The Telecommunications and Data Acquisition Progress Report 42-101

January–March 1990

E. C. Posner
Editor

May 15, 1990

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

The Telecommunications and Data Acquisition Progress Report 42-101

January–March 1990

E. C. Posner
Editor

May 15, 1990

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

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

Preface

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

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

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

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

OSO Tasks:
- DSN Advanced Systems
  - Tracking and Ground-Based Navigation
  - Communications, Spacecraft-Ground
  - Station Control and System Technology
  - Network Data Processing and Productivity
- DSN Systems Implementation
  - Capabilities for Existing Projects
  - Capabilities for New Projects
  - New Initiatives
  - Network Upgrade and Sustaining
- DSN Operations
  - Network Operations and Operations Support
  - Mission Interface and Support
  - TDA Program Management and Analysis
- Communications Implementation and Operations
- Data and Information Systems
- Flight-Ground Advanced Engineering

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

DSN Systems Implementation

CAPABILITIES FOR EXISTING PROJECTS

CAPABILITIES FOR NEW PROJECTS

TDA PROGRAM MANAGEMENT AND ANALYSIS

GOLDSTONE SOLAR SYSTEM RADAR

Goldstone Intracomplex Connected Element Interferometry

C. D. Edwards
Tracking Systems and Applications Section

Interferometric observations of the radio source pair 3C 84 and OE 400 have been made on the 21-km baseline between DSS 13 and DSS 15 to explore the angular navigation potential of intracomplex connected element interferometry (CEI). The differential phase-delay observable formed from pairs of 3-minute scans exhibited a precision of 1 psec, while the actual scatter of the phase-delay residuals for eleven scans over the 90-minute observing session was about 10 psec, consistent with the expected few-millimeter fluctuations in the wet tropospheric path delay. Fitting for the position of OE 400 relative to 3C 84 yielded an error ellipse with a semi-minor axis of 60 nrad. Given the short data arc in this experiment, the orthogonal direction in the plane of the sky is not well determined; however, a second baseline or a data arc spanning a larger fraction of the source mutual visibility window could provide simultaneous determination of both right ascension and declination. Examination of the phase-delay residuals supports the accuracy of the cycle ambiguity resolution. However, reliable phase ambiguity resolution will pose the most significant challenge to routine use of CEI for spacecraft tracking, particularly when the a priori spacecraft source position is not well known. Several approaches for ambiguity resolution are briefly outlined.

I. Introduction

Connected element interferometry (CEI) can provide accurate angular tracking from short intracomplex baselines by making use of the very precise phase-delay data type. A number of operational advantages result from being able to form the angular tracking observable from data collected within a single Deep Space Communications Complex (DSCC), including the potential for real-time data processing and fringe verification, as well as near-real-time delivery of tracking observables to the orbit determination navigation software. Some of the motivations for CEI are discussed in [1]. Previous intracomplex observations at Goldstone on the 6-km baseline between DSS 13 and DSS 12 have been reported [1,2].

CEI could provide an efficient angular tracking capability during mission phases where the full precision of intercontinental very long baseline interferometry (VLBI) is not required. A 50–100 nrad intracomplex angular tracking capability on baselines of 20 km would represent an improvement over the current few-hundred-nrad angular accuracy of Doppler tracking. In addition, the CEI observable would not require long, continuous data arcs

and would not suffer from the well-known singularity of the Doppler data type near zero declination. Increasing the CEI baseline lengths to 100 km could further improve angular accuracy to the 10–20 nrad level. Intercontinental VLBI would still provide the highest angular capabilities for specific mission requirements. Current VLBI accuracy is at the 30–50 nrad level for the operational Block I VLBI system,[1] while future systems may deliver nanoradian-level accuracy relative to the quasar frame [3] and 10-prad accuracy for relative angular tracking of two or more spacecraft within the same primary antenna beam [4].

In this article, the first set of differential phase-delay observations on the 21-km baseline between DSS 13 and DSS 15 is presented. This is the longest baseline currently available within any of the DSCCs. DSS 13 is the Deep Space Network (DSN) Advanced Systems Program's 26-m research and development antenna, while DSS 15 is a high-efficiency, 34-m antenna used primarily for operational DSN spacecraft tracking. Differential spacecraft-quasar observations have been simulated by observing a pair of quasars, 3C 84 and OE 400, separated by 6 deg in the plane of the sky. After describing the observing schedule and the data acquisition system, the observable formation will be reviewed. The resulting differential phase-delay residuals will be examined in light of previously reported calculations which suggest that tropospheric fluctuations should dominate the CEI differential phase observable [5].

Weighting the data to reflect the observed scatter, the differential phase-delay residuals will be used to solve for the relative positions of the two quasars. The resulting source position error ellipse characterizes the angular information content of this intracomplex CEI pass. Since the a priori relative source positions are well known and based on many years of high-accuracy VLBI observations, the size of the estimated relative position shift also serves as a consistency check on the CEI results.

Finally, potential problems relating to reliable ambiguity resolution will be discussed, along with a brief outline of several alternative approaches to ambiguity resolution.

II. Experiment Description

Previous intracomplex phase-delay observations at Goldstone have been made on the 6-km baseline between DSS 12 and DSS 13, using a fiber-optic link between those stations to operate them coherently [1,2]. These observations were encouraging in that phase connection was reliably achieved and the final phase-delay residuals were just a few millimeters, but the extremely short baseline limited the angular resolution of these observations to many hundreds of nrad. When the fiber-optic link at Goldstone was extended to SPC-10, where both DSS 14 and DSS 15 are located, a 21-km baseline between DSS 13 and the SPC-10 antennas became available. This longer baseline offers the potential for much greater angular accuracy. With the goals of understanding the limiting phase-delay errors, demonstrating phase ambiguity resolution, and quantifying the angular accuracy on this 21-km baseline, an experiment was planned and scheduled.

Interferometric data were acquired on May 27, 1988, from 2130–2300 UT, at DSS 13 and DSS 15 in the Goldstone DSCC. Observations were made at both 2.3 GHz (S-band) and 8.4 GHz (X-band) to allow calibration of the effects of charged particles. A hydrogen maser at SPC-10 provided the frequency reference for DSS 15; the analog fiber-optic link was used to transfer this frequency reference to DSS 13, allowing the two stations to be operated coherently. Previously reported tests of the fiber-optic link indicate that the link stability $\Delta\nu/\nu$ is well below 10^{-14} for averaging times of several hundred seconds [6].

The Block 0 VLBI data acquisition system was used to single-bit quantize, sample, format, and record the incoming signals at each station [7]. The Block 0 system records a single 2-MHz baseband signal, sampled with 1-bit quantization at the Nyquist rate, resulting in a 4-Mbit/sec data rate. This 2-MHz bandwidth was time-multiplexed over six separate sky frequencies, three at S-band and three at X-band. Each S-band channel was observed for 0.4 sec, while each X-band channel was observed for 1.6 sec, during each 6-second multiplexing cycle. (The longer X-band dwells were chosen because the X-band observable carries a greater weight in the final S/X linear combination.)

A. Frequency Constraints on Short Baselines

Table 1 shows the sky frequencies and channel dwell times used in this experiment. Note the 200-Hz offset between the local oscillator (LO) frequencies at the two stations. This is required due to the small differential Doppler shift between stations, and can be understood as follows: The sidereal rotation of the Earth induces a Doppler shift, for observations in the direction of a given radio source, at each station of an interferometer. For short baseline observations, the differential Doppler shift ν_D between the two stations becomes very small, typically less than 10 Hz

[1] J. B. Thomas, "An Error Analysis for Galileo Angular Position Measurements with the Block I ΔDOR System," JPL Engineering Memorandum 335-26 (internal document), Jet Propulsion Laboratory, Pasadena, California, November 11, 1981.

at 2.3 GHz on a 21-km baseline. Adding a local oscillator offset ν_{OFF} ensures that when the signals from the two stations are cross-correlated (i.e., multiplied), the resulting fringe frequency, $\nu_F = \nu_D + \nu_{OFF}$, will be well separated from dc. When the fringes are subsequently counter-rotated, or "stopped," by multiplying them with a model sinusoid with frequency $\tilde{\nu}_F \approx \nu_F$, both sum and difference frequencies are generated, at roughly $2\nu_F$ and 0, respectively. In the subsequent coherent integration, the high-frequency term will be attenuated if the integration time $\tau \gg 1/(2\nu_F)$, leaving only the stopped fringe phase, as desired. With the coherent integration time of 0.2 sec used at the correlator, the 200-Hz LO shift ensures that this is the case.

Phase calibration tones were injected at each station at both S-band and X-band frequencies to calibrate temporal fluctuations in the instrumental phase. Here again, some modifications to standard VLBI procedure were required due to the short baselines. If the same calibration tone frequencies are used at both stations, the cross-correlation of these phase calibration tones will produce a signal with very nearly the same frequency as the actual quasar fringes, even when an LO offset is used, due to the very small differential Doppler shift between stations on a short baseline. For this reason, different calibration tone frequencies were used at each station in this experiment. The DSN phase calibration system produces calibration tones at integral multiples of $5/N$ MHz, where N is user-selectable in the range 5–99. A value of $N = 7$ was used for DSS 13, and $N = 8$ for DSS 15. This choice placed at least three calibration tones in each 2-MHz channel. All tones in each channel were phase-tracked and used during the correlation process to calibrate instrumental phase errors.

B. Observations and Initial Processing

The experiment observing schedule consisted of repeated differential observations of the two radio sources 3C 84 and OE 400. These sources are separated by about 6 degrees in the plane of the sky. Table 2 summarizes the a priori J2000 positions of the two sources, based on many years of DSN VLBI observations. Each source was observed for 180 sec, with a slew time of 30–45 sec between observations. Both sources were setting during the period of observation; 3C 84 varied in elevation from 57° down to 42°, while OE 400 varied from 53° down to 40°. (On the DSS 13–DSS 15 baseline, the elevation angles at the two stations will always agree to 0.2 deg or better.) Table 3 lists the sequence of observations. One observation pair was lost due to a brief failure of the DSS 13 antenna pointing computer at 215945 GMT. All other scans were successful.

The recorded data, on Block 0 VLBI videocassettes, were processed at the JPL/Caltech Block II VLBI correlator [8]. The output of this initial processing is sine and cosine correlation sums for each of the six channels at a rate of one point per channel every six seconds, corresponding to the time-multiplexing cycle period. These correlation sums were then processed with the REVERT/PHASOR software package to fit for fringe amplitude and phase for each channel. Fringes were obtained for all of the observations for which data were obtained.

With three frequency channels recorded at S-band and three at X-band, it is possible to form group delay observables using the bandwidth synthesis (BWS) technique [9]. Although not a goal of this experiment, the BWS observable was formed as a check on the quality of the data. The BWS observable is formed by calculating the slope of phase versus frequency over a given spanned bandwidth:

$$\tau_{BWS} = \frac{\phi(\nu_1) - \phi(\nu_2)}{\nu_1 - \nu_2}$$

For the channel frequencies used in this experiment, the maximum spanned bandwidth $\nu_1 - \nu_2$ was 40 MHz at both S-band and X-band. The formal error on the X-band BWS delay observable was about 100 psec for 3C 84, and over 200 psec for OE 400, with somewhat higher errors at S-band due to the shorter dwell times. These errors correspond to path-delay errors on the order of 5 cm, or angular errors of several μrad on a 21-km baseline.

III. The Phase Observable and Ambiguity Resolution

To achieve angular accuracies of better than 100 nrad, the much more precise phase data type is required. The geometric component of the interferometric phase can best be thought of as a measure of the geometric delay in units of the observing wavelength. To make use of the interferometric phase, however, one must be able to resolve the integer cycle ambiguity associated with that data type. Further complicating the ambiguity resolution is the contribution of nongeometric errors to the phase, such as the unknown LO phase offset between stations, propagation media delays, and unknown instrumental delays and dispersions.

In principle, the BWS delay residuals contain information about errors in the correlator delay model that could be used to resolve the phase observable on a scan-by-scan basis. However, the formal errors on the BWS delay residuals in this experiment were on the order of an RF cycle (120 psec at X-band) or larger, and thus could

not be used to aid in ambiguity resolution. (It should be pointed out, however, that a wider BWS spanned bandwidth coupled with a larger total recorded bandwidth could provide sufficient BWS precision to enable cycle ambiguity resolution for each scan based on the BWS residual. This more robust approach to ambiguity resolution will be attempted in an experiment planned for early 1990 on this same baseline, using the wider bandwidth Block II VLBI system, which is capable of providing 10-psec BWS precision.)

Instead, for this experiment it was necessary to determine an a priori delay model which was accurate enough to resolve the cycle ambiguity directly. Using the MASTERFIT VLBI parameter estimation software [10] and the best a priori models for station locations and source positions, an (ambiguous) residual phase observable was calculated for each quasar observation. Consider observations of two sources, A and B, and let ϕ_A represent the total phase observable for source A, N_A the unknown integer cycle ambiguity, ω_{RF} the RF observing frequency, and $\tilde{\tau}_A$ the a priori model delay. Then the phase residual $\widehat{\phi}_A$ can be written

$$\widehat{\phi}_A = \phi_A + 2\pi N_A - \omega_{RF}\tilde{\tau}_A$$

A similar phase residual can be formed for the subsequent observation of source B:

$$\widehat{\phi}_B = \phi_B + 2\pi N_B - \omega_{RF}\tilde{\tau}_B$$

If the a priori model were sufficiently accurate, the integer cycle ambiguities N_A and N_B could each be determined by the requirement that the phase residuals be near zero. However, the above-mentioned nongeometric model uncertainties are typically on the order of an RF cycle or more, generally preventing this.

By differencing the two phase observations, many of the model uncertainties are canceled or reduced, thereby allowing the determination of the *relative* cycle ambiguity $N_A - N_B$. Forming a differential observable yields

$$\begin{aligned}\Delta\widehat{\phi}_{A-B} &= (\phi_A - \phi_B) + 2\pi(N_A - N_B) - \omega_{RF}(\tilde{\tau}_A - \tilde{\tau}_B) \\ &\equiv \Delta\phi_{A-B} + 2\pi\Delta N_{A-B} - \omega_{RF}\Delta\tilde{\tau}_{A-B}\end{aligned}$$

Many potential error sources in the delay models $\tilde{\tau}$ are reduced or eliminated by this differencing. The unknown offset of the LO phases between the two stations, which manifests itself as a clock offset for the single-source observable, cancels completely in the differential observable. Geometric errors are also largely reduced by differencing the phase delays for angularly close sources. A baseline uncertainty of size ΔB can cause an error of up to $\Delta B/c$ in the delay model for an individual source, where c is the speed of light. However, when two sources with a small angular separation of $\Delta\theta$ are differenced, much of this geometric error cancels. The resulting error in the differenced model delay will be bounded by $\Delta B\Delta\theta/c$, where $\Delta\theta$ is expressed in radians. For example, a 6-deg angular source separation will cause geometric model errors to be reduced by about a factor of 10 in a differential observable. Thus even a 2-cm baseline error, representing a full half-cycle error for X-band observations, would be reduced to a 2-mm maximum path-delay error for a differential observable between sources separated by 6 deg, and would not complicate the relative phase connection.

Similarly, propagation media errors are greatly reduced by differencing observables for angularly close sources. Any uncertainty in the overall tropospheric delay at each station is greatly attenuated due to several factors: The short baseline causes the total zenith delays at the two stations to be highly correlated; the short baseline also causes the two stations to observe a source at nearly the same elevation angle; and the angular proximity of the sources also causes the elevation angles for the two sources to be nearly the same at each station. The remaining error is predominantly due to small-scale temporal and spatial fluctuations in the tropospheric delay, on the time scale of the time between scans, and on the spatial scale of the angular distance between sources, projected to the tropospheric scale height of several kilometers.

The relative cycle ambiguity ΔN is given by

$$\Delta N = \mathrm{nint}\left(\frac{\omega_{RF}\Delta\tilde{\tau} - \Delta\phi}{2\pi}\right)$$

where nint is the "nearest integer" function, and where the subscript $A - B$ has been dropped. The reliability of the phase ambiguity determination is reflected in the size of the final phase residuals. By the above choice of ΔN, the residuals will lie between $\pm 1/2$ cycle; for the phase connection to be deemed reliable, however, the residuals should be distributed in a peak around zero and be well separated from $\pm 1/2$ cycle.

For this condition to be satisfied, two criteria must be met: both the precision of the differential phase observable $\Delta\phi$ and the error in the a priori differential model delay $\Delta\tilde{\tau}$ must be well below 1/2 cycle of phase. The first criterion is usually satisfied just by the requirement that the sources be detected with good SNR since the statistical phase error, expressed in radians, is roughly 1/SNR.

The second criterion is more difficult to ensure. Given known uncertainties in station location and source position, the geometric component of $\Delta\tilde{\tau}$ can be calculated reliably. However, contributions of unmodeled errors, such as stochastic troposphere fluctuations or antenna deformation, are more difficult to estimate. The approach in this article is to simply determine the phase connection using the best a priori information, and then evaluate the reliability of the phase connection by examining the distribution of phase residuals $\Delta\hat{\phi}$.

Figure 1 shows histograms of the S-band and X-band phase residuals after ambiguity resolution. The residuals show a very clear clustering about zero, with a root-mean-square (rms) value below 1/10th of a cycle for both bands. Quantitatively determining the reliability of the ambiguity resolution requires some assumption about the underlying probability distribution of the phase residual error sources. If it were assumed that the phase residual error was due to a Gaussian error source with a standard deviation of 0.1 cycle, then it would be possible to calculate the likelihood that all of the integer cycle ambiguities had been correctly resolved. The probability that a single observation was correctly resolved would simply be the probability that the observation's phase error was less than 0.5 cycle. For the case of $\sigma = 0.1$ cycle, this is just the integrated Gaussian probability distribution out to $\pm 5\ \sigma$, equal to 0.9999994. For the eleven observations, this gives a cumulative probability of 0.999993 that all eleven observations have been correctly resolved, if it is assumed that the observations are uncorrelated. Unfortunately, the error sources contributing to the observed phase residuals most likely have significant nonGaussian tails that would drastically change this conclusion, increasing the probability of an incorrect integer cycle determination. Obtaining a better characterization of the probability distribution of phase errors, and thus determining more realistically the reliability of cycle ambiguity resolution, is a primary motivation for collecting a much larger database of observations on this intracomplex baseline. At this time, all that can be said is that the distribution of phase residuals after ambiguity resolution strongly suggests that the phase ambiguity resolution was successful for this experiment.

In the final stage of observable formation, the S-band and X-band phase residuals are linearly combined to form an S/X residual phase-delay observable, free from the dispersive effects of charged particles:

$$\Delta\hat{\tau}_{S/X} = \left(\frac{\omega_X^2}{\omega_X^2 - \omega_S^2}\right)\frac{\Delta\hat{\phi}_X}{\omega_X} - \left(\frac{\omega_S^2}{\omega_X^2 - \omega_S^2}\right)\frac{\Delta\hat{\phi}_S}{\omega_S}$$

Figure 2 shows the S/X-corrected differential phase-delay residuals as a function of time for the eleven source pair observations. The rms value of the residuals is 10 psec, or about 3 mm of path delay. The formal statistical error on each point is only about 1 psec; the observed scatter is expected to be due primarily to stochastic spatial and temporal fluctuations of the wet tropospheric path delay at each station. To account for these tropospheric fluctuations, the actual error bars shown on the data points have been inflated to yield a reduced χ^2 equal to 1, as described in Section IV. A previous article [5] calculated the size of such fluctuations and their impact on differential phase-delay observations in CEI, based on numerical integrations of a Kolmogorov turbulence model for atmospheric fluctuations [11]. Using the model described in that article and the relevant parameters of these observations (namely, a 21-km baseline, 6-deg source separation, 3-minute scan and 40-sec slew time, and the range of elevation angles occurring during this experiment) the expected scatter in the differential phase delay is calculated to range from 3.5 to 4.2 mm over the observation period, increasing as the sources set in elevation. The model has some uncertainty due to daily variations in the wind speed and scale height of the wet troposphere. Nonetheless, this excellent level of agreement supports the hypothesis that tropospheric effects are indeed a dominant error source for the CEI differential phase observable.

IV. Relative Source Position Accuracy

The most direct way of demonstrating the angular accuracy of these observations is to solve for the relative angular positions of the two sources, just as one would solve for a spacecraft position relative to a reference quasar. Doing this, one obtains an estimated position correction for one of the quasars, along with a covariance describing the uncertainty in that estimated position. The a priori source positions of both sources in this experiment are known to an accuracy of about 10 nrad, based on many years of observations on long baselines. Thus a consistency check on the estimated position is that it agree, to within the estimated covariance, with the a priori position.

(There may be concern that structure in the radio source brightness distribution could cause the apparent source position to differ on long versus short baselines. Although source structure issues will not be addressed in this article, future CEI observations of a number of source pairs would provide a data set for quantifying this effect. If long- and short-baseline apparent source positions differed significantly, it would be necessary to develop a distinct CEI source position catalog based solely on short-baseline observations.)

The MASTERFIT VLBI software was again used, in this case to estimate a position correction to the source OE 400. As mentioned earlier, the formal statistical errors on the phase-delay data points are only about 1 psec, while the actual scatter of 10 psec is consistent with the expected level of tropospheric fluctuations. To take this into account, an error contribution of about 10 psec was added in quadrature to each differential observation's formal error to obtain the final data weight used in the source position parameter estimation. The size of this supplemental error contribution was determined by requiring the reduced χ^2 of the final fit to equal one. In this way, the actual level of fluctuations observed in the data can be incorporated into the data weights and thereby into the uncertainty in the final source position determination. Given that atmospheric dynamics can vary significantly from day to day, this approach is probably more robust and reliable than trying to specify the data weights based solely on some model of tropospheric fluctuations. Nonetheless, as described earlier, this supplemental phase-delay error contribution, determined empirically, agrees fairly well with the predictions of a statistical model of tropospheric fluctuations.

Using these empirically determined data weights, the eleven differential phase-delay residuals served as input to a weighted least-squares adjustment of the right ascension and declination of OE 400. No other parameters were estimated in this process. The OE 400 source position was essentially unconstrained: the a priori right ascension and declination uncertainties were set at 1 radian. Figure 3 shows the resulting 1-σ error ellipse for the estimated OE 400 position correction. The vertical axis is the declination shift $\Delta\delta$, while the horizontal axis represents the right ascension shift expressed as an arclength: $\Delta\alpha \times \cos\delta$. The error ellipse is very elongated, with a semi-major axis of about 1165 nrad, due to the short data arc. However, in the direction of the average baseline projection during the experiment, the semi-minor axis is 60 nrad. In addition, the 1-σ error ellipse is consistent with the a priori source position, which in this experiment was well known.

These results represent the first time that angular measurements from a single DSN complex have provided accuracies below 100 nrad. The high eccentricity of $\epsilon = 0.90$ for the error ellipse derived from this data set reflects the short span of data. Each single phase-delay observation only contains information for one direction in the plane of the sky, namely, along the projection of the interferometer baseline in the radio source direction. Over the 90-minute duration of this experiment, the baseline projection in the direction toward OE 400 rotates only a small amount, about 10°. Because the experiment duration was so short, the full range of baseline orientation was not sampled, and thus the resulting source position will have a much better determination in one direction—namely, the projected baseline direction at the central epoch of the experiment, with a much larger uncertainty in the orthogonal direction.

To illustrate this, Fig. 4 shows the east–west and north–south components of the projected DSS 13–DSS 15 baseline in the direction of OE 400, over the full 15.5-hour mutual visibility window of this northern declination source. (This can be thought of as the baseline length and orientation, as viewed from the quasar.) The 90-minute observation period represented by the current data set is indicated by the heavy line. The limited range of baseline orientation during the 90-minute observation period underlies the large eccentricity of the estimated OE 400 position error ellipse in this experiment. But over the entire mutual visibility window, the baseline rotates through well over 90°, and hence CEI observations spanning this entire period would provide strong solutions for both components of sky position. Previous analyses[2,3] have shown that CEI observations at Goldstone on the DSS 13–DSS 15 baseline, collected throughout the mutual visibility period, can provide good determination of both α and δ for northern declination sources.

Adding a second baseline, orthogonal to the DSS 13–DSS 15 baseline, would also enable good determination of both components of sky position. This solution has the added advantage that both components could be determined simultaneously from observations taken at a single epoch, instead of combining observations that are separated by many hours—an important advantage in the case of tracking a spacecraft with a poorly determined orbit.

V. Reliability of Ambiguity Resolution

Successful cycle ambiguity resolution is the key to unlocking the high precision of the phase data type. Much more work needs to be done to understand the reliability of ambiguity resolution as a function of baseline length, angular source separation, elevation angle, and temporal scan separation. Nonetheless, one encouraging conclusion

[2] M. H. Finger and C. D. Edwards, "Relative CEI Navigation Performance of Goldstone Intracomplex Baselines," JPL Interoffice Memorandum 335.3-88-116 (internal document), Jet Propulsion Laboratory, Pasadena, California, October 20, 1988.

[3] S. W. Thurman, "Information Content of a Single Pass of Phase-Delay Data From a Short Baseline Connected Element Interferometer," JPL Engineering Memorandum 314-479 (internal document), Jet Propulsion Laboratory, Pasadena, California, December 13, 1989.

of this analysis is that differential phase-delay errors do not seem to grow linearly with baseline length. Limited data on a 6-km, a 21-km, and a 253-km baseline are now available, and in each case, the final delay residual scatter has been at about the 10-psec level. This is consistent with the hypothesis that the scatter is dominated at each site by rapid temporal fluctuations, which are uncorrelated between stations on the short time scale of a differential observation, and therefore are independent of baseline length. In each of these experiments, the a priori source positions were well known; thus the ambiguity resolution was only limited by the size of unmodeled components of the differential delay, particularly these rapid tropospheric fluctuations. At S-band and X-band, this 10-psec level of unmodeled phase-delay uncertainty does not prevent ambiguity resolution. However, at 32 GHz (Ka-band), the delay ambiguity is only about 30 psec. These same errors would then represent a much larger fraction of an RF cycle, seriously complicating phase connection.

Ambiguity resolution also becomes more difficult when the a priori source position of one of the sources is less well known. In the "direct" ambiguity resolution strategy used in this article, in which the integer cycle ambiguity is determined solely on the a priori delay model, one clearly must know the source position in the projected baseline direction to better than half of a fringe spacing. On a 21-km baseline, the minimum fringe spacing is 6.2 μrad at 2.3 GHz and 1.7 μrad at 8.4 GHz. Doppler tracking can usually determine a spacecraft ephemeris with sufficient accuracy to satisfy this constraint during periods of cruise. However, at encounter or during other periods of high spacecraft dynamics, this constraint may be more difficult to satisfy.

Several approaches can be used to improve the reliability of ambiguity resolution in the case that the a priori model delay error is too large to allow "direct" ambiguity resolution. All of them involve starting with a lower accuracy data type, but one whose cycle ambiguity can be reliably resolved. The information in this lower accuracy observable is then used to assist in resolving the cycle ambiguity associated with the higher accuracy data type. Some of these approaches have been discussed previously and are outlined below.[4] What follows is a sketch of possible techniques; more work is required to fully analyze and explore how each of them would perform in an operational CEI system.

[4] S. Thurman, "Continuation of the Connected Element Interferometry System Study," JPL Interoffice Memorandum 314.5-1367 (internal document), Jet Propulsion Laboratory, Pasadena, California, October 20, 1988.

A. BWS Aiding

As mentioned previously, with sufficient spanned bandwidth, BWS group delay residuals can be determined with formal errors of less than a cycle of RF phase. In that case, if dispersive errors can be accurately calibrated and removed, the group delay residual can be used to correct the delay model and aid in ambiguity resolution. How precise must the BWS observable be to aid in ambiguity resolution? Consider a system with two BWS channels, separated in frequency by a spanned bandwidth of ν_{BWS}, and with a formal phase error of σ_ϕ cycles. Requiring that the BWS observable precision be less than 1/6th of an RF cycle, to ensure reliable ambiguity resolution, yields:

$$\frac{\sqrt{2}\sigma_\phi}{\nu_{BWS}} < \frac{1}{6\nu_{RF}}$$

With a 400-MHz spanned bandwidth and a 5-mcyc phase error in each channel, it would then be possible to resolve RF cycles at X-band. It is important to keep in mind that this argument has ignored the effects of dispersive errors due to the ionosphere, solar plasma, or instrumentation, which contribute differently to the phase and group delay data types.

For use in spacecraft tracking, this technique of BWS-aided ambiguity resolution will require large spanned bandwidths on the spacecraft. However, the current X-band deep space downlink frequency allocation is only 40 MHz wide, probably insufficient for resolving the X-band or even the S-band phase ambiguity. At Ka-band, however, a full 500-MHz downlink allocation exists and could aid ambiguity resolution. For example, with a spacecraft incorporating a dual-band X/Ka downlink, and with widely spaced VLBI tones at Ka-band, the accurate Ka-band BWS observable could be used to resolve the X-band phase observable. Such a scenario has been proposed for same-beamwidth interferometric tracking of two spacecraft at Mars with simultaneous X/Ka downlinks [12].

B. Band-to-Band Aiding

Just as one can "step" from the BWS observable, with its spanned bandwidth ν_{BWS}, up to the phase observable at frequency ν_{RF}, one can also imagine using the phase observable at a lower frequency band to aid the phase ambiguity resolution at a higher frequency band, e.g., using an S-band phase residual to resolve the X-band phase ambiguity. Again, the key to making this work is the requirement that dispersive errors represent much less than a cycle of phase at the higher frequency. Instrumental phase calibrations can probably reduce instrumental dispersion to a

few degrees of phase, in which case they would not pose an obstacle to stepping from one band to the next. In that case, the biggest dispersive error will be due to the effects of charged particles in the ionosphere and solar plasma. Since charged-particle delays scale as $1/\nu^2$, whereas cycle ambiguities scale as $1/\nu$, charged particles will represent a four times more serious error in stepping from S-band to X-band than in stepping from X-band to Ka-band.

The current understanding of ionospheric errors and, in particular, spatial and temporal fluctuations in ionospheric delays is not as well developed as the corresponding understanding of tropospheric errors. Empirically, however, the concept of S-to-X-band aiding in the current data set can be tested by asking the question: if the final S-band phase delays had been used instead of the a priori delay model to resolve the X-band phase, would the same integer cycle ambiguities have been determined? Examining the S-band phase delay residuals, it was found that ten of the eleven X-band observables would have been resolved identically. However, one observation (the outlier in the S-band histogram of Fig. 1) would have had a 1-cycle shift in the final X-band observable. Presumably the "direct" solution in this case is the correct one, not the "S-to-X-aided" solution, based on the final distribution of X-band residuals for the repeated observations of the source pair. It is not known whether the S-band outlier was caused by an uncalibrated instrumental fluctuation or by actual fluctuations in the ionosphere above Goldstone.

C. Baseline Aiding

Going to shorter baselines increases the fringe spacing on the sky, and thus relaxes the a priori source position requirements for phase connection (at the cost of lower angular precision). One can imagine a series of progressively longer baselines, where the phase residual on each baseline would be used to aid the ambiguity resolution of the next longer baseline. If the a priori position knowledge is sufficient to resolve the phase ambiguity on the shortest baseline, and if each baseline's phase-delay residual is sufficiently accurate to resolve the ambiguity on the next longer baseline, then ultimately the phase ambiguity on the longest baseline can be resolved. This technique works best if the baselines are exactly parallel, since each baseline only measures instantaneously one component of the source position.

Consider a linear array of five antennas with coordinates of 0, 1, 5, 25, and 125 km along some line. At X-band, the 1-km baseline formed by the first two antennas would correspond to a very large minimum fringe spacing of 40 μrad. Phase connection on this baseline would require only very crude a priori source position information. If the phase observable on this 1-km baseline were sufficient to determine the source position (along the projected baseline direction) to a 1-σ uncertainty of 1/30th of a fringe spacing, that improved source position would enable reliable ambiguity resolution on the 5-km baseline, i.e., the 3-σ source position error would contribute less than half a cycle of phase on the 5-km baseline. This process would continue until the 125-km baseline was reached. A 1/30th of a cycle error on the 125-km baseline would correspond to a 10-nrad angular position determination at X-band.

Each of these techniques has inherent strengths and weaknesses in terms of requirements of the CEI system. The BWS or band-to-band techniques could help to resolve ambiguities at Ka-band in the likely event that stochastic tropospheric fluctuations prevent direct Ka-band ambiguity resolution. All three techniques could help in the case of tracking an object with very poor a priori angular position information. Perhaps the most robust solution would utilize several of these approaches at once: for instance, an array of baselines receiving simultaneous X/Ka downlinks, with widely spaced Ka-band tones for generating a high precision group delay data type.

VI. Summary

A 90-minute observation pass on the DSS 13–DSS 15 baseline has yielded eleven differential phase-delay observables with formal errors of 1 psec and an rms scatter of 10 psec, roughly the level predicted by statistical models of tropospheric fluctuations. These data determined the relative source positions to an accuracy of 60 nrad in one component of sky position. This is the highest angular accuracy achieved to date from observations within a single DSN tracking complex. The other component of source position was not well determined, due to the short observing period. However, similar observations spanning the mutual visibility window for this source would provide comparable accuracy for both components of sky position. Placing a quantitative value on the probability of a cycle error is difficult without a better understanding of the probability distributions of the underlying error sources, although examination of the distribution of phase residuals suggests that the cycle ambiguity resolution was successful.

Developing techniques to ensure highly reliable ambiguity resolution is a prerequisite to incorporating CEI into operational spacecraft tracking. At Ka-band, the few-millimeter level of stochastic tropospheric fluctuations

observed in this experiment would seriously complicate ambiguity resolution. In addition, large uncertainties in a priori source position would also hamper ambiguity resolution. Several techniques have been outlined for using less accurate, but more easily resolved, data types to aid in resolving a higher accuracy observable. Data collected in the last few months and over the coming year should allow some of these techniques to be tested and evaluated.

Acknowledgments

I would like to thank Lyle Skjerve, Ben Johnson, and Earl Lobdell for their invaluable efforts in acquiring these data, George Lutes and Lute Maleki for their assistance in using the fiber-optic link between DSS 13 and DSS 15, Dick Branson and Ojars Sovers for implementing and testing some new features in the MASTER-FIT parameter estimation software, and Jim Ulvestad, Bill Folkner, Mark Finger, and Sam Thurman for their comments on a preliminary draft of this article.

References

[1] C. D. Edwards, "Short Baseline Phase Delay Interferometry," *TDA Progress Report 42-91,* vol. July–September 1987, Jet Propulsion Laboratory, Pasadena, California, pp. 46–56, September 15, 1987.

[2] C. D. Edwards, "Angular Navigation on Short Baselines Using Phase Delay Interferometry," *IEEE Transactions on Instrumentation and Measurement,* vol. 38, pp. 665–667, 1989.

[3] 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, April 15, 1988.

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

[5] C. D. Edwards, "The Effect of Spatial and Temporal Wet-Troposphere Fluctuations on Connected Element Interferometry," *TDA Progress Report 42-97,* vol. January–March 1989, Jet Propulsion Laboratory, Pasadena, California, pp. 47–57, May 15, 1989.

[6] G. Lutes and A. Kirk, "Reference Frequency Transmission Over Optical Fibers," *TDA Progress Report 42-87,* vol. July–September 1986, Jet Propulsion Laboratory, Pasadena, California, pp. 1–9, September 15, 1986.

[7] B. G. Clark, "The NRAO Tape-Recorder Interferometer System," *Proc. IEEE,* vol. 61, pp. 1242–1248, 1973.

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

[9] A. E. E. Rogers, "Very Long Baseline Interferometry with Large Effective Bandwidth for Phase-delay Measurements," *Radio Science,* vol. 22, pp. 1239–1247, 1970.

[10] O. J. Sovers and J. L. Fanselow, "Observation Model and Parameter Partials for the JPL VLBI Parameter Estimation Software 'MASTERFIT'-1987," JPL Publication 83-39, Rev. 3, Jet Propulsion Laboratory, Pasadena, California, December 15, 1987.

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

[12] J. S. Border and W. M. Folkner, "Differential Spacecraft Tracking by Interferometry," *Proceedings of the CNES International Symposium on Space Dynamics,* Toulouse, France, November 6–10, 1989.

Table 1. Frequency configuration

Channel	DSS 13 frequency, MHz	DSS 15 frequency, MHz	Sideband, upper/lower	Dwell time, sec
1	2304.9999	2305.0001	LSB	0.4
2	2264.9999	2265.0001	USB	0.4
3	2297.9999	2298.0001	LSB	0.4
4	8439.9999	8440.0001	LSB	1.6
5	8399.9999	8400.0001	USB	1.6
6	8432.9999	8433.0001	LSB	1.6

Table 2. The a priori J2000 source positions

Source	Right ascension, h m s	Declination, o ′ ″
3C 84	03 19 48.160328	41 30 42.10581
OE 400	03 03 35.242181	47 16 16.27738

Table 3. Observation sequence

Observation no.	Source name	Start − stop, GMT	Average elevation, deg
1	3C 84	213100 - 213400	56.7
2	OE 400	213442 - 213742	53.2
3	3C 84	213824 - 214124	55.3
4	OE 400	214206 - 214506	51.9
5	3C 84	214542 - 214842	54.9
6	OE 400	214924 - 215224	50.7
7	3C 84	215300 - 215600	52.6
8	OE 400	215642 - 215942	49.5
9[a]	3C 84	220018 - 220318	51.2
10[a]	OE 400	220354 - 220654	48.2
11	3C 84	220730 - 221030	49.9
12	OE 400	221106 - 221406	47.0
13	3C 84	221442 - 221742	48.5
14	OE 400	221818 - 222118	45.8
15	3C 84	222154 - 222454	47.2
16	OE 400	222530 - 222830	44.6
17	3C 84	222906 - 223206	45.9
18	OE 400	223242 - 223542	43.4
19	3C 84	223618 - 223918	44.5
20	OE 400	223954 - 224254	42.2
21	3C 84	224330 - 224630	43.2
22	OE 400	224706 - 225006	40.9
23	3C 84	225042 - 225342	41.9
24	OE 400	225418 - 225718	39.7

[a] These two observations were lost due to a malfunction of the antenna pointing computer at DSS 13. (Observations are grouped in pairs used to form the final differential observables.)

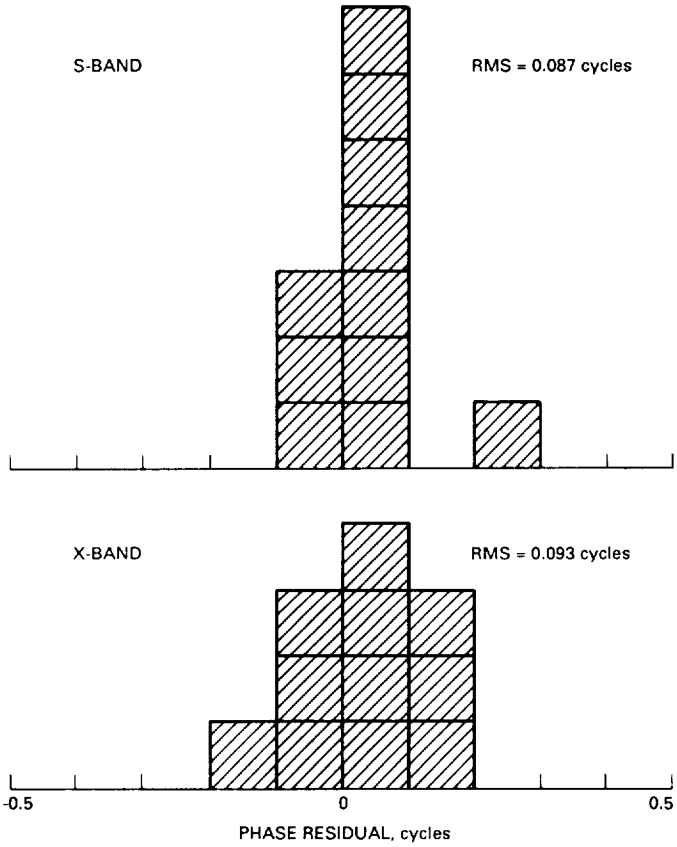


Fig. 1. S-band and X-band phase residuals after ambiguity resolution. The rms width of the distribution is 0.087 cycle at S-band and 0.093 cycle at X-band.

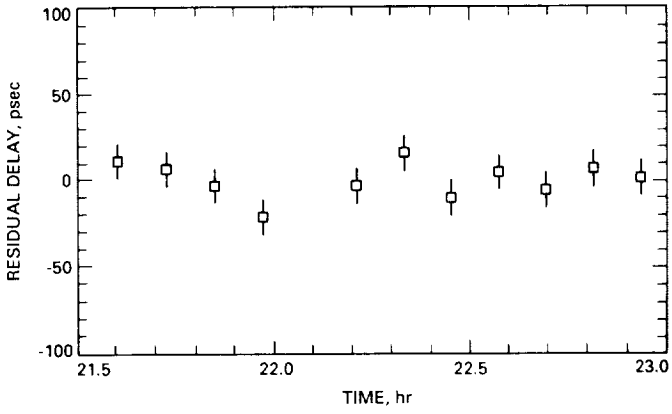


Fig. 2. Final S/X-combined differential phase-delay residuals as a function of time over the 90-minute observation.

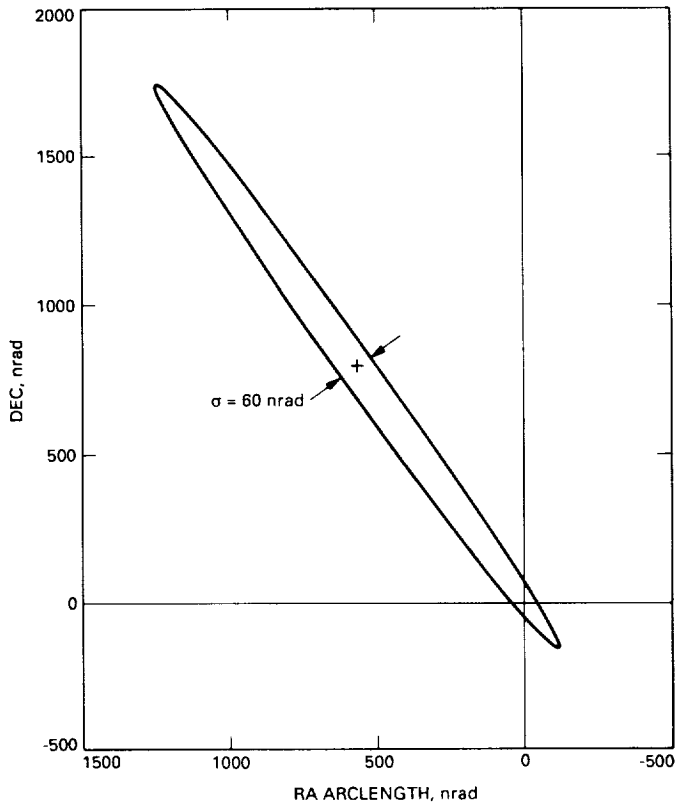


Fig. 3. Estimated source position correction and error ellipse for OE 400, based on a weighted least-squares fit to the phase-delay residuals shown in Fig. 2. The orientation of the ellipse corresponds to the average projected baseline direction over the 90-minute observation period.

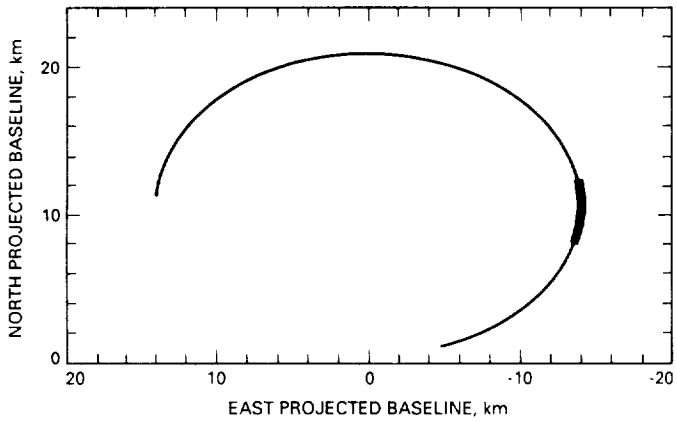


Fig. 4. The projection of the DSS 13–DSS 15 baseline on the plane of the sky, in the direction of the radio source OE 400, over the full mutual visibility window of over 15 hours. The portion of the visibility window represented by the 90-minute observation reported here is shown with a heavier line; during this short period, the projected baseline only rotates through about 10 degrees, and so the resulting position determination is much stronger in one direction on the plane of the sky than in the other.

N90-26219

TDA Progress Report 42-101 May 15, 1990

A Reduced-Dynamic Technique for Precise Orbit Determination

S. C. Wu, T. P. Yunck, and C. L. Thornton
Tracking Systems and Applications Section

Observations of the Global Positioning System (GPS) will enable a reduced-dynamic technique for achieving subdecimeter orbit determination of Earth-orbiting satellites. With this technique, information on the transition between satellite states at different observing times is furnished by both a formal dynamic model and observed satellite positional change (which is inferred kinematically from continuous GPS carrier-phase data). The relative weighting of dynamic and kinematic information can be freely varied. Covariance studies show that in situations where observing geometry is poor and the dynamic model is good, the model dominates determination of the state transition; where the dynamic model is poor and the geometry strong, carrier phase governs the determination of the transition. When neither kinematic nor dynamic information is clearly superior, the reduced-dynamic combination of the two can substantially improve the orbit-determination solution. Guidelines are given here for selecting a near-optimal weighting for the reduced-dynamic solution, and sensitivity of solution accuracy to this weighting is examined.

I. Introduction

The Defense Department's Global Positioning System (GPS) will be in full operation, with at least 18 satellites in orbit, by the end of 1992. The two precise radiometric data types available from GPS, P-code pseudorange and continuous carrier phase, will provide accurate positioning for users on the Earth's surface and in Earth orbit [1]. Advanced differential techniques incorporating GPS data from a global network of ground reference sites now promise to provide the subdecimeter orbit determination accuracy being sought for a growing number of scientific remote-sensing satellites [2-4]. One such satellite is the U.S.–France Ocean Topography Experiment, TOPEX/POSEIDON, to be launched in June of 1992. TOPEX, which will fly in a circular orbit at an altitude of 1,336 km [5], has a formal accuracy requirement of 13 cm for the continuous determination of its geocentric altitude; in fact, TOPEX ocean science would benefit from an altitude accuracy comparable to the 2.5-cm precision of its radar altimeter. A number of proposed international missions for the mid- to late 1990s with high-Earth elliptically orbiting radio interferometric platforms would also benefit from a decimeter-level orbit determination capability. The Deep Space Network has responsibility for GPS-based tracking of TOPEX and for tracking of other satellites in high-Earth orbit.

Differential GPS tracking can take many forms. The simplest is relative point positioning using instantaneous differential pseudorange measurements to four or more GPS satellites. The accuracy of this *geometric* relative positioning is limited primarily by measurement noise and

GPS ephemeris error, magnified by position dilution of precision (PDOP), which is related to observing geometry [1]. For a low-Earth orbiter, instantaneous differential positioning accuracy is expected to approach one meter [6]. In applications where more accurate orbit knowledge is needed, a long-arc *dynamic* solution may be suitable. With this approach, the satellite dynamics are constrained by physical models, and noisy instantaneous measurements obtained over a period of time are combined to improve precision and to yield greater information on the user state at a single epoch [2,3]. For greatest accuracy, all GPS orbits and some ground receiver positions are also adjusted; all solutions are obtained in a reference frame established by a small set of fixed (unadjusted) ground sites. In the dynamic solution, the transition from satellite states at different measurement times to the state at the solution epoch is furnished by integration of the equations of motion, which are governed by the forces (dynamics) acting on the satellites over the time of interest. Any mismodeling of these dynamics will result in systematic errors in the state solution—errors which tend to grow as the data arc length increases.

In [4], a long-arc non-dynamic (or *kinematic*) tracking technique is proposed. With this technique, the instantaneous user satellite positions are again determined by differential GPS pseudorange measurements; however, information about the transition between positions at different times is furnished by the satellite positional change as inferred from observations of continuous differential GPS carrier-phase data. These data enable many point position solutions to be smoothed over long data arcs. Since the GPS carrier phase can be measured extremely accurately (to 1 mm or better in 1 sec), with favorable observing geometry it can provide nearly ideal state-transition information. And because the kinematic solution is fundamentally geometric, the solution is free of dynamic modeling errors. In exchange, however, the solution has a high sensitivity to the continuously changing observing geometry. To maintain decimeter orbit accuracy, strong observing geometry must be ensured by providing sufficient receiving channels on the orbiter and a sound global network of ground receivers [4].

Kinematic tracking discards dynamic orbit models and the associated information entirely. That, indeed, is one of its principal attractions. Not only is sensitivity to model errors eliminated, but the complexity of the solution process is greatly reduced. Nevertheless, one can expect in general that an optimal synthesis of dynamic and kinematic information will offer advantages. In the reduced-dynamic technique, first proposed in [7], both the dynamic and kinematic methods of state-transition determination are used with carefully selected relative weighting. The weight on the dynamic information is controlled through adjusting a set of three process-noise parameters representing a fictitious 3-D force on the user satellite. These parameters include the a priori uncertainty σ_o, the steady-state uncertainty σ, and the correlation time τ.

Appending process-noise parameters to the satellite force model is attractive since, although the fictitious force is piecewise constant and therefore discontinuous between batches, the satellite state components, derived by integrating the noise parameters with the force model, remain continuous. This fits very naturally with the traditional dynamic formulation. Alternatively, one could add process noise to the satellite state itself (position or velocity). While this works well in a kinematic solution, it renders the satellite states discontinuous between batches. This complicates incorporation of a deweighted dynamic model in a reduced-dynamic formulation and is ill-suited to many science applications in which continuous orbits are required.

The dynamic, kinematic, and reduced-dynamic techniques are compared qualitatively in Fig. 1. The dynamic solution (top) adjusts the fewest parameters, preserving maximum data strength and yielding the lowest formal error (error due to data noise), but can suffer a large systematic error from mismodeled dynamics. The kinematic solution (center) eliminates modeling error, but the orbit transition is determined entirely from the observations, data strength is depleted, and the formal error can grow large. The reduced-dynamic solution (bottom) optimally combines the two techniques to achieve the lowest overall error.

Here, a Kalman filter formulation of the reduced-dynamic technique is first presented, then a covariance analysis carried out to evaluate its performance in comparison with the dynamic and kinematic approaches is described. Guidelines for selecting a near-optimal reduced-dynamic weighting are given, sensitivity to this weighting is assessed, and practical aspects of the technique are discussed.

II. Formulation

The reduced-dynamic technique can be described mathematically in terms of a Kalman sequential filter formulation. This involves two steps: *a time update*, which makes use of a state-transition model to propagate the satellite state estimate and covariance from one time batch to the next, and *a measurement update*, which incorporates a new batch of measurements. These two steps alternate until all data batches are incorporated.

A. Time Update

Let $\hat{\mathbf{x}}_j$ be the user satellite state estimate at time t_j, using data up to the time t_j, and $\tilde{\mathbf{x}}_{j+1}$ the predicted state estimate at time t_{j+1}, using data only up to t_j; let $\phi_x(j+1,j)$ denote the state transition from t_j to t_{j+1}. Process noise parameters $\mathbf{p}$ representing a fictitious 3-D force on the user satellite are now introduced. This gives the following dynamic state-transition model for the augmented state $\mathbf{X}=[\mathbf{x},\mathbf{p}]^T$ and its associated covariance P [8]:

$$\tilde{\mathbf{X}}_{j+1} = \phi_j \tilde{\mathbf{X}}_j + B\mathbf{w}_j \tag{1}$$

and

$$\tilde{P}_{j+1} = \phi_j \hat{P}_j \phi_j^T + BQ_jB^T \tag{2}$$

where

$$\phi_j = \begin{bmatrix} \phi_x(j+1,j) & \phi_{xp}(j+1,j) \\ 0 & M_j \end{bmatrix} \tag{3}$$

$$B = \begin{bmatrix} 0 \\ I_p \end{bmatrix} \tag{4}$$

The transition matrix relating $\tilde{\mathbf{x}}_{j+1}$ to the process noise parameters $\mathbf{p}_j$ is $\phi_{xp}(j+1,j)$; M_j is a 3 × 3 diagonal matrix with its ith element

$$m_i = \exp\left[-(t_{j+1}-t_j)/\tau_i\right] \tag{5}$$

$\mathbf{w}_j$ is a white-noise process of covariance Q_j, which, for convenience, is assumed diagonal with its ith element $q_i = (1-{m_i}^2){\sigma_i}^2$; I_p is a unit matrix; τ_i is the correlation-time constant, which controls the decay rate of the correlation between time batches; and σ_i is the steady-state uncertainty, which is equal to the root-mean-squared (rms) value of the process-noise uncertainty after a long time. Both σ_i and τ_i can be selected to be the same for all i in this application, so the subscript i will be dropped from now on. The relative weighting of the dynamics is varied by selecting different values for the steady-state uncertainty σ, the correlation time τ for the process-noise parameters, and the a priori uncertainty σ_o, which is the initial error of the parameters. Increasing τ and decreasing σ_o and σ increases the weight on the dynamic information. When $\tau \to \infty, \sigma \to 0$ and $\sigma_o \to 0$, the technique reduces to conventional dynamic tracking; when $\tau \to 0, \sigma \to \infty$ and $\sigma_o \to \infty$, it becomes purely kinematic. It follows that an optimal reduced-dynamic solution must be as good as or better than both the purely dynamic and purely kinematic solutions.

B. Measurement Update

The model for a measurement update in the reduced-dynamic technique is the same as in the dynamic or kinematic techniques, with the exception that $\mathbf{X}$ and P are now associated with the augmented state. Thus,

$$\hat{\mathbf{X}}_j = \tilde{\mathbf{X}}_j + G_j(\mathbf{z}_j - A_j\tilde{\mathbf{X}}_j) \tag{6}$$

and

$$\hat{P}_j = \tilde{P}_j - G_jA_j\tilde{P}_j \tag{7}$$

where $\mathbf{z}_j$ is the measurement vector at time t_j; A_j is the matrix of the corresponding measurement partials with $\mathbf{X}_j$; and G_j is the Kalman gain given by

$$G_j = \tilde{P}_jA_j^T(A_j\tilde{P}_jA_j^T + R_j)^{-1} \tag{8}$$

with R_j being the error covariance of $\mathbf{z}_j$.

The above models have been formulated in terms of *current* state for clarity. A *pseudoepoch* state, *U-D* factorized formulation [8][1] of these models has been implemented in the GPS analysis software system, OASIS, developed at JPL [9].[1]

III. Covariance Analysis

A. Assumptions and Approach

A covariance analysis comparing the accuracy of TOPEX altitude determination expected with the reduced-dynamic, dynamic, and kinematic techniques is presented next. Assume a constellation of 18 GPS satellites placed in six orbit planes. GPS P-code pseudorange and carrier phase data are acquired by a receiver on TOPEX and by six globally distributed ground sites. Data noise, after a 5-minute integration and dual-frequency correction for ionospheric delay, is put at 5 cm and 0.5 cm for pseudorange and carrier phase, respectively, which is consistent with the performance of modern GPS receivers. Carrier phase biases are adjusted with a large a priori uncertainty. A 2-hour data arc covering a full TOPEX orbit is used initially. The ground track of the TOPEX orbit and the positions of the six ground sites are shown in Fig. 2. Other error sources evaluated are given in Table 1.

[1] S. C. Wu, W. I. Bertiger, J. S. Border, S. M. Lichten, R. F. Sunseri, B. G. Williams, P. J. Wolff, and J. T. Wu, *OASIS Mathematical Description, V. 1.0*, JPL D-3139 (internal document), Jet Propulsion Laboratory, Pasadena, California, April 1, 1986.

The clocks on all GPS satellites and at all but one ground site are modeled as white process noise and thus are adjusted independently at each time point. This clock model is comparable to using doubly differenced data and is the most conservative (pessimistic) model we can use, since it maximizes the number of adjusted parameters and hence the formal error. The gravity-error model is derived by scaling the difference between corresponding coefficients, up to degree and order 20, of two different Earth models, GEM-10 and GEM-L2 [10,11]. This form of gravity-error model is convenient to implement, is easily varied with a single scaling factor, and has proven reliable in numerous studies over the years. Alternatively, one can employ the covariance matrix associated with a single Earth model. Our own experiments have shown that a 50 percent scaling of GEM-10 − GEM-L2 approximates the error obtained with the covariance matrix from the gravity model, GEM-T1 [12], which is one of the best current models.

The 1-cm zenith troposphere error assumes the use of a water-vapor radiometer at each ground site. An earlier analysis showed that the error from mismodeling atmospheric drag is less than 1 mm for TOPEX over several-hour arcs of data [3]; therefore, drag is not included here. Such potential error sources as thermal imbalances and outgassing; instrumental delay variations (in the GPS receivers) not common to all signals; and imperfect knowledge of the satellite center of mass, the GPS antenna phase center, and platform attitude are being carefully controlled for TOPEX and are expected to be below 1 cm. One other potentially serious error source is GPS signal multipath. Because multipath is not readily treated by covariance analysis, separate simulation studies, incorporating all major reflecting surfaces, antenna gain patterns, and receiver tracking characteristics have been carried out. The result is that while the instantaneous multipath can at times look alarming (tens of centimeters on pseudorange and up to 1 cm on the carrier), it oscillates with periods of minutes or less and therefore averages down quickly; the net orbit error is typically at the 1-cm level or below after a few hours of averaging.

The inherent GPS data strength allows accurate simultaneous adjustment of GPS orbits and all but three ground sites [2–4]. Because the GPS satellites are at an altitude where dynamic mismodeling is very small, their solutions remain dynamic in all three approaches. Since we know that GPS orbits will be routinely tracked with high accuracy by a global network, a tight a priori uncertainty is applied to GPS states. By contrast, to examine the strength of each solution technique, a large a priori error is placed on TOPEX. In actual operation, a much tighter TOPEX a priori error could be used.

JPL's recently developed Deep Space Network (DSN) Rogue GPS receiver is currently planned for use at all ground reference sites. The Rogue can track up to eight GPS satellites simultaneously, while the operational TOPEX receiver will have a six-satellite capacity. For study purposes it is first assumed that the TOPEX and ground receivers can observe *all* visible GPS satellites (typically six or seven); then cases with lesser TOPEX receiver capacity are examined. The three tracking techniques will be assessed here by comparing the TOPEX altitude errors over the entire data span. For this, the *pseudoepoch-state* covariances of TOPEX are first smoothed backward and then mapped to all time points when data are taken. Comparison is made between the rms errors calculated over the entire data span.

B. Results

In a preliminary study not shown here, we examined the limiting cases of the reduced-dynamic technique. With τ set to 0 and both σ_o and σ set to a large number, the error estimate indeed approached the kinematic solution. With τ set to a large number and both σ_o and σ set to 0, it yielded the dynamic estimate. Here, a series of intermediate values for τ, σ_o, and σ is studied. In general, when τ is long compared to the batch size, the results vary with the *batch-to-batch* uncertainty $\sigma_{bb} = (1 - m^2)^{1/2}\sigma$, rather than with the steady-state uncertainty σ and τ individually. (The batch-to-batch uncertainty is the one-sigma change in value from one time batch to the next that is allowed for the process-noise parameters.) In the rest of this analysis, a batch size of 5 minutes and a constant $\tau = 15$ minutes are used for all cases with the reduced-dynamic approach; only $\sigma_o = \sigma$ is varied to yield a nearly optimal solution.

Figure 3 shows the TOPEX altitude error as a function of the percentage of the GEM-10 − GEM-L2 error, for various values of σ. Also included are the results with dynamic tracking ($\sigma = 0, \tau \to \infty$) and kinematic tracking ($\sigma \to \infty, \tau = 0$). It is clear that for any finite dynamic-model error (in this case dominated by gravity), a range of σ exists over which TOPEX altitude error is lower than with either the dynamic or kinematic solutions. In other words, the reduced-dynamic technique is superior provided that the dynamic model is properly weighted. A procedure for estimating the proper weight is outlined later.

Figure 4 compares the reduced-dynamic solution with the dynamic and kinematic solutions for three different viewing capacities for the TOPEX receiver: four, five, and

all GPS satellites (typically six, seldom more than seven) visible above a 90-deg zenith angle. In the cases with restricted receiver capacity, the GPS satellites observed are carefully selected to minimize satellite switches over the observing period (thereby maximizing continuity in carrier phase measurements) while still maintaining good observing geometry (low PDOP). The gravity error is fixed at 50 percent of the difference between GEM-10 and GEM-L2. A near-optimal weight (σ = 0.5 μm/sec^2) is used for the reduced-dynamic solution in all three cases.

When the TOPEX receiver can track all visible GPS satellites, geometry is always strong and kinematic tracking is effective; incorporating additional dynamic information through the reduced-dynamic technique improves accuracy by only 1 cm. A lower gravity error, perhaps achieved through gravity tuning, would of course improve reduced-dynamic performance. At the other extreme, when the receiver can track only four GPS satellites, geometry is often poor and dynamic tracking is far superior to kinematic; the optimal reduced-dynamic combination again offers little advantage. If, however, the gravity error is doubled, as would be the case with a lower orbit, the error with dynamic tracking nearly doubles, to 24 cm, while the reduced-dynamic performance degrades only moderately, to 16 cm, illustrating that even weak geometry can be of value when dynamics are poorly known. In the intermediate case, with TOPEX tracking up to five satellites at once, the dynamic and kinematic solutions are better balanced, achieving 12 and 16 cm, respectively. The reduced-dynamic combination improves this to 9 cm. In general, the reduced-dynamic technique is of greatest value when the kinematic and dynamic solutions are comparable.

Dynamic tracking performance degrades over regions where gravity is poorly known (for example, over the oceans). Kinematic performance, on the other hand, can vary dramatically with changing observing geometry. In the reduced-dynamic solution, the two techniques complement one another, and the solution is better balanced. This is illustrated in Fig. 5, which compares TOPEX altitude accuracy over the whole orbit (two hours), using the three techniques. In this case both a TOPEX receiver tracking five GPS satellites and the 50 percent of GEM-10 − GEM-L2 gravity error are assumed. Both the dynamic and kinematic solutions show peak errors of 25 cm or higher at some points. The reduced-dynamic solution, with a near-optimal weight (σ = 0.5 μm/sec^2), smooths these peaks and remains below 13 cm for the entire period. Reduction of the error peaks results from a balance of state-transition information between dynamics and kinematics. When the information from one source is weak, the Kalman filter places greater weight on the other to minimize the overall error. To further illustrate this trade, Fig. 6 breaks down the TOPEX altitude error into its contributing components at the three times (20, 60, and 90 minutes past epoch) when either the dynamic or the kinematic error grows large. The balance of transition information in the reduced-dynamic solution has resulted in a more uniform contribution from all error components.

In the examples up to this point, a 2-hour tracking arc has been used. In general, as the span is increased, the effects of data noise and troposphere are reduced while the dynamic modeling error grows. In a purely dynamic solution, the effect of increasing model error soon overtakes the decreasing data error and the overall error tends to grow with data span. In the optimal reduced-dynamic solution, however, the estimator continuously shifts weight to the increasingly strong data, away from dynamics, as the span increases. This deweighting of dynamics is a natural consequence of the estimation process; no change in σ is needed since the optimal σ applies to a specific dynamic-model error, independent of data span. As a result, with optimal weighting, the overall performance will tend to improve with increased data span.

Figure 7 compares the TOPEX altitude error when 2-hour and 4-hour data spans are used. The longer data span reduces the error over the initial 2-hour period to 7 cm from the 8.9 cm obtained with the original 2-hour arc. An examination of the error breakdown shows a reduction in gravity error, reflecting the deweighted dynamics, as well as in other errors. Although spans longer than four hours have not been studied for TOPEX, we expect that the error will reduce monotonically with data span. Because the weight on the dynamic model decreases with longer data span, a reduced-dynamic solution will tend to a kinematic solution as the span is increased. Note that this is true only on the assumption that a fixed dynamic model is used, independent of the data arc length. If the model is improved through tuning or other efforts, the optimal weight for a given data span will shift back toward the dynamic solution [13].

C. Other Applications

TOPEX nicely illustrates the benefits of reduced-dynamic tracking since its altitude of more than 1,300 km, six-satellite receiver capacity, and relatively compact dimensions permit both good observing geometry and reasonably well-modeled dynamics. A far greater modeling challenge is presented by several other current or planned NASA space platforms: the large (14 m) platforms of the polar-orbiting Earth Observing System (EOS), which will carry heavy slewing instruments and fly at 700 km; the actively maneuvering space shuttle, at altitudes as low as

300 km; and the sprawling (155 m) Space Station Freedom, at about 400 km. All will eventually carry experiments seeking tracking accuracies better than 10 cm. Indeed, a recent international workshop on space geodesy set a goal of "no more than 1 cm rms error, single pass, without orbit discontinuities" [14] for tracking future orbiting ocean altimeters, such as the one that will fly on EOS.

Since one cannot expect to approach centimeter or even decimeter accuracy in modeling the dynamics of such ungainly platforms, the optimal orbit solution strategy will be almost purely kinematic. To maximize performance under kinematic tracking, geometric observing strength must be maximized. With that in mind, one can take the examples of EOS (98-deg inclination, 705-km altitude) and Space Station Freedom (28 deg, 400 km) and carry out covariance studies under a more robust set of assumptions: the GPS constellation is increased to 24 satellites, as is expected to occur by 1995; the ground network is expanded to 10 sites; the flight receiver is extended to track all satellites in view down to the Earth limb (typically a dozen or more); and the three fixed ground sites are assumed known to 2 cm in each component, which is expected to be achieved or surpassed by very long baseline interferometry within the next few years [14].

Figure 8 plots the rms position errors estimated for a purely kinematic solution for arc lengths of 2, 4, and 8 hrs. Under the revised assumptions, the observing geometry is so consistently strong that few-centimeter accuracy is achieved continuously, and the rms error approaches 1 cm per component after 8 hrs. Despite the widely different orbits and dynamics, performance is virtually indistinguishable between EOS and the Space Station, reflecting the full emphasis on geometric strength (which differs negligibly between the two) over dynamics. Note that the purely kinematic position solution is referred to the phase center of the orbiter's GPS antenna, which can be calibrated with submillimeter accuracy, rather than to the platform center of mass, which can be difficult to pinpoint on a large and variable structure. Because the dynamics of these orbiters are so poorly known, optimal reduced-dynamic solutions would improve the kinematic results by only 1 or 2 mm—at a great cost in computation.

For satellites in high-Earth elliptical orbits, which might cover a range of altitudes from 1,000 km to 20,000 km or higher, a variety of different process-noise force models for reduced-dynamic tracking can be employed. At the lower altitudes, models similar to those adopted for TOPEX can be utilized, while at the high altitudes (where gravity and drag are insignificant), low levels of process noise with long time constants are more appropriate to absorb unmodeled accelerations that are due to gas leaks and solar radiation pressure. Covariance analyses to study orbit determination performance for high-Earth elliptical orbiters using GPS will be reported in a future article; however, preliminary analyses indicate that orbit accuracies at the several-decimeter level should be possible even at altitudes of 20,000 km or higher.[2]

IV. Weighting the Dynamic Model

For applications such as TOPEX where geometry and dynamics are more balanced, a procedure is needed to estimate the weight for the dynamic model, specified by σ with any adopted τ, that minimizes the rms orbit error. This may be difficult to do precisely, since the quality of dynamic models is not always well understood; often, in fact, unsuspected modeling errors are present. Fortunately, the sensitivity to a departure from optimal weighting appears to be low. This is illustrated in Fig. 9, in which the TOPEX altitude error with the reduced-dynamic technique is shown as a function of the level of gravity error for three different weightings. Dynamic and kinematic results are shown for comparison. The weight $\sigma = 0.5\ \mu\text{m/sec}^2$ is nearly optimal for a gravity error of 50 percent of GEM-10 − GEM-L2. The two other curves are for values of σ a factor of 4 larger ($\sigma = 2\ \mu\text{m/sec}^2$) and smaller ($\sigma = 0.125\ \mu\text{m/sec}^2$). This wide range of suboptimal weights increases the TOPEX altitude error by only 0.3 cm at one end and 1.5 cm at the other. In other words, the performance is fairly insensitive to suboptimal σ.

In practice, an approximate weight can be estimated in advance through a covariance analysis. This is done by a series of filter runs, each using a different weight, simulating the actual measurement and estimation scenario. Realistic data noise and models for all significant (including dynamic) error sources must be considered and their effects on the orbit determination evaluated. The weight resulting in the lowest estimated rms orbit error is the best estimate of the optimal weight for actual data processing.

A misjudgment of the dynamic-model error will, of course, result in selection of a suboptimal weight; however, one can take care to minimize the effect of such a misjudgment with the following strategy: The assumed dynamic error model is used first to predict the performance of both the dynamic and kinematic solutions. If either of these appears far superior to the other, say, by a factor of 3 or more, the slight improvement that would

[2] S. M. Lichten, "Orbit Determination for High-Earth Elliptical Orbiters Using GPS I. Initial Results for VSOP," JPL Section 335 IOM 335.4-89-123 (internal document), Jet Propulsion Laboratory, Pasadena, California, November 15, 1989.

result by combining the two approaches with the reduced-dynamic technique may not justify the extra effort, and the simpler form can be adopted. If neither technique is far superior, a weight departing from the predicted optimum in a direction favoring the kinematic (i.e., larger σ) will be prudent. Such a bias in the weighting can reduce the more damaging effect of dynamic error if the error is larger than expected. This point is illustrated in Table 2, which summarizes the effect of weight misjudgment by a factor of 4 on TOPEX altitude determination for two different levels of gravity error. These results suggest that a weight biased in favor of the kinematic approach is preferable when the level of dynamic error cannot be well determined.

V. Other Considerations

In the analysis of the reduced-dynamic solution, a fictitious 3-D force on TOPEX was treated as process noise and adjusted together with TOPEX and GPS states. Introduction of the force is merely for the purpose of changing the filter model to reduce that model's reliance on the dynamic model. Since, in the real world, this force does not exist, its presence in the formulation adds an error source in the estimation process, causing the formal error to be overestimated. To remove this effect, an "evaluation" run[3] of the filter is needed. In an evaluation run, the filter model would be specified as before, including the fictitious force, but the contribution of this force in the "truth" model would be ignored. Such an algorithm is fairly complicated when a smoothing process is required, because of the dynamic process-noise parameters involved. For a fair estimate of this effect, evaluation runs with the process-noise force replaced by correlated piecewise-constant (in time) forces have been made, thus avoiding the need of smoothing. These runs show that the spurious increase in the formal error due to the fictitious force is only a few millimeters; the corresponding increase in the total error is even smaller, typically 1 to 2 mm.

VI. Conclusions

A reduced-dynamic technique for determining the orbits of Earth satellites is made possible with observations of the Global Positioning System. In this technique, satellite-state-transition information obtained from both a dynamic model and continuous GPS carrier phase observations is optimally combined to improve orbit determination accuracy. Analysis indicates that a significant improvement can be expected when neither of the two types of state transition information is far superior to the other. Performance is not highly sensitive to the relative weighting between dynamic and kinematic information. When the actual level of dynamic model error is uncertain, an additional deweighting of the dynamic model is recommended; this would prevent an inordinately large error resulting from larger-than-expected dynamic model error.

Although a tracking arc of 2 hours was used for most of the reduced-dynamic analysis, a 4-hour span was examined to illustrate the improved performance with increased data. Further improvement can be expected with longer data spans due to reduction of the effects of data noise and random error in tropospheric delay modeling. The effects of increasing dynamic-model error will be automatically controlled in the reduced-dynamic solution by further deweighting and will thus remain at a low level. Because the weight on the dynamic model is lowered with growing data strength, a reduced-dynamic solution will gradually approach the kinematic solution as the data span increases, provided that a fixed dynamic model is used.

Reduced-dynamic tracking can be used with any Earth satellite that can adequately observe GPS. The altitude range over which reduced-dynamic tracking provides useful improvement over dynamic and kinematic tracking will depend on the actual level of the dynamic-model uncertainty. It is expected that satellites at altitudes between 400 and 2,000 km will receive the greatest benefit. Above this range, GPS observability diminishes while dynamic-model errors decrease markedly, so greater reliance on dynamic tracking will be favored. Below this range, uncertainties in gravity and atmospheric drag become so great that a kinematic solution may be favored for simplicity. For some exceptional vehicles, other considerations apply. The actively maneuvering space shuttle may receive no benefit from dynamics at any altitude, while drag-compensated satellites may exploit dynamics at even the lowest orbit altitudes.

In designing a GPS-based precise tracking system for an Earth orbiter, there is a simple trade-off between modeling accuracy and geometric strength to consider. Where the models are strong, the geometry can be relaxed and the flight and ground systems kept relatively simple. Where the models are weak, as will be the case with a number of dynamically complex missions in the future, the geometry must be strengthened. In any case, the global coverage and unique mix of data types offered by GPS ensure that there will be a practical system design and solution strategy that can deliver orbit accuracies well under a decimeter for any low-Earth satellite.

[3] C. L. Thornton, "Triangular Covariance Factorizations for Kalman Filtering," TM 33-798 (internal document), Jet Propulsion Laboratory, Pasadena, California, October 1976.

References

[1] R. J. Milliken and C. J. Zoller, "Principles of Operation of NAVSTAR and System Characteristics," *Navigation*, vol. 2, no. 2, pp. 95–106, Summer 1978.

[2] T. P. Yunck, W. G. Melbourne, and C. L. Thornton, "GPS-Based Satellite Tracking System for Precise Positioning," *IEEE Trans. Geosci. Remote Sens.*, vol. GE-23, no. 4, pp. 450–457, July 1985.

[3] S. M. Lichten, S. C. Wu, J. T. Wu, and T. P. Yunck, "Precise Positioning Capabilities for TOPEX Using Differential GPS," paper AAS-85-401, AAS/AIAA Astrodynamics Specialists Conference, Vail, Colorado, August 1985.

[4] T. P. Yunck and S. C. Wu, "Non-Dynamic Decimeter Tracking of Earth Satellites Using the Global Positioning System," paper AIAA-86-0404, presented at the AIAA 24th Aerospace Sciences Meeting, Reno, Nevada, January 1986.

[5] G. H. Born, R. H. Stewart, and C. A. Yamarone, "TOPEX—A Spaceborne Ocean Observing System," in *Monitoring Earth's Ocean, Land, and Atmosphere from Space—Sensors, Systems, and Applications*, A. Schnapf (ed.), New York: AIAA, Inc., pp. 464–479, 1985.

[6] W. I. Bertiger and T. P. Yunck, "The Limits of Direct Satellite Tracking with GPS," *Navigation*, vol. 37, no. 1, Spring 1990 (to appear).

[7] T. P. Yunck, S. C. Wu, J. T. Wu, and C. L. Thornton, "Precise Tracking of Remote Sensing Satellites with the Global Positioning System," *IEEE Trans. Geosci. Remote Sens.*, vol. GE-28, no. 1, pp. 108–116, January 1990.

[8] G. J. Bierman, *Factorization Methods for Discrete Sequential Estimation*, New York: Academic Press, 1977.

[9] S. C. Wu and C. L. Thornton, "OASIS—A New GPS Covariance and Simulation Software System," in *Proceedings of the First International Symposium on Precise Positioning with GPS*, pp. 337–345, May 1985.

[10] F. J. Lerch, S. M. Klosko, R. E. Laubscher, and C. A. Wagner, "Gravity Model Improvement Using Geos 3 (GEM 9 and 10)," *J. Geoph. Res.*, vol. 84, no. B8, pp. 3897–3916, July 1979.

[11] F. J. Lerch, S. M. Klosko, G. B. Patel, and C. A. Wagner,"A Gravity Model for Crustal Dynamics (GEM-L2)," *J. Geoph. Res.*, vol. 90, no. B11, pp. 9301–9311, September 1985.

[12] F. J. Lerch, J. G. Marsh, S. M. Klosko, G. B. Patel, and D. S. Chinn, "The GEM-T1 Gravity Model: An Error Assessment," *EOS, Trans. Am. Geophys. Union*, vol. 68, no. 44, p. 1247, November 3, 1987.

[13] W. G. Melbourne, T. P. Yunck, and S. C. Wu, "GPS-Based Precision Positioning of Earth Orbiting Remote Sensing Systems," paper AAS-86-398, AAS 33rd Annual Meeting—Aerospace: Century XXI, Boulder, Colorado, October 1986.

[14] I. I. Mueller and S. Zerbini (eds.), *The Interdisciplinary Role of Space Geodesy*, New York: Springer-Verlag, 1989.

Table 1. Error model and other assumptions used in covariance analysis

Model component	Assumption
User satellite	TOPEX (1,334 km in altitude)
Number of stations	6 (cf. Fig. 2)
Number of GPS satellites	18
Cut-off elevation	10 deg at stations 0 deg at TOPEX
Data type	P-code pseudorange Carrier phase
Data span	2 hours
Data interval	5 minutes
Data noise	5 cm (pseudorange) 0.5 cm (carrier phase)
Carrier phase bias	10 km (adjusted)
Clock bias	3 μsec (adjusted as white process noise)
TOPEX epoch state	2 km; 2 m/sec (adjusted)
GPS epoch states	2 m; 0.2 mm/sec (adjusted)
Station location	5 cm each component
Zenith troposphere	1 cm
Earth's GM	1 part in 10^8
Gravity	Scaled GEM-10–GEM-L2 (see text)
Solar pressure	10 percent

Table 2. Effects of weight misjudgment on TOPEX altitude determination accuracy

Gravity error	$\frac{1}{4}$ × optimal σ	optimal σ	4 × optimal σ
50 percent	10.4 cm	8.9 cm	9.2 cm
100 percent	12.3 cm	9.7 cm	10.1 cm

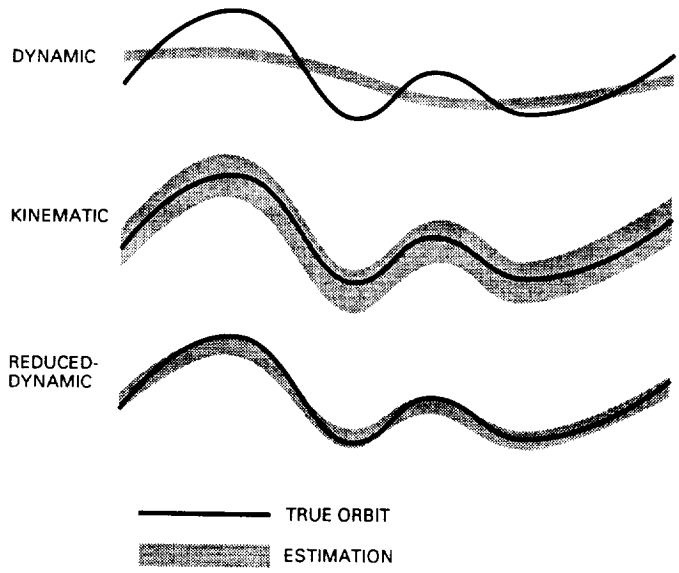


Fig. 1. Qualitative comparison of dynamic, kinematic, and reduced-dynamic tracking performances.

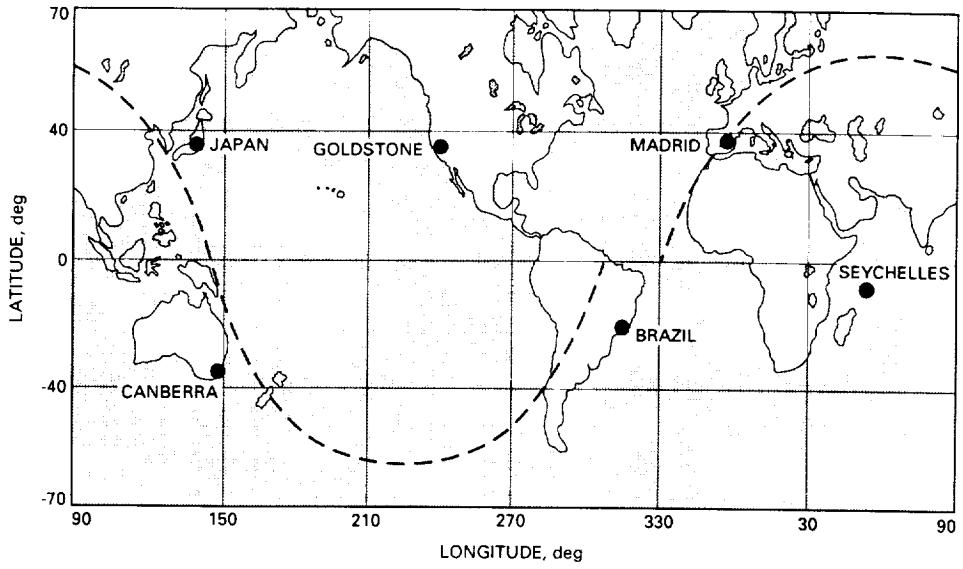


Fig. 2. Ground track of TOPEX orbit, and ground stations used in error analysis.

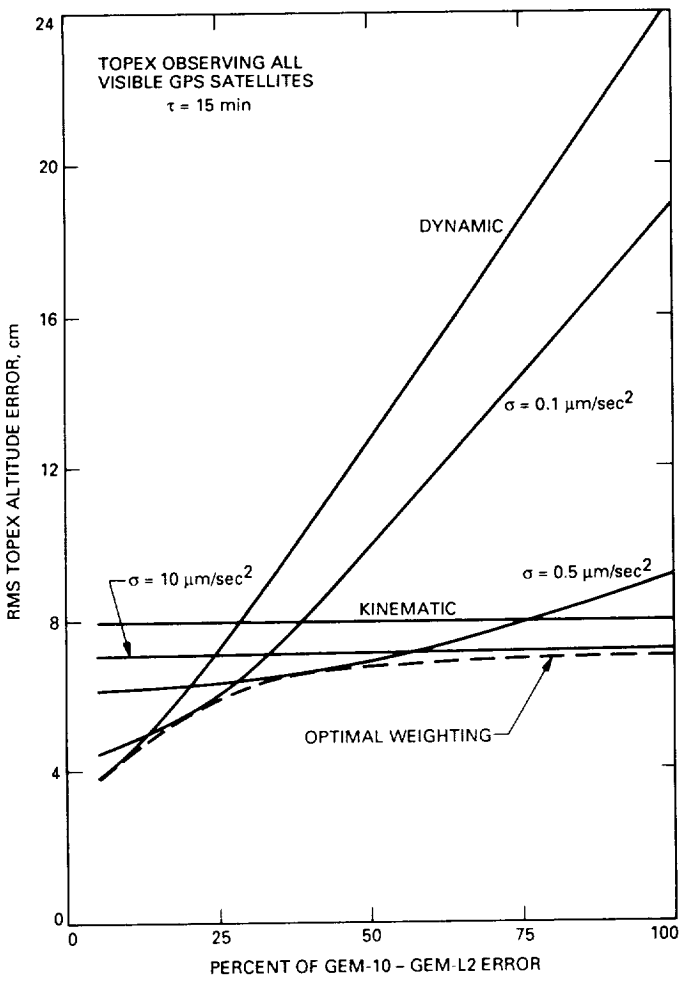


Fig. 3. Performance of reduced-dynamic technique with different weighting on dynamic model.

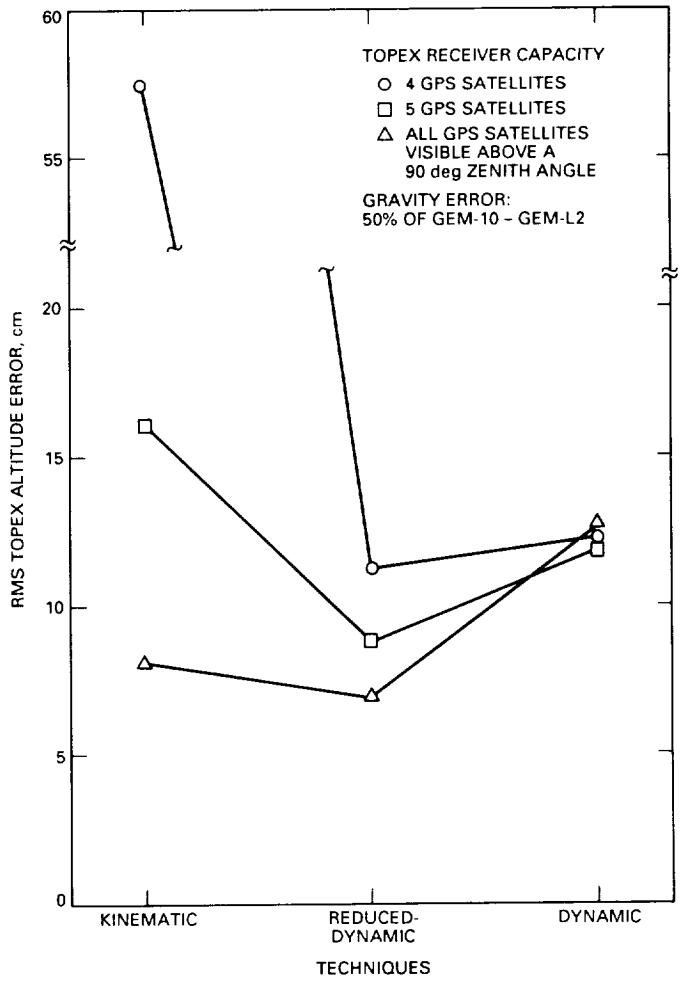


Fig. 4. Performance of dynamic, reduced-dynamic, and kinematic techniques.

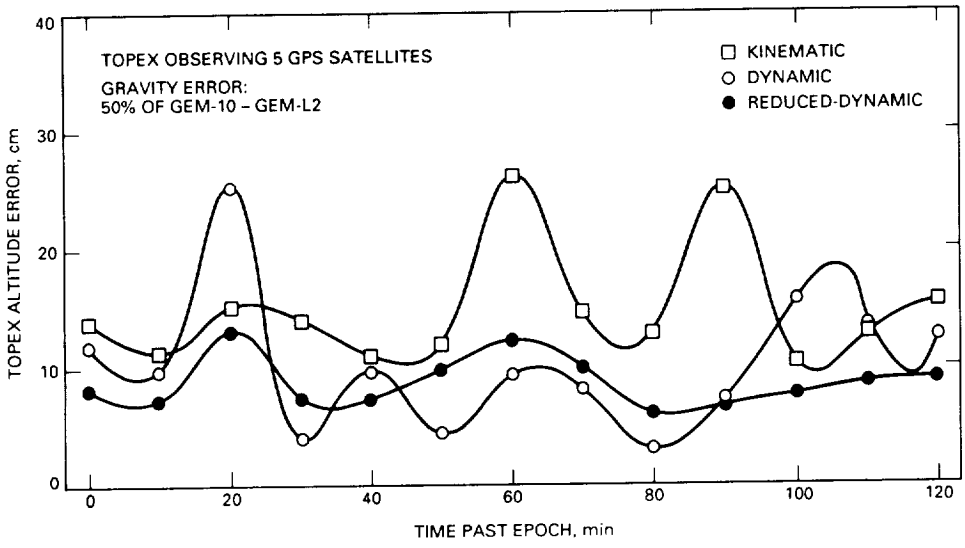


Fig. 5. Variation of TOPEX altitude error over time.

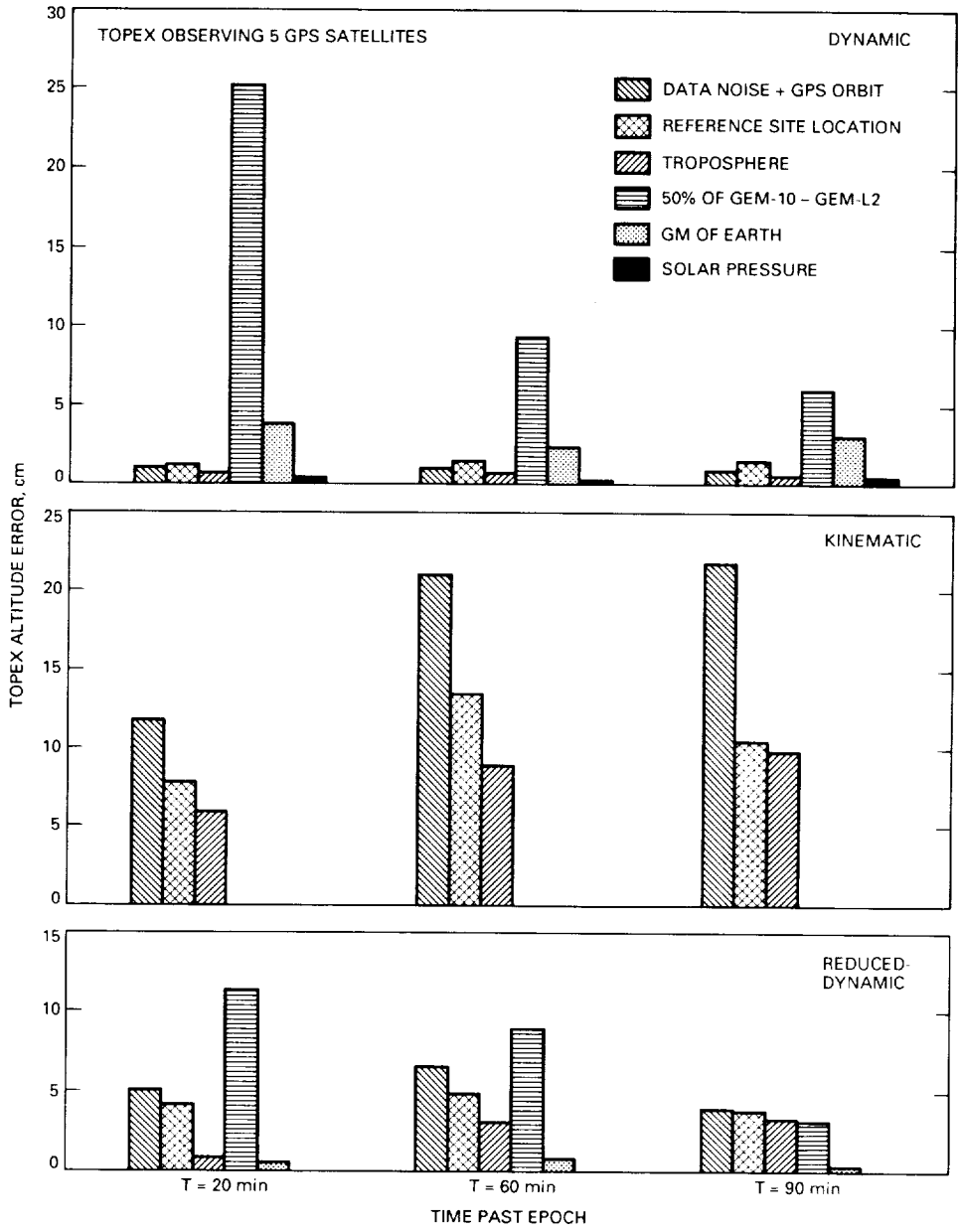


Fig. 6. Breakdown of TOPEX altitude error at three different epochs.

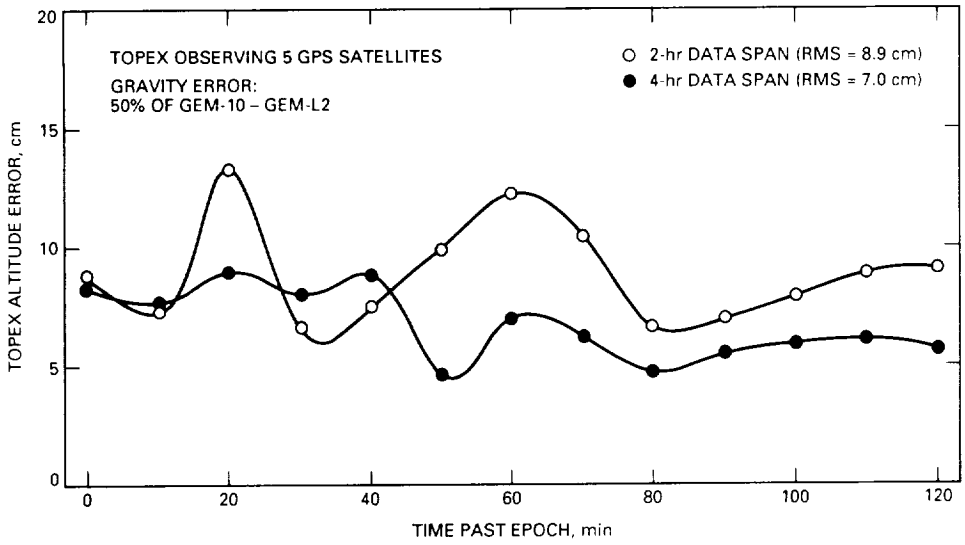


Fig. 7. Comparison of reduced-dynamic tracking performance with different data spans.

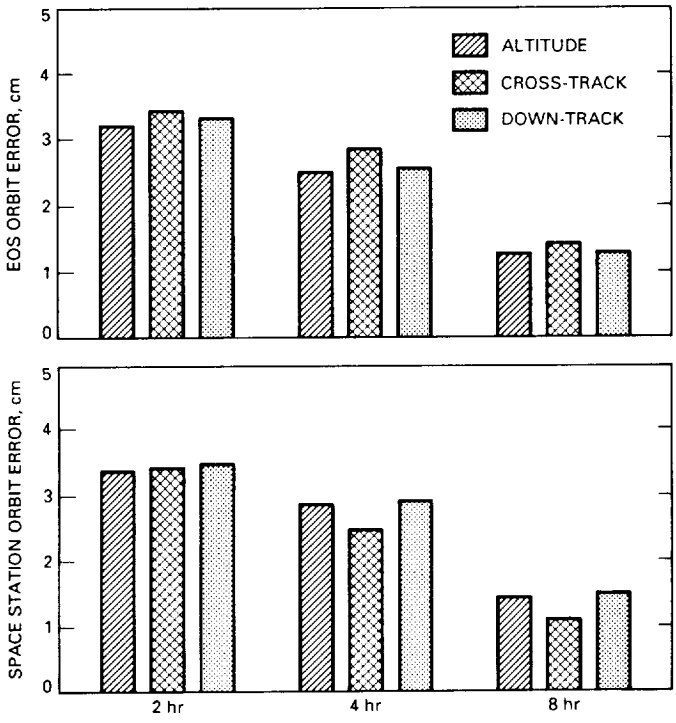


Fig. 8. Estimated kinematic tracking accuracies for EOS and Space Station, both with a robust GPS observing system.

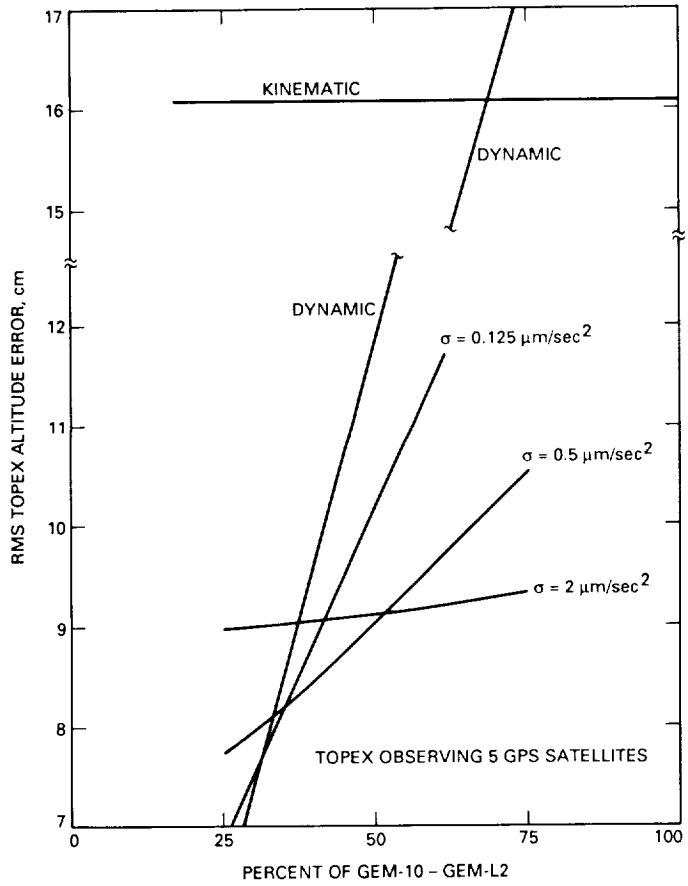


Fig. 9. Insensitivity of TOPEX accuracy to suboptimal weighting.

Information Content of a Single Pass of Phase-Delay Data From a Short Baseline Connected Element Interferometer

S. W. Thurman
Navigation Systems Section

This article presents an analytic development of the information array obtained with a single tracking pass of phase-delay measurements made from a short baseline interferometer. Phase-delay observations can be made with great precision from two antennas using a single, common distributed frequency standard, hence the name "connected element." With the information array, closed-form expressions are developed for the error covariance in declination and right ascension. These equations serve as useful tools for analyzing the relative merits of candidate station locations for connected element interferometry (CEI). The navigation performance of a short baseline interferometer located at the Deep Space Network's (DSN's) Goldstone complex is compared with that which is presently achievable using Very Long Baseline Interferometry (VLBI) over intercontinental baselines. The performance of an intracomplex pair of short baselines formed by three stations is also investigated, along with the use of a single baseline in conjunction with conventional two-way Doppler data. The phase-delay measurement accuracy and data rate used in the analysis are based on the expected performance of an experimental connected element system presently under construction at Goldstone. The results indicate that the VLBI system that will be used during the Galileo mission can determine the declination and right ascension of a distant spacecraft to an accuracy of 20–25 nrad, while the CEI "triad" system and the combination CEI-Doppler system are both capable of 30–70 nrad performance.

I. Introduction

At the present time, wideband delta Very Long Baseline Interferometry (ΔVLBI) is the most accurate angular measurement tool of all the radio metric data types used in deep space navigation. The wideband ΔVLBI system designed for the Galileo mission is capable of angular measurement accuracy in the 20–30 nrad range, which corresponds to a position uncertainty of about 3.0–4.5 km per astronomical unit (AU).[1] Connected element interferometry (CEI) is a new radio interferometric technique in which a common frequency standard, distributed through

[1] *Galileo Navigation Plan*, Galileo Project Document 625-566, Rev. A (internal document), Jet Propulsion Laboratory, Pasadena, California, October 1989.

a fiber-optic communications link to two stations 10–100 km apart, is used to make very precise measurements of the time delay of an incoming signal between the two stations [1,2]. The proximity of the stations makes it possible to observe a spacecraft for 9 to 14 hours, as opposed to the typical 0-to-4-hour view period obtainable with an intercontinental baseline.

In order to compare the relative capabilities of CEI and wideband ΔVLBI, a simple model of the differenced one-way range (DOR) observable, which is theoretically equivalent to both CEI phase-delay and wideband ΔVLBI measurements, was developed and used to form the information array for each data type. The two information arrays were then inverted to obtain closed-form expressions for the error covariance of spacecraft declination and right ascension associated with each tracking model. The use of CEI augmented with two-way Doppler is also investigated, using a simple analytic model for the information contained in a Doppler tracking pass in combination with the CEI information equations. These information models provide a common framework for a comparison of the capabilities of these data types, as a function of some of the key parameters that have an impact on navigation performance.

II. Analysis

Both wideband ΔVLBI and CEI measure the differenced one-way range from a spacecraft to two separate stations, which can be expressed simply as

$$\mathrm{DOR} = \underline{B} \cdot \underline{S} \tag{1}$$

where

$\underline{B}$ = baseline vector $= \underline{r}_1 - \underline{r}_2$

$\underline{r}_1$ = position vector of station 1

$\underline{r}_2$ = position vector of station 2

$\underline{S}$ = unit vector in spacecraft direction

The station locations and baseline components are expressed in cylindrical coordinates in an Earth-fixed frame:

$$\underline{B} = (r_B \cos\lambda_B,\ r_B \sin\lambda_B,\ z_B)$$

where

r_B = baseline component perpendicular to Earth's spin axis

z_B = baseline component parallel to Earth's spin axis

λ_B = baseline longitude

The baseline coordinates as a function of the station coordinates are as follows:

$$\begin{aligned} r_B &= \left[r_1^2 + r_2^2 - 2r_1 r_2 \cos(\lambda_1 - \lambda_2)\right]^{1/2} \\ z_B &= z_1 - z_2 \\ \lambda_B &= \mathrm{Atan}\left[(r_1 \sin\lambda_1 - r_2 \sin\lambda_2)/(r_1 \cos\lambda_1 - r_2 \cos\lambda_2)\right] \end{aligned} \tag{2}$$

where

r_1, r_2 = station location components perpendicular to Earth's spin axis

z_1, z_2 = station location components parallel to Earth's spin axis

λ_1, λ_2 = station longitudes

Using Eq. (2), the DOR observable, Eq. (1), and its partial derivatives can be expressed as a function of the baseline components and the spacecraft coordinates:

$$\mathrm{DOR} = r_B \cos\delta \cos H_B + z_B \sin\delta \tag{3}$$

where

δ = spacecraft declination

$H_B = \alpha_B - \alpha$ (baseline hour angle)

α_B = baseline right ascension

α = spacecraft right ascension

$$\begin{aligned} \partial\mathrm{DOR}/\partial\delta &= -r_B \sin\delta \cos H_B + z_B \cos\delta \\ \partial\mathrm{DOR}/\partial\alpha &= r_B \cos\delta \sin H_B \end{aligned} \tag{4}$$

A. Information Array for DOR Observables

Due to the manner of its implementation, wideband ΔVLBI is typically used to form a single measurement during the time period in which the spacecraft is in view at both stations. Since the baseline is only sensitive to angular displacement in one direction, wideband ΔVLBI observations are made in pairs using two baselines that are as near orthogonal as possible. This is the approach employed for Galileo navigation, with a pair of wideband ΔVLBI measurements being made every three days on average

using the Goldstone–Madrid and Goldstone–Canberra DSN baselines.

The development given below establishes an approximate analytic representation of the error covariance for a pair of wideband ΔVLBI measurements, using the DOR observable model. Since ΔVLBI observations are very sensitive to the effects of the troposphere on the received radio signal, it is desirable to make the measurements at the highest elevation possible for both antennas. To account for this in some respect, the baseline hour angle H is assumed to be 90 deg for both DOR measurements. This assumption simplifies the error covariance equations and has no significant effect on the results obtained from them.

The information content for each DOR measurement is formed using the partial derivatives given in Eq. (4); the two are then summed together to form the complete information array for the two observations. The information content of a single DOR measurement is as follows:

$$I_{\mathrm{DOR}} = \begin{bmatrix} \left(\partial \mathrm{DOR}/\partial\delta\right)^2 & \left(\partial \mathrm{DOR}/\partial\delta\right)\left(\partial \mathrm{DOR}/\partial\alpha\right) \\ \left(\partial \mathrm{DOR}/\partial\delta\right)\left(\partial \mathrm{DOR}/\partial\alpha\right) & \left(\partial \mathrm{DOR}/\partial\alpha\right)^2 \end{bmatrix} \left(2/\sigma^2_{\mathrm{DOR}}\right) \tag{5a}$$

Assuming that no a priori information exists, the error covariance matrix associated with the estimates of δ and α obtained from the DOR observation pair is simply the inverse of the information array:

$$\Gamma_{\mathrm{DOR}} = E\left\{\left(\underline{\mathrm{X}} - \underline{\hat{\mathrm{X}}}\right)\left(\underline{\mathrm{X}} - \underline{\hat{\mathrm{X}}}\right)^{\mathrm{T}}\right\}$$

$$= \begin{bmatrix} \sigma^2_\delta & \sigma^2_{\delta\alpha} \\ \sigma^2_{\delta\alpha} & \sigma^2_\delta \end{bmatrix} = \left(I_{\mathrm{DOR}}\right)^{-1} \tag{5b}$$

where

$$\underline{\mathrm{X}}^{\mathrm{T}} = (\delta, \alpha) \text{ (true values)}$$

$$\underline{\hat{\mathrm{X}}}^{\mathrm{T}} = (\hat{\delta}, \hat{\alpha}) \text{ (estimated values)}$$

Substituting Eq. (4) for Eq. (5a) in each measurement, then adding the two and inverting the resulting DOR information array yields the desired expressions for the elements of the error covariance matrix:

$$\sigma^2_\delta = \left\{ \frac{\left(r^2_{B1} + r^2_{B2}\right)}{\left[\left(r^2_{B1} + r^2_{B2}\right)\left(z^2_{B1} + z^2_{B2}\right) - \left(r_{B1}z_{B1} + r_{B2}z_{B2}\right)^2\right]\cos^2\delta} \right\} \left(\sigma^2_{\mathrm{DOR}}/2\right) \tag{6}$$

$$\sigma^2_\alpha = \left\{ \frac{\left(z^2_{B1} + z^2_{B2}\right)}{\left[\left(r^2_{B1} + r^2_{B2}\right)\left(z^2_{B1} + z^2_{B2}\right) - \left(r_{B1}z_{B1} + r_{B2}z_{B2}\right)^2\right]\cos^2\delta} \right\} \left(\sigma^2_{\mathrm{DOR}}/2\right) \tag{7}$$

$$\sigma^2_{\delta\alpha} = \left\{ \frac{-\left(r_{B1}z_{B1} + r_{B2}z_{B2}\right)}{\left[\left(r^2_{B1} + r^2_{B2}\right)\left(z^2_{B1} + z^2_{B2}\right) - \left(r_{B1}z_{B1} + r_{B2}z_{B2}\right)^2\right]\cos^2\delta} \right\} \left(\sigma^2_{\mathrm{DOR}}/2\right) \tag{8}$$

where

r_{B1}, z_{B1} = radial and z-height components for baseline 1

$r_{B2}, = z_{B2}$ = radial and z-height components for baseline 2

σ^2_{DOR} = variance of DOR observations

B. Information Array for CEI Tracking Pass

The CEI phase-delay observable, like the wideband ΔVLBI observable, is a measurement of differenced one-way range, so the information array for a single measurement is given by Eq. (5a). The information array and associated error covariance matrix for a pass containing N measurements, assuming a constant measurement variance of $\sigma^2{}_{\mathrm{CEI}}$, is then

$$I_{\mathrm{CEI}} = \left[\left(N/\sigma^2_{\mathrm{CEI}}\right) \sum_{i=1}^{N} \left(\partial \mathrm{DOR}_i/\partial \underline{X}\right) \left(\partial \mathrm{DOR}_i/\partial \underline{X}\right)^{\mathrm{T}} \right] \tag{9a}$$

$$\Gamma_{\mathrm{CEI}} = \left(I_{\mathrm{CEI}}\right)^{-1} \tag{9b}$$

where, once again

$$\underline{X}^{\mathrm{T}} = (\delta, \alpha)$$

Equation (9b) is analogous to Eq. (5b) for the DOR observation pair, and it also is based on the assumption that no a priori information exists. The summation in Eq. (9a) can be replaced with integration, assuming continuous observations. The resulting integrals can then be adjusted by a constant to account for the fact that the observations are discrete. For example, the information element for declination becomes

$$I_{\mathrm{CEI}\ 11} = \left(1/\omega^2{}_{\mathrm{CEI}}\Delta t\right) \int_{\mathrm{A}}^{\mathrm{B}} \left(\partial \mathrm{DOR}/\partial \delta\right)^2 dH \tag{10}$$

where

ω = Earth rotation rate

Δt = time interval between measurements

The variable Δt forces the integrated information element to be equivalent to the corresponding summed element. Since the variable of integration is the baseline hour angle H_B, careful consideration must be given to the choice of the limits A and B. The baseline hour angle can be defined in terms of the hour angle of one end of the baseline, chosen to be station 2 here, as follows:

$$H_B = H + \phi \tag{11}$$

where

H_B = baseline hour angle

H = station 2 hour angle

ϕ = baseline longitude relative to station 2

$$\phi = \lambda_B - \lambda_2 = \mathrm{Atan}\left[r_1 \sin(\lambda_1 - \lambda_2)/(r_1 \cos(\lambda_1 - \lambda_2) - r_2)\right] \tag{12}$$

Equation (12) was obtained by setting λ_2 to zero in Eq. (2) and replacing λ_1 with $(\lambda_1 - \lambda_2)$. Using Eqs. (11) and (12), the limits of integration for the elements of the information array can be expressed in terms of the hour angle of station 2. If it is further assumed that the two stations comprising the connected element are near enough to each other such that the hour angle for both stations is roughly the same, then the limits can be expressed in terms of a single station hour angle and the baseline angle ϕ:

$$A = -H + \phi$$

$$B = H + \phi \tag{13}$$

where

H = tracking pass half-width

The information array for a single, symmetric (in H) CEI tracking pass is then

$$I_{CEI} = \begin{bmatrix} I_{\mathrm{CEI}\ 11} & I_{\mathrm{CEI}\ 12} \\ I_{\mathrm{CEI}\ 12} & I_{\mathrm{CEI}\ 22} \end{bmatrix}$$

$$I_{\mathrm{CEI}\ 11} = \left(J_1 r_B^2 \sin^2\delta + J_2 z_B^2 \cos^2\delta - 2J_3 r_B z_B \sin\delta \cos\delta\right) / \left(\omega \sigma^2_{\mathrm{CEI}} \Delta t\right) \tag{14}$$

$$I_{\mathrm{CEI}\ 12} = \left(J_4 r_B z_B \cos^2\delta - J_5 r_B^2 \sin\delta \cos\delta\right) / \left(\omega \sigma^2_{\mathrm{CEI}} \Delta t\right)$$

$$I_{\mathrm{CEI}\ 22} = \left(J_6 r_B^2 \cos^2\delta\right) / \left(\omega \sigma^2_{\mathrm{CEI}} \Delta t\right)$$

where

$$
\begin{aligned}
J_1 &= H + (1/2)\sin 2H \cos 2\phi \\
J_2 &= 2H \\
J_3 &= 2\cos\phi \sin H \\
J_4 &= 2\sin\phi \sin H \\
J_5 &= \sin 2\phi \sin 2H \\
J_6 &= H - (1/2)\sin 2H \cos 2\phi
\end{aligned} \tag{15}
$$

It must be noted that Eqs. (14) and (15) were developed assuming that the declination δ and the right ascension α of the spacecraft are constant during the tracking period. While this is a reasonable assumption for a single pass, which takes place during the course of several hours, it is not applicable when it is desired to accumulate a sequence of CEI tracking passes taking place over a period of several days. To do so requires a more sophisticated model for the spacecraft state variables which includes the rates of change of α and δ.

The error covariance of the spacecraft angular coordinates can now be obtained by inverting the CEI information array, Eq. (14):

$$
\sigma_\delta^2 = \frac{H - (1/2)\sin 2H \cos 2\phi}{D_1 r_B^2 \sin^2\delta + D_2 z_B^2 \cos^2\delta + D_3 r_B z_B \sin\delta\cos\delta} \left(\omega\sigma_{\mathrm{CEI}}^2 \Delta t\right) \tag{16}
$$

$$
\sigma_\alpha^2 = \frac{[H + (1/2)\sin 2H\cos 2\phi]\, r_B^2 \sin^2\delta + 2H z_B^2 \cos^2\delta - 4\cos\phi \sin H r_B z_B \sin\delta\cos\delta}{D_1 r_B^4 \sin^2\delta\cos^2\delta + D_2 r_B^2 z_B^2 \cos^4\delta + D_3 r_B^3 z_B \sin\delta\cos^3\delta} \left(\omega\sigma_{\mathrm{CEI}}^2 \Delta t\right) \tag{17}
$$

$$
\sigma_{\delta\alpha}^2 = \frac{2\sin\phi\sin H z_B \cos\delta - \sin 2\phi \sin 2H r_B \sin\delta}{D_1 r_B^3 \sin^2\delta\cos\delta + D_2 r_B z_B^2 \cos^3\delta + D_3 r_B^2 z_B \sin\delta\cos^2\delta} \left(\omega\sigma_{\mathrm{CEI}}^2 \Delta t\right) \tag{18}
$$

where

$$
\begin{aligned}
D_1 &= H^2 - \sin^2 2H \left[\sin^2 2\phi + (1/4)\cos^2 2\phi\right] \\
D_2 &= 2H\left[H - (1/2)\sin 2H\cos 2\phi\right] - 4\sin^2\phi\sin^2 H \\
D_3 &= \sin H \sin 2H \left(2\sin\phi\sin 2\phi + \cos\phi\cos 2\phi\right) \\
&\quad - 2H\cos\phi\sin H
\end{aligned} \tag{19}
$$

C. Information Array for Doppler Tracking Pass

The Doppler information model given below was developed by Hamilton and Melbourne in their classic 1966 paper [3]. Over the course of a single tracking period, the range-rate to a distant spacecraft observed at a station can be closely approximated by the following expression:

$$
\dot\rho = \dot r + \omega r_s \cos\delta \sin H \tag{20}
$$

where

$\dot\rho$ = station–spacecraft range-rate

$\dot r$ = spacecraft geocentric range-rate

r_s = station component perpendicular to Earth's spin axis

In Eq. (20), H and δ are the spacecraft hour angle and declination, respectively, just as in the DOR and CEI tracking models. The Doppler signal is seen to be a function of the three spacecraft coordinates $\dot r$, δ, and α. The information array for a single, symmetric pass ($-H$ to $+H$) of data is

$$
\mathrm{I_{DOP}} = \left[(N/\sigma_{\dot\rho}^2)\sum_{i=1}^{N} (\partial\dot\rho_i/\partial\underline{\mathrm{X}})\,(\partial\dot\rho_i/\partial\underline{\mathrm{X}})^{\mathrm{T}}\right] = \mathbf{A}^{\mathrm{T}}\mathbf{J}\mathbf{A} \tag{21}
$$

where

$$
\underline{\mathrm{X}}^{\mathrm{T}} = (\dot r, \delta, \alpha)
$$

$$
\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -\omega r_s \sin\delta & 0 \\ 0 & 0 & -\omega r_s \cos\delta \end{bmatrix} \tag{22}
$$

$$\mathbf{J} = \begin{bmatrix} 2H & 0 & 2\sin H \\ 0 & H - (1/2)\sin 2H & 0 \\ 2\sin H & 0 & H + (1/2)\sin 2H \end{bmatrix} \times \left(\frac{1}{\omega \sigma_{\dot{\rho}}^2 \Delta t} \right) \tag{23}$$

Inversion of Eq. (21) yields expressions for the variance of the declination and right ascension estimates obtained from the tracking pass. In the equations given below, the effects of timing and station location errors on the estimates are included.

$$\sigma_\delta^2 = (1/\sin^2\delta) \left\{ \frac{(\sigma_{\dot{\rho}}^2 \Delta t/\omega r_s^2)}{[H - (1/2)\sin 2H]} + \cos^2\delta \, (\sigma_{r_s}^2/r_s^2) \right\} \tag{24}$$

$$\sigma_\alpha^2 = (\sigma_{\dot{\rho}}^2 \Delta t/\omega r_s^2 \cos^2\delta) \times \left\{ H/\left[H^2 + (1/2)H\sin 2H - 2\sin^2 H\right] \right\} + \omega^2 \sigma_\tau^2 + \sigma_\delta^2 \tag{25}$$

$$\sigma_{\delta\alpha}^2 = 0 \tag{26}$$

where

$\sigma_{\dot{\rho}}^2$ = range-rate measurement variance

Δt = time interval between data points

σ_τ^2 = variance of timing error

σ_λ^2 = variance of station longitude error

D. Tracking Pass Width as a Function of Declination

Before comparing CEI and VLBI navigation performance, the selection of an appropriate value for the tracking pass half-width, H, must be addressed. Edwards [2] has shown that random fluctuations in the troposphere are probably the dominant error source in determining CEI phase-delay measurement accuracy, and that the statistical uncertainty of this effect is inversely proportional to the elevation angle at which the spacecraft is being observed. His analysis indicates that the troposphere-induced measurement uncertainty drops off exponentially with increasing elevation, with most occurring within the elevation angle at which the spacecraft is being observed. His analysis indicates that the troposphere-induced measurement uncertainty drops off exponentially with increasing elevation, with most occurring within the elevation range 0–30 deg. To account for this variation, the tracking pass half-width H can be made to vary with declination such that the spacecraft is only observed when its elevation angle is above some minimum. An elevation cutoff of 30 deg was chosen for use here, based on the above discussion. At this elevation, the phase-delay uncertainty due to tropospheric fluctuations is about 5 mm. Admittedly this is a somewhat arbitrary selection; nevertheless, the use of this value should yield some reasonable indication of how the navigation performance obtained from CEI varies with declination.

To some degree this same argument also applies to Doppler data since they are also affected adversely by random tropospheric fluctuations. An elevation cutoff of 15 deg has been commonly used during past flight projects—and was used below for all scenarios involving Doppler. Since Doppler data can be readily acquired by all of the stations within the Deep Space Network (DSN), which are located in both the northern and southern hemispheres, the declination of the spacecraft being tracked is of little importance in terms of its effect on tracking pass width.

Elevation can be expressed directly as a function of declination, hour angle, and station location as

$$h = \mathbf{A} \sin\left[(r_s/\zeta)/\cos\delta\cos H + (z/\zeta)\sin\delta\right] \tag{27}$$

where

h = elevation angle

r_s = station location spin radius, as in Eq. (2)

z = station location z-height, Eq. (2)

$\zeta = (r_s^2 + z^2)^{1/2}$

Figure 1 shows the variation in tracking pass half-width with declination for minimum elevations of 10–30 deg. As seen in Fig. 1, the reduction in pass width becomes especially pronounced as declination drops below zero. These curves, or curves corresponding to other stations when appropriate, were used to determine the value of H to use in all of the results involving CEI and Doppler data which follow.

III. Results

A reference VLBI tracking scenario was created using the DOR error covariance equations, Eqs. (6) through (8),

to obtain the uncertainty for a pair of observations taken from the baselines formed by DSN Goldstone–Madrid (DSS 14–63) stations and Goldstone-Canberra (DSS 14–43) stations. These two baselines are the most nearly orthogonal set possible within the Deep Space Network. The measurement accuracy assumed for both DOR observations was 14 cm (1σ), which is the performance expected for the wideband ΔVLBI system carried by the Galileo spacecraft.[2] The baseline coordinates and results for this case are given in Table 1. In this scenario it was assumed that $\delta = 0$, making this a best-case result.

Actual wideband ΔVLBI measurements are constructed by differencing a spacecraft DOR observation with a second DOR observation of an extragalactic radio source (EGRS), which is normally a quasar. This results in the cancellation, or near-cancellation, of the effects of certain error sources that are common to both stations. The locations of quasars used for ΔVLBI are usually very well known, but not exactly known, and so this uncertainty must be considered in determining the overall accuracy of ΔVLBI measurements. Current quasar catalogs are known to be accurate to 5–15 nrad [4]. To account for this uncertainty in some sense, the results in Table 1 consist of the root-sum-square uncertainty of the right ascension (RA) and declination (DEC) uncertainties obtained from the ΔVLBI measurements plus a 15-nrad quasar location uncertainty.

A CEI tracking scenario, also for $\delta = 0$, was constructed for comparison, using a fictitious north–south baseline ($\phi = 0$) 20.5 km in length, with one end being DSS 15 at Goldstone. The baseline coordinates and results for this scenario are given in Table 2. A scale map of the Goldstone Deep Space Communications Complex (DSCC) appears in Fig. 2, showing DSS 15 at the north end.[3] Admittedly, the construction of a 20.5-km north–south baseline would require a station located outside of the present boundaries of the Goldstone DSCC, shown in the lower left-hand corner of Fig. 2. This fictitious station would not be far outside the current boundary, though, and since the construction of any operational CEI system would probably depend upon the construction of new stations in any event, this assumed station location is reasonable for the purposes of this analysis.

Actual CEI measurements will consist of a spacecraft phase-delay measurement differenced with a quasar phase-delay measurement, similar to ΔVLBI observations, and for essentially the same reasons—to obtain cancellation of common error sources. As with the VLBI results in Table 1, the CEI accuracy figures in Table 2 also contain a 15-nrad quasar location uncertainty in both RA and DEC, combined in a root-sum-square sense with the uncertainties obtained from the measurements themselves. It should be noted here that quasar location uncertainty was not a significant contributor to navigation error uncertainty in any of the VLBI or CEI cases considered herein.

Table 2 illustrates an important point: a single north–south CEI baseline can determine both declination and right ascension. Figures 3 and 4 show the effects of rotating the baseline away from a north–south orientation as a function of spacecraft declination. By the conventions chosen in defining ϕ, Eq. (12), a negative value of ϕ implies a rotation of the baseline to the west from north, while a positive value of ϕ implies a rotation to the east from north. Only curves for positive values of ϕ are shown, since a negative ϕ yields a covariance of the same magnitude as the corresponding ϕ of opposite sign. From these two figures, it can be seen that rotating the baseline away from a north–south orientation causes some degradation in both σ_δ and σ_α, but it is not significant except for low (less than $-10°$) declination and for baseline angles greater than roughly $\pm 20°$. Figure 4 shows that the capability of a north–south baseline to determine RA decreases rapidly with decreasing declination. Finger and Edwards,[4] in their numerical investigation of CEI tracking performance from Goldstone baselines, obtained similar results, but their analysis was restricted to declinations of $-23°$, $0°$, and $+23°$.

If a third station were added to the single baseline two-station CEI system envisioned thus far, two separate baselines could be formed into a connected element "triad." Figure 5 is a diagram showing such a hypothetical system at Goldstone, with the third station situated just outside the southeast corner of the complex. This system incorporates an east–west baseline with the original north–south baseline. Table 3 contains the baseline coordinates of the system and the tracking results which would be obtained for a spacecraft at $0°$ declination. In these scenarios, the error covariance for the combined tracking solution from the two baselines is computed by adding the information arrays obtained for each individual baseline, then inverting the result. The combined information from the two baselines reduces the RA uncertainty obtained with just the north–south baseline from 107 nrad to about 42 nrad,

[2] Ibid.

[3] Based on *Directory of Goldstone Buildings and Facilities*, JPL Project Document 890-165, Rev. A (internal document), Jet Propulsion Laboratory, Pasadena, California, October 1, 1989.

[4] M. H. Finger and C. D. Edwards, "Relative CEI Navigation Performance of Goldstone Intracomplex Baselines," Interoffice Memorandum 335.3-88-116 (internal document), Jet Propulsion Laboratory, Pasadena, California, October 20, 1988.

a factor of 2.5 improvement. This suggests that a CEI triad system could determine both RA and DEC to better than 50 nrad, using data taken only by stations at a single complex—in this case, Goldstone.

When used in conjunction with other data types, a single north–south CEI baseline may yet be a valuable navigation tool. The use of X-band (8.4 GHz) frequencies for both the uplink and downlink communications legs with the Magellan spacecraft, and later for Galileo and Mars Observer, should yield Doppler measurements of range-rate accurate to 0.1 mm/sec or better, except in cases where the Sun-Earth-probe angle is less than 15°.[5] Current Doppler accuracy for S-band (2.3 GHz) spacecraft communications is in the 0.5–1.0 mm/sec range. Table 4 describes the navigation accuracy which would be obtained with a combination of X-band Doppler passes from stations at the DSN sites in Madrid, Spain, and Canberra, Australia, with CEI data from Goldstone, again for a spacecraft at 0° declination. The values used in Table 4 for timing error (σ_τ) and station spin radius (σ_{r_s}) and longitude errors (σ_λ) are somewhat optimistic by current standards, but they should be attained within the next few years. As in the CEI triad scenario, the error covariance for the combined CEI-Doppler solution is simply the inverse of the summed information obtained from each data type. The addition of the Doppler data in this case results in a factor of 2 improvement in RA uncertainty over the performance of the north–south CEI baseline alone, from 107 nrad to 53 nrad.

Figure 6 is a comparison of the error uncertainty ellipses for the tracking scenarios given in Tables 1–4. The orientations of the VLBI ellipse reflects the orientations of the baselines that comprise the system, while it is apparent that the CEI systems and the CEI–Doppler combination both produce uncorrelated error covariances for δ and α. Table 5 provides a comparison of the navigation performance obtained from each of the three tracking systems discussed here, ΔVLBI, CEI triad and CEI-Doppler, as declination varies from −20° to +20°. Figures 7 and 8 present the information in Table 5 graphically. In these tables and figures, "CEI+Dop" refers to the tracking system described in Table 4, which uses data from a single north–south Goldstone CEI baseline in combination with Doppler data from Madrid and Canberra. Remember that in all of these cases, the tracking pass half-width is also changing with declination, so the variations seen in Table 4 are actually a function of two parameters, H and δ, not just δ.

[5] P. W. Kinman and J. G. Meeker, "Two-Way Coherent Doppler Measurement Accuracy for Venus Radar Mapper," Interoffice Memorandum 3392-84-94 (internal document), Jet Propulsion Laboratory, Pasadena, California, July 6, 1984.

IV. Conclusions

This study used simple, approximate mathematical models to investigate the relative tracking performance obtainable with two interferometric methods—a hypothetical CEI triad system at Goldstone and the wideband ΔVLBI system which will be used by the Galileo spacecraft—and a third hybrid system that uses a single CEI baseline in conjunction with conventional Doppler. The analysis performed indicates that the Galileo ΔVLBI system is capable of determining both RA and DEC to 20–25 nrad throughout the ecliptic plane. The hypothetical Goldstone intracomplex CEI system, which comprises three stations, is capable of producing sub-50–nrad estimates of DEC throughout the ecliptic plane, and RA estimates of 50 nrad accuracy or better down to a declination of about −10°. Data from even a single Goldstone connected element interferometer, when combined with X-band Doppler data from two other DSN stations, are capable of determining DEC to better than 50-nrad throughout the ecliptic plane and RA to 50–70 nrad.

The results of this study show that intracomplex CEI is a viable alternative to ΔVLBI for missions requiring spacecraft angular coordinate determination to an accuracy of 30–50 nrad. While the navigation accuracy achievable with intracomplex CEI from the Goldstone DSCC suffers somewhat at very low declinations, it has been shown that even a single 20.5-km baseline at Goldstone can potentially deliver better than 70-nrad accuracy when used in conjunction with X-band (8.4 GHz) Doppler. If a CEI system were also built at the DSN Canberra complex, this low declination problem would be eliminated. Considering the speed and reliability with which phase-delay data can be generated and processed for navigation, CEI can provide the DSN with a tremendous capability, in terms of both the number of missions that can be supported simultaneously and the navigation accuracy that can be achieved.

References

[1] C. D. Edwards, "Angular Navigation on Short Baselines Using Phase Delay Interferometry," *IEEE Transactions on Instrumentation and Measurement*, vol. 38, no. 2, pp. 665–667, April 1989.

[2] C. D. Edwards, "The Effect of Spatial and Temporal Wet-Troposphere Fluctuations on Connected Element Interferometry," *TDA Progress Report 42-97*, vol. January–March 1989, Jet Propulsion Laboratory, Pasadena, California, pp. 47–57, March 15, 1989.

[3] T. W. Hamilton and W. G. Melbourne, "Information Content of a Single Pass of Doppler Data from a Distant Spacecraft, *JPL Space Programs Summary*, vol. 3, no. 37–39, Jet Propulsion Laboratory, Pasadena, California, pp. 18–23, March–April 1966.

[4] 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, June 1988.

Table 1. VLBI tracking scenario results ($\delta = 0$)

Baseline	r_B, km	z_B, km
Goldstone–Madrid	8378.986	−341.545
Goldstone–Canberra	7620.841	7351.800

Measurement accuracy σ_{DOR} = 14 cm (approx. 20 nrad)

Quasar direction uncertainty (1σ) = 15 nrad (in both declination and right ascension)

σ_δ = 24.2 nrad
σ_α = 19.4 nrad

Table 2. CEI tracking scenario results ($\delta = 0$)

Baseline: r_B = 11.467 km, z_B = 17.000 km, ϕ = 0, length = 20.61 km
(Station 1 is DSS 15, Goldstone, California)

Tracking pass half-width H = 52.3°
(elevation cutoff = 30°)

Time interval between measurements Δt = 360 sec

Earth rotation rate $\omega = 7.292 \times 10^{-5}$ rad/sec

Measurement accuracy σ_{CEI} = 5 mm (approx. 250 nrad)

Quasar direction uncertainty (1σ) = 15 nrad

σ_δ = 38.4 nrad
σ_α = 107 nrad

Table 3. CEI triad system performance ($\delta = 0$)

Baseline	r_B, km	z_B, km	ϕ, deg	Length, km
1: North–south	11.647	17.000	0.0	20.61
2: East–west	12.000	0.0	90.0	12.00

For all baselines:
H = 52.3° (elevation cutoff = 30°), Δt = 360 sec, σ_{CEI} = 5 mm

$\omega = 7.292 \times 10^{-5}$ rad/sec
Quasar direction uncertainty (1σ) = 15 nrad

σ_δ = 31.6 nrad
σ_α = 42.4 nrad

Table 4. Combined X-band Doppler–CEI performance ($\delta = 0$)

Doppler data: Station	r_B, km	z, km	λ, deg	H, deg
DSS 42, Canberra	5205.352	−3674.582	148.9813	71.5
DSS 62, Madrid	4860.817	4116.906	355.6322	70.2

CEI baseline:
r_B = 11.467 km, z_B = 17.000 km, ϕ = 0°
H = 52.3° (elevation cutoff = 30°), Δt = 360 sec, σ_{CEI} = 5 mm

Δt = 60 sec, $\sigma_{\dot\rho}$ = 0.1 mm/sec, elevation cutoff = 15°
$\omega\sigma_\tau$ = 20 nrad, σ_λ = 60 nrad, σ_{r_s} = 30 cm

σ_δ = 35.3 nrad
σ_α = 52.9 nrad

Table 5. Relative performance comparison between systems

δ(°)	DEC uncertainty, nrad			RA uncertainty, nrad		
	VLBI	CEI triad	CEI+Dop	VLBI	CEI triad	CEI+Dop
−20	25.2	40.7	46.6	19.9	68.3	67.5
−15	24.8	35.2	39.7	19.7	55.1	62.7
−10	24.4	32.9	37.4	19.5	48.7	58.5
−5	24.3	31.9	36.1	19.5	44.9	55.3
0	24.2	31.6	35.3	19.4	42.4	52.9
5	24.3	31.7	34.8	19.5	40.9	51.2
10	24.4	32.3	34.7	19.5	40.0	50.0
15	24.8	33.2	34.6	19.7	39.6	49.3
20	25.2	34.5	34.7	19.9	39.6	48.9

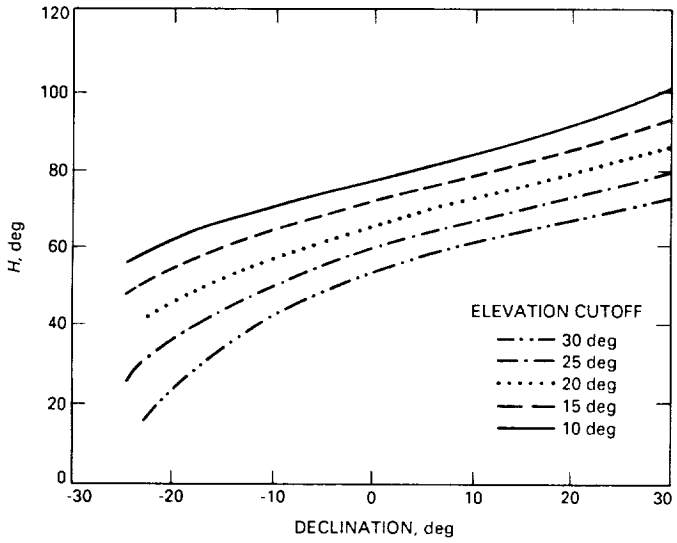


Fig. 1. Tracking pass half-width versus spacecraft declination for different elevation cutoffs (DSS 15, Goldstone, California).

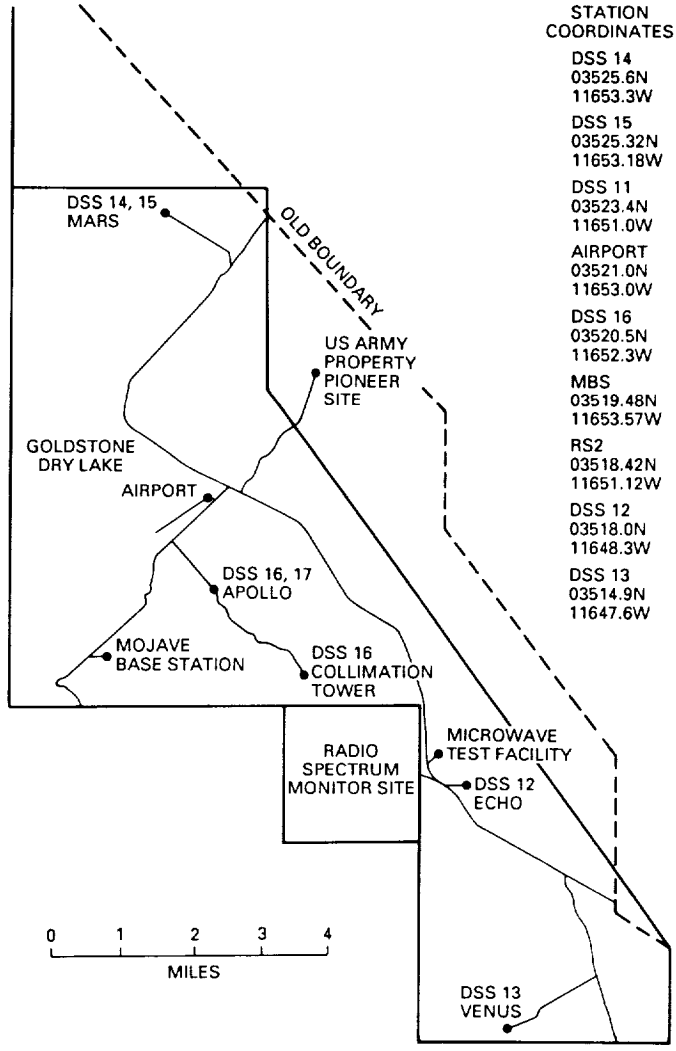


Fig. 2. Goldstone Deep Space Communications Complex.

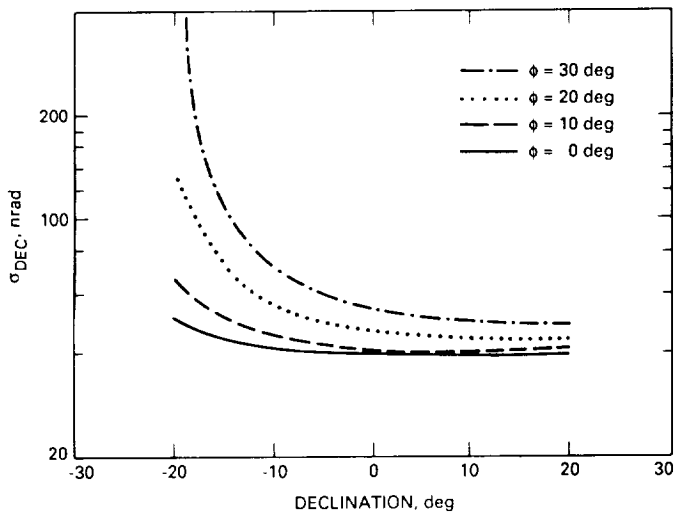


Fig. 3. DEC uncertainty versus declination for a north-south Goldstone CEI baseline.

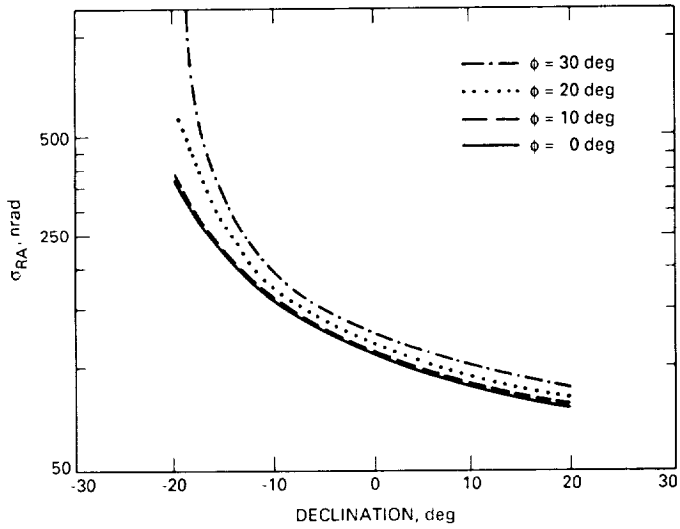


Fig. 4. RA uncertainty versus declination for a north-south Goldstone CEI baseline.

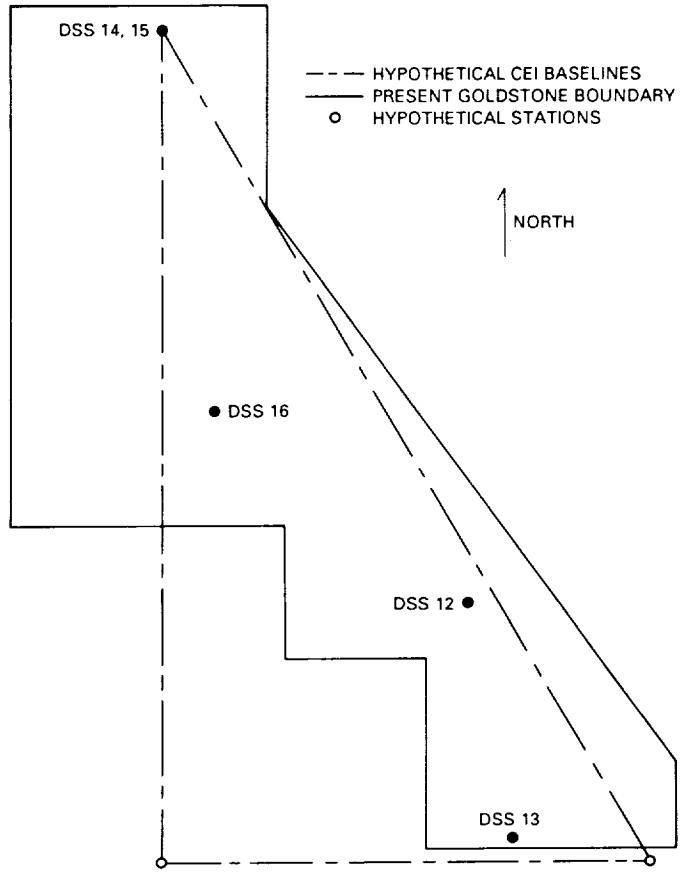


Fig. 5. Layout of hypothetical Goldstone CEI triad system.

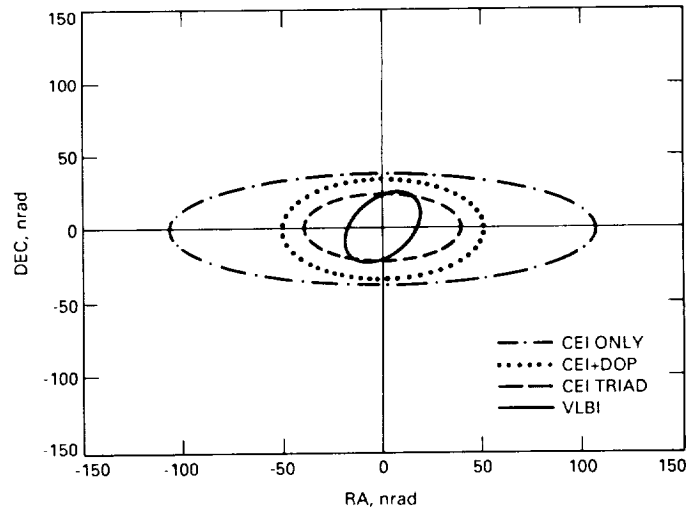


Fig. 6. Uncertainty ellipses for CEI and VLBI tracking scenarios ($\delta = 0°$).

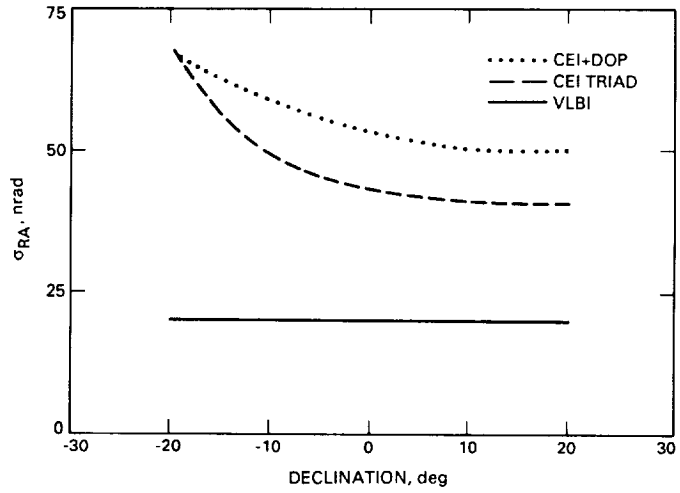


Fig. 7. RA uncertainty versus declination for VLBI, CEI triad, and CEI+Doppler systems.

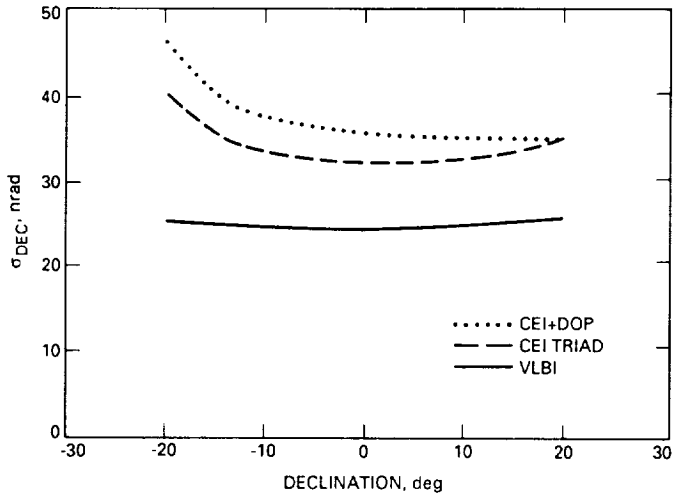


Fig. 8. DEC uncertainty versus declination for VLBI, CEI triad, and CEI+Doppler systems.

Applications of Different Design Methodologies in Navigation Systems and Development at JPL

S. W. Thurman
Navigation Systems Section

The NASA/JPL deep space navigation system consists of a complex array of measurement systems, data processing systems, and support facilities, with components located both on the ground and on board interplanetary spacecraft. From its beginnings nearly 30 years ago, this system has steadily evolved and grown to meet the demands for ever-increasing navigation accuracy placed on it by a succession of unmanned planetary missions. Principal characteristics of this system are its unique capabilities and great complexity. In this article, three examples in the design and development of interplanetary space navigation systems are examined in order to make a brief assessment of the usefulness of three basic design theories, known as normative, rational, and heuristic. Evaluation of the examples indicates that a heuristic approach, coupled with rational-based mathematical and computational analysis methods, is used most often in problems such as orbit determination strategy development and mission navigation system design, while normative methods have seen only limited use in such applications as the development of large software systems and in the design of certain operational navigation subsystems.

I. Introduction

The ability to accurately navigate a spacecraft often plays a significant role in the success or failure of unmanned interplanetary space missions. In the context of the NASA planetary exploration program, navigation is defined as the process of determining the current and predicted flight path of a space probe, and controlling that flight path to meet stated mission objectives [1,2]. The navigation system to be used for each mission is developed and configured during pre-flight mission design and planning, based on the trajectory design developed as part of the same process. The result is a mission-specific system which is composed of several basic "building blocks" common to all missions and elements unique to the mission at hand.

Figure 1 shows a simplified diagram of the "generic" deep space navigation system.[1] As is evident from Fig. 1,

[1] C. E. Kohlhase, "Navigation Systems Overview," Presentation Viewgraphs, Navigation Systems Section, Jet Propulsion Laboratory, Pasadena, California, March 16, 1973.

the principal traits of the system are its great complexity and unique capabilities (it is a special purpose system and essentially the only one of its kind). Each planetary mission has its own special capabilities, flight path, and science objectives, which make it necessary to develop a custom navigation system for every mission flown, even though these systems share many elements in common such as the tracking facilities of the Deep Space Network (DSN). Some examples of the building blocks used in deep space navigation are orbit determination and maneuver analysis software, spacecraft propulsion and imaging systems, and radio tracking networks (primarily the DSN).

What follows is a brief evaluation of three examples that show how different system design methodologies—normative, rational, and heuristic—are used in the design and development of navigation system software, orbit determination strategy and methods, and the design of navigation systems for specific missions. Before beginning, a brief description of each of the three design theories is in order. The descriptions given here are those put forth by Rechtin [3]. A normative theory is one in which system design is accomplished by following a set of rules and principles that are rooted in the values of the creator(s) of the theory. The definitions of what constitutes "good" and "bad" designs are largely judgmental pronouncements. A rational design theory is based on the idea that some generalized set of procedures for design and problem solving can be used to develop any system design, regardless of the system's purpose or functions. Rational theories typically make extensive use of mathematical analysis-based tools, such as the calculus of variations and probability theory. The heuristic approach to design shares some commonality with both the normative and rational approaches, but it is based more on insights and guidelines derived from experience rather than on rules and pronouncements or mathematical methods. Studying past and present applications of these design methodologies may provide some insight into their use for architecting the navigation systems of future planetary missions.

II. Example 1: Development of the JPL Orbit Determination Program

The Orbit Determination Program (ODP) is actually a large set of programs used to process radio tracking data and spacecraft onboard optical data, then construct a flight path which fits all of the observational data included in the solution. The design and evolution of this system provide an opportunity to evaluate the usefulness of some of the software engineering disciplines in developing very large, computationally intensive software.

Although it is difficult to rigidly classify many software engineering methods as being solely normative, rational, or heuristic, certain ones can be identified fairly closely with a single design methodology. One of the better known techniques in software engineering is "top-down" development, which consists of a set of guidelines, based on experience, for recursively partitioning a large problem into smaller ones in a hierarchical manner. Top-down development is primarily heuristic in nature [4]. Structured coding, on the other hand, is a programming technique that is somewhat heuristic, but has many traits that match the description of a normative theory given above; that is, it consists of a set of rules and principles that are heavily experience-based and largely judgmental.

The basic idea behind structured programming is that code written with only a specific, well-defined set of control constructs and principles will result in the best possible program—best being defined primarily in terms of readability and ease of maintenance, which were considered by the creators of the discipline to be of overriding importance [5]. If several different programmers were given a program design and using structured coding asked to write code to implement it, the end product arrived at by each programmer should be nearly the same. Other important criteria deemed good in structured programming are small, single-purpose program modules, each having only one entry point and one exit point.

The single most outstanding trait of the ODP throughout its 27-year history is its tremendous complexity. Even in its original form in 1962, the ODP contained programs capable of calculating spacecraft trajectories throughout a sizeable portion of the solar system and solving for up to 63 parameters using 13 different tracking data types [6,7]. While the description which follows is of the original first-generation ODP, the basic structure and organization of the system is preserved in the current ODP, even though the system is run differently on today's computers.

The basic design of the system can be seen in Fig. 2, which shows a high-level view of the organization of ten programs, known as links, comprising the first-generation ODP [7]. Figure 2 shows that the ODP was organized in a top-down, hierarchical fashion even in the early 1960's when software engineering was still in its infancy. The system was broken down in such a way that each major function in the orbit determination process was performed by a stand-alone program. Each program used output files generated by the previous programs as inputs. A single set of user-supplied instructions was used to execute all of the programs, which were run sequentially. This design espouses such heuristics as the matching of form to

function and the use of system elements with high internal complexity and low external complexity [3], resulting in a system that performed many complicated functions but was relatively simple to use.

Of primary importance for programs of such size and complexity is the efficient use of computer memory (this was especially true in the first ODP, which was run on an IBM 7090 computer with limited memory) and the minimization of communications between subprograms. Minimizing communications also minimizes the amount of time required to run the various links, as the individual programs which make up the ODP are known. Even when using modern computers with virtual memory capabilities (the ability to run a program of unlimited size), in order to minimize run time, it is important to keep to a minimum the memory required to run a given program so that the computer spends as little time as possible swapping inactive program segments and active segments in and out of its memory.

The modern ODP system, now in its third generation at JPL, was designed and written using top-down design and structured coding principles. It has been used successfully in all planetary missions dating back to the Mariner 6 and 7 missions to Mars in 1969: Mariner–Mars '69, Mariner–Mars '71, Mariner–Venus–Mercury '73, Pioneers 10 and 11, Viking, Voyager, Pioneer–Venus Orbiter, and currently the Magellan and Galileo missions. However, there are two goals of the ODP that have continually been in conflict with the rules of structured coding—minimizing run time and storage use. The ODP source code was written in FORTRAN, and extensive experience in the use of FORTRAN for computationally intensive applications has led to the development of heuristic guidelines for the optimization of FORTRAN programs in terms of run time and storage use [8]. Examples of guidelines for storage and run time optimization are the use of COMMON storage blocks for sharing data among multiple subprograms, minimizing memory usage and subprogram communication, and the use of a minimum number of subprograms to accomplish necessary tasks, since communication among subprograms is very slow relative to most operations performed by FORTRAN.

Some ODP links make extensive use of COMMON, have many subprograms that are quite large (thousands of lines), and have many subprograms with multiple entry points. While the actual code sometimes differs from what is considered to be good code according to structured programming rules, it was implemented in this way so that the links could be run quickly (run times in minutes), and so the software could be accommodated on computers with limited memory capacity. In the early history of the ODP, the concern with regards to memory usage was caused by the limited capability of the mainframe computers in use at the time (the IBM 7090). More recently, the motivation for minimizing storage usage was the desire to use the ODP on smaller minicomputers for dedicated use by designated groups of users. The rigorous use of structured code, with its small, self-contained single-purpose modules, may have yielded an ODP system that would be more readable and maintainable than it is now; but this probably would be of little comfort to the users of the software when faced with run times measured in hours and the need for large, expensive mainframe computers to run the ODP system.

The implementation of the ODP is a balance between the good qualities (readability, maintainability) of structured coding and the special considerations of memory and processing time requirements. It can be seen that while structured programming did play a role in the development of the ODP, the architects of this system tempered the rules of structured coding with previous experience, a more heuristic approach to programming.[2] In summary, the ODP is a result of compromises made by its architects, who had to reconcile the requirements of their system with the rules and guidelines of their software engineering tools.

The modern ODP has evolved into a multi-mission orbit determination tool. It has also been adapted for use on a variety of computers, from mainframes to desk-top workstations. For example, the DSN Multi-Mission Navigation Team runs the ODP system on a VAX 8530 minicomputer, supporting missions for NASA and several international space agencies. The Galileo Navigation Team is running the ODP in its traditional "home" environment, a UNIVAC 1108 mainframe. The Magellan Navigation Team, on the other hand, is presently running the ODP on its own dedicated computer system, centered around a Sun 3/150 workstation. Although each of these user groups is running the same set of programs, each group uses the ODP in a different manner, that is, with different input and output data to meet its own special needs.

III. Example 2: The Orbit Determination Strategy Design Process

This example provides a brief illustration of the process through which orbit determination strategy is conceived and developed. The development of an orbit determination strategy basically consists of choosing such things as

[2] Interview with John E. Ekelund, ODP cognizant engineer, Navigation Systems Section, Jet Propulsion Laboratory, Pasadena, California, November 2, 1989.

the quantity and types of measurements to be used (Earth-based ranging, Doppler, spacecraft onboard optical, etc.), the type of estimation algorithm employed (batch, sequential, reduced-order, etc.), and the mathematical model to be used by the estimation algorithm. This activity usually takes place as part of a feasibility study, when a new mission is being considered. At this early stage, there is usually little consideration given to any constraints that may be imposed by a specific spacecraft, since a firm spacecraft design is usually not in existence yet. The result of this process is usually a set of preliminary navigation accuracy requirements and an orbit determination strategy (or set of strategies) capable of meeting them. As might be expected, these requirements are often a compromise between what is desired and what is really possible.

Because of its complicated nature, the design and development of orbit determination strategy is primarily a heuristic process, but one that also makes extensive use of rational-based analysis tools to serve as a guide. The description that follows delineates why this is so and gives a brief history of this field as practiced at JPL, using a few specific examples for illustration. Orbit determination is a classic example of the limitations of mathematical tools: There are none that can tell the analyst what the best navigation system will be for a given spacecraft flying a given trajectory. The navigation analyst can choose from a number of different kinds of measurements, each with its own strengths and weaknesses, which can be combined into an orbit determination strategy in a great variety of ways. The number of possibilities quickly destroys any hopes of constructing a mathematical search procedure to seek out and find the "best" system.

The mathematical tools used to exercise and evaluate candidate strategies normally come from optimal estimation theory, a body of knowledge describing how to obtain the best possible estimate of a system given a mathematical description of that system [9,10]. Estimation theory is rational in nature; subject to certain assumptions, it theoretically specifies the estimation algorithm that will yield the best estimate, in a statistical sense, of any system which can be described using a basic mathematical framework. The user only needs a mathematical model for the system in question and some knowledge of pertinent mathematical methods (matrix algebra, linear systems of differential equations) to define the optimal estimator. In orbit determination, the system consists of a set of parameters describing the spacecraft trajectory, tracking station locations, and numerous error sources.

Although it is very powerful, optimal estimation theory has two fundamental limitations in orbit determination applications (and many other fields for that matter). The first is that it gives its user only one accuracy estimate at a time for a single measurement strategy; for example, it does not indicate how much performance improvement would be obtained by the addition of more data and/or different types of measurements in the solution. The second limitation is that the optimal estimates computed are only correct if the mathematical model of the system is correct, which in practice it never really is.

The sensitivity of the results obtained from the estimation techniques used by JPL to unmodeled or poorly modeled parameters were discovered early on in the planetary exploration program. Spacecraft such as Mariner-Mars '65 (Mariner 4) experienced unexpected deviations from the estimated flight path, which were later determined to be caused by small gas leaks in the valves of the attitude control thrusters [11]. Gas leaks and small thruster misalignments—as well as many other effects which are known to exist but are extremely difficult to model—are now known to be present on all interplanetary spacecraft. Another example is the effect of small variations in the Earth's rotation rate on the timing of tracking measurements. A mathematical method, known as consider state analysis, was developed to estimate the effects on orbit determination accuracy of parameters that were known or suspected to influence the problem but were too poorly known to be estimated themselves [7].[3]

While consider state analysis is capable of estimating the effects of parameters not present in the system model, it provides no guide to the navigation analyst about how to change the orbit determination strategy to minimize the influence of these parameters. The general rules given by optimal estimation theory which apply to the behavior of changes in solution accuracy no longer apply when consider states are taken into account; for example, according to optimal estimation theory, if more data are added to a trajectory solution, then the accuracy of that solution must increase (it does not say how much, just that it must increase). Also, if the data used in determining a trajectory are made to be more accurate, then the accuracy of the resulting estimate must improve. The author has personally experienced the contradictions to these well-established rules which commonly arise in consider state analysis.

As discussed above, there are no well-defined rules that tell the analyst which direction to follow when searching

[3] N. D. Panagiotacopulos, *An Introduction to JPL's Orbit Determination Program*, JPL Document 1846-37 (internal document), Jet Propulsion Laboratory, Pasadena, California, pp. 28-32, May 21, 1974.

for a good orbit determination strategy. Over the years, a small body of knowledge has been collected on the effects of certain poorly modeled parameters, such as spacecraft gas leaks, station location errors and clock errors on deep space navigation measurements. This knowledge, in the form of heuristics developed from flight experience and analysis of simple analytic approximations of tracking measurements, is used by navigation analysts when designing and developing orbit determination strategies. These heuristics are occasionally written down, usually as guidelines gleaned from the analysis of simple models ([12,13] are two examples), but they exist mostly in the minds of the individuals who learned through a great deal of experience how to use them.

IV. Example 3: Navigation System Design for Mission Operations

This final example describes an activity that begins roughly at the point where preliminary orbit determination strategy design, described in Example 2, leaves off. At this point in mission planning, a set of navigation requirements and a tentative orbit determination strategy, or set of strategies, has been developed as a part of preliminary mission design activities. Both the requirements and the orbit determination strategy chosen to meet them may be modified somewhat during the process of designing the operational navigation system. The end product of this design phase is a detailed navigation plan, specifying the number and type of measurements needed throughout the mission, the model to be used operationally in the orbit determination software (the ODP), and an exhaustive set of computer simulations demonstrating compliance with the requirements over the entire mission (see Mohan and Kirhofer for examples).[4,5]

As in so many other aspects of space navigation, the essence of operational navigation system design is to bring order to a very complex situation. Although there may be similarities among different planetary missions, no two are ever the same; consequently the navigation system designed for each spacecraft must be a special purpose system that is tailored to meet the specific requirements of a particular mission. This implies that mission navigation system design is heuristic in nature, requiring knowledgeable system architects who use their previous experience to provide guidelines which make sense in the context of the problem at hand [3], but who do much more than just follow a set of rules based on previous experience (a normative approach) or use a formula which yields a good system (a rational approach).

Figure 3 shows the functional organization of the navigation system used in the Viking mission to Mars [14], known as the Flight Path Analysis Group (FPAG). This system consisted of people, hardware, and software. The navigation team, which was responsible for carrying out all of these functions, was organized in the same manner as seen in Fig. 3. Each functional block was implemented as a group of people and equipment whose job it was to perform all of the functions in that block. Just as in the architecture of the ODP, form matched function. Notice that the architecture of the navigation system in Fig. 3 has some elements that are common to all deep space missions, such as the tracking data conditioning team, while other elements present are designed to fulfill the specific needs of the Viking mission, such as the lander flight path analysis team. Obviously, one would not expect to find such a group in the Voyager navigation team, whose mission consists solely of planetary flybys. This brief example suggests that operational navigation systems have some components which are unique to a specific mission and others which are very nearly the same across different missions.

While the navigation system for a given mission will not be exactly the same as that of another, there are certain subsystems within the overall navigation system which change very little from mission to mission, as pointed out above. Subsystems that have become formalized structures can be used in virtually all missions with very few changes required. Continuing with the previous example, a subset of the navigation system used for the Viking mission is shown in Fig. 4 (also see [14]). This diagram shows the "flow" of the orbit determination process as envisioned and implemented by the Viking navigation team. This design is the result of many years of flight experience. The majority of Fig. 4 depicts functions performed with the orbit determination software. Since the orbit determination process requires infrequent changes (which may be brought about with the introduction of a new tracking data type, for example), the design of the orbit determination subsystem for a new mission is more of a normative process; it consists of merely arranging the required functions, shown in Fig. 4, in the proper sequence and station in the operations flow.

A good example of the applicability of the orbit determination process (Fig. 4) to a variety of different missions can be found in the DSN Multi-Mission Naviga-

[4] S. N. Mohan, *Magellan Navigation Plan*, Magellan Project Document 630-51 (internal document), Revision B, Jet Propulsion Laboratory, Pasadena, California, March 23, 1988.

[5] *Galileo Navigation Plan*, Galileo Project Document 625-566 (internal document), Revision A, Jet Propulsion Laboratory, Pasadena, California, October 1989.

tion Team's activities. As mentioned earlier, this group supports a large number of both domestic and international missions. The orbit determination operations architecture implemented by the DSN Multi-Mission Navigation Team consists of a basic set of tasks or functions comprising a "generic" mission operations scenario, which looks almost exactly like the diagram in Fig. 4. Although this "generic" architecture is modified slightly to meet the needs of each mission, it is the foundation upon which all mission-specific orbit determination systems used operationally are built.

V. Summary and Conclusions

Three examples of system design and development used in the field of deep space navigation have been briefly examined to make some assessment of the role of the normative, rational, and heuristic design theories in this area. The examples studied were the design of the JPL orbit determination software system, the design and evaluation of orbit determination strategies, and the design of the operational navigation system and orbit determination subsystem for the Viking mission to Mars. The examples show that all three of the design theories studied are used in some capacity in this field.

One of the primary characteristics of heuristic design theory is that the architect using it must possess a body of expertise that is relevant to the specific context of the system to be designed. By this standard, the examples considered indicate that while the design of certain subsystems can be done using a normative approach to design, not requiring a great deal of expertise from the architect, the design of deep space navigation systems at JPL has been accomplished using a primarily heuristic approach due to the complexity of the problems and the specialized nature of the functions to be performed. There are some rational theories, such as optimal estimation theory, which play an important but supporting role in the design process. The ultimate test of a design theory's effectiveness is the success or failure of the resulting systems; the success of a great number of unmanned planetary missions, from Mariner to Voyager, appears to indicate that the heuristic approach to design has produced deep space navigation systems which have worked very well.

References

[1] J. F. Jordan, "Deep Space Navigation Systems and Operations," European Space Agency International Symposium on Spacecraft Flight Dynamics, May 1981.

[2] L. J. Wood and J. F. Jordan, "Interplanetary Navigation in the 1980's and 1990's," Paper AAS 81–113, AAS/AIAA Astrodynamics Conference, Lake Tahoe, Nevada, August 1981.

[3] E. Rechtin, *Systems Architecting*, University of Southern California, Los Angeles, California, 1989 (in press).

[4] P. T. Ward and S. J. Mellor, *Structured Development for Real-Time Systems, Vol. 1: Introduction and Tools*, New York: Yourdon Press, 1985.

[5] T. de Marco, *Concise Notes on Software Engineering*, New York: Yourdon Press, 1979.

[6] D. B. Holdridge, *Space Trajectories Program for the IBM 7090 Computer*, Technical Report No. 32–223, Jet Propulsion Laboratory, Pasadena, California, March 2, 1962.

[7] M. R. Warner, M. W. Nead, and R. H. Hudson, *The Orbit Determination Program of the Jet Propulsion Laboratory*, Technical Memorandum No. 33–168, Jet Propulsion Laboratory, Pasadena, California, March 18, 1964.

[8] M. Metcalf, *FORTRAN Optimization*, New York: Academic Press, 1982.

[9] J. D. Anderson, "Trajectory Determination from Observation Data," from *Recent Developments in Space Flight Mechanics*, vol. 9, *American Astronautical Society Science and Technology Series*, pp. 133–158, 1966.

[10] A. Gelb, ed., *Applied Optimal Estimation*, Cambridge, Massachusetts: M.I.T. Press, 1974.

[11] G. W. Null and H. J. Gordon, *The Mariner IV Flight Path and Its Determination From Tracking Data*, Technical Report 32–1108, Jet Propulsion Laboratory, Pasadena, California, pp. 8-17, August 1, 1967.

[12] T. W. Hamilton and W. G. Melbourne, "Information Content of a Single Pass of Doppler Data From a Distant Spacecraft," *JPL Space Programs Summary No. 37–39*, vol. III, Jet Propulsion Laboratory, Pasadena, California, pp. 18–23, May 31, 1966.

[13] J. F. Jordan, G. A. Madrid, and G. E. Pease, "The Effects of Major Error Sources on Planetary Spacecraft Navigation Accuracies," AIAA Paper 70–1077, AAS/AIAA Astrodynamics Conference, Santa Barbara, California, August 19–21, 1970.

[14] W. J. O'Neil, *Viking Navigation*, JPL Publication 78–38, Jet Propulsion Laboratory, Pasadena, California, November 15, 1979.

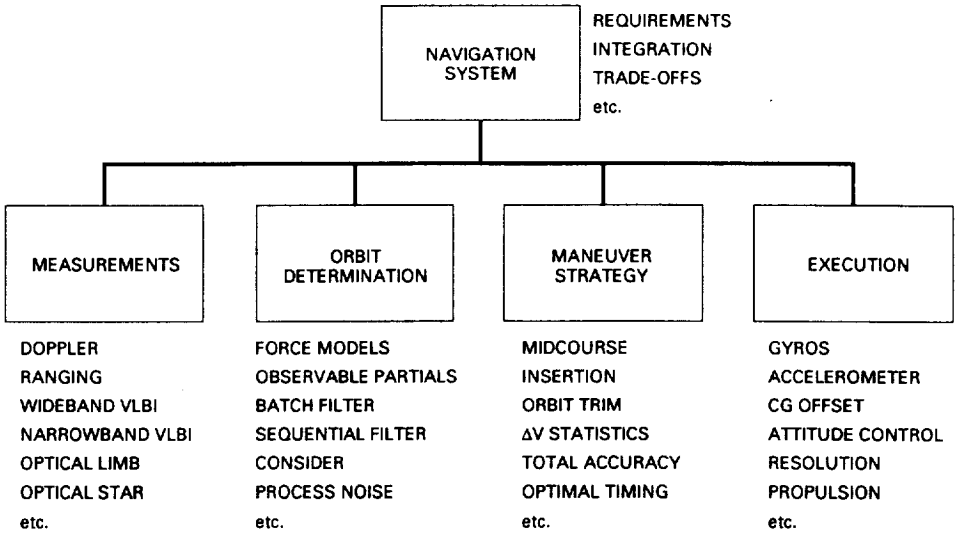


Fig. 1. Composition of generic navigation system.

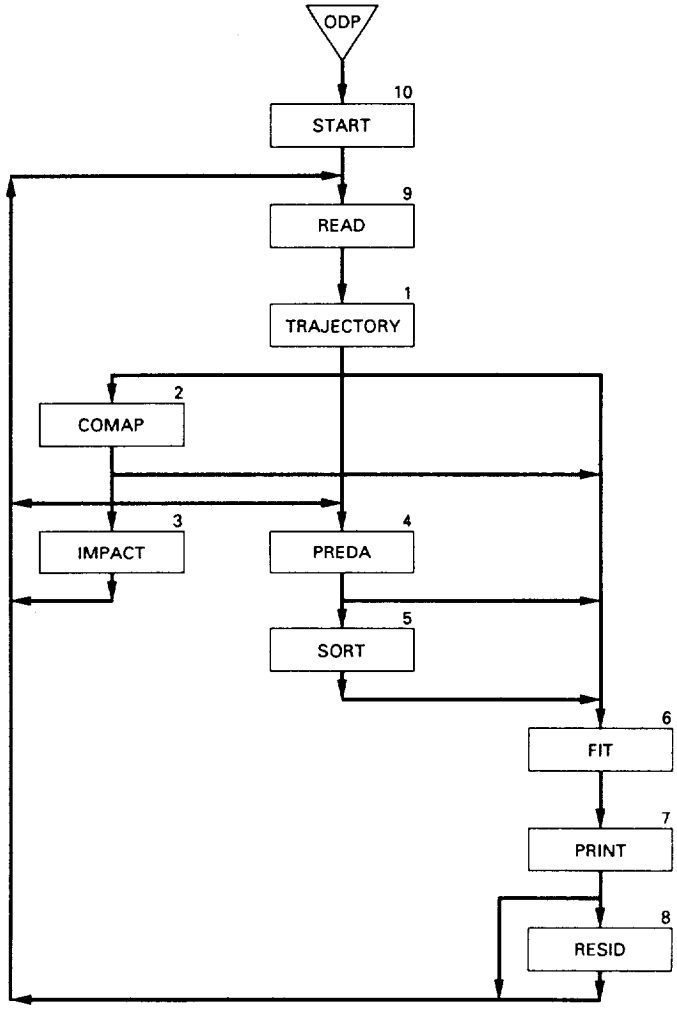


Fig. 2. High-level view of first-generation ODP structure.

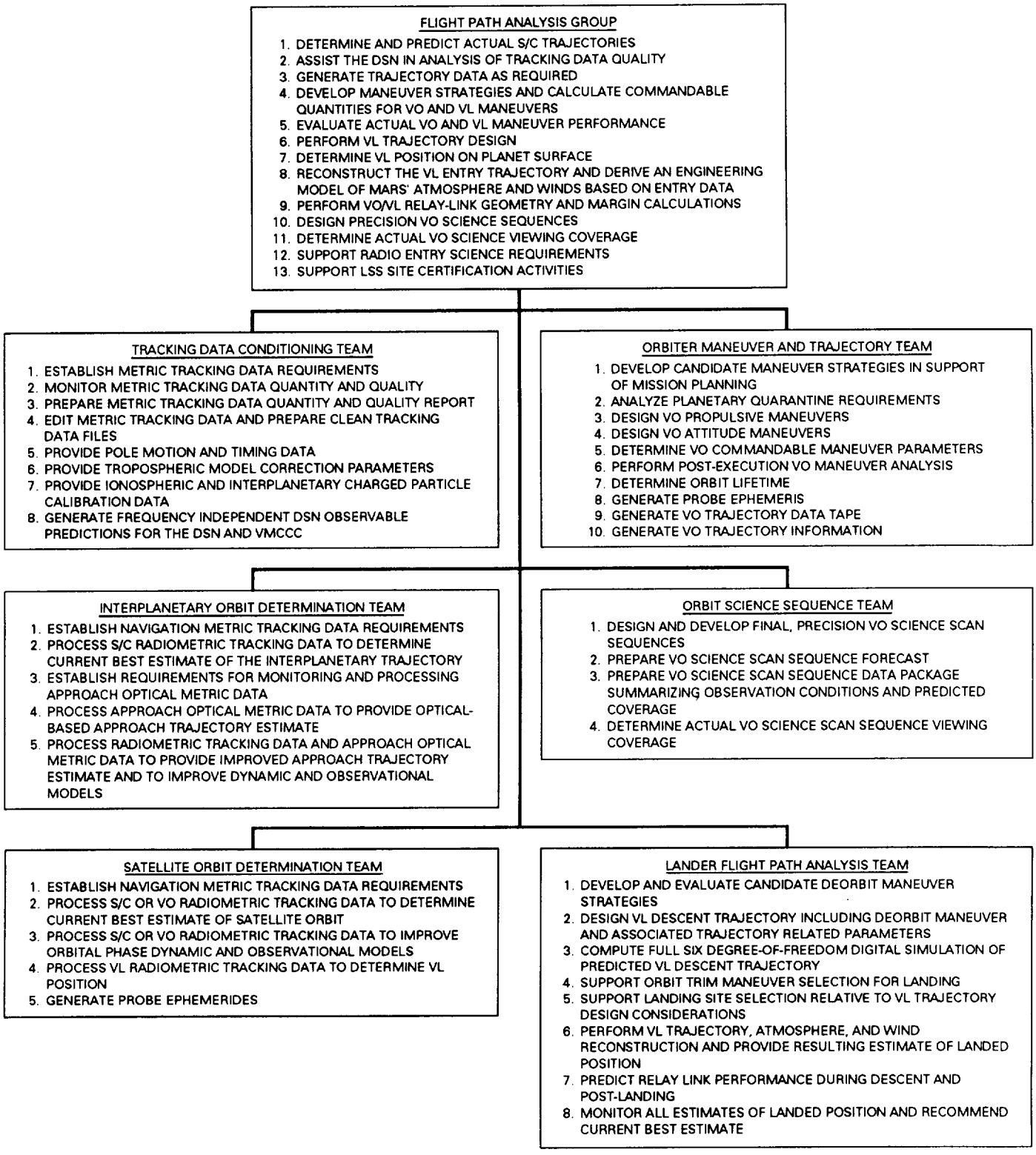


Fig. 3. Navigation system design for the Viking Mission.

1.

FOLDOUT FRAME

2.

FOLDOUT FRAME

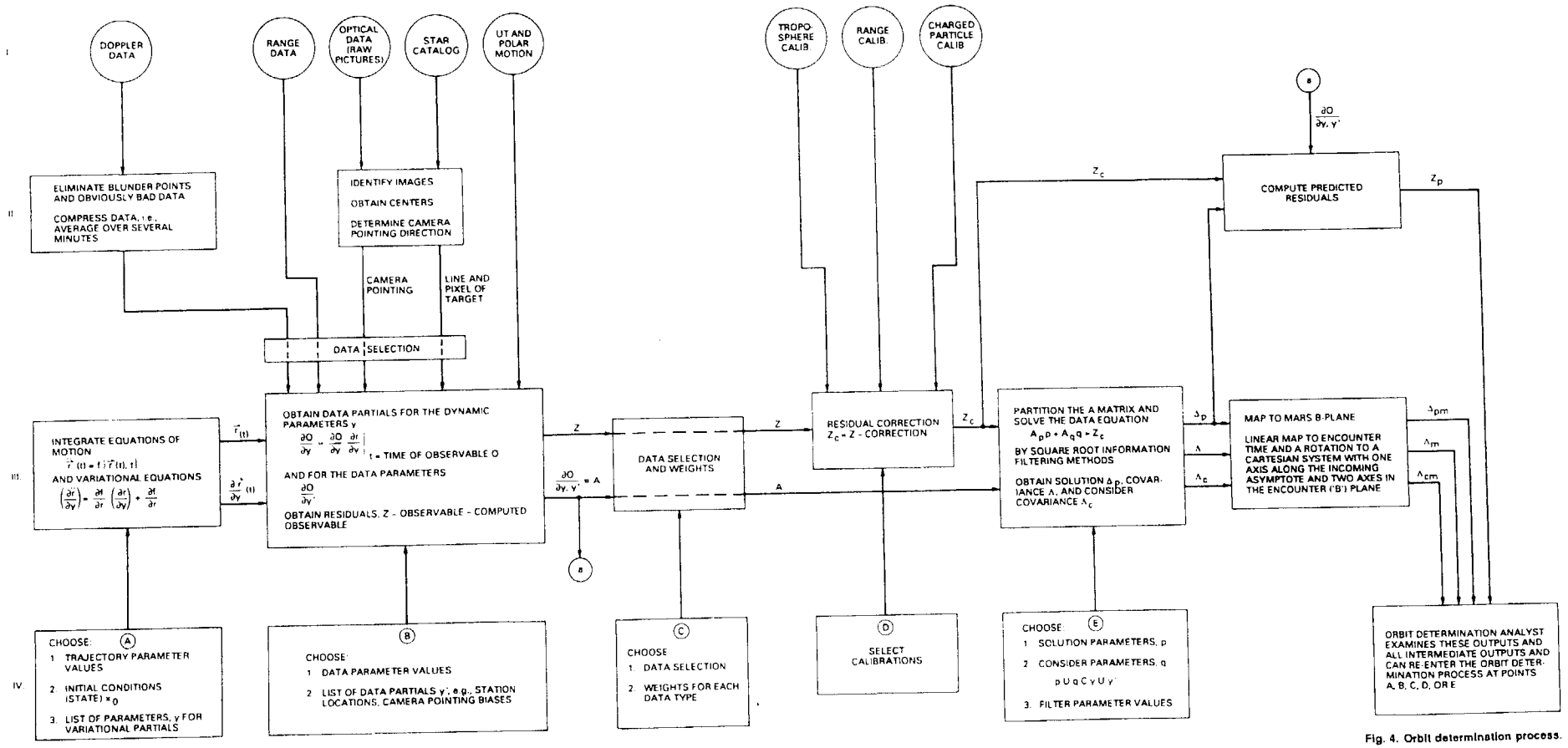


Fig. 4. Orbit determination process.

48

TDA Progress Report 42-101

May 15, 1990

Inhomogeneous Electric Field Effects in a Linear RF Quadrupole Trap

R.K. Melbourne, J.D. Prestage, and L. Maleki
Communications Systems Research Section

This article presents the exact potential corresponding to confining fields inside a linear rf quadrupole particle trap of finite length. The analytic expression for the trapping potential is derived by introducing a linear trap employing a relatively simple cylindrical geometry and solving Laplace's equation for the trap electrodes. The finite length of linear traps results in field distortion near the trap ends. An exact analytic determination of the fields is useful because the profile of the trapped ion cloud is highly dependent on the fields confining it. It is shown that near the ends of the trap, the effective potential arising from the rf fields acts to propel particles out of the trap, and further, that the addition of a dc bias generates an inhomogeneous field in the trap that influences the particles both perpendicularly to and along the trap's long axis.

I. Introduction

Trapped $^{199}Hg^{+}$ ion standards are presently the most stable frequency standard developed for averaging times $\geq 10^4$ seconds. Trapped ions can be stored for long periods, essentially isolated from the outside environment. This arrangement serves to reduce the perturbation to the ions' atomic-energy levels and hence minimize frequency fluctuation.

Currently, the Time and Frequency Systems Research Group is developing a prototype linear ion trap consisting of four parallel cylindrical rod electrodes enclosed between two end electrodes (Fig. 1). This trap is preferable over the conventional spherical trap in frequency-standard applications because of its 20-times-larger ion storage capacity. The number of ions determines an important trade-off parameter between the signal-to-noise ratio and the second-order Doppler shift arising from the motion of the ions caused by the trapping field. The geometry of the linear trap improves clock peformance by allowing an increase of the number of trapped ions without degradation of the frequency stability [1].

Particle motion is directly affected by the electric fields generated within the trap. A time-varying quadrupole electric field is produced in the linear rf trap by an ac voltage applied to the four rods such that any rod's nearest neighbor has the opposite polarity. The rf field provides the trapping force necessary to confine particles in the radial direction (see Fig. 1). In order to prevent trapped particles from escaping out the ends of the trap, a dc voltage is applied to the two "endcaps," producing an exponentially

decaying field directed inward along the the axis of cylindrical symmetry of the trap (i.e, along the z axis). The linear trap is also capable of containing larger macroparticles (such as alumina or other substances) [2]. In this case, however, static electric fields must be incorporated in addition to the rf and endcap fields to offset gravitational forces on the macroparticles, which are comparable to the trapping forces. If the trap is oriented such that the trap axis is horizontal with the rods at the corners of a square when viewed down the trap axis, a dc voltage can be applied to the top two rods of the quadrupole to provide this offset.

RF fields in the radial direction of a linear trap have been analyzed extensively by assuming the trap to be infinitely long [3]. In actuality, however, the traps are of finite length, and thus both static and time-varying fields emanating from the rod electrodes will be directed longitudinally along the axis of the trap, as well as radially. Near the ends of the trap, field distortion occurs. In an ion trap, this can subsequently affect the shape of the trapped ion cloud and thus ion motion. To optimize clock performance, an exact determination of the fields generated within the trap, including the "end effects," is necessary. This then allows for the calculation of the surface shape of the ion plasma contained within the trap. In this article, an analytic model is developed for one implementation of a linear trap to determine static fields produced by the end electrodes and time-varying electric fields everywhere inside a finite-length linear trap with cylindrical geometry. The model consists of a hollow cylinder, partitioned lengthwise into four sections (analogous to the four rods) and two endcaps. The simple geometry of the cylinder allows the calculation of the fields within the cylinder accurately (Fig. 2).

Although consideration of a dc offset to counterbalance gravitational forces is unnecessary when modeling a linear trap designed for the storage of ions, it cannot be neglected when discussing a trap that contains macroparticles. The primary purpose of applying the dc offset is to counterbalance gravitational forces on macroparticles. However, if the bias is added to the rod electrodes, electric fields will be additionally generated along the axis of the trap in a manner that acts to longitudinally localize the particles in the trap center or to propel them out the ends of the trap, depending on the polarity of the dc offset. A second model is presented in this article to determine the effects of the fields generated by a dc voltage applied to two rods in the linear trap. Because the focus of this model is solely to demonstrate the effects of applying a dc offset to the trap, the contribution from the rf trapping fields and the fields generated by the end electrodes are neglected, although the complete determination of the potential inside a cylindrical macroparticle trap is possible by combining the potentials derived from the two models.

First presented is the conventional trapping theory for rf quadrupoles previously developed in the literature. Then, the potential everywhere inside a linear ion trap with cylindrical geometry is derived by modeling the trap after a four-sectioned hollow cylinder of finite length enclosed by two endcaps. In Section IV, the potential is determined inside a second model that consists of a hollow cylinder split along its length into two halves. The top half is at a voltage V_o and the bottom half and endcaps are grounded. This model accounts for the effect of a dc-offset potential applied to the rod electrodes in a linear macroparticle trap. The conclusion summarizes the results of the two models.

II. Trapping Theory

The linear trap typically employs two dc-biased "endcap" electrodes to confine the particles longitudinally, and four rods that produce a time-varying electric field to contain them in the transverse or radial direction. The voltage applied to the rods is such that at a given instant, any pair of adjacent rods always has opposite polarity (Fig. 3). The potential inside the trap near the central axis (i.e., $x, y \ll R$) is given by

$$V(x,y) = \frac{[V_o(x^2 - y^2)\cos(\Omega t)]}{2R^2} \tag{1}$$

where x and y denote the particle's position in the radial direction, R represents the distance between the central axis and the surface of the rods, and Ω is the driving frequency. The resultant time-varying electric fields produce a trapping force that increases linearly in the radial direction. A single charged particle within the trapping region will be subjected to a time-averaged force towards the central axis of the trap ($r = \sqrt{x^2 + y^2} = 0$) where the field intensity is a minimum. A nonzero time-averaged restoring force for a single particle is derivable from a pseudopotential, the effective potential existing in an inhomogeneous electric rf field [4]

$$\psi(x,y) = \frac{q[E_o(x,y)^2]}{4m\Omega^2} \tag{2}$$

In the case of an infinitely long linear trap, Eq. (2) modifies to

$$\psi(x,y) = \frac{qV_o^2}{4m\Omega^2 R^4}(x^2 + y^2) \tag{3}$$

where q is the charge on the particle and x and y are the particle's position in the radial direction averaged over one period $T = (2\pi/\Omega)$. The force generated from this pseudopotential produces particle dynamics that superimpose a slower oscillating motion characterized by frequency ω, upon a faster micromotion due to the driving frequency Ω. Stable confinement of the particles occurs when $\omega \ll \Omega$. The frequency ω for radial particle oscillation can be derived from Eq. (3) as

$$\omega^2 = \frac{q^2 {V_o}^2}{2m\Omega^2 R^4} \tag{4}$$

Although the particle is in dynamic equilibrium along the radial direction, it is still free to traverse the longitudinal (z) axis. If there is no constraining force along this axis and the trap contains more than one particle, the coulomb repulsion between the particles tends to propel them out the ends of the trap. To counterbalance these forces, dc-biased electrodes are incorporated at either end of the trap in both ion and macroparticle systems.

III. Analytical Model for the Potential Inside a Linear Trap

The complete electric field inside the linear ion trap consists of a static field arising from applied voltage on the endcaps and the time-varying fields resulting from a voltage arrangement on the trap rods. Unfortunately, the analytic determination of the net electric field for a finite-length four-rod linear trap is formidable and, as yet, only an approximation of the field exists by considering the trap to be infinitely long. This approximation neglects the effect trap ends have on the generated rf fields that influence the dynamics of trapped ions. Rather than neglecting the end effects, in the model described here the fields generated by the four-rod geometry are approximated by employing a hollow cylinder partitioned into four sections with endcaps at either end (Fig. 2). While providing a rough approximation to the four-rod trap, the model primarily represents a new linear-trap geometry that can be accurately modeled to determine the exact confining field, including any significant end effects. An exact analytic determination of the fields is highly useful since the shape of the trapped ion cloud depends strongly on the fields surrounding it. Because the ions contribute significantly to the electric field within the trap, only certain density profiles are acceptable. With a complete picture of the fields generated by the cylindrical shell and endcaps, Poisson's equation can be solved numerically to investigate the shape of the ion plasma. By virtue of its design, the cylindrical geometry also conveniently provides shielding against charge buildup that might occur in regions external to the trap such as lenses or mirrors. The pseudopotential arising from the rf fields and the potential generated from the endcaps are solved separately. Superimposing these two solutions produces the net potential everywhere inside the cylindrical trap.

The pseudopotential is determined by employing the four-sectored cylindrical model in Fig. 2 with grounded endcaps and a surface voltage V_o applied such that $V_o(\theta) = -V_o(\theta + \pi/2)$. This voltage arrangement represents the applied voltage on the four-rod quadrupole at a given instant. Starting with Laplace's equation in cylindrical coordinates, a general solution of the product form is assumed:

$$\Phi(r, \theta, z) = R(r)Q(\theta)Z(z) \tag{5}$$

It can be shown that for this type of geometry [5]

$$\Phi(r, \theta, z) = [A\cos(m\theta) + B\sin(m\theta)][CI_m(kr) + DK_m(kr)][E\sin(kz) + F\cos(kz)] \tag{6}$$

where I_m and K_m are the modified Bessel functions. The remaining coefficients are constants to be determined. With the requirements that $F, D = 0$ and

$$k_n = \frac{n\pi}{L} \tag{7}$$

the equation for Φ takes on the modified general form

$$\Phi(r, \theta, z) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \left[(A_{mn}\cos(m\theta) + B_{mn}\sin(m\theta)) \left(I_m\left(\frac{n\pi r}{L}\right)\right) \left(\sin\left(\frac{n\pi z}{L}\right)\right)\right] \tag{8}$$

The asymmetric voltage is applied to the four cylinder sections on the cylindrical surface ($r = R$) such that:

$$V(\theta, z) = \left\{ \begin{array}{ll} V_o, & -\pi < \theta < \frac{-\pi}{2} \\ -V_o, & \frac{-\pi}{2} < \theta < 0 \\ V_o, & 0 < \theta < \frac{\pi}{2} \\ -V_o, & \frac{\pi}{2} < \theta < \pi \end{array} \right\} \tag{9}$$

The potential on the surface is thus given as

$$V(\theta,z)=\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\Bigg[(A_{mn}\cos(m\theta)+B_{mn}\sin(m\theta))\left(I_m\left(\frac{n\pi R}{L}\right)\right)\times\left(\sin\left(\frac{n\pi z}{L}\right)\right)\Bigg] \quad (10)$$

with the coefficients

$$A_{mn}=0$$

$$B_{mn}=\frac{32V_o}{mn\pi^2}\left(I_m\left(\frac{n\pi R}{L}\right)\right)^{-1} \quad (11)$$

Substituting these constants back into Eq. (8), and multiplying Φ by $\cos(\Omega t)$ to take into account that it varies with a frequency Ω, the time-varying potential due to the surface voltage is determined:

$$\Phi=\frac{32V_o}{\pi^2}\sum_{m}^{\infty}\sum_{n}^{\infty}\left[\left(\frac{1}{mn}\frac{I_m\left(\frac{n\pi r}{L}\right)}{I_m\left(\frac{n\pi r}{L}\right)}\right)\times\sin(m\theta)\sin\left(\frac{n\pi z}{L}\right)\right]\cos(\Omega t) \quad (12)$$

where $m=2,6,10,\ldots$ and $n=1,3,5,\ldots$. Providing the secular frequency is much less than the frequency of micromotion ($\omega \ll \Omega$), the macromotion of the particle can be described by the time-independent pseudopotential discussed in Section II

$$\psi=\frac{q}{4m\Omega^2}(\nabla\Phi)^2 \quad (13)$$

and incorporating Eq. (12),

$$\psi=\frac{q}{4m\Omega^2}\left(A^2(r,\theta,z)+B^2(r,\theta,z)+C^2(r,\theta,z)\right) \quad (14)$$

where

$$A(r,\theta,z)=\frac{32V_o}{\pi L}\sum_{m}^{\infty}\sum_{n}^{\infty}\left[\frac{1}{m}\frac{I_m\left(\frac{n\pi r}{L}\right)}{I_m\left(\frac{n\pi R}{L}\right)}\times\sin(m\theta)\cos\left(\frac{n\pi z}{L}\right)\right]$$

$$B(r,\theta,z)=\frac{32V_o}{r\pi^2}\sum_{m}^{\infty}\sum_{n}^{\infty}\left[\frac{1}{n}\frac{I_m\left(\frac{n\pi r}{L}\right)}{I_m\left(\frac{n\pi R}{L}\right)}\times\cos(m\theta)\sin\left(\frac{n\pi z}{L}\right)\right]$$

$$C(r,\theta,z)=\frac{32V_o}{2L\pi}\sum_{m}^{\infty}\sum_{n}^{\infty}\left[\frac{1}{m}\times\frac{I_{m-1}\left(\frac{n\pi r}{L}\right)+I_{m+1}\left(\frac{n\pi r}{L}\right)}{I_m\left(\frac{n\pi R}{L}\right)}\times\sin(m\theta)\sin\left(\frac{n\pi z}{L}\right)\right] \quad (15)$$

The inhomogeneous electric field arising from this effective potential produces a net force on a single particle that, averaged over one period of the micromotion ($T=\Omega/2\pi$), is directed towards $r=0$ in the radial direction and out the ends of the trap in the z direction. The z dependence of the pseudopotential is demonstrated by Figs. 4(a) and 4(b), which represent examples of $\psi(z)$ at two different positions of θ. In both of the examples, $r=R/8$ with $V_{ac}=200$ V, $\Omega/2\pi=60$ Hz, $R=0.008$ m, and $L=0.06$ m. A 5 μm-diameter alumina particle with an approximate positive charge-to-mass ratio of 0.0053 C/kg is used to evaluate Eq. (14).

The magnitude of $\psi(\theta)$ for fixed values of r and z is sinusoidal and oscillates from a maximum in regions near the boundaries of each of the four sections of the cylinder ($\theta=0,\pi/2,\pi,2\pi$) to a minimum along the centers of the cylinder sections ($\theta=\pi/4,3\pi/4,5\pi/4,7\pi/4$) (Fig. 5). It is interesting to note that the pseudopotential in the cylindrical model indicates an electric field in θ that is actually 45 deg out of phase with the electric field generated in the linear four-rod trap. In the latter case, the pseudopotential is a maximum in the vicinity of the rods (located at $\theta=\pi/4,3\pi/4,5\pi/4,7\pi/4$) and a minimum between them. This discrepancy arises because the spatial separation of the rods results in weaker fields between any two rods in contrast to the fields between the cylinder sections whose boundaries meet and give rise to large electric fields. The fact that the pseudopotential is a minimum at the section centers in the cylindrical linear trap suggests that small viewing holes may be added in this vicinity without significant disturbance of the fields within.

In order to obtain the total potential inside the cylinder, the potential arising from the endcaps must be determined and superimposed on the pseudopotential. This is accomplished by solving Laplace's equation for a cylinder

with a grounded surface and endcaps held at a constant voltage V_e. The resulting potential inside the cylinder for this configuration is

$$\Phi_e(r,z) = 2V_e \sum_{m=1}^{\infty} \left[\frac{J_0\left(\frac{P_m r}{R}\right)}{P_m J_1(P_m)} \left(\frac{\cosh\left(\frac{P_m(z-L/2)}{R}\right)}{\cosh\left(\frac{P_m L}{2R}\right)} \right) \right] \tag{16}$$

where J_0 and J_1 are the *zero*th and first Bessel functions, respectively, and P_m is the mth root of J_0. Adding Eqs. (14) and (16) determines the net potential inside a finite cylinder of length L and radius R

$$\Psi_{net}(r,\theta,z) = \psi + \Phi_e \tag{17}$$

The net potential can be interpreted geometrically by observing that it is composed of the pseudopotential whose curve, as demonstrated in Figs. 4(a) and 4(b), rises for fixed values of r and θ from a minimum at $z = 0$ and $z = L$ to a maximum in the center of the trap, and an "endcap" potential that varies from a constant V_1 on the ends to a minimum in the center of the trap (Fig. 6). By controlling the amount of voltage applied to the endcaps, Ψ_{net} can be adjusted. The model suggests that there is an optimum endcap voltage that maximizes the continuity of the net potential in the z direction. Figure 7 demonstrates the z dependence of Ψ_{net} when Φ is greater than ψ.

IV. Analytical Model for a DC Field Perpendicular to the Cylindrical Axis

In a horizontal configuration, the introduction of a dc voltage on the top two rods of the linear quadrupole trap generally serves to counter the forces of gravity on macroparticles. Because the rods are of finite length, however, the electric field is not directed exclusively in the radial direction as in the case of infinite rod length. Instead, a distortion of the field lines occurs at the ends of the rods, giving rise to a field component in the longitudinal direction. The effects on particle dynamics of an asymmetric dc bias may be understood by again considering the four-rod quadrupole to be a hollow cylindrical trap split into two sections; the z axis is coincident with the axis of the linear quadrupole trap. Figure 8 depicts the cylindrical model with radius R and length L. To model the effects of the offset in the trap, a dc voltage is applied on the top half of the cylinder and the bottom half and endcaps are grounded. This potential can be added to Ψ_{net} developed in the previous section to provide a full solution of the potential inside a linear trap when a dc bias is added. The potential due to the dc offset is solved in a similar manner as the problem in Section II, and the results are

$$\Phi(r,\theta,z) = \frac{2V_o}{\pi} \sum_{n\ odd}^{\infty} \left[\left(\frac{\sin\left(\frac{n\pi z}{L}\right)}{n} \right) \left(\frac{I_0\left(\frac{n\pi r}{L}\right)}{I_0\left(\frac{n\pi R}{L}\right)} + \sum_{m\ odd}^{\infty} \frac{4}{m\pi} \frac{I_m\left(\frac{n\pi r}{L}\right)}{I_m\left(\frac{n\pi R}{L}\right)} \sin(m\theta) \right) \right] \tag{18}$$

Figure 9 represents a computer evaluation for this function of the first 22 values of m and n, respectively. Near the central axis of the trap ($r = 0$), the function varies smoothly, falling off to zero at $z = 0$ and $z = L$. The resultant electric field at $r = 0$ is

$$-\frac{\partial\Phi(r,\theta,z)}{\partial z} = E_z = \frac{-2V_o}{L} \sum_{n\ odd}^{\infty} \cos\left(\frac{n\pi z}{L}\right) \left[I_o\left(\frac{n\pi R}{L}\right) \right]^{-1} \tag{19}$$

The surface voltage V_o is inherently positive in Eq. (19). By referring to Figs. 10(a) and (b), it is evident that if a single negatively charged particle is subjected to this field, the force on that particle will be directed inward towards the center of the trap ($z = L/2$), growing abruptly from zero to a maximum as it approaches either end. The potential inside the trap could have been determined similarly by solving Laplace's equation for a voltage V_o applied to the bottom half of the cylinder instead of the top. In this case, if the particle had been taken to be positive, the net effect would still support the particle against gravity, but would accelerate the particle out one of the ends of the cylinder.

The analogy to the linear cylindrical trap is apparent. In order to counterbalance the gravitational effects on the particle, the dc voltage applied to the top rods must be of the opposite sign to that of the particle. But as demonstrated above, this condition inherently gives rise to an inhomogeneous force on the particle that is directed longitudinally inward, towards the center of the trap. Likewise, a gravitational offset can also be accomplished by applying a voltage to the bottom two trap rods that is the same polarity as the particle. This, however, additionally produces an electric field that acts to force the particle out the ends of the trap.

Farther out along the radial direction, the potential increases in magnitude and the function assumes a more rectangular shape. In the limit at $r = R$, V_o spans the

rectangular shape. In the limit at $r = R$, V_o spans the length of the cylinder and the function discontinuously jumps to 0 at either end. The electric field evaluated on the surface is

$$-\frac{\partial\Phi(\theta,z)}{\partial z} = E_z = -\frac{2V_o}{L}\sum_{n\ odd}^{\infty}\left[\cos\left(\frac{n\pi z}{L}\right)\right.$$
$$\left.\times\left(1+\frac{4}{\pi}\sum_{m\ odd}^{\infty}\frac{1}{m}\sin\left(\frac{m\pi}{2}\right)\right)\right] \quad (20)$$

where θ is again taken to be a constant at $\pi/2$. The model indicates that the electric field in the z direction towards the center of the trap depends on the radial position of the particle. This dependence is demonstrated by comparing Figs. 10(a) and (b).

The cylindrical model approximates the field accurately and yields insight into the influences on the particle in the longitudinal direction from the gravitational offset. For the case of many trapped particles, the model suggests that in some instances a bias on the end electrodes is not really necessary to counterbalance coulomb repulsion between particles, if a dc offset on the top rods is added. In this case, the particles will distribute themselves in such a manner as to create an equilibrium situation, balancing the coulomb forces and the force arising from the applied voltage.

V. Conclusion

This article has presented a model for the rf fields in a linear trap by considering a four-sector cylindrical geometry. This geometry approximates a finite linear trap with rod electrodes. It also represents a new configuration for an ion trap that is amenable to analytic determination of the confining fields. Using the model for the finite-length linear cylindrical trap, it was shown that confining fields have a longitudinal component as well as a radial component. The z component of the rf electric field produces an extra term in the pseudopotential not present in the ideal, infinitely long trap. The details of the end effects are important in ion confinement. The analytic model indicates that the pseudopotential arising from the rf fields gives rise to a time-averaged force towards the center of the trap in the radial direction and away from the center in the z direction. This latter effect, combined with the ions' own coulomb interaction, will accelerate the particles out of the trap if biased endcaps are not present. The density of the ion cloud can be determined numerically by solving Poisson's equation for the fields generated by the trap electrodes encompassing the ion plasma. Future research is planned for the evaluation of the plasma shape in a linear cylindrical trap.

When a dc bias potential is applied to the trap electrodes (for example, in the case of a macroparticle trap where the gravitational force is counterbalanced with a dc offset), an electric field is generated in both the radial and longitudinal directions. An analytic model has been developed that utilizes a cylindrical geometry to determine these fields. According to the model, a particle of the opposite sign to an applied dc voltage on the upper electrodes will feel an inwardly directed force along the z axis in addition to the transverse force. On the other hand, a particle with the same polarity as an applied dc voltage on the lower electrodes will be forced upwards in the transverse direction and towards the trap ends in the longitudinal direction.

Acknowledgment

The authors wish to acknowledge the use of the JPL Cray supercomputer to perform these simulations.

References

[1] J. D. Prestage, G. J. Dick, and L. Maleki, "New Ion Trap for Frequency Standard Applications," *J. Appl. Phys.*, vol. 66, no. 3, pp. 1013–1017, 1989.

[2] R. F. Wuerker, H. Shelton, and R. V. Langmuir, "Electrodynamic Containment of Charged Particles," *J. Appl. Phys.*, vol. 30, no. 3, pp. 342–349, 1959.

[3] G. R. Janik, J. D. Prestage, and L. Maleki, "Simple Analytic Potentials for Linear Ion Traps," *TDA Progress Report 42-99*, vol. July–September 1989, Jet Propulsion Laboratory, Pasadena, California, pp. 12–19, November 15, 1989.

[4] H. G. Dehmelt, "Radiofrequency Spectroscopy of Stored Ions, I: Storage," *Adv. At. Mol. Phys.*, vol. 3, pp. 53–154, 1967.

[5] J. D. Jackson, *Classical Electrodynamics*, 2nd edition, New York: John Wiley and Sons, Inc., 1975.

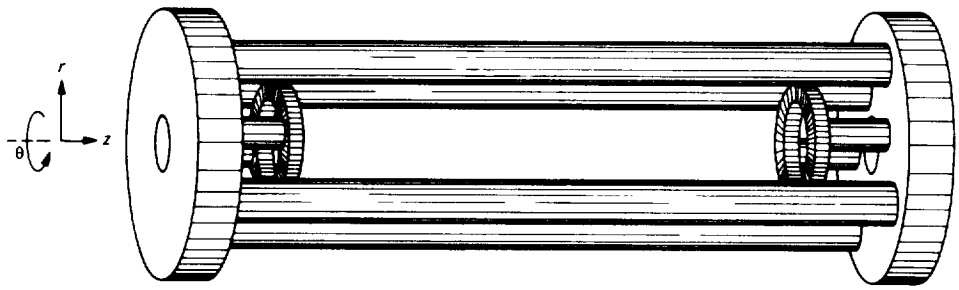


Fig. 1. Linear four-rod quadrupole trap.

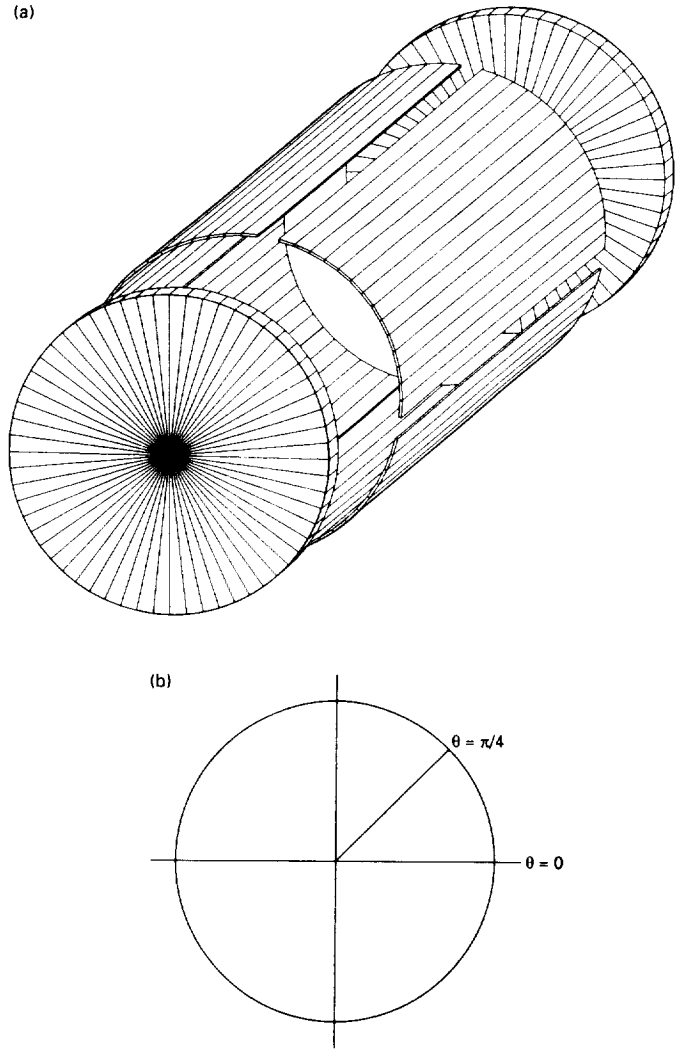


Fig. 2. Linear four-sectored quadrupole trap: (a) three-dimensional view of the trap with endcaps; and (b) orientation of the cylinder sections as viewed down the z axis. Section boundaries meet at $\theta = 0$, $\pi/2$, π, and $3\pi/2$.

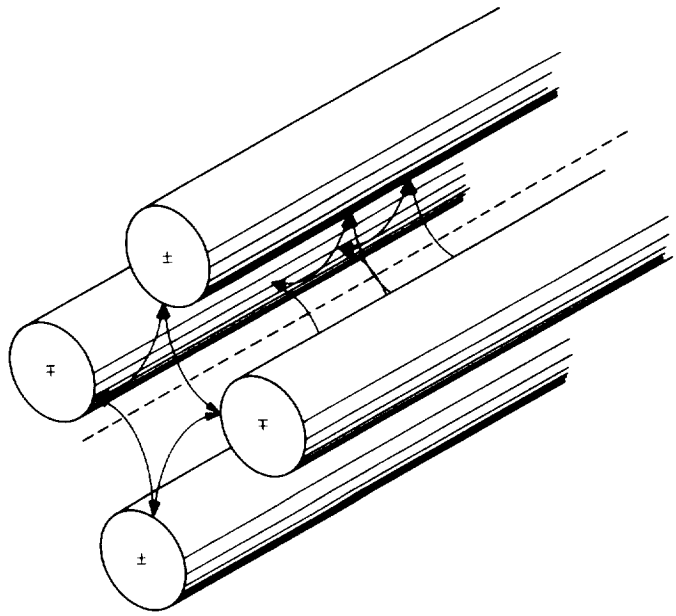


Fig. 3. Four-rod quadrupole. The voltage on each rod's nearest neighbor has opposite polarity.

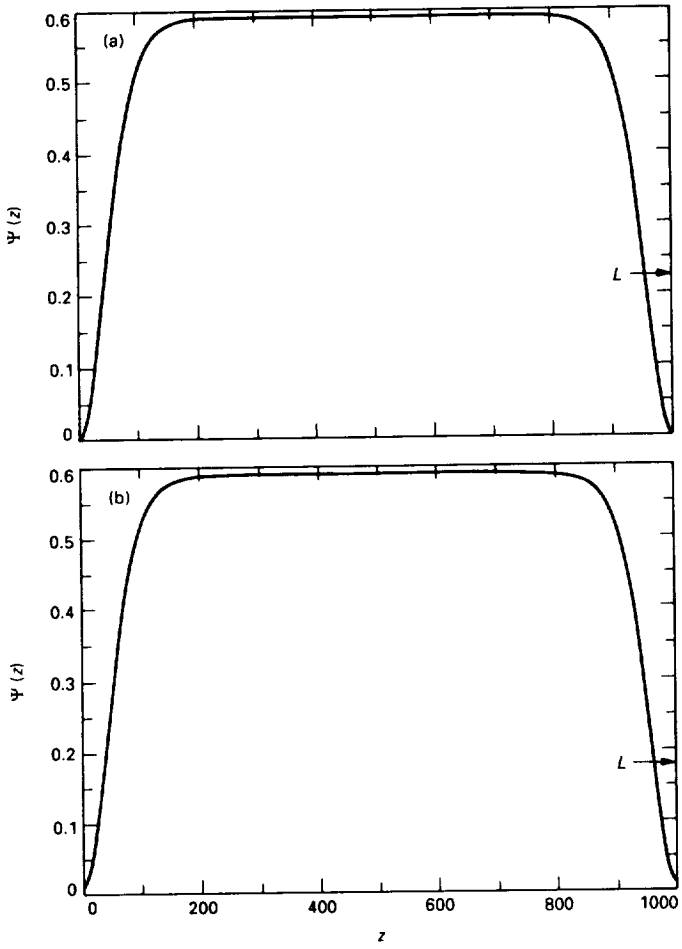


Fig. 4. The pseudopotential $\psi(z)$ at two positions of θ: (a) z dependence of $\psi(r,\theta,z)$ evaluated at $r = R/8$ and $\theta = 0$ for a single alumina particle with $e/m = 0.0053$ C/kg and trapping parameters $V_{ac} = 200$ V and $\Omega/2\pi = 60$ Hz; and (b) $\psi(r,\theta,z)$ evaluated at $r = R/8$ and $\theta = \pi/4$.

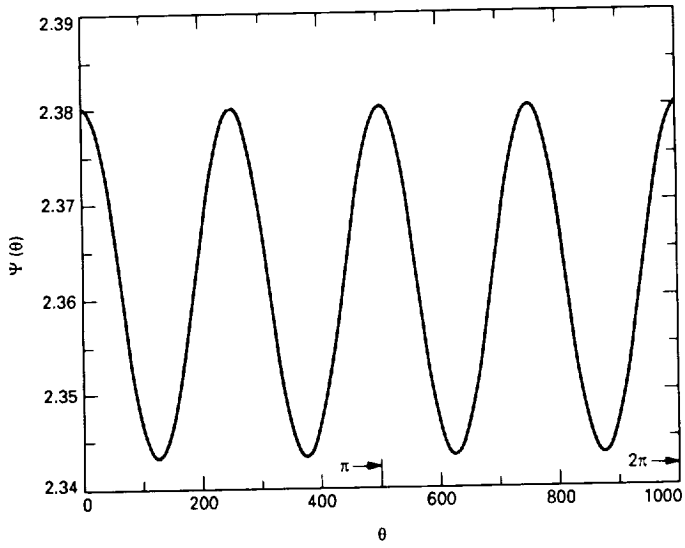


Fig. 5. Angular dependence of $\psi(r,\theta,z)$ evaluated at $r = R/4$ and $z = L/2$ for a single alumina particle with $e/m = 0.0053$ C/kg and trapping parameters $V_{ac} = 200$ V and $\Omega/2\pi = 60$ Hz.

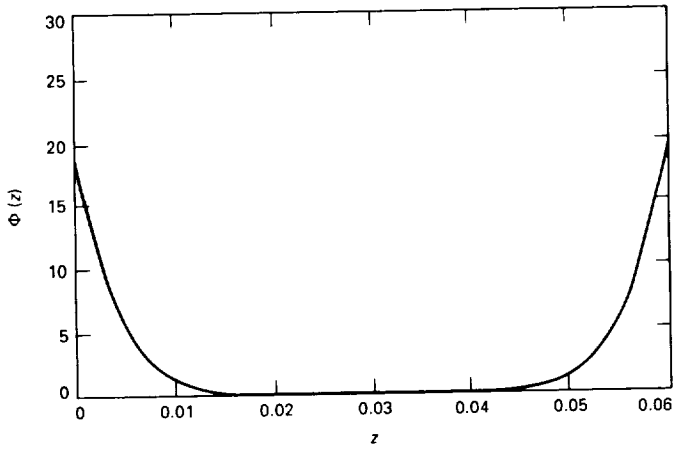


Fig. 6. The longitudinal (z) dependence of $\Phi_e(r,z)$ evaluated at $r = 0$ with end electrode voltage $V_1 = 20$ V.

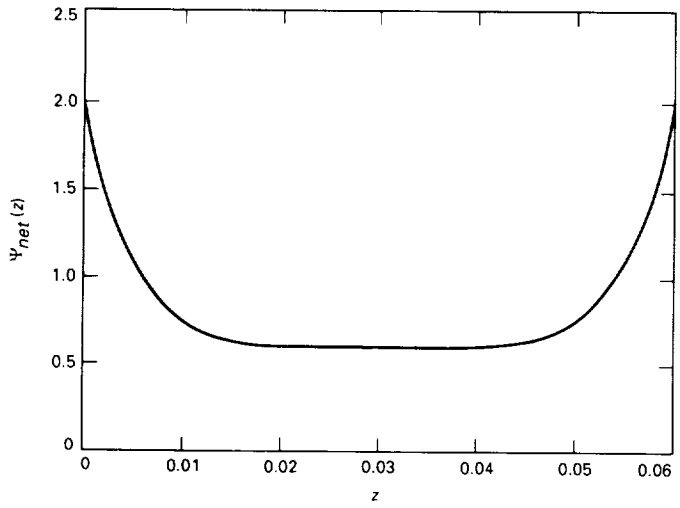


Fig. 7. The longitudinal *(z)* dependence of $\Psi_{net}(r,\theta,z)$ evaluated at $r = R/8$ and $\theta = \pi/4$ for a single alumina particle with $e/m = 0.0053$ C/kg and trapping parameters $V_{ac} = 200$ V, $V_1 = 2$ V, and $\Omega/2\pi = 60$ Hz.

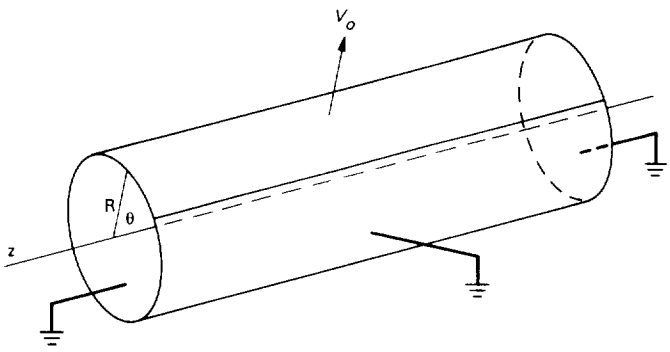


Fig. 8. Cylindrical model for a linear particle trap with a dc bias added to the top two rods of the trap.

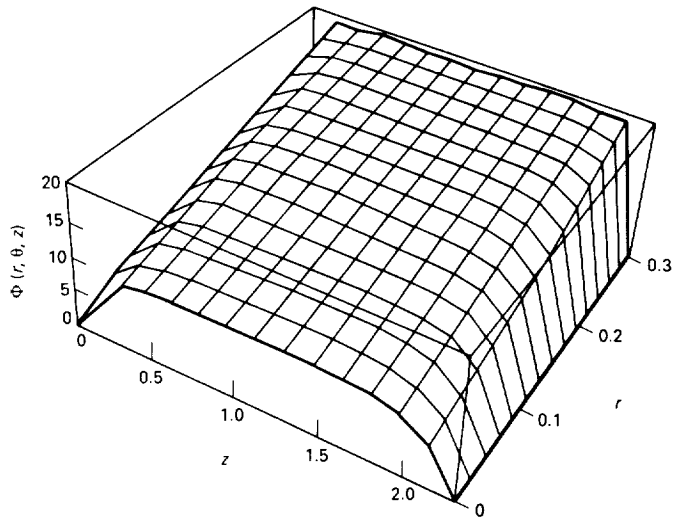


Fig. 9. The potential $\Phi(r,\theta,z)$ due to a dc offset inside the cylindrical model evaluated at $\theta = \pi/2$ with an applied voltage $V_o = 20$ V.

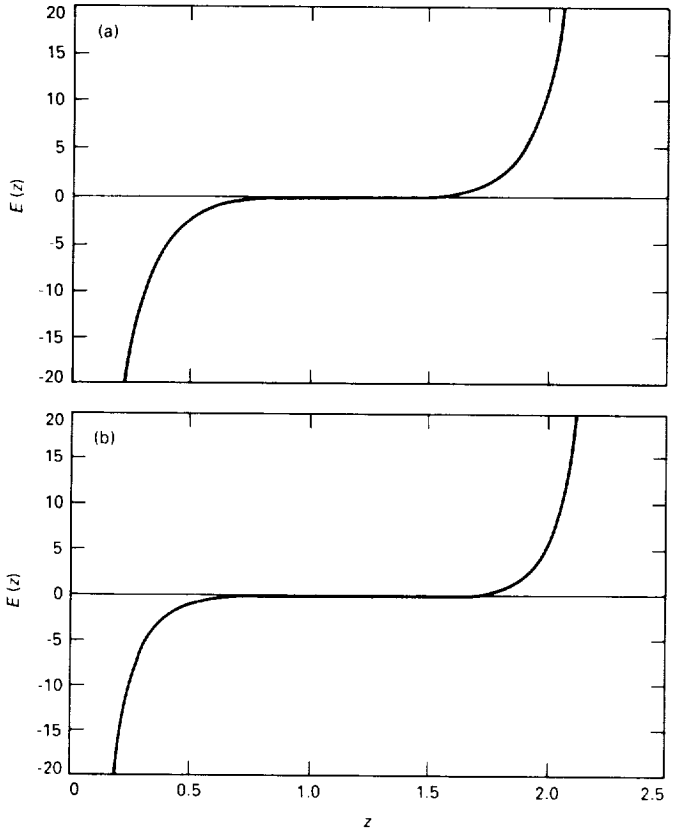


Fig. 10. The electric field $E(z)$ within the cylindrical model: (a) at a fixed $r = 0$, $\theta = \pi/2$, and $V_o = 20$ V; and (b) at a fixed $r = 3R/4$, $\theta = \pi/2$, and $V_o = 20$ V.

N90-26223

TDA Progress Report 42-101

May 15, 1990

A Study of the Influence of Hg(6 $^3P_2$) Population in a Low-Pressure Discharge on Mercury Ion Emission at 194.2 nm

L. Maleki, B. J. Blasenheim, and G. R. Janik
Communications Systems Research Section

A low-pressure mercury-argon discharge, similar to the type existing in the mercury lamp for the trapped-ion standard, is probed with a new technique of laser spectroscopy to determine the influence of the IIg(6 $^3P_2$) population on discharge emission. The discharge is excited with inductively coupled rf power. Variations in the intensity of emission lines in the discharge were examined as $\lambda = 546.1$-nm light from a continuous-wave (CW) laser excited the IIg(6 $^3P_2$) to (7 $^3S_1$) transition. The spectrum of the discharge viewed in the region of laser irradiation showed increased emission in $\lambda = 546.1$, 435.8, 404.7, 253.7, and 194.2 nm lines. Other lines in IIg I exhibited a decrease in emission. When the discharge was viewed outside the region of laser irradiation, all lines exhibited an increased emission. Based on these results, it is concluded that the dominant mechanism for the excitation of higher lying levels of mercury is the electron-impact excitation via the $^3P_2$ level. The depopulation of this metastable level is also responsible for the observed increase in the electron temperature when the laser irradiates the discharge. It is also concluded that the $^3P_2$ metastable level of mercury does not play a significant role in the excitation of the $^2P_{1/2}$ level of mercury ion.

I. Introduction

Mercury-argon discharges produce the light emission at 194.2 nm required for optical pumping in the trapped-mercury-ion frequency standard currently under development at the Jet Propulsion Laboratory (JPL). These discharges are one of the most widely studied examples of low-temperature plasmas. During the past 70 years or so, numerous investigators have studied the emission and absorption properties of mercury-rare gas discharges. The significance of these discharges to the lighting industry in particular has been responsible for extensive investigations, spanning practically all aspects of radiative, collisional, and plasma processes of this type of discharge.

The advent of laser spectroscopy has produced yet another tool for the investigation of mercury-rare gas discharges. In the authors' laboratory, for instance, lasers

have been used to study saturated absorption, optogalvanic, and phase-conjugate spectra in the mercury-argon discharge [1]. Other investigators have also examined laser-induced fluorescence (LIF) and laser-induced stimulated emission (LISE) in this type of discharge [2].

Despite the wealth of data on various parameters of mercury-argon discharges, a number of questions regarding this system have remained poorly answered. In particular, the significance of a metastable population in the discharge, while qualitatively understood in a general way, has been difficult to determine specifically. Various measurements have been made to determine the concentration and the influence of the population of the ^3P levels of mercury in the line emission and ionization of the discharge [3–6]. Yet definitive conclusions regarding the specifics of the effect of the metastable state population on the rate of ionization and excitation of atoms, molecules, excimers, and ions that may be present in mercury-rare gas discharges have been difficult to model because of a lack of data or understanding of all intervening processes.

Laser spectroscopic studies have provided further understanding of this problem, but have generated other questions. This is because modification of one of the discharge parameters by the laser can influence other parameters, sometimes in a nonlinear manner. Thus the details of the interaction of the laser light with discharges is not always well understood.

One of the most significant areas of data deficiency has been information delineating the role of metastables in the excitation process of the mercury ion. In fact, most models devised to provide a theoretical account of mercury-argon discharges [7–9] do not include excitation rates of the ion in determining plasma parameters. Yet the ion emission at 194.2 nm is the process of greatest interest in connection with the development of an efficient and reliable optical pumping light source for the trapped-mercury-ion frequency standard [10].

Thus, another study has been undertaken to examine the influence of a metastable population on the excitation processes in the mercury-rare gas discharges. In this study, laser spectroscopy of the discharge is utilized in a new way. Whereas LIF and LISE rely on the observation of light emitted from an upper energy level which is optically connected by the laser from a lower level, this laser-induced emission (LIE) spectroscopy examines the response of all lines in the spectrum of the discharge to the population changes in a metastable state that are created by the laser optical field. The major feature of this technique is that all excitation processes affected by the metastable population can be studied, including possible effects on the excitation of mercury atoms, ions, and the buffer gas.

Since results of investigations with discharges depend on the specific parameters, including the concentration of mercury and the rare gas, as well as the discharge excitation mode, they are often limited in their applicability. For this study, a low-pressure rf-excited discharge and the technique of laser spectroscopy were chosen to make results applicable to a general class of discharges that are deemed to be well behaved and relatively well understood. The rf discharge in the glow regime, with argon at less than one torr of pressure and mercury at about 30 degrees centigrade, is readily amenable to modeling. The inductive coupling of the rf power further simplifies the system through elimination of electrodes. The choice of low-power continuous-wave (CW) laser radiation permits additional simplification of the system under study through elimination of higher order light-induced processes, and provides a well-defined interaction channel for the laser and a discharge constituent. It is hoped that the study will assist in understanding of the significance of the metastable populations in the modeling of mercury-argon glows.

In the following section, the details of the experiment are outlined; the results are presented in Section III. The interpretation of the experimental results and the discussion are in Section IV. Finally, Section V provides a summary, conclusion, and an outline of plans for further study.

II. The Experiment

The experimental apparatus in this study is schematically depicted in Figs. 1 and 2. Light generated at $\lambda = 546.1$ nm by a CW dye laser was introduced into a quartz discharge cell. The cylindrical cell had a diameter of 2.5 cm and a length of 10 cm. The cell contained a few milligrams of mercury, and argon at 800 mtorr of pressure. A similar cell containing neon at 500 mtorr instead of argon was also used. The discharge cell was inductively coupled to an rf oscillator, which also provided for the monitoring of the optogalvanic signal. The laser beam passed through a rectangular aperture to produce a ribbon of light which was then directed to the discharge cell along the axis.

Discharge emission was monitored by a 0.25-m monochromator with rectangular slits parallel to both the axis of the cell and laser light ribbon. The monochromator was blazed for 550 nm, and thus could not be effectively used for wavelengths below 190 nm or above 850 nm. The monochromator was mounted on a micrometer-driven translation stage for measuring discharge emission at various positions relative to the center of the laser light rib-

bon. A photomultiplier tube at the output port of the monochromator produced the detected signal that was fed to the input of a lock-in amplifier. The lock-in reference was provided by a chopper which chopped the laser beam on the way to the discharge cell. The output of the lock-in was fed to a chart recorder which recorded the spectral signals.

III. Results

Results of the experiment are summarized in Table 1 and Figs. 3 and 4. The figures show the spectrum of the Hg-Ar discharge outside and inside the laser beam, respectively. Intensities given in the table were obtained directly from the spectra, and were not corrected for the spectral efficiency of the photomultiplier detector or the monochromator. Under the conditions of the experiment described in the previous section, a density of $1.7 \times 10^{17}/m^3$ for the $^3P_2$ population in the discharge was measured using the absorption technique [14].

Because a discharge of the type under study does not satisfy the condition of local thermal equilibrium (LTE), calculation of the electron temperature using emission data of the discharge is not expected to give accurate results [14]. Nevertheless, an attempt was made to obtain an approximate electron temperature by using the emission data. The approximation was $T_e = 9600$ K in the discharge without the laser optical field.

When the monochromator monitored light emission of the discharge in a field of view within the region illuminated by the laser, the following lines exhibited an increase in their emission: 546.1, 435.8, 404.7, 253.7, and 194.2 nm. All other transitions from Hg I exhibited a reduction in light emission. When the monochromator viewed a region outside the ribbon of laser light, all lines in the emission exhibited an increase in intensity.

When the monochromator moved away from the axis of the cell where the laser illuminates the discharge, and towards the edge of the plasma, the intensity of the light emission due to the interaction of the laser with the plasma changed in a continuous manner. In fact, the change, which starts with a decrease and proceeds to a gradual increase in the light emission, is a very sensitive function of the position of the monochromator. Figure 5 exhibits such data for the 313.2- and 312.6-nm lines. Other lines in the Hg I spectrum (with the exception of those that increased in intensity and are listed above) behaved similarly. For different lines, however, the position of the point where the transition was made from an increase to a decrease in the emission varied along the translation axis.

IV. Discussion

This section examines the ramifications of the experimental results for the role of the metastable levels of mercury in a discharge of the type described above. The discussion is specifically aimed at the mercury-argon discharge. But since the results with the mercury-neon discharge are qualitatively the same as the Hg-Ar discharge, all discussions apply to the case of Hg-Ne discharge as well. Three specific points are addressed:

(1) the process responsible for changes in the electron temperature in the discharge,

(2) the role of the $^3P_2$ level in the excitation of the higher lying levels of Hg I, and

(3) the role of the ^3P metastables in the ionization process in the discharge.

When the discharge is irradiated with $\lambda = 546.1$-nm light, the $^3P_2$ to $^3S_1$ transition is excited (see Fig. 6). Since the lifetime of the upper level is short (8 ns), excited $^3S_1$ levels quickly decay to all levels of the 6 ^3P manifold. In this way the population of the $^3P_2$ level is cycled to the $^3P_1$ and $^3P_0$ levels. Some of the population is cycled to the ground state through the $^3P_1$ level. Thus, the emission from the $^3S_1$ level at 404.7, 435.8, and 546.1 nm is expected to increase, and is observed to do so. The emission of the 253.7-nm line from the $^3P_1$ level is also expected, and is observed, to behave similarly.

It is known that one effect of the laser on the discharge is to increase the electron temperature T_e. This is readily verifiable from the observed sign of the optogalvanic signal in previous work done by the authors and in other studies [1,2]. The effect of the laser on the electron temperature, determined in the same manner as described above to obtain T_e, was to increase T_e by less than 5 percent.

While there is no disagreement on the sign of the change in the electron temperature with laser irradiation, there is no universal agreement as to its origin. The general view is that T_e increases in the discharge since ionization rate R_i is directly dependent on the population of the $^3P_2$ level [8]. Since the laser irradiation depopulates this level, its effect on the discharge is to reduce R_i. The electric field subsequently increases to raise T_e, thus increasing ion production and maintaining plasma neutrality. The increase in the electron temperature then compensates for the loss of ion production created by the laser, and the ion density remains a constant. This explanation is generally given for dc discharges where the current through the discharge is held fixed. Its applicability to an rf glow discharge, where the assumption of the constancy of the current is not necessarily held, can be questioned.

The results of the study discussed here, however, do not support this reasoning. It was observed that there was an increase of 14.0 percent in the emission of the excited state ions at 194.2 nm with the Hg-Ar discharge in the region irradiated with the laser, an increase of 9.5 percent outside the region, and an increase of 7.5 percent both inside and outside the region with the Hg-Ne discharge. The 194.2-nm line corresponds to the $6p^2\ ^2P_{1/2} \rightarrow 6s\ ^2S_{1/2}$ transition in the ion. If the excitation of the ion is assumed to be the result of electron-impact excitation from the ground $6s^2\ \ ^2S_{1/2}$ state, then the excitation rate can be written as $R_i = n_e n_i K$, where n_e and n_i are the electron and ion densities, respectively, and K is the rate coefficient and is related to T_e. An increase in R_i means that one or more of the parameters will have to increase.

It is believed that the major influence of the laser on the electron parameters in the discharge is the increase in T_e, and not in the density n_e. This belief is based on the observed variation of the light intensity of different transitions while the laser illuminates the discharge. Since the intensity of transitions excited by electron collisions is directly proportional to n_e, one would expect a constant change in the intensity of all transitions when n_e increases. It was observed, however, that there were variations in the intensity ranging from 4 to 18 percent for different transitions. On the other hand, such a variation in the intensity is expected if the electron temperature change is the primary result. This is because various transitions have different excitation cross sections for the same electron energy.

It could be argued that the decrease in the metastable population and the accompanying rise in T_e does in fact leave the ground-state ion production a constant, but that the increase in electron temperature has a larger proportionate effect in the excitation of the ion. This could be true if the cross section of the excitation of the ion has a very rapid rise in slope with energy. Unfortunately, the shape of the cross-section curve for this transition is not accurately known at energies corresponding to the high-energy tail of the electron energy distribution of the discharge. A comparison with data on electron-impact excitation of the $^2P_{3/2}$ level [11] hints that such a dependence is conceivable only at threshold energies.

An increase in the number of electrons in the high-energy tail of the electron energy distribution could, however, account for the observed increase in the 194.2-nm emission. Using an equilibrium value of $T_e = 9600$ K and an increase of 3 percent in T_e due to laser irradiation, it was calculated that there was an increase of approximately 14 percent in the number of electrons with energy above the first excitation threshold of mercury, assuming a Maxwellian electron energy distribution. Thus it is possible to attribute, at least partially, the observed increase in the ion emission at 194.2 nm to the increase in the number of high-energy electrons that could produce excited ions by collision with ground-state ions.

Nonetheless, the increase in the electron temperature with laser irradiation is evidently related to the metastable population depletion. Hence, two questions still remain:

(1) What mechanism produces the elevation in the electron temperature?

(2) What other processes may contribute to the observed increase in the 194.2-nm emission of the ion when the laser optical field is applied to the discharge?

It is proposed that the principal mechanism responsible for the increase in electron temperature is the process of two-step electron-impact excitation of the higher lying states: the first excitation occurs from the ground level to the metastable $^3P_2$ level, and a second electron impact excites the atom from the metastable to the higher lying level. The observed decrease in the intensity of the light emission from virtually all higher lying transitions with laser radiation is evidence for the importance of this channel for electron-impact excitation in the discharge. Since inelastic collisions of electrons with discharge particles siphon the electron energy, the diminishing of the electron-impact excitation channel caused by the laser through depopulation of the $^3P_2$ level results in the increase of the average energy of the electrons.

A corollary to the above argument is that the size of the cross section for the excitation of the higher lying states from this metastable level is significantly larger than that for the direct excitation from the ground state, or excitation out of the other levels in the $6\ ^3P$ manifold, whose population is also increased with the applied laser field. The particular finding regarding the role of metastable levels in the excitation of higher lying levels in mercury is well known. The results of this study, however, provide a direct observation of this effect, and also point to the relative significance of the $^3P_2$ level in comparison to the other metastables in the manifold. A quantitative measure of the relative strength of the direct versus the two-step excitation of the levels via the metastables may also be possible with the LIE technique.

The observation that the ion transition is apparently independent of the $^3P_2$ population is somewhat surprising. This is because there have been suggestions in the

past [8,11] that a significant mechanism for the production of mercury ions is via the autoionization of high-lying levels that converge to the 6 ^2D manifold of the ion. The autoionizing levels themselves are expected to be primarily populated by electron collision with the metastable $^3P_2$ level of the atom, since their direct excitation would require higher electron energies not available in a low-temperature discharge.

The observation of the increase in the ion emission with the decrease in the metastable population also apparently weakens the case for a suggested [8] associative ionization process involving the $^3P_2$ level in collision with other metastables:

$$\mathrm{Hg}(6\,^3P_2) + \mathrm{Hg}(6\,^3P_2) \rightarrow \mathrm{Hg}^+ + \mathrm{Hg}(6\,^1S_0) + e \qquad (1)$$

If this were an important mechanism for ion production in discharges of the type in this study, the laser depopulation of the $^3P_2$ would produce a decrease in the intensity of the 194 transition, rather than the observed increase.

The observation of an increase in the 194.2-nm ion emission in parallel with the decrease in the $^3P_2$ population in the laser-irradiated discharge leads to the conclusion that the $^3P_0$ and the $^3P_1$ levels, whose populations increase with the applied laser field, may have a more significant role in the ionization process than the $^3P_2$ level. This conclusion is in contrast to previous works [7,12] where the $^3P_2$ level is regarded to be at least as important as the other metastables in ion production. The $^3P_1$ level, on the other hand, could be the major contributor to the ionization process in the discharge through processes such as

$$\mathrm{Hg}(6\,^3P_1) + \mathrm{Hg}(6\,^3P_1) \rightarrow \mathrm{Hg}(6p'\,^1P^o) + \mathrm{Hg}(6\,^1S_0)$$

$$\mathrm{Hg}(6p'\,^1P^o) + e \rightarrow \mathrm{Hg}^+ + 2e \qquad (2)$$

The population of the 6 $^3P_1$ level may also increase outside of the laser beam in the discharge, thus providing an explanation for the observed increase in the 194.2-nm emission away from the discharge axis. This is possible through the absorption of the excess 254-nm light created by the laser. As mentioned above, 254-nm line emission was observed to increase both inside and outside the region of laser illumination. Thus, radiation transport increases the population of the $^3P_1$ level outside the region illuminated by the laser.

It should be emphasized that the second process above is believed by Vriens et al. [8] to have a significantly smaller rate than the first process, which was excluded as an explanation for the results. If this first process does in fact have a higher rate than the second process, then other processes involving the $^3P_1$ and $^3P_0$ level must be examined to explain the observed behavior of the discharge emission. Candidate processes include electron-impact excitation of highly excited states, which are in turn produced by collision of atoms in the $^3P_{1,0}$ states [4]. Collisions involving these states have also been found to be responsible for the formation of molecular ions [13], which could subsequently result in 194.2-nm emission by electron-impact dissociation.

The results relating to the change of the intensity with the position of the monochromator (depicted in Fig. 5) are readily explained. When the monochromator views the region illuminated by the laser, only the portion of the observed light intensity affected by the depletion of the metastable population is observed. Note here that the decreased light intensity is the sum of the decrease due to the depletion of the metastable and increased light intensity due to the increase in the electron temperature.

As the monochromator is moved away from the axis of the discharge, it simultaneously views the region that is illuminated by the laser and regions outside of the illuminated area. Depending on the size of the two regions within the field of view of the monochromator, and depending on the ratio of the rate of direct excitation over excitation involving the $^3P_2$ level, the intensity of each transition will change. Obviously, there will also be a region where the two processes exactly balance each other out, and no change in intensity will be observed, as seen in Fig. 5.

V. Summary and Conclusion

This study examined the role of the $^3P_2$ metastable of mercury in a low-pressure discharge. The study used a simple discharge at low mercury and buffer gas pressure, together with the inductively coupled rf excitation mode, in order to make the results of the study more generally applicable. The discharge was then probed with a new technique of laser spectroscopy to study the effect of the metastable population on all discharge emissions.

The results have led to essentially three conclusions:

(1) The $^3P_2$ level plays a significant role in electron-impact excitation of high-lying levels of mercury.

(2) The observed increase in electron temperature with the applied optical field is due to the diminishing of

the electron-excitation channel in collision with the $^3P_2$ metastables, which increases the average energy of the electron distribution in the discharge.

(3) The $^3P_2$ level appears to play a smaller role in the ionization rate than the $^3P_1$ and $^3P_0$ levels.

The second conclusion implies that a quenching of the $^3P_2$ population may increase the electron energy and the emission of the 194.2-nm ion line and 253.7-nm line of the neutral mercury. This implies that more efficient lamps for the mercury-ion standard may be designed through approaches that result in quenching the population of the $^3P_2$ level. This may be accomplished, for example, by adding a small amount of a molecular gas such as nitrogen that could quench the metastable level by collision. The last conclusion is in contrast to the general view regarding the role of metastables in the ionization processes in mercury-rare gas discharges. The article pointed out candidate processes involving $^3P_{1,0}$ levels that could lead to the observed 194.2-nm emission of the mercury ion. Further studies are required to ascertain the role of metastable levels in the ionization process in mercury-argon discharges.

Laser-induced emission spectroscopy may be useful in making quantitative measurements of the relative rates and cross sections for the direct and two-step collisional processes involving the $^3P_2$ level. This could be accomplished if the magnitude of metastable population in the field of view of the monochromator could be determined. Future studies to employ the technique described here, together with studies employing other buffer gases, are currently being planned. Processes involving $^3P_{1,2}$ collisions that were identified as possible mechanisms for the observed increase in 194.2-nm emission require further investigations to develop a better understanding of their role in low-pressure glow discharges.

Acknowledgments

The authors wish to thank J. M. Christman and J. L. Streete for their contributions in the initial stages of this experiment.

References

[1] W. H. Richardson, L. Maleki, and E. Garmire, "Experimental and Theoretical Investigation of the Optogalvanic Effect in a Hg-Ar Discharge," *Phys. Rev. A*, vol. 36, no. 12, pp. 5713–5728, December 15, 1987.

[2] P. van de Weijer and R. M. M. Cremers, "Pulsed Optical Pumping in Low Pressure Mercury Discharges," in *Radiative Processes in Discharge Plasmas*, J. M. Proud and L. H. Laussen, editors, New York: Plenum, 1986.

[3] M. Koedam and A. A. Kruithof, "Transmission of the Visible Mercury Triplet by the Low Pressure Mercury-Argon Discharge: Concentration of the 6 ^3P States," *Physica*, vol. 28, no. 1, pp. 80–100, January 1962.

[4] P. van de Weijer and R. M. M. Cremers, "Experimental Determination of 6 ^3P – 6 ^3P Collisional Excitation Cross Sections for Line Emission in the Positive Column of dc Mercury Discharges," *J. Appl. Phys.*, vol. 53, no. 3, pt. 1, pp. 1401–1408, March 1982.

[5] P. D. Johnson, "Excitation of Hg^+ 1942 Å in the High-Current–Low-Pressure Discharge," *Appl. Phys. Lett.*, vol. 18, no. 9, pp. 381–382, 1971.

[6] T. Lin and T. Goto, "Hg^+ Ion Density in Low-Pressure Ar-Hg Discharge Used for a Mercury Lamp," *J. Appl. Phys.*, vol. 66, no. 7, pp. 2779–2282, October 1, 1989.

[7] M. A. Cayless, "Theory of the Positive Column in Rare-Gas Discharge," *Brit. J. Appl. Phys.*, vol. 14, no. 12, pp. 863–869, December 1963.

[8] L. Vriens, R. A. J. Keijser, and F. A. S. Lighthart, "Ionization Processes in the Positive Column of the Low-Pressure Hg-Ar Discharge," *J. Appl. Phys.*, vol. 49, no. 7, pp. 3807–3813, July 1978.

[9] J. T. Dakin, "A Model of Radial Variations in the Low-Pressure Mercury-Argon Positive Column," *J. Appl. Phys.*, vol. 60, no. 2, pp. 563–570, July 15, 1986.

[10] J. D. Prestage, G. J. Dick, and L. Maleki, "The JPL Trapped Ion Frequency Standard Development," *Proceedings of the 19th Annual Precise Time and Time Interval Applications and Planning Meeting*, R. L. Sydnor, editor, Washington, D.C.: U.S. Naval Observatory, pp. 285–297, 1987.

[11] D. H. Crandall, R. A. Phaneuf, and G. H. Dunn, "Electron Impact Excitation of Hg," *Phys. Rev. A*, vol. 11, no. 4, pp. 1223–1232, April 1975.

[12] J. Polman, J. E. Van der Werf, and P. C. Drop, "Nonlinear Effects in the Positive Column of a Strongly Modulated Mercury-Rare Gas Discharge," *J. Phys. D*, vol. 5, no. 2, pp. 266–279, February 1972.

[13] K. L. Tan and A. von Engel, "Energy transfer from excited mercury atoms to electrons," *Proc. R. Soc. London Ser. A*, vol. 324, pp. 183–200, 1971.

[14] F. Cabbanes and J. Chapelle, "Spectroscopic Plasma Diagnostics," in *Reactions Under Plasma Conditions*, M. Venugopalan, editor, pp. 367–469, New York: Wiley-Interscience, 1971.

Table 1. Experimental results

Wavelength, nm	Transition	Percent change Ar on-axis	Percent change Ar off-axis	Percent change Ne on-axis	Percent change Ne off-axis
690.7	$8^3P_2 - 7^3S$	−43	24	−20	4
579.0	$6^1D_2 - 6^1P$	−6.3	6.0	−7.8	4
	$6^3D_1 - 6^1P_1$				
577.0	$6^3D_2 - 6^1P_1$	−7.4	6.1	−16	2.5
546.1	$7^3S - 6^3P_2$	175	18	363	33
435.8	$7^3S - 6^3P_1$	137	50	213	31
407.8	$7^1S - 6^3P_1$	−7	13	−4	5
404.7	$7^3S - 6^3P_0$	311	32	152	36
390.6	$8^1D - 6^1P$	−18	10	−7	5
366.3	$6^1D - 6^3P_2$	−13	6.2	−3.9	5.9
	$6^3D_1 - 6^3P_2$				
365.4	$6^3D_2 - 6^3P_2$	−21	6	−11	5
365.0	$6^3D_3 - 6^3P_2$	−7	3.5	−8.6	4
334.1	$8^3S - 6^3P_2$	−17	11	−6.7	5.2
313.2	$6^1D - 6^3P_1$	−11	5.4	−4.7	5.8
	$6^3D_1 - 6^3P_1$				
312.6	$6^3D_2 - 6^3P_1$	−18	4.7	−9.5	4.4
302.3	$7^3D_1 - 6^3P_2$	−8.2	8.6	−4.5	7.8
	$7^3D_2 - 6^3P_2$				
	$7^3D_3 - 6^3P_2$				
296.7	$6^1D - 6^3P_0$	−12	6	−9	4
	$6^3D_1 - 6^3P_0$				
292.5	$9^3S - 6^3P_2$	−13	9	−6	6
289.4	$8^3S - 6^3P_1$	−13	8	−7	5
280.5	$8^1D - 6^3P_2$	−11	10	−6	7
	$8^3D_1 - 6^3P_2$				
	$8^3D_2 - 6^3P_2$				
	$8^3D_3 - 6^3P_2$				
276.0	$10^3S - 6^3P_2$	−14	7	−5	5
275.3	$8^3S - 6^3P_0$	−15	9	−8	5
269.9	$9^3D_3 - 6^3P_2$	−6	13	−4	7
265.4	$7^1D - 6^3P_1$	−14	7	−7	7
	$7^3D_1 - 6^3P_1$				
	$7^3D_2 - 6^3P_1$				
257.6	$9^3S - 6^3P_1$	−16	8	−7	5
253.7	$6^3P_1 - 6^1S$	12.3	3.0	4.6	2.1
248.3	$8^1D_2 - 6^3P_1$	−13	7	−7	6
	$8^3D_1 - 6^3P_1$				
	$8^3D_2 - 6^3P_1$				
239.9	$9^3D_2 - 6^3P_1$	−10	8	−7	7
237.8	$8^3D_1 - 6^3P_0$	−11	7	−1	6
235.2	$10^3D_2 - 6^3P_1$	−10	10	−5	10
230.2	$9^3D_1 - 6^3P_0$	−14	10	−7	8
194.2	$6^2P_{1/2} - 6^2S_{1/2}$ (ion transition)	14	10	7	7

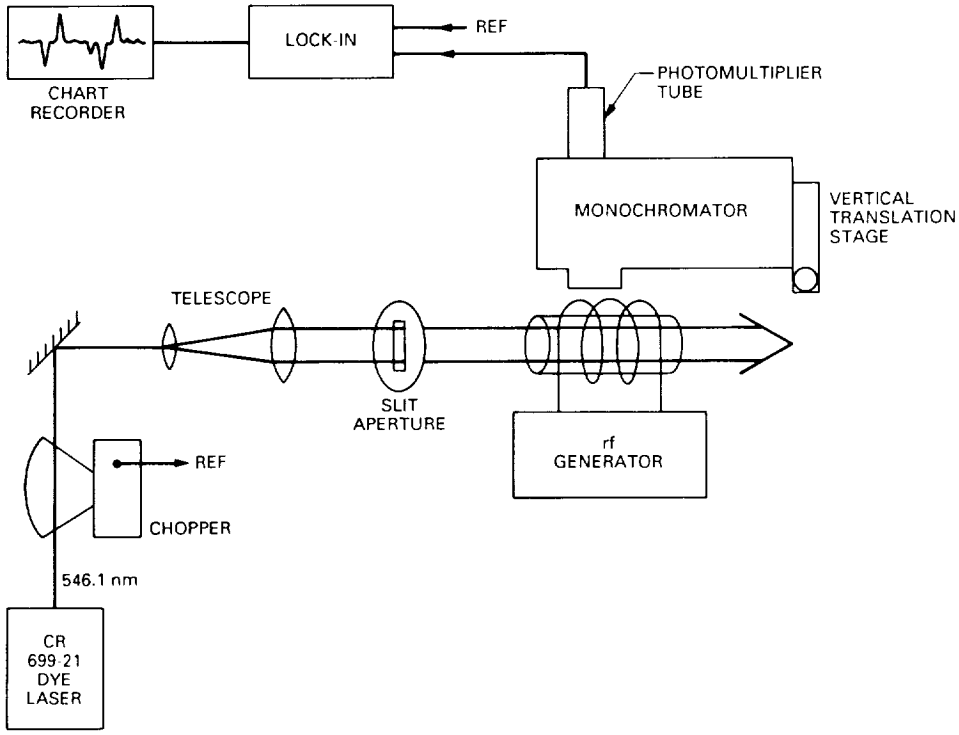


Fig. 1. Schematic of the apparatus.

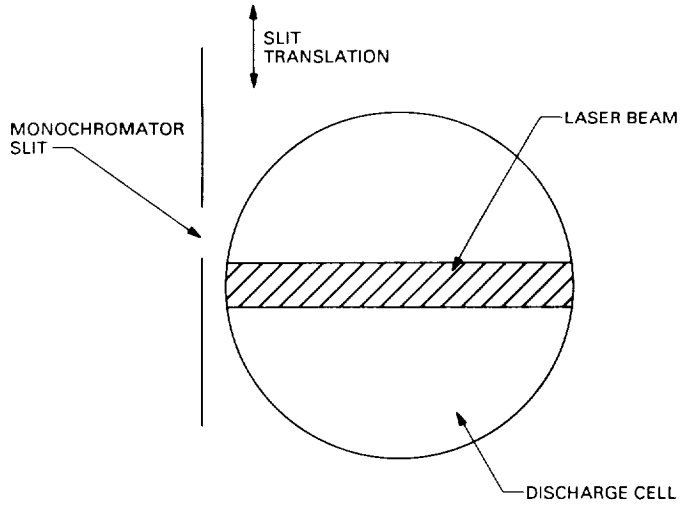


Fig. 2. Detailed view of the observation region.

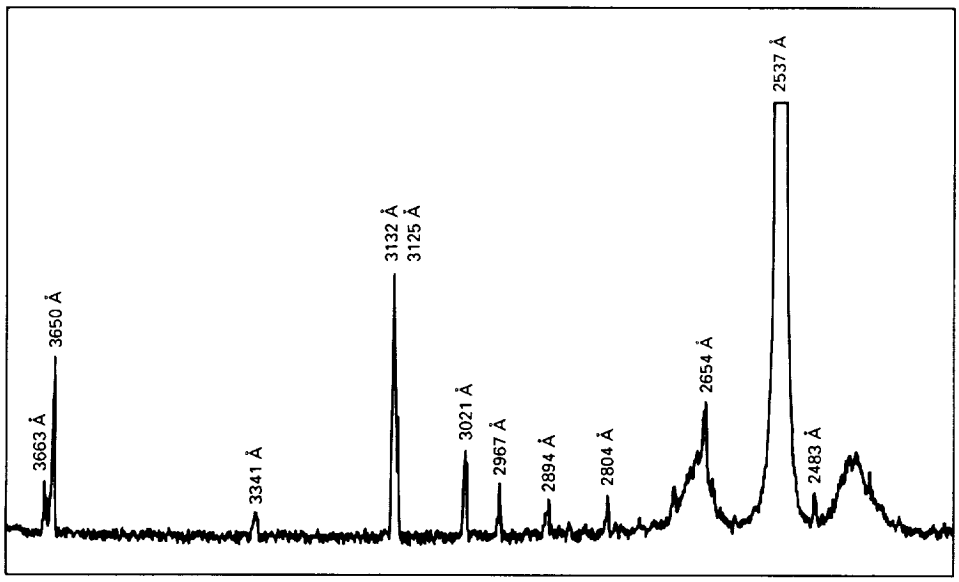


Fig. 3. A partial scan of the monochromator while it is viewing outside the laser ribbon. Positive peaks indicate that the emission increases while the laser is present. The two broad peaks are artifacts of the monochromator connected with the resonance line.

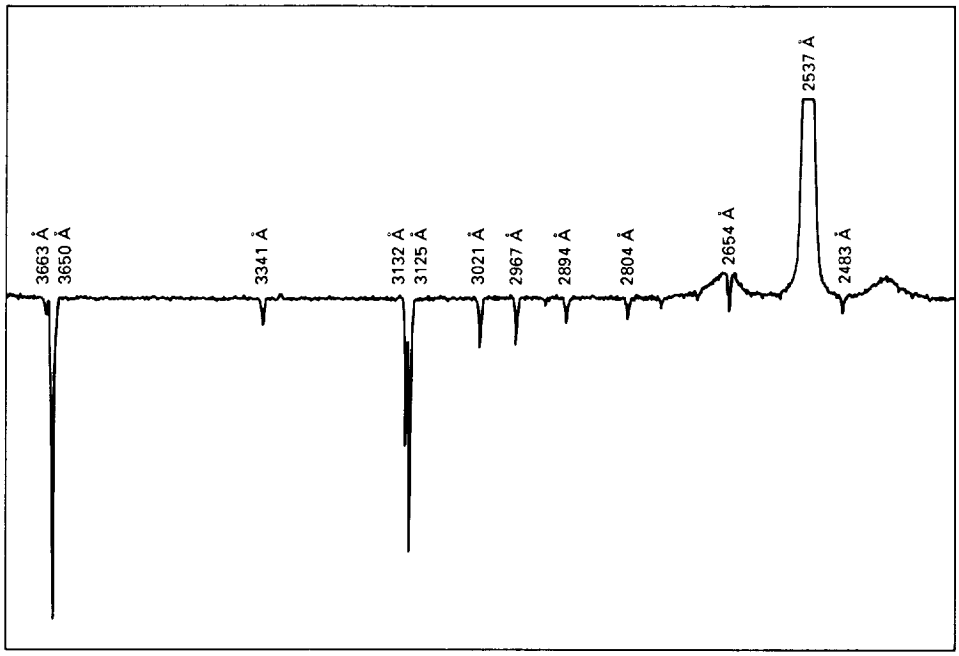


Fig. 4. A partial scan over the same range as in Fig. 3, but viewing within the laser ribbon. All lines except the resonance at 2537 Å have reversed, indicating that they now reduce in intensity while the laser is present.

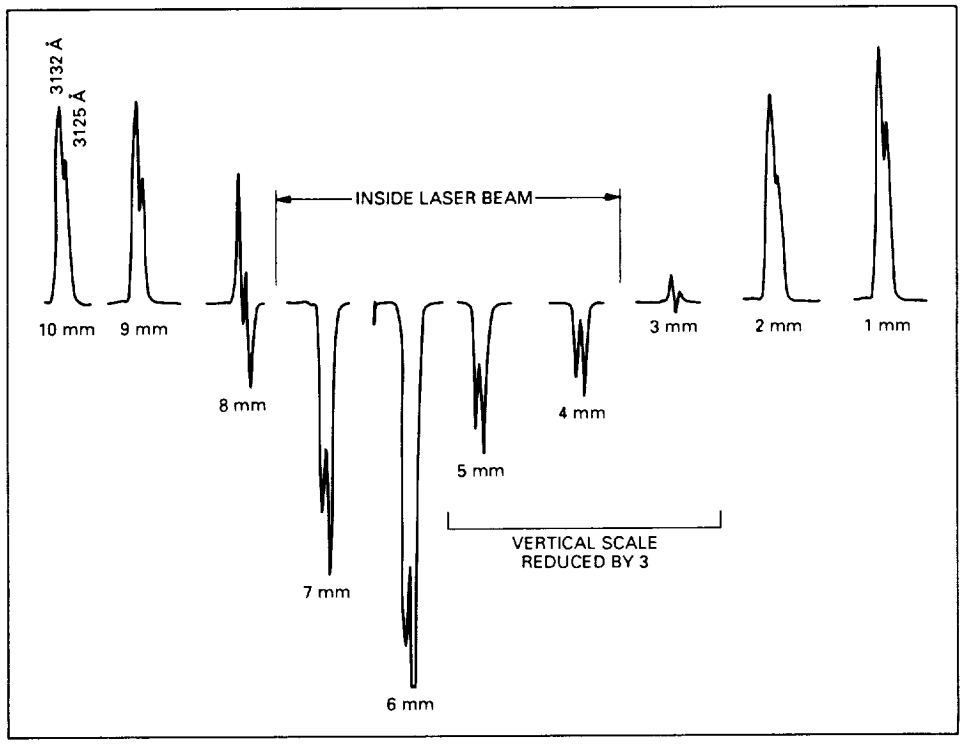


Fig. 5. Short spectral scans over the 3132, 3125 Å doublet as a function of monochromator translation at intervals of 1 mm. The reversal while viewing within the laser beam is clearly apparent.

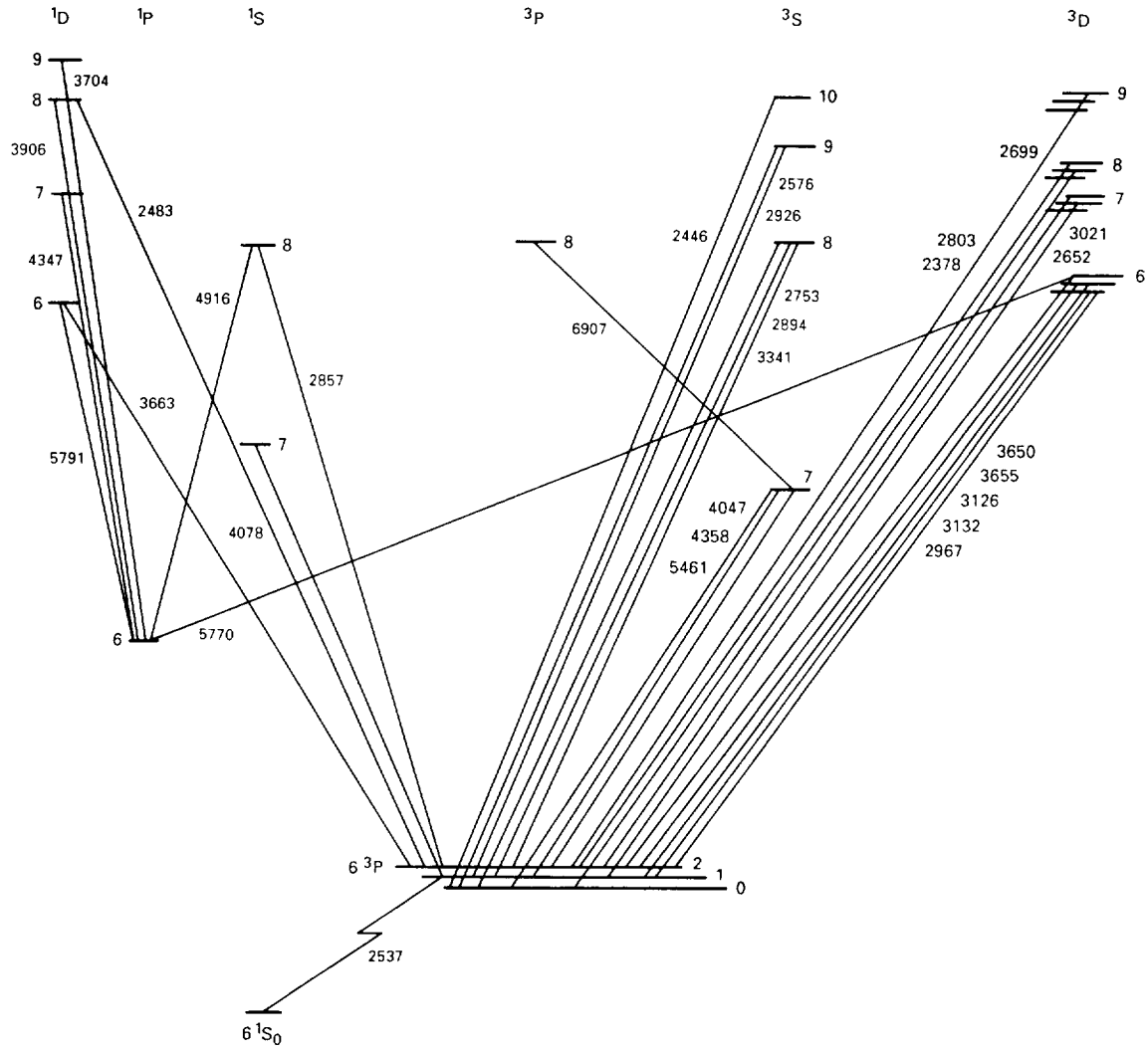


Fig. 6. A Grotrian diagram showing all the transitions observed in Hg I.

Application of Inertial Instruments for DSN Antenna Pointing and Tracking

D. B. Eldred, N. M. Nerheim, and K. G. Holmes
Guidance and Control Section

The feasibility of using inertial instruments to determine the pointing attitude of NASA's Deep Space Network antennas is examined. The objective is to obtain 1-mdeg pointing knowledge in both blind pointing and tracking modes to facilitate operation of the Deep Space Network 70-m antennas at 32 GHz. A measurement system employing accelerometers, an inclinometer, and optical gyroscopes is proposed. The initial pointing attitude is established by determining the direction of the local gravity vector using the accelerometers and the inclinometer, and the Earth's spin axis using the gyroscopes. Pointing during long-term tracking is maintained by integrating the gyroscope rates and augmenting these measurements with knowledge of the local gravity vector. A minimum-variance estimator is used to combine measurements to obtain the antenna pointing attitude. A key feature of the algorithm is its ability to recalibrate accelerometer parameters during operation. A survey of available inertial instrument technologies is also given.

I. Introduction

Currently, the Deep Space Network (DSN) antennas are pointed using either precision angle encoders mounted at the reflector azimuth and elevation axes, or a Master Equatorial (ME), which optically determines the angular difference between the antenna and a mirror, which is normal to the boresight and attached to an intermediate reference structure (IRS) on the back side of the primary reflector. The ME is more accurate than the angle encoders; thus the 70-m antennas, which have the most stringent pointing requirements, use MEs. The smaller 26- and 34-m antennas, which have less stringent pointing requirements than the 70-m antennas, rely solely on the angle encoders. Even if the encoders or ME were error-free, there would still be pointing errors arising from uncertainty in the modeling of the IRS, caused by gravity, wind, and thermal effects, as well as by parameter errors, nonlinearities, and model truncation. For example, wind and gravity distort the antenna dish-bearing structures as well as the ME tower itself, which causes errors in the virtual reference plane on the dish relative to the ME mirror

plane and thereby alters the antenna boresight. A strategy has been proposed and analyzed [1] which uses the Spatial High Accuracy Position Encoding Sensor (SHAPES) developed at the Jet Propulsion Laboratory (JPL) to measure distortion in the antenna dish. When combined with a maximum-likelihood estimator, the method can be used to determine the best-fit antenna boresight. However, this sensor still has to be referenced to a known and stable reference plane. There are approximately 25 ft of structure between the ME and the front surface of the 70-m antenna, and translation of a pointing attitude at the ME to a pointing attitude at the reflector surface requires an accurate model of this connecting structure, which does not exist. Presently, gravity-induced sag, systematic component errors, and misalignments are largely removed while other structural effects are treated as random errors and are left uncompensated.

The current study was undertaken to investigate the feasibility of using an inertial pointing system to achieve pointing and tracking performance consistent with operation of the 70-m antennas at 32 GHz. The system would ideally function as a "black box" that outputs pointing attitude upon request. It would be mounted unobtrusively close to the front surface of the main antenna reflector (Fig. 1). This location would eliminate the error introduced by uncertainty in the model of the structure between the ME and the antenna surface, and would provide an antenna pointing reference that can be used to steer the antenna or from which SHAPES or another surface-sensing sensor could measure. The system would be required to operate in real time and provide capabilities for both blind pointing and precision tracking. A pointing-vector sensing accuracy of 1 mdeg root mean square (rms) or better over a 10-hr tracking period is used as a performance target in the study and serves as the basis for sorting out the potential inertial instrument candidates. This performance requirement is driven by the narrow radio-frequency beam resulting from 32-GHz operation of the 70-m antennas. Additional factors that were considered include reliability, maintenance requirements, and instrument cost.

The investigation includes both the development and analysis of concepts for determining pointing attitude and an assessment of state-of-the-art instruments. Critical issues include attitude initialization and on-line compensation for pointing system errors throughout antenna tracking periods. It was recognized at the onset of the study that the requirements could not be met easily with the available instruments. However, by using multiple instrument arrays, common-mode rejection, and parameter-estimation processing algorithms in real time, the instrument deficiencies are largely overcome, particularly in the determination of the elevation attitude. Simulation results using representative instrument parameters are analyzed to illustrate the type of performance one might expect using currently available instruments operating in a DSN antenna environment.

II. Concept Overview

The proposed instrument to effect the pointing functions combines three single-axis gyroscopes in a three-axis configuration (Fig. 2), four accelerometers (Fig. 3), and an inclinometer. The gyroscopes are arranged with their sensitive axes aligned along three mutually orthogonal axes, with one gyroscope nominally along the antenna boresight axis and the other two perpendicular to this. The accelerometers are configured as a plane array perpendicular to the antenna elevation axis and with their sensitive axes oriented at right angles to one another. The inclinometer is mounted such that it is sensitive to any tilt from horizontal of the antenna elevation axis.

The gyroscopes constitute the most important part of the inertial instrument. Gyroscopes measure either angular rate in an inertial frame of reference or the integral of this rate; thus their outputs can be integrated to track any change in antenna attitude due to movement from an initial position. However, due to instrument errors, the accuracy of the computed attitude degrades with time, and the gyroscopes must occasionally be recalibrated against a known reference. This can be done, for example, immediately before a tracking sequence.

The proposed scenario for determining the absolute antenna attitude for gyroscope initialization involves determination of the direction of the local gravity vector and the Earth's spin axis in the antenna local coordinate frame. The gravity vector is determined using the accelerometers and the inclinometer. The accelerometers measure the component of gravitational acceleration along their sensitive axes, which varies according to the elevation angle of the antenna. The inclinometer directly measures the tilt of the elevation axis. The direction of the Earth's spin vector is determined by clamping the antenna's elevation and azimuth axes. With the antenna fixed in the Earth local reference frame, the only remaining angular rate is that of the Earth, which is observed in the gyroscope outputs.

The instrument is configured as a strapdown system in a single assembly, i.e., the individual components are rigidly attached to a holding fixture, which in turn is securely attached to the antenna dish and moves with it. The alternative to a strapdown system is one that is mounted

on an inertially stabilized platform. This alternative configuration can offer added accuracy but was not seriously considered because of its far greater expense, complexity, and maintenance requirements.

III. Theory

A. Elevation Determination

The following describes the principle of operation for determining the elevation using the accelerometers. An accelerometer measures the component of acceleration along its sensitive axis and is insensitive to cross-axis accelerations (within limits). This principle can be applied to determine its orientation relative to a local gravity vector, from which an inertial geocentric vertical can be obtained. For example, the ideal accelerometer in a gravitational field shown in Fig. 4 would have as its output

$$y = g \sin(\theta) \tag{1}$$

which can be inverted to yield the elevation angle

$$\theta = \sin^{-1}\left(\frac{y}{g}\right) \tag{2}$$

Here, y is the accelerometer output, g is the local acceleration of gravity, and θ is the angle between the accelerometer and local horizontal. An advantage of using accelerometers for elevation measurement is that they have a wide range of motion; since many high-quality accelerometers are capable of measuring greater than ± 1 g, the effective range for angle measurement is ± 90 deg. A second advantage is that they can be relatively inexpensive, so that several may be used in an inertial measurement system to provide maximum sensitivity and measurement redundancy, with little impact on total cost.

The inclinometer is used to determine how far the accelerometer array deviates from vertical. An inclinometer is a device that directly measures an angle with high accuracy but may have a limited range. This prevents them from being used as alternatives to the accelerometers. In the following analysis, the inclinometer is ignored and the elevation axis is assumed to lie within the horizontal plane. To include the effect of the tilt of the elevation axis would obscure the analysis while not fundamentally changing the algorithms.

The outputs from the four accelerometers shown in Fig. 3 can be written as

$$y_1 = \alpha_1 + (1 + \beta_1) g \sin(\theta) + \nu_1 \tag{3}$$

$$y_2 = \alpha_2 + (1 + \beta_2) g \cos(\theta) + \nu_2 \tag{4}$$

$$y_3 = \alpha_3 - (1 + \beta_3) g \sin(\theta) + \nu_3 \tag{5}$$

$$y_4 = \alpha_4 - (1 + \beta_4) g \cos(\theta) + \nu_4 \tag{6}$$

Here, θ is the antenna elevation angle, y_i is the output signal, α_i is the bias, β_i is the gain factor, and ν_i is the error associated with the ith accelerometer. For the moment, it is assumed that the biases and gains are known quantities, obtained from calibration of the individual accelerometers.

Equations (3) through (6) can be conveniently represented as the single symbolic vector equation

$$y = h(\theta) + \nu \tag{7}$$

where y, h, and ν are 4×1 vectors with obvious connections to the terms in Eqs. (3) through (6).

Since there are four equations but only one unknown, the system is overdetermined. Accordingly, a minimum-variance estimator is used to determine the elevation angle. It may be assumed, for lack of better knowledge of the error distribution, that the errors ν_i are zero mean and uncorrelated, and have variances given by ${\sigma_i}^2$. Then the covariance R of the measurement error vector ν is given by

$$R \equiv E(\nu\nu^T) = \begin{bmatrix} {\sigma_1}^2 & 0 & 0 & 0 \\ 0 & {\sigma_2}^2 & 0 & 0 \\ 0 & 0 & {\sigma_3}^2 & 0 \\ 0 & 0 & 0 & {\sigma_4}^2 \end{bmatrix} \tag{8}$$

The minimum-variance estimate $\hat{\theta}$ of the elevation angle θ minimizes the cost functional

$$J = \frac{1}{2}(y - h(\theta))^T R^{-1} (y - h(\theta)) \tag{9}$$

which amounts to weighting each of Eqs. (3) through (6) according to its expected error variance. Calculus of variations yields the necessary condition

$$\left(\frac{\partial h}{\partial \theta}\right)^T R^{-1}(y - h(\theta)) = 0 \text{ at } \theta = \hat{\theta} \tag{10}$$

Unfortunately, Eq. (10) is nonlinear and cannot easily be solved directly. To make the problem tractable, the equation is linearized about a close solution θ_0 as

$$h(\theta) \approx h(\theta_0) + \frac{\partial h(\theta_0)}{\partial \theta}\,(\theta - \theta_0) \tag{11}$$

Substituting for the four-vector, given by

$$H(\theta_0) \equiv \frac{\partial h(\theta_0)}{\partial \theta} \tag{12}$$

the minimum-variance estimate is obtained, after some algebra:

$$\hat{\theta} \approx \theta_0 \; + \; \left(H(\theta_0)^T R^{-1} H(\theta_0)\right)^{-1} H^T(\theta_0) R^{-1} \left(y - h(\theta_0)\right) \tag{13}$$

Owing to the linear approximation used in the estimator, accuracy is assured only in a small neighborhood of the true elevation angle θ, and so the initial guess θ_0 must be "close" to the true angle, or at least close to the estimate $\hat{\theta}$, since θ is unknown. One way to insure this is to apply the algorithm iteratively. Starting with an initial estimate θ_0 for the elevation angle (obtained from encoders or prior knowledge about what θ should be), $h(\theta)$ is linearized according to Eq. (12) to obtain $H(\theta_0)$. Using this quantity, Eq. (13) is used to obtain a better estimate for the elevation angle. This in turn is used as the new starting estimate for the algorithm, and the process is repeated until no further improvement is observed. In practice, the algorithm has been observed to converge after several iterations, reflecting the slow rate of change of $H(\theta)$ as a function of angle θ.

It is possible to obtain an expression for the covariance of the estimate if the measurement errors can be characterized [3]. It is assumed that the error vector ν is gaussian distributed random and zero mean, which is reasonable for independent analog transducers subject to external noise such as the accelerometers. Then the covariance is given by

$$P_\theta = E(\hat{\theta} - \theta)^2 \tag{14}$$

$$= \left(H^T(\hat{\theta}) R^{-1} H(\hat{\theta})\right)^{-1} \tag{15}$$

If for the moment it is assumed that all the accelerometers have zero bias ($\alpha_i = 0$), unit gain ($\beta_i = 0$), and equal error variances ($\sigma_i = \sigma_a$), then it can be shown after some algebra that the expected standard deviation of the estimate is given by

$$\sigma_\theta = \sqrt{P_\theta} = \frac{\sigma_a}{\sqrt{2}\, g} \tag{16}$$

where again, σ_a is the standard deviation of a single accelerometer, g is the acceleration of gravity, and σ_θ is the predicted standard deviation of the resulting angle measurement.

At this point it is appropriate to discuss some of the advantages of the four-accelerometer configuration. First, the expected angle error given by Eq. (16) is independent of elevation angle θ. In other words, the instrument is uniformly sensitive at all attitudes. This is a consequence of having the instruments mounted along orthogonal axes; although it is possible to determine elevation angle with a single accelerometer, the resulting error in the estimate varies with angle and even diverges when the accelerometer is oriented parallel to gravity. Second, the configuration provides for common-mode rejection. In the estimator, the outputs from two accelerometers in a back-to-back pair are essentially subtracted from one another. Thus, any unintended response common to both accelerometers is effectively negated. The advantage of back-to-back instruments goes even one step further by eliminating some nonlinear terms. In general, the most significant nonlinear term in an accelerometer is the quadratic term, commonly referred to as rectification. This term could arise, for example, in a dynamic environment where seismic vibrations were present. Again, since the outputs from the back-to-back accelerometers are essentially differenced in the algorithm, the rectification term nominally disappears and disappears completely if the accelerometers have identical characteristics.

The description of the minimum-variance estimator used to determine elevation angle from the accelerometer outputs is complete except for the corrections required to accommodate deviations of the gravity vector from the true geocentric vertical. However, parameter estimation using an extended Kalman filter is proposed that provides dramatic improvement of the performance. Because of the common-mode rejection, the largest remaining errors in the accelerometer outputs are caused by errors in the biases α_i and the gains β_i. These errors tend to vary from day to day; thus they cannot be removed consistently by means of a one-time calibration. However, as will be shown, it is possible to obtain estimates of them using the accelerometer outputs. The following description illustrates the underlying principle.

For a given antenna elevation, there are four outputs from the accelerometers. The number of unknowns is nine, which includes the elevation angle, four biases, and four gains. The four measurements are linear combinations of the nine unknowns, in a linearized version of the system. The number of unknowns exceeds the number of measurements, so the system is underdetermined.

Now, consider the antenna at a second, different elevation angle. A second unknown is added—the second elevation angle. Four more outputs from the accelerometers are available. These additional outputs are different linear combinations of the unknowns than the first set of measurements since the system equations are linearized about a different angle θ. Thus, at this point the number of unknowns is 10 and the number of measurements is 8. With the addition of measurements at a third elevation, the number of equations will exceed the number of unknowns, and all unknown parameters can be estimated.

In practice, the parameter estimation is implemented using an extended Kalman filter [3]. The components of the state vector at time t_i are defined as the elevation angle, the four biases, and the four gains:

$$x_i \equiv (\theta\ \alpha_1\ \alpha_2\ \alpha_3\ \alpha_4\ \beta_1\ \beta_2\ \beta_3\ \beta_4)_i^T \tag{17}$$

In the extended Kalman filter formulation, it is assumed that the state update and measurement equations can be written as a single-stage transition with zero-mean gaussian distributed random noise inputs:

$$x_{i+1} = x_i + \omega_i;\ E(\omega_i \omega_i^T) = Q \tag{18}$$

$$y_i = h(x_i) + \nu_i;\ E(\nu_i \nu_i^T) = R \tag{19}$$

Here, ω_i is the process noise and ν_i is the measurement noise. An extended Kalman filter for estimating the state in this system is given by

$$\hat{x}_{i+1}^- = \hat{x}_i^+$$

$$(\hat{x}_i^-)_1 = \hat{\theta}_i$$

$$H = \frac{\partial h(x_i^-)}{\partial x} \tag{20}$$

$$M_i = P_i + Q$$

$$P_i = M_i - M_i H_i^T (H_i M_i H_i^T + R)^{-1}\ H_i M_i$$

$$K_i = P_i\ H_i^T R^{-1}$$

$$x_i^+ = x_i^- + K_i(y_i - h(x_i^-))$$

Here, x_i^- is the state estimate before the Kalman filter update, and x_i^+ is the estimate after update; $\hat{\theta}$ denotes the minimum-variance estimate derived earlier, K_i is the Kalman filter gain, M_i is the 9×9 state error covariance matrix estimate before update, and P_i is the covariance matrix after update. Of course, all dimensions are compatible with the nine-element state vector and the four-element measurement vector.

The validity of the extended Kalman filter hinges on whether or not the linearization is performed about the correct trajectory. For this reason, the minimum-variance estimate of the elevation angle $\hat{\theta}$ is used in the filter so that the best possible parameter estimates are used before linearization. In the simulation, the benefit of this modification was observed in the sense that angle estimation is more accurate using the minimum-variance estimate than not using it.

The covariance propagation equations as given in Eq. (20) are well known to be sensitive to small errors introduced from computational errors, and it is a common problem that positive definiteness in the covariance matrices can be lost, which can cause divergence of the filter. To insure against this sort of problem, a square-root algorithm was used to propagate the covariances [2].

The parameter estimation works best when the antenna is rotated through as wide a range of elevation as possible. This suggests an initial calibration to initialize the accelerometers, which is best accomplished prior to the start of a tracking period. The calibration sequence would require the antenna to be elevated from horizontal to vertical. Even during tracking, the changes in elevation which occur are sufficient to allow the filter to significantly improve the accuracy of the elevation estimate.

B. Gyroscope Initialization

In order to obtain the antenna attitude angle using the gyroscopes, the measured angular rates are integrated over time. Errors in the gyroscope outputs, which may include both systematic and random errors, are simultaneously integrated, resulting in corresponding errors in the computed attitude. Thus it is necessary to reinitialize the gyroscopes occasionally to determine their integration constants. To initialize the gyroscopes, the antenna's attitude in inertial space must be known at some instant in time, at which point the offset between the computed attitude and actual attitude can be computed. The offset is then subtracted from all subsequent computed attitudes to provide a more accurate attitude estimate. It is not critical where the antenna is pointed to obtain this initial attitude, since it can be slewed to a desired target afterwards using the gyroscopes and the accelerometers to guide the antenna along its desired trajectory.

Determining the antenna's attitude is equivalent to determining the coordinate transformation between the antenna's local coordinate system and some known reference system, e.g., inertial or Earth-based. The most commonly used transformation is a 3×3 rotation matrix, which is

composed of nine elements that consist of the direction cosines between the coordinate axes in the local system and the corresponding axes in the reference system. An alternative formulation using quaternions has computational advantages for real-time computations [2] and will likely be used in an implementation of the proposed concept; however, quaternions are not used in the analysis presented here and therefore are not discussed further in this article. Of the nine parameters that constitute a rotation matrix, only three are independent. Thus to evaluate the rotation matrix, or equivalently, the antenna attitude, at least three parameters must be determined.

Two of the three required parameters can be determined by establishing the direction, in the antenna reference system, of a vector whose direction is known in the absolute reference frame a priori. For example, the direction of any vector originating from the center of the Earth is specified by the two parameters of latitude and longitude. Once the direction of a second, different vector is known, the coordinate transformation is overspecified and can be estimated by combining the measurements in a minimum-variance estimator. One example of a vector that could be used to determine absolute antenna attitude is the local gravity vector. Its direction can be estimated using the algorithms described in the previous section. A second candidate vector is the Earth's spin vector. This can be determined using the gyroscopes, which fundamentally measure angular rate about their sensitive axes. However, one must be careful to accommodate any rate caused by motion of the antenna relative to the Earth. A third candidate vector is a ground-based optical beacon. The antenna could be pointed towards a beacon whose position is known from a prior survey, using a star tracker in a closed-loop system. This technique has the disadvantage that a star tracker is required and is also subject to errors caused by refraction in the atmosphere. Yet another candidate vector for initialization is a celestial object such as a star or radio source. The primary disadvantage of determining the direction of such a vector is that the antenna must actively track the object to determine its direction; additional disadvantages include errors caused by atmospheric refraction and the necessity for a star tracker (for an optical source).

The proposed scenario for initializing the gyroscopes uses determination of the local gravity vector and the Earth's spin vector to determine the required coordinate transformation. The scenario requires temporarily locking the antenna in its elevation and azimuth axes to hold the antenna rate relative to Earth to zero, at least in a mean sense if seismic vibrations are present, so that the accelerometers and gyroscopes can be averaged over a period of time to obtain the best possible accuracy. It was recognized from the beginning that the requirements on the gyroscopes would be severe to determine the Earth's spin vector with sufficient accuracy for 1-mdeg pointing. It is estimated from the statistical properties of available gyroscopes that the antenna will have to be kept stationary for approximately 15 minutes in order to achieve this level of pointing knowledge. This is a burden on antenna operations which must ultimately be weighed against the disadvantages of alternative initialization strategies such as those utilizing star trackers.

A minimum-variance estimator is used to determine the direction of the Earth's spin axis using the outputs from the three gyroscopes. Figure 5 shows the relative orientations of the antenna local reference frame, the gyroscopes, and the Earth's spin axis. The corresponding measurement equations are

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} \Omega_1 \\ \Omega_2 \\ \Omega_3 \end{pmatrix} + \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix} \tag{21}$$

Here, Ω_i is the component of the Earth's spin vector along the ith coordinate axis, y_i is the rate output associated with the ith gyroscope, and ν_i is the corresponding rate-error term. The gyroscope errors ν_i consist of both random errors such as might be caused by electronic noise, and deterministic components such as those caused by imperfect calibration, drift in gyroscope parameters, uncompensated temperature effects, etc. It is reasonable to assume that since the gyroscopes are identical within manufacturing tolerances, the expected variances of the error terms should be approximately equal, at least on the average:

$$E(\nu_1{}^2) = E(\nu_2{}^2) = E(\nu_3{}^2) = R \tag{22}$$

The corresponding minimum-variance estimate of the Earth spin vector, $\hat{\Omega}$, is found by minimizing the cost functional

$$J = \frac{1}{2R} \sum_{i=1}^{3} (y_i - \Omega_i)^2 \tag{23}$$

from which the components of the spin vector estimate can be obtained

$$\hat{\Omega}_i = y_i \tag{24}$$

Because the measurements are taken along independent axes, the error in the estimate $\hat{\Omega}$ is given, where I is the 3×3 identity matrix, by

$$E(\hat{\Omega} - \Omega)(\hat{\Omega} - \Omega)^T = RI \tag{25}$$

Within one standard deviation, the estimate lies within a sphere of radius $\sqrt{R}$ of the true spin vector, as illustrated in Fig. 6. It can be seen that the angular error in the estimate of the Earth's spin vector is bounded by

$$\sigma_s \leq \frac{\sqrt{R}}{|\Omega|} \tag{26}$$

Here, σ_s is the standard deviation of the angular error in the estimated spin vector.

With knowledge of the Earth's spin vector and the local gravity vector and estimates of their respective errors, it is possible to construct an estimator that combines the local gravity vector with the Earth spin vector to determine the rotation matrix Q which relates the antenna local reference frame to the Earth local frame, i.e.,

$$e_i = Qe_i' \tag{27}$$

Here, e_i is the ith unit vector in the Earth or "unprimed" frame and e_i' is the corresponding unit vector in the antenna or "primed" frame. The rotation matrix Q can be written as the product of two rotations as

$$Q = Q_1 * Q_0 \tag{28}$$

where Q_0 is the known a priori estimate and Q_1 is a rotation matrix involving small angles that needs to be determined. An intermediate "double-primed" reference frame is defined by

$$e_i'' = Q_0 e_i' \tag{29}$$

where components e_i'' consist of a priori estimates of the Earth-based unit vectors e_i. Thus, it is possible to write

$$e_i = Q_1 e_i'' \tag{30}$$

If Q_0 is a close approximation to Q, then Q_1 can be written using a small-angle approximation. Figure 7 shows the "unprimed" and "double-primed" frames, and it can be seen that for small rotations, Q_1 has the form

$$Q_1 \approx \begin{pmatrix} 1 & -\alpha_3 & \alpha_2 \\ \alpha_3 & 1 & -\alpha_1 \\ -\alpha_2 & \alpha_1 & 1 \end{pmatrix} \tag{31}$$

Here, α_i is the rotation angle about the ith coordinate axis in the "unprimed" frame. The rotation angles α_i, and hence Q_1, can be estimated using a minimum-variance estimator that combines the estimate of the Earth spin vector with the estimate for the local vertical. After some algebra, a state-measurement equation can be developed that has the familiar form

$$y = H\alpha + \nu \tag{32}$$

Here, y is a 6×1 vector whose elements consist of the components of the normalized estimated gravity and spin vectors, H is a 6×3 matrix whose elements are functions of the known rotation matrix Q_0, α is a 3×1 vector whose elements consist of the rotation angles α_i, and ν is a 6×1 error vector. Using Eq. (32), the minimum-variance estimator is developed in the usual manner and has the same form as Eq. (13), the minimum-variance estimator for the antenna elevation angle.

The estimator obtained from Eq. (32) remains valid only as long as the linear approximation given by Eq. (31) for Q_1 remains accurate. This will be true only if the initial approximation to the transformation matrix Q_0 is close to the actual transformation Q. A related problem is that the matrix Q computed using Eq. (28) tends to lose orthogonality because of the small-angle approximation made in Eq. (31). The approach used to address these problems is to use an iterative strategy to determine Q, as follows. An initial approximation Q_0 to the rotation matrix Q is made using angle encoder readouts, which provide a very good initial value. The minimum-variance estimator is used to determine the correction term Q_1. From this, the rotation matrix Q is obtained according to Eq. (28). At this point, Q is reorthogonalized. The process is repeated several times, using the most recently calculated Q as the new Q_0.

The covariance of the estimate can be derived analytically, and is given by

$$P = E(\alpha\alpha^T) = \begin{pmatrix} \frac{R_1 R_2}{R_1+R_2} & 0 & 0 \\ 0 & R_1 & R_1\tan(\phi) \\ 0 & R_1\tan(\phi) & R_1\tan^2(\phi) + \frac{R_2}{\cos^2(\phi)} \end{pmatrix} \tag{33}$$

Here, ϕ is the latitude angle of the antenna, R_1 is the variance of the normalized Earth spin vector (in rad^2) and R_2 is the variance of the normalized local gravity vector (again, in rad^2). From Eq. (33), one can determine the accuracy of the estimate as a function of the accuracy with which the spin and gravity vectors are estimated. This can

be used in reverse to determine the requirements on the gyroscopes and the accelerometers, as will be shown below.

C. Slewing and Tracking

Once the gyroscopes have been initialized, the antenna must be slewed into a position that will intersect the desired tracking trajectory. The speed and accuracy with which the antenna can be slewed is governed by a number of factors, including the rate at which the antenna can be moved, the time that can be tolerated to perform a slew, and the buildup of errors caused by gyroscope errors and sampling rate. Tracking a target is virtually identical conceptually to slewing, except that the slew rate is nearly zero because celestial targets remain essentially stationary in an inertial frame.

Since the baseline configuration uses a ring-laser gyroscope (see below), it is considered in the following discussions of slewing and tracking. An optical gyroscope outputs an integrated angle; thus, to obtain the rate used for the Earth spin axis determination, successive outputs must be differenced. A simplified model for the output of a single-axis optical gyroscope at time t_i can be written as

$$y_i = B_i + \epsilon_d t_i + G(1 + \epsilon_g)\theta_i + \eta_i \tag{34}$$

The quantities are defined as: y_i is the gyroscope output at time $t = t_i$, B_i is the bias drift, ϵ_d is the error in the drift rate, t_i is the time elapsed since gyroscope initialization, G is the gyroscope gain, ϵ_g is the error in the gain, θ_i is the total angle through which the gyroscope has rotated about its sensitive axis, and η_i is the measurement noise, which may be taken to be random. In addition, the gyroscope bias B_i is subject to a random-walk phenomenon caused by quantum-mechanical effects or shot noise:

$$B_{i+1} = B_i + \gamma_i \tag{35}$$

Here, γ_i is a random process noise. Equations (34) and (35) govern the growth of errors in a single optical gyroscope. By inspection, the growth of the error terms with time is as follows. Error caused by drift in the bias is a random-walk process and is proportional to $\sqrt{t}$. Error caused by rate drift is proportional to t. That due to gain error is proportional to the change in angle between sampling periods, which is proportional to the antenna slew rate and inversely proportional to the sampling rate. That due to measurement error is proportional to $1/\sqrt{t}$. Figure 8 shows the growth of the error terms with time for a representative optical gyroscope whose parameters are based on conversations with and specifications provided by several gyroscope manufacturers, and are given in Table 1. In this example, the error contribution from the measurement noise dominates for elapsed times less than about 1 second, while the gain error and bias random-walk terms dominate at larger elapsed times. Note that the contribution from gain error at the longer elapsed times can, in general, be reduced by slowing the slew rate until the bias random walk dominates the error.

An additional source of error during slewing and tracking arises from the fact that rotations are not commutative, except in the limit of infinitesimal rotations [4]. Because of this, the trajectory of the antenna must be known in order to correctly interpret the gyroscope outputs; individually integrating the rate outputs from the three gyroscopes will produce an incorrect result. Since the gyroscopes are sampled at discrete times, the antenna trajectory is only approximated. For example, suppose an antenna undergoes an attitude change resulting from successive rotations about two different axes

$$R = R_1(\theta_1) * R_2(\theta_2) \tag{36}$$

If these rotation angles θ_1 and θ_2 are small, then it is possible to make the approximation

$$R_i(\theta_i) \approx I + \Delta_i(\theta_i) \tag{37}$$

where $\Delta_i(\theta_i)$ is linear in θ_i and $|\Delta_i| \leq c|\theta_i|$, where c is some constant that is close to 1. A second, different rotation that gives the identical gyroscope outputs is given by

$$R' = R_2(\theta_2) * R_1(\theta_1) \tag{38}$$

and differs from R by

$$R' - R \approx \Delta_2\Delta_1 - \Delta_1\Delta_2 \tag{39}$$

Substituting the above bound on Δ_i into Eq. (39), one obtains an expression for a bound on the error in the rotation

$$| R' - R | \leq 2c | \theta_1\theta_2 | \tag{40}$$

The accumulated error from n rotations, each of magnitude θ, is thus bounded by approximately

$$| R' - R | \leq 2cn\theta^2 \tag{41}$$

This relation cannot be used directly to determine the error caused by noncommutation of rotations because the constant c depends on the trajectory followed, which depends on the vibration environment of the antenna. The

determination of the antenna vibration environment is beyond the scope of this article, as it requires a simulation that includes the antenna dynamics. Nevertheless, Eq. (41) does illustrate the growth of this type of error; the error is quadratic with rotation angle and linear with the number of rotations. Furthermore, once c is determined, Eq. (31) can be used to specify the maximum allowable angular excursion θ that can be tolerated between gyroscope sampling periods, and consequently, the maximum allowable slew rate and the required sampling rate. For example, using Eq. (41) with $c = 1$, it can be shown that the number of samples n required to perform a 20-deg slew at a rate of 0.1 deg/sec is about 250,000 over the period of the slew, which is 200 seconds. Thus the sample rate is bounded by about 1250 Hz. Since Eq. (31) represents an upper bound, this sample rate is probably faster than it needs to be.

Since the direction of the local gravity vector can be determined and is available, it is reasonable to incorporate this information into yet another minimum-variance estimator that combines it with the attitude estimate obtained from integrating the gyroscope rates. This estimator has essentially the same structure and derivation as that described in Eqs. (27) through (31), but with two differences: first, the estimated boresight of the antenna which is obtained from integrating the gyroscopes is used in place of the estimate of the Earth's spin vector; and second, the gyroscope integration provides additional information consisting of the antenna rotation about its boresight, for use by the estimator. In this case, the covariance of the minimum-variance estimate is given by

$$P = \begin{pmatrix} \frac{R_1 R_2}{R_1 + R_2} & 0 & 0 \\ 0 & \frac{R_2(R_1 + R_2 \sin^2(\phi))}{R_1 + R_2} & \frac{R_2{}^2 \cos(\phi)\sin(\phi)}{R_1 + R_2} \\ 0 & \frac{R_2{}^2 \cos(\phi)\sin(\phi)}{R_1 + R_2} & \frac{R_2(R_1 + R_2 \cos^2(\phi))}{R_1 + R_2} \end{pmatrix} \tag{42}$$

Again, this expression for the attitude covariance can be used to determine instrument requirements. In Eq. (42), the angle ϕ does not refer to the antenna latitude but instead refers to the angle between the antenna boresight and the local gravity vector.

D. Instrument Requirements

In this section, the expressions for the error covariances are used to obtain error bounds for the accelerometers and gyroscopes. These bounds are obtained by requiring that the norm of the attitude covariance matrix P given in Eqs. (33) and (42) be smaller than the required pointing variance.

First the requirements for initialization are considered. The relevant instrument parameters are absolute accelerometer accuracy in μg, and the gyroscope rate accuracy in deg/hr (for example). It can be shown from Eq. (33) that the following inequality holds true:

$$| P | < R_1(1 + \tan^2(\phi)) + \frac{R_2}{\cos^2(\phi)} \tag{43}$$

Specifying that $|P| \leq 1$ mdeg^2, and substituting 1 mdeg = 17.5×10^{-6} rad, requires both that

$$R_1(1 + \tan^2(\phi)) \leq (17.5 \times 10^{-6})^2 \tag{44}$$

and

$$\frac{R_2}{\cos^2(\phi)} \leq (17.5 \times 10^{-6})^2 \tag{45}$$

Recall that R_1 is the angle variance of the normalized local gravity vector; thus, applying Eq. (16) one obtains

$$\sigma_a \leq \sqrt{2} g \cos(\phi) \times 17.5 \times 10^{-6} \tag{46}$$

as a requirement for the accelerometer accuracy σ_a. Substituting for the latitude of the Goldstone complex, $\phi = 35$ deg, the accelerometer requirement becomes

$$\sigma_a \leq 20\ \mu g \tag{47}$$

In other words, the accelerometers must be accurate in an absolute sense to better than 20 μg rms. This accuracy applies to the accelerometers after the extended Kalman filter algorithm has been applied to remove gain and bias errors, so the devices actually used may have relaxed specifications compared to Eq. (47).

The required gyroscope accuracy is given by

$$\frac{R_2}{\cos^2(\phi)} \leq (17.5 \times 10^{-6})^2 \tag{48}$$

From Eq. (26), $R_2 = R/\Omega^2$, where $R = \sigma_r^2$ is the variance of the gyroscope rate and Ω is the magnitude of the Earth's spin rate. Thus, the accuracy requirement on the gyroscope rate is given by

$$\sigma_r \leq \Omega \cos(\phi) \times 17.5 \times 10^{-6} \tag{49}$$

$$\sigma_r \leq 2.1 \times 10^{-4} \text{ deg/hr} \tag{50}$$

It must be recognized that the requirements specified above for the accelerometers and gyroscopes represent upper bounds; when errors from both terms occur simultaneously, each individual term must be correspondingly reduced.

Next, the requirements for tracking at the 1-mdeg level are given. It can be shown, after considerable algebra, that the norm of the covariance matrix for tracking given by Eq. (42) is bounded by

$$| P | \leq R_2 \tag{51}$$

or, imposing the condition that $|P| \leq 1$ mdeg^2, one obtains the necessary condition for the gyroscope drift

$$\sigma_\theta \leq 1 \text{ mdeg} \tag{52}$$

This bound σ_θ is the maximum angular error that can be tolerated from the integration of a single gyroscope at a low slew rate. From Eq. (52), the maximum error in tracking is completely determined by the accuracy in the gyroscopes, and the accelerometers play no role in fixing this bound. This reflects the fact that rotation of the antenna about the local gravity vector can only be sensed by the gyroscopes. This does not indicate, however, that the accelerometers do not benefit tracking performance; on the contrary, they help significantly to determine antenna attitude within the elevation plane.

While the requirements presented above are representative of the performance levels needed to satisfy 1-mdeg attitude knowledge, the system is sufficiently complex that these requirements are overly simplistic. In fact, the true performance of the system for a given set of instruments must be determined via simulation, because of the complex interactions among the various estimation algorithms and the mix of stochastic and systematic error sources.

IV. Component Selection

A. Introduction

A large part of the effort in the feasibility study has been focused on collecting data from the various manufacturers on the performance of available and projected state-of-the-art instruments, including gyroscopes, accelerometers, and inclinometers. This information was needed in order to conduct simulations using representative instrument parameters, and also can be used to make preliminary recommendations for component selection. The instruments suitable for the present application were developed primarily to meet the need for precision inertial guidance for missile systems and other military applications, with some development traceable to spaceflight requirements. The inertial instruments are typically integrated into a single package, either as an inertial platform that retains its attitude in inertial space or as a strapdown system that is rigidly secured to the vehicle. Inertial platforms can be more accurate than strapdown systems and are generally used for long-range navigational systems that must maintain precision for long periods of time. Because of the high cost of inertial platforms, however, strapdown systems are favored for the present application.

The primary sources of information concerning the instrument characteristics were the instrument vendors. For the most part, classified sources were not used in the study. An excellent summary of modern inertial instruments is contained in [5].

B. Gyroscopes

1. Overview. Gyroscopes can be classified as either optical or mechanical. In addition to the familiar spinning-wheel gyroscopes, mechanical gyroscopes include devices such as the hemispherical-resonator gyroscope [6] and the experimental magnetic-resonance gyroscope [5], which do not have any moving parts. Because of their more mature status, the only mechanical devices considered here are the spinning-mass gyroscopes. In spite of the notable lack of a rotating component in the optical devices, the term "gyroscope" has been retained for the optical devices.

Descriptions of the spinning-wheel gyroscopes can be found in standard textbooks such as [7]. There are two major classifications of spinning-wheel gyroscopes. Those constrained to precess only about one axis (the "output" axis) are called single-degree-of-freedom (SDOF) or single-axis gyroscopes, while those that can precess about two axes are referred to as two-axis or two-degree-of-freedom devices. This precession is sensed by the instrument and a torque is generated to counter the rotation about the output axis. Extensive research and development has been devoted to all aspects of precision gyroscopes. As a result, the spinning-wheel gyroscope is much more mature than other types, and consequently shows little potential for radical improvements over the current performance levels. The mechanical gyroscopes considered include the SDOF floated gyroscopes, dynamically tuned gyroscopes, and electrostatically suspended gyroscopes.

Optical gyroscopes are less mature than mechanical gyroscopes and at present are characterized by somewhat larger errors. However, the lack of moving parts in the optical gyroscope provides the basis for a number of potential advantages that include lower initial cost, reduced maintenance and cost of ownership, shorter warmup times,

large dynamic range, greater long-term stability, and insensitivity to acceleration, shock, and vibration. The optical gyroscopes include various versions of the ring-laser gyroscope (RLG) [8] and the interferometric fiber-optic gyroscope (IFOG) [9]. Both types are currently undergoing intense development and show the potential for continued improvement in performance. Of the two, the ring-laser gyroscope represents the more mature technology and is widely used on commercial aircraft as well as for military applications. On the other hand, the fiber-optic gyroscopes are theoretically capable of better performance. An IFOG has been chosen for the guidance system of the upcoming CRAF/Cassini Mariner Mark II space mission.

2. Mechanical gyroscopes.

a. The single-degree-of-freedom-floated gyroscope. The single-degree-of-freedom floated gyroscope (SDOFF) consists of a spinning wheel mounted in a sealed cylindrical float with its spin axis perpendicular to the axis of the cylinder. The cylinder is mounted within a case with bearings that constrain its motion to rotation about the cylinder axis. This axis is also the output axis of the gyroscope. The space between the float and the case is filled with a viscous fluid. To minimize acceleration effects, the fluid is selected to provide neutral buoyancy for the float. The gyroscope input axis is mutually perpendicular to the wheel spin axis and the float axis.

In operation, rotation about the gyroscope input axis produces a torque about the output axis. If the gyroscope is configured as rate-integrating, the float angular rate is such that the viscous drag on the float just balances the gyroscopic torque. Thus, the float rotation rate is proportional to the angular rate about the input axis and the angular deflection of the float is proportional to the time integral of this rate. It is common practice to detect float rotation and apply a countertorque to return the float to its null position. The applied torque is proportional to the rotation rate about the input axis and the angular position of the case is found as the integral of this rate.

Development of the SDOFF gyroscopes during the past 40 years has been mainly in response to the inertial guidance needs for aircraft and ship navigational systems and more recently for missile guidance and control. Examples of the results of this effort for the most precise applications include the use of hydrodynamic gas bearings for wheel support and magnetic suspensions of the float output axis. Also, considerable effort has been devoted to the development of a single-species flotation fluid to avoid errors caused by stratification of mixed-polymer fluids. The need for precision pointing and tracking of directed-energy weapons has placed an emphasis on low noise.

Both Charles Stark Draper Laboratories (CSDL), previously the MIT Instrumentation Laboratory [10], and Northrop Corporation [11] have developed SDOFF gyroscopes that appear adequate for the antenna pointing requirements. Typical performance of the fourth-generation gyroscopes are drift stability of 0.0001 deg/hr over a 24-hr period, and rate uncertainty, which is primarily due to the gyroscope electronics, of 1.6×10^{-5} deg/hr. A simplified cutaway drawing of the CSDL Fourth Generation Technology Demonstration Device is shown in Fig. 9.

b. The dynamically tuned gyroscope. The dynamically tuned gyroscope (DTG) is a spinning-wheel gyroscope that avoids the use of temperature-sensitive flotation fluids that is characteristic of the SDOFF. The DTG was developed in the early 1960s for applications that required medium accuracy in moderately severe environments. Successful applications include use for the inertial reference units of the Voyager spacecraft (DRIRU I) and the Magellan spacecraft (SKIRU) [12]. Another DTG is DRIRU II [13] developed by Teledyne for NASA as the NASA "standard" inertial reference unit.

Although development of the DTG is continuing, high-grade DTGs do not match the accuracy of precision SDOFF gyroscopes and probably would not satisfy the antenna pointing requirements. A cross section of the Teledyne DTG is shown in Fig. 10.

c. The electrostatically suspended gyroscope. Development of the electrostatically suspended gyroscope (ESG) [14] began during the late 1950s, and 0.001-deg/hr gyroscope system performance was demonstrated by 1971. The ESG has undergone continued improvement since then. A major attribute of this system, which has found application as an accurate submarine navigational system, is its long-term stability and high precision.

The ESG is a free-rotor system consisting of a metallic sphere that is spun up to about 3600 rev/sec and electrically suspended in a hard vacuum. The operating period is measured in years. The spin axis remains fixed in inertial space and its position relative to the case is determined by pickoffs. The ESG is produced by both Honeywell and Rockwell International. The Honeywell instrument is made with a hollow sphere and the spin axis is measured optically by observation through a window. The Rockwell device is made with a solid sphere whose center of mass is slightly offset from the geometric center, which causes a slight wobble that can be detected optically, thereby facilitating determination of the spin axis.

The ESG is generally regarded as the most accurate gyroscope available for shipboard navigational systems when

used on a stabilized platform. Unfortunately, the performance figures are classified. It is also a very expensive system and for this reason was not considered for the present application.

3. Optical gyroscopes. Optical gyroscopes are based on the Sagnac effect, which predicts that two optical waves traveling in opposite directions around identical closed paths will experience pathlength differences that are proportional to the rotation rate of the closed loop [15]. The process is shown schematically in Fig. 11, in which optical waves are introduced into a circular loop at point A and traverse a common path in opposite directions. If the loop (and point A) rotate in the clockwise direction, the clockwise wave requires a longer time to arrive back at point A than does the counterpropagating counterclockwise wave. It can be shown that the time difference δt for the arrival times of the two waves at point A and the associated pathlength difference δL are given by

$$\delta t = \left(\frac{LD}{c^2}\right)\Omega \tag{53}$$

and

$$\delta L = c\delta t = \left(\frac{LD}{c}\right)\Omega \tag{54}$$

where D is the loop diameter, $L = \pi D$ is the optical pathlength, Ω is the loop rotation rate, and c is the speed of light. A more general derivation that includes noncircular paths shows that LD may be replaced by $4A$, where A is the area enclosed by the optical loop.

The different types of optical gyroscopes use different techniques to measure the rotation-induced pathlength difference. The optical gyroscopes considered here are the ring-laser gyroscope and the interferometric fiber-optic gyroscope, as represented by the JPL fiber-optic rotation sensor (FORS) [16]. The ring-laser gyroscope takes advantage of the lasing characteristics of an optical cavity between two cavity resonance frequencies. For a fiber-optic gyroscope, L is the total length of optical fiber comprising the optical path. The rotation-induced pathlength difference is sensed as an optical phase difference of the two waves. The phase difference $\delta\phi$ is given by

$$\delta\phi = \frac{2\pi\delta L}{\lambda} = \left(\frac{2\pi LD}{\lambda c}\right)\Omega \tag{55}$$

where λ is the optical wavelength.

It is expected that continued development will continue to reduce errors; however, a fundamental limit to performance of optical gyroscopes associated with photon statistics results in irreducible random-walk errors. This limit does not exist for mechanical gyroscopes, whose errors are caused mainly by the electronic processing.

a. The ring-laser gyroscope. A block diagram of a ring-laser gyroscope is shown in Fig. 12. As the ring laser rotates, the Sagnac effect results in an effective pathlength difference, given by Eq. (54), for waves that propagate in opposite directions. The pathlength changes can be quite small. For the present application, which requires sensing a rotation rate as small as 10^{-4} deg/hr to determine the spin axis to within 1 mdeg, Eq. (54) implies a pathlength difference of about 10^{-19} m for a typical RLG with optical loop area $A = 0.01$ m^2. This indicates the very extreme gyroscope sensitivity required and indicates a need to integrate the gyroscope output for a period of time to accomplish initialization.

The wavelength and frequency of the clockwise and counterclockwise waves in RLGs adjust to satisfy the laser resonance condition that requires an integer number (m) of optical wavelengths λ within the optical cavity. Thus, $\lambda m = L$, or using $c = \lambda\nu$, one obtains the frequency difference of the two laser modes with path length difference δL as

$$\nu_+ - \nu_- = \delta\nu = \frac{\nu}{L}\delta L = \left(\frac{4A}{\lambda L}\right)\Omega \tag{56}$$

The frequency difference is detected as interference fringes as portions of the counterpropagating waves combine on an optical detector. One fringe passes a point on the detector during a time interval of $(\delta\nu)^{-1}$. The output from the RLG is the fringe count and is traditionally represented as the angular motion expressed as arcseconds. A ring-laser gyroscope scale factor ($\lambda L/4A$) of about 1.5 arcsec/count is typical.

The fundamental limit to performance of any RLG is associated with statistical fluctuations of spontaneous emissions from the laser. This error mechanism results in "angle random walk," a quantity characterized by the random-walk coefficient. At the present time, RLG performance is nearly at the quantum limit imposed by the photon statistics.

The key error mechanisms that can be suppressed or compensated are bias and scale factor nonlinearity. Bias is the measured rate when gyroscope rotation rate in inertial space is zero. Scale factor, as indicated above, is the

ratio of the gyroscope rotation rate and the indicated output. At low rotation rates, the counterpropagating waves couple together because of light scattered from each wave to produce a dead zone over which the effective output is zero. This phenomenon is known as "lock-in" and represents the major source of scale factor nonlinearity. Various ways to reduce the scattered light and to compensate for lock-in have been developed.

Ring-laser gyroscopes are classified as either two-wave or four-wave devices. The light waves are linearly polarized in two-wave RLGs and circularly polarized in the four-wave types. Various types of two-wave devices have been developed to compensate for lock-in. The most common techniques to overcome lock-in of two-wave RLGs are body dithering and optical-pathlength dithering. In each, an oscillating bias is applied to produce a sensed rate that is larger than the lock-in for most of the dither period. Dither noise, which used to dominate RLG performance, has been reduced to very low levels. Dithered RLGs are produced by Honeywell, Rockwell International, and Kearfott, among others.

The four-wave RLG, which is produced by Litton and marketed as a ZLG (zero lock-in laser gyroscope)[17], uses an optical method to avoid lock-in. The technique that is used to bias the four-wave gyroscope provides for common-mode rejection of the drift and noise caused by the dc optical bias element.

The RLG output is the count of interference fringes that cross the detector. If the readout is not interpolated to indicate a small fraction of a fringe, a "quantization error" results. The dithered instruments generally use a technique that fixes the quantization error at 1/4 fringe. The resolution of the Litton ZLG is about 10^{-3} fringes.

Typical values of the best performance of present day ring-laser gyroscopes are a random-walk coefficient of $2 \times 10^{-4}\,\mathrm{deg}/\sqrt{\mathrm{hr}}$ and bias instability of 3×10^{-4} deg/hr. The projected performance is very near the fundamental limit with a random walk of $5 \times 10^{-5}\,\mathrm{deg}/\sqrt{\mathrm{hr}}$.

If the bias instability can be reduced to less than 10^{-4} deg/hr, the tracking requirements for a 10-hr period can be met by an RLG. The initialization procedure entails holding the antenna at a fixed position for a period of time. If the jitter of the stationary antenna is zero-mean, the projected random walk would permit the required rate accuracy of 10^{-4} deg/hr with an integration time of 15 minutes. During this period, approximately one-tenth of a fringe will cross the detector.

b. The fiber-optic gyroscope. Fiber-optic gyroscopes are classified as passive devices because the light source is not an integral part of the optical path. Unlike the ring-laser gyroscope, the fiber-optic gyroscope is not subject to the lock-in phenomenon. There are two types of fiber-optic gyroscopes, the interferometric fiber-optic gyroscope (IFOG) [18] and the resonant fiber-optic gyroscope (RFOG). The RFOG is presently under development at CSDL and will not be discussed here. The IFOG is under development at a number of universities and industrial houses. The IFOG discussed below is being developed at JPL as the fiber-optic rotation sensor (FORS).

The motivation for developing fiber-optic gyroscopes is that the all solid-state construction has potential for highly reliable devices characterized by low weight, low power and long life. Once developed, fiber-optic gyroscopes are expected to have a distinct cost advantage over other gyroscopes. The FORS, which is the present baseline inertial reference unit for the CRAF/Cassini Mariner Mark II (MMII) space mission, consists of an integrated optic chip to which the fiber coil, edge-emitting laser diode source, and detector are connected. Figure 13 shows a diagram of the JPL FORS.

In contrast to the ring-laser gyroscope, which measures the Sagnac effect due to a single passage around a closed optical path, the interferometric fiber-optic gyroscope measures the Sagnac effect in a fiber coil having many turns. Because the source is not part of the loop, the frequency of both waves remains constant and the Sagnac effect is measured as a phase difference of the two counterpropagating waves. The phase difference is related to the pathlength as shown in Eq. (55). For an N-turn coil, Eq. (55) may be written as

$$\Delta\phi = 2\pi \left(\frac{4NA}{\lambda c}\right) \Omega \tag{57}$$

where A is the area enclosed by a single loop of fiber.

IFOGs may be operated open-loop or closed-loop by compensating the Sagnac phaseshift through the introduction of a nonreciprocal phaseshift of opposite polarity within the fiber loop. A common method to accomplish the phase-nulling effect is to apply the shift at one end of the loop using a technique known as serrodyne phase modulation. In FORS, the effect of phase nulling is read by sampling a part of the counterpropagating beams using a method called optical beat detection. This results in a scale factor differing from that of the RLG, Eq. (56), by only a factor of n, the refractive index of the fiber.

C2

Detector shot noise provides the fundamental limit to the IFOG performance and results in a random-walk error characterized by a random-walk coefficient, as for the RLG. Shot noise is reduced by increasing the optical power on the detector; however, there is an upper limit to the laser power because of errors due to nonlinear optical effects in the fiber.

Rayleigh backscattering results in short-term random noise that is corrected by the use of broadband optical sources, such as the superluminescent diodes or edge-emitting diodes as used in FORS. Error sources associated with long-term bias stability have been identified and are presently undergoing intense study.

The demonstrated random-walk coefficient (RWC) of FORS is close to the quantum limit for the configuration used. With $\lambda = 1.3$ μm, $L = 1$ km, and $D = 0.1$ m, the measured RWC is 7×10^{-4} deg/root hr. The bias instability goal for the MMII application is 10^{-3} deg/hr.

Additional development is required for antenna-pointing applications. The probable approach would be to increase the fiber length and the LD product to provide a RWC comparable to that of the projected value of the RLG. To be useful for the present application, the bias instability must be decreased by at least an order of magnitude.

C. Accelerometers

1. Overview. Accelerometers are needed to determine the local vertical during the initialization step, and are also used to determine antenna elevation. Accelerometers may also be used to measure the tilt of the elevation axis; however, an inclinometer can perform that function more cheaply and accurately.

The heart of an accelerometer is a proof mass that is constrained to move along a single sensitive axis. A variety of ingenious methods have been used to sense the force on the proof mass and to provide an output signal that is proportional to the input acceleration. Nearly all accelerometers operate in a closed-loop configuration, in which movement of the proof mass is sensed and a signal is generated to restore the proof mass to its null position. An exception is the vibrating beam accelerometer, which operates open-loop and senses the acceleration as the change in frequency of force-sensitive vibrating quartz beams.

Accelerometers are commercially available with a wide range of performance and cost. Reported resolution values vary from a fraction of a μg for high-grade navigational instruments to over a hundred μg for more commonplace applications such as construction work. Some of the accelerometers with capabilities sufficient for the proposed antenna pointing application are discussed below.

2. The gyroscope-based accelerometer. The pendulous integrating gyroscope accelerometer (PIGA) has a pendulous proof mass attached to the spin axis of a single-degree-of-freedom mechanical gyroscope. Acceleration along the sensitive axis results in a force on the proof mass and produces a torque on the gyroscope output axis. The gyroscope is mounted on a member that is rotated in response to the gyroscope output signal to produce a gyroscopic torque that balances the torque caused by the input acceleration. The PIGA output signal is proportional to the time integral of the input acceleration.

The PIGA is probably the highest performance accelerometer available. It has undergone continuous development since World War II by CSDL and the MIT Instrumentation Laboratory. It is presently used in high-performance strategic missiles and can serve as a high-resolution gravimeter. The performance of existing PIGAs includes a resolution of better than 10^{-6} g with prospects of a resolution better than 10^{-8} g projected for future instruments. This performance significantly exceeds the proposed DSN antenna application; however, because of its cost, the PIGA is not considered a candidate for the present application.

3. Force-rebalance accelerometers. The general class of force-rebalance accelerometers contains a proof mass supported by a flexure and constrained to move along a single axis. The position of the proof mass is detected and the mass is restored to its null position by a rebalance force generated by a control loop. The current in the control loop, which is proportional to the input acceleration, provides the output signal. Performance characteristics include the instrument resolution as well as long-term and short-term errors in the bias and gain parameters. As discussed above, the values of the gain and bias parameters may be estimated prior to and during normal operation by using the minimum-variance estimators and extended Kalman filtering algorithms. When this is done, the uncompensated random variations of the bias and gain determine the accuracy of the instrument.

Manufacturers of force-rebalance accelerometers include Bell Aerospace, Textron, Incosym, Kearfott, Litton, Northrop, Rockwell International, Sundstrand, and Schaevitz. Many instruments in this category, which feature a resolution that varies from 1 μg to about 10 μg, meet the requirements for the present pointing application.

4. Vibrating-beam accelerometers. The basic sensing element of the vibrating-beam accelerometer (VBA) is a force-sensitive, vibrating quartz crystal beam that changes its resonant frequency in response to axial tension and compression. Forces are derived from accelerations applied to a pendulously supported proof mass. The proof mass is restrained along the sensitive axis by two quartz beams so that one beam is placed in tension and the other in compression as a result of acceleration. The VBA output signal is proportional to the frequency difference of the two beams. The "push-pull" arrangement of the quartz beams results in common-mode rejection of most of the error sources. A schematic of the VBA [19], which operates open-loop, is shown in Fig. 14.

Acceleration on the VBA's flexure-mounted proof mass places one of the vibrating quartz beams in tension and another in compression. The VBA output is the difference in frequency of the two beams. Thus, the crystal beams and oscillator circuits replace the torquer coils, magnets, and capture electronics of the conventional force-rebalance accelerometer.

The VBA has been under development at the Kearfott Division of the Singer Company (now the Astronautics Corp. of America). Short-term bias stabilities of 1 μg and scale factor stabilities of 1 part per million (ppm) have been demonstrated with the VBA. This performance is well within the requirements of the present application; however, the instrument is not yet available commercially.

D. Inclinometers

The purpose of an inclinometer in the present application is to measure the level of the elevation axis. Common inclinometers make use of force-rebalance accelerometers or the level assumed by a liquid in a suitable enclosure. One simple concept for an inclinometer is based on the carpenter's bubble level. Spectron Glass and Electronics markets a precision version of the bubble concept as an "Electrolytic Tilt Sensor." These units are one-piece glass enclosures partially filled with an electrolyte and constructed with platinum terminals and contacts that form two arms of an electrical circuit. Part of each arm includes the electrolyte as a resistive element. As the sensor is tilted and electrolyte flows from one side to another, the resistance of one arm increases while the other decreases. This change is sensed as a voltage change. The most precise model covers the range of $\pm 1/2$ deg with resolution of less than 1 mdeg. A disadvantage for the present application is that the high precision of the sensor is only available when the inclinometer is mounted precisely "on top" of the elevation axis. As the antenna is elevated, the sensor rolls away from its preferred position, and the accuracy decreases. It would be necessary to use the inclinometer on a gimbal to achieve maximum resolution.

Inclinometers that use dielectric liquids as an integral part of the device are also offered by Schaevitz. A change in angle is measured as a change in capacitance between two plates as the liquid flows to a new position between the plates. The resolution is about 1 mdeg and null repeatability is 5 mdeg. The total angular range is ± 60 deg.

Inclinometers that use closed-loop force-rebalance linear accelerometers are available from a number of vendors. For example, inclinometers with a resolution of 0.1 arcsec and a total range of ± 90 deg are offered by Schaevitz and Sundstrand. The output offset at zero tilt can reach 50 mdeg for the 90 deg range and 2 mdeg when the range is decreased to ± 1 deg.

E. Recommendations

Currently, the best mechanical gyroscopes outperform the best optical gyroscopes. The single-degree-of-freedom floated gyroscope and the electrostatically suspended gyroscope are both mechanical devices that appear capable of satisfying the requirements for inertial pointing. However, optical gyroscopes enjoy continued improvement through intense development, and are expected to surpass the mechanical gyroscopes in performance in the near future. In contrast, the mechanical devices represent a mature technology, and dramatic improvements in performance are unlikely to occur. In anticipation of these future developments, ring-laser gyroscopes have been chosen in the baseline for the inertial instrument. Optical gyroscopes potentially have additional advantages over mechanical gyroscopes, including lower initial cost, greater reliability, and immunity to vibration and gravity loading.

A number of force-rebalance accelerometers appear to be capable of inertial pointing. These devices are relatively inexpensive, off-the-shelf items. Stability of the accelerometer parameters remains a concern since the extended Kalman filter is limited in its ability to estimate rapidly changing instrument parameters. Another concern is the level of hysteresis that might be encountered during the course of antenna tracking. Vibration levels on the antenna and electronic noise also have to be addressed. Even if some unforeseen difficulties rule out the use of available force-rebalance accelerometers, the emerging vibrating-beam accelerometer, when it becomes available, should be a viable option.

Small-angle inclinometers of various types are available, so no specific recommendation is made. There is a problem that the inclinometers lose accuracy when rotated about

their insensitive axes. This would occur, for example, when the antenna elevation was changed. A solution to this problem has been developed which involves mounting the inclinometer on a simple single-axis pendulum-stabilized platform.

V. Simulation Results

Simulations were conducted to illustrate the behavior of the algorithms presented in the article and to predict the performance of inertial instruments, given a set of representative instrument parameters. The simulations were written in Pro-Matlab by The Mathworks, a well-known commercial software product that uses matrices as its fundamental data type, and were run on a Sun 3/60 desktop workstation.

The gyroscope parameters used in the simulations were previously given in Table 1, and the accelerometer parameters are summarized in Table 2. These instrument parameters are fictitious in the sense that they do not correspond to any specific devices, but instead reflect "typical" parameters for current or projected near-term state-of-the-art devices, based on conversations with and data sheets provided by the manufacturers. The purpose of the simulations is to illustrate the feasibility of an inertial pointing system, not to evaluate specific instruments.

The first two simulations, shown in Figs. 15 and 16, illustrate the behavior of the extended Kalman filter which is used to estimate the accelerometer parameters. For this example, the antenna is rotated in elevation from horizontal to vertical. The initial bias and gain errors were initialized with a random-number generator which assigned them values with standard deviations of 100 μg and 100 ppm rms. The two figures show that even with this level of initial error, the bias and gain parameters for this example can be estimated with an accuracy much smaller than 1 μg following the antenna rotation.

Figure 17 illustrates the performance of the elevation-estimation algorithm that can be expected when the antenna is tracking a target. It is assumed that the antenna is located at the same latitude as the antenna complex at Goldstone, and that the tracked object is inertially stationary and elevated 45 deg from the Earth's equator. The tracking period is set at 10 hr. At the beginning of the tracking period, the accelerometer bias and gain errors are set to zero, reflecting the fact that these parameters are initially known from the previous calibration using the extended Kalman filter. However, the accelerometer model that was chosen allows random walk in the bias and gain parameters. The extended Kalman filter successfully estimates these parameters during target tracking, resulting in elevation error smaller than about 2 μrad (0.1 mdeg). The results are impressive but it must be recognized that several factors which would degrade the accuracy have not been accounted for, including seismic vibrations, accelerometer misalignments, nonlinearities, and electronic noise.

The final two simulations shown in Figs. 18 and 19 illustrate the tracking performance of the entire inertial instrument. Again the antenna is assumed to be located at Goldstone, and the target is inertially stationary and elevated 45 deg from the Earth's equator. Fig. 18 shows the tracking performance in terms of boresight error when only the gyroscopes are used to integrate attitude. The nearly linear growth of error with time indicates that the predominant source of error in this case is the rate-bias error, which is typical of gyroscopes. In Fig. 19, the accelerometer and inclinometer measurements were incorporated into the attitude determination using a minimum-variance estimator to combine the outputs with those of the gyroscopes and the extended Kalman filter to determine antenna elevation. The results demonstrate reduced antenna boresight error, particularly during the middle of the 10-hr tracking period when the antenna is closest to vertical, and thus the elevation determination is nearly redundant with the gyroscope output.

VI. Summary

A system using inertial instruments for pointing of the Deep Space Network antennas has been described. The proposed configuration includes a three-axis gyroscope, four accelerometers, and an inclinometer in a strapdown system. The system can be used both to initialize the antenna pointing attitude and to track an object for prolonged periods of time.

The concept for initializing the gyroscopes requires determination of the local gravity vector and the Earth's spin axis in the antenna local coordinate frame. The gravity vector is determined using the accelerometers to measure elevation angle and the inclinometer to measure cross-axis elevation. An advanced algorithm employing both a minimum-variance estimator and an extended Kalman filter is used to combine the measurements to determine elevation angle, and also to estimate instrument parameters, including biases and gains, while the instrument is in operation. The Earth's spin axis is determined using the gyroscopes, with the antenna held stationary.

Once the antenna attitude is known, it can be slewed to a desired target and held there by using the gyroscopes to

integrate angular rate to obtain attitude and using knowledge of the local vertical to augment this estimate, again combining the measurements in a minimum-variance estimator.

The article continues with a survey of available state-of-the art technology, and makes recommendations towards component selection. In particular, the ring-laser gyroscope appears attractive as the candidate gyroscope, and relatively inexpensive accelerometers appear to fulfill the elevation-determination requirements. The article concludes with a sampling of the simulation results illustrating the functions and performance of an idealized instrument on a DSN antenna. The results show pointing performance close to 1 mdeg, indicating that this level of performance may be possible, though not necessarily easy to achieve, with an inertial instrument.

References

[1] R. E. Scheid, "Precision Pointing Compensation for DSN Antennas with Optical Distance Measuring Sensors," *TDA Progress Report 42-97*, vol. January–March 1989, Jet Propulsion Laboratory, Pasadena, California, pp. 127–140, May 15, 1989.

[2] J. R. Wertz, editor, *Spacecraft Attitude Determination and Control*, Boston: D. Reidel Publishing Company, 1984.

[3] A. E. Bryson and Y. C. Ho, *Applied Optimal Control*, Washington, D.C.: Hemisphere Publishing Company, 1975.

[4] H. Goldstein, *Classical Mechanics*, Dallas: Addison-Wesley Publishing Company, 1950.

[5] R. R. Ragan, editor, "Inertial Technology for the Future," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-20, no. 4, pp. 414–440, July 1984.

[6] E. J. Loper and D. D. Lynch, "Projected System Performance Based on Recent HRG Test Results," *Proceedings of the IEEE/AIAA 5th Digital Avionics Systems Conference*, Seattle, Washington, October 31–November 3, 1983.

[7] P. H. Savet, editor, *Gyroscopes: Theory and Design*, New York: McGraw-Hill, 1961.

[8] W. W. Chow, J. Gea-Banacloche, L. M. Pedrotti, 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.

[9] S. Ezekial and H. J. Ardity, editors, *Fiber-Optic Rotation Sensors*, Berlin, Heidelberg, and New York: Springer-Verlag, 1982.

[10] C. Kochakian, R. McKern, and J. Negro, "Performance Characterization of Low Noise Gyro," Thirteenth Biennial Guidance Test Symposium, Central Inertial Test Facility, Holloman AFB, New Mexico, August 1987.

[11] *The GI-G6G-PMM: A Permanent Magnet Rate Integrating Gyro*, Northrop Precision Products Division, Company Brochure D00017, 1989.

[12] C. O. Swanson, J. Kass, and A. Greiner, "Line-of-Sight Stabilization for the Dynamic Magellan Spacecraft," *12th Annual AAS Guidance and Control Conference*, AAS 89-045, Keystone, Colorado, February 4–8, 1989.

[13] R. B. Irvine and J. W. Ritter, "DRIRU II: The NASA Standard High Performance Inertial Reference Unit," *Annual Rocky Mountain Guidance and Control Conference*, AAS 79-021, Keystone, Colorado, February 24–28, 1979.

[14] R. R. Warzynski and R. L. Ringo, "The Evolution of ESG Technology," *Proceedings of the AGARD Conference on Inertial Navigation Components and Systems*, Paper 13, October 2–5, 1972.

[15] G. Joos, *Theoretical Physics*, 2nd edition, New York: Hafner Publishing Co., 1950.

[16] R. Bartman, B. Youmans, and N. Nerheim, "Integrated Optics Implementation of a Fiber Optic Rotation Sensor: Analysis and Development," paper 719-19, *SPIE Proceedings*, vol. 719, Cambridge, Massachusetts, September 21–26, 1987.

[17] M. Fernandez, R. Ebner, and N. Dahlen, "Zero-Lock Laser Gyro," *12th Annual AAS Guidance and Control Conference*, AAS 89-024, Keystone, Colorado, February 4–8, 1989.

[18] V. Vali and R. W. Shorthill, "Fiber Laser Gyroscopes," *SPIE*, vol. 77, pp. 110–115, 1977.

[19] W. C. Albert, "Vibrating Quartz Crystal Beam Accelerometer," *ISA 28th International Instrumentation Symposium*, vol. 28, no. 1, pp. 33–44, 1982.

Table 1. Parameters used to simulate gyroscope behavior

Quantity	Value	Comment
Angle random walk	$(0.1)^2$ mdeg^2/hr	Gaussian distributed error
Drift uncertainty	0.1 mdeg/hr 0-p	Systematic error may be reduced via parameter estimation
Scale factor uncertainty	5 ppm	Systematic error may be reduced via parameter estimation, but error is small except during slew
Resolution/noise	0.1 mdeg 0-p	Measurement error effect is reduced by averaging data over time

Table 2. Accelerometer parameters used for simulating elevation determination

Quantity	Value
Accelerometer noise	$(1.5\ \mu g)^2$
Angle process noise	$(20$ mdeg$)^2$/sec
Bias process noise	$(100\ \mu g)^2$/10 hr
Gain process noise	$(100\ \mu g)^2$/10 hr
Time step used for simulation	10 hr/450

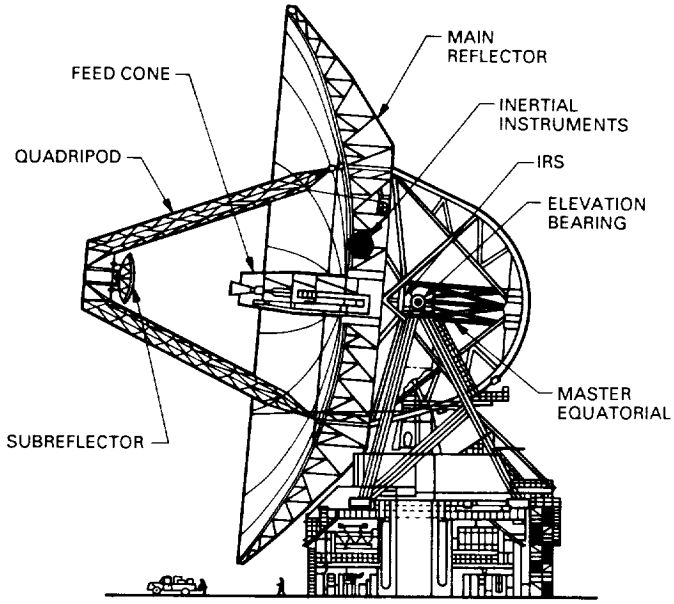


Fig. 1. A 70-m DSN antenna showing potential location of inertial instrument package.

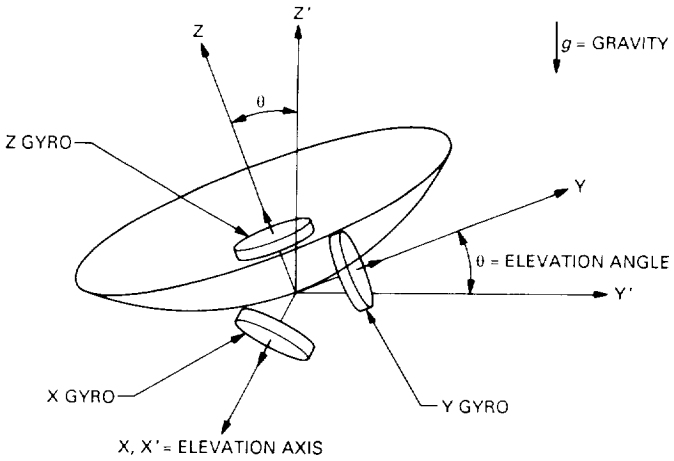


Fig. 2. Orientation of gyroscopes on antenna.

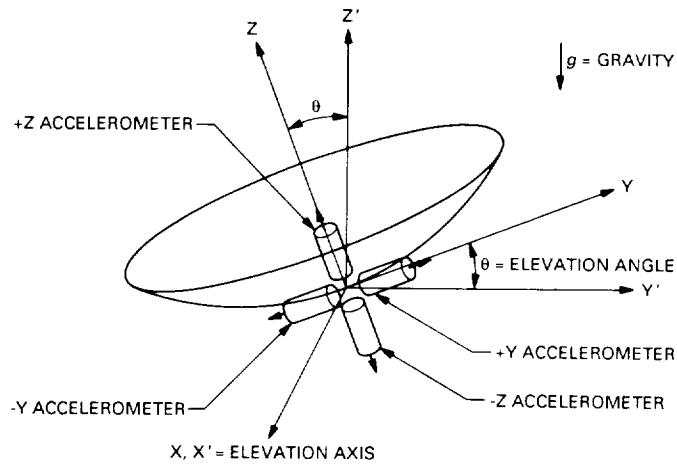


Fig. 3. Orientation of accelerometers on antenna.

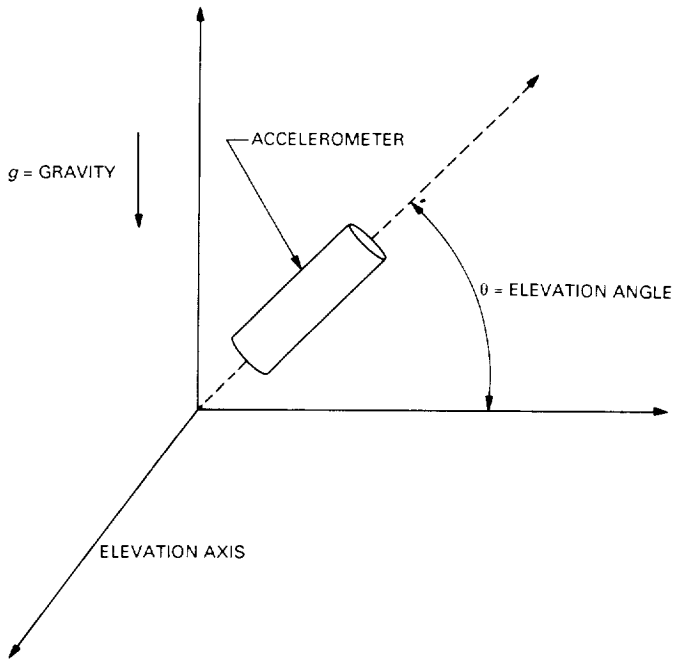


Fig. 4. Single accelerometer under the influence of gravity. Output depends on the angle between the accelerometer sensitive axis and the gravity field.

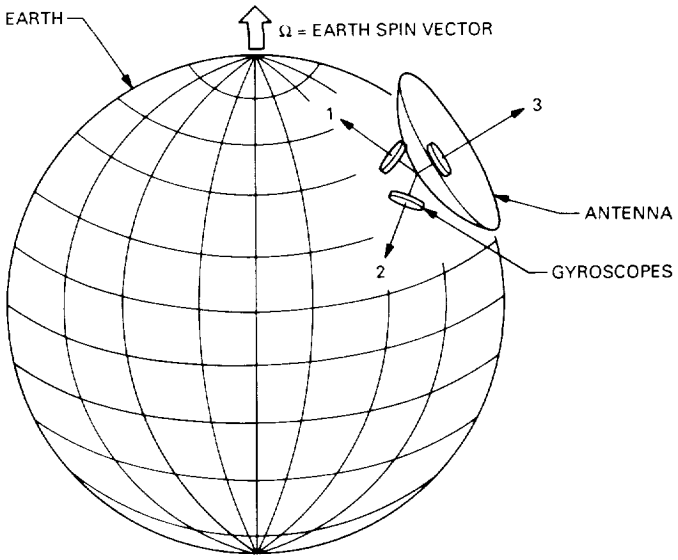


Fig. 5. Orientation of antenna and gyroscopes with respect to the Earth and its spin vector.

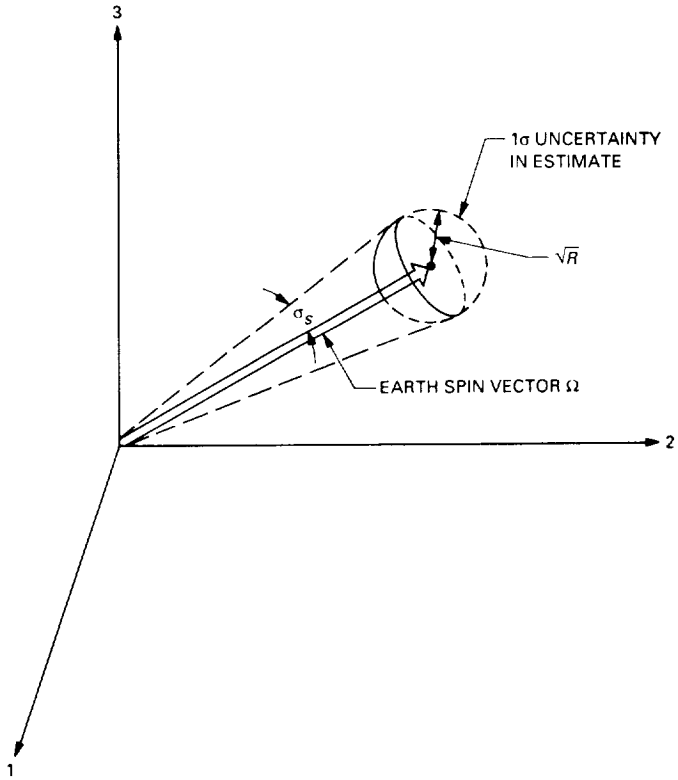


Fig. 6. Uncertainty in the estimate of the Earth's spin vector determined from a minimum-variance estimator.

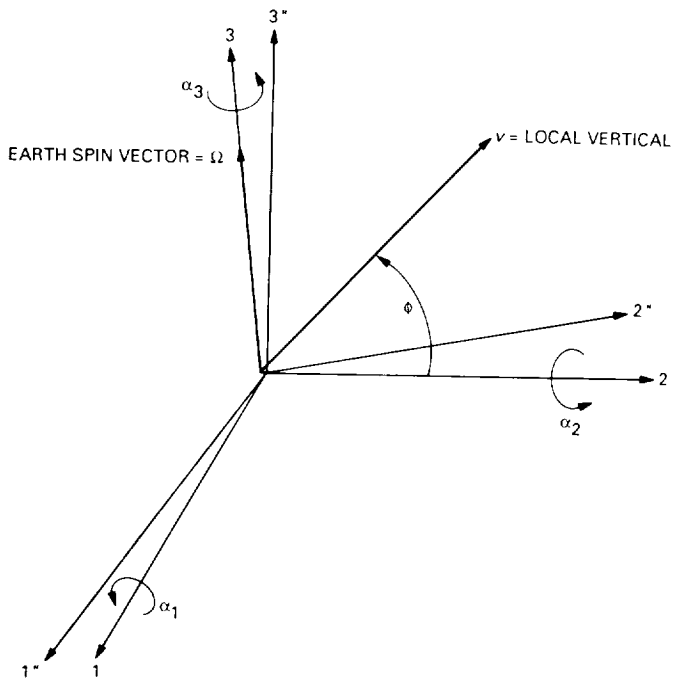


Fig. 7. Coordinate system for definition of small angle corrections. The unprimed frame is attached to Earth, and the double-primed frame is attached to the antenna.

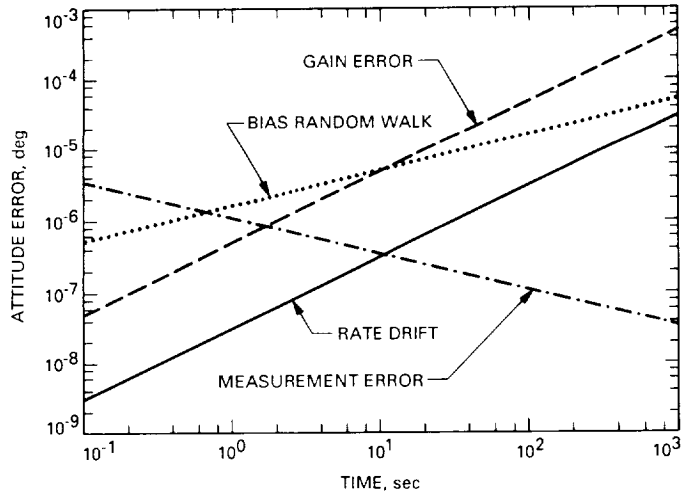


Fig. 8. Summary of error growth terms resulting from integration of a gyroscope.

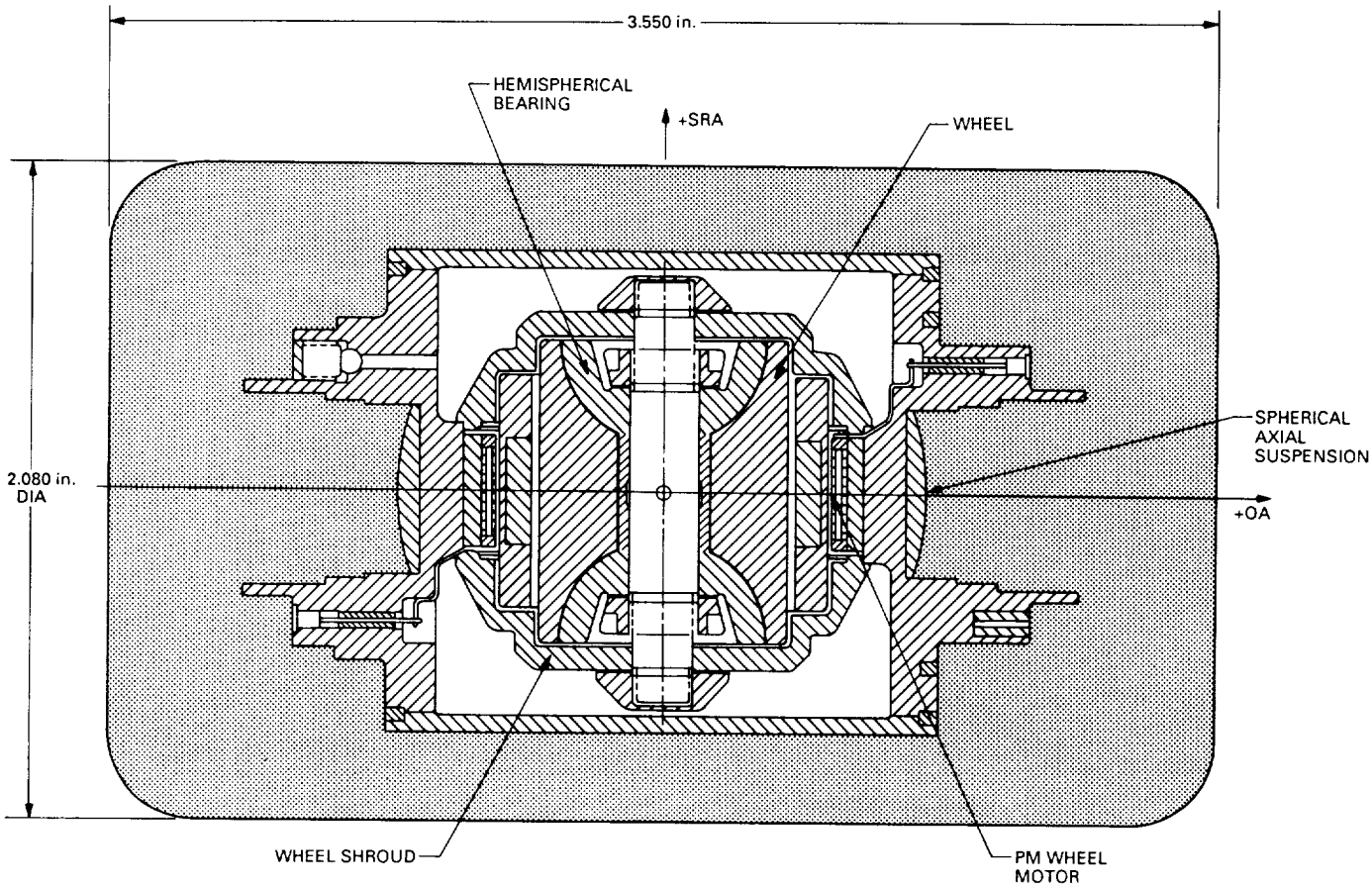


Fig. 9. Simplified drawing of a CSDL fourth generation single-degree-of-freedom floated (SDOFF) gyroscope.

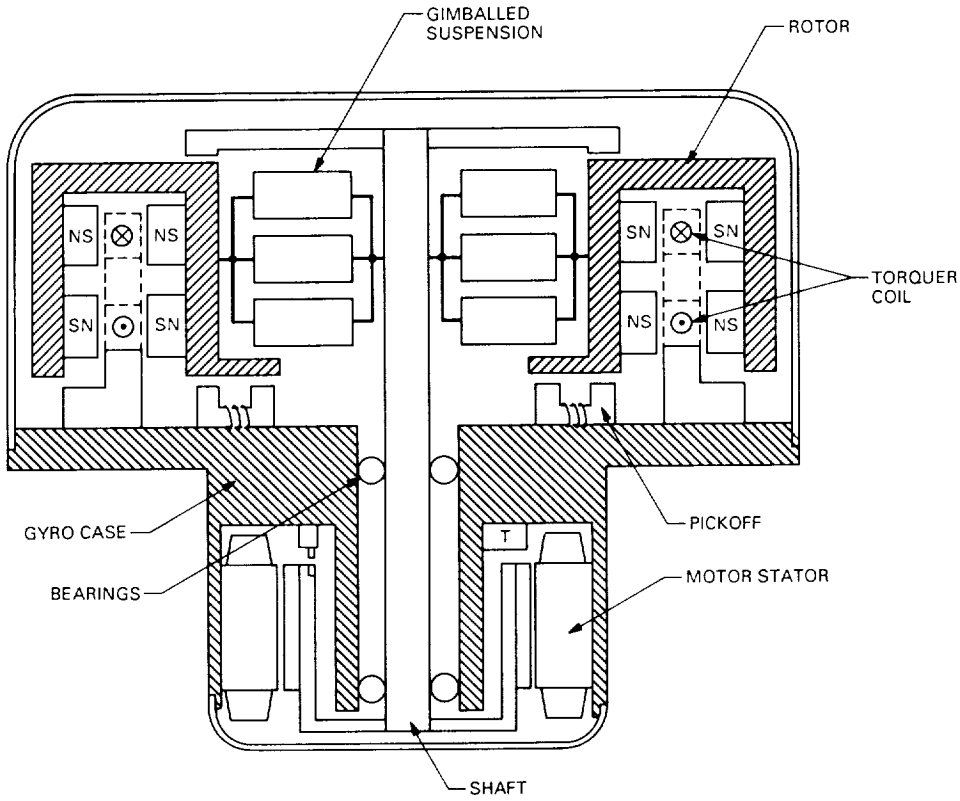


Fig. 10. Cross section of Teledyne's dynamically tuned gyroscope that is used in the NASA-developed dry rotor inertial reference unit (DRIRU II).

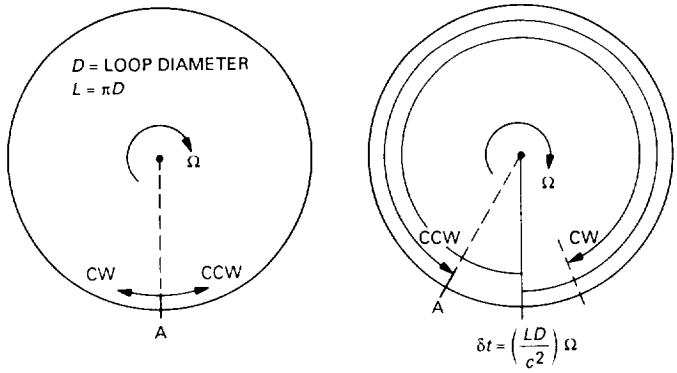


Fig. 11. The Sagnac effect. Optical waves launched in the same direction as the loop rotation require a longer time to complete a revolution than do the counterpropagating waves.

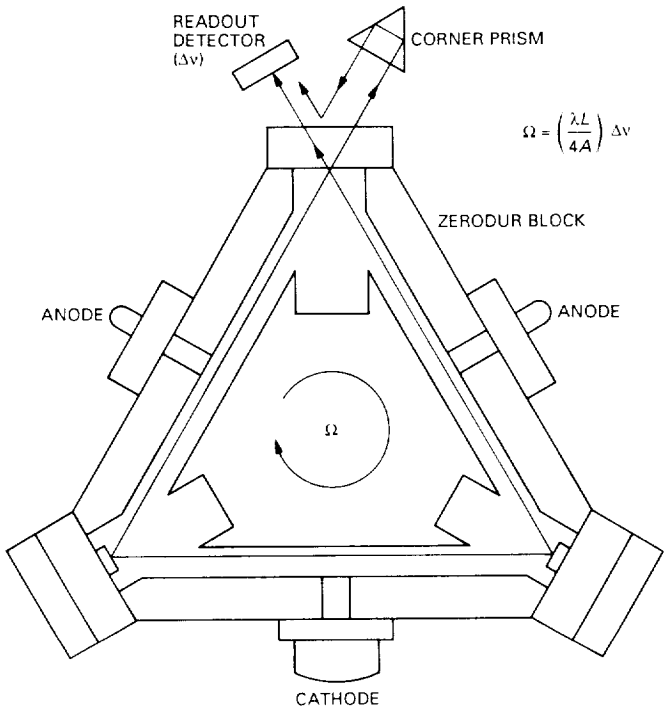


Fig. 12. A ring-laser gyroscope (RLG). Gyroscope rotation changes the optical frequency of counterpropagating waves.

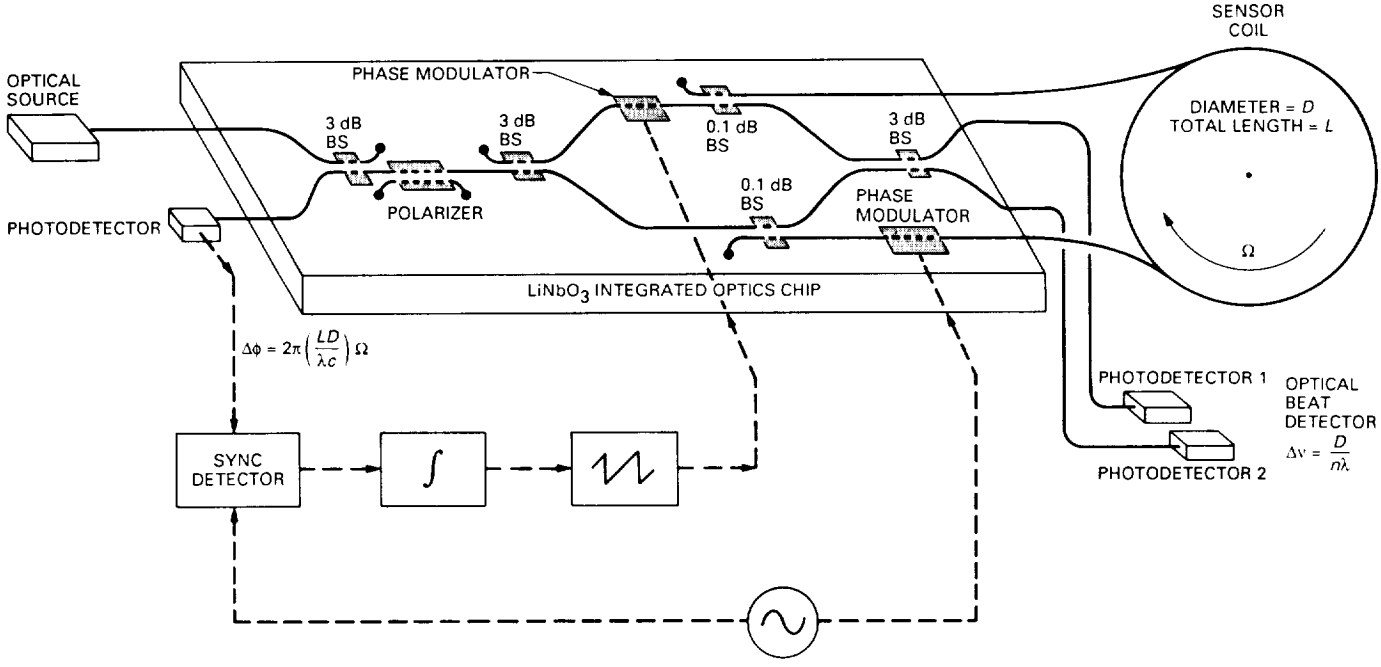


Fig. 13. The fiber-optic rotation sensor (FORS), a fiber-optic gyroscopic under development at the Jet Propulsion Laboratory.

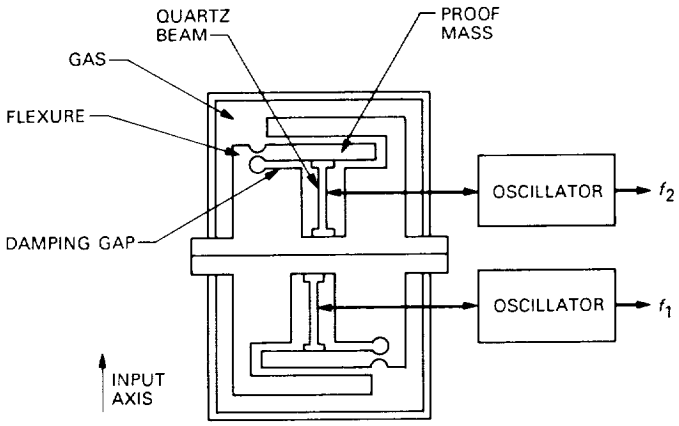


Fig. 14. A vibrating-beam accelerometer. Acceleration of the proof mass places one beam in tension and the other in compression. Acceleration is sensed as a change in the resonant vibrational frequency of the quartz beams.

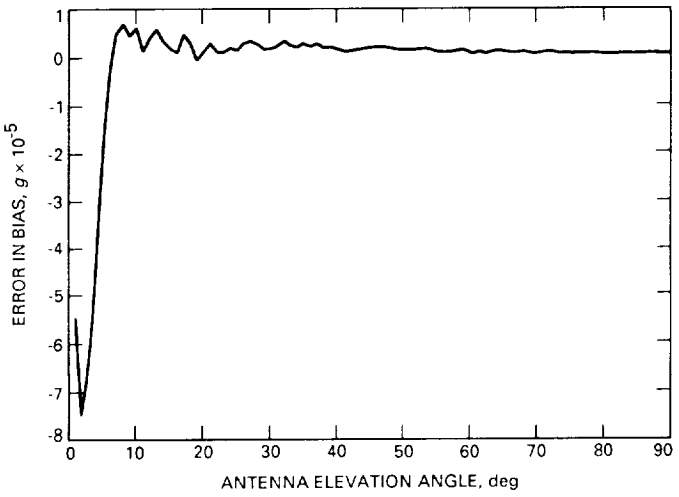


Fig. 15. Convergence of accelerometer bias estimate using an extended Kalman filter to estimate parameters. The antenna is elevated by 90 deg to provide maximum ease of identification.

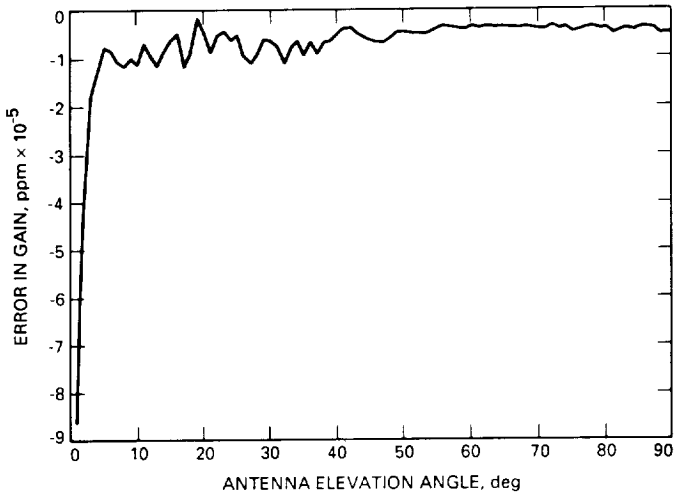


Fig. 16. Convergence of accelerometer gain estimate using an extended Kalman filter.

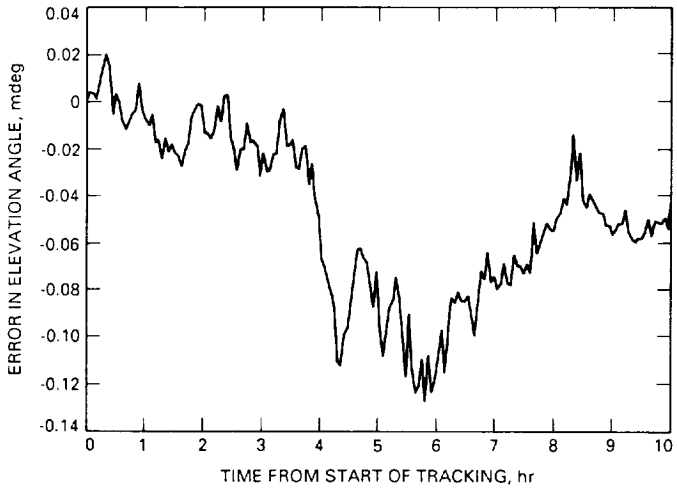


Fig. 17. Performance of a combined minimum-variance estimator and extended Kalman filter for estimating elevation angle during a 10-hr track of an inertially stationary object elevated 45 deg from Earth's equator.

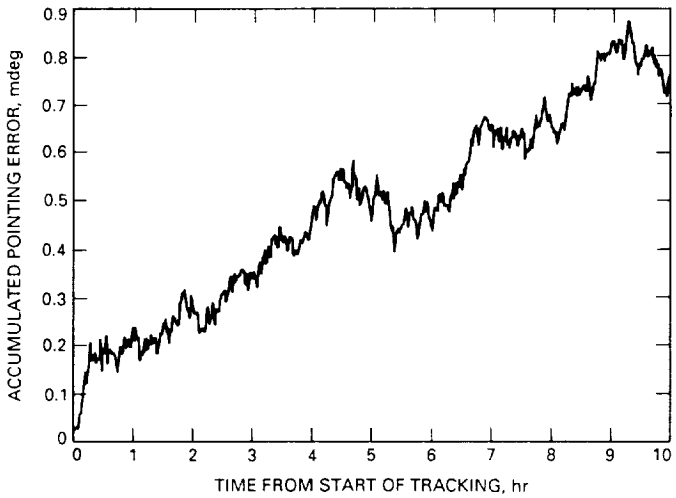


Fig. 18. Propagation of attitude errors obtained from three-axis gyroscope integration. Initial attitude error is set to zero.

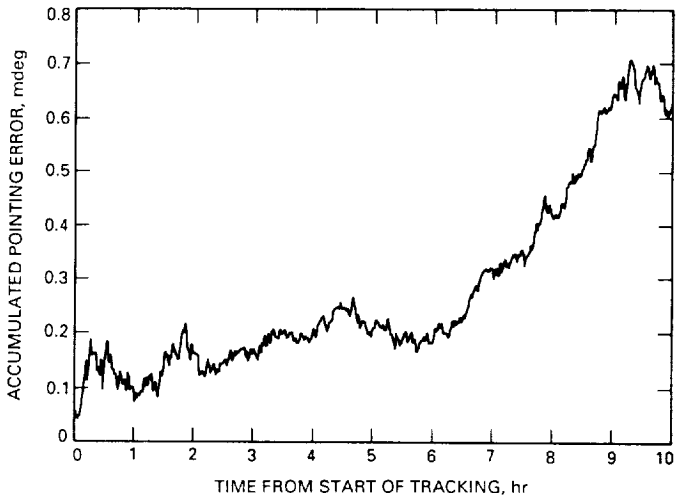


Fig. 19. Performance of attitude estimator that combines gyroscope integration with knowledge of local vertical. The additional information improves attitude accuracy.

Design and Performance Analysis of the DSS-13 Beam Waveguide Antenna

T. Veruttipong, W. Imbriale, and D. Bathker
Ground Antenna and Facilities Engineering Section

A new 34-m research and development antenna is currently being constructed prior to introducing beam waveguide (BWG) antennas and Ka-band (32 GHz) frequencies into the NASA/JPL Deep Space Network. The new 34-m antenna, fed with either a center or bypass BWG, will lose less than 0.2 dB (excluding surface root mean square and mirror misalignment losses), as compared with a standard-fed Cassegrain antenna at X- (8.4 GHz) and Ka-bands. The antenna is currently under construction and is scheduled to be completed in July 1990. Phase 1 of the project is for independent X- and Ka-band receive-only tests. Phase 2 of the project is for simultaneous S- (2.3 GHz) and X-band or X- and Ka-band operation, and the design is currently under way.

I. Introduction

Feeding a large low-noise, ground-based antenna via a beam waveguide (BWG) system has several advantages over placing the feed directly at the focal point of a dual-reflector antenna. For example, significant simplifications are possible in the design of high-power, water-cooled transmitters and low-noise cryogenic amplifiers, since these systems do not have to rotate, as in a normally fed dual reflector. Furthermore, these systems and other components can be placed in a more accessible location, which leads to improved service and availability. Also, the losses and noise degradation associated with rain on the feedhorn radome are eliminated because the feedhorn can be sheltered from the weather.

The design of the new 34-m BWG antenna at DSS 13 is based upon geometrical optics (GO) criteria introduced by Mizusawa and Kitsuregawa in 1973 [1], which guarantee a perfect image from a reflector pair. Since it may be desirable to retrofit existing antennas with a BWG, as well as construct new antennas, there are two independent BWG designs built into the research and development antenna. The first, termed a bypass design, places the BWG outside one of the elevation bearings on the rotating azimuth platform, thereby retaining the existing elevation wheel and counterweight subassembly, suitable for retrofit applications. The second, a center design, places the BWG through the center of the main reflector, inside the elevation bearings, and through the azimuth axis into a pedestal room located below the antenna. The centerline design is preferred, given new construction. The bypass design uses a pair of paraboloidal sections and two flat mirrors, whereas the center design uses the same four-mirror design (although not physically the same four mirrors) above the azimuth bearing with a flat plate and an ellipsoidal section

that functions as a beam magnifier in the pedestal room. A beam magnifier is required since the pair of paraboloids requires a 29-dBi gain horn to feed, whereas at the lower frequencies a 29-dBi gain horn would be too large to fit in the pedestal room. The ellipsoid design allows the use of smaller 22-dBi gain horns in the pedestal room.

Although the upper four reflectors in either the bypass or centerline BWG satisfy the Mizusawa criteria, the single (curved) ellipsoidal mirror in the pedestal does not. Hence, use of the overall six-mirror, three of which are curved, centerline system introduces a small beam distortion (imperfect imaging). It would require a second ellipsoidal mirror in the pedestal to obtain perfect imaging (in the GO limit). The second ellipsoidal mirror, to be fully compensating for the first, would defeat the (here) necessary beam magnifier function, and is therefore not used.

The microwave antenna gain performance was analyzed using an appropriate combination of physical optics (PO)/spherical wave expansion (SWE) and geometrical theory of diffraction (GTD) software. The initial operation (Phase 1) of the DSS-13 project is for independent X- and Ka-band receive modes, and performance predictions for these frequencies will be given below.

A. Dual-Shaped Reflector Design

The DSN presently operates three 34-m high-efficiency (HEF) dual-shaped reflector antennas with a dual band feed (2.3/8.4-GHz), which has a far-field gain of 22.4 dBi that is conventionally located at the Cassegrain focal point (see Fig. 1). The structures were designed prior to BWG requirements and feature a continuous elevation axle and a carefully designed elevation-wheel substructure. The elevation wheel is supported by an alidade that rotates on a circular azimuth track. To minimize the cost of developing a new 34-m BWG antenna, as much as possible of the existing structure design was to be used (see Fig. 2). Through the use of a clever mechanical design, the elevation-tipping structure was modified to accommodate a central BWG inside the elevation bearings. To provide clear access for an 8-ft (2.44-m) diameter, center-fed BWG, the main reflector backup trusses are connected to a revised elevation wheel via the integral ring girder, or IRG. The IRG is a toroidal structure, an octagonal space truss with a square cross section, approximately 290 in. (~7.4 m) in maximum radius and 80 in. (~2.0 m) high. It is interwoven with, but separate from, the conventional rib-and-ring backup structure. In order to minimize the distortion of the main reflector surface under gravity loading, the reflector connections to the elevation-wheel structure were selected to provide equal stiffness supports. This is achieved by grouping eight equally spaced reflector radial ribs into four pairs and connecting each pair to the IRG top plane at alternate vertices of the octagon. The vertices lying on the elevation axis, however, are reserved for supporting the IRG at the two elevation-bearing points. The counterweight and single elevation bullgear lie on a plane orthogonal to the elevation axis. The entire tipping structure (including the main reflector, elevation wheel, subreflector, and its support) is weight-balanced about the elevation axis.

Selection of the previously designed HEF-reflector structure fixes the focal length/diameter (f/D) of the main reflector surface. The reflector shape is free to be different from the HEF design, but had to be within an adjustable tolerance (~1 in.) of the existing surface to allow use of the existing design drawings.

B. Feed Selection

GO was used to design the upper portion of the centerline BWG system (mirrors M1 to M4). As shown in Fig. 3, the first mirror, M1, has azimuth and elevation rotations together with the main reflector and subreflector. A plane surface is used for M1 to ensure an imaged feed pattern that is independent of the elevation angle of the antenna. Mirrors M2 and M3 are sections of paraboloids, and the system is designed so that a feed placed at F_2 (in the GO limit) is perfectly imaged at F_1.

An imaged feed pattern at F_1 is used to illuminate the subreflector with a narrow-angle high-gain (~29-dBi) pattern. This configuration is used because of the large distance between the subreflector and the first BWG mirror (M1), and also because the size of M1 (as well as M2) is smaller than the subreflector. The position of the focal point F_1 in Fig. 3 must be close to M1 to achieve acceptable spillover losses at the subreflector, as well as M1 and M2. Normally F_1 is in the neighborhood of the main reflector vertex with an 8–9 deg half-cone angle of illumination at the subreflector, as compared with 17 deg for the normal Cassegrain feed of the 34-m HEF antenna.

The diameter of the subreflector D_S is determined by the size of the main reflector. According to [2], a subreflector diameter not exceeding 1/10 of a main reflector is normally selected for good radiation efficiency of the antenna. For a 34-m antenna, a subreflector diameter of 3.43 m (135 in.) was chosen. The illumination angle θ_S is determined next. For the same f/D ratio as the HEF and D_S = 135 in., the distance L_1 = 593.1 in. (~15 m) is obtained. Iterations are needed for a calculation of θ_S and the location of F_1. Known parameters are D_S = 135 in.,

L_1 = 593.1 in., and D_2 = 94 in. (~2.4 m). Variable parameters are 8.0 deg $\leq \theta_S \leq$ 9.0 deg, 105 in. $\leq L_2 \leq$ 110 in. (~2.7 m), and 9.5 deg $\leq \theta_m \leq$ 11.0 deg. The angle θ_m is the illumination angle at M2 with an edge taper of about −23 dB. The results of iterations of GO ray geometry between D_S and D_2 are θ_S= 8.7 deg, θ_m = 10.4 deg, L_0= 441.11 in. (~11.2 m) and L_2 = 108.01 in. (~2.7 m). The GO focal length of M2 is selected as 260 in. (6604 mm). The exact dimensions are somewhat arbitrary, but are constrained by the M2 mirror-projected diameter (96-in. limit). The BWG shroud, or tube diameter, was chosen because the tube effects at S-band would be small. It is now necessary to design a horn that has an approximately −18 to −20-dB taper at 8.7 deg (the illumination of the subreflector) and minimal spillover past 10.4 deg (the illumination of the BWG mirror).

An important design parameter is the horn's flare angle. Figure 4 shows the patterns and efficiencies (spillover times phase efficiency) for several different horn-flare angles and, as can be seen, the patterns are not very sensitive to the flare angle. For that reason and because the JPL standard feedhorn has a flare angle of 6.25417 deg, it was decided to examine the standard flare angle, since existing feedhorns or feedhorn designs could be utilized.

Various horn sizes with the JPL standard angle of 6.25417 deg and frequency = 8.45 GHz were investigated. The goal was to find a horn with a −18-dB taper at θ = 8.7 deg (near-field distance of 425 in., ~10.8 m) and a −23-dB taper at θ = 10.4 deg (near-field distance of 260 in., ~6.6 m). The distances 260 in. and 425 in. are for a high-gain horn illuminating the BWG mirror M2 and the subreflector, respectively. The combined phase and spillover efficiencies ($\eta_{phase} \times \eta_{spill}$) should be optimally between 8.7 and 10.4 deg. The results from various trials show that a 19-in. (~ 483-mm) aperture diameter at X-band, with dimensions shown in Fig. 5, gives a radiation pattern that meets these goals. Figures 6(a) and 6(b) show amplitude, phase, and efficiency plots of the 19-in. X-band aperture diameter at r = 425 in. and 260 in., respectively. From Fig. 6(a), the edge taper at the rim of the subreflector (θ = 8.7 deg, r = 425 in.) is equal to −18.7 dB, which is within the desirable range of −18 dB to −20 dB. The combined phase and spillover efficiency is about 96.4 percent, where the maximum efficiency is about 97.8 percent, at θ =~11.5 deg. This is a typical design point for a dual-shaped system, since the use of the maximum efficiency point results in a larger subreflector. It should be noted that the results for a 21-in. (~533-mm) aperture were very similar to those of the 19-in. aperture, but the 21-in. aperture results in a horn that is 11 in. (~280 mm) longer at X-band; hence, the smaller design was chosen.

From Fig. 6(b), the GO edge taper at the rim of any BWG mirror (at r = 260 in.) is ~ −23.6 dB at θ = 10.4 deg, with 96.5 percent efficiency. The maximum efficiency is equal to 96.7 percent at θ = 9.8 deg, which falls between the desired values of 8.7 and 10.4 deg. Because the 19-in. X-band horn has radio frequency (RF) radiation characteristics that meet the requirements, it is therefore used in the design of BWG mirrors and synthesis of dual-shaped reflectors.

In similar analyses, the +22-dbi horn required for the F_3 pedestal room focus was examined at a range of 165 in., as discussed in section I.D.

C. Dual-Shaped Reflector Design

An essential requirement of the 34-m main reflector is to maintain the newly designed BWG antenna surface contour within ±0.5 in. (13 mm) of the previously designed HEF antenna. This makes it possible to retain the existing backup structure and adjust the individual reflecting panels with existing standoffs to fit the newly designed contour.

The X-band feedhorn pattern at r = 425 in. (the mean distance to the subreflector) is used as an input pattern to a high-resolution synthesis program developed by Galindo-Israel.[1] The input parameters were similar to the HEF antenna design and are shown in Figs. 7(a), 7(b), and 7(c). The maximum difference between the main reflector surfaces of DSS 13 and the HEF antenna is 0.43 in. (11 mm), allowing HEF antenna panel forming tooling and standoff hardware to be reused.

D. Center-Fed BWG

This section provides technical details for the design of the centerline BWG feed system (Fig. 3). GO was used to design the centerline system of the first four mirrors (M1 to M4), while PO was applied to the design of the curved mirror (M5) in the pedestal room. As shown in Fig. 3, the first mirror, M1, has azimuth and elevation rotations together with the main reflector and subreflector. A plane surface is used for M1 to ensure an imaged feed pattern that is independent of the elevation angle of the antenna. Mirrors M2, M3, and M4 are attached to that part of the structure with azimuth rotation only. The last two mirrors, M5 and M6, are stationary on the ground in the pedestal room.

[1] V. Galindo-Israel, "Circularly Symmetric Dual-Shaped Reflector Antenna Synthesis With Interpolation Software-User Manual" (internal document), Jet Propulsion Laboratory, Pasadena, California, January 1988.

For a long RF ray path, curved mirrors are needed to refocus and guide energy from F_1 to F_2 (and later on to F_3) with acceptable spillover loss. It is preferable to have two curved mirrors arranged so that Mizusawa's conditions [1] are met, thus achieving minimum cross-polarization at high frequency. Oversize flat plates are used merely to direct the RF beam into desired directions without changing any other characteristic of the RF beam.

The choice of two identical paraboloidal sections for M2 and M3 has the following advantages:

(1) In the GO limit, a circularly symmetric input pattern still retains the original symmetrical shape after reflection through both (or all four) surfaces.

(2) Since there is no focal point between the two curved mirrors, as there would be with ellipsoids, RF performance is not sensitive to the spacing (L_3) between the two mirrors, provided that the spillover loss remains small.

(3) A high-pass (filter) RF performance is obtained with very good X-band performance for 8-ft mirrors (<0.1 dB for this path).

(4) It is possible to have four identical mirrors (two for center-fed and two for bypass) when paraboloidal surfaces are used in the design.

(5) Identical mirrors are obviously more economical.

The centerline BWG paraboloidal mirrors are positioned so that feedhorns and instrumentation packages can be either in an alidade location (not presently planned for implementation) or the pedestal room. Spacing between the two paraboloids, L_3 = 360 in. (~9144 mm), is chosen to allow enough headroom for vertical orientation of S-, X-, and Ka-band/amplifier subassemblies. Also, the S-band spillover loss at this distance is acceptably small. A flat plate, M4, reflects the RF beam downward along the antenna azimuth axis to the pedestal room, with focal point F_2 about 85 in. (~2 m) above the azimuth floor and about 195 in. (~5 m) above the pedestal room ceiling.

A significant decision was whether to locate the feeds on the alidade at focal point F_2 (requiring 29-dBi gain feeds) or in the pedestal room under the antenna, using focal point F_3. Despite an additional RF loss going from F_2 to F_3, the clear advantages of using the pedestal room (more available space, no cable wrap across the azimuth axis, smaller feeds required, etc.) led to its selection. The stable environment of the pedestal room was a major determinant.

Only X- and Ka-bands are planned for Phase I operation of DSS 13. However, the design must have capabilities for future S/X-, X/Ka-, C-, and Ku-band operations (S-band is 2 GHz, C-band is 4–6 GHz, and Ku-band is 13–15 GHz). Low-gain horns (~22 dBi) are desirable for all frequency bands. A basic layout for the RF design in the pedestal room is given in Fig. 2. Mirror M5 is an ellipsoidal surface used for magnifying gain (reducing beamwidth) from 22 to 29 dBi and switching among various horns by rotating M5 about the antenna aximuth. Mirror M6 is a flat plate used to reflect the RF beam from a vertically positioned feedhorn to M5, with an angle θ = 60 deg. The 60-deg angle is preferred because the existing JPL dichroic plate is designed with a 30-deg incident angle (equivalent to θ = 60 deg). Therefore, the θ = 60 deg angle will be convenient for simultaneous operation (S/X- and X/Ka-band) while reusing the existing JPL dichroic technology. Even though a smaller angle of θ would yield a more symmetric beam pattern, angles smaller than 50 deg will have shadowing problems among M5, M6, and the feedhorn. The curvature of M5 is determined by placing the near-field phase center of the 22-dBi X-band horn at a focal point of M5 (F_3) and calculating the field at M3 by using PO. Iteration continues by changing the surface curvature of M5 until the scattered field has an average edge taper at M3 of about −23 dB. The mirror M5 is adjusted vertically until the best-fit phase center of the scattered field of M5 overlays F_2. The curvature and position of M5 are designed at X-band, and there is no vertical adjustment of the mirror for other bands. The Ka-band horn (or other high-frequency horns) must be defocussed and the gain increased slightly (from 22 to 23 dBi) to approximate the same edge taper and best-fit phase center as at X-band. The detailed RF design layouts in the pedestal room for X- and Ka-bands are shown in Figs. 8 and 9. There are small lateral translations of the feedhorns to compensate for radiation pattern asymmetry due to the surface curvature of mirror M5.

The theoretical performance of the BWG system is determined by using various combinations of analytical software, as described in [3] and [4]. Figure 10 shows the measured pattern of the input of the 22-dBi horn fed at F_3, the calculated output of the ellipsoid at F_2, and demonstrates the X-band gain-magnifying (beamwidth-reducing) property of the ellipsoid. Figure 11 shows the X-band output of the BWG at F_1 compared with both the calculated input at F_2 and the measured 29-dBi horn. Figure 12 shows a comparison of the E- and H-planes of the BWG output. The system is designed to image the 29-dBi horn of Fig. 5 at the input to the dual-reflector system. Figure 13 shows the input and output of the BWG at Ka-band and illustrates the nearly perfect imaging properties of the

paraboloid pair. Figure 14 is a comparison of the 29-dBi horn and the BWG feed for the dual-reflector system at X-band.

E. Bypass BWG Design

A general layout of the bypass BWG is shown in Fig. 15(a). All mirrors rotate in the elevation plane except M10. To allow enough clearance between mirror M10 and the elevation bearing, a bypass BWG vertical tube must be positioned at Hoop 6, which is about 403 in. (~10.2 m) from the antenna center line. A retractable flat plate, M7, is out when the center-fed BWG mode is used. Mirrors M8 and M9 are paraboloidal surfaces positioned to satisfy Mizusawa's conditions. The mirrors M7, M8, and M9 are attached to and move together with the main reflector structure. A flat mirror, M10, is attached to an elevation bearing; it is not rotated with elevation rotation (but moves with azimuth rotation) in order to have a focal point F_4 always pointing straight downward to the alidade platform. By carefully adjusting L_5 and L_6 so that the distance from F_1 to the mirror M8 is equal to 260 in. (6604 mm), the paraboloidal mirrors M8 and M9 are identical to the mirrors M2 and M3 in the center-fed BWG design. Thus, there are four identical curved mirrors in this double BWG feed system.

The value of L_5 used in this design is 290.645 in. (7382 mm), which is the same as the spacing between mirrors M8 and M9. There is also enough clearance between an incident ray at the lower rim of M8 and the rim of the opening hole on the surface of the main reflector. Figure 15(b) shows detailed dimensions of the bypass BWG design. Observe that the bypass performs slightly better than the center BWG, due to the absence of the ellipsoidal magnifier mirror and the shorter main path (290 versus 360 in.).

F. Microwave Performance Summary

Table 1 lists the BWG losses at X- and Ka-band for both BWG systems and shows the reference HEF performance. The loss due to spillover was calculated with the assumption that the mirrors are in free space and that the energy not impinging on the mirrors is lost.

II. Conclusions

The new 34-m antenna fed with either a center or bypass BWG will lose 0.2 dB or less, as compared with a standard-fed Cassegrain antenna at X- and Ka-bands. The antenna is currently under construction and is scheduled to be completed in July 1990. Phase 1 of the project will provide independent X- and Ka-band receive-only operation. Phase 2 of the project will provide for simultaneous S- and X-band or X- and Ka-band operation, and the design is currently under way.

Acknowledgments

The authors would like to thank M. Esquivel for calculating the horn patterns and J. Cucchissi for supplying information on the IRG. J. Withington supplied highly accurate and crucial experimental checks of BWG performances by use of a 1/4 scale model and special data recording and analysis techniques. We would also like to thank the JPL Supercomputing Project for computer time on the Cray X-MP/18.

References

[1] M. Mizusawa and T. Kitsuregawa, "A Beam Waveguide Feed Having a Symmetric Beam for Cassegrain Antennas," *IEEE Trans. Antennas and Propagat.*, vol. AP-21, pp. 884–886, November 1973.

[2] M. Mizusawa, "Effect of Scattering Pattern of Subreflector on Radiation Characteristics of Shaped-Reflector Cassegrain Antenna," *Trans. IECE,* vol. 52-B, Japan, pp. 78–85, February 1969.

[3] T. Veruttipong, J. Withington, V. Galindo-Israel, W. A. Imbriale, and D. A. Bathker, "Design Considerations for Beam Waveguide in the NASA Deep Space Network," *IEEE Trans. Antennas Propagat.*, vol. AP-36, pp. 1779–1787, December 1988.

[4] V. Galindo-Israel, T. Veruttipong, and W. Imbriale, "GTD, Physical Optics, and Jacobi-Bessel Diffraction Analysis of Beam Waveguide Ellipsoids," *IEEE AP-S International Symposium Digest,* vol. 2, Philadelphia, Pennsylvania, pp. 643–646, June 1986.

Table 1. Beam waveguide performance*

Frequency, GHz	Gain, dBi (100 percent efficient)	HEF Cassegrain	DSS-13 Bypass BWG		DSS-13 Center-fed BWG		
		Gain, dBi	Gain, dBi	Portion due to spillover	Gain, dBi	Paraboloid spill portion	Ellipsoid spill portion
8.45 X-band	69.57	69.21	69.13	−0.06	69.06	−0.05	−0.06
31.4 Ka-band	80.98	80.62	80.55	−0.06	80.42	−0.03	−0.03

*Losses due to surface rms, BWG mirror misalignments, subreflector support blockage, and feed system I^2R are not included.

ORIGINAL PAGE
BLACK AND WHITE PHOTOGRAPH

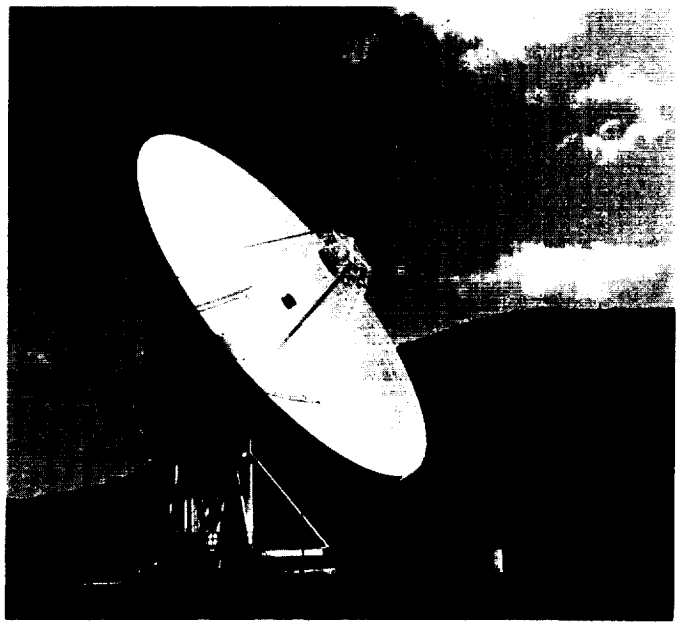
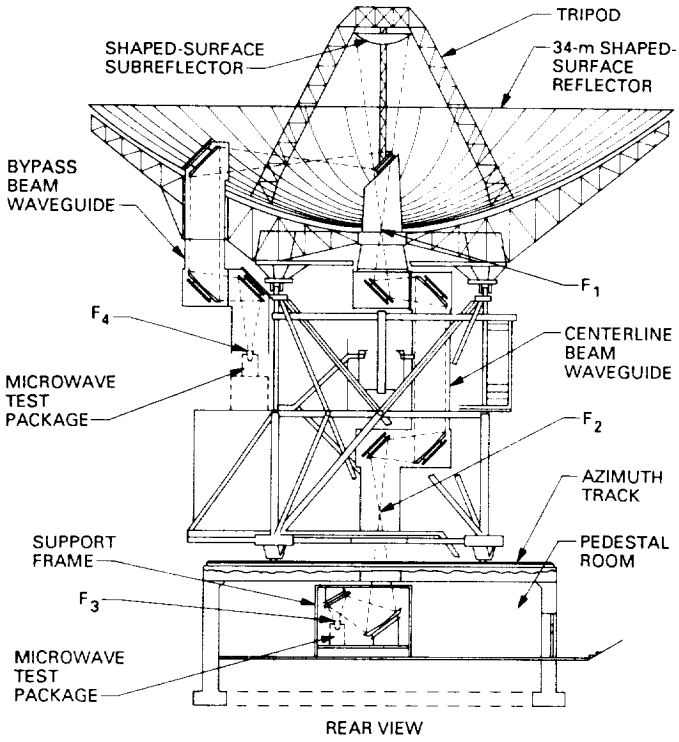


Fig. 1. The 34-m high-efficiency-frequency antenna.



NOTE: IN THIS CONFIGURATION, THE BYPASS BWG IS ACTIVE

Fig. 2. New 34-m beam waveguide antenna.

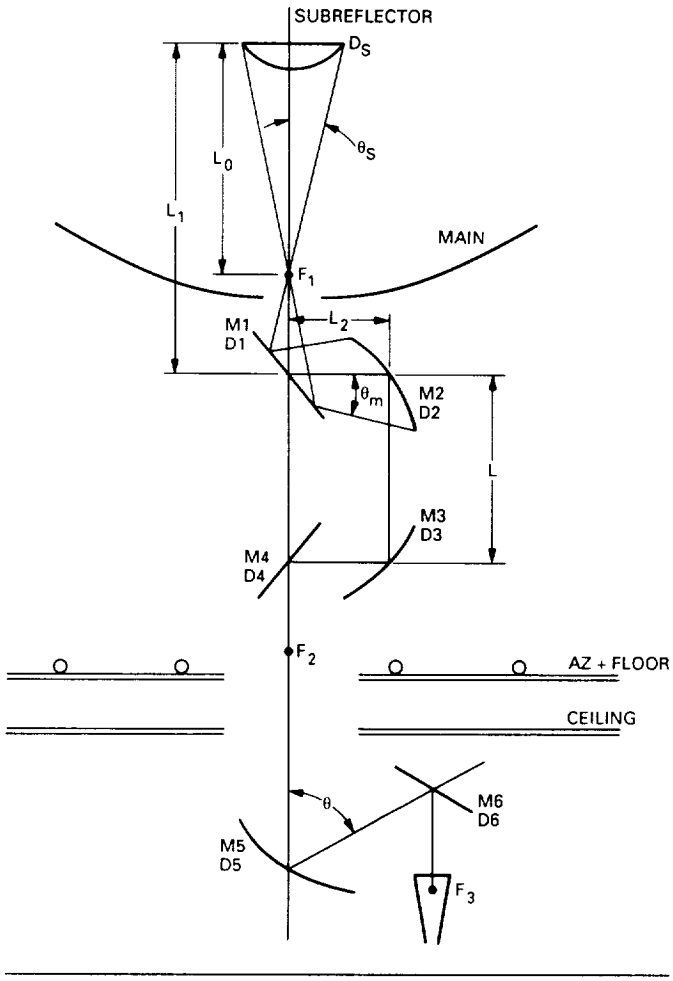


Fig. 3. A center-fed beam waveguide layout.

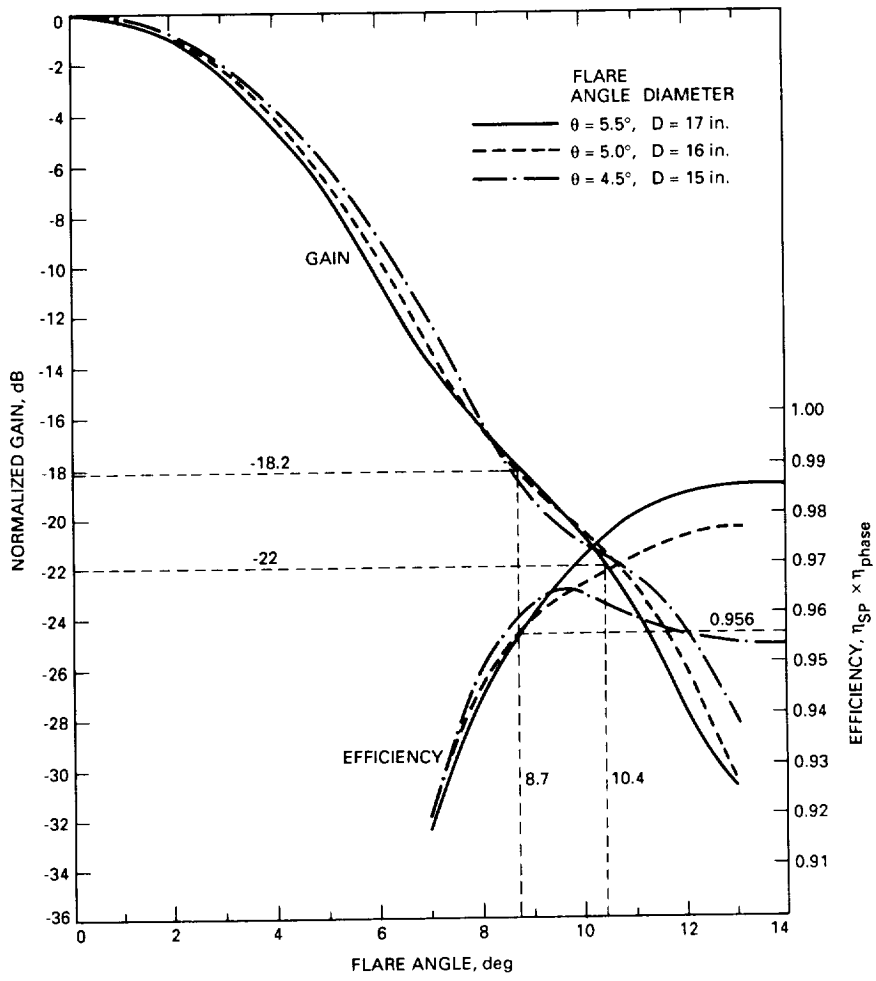


Fig. 4. Gain and efficiency versus flare angle.

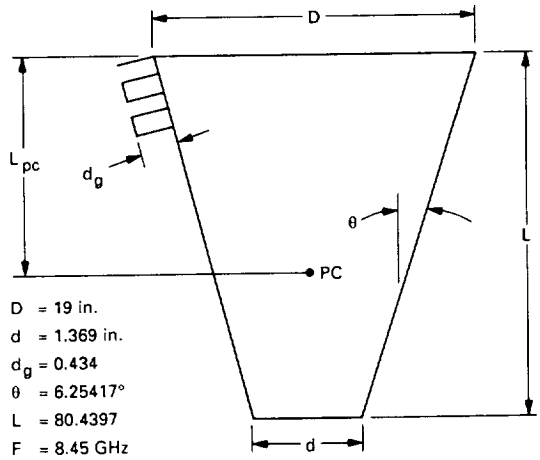


Fig. 5. Horn dimensions.

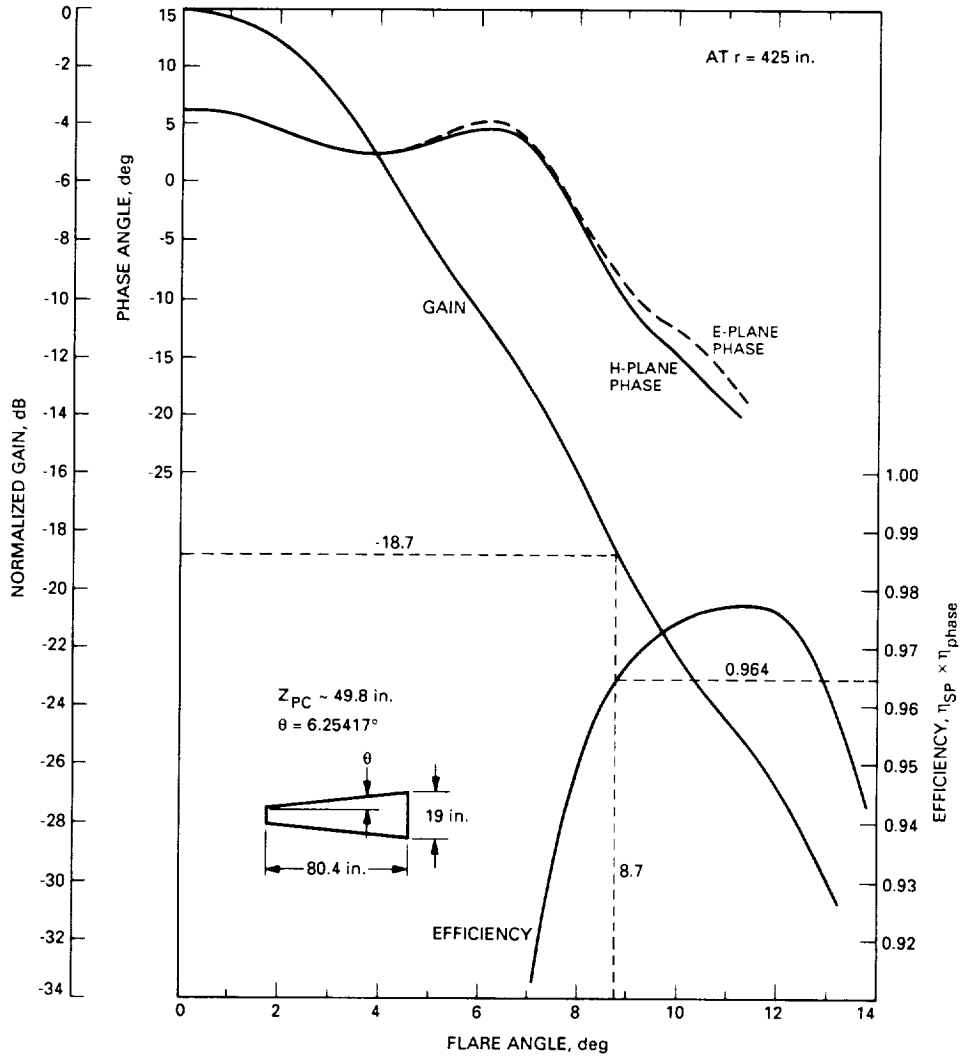


Fig. 6(a). Horn selection study (at r = 425 in., subreflector), existing JPL flare angle.

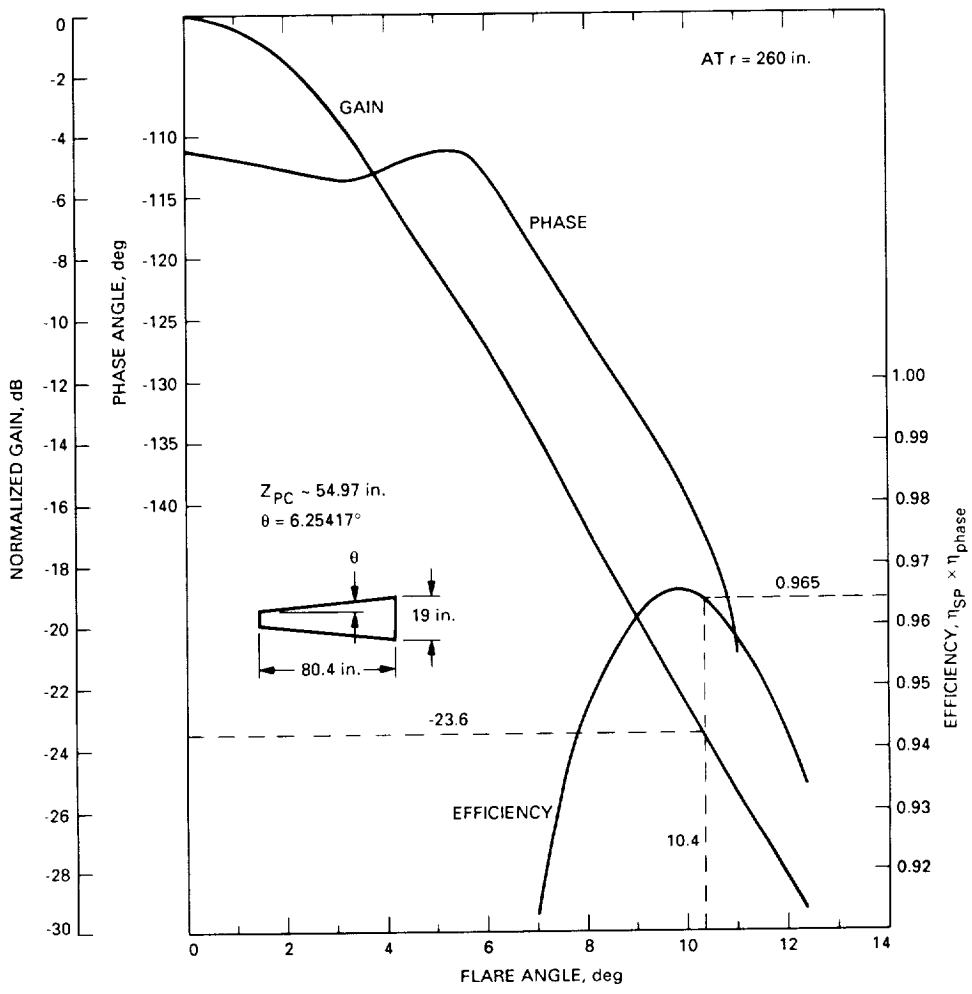


Fig. 6(b). Horn selection study (at r = 260 in., beam waveguide), existing JPL flare angle.

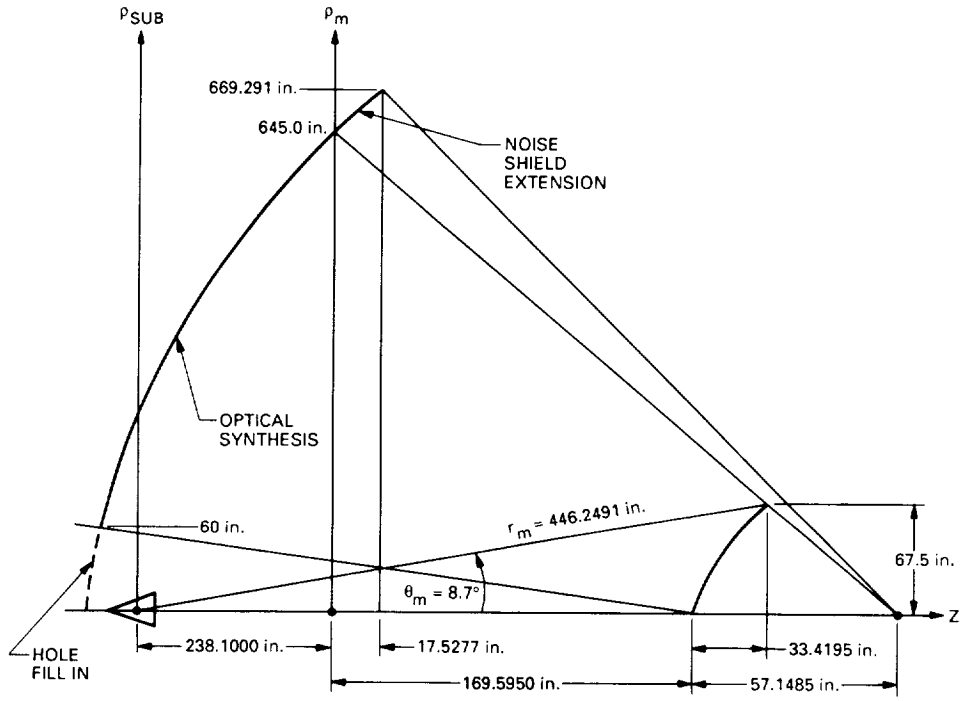
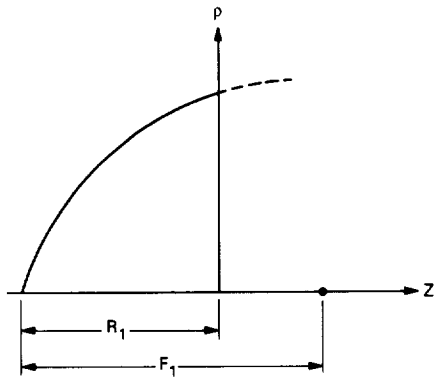
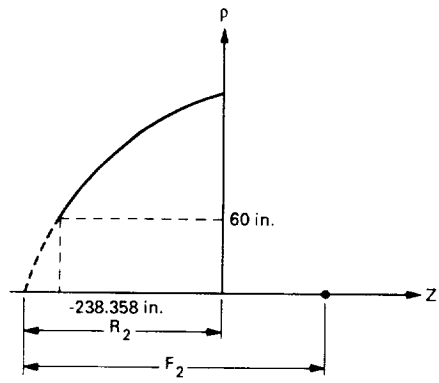


Fig. 7(a). The 34-m beam waveguide antenna geometrical synthesis parameters.



$$Z_{Noise} = -R_1 + \frac{\rho^2}{4F_1};\ R_1 = 228.48 \text{ in.},\ F_1 = 455.22 \text{ in.}$$

Fig. 7(b). Equivalent paraboloid for the noise shield.



$$Z_{Hole} = -R_2 + \frac{\rho^2}{4F_2};\ R_2 = 240.55 \text{ in.},\ F_2 = 410.14 \text{ in.}$$

Fig. 7(c). Equivalent paraboloid for the hole.

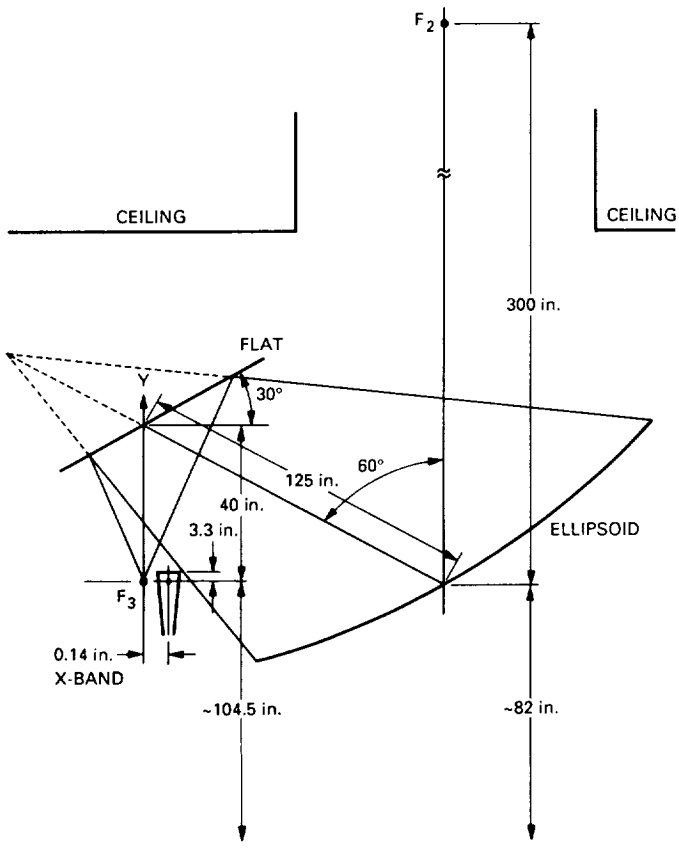


Fig. 8. Pedestal feed system for X-band.

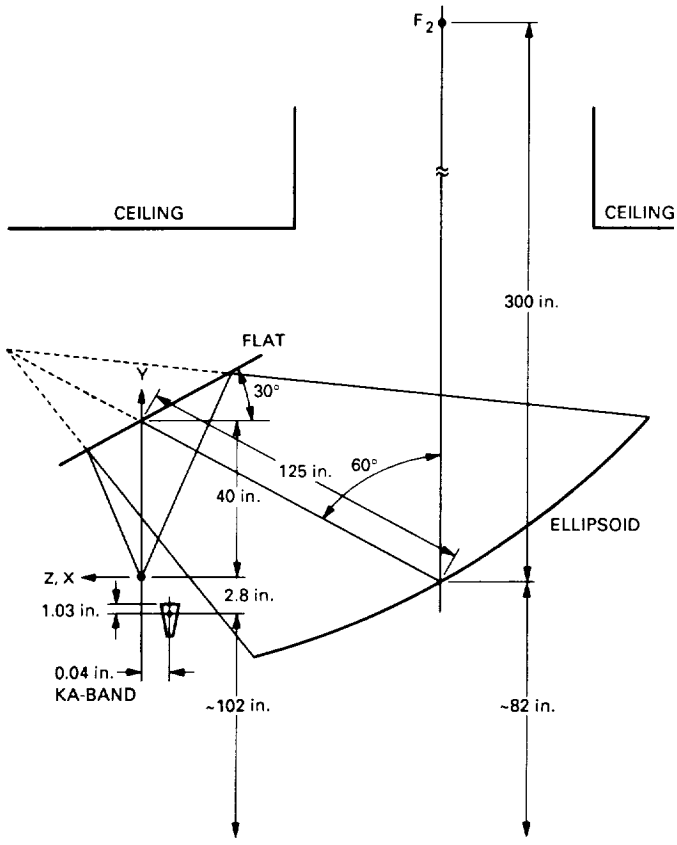


Fig. 9. Pedestal feed system for Ka-band.

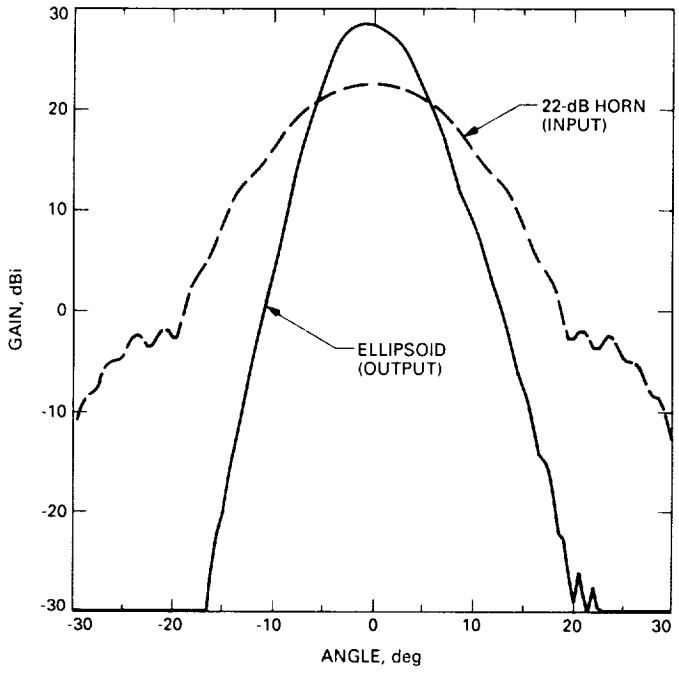


Fig. 10. Beam magnifier ellipse.

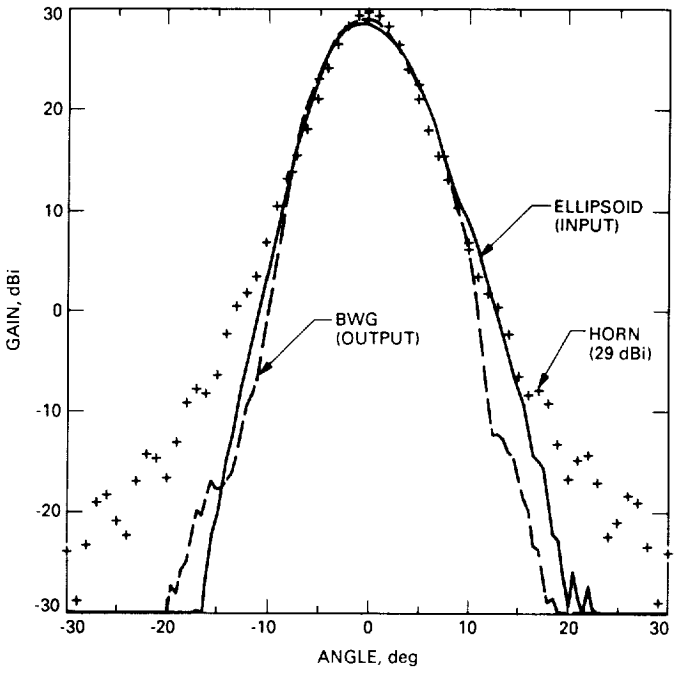


Fig. 11. Center-fed output beam waveguide.

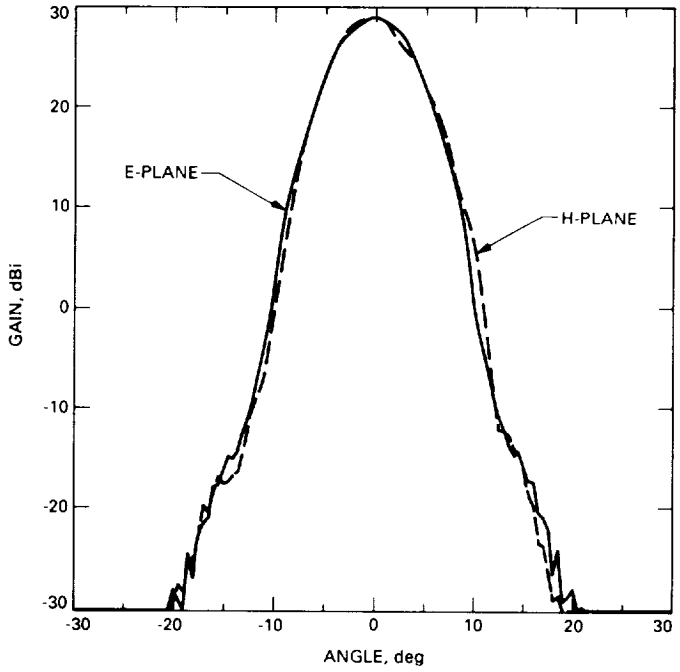


Fig. 12. X-band E- and H-plane beam waveguide output.

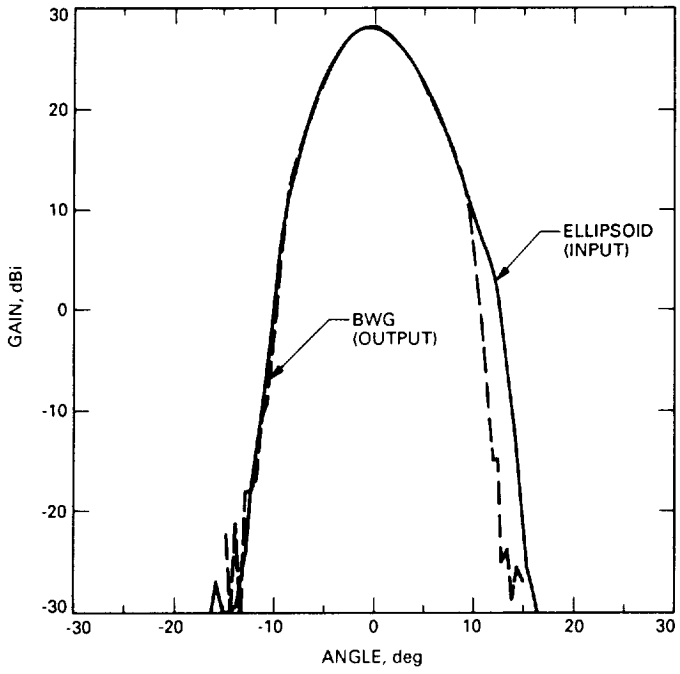


Fig. 13. Center-fed beam waveguide output (Ka-band)

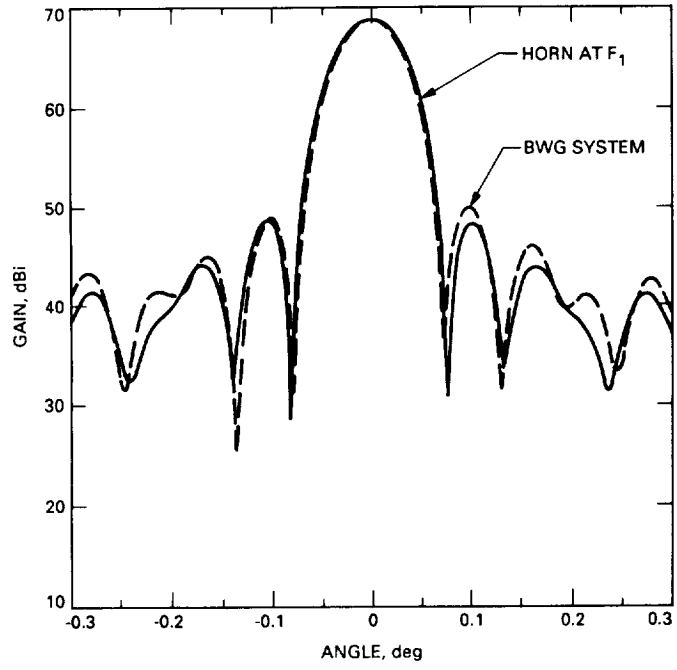
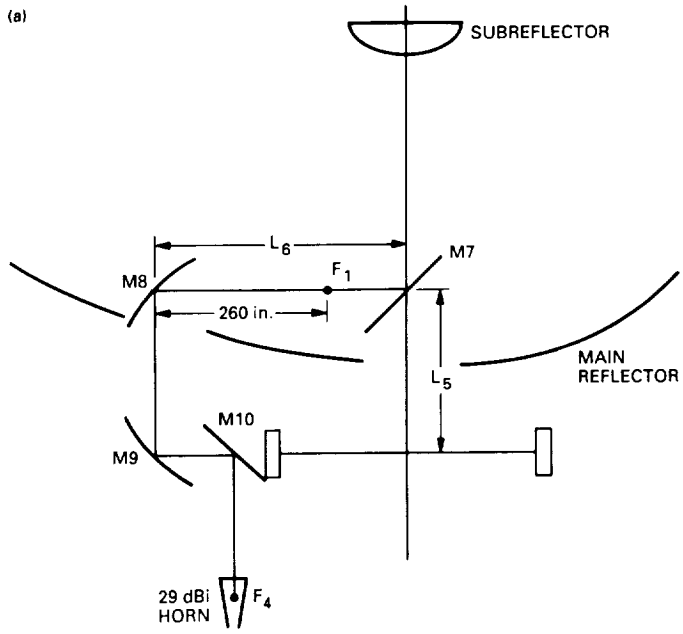


Fig. 14. Comparison of horn versus beam waveguide feeding dual-reflector system.



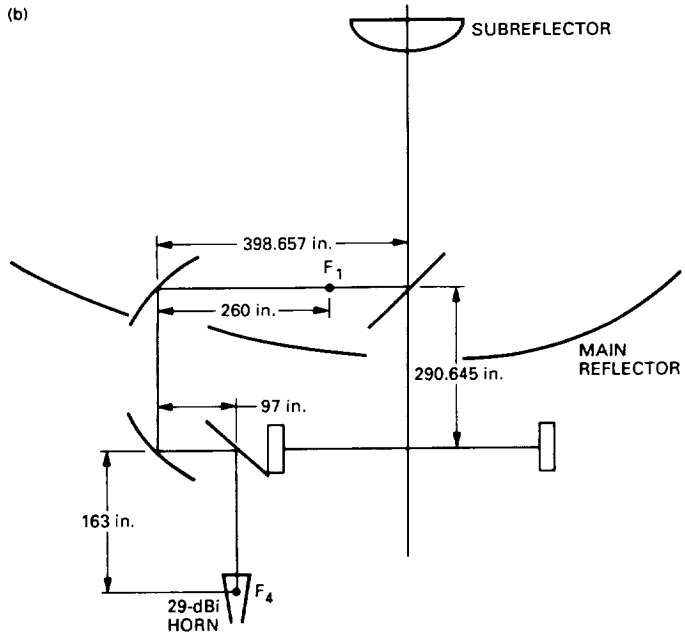


Fig. 15. Bypass beam waveguide geometry: (a) general layout; (b) detailed dimensions of the bypass beam waveguide design.

A Two-Telescope Receiver Design for Deep Space Optical Communications

K. Shaik
Communications Systems Research Section

A two-telescope system for a deep space optical communication receiver is proposed. To substantially improve the overall receiver efficiency, the design reserves the large telescope for signal-detection purposes only. A small diffraction-limited telescope is introduced to accomplish the task of navigational tracking.

I. Introduction

A 10-m hexagonally segmented Cassegrain optical telescope has been proposed for use as an Earth-vicinity receiver for laser communications from planetary spacecraft [1–3]. Such a reception station on the ground may provide, for example, a channel capacity of 5 Mbits/sec for a 0.3-m-diameter transmitter at Mars. Even though the first reception station will be ground based, the goal is to develop and demonstrate technology which can be used to put a telescope into Earth orbit to avoid the deleterious effects of the atmosphere on optical communications. The technology has to be economical enough to allow replication for either a ground- or a space-based network.

The present design requires a demonstration of direct detection of optical signals for typical data rates from planetary spacecraft in the presence of considerable solar background interference, and also the ability to acquire and track spacecraft signals relative to the stellar background by means of a single 10-m telescope. Since the two functions have somewhat different requirements, the necessary compromises are likely to make the system less efficient. This difficulty can be avoided by including a well-corrected second telescope that is roughly one meter in diameter at a relatively small cost to the system, to accomplish the acquisition and navigational tracking of spacecraft, which leaves the larger telescope for signal-detection purposes only. The following sections discuss the motivation for and the design aspects of a two-telescope optical reception station.

Note that the introduction of a small, well-corrected telescope cannot contend or compete with the signal-detection capabilities of the large telescope. For short signal pulses, the small telescope will become photon starved for most ranges of interest. It is only when the integration times can be long, as in the case of navigational tracking, that the small telescope becomes a useful tool by providing a good signal-to-noise-ratio (SNR) for daytime observations. The introduction of a small telescope also provides an opportunity to optimize the design of the large telescope for communications. The following discussion focuses only on the navigational tracking of the spacecraft and describes derivative benefits for the receiver system if a two-telescope approach is adopted.

II. System Design with Two Telescopes

A. Motivation

The cost of a telescope depends heavily on the size and the surface accuracy of the primary mirror. A large collector area (primary) is needed to ensure that there are enough photons per signal pulse for the detector. The image quality, however, can be sacrificed by accepting a high-surface-tolerance primary. The present design and cost analysis permit a primary with a root-mean-square tolerance of about four wavelengths in the visible range. Also, the design envisions a fast primary ($f/0.5$), and no active or adaptive figure-control mechanism for the correction of gravity sag and other tilt errors. Such decisions considerably reduce the cost of a complete system; however, the expected spot size at the focus becomes quite large. Calculations show that the contribution to the blur circle by the mirror surface roughness alone will have a half-width, half-maximum (HWHM) angle of about 50 μrad for a point source. A further deterioration of similar magnitude is expected due to the gravity sag, atmospheric seeing, and other errors, making the eventual system HWHM angle for the blur circle about 100 μrad.

Radiant intensity (watt/steradian) at the focus is proportional to A_c/ϕ^2, where A_c is the area of collection for the mirror and ϕ is the blur circle HWHM angle for a point source. The HWHM blur angle for a diffraction-limited 1-m telescope is largely determined by atmospheric conditions. Hence, everything else being the same, the radiant intensity for a diffraction-limited 1-m telescope with a nominal $\phi \simeq 3$ μrad will be about 10 times higher than that for a 10-m telescope with $\phi \simeq 100$ μrad. Under adverse atmospheric conditions (seeing $\phi \simeq 10$ μrad), which may occur less than 10 percent of the time during the midmornings and the late afternoons at well-chosen sites, the smaller telescope can do no worse than the larger telescope. Such extreme conditions, however, can be avoided by scheduling the acquisition and tracking events at other times.

For the large, one-telescope system, when the task of navigational tracking of the spacecraft coordinates at the tracking detector is to be performed simultaneously with the gathering of signal photons at the communications detector, only a small fraction of the collected light can be directed to the tracking detector without jeopardizing the prime function of the telescope. Assuming 10 percent of the light for the one-telescope system going to the tracking detector, the radiant intensity for the small telescope, which is dedicated to navigational tracking only, can be 100 times higher than that of the large telescope. However, calculations in Section B of this article ignore this fact and assume that all the light collected by the large telescope is available for navigational tracking.

Use of a two-telescope arrangement provides a number of opportunities to optimize the design of each of the telescopes for its designated function. Some of these refinements, which may substantially improve the system sensitivity, are discussed in the following paragraphs.

The less efficient broadband optical coating required to image starfields for acquisition and tracking for the 10-m telescope can be discarded in favor of a highly efficient narrowband optical coating specific to the signal wavelength. The reflectance per surface can go up from about 0.75 to 0.98, resulting in (1) an increase of more than 1 dB per surface in the available radiant intensity at the signal detector, and (2) a substantial decrease in out-of-band background noise, which may help simplify the post-tertiary optical design.

The stray-light rejection capabilities for the system can be much higher. Since the small telescope, with its broad field of view (FOV), is now used for blind pointing, acquisition, and tracking, the FOV of the large telescope can be decreased to limit stray light, albeit at the expense of introducing some complexity by requiring a mechanism to transfer accurate position coordinates from the small to the large telescope.

Indeed, some form of restriction on the telescope passband or on the FOV, or both, may be necessary to protect the post-tertiary optics for a one-telescope receiver. Under certain conditions, the collected energy for the 10-m telescope for daytime observations can be several watts. Special schemes may have to be devised to protect the detectors and to dissipate the heat generated.

Post-tertiary optics are further simplified as there is no need for a removable beam deflector with its random mechanical placement errors to direct the beam to the tracking detector, when intermittent acquisition and tracking updates are required. The case of a permanent beam splitter for continuous tracking has already been discussed above.

B. Telescope Performance for Navigational Tracking

The baseline parameter values for deep space optical communications envisioned for Moon-Earth, Mars-Earth, Saturn-Earth, and the Thousand Astronomical Units (TAU) missions are given respectively in columns A, B, C, and D of Table 1. Using the parameter values given in Table 1, the received signal flux and the corresponding number of photoelectrons per second generated at the

detector (as well as a number of other relevant quantities) can be calculated by the computer program [4]. Data from this program are obtained for the typical link profiles described in Table 1, which are then used in the SNR formula developed below to calculate expected SNR for the telescopes in question in the presence of typical daytime and nighttime background noise.

1. SNR calculation. If n_s is the total number of signal photoelectrons per pixel at the detector, and n_b, n_{th}, n_d, and n_r represent sky background, thermal, dark-current, and readout noise photoelectrons per image spot size of a point source respectively, then the SNR is defined as

$$\text{SNR} = \frac{n_s}{[n_s + n_b + n_{th} + n_d + n_r]^{1/2}} \tag{1}$$

For observations in the visible spectral range, the noise contribution from the thermal and the dark-current noise is very small compared to the readout and the sky background noise, and will be ignored. Define the following variables:

λ = signal wavelength, nm

t = integration time, sec

η_a = Earth-atmosphere transmission efficiency

η_d = detector quantum efficiency

η_r = receiver optics efficiency

$\delta\lambda$ = source-laser bandwidth, nm

$\Delta\lambda$ = filter bandwidth centered about λ and $\delta\lambda \subseteq \Delta\lambda$, nm

f_L = signal-laser spectral flux, W/(m$^2\cdot$ nm)

A_s = noise-equivalent sky area, sr

F_s = sky-background spectral flux per solid angle, W/(m$^2\cdot$ nm $\cdot$ sr)

N = average number of photons/joule about wavelength λ

Neglecting n_{th} and n_d, and expressing the signal and the noise photocounts in terms of the variables defined above, Eq. (1) can be rewritten as

$$\text{SNR} = \frac{K f_L \delta\lambda}{[K(f_L\delta\lambda + A_s F_s \Delta\lambda) + n_r]^{1/2}} \tag{2a}$$

where A_c is the primary collector area as defined earlier, and $K = N\eta_d\eta_r\eta_a A_c t$.

Photoelectron counts per second per pixel due to the readout noise are about seven counts for a typical charge-coupled device detector. Allowing for a 50-msec integration time and a 4-pixel image size for a point source, the expected number of photoelectrons from the readout noise is found to be about 1.5 photoelectrons per integration time. Using nighttime sky background radiance given in Table 1, the corresponding number of photoelectrons for the small telescope is found to be 0.2 per 50-msec. The number of photoelectrons due to the sky background collected by the small telescope during the daytime and by the large telescope during the daytime as well as the nighttime is very large compared to all other noise sources discussed above. Therefore, for the large telescope at all times and for the small telescope during the daytime, the above equation can be further reduced to

$$\text{SNR} = \frac{K^{1/2} f_L \delta\lambda}{[f_L\delta\lambda + A_s F_s \Delta\lambda]^{1/2}} \tag{2b}$$

Determination of the effective noise area, A_s, that contributes to sky background noise is a difficult problem. It has been shown that $A_s = 2.28\pi\phi^2$ is a good approximation as long as the pixel size is less than the blur radius (ϕ in radians), and a two-dimensional Gaussian point-spread function is assumed [5–7]. This expression for the noise-equivalent area will be used to calculate specific results below.

The resonant frequency of the telescope structure, in the absence of active controls and involved post-processing of signals, determines the longest integration time. The maximum resonant frequency for the 10-m telescope structure, which may also house the tracking telescope, is expected to be about 5 Hz. Using a factor of four for safety, an integration time of 50 msec (20 Hz) can be used. Further improvement in the SNR is possible if some shuttering scheme is devised to prevent background light from reaching the detector when the signal pulse is not expected. The length of time which the detector can remain shuttered can be a large fraction of 50 msec, as the pulsewidth is likely to be much smaller ($\sim$ 10 ns) than the dead time between pulses ($\sim$ 100 μsec). However, no such shuttering scheme is assumed in the following examples.

An optical communications link program in [4] is used to calculate the expected signal power and the corresponding photoelectrons generated at the detector for the link profiles given in Table 1 for both the 1-m and the

10-m telescopes. These results are then used in conjunction with Eqs. (2a) and (2b) above to obtain SNR estimates for both the daytime and the nighttime sky-background radiances. In the case of nighttime calculations for the Moon-Earth, Mars-Earth, and the Saturn-Earth links, the presence of the Moon, Mars, and Saturn, respectively, in the background is assumed. A number of additional calculations using the TAU link profile for ranges between 50 and 1000 AU was also made.

The results for the nighttime and the daytime performance of the two telescopes are summarized in Figs. 1 and 2. In these figures, links A, B, and C assume the presence of the Moon, Mars, and Saturn, respectively, in the receiver field of view. Links D_1, D_2, D_3, and D_4 represent link-profile D in Table 1 with 50, 100, 500, and 1000 AU range, respectively. The SNR for the 1-m telescope for link profiles A through C for both the daytime and the nighttime observations remains higher than unity. Calculations with the TAU profile for the 1-m telescope were made only up to a range of 500 AU, where the signal photoelectrons begin to be comparable to the threshold of detection. In all cases except the daytime result at 500 AU, the SNR is found to be better than unity for the 1-m telescope. For the 10-m telescope, the nighttime SNR is higher than with the 1-m telescope for all link profiles considered. The daytime SNR for the Earth-Saturn and the TAU links, however, falls short of unity. Note that use of the same expression for A_s and the assumption that the entire signal power is focused on the tracking detector for the 10-m telescope may not be accurate, thereby overestimating the SNR for the larger telescope.

2. Navigational tracking and star magnitudes. To gain a sense of the magnitude of a star that may be visible to the telescope for navigation, Eq. (2b) may be expressed, in terms of star magnitudes, as follows:

$$m = -2.5 \log \left[\frac{(\mathrm{SNR})^2}{2Kn} \left\{ 1 + \left(1 + \frac{4KNF_sA_s}{(\mathrm{SNR})^2} \right)^{1/2} \right\} \right] \tag{3}$$

where m is the visual magnitude of a star and $n = 10^8$ photons/(sec·m^2 · nm) is the canonical value at the top of the atmosphere for a zero-magnitude A0 star at 550 nm. Note that the passband for both the navigational star and the background noise is limited by $\Delta\lambda = 200$ nm. This passband essentially represents the width of a V-filter, which is designed to cover the visible spectral range, and is widely used for star-tracking purposes. For SNR = 1, the above equation can be simplified to give

$$m = -2.5 \log \left[\frac{1}{2Kn} \left\{ 1 + (1 + 4KNF_sA_s)^{1/2} \right\} \right] \tag{4}$$

For the typical parameter values pertaining to the tracking telescope and the communication channels given in Table 1, a star brighter than magnitude 13 will be visible to the 1-m tracking telescope with an SNR of unity before the threshold of detection is reached for nighttime observation. The corresponding number for the 10-m telescope is found to be magnitude 18. During the daytime, the small telescope can see a star of magnitude 10 or brighter with an SNR of unity, which is better than the performance of the large telescope with a large blur radius by one visual magnitude.

C. Other Design Aspects

The size of the proposed secondary for the 10-m telescope is 1.1 m. The 1-m acquisition and tracking telescope can fit nicely on top of the secondary, but well inside the proposed sunshade for the large telescope to use it as protection against direct sunlight. This will make for a compact system. The two telescopes can be mounted rigidly with respect to each other and aligned, making the problem of angle transfer relatively simple.

It may also be possible to use the 1-m telescope as the uplink transmitter by including an uplink laser with steering mirrors for the point-ahead.

III. Conclusion

In a two-telescope design for a deep space optical communication reception station, each of the two telescopes can be optimized for its appointed function. This configuration provides a better daytime SNR and an adequate nighttime SNR for most missions within the solar system. However, it fails to perform well for the TAU link because the photon-collection capabilities of the small navigational telescope are limited. With a two-telescope system, however, a detection-sensitivity gain of several decibels can be achieved by specific tailoring of the large telescope for communications. It is concluded that for most planetary missions, adopting a two-telescope system, which may raise the system cost by about two percent, can substantially improve system performance.

Acknowledgment

The author wishes to thank James R. Lesh for his useful comments and helpful suggestions.

References

[1] E. L. Kerr, "Strawman Optical Reception Development Antenna (SORDA)," *TDA Progress Report 42-93*, vol. January–March 1988, Jet Propulsion Laboratory, Pasadena, California, pp. 97–110, May 15, 1988.

[2] E. L. Kerr, "An Integral Sunshade for Optical Reception Antennas," *TDA Progress Report 42-95*, vol. July–September 1988, Jet Propulsion Laboratory, Pasadena, California, pp. 180–195, November 15, 1988.

[3] E. L. Kerr, "Architectural Design of a Ground-based Deep-space Optical Reception Antenna," submitted for publication in *Opt. Eng.*

[4] W. K. Marshall and B. D. Burk, "Received Optical Power Calculations for Optical Communications Link Performance Analysis," *TDA Progress Report 42-87*, vol. July–September 1986, Jet Propulsion Laboratory, Pasadena, California, pp. 32–40, November 15, 1986. (The program, called OPTI or *Optical Communications Link Analysis Program*, program number NPO 17444, is available from COSMIC, 382 East Broad Street, University of Georgia, Athens, GA 30602.)

[5] I. R. King, "Accuracy of Measurement of Star Images on a Pixel Array," *Publications Astron. Soc. Pacific*, vol. 95, no. 564, San Francisco, California, pp. 163–168, February 1983.

[6] S. Faber, "The Optimum Match Between Image Size and the Size of Detector Pixels: The Choice of Angular Scale," *Keck Observatory Report No. 61*, University of California, Berkeley, Lawrence Berkeley Laboratory, June 1981.

[7] J. Nelson, T. Mast, and S. Faber, "The Design of the Keck Observatory and Telescope (Ten-Meter Telescope)," *Keck Observatory Report No. 90*, University of California, Berkeley, Lawrence Berkeley Laboratory, January 1985.

Table 1. Optical communication link parameters

Parameter name	Value[a]			
	A	B	C	D
Transmitter characteristics:				
Aperture diameter, m	0.05	0.3	0.3	1.0
Central obscuration diameter, m	0	0.06	0.06	0.2
Optics efficiency	0.65	0.65	0.65	0.65
Beam width, μrad	15	3	3	0.55
Root-mean-square pointing bias error, μrad	1.5	0.3	0.3	0.05
Root-mean-square pointing jitter, μrad	1.5	0.3	0.3	0.05
Average laser output power, W	0.1	2	2	10
Laser wavelength, nm	850	532	532	532
Laser pulse-width, ns	10	10	10	10
Channel characteristics:				
Earth atmosphere transmission	0.5	0.5	0.5	0.5
Range, astronomical units	0.0026	1.5	10.5	$\leq$1000
Daytime typical sky background radiance, $W/(m^2 \cdot sr \cdot nm)$	0.2	0.2	0.2	0.2
Nighttime typical sky background radiance,[b] $10^{-8} W/(m^2 \cdot sr \cdot nm)$	0.476	0.476	0.476	0.476
1-meter telescope characteristics:				
Aperture diameter, m	1	1	1	1
Central obscuration diameter, m	0.3	0.3	0.3	0.3
Nominal HWHM image radius for a point source, μrad	3	3	3	3
Optics efficiency	0.4	0.4	0.4	0.4
Detector quantum efficiency	0.5	0.5	0.5	0.5
Filter bandwidth for spacecraft tracking, nm	0.1	0.1	0.1	0.1
Filter bandwidth for navigational startracking, nm	200	200	200	200
10-meter telescope characteristics:				
Aperture diameter, m	10	10	10	10
Equivalent obscuration diameter, m	4.3	4.3	4.3	4.3
Nominal HWHM image radius for a point source, μrad	100	100	100	100
Optics efficiency	0.4	0.4	0.4	0.4
Detector quantum efficiency	0.5	0.5	0.5	0.5
Filter bandwidth for spacecraft tracking, nm	0.1	0.1	0.1	0.1
Filter bandwidth for navigational star tracking, nm	200	200	200	200

[a] The letters A, B, C, and D represent Moon-Earth, Mars-Earth, Saturn-Earth, and the Thousand Astronomical Unit (TAU) communication link, respectively.

[b] The nighttime sky background radiance given here corresponds to a star of visual magnitude 21.25 per arcsec^2 area in the sky.

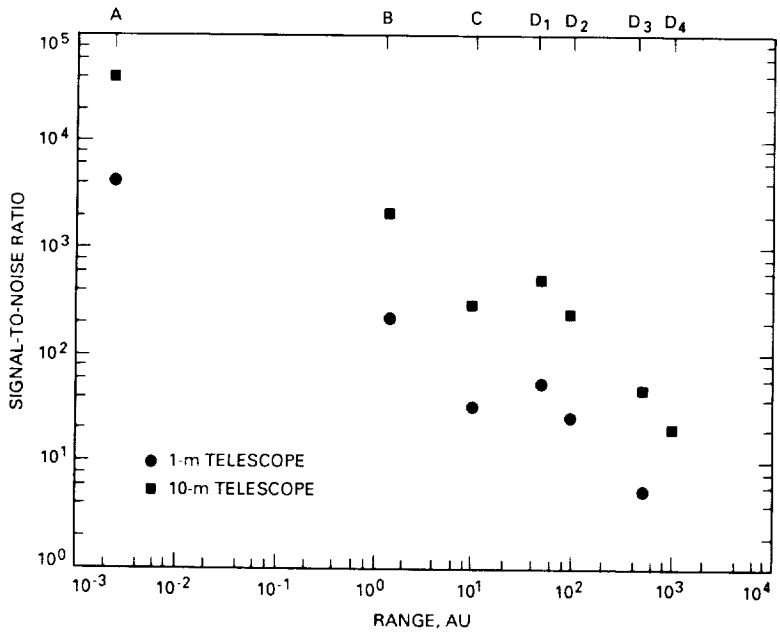


Fig. 1. Nighttime SNR with a 50-msec integration time for 1-m and 10-m telescopes, which are compared for various communication links (telescope characteristics are given in Table 1).

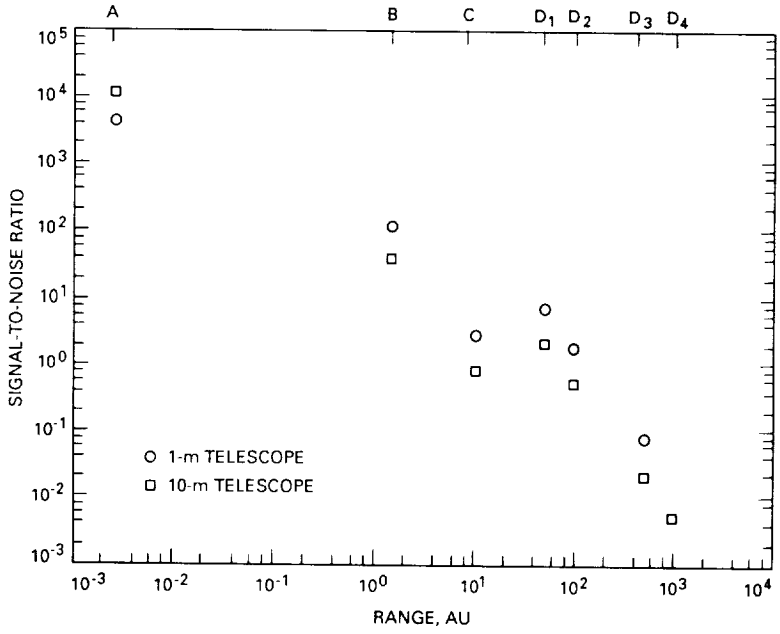


Fig. 2. Daytime SNR with a 50-msec integration time for 1-m and 10-m telescopes, which are compared for various communication links (telescope characteristics are given in Table 1).

Preliminary Error Budget for an Optical Ranging System: Range, Range Rate, and Differenced Range Observables

W. M. Folkner and M. H. Finger
Tracking Systems and Applications Section

Future missions to the outer solar system or human exploration of Mars may use telemetry systems based on optical rather than radio transmitters. Pulsed laser transmission can be used to deliver telemetry rates of about 100 kbits/sec with an efficiency of several bits for each detected photon. This article discusses navigational observables that can be derived from timing pulsed laser signals. Error budgets are presented based on nominal ground station and spacecraft-transceiver designs. Assuming a pulsed optical uplink signal, two-way range accuracy may approach the few-centimeter level imposed by the troposphere uncertainty. Angular information can be achieved from differenced one-way range using two ground stations with the accuracy limited by the length of the available baseline and by clock synchronization and troposphere errors. A method of synchronizing the ground station clocks using optical ranging measurements is presented. This could allow differenced range accuracy to reach the few-centimeter troposphere limit.

I. Introduction

Advanced spacecraft require increasingly higher telemetry rates to cope with the needs of more sophisticated scientific instruments. Increasing the frequency of the telemetry carrier generally increases the data rate by providing more power at the detector due to reduced diffraction-limited beam divergence. Changing the carrier from the radio to the optical band can improve the received-to-transmitted power ratio due to the much higher frequency. Coupled with conversion efficiency approaching 50 percent for solid-state lasers, there is a potential for reduced spacecraft weight and power requirements for a given data rate [1]. New modulation techniques can be employed with an optical telemetry system. Pulse-position modulation (PPM), in which characters are encoded in the arrival times of narrow laser pulses, is being considered for use in the outer solar system.

Spacecraft navigation is based largely on observables derived from the telemetry system. Several observables are available from an optical PPM system. In this article, the potential accuracy of tracking observables derived from the timing information available from PPM signals is examined. These observables include two-way range and differenced one-way range (DOR). The range rate can be derived from the range measurements. These observables correspond to existing radio metric data types.

While optical telemetry systems are under development, there may be a number of "hybrid" missions where normal radio metric data are available. The optical tracking techniques should be explored during these missions since they may provide greater accuracy and may eventually provide stand-alone capability. A new tracking data type is an astrometric angular measurement between the spacecraft and a target solar-system body.[1] Since most solar-system bodies are not bright in the radio band, this data type has not previously been available. Astrometric tracking is not considered in this article.

Optical systems suffer much more than radio systems from weather and daytime degradation. Sufficient reduction of daytime skylight may be attained through use of sun shades and narrow bandpass filters [2,3]. The use of a number of sites a few hundred kilometers apart may provide adequate freedom from cloud cover. Investigations into cloud statistics at candidate sites are proceeding [4]. The alternative to acceptable ground stations is the use of an orbiting communications facility. This alternative will be more expensive and more difficult to maintain and upgrade. For the purposes of this article, it is assumed that ground stations will provide adequate communications capability. The optical tracking observables will be available under standard telemetry conditions, including clear daytime conditions.

The accuracy of optical observables will depend on the parameters of the telemetry system. However, there is as yet no well defined design of a complete optical telemetry system. A nominal system providing a data rate of 100 kbits/sec from Saturn has been the basis of studies for an optical spacecraft transceiver and ground receiving stations [2,3,5]. The parameters from these previous studies have been adopted here, with extensions where necessary, to examine the tracking observable accuracies. Some specific areas where additional assumptions were made are the definition and performance of the detectors for the spacecraft and ground station and the effect of the troposphere on the uplink signals. Given signal-to-noise ratio and detector performance sufficient for telemetry operation, the DOR accuracy is likely to be limited by troposphere and clock synchronization uncertainties, while two-way range accuracy should be limited by troposphere and spacecraft transceiver uncertainties. Designing the transceiver with attention to minimization of the range error may allow two-way range accuracy of ≤ 10 cm to be achieved.

The following section briefly describes the PPM telemetry system and the method proposed for deriving two-way range and DOR. The parameters of the telemetry system assumed for this study are presented in Section III. The errors in the optical tracking observables include the uncertainty in pulse arrival time (at the spacecraft and at the ground station), instrumental delay uncertainties, media delay uncertainty, and clock error. Station location and Earth orientation uncertainties are not included, since for spacecraft navigation these errors are counted separately. The uncertainty in the pulse arrival time is due to photon statistics and instrumental effects. The photon statistical errors are discussed in Section IV. The success of satellite laser ranging (SLR) suggests that the instrumental errors can reach the subcentimeter level. In Sections V, VI, and VII, preliminary error budgets for the optical tracking observables are presented. Synchronization of ground station clocks is an important error source for DOR. In Section VIII, a method to synchronize ground station clocks through a combination of two-way and three-way range measurements is presented.

II. Range Measurements from the PPM Telemetry System

Figure 1 is a simplified diagram of the components of the optical telemetry system. The spacecraft laser transmits pulses at a wavelength of 532 nm through the dichroic splitter and main objective towards the Earth. The beam diameter at the Earth is comparable to the diameter of the Earth (for the nominal spacecraft system at Saturn). This requires the spacecraft to transmit to the predicted position of the Earth. The spacecraft main objective is used to collect incoming light from the Earth, most of which is reflected by the two dichroic splitters into the tracking detector. The position of the Earth's image on the tracking detector is used to determine the location of the Earth and compute the correct pointing angle for transmission. Steering optics (not shown) are used to deflect the outgoing laser beam by the correct amount. The tracking detector uses the visible portion of the incoming light. The infrared portion is transmitted by the second dichroic splitter for use by the uplink detector.

The ground receiver derives the telemetry from timing pulse arrivals. Since no imaging is necessary, the receiver can consist of a number of objectives added incoherently. The Strawman receiver design consists of segments of about 1-m diameter with a total collecting area equivalent to a 10-m-diameter collector. A narrow band filter (not shown) is inserted prior to the downlink detector to subtract most of the light from the sky.

The uplink transmitter sends narrow pulses of infrared light at 1060 nm for the uplink telemetry. The

[1] G. Null, "Unique Earth-based Optical-Navigation Data Types," JPL IOM 314.5-1292 (internal document), Jet Propulsion Laboratory, Pasadena, California, September 2, 1988.

uplink transmitter need not be larger than 10 cm in diameter since the troposphere smears the uplink beam by an amount corresponding to the diffraction limit of an objective diameter of 10 cm or smaller, depending on weather conditions. For this article, an uplink data rate of 0.5 kbits/sec with 10-ns-pulse widths and a PPM decoder on the spacecraft has been assumed. However, for ranging purposes, no telemetry is required; only the ability to detect uplink pulses and trigger the downlink laser is necessary. Even if no optical uplink telemetry capability is needed, the addition of an uplink detector for ranging purposes could be included.

PPM encoding/decoding is illustrated in Fig. 2. A clock is used to define slots (time intervals) of equal width. The transmitting laser is triggered in the middle of slot n, where n ranges from 1 to the alphabet size 2^N. The receiving station clock is synchronized to the transmitter clock to determine the time corresponding to the beginning of slot 1. The detection of the pulse determines which PPM symbol (out of 2^N possible symbols) has been received. The difference between the center of the received pulse and the center of the slot is measured and used to keep the receiver clock synchronized with the telemetry stream.

The detection of each pulse determines N bits of information. In principle, N bits can be received for a single detected photon. The desired telemetry rate and slot width determine the number of pulses per second required. The telemetry also includes error correction bits to achieve a given error tolerance. The nominal downlink requirements are a 100 kbits/sec data rate and a bit error rate of 10^{-3}. Employing 7/8 Reed-Soloman encoding with these requirements results in a raw telemetry rate of 114.35 kbits/sec.[2] With 8 bits per PPM symbol, 14,294 pulses/sec are needed to send 114.35 kbits/sec.

Figure 3 is a functional block diagram of the telemetry and ranging system. It is proposed that for ranging purposes the spacecraft can be put into a mode in which a single downlink pulse is sent for each detected uplink pulse. This will interrupt the telemetry for the period of the measurement, which may be on the order of a few seconds. In this mode, the downlink pulse rate will be reduced to the uplink rate. To avoid delay errors associated with the PPM decoding and encoding steps, a bypass switch can be used in ranging operation. DOR measurements can be done by sampling the regular downlink telemetry.

A two-way range measurement is derived from the time interval between the transmission of an uplink pulse and the reception of the corresponding downlink pulse. The time of the uplink pulse is determined by deflecting part of the uplink signal into a separate detector and time tagging the output following an amplifier and shaping circuit. This method avoids the uncertainty in the laser response that might be a problem for a high-power uplink laser. The spacecraft receives the uplink pulse at the uplink detector. The shaping amplifier produces a signal that, fed through the range switch, can be used to trigger the laser. The signal adder allows the spacecraft laser to be triggered by uplink ranging pulses or downlink telemetry signals. The returned ranging pulse is detected by the downlink detector, followed by an amplifier, and time tagged. A two-way range measurement is derived from the difference between the time tags for each uplink pulse and its corresponding downlink pulse. Range ambiguities can be easily resolved by imposing a pattern (ranging code) on the times of the uplink pulses.

Many error terms will be reduced by the square root of the number of measurements, so that in a short time the range error will be dominated by systematic effects such as the error in the troposphere model and instrumental biases. The number of measurements per second will be set by the uplink laser pulse rate. In the example included here, the uplink pulse rate of 63 Hz supplies a telemetry rate of 0.5 kbits/sec given a PPM alphabet size of 256 (8 bits/pulse).

DOR would be derived by time tagging downlink pulses from the downlink telemetry stream at two ground stations. The DOR error is likely to be dominated by the uncertainty in clock synchronization between the two stations. The angular accuracy depends on the separation of the ground stations. For the diffraction-limited spacecraft objective of 30-cm diameter, the beam diameter at the Earth is 650 km when the spacecraft is 1 AU from Earth. The beam diameter increases to 6500 km at 10 AU. However, a focal length adjustment may be included on the spacecraft to optimize power delivered to the ground station. This possibly could be used to deliberately spread the beam for DOR measurements to increase the available baseline at some cost in signal-to-noise ratio.

III. Nominal System Parameters

The parameters listed in Table 1 for the PPM downlink from Saturn are taken from the McDonnell Douglas transceiver study [5] and Kerr's ground station design [2,3]. The Communications Systems Research Section optical link analysis program [6] was employed to design a nominal uplink at 0.5 kbits/sec. The uplink parameters assumed for this telemetry system are given in Table 2. The

[2] W. K. Marshall, "Using the Link Analysis Program with R-S Encoded Links," JPL IOM 331-86.6-202 (internal document), Jet Propulsion Laboratory, Pasadena, California, August 1,1986.

uplink power received will be affected by beam steering and spreading due to the troposphere. These effects were estimated by using a small uplink transmitter objective (corresponding to 2-arcsec seeing) along with 0.5-arcsec steering bias and 1-arcsec steering jitter.

The parameters listed in Tables 1 and 2 have been chosen to provide a sufficient signal-to-noise ratio to meet telemetry rate requirements in the absence of detector limitations. The important parameters from the ranging perspective are the slot width, the number of signal photons received per pulse, and the number of background photons per slot. The downlink background rate is due to the daytime sky irradiance. Here a value of 100 W/(μm m^2 steradian) at 532 nm was used for the sky spectral irradiance [7]. Including the radiance of Saturn as a background source would increase the background rate by less than 5 percent [8]. The uplink background rate is due to Earthshine. The spacecraft objective, which is assumed to be diffraction limited, resolves the Earth out to 9.9 AU. That is, the Earth fills the field of view of the spacecraft telescope out to that distance. Thus, the background power received by the spacecraft is independent of the distance from the Earth up to that point. Therefore, the Earth has been treated as an extended background source with spectral radiance 6.5 W/(μm m^2 steradian) at 1060 nm [9]. The background due to scattered sunlight, which will depend on the sun shield design and objective quality for the spacecraft transceiver, has been ignored.

The proper detectors for use at the ground station and on the spacecraft have yet to be determined. Photomultiplier tubes (PMTs) are used in satellite and lunar laser ranging with timing accuracy better than the media error [10]. Daytime SLR has been performed with PMTs with good accuracy, although the gain of the PMT was reduced due to the high background rate [11]. The quantum efficiency of some PMTs at the downlink wavelength (532 nm) is about 0.1 and may be high enough for telemetry operation; if not, the quantum efficiency may be increased through internal reflection techniques [12]. At the uplink wavelength (1060 nm), the quantum efficiency of PMTs is too low for use on the spacecraft. Instead, over-biased avalanche photodiodes (APDs) are being investigated as candidate detectors. These devices can have large quantum efficiency at the uplink wavelength [13], very fast response time [14], and low dark-count rate [15]. The dead time of these APDs will probably make them unsuitable for use on the ground station, but, since the uplink background rate is lower, the APD dead time may be tolerable on the uplink.

For the rest of this article, the ground station detector is assumed to be a PMT with no dead time, linear response, and an impulse response time much less than 10 nsec. The uplink detector is assumed to be an APD with fast response time and dead time that is long compared to the PPM symbol length (2.56 μsec). The APD is activated shortly before the expected time of arrival of the PPM symbol and detects the arrival of the next incident photon.

IV. Photon Statistics of Pulse Arrival

The error in determining the time of pulse arrival depends on the detection strategy and the instrumental parameters. This section presents the errors due to photon statistics for particular uplink and downlink pulse-detection algorithms. These algorithms are shown to be adequate to derive precise range observables; they may not suffice to derive the needed telemetry accuracy.

Statistical variations in the time of arrival of the first photon in an uplink pulse result in jitter in the time of transmission of the corresponding downlink pulse, while background photons at the spacecraft can result in premature downlink pulses. Since the background rate at the spacecraft is relatively low, the detector can trigger on the first detected photon. The transmitted pulse shape is expected to have a flat top centered in the slot with some rise and fall time. A particular example of a normalized pulse shape is shown in Fig. 4, which is given by

$$f(\tau) = \frac{1}{7 \text{ nsec}} \times \begin{cases} 0 & \tau < 0 \\ \tau/2 \text{ nsec} & 0 < \tau < 2 \text{ nsec} \\ 1 & 2 \text{ nsec} < \tau < 7 \text{ nsec} \\ (9 \text{ nsec} - \tau)/2 \text{ nsec} & 7 \text{ nsec} < \tau < 9 \text{ nsec} \\ 0 & 9 \text{ nsec} < \tau \end{cases} \quad (1)$$

which defines a 9-nsec-wide pulse to fit inside the 10-nsec slot. Detecting the first photon instead of estimating the center of the pulse introduces a bias that can be calibrated out.

To determine the photon statistical error in triggering on the first detected photon, the probability distribution of the time of detection of the first photon is derived. The detector is activated at time t_{gate} before t_0, the time of arrival of the pulse. Let $P_1(t)$ be the probability that the first photon arrives at time t, where $t_{\text{gate}} < t$. The mean number of detected photons expected in the interval from t_{gate} to a later time t_1 is

$$N(t_1) = \int_{t_{gate}}^{t_1} [B_u + S_u f(t - t_0)]dt \qquad (2)$$

where B_u is the background rate as seen by the uplink detector and S_u is the average number of detected signal photons per pulse. According to Poisson statistics, the probability of no photons being detected in the interval t_{gate} to t_1 is $e^{-N(t_1)}$ [16]. Then the integral of $P_1(t)$ is given by

$$\int_{t_{gate}}^{t_1} P_1(t)dt = 1 - e^{-N(t_1)} \qquad (3)$$

since the probability that the first photon arrived by t_1 is 1 minus the probability of no photons arriving before time t_1. Thus, $P_1(t)$ is given by

$$P_1(t) = \frac{\partial}{\partial t_1}\left(1 - e^{-N(t_1)}\right)\Big|_{t_1=t} \qquad (4)$$

The distribution $P_1(t)$ for the pulse shape given above and the uplink source and background rates given in Section II are shown in Fig. 5. The value of 1.2 μsec for $t_0 - t_{gate}$ has been used as an example. Nine-tenths percent of the distribution is due to the detection of a background photon before the pulse's arrival, while less than one-tenth percent is due to background photons being detected after no signal photons were detected. The remaining portion of the distribution is contained in a narrow peak with a mean of 2.1 nsec after the pulse arrival and a variance of (1.3 nsec)^2. This error will decrease as 1 over the square root of the number of measurements.

An important consequence of the first-photon detection algorithm is that the mean of the distribution depends on the power level of the received signal. If the number of detected photons per pulse changes by 10 percent, the distribution mean changes by 0.19 nsec. An estimate of the detected uplink power level can be made by analyzing the shape of the range distribution. Alternatives to doing this include shortening the uplink pulses, increasing the uplink laser power (which lowers the first photon shift with respect to power level change), or using a detector that can respond to more than one photon per uplink pulse.

At the ground station, the background rate is too high to simply detect the arrival of the first photon in the pulse. If the arrival of each photon could be time tagged, the maximum likelihood estimator could be used to estimate the pulse arrival time [17]. This procedure consists of maximizing the likelihood function L where

$$ln(L) = \sum_{k=1}^{n} ln\left(1 + f(t_k - \hat{t}\,)\frac{S_d}{B_d}\right) + \text{constant} \qquad (5)$$

where L is the a posteriori probability that a set of photon detection times t_k would result from a pulse arriving at time $\hat{t}$ given the background rate B_d, the expected number of photons in the pulse S_d, and the number n of photons detected in the gating window.

Although time tagging the arrival of each photon might be prohibitively difficult, Eq. (5) can be viewed as specifying an optimal pulse-shaping filter that has an impulse response proportional to $ln(1 + f(t)S_d/B_d)$. Detection of the peak time of a signal resulting from the detector filtered in this manner results in the maximum likelihood estimate of the pulse arrival time.

The statistical error in this estimation of the pulse arrival time is given by [17]

$$\frac{1}{\sigma_i^2} = \int \frac{\left[S_d \frac{\partial f(t)}{\partial t}\right]^2}{B_d + S_d f(t)} dt \qquad (6)$$

For the pulse shape given by Eq. (1) and the downlink signal and background rates given in Section II, this results in a single measurement timing error of 0.38 nsec.

V. Two-Way Range Error Budget

An analysis of detected minus predicted downlink pulse arrival times will reflect the combined effects of error in the time tagging of the uplink pulse, pulse detection jitter at the spacecraft, the downlink pulse arrival time error, and media delay errors. A small percentage of the differences will be spread out in a flat distribution due to background photons at the spacecraft, but 99 percent will be contained in a central peak. Differences outside this region of the distribution will be rejected in the analysis algorithm. Estimation of the center of this peak gives the spacecraft range and range error.

The components of the range error are listed in Table 3. Most of the error terms are random and will be reduced by the square root of the number of independent measurements. A separate measurement can be made for each uplink pulse occurring at the rate of 63 Hz. The largest random errors are due to photon statistics, so each measurement should be independent for those errors. The random instrumental errors may decrease more slowly since there may be correlation times longer than the interval between pulses. A bias error is introduced by the uncertainty in the troposphere delay derived from a model of the

troposphere and meteorological measurements. Other systematic errors include the ground station and transceiver instrumental biases. Instrumental errors are often given in time units. Since each time error contributes to the round-trip light time error, the range error is given by 1/2 the time error times the speed of light.

The start time error is associated with the time tagging of a sample of the uplink pulse. Relying on the experience of SLR measurements, it is supposed that the ground station systematic errors can be calibrated, by using a fixed reference or a well-known Earth-orbiting target, to a level below the troposphere bias error. The random detection errors come from the photon statistics, the PMT jitter, and the electronics jitter. The photon statistical error for the downlink reception derived in Section IV was 0.38 nsec, corresponding to 5.7 cm. For simplicity, the same value is used for the statistical component of the start time, although a stronger signal could be available. The PMT jitter for a 27 photoelectron event of 3.3 cm is taken from SLR measurements [10]. SLR systems may use a constant-fraction discriminator (CFD) following the PMT, with 0.2 cm error [10]. Here a value of 100 ps or 1.5 cm is included, representing 1/10 of the rise time of a shaping amplifier with 1 GHz bandwidth. The Hewlett Packard 5370A time interval unit has 100 ps accuracy [10], which is included here as the time-tag error.

The troposphere bias error is 1 cm at zenith for the SLR troposphere model [18]. The error will be different on the uplink and the downlink since several hours may separate them. Two-color measurements, possibly to an Earth-orbiting satellite, could be used to reduce the troposphere bias error if required. The fluctuating component of the troposphere is not very important for the two-way range measurements [19].

The photon statistical error at the spacecraft was discussed in Section IV. The 1.3-nsec error corresponds to a range error of 19 cm. The dark-count rate of the detector is ignored here since for some APDs the dark count can be made as low as 100/sec [15], which is lower than the uplink background rate. For the APD jitter, a value of 2.3 cm is used, which is the upper limit reported by Cova [14]. The output from an over-biased APD is assumed to be input to a CFD with a jitter of 0.2 cm. For ranging purposes, the discriminator output could be input directly to the laser modulator driver. The path time from the APD to the modulator driver can be quite short, perhaps 10–20 nsec. With proper temperature control and testing, the systematic delay error may be about 1 percent of the delay, corresponding to 3 cm. This value is entered as the electronics bias in Table 3. The error between triggering the laser modulator and the output light pulse depends on the laser design. Table 3 includes 23 cm (1.5 nsec) for the random error that is reported for a particular cavity-switched laser used for SLR [10]. The systematic error in the modulator driver can be calibrated out in SLR systems. For the spacecraft laser, an appropriate error will have to be derived from tests on candidate lasers. A value of 0.2 nsec (10 percent of the rise time) is included in Table 3 for the systematic laser error.

The downlink troposphere error is the same as the uplink error used above. The reception error is identical to the start-time error. The final error term in Table 3 is the clock error in keeping track of time between range transmission and reception. If the clock accuracy is 1 part in 10^{14} for the round-trip light time to Saturn, the range error is 10^{-14} times the Earth–Saturn distance (10 AU), which is 1.5 cm. Current hydrogen maser stabilities are better than this over the round-trip light time.[3]

Examination of the terms in Table 3 shows that the random errors drop below the few-centimeter level after a few seconds, assuming 63 independent measurements per second. Due to correlations between measurements, the instrumental errors may drop more slowly, but SLR results suggest that the random instrumental errors drop below the few-centimeter level in a reasonable time period (a few minutes). The systematic errors listed are at the few-centimeter level, indicating that an overall range error of less than 10 cm may be possible. This is comparable to the radio metric accuracy range one might achieve with a dual-frequency digitized system [20]. The most important unknown systematic effects are the detection algorithm and the laser modulator response. Tests of candidate spacecraft detectors and lasers will have to be performed to substantiate these numbers.

VI. Range-Rate Error Budget

The two-way range algorithm used above returns 63 range points/sec. During 1 sec, the spacecraft moves on the order of 1 km, so a model for the spacecraft motion is needed to fit expected versus predicted range points. The difference in the range points can be fit to the model to produce a measurement of range rate. At present, radio metric Doppler data are the simplest and most reliable data type used in spacecraft navigation. However, the orbit analysis utilizes a Doppler measurement by inte-

[3] B. Gutierrez-Lucas, *Deep Space Network/Flight Project Interface Design Handbook, Volume I: Existing Capabilities*, JPL Publication 810-5, Rev. D, Section FTS-10 (internal document), Jet Propulsion Laboratory, Pasadena, California, 1988.

grating the measured velocity over a time interval to produce a difference in range at the times of the interval end points.[4] If the optical telemetry system provides accurate range reliably, there may be no need to derive a range-rate measurement.

A least-squares fit to a sequence of (independent) range points equally spaced in time (at 63 Hz) can be used to find the error in range rate. The error components of each range point are given in Table 3. However, most of the range bias terms do not contribute to the range-rate error. The troposphere bias enters only due to a change in elevation of the spacecraft; the resulting error is small compared to the fluctuating troposphere. The systematic instrumental errors contribute depending on their time dependence. Here they are ignored, and only the random range errors are considered as contributing to the range-rate error. The range-rate error σ_v is given by [16]

$$\sigma_v^2 = \frac{12\sigma_r^2}{\Delta t^2 N^3} \tag{7}$$

where σ_r^2 is the root-sum-square (RSS) of the random range point errors, Δt is the time interval between range points, and $N \gg 1$ is the number of range points. The value of σ_r from Table 3 is 35 cm. With a rate of 63 Hz, the range-rate error is

$$\sigma_v = 1.5 \text{ cm/sec} \left(\frac{T}{\text{sec}}\right)^{-1.5} \tag{8}$$

The error derived from Eq. (8) is less than 1 mm/sec after a 30-sec integration time.

VII. DOR Error Budget

For a DOR measurement, two stations time tag the reception of downlink pulses. Differences in arrival times between the two stations are calculated on a pulse-by-pulse basis, and the distribution of these differences is analyzed to determine the geometric delay. The photon statistical error on this delay results purely from the downlink time-of-arrival error at both stations. Since synchronization with the uplink is not required, the measurement can be made directly on the telemetry signal.

The errors for the DOR measurement are listed in Table 4. Since the downlink pulse rate is 14,000/sec, the statistical terms are insignificant after a 1-sec integration time. The dominant errors are the clock offset error and the troposphere bias at each station. The global positioning system (GPS) is expected to be able to provide time calibration to 1 nsec between stations [21], which corresponds to the 30 cm in Table 4. The GPS is also expected to establish the geocentric station coordinates to a few centimeters [22]. One alternative to GPS clock synchronization is fiber-optic connections between stations. Extending the fibers to several-hundred-kilometer baselines with the necessary stability is not currently possible, but may be in the future. Another alternative means of clock synchronization is to use a combination of two-way and three-way range measurements, as discussed in the next section. This method may be able to establish intercomplex timing to the few-centimeter level.

Assuming that the clock synchronization problem can be solved, optical DOR measurements with few-centimeter accuracy should be possible. The angular accuracy will then be limited by the length of the available baseline. A 3-cm DOR measurement over a 300-km baseline corresponds to a 100-nrad angular measurement, which is not as accurate as present very-long baseline interferometry measurements. Better angular resolution requires either lower troposphere errors (assuming the clock error can be reduced to the troposphere level) or longer baselines. A 3000-km baseline could fit in the continental U.S. and potentially could provide measurements of about 10 nrad. Mutual station visibility would be a concern since optical measurements are more susceptible to poor weather conditions.

VIII. Clock Synchronization Using Range Measurements

This section describes a method for clock synchronization that is much like a differenced two-way range measurement. In fact, a differenced two-way range measurement can be derived from the same information. This particular set of measurements avoids the problem of short station overlap interval that occurs for a more straightforward differenced two-way range measurement [23]. This synchronization method is not unique to optical systems and may have been considered for radio metric systems. There are advantages to using this technique for the optical system: using the same instrumentation avoids local time-transfer stability problems, and this method could allow continual pointing at the spacecraft. The method is only outlined below. An error analysis would have to take into account the uncertainties in the motions of the Earth and the spacecraft, which are neglected here.

[4] T. Moyer, *Mathematical Formulation of the Double-Precision Orbit Determination Program (DPODP)*, JPL Technical Report 32-1527 (internal document), Jet Propulsion Laboratory, Pasadena, California, May 1971.

The sequence of measurements is shown in Fig. 6. Station A keeps time t, and station B keeps time t'. The difference in the clocks $\Delta\tau = t - t'$ is to be determined. Initially, the spacecraft sends a pulse that is detected by ground stations A and B. Each ground station records the time of pulse arrival (Fig. 6a). Station A records the time as t_1, while station B records the time as t'_1. After reception of the spacecraft pulse, each station sends up a pulse to the spacecraft, recording the times as t_2 and t'_2 (Fig. 6b). The time between downlink pulse reception and uplink transmission is taken to be short enough that the troposphere path delays at each station (p_A and p_B) are the same for the uplink and the downlink. The spacecraft returns each uplink pulse after an instrumental path delay s. The times of reception of these two returned pulses are recorded at station A as t_{3A} and t_{3B} (Fig. 6c). In the meantime, station B could have set.

For the present, the rotation of the Earth and the motion of the spacecraft are ignored. Then two time-tag differences, D_1 and D_2, are formed as

$$D_1 = t_{3A} - t_2 = 2r_A/c + p_A + p_3 + s \tag{9}$$

$$D_2 = t_{3B} - t'_2 = r_A/c + r_B/c + p_B + p_3 + s + \Delta\tau \tag{10}$$

where r_A is the distance from station A to the spacecraft, r_B is the distance from station B to the spacecraft, c is the speed of light, and p_3 is the media delay occurring on the final downlink from the spacecraft to station A (assumed to be the same for both downlinks). Taking the difference between these two values gives

$$D_3 = D_1 - D_2 = r_A/c - r_B/c + p_A - p_B - \Delta\tau \tag{11}$$

The regular DOR measurement gives

$$D = t_1 - t'_1 = r_A/c - r_B/c + p_A - p_B + \Delta\tau \tag{12}$$

Summing D and D_3 gives a result equivalent to a differenced two-way range measurement. Subtracting D_3 from D gives twice the clock offset $\Delta\tau$. Note that the tropospheric and geometric delays cancel out in the determination of the clock offset.

For this discussion of clock synchronization, instrumental delays were ignored along with Earth and spacecraft motions. However, since the method involves the same instrumentation as range measurements, the clock synchronization should approach the level of the range error. In fact, the error may be less since some of the error terms cancel out in the differences. Further analysis will have to be performed to make a better assessment of the clock synchronization accuracy. In practice, a method that involves three one-way light times to Saturn may not be desirable. It may be preferable to use a nearer spacecraft for clock synchronization and to perform DOR based on this clock synchronization on the deep-space probe.

IX. Discussion

This article has attempted to consider in some detail the potential accuracy of optical two-way range, range rate, and DOR. The ranging system assumed for this analysis relies almost entirely on telemetry system hardware and can be implemented with little impact on telemetry system design. This analysis shows that, given a reasonable telemetry system, the photon statistical errors on the resulting range and DOR observations fall below the few-centimeter level within a few seconds of observation time. Other sources of random error exist, but can be reduced to the centimeter level within a few minutes of observation.

For two-way range measurements, the most important error sources are the troposphere delay uncertainty and the error in the spacecraft delay calibration. Since the spacecraft transceiver is in an early stage of development, the bias errors cannot be properly characterized here. However, there is reason to believe that careful design and calibration of the transceiver will allow two-way range accuracy to be 10 cm or less. For DOR measurements, the major errors are due to the troposphere and the calibration of the offset between station clocks. A method of measuring this clock offset has been proposed. Further work is needed to determine if it is feasible. If this or some other method of clock synchronization allows station clocks to be calibrated to the 30-ps level, then angular accuracy using DOR will be limited by the few-centimeter troposphere error and the length of the available baseline.

Acknowledgments

The authors would like to thank J. M. Davidson and C. C. Chen for many suggestions that contributed to this work.

References

[1] J. Lesh, "Deep Space Optical Communications Development Program," *Proc. SPIE*, vol. 756, pp. 8–11, 1987.

[2] E. L. Kerr, "Strawman Optical Reception Development Antenna (SORDA)," *TDA Progress Report 42-87*, vol. July–September 1986, Jet Propulsion Laboratory, Pasadena, California, pp. 97–110, November 15, 1986.

[3] E. L. Kerr, "Fraunhofer Filters to Reduce Solar Background for Optical Communications," *TDA Progress Report 42-93*, vol. January–March 1988, Jet Propulsion Laboratory, Pasadena, California, pp. 48–55, May 15, 1988.

[4] K. Cowles, "A Visibility Characterization Program for Optical Communication Through the Atmosphere," *TDA Progress Report 42-97*, vol. January–March 1989, Jet Propulsion Laboratory, Pasadena, California, pp. 221–225, May 15, 1989.

[5] S. G. Lambert et al., *Design and Analysis Study of a Spacecraft Optical Transceiver Package*, McDonnell Douglas Report for JPL Contract No. 957061, St. Louis, Missouri, 1985.

[6] W. K. Marshall and B. D. Burk, "Received Optical Power Calculations for Optical Communications Link Performance Analysis," *TDA Progress Report 42-87*, vol. July–September 1986, Jet Propulsion Laboratory, Pasadena, California, pp. 32–40, November 15, 1986.

[7] *Parametric Analysis of Microwave and Laser Systems for Communication and Tracking*, Hughes Aircraft Co., Report No. P67-09, 1966. Referenced in W. K. Pratt, *Laser Communications Systems*, New York: John Wiley and Sons, 1969.

[8] R. C. Ramsey, "Spectral Irradiance from Stars and Planets, above the Atmosphere, from 0.1 to 100.0 Microns," *Applied Optics*, vol. 1, pp. 465–472, 1962.

[9] I. L. Goldberg, *Radiation from Planet Earth*, U.S. Army Signal and Research Development Lab. Report 2231, AD-266-790, September 1961.

[10] J. J. Degnan, "Satellite Laser Ranging: Current Status and Future Prospects," *IEEE Trans. Geoscience and Remote Sens.*, vol. GE-23, pp. 398–413, 1985.

[11] M. Sasaki, "Satellite Laser Ranging at the Simosato Hydrographics Observatory and Its Preliminary Results," *J. Geodetic Soc. of Japan*, vol. 30, pp. 29–40, 1984.

[12] S. N. Del'nova, V. I. Markov, N. K. Smirnov, and N. A. Krivokul'skaya, "Characteristics of Multialkali Photomultipliers with Total Internal Reflection," *Instruments and Experimental Techniques*, vol. 24, pp. 1269–1272, 1981.

[13] Data sheet for RCA C30954E photodiode, RCA Corporation, Ste. Anne de Bellevue, Quebec, Canada H9X-3L3, 1979.

[14] S. Cova, A. Longoni, and A. Andreoni, "Towards Picosecond Resolution with Single-Photon Avalanche Diodes," *Rev. Sci. Inst.*, vol. 52, pp. 408–412, 1981.

[15] R. G. W. Brown, K. D. Ridley, and J. G. Rarity, "Characterization of Silicon Avalanche Photodiodes for Photon Correlation Measurements. 1: Passive Quenching," *Applied Optics*, vol. 25, pp. 4122–4126, 1986.

[16] P. R. Bevington, *Data Reduction and Error Analysis for the Physical Sciences*, New York: McGraw-Hill, 1969.

[17] M. Fisz, *Probability Theory and Mathematical Statistics*, New York: John Wiley and Sons, 1963.

[18] J. B. Abshire and C. S. Gardner, "Atmospheric Refractivity Corrections in Satellite Laser Ranging," *IEEE Trans. Geoscience and Remote Sens.*, vol. GE-23, pp. 414–425, 1985.

[19] C. S. Gardner, "Effects of Random Path Fluctuations on the Accuracy of Laser Ranging Systems," *Applied Optics*, vol. 15, pp. 2539–2545, 1976.

[20] L. E. Young, "Improved Ranging Systems," Workshop on Relativity and Gravitational Experiments, Annapolis, Maryland, June 1988 (unpublished).

[21] G. Blewitt, "Carrier Phase Ambiguity Resolution for the Global Positioning System Applied to Geodetic Baselines up to 2000 km," *J. Geophysical Research*, vol. 94, pp. 10187–10203, 1989.

[22] R. P. Mala and S.-C. Wu, "Deriving a unique reference frame for GPS measurements," *IEEE Position Location and Navigation Symposium Record*, pp. 177–184, 1988.

[23] T. H. Taylor, J. K. Campbell, R. A. Jacobson, B. Moultrie, R. A. Nichols, and J. E. Riedel, "Performance of Differenced Range Data Types in Voyager Navigation," *Journal of Guidance, Control, and Dynamics*, vol. 7, pp. 301–306, May–June 1984.

Table 1. Downlink telemetry system parameters

Parameter	Value
Wavelength, μm	0.532
Average laser output power, W	2.000
Diameter of XMTR aperture, m	0.300
Obscuration diameter of XMTR, m	0.060
Transmitter optics efficiency	0.650
XMTR pointing bias error, μrad	0.300
XMTR rms pointing jitter, μrad	0.300
Diameter of RCVR aperture, m	10.000
Obscuration diameter of RCVR, m	4.300
Receiver optics efficiency	0.380
Narrowband filter transmission factor	0.400
Filter spectral bandwidth, Å	0.300
Detector quantum efficiency	0.350
Detector field-of-view angle, μrad	100.000
Background, W/μm m^2 str	100.000
Number of slots per PPM symbol	256.000
Data rate, kbits/sec	114.350
Pulse rate, number of symbols/sec	14294.000
Slot width, nsec	10.000
Distance between XMTR and RCVR, AU	10.000
Atmospheric transmission factor	0.500
Detected signal photons/pulse	27.600
Detected background photons/slot	2.150

Table 2. Uplink telemetry system parameters

Parameter	Value
Wavelength, μm	1.06
Average laser output power, W	110.00
Diameter of XMTR aperture, m	0.05
Obscuration diameter of XMTR, m	0.00
Transmitter optics efficiency	0.65
XMTR pointing bias error, μrad	2.50
XMTR rms pointing jitter, μrad	5.00
Diameter of RCVR aperture, m	0.30
Obscuration diameter of RCVR, m	0.06
Receiver optics efficiency	0.65
Narrowband filter transmission factor	0.80
Filter spectral bandwidth, Å	10.00
Detector quantum efficiency	0.10
Detector field-of-view angle, μrad	8.62
Background, W/μm m^2 str	6.50
Number of slots/PPM symbol	256.00
Data rate, kbits/sec	0.50
Pulse rate, number of symbols/sec	63.00
Slot width, nsec	10.00
Distance between XMTR and RCVR, AU	10.00
Atmospheric transmission factor	0.50
Detected signal photons/pulse	5.47
Detected background photons/slot	7.14×10^{-5}

Table 3. Two-way range error budget[a]

Error source	Value
Start time error	
Instrumental biases	<3.0 cm
Detection statistics	5.7 cm/$\sqrt{N}$
PMT jitter	3.3 cm/$\sqrt{N}$
Amplifier jitter	1.5 cm/$\sqrt{N}$
Time-tag error	1.5 cm/$\sqrt{N}$
Uplink troposphere error	
Bias component	1.0 cm/sin E
Fluctuating component	0.1 cm/$\sqrt{(N \sin E)}$
Spacecraft turnaround error	
Detection statistics	19.0 cm/$\sqrt{N}$
APD jitter	2.3 cm/$\sqrt{N}$
Discriminator jitter	1.5 cm/$\sqrt{N}$
Electronics bias	3.0 cm
Modulator random error	23.0 cm/$\sqrt{N}$
Modulator bias error	7.5 cm
Downlink troposphere error	
Bias component	1.0 cm/sin E
Fluctuating component	0.1 cm/$\sqrt{(N \sin E)}$
Reception time error	
Instrumental biases	<3.0 cm
Detection statistics	5.7 cm/$\sqrt{N}$
Detector jitter	3.3 cm/$\sqrt{N}$
Amplifier jitter	1.5 cm/$\sqrt{N}$
Time-tag error	1.5 cm/$\sqrt{N}$
Clock rate error	<1.5 cm

[a] N is the number of independent measurements, which may be different for instrumental and troposphere errors (which may have correlated noise) and detection errors (which are purely statistical). The elevation angle is E.

Table 4. DOR error budget[a]

Error source	Value
Station 1 detection error	
Instrumental bias	<3.0 cm
Detection statistics	5.7 cm/$\sqrt{N}$
Detector jitter	3.3 cm/$\sqrt{N}$
Amplifier jitter	1.5 cm/$\sqrt{N}$
Time-tag error	1.5 cm/$\sqrt{N}$
Troposphere bias	1.0 cm/sin E
Troposphere fluctuation	0.1 cm/$\sqrt{(N \sin E)}$
Station 2 detection error	
Instrumental bias	<3.0 cm
Detection statistics	5.7 cm/$\sqrt{N}$
Detector jitter	3.3 cm/$\sqrt{N}$
Amplifier jitter	1.5 cm/$\sqrt{N}$
Time-tag error	1.5 cm/$\sqrt{N}$
Troposphere bias	1.0 cm/sin E
Troposphere fluctuation	0.1 cm/$\sqrt{(N \sin E)}$
Clock offset error	30.0 cm

[a] N is the number of independent measurements, and E is the elevation angle.

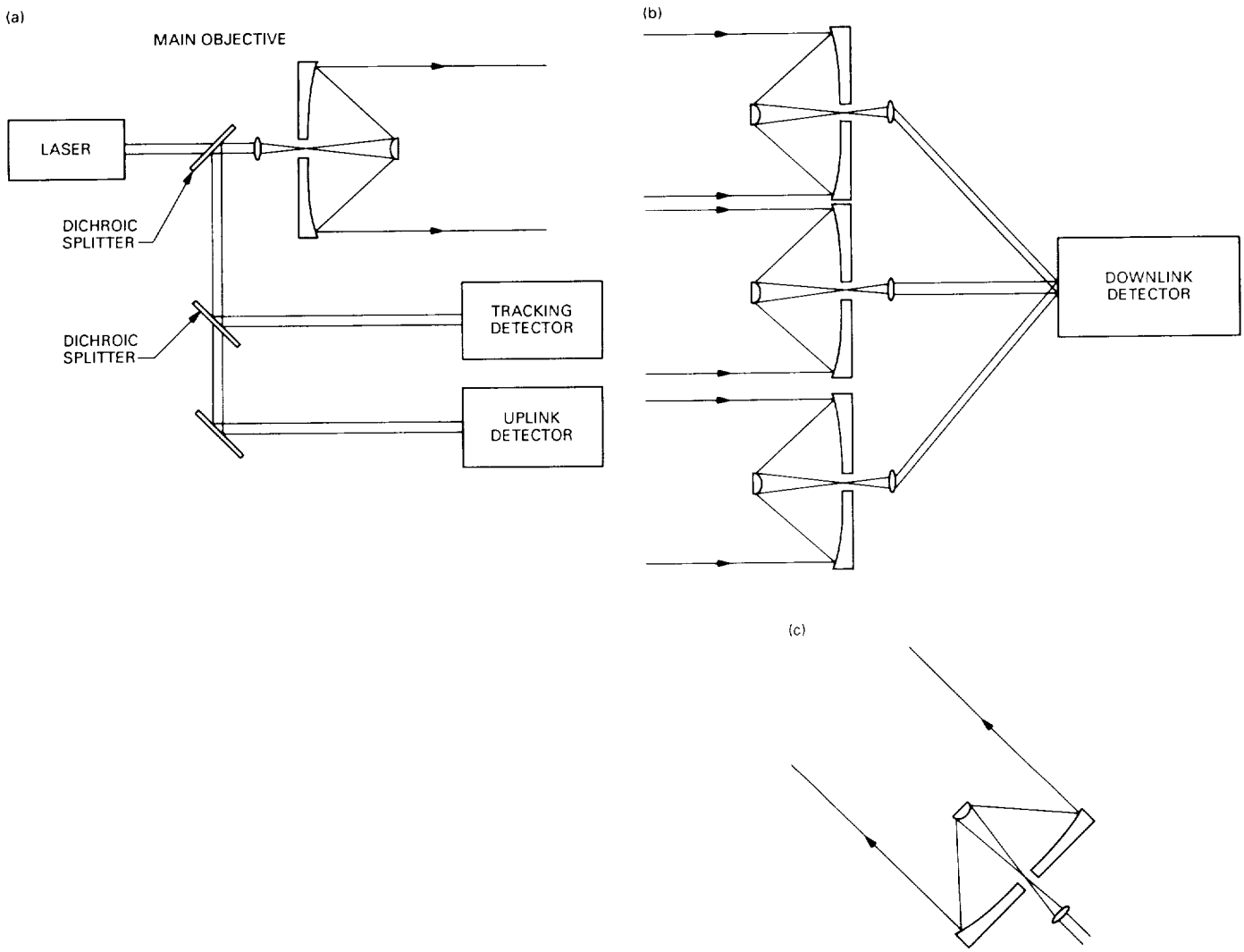


Fig. 1. Simplified optical diagram for PPM telemetry system: (a) spacecraft transceiver; (b) ground receiver; and (c) uplink transmitter.

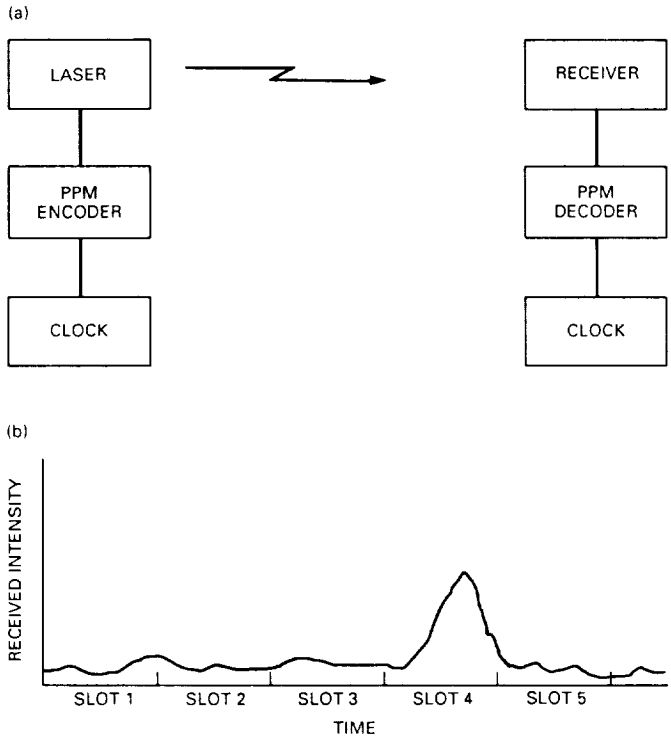


Fig. 2. PPM modulation method schematic: (a) telemetry system; and (b) received intensity diagram.

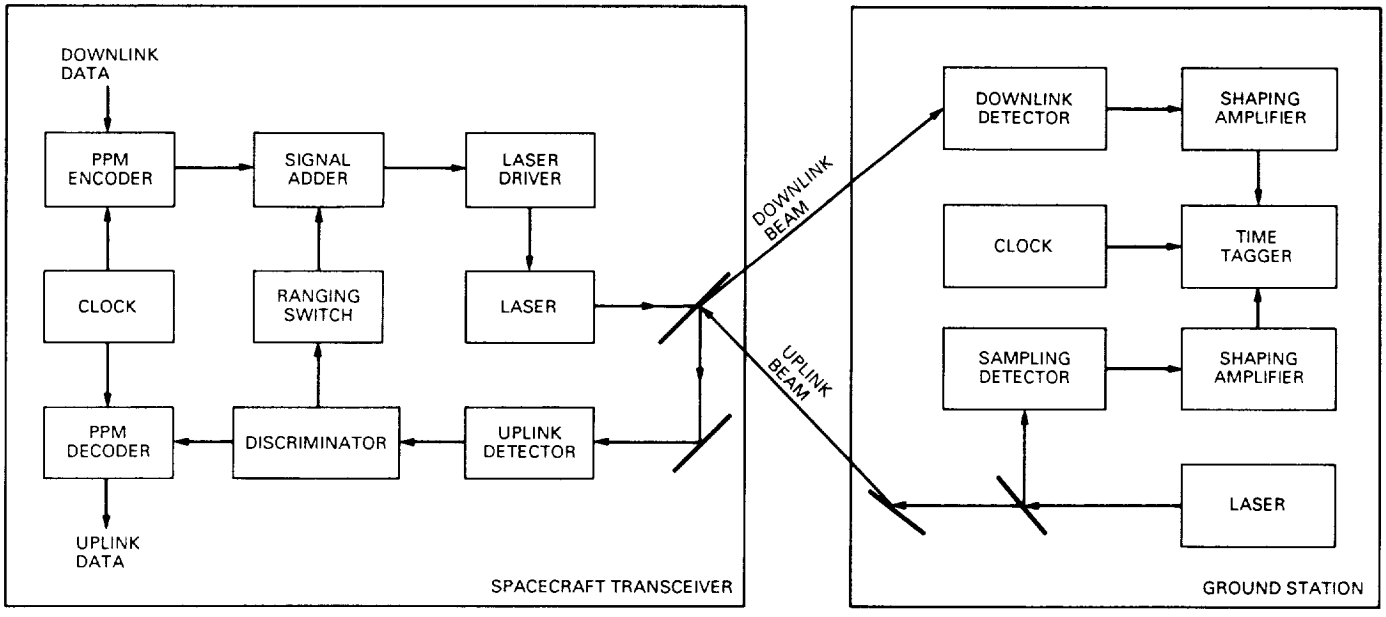


Fig. 3. Two-way range measurement.

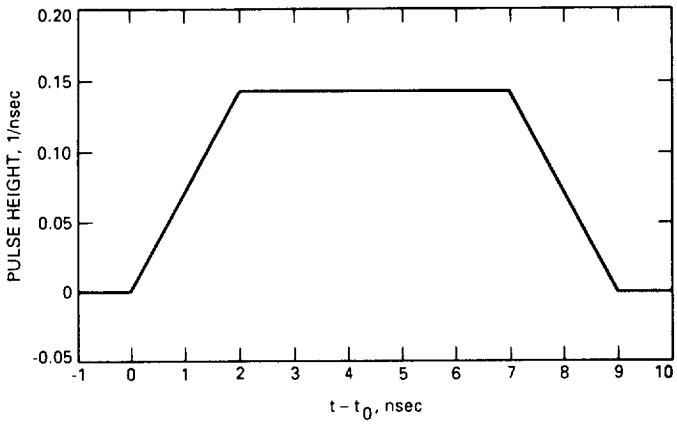


Fig. 4. Normalized laser output pulse shape.

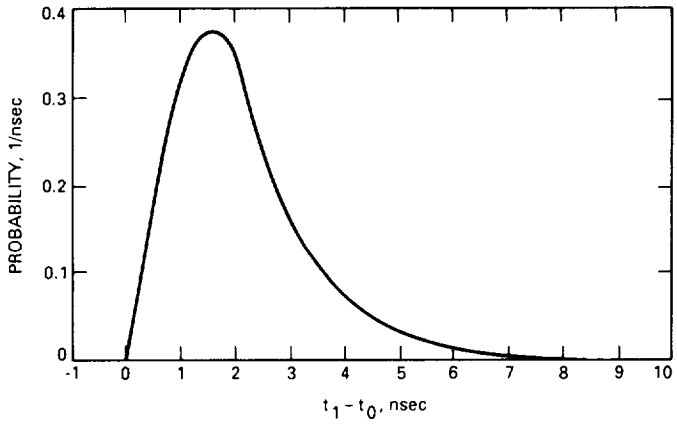


Fig. 5. Normalized probability distribution for detection of first photon.

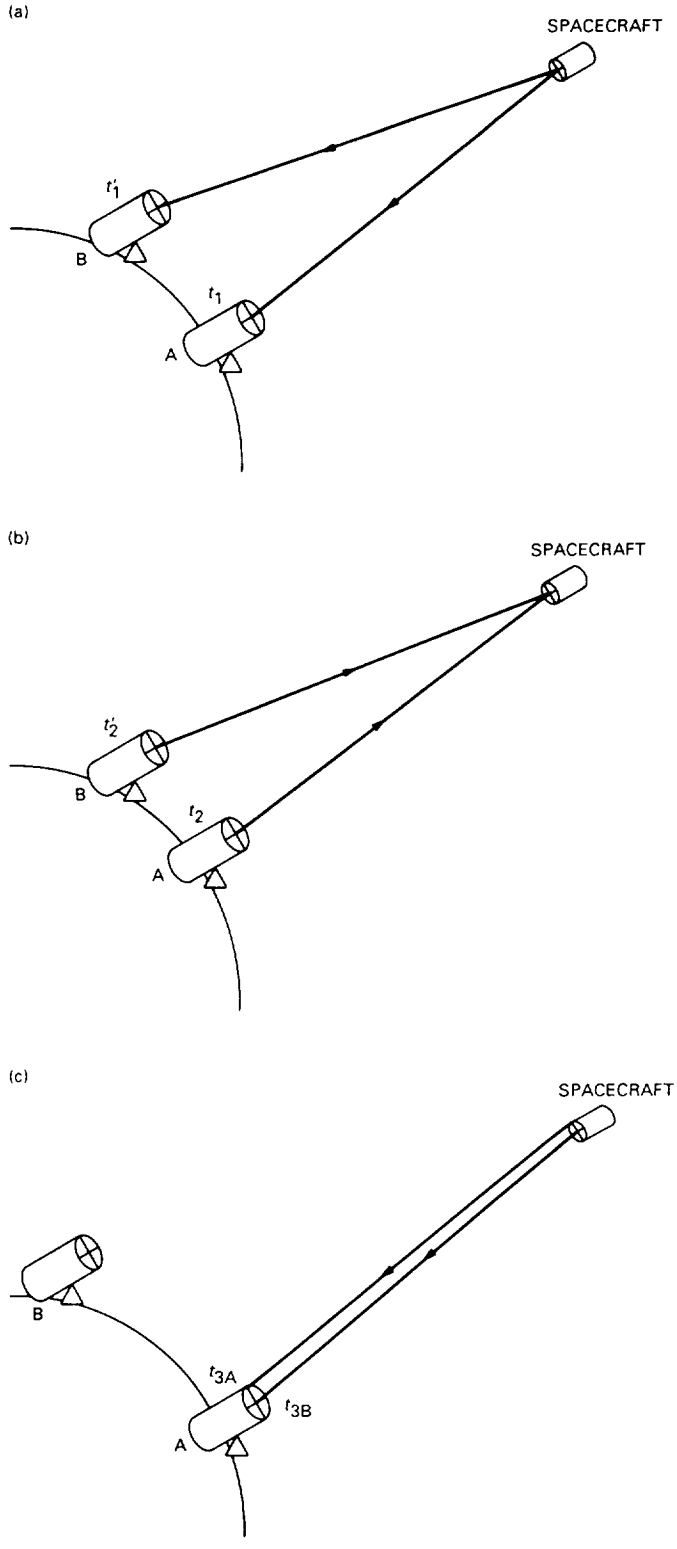


Fig. 6. Sequence of events (time tags) for clock synchronization: (a) initial downlink; (b) both stations send uplink; and (c) station A records time of return for both uplinks.

N90-26228

TDA Progress Report 42-101 May 15, 1990

Initial Results on Fault Diagnosis of DSN Antenna Control Assemblies Using Pattern Recognition Techniques

P. Smyth
Communications Systems Research Section

J. Mellstrom
Ground Antennas and Facilities Engineering Section

This article describes initial results obtained from an investigation of using pattern recognition techniques for identifying fault modes in the Deep Space Network (DSN) 70-m antenna control assembly. It describes the overall background to the problem, outlining the motivation and potential benefits of this approach. In particular, it describes an experiment in which fault modes were introduced into a state-space simulation of the antenna control loops. By training a multilayer feed-forward neural network on the simulated sensor output, classification rates of over 95 percent were achieved with a false alarm rate of zero on unseen test data. It concludes that although the neural classifier has certain practical limitations at present, it also has considerable potential for problems of this nature.

I. Background and Motivation

Very accurate and precise pointing is a characteristic of the Deep Space Network (DSN) antennas. Some recent pointing problems have led to an interest in investigating automated methods of fault detection and identification within the antenna control assembly (ACA). The ACA for the 70-m antenna is a two-axis (azimuth and elevation) digital control system. In its simplest configuration, each axis controller consists of several servo-valve-controlled hydraulic motors, countertorque motors, gears, analog electronics (power amplifiers, analog compensation, filters), tachometers, an encoder, a digital computer, and various digital interfaces. It gets more complicated if the antenna is operated in precision mode, in which the 70-m antenna position is slaved to that of a precision pointing mechanism called the master equatorial. Clearly, there are many hydraulic, electrical, mechanical, hydromechanical, and electromechanical components that may be subject to wear, degradation, and aging. Identifying the source of pointing degradation within the ACA is not a trivial problem.

Furthermore, although excellent performance of the ACA is critical for good antenna pointing, it is only a part of the complex interaction of people, procedures, and equipment that affects pointing. To track down a pointing problem through all this can sometimes be a very difficult task. As a result, component degradation often goes unnoticed, resulting in suboptimal system performance. No fault identification action is taken until the X-band pointing requirements are no longer met or catastrophic failure occurs. It was recently reported that the antenna subsystem functional requirements for test or diagnostic capabilities have not been fully met [1].

According to the *Deep Space Network System Functional Requirements,* network equipment shall be designed to have a service life of at least 10 years.[1] However, the *Deep Space Network Long Range Plan* indicates that existing antennas will be operating well into the 21st century with greater availability (99 percent), lower crew sizes (35 percent of 1992 levels), and at much higher frequencies (Ka-band).[2] The implications of these goals are that (1) the equipment related to antenna pointing must always operate at near optimal performance levels, (2) scheduled maintenance times must be reduced, (3) equipment failures must be eliminated, and (4) these must be accomplished with a reduction in the personnel available for monitoring, diagnostics, repair, and maintenance. Recognizing this, it was identified in the *Deep Space Network Long Range Plan* that over the next 20–30 years the DSN must develop computer-aided maintenance and expert systems capability.

The objective of maintenance is to keep equipment operating in a nominal condition. Historically, maintenance has meant the periodic inspection, replacement, and rebuilding of equipment that is critical to system performance. However, this strategy is expensive because it results in downtime to replace equipment that may be operating nominally, and it still does not guarantee against catastrophic failure. A more effective strategy is to schedule repairs based on the operating condition of the system. An automatic monitoring system that can detect deviations from the nominal system state and identify the source of the deviation is a more desirable method of scheduling maintenance, maintaining optimal performance, and avoiding catastrophic failure.

As indicated above, a suitable system for an investigation of automated fault detection and identification is the ACA of the 70-m antenna mechanical subsystem (ANT). More fully automated fault detection and identification clearly would assist current DSN operations and is absolutely necessary for future operations.

II. ACA System Model and Fault Simulation

For this investigation, the 70-m antenna azimuth drive was simulated operating in nominal condition and four fault conditions. These simulations were repeated at three different angular velocities: 0.0, 4.0, and 40.0 mdeg/sec. The rates were chosen to emulate the range of rates encountered in the azimuth drive during a spacecraft track. At low elevation angles, the azimuth rate is very small. As elevation angle increases, azimuth rate also increases. The azimuth drive of the 70-m antenna was simulated on MatrixX simulation software. MatrixX is a commercial engineering analysis and control design software package. It incorporates most of the matrix analysis functions in EISPACK and LINPACK. It also has a graphical environment for simulation of discrete and continuous models.

The model described in this article is similar to that reported in [2] and is very briefly described here. For detailed information, readers are referred to the original paper. A block diagram of the model is shown in Fig. 1. The model was a hybrid continuous and discrete time model. The antenna servo controller (ASC) in this simulation consisted of a discrete-time-state feedback control algorithm and a steady-state Kalman filter. Its inputs were the commanded position and position feedback (measured and quantized by a 20-bit encoder). The ASC outputs were the position estimate, rate command, and quantized rate command, a 12-bit digital-to-analog (D/A) conversion labelled DAC Out. The rate loop amplifier represented all the analog electronics, with inputs of rate command and tachometer voltage feedback, and valve current as output. The tachometer voltage feedback represented four tachometers, one for each drive motor. The valve converted an electrical signal to hydraulic flow. Its inputs and outputs were valve (coil) current and valve (hydraulic) flow, respectively. The motor model represented four hydraulic motors. The inputs were valve flow and load torque. The outputs were motor rate, tachometer rate, and differential hydraulic pressure. The structure model was a seventh-order model incorporating the dominant modes of the structure and gearboxes. Its inputs were motor rate and wind disturbance torque. Its outputs were structure position referenced at the encoder and load torque on the axis.

The model incorporates the nonlinearities of static and coulomb friction in the motors, deadband and hysteresis in the valve, position quantization (encoder), and control effort quantization (D/A conversion). At low antenna angular velocities, these nonlinearities are significant and make system analysis very difficult. Since the nonlinearities are discontinuous, it is not possible to get a linear approximation that is valid at low angular velocities. Unfortunately, almost all operation of DSN antennas is at angular velocities from 0.0 to 5.0 mdeg/sec.

[1] *Deep Space Network System Functional Requirements General Requirements and Policies Through 1988,* JPL Document 820-20, vol. 1, Rev. A (internal document), Jet Propulsion Laboratory, Pasadena, California, March 1, 1988.

[2] *Deep Space Network Long Range Plan,* JPL Document 801-1 (internal document), Jet Propulsion Laboratory, Pasadena, California, March 15, 1989.

The faults simulated for this investigation were faults that have actually occurred at one or more of the 70-m antennas. When these faults have occured at the antenna, they have been severe enough to affect antenna pointing, yet subtle enough to be very difficult to diagnose. Part of the difficulty is due to the effect of nonlinearities at operational velocities. Signals obtained at the antenna have such a complex structure in the time domain that it is often very difficult for operations personnel or an engineer to diagnose the fault.

The faults chosen for this investigation, how they were simulated, and their relationships to the actual antenna are described below:

(1) Tachometer failure: This corresponds to a break in a tachometer winding or another electrical connection. There are tachometers associated with each drive motor. Voltage ripple, inherent in any tachometer and/or caused by torque ripple of the motor, is reduced and some failure robustness is achieved by averaging the tachometers. The loss of one tachometer reduces the gain and bandwidth of the rate loop. As a result, the servo will not follow a command as quickly, and disturbances will not be rejected as well. This was simulated by reducing the tachometer voltage by one-fourth.

(2) Increased valve deadband: This corresponds to wear of the surfaces in the hydraulic valve. Very precise machining is required to manufacture a low-deadband valve. Flow of the hydraulic fluid wears these surfaces, especially if the fluid is carrying particulates. Greater deadband increases the limit cycle behavior of the servo. A limit cycle may be unavoidable even in the nominal case, but it reduces pointing performance and increases drive mechanical wear. This was simulated by increasing the deadband in the valve by a factor of 2.

(3) Increased static friction: The significant sources of static friction in the ACA are the valve, the motor, and the gear reducers. It is also caused by wear. The result of increased friction is increased limit cycling. For this investigation, static friction was simulated in the motor. The fault condition corresponded to increasing the static friction by a factor of 2.

(4) Tachometer noise: Tachometer noise corresponds to brush wear and/or bearing wear. It was simulated as additive Gaussian noise with zero mean and standard deviation that increased with velocity.

III. Classification Experiment

As described above, the data for the classification experiment were generated by introducing fault modes into the control-loop simulation model. In addition, data were obtained for normal operation in the absence of any of the four fault modes. Hence, there are in effect five classes. For each class, the system was simulated at three angular velocities, namely, 0, 4, and 40 mdeg/sec over a time span of 20 seconds for each rate, with a sampling resolution of 200 Hz. This yielded $4000 \times 3 \times 5 = 60,000$ data vectors in total. Each data vector has eight components, corresponding to eight system outputs or observable sensors in the simulator. These outputs are antenna rate, differential pressure, valve flow, encoder, rate command, position estimate, valve current, and tachometer voltage. Figure 2 shows a plot of these outputs over 20 seconds at a rate of 4 mdeg/sec under normal operation (no faults). As mentioned above, this corresponds to 4000 data points for each component of the output vector (for a particular class at a given rate). Figures 3, 4, 5, and 6 show output plots at the same rate for the four different faults, namely, tachometer failure, increased deadband, increased static friction, and tachometer noise, in that order. Clearly, the problem of discriminating the individual fault conditions from normal behavior is nontrivial, based on visual inspection of the waveforms. The problem is as follows: given part of the data, say the first 2000 points, derive a classification algorithm that can classify as accurately as possible the remainder of the data.

IV. Feature Generation

Although in principle it would be possible to use the 60,000 input vectors directly as input to a classifier, it is generally considered in the statistical pattern-recognition literature to be a better idea to generate "features" by preprocessing the data. Essentially, the aim is to transform the data into a feature domain, where the features possess greater discriminatory power than the values of the raw data do. Heuristic motivation for this technique comes from the observation that biological systems such as the human visual system use this approach. In addition, there are rigorous statistical arguments that show it is important to make as efficient use of the available data as possible, and transformation to a good feature domain promotes such efficiency. As an example, it might be desirable to transform the data to the frequency domain for a more efficient representation.

Although automated feature discovery systems exist (based on expansions such as the Karhunen-Loeve transform), by and large the technique that works best in prac-

tice is manual feature generation. In effect, by defining features thought to possess useful discriminatory power, the classifier is helped up the learning curve; in a statistical sense, this is equivalent to a prior bias on the hypothesis space.

For this particular problem, the initial study chose to define simple time-domain features such as the mean and the range. The motivation for this choice was to investigate how well one could classify the data by only using these very simple statistical indicators; as shall be seen, one can do surprisingly well. An arbitrary choice was made of a window size of 128 over which these features were estimated, which resulted in a reduction of the number of input data vectors from 60,000 to 465. Another advantage of the simple estimators over more sophisticated techniques was their robustness over small sample sizes; i.e., the variance of these estimators could be expected to be lower than Fourier-based estimators for the same amount of data. In turn, more robust estimators would lead to better generalization performance on unseen data.

For each of the differential pressure, valve current, and tachometer voltage outputs, the range, mean, and variance in each window were estimated, giving nine features. The slope of the encoder and position estimate and the mean of the rate command were also estimated, giving a total of 12 features in all. The data from the antenna rate and valve flow outputs were not used in this experiment, as they are not directly measurable in the actual physical control assembly in the stations.

It is instructive to view the discriminatory power of some of these features. In Fig. 7(a), the normalized values of the tachometer-voltage-mean feature as a function of the class values are plotted. The class numbers correspond to the four fault conditions described earlier, with class 5 being the absence of any fault, or normal conditions. It can be seen that this feature contains some discriminatory power for classes 1 and 3, but otherwise not much class information can be distinguished. Figure 7(b) shows a similar plot of the valve-current variance where class 4 (increased tachometer noise) is the only distinguishable class; naturally, the variance-based features possess the capability of discriminating such a class. In general, most of the other features possess even less discriminatory power on their own. Hence, whatever discriminatory power these features possess as a group will only be discovered by a classifier that can effectively combine these features into composite functions; i.e., it would be expected that, say, a simple linear discriminant classifier would not do very well on this problem.

V. Choosing a Classifier

In pattern recognition, there is a wide variety of different algorithms available for generating classification models from data. Among the most widely used methods are nearest-neighbor classifiers, Bayesian models, and, more recently, multilayer feed-forward perceptrons (neural networks). What is perhaps not so well known is that many of these schemes perform equally well across a broad range of problems if evaluated in terms of classification-error performance alone. In other words, the difference between these various schemes in terms of classification accuracy has been empirically shown to be often minimal [3,4]. What often matters then in choosing a classifier technique are other considerations, such as the efficiency of the learning algorithm, ease of implementation, amount of prior knowledge required, etc. For example, the nearest-neighbor classifier is easy to use, but can be very inefficient in terms of memory requirements to implement. The Bayesian approach, for problems involving nondiscrete or continuous-valued data in particular, often requires significant prior knowledge regarding the distribution of the data; for the antenna problem, since the plant under observation is essentially nonlinear, little can be said a priori regarding the distribution of parameters such as the range and variance of the outputs.

Hence, for the initial study at least, a neural network classifier was chosen. The classification of relatively "low-level" time varying waveforms, where there was little prior knowledge about the underlying form of the probability density functions, was considered a suitable problem for the neural approach [5]. Problems that appear to be similar in nature to human perceptual tasks intuitively seem to be typically well matched to connectionist models. In addition, a public-domain algorithm coded in C for exactly this purpose was available (and will be described in more detail in the next section), making it very easy to experiment with the neural approach; i.e., no coding effort was required. It is also worth noting at this point that in this small-scale initial study, the primary interest was in getting an idea of the scale of the problem; e.g., is it possible to classify these waveforms using very simple features?

VI. Conjugate-Gradient Neural Learning Algorithms

The well-known backpropagation algorithm [6] for training multilayer feed-forward neural networks is somewhat wasteful of computational resources, and it is relatively well known that practitioners resort to various unpublished "tricks" to speed up the algorithm in practice. Hence, until recently, although impressive results had been reported in the literature from using this algorithm, it was

not practical to experiment with it without a significant investment in initial effort. However, recent results have taken a broader view of the algorithm, and by utilizing prior work in conventional optimization theory and practice, more standard and conventional approaches to backpropagation have developed. In particular, the algorithm used in this experiment is described by Barnard and Cole [7], which in turn is an application of a conjugate-gradient optimization algorithm of Powell [8]. The algorithm will not be described in detail here except to note a few practical points; the interested reader is referred to the original papers. As described by Barnard and Cole, the conjugate-gradient algorithm is usually able to locate the minimum of a multivariate function much faster than a pure gradient-descent technique. In practice, it was found that the algorithm performs consistently well on a variety of classification problems. Of course, with these techniques there is no guarantee of convergence to the global optimum, but again, in practice the algorithm has consistently generated near-optimal solutions.

A factor that is often glossed over in the literature is the choice of neural architecture. This prior choice of a network model is suboptimal in general, and one would prefer to have the algorithm automatically select the appropriate size architecture from the data. A number of research groups are pursuing this goal, but as yet there are no widely accepted robust algorithms available. Hence, in practice, one must choose a network architecture for the problem at hand, i.e., the number of "hidden" layers and number of "hidden units" at each layer. For this experiment, attention is restricted to three-layer models (i.e., one hidden layer). The Appendix describes in more detail the exact nature of the three-layer networks under consideration. Note that there are many other variations of neural network architectures, such as recurrent networks and Boltzmann machines. The three-layer network is the simplest of these models with universal approximation capabilities; i.e., in principle, it can approximate any function, given enough hidden units.

VII. Results of the Classification Experiment

As described earlier, the original simulator output data were preprocessed into 465 feature vectors, with 12 feature components in each vector. This gave 93 data vectors per class. On closer inspection of the data, it was decided that the transient portions of the waveforms could safely be eliminated from consideration. In practice, one would in effect implement a hierarchical classifier, where the data were initially classified as either transient or nontransient. In addition, it was decided that the low-rate case of rate = 0 was a special case, and since large portions of the waveform at this rate contained no information at all, including them in the experiment would not yield meaningful results. Hence, only the nontransient, nonzero-rate data were looked at. This resulted in further data reduction to 260 data vectors.

The experiment consisted of generating two disjoint (roughly equally sized) subsets of the original data, calling one the training set, the other the test set. The conjugate-gradient algorithm was run on the training set, and the resulting three-layer network was used to classify the data in the "unseen" test set. After eight runs of this nature on randomly chosen training and testing disjoint subsets of roughly equal size, the resulting mean classification accuracy was 95.1 percent with almost no variance. Figure 8 shows a so-called "confusion matrix" for one of the networks. The left-hand column denotes the true value of the class; the top row denotes the network's estimate. Hence, a perfect network would have all of its entries in the diagonal; an entry in location i, j indicates the number of test points of class i that were classified as j. Remembering that class 5 is normal behavior, it can be seen that the false alarm rate is zero; i.e., no normal windows are incorrectly classified as a fault condition. In addition, it can be seen that the network has trouble classifying only one class, namely, tachometer failure. The network tends to confuse it with either increased static friction (class 3) or normal mode (class 5). Apart from this class, it performs perfectly.

The results of this simple classification experiment are surprisingly good in the context of pattern recognition. In general, for a given set of features and a class variable, there is a theoretical upper bound (the Bayes optimal rate) on the classification accuracy that is attainable. For example, if the features are completely independent of the class variable, then the optimal strategy is always to choose the most likely class and, hence, the optimal rate is the prior probability of this class. Since in practice the upper bound on performance is often considerably less than 100 percent, a figure of 95 percent is quite respectable for an initial experiment.

VIII. Conclusions

The result of the neural network classification experiment is promising. Even though the faults were only single-mode failures of a simple nature, and only on a simulator, one has reason to believe that the real problem may be amenable to these techniques when one takes into account that the classifier as implemented did not use any of a wide variety of additional information that was available. For example, by treating the data vectors (windows)

as random samples, all sequential information in the waveform was ignored; i.e., in practice one would use memory in the classifier to weight the current classification decision based on previous decisions (effectively using a "smoothness" assumption on the occurrence of faults over time).

Of course, the neural network approach has inherent drawbacks also. It may be difficult to ascertain which features, or combinations of features, are contributing most to the classification accuracy, although for three-layer networks there exist visual analysis techniques for this purpose. In addition, training the network on a Sun-3/260 typically consumed about 1.5 hours of computation (with no other processes running except for Unix overhead), while the training data correspond to only 10 seconds of actual elapsed (simulated) time. Hence, it is difficult to see the implementation of actual, practical neural network algorithms, which learn in real time in the field, until very large scale integrated (VLSI) neural hardware becomes available.

IX. Future Work

In general, the problem of real-time predictive and diagnostic monitoring of the antenna control assembly is quite a challenging one. It would be naive to expect that a simple "static" classifier, such as that presented in this paper, would be robust enough to work in the field. In particular, the assumption that there are clearly defined fault classes will probably not hold up in practice, so that approaches such as unsupervised classification techniques (in which the training data have no class labels) will need to be considered. In addition, there are a number of problems, both at the theoretical and implementation levels, with developing an autonomous monitoring system. These include, for example, issues of memory (when should the system discard old data?), validation (how can one verify or quantify the operation of such a system in a nonintrusive manner?), etc. Once these algorithmic issues are dealt with, it may be possible to develop dedicated VLSI hardware specifically for antenna-control-assembly monitoring.

It is proposed that these problems be addressed by using a phased approach, applying existing technologies to prototype systems in-lab, and experimenting with DSS 13 facilities. In this manner, the feasibility of these techniques can be proven without incurring significant risk, and the prototype can be gradually transferred to the DSN operations environment in a relatively low-cost manner.

References

[1] W. O. Wood, "Report on the Maintenance of RF Performance at the DSN 70-Meter Antennas," prepared for the Bendix Field Engineering Corporation, Columbia, Maryland, September 30, 1989.

[2] R. E. Hill, "Dynamic Models for Simulation of the 70-M Antenna Axis Servos," *TDA Progress Report 42-95*, vol. July–September 1988, Jet Propulsion Laboratory, Pasadena, California, pp. 32–50, November 15, 1988.

[3] Y. Lee and R. P. Lippmann, "Practical characteristics of neural network and conventional pattern classifiers on artificial and speech problems," presented at the 1989 IEEE Neural Information Processing Conference, Denver, Colorado, December 1989.

[4] S. M. Weiss and I. Kapouleas, "An empirical comparison of pattern recognition, neural nets, and machine learning classification methods," *Proceedings of IJCAI 1989*, Palo Alto, California, pp. 781–787, 1989.

[5] P. Smyth, "Automated Monitor and Control for Deep Space Network Subsystems," *TDA Progress Report 42-98*, vol. April–June 1989, Jet Propulsion Laboratory, Pasadena, California, pp. 110–120, August 15, 1989.

[6] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," in *Parallel Distributed Processing, Vol. 1*, Cambridge, Massachusetts: MIT Press, pp. 318–362, 1986.

[7] E. Barnard and R. Cole, "A neural net training program based on conjugate-gradient optimization," Oregon Graduate Centre Technical Report No. CSE 89-014, Beaverton, Oregon, 1989.

[8] M. J. D. Powell, "Restart procedures for the conjugate gradient method," *Mathematical Programming*, vol. 12, pp. 241–254, April 1977.

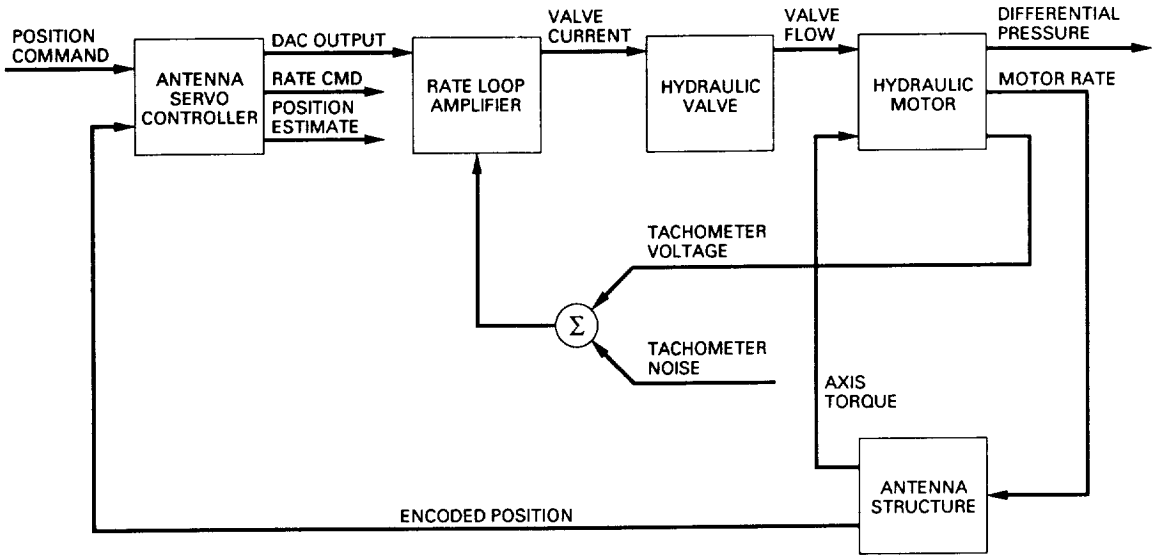


Fig. 1. The 70-m azimuth ACA simulation model.

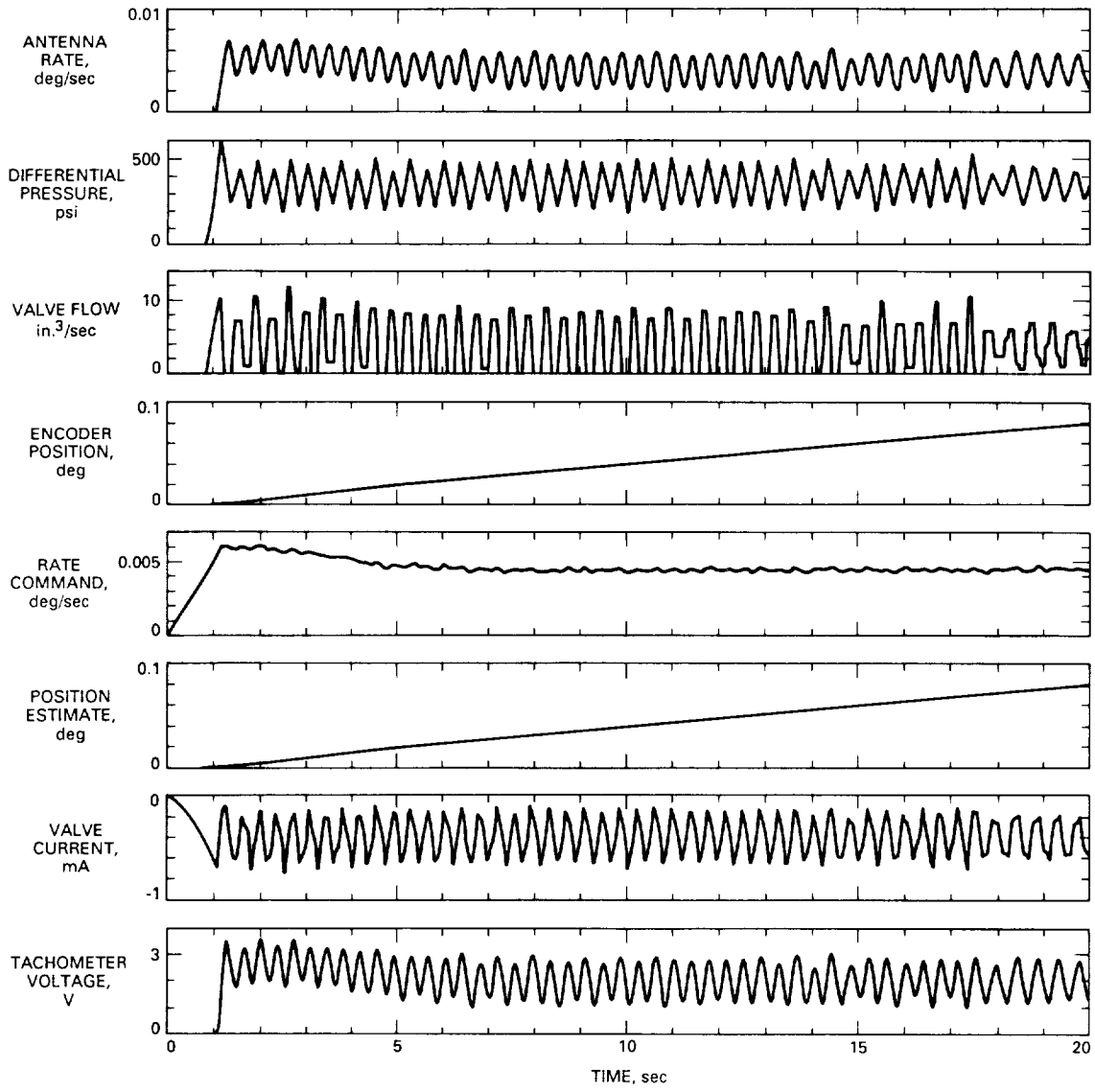


Fig. 2. Plot of each of eight features, sampled at 200 Hz for 20 sec, at rate = 4 mdeg/sec, under normal conditions (no faults).

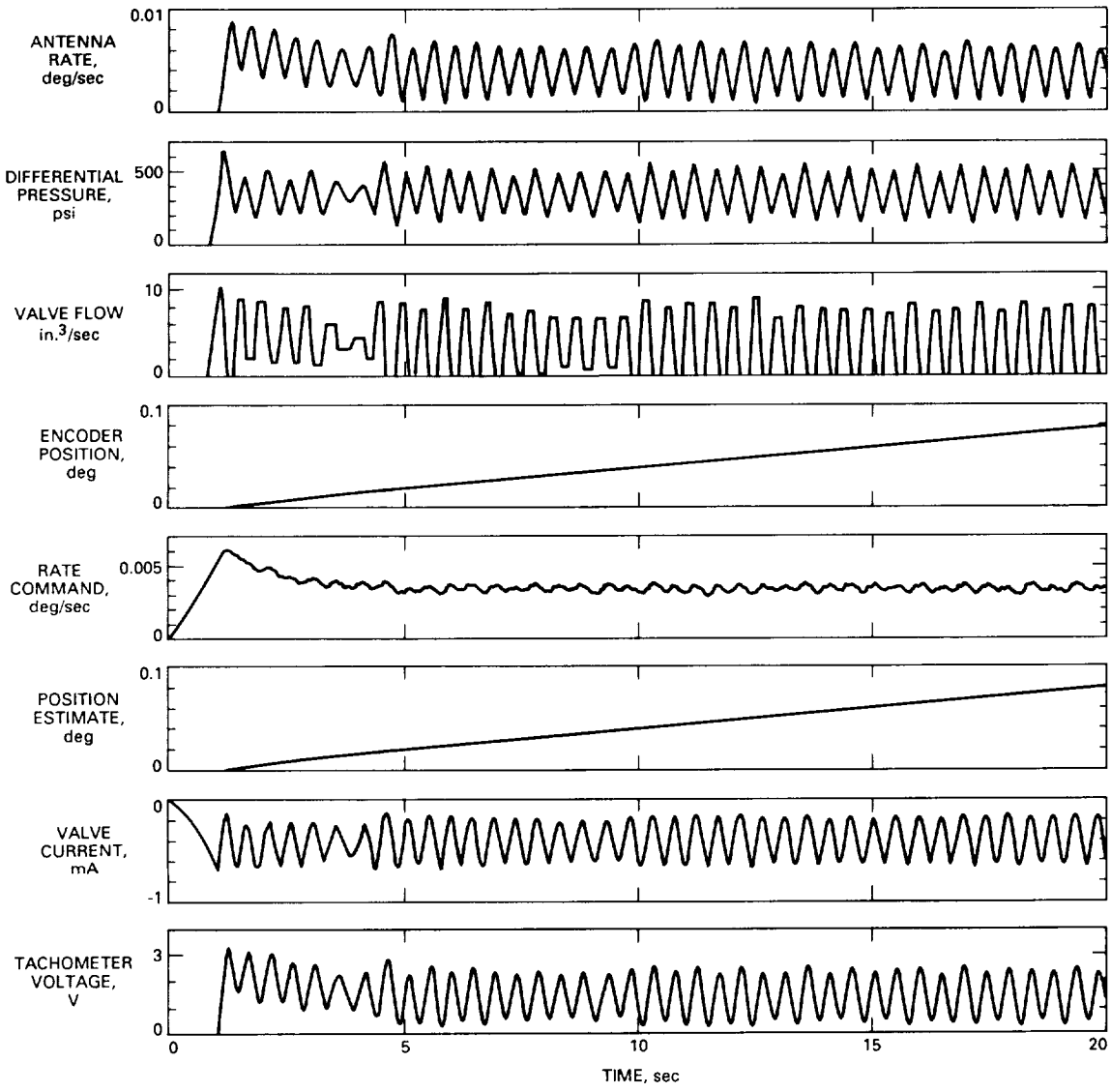


Fig. 3. Plot of each of eight features, sampled at 200 Hz for 20 sec, at rate = 4 mdeg/sec, with fault 1 (tachometer failure).

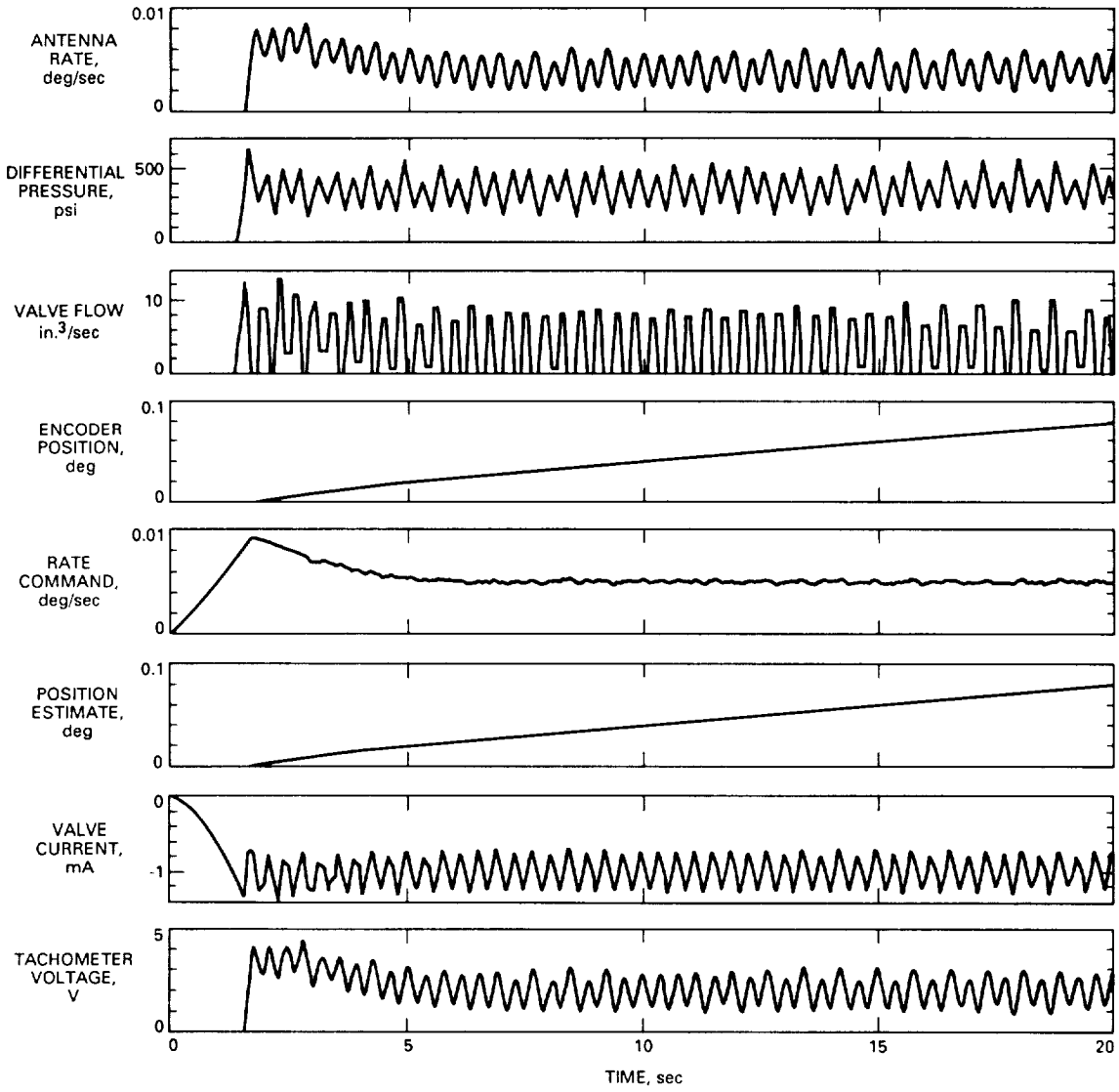


Fig. 4. Plot of each of eight features, sampled at 200 Hz for 20 sec, at rate = 4 mdeg/sec, with fault 2 (increased deadband).

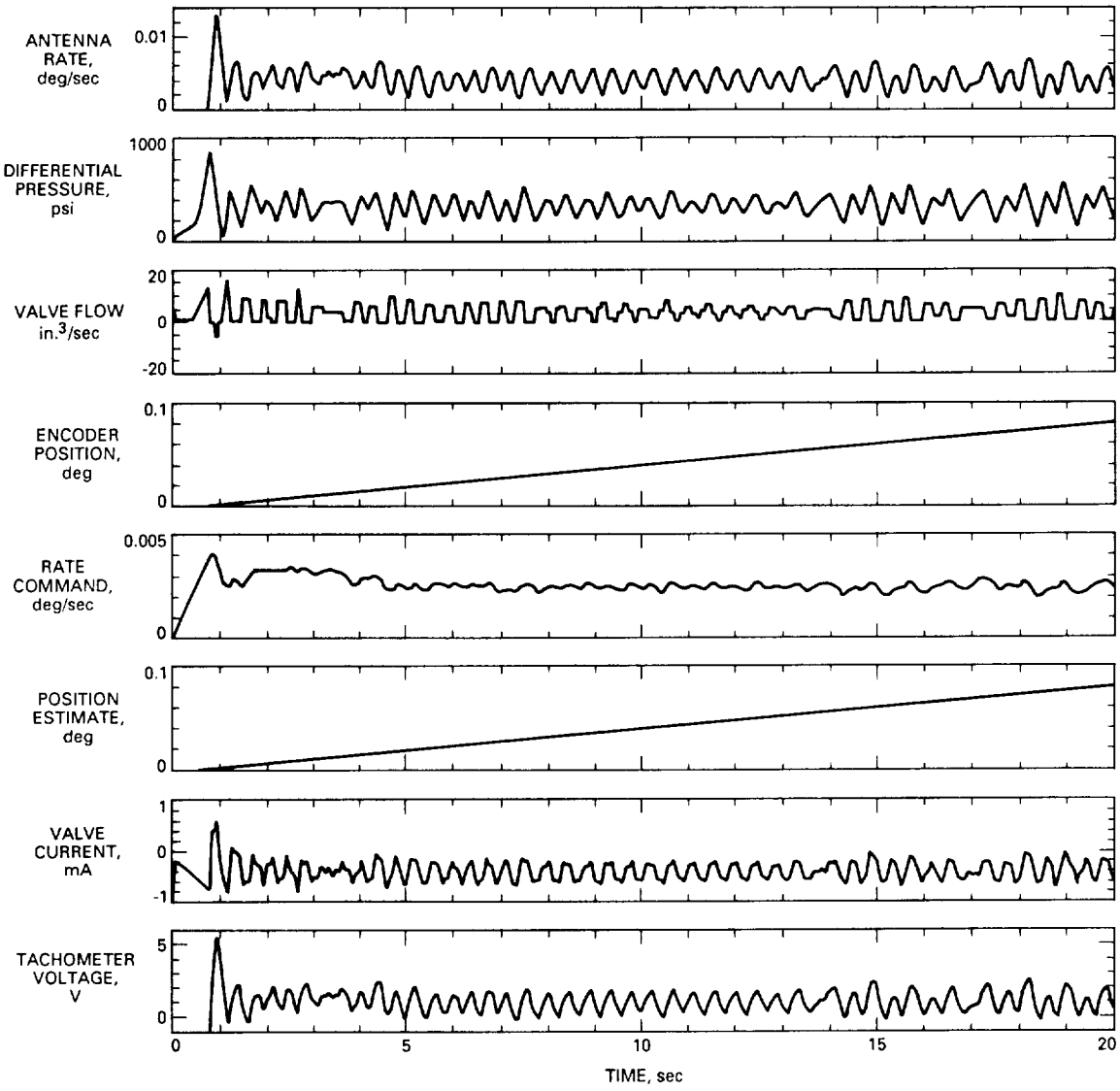


Fig. 5. Plot of each of eight features, sampled at 200 Hz for 20 sec, at rate = 4 mdeg/sec, with fault 3 (increased static friction).

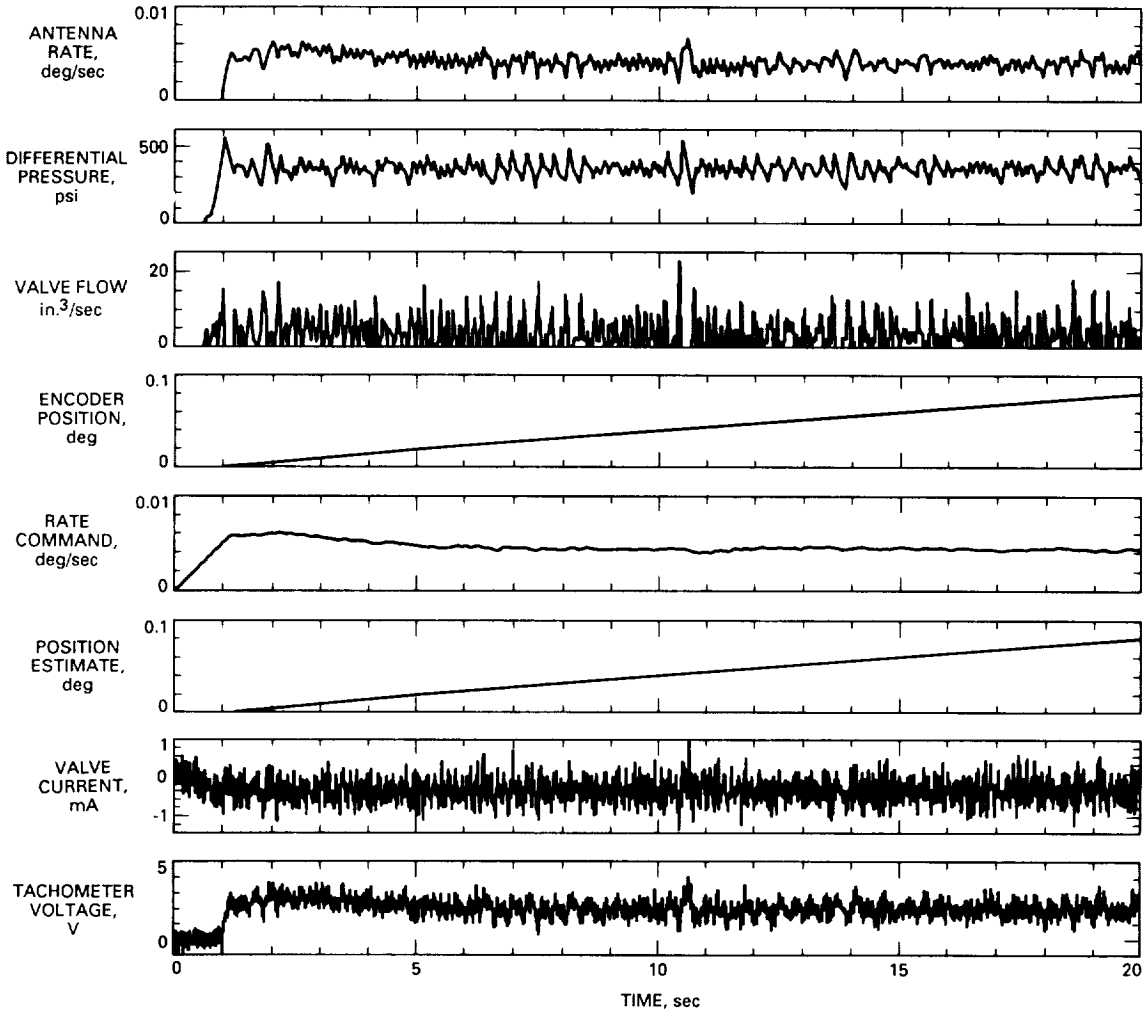


Fig. 6. Plot of each of eight features, sampled at 200 Hz for 20 sec, at rate = 4 mdeg/sec, with fault 4 (tachometer noise).

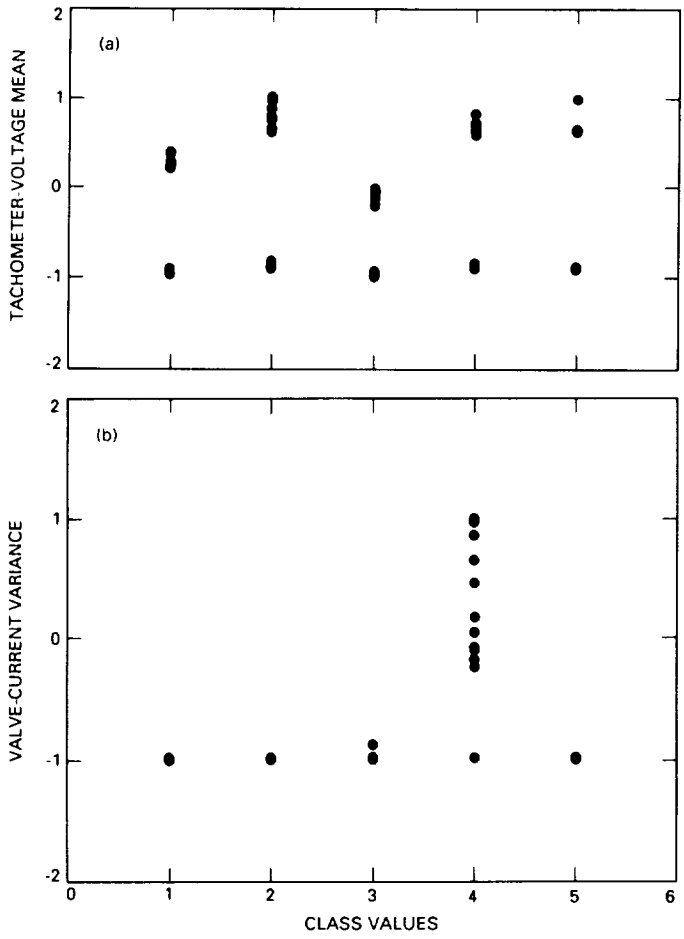


Fig. 7. Plots of (a) tachometer-voltage mean (normalized to ±1) versus class values; and (b) valve-current variance (normalized to ±1) versus class values.

CLASS AS ESTIMATED BY THREE-LAYER NEURAL NETWORK

ACTUAL TEST CLASS	1	2	3	4	5
1	18	0	3	0	3
2	0	24	0	0	0
3	0	0	25	0	0
4	0	0	0	28	0
5	0	0	0	0	25

CLASS 1: TACHOMETER FAILURE
CLASS 2: INCREASED DEADBAND
CLASS 3: INCREASED STATIC FRICTION
CLASS 4: TACHOMETER NOISE
CLASS 5: NOMINAL (NO FAULTS)

Fig. 8. Confusion matrix with 6 errors resulting from testing a particular neural network on independent test data of size 126 (percentage error = 4.8).

Appendix

Three-Layer Networks

Figure A-1 shows an example of a network. The input nodes are labeled n_i, the hidden nodes are labelled h_i, and the output layers are labelled o_i. In general, there are $K+1$ input units, where K is the number of features (12 in this case). The extra node is always in the "on" state, providing a threshold capability. Similarly, there are m output nodes, where $m = 5$ is the number of classes.

The number of hidden units was chosen arbitrarily in these experiments, but an empirically found rule of thumb to have between 1.5 and 2 times the number of input units typically worked well. The size of this hidden layer can influence the classifier performance critically: too many hidden units, and the network overfits the data (i.e., the estimation error will be large), whereas too few hidden units leaves the network with insufficient representational power (i.e., the approximation error term is large). With the weight from input unit n_i to hidden unit h_j as w_{ij}, each hidden unit calculates a weighted sum and passes the result through a nonlinear sigmoid function $F()$, i.e.,

$$a(h_j) = F\left(\sum_{i=1}^{i=K+1} w_{ij}\, a(n_i)\right)$$

where $a(n_i)$ is the activation of input unit i—typically, the actual value of feature i normalized to the range $+1, -1$. The function $F(x)$ is defined as

$$F(x) = \frac{1}{1+e^{-x}}$$

Output unit k, $1 \leq k \leq 5$ calculates a similar weighted sum using the weights w_{jk} between the jth hidden unit and the kth output unit, i.e.,

$$a(o_k) = \sum_j w_{jk}\, a(h_j)$$

A classification decision is made by choosing the output unit with the largest activation for a given set of inputs (feature values); i.e., choose class k such that

$$k = \arg\max_i \{a(o_i)\}$$

Hence, the optimization problem is to find the best set of weights such that the mean-square prediction error on the training data is as small as possible. Note that strictly speaking, from a statistical point of view, this is not the appropriate criterion, as the error on the training data may be an overly optimistic estimator of the true error of the classifier on unseen samples. Nonetheless, provided the number of free parameters in the network is at least an order of magnitude less than the number of training data points available, this minimization of training error is a reasonably robust procedure in practice.

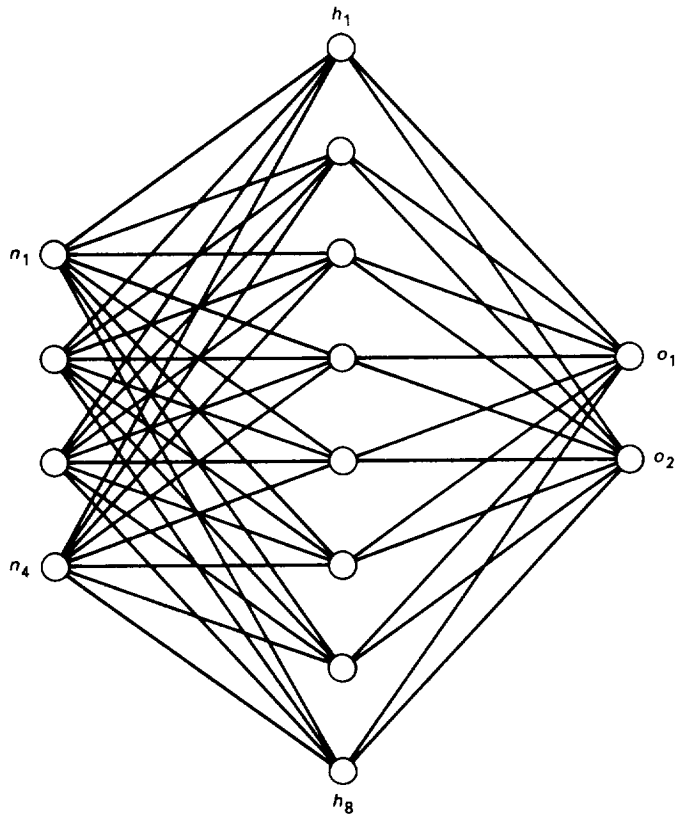


Fig. A-1. An example of a three-layer feed-forward neural network.

DSS 15, 45, and 65 34-Meter High-Efficiency Antenna Radio Frequency Performance Enhancement by Tilt Added to the Subreflector During Elevation Angle Changes

M. S. Katow[1]
Ground Antenna and Facilities Engineering Section

The focusing adjustments of the subreflector of an az-el Cassegrainian antenna that use only linear motions have always ended in lateral offsets of the phase centers at the subreflector's focus points at focused positions, which have resulted in small gain losses. This article describes how lateral offsets at the two focus points were eliminated by tilting the subreflector, resulting in higher radio frequency (RF) efficiencies at all elevation angles rotated from the rigging angle.

I. Introduction

At the 45-deg elevation angle, the subreflector of the shaped Cassegrainian geometry radio frequency (RF) system of the 34-m high-efficiency antenna is focused for maximum RF gain. Any change in the elevation angle results in gravity loading displacements to the reflective RF surfaces and subsequent RF-gain loss due to the defocused condition of the Cassegrainian system.

An analysis method for determining the lateral and axial translations of the subreflector required to maintain its optimum focused position throughout the elevation angle range, without the ability to tilt the axis of the subreflector, was described in [1]. Lateral offsets of the phase centers existed at the optimum focused positions, which resulted in some RF-gain loss. This article describes how the addition of the tilting adjustment of the subreflector minimized the lateral offsets at the phase centers of the secondary reflector. This correction resulted in smaller RF-gain losses with elevation angle motions from the 45-deg rigging elevation angle.

II. Shaped Cassegrainian RF Geometry

The 34-m shaped Cassegrainian RF geometry, set at the 45-deg elevation angle, is illustrated in Fig. 1. Like a paraboloid, the inner rays reflected from the subreflector are moved outward, resulting in more uniform RF-amplitude distribution across the main reflector for increased RF efficiency. Also, the equivalent focal length, as defined in a paraboloid, increases from about 11.02 m

[1] M. S. Katow, who is assigned on contract to the Ground Antenna and Facilities Engineering Section, is an employee of Planning Research Corporation, McLean, Virginia.

(434 in.) at the center to about 11.81 m (465 in.) at the outer edge of the reflector.

Assuming a single phase center, the movement of the primary phase center of the main reflector during the elevation angle change was determined by best fitting a paraboloid to the gravity-loaded displacements of the reflector. This was done by using both the longer focal length to represent the major part of the reflector surface and the root mean square (rms) program [2] to best fit a paraboloid to the reflector distortions for the gravity loading changes from a 45-deg to a 90-deg elevation angle. The change is defined in Fig. 2. The gravity distortions of the reflector were computed using the IDEAS program [3] along with the displacement of the nodes supporting the subreflector on the quadrupod. The field-measured displacements[2] between the subreflector and its quadrupod supporting nodes with dial indicators were then added to the computed displacements to determine the deflected position of the subreflector for the same elevation angle change as for the reflector. The deflected position, as calculated by procedures described in the following paragraphs, is also shown in Fig. 2.

III. Subreflector Gravity Displacements

To measure the deflections of the subreflector's supporting and controlling mechanism, dial indicators were installed, using special fixtures, between the apex of the quadrupod and the top of the subreflector as shown in Fig. 3.[2] These dial indicators were set to 12.7 mm (0.5 in.) at 90-deg elevation, and the deflection figures were read as the antenna was rotated in elevation angle down from a 90-deg elevation angle to a 6-deg elevation angle. The readings are plotted in Fig. 4 for dial indicator No. 5.[3] The No. 1, No. 2, and No. 3 dial-indicator readings are shown in Fig. 5, where dial readings were also set at 12.7 mm (0.5 in.) at 90-deg elevation, and the displacements were read as the elevation angle was reduced to 6 deg from 90 deg.[3]

The computed displacements of the structural nodes of the quadrupod's apex for the gravity loading change from a 45-deg to a 90-deg elevation angle are given in Table 1. The computed displacements portion of the subreflector is shown in Fig. 6, where the deflections of the quadrupod's supporting nodes are resolved into the displaced position of the top surface of the subreflector and its central axis, with no displacements assumed in the subreflector's supporting and controlling mechanism. To account for the deflections in the subreflector's supporting and controlling mechanism, the dial indicator readings of Figs. 4 and 5 resolved into displacements are shown in Fig. 7 for a 90-deg elevation with the subreflector set on axis at a 45-deg elevation. The summed or total displacement of the subreflector is also shown in Fig. 2 for a 90-deg elevation with a 0 setting at a 45-deg elevation.

IV. Subreflector Supporting and Controlling Mechanism

The subreflector is supported by an assembly of three parallel plates in which the top plate is supported by three vertical jacks attached to the quadrupod. The top view of the vertical jacks' positions are shown in Fig. 8. The three plates are interconnected by slides and horizontal jacks to provide relative motion in the X- and Y-axes. Position indicating transducers parallel to the jacks provide a servo system where tables inserted in a microprocessor control the position of the jacks according to the elevation angle of the antenna.

V. Axial (Z)-Focus Curve Table

With the availability of field-measured Z-focus curve from the DSS 15 antenna, the focus-curves data from A. Freiley were used (Fig. 9).[4] The initially computed focus curve, designated as autofocus, was used in the DSS 15 servo controlling system. The measured focus curve indicated more deflection than was computed in the subreflector/reflector system and also showed an error in Z-positioning at a 45-deg elevation of −1.02 mm (−0.04 in.). A "measured" focus line is drawn through the experimental data. In order to develop an operating table for the microprocessor, it follows that:

$$Z = K\,(\sin(\text{elevation angle}) - \sin(\text{rigging el angle})) + \text{offset}$$

The fit of this curve with $K = 9.65$ mm (0.380 in.), with an offset of −1.0 mm (−0.039 in.), is shown by "X" points in Fig. 9. Reduced to table form, this curve when no tilting of the subreflector is done is shown in Table 2. Table 2 also delineates the curve with no offset, to be used when the zero-C focused position of the subreflector corresponds with the dial indicators.

[2] B. Parvin, "DSS-15 S/R Position Test," JPL IOM 3324-86-31 (internal document), Jet Propulsion Laboratory, Pasadena, California, April 21, 1986.

[3] Data by Ben Parvin.

[4] A. Freiley, "Preliminary Results of the 34M H.E. Antenna X-Band Performance Measurements at DSS-15," JPL IOM 3334-0885-023 (internal document), Jet Propulsion Laboratory, Pasadena, California, May 31, 1985.

VI. Subreflector Tilt and Y-Lateral Adjustments

Although the complex quadrupod assembly is modeled closely in the computing process, the deflections usually are larger than computed, as illustrated by Fig. 9. To address this problem, a 10-percent addition will be made to the computed and measured displacements of the subreflector. Also, to simplify the calculations, the assumption is made that the lateral adjusting jackscrews and the Z-axial jackscrews act on a plane on the same level as the top plate of the controlling mechanism. In other words, no interaction is assumed. This was thought to be consistent with the accuracy requirements.

Figure 10 delineates the total tilt and the Y-lateral adjustments required at a 90-deg elevation angle, starting from el = 45 deg, to focus the subreflector to the phase center of the primary RF feed horn and the primary focus of the best-fit paraboloid:

For el = 45 deg to el = 90 deg,

$$\text{tilt change} = \frac{3.101 - 2.298}{689.643} = 0.001164 \text{ rad}$$

For el = 45 deg to el = 90 deg,

$$\text{Y-travel} = -3.1001 + (31.75 \times 0.001164) = -3.064 \text{ cm}$$

When tilting the subreflector from el = 45 deg to 90 deg, the jackscrew length changes. Jackscrew No. 1 must be raised as follows:

$$0.001164 \times 1.194 \text{ m} = 1.390 \text{ mm } (0.055 \text{ in.})$$

Jackscrews No. 2 and No. 3 must be lowered as follows:

$$0.001164 \times -0.597 \text{ m} = -0.695 \text{ mm } (-0.0275 \text{ in.})$$

It turns out that the cosine function of the elevation angle accurately describes the tilting of the subreflector [1] from 0 deg to 90 deg. For jackscrew No. 1,

$$Z_1 = K_2 \text{ (cos (el) } - \text{ cos (rigging angle))} + \text{Z-axial focus}$$

For jackscrews No. 2 and No. 3,

$$Z_{23} = \frac{-K_2 \text{ (cos (el) } - \text{ cos (rigging angle))}}{2} + \text{Z-axial focus}$$

K_2 = total jack travel constant for 0-deg to 90-deg elevation angle change

$$= 1.390/\cos 45 \text{ deg} = 1.966$$

The summed jackscrew positions for tilt and axial focus corrections computed from the above equations are delineated in Tables 3 and 4. These table values can be incorporated in the microprocessor for controlling the subreflector jackscrews for axial and tilting adjustments during the elevation angle changes.

For the −3.064 cm (−1.206 in.) Y-lateral adjustment of the subreflector (Fig. 10) required to maintain focus at el = 90 deg after the elevation angle change from el = 45 deg, the total Y-lateral travel equals −3.064 cm (−1.206 in.) divided by cosine 45 deg or −4.333 cm (−1.706 in.), which equals Ky. It follows that the equation controlling the Y-lateral adjustment equals

$$Y = K_y \text{ (cos (El) } - \text{ cos (45 deg))}$$

Finally, the Y-lateral positions shown in Table 5, computed from the above equation, can be incorporated in the microprocessor to control the Y-lateral motion of the subreflector.

VII. Conclusions

The described tables were installed first in the DSS 45 subreflector control system, and the measured efficiency data[5] are shown in Fig. 11, where the RF efficiencies ranged from about 67 to 70 percent at X-band.[6]

Although the measured RF efficiency change from a 45-deg elevation angle to a 10-deg elevation angle indicates only about a 3 percent drop, some field-checking focusing at the low angles should be done for possible improved RF-efficiency values. These field readings could then be used to upgrade the focusing tables in the subreflector's controller to enhance future operations required when higher RF frequencies are used.

---

[5] D. Bathker, "Data with Graham Baines, DSS-45," JPL IOM 3331-88-061 (internal document), Jet Propulsion Laboratory, Pasadena, California, June 30, 1988.

[6] Figure 11 reports measured efficiency according to existing standards (*DSN Radio Source List for Antenna Calibration*, JPL D-3801, Rev. B (internal document), Jet Propulsion Laboratory, Pasadena, California, September 25, 1987) and includes slight atmospheric extinction. The source flux and size standards within JPL D-3801, Rev. B are currently being revised such that the true peak antenna efficiency, sans atmosphere, is slightly over 75 percent. A revised *DSN Radio Source List* is being produced.

References

[1] R. D. Hughes and M. S. Katow, "Subreflector Focusing Techniques Applied to New DSS-15 and DSS-45 34-Meter Antennas," *TDA Progress Report 42-80*, vol. October–December 1984, Jet Propulsion Laboratory, Pasadena, California, pp. 83–90, February 15, 1985.

[2] M. S. Katow and L. W. Schmele, "Antenna Structures: Evaluation Techniques of Reflector Distortions," in *Supporting Research and Advanced Development, Space Programs Summary*, 37-40, vol. IV, Jet Propulsion Laboratory, Pasadena, California, pp. 176–184, April 15, 1967.

[3] R. Levy, "Optimization of Antenna Structure Design," *Eighth Conference on Electronic Computation*, ASCE, Houston, Texas, pp. 114–129, February 1983.

Table 1. Displacements of structural nodes of the quadrupod's apex

Node no.	Displacements		
	X, mm/in.	Y, mm/in.	Z, mm/in.
16131	0.23/0.009	17.70/0.697	−1.96/−0.077
16141	0.23/0.009	17.80/0.701	−1.96/−0.077
15130	−0.25/−0.010	17.70/0.697	−0.30/−0.012
15140	−0.25/−0.010	17.80/0.701	−0.30/−0.012

Table 2. Z-axial focus curve

Seq	Elevation angle, deg	sin, el	sin, el − sin, 45 deg	Z with K = 9.65 K = 0.38		Z with 1.02(−0.4) offset	
				mm	in.	mm	in.
1	0	0.0	−0.7071	−6.82	−0.269	−7.85	−0.309
2	10	0.1736	−0.5335	−5.16	−0.203	−6.17	−0.243
3	20	0.3420	−0.3651	−3.53	−0.139	−4.55	−0.179
4	30	0.5000	−0.2071	−2.01	−0.079	−3.02	−0.119
5	40	0.6428	−0.0643	−0.61	−0.024	−1.63	−0.064
6	45	0.7071	0.0	0.0	0.0	−1.02	−0.040
7	50	0.7660	0.0589	0.56	0.022	−0.46	−0.018
8	60	0.8660	0.1589	1.52	0.060	0.51	0.020
9	70	0.9397	0.2326	2.24	0.088	1.22	0.48
10	80	0.9848	0.2777	2.69	0.106	1.68	0.066
11	90	1.000	0.2929	2.82	0.111	1.80	0.71

Table 3. Jack 1 tilt corrections Z-axial focus

Seq	Elevation angle, deg	cos, el	cos, el − cos, 45 deg	(1) Tilt correction for jack no. 1		(2) Z−axial focus for no tilt	
				mm	in.	mm	in.
1	0	1.0	0.2929	−0.58	−0.023	−6.83	−0.269
2	10	0.9848	0.2777	−0.53	−0.021	−5.16	−0.203
3	20	0.9397	0.2326	−0.46	−0.018	−3.53	−0.139
4	30	0.8660	0.1589	−0.30	−0.012	−2.01	−0.079
5	40	0.7660	0.0589	−0.13	−0.005	−0.61	−0.024
6	45	0.7071	0.0	0.0	0.0	0.0	0.0
7	50	0.6428	0.0643	0.13	0.005	0.56	0.022
8	60	0.5000	−0.2071	0.41	0.016	1.52	0.060
9	70	0.3420	−0.3651	0.71	0.028	2.24	0.088
10	80	0.1736	−0.5335	1.07	0.042	2.69	0.106
11	90	0.0	−0.7071	1.40	0.055	2.82	0.111

Table 4. Focused jack positions

Seq	Elevation angle, deg	(3)[a] Jack no. 1 Z-position		(4)[b] Jacks no. 2 and no. 3 Z-position	
		mm	in.	mm	in.
1	0	−7.41	−0.292	−6.55	−0.258
2	10	−5.69	−0.224	−4.90	−0.193
3	20	−3.99	−0.157	−3.30	−0.130
4	30	−2.31	−0.091	−0.85	−0.073
5	40	−0.74	−0.029	−0.53	−0.021
6	45	0.0	0.0	0.0	0.0
7	50	0.69	0.027	0.51	0.020
8	60	1.93	0.076	1.32	0.052
9	70	2.95	0.116	1.88	0.074
10	80	3.76	0.148	2.16	0.085
11	90	4.22	0.166	3.52	0.139

[a] Column (3) = column (1) + column (2).

[b] Column (4) = $-\frac{\text{column(1)}}{2}$ + column (2).

Table 5. Y-lateral adjustments

Seq	Elevation, deg	cos, el	(a) cos, el − cos, el_{45}	(a) × 1.706 = Y-lateral position
1	0	1.0	0.2929	0.500
2	15	0.9659	0.2588	0.442
3	30	0.8660	0.1589	0.159
4	45	0.7071	0.0	0.0
5	60	0.5000	−0.2071	−0.207
6	75	0.2588	−0.4483	−0.765
7	90	0.0	−0.7071	−1.206

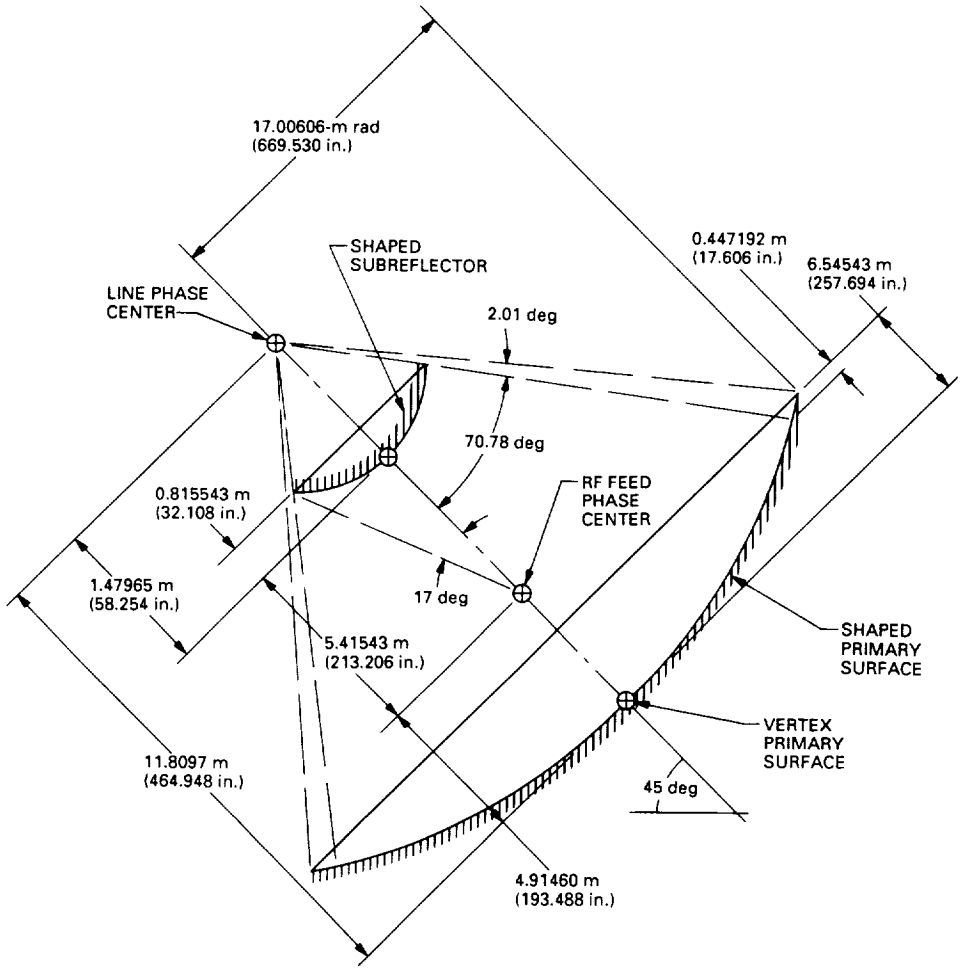


Fig. 1. The 34-m shaped Cassegrainian RF geometry at 45-deg elevation angle.

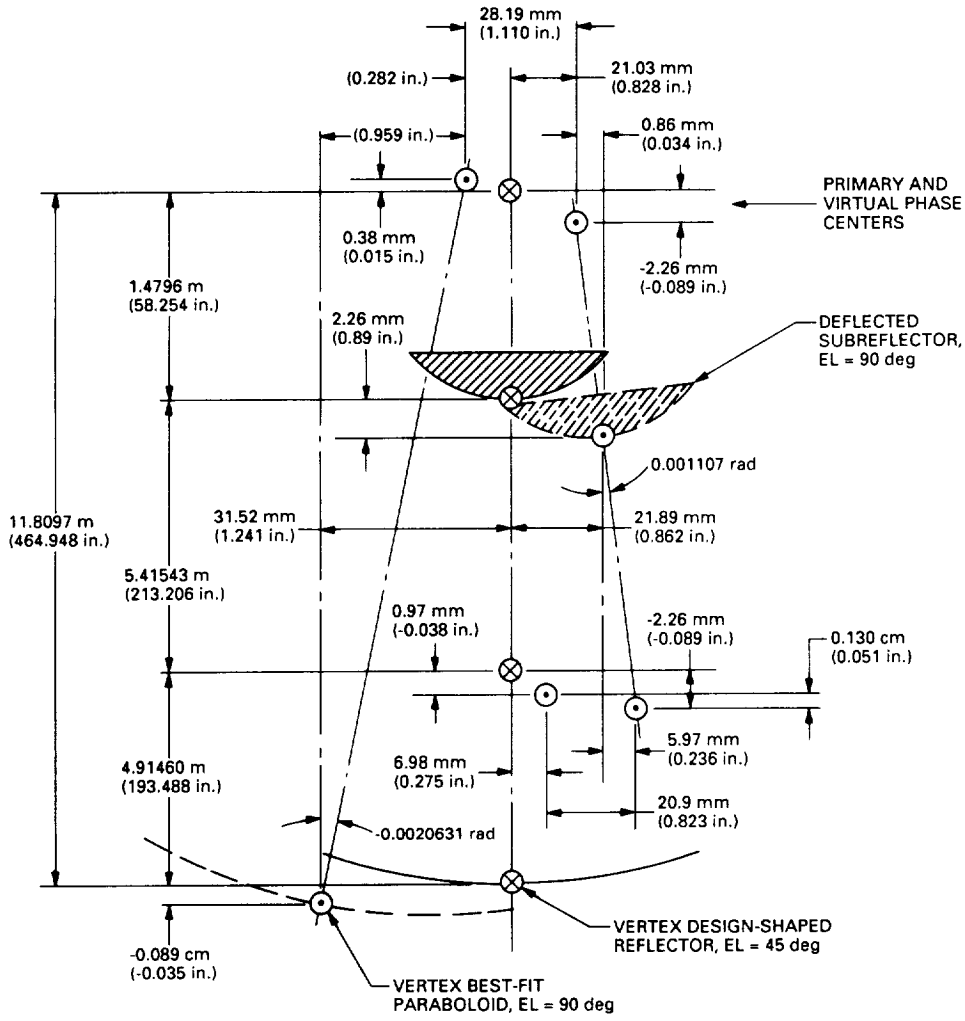


Fig. 2. Cassegrainian displacements at el = 90 deg with surfaces set at el = 45 deg.

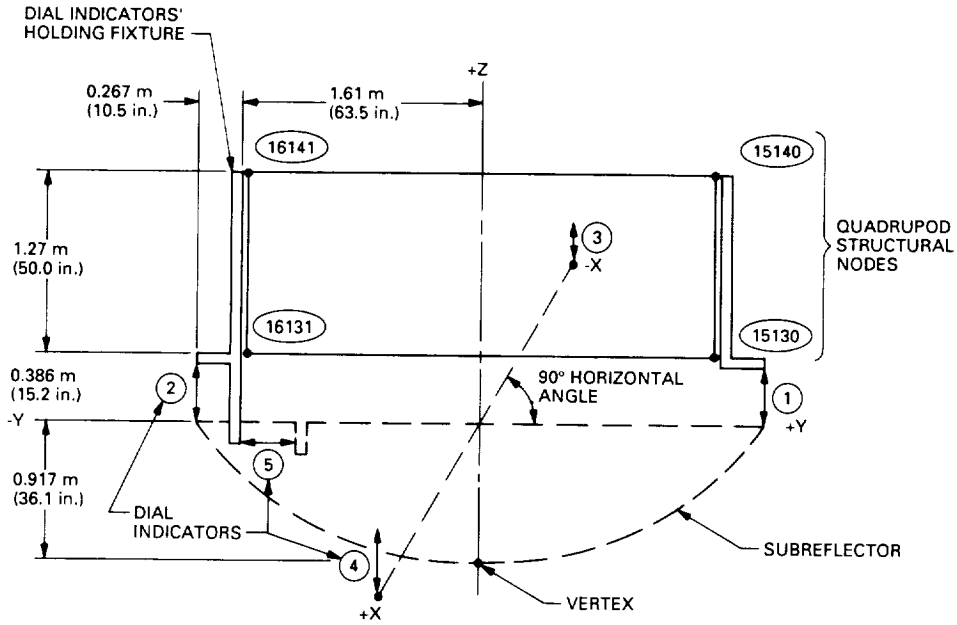


Fig. 3. Dial indicator locations.

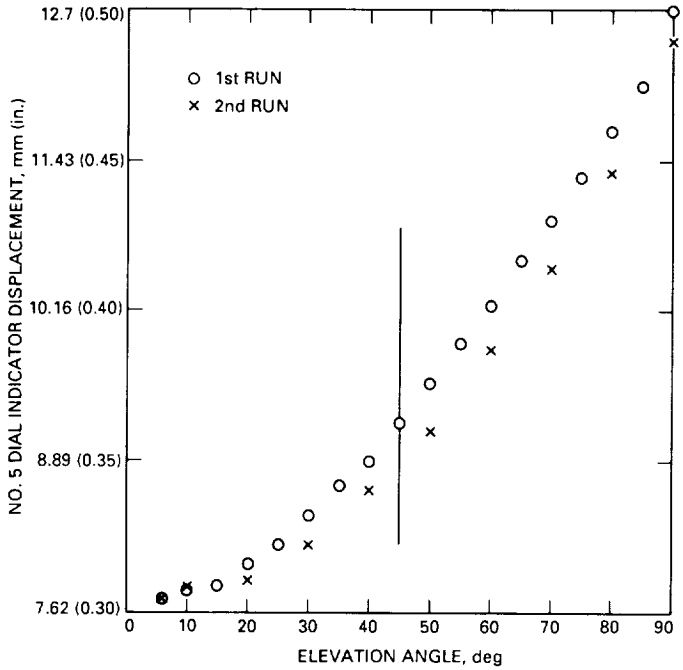


Fig. 4. The 34-m az-el subreflector field deflections from apex, No. 5 dial indicator (shown in Fig. 3), lateral displacements.

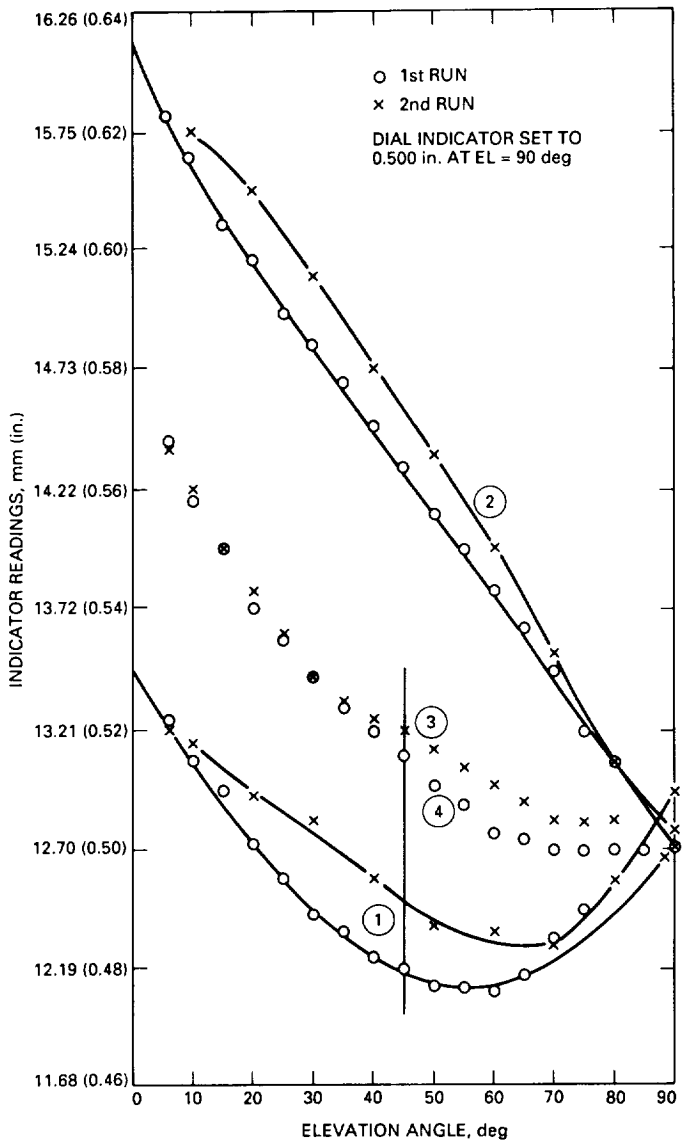


Fig. 5. The 34-m az-el subreflector field deflections from apex, No. 1, No. 2, No. 3, and No. 4 dial indicators (shown in Fig. 3).

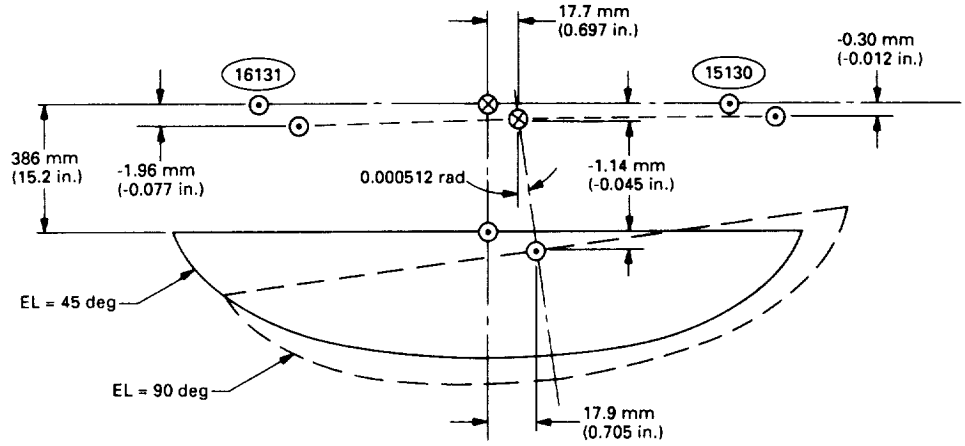


Fig. 6. Computed displacements of subreflector at el = 90 deg, set at 45 deg.

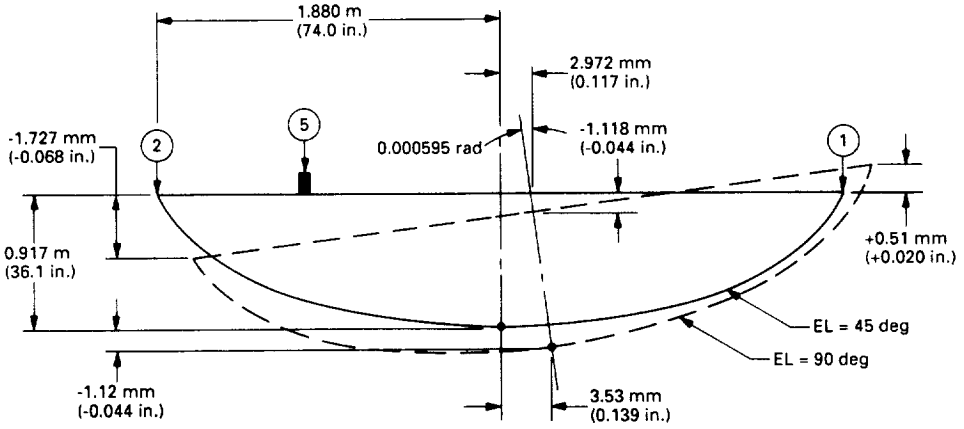


Fig. 7. Dial-indicated displacements of subreflector at el = 90 deg, set at 45 deg.

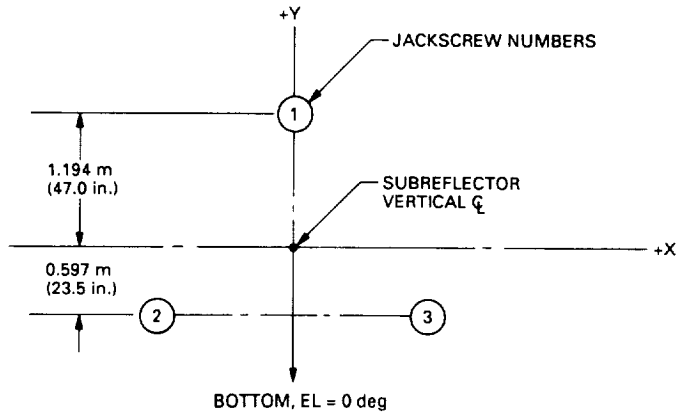


Fig. 8. X-Y jackscrew locations on quadrupod.

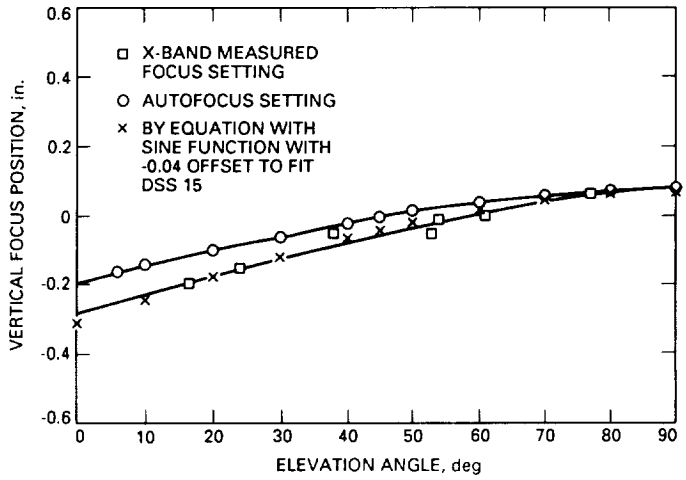


Fig. 9. The 34-m field data, DSS 15.

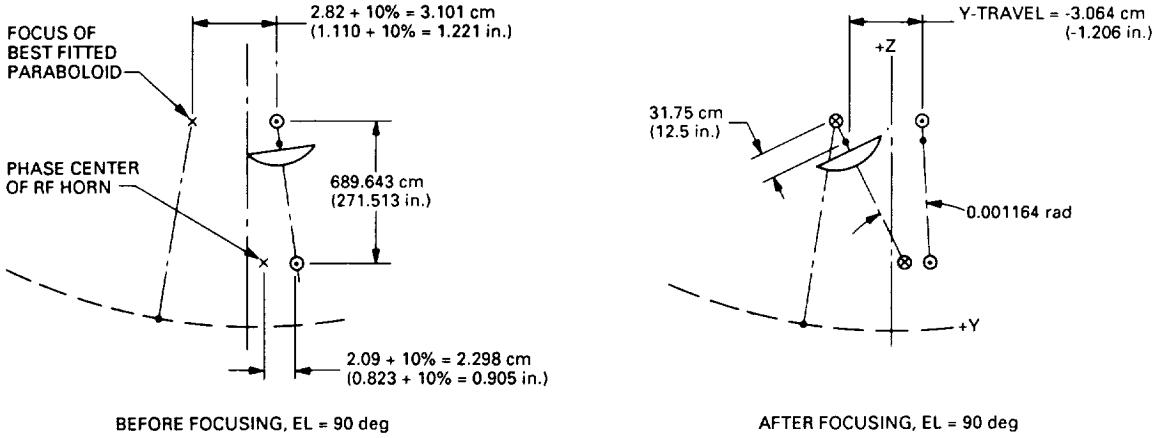


Fig. 10. Subreflector lateral and tilt focusing adjustments, el = 45 deg to el = 90 deg.

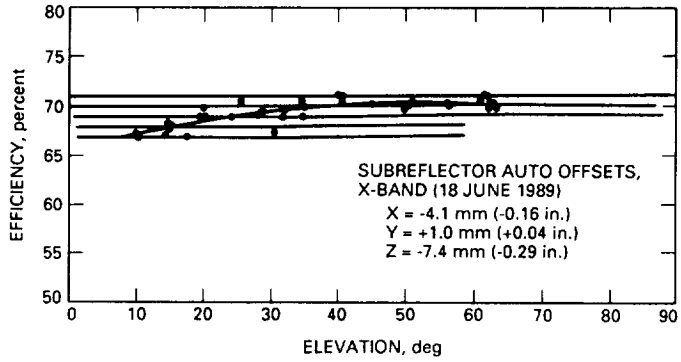


Fig. 11. DSS 45 measured efficiency.

QPSK Loop Lock Detection in the Advanced Receiver

A. Mileant
Telecommunications Systems Section

S. Hinedi
Communications Systems Research Section

The Advanced Receiver (ARX II) currently being developed uses a Costas cross-over loop to acquire and track the phase of an incoming quadrature phase-shift-keyed (QPSK) signal. This article describes the performance of the QPSK lock detector to be implemented, taking into account the phase jitter in the tracking loop. Simulations are used to verify the results of the analysis.

I. Introduction

The Advanced Receiver II (ARX II) [1] is currently being developed to demodulate signals from deep space spacecraft. In addition to processing binary phase-shift-keyed (BPSK) signals, the ARX II will acquire and track quadrature phase-shift-keyed (QPSK) signals. The tracking as well as the acquisition performance of several QPSK loops has been investigated [2], and it was determined that the Costas cross-over loop offered a "good" compromise between implementational complexity and relative performance. The other two candidates were the maximum a posteriori (MAP) estimation and the generalized Costas loops. This article describes a lock detector for the QPSK loops. The analysis is general in that it is applicable to all three QPSK loops, and the results are validated by computer simulations. The general QPSK carrier-tracking loop structure and the lock detector are both shown in Fig. 1. When this loop is implemented, an extra accumulator is present in front of the loop filter to reduce the loop update rate and improve the loop signal-to-noise ratio (SNR). The lock detector structure can be generalized to accommodate any multiple phase-shift-keying (MPSK) signal as described in [3].

II. Lock Detection Analysis

The received QPSK signal can be modeled by

$$r(t) = \sqrt{P_D}a(t)\sin(\omega_0 t + \theta) + \sqrt{P_D}b(t)\cos(\omega_0 t + \theta) + n(t) \quad (1)$$

where

$$a(t) = \sum_k a_k p(t - kT_s)$$

and

$$b(t) = \sum_k b_k p(t - kT_s)$$

are the independent in-phase and quadrature data streams with a_k and b_k the binary ± 1 random data and $p(t)$ the

non-return-to-zero (NRZ), or Manchester, pulse of duration over T_s seconds. The data power is given by P_D and the incoming phase by θ. The signal is received in the presence of noise that can be expressed as

$$n(t) = \sqrt{2}n_c(t)\cos(\omega_0 t + \theta) - \sqrt{2}n_s(t)\sin(\omega_0 t + \theta) \quad (2)$$

where $n_c(t)$ and $n_s(t)$ are random processes with a two-sided noise spectral density of $N_0/2$. The input signal is mixed with the reference signals

$$x_c(t) = \frac{2\sqrt{P_D}}{N_0}\cos(\omega_0 t + \hat{\theta}) \quad (3a)$$

and

$$x_s(t) = \frac{2\sqrt{P_D}}{N_0}\sin(\omega_0 t + \hat{\theta}) \quad (3b)$$

to produce, after integrating and dumping, the samples

$$\begin{aligned} r_{sk} &= \int_{kT_s}^{(k+1)T_s} r(t)x_s(t)\,dt \\ &= R\left(a_k\cos\phi - b_k\sin\phi - \frac{x_{1k}}{\sqrt{R}}\cos\phi - \frac{x_{2k}}{\sqrt{R}}\sin\phi\right) \\ r_{ck} &= \int_{kT_s}^{(k+1)T_s} r(t)x_c(t)\,dt \\ &= R\left(a_k\sin\phi + b_k\cos\phi - \frac{x_{1k}}{\sqrt{R}}\sin\phi + \frac{x_{2k}}{\sqrt{R}}\cos\phi\right) \end{aligned} \quad (4)$$

where $R \triangleq \frac{P_D T_s}{N_0}$ is the symbol SNR, x_{1k} and x_{2k} are two zero-mean, independent white Gaussian random sequences with unity variance, and $\phi \triangleq \theta - \hat{\theta}$ is the phase estimation error. The reference signals of Eq. (3a) and Eq. (3b) need to be normalized by $\sqrt{P_D}/N_0$ for only the MAP estimation loop, which contains the "hyperbolic tangent" nonlinearity. For the other two loops, the amplitude can be a constant independent of $\sqrt{P_D}/N_0$. The tracking performance of the three QPSK tracking loops has been derived elsewhere [3], and the results are summarized in Appendix A.

The lock detector algorithm used in the ARX II is summarized in Appendix B for a general MPSK signal. In the specific case of a QPSK signal, the algorithm reduces to

$$z = \sum_{k=1}^{M} y_k \gtrless \tau \quad (5a)$$

where z is the detector's signal obtained from samples y_k

$$y_k \triangleq \left(r_{ck}^2 - r_{sk}^2\right)^2 - \left(2r_{ck}r_{sk}\right)^2 \quad (5b)$$

It is shown in Appendix C that the mean and the variance of the detector signal z can be expressed as

$$\mu_z = -4MR^4\overline{\cos 4\phi} \quad (6a)$$

and

$$\begin{aligned} \sigma_z^2 = {} & 2M\left[m_8 + 19m_{44} - 12m_{62}\right] \\ & + 4M\left[\frac{1}{M}\sum_{k=1}^{M-1} c_k d_k - Mm_4^2 - 9m_2^4 + 6m_4 m_2^2\right] \end{aligned} \quad (6b)$$

where ϕ denotes the carrier phase error; $m_i = \mathbf{E}\{r_c^i\} = \mathbf{E}\{r_s^i\}$ the ith moment; $m_{ij} = \mathbf{E}\{r_c^i r_s^j\}$ the ijth cross-moment; and c_k, d_k are constants defined in Appendix C. (The overbar and $\mathbf{E}$ indicate expected values.) Note that at high-loop SNR (i.e., when $\sigma_\phi \to 0$), the mean and variance reduce to

$$\mu_z = -4MR^4 \quad (7a)$$

and

$$\sigma_z^2 = 64MR^8\left[\frac{2}{R} + \frac{9}{R^2} + \frac{12}{R^3} + \frac{3}{R^4}\right] \quad (7b)$$

from which the ideal detector's SNR can be derived. At lower loop SNRs, the phase jitter might not be negligible and might result in a degraded detector SNR. All the derivations have been summarized in Appendix C for reference.

III. Probability of Detection and of False Indication

During carrier lock detection, each z sample is compared with a predefined threshold τ, and the lock detector decides that the loop is in-lock when z exceeds τ (i.e., $z > \tau$). It is possible that even when the loop is still not locked or when no signal is present, z will occasionally be larger than τ. In this case, the lock detector will mistakenly declare an in-lock condition. The probability of this event (probability of false indication) is

$$P_f = \frac{1}{\sqrt{2\pi\sigma_{z0}^2}} \int_\tau^\infty \exp\left(-\frac{(z-\mu_{z0})^2}{2\sigma_{z0}^2}\right) dz$$
$$= \frac{1}{2}\operatorname{erfc}\left(\frac{\tau-\mu_{z0}}{\sqrt{2\sigma_{z0}^2}}\right) \quad (8)$$

where μ_{z0} and σ_{z0}^2 are the mean and variance of the lock detector signal in the out-of-lock state and $\operatorname{erfc}(x)$ is the complementary error function ($\operatorname{erfc}(x) = 1 - \operatorname{erf}(x)$, where $\operatorname{erf}(x)$ is the error function). Equation (8) assumes that z is a Gaussian random variable since z is the sum of many samples of equal variance. When the loop is not locked on the signal, the input can be modeled as pure noise to produce

$$\mu_{z0} = 0$$
$$\sigma_{z0}^2 = 192MR^4 \quad (9)$$

On the other hand, when the loop is locked on the signal, the probability of detection is given by

$$P_d = \frac{1}{\sqrt{2\pi\sigma_z^2}} \int_\tau^\infty \exp\left(-\frac{(z-\mu_z)^2}{2\sigma_z^2}\right) dz$$
$$= \frac{1}{2}\operatorname{erfc}\left(\frac{\tau-\mu_z}{\sqrt{2\sigma_z^2}}\right) \quad (10)$$

where μ_z and σ_z^2 are given by Eq. (6). Alternatively, the threshold can be eliminated from Eq. (10) and the probability of detection expressed as

$$P_d = \frac{1}{2}\operatorname{erfc}\left(\frac{\sigma_{z0}}{\sigma_z}\operatorname{erfc}^{-1}(2P_f) - \sqrt{\frac{SNR_z}{2}}\right) \quad (11)$$

where $SNR_z = \mu_z^2/\sigma_z^2$ denotes the detector SNR. The phase jitter in the tracking loop degrades the detector's SNR by a factor D,

$$D = \frac{SNR_z}{SNR_{z,ideal}} \quad (12)$$

where $SNR_{z,ideal}$ is the detector SNR if an infinite-loop SNR is assumed; i.e., there is no phase jitter ($SNR_{z\,(ideal)}$ is computed by using the high-SNR approximations for μ_z and σ_z^2 given by Eq. (7)). In Appendix A, it is shown that at high-loop SNR the detector's SNR reduces to

$$SNR_{z,ideal} = \frac{MR}{4(2 + \frac{9}{R} + \frac{12}{R^2} + \frac{3}{R^3})} \quad (13)$$

For a given M, loop SNR, ρ, ($\rho = 1/\sigma_\phi^2$ where σ_ϕ^2 is given by Eq. (A-13)), and P_f, the detector's SNR, has to be increased roughly by the factor $1/D$ in order to achieve the desired probability of detection P_d.

Given P_d and P_f, we can solve for the number of required detector samples M, namely,

$$M = \frac{2}{\mu_y^2}\left[\sigma_{y0}\operatorname{erfc}^{-1}(2P_f) - \sigma_y\operatorname{erfc}^{-1}(2P_d)\right]^2 \quad (14)$$

The threshold τ is obtained by solving Eq. (8) and setting it equal to

$$\tau = \sqrt{2\sigma_{z0}^2}\operatorname{erfc}^{-1}(2P_f) \quad (15)$$

where $\operatorname{erfc}^{-1}(\cdot)$ is the inverse complementary error function. When the loop is in lock, it can be argued via the central-limit theorem that the random variable z is approximately Gaussian, with mean and variance as obtained earlier.

IV. Discussion and Numerical Results

Computer simulations were used to validate the various assumptions made in the analysis. Figure 2 depicts the detection probability as a function of loop SNR when $P_f = 10^{-4}$. It is clear that the detection probability degrades significantly at low-loop SNR and the degradation itself is a function of the desired detection probability $\overline{P_d}$ (which is computed assuming zero phase jitter). The degradation is better shown in Fig. 3, which depicts the loss in detector SNR as a function of the loop SNR. At 20-dB loop SNR, the degradation is about 1 dB, but it increases to 2.5 dB as the loop SNR is reduced to 15 dB. In both figures, the simulations and theory are in agreement for all three loops. Note that the degradation in detector SNR for BPSK was about 1 dB for 15-dB loop SNR [5].

When the detection probability was plotted as a function of symbol SNR, different performances for the various loops were expected, as the loop SNRs are different when both the data rate and the loop bandwidth are fixed. Figure 4 clearly depicts that effect, as the MAP estimation loop attains a higher detection probability than the Costas cross-over loop because of its higher loop SNR. Finally, Fig. 5 shows both the nominal and the actual (16-dB loop SNR) detection probabilities as a function of the number of samples M and the corresponding threshold. This kind of curve is very useful when one is designing the operating parameters of the lock detector.

Figure 6 depicts the probability of detection versus the detector's SNR for several values of probability of false lock. This figure is useful in determining the required detector's SNR for a desired detector's performance. Finally, Fig. 7 plots the detector's SNR versus E_s/N_0 for $M = 300$. The detector's SNR for other values of M can be computed from the following relation:

$$SNR_{z,M} = SNR_{z,300} + 10\log_{10}\left(\frac{M}{300}\right) \text{ dB} \qquad (16)$$

As a design example, suppose that a QPSK detector operates at $P_f = 10^{-4}$ and $P_d = 0.99$, and that the signal rate is 10,000 symbols per second with $E_s/N_0 = 1.0$ dB. Using Fig. 6, it can be seen that the detector's SNR should be about 12 dB. But, Fig. 7 states that at $E_s/N_0 = 1$ dB, $M = 300$ and loop SNR = 16 dB (low-loop SNR case), the detector's SNR will actually be about 5.0 dB. Equation (16) is used to compute the number of samples, M, required to achieve 12 dB of the detector's SNR, which results in $M = 300 \times 10^{(12-5.0)/10} = 1504 = 0.15$ sec, and Eq. (15) is used to compute the threshold τ needed for $P_f = 10^{-4}$. Alternatively, Fig. (5) can be used to find the required threshold value, which turns out to be $\tau = 1300$, assuming that the output of the integrate-and-dump devices is scaled by $1.0/(E_s/N_0)$.

V. Conclusion

This article analyzed the lock detector performance for the QPSK carrier-tracking loop in the ARX II. The analysis is general and is applicable to other QPSK loops. Both analysis and simulation were used to show that the effect of phase jitter on the detector SNR can be significant (as much as two decibels), especially at low-loop SNR.

References

[1] S. Hinedi, "A Functional Description of the Advanced Receiver," *TDA Progress Report 42-100*, vol. October–December 1989, Jet Propulsion Laboratory, Pasadena, California, pp. 131–149, February 15, 1990.

[2] S. Hinedi and B. Shah, "QPSK Carrier-Acquisition Performance in the Advanced Receiver II," *TDA Progress Report 42-100*, vol. October–December 1989, Jet Propulsion Laboratory, Pasadena, California, pp. 150–159, February 15, 1990.

[3] W. Lindsey and M. Simon, *Telecommunications Systems Engineering*, Englewood Cliffs, New Jersey: Prentice Hall, 1973.

[4] M. K. Simon, "Further Results on Optimum Receiver Structure for Digital Phase and Amplitude Modulated Signals," *Conference Record: 1978 International Conference on Communications*, Toronto, Canada, pp. 42.1.1–1.7, June 4–7, 1978.

[5] A. Mileant and S. Hinedi, "Costas Loop Lock Detection in the Advanced Receiver," *TDA Progress Report 42-99*, vol. July–September 1989, Jet Propulsion Laboratory, Pasadena, California, pp. 72–89, November 15, 1989.

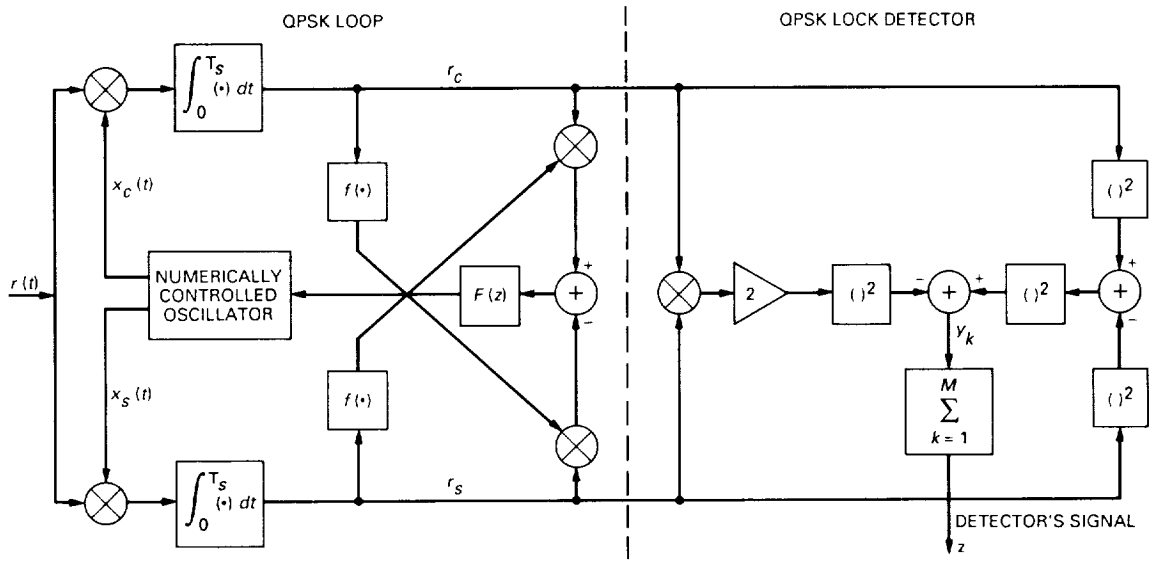


Fig. 1. Implementation of the QPSK loop with lock detector.

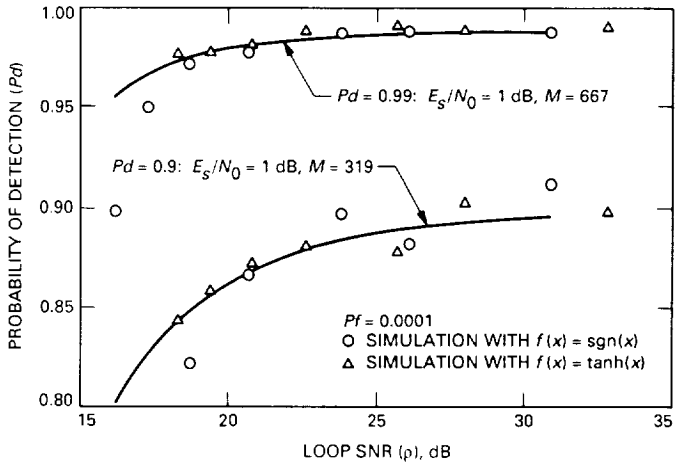


Fig. 2. Probability of detection versus loop SNR.

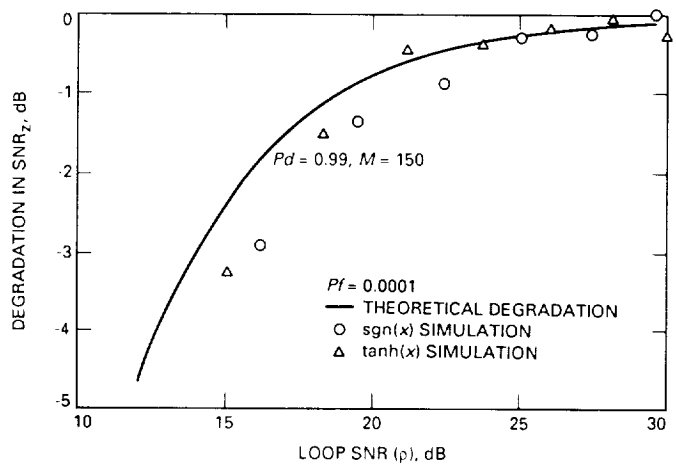


Fig. 3. Degradation in detector SNR versus loop SNR.

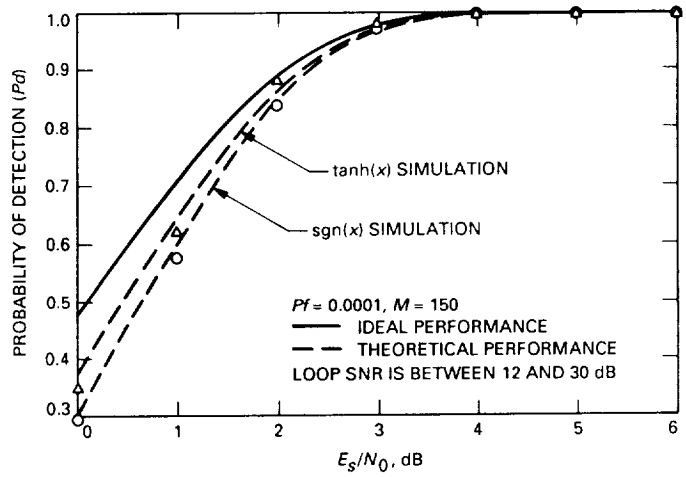


Fig. 4. Probability of detection versus E_s/N_0. (Theoretical performance assumes degradation due to phase jitter in the loop.)

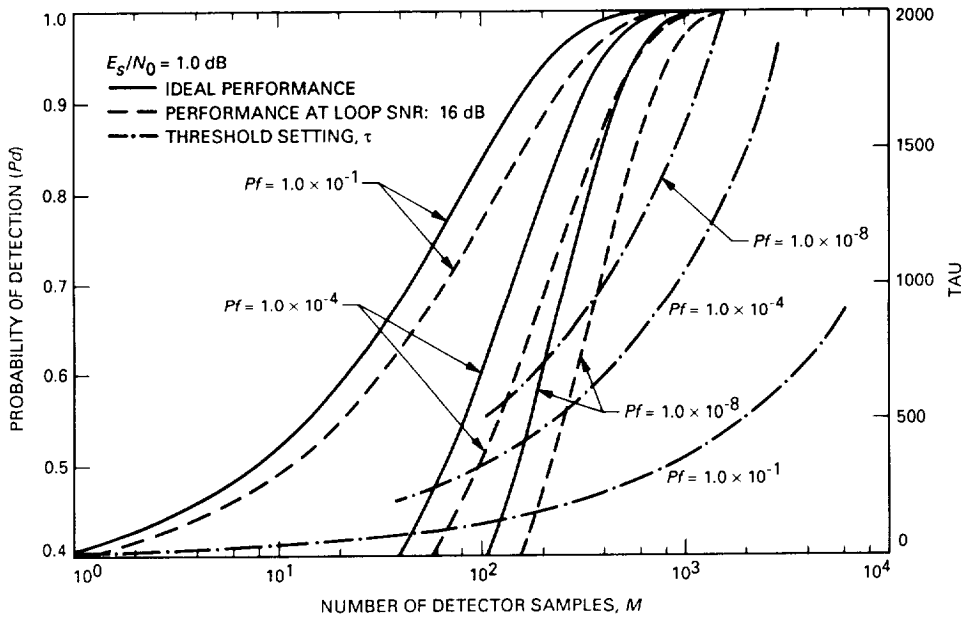


Fig. 5. Probability of detection versus number of detector samples. (The values of τ are given with the assumption that the outputs of the integrate-and-dump devices are scaled by $1/(E_s/N_0)$.)

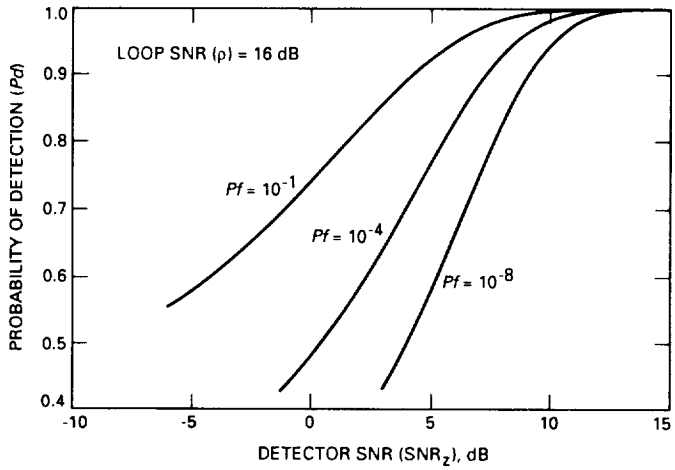


Fig. 6. Probability of detection versus detector SNR.

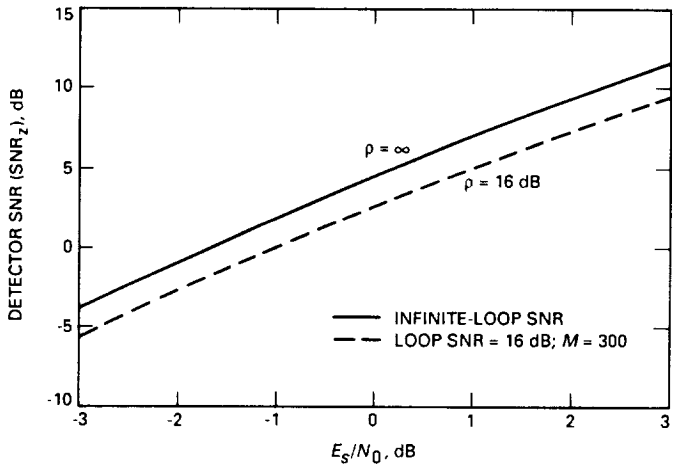


Fig. 7. Detector SNR versus E_s/N_0.

Appendix A
Tracking Performance of Three QPSK Carrier Loops

Depending on the loop, the samples r_c and r_s of Eq. (4) are processed to produce the error signal that forms the input to the loop filter. In general, that error signal can be expressed as

$$\epsilon(\phi) = r_{ck} f(r_{sk}) - r_{sk} f(r_{ck}) \tag{A-1}$$

where $f(x)$ is some nonlinear odd function that defines the loop. Three possible functions will be considered:

$$f(x) = \begin{cases} \operatorname{sgn}(x) \\ x^3 \\ \tanh(x) \end{cases} \tag{A-2}$$

In order to predict the performance of the loops, we need to compute the "squaring loss," S_L, which is defined by

$$S_L \triangleq \frac{(\mathbf{E}\{\epsilon'(0)\})^2}{\mathbf{E}\{\epsilon^2(0)\}} \tag{A-3}$$

where $\epsilon(0)$ and $\epsilon'(0)$ are the error signal and its slope, respectively, evaluated at $\phi = 0$. Note that $\mathbf{E}\{\epsilon^2(0)\} = \overline{N^2}$ equals the variance of the open-loop phase error process, σ_N^2. From this, one obtains the equivalent single-sided noise spectral density, N_0' , using the relation $N_0' = 2\overline{N^2}T_s$, where T_s is the symbol time. The loop responds to the mean of the error signal, which is given by

$$\begin{aligned} \overline{\epsilon(\phi)} = \mathbf{E}\Big\{ R(a_k \sin\phi + b_k \cos\phi) f\left[R\left(a_k \cos\phi - b_k \sin\phi + (-x_{1k}\cos\phi - x_{2k}\sin\phi)/\sqrt{R}\right)\right]\Big\} \\ - \mathbf{E}\Big\{ R(a_k \cos\phi - b_k \sin\phi) f\left[R\left(a_k \sin\phi + b_k \cos\phi + (-x_{1k}\sin\phi + x_{2k}\cos\phi)/\sqrt{R}\right)\right]\Big\} \end{aligned} \tag{A-4}$$

where the expectation is taken first over the random data and then over the thermal noise. Using Eq. (A-4), the expected value of the slope of the error signal can be expressed as

$$\begin{aligned} \overline{\epsilon'(0)} \triangleq \overline{\left.\frac{d\epsilon(\phi)}{d\phi}\right|_{\phi=0}} &= R\mathbf{E}\Big\{ a_k f\left[R(a_k - x_{1k}/\sqrt{R})\right] + b_k f\left[R(b_k + x_{2k}/\sqrt{R})\right]\Big\} \\ &+ R^2\mathbf{E}\Big\{ b_k f'\left[R(a_k - x_{1k}/\sqrt{R})\right](-b_k - x_{2k}/\sqrt{R})\Big\} \\ &- R^2\mathbf{E}\Big\{ a_k f'\left[R(b_k + x_{2k}/\sqrt{R})\right](a_k - x_{1k}/\sqrt{R})\Big\} \end{aligned} \tag{A-5}$$

Averaging the above equation over the random data, one obtains

$$\overline{\epsilon'(0)} = 2R\overline{f[R - \sqrt{R}x]} - 2R^2\overline{f'[R - \sqrt{R}x]} \tag{A-6}$$

where the overbar denotes expectation over the Gaussian random variable x with zero mean and unit variance. After some algebra, we find the expected value of $\epsilon^2(0)$, namely,

$$\mathbf{E}\{\epsilon^2(0)\} = 2R\left[(R+1)\mathbf{E}\{f^2[R-\sqrt{R}x]\} - \left(\mathbf{E}\{xf[R-\sqrt{R}x]\} - \sqrt{R}\mathbf{E}\{f[R-\sqrt{R}x]\}\right)^2\right] \tag{A-7}$$

In order to proceed further, the non-linearity $f(x)$ needs to be specified. For example, $f(x) = \text{sgn}(x)$ ($f'(x) = 2\delta(x)$) for the Costas cross-over loop, in which case we obtain

$$\mathbf{E}\{\text{sgn}(R-\sqrt{R}x)\} = \text{erf}\left(\sqrt{R/2}\right) \tag{A-8a}$$

$$\mathbf{E}\{2\delta(R-\sqrt{R}x)\} = \sqrt{\frac{2}{\pi R}}e^{-R/2} \tag{A-8b}$$

$$\mathbf{E}\{x\,\text{sgn}(R-\sqrt{R}x)\} = -\sqrt{\frac{2}{\pi}}e^{-R/2} \tag{A-8c}$$

$$\mathbf{E}\,\text{sgn}^2(R-\sqrt{R}x)\} = 1 \tag{A-8d}$$

which result in

$$S_L = \frac{\left[\text{erf}\,\sqrt{R/2} - \sqrt{\frac{2R}{\pi}}e^{-R/2}\right]^2}{R+1-\left(\sqrt{R}\,\text{erf}\,\sqrt{R/2} + \sqrt{\frac{2}{\pi}}e^{-R/2}\right)^2} \tag{A-9}$$

On the other hand, for the generalized Costas loop, $f(x) = x^3$ ($f'(x) = 3x^2$) and we have

$$\mathbf{E}\{(R-\sqrt{R}x)^3\} = R^2(R+3) \tag{A-10a}$$

$$\mathbf{E}\{3(R-\sqrt{R}x)^2\} = 3R(R+1) \tag{A-10b}$$

$$\mathbf{E}\{x(R-\sqrt{R}x)^3\} = -R^{3/2}(R+1) \tag{A-10c}$$

$$\mathbf{E}\{f^2(x)\} = R^3(R^3+15R^2+45R+15) \tag{A-10d}$$

$$S_L = \frac{1}{1+\frac{9}{2R}+\frac{6}{R^2}+\frac{3}{2R^3}} \tag{A-11}$$

In the MAP estimation loop, the nonlinearity is given by $f(x) = \tanh(x)$ ($f'(x) = \text{sech}^2(x)$), which prevents a closed-form expression, i.e.,

$$S_L = \frac{\left[\overline{\tanh(R-\sqrt{R}x)} - R\overline{\text{sech}^2(R-\sqrt{R}x)}\right]^2}{(R+1)\overline{\tanh^2(R-\sqrt{R}x)} - \left[\overline{x\tanh(R-\sqrt{R}x)} - \sqrt{R}\,\overline{\tanh(R-\sqrt{R}x)}\right]^2} \tag{A-12}$$

The variance of the phase error process in all three loops can be expressed as

$$\sigma_\phi^2 = \frac{N_0' B_L}{[\epsilon'(0)]^2} = \frac{1}{\rho_c S_L} = \frac{B_L}{r_{sy} R S_L} \tag{A-13}$$

where $\rho_c \triangleq \frac{P_D}{N_0 B_L}$, B_L is the loop bandwidth, r_{sy} is the symbol rate, and $R = E_s/N_0$ is the symbol SNR. Note that for $R \approx 1$ dB, $S_L \ll 1$. We should comment that the above variance σ_ϕ^2 is the variance relative to ω_0, which is used to compute the telemetry performance. The variance that determines acquisition and the cycle slip process is relative to $4\omega_0$, and hence will be 16 times larger.

Appendix B
Error and Detector Signals for MPSK

Let r_c be the "in-phase" and r_s the "quadrature" outputs of the integrate-and-dump devices of a MPSK loop: Ref. [4] proposes the following algorithms for deriving the error signal and the detector's signals for the various MPSK tracking loops.

Derivation algorithms

M	Error signal	Detector signal
2	$\epsilon_2 = r_c \times r_s$	$d_2 = r_c^2 - r_s^2$
4	$\epsilon_4 = d_2 \times \epsilon_2$	$d_4 = d_2^2 - \epsilon_2^2$
8	$\epsilon_8 = d_4 \times \epsilon_4$	$d_8 = d_4^2 - \epsilon_4^2$
N	$\epsilon_N = d_{\frac{N}{2}} \times \epsilon_{\frac{N}{2}}$	$d_N = d_{\frac{N}{2}}^2 - \epsilon_{\frac{N}{2}}^2$

Appendix C
QPSK Lock-Detector Statistics

The signal for the lock detector is derived using the following algorithm

$$z = \sum_{k}^{M} y_k \gtrless \tau$$

where

$$y_k \triangleq \left(r_{ck}^2 - r_{sk}^2\right)^2 - \left(2r_{ck}r_{sk}\right)^2 \tag{C-1}$$

In order to compute the mean and the variance of z, we need the second through the eighth moments of r_{ck} and r_{sk}. The odd moments are all zero (since the noise is Gaussian), and hence only the even moments need to be computed. To simplify the notation, the subscript k will be dropped from the next several equations. Taking the expected value of Eq. (4), the even moments are given by

$$m_2 \triangleq \overline{r_c^2} = \overline{r_s^2} = R^2\left(1 + \frac{1}{R}\right) \tag{C-2}$$

$$m_4 \triangleq \overline{r_c^4} = \overline{r_s^4} = R^4\left(r - \frac{1}{2}\overline{\cos 4\phi}\right) \tag{C-3}$$

$$m_6 \triangleq \overline{r_c^6} = \overline{r_s^6} = R^6\left(\frac{1}{2}(5 - 3\overline{\cos 4\phi}) + \frac{15}{2R}(3 - \overline{\cos 4\phi}) + \frac{15}{R^2}\left(3 + \frac{1}{R}\right)\right) \tag{C-4}$$

$$\begin{aligned} m_8 \triangleq \overline{r_c^8} = \overline{r_s^8} = R^8\Bigg[&\frac{1}{16}\left(105 - 5\overline{\cos 4\phi} - 84\overline{\cos 4\phi}\right) \\ &+ \left(\frac{7}{2R}(5 + 3\overline{\cos 4\phi}) + 15(1 - \overline{\cos 4\phi})\right)\left(\frac{105}{2R^2}(3 + \overline{\cos 4\phi}) + 3(1 - \overline{\cos 4\phi})\right) + \frac{420}{R^3} + \frac{105}{R^4}\Bigg] \end{aligned} \tag{C-5}$$

where

$$r = 3\left(\frac{1}{2} + \frac{2}{R} + \frac{1}{R^2}\right) \tag{C-6}$$

At high-loop SNR, $\phi \rightarrow 0$ and m_4, m_6, m_8 reduce to

$$R^4\left(1 + \frac{6}{R} + \frac{3}{R^2}\right) \tag{C-7}$$

$$R^6\left(1 + \frac{15}{R} + \frac{45}{R^2} + \frac{15}{R^3}\right) \tag{C-8}$$

$$R^8\left(1 + \frac{28}{R} + \frac{210}{R^2} + \frac{420}{R^3} + \frac{105}{R^4}\right) \tag{C-9}$$

respectively. Cross-moments for $i = j$ are also needed and are given by

$$m_{22} \triangleq \mathbf{E}\{r_c^2 r_s^2\} = R^4 \left(\frac{1}{2}(1 + \overline{\cos 4\phi}) + \frac{2}{R} + \frac{1}{R^2} \right) \tag{C-10a}$$

$$m_{44} \triangleq \mathbf{E}\{r_c^4 r_s^4\} = R^8 \left(r^2 + \frac{1}{8}(1 + \overline{\cos 8\phi}) - r\overline{\cos 4\phi} - \frac{18}{R^2}(1 - \overline{\cos 4\phi}) \right) \tag{C-10b}$$

$$m_{62} \triangleq \mathbf{E}\{r_s^6 r_c^2\} = \mathbf{E}\{r_c^6 r_s^2\} = R^8 \left(-5 + 6\overline{\cos 4\phi} + \frac{1}{R}(25 - 9\overline{\cos 4\phi}) + \frac{15}{2R}(9 - \overline{\cos 4\phi}) + \frac{60}{R^3} + \frac{15}{R^4} \right) \tag{C-10c}$$

Note that as $\sigma_\phi \to 0$, $m_{22} \to m_2^2$, $m_{44} \to m_4^2$, and $m_{62} \to m_6 m_2$. Now, the mean value and the variance of the detector's signal can be computed. The mean is given by

$$\mu_z = \sum_{k=1}^{M} \mathbf{E}\{y_k\} = M\mu_y \tag{C-11}$$

where

$$\mu_y = \overline{r_c^4} + \overline{r_s^4} - 6\overline{r_c^2 r_s^2} \tag{C-12}$$

Using the moments just obtained gives

$$\mu_y = -4R^4\overline{\cos 4\phi} \tag{C-13}$$

The next step is to find the moment of z^2. Equation (C-1) gives

$$z^2 = \left(\sum_{i=1}^{M} r_{ci}^4 + \sum_{i=1}^{M} r_{si}^4 - 6\sum_{i=1}^{M} r_{ci}^2 r_{si}^2 \right)^2 = \sum_{i=1}^{M}\sum_{j=1}^{M} r_{ci}^4 r_{cj}^4 + \sum_{i=1}^{M}\sum_{j=1}^{M} r_{si}^4 r_{sj}^4$$
$$+ 36\sum_{i=1}^{M}\sum_{j=1}^{M} r_{ci}^2 r_{cj}^2 r_{si}^2 r_{sj}^2 + 2\sum_{i=1}^{M}\sum_{j=1}^{M} r_{ci}^4 r_{sj}^4 - 12\sum_{i=1}^{M}\sum_{j=1}^{M} r_{ci}^4 r_{cj}^2 r_{sj}^2 - 12\sum_{i=1}^{M}\sum_{j=1}^{M} r_{si}^4 r_{cj}^2 r_{sj}^2 \tag{C-14}$$

Let $z^2 = S_1 + S_2$, where S_1 is the sum of all the terms when $i = j$ and S_2 is the sum of all the terms for $i \neq j$. Then for $i = j$

$$\overline{r_{ci}^4 r_{cj}^4} = \overline{r_{si}^4 r_{sj}^4} = m_8 \tag{C-15a}$$

$$\overline{r_{ci}^4 r_{sj}^4} = m_{44} \tag{C-15b}$$

$$\overline{r_{ci}^6 r_{sj}^2} = \overline{r_{si}^6 r_{cj}^2} = m_{62} \tag{C-15c}$$

and

$$S_1 = 2M(m_8 + 19m_{44} - 12m_{62}) \tag{C-16}$$

C-3

For $i \neq j$, let $k \triangleq \mid i-j \mid$. Note that $\overline{r_{ci}^2 r_{cj}^2} = m_2^2$ because the random data make the second moments of r_{ci}^2 independent of r_{cj}^2. The double sums can be converted into single sums as follows:

$$S_2 = 2\sum_{k=1}^{M-1} c_k \overline{r_{ci}^4 r_{cj}^4} + 36\sum_{k=1}^{M-1} c_k \overline{r_{ci}^2 r_{cj}^2 r_{si}^2 r_{sj}^2} + 2\sum_{k=1}^{M-1} c_k \overline{r_{ci}^4 r_{sj}^4} - 24\sum_{k=1}^{M-1} c_k \overline{r_{ci}^4 r_{cj}^2 r_{sj}^2}$$

$$= 4\sum_{k=1}^{M-1} c_k d_k + 36(M-1)Mm_2^4 - 24(M-1)Mm_4 m_2^2 \tag{C-17}$$

where $c_k = 2(M-k)$ for $k = 1, 2, \ldots M-1$. Note that $\sum_{k=1}^{M-1} c_k = M^2 - M$ (and $c_0 = M$). The variance of z can now be found from the relation

$$\sigma_z^2 = \overline{z^2} - (\overline{z})^2 = S_1 + S_2 - \mu_z^2 \tag{C-18}$$

Using the previous equations, it can be shown that

$$\sigma_z^2 = 2M\left[m_8 + 19m_{44} - 12m_{62}\right] + 4M\left[\frac{1}{M}\sum_{k=1}^{M-1} c_k d_k - Mm_4^2 - 9m_2^4 + 6m_4 m_2^2\right] \tag{C-19}$$

where

$$d_k \triangleq \overline{x_{ci}^4 x_{cj}^4}$$

such that

$$\begin{aligned} k &= \mid i-j \mid \\ &= \overline{\left(r - \frac{1}{2}\cos 4\phi_i\right)\left(r - \frac{1}{2}\cos 4\phi_j\right)} \\ &= (r^2 - re^{-8\sigma_\phi^2} + f_k/4) \end{aligned} \tag{C-20}$$

and

$$f_k \triangleq \overline{\cos 4\phi_i \cos 4\phi_j} \tag{C-21}$$

At high-loop SNR, the mean and variance become

$$\mu_{z,ideal} = -4MR^4 \tag{C-22a}$$

and

$$\sigma_{z,ideal}^2 = 64MR^8\left(\frac{2}{R} + \frac{9}{R^2} + \frac{12}{R^3} + \frac{3}{R^4}\right) \tag{C-22b}$$

which result in

$$SNR_{z,ideal} = \frac{MR}{4(2 + \frac{9}{R} + \frac{12}{R^2} + \frac{3}{R^3})} \tag{C-23}$$

So far, no assumption has been made regarding the statistics of the phase error process. Assuming that the phase error is indeed Gaussian with second-order density,

$$\begin{aligned} p(\phi_i, \phi_j, \tau) &\approx \frac{1}{2\pi\sqrt{R^2(0) - R^2(\tau)}} \exp\left(-\frac{R(0)\phi_i^2 - 2R(\tau)\phi_i\phi_j + R(0)\phi_j^2}{2(R^2(0) - R^2(\tau))}\right) \\ &= \frac{1}{2\pi\sigma_\phi^2\sqrt{1 - C^2(\tau)}} \exp\left(-\frac{\phi_i^2 - 2C(\tau)\phi_i\phi_j + \phi_j^2}{2\sigma_\phi^2(1 - C^2(\tau))}\right) \end{aligned} \tag{C-24}$$

where $R(\tau)$ is the correlation function and can be expressed in terms of the normalized correlation function $C(\tau)$ as $R(\tau) = \sigma_\phi^2 C(\tau)$ where

$$C(\tau_k) = \left(1 - \frac{\mid B_L\tau_k \mid}{0.91}\right) \exp(-1.25B_L\tau_k) \tag{C-25}$$

B_L is the one-sided loop bandwidth and σ_ϕ^2 the closed-loop error variance.

$$\tau_k \triangleq \tau_{ij} = T_s \mid i - j \mid = \mid t_i - t_j \mid \tag{C-26}$$

With this Gaussian assumption, the expected value of $\cos b\phi$ will be $e^{\frac{-b^2\sigma_\phi^2}{2}}$. Hence, we can substitute in all the above equations $\overline{\cos 4\phi}$ with $e^{-8\sigma_\phi^2}$ and $\overline{\cos 8\phi}$ with $e^{-32\sigma_\phi^2}$. Moreover,

$$\overline{\cos 4\phi_i \cos 4\phi_j} = e^{-16\sigma_\phi^2} \cosh(16\sigma_\phi^2 C(\tau_k)), \quad i \neq j \tag{C-27}$$

Long-Range Planning Cost Model for Support of Future Space Missions by the Deep Space Network

J. S. Sherif
Software Product Assurance Section
and
California State University at Fullerton

D. S. Remer
TDA Planning Section
and
Harvey Mudd College of Engineering and Science

H. R. Buchanan
Radio Frequency and Microwave Subsystems Section

This article suggests a simple model to do long-range planning cost estimates for Deep Space Network (DSN) support of future space missions. The model estimates total DSN preparation costs and the annual distribution of these costs for long-range budgetary planning. The cost model is based on actual DSN preparation costs from four space missions: Galileo, Voyager (Uranus), Voyager (Neptune), and Magellan. The model was tested against the four projects and gave cost estimates that range from 18 percent above the actual total preparation costs of the projects to 25 percent below.

The model was also compared to two other independent projects: Viking and Mariner Jupiter/Saturn (MJS later became Voyager). The model gave cost estimates that range from 2 percent (for Viking) to 10 percent (for MJS) below the actual total preparation costs of these missions.

A rule of thumb based on these six missions is that the average annual DSN preparation cost is $7.2 million in 1987 dollars.

I. Introduction

Many times, the Office of Telecommunications and Data Acquisition (TDA) is required to provide quick, rough budgetary estimates for potential future projects. Because of the lack of definition at the very early stages, it has been very difficult to make meaningful long-range planning estimates. The purpose of this modeling effort is to improve the process for these estimates by providing supportable values in proper context, and thereby gaining time for developing the carefully thought out and reviewable cost estimates that are required before cost commitments are made.

This section describes the objectives of this article and gives a brief description of the Deep Space Network (DSN), its role in supporting the National Aeronautics and Space Administration (NASA) exploration of space, and its basic services to space missions. Then, an overview of the article is presented.

A. Study Objective

The objective of this study is to develop a model that can be used in the early planning stages to estimate both the DSN cost to prepare for future space-mission support and the allocation of these costs over the life of the project. DSN preparation costs for a mission occur when minimum requirements for a supported new mission lie outside the installed DSN capability. This proposed model is useful for long-range budgetary planning, but does not replace the need for a detailed cost estimate. Also, the model provides "ballpark" numbers that can be used to check detailed "grass roots" estimates. The original approach to this study was to develop a mathematical model for the cost of a DSN project, to be a function of time and of various cost categories such as a new uplink frequency, a telemetry upgrade, etc. However, as the data analysis began to develop, it was noted that a simpler model using only the length of the project as a parameter could effectively represent the data.

B. The Deep Space Network

The NASA DSN is a multimission telecommunications and radio-metric data facility used to support NASA's exploration of space, research in space science, and advanced technology investigations. The Network has facilities located on three continents (North America, Europe, and Australia), with tracking complexes at intervals of 120 degrees of longitude. The Network's basic services are (1) reception of telemetry from spacecraft, (2) transmission of commands to spacecraft, (3) measurement of radio-metric data for spacecraft navigation, and (4) radio science measurements.

C. Overview of Article

In Section II, the purpose of each of the four space missions is summarized, and the TDA modifications that were required to support these four missions are described. The methodology for collecting the data and the cost history are summarized in Section III. The cost models developed using this cost data are presented in Section IV. In Section V, the results from the models are compared to the actual data. In Section VI, the use of the model for future-cost estimates for budgetary planning is described. Finally, as an "external" check, in Section VII, the model is compared to two independent projects, Viking and Mariner Jupiter/Saturn (MJS).

II. Background of Space Missions

Presented in this section is a brief overview of the four missions that were analyzed for DSN cost modeling: Galileo, Voyager (Uranus), Voyager (Neptune), and Magellan.

A. Galileo Space Mission

The Galileo spacecraft is a Jupiter orbiter that includes a probe for penetrating the atmosphere of Jupiter. Galileo's launch was planned for 1986, but due to the Challenger (STS-51L) loss, the launch date was delayed to 1989. This delay has resulted in a skewed cost profile for the DSN preparation for the mission. The primary objectives of the Galileo space mission are to investigate the chemical composition and physical state of Jupiter's atmosphere and satellites and to study the structure and dynamics of Jupiter's magnetic field. This space mission is the first outer-planet mission that will (1) perform a detailed observation of Jupiter's system, (2) send an orbiter, which has a probe to penetrate the atmosphere of Jupiter and which will provide 22 months of orbital operations to map Jupiter's surface, and (3) use a dual-spin spacecraft and a complex Venus-Earth-Earth-Gravity-Assist (VEEGA) trajectory. The spacecraft carries 19 instruments, 12 by the orbiter and 7 by the probe. Galileo was launched on October 18, 1989 [1,2,3].

The major improvements to the DSN associated with the Galileo mission are narrow channel bandwidth (NCB), very long baseline interferometry (VLBI) equipment at the 70-meter stations, and both NCB and wide channel band-

width (WCB) VLBI equipment at the 34-meter stations. These provide the needed improvement in the navigation capabilities. Other significant upgrades in the radio metrics and radio science categories were the Block II Meteorological Monitor Assembly (MMA), improved frequency standards, and a variable second local oscillator for the Multimission Receiver (MMR).[1]

B. Voyager (Uranus) and Voyager (Neptune)

Voyager 2 was launched on August 20, 1977, on a flight path that would allow it to become the first spacecraft from Earth to observe the planets Uranus and Neptune. In this study, the DSN preparation costs for the Voyager (Uranus) and the Voyager (Neptune) missions have been considered separately from the costs for the basic MJS mission. Voyager 2 made its closest approach to Uranus on January 24, 1986, passing within 110,000 kilometers of the planet's center. The main objective of Voyager (Uranus) was to provide a basic characterization of Uranus, its satellites, and its rings, which it did very well [4,5].

On August 24, 1989, Voyager (Neptune) sailed over the north pole of Neptune, within about 4850 kilometers of the visible cloud tops. The Neptune encounter was Voyager 2's closest encounter with any object in its 12-year trip to the outer solar system. The objectives of this mission were to provide data on magnetic fields and charged particles at Neptune, to probe deep into Neptune's atmosphere with Voyager's radio waves, and to search for new rings and satellites.

One of the major improvements to the DSN needed for the Voyager-Uranus encounter was the decrease in receiver threshold to compensate for the ever-increasing distance of the spacecraft from Earth. This effort included: the 34-m/64-m array, the Parkes Observatory equipment, and DSN arraying in Australia; and new 34-m antenna/microwave/low-noise amplifier equipment and baseband-combining equipment. A new 400-kW transmitter for uplink commanding was also provided for spacecraft-emergency purposes. New telemetry formats were accommodated with new capabilities in software and hardware for correlation of received data. Upgrades in the frequency standards and coherent reference generators (CRGs) were made. Open-loop receiver hardware was provided.

The Voyager-Neptune encounter preparations for the DSN carried on many of the same tasks as were required for the Uranus encounter. The 64-m antennas were expanded to 70 m, including an improved-precision antenna reflector surface for both S- and X-band improved performance. Improvements were made to the X-band low-noise maser amplifiers, and antenna-mounted cold backup X-band masers were installed on the 34-m high-efficiency (HEF) antennas. The planetary ranging assembly (PRA) computers were replaced and a precision power monitor (PPM) was implemented at Signal Processing Center, SPC-60. A new monitor and control system and noise-adding radiometer were designed and supplied for the Parkes antenna in support of the Neptune encounter. Also, new additions to the total array were implemented—one incorporating the Japanese 64-m antenna at Usuda and the other the National Radio Astronomy Observatory (NRAO) radio astronomy Very Large Array (VLA) in New Mexico. JPL supplied the 27 VLA antennas with new X-band feed horns and dual-channel X-band solid-state low-noise amplifiers. The Usuda antenna was supplied with an ultra-low-noise maser amplifier and backup by the Jet Propulsion Laboratory (JPL).[1]

C. Magellan Space Mission

The Magellan space mission to Venus was originally conceived as the Venus Orbiting Imaging Radar (VOIR). As first envisioned in 1980, the VOIR spacecraft was to carry a high-resolution synthetic aperture radar (SAR) and six other instruments, most of them for atmospheric studies. The VOIR was de-scoped to the single-instrument Venus Radar Mapper (VRM) in 1984 and then renamed Magellan in 1986. It is managed for NASA by JPL.

The primary objective of this mission is to investigate the origin and evolution of Venus by obtaining a global radar image of the planet. The spacecraft will perform two types of investigations: radar and gravity. The radar investigations will produce (1) continuous images of at least 70 percent of the planet with no systematic gaps except for one pole and with a surface resolution of at least 1 kilometer and (2) a global topographic map with a range of resolution commensurate with the SAR range of resolution. The gravity investigation will measure the distribution of gravity potential around Venus [6,7].

The new requirements on the DSN imposed by the Magellan mission include the modifications (seven subsystems) for providing an operational 20-kW X-band uplink at three 34-m stations. High-density recording systems will be implemented to accommodate the Magellan data rates. The baseband assemblies are being upgraded. Implementation of hardware transfer-level frame synchronization of telemetry data is being done. Magellan-related

[1] "Narratives covering FY'82, FY'86, and FY'88," TDA work authorization documents (WADs) (internal documents), Jet Propulsion Laboratory, Pasadena, California, 1983, 1987 and 1989.

predicts capability is being incorporated in the network support subsystem (NSS).

The Block IV receiver at Compatibility Test Area 21 (CTA 21) was upgraded to accommodate X-band Doppler. The closed-loop receivers in the 34-m and 70-m subnets have been provided with rapid acquisition capability to accommodate the multiple Magellan acquisitions during each scheduled track, due to planetary occultations. Engineering and equipment are being provided for various VLBI upgrades, including an improved radio source catalog and modifications to the delta VLBI software to accommodate the low Sun-Earth-Probe (SEP) angles.[2]

III. Cost Data

The annual cost obligations used in this article are taken from Telecommunications and Data Acquisition (TDA) Work Authorization Documents (WAD Obligations Performance Reports), and do not include construction of facilities cost, spacecraft cost, transportation cost, and/or other logistics costs.[3] All costs used in this article are adjusted for inflation to 1987 dollars using the NASA inflation index. The DSN costs for each project are collected into the following subsystem upgrade categories:

(1) (M/O) maintenance and operations

(2) (D/L) downlink frequency

(3) (U/L) uplink frequency

(4) (TEL) telemetry upgrade

(5) (G/T) gain over system-noise temperature

(6) (CMR) upgrade command rate

(7) (CMP) upgrade the effective radiated power

(8) (R/M) radiometric accuracy upgrade

(9) (R/S) radio-science stability upgrade

(10) (VLBI) very long baseline interferometry system

(11) (OTH) other

Inspection of the 11 category costs revealed that for each mission, costs are primarily assignable to three or four categories. This is summarized in graphical form in Fig. 1. If a future mission calls for a significant upgrade in a particular category, these historical category data can be helpful in predicting mission costs.

The total cost data for the four space missions, Galileo, Voyager (Uranus), Voyager (Neptune), and Magellan, is summarized in Table 1. For example, Table 1 shows for Galileo the annual costs from fiscal year (FY) 1979 to FY 1988, mission total cost ($47,186K), annual average cost ($4,719K), the standard deviation of annual costs ($2,673K), the maximum annual cost ($8,275K), the minimum annual cost ($1,052K), and the range ($7,223K). A comparison of the four space missions gives the following results:

Mission	Total DSN preparation cost, $M	Average yearly cost, $M	Maximum yearly cost, $M
Galileo	47.2	4.7	8.3
Voyager (U)	35.9	7.2	13.1
Voyager (N)	36.0	9.0	12.3
Magellan	32.5	8.1	13.8
Average	37.9	7.3	11.8
Standard deviation	6.3	1.8	2.4

It can be seen that the average yearly cost and also the maximum yearly cost are fairly close for the last three projects.

The TDA costs for DSN preparation that are covered in this article are relatively small as compared to the total mission costs, as shown below [2,6,8].

Mission	Total mission costs: 1987, $M	DSN preparation costs: 1987, $M	DSN/total, %
Galileo	1006 [2]	47.2	4.7
Voyager (U) Voyager (N) Voyager (JS)	841.2 [8]	169.4	20.1
Magellan	413 [6]	32.5	7.8

[2] "Narratives covering FY'82, FY'86, and FY'88," TDA work authorization documents (internal documents).

[3] "Obligations Performance Reports, 1972–1988," TDA work authorization documents (WADs) (internal documents), Jet Propulsion Laboratory, Pasadena, California, 1973–1989.

IV. Cost Models

A number of models were compared to the cost data. The Rayleigh distribution [9] did a good job of describing the DSN preparation cost for each of the four space missions and also for the composite, which is the average cost of the four missions.

A. Galileo Space Mission

The total DSN preparation cost for Galileo was fitted to the Rayleigh distribution, and the following model was obtained:

$$Y_t = 1983t \exp(-0.0224t^2)$$

where Y_t is the cost for DSN preparation to support Galileo in year t and $(t = 1, 2, \ldots, n)$, t is the number of the year in the life of the DSN preparation for Galileo support, n is the total years of the DSN preparation, and the total cost of the DSN preparation is $\sum Y_t$. The model shows a coefficient of determination (R^2) of 71 percent (R^2 indicates the precision of the model, or the amount of variability in the total cost that can be explained by the model [10]). The cost data and the costs predicted by the model are shown in Table 2 and Fig. 2.

B. Voyager (Uranus), Voyager (Neptune), and Magellan Space Missions

Cost models were developed for DSN preparation for three other space missions, Voyager (Uranus), Voyager (Neptune), and Magellan. The Magellan model follows a different form since its annual costs are still increasing (see Section IV.C.2). These models and their R^2s are

Mission	Model	R^2, %
Galileo	$Y_t = 1983t \exp(-0.0224t^2)$	71
Voyager (U)	$Y_t = 6583t \exp(-0.0861t^2)$	92
Voyager (N)	$Y_t = 4879t \exp(-0.0308t^2)$	75
Magellan	$Y_t = -9358 + 10574t - 1193t^2$	99

C. Composite Model

1. Total mission period. A general long-range planning cost model for support of future space missions by the DSN has also been developed. This model is called the "composite model," since it uses the average annual DSN preparation cost data of the above four space missions. The data from Table 1 are rearranged to a "year of mission" format in Table 3 and averaged by year as shown in Table 4. The resulting composite model is

$$Y_t = 3613t \exp(-0.0311t^2) \quad (1)$$

and the total cost of a mission is

$$Y(\text{total}) = \sum Y_t \quad (2)$$

The composite model has a goodness of fit (R^2) of 87 percent, and, therefore, could be used to give cost estimates for preliminary budgetary planning for future space missions. Table 3 and Fig. 3 show the actual cost data of the four missions and costs as predicted by the model. Table 4 and Fig. 4 show the actual composite average cost data of the four missions and costs as predicted by the model.

2. Growth period. During the growth period (or early stages) of DSN preparation for a space mission, the DSN annual preparation costs are generally increasing—for example, during the first two years of Voyager (U), the first three years of Voyager (N), the first four years of Magellan, and the first five years of Galileo. To get better accuracy in predicting the DSN annual preparation costs during the growth period, the following model is used:

$$Y_t = -4738 + 8786t - 1322t^2 \quad (3)$$

where Y_t is the cost for DSN to support the space mission in year t. The growth model shows a coefficient of determination R^2 of 99 percent. A comparison of the DSN actual average annual preparation costs of the four missions versus those same costs as predicted by the growth model is as follows:

Year, Y_t	Average annual cost: actual, $K	Average annual cost: model, $K	Model minus actual, Δ in $K	Error, %, Δ/actual
1	2652	2726	+ 74	+ 3
2	7769	7546	− 223	− 3
3	9501	9722	+ 221	+ 2
4	9331	9254	− 77	− 1
Average/ year	7313	7312	− 1.25	+ 0.25

A comparison of the composite model derived from all four missions, Eq. (1), the growth model, Eq. (3), and the actual average annual costs of the four missions for the first four years gives the following:

Year, Y_t	Annual cost: composite model, $K	Annual cost: growth model, $K	Actual average annual cost, $K
1	3502	2726	2652
2	6381	7546	7769
3	8193	9722	9501
4	8787	9254	9331
Average/ year	6716	7312	7313

On the average, the annual cost for the DSN to support a mission during the growth period is about $7 million.

V. Back-Testing the Composite Model

The composite model was checked with four missions: Galileo, Voyager (U), Voyager (N), and Magellan. The actual total preparation costs of the missions and those costs as predicted by the composite model are shown below:

Space mission	Preparation cost: actual, $M	Total cost: model, $M	Model minus actual, Δ in $M	Error, %, Δ/ actual
Galileo	47.2	55.9	+ 8.7	+ 18.0
Voyager (U)	35.9	35.2	− 0.7	− 1.9
Voyager (N)	36.0	26.9	− 9.1	− 25.0
Magellan	32.5	26.9	− 5.6	− 17.0
Average/ mission	37.9	36.3	− 1.6	

The actual preparation costs for Voyager (U), Voyager (N), and Magellan were larger than those predicted by the model. However, Galileo actual preparation costs were less than the model predicted. This is probably a result of the Galileo launch slipping from 1986 to 1989 because of the Challenger loss. Therefore, a larger value for n was used in the model, due to this slippage. On the average, the difference between a mission actual total preparation cost and that predicted by the composite model is about $1.6 million, or 4 percent of actual total preparation cost.

The difference between the actual average annual cost and the predicted average annual cost is $730,000, as shown below.

Space mission	Average annual cost: actual, $M	Average annual cost: model, $M	Model minus actual, Δ in $M	Error, %, Δ/ actual
Galileo	4.72	5.59	+ 0.87	+ 18.0
Voyager (U)	7.18	7.04	− 0.14	− 1.9
Voyager (N)	9.00	6.72	− 2.27	− 25.0
Magellan	8.12	6.72	− 1.39	− 17.0
Annual grand average	7.25	6.50	− 0.73	

VI. How to Use the Model

The long-range planning cost model for support of future space missions by the DSN is developed from historical cost data as a composite cost average of four space missions: Galileo, Voyager (Uranus), Voyager (Neptune), and Magellan. The model is

$$Y_t = 3613t \exp(-0.0311t^2)$$

where Y_t is the cost in year t for DSN preparation to support a mission ($t = 1, 2, \ldots, n$), and n is the total number of years for DSN preparation. The total DSN preparation cost is $= \sum Y_t$.

For example, to estimate the DSN preparation budget for a future project, sum the annual preparation costs predicted by the model over the life of the project (n), as shown below:

Year, t	Composite model annual predicted cost, Y_t, \$M	Predicted DSN preparation total cost $\sum Y_t$, n years						
		4	5	6	7	8	9	10
1	3.5							
2	6.4							
3	8.2							
4	8.8	26.9						
5	8.3		35.2					
6	7.1			42.3				
7	5.5				47.8			
8	3.9					51.7		
9	2.6						54.3	
10	1.6							55.9

For example, a DSN preparation of four years is predicted to have a total cost of \$26.9 million, and a total cost of \$35.2 million is predicted for a DSN preparation of five years, and so on. It should be noted that since the historical data covered DSN tasks from $n = 4$ to 10 years, extrapolating outside that range of years should be avoided. It should also be cautioned that there are missions differing considerably in scope and/or complexity from the mission set analyzed here.[4] In such cases the required DSN equipment might well mean a cost greater (or less) than that indicated by the model. Obviously, good judgment must be used.

VII. Results

A. Comparison With Other Independent Missions

The composite model developed from the four missions was also tested against two other independent missions, Viking and Mariner Jupiter/Saturn (MJS, later renamed Voyager).

DSN preparation time for Viking was $7\frac{1}{4}$ years—from 1971 through 1977, including the fiscal year transition quarter of 1976–1977. Using the table in Section VI, \$47.8 million is obtained for $n = 7$ years, and one-fourth of the difference between $n = 7$ and $n = 8$ is added, to get \$1 million. The total predicted cost from the model is therefore \$48.8 million, as compared to the actual cost of \$49.7 million.[5,6]

DSN preparation time for MJS was harder to define. It was recognized before the Jupiter encounter that the MJS mission would require a significant enhancement of the received signal for the Saturn encounter. It appears reasonable to apportion the DSN preparation costs for the MJS mission into two phases: (1) $8\frac{1}{4}$ years' duration (1972 through 1979, including the 1976/1977 fiscal year transition quarter), and (2) 5 years' duration (1977 through 1981). Using this model of MJS, the total predicted cost is \$87.6 million. The actual value was \$97.5 million.[5,6]

Note that the annual average of these two projects is \$7.15 million, which is essentially the same as the \$7.25 million average for the four projects used to develop the model. Therefore, a rule-of-thumb is that DSN preparation for a mission has an annual average preparation cost of \$7.2 million in 1987 dollars. The results are summarized below.

Space mission	Total cost: actual, \$M	Total cost: model, \$M	Model minus actual, Δ in \$M	Error, %, Δ/ actual
MJS	97.5[5,6]	87.6	− 9.9	− 10.2
Viking	49.7[5,6]	48.8	− 0.9	− 1.8

A comparison between the actual average annual preparation cost and the predicted average annual cost of the missions is \$500,000, as shown below.

Space mission	Average annual cost: actual, \$M	Average annual cost: model, \$M	Model minus actual, Δ in \$M
MJS	7.4	6.6	− 0.8
Viking	6.9	6.7	− 0.2
Annual grand average	7.15	6.65	− 0.50

[4] J. W. Layland, private communication.

[5] "Narratives covering FY'82, FY'86, and FY'88," TDA work authorization documents (internal documents).

[6] J. W. Layland, private communication.

B. Total DSN Preparation Cost Composite Model

The composite model for total DSN preparation cost presented here is obviously a simple model that has only time duration as a parameter. However, the model does a reasonable job for representing the actual preparation costs for Galileo, Voyager (U), Voyager (N), and Magellan and also for the two independent projects Mariner Jupiter/Saturn (MJS) and Viking. This model could be used for long-range planning cost estimates for budgetary planning of DSN support of future space missions. The model can also be used to check "grass roots" detailed cost estimates. Based on our back-testing the actual four projects against the model, the results are in the range of 18 percent above actual costs to 25 percent below actual costs. This model should only be used for projects comparable in scope (4 to 10 years, and $25 million to $55 million).

References

[1] *Project Galileo*, JPL Fact Sheet, Jet Propulsion Laboratory, Pasadena, California, September 14, 1988.

[2] *Space Exploration: Cost, Schedule and Performance of NASA's Galileo Mission to Jupiter*, GAO/NSIAD-88-138FS, U.S. General Accounting Office, Washington, D.C., May 27, 1988.

[3] *The Galileo Messenger*, JPL 410-16, issues 1-19, Jet Propulsion Laboratory, Pasadena, California, 1981-1987.

[4] *DSN: Deep Space Network*, JPL Fact Sheet, Jet Propulsion Laboratory, Pasadena, California, November 6, 1986.

[5] *Voyager Bulletin*, JPL 410-15, status reports 67-84, Jet Propulsion Laboratory, Pasadena, California, 1981-1989.

[6] *Space Exploration: Cost, Schedule and Performance of NASA's Magellan Mission to Venus*, GAO/NSIAD-88-130FS, U.S. General Accounting Office, Washington, D.C., May 27, 1988.

[7] V-Gram, *Magellan Project*, JPL 410-11, issues 6-14, Jet Propulsion Laboratory, Pasadena, California, 1986-1989.

[8] JPL Office of Space Science, *Project Management Report (1977)* and *Project Management Report (1986)*, Agency Code 802, Jet Propulsion Laboratory, Pasadena, California, 1977 and 1986.

[9] B. Boehm, *Software Engineering Economics*, Englewood Cliffs, New Jersey: Prentice Hall, 1981.

[10] J. T. McClave and P. G. Benson, *Statistics* (4th ed.), New York: Macmillan Publishing Co., 1988.

Table 1. DSN preparation costs summary by fiscal year (1987 $K; FY'79–FY'88)

Mission	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	Total	Average	Standard deviation	Max. (Y)	Min. (Y)	Range
Galileo	1,052	3,822	4,352	4,875	7,010	6,833	8,012	8,275	1,278	1,677	47,186	4,719	2,673	8,275 ('86)	1,052 ('79)	7,223
Voyager (Uranus)				4,482	13,124	8,339	6,211	3,705			35,861	7,172	3,774	13,124 ('83)	3,705 ('86)	9,419
Voyager (Neptune)							4,984	7,308	12,328	11,429	36,049	8,012	3,462	12,328 ('87)	4,984 ('85)	7,344
Magellan							88	6,823	11,824	13,788	35,523	8,131	6,111	13,788 ('88)	88 ('85)	13,700

Table 2. Galileo preparation costs versus Galileo model (1987 $K; FY'79–FY'88)

	FY'79	FY'80	FY'81	FY'82	FY'83	FY'84	FY'85	FY'86	FY'87	FY'88	Total	Average	Standard deviation	Max.	Min.	Range
Costs (C)	1,052	3,822	4,352	4,875	7,010	6,833	8,012	8,275	1,278	1,677	47,186	4,719	2,673	8,275	1,052	7,223
Model (M)	1,939	3,626	4,863	5,543	5,664	5,312	4,632	3,783	2,908	2,111	40,381	4,038	1,382	5,664	1,939	3,725
Δ(M−C)	887	−196	511	668	−1,346	−1,521	−3,380	−4,429	1,630	434	−6,805	−681				

Table 3. DSN actual preparation costs and composite model costs (1987 $K)

Actual	Yr. 1	Yr. 2	Yr. 3	Yr. 4	Yr. 5	Yr. 6	Yr. 7	Yr. 8	Yr. 9	Yr. 10	Total cost	Avg. cost/yr.
Galileo	1,052	3,822	4,352	4,875	7,010	6,833	8,012	8,275	1,278	1,677	47,186	4,719
Voyager (Uranus)	4,482	13,124	8,339	6,211	3,705						35,861	7,172
Voyager (Neptune)	4,984	7,308	12,328	11,429							36,049	9,012
Magellan	88	6,823	11,824	13,788							32,523	8,131
Model (composite)	**Yr. 1**	**Yr. 2**	**Yr. 3**	**Yr. 4**	**Yr. 5**	**Yr. 6**	**Yr. 7**	**Yr. 8**	**Yr. 9**	**Yr. 10**	**Total cost**	**Avg. cost/yr.**
Galileo	3,502	6,381	8,193	8,787	8,302	7,076	5,510	3,949	2,619	1,611	55,930	5,593
Voyager (Uranus)	3,502	6,381	8,193	8,787	8,302						35,165	7,033
Voyager (Neptune)	3,502	6,381	8,193	8,787							26,863	6,716
Magellan	3,502	6,381	8,193	8,787							26,863	6,716

Table 4. Average of four projects' preparation costs versus composite model preparation costs (1987 $K; FY'79–FY'88)

	FY'79	FY'80	FY'81	FY'82	FY'83	FY'84	FY'85	FY'86	FY'87	FY'88	Total	Average	Standard deviation	Max.	Min.	Range
Costs (C)	2,652	7,769	9,211	9,076	5,358	6,833	8,012	8,275	1,278	1,677	60,141	6,014	3,081	9,211	1,278	7,933
Model (M)	3,502	6,381	8,193	8,787	8,302	7,076	5,510	3,949	2,619	1,611	55,933	5,593	2,558	8,787	1,611	7,176
Δ(M−C)	850	−1,388	−1,018	−289	2,944	243	−2,502	−4,326	1,341	−66	−4,208	−421				

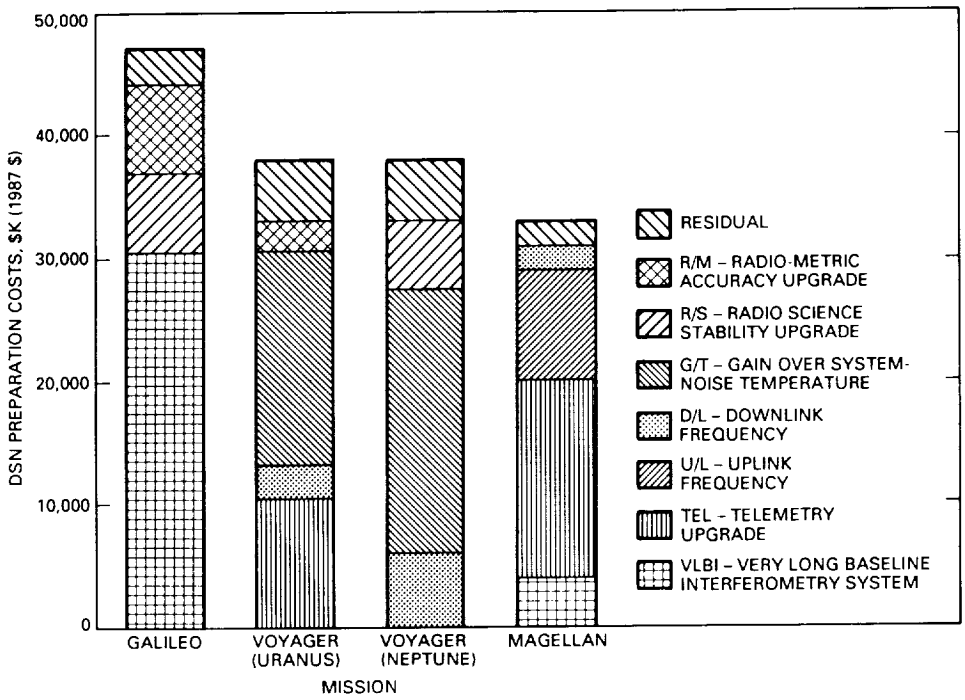


Fig. 1. DSN preparation costs by category.

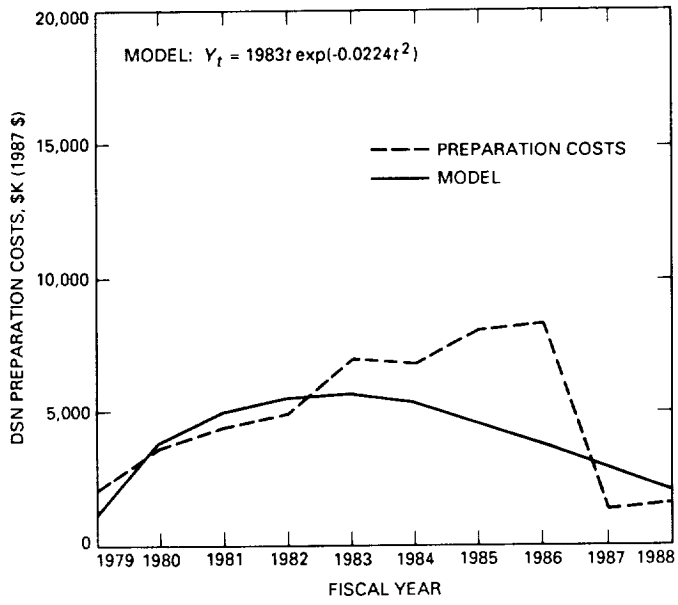


Fig. 2. Actual Galileo preparation costs versus costs predicted by Galileo model (fiscal years 1979–1988).

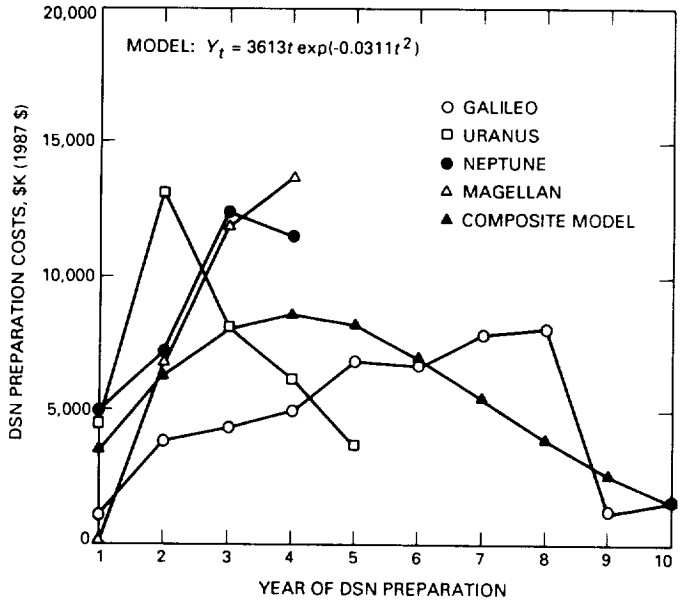


Fig. 3. Actual DSN preparation costs versus costs predicted by composite model.

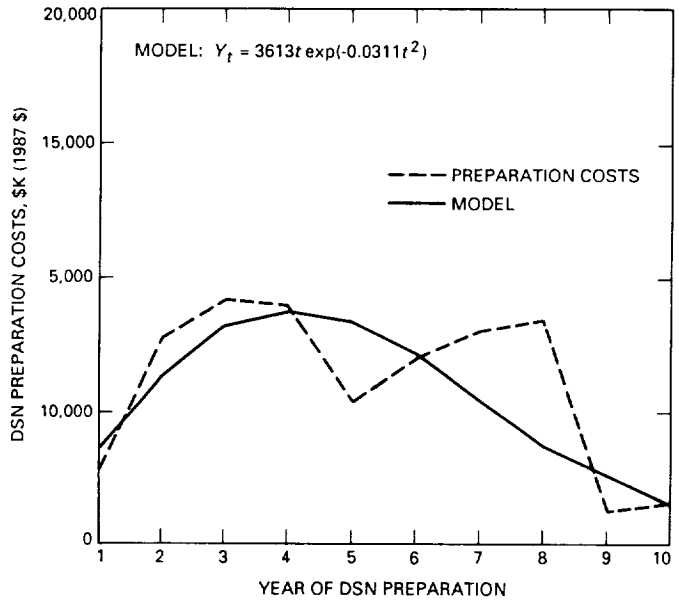


Fig. 4. Average DSN preparation costs for four projects versus costs predicted by composite model.

Rings of Earth Detected by Orbital Debris Radar

R. Goldstein
Telecommunications Science and Engineering Division

L. Randolph
TDA Engineering Office

Small particles moving at an orbital velocity of 7.6 kilometers per second can present a considerable hazard to human activity in space. For astronauts outside of the protective shielding of their space vehicles, such particles can be lethal. The powerful radar at NASA's Goldstone Deep Space Communications Complex has been used to monitor such orbital debris. This radar can detect metallic objects as small as 1.8 mm in diameter at 600 km altitude. The results of the preliminary survey show a flux (at 600 km altitude) of 6.4 objects per square kilometer per day of equivalent size of 1.8 mm or larger. Forty percent of the observed particles appear to be concentrated into two orbits. An orbital ring with the same inclination as the radar (35.1 degrees) is suggested. However, an orbital band with a much higher inclination (66 degrees) is also a possibility.

I. Orbital Debris

In support of Space Station Freedom's shielding design, NASA, through its Office of Space Tracking and Data Systems, has been obtaining orbital debris information from various sources. To augment these data, two experiments have been conducted by JPL to obtain information on debris less than 10 cm in diameter. The first experiment was conducted by Tommy Thompson of JPL and Donald Campbell of Cornell University at the Arecibo Observatory in Puerto Rico, and was able to detect debris down to 5 mm in diameter. The second experiment, which is described here, utilized a radar of a shorter wavelength, which can thereby detect smaller particles.

II. The Radar

The radar used for this survey is at the Goldstone Deep Space Communications Complex. An X-band (8.5 GHz) radar transmitter connected to the 70-m dish antenna, DSS 14, was used to illuminate the orbital particles with 3-ms pulses (with an 8-ms repetition period) of otherwise unmodulated microwave energy. A receiver was connected

to a smaller 26-m antenna, DSS 13, located 21.6 km from the transmitter. The antennas were pointed so that the beams intersected 600 km above the midpoint between the two antennas. The radar parameters were:

Transmitter power = 340 kW (average)

Transmitter antenna gain = 2.3×10^7 (over isotropic)

Receiver antenna area = 191 m^2 (effective)

System noise temperature = 24.3 K

Wavelength = 3.5 cm

Figure 1 shows the geometry of the experiment. The cross-sectional area of the antenna beam intersection was 11.6 km^2. Particles in near circular orbits at 600 km altitude take 51 ms to traverse the beam intersection.

III. Observations

Two- to five-hour observation periods were scheduled on an irregular basis from March 22 to October 16, 1989. Altogether, 48 hours of data were collected over 15 separate days or nights.

IV. Signal Processing

Received signals were filtered to a bandwidth of 10 kHz and stored on digital magnetic tape. Subsequently, blocks of 3-ms duration were Fourier transformed and accumulated over 48 ms. The resulting power spectra were compared to a threshold to decide if there had been a "hit" or if there were only noise in the data.

The threshold was determined from a chi-squared distribution of 12 degrees of freedom such that, on average, there would be one false alarm per 5 hours (375,000 spectra). The resulting signal-to-noise ratio (3.87) corresponds to a radar cross section of 5.8×10^{-3} mm^2, which was chosen as the detection limit. At this wavelength, a metallic sphere of 1.8-mm diameter has this radar cross section.

The number of reported events was reduced by the expected number of false alarms. Thus, the reported flux is a lower limit; there must have been undetected small particles.

V. Data

The spectrograms for a representative event are shown in Fig. 2. Five spectrograms are plotted in a time sequence separated by 48 ms (the approximate time within the beam intersection). Line-of-sight velocity is the abscissa; relative power density is the ordinate. The line-of-sight velocity component of ± 88 m/sec corresponds to a frequency range of ±5 kHz. The signal-to-noise ratio for this event was 6.5, somewhat over the threshold and corresponding to a metallic sphere of 1.9-mm diameter.

Figure 3 is a cumulative distribution of all the particles detected in the survey, plotted according to radar cross section. Along the upper abscissa is the size of the equivalent metallic sphere of the same radar cross section. Note that the actual particles can be larger than those indicated (i.e., nonmetallic particles).

VI. Swarms

For most observation days, one or two events per hour were recorded. During some hours, however, the rate was as high as 15. What happened on those days to produce swarms of events?

Figure 4 gives the results of one such day. The time for each event is plotted against the line-of-sight velocity. The velocity is measured from the spectrogram and represents the Doppler shift generated by the (small) radial velocity of the object when it crossed the antenna beam. The signal-to-noise ratio of each event is marked on the figure. Within the accuracy of the velocity estimates, one straight line goes through most of the points. These separate events are, therefore, strongly related; they must have a common origin. However, during the hour-long lifetime of such a swarm, the Earth has rotated 15 degrees. How is it possible for this family of objects to remain so long in an antenna beam of only 315-m width at that altitude?

Figure 5 presents one answer: the objects are in a common orbit (a ring) that has the same inclination as the latitude of the radar. Then, the Earth's rotation carries the radar in a direction that is parallel to the putative ring. Such an hypothesis should be easy to test. All one needs to do is to repeat the experiment the next sidereal day, when the radar has rotated back under the ring. Unfortunately, the observation schedule has so far been too irregular, averaging only two tracks per month.

During the survey, six occasions were found where the velocity–time plots could be essentially connected by one straight line. For each of these occasions, the time of the zero crossing of the straight line is plotted against the day of the year on which it occurred. The result is given in Fig. 6.

A single straight line connects five of the six points. This indicates that the same orbit was observed five times. The slope of the line represents the precession of the orbit, which was 3.0 degrees per day. The orbit is prograde. The sixth point of Fig. 6 represents a second orbit; one observed only once. The velocity–time line for that swarm had the opposite slope from the other five.

The observed precession rate seems too low for an orbit of 35-degree inclination. It matches, instead, what one would expect for 66 degrees.

A second possibility is that of a satellite at the higher inclination that has broken up in orbit. The debris might then have spread into a band that would be in the Goldstone beam for the required hour.

The breakup must have been rather violent, since so many small particles were produced. Such a process would tend to randomize the Doppler shifts, which is contrary to the linear relationship which was found.

The dotted circle on Fig. 6 represents one day when observations were made at the correct time to see the swarm, but no swarm was seen. One that day, the receiver's circular polarization was reversed. Large, irregular objects and flake or wire shapes can reverse the sense of circular polarization. The observed swarm was not composed of these.

VII. Conclusions

From the observed average flux of 6.4 events/km^2/day, the expected time for a human-sized object to be hit is 214 years. The danger, however, appears to depend on location. Within the ring volume, the expected time drops to 59 years. And at the convergence point (if it exists) of the orbital band, the expected time is only 1.5 months. In any case, the situation is continually deteriorating as more and more debris is deposited into orbit. Furthermore, there is an undetermined, but apparently larger, flux of particles smaller than 1.8 mm that could pose a hazard.

Acknowledgments

The authors thank Don Kessler for helpful discussions about orbital dynamics, and the Goldstone crew for the many hours spent in preparing and running this experiment.

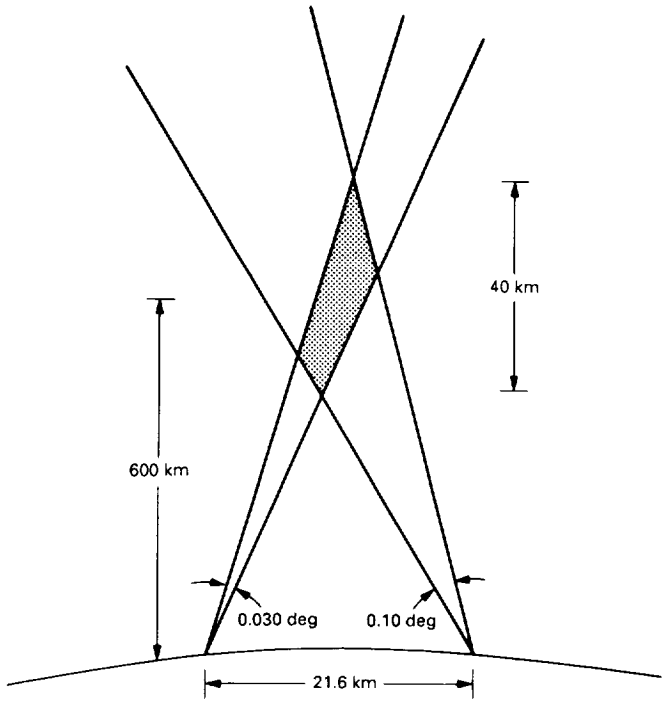


Fig. 1. The intersection of the two antenna beams. The common area of 11.6 km^2 was centered at an altitude of 600 km.

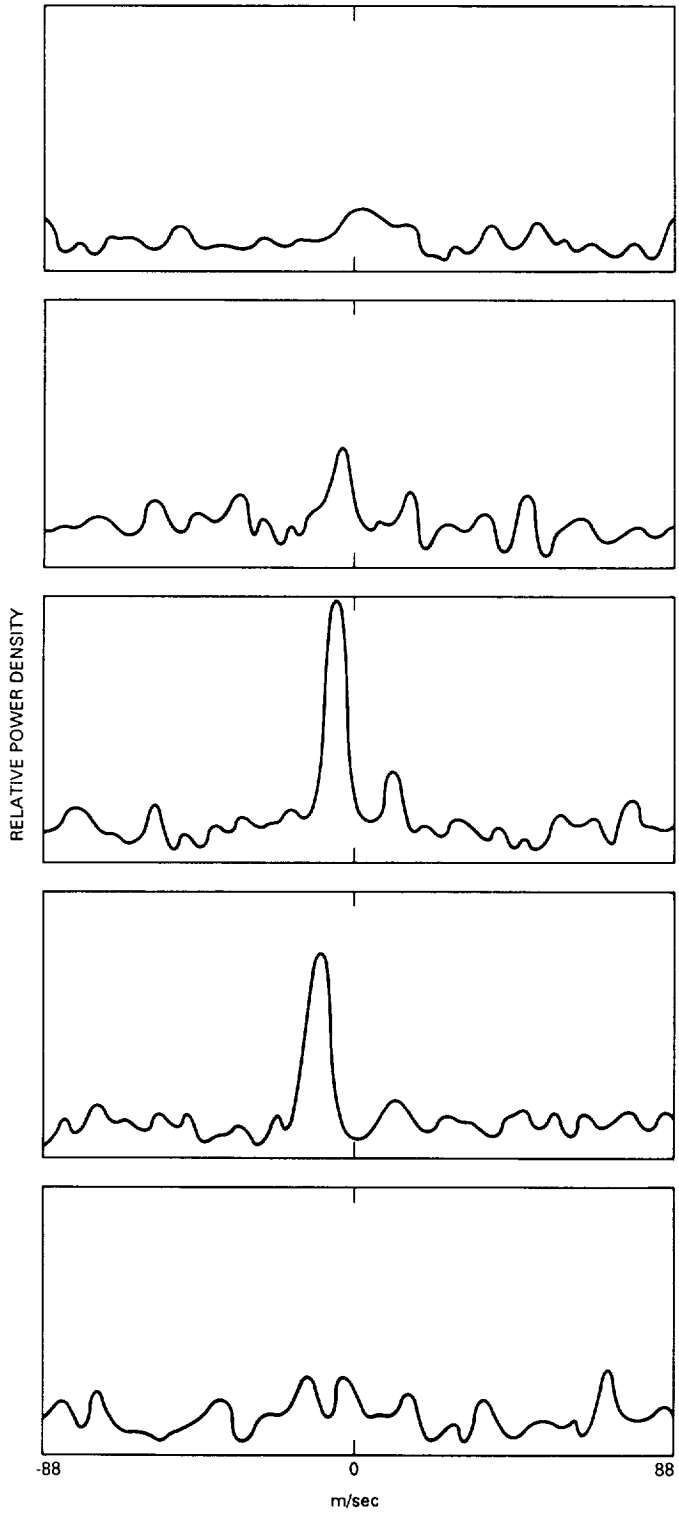


Fig. 2. Spectrograms separated in time by 48 ms. An object of equivalent diameter of 1.9 mm appears to have passed through the intersection of the two antenna beams.

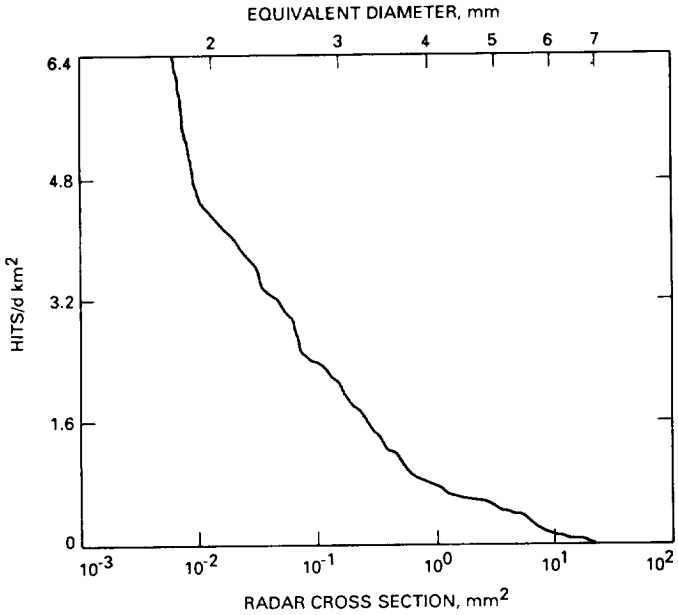


Fig. 3. Cumulative distribution for the flux of orbital debris. For each radar cross section, the plot gives the flux of particles of that cross section or larger.

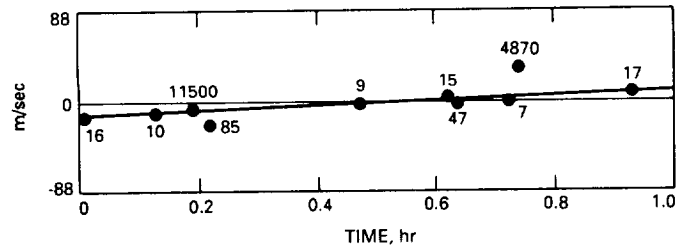


Fig. 4. Record of events during one hour on September 4, 1989. One straight line fits the data so well that the events must be related. Numbers indicate the signal-to-noise ratio.

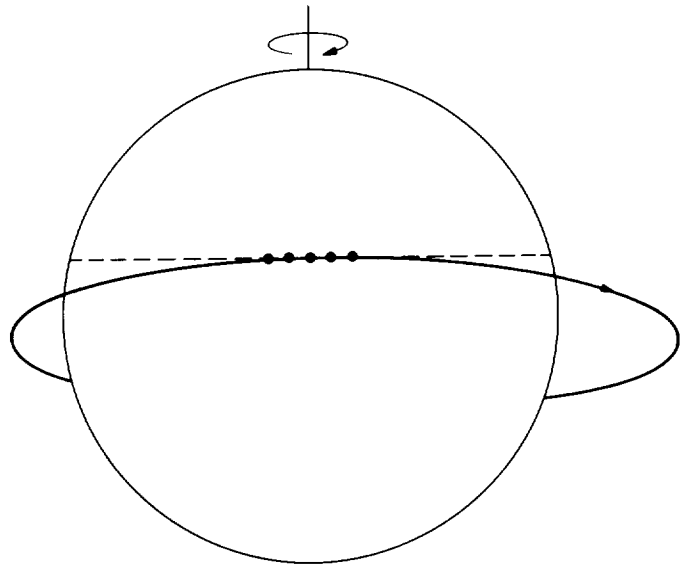


Fig. 5. Diagram showing how an antenna fixed to the rotating Earth can illuminate an orbital ring for an extended period of time.

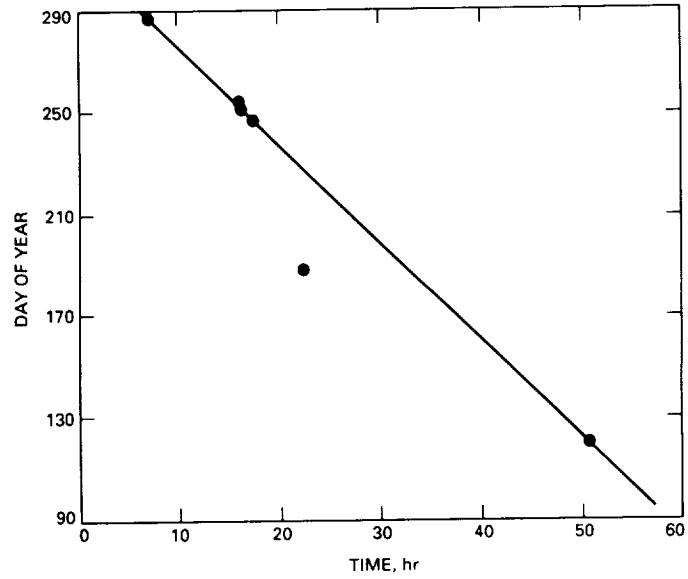


Fig. 6. Day-of-year versus time-of-day plot shows that five of the six observed swarms were repeat observations of the same orbital ring, precessed by 3.0 degrees per day.

Errata

The following tables were inadvertently left out of W. M. Folkner and M. H. Finger's article, "Photon Statistical Limitations for Daytime Optical Tracking," that appeared in the *Telecommunications and Data Acquisition Progress Report 42-99*, vol. July–September 1989, November 15, 1989:

Table 1. Characteristics of nominal astrometric tracking system

Receiver characteristics	
Aperture diameter d_r, m	1.0
Filter bandwidth $\Delta\lambda$, nm	100.0
Atmospheric transmission η_a	0.5
Obscuration factor η_{ra}	0.9
Optics efficiency η_{ro}	0.8
Filter transmission η_e	0.8
Detector efficiency η_d	0.5
Transmitter characteristics	
Laser power P, W	2.00
Laser wavelength λ, nm	532.00
Transmitter diameter d_t, m	0.30
Distance from Earth, au	10.00
Obscuration factor η_{ta}	0.71
Pointing efficiency η_{tp}	0.84
Optics efficiency η_{to}	0.65

Table 2. Integration times needed for photon statistical error to reach a specified angular accuracy, considering a star of magnitude m = 8 and a 2-W laser on the spacecraft. Several different filter bandpasses are considered. Note that the 0.03-nm filter has a different transmission efficiency.

Source	Star	Star	S/C	S/C	S/C
m_v or power	8	8	2 W	2 W	2 W
Bandpass, nm	100	1.0	100	1.0	0.03
Count rate, photons/s	6.8×10^5	6.8×10^3	2.6×10^4	2.6×10^4	2.6×10^4
Background rate, photons/s	2.6×10^9	2.6×10^7	2.6×10^9	2.6×10^7	3.8×10^5
T for 50 nrad, s (no background)	0.0099	0.99	0.26	0.26	0.52
T for 5 nrad, s (no background)	0.99	99	26	26	52
T for 50 nrad, s (with background)	38	3.8×10^5	2.6×10^4	260	15
T for 5 nrad, s (with background)	3.8×10^3	3.8×10^7	2.6×10^6	2.6×10^4	1.5×10^3

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No. TDA PR 42-101	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle The Telecommunications and Data Acquisition Progress Report 42-		5. Report Date May 15, 1990 6. Performing Organization Code
7. Author(s) Editor: E.C. Posner		8. Performing Organization Report No TDA PR 42-101
9. Performing Organization Name and Address JET PROPULSION LABORATORY California Institute of Technology 4800 Oak Grove Drive Pasadena, California 91109		10. Work Unit No. 11. Contract or Grant No. NAS7-918 13. Type of Report and Period Covered TDA Progress Report January-March 1990
12. Sponsoring Agency Name and Address NATIONAL AERONAUTICS AND SPACE ADMINISTRATION Washington, D.C. 20546		14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

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

In the search for extraterrestrial intelligence, 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, with support by the Office of Space Operations for station support time. Finally, tasks funded under the JPL Director's Discretionary Fund and the Caltech President's Fund which involve the TDA Office are included.

17. Key Words (Selected by Author(s))	18. Distribution Statement Unlimited/Unclassified		
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages	22. Price