacecraft is nearly linear, hence, the angular rate coordinates of the spacecraft are assumed to be constants for this analysis. Although these assumptions do not significantly affect the formulation of the original differenced range observable arrived at in [1], they do impact the differenced Doppler model; consequently, the observation partial derivatives required for information content and sensitivity analyses become more involved computationally.

A. General Expressions for the Observables

The approximate differenced range observable model is taken to be

$$\Delta\rho \approx \Delta\rho_{\mathrm{G}} + \tau_{\mathrm{tro}} + \tau_{\mathrm{ion}} + \tau_{\mathrm{clk}} \tag{1}$$

where

$\Delta\rho_G$ = differenced range term based purely on geometry

τ_{tro} = delay due to static troposphere calibration errors

τ_{ion} = delay due to static ionosphere calibration errors

τ_{clk} = delay due to station clock and frequency offset and signal-path calibration errors

All delay terms are assumed to be in distance units (i.e., the speed of light factor is dropped for notational convenience). From this formulation, an approximate differenced range-rate observable, proportional to the differenced Doppler observable, follows directly via a time-derivative of Eq. (1), yielding

$$\Delta\dot{\rho} \approx \Delta\dot{\rho}_G + \dot{\tau}_{tro} + \dot{\tau}_{ion} + \dot{\tau}_{clk} \tag{2}$$

in which

$\Delta\dot{\rho}_G$ = differenced Doppler geometric term

$\dot{\tau}_{tro}$ = delay-rate due to the troposphere calibration errors

$\dot{\tau}_{ion}$ = delay-rate due to the ionosphere calibration errors

$\dot{\tau}_{clk}$ = delay-rate due to station and frequency offset calibration errors.

B. Differenced Range and Doppler Geometric Terms

An analytic expression for the differenced range geometric value can be attained by considering the illustrations provided in Figs. 1 and 2. (Fig. 3 illustrates the differenced Doppler and range data acquisition scheme.) From these figures, it is seen that the differenced range geometric term can be expressed as

$$\Delta\rho_G = \mathbf{B} \cdot \left(\frac{\mathbf{r}}{r}\right) = r_B \cos\delta \cos H_B + z_B \sin\delta \tag{3}$$

where

$\mathbf{B}$ = baseline vector between co-observing deep-space stations

$\mathbf{r}$ = spacecraft geocentric position vector

r_B = baseline component normal to the spin axis of Earth

z_B = baseline component parallel to the spin axis of Earth

H_B = baseline hour angle, $\alpha_B - \alpha$

α_B = baseline right ascension, $\alpha_g + \lambda_B$

α_g = Greenwich right ascension

λ_B = baseline longitude

α = spacecraft right ascension

δ = spacecraft declination

The cylindrical baseline coordinates can be expressed as functions of the individual station coordinates, which are defined in Fig. 2, as follows:

$$\left.\begin{aligned} r_B &= \sqrt{(r_{s_1} + r_{s_2})^2 - 2r_{s_1}r_{s_2}\left[1 + \cos(\lambda_1 - \lambda_2)\right]} \\ z_B &= z_{h_1} - z_{h_2} \\ \lambda_B &= \tan^{-1}\left(\frac{r_{s_1}\sin\lambda_1 - r_{s_2}\sin\lambda_2}{r_{s_1}\cos\lambda_1 - r_{s_2}\cos\lambda_2}\right) \end{aligned}\right\} \quad (4)$$

where

r_{s_1}, r_{s_2} = station distances from Earth's spin axis

z_{h_1}, z_{h_2} = station distances (height) from Earth's equator

λ_1, λ_2 = station longitudes

Conservative values of station location and baseline coordinate data for three representative DSN stations and their associated baselines are provided in Table 1.[1]

The spacecraft geocentric angular coordinates are approximated by a Taylor series expansion about some reference epoch, t_0

$$\delta = \delta_0 + \dot{\delta}_0(t - t_0) + \cdots$$

$$\alpha = \alpha_0 + \dot{\alpha}_0(t - t_0) + \cdots$$

where

δ_0, α_0 = spacecraft declination and right ascension epoch

$\dot{\delta}_0, \dot{\alpha}_0$ = spacecraft epoch declination rate and epoch right ascension rate

$t - t_0$ = elapsed time from epoch

Higher order terms are not modeled.

A time derivative of Eq. (3) yields the analytic differenced range-rate (Doppler) geometric value

$$\Delta\dot{\rho}_G = -r_B(\dot{H}_B \cos\delta \sin H_B + \dot{\delta}\sin\delta\cos H_B) + z_B\dot{\delta}\cos\delta \quad (5)$$

where

$\dot{H}_B$ = time rate of change of the baseline hour angle

$= \omega - \dot{\alpha}_0$

ω = Earth's rotation rate

$\dot{\delta} = \dot{\delta}_0$

C. Static Troposphere Delay and Delay Rate

Sophisticated empirical models have been developed for tropospheric path delay effects [3].[2] For this study, a much simpler model is used, but one which is still accurate to within about 10 percent of the actual static troposphere path delay for the station–spacecraft elevation angle range of interest ($\geq$10 deg).

A simple troposphere delay model which yields results commensurate with the more complicated empirical models, for elevation angles in excess of about 5 deg, is given by

$$\tau_{tro} = \frac{\tau_{z_{tro}}}{\sin\gamma} \quad (6)$$

where

$\tau_{z_{tro}}$ = zenith troposphere delay

γ = station–spacecraft topocentric elevation angle

[1] T. D. Moyer, "Station Location Sets Referred to the Radio Frame," JPL Interoffice Memorandum 314.5-1334 (internal document), Jet Propulsion Laboratory, Pasadena, California, February 24, 1989.

[2] R. K. Russell, "Computation of Troposphere Partial Derivatives," JPL Technical Memorandum 391-277 (internal document), Jet Propulsion Laboratory, Pasadena, California, February 3, 1972.

In Eq. (6), the zenith delay term, $\tau_{z_{tro}}$, is assumed to represent the *total* zenith troposphere delay. In the empirical models, the delay is typically broken down into separate *wet* and *dry* components.

The model used for differenced range measurements, using Eq. (6), is then given by

$$\tau_{tro} = \left(\frac{\tau_{z_{tro}}}{\sin \gamma}\right)_{sta_1} - \left(\frac{\tau_{z_{tro}}}{\sin \gamma}\right)_{sta_2} \tag{7}$$

The subscripts sta_i, for $i = 1$ and 2, are used to denote each participating tracking station.

A useful modification of Eq. (7) is to express $\sin \gamma$ for each station as a function of spacecraft declination and individual station hour angle, which is accomplished by the following relation:

$$\sin \gamma = \frac{1}{r_{sta}} (r_s \cos \delta \cos H + z_h \sin \delta) \tag{8}$$

where

r_s = station spin radius

z_h = station z-height

H = station hour angle

$r_{sta} = \sqrt{r_s^2 + z_h^2}$

This expression is quite useful when seeking to derive the partial derivatives needed for the sensitivity analyses described later in this article. It should also be noted that each individual station hour angle does not have to be explicitly known as it can be extracted from the baseline hour angle. Let

$$H_2 = H_B - \varphi \tag{9}$$

where

H_2 = hour angle of station 2

φ = baseline longitude relative to station 2

$= \lambda_B - \lambda_2$

Then

$$H_1 = H_2 - (\lambda_2 - \lambda_1) \tag{10}$$

Deriving the troposphere delay-rate model simply requires a time derivative of Eq. (7), thereby yielding

$$\dot{\tau}_{tro} = \left(\frac{\partial \tau_{tro}}{\partial \gamma} \dot{\gamma}\right)_{sta_1} - \left(\frac{\partial \tau_{tro}}{\partial \gamma} \dot{\gamma}\right)_{sta_2} \tag{11a}$$

in which

$$\frac{\partial \tau_{tro}}{\partial \gamma} = \frac{-\tau_{z_{tro}} \cos \gamma}{\sin^2 \gamma} \tag{11b}$$

Differentiating Eq. (8) with respect to time, and solving for $\dot{\gamma}$, gives

$$\dot{\gamma} = \frac{1}{r_{sta} \cos \gamma} \left[-r_s \left(\dot{\delta} \sin \delta \cos H + \dot{H} \cos \delta \sin H\right) + z_h \dot{\delta} \cos \delta\right] \tag{12}$$

D. Static Ionosphere Delay and Delay Rate

In the previous subsection, it was argued that a very simple, yet accurate, approximation can be used to represent the static troposphere path delay and delay rate. Unfortunately, the behavior of the static ionosphere delay does not lend itself to approximation so easily, as it is dependent upon the location and movement of the Sun with respect to the station–spacecraft line of sight, as well as other elevation-independent parameters [4]. A simple model does exist, however, which approximates the behavior of ionosphere delay as a function of elevation for an "average" homogeneous ionosphere, and it is this model that is utilized for this study.[3]

The ionosphere delay model associated with the differenced range data type is taken to be

$$\tau_{ion} = \left(\frac{A \tau_{z_{ion}}}{B + \sin \gamma}\right)_{sta_1} - \left(\frac{A \tau_{z_{ion}}}{B + \sin \gamma}\right)_{sta_2} \tag{13}$$

where

$A = 1.15$, $B = 0.15$ } empirical constants that are derived from ray-tracing methods

[3] R. K. Russell, "Computation of Ionosphere Partial Derivatives," JPL Technical Memorandum 391-291 (internal document), Jet Propulsion Laboratory, Pasadena, California, February 29, 1972.

$\tau_{z_{ion}}$ = zenith ionosphere delay

γ = station–spacecraft topocentric elevation angle

The ionosphere delay-rate model for the differenced Doppler is arrived at by differentiating Eq. (13) with respect to time, which gives

$$\dot{\tau}_{ion} = \left(\frac{\partial \tau_{ion}}{\partial \gamma}\dot{\gamma}\right)_{sta_1} - \left(\frac{\partial \tau_{ion}}{\partial \gamma}\dot{\gamma}\right)_{sta_2} \tag{14a}$$

in which

$$\frac{\partial \tau_{ion}}{\partial \gamma} = \frac{-A\ \tau_{z_{ion}} \cos\gamma}{(B + \sin\gamma)^2} \tag{14b}$$

E. Clock Offset and Rate

The station clock offset is modeled as a random ramp, which consists of a random bias term to account for clock offset calibration errors in the ground instrumentation (e.g., hydrogen masers) together with station signal-path calibration errors, and a rate term representing the frequency offset calibration error between the two tracking stations participating in the three-way link.

Mathematically, the clock offset (delay) model can be expressed as

$$\tau_{clk} = \underbrace{b_o + b_{sp}}_{b_T} + f_o(t - t_0) \tag{15}$$

where

b_T = total clock bias, offset (b_o) plus signal-path (b_{sp}), between co-observing stations

f_o = frequency offset between co-observing stations

Second-order effects, such as frequency drift are neglected.

It is easy to derive the clock delay-rate model by virtue of a time-derivative of Eq. (15), which yields

$$\dot{\tau}_{clk} = f_o \tag{16}$$

III. Information Content Analysis

The partial derivatives of any data type represent, to first order, the ability of that data type to sense changes in a spacecraft trajectory. The "information content" of a particular data type is effectively described by the characteristics and behavior of its partial derivatives, and refers to the ability of a data type to determine the various elements that constitute a spacecraft trajectory model. This study specifically explores the information content of differenced range and Doppler for determining geocentric spacecraft angular and angular rate coordinates.

A. Differenced Range and Doppler Partial Derivatives and Error Analysis Formulation

A "linear" model is assumed for the regression equation expressed by

$$\mathbf{z} = A_x \mathbf{x} + \nu \tag{17}$$

where

$\mathbf{z} = [z_1, z_2, \cdots, z_N]^{\mathrm{T}}$, vector of N observations

$\mathbf{x} = \left[\delta_0, \alpha_0, b_{T_1}, b_{T_2}, \cdots, b_{T_{npass}}, \dot{\delta}_0, \dot{\alpha}_0, f_o\right]^{\mathrm{T}}$, vector of parameters to be estimated[4]

$\nu = [\nu_1, \nu_2, \cdots, \nu_N]^{\mathrm{T}}$, vector of N independent Gaussian measurement errors[5]

and A_x is the matrix of vector partial derivatives or partials of the observable, at the time of observation, with respect to the estimated parameter set:

$$A_x = \begin{pmatrix} \partial z_1/\partial \mathbf{x} \\ \partial z_2/\partial \mathbf{x} \\ \cdots \\ \partial z_N/\partial \mathbf{x} \end{pmatrix}$$

In this analysis, the observation set $\mathbf{z}$ contains differenced range and differenced Doppler measurements.

For a weighted least-squares estimator, the statistics associated with the estimation error can be readily computed by using the partial derivative matrix, A_x. A weighted least-squares estimate is one that minimizes the

[4] The subscript *npass* for the clock bias parameter indicates that an independent bias is assumed for each successive tracking pass, since multiple passes are evaluated in this study. An alternative method would be to model the bias as an exponentially correlated process noise parameter with an appropriately selected time constant. The frequency offset is assumed to be stable enough over several passes to warrant only a single representative parameter.

[5] Specifically, the components of ν are assumed to be independent, zero-mean, Gaussian random variables.

weighted sum of squares of the deviations between the actual and predicted (computed) measurements expressed by the scalar, quadratic cost function Q, written as

$$Q = \frac{1}{2}[\mathbf{z} - A_x\hat{\mathbf{x}}]^{\mathrm{T}} W [\mathbf{z} - A_x\hat{\mathbf{x}}] \tag{18}$$

in which $\hat{\mathbf{x}}$ is the "best" estimate of the unknown parameter vector $\mathbf{x}$ (i.e., minimizes the cost function Q) and W is taken to be a symmetric, positive definite weighting matrix. For the special case in which $W = \Gamma_\nu^{-1}$, where Γ_ν is the covariance matrix associated with the data noise vector, ν, the estimate $\hat{\mathbf{x}}$ that minimizes Q is the unbiased, minimum-variance estimate of $\mathbf{x}$, and is given by

$$\hat{\mathbf{x}} = \left(A_x^{\mathrm{T}}\ \Gamma_\nu^{-1} A_x\right)^{-1} A_x^{\mathrm{T}}\ \Gamma_\nu^{-1}\mathbf{z} \tag{19}$$

An important mathematical entity used in performing a statistical error analysis is the information matrix or information array. The information array associated with the estimated parameter set, denoted herein as J_x, can be determined computationally from the matrix of observation partials and the variances associated with each measurement:[6]

$$J_x \equiv A_x^{\mathrm{T}} W A_x = \sum_{i=1}^{N} \frac{1}{\sigma_{\nu_i}^2}\left(\frac{\partial z_i}{\partial \mathbf{x}}\right)^{\mathrm{T}}\left(\frac{\partial z_i}{\partial \mathbf{x}}\right) \tag{20}$$

Equation (20) assumes that the weighting matrix W is diagonal with the ith diagonal element w_i being equal to $1/\sigma_{\nu_i}^2$.

In many of the applications, some initial knowledge is available about a particular parameter or parameters of interest in the form of an unbiased a priori estimate. A priori statistics and the regression equation [Eq. (17)] can be combined to derive a modified form of the weighted least-squares estimator, expressed as

$$\hat{\mathbf{x}} = \left(\tilde{J}_x + A_x^{\mathrm{T}}\ \Gamma_\nu^{-1} A_x\right)^{-1} A_x^{\mathrm{T}}\ \Gamma_\nu^{-1}\mathbf{z} \tag{21}$$

The term $\tilde{J}_x$ denotes the a priori information array and is usually taken to be equal to the inverse of $\tilde{\Gamma}_x$, the initial covariance matrix for $\mathbf{x}$.[7]

From this development, the estimated or computed error covariance matrix, denoted $\Gamma_{x_{comp}}$, can readily be determined

$$\Gamma_{x_{comp}} \equiv E\left[(\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^{\mathrm{T}}\right] = \left(\tilde{J}_x + J_x\right)^{-1} \tag{22}$$

For differenced range measurements, the observation partials are found to be

$$\begin{aligned}\left(\frac{\partial \Delta\rho}{\partial \mathbf{x}}\right)^{\mathrm{T}} &= \frac{\partial \Delta\rho}{\partial\left[\delta_0, \alpha_0, b_{T_1}, b_{T_2}, \ldots, b_{T_{npass}}, \dot{\delta}_0, \dot{\alpha}_0, f_o\right]} \\ &= [-r_B \sin\delta \cos H_B + z_B \cos\delta, \\ &\quad r_B \cos\delta \sin H_B, 1, 1, \ldots, 1_{npass}, \\ &\quad (\partial\Delta\rho/\partial\delta_0)(t - t_0),\ (\partial\Delta\rho/\partial\alpha_0)(t - t_0), \\ &\quad t - t_0]\end{aligned} \tag{23}$$

and follow directly from the observable models provided in Section II.

For the case of differenced Doppler measurements, by letting $\dot{H}_B \rightarrow \omega$ (since $\dot{\alpha}_0 \ll \omega$) and substituting $\dot{\delta} = \dot{\delta}_0$, the observation partials are found to be

$$\begin{aligned}\left(\frac{\partial \Delta\dot{\rho}}{\partial \mathbf{x}}\right)^{\mathrm{T}} &= \frac{\partial \Delta\dot{\rho}}{\partial\left[\delta_0, \alpha_0, b_{T_1}, b_{T_2}, \ldots, b_{T_{npass}}, \dot{\delta}_0, \dot{\alpha}_0, f_o\right]} \\ &= \left[r_B\left(\omega \sin\delta \sin H_B - \dot{\delta}_0 \cos\delta \cos H_B\right) - z_B\dot{\delta}_0 \sin\delta, \right. \\ &\quad r_B\left(\omega \cos\delta \cos H_B - \dot{\delta}_0 \sin\delta \sin H_B\right), 0, 0, \ldots 0_{npass}, \\ &\quad (-r_B \sin\delta \cos H_B + z_B \cos\delta) + (\partial\Delta\dot{\rho}/\partial\delta_0)(t - t_0), \\ &\quad \left. r_B \cos\delta \sin H_B + (\partial\Delta\dot{\rho}/\partial\alpha_0)(t - t_0),\ 1\right]\end{aligned} \tag{24}$$

[6] The information array can also be computed analytically using an integral approximation if certain assumptions are made about the sampling rate of the data. This, in fact, was the method employed in the earlier companion studies [1,2]. Unfortunately, for this analysis, the observable models and associated partials are significantly more involved and do not lend themselves to easy analytical derivation of the information arrays.

[7] Generally speaking, the a priori information array need not be invertible. Furthermore, Eq. (21) represents a "normalized" form of the weighted least-squares estimator in which estimates are made of the corrections to the a priori values, hence, the a priori estimate of the parameter vector $\mathbf{x}$ is assumed to be zero.

B. Error Covariance Calculations (Part 1)

Recall that the baseline hour angle varies linearly with time and can be expressed as

$$H_B = H_{B_0} + \omega (t - t_0) \tag{25}$$

where

H_{B_0} = epoch baseline hour angle

$\omega t_0 = \alpha_0$

For this study, a "symmetric" tracking pass was assumed about H_{B_0} from which lower and upper limits on the baseline hour angle were used to accumulate the differenced range and Doppler information array given in Eq. (20). The lower and upper baseline hour angle limits, H_{B_l} and H_{B_u}, respectively, were taken to be

$$H_{B_l}, H_{B_u} = H_{B_0} - \Psi, H_{B_0} + \Psi \tag{26}$$

where

$\Psi \equiv$ tracking pass half-width.

A suitable choice for the tracking pass width 2Ψ was made for each candidate baseline by constraining the minimum elevation angle at each participating station to be approximately 10 deg. A more detailed description of this geometric dependence and suitable choices of 2Ψ for the DSN Canberra–Goldstone and Madrid–Goldstone baselines are provided in [1]. For the Canberra–Goldstone baseline, it was found that a constant tracking pass half-width value of 30 deg (about four hours' worth of continuous tracking) could be obtained for the range of spacecraft declination angles studied: -20 deg $\leq \delta \leq 20$ deg. In terms of data noise characteristics, the data accuracy for the differenced Doppler and range measurements were taken to be somewhat conservative by the standards of earlier analyses; specifically, measurement uncertainties were taken to be $\sigma_{\Delta\dot{\rho}} = 0.15$ mm/sec and $\sigma_{\Delta\rho} = 30.0$ cm. A 60-sec sampling rate was used for differenced Doppler and one differenced range point was acquired every 300 sec.

No a priori statistics were assumed for the spacecraft angular coordinate parameters to be estimated by the filter $(\delta_0, \alpha_0, \dot{\delta}_0, \dot{\alpha}_0)$. Conversely, a priori information was assumed to be available for the clock bias (offset/signal path) and frequency offset parameters, based on extrapolations of current DSN ranging and calibration system capabilities. The a priori information array $\tilde{J}_x$ was thus taken to be

$$\tilde{J}_x = \text{diag}\left[0, 0, \left(\frac{1}{\sigma_{b_{T_1}}}\right)^2, \left(\frac{1}{\sigma_{b_{T_2}}}\right)^2, \cdots, \left(\frac{1}{\sigma_{b_{T_{npass}}}}\right)^2, 0, 0, \left(\frac{1}{\sigma_{f_o}}\right)^2\right] \tag{27}$$

where

$\sigma_{b_{T_i}}$ = one-sigma a priori clock offset/signal-path uncertainty for the ith tracking pass

σ_{f_o} = one-sigma a priori frequency offset uncertainty

One-sigma uncertainties for the clock offset/signal-path and frequency offset parameters were assumed to be $\sigma_{b_{T_i}} =$ 10 nsec (~300 cm) and $\sigma_{f_o} = 0.02$ mm/sec, respectively.

Preliminary error analysis results for differenced range tracking from the DSN Canberra–Goldstone baseline suggest that the data are capable of determining spacecraft angular position coordinates to a precision of about 0.04 to 0.3 μrad (computed-only results) at the conclusion of five successive tracking passes, over the range of declinations investigated. The geocentric angular rate terms are seen to be determined within a range of 10 to 20 $\times 10^{-12}$ rad/sec. These results illustrate the difficulty which differenced range data have in being able to accurately sense spacecraft angular motion in the absence of other data types and a priori information. Differenced Doppler measurements were consequently used to augment the differenced range with computed-only results, which suggested that the two data types, when used in concert, can yield precisions on the order of 0.02 to 0.2 μrad for the geocentric angular coordinates and a reduction in the angular rate uncertainty to 3 to 8 $\times 10^{-12}$ rad/sec, again, at the conclusion of five successive tracking passes. The full set of these results is summarized in Table 2, which provides the one-sigma uncertainties for estimated parameters over the course of five tracking passes.

It is interesting that the best performance is seen at the smaller declination magnitudes. Although this is contrary

to the behavior seen in earlier studies [1,2], one must remember that the spacecraft angular coordinates were assumed to remain "fixed" over the evolution of tracking passes considered, whereas in this study, the data must also attempt to determine the angular rate terms. Without adequate initial knowledge of these parameters, the filter, hence the data, must work extremely hard not only to determine the spacecraft angular coordinates and rates, but the clock/clock-rate offsets as well.

It is well known that unmodeled delays due to clock offset and station signal path calibration errors can be a major factor preventing differenced range data from yielding angular precisions comparable to those of ΔVLBI data, and the addition of differenced Doppler data will not necessarily help, as they are nearly insensitive to clock offsets. For this reason, another error covariance calculation was made with a "tighter" a priori knowledge of the clock biases, assuming an advanced DSN ground calibration system emplaced, e.g., a DSN Global Positioning System (GPS) Ground Calibration System.[8] For these cases, the one-sigma a priori uncertainty in the clock offset was assumed to be 1 nsec (~30 cm). Because the epoch declination estimate is most affected by the uncertainty in the station clock offset, only the improvement in declination precision is shown (see Fig. 4). The results shown in Fig. 4 indicate that the ability of the differenced range data to determine the clock bias parameters is relatively weak in the near-zero declination regime; this is reflected in the more dramatic improvement seen for the tighter clock synchronization value, as is evident in the same figure, and the lesser improvement seen for the higher declination magnitudes. The inability of the filter to reduce the uncertainty in frequency offset is a reflection of its current highly precise calibration value.

A similar assessment was made for a DSN Madrid–Goldstone baseline, however, there are some important differences in terms of viewing constraints and tracking geometries that need to be noted before summarizing the error analysis results. The fact that both DSN complexes (Madrid and Goldstone) are located in the Northern Hemisphere severely restricts the tracking-pass width, and hence, the amount of available data at low declinations ($\delta \leq 5$ deg). The result, as seen in the earlier companion studies, is a severe degradation in precision at these lower declination angles. Furthermore, there is an inherent difficulty in being able to determine the spacecraft declination in the near-zero region resulting from the relatively small magnitude of the baseline z-height component, as compared with the Canberra–Goldstone baseline (see Table 1). The available overlap increases dramatically for the higher declination magnitudes ($\delta \leq 10$ deg). For this analysis, a tracking-pass half-width value of 30 deg was selected for $\delta = 10$ deg, and a value of 37 deg was chosen for the $\delta = 20$-deg case, which translates to about four and five hours of continuous tracking, respectively. (Note that the lower declination magnitude cases are ignored for this baseline for reasons just cited.)

Estimation statistics for the Goldstone–Madrid baseline are summarized in Table 3, in which identical assumptions on data-sampling rate and measurement accuracy characteristics are made as for the Canberra–Goldstone study, as well as on the a priori statistics. Results suggest that the differenced data types can together deliver about 0.2- to 0.5-μrad precision for the geocentric angular coordinates and about 3×10^{-12} to 40×10^{-12}-rad/sec precision for the angular rates, at the conclusion of five successive tracking passes. Clearly, the Canberra–Goldstone baseline results are superior for the $\delta = 10$-deg case by about a factor of three in which the same tracking-pass half-width value was assumed ($\Psi = 30$ deg). For the 20-deg case, on the other hand, better performance is seen for the Madrid–Goldstone baseline in terms of being able to determine the epoch declination and measurement biases. This should not be surprising, however, as a larger tracking-pass half-width value was employed ($\Psi = 37$ deg) than with the Canberra–Goldstone case. In fact, even with the increase in data volume, the determination of the epoch right ascension parameter and angular rate terms was poorer than with the Canberra–Goldstone baseline. Although the Madrid–Goldstone baseline offers a greater availability of data over the Canberra–Goldstone baseline, as the spacecraft declination angle increases, the latter offers the more favorable results in terms of angular precision.

IV. Sensitivity Analysis

A useful analysis tool is the sensitivity matrix method, which is frequently used in orbit determination error analyses and provides a means to distinguish among the effects of several different unmodeled systematic error sources on the parameter estimates [5,6]. Knowledge of the sensitivity matrix enables one to compute the full-consider error covariance matrix, which accounts for the computed uncertainty due purely to random measurement noise *plus* the uncertainty induced by unmodeled consider parameters.[9]

[8] S. M. Litchen, "GPS-Based DSN Calibration System (RTOP 61)," (presentation viewgraphs), Office of Spaceflight Operations Advanced Systems Review, DSN Advanced Systems Program, Jet Propulsion Laboratory, Pasadena, California, June 18–19, 1991.

[9] Recall that a consider parameter is treated by the filter as an unmodeled systematic error source.

A. Filter Augmentation

The original regression equation, Eq. (17), is augmented with an additional term, which results in a regression equation of the form

$$\mathbf{z} = A_x\mathbf{x} + A_y\mathbf{y} + \nu \tag{28}$$

where

$\mathbf{z} = [z_1, z_2, \cdots, z_N]^{\mathrm{T}}$, vector of N observations

$\mathbf{x} = \left[\delta_0, \alpha_0, b_{T_1}, b_{T_2}, \cdots, b_{T_{npass}}, \dot{\delta}_0, \dot{\alpha}_0, f_o\right]^{\mathrm{T}}$, vector of original estimated parameters

$\mathbf{y} = \left[r_B, z_B, \lambda_B, \tau_{z_{tro_1}}, \tau_{z_{tro_2}}, \tau_{z_{ion_1}}, \tau_{z_{ion_2}}\right]^{\mathrm{T}}$, vector of considered parameters

$\nu = [\nu_1, \nu_2, \cdots, \nu_N]^{\mathrm{T}}$, vector of N independent Gaussian measurement errors

and A_x is, again, the matrix of observation partials with respect to the estimated parameter set, while A_y is the matrix of observation partials with respect to the considered parameter set:

$$A_x = \begin{pmatrix} \partial z_1/\partial \mathbf{x} \\ \partial z_2/\partial \mathbf{x} \\ \cdots \\ \partial z_N/\partial \mathbf{x} \end{pmatrix}, \qquad A_y = \begin{pmatrix} \partial z_1/\partial \mathbf{y} \\ \partial z_2/\partial \mathbf{y} \\ \cdots \\ \partial z_N/\partial \mathbf{y} \end{pmatrix}$$

The sensitivity matrix, denoted S_{xy}, is defined to be

$$S_{xy} \equiv \frac{\partial \hat{\mathbf{x}}}{\partial \mathbf{y}} = \frac{\partial \hat{\mathbf{x}}}{\partial \mathbf{z}}\left(\frac{\partial \mathbf{z}}{\partial \mathbf{y}}\right)$$

$$= \Gamma_{x_{comp}} J_{xy} \tag{29}$$

where

$$J_{xy} \equiv A_x^{\mathrm{T}} W A_y = \sum_{i=1}^{N} \frac{1}{\sigma_{\nu_i}^2}\left(\frac{\partial z_i}{\partial \mathbf{x}}\right)^{\mathrm{T}}\left(\frac{\partial z_i}{\partial \mathbf{y}}\right) \tag{30}$$

Once again, it is assumed in Eq. (30), that the weighting matrix W, is diagonal.

For differenced range measurements, based on the observable development given in Section II and a substitution of the empirical ray-tracing constants for the static ionosphere delay model, the consider observation partials are found to be

$$\left(\frac{\partial \Delta\rho}{\partial \mathbf{y}}\right) = \frac{\partial \Delta\rho}{\partial\left[r_B, z_B, \lambda_B, \tau_{z_{tro_1}}, \tau_{z_{tro_2}}, \tau_{z_{ion_1}}, \tau_{z_{ion_2}}\right]^{\mathrm{T}}}$$

$$= \Big[\cos\delta\cos H_B,\ \sin\delta,\ -\partial\Delta\rho/\partial\alpha_0,$$

$$r_{sta_1}/(r_{s_1}\cos\delta\cos H_1 + z_{h_1}\sin\delta),$$

$$-r_{sta_2}/(r_{s_2}\cos\delta\cos H_2 + z_{h_2}\sin\delta),$$

$$1.15r_{sta_1}/(0.15r_{sta_1} + r_{s_1}\cos\delta\cos H_1 + z_{h_1}\sin\delta),$$

$$-1.15r_{sta_2}/(0.15r_{sta_2} + r_{s_2}\cos\delta\cos H_2 + z_{h_2}\sin\delta)\Big]^{\mathrm{T}} \tag{31}$$

In the case of differenced Doppler measurements, the consider observation partials are found, upon further substitution of $\dot{H}_B = \omega$ and $\dot{\delta} = \dot{\delta}_0$, to be

$$\left(\frac{\partial \Delta\dot{\rho}}{\partial \mathbf{y}}\right) = \frac{\partial \Delta\dot{\rho}}{\partial \left[r_B, z_B, \lambda_B, \tau_{z_{tro_1}}, \tau_{z_{tro_2}}, \tau_{z_{ion_1}}, \tau_{z_{ion_2}}\right]^{\mathrm{T}}}$$

$$= \Bigg[-\omega \cos\delta \sin H_B - \dot{\delta}_0 \sin\delta \cos H_B,\ \dot{\delta}_0 \cos\delta,\ -\partial\Delta\dot{\rho}/\partial\alpha_0,$$

$$\frac{1}{r_{sta_1} \sin^2 \gamma_{s_1}} \left[r_{s_1}\left(\dot{\delta}_0 \sin\delta \cos H_1 + \omega \cos\delta \sin H_1\right) - z_{h_1}\dot{\delta}_0 \cos\delta\right],$$

$$\frac{1}{r_{sta_2} \sin^2 \gamma_{s_2}} \left[-r_{s_2}\left(\dot{\delta}_0 \sin\delta \cos H_2 + \omega \cos\delta \sin H_2\right) + z_{h_2}\dot{\delta}_0 \cos\delta\right],$$

$$\frac{1.15}{r_{sta_1}(0.15 + \sin\gamma_{s_1})^2} \left[r_{s_1}\left(\dot{\delta}_0 \sin\delta \cos H_1 + \omega \cos\delta \sin H_1 + \omega\cos\delta\sin H_1\right) - z_{h_1}\dot{\delta}_0 \cos\delta\right],$$

$$\frac{1.15}{r_{sta_2}(0.15 + \sin\gamma_{s_2})^2} \left[-r_{s_2}\left(\dot{\delta}_0 \sin\delta \cos H_2 + \omega \cos\delta \sin H_2\right) + z_{h_2}\dot{\delta}_0 \cos\delta\right]\Bigg]^{\mathrm{T}} \tag{32}$$

The computation of the total or full-consider error covariance matrix, denoted $\Gamma_{x_{cons}}$, is given by

$$\Gamma_{x_{cons}} = \Gamma_{x_{comp}} + S_{xy}\ \Gamma_y\ S_{xy}^{\mathrm{T}} \tag{33}$$

where

$\Gamma_{x_{comp}}$ = computed error covariance matrix

S_{xy} = sensitivity matrix

Γ_y = consider parameter covariance matrix

The considered parameters are assumed to be uncorrelated, thus Γ_y is diagonal, with the associated consider variances as principal diagonal entries:

$$\Gamma_y = \mathrm{diag}\left[\sigma_{r_B}^2, \sigma_{z_B}^2, \sigma_{\lambda_B}^2, \sigma_{\tau_{z_{tro_1}}}^2, \sigma_{\tau_{z_{tro_2}}}^2, \sigma_{\tau_{z_{ion_1}}}^2, \sigma_{\tau_{z_{ion_2}}}^2\right] \tag{34}$$

This leads to the introduction of the perturbation matrix, denoted P_{xy}, and defined as

$$P_{xy} \equiv S_{xy}\sqrt{\Gamma_y} = S_{xy}\ \mathrm{diag}\left[\sigma_{r_B}, \sigma_{z_B}, \sigma_{\lambda_B}, \sigma_{\tau_{z_{tro_1}}}, \sigma_{\tau_{z_{tro_2}}}, \sigma_{\tau_{z_{ion_1}}}, \sigma_{\tau_{z_{ion_2}}}\right] \tag{35}$$

The perturbation matrix is another useful analysis tool as it indicates the one-sigma perturbation of each estimated parameter due to each consider parameter [6]. This information can be used to evaluate the impact of each individual consider parameter on the estimated parameter uncertainties or used to lump the effects into various groups of error sources; in this case, into observing platform errors and transmission media delay calibration errors.

The one-sigma uncertainties for the considered parameters are taken to be

$$\left.\begin{aligned} \sigma_{r_B} &= 12.6\ \mathrm{cm} \\ \sigma_{z_B} &= 13.0\ \mathrm{cm} \\ \sigma_{\lambda_B} &= 39.8\ \mathrm{nrad} \end{aligned}\right\} \text{(observing platform)}$$

$$\left.\begin{aligned} \sigma_{\tau_{tro_{1,2}}} &= 4.0\mathrm{cm} \\ \sigma_{\tau_{ion_{1,2}}} &= 5.0\mathrm{cm} \end{aligned}\right\} \text{(transmission media)}$$

These baseline coordinate uncertainties are representative of the combined uncertainty due to current levels of relative station location error and Earth orientation calibration error [7]. The troposphere zenith delay values are consistent with current DSN calibration capabilities, and the ionosphere zenith delay values reflect that X-band (8.4-GHz) radio link frequencies are assumed.

B. Error Covariance Calculations (Part 2)

Several cases were run to determine the effects of unmodeled systematic errors on the differenced Doppler and range data. In Tables 4 and 5, error statistics for the reference DSN Canberra–Goldstone and Madrid–Goldstone baseline cases using differenced Doppler together with differenced range over the evolution of five successive tracking passes are shown. These results reflect the *total* error in which the uncertainty due to the consider parameters is combined with the estimated parameter uncertainty due to measurement noise to better reflect "real world" results. Here, it is seen that the consider parameter effects can be quite substantial, even after several passes of data have been acquired. Although the estimated clock delay terms are only marginally affected by the unmodeled error sources, the parameters constituting the spacecraft angular motion are more severely impacted.

For the Canberra–Goldstone baseline, the most affected parameter is seen to be the epoch right ascension in which up to a seven-fold degradation in performance is evident at zero declination (see Tables 2 and 4) and about a four-fold degradation at the lowest and highest declination magnitudes (recall that symmetric passes are assumed). Degradation was less severe for the estimated epoch declination parameter yet still significant; about an 85-percent degradation at the extreme declination magnitudes and about a 23-percent degradation in the zero declination region. The angular rate parameters are degraded by about a factor of two to three-and-a-half times the nominal (computed) values over the selected declination range. In the case of the Madrid–Goldstone baseline, the error statistics suggest that the most heavily impacted parameter is epoch declination, especially for the 10-deg declination case; the resulting performance is shown to be quite poor, even after five successive tracking passes. In the 20-deg declination case, an approximate three-fold degradation in epoch declination determination is seen over the computed-only results, and the determination of epoch right ascension is degraded by about one-and-a-half times the estimated result. To gain insight into whether the observed degradation was due primarily to baseline coordinate errors or transmission media delay calibration errors, the one-sigma perturbations due to each consider parameter were computed (i.e., the perturbation matrix). The perturbations due to "lumped" error sources are shown in Fig. 5 for the Canberra–Goldstone baseline case. Clearly, the troposphere is the dominant error source in almost all cases. Generally speaking, the effects of baseline coordinate uncertainties are not seen to be significant.

As a result of these findings, a focus on reducing the effects of tropospheric path delay errors was made[10] that was attempted in two ways: (1) using a more accurate zenith delay calibration and (2) raising the elevation cutoff angles at the two stations constituting the baseline. Another possible strategy, which would require further research, would be the use of an elevation-dependent data-weighting function, such as the one that was recently developed for two-way X-band (8.4-GHz) Doppler [8,9]. The results obtained with these techniques are illustrated in Fig. 6, for the most severely affected estimated parameter—epoch right ascension. Here, it is seen that the most dramatic improvement results from a smaller zenith delay calibration error value, taken to be 2 cm in this case, which could be achieved by an improved DSN ground calibration system. The more interesting curve results from the higher elevation cutoff (approximately 15 deg) in which the reduction in uncertainty is also seen to be substantial. These results indicate that an elevation cutoff of 15 deg (or perhaps higher) could yield significantly improved performance over the results given in Tables 4 and 5 for a 10-deg cutoff, even without the benefit of an improved troposphere calibration technique.

V. Remarks

Differenced Doppler and range measurements acquired from multiple baselines were not addressed in this analysis. Results from the earlier companion studies suggested that if knowledge of DSN station clock offsets was assumed to be relatively large, angular precision for cases involving longer differenced data arcs from a single baseline were superior to the shorter data arcs acquired from two different baselines [1,2]. This, of course, was a reflection of a greater amount of data available to the filter for estimating clock offsets. For studies involving more optimistic assumptions about the DSN's ability to calibrate station clock offsets, better performance was observed for the dual baseline cases over the single baseline cases, despite the shorter data arcs used from each of the two individual baselines. Another tracking strategy that has been

[10] For the Canberra–Goldstone baseline case only.

suggested is to alternate data acquisition from the DSN baselines on a per-pass basis in an effort to further reduce the sensitivity of the differenced data types to transmission media delay effects; this, however, will require further study.[11]

To provide a reference point for comparison with the differenced range and Doppler results, angular precision and angular rate precision estimates were computed for both wideband ΔVLBI (ΔDOR) and narrowband ΔVLBI (ΔDOD) data acquired from a single baseline over a period of a few days. In these calculations, it was assumed that one ΔDOR measurement and one ΔDOD measurement were acquired simultaneously each day from the DSN Goldstone–Canberra baseline for five successive days. The measurement accuracies assumed for these data were 20 cm for ΔDOR, and 0.05 mm/sec for ΔDOD; these measurement accuracies are representative of the performance that can be achieved at X-band (8.4-GHz) frequencies.[12] The results for five different declination values ranging from −20 deg to +20 deg are given in Table 6. The baseline hour angle for each pair of ΔDOR/ΔDOD measurements was chosen so that a spacecraft at the specified declination angle would be observed at or near the maximum elevation angle from both the Goldstone and Canberra complexes. Small departures of up to 10 deg in the baseline hour angle away from this configuration were intentionally made so as to vary the observing geometry some, although no attempt was made to choose the baseline hour angles for each day in such a way as to optimize the results.

Compared with the differenced range and Doppler results given in Tables 4 and 5, the data in Table 6 show comparable angular precision (about 0.07 to 0.04 μrad), but much better angular rate precision than the differenced range and Doppler data (0.5×10^{-12} to 1.0×10^{-12} rad/sec as opposed to 7×10^{-12} to 25×10^{-12} rad/sec). A comparison of Table 6 with the ideal (no systematic error source effects) differenced range/Doppler precision data given in Tables 1 and 2 indicates that the ideal angular precision obtained with differenced range/Doppler data is much closer to that obtained with ΔVLBI data, but that the ΔVLBI data still yield the greatest angular precision even in this scenario.

VI. Conclusions

The simple three-parameter analytic differenced range and Doppler observable models that were developed in the previous studies were refined in this analysis into more-detailed six-parameter observable models in order to ascertain the ability of these data types to estimate both spacecraft angular coordinates and DSN clock/frequency offsets, as well as spacecraft angular rate coordinates, over a tracking period of a few days. Furthermore, the expanded models incorporated systematic station location and transmission media (troposphere/ionosphere) calibration errors which were not addressed previously.

Error covariance calculations suggested that a few differenced Doppler plus ranging passes were capable of yielding angular position estimates with a precision on the order of 0.1 to 0.4 μrad, and an angular rate precision on the order of 3 to 25×10^{-12} rad/sec—this in the absence of any a priori statistical information on the coordinate parameters. Results based on sensitivity analysis calculations suggested that the most dominant systematic error source in most of the tracking scenarios that were investigated was troposphere zenith delay calibration error. As expected, the differenced Doppler data were found to be more sensitive to troposphere calibration errors than the differenced range data. However, it was also discovered that by raising the elevation cutoff to 15 deg at both stations constituting the baseline, the effects due to troposphere calibration errors were significantly reduced. A similar, yet even more dramatic improvement was seen when an improved zenith delay value was chosen based on an advanced DSN ground calibration system employing GPS measurements. These quantitative results were based strictly on the differenced data types themselves, as no additional radio metric data types were assumed, e.g., two-way (coherent) Doppler and/or range. It must be remembered that in an operational environment, the differenced data types would be used in conjunction with conventional line-of-sight tracking data types (two-way range/Doppler), however, further study is needed to establish navigation accuracy estimates for more realistic scenarios. For comparison purposes, error covariance calculations were also performed using wideband ΔVLBI (ΔDOR) and narrowband ΔVLBI (ΔDOD) data which yielded angular precisions on the order of 0.07 to 0.4 μrad, and angular rate precisions on the order of 0.5 to 1.0×10^{-12} rad/sec.

[11] W. M. Folkner, Tracking Systems and Applications Section, personal communication, Jet Propulsion Laboratory, Pasadena, California, January 22, 1992.

[12] J. S. Border, "Analysis of ΔDOR and ΔDOD Measurement Errors for Mars Observer Using the DSN Narrow Channel Bandwidth VLBI System," Interoffice Memorandum 335.1-90-026 (internal document), Jet Propulsion Laboratory, Pasadena, California, May 15, 1990.

Acknowledgments

The authors thank Carl Christensen and John McNamee for their useful comments and suggestions in review of this article.

References

[1] S. W. Thurman, "Deep-Space Navigation With Differenced Data Types, Part I: Differenced Range Information Content," *TDA Progress Report 42-103*, July–September 1990, Jet Propulsion Laboratory, Pasadena, California, pp. 47–60, November 15, 1990.

[2] S. W. Thurman, "Deep-Space Navigation With Differenced Data Types, Part II: Differenced Doppler Information Content," *TDA Progress Report 42-103*, July–September 1990, Jet Propulsion Laboratory, Pasadena, California, pp. 61–69, November 15, 1990.

[3] S. C. Wu, "Atmospheric Media Effects on ARIES Baseline Determination," *TDA Progress Report 42-61*, November–December 1980, Jet Propulsion Laboratory, Pasadena, California, pp. 1–6, February 15, 1981.

[4] J. A. Klobuchar, *A First-Order, Worldwide, Ionosphere Time Delay Algorithm*, Document ARCRL-TR-75-0502, Air Force Cambridge Research Laboratories, Hanscom Air Force Base, Massachusetts, September 25, 1975.

[5] S. R. McReynolds, "The Sensitivity Matrix Method for Orbit Determination Error Analysis, With Applications to a Mars Orbiter," *JPL Space Programs Summary 37-56*, vol. 3, January–February 1969, pp. 85–87, March 31, 1969.

[6] G. J. Bierman, *Factorization Methods for Discrete Sequential Estimation*, San Diego, California: Academic Press, 1977.

[7] S. W. Thurman, "DSN Baseline Coordinate and Station Location Errors Induced by Earth Orientation Errors," *TDA Progress Report 42-103*, July–September 1990, Jet Propulsion Laboratory, Pasadena, California, pp. 40–46, November 15, 1990.

[8] J. S. Ulvestad and S. W. Thurman, "Orbit-Determination Performance of Doppler Data for Interplanetary Cruise Trajectories, Part I: Error Analysis Methodology," *TDA Progress Report 42-108*, October–December 1991, Jet Propulsion Laboratory, Pasadena, California, pp. 31–48, February 15, 1992.

[9] J. S. Ulvestad, "Orbit-Determination Performance of Doppler Data for Interplanetary Cruise Trajectories Part II: 8.4-GHz Performance and Data-Weighting Strategies," *TDA Progress Report 42-108*, October–December 1991, Jet Propulsion Laboratory, Pasadena, California, pp. 49–65, February 15, 1992.

Table 1. DSN station and baseline cylindrical coordinates.

Station	Location	r_s, km	z_h, km	λ, deg
DSS 14	Goldstone	5203.997	3677.052	243.1105
DSS 43	Canberra	5205.251	−3674.749	148.9813
DSS 63	Madrid	4862.451	4115.109	355.7520
Baseline	**Length, km**	r_B, km	z_B, km	λ_B, deg
DSS 43–14	10,588.966	7620.841	7351.801	286.0523
DSS 63–14	8,390.430	8378.986	−438.057	210.7265

Table 2. Differenced Doppler plus range angular precision using DSN Canberra–Goldstone baseline (computed-only results, 1σ).

Estimated parameter	Number of passes				
	1	2	3	4	5
			$\delta = -20$ deg		
δ_0, nrad	566.01	387.40	311.33	267.26	237.83
α_0, nrad	307.18	187.04	139.82	114.64	99.09
$\dot{\delta}_0$, prad/sec	21.23	13.40	10.36	8.75	7.76
$\dot{\alpha}_0$, prad/sec	10.13	6.11	4.50	3.62	3.06
			$\delta = -10$ deg		
δ_0, nrad	440.24	308.26	250.51	216.38	193.23
α_0, nrad	112.64	66.59	48.51	38.91	33.02
$\dot{\delta}_0$, prad/sec	12.08	7.55	5.85	4.98	4.47
$\dot{\alpha}_0$, prad/sec	10.23	5.96	4.24	3.31	2.73
			$\delta = 0$ deg		
δ_0, nrad	411.37	289.98	236.39	204.53	182.82
α_0, nrad	56.57	32.66	23.10	17.89	14.61
b_{T_1}, cm	300.00	212.15	173.22	150.02	134.18
b_{T_2}, cm	–	212.15	173.23	150.02	134.18
b_{T_3}, cm	–	–	173.23	150.02	134.18
b_{T_4}, cm	–	–	–	150.04	134.19
b_{T_5}, cm	–	–	–	–	134.21
$\dot{\delta}_0$, prad/sec	9.89	6.12	4.74	4.05	3.66
$\dot{\alpha}_0$, prad/sec	10.24	5.91	4.18	3.24	2.64
f_o, mm/sec	0.02	0.02	0.02	0.02	0.02
			$\delta = 10$ deg		
δ_0, nrad	438.62	307.27	249.77	215.77	192.70
α_0, nrad	110.26	65.28	47.63	38.27	32.53
$\dot{\delta}_0$, prad/sec	12.07	7.55	5.85	4.98	4.47
$\dot{\alpha}_0$, prad/sec	10.23	5.96	4.25	3.31	2.73
			$\delta = 20$ deg		
δ_0, nrad	561.34	384.56	309.20	265.52	236.33
α_0, nrad	304.68	185.80	139.10	114.18	98.81
b_{T_1}, cm	299.97	212.12	173.19	149.98	134.14
b_{T_2}, cm	–	212.12	173.19	149.98	134.14
b_{T_3}, cm	–	–	173.20	149.99	134.15
b_{T_4}, cm	–	–	–	150.01	134.15
b_{T_5}, cm	–	–	–	–	134.18
$\dot{\delta}_0$, prad/sec	21.19	13.38	10.35	8.74	7.76
$\dot{\alpha}_0$, prad/sec	10.14	6.11	4.50	3.62	3.06
f_o, mm/sec	0.02	0.02	0.02	0.02	0.02

Table 3. Differenced Doppler plus range angular precision using DSN Madrid–Goldstone baseline (computed-only results, 1σ).

Estimated parameter	Number of passes				
	1	2	3	4	5
	$\delta = 10$ deg				
δ_0, nrad	1207.32	762.43	594.71	509.12	458.51
α_0, nrad	393.96	277.10	225.54	194.87	173.97
$\dot{\delta}_0$, prad/sec	84.50	59.22	48.11	41.53	37.04
$\dot{\alpha}_0$, prad/sec	14.32	8.38	6.03	4.76	3.98
	$\delta = 20$ deg				
δ_0, nrad	324.05	224.20	190.60	175.09	166.63
α_0, nrad	379.64	266.01	215.31	184.88	163.99
b_{T_1}, cm	294.50	206.39	167.03	143.40	127.18
b_{T_2}, cm	–	206.39	167.04	143.41	127.18
b_{T_3}, cm	–	–	167.04	143.41	127.18
b_{T_4}, cm	–	–	–	143.43	127.19
b_{T_5}, cm	–	–	–	–	127.21
$\dot{\delta}_0$, prad/sec	39.77	27.82	22.50	19.31	17.12
$\dot{\alpha}_0$, prad/sec	8.23	4.88	3.58	2.90	2.49
f_o, mm/sec	0.02	0.02	0.02	0.02	0.02

Table 4. Differenced Doppler plus range angular precision using DSN Canberra–Goldstone baseline (full-consider results, 1σ).

Estimated parameter	Number of passes				
	1	2	3	4	5
	$\delta = -20$ deg				
δ_0, nrad	675.03	534.78	483.21	456.86	441.18
α_0, nrad	498.46	435.05	417.21	409.79	406.10
$\dot{\delta}_0$, prad/sec	33.37	29.02	27.74	27.18	26.88
$\dot{\alpha}_0$, prad/sec	15.42	13.14	12.48	12.20	12.05
	$\delta = -10$ deg				
δ_0, nrad	474.99	356.27	307.83	280.95	263.76
α_0, nrad	201.46	179.86	174.03	171.68	170.53
$\dot{\delta}_0$, prad/sec	19.49	17.06	16.38	16.09	15.94
$\dot{\alpha}_0$, prad/sec	15.57	13.16	12.48	12.19	12.05
	$\delta = 0$ deg				
δ_0, nrad	431.60	318.02	270.05	242.64	224.64
α_0, nrad	128.48	119.89	117.64	116.73	116.27
b_{T_1}, cm	300.00	212.15	173.22	150.02	134.18
b_{T_2}, cm	–	212.15	173.23	150.02	134.18
b_{T_3}, cm	–	–	173.23	150.02	134.18
b_{T_4}, cm	–	–	–	150.04	134.19
b_{T_5}, cm	–	–	–	–	134.21
$\dot{\delta}_0$, prad/sec	16.45	14.50	13.97	13.76	13.65
$\dot{\alpha}_0$, prad/sec	15.53	13.09	12.41	12.12	11.98
f_o, mm/sec	0.02	0.02	0.02	0.02	0.02
	$\delta = 10$ deg				
δ_0, nrad	472.99	354.75	306.47	279.66	262.50
α_0, nrad	119.78	178.97	173.38	171.12	170.02
$\dot{\delta}_0$, prad/sec	19.48	17.05	16.37	16.08	15.93
$\dot{\alpha}_0$, prad/sec	15.53	13.11	12.43	12.14	12.00
	$\delta = 20$ deg				
δ_0, nrad	668.69	529.70	478.56	452.43	436.90
α_0, nrad	496.29	433.83	416.28	409.00	405.41
b_{T_1}, cm	299.98	212.16	173.27	150.13	134.39
b_{T_2}, cm	–	212.16	173.28	150.14	134.39
b_{T_3}, cm	–	–	173.29	150.14	134.39
b_{T_4}, cm	–	–	–	150.16	134.40
b_{T_5}, cm	–	–	–	–	134.42
$\dot{\delta}_0$, prad/sec	33.32	28.99	27.72	27.17	26.87
$\dot{\alpha}_0$, prad/sec	15.34	13.05	12.38	12.10	11.95
f_o, mm/sec	0.02	0.02	0.02	0.02	0.02

Table 5. Differenced Doppler plus range angular precision using Madrid–Goldstone baseline (full-consider results 1σ).

Estimated parameter	Number of passes				
	1	2	3	4	5
	$\delta = 10$ deg				
δ_0, nrad	2264.36	2057.61	1997.18	1968.97	1952.06
α_0, nrad	410.51	300.00	253.87	228.87	214.37
$\dot{\delta}_0$, prad/sec	92.11	68.40	57.76	51.16	46.45
$\dot{\alpha}_0$, prad/sec	26.52	23.83	23.09	22.77	22.61
	$\delta = 20$ deg				
δ_0, nrad	624.82	578.70	565.86	560.16	556.92
α_0, nrad	387.24	285.02	251.82	244.03	249.26
b_{T_1}, cm	299.48	221.39	197.27	192.83	198.24
b_{T_2}, cm	–	221.39	197.27	192.83	198.24
b_{T_3}, cm	–	–	197.27	192.83	198.24
b_{T_4}, cm	–	–	–	192.84	198.24
b_{T_5}, cm	–	–	–	–	198.24
$\dot{\delta}_0$, prad/sec	40.22	28.26	23.29	21.01	20.32
$\dot{\alpha}_0$, prad/sec	15.65	14.17	13.77	13.60	13.52
f_o, mm/sec	0.02	0.02	0.02	0.02	0.02

Table 6. ΔVLBI angular precision (1σ) using DSN Goldstone–Canberra baseline (results for five consecutive passes).

Estimated parameter	Number of passes				
	−20	−10	0	10	20
	$\delta = 10$ deg				
δ_0, nrad	71.0	115.0	235.0	384.0	152.0
α_0, nrad	92.6	127.0	236.0	369.0	150.0
$\dot{\delta}_0$, prad/sec	0.55	1.35	0.98	1.22	0.51
$\dot{\alpha}_0$, prad/sec	0.56	1.30	1.00	1.18	0.53

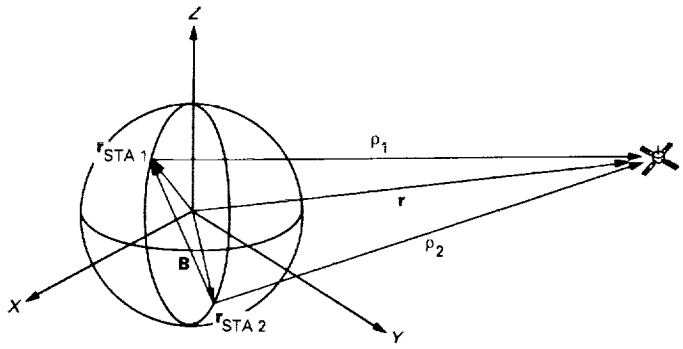


Fig. 1. Differenced range measurement geometry.

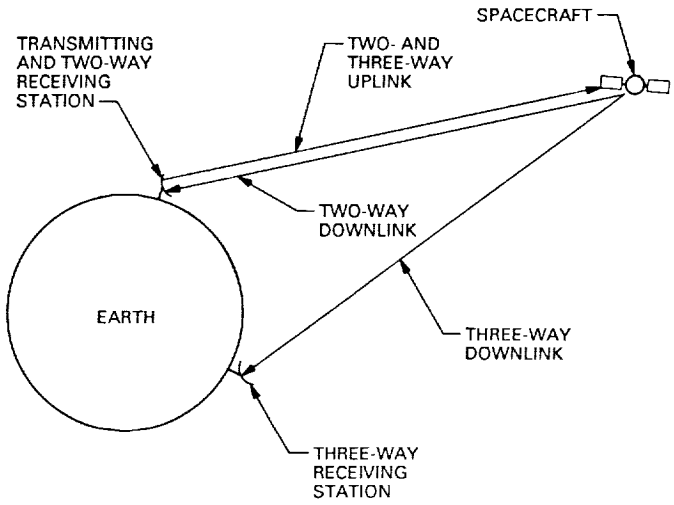


Fig. 3. Differenced Doppler and range data acquisition scheme.

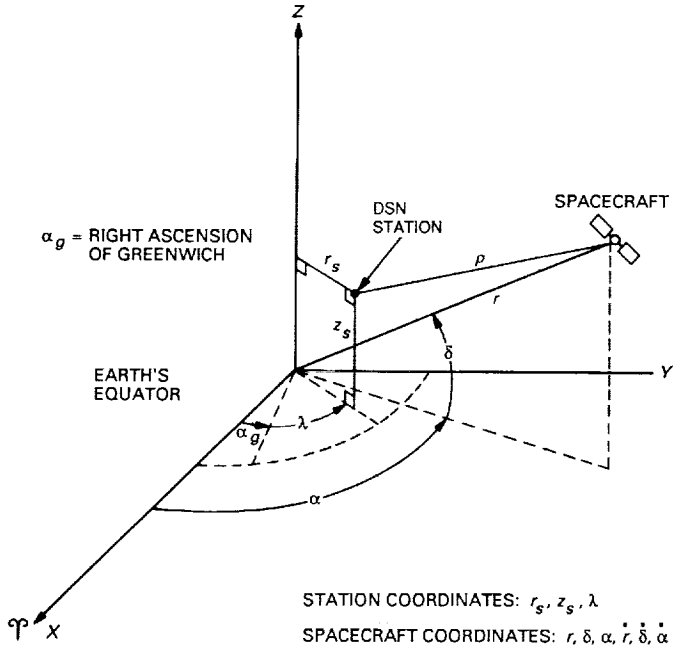


Fig. 2. Spacecraft and DSN station coordinates.

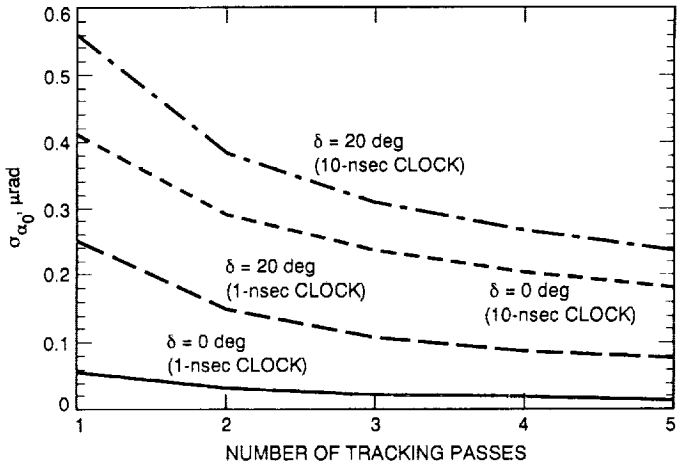


Fig. 4. Comparison of Canberra–Goldstone baseline declination precision for varying clock bias values.

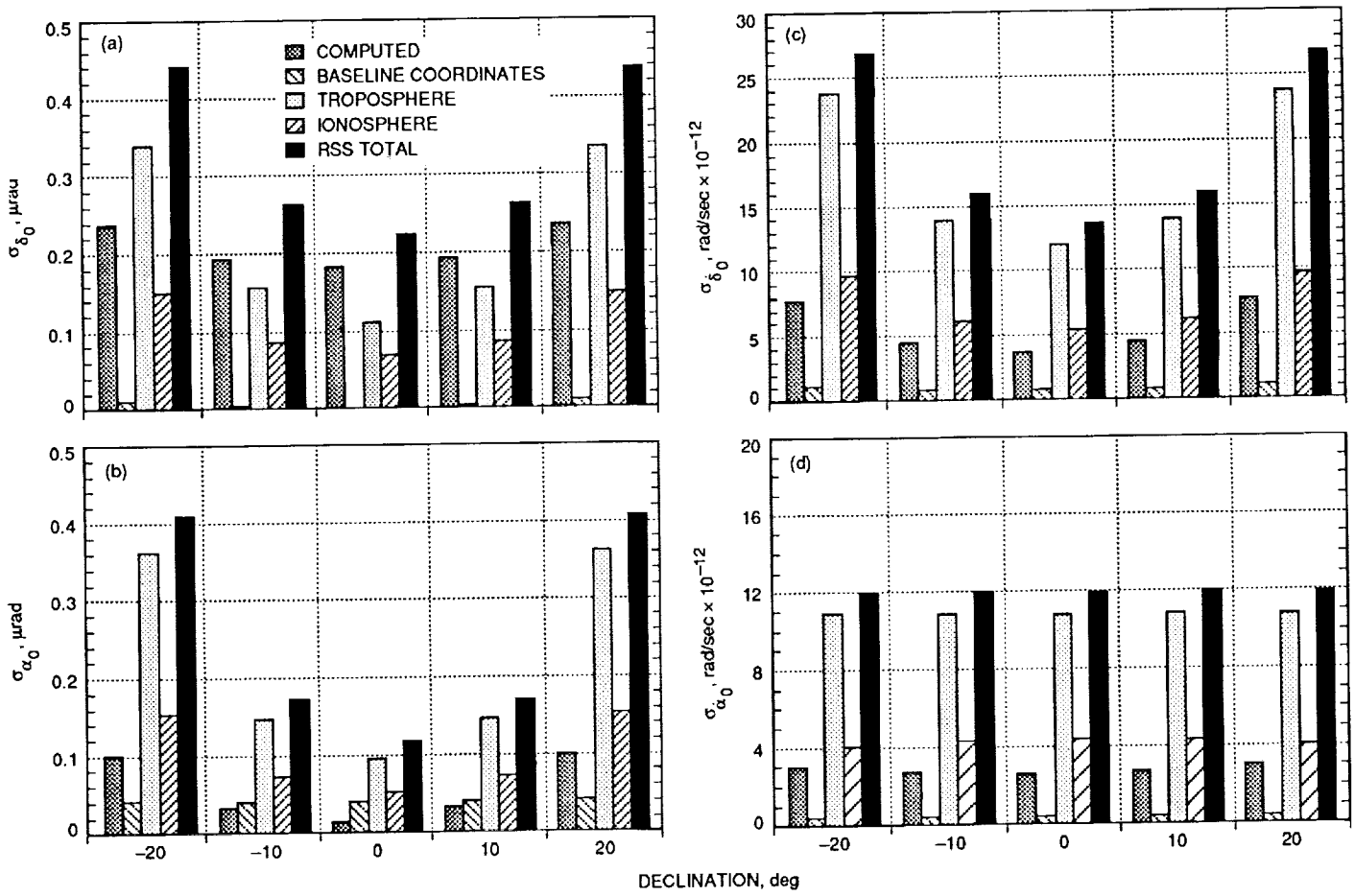


Fig. 5. Angular precision in terms of individual error sources for DSN Canberra–Goldstone baseline: (a) epoch declination uncertainty; (b) epoch right ascension uncertainty; (c) epoch declination rate; and (d) epoch right ascension rate.

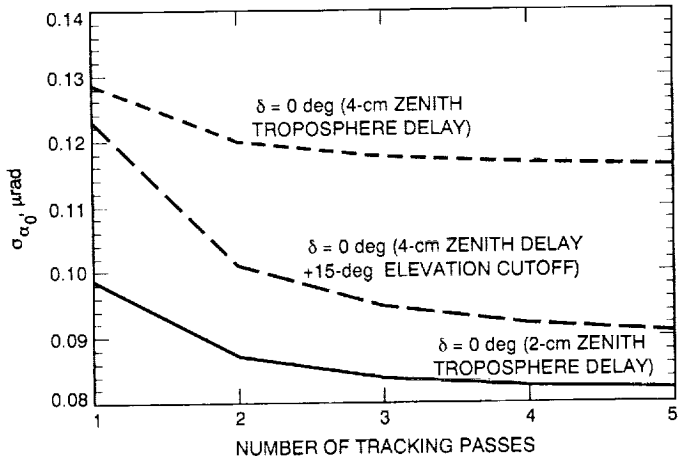


Fig. 6. Performance comparison for various troposphere path-delay reduction schemes using DSN Canberra–Goldstone baseline.

Orbiter–Orbiter and Orbiter–Lander Tracking Using Same-Beam Interferometry

W. M. Folkner and J. S. Border
Tracking Systems and Applications Section

Two spacecraft orbiting Mars will subtend a small angle as viewed from Earth. This angle will usually be smaller than the beam width of a single radio antenna. Thus the two spacecraft may be tracked simultaneously by a single Earth-based antenna. The same-beam interferometry (SBI) technique involves using two widely separated antennas, each observing the two spacecraft, to produce a measurement of the angular separation of the two spacecraft in the plane of the sky. The information content of SBI data is thus complementary to the line-of-sight information provided by conventional Doppler data. The inclusion of SBI data with the Doppler data in a joint orbit estimation procedure can desensitize the solution to gravity mismodeling and result in improved orbit determination accuracy. This article presents an overview of the SBI technique, a measurement error analysis, and an error covariance analysis of some examples of the application of SBI to orbit determination. For hypothetical scenarios involving the Mars Observer and the Russian Mars '94 spacecraft, orbit determination accuracy improvements of up to an order of magnitude are predicted, relative to the accuracy that can be obtained by using only Doppler data acquired separately from each spacecraft. Relative tracking between a Mars orbiter and a lander fixed on the surface of Mars is also studied. Results indicate that the lander location may be determined to a few meters, while the orbiter ephemeris may be determined with accuracy similar to the orbiter–orbiter case.

I. Introduction

Measurements of the radio signal emitted by a spacecraft orbiting another planet provide information about the spacecraft's position and velocity. A single Earth-based tracking station can directly measure line-of-sight range rate. The spacecraft trajectory can be inferred from an analysis of the time signatures imposed by the spacecraft acceleration due to gravity and by the change in geometry due to the orbital motion of the Earth and target planet. The orbit determination accuracy that can be achieved is limited by measurement system errors and errors in the spacecraft force models. The former include ground instrumental errors, clock instability, uncertain

Earth orientation, uncalibrated media delays, and errors in tracking station locations. The mismodeled forces include uncertainties in the gravity field of the central body, the force due to solar pressure, and atmospheric drag. For a low orbiter, with a period of a few hours, the uncertainties in the gravity field often dominate the orbit determination uncertainty.

Differential measurements are affected less by ground system and media errors, and desensitize orbit solutions to spacecraft force modeling errors. Differential measurements studied for low orbiters such as Magellan or Mars Observer (MO) include differenced-Doppler and delta differential one-way Doppler (ΔDOD). Differenced-Doppler involves the use of two tracking stations to measure the Doppler shift of the spacecraft carrier, with one of the stations providing a stable uplink frequency. A clock rate offset between the two stations introduces a systematic error into this measurement. The ΔDOD measurement subtracts the differential spacecraft Doppler from the differential frequency shift of a quasar to calibrate the station clock offset. Both of these measurement types give, essentially, a measurement of the spacecraft velocity in the plane of the sky (the plane perpendicular to the Earth–spacecraft line of sight). This type of measurement has been shown to improve orbit determination accuracy in case studies for Magellan [1].

For a number of upcoming interplanetary missions, primarily directed towards Mars, two spacecraft may simultaneously be in orbit about the same planet. This opens the possibility of differential measurements between the two spacecraft instead of between a spacecraft and a quasar. The angle between the two spacecraft as viewed from Earth will be much smaller than the usual spacecraft–quasar angle, which is typically 10 deg. Since many measurement errors scale with the angular separation, the spacecraft–spacecraft measurement is potentially much more accurate than conventional interferometric measurements. If the two spacecraft lie within the beam width of a single antenna, as will often be the case, the carrier phases of both spacecraft can be simultaneously tracked. This use of phase rather than group delay (or delay rate) further increases the measurement accuracy. It is predicted that the same-beam interferometry (SBI) measurement accuracy could be up to three orders of magnitude better than that for conventional spacecraft–quasar interferometry [2,3].

The utility of SBI measurements will depend on a number of factors, such as the data arc length, data accuracy and weighting, orbital geometry, and gravity modeling uncertainties. The analysis presented here is not meant to be definitive but rather to show the relative power of adding SBI data to nominal Doppler tracking strategies. The sample cases presented below are based on the Russian Mars '94 mission arriving at Mars while MO is nearing the end of its prime mission. An earlier opportunity to perform SBI measurements and demonstrate their utility for orbit determination occurred in August 1991 when Magellan joined Pioneer 12 in orbit about Venus [4].

A more futuristic case involves relative tracking with respect to landers on the surface of Mars. There are many possible applications which then arise, including lander–rover tracking for rover position estimation, or lander–spacecraft tracking for spacecraft approaching the planet. The potential use of SBI to determine relative lander–rover positions at the meter level has been discussed briefly elsewhere [5]. The only case involving a lander included below involves the use of SBI for tracking an orbiting spacecraft with respect to a lander.

II. Same-Beam Interferometry Technique

The SBI measurement of two spacecraft is depicted in Fig. 1. Two ground stations measure the phase of the carrier signal from each spacecraft as a function of time. The measured phase, when differenced between spacecraft and differenced between ground stations, gives an instantaneous measurement of the separation of the two spacecraft in the plane of the sky (in the direction along the projected baseline). The relative measurements can be included in a joint orbit estimation process along with Doppler data obtained from each of the two spacecraft. The use of all the data in a single estimation process ties both orbits to the center of gravitation through the dynamic signatures in the data.

The SBI measurement is similar to a delta differential one-way range (ΔDOR) measurement [6] in that an observable is formed from the observation of two sources at two widely separated ground stations. The ΔDOR measurement determines the spacecraft differential one-way range and compares this with the interferometric delay of an angularly nearby quasar to calibrate the ground station clocks and other common mode errors. The ΔDOR measurement uses tones modulated on the downlink carrier to determine the group delay, with a precision of a fraction of the wavelength corresponding to the spanned bandwidth. This wavelength is about 7.5 m for the 40-MHz spanned bandwidth for X-band (8.4-GHz) ΔDOR measurements. The SBI measurement has the advantage of a much smaller angular separation between the two spacecraft in orbit (a fraction of a milliradian) than the

spacecraft–quasar separation for a ΔDOR measurement (typically 10 deg or about 175 mrad). The SBI measurements determine the phase delay to a precision of a fraction of the carrier wavelength, which is 3.6 cm at X-band. The combination of smaller angular separation and the use of phase delay rather than group delay results in a theoretical accuracy for SBI measurements of 0.2 mm compared with the 14-cm accuracy of conventional X-band ΔDOR measurements.

SBI has some operational advantages over conventional spacecraft interferometry. Since no quasar is used, there is no need to change the pointing of the antennas away from the spacecraft. Without the quasar, there is also no need for a cross-correlation step in the data processing. Appropriately designed receivers could simultaneously extract the phase measurements in real time. The phase measurements would then be processed much like conventional Doppler data.

The SBI data do not determine the doubly differenced carrier phase unambiguously. Each arc of SBI measurements begins with a phase bias, which consists of an integer number of cycles and a possible fractional phase due to imperfectly calibrated instrumental delays. Provided that the calibration of the station instrumentation is sufficient to determine the fractional phase to a level small as compared with the SBI data noise, the phase bias can be constrained to be an integer number of cycles. The phase bias must be provided from a priori information or else a phase bias must be estimated for each SBI data arc. The phase bias could be determined from group delay measurements (using widely spaced tones about the carrier) or from a sufficiently accurate a priori orbit solution. In particular, if the orbit solution using SBI data with the phase bias estimated as a real number can determine the bias to a small fraction of a cycle, then the phase bias can be confidently fixed to the nearest integer. A subsequent orbit solution will result in greater orbit determination accuracy.

An error budget for X-band SBI measurements is presented in Table 1. The error budget assumes a Sun–Earth–Probe (SEP) angle of 20 deg, a spacecraft separation angle $\Delta\theta$ of 100 μrad, an Earth–spacecraft distance of 1.4 AU, and a projected baseline length of 8000 km (a representative value for intercontinental baselines). The largest error contribution listed is from imperfect delay cancellation due to solar plasma. Nearly all charged-particle effects could be removed by dual-frequency measurements. However, the upcoming opportunities for SBI measurements are not assured of having two frequencies from both spacecraft. The particular case studied below occurs when the Russian Mars '94 mission arrives at Mars in 1995 while MO is still active. The Mars '94 spacecraft will have an X-band transmitter to provide a signal for ΔDOR measurements to be acquired by the DSN during cruise, but it will rely on C-band (6-GHz) transmissions for telemetry and conventional Doppler and range tracking. MO will use X-band for telemetry and navigation although it does have an experimental 34-GHz capability. Often the charged-particle effects will be less than those shown in Table 1 since the solar plasma effects will be less at SEP angles larger than 20 deg. In that respect the measurement error budget given in Table 1 is somewhat conservative. The terms in the error budget are briefly discussed below.

A. Solar Plasma

The solar plasma error has been calculated by using a thin-screen frozen turbulence model [7]. The differential delay error is computed numerically; sample results are shown in Fig. 2. It can be seen that for these small spacecraft separation angles, the differential delay error is approximately linear with separation angle and SEP angle. The applicability of this model to SBI measurements will need to be tested by taking some experimental data.

B. Ionosphere

Ionosphere calibration is provided to the DSN from Global Positioning System (GPS) measurements. The error in the Earth's ionospheric delay mapped to any line of sight after calibration is approximately 30 mm at X-band [8]. The differential delay error ε_d for two nearby lines of sight is taken to be

$$\varepsilon_d \text{ (mm)} = 30 \text{ mm} \times \Delta\theta \times F \times \sqrt{2} \qquad (1)$$

where $\Delta\theta$ is the spacecraft separation angle (in radians), F is a factor representing the derivative of the mapping function with respect to angle in the direction of $\Delta\theta$, and the $\sqrt{2}$ factor is introduced because there are independent errors at each station. The mapping functions represent the largest uncertainty in ionospheric calibrations and are not well known for the small separations needed for SBI data. The derivative of the mapping function used for GPS calibrations has a maximum value of 3.5/rad if the separation angle occurs in elevation. Here a value of 5/rad for the derivative of the mapping function was assumed to be somewhat more conservative. More study will be needed to better understand the ionospheric error for SBI.

C. Troposphere

The tropospheric error is represented by

$$\varepsilon_d \text{ (mm)} = \sqrt{2} \times \Delta\theta \times 40 \text{ mm} \times \cos(E)/\sin^2(E) \qquad (2)$$

where E is the elevation angle, $\Delta\theta$ is the difference in elevation angle between the two spacecraft, and the troposphere delay error is taken to be a zenith value of 40 mm mapped to lower elevations as $(1/\sin(E))$. For the troposphere error listed in Table 1, an elevation value of 15 deg and an elevation difference of 100 μrad is assumed for the two spacecraft. A factor of $\sqrt{2}$ accounts for the separate errors at two ground stations.

D. System Noise

The received signal contains the spacecraft signals and ground-receiver generated noise, which is proportional to the system operating temperature. The system noise error depends on the ratio of received signal power to noise power. The voltage signal-to-noise ratio (SNR_v) can be made higher by averaging over a longer time interval. The SBI phase error due to system noise is given by

$$\varepsilon_d \text{ (mm)} = \lambda\sqrt{2}/(2\pi SNR_v) \tag{3}$$

where λ is the X-band wavelength (36 mm). With nominal Mars '94 transmitter power of 1 W and antenna gain of 17 dB at X-band, a DSN 34-m antenna achieves an SNR_v value of 175 for an integration time of 5 min. The MO SNR_v will be approximately 10 times higher due to the greater effective transmitted carrier power. There is a separate error for Mars '94 (neglecting the MO SNR_v error) at each station, which results in the factor of $\sqrt{2}$.

E. Phase Dispersion

The SBI observable is generated by double differencing the measured phase of sinusoidal signals transmitted from two spacecraft and received at two stations. The ground receiver chains introduce phase shifts which depend on the Doppler shifted signal frequencies, and hence will in general be distinct for each station and for each spacecraft. Instrumental phase shifts can be divided into two categories: phase shifts which vary linearly with frequency (nondispersive) and phase shifts which have a nonlinear frequency dependence (dispersive). Nondispersive instrumental errors are estimated below in Subsection II. Dispersive errors are approximated by

$$\varepsilon_d \text{ (mm)} = 2 \times (0.5 \text{ deg}) \times \lambda/(360/\text{deg}) \tag{4}$$

where 0.5 deg is representative of the instrumental phase dispersion in the operational VLBI receiver system [9]. A separate error occurs for each spacecraft at each ground station leading to the factor of 2. The phase dispersion effects can be reduced by better instrumentation or very close spacecraft frequencies.

F. Oscillator Drift

An unknown offset between the transmitter frequencies of the two spacecraft will cause an error given by

$$\varepsilon_d \text{ (mm)} = c\tau \times \Delta f/f \tag{5}$$

where c is the speed of light, τ is the difference in reception time at the two stations (here assumed to be 10 msec), f is the nominal transmitter frequency for each spacecraft, and Δf is the unknown transmitter frequency offset. For two-way transmissions, where separate uplinks derived from independent frequency standards are used for the two spacecraft, an estimate of $\Delta f/f$ is provided by the expected accuracy of the station clock rate calibration, which is 5×10^{-14}. For one-way transmissions, line-of-sight Doppler measurements are used to estimate corrections to the nominal spacecraft onboard oscillator frequency. The accuracy to which the oscillator frequency can be estimated depends on the tracking coverage and on the oscillator stability; $\Delta f/f$ can typically be estimated with an accuracy of 2×10^{-12} for one-way transmissions.

G. Baseline

Since an angular measurement is derived from knowledge of the time of reception at two Earth ground stations, uncertainty in station location and Earth orientation degrades the interpretation of the SBI measurement. The Earth's pole orientation and rotation rate change randomly and must be monitored to maintain knowledge of these quantities. Currently at JPL, knowledge of the Earth's orientation is being maintained with an accuracy of 30 cm for real-time data analysis. For analysis of data more than two weeks old, the error in Earth orientation is less than 5 cm. The Earth orientation accuracy for real-time analysis could be improved if required and is expected to improve to the 5-cm level as measurements from the GPS are included in the coming years [10,11,12]. DSN station locations have been determined with an accuracy better than 5 cm by VLBI and satellite laser ranging [13,14]. Overall, the value of 7 cm is used to represent the baseline error due to station location and Earth orientation uncertainties. The SBI error is given by

$$\varepsilon_d \text{ (mm)} = 70 \text{ mm} \times \Delta\theta \tag{6}$$

H. Station Instrumentation

An uncalibrated group delay or clock offset in the ground station instrumentation causes a phase delay error of the form

$$\varepsilon_d \text{ (mm)} = (\dot{\rho}_1 - \dot{\rho}_2)\,\delta\tau_I \qquad (7)$$

where $\dot{\rho}_i$ is the line-of-sight range rate (mm/sec) between one station and spacecraft i, and $\delta\tau_I$ is the uncalibrated instrumentation delay (sec). This error varies slowly over a pass as the Doppler shift changes. The MO dynamics are used to bound this error, since Mars '94 is in a slower orbit. Over a 1-hr data arc, the range-rate change is bounded by 6×10^6 mm/sec. All station delays should be calibrated to 2×10^{-8} sec. The resulting SBI error due to nondispersive instrumental effects drifts by no more than 0.12 mm.

III. Orbiter–Orbiter Tracking Example

The Russian Mars '94 mission will nominally arrive at Mars in mid-1995 near the end of the MO primary mission. It will be possible to make SBI measurements at X-band by using Mars '94 and MO. Sample orbital elements for Mars '94 were chosen for the epoch June 1, 1995, at 19:50 UT. These, along with orbital elements for MO at the same epoch, are listed in Table 2. At this time, the Earth–Mars distance is 1.4 AU, the right ascension and declination of Mars as seen from Earth are 155.9 deg and 11.5 deg, respectively, and the SEP angle is 82.4 deg. A plot of the orbits in the plane of the sky as seen from Earth is shown in Fig. 3. The separation angle between the two spacecraft is always less than 100 μrad, so both spacecraft will lie within the X-band beamwidth of a 34-m antenna, which is 1.06 mrad.

Covariance analyses have been performed based on an early version of the MO Navigation Plan.[1] Both MO and Mars '94 were assumed to have area-to-mass ratios equal to 0.017 m^2/kg. For study purposes, a nominal 12-hr tracking arc was used, including Doppler data for both spacecraft from an antenna at the Goldstone, California, DSN complex. The data arc includes six orbits of MO and one orbit for Mars '94. The spacecraft modeling assumptions are outlined in Table 3. In each case, only the spacecraft epoch state is estimated. The dominant error in the following analyses is usually the "considered" (unadjusted) gravity field uncertainty. The gravity field uncertainty is based on an analysis of gravity calibration orbits for MO early in its mission.[1] These values will evolve as further studies are performed and as the MO mission progresses.

The following cases show orbit determination uncertainties over the 12-hr period during which Doppler data are scheduled. SBI data, when taken, also lie within this interval. This study is confined to investigating orbit determination accuracy for trajectory reconstruction purposes. Orbit determination for prediction purposes, which is of interest for mission operation, is more susceptible to force modeling assumptions and is not addressed here.

A. Doppler-Only Solutions

The orbit determination uncertainty for MO by using only Doppler data is presented in Fig. 4, which shows the rss position error during the 12-hr time interval. The Doppler data were weighted at 0.1 mm/sec for a 3-min integration time. The relatively low errors result from the multirevolution data span and the fairly low gravity field uncertainties derived from the gravity calibration orbit, which nonetheless still dominate the orbit determination uncertainty. For Mars '94, the Doppler data, spanning one orbit, were also weighted at 0.1 mm/sec for a 3-min integration time. The Doppler-only Mars '94 orbit determination uncertainty is shown in Fig. 5. The dominant error is the computed error (due to random measurement noise) rather than the considered gravity field error.

The Doppler-only orbit determination uncertainty for both MO and Mars '94 is limited by the ability of a single pass of Doppler data to determine the longitude of the ascending node in the plane of the sky. For the nearly circular MO orbit, the node uncertainty appears (in Fig. 4) as a twice-per-orbit signature. For the Mars '94 orbit, the node uncertainty shows up (in Fig. 5) as a once-per-orbit signature in the position uncertainty with maximum uncertainty at apoapsis.

B. Doppler Plus SBI

For this case, a joint orbit analysis was performed with SBI data employed in addition to Doppler data. SBI data were scheduled during the first four hours of the Doppler interval, during the mutual visibility period for Goldstone and the Madrid DSN complex, and for the last three hours of the Doppler interval, during the mutual visibility period for Goldstone and the Canberra DSN complex. The SBI data were weighted at 0.29 mm for a 5-min integration time. Phase biases were not estimated for this case since the Doppler-only results are sufficiently accurate to fix the phase biases, at least near the periapsis for Mars '94. The Doppler data were deweighted to 1 mm/sec to reduce sensitivity to gravity field errors in the estimation process. The orbit determination errors are shown in Figs. 6 and 7. The MO position uncertainty of about 10 m consists of approximately equal contributions of computed error and gravity field uncertainty. The SBI data accuracy of 0.29 mm (for 5-min integrations) equates to an effective

[1] P. B. Esposito, *Mars Observer Navigation Plan: Preliminary*, JPL D-3820 (internal document), Jet Propulsion Laboratory, Pasadena, California, December 16, 1988.

angular accuracy of about 36 prad, which corresponds to ~8 m at the 1.4-AU Earth–Mars distance. For MO, orbit determination accuracy is approaching the limit of the data accuracy. The Mars '94 position uncertainty of about 50 m is dominated by considered gravity field uncertainty.

C. One-Way Doppler Option for Mars '94

In order to track Mars '94 entirely from the DSN, one-way Doppler may have to be used instead of two-way Doppler. This is because the Mars '94 mission will not have two-way Doppler capability at X-band, instead relying on C-band Doppler from the C-band telemetry for routine navigation. Since the DSN will not support C-band, either Russian Doppler data or one-way Doppler at X-band are needed for orbit determination. The one-way Doppler case is also of interest since the DSN antennas can currently transmit only a single uplink frequency. By using one-way Doppler from the second spacecraft, telemetry and navigation for two spacecraft could be done from one antenna at each complex.

One-way Doppler accuracy will be limited by the stability of the oscillator on the Russian spacecraft. To examine the use of one-way X-band Doppler, the Mars '94 oscillator was assumed to have the characteristics shown in Fig. 8. This curve approximates the performance of the ultrastable oscillator on the Soviet Phobos spacecraft. One-way Doppler data were included with weight 0.75 mm/sec for a 3-min integration time, with a constant frequency offset and random walk in frequency modeled as estimated parameters to represent the spacecraft oscillator behavior.

Figure 9 shows the orbit determination accuracy for Mars '94 using only one-way Doppler for the 12-hr view period from Goldstone. The results are much worse than the two-way Doppler results of Fig. 5. However, the one-way Doppler, combined with MO two-way Doppler and SBI data, gives good results. Orbit determination accuracies that result from using 12 hr of two-way Doppler from MO weighted at 1 mm/sec, 12 hr of one-way Doppler from Mars '94, and SBI from both DSN baselines are shown in Figs. 10 and 11. In obtaining these results, phase biases were estimated. The MO accuracy shown in Fig. 10 is comparable to the results obtained when two-way Doppler and SBI data are used for both spacecraft. The Mars '94 results in Fig. 11 are not as good as the two-way Doppler-only solution for Mars '94 shown in Fig. 5. The orbital accuracy for these solutions is marginal in terms of being able to determine the correct phase biases for the SBI data. If the biases could be fixed (perhaps by using multiple tones), the orbit determination accuracy would be improved to the levels shown in Figs. 12 and 13.

IV. Orbiter–Lander Tracking Example

There are several potential missions which would place landers or rovers on the surface of Mars. The communications capability of the landers is not yet known. But in order to examine the utility of tracking an orbiter with respect to a lander, this study arbitrarily included a lander located at Mars latitude of +26 deg and longitude of 140 deg. This lander was assumed to be able to communicate with the Earth at X-band and be capable of supporting two-way Doppler tracking from DSN stations. The lander was tracked at the same time as the MO spacecraft in the above cases. Because of the relative rotation of Earth and Mars, the lander was visible from Goldstone for only the first four hours of the data period. An orbit solution covariance was calculated with 12 hr of Doppler data for MO, weighted at 1 mm/sec, 4 hr of Doppler data from the lander weighted at 1 mm/sec, and 4 hr of SBI data from the Madrid–Goldstone baseline weighted at 0.2 mm. The SBI phase bias was presumed to be fixed. The position of the lander with respect to the center of Mars was assumed to be known a priori to 100 km and was estimated along with the spacecraft state.

The resulting orbital accuracy for MO is shown in Fig. 14 and is comparable to the orbiter–orbiter tracking results shown in Fig. 10. The estimated lander position accuracy, given in Fig. 15, is a few meters in spin radius and longitude and 20 m in height (Z) above Mars' equator. This position accuracy is good enough that random orientation changes for Mars, analogous to terrestrial polar motion and rotation rate changes, will become observable. This suggests that orbiter–lander tracking, or differential tracking between multiple landers, can be used for studies of Mars rotation in addition to navigation.

Another tracking scenario studied included one-way Doppler from the lander (with the same assumed oscillator as for the one-way Mars '94 study above) in addition to two-way Doppler from MO and SBI data. The orbit determination errors for MO were found to be comparable to those shown in Fig. 14. However, the estimated lander position uncertainty increased to 20 m in longitude and spin radius and 80 m in Z. Thus, one-way Doppler from the lander is adequate for determination of the orbiter trajectory. For accurate location of the lander, either two-way Doppler must be used or some other strategy, such as the use of a longer data arc or a more stable oscillator, must be adopted.

V. Conclusion

Same-beam interferometry data, combined with two-way Doppler or a combination of two-way and one-way

Doppler, has the potential to be a powerful orbit determination data type and allows multiple spacecraft to be tracked simultaneously. Orbit determination studies using MO and the Russian Mars '94 spacecraft have predicted accuracy improvements of an order of magnitude or more over Doppler-only orbit solutions for short data arcs. Also, it should be possible to obtain 1-km orbit determinati accuracy for Mars '94 in downlink-only mode when track with respect to MO. SBI tracking of a lander on Mars and Mars orbiter can potentially yield position accuracies f Mars landers at the few-meter level by using single-d data arcs.

Acknowledgments

The authors thank A. S. Konopliv for help with the covariance software used and R. D. Kahn for assistance with the solar plasma error model.

References

[1] S. N. Mohan and P. B. Esposito, "Venus Radar Mapper Orbit Accuracy Analysis," paper 84-1986, presented at the AIAA/AAS Astrodynamics Conference, Seattle, Washington, August 1984.

[2] J. S. Border and W. M. Folkner, "Differential Spacecraft Tracking by Interferometry," paper CNES-89-145, presented at the CNES International Symposium on Space Dynamics, Toulouse, France, November 1989.

[3] J. S. Border and R. D. Kahn, "Relative Tracking of Multiple Spacecraft by Interferometry," paper 89-178, presented at the AAS/GSFC International Symposium on Orbital Mechanics and Mission Design, Greenbelt, Maryland, April 1989.

[4] W. M. Folkner and J. S. Border, "Orbit Determination for Magellan and Pioneer 12 Using Same-Beam Interferometry," paper AAS 91-393, presented at the AAS/AIAA Astrodynamics Specialist Conference, Durango, Colorado, August 19–22, 1991.

[5] R. D. Kahn, W. M. Folkner, J. S. Border, and A. J. Vijayaraghavan, "Position Determination of a Lander and Rover at Mars With Earth-Based Differential Tracking," *TDA Progress Report 42-108*, vol. October–December 1991, Jet Propulsion Laboratory, Pasadena, California, pp. 279–293, February 15, 1992.

[6] J. S. Border, F. F. Donivan, S. G. Finley, C. E. Hildebrand, B. Moultrie, and L. J. Skjerve, "Determining Spacecraft Angular Position with Delta VLBI: The Voyager Demonstration," paper 82-1471, presented at the AIAA/AAS Astrodynamics Conference, San Diego, California, August 1982.

[7] R. D. Kahn and J. S. Border, "Precise Interferometric Tracking of Spacecraft at Low Sun–Earth–Probe Angles," paper 88-0572, presented at the AIAA 26th Aerospace Sciences Meeting, Reno, Nevada, January 1988.

[8] G. E. Lanyi and T. Roth, "A Comparison of Mapped and Measured Total Ionospheric Electron Content Using Global Positioning System and Beacon Satellite Observations," *Radio Science*, vol. 23, no. 4, pp. 483–492, July–August 1988.

[9] N. C. Ham, "VLBI System (BLK 1) IF-Video Down Conversion Design," *TDA Progress Report 42-79*, vol. July–September 1984, Jet Propulsion Laboratory, Pasadena, California, pp. 172–188, November 15, 1984.

[10] A. P. Freedman, "Measuring Earth Orientation with the Global Positioning System," *Bulletin Geodesique*, vol. 65, pp. 53–65, 1991.

[11] S. M. Lichten, S. L. Marcus, and J. O. Dickey, "Sub-Daily Resolution of Earth Rotation Variations with Global Positioning System Measurements," *Geophys. Res. Lett.* (in press).

[12] U. J. Lindqwister, A. P. Freedman, and G. Blewitt, "Daily Estimates of the Earth's Pole Position with the Global Positioning System," *Geophys. Res. Lett.* (in press).

[13] J. R. Ray, C. Ma, J. W. Ryan, T. A. Clark, R. J. Eanes, M. M. Watkins, B. E. Schutz, and B. D. Tapley, "Comparison of VLBI and SLR Geocentric Site Coordinates," *Geophys. Res. Lett.*, vol. 18, no. 2, pp. 231–234, February 1991.

[14] M. H. Finger and W. M. Folkner, "A Determination of the Radio-Planetary Frame Tie and the DSN Tracking Station Locations," paper AIAA 90-2905, presented at the AIAA/AAS Astrodynamics Conference, Portland, Oregon, August 20–22, 1990.

Table 1. X-band same-beam error budget (for 5-min integration time and 20-deg Sun–Earth–Mars angle).

Error source	Error, mcycles	Error, mm	Error, prad
Solar plasma	6.2	0.22	28
Ionosphere	0.6	0.02	3
Troposphere	2.2	0.08	10
System noise	1.3	0.05	6
Phase dispersion	2.8	0.10	13
Spacecraft oscillator drift	0.2	0.01	1
Instrumentation	3.3	0.12	15
Baseline	0.2	0.01	1
RSS total	8.0	0.29	36

Table 2. Spacecraft orbital elements referred to the Martian equator of date.

Element	Mars Observer	Mars '94
Semimajor axis, km	3749.288	12650.0
Eccentricity	0.00382	0.7
Inclination, deg	92.821	105.0
Argument of perigee, deg	−90.0	−90.0
Longitude of ascending node, deg	322.773	0.0
Mean anomaly, deg	0.0	0.0
Period, hr	1.94	12.0
Epoch	June 1, 1995, 19:50 UTC	June 1, 1995, 19:50 UTC

Table 3. Assumptions for orbit determination covariance analysis.

Adjusted parameters	A priori sigma
Spacecraft epoch state position	10^4 km per component
Spacecraft epoch state velocity	10 km/sec per component
Unadjusted parameters	**A priori sigma**
Solar reflection coefficients	10 percent of nominal value
Atmospheric drag coefficient	20 percent of nominal value
Bias acceleration	10^{-12} km/sec^2 per component
Mars GM[a]	3.5×10^{-6} × nominal GM
Mars gravity field (spherical harmonics and mascons)	Errors from MO calibration orbit
Station locations (including UT1–UTC and polar motion)	7 cm per component
Zenith troposphere	4 cm
Line-of-sight ionosphere	3 cm

[a] GM = (Mass of Mars) + Newton's gravitational constant

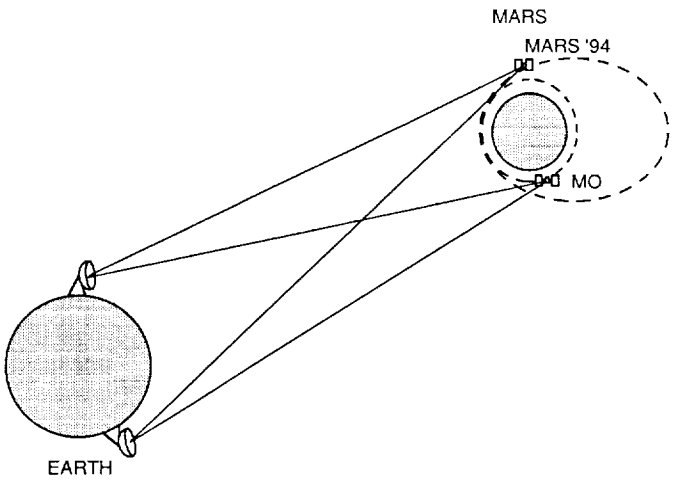


Fig. 1. Same-beam interferometry technique.

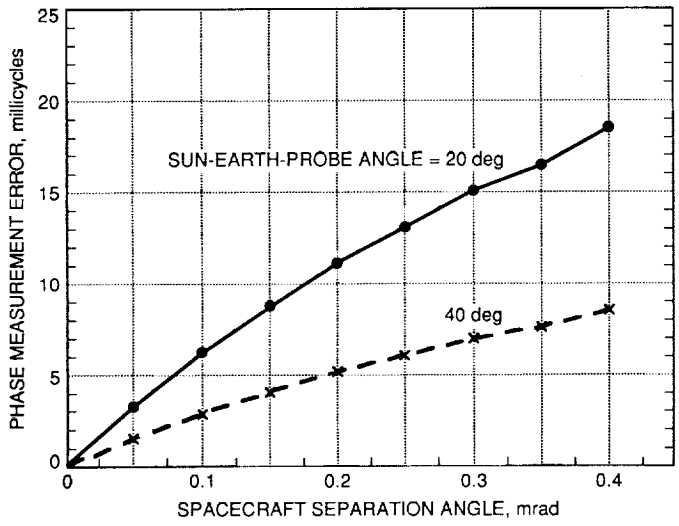


Fig. 2. Effect of solar plasma on X-band same-beam interferometry data for various spacecraft separation angles and Sun-Earth-Probe angles as calculated from model of Kahn (Ref. 5).

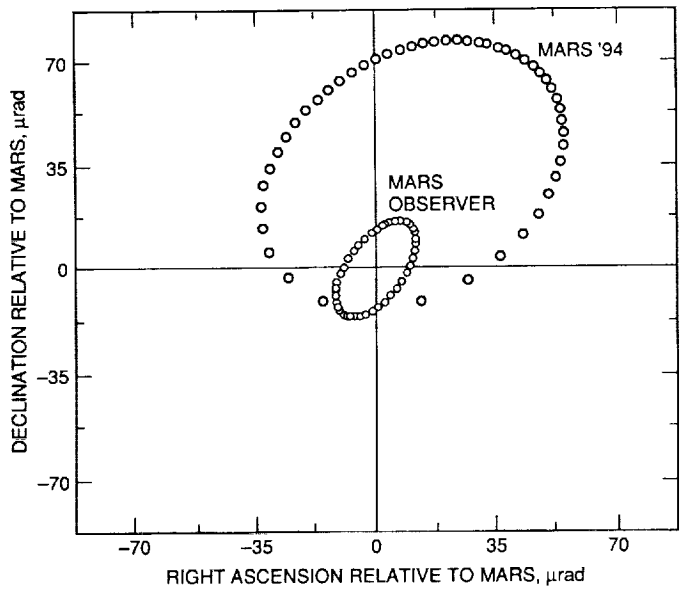


Fig. 3. MO and Mars '94 orbits centered about Mars as seen from Earth on 1-June-1995.

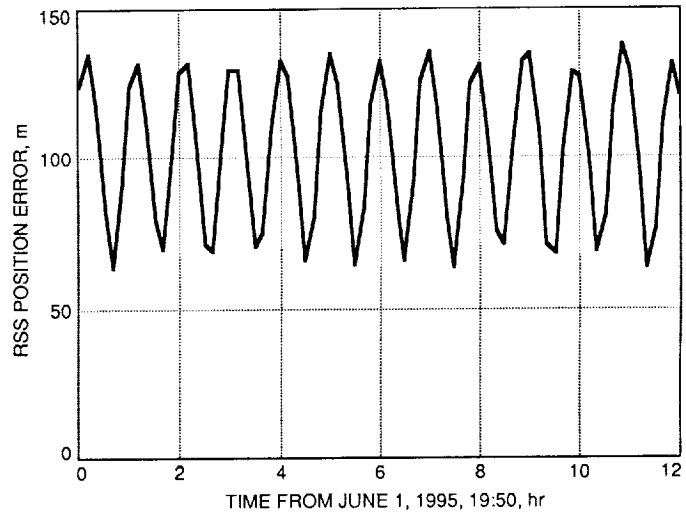


Fig. 4. Doppler-only orbit determination accuracy for Mars Observer.

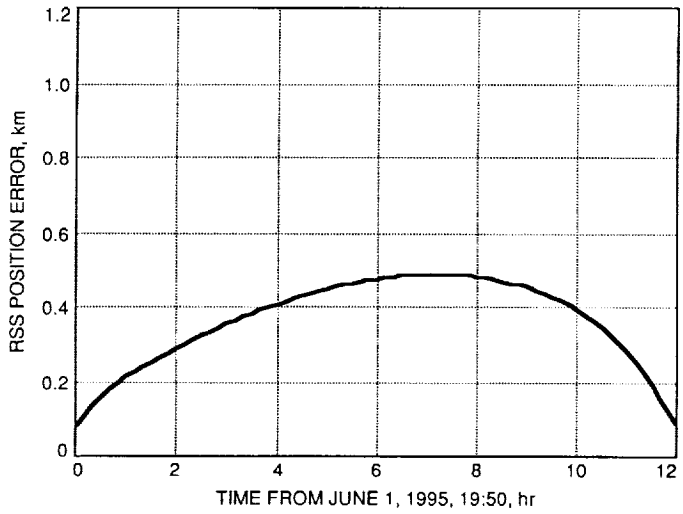


Fig. 5. Doppler-only orbit determination accuracy for Mars '94.

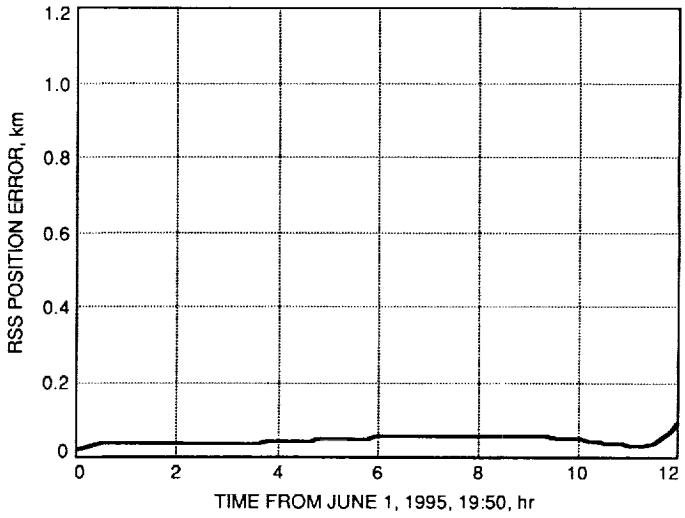


Fig. 7. Doppler plus SBI orbit determination accuracy for Mars '94.

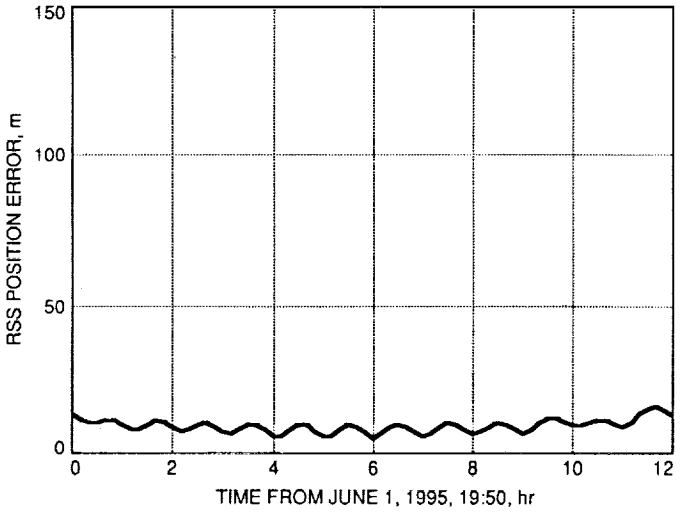


Fig. 6. Doppler plus SBI orbit determination accuracy for Mars Observer.

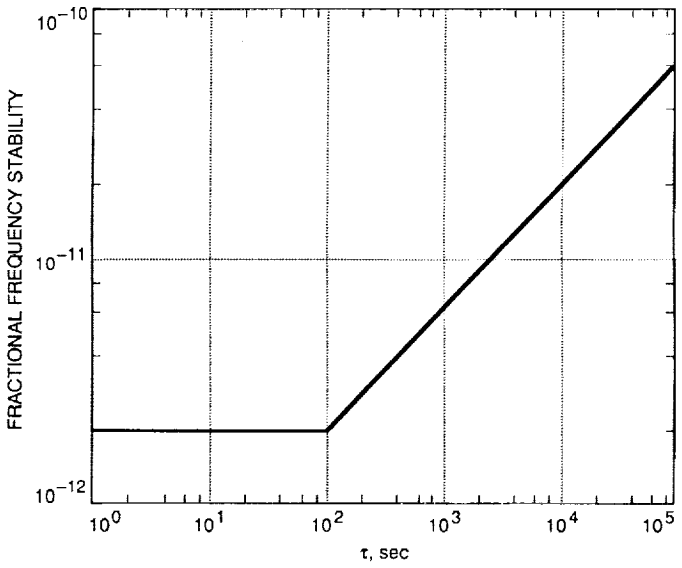


Fig. 8. Assumed stability for Mars '94 oscillator.

C-2

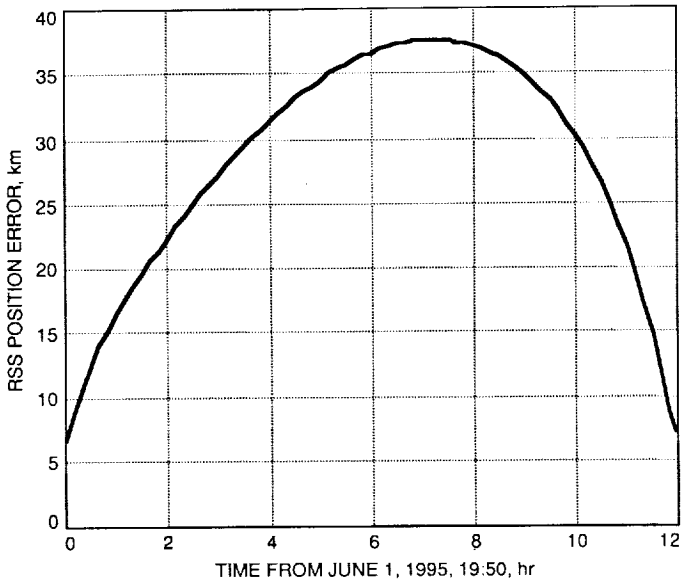


Fig. 9. One-way Doppler orbit determination accuracy for Mars '94.

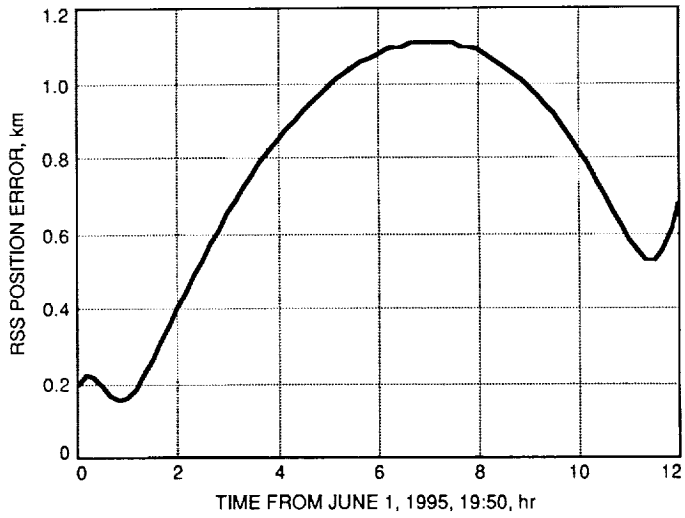


Fig. 11. Orbit determination accuracy for Mars '94 with one-way Doppler combined with bias-estimated SBI and two-way Doppler from Mars Observer.

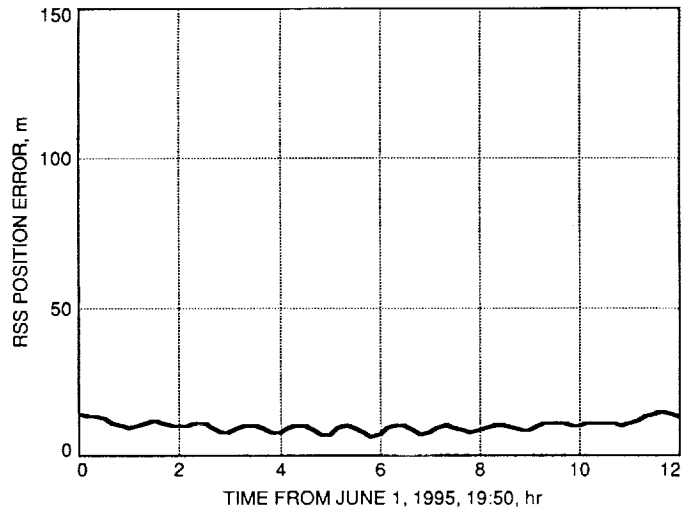


Fig. 10. Orbit determination accuracy for Mars Observer with two-way Doppler combined with bias-estimated SBI and one-way Doppler from Mars '94.

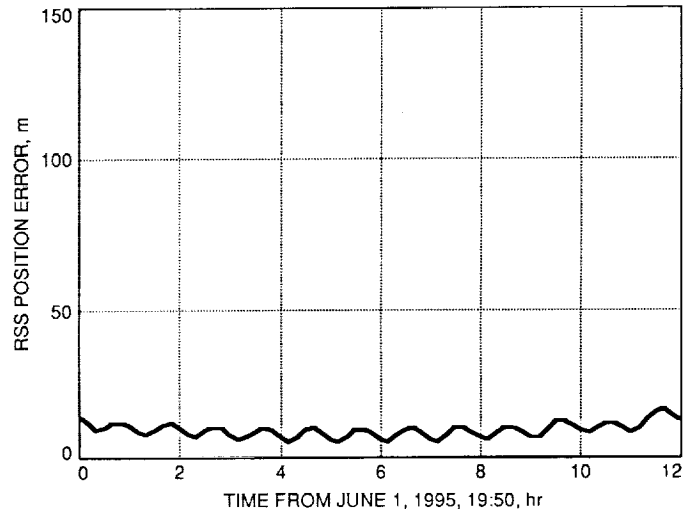


Fig. 12. Orbit determination accuracy for Mars Observer with two-way Doppler combined with bias-fixed SBI and one-way Doppler from Mars '94.

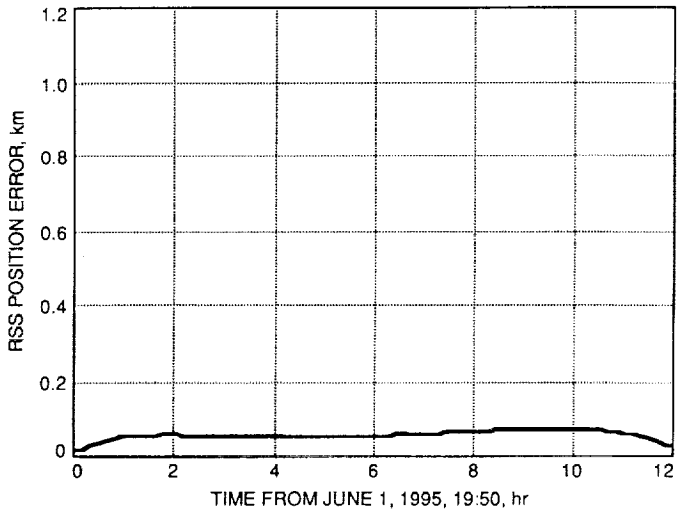


Fig. 13. Orbit determination accuracy for Mars '94 with one-way Doppler combined with bias-fixed SBI and two-way Doppler from Mars Observer.

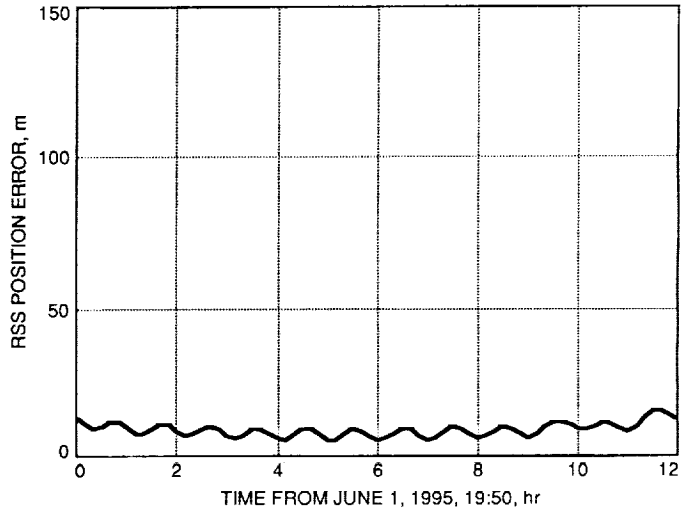


Fig. 14. Orbit determination accuracy for Mars Observer with two-way Doppler combined with SBI and two-way Doppler for the lander.

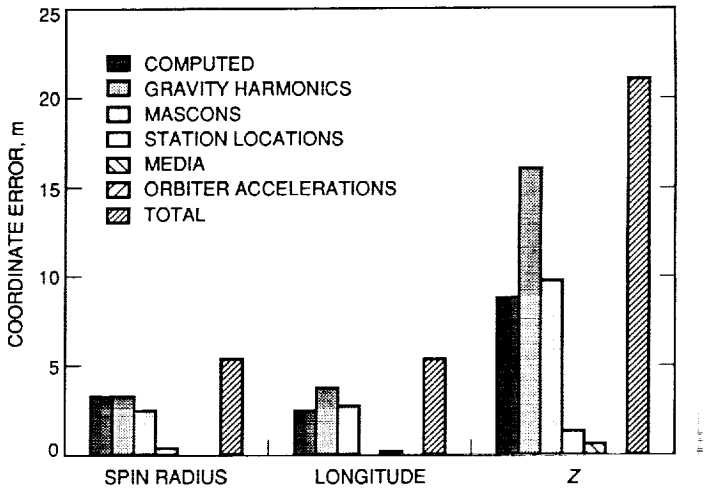


Fig. 15. Mars lander position determination using Doppler and SBI tracking with respect to an orbiter.

Geometric Distortion Analysis of a Wide-Field Astrograph

W. M. Owen, Jr.
Navigation Systems Section

S. B. Shaklan
Optical Sciences and Applications Section

Ground-based optical navigation seeks to determine the angular position of a star, Solar System body, or laser-emitting spacecraft relative to objects with well-known coordinates. Measurement accuracies of 25 nrad would make optical techniques competitive with current radio metric technology. This article examines a proposed design for a wide-field astrograph and concludes that the deviation of an image centroid from the ideal projection can be modeled to the desired accuracy provided that the field of view does not exceed 5 deg on a side.

I. Introduction

Astrometry, the science of measuring angular positions of celestial objects, is currently in renaissance thanks to new instrumentation such as the Multichannel Astrometric Photometer [1], the Hipparcos spacecraft [2], and the Mark III stellar interferometer [3]. These devices, whether orbiting or ground-based, have increased the precision of an angular measurement by two orders of magnitude compared to conventional photographic or transit-circle techniques.

Optical angular measurements accurate to 25 nrad (5 mas) could revolutionize JPL's spacecraft navigation as well. Ground-based observations of asteroid 951 Gaspra, while good to only 0.1 arcsec, nevertheless contributed greatly[1] to the successful flyby of that asteroid by Galileo in 1991. Measurements of the satellites of Jupiter and of Saturn could provide accurate ephemerides of these objects, thereby improving the power of onboard navigation images taken by the Galileo or Cassini spacecraft. And direct optical measurements of the laser light emitted by a future spacecraft can determine its trajectory position relative to its target.

One of the tasks necessary in defining a future optical navigation system is to assess the capability of various instruments to deliver relative positions at the 5-mas (25-nrad) level[2] over a relatively wide (5- or 6-deg) field of view. The instrument itself must perform at that level; the problem of measuring images in the focal plane is a separate issue.

This article reports on the optical characteristics of one candidate, a proposed wide-field astrograph designed by the United States Naval Observatory. This instrument

[1] D. K. Yeomans and M. S. Keesey, "Updated Orbit and Ephemeris for Asteroid 951 Gaspra," JPL Interoffice Memorandum 314.6-1332 (internal document), Jet Propulsion Laboratory, Pasadena, California, August 30, 1991. More detailed accounts of the Gaspra encounter navigation will be presented at the AIAA/AAS Astrodynamics Conference in Hilton Head, South Carolina, in August 1992.

[2] G. Null, "Wide-Field Telescope Selection Results," JPL Interoffice Memorandum 314.8-815 (internal document), Jet Propulsion Laboratory, Pasadena, California, December 4, 1991.

has a clear aperture of 36 cm, a focal length of 3.6 m, a plate scale of 57.3 arcsec/mm, and a 6-deg field of view; the focal length in particular is similar to those of the 51-cm astrographs at Lick Observatory and Yale University Observatory.

Analysis of this instrument involved simulating stellar images at various places in the focal plane, finding the centroids of each image, and developing a model to account for systematic differences between the observed positions of these centroids and their ideal positions as predicted by a pinhole projection. The observed centroids can deviate from the ideal projection by almost 0.4 arcsec, necessitating the use of higher order terms in the transformation model. However, most of the deviation can be accommodated by one third-degree radial term, and the remainder is on the order of 2–7 mas.

The authors conclude that the instrument under consideration will indeed be suitable for 5-mas astrometry over a 5-deg × 5-deg field of view.

II. Image Generation

The Controlled Optics Modeling Package (COMP) was used to simulate images in the focal plane, at steps of 0.5 deg (roughly 31 mm) from −3 deg to +3 deg in both x and y. The symmetry of the problem allowed generation of only those 28 images in the first quadrant for which $x \geq y$, marked by the filled dots in Fig. 1; the others follow from these. A total of $13^2 = 169$ image points was therefore effectively obtained.

The COMP program contains both ray trace and diffraction analysis capabilities. The advantages of this program are that it can automatically calculate highly aberrated or off-axis diffraction patterns while providing direct access to optical parameters for modification and sensitivity testing. COMP is not a design tool, but an analysis subroutine that can automatically generate sets of images or sensitivity tables.

For this astrograph study, 902 rays were traced through the system to produce each image. The rays were traced to the focal plane, then back-propagated to the exit pupil onto a spherical reference surface whose center of curvature was defined by the intersection of the chief ray and the focal plane. The spherical phase term was removed, and a Fourier transform was performed in order to calculate a monochromatic image. The image was stored as a 100 × 100 array of brightness values, with each number representing the intensity within a 1-μm square in the focal plane. Subsequent wavelengths were treated in the same fashion, except that the reference sphere was always centered on the same point as the first one. The five monochromatic images (at λ = 705.6, 650.0, 610.0, 587.6, and 546.0 nm) for each source position were then coadded to produce the final image files.

Images formed by the astrograph are diffraction limited on-axis, and show a slight coma at 3 deg off-axis. The first Airy ring is not broken, indicating weak aberration. Chromatic aberration was negligible.

One important question is to determine how much distortion is introduced by the finite number of rays, the finite number of Fourier grid points, the stepwise change in aperture representation due to vignetting, and pixellation effects in the image plane. One metric of the amplitude of these effects is to compare the ray centroids to diffraction centroids. One finds that diffraction centroids are within 0.8 mas of ray centroids throughout the field, which is well within the accuracy required for this study. A second metric is simply to increase the number of rays and grid points. Again, the distortion did not significantly change.

A second question involves the accuracy of the polychromatic approximation in this modeling. Because only five wavelengths were used to simulate a broad band, the composite Airy pattern is not a true representation of the broadband image. However, an empirical study showed that the major difference occurred only near the first and second ring minima. Because the images remain highly symmetric throughout the field, the small error does not significantly shift the centroids. Further, the ring minima regions are not particularly important in image centroiding.

III. Image Centroiding

Image centroids for each image file were determined by using a data number (DN) filter. Each image is modeled as the sum of a point-spread function (PSF) and a constant background; the PSF itself is smoothed from a square array of brightness values. The coordinates of the center of the PSF, the amplitude of the PSF, and the background level are the unknowns in an iterated linear least-square process. The on-axis image was used as the PSF for all images. The brightness values (DN's) in each pixel of an image form the observation set. The variance in the DN in each pixel was composed of two parts: a constant to represent read noise, and a part proportional to the DN level itself to represent photon statistics. The resulting

centroids carry a formal error of ± 0.025 μm in the focal plane (1.4 mas on the sky). The centroids were first computed in pixel and line coordinates within each image file, then transformed into millimeter coordinates in the focal plane.

IV. Plate Constants

Given the image centroids determined above and the ideal coordinates of each image, one seeks a transformation from measured to ideal coordinates. Following [4] or [5], denote the measured coordinates by (x, y) and the ideal coordinates by (ξ, η). Model the transformation as a polynomial in each coordinate:

$$\xi = \sum_i \sum_j a_{ij} x^i y^j$$

and similarly for η. The two coordinates are handled independently, and symmetry allows one to examine only ξ. The a_{ij} are the solution parameters (the plate constants, for in classical astrometry these are found for each photographic plate from reference star images on that plate). The best model is generally considered to be the one with the fewest parameters such that the fit is acceptable.

From symmetry, any displacement of an image centroid from its ideal location must be in the radial direction, as in this analysis all the astrograph lenses were perfectly aligned. Aberration theory in geometrical optics also dictates that the displacement must be proportional to an odd power of the separation r of the image from the optical axis. Thus, the leading nonlinear term in x is xr^2, which is represented in the plate model by the two coefficients a_{30} and a_{12}. Trial solutions confirmed that the only nonzero terms in the plate model were those with i odd and j even. Accordingly, one can construct a more restrictive model,

$$\xi = \sum_k b_k r^{2k} x$$

in which the only displacements allowed were proportional to odd powers of r.

Table 1 gives coefficients for various solutions: those in terms of a_{ij} on the top, and those in terms of b_k at the bottom. Each of the maximum degree numbers represents the highest degree of the solution; this is the largest value of $(i + j)$ or $(2k + 1)$ as appropriate. The formal errors of the coefficients are at most 1 in the final digit shown.

Table 2 summarizes the post-fit residuals for each case. The residuals generally reach a maximum in the corners of the field, but there is a secondary maximum not far from the origin. The table gives these values as well as the rms residual for all 169 images.

It is evident that low-order fits do not model the actual image centroid locations well in the corners of the full 6-deg × 6-deg field. The analysis was therefore repeated using only a 5-deg × 5-deg field by omitting the first and last row and column of images. The results appear in Tables 3 and 4. Now a third-degree fit meets the 5-mas requirement even in the corners of the field.

V. Conclusions and Discussion

The ideal projection is obviously not sufficient for a full 6-deg × 6-deg field, as distortions in this instrument reach 0.385 arcsec (in each coordinate) in the corners of the field. A third-degree fit removes most of the distortion, leaving residuals on the order of 2–7 mas or 10–35 nrad. In order to achieve the design goal of 25 nrad, it will be necessary to include the fifth-degree term or else to restrict the field of view to 5-deg square.

It is clear from the size of the residuals (Table 2) that it should be sufficient to model the astrograph by

$$\xi = b_0 x + b_1 r^2 x + b_2 r^4 x$$

the full generality of the a_{ij} should not be necessary. This is true, however, only if the optics are perfectly aligned. Any one of the five lenses can move in any direction by up to 8 μm without introducing more than 5 mas of additional aberration.[3] Random motions of this size are not likely (except for thermal expansion), as the lenses would be rigidly mounted in a cell. However, it is not at all clear that the assembly process can be controlled to that tolerance. Although one can expect stability, one should also expect the instrument to be misaligned, and these inevitable misalignments will modify the radial distortion pattern of the perfectly aligned system. Accordingly, the

[3] S. Shaklan and J. Yu, "Images Formed With 36-cm Astrograph: Results to Supersede OSAS DFM 91-166," JPL Interoffice Memorandum OSAS DFM 91-173 (internal document), Jet Propulsion Laboratory, Pasadena, California, November 8, 1991.

full model in terms of a_{ij} should be employed if the instrument is built.

In practice, most astrometrists are loath to include terms higher than third degree in their plate solutions, simply because most star fields do not contain enough reference stars to support the higher terms. Also, as mentioned above, in most cases the plate constants are determined independently for each plate, because non-instrumental effects such as atmospheric refraction and dispersion will randomly corrupt the instrumental distortions. These real-world considerations do not invalidate this study. Rather, the instrumental distortion model should be found by testing the optics in the laboratory, and the model parameters and their covariance used as a priori information when reducing actual observations.

Outside the scope of this article is the requirement tc measure image location to the 5-mas level. For this instrument, this requirement translates into a measurement precision (and accuracy) of about 0.1 μm. Maintaining this accuracy over a region nearly 40 cm on a side may be challenging.

In conclusion, the candidate astrograph design suffers from radial distortion but that can be modeled with one parameter provided that the field of view is 5 deg on a side. Including terms to fifth degree will produce a model good to 1 mas and will allow the use of a 6-deg field, but the coefficients may be difficult to determine in practice. The instrument design is therefore viable provided that detectors can be found to exploit its excellent optical qualities.

Acknowledgment

The work of Jeffrey Yu to help create the image files for this analysis is gratefully acknowledged.

References

[1] G. D. Gatewood, "The Multichannel Astrometric Photometer and Atmospheric Limitations in the Measurement of Relative Positions," *Astron. J.*, vol. 94, no. 1, pp. 213–224, July 1987.

[2] M. A. C. Perryman et al., *The Hipparcos Satellite*, European Space Agency, Publication SP–1111, vols. I–III, Paris, France, June 1989.

[3] M. Shao, M. M. Colavita, B. E. Hines, D. H. Staelin, D. J. Hutter, K. J. Johnston, D. Mozurkewich, R. S. Simon, J. L. Hershey, J. A. Hughes, and G. H. Kaplan, "The Mark III Stellar Interferometer," *Astron. Astrophys.*, vol. 193, pp. 357–371, 1988.

[4] A. König, "Astronomy with Astrographs," in *Astronomical Techniques: Vol. 2, Stars and Stellar Systems,* edited by A. Hiltner, Chicago: University of Chicago Press, pp. 461–486, 1962.

[5] H. Eichhorn, *Astronomy of Star Positions*, New York: Frederick Ungar Publishing Co., pp. 68–76, 1974.

Table 1. Solution parameters for 6-deg field.

Parameter	Maximum degree			
	1	3	5	7
a_{10}	0.9999655948	1.000008736	1.000009976	1.000010197
a_{30}		−0.0011189094	−0.0011969770	−0.0012239529
a_{12}		−0.0011063839	−0.0011871044	−0.0012123322
a_{50}			0.0011528	0.002152
a_{32}			0.0018920	0.003283
a_{14}			0.0009628	0.001705
a_{70}				−0.0117
a_{25}				−0.0196
a_{34}				−0.0147
a_{16}				−0.0070
b_0	0.9999655948	1.000008630	1.000009925	1.000010154
b_1		−0.0011116554	−0.0011895162	−0.001214510
b_2			0.00097727	0.0017079
b_3				−0.00615

Table 2. Residual statistics for 6-deg field.

Using a_{ij}	Maximum degree			
	1	3	5	7
RMS residual, μm	2.1277	0.0349	0.0043	0.0021
RMS residual, mas	121.91	2.00	0.25	0.12
Worst central residual, μm	3.233	0.073	0.014	0.007
Worst central residual, mas	185.2	4.2	0.8	0.4
Worst corner residual, μm	6.720	0.118	0.009	0.003
Worst corner residual, mas	385.0	6.8	0.5	0.2
Using b_k	Maximum degree			
	1	3	5	7
RMS residual, μm	2.1277	0.0369	0.0050	0.0027
RMS residual, mas	121.91	2.11	0.28	0.15
Worst central residual, μm	3.233	0.078	0.015	0.009
Worst central residual, mas	185.2	4.5	0.9	0.5
Worst corner residual, μm	6.720	0.125	0.010	0.004
Worst corner residual, mas	385.0	7.2	0.6	0.2

Table 3. Solution parameters for 5-deg field.

Using a_{ij}	Maximum degree			
	1	3	5	7
a_{10}	0.9999779128	1.000009343	1.000010126	1.000010264
a_{30}		−0.0011415289	−0.001211976	−0.00122835
a_{12}		−0.0011294587	−0.001200062	−0.00122944
a_{50}			0.0014983	0.001925
a_{32}			0.0023167	0.004635
a_{14}			0.0011932	0.002338
a_{70}				0.00
a_{25}				−0.04
a_{34}				−0.04
a_{16}				−0.01
b_0	0.9999779128	1.000009268	1.000010085	1.000010271
b_1		−0.0011344615	−0.0012033140	−0.001232187
b_2			0.00121582	0.002412
b_3				−0.0143

Table 4. Residual statistics for 5-deg field.

Using a_{ij}	Maximum degree			
	1	3	5	7
RMS residual, μm	1.0925	0.0153	0.0025	0.0014
RMS residual, mas	62.59	0.88	0.15	0.08
Worst central residual, μm	2.034	0.040	0.009	0.006
Worst central residual, mas	116.5	2.3	0.5	0.3
Worst corner residual, μm	3.841	0.052	0.005	0.003
Worst corner residual, mas	220.1	3.0	0.3	0.2
Using b_k	Maximum degree			
	1	3	5	7
RMS residual, μm	1.0925	0.0163	0.0030	0.0018
RMS residual, mas	62.59	0.94	0.17	0.10
Worst central residual, μm	2.034	0.055	0.010	0.005
Worst central residual, mas	116.5	3.2	0.6	0.3
Worst corner residual, μm	3.841	0.043	0.005	0.003
Worst corner residual, mas	220.1	2.5	0.3	0.2

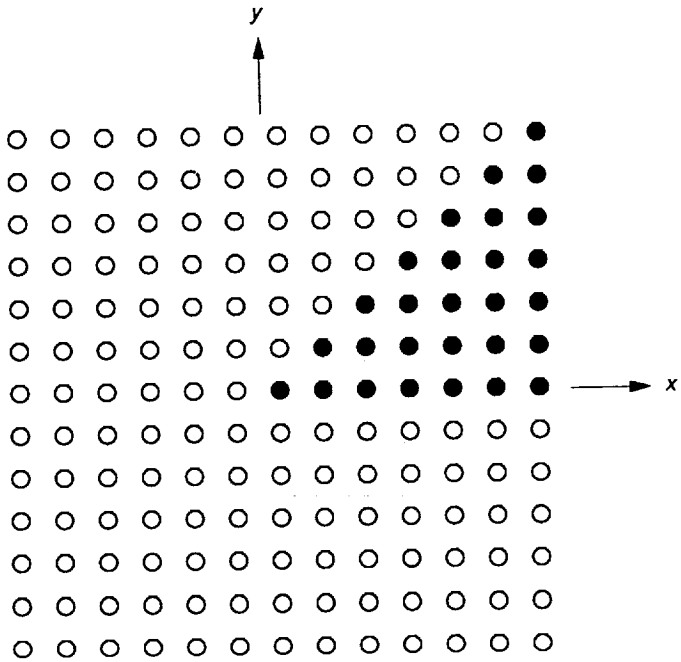


Fig. 1. Layout of images in the astrograph focal plane. Points are placed at 0.5-deg intervals from −3 to +3 deg in each coordinate. Filled circles correspond to images actually produced; open circles were derived from symmetry considerations.

57-35
104483
P-11

N92-29371

Applications of Inertial-Sensor High-Inheritance Instruments to DSN Precision Antenna Pointing

R. E. Goddard
Guidance and Control Section

Laboratory test results of the initialization and tracking performance of an existing inertial-sensor-based instrument are given. The instrument, although not primarily designed for precision antenna pointing applications, demonstrated an on-average 10-hour tracking error of several millidegrees. The system-level instrument performance is shown by analysis to be sensor limited. Simulated instrument improvements show a tracking error of less than 1 mdeg, which would provide acceptable performance, i.e., low pointing loss, for the DSN 70-m antenna subnetwork, operating at Ka-band (1-cm wavelength).

I. Introduction

Inertial sensors provide data on acceleration and spin vectors with respect to a stationary frame of reference. These data are processed by a strap-down instrument to provide the pointing coordinates of a large-structure antenna. The strap-down instrument is retrofit onto the primary reflector of the large-structure antenna, thus bypassing many error sources, such as gimbal encoders,[1] azimuth plane tilt, axis misalignments, and the Master Equatorial System.[2] A high-inheritance instrument is an existing instrument with minor modifications. Utilization of high-inheritance instruments yields cost savings in design, analysis, manufacturing, and test. In this article, the pointing performance of an inertial-sensor-based high-inheritance instrument is given. The inertial instrument is a navigator, capable of providing attitude (e.g., azimuth and elevation antenna coordinates) from the output of a triad of ring laser gyroscopes and accelerometers [1,2]. As a reference to antenna pointing requirements, the DSN 70-m antenna at Ka-band (1-cm wavelength) would have a half-power beamwidth of around 8 mdeg, so the goal would be errors in the region of a millidegree or less.

[1] C. Guiar, F. Lansing, R. Reynolds, and W. Merrick, "Sources of Pointing Error and Preliminary Estimates for the 64-Meter Antenna," JPL Interoffice Memorandum 3321-86-17 (internal document), Jet Propulsion Laboratory, Pasadena, California, February 28, 1986.

[2] H. McGinness, "Master Equatorial Pointing Errors at DSS 14," JPL Interoffice Memorandum 3553-84-006 (internal document), Jet Propulsion Laboratory, Pasadena, California, January 31, 1984.

II. Ring Laser Gyroscopes

Ring laser gyroscopes (RLG's) are optical devices, the operation of which is based on the Sagnac effect [3]. The

basic RLG components are given in Fig. 1. Two counter-rotating laser beams produce an interference pattern that is stationary in inertial space. Rotation of the RLG about the sensitive axis causes the interference pattern to pass over a photodetector. The output is electronically processed for zero-crossing detection and establishes one of the basic RLG error parameters: the quantization factor (Q). Two additional error sources complete the first-order error RLG model: angle random walk (q) caused by spurious emissions and bias instability (b) caused by out-of-plane mechanical reactions to the optical path-length controller.

Two stochastic error models of RLG's are shown in Figs. 2 and 3. These differ in the manner in which bias instability is modeled. In the ramp model (Fig. 2), instability is modeled as a random initial constant. At the beginning of each DSN tracking mode, a new random constant is assumed to initialize the rate integrator, thus producing a random ramp in position. This model is not useful when the bias instability changes significantly during the tracking mode (nominally 10 hr). When it is necessary to model a nonconstant bias instability, the exponential model (Fig. 3) is used. The RLG error statistics can be computed from the assumed form of the models.

Neglecting for the moment the quantization error Q, the sensor-limited performance can be computed in terms of the output variance σ_θ^2. Define an operator E[•] to be the statistical expectation of a random process. The output variance of the error models, assuming zero-mean disturbances, is $\mathrm{E}[\theta(t)\theta(t)]$. Evaluation of the statistics of Fig. 2 yields (see Appendix)

$$\sigma_\theta^2(t) = q^2 t + b^2 t^2 \tag{1}$$

The statistics of the model of Fig. 3 are also given in the Appendix as Eq. (A-2). In the following section, a method of identifying the strength of RLG error sources from recorded test data is given.

III. Allen Variance

Allen [4] developed a postprocessing technique to characterize the strength of noise terms in error models. The Allen Variance (AV) measure is defined as:

$$\sigma_{\theta_{\mathrm{AV}}}^2 = \frac{1}{2\tau^2}\mathrm{E}[(\theta_{k+2m} - 2\theta_{k+m} + \theta_k)^2] \tag{2}$$

where

$$\theta_k = \int^{k\tau_0} \dot{\theta}(t')dt' \tag{3}$$

where τ_0 is the sampling time and $\tau = m\tau_0$ is m samples. Typically, the AV assumes power-law spectral densities (PSD's) as uncorrelated error sources in angle rate:

$$s_{\dot{\theta}} = \sum_{\alpha=-2}^{\alpha=+2} h_\alpha f^\alpha \tag{4}$$

For RLG's, two error sources dominate and h_α is required to be nonzero only for $\alpha = 0$ (angle random walk) and $\alpha = -1$ (bias instability). The addition of the quantization noise PSD given by Papoulis [5] allows explicit computation of Eq. (2). The result for the noise-equivalent rate is

$$\sigma_{\dot{\theta}_{\mathrm{AV}}}^2(\tau) = 2(\log_e 2)b^2 + \frac{q^2}{\tau} + \frac{Q^2}{4\tau^2} \tag{5}$$

In terms of the noise-equivalent angle, Eq. (5) becomes

$$\sigma_{\theta_{\mathrm{AV}}}^2(\tau) = 2(\log_e 2)b^2\tau^2 + q^2\tau + \frac{Q^2}{4} \tag{6}$$

Thus the strength of noise sources is obtained by plotting the AV on a log–log plot and identifying each noise source by the slope.

Because Eq. (2) is the ensemble average over large numbers of identical experiments, practical methods of computing the AV from long-time series data on an individual RLG have been developed [6,7]. A practical approximation to the AV from N angle measurements is

$$\hat{\sigma}_{\theta_{\mathrm{AV}}}^2 = \frac{1}{2\tau^2(N-2m)} \sum_{k=1}^{N-2m} (\theta_{k+2m} - 2\theta_{k+m} + \theta_k)^2 \tag{7}$$

where N is the number of original time samples of length τ_0 and m is a parameter designating the size of the averaging interval (i.e., $\tau = m\tau_0$). The results of system-level (i. e., the sensors are installed in a mechanical subassembly, which is in turn installed into the inertial instrument) AV analyses on three laboratory RLG models[3] are given

[3] *Final Report, RL-34 Ring Laser Gyro Laboratory Evaluation for the Deep Space Network Antenna Application*, JPL Contract No. 959072 (internal document), Jet Propulsion Laboratory, Pasadena, California, November 28, 1991.

in Table 1 and arranged from best to worse. As can be seen, the best A-gyro is approximately twice as good as the C-gyro and represents improvement over several years of RLG development. These particular gyros were used in the pointing tests in Section IV. Simulation of RLG's is obtained by utilization of the noise parameters of Table 1 in the RLG error model in Fig. 3. These results are given in Fig. 4. Actual RLG error data from laboratory tests[4] are shown in Fig. 5 and verify the stochastic modeling.

IV. High-Inheritance Instrumentation

The instrument evaluated is a navigator, and as such possesses more processing capability than is actually needed in this application. The instrument computes attitude coordinates with respect to a local geographic frame (i.e., elevation, azimuth, and roll), latitude, longitude, and altitude. Azimuth and elevation only are required for DSN pointing applications. The instrument consists of a triad of RLG's, a triad of accelerometers, a sensor mounting subassembly, and processing electronics. A functional block diagram of the instrument is shown in Fig. 6. In this section, laboratory instrument test results are given for an existing navigator unit in initialization and tracking modes.

A. Laboratory Initialization Tests

Initialization is the process of determining the initial azimuth (heading) in the level geographic plane. Figure 7 defines the geographical coordinate system, which is referenced to geographic latitude vector and north. In practice, the local level plane is found by zeroing the x and y instrument accelerometers. Basically, the x and y gyro outputs are averaged until the bias instability (b) limit is reached, as shown in Eq. 5. The azimuth is found from

$$\tan^{-1}\left(\frac{g_y}{g_x}\right) = \alpha \tag{8}$$

where g_x and g_y are the averaged RLG outputs after subtraction of the Earth spin-vector projection in the level plane. This process is referred to as gyrocompassing.

The initialization tests were conducted on a Contraves 51C rate table with an air-bearing table. On top of this table was an Ultradex table for elevation adjustment. The Contraves table is installed on an isolation pad to attenuate any building coupling. An external precision metrology system provided system calibration in azimuth and elevation. Twenty-two gyrocompassing tests were conducted at nine different azimuths. The time required to reach the bias instability floor ranged from 2 to 6.25 hr. The results of laboratory initialization tests are given in Fig. 8 Summarizing, the average azimuth error over 22 tests was 0.86 mdeg.

B. Laboratory Tracking Tests

Twelve pseudo-tracking performance tests were conducted. The tests consisted of locking the Contraves and Ultradex tables on the precision pier and identifying as pointing errors all gyro outputs except those from Earth-spin components. Each test was run for 10 hr. The results are given in Table 2 and shown as an RMS error. The RMS error is the time averaged over 10 hr. The average RMS tracking error over the 10-hr interval was 3.83 mdeg.

C. Simulated Tracking Tests

To test improvements to the instrument, a simulation was coded in MatLab to emulate the instrument. A typical simulated tracking error trajectory is given in Fig. 9, which shows an end-of-track error of approximately 3 mdeg. The first simulated improvement is to change-out the lower performance gyros (B- and C-gyros in Table 1). A triad of "A" model RLG's will be used. Additionally, the scale factor Q can be reduced by a factor of 8 to reflect new threshholding electronics that are now available. With these changes, the simulated instrument tracking performance is shown in Fig. 10, which, when compared with Fig. 7, shows a performance improvement of 40 percent. The sensor-limited instrument performance confirmed analytically from Eq. (1) is 1.42 mdeg.

Thus, higher performance RLG's will significantly improve instrument performance. A search located RLG's with the following noise parameters: $q = 1.9 \times 10^{-4}$ (deg/root-hr) and $b = 4.0 \times 10^{-5}$ (deg/hr). Simulation results show that the error at the end of 10-hr tracks is less than 1 mdeg (Fig. 11).

V. Conclusion

The feasibility of utilizing inertial-sensor high-inheritance instrumentation for DSN large-structure antenna pointing was established. What was not accomplished includes

(1) Upgrade of the simulation to include accelerometer error models and Kalman filtering. In most applications, the instrument Kalman filter is run only during gyrocompassing because the instrument velocity

[4] *Final Report, RL-34 Ring Laser Gyro Evaluation for the Deep Space Network Antenna Application*, op. cit.

is known to be zero. In the DSN application, because the tracking rates are slow and nominally known, covariance improvement may be obtained by running the filter during tracks.

(2) Analysis to establish the optimum mounting location. The instrument must be mounted at a location, nominally on the rear primary reflector, that is highly correlated with the antenna pointing direction.

(3) Instrument modeling and filtering improvements. The DSN disturbance input/output modeling has not yet been incorporated into the simulation.

(4) Instrument field tests to validate the pointing error budget.

References

[1] C. Broxmeyer, *Inertial Navigation Systems*, New York: McGraw-Hill Book Company, 1964.

[2] *Advances in Inertial Navigation Systems and Components*, AGARD-AG-254, Chapter 7, pp. 7-1–7-37, available from NTIS, Springfield, Virginia, April 1981.

[3] W. W. Chow, V. E. Sanders, W. Schleich, and M. O. Scully, "The Ring Laser Gyro," *Reviews of Modern Physics*, vol. 57, no. 1, pp. 61–104, January 1985.

[4] D. Allan, P. Kartaschoff, J. Vanier, J. Vig, G. M. R. Winkler, and N. G. Yannoni, "Standard Terminology for Fundamental Frequency and Time Metrology," *Proceedings of the IEEE 42nd Annual Frequency Control Symposium*, Baltimore, Maryland, pp. 419–425, 1988.

[5] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, New York: McGraw-Hill, 1965.

[6] D. Sargent and B. O. Wyman, "Extraction of Stability Statistics From Integrated Rate Data," paper 80-1728, *Proceedings of the Guidance and Control Conference*, American Institute of Aeronautics and Astronautics, Danvers, Massachusetts, pp. 88–94, 1980.

[7] A. D. King, *Characterisation of Gyro In-Run Drift*, Edinburgh, Scotland: Ferranti Defence Systems Ltd., Navigation Systems Department, 1971.

Table 1. Allen Variance analysis on three RLG models.

Gyro models	Quantization (Q), mdeg/pulse	Angle random walk (q), mdeg/root-hr	Bias instability (b), mdeg/hr
A	0.106	0.30	0.16
B	0.106	0.42	0.20
C	0.106	0.59	0.40

Table 2. Laboratory tracking test results.

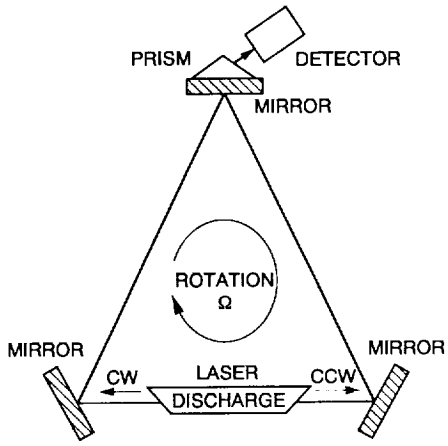
RMS azimuth error, mdeg	RMS elevation error, mdeg	RMS pointing, mdeg	Description
3.6111	0.8056	3.6667	El = 0 deg[a]
4.8889	1.2222	4.9444	El = 0 deg
5.5833	0.4722	5.6111	El = 0 deg[b]
3.4444	1.1111	3.5556	El = 0 deg
5.1944	1.0000	5.2222	El = 0 deg
3.7778	0.5833	3.7778	El = 0 deg
			Recalibration[c]
3.0000	0.7500	3.1111	El = 0 deg
1.2778	0.4722	1.3611	El = 0 deg[d]
3.1111	1.0556	3.2778	El = 60 deg
4.7778	0.8333	4.8333	El = 60 deg
2.8889	1.0278	3.0556	El = 60 deg
3.5833	1.0833	3.7500	El = 60 deg
—	—	3.8333	Average pointing over 12 tests
—	—	3.2283	Average after recalibration

[a] Typical.

[b] Worst case.

[c] One month from previous calibration.

[d] Best case.



SAGNAC EFFECT

COUNTER-PROPAGATING BEAMS TRAVEL IDENTICAL PATHS

ROTATION OF GYRO PRODUCES A PHASE DIFFERENCE OF THE TWO BEAMS

PHASE CHANGE = $\Delta\phi = (8\pi A/\lambda c)\Omega$

WHERE A = ENCLOSED AREA

λ = LASER WAVELENGTH

Ω = GYRO ROTATION RATE

c = SPEED OF LIGHT

PROPAGATION TIME ALONG CLOSED PATH OF LENGTH L IS $\tau = L/c$

RATE OF PHASE CHANGE = $2\pi\Delta f = \Delta\phi/\tau$

THUS, $2\pi\,\Delta f = \Delta\phi^*(c/L) = 2\pi\,(4A/\lambda L)\,\Omega$

WHERE Δf = OUTPUT FREQUENCY

TYPICAL SCALE FACTOR = 1.5 arcsec/count

(FOR CIRCULAR PATH OF RADIUS R, SCALE FACTOR = $\lambda/2R$)

Fig. 1. Ring laser gyro component diagram and operating principal.

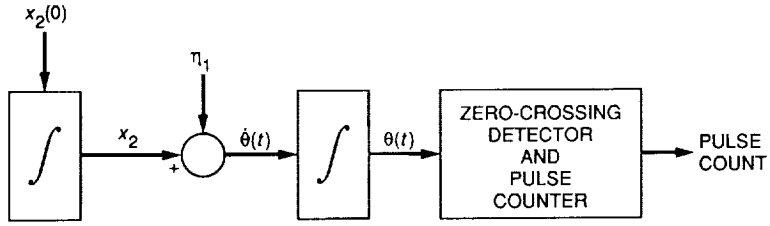


Fig. 2. Stochastic RLG model with ramp bias instability.

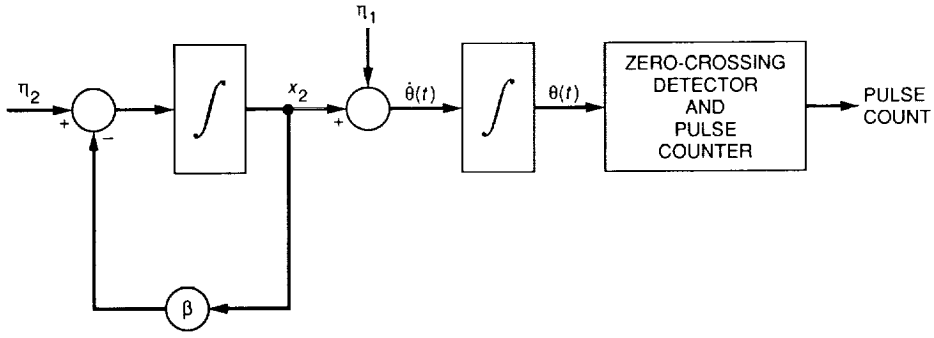


Fig. 3. Stochastic RLG model with exponential bias instability.

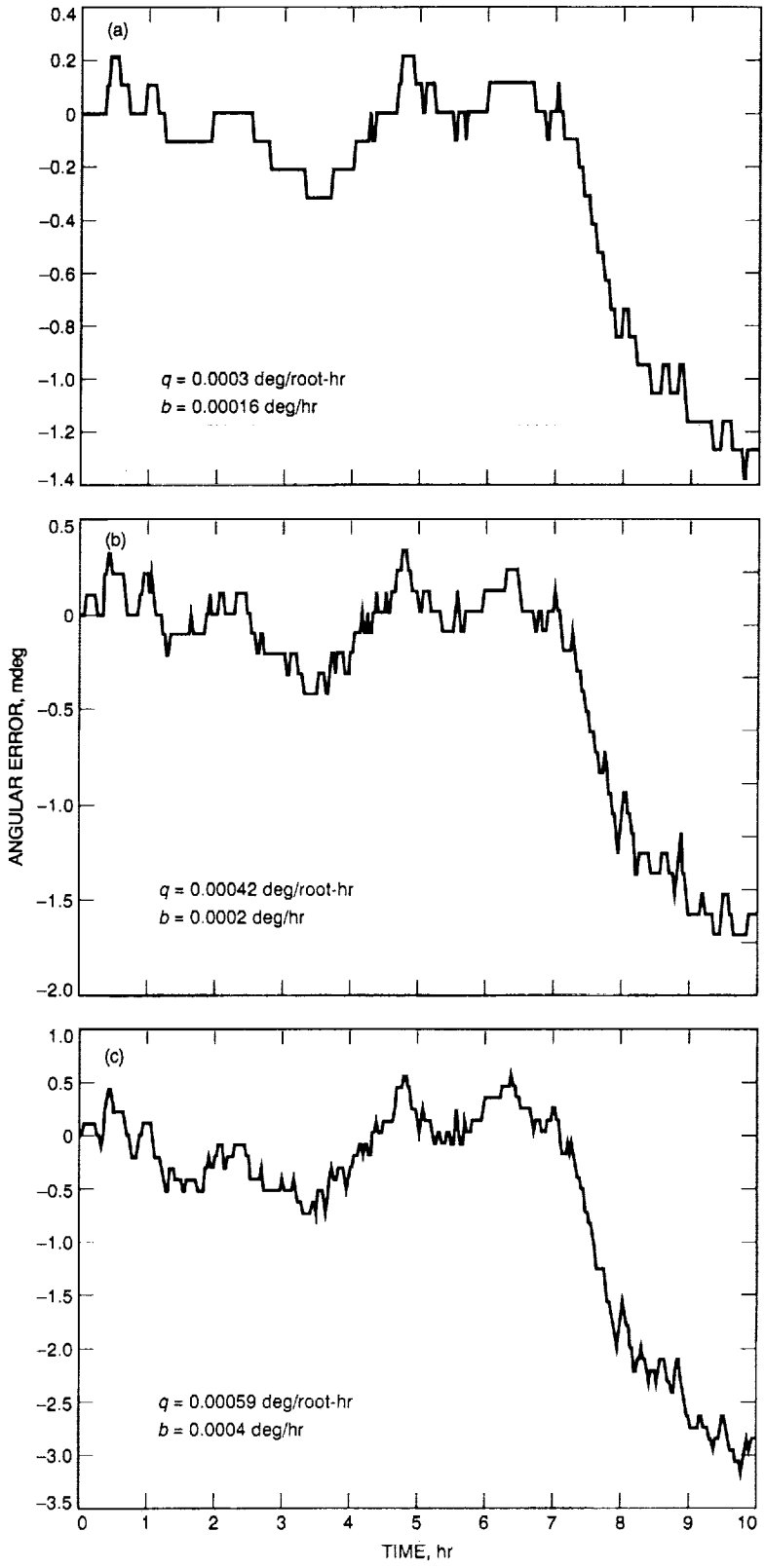


Fig. 4. Simulated gyro error with exponential bias instability: (a) A-gyro; (b) B-gyro; and (c) C-gyro.

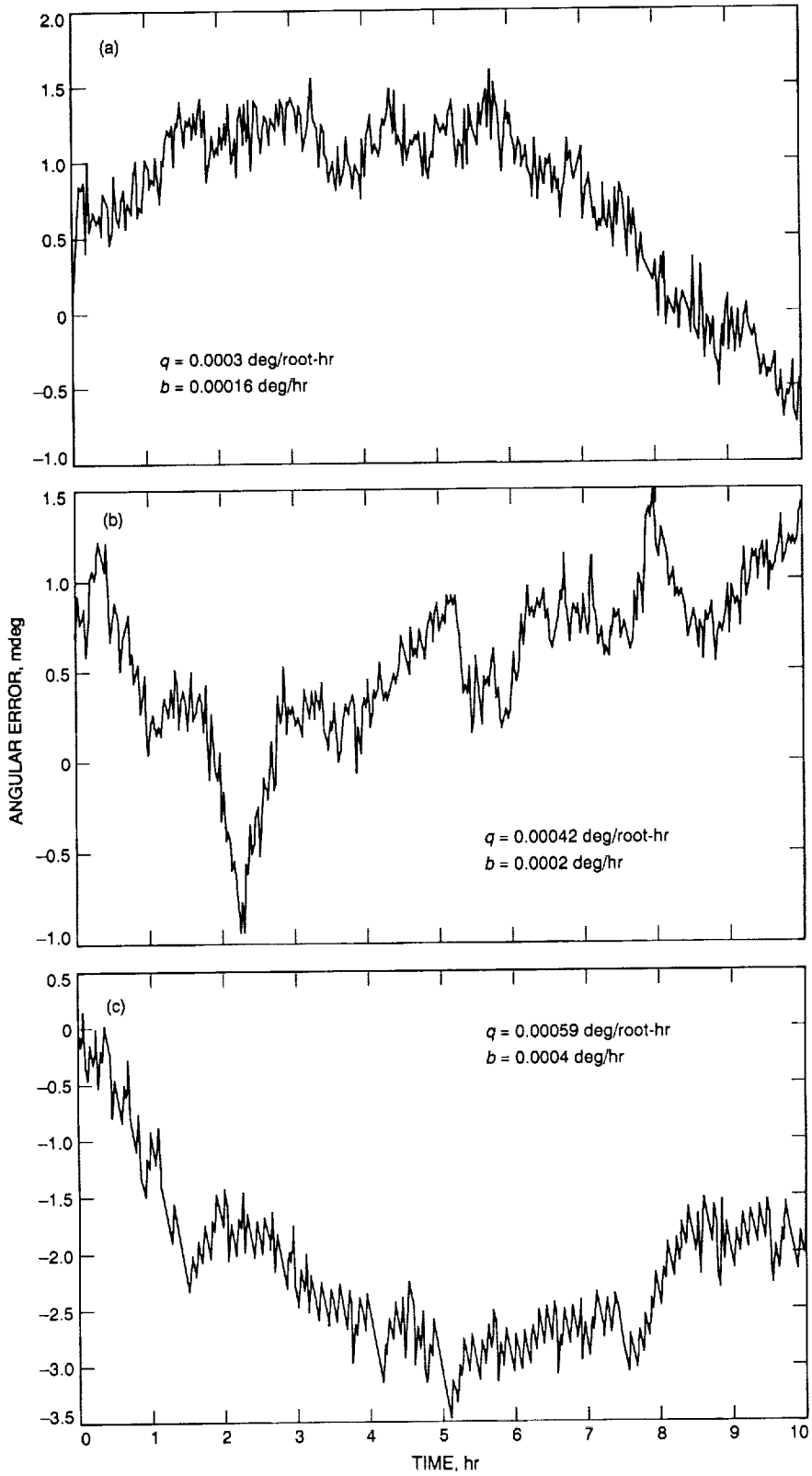


Fig. 5. Actual gyro laboratory test error: (a) A-gyro; (b) B-gyro; and (c) C-gyro.

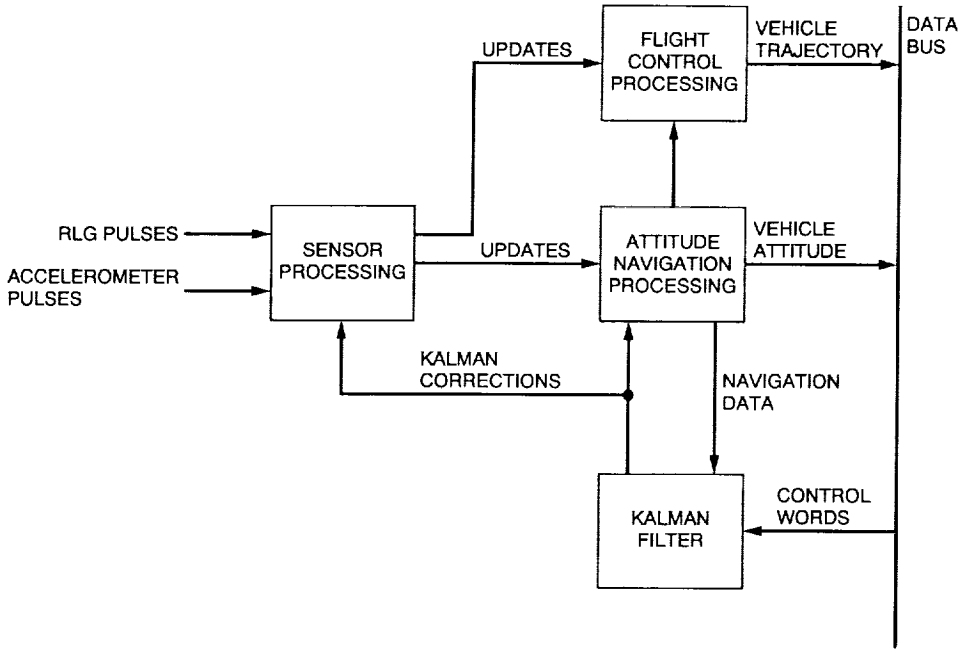


Fig. 6. Functional block diagram of inertial navigator.

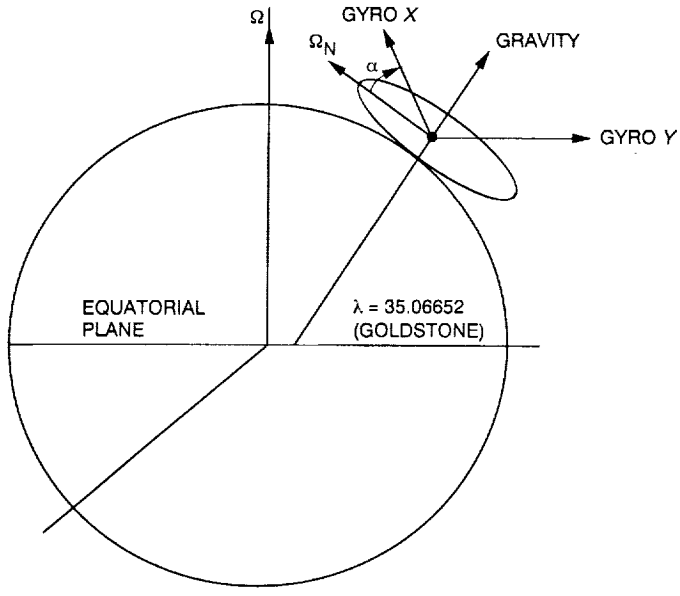


Fig. 7. Initialization frame.

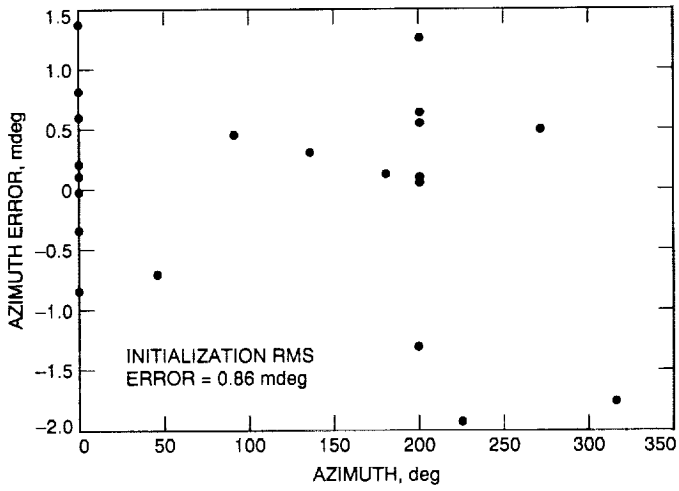


Fig. 8. Initialization-mode laboratory test results.

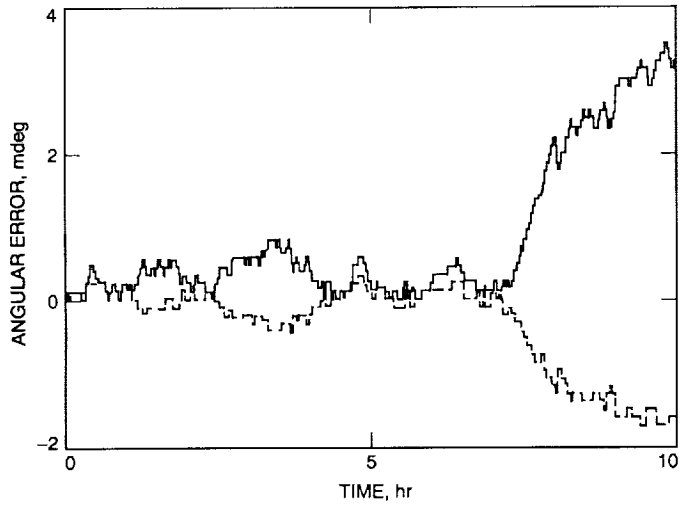


Fig. 9. Simulated system-level tracking error.

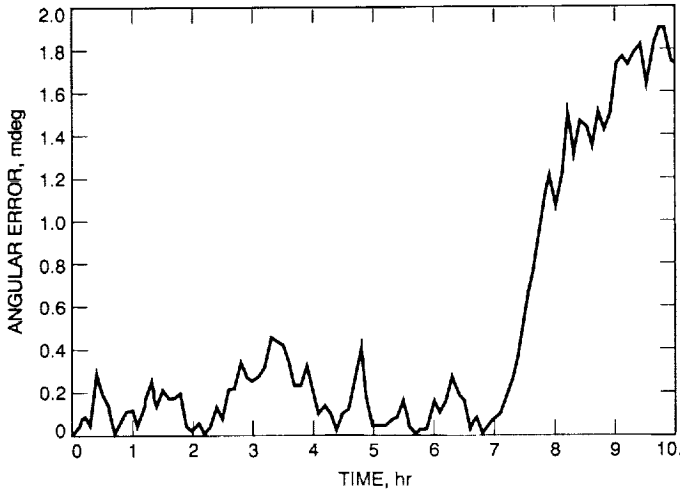


Fig. 10. Simulated tracking improvement with all A-gyros.

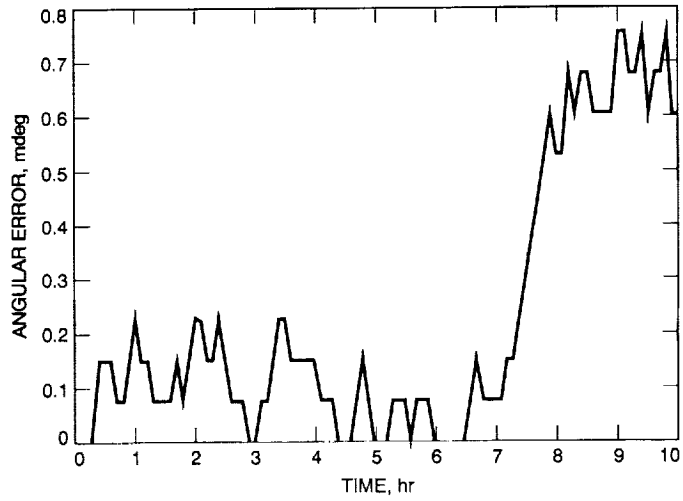


Fig. 11. Simulated tracking error with best performing gyros.

Appendix

The state equations of Fig. 2 are

$$\dot{\vec{x}} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \vec{x} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \eta_1$$

where $[x_1, x_2]^T$ *and* $x_1 = \theta$. Then

$$\dot{\vec{x}} = A\vec{x} + G\vec{\eta}$$

Let

$$\mathrm{E}\left[\vec{\eta}\vec{\eta}^T\right] = \sigma_{\eta_1}^2 \delta(\xi) \quad \text{(scalar)}$$

where η is a zero-mean white-noise process. Without tracking aids (fixes), the covariance of $\theta(t)$ increases in time. The linear covariance equation is

$$\dot{P} = AP + PA^T + GQG^T$$

and from previous notation,

$$Q = \begin{pmatrix} \sigma_{\eta_1}^2 & 0 \\ 0 & 0 \end{pmatrix}$$

Solving by assuming $\mathrm{E}[x_2(0)\theta(0)] = \mathrm{E}[\theta(0)\theta(0)] = \mathrm{E}[x_2(0)\eta_1(t)] = 0$,

$$\sigma_\theta^2(t) = E\left[x_2^2(0)\right] t^2 + \sigma_{\eta_1}^2 t \qquad \text{(A-1)}$$

In deriving Eq. (A-1), it is also understood that $x_2(0)$ and $\theta(0)$ are zero-mean.

The state equations of Fig. 3 are

$$\dot{\vec{x}} = \begin{bmatrix} 0 & 1 \\ 0 & -\beta \end{bmatrix} \vec{x} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \vec{\eta}$$

As before, let

$$\mathrm{E}\left[\vec{\eta}\vec{\eta}^T\right] = \begin{pmatrix} \sigma_{\eta_1}^2 & 0 \\ 0 & \sigma_{\eta_2}^2 \end{pmatrix} \delta(\xi)$$

be zero-mean and uncorrelated random processes and $x_2(0) = \theta(0) = 0$. Writing and solving the differential equations from the linear covariance equation yields

$$\sigma_\theta^2(t) = \sigma_{\eta_1}^2 t + \sigma_{\eta_2}^2 t + \left(\frac{\sigma_{\eta_2}^2}{2\beta^3}\right)\left(1 - e^{-2\beta t}\right)$$

$$+ \left(\frac{2\sigma_{\eta_2}^2}{\beta^3}\right)\left(e^{-\beta t} - 1\right) \qquad \text{(A-2)}$$

The correlation time β is computed, as a rule of thumb, to allow the gyro output uncertainty to reach the variance given by Eq. (A-1) at the end of a track.

Performance of a 12-GHz Fiber-Optic System for Beam-Waveguide Antenna Stability Testing

T. Y. Otoshi and M. M. Franco
Ground Antennas and Facilities Engineering Section

G. F. Lutes
Communications Systems Research Section

A 12-GHz fiber-optic system is a critical part of a test configuration that has been proposed for measuring the fractional frequency stability of the DSS-13 beam-waveguide (BWG) antenna. This fiber-optic system is used to carry Ku-band (12-GHz) signals from a reference antenna to the DSS-13 BWG pedestal room. Tests performed only on the fiber-optic system portion of the overall test configuration showed that the 12-GHz fiber-optic system (installed at DSS-13) has a frequency stability of about 1.1 × 10^{-16} for sampling time τ = 1000 sec for a nighttime run. This preliminary result establishes the lowest noise floor that can probably be achieved for the test configuration that will be used to measure the frequency stability of the DSS-13 BWG antenna.

I. Introduction

In a previous article [1], a new method for measuring the frequency stability of the DSS-13 BWG was presented. Figure 1 shows the proposed test configuration. The method requires the use of far-field signals in the 11.7- to 12.2-GHz region from geostationary satellites, a stable reference antenna, and a phase detector Allan variance measurement instrument [2]. By receiving the far-field signals simultaneously with a reference antenna and the 34-m antenna under test, the phase variations common to both paths cancel at the output of a mixer contained in the Allan variance measurement instrument [2]. An assumption is made that the time delay difference between the reference and test antennas is much less than the reciprocal of the smallest bandwidth of the signal after filtering. For example, if the bandwidth of the 12-GHz test signal is 1 MHz after filtering, the time delay difference between the test and the reference antennas must be much less than 1 microsecond. To enable accurate determination of the stability of the test path to parts in 10^{15}, it is desired that the path from the reference antenna to the Allan variance machine be very phase stable with fractional frequency stabilities on the order of one or two parts in 10^{16}. The employment of a 12-GHz fiber-optic system makes it possible to meet this stringent frequency stability requirement.

The purpose of this article is to present preliminary results of measurements that were made on the 12-GHz

fiber-optic portion of the overall measurement system that will ultimately be employed to test the BWG antenna.

II. Reference Path Test Configuration

Figure 2 shows the 10-ft reference antenna that was developed for this project. This 10-ft reference antenna is located in an outdoor environment and mounted on a solid concrete pad about 30 m from the center of the 34-m antenna (Fig. 3).

A 12-GHz fiber-optic system, installed at DSS 13, is used to carry the output of the reference antenna to a point near the F3 focal point in the pedestal room (see Fig. 1). The 12-GHz fiber-optic system consists of a transmitter, 150 ft of fiber-optic cable, a receiver, and associated amplifiers and short lengths of phase stable cables. The purpose of the transmitter is to receive the 12-GHz signal, use it to modulate a laser beam and then transmit it toward the receiver via the fiber-optic cable. About a 150-ft length of fiber-optic cable goes underground from the reference antenna into the pedestal room to the F3 focal point location where the Ku-band test package [3] is installed. The purpose of the receiver is to demodulate the 12-GHz signal on the laser beam. Short lengths of phase stable microwave cables are used to connect the receiver output to the Allan variance machine.

The new far-field method takes advantage of recent advances made in fiber-optic technology. Tests by Lutes and Logan [4] on a fiber-optic system at X-band (8.4 GHz) have demonstrated that a differential frequency stability of 7×10^{-16} has been obtained for sampling times of 10 sec. This stability value is about two orders of magnitude better than the frequency stability of typical hydrogen masers. The short-term stability is limited by the signal-to-noise ratio of the output signal.

For the proposed BWG antenna stability measurement method to yield useful and accurate data, it is desirable for the described 12-GHz reference path to have a fractional frequency stability of better than 1×10^{-16} or 2×10^{-16} for $\tau = 1000$ sec. It was demonstrated by Conroy [2] that the phase detector Allan variance instrument has a noise floor level of about 1×10^{-17} for $\tau = 1000$ sec. Therefore, the frequency stability of the reference path basically sets the noise floor of the measurement system. It is desirable that the reference path be 10 times more stable than the BWG antenna path stability to be measured.

As a preliminary step for evaluating the overall measurement system, the 12-GHz fiber-optic system (minus the reference antenna) was tested. The test block diagram is shown in Fig. 4. A 12-GHz signal from a frequency synthesizer is sent to a 10-dB coupler. One of the outputs from the coupler goes directly to the reference port of the Allan variance phase detector. The other output goes to the 12-GHz fiber-optic transmitter unit. An AM-optical signal emitted by the transmitter unit is sent to the reference antenna's outside environment via approximately 150 feet of fiber-optic cable. At the end of this cable, a fiber-optic connector is used to join this cable to another 150-ft length of fiber-optic cable going back to the receiver.[1] The modulated optical signal is demodulated by the receiver and the 12-GHz output signal is amplified by an external amplifier and routed to the test port of the Allan variance instrument. Thus, the configuration being tested is the 12-GHz fiber-optic system with a 300-ft length fiber-optic cable, part of which is exposed to outside temperature variations.

III. Test Results

The first test was made with the signal generator (see Fig. 4) being driven by its own internal free-running frequency reference. A second test was made with the signal generator being driven by a portable Cesium reference standard.

Figure 5 shows the Allan deviation results for the first test, which was run during both daytime and nighttime hours. Two 9-hour runs were made. Allan deviations for the day and night runs were, respectively, 4.9×10^{-16} and 1.1×10^{-16} for $\tau = 1000$ sec. It can be seen in Fig. 5 that a strange hump (a departure from linearity) occurred for the daytime run. This departure from linearity on the log–log plot can be explained by examining the corresponding phase versus time plot shown in Fig. 6. It can be seen in the plot that the phase change since the start of the test varied from 0 to 20 deg and then to about −5 deg over the duration of the test. Whenever there is a large departure from linear behavior in the phase curve slope, the Allan deviation will correspondingly be affected. The cause of the phase change in Fig. 6 can be explained by examining the weather data curve in Fig. 7. There is a strong correlation between the measured phase change (Fig. 6) and the outside air temperature change (Fig. 7). During the daytime hours, the outside temperature varied between 23.9 and 30 deg C. Examination of Fig. 6 shows that the phase varied from 0 to about −18 deg for the nighttime run. The outside air temperature varied between 22.4 and

[1] The transmitter and receiver modules, which incorporate features based on JPL research, were purchased from the Ortel Corporation of Alhambra, California.

24.6 deg C. The reason why the Allan deviation curve was more linear for the nighttime run than the daytime run is that the nighttime run phase-versus-time curve is nearly linear.

During the measurements, phase-versus-time data are stored in computer data files. The Allan deviation is computed from the phase data by using equations given in [2].

Even though most of the fiber-optic cable goes underground to a stable thermal environment, about 10 feet of the cable were exposed to the outside weather environment. The test results shown in Figs. 5 and 6 show that the fiber-optic cable is somewhat sensitive to outside air temperature changes and, for best results, the entire cable length needs to be in a thermally stable environment. After the tests described in this article were completed, the portion of fiber-optic cable exposed to the weather at the reference antenna location was wrapped with thermal insulation material.

Figure 8 shows the test data for the period when the signal source was driven by 5 MHz from a portable Cesium frequency standard. The Allan deviations for the daytime and nighttime runs were, respectively, about 4.5×10^{-16} and 3.5×10^{-16} for sampling times of 1000 sec. Figure 9 shows the corresponding phase-versus-time plot for the daytime run. Unfortunately, the phase-versus-time data for the nighttime run could not be saved due to a glitch that prevented the data file from being closed. Figure 10 shows the corresponding temperature data for the same time period. The air temperature changed from 26 to about 30 deg C for the daytime test. Even though the temperature variation was about as large as for the first set of tests (Figs. 5–7), the phase change was much smaller. The smaller variation of phase is attributed to the use of a more stable frequency source to perform the tests. These test results indicate that the Allan variance result depends upon the stability of the frequency source when there is a differential path length difference between the reference and the test paths. For this test, the differential path length was 300 feet of cable, whose velocity of propagation is about 66 percent of that for free space. Thus, a 300-ft length of this fiber-optic cable corresponds to a phase delay of approximately 460 nsec.

Tests and analyses will be performed to determine the effects of such parameters as differential delays, signal modulations (AM and FM) and bandwidth.

IV. Conclusions

Preliminary test results have been obtained on the 12-GHz fiber-optic portion of the test configuration that will be employed to test the BWG antenna frequency stability. These results indicate that the best performance that can be achieved with this test configuration is about 1.1×10^{-16}. It is expected that under wind conditions, the stability of the BWG antenna will be at least an order of magnitude worse.

The test results presented in this article indicate that microwave fiber-optic systems can be useful for many antenna applications, such as carrying microwave signals from the antenna reflector surface areas down to ground level.

A small part of the degradation of performance of the 12-GHz fiber-optic system, whose results were presented in this article, might be attributable to the external amplifiers and rf cables used in the test. Also, the differential delay between the reference and the test port might have introduced some degradations even though a Cesium standard was used. These effects are being investigated.

Acknowledgments

J. Garnica, L. Smith, J. Ney, and the DSS-13 crew fabricated and assembled the reference antenna shown in the photograph. R. Logan assisted in the checkout of the 12-GHz fiber-optic system.

References

[1] T. Y. Otoshi, "A Proposed Far-Field Method for Frequency-Stability Measurements on the DSS 13 Beam-Waveguide Antenna," *TDA PR 42-107*, vol. July-September, Jet Propulsion Laboratory, Pasadena, California, pp. 81–87, November 15, 1991.

[2] B. L. Conroy and D. Le, "Measurement of Allan Variance and Phase Noise at Fractions of a Millihertz," *Rev. Sci. Instruments*, vol. 61, no. 6, pp. 1720–1723, June 1990.

[3] T. Y. Otoshi, S. R. Stewart, and M. M. Franco, "A Portable Ku-Band Front-End Test Package for Beam-Waveguide Antenna Performance Evaluation," *TDA Progress Report 42-107*, vol. July-September 1991, Jet Propulsion Laboratory, Pasadena, California, pp. 73–80, November 15, 1991.

[4] G. F. Lutes and R. T. Logan, "Status of Frequency and Timing Reference Signal Transmission by Fiber Optics," *Proceedings of the 45th Annual IEEE Symposium on Frequency Control*, Los Angeles, California, pp. 679–686, May 29–31, 1991.

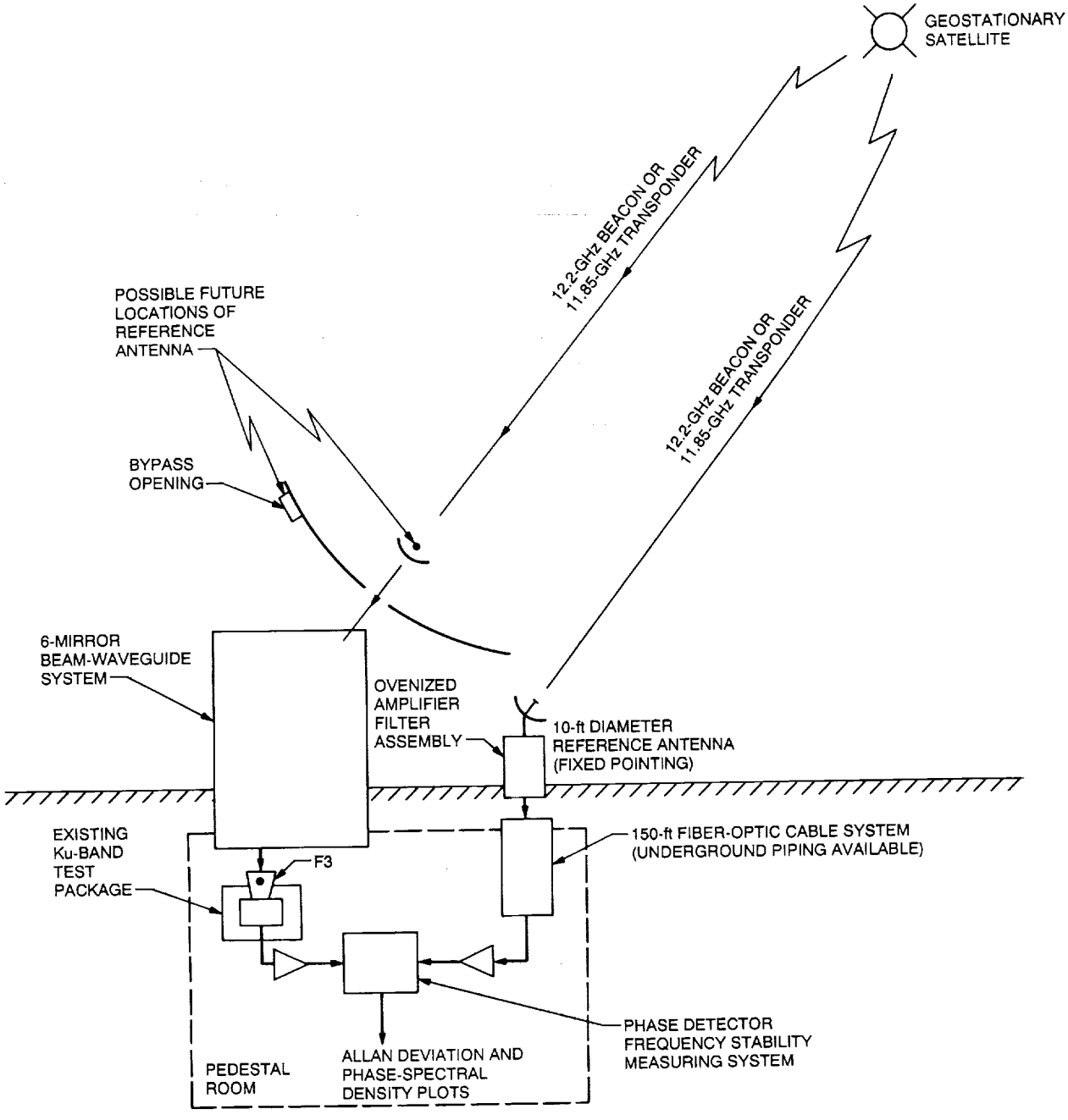


Fig. 1. Test configuration for measurement of the frequency stability of the BWG antenna.

Fig. 2. Newly installed 10-ft reference antenna.

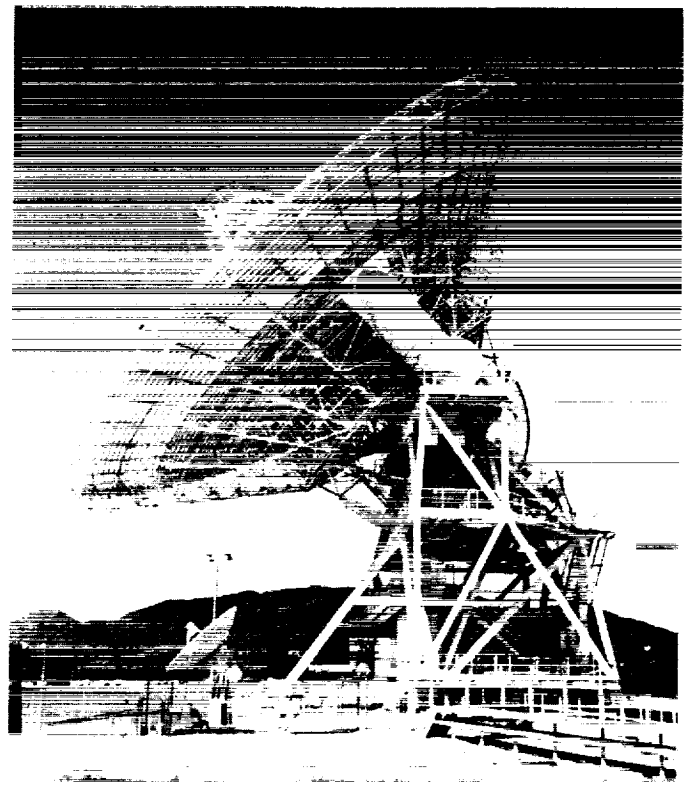


Fig. 3. View of the 34-m antenna and the 10-ft reference antenna pointed at a geostationary satellite at a 46.5-deg elevation angle.

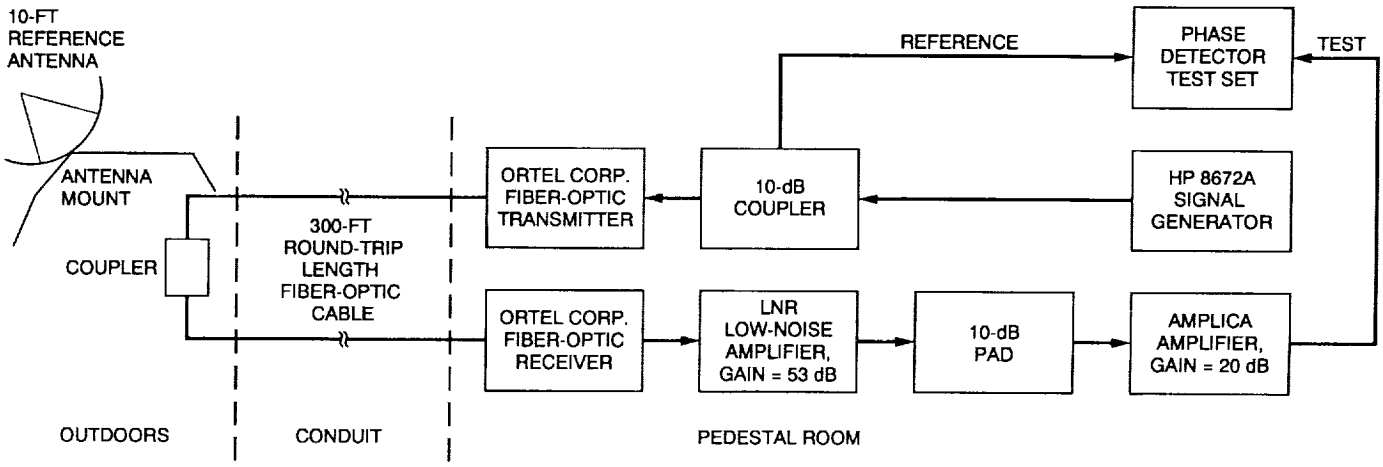


Fig. 4. Block diagram of the measurement setup to test the 12-GHz fiber-optic system that was installed at DSS 13.

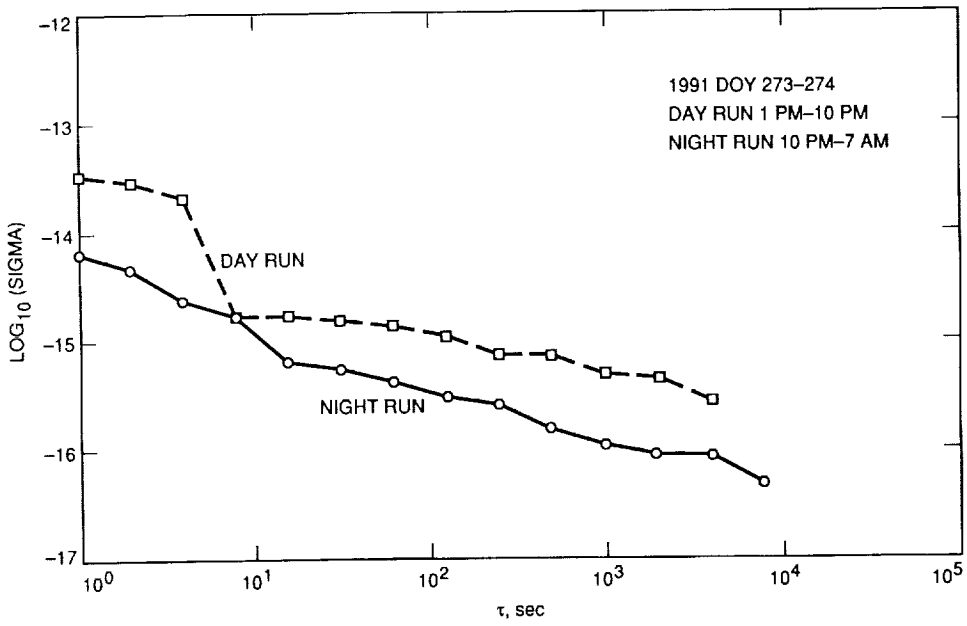


Fig. 5. Allan deviation characteristics of the 12-GHz fiber-optic system for day and night runs on 1991 DOY 273–274. The signal source was free-running. Test frequency was 12.198 GHz.

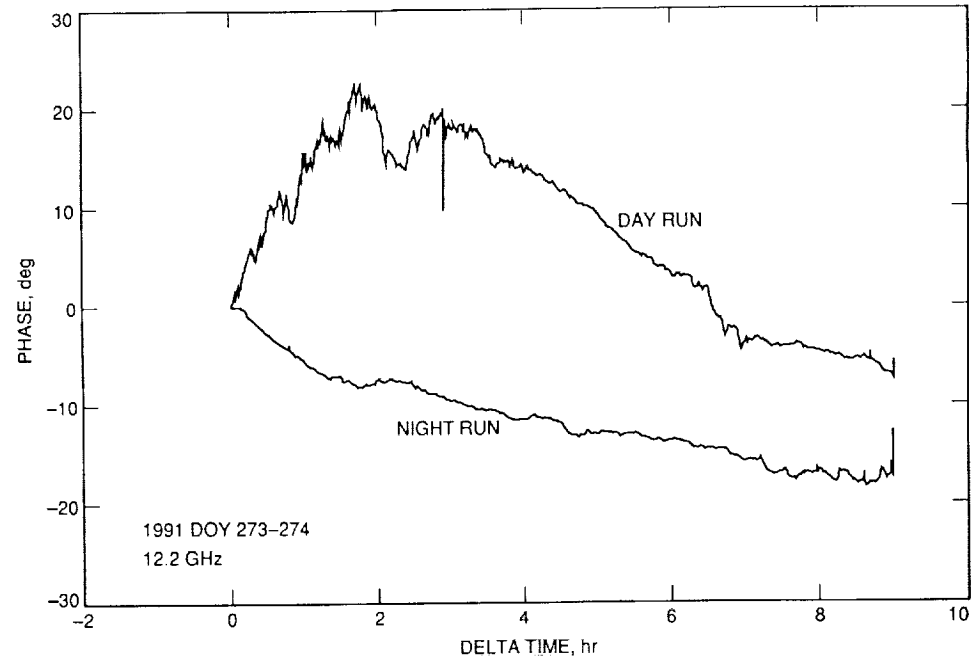


Fig. 6. Phase changes corresponding to the Allan deviation curves shown in Fig. 5.

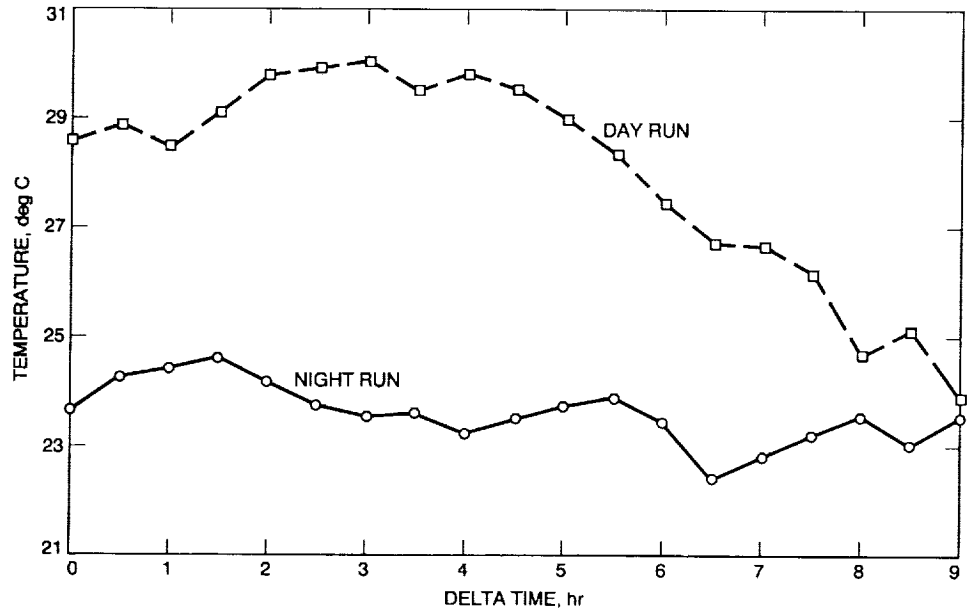


Fig. 7. Outside air temperature during measurements of phase changes shown in Fig. 6.

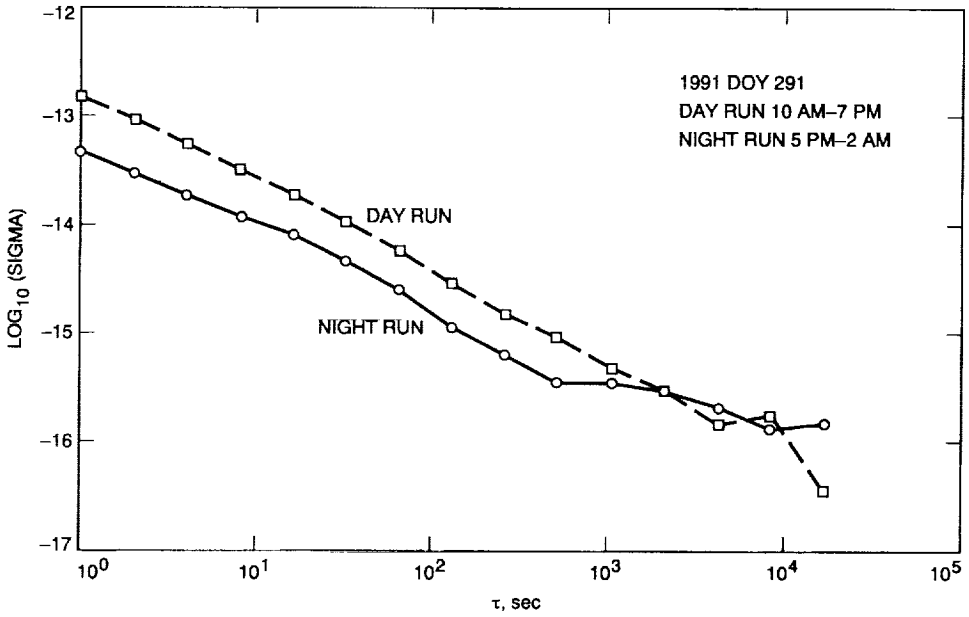


Fig. 8. Allan deviation characteristics of the 12-GHz fiber-optic system for day and night runs on 1991 DOY 291. The signal source was driven by 5 MHz from a portable Cesium reference standard. Test frequency was 12.248 GHz.

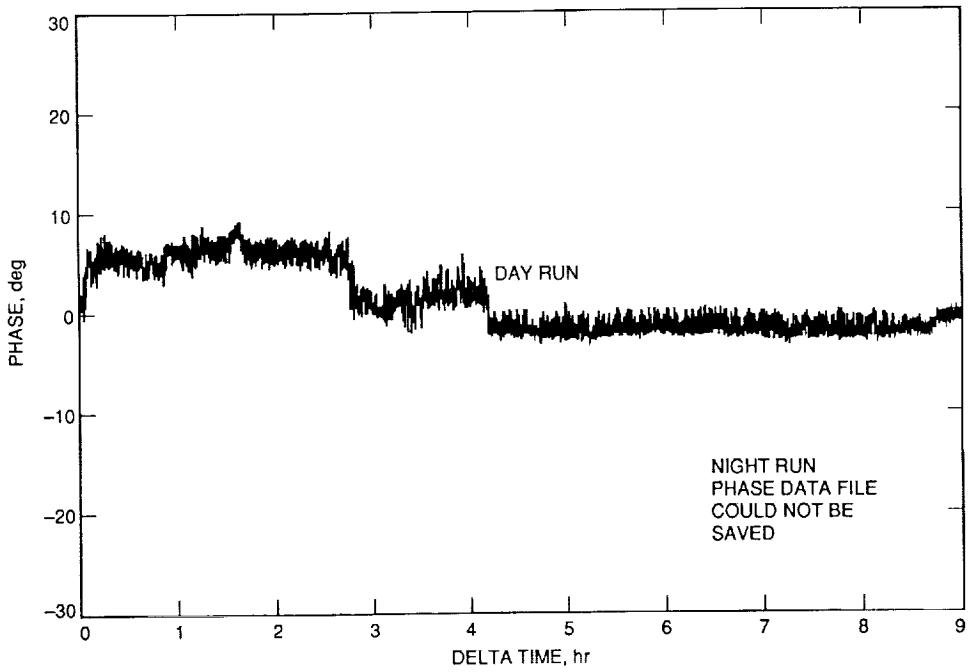


Fig. 9. Phase changes corresponding to the Allan deviation daytime curve shown in Fig. 8.

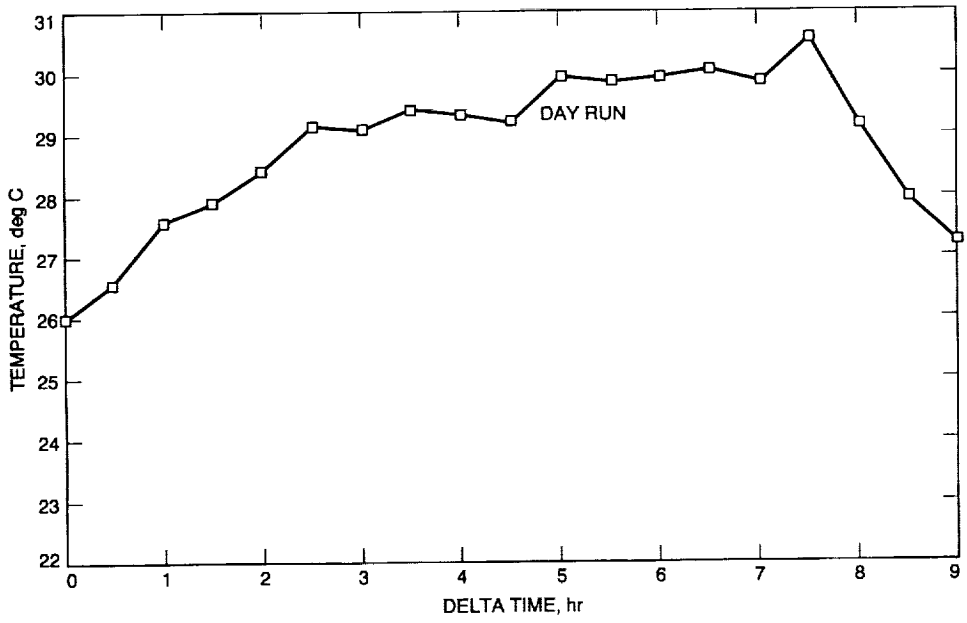


Fig. 10. Outside air temperature during measurements of phase changes shown in Fig. 9.

Improvements in X-Band Transmitter Phase Stability Through Klystron Body Temperature Regulation

R. M. Perez
Radio Frequency and Microwave Subsystems Section

This article describes the techniques used and experimental results obtained in improving transmitter stability by control of the klystron body temperature. Related work in the measurement of klystron phase control parameters (pushing factors) is also discussed. The contribution of waveguide temperature excursions to uplink phase stability is presented. Suggestions are made as to the direction of future work in this area.

I. Introduction

Long-term spacecraft tracking-link phase stability is an important parameter in the detection of gravity waves by Doppler methods. This link phase stability requirement dictated reducing the phase stability contribution of the DSN 20-kW X-band (7.167-GHz) uplink microwave transmitters to an Allan deviation of 5×10^{-15} for a 1000-second averaging period. This requirement was met in the DSN 20-kW X-band transmitters[1] by closely regulating the klystron cathode voltage. It was desired to decrease this figure by an order of magnitude to 1×10^{-16} to allow future operation with improved phase stability.

To achieve this goal a two-part approach was followed, which consisted of identifying and measuring the magnitude of the klystron phase-pushing factors and controlling those found to be significant contributors. Most prior transmitter phase stability improvement work has been concentrated on improving the klystron cathode (beam) voltage regulation, since the large effect of this parameter on the phase of the transmitter output signal is well established [1]. Efforts in this area have succeeded in regulating the beam voltage to about 0.01 percent over an 8-hour period and to nearly 0.003 percent over 1000 seconds, resulting in a beam voltage-induced phase shift on the order of 0.1 degree over eight hours of operation. In view of this, it was decided to concentrate on controlling other causes of phase instability.

II. Measurement of Phase-Pushing Factors

To establish the magnitude of the phase-pushing factors of the Varian 876P klystron used in the DSN 20-kW X-band transmitters, a measurement system (see Fig. 1) was configured. This system allows the independent variation of klystron operating parameters and the measure-

[1] R. Perez, "DSS 15 20 kW X-band Transmitter Test Results Regarding Stability Requirements," Interoffice Memorandum 3337-89-098 (internal document), Jet Propulsion Laboratory, Pasadena, California, October 3, 1989.

ment of subsequent changes in output signal phase. An HP 8510A network analyzer was used as the phase detector in conjunction with a coaxial switch, which enabled sampling of the phase of the klystron input and output signals and of a reference coaxial line. In this manner a differential phase measurement was made by using the phase of the reference line signal and the klystron input signal to normalize the phase of the klystron output signal. This resulted in the removal of any output signal phase fluctuations due to input signal or instrumentation instabilities from the phase-pushing factor data. Measurement of klystron operating parameters and temperatures was performed coherently with each phase measurement to permit later analysis. It should be noted that this configuration enabled control of the klystron body coolant temperature independent of the remaining cooling circuits and permitted the direct measurement of the klystron body temperature phase-pushing factor for the first time.

The magnitude of output signal phase fluctuation due to stimulus change was then measured for the following klystron parameters: beam voltage, drive RF level, body temperature, filament current, and focusing electromagnetic current. The results of these measurements may be found in Table 1.

III. Measurement of Phase Stability

After performing the pushing-factor measurements above, it was found that the klystron body temperature fluctuations showed the largest phase pushing of the parameters described above, given the typical operational variations encountered for each parameter. To stabilize the temperature of the klystron body, a refrigerated chiller (Neslab HX-75) capable of regulating the coolant temperature to 0.1 deg C was purchased. A system was then assembled (see Fig. 2) to perform long-term transmitter phase stability measurements (see [2] for a description of the stability measurement system-operating principles).

Phase stability measurements were then performed under two conditions: cooling the klystron body with the system heat exchanger and independent body cooling with the Neslab HX-75 refrigerated chiller. The results of these measurements may be found in Fig. 3 in the form of Allan deviations.

IV. Discussion of Results

The measurement of the klystron pushing factors revealed some significant points:

(1) The klystron output signal phase is most strongly affected by fluctuations in the beam voltage, with a value of 45 deg/kV; however, given the tight regulation of the beam power supply in the DSN 20-kW X-band transmitters, the beam voltage phase-pushing contribution becomes secondary to other factors.

(2) No phase pushing was detected due to changes in the klystron-focusing electromagnetic current or filament current. This should not be taken to mean that phase pushing by these parameters is not possible; rather, no phase change was observed due to a ± 10 percent change in focusing current and filament current, and a phase measurement resolution on the order of 0.1 degree. Wider excursions in these parameters carry a risk of exceeding the operating ratings of the klystron with ensuing damage. Normal drift of these parameters in operation is on the order of 1 to 2 percent; thus, it is unlikely that filament or focusing-current instabilities significantly affect overall transmitter stability.

(3) The temperature of the klystron body is a strong phase-pushing factor (see Fig. 4), with a value of 0.9 deg/deg C. It can be seen that given typical heat exchanger temperature fluctuations in the 5-deg-C range, this becomes the dominant klystron output phase-pushing factor in the presence of a tightly regulated beam voltage.

(4) A relationship was noticed between the output signal phase and the temperature of the transmitter output waveguide coolant. An analysis of the phase shift due to thermal expansion of the waveguide was performed (see Appendix), which confirmed this hypothesis. The expected phase pushing by waveguide temperature fluctuations agrees closely with the measured values (see Fig. 5).

(5) Cooling the klystron body independently with the Neslab HX-75 chiller resulted in a marked improvement in transmitter phase stability, particularly in the 10^2- to 10^4-second region critical to gravity wave detection [3]. At a 1000-second averaging time, the Allan deviation changed from $\sigma = 3.25 \times 10^{-16}$ without body temperature control to $\sigma = 5.88 \times 10^{-17}$ with body temperature control (see Fig. 3).

(6) A simulation was performed to investigate the extent to which the klystron pushing factors thus far identified account for the overall transmitter phase stability performance. The point-to-point variation in each parameter (beam voltage, body temperature, and waveguide temperature) from the test time series record was multiplied by its respective phase-pushing factor to obtain the phase step produced by

changes in each parameter. These phase changes were added algebraically to produce a composite phase file, the Allan variance of which was computed. It is postulated that the difference between the Allan variance of the synthetic phase file and the Allan variance measured for the whole transmitter is due to phase-pushing factors other than the variations in klystron beam voltage, body temperature, and waveguide temperature. A comparison of the results may be seen in Fig. 6. Note that the performance of the simulated transmitter (using only beam voltage and body and waveguide temperatures) is significantly superior to that of the actual transmitter, particularly for averaging times greater than 10 seconds. This strongly suggests that future gains in long-term transmitter phase stability are to be made by concentrating on components other than the klystron (the klystron filament and focus currents have no measured impact on output phase, as mentioned above). For example, examination of the phase stability of the RF control components and driver amplifier regarding temperature and possibly control voltages is likely to be fruitful. The amplitude stability of these components should also be examined, due to amplitude modulation–phase modulation (AM–PM) conversion in a saturated klystron.

V. Conclusion

The phase-pushing factors of the Varian 876P klystron used in the DSN 20-kW X-band transmitters have been measured. Tight temperature regulation of the klystron body temperature has resulted in appreciably improved transmitter phase stability, surpassing the goal of $\sigma = 1 \times 10^{-16}$ to 5.88×10^{-17} for the 1000-second averaging time set for the task. The effect of waveguide temperature changes on the phase of the transmitter output signal has been identified and verified analytically. Some promising areas for future work in transmitter stability improvement have been suggested.

References

[1] R. Kolby, "Evaluation of the VA876P Klystron for the 20-kW X-Band Uplink Transmitter," *DSN Progress Report 42-54*, vol. September–October, Jet Propulsion Laboratory, Pasadena, California, pp. 41–50, September 1979.

[2] B. Conroy and D. Le, "Simultaneous Measurement of Allan Variance and Phase Noise Spectrum at Fractions of a Millihertz," *Review of Scientific Instruments*, vol. 61, no. 6, pp. 1720–1723, June 1990.

[3] S. Nelson and J. Armstrong, "Gravitational Wave Searches Using the DSN," *TDA Progress Report 42-94*, vol. April–June 1988, Jet Propulsion Laboratory, Pasadena, California, pp. 75–85, August 15, 1988.

Table 1. Phase-pushing factors for the Varian 876P klystron.

Parameter	Phase-pushing factor
Beam voltage, deg/V	0.045
Drive level, deg/dB	2.3
Body temperature, deg/deg C	0.9
Filament current	None measured
Focusing current	None measured

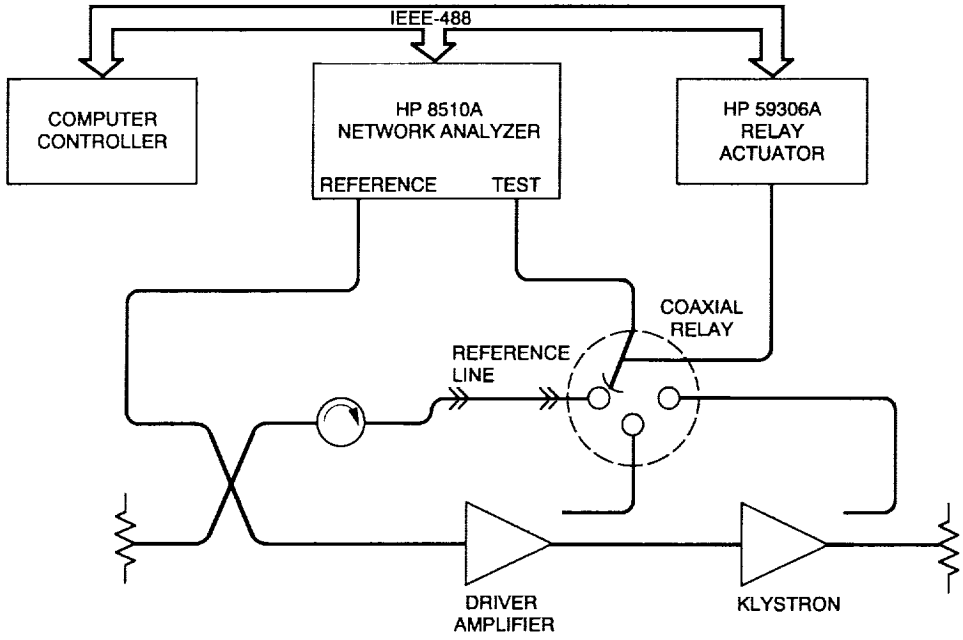


Fig. 1. Phase-pushing factor measurement system.

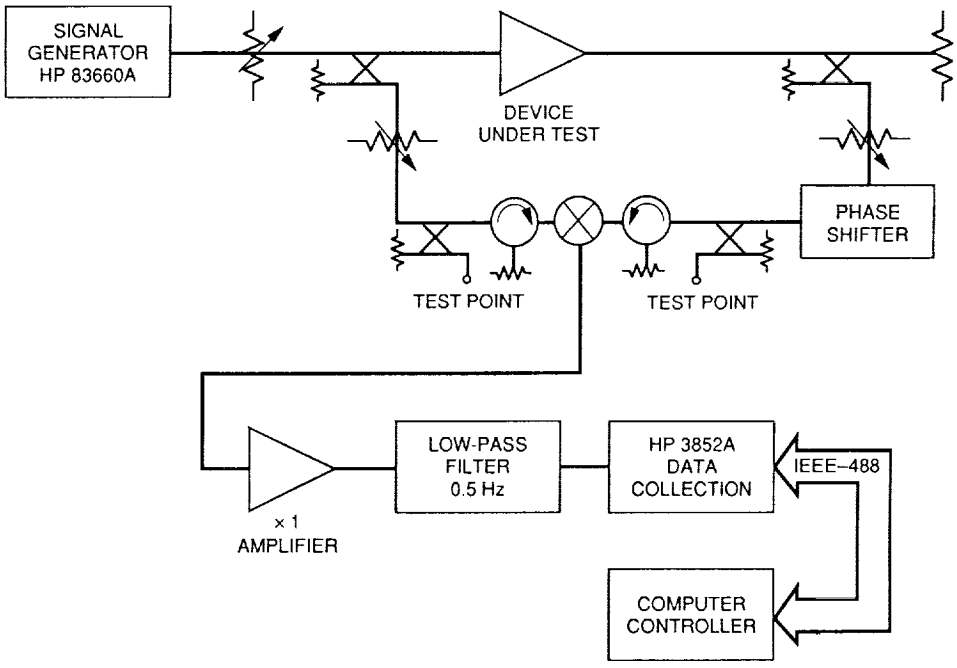


Fig. 2. Phase stability measurement system.

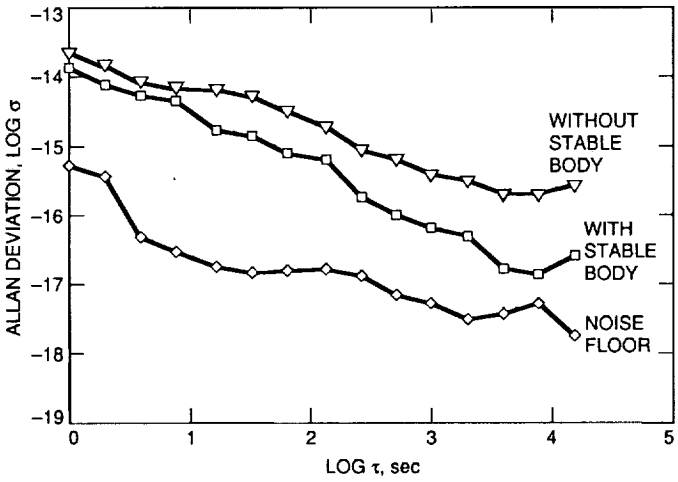


Fig. 3. Transmitter phase stability improvement.

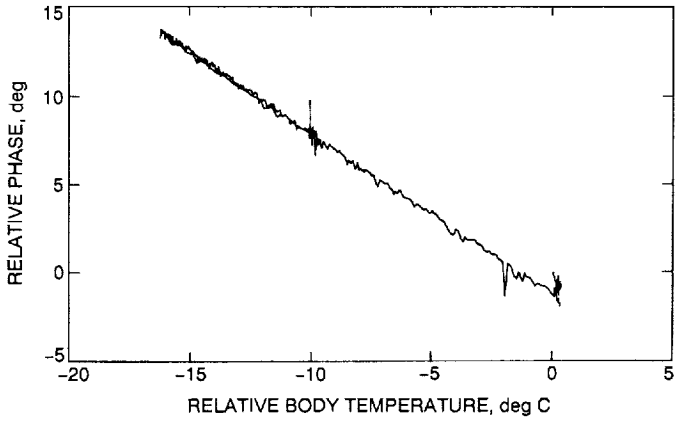


Fig. 4. Klystron body temperature pushing of phase.

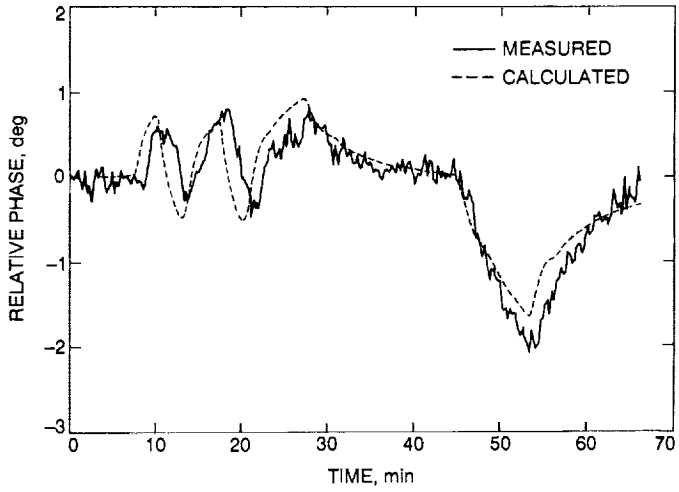


Fig. 5. Phase shift due to waveguide temperature fluctuations.

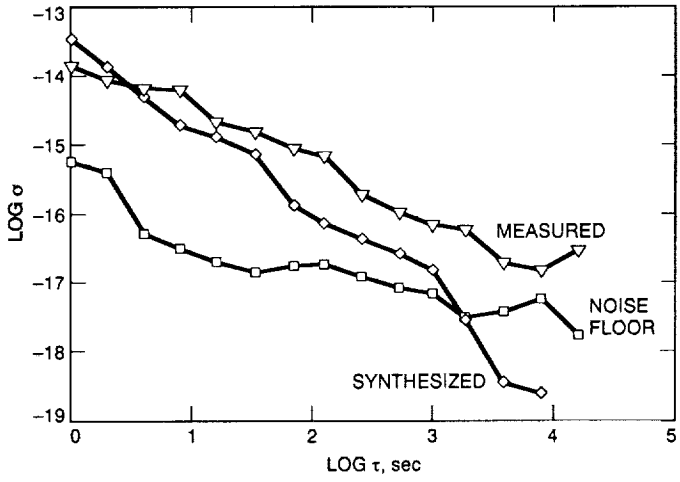


Fig. 6. Comparison of measured and synthesized Allan variances.

Appendix

Phase Shift of a Signal Due to Waveguide Temperature Changes

The guide wavelength of a signal in a rectangular waveguide and $TE_{1,0}$ mode is

$$\lambda_i = \frac{\lambda_0}{\sqrt{1 - \left(\frac{\lambda_0}{2A}\right)^2}} \tag{A-1}$$

where

λ_0 = free space wavelength

A = waveguide width

The phase shift for a waveguide of length l is

$$\phi_i = 360\frac{l}{\lambda_i} \text{ deg} \tag{A-2}$$

Given a thermal expansion coefficient for the waveguide material and a temperature change of ΔT, the resulting waveguide width change is

$$\Delta A = A\rho\Delta T \tag{A-3}$$

where

ρ = thermal expansion coefficient in ppm/C

ΔT = temperature change in deg C

which results in a new guide wavelength:

$$\lambda_f = \frac{\lambda_0}{\sqrt{1 - \left(\frac{\lambda_0}{2A + 2\Delta A}\right)^2}} \tag{A-4}$$

Likewise, the length of the waveguide section changes by

$$\Delta l = l\rho\Delta T \tag{A-5}$$

The expanded waveguide has a phase shift of

$$\phi_f = 360\frac{l + \Delta l}{\lambda_f} \text{ deg} \tag{A-6}$$

The phase shift induced by the temperature change is then

$$\Delta\phi = \phi_f - \phi_i \tag{A-7}$$

By substituting Eqs. (A-2) and (A-4) and simplifying,

$$\Delta\phi = 360\left(\frac{l + \Delta l}{\lambda_f} - \frac{l}{\lambda_i}\right) \text{ deg} \tag{A-8}$$

Noise-Optimal Control of HEMT LNA's for Compensation of Temperature Deviations

C. MacCarley[1], J. J. Bautista, and P. A. Willis
Radio Frequency and Microwave Subsystems Section

Noise-optimal control of high-electron mobility transistor low noise amplifier (HEMT LNA) bias voltage and current values was achieved at room temperature. The performance metric maximized was the amplifier gain divided by the amplifier input noise temperature, G/T_e. Additionally, the feasibility of automating the initial determination of bias settings was demonstrated in the laboratory. Simulation models of an HEMT were developed from available measurement data, installed on a Sun SPARC I workstation, and used in investigating several optimization algorithms. Simple tracking-type algorithms, which follow changes in optimum settings if started at or near the global optimum point, produced the best performance.

Implementation of the optimization algorithms was performed using a three-stage Field Effect Transistor (FET) LNA and an existing test apparatus. Software was written to control the bias settings of the first stage of the LNA and to perform noise and gain measurements by using the test apparatus. The optimization control was then integrated with existing test software to create a master test and optimization program for test apparatus use.[2]

I. Introduction

The prime objective of this work was to develop a method to maintain optimal bias voltage and current values of an HEMT LNA as the physical temperature varied. Since the input noise temperature of a multistage LNA is a function of both individual stage noise temperatures and gains, the optimal (i, v) values were defined to be those that maximize amplifier gain divided by the input noise temperature, G/T_e. These bias settings are temperature-dependent, and the amplifier performance degrades rapidly in the event of a cooling system failure. The amplifiers are normally cryogenically cooled to a physical temperature of 12 K.

[1] JPL Summer Facility Fellow, Radio Frequency and Microwave Subsystems Section.

[2] Source code for all programs discussed in this article is available in hard copy or electronic form from the authors by request.

Another objective was the automation of the initial in-laboratory setting of bias conditions for the LNA. The existing manual procedure is both time-consuming and possibly less than optimal due to practical limitations.

II. System Description

The experiment employed a three-stage HEMT LNA, an existing test apparatus, and an optimization algorithm realized as a program written in the IBM BASICA language. The procedure consisted of four steps prior to employing the test apparatus.

The first step was to characterize the HEMT LNA gain and noise temperature as a function of physical temperature for fixed bias conditions [1], as shown in Fig. 1. (Note that the curves of Fig. 1 are for fixed (i, v) points for each amplifier stage; their loci are obtained by varying the physical temperature.) The second step was to model the HEMT LNA to be optimally controlled as a $Zm + 2$ port system, as shown in Fig. 2.

The controller has available for feedback the system measurable output vector

$$y = \begin{bmatrix} T_e \\ G \\ T \end{bmatrix}$$

which generates a control vector

$$u = \begin{bmatrix} i_1 \\ \cdot \\ \cdot \\ \cdot \\ i_m \\ v_1 \\ \cdot \\ \cdot \\ \cdot \\ v_m \end{bmatrix}$$

such that some performance metric

$$\Psi = \Psi(T_e, G)$$

is maximized (or minimized).

Some forms (to be maximized) that jointly reflect the noise minimization and gain maximization objectives are

$$\Psi = \frac{G}{T_e} \text{ nonlinear} \tag{1}$$

$$\Psi = a_1 G + a_2 \left(\frac{1}{T_e}\right) \text{ linear, simple} \tag{2}$$

$$\Psi = \left[G, \frac{1}{T_e}\right] \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} G \\ \frac{1}{T_e} \end{bmatrix} \text{ linear, quadratic} \tag{3}$$

The simple ratio Eq. (1) performance metric (a key figure of merit for receive systems) has been used in all subsequent work. The ratio is subject to internal parameter variations related nonlinearily to the operating temperature, T. Two system variables may be measured indirectly by power measurements: noise and gain. A third input variable, HEMT LNA temperature, may be measured directly.

The Zm control values must be generated for an m-stage LNA (six for a three-stage LNA: three bias voltages and three bias currents). The required model is a dc-bias-dependent operating point model of the amplifier circuit and solid-state devices, which includes internal noise sources. Model data at this level of detail are not available for the amplifier, although several constant bias trends with temperature are known from laboratory tests [1,2,3].

The third step was to write and exercise simulation models of an m-stage HEMT until a model was found whose input-output behavior was a good fit to the behavior of an actual HEMT. After studying available information on the relationship between drain current and voltage and the resulting noise and gain of an HEMT stage, a simple first-order model was derived. The drain bias current I_D is regulated by the gate source voltage V_{GS}, and the model incorporates this direct dependency.

The transconductance g_m of the isolated class A stage is expressed as a function of the gate-source voltage V_{GS}, the drain-source voltage V_D, and the junction temperature T

$$g_m = g_{m,max} \left(\frac{V_D}{V_{D,max}}\right)^{k_1} \left(1 - \frac{V_{GS}}{V_P}\right)^{k_2} \left(\frac{1}{\frac{T}{T_0} + 1}\right)^{k_3}$$

where k_i represents the experimentally determined constants, T_0 is the nominal temperature 300 K, V_P is the pinch-off voltage, $V_{D,max}$ is the maximum drain-source voltage, and $g_{m,max}$ is the maximum (most negative)

transconductance. Typical values for preliminary setup purposes were

$$k_1 = \frac{1}{2}$$

$$k_2 = \frac{1}{2}$$

$$k_3 = 1$$

$$V_P = -0.8 \text{ volts}$$

$$V_{D,max} = 5 \text{ volts}$$

$$g_{m,max} = -10 \text{ amps/volt}$$

As a first-order estimate, the gain G and noise temperature T_e are functions of the amplifier transconductance g_m alone

$$G = k_4 g_m$$

$$T_e = T_{e,min}\sqrt{\frac{g_{m,max}}{g_m}}$$

where typical values were used for the constants:

$$k_4 = -1$$

$$T_{e,min} = 10 \text{ K}$$

Hard constraints on the controllable parameters were

$$0 < V_D \leq 5 \text{ volts}$$

and

$$-0.8 \leq V_{GS} \leq 0 \text{ volts}$$

The fourth step was to select the figure of merit and optimization algorithm and apply them to the simulated system. The simple gain-to-noise linear ratio ($\Psi = G/T_e$) was used. A simple tracking-type optimization algorithm was adopted. The operating points immediately surrounding the current point (assumed optimum) are searched for relative optimality. Thus, variations in the true optimum point (e.g., due to temperature changes) are faithfully tracked. The optimization method uses no knowledge of the actual system or the manifold of the performance metric. The system is treated purely as a black box. While this method is very inefficient for general optimum locating problems, it is a reasonable choice for problems of this type, in which the search is initiated at or near the known global optimum, and the objective is to track slow changes in the location of that optimum point.

Each iteration of 2^n points is tested by perturbing the values of each parameter, measuring gain and noise, and calculating the performance metric. For the simple case ($n = 2$), only two parameters V_D and V_{GS} are involved. The four adjoining points tested are

$$V_D \pm \Delta V_D$$

and

$$V_{GS} \pm \Delta V_{GS}$$

ΔV_D and ΔV_{GS} may be variable in size, starting at a large perturbation and decreasing as convergence to the optimum proceeds.

For the first-order circuit model, the simulation reliably located the optimum at the maximal limits $V_{GS} = 0$ and $V_D = 5.0$ volts. This was expected, since the simple relationships used for G and T_e are maximized and minimized, respectively, by increasing g_m. Thus, the optimization seeks to maximize g_m, which occurs at the limiting values of V_D and V_{GS}. Clearly, a more sophisticated model for the HEMT amplifier stage is needed that specifically incorporates higher order effects. These effects are not apparent from simple device physics and dc observations.

A simple second-order modification of the noise relationship was tested:

$$T_e = T_{e,min}\left[\left(\frac{g_{m,max}}{g_m}\right)^{1/2} + k_6 g_m^2\right]$$

where a range of values for k_6 was tested. For the case of $k_6 = 0.2$, the optimal values of V_{GS} and V_D were found to be 0 and 4.3 volts, respectively. The optimum now occurs away from the external limits.

It is doubtful that an actual amplifier stage behaves according to this model, especially when operated at cryogenic temperatures. Also, stage-to-stage interactions in a multistage LNA would further raise the level of model complexity. The task of determining an accurate higher

order model would be a major project alone, beyond the scope and time limitations of this work. It was concluded that bias optimization, in general, was potentially beneficial and worthy of further study using actual hardware, which is described in the remainder of this article.

III. Experimental Apparatus

An existing noise/gain test apparatus (JPL Automated Test Bench) was used for data acquisition and control of a three-stage Field Effect Transistor (FET) LNA. These experiments were conducted at room temperature. This apparatus provided the ability to measure frequency-specific amplifier noise and gain, and included two digital-to-analog converters for computer control of the first-stage drain current and voltage. The second- and third-stage biases were manually set to fixed values. A block diagram of the test apparatus is shown in Fig. 3.

The JPL automated bench test software (BENCH.BAS) written in IBM BASICA served as the LNA monitor and controller. Using the interface routines from this software, a program (OPT.BAS) was written in BASICA which used this apparatus to optimize the first-stage drain current and voltage with respect to the previously described performance metric, G/T_e. Eventually, all the features of BENCH.BAS were incorporated into OPT.BAS, which resulted in a single integrated test and first-stage optimization package for multistage LNA's. A user-friendly tutorial-type interface was also added to assist users in setting up and calibrating the apparatus and performing tests or optimization. (Use of the program is largely self-explanatory. Type "OPT" and follow the prompts and help menus. Complete documentation is also available in the text file OPT.MAN.)

The program makes power and gain measurements by using formulas based on the derivations below. In these, N is a physical power measurement made by the power meter circuit of the bench apparatus; T_e is a noise temperature, both in units of power and expressed as absolute temperature in K; G is power gain, unitless and linear (not in decibels); the superscript H refers to the hot noise source; C refers to the cold noise source; the subscript F refers to the postamplifier (including internal amplification in the bench apparatus); and the subscript e refers to the amplifier under test.

During calibration, only the postamplifier is connected in the signal path between the noise sources and the power meter. Power measurements in this configuration are denoted by the subscript 2. A hot and a cold power measurement is made for each anticipated test frequency

$$N_2^H = (T_H + T_F)\, G_F$$

$$N_2^C = (T_C + T_F)\, G_F$$

$$\Delta N_2 = N_2^H - N_2^C = (T_H - T_C)\, G_F$$

During a measurement or optimization operation, with the amplifier under test in the signal path ahead of the postamplifier, another hot and cold power measurement (denoted by the subscript 1) is made at each point:

$$N_1^H = ((T_H + T_e)\, G_e + T_F)\, G_F$$

$$N_1^C = ((T_C + T_e)\, G_e + T_F)\, G_F$$

$$\Delta N_1 = N_1^H - N_1^C = (T_H - T_C)\, G_e G_F$$

The power gain of the amplifier under test is calculated from

$$\frac{\Delta N_1}{\Delta N_2} = G$$

The power ratio Y, or Y factor, is defined as

$$Y = \frac{N_1^H}{N_1^C} = \frac{((T_H + T_e)\, G_e + T_F)\, G_F}{((T_C + T_e)\, G_e + T_F)\, G_F}$$

This expression is solved for the noise contribution of the amplifier under test (referenced to its input) in terms of the previous power measurements and the known source load temperatures T_H, T_C, and the noise temperature contribution of the postamplifier. The amplifier noise temperature is given by

$$T_e = \frac{T_H - YT_C}{Y - 1} - \frac{T_F}{G_e}$$

The previously simulated simple tracking optimization algorithm was employed, and the size of the perturbation for each parameter was made adaptive (in the latest version). As the algorithm converges closer and closer to the optimal operating point, the bias conditions are perturbed less and less until the optimum is reached. The stopping criteria tests for a reduction of the perturbation parameters

to a useful lower limit. This approach aids in the accuracy of the optimization, hopefully without compromising the ability to track slowly varying changes in the optimum point due to temperature variation. If the optimal point shifts too rapidly, a loss-of-lock situation might temporarily occur due to the slowed tracking ability of the optimization with the smaller parameter perturbation limits. Since no temperature dependency tests were performed with the test apparatus, it was not possible to determine if this was a legitimate concern.

Using the apparatus and software, the three-stage FET LNA operating at room temperature (300 K) and 2.3 GHz was tested and bias optimized. It was not possible to test for temperature effects on the optimal bias settings, or the optimal tracking performance. Rather, tests were performed using the apparatus to locate the optimal from arbitrary initial points in the bias condition vector space.

For starting points close to the global optimum and small initial search increments, the optimal bias drain voltage and current are usually found within 15 minutes of run time, at approximately 2.9 volts and 12 mA, respectively. For initial search increments greater than ± 1 volt for V_D or ± 5 mA for I_D, local optimums with performance metrics less than the global optimum were sometimes located. Even though the search increment is reduced by the algorithm during convergence to the optimum, large initial search increments typically resulted in random results. The reported G/T_e figure at each point is also very sensitive to the initial calibration and hot noise source temperature measurement, which typically varied as much as 5 percent between successive runs at identical conditions.

Two significant problems were encountered using this apparatus. The first was the problem of noise. For each test point, two power measurements are made: one with a room-temperature noise source connected to the input of the amplifier under test, and the other with a liquid nitrogen-cooled "cold" noise source at the input. The sources are selected by an electromagnetically actuated switch controlled by the program. The two power measurements and their difference provide the necessary information for calculating the noise generated by the amplifier itself, as well as the amplifier gain.

Since the progress of the optimization requires very accurate noise and gain measurements at each test point, even a small error component in the power measurements could seriously affect the progress of the optimization toward a global optimum. The result is a "random walk" in some neighborhood of the optimum, or convergence to a false optimum due to an erroneously high-gain or low-noise measurement.

The solution to this problem was to take several measurements at each point and average them together. The accuracy of the performance metric calculated at each point was substantially improved using this technique, at the expense of a proportional increase in the time required to converge to an optimum.

The second problem was one of convergence time. Upon each iteration, 2^n points must be tested. Then, several power measurements must be made at each point to form the average. The optimization process could become lengthy if one started at a point far from the optimum. The switching time of the electromechanical noise source selector switch and the integration time of the rms power meter circuit are the underlying time-consuming factors. By replacing the electromechanical switch with an electronically switchable noise source (or sources), and employing faster power measurement methods, the convergence time of the optimization could be proportionally reduced.

IV. General Considerations for Use of LNA Gain/Noise Optimization

There seem to be two general categories of uses for LNA gain/noise optimization: off-line optimization with open-loop compensation during actual service, or on-line optimization using closed-loop compensation. There are advantages and disadvantages associated with each approach.

A. Off-Line Method

Prior to actual service, optimum bias conditions for the LNA are determined at each temperature. An appropriate optimization algorithm is used at several fixed temperature settings to determine the (not necessarily unique) values of i_j and v_j, $j = 1, \cdots, m$, which maximize the given performance metric $\Psi(T_e, G)$ for the actual system.

The lack of a complete system model requires the use of the model-independent optimization method described in the previous section. Such optimization methods are characterized by search techniques that start from some initial parameter setting and converge to a global optimum. Since the actual system rather than a mathematical model is used, convergence may take a considerable amount of time.

Another limitation of search-based optimization algorithms is possible failure to converge or convergence to a local rather than global optimum. The problem is intrinsic to deterministic approaches. Alternatively, a Monte Carlo method might be applied, one that randomly accumulates knowledge about the system and identifies the global optimum. This approach increases the run time of the algorithm. Excessive run times could be expected for an eight-parameter problem (assuming that four stages are optimized), as well as for the practical problem of avoiding possible damage to the LNA by eliminating any unusual bias combinations.

Fuzzy logic methods were also considered for the optimization and/or control problem due to limited knowledge about the system to be controlled. The simplicity of fuzzy methods, which use a heuristic approach to control law construction, was attractive. When these methods are applied to the off-line optimization process, some rule generation process, such as differential competitive learning (DCL) [3,4], might be used to generate a rule base for a fuzzy optimization algorithm (if one could be synthesized). The rule base may be considered equivalent to the optimum bias lookup table.

From an input/output perspective, fuzzy logic controls simply map an m-vector of inputs to an r-vector of outputs. The mapping is nonlinear and no different from conventional nonlinear full-state feedback controls from an input/output point of view. The novelty and power of the method derive from the relative simplicity by which the input/output mapping is constructed by using heuristic information rather than precise models [5].

It is uncertain if an adequate number of sensed system variables from the LNA are available for the use of fuzzy control methods. Fuzzy logic controls seem to be most successful in situations where sufficient measurable system variables are accessible, and a smaller number of control inputs and outputs must be generated. In the present application, there are a maximum of three measurable system variables (noise, gain, and temperature), and as many as eight control outputs to be determined, each with synergistic effects on the system. Fuzzy controls are typically suboptimal. The degree of optimality of the control is determined by the accuracy of the off-line rule-generation process.

The only area of applicability of fuzzy methods to this task seems to be in the formation of the rule base or stored optimum bias information table for the off-line optimization or the adaptive optimization/control approaches. Minimum-knowledge learning procedures developed for fuzzy logic control programming might increase the efficiency of generating the tables which characterize the system.

Regardless of the optimization method, the result is a simple stored lookup table with one input T and up to eight outputs (all bias voltages and currents). This table is then used for real-time setting of the bias conditions as a function of T when the LNA is in actual use.

A general limitation of this approach is the potential that other disturbances and parameter variations could alter the validity of the open-loop table, or rule base, between the time it was determined and the time that it is actually used. Another limitation is that the amplifier must be taken off-line and run through a controlled temperature and possibly time-consuming optimal bias determination procedure. With the objective of testing the amplifier under conditions as close to actual as possible, the procedure is most accurately done on the actual field installation, although laboratory characterization may be the only practical approach.

The difficulties associated with the nonlinear multiparameter optimization have already been addressed, and it should be noted that these difficulties are common to both off-line and on-line optimization methods.

The noteworthy advantage of the off-line approach is that the system would not be disturbed during actual use by an optimization search routine. Measurement of noise and gain might also be more easily (or possibly only) accomplished off-line rather than during critical real-time usage.

B. On-Line Method

The amplifier is continuously optimized during actual in-field use. Starting with known optimum bias settings at the normal operating temperature (e.g., 12 K), the bias conditions are continuously optimized in such a way as to maximize $\Psi(T_e, G)$. A requirement of this approach is that both T_e (noise) and G (gain) are measurable in real time without adversely disturbing the normal operation of the receiving system.

Real-time optimization methods for partially or completely unknown systems involve the use of search or optimum tracking algorithms. These require that each input parameter (bias voltages and currents, in this case) of the actual system be periodically perturbed, and the measurable components of the performance metric (noise and

gain, in this case) be sampled upon each iteration. Herein lies a key limitation of this approach, since the actual system might not tolerate such periodic small perturbations of all parameters. Obviously such parameter variations would have to be kept as small as possible within the precision limits of the noise and gain measurements. Temperature feedback is not required, since the method would continuously correct for all (slow) disturbances and parameter variations not limited to temperature effects alone. If started at or near an optimum point, this approach could be expected to track the moving optimum as the operating temperature varied.

The most notable limitation of this method is the need to perturb bias conditions during actual operation. This may or may not be acceptable, depending on the size and tolerability of the perturbations. Also, for this approach to successfully maintain an optimum, all disturbances and parameter variations (mainly temperature related) must change at a sufficiently slow rate to permit tracking of the optimum by the algorithm. An abrupt change (e.g., one due to a change made manually by the operator) could leave the tracking algorithm lost, seeking the nearest local optimum, off the optimal locus for temperature variation. The inefficiency of the search algorithm could make this a nontrivial concern, especially if the available real-time computational bandwidth is limited.

Furthermore, there is no guarantee that by starting from a global optimum for the normal operating temperature, the temperature-optimal locus would continue to be globally optimal. A quantum change in bias settings might be required to hop to a new global optimum at some temperature. However, such an abrupt change in bias conditions may not be tolerable due to a need for glitch-free reception, especially during critical data acquisition events.

The clearest advantage of the on-line approach is that the actual system at the time of operation is optimized for best noise/gain performance. If the starting point is a global optimum for the normal temperature, and the system itself is continuous with temperature variations, the bias parameters could be expected to change continuously and smoothly, maintaining at least a locally optimal setting as the temperature or other amplifier parameters slowly vary. Finally, no downtime is required for off-line optimization table generation (or rule learning).

C. Adaptive Off-Line and On-Line Methods

Another approach is a combination of both on-line and off-line optimization and control methods. Off-line characterization is used to determine an initial optimal operation table. During actual operation of the amplifier, an on-line optimization algorithm continuously updates the table based on real-time tracking of the optimum and some long-term adaptive strategy. Temperature feedback is required.

Abrupt system changes are not tolerated, since the controller contains stored information about the previous optimal parameter settings for any given temperature. The learning capability of the controller frees it from the long-term accuracy limitations of the off-line optimization/open-loop control approach.

The problem of on-line parameter perturbation remains, and the need for preliminary determination of the optimal parameter table or rule base also exists. This approach is the most robust, probably the most effective, but also the most complicated to put into actual practice.

V. Conclusions

A computer simulation of a generic HEMT-based multistage LNA has been completed by utilizing a first-order approximate model of a single stage at room temperature. The simple analog ratio of the amplifier gain to the amplifier noise temperature was selected as a readily measurable optimization metric.

Several optimization methods for unknown systems were evaluated with the aid of the simulation. These included fuzzy logic control methods in which a temperature compensation table is implemented as a fuzzy rule base. The best results for this application were achieved with simple tracking-type optimization algorithms. The simulation results provided justification for proceeding with work on actual hardware.

A program, OPT, was written in BASICA under DOS 3.3 to control the JPL Automated Bench for use as a first-stage bias optimization apparatus. The previous BENCH software was merged with OPT to create a single LNA test and optimization package for the apparatus. User-friendly features were incorporated to assist operators in performing calibration, amplifier testing, and optimization.

A three-stage FET LNA operating at room temperature and 2.3 GHz was tested and bias optimized. For starting points close to the global optimum, the optimal bias drain voltage and current were consistently found within a few minutes of run time, at approximately 2.9 volts and

12 mA, respectively. Noise on the power measurements required that all optimization test points be averaged over at least five iterations. For initial search increments greater than ±1 volt for V_D or ±5 mA for I_D, local optimization with performance metrics less than the global optimum were sometimes located.

Bias optimization for maximizing the gain-to-noise ratio of an LNA appears to be feasible in the laboratory, as does construction of temperature compensation look-up tables for bias parameters. The applicability of on-line bias optimization during actual service of an LNA remains uncertain, and further experimentation is planned.

References

[1] K. H. Duh, W. Kopp, P. Ho, P. Chao, M. Ko, P. Smith, J. Ballingall, J. Bautista, and G. Ortiz, "32-GHz Cryogenically Cooled HEMT Low-Noise Amplifier," *IEEE Transactions on Electron Devices*, vol. 36, no. 8, pp. 1528–1535, August 1989.

[2] H. Fukui, "Optimal Noise Figure of Microwave GaAs MESFET's," *IEEE Transactions on Electron Devices*, vol. 26, pp. 1032–1037, July 1979.

[3] B. Kosko, "Neural Networks and Fuzzy Systems," Englewood Cliffs, New Jersey: Prentice-Hall, 1992.

[4] S. Chiu, S. Chand, D. Moore, and A. Chaudhary, "Fuzzy Logic for Control of Roll and Moment for a Flexible Wing Aircraft," *IEEE Control Systems*, vol. 11, pp. 42-48, June 4, 1991.

[5] K. Self, "Designing with Fuzzy Logic," *IEEE Spectrum*, pp. 42, 44, and 105, November 1990.

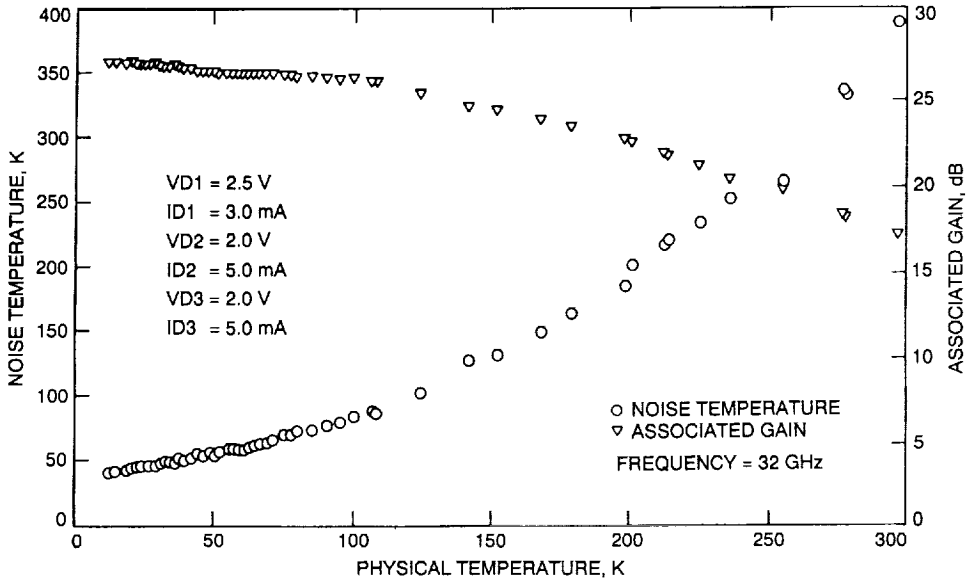
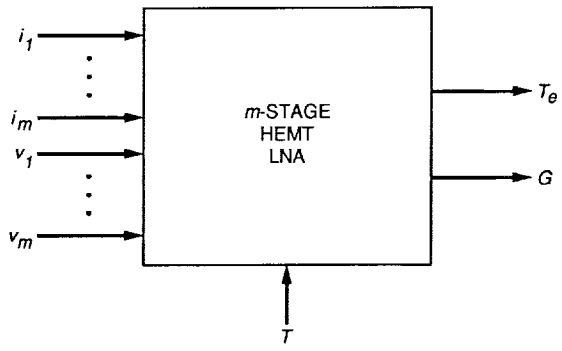


Fig. 1. Temperature dependency of noise and gain for a three-stage HEMT LNA.



m NUMBER OF AMPLIFIER STAGES
i_n BIAS CURRENT OF STAGE n
v_n BIAS VOLTAGE OF STAGE n
T_e EQUIVALENT AMPLIFIER INPUT NOISE TEMPERATURE
G OVERALL AMPLIFIER GAIN
T OPERATING TEMPERATURE (ABSOLUTE)

Fig. 2. System to be controlled or optimized.

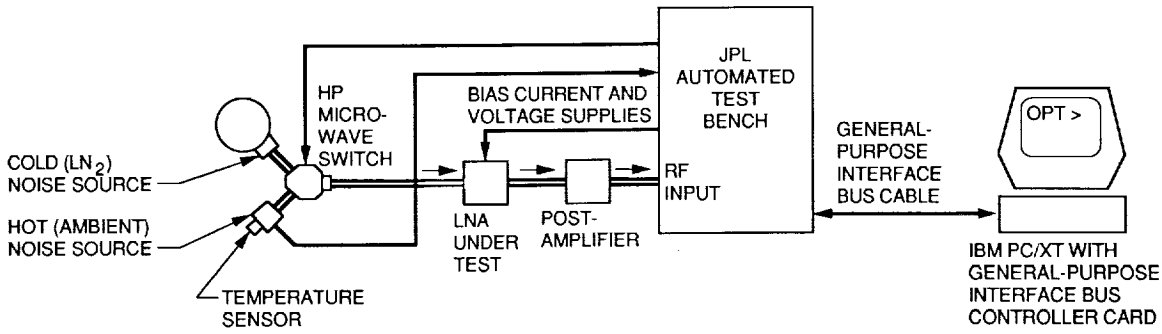


Fig. 3. LNA test and optimization apparatus.

Golay and Other Box Codes

G. Solomon[1]

The (24,12;8) extended Golay Code can be generated as a 6×4 binary matrix from the (15,11;3) BCH-Hamming Code, represented as a 5×3 matrix, by adding a row and a column, both of odd or even parity. The odd-parity case provides the additional 12th dimension. Furthermore, any three columns and five rows of the 6×4 Golay form a BCH-Hamming (15,11;3) Code. Similarly a (80,58;8) code can be generated as a 10×8 binary matrix from the (63,57;3) BCH-Hamming Code represented as a 9×7 matrix by adding a row and a column both of odd and even parity. Furthermore, any seven columns along with the top nine rows is a BCH-Hamming (63,57;3) Code.

A (80,40;16) 10×8 matrix binary code with weight structure identical to the extended (80,40;16) Quadratic Residue Code is generated from a (63,39;7) binary cyclic code represented as a 9×7 matrix, by adding a row and a column, both of odd or even parity.

I. Golay Code Properties

The (24, 12; 8) extended Golay Code possesses many properties. Solomon and Sweet [1] showed that it can be represented by a 6 × 4 binary matrix with equal row and column sums. Certain permutations of the matrix that keep the rows fixed give rise to at least three other boxes or matrices with identical row/column sum properties. These boxes can be used for "eyeball" decoding which avoids algebraic procedures. Here new properties of the extended Golay Code are further demonstrated.

A. Constructions

The (24,11;8) code in 6 × 4 matrix form is obtained from the BCH-Hamming (15,11;3) Code by adjoining row and column even parity. The BCH-Hamming Code here is expressed as a 5 × 3 matrix with entries in the *(i, j)* positions, $0 \leq i \leq 4$, $0 \leq j \leq 2$ corresponding to the coordinates $5i + 3j$ mod 15 of the code.

Let **A** be the BCH-Hamming (15,11;3) Code. The Mattson-Solomon (MS) polynomial for a code word $\mathbf{a} \in \mathbf{A} = (a_i; i = 0 \ldots 14)$ is given by

$$P_{\mathbf{a}}(z) = C_0 + \mathrm{Tr}\, Cz + \mathrm{Tr}\, Dz^3 + Ez^5 + E^2 z^{10}$$

where $C, D \in GF(16)$, $E \in GF(4)$, $C_0 \in GF(2)$, and $P_{\mathbf{a}}(\beta^i) = a_i$ for β a primitive 15th root of unity. Tr denotes the linear operator Trace in $GF(16)$. Tr $a = a + a^2 + a^4 + a^8$.

The parity check polynomial for the code is $(z + 1) \times f_1(z) f_3(z) f_5(z)$ where $f_i(z)$ is the irreducible polynomial over $GF(2)$ with β^i a root.

[1] Independent consultant to the Communications Systems Research Section.

The weight, $w(\mathbf{a}) \bmod 4$ for even-weight words $\mathbf{a}(C_0 = 0)$, is given by $w \bmod 4 = 2\Gamma_2(P_{\mathbf{a}}(x))$ where $\Gamma_2(P_{\mathbf{a}}(x)) = D^5 + D^{10} + E^3$ [2].

Now place the code words in 5×3 matrices (b_{ij}), $0 \le i \le 4$, $0 \le j \le 2$ corresponding to their values $5i + 3j \bmod 15$. The ith coordinate is entered thusly:

0	5	10
3	8	13
6	11	1
9	14	4
12	2	7

The MS polynomial expressed in the 5×3 setting, indexing each row by y in terms of the independent variable x, becomes $\mathrm{Tr}\, Dy^3 + \mathrm{Tr}'\,(E' + Cy + C^4y^4)x$; $E' = E^2$. Note again that for the rows, the trace is defined over $GF(4)$ as follows: $\mathrm{Tr}'\, a = a + a^2$ for $a \in GF(4)$.

Form the sum over the rows to give a sixth row with MS polynomial $\mathrm{Tr}'\, E'x$. Form the parity sum over the columns to obtain a 6×1 column, which is of course $\mathrm{Tr}\, Dy^3$; $y^6 = y$. The bottom row is indexed by $y = 0$, and the parity column corresponds to $x = 0$.

This is what is needed to prove the following results. Note that the coefficient of x is $E' + Cy + (Cy)^4$. This is the MS polynomial for a $(5, 3; 3)$ code indexed by $y^5 = 1$ over $GF(4)$. Adding a sixth row, one obtains a $(6, 3; 4)$ code indexed by $y^6 = y$ over $GF(4)$ as the coefficient of x. Note that the constant term in each row varies and is a $(5, 4; 2)$ binary code. It contributes the same values to the fourth parity column. Thus if one started with a BCH subcode of dimension 10 of even weight w with $w \bmod 4 = 2\Gamma_2$ where $\Gamma_2 = D^5 + D^{10} + E^3$, adjoining the parity rows and columns adjoins row and column code words whose weight modulo 4, $w \bmod 4 = D^5 + D^{10} + E^3$. So the total new weight $w' = 0 \bmod 4$.

This proves $w' \ge 8$. For if one started with $w = 4$, one has either $E^3 = 1$ and $D^5 + D^{10} = 1$, adding weight 4, or $E = 0$ and $D^5 + D^{10} = 0$, giving $D^5 = 1$, adding a column of weight 4.

One could also show easily that $w' \ge 8$ by noting that the coefficient of x is now a $(6, 3; 4)$ code over $GF(4)$, having adjoined an even parity row. Thus there are at least four rows of weight 2 each. The addition of the even parity column ensures $w \ge 8$ when the coefficient of $x = 0$. The new code words are of weights 8, 12, 16, and 24 in the 6×4 matrix code generated. Complementing these new code words still gives words of weights 8, 12, and 16, which takes care of the odd weight Hamming code words adding up to dimension 11.

The 12th dimension of the constructed code is obtained by adding odd parity row and column to the Hamming words. Thus the additional row (first 3 columns) is given by $1 + \mathrm{Tr}\, Ex$ and the additional column (upper 5 rows) is given by $1 + \mathrm{Tr}\, Dy$. The even weights are determined by $\Gamma_2 = E^3 + D^5 + D^{10}$. Consider the even weights equal to 4 and 6. If $E^3 = 0$ and 1, respectively, then weights 3 and 1, respectively, have been added to the bottom row. If $D^5 + D^{10} = 0$ and 1, respectively, then weights 1 and 3, respectively, have been added to the fourth column. If $D = 0$, then a 5 has been added.

Consider the case of $\Gamma_2 = E^3 + D^5 + D^{10} = 1$ or $w = 6$. Either $E \ne 0$, $D^5 = 1$, and weight 6 is added, or $E = 0$, $D^5 \ne 1$, and weight 2 is added. For $\Gamma_2 = 0$ or $w = 4$, $E \ne 0$ and $D^5 \ne 1$, so weight 4 is added, or $E = 0$ and $D^5 = 1$ and again weight 4 is added. In either case, adding odd parity row and column ensures that $w' = 0 \bmod 4$ and $w' \ge 8$.

This new code has weights 8, 12, 16, and 24. This is sufficient to guarantee that this code is the extended Golay Code by various uniqueness theorems in the literature. However, there is an explicit construction by Solomon and Sweet that does it.

B. The (24,12;8) Code Is the Golay Code

This formulation was first used by Solomon and Sweet [1]. The code has words of weight $w = 0 \bmod 4$ and is thus self-dual, has minimum distance 8, and contains the all one vector. This is the Golay Code. In fact, the correspondence between the coordinates of the cyclic code generated by the parity check polynomial and its representation as a 6×4 binary matrix is

0	2	1	3
4	12	7	10
9	22	6	11
16	15	8	19
20	21	18	13
17	∞	5	14

C. Encoding

Let $\mathbf{a}$ be a code word of length 24: $\mathbf{a} = (a_0, a_1, a_2, \ldots, a_{22}, \infty)$. Label positions $(0, 1, 2, \ldots, 22, \infty)$ generated by the recursion shift register rule

$$f(x) = x^{12} + x^{10} + x^7 + x^4 + x^3 + x^2 + x + 1$$

$$a_{n+12} = a_{n+10} + a_{n+7} + a_{n+4}$$

$$+ a_{n+3} + a_{n+2} + a_{n+1} + a_n$$

where $n = 0, 1, 2, 3, \ldots, 22$ and $a_\infty = \sum_{i=0}^{22} a_i$.

D. A Key Property

Theorem. Represent the Golay Code as a 6×4 binary matrix and consider any 5×3 submatrix obtained by removing one column and one row. This is a BCH-Hamming (15,11;3) Code. There is one proof and one verification.

Proof: Consider the 6×4 matrix with the top row deleted. Using the bottom parity check row and considering the first three columns, there is now a permuted 5×3 BCH-Hamming Code where the rows have been interchanged.

Note that the coefficient of x is the $(6,3;4)$ extended Reed-Solomon (RS) Code over $GF(4)$, which gives rise to the $(24,6;8)$ portion of the code. The map $y \rightarrow (1 + \alpha y + \alpha^2 y^4)$ is a permutation of this code that interchanges the top and bottom rows corresponding to $y = \beta^0 = 1$ and $y = 0$. For $\alpha = \beta^5$ a root of $x^2 + x + 1$, the second and fifth rows are interchanged and the third and fourth rows are fixed. Here y ranges over the values $y^6 = y$.

The remaining five dimensions, which are a function of C_0 and D in the BCH-Hamming Code, are easily seen to be manipulated so the weights stay the same. Since the code is clearly invariant under cyclic row cyclic permutations, this takes care of all subcodes with the first three columns fixed.

Now interchange the first column with the fourth rightmost parity column and the second with the third to obtain a BCH-Hamming Code still like the above in the top five rows. This interchange of columns is given by $x \rightarrow x + 1$.

This map takes the row indexed by y, Tr Dy+Tr' $(E' + Cy + (Cy)^4)x$, into a permuted row indexed by y, where $D \in GF(16)$ has been augmented: Tr' $(E' + Cy + (Cy)^4)$+ Tr Dy+ Tr' $(E' + Cy + (Cy)^4)x$. There clearly exists a D' such that D' = Tr' $(E' + Cy + (Cy)^4)$+Tr Dy for all values of y. Now clearly every three columns that occurred in the leftmost 5×3 matrix now occur in the newly formed 5×3 matrix. As the code is invariant under cyclic column permutation, the proof is complete. ❐

Verification: Postconjecture and preproof, a computer verification was performed by F. Pollara; this verification generated the identical weight distributions of the Hamming Code for each relevant permutation.

II. Extension of Results to (63,57;3) BCH-Hamming Code

Starting with the BCH-Hamming Code of length 63 in its 9×7 setting and using the MS polynomials for codes of lengths 63, 9, and 7, one obtains a 10×8 code of distance 8.

A. MS Polynomial for the BCH-Hamming Code in a 9 x 7 Setting

Let $f_1(x) = x^6 + x + 1$ be the primitive polynomial with β as a root. Then $f_i(x)$ is the irreducible polynomial with coefficients in $GF(2)$ with β^i as a root.

The BCH-Hamming Code in its MS polynomial form is written as

$$P(z) = C_0 + \sum \text{Tr } C_i z^i + \sum \text{Tr}' \; C_j z^j + C_{21} z^{21} + C_{21}^2 z^{42}$$

where $C_i \in GF(64)$; $i = 1, 3, 5, 7, 11, 13, 15, 23$; $C_j \in GF(8)$; $j = 9, 27$; and

$$\text{Tr}' \; a = a + a^2 + a^4; \; a \in GF(8)$$

Let $z \in GF(64)$ be a primitive root of $GF(64)$. Express $z = xy$ where

$$x = \beta^{9i}, \; 0 \le i \le 6; \; y = \beta^{7j}, \; 0 \le j \le 8$$

then

$$\text{Tr } Cz = \text{Tr } Cxy = \text{Tr}' \; [Cy + (Cy)^8]x$$

$$\text{Tr } C_3 z^3 = \text{Tr}' \; [C_3' y^6 + C_3'^8 y^3]x^6$$

$$C_3' = C_3^2$$

Replacing the letter by its primed letter to indicate a missing appropriate power,

$$\text{Tr } C_5 z^5 = \text{Tr}' \ (C'_5 y^2 + C'^8_5 y^7) x^6$$

$$\text{Tr } C_7 z^7 = \text{Tr} C'_7 y \text{Tr } C_9 z^9 = \text{Tr}' \ C'_9 x$$

$$\text{Tr } C_{11} z^{11} = \text{Tr}' \ (C'_{11} y^4 + C'^8_{11} y^5) x$$

$$\text{Tr } C_{13} z^{13} = \text{Tr}' \ (C_{13} y^4 + C^8_{13} y^5) x^6$$

$$\text{Tr } C_{15} x^{15} = \text{Tr}' \ (C'_{15} y^3 + C'^8_{15} y^6) x$$

$$\text{Tr } C_{23} z^{23} = \text{Tr}' \ (C'_{23} y^2 + C'^8_{23} y^7) x$$

$$\text{Tr } C_{27} z^{27} = \text{Tr}' \ C_{27} x^6$$

$$C_{21} z^{21} = C_{21} y^{21}$$

Recall that the Golay Code can be viewed in the MS polynomial formulation for lengths 6 or 4 as made up of components that are themselves RS Codes. Similarly express the Hamming Code here in MS polynomials of lengths 9 or 7. Recall that a binary codeword of length 7 has an MS polynomial of the form

$$P(x) = C_0 \ + \ \text{Tr}' \ (Cx \ + \ Dx^6)$$

$$\text{Tr}' \ C = C \ + \ C^2 \ + \ C^4; \ C \in GF(8)$$

Write the BCH-Hamming Code in all 57 dimensions as

$$C_0 + \text{Tr } C'_7 y^7 + C'_{21} y^{21} + C'^2_{21} y^{42}$$

$$+ \text{Tr}' \ (C'_9 + Cy + C^8 y^8 + C'_{11} y^4 + C'^8_{11} y^5$$

$$+ C'_{15} y^3 + C'^8_{15} y^6 + C'_{23} y^2 + C'^8_{23} y^7) x$$

$$+ \text{Tr}' \ (C_{27} + C'_3 y^6 + C'^8_{11} y^3 + C'_5 y^2$$

$$+ C'^8_5 y^7 + C_{13} y^4 + C^8_{13} y^5) x^6$$

where $C_i \in GF(64)$, $i = 1,3,5,7,11,13,15,23$; $C_j \in GF(8)$; $j = 9,27$.

Note that the coefficient of x in the above expression is (9,9;1) code over $GF(8)$, the coefficient of x^6 is a (9,7;3) RS code over $GF(8)$. The values $C_0 + \text{Tr } C'_7 y^7 + C'_{21} y^{21} + C'^2_{21} y^{42}$ taken over y form a (9,9;1) binary code. The minimum weight of this code is clearly 3. If the coefficient of x^6 is zero, then the minimum weight is given by a weight one word (coefficient of x) in the (9,9,1) code (giving rise to a weight 4 word) complemented by a value of C_0, which is 1 at that y position. If the coefficient of x^6 is nonzero, one again has a minimum weight 3 word.

Now extend the 9×7 matrix to 10×8 by adjoining even parity rows and columns. The 10th parity row is clearly the code word in MS form $C_0 + \text{Tr}' \ (C_9 x + C_{27} x^6)$. The coefficients $C_0 + \text{Tr } C'_7 y^7 + C'_{21} y^{21} + C'^2_{21} y^{42}$ in the row MS polynomials for $y^{10} = y$ now form a $(10,9;2)$ binary code. The coefficient of x is also a $(10,9;2)$ code over $GF(8)$. The coefficient of x^6 is a $(10,7;4)$ code over $GF(8)$. Adding the even parity column guarantees that the minimum weight w of the expanded code $w = 8$.

Thus the BCH-Hamming Code extends to a $(80,57;8)$ code. Finally, if the 10th row and 8th column are to be of odd parity, the minimum weight still is 8. It is obvious that the weight does not decrease this way. Consider words of weight $w \leq 7$ and try placing them in a 9×7 setting. Clearly $10 - w$ columns are at least zero and odd parity will certainly increase the weight this much.

B. BCH-Hamming Submatrices of the 10 x 8 Code

To show that every 9×7 submatrix of the 9×8 top portion of the matrix is also a Hamming Code, follow the technique used for the Golay Code. The map $x \rightarrow x + 1$ does an interchange of columns and replaces the first with the even parity column. A similar argument invoking cyclicity of the columns proves that the Hamming Code appears in every top 9×7 submatrix.

III. Quadratic Residue Code Properties (Box Codes)

The (63,39;7) binary cyclic code when extended by adding a row/column of odd and even parity has the weight structure of the extended (80,40;16) Quadratic Residue (QR) Code but is not isomorphic to it.

A. Constructions

The (80,39;16) in 10×8 matrix form is obtained from the (63,39;7) cyclic code by adjoining row and column even parity. The cyclic code here is expressed as a 9×7 matrix with entries in the (i,j) positions, $0 \leq i \leq 6$, $0 \leq j \leq 8$

corresponding to the coordinates $9i + 7j \bmod 63$ of the code.

Let **A** be the (63,39;7) cyclic code. The MS polynomial for a code word $\mathbf{a} \in \mathbf{A} = (a_i;\ i = 0 \ldots 62)$ is given by

$$P(z) = C_0 + \sum \mathrm{Tr}\ C_i z^i + \sum \mathrm{Tr}'\ C_j z^j + C_{21} z^{21} + C_{21}^2 z^{42}$$

where $C_i \in GF(64);\ i = 1,3,5,7,13;\ C_j \in GF(8);\ j = 9,27$; and

$$\mathrm{Tr}'\ a = a + a^2 + a^4;\ a \in GF(8)$$

Let $z \in GF(64)$ be a primitive root of $GF(64)$. Express $z = xy$ where

$$x = \beta^{9i},\ 0 \le i \le 6;\ y = \beta^{7j},\ 0 \le j \le 8$$

then

$$\mathrm{Tr}\ Cz = \mathrm{Tr}\ Cxy = \mathrm{Tr}'\ [Cy + (Cy)^8]x$$

$$\mathrm{Tr}\ C_3 z^3 = \mathrm{Tr}'\ [C_3' y^6 + C_3'^8 y^3]x^6$$

$$C_3' = C_3^2$$

Replacing the letter by its primed letter to indicate a missing appropriate power,

$$\mathrm{Tr}\ C_5 z^5 = \mathrm{Tr}'\ (C_5' y^2 + C_5'^8 y^7)x^6$$

$$\mathrm{Tr}\ C_7 z^7 = \mathrm{Tr}\ C_7' y \mathrm{Tr}\ C_9 z^9 = \mathrm{Tr}'\ C_9' x$$

$$\mathrm{Tr}\ C_{13} z^{13} = \mathrm{Tr}'\ (C_{13} y^4 + C_{13}^8 y^5)x^6$$

$$\mathrm{Tr}\ C_{27} z^{27} = \mathrm{Tr}'\ C_{27} x^6$$

$$C_{21} z^{21} = C_{21} y^{21}$$

Write the cyclic code in all 39 dimensions as

$$C_0 + \mathrm{Tr}\ C_7' y^7 + C_{21}' y^{21} + C_{21}'^2 y^{42}$$

$$+ \mathrm{Tr}'\ (C_9' + Cy + C^8 y^8)x + \mathrm{Tr}'\ (C_{27} + C_3' y^6 + C_3'^8 y^3$$

$$+ C_5' y^2 + C_5'^8 y^7 + C_{13} y^4 + C_{13}^8 y^5)x^6$$

where $C_i \in GF(64);\ i = 1,3,5,7,13;\ C_j \in GF(8);$ and $j = 9,27$. The generator polynomial for the code is

$$f_1(z) f_3(z) f_5(z) f_{13}(z)$$

where $f_i(z)$ is the irreducible polynomial over $GF(2)$ with β^i a root.

The weight, $w(\mathbf{a})$ mod 4 for even weight words $\mathbf{a}(C_0 = 0)$ is given by $w \bmod 4 = 2\Gamma_2(P_{\mathbf{a}}(x))$ where $\Gamma_2(P_{\mathbf{a}}(x)) = \sum_i C_i C_{-i} = \mathrm{Tr}'\ (C_7^9 + C_9 C_{54} + C_{21}^3)$ [2].

Now place the code words in 9×7 matrices (b_{ij}), $0 \le i \le 6,\ 0 \le j \le 8$ corresponding to their values $9i + 7j \bmod 63$. Note that the coefficient of x in the above expression is a (9,3;7) code over $GF(8)$, and the coefficient of x^6 is a (9,7;3) RS Code over $GF(8)$. The values $C_0 + \mathrm{Tr}\ C_7' y^7 + C_{21}' y^{21} + C_{21}'^2 y^{42}$ taken over y form a (9,9;1) binary code. The minimum weight of this code is clearly 7. If the coefficient of x and x^6 is zero, then the minimum weight is given by a weight one word (the coefficient of x) in the (9,9,1) code giving rise to a weight 7 word. If the coefficient of x^6 is nonzero, again there is a minimum weight 12 word, but, complemented by the constants, this can give rise to weight 9 at least.

Now extend the 9×7 matrix into a 10×8 matrix by adjoining even parity rows and columns. The 10th parity row is clearly the code word in MS form $C_0 + \mathrm{Tr}'\ (C_9 x + C_{27} x^6)$. The coefficients $C_0 + \mathrm{Tr}\ C_7' y^7 + C_{21}' y^{21} + C_{21}'^2 y^{42}$ in the row MS polynomials for $y^{10} = y$ now form a $(10,9;2)$ binary code. The coefficient of x is also a $(10,3;8)$ code over $GF(8)$. The coefficient of x^6 is a $(10,7;4)$ code over $GF(8)$. Adding the even parity column guarantees that the minimum weight w of the expanded code w equals 16.

Note that in the 10th row, $\Gamma_2 = \mathrm{Tr}'\ (C_9 C_{54})$. In the eighth column, $\Gamma_2 = \mathrm{Tr}'\ (C_7^9 + C_{21}^3)$. Thus coming from the length 63 cyclic code with $\Gamma_2 = \mathrm{Tr}'\ (C_7^9 + C_9 C_{54} + C_{21}^3)$ by adjoining a row and column of equal even parity, one has obtained a code with weights equal to 0 mod 4.

The 40th dimension of the constructed code is obtained by adding odd parity row and column to the code wc

and keeps the minimum weight and Γ_2 property. This new code has weights 16, 20, 24, 28, 32, 36, 40, 48, 56, and 64 and is self-dual. This is sufficient to guarantee that this code has the weight structure of the (80,40;16) extended QR Code [3].

B. Decoding

Place the code in its 10 x 8 box and compute row and column parities. Decide whether the code word is of even or odd row/column parity. If in doubt, assume first even and then try odd. Where a row is determined to have an odd number of errors, mark that row as an erasure. Otherwise assume an even number of errors in that row. To correct seven errors, there are at least three rows that must be correct, if seven row erasures are assumed. The coefficient of x is a (10,3;8) code, so one can extract that and generate a correct version. This leaves ten rows that are assumed to be BCH-Hamming Codes.

Only seven of these must be corrected to generate the entire word. Thus, if there is more than one error in each of four rows, but an odd number in the other three, then, with trial and error, $4 + 3 \times 3 = 13$ errors of a particular pattern can be corrected. If the 7-8-9 error patterns are such that three rows are clean, and at least four have single errors, then one can generate the (10,3;8) code over $GF(8)$ the coefficient of x and the (10,7;4) code over $GF(8)$ the coefficient of x^6. The rest emerges easily although this may require assuming first even and then odd parity of the row/column received code words. In general for error patterns of four or less, the row/column parity will be clear and the decoding simplified. In the event of even error patterns in the rows, one will have to decode the (10,3;8) code with some kind of modified RS decoding.

References

[1] G. Solomon and M. M. Sweet, "A Golay Puzzle," *Trans. Info. Theory*, vol. IT-29, no. 1, pp. 174–175, January 1983.

[2] G. Solomon and R. J. McEliece, "Weights of Cyclic Codes," *J. Comb. Theory*, vol. 1, no. 4, pp. 459–475, December 1966.

[3] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error Correcting Codes*, New York: North-Holland Publishing Co., 1977.

Figure of Merit for Direct-Detection Optical Channels

C.-C. Chen
Communications Systems Research Section

The capacity and sensitivity of a direct-detection optical channel are calculated and compared to those of a white Gaussian noise channel. Unlike Gaussian channels in which the receiver performance can be characterized using the noise temperature, the performance of the direct-detection channel depends on both signal and background noise, as well as the ratio of peak to average signal power. Because of the signal-power dependence of the optical channel, actual performance of the channel can be evaluated only by considering both transmit and receive ends of the systems. Given the background noise power and the modulation bandwidth, however, the theoretically optimum receiver sensitivity can be calculated. This optimum receiver sensitivity can be used to define the equivalent receiver noise temperature and calculate the corresponding G/T product. It should be pointed out, however, that the receiver sensitivity is a function of signal power, and care must be taken to avoid deriving erroneous projections of the direct-detection channel performance.

I. Introduction

Optical communication technology can offer potentially significant improvements in communication performance compared to current RF links. Much of the gain in optical receiver performance is due to the reduced operating wavelength and hence the increased antenna directivity. The small beam divergence resulting from the short operating wavelength implies that the received power needed for the communication link can be achieved with a much smaller transmit antenna aperture and a lower transmit power requirement. A smaller aperture also implies a lower weight communication package and hence an increased science payload capacity. Furthermore, a smaller beam divergence implies that the aperture size of future deep-space optical receiving terminals can be substantially smaller than the present-day Deep-Space Network (DSN) receivers.

For RF Earth receivers, an effective measure of the receiver performance is the ratio of the receiving antenna gain G_R to the system noise temperature T_{eq}. The received antenna gain G_R is proportional to the receiver aperture area and hence the received signal power. The channel efficiency, which is the maximum amount of information that can be relayed per unit of received energy,

directly related to the receiver noise temperature. For RF receivers, this G_R/T_{eq} ratio, known as the Earth receiver figure of merit [1], is directly proportional to the receiver signal-to-noise ratio (SNR). Given the transmitter and the link distance, the performance of the ground terminal can be compared by calculating the figure of merit without performing a complete end-to-end link analysis.

Since the G_R/T_{eq} figure of merit is a convenient parameter to compare the receiver performance, it is desirable to extend the concept to optical frequencies and to derive an equivalent figure of merit for optical receivers. For coherent optical receivers, the extension is straightforward. The resulting receiver noise temperature is given by $T_{op} = hf/k_B$, where hf is the energy of the photon and k_B is Boltzmann's constant. For direct-detection receivers, however, the concept of noise temperature cannot be applied directly. Although the minimum detectable power is equal to the energy of a single photon, the receiver performance can be effectively improved by trading the receiver sensitivity for increasing bandwidth [2,3]. In the limit of infinite bandwidth expansion, the direct-detection receiver can achieve very high channel efficiency that is limited only by the background noise level. Furthermore, unlike RF links where the channel efficiency depends only on the receiver noise temperature, direct-detection optical communication link performance is a function of both the signal and noise powers, and the ratio of peak-to-average signal power.

Because the performance of a direct-detection optical channel depends on the peak and average signal power in addition to the background noise, the performance of the direct-detection channel can be evaluated only if both ends of the link are defined. For a given background power and bandwidth, however, there exists an optimum condition under which the maximum amount of information can be transmitted across the channel per received signal photon. This optimal channel efficiency is achievable only under a particular set of signal power and bandwidth constraints. However, it allows one to derive an upper bound on the direct-detection channel performance.

The purpose of this article is to present a simple calculation of the receiver sensitivity for an ideal direct-detection optical channel subjected to a system bandwidth constraint. This optimal sensitivity can then be used to define the equivalent receiver noise temperature of the direct-detection channel. The equivalent G_R/T_{eq} parameter can then be calculated. Caution should be exercised, however, when using this ratio as it behaves quite differently from the corresponding ratio for RF systems.

II. Figure of Merit for Communication Links

An important figure of merit of an RF communication link is the channel efficiency C_E. The efficiency can be defined as the maximum amount of information that can be relayed per unit of received energy. For a white Gaussian noise channel with noise power spectral density N_0, the ideal channel efficiency is simply

$$C_E = \frac{\log_2 e}{N_0} \text{ (bits/joule)} \tag{1}$$

Given the amount of received signal power P_S, the channel capacity, which is defined as the maximum data rate that can be transmitted across the channel, is

$$R_{max} = P_S C_E = \frac{P_S}{N_0 \ln 2} = \frac{P_S}{k_B T_{eq} \ln 2} \text{ (bits/sec)} \tag{2}$$

Note that in the last equality we have replaced the noise power spectral density with $k_B T_{eq}$, where k_B is Boltzmann's constant and T_{eq} is the equivalent receiver noise temperature. The channel efficiency and channel capacity shown in Eqs. (1) and (2) are derived for additive white Gaussian noise (AWGN) channels without bandwidth constraint. Practical systems, with a limited bandwidth expansion factor, will have lower channel efficiency and capacity.

The received signal power P_S is a function of the transmit power, the transmitter and receiver parameters, and the link distance. With a diffraction-limited transmitter operating with aperture area A_T at a distance D, the signal power collected at the receiving terminal can be written as

$$\begin{aligned} P_S &= P_T \eta_T G_T \left(\frac{1}{4\pi D^2}\right) L_T(\theta_T) \eta_R A_R \\ &= P_T \eta_T G_T \left(\frac{\lambda}{4\pi D}\right)^2 L_T(\theta_T) \eta_R G_R \end{aligned} \tag{3}$$

where η_T and η_R are the transmitter and receiver efficiencies, G_T and G_R are the transmit and receive antenna gains, $L_T(\theta_T)$ is the pointing loss for an angular pointing error of θ_T, A_R is the receiver aperture area, and P_T is the transmit power. The transmit and receive antenna gains are defined as the ratio of far-field intensity when a signal is transmitted from the antenna to that which is radiated

from an isotropic radiator. For diffraction-limited apertures, the on-axis antenna gain is related to the aperture area and the operating wavelength λ by

$$G_T = 4\pi A_T/\lambda^2 \tag{4a}$$

$$G_R = 4\pi A_R/\lambda^2 \tag{4b}$$

Equation (3) shows that the total signal power received is proportional to the transmit and receive antenna gain and is inversely proportional to the link distance squared. The received power scales inversely with λ^2 because the far field intensity from a diffraction-limited transmitter is inversely proportional to λ^2. One should note that, although the received signal power is directly proportional to the receiver antenna gain in Eq. (4b), Eq. (3) does not imply that a diffraction-limited receiver must be used to collect the signal power. In fact, for a sufficiently large receiver field of view, the received power depends only on the total aperture area and not on the surface quality of the receiver. A nondiffraction-limited receiver, however, can admit much more noise power (Appendix A), or can degrade or even preclude the use of some forms of signal modulation.

By combining Eqs. (1) and (2), it is seen that the channel capacity is related to the transmit and receive parameters by

$$R_{\max} = \left(\frac{1}{4\pi D^2}\right) P_T \eta_T G_T L_T(\theta_T) \eta_R A_R \left(\frac{\log_2 e}{k_B T_{\text{eq}}}\right) \tag{5}$$

In general, it is desirable to derive a figure of merit which is independent of the link distance. For Gaussian channels, one such parameter is the product of channel capacity and link distance squared, $R_{\max} D^2$, which is given by

$$\begin{aligned} R_{\max} D^2 &= \left(\frac{1}{4\pi}\right) P_T \eta_T G_T L_T(\theta_T) \eta_R A_R \left(\frac{\log_2 e}{k_B T_{\text{eq}}}\right) \\ &= \left(\frac{\lambda}{4\pi}\right)^2 P_T \eta_T G_T L_T(\theta_T) \eta_R G_R \left(\frac{\log_2 e}{k_B T_{\text{eq}}}\right) \end{aligned} \tag{6}$$

The $R_{\max} D^2$ parameter defined in Eq. (6) is proportional to the transmit antenna gain, the receiver area, and is inversely proportional to the receiver noise temperature. For a given transmitter power and aperture area, this parameter is proportional to G_R/T_{eq}, which is commonly referred to as the Earth receiver figure of merit. The G_R/T_{eq} figure of merit is particularly useful in comparing different RF system performances. This is because for RF systems the transmitter aperture and available transmitter power usually do not vary much. As a result, the performance improvements are generally achieved by increasing aperture area, reducing operating wavelength, and lowering the receiver noise temperature; in other words, by improving the G_R/T_{eq} ratio.

III. Direct Detection Optical Channel

As was mentioned earlier, the channel efficiency of an ideal RF link, defined as the maximum amount of information that can be relayed across the channel per unit of received energy, is equal to $\log_2 e/k_B T_{\text{eq}}$ (bit/J). Note that the channel efficiency depends only on the receiver noise temperature T_{eq} and is independent of the signal power and aperture size. This simple expression for channel efficiency emerges because the only noise source present in the RF link model is the AWGN. An important aspect of the AWGN channel is that the error rate performance of the link depends only on the ratio of signal and noise powers (signal-to-noise ratio). Consequently, the G_R/T_{eq} parameter, which is proportional to the SNR, is a good measure of the receiver performance.

In contrast, the dominant source of noise in a direct-detection channel is the shot noise inherent in the signal, which cannot be modeled as AWGN. When detected using a square-law detector such as a photodiode, the shot-noise fluctuation can result in a fluctuation of the received photocount [4]. This self-noise fluctuation implies that, even when the amount of background noise admitted by the receiver is negligible, the number of photons collected (received signal energy) over a period ΔT can still fluctuate. It should be noted that the quantum fluctuation is also present in the RF receiver. However, such a fluctuation is usually ignored since the mean field is much larger than the rms fluctuation.

It can be shown that, for reception of a coherent signal field and multimode thermal radiation, the photocounts follow Poisson statistics [5]. For detection of single-mode background radiation, the photocounts follow Bose-Einstein statistics. For most direct-detection receivers, however, the large mode mismatch at the receiver implies that more than one spatial and temporal mode is being received and, in the limit of large mode mismatch, the background photocount process also exhibits Poisson statistics [5].

Because of the Poisson nature of the counting process, the direct-detection optical channel is also known as a Poisson channel. The received photocount over a given interval, ΔT, which contains contributions from both signal and background noise, can be modeled as a Poisson random variable with mean and variance given by [6]

$$< N > = (\lambda_S + \lambda_B)\Delta T$$

$$Var[N] = < N > = (\lambda_S + \lambda_B)\Delta T \qquad (7)$$

where λ_S and λ_B are the detected signal and background photocount rates measured in photons per second. The photocount rates are related to the power input by [5]

$$\lambda_S = \frac{\eta_D}{hf} P_S \qquad (8a)$$

$$\lambda_B = \frac{\eta_D}{hf} P_B \qquad (8b)$$

where η_D is the detector quantum efficiency, and hf is the energy of the photon at the operating wavelength. Equation (7) showed that the variation in the received photocount depends on both signal and background power. Because of this signal-dependent fluctuation, the receiver performance will depend on both signal and background power instead of a single quantity (signal-to-noise ratio).

The channel capacity and energy efficiency of an ideal direct-detection channel has been evaluated by several authors [2,3,7–9]. It has been shown that, for a direct detection channel with peak signal power P_S and an average to peak signal power ratio of σ, the channel capacity is given by [8,9]

$$R_{\max} = \log_2 e \, \lambda_S \left[\bar{q} \left(\frac{1+\rho}{\rho} \right) \ln \left(\frac{1+\rho}{\rho} \right) - (1-\bar{q}) \frac{1}{\rho} \ln \rho - \left(\frac{1+\bar{q}\rho}{\rho} \right) \ln \left(\frac{1+\bar{q}\rho}{\rho} \right) \right] \qquad (9a)$$

where

$$\rho = \lambda_S / \lambda_B = P_S / P_B \qquad (9b)$$

is the peak signal to background power ratio, and

$$\bar{q} = \min \left(\left(\frac{1+\rho}{\rho} \right)^{(1+\rho)/\rho} \rho^{1/\rho}/e - \rho, \sigma \right) \qquad (9c)$$

The quantity $\bar{q}$ is the ratio of average to peak signal power that achieves the channel capacity. When there is no average power constraint, that is, when

$$\left[\left(\frac{1+\rho}{\rho} \right)^{(1+\rho)/\rho} \rho^{1/\rho}/e - \rho \right] < \sigma$$

the channel capacity is limited only by the available peak signal power at the receiver. In this case the channel capacity is $\log_2 e \bar{q} \lambda_S$. Note that $\bar{q}\lambda_S$ is simply the average signal count rate. In other words, given the peak signal count rate λ_S, the maximum rate for which the information can be relayed across the channel is simply equal to the number of average received signal photons per second times $\log_2 e$. Equivalently, the channel can transmit $\log_2 e \approx 1.44$ bits per photon received.

On the other hand, when

$$\left[\left(\frac{1+\rho}{\rho} \right)^{(1+\rho)/\rho} \rho^{1/\rho}/e - \rho \right] > \sigma$$

the channel is said to be average power limited. The channel capacity decreases with decreasing average signal count rate, $\sigma\lambda_S$. However, for a constant peak signal count rate, the amount of information carried per signal photon, i.e., the channel efficiency, increases with decreasing average to peak power ratio, σ. Given the peak signal to background power ratio, ρ, the channel efficiency is given by

$$C_E = \frac{\log_2 e}{hf} \frac{R_{\max}}{\bar{q}\lambda_S} = \frac{\log_2 e}{hf} \left[\left(\frac{1+\rho}{\rho} \right) \ln \left(\frac{1+\rho}{\rho} \right) - \frac{1+\rho}{\rho} \ln \rho - \frac{1+\bar{q}\rho}{\bar{q}\rho} \ln(1+\bar{q}\rho) \right] \qquad (10)$$

The photon efficiency of the channel, which is defined as the maximum amount of information relayed through the channel per photon, is related to the channel efficiency by

$$C_{\text{ph}} = hf\, C_E$$

$$= \log_2 e \left[\left(\frac{1+\rho}{\rho} \right) \ln \left(\frac{1+\rho}{\rho} \right) - \frac{1+\rho}{\rho} \ln \rho - \frac{1+\bar{q}\rho}{\bar{q}\rho} \ln(1+\bar{q}\rho) \right] \quad (11)$$

In the limit as the average to peak signal power ratio goes to zero, the photon efficiency is given by

$$\lim C_{\text{ph}} = \log_2 e \times \left[\left(\frac{1+\rho}{\rho} \right) \ln \left(\frac{1+\rho}{\rho} \right) + \frac{1}{\rho} \ln \rho + \ln \rho - 1 \right] \quad (12)$$

Equations (9) through (12) illustrate an important aspect of the direct-detection channel; namely, the receiver sensitivity is a function of the average to peak signal power ratio (duty cycle), as well as the signal and noise power levels. Furthermore, one can trade bandwidth for receiver sensitivity by choosing modulation schemes with low average to peak power ratio. Plotted in Fig. 1 is the limiting receiver sensitivity as $\sigma \to 0$ versus the signal power to background noise ratio, ρ. As expected, the photon efficiency increases indefinitely as the amount of background noise decreases. Consequently, the limiting efficiency of a direct-detection channel is limited only by the amount of background noise the receiver collects.

Given the peak signal count rate λ_S, the channel capacity (data rate) can be optimized by letting the average count rate approach

$$\left[(1+\rho/\rho)^{(1+\rho)/\rho}\, \rho^{\frac{1}{\rho}}/e - \rho \right] \lambda_S$$

In this case, the limiting receiver sensitivity is $\log_2 e$ bits per photon. For near-Earth links where maximum data rate, and not maximum photon efficiency, is desired, the link should operate with average to peak signal power ratio equal to

$$\left[\left(\frac{1+\rho}{\rho} \right)^{(1+\rho)/\rho} \rho^{1/\rho}/e - \rho \right]$$

Since the channel capacity is achieved with a relatively high average to peak power ratio, semiconductor lasers, which can be modulated at a higher bandwidth, are more suitable for achieving high data rates. On the other hand, for a deep-space link where high power efficiency is desired, one should operate the link with a low average to peak power ratio, i.e., by choosing laser and modulation schemes such that $\sigma \to 0$, while holding the average power constant. Solid-state lasers, which can provide high peak power with a low duty cycle, are ideally suited for such applications.

IV. Bandwidth-Limited Direct-Detection Channel and Its Figure of Merit

In practice, the limiting performance given by Eqs. (9) through (12) cannot be achieved. Several factors contribute to the limitations on channel performance [10]. First, the timing resolution (bandwidth) of the receiver is limited by the response of the photodetector material, the timing resolution of the photodetector/preamplifier assembly, and the complexity of the decoding electronics. Furthermore, the bandwidth of the channel is affected by the modulation bandwidth of the laser. The maximum timing resolution at the receiver is bounded by the uncertainty principle to approximately 0.1 psec [10], and, in practice, the detector timing resolution is limited by the complexity of the electronics to approximately 1 nsec.

The amount of background noise admitted by the receiver is, in principle, limited only by the optical background noise. Even when the Sun (6000 K) is in the field of view, a diffraction-limited receiver with a 1-GHz-wide predetection filter will observe only 2.5×10^6 background counts/sec when operating at 532-nm wavelength. In practice, the nondiffraction-limited receiver admits much more background noise (Appendix A), and the predetection filter is usually much wider than the signal bandwidth. Furthermore, detector dark counts constitute an irreducible background level which is present even when no bright background object is absent. As a result, a practical receiver will admit much more background noise than the thermal background noise limit.

Another factor affecting channel performance is the average to peak signal power ratio. Although in principle an infinite peak to average power ratio is achievable, the maximum achievable peak to average power ratio for a solid state laser is limited by the laser parameters such as the pump power to threshold power ratio. A practical direct-detection channel will reach its complexity and bandwidth expansion limit long before it reaches the capacity limit.

As a result, a practical direct-detection channel will have performance that is much less than the theoretical thermal background noise limit as predicted by Pierce and Posner [3].

Shown in Table 1 is a list of projected parameters for direct-detection receivers. For short-term development support, a 1-m diffraction-limited telescope can be used to collect the downlink signal. The blur circle diameter of the 1-m diffraction-limited telescope is limited by atmospheric seeing to approximately 20 times its diffraction limit. Actual operational support for deep-space missions is planned by using several 10-m-class photon-bucket receivers to provide spatial diversity reception. The blur circle diameter of the 10-m photon-bucket is estimated to be approximately 2000 times more than an equivalent diffraction-limited aperture. The substantially worse surface quality implies that a much larger amount of background noise will be collected by the 10-m receiver. However, the larger collecting area can actually result in an improved system performance. A prototype research and development station, the Deep Space Optical Receiving Antenna (DSORA), is currently being studied by JPL, and a facility construction request for DSORA has been submitted to NASA for a projected 1997 start. The parameters shown in Table 1 are the projected parameters for DSORA.

It is seen from Eqs. (9) through (11) that the performance of direct-detection optical channels is determined by the average and peak signal powers, as well as the background noise power. Given a set of link parameters, the average signal and background powers can be calculated using Eq. (3) and Appendix A. For a constant average signal power, the channel performance can be improved by increasing the peak power to average signal power ratio. Since the receiver bandwidth limits the maximum rate at which arriving photons can be distinguished, the peak signal count rate, λ_S, of a practical direct-detection channel is simply equal to its receiver bandwidth limit. Based on the receiver bandwidth and average signal and background power, the channel efficiency and capacity of a band-limited direct-detection channel can be calculated. The equivalent noise temperature of the optical receiver can then be defined by equating the optimum channel efficiency with $\log_2 e/k_B T_{\text{eq}}$. That is,

$$C_{E_{\max}} \equiv \frac{C_{\text{ph}}}{hf} \equiv \frac{\log_2 e}{k_B T_{\text{eq}}} \text{ (bits/joule)} \qquad (13)$$

This equivalent noise temperature represents an upper bound of the direct-detection link performance. For a constant transmit power and aperture, the G/T figure of merit can be defined as:

$$(G/T)_{\text{optical}} = \eta_R \frac{(4\pi A_R/\lambda^2)}{T_{\text{eq}}} \qquad (14)$$

where T_{eq} is the equivalent noise temperature as given by Eq. (13), and η_R is the combined efficiency of the receiving optics and detector. The G/T ratio defined in Eq. (14) may also be interpreted as the equivalent G/T ratio needed to achieve similar performance for a coherent receiver. Since the direct detection receiver is not actually diffraction limited, we should refrain from calling the quantity in the numerator the receive antenna gain. One should note that the figure of merit as defined in Eq. (14) is calculated only under fixed signal and background power levels and receiver bandwidth. The figure of merit will vary as the link distance and power level change. Therefore, care should be taken when using the figure of merit to estimate the link performance at different power levels.

The projected performance of the direct-detection receiver is calculated and summarized in Table 2. The receiver figures of merit are calculated under night sky viewing and when Saturn is within the receiver field of view. The spectral irradiance of the night sky is assumed to be 5×10^{-6} W/m^2·sr·mm. Given the background irradiance, the receiver bandwidth, and the optics efficiency, the total amount of background power collected by the receiver can be calculated. The maximum photon efficiency can then be calculated by assuming a modulation format with very small average to peak signal power ratio ($\sigma \to 0$). Shown in Table 2 are the maximum photon efficiencies achievable given the bandwidth and background noise constraints. It is seen that, when viewing night sky only, a photon efficiency of 15.2 bits/photon can be achieved with the 1-m receiver, whereas 19.8 bits/photon can be achieved using the 10-m photon bucket. When Saturn is in the field of view, however, the efficiency decreases to 9.3 bits/photon for the 1-m receiver and 7.7 bits/photon for the 10-m photon bucket. The better performance of the 1-m receiver when Saturn is in the FOV is due primarily to the reduced background power. The equivalent noise temperature of the receiver is then calculated using Eq. (13). The calculated equivalent noise temperatures are between 1970 and 25700 K when viewing night sky and between 4200 and 5070 K when Saturn is in the field of view.

Based on the equivalent noise temperature, a G/T figure of merit can then be calculated. The calculated receiver figure of merit is between 86.1 and 88.2 dB/K for the 1-m photon bucket and between 109.5 and 113.6 dB/K

for the 10-m photon bucket. The variation is due to the reduction in receiver sensitivity due to the presence of background noise. One should note that the term G_R in Eq. (14) bears no relationship to the actual antenna gain, which is defined using its effective isotropic radiated power, since the photon bucket is not diffraction limited. Furthermore, the optimal receiver noise temperatures shown in Table 2 are calculated by assuming a low modulation duty cycle ($\sigma \rightarrow 0$) and a large bandwidth expansion factor. As a result, the limiting performance calculated in Table 2 is achievable only under a very low data rate. Practical direct-detection channels operate with a nonzero average to peak power ratio and hence have a higher noise temperature than that shown in Table 2.

V. Pulse Position Modulation for Direct-Detection Optical Channel

One practical modulation scheme which has low average to peak power ratio is pulse-position modulation (PPM), in which the information is conveyed through the channel by the time window in which the signal pulse is present [11,12]. In an M-ary PPM channel, each code word period is divided into M time slots. The transmit alphabet contains M symbols. Each symbol has a unique pulse location within the M time slots. At the receiving end, the decoder simply inspects the time windows and determines which time slot contains the signal pulse. Direct-detection PPM has been shown to be very effective in achieving high energy efficiency [13–15]. In the limit of infinite bandwidth expansion, i.e, at $M \rightarrow \infty$, PPM can achieve a photon efficiency limited only by the background noise level. Practical implementation constraints [10], however, limit the maximum order of PPM and hence impose an upper limit on the receiver sensitivity.

As expected, the performance of a PPM channel depends on the peak signal count rate, λ_S, the background count rate, λ_B, the modulation slot time, ΔT, and the order of modulation, M. Detailed derivation of the PPM channel capacity is given in Appendix B. Given the signal and background powers and slot time, ΔT, there exists a modulation order M that optimizes the channel efficiency. Shown in Fig. 2 are the maximum values of photon efficiency versus the average signal counts per slot for several background powers. Shown in Fig. 3 are the corresponding PPM alphabet sizes which achieve maximum photon efficiency. The maximum PPM order was artificially constrained to 256 to illustrate the effect of limited modulation bandwidth expansion. At high signal powers, the sensitivity of the receiver is limited by the modulation bandwidth, $1/\Delta T$, and the photon efficiency decreases with increasing average signal power. As the average power decreases, the photon efficiency can be improved by using a higher order PPM. If the maximum order of PPM is constrained, however, the photon efficiency will eventually reach a maximum. Further decrease in signal power will result in reducing photon efficiency. The sharp corners for the curves in Fig. 2 near 0.01 average signal counts/slot are due to the limiting PPM alphabet size.

It is seen from Fig. 2 that, in the limit of large signal power, the receiver sensitivity decreases. However, the channel capacity per slot, $R_{\max}\Delta T$, approaches a constant. This is because at high signal power, $R_{\max}$ is limited by the bandwidth. Shown in Fig. 4 is a plot of the capacity per slot versus the average signal count rate. Note that the limiting performance of approximately 0.53 bit/slot is achieved with average signal count per slot greater than 5.

The performance of the PPM modulated channel can be compared to that of a channel constrained only by the bandwidth. Shown in Table 3 are link design tables for the theoretically limited channel and two examples of PPM channels. It is seen that, for a Saturn return link using a 2-W transmitter and a 60-cm-diameter aperture, the maximum data rate sustainable is 52 Mbps. The maximum photon efficiency is 6.1 bits/photon and the equivalent noise temperature is 6,400 K. The capacity of an M-ary PPM channel, on the other hand, is somewhat lower. For a PPM channel with a 10-nsec slot time, the capacity is approximately 15 Mbps with an equivalent noise temperature of 22,170 K. For a PPM channel with a 1-nsec slot time, the capacity is 32 Mbps with an equivalent noise temperature of 10,240 K. Note that the effective photon efficiency and hence the receiver noise temperature change with modulation format. The $R \times D^2$ parameter can also be calculated based on realistic link parameters. Recall that the figure of merit defined in Eq. (14) is achievable only at the optimal signal level and modulation format. For a Saturn return link with the parameters shown in Table 3, the achievable photon efficiency is 6.1 bits/photon. This is very close to the 7.7-bits/photon limit based on the noise power. The resulting maximum $R \times D^2$ parameter is approximately 320 dB. For a PPM channel with a 1-nsec slot time, the $R \times D^2$ parameter is approximately 318.5 dB.

VI. Conclusion

The discussion of the direct-detection channel above illustrates the difficulty in defining a figure of merit of the direct-detection receiver. The Poisson statistics imply

that the system cannot be characterized using only the signal-to-noise ratio. Instead, receiver sensitivity depends on both signal and noise power. Furthermore, sensitivity depends on the modulation format. One can trade system bandwidth (via reducing the average to peak power ratio) for photon efficiency. As a result, the performance of the channel can be characterized only by defining both ends of the system contrary to the microwave case.

Although the actual performance cannot be defined without knowing the signal and background powers, one can derive the limiting channel capacity based on the background noise power and bandwidth constraints and, from the limiting channel performance, define an equivalent receiver noise temperature. The system performance calculated using the G/T parameters, therefore, can be used as an indicator of the limiting channel performance given the bandwidth and noise power constraints. Based on the projected receiver parameters shown in Table 1, it is seen that the maximum G/T ratio for an optical receiver in 10 to 15 years is 113.6 dB/K when operating under a night-sky background. When Saturn or Jupiter is within the field of view, the G/T ratio decreases to approximately 109.5 dB/K. The limiting G/T performance calculated above is achieved at low average to peak power ratio and, consequently, at a very low data rate. For the practical link shown in Table 3, the G/T ratio is between 103.0 dB/K for a 16-ary PPM channel with a 10-nsec slot time, and 108.4 dB/K for a 64-ary PPM with a 1-nsec slot time.

It should be emphasized that the G/T ratio calculated in this article is an indicator of the limiting channel performance given the bandwidth and noise power constraints. Practical system performance can be substantially improved by increasing the receiver bandwidth. Furthermore, unlike RF channels where the noise temperature can be easily measured using a known background source, the equivalent noise temperature of a direct-detection PPM channel depends on both peak signal to background power ratio, ρ, as well as the average to peak signal power ratio, σ. Changes in link parameters such as aperture size, link distance, and transmitter power can result in changes in signal and noise powers admitted by the receiver and hence in the equivalent noise temperature. As a result, care must be taken in applying the equivalent noise temperature and hence the G/T ratio in projecting direct-detection link performance.

Acknowledgments

The author would like to acknowledge the following individuals for their helpful discussions: C. Stelzried, S. Butman, J. Lesh, T. Yan, and R. Clauss.

References

[1] W. L. Morgan and G. D. Gordon, *Communications Satellite Handbook*, New York: John Wiley & Sons, 1989.

[2] J. R. Pierce, "Optical Channels: Practical Limits With Photon Counting," *IEEE Trans. Commm.*, vol. COM-26, no. 12, pp. 1819–1821, 1978.

[3] J. R. Pierce, E. C. Posner, and E. R. Rodemich, "The Capacity of the Photon Counting Channel," *IEEE Trans. Info. Theory*, vol. IT-27, no. 1, pp. 61–77, January 1981.

[4] P. L. Kelley and W. H. Kleiner, "Theory of Electromagnetic Field Measurement and Photoelectron Counting," *Phys. Rev.*, vol. 136, pp. A316–A334, 1964.

[5] S. Karp and J. R. Clark, "Photon Counting: A Problem in Classical Noise Theory," *IEEE Trans. Info. Theory*, vol. IT-16, no. 6, pp. 672–680, 1970.

[6] A. Papoulis, *Probability, Random Variables, and Stochastic Processes,* New York: McGraw-Hill, 1965.

[7] J. L. Massey, "Capacity, Cutoff Rate, and Coding for a Direct-Detection Optical Channel," *IEEE Trans. Commun.,* vol. COM-39, no. 11, pp. 1615–1621, November 1981.

[8] M. H. A. Davis, "Capacity and Cutoff Rate for Poisson-Type Channels," *IEEE Trans. Info. Theory,* vol. IT-26, no. 6, pp. 710–715, November 1980.

[9] A. D. Wyner, "Capacity and Error Exponent for the Direct Detection Photon Channel—Part I," *IEEE Trans. Info. Theory,* vol. 34, no. 6, pp. 1449–1961, November 1988.

[10] S. A. Butman, J. Katz, and J. R. Lesh, "Practical Limitations on Noiseless Optical Channel Capacity," *DSN Progress Report,* vol. 42-55, Jet Propulsion Laboratory, Pasadena, California, pp. 12–14, November 1979.

[11] R. Gagliardi and S. Karp, "M-ary Poisson Detection and Optical Communications," *IEEE Trans. Comm. Theory,* vol. CT-17, no. 2, p. 208, 1969.

[12] R. J. McEliece, "Practical Codes for Photon Communications," *IEEE Trans. Info. Theory,* vol. IT-27, no. 4, pp. 393–397, 1981.

[13] J. R. Lesh, "Capacity Limit of the Noiseless Energy-Efficient Optical PPM Channel," *TDA Progress Report 42-67,* vol. July–September, Jet Propulsion Laboratory, Pasadena, California, pp. 54–58, November 15, 1981.

[14] R. G. Lipes, "Pulse-Position Modulation Coding as Near-Optimum Utilization of Photon Counting Channels with Bandwidth and Power Constraints," *DSN Progress Report,* vol. 42-56, Jet Propulsion Laboratory, Pasadena, California, pp. 108–113, January 1980.

[15] D. L. Zwillinger, "Maximizing Throughput Over an Average-Power Limited and Band-Limited Optical Pulse Position Modulation Channel," *TDA Progress Report 42-62,* vol. April–June 1981, Jet Propulsion Laboratory, Pasadena, California, pp. 81–86, January 1981.

Table 1. Projected receiver parameters for a direct detection optical receiver.

Parameters	Near term	Projected in 10–15 years
Detector quantum efficiency, μm	0.5 at 0.532	0.8 at 0.532
Receiver optics efficiency	0.4	>0.5
Filter transmissivity	0.6	>0.8
Filter bandwidth, nm	0.1	<0.001
Receiver diameter, m	≈1 diffraction limited	10, photon bucket
Receiver spatial mode mismatch	≈400	$\approx 4 \times 10^6$
Temporal bandwidth mismatch	$\approx 5 \times 10^3$	≈12
Receiver bandwidth, MHz	200	1000
Maximum data rate, Mbps	106	530
Detector dark count rate	2,000	200
Operating wavelength, nm	532	532

Table 2. Loss factors for a direct detection ground receiver.

Time, years	Near term		In 10–15 years	
	Night sky only	Saturn in FOV	Night sky only	Saturn in FOV
Antenna gain, dB[a]	134.5	134.5	154.5	154.5
Efficiency, dB				
Optics efficiency	−6.2	−6.2	−4.0	−4.0
Detector quantum efficiency	−3.0	−3.0	−1.0	−1.0
Atmospheric attenuation[b]	−3.0	−3.0	−3.0	−3.0
Receiver bandwidth, MHz	200	200	1000	1000
Total background count rate	2000	1.2×10^5	400	1.8×10^6
Maximum photon efficiency, bits/photon	15.2	9.3	19.8	7.7
Equivalent noise temperature, K	2570	4200	1970	5070
Figure of merit, dB/K	88.2	86.1	113.6	109.5

[a] Antenna gains are calculated by assuming the telescopes have 18.5-percent central obscuration area.

[b] Assume single site, 65-percent weather confidence at 30-deg elevation.

Table 3. Projected link performance for a Saturn return link using a 1-GHz bandwidth receiver.

Parameters	Theoretical limit	MPPM (10-nsec slot time)	MPPM (1-nsec slot time)
Transmitter power, W	2	2	2
Transmitter aperture, cm	60	60	60
Transmitter optics efficiency	>0.5	>0.5	>0.5
Link distance, AU	10	10	10
Atmospheric attenuation	0.5	0.5	0.5
Receiver aperture, m	10, photon bucket	10, photon bucket	10, photon bucket
Receiver optics efficiency	>0.5	>0.5	>0.5
Narrow-band filter transmission	>0.8	>0.8	>0.8
Filter bandwidth, nm	<0.001	<0.001	<0.001
Detector quantum efficiency	0.5	0.5	0.5
Receiver spatial mode mismatch	4×10^6	$\approx 4 \times 10^6$	$\approx 4 \times 10^6$
Receiver bandwidth, MHz	1000	1000	1000
Modulation	N/A		
Slot width, nsec		10	1
PPM order		$M = 16$	$M = 64$
Background count rate	1.8×10^6 (Saturn in FOV)	1.8×10^6 (Saturn in FOV)	1.8×10^6 (Saturn in FOV)
Signal photocount rate	8.6×10^6	8.6×10^6	8.6×10^6
Effective photon efficiency, bits/photon	6.1	1.76	3.81
Effective noise temperature, K	6,400	22,170	10,240
Channel capacity, Mbps	52	15.1	32.7
G/T value, dB/K	108.4	103.0	106.4
$R_{max}D^2$, dB(m^2/sec)	320.7	315.3	318.7

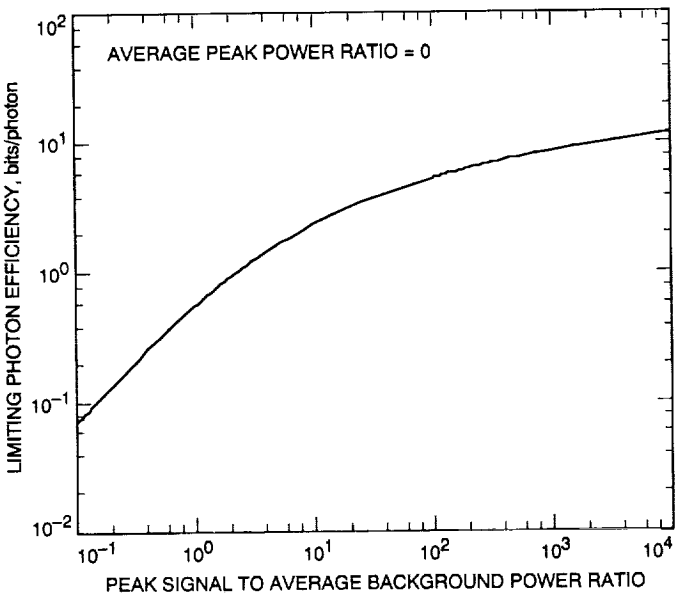


Fig. 1. Photon efficiency versus peak signal to average background power ratio in the limit of average power to peak power ratio equals 0 (i.e., with extremely low duty cycle).

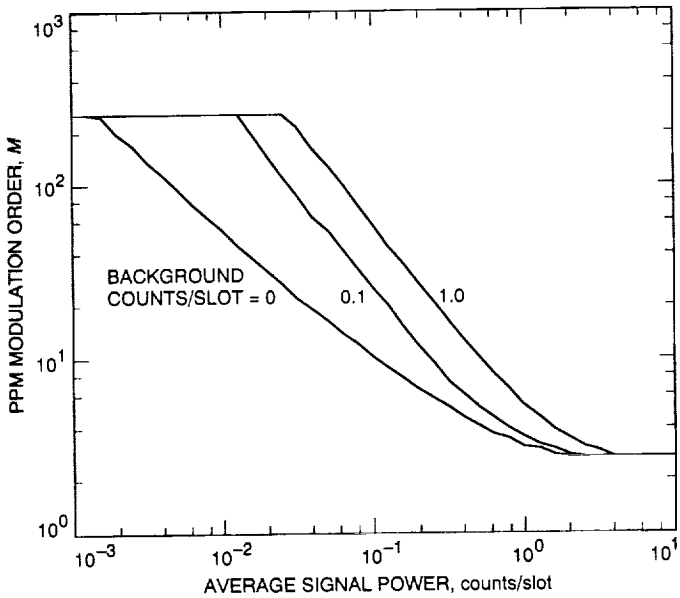


Fig. 3. PPM alphabet size that achieves maximum photon efficiency versus the average signal power for an *M*-ary PPM channel with constant slot width. Maximum PPM alphabet size is constrained to 256.

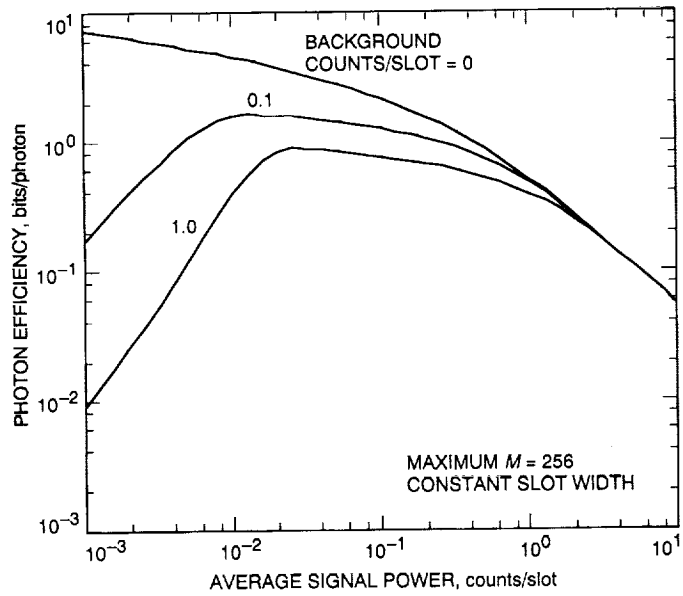


Fig. 2. Optimum photon efficiency versus average signal power for a PPM channel with constant slot width and maximum *M* constrained to 256.

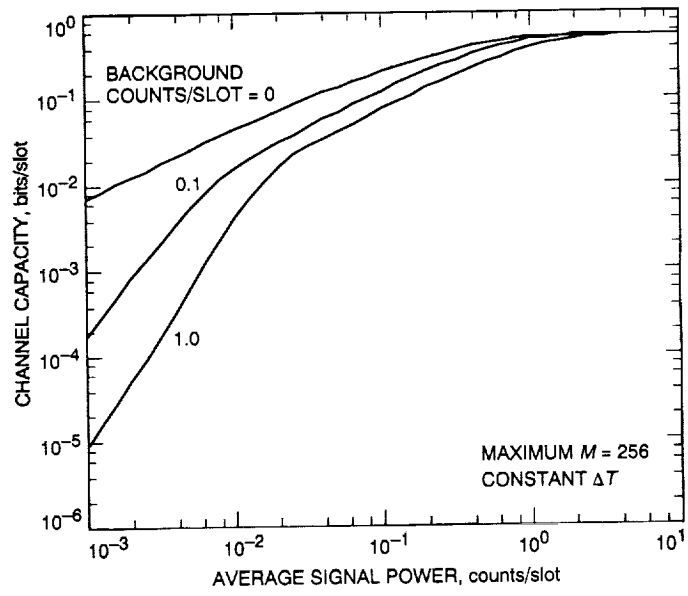


Fig. 4. Optimum channel capacity versus average signal power for an *M*-ary PPM channel for which the slot time is kept constant and the maximum *M* is constrained to 256. Note that in the limit of high signal power, the channel capacity approaches 0.53 bits/slot.

Appendix A

Received Background Power Calculation

The intensity pattern at the receiver focal plane can be related to the incoming signal amplitude and the wavefront quality of the primary aperture. For simplicity, it will be assumed that the incoming signal is a plane wave and that all the distortion due to the optical system can be summarized by the wavefront distortion of the primary aperture, $A(\mathbf{x})$, where $\mathbf{x}$ is the coordinate in the aperture plane. If it is further assumed that the incident light beam has unit intensity and is incident from an angle θ_S from the surface normal, the focal plane intensity pattern can be expressed as

$$I_f(\mathbf{r}) = \left(\frac{1}{\lambda f}\right)^2 \left|\mathcal{F}\left[A(\mathbf{x})\, e^{i\mathbf{k_S}\cdot\mathbf{x}}\right]_{\mathbf{r}/\lambda f}\right|^2 \cos\theta_S \qquad \text{(A-1)}$$

where $\mathbf{r}$ is the coordinate in the receiver focal plane, λ is the wavelength, f is the focal length of the optical system, and $\mathbf{k_S}$ is the incident wave vector. The factor $\cos\theta_S$ in Eq. (A-1) accounts for the reduction in signal intensity when the aperture is illuminated off angle. Equation (A-1) can be simplified by noting that the Fourier transform of a tilted wavefront results in a translation on the receiver focal plane:

$$I_f(\mathbf{r}) = \left(\frac{1}{\lambda f}\right)^2 \left|\mathcal{A}\left[\frac{\mathbf{r}-\mathbf{u_S}}{\lambda f}\right]\right|^2 \cos\theta_S \qquad \text{(A-2)}$$

where $\mathcal{A}(\mathbf{r})$ is the Fourier transform of the aperture distortion function $A(\mathbf{x})$. The translation on the focal plane, $\mathbf{u_S}$, is related to the angle of incidence by

$$|\mathbf{u_S}| = f\cos\theta_S \qquad \text{(A-3)}$$

The amount of power received by the detector is simply the integral of Eq. (A-1) over the detector area:

$$P_R = \int_r w_f(r) I_f(r) d^2r \qquad \text{(A-4)}$$

where $I_f(r)$ is the intensity distribution over the detector focal plane, and $w_f(r)$ is the detector aperture function; $w_f(r) = 1$ if $r < r_D$, and $w_f(r) = 0$ if $r > r_D$. The integration of Eq. (A-4) can be equally carried out in angular space. The detector's angular field of view, Ω_D, defines the boundary of integration, and Eq. (A-4) can be equivalently written as

$$P_R = \int_{\Omega_D} \frac{1}{\lambda^2}\left|\mathcal{A}\left[\left(\frac{\theta-\theta_S}{\lambda}\right)\right]\right|^2 \cos\Omega_S d\Omega \qquad \text{(A-5)}$$

The receiver aperture, in principle, can be used for transmitting a signal. The far-field amplitude can again be related to the wavefront at the aperture using a Fourier transform relationship. The antenna gain $G(\Omega)$ is defined as the intensity at far field versus that of an isotropic radiator. That is,

$$G(\Omega) = \frac{4\pi z^2}{P_{in}} I(\Omega) \qquad \text{(A-6)}$$

where P_{in} is the input optical power, and z is the propagation distance. The far field intensity pattern resulting from an aperture pattern of $A(\mathbf{x})$ can be given by

$$I(\theta) = \left(\frac{1}{\lambda z}\right)^2 \left|\mathcal{A}\left[\frac{\theta}{\lambda}\right]\right|^2 \qquad \text{(A-7)}$$

By substituting the far field intensity pattern into the expression for antenna gain, it is seen that the antenna gain can be related to the Fourier transform of the aperture by

$$G(\theta) = \frac{16}{d_R^2\lambda^2}\left|\mathcal{A}\left[\frac{\theta}{\lambda}\right]\right|^2 \qquad \text{(A-8)}$$

Given the relationship between antenna gain and signal intensity, the total amount of background power received can then be calculated. The amount of power radiated by a patch of sky with solid angle $d\Omega_S$ can be characterized by Planck's radiation law. The intensity of light as seen on the receiver aperture can be written as

$$I_{d\Omega_S} = \left(\frac{2hf}{\lambda^2}\right)\frac{1}{e^{hf/k_BT}-1}\eta B\Delta f d\Omega_S \quad \left(\mathrm{W/cm^2\mu m}\right) \qquad \text{(A-9)}$$

where the factor of 2 came from the two orthogonal polarization modes, and the parameter η is the emissivity of

the blackbody. By combining Eqs. (A-1) through (A-9), the total power collected on the receiver from a patch of background source that subtends the solid angle $d\Omega_S$ is given by

$$P_R(\theta_S)d\Omega_S = \int_{\Omega_D} \frac{1}{\lambda^4}\left|\mathcal{A}\left[\frac{\mathbf{r}-\mathbf{u_S}}{\lambda f}\right]\right|^2 \times \cos\theta_S \frac{2\eta h f B}{e^{hf/k_BT}-1} B\Delta f d\Omega_S d\Omega$$

$$= \int_{\Omega_D}\left(\frac{d_R^2}{16\lambda^2}\right) G(\theta-\theta_S) \times \cos\theta_S \frac{2\eta h f B}{e^{hf/k_BT(\Omega_S)}-1}\Delta f d\Omega d\Omega_S \qquad \text{(A-10)}$$

Integration of Eq. (A-10) over the entire sky then gives the expression of total background power collected.

$$P_N = \int\int_{\Omega_D}\left(\frac{d_R^2}{16\lambda^2}\right) G(\theta-\theta_S) \times \cos\theta_S \frac{2\eta h f B}{e^{hf/k_BT}-1} d\Omega d\Omega_S \qquad \text{(A-11)}$$

where the integration is performed first over the detector field of view, Ω_D, then over the extent of the background source distribution. For a nondiffraction-limited receiver, the integral given by Eq. (A-11) can be substantially larger than the power received by a single mode receiver. Consequently, a nondiffraction-limited receiver can admit much more noise than a diffraction-limited one.

A simple figure of merit of the optical receiver is the number of background noise modes it collects. This spatial mode mismatch factor, F_B, is given by

$$F_B = \int\int_{\Omega_D}\left(\frac{d_R^2}{16\lambda^2}\right) G(\theta-\theta_S)\cos\theta_S d\Omega d\Omega_S \qquad \text{(A-12)}$$

The number of background modes, F_B, varies from 1 for a diffraction-limited receiver to greater than 2×10^6 for the proposed 10-m photon-bucket receiver.

Appendix B

Capacity of Optical PPM Channel

The capacity per channel use for a general M-ary channel is defined as

$$C = I(\mathbf{x};\mathbf{y}) = H(\mathbf{y}) - H(\mathbf{y}|\mathbf{x}) \qquad \text{(B-1)}$$

where $\mathbf{x}$ and $\mathbf{y}$ represent the transmit and receive symbol sets, respectively, $I(\mathbf{x};\mathbf{y})$ is the mutual information function, $H(\mathbf{y})$ is the entropy of $\mathbf{y}$, and $H(\mathbf{y}|\mathbf{x})$ is the conditional entropy function. The capacity per transmission is given as the entropy of the received symbol set minus the amount of conditional entropy between the transmission and reception signal set. If the transmission is perfect, the conditional entropy between $\mathbf{x}$ and $\mathbf{y}$ is zero, and the capacity per transmission is simply equal to the entropy of the signal set.

For PPM systems, the demodulator can be implemented as follows: The receiver simply counts the number of photons that are received during the M time slots, and then chooses the slot with the largest photocount. If there is more than one slot having the largest count, the demodulator can either randomly choose between these slots, or it can assign a special erasure symbol to the output. In the first case, the PPM channel can be modeled as an M-ary symmetric channel with probability of correct reception p, and probability of erroneous decision q, for each of the remaining $(M-1)$ symbols. The probabilities p and q are given by

$$p = \frac{1}{M}e^{-K_S+MK_B} + \sum_{k=1}^{\infty}\frac{(K_S+K_B)^k}{k!}e^{-(K_S+K_B)} \times \left[\sum_{m=0}^{k-1}\frac{K_B^m}{m!}e^{-K_B}\right]^{M-1}\frac{1}{aM}\left[(1+a)^M - 1\right]$$

$$a \equiv \frac{K_B^k}{k!}\Big/\sum_{m=0}^{k-1}\frac{K_B^m}{m!} \qquad \text{(B-2a)}$$

$$q = (1-p)/(M-1) \qquad \text{(B-2b)}$$

where $K_S = \lambda_S \Delta T$ and $K_B = \lambda_B \Delta T$ are the average signal and background counts received during one time slot. By assuming that the M transmission symbols have equal probability, the capacity per word (channel use) for this M-ary symmetric channel is given by

$$C_{\text{MSC}} = \log M + p\log p + (M-1)q\log q \qquad \text{(B-3)}$$

Similarly, the PPM channel can also be modeled as an M-ary erasure channel with probability of correct transmission α, and probability of erasure γ. The probability of erroneous decoding to the remaining $(M-1)$ symbol is given by $\beta = (1-\alpha-\gamma)/(M-1)$. The probabilities can be given by

$$\alpha = \sum_{k=1}^{\infty}\frac{(K_S+K_B)^k}{k!}e^{-(K_S+K_B)} \times \left[\sum_{m=0}^{k-1}\left(\frac{K_B^m}{m!}e^{-K_B}\right)\right]^{M-1} \qquad \text{(B-4a)}$$

$$\beta = \sum_{k=1}^{\infty}\frac{K_B^k}{k!}e^{-K_B}\left[\sum_{m=0}^{k-1}\frac{K_B^m}{m!}e^{-K_B}\right]^{M-2} \times \left[\sum_{m=0}^{k-1}\frac{(K_S+K_B)^m}{m!}e^{-(K_S+K_B)}\right] \qquad \text{(B-4b)}$$

$$\gamma = 1-\alpha-(M-1)\beta \qquad \text{(B-4c)}$$

Again, if it is assumed that the transmission symbols have equal probabilities, the capacity per channel use is given by

$$C_{\text{MEC}} = (1-\gamma)\log\frac{M}{1-\gamma} + \alpha\log\alpha + (M-1)\beta\log\beta \qquad \text{(B-5)}$$

In the limit where the background noise is zero, the probability of error β goes to zero, and the capacity (per word) of the PPM channel reduces to the familiar form for the M-ary erasure channel studied by Pierce:

$$C_{\mathrm{MEC}} = \alpha \log M = (1 - e^{Ks}) \log M \qquad \text{(B-6)}$$

The capacity per channel use is slightly different for the two demodulation schemes. For this calculation, the M-ary PPM channel is modeled as an erasure channel.

Given the capacity per channel use and the expected signal count, the photon efficiency, C_{ph}, is given by

$$C_{\mathrm{ph}} = C_{\mathrm{MEC}}/K_S \qquad \text{(B-7)}$$

The capacity per second, R_{max}, is given by

$$R_{\mathrm{max}} = \frac{C_{\mathrm{MEC}}}{M\Delta T} \qquad \text{(B-8)}$$

and the capacity per slot, $C_{\Delta T}$, is simply

$$C_{\Delta T} = \frac{C_{\mathrm{MEC}}}{M} \qquad \text{(B-9)}$$

Optimizing the *G/T* Ratio of the DSS-13 34-Meter Beam-Waveguide Antenna

M. S. Esquivel
Ground Antennas and Facilities Engineering Section

Calculations using Physical Optics computer software were done to optimize the gain-to-noise-temperature (G/T) ratio of DSS 13, the DSN's 34-m beam-waveguide antenna, at X-band for operation with the ultra-low-noise amplifier maser system. A better G/T value was obtained by using a 24.2-dB far-field-gain smooth-wall dual-mode horn than by using the standard X-band 22.5-dB-gain corrugated horn.

I. Introduction

During Phase-I testing of the DSS-13 beam-waveguide (BWG) antenna, it was discovered that there was higher system noise temperature than expected. The high noise temperature was caused by the spillover losses of the BWG mirrors having a bigger effect than previously thought. It was experimentally determined that higher-gain horns would improve the *G/T* ratio of the antenna because of two reasons: (1) there would be a lower spillover loss in the BWG mirrors and, hence, less noise temperature, and (2) it was hoped that by using higher-gain horns the antenna gain would not decrease significantly.

The Physical Optics (PO) software was used to analyze the theoretical performance of the DSS-13 BWG antenna using three different horn patterns as inputs. The three patterns used were those of corrugated horns with far-field gains of 22.5 dB, 24.2 dB, and 26.1 dB. Each horn pattern was placed at different positions along the z-axis from F_3, the folded focal point of the basement ellipsoid (see Fig. 1). From the PO analysis, the spillover of the BWG mirrors and the gain of the DSS-13 BWG antenna could be obtained and, in this manner, various *G/T* values could be calculated. All of these calculations were done at 8.45 GHz.

This article presents the results of these calculations. These results were the basis on which a horn was designed to be used in the DSS-13 BWG antenna along with the ultra-low-noise amplifier (ULNA) X-band package. The final design chosen was that of a smooth-wall dual-mode (Potter) horn with a far-field gain of 24.2 dB.

II. Theoretical Calculations

A. Method

The theoretical horn patterns mentioned previously were calculated by using software that analyzes a circular waveguide device and computes the TE_{mn} and TM_{mn} modes at its ports.[1] The TE_{mn} and TM_{mn} modes at the

[1] D. Hoppe, *Scattering Matrix Program for Circular Waveguide Junctions* (internal document), Jet Propulsion Laboratory, Pasadena, California, March 9, 1987.

aperture of a horn are then used to compute the equivalent current distribution. The far-field pattern was calculated using the radiation integral formula. In this case, the standard JPL X-band corrugated horn with a 22.5-dB gain was modeled and its theoretical far-field pattern was computed. The pattern was then converted into a set of spherical wave expansion (SWE) coefficients using software created at JPL by Art Ludwig in the late 1960s [1]. The SWE coefficients were then used as the input to the PO software to obtain the gain of the antenna and the spillover of the BWG mirrors. The method used here was similar to the one used by the designers of the DSS-13 BWG antenna [2].

Besides the 22.5-dB-gain horn, two other corrugated horns (with far-field gains of 24.2 dB and 26.1 dB) were also modeled. These horns were created by simply adding extensions with the same flare angle to the existing 22.5-dB horn circular waveguide configuration geometry. Each of these horns then went through the same process described in the preceding paragraph.

To find the best focus position, each horn aperture was displaced by increments of 2λ ($\lambda = 3.55$ cm) from focal point F_3 in a direction that would bring it closer to mirror M_6 (along the +z-axis of Fig. 1). By doing this, the spillover was found to decrease and the directivity of the DSS-13 BWG antenna was found to increase slightly. This result in itself was surprising; the phase center of the feed as located in this system was not where expected from previous Phase-I design considerations.

The following formula, Eq. (1), was used to compute the total noise temperature from all the noise contributors:

$$N_t = \sum_{i=1}^{M} (T_{fact})_i (P_s)_i \qquad (1)$$

where N_t is the total noise temperature (K), $(T_{fact})_i$ is the temperature factor of the ith noise contributor (K), $(P_s)_i$ is the power factor of the ith spillover contributor, and M is the number of contributors.

For the DSS-13 BWG antenna, four noise contributors ($M = 4$) are identifiable as follows:

(1) BWG mirrors (see Table 4)[2]:

(a) Two lower BWG mirrors with $(T_{fact})_1$ equal to 300 K.

(b) Four upper BWG mirrors with $(T_{fact})_2$ equal to 270 K.

(2) The subreflector with $(T_{fact})_3$ equal to 5 K.[3]

(3) The 34-m main reflector with $(T_{fact})_4$ equal to 240 K.

Note that the calculations done in this article are applicable to the antenna pointing at the zenith.

The quantity $(P_s)_i$ in Eq. (1) is directly related to the spillover of each noise contributor by the formula in Eq. (2):

$$(P_s)_i = 1 - 10^{(S_p)_i/10} \qquad (2)$$

where $(S_p)_i$ is the spillover in decibels of each noise contributor computed by the PO software.

B. Results

After the above calculations were done, the results were plotted in Figs. 2(a)–2(c). Fig. 2(a) shows the efficiency of the DSS-13 BWG antenna for various horns as a function of the horn aperture position measured from F_3 (see Fig. 1). The efficiency plotted in Fig. 2(a) includes the BWG spillover losses and the other losses associated with the antenna as listed in Table 1. For example, the PO software predicts a gain of 68.933 dB for the 22.5-dB corrugated horn with its aperture located 2λ from F_3. This gain corresponds to an efficiency of 86.36-percent at $f = 8.45$ GHz (69.57 dB is equivalent to 100 percent). The 86.36-percent efficiency would then be placed in the "PO calculation" slot of Table 1 and multiplied by the subtotal efficiency (84.6 percent) to give a total efficiency of 73 percent. This result agrees well with the measured value of 72.4 percent reported.[4,5]

Figure 2(b) shows the total noise temperature of the DSS-13 BWG antenna due to the ULNA and the feeds as a function of horn aperture position with respect to

[2] D. A. Bathker, W. Veruttipong, T. Y. Otoshi, and P. W. Cramer, Jr., "Beam-Waveguide Antenna Performance Predictions With Comparisons to Experimental Results," *IEEE Transactions on Microwave Theory and Techniques, Special Issue: Microwaves in Space*, vol. 40, no. 6, edited by R. Dickinson, to be published in June 1992.

[3] D. Bathker and W. Veruttipong, *Excess Noise in BWG Antennas: Results of DSS-13 Shroud Testing and Revised DSS-13 Performance Predictions*, presentation viewgraphs, Jet Propulsion Laboratory, Pasadena, California, July 1991.

[4] D. A. Bathker, W. Veruttipong, T. Y. Otoshi, and P. W. Cramer, Jr., op. cit., p. 9.

[5] M. J. Britcliffe, ed., *DSS-13 Beam Waveguide Antenna Project: Phase I Final Report*, JPL D-8451 (internal document), Jet Propulsion Laboratory, Pasadena, California, p. 5–12, March 1991.

F_3. The total noise temperature was obtained by adding the baseline temperature of 11.054 K (see Table 2) of the ULNA to the contributions of the six BWG mirrors, the subreflector and the 34-m main reflector. As expected, the noise temperature decreases as higher-gain horns are used because they cause less spillover in the BWG mirrors than the lower-gain horns.

For example, for the 22.5-dB corrugated horn with its aperture located 2λ (7.1 cm) from F_3, the PO software predicts a spillover noise temperature of 9.41 K. This value was calculated by taking the spillovers computed by the PO software and using Eqs. (1) and (2). The measured BWG noise temperature of this horn is 8.9 K with the aperture placed about 8.4 cm from F_3.[6] As expected, the smaller displacement from F_3 gives the bigger noise temperature. The comparisons of antenna efficiency and BWG noise temperature with measured results give confidence that the method (specifically the values chosen for $(T_{fact})_i$ of the four noise contributors) that is used to do these calculations does work.

Figure 2(c) shows the G/T values of the DSS-13 BWG antenna for the horns as a function of their aperture position with respect to F_3. The G/T ratio is calculated by converting the efficiencies of Fig. 2(a) to decibels and then subtracting the total noise temperature (in decibels).

III. Horn Selection

Since the purpose of this study is to optimize the G/T ratio of the DSS-13 BWG antenna, Fig. 2(c) needs to be examined closely to select the proper horn. It seems that a corrugated horn with far-field gain between 24 and 26 dB is the best choice.

Because of cost considerations, it was decided to use an existing JPL design for a 22-dB smooth-wall dual-mode circular horn design instead of a corrugated feed. The smooth-wall dual-mode circular feed is usually referred to as the Potter horn, since P. D. Potter was the original designer [3]. The Potter horn has a smooth-wall flare section 54.193 cm in length with a horn-flare half angle of 6 deg 15 arcmin 15 arcsec. The radius at its aperture is 8.2487 cm.[7] The plan was to add an extension on top of the existing ULNA Potter horn that would give a far-field gain between 24 and 26 dB.

Since the extension was to be built at JPL and there were cost limitations, it was decided to modify the existing horn to have a far-field gain of 24.2 dB instead of 26 dB. To accomplish this, a 25.4-cm extension (with the same flare angle, input radius = 8.2487 cm and output radius = 11.033 cm) needed to be used. The original Potter horn has a phasing section that is 0.721 cm in length. The purpose of this phasing section is to fix the differential phase between the TE_{11} and TM_{11} modes at the aperture of the horn so that an equal beam width pattern is created.

Since a 25.4-cm extension has been added, the TE_{11} and TM_{11} will not be properly phased at the aperture. Making use of Eqs. (11) and (12) in [3], the new phasing section needs to be adjusted to a length of 6.48 cm. Figure 3 shows the geometry of this new 24.2-dB Potter horn.

To appreciate the importance of the phasing section, compare Figs. 4(a) and 4(b), which show the far-field patterns for a Potter horn with a 25.4-cm extension and phasing section lengths of 6.48 cm and 0.721 cm, respectively. Notice how the 6.48-cm phasing section gives equal beam widths out to $\theta = 15$ deg and a 24.2-dB gain while the 0.721-cm phasing section gives a smaller gain (23.8 dB) with unequal beam widths.

The new 24.2-dB Potter horn was then run through the PO software as the other horns were. The results are shown in Figs. 2(a)–2(c), under 24.2-dB Potter horn. Only four cases were run for this horn because results for the corrugated horns predicted that the optimum G/T ratio would be achieved when the horn aperture-to-F_3 displacement was about 10λ. Notice that the predicted G/T ratio for the 24.2-dB Potter horn is better than for all the horns. This is due to the fact that while the antenna efficiency is at the same level as for the 24.2-dB corrugated horn (see Fig. 2(a)), the noise temperature is about 0.7 K lower (see Fig. 2(b)). The noise temperature is less because the PO software predicts that the 24.2-dB Potter horn has nearly the same spillover as the 26.1-dB corrugated horn.

IV. Experiment

The new 24.2-dB Potter horn was built. Radiation patterns were measured at the JPL mesa antenna-range facility. Since these patterns were recorded on a strip chart, they were visually compared to the theoretical patterns of Fig. 4(a) and were found to be satisfactory. Then, in November 1991, the system noise temperature of the DSS-13 BWG antenna was measured using the 24.2-dB Potter horn with the ULNA. The horn's aperture was placed 8.61 λ (f = 8.45 GHz) from F_3. A value of N_t = 14.4 K

[6] D. Bathker and W. Veruttipong, op. cit.

[7] JPL Assembly Drawing D23835-13696 (internal document), Jet Propulsion Laboratory, Pasadena, California, 1969.

was observed.[8] This measured point is shown in Fig. 2(b). The predicted noise temperature for the 24.2-dB Potter horn at 8.61 λ is 12.78 K. At the present time, reasons for the small differences between the measured and calculated values are being investigated. Further work in this area is warranted.

V. Conclusions

A method has been developed that can be used to calculate the optimum-gain horn to be used in the DSS-13 34-m BWG antenna. The ability to predict the total system efficiency of the BWG system using this method was found to be highly accurate and agreed with measurements; however, work needs to be done to improve the noise-temperature prediction model. It was shown that the location of the position of the feed horn had to be modified such that the typical far-field phase center position is not used; rather, a near-field phase center located deeper inside the horn optimizes both the antenna noise temperature and the G/T ratio. This design was used to fabricate a Potter horn for the X-band ultra-low-noise amplifier maser system that was recently demonstrated at the DSS-13 34-m BWG antenna. The 1.62 K difference in measured and calculated noise temperatures of this demonstration warrants further study.

[8] H. K. Detweiler, JPL Internal Memorandum 3330-92-002 (internal document), Jet Propulsion Laboratory, Pasadena, California, January 7, 1992.

Acknowledgments

The author would like to thank the JPL Supercomputing Project for facilitating the use of Voyager, the new CRAY Y-MP2E/116 supercomputer. It would have been impossible to have completed this G/T ratio optimization study without the numerous CPU hours burned up on this powerful machine.

The noise-temperature models used in this article have been a direct result of the efforts of all the individuals from the JPL Ground Antennas and Facilities Engineering Section who have been a part of the DSS-13 BWG Antenna Project.

Finally, the author acknowledges Mark Gatti for many insightful discussions and for suggesting that this study be undertaken.

References

[1] A. C. Ludwig, "Near-Field Far-Field Transformations Using Spherical-Wave Expansions," *IEEE Transactions on Antennas and Propagation*, vol. AP-19, no. 2, pp. 214–220, March 1971.

[2] T. Veruttipong, W. Imbriale, and D. Bathker, "Design and Performance Analysis of the DSS-13 Beam Waveguide Antenna," *TDA Progress Report 42-101*, vol. January–March 1990, Jet Propulsion Laboratory, Pasadena, California, pp. 99–113, May 15, 1990.

[3] P. D. Potter, "A New Horn Antenna With Suppressed Sidelobes and Equal Beamwidths," *Microwave Journal*, vol. 6, pp. 71–78, June 1963.

Table 1. DSS-13 BWG antenna: overall gain accounting X-band at F_3.[a]

Element	Efficiency, percent	Notes
Main reflector:		
I^2R	0.99954	
Panel leakage	0.99992	
Gap leakage	0.9982	
RMS	0.977	Normal 0.45 mm (0.0177 in.)
Subreflector:		
I^2R	0.99954	
RMS	0.998	Normal 0.125 mm (0.005 in.)
Four BWG mirrors:		
I^2R	0.99807	
RMS	0.996	Each normal 0.125 mm (0.005 in.)
Two BWG mirrors:		
I^2R	0.99903	
RMS	0.997	Each normal 0.125 mm (0.005 in.)
BWG/Cassegrain VSWR	0.999	Estimated
Waveguide:		
I^2R	0.984	Loss = −0.070 dB
VSWR	0.99	
Feed support blockage	0.918	X-band, tripod, 2.7%, 1.55 factor
Bypass blockage	0.987	9-ft equivalent diameter
Pointing squint	0.9954	Scan loss = −0.02 dB
Mirror alignments	0.9994	Based on 0.99 at 32 GHz
Subtotal:	0.844491	
PO calculation	Variable	Includes BWG mirrors' spillover losses. Dependent on horn gain and position
Total efficiency	Variable	69.573 dB is 100% at 8.45 GHz. Dependent on horn gain and position

[a] This table is based on D. A. Bathker, W. Verruttipong, T. Y. Otoshi, and P. W. Cramer, Jr., op. cit., p. 21.

Table 2. DSS-13 BWG antenna: overall noise accounting X-band at F_3 at 90-deg elevation.[a]

Element	Noise, K	Notes
Cosmic background	2.5	Equivalent blackbody
Atmosphere	2.17	Goldstone (average clear)[b]
Four mirror I^2R/BWG	0.572	
Two mirror I^2R/BWG	0.272	
Main reflector:		
I^2R	0.136	
Panel leakage	0.02	
Gap leakage	0.5	
Subreflector:		
I^2R	0.134	
Tripod scatter	2	
Bypass scatter	0.07	
Subtotal	8.374	
ULNA	2.68	Based on roof measurements of ULNA and the new horn[c]
Baseline (ULNA plus subtotal)	11.054	
DSS-13 spillover:		
Two lower BWG mirrors	Variable	T_{fact} = 300 K
Four upper BWG mirrors	Variable	T_{fact} = 270 K
Subreflector	Variable	T_{fact} = 5 K
Main reflector	Variable	T_{fact} = 240 K
Total noise	Variable	Depends on horn gain and aperture position

[a] This table is based on D. A. Bathker, W. Verruttipong, T. Y. Otoshi, and P. W. Cramer, Jr., op. cit., p. 23.

[b] S. D. Slobin, "Atmospheric and Environmental Effects," Module TCI-40, Rev. C, *DSN/Flight Project Interface Design*, Document 810-5, Rev. D (internal document), Jet Propulsion Laboratory, Pasadena, California, 1991.

[c] M. Gatti, Manager, DSS-13 Beam-Waveguide Antenna Project, personal communication, Jet Propulsion Laboratory, Pasadena, California, January 1992.

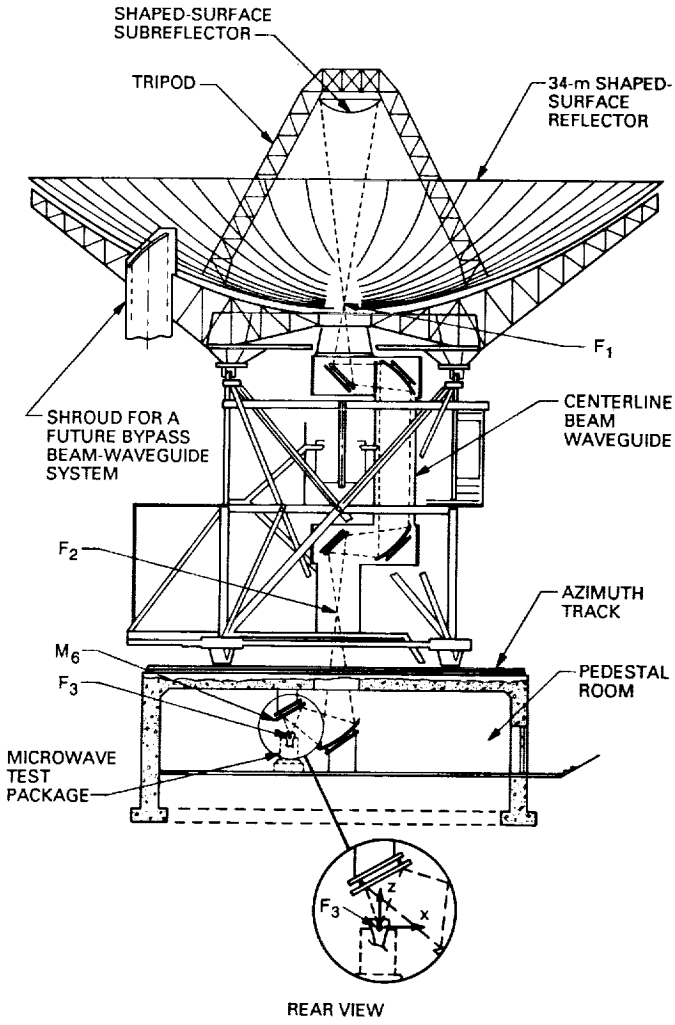


Fig. 1. Configuration of DSS-13 antenna.

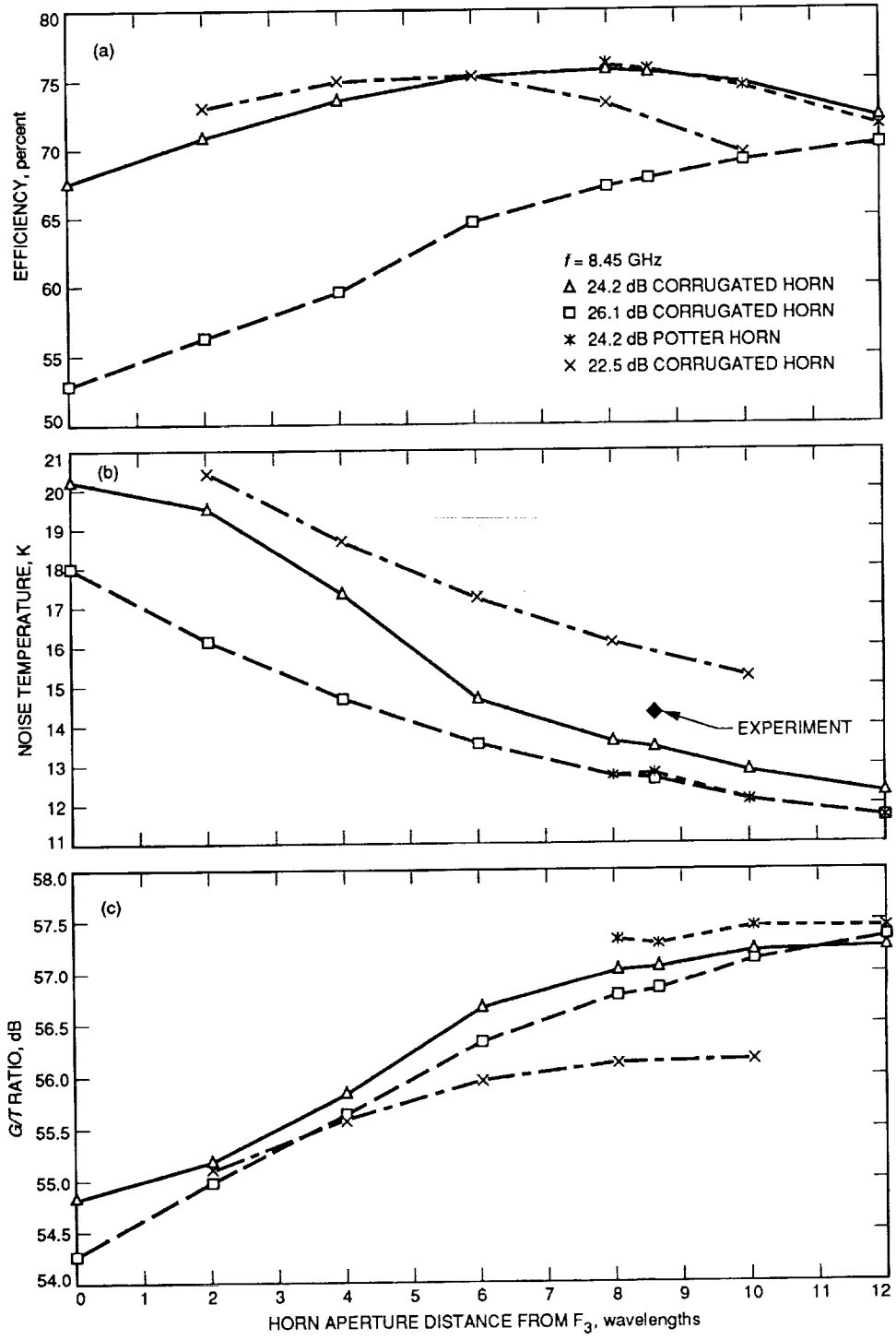


Fig. 2. DSS 13: (a) efficiency versus horn aperture distance from F_3; (b) system temperature versus horn aperture distance from F_3; and (c) *G/T* ratio versus horn aperture distance from F_3.

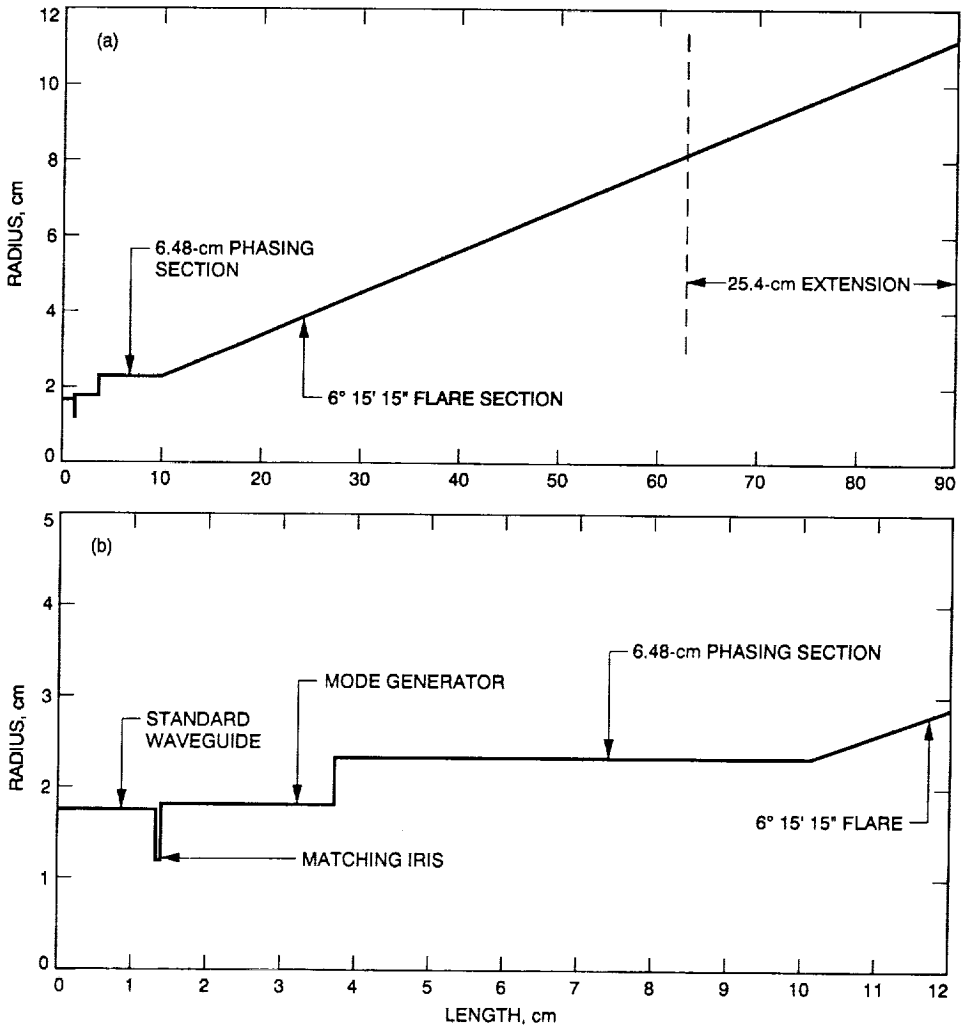


Fig. 3. X-band 24.2-dB Potter horn dimensions: (a) overall feed and (b) input section.

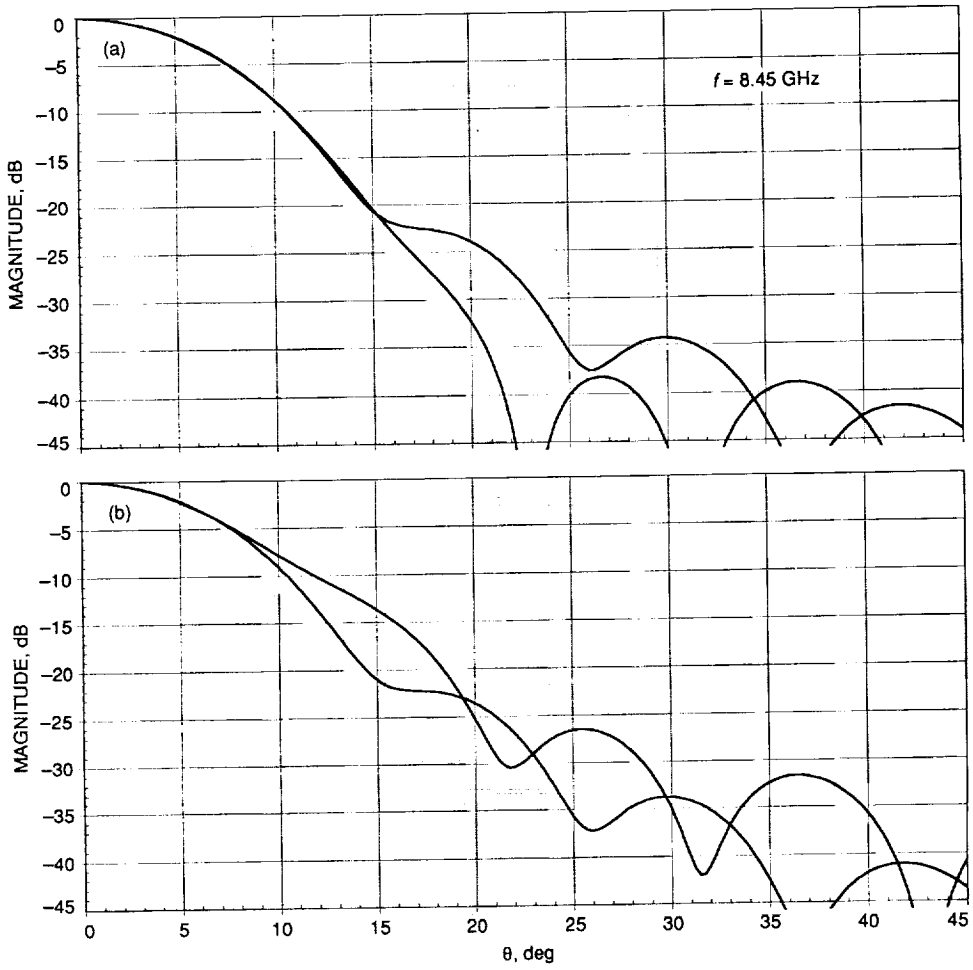


Fig. 4. X-band radiation patterns of the Potter horn with: (a) 24.2-dB gain: 6.48-cm phasing section and (b) 23.8-dB gain: 0.721-cm phasing section.

Dynamic Modeling of Fluid Transmission Lines of the DSN 70-Meter Antennas by Using a Lumped Parameter Model

R. D. Bartos
Ground Antennas and Facilities Engineering Section

Fluid transmission lines and fittings have been found to significantly affect the dynamic and steady state operation of the engineering system in which they are incorporated. Therefore, a better understanding of the operation of a system can be obtained by including the performance of the transmission lines and fittings within the system model. The most accurate model of a transmission line or fitting is obtained by using a distributed parameter model. However, a distributed parameter model tends to be very complex. This complexity can be avoided without significant loss of model accuracy by using a lumped parameter model when the length of the fluid path through the transmission line or fitting is short. This article develops a lumped parameter model for short fluid transmission lines and fittings, describes the conditions under which the model is valid, and presents the model parameters associated with the servo hydraulic system of the DSN 70-meter antennas.

I. Introduction

All fluid transmission lines and fittings are physically distributed systems because the resistance to fluid flow resulting from the fluid viscosity, inertial, and compressibility effects are spread out over the length of the component. Modeling fluid lines and fittings with distributed parameters generally results in a very complex model. Under certain conditions in which the line or fitting is considered to be short, a lumped parameter model may be used so that the mathematical equations describing the physical behavior are greatly simplified. A lumped parameter model approximates the behavior of the actual line or fitting without considering the distribution of the parameter over the length of the component. This article develops a theoretical model for fluid transmission lines and fittings from a lumped parameter point of view by using the methodology recommended by Meritt [1] and Watton [2]. A description of the conditions under which the lumped parameter model is valid is also included in the article. The model is subsequently used to mathematically describe the behavior of the control lines found within the servo hydraulic system of the DSN 70-meter antennas.

II. Theoretical Model Development

The lumped parameter model developed in this section provides a means of approximating the behavior of short transmission lines and fittings. The model is based on the equation of state for a fluid:

$$dP = \beta \frac{d\rho}{\rho_0} \tag{1}$$

where P is the fluid pressure, β is the effective bulk modulus, ρ is the density, and ρ_0 is the average density. The effective bulk modulus is given by the relationship

$$\frac{1}{\beta} = \frac{1}{\beta_c} + \frac{1}{\beta_f} \tag{2}$$

where β_f is the bulk modulus of the fluid and β_c is the effective bulk modulus of the container. β_f describes how the fluid density changes with respect to pressure, while β_c describes how the fluid container expands with respect to pressure. The fluid bulk modulus is determined experimentally for a liquid and is given by

$$\beta_f = \gamma \rho_0 R \theta_0 \tag{3}$$

for an ideal gas, where γ is the specific heat ratio, R is the ideal gas constant, and θ_0 is the average gas temperature. When the container is a circular conduit, the effective bulk modulus of the container, β_c, is given by the equation [2,3]

$$\frac{1}{\beta_c} = \frac{2}{E\delta}\left[\frac{(1+\nu)(a+\delta)^2 + (1-\nu)a^2}{(2a+\delta)}\right] \tag{4}$$

where a is the internal radius of the conduit, E is the modulus of elasticity, δ is the conduit wall thickness, and ν is Poisson's ratio. Note that the effective bulk modulus will change for different types of fluid containers. When several containers are combined together, as shown in Fig. 1, the system can be represented as a single container by using the equation

$$\frac{V_0}{\beta} = \frac{V_1}{\beta_1} + \frac{V_2}{\beta_2} + \frac{V_3}{\beta_3} + \cdots + \frac{V_i}{\beta_i} \tag{5}$$

where V_0 is the total volume, β is the effective bulk modulus, and β_i and V_i are the effective bulk modulus and volume of the ith container, respectively.

The development of a dynamic model for a short line or fitting using the lumped parameter representation involves expressing Eq. (1) as

$$\frac{\partial P}{\partial t} = \frac{\beta}{\rho_0} \times \frac{\partial \rho}{\partial t} \tag{6}$$

The continuity relationship for a control volume is given by

$$\frac{\partial \rho}{\partial t} = \frac{\rho_0}{V_0}\left[\sum Q_{in} - \sum Q_{out} - \frac{\partial V_0}{\partial t}\right] \tag{7}$$

where V_0 is the total volume, Q_{in} is the flow into the control volume, Q_{out} is flow out of the control volume, and $\partial V_0/\partial t$ is the boundary deformation of the control volume. If the volumetric expansion of the fluid container with respect to pressure is expressed in terms of β_c, then $\partial V_0/\partial t$ is zero. Substituting Eq. (7) into Eq. (6) and setting $\partial V_0/\partial t$ to zero yields

$$\frac{\partial P}{\partial t} = \frac{\beta}{V_0}\left[\sum Q_{in} - \sum Q_{out}\right] \tag{8}$$

The pressure drop across a fitting or pipe where the lumped parameter model is valid can be ignored when the pressure drop across the fluid line is negligible with respect to the system operating pressures. This article considers only fluid lines and fittings with negligible pressure drops. A detailed presentation on modeling pressure drops across fluid conduits can be found in [3].

The conditions under which the lumped parameter model is valid are formulated by considering the speed, c_0, at which pressure waves traverse the fluid component given by the equation

$$c_0 = \sqrt{\frac{\beta}{\rho_0}} \tag{9}$$

Hence, the time it takes a pressure wave to travel down and back along a line or fitting of length L is

$$\tau = \frac{2L}{c_0} \tag{10}$$

Therefore, a lumped parameter model is only valid when the relationship

$$\frac{2L}{c_0} < \frac{1}{f} \tag{11}$$

holds where f is the highest frequency of system oscillation that exists [2].

III. Model Parameters

The parameters of the model of the hydraulic control lines were determined for each control line of the 70-meter antenna servo hydraulic system based on the component layout of the 70-meter antenna servo hydraulic system shown in Figs. 2 and 3. The model parameters are presented in Table 1 and were computed by using Eqs. (2), (4), (5), and (8), along with the physical properties shown in Table 2. The maximum pressure drop between the servovalve and a hydraulic motor due to fluid viscosity is expected to be less than 69,000 Pa (10 PSI) under both tracking and slewing conditions. Hence, the pressure drop due to fluid viscosity is not modeled since it is not considered to be significant. Fluid accelerations in the lines also cause a pressure drop across the lines. However, the pressure drop across a line due to fluid acceleration is believed to have a negligible effect on system performance since the antenna accelerations are both small in magnitude and short in duration under tracking conditions. Therefore, the fluid pressure in each of the control lines is described by

$$\frac{\partial P}{\partial t} = K\left[\sum Q_{in} - \sum Q_{out}\right] \tag{12}$$

where

$$K = 4.756 \times 10^{10}\text{N/m}^5 \left(113.0\ \text{lbf/in.}^5\right)$$
[Azimuth control no. 1] (13)

$$K = 4.655 \times 10^{10}\text{N/m}^5 \left(110.6\ \text{lbf/in.}^5\right)$$
[Azimuth control no. 2] (14)

$$K = 5.220 \times 10^{10}\text{N/m}^5 \left(124.1\ \text{lbf/in.}^5\right)$$
[Elevation control no. 1] (15)

$$K = 5.216 \times 10^{10}\text{N/m}^5 \left(124.0\ \text{lbf/in.}^5\right)$$
[Elevation control no. 2] (16)

The summations of the inlet and outlet flow rates into a control line found in Eq. (12) must be specified as boundary conditions by using the servovalve and motor models.

The validity of the model presented in Eq. (12) is determined by using Eqs. (9) and (11). The largest distance between the servovalve and hydraulic motor is approximately 20.83 m (820 in.). From Eq. (9), and Tables 1 and 2, the speed of sound is determined to be approximately 1230 m/sec (48,500 in./sec). Commands are currently sent to the servovalve at a rate of 20 hertz, hence, Eq. (11) becomes

$$\frac{2L}{c_0} = 0.034\ \text{sec} < \frac{1}{f} = 0.05\ \text{sec} \tag{17}$$

which shows that a pressure wave is able to travel from the servovalve to the hydraulic motors and back to the servovalve between the time that two consecutive commands are sent to the servovalve. Therefore, the proposed model is valid.

IV. Conclusion

This article presented a method of approximating the performance of short fluid transmission lines and fittings by using a lumped parameter model and computed the parameters required to model the control lines of the DSN 70-meter antenna servo hydraulic system. It has been found to be more advantageous to use a lumped parameter model than the distributed parameter model because of the mathematical simplicity of the lumped parameter model. This simplicity allows the lumped parameter model to be implemented easily within computer simulation packages.

References

[1] H. E. Merritt, *Hydraulic Control Systems*, New York: John Wiley and Sons, 1967.

[2] J. Watton, *Fluid Power Systems Modeling, Simulation, Analog, and Microcomputer Control*, New York: Prentice Hall, 1989.

[3] F. M. White, *Fluid Mechanics*, 2nd ed., New York: McGraw-Hill, 1986.

Table 1. Transmission line parameters for the control lines of the DSN 70-meter antennas.

Transmission line	Total volume	Effective bulk modulus
Azimuth control no. 1	0.02710 m^3 (1653.6 $in.^3$)	1.2888 x 10^9 Pa (1.8693 x 10^5 psi)
Azimuth control no. 2	0.02770 m^3 (1690.4 $in.^3$)	1.2895 x 10^9 Pa (1.8703 x 10^5 psi)
Elevation control no. 1	0.02455 m^3 (1498.1 $in.^3$)	1.2815 x 10^9 Pa (1.8587 x 10^5 psi)
Elevation control no. 2	0.02457 m^3 (1499.6 $in.^3$)	1.2816 x 10^9 Pa (1.8588 x 10^5 psi)

Table 2. Physical properties.

Fluid bulk modulus	1.379 x 10^9 Pa (2 x 10^5 psi)
Fluid density	850 kg/m^3 (7.95 x 10^{-5} $lbf-sec^2/in.^4$)
Fluid kinematic viscosity	27.7 m^2/sec (1090 $in.^2$/sec)
Inside diameter of 1-in. pipe	2.070 cm (0.815 in.)
Outside diameter of 1-in. pipe	3.340 cm (1.315 in.)
Inside diameter of 1.5-in. pipe	3.399 cm (1.338 in.)
Outside diameter of 1.5-in. pipe	4.826 cm (1.900 in.)
Inside diameter of 2-in. pipe	4.285 cm (1.687 in.)
Outside diameter of 2-in. pipe	6.032 cm (2.375 in.)
Modulus of elasticity of stainless steel pipe	2.0684 x 10^{11} Pa (30 x 10^6 psi)
Poisson's ratio for stainless steel	0.25
Volume of 1.5-in. tee	2.147 x 10^{-4} m^3 (13.1 $in.^3$)
Volume of 2-in. tee	4.015 x 10^{-4} m^3 (24.5 $in.^3$)
Volume of 1-in. hose, 66-in. long	8.495 x 10^{-4} m^3 (51.8 $in.^3$)
Bulk modulus of hose	4.32 x 10^9 Pa (6.26 x10^5 psi)

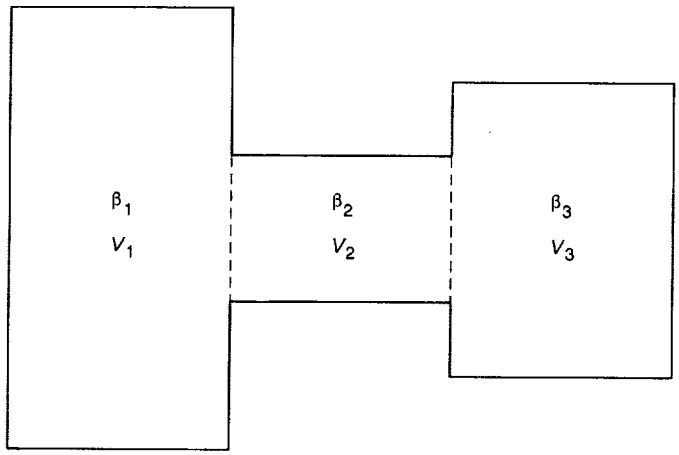


Fig. 1. Combination of fluid containers.

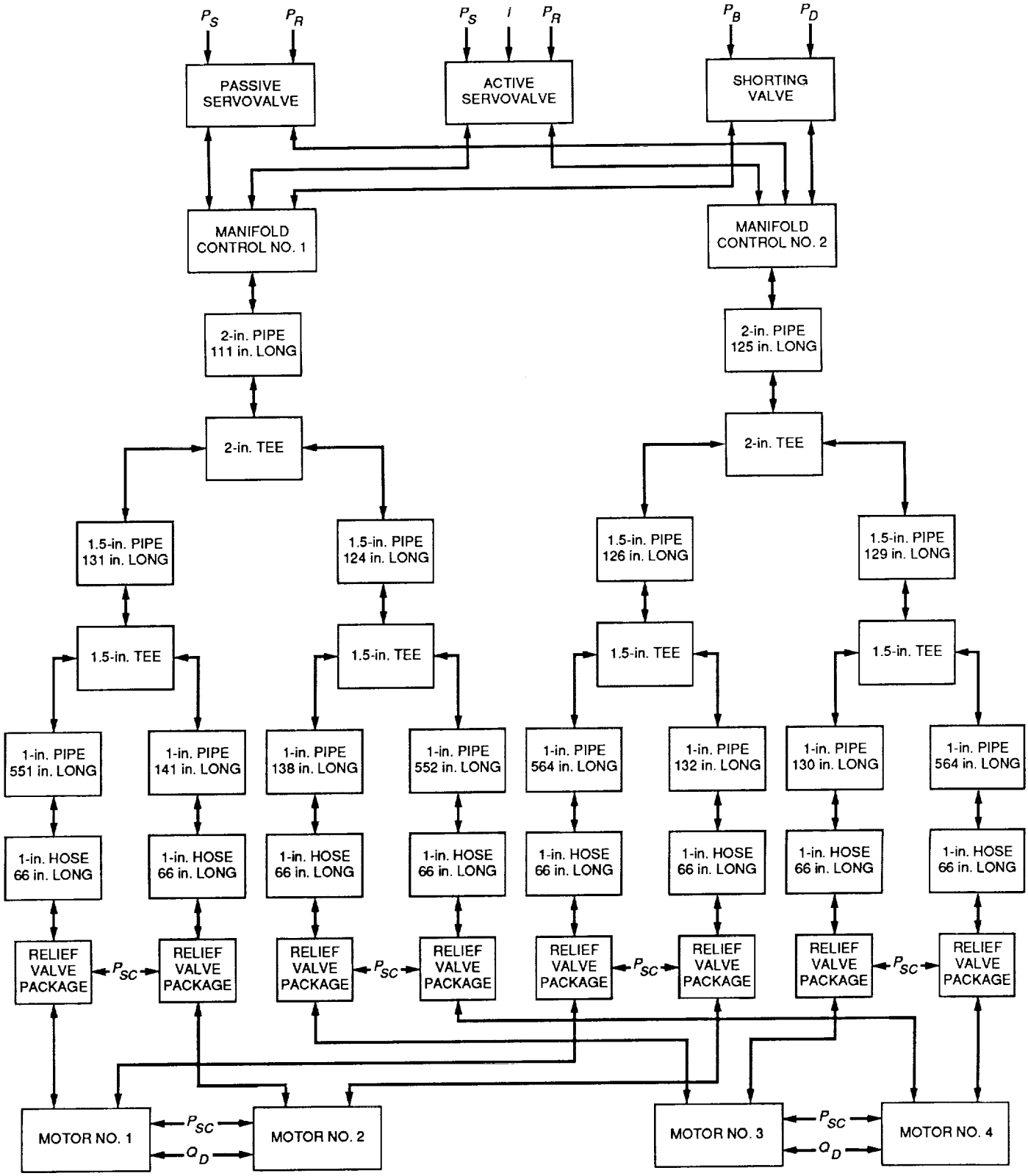


Fig. 2. DSN 70-meter antenna servo hydraulic system azimuth component layout.

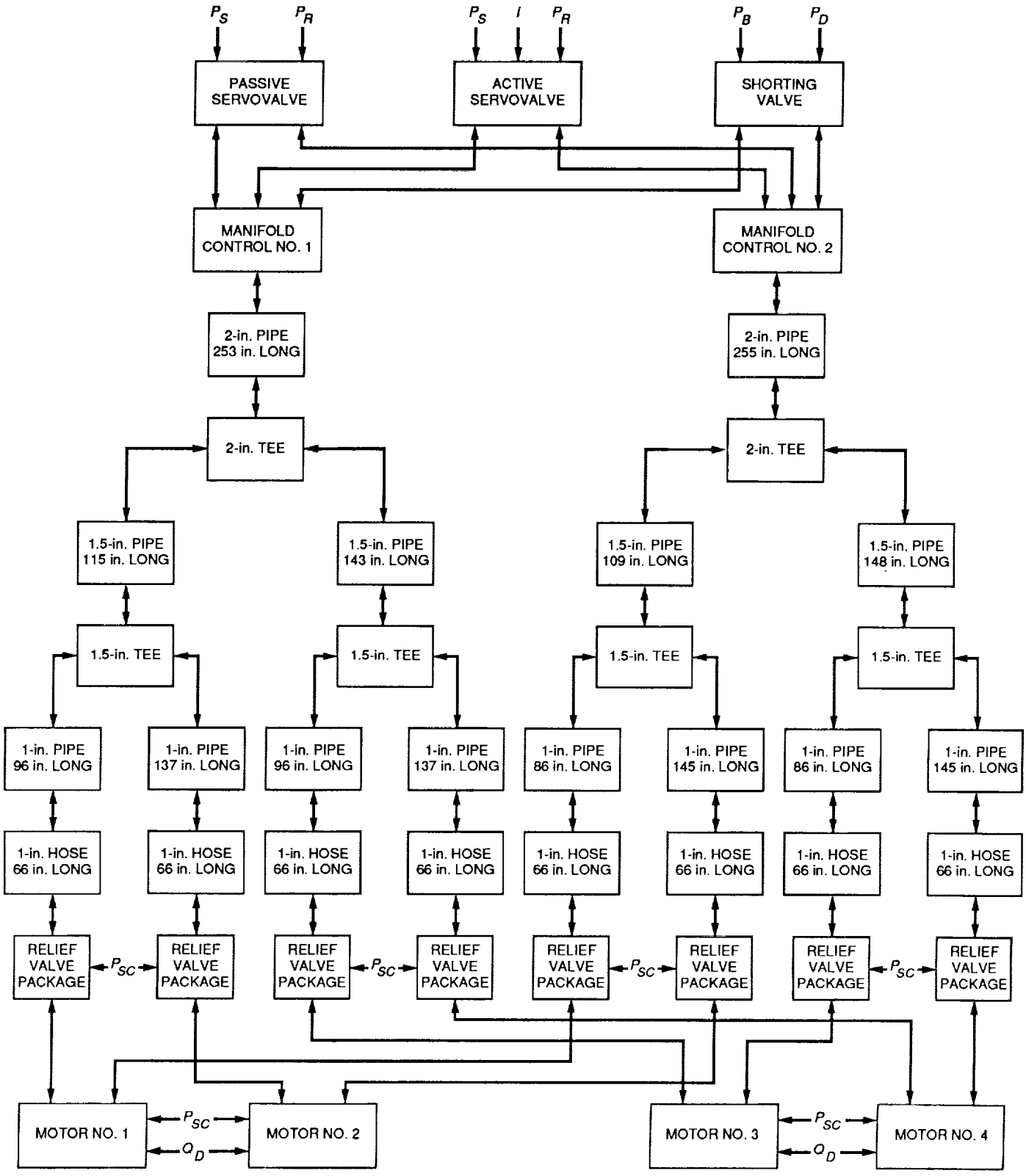


Fig. 3. DSN 70-meter antenna servo hydraulic system elevation component layout.

Appendix

Glossary

- a Internal radius of conduit
- c_0 Average speed of sound in a fluid
- E Modulus of elasticity
- f Highest system frequency, in hertz
- i Index, current
- K Constant associated with a particular fluid line
- L Length of fluid transmission line or fitting
- P Pressure
- P_B Brake pressure
- P_D Drain pressure
- P_R Return pressure
- P_S Supply pressure
- P_{SC} Supercharge pressure
- Q_D Volumetric flow rate into case drain
- Q_{in} Volumetric flow rate into a control volume
- Q_{out} Volumetric flow rate out of a control volume
- R Constant in ideal gas law
- V_i Volume of the ith container
- V_0 Total volume of a control volume
- β Effective bulk modulus
- β_c Effective bulk modulus of the container
- β_f Bulk modulus of the fluid
- β_i Bulk modulus of the ith container
- γ Specific heat ratio
- δ Conduit wall thickness
- θ_0 Average gas temperature
- ν Poisson's ratio
- ρ Fluid density
- ρ_0 Average fluid density
- τ Time delay

Bandwidth Compression of Noisy Signals With Square-Wave Subcarrier

Y. Feria and J. Statman
Communications Systems Research Section

This article discusses a method for downconverting the square-wave subcarrier of spacecraft signals, such as the one from Galileo, which results in a compression bandwidth that lowers the sample rate significantly. The study is focused on three issues. The first is the selection of an adequate down-mixing signal for the resulting signal to have a format similar to that of the original signal, except at a lower subcarrier frequency. The second is the control of the noise level so that the signal to noise ratio is not degraded due to the downconversion. The third is to determine the bandwidth of the downconverted signal considering the uncertainty of the residual carrier frequency.

I. Introduction

A typical downconverted spacecraft signal (e.g., the Galileo signal) has a square-wave subcarrier whose frequency is much higher than the modulating data bandwidth, as shown in Fig. 1 (a). Note that the residual carrier frequency is much smaller than the subcarrier frequency and is not necessarily zero, which results in the dual spectra centered at $-f_r$ and $+f_r$. If the signal is sampled at this point, the sample rate needs to be at least twice the frequency of the highest harmonic considered to have a significant amount of power, plus the residual carrier frequency and the single-sided data bandwidth. This implies that excessive equipment is needed for data processing and storage. For limited resources, this may even mean loss of data. On the other hand, if the square-wave subcarrier could be downconverted to a much lower frequency, as shown in Fig. 1 (b), then the sample rate can be reduced significantly, which leads to a smaller amount of data storage and more efficient data processing.

This article presents a method to downconvert the square-wave subcarrier, which includes finding a down-mixing signal and controlling the noise level. In the next section, the downconversion in the absence of noise will be studied. This will be followed by an analysis that considers noise.

II. Downconversion in the Absence of Noise

In order to use the existing telemetry recovery equipment, the downconverted subcarrier needs to preserve the

square-wave form at a lower subcarrier frequency. This can be achieved by a downconverting procedure shown in Fig. 2 (a), where the mixing signal $y(t)$ and the parameters of the lowpass filter (LPF) need to be carefully chosen.

A. Downconversion of the Subcarrier

The square-wave subcarrier, $D_2(t)$, can be represented in its Fourier series form, assuming that $D_2(t)$ has a unit amplitude and phase angle, θ_{sc}:

$$D_2(t) = \frac{4}{\pi}\sum_{n=1}^{\infty}\frac{1}{2n-1}\sin[(2n-1)(\omega_{sc}t+\theta_{sc})] \quad (1)$$

where $\omega_{sc} = 2\pi f_{sc}$.

In a practical system, it suffices to consider only a small number of terms of $D_2(t)$, typically up to the third or fifth harmonic because the power in the higher harmonics diminishes quickly.

Mixing $D_2(t)$ down can be achieved by multiplying $D_2(t)$ by a signal, $y(t)$, and then passing the product, $g_1(t) = D_2(t)y(t)$, through a low-pass filter, as shown in Fig. 2 (a). Assuming that the signal, $y(t)$, has the form

$$\begin{aligned} y(t) &= 2\ \mathrm{sgn}(\omega_{sc}-\omega_1)\sum_{m=1}^{N}\cos[(2m-1)(\omega_1 t+\theta_1)] \\ &= 2\ \mathrm{sgn}(\omega_{sc}-\omega_1)\cos(\omega_1 t+\theta_1) \\ &\quad + \cos 3(\omega_1 t+\theta_1) + \cos 5(\omega_1 t+\theta_1) \\ &\quad + \cdots + \cos[(2N-1)(\omega_1 t+\theta_1)] \end{aligned} \quad (2)$$

where $|\omega_{sc}-\omega_1|$ is very small. Note that Eq. (2) is not a Fourier expansion of a square wave. The authors will show that this signal can mix the square-wave subcarrier down to a much lower frequency.

Note that $y(t)$ can only have a finite number of terms to ensure convergence. This implies that the downconverted signal will not be a true square wave, instead only N number of harmonics will remain, with the highest harmonic being the $2N-1$. In the remainder of the discussion, it will be assumed that $N \leq 3$. By expanding $g_1(t) = D_2(t)y(t)$, one obtains

$$\begin{aligned} g_1(t) &= \left[\frac{4}{\pi}\sum_{n=1}^{\infty}\frac{1}{2n-1}\sin[(2n-1)(\omega_{sc}t+\theta_{sc})]\right] 2\ \mathrm{sgn}(\omega_{sc}-\omega_1)[\cos(\omega_1 t+\theta_1)+\cos 3(\omega_1 t+\theta_1)+\cos 5(\omega_1 t+\theta_1)] \\ &= \mathrm{sgn}(\omega_{sc}-\omega_1)\frac{4}{\pi}\Big\{\sin[(\omega_{sc}-\omega_1)t+\theta_{sc}-\theta_1] + \frac{1}{3}\sin 3[(\omega_{sc}-\omega_1)t+\theta_{sc}-\theta_1] + \frac{1}{5}\sin 5[(\omega_{sc}-\omega_1)t+\theta_{sc}-\theta_1] \\ &\quad + \sin[(\omega_{sc}-3\omega_1)t+\theta_{sc}-3\theta_1] + \sin[(\omega_{sc}-5\omega_1)t+\theta_{sc}-5\theta_1] + \sin[(\omega_{sc}+\omega_1)t+\theta_{sc}+\theta_1] \\ &\quad + \sin[(\omega_{sc}+3\omega_1)t+\theta_{sc}+3\theta_1] + \sin[(\omega_{sc}+5\omega_1)t+\theta_{sc}+5\theta_1] + \frac{1}{3}\sin[(3\omega_{sc}-\omega_1)t+3\theta_{sc}-\theta_1] \\ &\quad + \frac{1}{3}\sin[(3\omega_{sc}+\omega_1)t+3\theta_{sc}+\theta_1] + \frac{1}{3}\sin[(3\omega_{sc}+3\omega_1)t+3\theta_{sc}+3\theta_1] + \frac{1}{3}\sin[(3\omega_{sc}-5\omega_1)t+3\theta_{sc}-5\theta_1] \\ &\quad + \frac{1}{3}\sin[(3\omega_{sc}+5\omega_1)t+3\theta_{sc}+5\theta_1] + \frac{1}{5}\sin[(5\omega_{sc}-\omega_1)t+5\theta_{sc}-\theta_1] + \frac{1}{5}\sin[(5\omega_{sc}+\omega_1)t+5\theta_{sc}+\theta_1] \\ &\quad + \frac{1}{5}\sin[(5\omega_{sc}-3\omega_1)t+5\theta_{sc}-3\theta_1] + \frac{1}{5}\sin[(5\omega_{sc}+3\omega_1)t+5\theta_{sc}+3\theta_1] + \frac{1}{5}\sin[(5\omega_{sc}+5\omega_1)t+5\theta_{sc}+5\theta_1] + \cdots\Big\} \end{aligned} \quad (3)$$

By passing $D_2(t)y(t)$ through a lowpass filter with a proper cutoff frequency, the first three terms will remain.

$$z_1(t) = \text{sgn}(\omega_{sc} - \omega_1)\frac{4}{\pi}\left\{\sin[(\omega_{sc} - \omega_1)t + \theta_{sc} - \theta_1]\right.$$

$$+ \frac{1}{3}\sin 3[(\omega_{sc} - \omega_1)t + \theta_{sc} - \theta_1]$$

$$\left. + \frac{1}{5}\sin 5[(\omega_{sc} - \omega_1)t + \theta_{sc} - \theta_1]\right\} \quad (4)$$

which is similar to $D_2(t)$ except at a lower frequency.

The downconverting procedure can clearly be shown by using the following numerical example: The square-wave subcarrier that is used in the signal from Galileo has a fundamental frequency, 22.5 KHz, and it can be represented by the first three harmonics, assuming the phase of the subcarrier, $\theta_{sc} = 0$,

$$D_2(t) = \sin(22.5 \times 2\pi t) + \frac{1}{3}\sin(67.5 \times 2\pi t)$$

$$+ \frac{1}{5}\sin(112.5 \times 2\pi t)$$

Considering down-mixing $D_2(t)$ with a signal of the form shown in Eq. (2), where $\omega_1 = 24.75 \times 2\pi$ and $N = 3$

$$y(t) = -2\cos(24.75 \times 2\pi t) - 2\cos(74.25 \times 2\pi t)$$

$$- 2\cos(123.75 \times 2\pi t)$$

The product of $D_2(t)$ and $y(t)$ is

$$g_1(t) = D_2(t)y(t)$$

$$= \sin(2.25 \times 2\pi t) + \frac{1}{3}\sin(6.75 \times 2\pi t)$$

$$+ \frac{1}{5}\sin(11.25 \times 2\pi t) - \frac{1}{5}\sin(38.25 \times 2\pi t)$$

$$- \frac{1}{3}\sin(42.75 \times 2\pi t) - \sin(47.25 \times 2\pi t)$$

$$+ \sin(51.75 \times 2\pi t) + \frac{1}{3}\sin(56.25 \times 2\pi t)$$

$$- \frac{1}{5}\sin(87.75 \times 2\pi t) - \frac{1}{3}\sin(92.25 \times 2\pi t)$$

$$- \sin(96.75 \times 2\pi t) + \sin(101.25 \times 2\pi t)$$

$$- \frac{1}{5}\sin(137.25 \times 2\pi t) - \frac{1}{3}\sin(141.75 \times 2\pi t)$$

$$- \sin(146.25 \times 2\pi t) - \frac{1}{5}\sin(186.75 \times 2\pi t)$$

$$- \frac{1}{3}\sin(191.25 \times 2\pi t) - \frac{1}{5}\sin(236.25 \times 2\pi t)$$

where t is in milliseconds. It is clear that if an ideal lowpass filter is used with a cutoff frequency, f_L, being 11.25 KHz $< f_L <$ 38.25 KHz, then the output becomes

$$z_1(t) = \sin(2.25 \times 2\pi t) + \frac{1}{3}\sin(6.75 \times 2\pi t)$$

$$+ \frac{1}{5}\sin(11.25 \times 2\pi t)$$

which is similar to $D_2(t)$, except that the fundamental frequency is reduced from 22.5 to 2.25 KHz. Clearly, if the subcarrier is modulated by a slow-changing data sequence, a similar downconversion can occur.

It is important to notice that the resulting waveform represents a square wave due to the maintenance of proper frequency, phase, and amplitude relationships. It is distorted only by the truncation of the sequence to a finite length.

B. Downconversion of Subcarrier in the Presence of Data and a Residual Carrier

In the case where the residual carrier and the data are present, a typical downconverted spacecraft signal has in-phase and quadrature components of the form

$$x(t) = A\sin(\omega_r t + \theta_r) + BD_1(t)D_2(t)\cos(\omega_r t + \theta_r)$$

where A and B denote the amplitudes, $\omega_r = 2\pi f_r$ is the angular frequency of the residual carrier, $D_1(t)$ is the data

with bandwidth B_1, and $D_2(t)$ is the square-wave subcarrier with the frequency, f_{sc}. Without loss of generality, one assumes the residual carrier phase angle to be zero ($\theta_r = 0$) in the discussion that follows.

To keep the residual carrier present after the downconversion, the downconverting signal, $y(t)$, needs to be added to 1, as shown in Fig. 2 (b).

The lowpass filter input can be written as

$$\begin{aligned} g_2(t) &= x(t)[y(t)+1] \\ &= x(t)y(t) + x(t) \\ &= A\sin(\omega_r t) + BD_1(t)D_2(t)\cos(\omega_r t) \\ &\quad + A\sin(\omega_r t)y(t) + BD_1(t)D_2(t)\cos(\omega_r t)y(t) \end{aligned} \tag{5}$$

The first term on the right-hand side of Eq. (5) is the residual carrier with a very low frequency which will remain after low-pass filtering. The second and the third terms are centered on f_{sc} and f_1, respectively. Both f_{sc} and f_1 are much higher than $5|f_{sc}-f_1|$, which is the highest frequency of the desired terms, so the second and the third terms can be filtered out. The last term is

$$[BD_1(t)\cos(\omega_r t)] \times [D_2(t)y(t)]$$

Passing $D_2(t)y(t)$ through an LPF results in $z_1(t)$, as shown in Section II.A; passing the product through the LPF will result in $BD_1(t)\cos(\omega_r t)z_1(t)$, since $BD_1(t)\cos\times(\omega_r t)$ has a narrow bandwidth.

In the overall picture, the resulting LPF output, $z_2(t)$, is

$$\begin{aligned} z_2(t) &= A\sin(\omega_r t) + BD_1(t)\cos(\omega_r t) \\ &\quad \times [\sin(|\omega_{sc}-\omega_1|t+\theta) + \frac{1}{3}\sin 3(|\omega_{sc}-\omega_1|t+\theta) \\ &\quad + \frac{1}{5}\sin 5(|\omega_{sc}-\omega_1|t)+\theta)] \end{aligned} \tag{6}$$

where $\theta = \theta_{sc} - \theta_1$. The obtained output is similar to the original signal, $x(t)$, except that the fundamental frequency of the subcarrier in $z_2(t)$ is much lower than that in $x(t)$, the phase has been shifted from θ_{sc} to $\theta_{sc}-\theta_1$, and the subcarrier in $z_2(t)$ does not have an infinite number of terms.

C. Conditions on the Down-Mixing Signal Frequency

One condition on the down-mixing-signal frequency, f_1, is

$$|f_{sc} - f_1| > 2|f_r| + \frac{B_1}{2} \tag{7}$$

where f_r is the residual-carrier frequency, f_{sc} is the subcarrier frequency, and B_1 is the data-signal bandwidth, so that the down-converted signal does not mix up with the residual carrier.

To properly choose a down-mixing signal frequency, f_1, it is also necessary to find the lowest frequency of the undesirable term(s), and choose f_1 so that the lowest frequency of the undesirable terms is higher than the highest frequency of the desired terms; so they can be filtered out or kept, respectively.

From Eq. (3) in Section II.A, it can be seen that if $f_{sc} < f_1$, the lowest undesirable frequency is $(2N+1)f_{sc} - (2N-1)f_1 - |f_r| - B_1/2$, and the highest desired frequency is $(2N-1)(f_1 - f_{sc}) + |f_r| + B_1/2$. For there to be no aliasing between the desired terms and undesirable ones, it is necessary that

$$\begin{aligned} (2N-1)(f_1 - f_{sc}) + |f_r| + \frac{B_1}{2} &< (2N+1)f_{sc} \\ &\quad - (2N-1)f_1 - |f_r| - \frac{B_1}{2} \end{aligned} \tag{8}$$

Rearranging Inequality (8) and combining it with Inequality (7), one has

$$\begin{aligned} f_{sc} + 2|f_r| + \frac{B_1}{2} &< f_1 \\ &< \frac{2N}{2N-1}f_{sc} - \frac{1}{2N-1}\left(|f_r| + \frac{B_1}{2}\right) \end{aligned} \tag{9}$$

Similarly, if $f_{sc} > f_1$, then the lowest undesirable frequency becomes $(2N-1)f_1 - (2N-3)f_{sc} - |f_r| - B_1/2$, and the highest desired frequency, $(2N-1)(f_{sc}-f_1) + |f_r| + B_1/2$. In this case, the following is required,

$$\frac{2N-2}{2N-1}f_{sc} + \frac{1}{2N-1}\left(|f_r| + \frac{B_1}{2}\right) < f_1$$

$$< f_{sc} - 2|f_r| - \frac{B_1}{2} \quad (10)$$

As an example, let $N = 3$ and $f_1 > f_{sc}$, then Inequality (9) becomes

$$f_{sc} + 2|f_r| + \frac{B_1}{2} < f_1 < \frac{6}{5}f_{sc} - \frac{1}{5}|f_r| - \frac{1}{10}B_1 \quad (11)$$

For the signal from Galileo, $f_{sc} = 22.5$ KHz, and assuming that the data bandwidth $B_1 = 1$ KHz, $f_r = 1$ KHz, then 25 KHz $< f_1 <$ 26.7 KHz.

D. Conditions on LPF Cutoff Frequency

The cutoff frequency of the LPF needs to be between the highest frequency of the desired term and the lowest frequency of the undesirable term. Assuming that an ideal LPF is used, the cutoff frequency of the LPF, f_L, should be

$$(2N-1)|f_1 - f_{sc}| + |f_r| + \frac{B_1}{2} < f_L$$

$$< \min\left\{f_{sc} - |f_r| - \frac{B_1}{2}, f_1 - |f_r|\right\}$$

where $f_{sc} = \omega_{sc}/(2\pi)$ is the fundamental frequency of the square-wave subcarrier, $D_2(t)$, $f_1 = \omega_1/(2\pi)$ is the fundamental frequency of the mixing signal, $y(t)$, and N is the number of harmonics considered for the square wave.

Using the example of the Galileo signal again, if the cutoff frequency of the lowpass filter, $f_L < f_1 = 26$ KHz, then the condition that f_L has to be less than $7f_{sc} - 5f_1 - |f_r| - B_1/2 = 26$ KHz is satisfied, which implies that all the undesirable terms will be filtered out.

In summary, it is possible to downconvert the subcarrier to a much lower frequency, which leads to a much lower sample rate when an ideal lowpass filter is used in the absence of noise.

III. In the Presence of Noise

Assuming now that the signal is contaminated by an additive white noise, the input to the downconverter is: $x(t) + \mathbf{n}(t)$, where $\mathbf{n}(t)$ is additive white Gaussian noise.[1] The input to the lowpass filter will be:

$$[x(t) + \mathbf{n}(t)][y(t) + 1] = x(t)y(t) + x(t) + \mathbf{n}(t)y(t) + \mathbf{n}(t)$$

The third term in the above expression needs to be expanded

$$\mathbf{n}(t)y(t) = \mathbf{n}(t)2[\cos(\omega_1 t + \Theta_1) + \cos 3(\omega_1 t + \Theta_1) + \cos 5(\omega_1 t + \Theta_1)]$$

where $\mathbf{n}(t)$ is a stationary additive white noise with zero mean, and autocorrelation

$$R_{nn}(t+\tau, t) = \frac{N_0}{2}\delta(\tau)$$

The product of $\mathbf{n}(t)y(t)$ results in a stochastic process with zero mean and autocorrelation [1]

$$R(t+\tau, t) = \mathrm{E}\{\mathbf{n}(t+\tau)\mathbf{n}^*(t)y(t+\tau)y^*(t)\}$$

$$= R_{nn}(\tau)\mathrm{E}\{y(t+\tau)y(t)\} \quad (12)$$

since $\mathbf{n}(t)$ and Θ_1 are independent. By expanding the term $y(t+\tau)y(t)$, one has

$$\begin{aligned} y(t+\tau)y(t) &= 4\{\cos[\omega_1(t+\tau)+\Theta_1] + \cos 3[\omega_1(t+\tau)+\Theta_1] \\ &\quad + \cos 5[\omega_1(t+\tau)+\Theta_1]\}[\cos(\omega_1 t + \Theta_1) \\ &\quad + \cos 3(\omega_1 t + \Theta_1) + \cos 5(\omega_1 t + \Theta_1)] \\ &= 2\{\cos\omega_1\tau + \cos[\omega_1(2t+\tau) + 2\Theta_1] \\ &\quad + \cos[\omega_1(-2t+\tau) - 2\Theta_1] + \cos[\omega_1(4t+\tau) + 4\Theta_1] \\ &\quad + \cos[\omega_1(-4t+\tau) - 4\Theta_1] + \cos[\omega_1(6t+\tau) + 6\Theta_1] \\ &\quad + \cos(\omega_1(2t+3\tau) + 2\Theta_1) + \cos[\omega_1(4t+3\tau) + 4\Theta_1] \\ &\quad + \cos\omega_1(3\tau) + \cos[\omega_1(6t+3\tau) + 6\Theta_1] \end{aligned}$$

[1] All random variables and stochastic processes are boldfaced.

$$+ \cos[\omega_1(-2t+3\tau) - 2\Theta_1] + \cos[\omega_1(8t+3\tau) + 8\Theta_1]$$

$$+ \cos[\omega_1(4t+5\tau) + 4\Theta_1] + \cos[\omega_1(6t+5\tau) + 6\Theta_1]$$

$$+ \cos[\omega_1(2t+5\tau) + 2\Theta_1] + \cos[\omega_1(8t+5\tau) + 8\Theta_1]$$

$$+ \cos\omega_1(5\tau) + \cos[\omega_1(10t+5\tau) + 10\Theta_1]\}$$

By observing $y(t+\tau)y(t)$, one can see that there are two types of functions that are adding: $\cos(p\tau)$ and $\cos(qt + p\tau + q\theta_1)$, where p and q are integers that belong to the sets $\{1,3,5\}$ and $\{\pm 2, \pm 4, 6, 8, 10)$, respectively. The first type of function is deterministic whose expectation is itself. To evaluate the expectation of the second type of function, one assumes that Θ_1 is a random variable with uniform distribution in $[-\pi, \pi]$. Then the expectation becomes [2].

$$\begin{aligned} \mathrm{E}\{\cos(qt + p\tau + q\Theta_1)\} &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \cos(qt + p\tau + q\Theta_1)d\Theta_1 \\ &= \frac{1}{2\pi q}[\sin(qt + p\tau + q\pi) \\ &\quad - \sin(qt + p\tau - q\pi)] \\ &= 0 \end{aligned}$$

since q is an even number. So the autocorrelation function of the process $\mathbf{n}(t)y(t)$ becomes

$$R(\tau) = N_0\delta(\tau)[\cos(\omega_1\tau) + \cos(3\omega_1\tau) + \cos(5\omega_1\tau)]$$

The power spectrum of the process $\mathbf{n}(t)y(t)$, $S(\omega)$, is the Fourier transform of the autocorrelation function, $R(\tau)$ [2],

$$S(\omega) = \mathcal{F}\{R(\tau)\} = 3N_0 \qquad (13)$$

This implies that the noise level will be increased by about 7.8 dB/Hz after the downconversion, which is not acceptable. One solution to this problem is to put $N+1$ bandpass filters before the downconversion, in order to select only the residual carrier and the subcarrier harmonics, as illustrated in Fig. 3.

If three harmonics are considered for the subcarrier, then four ideal bandpass filters connected in parallel are needed. The first one is used to keep the residual carrier, with its center frequency at $|f_r|$. The other three filters are used for the three harmonics of the square-wave subcarrier and the data signal around them. Their center frequencies should be the appropriate harmonics of the subcarrier [see Fig. 1(a)]. The bandwidth of these harmonic filters, f_B, should be as narrow as possible with the condition $f_B \geq B_1 + 2|f_r|$ so that the data signal is allowed to pass. However if the bandwidth of the BPF is too narrow, there will be colored noise after the downconversion. To obtain the so-called white noise[2] after the downconversion, the BPF bandwidth should be $f_B = (f_L - |f_r|)\,N$.

Note that the above analysis ignores the noise power of the output of the bandpass filter for the residual carrier since its bandwidth is much smaller than that of the data signal.

IV. Boundaries on Downconversion

The selection of the downconversion circuit and associated filters must consider the following boundaries to ensure no aliasing of spectra. Assuming that all the filters are ideal, the conditions are summarized as follows.

The cutoff frequency of the lowpass filter has to be

$$\begin{aligned} (2N-1)|f_1 - f_{sc}| + |f_r| + \frac{B_1}{2} &< f_L \\ &< \min\left\{f_1 - |f_r|, f_{sc} - |f_r| - \frac{B_1}{2}\right\} \end{aligned}$$

The lower bound is to keep the $2N-1$th harmonic of the square wave plus the data around it, and the upper bound is to eliminate all the other undesirable terms.

The bandwidth of the bandpass filters has to be

$$f_B \geq B_1 + 2|f_r|$$

and

$$f_B = \frac{f_L - |f_r|}{N}$$

to keep the noise level as low as if no downconversion is ever done.

[2] Flat average power spectrum within the considered band.

The mixing signal can be

$$y(t) + 1 = 2 \text{ sgn } (\omega_{sc} - \omega_1) \{\cos(\omega_1 t) + \cos(3\omega_1 t) + \cos(5\omega_1 t) + \cdots + \cos[(2N-1)\omega_1 t]\} + 1$$

The fundamental frequency of $y(t)$ has to be either

$$f_{sc} + 2|f_r| + \frac{B_1}{2} < f_1 < \frac{2N}{2N-1} f_{sc} - \frac{1}{2N-1}(|f_r| + \frac{B_1}{2}) \quad (14)$$

or

$$\frac{2N-2}{2N-1} f_{sc} + \frac{1}{2N-1}(|f_r| + \frac{B_1}{2}) < f_1 < f_{sc} - 2|f_r| - \frac{B_1}{2} \quad (15)$$

The lower bound in Inequality (14), the upper bound in Inequality (15), is for the downconverted signal not to overlap with the residual carrier and between the harmonics, and the upper bound in Inequality (14), the lower bound in Inequality (15), is for the LPF to be able to filter out the undesirable terms.

As a conclusion, the single-sided bandwidth of the down converted signal, $BW/2$, is

$$\frac{BW}{2} = (2N-1)|f_{sc} - f_1| + |f_r| + \frac{B_1}{2}$$

and Figs. 4 and 5 illustrate this bandwidth with B_1 and f_r as variables.

V. Conclusions

This article discussed the possibility of downconverting the square-wave subcarrier of signals such as those sent from Galileo. A practical method is given to compress the bandwidth of the square-wave subcarrier by using a finite number of harmonics, where most of the received signal power is located.

References

[1] A. Papoulis, *Signal Analysis*, New York: McGraw-Hill, p. 300, 1977.

[2] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, New York: McGraw-Hill, pp. 142 and 338, 1965.

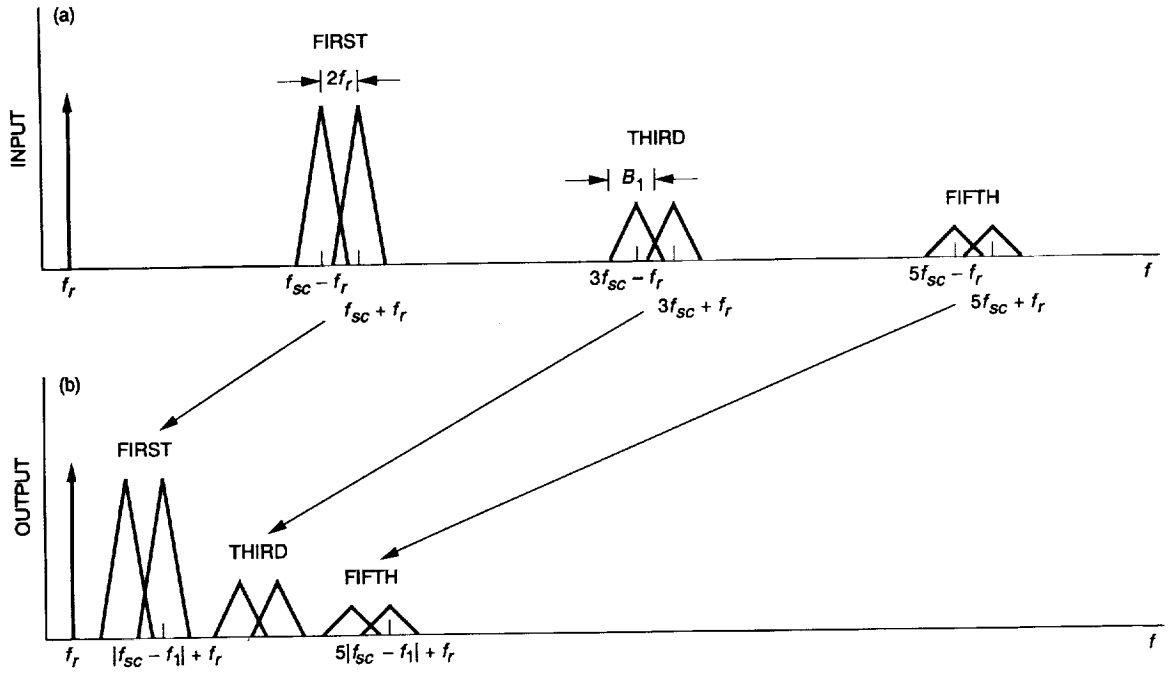


Fig. 1. Bandwidth compression shown in the frequency domain: (a) Spectrum before downconversion and (b) Spectrum after downconversion.

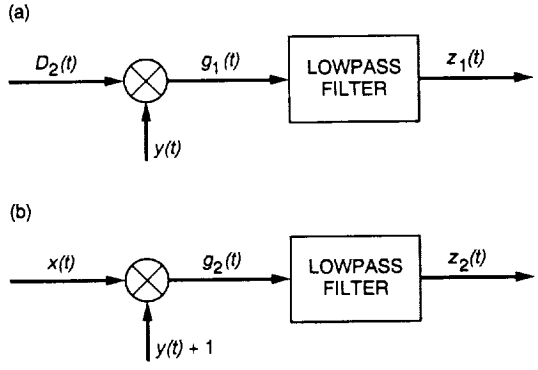


Fig. 2. The downconversion in the absence of noise: (a) Square-wave subcarrier only and (b) Residual carrier and data present.

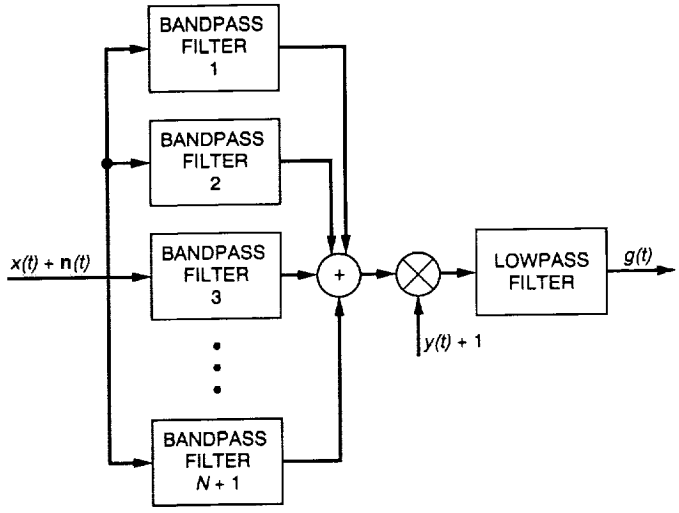


Fig. 3. The downconverter in the presence of noise.

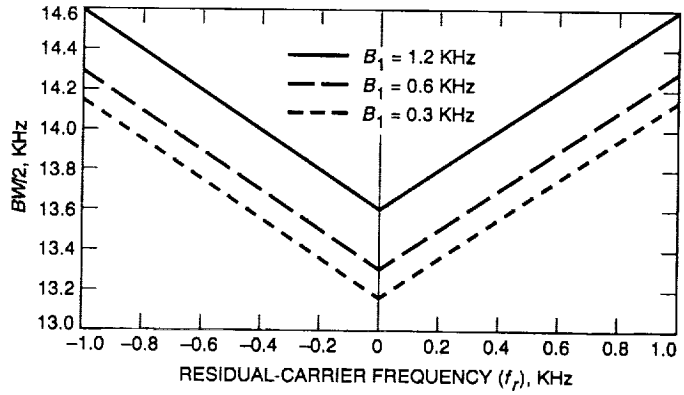


Fig. 4. The single-sided bandwidth of the downconverted signal. Assumptions: (a) ideal lowpass filter and bandpass filter and (b) Constant f_1, $|f_{sc} - f_1| = 2.6$ KHz.

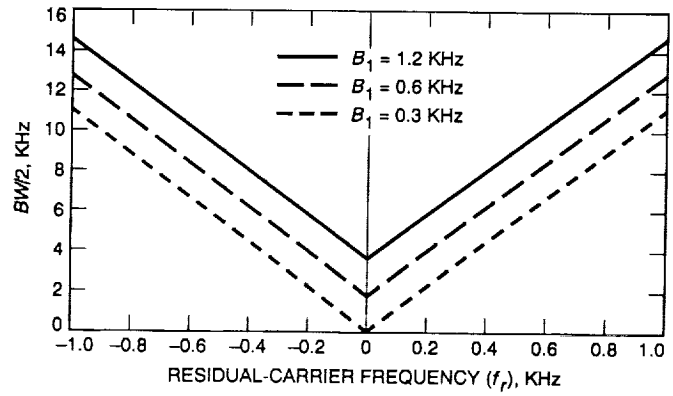


Fig. 5. The minimal single-sided bandwidth of the downconverted signal. Assumption: ideal lowpass filter and bandpass filter.

An OSI Architecture for the Deep Space Network

W. R. Heuser
Advanced Information Systems Section

This article presents an Open Systems Interconnection (OSI) architecture developed for the Deep Space Network. An historical review is provided to establish the context for current United States Government policy on interprocessor communication standards. An introduction to the OSI architecture, its seven-layer approach, and an overview of application service entities are furnished as a tutorial. Finally, the results of a prototype system developed for monitor and control of a Deep Space Station are also presented.

I. Introduction

The Deep Space Network (DSN) has, from its inception, employed automation whenever possible to support its mission: telecommunications with spacecraft in deep space. Over the past 25 years, advances in computer automation have accounted for substantial increases in operational productivity. Today, the Deep Space Stations are operated with a fraction of the personnel required 20 years ago. The DSN has evolved into a distributed computer system with all of its elements interconnected. The software programs (called application programs) used to operate the DSN are critically dependent on the software and hardware that support the exchange of data among DSN computers.

The development, integration, management, and maintenance of large distributed computer software systems is a costly enterprise. Although the price of computer hardware has dropped significantly over the past decade, the cost of software development has increased. In the DSN, software has become a major cost element. For example, the Signal Processing Center (SPC) modification project is a multimillion dollar effort and almost 80 percent is earmarked for software development. The DSN is not alone in facing the financial burden of the software development effort. The proliferation of distributed systems is a challenge facing government and the private sector.

Over the past two years, the DSN's Information Systems Division has conducted research into a distributed computer software architecture for the DSN that offers a solution to the high cost of distributed systems. The Open Systems Interconnection (OSI) architecture is an approach to interprocessor communications developed by the International Organization for Standardization (ISO) and is the focus of that research effort. The reasons for this focus are

(1) The OSI architecture meets the functional requirements of the DSN.

(2) The OSI architecture is designed to decouple the application program from the communications process. With OSI standards, the DSN can purchase commercial communication services and focus its resources on specific DSN applications.

(3) The OSI architecture has been adopted by the United States Government as the standard for distributed government computer systems [1].

(4) The OSI architecture has been selected by the Consultative Committee for Space Data Systems (CCSDS) as the model for the international space community [2].

An OSI-based architecture is expected to reduce the cost of distributed computer systems by providing standard communication interfaces for application programs. Market forces are driving the development of commercial products that provide communication services in accordance with the international standards. Products that meet the standards provide interprocessor communications between computers from different vendors. Competition between vendors will drive costs down.

This article is the result of two years of extensive research. It is written to provide a general education on OSI as well as to present the results of the research and prototype effort. To accommodate a wide audience with different levels of knowledge, the article is composed of four independent sections to permit selective reading. The sections are

(1) Section II: An Historical Perspective and Government Policy. An historical perspective on interprocessor communications is presented to provide an understanding of how the technology and government policy evolved.

(2) Section III: The OSI Basic Reference Model and Application Entities. An overview of the OSI Basic Reference Model is presented with a description of each layer's function and a description of the current OSI application entities.

(3) Section IV: An OSI Architecture for the DSN. An overview of an OSI-based DSN architecture is described. The DSN's data flow is partitioned between several application entities, and areas for additional research are identified.

(4) Section V: A DSCC Monitor and Control Prototype. A detailed report on an OSI-based Monitor and Control architecture for DSN tracking stations is presented. The potential for major cost savings is analyzed and the results of a prototype effort are detailed.

This approach is necessary because, in order to understand where one is and where one is going, it is necessary to understand where one has been. To achieve a comprehensive solution, one must understand the overall approach and the total problem. Finally, no large-scale implementation should be initiated without a small-scale prototype to verify the approach.

II. An Historical Perspective and Government Policy

In the late 1960s, the Defense Advanced Research Projects Agency (DARPA) recognized not only the demand for direct computer data exchange but the necessity for industrial standards that would be independent of the computer manufacturer. The Department of Defense (DoD) was already experienced in the business of computer data exchange and the inability of systems built by different companies to communicate. An early example was the 1950s Nike Ajax air defense system and the North American early warning radar system. The air defense system was developed under Army supervision and the radar system was developed under the Air Force [3]. When the time came to exchange data for a coordinated defensive response, the two systems were incompatible. With this historical perspective, DARPA began a long-term research effort through university and industrial laboratories to develop standards for interprocessor communications.

The first steps involved simple systems with two computers from the same manufacturer, connected by a cable. Addressing and protocol issues arose with the addition of the third and then the fourth computer. Still more problems arose with the introduction of computers from different manufacturers. Multiple cables were replaced by the development of media access control technology, and slowly the term local area network (LAN) emerged from the laboratory to the commercial world. DARPA's efforts culminated in 1983 with the publication of two DoD standards for interprocessor communications: Transmission Control Protocol (TCP) [4] and Internet Protocol (IP) [5]. Today, work continues in the university and commercial arenas, expanding what has become known as the DoD suite of TCP/IP standards.

The effort to develop standards for the computer industry is by no means a DARPA-exclusive activity. A number of different organizations, such as the Consultative Committee for International Telegraph and Telephone (CCITT) and the Institute of Electrical and Electronics Engineers (IEEE), are engaged in the standards development effort. These organizations have provided a wide variety of standards for different telecommunications media and media access techniques. The American National Standards Institute (ANSI) has contributed standards for character encoding and computer languages.

A more comprehensive and general solution to the problems of interprocessor communications is under development by the International Organization for Standardization (ISO). ISO began its effort in the mid-1970s and has

built on the lessons learned by DARPA. Called Open Systems Interconnection (OSI), the ISO approach provides for rapid changes in computer hardware and telecommunications technology with an architecture that insulates the user program from the manufacturer-dependent communication services. In addition, ISO recognized that the cost of software development was exceeding the cost of computer hardware, a significant change from 20 years earlier. Consequently, the OSI architecture was also designed to allow software systems to transcend computer hardware changes and communication systems evolution.

The growing costs to the federal government of computer hardware and software prompted the Office of Management and Budget in 1984, through the Office of the Chief Executive, to commission a study by the National Academy of Sciences on the status of interprocessor communication standards. The report *Transport Protocols for Department of Defense Data Networks* was published by the National Research Council (NRC) in February 1985 [6], with the recommendation that the United States Department of Defense adopt the International Organization for Standardization's OSI as the basis for all interprocessor communications. The NRC committee made its recommendation based primarily on two considerations:

(1) The ISO and DoD protocols are basically equivalent at the transport level.

(2) The worldwide market demand for ISO protocols is far larger than the market for the DoD protocol suite.

In response, the Department of Defense adopted OSI protocols as costandards with the DoD standards, with plans to make them "the sole mandatory interoperable protocol suite" [7].

In 1987, the United States Congress passed the Computer Security Act [8], which established the National Bureau of Standards (NBS) as the sole government agency responsible for the development of computer standards. This legislation also created a new category of Federal Information Processing Standards (FIPS's) called compulsory standards and modified the federal property laws to impose compulsory standards on the procurement of all federal government property. In April 1987, the first draft of the Government Open Systems Interconnection Profile (GOSIP) was released for comment. In August 1988, the draft document became the first compulsory Federal Information Processing Standard—146 [1].

The motivations for GOSIP are clearly stated in its introduction. "In the past, vendor-specific implementations of data communications protocols led to isolated domains of information, very difficult and expensive to bridge." Through GOSIP, "... the government expects to realize significant savings through reducing duplicate circuits and wiring, training, custom software, workstations, and custom hardware interfaces."

The Department of Veterans Affairs (VA) provides an excellent example of why GOSIP is necessary. The VA is a very large government agency, spread around the world, with computer networks numbering in the hundreds. Given the size of the VA, it is impossible for Congress to allocate funds to replace all the VA networks with a proprietary computer network in a single budget appropriation. In addition, the competitive procurement process makes it impossible to guarantee a single vendor source spread over many procurement cycles. Consequently, a vendor-independent standard is the only mechanism available for the acquisition of network products and services that ensure interoperability, and the selection of OSI provides the United States Government with an internationally recognized standard supported by a wide variety of vendors [9].

The National Institute of Standards and Technology (NIST, formerly the National Bureau of Standards) is responsible for the development of OSI network services in the government. Through GOSIP, NIST is creating the government market for commercial OSI products. NIST sponsors quarterly meetings of the NIST Workshop for Implementors of OSI to assist industry in the development of these OSI products. In addition, NIST is coordinating the conformance testing effort to assist government agencies in determining the interoperability of commercial products.

III. The OSI Basic Reference Model and Application Entities

A. Basic Reference Model

The objective of the OSI architecture is a system where user programs or application programs can employ the resources of any processor in a network without concern for the communications process or the computer hardware. To achieve this objective, ISO adopted a layered approach based on the functional partitioning of the communications process. The Basic Reference Model [10] emerged as an international standard (ISO 7498) in 1984 and defined seven layers, the so-called seven layer cake (see Fig. 1). Each layer has been assigned a specific set of services and an associate set of protocols. The seven layers are summarized below [11,12].

1. **The application layer.** This layer provides the user program with an interface to an OSI system. In this case, the user is any computer program or application program that requires interprocessor communications. A number of common communication functions have been identified and grouped into so-called application entities with standard services and protocols. The application entities are the key to open systems and will be discussed in more detail later.

2. **The presentation layer.** This layer (ISO 8823/9576) provides a common representation of application data that is communicated between application entities. Common representation refers to the encoding of data and the order of bits and bytes exchanged between processors. For example, the exchange of data between a processor using American Standard Code for Information Interchange (ASCII) encoded characters and a processor using Extended Binary Coded Decimal Interchange Code (EBCDIC) encoded characters requires a data translation before the information can be utilized. Another example of a presentation issue is the exchange of data between a 32-bit/word computer processor and a 16-bit/word processor where the bit and byte ordering must be adjusted for the correct interpretation of the data.

3. **The session layer.** This layer (ISO 8327/9548) provides dialogue services for those functions that require the establishment of a connection and synchronization between any two machines prior to the exchange of information. This layer provides the "Are you there"–"Yes I am" exchange prior to the exchange of application data.

4. **The transport layer.** This layer (ISO 8073/8602) provides those services required for the reliable end-to-end transmission of data. The issues of error detection, error recovery, and multiplexing are network dependent, and the transport layer manipulates the underlying network services to provide the appropriate quality of service (QOS).

5. **The network layer.** This layer (ISO 8473) provides the networkwide (or internetwork) services required to transmit, route, and relay data between computers without regard for the communications medium. In networks composed of several segments connected with bridges or independent networks connected through gateways, the network layer provides the services and protocols necessary to deliver the data to its final destination.

6. **The data-link layer.** This layer (ISO 8802-x) provides support for the physical medium employed in the transmission of data. ISO has established standards for a wide variety of transport media including four types of local area networks: an Ethernet (ISO 8802-3), a token bus (ISO 8802-4), a token ring (ISO 8802-5), and a Fiber Distributed Data Interface (FDDI).

7. **The physical layer.** This layer is the physical communication medium that supports the transmission of bits and is concerned with the electrical interface to the equipment supporting the transmission.

The process of establishing ISO standards is a difficult and time-consuming effort that involves committees from all participating nations. A standard begins as a draft proposal (DP), moves to a Draft International Standard (DIS), and achieves International Standard (IS) status over a period of three to six years. The key standards for the lower six layers (presentation layer to physical layer) have been International Standards (IS's) for more than three years. Some of the application layer entities are International Standards and others continue under development [12].

B. Application Entities

Much of the standards effort is now focused on the application entities, which are the key to the OSI architecture. Using the appropriate application entity, standardized software functions provide access to the OSI services required for interprocessor communications. The application entities are summarized below.

1. **The Manufacturing Message Specification entity.** This application entity (MMS, ISO 9506) provides a set of services developed from the Manufacturing Automation Protocol (MAP) initiative at General Motors in the early 1980s. Designed as a process control standard, the MMS Standard achieved IS status in 1989 and is aimed at direct interprocessor communications among machines on a factory floor. In the factory environment, an assembly line can be composed of hundreds of machines performing complex, precision tasks all as a unit. If any element of the assembly line fails to perform or performs at less than optimum performance, the assembly line as a whole fails to perform. The interprocessor communication provided by MMS is designed to support this type of environment. The application of MMS to the DSN is discussed in detail later [13].

2. **The File Transfer, Access and Management entity.** This application entity (FTAM, ISO 8571) provides a service and protocol standard to access and manage files in an open system and was one of the first application entities to achieve IS status in 1988. Using FTAM, a user program can open, read, write, and close files on

another processor just as though it were manipulating a local file. FTAM also provides services for copying files and obtaining file directories from remote systems. The application of FTAM services might greatly simplify the problems associated with centralized data recording at the Deep Space Communication Complexes (DSCC's). For example, a telemetry processor can open a file for recording on a remote network file server at the beginning of a spacecraft support pass, write the telemetry data to that file, and close the file at the completion of the activity. Playback could be accomplished in a similar fashion using the FTAM Read services.

3. The Network Management entity. This entity (NM, ISO 9595/9596) developed as a result of the growing use of distributed systems and the lack of mechanisms to monitor and manage the communications resources of networks. Though network management is still in the draft proposal stage of the standards process, five service elements have been identified as the core of this application entity:

(1) Performance will be monitored at each layer of the seven-layer architecture, providing information on byte counts, time delays, data rates, and other statistical information related to the performance of the communication process.

(2) Error Reporting and Logging will detect errors and failures in the communications process.

(3) Security guards against unauthorized access.

(4) Configuration Management will provide for the allocation and assignment of communications resources.

(5) Accounting Management will provide audit services [14–16].

4. The Message Handling Systems Service entity. This entity (MHS, ISO 10021), unlike other application layer entities, is a collection of international standards that together form the basis for OSI electronic mail (the X.400 standards). Many of the standards under the umbrella of MHS are the result of a long collaboration between ISO and the CCITT standards effort. The distribution of DSN operational support messages is the result of an excellent match of functional requirements.

5. The Directory Services entity. This entity (DS, ISO 9594) is intended to provide a global interconnected directory for all types of OSI entities, individuals, distribution lists, application entities, and general agents using network communications services. An international standard since 1989, Directory Services provides address resolution based on logical processor names anywhere in the network.

6. The Remote Database Access entity. This entity (RDA) is a proposed standard stimulated by the expanding use of commercial database systems. Still in the draft proposal phase, RDA is intended to provide a set of standards for access to any "open systems" database through a set of standard functions and a standard Sequential Query Language (SQL).

7. The Virtual Terminal entity. This entity (VT, ISO 9040) is a standard for terminals and hosts to communicate across different networks without requiring that one side know the terminal characteristics handled by the other side. An International Standard since 1988, VT provides a generic set of terminal characteristics for communication, which can be mapped to local terminal characteristics for display.

8. The Job Transfer and Manipulation entity. This entity (JTM, ISO 8831/8832) is designed to support computer-to-computer communications for the purpose of performing work remotely. JTM developed as a spin-off of batch processing and provides the protocols necessary to transfer and perform jobs on processors in a distributed system.

One of the most important aspects of the layered architecture is the flexibility to select application layer services that can operate over a number of different data-link and physical layer transmission standards. For example, an organization can implement an application based on the Manufacturing Message Specification operating over an Ethernet (ISO 8802-3) and move to an FDDI as its data rate requirements change. This approach provides an evolutionary path for distributed systems while preserving the organization's software investment.

In addition, the OSI-layered architecture is open to expansion as new problems and technologies evolve. In the mid-1980s for example, the Massachusetts Institute of Technology developed X Windows to support a variety of computer terminals. The X Windows concept is now in the standards process and will emerge in the future as a companion standard to the Virtual Terminal application entity [17].

IV. An OSI Architecture for the DSN

An OSI architecture for the DSN would provide access to communication resources through a standard OSI application layer entity. The appropriate selection of the application entity is the key and is based on the type of data flow to be supported. A study of DSN data flow reveals two general categories of data flow:

(1) Real-time spacecraft data: acquired, recorded, and transmitted back to JPL to support flight projects (the end product of the DSN).

(2) Process control: the data exchange that supports all aspects of DSN operations.

Included in the category of real-time spacecraft data are subcategories such as the playback of recorded spacecraft data in nonreal time. Under the category of process control are subcategories including the transmission of support data and network control data. The partitioning of data flow into these two categories provides a starting point for the partitioning of the system.

Under the category of real-time spacecraft data, the DSN is responsible for recording the spacecraft data types at the DSCC's. In addition, the DSN is frequently required to relay the spacecraft data types back to JPL in near-real time. The services and protocols provided by FTAM will support the functional requirements for recording spacecraft data at the Deep Space Stations. The telemetry subsystems at the stations can use FTAM software to open a client–server relationship with a station data recording subsystem (commonly referred to as a network file server). The FTAM services F_OPEN, F_WRITE, and F_CLOSE would be used to open files on the station data recorder, write the spacecraft data to those files, and close the files at the completion of the spacecraft track. The playback of recorded spacecraft data would employ a similar approach using the F_READ FTAM service in place of the F_WRITE service. These services use reliable connection-oriented Transport Protocol Four (TP-4) [11].

The process control category is composed of a diverse list of data flow types:

(1) Monitor data flow throughout the DSN to inform operations personnel of the status of all operational elements. The monitor data report the position of antennas, the condition of receivers, the available communications facilities, and the values of many other detailed components.

(2) Control data flow to each of the diverse elements to effect change and allocate DSN resources to support specific activities. Anomalies detected at the subsystem level generate messages to alert operations personnel.

(3) Support data files are transferred throughout the DSN to provide configuration and control information at the subsystem level.

The diversity of data flow results in a partitioning of functional requirements across several OSI application entities.

(1) Station Monitor and Control. The data flow for monitor and control of the Deep Space Stations is supported with the services of the OSI MMS and is discussed in more detail in the next section.

(2) Communications Resource Management. The OSI Network Management Service (NMS) provides the tools necessary to manage the communication resources of an OSI-compatible network. The security element of network management provides the apparatus to extend access to the DSN to those functions that require DSN connection, while providing the necessary safeguards against unauthorized access.

Other elements of network management identify and automatically report communications performance, problems, and failures, and thus provide DSN operations with the tools necessary to maintain the flow of data. The allocation of DSN resources in support of specific missions is reported through the accounting element of network management, which provides detailed information on DSN support performance and feedback to the scheduling teams.

(3) Support and Command Files. The transfer of support and command data is supported with the services of FTAM.

(4) Resource Identification. The identification of resources within the system is handled through address resolution with OSI Directory Services (DS).

(5) Operational Message Services. Administrative messages supporting operations are distributed through OSI Message Handling Systems (MHS—better known as X.400 electronic mail).

The typical architecture for an OSI-based DSN subsystem is shown in Fig. 2. The OSI application entities operating in each subsystem will vary from subsystem to subsystem depending on functional requirements.

One troubling issue remains and requires further research and analysis: the real-time relay of spacecraft data from the Deep Space Stations to JPL. The DSN employs dedicated leased communication lines to support the flow of data between JPL and the Deep Space Stations. The leased lines are generally expensive, have limited capacity (bandwidth), and operate through satellites in geosynchronous orbit. Traditionally, NASA has developed elaborate protocols to support this communication facility. These protocols are connectionless and designed to maxi-

mize the spacecraft data delivered over the limited capacity lines. Error detection and recovery are handled at the equivalent of the ISO data-link layer.

OSI connection-oriented protocols have been implemented over satellite communication systems and operate effectively. However, these protocols employ transport-level error detection and recovery. Large delays are typically experienced in transport-level acknowledgments over satellite systems and have a major negative impact on data rates. Connectionless OSI protocols could be used, however, to provide a service similar to the present DSN system. At present, there is no OSI application entity based on connectionless protocols, but the standards are in place for connectionless presentation (ISO 9576), session (ISO 9548), and transport (ISO 8602) layers. Development of a real-time data delivery OSI application layer standard is needed. Such a standard would employ Link Level Control 3 (LLC3) under ISO 8802-2 to provide a reliable data-link service.

The most important aspect of this solution is the incorporation of the ISO 8473 network protocol which provides conformance to the CCSDS recommendations and the proposed NASA Communications Network (NASCOM) gateway services [2,18]. Commercial high-speed multiprotocol routers that support ISO 8473 would provide LAN-to-LAN interconnection over standard commercial communication services. These routers are inexpensive and the cost to interconnect DSN facilities and support its customers would be greatly reduced.

V. A DSCC Monitor and Control Prototype

The Monitor and Control System for the DSN DSCC's was selected for an OSI-based prototype system. The prototype activity has provided hands-on experience with OSI protocols and is expected to substantially reduce the risk of errors associated with future implementations. The selection of the Monitor and Control System for the prototype effort was the result of several factors:

(1) The station monitor and control requirements are well defined and documented.

(2) The functional requirements for station monitor and control and the service specification for the MMS protocol are a close match.

(3) Commercial software products built to the MMS standard are readily available.

(4) The software and hardware products required for a prototype are available within the constraints for the research effort.

(5) The anticipated replacement of the DSCC Monitor and Control (DMC) Subsystem computer hardware in the mid-1990s provides an opportunity to transfer OSI technology to the DSN with substantial cost savings.

The present Mark IVa Monitor and Control System is a distributed software architecture. The DMC Subsystem provides the operator interface to the Mark IVa System. In addition to the software on the DMC Subsystem computers, software is required on each individual subsystem (see Fig. 3). The DMC Subsystem software consists of roughly 220,000 lines of HAL/S source code and represents an investment of almost 100 work-years.[1] The replacement of the aging Modcomp Classics will require moving that subsystem software to a new computer, and the translation of the present software to a supported computer language will require an equivalent investment. In addition, a survey of four subsystems reveals that 40 to 50 percent of the total lines of code on any subsystem are dedicated to supporting the Mark IVa Monitor and Control System (see Table 1). Any enhancement of monitor and control services to expand automation will require additional software development and modifications to DMC and all other subsystems.

An OSI-based monitor and control architecture would make extensive use of commercial off-the-shelf (COTS) software packages. All the interprocessor communication software would use OSI-based commercial software packages. Most of the subsystem software required for the Mark IVa Monitor and Control System would be replaced with these OSI-based software packages. A commercial process-control software package would be employed to support the DMC Subsystem human/machine interface. The focus of DSN resources would shift from software development and maintenance to DSN unique applications.

A. Monitor and Control Functional Requirements

The Mark IVa upgrade (1983–1985) was the first application of a local area network (LAN) in the DSN to support centralized monitor and control. The LAN provides a communication highway to all of the station subsystems. The monitor and control computers provide software for the station personnel to control the subsystems and display subsystem status information. The subsystems provide software to interpret remote commands and generate status information. Using this approach, the Mark IVa architecture distributes elements of the Monitor and Con-

[1] S. Fowler, personal communication, Data Systems Section, Jet Propulsion Laboratory, Pasadena, California, April 1991.

trol System to all of the station subsystems. The Mark IVa architecture is schematically represented in Fig. 3.

An OSI-based architecture would redistribute the monitor and control functions and concentrate the human interface in the DMC Subsystem instead of the individual subsystems being controlled. Operator directives would be processed and displays generated by the DMC Subsystem and not by the individual subsystems. An MMS network server would send configuration and control information to the subsystems and poll the subsystems for status. A real-time database would link the human interface services with the MMS network server. Figure 4 is a graphical representation of this architecture. The key element in this approach is the MMS abstract model called the Virtual Manufacturing Device (VMD), which is used to describe the externally visible characteristics of a real device. Software modules that manipulate a real device are called VMD objects. These objects can be manipulated using MMS services such as context management, variable access, domain management, semaphore management, and event management. State changes detected in the real device and defined in the VMD model can trigger MMS services. Applied to the DSN, subsystems would be modeled as one or more VMD's and operated across the network through MMS services.

The major differences between the Mark IVa architecture and an OSI-based architecture are that in the OSI-based architecture

(1) The subsystem interface is defined as a virtual machine and establishes a basis for automation through machine-to-machine communications.

(2) The architecture removes the human interface from the individual subsystems and isolates the human interface in the DMC Subsystem.

A comparison of Mark IVa functional requirements partitioned to MMS protocol services is presented in Table 2. A complete list of the 86 MMS services can be found in Appendix A.

B. Prototype Configuration

The prototype configuration incorporates three subsystems that mirror the typical functions required for station operations. This configuration is shown schematically in Fig. 5 and consists of three computers performing the following functions:

(1) The Link Monitor and Control (LMC) Subsystem provides the human interface to operate the station.

(2) The Very Long Baseline Interferometry (VLBI) Subsystem operates the Wide Channel Band (WCB) equipment for signal processing and data recording. The WCB equipment is installed at DSS 13 and is typical of DSN subsystems: uniquely designed for NASA with no commercial equivalent.

(3) The Advanced Power Meter (APM) is a collection of hardware designed to measure antenna system temperatures without continuous recalibration. This system is under development in the DSN Telecommunications Division and is installed at DSS 13.

All of the computers used to operate the prototype are Intel 80286/80386-based systems operating under Microsoft DOS 3.3 (see Table 3).

The motivation for the prototype configuration is derived from a requirement for accurate flux measurements of radio sources used for spacecraft navigation.[2] The catalog of radio sources for navigation is being expanded and the radio flux stability of a source is one of the selection criteria for the catalog. The Mark IVa architecture does not provide a mechanism for the VLBI Subsystem to control the Precision Power Meter (PPM) Subsystem; flux measurements can be acquired only through manual operation of the PPM by station personnel. The prototype was designed to examine how subsystem automation would be enhanced through the application of MMS protocols. To accomplish this goal, the prototype provides for three client–server relationships:

(1) Link Monitor and Control Subsystem (client) to the VLBI Subsystem (server).

(2) Link Monitor and Control Subsystem (client) to the Advanced Power Meter (server).

(3) VLBI Subsystem (client) to the Advanced Power Meter (server).

The ease with which these relationships are established for control and data acquisition indicates that the MMS will simplify the software development effort in the DSN.

The computers are interconnected with an ISO 8802-4 carrier band token bus local area network (5-MHz bandwidth) using Concord Communications 1210 and 1215 boards. The lower six OSI stack layers (presentation [layer six] to physical [layer one]) are downloaded to the Concord board during a configuration process. The MMS software (layer seven) was supplied by Systems Integration Specialists Corporation (SISCO). All MMS services

[2] R. Linfield and C. Jacobs, personal communication, Jet Propulsion Laboratory, Pasadena, California, February 1989.

in the prototype employ Transport Protocol Class 4 (TP4), a connection-oriented transport protocol with error detection and recovery. A commercial software package, FactoryLink by U.S. Data, was purchased and integrated into the prototype to support the human interface functions for the Link Monitor and Control Subsystem.

C. Results

The allocation of station resources is supported with the application of MMS context management services. These services include the following basic functions:

(1) Initiate.

(2) Conclude.

(3) Cancel.

(4) Abort.

In addition, the protocol specification includes the following functions to support the client–server relationship:

(1) Initiate request.

(2) Initiate indication.

(3) Initiate response (positive and negative).

(4) Initiate confirm (positive and negative).

(5) Conclude request (positive and negative).

(6) Conclude indication.

(7) Conclude response (positive and negative).

(8) Conclude confirm (positive and negative).

A typical protocol exchange is shown in Fig. 6. To begin a session, a client initiates a connection (mv_init()), which triggers an indication (u_mllp_a_assoc_ind()) on the target server. The target server responds (u_mllp_a_assoc_resp()) and triggers a confirmation (u_mv_read_conf()) on the client. This exchange is characteristic of all MMS confirmed services. A functional addressing scheme similar to that implemented in the Mark IVa System was realized through the Application Reference Name (AR-Name) conventions in the MMS protocol.

The distribution of support files (predicts) employs the MMS File Management services:

(1) Copy.

(2) Obtain.

(3) Open.

(4) Read.

(5) Close.

(6) Rename.

(7) Delete.

Again, the protocol specification includes functions to support the client–server relationship: request, indication, response, and confirmation. As in Mark IVa, the prototype Monitor and Control System distributes predict (PR) files and standards and limits (SL) files to the appropriate subsystem prior to a scheduled activity. The MMS Obtain File service is used to transmit the file names to be copied by the subsystem from the DMC Subsystem. The Obtain File Indication triggers the MMS software on the subsystem to copy the file across the LAN (see Fig. 7). The prototype code for an MMS service call requires the destination, the source file name, and the destination file name:

send_file(subsystem, source file, destination file)

or more specifically,

send_file("VLBI", "PREDICTS.DAT",

"NOVA1987.DAT")

Subsystem Directives, Displays, Events, and Alarms are supported through the application of MMS Variable Access services:

(1) Read variable.

(2) Write variable.

(3) Information report.

(4) Get variable access attributes.

(5) Define named variable.

(6) Delete variable access.

(7) Define named variable list.

(8) Get named variable list attributes.

(9) Delete named variable list.

(10) Define named type.

(11) Get named type attributes.

(12) Delete named type.

These services can be used to read or write a wide range of MMS standard variable types defined in the protocol specification:

(1) Boolean—8 bits.

(2) Integer8—8 bits.

(3) Integer16—16 bits.

(4) Integer32—32 bits.

(5) Unsigned Integer8—8 bits.

(6) Unsigned Integer16—16 bits.

(7) Unsigned Integer32—32 bits.

(8) Floating point—32 bits.

(9) Double floating point—64 bits.

(10) ASCII String8—8 bytes.

(11) ASCII String16—16 bytes.

(12) ASCII String32—32 bytes.

(13) ASCII String64—64 bytes.

(14) ASCII String128—128 bytes.

(15) ASCII String256—256 bytes.

In addition, the protocol supports the definition of complex variable types that include arrays and data structures. For example, one can define a C language data structure for a VLBI Subsystem device called the IF Distributor. This device has two inputs (character strings), two attenuation controls (integer values), and two total power integrators (floating-point values):

```
struct IFD_TYPE { char  if_1_in[4],
                  char if_2_in[4],
                  short if_1_att,
                  short if_2_att,
                  float if_1_pwr,
                  float if_2_pwr } ifd_status
```

When defined as an MMS complex variable, the entire structure can be written or read across the network as though it were a single variable with the name "ifd_status" of type "IFD_TYPE." Arrays of simple variables and structures can be handled in the same way. In addition, structures can be nested, that is, a structure can be within a structure.

In an MMS-based Monitor and Control System, operator directives are entered, converted to data, and transmitted by the DMC Subsystem to the individual subsystems using the confirmed MMS Variable Write service. An MMS confirmed service requires a response from the server to the client to verify the success or failure of the service. For example, the VLBI Wide Channel Band Subsystem has an attenuation control that requires operator adjustment. In the prototype system, a change of attenuation to 23 dB is entered through the graphical user interface and the MMS Variable Write service is used to transmit the data across the LAN to the VLBI Subsystem. The prototype code for the MMS Variable Write service requires the subsystem destination, the variable name, the variable type, and the data:

```
write_named_var("VLBI", "if_1_att", "Integer16", 23)
```

Operator displays are constructed on the DMC Subsystem from data obtained across the LAN using the confirmed MMS Variable Read service. In the prototype, a polling system was implemented to read the data from each subsystem on a periodic basis. Again based on the earlier VLBI example, the prototype code for the MMS Variable Read service requires the subsystem source, the variable name and the variable type:

```
read_named_var("VLBI","if_1_att", "Integer16")
```

Event and Alarm conditions are reported across the LAN using the unconfirmed MMS Information Report (info_report) service, a variation on the Variable Write service. Unlike the Variable Read and Write services, which must be initiated by the client, the Information Report service can be initiated by a server. The protocol specification does not require the client to acknowledge an information report, therefore the service is unconfirmed. Though unconfirmed (see Fig. 8) at the application layer, the Information Report service employs TP4, a reliable transport protocol service. The first goal in the prototype effort was to simulate the functions in DSN operations today. The Mark IVa Event/Alarm messages consist of text information for the operator. The messages are transmitted by the subsystems to the DMC Subsystem with a category identification:

(1) PROMPT.

(2) PROGRESS advisory.

(3) COMPLETION advisory.

(4) DEVIATION advisory.

(5) WARNING alarm.

(6) CRITICAL alarm.

(7) EMERGENCY alarm.

In the prototype, seven MMS-named variables are defined with the same Mark IVa names, each as a 64-character ASCII string. The prototype code for the DSN Event Message service requires the subsystem destination, the event name, and the message:

event_msg(subsystem, type, message)

or more specifically,

event_msg("VLBI", "COMPLETION",

"Data recording started")

The commercial monitor and control software package employed in the prototype offers another approach to reporting event and alarm conditions. Commercial monitor and control packages provide a service to generate and manage alarms based on changes detected in their real-time database. Again using the VLBI Subsystem as an example, the VLBI Subsystem controller monitors all of its devices every 15 sec. If a deviation in the expected configuration is detected, the VLBI Subsystem software can use the MMS Information Report service to update the DMC Subsystem:

dsn_send_info_rpt(subsystem, var_name, var_type, data)

or more specifically,

dsn_send_info_rpt("VLBI", "if_1_att", "Integer16", 18)

The MMS network server on the DMC Subsystem updates the real-time database and the commercial alarm-management software triggers an operator alarm.

In the Mark IVa era, monitor data blocks are transmitted from each subsystem to the DMC Subsystem. In turn, the DMC Subsystem acts as a middleman and redistributes monitor data to all of the subsystems. The data contained in the monitor data block are individually negotiated in advance, in detail, down to the bit level. In an MMS-based architecture, the data acquired through Variable Read services to support DMC Subsystem displays replace the current monitor data blocks. The redistribution of monitor data is supported with Variable Write services. In addition, the careful application of direct subsystem-to-subsystem Variable Read services would eliminate the middleman function of the DMC Subsystem.

D. Prototype Performance

The performance of any computer system is highly dependent on the specific implementation and is not a function of the protocol alone. Some of the factors that impact system performance are

(1) Software architecture.

(2) Software implementation.

(3) Operating systems.

(4) Computer hardware.

(5) Communication medium.

(6) Supporting hardware.

The resource constraints for this research effort defined the hardware and software options available for the prototype. The selection of the token bus LAN was one consequence of the constraints. In addition, the effort did not permit the acquisition of a token bus network analyzer, the lack of which limited the range of performance tests. Given these limitations, performance tests were developed to provide a baseline for comparison with future implementations. These tests focus on the key services:

(1) Time required to establish a connection.

(2) File transfer data rates.

(3) Variable write data rates.

(3) Variable read data rates.

The average time required to establish a connection on the prototype system ranged from 0.2 to 0.3 sec. This measurement includes the time to build, transmit, and process all four elements of the specified protocol shown in Fig. 6. A direct comparison with the present system is difficult because today's DSN is based on connectionless protocols. However, the time required for an operator directive to be acknowledged in the present system ranges from 1 to 3 sec.

The performance results of the MMS Obtain File service are plotted in Fig. 9. These results indicate that rates of 4500 bytes/sec can easily be achieved. The difference in the performance of the two computers (8 MHz versus 6 MHz) reflects not only the central processing unit (CPU) performance but the difference in disk access speed of the two computers. The average disk access speed for the Everex (8-MHz) computer was 19.9 msec while the IBM AT (6-MHz) computer disk access speed was 37.9 msec. The Mark IVa support files range in size from 1000 to 80,000 bytes; the DMC Subsystem transmits these files at a rate of 1800 bytes/sec. Based on the prototype results, the implementation of MMS File Management services would double the present throughput performance.

The performance results for the Variable Access services are plotted in Figs. 10 and 11, and reflect their sensitivity to the CPU and memory speed of the prototype computers. The data rates reported for the prototype tests

are for application data. The number of bytes transmitted in an MMS message packet is a function of the specific MMS service and the application data. For example, the packet generated to write a single variable string of 1000 characters contains 1019 bytes. The packet generated to write an array of 500 16-bit integers (1000 bytes of data) contains 1948 bytes. The application layer data included to support the processing of the packet accounts for the difference in the packet size. (See Appendix B for a detailed example of the MMS message formulation.) Arrays of 16-bit integers, arrays of 32-bit floating point variables, and a series of ASCII strings were created to generate tests with packets of 4, 20, 200, 464, 732, and 1000 bytes of application data.

Once viewed as unnecessary protocol overhead, the application layer data perform a vital role in the communication process by defining all elements of the data packet. The benefits are realized through application software that is independent of the communication process. In the past, protocol design was motivated by communication bandwidth limitations. Today, modern LAN's have alleviated many of the bandwidth limitations. The performance results obtained for MMS Variable Access services are typical for packet transfer protocols. Note that Variable Reads and Writes of individual variables are less efficient than Reads and Writes of large data structures and arrays. Based on these results, an MMS-based Monitor and Control System should require subsystem VMD's designed to use data structures and arrays whenever possible and appropriate.

E. Intersubsystem Automation

One of the primary objectives of the prototype effort was the exploration of subsystem-to-subsystem automation using MMS protocols. For this effort, the software for the Advanced Power Meter was designed to operate in two modes:

(1) In single measurement mode, the power meter performs a simulated discrete system temperature measurement.

(2) In continuous mode, the power meter performs repeated system temperature measurements.

Software was developed for the VLBI Subsystem controller to establish a direct client–server relationship with the power meter controller. Under the command of the DMC Subsystem, the VLBI Subsystem can set the power meter operating mode (see Fig. 12). In single measurement mode, VLBI uses the MMS Variable Write service to trigger the power meter to perform a single system temperature measurement based on the contents of the VLBI predict schedule. On completion of the measurement, the power meter reports the results back to VLBI using the MMS Information Report service. In continuous mode, VLBI uses the MMS Variable Write service to set the power meter to continuous sampling mode, and the VLBI Subsystem performs an MMS Variable Read to obtain the latest results based on its predict schedule. The VLBI Subsystem and DMC Subsystem interfaces to the power meter are identical. The establishment of this relationship was straightforward and simple to implement.

F. Commercial Monitor and Control Packages

One of the benefits of implementing systems based on the standards is the availability of products designed to operate with the standards. In the process control world, there are a number of commercial products available for factory automation. These products provide graphical user interfaces for monitor and control, communication services to process control devices, logging of real-time performance data, operator alarm notification, and a variety of other services. The products are table driven and designed to be tailored and installed in any process control environment without software modifications. Several companies which produce these products have developed MMS interfaces to their systems. The U.S. Data product FactoryLink is one example. The FactoryLink DOS product (for IBM PC's) has been integrated in the prototype to provide operator control and a graphical user interface with time-ordered (trending) plots of real-time data [19].

G. Future Steps

The MMS prototype is the basis for a new Monitor and Control System to be developed and installed at DSS 13 in 1992. This system will provide centralized monitor and control for all of the core subsystems planned for the new 34-meter beam waveguide antenna. The DSS-13 effort will provide an expanded test bed to examine other MMS services such as

(1) Domain management.

(2) Semaphore management.

(3) Event management.

(4) Journal management.

Beyond the MMS effort, a full and complete evaluation of FTAM services and protocols is necessary before FTAM can be applied to centralized data recording at the DSN stations.

VI. Conclusions

The DSN can adopt an OSI architecture and would benefit from the application of OSI services and protocols in several areas.

(1) OSI would provide the DSN with a set of interprocessor communication standards that can be specified in all future implementations. Wide industrial support for OSI will insure a selection of vendors while providing compatibility with future implementations.

(2) The development costs for OSI products will be distributed over the worldwide market, reducing DSN costs for network services. Manufacturers will compete to have highly reliable, high performance products at a relatively low price.

(3) The application of the MMS would provide the foundation for increased automation of the DSN. MMS establishes the client-server relationship and services required to operate multiple computers operating as a single system.

(4) The application of FTAM services would provide the basis for network file servers in the DSN. FTAM services would be used to record data, play back data, and transfer files more reliably.

(5) The application of standard OSI protocols establishes an internetworking system based on the ISO 8473 internetwork protocol. Commercial Wide Area Network (WAN) bridges and routers would be employed to interconnect the distributed DSN facilities at reduced investment. In addition, the utilization of the ISO internetwork protocol would simplify the exchange of data between agencies, particularly as the international community moves to OSI.

(6) The application of OSI Network Management would provide the services to control communication resources, identify fault conditions, account for network utilization, and insure full security in an open system.

New products based on the OSI services are already under development. A number of companies are developing programmable logic controllers (PLC's) based on the OSI MMS standards; other companies are developing MMS servers to support commercial database products. Commercial products are already available to provide monitor and control systems for factories and their adaptation to OSI services is under way.

The cost of transition to OSI may be equal to, or even exceed, the cost of other networking solutions in the short term. However, the federal policy to adopt OSI is based on the long-term cost benefits (estimated by the National Research Council at 30 to 80 percent of the implementation cost for new computer systems) [6]. OSI-based MMS products have been introduced by companies such as Digital Equipment Corporation (DEC), International Business Machines (IBM), Hewlett-Packard, and Motorola for a wide range of computers, indicating industry's commitment to the standards. The cost of software development, test, and integration is thus distributed over their large customer base. In addition, the MMS Standard has been developed to meet the diverse demands of commercial industry and offers a spectrum of services that this research has shown to be more than adequate to meet the DSN requirements for DSCC monitor and control in the next century. Moreover, the layered OSI architecture would enhance the DSN's ability to cope with changing requirements and technologies.

In a broader sense, the adoption of ISO protocols would benefit the DSN by incorporating it into the world networking community. The Consultative Committee for Space Data Systems [2] has adopted the OSI architecture as the basis for international space data systems and the ISO 8473 network protocol is a key component of the CCSDS architecture. In addition, GOSIP requires all federal procurement of networking services to employ ISO protocols [1]. In response to GOSIP, the NASA Science Internet (NSI) is making the transition to OSI and NASCOM is planning a transition to OSI in the 1990s [18].[3] The adoption of an OSI architecture and the application of OSI protocols is necessary to meet the demands of DSN customers in a cost-effective manner.

[3] R. Nitzen, personal communication, NASA Headquarters, Washington, DC, April 1990.

References

[1] U.S. Department of Commerce, National Bureau of Standards, *Federal Information Processing Standards Publication—146, Government Open Systems Interconnection Profile*, Washington, DC, August 1988.

[2] Consultative Committee for Space Data Systems (CCSDS), *Advanced Orbiting Systems, Network and Data Links: Architectural Specification*, Blue Book, CCSDS 701.0-B-1, CCSDS Secretariat, NASA, Washington, DC, October 1989.

[3] C. Morgan, "Department of Defense Plans for Open Systems Interconnection," in *Case Studies in Implementing OSI* (tutorial), Interop Incorporated, Mountain View, California, October 1989.

[4] U.S. Department of Defense, *Military Standard Transmission Control Protocol*, MIL-STD-1776, Washington, DC, August 1983.

[5] U.S. Department of Defense, *Military Standard Internet Protocol*, MIL-STD-1777, Washington, DC, August 1983.

[6] National Research Council, *Transport Protocols for Department of Defense Data Networks*, PB85-176147, National Technical Information Service, Springfield, Virginia, February 1985.

[7] U.S. Department of Defense, *The Department of Defense Open Systems Interconnection (OSI) Implementation Strategy*, planning document, Reston, Virginia, May 1988.

[8] United States Congress, "Computer Security Act of 1987," Public Law 100-235, *Congressional Record*, vol. 133 (1987), approved January 8, 1988.

[9] R. C. Brooks, "Department of Veterans Affairs," in *Case Studies in Implementing OSI* (tutorial), Interop Incorporated, Mountain View, California, October 1989.

[10] International Organization for Standardization, *Information Processing Systems—Open Systems Interconnection—Basic Reference Model*, ISO 7498, American National Standards Institute, New York, October 1984.

[11] J. Henshall and S. Shaw, *OSI Explained, End-to-End Computer Communication Standards*, West Sussex, England: Ellis-Horwood Limited, pp. 10–20, 1988.

[12] K. G. Knightson, J. Larmouth, and T. Knowles, *Standards for Open Systems Interconnection*, New York: McGraw-Hill Book Company, pp. 12–19, 20–54, 259–301, 1988.

[13] International Organization for Standardization, *Information Processing Systems—Open Systems Interconnection—Manufacturing Message Specification*, ISO 9506, American National Standards Institute, New York, 1989.

[14] International Organization for Standardization, *Information Processing Systems—Open Systems Interconnection—Basic Reference Model, Part 4: Management Framework*, ISO 7498-4 (Draft International Standard), American National Standards Institute, New York, 1988.

[15] International Organization for Standardization, *Information Processing Systems—Open Systems Interconnection—Systems Management Overview*, ISO 10164 (Draft Proposal), American National Standards Institute, New York, 1990.

[16] International Organization for Standardization, *Information Processing Systems—Open Systems Interconnection—Structure of Management Information*,

ISO 10165 (Draft Proposal), American National Standards Institute, New York, 1990.

[17] R. Brennan, K. Thompson, and R. Wilder, "Mapping the X Window onto Open Systems Interconnection Standards," *IEEE Network Magazine*, vol. 4, no. 2, pp. 32–40, May 1991.

[18] *NASCOM Service Gateway Protocol Study*, Computer Sciences Corporation, Greenbelt, Maryland, July 1988.

[19] United States Data Corporation, *FactoryLink Software System*, Publication TP-FLDOS 1/90, Richardson, Texas, January 1990.

[20] P. Norton, *The Norton Utilities Version 5.0 User's Guide*, Peter Norton Computing, Inc., Santa Monica, California, pp. 266–270, 1990.

Table 1. Subsystem lines of code supporting the Mark IVa Monitor and Control System.

Subsystem	Lines of code	Percentage of code for M&C
Monitor and Control	220,000	100
VLBI	107,152	40
Command	23,644	40
Telemetry	60,940	50
Radio Science	53,335	40

Table 2. The functional requirements for Mark IVa Monitor and Control (on the left) and the equivalent MMS protocol services (on the right).

Mark IVa functional requirements	MMS protocol services
Resource allocation	Context Management
Supported with the distribution of Functional Address Tables by Complex Monitor and Control. These tables are used to identify subsystems engaged to support specific station activities. All communications between subsystems are based on addresses in these tables.	Supported with context (connection) management services through the Application Reference Name.
Distribution of support data	File Management
Support data are received by the DMC Subsystem from the Network Operation Control Center and redistributed to individual subsystems prior to a station activity. Subsystems that require support data must have software to accept and process support data blocks distributed by the DMC Subsystem.	Support data files are transmitted to and from the DMC Subsystem using the FTAM services provided in the MMS specification: File Copy, File Obtain, File Open, File Read, File Close.
Operator directives	Variable Access
Operator directives to configure or control the subsystems are entered by an operator at DSCC Monitor and Control (DMC) and processed into messages which are transmitted to the appropriate subsystem across the LAN. Subsystems process directive messages received from DMC to perform the requested functions.	Operator directives for subsystem configuration and control are entered on the DMC Subsystem, converted to the data required by the subsystem, and MMS Variable Write services deliver the data across the network to the subsystem.
Subsystem displays	Variable Access
Subsystem health and performance information is reported to the operator through subsystem displays. Subsystem display data blocks are generated and transmitted across the LAN by each subsystem on request from an operator. The DMC Subsystem is responsible for processing and presenting of subsystem displays for the operator.	Subsystem data for health and performance are obtained using MMS Variable Read services. The data are presented to the operator through graphical displays built by the DMC Subsystem.
Events and Alarms	Variable Access
Subsystem Event and Alarm conditions are generated and reported to DMC. The DMC Subsystem processes Event messages, displays the messages to the operator and logs the messages to an archive file.	Subsystem Event or Alarm conditions are reported to the DMC Subsystem using the MMS Information Report service. MMS Journal services log Events and Alarms to an archive file.
Monitor data	Variable Access
Monitor data are transmitted by each active subsystem to DMC based on negotiated interface agreements. In addition, the DMC Subsystem redistributes monitor data to all of the subsystems based on the negotiated interface agreements.	Monitor data are obtained using the MMS Variable Read services. The redistribution of monitor data is accomplished with MMS Variable Write services.

Table 3. The computer hardware used for the prototype is listed with its Norton Performance Index [20].

Subsystem	Computer	Central processing unit	Clock speed, MHz	Norton Performance Index
Link Monitor and Control (LMC)	BiLink 386	80386	20	23.0
Very Long Baseline Interferometry (VLBI)	Everex 1800	80286	8	7.7
Advanced Power Meter (APM)	IBM AT	80286	8	5.7

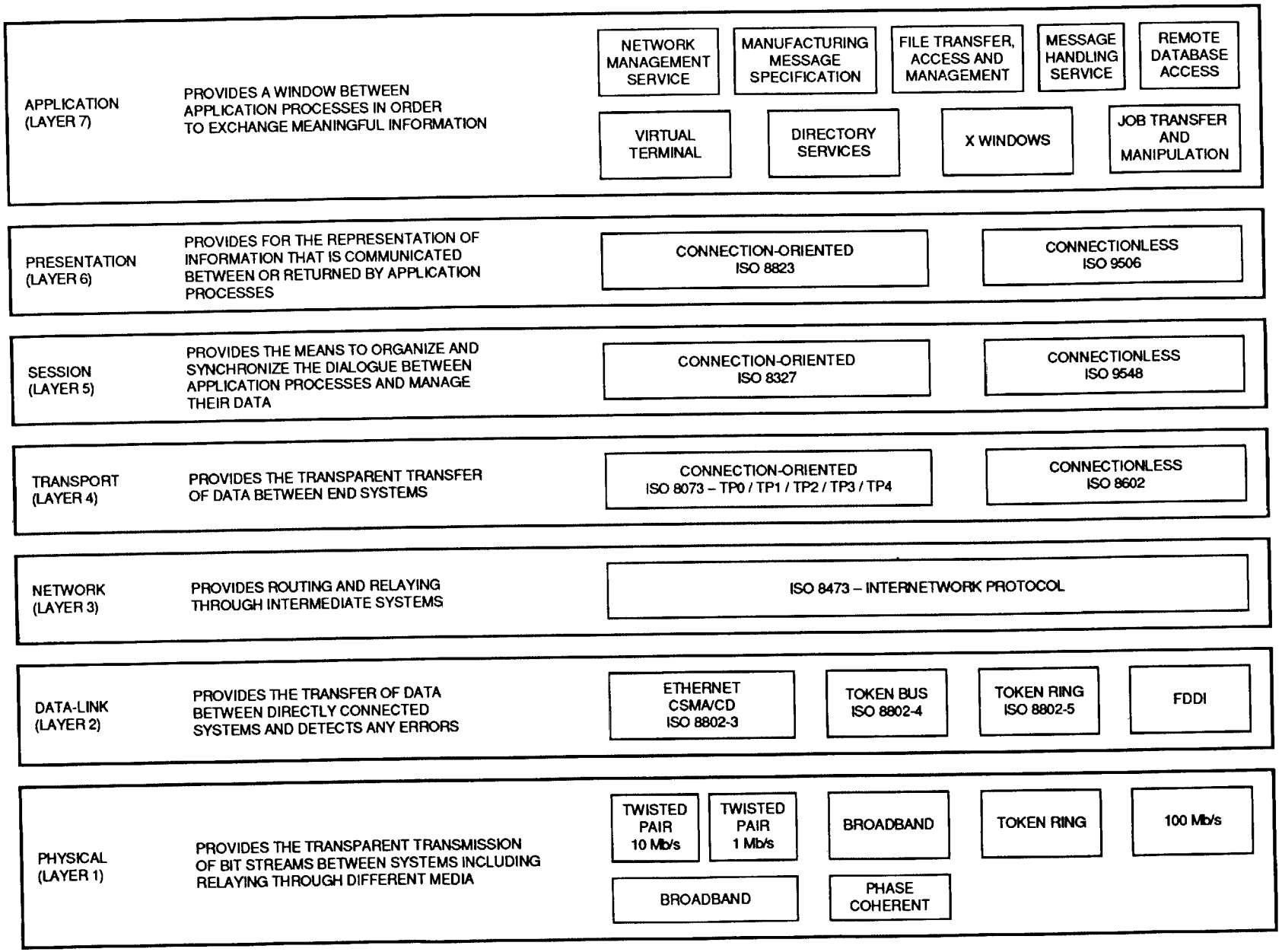


Fig. 1. Open Systems Interconnection architecture provides a wide range of services and protocols to support the growing demands on distributed systems.

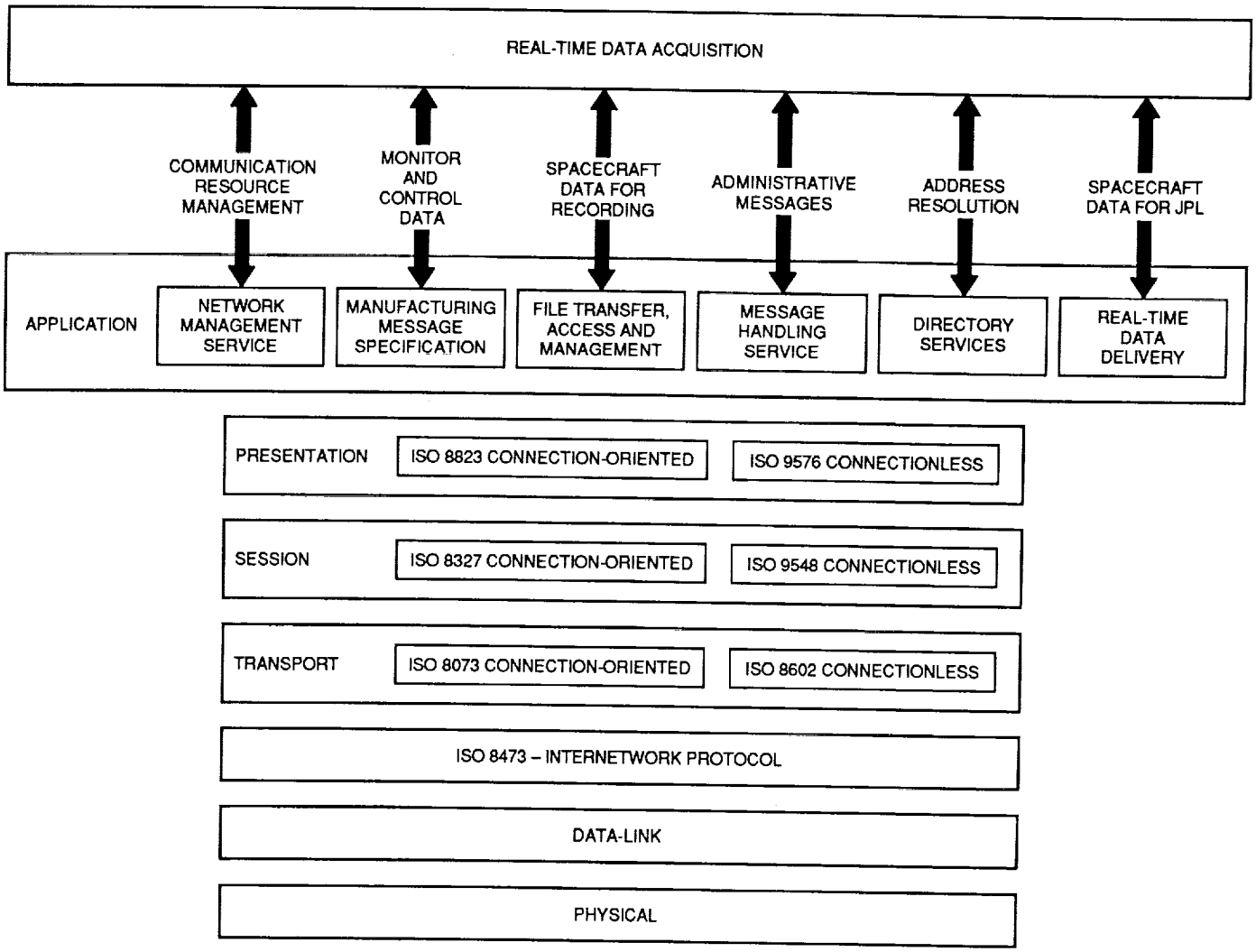


Fig. 2. An Open Systems Interconnection architecture will access communication services through application entities.

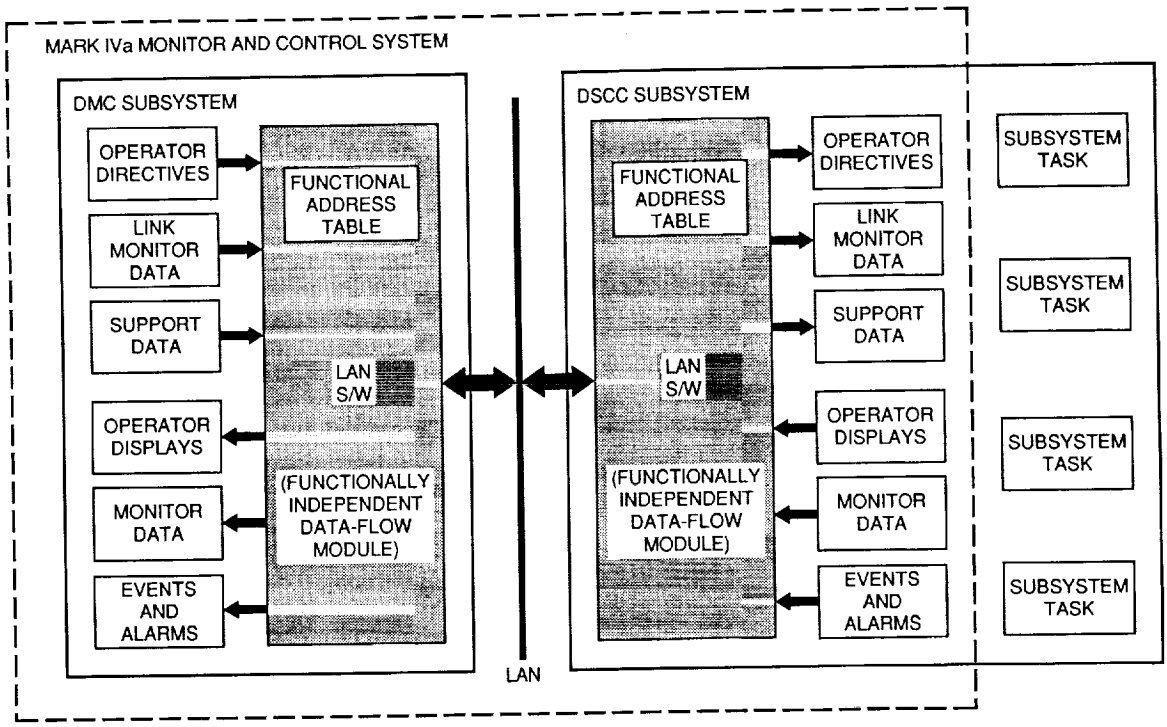


Fig. 3. A schematic representation of the Mark IVa monitor and control architecture, which requires software on each subsystem to support monitor and control.

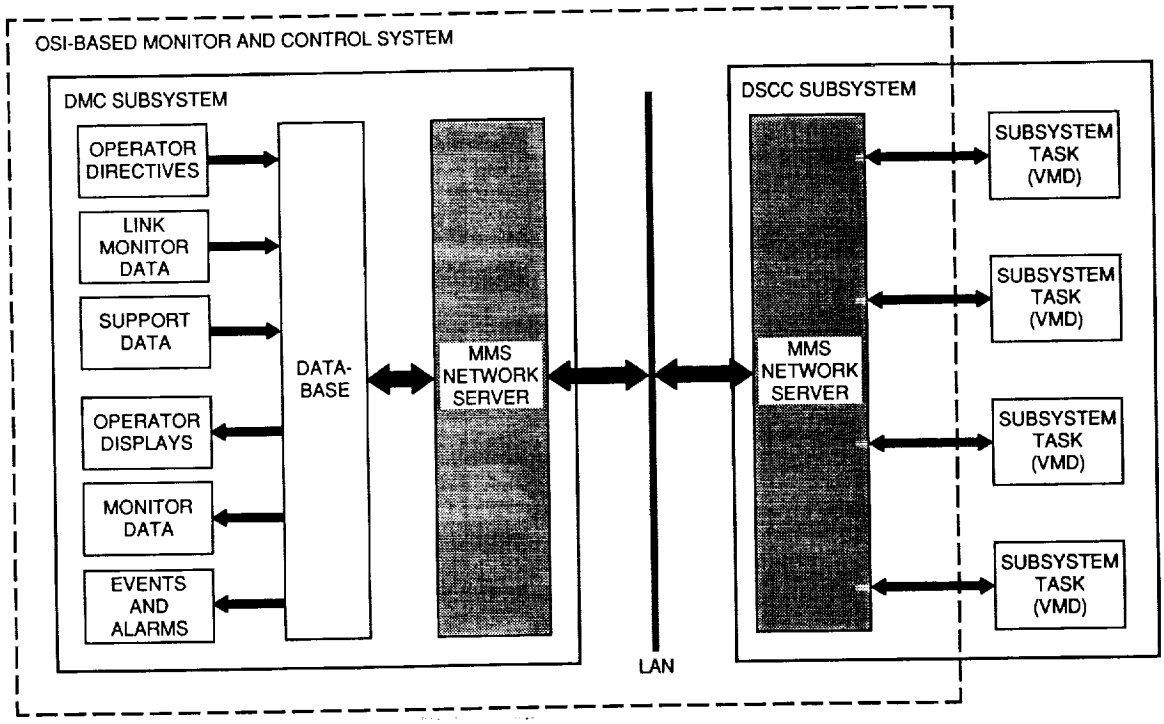


Fig. 4. A schematic representation of an OSI-based monitor and control architecture, which employs MMS to support all monitor and control functions.

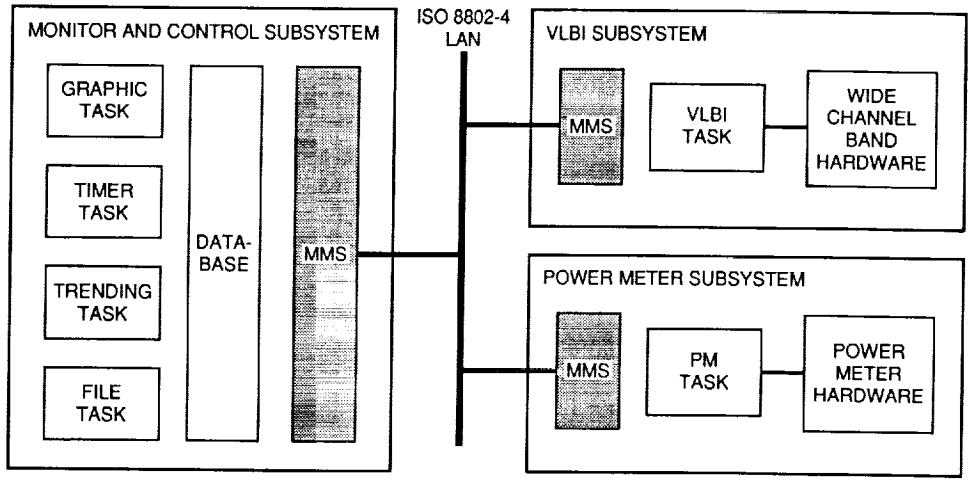


Fig. 5. The OSI monitor and control prototype consists of three subsystems, Link Monitor and Control, VLBI, and the Advanced Power Meter.

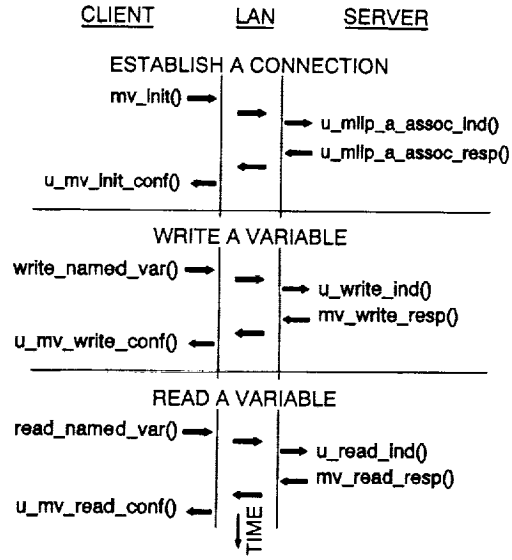


Fig. 6. The protocol exchange between client and server is shown for Context Management and Variable Access services as specified in the MMS Standard.

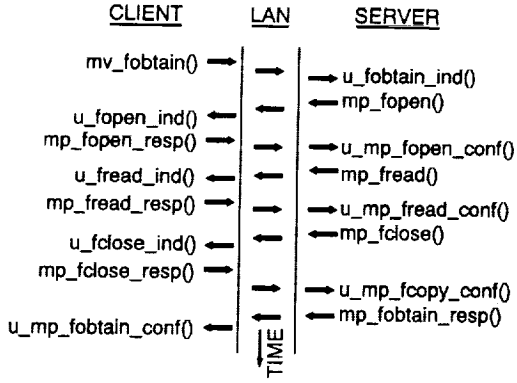


Fig. 7. The File Transfer, Access and Management protocol exchange for the Obtain File service is outlined above as specified in the MMS Standard.

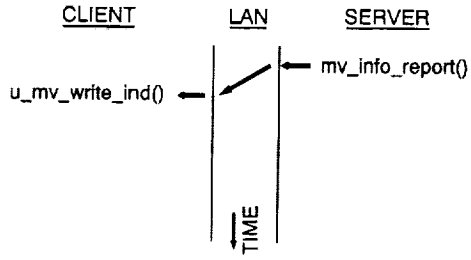


Fig. 8. The protocol exchange for the Information Report service as specified in the MMS Standard.

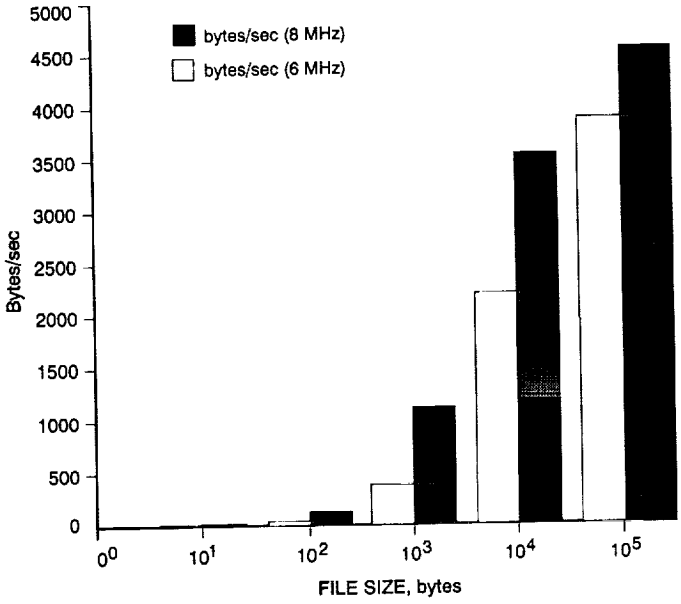


Fig. 9. The performance of the MMS Obtain File service in the prototype environment.

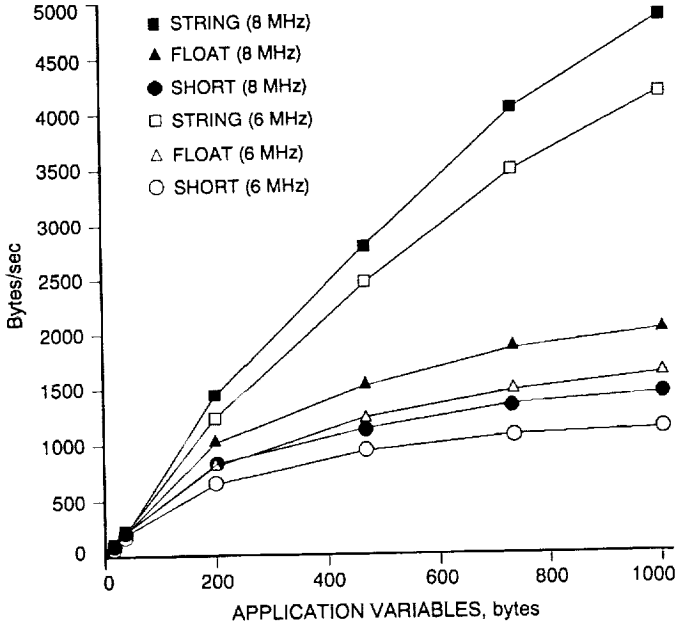


Fig. 10. The performance of the MMS Variable Write service in the prototype environment for application data in bytes/sec.

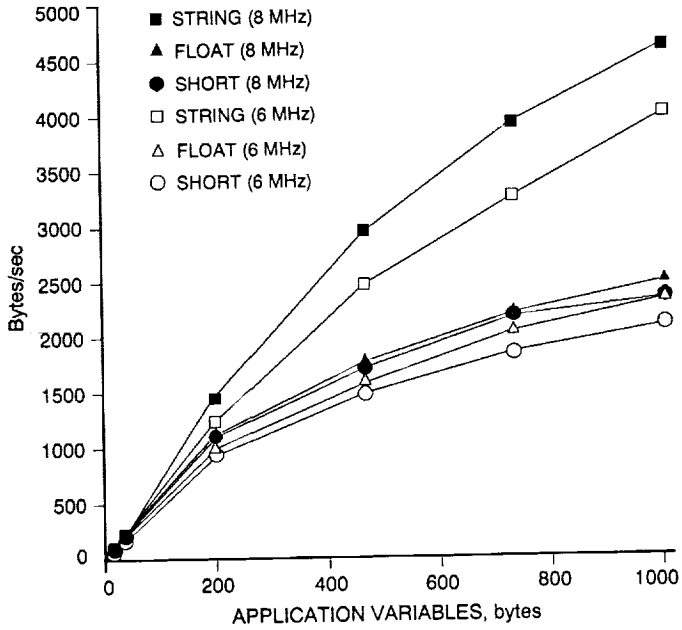


Fig. 11. The performance of the MMS Variable Read service in the prototype environment for application data in bytes/sec.

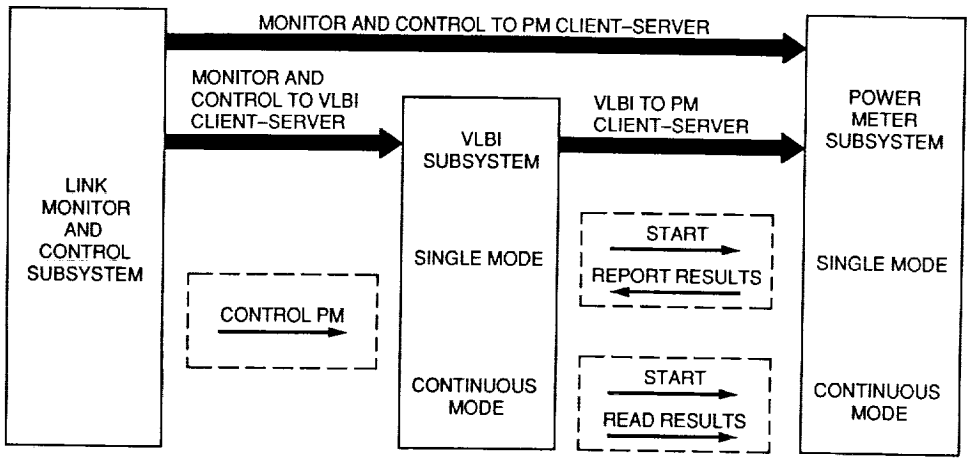


Fig. 12. The client–server relations for automation of the Advanced Power Meter in the prototype Monitor and Control Subsystem.

Appendix A

List of MMS Protocol Services

I. Connection Services

Initiate

Conclude

Cancel

Abort

II. VMD Support Services

Status

Unsolicited Status

GetNameList

Identify

Rename

GetCapabilityList

III. Domain Management Services

InitiateDownSequence

DownloadSegment

TerminateDownSequence

InitiateUploadSequence

UploadSegment

TerminateUploadSequence

RequestDomainDownload

RequestDomainUpload

LoadDomainContent

StoreDomainContent

DeleteDomain

GetDomainAttributes

IV. Program Invocation Management Services

CreateProgramInvocation

DeleteProgramInvocation

Start

Stop

Resume

Reset

Kill

GetProgramInvocationAttributes

V. Variable Access Services

Read

Write

InformationReport

GetVariableAccessAttributes

DefineNamedVariable

DefineScatteredAccess

GetScatteredAccessAttributes

DeleteVariableAccess

DefineNamedVariableList

GetNamedVariableListAttributes

DeleteNamedVariableList

DefineNamedType

GetNamedTypeAttributes

DeleteNamedType

VI. Semaphore Management Services

TakeControl

RelinquishControl

DefineSemaphore

DeleteSemaphore

ReportSemaphoreStatus

ReportPoolSemaphoreStatus

ReportSemaphoreEntryStatus

AttachToSemaphoreModifier

VII. Operator Communication Services

Input

Output

VIII. Event Management Services

DefineEventCondition

DeleteEventCondition

GetEventConditionAttributes

ReportEventConditionStatus

AlterEventConditionMonitoring

TriggerEvent

DefineEventAction

DeleteEventAction

GetEventActionAttributes

ReportEventActionStatus

DefineEventEnrollment

DeleteEventEnrollment

GetEventEnrollmentAttributes

ReportEventEnrollmentStatus

AlterEventEnrollment

EventNotification

AcknowledgeEventEnrollment

GetAlarmSummary

GetAlarmEnrollmentSummary

AttachToEventConditionModifier

IX. Journal Management Services

ReadJournal

WriteJournal

InitializeJournal

ReportJournalStatus

CreateJournal

DeleteJournal

X. File Management Services

ObtainFile

FileOpen

FileRead

FileClose

FileRename

FileDelete

FileDirectory

Appendix B

MMS Message Formulation

The precise formulation of an MMS message is dependent on the service and defined in the standard specification. The construction and interpretation of MMS messages in the prototype were performed by commercial MMS software, and detailed knowledge of the process is not required to employ the protocol. However, a brief examination of how one MMS message is defined and constructed will yield insight into the complexity and versatility of the protocol.

The process for the MMS Variable Write service begins with an MMS protocol data unit (PDU):

```
MMSpdu    ::  {
        confirmed-RequestPDU            [0]   IMPLICIT    Confirmed-RequestPDU,
        confirmed-ResponsePDU           [1]   IMPLICIT    Confirmed-ResponsePDU,
        confirmed-ErrorPDU              [2]   IMPLICIT    Confirmed-ErrorPDU,
        unconfirmed-PDU                 [3]   IMPLICIT    Unconfirmed-PDU,
        rejectPDU                       [4]   IMPLICIT    RejectPDU,
        cancel-Request-PDU              [5]   IMPLICIT    Cancel-RequestPDU,
        cancel-Response-PDU             [6]   IMPLICIT    Cancel-ResponsePDU,
        cancel-ErrorPDU                 [7]   IMPLICIT    Cancel-ErrorPDU,
        initiate-RequestPDU             [8]   IMPLICIT    Initiate-RequestPDU,
        initiate-ResponsePDU            [9]   IMPLICIT    Initiate-ResponsePDU,
        initiate-ErrorPDU              [10]   IMPLICIT    Initiate-ErrorPDU,
        conclude-RequestPDU            [11]   IMPLICIT    Conclude-RequestPDU,
        conclude-ResponsePDU           [12]   IMPLICIT    Conclude-ResponsePDU,
        conclude-ErrorPDU              [13]   IMPLICIT    Conclude-ErrorPDU
}
```

The confirmed-RequestPDU follows with the Write-Request:

```
Confirmed-RequestPDU        ::=  SEQUENCE    {
        invokeID    Unsigned32,
        listOfModifier      SEQUENCE    OF  Modifier    OPTIONAL,
        ConfirmedServiceRequest,
        [79]   CS-Request-Detail        OPTIONAL
        --  shall   not   be   transmitted     if   value    is   NULL
}

Write-Request       ::=  SEQUENCE    {
      variableAccessSpecification              VariableAccessSpecification,
       listOfData                               [0]   IMPLICIT    SEQUENCE    OF  DATA
      }
```

The variable access specification and the data follow:

```
VariableAccessSpecification          :: CHOICE   {
        listOfVariable              [0]  IMPLICIT   SEQUENCE   OF  SEQUENCE   {
                variableSpecification          VariableSpecification,
                alternateAccess                [5]  IMPLICIT   AlternateAccess      OPTIONAL
                },
        variableListName        [1]  ObjectName
        }
```

Each variable name, address and description is included in the PDU:

```
VariableSpecification        ::=  CHOICE   {
    name                                [0]  ObjectName,
    address                             [1]  Address,
    variableDescription                 [2]  IMPLICIT   SEQUENCE   {
        address                    Address,
        typeSpecification          TypeSpecification
        },
    scatteredAccessDescription          [3]  IMPLICIT   ScatteredAccessDescription,
    invalidated                         [4]  IMPLICIT   NULL
    }
```

The name is presented in the form of ObjectName:

```
ObjectName      ::=  CHOICE   {
        vmd-specific            [0]  IMPLICIT   Identifier,
        domain-specific         [1]  IMPLICIT   SEQUENCE   {
                domainID        Identifier,
                itemID          Identifier
                },
        aa-specific             [2]  IMPLICIT   Identifier
        }
```

The address parameter is presented in the form of Address:

```
Address     ::=  CHOICE   {
        numericAddress              [0]  IMPLICIT     Unsigned32,
        symbolicAddress             [1]  IMPLICIT     VisibleString,
        unconstrainedAddress        [2]   IMPLICIT    OCTET   STRING
        }
```

And the type specification takes the following form:

```
TypeSpecification        ::=  CHOICE   {
        typeName         [0]  ObjectName,
        array            [1]  IMPLICIT    SEQUENCE    {
                packed                  [0]  IMPLICIT    BOOLEAN    DEFAULT    FALSE,
                numberOfElement         [1]  IMPLICIT    Unsign32,
                elementType             [2]  TypeSpecification
                },
        structure            [2]  IMPLICIT    SEQUENCE    {
                packed                  [0]  IMPLICIT    BOOLEAN    DEFAULT    FALSE,
                components              [1]  IMPLICIT    SEQUENCE    OF  SEQUENCE    {
                  componentName          [0]  IMPLICIT    Identifier    OPTIONAL,
                componentType           [1]  TypeSpecification
                  }
              },
        -- Simple
        boolean              [3]  IMPLICIT    NULL,              --  BOOLEAN
        bit-string           [4]  IMPLICIT    Integer32,         --  BIT-STRING
        integer              [5]  IMPLICIT    Unsigned8,         --  INTEGER
        unsigned             [6]  IMPLICIT    Unsigned8,         --  UNSIGNED
        floating-point       [7]  IMPLICIT    SEQUENCE    {
                format-width          Unsigned8,       --  number    of  bits   in
                                                       --  fraction    plus   sign
              exponent-width          Unsigned8        --  size   of   exponent    in  bits
              },
        real                 [8]  IMPLICIT    SEQUENCE    {
                base             [0]  IMPLICIT    INTEGER(2|10),
                exponent         [1]  IMPLICIT    INTEGER,       --  max   number    of  octets
                mantissa         [2]  IMPLICIT    INTEGER        --  max   number    of  octets     },
          octet-string                [9]  IMPLICIT    Integer32,       --  OCTET-STRING
          visible-string              [10]  IMPLICIT    Integer32,      --  VISIBLE-STRING
          generalized-time            [11]  IMPLICIT    NULL,           --  GENERALIZEDTIME
          binary-time                 [12]  IMPLICIT    BOOLEAN,        --  BINARY-TIME
          bcd                         [13]  IMPLICIT    Unsigned8       --  BCD
      }
```

Finally, the data are specified:

```
Data    ::=  CHOICE   {
     --  context    tag  0  is  reserved    for  AccessResult
     array                    [1]    IMPLICIT    SEQUENCE    OF  Data,
     structure                [2]    IMPLICIT    SEQUENCE    OF  Data,
     boolean                  [3]    IMPLICIT    BOOLEAN,
     bit-string               [4]    IMPLICIT    BIT  STRING,
     integer                  [5]    IMPLICIT    INTEGER,
     unsigned                 [6]    IMPLICIT    INTEGER,
     array                    [7]    IMPLICIT    FloatingPoint,
     real                     [8]    IMPLICIT    REAL,
     octet-string             [9]    IMPLICIT    OCTET   STRING,
     visible-string           [10]   IMPLICIT    VisibleString,
     generalized-time         [11]   IMPLICIT    GeneralizedTime,
     binary-time              [12]   IMPLICIT    TimeOfDay,
     bcd                      [13]   IMPLICIT    INTEGER,
     booleanArray             [14]   IMPLICIT    BIT  STRING
     }
```

Once formulated into a PDU, the message is encoded in Abstract Syntax Notation One (ASN.1), then passed to the presentation layer and down through the lower layers of the protocol stack. At each step along the way, a protocol data unit is added by each layer until the entire message is formulated and transmitted on the physical network.

DSS-13 26-Meter Antenna Upgraded Radiometer System

C. T. Stelzried
TDA Technology Development

L. Skjerve
Tracking Systems and Applications Section

G. Bury
TDA Mission Support and DSN Opeations

The DSS-13 26-m antenna radiometer system has been upgraded with an IBM-compatible computer-controlled configuration with improved supporting hardware and software. Software has been generated to analyze results and correct for antenna mispointing, tropospheric loss, and other observing errors. This total power radiometer configuration provides a prototype for the new DSS-13 34-m antenna. The radiometer system is described in terms of the theory, instrumentation hardware, computer configuration, and operational features and performance. The system is used to obtain antenna efficiency and pointing model data and is useful for radio source calibrations required for radio astronomy. Some recent results are given.

I. Introduction

The highly productive DSS-13 26-m antenna radiometer system constructed by P. D. Batelaan et al. [1] more than 20 years ago became increasingly difficult to maintain due to a lack of replacement parts. This system has been replaced with an IBM-compatible computer-controlled configuration and improved supporting hardware and software. New calibration techniques provide increased measurement accuracy. This configuration is a prototype for the new DSS-13 34-m antenna radiometer system. The radiometer system improvements are described in terms of the theory, instrumentation hardware, computer configuration, and operational features and performance. Noise temperature measurement errors in a microwave radiometer system due to receiver nonlinearities are small when a well-designed system is maintained and operated within the design range of amplifier signal levels. Nevertheless, it is necessary to quantify residual measurement errors and provide corrections when precision measurements are desired. Comparisons of the measured system noise temperature with different input signal levels obtained by switching between the antenna and ambient load are used to verify system linearity and to correct for system nonlinearity [2].

Radiometer development from 1987 to 1992 for the DSS-13 26-m antenna at both S-band (2.295 GHz) and

X-band (8.420 GHz)[1,2,3,4] includes calibration methods and performance verification of the radiometer as used in the total power (TP) mode of operation using a single noise diode (ND). The radiometer system is useful for antenna efficiency and pointing offset measurements and radio source calibrations required for radio astronomy applications.

With proper receiver-amplifier signal levels and a power-meter detector, the linearity correction is very small. The power-meter output reading is highly linear with input power level, especially as compared with the square-law diode detectors designed in the 1970s. Using an ND at the receiver front end for determining system linearity is very useful for station equipment setup and performance monitoring. The measurements are all made from the control room for the operator's convenience. With good linearity, only verification but not correction for nonlinearity is necessary. This is usually not the case with presently available diode detectors, which, for accurate results, require correction for nonlinearity.

II. Total Power Radiometer System

Figure 1 is the DSS-13 26-m antenna block diagram for the antenna and control room; it shows the radiometer in the TP configuration. Simultaneous S- and X-band low-noise amplifiers are available with matching IF distribution amplifiers and power-meter detectors. The present two-channel radiometer system configuration is switchable for two S-band, two X-band, or one S-band and one X-band TP channels. The IBM PC system and microwave interface have been described previously.[5] Radiometer calibration is obtained by switching the low-noise amplifier input into the microwave ambient termination and recording the power-meter levels with the ND on or off.

Figure 2 shows measurements of system noise temperature on the antenna (T_{op}) using the DSS-13 26-m antenna S-band system in the radiometer TP configuration. The left side of Fig. 2 compares the uncorrected results of the diode detector with those of the power-meter. The right side compares the same data after corrections were applied for system nonlinearity. The agreement in T_{op} after correction is consistent with the nonlinear performance of the diode detector and the capability of the calibration technique to correct for nonlinearity. The power-meter detector is best used with strong and medium strength sources where accuracy is more important than minimum measurement resolution. The diode detector may be optimum for measurement resolution with some loss of accuracy associated with higher nonlinearity errors. Further performance testing will be required to compare the various combinations of S- and X-band amplifiers, TP and noise-adding radiometer (NAR) modes of operation, and power-meter and diode detectors. This comparison is needed to optimize the configuration for a particular measurement requirement. However, for most applications, the performance of the TP mode with a power-meter detector is satisfactory and justified by the configuration's simplicity.

The linearity correction factor for a power-meter configuration with proper amplifier level settings in the receiver system is typically less than 1 percent. For a well-designed system using the power-meter detector, the correction is unnecessary for most applications.

An example follows of the TP radiometer stability performance at the present stage of development. For December 14, 1989, data taken during radio source boresight observations, a measurement resolution of 0.006 K was obtained.[6] The radiometer parameters for these observations were T_{op} = 27.5 K and B (bandwidth) = 19.2 Mhz, with a measurement time of 5 sec; these parameters resulted in a theoretical measurement resolution of 0.003 K. Therefore, such parameters as computer dead time, system gain instability, changing tropospheric effects, and radio-frequency interference reduce radiometer performance by about a factor of 2 for this example.

Future increased computer speed is expected to improve radiometer resolution performance by the reduction of dead time.

III. Radiometer Calibration Sequences

In addition to preobserving calibrations for the radiometer, shorter calibrations using the same data sequence as the precalibrations are performed during the

[1] C. T. Stelzried, *DSS-13 Radiometer System Status and Performance*, JPL D-9291 (internal document), Jet Propulsion Laboratory, Pasadena, California, January 1990.

[2] C. T. Stelzried, *Microwave Radiometers*, JPL D-9295 (internal document), Jet Propulsion Laboratory, Pasadena, California, August 1990.

[3] L. J. Skjerve, *Preliminary Documentation for DSS-13 Radiometer Program*, JPL D-9292 (internal document), Jet Propulsion Laboratory, Pasadena, California, January 1990.

[4] G. Bury, *DSS-13 Block Diagram*, JPL D-9300 (internal document), Jet Propulsion Laboratory, Pasadena, California, January 1990.

[5] L. J. Skjerve, *DSS-13 26-m Antenna Radiometer Upgrade*, JPL D-9525 (internal document), Jet Propulsion Laboratory, Pasadena, California, March 1992.

[6] Stelzreid, *DSS-13 Radiometer System Status and Performance*, loc. cit.

observing session.[7] These calibrations are called minicals and are used to correct for the amplifier gain changes associated with TP operation and in effect provide a slow Dicke-like switching radiometer mode of operation.

IV. Antenna Efficiency

Antenna efficiency in the DSN is defined with the atmosphere removed. The 5-point antenna boresight calibration data method[8] is used to obtain radio source noise temperatures corrected for antenna mispointing.

The radio source noise temperatures for a 100-percent efficient antenna are evaluated from the source flux and angular size data.[9] The ratio of measured to known radio source noise temperature defines antenna efficiency. As a function of elevation angle, EL, the antenna efficiency is represented by a second-order fit,

$$EFF = C0 + C1 \times EL + C2 \times EL^2 \qquad (1)$$

For the DSS-13 26-m antenna, the 1991 results are given by

$$\left.\begin{aligned} C0 &= 0.56398, \\ C1 &= 0.0003018, \\ C2 &= -0.00000417 \end{aligned}\right\} \text{S-band}$$

$$\left.\begin{aligned} C0 &= 0.42120, \\ C1 &= 0.0006582, \\ C2 &= -0.00001867 \end{aligned}\right\} \text{X-band}$$

for 2.295 GHz (S-band) and 8.420 GHz (X-band), respectively. These represent maximum efficiencies of 0.569 at a 36.2-deg elevation angle and 0.427 at a 17.6-deg elevation angle for the S- and X-bands, respectively. Determination of these coefficients representing the change in efficiency with elevation angle is required to correct for the difference in elevation angle between the calibrators and unknown radio sources.

As discussed, it is necessary to correct for the radio source angular size relative to the antenna beamwidth. For precise calibrations, a further complication arises if the antenna beamwidth is a function of the elevation angle, as shown in Fig. 3.

V. Radio Source Calibrations

Antenna efficiency measurements depend on assumed known radio source fluxes. The radiometer system is also used to measure unknown sources, usually relative to known sources.

An observing program is currently underway to monitor S- and X-band flux of the radio source 1830-211.[10] Figure 4 shows a station observation log for 1991 DOY 352 data appropriate for 5-point on–off radio source data. The calibration data and observing sequence are computer controlled. This control includes antenna pointing, microwave switch control, and data collection. Data are alternately collected for the 1830-211 unknown source and calibrator known sources. A BASIC computer program compiled for use with IBM-compatible office computers and designated TPANxxxx (xxxx is currently version 0128)[11] is used to analyze the data. Figure 5 shows a printout of data file 06 for calibrator source 3C274 for DOY 352. Two data sequences, passes 1 and 2 for both X-band (IF1) and S-band (IF2), show the results of the analyzed observing data of the radio source temperature corrected for pointing offsets, atmospheric attenuation, antenna gain change with elevation, and source size.

The S- and X-band data are taken simultaneously with the radiometer system. The antenna pointing angular offsets for the 5-point data sequence are optimized for X-band, which has the narrowest beamwidth. The 5-point data sequence consists of an on-source measurement and off-source measurements with antenna angular offsets of ±0.5 X-band beamwidths and ±2.5 S-band beamwidths. The pointing offset is determined from the X-band data. The efficiency correction for the antenna pointing offset is applied directly for X-band and scaled by beamwidth for S-band. The S- and X-band beams are assumed to be coaligned.

Figure 6 shows the radiometer linearity and gain change as obtained from the minicals. The radiometer gain constants are updated after each minical, as the radiometer is presently operated. The gain correction is assumed to be

[7] Ibid.

[8] Ibid., Appendix C.

[9] M. Klein and A. Freiley, *DSN Radio Source List for Antenna Calibration*, JPL D-3801, Rev. B (internal document), Jet Propulsion Laboratory, Pasadena, California, September 25, 1987.

[10] M. Klein, personal communication, Space Physics and Astrophysics Section, Jet Propulsion Laboratory, Pasadena, California, November 1991.

[11] Skjerve, loc. cit.

the difference between the most recent minical and that computed from the curve fit at the time of the observation. Figure 7 shows a summary of the observations for X-band (IF1 = 12-MHz bandwidth). Some of the bad data are removed automatically by the computer (according to default criteria, such as pointing offsets, which result in greater than 15-percent correction to the source temperature) and further manually by the experimenter during postprocessing. For file 06, pass 1, X-band (IF1), the corrections modify the uncorrected on–off source temperature for 3C274 from 3.701 to 4.105 K. Figures 8 and 9 show the S- and X-band corrected fluxes measured during 1991 using this analysis. The standard deviations of the data relative to the straight-line fit are 0.28 and 0.20 flux units for the S- and X-bands, respectively.

Greater numbers of observed incidents of radio interference at S-band compared with those at X-band at DSS 13 are assumed to account for the higher S-band standard deviations. The high-electron mobility transfer (HEMT) amplifiers have gigahertz bandwidths with no filtering on their input. S- and X-band filtering at the input to the low-noise amplifiers will be provided in the future to reduce this effect.

Computer program TPAN0128 generates data files for all the raw and computed data. These files are further analyzed by the experimenter to provide the final scientific result.

VI. Conclusion

The TP radiometer mode of operation with the power-meter detector provides excellent measurement resolution and accuracy with a simple configuration. The TP mode utilizes modern stable amplifiers. Periodic radiometer calibrations during the observing period are used to remove long-term gain drifts. These minicals are used during post-data analysis for correcting the results. The NAR configuration is more complicated but should improve performance with amplifier gain changes.

The DSS-13 26-m antenna TP radiometer configuration continues to be upgraded. This system has been demonstrated during 1991 and continues operating in 1992 for engineering and science observations. Sample results obtained from the analysis program TPAN0128 are shown in Figs. 5, 7, 8, and 9.

Acknowledgments

Numerous people supported this activity and performed the measurements. DSS-13 personnel including J. Garnica, R. Rees, G. Farner, and R. Littlefair upgraded, modified, and documented the equipment and performed many hours of data taking and manipulation. C. Goodson and A. Price provided DSS-13 management support. M. Klein early on suggested collaboration between the radiometer development and data verification. This has been very beneficial to both activities and resulted in a universal radiometer system available for both station engineering activities, such as antenna boresighting and efficiency calibrations, and outside science application users.

References

[1] P. D. Batelaan, R. M. Goldstein, and C. T. Stelzried, "A Noise Adding Radiometer for Use in the DSN," *Space Programs Summary 37-65*, vol. 2, Jet Propulsion Laboratory, Pasadena, California, pp. 66–69, September 30, 1970.

[2] C. T. Stelzried, "Non-Linearity in Measurement Systems: Evaluation Method and Application to Microwave Radiometers," *TDA Progress Report 42-91*, vol. July–September 1987, Jet Propulsion Laboratory, Pasadena, California, pp. 57–66, November 15, 1987.

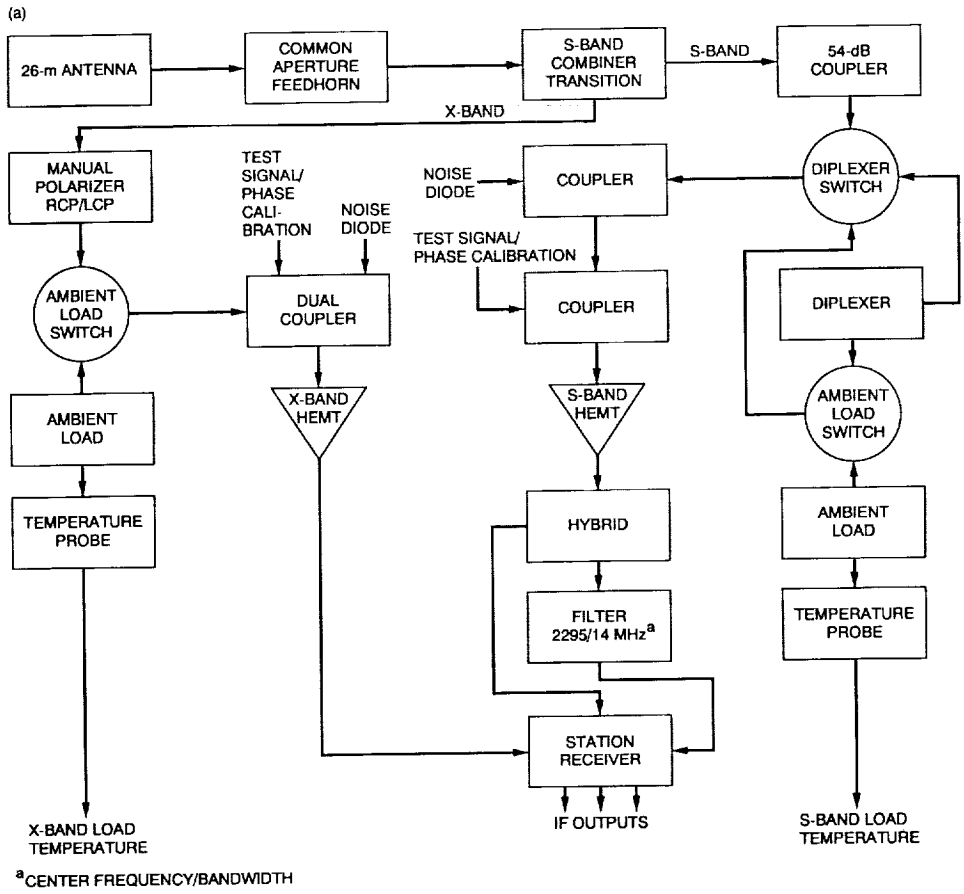


Fig. 1. DSS-13 26-m antenna system block diagram: (a) antenna feed cone assembly and (b) control room assembly.

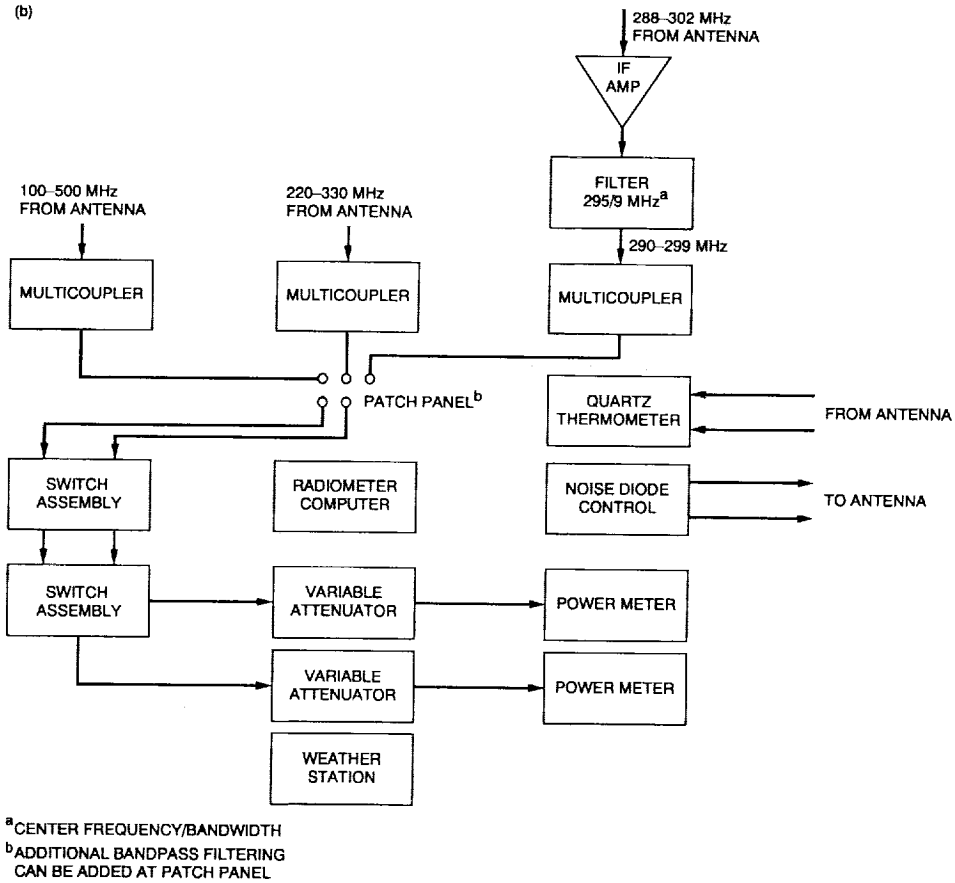

(b)
288–302 MHz
FROM ANTENNA
IF
AMP
FILTER
295/9 MHz[a]
290–299 MHz
100–500 MHz
FROM ANTENNA
220–330 MHz
FROM ANTENNA
MULTICOUPLER
MULTICOUPLER
MULTICOUPLER
PATCH PANEL[b]
QUARTZ
THERMOMETER
FROM ANTENNA
SWITCH
ASSEMBLY
RADIOMETER
COMPUTER
NOISE DIODE
CONTROL
TO ANTENNA
SWITCH
ASSEMBLY
VARIABLE
ATTENUATOR
POWER METER
VARIABLE
ATTENUATOR
POWER METER
WEATHER
STATION
[a]CENTER FREQUENCY/BANDWIDTH
[b]ADDITIONAL BANDPASS FILTERING
CAN BE ADDED AT PATCH PANEL

Fig. 1 (contd).

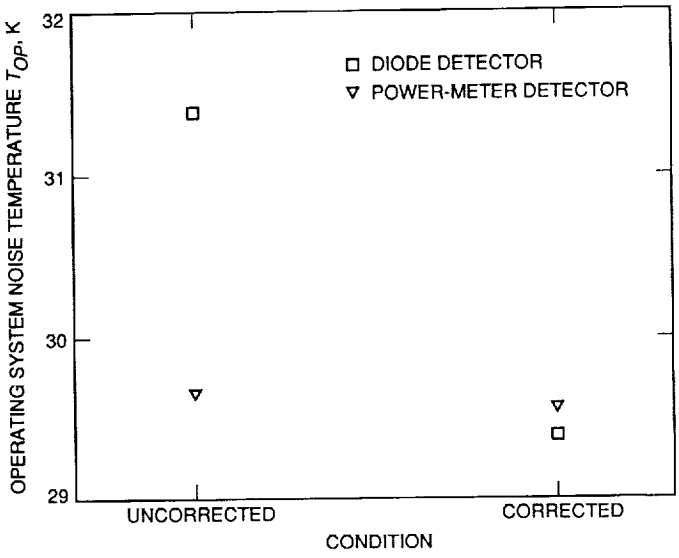


Fig. 2. Comparison of DSS-13 26-m 1988 DOY 271 S-band total power radiometer with power-meter and diode detector configuration noise temperature measurements uncorrected and corrected for nonlinearity.

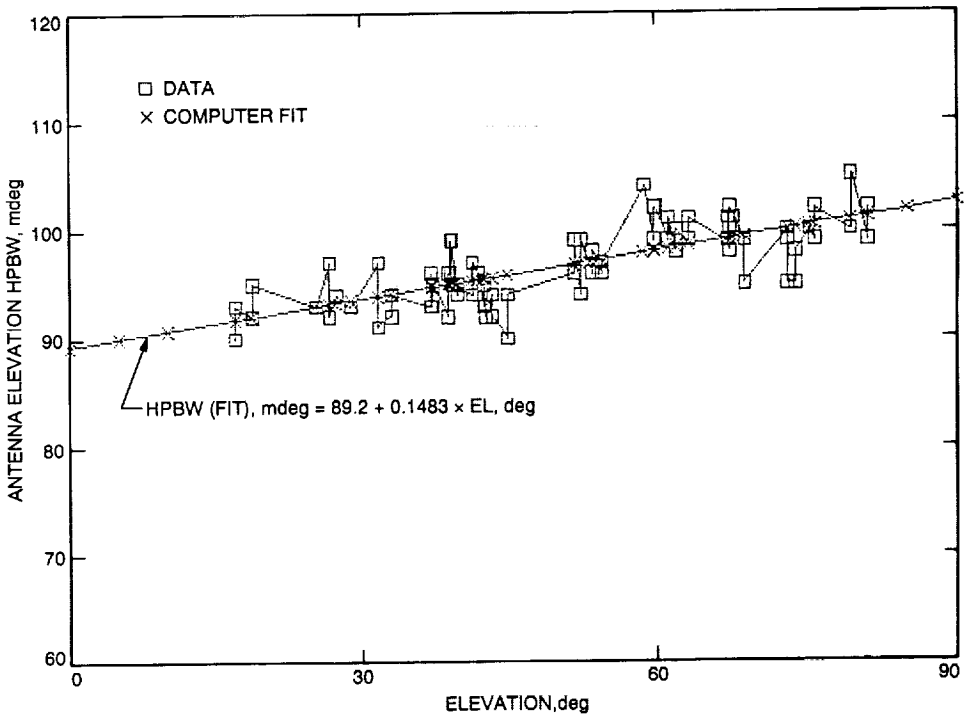


Fig. 3. DSS-13 26-m antenna elevation X-band half-power beamwidth (HPBW) versus elevation angle, December 14, 1989.

DSS13 OBSERVATION LOG

DATE 18 DEC 91 DOY 352 OPERATOR G. Bury PAGE 1 OF 3

SOURCE 1831-21 TYPE SCHEDULE X/S HA/DEC x AZ/EL

CHAN 1 RF FREQ 8437-8451 MHZ IF FREQ 337-351 MHZ (RCP) LCP DET PM

CHAN 2 RF FREQ 2290-2299 MHZ IF FREQ 290-299 MHZ (RCP) LCP DET PM

RF SOURCES OFF? Yes ATTENUATION: IF1 7 dBm /IF2 5 dBm FOCUS 258

FILE#	SOURCE	UT	AZ/EL	DIRECT OFFSETS AZ/EL	UT START TIMES OF SCHEDULED EVENTS				CONE TEMP#1/#2
1X91352	DR21	1624	052 / 20.5	115 / -0.44	1627	1632	1637	1642	18.29 / 18.0

BORESIGHT: BLIND? x MANUAL? ___ AUTO? ___ WIND? 4 MPH SKY? PARTLY CLOUDY MINICAL? x PRECAL? ___

comments: RFI BP 894.6 TC 8.02 RH 77.44

FILE#	SOURCE	UT	AZ/EL	DIRECT OFFSETS AZ/EL	UT START TIMES OF SCHEDULED EVENTS				CONE TEMP#1/#2
2X91352	DR21	1651	055/ 26.9	116/ -043	1709	1714	1719	1724	18.78 / 18.5

BORESIGHT: BLIND? x MANUAL? ___ AUTO? ___ WIND? 13 MPH SKY? PARTLY CLOUDY MINICAL? x PRECAL? ___

comments: RFI

FILE#	SOURCE	UT	AZ/EL	DIRECT OFFSETS AZ/EL	UT START TIMES OF SCHEDULED EVENTS				CONE TEMP#1/#2
3X91352	DR21	1737	58 / 32	116 / -042	1739	1744	1749	1754	1933 / 18

BORESIGHT: BLIND? x MANUAL? ___ AUTO? ___ WIND? 10.8 MPH SKY? PARTLY CLOUDY MINICAL? x PRECAL? ___

comments: RFI

FILE#	SOURCE	UT	AZ/EL	DIRECT OFFSETS AZ/EL	UT START TIMES OF SCHEDULED EVENTS				CONE TEMP#1/#2
4X91352	1831-21	1803	141 / 23	109 / .053	1804	1809	1814	1819	19.52 / 19

BORESIGHT: BLIND? x MANUAL? ___ AUTO? ___ WIND? 8.7 MPH SKY? PARTLY CLOUDY MINICAL? x PRECAL? ___

comments: RFI Edit UT 1819, 1822 Hung up

FILE#	SOURCE	UT	AZ/EL	DIRECT OFFSETS AZ/EL	UT START TIMES OF SCHEDULED EVENTS				CONE TEMP#1/#2
5X91352	1831-21	1824	146 / 25.9	108 / -052	1828	1833	1838	1843	19.25 / 19

BORESIGHT: BLIND? x MANUAL? ___ AUTO? ___ WIND? 9.7 MPH SKY? ___ MINICAL? x PRECAL? ___

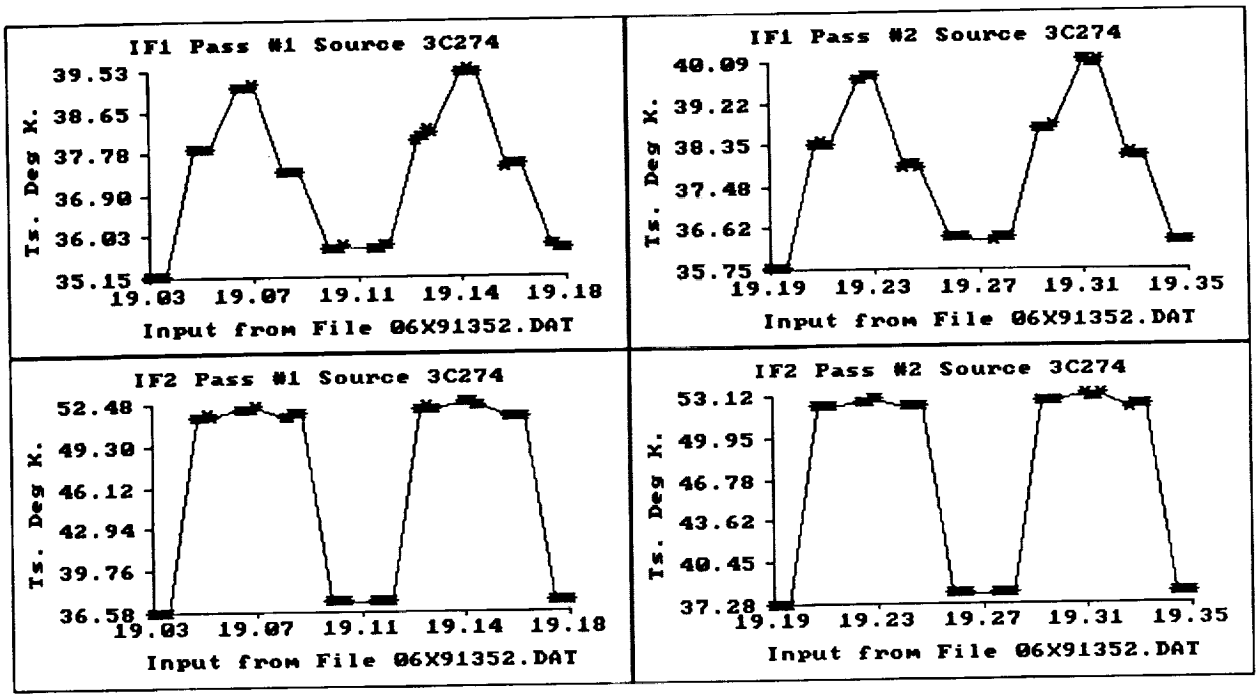
comments: RFI

FILE#	SOURCE	UT	AZ/EL	DIRECT OFFSETS AZ/EL	UT START TIMES OF SCHEDULED EVENTS				CONE TEMP#1/#2
6X91352	3C274	1902	267 / 24.8	105 / -029	1901	1906	1911	1916	18.99 / 19

BORESIGHT: BLIND? x MANUAL? ___ AUTO? ___ WIND? 8.9 MPH SKY? PARTLY CLOUDY MINICAL? x PRECAL? ___

comments: RFI

Fig. 4. DSS-13 26-m antenna observation log for 1991 DOY 352 (transcribed).



Total Power Radiometer Analysis Program Dated 28 January 1992

Source = 3C274	R.A. = 12:30:25.7	Dec. = +12:25:57		Fix Az Off	0.105		
IF1 Rx = NBX	IF2 Rx = NBS			Fix El Off	-0.029		
IF1 = 8445.000	IF1 BW = 12.000	IF2 = 2294.500		IF2 BW =	9.000		
Pass 1	MidPass 19:02:58	Az. = 268.910		El. =	23.590		
Pass 2	MidPass 19:15:29	Az. = 270.360		El. =	21.540		
DOY = 352	Temp.(C.) 11.195	Rh. = 82.345		Barom. =	894.050		

Input File = 06X91352.DAT

	IF1Pass1		IF1Pass2		IF2Pass1		IF2Pass2	
	Mean	StDev.	Mean	StDev.	Mean	StDev.	Mean	StDev.
Az.Corr.	-8.661	0.3658	-7.831	0.4789	-11.658	0.7403	-10.510	0.7773
Az.Total	96.339		97.169		93.342		94.490	
El.Corr.	2.276	0.1592	2.610	0.2452	2.499	1.4578	0.536	0.5753
El.Total	-26.724		-26.390		-31.499		-28.464	
HPBW Ha.	93.203	0.7691	89.645	0.8718	367.878	24.6272	368.846	13.1837
HPBW Dec.	89.927	1.5095	89.501	0.7084	334.579	17.0784	346.174	18.3827
System Top	35.617	0.0139	36.221	0.0132	37.122	0.0219	37.836	0.0173
Efficiency	0.429	0.0037	0.419	0.0035	0.561	0.0026	0.560	0.0018
Eff. Corr.	0.465		0.463		0.593		0.593	
Path/Zen. DB	0.108	0.0432	0.118	0.0432	0.079	0.0316	0.086	0.0316
Theo/Measure	0.330		0.354		0.253		0.226	
Perf. Test	2.519		2.376		2.300		1.629	
IsUnCorr.	3.701	0.0325	3.692	0.0310	15.116	0.0695	15.068	0.0496
Corr. Ptg.	1.024		1.021		1.002		1.002	
Corr. El.	1.002		1.001		1.001		1.002	
Corr. Atmos.	1.025		1.027		1.018		1.020	
Corr. Size	1.054		1.054		1.035		1.035	

Fig. 5. DSS-13 26-m antenna 1991 DOY 352 5-point measurement of radio source calibrator 3C274 used for 1830-211 flux determination.

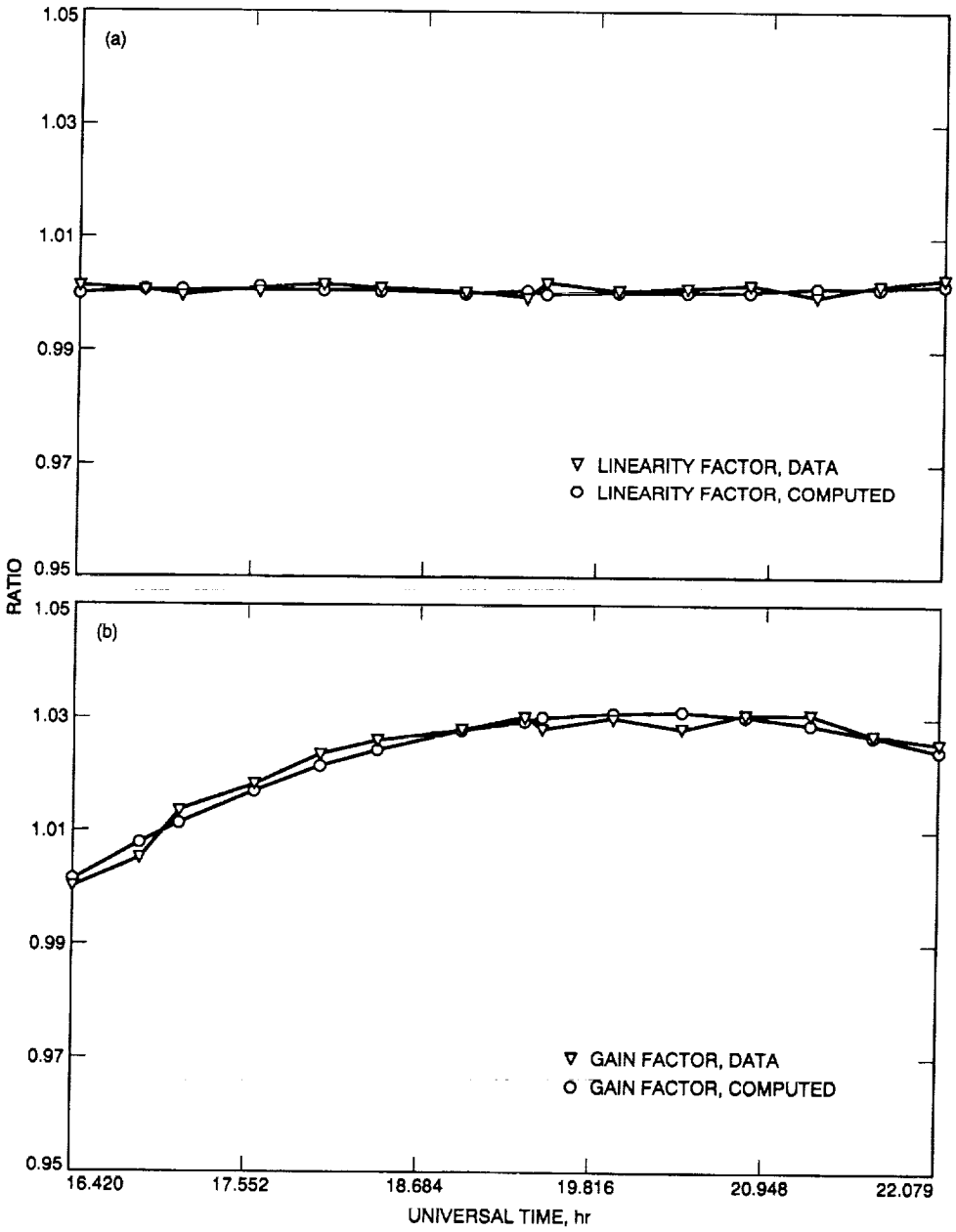


Fig. 6. DSS-13 26-m antenna radiometer X-band calibrations for radio source 1830-211: (a) linearity and (b) gain.

Radiometer Analysis Program Version TPAN0128.EXE Dated 28 January 1991
Summary of Experiment 12-18-1991 (DOY=352) Beginning at 16:26:46
For IF #1 8445 MHz and BandWidth of 12 MHz

RUN	Src.	MidP UT	El	Computed Corrections Ptg.	El.	Atm.	Size	Gain	Computed Ts Values UnCorr.	Corr	Ts100	Flux Meas.	Eff.
1	DR21	16.5	22										
1	DR21	16.7	24										
2	DR21	17.2	29										
2	DR21	17.4	31										
3	DR21	17.7	35										
3	DR21	17.9	36										
4	1831-	0.0	0										
4	1831-	0.0	0										
5	1831-	18.3	27	1.045	1.004	1.022	1.000	1.000	0.895	0.961	2.035	10.58	
5	1831-	18.7	28	1.054	1.005	1.021	1.000	1.001	0.782	0.847	1.793	9.33	
6	3C274	19.1	24	1.024	1.002	1.025	1.054	1.001	3.701	4.105	8.822		.465
6	3C274	19.3	22	1.021	1.001	1.027	1.054	1.001	3.692	4.091	8.822		.464
7	1831-	19.6	33	1.002	1.010	1.019	1.000	1.002	0.821	0.848	1.796	9.34	
7	1831-	19.8	33	1.006	1.010	1.019	1.000	1.002	0.804	0.833	1.765	9.18	
8	3C286	20.1	32	1.084	1.010	1.019	1.000	1.000	0.440	0.491	1.043		.471
8	3C286	20.3	30	1.075	1.007	1.020	1.000	1.000	0.450	0.498	1.043		.477
9	1831-	20.5	34										
9	1831-	20.7	34										
10	1831-	20.9	33	1.004	1.011	1.018	1.000	0.998	0.828	0.854	1.809	9.41	
10	1831-	21.1	33	1.013	1.010	1.019	1.000	0.998	0.804	0.837	1.773	9.22	
11	3C295	21.4	33	1.029	1.011	1.018	1.000	0.998	0.289	0.305	0.639		.477
11	3C295	21.5	32										
12	3C295	21.8	30	1.102	1.006	1.020	1.000	1.000	0.271	0.306	0.639		.479
12	3C295	21.9	28										
13	1831-	22.2	29										
13	1831-	22.4	28										
Averages				1.069	1.008	1.020	1.007	1.000					.472

Curve Fit Data (Gain Corr. Fit from MiniCal CSV Files)

El. Corr Used =	4.2120E-01	6.5820E-04	-1.8670E-05	0.0000E+00		
El. Corr Comp =	3.0331E-01	1.1451E-02	-1.8938E-04	0.0000E+00	Sig	3.8879E-03
Gain Corr Meas =	-8.5898E+08	1.9185E+08	-8.8720E+06	1.3579E+05	Sig	7.4146E+05

Max El. Used for El Corr. = 17.627 Computed = 30.234 Sigma 0.3888E-02
Max Eff. Used = 0.427 Computed = 0.476

X-Band Constants File Written JULY 91
Catalog = c:\analysis\TPSOURCE.ALL dated 25NOV91

Reduced Data for IF1 on Day 352 Data is Written to File F191352E.CSV

Source	Ts Corr. Mean	Ts Corr. Sigma	Flux Meas. Mean	Flux Meas. Sigma	Flux Known	Eff Corr. Mean	Eff Corr. Sigma	Ts100 Meas.	Ts100 Known
1831-21	0.863	0.044	9.509	0.486	45.884	0.465	0.001	1.828	8.822
3C274	4.098	0.007			5.427	0.474	0.003		1.043
3C286	0.494	0.003			3.324	0.478	0.001		0.639
3C295	0.306	0.001							

Radiometer Analysis Program Version TPAN0128.EXE Dated 28 January 1992
Performance Analysis of Experiment

File	Source	IF1P1	Status	IF1P2	Status	IF2P1	Status	IF2P2	Status
01X91352	DR21	14.593	Perf.	13.517	Perf.	%105.129	Perf.	78.635	Perf.
02X91352	DR21	12.421	Perf.	11.005	Perf.	76.996	Perf.	88.499	Perf.
03X91352	DR21	9.956	Perf.	10.881	Perf.	87.246	Perf.	84.108	Perf.
04X91352	1831-21	0.000	NOffs	0.000	NOffs	0.000	NOffs	0.000	NOffs
05X91352	1831-21	2.611	Good	1.683	Good	0.000	Oper.	11.685	Perf.
06X91352	3C274	2.512	Good	1.376	Good	2.300	Good	1.629	Good
07X91352	1831-21	2.265	Good	1.752	Good	4.128	Good	28.426	Perf.
08X91352	3C286	4.938	Good	1.752	Good	7.393	Perf.	1.571	Good
09X91352	1831-21	6.328	Ptg.	4.086	Ptg.	15.854	Perf.	11.015	Perf.
10X91352	1831-21	3.560	Good	5.148	Good	11.723	Perf.	6.531	Good
11X91352	3C295	2.153	Good	3.662	Ptg.	0.556	Good	2.857	Good
12X91352	3C295	1.202	Good	0.000	Oper.	2.041	Good	0.000	Oper.
13X91352	1831-21	0.000	Oper.	0.000	Oper.	42.551	Perf.	21.267	Perf.

Good = No Errors
Perf = Bad Performance (Theo/Actual)
Ptg = Eliminated for Bad Pointing Corr.
Oper = Operator Requested No Analysis of this Particular Run
Math = Error in Computing
NOffs = Insufficient Data
Source = No Source Name in File

Fig. 7. DSS-13 26-m antenna X-band (IF1) data summary of 1991 DOY 352 1830-211 radio source calibrations.

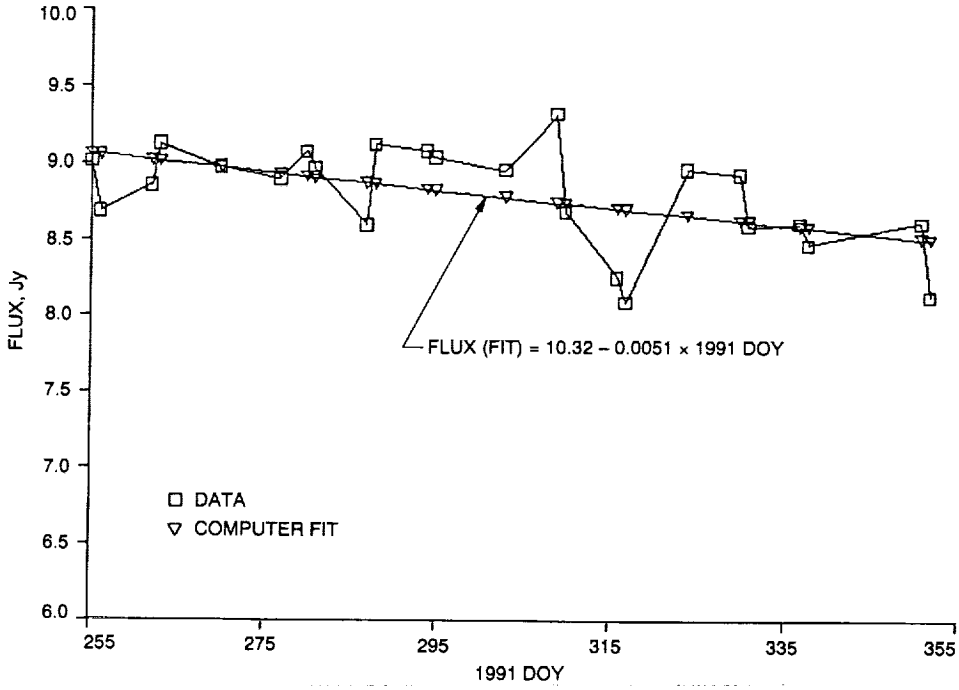


Fig. 8. 1830-211 radio source S-band flux as measured at DSS-13 26-m antenna during 1991.

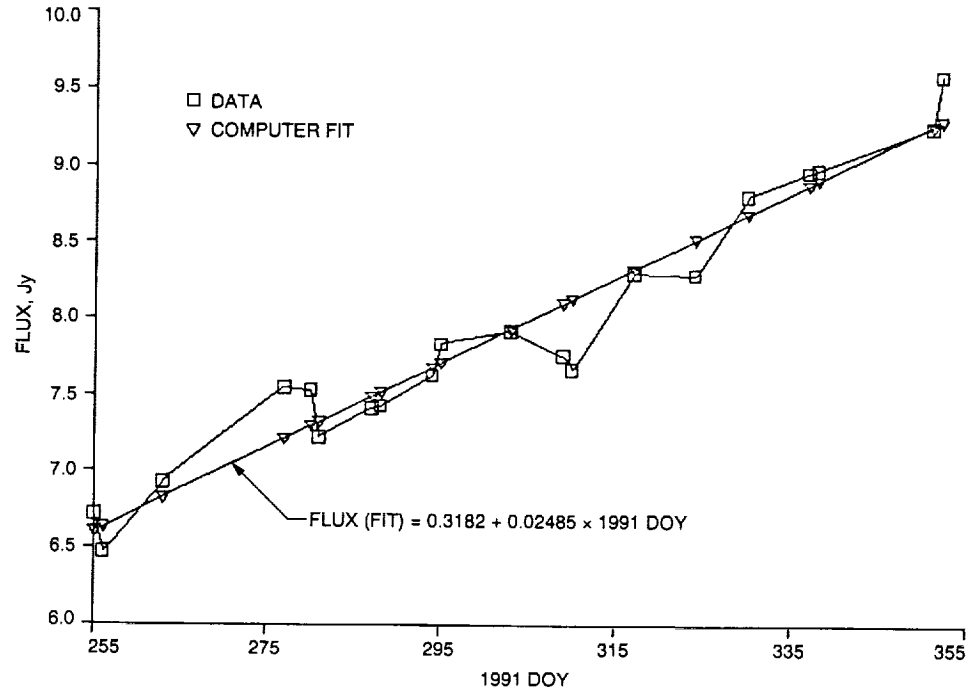


Fig. 9. 1830-211 radio source X-band flux as measured at DSS-13 26-m antenna during 1991.

Sequential Design of a Linear Quadratic Controller for the Deep Space Network Antennas

W. Gawronski
Ground Antennas and Facilities Engineering Section

A new linear quadratic controller design procedure is proposed for the NASA/JPL Deep Space Network antennas. The antenna model is divided into a tracking subsystem and a flexible subsystem. Controllers for the flexible and tracking parts are designed separately by adjusting the performance index weights. Ad hoc weights are chosen for the tracking part of the controller and the weights of the flexible part are adjusted. Next, the gains of the tracking part are determined, followed by the flexible controller final tune-up. In addition, the controller for the flexible part is designed separately for each mode; thus the design procedure consists of weight adjustment for small-size subsystems. Since the controller gains are obtained by adjusting the performance index weights, determination of the weight effect on system performance is a crucial task. A method of determining this effect that allows an on-line improvement of the tracking performance is presented in this article. The procedure is illustrated with the control system design for the DSS-13 antenna.

I. Introduction

A linear quadratic (LQ) controller design procedure for the Deep Space Network (DSN) antennas is presented. Alvarez and Nickerson [1] have used the LQ approach for controller design of the DSS-14 antenna. In the Alvarez and Nickerson approach, the gearbox flexible mode was included in the rigid-body model of the antenna. In recently designed antenna structures (such as the DSS-13 antenna), significant flexible deformations are observed during tracking operations. The antenna rate-loop model described in [2] consists of 21 flexible modes up to 10 Hz. Controllers for these antennas should suppress flexible motion while following the tracking command. The method presented in this article allows the design of a controller with a flexible motion suppression capability through sequential adjustment of the weights of the LQ performance index.

An LQ controller is optimal in the sense of minimization of the performance index. The tracking performance requirements are reflected in the definition of the performance index through proper adjustment of weights. Indeed, the closed-loop system performance depends heavily on the choice of the weighting matrix, as illustrated with the DSS-13 antenna in Fig. 1. In case 1, the weight 10 for the integral of the antenna position, the weight 1 for the position itself, and the weight 0 for the flexible modes have been chosen. The antenna performance, characterized in this case by its step response in Fig. 1 (solid line), shows

excessive flexible motion. In case 2, the weights are the same as those in the previous case, but the weights of the flexible modes are now set equal to 0.001. The closed-loop antenna performance in Fig. 1 (dashed line) shows a significant deterioration of the antenna tracking capabilities.

The procedure presented in [1], as well as other procedures frequently used in antenna design [3], separates controller design for the elevation and azimuth drives. This approach, effective for slow and/or rigid antennas, cannot be justified for fast and/or flexible antennas. In the latter case, the flexible properties of the full antenna significantly differ from the properties of the elevation-only or azimuth-only model of the antenna; thus the separate design of controllers for elevation and azimuth drives would result in system instability. For flexible antennas, there is a quasi-separation of the flexible and tracking motions. This property is used to simplify the controller design procedure. A controller for the flexible part is designed first, followed by a controller for the tracking part, with additional corrections of the controller for the flexible part. The design of the controller for the flexible part is of a sequential nature as well: a controller for each mode is designed separately. The design consists of weight adjustment; it is crucial, therefore, to accurately determine the effect of weight on system performance. The analysis of the impact of weight on system performance is presented in this article. It allows on-line improvement of the tracking performance. The procedure is illustrated with the control system design of the DSS-13 antenna.

II. Properties of a Generic DSN Antenna Model

In this section, study of the properties of an open-loop model (called also a rate-loop model) of a generic Deep Space Network antenna is based on the DSS-13 antenna model. This antenna represents the new generation of 34-meter-diameter antennas. Dynamics of these antennas include non-negligible flexible motion [2], which must be taken into account while designing the tracking controller.

The balanced state-space representation (A_p, B_p, C_p) of the DSS-13 antenna is derived in [2]. Its rate command input is denoted $u_p^T = [u_{pe}\ u_{pa}]$, where u_{pe} and u_{pa} are elevation and azimuth rate commands, respectively, and output is denoted $y_p^T = [y_{pe}\ y_{pa}]$, where y_{pe} and y_{pa} are elevation and azimuth angles. The state vector x_p includes integrator states $x_i^T = [x_{ie}\ x_{ia}]^T$ in elevation and azimuth (rate inputs and position outputs indicate the presence of integrators), and flexible coordinates x_f of dimension n_f; thus,

$$x_p^T = [x_i\ \ x_f^T] \tag{1a}$$

The respective state triple is obtained:

$$A_p = \begin{bmatrix} 0 & 0 \\ 0 & A_f \end{bmatrix},\quad B_p = \begin{bmatrix} B_{pt} \\ B_{pf} \end{bmatrix},\quad C_p = [C_{pt}\ C_{pf}] \tag{1b}$$

where 0 denotes a zero matrix of proper dimensions.

The matrix C_p, which describes a relationship between the balanced states of the rate-loop model and the elevation and azimuth angles, is small, typically $\|C_p\| < 10^{-3}$ (here and later $\|\cdot\|$ denotes a Euclidean norm). This means that the outputs of the rate-loop system (position angles of the antenna) are much smaller than its states. This property is used later in the controller design procedure.

For controller design purposes, the position angles of the antenna y_p are required to be the first states in the state-space representation. Thus the state x_p is transformed accordingly, so that the new state is

$$x_{pn}^T = [y_p^T\ x_f^T] \tag{2}$$

Since $y_p = C_p x_p$, one obtains the transformation P such that $x_{pn} = P x_p$, where

$$P = \begin{bmatrix} C_p \\ C_r \end{bmatrix} = \begin{bmatrix} C_{pt} & C_{pf} \\ 0 & I_{nf} \end{bmatrix} \tag{3}$$

and $C_r = [0\ I_{nf}]$ (I_{nf} is an identity matrix of dimension n_f). The new state-space representation (A_a, B_a, C_a) is obtained:

$$(A_a, B_a, C_a) = (PA_pP^{-1}, PB_p, C_pP^{-1}) \tag{4}$$

where

$$A_a = \begin{bmatrix} 0 & C_{pf}A_f \\ 0 & A_f \end{bmatrix},\quad B_a = \begin{bmatrix} B_y \\ B_f \end{bmatrix}$$

$$C_a = [I\ 0],\ B_y = C_pB_p \tag{5}$$

Additionally, for the controller design purposes, the plant is augmented with the state variables $y_i = [y_{ie},\ y_{ia}]^T$ —an integral of the elevation and azimuth position [1,2]. Thus, by defining the state vector x as

$$x = [x_t^T \; x_f^T]^T \tag{6}$$

where $x_t = [y_i^T \; y_p^T]^T$, one obtains finally the rate-loop representation (A, B, C):

$$A = \begin{bmatrix} A_t & A_{tf} \\ 0 & A_f \end{bmatrix}, \; B = \begin{bmatrix} B_t \\ B_f \end{bmatrix}, \; C = [C_t \; 0] \tag{7a}$$

where

$$A_t = \begin{bmatrix} 0_2 & I_2 \\ 0_2 & 0_2 \end{bmatrix}, \; A_{tf} = \begin{bmatrix} 0 \\ C_{pf} A_f \end{bmatrix}$$

$$B_t = \begin{bmatrix} 0 \\ C_p B_p \end{bmatrix}, \; C_t = [0_2 \; I_2] \tag{7b}$$

The rate-loop representation in Eq. (7a) is shown in Fig. 2, where the flexible and tracking parts are distinguished. In this representation, B_t is small in comparison with B_f (typically $\| B_t \| \; / \; \| B_f \| < 10^{-6}$. Also A_{tf} is small in comparison with A_t and A_f (typically $\| A_{tf} \| < 10^{-3}$, $\| A_f \| > 10$, and $\| A_t \| = 1$). Both properties are the result of a small value of $\| C_p \|$, shown earlier. Thus, the states of the tracking part are much weaker than the states of the flexible part. The strong and weak signal flows are shown in Fig. 2. The strong states of the flexible subsystem and the weak states of the tracking subsystem are shown in Fig. 3, which presents the transfer function plots of the rate-loop systems due to elevation rate command. This property is a foundation of the control design strategy described below.

III. Quasi-Separation of the Flexible and Tracking Subsystems

In the LQ design, the feedback $u = -Kx$ is determined such that the performance index J,

$$J = \int_0^\infty (x^T Q x + u^T R u) \; dt \tag{8}$$

is minimal. The minimum of J is obtained for the gain $K = R^{-1} B^T S$, where S is a solution of the Riccati equation [4]:

$$A^T S + S A - S B R^{-1} B^T S + Q = 0 \tag{9}$$

In the above equations, R is a positive definite input weight matrix, while Q is a positive semidefinite state weight matrix. It is assumed that $R = \rho I$, since both inputs (elevation and azimuth commands) are of equal importance. The further assumption that $\rho = 1$ is made without loss of generality. Divide S and K into parts related to the triple (A, B, C) in Eq. (7a):

$$S = \begin{bmatrix} S_t & S_{tf} \\ S_{tf}^T & S_f \end{bmatrix}, \; K = [K_t \; K_f] \tag{10}$$

so that Eq. (9) can be written as follows:

$$A_t^T S_t + S_t A_t - S_t B_t B_t^T S_t + Q_t - \Delta_{tf} = 0 \tag{11a}$$

$$A_t^T S_{tf} + S_{tf} A_f + S_t A_{tf} - K_t^T K_f = 0 \tag{11b}$$

$$A_f^T S_f + S_f A_f + S_f B_f B_f^T S_f + Q_f - \Delta_{ft} = 0 \tag{11c}$$

where

$$K_t = B_t^T S_t + B_f^T S_{tf}^T \tag{12a}$$

$$K_f = B_t^T S_{tf} + B_f^T S_f \tag{12b}$$

$$\Delta_{tf} = S_t B_t B_f^T S_{tf}^T + S_{tf} B_f K_t \tag{12c}$$

$$\Delta_{ft} = A_{tf}^T S_{tf} + S_{tf}^T A_{tf} - S_{tf}^T B_t K_f + S_f B_f B_t^T S_{tf} \tag{12d}$$

Taking a closer look at Eqs. (12), one can notice that there exist weights Q_t and Q_f such that the gain K_f depends on the flexible subsystem only. Namely, for a large enough matrix Q_f, such that $\| Q_f \| \gg \| \Delta_{ft} \|$, the solution S_f of Eq. (11c) is independent of the tracking subsystem, and for small matrix Q_t one obtains $\| B_t^T S_{tf} \| \ll \| B_f^T S_f \|$. In terms of Eq. (12b), the latter inequality means that the gain K_f depends only on the flexible subsystem. However, due to the weak–strong relationship between flexible and tracking subsystems, the situation is not quite symmetric: There are no Q_t and Q_f such that the gain K_t depends only on the tracking subsystem. To understand this, note that the term small has a different meaning for Q_f and Q_t. Magnitudes of a small matrix Q_f and a small matrix Q_t are of a different order, namely Q_f is small if $\| Q_f \| < 10^{-7}$

and Q_t is small if $\| Q_t \| < 1$. Therefore, increasing Q_t to obtain $\|Q_t\| \gg \|\Delta_{tf}\|$, one obtains $\|B_f^T S_{tf}\|$ and $\|B_t^T S_t\|$ of the same magnitude. According to Eq. (12a), the latter fact means that the gain K_t depends on the flexible subsystem as well as on the tracking subsystem, and the solution S_t of Eq. (11a) is dependent on the flexible subsystem. This property can be validated by observation of the closed-loop transfer functions for different weights as shown in Fig. 4. It follows from the plots that the variations of Q_f changed the properties of the flexible subsystem only, while the variations of Q_t changed the properties of both subsystems.

The independence of the flexible subsystem gains from the tracking subsystem properties is a consequence of small values of B_t, A_{tf}, and Q_t. However, it is required that the weight Q_t be large enough to achieve the required pointing performance. But the increase of Q_t causes the increasing dependency of the flexible subsystem gains on the tracking subsystem. This phenomenon in controller design changes the above independence into a quasi-independence (conditional independence). This property results in a separation of the flexible and tracking parts in the first stage of controller design. Thus the design consists of initial determination of the controller gains of the flexible subsystem followed by adjustment of weights of the tracking subsystem and a final tuning of the flexible weights.

IV. Properties of the LQ Controller for Flexible Structures

The properties of an LQ controller for a flexible subsystem are discussed in this section. In this application, a linear system with distinct complex conjugate pairs of poles and small real parts of the poles is considered a flexible structure. In the following, a balanced state-space representation of a flexible structure is discussed. The balanced representation of flexible structures is close (but not identical) to a modal one [5,6,7]. For LQ synthesis purposes, a balanced rather than modal representation is recommended since the balanced reduction (necessary in controller design) yields more accurate results than the modal reduction, especially for closely spaced poles [8].

Since the LQ controller for the flexible subsystem is determined separately from the tracking subsystem, in this section only the flexible subsystem is considered. Its state-space representation (A, B, C) is controllable and observable (the subscript f is dropped in this section for simplicity of notation), and its controllability (W_c) and observability (W_o) grammians are equal and diagonal, $W_c = W_o = \Gamma$, where Γ is a positive definite diagonal matrix that satisfies the following Lyapunov equations:

$$A\Gamma + \Gamma A^T + BB^T = 0, \quad A^T\Gamma + \Gamma A + C^T C = 0 \tag{13}$$

For a balanced flexible system with n components (or $2n$ states), the balanced grammian has the following form:

$$\Gamma \cong \mathrm{diag}(\gamma_1, \gamma_1, \gamma_2, \gamma_2, \cdots, \gamma_n, \gamma_n) \tag{14}$$

and the matrix A is almost block diagonal [6,7], with dominant 2×2 blocks on the main diagonal:

$$A \cong \mathrm{diag}(A_i), \quad A_i = \begin{bmatrix} -\zeta_i\omega_i & -\omega_i \\ \omega_i & -\zeta_i\omega_i \end{bmatrix}, \; i = 1, \cdots, n \tag{15}$$

where ω_i is the ith natural frequency of the structure and ζ_i is the ith modal damping. The combination of Eqs. (13) and (15) gives

$$\gamma_i(A_i + A_i^T) \cong -B_i B_i^T \cong -C_i^T C_i \tag{16}$$

For the LQ controller defined by Eqs. (8) and (9), it is assumed that

$$Q = \mathrm{diag}(q_i I_2), \text{ and } 0 < q_i \ll 1, \; i = 1, \cdots, n \tag{17}$$

Denote

$$\beta_i = \sqrt{1 + 2q_i\gamma_i/\zeta_i\omega_i} \tag{18}$$

then one obtains Proposition 1.

Proposition 1. $S \cong \mathrm{diag}(s_i I_2)$ is the solution of Eq. (9), where

$$s_i = -0.5\gamma_i^{-1}(1 - \beta_i), \quad i = 1, \cdots, n \tag{19}$$

Proof is presented in the Appendix.

The plots of s_i with respect to $\zeta_i\omega_i$ and γ_i are shown in Fig. 5. They show s_i increases with the weight q_i increase, and s_i decreases with γ_i or $\zeta_i\omega_i$ increase.

Next it will be shown that weighting as in Eq. (17) shifts the ith pair of complex poles of flexible structure, and leaves the remaining pairs of poles almost unchanged. Only the real part of the pair of poles is changed (moving the pole apart from the imaginary axis (see Fig. 6).

Proposition 2. For the weight Q as in Eq. (17) the closed-loop pair of flexible poles $(\lambda_{cri}, \pm j\lambda_{cii})$ is obtained from the open-loop poles $(\lambda_{ori}, \pm j\lambda_{oii})$:

$$(\lambda_{cri}, \pm j\lambda_{cii}) = (\beta_i \lambda_{ori}, \pm j\lambda_{oii}), \quad i = 1, \cdots, n \tag{20}$$

where β_i is defined in Eq. (18).

For proof see the Appendix.

The real part of the poles is shifted by β_i, while the imaginary part of the poles remains unchanged. The plots of β_i with respect to $\zeta_i\omega_i$ and γ_i are shown in Fig. 7. They show relatively large values of β_i even for small values of q_i, i.e., a significant pole shift to the left. Also, since β_i increases with γ_i and decreases with $\zeta_i\omega_i$, there is a significant pole shift for highly observable and controllable modes with small damping. In terms of the transfer function profile, the weight q_i suppresses the resonant peak at frequency ω_i while leaving the natural frequency unchanged (see Fig. 8 for $i = 1$). Due to weak coupling between modes, the assignment of one mode insignificantly influences other modes. Therefore, the weight assignment is performed either simultaneously for all modes or for each mode separately.

V. Controller Design Algorithm

The LQ controller configuration for the DSN antenna model is shown in Fig. 9(a). The tracking command y_c is compared with the antenna position y_p, and the error $\epsilon = y_p - y_c$ and the integral of the error are the controller inputs.

The procedure for the antenna LQ controller design is sequential. First, for the ad hoc (but relatively small) chosen weights of the tracking subsystem, the weights of the flexible subsystem are determined. Second, the adjustment of the weights of the tracking system is performed, followed by the final adjustment of the weights of the flexible system. The weights of the flexible subsystem are determined sequentially, simplifying the procedure.

The controller order is determined as a part of the weight tuning process. Only the modes that influence the plant performance are considered. If the number of flexible modes is n_f, the number of disregarded modes is n_o, and the size of the tracking system is n_t, then the controller order n_c is

$$n_c = n_t + 2(n_f - n_o) \tag{21}$$

The following LQ controller design algorithm is proposed:

(1) Determine the plant state-space representation, consisting of flexible and tracking parts, in the form of Eq. (7).

(2) Choose ad hoc but reasonably small weights for the tracking part $Q_t = Q_{\text{tah}}$.

(3) For each balanced coordinate of the flexible part, choose the weight q_i $(i = 1, \cdots, n_f)$, and define the weight matrix $Q_{fi} = \text{diag}(0,\ 0,\ \cdots,\ q_i,\ q_i,\ 0,\ 0,\ \cdots,\ 0)$ so that the closed-loop system performance for the weight $Q_i = \text{diag}(Q_{\text{tah}},\ Q_{fi})$ is maximized. For example, determine the weights q_i to impose the required pole shift or to suppress the ith resonant peak to the required level without depreciating other properties of the closed-loop transfer function. Note that for small values of q_i, only the ith pair of poles is shifted (to the left), and the remaining poles are almost unaffected. Disregard the modes for which the weighting does not improve the closed-loop system performance. The resulting weight for the flexible subsystem is

$$Q_f = \sum_{i=1}^{n_f} Q_{fi} \tag{22}$$

(4) For the already determined weight Q_f, tune weight Q_t to obtain improvements in tracking properties of the antenna.

(5) Adjust the weights of the flexible subsystem, if necessary.

VI. Closed-Loop System

Equations for a closed-loop system with the LQ controller, and with the LQ controller and observer, are derived.

The closed-loop system configuration is shown in Fig. 9(a). The equations for the plant triple (A, B, C), given by Eq. (7) are

$$\dot{x} = Ax + Bu_f, \quad y_p = C_p x, \quad x_f = C_f x \tag{23}$$

Denoting the output error ϵ and the integral of the error z, one obtains

$$\epsilon = y_p - u = C_p x - u, \quad \dot{z} = \epsilon \tag{24}$$

thus

$$u_f = k_f C_f x - k_o C_p x - k_i z + k_o u \tag{25}$$

Defining the closed-loop state $x_{cl} = [z\ x^T]^T$, one obtains the closed-loop equations

$$\dot{x}_{cl} = A_{cl} x_{cl} + B_{cl} u, \quad y_p = C_{cl} x_{cl} \tag{26a}$$

where

$$A_{cl} = \begin{bmatrix} 0 & C_p \\ -Bk_i & A - Bk_f C_f - Bk_o C_p \end{bmatrix}$$

$$B_{cl} = \begin{bmatrix} -I \\ Bk_o \end{bmatrix}, \tag{26b}$$

$$C_{cl} = [0\ C_p]$$

The antenna states are not directly measured; thus the state observer is included in the closed-loop system in Fig. 9(b). Based on plant Eq. (23), the estimator equations are obtained:

$$\dot{\hat{x}} = A\hat{x} + Bu_f + K_e(y_p - \hat{y}_p)$$
$$\hat{y}_p = C_p \hat{x}, \quad \hat{x}_f = C_f \hat{x} \tag{27}$$

The integrator equation

$$\dot{z} = \epsilon = y_p - u = C_p x - u \tag{28}$$

and the nodal equation

$$u_f = -k_f \hat{x}_f - k_o \epsilon - k_i z \tag{29}$$

along with Eqs. (23) and (27) give the equations for the closed-loop system with the state observer

$$\dot{x}_{co} = A_{co} x_{co} + B_{co} u, \quad y = C_{co} x_{co} \tag{30a}$$

where the closed-loop state is $x_{co} = [z\ x^T\ \hat{x}^T]^T$ and

$$A_{co} = \begin{bmatrix} 0 & C_p & 0 \\ -Bk_i & A - Bk_o C_p & -Bk_f C_f \\ -Bk_i & K_e C_p - Bk_o C_p & A - K_e C_p - Bk_f C_f \end{bmatrix}$$

$$B_{co} = \begin{bmatrix} -1 \\ Bk_o \\ Bk_o \end{bmatrix}, \quad C_{co} = [0\ C_p\ 0] \tag{30b}$$

VII. LQ Controller Design for the DSS-13 Antenna

The DSS-13 antenna model consists of two tracking states (azimuth and elevation angle), and 13 flexible modes (or 26 balanced states). The preliminary weights $q_{ie} = q_{pe} = q_{ia} = q_{pa} = 1$ for the tracking subsystem (for y_i and y_p) and zero weights for the flexible subsystem ($q_1 = q_2 = \cdots = q_{13} = 0$ for all 13 modes) have been chosen for the LQ controller design. The closed-loop system step response is presented in Fig. 10 (elevation and azimuth encoder reading due to elevation and azimuth command) and the magnitudes of the closed-loop transfer function in Fig. 11. Both figures show that flexible motion of the antenna is excessive, and should be damped out. This is achieved by adjusting weights for the flexible subsystem. For the same tracking weights as before, the weight for the first mode (2.32 Hz) is chosen to be $q_1 = 10^{-7}$, and the remaining weights are zero; this obtains the closed-loop system responses shown in Figs. 12 and 13. One can see that the 2.32-Hz resonance peak in the azimuth command response (Fig. 12) has disappeared, as well as most of the flexible motion in the azimuth step response (Fig. 13). The elevation motion is unaffected however, since the azimuth gearbox mode is almost nonexistent in the elevation motion.

The weight should be chosen carefully. Too small a weight (e.g., 3×10^{-9} in the case considered) will not suppress the resonant peak, Fig. 14(a). Too large a weight (e.g., 1×10^{-5}) will deteriorate the tracking performance: For the overweighted mode, the transfer function is pressed down within a wide frequency range, Fig. 14(b). The proper weight suppresses the resonant peak, leaving the other peaks unchanged (Fig. 8).

Next, assuming q_{ie}, q_{ia}, q_{pe}, q_{pa} and q_1 are as given above and setting a weight for the second mode (2.64 Hz) $q_2 = 10^{-7}$ while the remaining weights are zero, one obtains the LQ control system responses shown in Figs. 15 and 16. As a result of a nonzero weight, the 2.64-Hz resonant peak has disappeared in the elevation command motion.

Similar procedures have been applied for the third (4.26-Hz), fourth (3.77-Hz), fifth (7.88-Hz), sixth (4.47-Hz), seventh (3.38-Hz), eighth (5.98-Hz), ninth (7.92-Hz), and tenth (9.48-Hz) modes, with weight 10^{-7} for each mode. As a result, suppression of the remaining flexible motion and resonant peaks is observed in Figs. 17 and 18. Weights for the remaining modes (eleventh through thirteenth) have been set to zero. According to Eq. (21), the controller order is 24 for the plant order of 30.

The root locus of the closed-loop system due to weight variations of the 7.92-Hz mode is shown in Fig. 19. The figure shows the horizontal departure of poles into the left-hand side direction (stabilizing property). It confirms the properties of the weighted LQ design described previously.

In the next step, the tracking properties of the system are improved by proper weight setting of the tracking subsystem. Namely, setting the integral weight to $q_{ie} = q_{ia} = 70$ and the proportional weight to $q_{pe} = q_{pa} = 100$ improves the system tracking properties, as shown in Fig. 20 (small overshoot and settling time) and in Fig. 21 (extended bandwidth—up to 2 Hz). However, by improving the tracking properties, the transfer function has been raised dramatically in the frequency region of 1 to 3 Hz, which forces the first two modes located in this region to appear again in the step response. By sacrificing a bit of the tracking properties, the flexible motion in the step response is reduced. This is done by increasing slightly the weights of the flexible subsystem, setting them as follows: $q_1 = q_2 = q_3 = q_4 = q_5 = q_6 = 10^{-6}$, $q_7 = q_8 = 10^{-7}$, and $q_9 = q_{10} = 10^{-5}$. The closed-loop system response with the satisfactory tracking performance is shown in Figs. 22 and 23 (small overshoot, small settling time, and 1-Hz bandwidth are observed).

VIII. Conclusions

A new procedure for the DSN antenna controller design has been proposed. The antenna model is divided into flexible and tracking parts rather than into elevation and azimuth parts. In a sequential design strategy, a controller for the flexible subsystem is designed first, followed by a controller design for the tracking subsystem. This approach results in a significant improvement of the performance of the antenna closed-loop system through a sequential weight adjustment of the state vector. The properties of the weight adjustment have been quantified in this article. The controller reduction is inherent in this approach. The minimal-order controller is determined through monitoring the closed-loop performance for each flexible mode. The DSS-13 antenna tracking controller design has been used to illustrate the procedure.

Acknowledgment

The author would like to thank B. Parvin for his helpful comments.

References

[1] L. S. Alvarez and J. Nickerson, "Application of Optimal Control Theory to the Design of the NASA/JPL 70-Meter Antenna Axis Servos," *TDA Progress Report 42-97*, vol. January–March 1989, Jet Propulsion Laboratory, Pasadena, California, pp. 112–126, May 15, 1989.

[2] W. Gawronski and J. A. Mellstrom, "Modeling and Simulations of the DSS 13 Antenna Control System," *TDA Progress Report 42-106*, vol. April–June 1991, Jet Propulsion Laboratory, Pasadena, California, pp. 205–248, August 15, 1991.

[3] G. Biernson, *Optimal Radar Tracking Systems*, New York: Wiley-Interscience, 1990.

[4] H. Kwakernaak and R. Sivan, *Linear Optimal Control Systems*, New York: Wiley-Interscience, 1972.

[5] E. A. Jonckheere, "Principal Component Analysis of Flexible Systems—Open Loop Case," *IEEE Trans. on Automat. Control*, vol. AC-29, no. 12, pp. 1095–1097, December 1984.

[6] W. Gawronski and T. Williams, "Model Reduction for Flexible Space Structures," *Journal of Guidance, Control, and Dynamics*, vol. 14, no. 1, pp. 68–76, January 1991.

[7] W. Gawronski and J.-N. Juang, "Model Reduction for Flexible Structures," in *Control and Dynamics Systems*, edited by C. T. Leondes, vol. 36, pp. 143–222, New York: Academic Press, 1990.

[8] T. Williams and W. Gawronski, "Model Reduction for Flexible Spacecraft with Clustered Natural Frequencies," *Proceedings of the 3rd Annual Conference on Aerospace Computational Control*, JPL Publication 89-45, Vol. 2, Jet Propulsion Laboratory, Pasadena, California, p. 620, 1989.

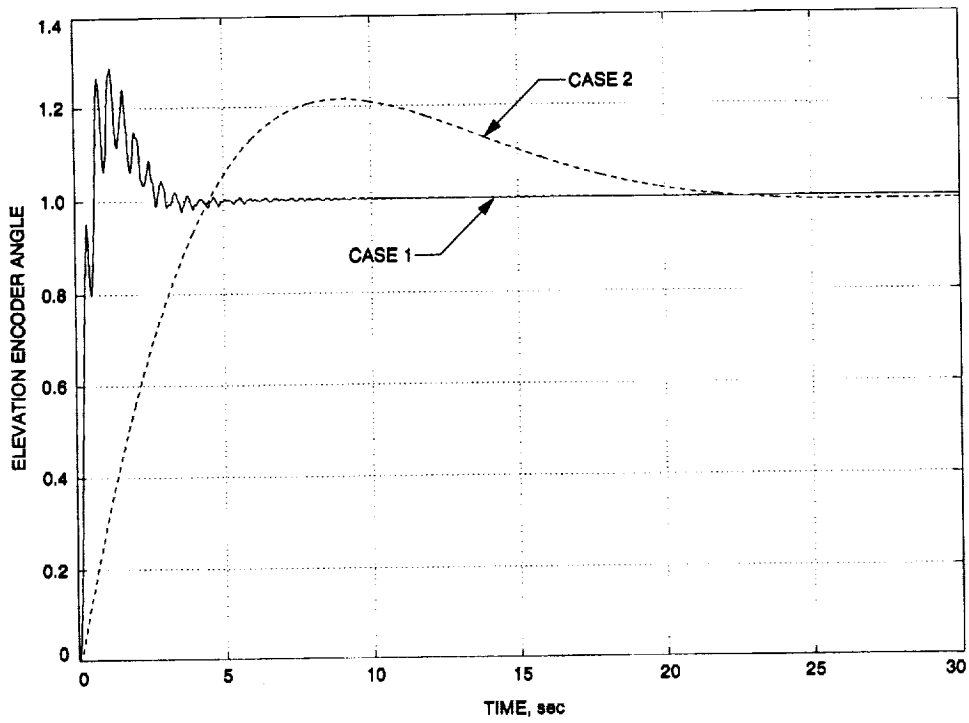


Fig. 1. Step response of the DSS-13 antenna with LQ controller for different weights.

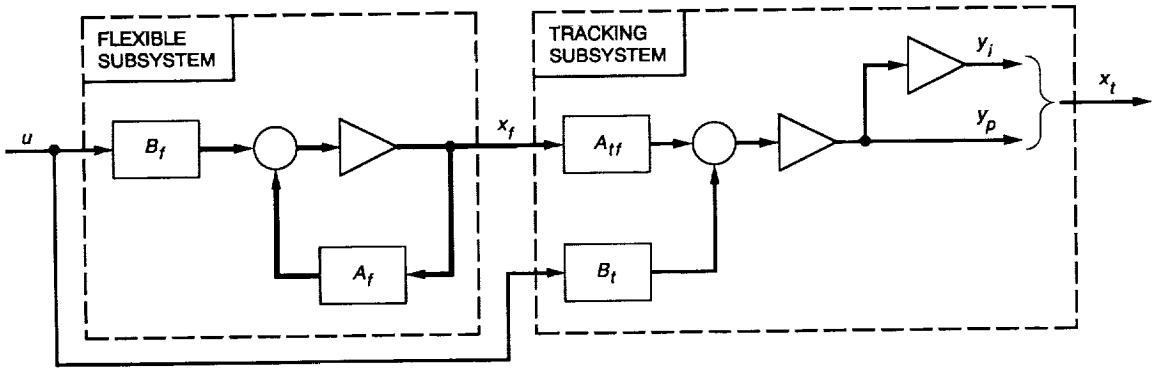


Fig. 2. System configuration.

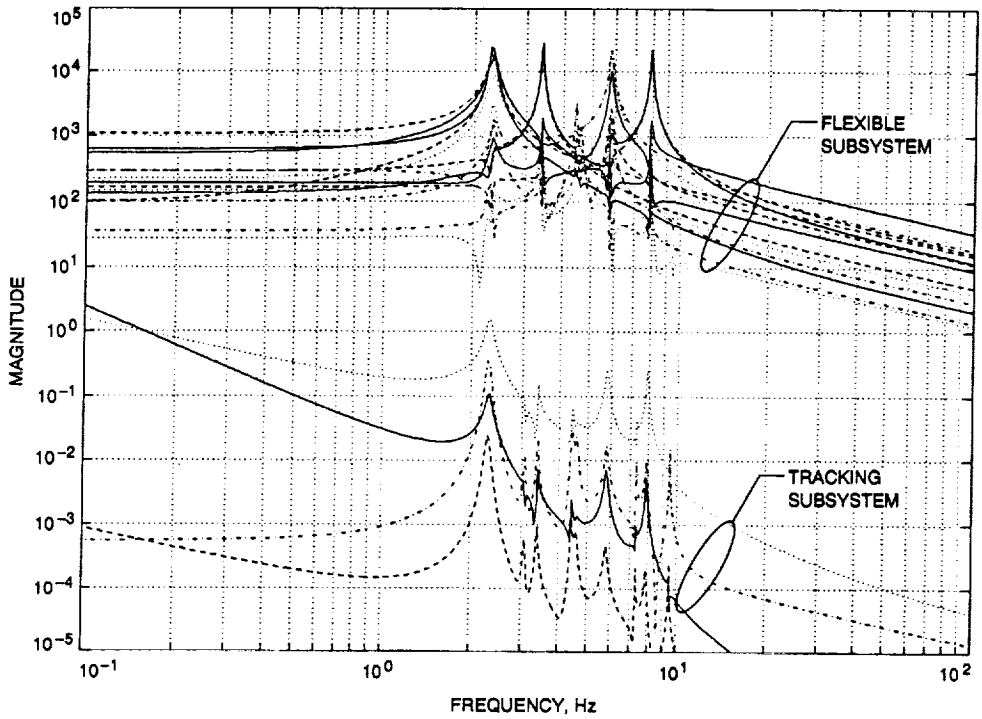


Fig. 3. Magnitude of transfer function of tracking and flexible subsystems for the elevation rate input.

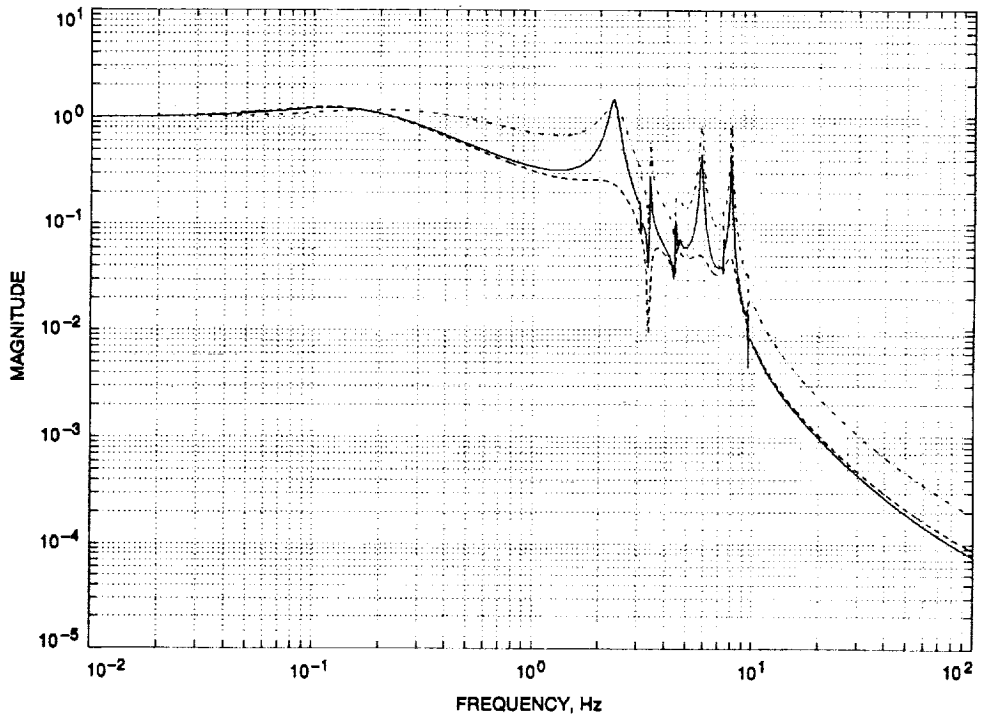


Fig. 4. Magnitude of transfer function of the closed-loop system for different weights: $Q_t = I_4$, $Q_f = 0 \times I_{26}$ (solid line); $Q_t = I_4$, $Q_f = 10^{-7} \times I_{26}$ (dashed line); and $Q_t = 10 \times I_4$, $Q_f = 0 \times I_{26}$ (dot-dashed line).

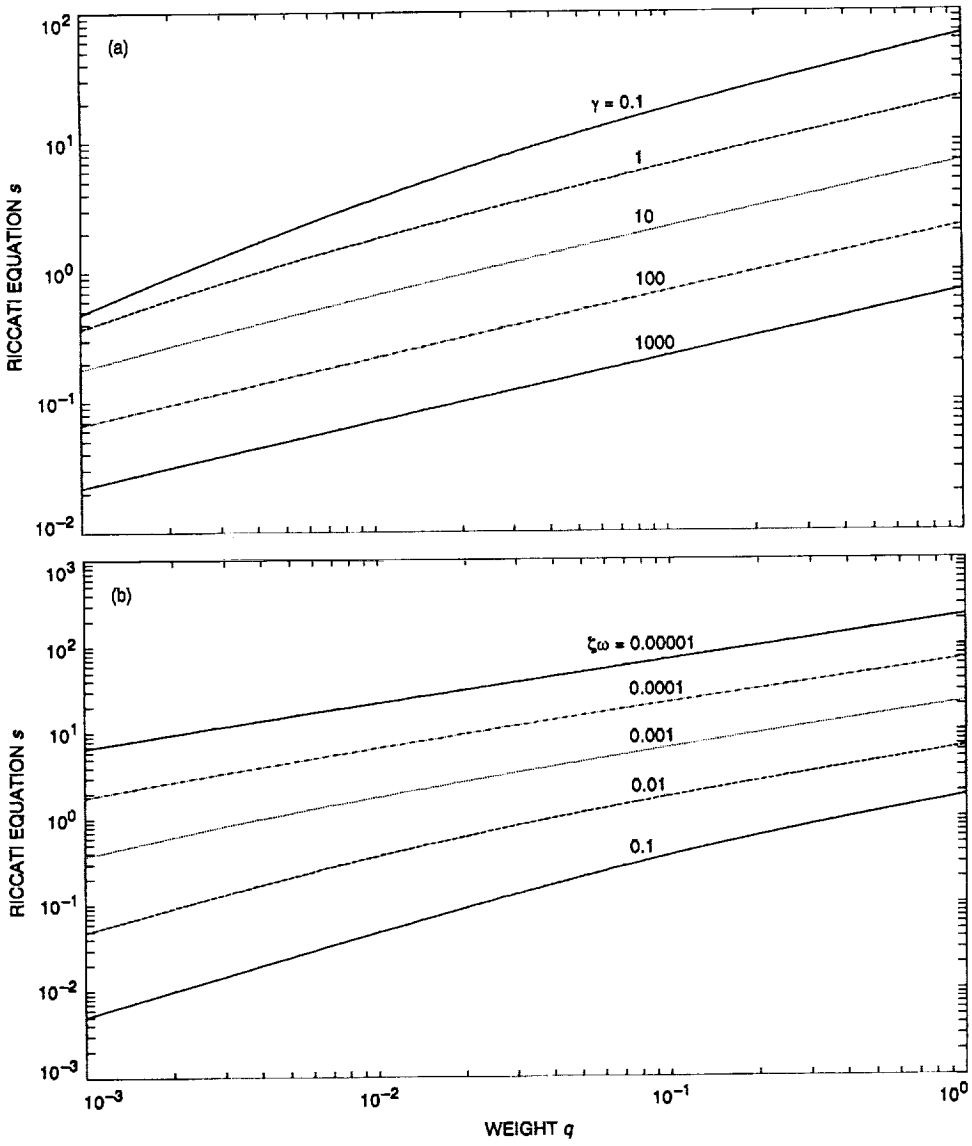


Fig. 5. Solution of the Riccati equation s versus: (a) weight q and Hankel singular value γ and (b) weight q and modal damping $\zeta\omega$.

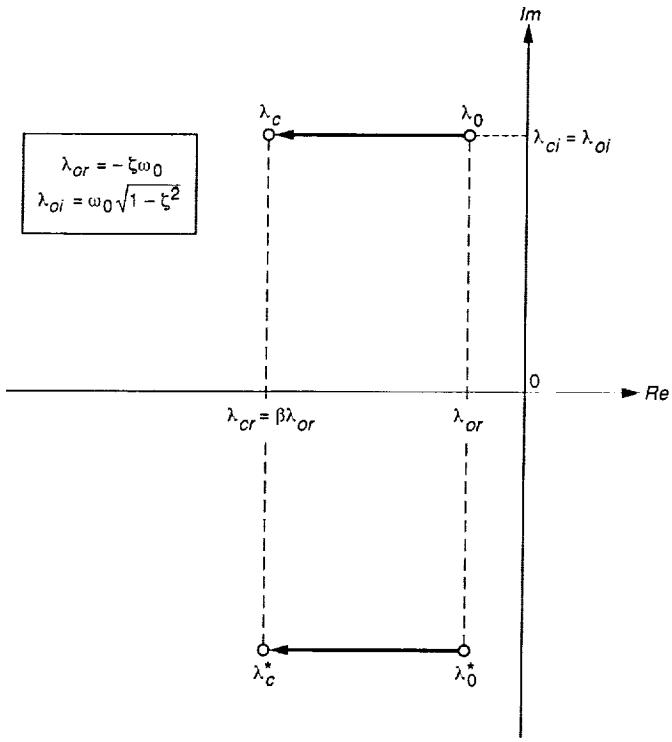


Fig. 6. Location of poles of a flexible structure.

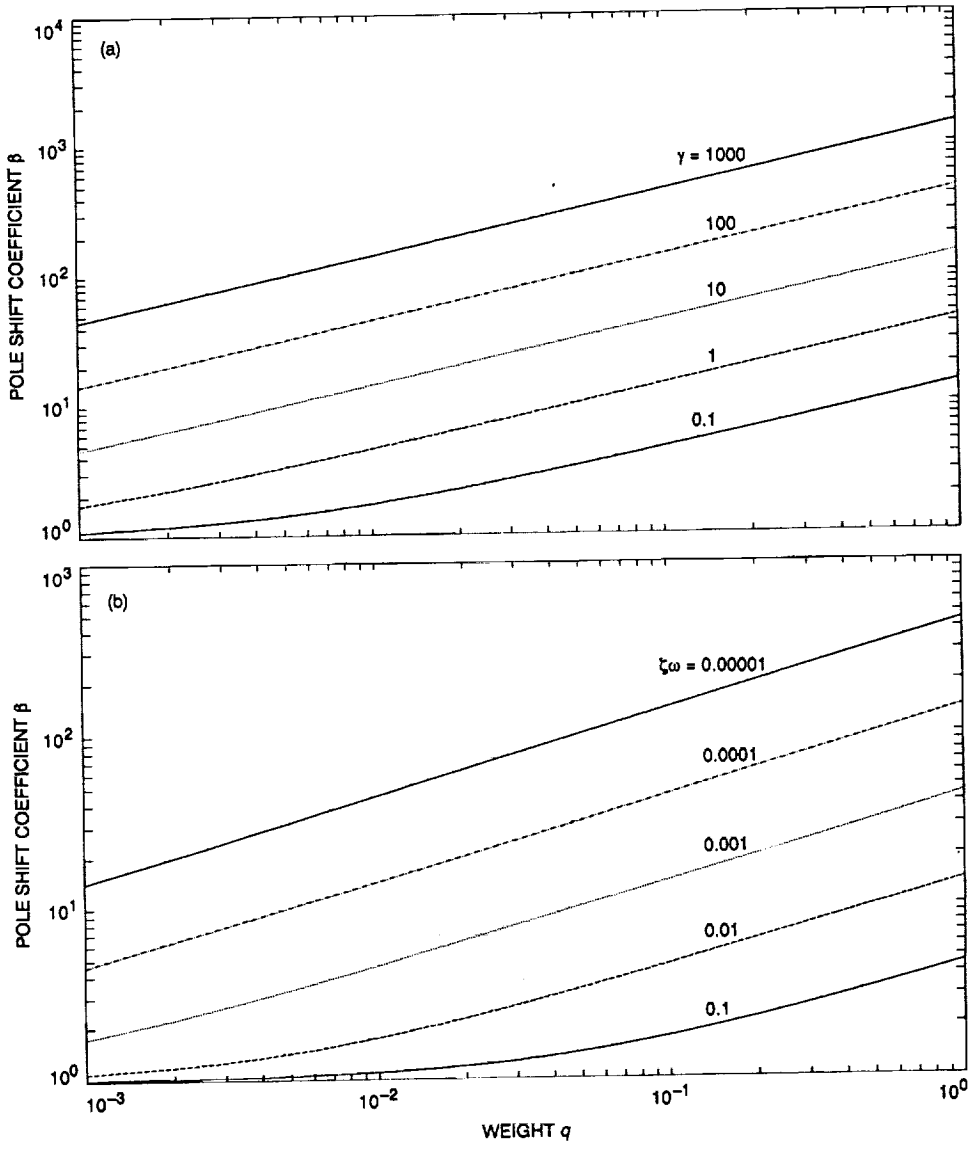


Fig. 7. Pole shift coefficient β versus: (a) weight q and Hankel singular value γ and (b) weight q and modal damping $\zeta\omega$.

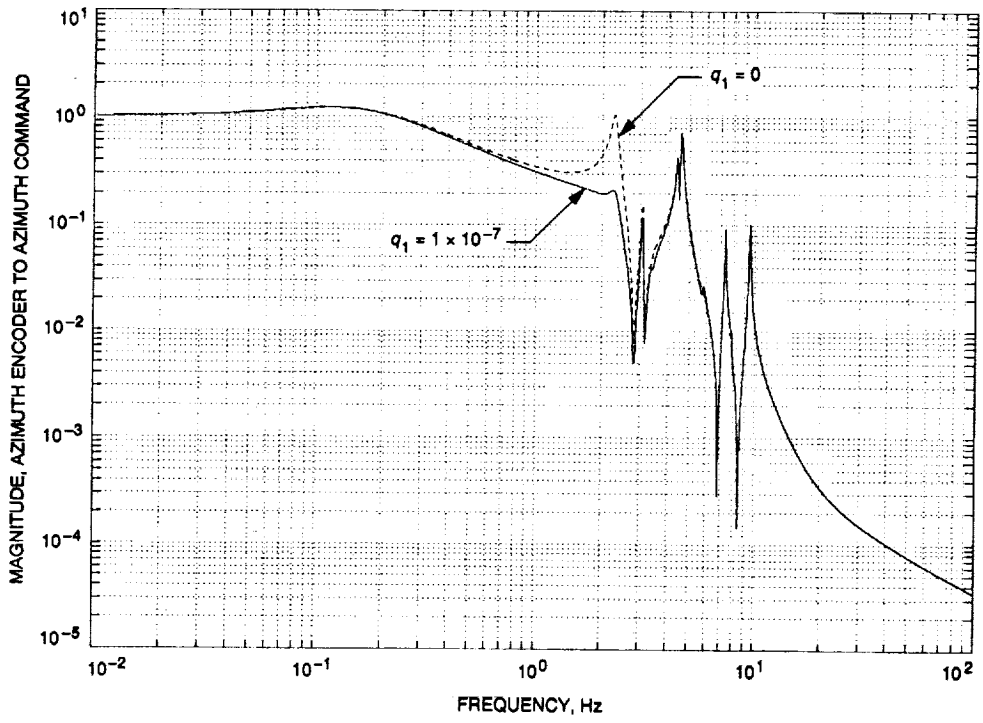


Fig. 8. Vibration suppression with a single weight.

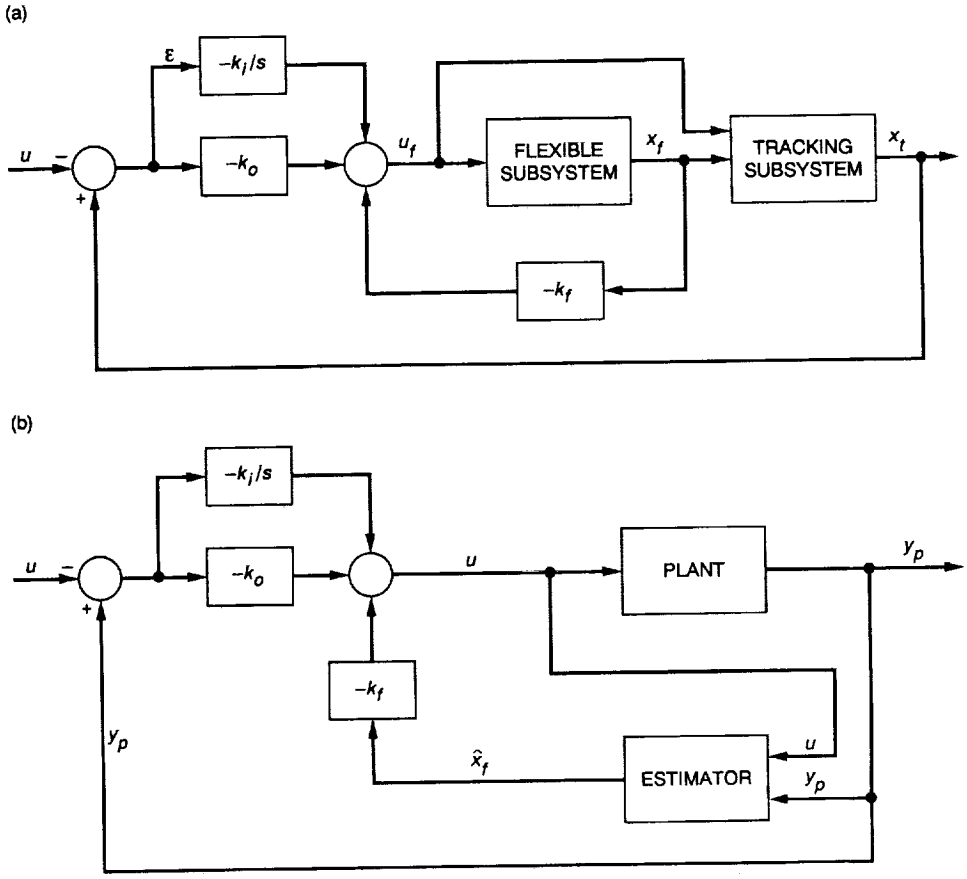


Fig. 9. LQ controller configuration: (a) antenna closed-loop system and (b) antenna closed-loop system with an observer.

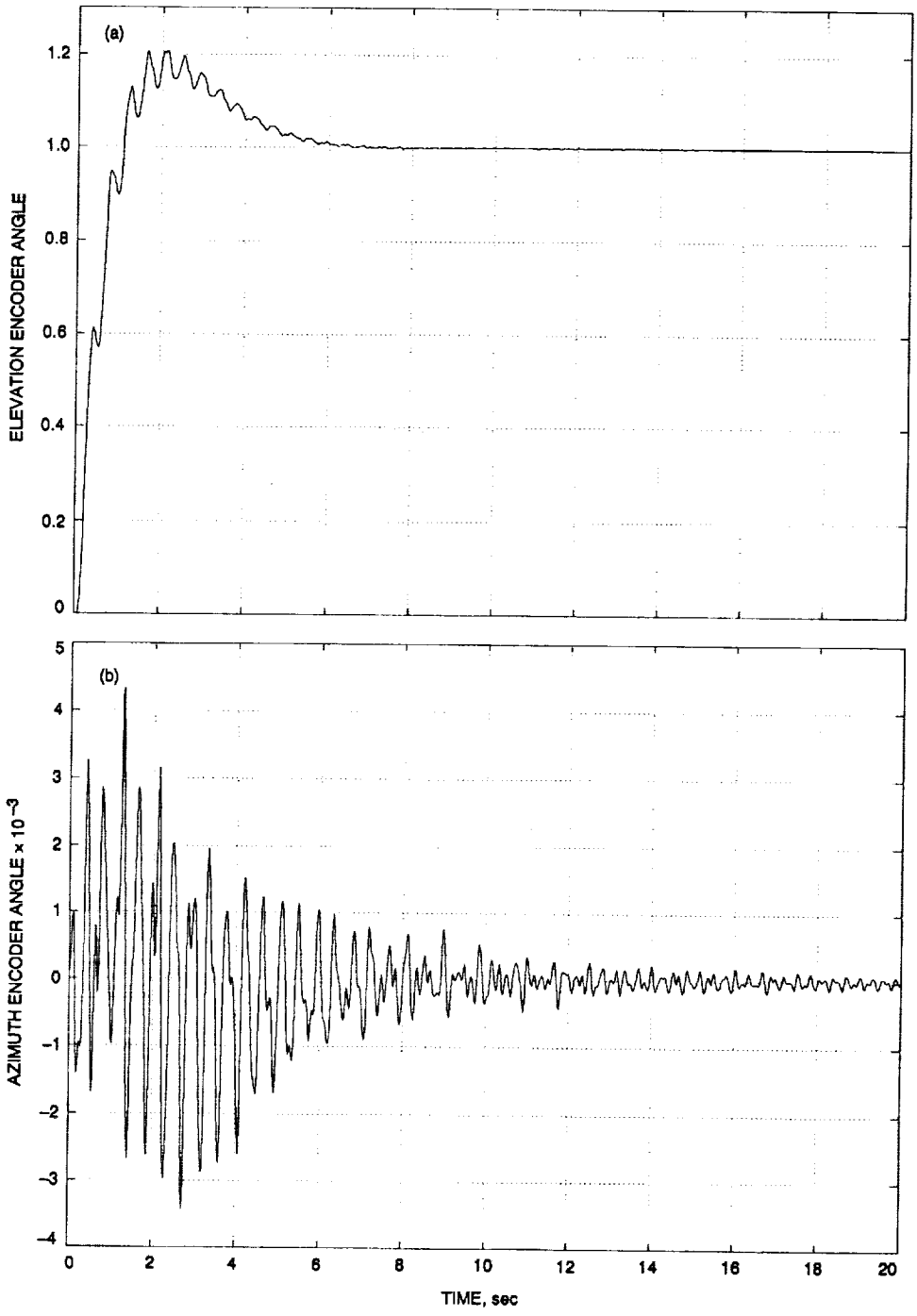


Fig. 10. Closed-loop response to step input for unit proportional and unit integral weights in azimuth and elevation, and zero weight for the flexible subsystem: (a) elevation encoder angle to elevation step command; (b) azimuth encoder angle to elevation step command; (c) azimuth encoder angle to azimuth step command; and (d) elevation encoder angle to azimuth step command.

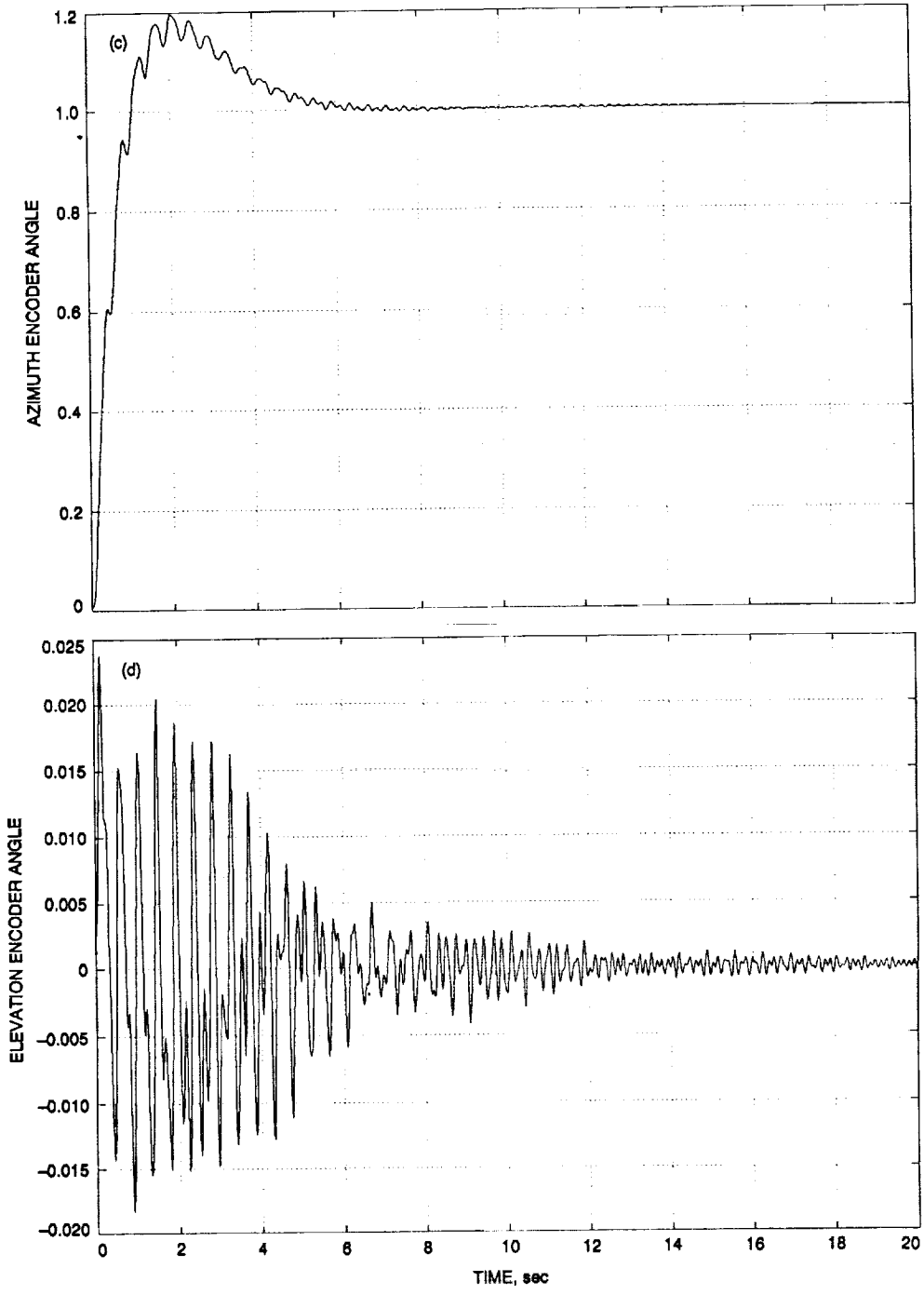

(c)
AZIMUTH ENCODER ANGLE
1.2
1.0
0.8
0.6
0.4
0.2
0
(d)
ELEVATION ENCODER ANGLE
0.025
0.020
0.015
0.010
0.005
0
−0.005
−0.010
−0.015
−0.020
0
2
4
6
8
10
12
14
16
18
20
TIME, sec

Fig. 10 (contd).

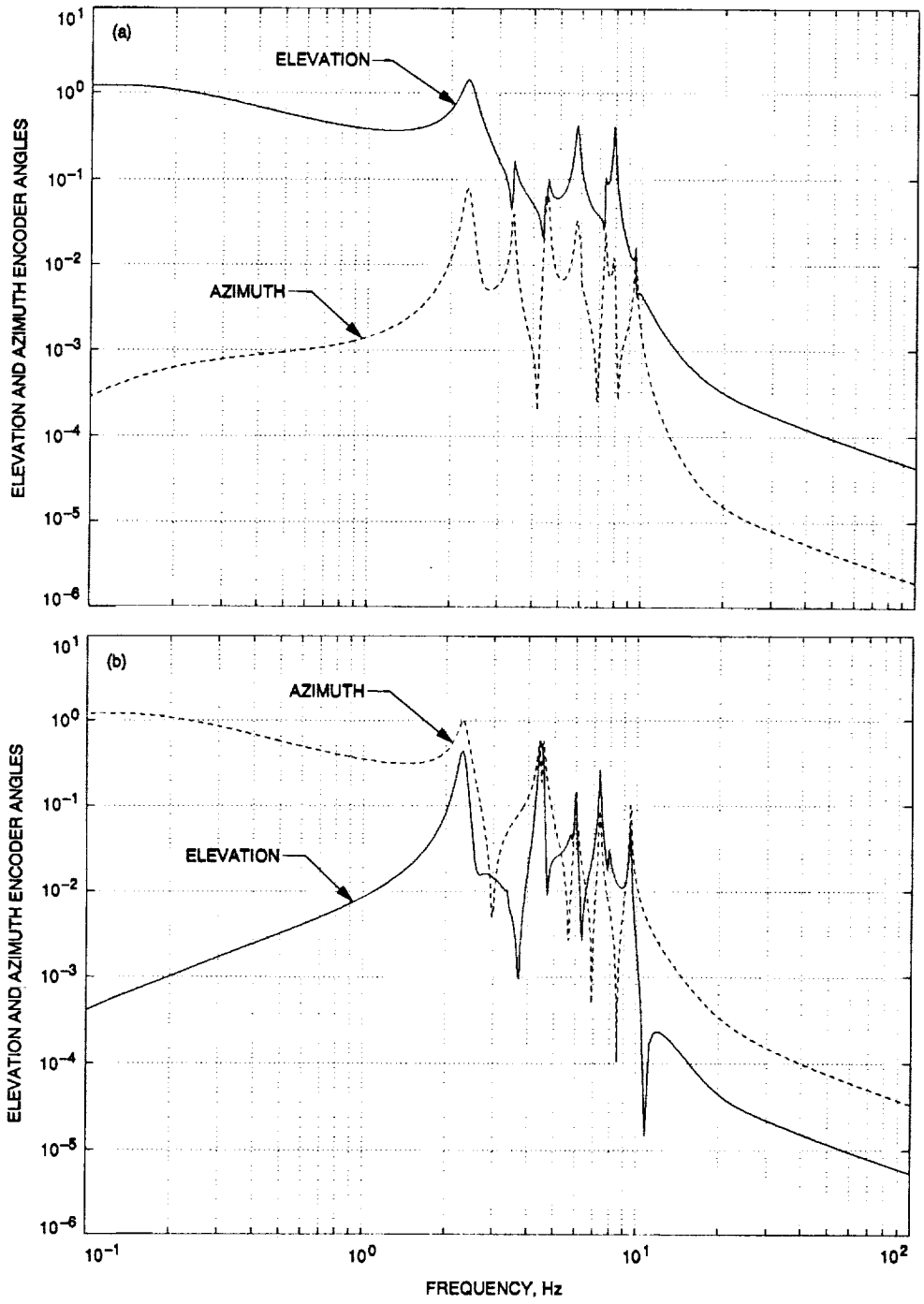


Fig. 11. Closed-loop transfer function for unit proportional and unit integral weights in azimuth and elevation, and zero weight for flexible subsystem: (a) elevation and azimuth encoder angles to elevation step command and (b) elevation and azimuth encoder angles to azimuth step command.

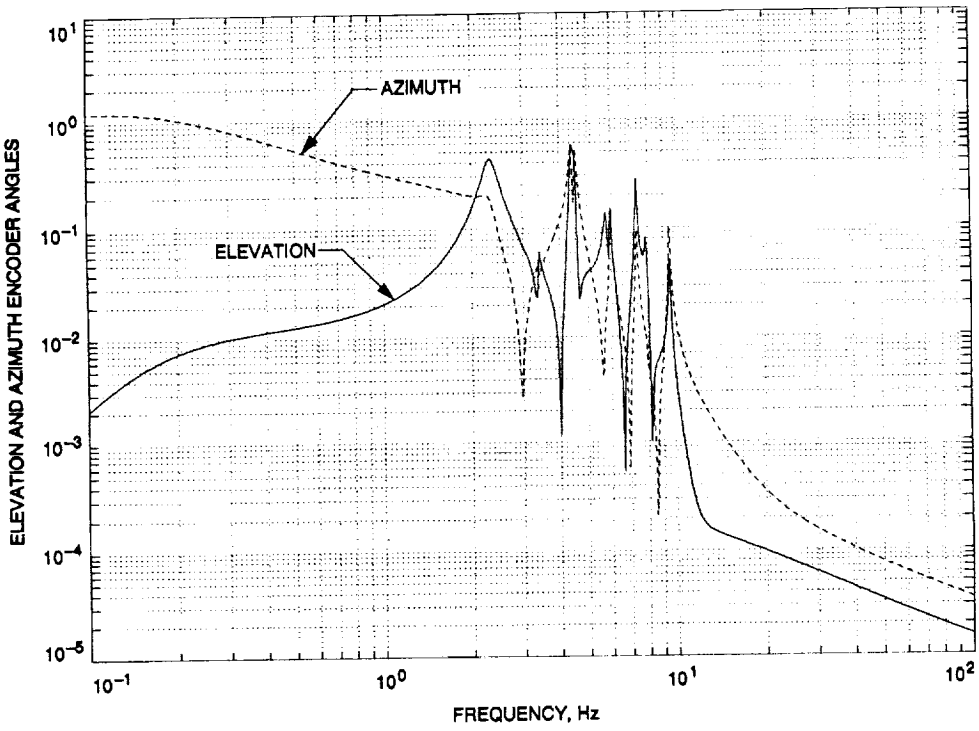


Fig. 12. Transfer function (azimuth and elevation encoder angles to azimuth command) for weights the same as those in Fig. 10, but $q_1 = 1 \times 10^{-7}$.

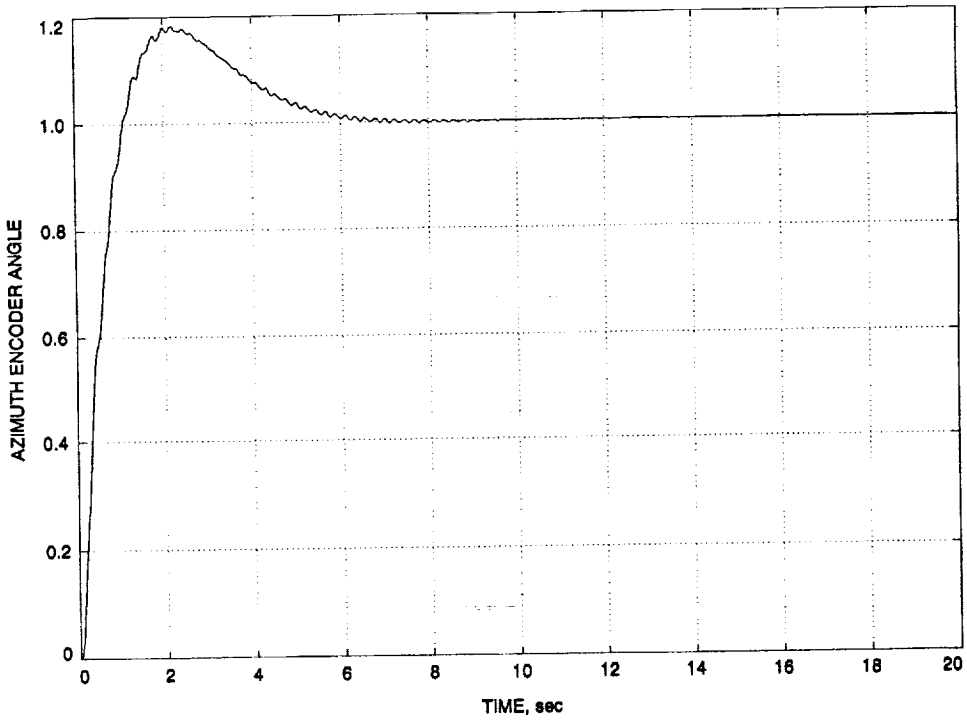


Fig. 13. Azimuth encoder angle step response to azimuth command for weights the same as those in Fig. 12.

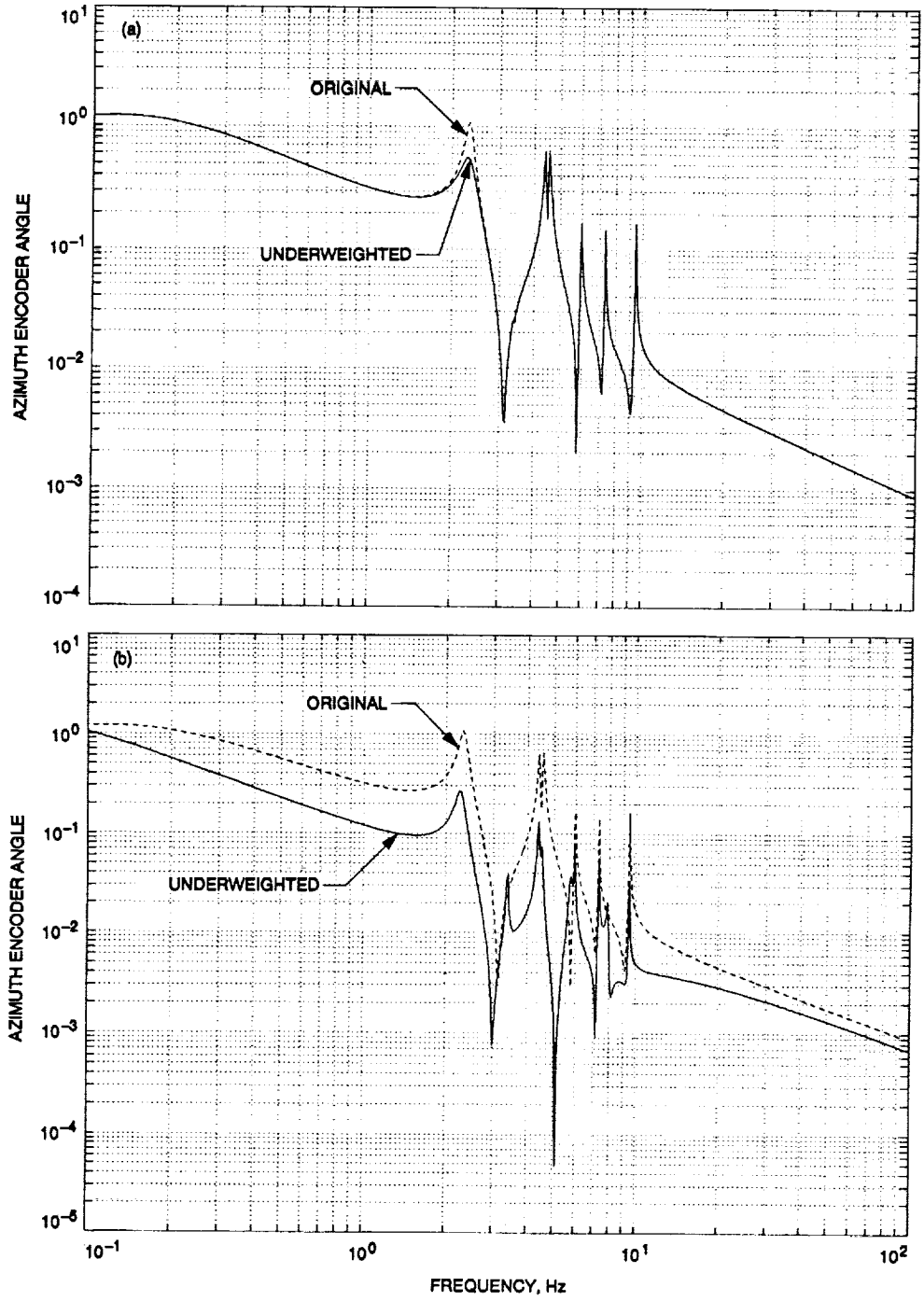


Fig. 14. The first component: (a) underweighted and (b) overweighted.

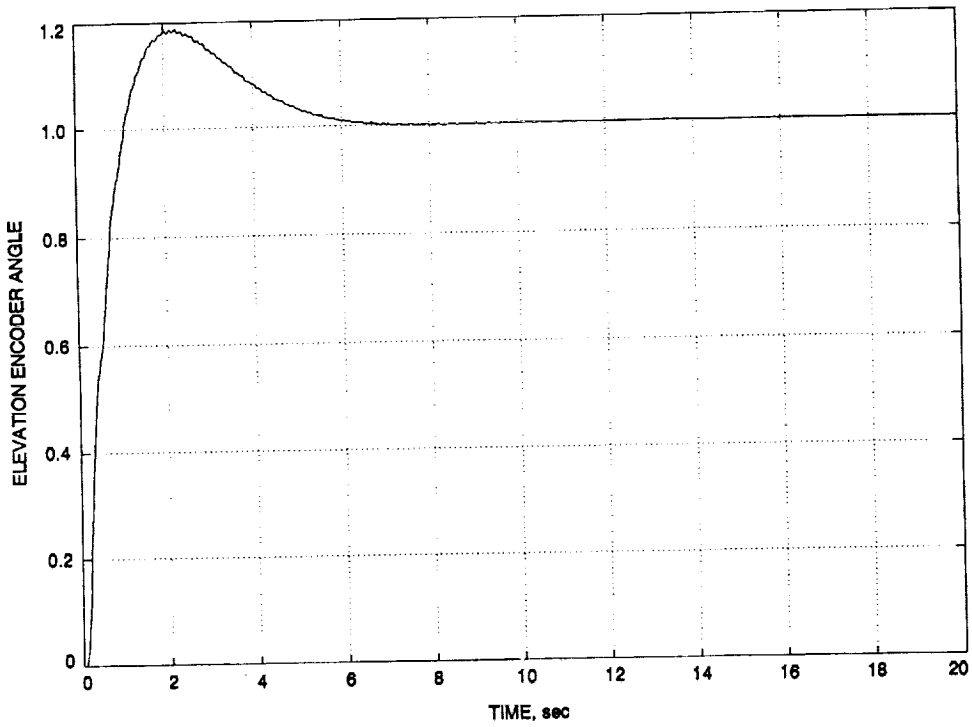


Fig. 15. Elevation encoder angle step response to elevation command for weights the same as those in Fig. 10, but $q_1 = q_2 = 1 \times 10^{-7}$.

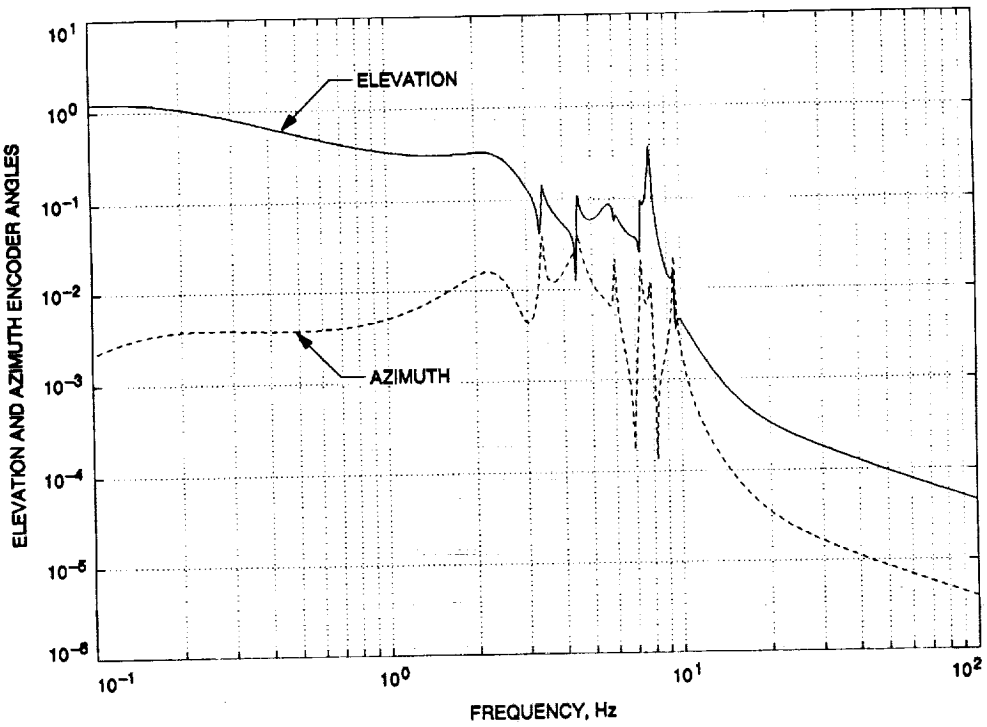


Fig. 16. Transfer function (azimuth and elevation encoder angles to elevation command) for weights the same as those in Fig. 15.

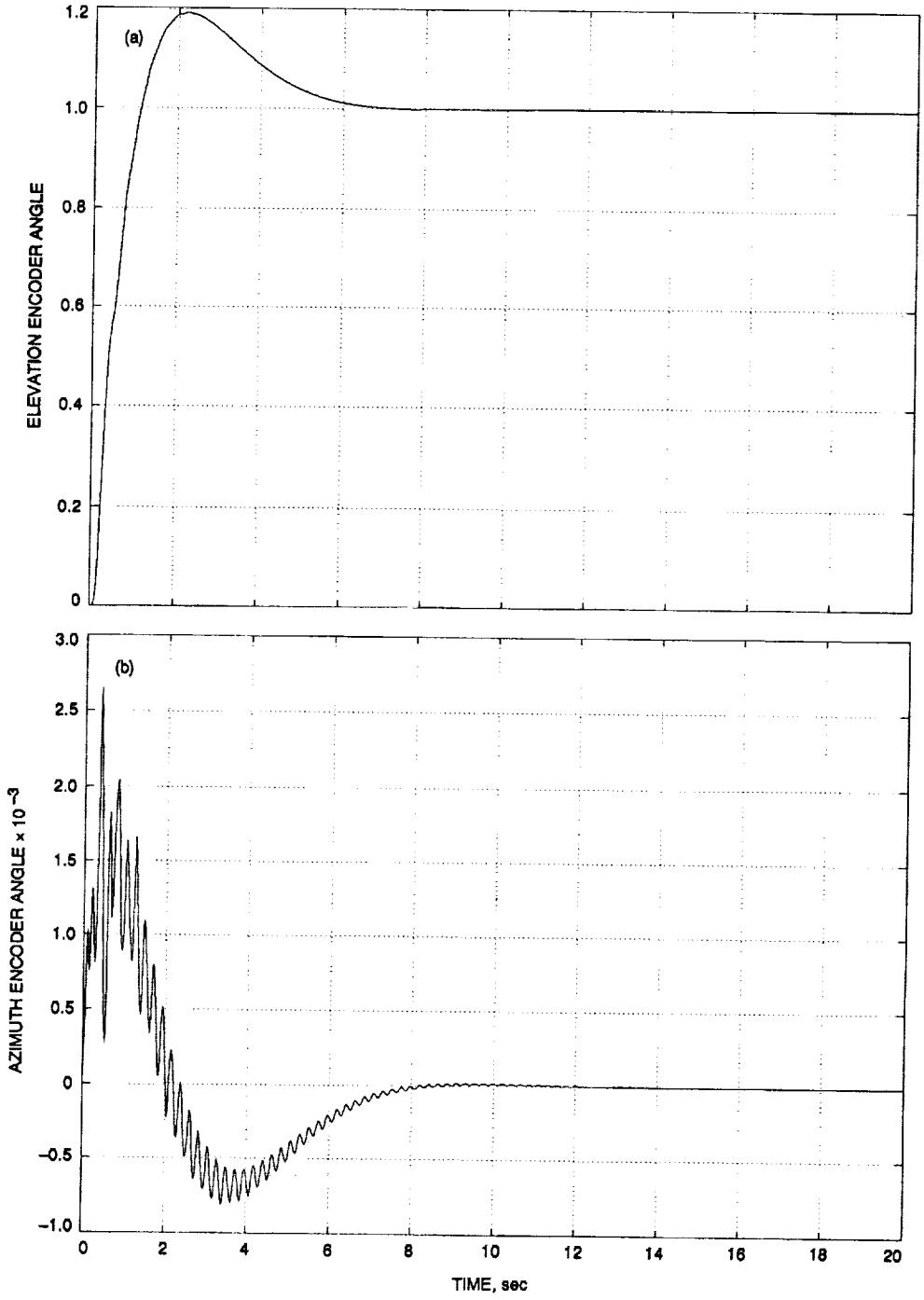


Fig. 17. Closed-loop response to step input for unit proportional and unit integral weights in azimuth and elevation, and weights for the flexible subsystem equal to 1×10^{-7}: (a) elevation encoder angle to elevation step command; (b) azimuth encoder angle to elevation step command; (c) azimuth encoder angle to azimuth step command; and (d) elevation encoder angle to azimuth step command.

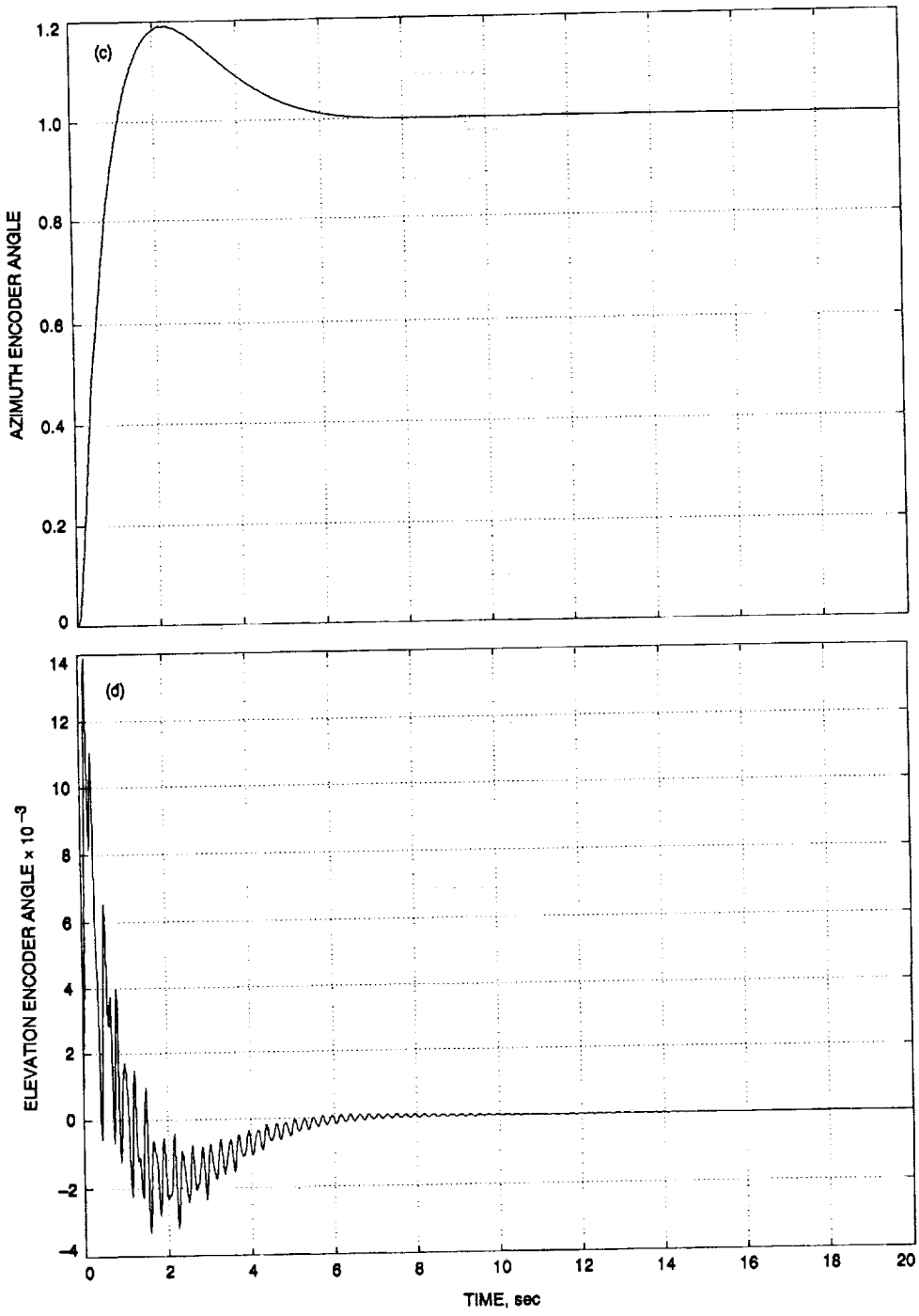

(c)
AZIMUTH ENCODER ANGLE
1.2
1.0
0.8
0.6
0.4
0.2
0
(d)
ELEVATION ENCODER ANGLE x 10 -3
14
12
10
8
6
4
2
0
-2
-4
0
2
4
6
8
10
12
14
16
18
20
TIME, sec

Fig. 17 (contd).

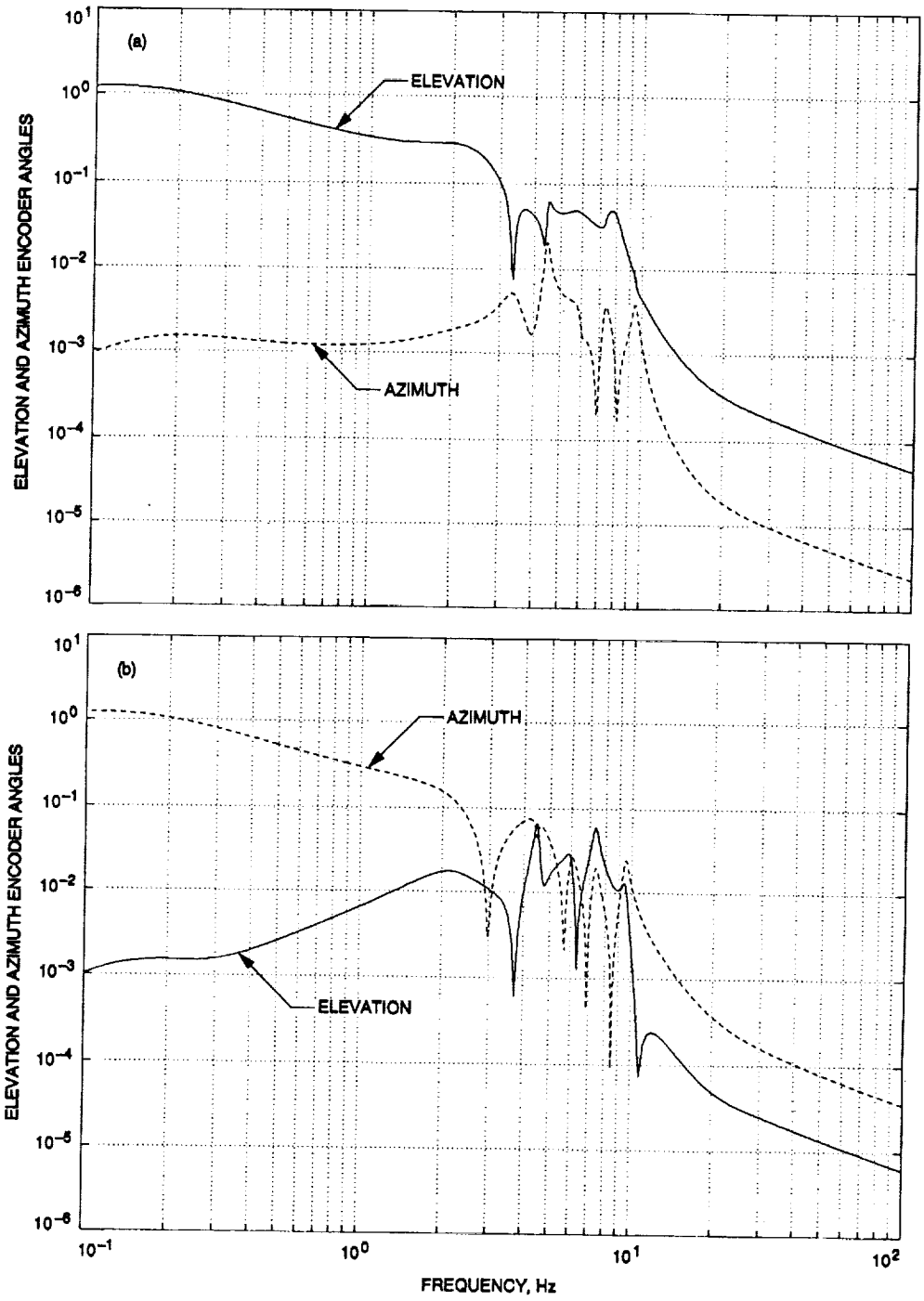


Fig. 18. Closed-loop transfer function for weights the same as those in Fig. 17: (a) elevation and azimuth encoder angles to elevation step command and (b) elevation and azimuth encoder angles to azimuth step command.

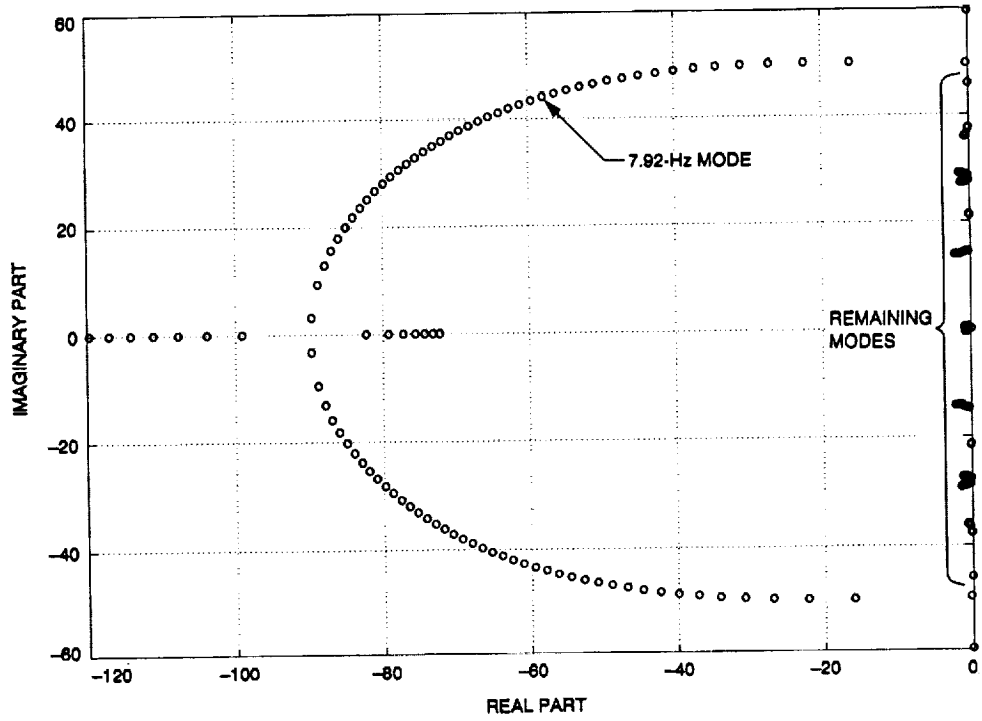


Fig. 19. Root locus for 7.92-Hz mode versus weight.

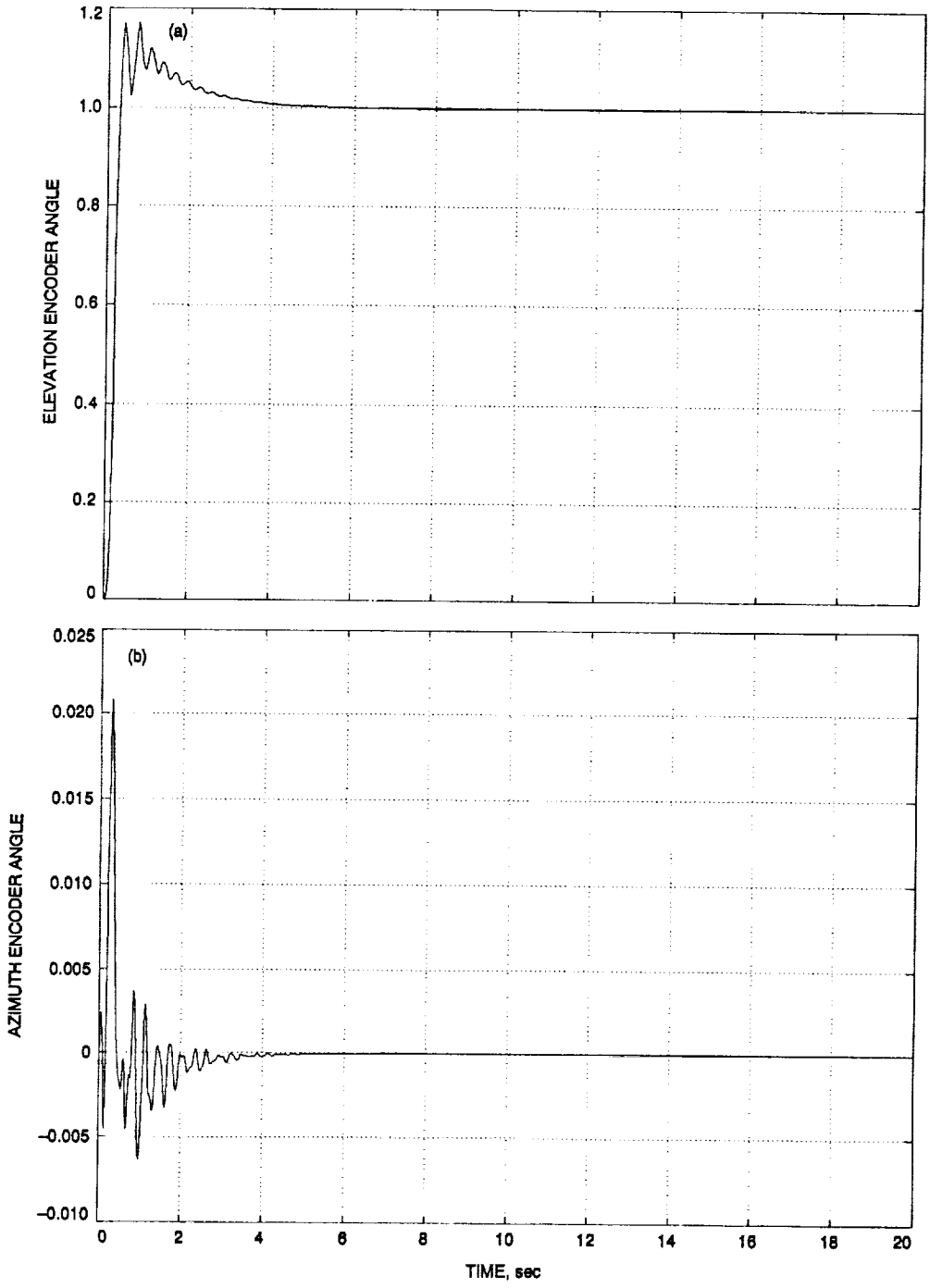


Fig. 20. Closed-loop response to step input for proportional weight 100 and integral weight 70 (both in azimuth and elevation), and weights for the flexible subsystem equal to 1×10^{-7}: (a) elevation encoder angle to elevation step command; (b) azimuth encoder angle to elevation step command; (c) azimuth encoder angle to azimuth step command; and (d) elevation encoder angle to azimuth step command.

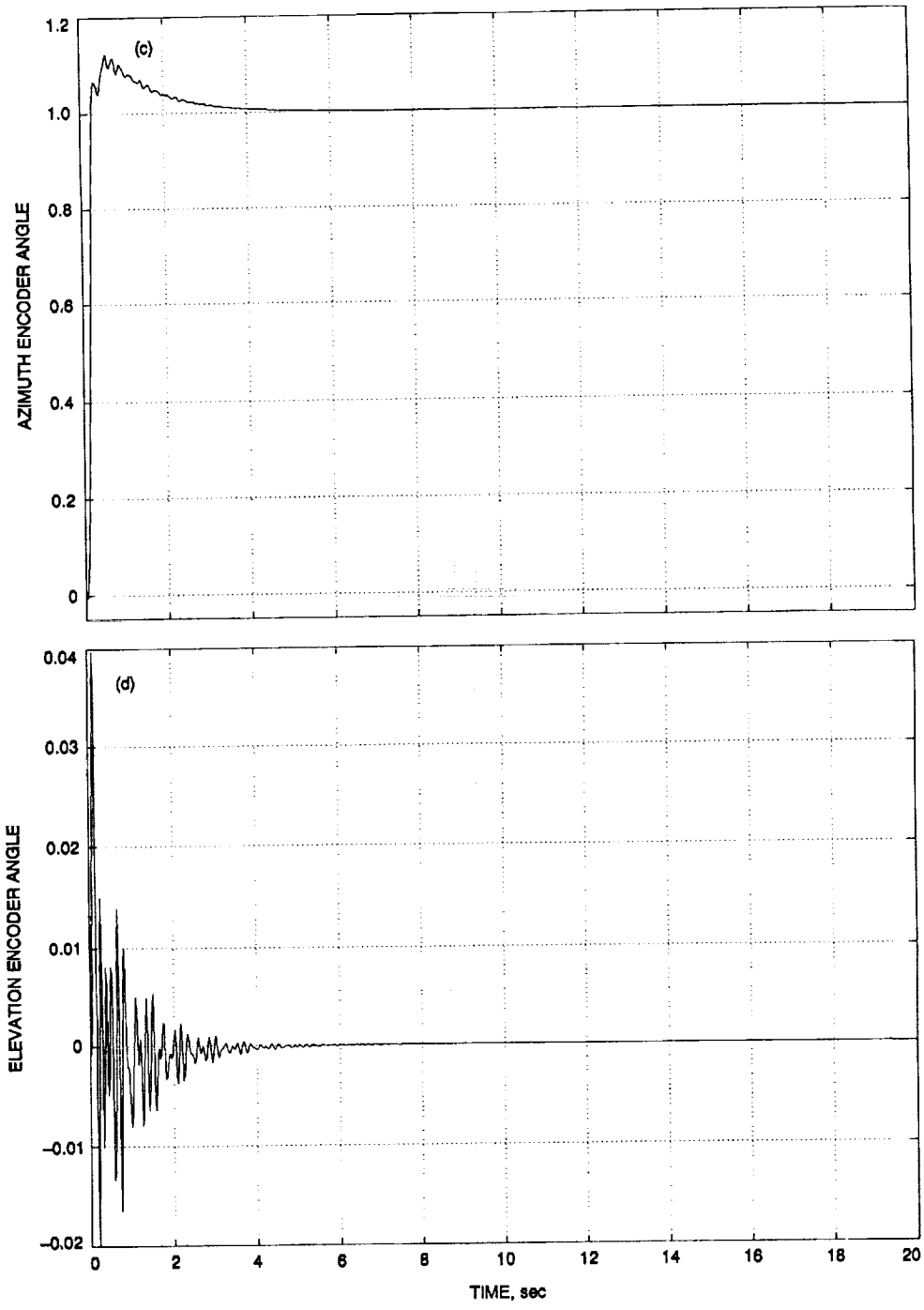

(c)
AZIMUTH ENCODER ANGLE
1.2
1.0
0.8
0.6
0.4
0.2
0
(d)
ELEVATION ENCODER ANGLE
0.04
0.03
0.02
0.01
0
−0.01
−0.02
0
2
4
6
8
10
12
14
16
18
20
TIME, sec

Fig. 20 (contd).

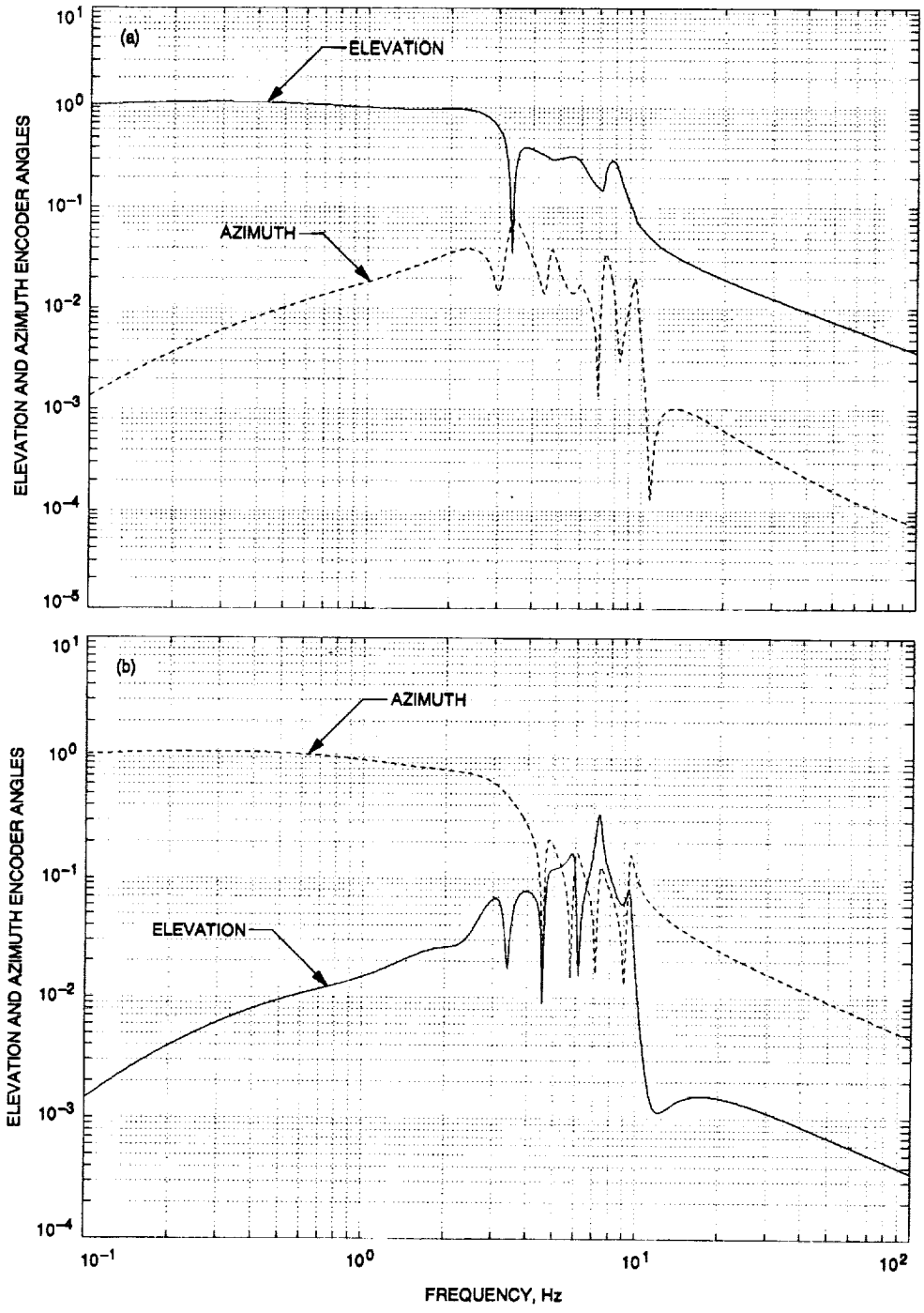


Fig. 21. Closed-loop transfer function for weights the same as those in Fig. 20: (a) elevation and azimuth encoder angles to elevation step command and (b) elevation and azimuth encoder angles to azimuth step command.

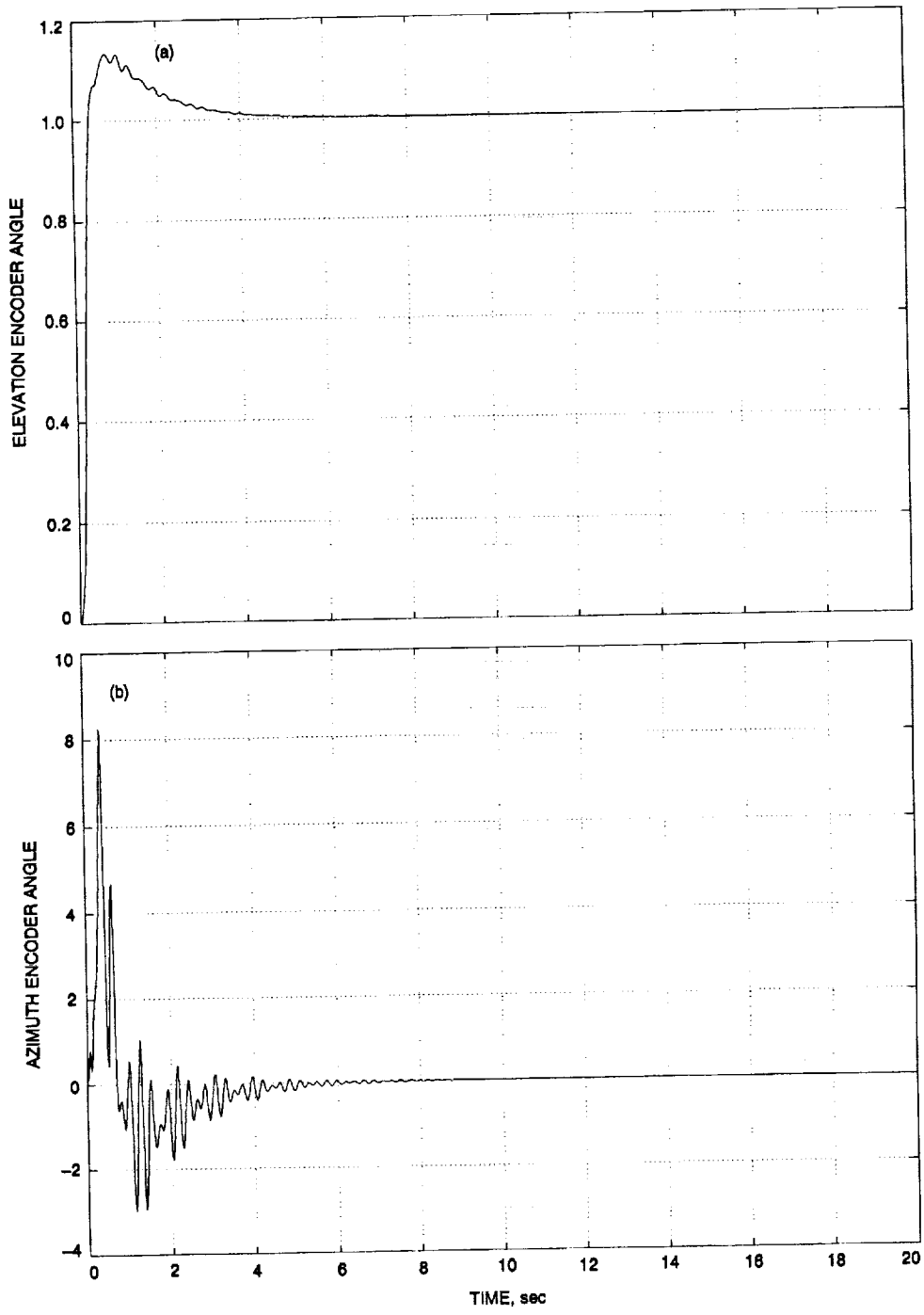


Fig. 22. Closed-loop response to step input for proportional weight 100 and integral weight 70 (both in azimuth and elevation), and weights for flexible subsystem $q_1 = q_2 = q_3 = q_4 = q_5 = q_6 = 1 \times 10^{-6}$, $q_7 = q_8 = 1 \times 10^{-7}$, and $q_9 = q_{10} = 1 \times 10^{-5}$: (a) elevation encoder angle to elevation step command; (b) azimuth encoder angle to elevation step command; (c) azimuth encoder angle to azimuth step command; and (d) elevation encoder angle to azimuth step command.

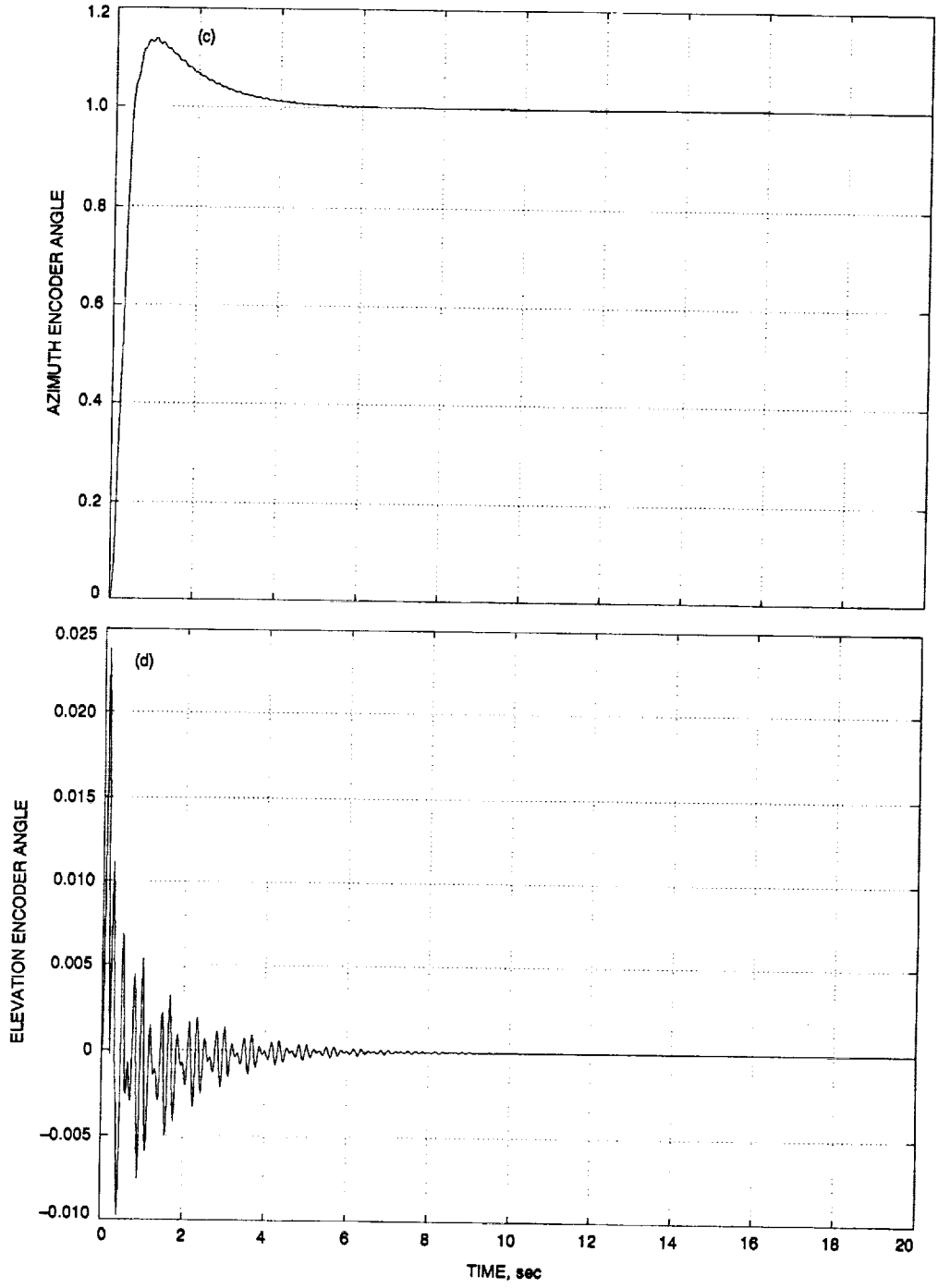

(c)
AZIMUTH ENCODER ANGLE
1.2
1.0
0.8
0.6
0.4
0.2
0
(d)
ELEVATION ENCODER ANGLE
0.025
0.020
0.015
0.010
0.005
0
−0.005
−0.010
0
2
4
6
8
10
12
14
16
18
20
TIME, sec

Fig. 22 (contd).

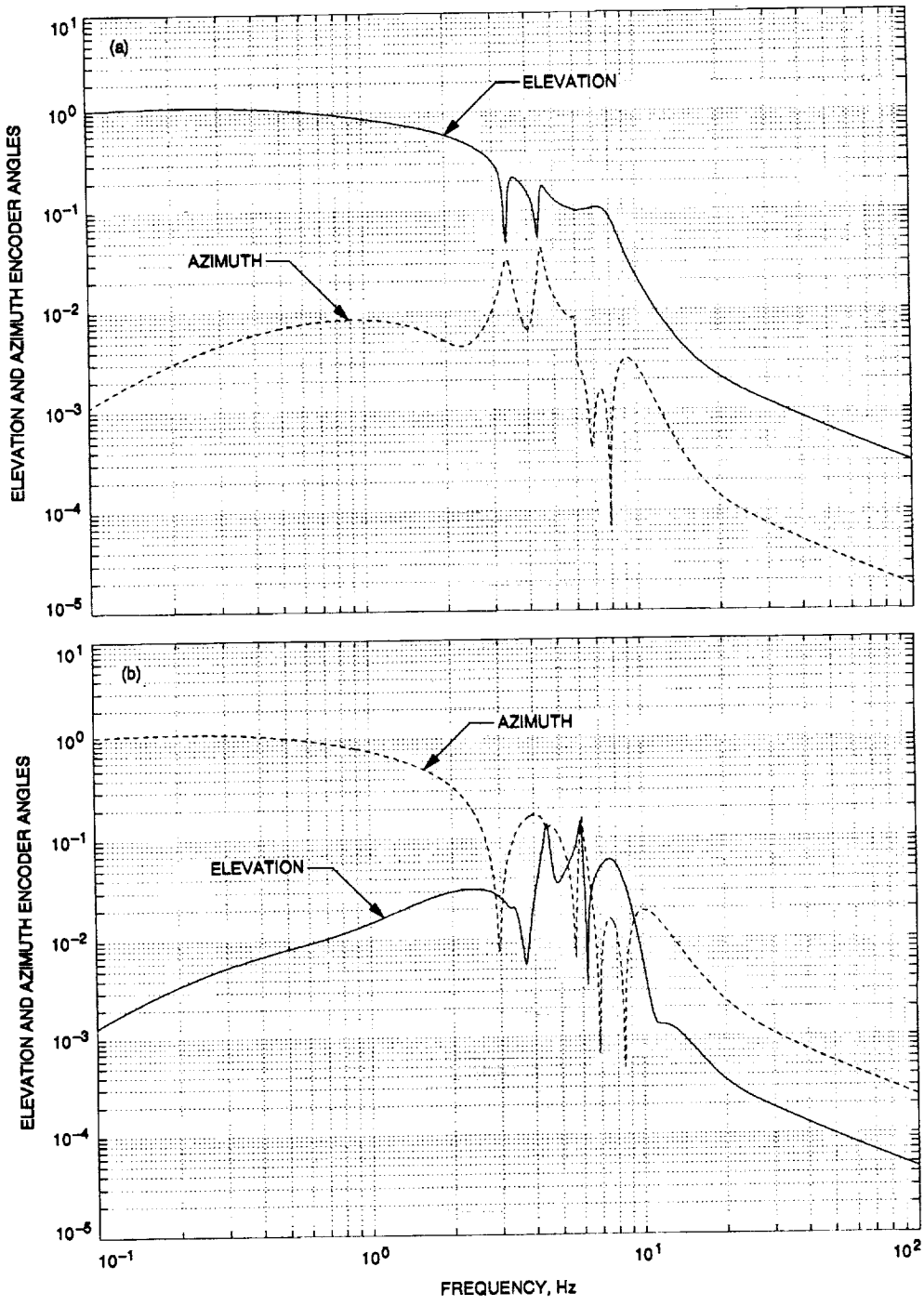


Fig. 23. Closed-loop transfer function for weights the same as those in Fig. 22: (a) elevation and azimuth encoder angles to elevation step command and (b) elevation and azimuth encoder angles to azimuth step command.

Appendix
Proofs

Proof of Proposition 1. For a flexible structure in the balanced representation, the state matrix A is diagonally dominant (with 2×2 blocks on the main diagonal), and for $R = I$ and Q as in Eq. (17), the solution S of the Riccati Eq. (9) is also diagonally dominant with 2×2 blocks S_i on the main diagonal:

$$S_i = s_i I_2, \quad s_i > 0, \quad i = 1, \cdots, n \tag{A-1}$$

Thus, Eq. (9) turns into a set of the following equations:

$$s_i(A_i + A_i^T) - s_i B_i B_i^T + q_i I_2 = 0, \quad i = 1, \cdots, n \tag{A-2}$$

For a balanced system $B_i B_i^T \cong -\gamma_i(A_i + A_i^T)$, see Eq. (16), and for $A_i + A_i^T = -2\zeta_i\omega_i I_2$, see Eq. (15). Therefore, Eq. (A-2) is now

$$s_i^2 + \gamma_i s_i - 0.5 q_i/\zeta_i\omega_i\gamma_i = 0, \quad i = 1, \cdots, n \tag{A-3}$$

There are two solutions of Eq. (A-3), but for $q_i = 0$ it is required that $s_i = 0$. Therefore, Eq. (19) represents the unique solution of Eq. (A-3).

Proof of Proposition 2. For small values of q_i, the matrix A of the closed-loop system is diagonally dominant: $A_o = \text{diag}(A_{oi})$, $i = 1, \cdots, n$, and

$$A_{oi} = A_i - B_i B_i^T s_i \tag{A-4}$$

By introducing Eq. (16) to Eq. (A-4), one obtains

$$A_{oi} = A_i + 2 s_i \gamma_i (A_i + A_i^T) \tag{A-5}$$

and introducing A_i as in Eq. (15) to Eq. (A-5) one obtains

$$A_{oi} = \begin{bmatrix} -\beta_i\zeta_i\omega_i & -\omega_i \\ \omega_i & -\beta_i\zeta_i\omega_i \end{bmatrix} \tag{A-6}$$

with β_i as in Eq. (18).

DSN Acquisition of Magellan High-Rate Telemetry Data

A. L. Berman and P. A. Au
TDA Mission Support and DSN Operations

The Magellan Project levied the stringent requirement of a 98-percent high-rate telemetry data capture rate on the Deep Space Network (DSN) during the Magellan Prime Mapping Mission. To meet this requirement, the DSN undertook extensive development of the DSN Telemetry System, as well as extensive DSN operations planning and test and training. In actuality, the DSN substantially exceeded the requirement by achieving a Prime Mapping Mission high-rate telemetry data capture rate of 99.14 percent.

This article details the DSN telemetry system development, and DSN operations planning and test and training. In addition, the actual high-rate telemetry data outages are comprehensively presented and analyzed.

I. Introduction

The preeminent science objective of the Magellan mission to Venus was the Synthetic Aperture Radar (SAR) mapping of 90 percent of the surface of Venus. To meet this objective, the Magellan Project levied a requirement on the Deep Space Network (DSN) to successfully acquire 98 percent of all available Magellan high-rate telemetry data (which contains the SAR mapping data). The DSN committed to a data capture rate of 96 percent and adopted a high-priority goal of achieving a 98-percent data capture rate.

In fact, the DSN achieved a composite Magellan high-rate telemetry data capture rate of 99.14 percent during the eight-month Magellan Prime Mapping Mission.

This article describes the DSN Telemetry System development and operations planning that allowed the DSN to capture over 99 percent of the available data. Also described are the detailed statistics of the DSN data outages.

II. Mission Overview

The Magellan spacecraft was launched on its 15-month trajectory to the planet Venus, from the Kennedy Space Center, at 11:46:00 Pacific Daylight Time (PDT) on May 4, 1989. Venus Orbit Insertion (VOI) occurred at 10:06:00 PDT (ground observed time) on August 10, 1990, but was not directly observed from Earth because the spacecraft was occulted by Venus. Shortly thereafter, at 10:06:38 PDT (ground observed time), the spacecraft emerged from behind Venus, which indicated that the spacecraft solid rocket motor burn had properly executed and had thereby placed Magellan into the desired (approximately) three-hour orbit about Venus.

Following the successful VOI, the Project began its in-orbit checkout (IOC) activities, which were nominally scheduled for three weeks, thus allowing a start of mapping operations on September 1, 1990. However, during the first IOC radar mapping test on August 17, the Magellan signal was lost and not fully recovered for approximately 12 hours. A second signal loss incident occurred on August 22, with the signal being recovered approximately 20 hours later.

During early September 1990, the Project continued to analyze the spacecraft signal disappearance incidents and took steps to lessen the likelihood of such incidents in the future. In the second week of September, the Project gradually restored the spacecraft to full functionality.

On September 15, 1990, at 9:29:42 PDT, the Magellan Project inaugurated mapping operations using the 34-meter antenna array at Goldstone. With the exceptions of superior conjunction (November 1990) and very occasional spacecraft malfunctions and Ground Control Team errors, DSN support of Magellan high-rate telemetry data, from September 15, 1990, through January 1992, has essentially been continuous.

III. DSN Telemetry System

The DSN Telemetry System at each Deep Space Communications Complex (DSCC) acquires, demodulates, decodes, and records the Magellan 268.8-Kbps high-rate telemetry data that contains the radar mapping data.

The major functions of the DSN Telemetry System can be stated as follows:

(1) Receive the radio frequency (RF) carrier and process to baseband.

(2) Demodulate the carrier and subcarriers.

(3) Perform symbol synchronization and data decoding.

(4) Perform frame synchronization to the transfer frame level.

(5) Format data and interlace status information.

(6) Generate Original Data Records.

(7) Transmit telemetry data from the Signal Processing Centers to the Space Flight Operations Center (SFOC) at JPL.

The DSN telemetry data flow path is shown in Fig. 1. The antenna collects the signal under the control of the Antenna Mechanical Subsystem. The low-noise maser amplifies the S-band (2297.962963-MHz) and/or X-band (8425.864197 MHz) signals in the Antenna Microwave Subsystem. The ultrasensitive S- and X-band receivers detect and downconvert the signals in the Receiver-Exciter Subsystem. The Telemetry Subsystem extracts the telemetry symbol streams from the downconverted signals, decodes the symbol streams into bit streams, and synchronizes them to the bit streams, which are the desired telemetry data. Finally, the Ground Communications Facility formats, records, transmits, and routes the telemetry data to the Project.

IV. DSN Telemetry System Development

The DSN began planning for Magellan mission support in 1982. It was almost immediately recognized that the Project intent of transmitting 268.8 Kbps of mapping data for two hours out of every three, continuously during the eight months necessary to fully map (one complete revolution of) Venus, would require very significant modifications to the then-existing DSN Telemetry System. The Magellan requirements that most significantly affected the DSN Telemetry System were as follows:

(1) The ability to simultaneously receive and process two coded downlink channels (one high-rate X-band and one low-rate S- or X-band). The high-rate X-band channel carries high-rate mapping telemetry, and the low-rate S- or X-band channel carries low-rate real-time engineering telemetry.

(2) The ability to receive, process, record, and simulate a high-rate coded telemetry downlink with a data rate of 268.8 Kbps.

(3) The ability to perform telemetry data acquisitions from the receiver through the frame synchronizer within one minute.

Figure 2 provides a more detailed diagram of the DSN Telemetry System at a DSCC and illustrates the assembly-level modifications that were necessary to support Magellan telemetry requirements.

To support a Magellan telemetry data acquisition requirement of one minute, extensive modifications were necessary for both the receiver and the baseband assembly (BBA), which acquires the telemetry subcarrier and performs the signal demodulation and symbol synchronization. In the receiver, a fast fourier transform (FFT) algorithm was added, which allows a ±1-KHz X-band signal uncertainty band (which encompasses the Magellan uncertainties) to be searched, with subsequent signal lockup in less than 20 seconds. In the BBA, the subcarrier acquisition bandwidth was broadened from ±0.75 to ±5.5 Hz

to encompass the Magellan uncertainties. An FFT-like acquisition algorithm was added, which initiates a subcarrier search utilizing a frequency loop; upon detection, the subcarrier is then switched to a phase locked loop (PLL) for subsequent tracking. Finally, hardware modifications were implemented in the BBA to increase overall processing speed. These BBA modifications provide subcarrier acquisition and processing in less than 40 seconds. Thus, the combination of receiver and BBA performance allows telemetry acquisition to be accomplished within the one-minute Magellan requirement, as compared with a typical premodification acquisition time of approximately 5–20 minutes.

To meet the Magellan requirement of a high-rate downlink data rate of 268.8 Kbps, the Maximum Likelihood Convolutional Decoder (MCD) assembly was modified to provide higher processing speed; the new maximum MCD data rate has been doubled from 135 to 270 Kbps. In addition, a second MCD was added to allow the simultaneous reception and processing of two downlink coded channels, as per the Magellan requirement.

A Frame Synchronizer subassembly was added to the DSN Telemetry Subsystem, thereby giving greater monitoring capability to DSN personnel at the DSCC's and the Network Operations Control Center. Finally, to record the high-rate mapping data, a new Digital Data Recording (DDR) assembly was added to the Ground Communications Facility. The new DDR increases the Telemetry System (single) magnetic tape ten-minute record capacity at 268.8 Kbps to over one hour, primarily by switching from 1650-bits/sec tape drives to 6250-bits/sec tape drives. In addition, the new DDR produces two tape copies simultaneously, allowing one copy to be air-freighted to the Magellan Project personnel in Pasadena, while the second remains at the DSCC as a backup in the event of a loss during shipment.

V. Orbital Operations Preparation

During the first half of 1990, intensive DSN and Magellan Project test and training activities were performed to prepare for orbital operations and, more specifically, for acquisition of the Magellan high-rate telemetry data. For the DSN, the most significant activities were the mission readiness test program and the cruise mapping test (CMT).

During January 1990, the DSN embarked on a four-month test and training program to prepare for mapping operations. Four major DSN mapping phase antenna configurations, including (1) 70-m stand-alone support, (2) 70-m and 34-m high-efficiency (HEF) X-band uplink support, (3) 34-m standard and 34-m HEF arrayed support, and (4) 70-m dual X-band subcarrier support, were tested several times at each DSCC. At the conclusion of these tests, on May 10, 1990, the DSN was adjudged to be operationally ready to support Magellan Mapping Operations.

During May 21–25, 1990, the Project performed the CMT, which was the final, full dress rehearsal for the upcoming mapping operations. During the four days of the CMT, the spacecraft continuously simulated all the events and high-rate telemetry data playbacks as if it were in a three-hour Venus orbit. DSN performance during this test was exemplary; 94 percent of all acquisitions were performed within the one-minute requirement, and 97.7 percent of the high-rate telemetry (simulated mapping) data was successfully acquired by the DSN. This high level of performance indicated that all DSN systems were operating correctly, and that the DSN operations personnel had achieved a high level of mapping operations proficiency.

VI. Telemetry System Performance During the Magellan Prime Mapping Mission

The Magellan Prime Mapping Mission began on September 15, 1990, and continued to May 15, 1991. During that time, one or more DSN stations continuously supported the Magellan spacecraft, with the exception of an approximately ten-day period centered about solar superior conjunction (November 2, 1990). During each 3.26-hour orbit, radar mapping data were acquired and recorded for approximately 37 minutes by the Magellan spacecraft at 804.6 Kbps. Subsequently these data were replayed (within the same orbit) in two separate, approximately 57-minute, playback periods at 268.8 Kbps. At the start of each playback period, the DSN had one minute available from the time of signal appearance to acquire the signal and lock up the end-to-end DSN Telemetry System to avoid the loss of any radar mapping data.

During the Prime Mapping Mission, the DSN kept detailed records, on a day-by-day basis, of the total minutes of data transmitted by the spacecraft and the total minutes of data lost due to failures within the DSN Telemetry System. For the entire Magellan Prime Mapping Mission, the high-rate telemetry data acquisition performance by the DSN was as follows:

Total data available = 172,509 min

Total data lost (all DSN) = 1,478 min

DSN data capture rate = 99.14 percent

The above calculations exclude (from total data available) all losses attributable to either spacecraft problems, Ground Control Team errors, solar effects, or terrestrial weather. Interim reports on the DSN cumulative data capture rate were published on a weekly basis (more frequently during the first month), and can be seen in Fig. 3. As can be seen in that figure, performance throughout the Magellan Prime Mapping Mission was remarkably stable, with the slight, but pronounced, upward trend probably attributable to increased proficiency on the part of station operations personnel. The average (single-incident) data loss was slightly over six minutes, with typically five occurrences per week, throughout the entire DSN. During the Prime Mapping Mission, there were only five failures that resulted in data losses greater than 30 minutes, and these are characterized in Table 1. Total data losses were substantially reduced by employing considerable redundancy; approximately 70 percent of the playbacks were supported by two totally independent Telemetry Systems (antenna through data recording), and even when only one antenna was engaged, approximately 90 percent of these (single-antenna) playbacks were supported by two independent sets of telemetry data-processing equipment (demodulation through recording).

To further characterize the nature of the data outages, the data outages were divided into "large" and "small" regimes, defined as follows:

$$\text{"large"} : 30 \text{ min} \leq \text{outage}$$

$$\text{"small"} : 30 \text{ sec} \leq \text{outage} \leq 30 \text{ min}$$

Figure 4 shows the cumulative "large" data loss total throughout the primary mapping mission, while Fig. 5 shows the average "large" incidents per month, and Fig. 6 shows the average "large" data loss per incident. By the end of the Prime Mapping Mission, it is seen that "large" losses accounted for a total data loss of slightly under 0.30 percent, with an average occurrence per month of approximately 0.7, and an average duration of approximately 90 minutes per incident. These are basically random occurrences, and these values probably would be characteristic of continuous DSN support of any flight project.

Figure 7 shows the cumulative "small" data loss total throughout the Prime Mapping Mission, while Fig. 8 shows the average "small" incidents per day, and Fig. 9 shows the average "small" data loss per incident. By the end of the Prime Mapping Mission, it is seen that "small" losses accounted for a total data loss of slightly under 0.60 percent, with an average occurrence per day of approximately 0.75, and an average duration of approximately 6 minutes per incident. Probably the most significant aspect of the "small" data losses is the rather smooth decrease in daily incidence from approximately 2.0 at the start of mapping, to approximately 0.75 at the conclusion. Undoubtedly, at least a significant portion of the decrease in "small" incidents is due to increase in operator proficiency, as the complex Magellan operational sequence was repeated many times each day at each DSCC.

Finally, statistics were kept on data unavailability due to terrestrial weather. Weather incidents in general were wind, rain, or snow. Figure 10 shows the cumulative weather data unavailability throughout the Prime Mapping Mission, while Fig. 11 shows the average data unavailability per weather incident. By the end of the Primary Mapping Mission, it is seen that weather accounted for a total data unavailability of approximately 0.20 percent, with an average data unavailability per incidence of approximately 25 minutes.

Table 1. Discrete DSN data losses in excess of 30 min.

Date	Complex	Loss, min	Failure characterization
October 1, 1990	Madrid	122	Hardware—complex-wide power failure
November 12, 1990	Canberra	147	Hardware—complex-wide failure of frequency and timing system
December 10, 1990	Canberra	48	Documentation—error in predicted acquisition frequency
April 10, 1991	Madrid	44	Hardware—antenna azimuth decoder power supply failure
April 25, 1991	Goldstone	108	Procedural—mislabled cables caused loss of connection to data recorders

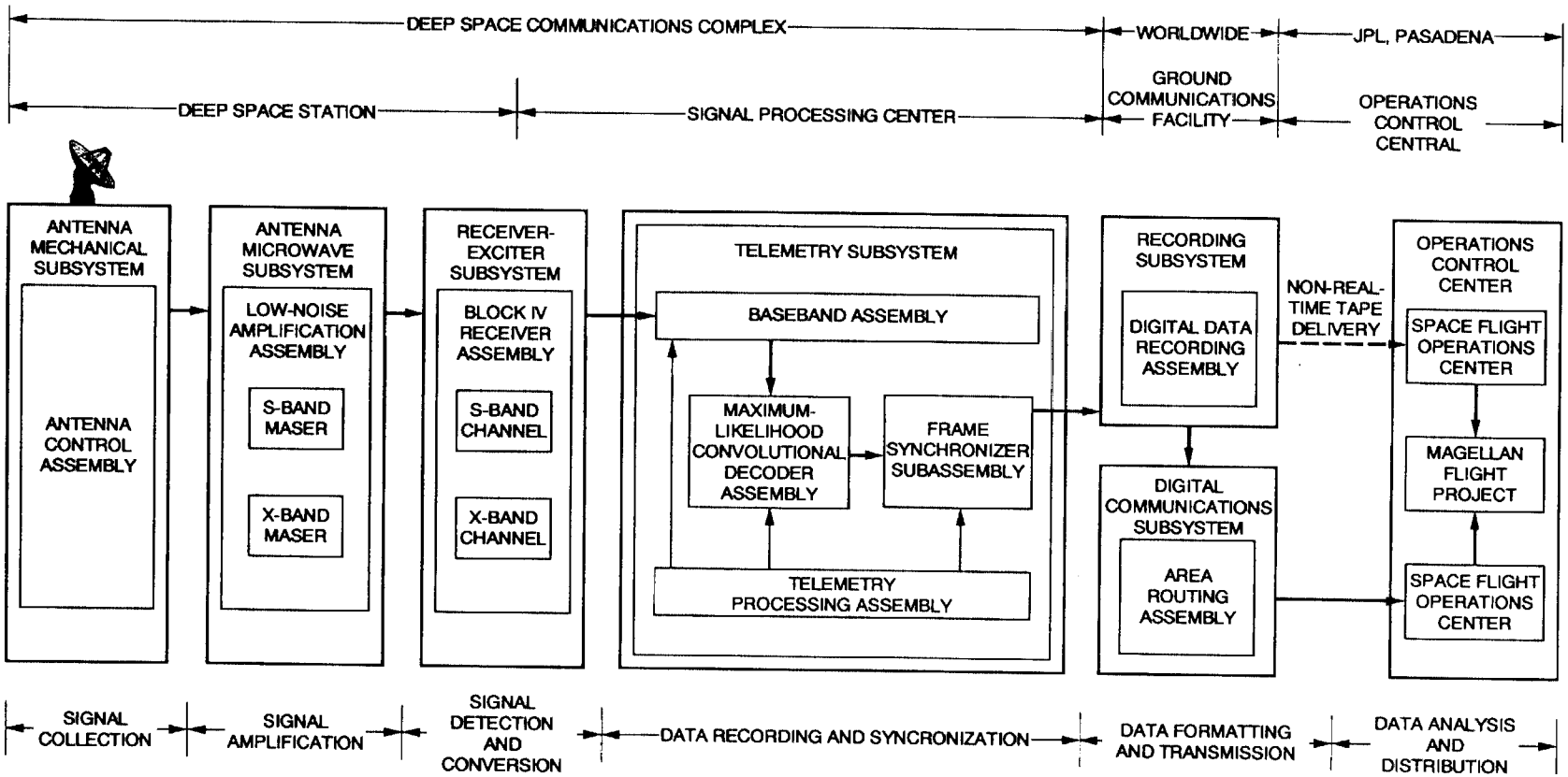


Fig. 1. Network telemetry data flow path.

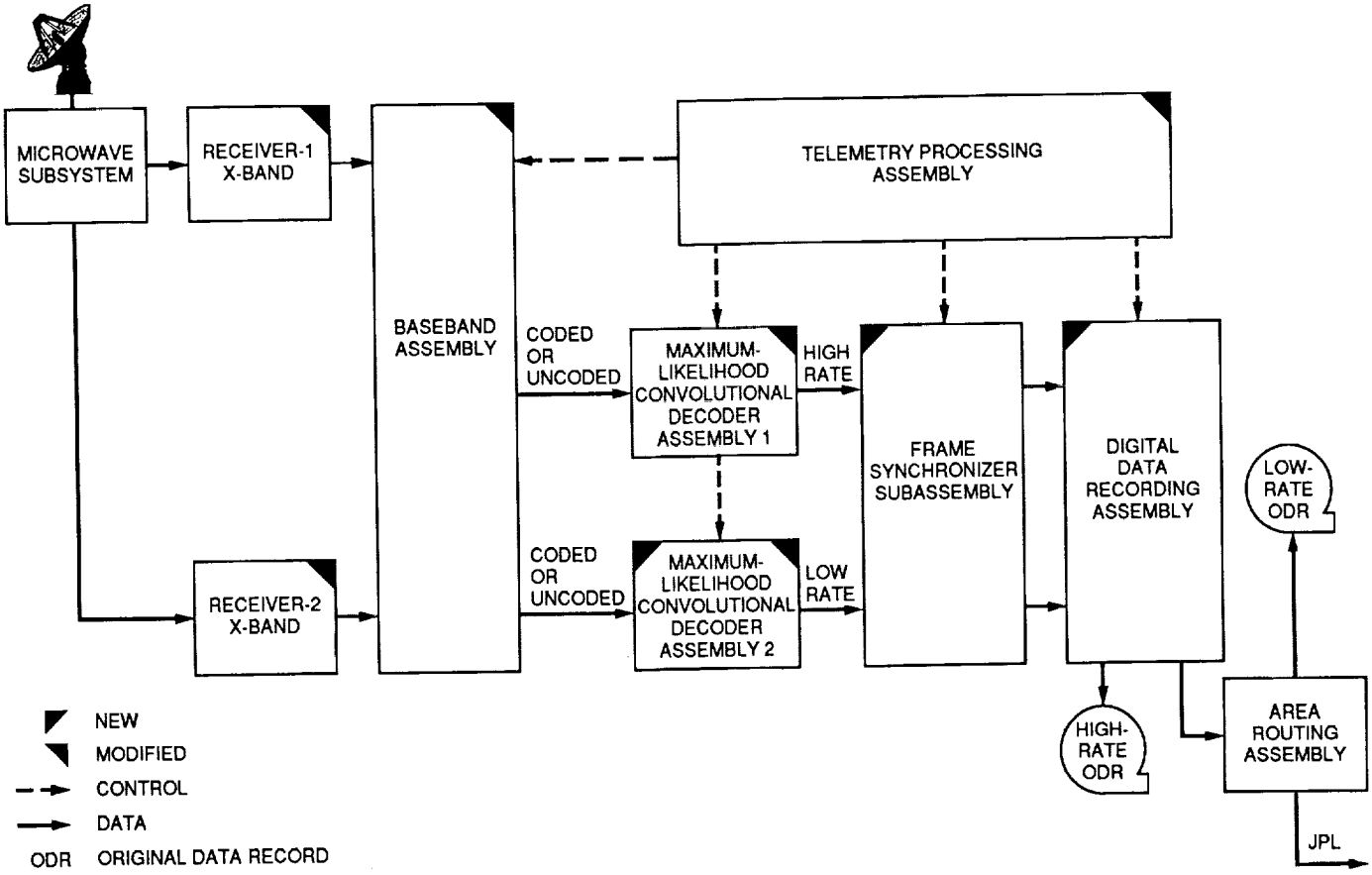


Fig. 2. Telemetry block diagram.

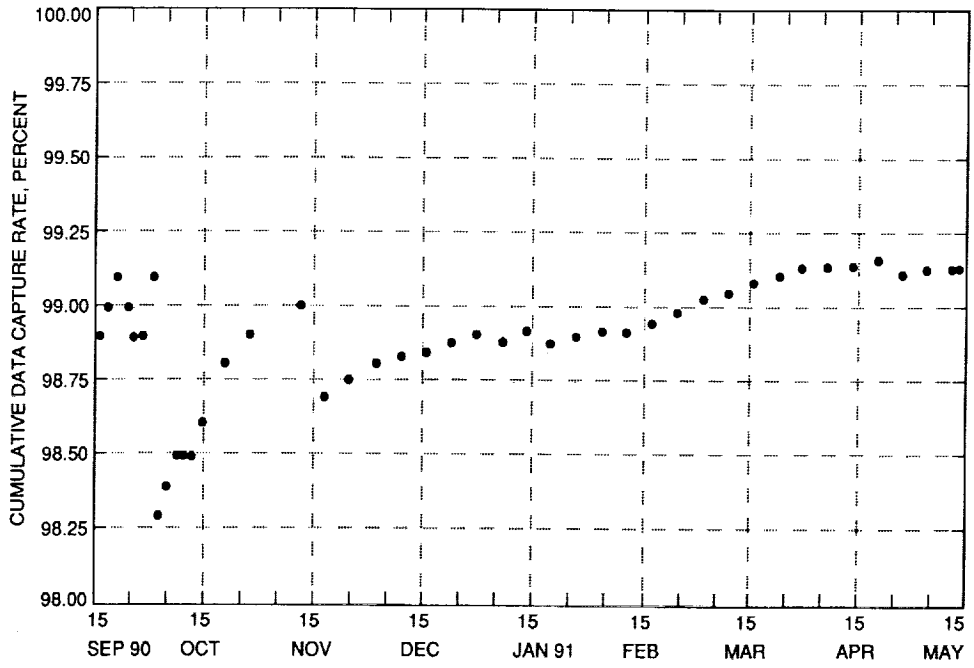


Fig. 3. Magellan high-rate telemetry cumulative data capture rate.

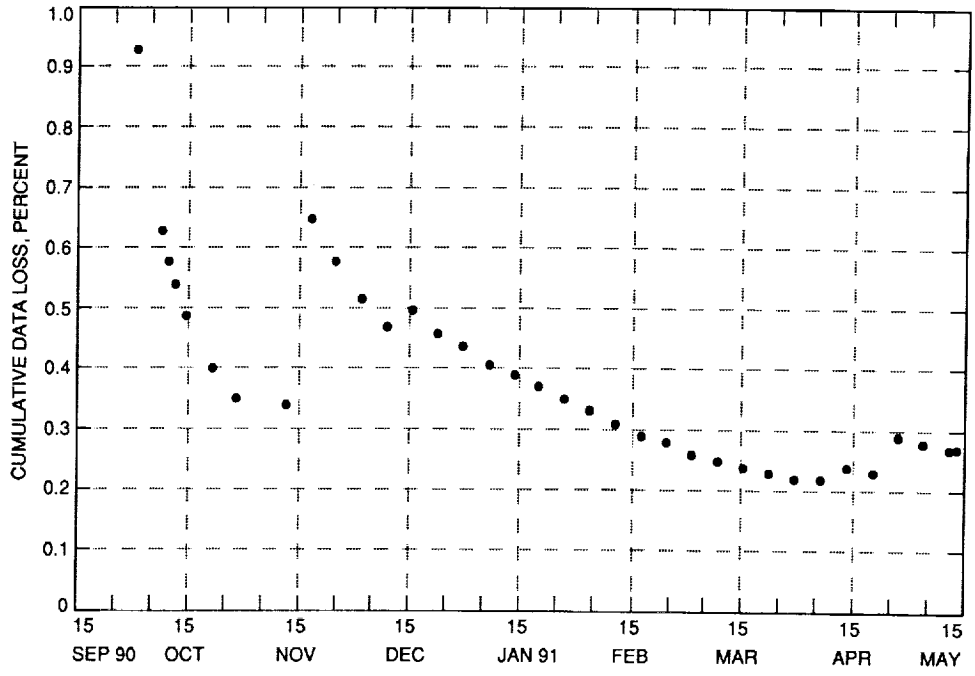


Fig. 4. Magellan high-rate telemetry, "large" data losses >30 minutes.

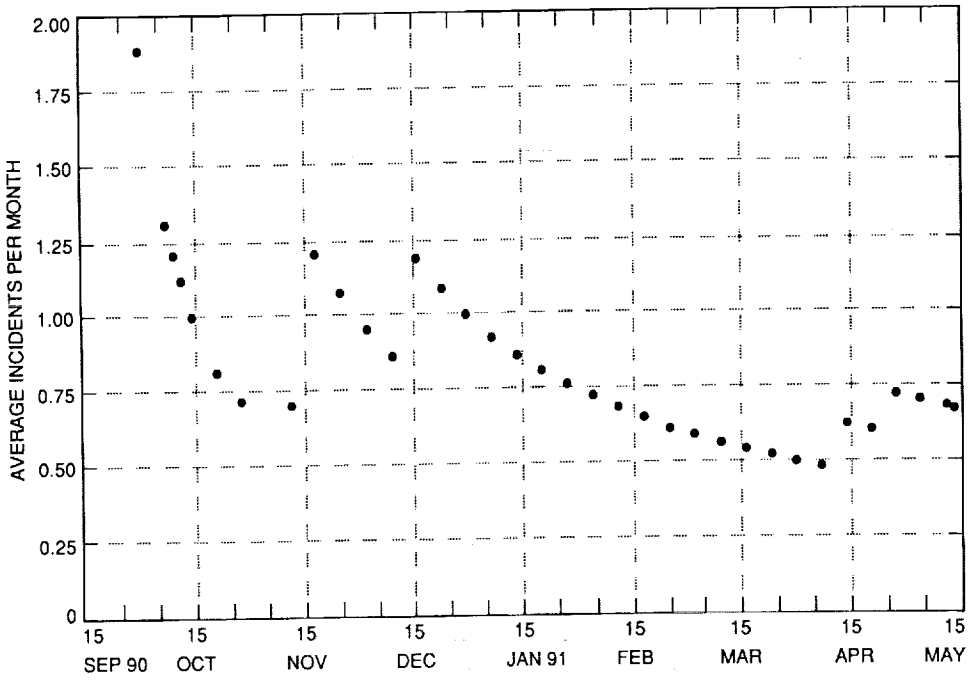


Fig. 5. Magellan high-rate telemetry, average "large" incidents per month.

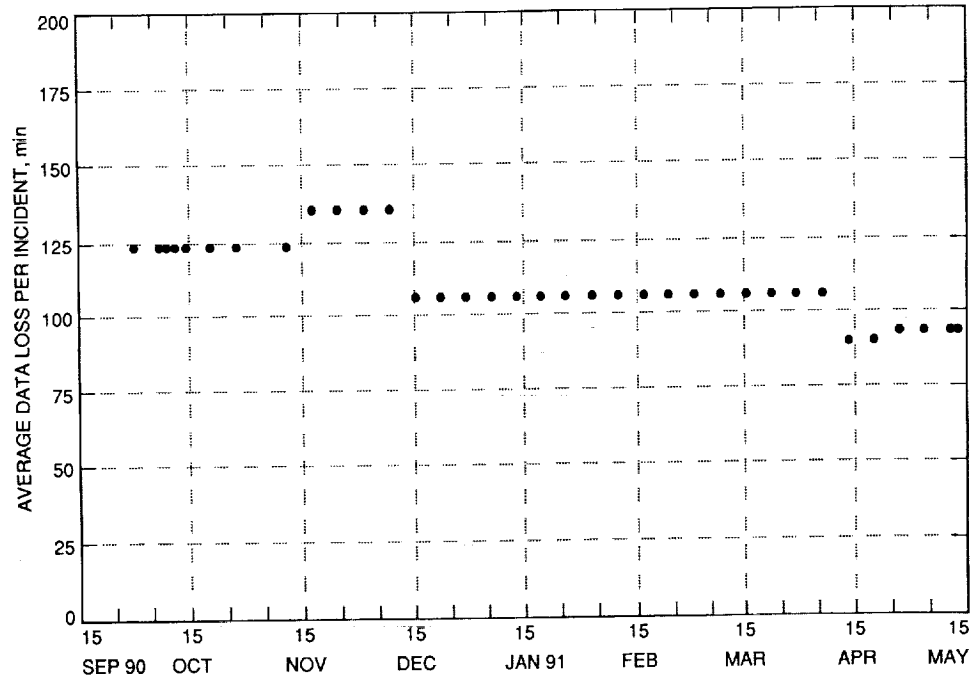


Fig. 6. Magellan high-rate telemetry, average "large" data loss per incident.

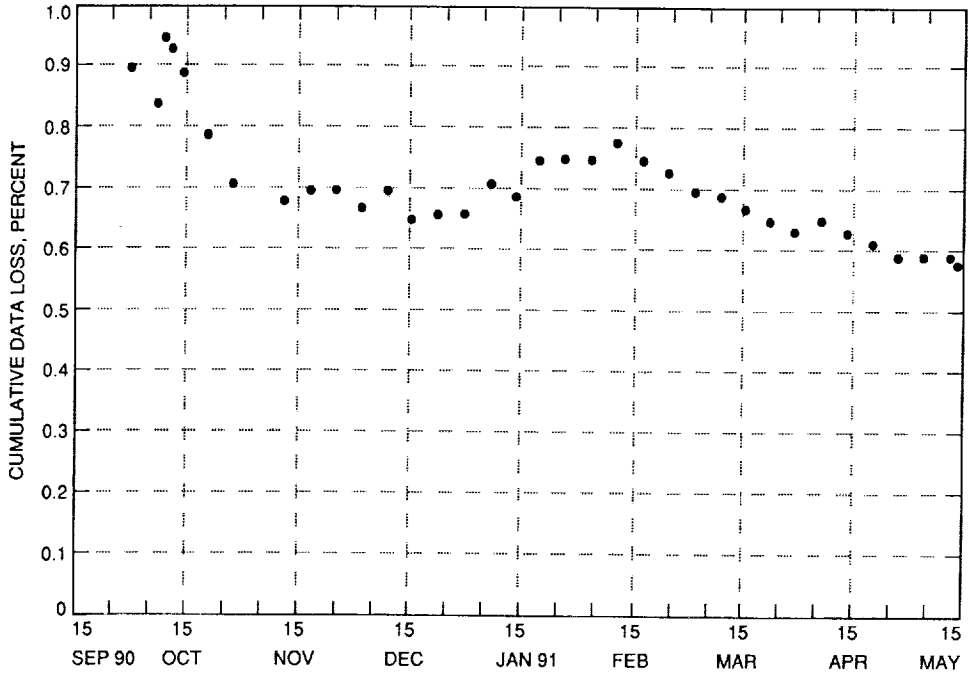


Fig. 7. Magellan high-rate telemetry, "small" data losses (<30 minutes).

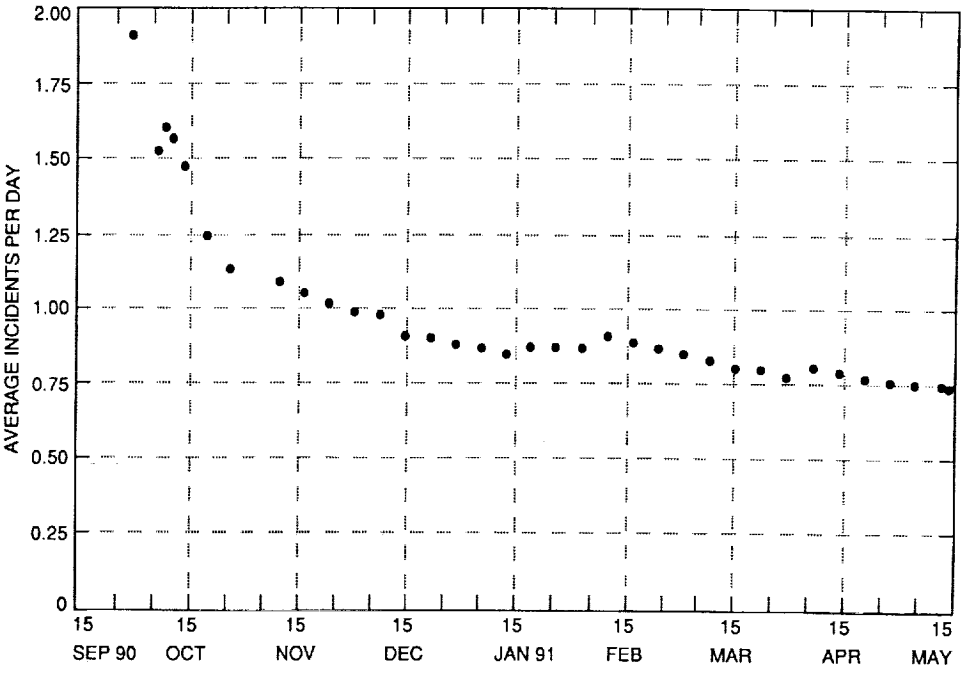


Fig. 8. Magellan high-rate telemetry, average "small" data incidents per day.

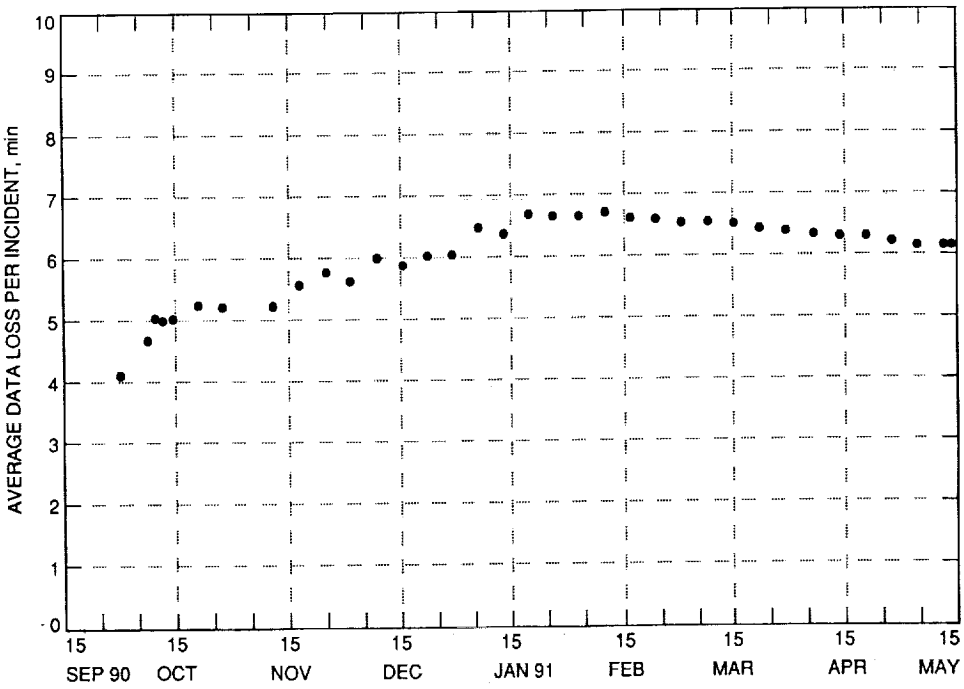


Fig. 9. Magellan high-rate telemetry, average "small" data loss per incident.

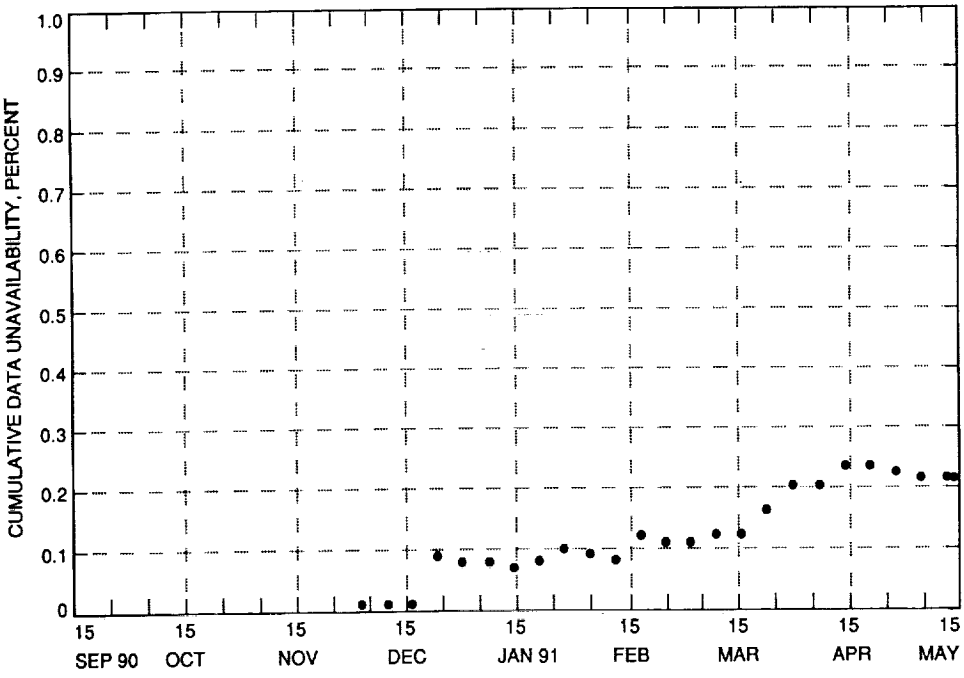


Fig. 10. Magellan high-rate telemetry, weather-induced data unavailability.

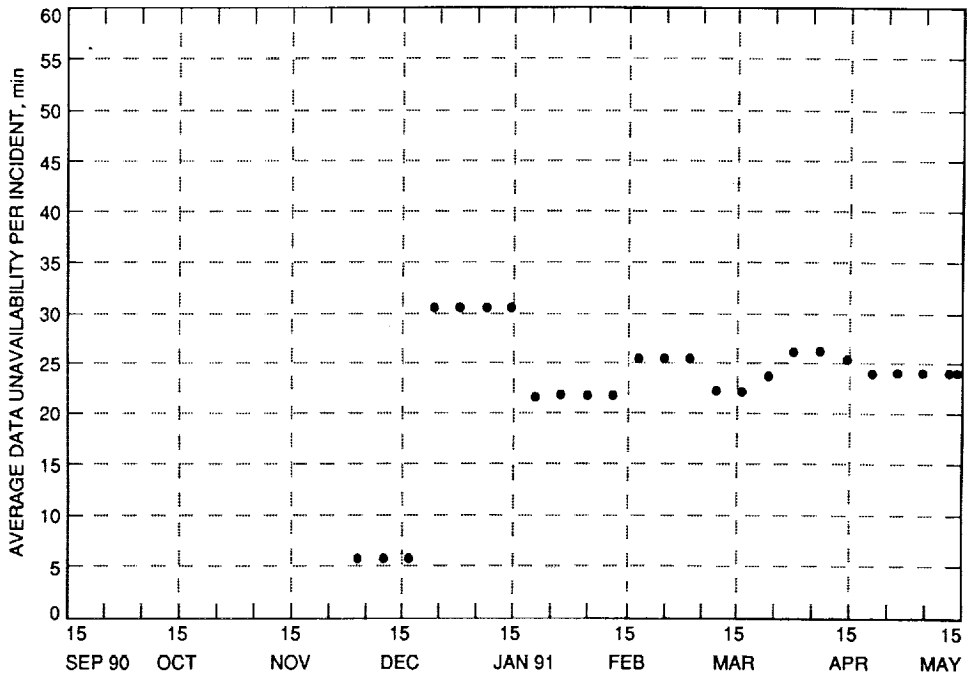


Fig. 11. Magellan high-rate telemetry, average weather-induced data unavailability per incident.

C-4

Galileo Early Cruise, Including Venus, First Earth, and Gaspra Encounters

P. E. Beyer and R. C. O'Connor
TDA Mission Support and DSN Operations

D. J. Mudgway[1]

This article documents DSN support for the Galileo cruise to Jupiter. The unique trajectory affords multiple encounters during this cruise phase. Each encounter had or will have unique requirements for data acquisition and DSN support configurations. An overview of the cruise and encounters through the asteroid Gaspra encounter is provided.

I. Introduction

The Galileo mission is designed to make long-term investigations of the Jovian system by using a spacecraft comprised of a Probe and an Orbiter. After being released on the initial approach to Jupiter, the Probe will enter the Jovian atmosphere and make in situ measurements; its data will be relayed to Earth by the Orbiter. The Orbiter will then enter orbit around Jupiter for an approximate two-year, ten-orbit tour of the Jovian system. This article will only address DSN support and operations during the early cruise phase, through the asteroid Gaspra encounter in October 1991. Earth-2 encounter, asteroid Ida encounter, Probe support, and Jupiter operations will be the topics of future articles. Figure 1 is included as a mission overview.

The Galileo spacecraft (Fig. 2) was launched on October 18, 1989, on the Space Shuttle Atlantis (STS-34). A two-stage Inertial Upper Stage (IUS) rocket was used to place the combined Orbiter and Probe on a Venus–Earth–Earth gravity assist (VEEGA) trajectory to Jupiter. The Galileo VEEGA trajectory is shown in Fig. 3. Following is a description of the Galileo cruise and the included DSN support, as well as planetary and asteroid encounter support during the cruise period.

The Galileo interplanetary cruise has been divided into the following phases for planning purposes:

(1) Earth–Venus cruise: October 29, 1989–February 19, 1990.

(2) Venus–Earth cruise: February 19, 1990–April 29, 1991.

(3) Earth–Earth cruise: April 29, 1991–March 15, 1993.

(4) Earth–Jupiter cruise: March 15, 1993–October 9, 1995.

Thus, the Venus encounter on February 9, 1990, occurred in Earth–Venus cruise; Earth-1 encounter on December 8, 1990, occurred in Venus–Earth cruise; the Gaspra asteroid encounter on October 29, 1991, occurred in Earth–Earth cruise; Earth-2 encounter on December 8,

[1] Consultant to the TDA Mission Support and DSN Operations Office.

1992, will also occur in Earth–Earth cruise; and the potential for an Ida asteroid encounter will occur in Earth–Jupiter cruise. The final decision to commit to an Ida encounter has not been made at the time of this writing.

II. DSN Configuration for Galileo Launch and Initial Acquisition

The first acquisition of the S-band (2295-MHz) downlink signal from the spacecraft by DSS 16, on October 18, 1989, marked the beginning of full DSN operational support for Galileo. This most significant event was the culmination of more than ten years of planning, and the start of the DSN support for the 1989 Galileo mission to Jupiter, which is planned to continue to November 7, 1997 (end of Prime Mission).

The DSN configuration required for Galileo launch and cruise support in accordance with the Support Instrumentation Requirements Document and the NASA Support Plan commitments is shown in Fig. 4. This configuration includes all the stations of the network that are required for support at various stages of the Galileo mission. The 26-m antennas performed a critical role in the initial acquisition sequences and the first Earth encounter in December 1990. They are not able to provide support again until the second Earth encounter in December 1992. Because the Galileo high-gain antenna (HGA) remained stowed for the first one and a half years of the mission, the early phases of the mission relied on one or the other of the two low-gain antennas (LGA's). Resulting minimal telecommunications link margins necessitated the use of the DSN 70-m antennas within a few weeks of launch to achieve the relatively high data rates desired for spacecraft and science instrument characterization.

The current DSN capabilities are adequate to support Galileo telemetry data rates from 10 bps to 134.4 kbps (convolutionally coded) and command at 32 bps (Manchester coded). Until the HGA is deployed, the spacecraft can support only S-band downlink and S-band uplink. When the HGA becomes available, an operational X-band (8415-MHz) downlink will be used for telemetry, and an X-band uplink capability will be demonstrated for command use.

Navigation and radio science data will be provided by the DSN radio metric systems by using Doppler, ranging, and very long baseline interferometery (VLBI) techniques. A new, high-precision navigation data type called delta differential one-way range (ΔDOR) will also be used by the DSN on this mission.

Wideband digital communications links among all three complexes and the Network Operations Control Center (NOCC) at JPL will allow the real-time transmission of telemetry data to the project at the maximum Galileo data rate of 134.4 kbps.

The NOCC has the capability to provide the project with permanent data records in tape or file form, from which errors and omissions have been removed in accordance with an agreed level of criticality. These are called Intermediate Data Records (IDR's) and are used by the project for further processing to extract engineering or science data and for archival purposes.

A period of intense test and training exercises carried out over the preceding nine months had prepared the DSN operations personnel for the critical initial acquisition phase of the mission. Because this was a Shuttle launch combined with an IUS injection vehicle, a high degree of coordination and technical interfacing with other centers was involved. To prepare for this, the DSN participated in the Galileo Joint Integrated Simulation Tests with Kennedy Space Center (KSC), Johnson Space Center, and the Air Force Consolidated Space and Test Center (CSTC). In these tests, state vectors were received from CSTC and processed by the DSN Navigation Team to produce predictions for eventual use by the initial acquisition station in making its first contact with the spacecraft. Data transfers were made in a timely manner and delivered to the correct destinations.

A Galileo/STS-34 Readiness Review was held on August 31, 1989, and the Review Board's assessment based on these results confirmed that the DSN was ready to support the launch.

A final Operational Readiness Test at Goldstone on September 15, using DSS 12 and DSS 16 in the cross-support mode, and a final proficiency test at DSS 46 to verify some last-minute modifications to the receiver/exciter on September 17, allowed the DSN to declare "ready for launch support" status. STS-34 was successfully launched from KSC on October 18, 1989, at 16:53:40 UTC, and the subsequent deployment of the IUS and Galileo spacecraft from the shuttle cargo bay on revolution 5 (rev. 5) was nominal.

Following deployment, all events of concern to the DSN were controlled by a very detailed Initial Acquisition Plan that provided DSN parameters for all possible combinations of launch time and actual deployment and injection conditions. The critical events in this period are shown in the IUS/Galileo nominal trajectory profile of Fig. 5. In

accordance with this plan, initial acquisition of the IUS S-band signal prior to IUS separation from the spacecraft was accomplished by DSS 16, followed 20 minutes later by acquisition of the Galileo S-band downlink first at DSS 16, and immediately thereafter at DSS 12. This provided continuous telemetry coverage of the critical IUS/Galileo separation event, which occurred 10 minutes later. The timely delivery of all the required state vectors from CSTC to the NOCC, enabled the DSN Navigation Team to run the necessary pointing predictions and transmit them to the stations for use in these time-critical sequences.

III. Earth–Venus Cruise

With the critical DSN-to-spacecraft telecommunications link firmly established in the two-way mode, the DSN settled down to continuous tracking passes at high data rates, while the project set about verifying and conditioning the spacecraft in accordance with the Earth–Venus (E–V) timeline shown in Fig. 6.

A command memory load consisting of 128 elements was uplinked over DSS 61 on October 25, and a Probe checkout over DSS 63, DSS 43, and DSS 14 on October 26 at 28.8 kbps captured 26.7 hours of Probe data. The first of many calibrations of the spacecraft radio-frequency system, called the radio-frequency system automatic gain control (RFSAGC) was carried out on October 27 over DSS 63. These events and the prevailing data rates are also shown in Fig. 6, together with a profile of DSN tracking required to support them.

As the Earth-to-spacecraft range continued to increase, the telemetry signal-to-noise ratio (SNR) began to fall rapidly, which necessitated a continuous demand for 70-m support to sustain even a 1200-bps telemetry data rate. Galileo's dependency on 70-m support created severe conflicts for other DSN users, and conflict resolution became a continuing critical activity. By mid-November the spacecraft changed from LGA1 to LGA2 in order to sustain the 1200-bps telemetry rate needed for the Trajectory Correction Maneuver 2 (TCM-2) and science instrument calibrations and checkout, which were scheduled for the month of December. TCM-2 was supported by the DSN without any problems and excellent radio metric data were provided to the Galileo Navigation Team.

On December 13 in the midst of this intense 70-m activity, the right-side inboard elevation bearing on the DSS-63 antenna failed, and the antenna was immobilized. The DSN responded immediately, and a team of JPL engineers arrived on-site on December 15 to start repair work. Under heavy pressure because of the rapidly approaching Venus encounter and working under extremely difficult winter conditions, the Team was able to report that the bearing replacement was complete and that the antenna was operational on January 25, 1990. This outage required Galileo to replan the science instrument checkout events, but no science data were lost, since active science events had not started at that time.

By the beginning of January 1990, the Project reported that TCM-2, which had been executed in December, was very satisfactory, based on the radio metric data that had been analyzed since then. These data were being used to refine the Venus targeting parameter for the Venus flyby in February 1990. The replanned science instrument calibrations executed during the Christmas period had been replayed over DSS 43 by then, and were found to be quite satisfactory despite some minor anomalies. Because of DSS-63 downtime, Galileo tracking time had been negotiated to minimize data loss through January 31, 1990. DSN support with the remainder of the Network continued to meet minimum requirements, using DSS 61 as a fill-in for DSS 63.

On January 1, DSS 15 returned to operational status, and the first of the 34-m X-band transmitters on this antenna was expected to be operational by January 15. This capability on all three high-efficiency (HEF) antennas had been committed for support later in the Galileo mission. The 34-m X-band uplink was to be used for a command demonstration. It was also considered to be a vital capability for the Gravitational Wave Experiment in 1994 and 1995. The 34-m HEF antennas were also to be used to generate high-precision (50-nanoradian) ΔDOR navigation data.

However, analysis had shown that a ΔDOR demonstration on S-band (using the ΔDOR tones already implemented in the spacecraft radio system for the later demonstration on X-band) could yield significantly improved navigation data for the Earth 1 encounter, which would occur well before the planned X-band became available from the spacecraft in May 1991. Consequently, a set of seven demonstration passes of S-band ΔDOR was planned, the first of which was scheduled for January 9 and would use the baseline pairs between DSS 14 and DSS 65, and DSS 14 and DSS 43.

DSN support for Galileo continued to work around the DSS-63 outage by relying heavily on support from DSS 61. The DSN status for the Venus encounter was now conditional upon DSS 63 returning to service. This was expected to occur on January 21, coincident with the return of Signal Processing Center (SPC) 60 to operations

after some extensive modifications of SPC 60 for purposes of relocation and integration of DSS 66.

While the bearing repair progressed at DSS 63, the DSN worked through seven of the S-band ΔDOR demonstration passes by using the baseline pairs of DSS 14 and DSS 65, and DSS 14 and DSS 43. One of the passes failed because of a software anomaly relating to tracking across a day-of-year change (which was subsequently corrected), and one was lost due to a spacecraft emergency. The other five yielded excellent data according to evaluation by the Galileo Navigation Team. The lessons learned by both the DSN and the Project from this demonstration were incorporated in the planning for the Earth encounter later in the year, when a much larger ΔDOR campaign was proposed.

The Galileo Project declared a spacecraft emergency on January 15 due to an unexpected switch to the fault protection mode, which resulted in a downlink change from 1200 to 10 bps. This occurred near the end of the Goldstone view period, and a Pioneer track was preempted to give Galileo additional time for assessment of the spacecraft's condition. The problem was rapidly diagnosed to be an incorrect command that had been sent earlier, and the spacecraft was commanded back to 1200 bps.

DSS-15 X-band uplink capability was declared to be operational on January 22, and the DSN began a series of operational verification tests to qualify it for future Galileo support. Operational Readiness Tests were being conducted simultaneously, with all elements of the DSN and Project to prepare for the rapidly approaching Venus encounter. Meanwhile, the DSS-63 elevation bearing replacement was on schedule for completion on January 25 at 14:00 UTC, with the full operational status on January 26 at 00:00 UTC following two satisfactory demonstration passes.

The two successful demonstration passes, plus two additional tracking passes finally resolved the remaining DSN concerns regarding DSS-63 support for the Galileo–Venus encounter, which was then scheduled for the period from February 8–17. Venus closest approach was to occur at 22:10 PST on February 9 at an altitude of approximately 17,000 km, during mutual view at DSS 43 and DSS 63. The gravity assist expected from the flyby was equivalent to a velocity increase ΔV of 2.2 km/sec. The downlink bit rate would be 40 bps at that time, but would increase to 1200 bps as soon as the link margins improved at the higher DSN antenna elevation angles.

The DSN made a special provision for critical Class 1 support throughout the Venus flyby phase. This included diesel generater power and communications backup, as well as spectrum protection during the important two-day memory readout period on February 13–14 when special selected infrared and video images would be returned at 1200 bps. On February 8 the DSN was declared ready to support the Venus encounter phase with no anomalies or exceptions being reported. Closest approach occurred at an altitude of 16,123 km at 05:58:48 UTC, February 10, 1990.

The Galileo encounter of Venus was supported by DSS 43 and DSS 63. Bad weather at DSS 43 degraded the already marginal telecommunications conditions for 1200-bps telemetry. However, excellent ranging and Doppler data were generated by both stations. Early in the DSS-63 pass, the spacecraft ranging channel was turned off to improve the telecommunications margin for the 1200-bps telemetry, which consisted only of engineering data.

Playback of selected images began with the DSS-14 pass on Tuesday, February 13. Because the limited data rate precluded a direct playback off the spacecraft tape recorder, the data were played into Command and Data Subsystem (CDS) memory and then transmitted to Earth in the engineering data format via memory readouts (MRO's). Although this introduced considerable complexity into the operational procedures for the DSN station operators, the procedure had received much attention during the pre-encounter test and training exercises, and was executed without problems.

About three hours into the DSS-14 pass, the 70-m antenna was put to "stow" because of high winds at Goldstone. DSS 43 was called up from a Pioneer pass and was able to resume the Galileo track at spacecraft rise, which occurred one and a half hours later. Approximately 1 hr and 23 min of imaging telemetry was lost, which amounted to about one third of the first Venus image. It was decided to try to recover these data in October, when the entire Venus sequence would be replayed under improved telecommunications conditions.

All the remaining MRO passes were carried out in the "ranging off," "listen only" diplexer bypass mode to maximize the SNR for the 1200-bps telemetry rate. MRO sequences continued through February 14, when critical coverage throughout the Network was lifted at the end of the DSS-43 pass. Reports from the Galileo science team commented very favorably on the quality of the data received, the excellent performance of the DSN in handling the complex operational procedures, and on the response to unexpected problems that occurred during the encounter period. Due to the low data rate, only three images and

limited other science data were planned from this MRO process. However, of utmost importance to the Project was the DSN Doppler and ranging data to be used by the Navigation Team for orbit determination, and by the Radio Science Team for Venus ephemeris improvement.

IV. Venus–Earth Cruise

Within a few days of the Venus encounter, the sustainable telemetry data rate on the 70-m antennas had dropped to 40 bps, and by March 7 it had fallen still further to 10 bps. This was due to the orientation of the spacecraft with the spin axis pointed to the Sun. The activity profile for this phase of the mission is shown in Fig. 7.

On March 12 the spacecraft switched to its alternate LGA, and because of the more favorable Earth–spacecraft sight line, the signal margins slowly started to improve. DSN support for Galileo continued at 10 bps without problems, although the B-string telemetry equipment at the stations experienced difficulty in maintaining lock due to the non-optimum use of the higher frequency subcarrier (360 KHz) for this extremely low bit rate. However, the Project elected at that time not to change the subcarrier to the design value of 22.5 KHz for other reasons.

In addition, the DSN began to experience a problem known as "false lock" due to the prevailing minimal SNR of the Galileo downlink telemetry signal. To alleviate this problem, a special operational procedure was developed in which receiver lock was first established and verified, before lockup of the telemetry processors was attempted. Although this procedure resulted in some delay to the delivery of the telemetry data, it avoided the excessive delays that were being caused by false lock conditions.

In early April, the project decided to raise the data rate to 40 bps, which required the DSN to operate the 70-m antennas in the "listen only" diplexer bypass mode to maximize the link margin. Even under these conditions telemetry performance was at times marginal.

Following repair of the DSS-63 elevation bearing, it had been decided to inspect and carry out a bearing maintenance program on the other two 70-m antennas to assure their availability for the balance of the Venus–Earth (V–E) segment of the mission. The impact of this plan on the Galileo mission sequences already loaded in the spacecraft, particularly the trajectory correction maneuvers shown in Fig. 7, was evaluated in some detail by the Project. Eventually, a workable plan was developed, and the elevation-bearing maintainence program was successfully completed prior to Earth-1 encounter.

The VEEGA mission design called for many periodic TCM's that depended on precision ranging and Doppler navigation data to be provided by the DSN. These data were provided as required, and met or exceeded the prelaunch commitments for quality and accuracy.

Encouraged by the results from the first S-band ΔDOR demonstration in January and February, the Galileo Navigation Team asked for DSN support for an S-band ΔDOR campaign that would yield a minimum of eight good passes on each baseline. These data would be used as the prime navigation data type for the Earth-1 encounter, and would have the potential for realizing a significant spacecraft fuel savings if the accuracies could be met. The DSN agreed to schedule about twice as many passes as needed, to ensure that the minimum requirements could be met.

Between August 21 and November 22, 1990, 27 ΔDOR passes were completed, 16 of which were on the north–south baseline (DSS 14–DSS 43) and 11 on the east–west baseline (DSS 63–DSS 14). Of this total, 23 were successful and yielded good data, while four failed for various operational reasons. Overall, however, the yield of good data far exceeded the minimum data return requested by the Project, and contributed significantly to the outstanding targeting accuracy achieved for the actual Earth-1 encounter.

A considerable effort went into planning an optimum operational strategy for DSN tracking through the rapid station transfers that would occur during the short time interval around closest approach to Earth (960 km). The DSN conducted extensive classroom training sessions in Pasadena for key personnel from all three complexes, and carried out a comprehensive series of Earth encounter simulation exercises, to ensure the highest possible level of proficiency for the encounter. By integrating this effort with the Project's plan for preparing the Galileo Flight Team, it was possible to make the most effective use of the limited DSN and spacecraft time available for test and training purposes, without compromising the level of proficiency required by all personnel.

The 34-m and 26-m stations at Canberra, Madrid, and Goldstone supported the operationally intense, four-hour period of Galileo Earth-1 encounter, from 18:00 UTC through 22:00 UTC on December 8, 1990, with no significant incidents or anomalies reported.

During this period, very high signal levels (greater than −100 dBm) were able to sustain telemetry data rates at 115.2 kbps and 134.4 kbps. One hundred percent of the available telemetry data was delivered to the Project.

The operationally complicated strategy to accommodate the high-angle and Doppler rates while rapidly transferring uplink and downlink from DSS 42, to DSS 61, to DSS 12, and back to DSS 42 in this short period, was carried out perfectly by all supporting elements of DSN operations.

The time-critical two-way transfer from DSS 42 to DSS 61 at the mutual low elevation of 6 degrees occurred at 18:52 UTC on schedule. DSS 61 continued to track through the spacecraft switch from coherent to noncoherent mode at 19:38 UTC until the spacecraft executed its stored command to switch from LGA1 to LGA2 at 20:15 UTC. The critical ground backup command to switch the LGA's was correctly transmitted at 20:16 UTC, as planned, prior to turning off the station transmitter. Four minutes later, at 20:20 UTC, the spacecraft reverted to the two-way coherent mode.

The expected drop in downlink signal strength due to the spacecraft antenna switch did not occur, and as a consequence DSS 61 was able to maintain lock on the downlink until 20:25 UTC when track was terminated due to approaching limits on the antenna slew rate. However, DSS 66, which was tracking in backup mode, was able to track down to spacecraft set, thereby providing an additional 15 minutes of telemetry data beyond that which was called for in the planned sequence. Earth closest approach occurred at 20:34:34 UTC at an altitude of 960 km over the Caribbean Sea, out of view of the DSN between Madrid set and Goldstone rise. Immediately upon spacecraft rise at Goldstone, DSS 12 acquired the one-way downlink with telemetry in lock at 134.4 kbps at 20:40 UTC as scheduled. Two-way acquisition was completed at 20:46:53 UTC.

The second time-critical transfer from DSS 12 to DSS 42 was planned for 21:10 UTC when the spacecraft's best lock frequency would be seen simultaneously by both stations. DSS 42 was able to acquire the incoming signal about three minutes early, and complete the transfer on time at 21:10 UTC.

From this point on, DSS 42 continued to provide tracking support at 134.4 kbps and 115.2 kbps until the end of the scheduled pass at 06:20:00 UTC the following day. Throughout this critical period, the DSN executed all the complex, time sensitive sequences without error, due in no small part to the well-planned test, training and documentation effort that preceded the real-time events.

The effort expended by the DSN on the ΔDOR campaign prior to the encounter was well justified when the Galileo Navigation Team estimated that the improved targeting errors resulting from use of the ΔDOR data translated into a spacecraft fuel savings of approximately 6 kg, an important factor for consideration in the eventual design of the Jupiter satellite tour.

In the immediate post-encounter period, the Project took advantage of the prevailing telecommunication link performance to return all the Earth encounter science data at high bit rates. However, as the Earth–spacecraft range increased, the link margin rapidly decreased, and the sustainable bit rate declined again to 40 bps by the end of January 1991. By that time, the DSN had resumed regular tracking support on the spacecraft as the Project commenced the final days of the Venus–Earth segment of the mission.

One of the final events of the Venus–Earth (VE) cruise was the unfurling of the HGA. This event was in the final VE sequence. However, on March 26 the spacecraft experienced a bus reset signal and consequently safed itself. The safing incident automatically canceled the on-board sequence. So, the HGA unfurling attempt was carried out with real-time commands on April 11, 1991.

The unfurling was not successful. Real-time observations by the Spacecraft Team detected a quick ramp-up in motor current to well above the expected value. The redundant deployment motors ran for eight minutes, at which time the sequence turned them off. A normal full deployment would have been completed in about three minutes. Then a microswitch should have cut off the motors. No microswitches tripped. Two other observations also pointed to an HGA anomaly. Namely, the spacecraft's spin rate decreased by less than that expected for a full deployment. Finally, the sun gate indicated an obscuration.

It was immediately determined that except for the HGA, the spacecraft was completely normal. An anomaly investigation team was then convened. Intensive modeling, analysis, and testing have been carried out and refined since then. The primary scenario is that three (of the eighteen) ribs are restrained in the stow position. Two additional ribs may have been involved initially and sequentially released. But no hardware is thought to have suffered any damage or structural failure.

The HGA anomaly team, now called the HGA Recovery Team, has determined that to free the stuck ribs, the antenna tower must be expanded and contracted to walk the guide pins out. A warming turn was conducted in May, then cooling turns in July and August (see Fig. 8)

were supported by the DSN. Each involved about 50 hours of near-continuous coverage, with many more tracks after the turn to analyze the results. To date, the turns have not been successful in freeing the ribs.

V. Earth–Earth Cruise Summary (Through Gaspra Encounter)

As a result of the HGA anomaly, the DSN lost two significant opportunities planned for early in the Earth–Earth cruise. First of these was a long constraint length Viterbi decoding demonstration planned for May and June. A constraint length = 15, rate = 1/4 convolutional code was implemented on the spacecraft to enable better performance for the Io encounter. This code was implemented only for X-band, which is dependent on the HGA. So the Big Viterbi Decoder implementation has continued without the use of flight data.

Also lost during this period was the X-band uplink demonstration. This was to be a one-station demonstration during opposition. This demonstration is in support of the Gravitational Wave Experiment. Two brief X-band uplink periods were undertaken in support of an HGA characterization, but were in no way beneficial to the Gravity Wave Experiment.

The spacecraft again entered safing on May 3, 1991. Extensive ground problem isolation activities were initiated during the DSS-43 pass to determine that the problem was not on the ground. After determining the state of the spacecraft, recovery plans were initiated, and the Project requested extra ground support. The DSN consequently provided continous coverage from May 3, 1991, when the problem was first noticed, through 9:00 P.M. PDT on May 4, 1991.

The rest of 1991 was devoted to two primary activities: trying to overcome the HGA anomaly, and preparing for the first asteroid encounter. The maneuvers to free the stuck HGA ribs have been discussed briefly above. Figure 8 is included as a time line for this period. More details on the Gaspra encounter support follow.

After the HGA anomaly, the DSN recommended to the Project that the DSN could better support the low data rates on the 22.5-kHz subcarrier. A demonstation was scheduled for July 23, but the spacecraft had gone into fault protection and reverted to a repeating memory dump at 10 bps sometime following the tracking support on July 19. This was discovered on July 22, 1991, and recovery was not accomplished until July 25. The 22.5-kHz subcarrier demonstration was rescheduled for August 23, 1991. It was successful, and improved performance led to the decision to operate on the 22.5-kHz subcarrier through the Gaspra optical navigation recovery period, the Gaspra encounter, and indefinitely during subsequent low data rate operations.

Planning for the asteroid Gaspra encounter necessitated an early decision to commit to an LGA encounter. This commitment involved much less optical navigation data than had been planned with the HGA, thereby making those data even more critical to the Navigation Team. Optical navigation data has become a key data type for the Galileo mission. DSN support to recover the data was crucial for Gaspra, as the data rate was reduced to 40 bps during this time frame. To recover the optical navigation data, it was necessary to play the imaging data from the tape recorder to the CDS, and then subsequently read the data out as a subset of the engineering data. This is analagous to the Venus MRO's, only at the reduced data rate. The DSN supported nearly 400 hours of these optical navigation playbacks, with no significant problems. As can be seen in Fig. 8, these playbacks occurred throughout September and October.

As for previous encounters, the DSN made special provision for critical Class I support throughout the Gaspra encounter, including the final time-critical optical navigation frame. MRT's were conducted at all three 70-m sites, and classroom training was provided in Pasadena for key Project and DSN personnel. A Gaspra Readiness Review was held by the Project on October 14, 1991, and the DSN was declared ready for support at that time.

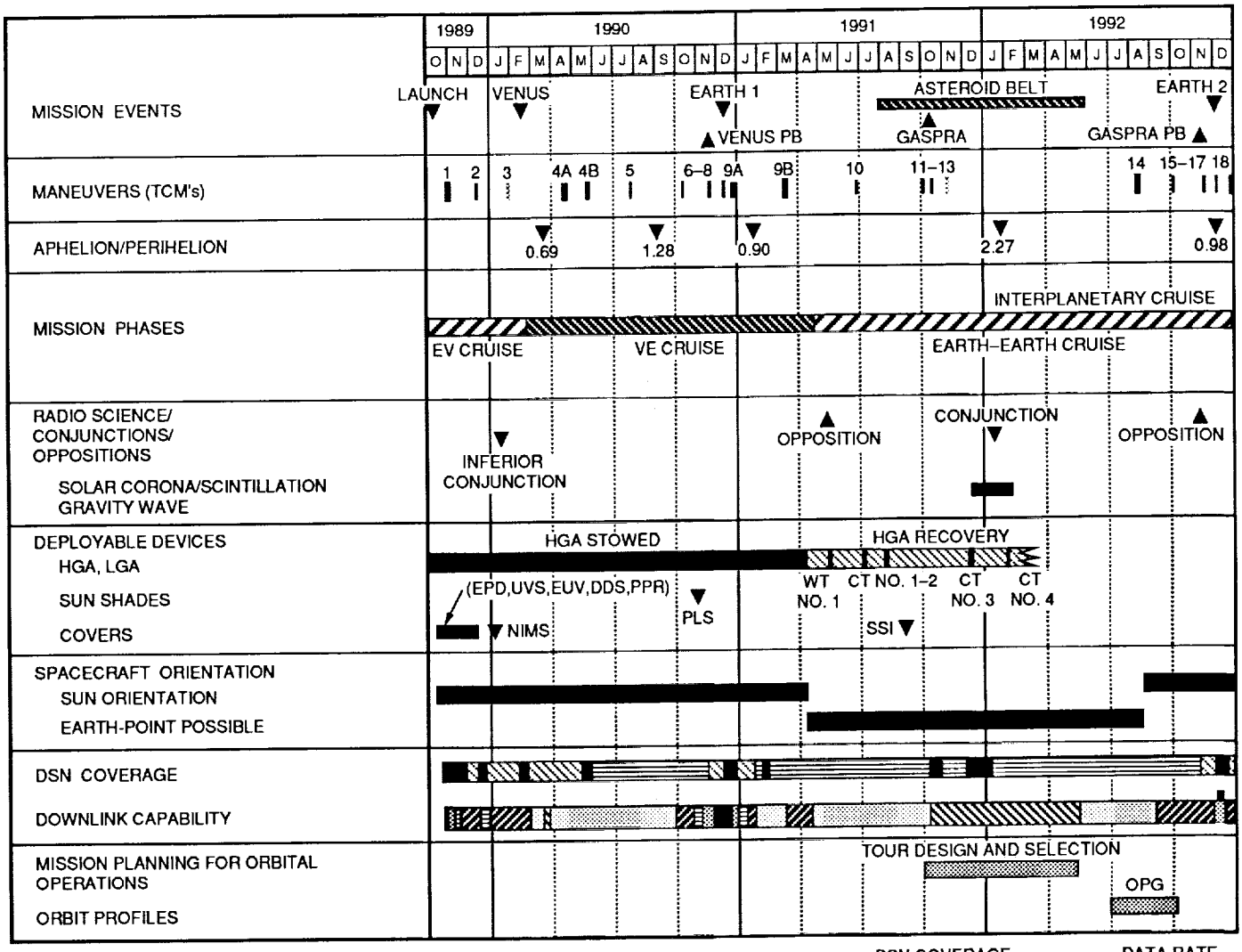
Accomplishing the encounter on the LGA also meant that no real-time science could be realized; only 10-bps engineering from a 70-m antenna was possible at closest approach. Furthermore, only one tape load of science data could be realized. In fact, one track was reserved for engineering data in case analysis of anomalous spacecraft events might be required. Despite this constraint, the Project was able to realize excellent science, and recover on the tape recorder essentially all the basic science goals. These data will be relayed to the ground when higher data rates are achievable during the Earth-2 flyby period.

Closest approach to Gaspra occurred on October 29, 1991. Galileo passed within 1601 km of the main-belt, S-type asteroid. The time of closest approach was 22:36:59 UTC.

Subsequent to the encounter, it was decided to attempt to recover one image of Gaspra in the same fashion as was

done for the early Venus playback data, and the aforementioned Gaspra optical navigation data. There were two contributing factors to this decision: (1) DSN performance was significantly better than had been predicted, which allowed 40-bps capability for a few weeks after Gaspra encounter; and (2) high probability was established for being able to recover the medium resolution image.

This image was about 77 lines long (out of 800 for a full image). When it was recovered within the first three days of scheduled DSN time, a decision was made to attempt to recover more images of Gaspra. Within the window of the time Galileo had reserved to recover the one image, it became possible to play back a total of four images of Gaspra, each through a different filter. DSS 43 was the only DSN antenna to be utilized for this effort because of significantly better elevations during the passes. The four images permitted a full-color Gaspra image. The 40-bps data rate was only supportable in the low-noise, diplexer bypass mode for this period.



1. TOUR SHOWN IS REPRESENTATIVE
2. GSOC OPERATIONS ARE ASSUMED TO BEGIN IN MARCH 1993
3. DOWNLINK CAPABILITY SHOWN IS FOR A 70-m DSN STATION, RANGING ON, 90-PERCENT CONFIDENCE. PRIOR TO FEBRUARY 1993, DOWNLINK IS ASSUMED TO BE VIA LGA-1 (S-BAND TRAVELING WAVE TUBE AMPLIFIER, HIGH POWER).

DSN COVERAGE
- CONTINUOUS
- 4–7 TRACKS/WEEK
- 2 TRACKS/WEEK

DATA RATE
- 134.4 kbps
- 28.8 kbps
- 7.68 kbps
- 1200 bps
- 40 bps
- 10 bps

CT	COOLING TURN	PB	PLAYBACK
DDS	DUST DETECTOR	PLS	PLASMA SUBSYSTEM
EPD	ENERGETIC PARTICLES DETECTOR	SSI	SOLID STATE IMAGING SUBSYSTEM
EUV	EXTREME ULTRAVIOLET SPECTROMETER	UVS	ULTRAVIOLET SPECTROMETER
OPG	ORBITAL PLANNING GUIDE	WT	WARMING TURN

Fig. 1. Galileo mission overview.

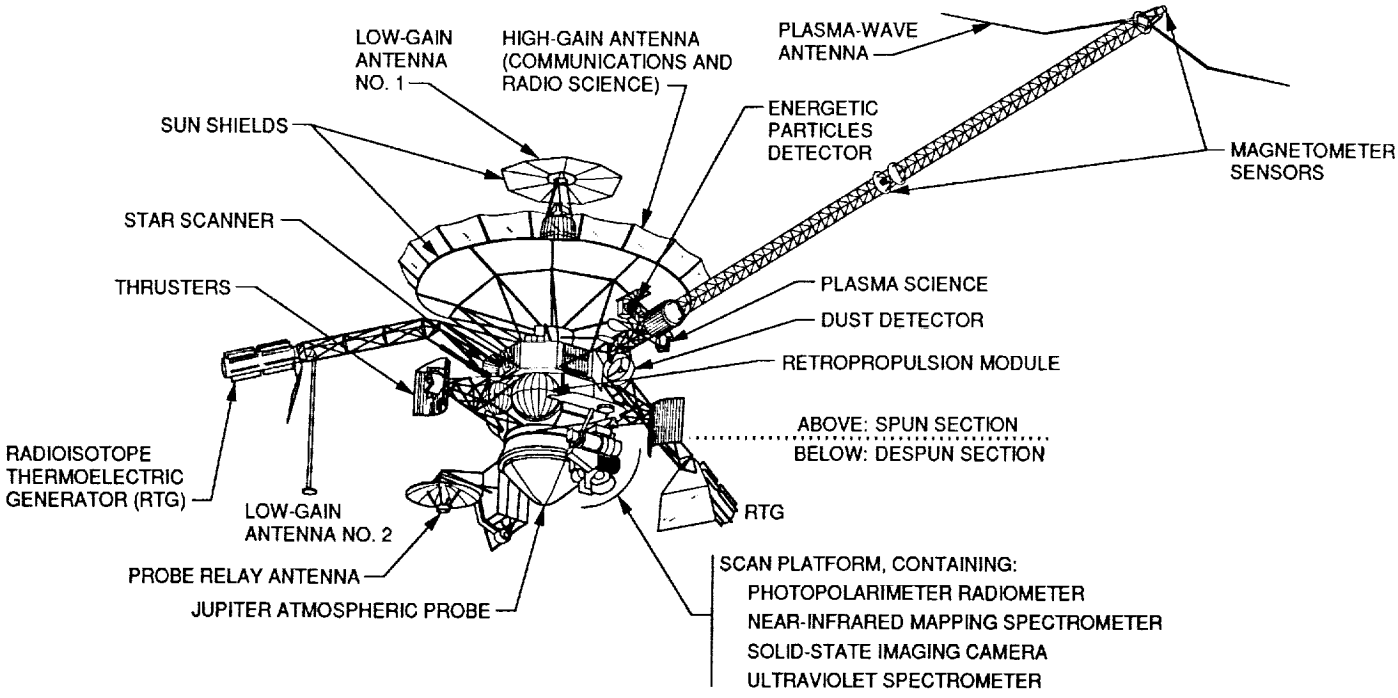


Fig. 2. Galileo spacecraft configuration.

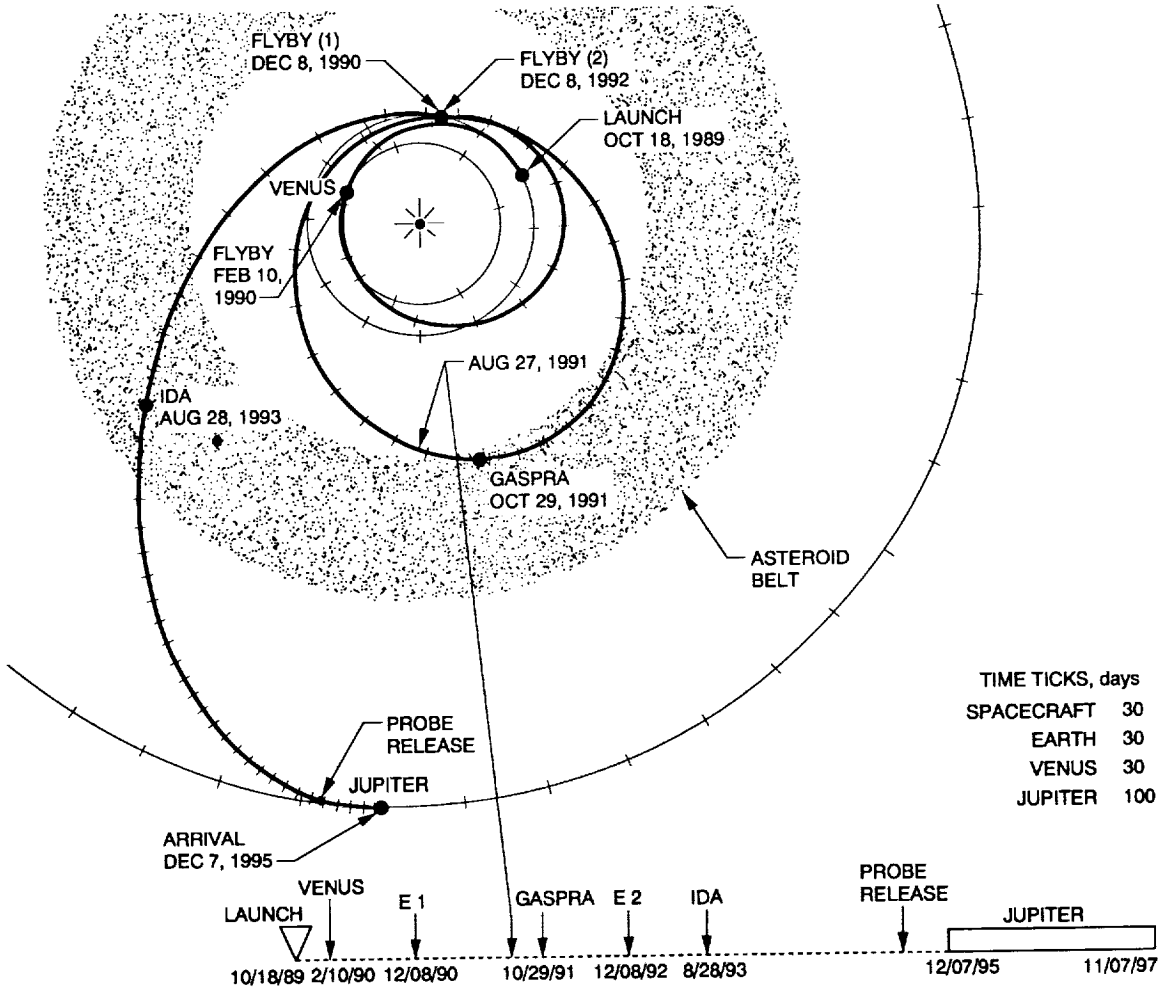
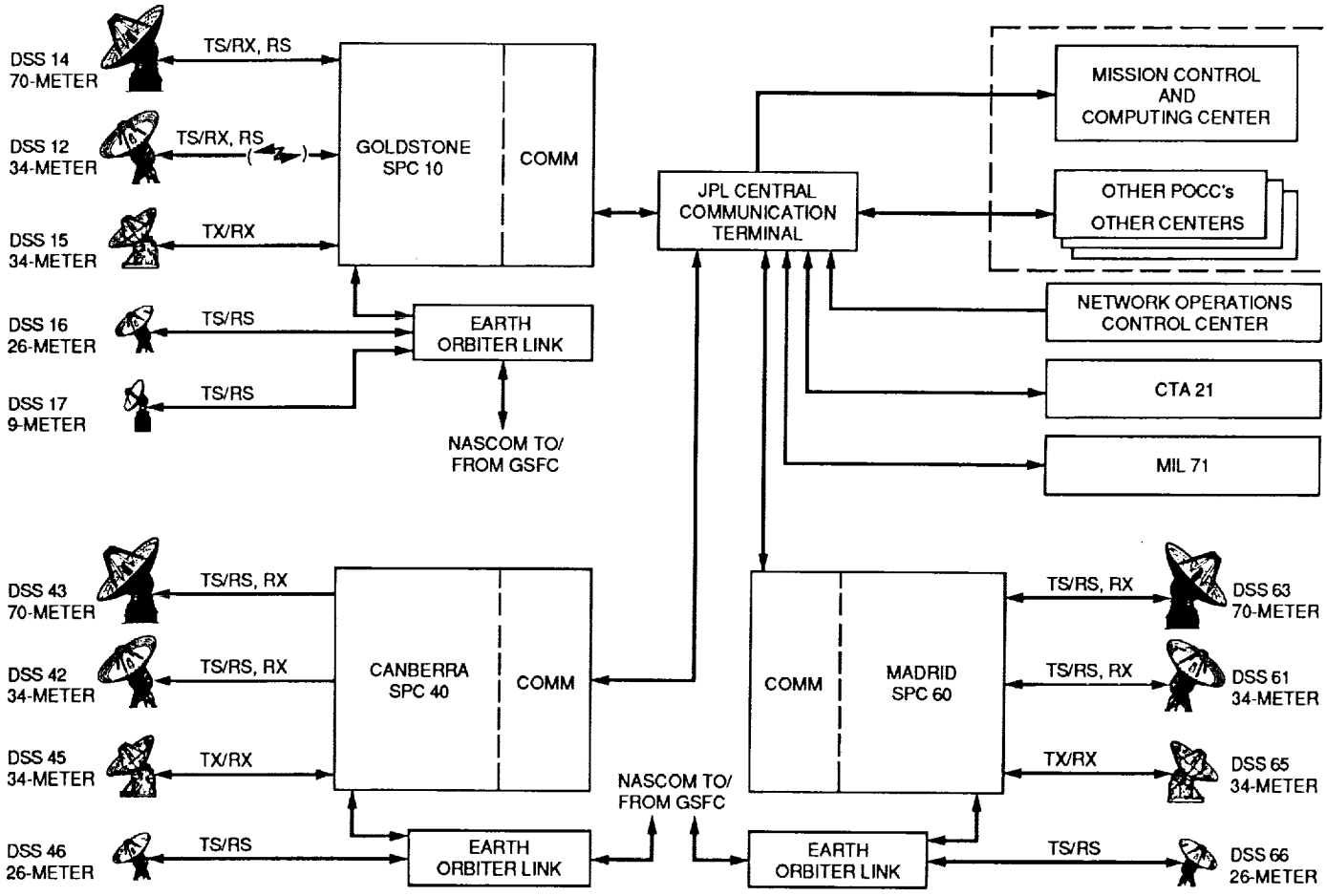


Fig. 3. Galileo trajectory.



COMM	GROUND COMMUNICATIONS	RS	RECEIVE S-BAND
CTA	COMPATIBILITY TEST AREA (PASADENA)	RX	RECEIVE X-BAND
DSS	DEEP SPACE STATION	SPC	SIGNAL PROCESSING CENTER
GSFC	GODDARD SPACE FLIGHT CENTER	TS	TRANSMIT S-BAND
MIL 71	MERRIT ISLAND LAUNCH AREA (FLORIDA)	TX	TRANSMIT X-BAND
POCC	PROJECT OPERATIONS CONTROL CENTER		

Fig. 4. 1989 DSN configuration for Galileo.

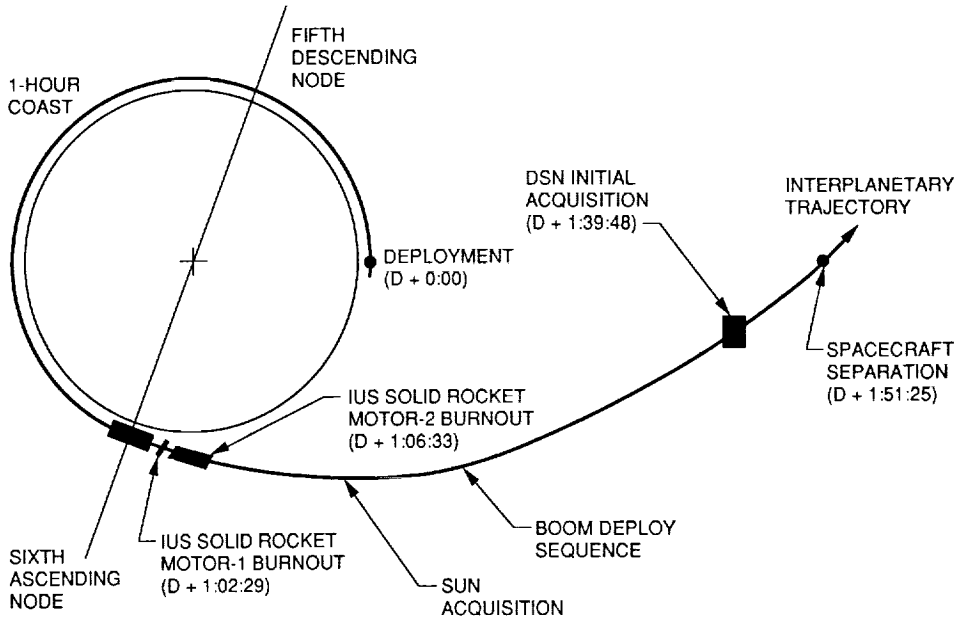


Fig. 5. IUS/Galileo nominal trajectory profile.

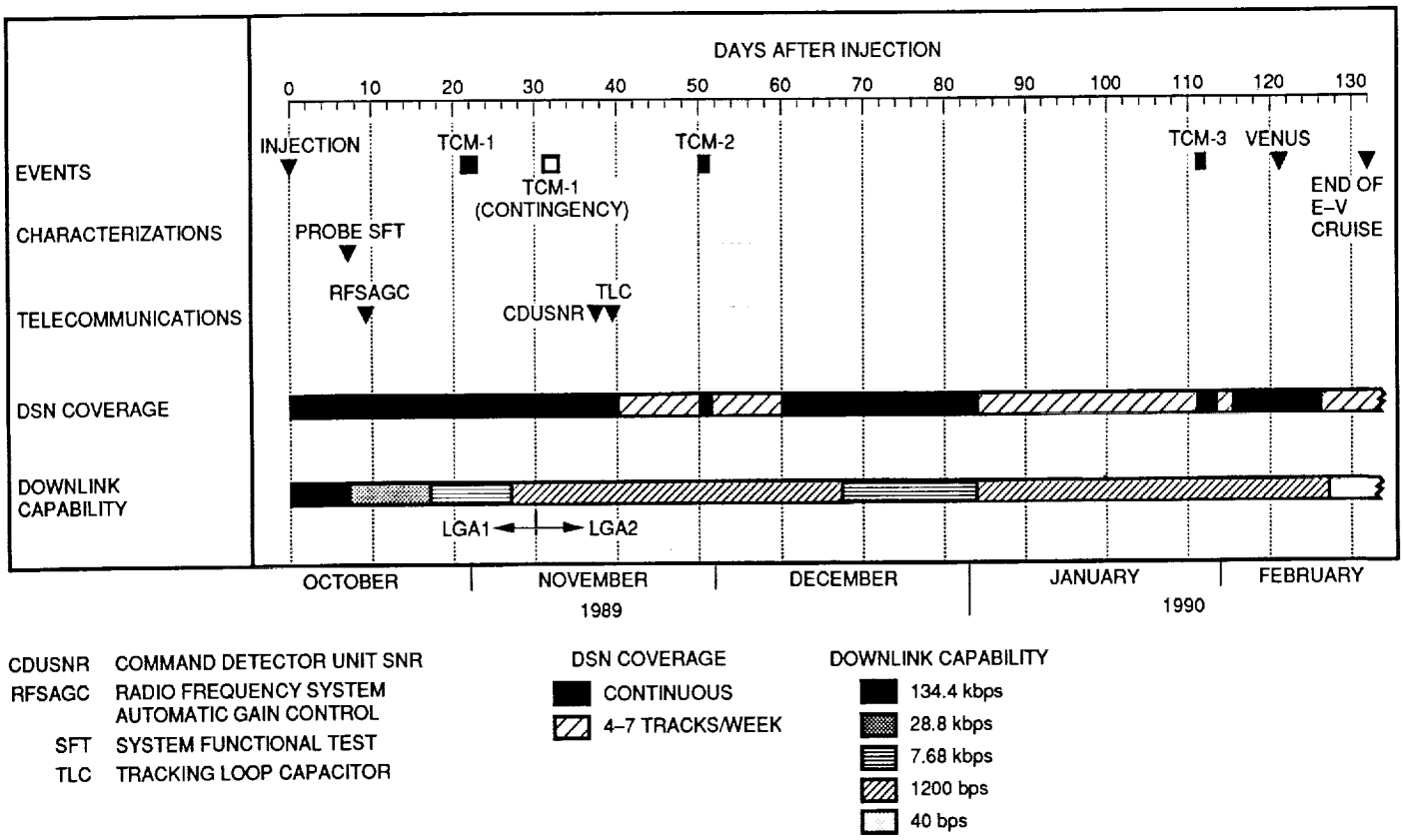
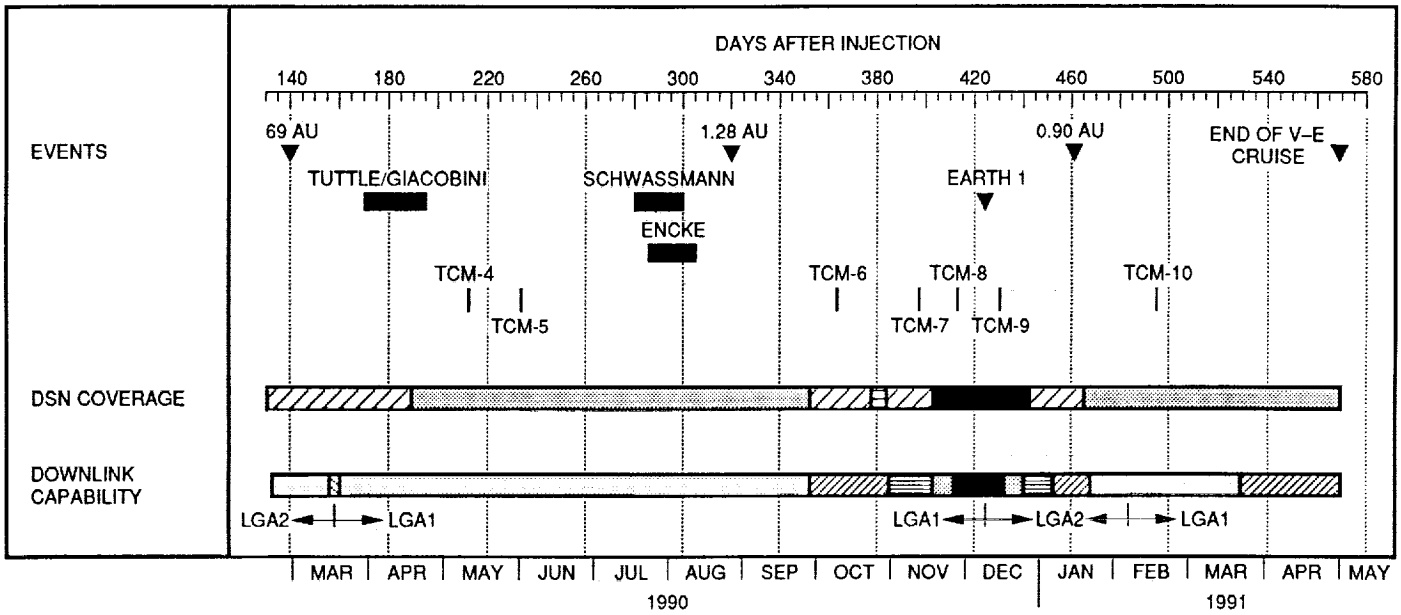


Fig. 6. Earth–Venus time line.



DSN COVERAGE
- CONTINUOUS
- 4–7 TRACKS/WEEK
- 2 TRACKS/WEEK
- ARRAY 1/DAY

DOWNLINK CAPABILITY
- 134.4 kbps
- 28.8 kbps
- 7.68 kbps
- 1200 bps
- 40 bps
- 10 bps

Fig. 7. Venus–Earth time line.

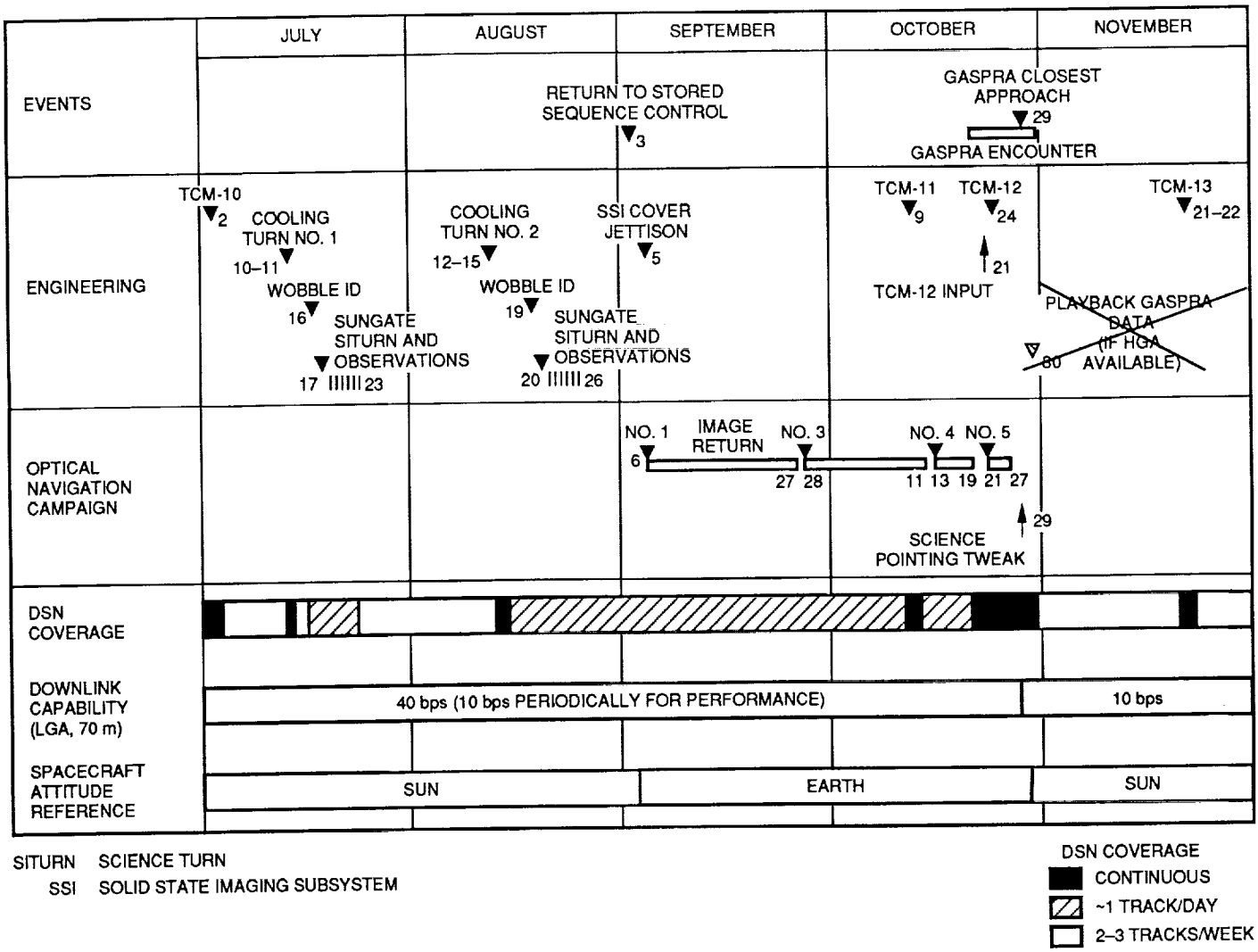


Fig. 8. Galileo events through Gaspra encounter, 1991.

Appendix

The total tracking support provided to the Galileo Project by the DSN in the period from launch (October 1989) through the Gaspra encounter (October 1991) is shown in Fig. A-1. The height of each bar represents the number of tracking hours scheduled for the month, with the lost time shown as a blank at the top, and the actual support time shown as shading within the bar.

During the review period, 8224.65 hours of support on all antennas was scheduled, of which 98.67 percent (8115.58 hours) was actually provided and 1.33 percent (109.07 hours) was lost for various operational and technical reasons. The large peaks in tracking time provided for launch in October–December 1989, Earth encounter in November–December 1990 and Gaspra encounter in October 1991 are clearly evident.

The telemetry data capture performance is also shown in Fig. A-1 as a profile at the the top of the figure. The vertical scale shows the monthly percentage of available data captured and delivered to the Project. A level better than 95 percent was maintained for the entire period.

The command performance is reflected in the total number of commands sent to the spacecraft and the number of aborts over the period October 1989–December 1991: The number of commands transmitted was 6202 and the number of commands aborted was 7.

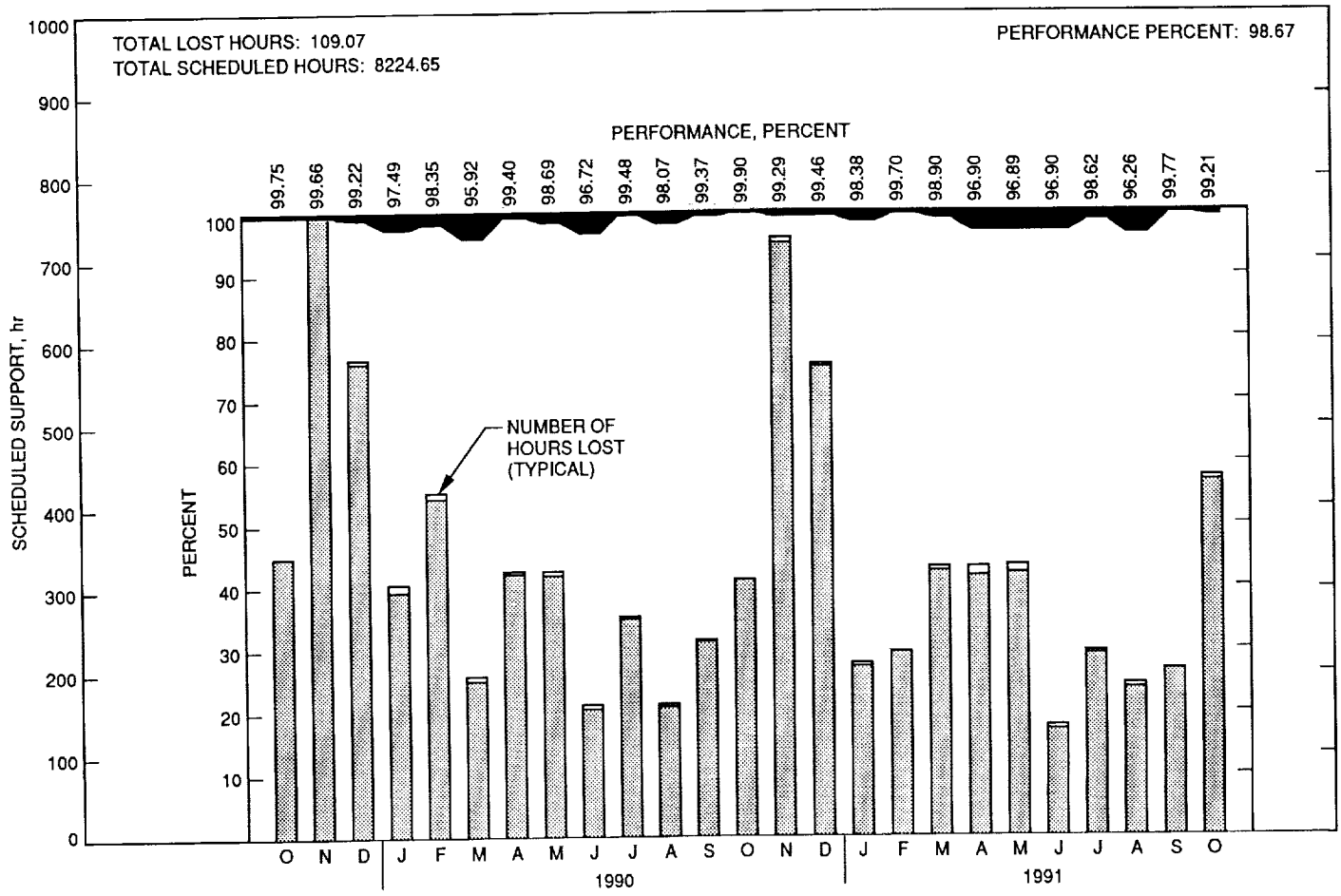


Fig. A-1. DSN telemetry performance.

Deep Space Network Information System Architecture Study

C. A. Beswick
Information Systems Division

R. W. Markley (Editor), D. J. Atkinson, and L. P. Cooper
Advanced Information Systems Section

R. C. Tausworthe
Information Systems Division

R. C. Masline
Science Data Systems Section

J. S. Jenkins
Control Center Data Systems Development Section

R. A. Crowe, J. L. Thomas, and M. J. Stoloff
DSN Data Systems Section

The purpose of this article is to describe an architecture for the DSN information system in the years 2000–2010 and to provide guidelines for its evolution during the 1990s. The study scope is defined to be from the front-end areas at the antennas to the end users (spacecraft teams, principal investigators, archival storage systems, and non-NASA partners). The architectural vision provides guidance for major DSN implementation efforts during the next decade. A strong motivation for the study is an expected dramatic improvement in information-systems technologies, such as: computer processing, automation technology (including knowledge-based systems), networking and data transport, software and hardware engineering, and human-interface technology.

The proposed Ground Information System has the following major features: unified architecture from the front-end area to the end user; open-systems standards to achieve interoperability; DSN production of level 0 data; delivery of level 0 data from the Deep Space Communications Complex, if desired; dedicated telemetry processors for each receiver; security against unauthorized access and errors; and highly automated monitor and control.

I. Introduction

A. Background

The Deep Space Network (DSN) is the largest, most sensitive scientific telecommunications and radio navigation network in the world. Its principal responsibilities are to support automated interplanetary spacecraft missions and radio and radar astronomy observations in the exploration of the solar system and the universe. The DSN also supports high-Earth orbiter, lunar, and shuttle missions.

The DSN is managed, technically directed, and operated for the National Aeronautics and Space Administration (NASA) by the Jet Propulsion Laboratory (JPL), California Institute of Technology, Pasadena, California. This study responds to the following action item developed by the 1990 JPL Telecommunications and Data Acquisition (TDA) NASA Office of Space Operations (OSO) Planning Workshop: "Perform a study of the DSN information system architecture and recommend guidelines for its evolution."

It is timely that such guidelines be developed because of such improvements in information technology as:

(1) Dramatic advances in computer hardware and software.

(2) Reduced size of computers and other processing elements due to very large-scale integration (VLSI).

(3) Dramatic increase in ground communication data rates using optical fiber.

(4) Improved ability to internetwork local area and wide area networks.

(5) Efforts by the U. S. Government and the international community to standardize network protocols.

(6) Efforts by the U. S. Government to standardize software.

(7) Advances in distributed computing.

(8) Advances in providing assistance to decision makers by means of artificial intelligence (AI) and visualization techniques.

(9) Advances in computer–human interfaces, including multimedia, interactivity, computer-supported cooperative work, and virtual reality.

New projects in NASA's interplanetary exploration program require increased communications and tracking capabilities that must be implemented while maintaining support for existing projects. There is also a requirement to increase DSN availability beyond its already high level. Additional missions will have a tendency to increase development, implementation, maintenance, and labor force costs, and it is believed that these new technologies will enable significant reduction of these costs.

B. Purpose

The purpose of the study is to develop an architecture for the Ground Information System (GIS) in the years 2000–2010 and to provide guidelines for its evolution during the 1990s. The study also provides a forecast of information systems technologies that are pertinent to the GIS. Where forecasted technologies are inadequate for the projected architectural needs of the GIS, the study recommends appropriate pathfinder work. A flexible transition approach is developed, with the expectation that most improvements to the architecture will appear evolutionary, although they are concentrated on a clearly defined vision of the future architecture.

C. Scope

In order to allow for creativity in the selection of architectural candidates, the information systems boundaries for the study were deliberately chosen without consideration for current organizational boundaries. Specifically, the scope was defined to be the Ground Information System (Fig. 1), which extends from the ground side of the antenna front-end assemblies to the end user (spacecraft teams, principal investigators, archival storage systems, and non-NASA partners). This choice keeps the scope within bounds (i.e., only ground systems), but allows flexibility in defining where ground processing functions are performed. For example, the current interface between the DSN and the Space Flight Operations Center (SFOC) is not considered to be a constraint. After an appropriate technical solution is determined, organizational decisions may be made (outside the scope of this study) and appropriate interfaces defined.

The GIS functions include processing (telemetry, tracking, command, radio science, and very long baseline interferometry [VLBI]), monitor and control of the front ends and the GIS, and data delivery and management. The focus of the study is on multimission functions; project-unique activities will be dealt with only as external requirements. Issues that provide both focus and motivation for the study include:

(1) The degree of automation and centralization of the monitor and control function.

(2) The degree of distribution of computing and control functions.

(3) The feasibility of implementing an open systems architecture.

(a) Cost and performance.

(b) Security.

(4) The modularity and reconfigurability of the GIS.

(5) Communication alternatives.

(6) The choice of functions to be generic multimission functions.

(7) The limits and capabilities of technology.

D. Evaluation Criteria

Architectural candidates have been developed that meet planned performance requirements and are evaluated with respect to the following criteria:

(1) Performance, including the degree of margin for unplanned mission support.

(2) Life-cycle cost, including development, implementation, maintenance, and workforce costs.

(3) Operability, including manageability and human-interface simplicity.

(4) Flexibility, evolvability, and growth potential.

(5) Availability, including fault-tolerance, reliability, and maintainability.

(6) Technical risk, including ease of transition.

Emphasis in this study is on general advantages and disadvantages of the candidate architectures relative to a baseline (the current architecture) or to the other competing candidates. Accurate measures of the above criteria (especially cost) require more design effort than is appropriate for a high-level architectural study. Sensitivity of the recommendations to technology advances is considered.

E. Organization of This Article

The remainder of this article is organized as follows:

Section II. Requirements: This section provides a functional partitioning of future requirements, and discusses the current functional architecture, system interfaces, future mission requirements and end-user interfaces.

Section III. Technology Forecast and Projected Impact: This section includes discussions of computers, software, data transport, human interfaces, and their potential impact on the DSN information system.

Section IV. Architectures: Processing, monitor and control, data transport, and software architecture discussions are included in this section.

Section V. Transition Approach: This section describes a plan for evolving in an orderly way from the current architecture to the envisioned architecture.

Section VI. Summary: This section highlights the most significant conclusions derived from evaluating the DSN requirements and technology, including recommendations for achieving the strategic objectives.

Appendix A. Acronyms: This section provides definitions of all acronyms used.

Appendix B. Glossary: This section provides more detailed descriptions of certain terms used.

II. Requirements

A. Functions

A functional partitioning of future requirements for the GIS is shown in Fig. 2. This figure illustrates functional interrelationships of the various subsystems and clarifies areas where functional commonality exists. An easy reference, it should help to ensure that any proposed design will address the functions needed to meet the long-term requirements. Several strategic goals for the GIS are described in Table 1. Some of these goals are included later (Section IV) as evaluation criteria when the architectures are discussed.

1. Processing. This section describes the existing general functions of the GIS. It is believed that future designs will embody most of the major functions that exist in the GIS today. There are three Deep Space Communications Complexes (DSCC's), each identical in terms of function, which are located in Goldstone, California; Madrid, Spain; and Canberra, Australia. Communication circuits that link JPL to the three global sites are managed by the Ground Communications Facility (GCF).

Each DSCC has a number of antenna stations distributed across its site with an associated set of electron-

ics. These antennas and their associated control rooms and electronics are called Deep Space Stations (DSS's). They perform a limited amount of analog signal processing before transmitting an analog signal to a collocated central area called the Signal Processing Center (SPC), shown in Fig. 3. If several antennas are arrayed to detect the usually weak spacecraft radio signal, the analog signals from several DSS's may be combined before further SPC processing. After the SPC has filtered and amplified the resultant analog signal, the telemetry, tracking, radio science, and VLBI data are digitally processed.

In the current system, the resulting digital data may be communicated to JPL over long-haul communications circuits, or stored on magnetic tape and shipped. Various types of processing take place at JPL in the SFOC, including deletion of duplicate data acquired by two DSCC's in the event of overlapping coverage during an encounter. Much of the current SFOC processing is mission-specific; however, it is expected that future missions will share identical standard data structures and coding techniques because of ongoing development of standards by the Consultative Committee for Space Data Systems (CCSDS).

a. Telemetry. Some spacecraft projects request that their data be immediately transmitted to JPL upon reception (i.e., real-time data). Rapid delivery depends on the availability of appropriate long-haul circuits. This type of data is delivered in the order that it is received at the antenna, delayed only by initial processing and the length of time it takes to traverse the network. In the event of errors in transmission (i.e., missing data blocks), retransmission may be requested later by the user—a process called postpass replay. Other projects prefer that data be stored on magnetic tape and either mailed back to JPL or replayed at a slower rate over long-haul communication circuits (i.e., non-real-time data). Non-real-time delivery is more economical when the spacecraft downlink rates are very high or volume is very large.

The DSCC Telemetry Subsystem (DTM) currently acquires a digital telemetry baseband signal from the receiver and converts it into the format required for further processing at the SFOC to produce a data stream useful to the end user. This telemetry processing function is accomplished through a sequence of transformations that optimizes the amount and correctness of data acquired in the presence of noise. Deep space missions have significant variations in the specific transformations that are applied. There are also variations due to the different noise environments of Earth orbiters and deep space missions.

(1) The Telemetry Source. Telemetry data originate in the spacecraft, where information from several sensors is multiplexed into a single binary stream and is then modulated onto a subcarrier of the downlink signal. Spacecraft engineering information may be multiplexed into this signal or onto a separate downlink subcarrier. The multiplexed stream is formatted into logical groups, or frames, for synchronization purposes.

The telemetry data may be transmitted directly without further coding if signal-to-noise ratios are high on the spacecraft–Earth link; however, it is usually necessary in deep space missions to code the telemetry data with additional bits and reconstruct the correct data packets on Earth. The two principal coding schemes used in deep space missions are convolutional, such as Viterbi codes, and block, such as Reed-Solomon (RS) codes.

Both coding schemes reduce bit errors for a given signal-to-noise ratio over uncoded transmissions, but the bandwidth cost is the need to transmit additional bits from the spacecraft. For example, the RS code used by Mars Observer has a block code word structure of 10,000 bits, in which 8720 are information bits, and the remaining 1280 are parity (or check) bits. The rate of this code is $R = 8720/10,000 = 0.87$; thus, 87 percent of the bits received are information.

Convolutional codes, on the other hand, are codes in which n channel symbols are transmitted for each information bit. Since each information bit leads to n transmitted channel symbols, the rate is $R = 1/n$. Typical rates are 1/2 for high-Earth orbiters and lunar missions and 1/6 for deep space missions. Thus, in cases where $R = 1/6$ and the symbols are represented by 8 bits, 1 information bit may lead to transmitting and processing up to 48 bits.

Block and convolutional codes can operate on multisymbol sets. For example, the RS code above treats each 8 bits of the code word as one symbol because of the algebraic method used for correcting burst errors. Thus, the RS code above is also referred to as a (255,223) 8-bit symbol RS code.

The implementation of a convolutional coder is simpler, less costly, and more reliable than a block coder for spacecraft operation. However, the information may be RS-coded prior to convolutional coding to make the telemetry data resilient to both random noise and burst errors. The overall rate of information bits per transmitted binary symbol is the product of the two coding processes. A more quantitative discussion of the characteristics of the

telemetry stream (as well as deep space telecommunications in general) may be found in [1].

(2) Subcarrier Demodulation and Symbol Synchronization. In order to extract the telemetry data at the station, it is necessary to reverse the processing steps performed on the spacecraft. First, the subcarrier is removed by standard heterodyne means; next, binary symbol synchronization is established. Each binary symbol is integrated for the duration of its symbol time. In the absence of noise, positive and negative integrated symbol values would be interpreted as ones and zeroes, respectively. However, in the presence of noise, such a decision may not always be correct. Rather than commit to a decision at this point, the convolutional decoders are designed to operate on soft symbols, i.e., the magnitudes of the outputs of the symbol integrator. Since the decoders work in digital logic, the soft symbols are quantized into 2^n levels per sample, which yields a number in the range $(-2^{n-1}, 2^{n-1} - 1)$ for each soft symbol.

In the future, subcarrier demodulation, symbol synchronization, and symbol-matched filtering will be accomplished by the planned-for Block V receiver. At that time, the input to telemetry processing will be a stream of soft symbols.

(3) Telemetry Decoding. The next step in telemetry processing is to transform the soft symbol stream into an information bit stream. For uncoded data (typical of older spacecraft and some Earth orbiters), there is a one-to-one correspondence between symbols and bits; symbols are converted to information bits by simply taking the sign bit of each soft symbol. However, when both convolutional and RS coding have taken place, two more processing stages are necessary.

Current convolutional decoders use 3-bit symbols (eight levels), but performance improvements are possible through the use of 8-bit symbols (256 levels of quantization). The logic speed (S) required to process telemetry data must accommodate the product $QI(1/R_{rs})(1/R_c)$ where Q is the soft symbol quantization rate in bits per symbol, I is the information bit rate, R_{rs} is the Reed-Solomon rate, and R_c is the convolutional code rate. For example, in the case where

$$Q = 8 \text{ bits per symbol}$$

$$I = 10^6 \text{ bps}$$

$$R_{rs} = 223/255$$

$$R_c = 1/6$$

The logic speed $S = 55$ Mbps

(4) Frame Synchronization. Each frame consists of a sequence of binary symbols; the beginning (or end) of a frame is identified by a unique bit pattern called the frame synchronization marker. Output from the frame synchronizer marks the data in such a way that subsequent telemetry transforms the data frame by frame.

For the DSN, frame synchronization is useful as a real-time data-quality monitor, and is a necessary step before doing RS decoding. Among the complications of frame synchronization are errors in the frame synchronization pattern, false synchronization (i.e., data in the frame that coincidentally match the frame synchronization pattern), reversed data (from spacecraft tape recorder dumps), and changes in frame length.

(5) Output Formatting. Corrected frames from the RS decoder must be transmitted or recorded. Traditional transmission requires that the data be formatted into 4800-bit NASA Communications Network (NASCOM) data blocks. Data in the blocks include the NASCOM header and trailer, the telemetry header, telemetry partial status data, and the actual telemetry data. Some missions have placed non-NASCOM requirements for the content and organization of telemetry headers, partial status, and data. Thus, output processing must support a variety of output block formats.

Newer missions, such as Mars Observer, will require the output blocks to be formatted in accordance with CCSDS recommendations. Processing is required to recognize CCSDS-compliant telemetry packets and provide the capability to route different packets to different destinations based on the virtual channel identifier in the packet header.

b. Tracking. The DSCC Tracking Subsystem (DTK) has the following primary functions:

(1) Control the transmitted uplink frequency to enable initial signal acquisition by the spacecraft receiver and maintain two-way communications in which the radio signal is transmitted from the ground to the spacecraft and relayed back to the ground receiver in a phase-coherent manner.

(2) Measure the round-trip travel time of light to the spacecraft, thereby enabling precise calculation of the spacecraft range.

(3) Measure the Doppler shift on the radio signal received from the spacecraft to give a precise indication of the spacecraft radial velocity.

Support information (generically called predictions or predicts), such as subsystem and instrument settings required for uplink signal control, is generated by the Network Support Subsystem (NSS) at JPL and sent to the DTK to support these functions; the DTK packages raw and computed Doppler shift data into data blocks for transmission back to the project navigation teams.

Frequency ramping commands go from the Metric Data Assembly (MDA) to a digitally controlled oscillator according to time-stepped predicts. The intention is to compensate for the Doppler shift at the spacecraft and keep the signal frequency received at the spacecraft within its closed-loop bandwidth. The MDA utilizes some predicts to construct acquisition profiles that are translated into frequency ramp commands aimed at establishing a two-way link with the spacecraft.

Ranging modulation from the current Sequential Ranging Assembly (SRA) is in the form of square waves or sine waves imposed on the subcarrier in a step-wise progression from highest to lowest frequency. The highest frequency now in use is 1 MHz and the lowest is 1 Hz. Capability is provided in the SRA to go to at least 2 MHz. The detection and exact timing of these signals coherently relayed back from the target spacecraft provides a precise estimate of the spacecraft range (via the round-trip travel time of light).

Doppler processing is done by the MDA to provide project navigation and network monitor personnel with a precise measurement of the Doppler shift at any given time. Counts timed to the thousandth of a cycle are provided for the S-band and X-band receivers as well as for the exciter reference frequency. Individual counts are taken with a granularity of 10 times per second. The MDA computes Doppler residuals from values provided by the NSS in a separate support-data file. Range, Doppler, and residual data are sent to the spacecraft navigation team.

In the future, the received signals will be digitized in the Block V receiver, and Doppler data will primarily contain the difference between the received frequency and a known reference frequency. The MDA will format the data and include the uplink frequency data and other ancillary data for transmission to the flight projects.

c. Command. The DSCC Command Subsystem (DCD) has as its primary function the transmission of commands to the spacecraft. As part of this function it receives spacecraft commands from JPL, developed by Project Operations Centers (POC's), processes those files, and transmits individual elements from selected files. The commands are received, verified, acknowledged, stored and subsequently transmitted to the spacecraft from the Command Processor Assembly (CPA) at each DSCC.

Actual spacecraft commanding is done by one of several methods of waveform modulation imposed upon the subcarrier by the Command Modulator Assembly (CMA), which has a maximum output of 2 kbps. The CPA, under direction from the POC, sends binary command elements to the CMA where they are modulated onto the subcarrier and transferred to the transmitter exciter. Commands are verified by a return line from the exciter to the CMA. Nonfavorable comparisons result in a cessation of the command process and a notification sent to the POC.

d. Radio Science. Radio science is the analysis of scientific information extracted from the variation in radio signals coming from a spacecraft. Depending on the information being deduced, the variation of interest might be the Doppler shift, phase, polarization, or the frequency spectrum.

The received signal can have variations from three causes:

(1) the motion of the spacecraft.

(2) the influence of the media through which the signal passes.

(3) deficiencies in the hardware used to capture the signal data.

Media that can affect the signal include the atmosphere, planetary rings, and solar corona. In order to detect the sought-for effects, instability in the ground system frequency standard must be minimized via ultrastable hardware and by the removal of identifiable noise during signal data processing.

Radio science signal processing performs the following actions:

(1) Accepts commands from, and provides status to, the operator.

(2) Obtains signal variation predictions from JPL.

(3) Provides tuning information to the receiver.

(4) Converts the analog signal to digital data.

(5) Records signal data and ancillary data to storage.

(6) Reconstitutes the analog signal for real-time monitoring.

(7) Transmits partial signal data to JPL for validation.

e. VLBI. This technique uses extragalactic radio sources to determine UT1, Earth polar motion, the relative positions of stations on the Earth, clock synchronization among the stations, and clock stability. VLBI is also used to maintain and enhance the JPL catalog of extragalactic radio sources. Delta-VLBI alternates between extragalactic radio sources and spacecraft radio sources to accurately determine the position of the spacecraft.

VLBI processing measures the time difference between the arrival of a radio signal at two or more different receiving stations on the Earth (Fig. 4). The signal source may be either a spacecraft or an extragalactic source. The time difference is determined by correlating the two received signals.

Narrow channel bandwidth (NCB) capability is normally utilized. However, for cataloging extragalactic radio sources, wide channel bandwidth (WCB) capability is utilized. The radio signals are transmitted by the spacecraft and by extragalactic sources; received by two or more physically distant DSN antennas; tuned and translated for several simultaneous frequencies by the receivers; digitized and stored by the signal processors (except for WCB data); and transmitted to JPL for correlation. The VLBI signal processors perform the following functions:

(1) Accept commands from, and provide status to, the operator.

(2) Provide tuning and control commands to both receivers.

(3) Convert the NCB analog signal to digital data.

(4) Record the NCB signal and store ancillary data on a disk.

(5) Store WCB ancillary data on tape.

(6) Transmit stored NCB data to JPL for correlation.

(7) Monitor the coherency of the signal and inserted tones.

2. Monitor and Control. The DSN Monitor and Control subsystem consists of JPL (Central Site) and DSCC components. A summary of the major functions at each site follows.

(1) JPL. The Network Operations Control Center (NOCC) is the present central-site focus for monitoring and is the primary interface between the DSN and the POC's. NOCC negotiates tracking schedules and generates sequences of events (SOE's), angle and performance predictions, and standards and limits. These support data are stored in databases and periodically distributed to the DSCC's. During spacecraft tracking, NOCC monitors the status, configuration, and performance of DSN equipment and subsystems. This function includes acquiring real-time data from each DSCC and monitoring all DSN missions. Monitoring is aided by real-time displays on workstations with alarm, event, and advisory messages, and access to historical logs.

(2) DSCC Site. Prior to an acquisition, equipment is configured to provide a logical processing sequence. During an acquisition, operations personnel issue directives through a workstation to control the subsystems. The functions are divided into two hierarchical subsystems: the Complex Monitor and Control (CMC) and the Link Monitor and Control (LMC).

The CMC subsystem is responsible for unassigned equipment pool monitoring, link assignment management, support data management and distribution to subsystems, multilink equipment monitoring, configuration files maintenance, and equipment calibration.

Whereas the CMC covers all links at a Complex, the LMC subsystem is concerned with only one tracking antenna and its associated ground data system elements. The main LMC functions are subsystem control by operator directives, health and status monitoring of subsystems, receipt and presentation of subsystem configuration displays, and monitor data distribution to POC's (for missions launched prior to 1991).

B. System Interfaces

In the following sections, the input and output interfaces to the GIS are identified and the data transfer rates are estimated in order to provide a basis for performance trade-offs in Section IV.A.

1. Input Interfaces. At the Central Site, inputs to the GIS include mission schedules and sequences of events from the POC's to the NOCC and spacecraft commands from the POC mission-support teams.

At the DSCC's, the inputs to the GIS include the spacecraft signals from the Block V receiver and monitor and control data from many devices and subassemblies in the front-end area.

a. Central Site

Planning. Mission schedules and sequences of events from the POC's are communicated to the NOCC. The data rates are generally low, less than 10 kbps.

Commands. Spacecraft commands are generated by the flight project planning and sequence team and transmitted to the DSCC Command subsystem through a central-site interface. These data rates are also usually less than 10 kbps.

b. DSCC Site

Antennas. There are currently four primary antennas at each complex: one 70-m antenna, two 34-m antennas, and one 26-m antenna. A 10-m antenna is currently being added. A new antenna every three years is forecasted, for a total of about seven to ten antennas at each complex by the years 2000–2010.

Block V Receiver. The Block V receiver is a planned DSN upgrade that will eventually be used in all SPC's. Each SPC will require one Block V receiver with sufficient channel processors to accommodate the subcarrier signals from each spacecraft. These signals may be transmitted on various bands (i.e., S- , X- , and possibly Ka-band in the proposed time frame) and different polarizations. During a given acquisition, one channel processor must be logically connected to one telemetry processor. In this study, approximately 22 channel processors and telemetry processors are estimated for the year 2000 and 32 in 2010.

A functional block diagram that also includes the telemetry subsystem is shown in Fig. 5. Each channel processor of the Block V receiver has a planned output of 26.4 Msps (megasymbols per second, with 8-bit symbols) for current X- and S-band missions; however, advancing to Ka-band communications will enable a 165-Msps planned output rate.

As discussed earlier in Section II.A, redundant bits are usually added to spacecraft telemetry for forward error correction. Viterbi code may be used for high-Earth orbiter and lunar communications at a 1/2 rate (2 bits for every information bit) and deep space missions at a 1/6 rate (6 bits for every information bit). RS code may also be used to spread the impact of burst noise on the space link; a typical overhead is about 14 percent. Figure 6 illustrates the effect that this coding has on the receiver-to-telemetry communication rate. At the Block V planned output rate, which handles a spacecraft-information rate between 4 and 11 Mbps—depending on the coding—the transfer rate to the telemetry processor is 26.4 symbols per second, or about 211 Mbps.

Each Block V receiver also provides Doppler and ranging data at less than 1 kbps, monitor data at less than 50 kbps (including spectral data), and radio science data at less than 1.2 Mbps (100 ksps with either 8 or 12 bits per sample).

2. Output Interfaces.

a. Central Site. The outputs of the GIS are usually transmitted to flight project teams and other users, or stored temporarily on portable media for off-line delivery to users (see Section II.D). Some data, such as telemetry data, may be transferred to archival systems, such as the Planetary Data System, for later distribution on demand. Outputs usually emanate from JPL; however, provisions are made for direct delivery from the DSCC to non-NASA POC's.

Telemetry. Most present deep space missions have low data rates in the 10- to 100-kbps range. High-Earth orbiters have data rates ranging from 100 kbps to several hundred Mbps. Future missions, such as the Space Exploration Initiative (SEI), will range from 10 Mbps for Mars support to 100 Mbps for lunar missions.

Tracking. The tracking subsystem provides metric data in packet format to the flight projects navigation team. These formatted packets are relatively low-rate, that is, less than 10 kbps.

Radio Science. Radio science data are produced by the Block V receiver. It is forecasted that two or three concurrent receivers may each be producing up to 1.2 Mbps in the forecasted time frame.

VLBI. VLBI is an emerging technique for computing spacecraft range. Considerable celestial radio-source mapping is required to ensure its success. While the ranging function is performed only occasionally, the mapping is done on a daily basis. The VLBI subsystem typically acquires 20 to 30 minutes of high-speed data into a file (e.g., 130 Mbps in the NCB subsystem) and plays it back slowly for processing. In an effort to improve reliability of the acquisition, it is envisioned that the proposed VLBI subsystem will acquire data in a quick-look mode for 10–20 seconds, process it, and transmit real-time corrections back to the station.

b. DSCC Site At the DSCC's, the outputs of the GIS include monitor and control data to many devices and subassemblies in the front-end area.

3. Availability. The total loss of telemetry data in the DSN from 1986 through 1990 amounted to 2.9 percent,

or 4814 hours, of the scheduled support time, according to a recent study. (See Table 2; the items listed in that table account for 98 percent of the total lost hours.) Steady improvement is evident from the yearly data: 3.9, 3.5, 3.5, 2.4, and 2.2 percent for these five years (decreasing an average of 12 percent per year). These percentages are derived from planned hours of operation that have been lost and do not include such operational delays as calibration and scheduled maintenance.

More interesting are the causes of these lost hours as attributed to the subsystems and assemblies. The 4814 hours of loss were attributed to 22 subsystems located from the front-end areas to the NOCC, plus radio-frequency interference and undetermined origins.

If a major effort is made in the future to improve the availability of these data subsystems, control interfaces and control assemblies, and the way the front-end subsystems are controlled, it is conceivable that a 40- to 50-percent reduction of lost telemetry hours can be achieved. As compared with a 2.2-percent telemetry hour loss in 1990, a reasonable goal can be set at 1-percent loss, or 99-percent availability, which is consistent with the 12-percent annual decrease already noted. Another major improvement in overall availability can be made by reducing operational delays in precalibration, postcalibration, and scheduled maintenance.

C. Future Mission Sets

Table 3 summarizes the new DSN missions set and includes deep space and high-Earth orbiter (HEO) missions. (The Earth Orbiting System [EOS] is not shown.) Approval is expected for three to five of these missions. International cooperation missions may be supported. Also, the DSN may be requested to serve as a contingency network for near-Earth missions, including Space Station Freedom and shuttle missions. Support to the following missions are not mandatory requirements but were considered in evaluating the architectures:

(1) SEI (Mars), 10 Mbps.

(2) SEI (Lunar), 100 Mbps.

(3) Tracking and Data Relay Satellite System (TDRSS) backup (e.g., EOS), up to 300 Mbps.

(4) Advanced Tracking and Data Relay Satellite System (ATDRSS) backup, up to 650 Mbps.

(5) International cross support, undefined.

Table 4 summarizes three views of the future data requirements. The architectures developed later in Section IV.A will use the moderate forecast for most of the evaluation, and one can consider the aggressive forecast as a possible extension.

D. End-User Interfaces

End-user interfaces fall into the following four categories: mission-specific teams, principal investigators, data archives, and cross-support that the DSN provides to other agencies.

Figure 7 shows the current end-user interfaces between the GIS (including the DSN and the SFOC) and typical flight projects. Also shown is the type of information flowing to and from the end users.

With continued development of the SFOC capability under the Multimission Operations Systems Office (MOSO), the relationship between the end users and the GIS will probably change significantly. These changes may cause related interfaces to change. The driving force for the changes is to reduce mission operations and data analysis costs. The approach is to identify and replace mission functions that could be performed by multimission tools and teams. One of those changes, which can already be seen in the Mars Observer and CRAF/Cassini plans, is the use of the remote Science Planning Operations Computer (SPOC). The SPOC acts as a remote extension of the SFOC by providing command, telemetry, and navigation data access, and planning capabilities to the project scientists. SFOC-based multimission operations teams and remote extensions of the SFOC will continue to evolve.

The interfaces for each class of end user are shown in Table 5. The table is organized to show the direction of data flow (to and from the GIS), the type of interface, and a brief description of the content or purpose of the interface. Note that for some of the interfaces the interface medium is magnetic tape. The GIS in years 2000–2010 will likely use other physical storage media.

1. Mission-Specific Teams. JPL planetary spacecraft have a great deal of similarity in their mission operations design. The particular names given to the various teams may sometimes be different, but when the functions are identified, the teams that most spacecraft projects agree upon are: Mission Control Team, Planning and Sequencing Team, Navigation Team, Spacecraft Team, and Radio Science Team.

2. Principal Investigators. The term principal investigator is applied to the cognizant individual who leads a science experiment. The team may include the principal investigator, coinvestigators, and interdisciplinary scientists. The investigators of the future, especially for the

exploration missions, may do most of their work for the project at their home institution. From there, they will submit observation requests, receive virtual channel data from their instruments as well as ancillary data they need to analyze their data, and return their reduced data to the project for archiving. They may also receive data on tape (or other media) if an instrument's data rate is too high, or too expensive to transport.

3. Data Archives. The primary data archive of concern to this architecture study is the Planetary Data System (PDS). This archive is a distributed system that has a central computer which houses the master catalog and the archive operations functions. Data are archived and maintained at the distributed nodes that are defined by various planetary science disciplines (e.g., planetary geology, fields and particles, etc.). The interface with the GIS is still under development; for example, for the Magellan project, the interface has been negotiated as a CD-ROM interface in which the project produces the compact disks and delivers them to the PDS. The future interface will probably be in some other electronic form.

4. Cross-Support Services. The GIS provides several forms of cross-support services, for example:

(1) Backup support for the TDRSS (i.e., for shuttle, Space Station Freedom [SSF], and EOS).

(2) Reimbursable support to international space agencies (initial acquisitions, tracking support, etc., for National Space Development Agency [Japan], Institute of Space and Astronautical Science [Japan], European Space Agency, and others).

(3) Arraying of antennas, such as was done for the Voyager encounters at Uranus and Neptune.

For backup and support to international agencies, the data exchange is through the use of CCSDS data packets.

III. Technology Forecast and Projected Impact

This section forecasts new and emerging technologies in the areas of computer hardware, software, data communications, and human interfaces that are of potential benefit to the DSN GIS in the 2000–2010 era.

A. A Look at the Future

The following perspectives from a research scientist, an operator, and a software developer are offered to stimulate creative thinking and are not constraints on the architectural study.

1. A Research Scientist's Perspective.

"I had a personal computer on my desk 10 years ago, so that much hasn't changed. The personal computer that is on my desk now has dramatically higher performance, a larger screen, and larger storage capacity than before; however, the biggest change has come in the way I use it. For me, and for most of the people I deal with, the personal computer has replaced the telephone in all sorts of ways. I have the regular personal computer applications I use every day—the spreadsheets, a word processor, and a couple of small databases—but what really makes the system useful for me is that it is linked into the local network I share with my staff, into national and international computer networks, and into the flight project office, where I can access the really large databases we depend on for up-to-date scientific analysis and monitor the latest input from my planetary instrument.

"Today I often have no idea where the data I'm using is actually stored. I don't really need to know. Sometimes I'll pull summary information from the Planetary Data System, combine it with up-to-the-minute status reports on a current mission, pass the whole package to my science analysis team for some 'what if ...' projections—all without having to think about the mechanics of how things really link together. For all I know the data comes by way of Siberia. When I share the data with my colleagues we can discuss it as we analyze it on our screens and propose and conduct analysis during 'electronic meetings.' We hear everyone's voice and see team members and review the data—our attention is focused on the screen.

"Perhaps the most pervasive difference today is that we are much less dependent upon technicians than we were in the past. Most of our analysis is produced by our professional staff—the knowledge workers—with the assistance of sophisticated software tools. Compared to just a few years ago, there is much less manual analysis to be done, because of our reliance on 'electronic assistants.' We often annotate results with voice and video clips and share it with colleagues for their review and comments.

"I have a lot more time today to focus on doing the 'right things,' because we have learned to use the systems to avoid spending time doing the 'wrong things'—things that don't contribute to the accomplishment of our research. We make fewer errors because we are better able to monitor activities, to do reasonableness checks. We spend less time looking for related data, particularly pieces of paper, because more of our files are stored in electronic form,

and we can do extremely rapid electronic searches, comparisons, and associations. The problems associated with performing analysis in stages have been minimized because we have replaced much of the paperwork with electronic transactions. There just isn't that much 'float' in the process anymore.

"If you look around this office, you can see that science still uses paper, but in many senses it is a different kind of paper: It's what you might call secondary paper—reports, studies, and day-to-day reports—the 'real-time' information is almost entirely on the screen. And as I've said, that means a lot less looking and waiting. I just call up. It's there when I need it, it's in a format that I can use, and it's reliable. That's most important, of course—that I can trust the information to be correct and current."

2. An Operator's Perspective.

"The most significant change in my job is that all the equipment is remotely operated from here in Pasadena, and my job involves monitoring a highly automated process. If any subsystem needs to be adjusted, I can do it from my workstation. Every action has explicit, reliable feedback. I can also query the subsystems and get immediate responses. The subsystems themselves are smarter: I can now control them with much higher level instructions; they are self-monitoring and self-diagnosing and I can place one subsystem under the direct control of another. The biggest difference I see in operations is that overall operations seem more integrated, and the networks and tools available to me through my console have made the system more responsive to what I need to do to control it. I make fewer errors because I am better able to monitor activities, to do reasonableness checks.

"The DSN supports more missions now than 10 years ago, and I am able to maintain several spacecraft-to-ground links through multiple antennas at the Complex. JPL now arrays antennas routinely, and is progressing towards using one antenna to communicate with several spacecraft.

"I use automated monitoring tools to keep a constant check on the operational status. Graphics and animation on my console present information to me in such a way that I can effectively monitor several links at once. When things go wrong, fault detection, isolation, and diagnosis tools help me to narrow down the possible problems, suggest alternative solutions and quickly get a message to the maintenance personnel at the Complexes. Expert knowledge is available from my console so I can benefit from having the 'experts' available to me whenever I need them. Maintenance people at each Complex have similar tools to help isolate and repair equipment.

"The procedures and manuals are now all available on the screen. Instead of having to sort through volumes of information to pull together the bits and pieces I need, I can now navigate through a library that has not only the documentation, but photographs, animated sequences, and voice notes all linked together. Getting the 'big picture' no longer involves a treasure hunt through documentation libraries and coworkers' notebooks. When I discover a problem in the documentation or procedures, I can instantaneously annotate the procedure, and even recommend changes.

"Embedded training programs enable me to practice the performance of difficult operations in an environment exactly like the operational one. Instead of sitting idly by when I'm in the middle of a long track, I can now place my console in a training mode. I'll run scenarios and practice procedures while background processes monitor the health and status of the space link. If any problem occurs, the system immediately notifies me and provides me with the necessary background information.

"Because of the training system, I am now proficient at performing operations tasks I would rarely work on a decade ago. I am better informed, understand the activities on a much deeper level, and am better equipped to deal with emergencies than before. I'm also more aware of what's important to our customers and can work with them to deliver the best possible data product.

"I now provide better problem reports to system developers with a quick and easy way to document problems: if a failure occurs, I dump all the appropriate information into a file and the file provides a history of events surrounding the failure and a clear picture of the health, status, and configuration of the station when it surfaced. This way, the developers and maintainers have first-hand information to work with—and problems can be evaluated to make sure they don't happen again."

3. A Software Developer's Perspective.

"International standards apply to the space and ground communications portion of the DSN, and this eliminates my need to create new communications software, and helps me focus more on what the user needs to see and control. International standards have made it easier for my division to assemble a substantial library of space-application modules, many developed by foreign programmers.

"Most of my software engineering efforts are focused on building maintainable code, and interfacing it to tested

and proven software. One major change is that there is greater emphasis on designing fault tolerance into the system. One way we do this is by integrating common monitor and control modules into the software architecture that enable remote monitoring of hardware and software performance.

"There is a mix of computer platforms in the Ground Information System; all use identical operating systems. Those that are roughly equal in capability can be dynamically assigned tasks. This makes the information system more dependable by making it tolerant to isolated computer failures. Network, computer, and software architectures are carefully designed to be interoperable and fault-tolerant. They must be—the DSN supports many more real-time users than it did ten years ago. These users are expecting portions of their data within seconds of capture and want to send real-time signals to their space instruments.

"The human interfaces that I design involve text, sound, and graphics—and sometimes include voice and video. The format of these data types has been standardized, and many human-interface application modules are available—either commercially or from software libraries. I work closely with human factors engineers and users to provide all the features they need to do their job at one location—at their workstation. The computer has also become the focus of communications between operators and engineers like myself, which simplifies their access to me.

"I obtain the majority of my application modules from software libraries. My development team tends to use object-oriented software modules; the porting of modules to other machines is easier and faster. One measure of my success as a programmer is how many people want to reuse my code for their applications; the usefulness of my code to other programmers is very important to me."

B. Computer Hardware

The 1990s will see extremely high-speed electronic systems that will be increasingly compact, mobile, and lightweight. High-speed computer systems will include multiple high-speed microprocessors with increased parallelism; memories with ever-increasing densities; and specialized high-performance chips, such as video chips with both analog and digital memory interfaces and voice-processor chips with adaptive recognition capability.

The growing power of functions at the chip level enables greater degrees of parallelism. It will become more feasible to dedicate a processor of the proper kind to each element of a system in subsumptive architectures with hierarchic arrangements of functions in increasingly detailed layers of computation, and systolic data-flow architectures with patterns of specialized processors arranged in data-flow orientation.

Initially, the DSN computers were JPL-built controllers, which gave way to minicomputers of several kinds, tending toward microcomputers today. Recently there has been some work to seek very-large-scale-integration solutions for some of the functions of the DSN.

1. Computers. The DSN will probably require microcomputer, minicomputer, and mainframe levels of capability in its future architecture. Mainframes, if used, will probably be located at the Central Site, with minicomputers and microcomputers distributed at the Central Site and DSCC's.

a. Microprocessors. Since the late 1970s, there has been a steady increase in the capability of microcomputer chips in execution speed, addressing capability, input and output (I/O) bandwidth, and instruction sets. This trend is expected to continue, with processing speed increasing by a factor of 10 or so every five years.

Compatibility between successive members of a manufacturer's line is reasonably good, at least at the level of users of commercial software packages and operating systems, although significant rework has often been necessary on the part of software suppliers to support a new version of a workstation microcomputer chip. The trend to graphical user interfaces (GUI's) has helped to disguise the extensive modification required to support new workstations.

Mainframes are emerging with higher speeds and larger storage capability, which solidifies the mainframe as the machine of choice for large numbers of users running database-oriented applications (e.g., in business). For scientific users running computation-oriented applications, the appearance of the new mainframes provides additional powerful analytic tools and the potential for new discoveries.

Today's microcomputers operate up to over 75 million instructions per second (MIPS), 22 million floating point operations per second (MFLOPS), with over 800 Mbytes of internal mass storage. External devices may extend the storage to over 10 Gbytes. So-called optical juke boxes extend the mass storage to over 300 Gbytes.

The points of demarcation among the performances of microcomputers, minicomputers, and mainframes, and be-

tween workstations and systems are not firm. NCube, Inc., for example, has introduced a computer capable of speeds up to 1 billion floating point operations per second (GFLOPS) using 512 processors, with up to 1 Gbyte of storage, all in one cubic foot of space and weighing about 60 pounds. The computer is housed in a portable chassis that fits into a personal computer carrying case.

Computer chip manufacturers predict that 1- to 2-GFLOPS capability will exist at the board level within ten years. The Intel Micro 2000, now in design, will contain 50–100 million transistors in four central processing units, two vector processing units, a human interface unit of 4 million transistors to provide high-resolution, full-motion graphics, and a 2-Mbyte cache to reduce off-chip reaches for data. In addition, this chip is planned to be upwardly compatible with the Intel 386 instruction set and would run all software that currently runs on that processor.

b. Parallel Processors. Parallel processors fall into two main categories: (1) single-instruction multiple-data (SIMD) machines, in which many processors work with the same program on vectors or arrays of data values, and (2) multiple-instruction multiple-data (MIMD) machines, in which each machine has its own program and the (more or less) independent machines communicate by means of I/O interconnections. The SIMD machines are particularly well-suited to calculation of multiple-degrees-of-freedom problems, such as orbit determination, while the MIMD machines are well-suited to simulation of complex events.

All these machines currently require extensive reworking of existing processing algorithms to take advantage of parallel processing; unfortunately, none of the existing implementations promises or even suggests compatibility with later versions of parallel machines.

2. Digital Signal Processors. Digital signal processing can be advantageously applied to several functions of the DSN. These include receiving, transmission, switching, encoding and decoding for error correction, security, storage, and retrieval.

Very-high-speed integrated circuits (VHSIC's) are expected to do complex signal processing dramatically faster with smaller chips. This development will enable advanced forms of modulation, synchronization, and coding. The realization of the Viterbi algorithm on a chip is an example of the application of this emerging technology.

Signal processors may operate as circuit boards managed by external computers. Current capabilities of such board-level devices include 150-MFLOPS 32-bit vector processing, with 400-Mbps access to memory. The ability to do fast floating-point operations permits a very flexible signal-processing design. The speed of commercial digital signal processing boards is expected to keep pace with computing speeds, growing at least by a factor of ten over the next decade.

3. Storage. Storage technology is rapidly increasing the capability to rapidly and reliably store and retrieve large volumes of data. Three types of devices are all current areas of research: solid state, magnetic, and optical.

In 1990, the semiconductor industry announced the first 8-Mbyte dynamic random-access memory on a single 10- × 20-mm chip containing 40 million electronic components. At present, 32-Mbyte and even 128-Mbyte solid-state storage chips are in development. Access times, meanwhile, are falling steadily; whereas about 80 nsec is common now, about 60 nsec will be the norm within one year in standard complementary metal oxide semiconductor (CMOS) technology, and 35 nsec may be achieved using bipolar CMOS within another year. JPL is presently developing a 4-Gbyte, 320-Mbps vertical block line device, with an expected latency under 100 msec. By the years 2000–2010, it is not unreasonable to project solid-state devices reaching 1 Gbyte on a single chip, with 20-nsec access.

Magnetic device performance will also continue to rise dramatically. Magnetic tape is expected to reach 1–10 terabytes per reel, with a transfer rate ranging from 300 Mbps to 1 Gbps by the 2000–2010 time frame. Magnetic disks are expected to reach 0.15 Gbyte/cm^2 by then, with only about 1-μsec latency. JPL is currently conducting research in the application of tunneling microscopy to magnetic memory, which could reach densities of about 1 terabyte/cm^2 by the year 2020.

A less dramatic rise is being predicted for optical storage technology. Currently, digital optical tape systems are capable of 50 Gbytes per cartridge and a 24-Mbps sustained data-transfer rate, with seek times of less than 4 sec per Gbyte. The optical diffraction limit will permit only up to about 1 terabyte per cartridge, and only about a 30-Mbps readout is expected. Similarly, optical disks are projected to limit at about 45 Mbytes/cm^2 with 1-μsec latency. Because of these limitations and the projected 3:1 superior performance of magnetic technology, it has been questioned whether there will be a market for high-performance optical storage technology.

4. Hardware Integration. The Department of Defense (DoD) is developing multichip-module fabrication

and packaging technology that drastically decreases the volume, power, heat, and cost of high-performance computational systems, while increasing their speed, reliability, and effectiveness. The concept of three-dimensional (3-D) packaging is not new, but only recently have vendors made significant progress in implementing the process. The DoD has identified multichip-module technology as an enabling technology and plans to establish multiple manufacturing sources for each type. Figure 8 illustrates the concept.

A number of techniques are being explored. A Hughes Aircraft 3-D computer uses 4-in. diameter wafers. Other systems utilize bare chips (the dies themselves), some of which are trimmed to remove line drivers. Line drivers are rarely needed since the distances are reduced and no pinning is involved. The systems are considerably cooler, since as much as 85 percent of the power budget and heat budget for a chip is used by line drivers.

The multichip modules or wafers are interconnected by very-high-speed links. For example, a Hughes module has an internal rate of 160 Gbps, and a Texas Instruments network operates at 20 Gbps on the module and at 800 Mbps between modules. The various packaging techniques used by vendors often utilize off-the-shelf commercial chips, such as static random-access memory, dynamic random-access memory, central processors, communications processors, and digital signal processors. Repair of these systems will usually be done by replacing a module or wafer. Since there are typically only three to five types of modules, the parts problem could be minimal.

The systems currently being developed are for weapons control. They are programmable (sometimes in multiple languages) and reconfigurable. Other potential applications presently under consideration include avionics control, general-purpose supercomputer workstations, and sensor analysis systems.

A system being developed by the Environmental Research Institute of Michigan has performance goals of 30 MFLOPS and 15 MIPS. It will fit in 200 cubic inches and is expected to cost about $85,000 in volume quantities. The same capability in 10 years is projected to cost less than $10,000 (in current year dollars).

5. Reliability. The GIS has a functional availability goal of over 99 percent in the envisioned time period. This may be approached by a fault-tolerant design with appropriate fault detection, isolation, and recovery technology. One key need is for highly reliable components in each subsystem. This section summarizes a forecast of computer-oriented reliability issues.

a. Current Architectures. In 1975, the mean time between failures (MTBF) for disk drives was only about 3000 hours. Today, the MTBF for the newest physically small units is quoted in the 200,000-hour range. Much of the increased reliability has been attributed to the reduction in size. The 1991 disks are 3.5-in. units that hold as much as ten times more data and have transfer rates five times faster than the 1975 units.

Failure rates for tape units have not improved as dramatically, but helical scan devices have jumped in reliability from 28,000 hours MTBF to 40,000 hours MTBF within the last year, validated by field experience.

The next most frequent failures encountered in computer systems occur in the cabling and connection equipment and on the boards themselves. Flexing and stress, due to temperature and mechanical variations, create socket and tracing breaks. Years of age on connective and board components becomes a major contributing factor. The incidence of memory and processor failures, after initial so-called infant mortality, has been so rare that industry figures are not usually published.

b. Advanced Architectures. The environmental conditions that DoD's advanced integrated architectures (such as wafer and composite module devices) are intended to survive exceed those found in the conventional computer room: mechanical shock tests up to 168,000 g's (required for impact, and rail guns), thermal cycling from liquid nitrogen to boiling oil temperatures, vacuum tests, and radiation tests.

The failures that are experienced are inside the chips and dies themselves. Chip failures occur due to operating temperatures and radiation, usually not to mechanical stress in these packages. The design process can provide for cooling by convection or conduction flow where necessary.

Most of these dense packaging designs have built-in redundancy. Some have 50-percent spare components; others have up to 200-percent redundancy in connectivity (much of which can be switched dynamically); and still others have fully active switching to bypass components that are failing. These design features further enhance the reliability of the unit.

It appears that new, dense packaging architectures present a number of advantages over present traditional designs because the number of connections is significantly reduced. Shock failures due to physical or thermal events are virtually eliminated. All these attributes contribute

to significant increases in the reliability of the solid-state units and systems.

C. Software

The 1980s witnessed the advent of major efforts to standardize many aspects of computing, including programming languages, operating system interfaces, and network protocols. By the year 2000, most of these efforts will have resulted in wide adoption of international standards. The standards process will provide a common basis on which to build distributed computer systems.

1. Open Software and Standards. A major concept is that of open systems. Although not rigorously defined, in this context the term open refers to the willingness of a computer or software vendor to offer products compatible with those available from its competitors, usually because they mutually agree on standards. Open systems benefit users because they yield greater independence in the procurement of hardware and software.

Planning for an open system makes it possible to implement highly portable, hardware-independent, scalable computing systems through the use of existing software technology, an awareness of trends, and by adhering to a plan for compatibility.

The following paragraphs summarize various developments in the evolution of open systems that will have important implications for the DSN GIS. Large-scale computer applications often rely on capabilities beyond those specified by a programming language definition. For example, file system services, process creation, and network I/O are support functions that require additional specification. The universe defined by this broader specification is called an application environment.

Several related (and mutually consistent) efforts to set a standard application environment are under way. The following discussion is based on the Open Software Foundation (OSF)/Application Environment Specification (AES). AES is itself based on various other standards (both adopted and emerging) as noted. The discussion is intended to indicate current directions in the evolution of open systems.

a. Programming Languages. At the present time, the C programming language is the language of choice for systems software. C is a middle-level language originally developed to replace assembly code for the most demanding systems' programming tasks. C language deals with the same sort of objects that most computers do, namely, characters, numbers, and addresses. It may be used in the GIS for applications that strongly manipulate operating system resources and the computer platform, such as operating systems, compilers, editors, and low-level I/O software.

C++, a superset of C, is a general-purpose language. C++ was designed for larger, more structured applications than C, and provides flexible and efficient facilities for defining new data types. C++ hides much of the application functionality behind object interfaces, hence it is referred to as an object-oriented language. It will be applicable to the GIS by providing a tool for the rapid development of highly portable, structured application programs. It is applicable to human-interface displays and database management systems.

Common LISP has proved to be a useful and stable platform for rapid prototyping and systems development in artificial intelligence and other areas. Artificial intelligence applications written in Common LISP are also successfully integrated with software written in other languages, including C, C++, Fortran, and Ada. The unique capabilities of Common LISP, which include memory and data structure management, flexible typing, symbolic representation, processing capabilities, and functional programming style, ensure that the language will remain the dominant language for most AI research and application software for the foreseeable future.

Extension languages support a new style of programming that has emerged in recent years called language-based programming. With this style, a programmer presented with requirements develops a special-purpose language in which a program that satisfies the requirements can be expressed succinctly and clearly. The ultimate program is written in this higher level, special-purpose language. This style allows easy experimentation, testing, and evolution of the application much more readily than do more traditional top-down, or structured, programming styles.

In the future, more and more programs will be designed and implemented as multilayer language systems, with the upper layers calling languages implemented in lower layers. High levels of abstraction will be provided by upper layers, allowing applications to be programmed in very few lines of code. Portability will be routinely achieved by defining one of the interlayer boundaries as a portability boundary, with machine-dependent code allowed only below the boundary. Any code above the boundary will port from one machine to another without change. Table 6 summarizes the languages that may be considered as open languages.

b. Operating Systems. The 1980s saw several important trends in scientific and engineering computing. The first and most important was the emergence of UNIX as the dominant operating system. The wide availability of UNIX also encouraged the development of highly distributed computing systems based on the client-server model.

The Institute of Electrical and Electronics Engineers (IEEE) has recently adopted Standard 1003.1-1990, the Portable Operating System Interface (POSIX.1). This document defines a C-language interface to an operating system (i.e., UNIX). POSIX.1 is also an International Organization for Standardization (ISO) standard (ISO 9945-1: 1990). The two major players in the open operating system field, UNIX International and the Open Software Foundation, have pledged compliance with POSIX.1 for their respective operating system products.

c. User Interfaces. The X Window System from Massachusetts Institute of Technology Project Athena has established itself as the dominant technology for windowing systems under UNIX. Standards for X Windows are developed under American National Standards Institute (ANSI) committee X3H3.

At a higher level, the most clearly drawn battle line in open systems is perhaps between the OSF/Motif and Sun OpenLook graphical user interfaces, At the present time Motif is proprietary software; OSF publishes the standard and the software itself must be licensed from OSF individually for each workstation. Efforts are under way to build a virtual interface supporting either technology.

d. Network Services. AES supports the Internet suite, plus other selected Internet services developed at the University of California, Berkeley, for UNIX networking. AES also includes support for selected Open Systems Interconnection (OSI) protocols.

From the standpoint of software architecture, the choice of network protocol is often incidental, since it is usually hidden by higher layers in the distributed computing environment (see Section III.C).

2. Distributed Computing. The concurrent explosive growth in the power of inexpensive computers and availability of network technology has led to widespread adoption of distributed systems in which large numbers of loosely coupled computers communicate over local area networks. The problems of designing software systems that can take advantage of diffuse computing power are many and derive mainly from the additional complexity of simply having more semi-independent components.

Any monolithic computer system can be distributed by the ad hoc insertion of system boundaries and corresponding communications support. A properly designed distributed application, by contrast, will be built upon a framework that provides a portable programming interface and subsumes much of the complexity that accompanies distributed computing.

OSF's Distributed Computing Environment (DCE) technology includes support for building distributed applications in the UNIX environment. The following discussion highlights the key capabilities of DCE, again to demonstrate the current state of open systems evolution.

a. Remote Procedure Call. Client-server computing, in which a process requests actions on the part of another process, is a useful abstraction for exploiting the power of distributed computing systems. With the necessary communications support, processes on different computers can be made to cooperate to bring greater processing power to a single task.

A remote procedure call (RPC) is a logical extension of the local procedure call model used by many programming languages. RPC technology allows the programmer to specify interactions among components of a distributed system through a convenient and consistent method. In addition to the obvious benefit of making distributed applications easier to write, RPC also insulates those applications from the details of specific processor architectures and network implementations. It therefore contributes to the portability of such applications. A well-written RPC application, for example, can operate among several vendors' systems, and use different network protocols (e.g., Internet or OSI) simultaneously without code changes.

b. Naming Service. The added flexibility that can be obtained from distributed computing leads immediately to a problem of information management. A client process, for example, needs to know how to contact a server process on another computer. If the binding of service to computer is deferred for as long as possible, reconfiguration can be accomplished, when necessary, with minimal side effects.

A naming service is a mechanism for storing and distributing such information throughout a network. Well-written applications should exploit the naming service for any information that may be subject to change. Source code is thereby insulated from such changes, and greater run-time configuration flexibility is afforded.

c. Threads. Figure 9 shows a process requesting data from an upstream server and responding to data requests

from a downstream client. The reading and writing activities are not directly coupled; a cache buffer is inserted so that a temporary decrease in output data rate does not cause a corresponding slowdown in input flow and processing. In addition to reading and writing data, the process also communicates with a monitor and control process. It has therefore, several semi-independent flows of control.

The general problem of multiplexing loosely synchronized activities has been addressed through a variety of mechanisms. Real-time systems typically provide a rich set of tools (scheduling primitives, interrupt service routines, etc.) but generally at the cost of increased complexity and reduced portability. Multitasking systems (e.g., UNIX) provide mechanisms based on interprocess communication, but these are often expensive in performance overhead.

Perhaps the most useful conceptual model for this problem is the thread. A thread is simply the flow of control within a process. Because all threads in a process exist in a single address space, they can communicate efficiently through shared memory. Synchronization among threads is accomplished via scheduling primitives provided by the threads implementation.

The threads model simplifies the design of multiplexed processes by decomposing them into a collection of threads whose temporal relationships are explicitly specified. Considering the example in Fig. 9, the decompositions might be as follows:

(1) Read. Wait for server input. Read and process data. Wait for room in output cache. Queue output. Repeat.

(2) Write. Wait for client request. Wait for data in output cache. Write data. Repeat.

(3) Monitor. Wait for timer to expire. Report program status. Repeat.

(4) Control. Wait for control input. Respond to directive. Repeat.

An IEEE standard for threads (POSIX 1003.4a) is currently in development.

d. Security Technology. Distributed computing systems need reliable methods for authentication, authorization, and the assurance of integrity and privacy of messages sent across the network. The DCE security technology is based on Project Athena's Kerberos (Version 5) service. Kerberos provides private authentication services and secure message service. The DCE authorization function is provided independently via POSIX-compliant access con trol lists.

e. Distributed File System. Another useful abstractior for building distributed systems is the distributed file system (DFS). A DFS is an efficient method for application-transparent sharing of file-oriented data among computers. While not appropriate for stream transfers, a distributed file system can nonetheless simplify certain data exchange problems, particularly distribution of software across the network.

3. Database Technology. The DSN has traditionally developed custom storage and retrieval methodologies for most of its database management systems (DBMS's). Experience at JPL and throughout industry demonstrates the wide applicability of relational database systems, particularly when implemented in a client–server mode of operation.

Relational DBM's. Such systems have been in use for approximately 20 years and have achieved reliability, performance, recoverability, operability, and security levels acceptable to most applications. Furthermore, related computer-aided software engineering (CASE) tools, program generators, report writers, windowing interfaces, distributed operations, and parallel operation are becoming available.

The Structured Query Language (SQL) has been accepted as an ANSI-standard relational database interface language. Virtually all relational database systems, and many non-relational database systems, provide an SQL interface.

Object-Oriented DBM's. Objects are data that have semantic definitions and procedures attached to them. They are examples of abstract data types. Object-oriented DBM's are designed to support the storage, manipulation, retrieval, and safety of objects. However, there is currently no single, consistent, accepted model or standard for object-oriented languages and systems. Object-oriented DBMS's may take years to be standardized.

Text DBM's. Most of the world's data exist in textual form. Text DBM's have been around since at least 1980. They were initially developed for use in the intelligence community for searching the enormous volume of public information found in publications. Text DBMS's are now emerging as tools to be applied in office automation, publication, and documentation. Text DBMS's suffer from the same weaknesses as object-oriented DBMS's. They have no single codification, definition, theory, or de facto standards. Current efforts in industry are being expended to

broaden the capabilities of text DBMS's to include image and graphical data.

Based on history, it appears that at least 20 years of intensive use and development are required to provide a user with a mature DBMS. Object-oriented and text DBMS's have neither the single solid theoretical base that relational systems have nor do they have a champion. Without such cornerstones the systems may not progress, mature, and become reliable before 2010. Therefore, based on history and present reality, relational DBMS's are viewed as the most likely candidates for database management systems for the DSN in the 2000–2010 era.

4. Algorithms and Heuristics.

a. AI and Knowledge Systems. Expert and knowledge-based systems, more generally called knowledge systems represent specialized areas of AI that have developed rapidly over the last ten years. It has been estimated that there are currently over 10,000 deployed, operational knowledge systems in the United States in many different areas of industry, services, and government. For example, at JPL, the Spacecraft Health Automated Reasoning Prototype (SHARP) knowledge system has been successfully applied to Voyager and Magellan real-time link analysis operations and Galileo operations.

Knowledge systems are computer systems that provide automated support to human workers, or high degrees of autonomy to large-scale systems. This technology has opened many entirely new classes of aerospace applications to computer automation, such as fault diagnosis, fault isolation and repair, resource planning and scheduling, design, and intelligent interfaces to computer systems. Knowledge systems extend the capability of information systems by manipulating data at many levels of abstraction in ways that were previously impossible or achieved only by humans.

The combination of high-performance hardware, software algorithms, heuristics, and technology from many areas of computer science allows knowledge systems to address application problems that were previously not amenable to computer-based solutions, such as dynamic replanning and expert fault diagnosis. In the GIS, knowledge systems will play a major, enabling role in the automation of operations, data management and analysis, and in the achievement of higher degrees of autonomy in the next generation of DSCC's.

b. Neural Networks and Fuzzy Logic. A neural network is an AI technique that is used to mimic the way the human brain works. A significant benefit of neural networks is their ability to exhibit acceptable computation performance in the presence of noisy or incomplete data. Although neural networks are most frequently implemented in software, the most impressive performance results from neural networks realized in hardware. Optics is a technology particularly well suited for implementing neural computers because of the relative ease with which a programmable, massive interconnection network can be optically synthesized. Experimental work along these lines is currently under way at the California Institute of Technology and elsewhere. At JPL, hardware neural network chips for resource allocation problems have been successfully developed using VLSI techniques.

Another interdisciplinary area of study related to AI and mathematics is fuzzy logic, originally proposed and mathematically formalized in the 1960s and only now undergoing widespread application. As an alternative to probabilistic reasoning, fuzzy logic provides methods for dealing with imprecise data and incomplete information in decision making. These techniques have been successfully used in a wide range of control applications, including space shuttle rendezvous and docking, elevator scheduling, and control of braking on subways. Like neural networks, fuzzy logic controllers can be implemented in special-purpose hardware, yielding impressive performance improvements over conventional controllers in many applications. Although these technologies are still experimental, it is expected that neural networks and fuzzy logic controllers in both hardware and software forms will find use in a variety of application problems in the GIS.

D. Data Transport

The data transport function transfers digital data among GIS computers over local and wide area networks. Government and industry are moving towards standard transport protocols, including message formats, for reliable and timely transmission. The goal is an environment where heterogeneous computers are capable of communicating among themselves by virtue of mutual implementation of ISO-standard protocols.

The protocols provide various services commonly organized into layers by the ISO OSI reference model (see Fig. 10). This section is roughly organized in terms of these layers. The protocols, while they are ISO standards, are often referred to as the OSI suite. For its part, the Federal government has chosen to require ISO-standard protocols. Within the large number of ISO protocols and options, the government has specified a subset of protocols for Federal networks and has published the list as the Government OSI Profile (GOSIP).

At the lower layers, protocols are required to move data reliably and accurately within local area networks (LAN's) and wide area networks (WAN's). The maximum transport speed of these networks depends on the physical characteristics of the medium, the transmitter and receiver technology, and the protocol processing speed. At intermediate layers, protocols must provide reliable process-to-process communication among different subnetworks. The goal of the intermediate layers is to interconnect multiple subnetworks to form a single logical network (an internetwork). The higher layers provide special services ranging from data compression to specific message and handshake procedures for file transfers and electronic mail.

The following sections summarize the current state of the art, forecast trends, and describe the impact of those trends:

1. Media. The physical properties of communications media limit the bit rate and transmission distance of digital links. Speed, limitations, and usage trends of several vital media in the GIS communications architecture are summarized in the following section. Common-carrier circuits are also considered in this section.

a. Wire. Twisted-pair wire is the most common medium-distance transmission medium (1–1000 m) and may be used to transmit data at rates up to 100 Mbps (depending on distance). It is often used for networks within buildings because twisted-pair wire is already a part of the installed telephone plant, relatively inexpensive, and easy to maintain. This medium will continue to be available because of the large installed base; however, it is primarily useful at older facilities.

It is expected that twisted-pair wire will be used in the future within locally instrumented areas to connect computers to peripherals and instrumentation by using such standard interchange standards as RS-232, RS-422, or IEEE 488. It will be an important medium for DSN monitor and control circuits.

Coaxial cable can support higher data rates than twisted pairs without emitting spurious radiation. Rates of hundreds of Mbps can be supported in both baseband and broadband implementations. Such current LAN standards as Ethernet 10 Base 2 (Thinnet) and 10 Base 5 (Standard) are relatively easy to implement. A popular monitor and control standard, the Manufacturing Automation Protocol (MAP), which includes a token-passing bus LAN standard (ISO 8802-4), is also implemented with coaxial cable.

The number of coaxial cable–based LAN's will likely decline because of the increased commercial availability of high-rate, low-cost optical fiber components. New-generation LAN's based on optical fiber, such as the Fiber Distributed Data Interface (FDDI), will become the LAN's of choice.

b. Microwave. Terrestrial and satellite microwave systems provide high-capacity channels (hundreds of Mbps) and relatively rapid installation. Although subject to weather degradation and environmental disturbances, they are not encumbered with installation problems; they are ideal for backup or alternative routing. Recent development of commercial LAN's that use 18-GHz signals for communications may be applicable; however, the radiation of wireless networks may interfere with sensitive experiments at the DSS's.

One problem with satellite communications is the propagation delay associated with transmitting via geostationary satellites. At heights of approximately 35,000 km, the one-way delay is about 0.25 sec, depending on the elevation of the satellite above the horizon. This delay may significantly affect data throughput if the network uses error-correction methods usually associated with terrestrial communications.

Trends show increasing use of very small aperture terminals (VSAT's) for WAN applications to communicate with terminals in unreachable areas and for mobile communications. The primary usefulness of microwave communications in the DSN will be to provide redundant WAN routing paths for critical missions and temporary communications.

c. Optical Fiber. Optical fiber has the bandwidth to enable terabit-per-second transmission rates. Fiber not only has high bandwidth, but it also has the benefits of low noise and signal attenuation (see Fig. 11), with the result that high-rate data can be transmitted over long distances (e.g., 2 km with multimode fiber and over 40 km with single-mode fiber). Fiber has advantages in security, safety, immunity to electromagnetic interference, and reduced weight and size.

While multimode fiber is less expensive, single-mode fiber is the medium of choice for high-speed digital transport, and the number of installations will likely grow. Research is focused on signaling at higher rates. Current signaling records are 350 Gbps. Another research thrust is in wavelength multiplexing; the result is increased utilization of the optical bandwidth by the establishing of multiple communications channels at different wavelengths over a single fiber. The primary impact on the GIS architecture will be higher data rates over LAN's and common-carrier WAN circuits.

d. Common Carriers and ISDN. There are two major trends that will affect the services offered by the common carriers: the trend to convert long-haul links to optical trunks and the offering of digital interfaces directly to the users. The first trend has been spurred by the adoption of the International Telegraph and Telephone Consultative Committee (CCITT) Synchronous Digital Hierarchy (SDH, formerly SONET) standards for optical networks and the second by the increasing availability of Integrated Services Digital Network (ISDN) central office switches. ISDN is an international standard for providing digital circuits directly to the user and will eventually replace current analog circuits for voice, data, and image (i.e., facsimile and video) transmission.

Synchronous Digital Hierarchy. The SDH standard for optical phone company networks will enable the use of a hierarchy of service data rates significantly higher than the current CCITT-standard G.702 hierarchy (Table 7). At the present time, many corporate users, including JPL's Information Systems Division, are using 45-Mbps DS-3 trunks to replace multiple 1.5-Mbps DS-1 circuits. It is expected that OC-1 (51.84 Mbps) will replace many of these circuits when SDH is fully implemented. Note that the SDH hierarchy goes to 2.4 Gbps. Most of the data speeds in this hierarchy are for internal telephone company support networks. Availability of these high-speed trunks for other users will depend on market demand.

Terrestrial and undersea installation of fiber will make these higher rates available between the DSCC's and JPL. Also, replacing satellite links with fiber will eliminate the typical 0.25-sec propagation delay associated with each satellite hop and will significantly reduce bit errors. It is predicted that by the year 2000 all global links associated with the GIS will be fiber.

Table 8 summarizes the major differences between satellite and fiber circuits. Based on industry projections, optical fiber leased-circuit cost for the year 2000 time frame may be as little as 10 percent of today's rates.

ISDN. For lower data rates (< 1.5 Mbps, narrowband ISDN), user service is provided on three types of circuits (Table 9). The D channels are used for common-channel signaling and (optional) packet transmission; the D channel has its own OSI protocol stack. The basic user B channel may carry voice, high-speed fax data, slow-speed video, or computer data; the protocol is at the user's discretion. Three kinds of connections may be set up with B channels: circuit-switched (similar to today's dial-up service), packet-switched (user is connected to a packet switch), and semipermanent (user has the equivalent of today's dedicated circuit).

Current commercial services include basic and primary-rate service (Table 10). These digital services are direct improvements over slower 9.6- to 19.2-kbps links currently used over the GIS, and their implementation will enable a higher volume of data traffic between the DSCC's and JPL.

Future offerings built on optical fiber capability will make higher data rates available in a service called broadband ISDN (B-ISDN). B-ISDN will eventually permit switched services up to 2.4 Gbps. This service will also enable video conferencing and transmission of high-definition television on a switched, on-demand basis. It is also expected to enable a variable amount of bandwidth on demand, so that the user will pay only for the bandwidth required at a given time.

2. Lower Layer Technology. Network technology can be organized into three general classes, depending on the network dimensions. Each class has a need for lower layer protocols. At the lowest level are networks that interconnect computer boards in high-speed real-time applications. At the next level, computers within buildings may be interconnected through LAN's. At the highest level are WAN's that extend over longer distances. WAN's usually depend on common carriers for the links among computers.

a. Processor-Board Buses. Board-level computers may be linked with serial or parallel buses. Simplicity and relatively low speeds can be met with serial copper buses. High-speed links over optical fiber using such technology as FDDI may be common by the year 2000.

Tightly coupled computer systems with very high data rates among the computers are better served with parallel buses. Current buses are 32 bits wide; 64-bit buses are emerging, with future growth expected to 256 bits. Several standard buses are summarized in Table 11.

It is expected that the number of tightly coupled processors will continue to grow to take advantage of parallel-processing techniques and processor miniaturization. Further, 100+ MIPS workstation computers are emerging that will be used in distributed processing applications, such as on-site, front-end processing of spacecraft signals and back-end high-volume imaging data processing. Research efforts will continue to extend bus widths and integrate more functions on a chip.

b. Local Area Networks. LAN's provide loosely coupled communications between computers typically within a 2-km distance. Their numbers have climbed because of the

growth in personal computers and the need of users to share resources and information. Several standard LAN's are summarized in Table 12.

Current data rates range from 4 to 100 Mbps. The FDDI network, which runs at 100 Mbps, is designed for high-speed mainframe-to-mainframe and LAN-backbone applications. It is expected that the speeds of these types of networks will continue to climb and their useful distances will increase with improvements in fiber-optic technology.

The most significant standard is FDDI. It has a 2-km range between nodes. Multimode fiber is specified in the standard to minimize implementation costs; however, as discussed in Section III.D, the use of higher cost single-mode fiber may permit extensions up to 40 km. The FDDI network is ideal for interconnecting other packet-oriented networks, such as Ethernet and token rings. FDDI-II, a significant departure from FDDI, is emerging with unique capabilities. With the ability to integrate circuit-switched traffic with bursty LAN traffic, FDDI-II can concurrently support data, digital voice, and video traffic.

Research and development efforts are focused on higher speed optical LAN's. Some vendors are developing wireless LAN's to operate at radio frequencies and in the infrared spectrum; this effort is not leading to new breakthroughs in speeds, only to increased flexibility in placing workstations within an office.

c. Wide Area Networks. The High-Level Data Link Control (HDLC) protocol is a common ISO-standard link protocol for providing reliable service at the data-link layer. Errors are detected by using a 16-bit cyclic redundancy check (CRC) algorithm. Error recovery may be done in one of two ways:

(1) Go-back-N: When an error is detected in a data frame, the receiver notifies the sender of the faulty frame; the sender retransmits it along with every frame that followed.

(2) Selective retransmission: When an error is detected, the receiver notifies the sender; the sender retransmits only the frame in question.

Most terrestrial networks use the Go-back-N algorithm. However, noise on satellite circuits highly degrades the performance of high-speed satellite links when using this algorithm because of the 0.5-sec round-trip delays; selective retransmission is preferred in such cases.

GOSIP specifies CCITT X.25 as the WAN protocol of choice. It is a three-layer (physical, data-link, and network) packet network standard. Its link-level protocol is link access procedure balanced (LAPB), a version of HDLC but without selective retransmission features. Therefore, it is limited to lower speed satellite links, or optical fiber links.

With X.25, data are packetized into basic 128-byte protocol data units (unless the overall network has been designed for larger packets). At the network layer, X.25 software establishes a path across multiple computers from the sender to the receiver through the use of a connection-request packet, and the data packets flow across this virtual circuit until the connection is terminated.

Frame relay is an emerging protocol that will enable higher throughput over common carrier circuits by reducing the number of acknowledgments, as compared with X.25. Research and development efforts are focused on minimizing delays due to protocol overhead and providing dynamic bandwidth allocation and optimal routing algorithms.

3. Internetwork Communications. When there is connectivity between LAN's and WAN's, additional users are able to share resources, access common databases, and exchange data. Standard internetwork protocols provide routing among the subnetworks and process-to-process communications, and lay the foundation for valuable application services, such as electronic mail, file management and file transfer. Research goals are to make details at this level of communication transparent to users, increase information throughput, and reduce delays with efficient software implementation architectures.

The most popular internetwork protocols today are two Internet protocols: Transmission Control Protocol and Internet Protocol. Together they are capable of providing interprocess communications among applications on widely separate computers and networks. TCP operates on the principle of virtual-circuit connectivity; that is, it establishes a path to the destination process with a connection request packet, transfers the data sequentially and reliably, and after completion of the transfer, terminates the connection.

GOSIP specifies a pair of ISO protocols that perform the same services as TCP/IP: Connectionless Network Protocol (CLNP) and Transport Protocol Class 4 (TP4). They are similar to TCP/IP but are not interoperable because of different addressing schemes and some other minor differences (see Table 13).

TCP/IP and TP4/CLNP protocols themselves are fairly efficient; however, many current implementations are in-

efficient because of how the code is executed and messages are buffered. Intermediate protocol layers constitute an area of continuing research to reduce delays in interprocess communications and increase performance.

4. Application Services. Protocols in the upper layers of the OSI model provide powerful application-oriented services for users of internetworks. They range from standard file transfer and electronic mail protocols to message formats for robot applications. Several are extremely important to the DSN and will be summarized here.

a. File Transfer. The Internet file management protocol is the File Transfer Protocol (FTP). It is a popular protocol for moving files from one host to another across a network. It is based on TCP/IP connections.

OSI provides a more powerful file-oriented protocol, File Transfer, Access, and Management (FTAM), that may be used for distributed database applications, document retrieval and updating (e.g., library information services), and special messaging systems that transfer long text messages.

FTAM is a two-party file transfer protocol; in other words, there is a controller of the file activity (initiator) who directs the action and a responder who responds to the initiator in a passive role. All file transfer and access operations occur between an initiator and responder over connection-oriented paths (virtual circuits, i.e., CLNP/TP4). Three-party file transfer is a future consideration that is expected to be available by the year 2000.

b. Electronic Mail. Electronic mail allows people to exchange text messages over networks. Internet uses the Simple Mail Transfer Protocol (SMTP). OSI specifies CCITT X.400 as its message-handling service (MHS).

MHS is composed of a series of message-transfer agents (intermediate computers) that are responsible for relaying a message from the sender to the receiver. The transfer agent serving the recipient need not be active when the message leaves the originator's transfer agent; the message can be stored at an intermediate location until the recipient becomes operational.

c. Network Management. At the present time, the Internet Simple Network Management Protocol (SNMP) is implemented on all new commercial bridges and routers. It is used to report the performance of network resources to a central monitoring location. The trend is toward implementing this protocol on all network elements, including workstations and personal computers.

The OSI equivalent is called the Common Management Information Protocol (CMIP). Although CMIP is specified in the Government Network Management Profile (GNMP), which is a network-management equivalent to GOSIP, OSI network management services are only in a state of definition and development today. The basic service provides a capability to report events remotely, read attributes of the network hardware, set a characteristic (an attribute that can be set to influence the operation of the network), set a characteristic if and only if the previous value was in a certain specified ranges, and demand actions of the resource.

SNMP and CMIP are used to report the status and performance of network elements (e.g., routers) to a network management system. Typically the network elements generate events that require immediate management attention. The network management system logs in the events, and depending on whether the option is requested, will keep track of their frequency, and notify a person when a threshold has been exceeded. The system will enable the following services: fault management, accounting management, configuration management, performance management, and security management. All these functions are expected to be fully developed by the year 2000.

d. Manufacturing Message Specification. Manufacturing Message Specification (MMS) is an OSI-unique application layer service that has no counterpart among the Internet protocols, and it illustrates the thoroughness of the OSI approach. MMS is a standard messaging service that may be used within application programs for monitor and control of instruments and robots. In a sense, it is the monitor and control equivalent of an electronic mail standard, only it provides a device interface (versus an electronic mail–human interface). MMS also uses the transport services of the OSI protocol stack.

5. Interfaces to Other Networks and GOSIP. The GIS, although internally protected, will pass data through communication servers to several external national networks, which in turn will connect to international, university, and research center networks. These include the NASCOM II, NASA Science Network, and Space Physics Analysis Network (SPAN). Because of international acceptance of the OSI protocols, it is expected that all these networks will interoperate with OSI protocols.

The CCSDS Advanced Orbiting Systems (AOS) architecture (CCSDS 701.0-B-1), which recommends standards for data transmission over special-purpose space data links and their associated interface networks also recommends OSI protocols. The CCSDS provides the following guid-

ance: "Wherever possible and technically appropriate, interfaces with the commercial network environment are provided so that the emerging infrastructure of commercial OSI may be exploited."

E. Human Interfaces

A variety of technologies that provide improved methods for allowing users to interface with computer systems will become available during the next decade. These interface technologies focus on improvement in the amount of information that the user can perceive from a given interface configuration. There are three general categories of interface technologies: graphic visualization, hypermedia, and AI.

1. Graphic Visualization. Graphic visualization may be of potential use in GIS human interfaces in the following areas:

(1) Representing multidimensional data in images on computer graphics displays and in a form that allows people to perceive, amplify, and interpret the data.

(2) Synthesizing data into animated models. Animation time may represent an actual time variable in the data, perhaps scaled, or it may be mapped from an arbitrary variable in the data, presenting novel opportunities for perception and discovery on the part of the user.

(3) Interactively applying experimental transformations and algorithms to explore the effects on the data (e.g., filters, Fourier transforms, color maps, projections, and nonlinear geometric mapping).

(4) Monitoring and debugging computer simulations of physical components and processes through the use of graphical representations of equipment (e.g., antennas, spacecraft, etc.).

(5) Implementing virtual reality, where a computer and peripheral devices create a complete sensory environment for three-dimensional visualization applicable to such interactive activities as design, training, and remote control.

2. Hypermedia. Hypermedia software technology enables a user to retrieve data in various formats in one or more display windows. The formats may include text, graphics, animation, digital audio, and/or video. The data may be retrieved nonlinearly, that is, users may interactively traverse levels of description in a file from high-level overview to low-level details, interactively trace cross-references within a file and retrieve nonlocal references to other files, or interactively connect with off-line databases and application programs.

Hypermedia may be useful in the following DSN areas:

(1) Software development. The use of hypermedia may minimize time that programmers typically spend tracing cross-references in code by automating the tracing of such cross-references. Multimedia comments, such as text, diagrams, audio, and video, may be inserted.

(2) Software engineering. Hypermedia enables such features as an encyclopedia of software components that supports software reuse, tracking of software versions and changes, hardware configuration, etc.

(3) Requirements, design, training, and reference documentation. These documents are voluminous, expansive, and full of cross-references, the kind of complex documents for which hypermedia was originally devised.

(4) Browsing and searching knowledge bases. Hypermedia is the natural technology for information retrieval from AI knowledge bases. The nonlinear structure of hyperdocuments can easily be made to mirror the inherently nonlinear structure of knowledge bases (as compared with the linear structure of relational databases).

3. Artificial Intelligence Support. AI techniques that help translate raw data into knowledge may be applied in the following ways:

(1) Context sensitivity. Interpreting events and actions based on the current context. In a scientific domain, for example, context sensitivity can filter data by using dynamic thresholding.

(2) Abstraction. Using AI techniques to abstract information and present summaries to the user. For scientists, AI can be applied to data sets to identify interesting features, summarize results and indicate areas that warrant additional review by the scientist. For engineering operators, AI can provide techniques for knowing what state the entire system is in and how ongoing activities are expected to affect that state.

IV. Architectures

The GIS architecture described and recommended in this section is intended to meet performance requirements in the 2000–2010 time frame and provide adequate margins; minimize development, implementation, and operational costs; take advantage of technology expected to be

available in the relevant time frame; and allow for a smooth transition from the current architecture.

The overall GIS architecture (Fig. 12) is designed as a single unified system from the front-end area to the end user. Processing activities and network resources (e.g., communication and database servers) are located at the most appropriate points to meet performance requirements, and standard networks are used to move data from one to another. Data buffers are provided at several points in the architecture, in each case insulating upstream parts of the architecture from the effects of failures downstream. In particular, buffering is provided at each DSCC to accommodate WAN outages.

Unbuffered real-time data exist only at the front end between the digital receivers and telemetry processors. Currently it is expected that the first level of buffering will be provided at the point where frames (or CCSDS packets) first exist, for example, at the output of the telemetry or radio science subsystem. Subsequent processing stages are based on client–server relationships. Data flow after the data are packetized may be temporarily buffered before transmission to control data flow. The goal at each stage is to perform a clearly defined transformation that minimizes overall operational costs, including data transport costs.

Multimission level 0 processing (LZP) is done at each DSCC, to the extent possible with locally available data. Final multimission level 0 merge (LZM) occurs at the Central Site (JPL), where it first becomes possible to consolidate data from all DSCC's into a single continuous data stream.

Project-specific processing is done at JPL (or at other locations, as negotiated with non-JPL DSN users). Data to be processed are made available to projects and other users in a uniform manner, using standardized protocols, procedures, and formats.

A monitor and control network complements the processing architecture, and its processors perform several major functions. In particular, a Network Control Facility (NCF) at the Central Site will remotely control the DSCC antennas through appropriate directives and maintain a management information base to enable status monitoring of all subsystems.

An appropriate version of the information base will be maintained at the DSCC's to automatically reconfigure antennas for each mission activity and automatically detect, isolate, and assist recovery from faults. A subset of this information will be available at several locations at the DSCC to provide a human interface when necessary for manual intervention.

The following is a summary of the proposed GIS architecture:

(1) Unified architecture from front-end area to end user.
   (a) Client-server model for each processing stage.
   (b) Unbuffered real-time acquisition only at the front end.
   (c) Standardized interprocess communication.

(2) Open-system standards to achieve interoperability.

(3) Direct delivery of level 0 data to the user from the DSCC's, if desired.

(4) Dedicated telemetry processors for each receiver.

(5) Security against unauthorized access and accidental errors.

(6) Highly automated monitor and control (M&C).
   (a) Centralized control of entire network.
   (b) Automated configuration of support activities.
   (c) Vastly improved human interface.
   (d) Automated decision support systems for network resource management, performance analysis, fault diagnosis, and contingency management.
   (e) Embedded intelligent computer-aided training systems.

The Ground Information System is logically organized into four reasonably distinct high-level functions: processing, monitor and control, data transport, and software. For ease of analysis, each of these functional areas is considered separately in the following sections.

A. Processing

In the early stages of the study, several classes of processing architectures were considered (see Table 14). Each of these architectural candidates (in its pure form) was rejected with the better features of each subsumed into the recommended architecture. The DSCC aspect of this architecture and the Central Site aspect are detailed in this section. The architecture includes elements of these

candidates, organized to minimize implementation and operational costs, to take advantage of technology expected to be available in the relevant time frame, and to allow for a smooth transition from the current architecture.

1. DSCC Site. The DSCC processing architecture performs initial level 0 processing of telemetry, and initial processing of radio science, VLBI, ranging and tracking, and command data. The estimated data rates for each of these subsystems are summarized in Table 15. In the DSCC architecture (Fig. 13), downconverted signals are routed to the Block V receiver, VLBI subsystem, or special acquisition equipment through an analog distribution switch.

In some cases processing is initiated by the Block V receiver and data are passed directly to a dedicated subsystem (telemetry and ranging). In other cases, such as tracking and radio science, data are routed over a network to an available subsystem. The data paths depend on the real-time data rate, with functions with higher rates requiring dedicated processors. VLBI processing is a dedicated string from the point of signal downconversion and digitization to the stage of formatting the data for transmission to JPL for correlation with data from another antenna.

Each subsystem transmits its formatted packets to a network communications server over the DSCC backbone LAN. Reliable OSI protocols are utilized for communication to JPL. The M&C network, not explicitly illustrated in Fig. 14, complements the processing architecture and is the means used to communicate configuration data, status, and commands among the subsystems.

The command subsystem stores commands received from JPL. The subsystem modulates a subcarrier to form a baseband signal that is transmitted to an exciter under the control of the M&C subsystem.

The Ranging Subsystem, closely coupled to the Block V receiver, receives predict data from JPL and uplinks appropriate ranging signals through the exciter. Test data from JPL are sent to the test subsystem and it, in turn, modulates a signal to the exciter for telemetry testing.

a. Telemetry Processing. The Block V receiver and its successors will feed data to the telemetry process in the proposed time frame. For analysis, it is assumed that there will be one Block V receiver at each Complex, with the potential for at least 22 channel processors supporting seven antennas, and by the year 2010, at least 32 for 10 antennas. Two combiners are expected by 2000 and three in 2010, with dedicated channel processors.

The goal of the architecture is to pass telemetry data directly from a Block V channel processor to the telemetry processor with virtually no delays (real-time), process the data, and deliver it in sequence and with no detectable errors to JPL (near-real-time). Short delays may be required for buffering in the communication links to detect and correct packet errors downstream.

There are three fundamental architectures that are considered for the channel processor–telemetry processing operations: dedicated, switched, and distributed. All architectures may operate in a fully automated mode.

In the dedicated processor architecture, Fig. 14(a), a separate set of electronics is attached to each Block V channel processor. Two or more telemetry processors may be dedicated for increased availability. Processors may be reconfigured for different missions by downloading software. High availability is achieved through processor redundancy and component redundancy within each processor.

The switched architecture, Fig. 14(b), interconnects the receivers and processors via a matrix switch, similar to the way that it is done today. While mechanical switches today suffer from low reliability, this proposed architecture may be envisioned with highly reliable switches (e.g., asynchronous transfer mode switches) in the proposed time frame. The switched architecture provides flexibility in configuration management. High availability is achieved by telemetry processor redundancy.

The distributed architecture, Fig. 14(c), interconnects the receivers and processors via a common multiple access network (e.g., a LAN), and provides connectivity for passing data from any receiver to any processor on the network. This approach envisions extensible use of general-purpose computers and commercial network technology integrated into a DSCC-distributed processing system. The distributed network provides extra flexibility in configuration management and enhances options for load balancing among machines. As with the switched architecture, high availability is achieved through computer platform redundancy.

There may be occasions for exceptional support for which dedicated high-speed acquisition equipment may be used, and the data stored on portable media (e.g., tapes or disks) for off-line delivery to the users. This capability was illustrated earlier in Fig. 13.

Ancillary data, such as received power levels and system temperature, are required for real-time processing. Some of these data are generated by the Block V receiver, transmitted to a network file server and distributed to each telemetry processor. Also, software configuration data (e.g., the type of decoding, etc.) must be transferred to each telemetry processor by Monitor and Control during configuration of a link.

Output from the telemetry processor is transmitted over the DSCC backbone LAN to the communication server. The communication server sends the data to the Central Site for merging, deletion of duplicate data, and completion of multimission level-zero processing. The data are then available to users for further processing over mission-unique networks.

In addition, the DSN antennas may be arrayed with remote antennas (e.g., Goldstone, California, antennas with the Very Large Array [VLA], in Socorro, New Mexico). Soft-symbol combining must be performed within the DSCC combiners. Other occasions may require GIS to transmit these same symbols to other partners in an array; it is assumed that Block V receiver-to-array partner communications will be specially arranged, and there will be little impact on the information system.

Information rates above 10 Mbps may be expected for Earth orbiters and lunar missions, and these usually require a convolutional code rate of 1/2. Deep space missions with rates less than 10 Mbps usually require code rates of 1/6. Thus, for orbiters and lunar missions, one megabit per second of information translates into about 18 Mbps of Block V channel processor output (assuming Reed-Solomon [255,223] code), and with deep space missions, one megabit per second translates into approximately 55 Mbps (see Section II.A). Table 16 translates the proposed spacecraft telemetry profile into a telemetry processing rate profile. These rates have a major impact on the architecture.

Dedicated Processing Evaluation. The dedicated architecture (Fig. 15) features Block V channel processors and combiners with dedicated telemetry processors. It is assumed that communications between the channel processor and the telemetry processor are over a dedicated backplane bus, although a dedicated LAN may be used. In contrast to the distributed architecture, the data path is dedicated—with no sharing. In contrast to the switched architecture, the path is permanent.

Specially designed protocols are assumed for data transfer. The optimal performance is summarized in Table 17. It can be seen that the potential for growth is excellent—even the 1980s bus technology meets the conservative mission-support forecast. The year 2000 bus technology meets the aggressive forecast.

The advantages of this architecture include:

(1) The architecture meets all year 2010 requirements, up to 450-Mbps backup support.

(2) Potential points of failure are the Block V channel processor, telemetry processor, and active interprocessor bus.

(3) Dual telemetry processors within each string improve fault tolerance.

Major risks include:

(1) Computers (I/O software) cannot keep up with bus speed.

(2) Custom-designed interprocessor communication software may lead to higher life-cycle software costs.

(3) Bus management (fault detection, recovery, etc.) will be nonstandard software, and may lead to high life-cycle costs.

(4) The cost of many telemetry processors may be high.

Switched Processing Evaluation. A switched telemetry architecture is illustrated in Fig. 16. The architecture features multiple Block V channel processors and combiners that feed telemetry data to one of many switch-selectable telemetry processors.

The switch is computer-controlled and provides a temporary dedicated path on demand. The highest speed of the path is the serial I/O speed of the channel processor/telemetry processor link. Currently, the best technology is about 100–130 Mbps.

A major issue is the generally poor reliability of mechanical switch technology. It is assumed that more reliable technology, such as the asynchronous transfer mode switches, will evolve by the year 2000 and will provide high reliability.

The advantages of the switched architecture include:

(1) Telemetry processors can be easily changed as mission-unique requirements are added.

(2) A minimum complement of telemetry processors reduces costs.

(3) Redundant telemetry processors increase availability.

The major risks include:

(1) Forecasted telemetry rates may exceed I/O capacity.

(2) The switches may be expensive.

(3) The switch is a single point of failure.

Distributed Processing Evaluation. A distributed telemetry architecture is illustrated in Fig. 17. The architecture features multiple Block V channel processors and combiners, each feeding telemetry data to one of many software-selectable telemetry processors.

An ultra-high-speed network provides reliable data transfer among the receivers and telemetry processors. The network is shared—it provides a data path for all acquisitions that are under way at any moment. The optimal LAN performance is summarized in Table 18, with an indication of the instantaneous aggregate information rate. It can be seen that the best current technology (100 Mbps) provides an aggregate capability of 1.8 Mbps; a 1-Gbps LAN would provide aggregate capability of 18 Mbps.

The advantages of the distributed architecture include:

(1) It may enable open interprocessor communications with commercial off-the-shelf (COTS) hardware and software interfaces.

(2) Telemetry processors can be easily changed as mission-unique requirements are added.

(3) FDDI, a dual counter-rotating ring, is available in 1990s technology that automatically recovers from faults.

The major risks include:

(1) An aggregate telemetry stream may exceed LAN capacity—a 100-Mbps LAN can only support an aggregate of 1.8 Mbps of information.

(2) A one-Gbps standard LAN may not be available.

(3) Computers (I/O software) cannot cannot communicate at the LAN speed. This is a 1990s area of research.

(4) This configuration has low growth potential.

Evaluation. The alternative telemetry architectures were evaluated according to the criteria described in Section I.D, and the results are summarized in Table 19.

Based on performance, growth potential, and high availability, a dedicated telemetry architecture is recommended. Additional study may show that adding redundant telemetry processors to each string will significantly improve availability. Primary concerns will be the cost of custom interprocess communication software and the cost of many processors.

A hybrid of the switched and dedicated architectures was considered. The hybrid involves an addressable bus that permits each receiver or combiner to talk to any telemetry processor over its own dedicated bus. The hybrid is based on the ability to provide a backplane bus from any channel processor to any telemetry processor under program control. Further study will determine whether the additional cost of this technique justifies the improvement in availability. A cost study should be made before a final commitment is given to the dedicated architecture.

b. Additional DSCC Processes.

Tracking. Tracking is generally a low-speed process whereby Doppler and ranging packets are sent from a ranging processor associated with the Block V receiver to the tracking processor. Its primary function is to take these packets and format them into a suitable product for navigation processing by using ancillary data and predict data from the M&C database. After formatting, the packets are transmitted to users, such as the spacecraft navigation team, through the communication server.

Command. Command processing is another low-rate process that takes commands sent to the DSCC from JPL and sends command modulation signals to the exciter and transmitter for an appropriate antenna. The command subsystem also sends a verification back to the mission control team at JPL.

Radio Science. Radio science packets are generated by the Block V receiver and transmitted to a radio science subsystem for formation and transmission to the Central Site. The potential rate from each receiver is 1.2 Mbps, and several experiments may acquire data simultaneously. After formatting, the packets are transmitted to the radio science team through the DSCC communication server.

VLBI. VLBI ranging is a high-rate process (approximately 130 Mbps) that typically digitizes and stores X- and S-band signals from radio sources or special tones from a spacecraft. The data are temporarily stored on a high-speed, high-volume storage device and later played back at a slower rate (e.g., 1.5 Mbps) to JPL for correlation with data simultaneously acquired at another DSCC.

The results are passed to flight projects for navigation processing. Approximately 20–30 min of 130-Mbps data are acquired at one time.

If required, quick-look data may be acquired during the first 15–20 sec and sent to the radio science team for real-time processing. If anomalies are discovered in the data, adjustments to the instruments may be performed with real-time commands.

In summary, the overall processing architecture has the following features:

(1) The architecture meets performance requirements in the forecasted time frame and provides an adequate margin. Each Block V channel processor is provided with at least two dedicated telemetry processors.

(2) DSCC on-site processing is performed using distributed open systems.

(3) Communications are based on open-systems standards.

(4) Final-product processing traffic flows to the DSCC backbone LAN and to the DSCC communication server. Subsystem-unique requirements (e.g., VLBI) are accommodated with intermediate high-performance processing and storage devices.

(5) Certain exceptional types of processing are acquired in real-time, but delivered off-line to JPL or direct to users (e.g., orbiting VLBI) via portable media.

2. Central Site (JPL). The Central Site performs two major GIS functions and provides an open systems interface to the multimission and mission-unique subnetworks (Fig. 18). A key characteristic is the high-speed DSN backbone LAN that permits client-server architectures.

The first GIS function is completion of level-zero processing, where duplicate data records from the DSCC's are deleted and the data steams are merged and prepared for delivery to higher level processing. This relatively high-rate function may involve distributed processing over a dedicated LAN (level-zero merge subnetwork). The output is delivered in near-real-time to multimission networks. Temporary buffering will be available for network flow control; i.e., to keep the level-zero merge processor from overrunning the multimission telemetry (SFOC) processor.

The second major Central Site function is monitor and control of the DSN. This is accomplished over a subnetwork performing network and multilink control. Network control includes the highest level functions in the control hierarchy, such as resource management, scheduling, and data network management. Operators will control the entire DSN through multilink control workstations (see Section IV.B).

A key architecture requirement is a large operations database for schedules and resource management, accessible by each control subsystem. The database server will also make DSCC equipment status data available to the end users for later analysis.

The DSN backbone LAN interfaces through a local communications server (router) to the SFOC backbone and related local networks. This interface is designed to provide network security. Because of the extensive use of open standards, subsystems and users will use simple and relatively inexpensive communications software to exchange data across this interface. Through the use of standard protocols, data residing in DSCC or Central Site databases, such as resource status at the time of acquisition, may be retrieved by the users on demand, assuming that the network administrator has authorized such access.

B. Monitor and Control

1. Centralized Control. In the proposed M&C architecture (Fig. 19), the DSN is controlled from a central location, the NCF. The control structure is hierarchical: network control (the highest level), multilink control, link control, and subsystem control.

The two higher levels reside at the NCF and are integrated human–computer operations. The lower two levels, link and subsystem control, are located at the individual DSCC's and are automated. In addition, there are site-resource controllers at each of the DSCC's that autonomously monitor the health and status of station resources when they are not assigned to a link. Operations personnel are located primarily at the NCF, with only maintenance and backup operations personnel located at each of the DSCC's.

A generic model for all control layers is presented in Fig. 20. (The link controller is pictured as an example.) The controller receives assignments and the necessary support data from its superior controller, delegates control to its subordinate elements, receives monitor information (health, status, configuration, performance, and events) from each of the elements under its control, and consolidates and reports monitor information to its superior. The controllers themselves perform a variety of functions, including anomaly detection, isolation, diagnosis, and correction. If an anomaly cannot be handled at the current

level of control, the controller requests support from its superior. If more information is needed, the controller requests additional data from its subordinates.

Subsystem controllers interface with the subassemblies and devices that constitute the actual subsystem. They essentially perform the same basic functions that current subsystem controllers perform, but with a greater ability to execute control sequences. Rather than the explicit, low-level directives used to interface with the subsystems in the current DSN, control messages will be at higher levels (e.g., "Configure for Voyager 2 one-way"). The subsystems will have the intelligence to interpret these higher level directives and perform the necessary actions. Any problems encountered in executing a higher level function will be reported up to the next level of control (the link controller if the equipment is assigned to a link, otherwise the site-resource controller).

Subsystem controllers will have extensive built-in test equipment and self-monitoring functions that will assess the health and status of the subsystems internally and alert the higher level M&C systems to any problem.

The primary functional unit for the DSN is a link—the series of equipment necessary to establish and maintain communications with a spacecraft or to meet a scientific objective. In current DSN operations, setting up a link to support a given operation (referred to as precalibration) is a time-consuming manual process. In the proposed architecture, not only will this task be automated, but the monitoring of the link during the track and the performance of any postcalibration activity will also be automated.

The computer-based link controller will initiate, monitor, and control all the actions necessary to accomplish an activity. The link controller will activate the given subsystem controllers, pass down any required support data, monitor the progress, and respond to any problems that occur. Overall link health, status, and performance will be reported to the next higher level of control.

The functions currently performed by link monitor and control (LMC) operators at each of the DSCC's will be automated or will migrate to an NCF multilink controller (MLC). Multilink control is an integrated human–computer function. With the addition of automated background monitoring capability, the MLC operators will now be able to control and monitor the health, status, and performance of several links. There are three modes in which multilink controllers will operate: coordinated, consolidated, and combination.

Coordinated activities are those which require that two or more DSS's operate together to support a customer request. Examples of coordinated activities are antenna arraying and VLBI. During coordinated activities, the multilink operators must ensure that the DSS's and associated equipment are working together to provide the requested service. In some cases, such as VLBI, this requires coordinating antennas at different DSCC's. In a consolidated mode, the multilink controller is responsible for several links that are operating independently. The controller can work with several different types of antennas performing several different types of services. In the combined mode, the multilink controller supports both coordinated and consolidated control.

Site-resource controllers are automated processes that monitor and control equipment that is not operating in a link. In addition, the site-resource controllers are responsible for receiving and storing any support data that is to be maintained on site, scheduling site-specific resources, and resolving any conflict that may arise concerning the disposition of site resources.

The NCF, located at JPL, is the primary control point for the entire DSN. It provides the interface with the DSN customers. The NCF schedules DSN resources, processes customer requests and support data into a format usable by the DSN, and resolves any conflicts that cannot be resolved at a lower level (primarily those which require interaction with DSN customers, or which involve DSN resources from several DSCC's). The NCF network controller subsystem initiates all DSN operations and monitors the health, status, and performance of the DSN. The network controller delegates more specific monitor and control functions to the lower level controllers. The functions of the NCF are described in more detail in the operations concept (see Section IV.B).

2. Automation. The proposed architecture is highly automated. At each level in the control hierarchy, routine and contingency operations are automated, while functions requiring human decision making are supported by automated decision aids. User interfaces with embedded help and training facilities enable human operators to perform complex functions efficiently and effectively despite varying levels of experience and training. The proposed architecture incorporates automation techniques to support a variety of functions as detailed in the following sections.

DSN resources are currently controlled by text strings (directives) manually entered at the DSCC LMC consoles. The status is monitored by operators who interpret the various responses that are provided by the subsystems.

In positive-control systems, all control actions have corresponding explicit verifications that they have been received and processed. A major problem with the existing M&C architecture is that it does not provide positive control, that is, operators must interpret a variety of disassociated information in order to determine the true response of the link equipment.

In the proposed architecture, control functions for routine and specific contingency operations are automated. The higher level controller only needs to give the command, and the lower level controller will execute the appropriate control sequences to support the requested activity. The higher level controller coordinates actions among lower level controllers and has immediate feedback on the status, configuration, health, and performance of the control activities and associated equipment. In the proposed architecture, automated monitoring of health, status, performance, and configuration, with fast and efficient handling of exceptions, will support situational awareness at all levels of control.

Situational awareness means that the controllers (human or computer) have accurate knowledge of the health, status, configuration, and performance of the equipment and processes that are under their control. Situational awareness is especially important in the multilink operations modes because controllers need a clear picture of ongoing operations in order to assess the situation and deal with unexpected events. Situational awareness depends heavily on background monitoring capability, accurate and timely reporting from lower level controllers, and a user interface that ensures that the most important information is brought to the attention of the controller quickly and reliably.

The architecture incorporates embedded training and help facilities that enable human operators to learn new skills, and improve and maintain existing ones. Operators are able to use the help facilities to aid in performing tasks. Help facilities include instructions on how to perform certain activities, how to access and interpret support data, and how to retrieve desired information. The help facilities will also assist operators at a higher level of control to interact with lower level controllers.

The training facilities are accessible from the operations consoles. Training consists of two types: tutoring on declarative knowledge (such as the description of a subsystem, or discussion of the scientific principles behind VLBI) and procedural knowledge (such as how to perform a track). In the procedural training mode, the operator is able to run scenarios that simulate the procedure under both optimal and anomalous conditions. The operator is able to activate a training session while performing actual operations and is alerted to any condition that requires human intervention.

The proposed architecture controls access to the DSN with tight security and safety features. In addition to the features provided by network security applications that prevent tampering with communications, an active security feature will be incorporated into monitor and control. Users (operators and experimenters who want to control DSN resources) will be modeled in a "client–server" relationship. Privileges to perform certain control functions will be restricted to only those users who are authorized. In addition, special security features will ensure that training activities are isolated from actual operations.

3. Operations Concept. The following paragraphs describe an operations scenario in terms of the control flow from the NCF to the DSCC's, and the monitor flow from the DSCC's to the NCF. This concept is summarized in Fig. 21.

a. Control Flow: Nominal Operations. There are two types of control flow: directions on what to do and authorization to do it. Standard operating procedures will be in place at the DSCC's and the NCF, and schedules will be transmitted in advance to the appropriate resource controllers. The authorization to perform operations will generally be given in real time. The following example describes the control flow for routine operations under nominal circumstances.

All DSN activities originate with a customer request for support. The NCF is responsible for coordinating these requests and iterating them with the various customers to develop a workable schedule. Scheduling is mainly the process of assigning an antenna (or group of antennas) to support a customer request.

Once the NCF has an acceptable schedule, the network controller assigns support activities to the multilink controllers. The multilink controllers, in turn, request the necessary link controllers and link equipment from the DSCC resource manager. The link controllers are configured to support the requested activities and the necessary support data are routed to them. The link controllers then perform an analogous function by configuring the subsystems assigned to their links. For example: The NCF may output a schedule of support activities that covers a 24-hr window. The network controller uses this schedule to create assignments for each of the MLC's. The assignments are distributed and authorization to execute the schedule is given.

MLC operations based on the NCF schedule are shown in detail in Fig. 22. In the figure, multilink controller A (MLC-A) has three activities to support: (1) a Voyager one-way track on DSS 15, (2) a Pioneer 10 ranging on DSS 43, and (3) a Galileo VLBI delta-differential one-way range (DDOR) on DSS's 14 and 63. MLC-A assigns four link controllers to configure the links for each of the activities. It also loads the appropriate procedures for performing its monitor functions and sets up the displays necessary to provide information to the human operator. For this particular set of assignments, MLC-A is working in a combined mode: It must coordinate the two DSS's assigned to the DDOR while consolidating control over the other two tracks.

The link controller (Fig. 23) assigned to the Voyager track loads the appropriate procedures from its procedure library and initializes the equipment assigned to its link. When it receives the authorization from MLC-A, it initiates the precalibration sequence in the subsystems. The link controller coordinates all the subsystem actions, based on the timeline associated with the particular track and the detailed timing information provided in the support data.

Each subsystem is initialized and, under the direction of the link controller, executes a series of steps necessary to configure for the Voyager track. The necessary procedures are in the subsystems' local memory. The link controller only needs to send a high-level instruction to the subsystem.

b. Monitor Flow: Nominal Operations. There are two fundamental types of monitor data: The first supports positive control by providing immediate feedback as to the results of control actions; the second provides health, status, and performance information regardless of any control actions. Therefore, monitoring has two components: knowing the status of ongoing operations and knowing the status of the resources. While control is a top-down function, monitoring is a bottom-up function. Information, in general, is sent from a lower level to its superior where it is interpreted, consolidated, summarized, and sent up the chain to the next level. The following example illustrates the flow of monitor data during normal operations continuing from the control scenario presented above.

A subsystem, e.g., the receiver, has completed a portion of its precalibration activities for the Voyager track. It sends a message to the link controller stating that it has completed the procedure and the resulting configuration, health, and status of the equipment is as desired.

The link controller receives this message, consolidates it with similar messages from other subsystems and determines that the precalibration sequence has been completed. The link controller then sends a message to the MLC-A reporting that precalibration has been completed and specifies the health, status, and configuration of the Voyager link as a whole.

MLC-A uses this information to create a summary health and status report for the network controller, which consolidates all the reports from the various MLC's and determines the overall health, status, configuration, and performance of the DSN as a whole.

c. Operations Under Anomalous Conditions. If a problem occurs during the course of an operational activity, it is handled at the lowest control level possible. For example, if a subsystem experiences a failed module, then the subsystem controller detects, isolates, diagnoses, and repairs or replaces the module (i.e., switches to a redundant unit). It then reports this event and subsequent change in configuration to the link controller.

If the subsystem cannot recover, it informs the link controller that the subsystem is not operating nominally. The link controller determines whether to proceed with the subnominal equipment or to replace the subsystem in the link, and informs MLC-A.

If the link controller decides to reconfigure the link, it requests resources from the site-resource controller. If they are available, then the change is made and reported to the MLC-A. If not, then the link controller requests a change from MLC-A.

MLC-A receives the status report from the link controller and notifies the network controller, as appropriate. If the link controller was unable to resolve the problem, MLC-A assesses the situation and determines whether to assign a new DSS to support the task, request that resources be pulled from some other task, or cancel the track. In the event that the decision is sensitive, MLC-A provides the necessary information to the network controller, which has ultimate decision-making authority over the DSN resources.

4. Backup Operations. It is important to provide for backup operations capability in the event of a natural disaster or other emergency. A complete failure-modes analysis was not performed; however, several types of failures were evaluated to identify how the proposed architecture could address such problems. The proposed architecture

makes use of redundancy, interoperability, and additional operations procedures for backup operations. To understand the nature of backup operations, it is important to first recognize some of the important features of the proposed architecture and some characteristics of operations:

(1) The link and multilink controllers use the same basic architecture. The differences are in the databases used to perform their activities; therefore, the functionality of the controller can be changed by loading a new database.

(2) The communications at the DSCC's will be very reliable. Redundant LAN's will ensure continuous communications at the DSCC's.

(3) Storage is relatively inexpensive. Support data and procedural information can be duplicated and archived at each DSCC.

(4) The amount of control information that must be sent in real time is minimal. In the event of an emergency, certain information can be sent over telephone lines.

(5) The DSN (and the flight projects it supports) use established procedures which are known a priori. For most types of operations, the basics are defined well in advance, although specific parameters or sequences might change at the last minute.

The three primary types of failures—loss of a resource, loss of communications, and loss of the NCF—are addressed next. In the event that one of the controllers (network, multilink, link, or subsystem) fails, the DSN will rely on redundant units to replace the failed unit. The major issue then becomes one of graceful transition during ongoing operations. There are a variety of strategies that could be employed, including hot backups (redundant, active units performing the same function), warm backups (equipment on-line and running, but not necessarily performing the same function), and cold spares with context switching (spare equipment that can be brought up but with some level of time delay). An evaluation of the different redundancy management techniques is beyond the scope of this article. However, it is possible to envision an overall DSN strategy that uses a variety of these techniques based on the priority of the activity being supported (routine operations versus encounter).

A centralized architecture is extremely sensitive to WAN communication failures. If a communication problem cuts communications between the DSCC's and the NCF, the DSCC's will not be able to complete a command link with the spacecraft or feed back mission data in real time, and will lose contact with the MLC's. The MLC's will be unable to coordinate activities, including those involving more than one DSCC, and the DSN will not be able to assess overall network performance and status.

In the event of a communications loss, the DSCC's will record incoming mission data for later playback, set up a link controller at each station to perform MLC functions for that station, call in backup operations personnel as required, and use telephone lines for voice and data communications with the NCF. Operations that have to be coordinated between more than one DSCC will be conducted using voice communications (if possible). In the event that no communications exist between the DSCC's and the NCF, and if voice links cannot be set up between DSCC's, then each of the DSCC's will operate individually until communications can be reestablished.

The greatest single threat to the coordinated operations of the DSN is loss of the NCF (a prime scenario involves a large earthquake that affects the Los Angeles basin). Loss of the NCF will result in the loss of the primary interface between the DSN and its various customers.

The proposed architecture supports near-real-time changes to operations due to customer requests. Therefore, in addition to precautions established to ensure operations through a major wide-area communications failure, special precautions must be taken to ensure that customers can still be supported. To preserve this capability, a limited-function backup NCF is proposed. The backup NCF will provide reactive scheduling capability and routing of support data. The backup NCF will also provide NASCOM connectivity in order to continue support to end users not affected by the same event that affected the NCF.

Each DSCC will have archives of spacecraft command sequences, adequate predicts, and other support data generation capability. As long as a voice link can be established between the DSCC's and the backup NCF, uplink support can continue if the necessary sequences are present in the DSCC archives.

C. Data Transport

At the DSCC, the GIS architecture depends on two subnetworks: the DSCC backbone LAN and the M&C LAN (Fig. 24). The backbone LAN is a key architectural feature; it provides connectivity between processes anywhere in the Complex.

The Central Site data transport architecture depends simply on a DSN backbone LAN. The data transport architecture features:

(1) Total interconnection of all processing and monitor and control resources.

(2) Standard protocols among networks.

(3) Optical-fiber medium for the backbone LAN's and a high-speed global WAN network.

(4) Dual, diverse circuits for the global WAN network.

(5) A standard network management protocol to enable fault management, configuration management, performance management, and security of the network resources.

1. DSCC Backbone LAN. The DSCC backbone LAN provides connectivity among all the subsystems and network resources (see Fig. 13). At the DSCC, these include such devices as subsystems (e.g., telemetry, tracking, etc.), communication servers to the global WAN, an M&C LAN interface, test equipment, and storage devices.

A high-speed (i.e., 100-Mbps) LAN will enable the backbone to carry all forecasted traffic, with the possible exception of a continuous stream of VLBI quick-look data (the network should be capable of handling short 15-sec bursts). A subsystem buffer is proposed for continuous VLBI data to avoid congestion. A 1-Gbps LAN will remove this constraint for the 2010 time frame. Based on forecasted LAN technology, FDDI is the best candidate for a year 2000 backbone. It has the following desirable features:

(1) 100-Mbps transmission.

(2) Fiber medium, to enable later growth to higher data rates and ensure low electromagnetic emissions in the vicinity of sensitive station equipment.

(3) Dual counter-rotating ring architecture for fault tolerance.

2. Monitor and Control LAN. The M&C LAN carries the following types of data (all generally at low rates): static parameters, such as directives to subsystems; dynamic parameters to processes, including received power, system temperature, weather, power meter, etc.; the health of processes, networks, and reports to the operator workstation; and fault-recovery directives to processes.

The forecasted LAN technology is not much different from what is available today, such as Ethernet and token-passing bus networks that use a coaxial cable medium. A 10-Mbps transmission rate will enable the LAN to carry all data in the forecasted time frame. The major improvements will be in the wide availability of application layer messaging services.

A likely candidate for the M&C LAN is MAP, a modern OSI-based network architecture designed for manufacturing automation. It incorporates MMS in the application layer (Section III.D). Other features include planned GOSIP 3.0 compliance, thus meeting open-systems interconnection requirements, and token-passing bus topology, with predictable performance as nodes are added, and a self-healing token-passing algorithm.

3. Global WAN. The Global WAN provides connectivity among the DSCC's, JPL, and potential international cross-support users. The flow across this network is lopsided—the flow of data inward from the Complexes to JPL is about three orders of magnitude greater than the outward flow.

Data from the DSCC's to JPL include the following:

(1) Level-zero telemetry data from each DSCC to the level-zero merge function, with the volume and rate depending on number of current missions being supported; e.g., 1–3 Mbps (conservative), 1–10 Mbps (moderate), 1–100 Mbps (aggressive).

(2) VLBI data, including both a short burst of quick-look data (130 Mbps) and slower playback data (1.5 Mbps).

(3) Tracking data derived from Doppler and ranging data (<64 kbps).

(4) Monitor and control data (<64 kbps).

Data from the Central Site to the DSCC's include the following:

(1) Spacecraft commands (<64 kbps).

(2) Monitor and control data (<64 kbps).

(a) High-level command language from multilink controllers to link controllers.

(b) Predicts and other directives for remote control.

(c) Database updates for hierarchical planning data.

(d) Fault-recovery commands.

(e) Test instructions.

Based on performance and bit error rates, optical fiber is an ideal medium for long-haul circuits; however, if disabled it may take a long time to repair. It is recommended that all global WAN circuits be dual, diverse circuits, with the diversity coming from either use of another medium (i.e., a satellite) or another physical fiber route from the DSCC to JPL. These two architectures are compared in

Table 20; it is concluded that a dual, diverse link provides significantly higher availability—an essential feature of the overall GIS architecture.

The following guidelines are suggested:

(1) The communication server should support at least two diverse DSCC–Central Site circuits.

(2) Circuit transmission rates should be negotiated on an as-needed basis, with opportunities to increase speeds on demand for certain missions.

(3) Global WAN protocols should provide intermediate and upper layer OSI services.

(4) Circuits must be full duplex to enable a path for acknowledgments.

(5) Network management software should manage the status of international networks and recover automatically.

4. DSN Backbone LAN. The DSN backbone LAN at JPL provides connectivity among the level-zero merge network, NCF, user networks and network resources. It is assumed that users will interface with the DSN through a common router and use common OSI services, such as X.400 electronic mail and FTAM distributed database capabilities.

A 100-Mbps backbone LAN will be adequate to carry expected 1- to 10-Mbps telemetry traffic. A DSCC buffer is proposed for VLBI data to avoid congestion. A 1-Gbps LAN will remove this constraint and will enable the backbone to carry 1- to 100-Mbps telemetry traffic. Based on forecasted LAN technology, FDDI is the best candidate for a year 2000 backbone and has the desirable features described in Section IV.C.

5. Internetwork Architecture. A typical network path from JPL to a DSCC may include the SFOC LAN, DSN LAN, global WAN, DSCC backbone, and M&C LAN. In the DSCC-to-JPL path there may be just as many networks. As discussed in Section III.C, intermediate-level protocols are required to route the packets and ensure reliable delivery to the destination.

Because high bit errors increase the frequency of the acknowledgment process for reliable stream-oriented protocols, long-haul circuit error rates and the data-link layer recovery algorithm significantly affect internetwork throughput. Satellite propagation delays also affect throughput. For example, using TCP over a satellite link with a typical data rate of 1.5 Mbps, 10^{-7} BER (coded), results in a maximum throughput of only about 87 percent; with a 10^{-6} BER, it drops to 40 percent. An immediate solution is to use a data link protocol with the selective retransmission option (Section III.D). When this is done, TCP or TP4 will deliver significantly higher throughput rates.

The emergence of fiber circuits will significantly improve throughput because 10^{-9} error rates are possible and propagation delays from the DSCC's to the Central Site are only about 0.04 sec. The maximum throughput is much closer to 100 percent.

6. Network Management. Each subsystem and communication resource connected to the DSCC and Central Site backbone LAN's will have a network interface capable of being remotely monitored and controlled. The ISO-standard protocol for network management, CMIP, is a key protocol within the proposed GNMP. CMIP will enable the use of commercial off-the-shelf software to manage network faults, configuration, performance, and security.

(1) Fault management alerts the network administrator at a DSCC or JPL when a fault is detected. It should provide fault isolation; it should accept and act upon error detection notifications, trace faults, and correct faults arising from abnormal operation.

(2) Configuration management enables the network administrator to exercise control over the configuration of the DSCC LAN, Central Site LAN, or global WAN. It will enable an administrator to close down nodes should a fault occur or workloads change.

(3) Performance management enables evaluation of the behavior of network and layer-entity resources (e.g., bridges and routers) and measures the effectiveness of communication activities. It can also adjust operating characteristics and generate network utilization reports by monitoring a resource's performance. An extension of this capability will enable monitoring subsystem (e.g., telemetry or tracking) performance.

(4) Security management provides for protection of network resources by including authorization facilities, access controls, encryption when required, authentication, maintenance, and examination of security logs.

D. Software

It is possible to list some of the desired characteristics of a software architecture. These features are necessary to ensure that the design complies with the stated goals of openness, interoperability, and cost-effectiveness. A precise specification of software architecture requires a larger scale design effort than is undertaken in this study.

(1) Open. The GIS should comply with national and international standards (when available) for programming languages, operating system interfaces, application environments, user interfaces, etc. This compliance has benefits in many areas, including portability and interoperability.

(2) Distributed. More computing power per dollar is available in highly distributed configurations than centralized. Furthermore, there are robustness benefits to distribution.

(3) Parallel. The design must minimize dead time in which a processor fails to service requests because it is blocking on some other incomplete action. The threads model provides a clean method for implementing parallelism.

(4) Configurable. Insofar as possible, multiple configurations for missions, antennas, etc., should be designed to be selected at run time. The name "table-driven" is often attached to this feature, but it can be considerably more general.

(5) Operable. Operators should be able to modify their interfaces, as appropriate. This also implies that the design should not be too tightly coupled with any particular user interface.

(6) Regular. A standard interface syntax, particularly for monitor and control, should be obeyed by all subsystems. This encourages the development of general-purpose tools and software reuse.

(7) Automated. Operators should not be burdened with routine commands.

(8) Secure. System security must be incorporated in the design from the beginning to avoid inconvenience and operator frustration.

In this section, a simple end-to-end flow of data is used to illustrate the application of client-servers, a key element of the software architecture. The flow is illustrated in Fig. 25. It may be useful to review the discussion in Section III.C. Responding to a stored sequence of events, the LMC subsystem begins configuration for a tracking pass. It consults the name server (via an RPC) to find the names of all processes required to build a logical series of telemetry processes (i.e., a telemetry string). It then asks a load-managing process to allocate a set of available processors and maps the desired processes onto them. The control process gives the new string assignments a unique name and passes this definition to the name server.

The LMC then starts each remote process (again using an RPC). The remote processes self-configure, using the name server to find the names of directly connected peers in the string. Each process also maintains its connection to the LMC. The RPC layers and connection-based network protocols provide reliable data transport between processes. In the event of network errors or other problems of short duration (approximately 0.1 sec), data integrity can be assured by protocol layers below application visibility. During normal operation, each process periodically saves its internal state to a monitor system (via an RPC or a distributed file system).

In the event of a longer failure (approximately 10 sec), a multithreaded process upstream of the failure continues to read and process input, and queues output; the writing thread is blocked. As the output queue enlarges, monitor messages are generated to alert both control processes and operators. When the problem clears, the writing thread unblocks and runs at maximum speed until the queue empties.

For a still longer failure (approximately 100 sec), the virtual circuit between the upstream server and its downstream client will expire. As before, input processing and output queuing continue. If the queue grows sufficiently large, it expands onto secondary (disk) storage, perhaps transparently via the host virtual memory system. The RPC layers contain error-recovery code that attempts to reestablish the connection; if successful, recovery takes place as described before.

In an even more extreme case, suppose that one computer in the string fails hard. When the complex controller determines that the problem is a hard system failure, it requests a backup assignment from the load manager and inserts its assignment into the string in place of the failed system. The replacement process negotiates a resynchronization with its peers (using the stored state information) and processing continues. At the end of the pass, the string assignments are dismantled and returned to the available pool.

V. Transition Approach

The transition strategy is to adopt the proposed architecture immediately and require all future implementations to conform. This strategy requires early development of requirements and design, and early demonstration of key elements.

In fiscal year (FY) 1992, a system design team should develop a top-level design that includes:

(1) Technical and mission requirements.

(2) Functional design.

(3) Management approach, including estimated costs and a funding plan.

(4) Specific standards to be employed (including protocols).

(5) Identification of key elements for early implementation.

Guidance should be provided for the design of the Block V receiver and new antenna, microwave, and receiver controllers, in order to minimize future rework.

During FY 1992 and 1993, the SPC, GCF, and NOCC upgrades should be evaluated for compatibility with the new architecture. Changes consistent with current requirements should be implemented as soon as it is practical. These changes should be viewed as parallel activities to the main architectural implementation effort and not as essential elements of it.

In FY 1993 and 1994, lower level system requirements should be developed. An end-to-end design should be produced in sufficient detail to allow a specification of the architectural infrastructure necessary for implementation. This infrastructure includes the overall data transport and software architectures described in Section IV.C and IV.D. Development of this infrastructure should begin as early as possible because it requires a significant investment in developing and procuring common hardware and software. This effort will require extensive prototyping of standard protocols and other features of open systems in order to ensure adequate performance margins and compatibility with system requirements. In FY 1996 implementation of the processing architecture should begin.

The architecture should be partitioned into pieces small enough to allow incremental implementation. Two potential implementation strategies are:

(1) End-to-end data flow strategy. Implement a segment of the GIS architecture, which includes a minimum capability for the entire end-to-end system.

(2) Facility-based strategy. Implement the system one facility at a time (e.g., DSCC).

Both approaches should employ an incremental implementation strategy. That is, they should be partitioned into smaller segments for final implementation. For example, the end-to-end strategy might be to implement only one mission and one system (e.g., tracking) as a first end-to-end segment. Other systems and missions would follow. This incremental strategy will work only if the investment in architectural infrastructure has been made as described above. In both approaches, the focus is clearly multimission; choosing one mission at a time for incremental implementation is only recommended to reduce the size of each incremental step.

The end-to-end approach is technically superior and is the recommendation of the study team because it allows the early testing and implementation of several end-to-end concepts that are crucial to the success of the proposed architecture. See Section V.D for further discussion of this approach.

The facility-based strategy has some advantages in terms of managing the scope of the implementation and the number of management interfaces for a given implementation segment. Cost issues may require consideration of this approach at later stages in the effort (e.g., after the top-level design has been completed and the architectural infrastructure has been developed).

Any implementation approach may necessitate that old systems operate for many years into the future. Using the end-to-end approach eliminates many of the requirements for gateways between old and new systems. Every effort will be made to minimize both the use of gateways and the number and duration of dual systems. Old missions and systems will be converted as soon as resources are available.

Specific transition approaches for processing, monitor and control, data transport, and software architectures are discussed in the following sections.

A. Processing

All the DSCC subsystems interface with a backbone LAN and communicate reliably to JPL through the use of standard protocols. Buffering requirements need to be identified for each of the subsystem data flows. The appropriate technology should be identified for each of the buffering requirements (0.1- , 10,- and 100-sec capability to meet the scenario described in Section IV.D), and a test-bed should be used to establish performance limits.

The processing architecture includes several high-speed systems that require special attention, including the telemetry and VLBI subsystems. Performance requirements need to be established for these subsystems concurrent with performance characterization of the networks (discussed in Section V.C).

A high-speed telemetry processor must be designed and prototyped to be integrated with the Block V receiver.

This is a critical requirement for the overall GIS architecture. This design effort should be undertaken as soon as possible so that feedback to the Block V effort will be timely (i.e., early 1992). The subsystem must have the following features:

(1) Standard backplane interface with the Block V (e.g., Futurebus+) for modularity and flexibility.

(2) Fault-tolerant architecture independent of the Block V (i.e., how many redundant systems are required? What are acceptable degraded modes? etc.).

(3) Standard interprocess communications interface with the network database server.

(4) Interface with the proposed M&C architecture for fault diagnosis and control.

(5) Standard interprocess interface with the Central Site merge process.

For VLBI quick-look processing, requirements for the volume and rate of data must be established. Work done on high-speed formatting and buffering of telemetry data will apply to this application.

The real-time telemetry interface with the Block V may be developed in parallel with an end-to-end near-real-time architecture. The subsystem must be designed to buffer data during downstream outages as described in the software architecture (Section IV.D). Much of the downstream architecture depends on an infrastructure provided by the data transport architecture (Section IV.C).

B. Monitor and Control

Goals for the proposed Monitor and Control System have been described extensively in Section IV.B. The following section explores the transition to be made from the existing operator-intensive, primarily manual system to one that is automated and has a higher level of service.

The transition approach for monitor and control is complicated because both implementation strategies (see the introduction to Section V), require that the Monitor and Control Subsystem manage processes that reside in both old and new equipment. The changes recommended to achieve the proposed M&C architecture fall into four basic areas: message communications, user interfaces, automation, and centralization.

The following sections describe the transition paths for evolving each of these areas to the new architecture. They are summarized in Fig. 26.

1. Message Communications. There are two required changes in message communications (messaging): addition of new networks and adoption of new data formats to make use of higher level open-systems network application services. Messaging transition states are illustrated in Fig. 26(a).

The subsystem transition states are complicated because it is unlikely that all the subsystems will transition at the same time, therefore several different transition states are possible.

The transition options for monitor and control are (1) the new formats would be converted to provide the same types of information as the old formats; (2) a new software layer would be added to the M&C system that can make full use of both the new and old data formats; or (3) a separate M&C system would be developed for new data formats and the old ones would be phased out as subsystems convert.

Transition Option 1 removes any near-term advantages of moving to the new formats, but is probably simplest to implement; Option 3 may result in serious operations problems by requiring operators to work with two incompatible M&C systems to accomplish an activity.

Option 2 is recommended although it requires modifications to the structure of M&C itself to support each additional new subsystem. All new subsystems should be required to use the new network and the data formats associated with an open systems architecture.

2. User Interfaces. The transition-state diagram for user interfaces, Fig. 26 (b), is relatively straightforward. The transition approach is to integrate new workstation-based GUI's with existing hardware and displays, and phase out the old interfaces as early as possible. Since the user interfaces can be updated without affecting any other areas, work can begin now to make the transition to the goal state.

3. Automation. The transition-state diagram for automation, Fig. 26 (c), is deceptively simple. Implementing the desired level of automation requires not only increases in the types of functions automated (e.g., anomaly detection, schedule execution, replanning, or fault diagnosis) but also increases in the scope of the domain in which these functions operate (e.g., all antennas or just the 70-m antenna; telemetry and ranging, but not commanding or VLBI). As both the functionality and domain grow, the domain in which higher levels of automation are possible increases because of the additional available information.

While automation will eventually exist at all levels of control in the DSN, the area with the highest payoff in the near term is in link operations (multilink operations are currently special-purpose operations only). Link operations are operator-intensive and substantial gains can be made by making link operations more efficient.

4. Centralization. Centralization is characterized by the ability to exert multilink control from the Central Site. Multilink control will depend on two major developments depicted in the state transition diagram, Fig. 26 (d): multilink capability and networks to support data communications. The recommended transition from a distributed control environment to a centralized one is to develop the capability for multilink control at the DSCC's and then transfer the capability to the Central Site.

C. Data Transport

The present data transport configuration involves heavy reliance on JPL-developed protocols. One of the major goals for the year 2000 will be to convert to an open-systems architecture based on standard protocols. The present configuration also depends on satellite communications from the DSCC's to JPL; these should be converted to fiber circuits.

Many other companies and government agencies are presently converting from a vendor-specific, or proprietary, network to an open architecture. Some are basing their new architectures on the Internet suite (TCP/IP and their related protocols). Internet products are presently widely available and provide many internetworking capabilities, although at a reduced level from that which is envisioned for OSI. Installing an end-to-end capability with TCP/IP is a reasonable step—it bases the network on standards. However, the ultimate GIS goal is greater capability with the OSI protocol suite.

Since JPL's starting point is a suite of JPL-unique protocols, a logical protocol choice for the internetwork could be either one of the following: an Internet suite, followed later by a conversion to OSI, or an immediate direct conversion to OSI. The transition will depend on the results of a cost and performance evaluation.

1. Global WAN. The data circuits from the DSCC's to JPL are primarily over satellite paths. The current circuits are 244 kbps (simplex) for telemetry data and 56 kbps (duplex) for all other data. The 244-kbps link uses a JPL-developed data-link protocol that does not provide an acknowledgment (basically a datagram protocol). Higher level application programs at JPL identify later (from several hours to several days) if any frames are missing, and request retransmission. The 56-kbps link uses a NASCOM data-link protocol. Both circuits encapsulate a JPL-developed network-layer protocol (JPL 890-201) that provides a header address for routing among facilities.

The GCF upgrade is in process, and it will affect the WAN circuits in the following ways:

(1) An upgrade to 512-kbps aggregate rate is expected.

(2) The simplex link-level protocol may be upgraded to include a selective retransmission feature to enable reliable transmission from a DSCC communications server to the Central Site communications server.

The year 2000 architecture focuses on the high-speed path from the DSCC's to JPL. This path is planned to be based on optical fiber, with its projected high speed, low error rate, and short delays. One of the most basic transformations to reach the proposed architecture will be the introduction of international optical fiber circuits.

Standard data-link and transport protocols that include error detection and correction are required for reliable (acknowledged) transmission. It is recommended that a data-link protocol with selective retransmission be implemented over current satellite links.

Another major recommendation is to introduce standard protocols at the intermediate layers, such as TCP/IP or OSI TP4 and CLNP. These may be installed simultaneously with the data link protocol. A key prerequisite to the effective use of these protocols is a relatively short time delay in the acknowledgment/retransmission process, which is provided by either optical-fiber circuits or the selective retransmission protocol with satellite communications (Section III.D). The high-speed WAN protocol implementation should be analyzed and thoroughly prototyped before installation. Once the underlying suite is installed, network-wide application-layer services can be implemented, such as file transfer and management, electronic mail, and network management. These services will provide a solid infrastructure for distributed applications.

2. DSCC Backbone LAN. The current DSCC backbone-equivalent LAN is an Ethernet that uses a JPL-developed protocol (890-131) for communication among the subsystems. The JPL protocol provides a user-defined quality of service that ranges from unreliable datagram service with no acknowledgments to reliable service with error detection and recovery.

An SPC upgrade is in process. It will affect the LAN in the following ways:

(1) A high-speed LAN will be dedicated to telemetry data. This LAN is planned to be an Ethernet using JPL 890-131 protocol with unreliable datagram service. It will transport data from the telemetry processors to the communications server.

(2) An SPC LAN will carry all other data using reliable service. This LAN is also an Ethernet using JPL 890-131.

The proposed DSCC backbone LAN architecture should be installed at the earliest opportunity. It will be the foundation on which the rest of the year 2000 DSCC architecture is based. FDDI is recommended because it is widely available now and meets the estimated performance requirements. A second recommendation is to introduce TCP/IP or OSI TP4/IP as the means by which to communicate among processes on the backbone. Conversion to standard intermediate-layer protocols will make other protocol services available, such as file transfer and network management.

3. Monitor and Control LAN. At the present time there is no dedicated M&C LAN; monitor and control data are communicated over the DSCC Ethernet with other scientific and engineering data. When the DSCC backbone LAN is installed, it is recommended that a parallel M&C LAN be simultaneously installed. A GOSIP-compliant architecture, such as MAP, is recommended. A MAP LAN similar to the DSS-13 M&C network should be extended to the SPC, and additional stations should systematically be brought on-line.

D. Software

Much of the software required to build the proposed data system is independent of the DSN's particular applications. For example, the low-level capabilities used to transport data (both spacecraft data and monitor and control messages) can be built generically and shared among subsystems. Common solutions are also possible for resource management, exploitation of parallelism, and other areas. These common system components should be prototyped and tested before detailed system design begins—subsystems can benefit from code-sharing only if the common software is available in advance. The following areas represent likely candidates for such an effort. Others are likely to emerge during the development process.

(1) Resource management. The client-server architecture necessitates some method of locating services whose mapping into a particular processor is deferred until run time. This specific problem is one of a class addressed by directory services (or naming services).

(2) Data transport. Processes in a distributed application can share common methods of connection management, data transport, and error recovery.

(3) Threads management. In general, processes in the GIS will require one or more data sources, one or more data sinks, and one or more monitor and control connections. The threads model provides a useful framework for multiplexing tasks without loss of portability.

(4) Monitor and control. The lower hierarchical levels of monitor and control are content-independent. All subsystems can share common methods for message passing and error handling, and generic message syntax rules.

(5) Software development environment. As a parallel effort, a common software development environment should be assembled. The results of the prototyping efforts should be a set of standards for development, plus tools for code control, configuration management, documentation, failure reporting, etc.

The end result of this software transition will be twofold: a software infrastructure that will allow DSN-specific applications to be built with minimal duplication of effort, and a development infrastructure that will assist application programmers in producing consistent, high-quality code in the common framework.

The early efforts should follow a spiral model of software development [2], and apply lessons learned in the design and implementation of the proposed GIS in an environment of rapid changes in the computer industry. For example, POSIX is relatively new (standardized in 1990), OSF DCE is only in a prerelease phase now, and software that implements OSI protocols is only beginning to appear on vendor price lists. A software development model that encourages prototyping will lend itself to more rapid convergence on a robust design.

When the system design team writes the lower level specifications, the quality of the resulting requirements will be enhanced by the experience gained during prototyping, and the subsequent implementations will be considerably streamlined by elimination of duplicated effort.

The system-wide consistency gained from a common software infrastructure will enable rapid and cost-efficient development of end-to-end implementations. This approach, therefore, lends itself more readily to the end-to-end approach discussed at the beginning of Section V. The facility-based approach, in contrast, would necessitate construction of gateways to interconnect old and new compo-

nents. These gateways lie outside the realm of open systems, and may divert resources from the effort to build an information system conforming with emerging standards.

E. Technology Development

This section highlights advanced development and engineering areas to be addressed during the transition period in order to implement the GIS as envisioned for the 2000–2010 era.

1. Advanced Development. The following is a summary of advanced development tasks that are required to achieve the envisioned architecture. In advanced development, JPL's role will be one of innovation, development, prototyping, demonstration, and incorporation into the DSN. Technology areas are summarized in Fig. 27.

Distributed Computing. Telemetry and VLBI subsystems require high-speed processing to decode and format the data. High-speed computing that can keep pace with input data rates is needed. Support data for antenna pointing and transmitter or receiver tuning may require intensive computation. Parallel computation techniques may be applied to these problems by sharing the workload over several processors with an appropriate software architecture. This approach involves coordinated use of multiple, possibly distributed, heterogeneous hardware and software computer systems on different aspects of a single problem. Additional investigation is needed into special cases where parallel processing should be used, and there must be prototyping to determine performance characteristics in those cases.

The availability, and projected availability, and capabilities of stable, portable parallel operating systems need to be continually tracked and evaluated as this area of computing matures. Research on distributed computing that incorporates bus I/O with dedicated telemetry processors is required. Investigation is needed into the use of multiple tightly coupled processors in 100+ MIPS computers for front-end signal processing in the GIS. Tools for custom-designed interprocessor communication software should be developed. Such tools are needed to help minimize any potential for increase in software life-cycle costs that could result from this special-purpose software.

VLSI and Optical Neural Processors, Including Hardware Packaging. Investigation of the application of high-density hardware integration and packaging techniques to DSN systems is required. Applications include board-level computing proposed for telemetry and VLBI real-time processing. The optimal field replaceable unit based on this technology needs to be established through reliability testing and evaluation. Integrated packaging concepts should be pursued, standards should be established, and multiple sources should be developed.

Research is needed on the application of neural network hardware to GIS automation functions, such as pattern recognition, performance compensation, and device control. One area of possible application is adaptive antenna stabilization and performance compensation for parts that deteriorate.

Data Storage Technology. High-speed buffering using advanced storage technology, such as high-density semiconductors, magnetic memory, and optical storage techniques, needs to be developed. The performance characteristics in GIS-type applications of these technologies need to be evaluated.

Data Management. DBMS's will be required to manage the larger volume of data and provide users with rapid access to associated mission-related data from the monitor and control information base. High-reliability servers will be required for mission data, remote monitor and control functions, and reference data for operator assistance.

Applied research into object-oriented database management systems and text-based database management systems should be pursued. The applicability of these technologies to the DSN should be evaluated through prototyping. Through prototyping, the performance of relational database management systems should be compared with the other varieties in selected DSN applications.

Software Engineering. Improved ability to manage development and implementation of new software-intensive systems is required. Needed work in this area includes development of methodologies, standards, models, and CASE tools; and fabricating an infrastructure to bind these together. Software artifact reuse, software process management, automatic programming tools, and programming environments will benefit from AI support and are particularly compelling. Automation of software testing tools to improve tracking of requirements and specifications is required. Improved software configuration and version management tools are needed. These have the ability to reduce coding time significantly and to improve the reliability of software code.

Standard graphics specification languages that allow portable hypermedia and visualization applications to be written should be investigated. This is an enabling technology for other software engineering disciplines presented in this article.

Protocols. Internetwork communication transparent to users, increased information throughput, and efficient implementation architectures need to be developed. GOSIP-based application prototyping of DSN services is required.

High-Speed WAN. Protocols for reliable high-speed Global WAN circuits from the DSCC's to JPL need to be developed and evaluated in order to improve bandwidth utilization in both directions and minimize costs for the return link.

ISDN Prototypes. Prototyping of ISDN applications for the DSN is required for global voice and facsimile communications. Applications of ISDN to manage temporary data surges should be pursued.

High-Speed LAN. Research on 1-Gbps LAN technology is required, including system integration with parallel and multicomputers, to address system performance issues. This 1-Gbps LAN technology is needed to mitigate the backbone LAN risk identified in Section IV.B and provide technology options for future distributed systems in the DSN.

Network Management. Research and development of network management systems is needed to develop a fault-tolerant data transport architecture. This is an ideal area for prototyping by the DSN because of the number of networks, the immaturity of the commercial software, and the high availability requirements of the GIS. This topic is strongly related to AI and automation technology needs.

AI and Automation. Advanced automation techniques, including AI, neural networks, and fuzzy logic are needed to increase the level of automated monitor and control at the stations, including link and resource management, system diagnostics, fault isolation, and automated recovery. New AI technology is needed in such areas as fault tolerance, knowledge-based systems, constraint-based problem solving, and computer-aided training.

Human Interfaces. Providing uniform GIS user and operator interfaces throughout the GIS should improve M&C operability, user interfaces, and training. Standard graphical (and other) user interfaces are needed and include:

(1) Multidimensional data in a form that allows people to perceive, amplify, and interpret the information content of the data.

(2) Data represented in an animation format.

(3) Transformations and algorithms introduced interactively to explore the effects on the data.

2. Engineering. The following is a summary of engineering capabilities required to achieve the envisioned architecture where JPL's role is to evaluate and adapt commercial technology for application in the GIS.

Engineering Laboratory. An Information System Engineering (ISE) Laboratory is required to evaluate commercial technologies for DSN utilization and prototype the key architectural features of the GIS. The scope should include computing environment, networking technology, automation architectures, expert systems, advanced workstations, and CASE tools. This laboratory will also provide a test-bed to evaluate platform-dependent characteristics of commercial processors; software; and data transport components, including such characteristics as WAN link throughput and LAN throughput, to help bridge the gap between theory and practice. This laboratory will be used in close conjunction with the Compatibility Test Area 21 to simulate the actual DSN operational environment.

Simulation. Simulation will be needed to estimate system loading on critical subsystems and networks. Some work on simulation is now in progress, which should provide the parametric answers needed to optimize the GIS configuration. The GIS will incorporate high-speed LAN's and WAN's. Efforts are needed to evaluate performance with different computer platforms and protocols and with their vendor-specific implementations.

VI. Summary

A unified architecture from the antenna front-end areas to the end user is proposed. Unbuffered real-time data will be processed only at the front end, and a client-server model will be employed at each subsequent processing stage. Open-system standards will be utilized to allow reliable (acknowledged) data transfer and achieve software interoperability.

These changes will substantially improve the interface between the SFOC and the DSN. Data will not be reacquired at SFOC and hence "post-pass replay" (Section II.A) will be eliminated for transmission problems of short duration. It is proposed that the DSN perform level-zero processing and merge multiple records prior to delivering the data to SFOC. Thus, the DSN would be responsible for eliminating artifacts associated with acquisition from the data stream (e.g., duplicate records).

The M&C architecture consists of four hierarchical layers: (1) network control, (2) multilink control, (3) link

control, and (4) subsystem control. The first two layers are assigned to the NCF at JPL and require limited operator support. The last two layers are typically performed at the Complex and are fully automated, requiring only backup and maintenance personnel.

The transition strategy will be to adopt the proposed architecture immediately. All future implementations should conform to the proposed architecture. This strategy requires early development of requirements and design, with a demonstration of key elements. The currently approved implementation program should be integrated with the proposed architecture.

In fiscal year (FY) 1992, a system design team should describe the system-level requirements, develop a functional design, estimate costs, and develop a funding plan. In FY 1993 and 1994, lower level system requirements should be developed. An end-to-end design should be produced in sufficient detail to allow a specification of the architectural infrastructure necessary for implementation. Development of this infrastructure should begin as early as possible. In FY 1995, the new Block V receiver, antenna, and microwave controllers should be implemented with the proposed M&C architecture, and in FY 1996 implementation of the processing architecture should begin.

An end-to-end data flow approach is recommended for partitioning the processing architecture into pieces that are small enough to implement incrementally. This approach allows early testing and implementation of several end-to-end concepts that are crucial to the success of the proposed architecture.

The architectural vision presented here provides guidance for major DSN implementation efforts during the next decade. It also provides focus for flight project mission operations systems developments. The technology forecast and impact assessment provide the basis for an expanded program of advanced development and prototyping activities. Taken together, the technology plan and architectural vision can be a significant contributor to programmatic planning by the Office of Telecommunications and Data Acquisition and MOSO and to future development activities in the Technical Divisions.

Acknowledgments

Valuable contributions to the study were made by the following people: M. Bailey, B. Beckman, J. Breidenthal, R. Burt, M. Carlton, R. Coffin, J. Dundas, S. Eng, J. R. Hall, T. Handley, R. Heuser, R. Holzman, C. Hung, W. Hurd, J. Kahr, K. Kimball, J. Leflang, P. Li, P. Liggett, D. McCornock, M. Parisi, T. Peng, J. Perry, G. Resch, M. Rodrigues, A. Salazar, J. Smith, B. Wilkinson, P. Willis, and H. Younger.

Table 1. Summary of long-range goals.

Attribute	Requirements
Performance	The normal operating mode should be near-real-time processing of telemetry, radio science, and VLBI data; capability should be provided for very high-speed acquisitions, with later off-line processing.
Availability	The GIS should have high availability due to fault-tolerant design that minimizes the impact of hardware, design, operations, and environmental faults; fault isolation should be assisted by computer software to reduce the mean time to repair.
Human operability	The GIS should operate with a minimum number of real-time operational personnel; tools needed to successfully manage an acquisition should be available to an operator at one workstation.
Flexibility and evolvability	The GIS must have the ability to rapidly and easily configure (and reconfigure) for a mission support activity; adequate performance margins should be built in at all levels to ensure flexibility to support unexpected missions.

Table 2. Leading causes of lost telemetry data, 1986–1990.

Subsystems	Hours
DSCC telemetry[a]	1398
Antenna[b]	1038
Radio frequency interference	734
Receiver-exciter	521
Microwave	482
Facility	158
Transmitter	115
DSCC monitor and control	73
Undetermined	65
NOCC support	49
DSCC tracking	45
GCF digital communications	43

[a] Major contributing assemblies are the telemetry processing assembly (1185 hr), the subcarrier demodulator assembly (110 hr), and the baseband assembly (42 hr).

[b] Major contributing assemblies are the antenna control assembly, antenna servo controller, and the subreflector controller (558 hr); and the antenna pointing assembly (60 hr).

Table 3. Current advocacy DSN mission set.

Mission	Earliest launch date	Likely convolutional code rate	Maximum date rate, kbps	Remarks
Radio astronomy (HEO)	1994	1/2	144 Mbps	Tape and engineering telemetry
Cassini	1995	1/6	156	Science and engineering telemetry
VLBI Space Operations Program (HEO)	1995	1/2	123 Mbps	Tape and engineering telemetry
Comet Rendezvous Asteroid Flyby (CRAF)	1996	1/6	180	Science and engineering telemetry and video
Advance Composition Explorer (E, HEO)	1996	1/2	100	Science and engineering telemetry
Lunar Observer	1997	1/2	500	Science and engineering telemetry and video
Near-Earth Asteroid Rendezvous (D)	1998	1/6	10	Science and engineering telemetry and video
Mars Global Network/Communications Orbiters	1998/01/03	1/6	10	Science and engineering telemetry
Solar Probe	2001	1/6	50	Science and engineering telemetry
Mars Reconnaissance/Communications Orbiter	2001	1/6	100	Science and engineering telemetry and video
Space Infrared Telescope Facility (HEO)	2001	1/2	135	Science and engineering telemetry, IR imagery (4 hr/ day at 1.1 Mbps)
Lunar aeronomy	2002	1/2	100	Science and engineering telemetry
Submillimeter Intermediate Mission (SMIM)	2002	1/2	1.0 Mbps	Spectroscopy data, highly elliptical orbit, 30-sec burst
Pluto flyby	2003	1/6	20	Science and engineering telemetry and video
Mars aeronomy	2005	1/6	100	Science and engineering telemetry
Mars Small Rover	2007	1/6	30	Science and engineering telemetry and video
Mars Rover Satellite Relay (SR)	2009	1/6	200	Science and engineering telemetry and video
Main Belt Asteroid Rendezvous (MBAR)	2011	1/6	100	Science and engineering telemetry and video
Large Deployable Reflector (HEO)	2012	1/2	5.0 Mbps	Real-time imagery
Neptune Orbiter/Probe (O/P)	2015	1/6	100	Science and engineering telemetry and video
Mars Reconnaissance/Communications Orbiters	-	1/6	100	Science and engineering telemetry and video
Thousand AU or Stellar Probe	2020	1/6	0.1	Science and engineering telemetry
Total missions to 2020	22			

Note: HEO = high-Earth orbiter; D = Discovery series small spacecraft; E = Explorer series

Table 4. Views of future telemetry requirements.

View	Telemetry data range, Mbps	Description
Conservative	1–3	End-to-end real-time flow, with limited backup mission capability, no manned-mission support, such as SEI, etc.
Moderate	1–10	End-to-end real-time flow with capability for Mars SEI, added backup capability, etc.
Aggressive	1–100	End-to-end real-time flow that can support lunar and Mars Space Exoloration Initiative, including additional backup capability, etc., over and above the moderate capability

Table 5. Ground Information System end users.

End-user interfaces	To/from Ground Information System	Data type	Content
Flight Project Mission Control Team	To/from	Voice	Real-time monitor and control
	From	Data block	DSN/SFOC configuration, performance, and status
Flight Project Planning and Sequencing Team	To	Data packet	Spacecraft commands
	To	Project sequence of events	Project sequence of events
	From	Data file	DSN system status, events, and fill tape
Flight Project Navigation Team	From	Data file	Radio metric data (range, Doppler, VLBI)
	To	Data file	Spacecraft trajectory
Flight Project Spacecraft Team	From	Data packets (virtual channel)	Spacecraft engineering data
Flight Project Principal Investigators	From	Data packet (virtual channel)	Instrument measurement data
	From	Data packet (virtual channel)	Instrument engineering data
	From	Tape-recorded data packet (virtual channel, high rate)	Instrument measurement data
Flight Project Radio Science Team	From	Data packet (digitized baseband)	Digitized spacecraft RF spectrum
	To	Data file	Trajectory information, planetary atmospheric model
Non-DSN antennas (cross support through NASA Communications Network) (e.g., TDRSS or Galileo German Space Operations Center)	From	Data packet	Spacecraft commands
	To	Data packet	Spacecraft telemetry (spacecraft engineering and instrument measurement data)
Non-DSN antennas (arraying) (e.g., Goldstone array with Very Large Array)	To	Base band signal, analog or digitized (real time)	Arraying of remote antennas with DSN
	To	Telemetry symbol, digital tape	
Planetary Data System	From	Data file	Archive of flight project data (including ancillary data)

Table 6. Summary of standard open programming languages.

Language	Standard	Comments
Ada	ANSI/MIL-STD (1815A-1983)	High-level language
C	ANSI (X3.150-1989)	Middle-level language for systems programming and real-time applications; also adopted by International Organization for Standardization
C++	ANSI standardization effort under way	Object-oriented language based on C
Fortran	ANSI 90	High-level language for numerical applications; new standard replaces Fortran 77 (ANSI X.39-1978)
Common LISP	X3J13	Major computer science research language; draft standard in public review
SCHEME	IEEE P1178	Used primarily for extension languages, computer graphics, and research

Table 7. Common-carrier circuit hierarchies.

Level	Current Asynchronous Digital Hierarchy, (CCITT G.702), Mbps		
	North America	Europe	Japan
1	1.544 (DS-1)	2.048	1.544
2	6.312 (DS-2)	8.448	6.312
3	44.736 (DS-3)	34.368	32.064
4	—	139.264	97.728

Synchronous Digital Hierarchy (CCITT G.707, G.708, and G.709), Mbps	
CCITT designation level	Data rate, Mbps
OC-1	51.84
OC-3	155.52
OC-9	466.56
OC-12	622.08
OC-18	933.12
OC-24	1244.16
OC-36	1866.24
OC-48	2488.32

Table 8. Satellite and optical fiber long-haul circuits.

Media	Typical rates	Typical bit error rate (uncoded)	One-way delay, sec	Comments
Satellite	9.6 kbps, 56 kbps, 64 kbps, 1.5 Mbps	10^{-4}	0.25	Uncoded BER typically 10^{-4}; coded 10^{-7}
Fiber	Planned SDH: 51 Mbps, 155 Mbps, 622 Mbps	10^{-9}	0.04	Service depends on availability of undersea circuits and international tariffs

Table 9. ISDN channels.

Channel	Speed, kbps
D	16 or 64
B	64
H	384, 1536, and 1920

Table 10. ISDN services.

ISDN service	Number of D channels	Number of B channels
Basic	1 (at 16 kbps)	2 (at 64 kbps)
Primary	1 (at 64 kbps)	23 (at 64 kbps U.S., Canada, and Japan); 30 (at 64 kbps Europe)

Table 11. Processor-board bus standards.

Bus	Standard	Speed, Mbps	Comments
Fiber Distributed Data Interface (FDDI)	ANSI X3T9.5	100	Optical fiber
VME bus	IEEE 1014-1987	320	32-bit standard, 64-bit (VME 64) available
Futurebus+	IEEE 896	25,000	32-bit bus available now; up to 256-bit parallel
High-performance peripheral interface (HIPPI)	Under consideration as an ANSI standard by the ANSI X3T9.3 Committee	800	

Table 12. Applicable LAN's and their characteristics.

LAN	Standard	Network topology	Speed, Mbps	Comments
Carrier-sense multiple access with collision detection (CSMA/CD)	International Organization for Standardization (ISO) 8802-3	Bus	10	Widely available as Ethernet
Token-passing bus	ISO 8802-4	Bus	5, 10	Process control applications
Token ring	ISO 8802-5	Ring	4, 16	IBM PC applications
FDDI	ANSI X3T9.5	Ring	100	Fault tolerant; optical fiber; 2 km (multimode); 40 km (single mode)
FDDI-II	ANSI, Planned 1993	Ring	100	Circuit-switched traffic multiplexed with packet traffic

Table 13. Comparison of internetwork protocols.

Open Systems Interconnection Protocol	ISO[a] standard	Internet protocol	Standard	Function
Connectionless Network Protocol	ISO 8473	Internet	MIL-STD-1777	Route data across different subnets
Transport Protocol Class 4	ISO 8073	Transmission Control	MIL-STD-1778	Provide reliable process-to-process data flow

[a] International Organization for Standardization (ISO).

Table 14. Early processing architecture candidates.

Candidates	Description
Buffering at the complex	Telemetry data are buffered at the Deep Space Communications Complex (DSCC) and delivered to a central facility for level-zero processing (no reacquisition)
Level-zero processing (LZP)	LZP is done at the DSCC and delivered to a central facility, or in some cases directly to the principal investigator; this option has been the focus of a separate study (ongoing)
Central processing (bent pipe)	As much processing (as reasonable) is done at a central facility
Dedicated strings	Dedicated strings of equipment are assigned to each antenna with little need for reconfiguration

Table 15. Processing requirements.

Process	Rate per process	Maximum number of concurrent processes	
		Year 2000	Year 2010
Telemetry (conservative)	1–3 Mbps	7	10
Telemetry (moderate)	1–10 Mbps	7	10
Telemetry (aggressive)	1–100 Mbps	7	10
Tracking	10 kbps	3	3
Command	10 kbps	3	3
Radio science	1.2 Mbps	3	3
VLBI (quick-look)	130 Mbps	1	1
VLBI (playback)	1.5 Mbps	1	1

Table 16. Telemetry processing requirements per subsystem.

Mission support forecast	Upper limit in spacecraft information rate, Mbps	Typical convolutional code rate	Block V receiver–telemetry data path, Mbps
Conservative	3	1/6	165
Moderate	10	1/6	550
Aggressive	100	1/2[a]	1830

[a] assuming support for lunar SEI, etc.

Table 17. Dedicated telemetry subsystem performance.

	1980s	1990s	2000s
Backplane bus capability	320 Mbps (VMEbus)	1 Gbps (Futurebus+)	25 Gbps (Futurebus+ evolution)
Individual spacecraft information rate	5.8 Mbps	18 Mbps	454 Mbps

Table 18. Distributed telemetry subsystem performance.

	1980s	1990s	2000s
Local Area Network capability	10 Mbps (Ethernet)	100 Mbps (FDDI)	1 Gbps (future technology)
Aggregate spacecraft information rate	180 kbps	1.8 Mbps	18 Mbps

Table 19. Evaluation of the major telemetry architecture options.

Criteria	Dedicated telemetry architecture	Switched telemetry architecture	Distributed telemetry architecture
Performance, growth potential	Easily accommodates growth; meets 2000–2010 requirements up to 450-Mbps backup support	Future growth depends on evolution of serial input and output (I/O) technology	Future growth depends on evolution of Local area network (LAN) technology
Life-cycle cost	Major interprocess communication costs affect life-cycle costs; fault-tolerant software has impact on life-cycle costs	Reduces number of telemetry processors, but requires a high-availability switch	OSI systems generally have lower life-cycle costs.
Operability	May use a highly specific operating system and require specially trained programmers	Simple programming	Operating system interface can be "open," e.g., POSIX; large pool of programmers
Flexibility, evolvability	Hardware communication system is closed (i.e., non-standard); may be difficult to add new functionality	I/O protocols may be standard, easy to add new processors	System is "open," with standard protocols
Availability, reliability, maintainability	Each Block V receiver has access to its own telemetry processor; fault-tolerant architecture: redundant processors	Mechanical matrix switches have generally low reliability; will require new technology; architecture handles telemetry faults by switching to redundant processor	Each Block V receiver has access to any telemetry processor; fault-tolerant architecture: built into LAN architecture
Technical risk	Monitor and Control needs careful design (low design risk)	High-availability switch is moderate–high risk	Availability of 1-Gbps LAN is moderate risk

Table 20. Comparison of alternative WAN architectures.

Criteria	Single circuit	Dual, diverse circuits
Performance and growth potential	No impact, probably would be fiber	Slight improvement: second link, not necessarily the same data rate, will be available for special support
Life-cycle cost	Least expensive	More expensive (two circuits)
Operability	Easiest to manage	Requires semiautomatic fault management to select alternate circuit
Flexibility and evolvability	Straightforward to negotiate with carrier	Straightforward to negotiate with carrier
Availability, reliability, and maintainability	Generally high reliability (fiber); however, when it goes down, it goes down for hours; recourse is tape backup	Highly reliable and available because of route diversity; when primary goes down, alternate circuit backs it up
Technical risk	International circuits may take longer to reestablish than domestic circuits	Small technology risk in network management area: need to know health of both links all the time

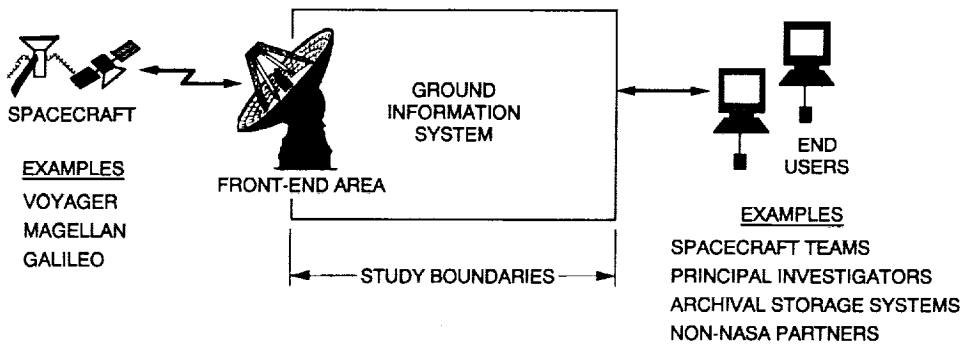


Fig. 1. Scope of architecture study.

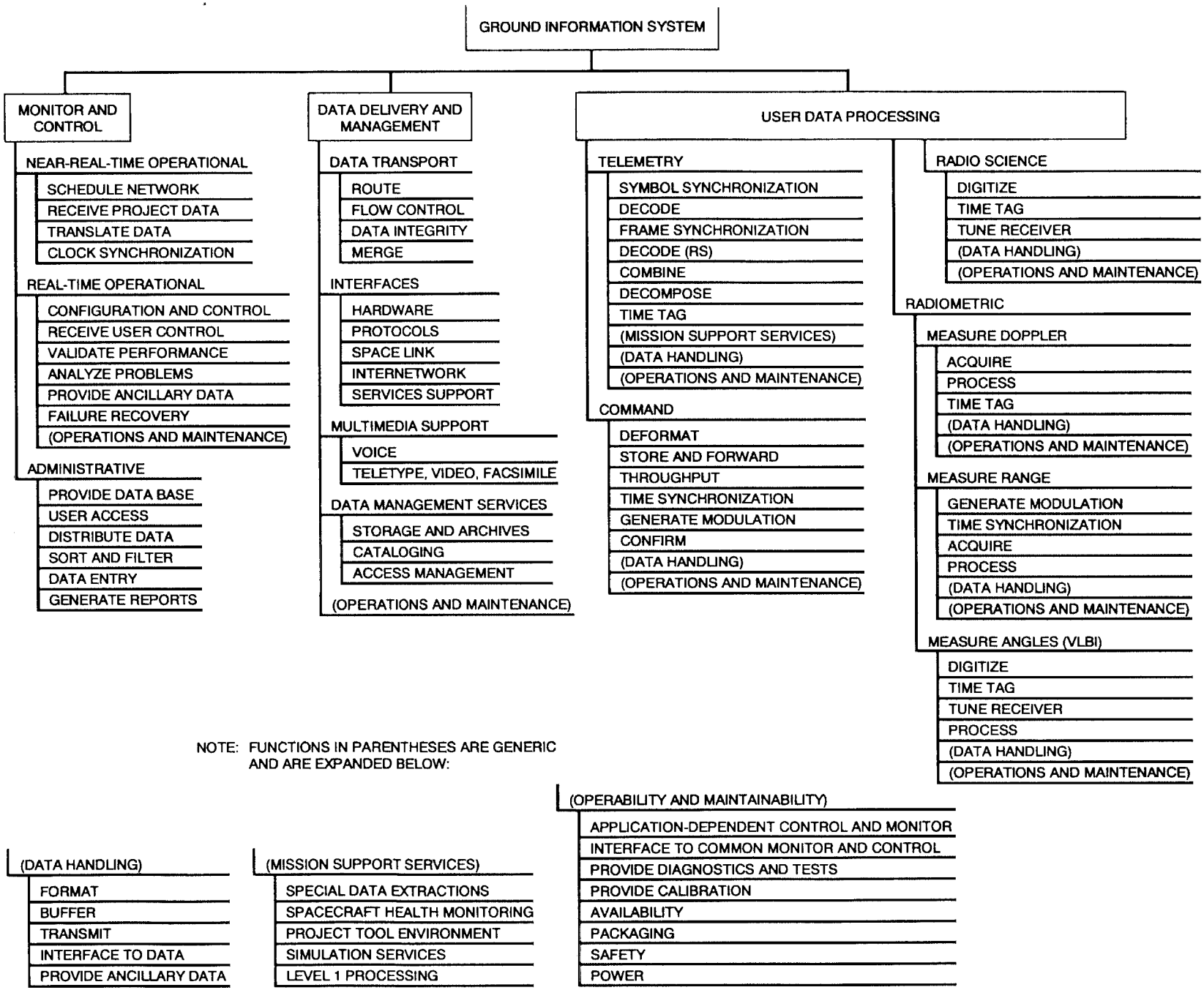


Fig. 2. Functional partition of the Ground Information System.

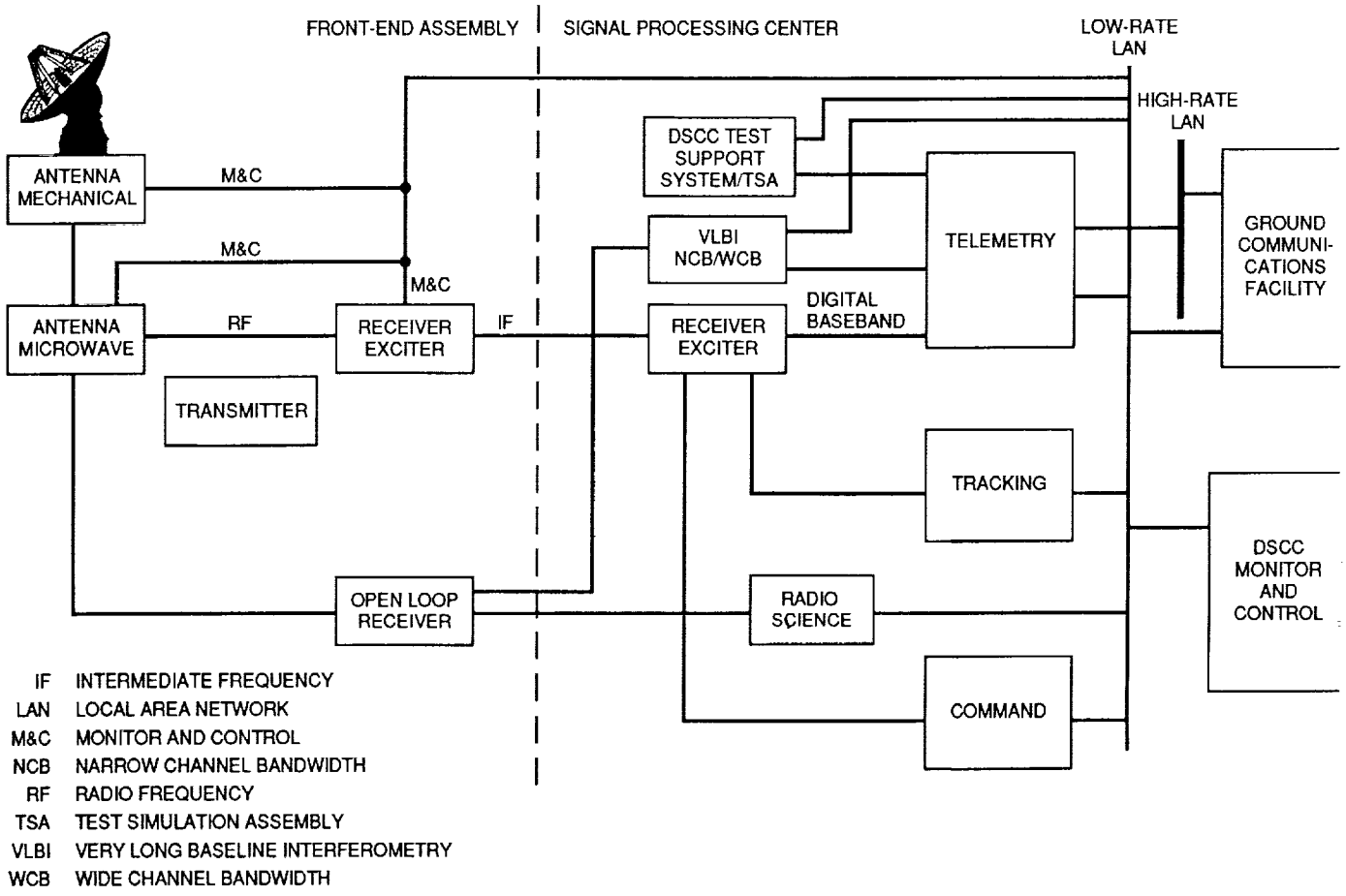


Fig. 3. Current Deep Space Station configuration.

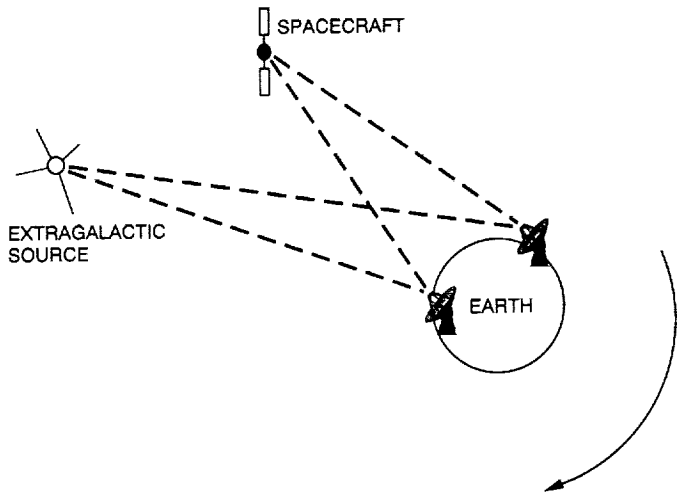


Fig. 4. Very long baseline interferometry.

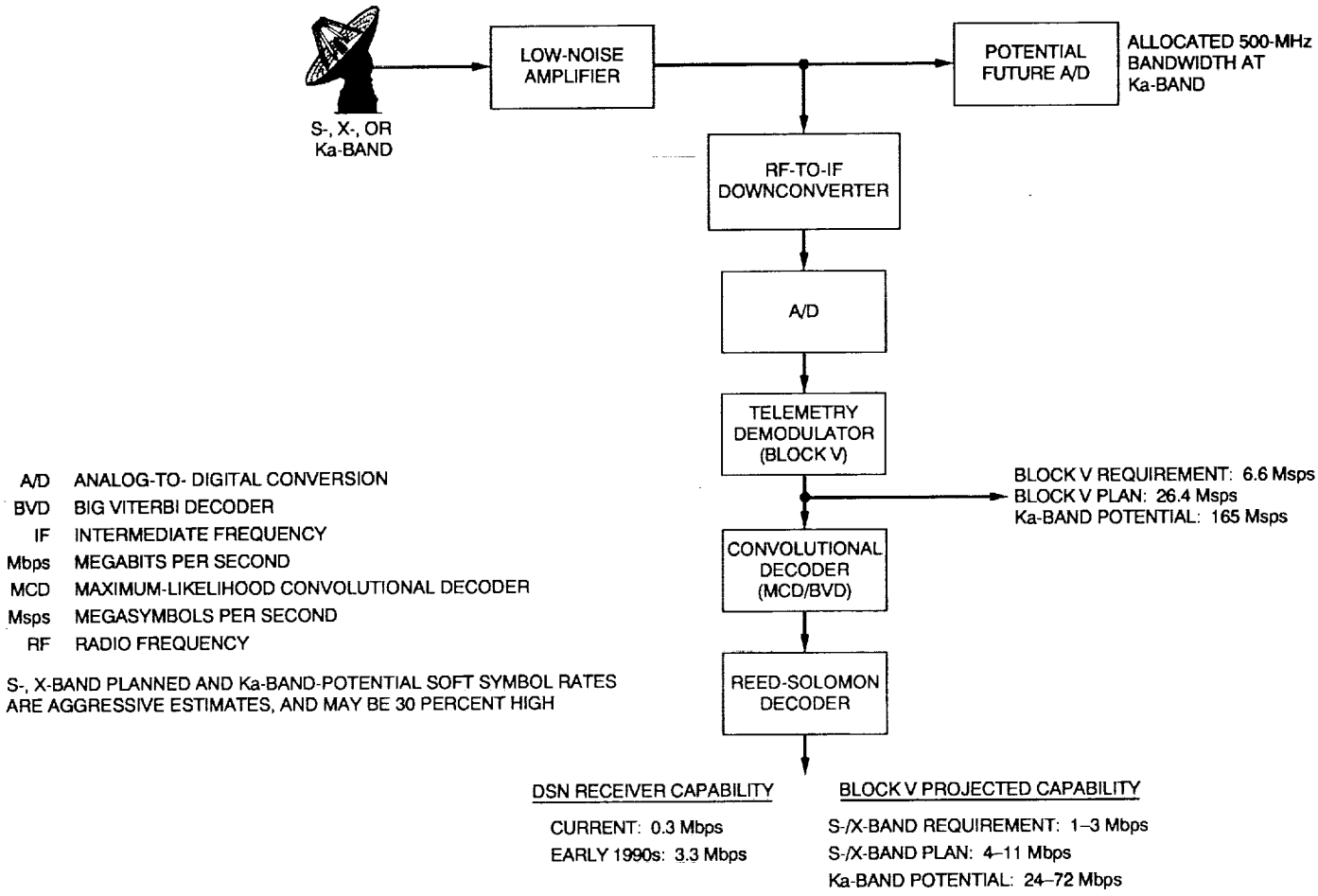


Fig. 5. Block V receiver and telemetry decoder.

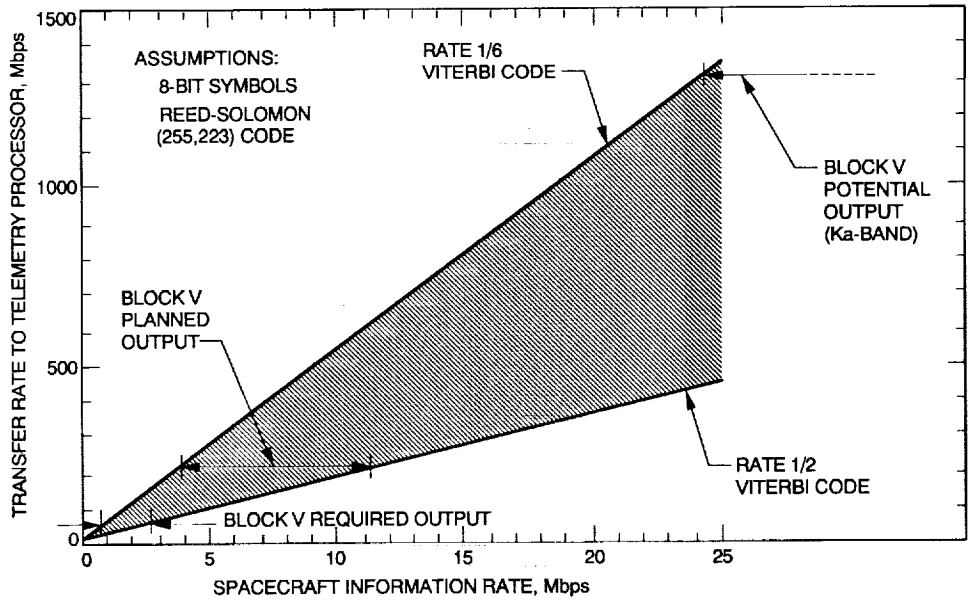


Fig. 6. Block V receiver transfer rate characteristics.

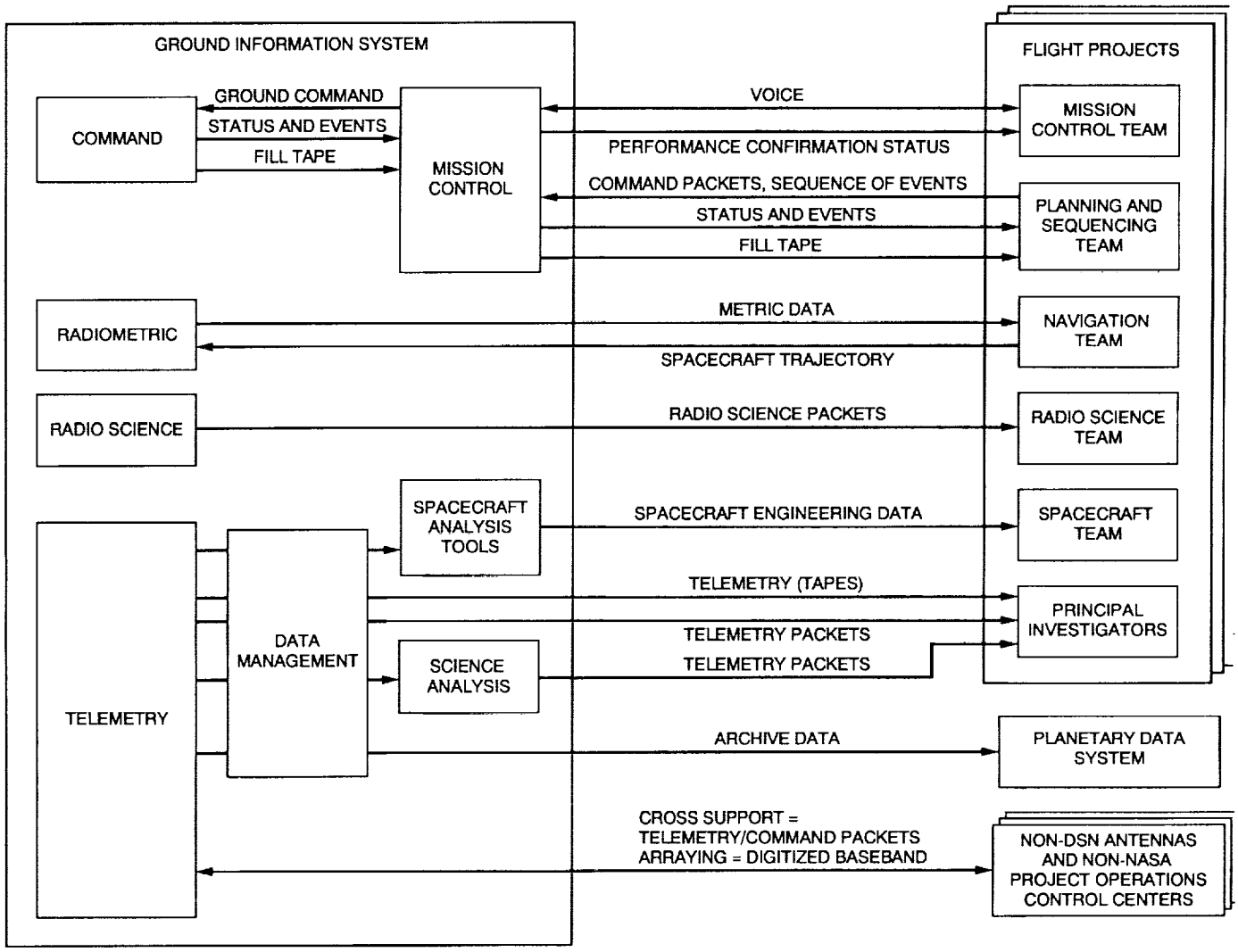


Fig. 7. External users of the Ground Information System.

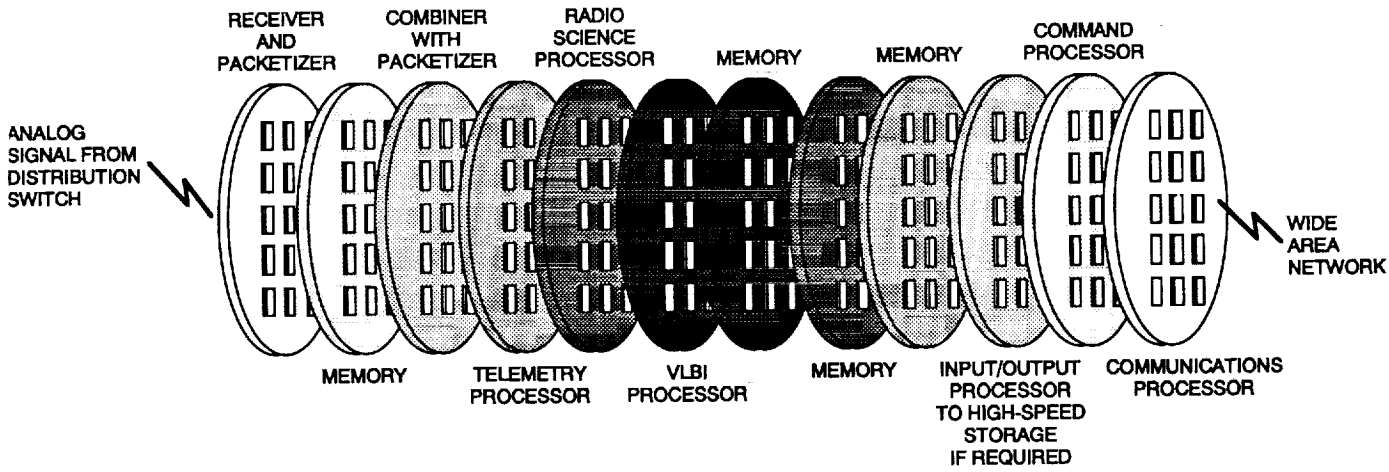


Fig. 8. Conceptual packaging of DSN processors.

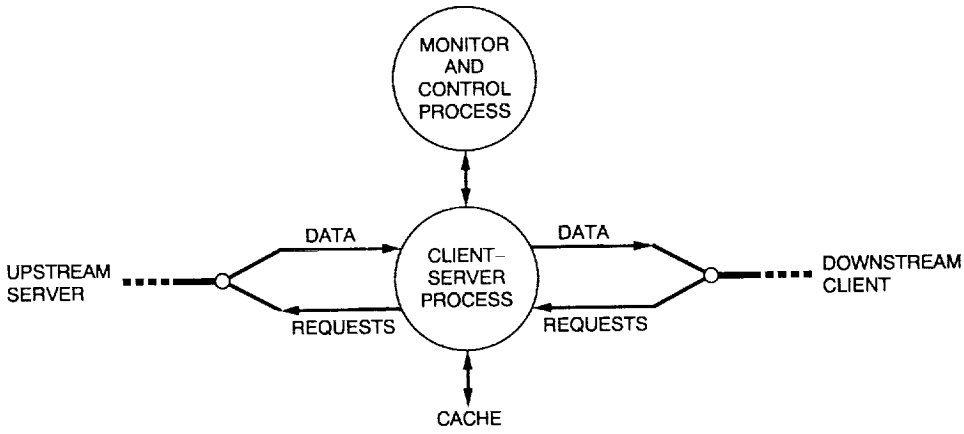


Fig. 9. Client-server model.

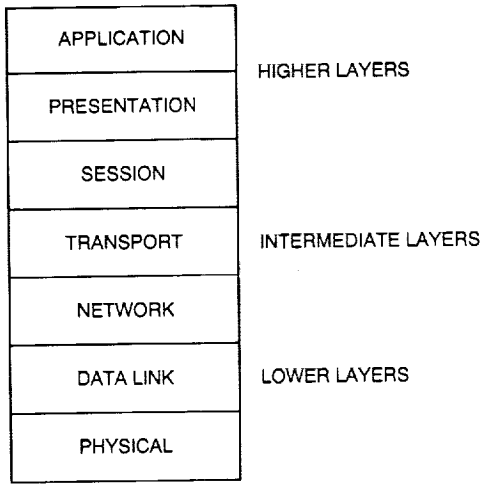


Fig. 10. Open-Systems Interconnection reference model.

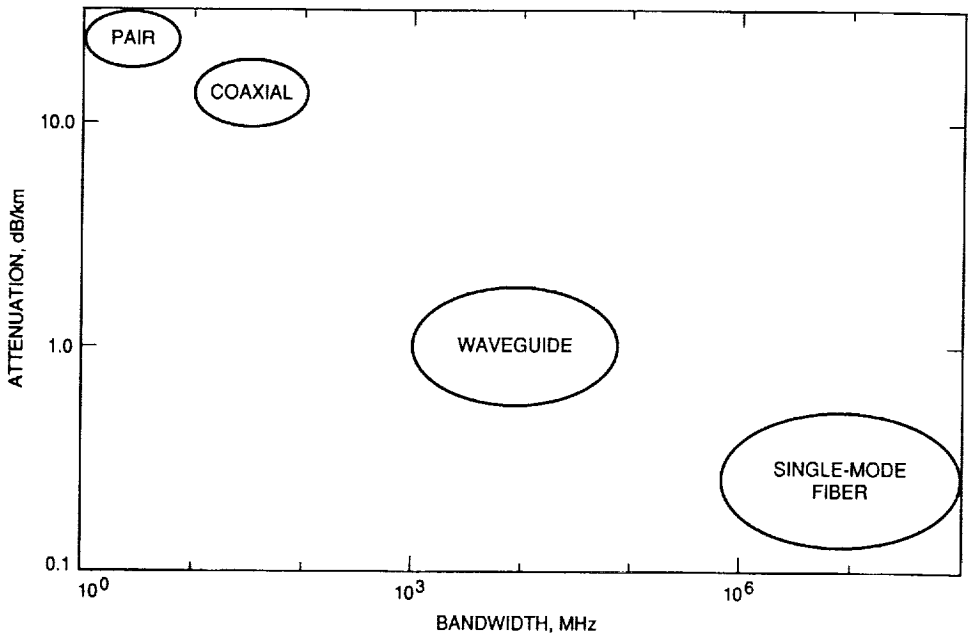


Fig. 11. Transmission properties of various media [3].

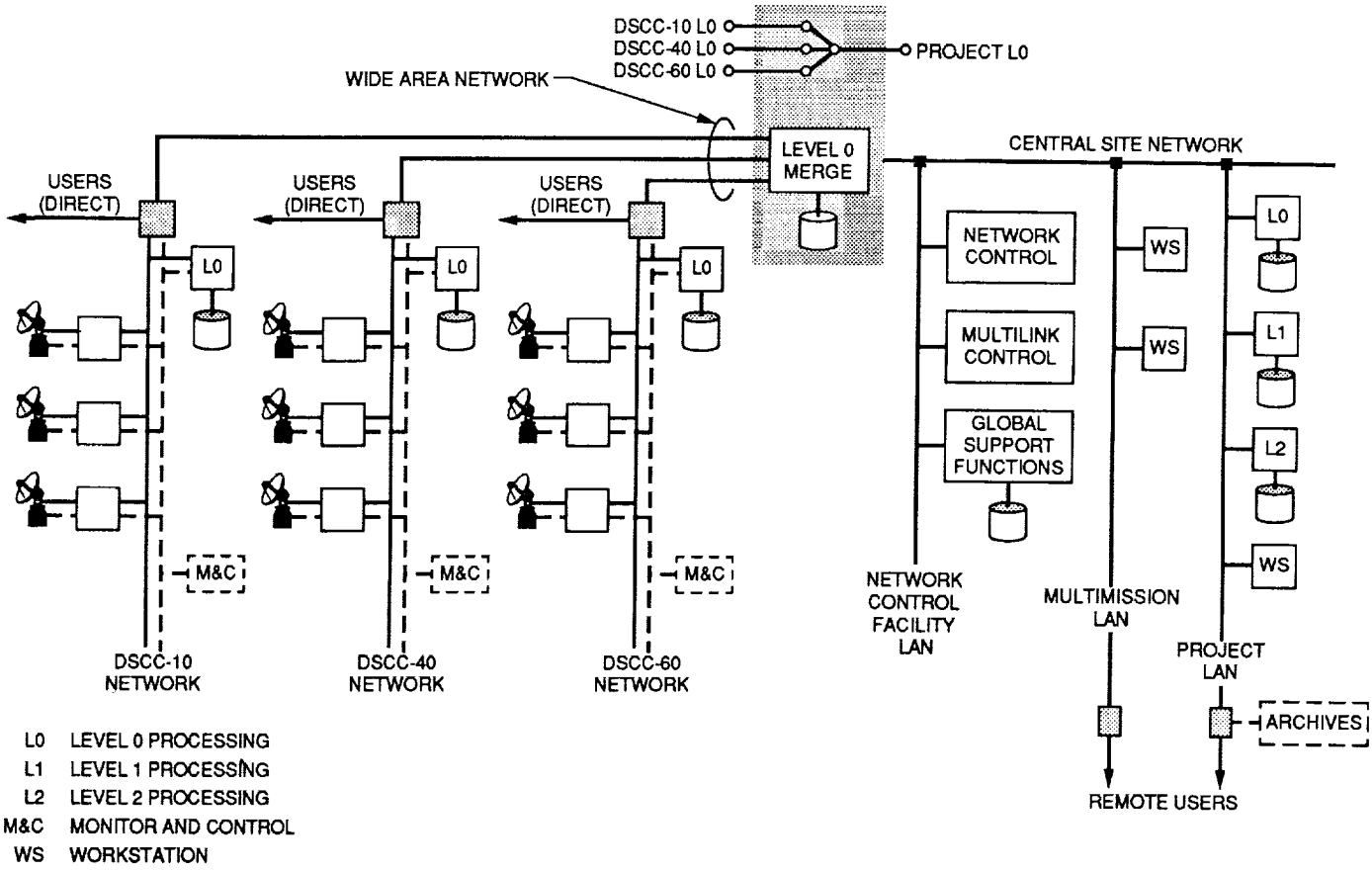


Fig. 12. Unified processing architecture.

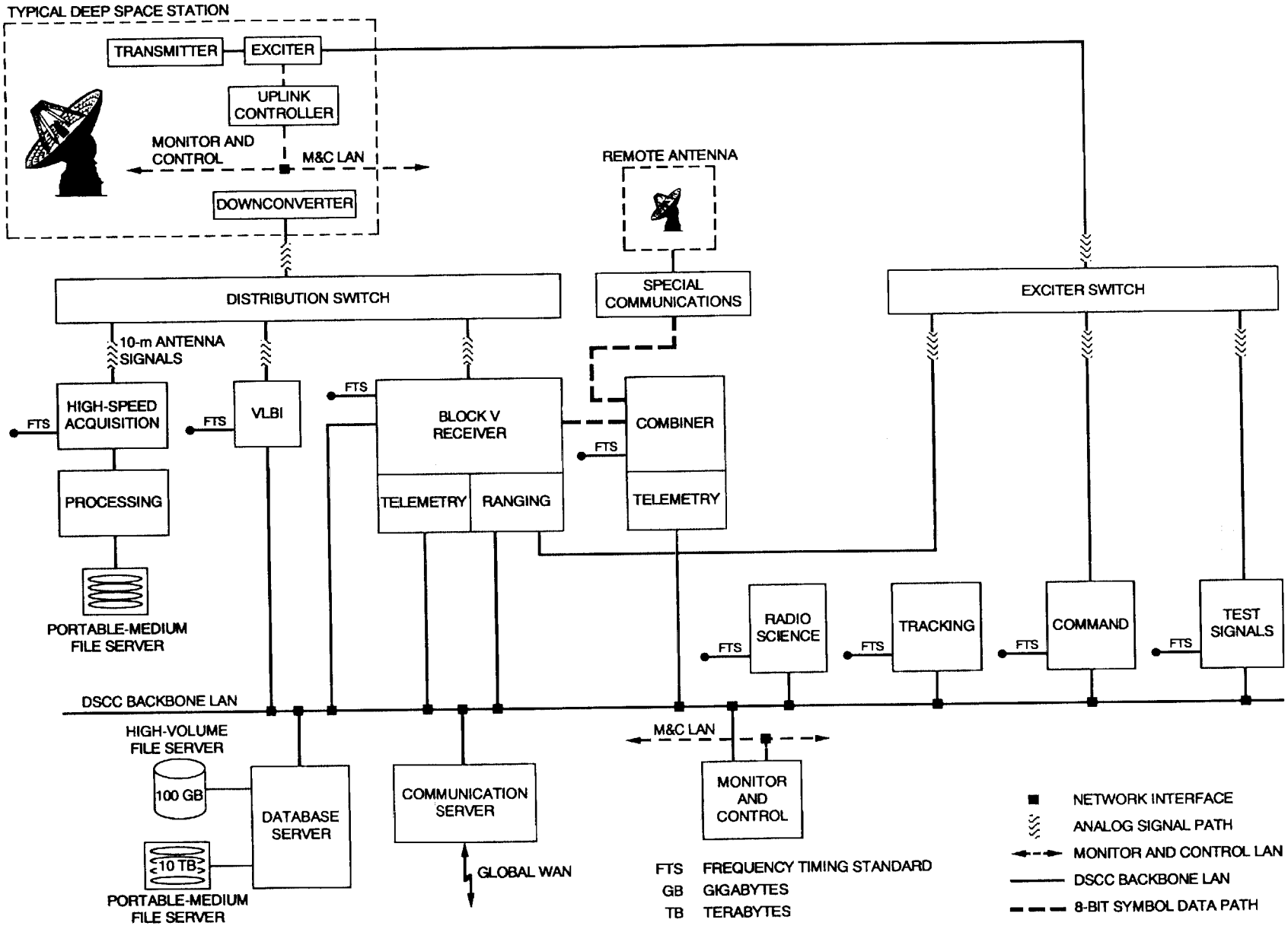


Fig. 13. DSCC processing architecture.

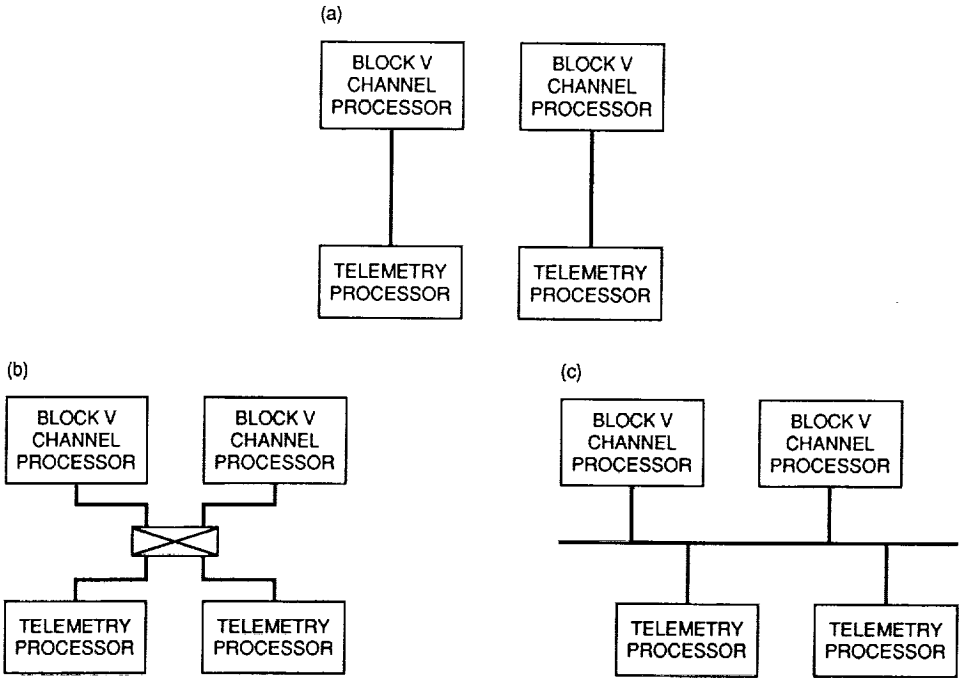


Fig. 14. Telemetry architecture candidates: (a) dedicated; (b) switched; and (c) distributed.

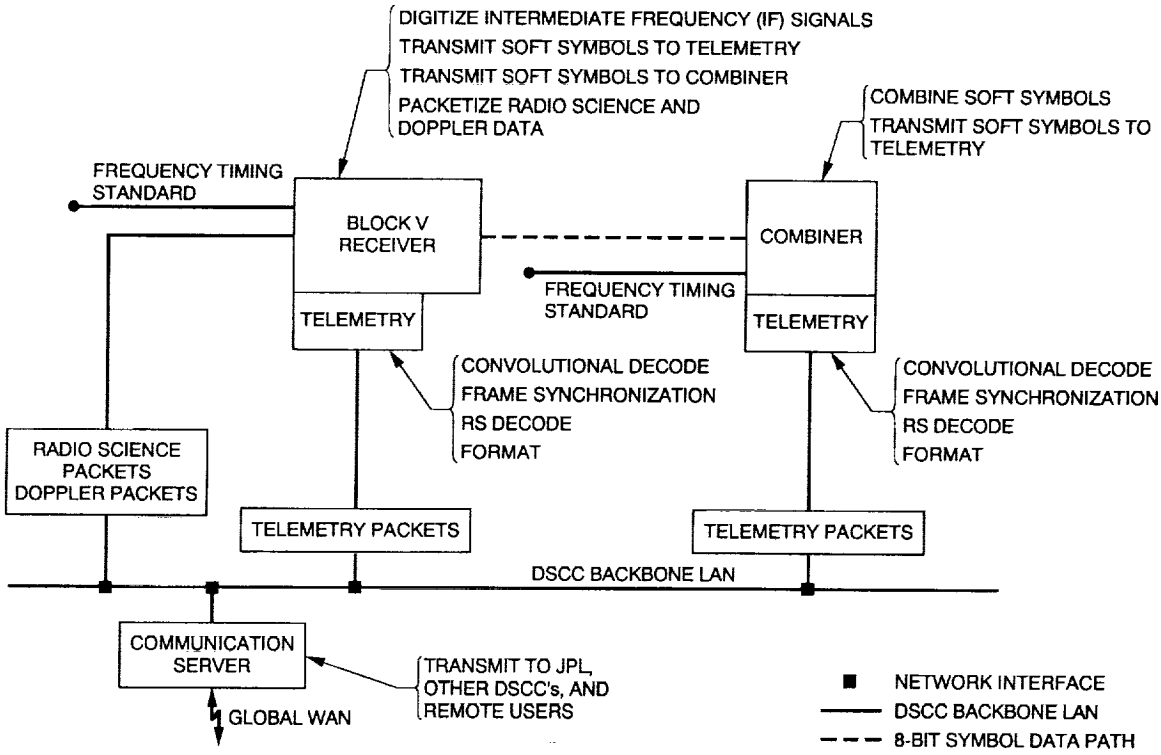


Fig. 15. Dedicated telemetry architecture.

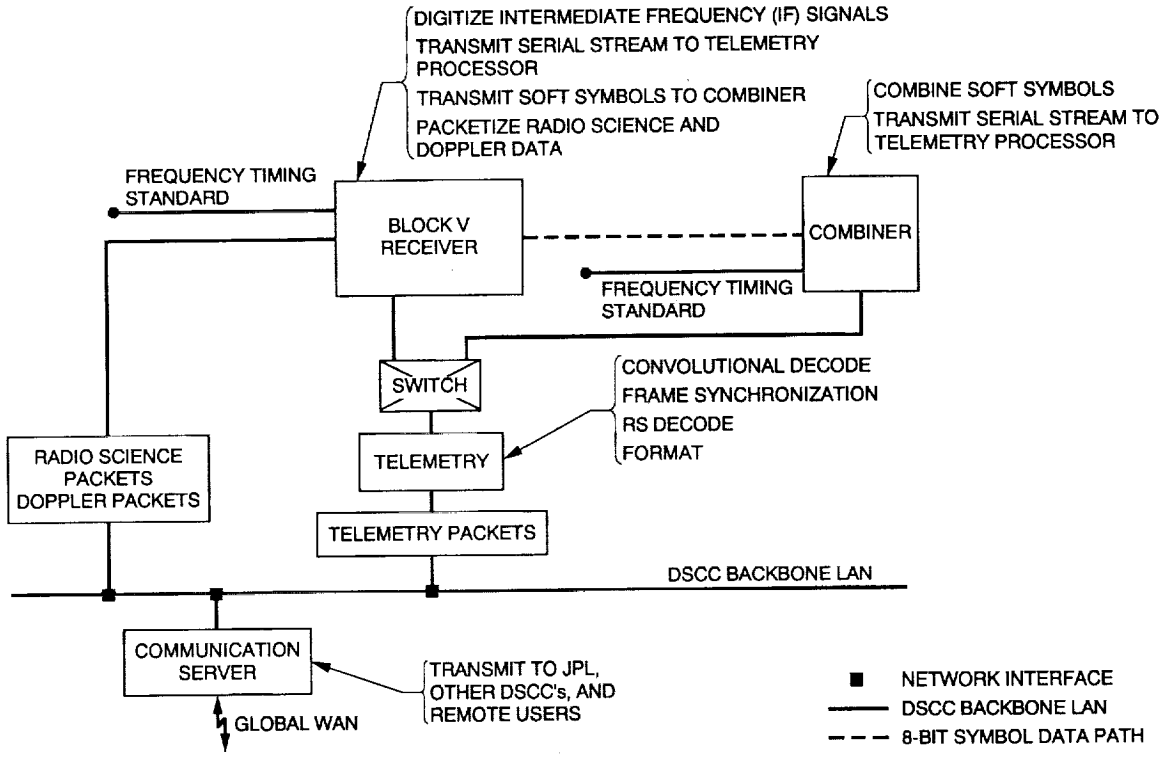


Fig. 16. Switched telemetry architecture.

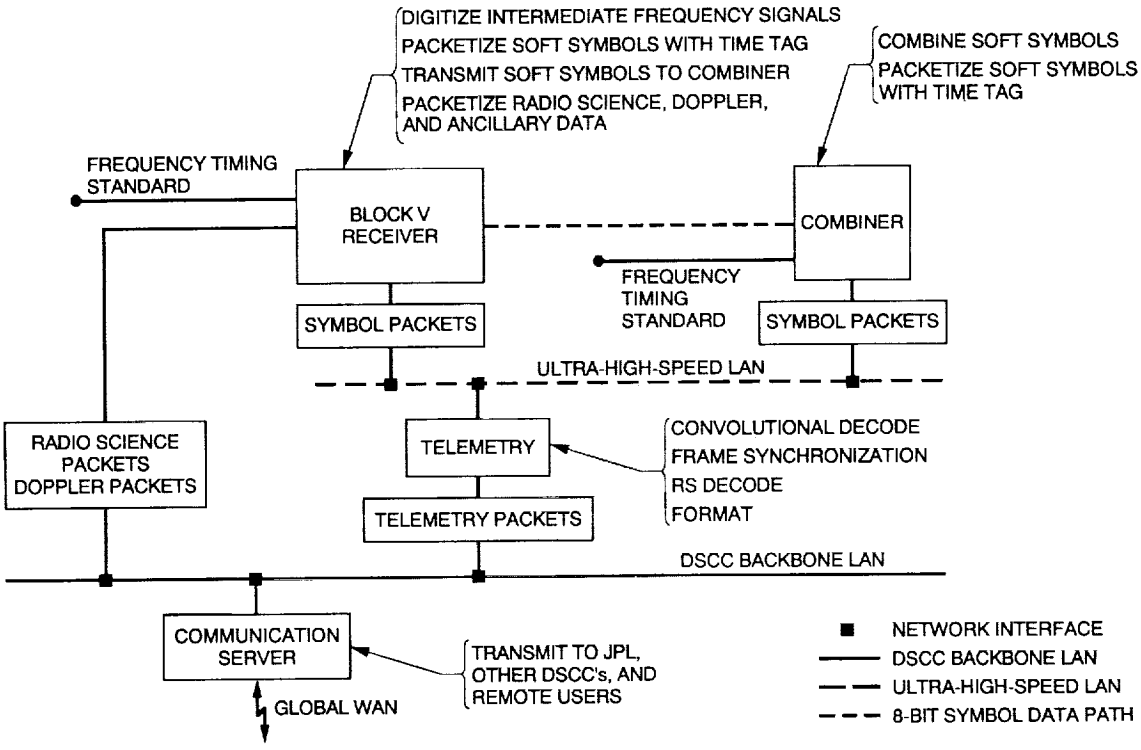


Fig. 17. Distributed telemetry architecture.

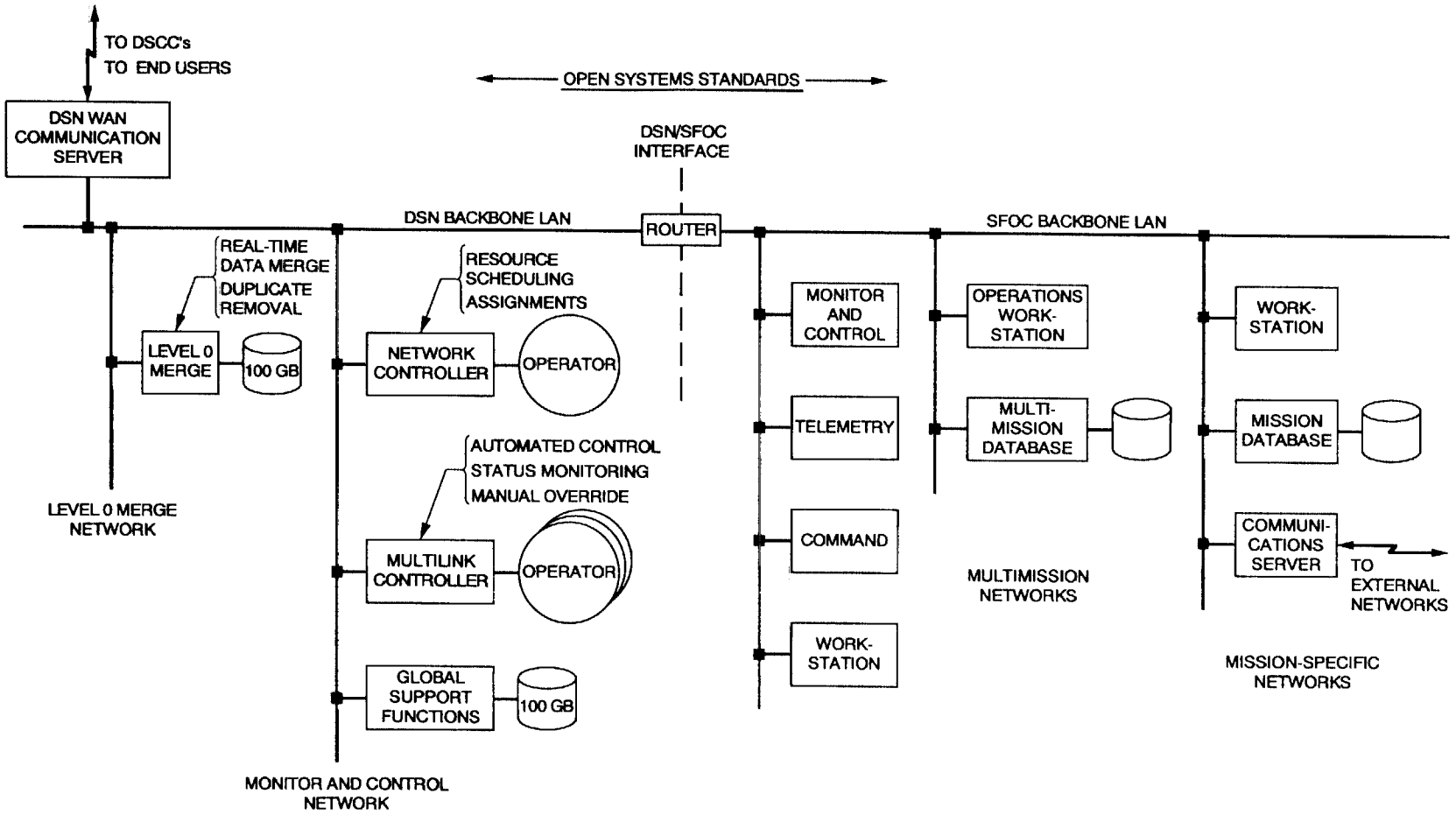


Fig. 18. Central Site architecture.

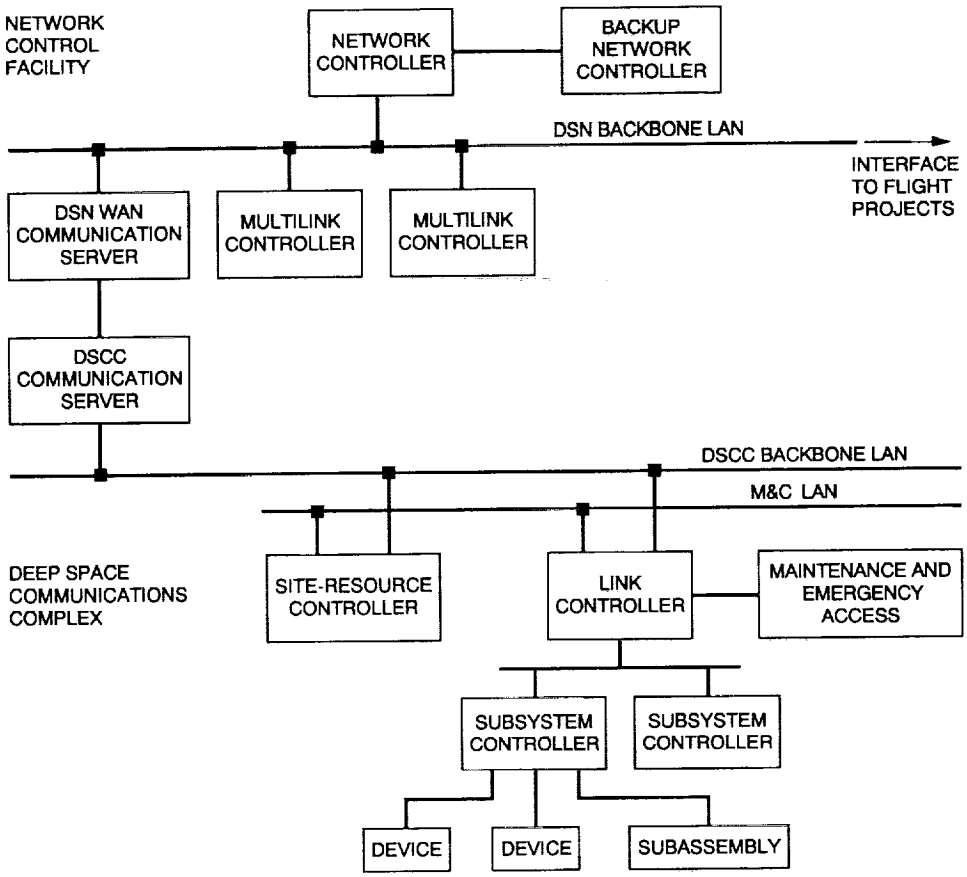


Fig. 19. Proposed Monitor and Control architecture.

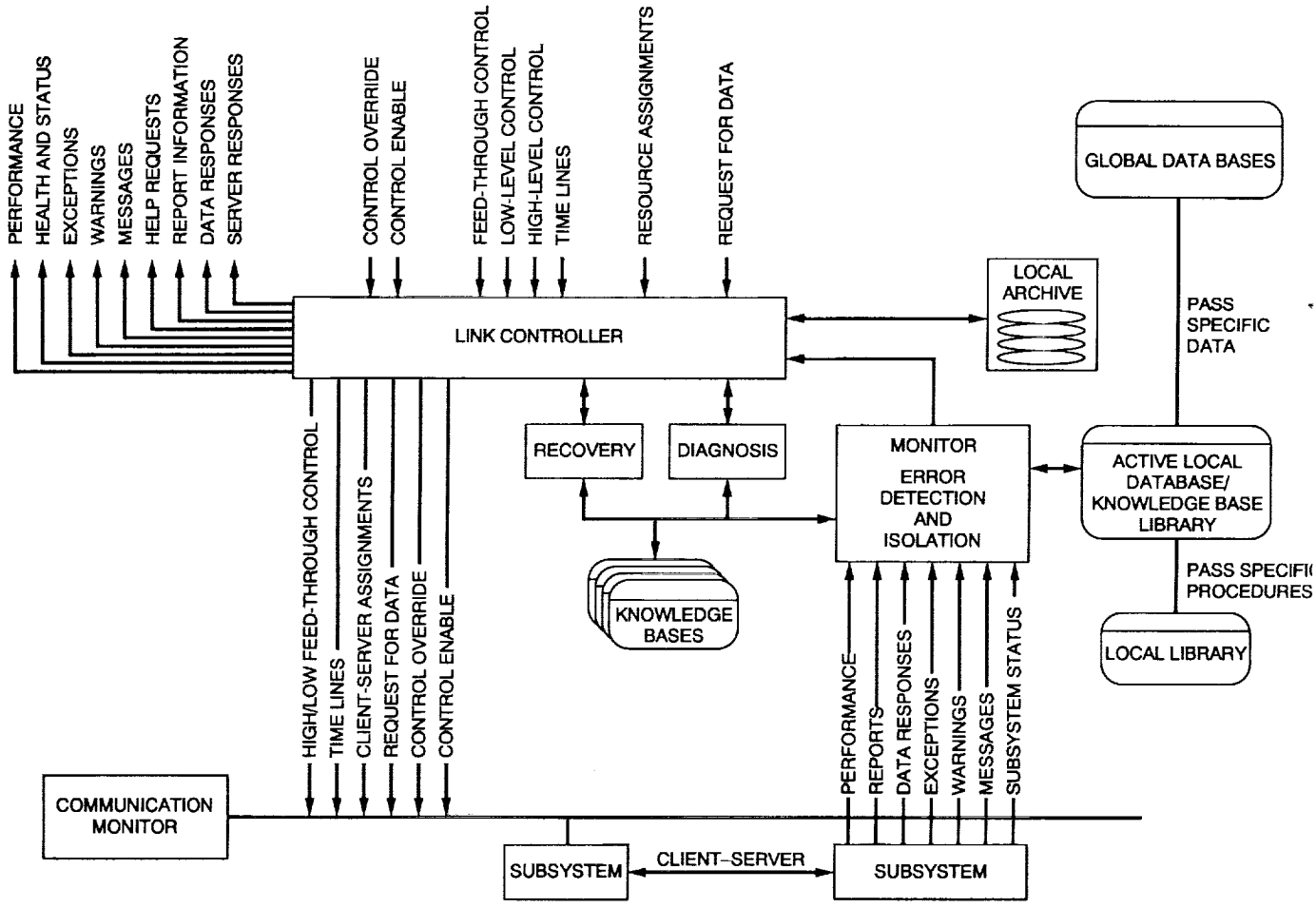


Fig. 20. A generic model for all control layers using the link controller as an example.

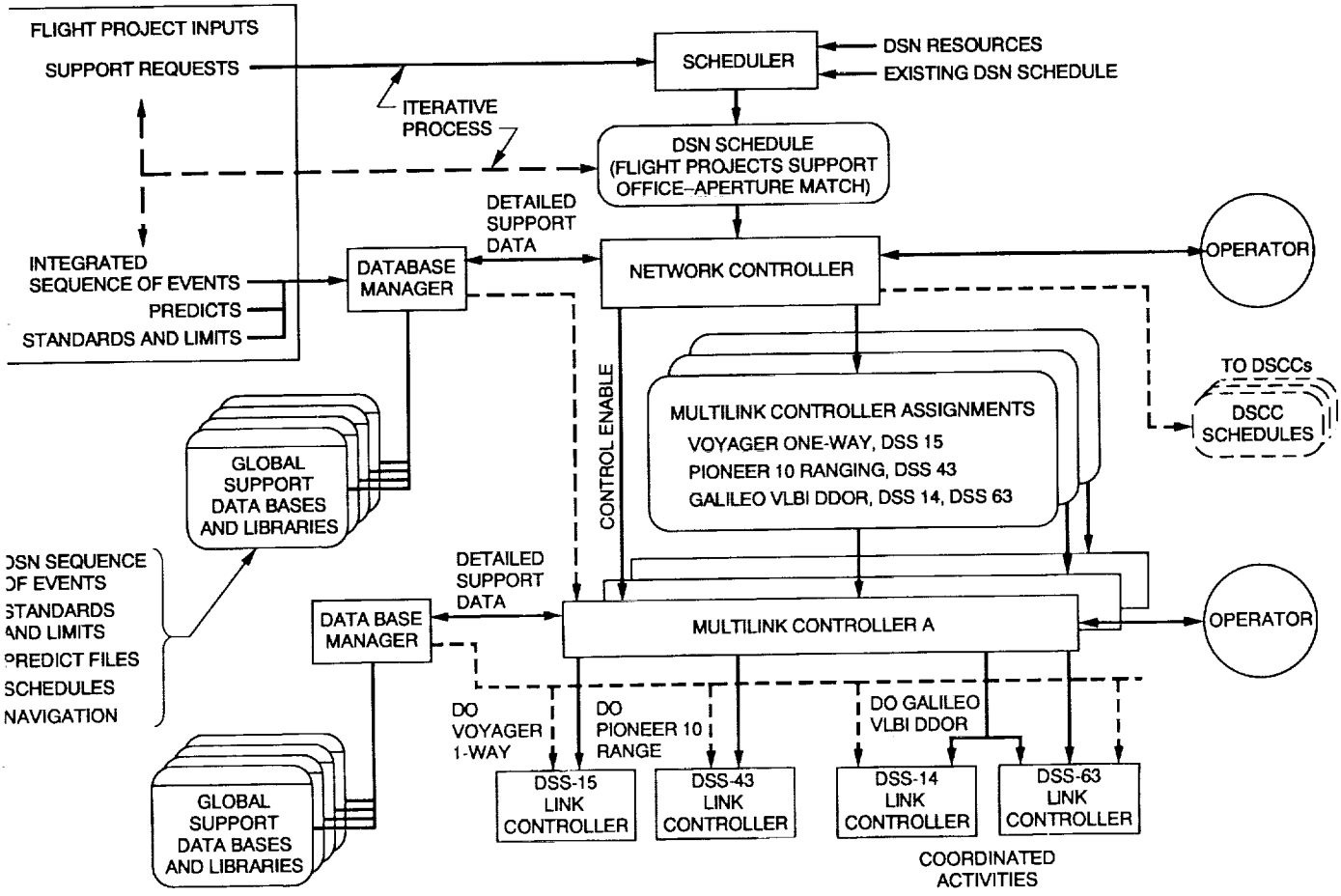


Fig. 21. Operations concept.

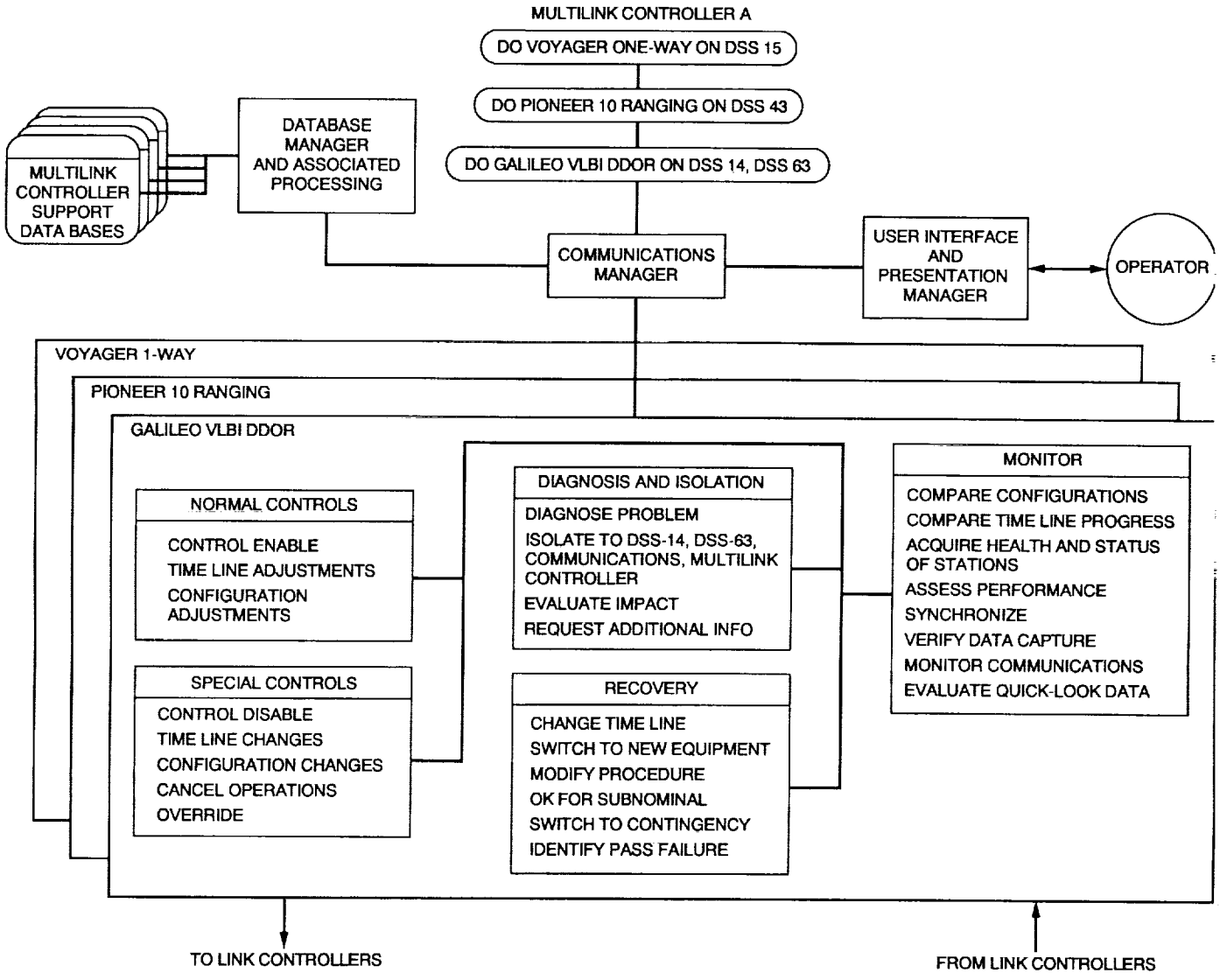
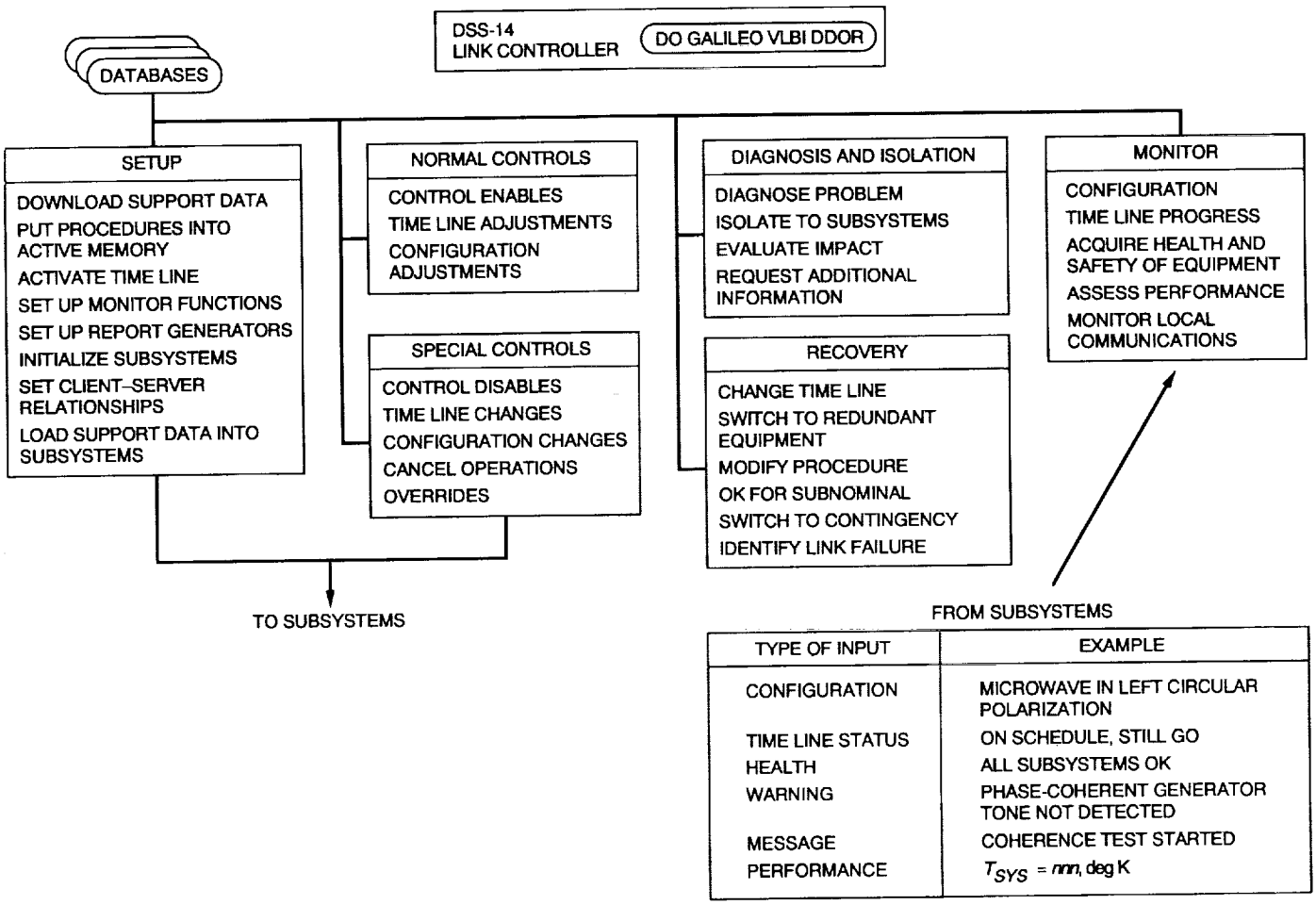


Fig. 22. Multilink controller operations.



TYPE OF INPUT	EXAMPLE
CONFIGURATION	MICROWAVE IN LEFT CIRCULAR POLARIZATION
TIME LINE STATUS	ON SCHEDULE, STILL GO
HEALTH	ALL SUBSYSTEMS OK
WARNING	PHASE-COHERENT GENERATOR TONE NOT DETECTED
MESSAGE	COHERENCE TEST STARTED
PERFORMANCE	T_{SYS} = *nnn*, deg K

Fig. 23. Link controller operations.

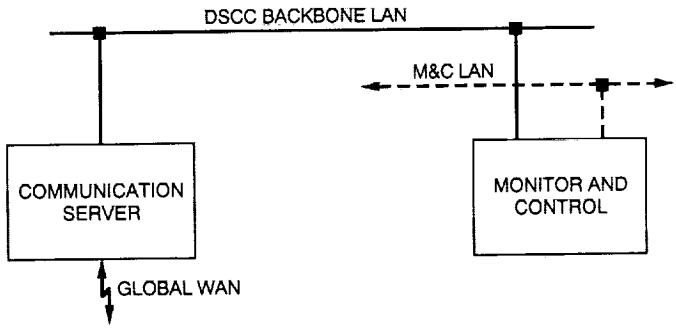


Fig. 24. Major DSCC subnetworks.

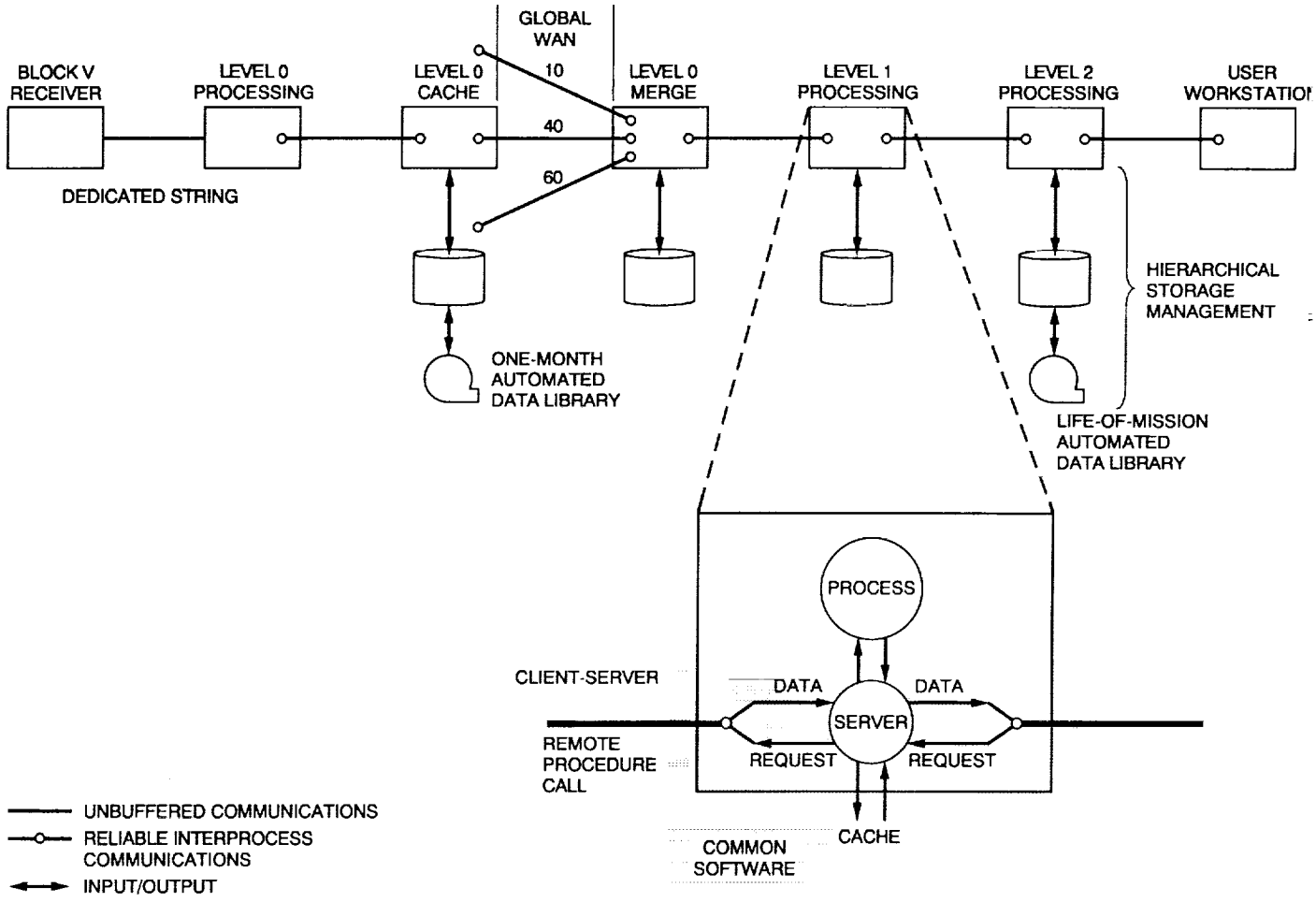


Fig. 25. Telemetry data flow with a client–server architecture.

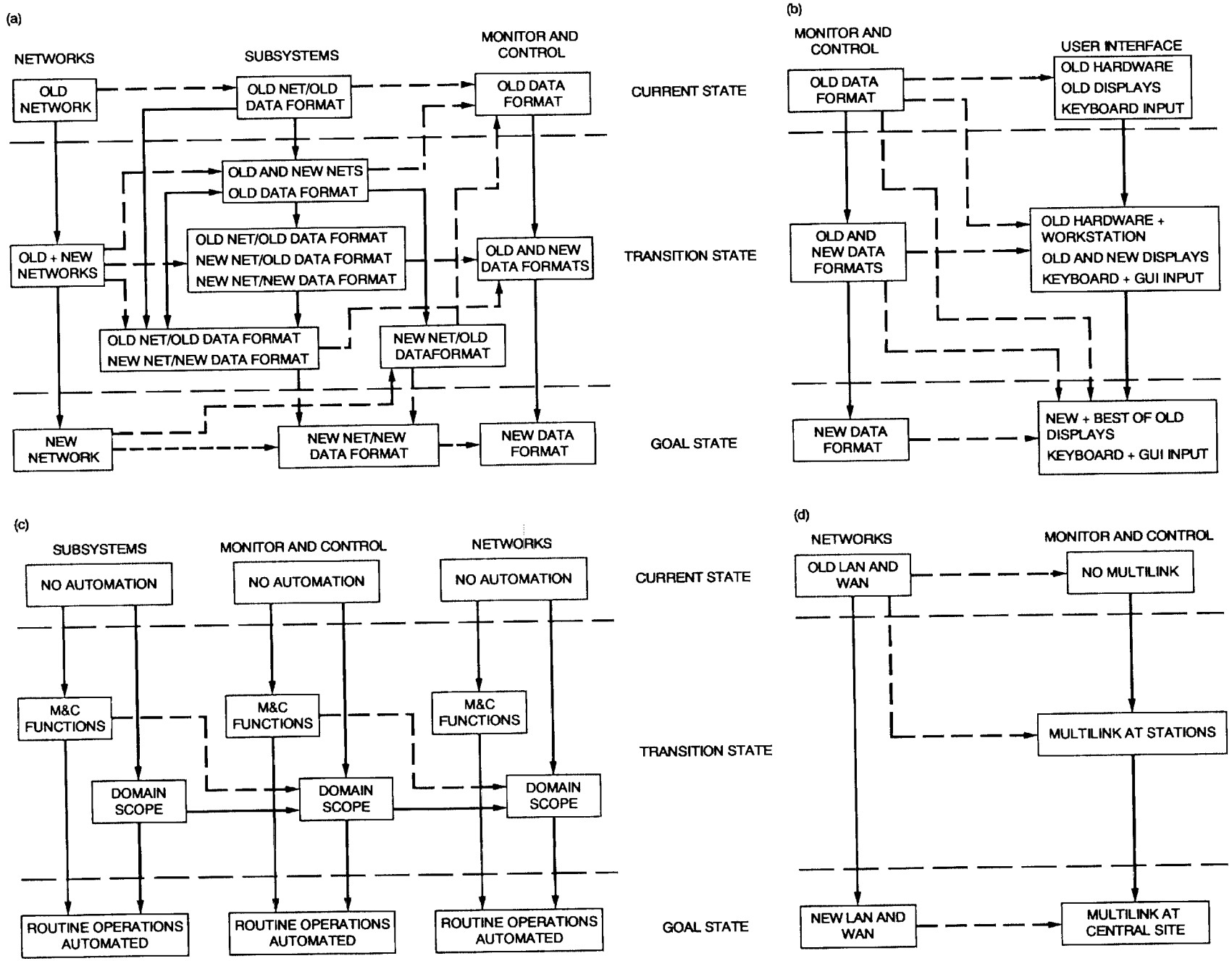


Fig. 26. Monitor and Control transition states: (a) message communications; (b) user interface; (c) automation; and (d) centralization.

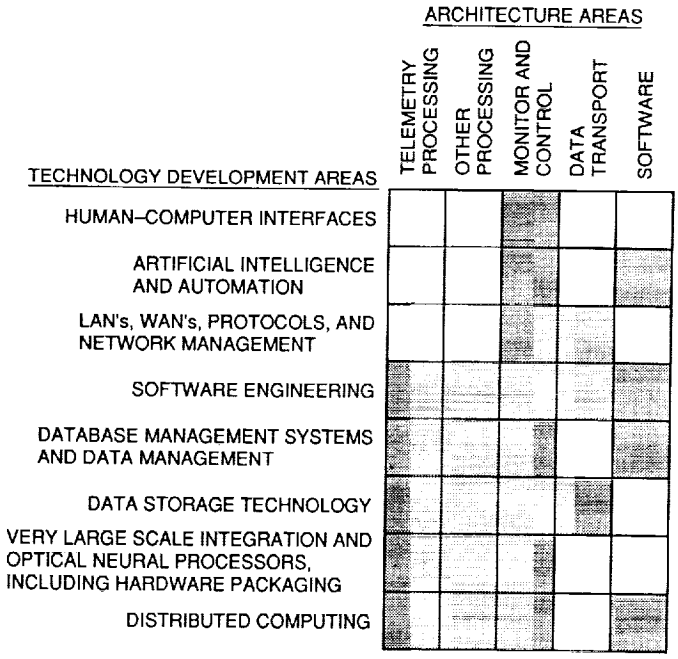


Fig. 27. Ground Information System architectures versus technology needs.

Appendix A

Acronyms

AES	Application Environment Specification
AGC	Automatic gain control
AI	Artificial intelligence
ANSI	American National Standards Institute
AOS	Advanced Orbiting System
ATDRSS	Advanced Tracking and Data Relay Satellite System
BER	bit error rate
bps	Bit per second
CASE	Computer-aided software engineering
CCITT	International Telegraph and Telephone Consultative Committee
CCSDS	Consultative Committee for Space Data Systems
CD/ROM	Compact disk/read-only memory
CLNP	Connectionless Network Protocol
CMA	Command Modulator Assembly
CMC	Complex Monitor and Control
CMIP	Common Management Information Protocol
CMIS	Common Management Information Service
CMOS	Complementary metal oxide semiconductor
COTS	Commercial off-the-shelf
CPA	Command Processor Assembly
CPN	CCSDS Principal Network
CRAF	Comet Rendezvous Asteroid Flyby
CRC	Cyclic redundancy check
CSMA/CD	Carrier-sense multiple access with collision detection
DBMS	Database management system
DCD	DSCC Command Subsystem
DCE	Distributed Computing Environment
DCO	Digitally Controlled Oscillator
DDOR	Delta differential one-way ranging
DFS	Distributed File System
DoD	Department of Defense
DSCC	Deep Space Communications Complex
DSS	Deep Space Station
DSN	Deep Space Network
DTK	DSCC Tracking Subsystem
DTM	DSCC Telemetry Subsystem
EOS	Earth Observing System
FDDI	Fiber Distributed Data Interface
FFT	Fast Fourier transform
FIPS	Federal Information Processing Standard
FTAM	File Transfer, Access, and Management
FTP	File Transfer Protocol
FY	Fiscal year
GCF	Ground Communications Facility
GFLOPS	Giga (10^9) floating point instructions per second
GIPS	Giga (10^9) instructions per second
GIS	Ground Information System
GLL	Galileo
GOSIP	Government OSI Profile
GNMP	Government Network Management Profile
GSOC	German Space Operations Center
GUI	Graphical user interface
HDLC	High-Level Data Link Control
HEO	High-Earth Orbiter
HIPPI	High-Performance Peripheral Interface
IEEE	Institute of Electrical and Electronics Engineers
I/O	Input/output
IP	Internet Protocol
ISDN	Integrated Services Digital Network
ISE	Information System Engineering
ISO	International Organization for Standardization
kbps	Kilobits per second (10^3 bps)
KBS	Knowledge-based system
LAN	Local area network
LAPB	Link Access Procedure-Balanced

LMC	Link Monitor and Control
LZM	Level-Zero Merge
LZP	Level-Zero Processing
M&C	Monitor and Control
MAP	Manufacturing Automation Protocol
Mbps	Megabits per second (106 bps)
MBAR	Main Belt Asteroid Rendezvous
MDA	Metric Data Assembly
MFLOPS	Mega (10^6) floating point instructions per second
MHS	Message Handling Service
MIB	Management Information Base
MIMD	Multiple instruction–multiple data
MIPS	Million instructions per second
MLC	Multilink Controller
MMS	Manufacturing Message Specification
MOSO	Multimission Operations Systems Office
Msps	Megasymbols per second (10^6 eight-bit symbols per second)
MTBF	Mean time between failures
NASA	National Aeronautics and Space Administration
NASCOM	NASA Communications Network
NCB	Narrow-Channel Bandwidth
NCF	Network Control Facility
NIST	National Institute for Standards and Technology
NOCC	Network Operations Control Center
NSS	Network Support Computer
OSF	Open Software Foundation
OSI	Open Systems Interconnection
OSO	Office of Space Operations
PAD	Packet assembler/disassembler
PCM	Pulse-coded modulation
POC	Project Operations Center
POSIX	Portable Operating System Interface for Computer Environments
RDA	Remote database access
RF	Radio frequency
RFI	Radio frequency interference
RPC	Remote procedure call
RS	Reed-Solomon
SDH	Synchronous Digital Hierarchy
SEI	Space Exploration Initiative
SFOC	Space Flight Operations Center
SHARP	Spacecraft Health Automated Reasoning Prototype
SIMD	Single instruction–multiple data
SMIM	Submillimeter Intermediate Mission
SMTP	Simple Mail Transfer Protocol
SNMP	Simple Network Management Protocol
SOE	Sequence of events
SONET	Synchronous Optical Network
SPAN	Space Physics Analysis Network
SPOC	Science Planning Operations Computer
SPC	Signal Processing Center
SQL	Structured Query Language
SR	Satellite relay
SRA	Sequential Ranging Assembly
SSF	Space Station Freedom
Tbyte	Terabyte (10^{12} bytes)
TCP	Transmission Control Protocol
TDA	Telecommunications and Data Acquisition
TDRSS	Tracking and Data Relay Satellite System
TP	Transport Protocol
TP4	Transport Protocol Class 4
TPDS	Tracking and Data Processing Subsystem
VHSIC	Very high speed integrated circuits
VLA	Very Large Array (National Science Foundation, Socorro, New Mexico)
VLBI	Very long baseline interferometry
VLSI	Very large scale integration
VSAT	Very small aperture terminal
WAN	Wide area network
WCB	Wide-Channel Bandwidth

Appendix B
Glossary

Application environment: Refers to all the software necessary to run an application except for the application-specific code. It includes the operating system, I/O routines, and network services and may include such items as a DBMS and graphical user interface (see Section III.C).

Architecture: A specification that defines the design of a system, and describes the structure, functions, and interfaces of its cooperating parts (see Section I.C).

Distributed: In network technology, this term refers to geographically dispersed computers that share common applications and information via local-area networks and wide-area networks (see Section IV.B).

Fault-tolerant: Systems that are able to detect a fault, report it, mask it, and then continue service while the faulty component is repaired off-line.

Dedicated: Systems that perform only one primary function, e.g., telemetry processing on the output of only one receiver (see Section IV.B).

Artificial intelligence: The technology whereby computer systems can be programmed to emulate special properties and capabilities attributable to human intelligence (see Section III.E).

Backbone network: A network designed to interconnect lower speed channels (see Section IV.D).

Central Site: The control center of the DSN, located at Jet Propulsion Laboratory (JPL) facilities in Pasadena, California (see Section IV.B).

Client-server: In a communications network, the client is the requesting machine and the server is the supplying machine (see section III.C.3).

Operating conditions: The conditions under which a system performs its functions.

Anomalous: A situation where operational procedures do not exist.

Backup: A situation where operational procedures are used to accommodate a failure of the primary system.

Nominal: In the operations concept of the M&C architecture, nominal operations refers to a situation where existing operational procedures can handle any problem.

Codes: As used in this study, codes are error-correcting bit patterns that are used to reconstruct a spacecraft signal into a form so that departures from this construction in the received signal can be automatically detected. This permits the automatic correction at the receiving terminal of some or all of the errors. Such codes require more signal elements (or bits) than are necessary to convey the basic information [4] (see Section II.A).

Data: The normal definition of data in this article refers to representation of information in binary digits (bits). However, in estimating overall data rates, data may also include redundant bits added for error detection and correction.

Near-real-time: Data transmitted directly from a sender to a receiver with possible delays on the order of seconds due to temporary buffering (see Section II.A).

Non-real-time: Data transmitted from a sender to a receiver with significant storing and shipping delays (on the order of hours or days). (See Section II.A.)

Real-time: Data that is transmitted directly from a sender to a receiver with virtually no delays.

Datagram: Data packet that is transmitted with sufficient information to reach its destination, but with no expectation of an acknowledgment.

Data protocol: A standardized set of rules that specifies the format, timing, sequencing, and/or error checking for data transmission.

End users: Recipients of DSN-produced data products. Recipients include spacecraft teams, principal investigators, archival storage systems, and non-NASA partners (see Section II.D).

Front end: The antenna equipment and electronics collocated with the antenna.

Fuzzy logic: Logic based on probabilities that statements are true rather than whether they are true or false (see Section III.C).

Gateway: A device for translating application services from one set of protocol standards to another, e.g., electronic mail from Internet to OSI.

Ground Information System (GIS): The DSN information processing system. In scope, it extends from the ground side of the front end to the end users. GIS functions include processing (telemetry, tracking, command, radio science and VLBI), monitor and control (of front end and processing), and data delivery and management, with the focus of this study on multimission operations (rather than on project-unique activities). (See Section I.C, II.D.)

Hypermedia: Technology that enables a user to review data by following a nonlinear path, for example, by jumping to related data items. The formats may include text, graphics, animation, digital audio, and video (see Section III.E).

ISE Laboratory: Information Systems Engineering Laboratory; a proposed laboratory for the envisioned GIS architecture which will be required to evaluate, prototype, and adapt commercial technologies for DSN utilization (see Section V.E).

Internet: As normally used in this study, a short form of "internetwork," a generic name for any network that interconnects two or more multivendor computers and networks (see Section III.D).

> *DoD Internet:* A DoD packet-switched network that uses military standard protocols for such internetwork services as ensuring message reliability, routing, file transfer, and electronic mail (see Section III.C).
>
> *CCSDS Internet:* Formal name for a spacecraft-to-user packet delivery service specified in CCSDS Advanced Orbiting Systems (AOS) architecture (CCSDS 701.0-B-1). (See Section III.D.)

Layers: Result of analyzing a system by dividing the system functions into groups of common services. Each layer specifies its own functions and assumes that lower level functions are provided. In "open systems" architecture, a collection of related functions that comprise one level of a hierarchy of communications functions.

Level-zero processing: Processing of spacecraft telemetry data to remove all artifacts of space data transport from the data and return it to the form it was in when it left the instrument aboard the spacecraft. Multimission level-zero processing (LZP) is done, to the extent possible, at each DSCC; final multimission level-zero merge (LZM) occurs at the Central Site.

NCF: Network Control Facility; an envisioned organization located at the Central Site, that exercises the highest level of monitor and control; it is planned to be the primary control point for the entire DSN (see Section IV.C).

Neural network: Parallel, distributed computing system of simple computational elements connected in a fashion which mimics, at a low level, the connections between neurons or systems of neurons in a brain (see Section III.C).

OSF: Open Software Foundation; a consortium of computer-product vendors committed to developing products that provide common application environments (see Section III.C.1).

OSI: Open-systems interconnection; ISO's reference model for enabling multivendor system to interoperate. The OSI protocols are a subset of the larger set of standard protocols established by the ISO.

> *GOSIP:* Government OSI profile; a list of OSI protocols specified by the Federal government for use in Federal networks.

Open systems: A computer application environment based on products from a vendor compatible with those available from its competitors, usually because they mutually agree on standards (see Section III.C).

Quick-look data: The subset of the instrument science data stream for which access and processing is necessary earlier than that normally provided. These data may be used for instrument health or safety monitoring (see Section II.B).

Virtual circuit: A network model in which data is received from the transport layer and delivered over the perfect connection: no errors, no duplicates, and all packets are delivered in sequence (see Section III.D).

Virtual reality: An environment in which computer and peripheral devices create a complete sensory environment dynamically controlled by the user (see Section III.E).

References

[1] J. H. Yuen, ed., *Deep Space Telecommunications Systems Engineering*, JPL Publication 82-76, Jet Propulsion Laboratory, Pasadena, California, July 1982.

[2] B. Boehm, "A Spiral Model of Software Development and Enhancement," *Computer*, vol. 21, no. 5, pp. 61–72, May 1988.

[3] P. S. Henry, "High-Capacity Lightwave LAN's," *IEEE Communications*, vol. 27, no. 10, p. 21, October 1989.

[4] U. S. Department of Commerce, Federal Information Processing Standards, FIPS Publication 146-1, Washington, DC, August 1988.

Generalized Probability Model for Calculation of Interference to the Deep Space Network Due to Circularly Earth-Orbiting Satellites

C. J. Ruggier
Telecommunications Systems Section

The probability of exceeding interference power levels and the duration of interference at the Deep Space Network (DSN) antenna is calculated parametrically when the state vector of an Earth-orbiting satellite over the DSN station view area is not known. A conditional probability distribution function is derived, transformed, and then convolved with the interference signal uncertainties to yield the probability distribution of interference at any given instant during the orbiter's mission period. The analysis is applicable to orbiting satellites having circular orbits with known altitude and inclination angle.

I. Introduction

Knowing the orbital parameters and transmitting power level of a potentially interfering satellite is prerequisite to estimating the received interference power level incident at any DSN tracking antenna. However, there are occasions when the satellite launch and orbit injection data are uncertain or unavailable at the time of the analysis. In this case, the satellite orbital position relative to the ground station antenna beam cannot be predicted using the usual deterministic analytical methods: A probabilistic approach to the analysis is more practicable. Reasonable probabilities of interference can be calculated if the geometrical and signal parameter uncertainties are defined for the orbiting satellite. The calculated probabilities can then be useful in determining whether any significant interference threats to the DSN will exist during the orbiting satellite mission period.

The probability of interference is sought from the probability of the orbiting satellite being incident at some point in the antenna beam, for some DSN tracking antenna pointing angle, during the period when the satellite is in view. This conditional probability function is derived from the geometrical relationships of the satellite's orbit. From the conditional probability, the off-beam antenna gain probability can be determined and then convolved with the interference signal transmission uncertainties to yield the probability of received interference power. Finally, the probability of received interference power is scaled by the ratio of view area contained within the satel-

lite's ground track coverage area and the probability of tracking time of the DSN antenna during a 24-hour period. The probability of the DSN station exceeding a given interference power level (density) at any given time can then be determined. In addition to the probability of interference, the period of the interference signal and its probability of occurrence is also considered, given that the interference event has occurred.

II. Rationale for the Interference Model

Figure A-1 (Appendix A) shows the geometrical relationships involved in the derivation of the probability model. It is assumed that the interfering satellite trajectory over the DSN station view area is constrained on the surface of a spherical sector with radius (R_e+h) and swept out by a given antenna off-beam angle, α. R_e is the mean Earth radius, and h is the altitude of the orbiting satellite above ground. Furthermore, the uncertainty of the satellite's orbiting position is equivalent to the satellite entering the view area randomly with uniformly distributed probability.

The probability model is derived from both the random and periodic variables which interplay during the encounter between the DSN station antenna beam and the orbiting satellite. There exist two fundamental random variables: the receiving antenna pointing angle and the satellite's position relative to the antenna off-beam angle. The antenna pointing angle variable can be further resolved into the elevation and azimuth components. The azimuth angle is primarily a function of the Earth's diurnal rate motion, and has been assumed to be nonrandom. However, the elevation angle is considered to be random and to have some probability density, relative to the spacecraft trajectory being tracked. Deterministic and periodic variations are associated with the azimuth angle, the angle of inclination of the Earth's axis with respect to the ecliptic (23.44 deg), and the rapid changes in the elevation angle at the rise and set times. Thus, if averaged over long periods of time, the periodic variations can be neglected without affecting the overall derived probabilities.

The circular orbiting system is also assumed to be ergodic, by virtue of the satellite's constant angular velocity in a circular orbit. This property enables one to transform the satellite's time-averaged periods within the view area into probabilities; these probabilities can then be used to scale the derived probability distribution function. Similarly, DSN spacecraft tracking period statistics can be utilized as additional probability scaling factors.

III. Statistical Interference Model

The probability density of the satellite interference signal power P_r at the DSN antenna is defined as

$$P(p_r) = \left[P(p_t) * P(g_a) * P(-L)\right] P_o P_V \qquad (1)$$

where * denotes convolution and

$P(p_t) \triangleq$ probability density function of the satellite transmit power (dBm), where the satellite antenna is assumed to have isotropic gain

$P(g_a) \triangleq$ probability density function of the DSN tracking off-beam antenna gain (dBi)

$P(-L) \triangleq$ probability density function of free space loss (−dB)

$P_o \triangleq$ probability of DSN antenna tracking time over a 24-hour period

$P_V \triangleq$ ratio of the view area contained within the total satellite ground track area

$= h/2R(\sin\xi)$

$h \triangleq$ satellite altitude, km

$R \triangleq$ R_e (Earth radius) $+h$, km

$\xi \triangleq$ satellite inclination angle, 0 deg $< \xi \leq$ 90 deg

The probability of exceeding the satellite interference signal power at the DSN antenna can be defined as the integral of $P\ (p_r)$:

$$\mathcal{P}\,(p_r) = \int_{P_r}^{\infty} P(p_r)\, dp_r$$

The probabilities arise from the uncertainty of the satellite's orbital position and transmission parameters. $P(p_t)$ can be represented by a triangular probability density function defined from the transmitter power minimum, maximum, and most probable values; $P(g_a)$ is the off-beam antenna gain probability density which is derived in this analysis; and $P(-L)$ is the uncertainty associated with the free space loss and can be represented by a uniform probability density function, with space loss values corresponding to line-of-sight range between the antenna and the interfering satellite. The aggregate effect of the system uncertainties can then be determined from the convolution of these probability densities. Note that when the

satellite transmit power variability is negligible, the probability density function can be replaced with a Dirac Delta function, which sums the mean value of p_t with the random variable values of $P(g_a)$.

The probability density $P(g_a)$, is the derivative of the probability distribution $\mathcal{P}(g_a)$. $P(g_a)$ is transformed using the International Radio Consultative Committee (CCIR) antenna gain pattern Eq. (2); $P(\alpha)$ represents the probability of the satellite being inside the antenna beam cone, bounded by the off-axis angle α and is derived in Appendix A.

The CCIR generalized antenna gain pattern is given as[1]

$$\left.\begin{aligned} g &= G_{\max},\ \text{dBi (maximum antenna gain)} \\ &\quad 0 \le \alpha < 1\ \text{deg} \\ g &= 32 - 25\log(\alpha),\ \text{dBi},\ 1\ \text{deg} \le \alpha < 48\ \text{deg} \\ g &= -10\ \text{dBi},\ 48\ \text{deg} \le \alpha \le 180\ \text{deg} \end{aligned}\right\} \quad (2)$$

The probability distribution of α is

$$\mathcal{P}(\alpha) = \int_0^{90} \mathcal{P}(\alpha \mid \epsilon)\, P(\epsilon)\, d\epsilon \quad (3)$$

where $\mathcal{P}(\alpha \mid \epsilon)$ is the conditional probability distribution of α, given ϵ, which is the DSN station antenna elevation angle.

$P(\epsilon)$ is the probability density function of the DSN antenna elevation angle. If this probability density function is not known, it can be approximated with a triangular probability density function, similar to that shown in Appendix B.

Figure A-1 in Appendix A shows the geometrical relationships used to determine $\mathcal{P}(\alpha \mid \epsilon)$. It is required to know the probability of finding the satellite anywhere inside the antenna beam bounded by the off-beam angle and the portion of the antenna beam which is in view of the satellite. The antenna beam cuts the satellite's orbital sphere at r_1 and r_2, for any given elevation angle ϵ; this sphere contains orbital surfaces bounded by an ellipse projection whose eccentricity is a function of the antenna pointing angle. Surface asymmetry is also affected by the variable off-beam angle α (representative of the satellite probable position inside the antenna beam cone). For th ease of calculation, first-order ellipses are assumed.

Furthermore, the portion of the antenna beam that blocked by the ground and out of sight of the satellit is a major consideration. Instead of directly calculatin the antenna beam cone surface area blockage, the ratio c the projected beam blockage to the total projected area i calculated and used as a scaling factor.

IV. Interference Period Probability

Given that the interference occurs, it becomes useful t determine the interference period and its associated prob ability. The probability of t seconds of interference for a given interference threshold level, corresponding to som fixed off-beam angle α_o, can be written as

$$\mathcal{P}(t) = \int_0^{\epsilon_{mp}} \left(\frac{p(\epsilon)}{t_{\alpha_o}}\right) d\epsilon = \frac{\epsilon_{mp}}{\epsilon_{max}}\, t_{\alpha_o} \quad (4$$

where

$t_{\alpha_o} \triangleq$ longest interference period at α_o

$\epsilon_{mp} \triangleq$ most probable antenna elevation angle

$\epsilon_{max} \triangleq$ maximum antenna elevation angle

the period t is computed piecewise to also account for the Earth-blocked segment of the antenna beam. For the unblocked view area,

$$t_{\alpha_o} = \frac{s}{V_s} = \cos^{-1}\left(1 + \frac{2r_1 r_2 \cos(2\alpha_o) - {r_1}^2 - {r_2}^2}{2R^2}\right) \times \frac{R}{V_s}, \qquad \alpha_o < \epsilon \le 90\ \text{deg} \quad (5)$$

where s is the longest possible path length (km) traversed by the satellite over the antenna beam cone bounded by α_o, and V_s is the mean orbital velocity of the satellite (km/sec). For the blocked view area,

$$t_{\alpha_o} = \frac{s}{V_s} = \cos^{-1}\left(1 + \frac{2r_1\rho\cos(\epsilon + \alpha_o) - {r_1}^2 - \rho^2}{2R^2}\right) \times \frac{R}{V_s}, \quad 0 \le \epsilon \le \alpha_o,\ \rho \triangleq \sqrt{R^2 - {R_e}^2} \quad (6)$$

[1] The equations apply to systems with antenna gains >48 dBi. Due to the very small probabilities associated with the angle α values between 0 deg and 1 deg, only values equal to or greater than 1 deg are considered in the model.

C-5

where the inverse cosine functions are in radians, and the space loss is averaged over the region bounded by a fixed antenna off-axis angle.

V. Application of Model

An example is now given of the model's application to determine the probability of interference when the satellite orbiting mission period and frequency band coincide with those of a DSN mission. Consider a circularly Earth-orbiting satellite and a DSN station with the following parameters:

Earth radius, R_e	6378 km
Satellite inclination angle, ξ	+75 deg
Satellite altitude above Earth, h	637.8 km
Satellite transmit power spectral density[2]	−26.9 dBm/Hz
Satellite orbital velocity	7.6 km/sec
Satellite/DSN station frequency band	8.0 GHz
DSN antenna diameter	70 m

P_o, the probability of tracking time over a 24-hour period, is assumed to be 1.0; P_V, the ratio of the view area, is calculated as 0.06. It is also assumed that the satellite transmitter is switched on throughout the satellite's Earth-orbiting mission period.

[2] Assume that satellite equivalent isotropic radiated power is at 1 W (0 dBW) and the transmit data bandwidth is 500 kHz.

By applying these specified parameters to the model, the derived $P(g_a)$ can be convolved with $P(p_t)$ and $P(-L)$ to yield the probability of exceeding a given power spectral density level. In this example $P(p_t)$ is treated as a constant and can be convolved as a Dirac Delta function. $P(-L)$ is derived in Appendix B.

Figures 1 and 2 show the probability of exceeding the total received power and power spectral density, respectively, for the above orbital and transmission parameters.

From Fig. 2, it can be noted that the probability of exceeding the CCIR recommended interference threshold level of −190 dBm/Hz is 4.5×10^{-6} or 0.00045 percent.

The interference periods and the probabilities of occurrence of those periods for the most probable antenna elevation angles are given in Table 1.

VI. Conclusion

It has been shown that when the instantaneous position of an Earth-orbiting satellite is unknown, one can determine the probability of interference to a DSN tracking station from the conditional probability of the satellite's signal being within the antenna beam cone, given any elevation angle. The analytical derivation of the interference probability function is made possible entirely from the known orbital geometrical properties and transmission parameters of the orbiting satellite.

The probability model provides an early first-order assessment of the potential interference to the DSN, and enables one to determine whether more comprehensive analysis will be necessary.

Table 1. Interference periods and probabilities.

Most probable elevation angle, deg	Interference period, sec	Probability, percent
15	102.8	0.16
25	65.2	0.42
35	44.1	0.88
45	32.3	1.50
55	25.5	2.40
65	21.5	3.30
75	19.3	4.30

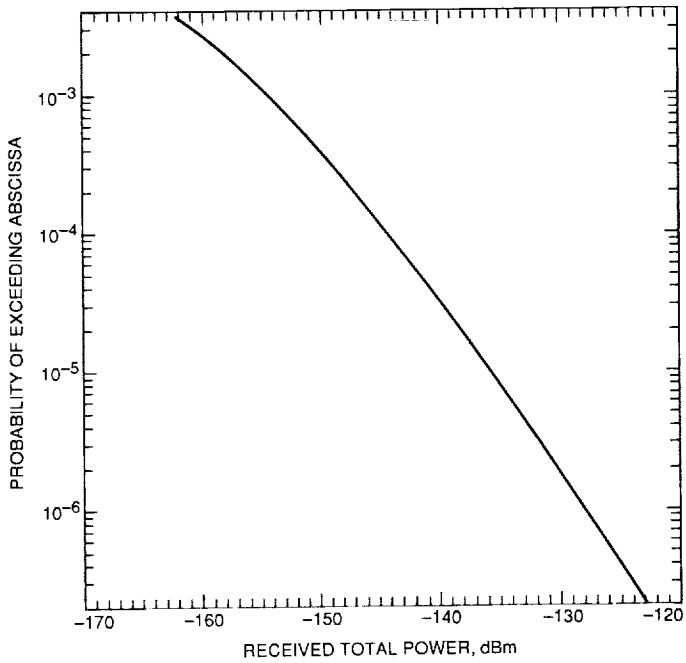


Fig. 1. Probability of exceeding the total received power.

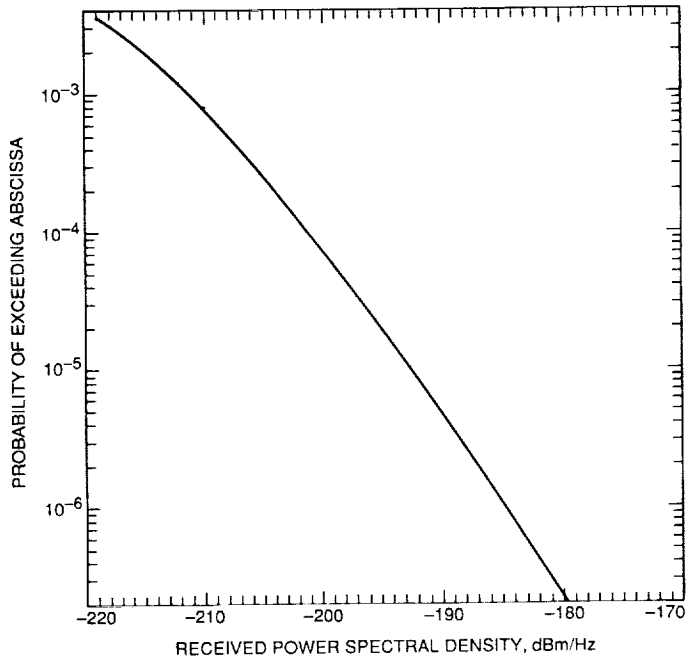


Fig. 2. Probability of exceeding the power spectral density.

Appendix A
Calculation of $P(g_a)$

The projected surface area is given by [1]

$$A = \frac{1}{2}\left[s'r' - c(r' - h')\right] \tag{A-1}$$

where

$$s' = 2r'\cos^{-1}\left(\frac{r' - h'}{r'}\right) \tag{A-2}$$

arc length of projected segment (not shown)

$$r' = r_1 \sin(\alpha) \tag{A-3}$$

radius of projected segment

$$c = \left[4(2h'r' - h'^2)\right]^{1/2} \tag{A-4}$$

chord of projected segment (not shown)

$$h' = \frac{r_1 \sin(\epsilon + \alpha)}{\cos(\epsilon)}$$

$$\text{for } 0 \le \epsilon \le \alpha,$$

$$1 \le \alpha \le 48 \text{ deg} \tag{A-5}$$

Case 1. In this case, $\epsilon \le \alpha$ (occurrence of view area blockage).

The conditional probability distribution of the satellite being in the antenna beam cone bounded by the off-beam angle α for a given elevation angle ϵ can be approximated by expressions for the blocked and unblocked view area. For the blocked view area, the conditional probability distribution $\mathcal{P}(\alpha \mid \epsilon)$ is given as

$$\mathcal{P}(\alpha \mid \epsilon) \approx \frac{S(\alpha, \epsilon) A_p(\alpha, \epsilon) \mid \sec \delta \mid}{2\pi R^2 [1 - R_e/R] A_T}$$

$$\text{for } \epsilon \le \alpha,\ 0 \le \epsilon \le \alpha,$$

$$1 \le \alpha \le 48 \text{ deg} \tag{A-6}$$

where $A_T = \pi ab$ is the total area of the ellipse projected from the orbital surface area bounded by the angle α; a and b are computed from Eqs. (A-12) and (A-13), respectively

$$R \triangleq R_e + h$$

and the visible projected area A_P is determined as

$$A_P = r_1^2 \sin^2(\alpha) \cos^{-1}\left[-\frac{\tan(\epsilon)}{\tan(\alpha)}\right]$$

$$- \left[2r_1^2 \frac{\sin(\alpha)\sin(\epsilon + \alpha)}{\cos(\epsilon)} - \frac{r_1^2 \sin^2(\epsilon + \alpha)}{\cos^2(\epsilon)}\right]^{1/2}$$

$$\times \frac{r_1}{2}\left[\sin(\alpha) - \frac{\sin(\epsilon + \alpha)}{\cos^2(\epsilon)}\right],$$

$$\text{for } 0 \le \epsilon \le \alpha,\ 1 \le \alpha \le 48 \text{ deg} \tag{A-7}$$

Note: The inverse sine function is in radians.

$$|\sec(\delta)| = \left|\frac{2xy}{y^2 + x^2 - z^2}\right| \tag{A-8}$$

where

$$x = \left(r_1^2 + r_2^2 - 2r_1 r_2 \cos 2\alpha\right)^{1/2}$$

$$y = \left[2r_1^2(1 - \cos 2\alpha)\right]^{1/2}$$

$$z = r_2 - r_1$$

and

$$r_1 = -R_e \sin(\epsilon + \alpha) + \left[R^2 - R_e^2 \cos^2(\epsilon + \alpha)\right]^{1/2} \tag{A-9}$$

$$r_2 = -R_e \sin(\epsilon - \alpha) + \left(R^2 - R_e^2 \cos^2(\epsilon - \alpha)\right)^{1/2} \tag{A-10}$$

The surface area covered by the orbiting satellite for a given α and ϵ is the surface area of the sphere (satellite orbital space) cut by the ellipse (antenna beam).

The surface integral $S(\alpha,\epsilon)$ is given as

$$S(\alpha,\epsilon) = 2\pi R^2 - 4R \int_0^{\pi/2} \left[R^2 - \frac{b^2(\alpha,\epsilon)}{M(\alpha,\epsilon)\sin^2\theta + 1}\right]^{1/2} d\theta \tag{A-11}$$

where

$$M = \frac{b^2 - a^2}{a^2}$$

and

$$a^2(\alpha,\epsilon) = \frac{r_1^2}{4}(\alpha,\epsilon) + \frac{r_2^2}{4}(\alpha,\epsilon) - \frac{r_1}{2}(\alpha,\epsilon)\, r_2(\alpha,\epsilon)\cos 2\alpha \tag{A-12}$$

$$b^2(\alpha,\epsilon) = \frac{r_1^2}{2}(\alpha,\epsilon)(1 - \cos 2\alpha) \tag{A-13}$$

By expanding $S(\alpha,\epsilon)$ as a power series [2],

$$S(\alpha,\epsilon) = 2R^2 \int_0^{\pi/2} \frac{b^2/R^2}{(M\sin^2\theta + 1)}\, d\theta + \frac{R^2}{2}\int_0^{\pi/2} \frac{(b^2/R^2)^2}{(M\sin^2\theta+1)^2}\, d\theta + \frac{R^2}{4}\int_0^{\pi/2} \frac{(b^2/R^2)^3}{(M\sin^2\theta+1)^3}\, d\theta + \cdots \tag{A-14}$$

Due to the fast convergence of the series, two terms will adequately approximate $S(\alpha,\epsilon)$ and can be further simplified as

$$S(\alpha,\epsilon) \approx \frac{\pi b^2}{\sqrt{1+M}}\left\{1 + \left(\frac{b^2}{R^2}\right)\frac{2+M}{8(1+M)}\right\} \tag{A-15}$$

Case 2. In this case, $\epsilon > \alpha$.

For the unblocked view area, the conditional probability distribution is given as

$$\mathcal{P}(\alpha \mid \epsilon) = \frac{S(\alpha,\epsilon)}{2\pi R^2\left(1 - \frac{R_e}{R}\right)}, \text{ for } \epsilon > \alpha,$$

$$\alpha < \epsilon \leq 90 \text{ deg}, \; 1 \leq \alpha \leq 48 \text{ deg} \tag{A-16}$$

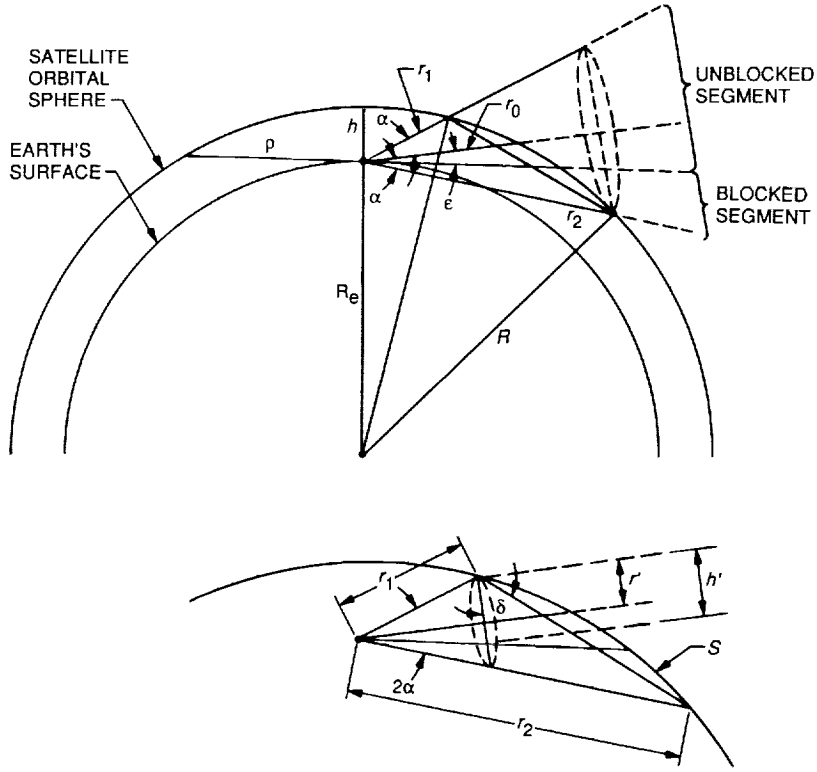
where S(α,ϵ) is given in Eq. (A-15). Then, from Eq. (3),

$$\mathcal{P}(\alpha) = \int_o^{90} \mathcal{P}(\alpha \mid \epsilon)\, p(\epsilon)\, d\epsilon$$

the probability density $P(\alpha)$ is obtained by differentiating $\mathcal{P}(\alpha)$; hence,

$$P(\alpha) = \frac{d}{d\alpha}[\mathcal{P}(\alpha)] \tag{A-17}$$

and $P(\alpha)$ is tranformed into $P(g_a)$ by using Eq. (2).



r_1, r_2 ≜ INTERFERENCE CONE BOUNDS

r_0 ≜ ANTENNA BORESIGHT

ρ ≜ SATELLITE ALTITUDE ABOVE GROUND (MAXIMUM DISTANCE)

h ≜ SATELLITE ALTITUDE ABOVE GROUND (MINIMUM DISTANCE)

R_e ≜ RADIUS OF EARTH

α ≜ ANTENNA OFF-BEAM ANGLE

ε ≜ ANTENNA ELEVATION ANGLE

S ≜ SURFACE AREA BOUNDED BY α

δ ≜ PROJECTION ANGLE

Fig. A-1. Orbital geometry of an Earth-orbiting satellite over a DSN station.

Appendix B

Consideration of $P(p_t)$ and $P(L)$

In the cases where the uncertainty of the satellite transmit power becomes a significant part of the analysis, a triangular probability density function can be used. The satellite transmit power probability density can be represented as [3]

$$P(p_t) = \begin{cases} \dfrac{2(p_t - a)}{(b-a)(c-a)} & \text{if } a \le p_t \le c \\ \dfrac{2(b - p_t)}{(b-a)(b-c)} & \text{if } c < p_t \le b \end{cases}$$

where

a $\triangleq$ minimum power

b $\triangleq$ maximum power

c $\triangleq$ most probable power

The variability of free space loss due to the continuous changes in the satellite's range to the DSN tracking antenna can introduce an additional uncertainty in the received interference power. This uncertainty becomes very significant for lower Earth-orbiting satellites where fluctuations in space loss can be several orders of magnitude.

The probability density function of the free space loss uncertainty $P(L)$ can be modeled as a uniform probability density:

$$P(L) = \begin{cases} \dfrac{1}{b-a} & \text{if } a \le L \le b \\ 0 & \text{otherwise} \end{cases}$$

where

$a \triangleq$ minimum loss $= 20 \log_{10}(4\pi/\lambda) + 20 \log_{10} h$, dB

$b \triangleq$ maximum loss $= 20 \log_{10}(4\pi/\lambda) + 10 \log_{10}(R^2 - R_e^2)$, dB

$\lambda \triangleq$ wavelength of interference signal, meters

$h \triangleq$ satellite altitude above ground, meters

$R_e \triangleq$ Earth's radius, meters

$R \triangleq R_e + h$, meters

References

[1] C. E. Jordan, *Reference Data for Engineers: Radio Electronics, Computer, and Communications*, 7th ed., 3rd printing, Indianapolis, Indiana: Howard W. Sams and Co., 1986.

[2] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, New York: Academic Press, 1980.

[3] A. M. Law and D. Kelton, *Simulation Modeling and Analysis*, New York: McGraw-Hill, 1982.

Ka-Band Feed Arrays for Spacecraft Reflector Antennas With Limited Scan Capability–An Overview

V. Jamnejad
Spacecraft Telecommunications Equipment Section

JPL and NASA are in the process of developing ground and spacecraft antenna systems at Ka-band frequencies for future deep space applications. The use of Ka-band (32-GHz down) communication will result in smaller ground and spacecraft antennas and associated equipment, and will provide larger bandwidths necessary for very high data rate communication and radio navigation. In this article, the use of a small phased array as a feed for a reflector antenna system with limited scan capability is addressed. Different feed and antenna configurations, as well as array architectures, are examined. Some theoretical and experimental parameters of a particular breadboard feed array developed by JPL and the University of Massachusetts are presented. Guidelines for the future direction of this effort are provided.

I. Introduction

JPL and NASA are in the process of developing ground and spacecraft antenna systems at Ka-band frequencies for future deep space mission applications, such as NASA's Solar Probe mission. Ka-band (32-GHz down) communication is planned to first supplement and perhaps ultimately supplant the X-band (8.5-GHz down) communication link presently used. This will result in smaller ground and spacecraft antennas and associated equipment, and will make available larger bandwidths necessary for very high data rate communication and radio navigation.

One crucial aspect of this development effort involves the design and development of advanced high-gain antennas on the spacecraft, which use state-of the-art monolithic microwave integrated circuits (MMIC) technology, solid state power amplifiers, and electronic beam steering. An important design consideration is that of the feed arrays for reflector antennas on the spacecraft. The feed array design depends on feed and reflector antenna architecture and configuration. A proposed method is the near-field Cassegrain (or Gregorian) dual reflector system in which a relatively small phased array is located very close to the subreflector, thus providing a plane wave incident on the subreflector.

The feed array is to be integrated with solid-state digital phase shifters and power amplifiers for transmission

and/or low noise amplifiers for reception. The advantages of using a small phased array in connection with a magnifying dual reflector system is threefold. First, it can provide vernier adjustment of the main reflector antenna beam by scanning the phased array beam incident on the subreflector. Second, the use of individual phase shifters and power amplifiers behind each radiating element means that the circuit losses at the dividers and phase shifters will occur prior to amplification and at very low power levels. Third, by using small solid-state amplifiers at the antenna element level, power from all amplifiers are in effect combined in space as opposed to the approach in which outputs from individual solid state amplifiers are circuit-combined at a substantial loss, before input to the array.

II. Near-Field Dual-Reflector Optics

A directly radiating phased array is usually used to obtain a robust electronically steered beam. To obtain a very narrow high-gain beam, however, a very large aperture array is required. However, a large phased array antenna is not a desirable option on a spacecraft due to its weight and loss, and the complexity of the required beam-forming network (BFN). Therefore, a large aperture reflector antenna is typically used on deep space probes and communication satellites. However, as discussed in [1], in a "happy medium" scenario, a small phased array can be combined with a large main reflector, via the intermediary of one or more small reflectors, to project a large image of the small array on the main reflector aperture.

The scan capability of the system is reduced, however, in inverse proportion to the magnification factor of the system, namely, the ratio of the reflector size to the array aperture size. For a magnification factor of 10, for example, an approximately ± 1-deg reflector beam scan is achieved for a ± 10 deg array beam scan. Thus, this concept of limited scan provides a practical way of increasing the gain of a small array at the expense of scan-angle coverage.

The small, or vernier, scan capability so provided for the high-gain reflector antenna will complement the large-scale antenna beam pointing provided by the spacecraft attitude control system, thus alleviating the need for mechanical vernier adjustments and the concomitant fuel consumption.

Figure 1 depicts two alternative scenarios. In Fig. 1(a), a symmetric confocal dual reflector Cassegrain system is presented in which both the reflector and the subreflector are paraboloids. In this arrangement, the small array is imaged by the subreflector onto the main reflector aperture. However, this arrangement provides a very poo scan capability, since the center of the array and the mai reflector are not conjugate points [1]. Furthermore, th blockage by the feed array can be substantial. The syste in Fig. 1(a) is a simple mechanical structure due to it symmetry and can be used in cases where the array feed is employed not necessarily for the scan ability but for the low-loss spatial power combining that the array provides, as will be discussed later.

In Fig. 1(b), the two confocal paraboloid reflectors are in an offset Gregorian arrangement. In this arrangement, the feed array is placed so that the center points of the main reflector and the feed array are conjugate points. That is, rays originating at one point pass through the other upon reflection from the small subreflector. This arrangement provides for a better scan capability. Furthermore, ideally, the possible phase errors due to small imperfections on the large main reflector can be easily compensated for by the phase variation of the array elements. Also, blockage effects have been completely circumvented. This is the more appropriate arrangement for a limited-scan reflector antenna system.

III. Feed Array Architecture

There are several advantages to using a small phased array in connection with a magnifying dual reflector system. This approach can provide vernier adjustment of the main reflector antenna beam by scanning the phased array beam incident on the subreflector. Furthermore, the use of individual phase shifters and power amplifiers behind each radiating element means that the circuit losses at the dividers and phase shifters will occur prior to amplification and at very low power levels. This active array approach is contrasted with a passive approach in which the amplification is performed before the corporate dividing network and phase-shifting stages. In addition, the active array approach is more reliable and provides for a more graceful degradation of the performance, since the failure of a few active components does not affect the entire array.

Finally, by using small solid-state amplifiers at the antenna element level, the power from all amplifiers are in effect combined in space as opposed to the approach in which output from individual solid state amplifiers are circuit-combined at a substantial loss before input to the array. The spatial combining thus circumvents this network loss. For example, consider the fact that at Ka-band frequencies, the dividing network loss for the 21-element array at 32 GHz, which is discussed below, could be as much as 5 dB and that of a 4-bit digital phase shifter can

be as much as 11 dB; the advantage of the active array approach, in which these losses occur before amplification, is quite evident. Indeed, this is the only realistic approach to obtaining high power transmission (few to tens of watts) using solid state amplifiers with individual output power levels of a few tenths of a watt.

The integration architecture of the amplifiers and the array is another very important aspect of the design. Three approaches to the array architecture are shown in Fig. 2. Each approach has certain advantages and disadvantages, and each is appropriate for certain types of radiating elements. One approach is a so-called brick architecture. The two-dimensional planar array is composed of planar layers. Each layer, in turn, is composed of a subarray of radiating elements. The phase-shifters, amplifiers and radiating elements are integrated on the same layer. Each layer is normal to the face of the planar array (hence the term brick). A two-dimensional array is obtained by stacking several layers of linear subarrays. A divider layer normal to the subarray layers connects them together. This approach is contrasted with a so-called tile approach where radiating elements are located on the planar face of the array in a two-dimensional grid and the amplifiers and phase shifters are arranged on separate layers stacked behind and parallel to the radiating layer. Hybrid approaches are also possible.

IV. Array Development

A two-dimensional Ka-band array has recently been developed at JPL and has been reported in the literature [2–8]. It is a power-combining array intended primarily for transmitter applications. The prototype array is designed for use in future spacecraft antenna applications where Ka-band is planned to supplement the X-band communication link.

The array uses MMIC 4-bit phase shifters and power amplifiers in conjunction with Vivaldi (exponentially tapered slot) elements operating at Ka-band frequency (32 GHz). As shown in Fig. 3, the two-dimensional array is composed of five planar layers (3 inner layers with 5 elements and 2 outer layers with 3 elements, for a total of 21 elements. Figure 4 is a photograph of the array hardware.

The power dividers and the array support frame were designed and fabricated in-house at JPL. The phase-shifters were procured from Honeywell. The very large-scale integration (VLSI) phase-control circuitry is designed by JPL and built at the Defense Advanced Research Projects Agency (DARPA) foundry, and the solid-state power amplifiers (pseudomorphic high-electron mobility transfer) were procured from Texas Instruments.

The design and fabrication of the array of tapered slots (Vivaldi) elements were performed via a contract with the University of Massachusetts [6–7]. Since Vivaldi elements are linearly polarized, an investigation was made to find ways to produce circular polarization. It was concluded that for a Vivaldi array, the most feasible method would be to place of a polarizer screen in front of the array, although other techniques, such as orthogonal and interleaved placement of the array elements, also offer possibilities that need to be further explored in the future.

V. Experimental Results

The 21-way divider fabricated by JPL has a loss of about 5 dB. The 4-bit phase shifters procured commercially have a loss of about 11 dB. As mentioned previously, however, these losses are irrelevant in an active array configuration where power amplifiers are located after the dividers and phase-shifters, right behind the antenna radiating elements. For this reason, the gain of the array will always be referred to a reference plane right after the power amplifiers and at the input to the microstrip feedlines of the antenna elements.

In this arrangement the only significant losses are the microstrip feedline and transition-to-element losses, the cross-polarization loss, and the loss due to amplitude and phase irregularities. These losses determine the efficiency of the array with respect to an ideal lossless antenna with no cross-polarization. The frequency region of primary interest is 31.5 to 32.5 GHz, with 32.0 GHz being the nominal center frequency (wavelength $\lambda = 9.37$ mm). Several different array element lengths and spacing were tried and tested. The achieved gain of the array with an element length of 2.3λ and interelement spacings of 1.22λ, was approximately 23 dB over the frequency region of interest. The average element and array beamwidths in the E and H planes were in the range of 30–40 deg and 7–9 deg, respectively. Even higher gains and narrower beams have been achieved by increasing the length of the elements. Figure 5(a–d) shows the results of an experimental study of the gain variation versus frequency for different array parameters.

Tests have been performed to obtain the far-field patterns and ascertain the scan capabilities of the array. Both individual layers and full two-dimensional arrays have been tested, and co- and cross-polarized patterns have been obtained. Figure 6 presents some typical far-field patterns.

An area of particular concern is the spurious radiation from the beam-forming network which tends to corrupt the sidelobe and cross-polarization regions of the patterns.

Another area of particular interest has been the determination of the phase centers of the array. It is shown that the phase centers of the array in the E- and H-plane are far apart, and this must be taken into account in feeding the reflector antenna. Figure 7 shows the phase variations in the E- and H-planes of the array within the vicinity of the phase centers. Figure 8 shows the experimentally obtained locations of the array phase centers in the two principal planes. There is no single phase center for the array; each cut through the pattern has a different phase center point. This property is not unique to this particular array. Rather, it seems to be a peculiarity of arrays composed of linear traveling-wave radiating elements. In any case, it must be taken into account where the array is used as a focal point feed for a reflector system. This effect, however, may not be significant in cases where it is used as a near-field array feed, and requires further investigation.

In the course of the array development, several array element configurations were designed and investigated. Gain improvements were obtained by proper tapering and increasing the length of the slots. The final configuration with an element length of 3.2λ and interelement spacing of 1.22λ (at 32.0 GHz) produced gain figures in the range of 23.4 to 24.5 dB for the frequency range of 31.5 to 32.5 GHz. The active element gain (i.e., gain in the presence of other elements) was measured in the range of 10.5 to 11.5 dB.

No appreciable improvement, however, was obtained by modifying the microstrip feedline. The measured loss due to the feeding structure (microstrip and feed transition) was about 1 dB. The overall loss due to the microstrip feedline, transition to element, cross-polarization, and amplitude and phase errors ranges from 1.5 to 2.5 dB over the frequencies of interest (56- to 70-percent efficiency). The measured array sidelobes are below 16 dB and the cross polarization level is below 18 dB. The total power lost t the cross-polarization was estimated to be approximatel 0.5 dB.

An attempt was also made to evaluate the mutual cou pling between adjacent array elements. This was done b inverse Fourier transforming the measured far-field pat terns to obtain the near field of the array. Only se lected pattern cuts [one-dimensional fast Fourier trans forms (FFT's)] were used and not for the entire patter (two-dimensional FFT). The results, in this case, have mostly qualitative value. It was shown, however, that th coupling can be significant.

A linear segment of the array has also been tested wit the phase shifters with promising results. Additional test including the phase shifters and amplifiers for the entir array would have to be performed in order to completely ascertain the active array performance.

VI. Conclusions

The design and breadboard of an active transmitting array at Ka-band frequencies and its application to a reflector antenna system have been described. The array is intended as a feed in a near-field Cassegrain or Gregorian dual reflector antenna system. It uses tapered slot (Vivaldi) elements in a brick architecture. The passive array (without the phase shifters and amplifiers) has been tested for RF performance. Its performance is acceptable, but there are some areas where improvements in the performance are desirable and possible. Particular attention must be paid to the losses at the input to the array elements and in the elements themselves. Variation of the array phase center in different azimuth planes is another area of concern and attention. The Vivaldi elements have a linear polarization and do not easily lend themselves to a circularly polarized arrangement. For lower losses and circular polarization, microstrip patch arrays in a tile configuration may prove to be more advantageous.

References

[1] C. Dragone and M. J. Gans, "Imaging Reflector Arrangements to Form a Scanning Beam Using a Small Array—Applicable to Satellite Antenna Design," *Bell System Technical Journal*, vol. 58, no. 2, pp. 501–515, February 1979.

[2] A. L. Riley, D. Rascoe, V. Lubecke, J. Huang, and L. Duffy, "Ka-Band MMIC Beam Steered Transmitter Array," *IEEE 1989 Microwave and Millimeter-Wave Monolithic Circuits Symposium Proceedings*, Long Beach, California, pp. 65–68, June 12–13, 1988.

[3] J. Huang, D. Rascoe, A. L. Riley, V. Lubecke, and L. Duffy, "A Ka-Band MMIC Phased Array Antenna," *1989 IEEE 1989 APS International Symposium Digest*, San Jose, California, pp. 1212–1215, June 26–30, 1989.

[4] D. Rascoe, R. Crist, A. L. Riley, T. Cooley, L. Duffy, D Antos, V. Lubecke, and W. Chew, "Ka-Band MMIC Beam Steered Planar Array Feed," *1990 IEEE MTT-S International Microwave Symposium Digest*, Dallas, Texas, pp. 809–814, May 8–10, 1990.

[5] J. Chang, D. H. Schaubert, K. S. Yngvesson, J. Huang, V. Jamnejad, D. Rascoe, and A. L. Riley, "Ka-Band Power-Combining MMIC Array," *1990 International Conference on Infrared and Millimeter Waves Digest*, Orlando, Florida, pp. 532–537, December 10–14, 1990.

[6] K. S. Yngvesson, Y.-S. Kim, T. L. Korzeniowski, E. L. Kollberg, and J. F. Johansson, "The Tapered Slot Antenna—A New Integrated Element for Millimeter Wave Applications," *IEEE Transactions on Microwave Theory and Techniques*, vol. 37, pp. 365–374, February 1989

[7] J. Chang, D. H. Schaubert, K. S. Yngvesson, J. Huang, V. Jamnejad, D. Rascoe, and A. L. Riley, "32-GHz Power-Combing TSA Array with Limited Sector Scanning," *IEEE 1990 APS International Symposium Digest*, pp. 1150-1153, Dallas, Texas, May 7–11, 1990.

[8] A. L. Riley, D. L. Rascoe, T. Cooley, L. Duffy, V. Jamnejad, and R. Thomas, "A Ka-band MMIC Array Feed Transmitter for Deep Space Applications," *IEEE Microwave and Millimeter-Wave Monolithic Circuits Symposium Proceedings*, Boston, pp. 11–14, June 1991.

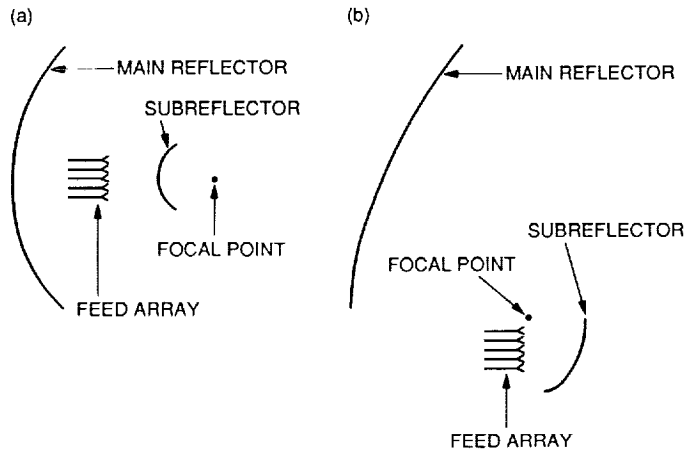


Fig. 1. Near-field array-fed dual reflector systems: (a) A symmetric Cassegrain system and (b) An offset Gregorian system

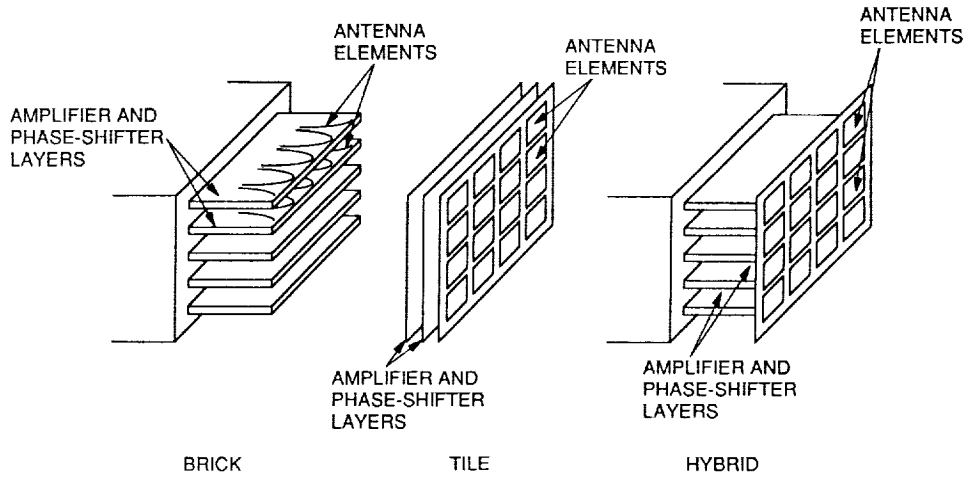


Fig. 2. Various feed array architectures.

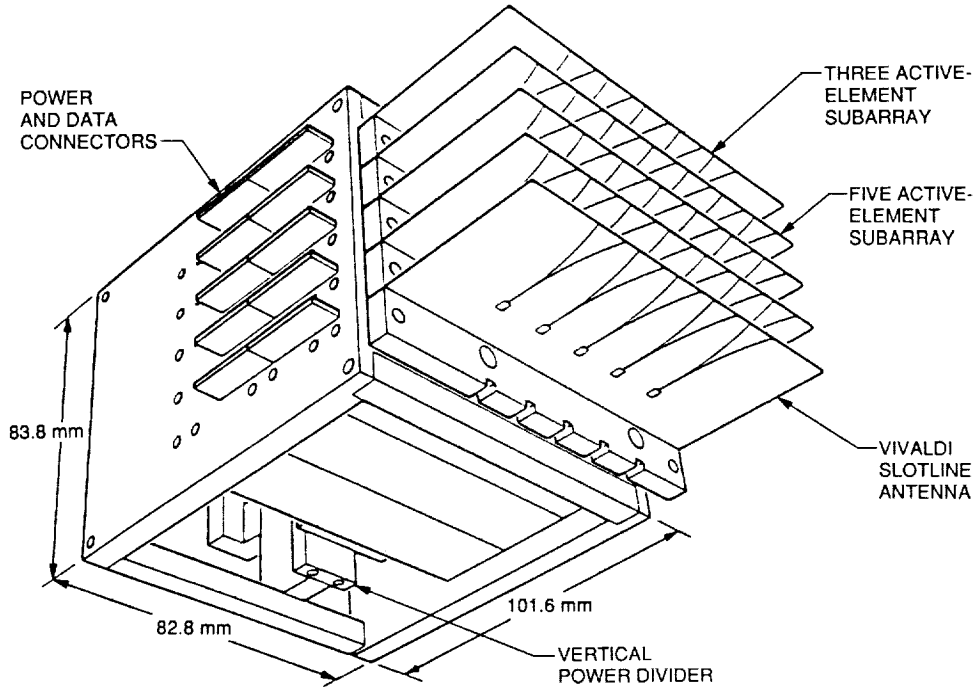


Fig. 3. A 21-element Ka-band Vivaldi array layout.

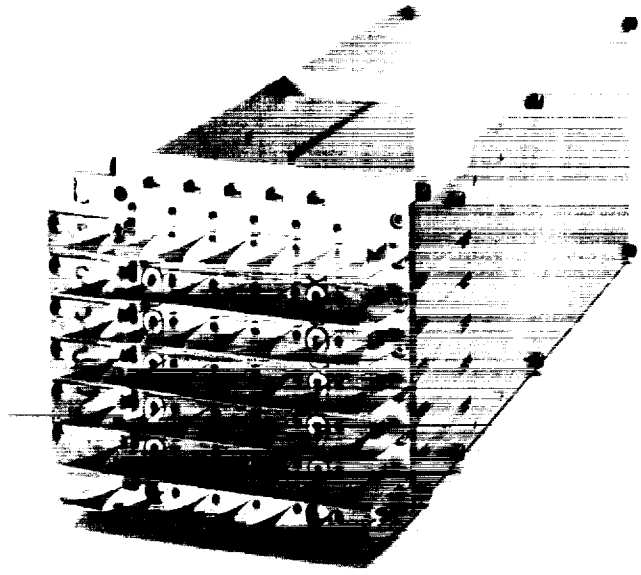


Fig. 4. A view of the 21-element Ka-band Vivaldi array breadboard.

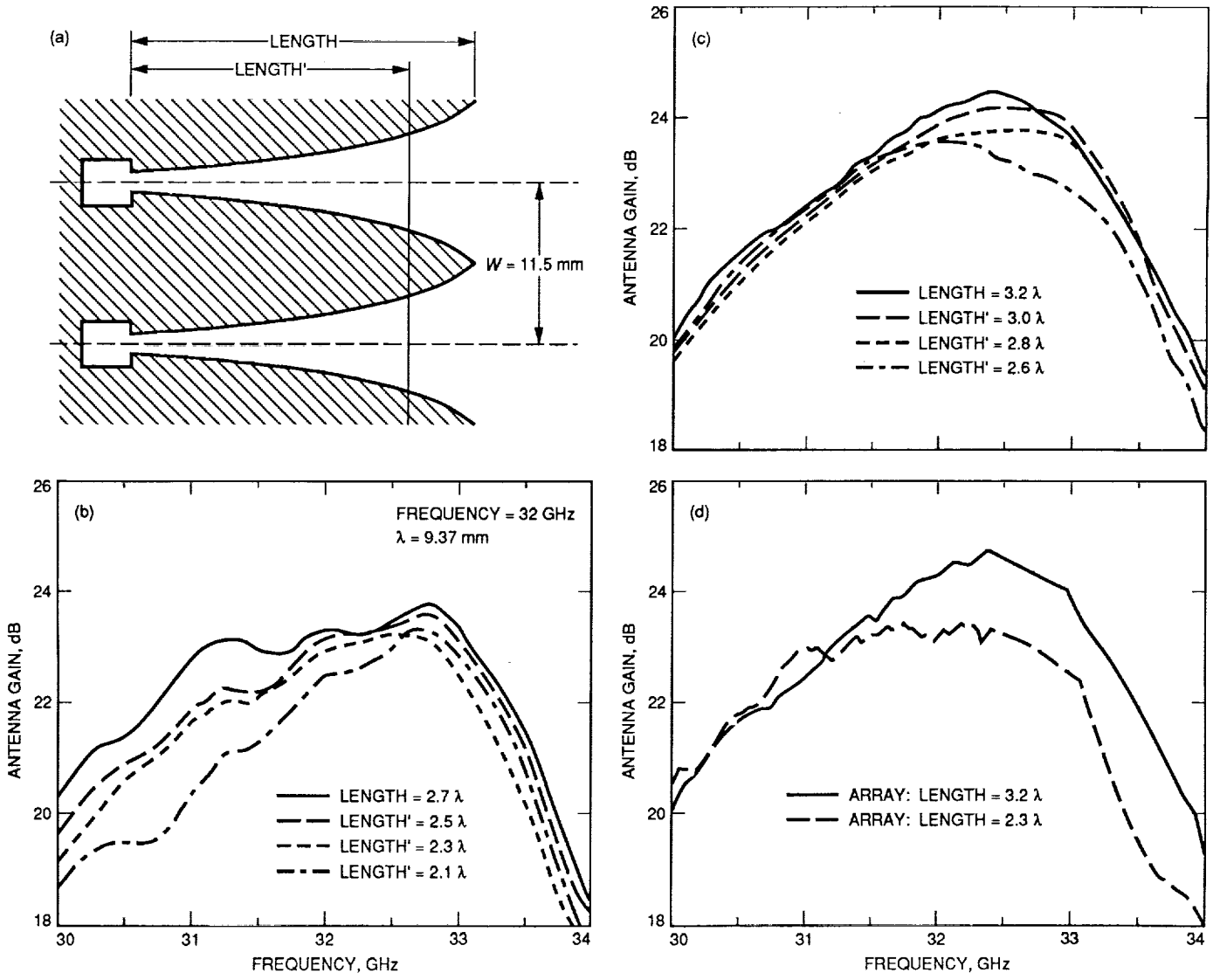
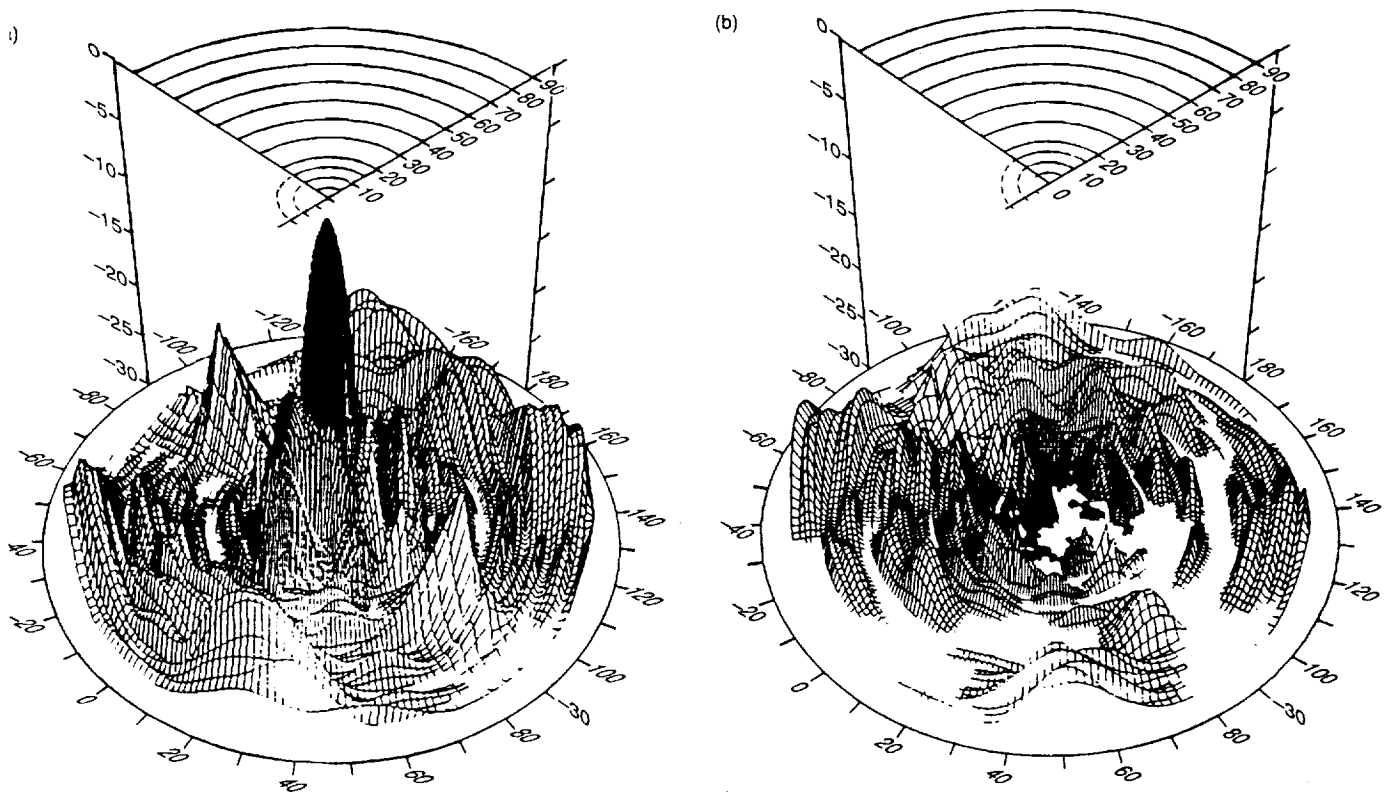


Fig. 5. Parametric study of gain versus frequency: (a) Array parameters; (b) Starting with L = 2.7 and cutting to shorter L' lengths; (c) Starting with L = 3.2 and cutting to shorter L' lengths; and (d) A gain comparison of element lengths of $L = 2.3\lambda$ and $L = 3.2\lambda$.



PEAK GAIN = 23 dB
MAXIMUM SIDELOBE = −16 dB
MAXIMUM CROSS-POLARIZATION = −19 dB

Fig. 6. Typical 3-dimensional far-field patterns of the Vivaldi Array: (a) Copolarized pattern and (b) Cross-polarized pattern.

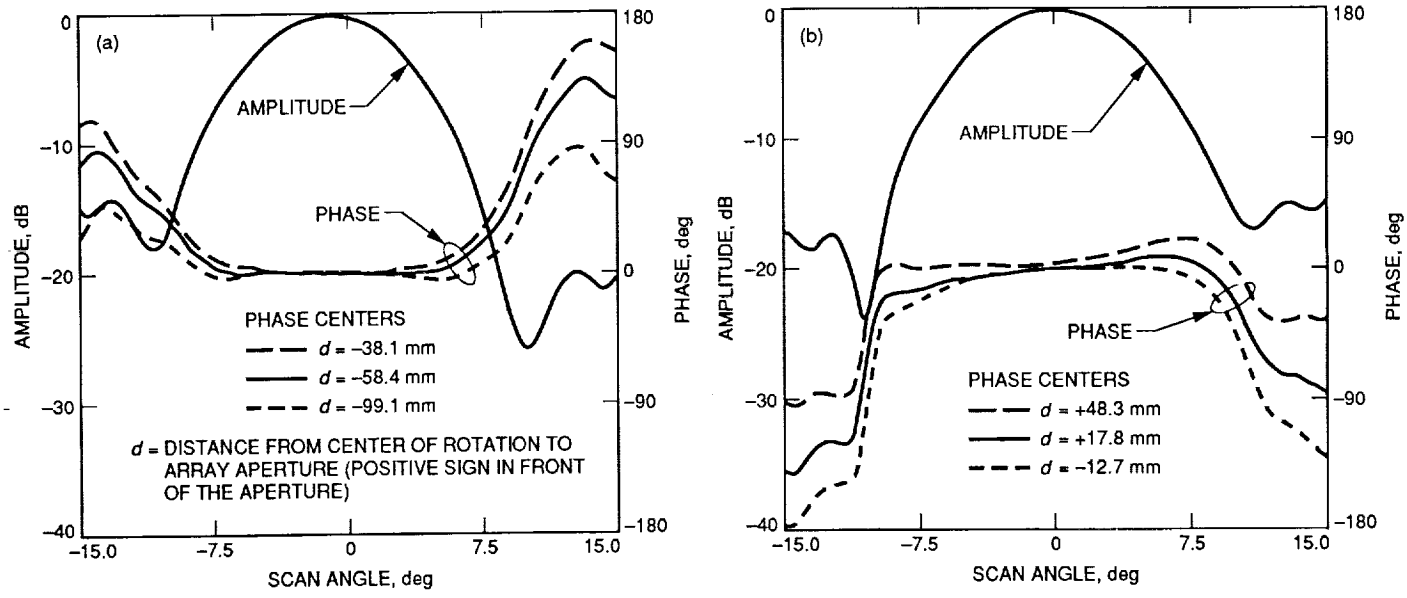


Fig. 7. Pattern phase variations near phase centers in two principal planes: (a) E-plane phase variations and (b) H-plane phase variations.

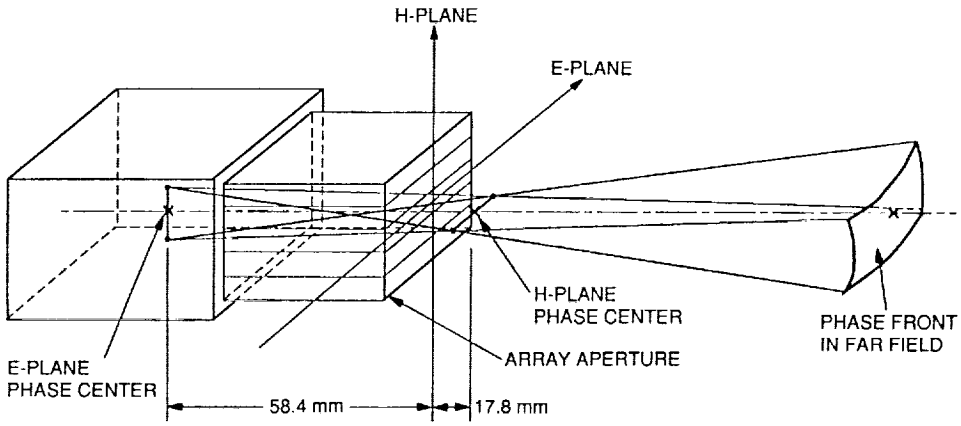


Fig. 8. Locations of the array phase centers in two principal planes.

DSN Observations of Titan

R. M. Goldstein
Telecommunications Science and Engineering Division

R. F. Jurgens
Communications Systems Research Section

When using DSS 14 in a monostatic configuration, radar observations of Titan show that Titan is a diffuse reflector with a relative radar cross section of 0.14 ±0.03. No hot spots were observed.

I. Background

Last summer, there was a favorable opportunity to observe Titan with the Goldstone radar. Detection of radar echoes and the study of their characteristics are expected to be very helpful in the design of the Cassini radar instrument and in planning for its data-acquisition scenarios.

Previous observations by Muhleman, Grossman, Butler, and Slade [1], using the Goldstone transmitter and the Very Large Array receivers, revealed reflectivities of 0.38, 0.78, and 0.25 percent on successive nights. Later measurements have shown that the hot spot may be variable in time and/or location.

A campaign was undertaken to use the upgraded Goldstone radar, in the monostatic mode, to observe such areas on Titan.

II. The Experiment

Sixteen nights at DSS 14 were initially scheduled for this experiment. This would have enabled the examination of each 22-deg time zone of longitude around Titan. An observation night consisted of one 2.5-hour period for transmission followed by an equal period for reception. The signal to noise ratio for a hot spot was expected to be greater than 10 for a single night's observation.

Each observation started with an antenna-pointing test using Saturn as a radiometric source. After verification that Saturn increased the system temperature by 1.6 kelvins, the antenna was swung to nearby Titan and transmission was started.

Frequency hopping was used, so that every 30 seconds the transmitter would jump 10 kHz in frequency. After three hops, the transmitter would return to the original frequency. This technique permits accurate background removal, which is required for detection of weak signals.

During the receive time, the signals were filtered to 40-kHz bandwidth, digitally sampled, Fourier transformed to 39-Hz resolution, detected, and summed. The summed power spectra for each hop period were stored on floppy disks. The above procedure was performed on-line with the newly installed orbital debris equipment.

A daytime test run, using the planet Mercury as a reflector, proved that all systems were working. Even so, all 16 nights of observing time were wasted. The problem, which will be described subsequently, was in the pointing of the antenna.

Because of the importance of Titan data, six additional nights of observations were scheduled. Five of these nights produced useful data. The (weighted) average radar characteristics were as follows:

Radiated power = 468 kW

Antenna diameter = 70 m

Antenna efficiency = 0.68

System temperature = 18.0 kelvins

Range to Titan = 9.22 AU

Wavelength = 3.5 cm

Figure 1 presents the results: one spectrogram for each night of successful observation. The spectrograms have been smoothed to a final resolution of 235 Hz. For clarity, the vertical scales are different for each night. A dotted reference curve shows the echo expected from the average relative cross section (0.35) reported by Muhleman, et al. [1].

The bottom curve of Fig. 1 is the weighted average of the other five observations and represents all the results of the experiment. It constitutes a weak (5.0σ) detection of Titan, and gives a radar relative cross section of 0.14 ± 0.03.

Titan appears to be a diffuse scatterer, with most of the disk contributing to the echo. There was no visible specular component, no central peak in the spectrogram, such as exists for the terrestrial planets. Neither was there any evidence of the sought-after hot spots.

III. Problems

The support given to this experiment by the Goldstone Radio Science group and the DSN personnel was excellent. There were, however, problems encountered whose solution would improve certain hardware and/or software deficiencies.

(1) For 15 of the scheduled nights, the antenna was off-point because the command "PLAN" was absent from the control stream. This command works like "Simon Says"; although the antenna moved away from Saturn when the coordinates were changed, it did not then move to Titan since Simon didn't say to do so. There was no feedback to the experimenters, so there was no way they could know a problem existed.

(2) On one of the remaining nights, the antenna was off-point because the time was entered with four digits instead of six. Once again, since there was no feedback about what the machine was actually doing, no one could know that the problem existed.

(3) On two occasions, the antenna went to brake, causing minor outages.

The Goldstone Solar System Radar would benefit from some provision for the experimenters to monitor the actual pointing of the antenna.

Reference

[1] D. O. Muhleman, A. W. Grossman, B. J. Butler, and M. A. Slade, "Radar Reflectivity of Titan," *Science*, vol. 248, pp. 975–980, May 25, 1990.

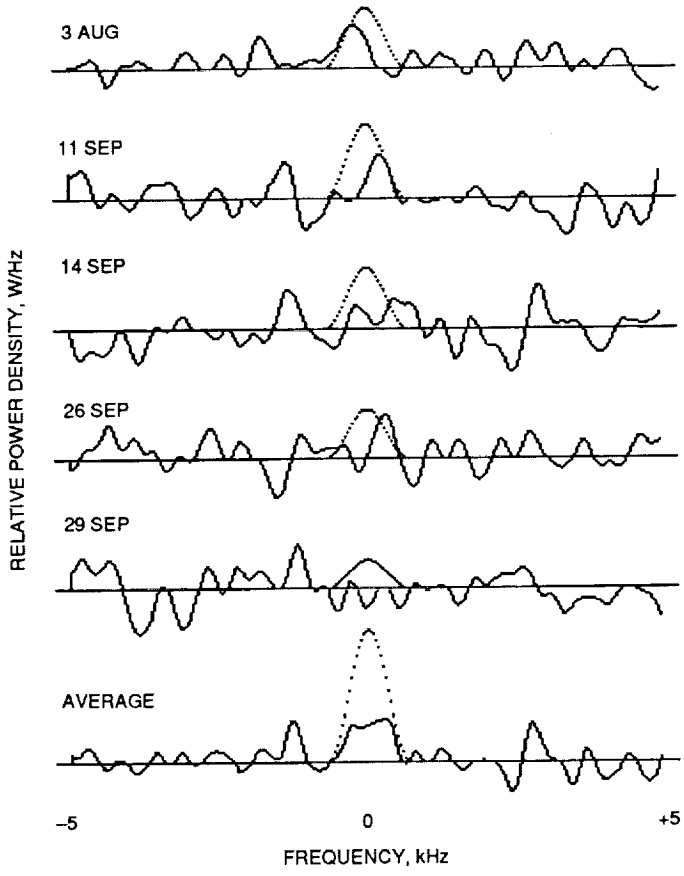


Fig. 1. Spectrograms for each night of successful observation.

Goldstone Solar System Radar Signal Processing

R. Jurgens, E. Satorius, and O. Sanchez
Communications Systems Research Section

A performance analysis of the planetary radar data acquisition system is presented. These results extend previous computer simulation analysis and are facilitated by the development of a simple analytical model that predicts radar system performance over a wide range of operational parameters. The results of this study are useful to both the radar systems designer and the science investigator in establishing operational radar data acquisition parameters which result in the best systems performance for a given set of input conditions.

I. Introduction

In a previous article [1], a system performance analysis of the Goldstone Solar System Radar (GSSR) high-speed data acquisition system was presented. For range-Doppler mapping applications, the system uses a binary phase-coded (BPC) transmitted waveform and the received echoes are complex basebanded, sampled, and demodulated with a replica of the transmitted BPC waveform. The system comprises high-speed digital front-end filters and complex demodulators as well as operator monitor and control, data recording, and real-time signal processing (in a general-purpose, VAX 11/780 host computer equipped with an FPS-5210 array processor). The real-time signal processing activities include the generation of range-Doppler maps.

The analysis in [1] was based on a computer simulation model of the digital baseband data paths. In this article, the previous results are extended with the development of a simple analytical model for the digital baseband data paths. This analytical model is based on a quantized Gaussian approximation for the digitized inputs to the baseband demodulators (correlators/accumulators) and predicts the system processing gain as a function of the various input and operational parameters. The result provides a characterization of system performance that is useful to both the radar system designer as well as the planetary radar science investigator.

In Section II of this article, an overview of the GSSR data acquisition ranging system and its theoretical (ideal) performance is provided, and in Section III the analytical model for the digital baseband data paths is presented. Then, in Section IV, predicted performance analysis results obtained with this model are presented and compared with the results obtained from the computer simulation model. Generally there is good agreement except for those parameter regions where the analytical model is not valid, e.g., in regions where signal clipping distortion occurs at the front-end A/D converters.

II. System Overview

Prior to developing the analytical model for the GSSR data acquisition system, a brief overview of the system architecture as well as a comprehensive review of its theoretical (ideal) performance in terms of the processing gain is presented.

A. System Architecture

As noted in [1], the GSSR data ranging system comprises multiple, complex baseband data channels. The block diagram for a single complex data path is presented in Fig. 1. In this article as in [1], the focus is on just one channel, which would correspond to monostatic, single polarization data collections.

Although limited to a single channel, the results presented here can be used to understand trends in multichannel system performance as a function of the various input system parameters.

To summarize from [1], both in-phase and quadrature channels are low-pass filtered (LPF) (to 6 MHz) and sampled (nominally at 40 MHz). The sampled data streams are then input to pre- and postbaud integrating filters. The prebaud filter integrates five input samples per output data sample and the postbaud filter integrates N_{BAUD} input samples per output data sample (N_{BAUD} is user specified).

The data paths out of the A/D converters are 8 bits wide, but are expanded to 16 bits after the prebaud integrators. A further expansion to 32 bits occurs after multiplication by the scaling constant so that the postbaud integration is carried out with 32-bit integer arithmetic. Following postbaud integration, the data are truncated to N_{BITS} (N_{BITS} is nominally 4) and are then correlated and accumulated to produce a stream of correlation data that are output to the system computer.

B. Ideal System Performance Analysis

To facilitate the presentation of the statistical performance model in Section III and to provide a point of comparison, a comprehensive review is presented here of the ideal (i.e., in the absence of A/D nonlinearities or digitization errors) processing gain for the GSSR data acquisition system. The basic function of the system (prior to the host computer) is to correlate contiguous segments of the received echoes with a replica of the transmitted BPC waveform:

$$\phi_i(\tau) = \int_{iT}^{(i+1)T} dt\; x(t)\; BPC(t+\tau) \qquad (1)$$

where $x(t)$ is the received complex basebanded data (just prior to the A/D converters) and $BPC(t)$ denotes the BPC waveform which is a periodic binary-valued ($BPC = \pm 1$), pseudo-noise (PN) sequence generated by a linear shift register [2]. T is the period of the BPC waveform:

$$T = \Delta T\; N_{CODE}$$

where N_{CODE} denotes the pseudo-noise (PN) code period and ΔT denotes the duration of a single code element (baud).

In carrying out the ideal system performance analysis, it is assumed that the input A/D converters perform only time sampling without amplitude quantization, i.e., the assumption of infinite A/D numerical precision. In this case, the time-sampled (complex) output from the A/D converters is represented simply by

$$x(k) \equiv x\,(kT_s)$$

where $F_s \equiv 1/T_s$ denotes the AD input sampling rate.

With reference to Fig. 1 but assuming that all operations depicted in Fig. 1 are performed with infinite numerical precision,[1] the sampled complex data, $x(k)$, are easily traced through the system. Specifically, at the output of the prebaud integrators (which are typically subsampled by a factor of 5),

$$x_{PRE}(k) = \sum_{n=0}^{4} x(5k-n) \qquad (2)$$

and at the output of the postbaud integrators (which are typically subsampled by a factor of N_{BAUD}),

$$x_{POST}(k) = \sum_{n=0}^{N_{BAUD}-1} x_{PRE}\,(N_{BAUD}k - n)$$

$$= \sum_{n=0}^{5N_{BAUD}-1} x\,(5N_{BAUD}k - n) \qquad (3)$$

Note that the sample rate out of the postbaud integrators is reduced by the factor $5N_{BAUD}$ from the input A/D sample rate, F_s.

Given the discrete-time representation of the complex data from Eq. 3, the correlation function (Eq. (1)) can be written in discretized form as:

[1] In the following, ignore all the operations in Fig. 1 that are connected with finite numerical precision implementation effects, i.e., the scaling multiplication and bit truncation operations.

$$c(i,k) \equiv \phi_i(k\Delta T)$$

$$= \sum_{n=iN_T}^{(i+1)N_T-1} x_{POST}(n) BPC\{5N_{BAUD}T_s n + k\Delta T\} \quad (4)$$

where N_T denotes the number of (decimated) time samples out of the postbaud integrator per PN code period, T. Given that the time interval between samples of $x_{POST}(k)$ is $5N_{BAUD}T_s$ (Eq. (3)), N_T is given by $N_T = T/(5N_{BAUD}T_s)$.

In Eq. (4), the authors tacitly assume that successive samples of the correlation function, $c(i,k-1)$ and $c(i,k)$, are separated in time by a code baud element, ΔT. In practice, this time separation can be a fraction of ΔT simply by controlling N_{BAUD}.

Prior to developing the ideal system performance analysis, a model for the input data, $x(t)$ was acquired. Here the authors assume that the received complex basebanded data are comprised of the received signal and additive noise:

$$x(k) = s(k) + n(k) \quad (5)$$

where $s(k)$ is modeled as a Doppler-shifted and scaled version of the transmitted BPC waveform:

$$s(k) = A\ BPC(kT_s)\, e^{j\theta} e^{2\pi j k \Delta f T_s} \quad (6)$$

A denotes the amplitude scaling (received amplitude level), θ is an arbitrary phase offset (assumed to be a random variable, uniformly distributed between $-\pi$ and π) and Δf denotes the Doppler shift.

Following postbaud integration (from Eqs. (3) and (5)),

$$x_{POST}(k) = \sum_{m=0}^{5N_{BAUD}-1} x(5N_{BAUD}k - m)$$

$$= \sum_{m=0}^{5N_{BAUD}-1} s(5N_{BAUD}k - m)$$

$$+ \sum_{m=0}^{5N_{BAUD}-1} n(5N_{BAUD}k - m)$$

$$\equiv s_{POST}(k) + n_{POST}(k) \quad (7)$$

Assuming that the time interval between samples of $x_{POST}(k)$, i.e., $5N_{BAUD}T_s$, is matched to a baud interval, ΔT (or less), and that the Doppler offset is much less than $1/\Delta T$,

$$s_{POST}(k) \approx 5N_{BAUD}\ s(5N_{BAUD}k)$$

$$= 5N_{BAUD}A \cdot BPC\{5N_{BAUD}kT_s\}\, e^{j\theta} e^{2\pi j 5N_{BAUD}k\Delta f T_s} \quad (8)$$

The signal-to-noise ratio (SNR) output from the postbaud integrators is thus given by

$$SNR_{POST} = \frac{\langle |s_{POST}|^2 \rangle}{\langle |n_{POST}|^2 \rangle} = \frac{\{5N_{BAUD}A\}^2}{\langle |n_{POST}|^2 \rangle} \quad (9)$$

where $< \cdot >$ denotes statistical expectation. From Eq. (7),

$$\langle |n_{POST}|^2 \rangle = \sum_{m,m'=0}^{5N_{BAUD}-1} \langle n(5N_{BAUD}k - m)\ n^*(5N_{BAUD}k - m') \rangle \quad (10)$$

Assuming that the input low-pass filters (see Fig. 1) are ideal and thus that the input power spectrum of the additive noise (just prior to the A/D converters) is constant out to the cutoff frequency, f_{CO}, of the filters,

$$\langle n(5N_{BAUD}k-m)\,n^*(5N_{BAUD}k-m')\rangle = \langle |n|^2\rangle \frac{\sin\{2\pi f_{CO}T_s(m-m')\}}{2\pi f_{CO}T_s(m-m')} \tag{11}$$

Substituting Eq. (11) into Eq. (10),

$$\langle |n_{POST}|^2\rangle = \langle |n|\rangle^2 \sum_{m,m'=0}^{5N_{BAUD}-1} \frac{\sin\{2\pi f_{CO}T_s(m-m')\}}{2\pi f_{CO}T_s(m-m')}$$

$$= 5N_{BAUD}\langle |n|^2\rangle \sum_{m=-(5N_{BAUD}-1)}^{5N_{BAUD}-1} \left(1-\frac{|m|}{5N_{BAUD}}\right) \frac{\sin\{2\pi f_{CO}T_s m\}}{2\pi f_{CO}T_s m} \tag{12}$$

Provided that the low-pass filtered noise coherence is significantly shorter than the combined pre- and postbaud integration time constant, $5N_{BAUD}T_s$, this approximation can be made:

$$OS \equiv \sum_{m=-(5N_{BAUD}-1)}^{5N_{BAUD}-1} \left(1-\frac{|m|}{5N_{BAUD}}\right) \frac{\sin\{2\pi f_{CO}T_s m\}}{2\pi f_{CO}T_s m}$$

$$\approx \sum_{m=-\infty}^{\infty} \frac{\sin\{2\pi f_{CO}T_s m\}}{2\pi f_{CO}T_s m} = \frac{1}{2f_{CO}T_s} \tag{13}$$

Plots of $\{2f_{CO}T_sOS\}$ versus $f_{CO}T_s$, for different values of $L \equiv 5N_{BAUD}-1$, are presented in Fig. 2. As is seen, the approximation in Eq. (13) is accurate to within, at most, a 10-percent error over a wide range of parameter values. As a typical example, consider the nominal system parameters, $f_{CO}T_s = (6\text{ MHz})/(40\text{ MHz}) = 0.15$ and assume that the combined baud integrator time constant is matched to a 1-μsec baud BPC code element, i.e., $5N_{BAUD} = \Delta T/T_s = 1\mu\text{sec} \times 40\text{ MHz} = 40$. Then, it is seen from Fig. 2 that the approximation in Eq. (13) is accurate to within approximately 2.5 percent ($2f_{CO}T_sOS \approx 0.975$), which corresponds to a difference of only 0.1 dB between the actual noise power, $\langle |n_{POST}|^2\rangle$, and its estimate based on the approximation in Eq. (13).

The factor OS is referred to as the "noise oversampling factor," which indicates the degree of dependence of the noise samples (out of the A/D). As f_{CO} approaches Nyquist, $1/2T_s$, there is no oversampling, and $OS = 1$. For $f_{CO} < 1/2T_s$, $OS > 1$, indicating noise oversampling beyond the minimum Nyquist rate. Substituting Eq. (12) (with Eq. (13)) into Eq. (9) yields the following expression for SNR at the output of the postbaud integrators:

$$SNR_{POST} = \frac{\{5N_{BAUD}A\}^2}{\langle |n_{POST}|^2\rangle} \approx \frac{5N_{BAUD}A^2}{OS\langle |n|^2\rangle} \equiv \frac{5N_{BAUD}}{OS} SNR_{IN} \tag{14}$$

where SNR_{IN} denotes the input SNR (referenced to the output of the ideal A/D converters).

To finally characterize the ideal system performance, the discretized correlation function (Eq. (4)) is used along with Eqs. (7) and (8):

$$c(i,k) = \sum_{m=iN_T}^{(i+1)N_T-1} x_{POST}(m)BPC\{5N_{BAUD}mT_s + k\Delta T\}$$

$$= \sum_{m=iN_T}^{(i+1)N_T-1} \left[5N_{BAUD}Ae^{j\theta}e^{2\pi 5N_{BAUD}m\Delta fT_s}BPC\{5N_{BAUD}mT_s\}\right.$$

$$\left.+n_{POST}(m)\right]BPC\{5N_{BAUD}mT_s + k\Delta T\}$$

$$\equiv S_i(k) + N_i(k) \tag{15}$$

where

$$S_i(k) = 5N_{BAUD}Ae^{j\theta}\sum_{m=iN_T}^{(i+1)N_T-1} e^{2\pi j5N_{BAUD}m\Delta fT_s}BPC\{5N_{BAUD}mT_s\}$$

$$\cdot\ BPC\{5N_{BAUD}mT_s + k\Delta T\} \tag{16a}$$

and

$$N_i(k) = \sum_{m=iN_T}^{(i+1)N_T-1} n_{POST}(m)BPC\{5N_{BAUD}mT_s + k\Delta T\} \tag{16b}$$

The system computer transforms the correlation data into range-Doppler maps, $RD(k,m)$, as follows:

$$RD(k,m) = \sum_{i=0}^{N_{BINS}-1} c(i,m)e^{-2\pi jki/N_{BINS}} \tag{17}$$

where N_{BINS} denotes the number of Doppler frequency bins. The maximum value of $RD(k,m)$ occurs when m is such that the local code replica is perfectly aligned with the received code and when $2\pi k/N_{BINS}$ coincides with the Doppler shift of the received signal, subsampled by $\{5N_{BAUD}N_T\}$, i.e.,

$$2\pi k/N_{BINS} = \{2\pi 5N_{BAUD}\Delta fT_s\}\,N_T \equiv \{\Delta\omega\}\,N_T$$

where $\Delta\omega$ denotes the Doppler shift of the received signal normalized by the decimated sample rate at the output of the postbaud integrators. Thus,

$$RD_{MAX} = \sum_{i=0}^{N_{BINS}-1} c(i,0)e^{-j(\Delta\omega N_T)i}$$

$$= \sum_{i=0}^{N_{BINS}-1} S_i(0)e^{-j(\Delta\omega N_T)i} + \sum_{i=0}^{N_{BINS}-1} N_i(0)e^{-j(\Delta\omega N_T)i}$$

$$\equiv S + N \tag{18}$$

By virtue of the definition of $S_i(k)$ (Eq. (16a)),

$$S = 5N_{BAUD}Ae^{j\theta}\sum_{i=0}^{N_{BINS}-1}\left\{\sum_{m=iN_T}^{(i+1)N_T-1}e^{j\Delta\omega m}\right\}e^{-j(\Delta\omega N_T)i}$$

$$= 5N_{BAUD}Ae^{j\theta}\frac{1-e^{j\Delta\omega N_T}}{1-e^{j\Delta\omega}}N_{BINS} \equiv 5N_{BAUD}Ae^{j\theta}\Psi(\Delta\omega; N_T)\,N_{BINS} \tag{19a}$$

where

$$\Psi(\Delta\omega; N_T) = \frac{1-e^{j\Delta\omega N_T}}{1-e^{j\Delta\omega}} \tag{19b}$$

The factor $\Psi(\Delta\omega; N_T)\,N_{BINS}$ represents the coherent integration time constant of the system. When $\Delta f \ll 1/T$, $\Delta\omega N_T \ll 1$, $\Psi(\Delta\omega; N_T) \approx N_T$, and the coherent integration time constant approaches $N_T N_{BINS}$.

To conveniently characterize the system performance, the authors define the system processing gain, which is simply the ratio of output to input SNR. The output SNR (SNR_0) is referenced with respect to the output range-Doppler map. Specifically,

$$SNR_0 = \frac{\langle|S|^2\rangle}{\langle|N|^2\rangle} \tag{20}$$

where (from Eq. (19))

$$\langle|S|^2\rangle = \{5AN_{BAUD}N_{BINS}\}^2\{\sin(\Delta\omega N_T/2)/\sin(\Delta\omega/2)\}^2 \tag{21}$$

and (from Eq. (18)),

$$\langle|N|^2\rangle = \sum_{i,i'=0}^{N_{BINS}-1}\langle N_i(0)N_{i'}^*,(0)\rangle = \sum_{i-0}^{N_{BINS}-1}\langle|N_i(0)|^2\rangle \tag{22}$$

The last equality in Eq. (22) follows from the independence of the noise samples, $N_i(k)$. From Eq. (16b),

$$\langle|N_i(0)|^2\rangle = \sum_{m,m'=iN_T}^{(i+1)N_T-1}BPC\{5N_{BAUD}mT_s\}\,BPC\{5N_{BAUD}m'T_s\}$$

$$\cdot\langle n_{POST}(m)n_{POST}^*(m')\rangle \tag{23}$$

But recall that the coherence of the input noise is sufficiently short that samples of $n_{POST}(m)$ are independent (i.e., the assumption leading to Eq. (13)). Thus, from Eqs. (12), (13), (22), and (23),

$$\langle |N|^2 \rangle = \sum_{i=0}^{N_{BINS}} \langle |N_i(0)|^2 \rangle = \sum_{i=0}^{N_{BINS}-1} \sum_{m=iN_T}^{(i+1)N_T-1} \langle |n_{POST}(m)|^2 \rangle$$

$$= \langle |n|^2 \rangle 5N_{BAUD}OSN_TN_{BINS} \tag{24}$$

and therefore, from Eqs. (21) and (24),

$$SNR_0 = \frac{\{5AN_{BAUD}N_{BINS}\}^2 \{\sin(\Delta\omega N_T/2)/\sin(\Delta\omega/2)\}^2}{\langle |n|^2 \rangle 5N_{BAUD}OSN_TN_{BINS}}$$

$$= \frac{1}{OS}5N_{BAUD}N_{BINS}\left[\frac{\{\sin(\Delta\omega N_T/2)\}^2}{N_T\{\sin(\Delta\omega/2)\}^2}\right]SNR_{IN} \tag{25}$$

Thus, for the ideal system performance gain,

$$PG_{IDEAL} \equiv \frac{SNR_0}{SNR_{IN}} = \frac{1}{OS}5N_{BAUD}N_{BINS}\left[\frac{\{\sin(\Delta\omega N_T/2)\}^2}{N_T\{\sin(\Delta\omega/2)\}^2}\right] \tag{26}$$

As $\Delta\omega$ approaches zero, PG_{IDEAL} can be expressed in terms of the system time-bandwidth product, i.e., from Eq. (13),

$$PG_{IDEAL}(\Delta\omega = 0) = 2f_{CO}T_s\ 5N_{BAUD}N_{BINS}N_T \equiv 2f_{CO}T_{COH} \tag{27}$$

where T_{COH} denotes the total coherent processing time and $(2f_{CO})$ is the system bandwidth. This agrees with the result presented in [1].

III. Statistical System Performance Model

Having provided a general basis for understanding ideal system performance, here is a simple, approximate performance analysis that takes into account finite numerical precision effects. Specifically, with reference to Fig. 1, it is seen that two sources of amplitude quantization errors arise: (1) at the A/D converters and (2) following the postbaud integrators. In the authors' simplified approximate analysis, the A/D amplitude quantization is ignored,[2] but the quantization after the postbaud integrators (at the input to the correlators) is taken into account. Thus, the quantized correlator outputs can be expressed in complex form, as follows:

$$c_Q(i,k) = \sum_{m=iN_T}^{(i+1)N_T-1} Q\left[A'e^{j(\Delta\omega m+\theta)}BPC\{\alpha m\} + n_{POST}(m)\right]BPC\{\alpha m + k\Delta T\} \tag{28}$$

where again $\Delta\omega \equiv 2\pi 5N_{BAUD}\Delta f T_s$, $A' \equiv 5AN_{BAUD}$, $\alpha = 5N_{BAUD}T_s$, and $Q[\cdot]$ denotes the nonlinear quantizer with $Q[x+jy] \equiv Q[x] + jQ[y]$ operation. In the following, it is assumed that $Q(x) = -Q(-x)$.

[2] This is reasonable since the A/D converters are nominally 8 bits wide (compared to nominally 4 bits at the correlator inputs) and the outputs from the A/D converters are smoothed by the pre- and postbaud integrators. Thus, the effects of A/D quantization on system performance (processing gain) are typically much less important than the effects of input correlator quantization.

As discussed in Section II (Eqs. (17) through (20)), system performance can be characterized in terms of the maximum range-Doppler map output, RD_{MAX}. For the case of quantized input correlator data,

$$RD_{MAX} = \sum_{i=0}^{N_{BINS}-1} c_Q(i,0)e^{-j(\Delta\omega N_T)i} \tag{29}$$

and the output SNR is defined as

$$SNR_0 = \frac{\langle |RD_{MAX}|^2 \rangle - \langle |RD_{MAX}|^2_{A'=0} \rangle}{\langle |RD_{MAX}|^2_{A'=0} \rangle} \tag{30}$$

where the subscript $A' = 0$ is used to denote the noise-only input condition.

Maximum map output, $\langle |RD_{MAX}|^2 \rangle$, can be expressed in terms of the statistics of the quantized correlator outputs, $c_Q(i,0)$:

$$\langle |RD_{MAX}|^2 \rangle = \sum_{i=0}^{N_{BINS}-1} \sum_{i'=0}^{N_{BINS}-1} \langle c_Q(i,0)c_Q^*(i',0) \rangle e^{-j(\Delta\omega N_T)(i-i')} \tag{31}$$

To proceed, the notation is simplified by expressing $c_Q(i,0)$ as

$$c_Q(i,0) = \sum_{m=iN_T}^{(i+1)N_T-1} \{Q[x_R(m)] + jQ[x_I(m)]\} BPC\{\alpha m\} \tag{32}$$

where

$$x_R(m) = A' \cos[\Delta\omega m + \theta]BPC(\alpha m) + Re\{n_{POST}(m)\} \tag{33a}$$

and

$$x_I(m) = A' \sin[\Delta\omega m + \theta]BPC(\alpha m) + Im\{n_{POST}(m)\} \tag{33b}$$

Assuming that $Re\{n_{POST}(m)\}$ and $Im\{n_{POST}(m)\}$ are uncorrelated zero-mean Gaussian noise components with

$$\left\langle [Re\{n_{POST}\}]^2 \right\rangle = \left\langle [Im\{n_{POST}\}]^2 \right\rangle \equiv \sigma^2$$

then note that x_R and x_I are independent random variables with conditional Gaussian densities:

$$p_R(x_R|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -[x_R - A'\cos(\Delta\omega m + \theta)BPC(\alpha m)]^2 / 2\sigma^2 \right\} \tag{34a}$$

and

$$p_I(x_I|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\left[x_I - A' \sin(\Delta\omega m + \theta)\, BPC(\alpha m)\right]^2 / 2\sigma^2\right\} \tag{34b}$$

In the subsequent development, the following moment functions, based on p_R and p_I, will be used:

$$\Phi_1(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} Q(u) \exp\left\{-\left[u - x\right]^2 / 2\sigma^2\right\} du \tag{35a}$$

and

$$\Phi_2(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \{Q(u)\}^2 \exp\left\{-\left[u - x\right]^2 / 2\sigma^2\right\} du \tag{35b}$$

By virtue of the odd property of $Q(u)$, it is easy to see that $\Phi_1(x)$ is also odd (with $\Phi_1(0) = 0$) and $\Phi_2(x)$ is even.

Now consider the evaluation of SNR_0. Specifically, from Eq. (32),

$$\begin{aligned}\left\langle c_Q(i,0) c_Q^*(i',0)\right\rangle = \sum_{m=iN_T}^{(i+1)N_T-1} \sum_{m'=i'N_T}^{(i'+1)N_T-1} \{&[\langle Q(x_R(m))\, Q(x_R(m'))\rangle \\ &+ \langle Q(x_I(m))\, Q(x_I(m'))\rangle] \\ &+ j\left[\langle Q(x_I(m))\, Q(x_R(m'))\rangle - \langle Q(x_R(m))\, Q(x_I(m'))\rangle\right]\} BPC(\alpha m) BPC(\alpha m')\end{aligned} \tag{36}$$

In evaluating Eq. (36), two cases are distinguished: $i = i'$ and $i \neq i'$. In the latter case, all terms in the expectations are conditionally independent (with respect to θ) and with Eq. (35),

$$\begin{aligned}\left\langle c_Q(i,0) c_Q^*(i',0)\right\rangle = \sum_{m=iN_T}^{(i+1)N_T-1} \sum_{m'=i'N_T}^{(i'+1)N_T-1} \{&[\langle \Phi_1(A' \cos(\Delta\omega m + \theta) BPC(\alpha m)) \\ &\cdot \Phi_1(A' \cos(\Delta\omega m' + \theta) BPC(\alpha m'))\rangle_\theta \\ &+ \langle \Phi_1(A' \sin(\Delta\omega m + \theta) BPC(\alpha m)) \Phi_1(A' \sin(\Delta\omega m' + \theta) BPC(\alpha m'))\rangle_\theta] \\ &+ j[\langle \Phi_1(A' \sin(\Delta\omega m + \theta) BPC(\alpha m))\, \Phi_1(A' \cos(\Delta\omega m' + \theta) BPC(\alpha m'))\rangle_\theta \\ &- \langle \Phi_1(A' \cos(\Delta\omega m + \theta) BPC(\alpha m)) \Phi_1(A' \sin(\Delta\omega m' + \theta) BPC(\alpha m'))\rangle_\theta]\} \\ &\cdot BPC(\alpha m) \cdot BPC(\alpha m'), \text{ for } i \neq i'\end{aligned} \tag{37}$$

where

$$\langle f(\theta)\rangle_\theta \equiv \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta)d\theta \tag{38}$$

By virtue of the oddness of $\Phi_1(x)$ (and recalling that $BPC = \pm 1$),

$$\Phi_1(xBPC(\alpha m))BPC(\alpha m) = \Phi_1(x), \text{ for any } x$$

Thus, from Eq. (37) and for $i \neq i'$,

$$\begin{aligned}\langle c_Q(i,0)c_Q(i',0)\rangle = \sum_{m=iN_T}^{(i+1)N_T-1}\sum_{m'=iN_T}^{(i'+1)N_T-1} &\{[\langle\Phi_1(A'\cos(\Delta\omega m+\theta))\Phi_1(A'\cos(\Delta\omega m'+\theta))\rangle_\theta \\ &+\langle\Phi_1(A'\sin(\Delta\omega m+\theta))\Phi_1(A'\sin(\Delta\omega m'+\theta))\rangle_\theta] \\ &+j[\langle\Phi_1(A'\sin(\Delta\omega m+\theta))\Phi_1(A'\cos(\Delta\omega m'+\theta))\rangle_\theta \\ &-\langle\Phi_1(A'\cos(\Delta\omega m+\theta))\Phi_1(A'\sin(\Delta\omega m'+\theta))\rangle_\theta]\} \\ \equiv \Gamma(i,i')\end{aligned} \tag{39}$$

When $i = i'$, the sum over m and m' in Eq. (36) is split into two parts: $m = m'$ and $m \neq m'$. The result for $\langle|c_Q(i,0)|^2\rangle$ is given by

$$\langle|c_Q(i,0)|^2\rangle = 2N_T\left(\langle\Phi_2(A'\cos\theta)\rangle_\theta - \langle\{\Phi_1(A'\cos\theta)\}^2\rangle_\theta\right) + \Gamma(i,i) \tag{40}$$

where the property if $f(\theta)$ is periodic in 2π, then $\langle f(\theta+x)\rangle_\theta = \langle f(\theta)\rangle_\theta$ for any x is exploited. From Eqs. (30), (31), (39), and (40),

$$SNR_0 = \frac{K}{2N_TN_{BINS}\,\Phi_2(0)} \tag{41a}$$

where

$$K = 2N_TN_{BINS}\left\{\langle\Phi_2(A'\cos\theta)\rangle_\theta - \Phi_2(0) - \langle\{\Phi_1(A'\cos\theta)\}^2\rangle_\theta\right\} + \sum_{i,i'}\Gamma(i,i')e^{-j(\Delta\omega N_T)(i-i')} \tag{41b}$$

Prior to presenting plots of SNR_0 in the next section, first consider some limiting cases. In particular, note that as $\Delta\omega$ approaches zero, SNR_0 reduces to

$$SNR_0(\Delta\omega = 0) = \{\langle\Phi_2(A'\cos\theta)\rangle_\theta/\Phi_2(0)\} - 1 + \left[\langle\{\Phi_1(A'\cos\theta)\}^2\rangle_\theta/\Phi_2(0)\right]\{N_TN_{BINS} - 1\} \tag{42}$$

As a consistency check, observe that in the limit of no quantization, i.e., $Q(u) = u$, Eq. (42) reduces back to Eq. (25) for the ideal case. Specifically, when $Q(u) = u$,

$$\Phi_1\{x|Q(u) = u\} = x \quad \text{and} \quad \Phi_2\{x|Q(u) = u\} = x^2 + \sigma^2$$

in which case Eq. (42) becomes

$$SNR_0\{\Delta\omega = 0|Q(u) = u\} = \left\{\left\langle A'^2 \cos^2\theta\right\rangle_\theta / \sigma^2\right\} N_T N_{BINS} = \left\{A'^2/2\sigma^2\right\} N_T N_{BINS} \tag{43}$$

Substituting

$$A' = 5N_{BAUD}A;\ \sigma^2 = \left\langle |n_{POST}|^2\right\rangle/2 = (5/2)N_{BAUD}OS\left\langle |n|^2\right\rangle \quad \text{and} \quad SNR_{IN} = A^2/\left\langle |n|^2\right\rangle$$

reveals that Eqs. (43) and (25) (with $\Delta\omega = 0$) are indeed equivalent.

Another interesting limit is that of 1-bit quantization, i.e.,

$$Q(u) = 0.5, u \geq 0 \quad \text{and} \quad Q(u) = -0.5, u < 0 \tag{44}$$

$$\Phi_2(x) = 0.25 \quad \text{(1-bit quantization)} \tag{45}$$

and thus from Eq. (42) (assuming $\Delta\omega = 0$):

$$SNR_0\{\Delta\omega = 0|\text{1-bit quantization}\} = 4\left\langle\left\{\Phi_1(A'\cos\theta)\right\}^2\right\rangle_\theta \{N_T N_{BINS} - 1\}$$

$$\approx 4N_T N_{BINS}\left\langle\left\{\Phi_1(A'\cos\theta)\right\}^2\right\rangle_\theta \tag{46}$$

In the small signal amplitude limit,

$$\Phi_1(A'\ \cos\ \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_0^{A'\cos\theta} e^{-u^2/2\sigma^2} du \approx \frac{A'\ \cos\ \theta}{\sqrt{2\pi\sigma^2}} \quad \text{(1-bit quantization)} \tag{47}$$

and thus Eq. (46) reduces to

$$SNR_0\{\Delta\omega = 0|\text{1-bit quantization}\} \approx 4N_T N_{BINS}\left\langle\left\{\Phi_1(A'\cos\theta)\right\}^2\right\rangle_\theta$$

$$\approx N_T N_{BINS}(2/\pi)\left\{A'^2/2\sigma^2\right\}$$

$$= N_T N_{BINS}(2/\pi)\left\{5N_{BAUD}/OS\right\} SNR_{IN} \tag{48}$$

(small signal amplitude limit)

Comparing Eqs. (48) and (25) (with $\Delta\omega = 0$), it is seen that 1-bit quantization provides a degradation in the small signal processing gain of $10 \log_{10}(2/\pi) \approx -2$ dB over ideal performance. As will be seen in the next section, system processing gain readily approaches ideal performance as the number of bits is increased.

IV. System Processing Gain Evaluation

In evaluating system performance, the system performance gain (defined by Eq. (26) but with SNR_0 computed from Eq. (41)) has been computed as a function of the various system parameters. In carrying out these calculations, assume an $L = 2^B$ level quantizer (B = number of bits) defined as follows:

$$Q(u) = INT\{u\} + 0.5,\ L/2 > u \geq 0;\ \ Q(u) = L/2 - 0.5,\ u \geq L/2;\ \text{ and }\ Q(u) = -Q(-u),\ u < 0$$

where $INT\{x\}$ denotes the integer part of x. An illustrative example of the system performance calculations is presented in Fig. 3 where plots of performance gain (from Eqs. (26) and (41)) versus SNR_{IN} are presented for various baud durations, ΔT, and for both B = 4-bit and 8-bit quantization. For all plots presented in Fig. 3,

$$(1/T_s) = 40 \text{ MHz};\ \ f_{CO} = 6 \text{ MHz};\ N_{BINS} = 64;\ 5N_{BAUD} = \Delta T/T_s;\ \ \Delta f = 0 \text{ Hz}$$

and the PN code period, N_T, is varied such that

$$PG_{IDEAL} = 2f_{CO}T_s 5N_{BAUD}\ N_{BINS}\ N_T \approx 8.2 \times 10^5\ (59 \text{ dB})$$

First, one notes in Fig. 3 the dependence of processing gain on the baud duration—especially for the 4-bit quantization cases and in the low SNR_{IN} regime. Short baud intervals require less postbaud integration (smaller values of N_{BAUD}) resulting in a large input noise fluctuation to the correlator quantizer. Larger noise fluctuation, however, populates a greater number of quantizer levels, thereby approaching the ideal processing gain. Vice versa, longer baud intervals require more integration in the postbaud filter so that the input noise fluctuation to the correlator quantizer is reduced. Fewer quantizer levels are populated and thus the processing gain approaches the 1-bit quantizer limit:

$$PG_{1-BIT} \approx PG_{IDEAL} - 2\text{dB} \approx 57 \text{ dB}$$

Another consequence of long postbaud integration is seen in the high SNR_{IN} region of the curves in Fig. 3. As SNR_{IN} is increased, especially for the longer baud intervals, the input samples to the correlators will saturate the quantizers. When this occurs, SNR_0 will not increase with increasing SNR_{IN}, thereby degrading the system processing gain. For shorter baud intervals, the magnitude of the input samples to the correlators are smaller for a given SNR_{IN}, and thus saturation will not occur except at the highest input signal levels.

Figure 4 presents plots of performance gain (again from Eqs. (26) and (41)) versus input noise level for various baud durations, ΔT, and for both B = 4-bit and 8-bit quantization. For all plots presented in Fig. 4,

$$SNR_{IN} = 10^{-3};\ (1/T_s) = 40 \text{ MHz};\ f_{CO} = 6 \text{ MHz};\ N_{BINS} = 256;\ 5N_{BAUD} = \Delta T/T_s;\ \Delta f = 0 \text{ Hz}$$

and the PN code period, N_T, is varied so that

$$PG_{IDEAL} = 2f_{CO}T_s 5N_{BAUD}N_{BINS}N_T \approx 3.1 \times 10^6 (64.9 \text{ dB})$$

Figure 4 shows that the smallest baud intervals ($\Delta T = 0.5$ μsec) achieve ideal performance sooner (at lower input noise levels) but also result in quantizer saturation sooner—as evidenced by the decrease in system processing gain at the higher noise levels. Vice versa, as the input noise level decreases, system processing gain at both 4-bit and 8-bit quantization reduces back to the 1-bit limit. In any case, from Fig. 4 it may be concluded that ideal system performance can be achieved only over a narrower range of input noise levels when 4-bit quantization is used versus 8-bit.

To further understand system performance, Fig. 5 presents plots of the average noise-only output power, P_{no} (derived from Eq. (40)):

$$P_{no} \equiv N_{BINS} \left\langle |c_Q(i,0)|^2_{A'=0} \right\rangle = 2N_T N_{BINS} \Phi_2(0) \tag{49}$$

versus the input noise level corresponding to both 4- and 8-bit correlator quantization. The other system parameters used in Fig. 5 are

$$(1/T_s) = 40 \text{ MHz};\ f_{CO} = 6 \text{ MHz};\ N_T = 255 \text{ and } 2047 \text{ (span of correlators)};\ N_{BINS} = 256; \text{ and}$$

$$5N_{BAUD} = 160 (N_T = 255) \text{ and } 20 (N_T = 2047)$$

Note that at low input noise levels, P_{no} approaches the 1-bit limit (from Eq. (45)):

$$P_{no} \{\text{1-bit quantization}\} = 0.5 N_T N_{BINS} \approx 45 \text{ dB } (N_T = 255) \text{ or } 54 \text{ dB } (N_T = 2047)$$

As the input noise level increases, more quantizer bits are toggled and P_{no} begins to increase linearly until quantizer saturation occurs. Since the onset of bit toggling occurs at lower input noise levels for 8-bit quantizers, note that the linear region is much broader for the 8-bit plots in Fig. 5. This is consistent with the superior performance of 8-bit correlator quantization as observed in Figs. 3 and 4.

Finally, in Fig. 6, model calculations of P_{no} (Eq. (49)) are compared with the results of computer simulations that utilize the testbed developed in [1]. This simulator models the digital high-speed data acquisition system as depicted in Fig. 1.[3] In Fig. 6 again, plots are presented of P_{no} versus the input noise level for both 4- and 8-bit correlator quantization; these plots correspond to the system parameter values

$$(1/T_s) = 40 \text{ MHz};\ f_{CO} = 6 \text{ MHz}; N_T = 255 \text{ (span of correlators)};\ N_{BINS} = 256; \text{ and } 5N_{BAUD} = 160$$

Note that as the input noise level increases, simulated performance diverges significantly from the model calculations, implying even more performance degradation at the higher input levels than would be predicted from the statistical model. This added degradation results from saturation of the input 8-bit A/D converters, which were not included in the statistical performance model. Thus, the model will provide, in general, only optimistic bounds on system performance.

V. Conclusions

This article has characterized the performance of the planetary radar data acquisition system in terms of its processing gain. It has provided performance bounds based on both the ideal implementation (no quantization) as well as an

[3] The computer simulation program, written in Fortran 77, was originally developed to run on the Radar 331 VAX 11/780 computer (without utilizing the FPS 5210 array processor). However, in generating the results presented here, the simulation software was ported over to the JPL Cray computer (originally the X-MP and now the Y-MP). In conducting the simulations presented here the authors observe that the run times for the Cray are at least 10 to 20 times faster than the corresponding 11/780 run times.

approximate statistical model that takes into account quantization at the correlator inputs. Basically, it is found that system performance is accurately characterized by the 1-bit correlator quantizer model at low input levels and for longer code baud element intervals, ΔT. As the input level increases and/or ΔT decreases, performance improves toward the ideal limit until quantizer saturation occurs (both the correlator quantizer as well as the input A/D converters). At this point, system performance varies rapidly as a function of the input level, from anywhere between the ideal limit to well below the 1-bit limit.

The results of this study are useful to both the radar systems designer as well as the science investigator in establishing operational radar data acquisition parameters which result in the best systems performance for a given set of input conditions.

References

[1] E. Satorius and S. Brokl, "Goldstone Solar System Radar Performance Analysis," *TDA Progress Report 42-93*, vol. January–March 1988, Jet Propulsion Laboratory, Pasadena, California, pp. 302–308, May 15, 1988.

[2] F. MacWilliams and N. Sloan, "Pseudo-Random Sequences and Arrays," *Proc. IEEE*, vol. 64, pp. 1715–1729, 1976.

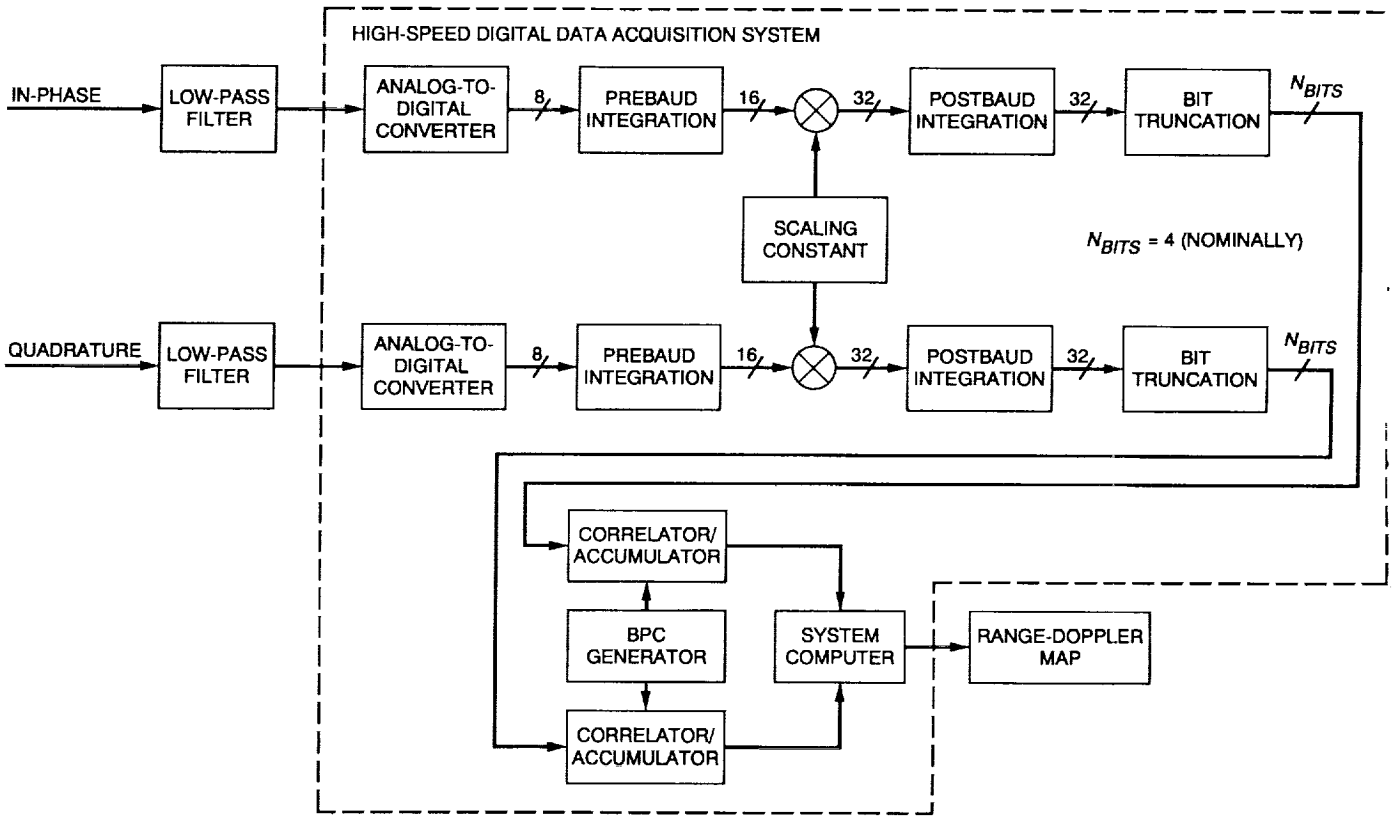


Fig. 1. System block diagram.

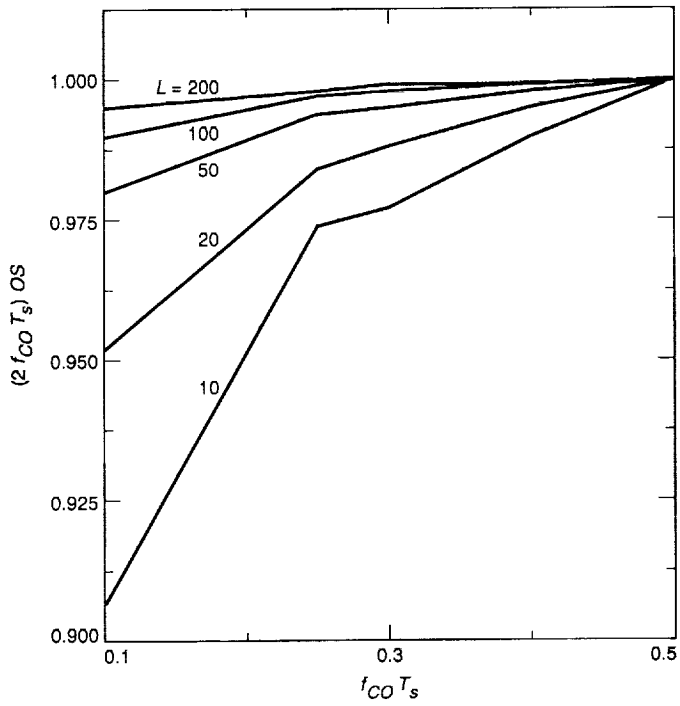


Fig. 2. Dependence of $(2f_{CO}T_s)OS$ versus $f_{CO}T_s$ and $L \equiv 5N_{BAUD} - 1$.

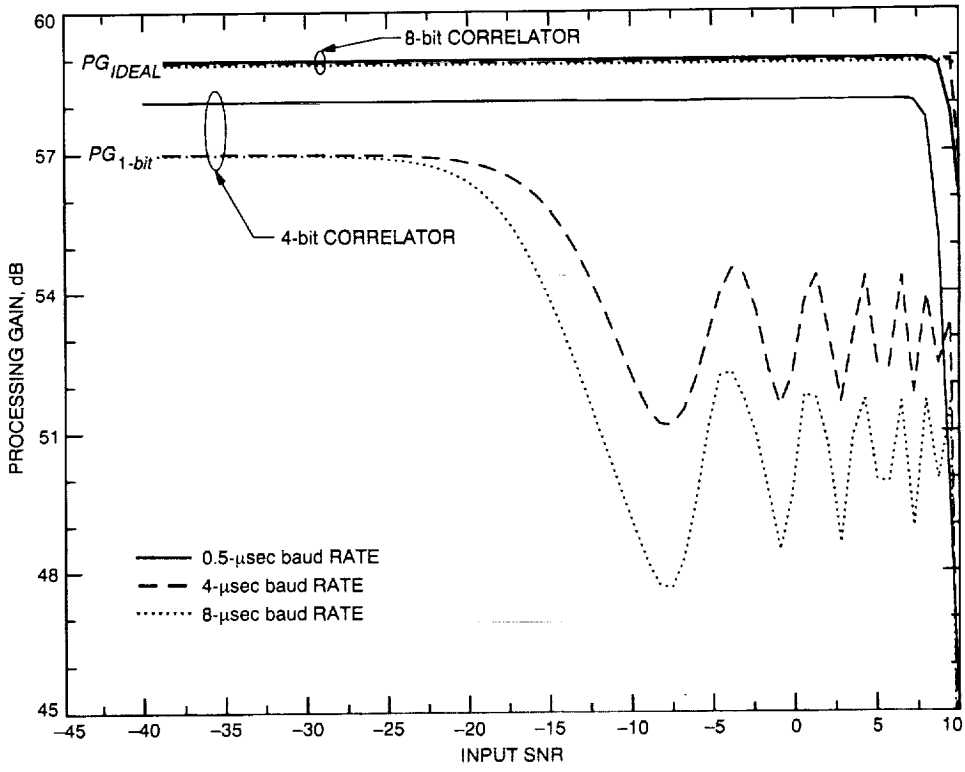


Fig. 3. Processing gain versus input SNR for 4-bit, 8-bit correlator quantization at constant time–bandwidth product (59 dB).

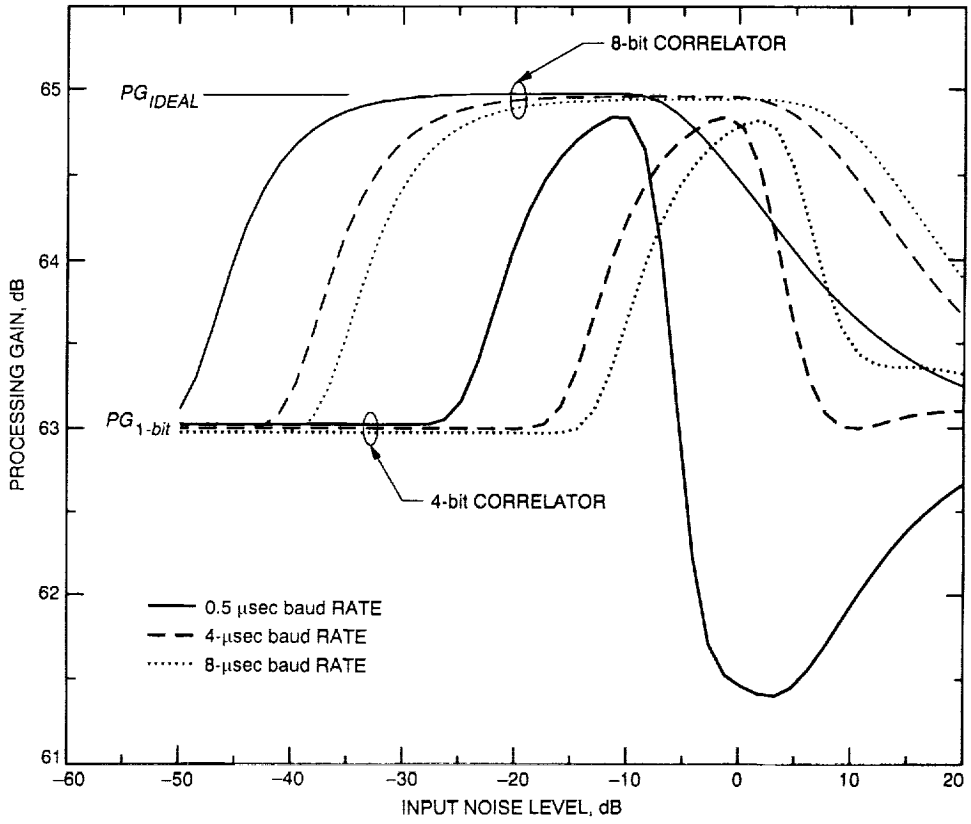


Fig. 4. Processing gain versus input noise level (decibels referenced to 0.3 × A/D full-scale voltage) at constant SNR (−30 dB) and time–bandwidth product (64.9 dB).

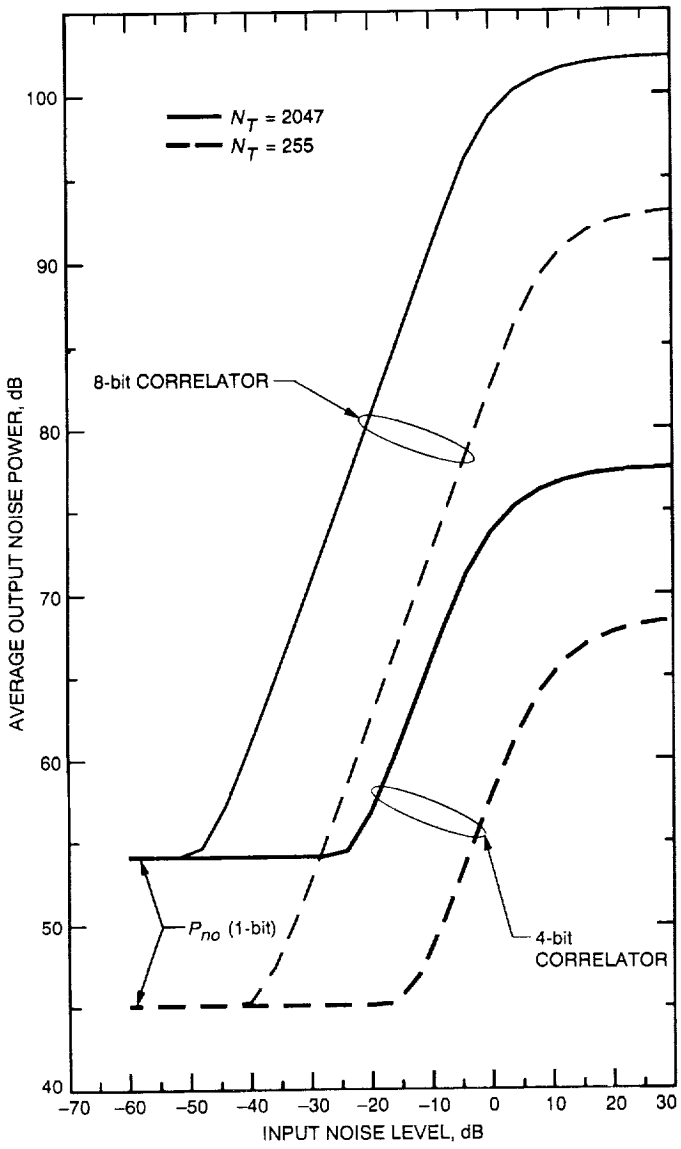


Fig. 5. Output noise versus input noise level (decibels referenced to 0.3 × A/D full-scale voltage) for 4- and 8-bit correlator quantization.

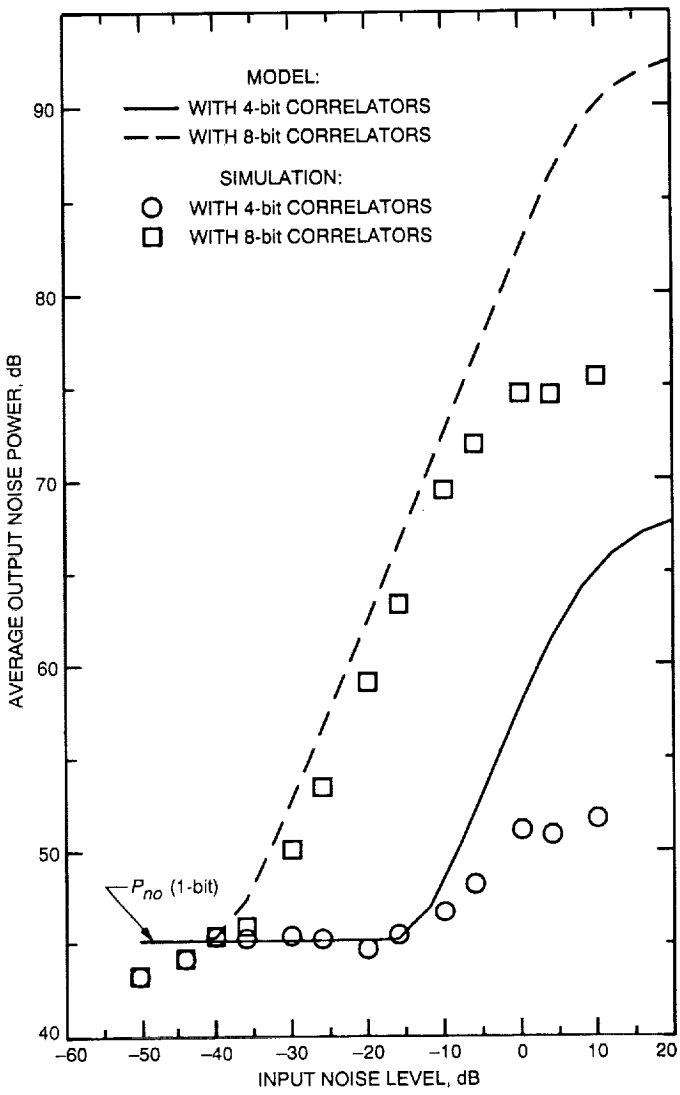


Fig. 6. Output noise versus input noise level (decibels referenced to 0.3 × A/D full-scale voltage) from statistical model and computer simulation.